Notas de Aula do Curso
ET584: Probabilidade 4

Leandro Chaves Rêgo, Ph.D.

2015.1

Prefácio

Estas notas de aula foram feitas para compilar o conteúdo de várias referências bibliográficas tendo em vista o conteúdo programático da disciplina ET584-Probabilidade 4 do curso de graduação em Estatística da Universidade Federal de Pernambuco. Em particular, elas não contém nenhum material original e não substituem a consulta a livros textos. Seu principal objetivo é dispensar a necessidade dos alunos terem que copiar as aulas e, deste modo, poderem se concentrar em entender o conteúdo das mesmas.

Recife, março de 2015.
Leandro Chaves Rêgo, Ph.D.

Conteúdo

Capítulo 1

Revisão de Sequências de Números Reais e Séries Numéricas

1.1 Sequências de Números Reais

Intuitivamente, uma sequência de números reais $x_1, x_2, x_3, \ldots$ é uma sequência de pontos da reta e o seu limite é um ponto do qual os pontos x_n tornam-se e permanecem arbitrariamente próximos, desde que se tome o índice n suficientemente grande.

Exemplo 1.1.1: Seja $x_n = 1 + \frac{1}{n}$, para $n = 1, 2, 3, \ldots$. Note que a medida que n cresce todos os pontos desta sequência se tornam arbitrariamente próximos de 1, que como veremos adiante é o limite desta sequência. ∎

Formalmente,

Definição 1.1.2: Uma sequência de números reais é uma função $x : \mathbb{N} \to \mathbb{R}$, definida no conjunto $\mathbb{N} = \{1, 2, 3, \ldots\}$ dos números naturais e tomando valores no conjunto $\mathbb{R}$ dos números reais. O valor $x(n)$, para todo $n \in \mathbb{N}$, será representado por x_n e chamado de n-ésimo termo da sequência. ∎

Escreveremos $(x_1, x_2, \ldots, x_n, \ldots)$, ou (x_n) para indicar a sequência x.

Não se deve confundir a sequência x com o conjunto $x(\mathbb{N})$ dos seus termos. Para este conjunto usaremos a notação $x(\mathbb{N}) = \{x_1, x_2, \ldots, x_n, \ldots\}$. A função x não é necessariamente injetiva: pode-se ter $x_m = x_n$ com $m \neq n$, ou seja, podem haver termos diferentes que assumem o mesmo valor, ou em outras palavras, podem haver termos repetidos em uma sequência.

Diz-se que a sequência (x_n) é limitada quando o conjunto dos seus termos é limitado, isto é, quando existem números reais a, b tais que $a \leq x_n \leq b$ para todo $n \in \mathbb{N}$. Quando uma sequência não é limitada, diz-se que ela é ilimitada.

Uma sequência (x_n) é limitada superiormente quando existe um número real b tal que $x_n \leq b$ para todo $n \in \mathbb{N}$. Analogamente, (x_n) é limitada inferiormente quando existe a real tal que $a \leq x_n$ para todo $n \in \mathbb{N}$. É fácil ver que uma sequência é limitada se, e somente se, ela for limitada inferiormente e superiormente. Por outro lado, existem algumas sequências ilimitadas que são limitadas inferiormente ou superiormente. O próximo exemplo, ilustra melhor a questão.

Exemplo 1.1.3: A sequência $x_n = 1 + \frac{1}{n}$ é limitada, pois por exemplo, temos que $0 \leq x_n \leq 3$ para todo $n \in I\!N$. Por outro lado, a sequência $x_n = n^2$ é ilimitada, mas limitada inferiormente pois $x_n \geq 0$ para todo n. Finalmente, a sequência $x_n = (-2)^n$ é ilimitada, não é limitada inferiormente nem superiormente. ∎

Dada uma sequência (x_n) de números reais, uma subsequência de (x_n) é um sequência (portanto, deve conter infinitos termos) cujos termos são termos da sequência (x_n) e a ordem em que estes termos aparecem na subsequência deve ser a mesma em que eles aparecem na sequência original (x_n).

Exemplo 1.1.4: Seja a sequência $x = (-2, 4, -8, 16, -32, 64, -128, \ldots)$, uma subsequência de x é $y = (4, 16, 64, 256, \ldots)$. Por outro lado, $z = (4, 16)$ não é uma subsequência de x, pois não é uma sequência já que possui apenas 2 termos. Também temos que $w = (4, -2, 16, -8, 64, -32, \ldots)$ não é uma subsequência de x já que os termos em w não aparecem na mesma ordem em que aparecem em x, ou seja, por exemplo, em x o termo -2 precede o termo 4, mas o mesmo não é verdade em w. ∎

Formalmente, dada uma sequência x, uma subsequência de x é a restrição da função x a um subconjunto infinito $I\!N' = \{n_1 < n_2 < \ldots < n_i < \ldots\}$ de $I\!N$. Escreve-se $x' = (x_n)_{n \in I\!N'}$, ou $(x_{n_1}, x_{n_2}, \ldots, x_{n_i}, \ldots)$ ou $(x_{n_i})_{i \in I\!N}$ para indicar a subsequência x'.

Uma sequência chama-se crescente (resp., decrescente) quando $x_1 < x_2 < x_3 < \ldots$ (resp., $x_1 > x_2 > x_3 > \ldots$). Se vale $x_n \leq x_{n+1}$ (resp., $x_n \geq x_{n+1}$) para todo n, a sequência diz-se não-decrescente (resp., não-crescente). As sequências crescentes, não-decrescentes, decrescentes e não-crescentes são chamadas sequências monótonas.

Exemplo 1.1.5: $x_n = 0$, para todo $n \in I\!N$. Ela é limitada, não-crescente e não-decrescente. Neste caso, temos que $x(I\!N) = \{0\}$. ∎

Exemplo 1.1.6: $x_n = 1$ para todo n ímpar; e $x_n = -1$ para todo n par. Ela é limitada, porém não é monótona, e temos $x(I\!N) = \{-1, 1\}$. ∎

Exemplo 1.1.7: $x_n = 1/n$ para todo $n \in I\!N$. Ela é monótona decrescente e limitada. ∎

1.1.1 Limite de uma sequência

Intuitivamente, dizer que o número real a é limite da sequência (x_n) significa afirmar que, para valores muito grandes de n, os termos x_n tornam-se e se mantém tão próximos de a quanto se deseje. Com um pouco mais de precisão: estipulando-se um "erro" por meio de um número real $\epsilon > 0$, existe um índice n_0 (que depende de ϵ, em geral, quanto menor o "erro" ϵ maior terá que ser o n_0) tal que todos os termos x_n que têm índice n maior que n_0 são valores aproximados de a com erro inferior a ϵ. Formalmente,

Definição 1.1.8: O número real a é limite da sequência (x_n) de números reais, e escreve-se $a = \lim_n x_n$, quando para todo número real $\epsilon > 0$, existe um número natural n_0 tal que $|x_n - a| < \epsilon$ (o que é equivalente a $x_n \in (a-\epsilon, a+\epsilon)$), sempre que $n > n_0$. ∎

Como a definição implica que para qualquer $\epsilon > 0$ arbitrário, a distância entre x_n e a se torna menor que ϵ para n suficientemente grande, podemos escrever de forma equivalente a definição da seguinte maneira: o número real a é limite da sequência (x_n) de números reais, quando para alguma constante K real positiva, temos que para todo número real $\epsilon > 0$, existe um número natural n_0 tal que $|x_n - a| < K\epsilon$, sempre que $n > n_0$. A equivalência se dá pelo fato que $K\epsilon$ também é um número positivo e pode se tornar tão pequeno quanto se deseje apenas fazendo ϵ ser um número pequeno também.

Observe que se $\lim_n x_n = a$, então qualquer intervalo $(a-\epsilon, a+\epsilon)$, de centro a e raio $\epsilon > 0$, contém todos os termos x_n da sequência, com exceção de no máximo um número finito de índices n (os termos de x_1 até x_{n_0}). Reciprocamente, se qualquer intervalo de centro a contém todos os x_n, salvo talvez um número finito de índices n, então $\lim x_n = a$.

Quando $\lim_n x_n = a$, diz-se que a sequência (x_n) converge para a, ou tende para a e escreve-se $x_n \to a$. Uma sequência que possui limite chama-se convergente. Do contrário, ela se chama divergente. Dentre as sequências divergentes destacamos duas que possuem limites infinitos:

Definição 1.1.9: Uma sequência (x_n) de números reais tem limite ∞ (resp., $-\infty$), e escreve-se $\lim_n x_n = \infty$ (resp., $\lim_n x_n = -\infty$), quando para todo número real $M > 0$, existe um número natural n_0 tal que $x_n > M$ (resp., $x_n < -M$), sempre que $n > n_0$. ∎

Exemplo 1.1.10: A sequência $x_n = 1/n$ para todo n converge para 0. Pois, dado qualquer $\epsilon > 0$ existe $n_0 > 1/\epsilon$, então para todo $n > n_0$, temos $1/n < 1/n_0 < \epsilon$, ou seja, $n > n_0 \Rightarrow |x_n - 0| < \epsilon$. ∎

Exemplo 1.1.11: Vamos provar que $\lim_n \frac{n^2+1}{3n+10} = \infty$. Para isto notamos que $\frac{n^2+1}{3n+10} > \frac{n^2}{3n+10}$, e que para $n \geq 10$, vale a desigualdade

$$\frac{n^2}{3n+10} \geq \frac{n^2}{3n+n} = \frac{n^2}{4n} = \frac{n}{4}.$$

Por sua vez, $n/4 > M$ se $n > 4M$. Portanto, tomando $n_0 = \max\{10, 4M\}$, teremos

$$n > n_0 \Rightarrow \frac{n^2+1}{3n+10} > M.$$

∎

Exemplo 1.1.12: A sequência $x_n = \frac{1}{n} + (-1)^n$ é divergente. Note que dado qualquer $\epsilon > 0$, para $n > \frac{1}{\epsilon}$ e par, temos que $|x_n - 1| < \epsilon$. Por outro lado, para $n > \frac{1}{\epsilon}$ e ímpar, temos que $|x_n + 1| < \epsilon$. Logo, a sequência fica oscilando entre vizinhanças dos números -1 e 1, para n grande. ∎

Exemplo 1.1.13: A sequência $x_{2n} = 1$ e $x_{2n-1} = \frac{(n+1)}{n}$, $n = 1, 2, 3, \ldots$ converge para 1. Pois, dado qualquer $\epsilon > 0$, existe $n_0 > 1/\epsilon$, então para todo $n > n_0$ e ímpar, temos

$$|\frac{n+1}{n} - 1| = |\frac{1}{n}| < \epsilon.$$

Por outro lado, para todo $n > n_0$ e par, temos $|x_n - 1| = 0 < \epsilon$. Portanto, $x_n \to 1$. ∎

Para facilitar o cálculo do limite de sequências, vamos recordar a noção de limite de funções reais. Intuitivamente, temos que dada uma função real $f(x)$ dizemos que o limite quando x tende a um número a é igual a L, se quando x se aproxima de a o valor de $f(x)$ se aproxima de L. Mais formalmente, temos que $\lim_{x\to a} f(x) = L$ se para todo "erro" $\epsilon > 0$ existe um $\delta > 0$ (que depende de ϵ) tal que para todo $x \in (a - \delta, a + \delta)$ que é diferente de a, temos que $f(x) \in (L - \epsilon, L + \epsilon)$. Por outro lado, quando queremos calcular o limite assintótico de uma dada função, estamos interessados em saber se quando x cresce arbitrariamente a função $f(x)$ tende a algum valor, deste modo dizemos que o limite quando x tende a infinito é igual a L, se para x grande o suficiente $f(x)$ se torna tão próximo de L quanto se queira. Mais formalmente, temos que $\lim_{x\to\infty} f(x) = L$ se para todo "erro" $\epsilon > 0$ existe um número natural $n_0 > 0$ (que depende de ϵ) tal que para todo número real $x > n_0$, temos que $f(x) \in (L - \epsilon, L + \epsilon)$.

Suponha que dada uma função real $f(x)$, uma sequência seja definida por $x_n = f(n)$ para todo $n \in I\!N$. Então, se $\lim_{x\to\infty} f(x) = L$, temos que para todo $\epsilon > 0$, existe um número natural $n_0 > 0$ tal que para todo número real $x > n_0$, temos que $f(x) \in (L - \epsilon, L + \epsilon)$. Como todo número natural é um número real, temos que para todo natural $n > n_0$, $x_n = f(n) \in (L - \epsilon, L + \epsilon)$. Logo, $\lim x_n = L$. Assim, toda vez que uma sequência (x_n) for uma restrição, para x natural, de uma função $f(x)$ definida para x real, ou $x > 0$, temos que se $\lim_{x\to\infty} f(x) = L$, podemos concluir que $x_n \to L$. Deste modo, podemos utilizar nosso conhecimento sobre limites de funções reais para calcularmos o limite de sequências. Em particular, podemos utilizar a regra de L'Hopital que diz que se $\lim_{x\to a} f(x) = 0$ e $\lim_{x\to a} g(x) = 0$, ou, se $\lim_{x\to a} f(x) = \infty$ e $\lim_{x\to a} g(x) = \infty$, então

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Exemplo 1.1.14: Seja $x_n = n(1-e^{-a/n})$. Vamos calcular o limite da função real $x(1-e^{-a/x})$ quando $x \to \infty$. Podemos reescrever esta função da seguinte maneira:

$$\frac{(1 - e^{-a/x})}{1/x}.$$

Note que tanto o numerador quanto o denominador convergem para zero quando $x \to \infty$. Utilizando a regra de L'Hopital, temos:

$$\lim_{x\to\infty} \frac{(1 - e^{-a/x})}{1/x} = \lim_{x\to\infty} \frac{(-ax^{-2}e^{-a/x})}{-x^{-2}} = \lim_{x\to\infty} ae^{-a/x} = a.$$

Portanto, o limite de x_n é igual a a. ∎

É importante ressaltar que mesmo que a sequência seja definida a partir da restrição de uma função real, o fato da sequência convergir para um certo limite L não implica que a função real tenderá a L quando $x \to \infty$. Por exemplo, considere a função real $f(x)$ tal que $f(x) = 0$ para $x \notin I\!N$ e $f(x) = 1$ para $x \in I\!N$, temos que $x_n = f(n) = 1$ para todo $n \in I\!N$. Logo, $x_n \to 1$, porém $\lim_{x\to\infty} f(x) \neq 1$.

A seguir provaremos alguns resultados sobre limites.

Teorema 1.1.15: *(Unicidade do limite). Se* $\lim_n x_n = a$ *e* $\lim_n x_n = b$*, então* $a = b$*.*

Prova: Seja $\lim_n x_n = a$. Dado qualquer número real $b \neq a$, mostraremos que não se tem $\lim_n x_n = b$. Para isso, tomemos $\epsilon = \frac{|b-a|}{2}$. Com essa escolha de ϵ, temos que os intervalos $(a - \epsilon, a + \epsilon)$ e $(b - \epsilon, b + \epsilon)$ são disjuntos. Ora, como $\lim_n x_n = a$, existe n_0 tal que $n > n_0$ implica que $x_n \in (a - \epsilon, a + \epsilon)$, e, portanto, $x_n \notin (b - \epsilon, b + \epsilon)$ para todo $n > n_0$. Logo, $\lim_n x_n \neq b$. ∎

Teorema 1.1.16: *Se* $\lim_n x_n = a$*, então toda subsequência de* (x_n) *converge para o limite* a*.*

Prova: Seja (x_{n_i}) uma subsequência de (x_n). Dado $\epsilon > 0$, existe $n_0 \in I\!N$ tal que $n > n_0 \Rightarrow |x_n - a| < \epsilon$. Como os índices da subsequência formam um subconjunto infinito, existe entre eles um $n_{i_0} > n_0$. Então, $n_i > n_{i_0} \Rightarrow n_i > n_0$, o que por sua vez implica que $|x_{n_i} - a| < \epsilon$. Logo $\lim_i x_{n_i} = a$. ∎

Observação 1.1.17: Há duas aplicações dos Teoremas 1.1.15 e 1.1.16. Uma delas é para mostrar que uma certa sequência não converge: basta obter duas subsequências de (x_n) com limites distintos. A outra é para determinar o limite de uma sequência (x_n) que, *a priori*, se sabe que converge: basta determinar o limite de alguma subsequência. Ele será o limite procurado. ∎

Exemplo 1.1.18: A sequência $(1, 0, 1, 0, 1, \ldots)$ não é convergente pois admite duas subsequências constantes que convergem para limites diferentes. ∎

Teorema 1.1.19: *Toda sequência convergente é limitada.*

Prova: Seja $a = \lim_n x_n$. Então, tomando $\epsilon = 1$, vemos que existe $n_0 \in I\!N$ tal que $n > n_0 \Rightarrow x_n \in (a-1, a+1)$. Consideremos o conjunto finito $F = \{x_1, x_2, \ldots, x_{n_0}, a-1, a+1\}$. Seja c o menor e d o maior elemento de F. Então, para $n \leq n_0$, é óbvio que $c \leq x_n \leq d$ e para $n > n_0$, temos que $c \leq a - 1 < x_n < a + 1 \leq d$. Então, todos os termos x_n da sequência estão contidos no intervalo $[c, d]$; logo a sequência é limitada. ∎

Dado um conjunto de números reais A, define-se como uma cota superior (resp. inferior) para A como sendo qualquer número real c tal que $c \geq x$ (resp. $c \leq x$) para todo $x \in A$. Por exemplo, se $A = (-1, 1]$, então qualquer número maior ou igual a 1 é uma cota superior para A e qualquer número menor ou igual a -1 é uma cota inferior para A. Define-se como o supremo (resp. ínfimo) de um conjunto A a menor (resp. maior) cota superior (resp. inferior) de A. No exemplo anterior, temos que $supA = 1$ e $infA = -1$. Note que o supremo e/ou o ínfimo de um conjunto, ao contrário de seu máximo e mínimo, não precisam ser elementos do conjunto. No exemplo, note que $infA \notin A$.

Teorema 1.1.20: *Toda sequência monótona limitada é convergente.*

Prova: Para fixar as idéias, seja $(x_1 \leq x_2 \leq \ldots \leq x_n \leq \ldots)$ uma sequência não-decrescente limitada. Tomemos $a = sup\{x_n : n = 1, 2, \ldots\}$. Afirmamos que $a = \lim_n x_n$. Com efeito, dado qualquer $\epsilon > 0$, como $a - \epsilon < a$, o número $a - \epsilon$ não é cota superior do conjunto dos x_n. Logo, existe algum $n_0 \in I\!N$ tal que $a - \epsilon < x_{n_0}$. Como a sequência é não-decrescente, $n > n_0 \Rightarrow x_{n_0} \leq x_n$ e, portanto, $a - \epsilon < x_n$. Como $x_n \leq a$ para todo n (pela definição de supremo), vemos que $n > n_0 \Rightarrow a - \epsilon < x_n < a + \epsilon$, ou seja, $\lim_n x_n = a$. ∎

1.1.2 Propriedades Aritméticas dos Limites

Estudaremos agora como se comportam os limites de sequências relativamente às operações aritméticas e às desigualdades.

Teorema 1.1.21: *Se* $\lim_n x_n = 0$ *e* (y_n) *é uma sequência limitada, então* $\lim_n x_n \cdot y_n = 0$ *(mesmo que não exista* $\lim_n y_n$*).*

Prova: Existe $c > 0$ tal que $|y_n| < c$ para todo $n \in I\!N$. Dado $\epsilon > 0$, como $\lim_n x_n = 0$, podemos encontrar $n_0 \in I\!N$ tal que $n > n_0 \Rightarrow |x_n| < \frac{\epsilon}{c}$. Logo, $n > n_0 \Rightarrow |x_n \cdot y_n| = |x_n| \cdot |y_n| < \frac{\epsilon}{c} \cdot c = \epsilon$. Isto mostra que $x_n \cdot y_n \to 0$. ∎

Exemplo 1.1.22: Qualquer que seja $x \in I\!R$, temos $\lim_n \frac{sen(nx)}{n} = 0$. Com efeito, $\frac{sen(nx)}{n} = sen(nx) \cdot \frac{1}{n}$, com $|sen(nx)| \leq 1$ e $\frac{1}{n} \to 0$. ∎

Teorema 1.1.23: *Se* $\lim_n x_n = a$ *e* $\lim_n y_n = b$*, então*

1. $\lim_n(x_n + y_n) = a + b$; $\lim_n(x_n - y_n) = a - b$;
2. $\lim_n(x_n \cdot y_n) = a \cdot b$;
3. $\lim_n(x_n/y_n) = a/b$ *se* $b \neq 0$ *e* $y_n \neq 0, \forall n$.

Prova: Para parte 1, dado $\epsilon > 0$ existem n_1 e n_2 em $I\!N$ tais que $n > n_1 \Rightarrow |x_n - a| < \frac{\epsilon}{2}$ e $n > n_2 \Rightarrow |y_n - b| < \frac{\epsilon}{2}$. Seja $n_0 = \max\{n_1, n_2\}$. Então, $n > n_0 \Rightarrow n > n_1$ e $n > n_2$. Logo $n > n_0$ implica:

$$\begin{aligned} &|(x_n + y_n) - (a + b)| = |(x_n - a) + (y_n - b)| \\ &\leq |x_n - a| + |y_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Isto prova que $\lim_n(x_n + y_n) = a + b$. O caso da diferença $x_n - y_n$ se trata do mesmo modo.

Para parte 2, temos $x_n y_n - ab = x_n y_n - x_n b + x_n b - ab = x_n(y_n - b) + (x_n - a)b$. Ora, (x_n) pelo Teorema 1.1.19 é uma sequência limitada e pela parte 1, temos que $\lim_n(y_n - b) = 0$. Logo, pelo Teorema 1.1.21, $\lim_n[x_n(y_n - b)] = 0$. Por motivo semelhante, $\lim_n[(x_n - a)b] = 0$. Assim, pela parte 1, já demonstrada, temos $\lim_n(x_n y_n - ab) = \lim_n[x_n(y_n - b)] + \lim_n[(x_n - a)b] = 0$, donde $\lim_n x_n y_n = ab$.

Para parte 3, notemos que, como pela parte 2, $y_n b \to b^2$, existe n_0 tal que $n > n_0 \Rightarrow y_n b > \frac{b^2}{2}$ (basta tomar $\epsilon = \frac{b^2}{2}$). Segue-se que, para todo $n > n_0$, $\frac{1}{y_n b}$ é um número positivo inferior a $\frac{2}{b^2}$. Logo, a sequência $(\frac{1}{y_n b})$ é limitada. Como temos que,

$$\frac{x_n}{y_n} - \frac{a}{b} = \frac{bx_n - ay_n}{y_n b} = (bx_n - ay_n)\frac{1}{y_n b}.$$

Como pelas partes 1 e 2, $\lim_n (bx_n - ay_n) = ab - ab = 0$, segue-se do Teorema 1.1.21 que $\lim_n (\frac{x_n}{y_n} - \frac{a}{b}) = 0$ e, portanto, $\lim_n \frac{x_n}{y_n} = \frac{a}{b}$. ∎

Observação 1.1.24: É claro que resultados análogos aos ítens 1 e 2, do Teorema 1.1.23 valem para qualquer número finito de sequências. Por exemplo, se $\lim_n x_n = a$, $\lim_n y_n = b$, e $\lim_n z_n = c$, então $\lim_n (x_n + y_n + z_n) = a + b + c$ e $\lim_n (x_n y_n z_n) = abc$. Contudo, deve-se tomar cuidado de não tentar aplicar o teorema para certas somas (ou produtos) em que o número de parcelas é variável e cresce acima de qualquer limite. Por exemplo, seja $s_n = \frac{1}{n} + \ldots + \frac{1}{n}$ (n parcelas). Então, $s_n = 1$ e, portanto, $\lim_n s_n = 1$. Por outro lado, cada parcela $\frac{1}{n}$ tem limite zero. Uma aplicação descuidada do Teorema 1.1.23 levaria ao absurdo de concluir que

$$\lim_n s_n = \lim_n 1/n + \ldots + \lim_n 1/n = 0 + \ldots + 0 = 0.$$

∎

Teorema 1.1.25: *Sejam $x_n \leq y_n$ para todo $n \in I\!N$, $\lim_n x_n = a$, e $\lim_n y_n = b$, então $a \leq b$.*

Prova: Suponha por contradição que $a > b$. Seja $\epsilon = \frac{a-b}{2}$. Então, por hipótese existem n_1 e n_2 tais que $n > n_1 \Rightarrow x_n \in (a - \epsilon, a + \epsilon)$ e $n > n_2 \Rightarrow y_n \in (b - \epsilon, b + \epsilon)$. Pondo $n_0 = \max\{n_1, n_2\}$, vemos que $n > n_0$ implica $y_n < b + \epsilon = \frac{a+b}{2} = a - \epsilon < x_n$, absurdo. ∎

Observação 1.1.26: O resultado análogo ao do Teorema 1.1.25 para desigualdades estritas **não é válido**. Ou seja, **não é verdade** que se $x_n < y_n$ para todo $n \in I\!N$, $\lim_n x_n = a$, e $\lim_n y_n = b$, então $a < b$. Por exemplo, seja $x_n = 0$ e $y_n = 1/n$ para todo $n \in I\!N$. Temos que $\lim_n x_n = \lim_n y_n = 0$. ∎

Teorema 1.1.27: *Sejam $x_n \leq z_n \leq y_n$ para todo $n \in I\!N$. Se $\lim_n x_n = \lim_n y_n = a$, então $\lim_n z_n = a$.*

Prova: Dado $\epsilon > 0$, existem n_1 e n_2 tais que $n > n_1 \Rightarrow x_n \in (a - \epsilon, a + \epsilon)$ e $n > n_2 \Rightarrow y_n \in (a - \epsilon, a + \epsilon)$. Pondo $n_0 = \max\{n_1, n_2\}$, vemos que $n > n_0$ implica $a - \epsilon < x_n \leq z_n \leq y_n < a + \epsilon$. Portanto, $\lim_n z_n = a$. ∎

Vamos a seguir provar que limites são preservados a aplicações de funções contínuas. Recorde que uma função $f : I\!R \to I\!R$ é contínua em $a \in I\!R$ se para todo $\epsilon > 0$, existe $\delta > 0$, tal que $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon$.

Teorema 1.1.28: *Se $\lim_n x_n = a$ e $g : I\!R \to I\!R$ é uma função contínua em a, então $\lim_n g(x_n) = g(a)$.*

Prova: Escolha $\epsilon > 0$, arbitrário. Como g é contínua em a, existe $\delta > 0$ tal que $|x - a| < \delta \Rightarrow |g(x) - g(a)| < \epsilon$. Por outro lado, como $x_n \to a$, temos que existe n_0 tal que $n > n_0 \Rightarrow |x_n - a| < \delta$. Portanto, para $n > n_0$, temos que $|g(x_n) - g(a)| < \epsilon$. Ou seja, $\lim_n g(x_n) = g(a)$. ∎

1.1.3 Valores de aderência, $\liminf$, $\limsup$

Definição 1.1.29: Um número real a chama-se *valor de aderência* de uma sequência (x_n) quando a é limite de alguma subsequência de (x_n). ∎

Exemplo 1.1.30: Se $\lim_n x_n = a$, então a é o único valor de aderência de (x_n). A sequência $(0, 1, 0, 2, 0, 3, \ldots)$ tem 0 como seu único valor de aderência, embora não seja convergente. A sequência $(0, 1, 0, 1, 0, \ldots)$ tem como valores de aderência 0 e 1. Seja $x_n = n$, a sequência (x_n) não possui valores de aderência. ∎

O próximo teorema mostra que um número real a é valor de aderência de uma sequência (x_n) se, e somente se, toda vizinhança de a contem infinitos termos de (x_n).

Teorema 1.1.31: *a é valor de aderência de (x_n) se, e somente se, para todo $\epsilon > 0$ e todo $n_0 \in I\!N$ existir $n \in I\!N$ tal que $n > n_0$ e $x_n \in (a - \epsilon, a + \epsilon)$.*

Prova: Suponha que a é um valor de aderência de (x_n). Então, existe uma subsequência (x_{n_i}) tal que $\lim_i x_{n_i} = a$, ou seja, para todo $\epsilon > 0$, existe n_{i_0}, tal que $i > i_0 \Rightarrow x_{n_i} \in (a - \epsilon, a + \epsilon)$. Então, dado qualquer n_0, como (x_{n_i}) contém infinitos termos de (x_n), existe $n_i > n_0$ tal que $i > i_0$ e, consequentemente, $x_{n_i} \in (a - \epsilon, a + \epsilon)$.

