Métodos Assintóticos em Estatística: Fundamentos e Aplicações

José Galvão Leite & Julio da Motta Singer

SÃO PAULO, Julho de 1990

Sumário

Prefácio

Estas notas são baseadas em cursos ministrados no Instituto de Matemática e Estatística da USP pelos professores Pranab K. Sen e José G. Leite em 1986 e 1989, respectivamente, e tem como objetivo apresentar de forma introdutória alguns fundamentos da teoria assintótica em Estatística. Procuramos manter as demonstrações num nível adequado para profissionais e alunos sem uma formação matemática sofisticada, e ilustrar os resultados através da aplicação em exemplos bastante freqüentes na prática. Obviamente, dadas as limitações de tempo e espaço, não podemos, nem pretendemos, abranger toda a gama de resultados disponíveis sobre o assunto. Na realidade, estas notas deveriam servir como material introdutório para a leitura de alguns textos clássicos nessa área, que listamos no Apêndice B. Algumas referências bibliográficas mais específicas, e cuja leitura requer um maior nível de sofisticação matemática estão apresentadas no Apêndice C.

Gostaríamos de agradecer aos organizadores do 9º SINAPE pela edição deste texto e pela oportunidade de apresentar o material na forma de um minicurso; também gostaríamos de agradecer aos colegas e alunos que nos ajudaram durante este empreendimento, dentre os quais destacamos Fábio Prates Machado e Pilar Loreto I. Zuazola. Finalmente gostaríamos de assumir total responsabilidade pelas possíveis imperfeições e de solicitar aos leitores que nos apressentassem críticas e sugestões para uma futura revisão.

José Galvão Leite

Julio da Motta Singer

São Paulo, julho de 1990.

Capítulo 1

Introdução

1.1 Apresentação e Motivação

O objetivo da inferência estatística é obter conclusões sobre algumas características de um conjunto de interesse, denominado *população*, com base na informação oriunda de um conjunto de dados disponíveis, denominado *amostra*. A base para o estabelecimento dessas conclusões são certos *modelos probabilísticos*, em relação aos quais as questões de interesse são especificadas. Em geral, vários modelos com diferentes níveis de complexidade podem ser propostos para o mesmo problema e a adoção de um ou outro depende, não só do grau de conhecimento que temos sobre a característica que está sendo investigada, como também dos objetivos do estudo. Examinemos um exemplo.

Exemplo 1.1.1 *Consideremos o problema de comparar as distribuições de altura de duas populações de adultos, com base em amostras aleatórias de tamanhos n_1 e n_2. Sejam F_X e F_Y as funções distribuição associadas às alturas dos indivíduos da primeira e da segunda população, respectivamente. Alguns modelos probabilísticos são:*

a) F_X e F_Y são Normais com médias μ_X e μ_Y e variância (comum) σ^2; as observações são as alturas $X_1, \ldots, X_{n_1}$ dos indivíduos da amostra da primeira população e as alturas $Y_1, \ldots, Y_{n_2}$ dos indivíduos da amostra da segunda população;

b) F_X e F_Y são distribuições contínuas e as observações são como em `1a)`*;*

c) F_X e F_Y são distribuições quaisquer e as observações correspondem ao número de indivíduos de cada uma das duas amostras que pertencem a cada um de r intervalos de altura.

Notemos que embora os modelos `1a)` e `1b)` sejam bastante utilizados para descrever problemas com estrutura semelhante, talvez o modelo `1c)` seja mais realista, pois só é possível medir alturas dentro de um certo nível de precisão (como por exemplo, 1 cm ou 1 mm).

Escolhido um modelo probabilístico para o problema em estudo, a análise estatística consiste em verificar sua adequabilidade (que, em geral, é uma tarefa difícil), estimar seus parâmetros e testar hipóteses sobre eles. As técnicas envolvidas são, muitas vezes, baseadas em critérios heurísticos (como Máxima Verossimilhança, Mínimos Quadrados, etc. . .) e grande parte do trabalho de pesquisa para sua efetiva aplicação prática está concentrada no estudo de suas propriedades estatísticas.

Por exemplo, sob o modelo `1a)`, um teste *exato* para a hipótese $\mathcal{H} : \mu_X = \mu_Y$ contra a alternativa $\mathcal{A} : \mu_X > \mu_Y$ pode ser realizado através da estatística $t = \frac{(\overline{X}-\overline{Y})}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ onde $\overline{X} = n_1^{-1} \sum_{i=1}^{n_1} X_i$, $\overline{Y} = n_2^{-1} \sum_{i=1}^{n_2} Y_i$ e $s^2 = \{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2\}/(n_1 + n_2 - 2)$ que obedece a uma distribuição t com $n_1 + n_2 - 2$ graus de liberdade e parâmetro de não-centralidade $\delta = (\mu_X - \mu_Y)/\sigma$.

Sob o modelo `1b)`, um teste *exato* para a hipótese $\mathcal{H} : F_X = F_Y$ contra a alternativa $\mathcal{A} : F_X(t) \leq F_Y(t), \ \forall \ t \in \mathbb{R}$, com $F_X \neq F_Y$ pode ser realizado através da estatística $W = \sum_{i=n_1+1}^{n_1+n_2} R_i$ onde $R_i, \ i = 1, \ldots, n_1 + n_2$ representam os postos correspondentes às observações $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ que, sob a hipótese nula, obedece à distribuição de Wilcoxon. Embora os percentis dessa distribuição possam ser calculados, sua obtenção é computacionalmente difícil para valores grandes de n_1 e n_2. No entanto, nessa situação, a distribuição da estatística $U = W - n_2(n_2 + 1)/2$ pode ser *aproximada* por uma distribuição Normal com média $n_1n_2/2$ e variância $n_1n_2(n_1 + n_2 + 1)/12$.

Sob o modelo `1c)`, a hipótese de *homogeneidade* das duas distribuições relativamente às proporções de observações nos r intervalos de altura pode ser testada através da estatística $Q = \frac{\sum_{i=1}^{2}\sum_{j=1}^{r}\{n_{ij}-n_{i.}n_{.j}/n_{..}\}^2}{n_{i.}n_{.j}/n_{..}}$ onde n_{ij} representa o número de observações da $i-$ésima população no $j-$ésimo intervalo de altura, $n_{i.} = \sum_{j=1}^{r} n_{ij}, \ n_{.j} = \sum_{i=1}^{2} n_{ij}$, e $n_{..} = \sum_{i=1}^{2}\sum_{j=1}^{r} n_{ij}, \ i = 1, 2; \ j = 1, \ldots, r$. A distribuição *exata* da estatística Q é intratável do ponto de vista prático, embora teoricamente ela possa ser obtida. No entanto, para amostras *suficientemente grandes* (isto é , $n_{1.}$ e $n_{2.}$ suficientemente grandes) a distribuição de Q pode ser *aproximada* por uma distribuição χ^2 com $r - 1$ graus de liberdade se a hipótese de homogeneidade for verdadeira.

A partir dessas considerações podemos notar que mesmo em um problema extremamente simples como aquele descrito no Exemplo 1.1.1, a utilização de distribuições aproximadas pode ser necessária para a análise estatística. Examinemos agora uma situação um pouco mais complexa:

Exemplo 1.1.2 *Consideremos o problema de estimar o tamanho N de uma população animal fechada, isto é , onde não haja nascimento (ou imigração) nem morte (ou emigração) através do* método da captura-recaptura. *Nesse contexto, o* procedimento da amostragem inversa *consiste em capturar, ao acaso e sem reposição, n_1 animais dessa população numa primeira amostragem, marcá-los e devolvê-los à população; após um certo tempo, selecionar uma segunda amostra casual* com reposição *até que sejam obtidos exatamente m animais marcados, onde m é um inteiro positivo, fixado arbitrariamente.*

Seja X_j o número de animais capturados entre o $(j-1)-$ésimo animal marcado (exclusive) e o $j-$ésimo animal marcado (inclusive), $j = 1, \ldots, m$. Então $X_1, \ldots, X_m$ são variáveis aleatórias independentes e identicamente distribuídas (i.i.d), com distribuição Geométrica de parâmetro n_1/N, isto é :

$$\mathbb{P}\{X_1 = x\} = (n_1/N)(1 - n_1/N)^{x-1} \quad , \; x = 1, 2, \ldots \tag{1.1.1}$$

Consequentemente, o tamanho da segunda amostra, isto é , o número de animais capturados até que sejam obtidos exatamente m marcados, $X = \sum_{j=1}^{m} X_j$ tem distribuição Binomial Negativa com parâmetros m e n_1/N, ou seja:

$$\mathbb{P}\{X = x\} = \binom{x-1}{m-1}\big(\frac{n_1}{N}\big)^m\big(1 - \frac{n_1}{N}\big)^{x-m} \quad , \; x = m, m+1, \ldots \tag{1.1.2}$$

Um estimador intuitivo para N corresponde à solução $\hat{N}$ da equação obtida ao igualarmos a proporção de animais marcados na população, antes da segunda amostragem, (n_1/N) à proporção de animais marcados na segunda amostra (m/X), ou seja:

$$\frac{n_1}{N} = \frac{m}{X} \Longrightarrow \hat{N} = \frac{n_1 X}{m} \tag{1.1.3}$$

que é aproximadamente igual ao estimador de Máxima Verossimilhança (***MV*** *) (veja Exercício 1.2.1).*

A questão fundamental no Exemplo 1.1.2 é quantificar o erro que cometemos ao estimar N através de $\hat{N}$; mais especificamente, isto pode ser traduzido na construção de um intervalo de confiança para N. Em princípio as propriedades estatísticas de $\hat{N}$ podem ser obtidas de (1.1.2); no entanto essa distribuição depende do parâmetro que queremos estimar, o que impossibilita sua utilização direta para fazermos inferências *exatas* sobre N. Para contornar esse problema podemos nos valer dos métodos aproximados que são estudados neste texto. Podemos mostrar, por exemplo, através da *Lei dos Grandes Números* que à medida que m aumenta, $\hat{N}$ se aproxima de N em algum sentido que esclareceremos adiante. Também é possível mostrar, utilizando o *Teorema Limite Central* e o *Teorema de Slutsky* que, para m suficientemente grande, a distribuição de $\hat{N}$ pode ser aproximada por uma distribuição Normal com média N e variância $\hat{N}(\hat{N} - n_1)/m$, o que nos permite construir um intervalo de confiança *aproximado* para N.

Ainda com relação a esse exemplo, uma outra questão de interesse diz respeito à comparação do estimador $\hat{N}$ com o estimador de **MV**; como as propriedades estatísticas *exatas* são de difícil obtenção nos dois casos, podemos nos basear no conceito de *eficiência assintótica* para nos decidir por um ou por outro.

Neste ponto da discussão já podemos perceber que soluções aproximadas são bastante freqüentes e muitas vezes as únicas disponíveis em problemas de inferência estatística. Além disso, pudemos notar que as aproximações dependem essencialmente do tamanho das amostras consideradas. Dessa forma, podemos dizer, a grosso modo, que o tema básico deste texto é o estudo do comportamento de estimadores e estatísticas de teste para grandes amostras. Nesse contexto, convém ressaltar que o termo "*métodos assintóticos* " constante do título destas notas está relacionado com o fato de que, em geral, os resultados desta área da Estatística são válidos no limite, quando o(s) tamanho(s) da(s) amostra(s) tende(m) ao infinito. Obviamente, na prática, trabalhamos com amostras finitas e é natural que consideremos questões do seguinte tipo:

Quão grandes devem ser n_1 e n_2 no Exemplo 1.1.1 e m no Exemplo 1.1.2 para que as aproximações sejam razoáveis ?

Embora não seja possível apresentarmos respostas totalmente satisfatórias para essa questão, o *Teorema de Berry-Esséen* e as expansões de *Edgeworth* nos fornecem *taxas* ou *velocidades de convergência* que permitem quantificar essas aproximações, pelo menos em algumas situações mais simples.

A intenção básica deste texto é discutir questões como as que apresentamos acima através de vários exemplos envolvendo inferência estatística. Procuramos dar mais ênfase aos conceitos e às aplicações e, nesse sentido, deixamos de apresentar as demonstrações que necessitam de ferramentas matemáticas mais sofisticadas. Nas demais seções deste capítulo apresentamos os conceitos de ordens de magnitude, expansões de Taylor e funções características, que são fundamentais para o desenvolvimento dos capítulos seguintes. No Capítulo 2 consideramos os principais modos de convergência estocástica e as Leis dos Grandes Números, que constituem a base probabilística para os chamados métodos assintóticos em Estatística. No Capítulo 3 concentramos nossa atenção nas diferentes versões do Teorema Limite Central que é o resultado mais importante para aplicações; também consideramos alguns resultados envolvendo a convergência em distribuição de funções de estatísticas com distribuição assintótica Normal, bem como o Teorema de Berry-Esséen e as expansões de Edgeworth, que em alguns casos são convenientes para quantificar as aproximações. No Capítulo 4 estudamos o comportamento assintótico dos principais tipos de estimadores e estatísticas de teste.

1.2 Ordens de magnitude de seqüências de números reais e vetores: a notação $O(\cdot)$, $o(\cdot)$.

Como veremos oportunamente, um problema frequente no estudo dos métodos assintóticos consiste na comparação de seqüências de variáveis aleatórias. A base para o estudo desse problema é a análise comparativa do *comportamento* de duas seqüências de números reais $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$, quando n tende ao infinito. Nesta seção apresentamos alguns resultados fundamentais sobre esse tema. O caso típico é aquele onde os termos gerais das seqüências de interesse não são dados explicitamente, mas uma (algumas) de suas propriedades é (são) conhecida (s). Por exemplo, seja $a_n = b_n^2, n \geq 1$ e vamos supor que $\{b_n\}_{n\geq 1}$ seja *limitada para todo n suficientemente grande*, ou mais especificamente, que existam um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$ tal que $|b_n| \leq K,\ \forall n \geq n_0$. Nosso problema é o de deduzir propriedades da seqüência $\{a_n\}_{n\geq 1}$. Neste caso, é claro que, $|a_n| \leq K^2,\ \forall n \geq n_0$ e, consequentemente, podemos deduzir que $\{a_n\}_{n\geq 1}$ também é limitada para todo n suficientemente grande. Para estudarmos situações mais complexas convém apresentar a seguinte definição:

Definição 1.2.1 *Sejam $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$ seqüências de números reais; então diremos que*

i) $a_n = O(b_n)$ se existirem um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$ tal que $|a_n/b_n| \leq K,\ \forall n \geq n_0$;

ii) $a_n = o(b_n)$ se para todo $\epsilon > 0$ existir um número inteiro positivo $n_0 = n_0(\epsilon)$ tal que $|a_n/b_n| < \epsilon,\ \forall n \geq n_0$.

Em outras palavras, diremos que $a_n = O(b_n)$ se a razão $|a_n/b_n|$ for limitada para todo n suficientemente grande e que $a_n = o(b_n)$ se $a_n/b_n \longrightarrow 0$ quando $n \longrightarrow \infty$. Em particular, $a_n = O(1)$ se existir um número real $K > 0$ tal que $|a_n| \leq K$ para todo n suficientemente grande e $a_n = o(1)$ se $a_n \longrightarrow 0$ quando $n \longrightarrow \infty$. Essencialmente, a Definição `1.2.1` está relacionada com a comparação das *ordens de magnitude* (ou de grandeza) de $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$. A grosso modo, afirmar que $a_n = O(b_n)$ corresponde a dizer que a ordem de magnitude de $\{a_n\}_{n\geq 1}$ é , *no máximo, igual* à de $\{b_n\}_{n\geq 1}$ para todo n suficientemente grande; analogamente, afirmar que $a_n = o(b_n)$ corresponde a dizer que a ordem de magnitude de $\{a_n\}_{n\geq 1}$ é *menor* que a de $\{b_n\}_{n\geq 1}$, para todo n suficientemente grande.

Exemplo 1.2.1 *Utilizando a Definição 1.2.1 note que:*

i) $n = o(n^2)$ pois $\frac{n}{n^2} = \frac{1}{n} \longrightarrow 0$, quando $n \longrightarrow \infty$.

ii) $n^{-1} = o(1)$ *pois* $\frac{n^{-1}}{1} = \frac{1}{n} \longrightarrow 0,$ *quando* $n \longrightarrow \infty$.

iii) $10n^2 + n = O(n^2)$ *pois* $\frac{10n^2+n}{n^2} = 10 + \frac{1}{n} \leq 11, \ \forall n \geq 1$.

iv) $n^2 = O(6n^2 + n)$ *pois* $\frac{n^2}{6n^2+n} < \frac{n^2}{6n^2} \leq \frac{1}{6}, \ \forall n \geq 1$.

v) $n = O(n^2)$ *pois* $\frac{n}{n^2} = \frac{1}{n} \leq 1, \ \forall n \geq 1$.

vi) $3n = o(n^2)$ *pois* $\frac{3n}{n^2} = \frac{3}{n} \longrightarrow 0,$ *quando* $n \longrightarrow \infty$.

No teorema seguinte apresentamos alguns resultados envolvendo a notação $O(\cdot)$ e $o(\cdot)$ que são bastante úteis em aplicações.

Teorema 1.2.1 *Sejam* $\{a_n\}_{n\geq 1}, \ \{b_n\}_{n\geq 1}, \ \{c_n\}_{n\geq 1},$ *e* $\{d_n\}_{n\geq 1}$ *seqüências de números reais;*

i) *se* $a_n = o(b_n)$, *então* $a_n = O(b_n)$;

ii) *se* $a_n = O(b_n)$ *e* $c_n = O(d_n)$, *então:*

  a) $a_n c_n = O(b_n d_n)$;

  b) $|a_n|^s = O(|b_n|^s)$ *para todo* $s > 0$;

  c) $a_n + c_n = O(\max\{|b_n|, |d_n|\})$;

iii) *se* $a_n = o(b_n)$ *e* $c_n = o(d_n)$, *então:*

  a) $a_n c_n = o(b_n d_n)$;

  b) $|a_n|^s = o(|b_n|^s)$ *para todo* $s > 0$;

  c) $a_n + c_n = o(\max\{|b_n|, |d_n|\})$;

iv) *se* $a_n = O(b_n)$ *e* $c_n = o(d_n)$, *então* $a_n c_n = o(b_n d_n)$;

v) *se* $a_n = O(b_n)$ *e* $b_n = o(c_n)$, *então* $a_n = o(c_n)$.

Demonstração: Provaremos apenas os itens `ii`), `iv`) e `v`); os demais ficam como exercício para o leitor.

`ii`) Pela Definição 1.2.1, existem números reais $K_1 > 0$ e $K_2 > 0$ e números inteiros positivos $n_1 = n_1(K_1)$ e $n_2 = n_2(K_2)$ tais que: $|a_n/b_n| \leq K_1, \ \forall n \geq n_1$ e $|c_n/d_n| \leq K_2, \ \forall n \geq n_2$. Logo, $|a_n c_n / b_n d_n| \leq K_1 K_2, \ \forall n \geq \max\{n_1, n_2\}$, o que prova `a`). Por outro lado, qualquer que seja $s > 0$, temos: $| \ a_n \ |^s \leq K_1^s \ | \ b_n \ |^s, \forall n \geq n_1$, o que prova `b`) . Finalmente, seja $m_n = \max\{|b_n|, |d_n|\}, \ n \geq 1$; então $|\frac{a_n + c_n}{m_n}| \leq \frac{|a_n|}{m_n} + \frac{|c_n|}{m_n} \leq \frac{|a_n|}{|b_n|} + \frac{|c_n|}{|d_n|} \leq K_1 + K_2, \ \forall n \geq \max\{n_1, n_2\}$, o que demonstra `c`) .

`iv`) Pela Definição 1.2.1 existem um número real $K > 0$ e um número inteiro positivo $n_1 = n_1(K)$ tais que $|a_n/b_n| \leq K, \ \forall n \geq n_1$; além disso, para todo $\epsilon > 0$ existe um número inteiro positivo $n_2 = n_2(\epsilon, K)$ tal que $|c_n/d_n| < \frac{\epsilon}{K}$, $\forall n \geq n_2$. Logo $|a_n c_n/b_n d_n| \leq K|c_n/d_n| < K\epsilon/K = \epsilon$, $\forall n \geq \max\{n_1, n_2\}$, e o resultado `iv )` segue.

`v )` Pela Definição 1.2.1 existem um número real $K > 0$ e um número inteiro positivo $n_1 = n_1(K)$ tais que $| \, a_n/b_n \, | \leq K$, $\forall n \geq n_1$. Por outro lado, qualquer que seja $\epsilon > 0$, existe um número inteiro positivo $n_2 = n_2(\epsilon, K)$ tal que $| \, b_n/c_n \, | < \epsilon/K$, $\forall n \geq n_2$. Logo: $| \, a_n/c_n \, | = | \, a_n/b_n \, || \, b_n/c_n \, | \leq K| \, b_n/c_n \, | < K\epsilon/K = \epsilon$, $\forall n \geq \max\{n_1, n_2\}$ o que completa a demonstração.

Outras propriedades bastante úteis são apresentadas no final desta seção sob a forma de exercícios.

Exemplo 1.2.2 *Utilizando os resultados do Teorema 1.2.1 podemos verificar que se* $a_n = o(1) + O(n^{-\frac{1}{3}}) + O(n^{-2})$ *então* $a_n = o(1)$. *Nesse sentido note que se* $b_n = O(n^{-\frac{1}{3}})$ *e* $c_n = O(n^{-2})$, *pelo item* ***iic)*** *temos:* $b_n + c_n = O(\max\{n^{-\frac{1}{3}}, n^{-2}\}) = O(n^{-\frac{1}{3}})$; *fazendo* $d_n = n^{-\frac{1}{3}}$, $n \geq 1$, *e usando a Definição 1.2.1 temos* $d_n = o(1)$, *e então pelo item* ***v)*** *segue que* $a_n = o(1) + b_n + c_n = o(1) + O(o(1)) = o(1) + o(1)$; *finalmente pelo item* ***iiic)*** *temos* $a_n = o(1) + o(1) = o(1)$.

Com a finalidade de estender esses conceitos para o caso vetorial, consideremos a seguinte definição:

Definição 1.2.2 *Sejam* $\{\boldsymbol{a}_n\}_{n\geq 1}$ *uma seqüência de vetores de* $\mathbb{R}^p, (p \geq 2)$ *e* $\{b_n\}_{n\geq 1}$ *uma seqüência de números reais; então diremos que:*

i) $\boldsymbol{a}_n = O(b_n)$ *se* $|| \, \boldsymbol{a}_n \, || = O(b_n)$;

ii) $\boldsymbol{a}_n = o(b_n)$ *se* $|| \, \boldsymbol{a}_n \, || = o(b_n)$.

A seguir apresentamos um resultado de grande importância prática porque nos permite reduzir o caso vetorial ao caso unidimensional, onde, em geral, as demonstrações são mais simples.

Teorema 1.2.2 *Sejam* $\{\boldsymbol{a}_n\}_{n\geq 1}$ *uma seqüência de vetores de* $\mathbb{R}^p$ *e* $\{b_n\}_{n\geq 1}$ *uma seqüência de números reais;*

i) $\boldsymbol{a}_n = O(b_n)$ *se e somente se* $a_{nj} = O(b_n)$, $j = 1, \ldots, p$;

ii) $\boldsymbol{a}_n = o(b_n)$ *se e somente se* $a_{nj} = o(b_n)$, $j = 1, \ldots, p$.

Demonstração: Provaremos apenas o item `i )` ; a demonstração de `ii )` pode ser feita de forma bastante semelhante. Suponhamos que $\boldsymbol{a}_n = O(b_n)$. Da Definição 1.2.2 sabemos que existem um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$ tal que $||\boldsymbol{a}_n||/|b_n| \leq K$, $\forall n \geq n_0$. Logo, para $j = 1, 2, \ldots, p$ temos:

$$|\frac{a_{nj}}{b_n}| = \frac{|a_{nj}^2|^{\frac{1}{2}}}{|b_n|} \leq \frac{|| \, \boldsymbol{a}_n \, ||}{|b_n|} \leq K, \ \forall n \geq n_0.$$

Consequentemente $a_{nj} = O(b_n), j = 1,\ldots,p$. Reciprocamente, se $a_{nj} = O(b_n)$, $j = 1,\ldots,p$, para cada j existem um número real $K_j > 0$ e um número inteiro positivo $n_j = n_j(K_j)$ tais que $|a_{nj}/b_n| \leq K_j,\ \forall n \geq n_j$. Sejam $K = \max\{K_j, j = 1,\ldots,p\}$ e $n_0 = \max\{n_j, j = 1,\ldots,p\}$. Então

$$\frac{\|\ \boldsymbol{a}_n \ \|}{|b_n|} = \Big(\sum_{j=1}^{p} \frac{a_{nj}^2}{b_n^2}\Big)^{\frac{1}{2}} \leq \Big(\sum_{j=1}^{p} K_j^2\Big)^{\frac{1}{2}} \leq p^{\frac{1}{2}} K,\ \forall n \geq n_0$$

e, pela Definição `1.2.2` , temos $\mathbf{a}_n = O(b_n)$.

Exercícios

Exercício 1.2.1 *Seja X uma variável aleatória com distribuição de probabilidade Binomial Negativa com parâmetros m ($m \geq 1$, inteiro) e $p(0 < p < 1)$, isto é ,*

$$\mathbb{P}(X = x) = \binom{x-1}{m-1} p^m (1-p)^{x-m}, \quad x = m, m+1, \ldots \quad .$$

*i) Prove que o estimador de **MV** de p é $\frac{m}{X}$;*

*ii) utilizando o item anterior, prove que o estimador de **MV** de N (o tamanho de uma população animal para o modelo descrito no Exemplo `1.1.2`) é aproximadamente igual a $\frac{n_1 X}{m}$. Note que N é , no caso, um número* inteiro *maior ou igual a n_1.*

Exercício 1.2.2 *Prove os itens `i )` e `iii )` do Teorema 1.2.1 .*

Exercício 1.2.3 *Prove o item `ii )` do Teorema 1.2.2 .*
Sejam $\{a_n\}_{n\geq 1}$, $\{b_n\}_{n\geq 1}$ e $\{c_n\}_{n\geq 1}$ seqüências de números reais.

Exercício 1.2.4 *Prove as seguintes propriedades envolvendo as ordens de magnitude $O(\cdot)$ e $o(\cdot)$.*

i) $O(O(a_n)) = O(a_n)$;

ii) $O(o(a_n)) = o(O(a_n)) = o(o(a_n)) = o(a_n)$;

iii) $O(a_n) + O(a_n) = O(a_n) + o(a_n) = O(a_n)$;

iv) $o(a_n) + o(a_n) = o(a_n)$;

v) $\big(O(a_n)\big)^2 = O(a_n^2)$.

Exercício 1.2.5 *Prove ou dê um contra-exemplo: se $a_n = O(b_n)$, então $a_n = o(b_n)$.*

Exercício 1.2.6 *Prove que, se $a_n = O(b_n)$ e $b_n = O(c_n)$, então*

i) $-a_n = O(b_n)$;

`ii )` $a_n + b_n = O(c_n)$

Exercício 1.2.7 *Prove que* $\log n = o(n^{\alpha})$ *para todo* $\alpha > 0$.

Exercício 1.2.8 *Prove que* $1 - (1 - \frac{1}{n})(1 - \frac{2}{n}) = O(n^{-1}) = o(n^{-\frac{1}{2}})$

Exercício 1.2.9 *Suponha que* $a_n = O(b_n)$, *com* $b_n > 0$, $\forall n \geq 1$.

i) Então mostre que $n^{-1} \sum_{i=1}^{n} a_i = O(n^{-1} \sum_{i=1}^{n} b_i)$;

ii) admita também que $a_n \geq 0$, $\forall n \geq 1$; *então mostre que*

$$(\prod_{i=1}^{n} a_i)^{\frac{1}{n}} = O(\{ \prod_{i=1}^{n} b_i \}^{\frac{1}{n}}).$$

1.3 Expansões de Taylor

Uma das ferramentas utilizadas com frequência no estudo de métodos assintóticos é a aproximação de uma função real f de variável real por um polinômio. Embora a expansão de f em série de potências em torno de um ponto x_0 seja importante sob o ponto de vista teórico, o que realmente interessa nas aplicações é o estudo do resto proveniente da aproximação de f por um polinômio. Nesta seção apresentaremos alguns conceitos e resultados básicos sobre esse tópico, além de considerarmos uma extensão dos conceitos de ordens de magnitude $O(\cdot)$ e $o(\cdot)$ de seqüências de números reais para funções de variável real.

Dentre as várias maneiras de aproximarmos uma dada função f por um polinômio, aquela que mais nos interessa corresponde ao caso em que

i) f coincide com o polinômio em um determinado ponto x_0;

ii) a $k-$ésima derivada de f coincide com a $k-$ésima derivada do polinômio no ponto x_0, $k = 1, 2, \ldots, n$. Por exemplo, se f for derivável no ponto x_0, o polinômio P_1 de grau no máximo 1 tal que $P_1(x_0) = f(x_0)$ e $P_1'(x_0) = f'(x_0)$ é dado por $P_1(x) = f(x_0) + f'(x_0)(x - x_0)$. Geometricamente, P_1 corresponde à reta tangente à curva associada à função f no ponto $\big(x_0, f(x_0)\big)$. Se D_f for o domínio de f, então para todo $x \in D_f$, o resto ou erro $R_1(x)$ da aproximação de $f(x)$ por $P_1(x)$ é dado por

$$R_1(x) = f(x) - P_1(x) = f(x) - f(x_0) - f'(x_0)(x - x_0)$$

Então, para todo $x \neq x_0$, temos

$$\frac{R_1(x)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0)$$

e portanto, $\lim\limits_{x \longrightarrow x_0} \frac{R_1(x)}{x - x_0} = f'(x_0) - f'(x_0) = 0$, isto é , o resto $R_1(x)$ tende a zero *mais rapidamente* que $x - x_0$ quando $x \longrightarrow x_0$. Nesse sentido o polinômio P_1 é uma aproximação de f em torno de x_0 e é denominado *polinômio de Taylor*, de ordem 1, de f em torno de x_0. Quando $f''(x_0)$ existe, podemos aproximar f por um polinômio P_2 tal que $P_2(x_0) = f(x_0)$, $P_2'(x_0) = f'(x_0)$ e $P_2''(x_0) = f''(x_0)$, ou seja,

$$P_2(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

Nesse caso, a reta tangente associada a P_2 coincide com a reta tangente associada a f, no ponto $\big(x_0, P_2(x_0)\big) = \big(x_0, f(x_0)\big)$, bem como P_2 e f tem a mesma *curvatura* no referido ponto. Nesse sentido, P_2 é uma melhor aproximação que P_1 para a função f

no ponto x_0, e é denominado polinômio de Taylor, de ordem 2, de f em torno de x_0. Para todo $x \in D_f$ o resto da aproximação de f por P_2 é dado por:

$$R_2(x) = f(x) - P_2(x) = f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{f''(x_0)}{2}(x - x_0)^2.$$

Então, para todo $x \neq x_0$,

$$\frac{R_2(x)}{(x - x_0)^2} = \frac{f(x) - f(x_0) - f'(x_0)(x - x_0) - f''(x_0)(x - x_0)^2/2}{(x - x_0)}$$

e como f é contínua em x_0, temos $\lim\limits_{x \longrightarrow x_0} \frac{R_2(x)}{(x-x_0)^2} = \frac{0}{0}$. Pela regra de L'Hospital segue que

$$\begin{aligned}
\lim_{x \longrightarrow x_0} \frac{R_2(x)}{(x - x_0)^2} &= \lim_{x \longrightarrow x_0} \frac{f'(x) - f'(x_0) - f''(x_0)(x - x_0)}{2(x - x_0)} = \\
&= \frac{1}{2} \lim_{x \longrightarrow x_0} \Big\{ \frac{f'(x) - f'(x_0)}{x - x_0} - f''(x_0) \Big\} = \frac{1}{2} \big\{ f''(x_0) - f''(x_0) \big\} = 0
\end{aligned}$$

Consequentemente podemos concluir que $R_2(x) \longrightarrow 0$ *mais* rapidamente que $(x - x_0)^2$ quando $x \longrightarrow x_0$.

A título de ilustração, consideremos a função $f(x) = e^x$. Os polinômios de Taylor, de ordem 1 e 2 de f, em torno de zero são, respectivamente, $P_1(x) = 1 + x$ e $P_2(x) = 1 + x + x^2/2$, e estão representados na Figura `1.3.1`, onde fica clara a idéia de que a aproximação de f por P_2 em torno de zero é melhor do que a aproximação de f por P_1 em torno desse mesmo ponto.

Figura 1.3.1: Aproximações polinomiais de $f(x) = e^x$ em uma vizinhança de zero.

Nessa altura parece evidente que podemos melhorar a precisão da aproximação de uma função f, em uma vizinhança de um ponto x_0, por um polinômio P tal que $P^{(k)}(x_0) = f^{(k)}(x_0)$, para $k = 0, 1, 2, \ldots, n$, $\big(f^{(0)}(x_0) = f(x_0)\big)$, na medida em que considerarmos valores de n cada vez maiores. O polinômio P se denomina o polinômio de Taylor, de ordem n, de f em torno de x_0.

Mais precisamente, seja f uma função derivável até a ordem n em um ponto x_0. O polinômio de Taylor, de ordem n, de f em torno de x_0 é o polinômio

$$P(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \tag{1.3.1}$$

P é um polinômio de grau menor ou igual a n e satisfaz as condições

$$P(x_0) = f(x_0), \; P'(x_0) = f'(x_0), \; \ldots, \; P^{(n)}(x_0) = f^{(n)}(x_0).$$

Se $x_0 = 0, P$ também é denominado polinômio de MacLaurin, de ordem n, de f.

Exemplo 1.3.1 *Consideremos a função* $f(x) = e^x$*. Então* $f^{(k)}(x_0) = e^{x_0}$*, para todo* x_0 *real e* $k = 0, 1, 2, \ldots$*. Se* $x_0 = 0$ *o polinômio (1.3.1) pode ser expresso como*

$$P(x) \;=\; \sum_{k=0}^{n} \frac{x^k}{k!} \;=\; 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}.$$

Se $x_0 = 1$ *o polinômio (1.3.1) pode ser expresso na forma* $P(x) = \sum_{k=0}^{n} \frac{e}{k!}(x-1)^k$.

Para todo $x \in D_f$, seja $R_n(x) = f(x) - P(x)$ o resto ou o erro da aproximação de $f(x)$ por $P(x)$. Logo, se f for derivável até a ordem n em um ponto x_0, temos

$$f(x) \;=\; \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}\,(x - x_0)^k \;+\; R_n(x) \tag{1.3.2}$$

para todo $x \in D_f$. A expressão (1.3.2) é denominada fórmula de Taylor, com resto R_n, de f em torno de x_0. Nós provamos, para $n = 1$ e 2, que $R_n(x)$ tende a zero *mais rapidamente* que $(x - x_0)^n$ quando x tende à x_0. Deixamos, como exercício para o leitor, provar que este resultado vale para todo $n \geq 1$.

Como veremos nos capítulos subsequentes, muitos resultados importantes dependem de um *controle* sobre o resto da aproximação de f por P. O teorema que apresentamos a seguir tem bastante utilidade no estudo do comportamento desse resto.

Teorema 1.3.1 *Seja* f *uma função real de variável real derivável até a ordem* $n+1$ *em um intervalo* I *e sejam* $x_0,\ x \in I$*. Então existe pelo menos um ponto* c *entre* x_0 *e* x *tal que*

$$f(x) \;=\; \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}\,(x - x_0)^k \;+\; \frac{f^{(n+1)}(c)}{(n+1)!}\,(x - x_0)^{n+1} \tag{1.3.3}$$

Em outras palavras, o resto na fórmula de Taylor (1.3.2) é dado por

$$R_n(x) \;=\; \frac{f^{(n+1)}(c)}{(n+1)!}\,(x - x_0)^{n+1} \tag{1.3.4}$$

Observamos que, quando $n = 0$, este teorema é precisamente o teorema do valor médio. A prova desse teorema pode ser encontrada em Courant (1958, cap VI).

Exemplo 1.3.2 *De acordo com o Exemplo (1.3.1) o polinômio de Taylor, de ordem* n*, de* $f(x) = e^x$ *em torno de zero é*

$$P(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$$

e, pelo Teorema 1.3.1, para todo $x \neq 0$, existe um ponto c entre 0 e x tal que

$$R_n(x) = e^x - P(x) = \frac{e^c}{(n+1)!}\ x^{n+1}.$$

Logo, se $|x| < 1$ temos

$$|\ e^x - P(x)\ | \ \leq \ \frac{e}{(n+1)!}$$

e consequentemente $P(x)$ é uma aproximação para e^x com um erro menor que $\frac{e}{(n+1)!}$. Quanto maior for n, melhor será essa aproximação, pois $\frac{e}{(n+1)!} \longrightarrow 0$, quando $n \longrightarrow \infty$.

A expressão (1.3.4) para o resto foi deduzida por Lagrange e é especialmente usada quando conhecemos a ordem de grandeza da $(n+1)$-ésima derivada de f, como no Exemplo (1.3.2). Uma outra expressão para o resto é

$$R_n(x) = \frac{1}{n!} \int_{x_0}^{x} (x-y)^n\ f^{(n+1)}(y)\ dy \tag{1.3.5}$$

(Veja exercício 1.3.10).

Vimos que o resto R_n na fórmula de Taylor (1.3.2) tende a zero *mais rapidamente* que $(x - x_0)^n$ quando x tende à x_0. Isto significa que, para x próximo de x_0, a ordem de magnitude (grandeza) de $R_n(x)$ é menor que a de $(x - x_0)^n$.

Este tipo de comparação das ordens de magnitude de duas funções reais de variável real, em uma vizinhança de um determinado ponto, é comum em muitos problemas envolvendo métodos assintóticos. Nesse contexto é conveniente considerar extensões para funções, dos conceitos $O(\cdot)$ e $o(\cdot)$ apresentados na Seção `1.2` em termos de seqüências de números reais.

