Universidade Estadual de Campinas
Departamento de Estatística — IMECC
Disciplina: MI401 – Probabilidade
Professor: Dr. Diego Fernando de Bernardini (vulgo "Come Mal")
Período: 1° semestre de 2019
Acadêmico: André Felipe B. Menezes

Notas de Probabilidade

Este documento contém um resumo das notas de aulas da disciplina MI401 - Probabilidade do mestrado em estatística na Unicamp.

O professor literalmente transcreveu na lousa o livro "Probabilidade: Um curso em nível intermediário" do Barry Ree James.

Um fato tanto curioso e lamentável: Barry James veio a falecer quando estávamos no meio da disciplina. E então uma dúvida surgiu: Será que ele resolveu todos os exercícios de seu livro?

Campinas
Agosto de 2019

Sumário

1 Definições Básicas

Nesta primeira parte do curso foi apresentado definições sobre eventos, eventos aleatórios, álgebra, σ-álgebra, axiomas de Kolmogorov, propriedades de probabilidade, modelo probabilístico, espaço de probabilidade, probabilidade condicional, teorema da probabilidade composta, teorema da probabilidade total e independência.

Nos dois primeiros dias definimos dois experimentos aleatórios que foram utilizados como exemplos para definição intuitiva de espaço amostral, evento e cálculo de probabilidades. Uma revisão da notação de conjuntos para a linguagem de eventos foi apresentada. Assim, vale lembrar que:

- $A \cup B$
- $A \cap B$
- A^c
- $A \subset B$: ocorrência de A implica na ocorrência de B;
- $A \cap B = \varnothing$: A e B são mutuamente exclusivos;
- $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$

A seguinte definição foi apresentada: *um evento A qualquer a qual conseguimos atribuir probabilidade é chamado de **evento aleatório**.*

1.1 σ-álgebra

Uma classe $\mathbb{A}$ de subconjuntos de um conjunto não vazio Ω satisfazendo as seguintes propriedades

A1 $\Omega \in \mathbb{A}$

A2 Se $A_n \in \mathbb{A}$ então $A_n^c \in \mathbb{A}$

A3 Se $A_n \in \mathbb{A}$ para n = 1,2,3, ..., então $\bigcup_{i=1}^{\infty} A_i \in \mathbb{A}$

é denominada de σ-álgebra de subconjuntos de Ω.

Considerando as propriedades A2 e A3 satisfeita e utilizando a leis de De-Morgan é possível mostrar que $\bigcap_{i=1}^{\infty} A_i \in \mathbb{A}$.

Na sequência, dois exemplos de experimentos foram apresentados e definições de σ-álgebra foram discutidas.

Além disso, importante destacar que denotaremos $\mathcal{B}$ como sendo a σ-álgebra de Borel na reta, isto é, $\mathcal{B}$ é a menor σ-álgebra contendo todos os intervalos.

1.2 Construção Axiomática de Probabilidade (Kolmogorov)

Considerando a σ-álgebra $\mathcal{A}$ de eventos aleatórios. Vamos supor que a todo $A \in \mathcal{A}$ seja associado um número real $P(A)$, chamado de probabilidade de A, de modo que os seguintes axiomas sejam satisfeitos:

1. $P(A) > 0$;
2. $P(\Omega) = 1$;
3. Se $A_1, A_2, \ldots \in \mathcal{A}$ são disjuntos então

$$P\left(\bigcup_{i=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} P(A_n) \tag{1}$$

O Axioma 3 é conhecido como σ-aditividade. Quando temos uma união finita, então denominados o axioma 3 de *Aditividade Finita*.

Portanto uma função P definida na σ-álgebra $\mathbb{A}$ e satisfazendo os Axiomas 1,2 e 3 chama-se uma medida de probabilidade em $\mathcal{A}$.

Outro Axioma importante que decorre dos axiomas 1, 2 e 3 (para o caso finito) é o seguinte

4. Se a sequência $(A_n)_{n\geq 1}$, onde $A_n \in \mathbb{A}$ decresce para o vazio, então $P(A_n) \to 0$ quando $n \to \infty$.

Portanto para provar que P é uma probabilidade em um σ-álgebra $\mathbb{A}$ basta verificar os axiomas 1,2 e 3 ou 1, 2, 3 (caso finito) e 4.

Algumas propriedades decorrentes dos axiomas de Kolmogorov são:

P1. $P(A^c) = 1 - P(A)$

P2. $0 \leq P(A) \leq 1$

P3. $A_1 \subset A_2 \to P(A_1) \leq P(A_2)$

P4. $P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$

P5. $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$

P6. Continuidade de probabilidade:

Se $A_n \uparrow A$ então $P(A_n) \uparrow P(A)$

Se $A_n \downarrow A$ então $P(A_n) \downarrow P(A)$.