Reciprocamente, suponha que para todo $\epsilon > 0$ e todo $n_0 \in I\!N$ exista $n \in I\!N$ tal que $n > n_0$ e $x_n \in (a - \epsilon, a + \epsilon)$. Vamos construir uma subsequência (x_{n_i}) tal que $\lim_i x_{n_i} = a$, mais especificamente, vamos construir uma subsequência tal que $x_{n_i} \in (a - 1/i, a + 1/i)$. Por suposição, existe n_1 tal que $x_{n_1} \in (a - 1, a + 1)$, vamos definir os demais termos da subsequência por indução. Suponha que exista $n_i > n_{i-1}$ tal que $x_{n_i} \in (a - 1/i, a + 1/i)$, queremos provar que existe $n_{i+1} > n_i$ tal que $x_{n_{i+1}} \in (a - \frac{1}{i+1}, a + \frac{1}{i+1})$. Por suposição, para $\epsilon = \frac{1}{i+1}$ e $n_0 = n_i$, existe um $n > n_0$ tal que $x_n \in (a - \frac{1}{i+1}, a + \frac{1}{i+1})$. Chamemos este n de n_{i+1}, e construímos a desejada subsequência. Então, temos que a é limite desta subsequência (x_{n_i}) e, portanto, valor de aderência de (x_n). ∎

Seja (x_n) uma sequência **limitada** de números reais. Mostraremos que o conjunto de valores de aderência de (x_n) não é vazio, que entre eles existe um que é o menor de todos e outro que é o maior, e que a sequência converge se, e somente se, possui apenas um valor de aderência. Suponha que $\alpha \leq x_n \leq \beta$ para todo $n \in I\!N$. Escrevamos $X_n = \{x_n, x_{n+1}, \ldots\}$. Temos $[\alpha, \beta] \supset X_2 \supset \ldots \supset X_n \supset \ldots$ Logo, definindo $a_n = \inf X_n$ e $b_n = \sup X_n$, temos

$$\alpha \leq a_1 \leq a_2 \leq \ldots \leq a_n \leq \ldots \leq b_n \leq \ldots \leq b_2 \leq b_1 \leq \beta$$

Como toda sequência monotônica e limitada é convergente, temos que $a_n \to a$ e $b_n \to b$. Escreve-se $a = \liminf x_n$ e $b = \limsup x_n$ e diz-se que a é o *limite inferior* e que b é o *limite superior* da sequência (x_n). Como $a_n \leq b_n$, tem-se $\liminf x_n \leq \limsup x_n$.

Exemplo 1.1.32: Sejam $x_{2n-1} = -\frac{1}{n}$ e $x_{2n} = 1 + \frac{1}{n}$. Verifica-se sem dificuldade que $\inf X_{2n-2} = \inf X_{2n-1} = -\frac{1}{n}$ e $\sup X_{2n-1} = \sup X_{2n} = 1 + \frac{1}{n}$. Logo, $\liminf x_n = 0$ e $\limsup x_n = 1$ e estes são os dois únicos valores de aderência da sequência (x_n). ∎

Teorema 1.1.33: *Seja (x_n) uma sequência limitada. Então, $\liminf x_n$ é o menor valor de aderência e $\limsup x_n$ é o maior valor de aderência de (x_n).*

Prova: Provaremos inicialmente que $a = \liminf x_n$ é valor de aderência de (x_n). Para isto, usaremos o Teorema 1.1.31, e mostraremos que dados $\epsilon > 0$ e $n_0 \in I\!N$ arbitrários, existe $n \in I\!N$ tal que $n > n_0$ e $x_n \in (a - \epsilon, a + \epsilon)$. Como $a = \lim_n a_n$, existe $n_1 > n_0$ tal que $a - \epsilon < a_{n_1} < a + \epsilon$. Como $a_{n_1} = \inf X_{n_1}$, segue-se da última igualdade que $a + \epsilon$ (sendo maior que a_{n_1}) não é cota inferior de X_{n_1}. Logo, existe $n \geq n_1$ tal que $a_{n_1} \leq x_n < a+\epsilon$. Isto nos dá $n > n_0$ com $a-\epsilon < x_n < a+\epsilon$. Mostremos agora que nenhum número $c < a$ pode ser valor de aderência de (x_n). Ora, como $a = \lim_n a_n$, segue-se de $c < a$ que existe $n_0 \in I\!N$ tal que $c < a_{n_0} \leq a$. Como $a_{n_0} = \inf X_{n_0}$, concluímos que $n \geq n_0 \Rightarrow c < a_{n_0} \leq x_n$. Tomando $\epsilon = a_{n_0} - c$, vemos que $c+\epsilon = a_{n_0}$, logo o intervalo $(c-\epsilon, c+\epsilon)$ não contém termo x_n algum com $n \geq n_0$. Isto exclui a possibilidade de c ser valor de aderência de (x_n). A demonstração para $\limsup$ se faz de modo semelhante. ∎

Corolário 1.1.34: *Uma sequência limitada de números reais (x_n) é convergente se, e somente se, $\liminf x_n = \limsup_n x_n$, isto é, se, e somente se, possui um único valor de aderência.*

Prova: Se (x_n) convergir para a, então vimos que a é o único valor de aderência. Portanto, $\liminf x_n = \limsup_n x_n = a$. Se $\liminf x_n = \limsup_n x_n = a$, então suponha que x_n não convirja para a. Logo, existe $\epsilon > 0$, tal que para todo $n_0 \in N$ existe $n > n_0$ tal que $x_n \notin (a - \epsilon, a + \epsilon)$. Então existe uma subsequência de (x_n) cujos termos não estão no intervalo $(a - \epsilon, a + \epsilon)$. Pelo Teorema 1.1.33, esta subsequência possui valores de aderência que são valores de aderência de (x_n) e estão fora do intervalo $(a-\epsilon, a+\epsilon)$, uma contradição. ∎

1.1.4 Sequências de Cauchy

Provamos anteriormente que toda sequência monótona limitada é convergente. Isto nos permite concluir que uma sequência possui limite mesmo sem conhecermos o valor deste limite. Veremos agora o critério de Cauchy, que nos dá uma condição necessária e suficiente para a convergência de números reais.

Definição 1.1.35: Uma sequência (x_n) de números reais é uma sequência de Cauchy quando dado qualquer $\epsilon > 0$, existe um $n_0 \in I\!N$ tal que $n > n_0$ e $m > n_0$ implica $|x_m - x_n| < \epsilon$. ∎

A fim de que (x_n) seja uma sequência de Cauchy, exige-se que seus termos x_m, x_n, para valores suficientemente grandes de índices n e m, se aproximem e permaneçam arbitrariamente próximos uns dos outros. Compare com a definição de limite, onde se exige que os termos x_n se aproximem e permaneçam arbitrariamente próximos de um número real a dado *a priori*. Aqui se impõe uma condição apenas sobre os termos da própria sequência.

Teorema 1.1.36: *Toda sequência convergente é de Cauchy.*

Prova: Seja $lim_n x_n = a$. Então, dado $\epsilon > 0$ existe n_0 tal que $n > n_0 \Rightarrow |x_n - a| < \epsilon/2$ e $m > n_0 \Rightarrow |x_m - a| < \epsilon/2$. Logo, $m, n > n_0 \Rightarrow |x_m - x_n| \leq |x_m - a| + |x_n - a| < \epsilon/2 + \epsilon/2 = \epsilon$, ou seja (x_n) é uma sequência de Cauchy. ∎

Intuitivamente: se $\lim_n x_n = a$ então, para valores grades de n, os termos x_n se aproximam de a, e portanto necessariamente aproximam-se uns dos outros.

Teorema 1.1.37: *Toda sequência de Cauchy de números reais é convergente.*

Prova: Iremos provar este teorema utilizando dois Lemas.

Lema 1.1.38: *Toda sequência de Cauchy é limitada.*

Prova: Seja (x_n) uma sequência de Cauchy. Tomando $\epsilon = 1$, obtemos $n_0 \in I\!N$ tal que $m, n > n_0 \Rightarrow |x_m - x_n| < 1$. Em particular para $m = n_0 + 1$, $n > n_0 \Rightarrow |x_{n0+1} - x_n| < 1$, ou seja, $n > n_0 \Rightarrow x_n \in (x_{n0+1} - 1, x_{n0+1} + 1)$. Sejam α o menor e β o maior elemento do conjunto $X = \{x_1, x_2, \ldots, x_{n_0}, x_{n0+1} - 1, x_{n0+1} + 1\}$. Então, $x_n \in [\alpha, \beta]$ para todo $n \in I\!N$, logo (x_n) é limitada. ∎

Lema 1.1.39: *Se uma sequência de Cauchy (x_n) possui um valor de aderência $a \in I\!R$, então $\lim_n x_n = a$.*

Prova: Dado $\epsilon > 0$, como (x_n) é uma sequência de Cauchy, existe $n_0 \in I\!N$ tal que $m, n > n_0 \Rightarrow |x_m - x_n| < \frac{\epsilon}{2}$. Como a é valor de aderência de (x_n), existe também $n_1 > n_0$ tal que $|x_{n_1} - a| < \epsilon/2$. Portanto, $n > n_0 \Rightarrow |x_n - a| \leq |x_n - x_{n_1}| + |x_{n_1} - a| < \epsilon$. Isto mostra que $\lim_n x_n = a$. ∎

Então, seja (x_n) uma sequência de Cauchy. Pelo Lema 1.1.38, ela é limitada. Logo, pelo Teorema 1.1.33, possui um valor de aderência e segue do Lema 1.1.39 que (x_n) converge. ∎

1.2 Séries de Números Reais

Nesta seção, estenderemos a operação de adição de modo a atribuir significado a uma igualdade do tipo $\frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n} + \ldots = 1$, na qual o primeiro termo é uma "soma" com uma infinidade de parcelas. É claro que não tem sentido somar uma sequência infinita de números reais. O que o primeiro membro da igualdade acima exprime é o limite $\lim_n(\frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n})$. A afirmação contida naquela igualdade significa que para todo $\epsilon > 0$ existe n_0 tal que, para todo $n > n_0$, a soma $\frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n}$ difere de 1 por menos de ϵ.

Definição 1.2.1: Seja (a_n) uma sequência de números reais. A partir dela, formamos uma nova sequência (s_n) cujos elementos são as somas

$$s_1 = a_1, s_2 = a_1 + a_2, \ldots, s_n = a_1 + \ldots + a_n,$$

que são chamados de soma parcial ou reduzida da série $\sum a_n$. A parcela a_n é chamada o n-ésimo termo ou o termo geral da série. ∎

Se existir o limite

$$s = \lim s_n = \lim_{n\to\infty}(a_1 + \ldots + a_n),$$

diremos que a série $\sum a_n$ é convergente e o limite s será chamado a soma da série. Escreveremos então

$$s = \sum a_n = \sum_{n=1}^{\infty} a_n = a_1 + a_2 + \ldots + a_n + \ldots.$$

Se a sequência de somas parciais não convergir, diremos que a série $\sum a_n$ é divergente.

Observação 1.2.2: Toda sequência (x_n) de números reais pode ser considerada como a sequência das reduzidas de uma série. Basta tomar $a_1 = x_1$ e $a_{n+1} = x_{n+1} - x_n$ para todo $n \in I\!N$. Então, $a_1 + \ldots + a_n = x_1 + (x_2 - x_1) + \ldots + (x_n - x_{n-1}) = x_n$. A série $\sum a_n$ assim obtida converge se, e somente se, a sequência (x_n) é convergente. No caso afirmativo, a soma desta série é igual a $\lim_n x_n$. Assim falando, pode-se dar a impressão de que a teoria das séries coincide com a teoria dos limites de sequências. Isto não é verdade, pelo seguinte motivo. Ao estudar a série cujas reduzidas são s_n, estaremos deduzindo suas propriedades a partir das diferenças $a_n = s_n - s_{n-1}$. Em vez de tomar como ponto de partida o comportamento dos números s_n, concentraremos atenção sobre os termos a_n. ∎

A primeira condição necessária para convergência de uma série é que seu termo geral tenda para zero.

Teorema 1.2.3: *Se* $\sum a_n$ *é uma série convergente, então* $\lim_n a_n = 0$.

Prova: Seja $s_n = a_1 + \ldots + a_n$. Então, existe $s = \lim_n s_n$. Evidentemente, tem-se também $s = \lim_n s_{n-1}$. Logo, $0 = s - s = \lim_n s_n - \lim_n s_{n-1} = \lim_n (s_n - s_{n-1}) = \lim_n a_n$. ∎

Exemplo 1.2.4: A recíproca do Teorema 1.2.3 é falsa. O contra-exemplo clássico é dado pela *série harmônica* $\sum \frac{1}{n}$. Seu termo geral, $\frac{1}{n}$, tende para zero mas a série diverge. Com efeito, temos

$$\begin{aligned} s_{2^n} &= 1 + \frac{1}{2} + (\frac{1}{3} + \frac{1}{4}) + \ldots + (\frac{1}{2^{n-1}+1} + \cdots \frac{1}{2^n}) \\ &> 1 + \frac{1}{2} + \frac{2}{4} + \ldots + \frac{2^{n-1}}{2^n} = 1 + n \cdot \frac{1}{2}. \end{aligned}$$

Segue-se que $\lim_n s_{2^n} = +\infty$ e, por conseguinte, como s_n é monotonicamente crescente, temos $\lim_n s_n = +\infty$. Resulta daí que, para $0 < r < 1$, a série $\sum \frac{1}{n^r}$ diverge, pois $\frac{1}{n^r} > \frac{1}{n}$ para todo $n > 1$. ∎

Exemplo 1.2.5: A *série geométrica* $\sum_{n=0}^{\infty} a^n$ é divergente quando $|a| \geq 1$, pois neste caso seu termo geral não tende a zero. Quando $|a| < 1$, a série geométrica converge, pois

$$\begin{aligned} & s_n - as_n = (1 + a + \ldots + a^n) - (a + a^2 + \ldots + a^{n+1}) = 1 - a^{n+1} \\ & \Rightarrow s_n(1-a) = 1 - a^{n+1} \Rightarrow s_n = \frac{1-a^{n+1}}{1-a}. \end{aligned}$$

Então, $\sum_{n=0}^{\infty} a^n = \lim_n s_n = \frac{1}{1-a}$. ∎

Exemplo 1.2.6: A série $\sum_{n=1}^{\infty}(-1)^{n+1} = 1 - 1 + 1 - 1 + \ldots$ é divergente pois seu termo geral não tende a zero. Suas somas parciais de ordem ímpar são iguais a 1 e as de ordem par são iguais a zero. ∎

Uma série $\sum a_n$ pode divergir por dois motivos. Ou porque as reduzidas $s_n = a_1 + \ldots + a_n$ não são limitadas ou porque elas oscilam em torno de alguns valores de aderência. Quando os termos da série têm todos o mesmo sinal, esta última possibilidade não ocorre, pois, neste caso, as reduzidas formam uma sequência monótona. A seguir nós estudaremos alguns critérios de convergência de séries.

Teorema 1.2.7: *Seja $a_n > 0$ para todo $n \in I\!N$. A série $\sum a_n$ converge se, e somente se, as somas parciais $s_n = a_1 + \ldots + a_n$ formam uma sequência limitada.*

Prova: Sendo $a_n > 0$, temos $s_1 \leq s_2 \leq s_3 \leq \ldots$; logo a sequência (s_n) sendo monótona converge se, e somente se, é limitada. ∎

Dada uma série de termos não negativos $a_1, a_2, \ldots$, suponha que os termos sejam reindexados numa outra ordem qualquer, $a'_1, a'_2, \ldots$, de forma que a'_1 pode ser a_{15}, a'_2 pode ser a_1, etc. Então, como os termos são todos não negativos, a nova soma parcial $s'_n = a'_1 + \ldots + a'_n$ é dominada por alguma soma parcial s_m com $m > n$. Se a série original converge para s, teremos $s'_n \leq s_m \leq s$, logo s'_n é limitada e portanto convergente. Seu limite s' é seu supremo, de sorte que $s' \leq s$. Mas a série original pode também ser interpretada como obtida de $\sum a'_n$ por reindexação de seus termos a'_n, logo temos também $s \leq s'$. Concluímos que uma série de termos não negativos que converge tem a mesma soma, independente da ordem de seus termos. É fácil ver também que se a série de termos não negativos diverge, ela será sempre divergente, não importa a ordem de seus termos.

O próximo teorema estabelece mais uma caracterização de séries convergentes e divergentes.

Teorema 1.2.8: *Seja $\sum_{n=1}^{\infty} a_n$ uma série de termos não-negativos. Então,*

(a) se $\sum_{n=1}^{\infty} a_n < \infty$, então $\lim_k \sum_{n=k+1}^{\infty} a_n = 0$;

(b) se $\sum_{n=1}^{\infty} a_n = \infty$, então $\forall k$, $\sum_{n=k}^{\infty} a_n = \infty$.

Prova: Para parte (a), como $s_k = \sum_{n=1}^{k} a_n$ é uma sequência monótona não-decrescente e limitada, então ela é convergente. Logo, pelo critério de Cauchy para sequências temos que $\forall \epsilon > 0$, $\exists m$ tal que $k, p > m \Rightarrow |s_p - s_k| < \epsilon$. Assumindo sem perda de generalidade que $p > k$, temos que $\forall \epsilon > 0$, $\exists m$ tal que $k, p > m \Rightarrow |\sum_{n=k+1}^{p} a_n| < \epsilon$. Fazendo $p \to \infty$, temos que $\forall \epsilon > 0$, $\exists m$ tal que $k > m \Rightarrow |\sum_{n=k+1}^{\infty} a_n| \leq \epsilon$, ou seja, $\sum_{n=k+1}^{\infty} a_n \to 0$.

Para parte (b), suponha por contradição que $\sum_{n=1}^{\infty} a_n = \infty$ e que exista k tal que $\sum_{n=k}^{\infty} a_k < \infty$. Seja $L = \sum_{n=1}^{k-1} a_n$, então $\sum_{n=1}^{\infty} a_n = L + \sum_{n=k}^{\infty} a_k < \infty$, uma contradição. ∎

1.2.1 Critérios de Convergência

Um dos problemas centrais no estudo das séries consiste em saber se uma dada série converge ou não. Há vários critérios para se testar a convergência de uma série, nós vamos destacar dois dos mais importantes: o teste da comparação e o teste da razão.

Teste de Comparação

Teorema 1.2.9: *Sejam* $\sum a_n$ *e* $\sum b_n$ *duas séries de termos não negativos. Suponhamos ainda que a primeira seja dominada pela segunda,* $a_n \leq b_n$ *para todo* $n \in I\!N$. *Então,*

(i) $\sum b_n$ *converge* $\Rightarrow \sum a_n$ *converge;*

(ii) $\sum a_n$ *diverge* $\Rightarrow \sum b_n$ *diverge.*

Prova: As somas parciais das séries dadas $s_n = a_1 + \ldots + a_n$ e $t_n = b_1 + \ldots + b_n$ são sequências não decrescentes, satisfazendo à desigualdade $s_n \leq t_n$, pois $0 \leq a_n \leq b_n$. No caso (i), t_n converge para um certo limite t, então $s_n \leq t$ para todo n, ou seja, s_n é uma sequência monótona limitada e, portanto, convergente.

Para provar (ii), raciocinamos por absurdo: se $\sum b_n$ convergisse, então, pela parte (i), $\sum a_n$ também teria de convergir, contrariando a hipótese. ∎

Exemplo 1.2.10: Um modo de provar a convergência da série $\sum \frac{1}{n!}$ consiste em compará-la com a série geométrica de razão $1/2$. Observemos que

$$\frac{1}{n!} = \frac{1}{2 \cdot 3 \ldots n} < \frac{1}{2 \cdot 2 \ldots 2} = \frac{1}{2^{n-1}},$$

donde se vê que a série dada é dominada pela série $\sum \frac{1}{2^{n-1}}$. Como esta é convergente, concluímos que a série original também é convergente. ∎

Exemplo 1.2.11: Vamos provar que a série $\sum \frac{1}{n^r}$ é convergente quando $r > 1$. Para isso, majoramos as somas parciais da série, diminuindo os denominadores de seus termos, de acordo com o seguinte esquema:

$$\begin{aligned}
&1 + (\frac{1}{2^r} + \frac{1}{3^r}) + (\frac{1}{4^r} + \frac{1}{5^r} + \frac{1}{6^r} + \frac{1}{7^r}) + \ldots \\
&\leq 1 + (\frac{1}{2^r} + \frac{1}{2^r}) + (\frac{1}{4^r} + \frac{1}{4^r} + \frac{1}{4^r} + \frac{1}{4^r}) + \ldots \\
&= 1 + \frac{2}{2^r} + \frac{4}{4^r} + \ldots = 1 + \frac{1}{2^{r-1}} + \frac{1}{4^{r-1}} \\
&= 1 + (\frac{1}{2^{r-1}}) + (\frac{1}{2^{r-1}})^2 + \ldots
\end{aligned}$$

Isto nos mostra que $\sum \frac{1}{n^r}$ é dominada pela série geométrica de razão $q = \frac{1}{2^{r-1}} < 1$, que é convergente. ∎

Exemplo 1.2.12: A série $\sum_{k=1}^{\infty} \frac{1}{k} sen \frac{1}{k}$ é convergente, pois como para todo $x \in (0, \pi/2)$, $sen x < x$, temos que para todo $k \geq 1$:

$$0 \leq \frac{1}{k} sen \frac{1}{k} \leq \frac{1}{k} \cdot \frac{1}{k}.$$

Como $\sum_{k=1}^{\infty} \frac{1}{k^2}$ é convergente, segue do critério da comparação que $\sum_{k=1}^{\infty} \frac{1}{k} sen \frac{1}{k}$ é convergente. ∎

Exemplo 1.2.13: A série $\sum_{k=0}^{\infty}\frac{k}{k^2+2k+1}$ é convergente ou divergente? Justifique.

Solução:

$$\frac{k}{k^2+2k+1}=\frac{1}{k}\cdot\frac{1}{1+\frac{2}{k}+\frac{1}{k^2}}.$$

Para todo $k\geq 1$, $1+\frac{2}{k}+\frac{1}{k^2}\leq 4$ e, portanto, para todo $k\geq 1$,

$$\frac{k}{k^2+2k+1}\geq\frac{1}{4k}.$$

Como $\sum_{k=1}^{\infty}\frac{1}{4k}=\infty$, resulta que $\sum_{k=0}^{\infty}\frac{k}{k^2+2k+1}=\infty$ e, portanto, a série é divergente. ∎

Teste da Razão

Teorema 1.2.14: *Seja $\sum a_n$ uma série de termos positivos tal que a_{n+1}/a_n converge para um certo limite r. Então, a série converge se $r<1$ e diverge se $r>1$.*

Prova: Supondo $r<1$, seja $\epsilon>0$ tal que $c=r+\epsilon<1$. Como $a_{n+1}/a_n\to r$, existe um índice N suficientemente grande tal que, para $n\geq N$,

$$r-\epsilon<a_{n+1}/a_n<r+\epsilon=c.$$

Fazendo n sucessivamente igual a N, $N+1$, $N+2,\ldots$, essa desigualdade nos dá

$$\begin{aligned}&a_{N+1}<a_Nc,\\&a_{N+2}<a_{N+1}c<a_Nc^2,\end{aligned}$$

em geral, $a_{N+n}<a_Nc^n$, de modo que a série $\sum_{n=N+1}^{\infty}a_n$ é dominada pela série geométrica $a_N\sum_{n=1}^{\infty}c^n$. Como $c<1$, esta série converge, logo o mesmo ocorre com a série original, pelo teste de comparação.

Ao contrário, se $r>1$, então, dado $\epsilon=r-1$, a partir de certo índice $n=N$ teremos

$$r-\epsilon<a_{n+1}/a_n<r+\epsilon.$$

Como $r-\epsilon=1$, a primeira desigualdade acima nos dá $a_{n+1}>a_n$ a partir de $n=N$. Portanto, $a_N<a_{N+1}<a_{N+2}<\ldots$ e a série original diverge para ∞. ∎

Exemplo 1.2.15: A série $\sum_{k=0}^{\infty}\frac{2^k}{k!}$ é convergente ou divergente? Justifique.

Solução: Como $a_k=\frac{2^k}{k!}$, temos

$$\frac{a_{k+1}}{a_k}=\frac{\frac{2^{k+1}}{(k+1)!}}{\frac{2^k}{k!}}=\frac{2}{k+1}.$$

Segue que $\lim_k a_{k+1}/a_k=0$, então, pelo critério da razão, a série $\sum_{k=0}^{\infty}\frac{2^k}{k!}$ é convergente. ∎

1.2.2 Convergência Absoluta

Definição 1.2.16: Diz-se que uma série $\sum a_n$ converge absolutamente, ou é absolutamente convergente, se a série $\sum |a_n|$ é convergente. ∎

Teorema 1.2.17: *Toda série absolutamente convergente é convergente. Além disso, a soma da série independe da ordem em que se consideram os termos.*

Prova: Sejam p_n a soma dos termos a_r positivos e q_n a soma dos valores absolutos dos termos a_r negativos, onde, em ambos os casos, $r \leq n$. Então, as somas parciais das séries $\sum |a_n|$ e $\sum a_n$ são dadas por

$$T_n = |a_1| + |a_2| + \ldots + |a_n| = p_n + q_n$$

e

$$S_n = a_1 + a_2 + \ldots + a_n = p_n - q_n,$$

respectivamente. As sequências $(T_n), (p_n)$ e (q_n) são não decrescentes, a primeira delas converge, por hipótese, digamos, para T. Ao mesmo tempo, $p_n \leq T_n \leq T$ e $q_n \leq T_n \leq T$, logo p_n e q_n também convergem, digamos para p e q, respectivamente. Concluímos que (S_n) também converge: $S_n = p_n - q_n \to p - q$.

Para demonstrar a segunda parte do teorema, basta notar que p_n e q_n são somas parciais de séries de termos não negativos, cujas somas independem da ordem em que se considerem seus termos. ∎

Exemplo 1.2.18: A série $-1 + \frac{1}{4} - \frac{1}{9} + \ldots = \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$ é convergente, já que ela é absolutamente convergente. ∎

Teste da Razão Para Séries de Termos Quaisquer

Teorema 1.2.19: *Seja a série $\sum_{k=0}^{\infty} a_k$, com $a_k \neq 0$ para todo natural k. Suponhamos que $\lim_k |\frac{a_{k+1}}{a_k}| = r$. Então, a série converge se $r < 1$ e diverge se $r > 1$.*

Prova: Se $r < 1$, a série $\sum_{k=0}^{\infty} |a_k|$ será convergente pelo teste da razão; logo $\sum_{k=0}^{\infty} a_k$ será, também, convergente.

Se $r > 1$, existirá um natural p tal que $k \geq p \Rightarrow \frac{|a_{k+1}|}{|a_k|} > 1$. Então, para todo $k > p$, $|a_k| > |a_p|$. Como $a_p \neq 0$, $\lim_k |a_k|$ não poderá ser zero e o mesmo acontecerá, então, com $\lim_k a_k$. Pelo critério do termo geral, a série $\sum_{k=0}^{\infty} a_k$ será divergente. ∎

Exemplo 1.2.20: Determine x para que a série $\sum_{n=1}^{\infty} nx^n$ seja convergente.

Solução: Para $x = 0$ a soma da série é zero; logo convergente. Suponhamos então que $x \neq 0$ e apliquemos o critério da razão.

$$\lim_n |\frac{(n+1)x^{n+1}}{nx^n}| = |x| \lim_n \frac{n+1}{n} = |x|.$$

Segue do critério da razão, que a série é convergente para $|x| < 1$ e divergente para $|x| > 1$. Para $|x| = 1$, a série é divergente pelo critério do termo geral. ∎

Exemplo 1.2.21: Determine x para que a série $\sum_{n=1}^{\infty}\frac{x^n}{n!}$ seja convergente.

Solução: Para $x = 0$ a soma da série é zero; logo convergente. Suponhamos então que $x \neq 0$ e apliquemos o critério da razão.

$$\lim_n |\frac{n!x^{n+1}}{(n+1)!x^n}| = |x| \lim_n \frac{1}{n+1} = 0.$$

Segue do critério da razão, que a série é convergente para todo x real. ∎

Exemplo 1.2.22: Determine x para que a série $\sum_{n=1}^{\infty}\frac{n!x^n}{n^n}$ seja convergente.

Solução: Para $x = 0$ a soma da série é zero; logo convergente. Suponhamos então que $x \neq 0$ e apliquemos o critério da razão.

$$\lim_n |\frac{(n+1)!n^n x^{n+1}}{n!(n+1)^{n+1}x^n}| = |x| \lim_n \frac{(n+1)n^n}{(n+1)^{n+1}} = |x| \lim_n (\frac{n}{n+1})^n = \frac{|x|}{e}.$$

Segue do critério da razão, que a série é convergente para todo $|x| < e$ e divergente para $|x| > e$.

Se $|x| = e$, utilizando a aproximação de Stirling, segundo a qual $\lim_{n\to\infty}\frac{n!}{(\frac{n}{e})^n\sqrt{2\pi n}} = 1$, temos que

$$\begin{aligned}
lim_{n\to\infty}|a_n| &= \lim_{n\to\infty}\frac{n!e^n}{n^n}\\
&= \lim_{n\to\infty}\frac{n!e^n}{n^n}\frac{(\frac{n}{e})^n\sqrt{2\pi n}}{(\frac{n}{e})^n\sqrt{2\pi n}}\\
&= \lim_{n\to\infty}\frac{n!}{(\frac{n}{e})^n\sqrt{2\pi n}}\lim_{n\to\infty}\frac{e^n(\frac{n}{e})^n\sqrt{2\pi n}}{n^n}\\
&= 1\lim_{n\to\infty}\sqrt{2\pi n} = \infty.
\end{aligned}$$

Portanto, o termo geral da série diverge, logo a série diverge. ∎

1.2.3 Ordens de Magnitude

Quando duas funções f e g são tais que o quociente $\frac{f(x)}{g(x)}$ tende a zero com x tendendo a um certo x_0, dizemos que f é de ordem pequena em relação a g, para $x \to x_0$ e escrevemos

$$f(x) = o(g(x)), x \to x_0.$$

Por exemplo, $sen^2 x = o(x)$ e $cos(\frac{1}{x}) = o(\frac{1}{x})$ para $x \to 0$, pois ambos quocientes, $\frac{sen^2 x}{x}$ e $\frac{cos(1/x)}{1/x}$ tendem a zero com $x \to 0$.