Definição 1.3.1 *Sejam f e g funções reais de variável real e x_0 um número real. Diremos que*

i) $f(x) = O\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0$ *se existirem números reais* $K > 0$ *e* $\delta > 0$ *tais que* $|\ f(x)/g(x)\ | \leq K$, *para todo* x *satisfazendo a condição* $0 < |x - x_0| < \delta$;

ii) $f(x) = O\big(\ g(x)\ \big)$ *quando* $x \longrightarrow +\infty$ $(-\infty)$ *se existirem números reais* $K > 0$ *e* $M > 0$ $(M < 0)$ *tais que* $|\ f(x)/g(x)\ | \leq K$, *para todo* $x > M$ $(x < M)$;

iii) $f(x) = o\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0$ *se para todo número* $\epsilon > 0$ *arbitrário existe um número real* $\delta > 0$ *tal que* $|\ f(x)/g(x)\ | < \epsilon$, *para todo* x *satisfazendo a condição* $0 < |\ x - x_0\ | < \delta$;

iv) $f(x) = o\big(\ g(x)\ \big)$ *quando* $x \longrightarrow +\infty$ $(-\infty)$ *se para todo número* $\epsilon > 0$ *arbitrário existe um número real* $M > 0$ $(M < 0)$ *tal que* $|\ f(x)/g(x)\ | < \epsilon$, *para todo* $x > M$ $(x < M)$.

Em outras palavras, $f(x) = O\big(\ g(x)\ \big)$ quando $x \longrightarrow x_0$ se $f(x)/g(x)$ for limitada em uma vizinhança de x_0; $f(x) = O\big(\ g(x)\ \big)$ quando $x \longrightarrow +\infty\ (-\infty)$ se $f(x)/g(x)$ for limitada para todo x suficientemente grande (pequeno) e $f(x) = o\big(\ g(x)\ \big)$ quando $x \longrightarrow x_0$ $(x \longrightarrow \pm\infty)$ se $\lim\limits_{x \longrightarrow x_0} \frac{f(x)}{g(x)} = 0$ $\big(\lim\limits_{x \longrightarrow \pm\infty} \frac{f(x)}{g(x)} = 0\big)$. Em particular, $f(x) = O\ (1)$ quando $x \longrightarrow x_0$ se $f(x)$ for limitada em uma vizinhança de x_0; $f(x) = O\ (1)$ quando $x \longrightarrow +\infty\ (-\infty)$ se $f(x)$ for limitada para todo x suficientemente grande (pequeno) e $f(x) = o(1)$ quando $x \longrightarrow x_0$ $(x \longrightarrow \pm\infty)$ se $\lim\limits_{x \longrightarrow x_0} f(x) = 0$ $\big(\lim\limits_{x \longrightarrow \pm\infty} f(x) = 0\big)$. Na seqüência, toda relação da forma $f(x) = h(x) + O\big(\ g(x)\ \big)$ quando $x \longrightarrow x_0\ (\pm\infty)\big($ ou $f(x) = h(x) + o\big(\ g(x)\ \big)$ quando $x \longrightarrow x_0\ (\pm\infty)\big)$ será interpretada como $f(x) - h(x) = O\big(\ g(x)\ \big)$ quando $x \longrightarrow x_0\ (\pm\infty)$ $\big($ ou $f(x) - h(x) = o\big(\ g(x)\ \big)\big)$ quando $x \longrightarrow x_0\ (\pm\infty)\big)$.

Exemplo 1.3.3 *i)* $\frac{7x^2+10x}{x^2} = 7 + \frac{10}{x} \leq 17,\ \forall x \geq 1,$ *o que implica* $7x^2 + 10x = O(x^2)$ *quando* $x \longrightarrow 2$;

ii) $\frac{7x^3+10x^2}{x} = 7x^2 + 10x,\ \forall x \neq 0,$ *o que implica* $\frac{7x^3+10x^2}{x} \longrightarrow 0$ *quando* $x \longrightarrow 0$, *ou* $7x^3 + 10x^2 = o(x)$ *quando* $x \longrightarrow 0$;

iii) $10(x-1)^2 + 4(x-1) \longrightarrow 0$ *quando* $x \longrightarrow 1$, *o que implica* $10(x-1)^2 + 4(x-1) = o(1)$ *quando* $x \longrightarrow 1$;

iv) pelo item (***iii*** *),* $10(x-1)^2 + 4(x-1) = O(1)$ *quando* $x \longrightarrow 1$;

v) $1/x \longrightarrow 0$ *quando* $x \longrightarrow \infty$, *o que implica* $1/x = o(1)$ *quando* $x \longrightarrow \infty$.

No teorema seguinte apresentamos alguns resultados bastante úteis em cálculos envolvendo os símbolos O e o.

Teorema 1.3.2 *Sejam* $f_1,\ f_2,\ g_1,\ g_2$ *e* g *funções reais de variável real e* x_0 *um número real;*

i) se $f_1(x) = O\big(\ g(x)\ \big)$ *e* $f_2(x) = O\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$, *então* $f_1(x) + f_2(x) = O\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$;

ii) se $f_1(x) = o\big(\ g(x)\ \big)$ *e* $f_2(x) = o\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$, *então* $f_1(x) + f_2(x) = o\big(\ g(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$;

iii) se $f_1(x) = O\big(\ g_1(x)\ \big)$ *e* $f_2(x) = O\big(\ g_2(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$, *então* $f_1(x)f_2(x) = O\big(\ g_1(x)\ g_2(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$;

iv) se $f_1(x) = o\big(\ g_1(x)\ \big)$ *e* $f_2(x) = o\big(\ g_2(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$, *então* $f_1(x)f_2(x) = o\big(\ g_1(x)\ g_2(x)\ \big)$ *quando* $x \longrightarrow x_0\ (\pm\infty)$.

Demonstração : Provaremos apenas os itens `i )` e `iv )` quando $x \longrightarrow x_0$; os casos em que $x \longrightarrow +\infty\ (-\infty)$ e os demais itens ficam como exercícios para o leitor.

`i`) Pela Definição `1.3.1` existem números reais $K_1 > 0,\ K_2 > 0,\ \delta_1 > 0$ e $\delta_2 > 0$ tais que $\mid f_1(x)/g(x) \mid \leq K_1$, para todo x satisfazendo a condição $0 < |x - x_0| < \delta_1$, e $\mid f_2(x)/g(x) \mid \leq K_2$, para todo x satisfazendo a condição $0 < |x - x_0| < \delta_2$. Logo,

$$\mid \frac{f_1(x) + f_2(x)}{g(x)} \mid \ \leq \ \mid \frac{f_1(x)}{g(x)} \mid + \mid \frac{f_2(x)}{g(x)} \mid \leq K_1 + K_2,$$

para todo x satisfazendo a condição $0 < |x - x_0| < \delta = \min\{\delta_1, \delta_2\}$, o que prova `i )`.

`iv`) Pela Definição `1.3.1`, para todo número $\epsilon > 0$ arbitrário existem números reais $\delta_1 > 0$ e $\delta_2 > 0$ tais que $\mid f_1(x) \ / \ g_1(x) \mid < \sqrt{\epsilon}$, para todo x satisfazendo a condição $0 < |x - x_0| < \delta_1$, e $\mid f_2(x) \ / \ g_2(x) \mid < \sqrt{\epsilon}$, para todo x satisfazendo a condição $0 < |x - x_0| < \delta_2$. Logo,

$$\mid \frac{f_1(x)\ f_2(x)}{g_1(x)\ g_2(x)} \mid = \mid \frac{f_1(x)}{g_1(x)} \mid \mid \frac{f_2(x)}{g_2(x)} \mid < \epsilon,$$

para todo x satisfazendo a condição $0 < |x - x_0| < \delta = \min\{\delta_1, \delta_2\}$ e o resultado `iv )` segue.

Pelo que vimos anteriormente, se uma função f for derivável até a ordem n em um ponto x_0, o resto R_n na fórmula de Taylor (1.3.2) é tal que $R_n(x) = o\big((x - x_0)^n\big)$ quando $x \longrightarrow x_0$; consequentemente (1.3.2) pode ser escrita como

$$f(x) \ = \ \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \ + \ o\big((x - x_0)^n\big) \quad \text{quando} \ \ x \longrightarrow x_0 \tag{1.3.6}$$

Como R_n também é tal que $R_n(x) = O\big((x - x_0)^n\big)$ (veja Exercício 1.3.6) a fórmula de Taylor (`1.3.2`) também pode ser escrita como:

$$f(x) \ = \ \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!} \ (x - x_0)^k \ + \ O\big((x - x_0)^n\big) \quad \text{quando} \ \ x \longrightarrow x_0 \tag{1.3.7}$$

Finalmente, para concluir esta seção faremos alguns comentários sobre *a série de Taylor* de uma função em torno de um ponto. Suponhamos que uma função real de variável real f, definida em algum intervalo aberto em torno de um ponto x_0, tenha derivadas de todas as ordens nesse intervalo. A série de Taylor de f em torno de x_0 é :

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} \ (x - x_0)^k \ = \ f(x_0) \ + \ \sum_{k=1}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{1.3.8}$$

Claramente, se $x = x_0$ a série (1.3.8) converge para $f(x_0)$. Nesse ponto surgem duas questões:

i) A série (1.3.8) converge para qualquer $x \neq x_0$?

ii) Se ela for convergente para algum $x \neq x_0$, então ela convergirá para $f(x)$?

As respostas à ambas estas questões não são sempre afirmativas. Por exemplo, com relação à questão `i )` considere a função $f(x) = \frac{1}{1-x}$, $x \neq 1$. A série de Taylor de f em torno de zero é a série geométrica $\sum\limits_{k=0}^{\infty} x^k$, que converge para $f(x)$ se $|x| < 1$ e é divergente se $|x| \geq 1$. Com relação à questão `ii )` considere a função $f(x) = e^{-\frac{1}{x^2}}$ para $x \neq 0$ e $f(0) = 0$. A série de Taylor de f em torno de zero converge para zero, para todo x real (veja Exercício 1.3.13). Logo, ela converge para $f(x)$ somente se $x = 0$, isto é , a série de Taylor de f em torno de zero representa f somente na origem. Em particular, o seguinte teorema nos dá uma condição suficiente para que uma função possa ser representada por sua série de Taylor.

Teorema 1.3.3 *Suponhamos que uma função real de variável real f tenha derivadas de todas as ordens em um intervalo $(x_0 - r, x_0 + r)$. Se existirem um número real positivo M (que pode depender de x_0) e um número inteiro positivo n_0 tais que, para todo $n \geq n_0$,*

$$| \, f^{(n)}(x) \, | \leq M, \quad \textit{para todo} \quad x \in (x_0 - r, x_0 + r) \tag{1.3.9}$$

então, para todo $x \in (x_0 - r, x_0 + r)$ a série de Taylor de f em torno de x_0 converge para $f(x)$, isto é ,

$$f(x) \; = \; \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

Demonstração : Para todo $x \in (x_0 - r, x_0 + r)$, pela fórmula de Taylor (1.3.3), temos que

$$f(x) \; = \; \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \; + \; \frac{f^{(n)}(c_n)}{n!}(x - x_0)^n \; , \quad \forall n \geq 1$$

onde c_n está entre x e x_0. Como, para todo $n \geq n_0$,

$$\left| \, \frac{f^{(n)}(c_n)}{n!}(x - x_0)^n \, \right| \leq \frac{M|x - x_0|^n}{n!} \quad \text{e} \quad \frac{M|x - x_0|^n}{n!} \longrightarrow 0 \text{ quando } \; n \longrightarrow \infty,$$

segue que $\frac{f^{(n)}(c_n)}{n!}(x - x_0)^n \longrightarrow 0$ quando $n \longrightarrow \infty$, o que prova o teorema.

Exemplo 1.3.4 *Consideremos a função $f(x) = e^x$. Como, para todo número real $r > 0$ arbitrário e para todo $n \geq 1$, $| \, f^{(n)}(x) \, | = e^x \leq e^r$, para todo $x \in (-r, r)$, a relação (1.3.9) se verifica para $n_0 = 1$, $M = e^r$ e $x_0 = 0$. Logo, pelo Teorema 1.3.3,*

$$e^x \; = \; \sum_{k=0}^{\infty} \frac{x^k}{k!} \; = \; 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \ldots \tag{1.3.10}$$

para todo $x \in (-r, r)$. Como r é arbitrário, (1.3.10) é valido para todo x real.

Gostariamos de ressaltar que, embora o estudo da convergência da série de Taylor de uma função em torno de um ponto seja importante para certos objetivos, ele não é essencial para nossos propósitos. Em geral basta conhecermos algumas propriedades do resto da fórmula de Taylor (1.3.2) para atingirmos nossas finalidades.

Exercícios

Exercício 1.3.1 *Determine o polinômio de Taylor, de ordem 2, de f em torno de x_0 nos casos:*

i) $f(x) = \frac{1}{1-x}$, $(x \neq 1)$ *e* $x_0 = 0$;

ii) $f(x) = \log(1+x)$, $(x > -1)$ *e* $x_0 = 0$;

iii) $f(x) = \frac{1}{1-x^2}$, $(x \neq \pm 1)$ *e* $x_0 = 0$;

iv) $f(x) = \sqrt{x}$, $(x > 0)$ *e* $x_0 = 0$.

Exercício 1.3.2 *Verifique que*

i) $1 - x + x^2 - x^3 + \cdots + (-1)^n x^n$ *é o polinômio de Taylor, de ordem n, de* $1/(1+x)$, $(x \neq -1)$ *em torno de* $x_0 = 0$;

ii) $x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^n \frac{x^{n+1}}{n+1}$ *é o polinômio de Taylor, de ordem $n+1$, de* $\log(x+1)$, $(x > -1)$, *em torno de* $x_0 = 0$.

Exercício 1.3.3 *Seja $f(x) = (1+x)^\alpha$, onde $\alpha \neq 0$ é um número real. Determine o polinômio de Taylor de f em torno de $x_0 = 0$ e dê a expressão do resto em função da derivada de ordem $n+1$.*

Exercício 1.3.4 *Seja f uma função derivável até $2^{\underline{a}}$ ordem em um intervalo I e seja $x_0 \in I$. Mostre que existe uma função $g(x)$ definida em I tal que, para todo $x \in I$,*

$$f(x) \;=\; f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + g(x)(x - x_0)^2 \;, \quad \text{com} \lim_{x \longrightarrow x_0} g(x) = 0.$$

Exercício 1.3.5 *Prove os itens **ii)** e **iii)** do Teorema 1.3.2 quando $x \longrightarrow x_0$ e os itens*
***i)- iv)** do Teorema quando $x \longrightarrow +\infty$ $(-\infty)$.*

Exercício 1.3.6 *As igualdades abaixo descrevem, de forma abreviada, algumas das regras utilizadas em cálculos envolvendo os símbolos O e o. Por exemplo, a igualdade **i)** significa que se $f(x) = o\big(\, g(x) \,\big)$, então $f(x) = O\big(\, g(x) \,\big)$, quando $x \longrightarrow x_0$ $(\pm\infty)$. Prove que, quando $x \longrightarrow x_0$ $(\pm\infty)$:*

i) $o\big(\, g(x) \,\big) \;=\; O\big(\, g(x) \,\big)$;

ii) $O\big(O(\ g(x)\ )\big) \ = \ O(\ g(x)\ );$

iii) $O\big(o(\ g(x)\ )\big) \ = \ o(\ g(x)\ );$

iv) $o\big(O(\ g(x)\ )\big) \ = \ o(\ g(x)\ );$

v) $o(\ g_1(x)\ )\ O(\ g_2(x)\ ) \ = \ o(\ g_1(x)\ g_2(x)\ );$

vi) se $c \neq 0$ *for uma constante, então*

$$O(\ c\ g(x)\ ) \ = \ c\ O(\ g(x)\ );$$

$$o(\ c\ g(x)\ ) \ = \ c\ o(\ g(x)\ ).$$

Exercício 1.3.7 *Se* a *e* b *forem números reais estritamente positivos prove que*

$$\lim_{x \longrightarrow 0} \frac{a^x - b^x}{x} \ = \ \log \frac{a}{b}$$

Sugestão*:* $a^x \ = \ e^{x \log a}$ *e* $e^x = 1 + x + o(x)$ *quando* $x \longrightarrow 0.$

Exercício 1.3.8 *Prove que, quando* $x \longrightarrow 0,$

i) $e^x \ = \ 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + o(x^n);$

ii) $\text{sen}\ x \ = \ x - \frac{x^3}{3!} + \frac{x^5}{5!} + o(x^6);$

iii) $\cos x \ = \ 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + o(x^5);$

iv) $\log\ (1 + x) \ = \ x - \frac{1}{2}x^2 + o(x^2).$

Exercício 1.3.9 *Seja* f *uma função real derivável até a ordem* n *em um ponto* x_0*. Prove que* $R_n(x) \ = \ o\big((x - x_0)^n\big)$ *quando* $x \longrightarrow x_0$*, onde* R_n *é o resto da fórmula de Taylor (1.3.2).*

Exercício 1.3.10 *Prove que o resto* R_n *na fórmula de Taylor (1.3.2) pode ser expresso como*

$$R_n(x) \ = \ \frac{1}{n!} \int_{x_0}^{x} (x - y)^n\ f^{(n+1)}\ (y)\ dy$$

Sugestão*:* $f(x) - f(x_0) \ = \ \int\limits_{x_0}^{x} f'\ (y)\ dy$*; use integração por partes para obter* $f(x) \ = \ f(x_0) + f'(x_0)(x - x_0) + \int\limits_{x_0}^{x} (x - y)\ f''\ (y)\ dy$ *e repita o processo.*

Exercício 1.3.11

i) Prove a seguinte identidade, válida para todo $t \neq -1$:

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \cdots + (-1)^{n-1} t^{n-1} + \frac{(-1)^n t^n}{1+t};$$

ii) integre ambos os lados para obter a fórmula

$$\log\,(1+x) \;=\; \sum_{k=1}^{n} (-1)^{k-1}\,\frac{x^k}{k} \;+\; (-1)^n \int_0^x \frac{t^n}{1+t}\,dt \quad \textit{se} \;\; x > -1;$$

iii) prove que

$$\Big|\int_0^x \frac{t^n}{1+t}\,dt\,\Big| \;\leq\; \frac{|x|^{n+1}}{1-|x|}\,\Big(\frac{1}{n+1}\Big) \quad \textit{se} \;\; |x| < 1;$$

iv) conclua que

$$\log\,(1+x) \;=\; \sum_{k=1}^{\infty} (-1)^{k-1}\,\frac{x^k}{k} \quad \textit{se} \;\; |x| < 1.$$

Exercício 1.3.12 *Prove que*

i) se $a > 0,\; a^x \;=\; \sum_{k=0}^{\infty} \frac{(\log a)^k}{k!}\, x^k$ *para todo* x *real;*

Sugestão*:* $a^x \;=\; e^{x \log a}$;

ii) $\frac{1}{2-x} \;=\; \sum_{k=0}^{\infty} \frac{x^k}{2^{k+1}}$ *se* $|x| < 2$;

iii) $e^{-x^2} \;=\; \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\, x^{2k}$ *para todo* x *real* ;

iv) $\log\,\sqrt{\frac{1+x}{1-x}} \;=\; \sum_{k=0}^{\infty} \frac{x^{2k+1}}{2^{k+1}}$ *se* $|x| < 1$.

Exercício 1.3.13 *Seja* f *a função definida por* $f(x) = e^{\frac{-1}{x^2}}$ *se* $x \neq 0$ *e* $f(0) = 0$.

i) Mostre que f *tem derivadas de todas as ordens em toda a reta real;*

ii) mostre que $f^{(k)}(0) = 0,\; \forall k \geq 1$.

(A série de Taylor de f *em torno de zero converge em toda a reta, mas ela representa* f *somente na origem)*

1.4 Funções Características

Nesta seção apresentaremos a definição e algumas propriedades básicas das funções características, ferramentas muito úteis no estudo da convergência em distribuição de seqüências de variáveis e vetores aleatórios.

Se X e Y forem variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$, então $Z = X + i\ Y$ será uma variável aleatória assumindo valores no conjunto dos números complexos, $\mathbb{C}$. Se $E(X) < \infty$ e $E(Y) < \infty$, a esperança de Z pode ser definida como $E(Z) = E(X) + i\ E(Y)$. Dada uma variável aleatória X com função distribuição F_X, podemos construir pela fórmula de Euler, a variável aleatória complexa
$e^{iX} = \cos X + i\ \text{sen} X$. Como as variáveis aleatórias $\cos X$ e $\text{sen} X$ tem esperanças finitas, $E(e^{iX}) = E(\cos X) + iE(\text{sen} X)$ 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 está bem definida. Utilizando esses fatos podemos considerar a seguinte definição:

Definição 1.4.1 *Seja X uma variável aleatória. A função característica de X é a função $\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}$ definida por $\varphi_X(t) = \int e^{itx} dF_X(x)$ para todo $t \in \mathbb{R}$.*

As seguintes propriedades são importantes nas aplicações:

Propriedade 1.4.1 $\mid \varphi_X(t) \mid \leq 1$ *para todo* $t \in \mathbb{R}$.

Demonstração:

$$\mid \varphi_X(t) \mid = \mid \int e^{itx} dF_X(x) \mid \leq \int |e^{itx}| dF_X(x) =$$

$$= \int \left[\cos^2(tx) + \text{sen}^2\ (tx) \right]^{\frac{1}{2}} dF_X(x) = \int dF_X(x) = 1.$$

Propriedade 1.4.2 $\varphi_X(0) = 1$ *e* $\overline{\varphi_X}(t) = \varphi_X(-t)$ *para todo* $t \in \mathbb{R}$, *onde* $\overline{\varphi_X}(t)$ *é o complexo conjugado de* $\varphi_X(t)$.

Demonstração: $\varphi_X(0) = E(e^{i0X}) = E(1) = 1$;

$$\overline{\varphi_X}(t) = E\big(\cos(tX)\big) - iE\big(\ \text{sen}(tX)\big) = E\big(\cos(-tX)\big) + iE\big(\ \text{sen}(-tX)\big) = \varphi_X(-t).$$

Propriedade 1.4.3 φ_X *é uniformemente contínua.*

Demonstração: Considere a função

$$h(u) \;=\; \int \mid e^{iux} - 1 \mid dF_X(x),\;\; u \in \mathbb{R}.$$

Pelo Teorema da Convergência Dominada (ver Bartle (1966), por exemplo) $h(u_n) \longrightarrow 0$ para toda seqüência $(u_n)_{n\geq 1}$ tal que $u_n \longrightarrow 0$, o que implica $\lim\limits_{u\longrightarrow 0} h(u) = 0$. Logo, para todo $\epsilon > 0$ existe $\delta = \delta(\epsilon) > 0$ tal que, para todo $t, s \in \mathbb{R}$ com $|t-s| < \delta$ temos $h(t-s) < \epsilon$ e $\mid \varphi(t) - \varphi(s) \mid \;=\; \mid \int \big(e^{itx} - e^{isx}\big) dF_X(x) \mid \;\leq\; \int \mid e^{itx} - e^{isx} \mid dF_X(x) \;=\; \int \mid e^{isx} \mid \mid e^{i(t-s)x} - 1 \mid dF_X(x) \;=\; \int \mid e^{i(t-s)x} - 1 \mid dF_X(x) \;=\; h(t-s) \;<\; \epsilon$.

Propriedade 1.4.4 *Se X e Y forem variáveis aleatórias independentes, então $\varphi_{X+Y}(t) \;=\; \varphi_X(t)\; \varphi_Y(t)$ para todo $t \in \mathbb{R}$.*

Demonstração: $\varphi_{X+Y}(t) \;=\; E\big(e^{it(X+Y)}\big) \;=\; E\big(e^{itX} e^{itY}\big) \;=\; E\big(e^{itX}\big) E\big(e^{itY}\big) \;=\; \varphi_X(t)\; \varphi_Y(t)$, para todo $t \in \mathbb{R}$.

A propriedade 1.4.4 pode ser facilmente generalizada:

Propriedade 1.4.4-A Se $X_1, X_2, \ldots, X_n$ forem variáveis aleatórias independentes, então

$$\varphi_{\sum\limits_{j=1}^{n} X_j}(t) \;=\; \prod_{j=1}^{n} \varphi_{X_j}(t) \quad ,\; \forall t \in \mathbb{R}.$$

A função característica de uma variável aleatória é determinada pela sua função distribuição pois $\varphi_X(t) \;=\; \int e^{itx} dF_X(x)$.

A recíproca corresponde à seguinte propriedade:

Propriedade 1.4.5 *A função característica de uma variável aleatória X determina a função distribuição de X.*

Esta propriedade é decorrente da "fórmula da inversão", cuja demonstração pode ser encontrada em Gnedenko (1969), por exemplo:

Seja X uma variável aleatória com função distribuição F e função característica φ. Se x e y forem pontos de continuidade de F tais que $x < y$, então

$$F(y) - F(x) \;=\; \frac{1}{2\pi} \lim_{u\longrightarrow\infty} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it}\; \varphi(t)\; dt$$

Demonstração: Usando a fórmula da inversão, para todo $z \in \mathbb{R}$, temos

$$F(z) \;=\; \lim_{y\downarrow z} \lim_{x\longrightarrow -\infty} \big(F(y) - F(x)\big) \;=\; \lim_{y\downarrow z} \lim_{x\longrightarrow -\infty} \lim_{u\longrightarrow\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it} \varphi(t) dt$$

Propriedade 1.4.6 *Uma variável aleatória X tem distribuição simétrica em torno de zero, isto é* $\mathbb{P}\{X \leq x\} \ = \ \mathbb{P}\{X \geq -x\}, \ \forall x \in \mathbb{R}$, *se e somente se* $\varphi_X(t)$ *for real,* $\forall t \in \mathbb{R}$.

Demonstração: X é simétrica em torno de zero se e somente se $F_X = F_{-X} \iff \varphi_X = \varphi_{-X}$, que é equivalente a:

$$\varphi_X(t) \ = \ \varphi_{-X}(t), \ \ \forall t \in \mathbb{R} \tag{1.4.1}$$

Mas $\varphi_{-X}(t) \ = \ E(e^{-itX}) \ = \ \varphi_X(-t) \ = \ \overline{\varphi_X}(t)$; consequentemente (1.4.1) é equivalente a $\varphi_X(t) \ = \ \overline{\varphi_X}(t), \ \forall t \in \mathbb{R}$ que por sua vez é equivalente a $\varphi_X(t) \in \mathbb{R}, \ \forall t \in \mathbb{R}$

Propriedade 1.4.7 *Dados* $a, b \in \mathbb{R}$ *arbitrários,* $\varphi_{aX+b}(t) \ = \ e^{ibt}\varphi_X(at)$ *para todo* $t \in \mathbb{R}$.

Demonstração:

$$\varphi_{aX+b}(t) \ = \ E\big(e^{it(aX+b)}\big) \ = \ E\big(e^{iatX}e^{ibt}\big) \ = \ e^{ibt}E(e^{iatX}) \ = \ e^{ibt}\varphi_X(at).$$

Propriedade 1.4.8 *Se* $E|X|^n < \infty$, *então* φ_X *tem* n *derivadas contínuas, com* $\varphi_X^{(k)}(t) \ = \ \int (ix)^k e^{itx} dF_X(x), \ \ k = 1, 2, \ldots, n$. *Particularmente,* $\varphi_X^{(k)}(0) \ = \ i^k E(X^k)$.

Demonstração: Para quaisquer t e $h \in \mathbb{R}, \ h \neq 0$

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} \ = \ \int \frac{e^{i(t+h)x} - e^{itx}}{h} dF_X(x) \ =$$

$$\int e^{itx}\frac{(e^{ihx} - 1)}{h} dF_X(x) \ = \ E\Big[\ e^{itX}\frac{(e^{ihX} - 1)}{h} \ \Big].$$

Como $\frac{e^{ihx}-1}{h} \longrightarrow ix$ quando $h \longrightarrow 0$, para todo $x \in \mathbb{R}$, temos

$$e^{itX}\Big(\frac{e^{ihX} - 1}{h}\Big) \longrightarrow iXe^{itX} \text{ quando } h \longrightarrow 0.$$

Por outro lado, como para todo $x \in \mathbb{R}$,

$$\Big| \ \frac{e^{ihx} - 1}{h} \ \Big| \ = \ \frac{| \ \int_0^h ixe^{iux}du \ |}{h} \ = \ |x| \ \Big| \ \frac{\int_0^h e^{iux}du}{h} \ \Big| \ \leq$$

$$\leq \ |x| \ \frac{\int_0^h |e^{iux}|du}{h} \ = \ |x| \ \frac{\int_0^h du}{h} \ = \ |x| \ ,$$

$$\text{então} \quad \Big| \ e^{itX}\frac{(e^{ihX} - 1)}{h} \ \Big| \ \leq \ |X|.$$

Logo, supondo $E|X| < \infty$, e utilizando o Teorema da Convergência Dominada temos

$$\varphi'_X(t) \ = \ \lim_{h \longrightarrow 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} \ =$$

$$= \lim_{h \longrightarrow 0} E\Big[\ \frac{e^{itX}(e^{ihX}-1)}{h} \ \Big] \ = \ E\big(iXe^{itX}\big) \ = \ \int ixe^{itx}dF_X(x)$$

A continuidade de $\varphi'(t)$ no ponto t decorre dos fatos: $\lim\limits_{s \longrightarrow t} ixe^{isx} \ = \ ixe^{itx}$ e $|ixe^{isx}| \ = \ |x|$

O resultado segue por indução em n.

Exemplo 1.4.1

a) Se $X \sim$ Poisson (λ), então $\varphi_X(t) \ = \ exp \ [\ \lambda\big(\ exp \ (it) - 1\big) \ ] \ = \ e^{\lambda(e^{it}-1)}$

b) Se $X \sim$ Binomial (n,p), então $\varphi_X(t) \ = \ (pe^{it}+q)^n$, *onde* $q = 1-p$.

c) Se $X \sim$ Geométrica (p), então $\varphi_X(t) \ = \ \frac{pe^{it}}{1-qe^{it}}$

d) Se $X \sim$ Binomial Negativa (r,p), então $\varphi_X(t) \ = \ \big(\frac{p}{1-qe^{it}}\big)^r$

e) Se $X \sim$ Uniforme $[\ a,b \ ]$, então $\varphi_X(t) \ = \ \frac{e^{itb}-e^{ita}}{it(b-a)}$

f) Se $X \sim$ Normal (μ,σ^2), então $\varphi_X(t) \ = \ e^{it\mu-\frac{1}{2}t^2\sigma^2}$

g) Se $X \sim$ Exponencial (λ), então $\varphi_X(t) \ = \ \big(1-\frac{it}{\lambda}\big)^{-1}$

h) Se $X \sim$ Gama (r,λ), então $\varphi_X(t) \ = \ \big(1-\frac{it}{\lambda}\big)^{-r}$

i) Se $X \sim$ Qui-Quadrado (n), então $\varphi_X(t) \ = \ \big(1-2it\big)^{-\frac{n}{2}}$

O conceito de função característica pode ser generalizado para vetores aleatórios. Nesse sentido seja $\mathbf{X} \ = (X_1, X_2, \ldots, X_p)$ um vetor aleatório $p-$dimensional $(p \geq 1)$.

Definição 1.4.2 *A função característica de $\boldsymbol{X}$ é a função $\varphi_{\boldsymbol{X}} : \mathbb{R}^p \longrightarrow \mathbb{C}$, definida por*

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) \ = \ E\Big(\ e^{i\sum\limits_{j=1}^{p} t_j X_j} \ \Big) \ = \ E\Big(\ e^{i\boldsymbol{t}\boldsymbol{X}^t} \ \Big), \quad \forall \boldsymbol{t} \ = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p.$$

A função característica de um vetor aleatório tem propriedades análogas às propriedades enunciadas para a função característica de uma variável aleatória. Quanto à Propriedade 1.4.5, existe uma "fórmula de inversão" para a função característica de um vetor aleatório (ver Gnedenko (1969), por exemplo). Essa fórmula nos permite provar o seguinte resultado.

Teorema 1.4.1 (Unicidade) *Se $\boldsymbol{X}$ e $\boldsymbol{Y}$ forem vetores aleatórios $p-$dimensionais tais que $\varphi_{\boldsymbol{X}}\ (\boldsymbol{t}\ ) \ = \varphi_{\boldsymbol{Y}}\ (\boldsymbol{t}\ )$ para todo $\boldsymbol{t} \ \in \mathbb{R}^p$, então $\boldsymbol{X}$ e $\boldsymbol{Y}$ tem a mesma distribuição.*

Em outras palavras, a função característica determina a distribuição e temos:

$$\varphi_{\mathbf{X}} = \varphi_{\mathbf{Y}} \iff F_{\mathbf{X}} = F_{\mathbf{Y}} \ .$$

O próximo teorema nos fornece um critério para verificação da independência de dois vetores aleatórios em termos de suas funções características.

Teorema 1.4.2 *Sejam* $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ *e* $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_q)$ *vetores aleatórios.* $\boldsymbol{X}$ *e* $\boldsymbol{Y}$ *são independentes se, e somente se,*

$$\varphi_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{s},\boldsymbol{t}) = \varphi_{\boldsymbol{X}}(\boldsymbol{s})\, \varphi_{\boldsymbol{Y}}(\boldsymbol{t}) \ \ \forall\, \boldsymbol{s} \in \mathbb{R}^p, \ \forall \boldsymbol{t} \in \mathbb{R}^q. \tag{1.4.2}$$

Demonstração: Demonstraremos somente que a independência de $\mathbf{X}$ e $\mathbf{Y}$ implica (`1.4.1`). A demonstração da recíproca desse resultado envolve conceitos da Teoria da Medida e Integração e pode ser encontrada em Breiman (1968, cap 8), por exemplo. Suponhamos, então, que $\mathbf{X}$ e $\mathbf{Y}$ sejam vetores aleatórios independentes. Logo, as variáveis aleatórias $e^{i\mathbf{s}\,\mathbf{X}^t}$ e $e^{i\mathbf{t}\,\mathbf{Y}^t}$ são independentes, $\forall \mathbf{s} \in \mathbb{R}^p$, $\forall \mathbf{t} \in \mathbb{R}^q$, o que implica

$$\varphi_{(\mathbf{X},\mathbf{Y})}(\mathbf{s},\mathbf{t}) = E\big(e^{i(\mathbf{s}\mathbf{X}^t + \mathbf{t}\mathbf{Y}^t)} \big) = E\big(e^{i\mathbf{s}\mathbf{X}^t}\, e^{i\mathbf{t}\mathbf{Y}^t} \big) =$$

$$= E\big(e^{i\mathbf{s}\mathbf{X}^t} \big)\, E\big(e^{i\mathbf{t}\mathbf{Y}^t} \big) = \varphi_{\mathbf{X}}(\mathbf{s})\, \varphi_{\mathbf{Y}}(\mathbf{t}), \ \ \forall \mathbf{s} \in \mathbb{R}^p, \ \forall \mathbf{t} \in \mathbb{R}^q$$

e o teorema está provado.

O Teorema 1.4.2 pode ser generalizado para qualquer número finito de vetores aleatórios (possivelmente de diferentes dimensões).

Em particular, se $X_1, X_2, \ldots, X_n$ forem variáveis aleatórias (vetores aleatórios unidimensionais) temos: $X_1, X_2, \ldots, X_n$ são independentes se, e somente se,

$$\varphi_{(X_1,X_2,\ldots,X_n)}(\mathbf{t}) = \prod_{j=1}^{n} \varphi_{X_j}(t_j) \ \ , \quad \forall\, \mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n.$$

Finalizando esta seção apresentamos o seguinte resultado devido a Cramér e Wold, que reduz o problema da determinação de uma distribuição multidimensional (multivariada) ao caso unidimensional.

Teorema 1.4.3 (Cramér - Wold) *A distribuição de um vetor aleatório* $\boldsymbol{X} = (X_1, \ldots, X_p)$ *é completamente determinada pelo conjunto das distribuições de todas as combinações lineares* $\boldsymbol{t}\,\boldsymbol{X}^t = \sum_{j=1}^{p} t_j X_j$, *onde* $\boldsymbol{t} \in \mathbb{R}^p$.

Demonstração: Como, pelo Teorema 1.4.1 (Teorema da Unicidade), a distribuição de $\mathbf{X}$ é determinada pela função característica $\varphi_{\mathbf{X}}$, o resultado segue do fato:

$$\varphi_{\mathbf{X}}(\mathbf{t}\ ) \ = \ E(\ e^{i\mathbf{t}\ \mathbf{X}\ ^t}\ ) \ = \ \varphi_{\mathbf{t}\ \mathbf{X}^t}(1)\ , \quad \forall\ \mathbf{t}\ \in \mathbb{R}^p.$$

Exercícios

Exercício 1.4.1 *Se $X \sim$ Poisson (λ), prove que a função característica de X é*

$$\varphi_X(t) \ = \ e^{\lambda(e^{it}-1)} \quad , \quad \forall\ t \in \mathbb{R}.$$

Exercício 1.4.2 *Seja $X \sim$ Bernoulli (p), isto é , $\mathbb{P}(X=0) \ = \ 1-p \ = \ q$ e $\mathbb{P}(X=1) \ = \ p$.*

i) Determine a função característica de X;

ii) se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de X, sabe-se que

$$S_n \ = \ X_1 + X_2 + \cdots + X_n \ \sim \ \textit{Binomial}\ (n,p).$$

*Utilize o item **i**) para provar que a função característica de S_n é*

$$\varphi_{S_n}(t) \ = \ \big(\ pe^{it} + q\ \big)^n, \qquad \forall\ t \in \mathbb{R}.$$

Exercício 1.4.3

i) Se $X \sim \mathcal{N}(0,1)$ prove que a função característica de X é

$$\varphi_X(t) \ = \ e^{-\frac{1}{2}t^2}\ , \quad \forall\ t \in \mathbb{R}.$$

Sugestão*:* $\displaystyle\int_{-\infty-it}^{\infty-it} e^{-u^2/2}\ du \ = \ \sqrt{2\pi}\ , \ \forall t \in \mathbb{R};$

*ii) utilize o item **i**) para provar que, se $X \sim \mathcal{N}(\mu, \sigma^2)$ então a função característica de X é $\varphi_X(t) \ = \ e^{it\mu - \frac{1}{2}\sigma^2 t^2}\ , \ \ \forall\ t \in \mathbb{R}$.*

Exercício 1.4.4 *Suponha que X e Y sejam variáveis aleatórias identicamente distribuídas. Prove que*

i) se X e Y forem independentes, então $X-Y$ tem distribuição simétrica em torno de zero;

ii) se X e Y assumem somente dois valores, então $X-Y$ tem distribuição simétrica em torno de zero.

Exercício 1.4.5 *Prove, usando funções características, que* *i) se $X \sim$ Binomial (n_1, p), $Y \sim$ Binomial (n_2, p) e X e Y forem independentes, então $X + Y \sim$ Binomial $(n_1 + n_2, p)$;*

ii) se $X \sim$ Poisson (λ_1), $Y \sim$ Poisson (λ_2) e X e Y forem independentes, então $X + Y \sim$ Poisson $(\lambda_1 + \lambda_2)$;

iii) se $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, e X e Y forem independentes, então, $aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ para quaisquer $a, b \in \mathbb{R}$.