Nesse ponto somos capazes de definir um **modelo probabilístico** e um **espaço de probabilidade**.

O **modelo probabilístico** é constituído dos seguintes elementos:

- O espaço amostral Ω, isto é, o conjunto não vazio de resultados possíveis do experimento;
- Uma σ-álgebra $\mathbb{A}$ de eventos aleatórios;
- Uma probabilidade definida na σ-álgebra $\mathbb{A}$

De forma redundante, mas com outras palavras, o **espaço de probabilidade** é um trio $(\Omega, \mathbb{A}, P)$ em que

- Ω é um conjunto não vazio;
- $\mathbb{A}$ é uma σ-álgebra de subconjuntos de Ω;
- P é uma probabilidade em $\mathbb{A}$.

1.2.1 Probabilidade Condicional

Seja $(\Omega, \mathbb{A}, P)$ um espaço de probabilidade. Se $B \in \mathbb{A}$ e $P(B) > 0$, então a probabilidade condicional de A dado B é definida por

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad A \in \mathbb{A}. \tag{2}$$

A partir de (2) temos o Teorema da Multiplicação ou Teorema da Probabilidade Composta diz que

- $P(A \cap B) = P(A)\, P(B \mid A) = P(B) P(A \mid B), \qquad \forall A, B \in \mathbb{A}$
- $P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \ldots A_{n-1})$, $\forall A_1, A_2, \ldots, A_n \in \mathbb{A}\ \forall\ n = 2, 3, \ldots$.

O Teorema da Probabilidade Total diz que se uma sequencia (finita ou enumerável) de eventos aleatórios $A_1, A_2, \ldots$ formam um partição do espaço amostral Ω então

$$P(B) = \sum_i P(A_i)\, P(B \mid A_i), \forall B \in \mathbb{B}. \tag{3}$$

A partir dos dois teoremas enunciados chegamos a fórmula de Bayes, isto é, a probabilidade condicional de A_i dado B.

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\, P(B \mid A_i)}{\sum\limits_j P(A_j)\, P(B \mid A_j)} \tag{4}$$

1.2.2 Independência

Dois eventos A e B são estocastimente independentes se

$$P(A \cap B) = P(A)\,P(B) \tag{5}$$

Deste fato decorre que os eventos $A \cap B^c$, $A^c \cap B$ e $A^c \cap B^c$ também são independentes.

Generalizando 5 considere os eventos aleatórios $A_i, i \in I$, sendo I um conjunto de índices, eles serão independentes 2 a 2 (a pares) se

$$P(A_i \cap A_j) = P(A_i)\,P(A_j), \quad \forall i, j, \in I,\ i \neq j \tag{6}$$

A independência a pares não implica a independência coletiva. Dessa foram temos uma terceira definição.

Os eventos $A_1, A_2, \ldots, A_n (n \geq 2)$ são coletivamente ou estocastimente independentes se

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1})\,P(A_{i_2})\,\ldots P(A_{i_m}) \tag{7}$$

$\forall 1 \leq i_1 \leq i_2 \leq \ldots \leq i_m \leq n$, $\forall m = 2, 3, \ldots, n$.

Ou seja, se todas as combinações de eventos satisfazem a regra produto.

2 Variáveis Aleatórias e Funções de Distribuições

Uma variável aleatória X em um espaço de probabilidade $(\Omega,\ \mathbb{A},\ P)$ é uma função real definida no espaço Ω tal que o evento $[X \leq x] := \{\omega \in \Omega : X(\omega) \leq x\}$ é um evento aleatório para todo $x \in \mathbb{R}$, ou seja, $X : \Omega \to \mathbb{R}$ é a variável aleatória se $[X \leq x] \in \mathbb{A}, \forall x \in \mathbb{R}$. Na linguagem da teoria da medida dizemos que X é uma função mensuravel a $\mathbb{A}$.

Dessa definição vemos que a variável aleatória mapeia o espaço amostral Ω nos reais, isto é, fornece valores reais aos eventos do espaço amostral. Em outras palavras a variável aleatória associa um numero real $X(\omega)$ ao elemento $\omega \in \Omega$.

A função distribuição da variável aleatória X, denotada por F_X ou F é definida por

$$F_X(x) = P[X \leq x], \qquad x \in \mathbb{R} \tag{8}$$

As seguintes propriedades definem uma função de distribuição

1. $x < y \to F(x) \leq F(y)$, isto é, F é não decrescente.