Quando apenas sabemos que o quociente permanece limitado numa vizinhança de x_0, isto é, quando existem números positivos δ e M tal que se $|x - x_0| < \delta$, então $\frac{|f(x)|}{|g(x)|} \leq M$, dizemos que f é de ordem grande em relação a g, para $x \to x_0$ e escrevemos

$$f(x) = O(g(x)), x \to x_0.$$

Por exemplo, $e^x - 1 - x = O(x^2)$ e $senx - x = O(x^3)$ para $x \to 0$, pois usando L'Hopital, temos que os quocientes $\frac{e^x-1-x}{x^2}$ e $\frac{senx-x}{x^3}$ tendem a $1/2$ e $-1/6$, respectivamente, quando x tende a zero.

Note que $f(x) = o(g(x)) \Rightarrow f(x) = O(g(x)), x \to x_0$, mas a recíproca não é verdadeira.

No caso de sequências de números reais, também podemos analisar o comportamento comparativo de duas sequências $\{a_n\}_{n\geq 1}$ e $\{b_n\}_{n\geq 1}$, quando n tende ao infinito. Dizemos que $a_n = o(b_n)$ se $\lim \frac{a_n}{b_n} = 0$ e dizemos que $a_n = O(b_n)$ se existir um número inteiro positivo n_0 tal que a subsequência de $\frac{|a_n|}{|b_n|}$ que contém todos os termos a partir de n_0 seja limitada. Em particular, temos que se (b_n) for uma sequência constante $b_n = c$, para todo n, então $a_n = o(c)$ se $a_n \to 0$ e $a_n = O(c)$ se (a_n) for uma sequência limitada.

Exemplo 1.2.23:

1. $n^k = o(e^n)$, para todo k.
2. $\log n = o(n^k)$, para todo $k > 0$.
3. $10n^2 + n = O(n^2)$.

▮

1.3 Série de Taylor

As funções polinomiais são as mais simples quando se quer calcular seus valores, derivá-las ou integrá-las. A possibilidade de aproximar funções por polinômios é de suma importância, pois permite obter propriedades das funções em termos de propriedades análogas dos polinômios que as aproximam.

Vamos considerar o problema de aproximar a função f, numa vizinhança de $x = 0$, por um polinômio de grau n:

$$p_n(x) = a_0 + a_1 x + \ldots + a_r x^r + \ldots + a_n x^n.$$

Suponha que f seja derivável em $x = 0$ até ordem n. Observamos que:

$$p_n'(x) = a_1 + 2a_2 x + \ldots + r a_r x^{r-1} + \ldots + n a_n x^{n-1}$$
$$p_n''(x) = 2a_2 + 6a_3 x + \ldots + r(r-1) a_r x^{r-2} + \ldots + n(n-1) a_n x^{n-2}$$

Em geral,

$$p_n^{(r)}(x) = r! a_r + \ldots + n(n-1)\ldots(n-r+1) a_n x^{n-r}.$$

Portanto, fazendo $x = 0$ nessa expressão, obtemos $p_n^{(r)}(0) = r! a_r$, $r = 0, 1, 2, \ldots, n$.

Como queremos aproximar f por p_n em $x = 0$, queremos que todas as derivadas até ordem n dessas funções em $x = 0$ coincidam, ou seja, que elas se toquem $(f(0) = p_n(0))$ no ponto $x = 0$, tenham a mesma inclinação $(f'(0) = p_n'(0))$ neste ponto, e assim por diante. Então, segue-se que

$$a_r = \frac{p_n^{(r)}(0)}{r!} = \frac{f^{(r)}(0)}{r!}.$$

Então, temos que

$$p_n(x) = \sum_{r=0}^{n} \frac{f^{(r)}(0)}{r!} x^r,$$

onde $f^{(0)}(x) = f(x)$.

Este é chamado de *polinômio de Taylor* de ordem n da função f em torno de $x = 0$. Sua importância reside no teorema que enunciamos e provamos a seguir.

Teorema 1.3.1: *Seja f uma função derivável até a ordem $n+1$, numa vizinhança V de $x = 0$. Então, o polinômio p_n aproxima f em V com erro ou resto dado por*

$$R_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(c_n)x^{n+1}}{(n+1)!},$$

onde c_n é um número compreendido entre 0 e x.

Prova: Começaremos enunciando e provando o seguinte Lema, conhecido como Teorema do Valor Médio Generalizado.

Lema 1.3.2: *Se F e G são funções deriváveis num intervalo (a, b), contínuas em $[a, b]$, com $G(a) \neq G(b)$ e $G'(x) \neq 0$ para $x \in (a, b)$, então existe $c \in (a, b)$ tal que*

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(c)}{G'(c)}$$

Prova: Considere a função

$$H(x) = (F(b) - F(a))(G(x) - G(a)) - (G(b) - G(a))(F(x) - F(a)).$$

Então, H é contínua em $[a, b]$, derivável em (a, b), e $H(a) = H(b) = 0$. Logo pelo Teorema do Valor Médio existe $c \in (a, b)$ tal que $H(b) - H(a) = 0 = H'(c)(b - a)$, ou seja, $H'(c) = 0$. Portanto, existe c tal que

$$(F(b) - F(a)G'(c) - (G(b) - G(a))F'(c) = 0.$$

Como $G(b) \neq G(a)$, o resultado está provado. ∎

Usaremos repetidamente o Teorema do Valor Médio Generalizado para provar o teorema. Seja $F(x) = f(x) - p_n(x)$, $G(x) = x^{n+1}$, $a = 0$, e $b = x$. Então aplicando o Lema notando que $f(0) = p_n(0)$, obtemos

$$\frac{R_n(x)}{x^{n+1}} = \frac{f(x) - p_n(x)}{x^{n+1}} = \frac{f'(c) - p'_n(c)}{(n+1)c^n},$$

onde c está entre 0 e x. Aplicando novamente o Lema com $F(x) = f'(x) - p'_n(x)$, $G(x) = (n+1)x^n$, $b = c$ e $a = 0$, temos (note que $f'(0) - p'_n(0) = 0$)

$$\frac{R_n(x)}{x^{n+1}} = \frac{f'(c) - p'_n(c)}{(n+1)c^n} = \frac{f''(c_1) - p''_n(c_1)}{(n+1)nc_1^{n-1}},$$

onde c_1 está entre 0 e c. Continuando desta maneira, levando sempre em conta que $f_n^{(r)}(0) = p_n^{(r)}(0)$, para $0 \leq r \leq n$ e o fato que $p_n^{(n+1)}(y) = 0$ para todo y real, obtemos

$$\frac{R_n(x)}{x^{n+1}} = \frac{f^{(n+1)}(c_n)}{(n+1)!},$$

onde c_n está entre 0 e x. ∎

A fórmula

$$f(x) = \sum_{r=0}^{n} \frac{f^{(r)}(0)}{r!} x^r + R_n(x)$$

é chamada de série, expansão ou desenvolvimento de Taylor de ordem n da função f em torno de $x = 0$.

Podemos generalizar este resultado e obter a série de Taylor de uma função em torno de um outro ponto qualquer $x = a$, onde a não é necessariamente igual a zero. Este problema se reduz facilmente ao problema tratado anteriormente, introduzindo-se a variável $h = x - a$ e a função $g(h) = f(a+h) = f(x)$. Dessa maneira a variável x se aproxima de a se, e somente se, h se aproxima de 0. Suponha que f seja derivável até ordem $n+1$ numa vizinhança de $x = a$, digamos $|x - a| < \delta$. Então, g terá $n+1$ derivadas em $|h| < \delta$. Além disso, $g^{(r)}(h) = f^{(r)}(a+h) = f^{(r)}(x)$, $0 \leq r \leq n+1$. Portanto, a série de Taylor de g de ordem n em torno de $h = 0$ é:

$$g(h) = \sum_{r=0}^{n} \frac{g^{(r)}(0)}{r!} h^r + \frac{g^{(n+1)}(c)}{(n+1)!} h^{n+1},$$

e pode ser reescrita como

$$f(x) = \sum_{r=0}^{n} \frac{f^{(r)}(a)}{r!} (x-a)^r + \frac{f^{(n+1)}(c')}{(n+1)!} (x-a)^{n+1},$$

onde $c' = a + c$ é um número entre a e x, do mesmo modo que c é um número entre 0 e h. Esta fórmula é chamada de série, expansão, ou desenvolvimento de Taylor de ordem n da função f em torno do ponto $x = a$, e $\sum_{r=0}^{n} \frac{f^{(r)}(a)}{r!}(x-a)^r$ é chamado de polinômio de Taylor de ordem n de f em torno de $x = a$.

Se a função $f^{(n+1)}$ for limitada por uma constante K numa vizinhança de $x = a$, isto é, $|f^{n+1}(x)| \leq K$, para $|x - a| \leq \delta$, então $R_n(x) = O((x-a)^{n+1})$ ou $R_n(x) = o((x-a)^n)$ com $x \to a$. Desse modo se uma função f possui derivada de ordem n numa vizinhança de $x = a$ para todo natural n, temos que sua série de Taylor é dada por:

$$\sum_{r=0}^{\infty} \frac{f^{(r)}(a)}{r!} (x-a)^r.$$

Exemplo 1.3.3: Vamos obter a série de Taylor de ordem n da função $f(x) = \ln(1+x)$, $x > -1$, em torno de $x = 0$.

Solução: Note que $f(0) = \ln(1) = 0$, $f'(x) = \frac{1}{1+x} = (1+x)^{-1}$, $f''(x) = -1(1+x)^{-2}$, e que em geral temos: $f^{(r)}(x) = (-1)^{r-1}(r-1)!(1+x)^{-r}$. Portanto, $f^{(r)}(0) = (-1)^{r-1}(r-1)!$, e a série de Taylor de ordem n de f em torno de $x = 0$ é:

$$f(x) = \sum_{r=1}^{n} \frac{(-1)^{r-1}}{r}(x)^r + \frac{(-1)^n(1+c)^{-(n+1)}}{(n+1)}(x)^{n+1},$$

onde c está entre 0 e x. ∎

Exemplo 1.3.4: Vamos obter a série de Taylor de ordem n da função $f(x) = \frac{1}{x}$, $x > 0$, em torno do ponto $a = 2$.

Solução: Note que $f(2) = 1/2$, $f'(x) = -x^2$, $f''(x) = 2x^{-3}$, e que em geral temos: $f^{(r)}(x) = (-1)^r r! x^{-r-1}$. Portanto, $\frac{f^{(r)}(2)}{r!} = \frac{(-1)^r}{2^{r+1}}$, e a série de Taylor de ordem n de f em torno de $x = 2$ é:

$$f(x) = \sum_{r=0}^{n} \frac{(-1)^r}{2^{r+1}}(x-2)^r + \frac{(-1)^{n+1}}{c^{n+2}}(x-2)^{n+1},$$

onde c é um número entre 2 e x. ∎

Exemplo 1.3.5: Fórmula de Euler. Neste exemplo usaremos séries de Taylor para demonstrar a fórmula de Euler: $e^{ix} = \cos(x) + i\,\mathrm{sen}(x)$, onde $i = \sqrt{-1}$. Note que para qualquer inteiro r, temos

$$\frac{d^r e^{ix}}{dx^r} = i^r e^{ix};$$
$$\frac{d^r \cos(x)}{dx^r} = \begin{cases} (-1)^{\frac{r+1}{2}}\,\mathrm{sen}(x) & \text{se } r \text{ for ímpar,} \\ (-1)^{\frac{r}{2}}\cos(x) & \text{se } r \text{ for par;} \end{cases}$$
$$\frac{d^r \,\mathrm{sen}(x)}{dx^r} = \begin{cases} (-1)^{\frac{r-1}{2}}\cos(x) & \text{se } r \text{ for ímpar,} \\ (-1)^{\frac{r}{2}}\,\mathrm{sen}(x) & \text{se } r \text{ for par.} \end{cases}$$

Então, temos as seguintes expansões de Taylor em torno de $x = 0$:

$$e^{ix} = \sum_{r=0}^{\infty} \frac{i^r}{r!}x^r = \sum_{r=0}^{\infty} \frac{(-1)^r}{2r!}x^{2r} + i\sum_{r=0}^{\infty} \frac{(-1)^r}{(2r+1)!}x^{2r+1};$$
$$\cos(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{2r!}x^{2r};$$
$$\mathrm{sen}(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{(2r+1)!}x^{2r+1}.$$

Logo, podemos concluir que $e^{ix} = \cos(x) + i\,\mathrm{sen}(x)$. ∎

Capítulo 2

Convergência Estocástica

2.1 Seqüência de Eventos

A definição de conceitos de convergência de variáveis aleatórias depende de manipulações de seqüências de eventos. Seja $A_n \subseteq \Omega$, define-se:

$$\inf_{k\geq n} A_k = \cap_{k=n}^{\infty} A_k, \ \sup_{k\geq n} A_k = \cup_{k=n}^{\infty} A_k$$
$$\liminf_n A_n = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k$$
$$\limsup_n A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k.$$

O limite de uma seqüência de eventos é definido da seguinte maneira: se para alguma seqüência (B_n) de eventos $\liminf_n B_n = \limsup_n B_n = B$, então B é chamado de limite de (B_n) e nós escrevemos $\lim_n B_n = B$ ou $B_n \to B$.

Exemplo 2.1.1: $\liminf[0, \frac{n}{n+1}) = \limsup[0, \frac{n}{n+1}) = [0, 1)$

Teorema 2.1.2: *Seja (A_n) uma seqüência de eventos de Ω.*

(a) $\omega \in \limsup A_n$ se, e somente se, $\omega \in A_k$ para um número infinito de índices k.

(b) $\omega \in \liminf A_n$ se, e somente se, $\omega \notin A_k$ para um número finito de índices k.

Prova: Para parte (a), note que $\omega \in \limsup A_n$, se, e somente se, para todo n, $\omega \in \cup_{k=n}^{\infty} A_k$, ou seja, se, e somente se, para todo n existe $n' \geq n$ tal que $\omega \in A_{n'}$. Como isto é válido para todo n, temos que isto é equivalente a existência de um número infinito de índices k tais que $\omega \in A_k$.

A prova da parte (b) é similar. ∎

A seguir descreveremos algumas propriedades do lim inf e lim sup de uma seqüência de eventos.

1. $\liminf A_n \subseteq \limsup A_n$

   Este fato é uma simples conseqüência do Teorema 2.1.2, pois se $\omega \in \liminf A_n$, ω não pertence apenas a um número finito de eventos A_k's, e conseqüentemente pertence a um número infinito deles. Logo, $\omega \in \limsup A_n$.

2. $(\liminf A_n)^c = \limsup A_n^c$

   Este fato decorre aplicando a Lei de De Morgan duas vezes:

$$(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty}(\cap_{k=n}^{\infty} A_k)^c = \cap_{n=1}^{\infty}(\cup_{k=n}^{\infty} A_k^c).$$

Seqüências Monotônicas

Uma seqüência de eventos (A_n) é monotônica não-decrescente (resp., não-crescente) se $A_1 \subseteq A_2 \subseteq \ldots$ (resp, $A_1 \supseteq A_2 \supseteq \ldots$). Denotaremos por $A_n \uparrow$ (resp., $A_n \downarrow$) uma seqüência não-decrescente (resp. não-crescente) de eventos.

Teorema 2.1.3: *Suponha que (A_n) é uma seqüência monotônica de eventos. Então,*

1. *Se $A_n \uparrow$, então $\lim_n A_n = \cup_{n=1}^{\infty} A_n$.*
2. *Se $A_n \downarrow$, então $\lim_n A_n = \cap_{n=1}^{\infty} A_n$.*

Conseqüentemente, como para qualquer seqüência B_n, temos $\inf_{k \geq n} B_k \uparrow$ e $\sup_{k \geq n} B_k \downarrow$, segue que:

$$\liminf B_n = \lim_n (\inf_{k \geq n} B_k), \ \limsup B_n = \lim_n (\sup_{k \geq n} B_k)$$

Prova: Para provar (1), precisamos mostrar que $\liminf A_n = \limsup A_n = \cup_{n=1}^{\infty} A_n$. Como $A_j \subseteq A_{j+1}$, temos $\cap_{k \geq n} A_k = A_n$, e portanto,

$$\liminf A_n = \cup_{n=1}^{\infty}(\cap_{k \geq n} A_k) = \cup_{n=1}^{\infty} A_n.$$

Por outro lado, temos,

$$\begin{aligned} \limsup A_n &= \cap_{n=1}^{\infty}(\cup_{k \geq n} A_k) \subseteq \cup_{k=1}^{\infty} A_k \\ &= \liminf A_n \subseteq \limsup A_n. \end{aligned}$$

Logo, temos igualdade acima, ou seja, $\limsup A_n = \cup_{k=1}^{\infty} A_k$.

A prova de (2) é similar. ∎

Exemplo 2.1.4:

1. $\lim_n [0, 1 - \frac{1}{n}] = \cup_{n=1}^{\infty} [0, 1 - \frac{1}{n}] = [0, 1)$.
2. $\lim_n [0, 1 + \frac{1}{n}) = \cap_{n=1}^{\infty} [0, 1 + \frac{1}{n}) = [0, 1]$.
3. $\lim_n (\frac{n}{n+1}, \frac{n}{n-1}) = \cap_{n=1}^{\infty} (\frac{n}{n+1}, \frac{n}{n-1}) = \{1\}$.

Exemplo 2.1.5: Sejam A_n, A, B_n, B eventos em Ω. Mostre que:

1. se $\lim_n A_n = A$, então $\lim_n A_n^c = A^c$.

   Solução: $\liminf A_n^c = (\limsup A_n)^c = A^c$ e $\limsup A_n^c = (\liminf A_n)^c = A^c$.

2. $\limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n$.

   Solução: Se $\omega \in \limsup(A_n \cup B_n)$, então $\omega \in (A_k \cup B_k)$ para infinitos índices k. Logo, temos que $\omega \in A_k$ para infinitos índices k, ou $\omega \in B_k$ para infinitos índices k. Portanto, temos $\omega \in \limsup A_n$ ou $\omega \in \limsup B_n$, ou seja, $\omega \in \limsup A_n \cup \limsup B_n$.

   Reciprocamente, se $\omega \in \limsup A_n \cup \limsup B_n$, então $\omega \in \limsup A_n$ ou $\omega \in \limsup B_n$. Logo, temos que $\omega \in A_k$ para infinitos índices k, ou $\omega \in B_k$ para infinitos índices k, ou seja, $\omega \in (A_k \cup B_k)$ para infinitos índices k. Portanto, $\omega \in \limsup(A_n \cup B_n)$.

3. Não é verdade que $\liminf(A_n \cup B_n) = \liminf A_n \cup \liminf B_n$.

   Solução: Vamos construir um contra-exemplo: Suponha que $A \cap B = \emptyset$, $A_n = A \neq \emptyset$ e $B_n = B \neq \emptyset$ para n par; e $A_n = B$ e $B_n = A$ para n ímpar. Como $A_n \cup B_n = A \cup B$ para todo n, é fácil ver que $\liminf(A_n \cup B_n) = A \cup B$. Também é fácil ver que $\liminf A_n = \liminf B_n = A \cap B = \emptyset$, pois somente os $\omega's$ em $A \cap B$ não ocorrem para um número finito de índices n tanto na seqüência A_n quanto na seqüência B_n. Então, $A \cup B = \liminf(A_n \cup B_n) \neq \emptyset = \liminf A_n \cup \liminf B_n$.

4. se $A_n \to A$ e $B_n \to B$, então $A_n \cup B_n \to A \cup B$ e $A_n \cap B_n \to A \cap B$.

   Solução: Pela parte (2), temos que

   $$\limsup A_n \cup B_n = \limsup A_n \cup \limsup B_n = A \cup B,$$

   e pela propriedade (1) de $\liminf$ e $\limsup$, temos

   $$\liminf A_n \cup B_n \subseteq \limsup A_n \cup B_n = A \cup B.$$

   Resta-nos provar que $A \cup B \subseteq \liminf A_n \cup B_n$. Suponha que $\omega \in A \cup B$, então $\omega \in \liminf A_n$ ou $\omega \in \liminf B_n$, ou seja, ω não pertence a um número finito de A_k's, ou ω não pertence a um número finito de B_k's. Logo, ω não pertence a um número finito de $A_k \cup B_k$'s. Portanto, $\omega \in \liminf A_n \cup B_n$. Então, $A_n \cup B_n \to A \cup B$.

   Utilizando os ítens anteriores e a Lei de De Morgan, temos:

   $$\begin{aligned} A \cap B &= (A^c \cup B^c)^c = (\lim A_n^c \cup \lim B_n^c)^c = \\ &= (\lim A_n^c \cup B_n^c)^c = \lim(A_n^c \cup B_n^c)^c = \lim A_n \cap B_n. \end{aligned}$$

2.1.1 Borel-Canteli

A seguir vamos enunciar e provar um importante Lema, conhecido como Lema de Borel-Cantelli, que trata da probabilidade da ocorrência de um número infinito de eventos.

Lema 2.1.6: *Sejam $A_1, A_2, \ldots$ eventos aleatórios em $(\Omega, \mathcal{A}, P)$, ou seja, $A_n \in \mathcal{A}, \forall n$.*

(a) Se $\sum_{n=1}^{\infty} P(A_n) < \infty$, então $P(A_n$ infinitas vezes $) = 0$.

(b) Se $\sum_{n=1}^{\infty} P(A_n) = \infty$ e os eventos A_n's são independentes, então

$$P(A_n \text{ infinitas vezes }) = 1.$$

Obervação: O ítem (b) não vale necessariamente sem independência. Por exemplo, seja $A_n = A, \forall n$, onde $0 < P(A) < 1$. Então, $\sum P(A_n) = \infty$ mas o evento $[A_n$ infinitas vezes$] = A$ e $P(A_n$ infinitas vezes$) = P(A) < 1$.
Prova: Para parte (a), se $\sum P(A_n) < \infty$, então $\sum_{k=j}^{\infty} P(A_k) \to 0$ quando $j \to \infty$. Mas

$$[A_n \text{ infinitas vezes}] \subseteq \cup_{k=j}^{\infty} A_k, \forall j,$$

logo

$$P(A_n \text{ infinitas vezes}) \leq P(\cup_{k=j}^{\infty} A_k) \leq \sum_{k=j}^{\infty} P(A_k) \to 0.$$

Portanto, $P(A_n$ infinitas vezes$) = 0$.

Para parte (b), basta provar que

$$P(\cup_{k=n}^{\infty} A_k) = 1, \forall n$$

(pois sendo $[A_n$ infinitas vezes$] = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k$ a intersecção de um número enumerável de eventos de probabilidade 1, é também de probabilidade 1). Para tanto, seja $B_n = \cup_{k=n}^{\infty} A_k$. Então B_n contém $\cup_{k=n}^{n+m} A_k$ para todo m, e

$$B_n^c \subseteq (\cup_{k=n}^{n+m} A_k)^c = \cap_{k=n}^{n+m} A_k^c.$$

Logo para todo m,

$$1 - P(B_n) = P(B_n^c) \leq P(\cap_{k=n}^{n+m} A_k^c) = \prod_{k=n}^{n+m} P(A_k^c) = \prod_{k=n}^{n+m} (1 - P(A_k)).$$

Como $1 - p \leq e^{-p}$ para $0 \leq p \leq 1$, temos

$$1 - P(B_n) \leq \prod_{k=n}^{n+m} e^{-P(A_k)} = exp(-\sum_{k=n}^{n+m} P(A_k)) \to 0$$

quando $m \to \infty$, pois $\sum_{k=n}^{n+m} P(A_k) \to \infty$ quando $m \to \infty$. Logo $P(B_n) = 1, \forall n$. ∎

Exemplo 2.1.7: Se sabemos que para uma dada coleção de eventos $\{A_k\}$, as suas probabilidades individuais satisfazem $P(A_k) \leq \frac{1}{k^2}$, então podemos concluir que infitos desses vezes ocorrem com probabilidade zero ou, que apenas um número finito deles ocorrem com probabilidade 1. Podemos reesecrever isso da seguinte forma: existe um instante aleatório N tal que, com probabilidade 1, nenhum dos A_k ocorrem para $k > N$. É importante ressaltar que nós podemos chegar a essa conclusão sem saber nada sobre as interações entre esses eventos como as que são expressas por probabilidades de papres de eventos $P(A_i \cap A_j)$. Contudo, se apenas sabemos que $P(A_k) > 1/k$, então não podemos concluir nada baseados no Lema de Borel-Cantelli. Se soubermos que os eventos são mutuamente independentes, então sabendo que $P(A_k) > 1/k$, podemos concluir que infinitos A_k ocorrem com probabilidade 1.

Exemplo 2.1.8: Considere uma seqüência de variáveis aleatórias $X_1, X_2, X_3, \ldots$. Podemos usar o Lema de Borel-Cantelli para determinar a probabilidade que $X_k > b_k$ infinitas vezes para qualquer seqüência de números reais $\{b_k\}$. Note que $P(X_k > b_k) = 1 - F_{X_k}(b_k)$. Logo, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) < \infty,$$

então, não importa qual a distribuição conjunta das variáveis aleatórias $\{X_k\}$, temos que o evento $\{X_k > b_k\}$ só ocorrerá para um número finito de índices k. Por outro lado, se

$$\sum_{k=1}^{\infty} P(X_k > b_k) = \sum_{k=1}^{\infty} 1 - F_{X_k}(b_k) = \infty,$$

então precisaríamos de informação adicional sobre a distribuição conjunta das variáveis aleatórias $\{X_k\}$ para determinar se os eventos $\{X_k > b_k\}$ ocorrem um número finito ou infinito de vezes.

Exemplo 2.1.9: Considere uma moeda não necessariamente honesta com probabilidade de cara igual a p, onde $0 < p < 1$. Se esta moeda for jogada um número infinito de vezes de maneira independente, qual a probabilidade da seqüência $(cara, cara, coroa, coroa)$ aparecer um número infinito de vezes? Justifique sua resposta.

Solução: Seja X_i o resultado do i-ésimo lançamento da moeda. Defina o evento $A_i = \{X_i = cara, X_{i+1} = cara, X_{i+2} = coroa, X_{i+3} = coroa\}$, queremos calcular $P(A_i$ infinitas vezes$)$. Note que para todo i, temos $P(A_i) = p^2(1-p)^2 > 0$. Não podemos aplicar diretamente o lema de Borel Cantelli, pois os eventos A_i's não são independentes, visto que, por exemplo, ambos A_1 e A_2 dependem de X_2, X_3, X_4. Considere a seguinte subseqüência da seqüência de eventos (A_i) tal que $B_i = A_{4i-3}$. Como os eventos B_i's dependem de famílias disjuntas de variáveis aleatórias independentes, eles são independentes. Além disso temos que $P(B_i) = p^2(1-p)^2 > 0$. Logo, $\sum_i P(B_i) = \infty$. Portanto, Borel-Cantelli implica que $P(B_i$ infinitas vezes$) = 1$. Como (B_i) é uma subseqüência de (A_i), temos que

$$[B_i \text{ infitas vezes}] \subseteq [A_i \text{ infinitas vezes}].$$

Portanto, $P(A_i$ infinitas vezes$) = 1$.

2.2 Covergência de Variáveis Aleatórias

Seguindo uma interpretação freqüentista, probabilidade está relacionada com a freqüência relativa de eventos no longo prazo. A matemática para estudar o longo prazo é a dos limites. Mas quando se trata de funções, existem vários tipos de limites (por exemplo, pontual, uniforme, em quase todo lugar). O mesmo ocorre quando consideramos limites de variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, visto que variáveis aleatórias são funções reais cujo domínio é Ω.

Relembrando: Seja $(\Omega, \mathcal{A})$ um espaço mensurável. Uma função $X : \Omega \to R$ é chamada de variável aleatória se para todo evento Boreliano B, $X^{-1}(B) \in \mathcal{A}$. Nós recordamos que um evento Boreliano é qualquer evento pertencente à σ-álgebra de Borel, onde a σ-álgebra de Borel é a menor σ-álgebra contendo intervalos da forma $(-\infty, x]$ para todo $x \in R$.

2.2.1 Tipos de Convergência

Vamos a seguir descrever vários tipos de convergência estocástica, ilustrando com exemplos cada tipo de convergência, e depois provaremos algumas relações entre os vários tipos de convergência. Sejam $Y, Y_1, Y_2, \ldots$ variáveis aleatórias definidas em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

Convergência Quase Certa

Definição 2.2.1: A seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$ converge *quase certamente (ou com probabilidade 1)* para a variável aleatória Y se

$$P(\{w : \lim_{n\to\infty} Y_n(w) = Y(w)\}) = 1.$$

Notação: $Y_n \to Y$ cp1.

Então se uma seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$ converge quase certamente para Y não significa que para todo $w \in \Omega$, $Y_n(w) \to Y(w)$, apenas o que se sabe é que a probabilidade do evento $D = \{w : Y_n(w) \nrightarrow Y(w)\}$ é nula. D é chamado de conjunto de exceção.