Exercício 1.4.6 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim \mathcal{N}(0,1)$;*

i) prove que $\overline{X}_n$ e $\left(X_1 - \overline{X}_n, X_2 - \overline{X}_n, \ldots, X_n - \overline{X}_n\right)$ são independentes.
Sugestão*: utilize o Teorema **1.4.2** e a identidade*

$$\sum_{j=1}^{n} t_j(X_j - \overline{X}_n) \; + \; t\,\overline{X}_n \; = \; \sum_{j=1}^{n} (t_j - \bar{t} \; + \; t/n)\; X_j,$$

onde $\bar{t} \; = \; \frac{1}{n} \sum_{j=1}^{n} t_j, \quad \forall\; (t, t_1, t_2, \ldots, t_n) \in \mathbb{R}^{n+1}$;

*ii) utilize o item **i**) para provar que $\overline{X}_n$ e S_n^2 são variáveis aleatórias independentes;*

iii) prove que $(n-1)S_n^2 \sim \chi^2_{(n-1)}$
Sugestão*: utilize a identidade* $\sum_{j=1}^{n} X_j^2 \; = \; \sum_{j=1}^{n} (X_j - \overline{X}_n)^2 \; + \; n\overline{X}_n^{\,2}$ *e o fato de que* $\varphi_{X^2}(t) \; = \; (1-2it)^{-\frac{1}{2}}, \quad \forall\; t \in \mathbb{R}$.

Exercício 1.4.7 *Suponha que $\boldsymbol{X} \; = \; (X_1, X_2, \ldots, X_q)$ tenha distribuição Multinomial com parâmetros $p_1, p_2, \ldots, p_q$ e n, isto é ,*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q) \; = \; \frac{n!}{x_1! x_2! \ldots x_q!}\; p_1^{x_1}\; p_2^{x_2}\; \ldots\; p_q^{x_q},$$

onde $x_1, x_2, \ldots, x_q$ são inteiros não negativos tais que $\sum_{j=1}^{q} x_j \; = \; n$. Prove que a função característica de $\boldsymbol{X}$ é dada por

$$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) \; = \; \left(\; p_1 e^{it_1} \; + \; p_2 e^{it_2} \; + \; \ldots \; + \; p_q e^{it_q}\; \right)^n, \quad \forall\; \boldsymbol{t} \; \in \mathbb{R}^q.$$

Exercício 1.4.8 *Seja (X, Y) um vetor aleatório com vetor de médias $\boldsymbol{\mu} = (\mu_1, \mu_2)$ e matriz de covariância*

$$\boldsymbol{\Sigma} \; = \; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

(ρ é o coeficiente de correlação entre X e Y).

Pode-se provar que, se (X, Y) tem distribuição Normal bivariada, então a função característica de (X, Y) é dada por

$$\varphi_{(X,Y)}(\boldsymbol{t}) \;=\; e^{i\boldsymbol{t}\boldsymbol{\mu}^t \;-\; \frac{\boldsymbol{t}\sum \boldsymbol{t}^t}{2}} \;,\;\; \forall\; \boldsymbol{t} \;=\; (t_1, t_2) \in \mathbb{R}^2.$$

i) Suponha que (X, Y) tenha distribuição Normal bivariada; prove que X e Y são independentes se, e somente se, $\rho = 0$;

ii) prove que, (X, Y) tem distribuição Normal bivariada se, e somente se,

$$t_1\, X \;+\; t_2\, Y \;\sim\; \mathcal{N}\left(\, \boldsymbol{t}\boldsymbol{\mu}^t, \boldsymbol{t}\sum \boldsymbol{y}^t\, \right), \quad \forall\; \boldsymbol{t} = (t_1, t_2) \in \mathbb{R}^2.$$

Capítulo 2

Convergência Estocástica

2.1 Motivação

As regras de decisão geralmente empregadas em inferência estatística envolvem a consideração de estimadores de parâmetros de interesse ou de estatísticas de teste para hipóteses sobre os modelos adotados. Essencialmente, essas *estatísticas* são funções das observações amostrais e como tal são variáveis aleatórias. O mínimo que podemos exigir de uma estatística empregada em alguma regra de decisão é que suas flutuações aleatórias diminuam à medida que o tamanho amostral aumente, ou, em outras palavras, que ela seja *consistente*. No caso da estimação de algum parâmetro de interesse a consistência implica uma crescente *proximidade* entre os valores do estimador e o verdadeiro valor do parâmetro com o aumento do tamanho da amostra. Obviamente, o conceito de proximidade, neste caso, precisa ser definido precisamente de forma a incorporar a natureza estocástica do estimador. Se definirmos uma *distância* (ou *norma*) conveniente entre o estimador e o parâmetro, o conceito de consistência estará relacionado com a forma segundo a qual essa distância converge para zero, quando o tamanho da amostra cresce. Mais especificamente, se $\boldsymbol{T}_n$ for um estimador de um parâmetro $\boldsymbol{\theta} \in \mathbb{R}^p$ $(p \geq 1)$ baseado em uma amostra de tamanho $n \geq 1$ e $||\cdot||$ representar uma norma em $\mathbb{R}^p$, diremos que a seqüência $\{\boldsymbol{T}_n\}_{n\geq 1}$ é consistente se $||\boldsymbol{T}_n - \boldsymbol{\theta}||$ convergir para zero de algum modo bem definido, quando $n \longrightarrow \infty$. O objetivo básico deste capítulo é definir esses modos de convergência de uma forma precisa. Com essa finalidade, na Seção 2.2 estenderemos os conceitos de ordens de magnitude de seqüências de números reais e funções de variável real para seqüências estocásticas; na Seção 2.3 especificaremos os diferentes modos de convergência estocástica, definindo convergência em probabilidade, quase certa, em média e em distribuição de seqüências de variáveis aleatórias. Finalmente, na Seção 2.4 aplicaremos esses conceitos através das chamadas Leis dos Grandes Números, que essencialmente estabelecem condições para a validade das diversas formas de consistência.

2.2 Ordens de magnitude de seqüências estocásticas: a notação $O_p(\cdot),\ o_p(\cdot)$.

No que segue, salvo menção explícita em contrário, quando nos referirmos a uma seqüência de variáveis aleatórias (vetores aleatórios) estaremos nos referindo a uma seqüência de variáveis aleatórias (vetores aleatórios) definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$.

Para comparar seqüências de variáveis aleatórias precisamos incorporar seu caráter aleatório nos conceitos de ordens de magnitude estudados na Seção 1.2. Nesse sentido, apresentamos inicialmente a seguinte definição:

Definição 2.2.1 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias e $\{b_n\}_{n\geq 1}$ uma seqüência de números reais (ou variáveis aleatórias). Diremos que*

i) $X_n = O_p(b_n)$ se para todo número real $\eta > 0$ existirem um número real positivo $K = K(\eta)$ e um número inteiro positivo $n_0 = n_0(\eta)$, tais que

$$\mathbb{P}\Big(\ |\ X_n b_n\ |\ \geq K\ \Big)\ \leq\ \eta, \quad \forall\ n \geq n_0;$$

ii) $X_n = o_p(b_n)$ se para todo número real $\epsilon > 0$ e para todo número real $\eta > 0$ existir um número inteiro positivo $n_0 = n_0(\epsilon, \eta)$, tal que

$$\mathbb{P}\Big(\ |\ X_n b_n\ |\ \geq \epsilon\ \Big)\ <\ \eta, \quad \forall\ n \geq n_0.$$

Em outras palavras, diremos que $X_n = O_p(b_n)$ se a seqüência $\{X_n b_n\}_{n\geq 1}$ for limitada *em probabilidade*, para todo n suficientemente grande e que $X_n = o_p(b_n)$ se, para todo número real $\epsilon > 0$, $\mathbb{P}\Big(\ |\ X_n b_n\ |\ \geq\ \epsilon\ \Big) \longrightarrow 0$ quando $n \longrightarrow \infty$. Em particular, $X_n = O_p(1)$ se para todo número real $\eta > 0$ existir um número real $K > 0$ tal que $\mathbb{P}\Big(\ |\ X_n\ |\ \geq K\ \Big)\ \leq\ \eta$, para todo n suficientemente grande e $X_n = o_p(1)$ se, para todo número real $\epsilon > 0$, $\mathbb{P}\Big(\ |\ X_n\ |\ \geq \epsilon\ \Big) \longrightarrow 0$ quando $n \longrightarrow \infty$.

Exemplo 2.2.1 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $\mu = E(X_1)$, $\sigma^2 =$ Var $(X_1) < \infty$ e $T_n = \sum_{j=1}^{n} X_j = n\overline{X}$,* $\forall\ n \geq 1$. *Pela desigualdade de Chebychev (Teorema* `A.1` *do Apêndice* `A` *), para todo $\eta > 0$,*

$$\mathbb{P}\Big(\ n^{\frac{1}{2}}|\ \overline{X}_n - \mu\ | \geq \frac{\sigma}{\eta^{\frac{1}{2}}}\ \Big)\ =\ \mathbb{P}\Big(\ n^{-\frac{1}{2}}|\ T_n - n\mu\ | \geq \frac{\sigma}{\eta^{\frac{1}{2}}}\ \Big)\ =$$

$$= \ \mathbb{P}\big(\mid T_n - n\mu \mid \geq \sigma \ (\frac{n}{\eta})^{\frac{1}{2}} \ \big) \ \leq \ \frac{\eta \ Var \ (T_n)}{n\sigma^2} \ = \ \eta, \quad \forall \ n \geq 1.$$

Logo, $\overline{X}_n - \mu \ = \ O_p(n^{-\frac{1}{2}})$ *ou* $T_n - n\mu \ = \ O_p(n^{\frac{1}{2}})$.

Exemplo 2.2.2 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias identicamente distribuídas com* $X_1 \sim$ *Exponencial* (1)*. Para todo* $\epsilon > 0$ *e para todo* $n > 1$*,*

$$\mathbb{P}\big(\mid X_n \log n \mid \geq \epsilon \ \big) \ = \ \mathbb{P}\big(\ X_1 \geq \epsilon \log n \ \big) \ = \ e^{-\epsilon \log n} \ = \ n^{-\epsilon} \ \longrightarrow 0 \quad \textit{quando } n \longrightarrow \infty.$$

Logo, $X_n = o_p(\log n)$.

No teorema seguinte apresentamos alguns resultados envolvendo a notação $O_p(\cdot)$ e $o_p(\cdot)$.

Teorema 2.2.1 *Sejam* $\{ \ X_n \ \}_{n\geq 1}$, $\{ \ Y_n \ \}_{n\geq 1}$ *seqüências de variáveis aleatórias e* $\{ \ a_n \ \}_{n\geq 1}$, $\{ \ b_n \ \}_{n\geq 1}$ *seqüências de números reais (ou variáveis aleatórias);*

i) se $X_n = o_p(a_n)$, *então* $X_n = O_p(a_n)$;

ii) se $X_n = O_p(a_n)$ *e* $Y_n = O_p(b_n)$, *então:*

a) $X_nY_n = O_p(a_nb_n)$;

b) $\mid X_n \mid^s = O_p(\mid a_n \mid^s), \ \forall \ s > 0$;

c) $X_n + Y_n = O_p\big(\ \max\{|a_n|, |b_n|\}\big)$;

iii) se $X_n = o_p(a_n)$ *e* $Y_n = o_p(b_n)$ *então:*

(a) $X_nY_n = o_p(a_nb_n)$;

(b) $\mid X_n \mid^s = o_p(\mid a_n \mid^s), \ \forall \ s > 0$;

(c) $X_n + Y_n = o_p\big(\ \max\{|a_n|, |b_n|\}\big)$;

iv) se $X_n = O_p(a_n)$ *e* $Y_n = o_p(b_n)$, *então:* $X_nY_n = o_p(a_nb_n)$.

Demonstração: Provaremos apenas os itens `ii`) e `iv`); os demais ficam como exercício para o leitor.

`ii`) Pela Definição 2.2.1, para todo número real $\eta > 0$ existem números reais positivos K_1 e K_2 e existem números inteiros positivos n_0 e n_1 tais que:

$$\mathbb{P}\big(\mid X_na_n \mid \geq K_1 \ \big) \leq \eta 2, \ \forall n \geq n_0 \text{ e } \mathbb{P}\big(\mid Y_nb_n \mid \geq K_2 \ \big) \leq \eta 2, \ \forall n \geq n_1.$$

Logo, $\forall \ n \geq \max\{n_0, n_1\}$,

$$\mathbb{P}\big(\mid \frac{X_nY_n}{a_nb_n} \mid \geq K_1K_2 \ \big) \leq \mathbb{P}\big(\mid \frac{X_n}{a_n} \mid \geq K_1 \text{ ou } \mid \frac{Y_n}{b_n}| \geq K_2 \ \big) \leq$$

$$\mathbb{P}\Big(\mid \frac{X_n}{a_n} \geq K_1 \Big) \ + \ \mathbb{P}\Big(\mid \frac{Y_n}{b_n} \mid \geq K_2 \Big) < \eta,$$

o que prova `ii a`).

Por outro lado, $\forall s > 0$, temos

$$\mathbb{P}\Big(\frac{|X_n|^s}{|a_n|^s} \geq K_1^s \Big) \ = \ \mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K_1 \Big) < \eta, \ \forall n \geq n_0,$$

o que prova `ii b`).

Finalmente, sejam $m_n = \max\{|a_n|, |b_n|\}$, $\forall n \geq 1$ e $K = \max\{K_1, K_2\}$; então pelo Teorema A.2 do Apêndice A, $\forall \ n \geq \max\{n_0, n_1\}$,

$$\mathbb{P}\Big(\mid \frac{X_n + Y_n}{m_n} \mid \geq 2K \Big) \leq \mathbb{P}\Big(\mid \frac{X_n}{m_n} \mid \geq K \Big) \ + \ \mathbb{P}\Big(\mid \frac{Y_n}{m_n} \mid \geq K \Big) \leq$$

$$\mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K \Big) \ + \ \mathbb{P}\Big(\mid \frac{Y_n}{b_n} \mid \geq K \Big) \ \leq$$

$$\mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K_1 \Big) \ + \ \mathbb{P}\Big(\mid \frac{Y_n}{b_n} \mid \geq K_2 \Big) \ \leq \ \eta,$$

o que demonstra `ii c`).

`iv`) Pela Definição 2.2.1, para quaisquer números reais $\epsilon > 0$ e $\eta > 0$ existem um número real $K > 0$ e números inteiros positivos n_0 e n_1, tais que

$$\mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K \Big) \ \leq \ \frac{\eta}{2}, \ \forall n \geq n_0 \quad \text{e} \quad \mathbb{P}\Big(\mid \frac{Y_n}{b_n} \mid \geq \epsilon K \Big) \ < \ \frac{\eta}{2}, \ \forall n \geq n_1.$$

Logo, $\forall \ n \geq \max\{n_0, n_1\}$,

$$\mathbb{P}\Big(\mid \frac{X_n Y_n}{a_n b_n} \mid \geq \epsilon \Big) \ \leq \ \mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K \ \text{ ou } \ \mid \frac{Y_n}{b_n} \mid \geq \epsilon K \Big) \ \leq$$

$$\mathbb{P}\Big(\mid \frac{X_n}{a_n} \mid \geq K \Big) \ + \ \mathbb{P}\Big(\mid \frac{Y_n}{b_n} \mid \geq \epsilon K \Big) \ < \ \eta,$$

o que completa a demonstração.

Outras propriedades são apresentadas no final desta seção sob a forma de exercícios.

No sentido de estender esses conceitos para o caso vetorial, consideremos a seguinte definição:

Definição 2.2.2 *Sejam* $\{\ \boldsymbol{X}_n\ \}_{n\geq 1} = \{(X_{n1}, X_{n2}, \ldots, X_{np})\}_{n\geq 1}$ *uma seqüência de vetores aleatórios* $p-$*dimensionais* $(p \geq 2)$ *e* $\{\ b_n\ \}_{n\geq 1}$ *uma seqüência de números reais. Diremos que*

i) $\boldsymbol{X}_n = O_p\ (b_n)$ *se* $||\ \boldsymbol{X}_n\ || = O_p\ (b_n)$;

ii) $\boldsymbol{X}_n = o_p\ (b_n)$ *se* $||\ \boldsymbol{X}_n\ || = o_p\ (b_n)$;

A seguir apresentamos um resultado que pode ser utilizado para reduzir o caso vetorial ao caso unidimensional.

Teorema 2.2.2 *Sejam* $\{\ \boldsymbol{X}_n\ \}_{n\geq 1}$ *uma seqüência de vetores aleatórios* $p-$*dimensionais e* $\{\ b_n\ \}_{n\geq 1}$ *uma seqüência de números reais; então:*

i) $\boldsymbol{X}_n = O_p\ (b_n)$ *se, e somente se,* $X_{nj} = O_p\ (b_n)$ *para* $j = 1, 2, \ldots, p$;

ii) $\boldsymbol{X}_n = o_p\ (b_n)$ *se, e somente se,* $X_{nj} = o_p\ (b_n)$ *para* $j = 1, 2, \ldots, p$.

Demonstração: Provaremos apenas o item `i`); deixamos a demonstração de `ii`) como exercício para o leitor. Pela Definição 2.2.2, para todo número real $\eta > 0$ existem um número real $K > 0$ e um número inteiro positivo n_0, tais que

$$\mathbb{P}\ \Big(\ \frac{||\ \boldsymbol{X}_n\ ||}{|\ b_n\ |} \geq K\ \Big)\ \leq\ \eta, \quad \forall n \geq n_0.$$

Logo, para $j = 1, 2, \ldots, p$, temos:

$$\mathbb{P}\ \Big(\ \frac{|\ X_{nj}\ |}{|\ b_n\ |} \geq K\ \Big)\ \leq\ \mathbb{P}\ \Big(\ \frac{||\ \mathbf{X}_n\ ||}{|\ b_n\ |} \geq K\ \Big)\ \leq\ \eta, \quad \forall n \geq n_0.$$

Consequentemente $X_{nj} = O_p(b_n), j = 1, 2, \ldots, p$. Agora vamos supor que $X_{nj} = O_p(b_n), j = 1, 2, \ldots, p$; então para cada j e para todo número $\eta > 0$ existe um número real $K_j > 0$ e um número inteiro positivo n_j, tal que

$$\mathbb{P}\ \Big(\ \Big|\ \frac{X_{nj}}{b_n}\ \Big|\ \geq\ K_j\ \Big)\ \leq\ \eta p, \quad \forall n \geq n_j.$$

Sejam $K = \max\{K_j, j = 1, \ldots, p\}$ e $n_0 = \max\{n_j, j = 1, \ldots, p\}$. Então, pelo Teorema A.2 do Apêndice A temos:

$$\mathbb{P}\ \Big(\ \frac{||\ \mathbf{X}_n\ ||}{|b_n|}\ \geq\ p^{\frac{1}{2}}K\ \Big)\ =\ \mathbb{P}\ \Big(\ \sum_{j=1}^{p} \frac{X_{nj}^2}{b_n^2}\ \geq\ pK^2\ \Big)\ \leq$$

$$\sum_{j=1}^{p} \mathbb{P}\ \Big(\ \frac{X_{nj}^2}{b_n^2}\ \geq\ K^2\ \Big)\ \leq\ \sum_{j=1}^{p} \mathbb{P}\ \Big(\ \frac{|\ X_{nj}\ |}{|b_n|}\ \geq\ K_j\ \Big)\ \leq\ \eta, \quad \forall n \geq n_0$$

e, pela Definição 2.2.2, temos $\mathbf{X}_n \ = \ O_p \ (b_n)$.

No Exemplo 2.2.1 vimos que, se $\{X_n\}_{n\geq 1}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $\mu = E(X_1)$ e $\sigma^2 =$ Var $(X_1) < \infty$, então $\overline{X}_n - \mu = O_p(\ n^{-\frac{1}{2}} \ )$, onde $n^{-\frac{1}{2}} \longrightarrow 0$ quando $n \longrightarrow \infty$. Nesse ponto surge a seguinte questão: se f for uma função real de variável real, o que podemos dizer sobre a *ordem de magnitude em probabilidade* da seqüência $\{ \ f(\overline{X}_n) \ \}_{n\geq 1}$?

A resposta a esta questão pode ser obtida através do seguinte teorema:

Teorema 2.2.3 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias, $\{a_n\}_{n\geq 1}$ uma seqüência de números reais positivos com $a_n \longrightarrow 0$ quando $n \longrightarrow \infty$, e f uma função real de variável real derivável até a ordem k em um intervalo que contém um ponto a. Se $X_n - a = O_p(a_n)$, então*

$$f(X_n) \ = \ \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} \ (X_n - a)^j \ + \ o_p(a_n^k).$$

Demonstração: Pela fórmula de Taylor,

$$f(x) \ = \ \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} \ (x - a)^j \ + \ R_k(x),$$

onde $R_k(x) = o\big((x-a)^k\big)$ quando $x \longrightarrow a$ (veja (1.3.2) e (1.3.6)). Além disso, pela Definição 2.2.1, para todo número real $\eta > 0$ existem um número real $K > 0$ e um número inteiro positivo n_0, tal que $\mathbb{P}\ \big(\ |X_n - a| \ < \ K a_n \ \big) \ \geq \ 1 - \eta, \ \ \forall \ n \geq n_0$. Por outro lado, pela Definição 1.3.1, para todo número real $\epsilon > 0$ existe um número real $\delta > 0$ tal que $|R_k(x)||x-a|^k \ < \ \epsilon K^k$, para todo x satisfazendo a condição $0 < |x - a| < \delta$ e, como $a_n \longrightarrow 0$ quando $n \longrightarrow \infty$, existe um número inteiro positivo n_1 tal que $K a_n < \delta$, $\forall n \geq n_1$. Logo,

$$A_n \ = \ \{ \ x : \ \frac{|R_k(x)|}{a_n^k} \ < \ \epsilon \ \} \ \supseteq \ \{ \ x : \ \frac{|x-a|}{a_n} \ < \ K \ \} \ = \ B_n, \ \ \forall \ n \geq n_1$$

o que implica:

$$\mathbb{P}\ \Big(\ \frac{|R_k(X_n)|}{a_n^k} \ < \ \epsilon \ \Big) \ = \ \mathbb{P}\big(\ X_n \in A_n \ \big) \ \geq \ \mathbb{P}\big(\ X_n \in B_n \ \big) \ \geq \ 1 - \eta, \ \forall \ n \geq \max\{n_0, n_1\};$$

ou seja, $R_k(X_n) \ = \ o_p(\ a_n^k \ )$ o que prova o teorema.

Exemplo 2.2.3 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $\mu = E(X_1), \sigma^2 =$ Var $(X_1) < \infty$ e f uma função real de* *variável real derivável em um intervalo que contém μ. Como $\overline{X}_n - \mu \;=\; O_p(\ n^{-\frac{1}{2}}\ )$, (ver Exemplo 2.2.1) então, pelo Teorema 2.2.3,*

$$f(\overline{X}_n) \;=\; f(\mu) \;+\; f'(\mu)(\overline{X}_n - \mu) \;+\; o_p(\ n^{-\frac{1}{2}}\ ).$$

Por outro lado, como $f'(\mu)(\overline{X}_n - \mu) \;+\; o_p(\ n^{-\frac{1}{2}}\ ) = O_p(\ n^{-\frac{1}{2}}\ )$ segue que

$$f(\overline{X}_n) = f(\mu) \;+\; O_p(\ n^{-\frac{1}{2}}\ ).$$

Exercícios

Exercício 2.2.1 *Prove os itens **i)** e **iii)** do Teorema 2.2.1.*

Exercício 2.2.2 *Prove o item **ii)** do Teorema 2.2.2.*

Exercício 2.2.3 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias com distribuição de probabilidade dada por:*

$$\mathbb{P}(X_n = 1) \;=\; 1n \quad e \quad \mathbb{P}(X_n = 0) \;=\; 1 - 1n, \;\forall\; n \geq 1.$$

Prove que $X_n = o_p(1)$.

Exercício 2.2.4 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $X_1 \sim$ Exponencial (λ). Seja $Y_n = \min\{X_j, j = 1, 2, \ldots, n\}, n \geq 1$. Prove que $Y_n \;=\; O_p\ (\ n^{-1}\ ) \;=\; o_p\ (\ n^{\alpha}\ ),\ \forall\ \alpha > -1$.*

Sugestão*: $Y_n \sim$ Exponencial $(n\lambda),\ \forall\ n \geq 1$.*

Exercício 2.2.5 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias com $\mu_n = E(X_n)$ e $\sigma_n^2 = Var\ (X_n) < \infty, \forall n \geq 1$. Prove que $X_n - \mu_n \;=\; O_p\ (\ \sigma_n\ )$.*

Sugestão*: Use a desigualdade de Chebychev.*

Exercício 2.2.6 *Prove que se, $\{X_n\}_{n\geq 1}$ for uma seqüência de variáveis aleatórias e $\{a_n\}_{n\geq 1}$ for uma seqüência de números reais tais que $E(X_n^2) \;=\; O(a_n^2)$, então $X_n \;=\; O_p\ (\ a_n\ )$.*

Sugestão*: Use a desigualdade de Chebychev.*

Exercício 2.2.7 *Se $\{X_n\}_{n\geq 1}$ for uma seqüência de variáveis aleatórias e $\{a_n\}_{n\geq 1}$ for uma seqüência de números reais tais que, $E(X_n) \;=\; O(a_n)$ e $Var\ (X_n) \;=\; O\ (a_n^2)$, então $X_n \;=\; O_p\ (\ a_n\ )$.*

Sugestão*: Prove que $E(X_n^2) \;=\; O(\ a_n^2\ )$ e use o Exercício 2.2.6.*

Exercício 2.2.8 *Se $\{X_n\}_{n\geq 1}$ for uma seqüência de variáveis aleatórias tal que $E(X_n) \longrightarrow c$, onde c é uma constante e $Var\ (X_n) \longrightarrow 0$ quando $n \longrightarrow \infty$, prove que $X_n - c \;=\; o_p\ (1)$.*

Sugestão*: Use a desigualdade de Chebychev.*

Exercício 2.2.9 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. *Prove que*

$$S_n^2 - \sigma^2 \;=\; O_p\;\Big(\;\frac{1}{\sqrt{n-1}}\;\Big),$$

onde $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Sugestão*:* $(n-1)\; S_n^2\sigma^2 \;\sim\; \chi^2_{(n-1)}$ *e use o Exercício 2.2.6.*

Exercício 2.2.10 *Sejam* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias e* $\{a_n\}_{n\geq 1},\ \{b_n\}_{n\geq 1}$ *seqüências de números reais. Prove que:*

i) se $a_n \;=\; O(b_n)$ *e* $X_n \;=\; O_p\;(a_n)$, *então* $X_n \;=\; O_p\;(b_n)$;

ii) se $a_n \;=\; o(b_n)$ *e* $X_n \;=\; O_p\;(a_n)$, *então* $X_n \;=\; o_p\;(b_n)$;

iii) se $a_n \;=\; O(b_n)$ *e* $X_n \;=\; o_p\;(a_n)$, *então* $X_n \;=\; o_p\;(b_n)$;

iv) se $a_n \;=\; o(b_n)$ *e* $X_n \;=\; o_p\;(a_n)$, *então* $X_n \;=\; o_p\;(b_n)$;

Exercício 2.2.11 *Sejam* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias,* $\{a_n\}_{n\geq 1}$ *uma seqüência de números reais positivos com* $a_n \longrightarrow 0$ *quando* $n \longrightarrow \infty$, *e* f *uma função real de variável real derivável até a ordem* k *em um intervalo que contém um ponto* a. *Prove que: se* $X_n - a \;=\; O_p\;(a_n)$, *então* $f(X_n) \;=\; \sum_{j=0}^{k-1}\frac{f^{(j)}(a)}{j!}\;(X_n - a)^j \;+\; O_p\;(a_n^k)$.

Sugestão*: Use o Teorema 2.2.3.*

2.3 Modos de Convergência Estocástica

O conceito de convergência de seqüências de variáveis aleatórias admite diferentes formas. Estes conceitos estão ligados à própria natureza estocástica das seqüências em estudo e sua utilização depende essencialmente dos objetivos que temos em mente. Nesta seção apresentaremos *formalmente* as definições dos diferentes modos de convergência estocástica, indicando suas inter-relações e tentando identificar sua importância nas diferentes aplicações.

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

Para indicar que $\{X_n\}_{n\geq 1}$ converge em probabilidade para X usaremos a notação: $X_n \xrightarrow{\mathbb{P}} X$. Essencialmente, $X_n \xrightarrow{\mathbb{P}} X$ significa que, para todo n suficientemente grande, X_n e X são aproximadamente iguais com alta probabilidade; pela Definição 2.2.1, também podemos notar que, $X_n \xrightarrow{\mathbb{P}} X$ é equivalente à $X_n - X = o_p(1)$.

Exemplo 2.3.1 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $\mu = E(X_1)$ *e* $\sigma^2 = Var(X_1) < \infty$.

Como $\overline{X}_n - \mu = O_p(n^{-\frac{1}{2}})$ *(ver Exemplo* **2.2.1** *) e* $n^{-\frac{1}{2}} = o(1)$ *segue que* $\overline{X}_n - \mu = o_p(1)$ *(ver Exercício 2.2.10), ou seja, que* $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$.

Exemplo 2.3.2 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias tal que* $\mu_n = E(X_n) \longrightarrow c$, *onde* c *é uma constante e* $\sigma_n^2 = Var(X_n) \longrightarrow 0$ *quando* $n \longrightarrow \infty$. *Pela desigualdade de Chebychev (Teorema A.1 do Apêndice A), para todo número real* $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{E\big((X_n - c)^2\big)}{\epsilon^2} = \frac{\sigma_n^2 + (\mu_n - c)^2}{\epsilon^2} \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Logo, $X_n \xrightarrow{\mathbb{P}} c$.

A definição seguinte generaliza esse conceito para vetores aleatórios.

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

Para indicar que $\{\mathbf{X}_n\}_{n\geq 1}$ converge em probabilidade para $\mathbf{X}_0$ usaremos a notação: $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}_0$. Logo, $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}_0$ é equivalente a $||\mathbf{X}_n - \mathbf{X}_0|| = o_p(1)$.

O teorema que apresentamos a seguir tem bastante importância prática, porque permite reduzir o estudo da convergência em probabilidade no caso multidimensional para o caso unidimensional, onde, em geral, os resultados são mais facilmente aplicáveis.

Teorema 2.3.1 *Sejam* $\{\boldsymbol{X}_n\}_{n\geq 1}$ *uma seqüência de vetores aleatórios e* $\boldsymbol{X}_0$ *um vetor aleatório. Então:*

$$\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}_0 \text{ se e somente se } X_{nj} \xrightarrow{\mathbb{P}} X_{0j}, \quad \text{para } j = 1, 2, \ldots, p.$$

Demonstração: Suponhamos que $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}_0$. Então, para todo número real $\epsilon > 0$ e para $j = 1, 2, \ldots, p$, $\mathbb{P}(|X_{nj} - X_{0j}| \geq \epsilon) \ \leq \ \mathbb{P}(||\mathbf{X}_n - \mathbf{X}_0|| \geq \epsilon) \longrightarrow 0$ quando $n \longrightarrow \infty$, o que implica $X_{nj} \xrightarrow{\mathbb{P}} X_{0j}$, $j = 1, 2, \ldots, p$. Reciprocamente, suponhamos que $X_{nj} \xrightarrow{\mathbb{P}} X_{0j}$ para $j = 1, 2, \ldots, p$. Então, para todo número real $\epsilon > 0$ e para $j = 1, 2, \ldots, p$,

$$\mathbb{P}(|X_{nj} - X_{0j}| \geq \frac{\epsilon}{\sqrt{p}}) \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Logo, pelo Teorema A.2 do Apêndice A,

$$\mathbb{P}(||\mathbf{X}_n - \mathbf{X}_0|| \geq \epsilon) = \mathbb{P}(||\mathbf{X}_n - \mathbf{X}_0||^2 \geq \epsilon^2) = \mathbb{P}\big(\sum_{j=1}^{p}(X_{nj} - X_{0j})^2 \geq \epsilon^2\big) \leq$$

$$\sum_{j=1}^{p}\mathbb{P}\big((X_{nj} - X_{0j})^2 \geq \frac{\epsilon^2}{p}\big) = \sum_{j=1}^{p}\mathbb{P}\big(|X_{nj} - X_{0j}| \geq \frac{\epsilon}{\sqrt{p}}\big) \longrightarrow 0 \ \text{ quando } \ n \longrightarrow \infty,$$

o que implica $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}_0$.

Dadas uma seqüência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ e uma variável aleatória X (possivelmente degenerada) lembremos que, $X, X_1, X_2, \ldots$ estão definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$. Denotemos por $(X_n \longrightarrow X)$ o evento $\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)$ quando $n \longrightarrow \infty\}$. Então podemos definir um segundo modo de convergência estocástica.

Definição 2.3.3 *Uma seqüência de variáveis aleatórias* $\{X_n\}_{n\geq 1}$ converge quase certamente *(* ou converge em quase toda parte *) para uma variável aleatória* X *se*

$\mathbb{P}(X_n \longrightarrow X) = 1.$

Para indicar que $\{X_n\}_{n\geq 1}$ converge quase certamente para X usaremos a notação: $X_n \xrightarrow{q.c.} X$. O teorema que apresentamos a seguir é uma ferramenta bastante útil no estudo da convergência quase certa de variáveis aleatórias.

Teorema 2.3.2 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias. Então,* $X_n \xrightarrow{q.c.} 0$ *se e somente se* $\mathbb{P}\big(\ |X_n| \geq \frac{1}{m}$ *infinitas vezes* $\big) = 0, \ m = 1, 2, \ldots$

Observação:

$$\Big(\ |X_n| \geq \frac{1}{m} \text{ infinitas vezes } \Big) = \limsup_{n \longrightarrow \infty} \Big(\ |X_n| \geq \frac{1}{m}\ \Big) = \bigcap_{n \geq 1} \bigcup_{k \geq n} \Big(\ |X_k| \geq \frac{1}{m}\ \Big)$$

Demonstração: Sejam $A_m = \bigcap_{n\geq 1} \bigcup_{k \geq n} \big(\ |X_k| \geq \frac{1}{m}\ \big),\ m \geq 1$ e $A = \bigcup_{m \geq 1} A_m$. Como $A^c = (X_n \longrightarrow 0)$ então, $\mathbb{P}(X_n \longrightarrow 0) = 1$ se e somente se $\mathbb{P}(A) = 0$, ou equivalentemente se e somente se $\mathbb{P}(A_m) = 0,\ m = 1, 2, \ldots$ e o teorema está provado.

Os dois resultados seguintes são conseqüências desse teorema.

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

Demonstração: Pelo Teorema 2.3.2, $X_n \xrightarrow{q.c.} X$, ou equivalentemente $X_n - X \xrightarrow{q.c.} 0$, se e somente se $\mathbb{P}\big(\ |X_n - X| \geq \frac{1}{m}$ infinitas vezes $\big) = 0$ para $m = 1, 2, \ldots$, ou equivalentemente se e somente se $\mathbb{P}\big(\ |X_n - X| \geq \epsilon$ infinitas vezes $\big) = 0$ para todo número real $\epsilon > 0$. Então, $X_n \xrightarrow{q.c.} X$, se e somente se para todo número real $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} \mathbb{P}\big(\ |X_k - X| \geq \epsilon \text{ para algum } k \geq n\ \big) = \lim_{n \longrightarrow \infty} \mathbb{P}\Big(\bigcup_{k \geq n} (\ |X_k - X| \geq \epsilon\ )\ \Big) =$$

$$\mathbb{P}\Big(\lim_{n \longrightarrow \infty} \bigcup_{k \geq n} (\ |X_k - X| \geq \epsilon\ )\ \Big) = \mathbb{P}\Big(\bigcap_{n \geq 1} \bigcup_{k \geq n} (\ |X_k - X| \geq \epsilon\ )\ \Big) =$$

$$\mathbb{P}\big(\ |X_k - X| \geq \epsilon \text{ infinitas vezes } \big) = 0.$$

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

Demonstração: Suponhamos que $\sum_{n=1}^{\infty} \mathbb{P}\big(\ |X_n - X| \geq \epsilon\ \big) < \infty$ para todo $\epsilon > 0$. Então, pelo Lema de Borel-Cantelli (Teorema A.3 do Apêndice A), $\mathbb{P}\big(\ |X_n - X| \geq \epsilon$ infinitas vezes $\big) = 0$, para todo número real $\epsilon > 0$, e o resultado segue do Teorema 2.3.2.

Exemplo 2.3.3 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias cujas distribuições de probabilidade são dadas por:*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad e \quad \mathbb{P}(X_n = n^2) = \frac{1}{n^2} \quad , \ \forall n \geq 1.$$

Para todo número real $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(\ X_n \geq \epsilon\ ) \leq \sum_{n=1}^{\infty} \mathbb{P}(\ X_n = n^2\ ) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

e, pelo Teorema 2.3.4, $X_n \xrightarrow{q.c.} 0$.

Exemplo 2.3.4 *Sejam* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $X_1 \sim$ *Exponencial* (1). *Vimos no Exemplo* 2.2.2 *que* $X_n \log n \xrightarrow{\mathbb{P}} 0$. *Por outro lado, dado um número real* ϵ, $0 < \epsilon \leq 1$, *os eventos* $\big(X_n \log n \geq \epsilon\big)$, $n = 2, 3, \ldots$ *são independentes e* $\sum\limits_{n=2}^{\infty} \mathbb{P}\big(X_n \log n \geq \epsilon\big) = \sum\limits_{n=2}^{\infty} 1 n^{\epsilon} = \infty$. *Logo, pelo Lema de Borel-Cantelli (Teorema A.3 do Apêndive A),*

$$\mathbb{P}\Big(\ X_n \log n \geq \epsilon \textit{ infinitas vezes }\ \Big) = 1,$$

o que implica, pelo Teorema 2.3.2, que $X_n \log n \overset{q.c.}{\not\longrightarrow} 0$.

O exemplo anterior ilustra o fato de que convergência em probabilidade não implica 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.

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

Demonstração: Para todo número real $\epsilon > 0$, temos:

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \mathbb{P}(|X_k - X| \geq \epsilon \text{ para algum } k \geq n),\ \forall n \geq 1$$

e o resultado segue do Teorema 2.3.3.