2. Se $x_n \downarrow x$ então $F(x_n) \downarrow F(x)$, isto é, F é contínua a direita.

3. Se $x_n \downarrow -\infty$ então $F(x) \downarrow 0$.

   Se $x_n \uparrow +\infty$ então $F(x) \uparrow 1$.

   Logo, $F(-\infty) = 0$ e $F(+\infty) = 1$.

2.1 Tipos de Variáveis Aleatórias

Existem 4 tipos de variáveis aleatórias. As variáveis aleatórias discretas, absolutamente contínuas, mistas e singulares. Entretanto, as variáveis aleatórias singulares não possuem aplicações práticas no mundo, isto é, nenhum fenômeno pode ser descrito por este tipo de variável aleatória (pelo menos conforme meu pouco conhecimento).

1. Uma variável aleatória X é discreta se toma um número finito ou enumerável de valores, i. e., se existe um conjunto finito ou enumerável $\{x_1, x_2, \ldots\} \subset \mathbb{R}$ tal que $X(\omega) \in \{x_1, x_2, \ldots\} \forall \omega \in \Omega$.

   A função $p(x_i) = P(X = x_i)$ é denominada de função de probabilidade de X.

   Neste caso, o evento $[X \leq x] = \bigcup\limits_{i:x_i \leq x} [X \leq x_i]$ assim,

$$F_X(x) = \sum_{i:x_i \leq x} P(X = x_i) = \sum_{i:x_i \leq x} p(x_i) \tag{9}$$

2. Uma variável aleatória X é absolutamente contínua se existe uma função $f(x) \geq 0$ tal que

$$F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad \forall x \in \mathbb{R}. \tag{10}$$

A função $f(x)$ é denominada de função densidade. Neste caso $F_X(x)$ é absolutamente contínua e $f(x) = F'(x)$.

3. De grosso modo, uma variável aleatória do tipo mista se sua função de distribuição for uma mistura de variáveis discreta e contínua.

2.2 Distribuição de uma Variável Aleatória

Se X é variável aleatória definida no espaço de probabilidade $(\Omega, \mathbb{A}, P)$ então o evento

$$[X \in B] := \{\omega \in \Omega : X(\omega) \in B\} \tag{11}$$

é evento aleatório para todo Boreliano, isto é, $[X \in B] \in \mathbb{A}$, $\forall B \in \mathcal{B}$, em que $\mathcal{B}$ é a σ-algebra de Borel na reta.

A probabilidade P_X, definida na σ-álgebra de Borel por $P_X(B) = P(X \in B)$ é chamada distribuição de X.

1. Caso discreto:

$$P_X(B) = \sum_{i:x_i \in B} P(X = x_i), \quad \forall B \in \mathcal{B} \tag{12}$$

2. Caso contínuo:

$$P_X(B) = \int_B f(x)\,\mathrm{d}x, \quad \forall B \in \mathcal{B} \tag{13}$$

Representações da distribuição da variável aleatória X são dadas por:

- Função de distribuição F_X;
- Função de densidade $f(x)$ para caso absolutamente contínuo;
- Função de probabilidade $p(x_i)$ para o caso discreto;
- Função característica.

Se X tem densidade $f_X(x)$ então $Y = b\,X + c$, com $b > 0$ e $c \in \mathbb{R}$ tem densidade dada por

$$f(y) = \frac{1}{b} f_X\left(\frac{y-c}{b}\right), \quad y \in \mathbb{R}. \tag{14}$$

em que c e b são chamados parâmetros de locação e escala, respectivamente.

Discutimos as seguintes distribuições: Normal, Gama, Binomial, Poisson e Geométrica.

2.3 Vetores Aleatórios

Um vetor $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ em que os elementos são variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathbb{A}, P)$ é denominado vetor aleatório (ou variável aleatória n-dimensional).

A função de distribuição $F = F_{X_1,\ldots,X_n}$ de um vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)$ é definida por

$$F(\boldsymbol{x}) = F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \quad \forall\, (x_1, \ldots, x_n) \in \mathbb{R}^n \tag{15}$$

O vetor aleatório $\mathbf{X}$ é uma função definida no espaço amostral Ω assumindo valores em $\mathbb{R}^n$, isto é, $\mathbf{X} : \Omega \to \mathbb{R}^n$.

As seguintes propriedades definem uma função de distribuição conjunta das variáveis aleatórias $X_1, X_2, \ldots, X_n$:

1. Se $x < y \to F(x, x_2, \ldots, x_n) \leq F(y, x_2, \ldots, x_n)$, isto é, F é não decrescente em cada uma das variáveis.

2. Se $y_m \downarrow x_1$ quando $m \to \infty$, então $F(y_m, x_2, \ldots, x_n) \downarrow F(x_1, x_2, \ldots, x_n)$, isto é, F é contínua a direita em cada uma das variáveis.