Exemplo 2.2.2: Considere uma variável aleatória Z tal que $P(\{w : 0 \leq |Z(w)| < 1\}) = 1$. Seja $X_n(w) = Z^n(w)$, então $X_n(w) \to 0$ cp1; note que o conjunto de exceção é $D = \{w \in \Omega : |Z(w)| \geq 1\}$ e que $P(D) = 0$. $\Box$

Podemos obter uma definição alternativa para convergência quase-certa, observando que, pela definição de limite de sequências de números reais, para um dado $w \in \Omega$ fixo, temos que $\lim_n Y_n(w) = Y(w)$ se, e somente se, para todo $k \in \mathbb{N}$, existir N tal que para todo $n \geq N$, temos $|Y_n(w) - Y(w)| < \frac{1}{k}$. Portanto:

$$\{w : \lim_n Y_n(w) = Y(w)\} = \{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| < \frac{1}{k}\}.$$

Então, $Y_n \to Y$ cp1 se, e somente se,

$$P(\{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| < \frac{1}{k}\}) = 1.$$

Isto é equivalente a:

$$P(\{w : \cap_{k=1}^{\infty} \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} |Y_n(w) - Y(w)| \geq \frac{1}{k}\}) = 0.$$

Defina $A_{n,k} = \{w : |Y_n(w) - Y(w)| \geq \frac{1}{k}\}$. Então para cada k fixo, temos que

$$\limsup_n A_{n,k} = \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} A_{n,k}.$$

Logo, $Y_n \to Y$ cp1 se, e somente se, para todo $k \in \mathbb{N}$,

$$P(\limsup_n A_{n,k}) = 0.$$

Exemplo 2.2.3: Seja $\{X_n\}_{n\geq 3}$ uma seqüência de variáveis aleatórias independentes com distribuição de probabilidade dada por:

$$P(X_n = 0) = 1 - \frac{1}{\log n} \text{ e } P(X_n = n) = \frac{1}{\log n}, \forall n \geq 3.$$

Mostre que $X_n \nrightarrow 0$ cp1.

Solução: Para qualquer ϵ tal que $0 < \epsilon < 1$, temos que

$$P(|X_n| > \epsilon) = P(X_n = n) = \frac{1}{\log n}.$$

Logo, $\sum_n P(|X_n| > \epsilon) = \sum_n \frac{1}{\log n} = \infty$. Então, o Lema de Borel-Cantelli implica que $P(|X_n| > \epsilon$ infinitas vezes$) = 1$, portanto com probabilidade 1, $X_n \nrightarrow 0$.

Exemplo 2.2.4: Considere $\{X_n : n \geq 1\}$ uma seqüência de variáveis aleatórias i.i.d. com função de distribuição F. Suponha que $F(x) < 1$, para todo $x < \infty$. Defina $Y_n = \max(X_1, X_2, \ldots, X_n)$. Vamos verificar que $Y_n \to \infty$ cp1.

Inicialmente, observe que para cada $\omega \in \Omega$, as variáveis Y_n formam uma seqüência não-decrescente de números reais. Seja M um número real, temos

$$\begin{aligned}
&P(Y_n \leq M : n = 1, 2, \ldots) \leq P(Y_n \leq M : n = 1, 2, \ldots, k) = P(Y_k \leq M) \\
&= P(\max(X_1, X_2, \ldots, X_k) \leq M) = P(X_1 \leq M, X_2 \leq M, \ldots X_k \leq M) \\
&= \prod_{n=1}^{k} P(X_n \leq M) = F^k(M), \forall k \geq 1.
\end{aligned}$$

Fazendo $k \to \infty$, temos que para todo M finito,

$$P(\lim_n Y_n \leq M) = P(Y_n \leq M : n = 1, 2, \ldots) = 0;$$

pois $F^k(M)$ tende a zero, quando $k \to \infty$. Dessa forma, o conjunto dos $w \in \Omega$, em que $\lim_n Y_n(w)$ é finito, tem probabilidade zero e, portanto, $Y_n \to \infty$ cp1.

Convergência na r-ésima Média

Definição 2.2.5: A seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$ converge *na r-ésima Média*, onde $r > 0$, para a variável aleatória Y se

$$\lim_{n\to\infty} E|Y_n - Y|^r = 0.$$

Notação: $Y_n \to^r Y$.

Se $r = 2$ este tipo de convergência é freqüentemente chamado de *convergência em média quadrática.*

Exemplo 2.2.6: Sejam $Z, X_1, X_2, \ldots$ variáveis aleatórias tais que

$$X_n = \frac{n}{n+1}Z,$$

então $X_n \to^2 Z$ se $EZ^2 < \infty$, mas não em caso contrário. $\Box$

Exemplo 2.2.7: Considere a seqüência de variáveis aleatórias definidas no Exemplo 2.2.3. Mostre que $X_n \nrightarrow^r 0$, para todo $r > 0$.

Solução: Temos que

$$E|X_n|^r = n^r P(X_n = n) = \frac{n^r}{\log n} \to \infty.$$

Logo, $X_n \nrightarrow^r 0$.

Pode-se provar que se $X_n \to^r X$, então $X_n \to^s X$ para $s < r$.

Convergência em Probabilidade

Definição 2.2.8: A seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$ converge *em probabilidade* para a variável aleatória Y se $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(\{w : |Y_n(w) - Y(w)| > \epsilon\}) = 0.$$

Notação: $Y_n \to^P Y$.

A intuição por trás desta definição é que para n muito grande a probabilidade de que Y_n e Y sejam bem próximas é bastante alta.

Exemplo 2.2.9: Considere a seqüência de variáveis aleatórias definidas no Exemplo 2.2.3. Mostre que $X_n \to^P 0$. **Solução:** Temos que para $0 < \epsilon < 1$, $P(|X_n| > \epsilon) = P(X_n = n)$ e para $\epsilon \geq 1$, $P(|X_n| > \epsilon) \leq P(X_n = n)$. Como $P(X_n = n) = \frac{1}{\log n} \to 0.$, temos que $\forall \epsilon > 0$, $\lim P(|X_n| > \epsilon) = 0$. Portanto, $X_n \to^P 0$.

Exemplo 2.2.10: Considere $X, X_1, X_2, \ldots$ onde as varáveis aleatórias têm distribuição normal conjunta, todas com média 0 e matriz de covariância parcialmente descrita por

$$COV(X, X) = COV(X_n, X_n) = 1, COV(X, X_n) = 1 - \frac{1}{n}.$$

Seja $Y_n = X_n - X$, como Y_n é uma combinação linear de variáveis aleatórias com distribuição normal, ela também possui distribuição normal. Precisamos determinar então sua média e sua variância. Mas $EY = E(X_n - X) = EX_n - EX = 0$ e

$$VarY = EY^2 = E(X_n - X)^2 = EX_n^2 - 2EX_nX + EX^2 = 1 - 2(1 - \frac{1}{n}) + 1 = \frac{2}{n}.$$

Portanto, $Y_n \sim \mathcal{N}(0, \frac{2}{n})$. Então,

$$P(|X_n - X| > \epsilon) = P(|Y_n| > \epsilon) = 2P(Y_n > \epsilon) = 2\int_\epsilon^\infty \frac{\sqrt{n}}{\sqrt{4\pi}} e^{-\frac{ny^2}{4}} dy = 2\int_{\epsilon\sqrt{\frac{n}{2}}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Logo, $\forall \epsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$, ou seja, $X_n \to^P X$. $\Box$

Convergência em Distribuição

O último tipo de convergência estocástico que mencionamos não é exatamente uma noção de convergência das variáveis aleatórias propriamente ditas, mas uma noção de convergência de suas respectivas funções de distribuição acumuladas.

Definição 2.2.11: A seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$, converge *em distribuição* para a variável aleatória Y se para todo ponto x de continuidade de F_Y

$$\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x).$$

Notação: $Y_n \to^D Y$.

Exemplo 2.2.12: Seja $\{X_n : n \geq 1\}$ uma seqüência de variáveis aleatórias independentes com distribuição Uniforme em $(0, b)$, $b > 0$. Defina $Y_n = \max(X_1, X_2, \ldots, X_n)$ e $Y = b$. Vamos verificar que $Y_n \to^D Y$. Temos

$$F_{Y_n}(y) = P(\max(X_1, X_2, \ldots, X_n) \leq y) = F^n_{X_1}(y) = \begin{cases} 0 & \text{se } y < 0, \\ (\frac{y}{b})^n & \text{se } 0 \leq y < b, \\ 1 & \text{se } y \geq b. \end{cases}$$

Fazendo n tender ao infinito, temos que

$$\lim_n F_{Y_n}(y) = \begin{cases} 0 & \text{se } y < b, \\ 1 & \text{se } y \geq b, \end{cases}$$

que corresponde à função de distribuição de Y e, portanto, $Y_n \to^D Y$.

Deve-se ficar atento que convergência em distribuição não implica nada em relação aos outros tipos de convergência. Uma seqüência convergindo em distribuição para uma variável aleatória X também converge em distribuição para qualquer outra variável aleatória Y tal que $F_Y = F_X$. O próximo exemplo serve para ilustrar melhor este fato.

Exemplo 2.2.13: Se uma seqüência de variáveis aleatórias $Y_1, Y_2, \ldots$ é independente e identicamente distribuída de acordo com F, então para todo n tem-se que $F_{Y_n} = F$, logo a seqüência converge em distribuição para qualquer variável aleatória X tal que $F_X = F$. Claro, como a seqüência é independente, os valores de termos sucessivos são independentes e não exibem nenhum comportamento usual de convergência. □

O requisito de continuidade, mencionado na definição acima, se justifica para evitar algumas anomalias. Por exemplo, para $n \geq 1$ seja $X_n = \frac{1}{n}$ e $X = 0$, para todo Ω. Parece aceitável que deveríamos ter convergência de X_n para X, qualquer que fosse o modo de convergência. Observe que

$$F_n(x) = \begin{cases} 0 & \text{se } x < \frac{1}{n}, \\ 1 & \text{se } x \geq \frac{1}{n}, \text{ e} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ 1 & \text{se } x \geq 0. \end{cases}$$

Portanto, como $\lim_n F_n(0) = 0 \neq F(0) = 1$, não temos $\lim_n F_n(x) = F(x)$ para todo $x \in I\!R$. Desse modo se houvesse a exigência de convergência em todos os pontos, não teríamos convergência em distribuição. Entretanto, note que para $x \neq 0$, temos $\lim_n F_n(x) = F(x)$ e, como o ponto 0 não é de continuidade de F, concluímos que $X_n \rightarrow^D X$.

Um exemplo mais complexo de convergência em distribuição pode ser visto na análise do limite de

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i),$$

onde X_i's são variáveis aleatórias independentes e identicamente distribuídas. Neste, o Teorema Central do Limite afirma que se $VAR(X_i) = \sigma^2 < \infty$, então S_n converge em distribuição para qualquer variável aleatória com distribuição $\mathcal{N}(0, \sigma^2)$.

O próximo teorema estabelece duas condições suficientes para que uma seqüência de variáveis aleatórias convirja em distribuição.

Teorema 2.2.14: *Seja $X, X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias:*

(a) Se $X, X_1, X_2, \ldots$ são variáveis aleatórias discretas com $P(X_n = x_i) = p_n(i)$ e $P(X = x_i) = p(i)$, onde $p_n(i) \rightarrow p(i)$ quando $n \rightarrow \infty$ para todo $i = 0, 1, 2, 3, \ldots$, então $X_n \rightarrow^D X$.

(b) Se $X, X_1, X_2, \ldots$ são variáveis aleatórias absolutamente contínuas com densidades dadas respectivamente por $f, f_1, f_2, f_3, \ldots$, onde $f_n(x) \rightarrow f(x)$ quando $n \rightarrow \infty$ em quase todo lugar, então $X_n \rightarrow^D X$.

Prova: Fora do escopo deste curso. ∎

O próximo exemplo mostra que se uma seqüência de variáveis aleatórias discretas converge em distribuição, não necessariamente sua função probabilidade de massa converge.

Exemplo 2.2.15: Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias tais que $P(X = 0) = 1$ e $P(X_n = 1/n) = 1$. Então, temos $F_X(x) = 1$ se $x \geq 0$, e $F_X(x) = 0$ caso contrário; e $F_{X_n}(x) = 1$ se $x \geq 1/n$ e $F_{X_n}(x) = 0$ caso contrário. Logo, $F_{X_n}(x) \rightarrow F_X(x), \forall x \neq 0$, ou seja, $X_n \rightarrow^D X$. Porém, $p(0) = 1 \neq 0 = \lim_n p_n(0)$.

O próximo exemplo mostra que se uma seqüência de variáveis aleatórias absolutamente contínuas converge em distribuição, não necessariamente sua função densidade de probabilidade converge.

Exemplo 2.2.16: Considere uma seqüência de variáveis aleatórias $X, X_1, X_2, \ldots$ com função de distribuição acumuladas dadas respectivamente por $F, F_1, F_2, F_3, \ldots$, onde

$$F_n(x) = \begin{cases} 0 & \text{, se } x \leq 0 \\ x(1 - \frac{\text{sen} 2n\pi x}{2n\pi x}) & \text{, se } 0 < x \leq 1 \\ 1 & \text{, se } x > 1; \end{cases}$$

e

$$F(x) = \begin{cases} 0 & \text{, se } x \leq 0 \\ x & \text{, se } 0 < x \leq 1 \\ 1 & \text{, se } x > 1. \end{cases}$$

Então F_n e F são absolutamente contínuas com densidade dada por

$$f_n(x) = \begin{cases} 1 - \cos 2n\pi x & \text{, se } 0 \leq x \leq 1 \\ 0 & \text{, caso contrário;} \end{cases}$$

e

$$f(x) = \begin{cases} 1 & \text{, se } 0 < x \leq 1 \\ 0 & \text{, caso contrário.} \end{cases}$$

É fácil ver que $F_n(x) \rightarrow F(x), \forall x \in I\!R$. Contudo, $f_n(x) \nrightarrow f(x)$.

2.2.2 Relação Entre os Tipos de Convergência

A primeira relação que iremos provar é que convergência quase certa implica convergência em probabilidade.

Teorema 2.2.17: $X_n \rightarrow X$ *cp1* $\Rightarrow X_n \rightarrow^P X$.

Prova: Para provar que convergência quase certa implica em convergência em probabilidade, considere a seguinte família de eventos

$$A_{n,\epsilon} = \{w : |X_n(w) - X(w)| \leq \epsilon\}.$$

Logo, pela interpretação de convergência pontual,

$$C = \{w : X_n(w) \rightarrow X(w)\} = \cap_{\epsilon>0} \cup_{N=1}^{\infty} \cap_{n \geq N} A_{n,\epsilon}.$$

Se $X_n \rightarrow X$ cp1, então $P(C) = 1$. Equivalentemente, pela Lei de De Morgan,

$$D = C^c = \cup_{\epsilon>0} D_\epsilon, \text{ onde } D_\epsilon = \cap_{N=1}^{\infty} \cup_{n \geq N} A_{n,\epsilon}^c,$$

e

$$P(\cup_{\epsilon>0} D_\epsilon) = 0.$$

Portanto, convergência quase certa implica que $\forall \epsilon > 0$, $P(D_\epsilon) = 0$. Seja $F_N = \cup_{n \geq N} B_n$. Note que $F_N \downarrow$. Logo, $\lim_N F_N = \cap_{N=1}^{\infty} \cup_{n \geq N} B_n$. Portanto, pelo axioma da continuidade monotônica da probabilidade, tem-se que

$$P(\cap_{N=1}^{\infty} \cup_{n \geq N} B_n) = \lim_{N \rightarrow \infty} P(\cup_{n \geq N} B_n).$$

Então,

$$0 = P(D_\epsilon) = \lim_{N \rightarrow \infty} P(\cup_{n \geq N} A_{n,\epsilon}^c) \geq$$
$$\lim_{N \rightarrow \infty} P(A_{N,\epsilon}^c) = \lim_{N \rightarrow \infty} P(|X_N(w) - X(w)| > \epsilon).$$

Portanto, $X_n \rightarrow^P X$. ∎

O próximo teorema prova que convergência na r-ésima média implica convergência em probabilidade.

Teorema 2.2.18: *$X_n \to^r X \Rightarrow X_n \to^P X$.*

Prova: Primeiro note que $\frac{|X_n-X|^r}{\epsilon^r} \geq I_{\{w:|X_n-X|>\epsilon\}}$. Logo, tem-se que

$$E(\frac{|X_n - X|^r}{\epsilon^r}) \geq E(I_{\{w:|X_n-X|>\epsilon\}}),$$

ou seja,

$$\frac{E(|X_n - X|^r)}{\epsilon^r} \geq P(\{w : |X_n - X| > \epsilon\}).$$

Se $X_n \to^r X$, tem-se que $\lim_{n\to\infty} E(|X_n - x|^r) = 0$. Então, para todo $\epsilon > 0$

$$\lim_{n\to\infty} P(\{w : |X_n - X| > \epsilon\}) = 0,$$

ou seja, $X_n \to^P X$. ∎

O próximo exemplo prova que nem convergência em probabilidade, nem convergência na r-ésima média implicam convergência quase certa.

Exemplo 2.2.19: Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$, e considere a seqüência de intervalos definida por

$$I_{2^m+i} = [\frac{i}{2^m}, \frac{i+1}{2^m}],$$

para $m = 0, 1, 2, \ldots$ e $i = 0, 1, \ldots, 2^m - 1$.

Note que tem-se 2^m intervalos de comprimento 2^{-m} que cobrem todo o intervalo $[0, 1]$, e o comprimento dos intervalos fica cada vez menor tendendo a 0. Definamos

$$Y_n(w) = \begin{cases} 1 & \text{se } X(w) \in I_n, \\ 0 & \text{se } X(w) \notin I_n. \end{cases}$$

A seqüência $Y_1, Y_2, \ldots$ converge em probabilidade para 0, pois para $0 < \epsilon \leq 1$,

$$P(|Y_n| \geq \epsilon) = P(Y_n = 1) = P(X \in I_n),$$

e esta probabilidade, que é igual ao comprimento de I_n, converge para zero quando $n \to \infty$.

Esta seqüência também converge na r-ésima média para todo $r > 0$, visto que $E(|Y_n|^r) = P(Y_n = 1) \to 0$ quando $n \to \infty$. Logo, Y_n converge na r-ésima média para 0.

Porém para todo $w \in \Omega$, $Y_n(w) = 1$ para um número infinito de n's e $Y_n(w) = 0$ para um número infinito de n's. Portanto, $Y_n(w)$ não converge para todo w, o que implica que Y_n não converge quase certamente. □

O próximo teorema estabelece mais uma relação entre convergência quase certa e convergência em probabilidade.

Teorema 2.2.20: *$X_n \to^P X$ se, e somente se, toda subseqüência $\{X_{n_k}\}$ possui uma outra subseqüência $\{X_{n_{k(i)}}\}$ tal que $X_{n_{k(i)}} \to X$ cp1 para $i \to \infty$.*

Prova: Suponha que $X_n \to^P X$, então dada qualquer subseqüência $\{X_{n_k}\}$, escolha uma outra subseqüência $\{X_{n_{k(i)}}\}$ tal que $j \geq k(i)$ implica que $P(|X_{n_j} - X| \geq i^{-1}) < 2^{-i}$. Em particular, temos que $P(|X_{n_{k(i)}} - X| \geq i^{-1}) < 2^{-i}$. Seja $A_i = \{|X_{n_{k(i)}} - X| \geq i^{-1}\}$, então $\sum_{i=1}^{\infty} P(A_i) < \sum_{i=1}^{\infty} 2^{-i} = 1 < \infty$. Logo, pelo Lema de Borel-Cantelli, temos que $P(A_i$ infinitas vezes$) = 0$, ou seja, $P(A_i$ finitas vezes$) = 1$. Portanto, $|X_{n_{k(i)}} - X| < i^{-1}$ exceto para um número finito de i's com probabilidade 1. Portanto, $X_{n_{k(i)}} \to X$ cp1.

Se X_n não converge para X em probabilidade, existe um $\epsilon > 0$ e uma subseqüência $\{X_{n_k}\}$ tal que $P(|X_{n_k} - X| > \epsilon) > \epsilon$. Logo nenhuma subseqüência de $\{X_{n_k}\}$ pode convergir para X em probabilidade, logo pelo Teorema 2.2.17, nenhuma subseqüência converge para X quase certamente. ∎

O próximo exemplo mostra que convergência em probabilidade não implica convergência na r-ésima média

Exemplo 2.2.21: Seja X uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$. Considere a seguinte seqüência de varáveis aleatórias

$$Y_n(w) = \begin{cases} 2^n & \text{se } X(w) \in (0, \frac{1}{n}), \\ 0 & \text{se } X(w) \notin (0, \frac{1}{n}). \end{cases}$$

Então, $P(|Y_n| > \epsilon) = P(X(w) \in (0, \frac{1}{n})) = \frac{1}{n} \to 0$, mas $E(|Y_n|^r) = 2^{nr}\frac{1}{n} \to \infty$.

O próximo teorema trata da relação entre convergência em distribuição e convergência em probabilidade.

Teorema 2.2.22: *As seguintes relações entre os tipos de convergência são válidas:*

(a) $X_n \to^P X \Rightarrow X_n \to^D X$

(b) Se $X_n \to^D c$, *onde* c *é uma constante, então* $X_n \to^P c$.

Prova: Para parte (a), suponha que $X_n \to^P X$ e seja x um ponto de continuidade de F_X. Queremos provar que $F_{X_n}(x) \to F_X(x)$ quando $n \to \infty$.

Como para $\epsilon > 0$, $X_n \leq x \Rightarrow X \leq x + \epsilon$ ou $|X - X_n| > \epsilon$, temos $\{w : X_n(w) \leq x\} \subseteq \{w : X(w) \leq x + \epsilon\} \cup \{w : |X_n(w) - X(w)| > \epsilon\}$. Logo,

$$F_{X_n}(x) = P(X_n \leq x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Por outro lado, $X \leq x - \epsilon \Rightarrow X_n \leq x$ ou $|X_n - X| > \epsilon$ de modo que

$$F_X(x - \epsilon) \leq F_{X_n}(x) + P(|X_n - X| > \epsilon).$$

Juntando as duas desigualdades, temos que $\forall \epsilon > 0$, and $\forall n$,

$$F_X(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq F_X(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Como $X_n \to^P X$, para qualquer $\delta > 0$, existe N tal que para $n \geq N$, temos que

$$F_X(x - \epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \delta.$$

Finalmente, como x é ponto de continuidade de F_X, para ϵ suficientemente pequeno, temos que
$$F_X(x) - 2\delta \leq F_X(x-\epsilon) - \delta \leq F_{X_n}(x) \leq F_X(x+\epsilon) + \delta \leq F_X(x) + 2\delta.$$
Ou seja, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.

Para parte (b), suponha que $X_n \to^D c$. Note que a função de distribuição de uma variável aleatória constante c é:
$$F_c(x) = \begin{cases} 1 & \text{se } x \geq c, \\ 0 & \text{se } x < c. \end{cases}$$

Pela convergência em distribuição, tem-se que $\lim_{n\to\infty} F_{X_n}(x) = 0$, se $x < c$ e $\lim_{n\to\infty} F_{X_n}(x) = 1$, se $x > c$. Logo, para $\epsilon > 0$,
$$P(|X_n - c| \leq \epsilon) = P(c-\epsilon \leq X_n \leq c+\epsilon) \geq P(c-\epsilon < X_n \leq c+\epsilon) =$$
$$F_{X_n}(c+\epsilon) - F_{X_n}(c-\epsilon) \to 1 \text{ quando } n \to \infty.$$

Ou seja, $\forall \epsilon > 0$, $\lim_{n\to\infty} P(|X_n - c| > \epsilon) = 0$. ∎

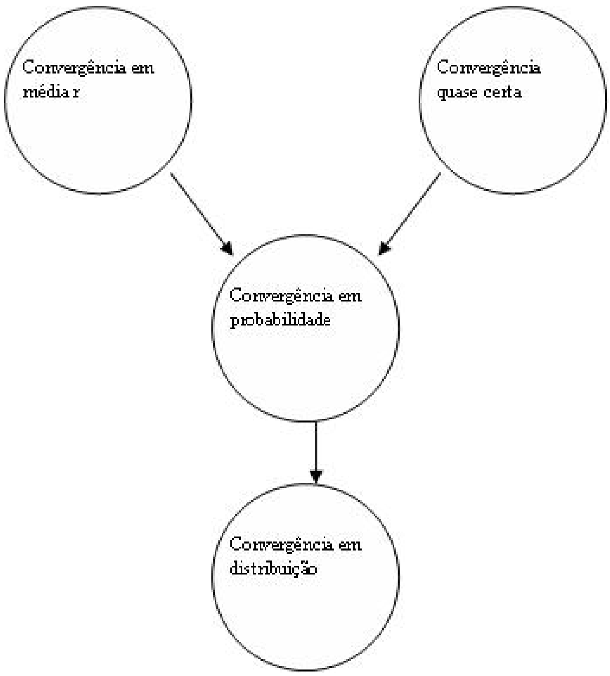

Figura 2.1: Relação entre os tipos de convergência.

A Figura 2.1 resume a relação entre os tipos de convergência.

Exemplo 2.2.23: Para $n \geq 1$, $X_n \sim U(0,1)$ são variáveis aleatórias i.i.d. Defina $Y_n = \min(X_1, X_2, \ldots, X_n)$ e $U_n = nY_n$. Mostre que

(a) $Y_n \to^P 0$,

(b) $U_n \to^D U$, sendo $U \sim Exp(1)$.

Solução: Para parte (a), note que

$$P(|Y_n| > \epsilon) = P(Y_n > \epsilon) = P(X_1 > \epsilon, X_2 > \epsilon, \ldots, X_n > \epsilon).$$

Como os X_n são independentes temos que a última expressão é igual a

$$(P(X_1 > \epsilon))^n = (1-\epsilon)^n.$$

Como $(1-\epsilon)^n \to 0$ quando $n \to \infty$, temos que $Y_n \to^P 0$.

Para parte (b), note que

$$F_{U_n}(x) = P(U_n \leq x) = 1 - P(U_n > x) = 1 - P(nY_n > x) = 1 - P(Y_n > x/n)$$

De acordo com a parte (a), esta expressão é igual a $1-(1-x/n)^n$, que por sua vez converge para $1-e^{-x}$ quando $n \to \infty$, que é igual a $F_U(x)$.

2.3 Convergência de Vetores Aleatórios

Para o caso vetorial as definições de convergência sofrem algumas adaptações. Para as convergências quase certa e em probabilidade, precisamos avaliar a proximidade entre os vetores aleatórios X_n e X pelo comportamento da norma da diferença entre eles. Em geral, essa norma é calculada por $||X_n - X|| = (\sum_{j=1}^{k}(X_{nj} - X_j)^2)^{1/2}$, onde k é a dimensão dos vetores e X_{nj} a coordenada j do vetor X_n. Pode-se verificar que a convergência do vetor aleatório, quase certamente ou em probabilidade, ocorre se, e somente se, existir a mesma convergência em cada uma das variáveis que compõe o vetor aleatório. Dessa forma, o caso multidimensional pode ser estudado a partir de repetidas aplicações do caso univariado.

Para convergência em distribuição de vetores aleatórios, requeremos que a função de distribuição conjunta $F_n(x)$ convirja para $F(x)$, em todos os pontos de continuidade da função F. Entretanto, lembremos que da função de distribuição conjunta podemos obter as marginais, mas o caminho inverso nem sempre é possível. Por essa razão, diferentemente das convergências quase certa e em probabilidade, não podemos reduzir o estudo da convergência em distribuição de vetores aleatórios, ao comportamento das suas respectivas coordenadas. Não temos equivalência, mas apenas implicação, em uma das direções. Ou seja, se o vetor converge em distribuição então cada componente também converge em distribuição, para a correspondente marginal da função de distribuição limite. Entretanto a recíproca não é em geral, verdadeira.

Capítulo 3

Funções Características

3.1 Motivação

Em matemática e suas aplicações, é sempre valioso ter maneiras alternativas de representar o mesmo objeto matemático. Uma analogia pode ser que um conjunto de vetores pode ser representado em vários sistemas de coordenadas. No nosso caso de probabilidade, o conceito básico é o de uma medida de probabilidade P que dá um valor real numérico a um conjunto de eventos em uma σ-álgebra. Para X uma variável aleatória, sabe-se que existem outras maneiras de representar a probabilidade P, como por exemplo através de sua função de distribuição acumulada F_X. Se X for uma variável aleatória discreta, pode-se equivalentemente representar P pela função de probabilidade de X, p_X. Se X for absolutamente contínua, então P pode ser representada pela função densidade de probabilidade de X, f_X. Uma função característica φ_X de uma variável aleatória X é uma outra maneira de representar P. Algumas vantagens do uso da função característica são: pode-se calcular os momentos de uma variável aleatória X diferenciando-se a função característica (o que geralmente é mais simples que usar diretamente as definições de momento que envolvem integrais), pode-se calcular mais facilmente a distribuição de soma de variáveis aleatórias independentes, e finalmente o uso de funções características ajuda na prova de uma família de Teoremas Centrais do Limite que ajudam a explicar a prevalência de distribuições normal ou Gaussianas na Natureza.