O conceito de convergência quase certa também pode ser facilmente estendido para o caso vetorial, como podemos ver através da seguinte definição:

Definição 2.3.4 *Diremos que a seqüência* $\{\boldsymbol{X}_n\}_{n\geq 1} = \{(X_{n1}, X_{n2}, \ldots, X_{np})\}_{n\geq 1}$ *de vetores aleatórios* converge quase certamente *(* ou converge em quase toda parte *) para um vetor aleatório* $\boldsymbol{X}_0$ *(possivelmente degenerado) se a seqüência de variáveis aleatórias* $\{||\boldsymbol{X}_n - \boldsymbol{X}_0||\}_{n\geq 1}$ convergir quase certamente *para zero.*

Usaremos a notação $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0$ para indicar que $\{\boldsymbol{X}_n\}_{n\geq 1}$ converge quase certamente para $\boldsymbol{X}_0$. Logo, $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0$ se $\mathbb{P}(\ ||\boldsymbol{X}_n - \boldsymbol{X}_0|| \longrightarrow 0\ ) = 1$.

O próximo teorema também pode ser utilizado para reduzir o estudo da convergência quase certa no caso multidimensional para o caso unidimensional.

Teorema 2.3.6 *Sejam* $\{\boldsymbol{X}_n\}_{n\geq 1}$ *uma seqüência de vetores aleatórios e* $\boldsymbol{X}_0$ *um vetor aleatório. Então:*

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

Demonstração: Suponhamos que $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0$. Então, como

$$(X_{nj} \longrightarrow X_{0j}) \supset (||\boldsymbol{X}_n - \boldsymbol{X}_0|| \longrightarrow 0) \text{ para } j = 1, 2, \ldots, p,$$

segue que $\mathbb{P}(X_{nj} \longrightarrow X_{0j}) = 1$ para $j = 1, 2, \ldots, p$.

Suponhamos agora que $X_{nj} \xrightarrow{q.c.} X_{0j}$ para $j = 1, 2, \ldots, p$. Seja $\mathcal{C} = \bigcap_{j=1}^{p} (X_{nj} \longrightarrow X_{0j})$. Então $\mathbb{P}(\mathcal{C}) = 1$ e o resultado segue do fato que $(||\mathbf{X}_n - \mathbf{X}_0|| \longrightarrow 0) \supset \mathcal{C}$.

Uma questão de interesse, tanto sob o ponto de vista teórico quanto sob o ponto de vista aplicado, diz respeito ao comportamento de seqüências de variáveis aleatórias convergentes (em probabilidade ou em quase toda parte) com relação aos momentos (quando existem) . Suponhamos, por exemplo, que uma seqüência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ converge quase certamente para uma variável aleatória X e que $E(|X_n|) < \infty$, $\forall n \geq 1$. Será que $E(|X|) < \infty$ e $E(X_n) \longrightarrow E(X)$ quando $n \longrightarrow \infty$? A resposta a esta questão nem sempre é afirmativa como ilustra o exemplo seguinte:

Exemplo 2.3.5 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias cujas distribuições de probabilidade são dadas por:*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad e \quad \mathbb{P}(X_n = n^2) = \frac{1}{n^2}, \ \forall n \geq 1.$$

No Exemplo 2.3.3 vimos que $X_n \xrightarrow{q.c.} 0$. *Contudo,* $E(X_n) \longrightarrow 1$ *quando* $n \longrightarrow \infty$. *Por outro lado,* $X_n \xrightarrow{q.c.} 0$ *implica* $X_n^2 \xrightarrow{q.c.} 0$, *mas* $E(X_n^2) = n^2 \longrightarrow \infty$ *quando* $n \longrightarrow \infty$.

Nesse contexto, um terceiro modo de convergência de seqüências de variáveis aleatórias pode ser introduzido:

Definição 2.3.5 *Sejam* $X, X_1, X_2, \ldots$ *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 seqüência* $\{X_n\}_{n\geq 1}$ converge em média de ordem r *para* X *se* $E(|X_n - X|^r) \longrightarrow 0$ *quando* $n \longrightarrow \infty$.

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

Exemplo 2.3.6 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $\mu = E(X_1)$ *e* $\sigma^2 = Var(X_1) < \infty$. *Então,* $E(\overline{X}_n - \mu)^2 = Var(\overline{X}_n) = \sigma^2 n \longrightarrow 0$ *quando* $n \longrightarrow \infty$ *e consequentemente,* $\overline{X}_n \xrightarrow{m.2.} \mu$.

O Exemplo 2.3.5 mostra que convergência em quase toda parte (ou em probabilidade) não implica convergência em média ou em média quadrática. Contudo, vale o seguinte resultado:

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

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

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

Demonstração: Pela Desigualdade de Chebychev (Teorema A.1 do Apêndice A) temos que, para todo número real $\epsilon > 0$, $\mathbb{P}(|X_n - X| \geq \epsilon) \leq E(|X_n - X|^r)\epsilon^r$, $\forall n \geq 1$. Logo,

i) como $E(|X_n - X|^r) \longrightarrow 0$ quando $n \longrightarrow \infty$, então $\mathbb{P}(|X_n - X| \geq \epsilon) \longrightarrow 0$ quando $n \longrightarrow \infty$, $\forall \epsilon > 0$, o que prova o item i);

ii) como $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{1}{\epsilon^r} \sum_{n=1}^{\infty} E(|X_n - X|^r) < \infty$, $\forall \epsilon > 0$, então, pelo Teorema 2.3.4, $X_n \xrightarrow{q.c.} X$ o que prova o item ii).

Consideremos situações em que nosso interesse esteja relacionado com a estimação de algum parâmetro. Os conceitos de convergência estocástica discutidos até aqui constituem ferramentas úteis para avaliar se os estimadores propostos se "aproximam" do verdadeiro valor do parâmetro, com o aumento do tamanho da amostra na qual eles estão baseados. Sob o ponto de vista prático, no entanto, a verificação dessa propriedade pode não ser suficiente; em muitas situações, a construção de intervalos de confiança (aproximados) para o parâmetro é necessária e para isso é preciso estudar o comportamento da distribuição dos estimadores em questão à medida que aumenta o tamanho da amostra na qual eles estão baseados. Esse novo aspecto do comportamento estatístico de estimadores para grandes amostras está relacionado com o conceito de *convergência em distribuição* (ou em lei), que essencialmente envolve a convergência de uma seqüência de funções distribuição. Antes de definirmos este modo de convergência, consideremos o seguinte exemplo:

Exemplo 2.3.7 *Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias definidas em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ por: $X \equiv 0$ e $X_n = 1n$, $n = 1, 2, \ldots$. Claramente $X_n \xrightarrow{q.c.} X$ e $X_n \xrightarrow{\mathbb{P}} X$ e além disso para $n = 1, 2, \ldots$ as funções distribuição de X_n e X são dadas, respectivamente, por $F_n(x) = 0$ se $x < 1n$, $F_n(x) = 1$ se $x \geq 1n$ e $F(x) = 0$ se $x < 0$, $F(x) = 1$ se $x \geq 0$. Como nosso interesse está dirigido para o comportamento da seqüência de funções distribuição $\{F_n\}_{n\geq 1}$, observemos que, para todo $x < 0, F_n(x) \longrightarrow F(x) = 0$ quando $n \longrightarrow \infty$ e para todo $x > 0, F_n(x) \longrightarrow F(x) = 1$ quando $n \longrightarrow \infty$, mas $F_n(0) \longrightarrow 0 \neq F(0) = 1$. Logo, observando que o único ponto de descontinuidade de F é o ponto zero, podemos concluir que:*

i) para todo ponto de continuidade*, x, de F, a seqüência* $\{F_n(x)\}_{n\geq 1}$ *converge para* $F(x)$;

ii) a seqüência $\{F_n(0)\}_{n\geq 1}$ não converge *para* $F(0)$.

Portanto, se tivermos a intenção de definir um modo de convergência para seqüências de variáveis aleatórias, baseado na convergência de suas funções distribuição, devemos tomar cuidado com os pontos de descontinuidade da função distribuição limite. Isto motiva a seguinte definição:

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

Para indicar que $\{X_n\}_{n\geq 1}$ converge em distribuição para X usaremos a notação: $X_n \xrightarrow{\mathcal{D}} X$.

Exemplo 2.3.8 *Seja $\{X_n\}_{n\geq 1}$ umna seqüência de variáveis aleatórias independentes e identicamente distribuídas com $X_1 \sim$ Exponencial (λ). Seja $Y_n = \min\{X_1, \ldots, X_n\}$, para todo $n \geq 1$. Como $Y_n \sim$ Exponencial $(n\lambda)$, então $nY_n \sim$ Exponencial (λ), para todo $n \geq 1$. Logo, $nY_n \xrightarrow{\mathcal{D}} Y_1$.*

Também convém notar que:

i) Se $X_n \xrightarrow{\mathcal{D}} X$ e F é contínua, então, pela Definição 2.3.6, para todo x real $\{F_n(x)\}_{n\geq 1}$ converge para $F(x)$. Além disso, é possível mostrar que $\{F_n\}_{n\geq 1}$ *converge uniformemente* para F.

ii) Como convergência em distribuição é definida em termos da convergência das funções distribuição, não é necessário que as variáveis aleatórias mencionadas na Definição 2.3.6 estejam definidas em um mesmo espaço de probabilidade.

Uma condição necessária e suficiente para que uma seqüência de variáveis aleatórias convirja em distribuição, (em termos de suas funções características) é dada pelo seguinte teorema:

Teorema 2.3.8 *Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com funções distribuição $F, F_1, F_2, \ldots$ e funções características $\varphi, \varphi_1, \varphi_2, \ldots$, respectivamente. Então, $X_n \xrightarrow{\mathcal{D}} X$ se, e somente se, para todo $t \in \mathbb{R}$, $\varphi_n(t) \longrightarrow \varphi(t)$ quando $n \longrightarrow \infty$.*

Demonstração: Suponhamos que $X_n \xrightarrow{\mathcal{D}} X$. Então, pelo Teorema de Helly-Bray (Teorema A.4 do Apêndice A), para todo $t \in \mathbb{R}$,

$$\varphi_n(t) = E(e^{itX_n}) = \int e^{itx} dF_n(x) =$$

$$\int \cos(tx) dF_n(x) + i \int \operatorname{sen}(tx) dF_n(x) \longrightarrow \int \cos(tx) dF(x) + i \int \operatorname{sen}(tx) dF(x) =$$

$$E(e^{itX}) = \varphi(t) \text{ quando } n \longrightarrow \infty.$$

Suponhamos agora que, para todo $t \in \mathbb{R}$, $\varphi_n(t) \longrightarrow \varphi(t)$ quando $n \longrightarrow \infty$. Como φ é contínua no ponto zero, segue pelo Teorema da Continuidade de Paul Lévy (Teorema `A.5` do Apêndice `A`) que, para todo ponto de continuidade x, de F, $F_n(x) \longrightarrow F(x)$ quando $n \longrightarrow \infty$, o que prova o teorema.

Como vimos anteriormente, convergência quase certa ou convergência em média de ordem $r(r > 0)$ implicam convergência em probabilidade, mas a recíproca não vale, em geral. O seguinte teorema nos mostra que dentre os modos de convergência definidos acima, convergência em distribuição é o mais fraco.

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

Demonstração: Suponhamos que $X_n \xrightarrow{\mathbb{P}} X$. Sejam $F, F_1, F_2, \ldots$ as funções distribuição de $X, X_1, X_2, \ldots$, respectivamente, e seja x um ponto de continuidade de F. Para todo número real $\epsilon > 0$,

$$(X_n \leq x) = (X_n \leq x, X \leq x + \epsilon) \bigcup (X_n \leq x, X > x + \epsilon).$$

Como

$$(X_n \leq x, X \leq x + \epsilon) \subset (X \leq x + \epsilon) \text{ e } (X_n \leq x, X > x + \epsilon) \subset (|X_n - X| > \epsilon)$$

segue que

$$(X_n \leq x) \subset (X \leq x + \epsilon) \bigcup (|X_n - X| > \epsilon).$$

Logo,

$$F_n(x) = \mathbb{P}(X_n \leq x) \leq F(x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Por outro lado,

$$(X \leq x - \epsilon) = (X \leq x - \epsilon, X_n \leq x) \bigcup (X \leq x - \epsilon, X_n > x).$$

Como

$$(X \leq x - \epsilon, X_n \leq x) \subset (X_n \leq x) \text{ e } (X \leq x - \epsilon, X_n > x) \subset (|X_n - X| > \epsilon)$$

segue que

$$(X \leq x - \epsilon) \subset (X_n \leq x) \bigcup (|X_n - X| > \epsilon).$$

Logo,

$$F(x - \epsilon) = \mathbb{P}(X \leq x - \epsilon) \leq F_n(x) + \mathbb{P}(|X_n - X| > \epsilon)$$

Das desigualdades acima temos que, para todo $\epsilon > 0$ e para todo inteiro $n \geq 1$,

$$F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Como $\mathbb{P}(|X_n - X| > \epsilon) \longrightarrow 0$ quando $n \longrightarrow \infty$, segue que

$$F(x-\epsilon) \leq \liminf_{n\longrightarrow\infty} F_n(x) \leq \limsup_{n\longrightarrow\infty} F_n(x) \leq F(x+\epsilon)$$

Finalmente, como x é um ponto de continuidade de F, segue que

$$F(x) = \lim_{\epsilon\longrightarrow 0} F(x-\epsilon) \leq \liminf_{n\longrightarrow\infty} F_n(x) \leq \limsup_{n\longrightarrow\infty} F_n(x) \leq \lim_{\epsilon\longrightarrow 0} F(x+\epsilon) = F(x),$$

ou seja, $\lim_{n\longrightarrow\infty} F_n(x) = F(x)$ o que prova o teorema.

O exemplo seguinte mostra que, em geral, convergência em distribuição não implica convergência em probabilidade.

Exemplo 2.3.9 *Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas com $X_1 \sim$ Bernoulli (12), isto é , $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = 12$. Então claramente, $X_n \xrightarrow{\mathcal{D}} X$. Logo, como $|X_n - X| \sim$ Bernoulli (12), para todo $n \geq 1$, segue que, para todo número real ϵ, $\;0 < \epsilon < 1$, $\;\mathbb{P}(|X_n - X| \geq \epsilon) = \mathbb{P}(|X_n - X| = 1) = 12$, para todo $n \geq 1$ e consequentemente $X_n \overset{\mathbb{P}}{\not\longrightarrow} X$.*

Contudo, vale o seguinte teorema:

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

Demonstração: Suponhamos que X seja uma variável aleatória degenerada em um ponto c e que $X_n \xrightarrow{\mathcal{D}} X$. Sejam $F, F_1, F_2, \ldots$ as funções distribuição de $X, X_1, X_2, \ldots$, respectivamente. Então, $F(x) = 0$ se $x < c$, $F(x) = 1$ se $x \geq c$ e, como todo ponto $x \neq c$ é um ponto de continuidade de F, segue que $F_n(x) \longrightarrow F(x) = 0$ se $x < c$ e $F_n(x) \longrightarrow F(x) = 1$ se $x \geq c$, quando $n \longrightarrow \infty$. Logo, para todo número real $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|X_n - c| > \epsilon) \leq \mathbb{P}(X_n > c+\epsilon) + \mathbb{P}(X_n \leq c-\epsilon) = 1 - F_n(c+\epsilon) - F_n(c-\epsilon) \longrightarrow 0$ quando $n \longrightarrow \infty$, o que prova o teorema.

Um resultado que será utilizado em muitas aplicações é dado pelo seguinte teorema:

Teorema 2.3.11 *Se uma seqüência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ converge em distribuição para uma variável aleatória X, então $X_n = O_p(1)$.*

Demonstração: Sejam $F_n,\ n = 1, 2, \ldots$ e F as funções distribuição de X_n e X, respectivamente. Para todo número real $\eta > 0$ existe um número real $K > 0$, tal que $-K$ e K são pontos de continuidade de F e $F(K) - F(-K) \geq 1 - \eta$. Como $X_n \xrightarrow{\mathcal{D}} X$, existem números inteiros positivos $n_0 = n_0(\eta, K)$ e $n_1 = n_1(\eta, K)$ tais que,
$F(-K) - \eta < F_n(-K) < F(-K) + \eta$ para todo $n \geq n_0$ e $F(K) - \eta < F_n(K) < F(K) + \eta$ para todo $n \geq n_1$. Logo, para todo $n \geq \max\{n_0, n_1\}$,

$$\mathbb{P}(|X_n| \leq K) \geq \mathbb{P}(-K < X_n \leq K) = F_n(K) - F_n(-K) > F(K) - \eta - F(-K) - \eta \geq 1 - 3\eta.$$

Como η é arbitrário o teorema está provado.

A seguinte definição é uma generalização da Definição 2.3.6 para o caso vetorial:

Definição 2.3.7 *Sejam* $\boldsymbol{X}_0 = (X_{01}, X_{02}, \ldots, X_{0p})$ *e* $\boldsymbol{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np}),\ n = 1, 2, \ldots$ *vetores aleatórios* $p-$*dimensionais cujas funções distribuição são* $F_{\boldsymbol{X}_0}, F_{\boldsymbol{X}_1}, F_{\boldsymbol{X}_2}, \ldots,$ *respectivamente. Diremos que a seqüência* $\{\boldsymbol{X}_n\}_{n\geq 1}$ converge em distribuição *(* ou em lei *) para* $\boldsymbol{X}_0$ *se, para todo ponto* $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$ *de continuidade de* $F_{\boldsymbol{X}_0}, F_{\boldsymbol{X}_n}(\boldsymbol{x}) \longrightarrow F_{\boldsymbol{X}_0}(\boldsymbol{x})$ *quando* $n \longrightarrow \infty$.

Para indicar que $\{\boldsymbol{X}_n\}_{n\geq 1}$ converge em distribuição para $\boldsymbol{X}_0$ usaremos a notação: $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$.

O próximo teorema é uma extensão do Teorema 2.3.8 para o caso multidimensional.

Teorema 2.3.12 *Sejam* $\boldsymbol{X}_0 = (X_{01}, \ldots, X_{0p})$ *e* $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{np}),\ n = 1, 2, \ldots$ *vetores aleatórios* $p-$*dimensionais cujas funções características são* $\varphi_{\boldsymbol{X}_0}, \varphi_{\boldsymbol{X}_1}, \varphi_{\boldsymbol{X}_2}, \ldots,$ *respectivamente. Então,* $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$ *se e somente se, para todo* $\boldsymbol{t} \in \mathbb{R}^p$, $\varphi_{\boldsymbol{X}_n}(\boldsymbol{t}) \longrightarrow \varphi_{\boldsymbol{X}_0}(\boldsymbol{t})$ *quando* $n \longrightarrow \infty$.

Demonstração: Ver Billingsley (1968), por exemplo.

Geralmente, as aplicações diretas da Definição 2.3.7 e do Teorema 2.3.12 no estudo da convergência em distribuição de seqüências de vetores aleatórios, apresentam sérias dificuldades técnicas. Contudo, o seguinte resultado devido a Cramér e Wold, reduz o problema da convergência em distribuição no caso multidimensional ao caso unidimensional.

Teorema 2.3.13 (Cramér - Wold) *Sejam* $\boldsymbol{X}_0 = (X_{01}, \ldots, X_{0p})$ *e* $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{np}),$ $n = 1, 2, \ldots$ *vetores aleatórios* $p-$*dimensionais. Então* $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$ *se e somente se*

$$\boldsymbol{t}\boldsymbol{X}_n^t = \sum_{j=1}^{p} t_j X_{nj} \ \xrightarrow{\mathcal{D}}\ \sum_{j=1}^{p} t_j X_{0j} = \boldsymbol{t}\boldsymbol{X}_0^t,\ \forall \boldsymbol{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p.$$

Demonstração: Suponhamos que $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$. Então, pelo Teorema 2.3.12, para todo $\boldsymbol{t} \in \mathbb{R}^p$ a sequência das funções características $\{\varphi_{\boldsymbol{t}\boldsymbol{X}_n^t}\}_{n\geq 1}$ das variáveis aleatórias $\boldsymbol{t}\boldsymbol{X}_n^t$ é tal que, para todo $s \in \mathbb{R}$,

$$\varphi_{\boldsymbol{t}\boldsymbol{X}_n^t}(s) = E(e^{is\boldsymbol{t}\boldsymbol{X}_n^t}) = \varphi_{\boldsymbol{X}_n}(s\boldsymbol{t}) \longrightarrow \varphi_{\boldsymbol{X}_0}(s\boldsymbol{t}) = E(e^{is\boldsymbol{t}\boldsymbol{X}_0^t}) = \varphi_{\boldsymbol{t}\boldsymbol{X}_0^t}(s) \text{ quando } n \longrightarrow \infty.$$

Logo, pelo Teorema 2.3.8, $\boldsymbol{t}\boldsymbol{X}_n^t \xrightarrow{\mathcal{D}} \boldsymbol{t}\boldsymbol{X}_0^t$, $\forall \boldsymbol{t} \in \mathbb{R}^p$.

Suponhamos agora que para todo $\boldsymbol{t} \in \mathbb{R}^p$, $\boldsymbol{t}\boldsymbol{X}_n^t \xrightarrow{\mathcal{D}} \boldsymbol{t}\boldsymbol{X}_0^t$. Então, pelo Teorema 2.3.8, a seqüência das funções características $\{\varphi_{\boldsymbol{X}_n}\}_{n\geq 1}$ dos vetores $\boldsymbol{X}_n$ é tal que,para todo $\boldsymbol{t} \in \mathbb{R}^p$,

$$\varphi_{\boldsymbol{X}_n}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}\boldsymbol{X}_n^t}) = \varphi_{\boldsymbol{t}\boldsymbol{X}_n^t}(1) \longrightarrow \varphi_{\boldsymbol{t}\boldsymbol{X}_0^t}(1) = E(e^{i\boldsymbol{t}\boldsymbol{X}_0^t}) = \varphi_{\boldsymbol{X}_0}(\boldsymbol{t}) \text{ quando } n \longrightarrow \infty.$$

Logo, pelo Teorema 2.3.12, $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$ o que completa a demonstração.

Logo, convergência em distribuição de uma seqüência de vetores aleatórios $\{\boldsymbol{X}_n\}_{n\geq 1}$ para um vetor aleatório $\boldsymbol{X}_0$ é equivalente à convergência em distribuição das seqüências das combinações lineares $\{\boldsymbol{t}\boldsymbol{X}_n^t\}_{n\geq 1}$, para $\boldsymbol{t}\boldsymbol{X}_0^t$, para todo $\boldsymbol{t} \in \mathbb{R}^p$.

Para encerrar esta seção consideremos dois tipos de problemas, envolvendo convergência estocástica, com os quais nos deparamos freqüentemente:

i) suponhamos que $\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ sejam vetores aleatórios $p-$dimensionais tais que $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0$ ou $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}_0$ ou $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$ e seja f uma função de $\mathbb{R}^p$ com valores em $\mathbb{R}^m$. O que podemos afirmar sobre a convergência (ou não) da seqüência $\{f(\boldsymbol{X}_n)\}_{n\geq 1}$?

ii) se $\{\boldsymbol{X}_n\}_{n\geq 1}$ e $\{\boldsymbol{Y}_n\}_{n\geq 1}$ são seqüências de vetores aleatórios $p-$dimensionais tais que $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$ e $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y}_0$ o que podemos concluir sobre a convergência da seqüência $\{\boldsymbol{X}_n + \boldsymbol{Y}_n\}_{n\geq 1}$?

Com relação a `i`) veremos que se f for contínua, então $\{f(\boldsymbol{X}_n)\}$ converge para $f(\boldsymbol{X}_0)$ do mesmo modo que $\{\boldsymbol{X}_n\}_{n\geq 1}$ converge para $\boldsymbol{X}_0$; com relação a `ii`), veremos que, se $\boldsymbol{Y}_0 = \boldsymbol{c}$, onde $\boldsymbol{c}$ é um vetor constante de $\mathbb{R}^p$, então $\boldsymbol{X}_n + \boldsymbol{Y}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0 + \boldsymbol{c}$.

Teorema 2.3.14 *Sejam* $\boldsymbol{X}_0 = (X_{01}, X_{02}, \ldots, X_{0p})$ *e* $\boldsymbol{X}_n = (X_{n1}, X_{n2}, \ldots, X_{np})$, $n = 1, 2, \ldots$ *vetores aleatórios* $p-$*dimensionais definidos em um mesmo espaço de probabilidade* $(\Omega, \mathcal{F}, \mathbb{P})$ *e* $f : \mathbb{R}^p \longrightarrow \mathbb{R}^m$ $(m \geq 1)$ *uma função contínua. Então*

i) $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0 \implies f(\boldsymbol{X}_n) \xrightarrow{q.c.} f(\boldsymbol{X}_0)$;

ii) $\boldsymbol{X}_n \xrightarrow{\mathbb{P}} \boldsymbol{X}_0 \implies f(\boldsymbol{X}_n) \xrightarrow{\mathbb{P}} f(\boldsymbol{X}_0)$;

iii) $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0 \implies f(\boldsymbol{X}_n) \xrightarrow{\mathcal{D}} f(\boldsymbol{X}_0)$.

Demonstração: Demonstraremos o teorema para o caso $m = p = 1$; a demonstração para o caso geral é análoga.

`i`) Suponhamos que $X_n \xrightarrow{q.c.} X_0$. Então o evento

$$\mathcal{C} = \{\omega \in \Omega : X_n(\omega) \longrightarrow X_0(\omega) \text{ quando } n \longrightarrow \infty\}$$

é tal que $\mathbb{P}(\mathcal{C}) = 1$ e, como f é contínua,

$$\mathcal{C} \subset \{\omega \in \Omega : f\big(X_n(\omega)\big) \longrightarrow f\big(\mathbf{X}_0(\omega)\big) \text{ quando } n \longrightarrow \infty\}.$$

Logo, $f(X_n) \xrightarrow{q.c.} f(X_0)$ o que prova `i`).

`ii`) Suponhamos que $X_n \xrightarrow{\mathbb{P}} X_0$. Queremos mostrar que $f(X_n) \xrightarrow{\mathbb{P}} f(X_0)$. Vamos admitir que $f(X_n) \not\xrightarrow{\mathbb{P}} f(X_0)$; então, para algum $\epsilon > 0$ e algum $\eta > 0$ existe uma subseqüência $\{f(X_{n_k})\}_{k\geq 1}$ da sequência $\{f(X_n)\}_{n\geq 1}$ tal que,

$$\mathbb{P}\big(|f(X_{n_k}) - f(X_0)| \geq \epsilon\big) > \eta, \ \forall k = 1, 2, \dots \tag{2.3.1}$$

Mas $X_n \xrightarrow{\mathbb{P}} X_0$ implica $X_{n_k} \xrightarrow{\mathbb{P}} X_0$ e então, pelo Teorema A.6 do Apêndice A, existe uma subseqüência $\{X_{n_{k_j}}\}_{j\geq 1}$ de $\{X_{n_k}\}_{k\geq 1}$ tal que $X_{n_{k_j}} \xrightarrow{q.c.} X_0$ quando $j \longrightarrow \infty$. Logo, pelo item `i`), $f(X_{n_{k_j}}) \xrightarrow{q.c.} f(X_0)$ e, como convergência quase certa implica convergência em probabilidade, temos que $f(X_{n_{k_j}}) \xrightarrow{\mathbb{P}} f(X_0)$ o que contradiz (2.3.1). Portanto, $f(X_n) \xrightarrow{\mathbb{P}} f(X_0)$ o que prova `ii`).

`iii`) Suponhamos que $X_n \xrightarrow{\mathcal{D}} X_0$ Como, para todo $t \in \mathbb{R}$, as funções $\cos\big(tf(x)\big)$ e $\text{sen}\big(tf(x)\big)$ são contínuas e limitadas em $\mathbb{R}$, segue que a seqüência das funções características de $f(X_n)$, $\{\varphi_{f(X_n)}\}_{n\geq 1}$, converge para a função característica de $f(X_0)$, $\varphi_{f(X_0)}$. Com efeito, para todo $t \in \mathbb{R}$, pelo Teorema de Helly-Bray (Teorema `A.4` do Apêndice `A`), temos que

$$\varphi_{f(X_n)}(t) = \int e^{itf(x)} \, dF_{X_n}(x) =$$

$$\int \{\cos\big(tf(x)\big) + i\text{sen}\big(tf(x)\big)\} dF_{X_n}(x) \longrightarrow \int \{\cos\big(tf(x)\big) + i\text{sen}\big(tf(x)\big)\} dF_{X_0}(x)$$

$$= \int e^{itf(x)} \, dF_{X_0}(x) = \varphi_{f(X_0)}(t) \text{ quando } n \longrightarrow \infty.$$

Logo, pelo Teorema 2.3.8, $f(X_n) \xrightarrow{\mathcal{D}} f(X_0)$ e o teorema está demonstrado.

Observações

`i`) O item `iii`) é conhecido como Teorema de Sverdrup.

`ii`) mesmo quando f não é definida e contínua em todo $\mathbb{R}^p$, pode-se provar que o teorema ainda vale; para isto, basta que f seja Borel mensurável e que, para algum conjunto mensurável $A \subset \mathbb{R}^p$, seja contínua em A, com $\mathbb{P}(\mathbf{X}_0 \in A) = 1$.

Exemplo 2.3.10

i) Se $\{X_n\}_{n\geq 1}$ *for uma seqüência de variáveis aleatórias tal que* $X_n \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)$, *então* $X_n^2 \xrightarrow{\mathcal{D}} Z^2 \sim \chi^2_{(1)}$;

ii) se $\{(X_n, Y_n)\}_{n\geq 1}$ *for uma seqüência de vetores aleatórios bidimensionais tal que* $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X_0, Y_0)$, *onde* (X_0, Y_0) *tem distribuição Normal Bivariada com vetor de médias* $\boldsymbol{0} = (0, 0)$ *e matriz de covariância* $\boldsymbol{I}_2$, *então* $X_n Y_n \xrightarrow{\mathcal{D}} W \sim$ *Cauchy.*

Os dois próximos teoremas são conseqüências imediatas do Teorema 2.3.14:

Teorema 2.3.15 *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 $X_n \xrightarrow{\mathbb{P}} X$ *e* $Y_n \xrightarrow{\mathbb{P}} Y$, *então* $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$ *e* $X_n Y_n \xrightarrow{\mathbb{P}} XY$.

Demonstração: Provaremos apenas o item `i`); a prova do item `ii`) fica como exercício para o leitor.

`i`) Suponhamos que $X_n \xrightarrow{q.c.} X$ e $Y_n \xrightarrow{q.c.} Y$. Então, pelo Teorema 2.3.6, temos $(X_n, Y_n) \xrightarrow{q.c.} (X, Y)$. Consideremos agora as funções $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ e $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ definidas por: $f(x, y) = x + y$ e $g(x, y) = xy$, para todo $(x, y) \in \mathbb{R}^2$. Logo, como f e g são contínuas, segue, pelo Teorema 2.3.14, que

$$f(X_n, Y_n) = X_n + Y_n \xrightarrow{q.c.} f(X, Y) = X + Y$$

e

$$g(X_n, Y_n) = X_n Y_n \xrightarrow{q.c.} g(X, Y) = XY$$

o que prova `i`).

Teorema 2.3.16 (Slutsky) *Sejam* $X, X_1, X_2, \ldots$ *e* $Y_1, Y_2, \ldots$ *variáveis aleatórias tais que* $X_n \xrightarrow{\mathcal{D}} X$ *e* $Y_n \xrightarrow{\mathbb{P}} c$, *onde* c *é uma constante. Então,*

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

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

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

Demonstração: Provaremos os itens `i`) e `ii`); a prova do item `iii`) fica como exercício para o leitor. Suponhamos que $X_n \xrightarrow{\mathcal{D}} X$ e $Y_n \xrightarrow{\mathbb{P}} c$.

`i`) Seja t um ponto de continuidade da função distribuição de $X+c, F_{X+c}$. Como $F_{X+c}(t) = F_X(t-c)$ então $t-c$ é um ponto de continuidade de F_X, a função distribuição de X. Para todo $\epsilon > 0$ tal que $t-c+\epsilon$ e $t-c-\epsilon$ também sejam pontos de continuidade de F_X, temos:

$$F_{X_n+Y_n}(t) = \mathbb{P}(X_n + Y_n \leq t) =$$

$$\mathbb{P}(X_n + Y_n \leq t, |Y_n - c| < \epsilon) + \mathbb{P}(X_n + Y_n \leq t, |Y_n - c| \geq \epsilon) \leq$$

$$\mathbb{P}(X_n + Y_n \leq t, |Y_n - c| < \epsilon) + \mathbb{P}(|Y_n - c| \geq \epsilon) \leq$$

$$\mathbb{P}(X_n \leq t - c + \epsilon) + \mathbb{P}(|Y_n - c| \geq \epsilon) =$$

$$F_{X_n}(t - c + \epsilon) + \mathbb{P}(|Y_n - c| \geq \epsilon), \ \forall n \geq 1,$$

o que implica

$$\begin{aligned} \limsup_{n \longrightarrow \infty} F_{X_n+Y_n}(t) &\leq \limsup_{n \longrightarrow \infty} F_{X_n}(t-c+\epsilon) + \limsup_{n \longrightarrow \infty} \mathbb{P}(|Y_n - c| \geq \epsilon) \\ &= F_X(t - c + \epsilon) \end{aligned} \tag{2.3.2}$$

Analogamente,

$$F_{X_n}(t - c - \epsilon) = \mathbb{P}(X_n \leq t - c - \epsilon) =$$

$$\mathbb{P}(X_n \leq t - c - \epsilon, |Y_n - c| < \epsilon) + \mathbb{P}(X_n \leq t - c - \epsilon, |Y_n - c| \geq \epsilon) \leq$$

$$\mathbb{P}(X_n + Y_n \leq t) + \mathbb{P}(|Y_n - c| \geq \epsilon) = F_{X_n+Y_n}(t) + \mathbb{P}(|Y_n - c| \geq \epsilon), \ \forall n \geq 1,$$

o que implica

$$\begin{aligned} F_X(t-c-\epsilon) &\leq \liminf_{n \longrightarrow \infty} F_{X_n+Y_n}(t) + \liminf_{n \longrightarrow \infty} \mathbb{P}(|Y_n - c| \geq \epsilon) \\ &= \liminf_{n \longrightarrow \infty} F_{X_n+Y_n}(t) \end{aligned} \tag{2.3.3}$$

De (2.3.2) e (2.3.3) segue que

$$F_X(t-c-\epsilon) \leq \liminf_{n \longrightarrow \infty} F_{X_n+Y_n}(t) \leq \limsup_{n \longrightarrow \infty} F_{X_n+Y_n}(t) \leq F_X(t-c+\epsilon) \tag{2.3.4}$$

Como $t-c$ é um ponto de continuidade de F_X, $\lim_{\epsilon \longrightarrow 0} F_X(t-c-\epsilon) = \lim_{\epsilon \longrightarrow 0} F_X(t-c+\epsilon) = F_X(t-c)$ e portanto, de (2.3.4), segue que $\lim_{n \longrightarrow \infty} F_{X_n+Y_n}(t) = F_{X+c}(t)$ o que prova `i`).

`ii`) Seja $\mathbf{V}_n = (X_n, Y_n)$, $n = 1, 2, \ldots$. Para todo $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, $t_1 X_n \xrightarrow{\mathcal{D}} t_1 X$ e $t_2 Y_n \xrightarrow{\mathbb{P}} t_2 c$. Então, pelo item `i`), $t_1 X_n + t_2 Y_n \xrightarrow{\mathcal{D}} t_1 X + t_2 c$; logo, pelo Teorema de Crámer-Wold (Teorema 2.3.13), $\mathbf{V}_n \xrightarrow{\mathcal{D}} (X, c)$. Consideremos agora a função $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ definida por: $f(x, y) = xy$, para todo $(x, y) \in \mathbb{R}^2$. Como f é contínua, segue, pelo Teorema 2.3.14, que $f(\mathbf{V}_n) = X_n Y_n \xrightarrow{\mathcal{D}} f(X, c) = cX$ o que prova `ii`).

Exercícios

Exercício 2.3.1 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias tais que $Cov(X_i, X_j) = 0$, $i \neq j$ com variâncias uniformemente limitadas (i.e., existe uma constante $c > 0$ tal que $Var(X_n) \leq c$, para todo $n \geq 1$) e $S_n = \sum_{j=1}^{n} X_j$, para todo $n \geq 1$.*

Prove que $\{S_n - E(S_n)\}n \xrightarrow{\mathbb{P}} 0$.

Exercício 2.3.2 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes com distribuições de probabilidade dadas por:*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \quad e \quad \mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \forall n \geq 1.$$

Mostre que $X_n \xrightarrow{\mathbb{P}} 0$ mas $\mathbb{P}(X_n \longrightarrow 0) = 0$.

Exercício 2.3.3 *Sejam $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas e $Y_n = X_n/n$, para todo $n \geq 1$;*

i) prove que $Y_n \xrightarrow{\mathbb{P}} 0$;

ii) se X for uma variável aleatória, prove que

$$\sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(|X| \geq n)$$

Sugestão*: $0 \leq [\![|X|]\!] \leq |X| \leq 1 + [\![|X|]\!]$, onde $[\![x]\!]$ denota o menor número inteiro não superior a x;*

*iii) utilizando o item **ii**) prove que, para todo número real $\epsilon > 0$,*

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n) \geq \epsilon) \leq \frac{E(|X_1|)}{\epsilon} \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(|Y_n| \geq \epsilon);$$

*iv) utilizando o item **iii**) prove que, $Y_n \xrightarrow{q.c.} 0$ se e somente se $E|X_1| < \infty$.*

Exercício 2.3.4 *Seja* $\{X_n\}_{n\geq 2}$ *uma seqüência de variáveis aleatórias com distribuições de probabilidade dadas por:*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{\log n} \quad e \quad \mathbb{P}(X_n = n) = \frac{1}{\log n}, \ \forall n \geq 2.$$

Prove que $X_n \xrightarrow{\mathbb{P}} 0$ *mas* $X_n \overset{m.r.}{\not\longrightarrow} 0$, *para todo* $r > 0$.

Exercício 2.3.5 *Sejam* $X, X_1, X_2, \ldots$ *variáveis aleatórias tais que para todo* $n \geq 1$

$$\mathbb{P}(a < X < b) = \mathbb{P}(a < X_n < b) = 1, \quad onde \ -\infty < a < b < +\infty.$$

Prove que, se $X_n \xrightarrow{\mathbb{P}} X$, *então* $X_n \xrightarrow{m.r.} X$, *para todo* $r > 0$.

Exercício 2.3.6 *Sejam* $\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ *vetores aleatórios* $p-$*dimensionais. Prove que* $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{X}_0$ *se e somente se, para todo número real* $\epsilon > 0$,

$$\mathbb{P}\big(\ ||\boldsymbol{X}_k - \boldsymbol{X}_0|| < \epsilon, \quad para \ todo \ k \geq n \ \big) \longrightarrow 1 \quad quando \ n \longrightarrow \infty.$$

Exercício 2.3.7 *Sejam* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *uma função definida por:*

$$f(t) = t - 1 \quad se \ t < 0 \quad e \quad f(t) = t + 1 \quad se \ t \geq 0.$$

Sejam $X, X_1, X_2, \ldots$ *variáveis aleatórias tais que* $X = 0$ *com probabilidade 1 e* $X_n = -1n$ *com probabilidade 1,* $n = 1, 2, \ldots$. *Prove que* $X_n \xrightarrow{\mathbb{P}} X$, *mas* $f(X_n) \overset{\mathbb{P}}{\not\longrightarrow} f(X)$.

Exercício 2.3.8 *Prove o item* ***ii*** *) do Teorema 2.3.15.*

Exercício 2.3.9 *Prove o item* ***iii*** *) do Teorema 2.3.16.*

Exercício 2.3.10 *Sejam* $\boldsymbol{X}_0 = (X_{01}, \ldots, X_{0p})$, $\boldsymbol{X}_n = (X_{n1}, \ldots, X_{np})$ *e* $\boldsymbol{Y}_n = (Y_{n1}, \ldots, Y_{np})$, $n = 1, 2, \ldots$ *vetores aleatórios* $p-$*dimensionais e seja* $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)$ *um vetor de* $\mathbb{R}^p$.

Suponha que (***a***) $\boldsymbol{X}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0$. *e* (***b***) $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{c}$.

i) Prove que $(\boldsymbol{X}_n, \boldsymbol{Y}_n) \xrightarrow{\mathcal{D}} (\boldsymbol{X}_0, \boldsymbol{c})$;

ii) utilize o item ***i*** *) para mostrar que*

$$\boldsymbol{X}_n + \boldsymbol{Y}_n \xrightarrow{\mathcal{D}} \boldsymbol{X}_0 + \boldsymbol{c} \quad e \quad \boldsymbol{X}_n \boldsymbol{Y}_n^t \xrightarrow{\mathcal{D}} \boldsymbol{X}_0 \ \boldsymbol{c}^t;$$

*iii) o que se pode concluir em (****i****) se a suposição (****b****) for substituída por :* $\boldsymbol{Y}_n \xrightarrow{\mathcal{D}} \boldsymbol{c}$?

*iv) o que se pode concluir em (****i****) se a suposição (****b****) for substituída por* $\boldsymbol{Y}_n \xrightarrow{\mathbb{P}} \boldsymbol{Y}_0 = (Y_{01}, \ldots, Y_{0p})$, *onde* $\boldsymbol{Y}_0$ *é um vetor aleatório* $p-$*dimensional ?*

2.4 Lei dos Grandes Números

Embora os conceitos de convergência estocástica discutidos na seção anterior sejam definidos para seqüências quaisquer de variáveis aleatórias, sua aplicação prática, em geral, requer a consideração de casos mais específicos. Nesta seção, concentramos nossa atenção nas chamadas *Lei dos Grandes Números*, que essencialmente são resultados sobre a *convergência quase certa*, ou *em probabilidade* de seqüências de estatísticas que podem ser expressas como médias de variáveis aleatórias (ou vetores aleatórios). Uma justificativa básica para tal particularização está no fato de que esse tipo de estatística ocorre com grande frequência na maioria dos modelos usualmente empregados na prática. A *convergência em distribuição* de seqüências de médias de variáveis (ou vetores) aleatórios é o tópico abordado no Capítulo 3.

Mais especificamente, sejam $X_1, X_2, \ldots$ variáveis aleatórias definidas em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ tais que $E(|X_n|) < \infty,\ n = 1, 2, \ldots$ e sejam $T_n = \sum_{j=1}^{n} X_j,\ n = 1, 2, \ldots$. Diremos que a seqüência $\{X_n\}_{n\geq 1}$ satisfaz a *Lei Fraca dos Grandes Números* se

$$\frac{T_n - E(T_n)}{n} \xrightarrow{\mathbb{P}} 0,$$

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, pelo Teorema 2.3.5, ela também satisfaz a Lei Fraca dos Grandes Números. (a Lei Forte é na realidade mais "forte"). Observamos que, se as variáveis aleatórias $X_n, n = 1, 2, \ldots$, tem a mesma média finita μ, então $E(T_n)n = \mu$, e $\{X_n\}_{n\geq 1}$ satisfaz a Lei Fraca (ou Forte) se, e somente se,

$$T_n n \xrightarrow{\mathbb{P}} \mu \quad (\text{ ou } T_n n \xrightarrow{q.c.} \mu).$$

O restante desta seção é destinado à demonstração da validade das Leis dos Grandes Números sob diferentes suposições sobre a distribuição das variáveis aleatórias envolvidas; basicamente essas suposições envolvem restrições sobre os momentos e sobre a independência dessas variáveis.

Teorema 2.4.1 (Lei Fraca dos Grandes Números de Chebychev) *Seja $X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias independentes com médias $\mu_1, \mu_2, \ldots$ e variâncias*

$\sigma_1^2, \sigma_2^2, \ldots,$ *respectivamente. Se* n^{-2} $Var(T_n) = n^{-2} \sum_{j=1}^{n} \sigma_j^2 \longrightarrow 0$ *quando* $n \longrightarrow \infty$, *então* $\{X_n\}_{n \geq 1}$ *satisfaz a Lei Fraca dos Grandes Números, ou seja,* $n^{-1}\{T_n - \sum_{j=1}^{n} \mu_j\} \xrightarrow{\mathbb{P}} 0.$

Demonstração: Pela desigualdade de Chebychev (Teorema A.1 do Apêndice A) temos que, para todo $\epsilon > 0$,

$$\mathbb{P}\Big(\Big|\frac{T_n - E(T_n)}{n}\Big| \geq \epsilon \Big) = \mathbb{P}\Big(\big|T_n - E(T_n)\big| \geq n\epsilon \Big) \leq \frac{\mathrm{Var}(T_n)}{n^2 \epsilon^2} \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Logo, $\big(T_n - E(T_n)\big)n = (T_n - \sum_{j=1}^{n} \mu_j)n \xrightarrow{\mathbb{P}} 0.$

Note que nesse caso, as variáveis aleatórias não precisam ter esperanças e variâncias iguais para a validade da Lei Fraca dos Grandes Números. Obviamente, se $\{X_n\}_{n \geq 1}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E(X_1) = \mu$ e $\mathrm{Var}(X_1) = \sigma^2 < \infty$, então $\mathrm{Var}(T_n)n^2 = n\sigma^2 n^2 = \sigma^2 n \longrightarrow 0$ quando $n \longrightarrow \infty$ e $\{X_n\}_{n \geq 1}$ satisfaz a Lei Fraca dos Grandes Números.

Embora o caso de variáveis aleatórias identicamente distribuídas seja o mais freqüente, note que essa condição não é necessária; na realidade basta que todas as variáveis aleatórias tenham a mesma média e a mesma variância. Khintchine provou a Lei Fraca dos Grandes Números supondo apenas $E(X_n) < \infty,\ n = 1, 2, \ldots$; no entanto como veremos a seguir, sua demonstração exige que todas as variáveis aleatórias $X_1, X_2, \ldots$ sejam identicamente distribuídas.

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* $E(X_1) = \mu < \infty$, *então* $T_n n \xrightarrow{\mathbb{P}} \mu$.

Demonstração: Sejam $\varphi_{\frac{T_n}{n}}$ e φ_{X_1} as funções características de $T_n n$ e X_1, respectivamente. Logo, para todo $t \in \mathbb{R}$:

$$\varphi_{\frac{T_n}{n}}(t) = E\big(e^{it\frac{T_n}{n}} \big) = \varphi_{T_n}(tn) = \varphi_{\sum_{j=1}^{n} X_j}(tn) = \prod_{j=1}^{n} \varphi_{X_j}(tn) = \big[\varphi_{X_1}(tn) \big]^n.$$

Considerando a expansão de Taylor até a 1ª ordem de φ_{X_1} em torno de zero e observando que $\varphi_{X_1}(0) = 1$ e $\varphi'_{X_1}(0) = i\mu$ temos:

$$\varphi_{X_1}(tn) = \varphi_{X_1}(0) + \varphi'_{X_1}(0)tn + o(tn) = 1 + i\mu\frac{t}{n} + to(1n), \quad \forall t \in \mathbb{R}.$$

Então, para todo $t \in \mathbb{R}$,

$$\varphi_{T_n n}(t) = [\varphi_{X_1}(tn)]^n = \big(1 + i\mu\frac{t}{n} + t\ o\ (1n)\big)^n \longrightarrow e^{it\mu} = \varphi_\mu(t) \ \text{ quando } n \longrightarrow \infty,$$

onde φ_μ é a função característica da variável aleatória degenerada no ponto μ. Logo, pelo Teorema 2.3.8, $T_n n \xrightarrow{\mathcal{D}} \mu$ e, pelo Teorema 2.3.10, $T_n n \xrightarrow{\mathbb{P}} \mu$.

Exemplo 2.4.1 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $E(X) = \mu$ e $Var(X) = \sigma^2 < \infty$. Sejam $S_n^2 = (n-1)^{-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$ a variância amostral e $\overline{X}_n$ a média amostral. Então, pela Lei Fraca dos Grandes Números de Khintchine, $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$. Com relação a seqüência $\{S_n^2\}_{n\geq 1}$ temos que $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$. Com efeito, como $\{X_n^2\}_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E(X_1^2) = Var(X_1) + \mu^2$, pela Lei Fraca dos Grandes Números de Khintchine, segue que $\sum_{j=1}^{n} X_j^2 n \xrightarrow{\mathbb{P}} \mu^2 + \sigma^2$. Logo, como $\frac{n-1}{n} \longrightarrow 1$ quando $n \longrightarrow \infty$ e como, pelo Teorema 2.3.14, $\overline{X}_n^2 \xrightarrow{\mathbb{P}} \mu^2$ temos, pelo Teorema 2.3.15, que:*

$$S_n^2 = \frac{n}{n-1} \frac{\sum_{j=1}^{n} X_j^2}{n} - \frac{n}{n-1} \overline{X}_n^2 \xrightarrow{\mathbb{P}} \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Observamos também que, pelo Teorema 2.3.14, $S_n \xrightarrow{\mathbb{P}} \sigma$.

O resultado seguinte, devido a Markov, permite mostrar que a Lei Fraca dos Grandes Números ainda vale para seqüências de variáveis aleatórias independentes, porém não necessariamente identicamente distribuídas mesmo sem a exigência do segundo momento finito.

Teorema 2.4.3 (Lei Fraca dos Grandes Números de Markov) *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com médias $\mu_1, \mu_2, \ldots$ finitas e tais que, para algum $\delta, 0 < \delta \leq 1$, vale a condição de Markov: $n^{-(1+\delta)} \sum_{j=1}^{n} E|X_j - \mu_j|^{1+\delta} \longrightarrow 0$ quando $n \longrightarrow \infty$. Então $n^{-1}\{T_n - \sum_{j=1}^{n} \mu_j\} \xrightarrow{\mathbb{P}} 0$.*

Demonstração: A demonstração desse Teorema pode ser feita utilizando a mesma técnica usada na demonstração do Teorema 2.4.2. Ver Wasan (1972) por exemplo.

Note que o caso $\delta = 1$ corresponde à Lei Fraca dos Grandes Números de Chebychev (Teorema 2.4.1).

Exemplo 2.4.2 *Consideremos uma seqüência de variáveis aleatórias independentes* $\{X_n\}_{n\geq 1}$ *tal que para todo* $n \geq 1$,

$$X_n = \begin{cases} -2^{n(1-\epsilon)} & \textit{com probabilidade } 2^{-n-1} \\ 0 & \textit{com probabilidade } 1-2^{-n} \\ 2^{n(1-\epsilon)} & \textit{com probabilidade } 2^{-n-1} \end{cases}$$

onde $0 < \epsilon < 1$ *é um número real. Então* $E(X_n) = 0$, $Var(X_n) = \sigma_n^2 = 2^{n(1-2\epsilon)}$ *e*

$$n^{-2}\sum_{j=1}^{n}\sigma_j^2 = n^{-2}\sum_{j=1}^{n}2^{j(1-2\epsilon)} = \frac{2^{1-2\epsilon}\ (2^{n(1-2\epsilon)}-1)}{n^2\ (2^{1-2\epsilon}-1)}$$

Logo, se $\epsilon \geq \frac{1}{2}$ *segue que* $n^{-2}\sum_{j=1}^{n}\sigma_j^2 \longrightarrow 0$ *quando* $n \longrightarrow \infty$ *e, pela Lei Fraca dos Grandes Números de Chebychev (Teorema 2.4.1),* $n^{-1}T_n \xrightarrow{\mathbb{P}} 0$. *Por outro lado, se* $\epsilon < \frac{1}{2}$, *então* $n^{-2}\sum_{j=1}^{n}\sigma_j^2 \longrightarrow \infty$ *quando* $n \longrightarrow \infty$ *e a Lei Fraca dos Grandes Números de Chebychev não pode ser aplicada. Contudo, como*

$$E|X_n|^{1+\delta} = 2^{n(1-\epsilon)(1+\delta)-n} \leq 1 \quad \textit{se} \quad (1-\epsilon)(1+\delta) \leq 1,$$

escolhendo $\delta = \epsilon$ *temos que* $E|X_n|^{1+\delta} < 1, \ \forall n \geq 1$. *Portanto,*

$$n^{-(1+\delta)}\sum_{j=1}^{n}E|X_j|^{1+\delta} < \frac{n}{n^{1+\epsilon}} = n^{-\epsilon} \longrightarrow 0 \quad \textit{quando } n \longrightarrow \infty$$

e, conseqüentemente, vale a Lei Fraca dos Grandes Números de Markov (Teorema 2.4.3), ou seja, $n^{-1}T_n \xrightarrow{\mathbb{P}} 0$.

Teorema 2.4.4 (Lei Forte dos Grandes Números de Kolmogorov) *Seja* $X_1, X_2, \ldots$ *uma seqüência de variáveis aleatórias independentes com médias* $\mu_1, \mu_2, \ldots$ *e variâncias* $\sigma_1^2, \sigma_2^2, \ldots$, *respectivamente.*

$$\textit{Se} \ \sum_{j=1}^{\infty}\frac{\sigma_n^2}{n^2} < \infty, \quad \textit{então} \ \ n^{-1}\{T_n - E(T_n)\} = n^{-1}(T_n - \sum_{j=1}^{n}\mu_i) \xrightarrow{q.c.} 0.$$

Demonstração: Por simplicidade suponhamos $\mu_n = 0, n = 1, 2, \ldots$; no caso geral basta considerar as variáveis aleatórias $X_n - \mu_n, n = 1, 2, \ldots$.

Como $\{X_n n\}_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes com $E(X_n n) = 0$, $n = 1, 2, \ldots$ e $\sum_{n=1}^{\infty} \mathrm{Var}(X_n n) = \sum_{n=1}^{\infty} \sigma_n^2 n^2 < \infty$ então, pelo Teorema A.7 do Apêndice A, $\sum_{n=1}^{\infty} X_n n$ converge quase certamente para uma variável aleatória X. Logo, pelo Lema de Kronecker (Teorema A.8 do Apêndice A), $n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{q.c.} 0$.

Em particular, se $\{X_n\}_{n\geq 1}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E(X_1) \leq \mu$ e $\mathrm{Var}(X_1) = \sigma^2$, então $\sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2\sigma^2}{6} < \infty$. Logo, $T_n n \xrightarrow{q.c.} \mu$.

O próximo resultado permite relaxar a condição sobre a existência do segundo momento das variáveis aleatórias envolvidas, porém exige que elas sejam identicamente distribuídas.

Teorema 2.4.5 (Lei Forte dos Grandes Números de Khintchine) *Seja $\{X_n\}_{n\geq 1}$ uma seqüência 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 $c = E(X_1)$.*

Demonstração: Ver Rao (1984) por exemplo.

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

Teorema 2.4.6 *Sejam $\{\boldsymbol{X}_n\}_{n\geq 1} = \{(X_{n1}, X_{n2}, \ldots, X_{np})\}_{n\geq 1}$ uma seqüência de vetores aleatórios $p-$dimensionais independentes e identicamente distribuídos tal que, $E(|X_{1j}|) < \infty$, $j = 1, 2, \ldots, p$, $\boldsymbol{\mu} = E(\boldsymbol{X}_1) = (\mu_1, \mu_2, \ldots, \mu_p)$ e $\overline{\boldsymbol{X}}_n = n^{-1} \sum_{j=1}^{n} \boldsymbol{X}_j$, $n = 1, 2, \ldots$. Então $\boldsymbol{X}_n \xrightarrow{q.c.} \boldsymbol{\mu}$.*

Demonstração: Para todo $j = 1, 2, \ldots, p$ a seqüência $\{X_{nj}\}_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas tal que $E(|X_{1j}|) < \infty$ e $E(X_{1j}) = \mu_j$. Então, pela Lei Forte dos Grandes Números de Khintchine (Teorema 2.4.5), $n^{-1} \sum_{k=1}^{n} X_{kj} \xrightarrow{q.c.} \mu_j$. Logo, pelo Teorema 2.3.6, $\overline{\boldsymbol{X}}_n \xrightarrow{q.c.} \boldsymbol{\mu}$.

Exemplo 2.4.3 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $E(X) = \mu$ e $Var(X) = \sigma^2 < \infty$. Sejam $S_n^2 = (n-1)^{-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$ a variância amostral e $\overline{X}_n$ a média amostral. Seja $\boldsymbol{Y}_n = \big((X_n - \mu)^2, (X_n - \mu)\big)$, $n = 1, 2, \ldots$. A seqüência $\{\boldsymbol{Y}_n\}_{n\geq 1}$ é uma seqüência de vetores bidimensionais independentes e identica-*

mente distribuídos com $\boldsymbol{\mu} = E(\boldsymbol{Y}_1) = (\sigma^2, 0)$. *Logo, pelo Teorema 2.4.6,*

$$\overline{Y}_n = \Big(\ n^{-1}\sum_{j=1}^{n}(X_j-\mu)^2, (\overline{X}_n-\mu)\ \Big) \xrightarrow{q.c.} (\sigma^2, 0).$$

Note também que, como a função $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *definida por:* $f(x_1, x_2) = x_1 - x_2^2$, $\forall (x_1, x_2)\ \in \mathbb{R}^2$, *é contínua, segue pelo Teorema 2.3.14, que*

$$f(\overline{\boldsymbol{Y}}_n) = n^{-1}\sum_{j=1}^{n}(X_j-\mu)^2 - (\overline{X}_n-\mu)^2 \xrightarrow{q.c.} f(\sigma^2, 0) = \sigma^2.$$

Portanto, como $(n-1)n \longrightarrow 1$ *quando* $n \longrightarrow \infty$, *então*

$$S_n^2 = \frac{n}{n-1}\Big[\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2 - (\overline{X}_n-\mu)^2\Big] = \frac{n}{n-1} f(\overline{\boldsymbol{Y}}_n) \xrightarrow{q.c.} \sigma^2.$$

Em todos os Teoremas dessa seção foi exigida a condição de *independência* das variáveis aleatórias envolvidas. é possível mostrar que as Leis dos Grandes Números ainda valem em certas situações onde essa condição não é verificada. Por exemplo, suponhamos que $\{X_n\}_{n\geq 1}$ seja uma seqüência de variáveis aleatórias não correlacionadas duas a duas, isto é , $\mathrm{Cov}(X_i, X_j) = 0,\ i \neq j,\ i, j = 1, 2, \ldots$, e com $\sigma_j^2 = \mathrm{Var}(X_j)$ uniformemente limitadas, isto é , $\sigma_j^2 \leq c$, onde c é uma constante, $j = 1, 2, \ldots$. Neste caso vale a Lei Fraca dos Grandes Números (Exercício 2.4.10).

Exercícios

Exercício 2.4.1 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $E(X_1) = \mu$ *e* $Var(X_1) = \sigma^2,\ 0 < \sigma^2 < \infty$. *Prove que*

$$\frac{\sum\limits_{j=1}^{n} X_j}{\sqrt{n\sum\limits_{j=1}^{n} X_j^2}} \xrightarrow{q.c.} \frac{\mu}{\sqrt{\mu^2+\sigma^2}}.$$

Exercício 2.4.2 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por:*

$$\mathbb{P}(X_n = n^{\alpha}) = \mathbb{P}(X_n = -n^{\alpha}) = \frac{1}{2},\ n = 1, 2, \ldots\ .$$

Prove que, se $0 < \alpha < \frac{1}{2}$ *então* $n^{-1}\sum\limits_{j=1}^{n} X_j \xrightarrow{q.c.} 0$.

Exercício 2.4.3 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $X_1 \sim \mathcal{N}(0,1)$. *Qual o limite quase certo de* $\frac{\sum_{j=1}^{n} X_j^2}{\sum_{j=1}^{n}(X_j-1)^2}$ *?*

Exercício 2.4.4 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por:*

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2}, \; n = 1, 2, \dots \; .$$

Prove que $\{X_n\}_{n\geq 1}$ *não satisfaz a Lei dos Grandes Números.*

Exercício 2.4.5 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias tais que* $E(X_n) = 0$, $\mathbb{P}(|X_n| \leq a) = 1$ *para algum* $a > 0, \; n = 1, 2, , \dots$ *e*

$$Var\Big(\frac{X_1 + X_2 + \dots + X_n}{n}\Big) \longrightarrow b \neq 0 \quad quando \quad n \longrightarrow \infty.$$

Prove que $\{X_n\}_{n\geq 1}$ *não satisfaz a Lei dos Grandes Números.*

Exercício 2.4.6 *Decida se a Lei dos Grandes Números vale para as seqüências de variáveis aleatórias independentes* $\{X_n\}_{n\geq 1}$, *cujas distribuições de probabilidade são dadas por:*

i) $\mathbb{P}(X_n = \pm 2^n) = \frac{1}{2}$;

ii) $\mathbb{P}(X_n = \pm 2^n) = 2^{-(2n+1)}, \; \mathbb{P}(X_n = 0) = 1 - 2^{-2n}$;

iii) $\mathbb{P}(X_n = \pm n) = \frac{1}{2\sqrt{n}}, \; \mathbb{P}(X_n = 0) = 1 - \frac{1}{\sqrt{n}}$.

Exercício 2.4.7 *Considere* n *ensaios de Bernoulli independentes com probabilidade de sucesso* p
$(0 < p < 1)$ *em cada ensaio. Seja* T_n *o número de vezes dentre os* n *ensaios em que um sucesso é imediatamente seguido de um fracasso. Prove que* $T_n n \xrightarrow{\mathbb{P}} c$, *onde* c *é uma constante. Identifique a constante* c.

Exercício 2.4.8 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $E(X_1) = \mu$, $Var(X_1) = \sigma^2$ *e* $E(X_1 - \mu)^4 = \sigma^4 + 1$. *Qual o limite quase certo de* $\sum_{j=1}^{n} X_j^2 n$ *?*

Exercício 2.4.9 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por:*

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2} \quad e \quad \mathbb{P}(X_n = n) = \frac{1}{n^2}, \; n = 1, 2, \dots \; .$$

Vale a Lei dos Grandes Números neste caso ?

Exercício 2.4.10 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias tais que* $Cov(X_i, X_j) = 0$, *para todo* $i \neq j$, $i, j = 1, 2, \ldots$ *e* $Var(X_j) = \sigma_j^2 < c$ *onde* c *é uma constante,* $j = 1, 2, \ldots$. *Prove que essa seqüência satisfaz a Lei Fraca dos Grandes Números.*

Capítulo 3

Teorema Limite Central

3.1 Introdução

Como vimos no capítulo anterior, o conceito de convergência em distribuição é de fundamental importância em aplicações estatísticas, pois permite (pelo menos do ponto de vista teórico) a construção de intervalos de confiança aproximados para os parâmetros de interesse e de testes aproximados para hipóteses sobre esses parâmetros. Quando as estatísticas em questão correspondem a somas de variáveis aleatórias independentes, esses procedimentos aproximados baseados em métodos assintóticos são especialmente atraentes do ponto de vista prático, pois, sob condições bastante gerais, essas somas, devidamente padronizadas, convergem fracamente para variáveis aleatórias com distribuição Normal de média zero e variância unitária. Esse resultado, que na literatura estatística é conhecido como *Teorema Limite Central*, é o tema deste capítulo. Na Seção 3.2 apresentaremos algumas versões desse teorema, caracterizadas por diferentes suposições sobre as distribuições das variáveis aleatórias que geram as estatísticas em consideração; como no caso das Leis dos Grandes Números, essas suposições estão essencialmente relacionadas com os momentos das variáveis aleatórias mencionadas. Na Seção 3.3 consideramos situações em que as estatísticas de interesse podem ser decompostas em dois componentes, um dos quais corresponde a uma soma de variáveis aleatórias independentes e outro que converge em probabilidade para zero; nesse caso, o Teorema Limite Central pode ser empregado em conjunto com o Teorema de Slutsky para mostrar a convergência em distribuição dessas estatísticas para uma variável aleatória com distribuição $\mathcal{N}(0,1)$. Também consideramos situações onde o interesse recai na convergência fraca de *funções* de estatísticas que (devidamente padronizadas) têm distribuição assintoticamente Normal; nesse contexto destacamos o conhecido *Método Delta* e as *transformações estabilizadoras da variância*. Na Seção 3.4 fazemos alguns comentários sobre taxas de convergência.

Finalmente gostaríamos de ressaltar que existem outras versões do Teorema Limite Central aplicáveis a seqüências de variáveis aleatórias dependentes; entre elas destacamos aquelas referentes a seqüências de variáveis aleatórias permutáveis ou com estrutura de Martingal. Esses resultados embora bastante importantes, não são compatíveis com o nível adotado nestas notas. O leitor interessado poderá consultar Billingsley (1961) ou Teicher e Chow (1978), por exemplo, para maiores detalhes e outras referências.

3.2 Principais Versões do Teorema Limite Central

Dada uma seqüência de variáveis aleatórias $\{X_n\}_{n\geq 1}$ usaremos a notação $T_n = \sum\limits_{j=1}^{n} X_j,\ s_n^2 = \mathrm{Var}(T_n),\ \overline{X}_n = T_n/n,\ n = 1, 2, \dots$. Para indicar que uma seqüência de variáveis aleatórias $\{Y_n\}_{n\geq 1}$ converge em distribuição para uma variável aleatória Normal, com média μ e variância σ^2, usaremos a notação $Y_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2)$.

Apresentamos a seguir a versão clássica do Teorema Limite Central.

Teorema 3.2.1 *(Teorema Limite Central para variáveis aleatórias independentes e identicamente distribuídas) Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $\mu = E(X_1)$ e $\sigma^2 = Var(X_1),\ 0 < \sigma^2 < \infty$. Então,*

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

Demonstração: Seja $Z_n = (T_n - E(T_n))/s_n, n \geq 1$. Pelo Teorema 2.3.8, para mostrar que $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ basta mostrar que a seqüência das funções características de Z_n, $\{\varphi_{Z_n}\}_{n\geq 1}$, é tal que, para todo $t \in \mathbb{R}$,

$$\varphi_{Z_n}(t) \longrightarrow \varphi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}} \quad \text{quando} \quad n \longrightarrow \infty.$$

Seja φ a função característica de $X_1 - \mu$. Então, para todo $t \in \mathbb{R}$,

$$\begin{aligned}
\varphi_{Z_n}(t) = E(e^{itZ_n}) &= E\big(e^{\frac{it}{\sigma\sqrt{n}} \sum\limits_{j=1}^{n} (X_j - \mu)} \big) \\
&= \varphi_{\sum\limits_{j=1}^{n}(X_j-\mu)}(t/\sigma\sqrt{n}) \\
&= \prod_{j=1}^{n} \varphi_{X_j-\mu}(t/\sigma\sqrt{n}) = \varphi^n(t/\sigma\sqrt{n})
\end{aligned} \tag{3.2.1}$$

Utilizando a Fórmula de Taylor, para todo $t \in \mathbb{R}$,

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''\big(\xi(t)\big)t^2,$$

onde $|\xi(t)| \leq |t|$. Como $\varphi(0) = 1,\ \varphi'(0) = iE(X_1 - \mu) = 0$ e $\varphi''(0) = i^2E(X_1 - \mu)^2 = -\sigma^2$, segue que, para todo $t \in \mathbb{R}$,

$$\begin{aligned}
\varphi(t) &= \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 + \frac{1}{2}\{\varphi''\big(\xi(t)\big) - \varphi''(0)\}t^2 \\
&= 1 - \frac{1}{2}\sigma^2 t^2 + \frac{1}{2}\{\varphi''\big(\xi(t)\big) - \varphi''(0)\}t^2
\end{aligned} \tag{3.2.2}$$

De (3.2.1) e (3.2.2) temos que, para todo $t \in \mathbb{R}$,

$$\varphi_{Z_n}(t) = \big[1 - \frac{t^2}{2n} + \frac{1}{2}\{\varphi''\big(\xi(\frac{t}{\sigma\sqrt{n}})\big) - \varphi''(0)\}\frac{t^2}{\sigma^2 n}\big]^n.$$

Como φ'' é contínua (Veja Propriedade 1.4.8) temos que, para t fixado,

$$\big\{\varphi''\big(\xi(\frac{t}{\sigma\sqrt{n}})\big) - \varphi''(0)\big\}\frac{t^2}{\sigma^2 n} = o\ (\frac{t^2}{n}).$$

Portanto, para todo $t \in \mathbb{R}$,

$$\varphi_{Z_n}(t) = \big[1 - \frac{t^2}{2n} + o\ (\frac{t^2}{n})\big]^n \longrightarrow e^{-t^2/2} \ \text{ quando } \ n \longrightarrow \infty,$$

o que prova o teorema

Como uma conseqüência imediata desse Teorema temos:

Teorema 3.2.2 *(Teorema Limite Central de De Moivre e Laplace): Seja T_n o número de sucessos em n ensaios de Bernoulli independentes com probabilidade de sucesso p, $0 < p < 1$, em cada ensaio. Então,*

$$\frac{T_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Demonstração: Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis independentes e identicamente distribuídas com $X_1 \sim$ Bernoulli (p). Então, $\mu = E(X_1) = p$ e $\sigma^2 = \text{Var}(X_1) = p(1-p)$. Logo, pelo Teorema 3.2.1, segue que

$$\frac{T_n - E(T_n)}{s_n} = \frac{T_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Exemplo 3.2.1 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $\mu = E(X)$ e $\sigma^2 = Var(X)$, $0 < \sigma^2 < \infty$. Seja $\overline{X}_n$ a média amostral. Então,*

$$\sqrt{n}\ (\overline{X}_n - \mu) = \sigma\big(\ \frac{T_n - n\mu}{\sigma\sqrt{n}}\ \big) \xrightarrow{\mathcal{D}} \sigma\mathcal{N}(0,1) = \mathcal{N}(0,\sigma^2)$$

Exemplo 3.2.2 *Sejam $\{X_n\}_{n\geq 1}$ e $\{Y_n\}_{n\geq 1}$ seqüências de variáveis aleatórias independentes e identicamente distribuídas, respectivamente, com $E(X_1) = 0$, $\sigma^2 = E(X_1^2)$, $0 < \sigma^2 < \infty$ e $E(Y_1) = \mu$. Então, pelo Exemplo 3.2.1, $\sqrt{n}\ \overline{X}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$ e, pela Lei Forte dos Grandes Números de Khintchine (Teorema 2.4.5), $\overline{Y}_n \xrightarrow{\mathbb{P}} \mu$. Logo, pelo Teorema de Slutsky (Teorema 2.3.16), $\sqrt{n}\ \overline{X}_n + \overline{Y}_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu,\sigma^2)$.*

Teorema 3.2.3 *(Teorema Limite Central de Lindeberg) Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes com* $\mu_n = E(X_n)$ *e* $\sigma_n^2 = Var(X_n) < \infty$, *onde pelo menos um* $\sigma_n^2 > 0$. *Se, para todo* $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} \frac{1}{s_n^2} \sum_{j=1}^{n} \int\limits_{|x-\mu_j|>\epsilon s_n} (x-\mu_j)^2 \; dF_{X_j}(x) \; = \; 0 \tag{3.2.3}$$

então, $\big(\; T_n - E(T_n) \; \big)/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Demonstração: A demonstração desse teorema consiste em provar que a seqüência das funções características de $Z_n = \big(\; T_n - E(T_n) \; \big)/s_n$, $\{\varphi_{Z_n}\}_{n\geq 1}$ converge para a função característica da variável aleatória Normal com média zero e variância um, isto é , que, para todo $t \in \mathbb{R}$,

$$\varphi_{Z_n}(t) = E\big(\; e^{\frac{it}{s_n} \sum\limits_{j=1}^{n}(X_j-\mu_j)} \; \big) = \varphi_{\sum\limits_{j=1}^{n}(X_j-\mu_j)} \; (t/s_n) = \prod_{j=1}^{n} \; \varphi_{(X_j-\mu_j)} \; (t/s_n) \longrightarrow e^{-t^2/2}$$

quando $n \longrightarrow \infty$, e pode ser vista em Gnedenko (1969), por exemplo.

Observações:

i) A condição (3.2.3) é chamada *condição de Lindeberg* e como, para todo $\epsilon > 0$,

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2 =$$

$$\sum_{j=1}^{n} \int\limits_{|x-\mu_j|\leq \epsilon s_n} (x-\mu_j)^2 \, dF_{X_j}(x) + \sum_{j=1}^{n} \int\limits_{|x-\mu_j|> \epsilon s_n} (x-\mu_j)^2 \, dF_{X_j}(x),$$

ela é equivalente à condição:

$$\lim_{n\longrightarrow\infty} \frac{1}{s_n^2} \int\limits_{|x-\mu_j|\leq \epsilon s_n} (x-\mu_j)^2 \, dF_{X_j}(x) = 1 \tag{3.2.4}$$

ii) A condição de Lindeberg implica que

$$\max_{1\leq j\leq n} \frac{\sigma_j^2}{s_n^2} \longrightarrow 0 \quad \text{quando } n \longrightarrow \infty.$$

De fato, para todo $\epsilon > 0$, para todo $n \geq 1$ e para todo $j,\ 1 \leq j \leq n$,

$$\frac{\sigma_j^2}{s_n^2} = \frac{1}{s_n^2} \int_{-\infty}^{\infty} (x-\mu_j)^2 dF_{X_j}(x) =$$

$$\frac{1}{s_n^2} \int\limits_{|x-\mu_j|\leq \epsilon s_n} (x-\mu_j)^2 dF_{X_j}(x) + \frac{1}{s_n^2} \int\limits_{|x-\mu_j|> \epsilon s_n} (x-\mu_j)^2 dF_{X_j}(x) \leq$$

$$\frac{1}{s_n^2} \int_{-\infty}^{\infty} \epsilon^2 s_n^2 dF_{X_j}(x) + \frac{1}{s_n^2} \int\limits_{|x-\mu_j|> \epsilon s_n} (x-\mu_j)^2 dF_{X_j}(x).$$

Logo, pela condição de Lindeberg,

$$\max_{1\leq j\leq n} \frac{\sigma_j^2}{s_n^2} \leq \epsilon^2 + \frac{1}{s_n^2} \sum_{j=1}^{n} \int\limits_{|x-\mu_j|> \epsilon s_n} (x-\mu_j)^2 dF_{X_j}(x) \longrightarrow \epsilon^2 \quad \text{quando } n \longrightarrow \infty.$$

Como ϵ é arbitrário, segue que

$$\max_{1\leq j\leq n} \frac{\sigma_j^2}{s_n^2} \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Logo, como $\big(T_n - E(T_n) \big)/s_n = \sum_{j=1}^{n} \big(\frac{X_j - \mu_j}{s_n} \big)$ e, pela condição de Lindeberg temos que

$$\max_{1 \le j \le n} \mathrm{Var}\Big(\frac{X_j - \mu_j}{s_n} \Big) = \max_{1 \le j \le n} \frac{\sigma_j^2}{s_n^2} \longrightarrow 0 \ \text{ quando } n \longrightarrow \infty,$$

então, a condição de Lindeberg implica que $\big(T_n - E(T_n) \big)/s_n$ é uma soma de variáveis aleatórias com *variâncias uniformemente pequenas*, para todo n suficientemente grande.

iii) Por outro lado, a condição de Lindeberg implica que, para todo $\epsilon > 0$,

$$\max_{1 \le j \le n} \mathbb{P}\big(|X_j - \mu_j| > \epsilon s_n \big) \longrightarrow 0 \quad \text{quando } n \longrightarrow \infty.$$

Com efeito, para todo $\epsilon > 0$, para todo $n \ge 1$ e para todo $j,\ 1 \le j \le n$, temos:

$$\frac{\sigma_j^2}{s_n^2} = \frac{1}{s_n^2} \int_{-\infty}^{\infty} (x - \mu_j)^2 dF_{X_j}(x) \ge$$

$$\frac{1}{s_n^2} \int_{|x-\mu_j| > \epsilon s_n} (x - \mu_j)^2 dF_{X_j}(x) \ge \frac{1}{s_n^2} \int_{|x-\mu_j| > \epsilon s_n} \epsilon^2 s_n^2 dF_{X_j}(x) =$$

$$\epsilon^2\, \mathbb{P}\big(|X_j - \mu_j| > \epsilon s_n \big), \quad \text{ou seja,} \quad \mathbb{P}\big(|X_j - \mu_j| > \epsilon s_n \big) \le \frac{\sigma_j^2}{\epsilon^2 s_n^2}.$$

Logo, pela observação `ii`),

$$\max_{1 \le j \le n} \mathbb{P}\big(|X_j - \mu_j| > \epsilon s_n \big) \le \frac{1}{\epsilon^2} \max_{1 \le j \le n} \frac{\sigma_j^2}{s_n^2} \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

iv) Em particular, se $\{X_n\}_{n \ge 1}$ for uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $\mu = E(X_1)$ e $\sigma^2 = \mathrm{Var}(X_1),\ 0 < \sigma^2 < \infty$, então, para todo $\epsilon > 0$,

$$\frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x-\mu_j| > \epsilon s_n} (x - \mu_j)^2 dF_{X_j}(x) = \frac{1}{n\sigma^2} \sum_{j=1}^{n} \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF_{X_1}(x) =$$

$$\frac{1}{\sigma^2} \int_{|x-\mu| > \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF_{X_1}(x) = \frac{1}{\sigma^2} \Big[\sigma^2 - \int_{|x-\mu| \le \epsilon\sigma\sqrt{n}} (x - \mu)^2 dF_{X_1}(x) \Big] =$$

$$1 - \frac{1}{\sigma^2} \Big[\int_{\mu-\epsilon\sigma\sqrt{n}}^{\mu} (x - \mu)^2 dF_{X_1}(x) + \int_{\mu}^{\mu+\epsilon\sigma\sqrt{n}} (x - \mu)^2 dF_{X_1}(x) \Big] \longrightarrow 1 - \frac{1}{\sigma^2}\sigma^2 = 0$$

quando $n \longrightarrow \infty$. Logo, como vale a condição de Lindeberg, o Teorema 3.2.1 é , no fundo, uma conseqüência do Teorema 3.2.3.