Uma função geratriz de momento é uma outra representação alternativa da distribuição de uma variável aleatória. As vantagens desta representação são as mesmas da função característica, mas como a função característica é mais robusta (no sentido que ela sempre existe), nós focaremos no uso da mesma, e apenas no final deste capítulo mencionaremos a definição de uma função geratriz de momento.

3.2 Definição

Definição 3.2.1: A *função característica* φ_X de uma variável aleatória X é dada por:

$$\varphi_X(t) = Ee^{itX} = E\cos(tX) + iE\,\mathrm{sen}(tX), \text{ onde } i \doteq \sqrt{-1}.$$

Note que como $\cos(tX)$ e $\text{sen}(tX)$ são variáveis aleatórias limitadas, a esperança na definição acima é finita e, conseqüentemente, a função característica de qualquer variável aleatória é bem definida. Note também que de acordo com esta definição, a função de distribuição acumulada determina a função característica de uma variável aleatória.

No caso particular de uma variável aleatória discreta, temos:

$$\varphi_X(t) = \sum_k e^{itx_k} p(x_k),$$

onde $p(x_k)$ é a função probabilidade de X.

Analogamente, se X for uma variável aleatória contínua, temos:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx,$$

onde $f_X(x)$ é a função densidade de probabilidade de X.

Observação 3.2.2: A função característica de uma variável aleatória contínua é a transformada de Fourier da densidade de probabilidade de X. ∎

3.2.1 Propriedades

Antes de enunciarmos e provarmos algumas propriedades da função característica, vamos enunciar dois teoremas importantes que tratam da convergência de esperanças de variáveis aleatórias.

Teorema 3.2.3: Teorema da Convergência Monótona. *Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias. Se $0 \leq X_n \uparrow X$, então, $EX_n \uparrow EX$.*

Teorema 3.2.4: Teorema da Convergência Dominada. *Sejam $Y, X, X_1, X_2, \ldots$ variáveis aleatórias. Considere que Y seja integrável, $|X_n| \leq Y$ e $X_n \to X$. Assim X e X_n são integráveis e $EX_n \to EX$.*

O próximo exemplo mostra que nem sempre $X_n \to X \Rightarrow EX_n \to EX$.

Exemplo 3.2.5: Seja $Y \sim U(0,1)$. Considere a seguinte seqüência $\{X_1, X_2, \ldots\}$ de variáveis aleatórias: $X_n(\omega) = n$ se $Y(\omega) \in (0, 1/n)$ e $X_n(\omega) = 0$ em caso contrário. Então, temos que $X_n(\omega) \to 0, \forall\omega$. Mas, $EX_n = 1 \neq 0 = E0$, ou seja, $EX_n \nrightarrow 0$.

A seguir listamos algumas propriedades da função característica.

P1. A função característica é limitada por 1: $|\varphi_X(t)| \leq 1, \forall t \in R$.

Prova: Como pela desigualdade de Jensen, $E^2\cos(tx) \leq E\cos^2(tx)$ e $E^2\,\text{sen}(tx) \leq E\,\text{sen}^2(tx)$, temos

$$|\varphi_X(t)| = \sqrt{E^2\cos(tX) + E^2\,\text{sen}(tX)} \leq \sqrt{E(\cos^2(tX) + \text{sen}^2(tX))} = E1 = 1.$$

∎

P2. A função característica assume o valor 1 no ponto 0: $\varphi_X(0) = 1$.

Prova: $\varphi_X(0) = Ee^{i0X} = E1 = 1$. ∎

P3. $\overline{\varphi_X(t)} = \varphi_X(-t)$, onde $\bar{c}$ é o complexo conjugado de c. (Se $c = x + iy$, o seu complexo conjugado é $\bar{c} = x - iy$.)

Prova: $\varphi_X(-t) = E\cos(-tX) + iE\,\text{sen}(-tX) = E\cos(tX) - iE\,\text{sen}(tX) = \overline{\varphi_X(t)}$. ∎

P4. φ_X é uniformemente contínua na reta.

Prova: Uma função φ é uniformemente contínua, se para todo $\epsilon > 0$ existe $\delta > 0$ tal que para todo $t, s \in R$ $|\varphi(t) - \varphi(s)| < \epsilon$ quando $|t - s| < \delta$. Logo,

$$|\varphi(t) - \varphi(s)| = |E(e^{itx} - e^{isx})| \leq E|e^{isx}(e^{i(t-s)x} - 1)| = E|e^{i(t-s)x} - 1|.$$

Seja $h(u) = |e^{iux} - 1|$. Como $0 \leq |e^{iux} - 1| \leq 2$, 2 é integrável, e $\lim_{u\to 0} h(u) = 0$, pelo teorema da convergência dominada, temos que $\lim_{u\to 0} Eh(u) = 0$. Então, para todo $\epsilon > 0$ existe $\delta > 0$ tal que $|u| < \delta$ implica que $Eh(u) < \epsilon$, ou seja, para todo $\epsilon > 0$ existe $\delta > 0$ tal que $|t - s| < \delta$ implica que $|\varphi(t) - \varphi(s)| \leq E|e^{i(t-s)x} - 1| < \epsilon$. ∎

P5. Se X e Y são independentes, então $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t), \forall t \in R$.

Prova: $\varphi_{X+Y}(t) = Ee^{it(X+Y)} = E(e^{itX}e^{itY}) = E(e^{itX})E(e^{itY}) = \varphi_X(t) \cdot \varphi_Y(t)$. ∎

É fácil provar por indução que se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, então $\varphi_{X_1+\ldots+X_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t), \forall t \in R$.

P6. A função característica de uma variável aleatória determina a função de distribuição acumulada.

Esta propriedade decorre da fórmula da inversão: seja X uma variável aleatória F_X sua função de distribuição acumulada, φ_X sua função característica. Se x e y são pontos de continuidade de F_X tais que $x < y$, então

$$F_X(y) - F_X(x) = \frac{1}{2\pi} \lim_{u\to\infty} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it} \varphi_X(t)dt.$$

Em particular se $y = z + h$ e $x = z - h$, temos que

$$F_X(z + h) - F_X(z - h) = \frac{1}{\pi} \lim_{u\to\infty} \int_{-u}^{u} \frac{\text{sen}ht}{t} e^{-itz} \varphi_X(t)dt.$$

A prova da fórmula da inversão é longa e será omitida.

φ_X determina F_X é o teorema da unicidade que é um corolário da fórmula da inversão, pois esta implica que para todo $z \in R$,

$$F_X(z) = \lim_{y\downarrow z} \lim_{x\to-\infty} \lim_{u\to\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-itx} - e^{-ity}}{it} \varphi_X(t)dt$$

Podemos observar que se X for absolutamente contínua, temos que φ_X determina f_X:

$$f_X(x) = \lim_{h\to 0} \frac{F_X(x+h) - F_X(x-h)}{2h} = \lim_{h\to 0} \frac{1}{2\pi} \lim_{u\to\infty} \int_{-u}^{u} \frac{\text{sen} ht}{ht} e^{-itx} \varphi_X(t) dt$$

Como $\lim_{h\to 0} \frac{\text{sen}(ht)}{ht} = 1$, trocando a ordem dos limites temos que:

$$f_X(x) = \frac{1}{2\pi} \lim_{u\to\infty} \int_{-u}^{u} e^{-itx} \varphi_X(t) dt,$$

que é a transformada inversa de Fourier de $\varphi_X(t)$.

P7. A variável aleatória X tem distribuição simétrica em torno de 0 se, e somente se, $\varphi_X(t)$ é real para todo $t \in R$.

Prova: X é simétrica em torno de 0 se e somente se $P(X \leq x) = P(X \geq -x)$, $\forall x \in R$. Como $X \geq -x \Leftrightarrow -X \leq x$, nós temos que $F_X = F_{-X}$, ou seja, $\varphi_X = \varphi_{-X}$. Como

$$\varphi_{-X}(t) = Ee^{it(-X)} = Ee^{i(-t)X} = \varphi_X(-t) = \overline{\varphi_X(t)}.$$

Então, X é simétrica em torno de 0 se e somente se $\varphi_X(t) = \overline{\varphi_X(t)}$, ou seja, se $\varphi_X(t)$ é real para todo $t \in R$. ▮

P8. Se $E|X|^n < \infty$, então $\varphi_X^{(k)}(0) = i^k EX^k$ para $k \in \{1, \ldots, n\}$, de modo que a função característica é uma espécie de função geradora de momentos.

Prova: Suponhamos que X seja integrável; queremos provar que $\varphi'_X(t) = E(iXe^{itX})$.

Note que para $h \neq 0$, temos $\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = E(e^{itX} \frac{(e^{ihX}-1)}{h})$. Como $\frac{(e^{ihx}-1)}{h} \to ix$ quando $h \to 0$ (regra de L'Hopital), $\forall x \in R$, temos que o resultado decorre se pudermos trocar a ordem do limite e da esperança. Mas como para todo x,

$$|\frac{e^{ihx} - 1}{h}| = |\frac{\int_0^h ixe^{isx} ds}{h}| = |x| \cdot |\frac{\int_0^h e^{isx} ds}{h}| \leq |x|.$$

Portanto, como $|e^{itX}| = 1$, temos

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

Como X é integrável, o Teorema da Convergência Dominada implica que

$$\varphi'_X(t) = \lim_{h\to 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} =$$
$$\lim_{h\to 0} E(e^{itX} \frac{(e^{ihX} - 1)}{h}) = E(\lim_{h\to 0} e^{itX} \frac{(e^{ihX} - 1)}{h}) = E(iXe^{itX}).$$

Logo, $\varphi'_X(0) = iEX$. O restante da prova segue por indução em n. ▮

P9. Se $Y = aX + b$, onde a e b são números reais constantes, $\varphi_Y(t) = e^{itb}\varphi_X(at)$.

Prova: $\varphi_Y(t) = Ee^{itY} = Ee^{it(aX+b)} = Ee^{itb}e^{itaX} = e^{itb}Ee^{i(at)X} = e^{itb}\varphi_X(at)$. ∎

P10. $\varphi_X(t)$ é positiva definida. Isto é, para todo $n = 1, 2, \ldots$, tem-se

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi_X(t_j - t_k)z_j\overline{z_k} \geq 0,$$

para quaisquer números reais $t_1, t_2, \ldots, t_n$ e complexos $z_1, z_2, \ldots, z_n$.

Prova:

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi_X(t_j - t_k)z_j\overline{z_k}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}E(e^{iX(t_j-t_k)})z_j\overline{z_k}\\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}E(z_je^{iX(t_j)}\overline{z_k}e^{-iXt_k})\\
&= E(\sum_{j=1}^{n}\sum_{k=1}^{n}z_je^{iX(t_j)}\overline{z_ke^{iXt_k}})\\
&= E[(\sum_{j=1}^{n}z_je^{iX(t_j)})(\sum_{k=1}^{n}\overline{z_ke^{iXt_k}})]\\
&= E[(\sum_{j=1}^{n}z_je^{iX(t_j)})\overline{(\sum_{k=1}^{n}z_ke^{iXt_k})}]\\
&= E(|\sum_{j=1}^{n}z_je^{iX(t_j)}|^2) \geq 0
\end{aligned}$$

Portanto, φ_X é positiva definida. ∎

Exemplo 3.2.6: Se $\varphi_X(t) = \frac{1}{1+t^2}$, calcule $VarX$.

Solução: Diferenciando φ_X, temos $\varphi'_X(t) = \frac{-2t}{(1+t^2)^2}$. Diferenciando mais uma vez, $\varphi''_X(t) = \frac{-2(1+t^2)^2+2t(2(1+t^2)2t)}{(1+t^2)^4}$. Portanto, $EX = \frac{\varphi'_X(0)}{i} = 0$ e $EX^2 = \frac{\varphi''_X(0)}{i^2} = -(-2) = 2$. Logo, $VarX = EX^2 - (EX)^2 = 2$.

Exemplo 3.2.7: Se uma variável aleatória X tem função característica

$$\varphi_X(t) = \frac{(1-\alpha)^2}{1+\alpha^2 - 2\alpha\cos(t)}, \text{ para } 0 < \alpha < 1.$$

Calcule EX e $Var(X)$.

Solução: Diferenciando φ_X, temos

$$\varphi_X'(t) = \frac{-(1-\alpha)^2 2\alpha \operatorname{sen}(t)}{(1+\alpha^2-2\alpha\cos(t))^2}.$$

Diferenciando mais uma vez,

$$\varphi_X''(t) = \frac{(1+\alpha^2-2\alpha\cos(t))^2(-(1-\alpha)^2 2\alpha\cos(t)) + (1-\alpha)^2 2\alpha \operatorname{sen}(t)\frac{d}{dt}(1+\alpha^2-2\alpha\cos(t))^2}{(1+\alpha^2-2\alpha\cos(t))^4}.$$

Portanto, $EX = \frac{\varphi_X'(0)}{i} = 0$ e $EX^2 = \frac{\varphi_X''(0)}{i^2} = -(\frac{-2\alpha}{(1-\alpha)^2}) = (\frac{2\alpha}{(1-\alpha)^2})$. Logo, $VarX = EX^2 - (EX)^2 = (\frac{2\alpha}{(1-\alpha)^2})$.

Exemplo 3.2.8: Seja $\varphi(t) = \cos(at)$, onde $a > 0$. Mostraremos que φ é função característica, achando a distribuição correspondente. Já que assume valores reais, se φ fosse função característica de alguma variável aleatória X, então por P7, X possuiria distribuição simétrica em torno de zero. Com efeito teríamos $\cos(at) = \varphi(t) = E\cos(tX)$, pois a parte imaginária seria nula. Como $\cos(at) = \cos(-at)$, é evidente que uma distribuição simétrica concentrada nos dois pontos a e $-a$ corresponderia a função característica φ. Portanto, φ é função característica de X, se, e somente se, $P(X = a) = 1/2 = P(X = -a)$.

Exemplo 3.2.9: Sejam X_1 e X_2 duas variáveis aleatórias i.i.d. e seja $Y = X_1 - X_2$. Qual a função característica de Y?

Solução: Seja φ a função característica de X_1 e X_2. Por P9 e P3, temos que $\varphi_{-X_2}(t) = \varphi(-t) = \overline{\varphi(t)}$. Então, como X_1 e X_2 são independentes, por P5, temos que

$$\varphi_Y(t) = \varphi(t)\varphi_{-X_2}(t) = |\varphi(t)|^2.$$

Teorema 3.2.10: *Uma função contínua $\psi : \mathrm{R} \to \mathrm{C}$ com $\psi(0) = 1$ é função característica de alguma variável aleatória se, e somente se, ela for positiva definida.*

Prova: Conforme propriedades já demonstradas, se for função característica, é contínua, positiva definida e aplicada em 0, resulta o valor 1. A prova da recíproca será omitida. ∎

3.2.2 Exemplos de Funções Características

Bernoulli. Suponhamos que $X \sim Bernoulli(p)$, onde $P(X = 1) = p = 1 - P(X = 0)$. Então,

$$\varphi_X(t) = Ee^{itX} = pe^{it} + (1-p).$$

Poisson. Suponhamos que $X \sim Poisson(\lambda)$. Então,

$$\varphi_X(t) = Ee^{itX} = \sum_{n=0}^{\infty} e^{itn} e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^{it})^n}{n!} = e^{\lambda(e^{it}-1)}.$$

Uniforme. Suponhamos que $X \sim Uniforme(-a, a)$. Então, $f_X(x) = \frac{1}{2a}$ para $-a < x < a$, e $f_X(x) = 0$ caso contrário. Logo, se $t = 0$, então $\varphi_X(0) = 1$, e para $t \neq 0$,

$$\varphi_X(t) = Ee^{itX} = \int_{-a}^{a} \frac{e^{itx}}{2a} dx = \frac{1}{2a}(\frac{e^{ita} - e^{-ita}}{it}) = \frac{\text{sen}(ta)}{ta}.$$

Normal. Suponhamos que $X \sim N(0, 1)$. Então,

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx} e^{\frac{-x^2}{2}} dx = e^{\frac{-t^2}{2}} \frac{1}{\sqrt{2\pi}} \int e^{-\frac{(x-it)^2}{2}} dx = e^{\frac{-t^2}{2}},$$

onde esta última integral pode ser calculada utilizando o Teorema de Cauchy tendo em vista que $e^{\frac{-z^2}{2}}$ é uma função analítica no plano complexo.

Exponencial. Suponhamos que $X \sim Exp(\alpha)$. Então,

$$\varphi_X(t) = \int_0^\infty e^{itx} \alpha e^{-\alpha x} dx = \alpha \int_0^\infty e^{x(-\alpha+it)} dx = [\frac{\alpha}{-\alpha + it} e^{x(-\alpha+it)}]_0^\infty = \frac{\alpha}{\alpha - it}.$$

Exemplo 3.2.11: Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, seguindo o modelo de Poisson com parâmetro λ. Queremos obter a distribuição de $X_1 + X_2 + \ldots + X_n$.

Solução: Temos

$$\varphi_{X_1+\ldots+X_n}(t) = E(e^{it(X_1+\ldots+X_n)}) = \prod_{j=1}^{n} E(e^{itX_j}) = e^{n\lambda(e^{it}-1)}.$$

Portanto, $X_1 + X_2 + \ldots + X_n$ tem uma distribuição Poisson com parâmetro $n\lambda$.

3.3 Teorema da Continuidade de Levy

Nosso objetivo nesta seção é provar que $X_n \to^D X$ se, e somente se, $\varphi_{X_n}(t) \to \varphi_X(t), \forall t \in R$. Antes de provarmos a necessidade desta afirmação, considere a seguinte definição de convergência de funções de distribuição.

Definição 3.3.1: Seja $X, X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias com funções de distribuição acumuladas dadas respectivamente por $F, F_1, F_2, \ldots$. Diz-se que F_n *converge fracamente* para F, se $X_n \to^D X$. ∎

Teorema 3.3.2: Teorema de Helly-Bray. *Sejam $F, F_1, F_2, \ldots$ funções de distribuição. Se F_n converge fracamente para F, então*

$$\int g(x) dF_n(x) \to \int g(x) dF(x)$$

para toda função $g : R \to R$ contínua e limitada.

Observação 3.3.3: A integral $\int g(x)dF(x)$ é a esperança de $g(X)$, onde X é uma variável aleatória com função de distribuição F, ela é calculada utilizando a integral de Lebesgue-Stieltjes. No caso de F ser discreta, essa integral é equivalente a:

$$\sum_i g(x_i)p(x_i),$$

e quando F for absolutamente contínua com função densidade de probabilidade f, essa integral é equivalente a:

$$\int_{-\infty}^{\infty} g(x)f(x)dx,$$

onde esta última integral é a integral de Riemann. ▮

Prova: Para $-\infty < a < b < \infty$, onde a e b são pontos de continuidade de F,

$$|\int gdF_n - \int gdF| \leq |\int gdF_n - \int_a^b gdF_n| + |\int_a^b gdF_n - \int_a^b gdF| + |\int_a^b gdF - \int gdF| = I + II + III.$$

Seja $c = \sup_{x\in R}|g(x)| < \infty$ e seja $\epsilon > 0$. Então,

$$\begin{aligned} III = |\int_a^b gdF - \int gdF| &= |\int_{-\infty}^a gdF + \int_b^{\infty} gdF| \leq |\int_{-\infty}^a gdF| + |\int_b^{\infty} gdF| \\ &\leq \int_{-\infty}^a |g|dF + \int_b^{\infty} |g|dF \leq \int_{-\infty}^a cdF + \int_b^{\infty} cdF = c(F(a) + 1 - F(b)) \end{aligned}$$

Logo, para qualquer $\epsilon > 0$, podemos escolher a suficientemente pequeno e b suficientemente grande tal que $III < \epsilon$, pois $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$. Para esses valores de a e b, e para n suficientemente grande, como a e b são pontos de continuidade de F, e como F_n converge fracamente para F, temos que $I \leq c(F_n(a) + 1 - F_n(b)) < 2\epsilon$.

Consideremos agora II. Sejam a e b os pontos já escolhidos. Já que g é uniformemente contínua em $[a, b]$,[1] podemos escolher $x_0, x_1, \ldots, x_N$ tais que $a = x_0 < x_1 < \ldots < x_N = b$, onde x_i são pontos de continuidade de F e $|g(x) - g(x_i)| < \epsilon$ para todo $x \in [x_i, x_{i+1}]$, $i \in \{0, \ldots, N-1\}$. Então,

$$m_{ni} = (g(x_i) - \epsilon)(F_n(x_{i+1}) - F_n(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)dF_n(x) \leq (g(x_i) + \epsilon)(F_n(x_{i+1}) - F_n(x_i)) = M_{ni}$$

e

$$m_i = (g(x_i) - \epsilon)(F(x_{i+1}) - F(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)dF(x) \leq (g(x_i) + \epsilon)(F(x_{i+1}) - F(x_i)) = M_i.$$

Portanto,

$$m_{ni} - M_i \leq \int_{x_i}^{x_{i+1}} g(x)dF_n(x) - \int_{x_i}^{x_{i+1}} g(x)dF(x) \leq M_{ni} - m_i,$$

[1] Uma função g é uniformemente contínua em $[a, b]$ se para todo $\epsilon > 0$, existe $\delta > 0$ tal que para todo $x, y \in [a, b]$ se $|x - y| < \delta$, então $|g(x) - g(y)| < \epsilon$. É fácil provar que toda função contínua em um intervalo fechado é uniformemente contínua neste intervalo.

para $i \in \{0, \ldots, N-1\}$. Somando, temos

$$\sum_{i=0}^{N-1}(m_{ni} - M_i) \leq \int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x) \leq \sum_{i=0}^{N-1}(M_{ni} - m_i).$$

Quando $n \to \infty$, temos que $m_{ni} \to m_i$ e $M_{ni} \to M_i$, logo,

$$\sum_{i=0}^{N-1}(m_{ni} - M_i) \to \sum_{i=0}^{N-1}(m_i - M_i) = -2\epsilon(F(b) - F(a)) \geq -2\epsilon$$

e

$$\sum_{i=0}^{N-1}(M_{ni} - m_i) \to \sum_{i=0}^{N-1}(M_i - m_i) = 2\epsilon(F(b) - F(a)) \leq 2\epsilon$$

Como para n suficientemente grande temos que $|\sum_{i=0}^{N-1}(m_{ni} - M_i) - \sum_{i=0}^{N-1}(m_i - M_i)| < \epsilon$ e $|\sum_{i=0}^{N-1}(M_{ni} - m_i) - \sum_{i=0}^{N-1}(M_i - m_i)| < \epsilon$, segue que $\sum_{i=0}^{N-1}(m_{ni} - M_i) \geq -3\epsilon$ e $\sum_{i=0}^{N-1}(M_{ni} - m_i) \leq 3\epsilon$. Então, para n suficientemente grande, temos que $II \leq 3\epsilon$. Portanto, $|\int gdF_n - \int gdF| \leq 6\epsilon$ para n grande o suficiente. ∎

Como $\cos(tx)$ e $\text{sen}(tx)$ são funções contínuas e limitadas, tem-se que para t fixo

$$E(\cos(tX_n)) \to E(\cos(tX))$$

e

$$E(\text{sen}(tX_n)) \to E(\text{sen}(tX))$$

Logo, $\varphi_{X_n}(t) \to \varphi_X(t)$.

É fácil definir a função característica φ dada uma função de distribuição F: $\varphi(t) = \int e^{itx}dF(x), \forall t \in R$. O próximo teorema implica a suficiência do nosso objetivo nesta seção, ou seja, se $\varphi_{X_n} \to \varphi_X$, então $X_n \to^D X$.

Teorema 3.3.4: *Sejam $F_1, F_2, \ldots$ funções de distribuições e $\varphi_1, \varphi_2, \ldots$ suas funções características. Se φ_n converge pontualmente para um limite φ e se φ é contínua no ponto zero, então*

(a) existe uma função de distribuição F tal que $F_n \to F$ fracamente; e

(b) φ é a função característica de F.

Prova: Note que o teorema anterior implica que, sob as hipóteses, (a) implica (b). Para provar que F_n converge fracamente para alguma função de distribuição, vamos primeiro provar que para toda seqüência de funções de distribuição satisfazendo as condições do teorema, existem uma subseqüência $F_{n_1}, F_{n_2}, \ldots$ e uma função de distribuição F tais que $F_{n_j} \to F$ fracamente, quando $j \to \infty$. Provaremos isso em duas etapas:

(i) existem uma subseqüência $F_{n_1}, F_{n_2}, \ldots$ e uma função $F : R \to [0,1]$ tais que F é não-decrescente e contínua à direita e $F_{n_j}(x) \to F(x)$, quando $j \to \infty$, para todo x ponto de continuidade de F; e

(ii) $F(\infty) = 1$ e $F(-\infty) = 0$.

Para provar (i), usaremos o método da diagonalização. Sejam $r_1, r_2, \ldots$, uma enumeração dos racionais da reta. Considere a seguinte matriz:

$$\begin{array}{ccccc} F_1 & F_2 & F_3 & F_4 & \cdots \\ F_1^1 & F_2^1 & F_3^1 & F_4^1 & \cdots \\ F_1^2 & F_2^2 & F_3^2 & F_4^2 & \cdots \\ F_1^3 & F_2^3 & F_3^3 & F_4^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

Nesta matriz temos que a seqüência $(F_1^j, F_2^j, F_3^j, \ldots)$ contida na $(j+1)$-ésima linha da matriz é uma subseqüência da seqüência contida na j-ésima linha que converge no racional r_j, para $j \geq 1$. Note que como a seqüência $(F_1^{j-1}(r_j), F_2^{j-1}(r_j), F_3^{j-1}(r_j), \ldots)$ é uma seqüência limitada de números reais, ela possui uma subseqüência convergente; logo pode-se escolher a seqüência $(F_1^j, F_2^j, F_3^j, \ldots)$ indutivamente conforme descrito acima. Seja $F_{n_j} = F_j^j$, para $j \geq 1$, então temos que a subseqüência $(F_{n_j})_j$ converge em todos os racionais da reta. Chamemos o limite de $F(r_k)$, de modo que $F_{n_j}(r_k) \to F(r_k), \forall k$. É óbvio que $0 \leq F(r_k) \leq 1$ e que F é não decrescente nos racionais. Definamos F em x irracional por $F(x) = \lim_{r \downarrow x, r \text{ rational}} F(r)$. F assim definida é não-decrescente, mas não é necessariamente contínua à direita. Vamos provar que $F_{n_j}(x) \to F(x)$ para todo ponto x de continuidade de F. Suponha que x é um ponto de continuidade de F e sejam r' e r'' racionais tais que $r' < x < r''$ e $F(r'') - \epsilon < F(x) < F(r') + \epsilon$. Então,

$$F(x) - \epsilon < F(r') = \lim_{j\to\infty} F_{n_j}(r') \leq \liminf_{j\to\infty} F_{n_j}(x) \leq$$
$$\limsup_{j\to\infty} F_{n_j}(x) \leq \lim_{j\to\infty} F_{n_j}(r'') = F(r'') < F(x) + \epsilon$$

Como ϵ é arbitrário, temos $F_{n_j}(x) \to F(x)$ quando $j \to \infty$. Finalmente, podemos redefinir F nos seus pontos de descontinuidade de modo que F seja contínua à direita.

Para provar (ii), note que

$$\int_0^t \varphi_{n_j}(s) ds = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF_{n_j}(x) ds.$$

Mas como o integrando é limitado podemos trocar a ordem de integração, logo

$$\int_0^t \varphi_{n_j}(s) ds = \int_{-\infty}^{\infty} \int_0^t e^{isx} ds dF_{n_j}(x) = \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF_{n_j}(x)$$

Considere a função, $h(x) = \frac{e^{itx}-1}{ix}$ para $x \neq 0$ e $h(0) = t$. h é limitada e contínua e um argumento similar ao utilizado na prova do teorema anterior, pode ser utilizado para provar que quando $j \to \infty$

$$\int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF_{n_j}(x) = \int_{-\infty}^{\infty} h(x) dF_{n_j}(x) \to \int_{-\infty}^{\infty} h(x) dF(x) =$$
$$\int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} dF(x) = \int_0^t \int_{-\infty}^{\infty} e^{isx} dF(x) ds$$

Como $\varphi_{n_j}(t) \to \varphi(t)$, φ é contínua em zero, implica que φ é limitada e mensurável, então pelo teorema da convergência dominada, tem-se que

$$\int_0^t \varphi_{n_j}(s)ds \to \int_0^t \varphi(s)ds.$$

Igualando-se os limites iguais e dividindo-se por t, temos

$$\frac{1}{t}\int_0^t \varphi(s)ds = \frac{1}{t}\int_0^t \int_{-\infty}^{\infty} e^{isx}dF(x)ds, t \neq 0.$$

Fazendo $t \to 0$ e usando a continuidade em $s = 0$ das duas funções $\varphi(s)$ e $\int_{-\infty}^{\infty} e^{isx}dF(x)$, tem-se

$$\varphi(0) = \int_{-\infty}^{\infty} 1dF(x) = F(\infty) - F(-\infty).$$

Como $\varphi(0) = \lim_{n\to\infty}\varphi_n(0) = 1$, temos que $F(\infty) - F(-\infty) = 1$, ou seja, o que implica que $F(\infty) = 1$ e $F(-\infty) = 0$.

Para terminar a prova suponha por contradição que F_n não convirja fracamente para F, onde $F_{n_j} \to F$ fracamente. Então, existirão x, ponto de continuidade de F e uma subseqüência $F_{1'}, F_{2'}, \ldots$ tais que $F_{n'}(x) \to a \neq F(x)$. Como essa subseqüência também satisfaz as condições do teorema, (i) e (ii) implicam que existe uma subseqüência $F_{1''}, F_{2''}, \ldots$ e uma função de distribuição G tais que $F_{n''} \to G$ fracamente. Como F e G possuem a mesma função característica (φ), temos que $F = G$, ou seja $F_{n''}(x) \to a = G(x) = F(x)$, uma contradição. ∎

Exemplo 3.3.5: Suponha que X_n e Y_n são independentes para cada $n \geq 0$ e que $X_n \to^D X_0$ e $Y_n \to^D Y_0$. Prove que $X_n + Y_n \to^D X_0 + Y_0$.

Solução: Pelo Teorema da Continuidade sabemos que $\varphi_{X_n}(t) \to \varphi_{X_0}(t)$ e que $\varphi_{Y_n}(t) \to \varphi_{Y_0}(t)$. Como X_n e Y_n são independentes temos que $\varphi_{X_n+Y_n}(t) = \varphi_{X_n}(t)\varphi_{Y_n}(t)$. Portanto,

$$\lim_n \varphi_{X_n+Y_n}(t) = \lim_n(\varphi_{X_n}(t)\varphi_{Y_n}(t)) = \varphi_{X_0}(t)\varphi_{Y_0}(t) = \varphi_{X_0+Y_0}(t).$$

Logo, pelo Teorema da Continudade, temos que $X_n + Y_n \to^D X_0 + Y_0$.

Exemplo 3.3.6: Suponha que a variável aleatória X_n tenha distribuição Binomial, ou seja,

$$P(X_n = k) = \binom{n}{k} p_n^k(1-p_n)^{n-k}, k = 0, 1, 2, \ldots, n.$$

Se $p_n \to 0$ quando $n \to \infty$ de tal modo que $np_n \to \lambda > 0$, então

$$X_n \to^D Y,$$

onde $Y \sim Poisson(\lambda)$. Para verificar isto relembre que podemos representar uma variável aleatória Binomial como a soma de variáveis aleatórias Bernoulli i.i.d., então

$$\varphi_{X_n}(t) = Ee^{itX_n} = (1 - p_n + e^{it}p_n)^n = (1 + p_n(e^{it} - 1))^n = (1 + \frac{np_n(e^{it}-1)}{n})^n \to e^{\lambda(e^{it}-1)},$$

onde a expressão final é a função característica de uma variável aleatória $Poisson(\lambda)$. Portanto, pelo Teorema da Continuidade, $X_n \to^D Y$.

Exemplo 3.3.7: Suponha que $X_n \sim U(-n, n)$. Mostre que

(a) Mostre que $\lim_n \varphi_{X_n}(t)$ existe.

(b) Existe alguma variável aleatória X_0 tal que $X_n \to X_0$?

Solução: Como

$$\varphi_X(t) = \begin{cases} \frac{\text{sen}(nt)}{nt} & , \text{ se } t \neq 0 \\ 1 & , \text{ se } t = 0, \end{cases}$$

temos que

$$\lim_n \varphi_{X_n}(t) = \begin{cases} 0 & , \text{ se } t \neq 0 \\ 1 & , \text{ se } t = 0. \end{cases}$$

Para parte (b), note que pelo teorema da continuidade, temos que se existir tal que $\varphi_{X_0}(t)$ seria sua função característica. Como $\varphi_{X_0}(t)$ não é contínua não pode ser uma função característica.

Exemplo 3.3.8: Prove que se $X_n \to^D X$, então $aX_n + b \to^D aX + b$, para a e b constantes reais.

Exemplo 3.3.9: Seja $\{a_n : n \geq 1\}$ uma seqüência de números reais com $a_n \to a$ para a finito. Prove que se $X_n \to^D X$, onde $X \sim N(0, 1)$, então $X_n + a_n \to^D Y$, onde $Y \sim N(a, 1)$.

Exemplo 3.3.10: Se X_n e Y_n são independentes para $n \geq 0$, $X_n \to^D X_0$, e $Y_n \to^D Y_0$. Encontre o limite em distribuição de $X_n + Y_n$.

Exemplo 3.3.11: A seqüência de variáveis aleatórias $\{X_n : n \geq 1\}$ é tal que cada X_n segue o modelo geométrico de parâmetro p/n, para algum $0 < p < 1$. Verifique se ocorre $\frac{X_n}{n} \to^D X$ para alguma variável aleatória X e obtenha sua distribuição.

Solução: Temos $\varphi_{X_n}(t) = \frac{1-p/n}{1-\frac{p}{n}e^{it}}$, e $\varphi_{\frac{X_n}{n}}(t) = \varphi_{X_n}(\frac{t}{n}) = \frac{1-p/n}{1-\frac{p}{n}e^{i\frac{t}{n}}}$. Como $\lim \varphi_{\frac{X_n}{n}}(t) = \lim \frac{1-p/n}{1-\frac{p}{n}e^{i\frac{t}{n}}} = 1$. Temos que $\frac{X_n}{n} \to^D 0$.

3.4 Soma de um Número Aleatório de Variáveis Aleatórias

Nesta seção, nós estudaremos somas de um número aleatório de variáveis aleatórias, ou seja,

$$S = \sum_{i=0}^{N} X_i,$$

onde N é uma variável aleatória inteira e não negativa, e assume-se que ela é independente das parcelas X_i. Por exemplo, N pode ser o número de clientes, pacotes ou trabalhos chegando em uma fila em um dado intervalo de tempo e X_i pode ser o tempo necessário para finalizar o i-ésimo trabalho. S então seria o tempo total do serviço. Em nossas aplicações assumiremos que $N = 0$ significa que $S = 0$, ou seja, $X_0 = 0$ com função característica $\varphi_{X_0}(u) = 1$. Sabemos que $ES = E[E(S|N)]$ e que

$$E(S|N = n) = \sum_{i=0}^{n} E(X_i|N = n).$$

Como assumimos que N é independente de X_i, temos

$$E(S|N = n) = \sum_{i=0}^{n} EX_i.$$

Se as variáveis aleatórias $\{X_i, i > 0\}$ têm esperança igual a m, então $E(S|N = n) = nm$ e $ES = mEN$.

Para informações mais detalhadas sobre S, vamos calcular sua função característica φ_S assumindo que as variáveis aleatórias $\{N, X_1, X_2, \ldots\}$ são independentes:

$$\varphi_S(t) = Ee^{itS} = E(E(e^{itS}|N)).$$

Por outro lado, utilizando a hipótese de independência, podemos calcular,

$$E(e^{itS}|N = n) = E(\prod_{i=0}^{n} e^{itX_i}|N = n) = \prod_{i=0}^{n} \varphi_{X_i}(t).$$

Logo,

$$\varphi_S(t) = \sum_{n=0}^{\infty} P(N = n) \prod_{i=0}^{n} \varphi_{X_i}(t).$$

Se as parcelas $\{X_1, X_2, \ldots\}$ forem também identicamente distribuídas com função característica φ_X, então

$$\varphi_S(t) = \sum_{n=0}^{\infty} P(N = n)\varphi_X^n(t),$$

onde utilizamos o fato que $\varphi_X^0 = 1 = \varphi_{X_0}(t)$. Note que a função característica de N é:

$$\varphi_N(t) = \sum_{n=0}^{\infty} P(N = n)e^{itn} = \sum_{n=0}^{\infty} P(N = n)[e^{it}]^n.$$

Comparando as expressões de φ_S e φ_N, nós vemos que escolhendo t em $\varphi_N(t)$ de forma que $e^{it} = \varphi_X$, nós podemos reescrever:

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)).$$

Portanto, nós provamos o seguinte teorema:

Teorema 3.4.1: *Se N é uma variável aleatória inteira não-negativa, $S = \sum_{i=0}^{N} X_i, X_0 = 0$, onde $\{X_i, i \geq 1\}$ são i.i.d. com função característica comum φ_X, e elas são independentes de N que é descrita pela função característica φ_N, então*

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)).$$

Exemplo 3.4.2: Suponha que $N \sim Poisson(\lambda)$ representa o número de clientes que são atendidos em um dado tempo T. Suponha ainda que com probabilidade p o i-ésimo cliente fica satisfeito com o atendimento. Assuma que os clientes ficam satisfeitos com o serviço de maneira independente e que N, é independente da probabilidade que clientes ficam satisfeitos. Determine a distribuição de probabilidade de S o número total de clientes satisfeitos no tempo T.

Solução: Seja $X_i \sim Bernoulli(p)$, $i \geq 1$, a variável aleatória que descreve se o i-ésimo cliente ficou ou não satisfeito com o atendimento. Então temos,

$$S = \sum_{i=0}^{N} X_i,$$

onde $X_0 = 0$. Desta forma, sabemos que

$$\varphi_S(t) = \varphi_N(-i \log \varphi_X(t)),$$

onde $\varphi_X(t) = pe^{it} + (1-p)$ e $\varphi_N(t) = e^{\lambda(e^{it}-1)}$. Substituindo temos:

$$\varphi_S(t) = e^{\lambda(e^{i(-i\log(pe^{it}+(1-p)))}-1)} = e^{\lambda(pe^{it}+(1-p)-1)} = e^{p\lambda(e^{it}-1)}.$$

Pela unicidade da função característica, temos que $S \sim Poisson(p\lambda)$.

3.5 Função Característica de um Vetor Aleatório

Definição 3.5.1: Seja $\vec{X} = (X_1, \ldots, X_k)$ um vetor aleatório k-dimensional. A função característica de $\vec{X}$ é a função $\varphi_{\vec{X}} : I\!R^k \to C$ definida por

$$\varphi_{\vec{X}}(\vec{t}) = Ee^{i\vec{t}\cdot\vec{X}} = Eexp(i \sum_{j=1}^{k} t_j X_j).$$

$\varphi_{\vec{X}}$ é também chamada de função característica conjunta de $X_1, \ldots, X_k$.

A função característica multivariada tem propriedades análogas a todas as propriedades enunciadas para a função característica de uma variável aleatória. As propriedades P1–P4 e P7 são válidas com as óbvias modificações (a reta é substituída por $I\!R^k$). Para P5, supõe-se que $\vec{X}$ e $\vec{Y}$ sejam vetores de mesma dimensão. Sob esta condição, a independência de $\vec{X}$ e $\vec{Y}$ implica que

$$\varphi_{\vec{X}+\vec{Y}}(\vec{t}) = \varphi_{\vec{X}}(\vec{t})\varphi_{\vec{Y}}(\vec{t}).$$

Quanto a P6, também existe uma fórmula da inversão para a função característica multidimensional que pode ser usada para provar a unicidade da função característica:

Teorema 3.5.2: Teorema da Unicidade. *Se $\vec{X}$ e $\vec{Y}$ forem vetores aleatórios k-dimensionais tais que $\varphi_{\vec{X}}(\vec{t}) = \varphi_{\vec{Y}}(\vec{t}), \forall \vec{t} \in I\!R^k$, então $\vec{X}$ e $\vec{Y}$ têm a mesma distribuição. Em outras palavras, a função característica determina a distribuição, e podemos escrever: $\varphi_{\vec{X}} = \varphi_{\vec{Y}} \Leftrightarrow F_{\vec{X}} = F_{\vec{Y}}$.*

Analogamente a P8, correlações de ordem maiores podem ser facilmente calculadas diferenciando-se a função característica conjunta repetidamente. Formalmente, seja $p = \sum_{k=1}^{n} p_k$ para números naturais quaisquer p_k, temos

$$E(\prod_1^n X_k^{p_k}) = \frac{1}{i^p}\frac{\partial^p \varphi_{\vec{X}}(\vec{t})}{\partial t_1^{p_1}\cdots\partial t_n^{p_n}}|_{\vec{t}=\vec{0}}.$$

No caso particular de $\vec{X} = (X_1, X_2)$, temos que

$$EX_1X_2 = -\frac{\partial^2 \varphi_{X_1,X_2}(t_1,t_2)}{\partial t_1 \partial t_2}|_{t_1=t_2=0}.$$

Também é fácil analisar o comportamento da função característica multivariada de transformações lineares de vetores aleatórios em analogia a propriedade P9. (Assumiremos que um vetor $\vec{X}$ k-dimensional é uma matriz coluna com dimensão $k\times 1$. Deste modo $\vec{t}\cdot\vec{X} = (\vec{t})^T\vec{X}$.) Por exemplo, seja $\vec{Y} = \mathbf{A}\vec{X} + \vec{b}$, então

$$\begin{aligned}\varphi_{\vec{Y}}(\vec{t}) &= Ee^{i(\vec{t})^T\vec{Y}} = Ee^{i(\vec{t})^T(\mathbf{A}\vec{X}+\vec{b})}\\ &= E(e^{i(\vec{t})^T\vec{b}}e^{i(\mathbf{A}^T\vec{t})^T\vec{X}}) = e^{i(\vec{t})^T\vec{b}}\varphi_{\vec{X}}(\mathbf{A}^T\vec{t}),\end{aligned}$$

onde utilizamos o fato que $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ e que $e^{i(\vec{t})^T\vec{b}}$ não é aleatório e pode sair fora da operação de esperança.

Assim como é fácil obter a distribuição marginal dada uma distribuição conjunta de variáveis aleatórias, também é fácil obter a função característica de qualquer distribuição marginal. Para isso basta fazer todos os termos "extras" iguais a zero na função característica multivariada. Por exemplo, para as variáveis aleatórias X, Y, e Z, temos $Ee^{i(xX+yY)} = Ee^{i(xX+yY+0Z)}$, ou seja, $\varphi_{X,Y}(x,y) = \varphi_{X,Y,Z}(x,y,0), \forall (x,y)\in I\!R^2$.

Como no caso unidimensional, temos convergência em distribuição se, e somente se, as funções características convergem.

Teorema 3.5.3: *$\vec{X}_n \to^D \vec{X}$ se, e somente se, $\varphi_{\vec{X}_n}(\vec{t}) \to \varphi_{\vec{X}}(\vec{t}), \forall \vec{t}\in I\!R^k$.*

Prova: Omitida. ∎

Terminaremos nossa discussão de funções características multidimensionais considerando um critério para independência de vetores aleatórios.

Teorema 3.5.4: *Sejam $\vec{X} = (X_1,\ldots,X_m)$ e $\vec{Y} = (Y_1,\ldots,Y_n)$ vetores aleatórios, onde $m\geq 1, n\geq 1$. $\vec{X}$ e $\vec{Y}$ são independentes se, e somente se,*

$$\varphi_{X_1,\ldots,X_m,Y_1,\ldots,Y_n}(x_1,\ldots,x_m,y_1,\ldots,y_n) = \varphi_{\vec{X}}(x_1,\ldots,x_m)\varphi_{\vec{Y}}(y_1,\ldots,y_n),$$

para todo $(x_1,\ldots,x_m)\in I\!R^m$ e $(y_1,\ldots,y_n)\in I\!R^n$.

Prova: Suponhamos primeiro que $\vec{X}$ e $\vec{Y}$ sejam variáveis aleatórias X e Y $(m = 1, n = 1)$, com X e Y independentes. Então temos,

$$\varphi_{X,Y}(x, y) = Ee^{i(xX+yY)} = Ee^{ixX}e^{iyY} = Ee^{ixX}Ee^{iyY} = \varphi_X(x)\varphi_Y(y), \forall(x, y) \in I\!R^2.$$

Reciprocamente, suponha que $\varphi_{X,Y}(x, y) = \varphi_X(x)\varphi_Y(y)$ para todo $(x, y) \in I\!R^2$. Então a independência de X e Y é conseqüência do Teorema da Unicidade: se X e Y fossem independentes, elas teriam função característica conjunta $\varphi_{X,Y}(x, y) = \varphi_X(x)\varphi_Y(y)$ pela parte inicial desta demonstração. Se não fossem independentes, elas teriam uma função característica diferente, o que contraria a hipótese. Logo, são independentes.

A prova no caso geral é análoga e omitida. ∎

Um resultado semelhante vale para um número finito qualquer de vetores aleatórios. Consideremos o caso mais simples em que $X_1, \ldots, X_n$ são variáveis aleatórias. Então, temos $X_1, \ldots, X_n$ independentes se, e somente se,

$$\varphi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \prod_{j=1}^{n} \varphi_{X_j}(t_j), \forall(t_1, \ldots, t_n) \in I\!R^n.$$

3.6 Funções Geratrizes de Momento

Definição 3.6.1: Uma função geratriz de momento $\hat{F}_X(t)$ de uma variável aleatória X com função de distribuição F_X existe se,

$$\hat{F}_X(t) := Ee^{tX} < \infty, \forall t \in I,$$

onde I é um intervalo contendo 0 no seu interior.

O problema de utilizar funções geratrizes de momento é que elas nem sempre existem. Por exemplo, a função geratriz de momento de uma variável aleatória com distribuição de Cauchy não existe. Pode-se provar que a existência da função geratriz de momento é equivalente a cauda da distribuição de X ser limitada exponencialmente, ou seja, $P(|X| > x) \leq Ke^{-cx}$, para algum $K > 0$ e $c > 0$. Se a função geratriz de momento existe, pode-se provar que ela também determina a função de distribuição.

3.7 Teorema de Slutsky

Nesta seção, estudaremos o Teorema de Slutsky que trata do comportamento da soma e do produto de variáveis aleatórias, uma convergindo em distribuição e outra em probabilidade. Antes disso, iremos provar que funções contínuas preservam convergência.

Teorema 3.7.1: *Sejam $\{X_n : n \geq 1\}$ e X variáveis aleatórias com funções de distribuição $\{F_n : n \geq 1\}$ e F, respectivamente. Seja $g : I\!R \to I\!R$ uma função contínua. Então, se X_n converge para X quase certamente, em probabilidade ou em distribuição, o mesmo ocorre com $g(X_n)$ para $g(X)$, no mesmo modo de convergência.*

Prova: Suponha que $X_n \to X$ cp1. Então, existe um conjunto $A \in \mathcal{F}$ tal que $P(A) = 0$ e $X_n(w) \to X(w)$ para $w \in A^c$. Como g é contínua, $g(X_n(w)) \to g(X(w))$ para $w \in A^c$ e, portanto, $g(X_n) \to g(X)$ cp1.

Considere que $X_n \to^P X$ e vamos verificar que $g(X_n) \to^P g(X)$. Dado $\epsilon > 0$ arbitrário, fixemos m grande o suficiente tal que $P(|X| > m/2) < \epsilon$. A função g sendo contínua em $I\!R$, será uniformemente contínua no intervalo fechado $[-m, m]$, logo para $\epsilon' > 0$ arbitrário existe δ tal que $0 < \delta \leq m/2$ e se $x, y \in [-m, m]$ e $|x - y| < \delta$, então $|g(x) - g(y)| < \epsilon'$.

Observe que se $P(A_n) \to 1$, então $P(A_n \cap A) \to P(A)$, pois $P(A_n) + P(A) - 1 \leq P(A_n \cap A) \leq P(A)$ e $P(A_n) + P(A) - 1 \to P(A)$. Portanto, como $P(|X_n - X| < \delta) \to 1$, temos que $P(|X| \leq m/2, |X_n - X| < \delta) \to P(|X| \leq m/2) > 1 - \epsilon$. Mas

$$[|X| \leq m/2, |X_n - X| < \delta] \subseteq [|X| \leq m, |X_n| \leq m, |X_n - X| < \delta] \subseteq [|g(X_n) - g(X)| < \epsilon'],$$

logo $P(|g(X_n) - g(X)| < \epsilon') > 1 - 2\epsilon$ para n suficientemente grande. Como ϵ é arbitrário, temos que $P(|g(X_n) - g(X)| < \epsilon') \to 1$ quando $n \to \infty$, ou seja $g(X_n) \to^P g(X)$.

Finalmente, considere que $X_n \to^D X$. Pelo Teorema da Continuidade de Levy, para que $g(X_n) \to^D g(X)$, basta a convergência das respectivas funções características. Por definição,

$$\varphi_{g(X_n)}(t) = Ee^{itg(X_n)} = E\cos(tg(X_n)) + iE\,\mathrm{sen}(tg(X_n)).$$

Como as funções $\cos(tg(x))$ e $\mathrm{sen}(tg(x))$ são contínuas e limitadas na reta, para t fixo, decorre do Teorema de Helly-Bray que

$$\varphi_{g(X_n)}(t) \to E\cos(tg(X)) + iE\,\mathrm{sen}(tg(X)) = \varphi_{g(X)}(t), \forall t \in I\!R.$$

■

Teorema 3.7.2: *Considere $\{X_n : n \geq 1\}$, $\{Y_n : n \geq 1\}$ e X variáveis aleatórias tais que valem as convergências $X_n \to^D X$ e $Y_n \to^P c$, com c constante. Então,*

(i) $X_n + Y_n \to^D X + c$;

(ii) $X_n Y_n \to^D cX$;

(iii) *Se* $c \neq 0$, $\frac{X_n}{Y_n} \to^D \frac{X}{c}$, *desde que* $P(Y_n \neq 0) = 1$.

Prova: Prova de (i): Temos

$$\varphi_{X_n+Y_n}(t) = E(e^{it(X_n+Y_n)}) = E(e^{it(X_n+c)}) + E[(e^{itX_n})(e^{itY_n} - e^{itc})].$$

Por hipótese temos,

$$\lim_n E(e^{it(X_n+c)}) = \lim_n e^{itc}E(e^{itX_n}) = e^{itc}E(e^{itX}) = E(e^{it(X+c)}).$$

Observe que $|e^{itX_n}| = 1$ e, assim, vem

$$|E[(e^{itX_n})(e^{itY_n} - e^{itc})]| \leq E[|(e^{itX_n})(e^{itY_n} - e^{itc})|] = E[|(e^{itY_n} - e^{itc})|].$$

Seja $Z_n = |(e^{itY_n} - e^{itc})|$, temos $0 \leq Z_n \leq 2$. Logo, para $\epsilon > 0$, temos

$$E[|(e^{itY_n} - e^{itc})|] = EZ_n = E(Z_n I_{Z_n \leq \epsilon}) + E(Z_n I_{Z_n > \epsilon})$$
$$\leq \epsilon + 2E(I_{Z_n > \epsilon}) \leq \epsilon + 2P(Z_n > \epsilon).$$

Como Z_n é uma função contínua de Y_n e lembrando que funções contínuas preservam convergência em probabilidade, temos que $Z_n \to^P 0$, pois $Y_n \to^P c$. Nessas condições, para n grande o suficiente,

$$|E[(e^{itX_n})(e^{itY_n} - e^{itc})]| \leq E[|(e^{itY_n} - e^{itc})|] < 2\epsilon.$$

Logo, tomando o limite de $\varphi_{X_n+Y_n}(t)$ quando $n \to \infty$, concluímos a demonstração da parte (i).

Prova de (ii): Inicialmente consideramos $c = 0$ e vamos verificar que $X_n Y_n \to^P 0$, e conseqüentemente, $X_n Y_n \to^D 0$. Sejam $\epsilon, \delta > 0$ e $x < 0 < y$ pontos de continuidade de F_X tais que $F_X(y) - F_X(x) = P(x < X \leq y) > 1 - \delta$. Como $X_n \to^D X$, temos $P(x < X_n \leq y) = F_{X_n}(y) - F_{X_n}(x) > 1 - 2\delta$ para n suficientemente grande. Definamos $M = \max(y, -x)$, então a convergência em probabilidade de Y_n para zero implica que $P(|Y_n| < \frac{\epsilon}{M}) > 1 - \delta$ para n suficientemente grande. Logo para n suficientemente grande, temos

$$P(x < X_n \leq y, |Y_n| < \frac{\epsilon}{M}) > 1 - 3\delta.$$

Como $x < X_n \leq y$ e $|Y_n| < \frac{\epsilon}{M}$ implicam $|X_n Y_n| < \epsilon$, temos $P(|X_n Y_n| < \epsilon) > 1 - 3\delta$ para n grande o suficiente. Portanto, para todo $\epsilon > 0$, $P(|X_n Y_n| < \epsilon) \to 1$, ou seja, $X_n Y_n \to^P 0$.

Agora consideremos o caso c geral. Como $X_n Y_n = cX_n + (Y_n - c)X_n$ e $Y_n - c \to^P 0$. Pelo caso $c = 0$, temos que $(Y_n - c)X_n \to^P 0$. Além disso como cx é uma função contínua, temos $cX_n \to^D cX$. Como $X_n Y_n$ é a soma de dois termos, o primeiro dos quais converge para cX em distribuição, e o segundo para zero em probabilidade, o resultado é conseqüência da parte (i).

Prova de (iii): Como $1/x$ é contínua para $x \neq 0$, temos que $1/Y_n \to^P 1/c$. Agora, basta aplicar o ítem (ii). ∎

Capítulo 4

Lei dos Grandes Números

4.1 Motivação

Entre outras coisas, a Lei dos grandes Números nos permite formalizar a idéia que à medida que o número de repetições de um experimento cresce, a freqüência relativa f_A de algum evento A converge (quase certamente) para a probabilidade teórica $P(A)$. É este fato que nos permite estimar o valor da probabilidade de um evento A, baseado na freqüência relativa de A em um grande número de repetições de um experimento.

Por exemplo, se uma nova peça for produzida e não tivermos conhecimento anterior sobre quão provável será que a peça seja defeituosa, poderemos proceder à inspeção de um grande número dessas peças, digamos N, contarmos o número de peças defeituosas dentre elas, por exemplo n, e depois empregarmos n/N com uma aproximação da probabilidade de que uma peça seja defeituosa. O número n/N é uma variável aleatória, e seu valor depende essencialmente de duas coisas. Primeira, o valor de n/N depende da probabilidade básica, mas desconhecida, p de que uma peça seja defeituosa. Segunda, depende daquelas N peças que tenham sido inspecionadas. O que a Lei dos Grandes Números mostra é que se a técnica de selecionar as N peças for aleatória, então o quociente n/N convergirá quase certamente para p. (Evidentemente, a seleção das N peças é importante. Se fôssemos escolher somente aquelas peças que exibissem algum defeito físico externo, por exemplo, poderíamos prejudicar seriamente nossos cálculos.)

Mais formalmente, considere um experimento básico, com a variável aleatória X representando o valor de um característico numérico do resultado (no caso anterior, temos que X seria a função indicadora do evento A). Pensemos na realização deste experimento N vezes (N grande), de tal maneira que as realizações sejam independentes. Suponhamos que depois de cada realização do experimento registre-se o valor do característico numérico do resultado; chamemos este um valor observado. A Lei dos Grandes Números afirma que a média aritmética dos n valores observados converge, em certo sentido, para a média EX, quando $N \to \infty$.

Uma versão da Lei dos Grandes Números diz que se $X_1, X_2, \ldots$ são i.i.d. e integráveis, então

$$\frac{X_1 + \ldots + X_n}{n} \to EX_1.$$

Quando o tipo de convergência é convergência em probabilidade, chamamos de Lei Fraca dos Grandes Números, e quando temos convergência quase certa, chamamos de Lei Forte dos Grandes Números.

4.2 Lei Fraca dos Grandes Números

Na seção anterior, motivamos o resultado da Leis dos Grandes Números para variáveis aleatórias independentes e identicamente distribuídas. Nesta seção, analisaremos duas versões da Lei Fraca dos Grandes Números, na primeira delas não é necessário assumir que as variáveis aleatórias são identicamente distribuídas. Antes de enunciarmos esta lei, vamos relembrar uma importante desigualdade, conhecida como desigualdade de Chebyshev.

Teorema 4.2.1: Desigualdade de Chebyshev Generalizada. *Dado um conjunto A e uma função $g(x)$ tal que $\forall x\ g(x) \geq I_A(x)$, tem-se que $P(X \in A) \leq \min(1, Eg(X))$.*

Prova: Pela monotonicidade da Esperança, temos que $Eg(X) \geq EI_A(X) = P(X \in A)$. Mas, como a cota superior pode exceder 1, temos que $\min(1, Eg(X)) \geq P(X \in A)$. ▮

Corolário 4.2.2: Desigualdade (Original) de Chebyshev. *Seja X uma variável aleatória, então $P(|X - EX| \geq \epsilon) \leq \frac{VarX}{\epsilon^2}$.*

Prova: Escolha $A = \{x : |x| \geq \epsilon\}$ e $g(x) = \frac{x^2}{\epsilon^2}$. Note que $g(x) \geq I_A(x)$, então pelo teorema anterior, $P(X \in A) = P(|X| \geq \epsilon) \leq \frac{EX^2}{\epsilon^2}$. Substituindo X por $X - EX$, temos $P(|X - EX| \geq \epsilon) \leq \frac{VarX}{\epsilon^2}$. ▮

Note que a desigualdade de Chebyshev converte conhecimento sobre um momento de segunda ordem ou uma variância numa cota superior para a probabilidade da cauda de uma variável aleatória. Vamos usar esta desigualdade para provar a Lei Fraca dos Grandes Números de Chebyshev.