Exemplo 3.2.3 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por*

$$\mathbb{P}(X_n = -n) = \mathbb{P}(X_n = n) = \frac{1}{2}, \quad n = 1, 2, \dots \ .$$

Então, $\mu_n = E(X_n) = 0$, $\sigma_n^2 = Var(X_n) = n^2$ e

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2 = \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6} > \frac{2n^3}{6} = \frac{n^3}{3}.$$

Então, para todo número $\epsilon > 0$, temos que

$$\epsilon \ s_n > \frac{\epsilon \ n^{\frac{3}{2}}}{\sqrt{3}} = \frac{\epsilon \ n^{\frac{1}{2}}}{\sqrt{3}} \ n \geq n, \ \ \textit{para todo} \ n \geq \frac{3}{\epsilon^2}.$$

Logo, para todo $n \geq \frac{3}{\epsilon^2}$ e para todo j, $1 \leq j \leq n$,

$$\int\limits_{|x-\mu_j|>\epsilon s_n} (x-\mu_j)^2 dF_{X_j}(x) = \int\limits_{|x|>\epsilon s_n} x^2 \ dF_{X_j}(x) = 0.$$

Portanto, a condição de Lindeberg se verifica e $T_n/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Uma outra conseqüência do Teorema 3.2.3 é o Teorema Limite Central de Liapunov, que é muito útil no caso de as variáveis X_n terem momentos finitos de ordem maior que dois.

Teorema 3.2.4 *(Teorema Limite Central de Liapunov) Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes com $\mu_n = E(X_n)$, $\sigma_n^2 = Var(X_n) < \infty$, onde pelo menos um $\sigma_n^2 > 0$, e suponhamos que, para algum $\delta > 0$, $E|X_n - \mu_n|^{2+\delta} < \infty$, $n = 1, 2, \dots$. Se*

$$\lim_{n\longrightarrow\infty} s_n^{-2-\delta} \sum_{j=1}^{n} E|X_j - \mu_j|^{2+\delta} = 0 \qquad (3.2.5)$$

então $\big(\ T_n - E(T_n) \ \big)/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Demonstração: Suponhamos que vale a condição (3.2.5). Então, para todo $\epsilon > 0$,

$$\frac{1}{s_n^2}\sum_{j=1}^{n}\int\limits_{|x-\mu_j|>\epsilon s_n}(x-\mu_j)^2 dF_{X_j}(x) \leq \frac{1}{s_n^2}\sum_{j=1}^{n}\int\limits_{|x-\mu_j|>\epsilon s_n}(x-\mu_j)^2\,\frac{|x-\mu_j|^\delta}{\epsilon^\delta s_n^\delta}dF_{X_j}(x) =$$

$$\frac{1}{\epsilon^\delta s_n^{2+\delta}}\sum_{j=1}^{n}\int\limits_{|x-\mu_j|>\epsilon s_n}|x-\mu_j|^{2+\delta}dF_{X_j}(x) \leq \frac{1}{\epsilon^\delta s_n^{2+\delta}}\sum_{j=1}^{n}\int|x-\mu_j|^{2+\delta}dF_{X_j}(x) =$$

$$\frac{1}{\epsilon^\delta s_n^{2+\delta}}\sum_{j=1}^{n} E|X_j-\mu_j|^{2+\delta} \longrightarrow 0 \ \text{ quando } \ n \longrightarrow \infty$$

Logo, pelo Teorema 3.2.3, o teorema está provado.

Observação: A condição (3.2.5) é chamada *condição de Liapunov*. Como a condição de Lindeberg, a condição de Liapunov implica que $\max\limits_{1\leq j\leq n}\frac{\sigma_j^2}{s_n^2} \longrightarrow 0$ quando $n \longrightarrow \infty$. De fato, suponhamos que para algum $\delta > 0$,

$$s_n^{-2-\delta}\ \sum_{j=1}^{n} E|X_j-\mu_j|^{2+\delta} \longrightarrow 0 \ \text{ quando } n \longrightarrow \infty.$$

Como a função $f(x) = x^{\frac{2}{2+\delta}}$, $x > 0$, é côncava temos, pela Desigualdade de Jensen (Teorema A.9 do Apêndice A), que, para todo $n \geq 1$ e para todo j, $1 \leq j \leq n$,

$$\sigma_j^2 = E(X_j-\mu_j)^2 = E\big\{\big(\ |X_j-\mu_j|^{2+\delta}\ \big)^{\frac{2}{2+\delta}}\big\} \leq \{\ E|X_j-\mu_j|^{2+\delta}\ \}^{\frac{2}{2+\delta}}.$$

Logo,

$$\max_{1\leq j\leq n}\frac{\sigma_j^2}{s_n^2} \leq \max_{1\leq j\leq n}\ \Big\{\ \frac{E|X_j-\mu_j|^{2+\delta}}{s_n^{2+\delta}}\ \Big\}^{\frac{2}{2+\delta}} =$$

$$\Big\{\ \max_{1\leq j\leq n}\ \frac{E|X_j-\mu_j|^{2+\delta}}{s_n^{2+\delta}}\ \Big\}^{\frac{2}{2+\delta}} \leq \Big\{\ \frac{1}{s_n^{2+\delta}}\ \sum_{j=1}^{n} E|X_j-\mu_j|^{2+\delta}\ \Big\}^{\frac{2}{2+\delta}} \longrightarrow 0$$

quando $n \longrightarrow \infty$.

Exemplo 3.2.4 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por*

$$\mathbb{P}(X_n = n^\alpha) = \ \mathbb{P}(X_n = -n^\alpha) = \ \frac{1}{2}, \quad n = 1, 2, \ldots \ .$$

Então, $\mu_n = E(X_n) = 0$, $\sigma_n^2 = Var(X_n) = n^{2\alpha}$ *e* $s_n^2 = \sum_{j=1}^{n} j^{2\alpha}$. *Para todo* $\alpha \geq 0$, *pelo Teorema* A.10 *do Apêndice* A, $s_n^2/n^{2\alpha+1} \longrightarrow 1/(2\alpha+1)$ *quando* $n \longrightarrow \infty$. *Logo, fazendo* $\delta = 1$ *na condição de Liapunov, temos:*

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - \mu_j|^{2+\delta} = \frac{1}{s_n^3} \sum_{j=1}^{n} E|X_j|^3 = \frac{1}{s_n^3} \sum_{j=1}^{n} j^{3\alpha} =$$

$$\frac{\sum_{j=1}^{n} j^{3\alpha}}{n^{3\alpha+1}} \left(\frac{s_n}{n^{\alpha+\frac{1}{2}}}\right)^{-3} n^{-\frac{1}{2}} \longrightarrow \left(\frac{1}{3\alpha+1}\right) (2\alpha+1)^{\frac{3}{2}} \ 0 = 0 \quad \textit{quando } n \longrightarrow \infty.$$

Por outro lado, para $-\frac{1}{2} < \alpha < 0$, *fazendo* $\delta = 1$ *na condição de Liapunov, temos:*

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - \mu_j|^{2+\delta} = s_n^{-3} \sum_{j=1}^{n} j^{3\alpha} < s_n^{-3} \sum_{j=1}^{n} j^{2\alpha} = \frac{s_n^2}{s_n^3} =$$

$$\frac{1}{s_n} = \left(\frac{s_n}{n^{\alpha+\frac{1}{2}}}\right)^{-1} \frac{1}{n^{\alpha+\frac{1}{2}}} \longrightarrow (2\alpha+1)^{\frac{1}{2}} \ 0 = 0 \textit{ quando } n \longrightarrow \infty.$$

Portanto, para todo $\alpha > -\frac{1}{2}$, $T_n/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Se as variáveis aleatórias envolvidas forem limitadas, o seguinte teorema é uma conseqüência do Teorema Limite Central de Liapunov.

Teorema 3.2.5 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes tais que* $\mu_n = E(X_n)$, $\sigma_n^2 = Var(X_n)$ *e* $\mathbb{P}(a \leq X_n \leq b) = 1$, *com* $-\infty < a < b < \infty$, $n = 1, 2, \dots$. *Se* $s_n \longrightarrow \infty$ *quando* $n \longrightarrow \infty$, *então* $(T_n - E(T_n))/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Demonstração: Como $|X_j - \mu_j| \leq b - a$, $j = 1, 2, \dots$ segue que

$$E|X_j - \mu_j|^3 = E\big\{|X_j - \mu_j| \ (X_j - \mu_j)^2\big\} \leq (b-a) \ E(X_j - \mu)^2 = (b-a) \ \sigma_j^2.$$

Logo,

$$\frac{1}{s_n^3} \sum_{j=1}^{n} E|X_j - \mu_j|^3 \leq \frac{(b-a)}{s_n^3} \sum_{j=1}^{n} \sigma_j^2 = \frac{(b-a)s_n^2}{s_n^3} = \frac{(b-a)}{s_n} \longrightarrow 0 \text{ quando } n \longrightarrow \infty.$$

Logo, como a condição de Liapunov é satisfeita para $\delta = 1$ o teorema está provado.

Exemplo 3.2.5 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por*

$$\mathbb{P}(X_n = 1) = p_n \quad e \quad \mathbb{P}(X_n = 0) = 1 - p_n, \ \ 0 < p_n < 1, \ \ n = 1, 2, \ldots \ .$$

Logo, pelo Teorema anterior, se $\sum\limits_{n=1}^{\infty} Var(X_n) = \sum\limits_{n=1}^{\infty} p_n \ (1 - p_n) = \infty$, *então*

$$(T_n - E(T_n))/s_n = \frac{T_n - \sum\limits_{j=1}^{n} p_j}{(\sum\limits_{j=1}^{n} p_j(1-p_j))^{\frac{1}{2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Até este ponto tratamos do Teorema Limite Central para somas de variáveis aleatórias independentes. Contudo, as versões de Lindeberg e Liapunov (Teoremas 3.2.3 e 3.2.4) podem ser estendidas, por exemplo, para disposições duplas ("double arrays") de variáveis aleatórias, ou seja, para seqüências do tipo $\{X_{nk}, \ 1 \leq k \leq k_n\}_{n\geq 1}$ onde $k_n \longrightarrow \infty$ quando $n \longrightarrow \infty$, tais que, para cada $n \geq 1$, $X_{n1}, X_{n2}, \ldots, X_{nk_n}$ são variáveis aleatórias independentes. Se $k_n = n$, então $\{X_{kn}, \ 1 \leq k \leq n\}_{n\geq 1}$ se diz uma disposição triangular (" triangular array") de variáveis aleatórias.

Concluímos esta seção apresentando o Teorema Limite Central de Hájek - Šidak, que é especialmente utilizado em Análise de Regressão.

Teorema 3.2.6 *(Teorema Limite Central de Hájek - Šidak) Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $\mu = E(X_1)$ *e* $\sigma^2 = Var(X_1), \ 0 < \sigma^2 < \infty$. *Seja* $\{\boldsymbol{c}_n\}_{n\geq 1} = \{(c_{n1}, c_{n2}, \ldots, c_{nn})\}_{n\geq 1}$ *uma seqüência de vetores reais não todos nulos. Se* $\max\limits_{1\leq j\leq n} c_{nj}^2 / \sum\limits_{k=1}^{n} c_{nk}^2 \longrightarrow 0$ *quando* $n \longrightarrow \infty$ *(condição de Noether), então:*

$$Z_n = \frac{\sum\limits_{j=1}^{n} c_{nj} \ (X_j - \mu)}{(\sigma^2 \ \sum\limits_{j=1}^{n} c_{nj}^2)^{\frac{1}{2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Demonstração: Sejam $Y_{nj} = c_{nj} \ X_j, \ n \geq 1$ e $1 \leq j \leq n$. Então $\{Y_{nj}, \ 1 \leq j \leq n\}_{n\geq 1}$ é uma disposição tringular ("triangular array") de variáveis aleatórias tal que, para cada $n \geq 1$, $Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ são variáveis aleatórias independentes. Seja $T_n = \sum\limits_{j=1}^{n} Y_{nj}, \ n \geq 1$. Então, $E(T_n) = \mu \ \sum\limits_{j=1}^{n} c_{nj}, \ s_n^2 = \mathrm{Var}(T_n) = \sigma^2 \ \sum\limits_{j=1}^{n} c_{nj}^2$ e pelo que observamos acima, para mostrar que $Z_n = (T_n - E(T_n))/s_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, basta mostrar que a condição de Lindeberg esta satisfeita, isto é , basta mostrar que, para todo $\epsilon > 0$,

$$\frac{1}{s_n^2}\sum_{j=1}^{n} \int\limits_{|y-c_{nj}\mu|>\epsilon s_n} (y - c_{nj}\mu)^2 dF_{Y_{nj}}(y) \longrightarrow 0 \ \text{ quando } n \longrightarrow \infty.$$

Suponhamos então que, $\max\limits_{1\le j\le n} c_{nj}^2 / \sum\limits_{k=1}^{n} c_{nk}^2 \longrightarrow 0$ quando $n \longrightarrow \infty$. Logo, denotando $\mathbf{1}_A$ a função indicadora de um evento A, temos, para todo $\epsilon > 0$:

$$\frac{1}{s_n^2}\sum_{j=1}^{n} \int\limits_{|y-c_{nj}\mu|>\epsilon s_n} (y - c_{nj}\mu)^2 dF_{Y_{nj}}(y) = \frac{1}{s_n^2}\sum_{j=1}^{n} E\big[(Y_{nj} - c_{nj}\mu)^2 \ \mathbf{1}_{(|Y_{nj}-c_{nj}\mu|>\epsilon s_n)}\big] =$$

$$\frac{1}{s_n^2}\sum_{j=1}^{n} c_{nj}^2 E\big[(X_j - \mu)^2 \ \mathbf{1}_{(c_{nj}^2 \ (X_j-\mu)^2>\epsilon^2 s_n^2)}\big] \le$$

$$\frac{1}{s_n^2}\sum_{j=1}^{n} c_{nj}^2 E\big[(X_j - \mu)^2 \ \mathbf{1}_{\big((X_j-\mu)^2>\epsilon^2 s_n^2/\max\limits_{1\le j\le n} c_{nj}^2\big)}\big] =$$

$$\frac{1}{s_n^2}\sum_{j=1}^{n} c_{nj}^2 E\big[(X_1 - \mu)^2 \ \mathbf{1}_{\big((X_1-\mu)^2>\epsilon^2 s_n^2/\max\limits_{1\le j\le n} c_{nj}^2\big)}\big] =$$

$$\frac{1}{\sigma^2} E\big[(X_1 - \mu)^2 \ \mathbf{1}_{\big((X_1-\mu)^2>\epsilon^2 \ \sigma^2 \sum\limits_{j=1}^{n} c_{nj}^2/\max\limits_{1\le j\le n} c_{nj}^2\big)}\big] =$$

$$\frac{1}{\sigma^2} E\big[(X_1 - \mu)^2 \ \mathbf{1}_{\Big(|X_1-\mu|>\epsilon \ \sigma\big(\sum\limits_{j=1}^{n} c_{nj}^2/\max\limits_{1\le j\le n} c_{nj}^2\big)^{\frac{1}{2}}\Big)}\big] \longrightarrow 0 \text{ quando } n \longrightarrow \infty,$$

pois supusemos que $\sum\limits_{j=1}^{n} c_{nj}^2 / \max\limits_{1\le j\le n} c_{nj}^2 \longrightarrow \infty$ quando $n \longrightarrow \infty$, o que prova o teorema.

Exemplo 3.2.6 *Considere o modelo linear* $Y_i = \alpha + \beta X_i + \epsilon_i, \ i = 1, 2, \ldots$ *onde* $\epsilon_1, \epsilon_2, \ldots$ *são variáveis aleatórias independentes e identicamente distribuídas tais que* $E(\epsilon_1) = 0$ *e* $\sigma^2 = Var(\epsilon_1) < \infty$ *e* $X_1, X_2, \ldots$ *são números reais tais que* $\max\limits_{1\le i\le n}(X_i - \overline{X}_n)^2 / \sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2 \longrightarrow 0$ *com* $n \longrightarrow \infty$. *O estimador de mínimos quadrados de* β *é dado por*

$$\hat{\beta}_n = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)Y_i}{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2} = \frac{1}{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2} \ \sum_{i=1}^{n}(X_i - \overline{X}_n)\{\alpha + \beta X_i + \epsilon_i\} =$$

$$\beta + \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)}{\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2}\ \epsilon_i$$

e então podemos escrever:

$$\{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\}^{\frac{1}{2}}\ (\hat{\beta}_n - \beta) = \sum_{i=1}^{n} c_{ni}\ \epsilon_i,$$

onde

$$c_{ni} = (X_i - \overline{X}_n)/\{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\}^{\frac{1}{2}}.$$

Como $\max\limits_{1 \le i \le n} c_{ni}^2 / \sum\limits_{i=1}^{n} c_{ni}^2 \longrightarrow 0$ *quando* $n \longrightarrow \infty$, *pelo Teorema Limite Central de Hájek - Šidak (Teorema 3.2.6) podemos concluir que:*

$$\{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\}^{\frac{1}{2}}\ (\hat{\beta}_n - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Em termos práticos, podemos dizer que para todo n *suficientemente grande* $\hat{\beta}_n$ *tem, aproximadamente, uma distribuição Normal com média* β *e variância* $\sigma^2 / \sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$. *Note que não é necessário especificar a forma da distribuição de* ϵ_1 *para que esse resultado seja válido; também é conveniente lembrar que se essa distribuição for Normal, o resultado vale para qualquer* $n \ge 2$.

Concluindo esta seção, mostraremos que a versão clássica do Teorema Limite Central (Teorema 3.2.1) é válida para seqüências de *vetores aleatórios* independentes e identicamente distribuídos. Seja $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)$ um vetor aleatório $p-$dimensional com vetor de médias $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$ e matriz de covariância $\mathbf{\Sigma}$ finita. Se $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ então, a função característica de $\mathbf{Z}$ é dada por:

$$\varphi_{\mathbf{Z}}(\mathbf{t}) = e^{i\mathbf{t}\boldsymbol{\mu}^t - \frac{1}{2}\mathbf{t}\mathbf{\Sigma}\mathbf{t}^t}\ ,\quad \forall \mathbf{t}\ \in \mathbb{R}^p.$$

O seguinte teorema apresenta, em termos das distribuições de combinações lineares das componentes de um vetor aleatório $\mathbf{Z}$, uma condição necessária e suficiente para que ele tenha uma distribuição Normal multivariada. Esse resultado é bastante útil na demonstração do Teorema Limite Central para vetores aleatórios.

Teorema 3.2.7 *$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_p) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se, e somente se, para todo $\boldsymbol{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p$,*

$$\boldsymbol{t}\ \boldsymbol{Z}^t = \sum_{j=1}^{p} t_j\ Z_j \sim \mathcal{N}(\boldsymbol{t}\ \boldsymbol{\mu}^t, \boldsymbol{t}\ \boldsymbol{\Sigma}\boldsymbol{t}^t).$$

Demonstração; Suponhamos que $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Então, para todo $\mathbf{t} \in \mathbb{R}^p$, a função característica de $\mathbf{t}\ \mathbf{Z}^t$ é a função característica de uma variável aleatória Normal com média $\mathbf{t}\ \boldsymbol{\mu}^t$ e variância $\mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t$. De fato, para todo $s \in \mathbb{R}$:

$$\varphi_{\mathbf{t}\ \mathbf{Z}^t}(s) = E(e^{is\mathbf{t}\ \mathbf{Z}^t}) = \varphi_{\mathbf{Z}}(s\mathbf{t}) = e^{is\mathbf{t}\ \boldsymbol{\mu}^t - \frac{1}{2}s\mathbf{t}\ \boldsymbol{\Sigma}(s\mathbf{t})^t} = e^{is\mathbf{t}\ \boldsymbol{\mu}^t - \frac{1}{2}s^2\mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t}.$$

Logo, $\mathbf{t}\ \mathbf{Z}^t \sim \mathcal{N}(\mathbf{t}\ \boldsymbol{\mu}^t, \mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t)$.

Reciprocamente, suponhamos que, para todo $\mathbf{t} \in \mathbb{R}^p$, $\mathbf{t}\ \mathbf{Z}^t \sim \mathcal{N}(\mathbf{t}\ \boldsymbol{\mu}^t, \mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t)$. Então,

$$\varphi_{\mathbf{Z}}(\mathbf{t}\ ) = E(e^{i\mathbf{t}\ \mathbf{Z}^t}) = \varphi_{\mathbf{t}\ \mathbf{Z}^t}(1) = e^{i\mathbf{t}\ \boldsymbol{\mu}^t - \frac{1}{2}\mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t}$$

o que prova o teorema.

Teorema 3.2.8 *(Teorema Limite Central para seqüências de vetores aleatórios independentes e identicamente distribuídos) Seja $\{\boldsymbol{X}_n\}_{n\geq 1} = \{(X_{n1}, X_{n2}, \ldots, X_{np})\}_{n\geq 1}$ uma seqüência de vetores aleatórios $p-$dimensionais independentes e identicamente distribuídos. Sejam $\boldsymbol{\mu}$ o vetor de médias e $\boldsymbol{\Sigma}$ a matriz de covariância de $\boldsymbol{X}_1$, onde $\boldsymbol{\Sigma} = (\sigma_{ij})_{1\leq i,j\leq p}$ é finita. Então,*

$$\boldsymbol{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}\ , \boldsymbol{\Sigma}).$$

Demonstração: Pelos Teoremas de Cramér - Wold (Teorema 2.3.13) e Teorema 3.2.7 basta mostrar que, para todo $\mathbf{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n}\ \sum_{j=1}^{p}\ t_j(X_{ij} - \mu_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{t}\ \boldsymbol{\Sigma}\mathbf{t}^t).$$

Seja então $Y_i = \sum\limits_{j=1}^{p} t_j(X_{ij} - \mu_j)$, $i = 1, 2, \ldots, n$. Como $\{Y_n\}_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E(Y_1) = 0$ e

$$\mathrm{Var}(Y_1) = E\big[\big(\sum_{j=1}^{p} t_j(X_{ij} - \mu_j)\big)^2\big] =$$

$$E\big[\big(\sum_{j=1}^{p} t_j(X_{ij}-\mu_j)\big)\ \big(\sum_{k=1}^{p} t_k(X_{1k}-\mu_k)\big)\big] = E\big(\sum_{j=1}^{p}\sum_{k=1}^{p} t_j t_k (X_{1j}-\mu_j)(X_{1k}-\mu_k)\big) =$$

$$\sum_{j=1}^{p}\sum_{k=1}^{p} t_j t_k \mathrm{Cov}(X_{1j}, X_{1k}) = \sum_{j=1}^{p}\sum_{k=1}^{p} t_j t_k\ \sigma_{jk} = \mathbf{t}\ \mathbf{\Sigma t}^t,$$

segue, pelo Teorema 3.2.1, que

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{t}\ \mathbf{\Sigma t}^t).$$

Logo,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{p} t_j(X_{ij}-\mu_j) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{t}\ \mathbf{\Sigma t}^t).$$

o que prova o teorema.

Exercícios

Exercício 3.2.1 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $E(X_1) = 0$ *e* $\sigma^2 = E(X_1^2)$, $0 < \sigma^2 < \infty$. *Determine o limite em distribuição das sequências* $\{V_n\}_{n\geq 1}$ *e* $\{W_n\}_{n\geq 1}$, *onde*

$$\boldsymbol{i}\ )\ \ V_n = \frac{\sqrt{n}\sum\limits_{j=1}^{n} X_j}{\sum\limits_{j=1}^{n} X_j^2}\ ,\quad n\geq 1\ ; \qquad \boldsymbol{ii}\ )\ \ W_n = \frac{\sum\limits_{j=1}^{n} X_j}{\big(\ \sum\limits_{j=1}^{n} X_j^2\ \big)^{\frac{1}{2}}}\ ,\quad n\geq 1.$$

Exercício 3.2.2 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes tais que* $X_n \sim$ *Uniforme* $\big([0,n]\big)$. *Prove que a condição de Lindeberg está satisfeita e enuncie o Teorema Limite Central correspondente.*

Exercício 3.2.3 *Use o Teorema de Liapunov para provar que*

$$e^{-n}\ \sum_{j=0}^{n} \frac{n^j}{j!}\ \longrightarrow\ \frac{1}{2}\quad \textit{quando}\quad n\longrightarrow\infty.$$

Sugestão*: Considere uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com distribuição de Poisson de parâmetro um.*

Exercício 3.2.4 *Sejam* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $E(X_1) = 0$ *e* $E(X_1^2) = 1$, $\{Y_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes cujas distribuições de probabilidade são dadas por*

$$\mathbb{P}(Y_n = \pm n) = \frac{1}{2n^2}, \quad \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n^2}$$

e suponhamos que $\{X_n\}_{n\geq 1}$ *seja independente de* $\{Y_n\}_{n\geq 1}$.
Seja $T_n = \sum_{j=1}^{n}(X_j + Y_j)$. *Prove que*

$$\frac{T_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \textit{mas que a condição de Lindeberg não vale.}$$

Exercício 3.2.5 *Seja* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com* $\mu = E(X_1)$, $\sigma^2 = Var(X_1) < \infty$, *e* $E(X_1 - \mu)^4 = \sigma^4 + 1$. *Determine*

$$\lim_{n\longrightarrow\infty} \mathbb{P}\Big(\ \Big| \ \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2 \ - \ \sigma^2 \ \Big| \ \leq \ \frac{1}{\sqrt{n}} \ \Big).$$

Sugestão*: Use o Teorema Limite Central.*

3.3 Normalidade assintótica de estatísticas que não podem ser expressas como somas de variáveis aleatórias independentes

Em muitas situações práticas, as estatísticas de interesse não podem ser expressas como somas de variáveis aleatórias independentes e, conseqüentemente, o Teorema Limite Central não pode ser diretamente aplicado para demonstrar sua normalidade assintótica. Um exemplo típico é a estatística $t = \sqrt{n}(\overline{X}_n - \mu)/s_n$ onde $\overline{X}_n$ e s_n^2 são, respectivamente, a média e a variância amostrais correspondentes a uma amostra aleatória $X_1, \ldots, X_n$ de uma variável aleatória X tal que $\mu = E(X)$ e $\sigma^2 = \mathrm{Var}(X) < \infty$. Note que se soubermos que $X \sim \mathcal{N}(\mu, \sigma^2)$ então t terá distribuição t de Student com $n-1$ graus de liberdade, para todo $n \geq 2$. Nosso interesse aqui se relaciona com situações onde a forma da distribuição de X não é especificada. Nesta seção apresentaremos duas classes de estatísticas para as quais a normalidade assintótica pode ser demonstrada através de estratégias simples. A primeira é o objeto do seguinte teorema:

Teorema 3.3.1 *Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X, $T_n = T(X_1, X_2, \ldots, X_n)$ uma estatística e g uma função real de variável real tal que $E\big(g(X_1)\big) = \xi$ e $Var\big(g(X_1)\big) = \nu^2$, $0 < \nu^2 < \infty$. Suponhamos que $T_n = \mathcal{G}_n + R_n$, onde $\mathcal{G}_n = \sum_{j=1}^{n} g(X_j)$ e $n^{-\frac{1}{2}}\ R_n \xrightarrow{\mathbb{P}} 0$. Então,*

$$\frac{T_n - n\ \xi}{\nu\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Demonstração: A seqüência $\{g(X_n)\}_{n\geq 1}$ é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com $E\big(g(X_1)\big) = \xi$ e $\mathrm{Var}\big(g(X_1)\big) = \nu^2$. Então, pelo Teorema 3.2.1,

$$\frac{\mathcal{G}_n - n\ \xi}{\nu\ \sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Logo, como $R_n/(\nu\sqrt{n}) \xrightarrow{\mathbb{P}} 0$ e

$$\frac{T_n - n\ \xi}{\nu\ \sqrt{n}} = \frac{\mathcal{G}_n - n\ \xi}{\nu\ \sqrt{n}} + \frac{R_n}{\nu\ \sqrt{n}}$$

segue, pelo Teorema de Slutsky (Teorema 2.3.16), que

$$\frac{T_n - n\ \xi}{\nu\ \sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Exemplo 3.3.1 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $\mu = E(X)$, $\sigma^2 = Var(X) > 0$, $\mu_4 = E(X-\mu)^4 < \infty$ e $\gamma^2 = Var\big((X-\mu)^2\big) = \mu_4 - \sigma^4$.*

Consideremos $S_n^2 = (n-1)^{-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$ a variância amostral correspondente, e façamos $T_n = \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$. Então $T_n = \sum_{j=1}^{n} (X_j - \mu)^2 - n\ (\overline{X}_n - \mu)^2 = \mathcal{G}_n + R_n$, onde $\mathcal{G}_n = \sum_{j=1}^{n} (X_j - \mu)^2$ e $R_n = -n\ (\overline{X}_n - \mu)^2$.

Pelo Teorema 3.2.1,

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1).$$

Então, pelo Teorema 2.3.14,

$$\Big(\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)\Big)^2 \xrightarrow{\mathcal{D}} Z^2 \sim \chi^2_{(1)}$$

e, como $-\sigma^2/\sqrt{n} \longrightarrow 0$ quando $n \longrightarrow \infty$ segue, pelo Teorema de Slutsky (Teorema 2.3.16), que

$$\frac{R_n}{\sqrt{n}} = -\sqrt{n}(\overline{X}_n - \mu)^2 = (-1)\frac{\sigma^2}{\sqrt{n}}\ \Big(\ \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)\ \Big)^2 \xrightarrow{\mathcal{D}} 0$$

Logo, pelo Teorema 2.3.10, $R_n/\sqrt{n} \xrightarrow{\mathbb{P}} 0$ e, então, pelo Teorema 3.3.1*,*

$$\frac{T_n - n\ \sigma^2}{\sqrt{n}} = \frac{(n-1)\ S_n^2 - n\ \sigma^2}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\gamma^2).$$

Portanto, como $S_n^2/\sqrt{n} \xrightarrow{\mathbb{P}} 0$, temos, pelo Teorema de Slutsky (Teorema 2.3.16), que

$$\sqrt{n}(S_n^2 - \sigma^2) = \frac{(n-1)\ S_n^2 - n\ \sigma^2}{\sqrt{n}} + \frac{S_n^2}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\gamma^2).$$

Com referência ao exemplo acima, note que uma questão de interesse é saber de que forma a convergência em distribuição de $\sqrt{n}(S_n^2 - \sigma^2)$ pode ser utilizada para estudar a convergência em distribuição de $\sqrt{n}(S_n - \sigma)$. Mais especificamente, suponhamos que $\{T_n\}_{n\geq 1}$ seja uma seqüência de estatísticas tal que $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$ e que g seja uma função real de variável real. O que podemos afirmar sobre a convergência em distribuição de $\sqrt{n}\{g(T_n) - g(\mu)\}$? A resposta a esta questão será dada através dos seguintes teoremas, cujos resultados são conhecidos na literatura estatística como *Método Delta*.

Teorema 3.3.2 *Sejam* $\{T_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias,* $\mu \in \mathbb{R}$ *e* g *uma função real de variável real derivável em um intervalo que contém o ponto* μ*, com* $g'(\mu) \neq 0$. *Se* $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ *então,* $\sqrt{n}\{g(T_n) - g(\mu)\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(\ 0, [g'(\mu)]^2\ \sigma^2\ \big)$.

Demonstração: Suponhamos que $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Então, pelo Teorema `2.3.11`, $\sqrt{n}(T_n - \mu) = O_p(1)$, ou seja, $T_n - \mu = O_p(1/\sqrt{n})$ e, pelo Teorema `2.2.3`,

$$g(T_n) = g(\mu) + g'(\mu)\ (T_n - \mu) + o_p(\frac{1}{\sqrt{n}})$$

o que implica que

$$\sqrt{n}\big(\ g(T_n) - g(\mu)\ \big) = g'(\mu)\sqrt{n}\ (T_n - \mu) + o_p(1).$$

Logo, como $g'(\mu)\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\ 0, [g'(\mu)]^2\ \sigma^2\big)$, temos, pelo Teorema de Slutsky (Teorema 2.3.16), que

$$\sqrt{n}\ \big(\ g(T_n) - g(\mu)\ \big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\ 0, [g'(\mu)]^2\ \sigma^2\big).$$

Observação: Se $g'(\mu) = 0$ o resultado continua válido se a distribuição da variável aleatória degenerada no ponto zero for interpretada como sendo Normal com média zero e variância zero. Contudo, vale o seguinte Teorema:

Teorema 3.3.3 *Seja* $\{T_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias,* $\mu \in \mathbb{R}$ *e* g *uma função real de variável real derivável até a ordem* k $(k \geq 2)$ *em um intervalo que contém o ponto* μ*, com* $g^{(k)}(\mu) \neq 0$ *e* $g^{(j)}(\mu) = 0,\ j = 1, 2, \ldots, k-1$. *Se* $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$, *então*

$$n^{\frac{k}{2}}\ [g(T_n) - g(\mu)] \xrightarrow{\mathcal{D}} \frac{\sigma^k}{k!}\ g^{(k)}(\mu)\ \big[\ \mathcal{N}(0,1)\ \big]^k.$$

Demonstração: Suponhamos que $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Como vimos na demonstração do teorema anterior $T_n - \mu = O_p(\frac{1}{\sqrt{n}})$. Então, pelo Teorema 2.2.3,

$$g(T_n) = g(\mu) + \sum_{j=1}^{k} \frac{g^{(j)}(\mu)}{j!}\ (T_n - \mu)^j + o_p(n^{-\frac{k}{2}}) = g(\mu) + \frac{g^{(k)}(\mu)}{k!}\ (T_n - \mu)^k + o_p(n^{-\frac{k}{2}})$$

o que implica que

$$n^{\frac{k}{2}}\ \big(g(T_n) - g(\mu)\big) = \sigma^k \frac{g^{(k)}(\mu)}{k!}\ \big(\ \frac{\sqrt{n}}{\sigma}(T_n - \mu)\ \big)^k + o_p(1).$$

Logo, como pelo Teorema 2.3.14, $\left(\frac{\sqrt{n}}{\sigma}(T_n - \mu) \right)^k \xrightarrow{\mathcal{D}} \left[\mathcal{N}(0,1) \right]^k$, temos, pelo Teorema de Slutsky (Teorema 2.3.16), que

$$n^{\frac{k}{2}} \left(g(T_n) - g(\mu) \right) \xrightarrow{\mathcal{D}} \sigma^k \, \frac{g^{(k)}(\mu)}{k!} \, [\mathcal{N}(0,1)]^k.$$

Exemplo 3.3.2 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $\mu = E(X),\ \sigma^2 = Var(X) > 0,\ \mu_4 = E(X-\mu)^4 < \infty$ e $\gamma^2 = Var\big((X-\mu)^2\big) = \mu_4 - \sigma^4$. Consideremos o desvio padrão amostral $S_n = \{(n-1)^{-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2\}^{\frac{1}{2}}$; sabemos, pelo Exemplo 3.3.1, que $\sqrt{n}\ (S_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2)$. Seja $g(x) = x^{\frac{1}{2}},\ x > 0$. Logo, $g'(\sigma) = \frac{1}{2\sigma}$ e, pelo Teorema 3.3.2,*

$$\sqrt{n}\ (S_n - \sigma) = \sqrt{n}\ \{g(S_n^2) - g(\sigma^2)\ \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2/4\sigma^2).$$

Exemplo 3.3.3 *Suponhamos que $\{T_n\}_{n\geq 1}$ seja uma seqüência de variáveis aleatórias tal que $\sqrt{n}(T_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Seja $g(x) = x^2,\ x \in \mathbb{R}$. Logo, $g'(x) = 2x$ e $g''(x) = 2,\ x \in \mathbb{R}$. Se $\mu \neq 0$ então, pelo Teorema 3.3.2,*

$$\sqrt{n}\ (T_n^2 - \mu^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\ \mu^2\ \sigma^2);$$

Se $\mu = 0$, conforme a observação apresentada após o Teorema 3.3.2, temos que $\sqrt{n}\ T_n^2 \xrightarrow{\mathcal{D}} 0$ o que, pelo Teorema 2.3.10 implica $\sqrt{n}\ T_n^2 \xrightarrow{\mathbb{P}} 0$. (Veja Exercício 3.3.3). Se utilizarmos o Teorema 3.3.2 podemos concluir que

$$n\ T_n^2/\sigma^2 \xrightarrow{\mathcal{D}} \{\mathcal{N}(0,1)\}^2 = \chi^2_{(1)}$$

O teorema seguinte é uma generalização do Teorema 3.3.2 para o caso vetorial.

Teorema 3.3.4 *Sejam $\{\boldsymbol{T}_n\}_{n\geq 1} = \{(T_{n1}, T_{n2}, \ldots, T_{np})\}_{n\geq 1}$ uma seqüência de vetores aleatórios $p-$dimensionais, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p) \in \mathbb{R}^p$ e $g(\boldsymbol{x}) = g(x_1, x_2, \ldots, x_p)$ uma função com valores reais tal que $g^{\bullet}{}_j(\boldsymbol{x}) = \frac{\partial g(\boldsymbol{x})}{\partial x_j}$ existe em uma vizinhança de $\boldsymbol{\mu}$ e é contínua no ponto $\boldsymbol{\mu}$, para $j = 1, 2, \ldots, p$. Seja*

$$\boldsymbol{g}^{\bullet}(\boldsymbol{\mu}) = \big(g_j^{\bullet}(\boldsymbol{\mu})\big)_{1\leq j\leq p} = \big(\frac{\partial g(\boldsymbol{x})}{\partial x_j}\big|_{\boldsymbol{x}=\boldsymbol{\mu}}\big)_{1\leq j\leq p}$$

Se $\sqrt{n}\ (\boldsymbol{T}_n - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}\ , \boldsymbol{\Sigma})$ então, $\sqrt{n}\ \big(g(\boldsymbol{T}_n) - g(\boldsymbol{\mu})\big) \xrightarrow{\mathcal{D}} \mathcal{N}\ \big(\boldsymbol{0}, \boldsymbol{g}^{\bullet}(\boldsymbol{\mu})\ \ \boldsymbol{\Sigma}\ \boldsymbol{g}^{\bullet}(\boldsymbol{\mu})^t\ \big)$

Demonstração: Ver Serfling (1980, cap 3), por exemplo.