Teorema 4.2.3: Lei Fraca de Chebyshev *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes 2 a 2 com variâncias finitas e uniformemente limitadas (ou seja, existe c finito tal que para todo n, $VarX_n \leq c$). Então, $X_1, X_2, \ldots$ satisfazem a Lei Fraca dos Grandes Números:*

$$\frac{S_n - ES_n}{n} \to^P 0.$$

Prova: Precisamos provar que para todo $\epsilon > 0$,

$$P(\frac{|S_n - ES_n|}{n} \geq \epsilon) \to 0 \text{ quando } n \to \infty.$$

Como as variáveis aleatórias são independentes 2 a 2, temos que

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) \leq nc.$$

Pela desigualdade de Chebyshev, temos que

$$P(|S_n - ES_n| \geq n\epsilon) \leq \frac{Var(S_n)}{\epsilon^2 n^2} \leq \frac{c}{\epsilon^2 n} \to 0. \tag{4.1}$$

∎

Corolário 4.2.4: Lei dos Grandes Números de Bernoulli. *Consideremos uma seqüência de ensaios binomiais independentes, tendo a mesma probabilidade p de "sucesso" em cada ensaio. Se S_n é o número de sucessos nos primeiros n ensaios, então*

$$\frac{S_n}{n} \to^P p$$

Prova: Seja $X_n = 1$ se o n-ésimo ensaio é sucesso, $X_n = 0$ caso contrário. Então, $X_1, X_2, \ldots$ são i.i.d. e integráveis com média $\mu = p$. Como $VarX_n = p(1-p)$, a Lei Fraca de Chebyshev implica que $\frac{S_n - np}{n} \to^P 0$, ou, equivalentemente, $\frac{S_n}{n} \to^P p$. ∎

Podemos utilizar a Lei Fraca dos Grandes Números para responder a seguinte questão: quantas repetições de um experimento devemos realizar a fim de termos uma probabilidade ao menos $0,95$ para que a freqüência relativa difira de $p = P(A)$ por menos do que, digamos, 0,01? Utilizando a equação (4.1), onde S_n é o número de ocorrências do evento A em n realizações do experimento temos que $S_n/n = f_A$, $ES_n = np$, $VarS_n = np(1-p)$, e:

$$P(|f_A - p| \geq 0,01) \leq \frac{p(1-p)}{n(0,01)^2},$$

ou seja, queremos que $\frac{p(1-p)}{n(0{,}01)^2} \leq 0,05$, o que é equivalente a $n \geq \frac{p(1-p)}{0{,}05(0{,}01)^2}$. Substituindo os valores específicos de $0,05$ e $0,01$ por δ e ϵ, respectivamente, teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta \text{ sempre que } n \geq \frac{p(1-p)}{\delta(\epsilon)^2}.$$

Em muitos problemas, não conhecemos o valor de $p = P(A)$ e, por isso, não poderemos empregar o limite acima. Nesse caso, poderemos empregar o fato de que $p(1-p)$ toma seu valor máximo quando $p = 1/2$, e esse valor máximo é igual a $1/4$. Conseqüentemente, estamos certamente seguros se afirmamos que para $n \geq \frac{1}{4\epsilon^2\delta}$ teremos

$$P(|f_A - p| < \epsilon) \geq 1 - \delta.$$

Exemplo 4.2.5: Peças são produzidas de tal maneira que a probabilidade de uma peça ser defeituosa é p (admitida desconhecida). Um grande número de peças, digamos n, são classificadas como defeituosas ou perfeitas. Que valor deverá ter n de maneira que possamos estar 99% certos de que a freqüência relativa de defeituosas difere de p por menos de $0,05$?

Solução: Porque não conhecemos o valor de p, deveremos aplicar a última fórmula com $\epsilon = 0,05$, $\delta = 0,01$. Deste modo encontraremos que se $n \geq \frac{1}{4(0{,}05)^2 0{,}01} = 10.000$, a condição exigida será satisfeita.

A hipótese de variâncias finitas pode ser eliminada e o próximo teorema prova uma versão da Lei Fraca dos Grandes Números para variáveis aleatórias i.i.d. e integráveis.

Teorema 4.2.6: Lei Fraca de Khintchin. *Se $X_1, X_2, \ldots$ são i.i.d. e integráveis com média comum μ, então*
$$\frac{S_n}{n} \to^P \mu.$$

Prova: É conseqüência da Lei Forte de Kolmogorov e do fato que convergência quase certa implica convergência em probabilidade. ∎

Exemplo 4.2.7: Sejam $\{X_n : n \geq 1\}$ variáveis i.i.d. com média μ e variância σ^2, ambas finitas. Prove que $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \to^P \sigma^2$.

Solução: Note que

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \to^P \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 + \frac{1}{n}\sum_{i=1}^{n} -2\overline{X}X_i + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}^2$$
$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

Pela Lei Fraca dos Grandes Números, temos que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to^P EX_i^2 = \sigma^2 + \mu^2$$

e

$$\overline{X} \to^P EX_i = \mu.$$

Como funções contínuas preservam convergência, temos que

$$\overline{X}^2 \to^P (EX_i)^2 = \mu^2$$

e, consequentemente,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \to^P \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

4.3 Lei Forte dos Grandes Números

Antes de iniciarmos a prova da Lei Forte dos Grandes Números, vamos estudar um critério para integrabilidade de variáveis aleatórias.

Lema 4.3.1: *Seja X uma variável aleatória qualquer. Então,*

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

e, portanto, X é integrável se, e somente se, $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$.

Prova: Primeiro vamos considerar uma variável aleatória Y que assume valores inteiros não-negativos. Neste caso, temos que

$$EY = \sum_{k=1}^{\infty} kP(Y = k) = \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(Y = k),$$

trocando a ordem dos somatórios:

$$EY = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(Y = k) = \sum_{j=1}^{\infty} P(Y \geq j). \tag{4.2}$$

Se $x \geq 0$, seja $\lfloor x \rfloor$ a parte inteira de x. Então, a variável aleatória $\lfloor |X| \rfloor$ assume o valor k quando $k \leq |X| < k + 1$ e $0 \leq \lfloor |X| \rfloor \leq |X| \leq \lfloor |X| \rfloor + 1$, então pela monotonicidade e linearidade da esperança temos:

$$0 \leq E\lfloor |X| \rfloor \leq E|X| \leq 1 + E\lfloor |X| \rfloor.$$

Como $\lfloor |X| \rfloor$ é uma variável aleatória que só assume valores inteiros não-negativos, temos

$$E\lfloor |X| \rfloor = \sum_{n=1}^{\infty} P(\lfloor |X| \rfloor \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$$

logo

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

∎

Nosso próximo passo é provar uma extensão da desigualdade de Chebyshev.

Lema 4.3.2: *Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tais que $EX_k = 0$ e $VarX_k < \infty, k = 1, \ldots, n$. Então, para todo $\lambda > 0$,*

$$P(\max_{1 \leq k \leq n} |S_k| \geq \lambda) \leq \frac{1}{\lambda^2} VarS_n = \frac{1}{\lambda^2} \sum_{k=1}^{n} VarX_k,$$

onde $S_k = X_1 + \ldots + X_k$.

Prova: Queremos uma cota superior para $P(\max_{1\leq k\leq n} S_k^2 \geq \lambda^2)$. Para tanto, seja $A = [\max_{1\leq k\leq n} S_k^2 \geq \lambda^2]$. Vamos decompor A conforme a primeira vez que $S_k^2 \geq \lambda^2$, definamos:

$$\begin{aligned}
&A_1 = [S_1^2 \geq \lambda^2],\\
&A_2 = [S_1^2 < \lambda^2, S_2^2 \geq \lambda^2],\\
&A_k = [S_1^2 < \lambda^2, \ldots, S_{k-1}^2 < \lambda^2, S_k^2 \geq \lambda^2], \text{ para } 2 \leq k \leq n.
\end{aligned}$$

Então os A_k são disjuntos e $A = \cup_{k=1}^n A_k$. Logo, $I_A = \sum_{k=1}^n I_{A_k}$ e

$$S_n^2 \geq S_n^2 I_A = \sum_{k=1}^n S_n^2 I_{A_k} \Rightarrow ES_n^2 \geq \sum_{k=1}^n ES_n^2 I_{A_k}.$$

Queremos substituir S_n^2 por S_k^2 no somatório (pois $S_k^2 \geq \lambda^2$ em A_k, e não vale necessariamente $S_n^2 \geq \lambda^2$); o truque é escrever

$$S_n^2 = (S_n - S_k)^2 + S_k^2 + 2(S_n - S_k)S_k \geq S_k^2 + 2(S_n - S_k)S_k.$$

Portanto,

$$ES_n^2 I_{A_k} \geq ES_k^2 I_{A_k} + 2E((S_n - S_k)S_k I_{A_k}).$$

Como $S_n - S_k = X_{k+1} + \ldots + X_n$ e $S_k I_{A_k}$ depende só de $X_1, \ldots, X_k$, as duas são funções de famílias disjuntas de variáveis independentes, logo são independentes e a esperança fatora:

$$E((S_n - S_k)S_k I_{A_k}) = E(S_n - S_k)E(S_k I_{A_k}).$$

Como $E(S_n - S_k) = 0$, temos

$$ES_n^2 I_{A_k} \geq ES_k^2 I_{A_k} \geq E\lambda^2 I_{A_k} = \lambda^2 P(A_k). \tag{4.3}$$

Portanto,

$$ES_n^2 \geq \sum_{k=1}^n \lambda^2 P(A_k) = \lambda^2 P(A),$$

logo

$$P(A) \leq \frac{1}{\lambda^2} ES_n^2 = \frac{1}{\lambda^2} Var S_n.$$

∎

O próximo teorema é conhecido como **Primeira Lei Forte de Kolmogorov**.

Teorema 4.3.3: *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e integráveis, e suponha que*

$$\sum_{n=1}^{\infty} \frac{Var X_n}{n^2} < \infty.$$

Então, as X_n satisfazem a Lei Forte dos Grandes Números, ou seja,

$$\frac{X_1 + \ldots + X_n}{n} - \frac{(EX_1 + \ldots + EX_n)}{n} \to 0 \textit{ quase certamente.}$$

Prova: Suponhamos sem perda de generalidade que $EX_n = 0, \forall n$. Queremos mostrar que $\frac{S_n}{n} \to 0$ cp1, onde $S_n = X_1 + \ldots + X_n$. Para tanto, basta mostrar que

$$M_n = \max_{2^n < k \leq 2^{n+1}} \frac{|S_k|}{k} \to 0 \text{ cp1 quando } n \to \infty.$$

Provaremos isto em duas etapas:

(i) $\sum_{n=1}^{\infty} P(M_n \geq \frac{1}{m}) < \infty, \forall m = 1, 2, \ldots$; e

(ii) $M_n \to 0$ cp1.

Para (i), considere m fixo. Então, para todo n,

$$P(M_n \geq \frac{1}{m}) \leq P(\max_{2^n < k \leq 2^{n+1}} |S_k| \geq \frac{2^n}{m}) \leq$$
$$\leq P(\max_{1 < k \leq 2^{n+1}} |S_k| \geq \frac{2^n}{m}) \leq \frac{m^2}{4^n} \sum_{k=1}^{2^{n+1}} Var(X_k),$$

onde vale a última passagem pelo lema anterior. Seja $A_n = [M_n \geq \frac{1}{m}]$, então

$$\sum_{n=1}^{\infty} P(A_n) \leq m^2 \sum_{n=1}^{\infty} (\frac{1}{4^n} \sum_{k=1}^{2^{n+1}} Var(X_k)) = m^2 \sum_{k=1}^{\infty} \sum_{n:2^{n+1} \geq k} (\frac{1}{4^n} Var(X_k)) =$$
$$= m^2 \sum_{k=1}^{\infty} Var(X_k) \sum_{n:2^{n+1} \geq k} (\frac{1}{4^n}).$$

Observe que

$$\sum_{n:2^{n+1} \geq k} (\frac{1}{4^n}) = \sum_{n:n+1 \geq \log_2 k} (\frac{1}{4^n}) = \sum_{n=\lceil \log_2 k - 1 \rceil} (\frac{1}{4^n})$$
$$\frac{(1/4)^{\lceil \log_2 k - 1 \rceil}}{1 - 1/4} \leq \frac{(1/4)^{\log_2 k - 1}}{3/4} \leq \frac{16}{3k^2}.$$

Portanto,

$$\sum_{n=1}^{\infty} P(A_n) \leq \frac{16m^2}{3} \sum_{k=1}^{\infty} \frac{Var(X_k)}{k^2} < \infty.$$

Para (ii), note que por Borel-Cantelli, tem-se $P(A_n$ infintas vezes$) = 0$. Logo, para todo m, a probabilidade é 1 de que M_n assuma um valor $\geq \frac{1}{m}$ para somente um número finito de n's. Seja B_m o evento "M_n assuma um valor $\geq \frac{1}{m}$ para somente um número finito de n's", então $P(B_m) = 1, \forall m$, o que implica que $P(\cap_{m=1}^{\infty} B_m) = 1$, e (ii) resulta da equivalência entre os eventos $\cap_{m=1}^{\infty} B_m$ e $[M_n \to 0]$. ∎

O próximo exemplo ilustra uma aplicação da Primeira Lei Forte de Kolmogorov.

Exemplo 4.3.4: Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes com $X_n \sim Poisson(\sqrt{n})$, para cada $n \geq 1$. Calcule o limite quase-certo de $\overline{X}$.

Solução: Como $VarX_n = \sqrt{n}$, temos que

$$\sum_{n=1}^{\infty} \frac{VarX_n}{n^2} = \sum_{n=1}^{\infty} \frac{\sqrt{n}}{n^2} < \infty.$$

Logo, a primeira Lei Forte de Kolmogorov implica que

$$\overline{X} - \frac{EX_1 + \cdots + EX_n}{n} \to 0 \text{ cp1, ou seja}$$

$$\overline{X} - \frac{\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}}{n} \to 0 \text{ cp1.}$$

Pelo teste da integral, pode-se verificar que

$$\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n} \geq \int_0^n \sqrt{x}dx = \frac{2n^{3/2}}{3}.$$

Portanto,

$$\frac{\sqrt{1} + \sqrt{2} + \cdots + \sqrt{n}}{n} \geq \frac{2n^{1/2}}{3} \to \infty.$$

Logo, $\overline{X} \to \infty$ cp1.

Antes de enunciarmos e provarmos a Segunda Lei Forte de Kolmogorov, considere o seguinte lema:

Lema 4.3.5: *Seja X uma variável aleatória integrável com função de distribuição F. Então,*

$$\sum_{n=1}^{\infty} (\frac{1}{n^2} \int_{-n}^{n} x^2 dF(x)) < \infty.$$

Prova: Pelo teste da integral, note que para $j = 1, 2, \ldots$, temos

$$\begin{aligned} \sum_{n=j}^{\infty} \frac{1}{n^2} &= \frac{1}{j^2} + \sum_{n=j+1}^{\infty} \frac{1}{n^2} \\ &\leq \frac{1}{j^2} + \int_j^{\infty} \frac{1}{x^2} dx = \frac{1}{j^2} + \frac{1}{j} \leq \frac{2}{j}. \end{aligned}$$

Como

$$\int_{-n}^{n} x^2 dF(x) = \sum_{j=-n+1}^{n} \int_{j-1}^{j} x^2 dF(x),$$

temos

$$\sum_{n=1}^{\infty}(\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)) = \sum_{n=1}^{\infty}\sum_{j=-n+1}^{n}(\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)) =$$
$$= \sum_{j=1}^{\infty}\sum_{n=j}^{\infty}(\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)) + \sum_{j=-\infty}^{0}\sum_{n=|j|+1}^{\infty}(\frac{1}{n^2}\int_{j-1}^{j} x^2 dF(x)) \leq$$
$$\leq 2\sum_{j=1}^{\infty}\int_{j-1}^{j}\frac{x^2}{j}dF(x) + 2\sum_{j=-\infty}^{0}\int_{j-1}^{j}\frac{x^2}{|j|+1}dF(x).$$

Como $\frac{x^2}{j} \leq x$ em $(j-1,j]$, para $j \geq 1$, e $\frac{x^2}{|j|+1} \leq |x|$ em $(j-1,j]$, para $j \leq 0$, temos

$$\sum_{n=1}^{\infty}(\frac{1}{n^2}\int_{-n}^{n} x^2 dF(x)) \leq 2\sum_{j=1}^{\infty}\int_{j-1}^{j} x dF(x) + 2\sum_{j=-\infty}^{0}\int_{j-1}^{j} |x| dF(x) =$$
$$= 2\sum_{j=-\infty}^{\infty}\int_{j-1}^{j} |x| dF(x) = 2\int_{-\infty}^{\infty} |x| dF(x) = 2E|X| < \infty. \tag{4.4}$$

∎

A seguir enunciamos e provamos a **Segunda Lei Forte de Kolmogorov**.

Teorema 4.3.6: *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com $EX_n = \mu$. Então,*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \textit{ quase certamente.}$$

Prova: Suponhamos sem perda de generalidade que $\mu = 0$. Vamos truncar as variáveis X_n, definamos $Y_n = X_n I_{[-n<X_n\leq n]}$. Seja $Z_n = X_n - Y_n$, de modo que

$$\frac{X_1 + \ldots + X_n}{n} = \frac{Y_1 + \ldots + Y_n}{n} + \frac{Z_1 + \ldots + Z_n}{n}.$$

A prova terá três partes:

(a) $\frac{Z_1+\ldots+Z_n}{n} \to 0$ quase certamente (usaremos Borel-Cantelli);

(b) $\frac{Y_1+\ldots+Y_n}{n} - \frac{EY_1+\ldots+EY_n}{n} \to 0$ quase certamente (usaremos a Primeira Lei Forte e o Lema 4.3.5); e

(c) $\frac{EY_1+\ldots+EY_n}{n} \to 0$ (usaremos o Teorema da Convergência Dominada).

É fácil ver que (a), (b), e (c) implicam o teorema. Para provar (a), note que $Z_n \neq 0 \Leftrightarrow Y_n \neq X_n \Leftrightarrow X_n \notin (-n,n]$. Logo,

$$P(Z_n \neq 0) = P(X_n \notin (-n,n]) \leq P(|X_n| \geq n).$$

Mas os eventos $A_n = [Z_n \neq 0]$ satisfazem

$$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n) \leq E|X_1| < \infty.$$

Portanto, Borel-Cantelli implica que $P(A_n$ infinitas vezes$) = 0$, ou seja

$$P(Z_n \neq 0 \text{ infinitas vezes}) = 0.$$

Isso significa que

$$P(Z_n = 0 \text{ para todo } n \text{ suficientemente grande}) = 1.$$

Mas se $Z_n = 0$ para n suficientemente grande, então $Z_n \to 0$ e

$$\frac{Z_1 + \ldots + Z_n}{n} \to 0, \text{ logo } P(\frac{Z_1 + \ldots + Z_n}{n} \to 0) = 1.$$

Para provar (b), seja F a função de distribuição comum, $F = F_{X_n}$. Verifiquemos a condição da primeira Lei Forte de Kolmogorov para as variáveis aleatórias Y_n. Como $Y_n = X_n I_{[-n<X_n\leq n]}$, temos

$$Var(Y_n) \leq E(Y_n^2) = E(X_n^2 I_{[-n<X_n\leq n]}) = \int_{-n}^{n} x^2 dF(x).$$

Portanto,

$$\sum_{n=1}^{\infty} \frac{Var(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-n}^{n} x^2 dF(x) < \infty,$$

onde a última desigualdade decorre do Lema 4.3.5. Portanto, (b) decorre da primeira Lei Forte de Kolmogorov.

Para provar (c), é suficiente mostrar que $EY_n \to 0$. Mas,

$$EY_n = E(X_n I_{[-n<X_n\leq n]}) = E(X_1 I_{[-n<X_1\leq n]}) \to EX_1 = 0,$$

pelo teorema da convergência dominada que se aplica pois $|X_1|$ domina $X_1 I_{[-n\leq X_1\leq n]}$'s e é integrável. ∎

Exemplo 4.3.7: As variáveis X_n, $n \geq 1$, são independentes e todas têm distribuição Exponencial de parâmetro λ. Mostre que a seqüência $\{X_n^2 : n \geq 1\}$ satisfaz a Lei Forte dos Grandes Números.

Solução: De acordo com a Segunda Lei Forte de Kolmogorov, precisamos mostrar que EX_n^2 é finita para todo n. Como $EX_n^2 = VarX_n + (EX_n)^2 = \frac{2}{\lambda^2} < \infty$, temos que a seqüência $\{X_n^2 : n \geq 1\}$ satisfaz a Lei Forte dos Grandes Números.

Exemplo 4.3.8: Seja $\{X_n : n \geq 1\}$ uma seqüência de variáveis aleatórias i.i.d., seguindo o modelo Uniforme contínuo em $(0, 1)$. Calcule o limite, quase certo, para $\frac{1}{n}\sum_{k=1}^{n}(-\log(X_k))$ quando $n \to \infty$.

Solução: Vamos tentar usar a Lei Forte dos Grandes Números. Para isso, precisamos calcular $E(-\log X_k)$.

$$E(-\log X_k) = \int_0^1 -\log x dx = -x\log x|_0^1 + \int_0^1 dx = 1.$$

Portanto, temos que $\frac{1}{n}\sum_{k=1}^{n}(-\log(X_k)) \to 1$ cp1.

Por fim nós enunciaremos e provaremos a **Recíproca da Lei Forte de Kolmogorov**. A Lei Forte afirma que se as variáveis aleatórias X_n são integráveis, então $\frac{S_n}{n}$ converge para um limite finito $(= EX_1)$ com probabilidade 1. A recíproca diz que se as X_n não forem integráveis, então com probabilidade 1, $\frac{S_n}{n}$ não convergirá para um limite finito.

Teorema 4.3.9: *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas. Se $E|X_1| = \infty$, então, com probabilidade 1, a seqüência $\frac{|S_n|}{n}$ não é limitada.*

Prova: Se $E|X_1| = \infty$, então $E(\frac{|X_1|}{k}) = \infty$, para $k = 1, 2, \ldots$. De acordo com Lema 4.3.1, temos que

$$\sum_{n=1}^{\infty} P(\frac{|X_1|}{k} \geq n) = \infty, \forall k.$$

Como as variáveis X_n são identicamente distribuídas, temos

$$\sum_{n=1}^{\infty} P(\frac{|X_1|}{k} \geq n) = \sum_{n=1}^{\infty} P(\frac{|X_n|}{k} \geq n) = \sum_{n=1}^{\infty} P(\frac{|X_n|}{n} \geq k).$$

Por independência dos X_n, os eventos $A_n = [\frac{|X_n|}{n} \geq k]$ são independentes, e Borel-Cantelli implica

$$P(\frac{|X_n|}{n} \geq k \text{ infinitas vezes}) = 1, \forall k.$$

Fazendo $B_k = [\frac{|X_n|}{n} \geq k$ infinitas vezes$]$, temos $P(\cap_{k=1}^{\infty} B_k) = 1$, pois a intersecção de um número enumerável de eventos de probabilidade 1 também tem probabilidade 1. Mas o evento $\cap_{k=1}^{\infty} B_k$ é o evento "$\frac{|X_n|}{n} > k$ para um número infinito de n, para todo k", ou seja, é o evento "a seqüência $\frac{|X_n|}{n}$ é ilimitada." Para terminar a prova, basta mostrar que se $\frac{|X_n|}{n}$ é ilimitada, então $\frac{|S_n|}{n}$ também é ilimitada. Agora, com $S_0 = 0$, temos

$$\frac{|X_n|}{n} = \frac{|S_n - S_{n-1}|}{n} \leq \frac{|S_n|}{n} + \frac{|S_{n-1}|}{n},$$

para $n = 1, 2, \ldots$. Portanto, se $\frac{|X_n|}{n}$ é ilimitada, então $\frac{|S_n|}{n}$ é ilimitada ou $\frac{|S_{n-1}|}{n}$ é ilimitada. Mas,

$$\frac{|S_{n-1}|}{n} = \frac{|S_{n-1}|}{(n-1)} \cdot \frac{(n-1)}{n},$$

então $\frac{|S_{n-1}|}{n}$ é ilimitada se, e somente se, $\frac{|S_n|}{n}$ também for. ∎

4.4 Um Exemplo de Divergência das Médias

Uma variável aleatória tem distribuição de Cauchy de parâmetro a se, para $a > 0$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2}.$$

Assuma que X_n são i.i.d. segundo uma distribuição de Cauchy de parâmetro a. Seja $S_n = \frac{1}{n}\sum_{i=1}^{n} X_n$. Utilizando a definição e as propriedades da função característica pode-se provar que

$$\varphi_{X_n}(u) = e^{-a|u|}, \text{ e } \varphi_{S_n}(u) = e^{-a|u|}.$$

Então, as médias S_n são distribuídas exatamente como uma das parcelas da soma. Para $n \geq m$, após alguma manipulação algébrica, temos que

$$S_n - S_m = (1 - \frac{m}{n})([Z_{n,m}] - [Y_{n,m}]),$$

onde $Z_{n,m} = \frac{1}{n-m}\sum_{i=m+1}^{n} X_i$ e $Y_{n,m} = \frac{1}{m}\sum_{i=1}^{m} X_i$. Observe que como $Z_{n,m}$ e $Y_{n,m}$ são médias de conjuntos disjuntos de variáveis aleatórias independentes, elas são independentes uma da outra. Ainda mais, pelo resultado para φ_{S_n}, é o caso que elas são identicamente distribuídas com função característica igual a $e^{-a|u|}$. Seja $W_{n,m} = Z_{n,m} - Y_{n,m}$, nós vemos que $S_n - S_m = (1 - \frac{m}{n})W_{n,m}$. Contudo,

$$\varphi_{W_{n,m}}(u) = \varphi_{Z_{n,m}}(u)\varphi_{Y_{n,m}}(-u) = e^{-2a|u|}.$$

Então, $W_{n,m}$ tem uma distribuição fixa, não degenerada que é independente de n e m. Fixando, $n = 2m$, temos que

$$\varphi_{S_{2m}-S_m}(u) = e^{-a|u|}.$$

Portanto, quando $m \to \infty$, $S_{2m} - S_m$ não converge para zero, mas para todo m, tem uma distribuição Cauchy de parâmetro a. Portanto, S_n não satisfaz o critério de convergência de Cauchy e não é convergente.

Observe que isto não é um contra-exemplo a Lei Forte de Kolmogorov, tendo em vista que uma variável aleatória que tem distribuição de acordo com uma Cauchy não tem valor esperado definido, ou seja

$$EX = -\int_{-\infty}^{0} \frac{1}{\pi}\frac{a|x|}{a^2 + x^2}dx + \int_{0}^{\infty} \frac{1}{\pi}\frac{ax}{a^2 + x^2}dx,$$

é indefinido, visto que ambas as integrais são infinitas. Este exemplo serve para ilustrar que a suposição da existência de EX é necessária para a Lei Forte dos Grandes Números.

Capítulo 5

Teorema Central do Limite

5.1 Motivação

Consideremos uma seqüência de variáveis aleatórias independentes, $X_1, X_2, \ldots$, definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, e seja $S_1, S_2, \ldots$ a seqüência de somas parciais, definidas por $S_n = X_1 + X_2 + \ldots + X_n$. A Lei dos Grandes Números trata da convergência de $\frac{1}{n}(S_n - ES_n)$ para 0, quando $n \to \infty$, supondo que as variáveis aleatórias X_i's sejam integráveis. Quando a seqüência obedece à lei dos grandes números, existe uma tendência da variável aleatória $\frac{S_n}{n}$, a média amostral no caso de variáveis aleatórias independentes e identicamente distribuídas, para concentrar-se em torno de sua média. O Teorema Central do Limite prova que sob certas hipóteses gerais, a distribuição da média amostral padronizada tende à normal. O problema consiste em achar condições sob as quais

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} \to^D N(0, 1).$$

Resumidamente, estas condições exigem que cada parcela da soma contribua com um valor sem importância para a variação da soma, ou seja é muito improvável que qualquer parcela isolada dê uma contribuição muito grande para a soma.

O Teorema Central do Limite dá apoio ao uso da normal como distribuição de erros, pois em muitas situações reais é possível interpretar o erro de uma observação como resultante de muitos erros pequenos e independentes. Há também outras situações que o Teorema Central do Limite pode justificar o uso da normal. Por exemplo, a distribuição de alturas de homens adultos de certa idade pode ser considerada aproximadamente normal, pois a altura pode ser pensada como soma de muitos efeitos pequenos e independentes.

5.2 Teoremas e provas

Existem vários Teoremas Centrais do Limite que variam de acordo com as hipóteses sobre as distribuições das variáveis aleatórias X_i's na seqüência. Como teoremas centrais do limite tratam de convergência em distribuição e como, pelo Teorema da Continuidade de Levy, sabe-se que uma seqüência de variáveis aleatórias $Y_n \to^D Y$ se, e somente se, $\varphi_{Y_n} \to \varphi_Y$, a idéia será provar que a função característica de $\frac{S_n - ES_n}{\sqrt{VarS_n}}$ converge para $e^{\frac{-t^2}{2}}$ que é a função característica da $N(0,1)$. Nós iremos agora enunciar e provar alguns desses teoremas, começando pelo caso de variáveis aleatórias independentes e identicamente distribuídas.

Teorema 5.2.1: *Sejam $X_1, X_2, \ldots$ variáveis aleatórias iid com $E(X_n) = \mu$ e $Var(X_n) = \sigma^2$. Suponha que N é uma variável aleatória com distribuição $N(0,1)$. Se $S_n = X_1 + X_2 + \ldots + X_n$, então*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to^D N.$$

Prova: Sem perda de generalidade, seja $E(X_n) = 0$ e $E(X_n^2) = 1$ (caso este não seja o caso, pode-se provar o resultado para

$$X_i^* = \frac{X_i - \mu}{\sigma},$$

já que $E(X_i^*) = 0$ e $E(X_i^*)^2 = 1$).