Exemplo 3.3.4 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $\mu = E(X) \neq 0$, $\sigma^2 = Var(X) > 0$, $\mu_3 = E(X-\mu)^3$ e $\mu_4 = E(X-\mu)^4 < \infty$. Sejam $\nu = \sigma/\mu$ o* coeficiente de variação *de X e $\nu_n = S_n/\overline{X}_n$ o correspondente* coeficiente de variação amostral. *Consideremos o vetor $\boldsymbol{\mu} = (\mu, \sigma^2)$ e as estatísticas $\boldsymbol{T}_n = (\overline{X}_n, S_n^2)$. Pelo Exercício 3.3.4 sabemos que*

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\mu}) = \sqrt{n}(\overline{X}_n - \mu\ ,\ S_n^2 - \sigma^2) \longrightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

onde

$$\boldsymbol{\Sigma} \ = \ \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix}$$

A função g definida por: $g(x_1, x_2) = \sqrt{x}_2/x_1$, para todo $(x_1, x_2) \in \mathbb{R}^2$ com $x_1 \neq 0$ e $x_2 > 0$, é tal que $g^{\bullet}{}_1(x_1, x_2) = \partial g(x_1, x_2)/\partial x_1 = -\sqrt{x}_2/x_1^2$ e

$$g^{\bullet}{}_2(x_1, x_2) = \frac{\partial g(x_1, x_2)}{\partial x_2} = \frac{1}{2x_1\sqrt{x}_2}.$$

Então, $\boldsymbol{g}^{\bullet}(\boldsymbol{\mu}) = (-\sigma/\mu^2\ ,\ 1/2\mu\sigma)$ e, pelo Teorema 3.3.4, temos:

$$\sqrt{n}(\nu_n - \nu) \ = \ \sqrt{n}(\frac{S_n}{\overline{X}_n} - \frac{\sigma}{\nu}) =$$

$$\sqrt{n}\{g(\boldsymbol{T}_n) - g(\boldsymbol{\mu})\} \xrightarrow{\mathcal{D}} \mathcal{N}\big(\ 0\ , (-\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma})\ \boldsymbol{\Sigma}\ (-\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma})^t\ \big)$$

Mas,

$$(-\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma})\ \boldsymbol{\Sigma}\ (-\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma})^t = \frac{\sigma^4}{\mu^4} - \frac{\mu_3}{\mu^3} + \frac{\mu_4}{4\mu^2\sigma^2} - \frac{\sigma^2}{4\mu^2}$$

Logo,

$$\sqrt{n}(\nu_n - \nu) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\ 0\ ,\ \frac{\sigma^4}{\mu^4} - \frac{\mu_3}{\mu^3} + \frac{\mu_4}{4\mu^2\sigma^2} - \frac{\sigma^2}{4\mu^2}\big).$$

Para concluir esta seção, consideremos novamente o Exemplo 3.3.1 onde mostramos que $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \mu_4 - \sigma^4\big)$; esse resultado tem pouca utilidade prática para fazermos inferências sobre σ^2, pois as regiões críticas de testes de hipóteses sobre σ^2 ou os limites de intervalos de confiança para σ^2 baseados nessa distribuição assintótica dependerão do parâmetro desconhecido. De uma forma geral, é comum depararmos com situações onde, dada uma seqüência de estimadores $\{T_n\}_{n\geq 1}$ de um parâmetro $\theta \in \Theta \subset \mathbb{R}$, o Teorema Limite Central garante que $\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2)$ onde a variância assintótica γ^2 é uma função de θ, digamos $\gamma^2 = h(\theta)$. Para podermos fazer inferência sobre θ gostaríamos de obter uma função (transformação) $g : \mathbb{R} \longrightarrow \mathbb{R}$ tal que $\sqrt{n}\{g(T_n) - g(\theta)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2)$ onde c seja uma constante que não dependa de θ. Esse tipo de transformação é conhecido como *transformação estabilizadora da variância* e pode ser determinado a partir do seguinte resultado:

Teorema 3.3.5 *Se* $\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, h(\theta)\big)$ *e* $g(x)$ *é solução da equação diferencial* $\frac{dg}{dx} = \frac{c}{\sqrt{h(x)}}$, *onde* $c \neq 0$ *é uma constante que não depende de* θ, *então*

$$\sqrt{n}\{\ g(T_n) - g(\theta)\ \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2)$$

Demonstração: Suponhamos que g seja solução da equação diferencial $\frac{dg}{dx} = \frac{c}{\sqrt{h(x)}}$. Então,

$$g'(\theta) = \frac{dg(x)}{dx}\Big|_{x=\theta} = \frac{c}{\sqrt{h(\theta)}}$$

o que implica que $[\ g'(\theta)\ ]^2 = \frac{c^2}{h(\theta)} \neq 0$. Portanto, pelo Teorema 3.3.2,

$$\sqrt{n}\{\ g(T_n) - g(\theta)\ \} \xrightarrow{\mathcal{D}} \mathcal{N}(0, c^2).$$

Exemplo 3.3.5 *Sejam* $\{X_n\}_{n\geq 1}$ *uma seqüência de variáveis aleatórias tal que* $X_n \sim$ *Binomial* (n, p), $0 < p < 1$ *e* $T_n = X_n/n$. *Então,* $\sqrt{n}(T_n - p) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, p(1-p)\big)$. *A transformação estabilizadora da variância, g, é , nesse caso, solução da equação* $\frac{dg}{dx} = \frac{c}{\sqrt{x(1-x)}}$, $0 < x < 1$. *Então,* $g(x) = \int \frac{c}{\sqrt{x(1-x)}}dx$ *e, considerando a transformação* $x = sen^2 y$, *temos:*

$$g(x) = \int \frac{2\ c\ sen y \cos y}{sen y \cos y} dy = 2c \int dy = 2cy = 2c(\ arcsen\sqrt{x}).$$

Escolhendo, por conveniência, $c = 1/2$ *temos* $g(x) = \ arcsen\sqrt{x}$.

Logo, pelo Teorema 3.3.5,

$$\sqrt{n}(\ \ arcsen\sqrt{T}_n - \ arcsen\sqrt{p}\ ) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{4}).$$

e para todo n suficientemente grande,

$$(\ \ arcsen\sqrt{T}_n - \ arcsen\sqrt{p}\ ) \approx \mathcal{N}(0, \frac{1}{4n})$$

Esta aproximação é razoavelmente boa para $n \geq 10$. A partir dessa aproximação obtemos o seguinte intervalo de confiança para p, com coeficiente de confiança $1-\alpha$:

$$\left(sen^2\{\ arcsen\sqrt{T}_n - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}\}\ ,\ sen^2\{\ arcsen\sqrt{T}_n + \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}\} \right)$$

onde $z_{\frac{\alpha}{2}}$ é o percentil $100(1-\frac{\alpha}{2})$ de uma distribuição $\mathcal{N}(0,1)$.

Exercícios

Exercício 3.3.1 *Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias tais que $X_n \sim$ Binomial (n,p), $0<p<1$. Se $T_n = X_n/n, \quad n \geq 1$,*

i) prove que $\sqrt{n}(T_n - p) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, p(1-p)\big)$;

ii) determine o limite em distribuição de $\sqrt{n}(\frac{1}{T_n} - \frac{1}{p})$ e $\sqrt{n}\big(T_n(1-T_n) - p(1-p)\big)$

Exercício 3.3.2 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim$ Uniforme $\big([0,\theta]\big)$, onde $\theta > 0$. Prove que $\sqrt{n}\{\log(2\overline{X}_n) - \log(\theta)\}$ converge em distribuição. Identifique o limite.*

Exercício 3.3.3 *Seja $\{T_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias tal que $\sqrt{n}T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$. Prove que $\sqrt{n}T_n^2 \xrightarrow{\mathbb{P}} 0$ sem utilizar a observação feita após a demonstração do Teorema 3.3.2.*

Sugestão*: note que $\sqrt{n}\ T_n^2 = \sqrt{n}\ T_n\ T_n$ e prove que $\sqrt{n}T_n = O_p(1)$ e $T_n = o_p(1)$.*

Exercício 3.3.4 *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma variável aleatória X com $\mu = E(X),\ \sigma^2 = Var(X) > 0,\ \ \mu_3 = E(X-\mu)^3,\ $ e $\ \mu_4 = E(X-\mu)^4 < \infty$. Sejam $\overline{X}_n$ a média amostral e $S_n^2 = (n-1)^{-1}\sum\limits_{j=1}^{n}(X_j - \overline{X}_n)^2$ a variância amostral. Prove que*

$$\sqrt{n}\big(\overline{X}_n - \mu\ ,\ S_n^2 - \sigma^2\big) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}\ ,\boldsymbol{\Sigma})$$

onde

$$\boldsymbol{0}\ = (0,0) \quad e \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix}$$

Sugestão*:*

i) Mostre que

$$S_n^2 - \sigma^2 = (\frac{n-1}{n})\ (S_n^2 - \sigma^2) + \frac{1}{n}\ (S_n^2 - \sigma^2) = \frac{1}{n}\sum_{j=1}^{n} Y_j + R_n,$$

onde $Y_j = (X_j - \mu)^2 - \sigma^2,\ 1 \leq j \leq n$ e $\sqrt{n}\ R_n = o_p(1)$.

ii) Utilize o Teorema 3.3.1.

Exercício 3.3.5 *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de uma variável aleatória* X *com distribuição de Poisson* $(\theta), \ \theta > 0,$ *e seja* $\overline{X}_n$ *a média amostral correspondente.*

i) Prove que $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \theta);$

ii) determine uma transformação estabilizadora da variância e conclua que

$$\sqrt{n}\big(\sqrt{\overline{X}_n} - \sqrt{\theta} \ \big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{4}).$$

Exercício 3.3.6 *Seja* $X_1, X_2, \ldots, X_n$ *uma amostra aleatória de* $X \sim$ *Uniforme* $\big([0, \theta]\big), \ \theta > 0.$

i) Prove que $\sqrt{2}\ (2\ \overline{X}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{\theta^2}{3}).$

ii) determine uma transformação estabilizadora da variância e conclua que

$$\sqrt{n}\big(\ \log(2\ \overline{X}_n) - \log\theta \ \big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{3}).$$

3.4 Expansões relativas ao Teorema Limite Central

Seja $\{X_n\}_{n\geq 1}$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com função distribuição F, média μ, variância σ^2, finita e função característica φ. Suponhamos, sem perda de generalidade, que $\mu = E(X_1) = 0$. Vimos na Seção 3.2 que $Z_n = \sum_{j=1}^{n} X_j/(\sigma\sqrt{n}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ (Teorema `3.2.1`). Isto significa que, se $\{F_n\}_{n\geq 1}$ for uma seqüência das funções distribuição de Z_n e Φ for a função distribuição de uma variável aleatória Normal com média zero e variância um então, para todo $x \in \mathbb{R}$, $F_n(x) \longrightarrow \Phi(x)$ quando $n \longrightarrow \infty$. Além disso, como Φ é contínua em $\mathbb{R}$, $F_n \longrightarrow \Phi$ uniformemente, ou seja,

$$\lim_{n\longrightarrow\infty} \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| = 0$$

O que gostaríamos de saber é "quão rapidamente" ou "com que velocidade" $\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \longrightarrow 0$ quando $n \longrightarrow \infty$. Em outras palavras, gostaríamos de ter uma idéia do erro que cometemos ao utilizar as aproximações sugeridas pelo Teorema Limite Central para um tamanho de amostra n finito. O seguinte resultado responde essa questão, embora de forma não totalmente satisfatória.

Teorema 3.4.1 (Berry - Esséen) *Se* $\rho = \int_{-\infty}^{+\infty} |x|^3 dF(x) < \infty$ *então existe uma constante* $\mathcal{C} > 0$ *tal que*

$$\sup_{x\in\mathbb{R}} \mid F_n(x) - \Phi(x) \mid \ \leq \mathcal{C}\ \frac{\rho}{\sigma^3\ \sqrt{n}}\ , \quad \forall n \geq 1.$$

Demonstração: Ver Feller (1966), por exemplo.

Observações:

i) Note que $\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq O\ (1/\sqrt{n})$, isto é , a ordem de magnitude de $\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)|$ é menor ou igual a de $1/\sqrt{n}$.

ii) Feller (1966), Zolotarev (1967) e Van Beeck (1972) obtiveram, respectivamente, os seguintes valores para $\mathcal{C}$: $8,25$; $0,91$ e $0,7975$.

iii) Contudo, Esséen (1956) provou o seguinte resultado:

$$\limsup_{n\longrightarrow\infty} \Big[\ \frac{\sigma^3\ \sqrt{n}}{\rho}\ \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)|\ \Big] \ = \ \frac{3+\sqrt{10}}{6\ \sqrt{2\ \pi}} \cong 0,41,$$

que implica que $\mathcal{C}\ \leq 0,41$.

Se $E(|X_1|^j) < \infty,\ j = 3, 4, 5, \ldots, k$, uma outra forma de avaliar as taxas de convergência mencionadas acima está relacionada com as expansões de Edgeworth para F_n e f_n, a função densidade de probabilidade de Z_n (quando existir). No caso de $E(|X_1|^3) < \infty$, por exemplo, temos:

a) se ρ^r for integrável para algum $r \geq 1$ então, f_n existe para todo $n \geq r$ e

$$f_n(x) = \Phi'(x) + \frac{\mu_3}{6\ \sigma^3\ \sqrt{n}}(x^3 - 3x)\ \Phi'(x) + o\ (1/\sqrt{n}) \quad \text{uniformemente em } x,$$

onde $\mu_3 = E(X_1^3)$ e $\Phi'(x) = \frac{1}{\sqrt{2}\ \pi} e^{-\frac{x^2}{2}},\ x \in \mathbb{R}$.

b) se $\limsup\limits_{|t| \longrightarrow \infty}\ |\varphi(t)| < 1$, então

$$F_n(x) = \Phi(x) + \frac{\mu_3}{6\ \sigma^3\ \sqrt{n}}(1 - x^2)\ \Phi'(x) + o\ (\frac{1}{\sqrt{n}})$$

uniformemente em x.

Embora essas expansões permitam uma melhor avaliação das taxas de convergência do que aquela sugerida pelo Teorema de Berry - Esséen, elas têm a desvantagem de requerer o conhecimento dos momentos da distribuição geradora dos dados. O leitor interessado poderá consultar Cramér (1946, cap 17) para maiores detalhes.

Capítulo 4

Comportamento Assintótico de Estimadores e Estatísticas de Teste

4.1 Introdução

Consideremos uma família de espaços de probabilidade $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$, indexada pelo parâmetro $\theta \in \Theta \subset \mathbb{R}$ e seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de alguma distribuição específica (porém desconhecida) $\mathbb{P}_\theta$ dessa família. Nesse contexto, o objetivo da *inferência paramétrica* é o de propor e avaliar métodos de seleção de estatísticas $T_n = T_n(X_1, \ldots, X_n)$ apropriados para *estimar* θ (isto é , "adivinhar" seu verdadeiro valor) ou para *testar hipóteses* sobre θ (isto é , decidir se $\theta \in \Theta^\circ \subset \Theta$, ou não); no primeiro caso, T_n é chamada de *estimador* de θ (e é usualmente denotada $\hat{\theta}_n$) e no segundo caso, de *estatística de teste*.

Entre os métodos usualmente empregados para obter tais estimadores e estatísticas de teste, alguns tem um papel fundamental, principalmente devido a sua interpretação e facilidade computacional. Em muitos casos, contudo, suas propriedades estatísticas em amostras pequenas são difíceis de serem obtidas e conseqüentemente o estudo de seu comportamento assintótico tem grande importância prática. Neste capítulo apresentamos algumas propriedades assintóticas de estatísticas obtidas através dos métodos de Máxima Verossimilhança, Mínimos Quadrados e outros métodos comumente empregados em inferência paramétrica. Por razões de dificuldade técnica, damos mais ênfase aos problemas de estimação, que introduzimos a seguir.

Uma seqüência $\{\hat{\theta}_n\}_{n\geq 1}$ de estimadores (de uma parâmetro θ) é (fracamente) *consis-*

tente se $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$; por outro lado, se $\hat{\theta}_n \xrightarrow{q.c.} \theta$, a seqüência $\{\hat{\theta}_n\}_{n\geq 1}$ é dita *fortemente consistente*. Estimadores consistentes correspondem essencialmente a estimadores cuja precisão aumenta com o aumento do tamanho da amostra na qual eles estão baseados; a propriedade de consistência, em geral, pode ser demonstrada sob condições de regularidade bastante amplas com a utilização das técnicas estudadas no Capítulo 2. Para a maioria das aplicações práticas, no entanto, essa propriedade tem um valor limitado se não estiver acoplada a algum tipo de convergência em distribuição. Por esse motivo não consideraremos consistência isoladamente nestas notas, embora as demonstrações de convergência em distribuição geralmente necessitem de suposições mais fortes.

Na Seção 4.2 discutimos as propriedades assintóticas de estimadores de **MV**; na Seção 4.3 consideramos outros tipos de estimadores freqüentemente utilizados na prática; o conceito de *eficiência assintótica*, fundamental para a seleção do "melhor" estimador dentre um conjunto de candidatos é brevemente discutido na Seção 4.4 e finalmente, na Seção 4.5 indicamos como as propriedades assintóticas de algumas estatísticas de teste podem ser estudadas através dos métodos apresentados nos capítulos anteriores.

4.2 Comportamento Assintótico de Estimadores de Máxima Verossimilhança

O método de **MV** introduzido por Fisher em 1922 é certamente uma das técnicas mais utilizadas em estimação paramétrica. Em geral ele produz equações bastante convenientes do ponto de vista computacional e estimadores com boas propriedades estatísticas.

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com função densidade $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}$. 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 (regulares) os estimadores de **MV** podem ser obtidos através da maximização do *logaritmo da função de verossimilhança*, o que, em geral, equivale à obtenção da solução da *equação de verossimilhança*:

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

A função $\Lambda_n(\theta) = \Lambda_n(\theta; X_1, X_2, \ldots, X_n)$ é chamada de *função de estimação* quando considerada como função de θ e de *estatística escore* quando encarada como função de $X_1, X_2, \ldots, X_n$. Consideremos alguns exemplos:

Exemplo 4.2.1 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. A função de verossimilhança é dada por : $L_n(\theta) = (2\ \pi)^{-\frac{n}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2\}$ e $\sup_{\theta\in\mathbb{R}} L_n(\theta)$ corresponde a $\min_{\theta\in\mathbb{R}} \sum_{i=1}^{n}(X_i-\theta)^2$ o que implica que o estimador de **MV** de θ é $\hat{\theta}_n = \overline{X}_n$.*

Exemplo 4.2.2 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com distribuição Bernoulli (θ), $0 < \theta < 1$. A função de verossimilhança é $L_n(\theta) = \theta^{T_n}\ (1-\theta)^{n-T_n}$, onde $T_n = \sum_{i=1}^{n} X_i$ e a função de estimação é $\Lambda_n(\theta) = (T_n - n\theta)/\theta(1-\theta)$, uma função não linear de θ com uma única raiz $\hat{\theta}_n = T_n/n$.*

Exemplo 4.2.3 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com distribuição Cauchy (θ), isto é , com função densidade $f(x;\theta) = [\ \pi\ \{1+(x-\theta)^2\ \}\ ]^{-1}$, $\theta \in \mathbb{R}$. A função de verossimilhança é $L_n(\theta) = \pi^{-n} \prod_{i=1}^{n}\{1+(X_i-\theta)^2\}^{-1}$; a função de estimação, $\Lambda_n(\theta) = 2\ \sum_{i=1}^{n}(X_i-\theta)/\{1+(X_i-\theta)^2\}$ se comporta como um polinômio de grau $2n-1$ em θ e,* *conseqüentemente $L_n(\theta)$ pode ter vários pontos de mínimo e máximo. Nesse caso outras técnicas precisam ser consideradas para a determinação de estimador de **MV** de θ.*

Exemplo 4.2.4 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com distribuição Uniforme $([0, \theta])$, isto é , com função densidade $f(x;\theta) = \theta^{-1}\ \mathbf{1}_{[0,\theta]}(x)$. Aqui, a função de verossimilhança $L_n(\theta) = \theta^{-n}\ \mathbf{1}_{[0,\theta]}(X_{n:n})$, onde $X_{n:n} = \max(X_1, \ldots, X_n)$ não é diferenciável e a função de estimação não existe. No entanto, é claro que a função de verossimilhança é maximizada no ponto $\hat{\theta}_n = X_{n:n}$, que é o estimador de **MV** de θ.*

Note que, nos casos onde uma estatística suficiente (para θ) T_n existe, podemos escrever $L_n(\theta) = g(T_n|\theta)\ h(X_1, X_2, \ldots, X_n)$ e claramente, a maximização de $L_n(\theta)$ equivale à maximização de $g(T_n|\theta)$; como conseqüência, o estimador de **MV** $\hat{\theta}_n$ será uma função de T_n, somente. Note também, que todos os exemplos apresentados, com exceção de Exemplo 4.2.3, estão nessa categoria.

Em geral, os estimadores de **MV** não podem ser expressos explícitamente e necessitamos métodos iterativos para obtenção das raízes da equação de verossimilhança. 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 seguinte expansão de Taylor em torno de alguma estimativa inicial $\theta_n^{(\circ)}$:

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

onde θ_n^* é um valor entre $\hat{\theta}_n$ e $\theta_n^{(\circ)}$. Então, temos:

$$\hat{\theta}_n = \theta_n^{(\circ)} - \left\{\frac{\partial}{\partial\theta}\log L_n(\theta)|_{\theta=\theta_n^{(\circ)}}\right\}\ \left\{\frac{\partial^2}{\partial\theta^2}\log L_n(\theta)|_{\theta=\theta_n^*}\right\}^{-1} \tag{4.2.1}$$

Se escolhemos $\theta_n^{(\circ)}$ numa vizinhança de $\hat{\theta}_n$ (por exemplo, se $\theta_n^{(\circ)}$ for baseado num estimador consistente de θ), podemos escrever (4.2.1) como:

$$\hat{\theta}_n \cong \theta_n^{(\circ)} - \left\{\frac{\partial}{\partial\theta}\log L_n(\theta)|_{\theta=\theta_n^{(\circ)}}\right\}\ \left\{\frac{\partial^2}{\partial\theta^2}\log L_n(\theta)|_{\theta=\theta_n^{(\circ)}}\right\}^{-1} \tag{4.2.2}$$

O *método de Newton-Raphson* para obtenção de estimadores de **MV** de θ consiste na iteração de (4.2.2) com $\theta_n^{(\circ)}$ substituído pelo valor $\hat{\theta}_n$ obtido no passo anterior. Se, por outro lado, substituirmos $\frac{\partial^2}{\partial\theta^2}\log L_n(\theta)$ por $E\{\ \frac{\partial^2}{\partial\theta^2}\log L_n(\theta)\ \}$ em cada iteração, o procedimento é conhecido como *método "scoring" de Fisher.*

O teorema seguinte estabelece condições sob as quais estimadores de **MV** tem distribuição assintótica Normal.

Teorema 4.2.1 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função densidade $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}$ 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* $\int_{\mathbb{R}} H_j(x)dx < \infty, j = 1, 2.$

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

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

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

Demonstração: Primeiramente observamos que :

$$
\begin{aligned}
E\frac{\partial}{\partial\theta}\log f(X_1;\theta) &= \int_{\mathbb{R}} \frac{\partial}{\partial\theta}\log f(x;\theta)\ f(x;\theta)dx \\
&= \int_{\mathbb{R}} \frac{\partial}{\partial\theta}f(x;\theta)dx = \frac{\partial}{\partial\theta}\int_{\mathbb{R}} f(x;\theta)dx = 0 \qquad (4.2.3)
\end{aligned}
$$

Note que a ordem de integração e diferenciação em (4.2.3) é permutável pois `i`) permite uma aplicação direta do Teorema da Convergência Dominada de Lebesgue. Em seguida observemos que :

$$
\begin{aligned}
E\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta) &= \int_{\mathbb{R}} \frac{\partial}{\partial\theta}\Big[\ \frac{\partial}{\partial\theta}f(x;\theta)\ \{f(x;\theta)\}^{-1}\ \Big]\ f(x;\theta)dx \\
&= \int_{\mathbb{R}} \Big[\ \frac{\partial^2}{\partial\theta^2}f(x;\theta)\ \frac{1}{f(x;\theta)} - \{\frac{\partial}{\partial\theta}f(x;\theta)\ \frac{1}{f(x;\theta)}\}^2\ \Big]\ f(x;\theta)dx \\
&= \int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2}f(x;\theta)dx - I_\theta = \int_{\mathbb{R}} \frac{\partial}{\partial\theta}\{\frac{\partial}{\partial\theta}f(x;\theta)\}dx - I_\theta \\
&= \frac{\partial}{\partial\theta}\int_{\mathbb{R}} \frac{\partial}{\partial\theta}f(x;\theta)dx - I_\theta = -I_\theta. \qquad (4.2.4)
\end{aligned}
$$

Fazendo $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(X_i;\theta)$ e utilizando (4.2.3) e `ii a`) em conjunto com o Teorema Limite Central (Teorema 3.2.1) podemos concluir que $U_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_\theta)$. Por outro lado, fazendo $V_n = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)$ e utilizando (4.2.4) em conjunto com a Lei Forte dos Grandes Números de Khintchine (Teorema 2.4.5) podemos concluir que:

$$V_n \xrightarrow{q.c.} -I_\theta \qquad (4.2.5)$$

Então, para $|u| \leq K,\ 0 < K < \infty$, façamos :

$$\lambda_n(u) = \log L_n(\theta + n^{-\frac{1}{2}}\ u) - \log L_n(\theta) = \sum_{i=1}^{n}\{\ \log f(X_i;\theta + n^{-\frac{1}{2}}\ u) - \log f(X_i;\theta)\ \}$$

e consideremos a seguinte expansão de Taylor de $\log f(X_i;\theta+n^{-\frac{1}{2}}u)$ em torno do ponto θ :

$$\lambda_n(u) = u\ \frac{1}{\sqrt{n}}\ \sum_{i=1}^{n}\ \frac{\partial}{\partial\theta}\log f(X_i;\theta) + \frac{u^2}{2}\ \frac{1}{n}\sum_{i=1}^{n}\ \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta_n^*)$$

onde $\theta_n^* \in (\theta, \theta + n^{-\frac{1}{2}}\ u)$. Definindo

$$Z_n(u) = \frac{1}{n}\sum_{i=1}^{n}\{\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta_n^*) - \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)\}$$

temos

$$\lambda_n(u) = u\ U_n + \frac{u^2}{2}\ V_n + \frac{u^2}{2}\ Z_n(u). \tag{4.2.6}$$

Agora, dado $\delta > 0$, existe $n_0 = n_0(\delta)$ tal que $\frac{|u|}{\sqrt{n}} \leq \frac{K}{\sqrt{n}} < \delta$ para todo $n \geq n_0$. Conseqüentemente, para n suficientemente grande, temos :

$$\begin{aligned} |Z_n(u)| &\leq \frac{1}{n}\sum_{i=1}^{n}\ \sup_{\{h:|h|\leq|u|/\sqrt{n}\}}\ |\ \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta+h) - \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)\ | \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\ \sup_{\{h:|h|\leq\delta\}}\ |\ \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta+h) - \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)\ | \qquad (4.2.7) \end{aligned}$$

Utilizando a Lei Forte dos Grande Números de Khintchine (Teorema 2.4.5) e a suposição `ii b`), temos:

$$\frac{1}{n}\sum_{i=1}^{n}\ \sup_{\{h:|h|\leq\delta\}}\ |\ \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta+h) - \frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)\ | \xrightarrow{q.c.} \Psi_\delta,$$

que converge para zero com $\delta \longrightarrow 0$; como δ é arbitrário, concluímos por (4.2.7) que para $|u| \leq K$:

$$|\ Z_n(u)\ | \xrightarrow{q.c.} 0 \tag{4.2.8}$$

Agora, reescrevendo (`4.2.6`) como :

$$\lambda_n(u) = u\ U_n - \frac{u^2}{2}\ I_\theta + \frac{u^2}{2}\ \{V_n - I_\theta\} + \frac{u^2}{2}\ Z_n(u)$$

e utilizando (4.2.5) e (4.2.8) temos para $|u| \leq K$:

$$\lambda_n(u) = u\ U_n - \frac{u^2}{2}\ I_\theta + o_p(1), \tag{4.2.9}$$

cujo máximo com respeito a u é obtido no ponto

$$\hat{u} = \frac{U_n}{I_\theta} \ + \ o_p(1) \tag{4.2.10}$$

Da definição de $\lambda_n(u)$ concluímos que este ponto também corresponde ao máximo de $L_n(\theta + n^{-\frac{1}{2}}\ u)$ que é o estimador de **MV** de $\theta, \hat{\theta}_n$. Então :

$$\hat{\theta}_n = \theta + n^{-\frac{1}{2}}\ \hat{u} = \theta + n^{-\frac{1}{2}}\ \{\frac{U_n}{I_\theta} + o_p(1)\ \}$$

o que implica que $\sqrt{n}(\hat{\theta}_n - \theta) = U_n/I_\theta + o_p(1)$; finalmente o Teorema de Slutsky (Teorema **??**) nos permite concluir que $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_\theta^{-1})$, encerrando a demonstração.

Esse enfoque para a demonstração da normalidade assintótica de estimadores de **MV** foi desenvolvido por LeCam (1957). Sua maior contribuição está relacionada com o relaxamento da suposição da existência da terceira derivada do logaritmo da função densidade exigida na demonstração clássica de Cramér (1946, cap 33).

A consistência do estimador de **MV** segue diretamente do resultado dos Teoremas 2.3.11 e 4.2.1. Demonstrações de consistência sob condições menos restritivas podem ser encontradas em diversas fontes, dentre as quais destacamos Cramér (1946, cap 33).

Exemplo 4.2.5 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com distribuição $\mathcal{N}(\theta, 1)$; como vimos no Exemplo* `4.2.1` *o estimador de **MV** de θ é $\overline{X}_n$. Além disso, temos : $\frac{\partial}{\partial\theta}\log f(x;\theta) = x - \theta$, $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = -1$, $\Psi_\delta = 0$ e $I_\theta = 1$. Então, pelo Teorema 4.2.1 segue que $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.*

Exemplo 4.2.6 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com distribuição Cauchy (θ), ou seja, com função densidade dada por*
$f(x;\theta) = [\ \pi\ \{1 + (x-\theta)^2\ \}\ ]^{-1}$, $\theta \in \mathbb{R}$; então temos :

$$\frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{2\ (x-\theta)}{1+(x-\theta)^2},$$

$$\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = \frac{-2}{1+(x-\theta)^2} + \frac{4\ (x-\theta)^2}{\{1+(x-\theta)^2\}^2},$$

$$U_n = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{(X_i-\theta)}{1+(X_i-\theta)^2} \quad e \quad I_\theta = 2.$$

As condições `i` *) e* `ii a` *) do Teorema 4.2.1 seguem; como $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ é limitada e contínua, a condição* `ii b` *) também é satisfeita. Então $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{2})$. Note que, neste caso, obtivemos a distribuição assintótica do estimador de **MV** sem calculá-lo explicitamente.*

O teorema seguinte corresponde a uma generalização do Teorema 4.2.1 para o caso multi-paramétrico.

Teorema 4.2.2 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com função densidade $f(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ satisfazendo as seguintes condições :*

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

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

a) a matriz de informação de Fisher*,*

$$\boldsymbol{I_\theta} = E\Big\{ \ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_1;\boldsymbol{\theta}) \ \Big\}^t \ \Big\{ \ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_1;\boldsymbol{\theta}) \ \Big\},$$

onde $\frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_1;\boldsymbol{\theta}) = \Big[\ \frac{\partial}{\partial \theta_1} \log f(X_1;\boldsymbol{\theta}), \ldots, \frac{\partial}{\partial \theta_q} \log f(X_1;\boldsymbol{\theta}) \ \Big]$, é finita e positiva definida.

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

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

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

Demonstração: Definamos, para $||\boldsymbol{u}|| \leq K,\ 0 < K < \infty$:

$$\lambda_n(\boldsymbol{u}) = \sum_{i=1}^{n} \{ \ \log f(X_i;\boldsymbol{\theta} + n^{-\frac{1}{2}}\boldsymbol{u}) - \log f(X_i;\boldsymbol{\theta}) \ \}$$

e consideremos os mesmos passos utilizados no Teorema 4.2.1 para mostrar que : $\lambda_n(\boldsymbol{u}) = \boldsymbol{U}_n\boldsymbol{u}^t - \frac{1}{2}\boldsymbol{u}\boldsymbol{I_\theta}\boldsymbol{u}^t + o_p(1)$ onde $\boldsymbol{U}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(X_i;\boldsymbol{\theta})$; maximizando $\lambda_n(\boldsymbol{u})$ com respeito a $\boldsymbol{u}$ o resultado segue.

4.3 Comportamento Assintótico de Outros Tipos de Estimadores

O método de Máxima Verossimilhança é freqüentemente utilizado para estimação paramétrica, quando a forma da distribuição das variáveis aleatórias geradoras dos dados é conhecida. Nesta seção nós estudamos algumas propriedades assintóticas de alguns procedimentos de estimação alternativos que são usualmente empregados em situações onde não temos informação sobre a forma das distribuições envolvidas.

Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função densidade $f(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$, $q \geq 1$, $x \in \mathbb{R}$ e suponhamos que $EX_1^k = \mu_k' = h_k(\boldsymbol{\theta}) < \infty$, $k = 1, \ldots, q$; sejam também $m_{kn}' = n^{-1} \sum\limits_{i=1}^{n} X_i^k$, $k = 1, \ldots .q$. O estimador do Método de Momentos (**MM**) $\hat{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}$ é uma solução (em $\boldsymbol{\theta}$) das equações $h_k(\boldsymbol{\theta}) = m_{kn}'$, $k = 1, \ldots, q$. O seguinte teorema estabelece condições sob as quais estimadores de **MM** de $\boldsymbol{\theta}$ tem distribuição assintótica Normal.

Teorema 4.3.1 *Consideremos uma situação como a descrita acima e suponhamos que* $\mu_k' < \infty$, $k = 1, \ldots, q$. *Seja* $\boldsymbol{h}(\boldsymbol{\theta}) = [h_1(\boldsymbol{\theta}), \ldots, h_q(\boldsymbol{\theta})]$ *e vamos supor que* $\boldsymbol{H}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{h}(\boldsymbol{\theta})$ *seja tal que rank* $\boldsymbol{H}(\boldsymbol{\theta}) = q$ *e que seus elementos* $\frac{\partial}{\partial \theta_j} h_i(\boldsymbol{\theta})$, $i, j = 1, \ldots, q$ *sejam contínuos em* $\boldsymbol{\theta}$. *Então* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}\{\ \mathbf{0}, [\boldsymbol{H}^{-1}(\boldsymbol{\theta})]^t\ \boldsymbol{\Sigma}\ [\boldsymbol{H}^{-1}(\boldsymbol{\theta})]\ \}$ *onde* $\boldsymbol{\Sigma} = [\mu_{j+k}' - \mu_j'\ \mu_k']_{jk}$, $j, k = 1, \ldots, q$.