Seja $\varphi_n(t) = E(e^{it\frac{S_n}{\sqrt{n}}})$ e $\varphi(t) = E(e^{itX_1})$. Como a função característica de uma soma de variáveis aleatórias independentes é igual ao produto das funções características das variáveis aleatórias, tem-se que

$$\varphi_n(t) = (E(e^{it\frac{X_1}{\sqrt{n}}}))^n = \varphi^n(t/\sqrt{n}).$$

Como os dois primeiros momentos existem, φ possui duas derivadas contínuas. Então, utilizando a expansão de Taylor de φ e o fato que $\varphi^{(k)}(0) = i^k E(X_1^k)$, temos que

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

onde $|\theta(t)| \leq |t|$. Logo, como φ'' é contínua em 0, temos que $\varphi''(\theta(t)) - \varphi''(0) \to 0$ quando $t \to 0$. Então, tem-se

$$\varphi(t) = 1 - \frac{t^2}{2} + \frac{t^2}{2}e(t),$$

onde $e(t) = \varphi''(\theta(t)) + 1$ e $\lim_{t\to 0} e(t) = 0$. Então, para t fixo

$$\varphi^n(\frac{t}{\sqrt{n}}) = [1 - \frac{t^2}{2n} + \frac{t^2}{2n}e(\frac{t}{\sqrt{n}})]^n = [1 + \frac{-t^2}{2n}[1 - e(\frac{t}{\sqrt{n}})]]^n \to e^{\frac{-t^2}{2}},$$

quando $n \to \infty$, pois $[1 - e(\frac{t}{\sqrt{n}})] \to 1$ e para números complexos $c_n \to c \Rightarrow (1 + \frac{c_n}{n})^n \to e^c$ (Esse limite é conhecido como limite de Euler e sua prova será omitida). ∎

Um caso especial do Teorema Central do Limite para variáveis aleatórias independentes e identicamente distribuídas é quando estas variáveis são distribuídas de acordo com a distribuição de Bernoulli, este caso é conhecido como Teorema Central do Limite de De Moivre e Laplace.

Corolário 5.2.2: *Seja $X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias independentes e distribuídas de acordo com a distribuição de Bernoulli com parâmetro p, ou seja, $P(X_i = 1) = p = 1 - P(X_i = 0)$ para $0 < p < 1$. Então, se $S_n = X_1 + \ldots X_n$,*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \to^D N(0,1).$$

Prova: É imediata dado o teorema anterior, já que $E(X_i) = p$ e $E(X_i^2) = p$. ∎

Exemplo 5.2.3: Suponha que temos algumas voltagens de ruídos independentes, por exemplo $V_i, i = 1, 2, \ldots, n$, as quais são recebidas naquilo que se denomina um "somador". Seja V a soma das voltagens recebidas. Suponha também que cada variável aleatória V_i seja uniformemente distribuída sobre o intervalo [0,10]. Daí, $EV_i = 5$ volts e $VarV_i = \frac{100}{12}$. De acordo com o Teorema Central do Limite, se n for suficientemente grande, a variável aleatória

$$S = \frac{(V - 5n)\sqrt{12}}{10\sqrt{n}}$$

terá aproximadamente a distribuição $N(0, 1)$. Portanto, se $n = 20$, podemos calcular que a probabilidade de que a voltagem total na entrada exceda 105 volts da seguinte maneira:

$$P(V > 105) = P(\frac{(V - 100)\sqrt{12}}{10\sqrt{20}} > \frac{(105 - 100)\sqrt{12}}{10\sqrt{20}}) \backsimeq 1 - \Phi(0,388) = 0,352.$$

Agora analisaremos um resultado mais forte que dá condições gerais que garantem convergência da média amostral padronizada para normal: o Teorema Central do Limite de Lindeberg.

Teorema 5.2.4: *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $E(X_n) = \mu_n$ e $Var(X_n) = \sigma_n^2 < \infty$, onde pelo menos um $\sigma_i^2 > 0$. Sejam $S_n = X_1 + \ldots + X_n$ e $s_n = \sqrt{Var(S_n)} = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$. Considere a seguinte condição, conhecida como condição de Lindeberg,*

$$\forall \epsilon > 0, \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\epsilon s_n} (x - \mu_k)^2 dF_k(x) = 0.$$

Então, se a condição de Lindeberg é satisfeita

$$\frac{S_n - ES_n}{s_n} \to^D N(0, 1).$$

Antes de provarmos este teorema, vamos primeiro dar alguma intuição sobre a condição de Lindeberg. Esta condição diz que, para n grande, a parcela da variância devida às caudas das X_k é desprezível.

A condição de Lindeberg implica que as parcelas X_k da soma têm variâncias uniformemente pequenas para n grande, em outras palavras nenhuma parcela tem muito peso na soma. Formalmente, a condição de Lindeberg implica que $\max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2} \to 0$ quando $n \to \infty$. Para ver isto, observe que para todo k,

$$\frac{\sigma_k^2}{s_n^2} = \frac{1}{s_n^2} \int_{|x-\mu_k|\leq\epsilon' s_n} (x - \mu_k)^2 dF_k(x) + \frac{1}{s_n^2} \int_{|x-\mu_k|>\epsilon' s_n} (x - \mu_k)^2 dF_k(x) \leq$$

$$\frac{1}{s_n^2} \int_{|x-\mu_k|\leq\epsilon' s_n} (\epsilon' s_n)^2 dF_k(x) + \frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x-\mu_j|>\epsilon' s_n} (x - \mu_j)^2 dF_j(x) \leq$$

$$\frac{1}{s_n^2} \int_{-\infty}^{\infty} (\epsilon' s_n)^2 dF_k(x) + \frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x-\mu_j|>\epsilon' s_n} (x - \mu_j)^2 dF_j(x).$$

Este último termo não depende de k, pois a primeira parcela é igual a $(\epsilon')^2$. Portanto, temos

$$\max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2}\leq(\epsilon')^2+\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon' s_n}(x-\mu_k)^2dF_k(x),$$

que converge para $(\epsilon')^2$, pela condição de Lindeberg. Como isto vale para todo ϵ', temos $\max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2}\to 0$.

Portanto, o Teorema Central do Limite de Lindeberg pode ser aplicado para justificar o seguinte raciocínio: a soma de um grande número de pequenas quantidades independentes tem aproximadamente uma distribuição normal.

Exemplo 5.2.5: Vamos verificar neste exemplo que uma seqüência $X_1, X_2, \ldots$ de variáveis aleatórias i.i.d. com $EX_i=\mu$ e $VarX_i=\sigma^2$ satisfaz a condição de Lindeberg. Note que $s_n=\sqrt{VarS_n}=\sigma\sqrt{n}$. Então para $\epsilon>0$, e F a distribuição comum das variáveis aleatórias:

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n}(x-\mu_k)^2dF_k(x)=\frac{1}{n\sigma^2}\sum_{k=1}^{n}\int_{|x-\mu|>\epsilon\sigma\sqrt{n}}(x-\mu)^2dF(x)$$
$$=\frac{1}{n\sigma^2}n\int_{|x-\mu|>\epsilon\sigma\sqrt{n}}(x-\mu)^2dF(x).$$

Então, finalmente,

$$\lim_n\frac{1}{\sigma^2}\int_{|x-\mu|>\epsilon\sigma\sqrt{n}}(x-\mu)^2dF(x)=0.$$

Agora iremos provar o Teorema Central do Limite de Lindeberg.

Prova: Assim como no caso de variáveis aleatórias i.i.d., mostraremos que a função característica de $\frac{S_n-ES_n}{s_n}$ converge para $e^{\frac{-t^2}{2}}$.

Para tanto, fixemos $t\in R$. Usaremos duas versões da fórmula de Taylor aplicada à função $g(x)=e^{itx}$:

$$e^{itx}=1+itx+\theta_1(x)\frac{t^2x^2}{2},\text{ onde }|\theta_1(x)|\leq 1$$

e

$$e^{itx}=1+itx-\frac{t^2x^2}{2}+\theta_2(x)\frac{t^3x^3}{6},\text{ onde }|\theta_2(x)|\leq 1.$$

Seja $\epsilon>0$. Usando a primeira fórmula para $|x|>\epsilon$ e a segunda para $|x|\leq\epsilon$, podemos escrever e^{itx} da seguinte forma geral:

$$e^{itx}=1+itx-\frac{t^2x^2}{2}+r_\epsilon(x),$$

onde

$$r_\epsilon(x)=\begin{cases}(1+\theta_1(x))\frac{t^2x^2}{2} & \text{se } |x|>\epsilon,\\ \theta_2(x)\frac{t^3x^3}{6} & \text{se } |x|\leq\epsilon.\end{cases}$$

Portanto,

$$E(e^{it\frac{X_k-\mu_k}{s_n}}) = \int e^{it\frac{x-\mu_k}{s_n}} dF_k(x) = \int (1 + it\frac{x-\mu_k}{s_n} - \frac{t^2(\frac{x-\mu_k}{s_n})^2}{2} +$$
$$+r_\epsilon(\frac{x-\mu_k}{s_n}))dF_k(x) = 1 + itE(\frac{X_k-\mu_k}{s_n}) - \frac{t^2}{2}E((\frac{X_k-\mu_k}{s_n})^2) +$$
$$+\frac{t^2}{2}\int_{|x-\mu_k|>\epsilon s_n} (1+\theta_1(\frac{x-\mu_k}{s_n}))(\frac{x-\mu_k}{s_n})^2 dF_k(x) +$$
$$\frac{t^3}{6}\int_{|x-\mu_k|\leq\epsilon s_n} \theta_2(\frac{x-\mu_k}{s_n})(\frac{x-\mu_k}{s_n})^3 dF_k(x).$$

Como $EX_k = \mu_k$ e $Var(X_k) = \sigma_k^2$, temos

$$E(e^{it\frac{X_k-\mu_k}{s_n}}) = 1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k},$$

onde o resto $e_{n,k}$ satisfaz

$$|e_{n,k}| \leq t^2\int_{|x-\mu_k|>\epsilon s_n} (\frac{x-\mu_k}{s_n})^2 dF_k(x) + \frac{|t^3|}{6}\int_{|x-\mu_k|\leq\epsilon s_n} \epsilon(\frac{x-\mu_k}{s_n})^2 dF_k(x)$$
$$\leq \frac{t^2}{s_n^2}\int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t^3|}{6s_n^2}\int_{-\infty}^{\infty} (x-\mu_k)^2 dF_k(x).$$

Temos então,

$$\sum_{k=1}^{n} |e_{n,k}| \leq \frac{t^2}{s_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) + \frac{\epsilon|t^3|}{6}.$$

Pela condição de Lindeberg, a primeira parcela do termo à direita tende a zero quando $n \to \infty$. Logo, para n suficientemente grande,

$$\sum_{k=1}^{n} |e_{n,k}| \leq \frac{\epsilon|t|^3}{3}.$$

Vamos então escolher uma seqüência de ϵ's que converge para zero. Para $\epsilon = \frac{1}{m}$, existe n_m tal que para $n \geq n_m$,

$$\sum_{k=1}^{n} |e_{n,k}| \leq \frac{|t^3|}{3m}, \tag{5.1}$$

onde os restos $e_{n,k}$ são os determinados pela fórmula baseada em $\epsilon = \frac{1}{m}$. Portanto, existe uma seqüência de inteiros positivos $n_1 < n_2 < \ldots$ tal que (5.1) é satisfeita para $n_m \leq n < n_{m+1}$, onde para estes valores de n os restos são baseados em $\epsilon = \frac{1}{m}$. É importante lembrar durante o restante da prova que o valor de ϵ que determina o resto $e_{n,k}$ depende da posição de n em relação aos n_m. Temos, então,

$$\sum_{k=1}^{n} |e_{n,k}| \to 0 \text{ quando } n \to \infty.$$

Como X_i's são independentes,

$$\varphi_{\frac{S_n - ES_n}{s_n}}(t) = \prod_{k=1}^{n} E(e^{it\frac{X_k-\mu_k}{s_n}}) = \prod_{k=1}^{n}(1 - \frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}).$$

Para provar que o termo à direita converge para $e^{\frac{-t^2}{2}}$, usaremos o seguinte Lema sobre números complexos.

Lema 5.2.6: *Sejam $c_{n,k}$ números complexos tais que $\sum_{k=1}^{n} c_{n,k} \to c$ quando $n \to \infty$. Se*

$$\max_{1 \leq k \leq n} |c_{n,k}| \to 0 \text{ quando } n \to \infty$$

e

$$\sum_{k=1}^{n} |c_{n,k}| \leq M < \infty,$$

onde M é uma constante que não depende de n, então

$$\prod_{k=1}^{n}(1 + c_{n,k}) \to e^c \text{ quando } n \to \infty.$$

Prova: Nós omitimos a prova deste lema que pode ser encontrada no livro do Chung seção 7.1. ∎

Em nosso caso, sejam $c_{n,k} = -\frac{t^2\sigma_k^2}{2s_n^2} + e_{n,k}$ e $c = \frac{-t^2}{2}$. Temos que

$$\sum_{k=1}^{n} |c_{n,k}| \leq \frac{t^2}{2} + \sum_{k=1}^{n} |e_{n,k}| \to \frac{t^2}{2},$$

logo existe $M < \infty$ tal que $\forall n$, $\sum_{k=1}^{n} |c_{n,k}| < M$. Para aplicar o lema resta verificar a condição sobre o máximo

$$\max_{1 \leq k \leq n} |c_{n,k}| \leq \max_{1 \leq k \leq n} \frac{t^2\sigma_k^2}{2s_n^2} + \max_{1 \leq k \leq n} |e_{n,k}| \leq \max_{1 \leq k \leq n} \frac{t^2}{2} \frac{\sigma_k^2}{s_n^2} + \max_{1 \leq k \leq n} |e_{n,k}|$$

Como já provamos que os dois termos acima tendem a zero, a prova está terminada. ∎

Corolário 5.2.7: Teorema Central do Limite de Liapunov. *Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes tais que $EX_n = \mu_n$ e $VarX_n = \sigma_n^2 < \infty$ com pelo menos um $\sigma_j^2 > 0$. Seja $S_n = X_1 + \ldots + X_n$ e $s_n^2 = VarS_n$. Se existir $m > 0$ tal que*

$$\frac{1}{s_n^{2+m}} \sum_{k=1}^{n} E(|X_k - \mu_k|^{2+m}) \to 0 \text{ quando } n \to \infty,$$

então,

$$\frac{S_n - ES_n}{s_n} \to^D N(0,1).$$

Prova: Para provar este teorema, é suficiente verificar que as condições do Teorema de Liapunov implicam as condições do Teorema de Lindeberg. A condição de Lindeberg estabelece uma integral na região $|x-\mu_k| > \epsilon s_n, \epsilon > 0$. Nessa região, temos que $\frac{|x-\mu_k|}{\epsilon s_n} > 1$, o que por sua vez implica $\frac{|x-\mu_k|^m}{\epsilon^m s_n^m} > 1$. Desse modo, temos que:

$$\frac{1}{s_n^2}\sum_{k=1}^n \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 dF_k(x) \leq \frac{1}{s_n^2}\sum_{k=1}^n \int_{|x-\mu_k|>\epsilon s_n} (x-\mu_k)^2 \frac{|x-\mu_k|^m}{\epsilon^m s_n^m} dF_k(x)$$
$$= \frac{1}{\epsilon^m s_n^{2+m}}\sum_{k=1}^n \int_{|x-\mu_k|>\epsilon s_n} |x-\mu_k|^{2+m} dF_k(x) \leq \frac{1}{\epsilon^m s_n^{2+m}}\sum_{k=1}^n \int_{-\infty}^{\infty} |x-\mu_k|^{2+m} dF_k(x)$$
$$= \frac{1}{\epsilon^m s_n^{2+m}}\sum_{k=1}^n E|X_k-\mu_k|^{2+m}.$$

Mas a condição de Liapunov implica que o último termo tende a zero quando $n \to \infty$. Portanto, a condição de Lindeberg está satisfeita. ∎

Antes de verficarmos um exemplo do Teorema Central do Limite de Liapunov, vamos considerar o seguinte Lema.

Lema 5.2.8: *Para* $\lambda > 0$,

$$\frac{1}{n^{\lambda+1}}\sum_{k=1}^n k^\lambda \to \frac{1}{\lambda+1},$$

quando $n \to \infty$*, de maneira que* $\sum_{k=1}^n k^\lambda$ *é da ordem de* $n^{\lambda+1}$.

Prova: Como $x^\lambda \leq k^\lambda$ se $k-1 \leq x \leq k$, e $k^\lambda \leq x^\lambda$ se $k \leq x \leq k+1$, segue-se que

$$\int_{k-1}^k x^\lambda dx \leq \int_{k-1}^k k^\lambda dx = k^\lambda = \int_k^{k+1} k^\lambda dx \leq \int_k^{k+1} x^\lambda dx,$$

somando-se em k de 1 até n, temos

$$\int_0^n x^\lambda dx \leq \sum_{k=1}^n k^\lambda \leq \int_1^{n+1} x^\lambda dx.$$

Logo,

$$\frac{n^{\lambda+1}}{\lambda+1} \leq \sum_{k=1}^n k^\lambda \leq \frac{(n+1)^{\lambda+1}-1}{\lambda+1} \leq \frac{(n+1)^{\lambda+1}}{\lambda+1},$$

o que é eqüivalente a

$$\frac{1}{\lambda+1} \leq \frac{1}{n^{\lambda+1}}\sum_{k=1}^n k^\lambda \leq \frac{1}{\lambda+1}\cdot(\frac{n+1}{n})^{\lambda+1}.$$

Como $(\frac{n+1}{n})^{\lambda+1} \to 1$ quando $n \to \infty$, o lema está provado. ∎

Exemplo 5.2.9: Sejam $X_1, X_2, \ldots$, independentes, $X_n \sim U[-n, n]$. Prove que $\frac{S_n - ES_n}{s_n} \rightarrow^D N(0,1)$.

Solução: Vamos verificar a condição de Liapunov para $\delta = 1$. Temos

$$E|X_k - \mu_k|^3 = E|X_k|^3 = \frac{1}{2k}\int_{-k}^{k} |x|^3 dx = \frac{1}{k}\int_0^k x^3 dx = \frac{k^3}{4}.$$

Logo, o Lema anterior implica que $\sum_{k=1}^n E|X_k - \mu_k|^3$ é da ordem de n^4. Vamos determinar a ordem de s_n^3. Como $\mu_k = EX_k = 0$ e

$$\sigma_k^2 = Var(X_k) = EX_k^2 = \frac{1}{2k}\int_{-k}^{k} x^2 dx = \frac{k^2}{3}, \text{ temos}$$

$$s_n^2 = \sum_{k=1}^n \frac{k^2}{3}.$$

Portanto, aplicando o resultado do Lema, temos:

$$\frac{s_n^2}{n^3} \rightarrow \frac{1}{9}.$$

Então,

$$\lim_{n\to\infty} \frac{1}{s_n^3}\sum_{k=1}^n E|X_k - \mu_k|^3 = \lim_{n\to\infty}\left(\frac{n^{9/2}}{s_n^3} \cdot \frac{\sum_{k=1}^n E|X_k - \mu_k|^3}{n^4} \cdot \frac{1}{n^{1/2}}\right)$$
$$= 9^{3/2} \cdot \frac{1}{16} \cdot \lim_{n\to\infty} \frac{1}{n^{1/2}} = 0.$$

5.3 Teorema Central do Limite: Caso Multivariado

Concluímos dizendo que o Teorema Central do Limite também pode ser estendido ao caso de vetores aleatórios. Neste caso, tem-se que a distribuição da média amostral centrada converge para uma distriuição normal multivariada. A seguir, nós enunciamos formalmente o teorema sem prová-lo.

Teorema 5.3.1: *Seja $\vec{X}_1, \vec{X}_2, \ldots$ uma seqüência de vetores aleatórios k-dimensionais, independentes e identicamente distribuídos. Suponha que $\vec{X}_1$ tenha variância finita, e sejam $\vec{\mu}$ a média e Σ a matriz de covariância de $\vec{X}_1$. Seja $\overline{X}_n$ a média amostral, definida como a média aritmética dos vetores $\vec{X}_1, \ldots, \vec{X}_n$. Então,*

$$\sqrt{n}(\overline{X}_n - \vec{\mu}) \rightarrow^D N(\vec{0}, \Sigma), \text{ quando } n \rightarrow \infty.$$

5.4 Método Delta

O método Delta é um resultado que aumenta significativamente a relevância do Teorema Central do Limite. Antes de enunciarmos o teorema, vamos provar dois lemas. Dizemos que uma seqüência de variáveis aleatórias $\{Y_n\}$ é *limitada em probabilidade* se para todo $\epsilon > 0$, existir K e n_0 tal que $P(|Y_n| \leq K) > 1 - \epsilon$ para todo $n > n_0$.

Lema 5.4.1: *Se $\{Y_n\}$ converge em distribuição para uma variável aleatória com função de distribuição H, então a seqüência é limitada em probabilidade.*

Prova: Fixemos K_1 e K_2 pontos de continuidade de H tal que $H(K_1) > 1-\epsilon/4$ e $H(-K_2) < \epsilon/4$. Escolhamos n_0 tal que, $\forall n > n_0$,

$$H_n(K_1) > H(K_1) - \epsilon/4 > 1 - \epsilon/2$$

e

$$H_n(-K_2) < H(-K_2) + \epsilon/4 < \epsilon/2.$$

Então,

$$P(-K_2 \leq Y_n \leq K_1) \geq H_n(K_1) - H_n(K_2) > 1 - \epsilon.$$

O resultado está provado se escolhermos $K = \max(|K_1|, |K_2|)$. ∎

Lema 5.4.2: *Se $\{Y_n\}$ é limitada em probabilidade e $X_n = o(Y_n)$, então $X_n \to^P 0$.*

Prova: Dados quaisquer $\epsilon > 0$ e $\delta > 0$, precisamos mostrar que existe N tal que $P(|X_n| > \epsilon) < \delta$ para todo $n \geq N$. Como $\{Y_n\}$ é limitada em probabilidade, existe K e n_1 tal que $P(|Y_n| \leq K) > 1 - \delta$ para todo $n \geq n_1$. Como $X_n = o(Y_n)$, sabemos que existe n_2 tal que $\frac{|X_n|}{|Y_n|} < \frac{\epsilon}{K}$ para todo $n \geq n_2$. Façamos $N = \max(n_1, n_2)$, então para $n \geq N$, $|X_n| > \epsilon \Rightarrow |Y_n| > K$. Logo

$$P(|X_n| > \epsilon) \leq P(|Y_n| > K) < \delta.$$

∎

Teorema 5.4.3: *Se $\sqrt{n}(T_n - \theta) \to^D N(0, \tau^2)$, então*

$$\sqrt{n}[f(T_n) - f(\theta)] \to^D N(0, \tau^2[f'(\theta)]^2), \tag{5.2}$$

desde que $f'(\theta)$ exista e não seja zero.

Prova: Utilizaremos a versão da série de Taylor em torno de $T_n = \theta$ que diz que:

$$f(T_n) = f(\theta) + (T_n - \theta)f'(\theta) + o(T_n - \theta),$$

e então

$$\sqrt{n}[f(T_n) - f(\theta)] = \sqrt{n}(T_n - \theta)f'(\theta) + o(\sqrt{n}(T_n - \theta)).$$

O primeiro termo do lado direito converge em distribuição para $N(0, \tau^2[f'(\theta)]^2)$. Por outro lado, como $\sqrt{n}(T_n - \theta)$ converge em distribuição, pelo Lema 5.4.1, temos que $\sqrt{n}(T_n - \theta)$ é limitada em probabilidade. Então pelo Lema 5.4.2, $o(\sqrt{n}(T_n - \theta))$ converge para zero em probabilidade. O resultado portanto é uma conseqüência do Teorema de Slutsky. ∎

Este teorema pode parecer uma surpresa, já que se X é distribuído normalmente, a distribuição de $f(X)$, por exemplo, $1/X$, $\log X$, ou e^X não será tipicamente normal. A explicação para este paradoxo aparente pode ser encontrada na prova. Como $o(T_n - \theta) \rightarrow^P 0$, nós estamos quase certos que quando n for grande, T_n é aproximadamente linear, e uma função linear de uma variável normal é também normal. O processo de aproximar a diferença $f(T_n) - f(\theta)$ pela função linear $(T_n - \theta)f'(\theta)$ e o limite em (5.2) é chamado de *método delta*.

Exemplo 5.4.4: Para estimar p^2, suponha que temos a escolha entre

(a) n ensaios binomiais com probabilidade p^2 de sucesso; ou

(b) n ensaios binomiais com probabilidade p de sucesso.

Sejam X e Y o número de sucessos no primeiro e segundo tipo de ensaios, e suponha que como estimadores de p^2 nos dois casos, nós usaríamos X/n e $(Y/n)^2$, respectivamente. Então nós temos:

$$\sqrt{n}(\frac{X}{n} - p^2) \rightarrow^D N(0, p^2(1 - p^2))$$

e

$$\sqrt{n}((\frac{Y}{n})^2 - p^2) \rightarrow^D N(0, p(1 - p)4p^2).$$

Então, pelo menos para n grande, X/n será mais acurado que $(Y/n)^2$, desde que

$$p^2(1 - p^2) < p(1 - p)4p^2.$$

Dividindo ambos os lados por $p^2(1 - p)$, podemos ver que

$$\frac{X}{n} \text{ ou } \frac{Y^2}{n^2} \text{ é preferível se } p > 1/3 \text{ ou } p < 1/3, \text{ respectivamente.}$$

O método delta proporciona a base para derivar transformações que estabilizam a variância, ou seja, transformações que levem a uma variância assintótica que é independente do parâmetro. Suponha, por exemplo, que $X_1, \ldots, X_n$ são variáveis Poisson com parâmetro λ. Segue do Teorema Central do Limite que

$$\sqrt{n}(\overline{X} - \lambda) \rightarrow N(0, \lambda).$$

Para problemas de inferência que se referem a λ, é quase sempre inconveniente que λ ocorre não somente na esperança mas também na variância da distribuição limite. É portanto de interesse achar uma função f para a qual $\sqrt{n}[f(\overline{X}) - f(\lambda)]$ tende em distribuição para $N(0, c^2)$, onde c^2 não depende de λ. Em geral, suponha que $\sqrt{n}(T_n - \theta) \rightarrow^D N(0, \tau^2(\theta))$. Então, pelo método delta:

$$\sqrt{n}[f(T_n) - f(\theta)] \rightarrow^D N(0, \tau^2(\theta)(f')^2(\theta)),$$

desde que a derivada de f exista em θ e seja diferente de 0. A distribuição limite do lado direito terá portanto variância constante c^2 se $f'(\theta) = \frac{c}{\tau(\theta)}$. A transformação resultante é dita ser estabilizadora de variância.

Exemplo 5.4.5: Poisson. No caso de Poisson, temos $\theta = \lambda$ e $\tau(\theta) = \sqrt{\lambda}$. Logo,

$$f'(\lambda) = \frac{c}{\sqrt{\lambda}} \text{ ou } f(\lambda) = 2c\sqrt{\lambda}.$$

Fazendo $c = 1$, temos que

$$2\sqrt{n}(\sqrt{\overline{X}} - \sqrt{\lambda}) \to^D N(0,1).$$

Exemplo 5.4.6: Chi-Quadrado. Seja $Y_i = X_i^2$, onde as X_i's são i.i.d. $N(0,\sigma^2)$. Então, $EY_i = \sigma^2$ e $VarY_i = 2\sigma^4$ e pelo Teorema Central do Limite, temos

$$\sqrt{n}(\overline{Y} - \sigma^2) \to^D N(0, 2\sigma^4),$$

ou seja, $T_n = \overline{Y}$, $\theta = \sigma^2$, e $\tau^2(\theta) = 2\theta^2$. Logo,

$$f'(\theta) = \frac{c}{\sqrt{2}\theta} \text{ ou } f(\theta) = \frac{c}{\sqrt{2}} \log\theta.$$

Fazendo $c = 1$, vemos que

$$\sqrt{\frac{n}{2}} \log(\frac{\overline{Y}}{\sigma^2}) \to^D N(0,1).$$

Referências Bibliográficas

1. James, B. (1981), “Probabilidade: um curso em nível intermediário”, 2a. edição - Projeto Euclides

2. Magalhães, Marcos M. (2006), "Probabilidade e Variáveis Aleatórias", 2a. edição, edusp.

3. Leite, José G. e Singer, Julio da Mota (1990), "Métodos Assintóticos em Estatística: Fundamentos e Aplicações", 9o. Simpósio Nacional de Probabilidade e Estatística, IME-USP.

4. Lima, E. (1976), “Curso de Análise”, vol.1 - Projeto Euclides

5. Guidorizzi, H. L. (1994), "Um curso de Cálculo", Volume 4, Livros Técnicos e Científicos Editora.

6. Ávila, Geraldo (1994), "Cálculo - Funções de Várias Variáveis", Volume 2, Livros Técnicos e Científicos Editora.

7. Meyer, P. (1983), "Probabilidade - Aplicações à Estatística", 2a. edição, Livros Técnicos e Científicos Editora, Rio de Janeiro.

8. Lehman, E. (1999), "Elements of Large-Sample Theory", Springer-Verlag.