Demonstração: Consideremos a expansão de Taylor:

$$\boldsymbol{h}(\boldsymbol{\theta} + n^{-\frac{1}{2}}\ \boldsymbol{u}) = \boldsymbol{h}(\boldsymbol{\theta}) + n^{-\frac{1}{2}}\ \boldsymbol{u}\boldsymbol{H}(\boldsymbol{\theta}) + n^{-\frac{1}{2}}\ \boldsymbol{u}\{\boldsymbol{H}(\boldsymbol{\theta}^*) - \boldsymbol{H}(\boldsymbol{\theta})\} \tag{4.3.1}$$

onde $\boldsymbol{\theta}^* = \boldsymbol{\theta} + n^{-\frac{1}{2}}\ \gamma\ \boldsymbol{u}$, $0 \leq \gamma \leq 1$. Como $\boldsymbol{H}(\boldsymbol{\theta})$ é contínua, podemos re-escrever (4.3.1) na forma:

$$\sqrt{n}\{\boldsymbol{h}(\boldsymbol{\theta} + n^{-\frac{1}{2}}\ \boldsymbol{u}) - \boldsymbol{h}(\boldsymbol{\theta})\} = \boldsymbol{u}\boldsymbol{H}(\boldsymbol{\theta}) = o(1) \tag{4.3.2}$$

Fazendo $\boldsymbol{u} = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ obtemos $\boldsymbol{h}(\boldsymbol{\theta} + n^{-\frac{1}{2}}\ \boldsymbol{u}) = \boldsymbol{h}(\hat{\boldsymbol{\theta}}) = \boldsymbol{m}_n$. Então:

$$\sqrt{n}\{\boldsymbol{m}_n - \boldsymbol{h}(\boldsymbol{\theta})\} = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})\ \boldsymbol{H}(\boldsymbol{\theta}) + o_p(1) \tag{4.3.3}$$

Seja $\boldsymbol{\lambda} \in \mathbb{R}^q$, $\boldsymbol{\lambda} \neq 0$, um vetor arbitrário, porém fixo; logo:

$$\sqrt{n}\boldsymbol{\lambda}\{\boldsymbol{m}_n - \boldsymbol{h}(\boldsymbol{\theta})\}^t = \sqrt{n}\{(\lambda_1 m_{1n}' + \cdots + \lambda_q m_{qn}') - (\lambda_1 \mu_1' + \cdots + \lambda_q \mu_q')\} =$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\sum_{j=1}^{q} \lambda_j (X_i^j - EX_i^j)\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i$$

onde $U_i = \sum_{i=1}^{n} \lambda_j (X_i^j - EX_i^j)$ é tal que $EU_i = 0$ e

$$EU_i^2 = \sum_{j=1}^{q} \sum_{k=1}^{q} \lambda_j \lambda_k E(X_i^j - \mu'_j)(X_i^k - \mu'_k) = \sum_{j=1}^{q} \sum_{k=1}^{q} \lambda_j \lambda_k (\mu_{j+k} - \mu'_j \mu'_k) = \boldsymbol{\lambda} \boldsymbol{\Sigma} \boldsymbol{\lambda}^t < \infty.$$

Como os $U_i's$ são variáveis aleatórias independentes e identicamente distribuídas, segue pelo Teorema Limite Central (Teorema 3.2.1) que

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} U_i = \sqrt{n} \boldsymbol{\lambda} \{\boldsymbol{m}_n - \boldsymbol{h}(\boldsymbol{\theta})\}^t \xrightarrow{\mathcal{D}} \mathcal{N}(0, \boldsymbol{\lambda} \boldsymbol{\Sigma} \boldsymbol{\lambda}^t);$$

então, pelo Teorema 2.3.13 podemos concluir que

$$\sqrt{n}\{\boldsymbol{m}_n - \boldsymbol{h}(\boldsymbol{\theta})\} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

Finalmente, de (4.3.3), podemos escrever :

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = \sqrt{n}\{\boldsymbol{m}_n - \boldsymbol{h}(\boldsymbol{\theta})\} \boldsymbol{H}^{-1}(\boldsymbol{\theta}) + \boldsymbol{\varsigma}_n$$

onde $\boldsymbol{\varsigma}_n = o_p(1)$ e utilizando a versão multivariada do Teorema de Slutsky (Exercício 2.3.10) concluímos que :

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\boldsymbol{0}, [\boldsymbol{H}^{-1}(\boldsymbol{\theta})]^t \ \boldsymbol{\Sigma} \ [\boldsymbol{H}^{-1}(\boldsymbol{\theta})]\big)$$

Exemplo 4.3.1 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição $\mathcal{N}(\mu, \sigma^2)$; então*

$$\boldsymbol{m}_n = \frac{1}{n}(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2), \qquad \boldsymbol{h}(\boldsymbol{\theta}) = (\mu, \mu^2 + \sigma^2)$$

e

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 0 \\ 2\mu & 1 \end{bmatrix} ;$$

como $\mu'_3 = EX_1^3 = \mu^3 + 3\mu\sigma^2$ e $\mu'_4 = EX_1^4 = \mu^4 + 3\sigma^4 + 6\mu^2\sigma^2$, uma aplicação direta do Teorema 4.3.1 nos permite concluir que

$$\frac{1}{\sqrt{n}}(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2) \xrightarrow{\mathcal{D}} \mathcal{N}\big((\mu, \mu^2 + \sigma^2), \boldsymbol{\Gamma}\big)$$

onde

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mu^2 + \sigma^2 & \mu\sigma^2 - \mu^3 \\ \mu\sigma^2 - \mu^3 & \mu^4 + 3\sigma^4 - 2\mu^2\sigma^2 \end{bmatrix}$$

Convém observar que o método descrito acima pode ser utilizado para a obtenção da distribuição assintótica de $\boldsymbol{T}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(X_i)$ onde $\boldsymbol{g}(\boldsymbol{x}) = [g_1(\boldsymbol{x}), \dots g_q(\boldsymbol{x})]$ é uma função com valores em $\mathbb{R}^q$, se fizermos $\boldsymbol{h}(\boldsymbol{\theta}) = E\boldsymbol{T}_n$.

Exemplo 4.3.2 *Sejam $X_1, \dots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição $\mathcal{N}(\mu, \sigma^2)$; sejam também*

$$\boldsymbol{T}_n = \frac{1}{n}\{\sum_{i=1}^{n} X_i, \frac{n}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\} \quad e \quad \boldsymbol{h}(\boldsymbol{\theta}) = E\boldsymbol{T}_n = (\mu, \sigma^2).$$

Então $\boldsymbol{H}(\boldsymbol{\theta}) = \boldsymbol{I}_2$ e pelo Teorema 4.3.1 temos

$$\sqrt{n}(\overline{X}_n, S_n^2) \xrightarrow{\mathcal{D}} \mathcal{N}\big((\mu, \sigma^2), \boldsymbol{\Gamma}\big) \quad onde \quad \boldsymbol{\Gamma} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$$

Muitos métodos de estimação (de um parâmetro θ) podem ser formulados em termos da minimização (com respeito a θ) de expressões do tipo $\sum_{i=1}^{n} \rho(X_i; \theta)$ onde ρ é uma função arbitrária. Quando $\Psi(x; \theta) = \frac{\partial}{\partial\theta}\rho(x; \theta)$ existe, esses procedimentos são equivalentes à obtenção de soluções de equações implícitas da forma :

$$\sum_{i=1}^{n} \Psi(X_i; \theta) = 0 \tag{4.3.4}$$

Em geral, este é o caso para os estimadores de **MV**, onde

$$\rho(x; \theta) = -\log f(x; \theta) \quad \text{e} \quad \Psi(x; \theta) = -\frac{\partial}{\partial\theta} f(x; \theta)/f(x; \theta)$$

com $f(x; \theta)$ denotando a função densidade correspondente, ou para os estimadores de Mínimos Quadrados (**MQ**), onde $\rho(x; \theta) = \{x - g(\theta)\}^2$ e $\Psi(x; \theta) = -2\{x - g(\theta)\}g'(\theta)$, com g e g' denotando uma função conveniente e sua derivada, respectivamente.

Estimadores definidos através de equações implícitas da forma (4.3.4) são chamados *estimadores M* (do tipo Máxima Verossimilhança) e têm interesse especial para estimação de *parâmetros de locação*, onde $\Psi(x; \theta) = \Psi(x - \theta)$, pois neste caso eles podem ser facilmente definidos de forma a produzir estimadores *robustos*, i.e., estimadores que tem propriedades estatísticas estáveis sob especificações bastante gerais para a distribuição geradora dos dados. Esse é o caso dos estimadores M definidos através da *função de Huber*:

$$\Psi(t) = \begin{cases} t & \text{se } |t| \leq k \\ k \text{ sinal } (t) & \text{se } |t| > k \end{cases}$$

onde k é uma constante positiva; essa função corresponde a

$$\rho(t) = \begin{cases} t^2/2 & \text{se } |t| \leq k \\ k|t| & \text{se } |t| > k. \end{cases}$$

Em geral, equações da forma (4.3.4) podem ter muitas raízes e elas precisam ser examinadas para verificarmos qual delas corresponde ao mínimo de $\sum_{i=1}^{n} \rho(X_i;\theta)$. Em muitos problemas de locação, contudo, a *função de estimação* $M_n(\theta) = \sum_{i=1}^{n} \Psi(X_i - \theta)$ é não-crescente em θ, e mesmo na presença de raízes múltiplas, o estimador M correspondente pode ser facilmente definido, por exemplo, como qualquer valor $\hat{\theta}_n \in [\hat{\theta}_{n_1}, \hat{\theta}_{n_2}]$ onde $\hat{\theta}_{n_1} = \sup\{\theta : M_n(\theta) > 0\}$ e $\hat{\theta}_{n_2} = \inf\{\theta : M_n(\theta) < 0\}$. Por razões técnicas nos restringiremos ao caso onde $M_n(\theta)$ tem uma única raiz $\hat{\theta}_n$. O leitor poderá consultar Huber (1980) para detalhes sobre situações mais gerais.

O teorema seguinte ilustra uma técnica bastante útil para demonstrar a normalidade assintótica se estimadores M.

Teorema 4.3.2 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função distribuição $F(x;\theta)$ e seja $\Psi(x-\theta)$ uma função não-crescente em θ. Vamos supor que θ_0 seja uma raiz isolada de $M(\theta) = \int \Psi(x-\theta)dF(x) = 0$ e que as seguintes condições estejam satisfeitas :*

i) a derivada $M'(\theta) = \frac{\partial}{\partial\theta}M(\theta)$ existe numa vizinhança de θ_0 e é tal que $M'(\theta_0) \neq 0$;

ii) $\int \Psi^2(x-\theta)dF(x) < \infty$ e é contínua no ponto θ_0.

Então, se $\hat{\theta}_n$ for uma solução de $M_n(\theta) = \sum_{i=1}^{n} \Psi(X_i - \theta) = 0$, temos :

a) $\hat{\theta}_n$ é consistente para θ_0;

b) $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \sigma^2(\theta_0)/[M'(\theta_0)]^2\}$ onde $\sigma^2(\theta_0) = Var\Psi(X_1 - \theta_0)$.

Demonstração: Como $\Psi(x-\theta)$ é não-crescente em θ, então $M(\theta)$ e $M_n(\theta)$ também é ; logo, dado $\epsilon > 0$, temos:

$$\begin{aligned}
\mathbb{P}\{\hat{\theta}_n - \theta_0 > \epsilon\} &= \mathbb{P}\{M_n(\theta_0+\epsilon) > 0\} = \mathbb{P}\{\frac{1}{n}\sum_{i=1}^{n} \Psi(X_i - \theta_0 - \epsilon) > 0\} \\
&= \mathbb{P}\{\frac{1}{n}\sum_{i=1}^{n} [\Psi(X_i - \theta_0 - \epsilon) - E\Psi(X_i - \theta_0 - \epsilon)] > -E\Psi(X_1 - \theta_0 - \epsilon)\} \\
&= \mathbb{P}\{\frac{1}{n}\sum_{i=1}^{n} Z_i > -E\Psi(X_1 - \theta_0 - \epsilon)\} \qquad (4.3.5)
\end{aligned}$$

onde $Z_i = \Psi(X_i - \theta_0 - \epsilon) - E\Psi(X_1 - \theta_0 - \epsilon),\ i = 1, \ldots, n$ são variáveis aleatórias independentes e identicamente distribuídas com médias zero. Como $M(\theta)$ é não-crescente, derivável numa vizinhança de θ_0 e $M(\theta_0) = 0$, então existe $\epsilon' > 0$ tal que $M(\theta_0 + \epsilon') < 0$. Logo:

$$\mathbb{P}\{\hat{\theta}_n - \theta_0 > \epsilon'\} = \mathbb{P}\{\frac{1}{n}\sum_{i=1}^{n} Z_i > -M(\theta_0 + \epsilon')\} \longrightarrow 0 \quad \text{com} \quad n \longrightarrow \infty$$

pela Lei Fraca dos Grande Números de Khintchine (Teorema `2.4.2`). Logo para todo $\epsilon > \epsilon' > 0$, temos $\mathbb{P}\{\hat{\theta}_n - \theta_0 > \epsilon\} \longrightarrow 0$ com $n \longrightarrow \infty$. Repetindo o raciocínio para $\mathbb{P}\{\hat{\theta}_n - \theta_0 < \epsilon\}$ podemos mostrar que $\mathbb{P}\{|\hat{\theta}_n - \theta_0| > \epsilon\} \longrightarrow 0$ com $n \longrightarrow \infty$ e o resultado `a`) segue.

Uma demonstração do resultado `b`) pode ser encontrada em Serfling (1980, cap 7), por exemplo.

Exemplo 4.3.3 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função distribuição F e seja $\Psi(x) = x$; então a solução de (4.3.4) define o estimador de Mínimos Quadrados (**MQ**) $\hat{\theta}_n = \overline{X}_n$.*

a) Se F corresponder à distribuição $\mathcal{N}(\theta, \sigma^2)$, então claramente $M'(\theta) = -1$ e $\int \Psi^2(x - \theta)dF(x) = \sigma^2$ e as condições `i`) e `ii`) do Teorema 4.3.2 estão satisfeitas; conseqüentemente, $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$.

b) Se F corresponder à distribuição Pearson tipo VII com parâmetros $m = (\nu + 1)/2,\ c = \sqrt{\nu}$ e θ, i. e. uma distribuição t de Student com $\nu = 3$ graus de liberdade, deslocada de θ, temos

$$M'(\theta) = \frac{\partial}{\partial \theta}\int \Psi(x - \theta)f(x)dx =$$

$$\frac{\partial}{\partial \theta}\int (x - \theta)\frac{\Gamma\{(\nu + 1)/2\}}{\sqrt{\pi\nu}\ \Gamma(\nu/2)}\ \{1 + \frac{(x - \theta)^2}{\nu}\ \}^{-(\nu+1)/2}dx = -1$$

e $\int \Psi^2(x - \theta)f(x)dx = \nu/(\nu - 2)$; então pelo Teorema 4.3.2 podemos concluir que $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \nu/(\nu - 2)\big)$.

c) Se F for como em `b`), porém com $\nu = 2$ não podemos aplicar o Teorema 4.3.2 dado que a condição `ii`) não está satisfeita; em tal caso um estimador robusto para θ pode ser considerado.

Exemplo 4.3.4 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com função densidade f simétrica com respeito a θ_0 e seja Ψ dada por (4.3). Então temos:*

$$M(\theta) = \int_{\theta-k}^{\theta+k} (x - \theta)f(x)dx + k\int_{\theta+k}^{\infty} f(x)dx - k\int_{-\infty}^{\theta-k} f(x)dx =$$

$$\int_{-k}^{k} xf(x+\theta)dx + k\big[\mathbb{P}\{X_1 \geq \theta + k\} - \mathbb{P}\{X_1 \leq \theta - k\}\big].$$

Claramente $M(\theta_0) = 0$ *e* $M'(\theta_0) = \int_{\theta_0-k}^{\theta_0+k} f(x)dx = \mathbb{P}\{\theta_0 - k \leq X_1 \leq \theta_0 + k\}$.

Como Ψ é limitada, $\sigma^2(\theta_0) = \int_{\theta_0-k}^{\theta_0+k} (x-\theta_0)^2 f(x)dx + 2k^2\mathbb{P}\{X_1 \geq \theta_0 + k\} < \infty$ independentemente de EX_1^2 ser finita ou não. Então a normalidade assintótica do correspondente estimador $M, \hat{\theta}_n$ segue, do Teorema 4.3.2.

4.4 Eficiência Assintótica de Estimadores

Como vimos nas seções anteriores, muitos tipos de estimadores usualmente empregados em análises estatísticas são consistentes e têm distribuição assintótica Normal. Então, para escolher um estimador dentre um conjunto de candidatos precisamos de algum critério adicional. Consideremos, por exemplo, seqüências $\{\hat{\theta}_n^{(1)}\}_{n\geq 1}$ e $\{\hat{\theta}_n^{(2)}\}_{n\geq 1}$ de estimadores tais que $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2)$ e $\sqrt{n}(\hat{\theta}_n^{(2)} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_2^2)$. Podemos utilizar as variâncias assintóticas σ_1^2 e σ_2^2 para definir a *eficiência assintótica relativa* de $\hat{\theta}_n^{(2)}$ com respeito a $\hat{\theta}_n^{(1)}$ como :

$$\text{EAR}(\ \hat{\theta}_n^{(2)} \mid \hat{\theta}_n^{(1)}\ ) = \sigma_1^2/\sigma_2^2$$

Obviamente, se $\text{EAR}(\hat{\theta}_n^{(2)}|\hat{\theta}_n^{(1)}) < 1$, $\hat{\theta}_n^{(1)}$ é mais eficiente que $\hat{\theta}_n^{(2)}$.

Exemplo 4.4.1 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com distribuição de Pareto, i.e., com função densidade*

$$f(x;\theta) = \theta^{-1}\ x^{-(1+\frac{1}{\theta})},\ x > 1,\ 0 < \theta < \frac{1}{2}.$$

Fazendo $\gamma = \theta^{-1}$, o que implica $\gamma > 2$ temos:

$$\mu_1' = EX_1 = \int_1^\infty x\ \gamma\ x^{-1-\gamma}dx = \gamma\int_1^\infty x^{-\gamma}dx = \frac{\gamma}{\gamma-1} = \frac{1}{1-\theta}$$

*Então, fazendo $\overline{X}_n = (1-\theta)^{-1}$ obtemos o estimador de **MM** de θ, $\overline{\theta}_n = 1 - 1/\overline{X}_n$. Para $h(\theta) = (1-\theta)^{-1}$, temos $H(\theta) = \frac{d}{d\theta}h(\theta) = (1-\theta)^2$. Além disso, observemos que:*

$$\mu_2' = EX_1^2 = \int_1^\infty x^2\ \gamma\ x^{-1-\gamma}dx = \frac{\gamma}{2-\gamma}\int_1^\infty (2-\gamma)x^{1-\gamma}dx = \frac{\gamma}{\gamma-2} = \frac{1}{1-2\theta}$$

Então, $VarX_1 = \sum = (1-2\theta)^{-1} - (1-\theta)^{-2} = \theta^2(1-2\theta)^{-1}(1-\theta)^{-2}$; logo, $\{H(\theta)\}^{-2}\sum = \theta^2(1-\theta)^2(1-2\theta)^{-1}$ e pelo Teorema 4.3.1 temos que

$$\sqrt{n}(\overline{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \theta^2(1-\theta)^2(1-2\theta)^{-1}\big).$$

*Por outro lado, o estimador de **MV** de θ é uma solução $\hat{\theta}_n$ de $\frac{\partial}{\partial\theta}\log L_n(\theta) = -n/\theta + \theta^{-2}\sum_{i=1}^n \log X_i = 0$ o que implica $\hat{\theta}_n = n^{-1}\sum_{i=1}^n \log X_i$. Agora observemos que:*

i) $\left|\frac{\partial}{\partial\theta}f(x;\theta)\right| = \left|-\theta^{-2}x^{-(1+\frac{1}{\theta})} + \theta^{-3}x^{-(1+\frac{1}{\theta})}\log x\right| < \theta^{-2}x^{-(1+\frac{1}{\theta})} + \theta^{-3}x^{-(1+\frac{1}{\theta})} = H_1(x)$ *que é integrável;*

ii) $\left|\frac{\partial^2}{\partial\theta^2} f(x;\theta)\right| = \left|-2\theta^{-3}x^{-(1+\frac{1}{\theta})} - 2\theta^{-4}x^{-(1+\frac{1}{\theta})}\log x + \theta^{-5}x^{-(1+\frac{1}{\theta})}\log^2 x\right| <$ $2\theta^{-3}x^{-(1+\frac{1}{\theta})}\{1 + \theta^{-1}\log x + \frac{1}{2}\theta^{-2}\log^2 x\} = H_2(x)$ *que é integrável;*

iii) $E\log X_1 = \theta$ *e* $E\log^2 X_1 = 2\theta^2$ *o que implica que*

$$I_\theta = E\{\frac{\partial}{\partial\theta}\log f(X_1;\theta)\}^2 = E\{\theta^{-2} - 2\theta^{-3}\log X_1 + \theta^{-4}\log^2 X_1\} = \theta^{-2} < \infty;$$

iv) $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = \theta^{-2} - 2\theta^{-3}\log x,$ *o que implica que :*

$$\Psi_\delta = E\{\sup_{\{h:|h|\le\delta\}}\left|(\theta+h)^{-2} - \theta^{-2} - 2(\theta+h)^{-3}\log X_1 + 2\theta^{-3}\log X_1\right|\} \le$$

$$\sup_{\{h:|h|\le\delta\}}\left|(\theta+h)^{-2} - \theta^{-2}\right| + 2E\log X_1 \sup_{\{h:|h|\le\delta\}}\left|(\theta+h)^{-3} - \theta^{-3}\right| =$$

$$\left|(\theta+\delta)^{-2} - \theta^{-2}\right| + \left|(\theta+\delta)^{-3} - \theta^{-3}\right|2\theta \longrightarrow 0 \quad com \quad \delta \longrightarrow 0$$

Então, pelo Teorema 4.2.1 temos $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$. *Conseqüentemente,* $EAR(\ \overline{\theta}_n \mid \hat{\theta}_n\ ) = (1-2\theta)(1-\theta)^{-2} < 1$, *indicando que o estimador de* ***MM*** *é assintoticamente menos eficiente que o estimador de* ***MV****.*

Sejam $X_1,\ldots,X_n$ variáveis aleatórias independentes e identicamente distribuídas com função densidade $f(x;\theta)$, $\theta \in \Theta \subset \mathbb{R}$ tal que $E\{\frac{\partial}{\partial\theta}\log f(X_1;\theta)\}^2 = I_\theta < \infty$ e consideremos uma seqüência $\{\hat{\theta}_n\}$ de estimadores não viciados de θ tal que $\mathrm{Var}\hat{\theta}_n < \infty$. Então notemos que, pela desigualdade de Cauchy-Schwarz :

$$E^2\{\sqrt{n}(\hat{\theta}_n - \theta)\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta)\} \le E\{\sqrt{n}(\hat{\theta}_n - \theta)\}^2\ E\{\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta)\}^2 \tag{4.4.1}$$

Como $E\{\sqrt{n}(\hat{\theta}_n - \theta)\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta)\} = 1$ e $E\{\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta)\}^2 = I_\theta$ segue que :

$$E\{\sqrt{n}(\hat{\theta}_n - \theta)\}^2 = n\mathrm{Var}\hat{\theta}_n \ge I_\theta^{-1} \tag{4.4.2}$$

que é conhecida como a *desigualdade de Fréchet - Cramér - Rao*. O limite I_θ^{-1} para a variância de todos os estimadores não viciados de θ segere a definição de *eficiência assintótica* de uma seqüência $\{\hat{\theta}_n\}_{n\ge 1}$ tal que $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma_\theta^2)$ como o quociente $I_\theta^{-1}/\sigma_\theta^2$. Se $\sigma_\theta^2 = I_\theta^{-1}$, o estimador correspondente é dito *assintoticamente eficiente.* Além disso, convém notar que sob as condições do Teorema 4.2.1 os estimadores **MV** são assintoticamente eficientes. De uma forma mais ampla, se observarmos que a igualdade em (4.4.1) ocorre se e somente se $\sqrt{n}(\hat{\theta}_n - \theta) = K\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta)$, com $K \ne 0$, segue que todos os estimadores $\hat{\theta}_n$ tais que $\sqrt{n}(\hat{\theta}_n - \theta) = K\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\log L_n(\theta) + o_p(1)$ são assintoticamente eficientes. Estimadores pertencentes a esta classe são chamados de *estimadores BAN* (Best Asymptotically Normal).

Exemplo 4.4.2 *Sejam* $X_1, \ldots, X_n$ *variáveis aleatórias independentes e identicamente distribuídas com distribuição* $\mathcal{N}(\mu, \sigma^2)$. *Como as suposições do Teorema* `4.2.1` *estão claramente satisfeitas, segue que o estimador de* $\boldsymbol{MV}$ *de* σ^2, $\sigma_n^2 = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ *é tal que* $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\sigma^4)$; *como* $S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = \hat{\sigma}_n^2 - \frac{\hat{\sigma}_n^2}{n} = \hat{\sigma}_n^2 + o_p(1)$ *podemos concluir que* S_n^2 *também é um estimador* BAN.

Esses resultados podem ser facilmente estendidos para o caso multiparamétrico, onde a desigualdade (`4.4.2`) é substituída por:

$$E\{n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t \ (\boldsymbol{\theta}_n - \boldsymbol{\theta})\} - \boldsymbol{I}_{\boldsymbol{\theta}^{-1}} = n\mathbf{Var}\hat{\boldsymbol{\theta}}_n - \boldsymbol{I}_{\boldsymbol{\theta}^{-1}} \text{ é não negativa definida.} \qquad (4.4.3)$$

4.5 Comportamento Assintótico de Estatísticas de Teste

Sejam $X_1, \ldots, X_n$ variáveis aleatórias identicamente distribuídas com função densidade $f(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ e vamos admitir que as suposições apresentadas no Teorema `4.2.2` estejam satisfeitas. Consideremos então o problema de testar a hipótese (simples)

$$\mathcal{H} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$$

onde $\boldsymbol{\theta_0}$ é um valor fixado, contra a alternativa $\mathcal{A} : \boldsymbol{\theta} \in \boldsymbol{\theta}^0 = \boldsymbol{\theta} - \{\boldsymbol{\theta}_0\}$. Essencialmente três estatísticas de testes são usualmente empregadas para testar $\mathcal{H}$:

i) a estatística de Wald: $Q_w = n \ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \ \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t$, onde $\hat{\boldsymbol{\theta}}_n$ é o estimador de **MV** (ou outro estimador BAN) de $\boldsymbol{\theta}_0$ e $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \ \partial \boldsymbol{\theta}^t} \log f(x_i; \boldsymbol{\theta})|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n}$;

ii) a estatística de Wilks (ou da Razão de Verossimilhança): $Q_v = -2 \ \log \Lambda_n = 2\{\log L_n(\hat{\boldsymbol{\theta}}_n) - \log L_n(\boldsymbol{\theta}_0)\}$ onde $\Lambda_n = L_n(\boldsymbol{\theta}_0) / \sup_{\boldsymbol{\theta} \in \Theta^\circ} L_n(\boldsymbol{\theta})$ é a razão de verossimilhança;

iii) a estatística de Rao (ou dos escores eficientes):

$$Q_r = n \ \boldsymbol{U}_n(\boldsymbol{x}; \boldsymbol{\theta}_0)^t \ \ \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1} \ \boldsymbol{U}_n(\boldsymbol{x}; \boldsymbol{\theta}_0)$$

onde

$$\boldsymbol{U}_n(\boldsymbol{X}; \boldsymbol{\theta}_0) = \ \frac{1}{n} \ \sum_{i=1}^{n} \ \frac{\partial}{\partial \boldsymbol{\theta}} \ \log f(X_i; \boldsymbol{\theta})|_{\boldsymbol{\theta} = \boldsymbol{\theta}_0};$$

note que esta estatística não exige o cálculo do estimador de **MV**, $\hat{\boldsymbol{\theta}}_n$.

Demonstraremos a seguir que sob a hipótese nula $\mathcal{H}$ as três estatísticas acima tem distribuição assintótica $\chi^2_{(q)}$. Nesse sentido, observemos que:

i) Pelo Teorema 4.2.1, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1})$, o que implica que

$$Q_w^* = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \boldsymbol{I}_{\boldsymbol{\theta}_0}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \chi^2_{(q)};$$

então, fazendo $\boldsymbol{x} = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ temos: $\frac{Q_w}{Q_w^*} = \boldsymbol{x}^t \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{x} / \boldsymbol{x}^t \boldsymbol{I}_{\boldsymbol{\theta}_0} \boldsymbol{x}$ e utilizando o Teorema A.10 do Apêndice A segue que:

$$\mathrm{ch}_q(\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1}) \leq \frac{Q_w}{Q_w^*} \leq \mathrm{ch}_1(\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1})$$

onde $\mathrm{ch}_q(\boldsymbol{M})$ e $\mathrm{ch}_1(\boldsymbol{M})$ denotam respectivamente a menor e a maior raiz característica de uma matriz $\boldsymbol{M}(q \times q)$. Agora, como pela Lei Fraca dos Grandes Números de Khintchine (Teorema 2.4.2) $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \xrightarrow{\mathbb{P}} \boldsymbol{I}_{\boldsymbol{\theta}_0}$, temos: $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{I}_q$ e conseqüentemente, $\mathrm{ch}_1(\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1}) \xrightarrow{\mathbb{P}} 1$ e $\mathrm{ch}_q(\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_n} \boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1}) \xrightarrow{\mathbb{P}} 1$. Logo podemos concluir que $Q_w/Q_w^* \xrightarrow{\mathbb{P}} 1$. Então, escrevendo $Q_w = Q_w^* \; Q_w/Q_w^*$ e utilizando o Teorema de Slutsky (Teorema 2.3.16) temos $Q_w \xrightarrow{\mathcal{D}} \chi^2_{(q)}$.

ii) Consideremos agora a seguinte expansão de Taylor:

$$\log L_n(\hat{\boldsymbol{\theta}}_n) =$$
$$\log L_n(\boldsymbol{\theta}_0) + \sqrt{n}\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \Big\{ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \Big\} +$$
$$\frac{n}{2}\ (\ \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\ ) \Big\{ \frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \log L_n\ (\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \Big\}\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{4.5.1}$$

onde $\boldsymbol{\theta}_n^*$ pertence ao segmento que une $\boldsymbol{\theta}_0$ e $\hat{\boldsymbol{\theta}}_n$. Em seguida, consideremos uma expansão de Taylor do primeiro termo entre { } na expressão (4.5.1):

$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} =$$
$$\frac{1}{\sqrt{n}} \frac{\partial}{\partial \boldsymbol{\theta}} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} - \frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0^{**}} \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) =$$
$$-\frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^{**}} \sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) \tag{4.5.2}$$

onde $\boldsymbol{\theta}_n^{**}$ pertence ao segmento que une $\boldsymbol{\theta}_0$ e $\hat{\boldsymbol{\theta}}_n$. Substituindo (4.5.2) em (4.5.1), obtemos:

$$\log L_n\ (\hat{\boldsymbol{\theta}}_n) = \log L_n(\boldsymbol{\theta}_0) - n\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t \Big\{ \frac{1}{n} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \log L_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^{**}} \Big\}\ (\ \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) +$$

$$\frac{n}{2}\ (\ \hat{\boldsymbol{\theta}}_n\ -\ \boldsymbol{\theta}_0\ )^t\ \{\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\ \log L_n(\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n^*}\}\ (\hat{\boldsymbol{\theta}}_n\ -\ \boldsymbol{\theta}_0\ ) \tag{4.5.3}$$

Como, pela Lei Fraca dos Grandes Números de Khintchine (Teorema 2.4.2), os dois termos entre { } na expressão (4.5.3) convergem em probabilidade (quase certamente) para $-\boldsymbol{I}_{\boldsymbol{\theta}_0}$, temos:

$$Q_v = 2\{\log L_n(\hat{\boldsymbol{\theta}}_n) - \log L_n(\boldsymbol{\theta}_0)\} = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^t\ \boldsymbol{I}_{\boldsymbol{\theta}_0}\ (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1) = Q_w^* + o_p(1)$$

e conseqüentemente, pelo Teorema de Slutsky (Teorema 2.3.16) segue que $Q_v \xrightarrow{\mathcal{D}} \chi^2_{(q)}$.

iii) Pelo Teorema Limite Central (Teorema 3.2.1) é claro que

$$\sqrt{n}\boldsymbol{U}_n(\boldsymbol{X};\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{\boldsymbol{\theta}});$$

conseqüentemente, $Q_R \xrightarrow{\mathcal{D}} \chi^2_{(q)}$.

Resultados semelhantes podem ser obtidos para testes de hipóteses compostas do tipo $\mathcal{H} : \boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}$. O leitor poderá consultar Rao (1973 , cap 6) ou Serfling (1980, cap 4) para maiores detalhes.

Apêndice A

Teoremas

Teorema A.1 (Desigualdade de Chebychev) *Sejam X uma variável aleatória e α um número real positivo tais que $E\big(\ |X|^\alpha\ \big) < \infty$. Então, para quaisquer números reais $\epsilon > 0$ e c,*

$$\mathbb{P}\big(\ |X - c| \geq \epsilon\ \big) \leq \frac{E\big(|X - c|^\alpha\big)}{\epsilon^\alpha}.$$

ß*Demonstração*: Seja F a função distribuição de X. Então,

$$E\big(|X - c|^\alpha\big) = \int_{-\infty}^{\infty} |x - c|^\alpha dF(x) \geq$$

$$\int_{|x-c|\geq\epsilon} |x - c|^\alpha dF(x) \geq \epsilon^\alpha \int_{|x-c|\geq\epsilon} dF(x) = \epsilon^\alpha\ \mathbb{P}\big(\ |X - c| \geq \epsilon\ \big),$$

o que prova o teorema.

Teorema A.2 *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias. Então, para todo número real $\epsilon > 0$,*

$$\mathbb{P}\big(\ |\sum_{j=1}^{n} X_j| \geq \epsilon\ \big) \leq \sum_{j=1}^{n} \mathbb{P}\big(\ |X_j| \geq \frac{\epsilon}{n}\ \big).$$

Demonstração: O resultado segue do fato de que para todo $\epsilon > 0$,

$$(\ |\sum_{j=1}^{n} X_j| \geq \epsilon\ ) \subseteq (\ \sum_{j=1}^{n} |X_j| \geq \epsilon\ ) \subseteq \bigcup_{j=1}^{n} (\ |X_j| \geq \frac{\epsilon}{n}\ )$$

Teorema A.3 (Lema de Borel - Cantelli) *Sejam $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\{A_n\}_{n\geq 1}$ uma seqüência de eventos de $\mathcal{F}$;*

i) se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, *então* $\mathbb{P}\big(A_n \text{ infinitas vezes }\big) = 0$;

ii) se $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, *e os* A_n *forem eventos independentes, então* $\mathbb{P}\big(A_n \text{ infinitas vezes }\big) = 1$.

Demonstração: Ver Breimann (1968).

Teorema A.4 (Teorema de Helly-Bray) *Sejam* $F, F_1, F_2, \ldots,$ *funções distribuição. Se, para todo ponto* x *de continuidade de* $F, F_n(x) \longrightarrow F(x)$ *quando* $n \longrightarrow \infty$, *então, para toda função contínua e limitada* $f : \mathbb{R} \longrightarrow \mathbb{R}$,

$$\int f(x)dF_n(x) \longrightarrow \int f(x)dF(x) \text{ quando } n \longrightarrow \infty$$

Demonstração: Ver Gnedenko (1969), por exemplo.

Teorema A.5 (Teorema da Continuidade de Paul Lévy) *Sejam* $F_1, F_2, \ldots$ *funções distribuição e* $\varphi_1, \varphi_2, \ldots$ *suas respectivas funções características, isto é ,* $\varphi_n(x) = \int e^{itx} dF_n(x)$ *para* $n = 1, 2, \ldots$ *. Se a seqüência* $\{\varphi_n\}_{n\geq 1}$ *convergir para uma função* φ *e* φ *for contínua no ponto zero, então existe uma função distribuição* F *tal que, para todo ponto* x *de continuidade de* $F, F_n(x) \longrightarrow F(x)$ *quando* $n \longrightarrow \infty$ *e* φ *é a função característica de* F.

Demonstração: Ver Gnedenko (1969), por exemplo.

Teorema A.6 *Sejam* $X, X_1, \ldots$ *variáveis aleatórias. Se* $X_n \xrightarrow{\mathbb{P}} X$, *então existe uma subseqüência* $\{X_{n_k}\}_{k\geq 1}$ *de* $\{X_n\}_{n\geq 1}$ *tal que* $X_{n_k} \xrightarrow{q.c.} X$.

Demonstração: Ver Bartle (1966), por exemplo.

Teorema A.7 *Seja* $X_1, X_2, \ldots$ *uma seqüência de variáveis aleatórias independentes com médias* $\mu_1, \mu_2, \ldots$ *e variâncias* $\sigma_1^2, \sigma_2^2, \ldots,$ *respectivamente. Se* $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$ *e* $\sum_{n=1}^{\infty} \mu_n$ *convergir, então* $\sum_{n=1}^{\infty} X_n$ *converge quase certamente para uma variável aleatória* X. *Além disso,* $E(X) = \sum_{n=1}^{\infty} \mu_n$ *e* $Var(X) = \sum_{n=1}^{\infty} \sigma_n^2$.

Demonstração: Ver Rao (1984), por exemplo.

Teorema A.8 (Lema de Kronecker) *Seja* $\{a_n\}_{n\geq 1}$ *uma seqüência de números reais tais que* $\sum_{n=1}^{\infty} (a_n/n)$ *converge. Então,*

$$\frac{1}{n} \sum_{i=1}^{n} a_i \longrightarrow 0 \quad \text{quando} \quad n \longrightarrow \infty$$

Demonstração: Ver Rao (1984), por exemplo.

Teorema A.9 (Desigualdade de Jensen) *Seja* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *uma função convexa. Se uma variável aleatória* X *é tal que* $E(|X|) < \infty$, *então,* $E\big(f(X)\big) \geq f\big(E(X)\big)$.

Demonstração: Ver Breiman (1968) por exemplo.

Teorema A.10 *Para todo* $\alpha > 0$,

$$\frac{1}{n^{\alpha+1}} \sum_{j=1}^{n} j^{\alpha} \longrightarrow \frac{1}{\alpha+1} \quad \textit{quando } n \longrightarrow \infty.$$

Demonstração: Para $j = 1, 2, \ldots, n$, temos :

$$\int_{j-1}^{j} x^{\alpha} dx \leq \int_{j-1}^{j} j^{\alpha} dx = j^{\alpha} = \int_{j}^{j+1} j^{\alpha} dx \leq \int_{j}^{j+1} x^{\alpha} dx$$

Então,

$$\int_{0}^{n} x^{\alpha} dx = \sum_{j=1}^{n} \int_{j-1}^{j} x^{\alpha} dx \leq \sum_{j=1}^{n} j^{\alpha} \leq \sum_{j=1}^{n} \int_{j}^{j+1} x^{\alpha} dx = \int_{1}^{n+1} x^{\alpha} dx$$

ou seja,

$$\frac{n^{\alpha+1}}{\alpha+1} \leq \sum_{j=1}^{n} j^{\alpha} \leq \frac{(n+1)^{\alpha+1} - 1}{\alpha+1}$$

Logo,

$$\frac{1}{\alpha+1} \leq \frac{1}{n^{\alpha+1}} \sum_{j=1}^{n} j^{\alpha} < \frac{1}{\alpha+1} \left[\left(\frac{n+1}{n}\right)^{\alpha+1} - \frac{1}{n^{\alpha+1}} \right] \longrightarrow \frac{1}{\alpha+1}$$

quando $n \longrightarrow \infty$, o que prova o Teorema.

Teorema A.11 *Sejam* $\mathbf{A}(q \times q)$ *uma matriz simétrica e* $\mathbf{B}(q \times q)$ *uma matriz positiva definida; então:*

$$\sup_{\mathbf{x} \in \mathbb{R}^q} \frac{\mathbf{x}^t \mathbf{A} \mathbf{x}}{\mathbf{x}^t \mathbf{B} \mathbf{x}} = ch_1(\mathbf{A}\mathbf{B}^{-1}) \quad e \quad \inf_{\mathbf{x} \in \mathbb{R}^q} \frac{\mathbf{x}^t \mathbf{A} \mathbf{x}}{\mathbf{x}^t \mathbf{B} \mathbf{x}} = ch_q(\mathbf{A}\mathbf{B}^{-1})$$

onde $ch_q(\mathbf{M})$ *e* $ch_1(\mathbf{M})$ *denotam respectivamente a menor e a maior raiz característica de uma matriz* $\mathbf{M}(q \times q)$.

Demonstração: Ver Rao (1973, cap. 1), por exemplo.

Apêndice B

Bibliografia Básica

Apêndice C

Bibliografia Complementar