3. Para todo i,
$$\lim_{x_i \to \infty} F(x_1, \ldots, x_n) = 0$$
ou seja, F é igual a zero quando uma das coordenadas vai para o infinito. Pois, o evento $[X_i \leq x_i] \downarrow \varnothing$.
$$\lim_{\forall i, x_i \to \infty} F(x_1, \ldots, x_n) = 1$$
ou seja, F é igual a um quando todas coordenadas tendem ao o infinito. Pois, o evento $[X_i \leq x_i] \uparrow \Omega$.

4. Para $I = (a, b]$ e $g : \mathbb{R}^k \to \mathbb{R}$ define-se o operador de diferença por:
$$\Delta_{k,I}\, g(x_i, \ldots, x_k) = g(x_1, \ldots, x_{k-1}, b) - g(x_1, \ldots, x_{k-1}, a).$$
Dessa forma, a quarta propriedade que caracteriza uma função de distribuição conjunta é:
$$\Delta_{1,I_1} \ldots \Delta_{n,I_n}\, F(x_1, \ldots, x_n) \geq 0, \quad \forall I_k = (a_k, b_k], \quad k = 1, \ldots, n.$$

Analogamente ao caso univariado dizemos que o vetor aleatório $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ é

1. **Discreto** se permite apenas um número finito ou enumerável de valores.

2. **Contínuo** se existe uma função $f(x_1, \ldots, x_n) \geq 0$ tal que

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\,\mathrm{d}t_1 \ldots \mathrm{d}t_n \quad \forall\,(x_1, \ldots, x_n) \in \mathbb{R}^n \tag{16}$$

   a função f é denominada densidade conjunta do vetor aleatório $\mathbf{X}$.

2.4 Distribuição de um Vetor Aleatório

Se $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ é um vetor aleatório n-dimensional definido no espaço de probabilidade $(\Omega, \mathbb{A}, P)$ então o evento

$$[X \in B] := \{\omega \in \Omega : X(\omega) \in B\} \tag{17}$$

é evento aleatório para todo Boreliano, isto é, $[X \in B] \in \mathbb{A}$, $\forall B \in \mathcal{B}^n$, em que $\mathcal{B}$ é a σ-algebra de Borel no $\mathbb{R}^n$.

A probabilidade P_X, definida em Borel $\mathcal{B}^n$ por $P_X(B) = P(X \in B)$ é denominada distribuição do vetor $\mathbf{X}$.

1. Caso discreto:

$$P_{\mathbf{X}}(B) = \sum_{i:\boldsymbol{x}_i \in B} P(\mathbf{X} = \boldsymbol{x}_i), \quad \forall B \in \mathcal{B}^n \tag{18}$$

2. Caso contínuo:

$$P_{\mathbf{X}}(B) = \int_B \cdots \int_B f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \ldots x_n, \quad \forall B \in \mathcal{B}^n \tag{19}$$

2.5 Independência

As variáveis aleatórias $X_1, \ldots, X_n$ do vetor aleatório $\mathbf{X}$ são independentes, se e somente se os eventos formados por qualquer grupo de variáveis distintas forem independentes.

Portanto, as variáveis $X_1, \ldots, X_n$ são coletivamente independentes se

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i) \tag{20}$$

Dessa definição decore que para todo família de variáveis independentes qualquer subfamília é também formada por variáveis independentes.

Podemos ir além observando que as variáveis $X_1, \ldots, X_n$ são independentes se sua função de distribuição conjunta fatora e é o produto das função de distribuição individuais.

Assim, podemos postular dois critérios para identificar a independência de variáveis aleatórias:

1. Se $X_1, \ldots, X_n$ são independentes então

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

2. Reciprocamente, se existem função $F_1, \ldots F_n$ tais que

$$\lim_{x \to \infty} F_i(x) = 1 \quad \forall i \tag{21}$$

e

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

então $X_1, \ldots, X_n$ são independentes e $F_i = F_{X_i}$, $\forall i = 1, \ldots, n$.

Particularmente, no caso de um vetor aleatório contínuo podemos definir mais dois critérios:

1. Se $X_1, \ldots, X_n$ são independentes com densidades $f_{X_1}, \ldots, f_{X_n}$ então a função

$$f_{X_1,\ldots,X_n} = \prod_{i=1}^{n} f_{X_i}(x_i), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

é a densidade conjunta do vetor aleatório $\mathbf{X}$.

2. Reciprocamente, se $X_1, \ldots, X_n$ densidade conjunta satisfazendo

$$f_{X_1,\ldots,X_n} = \prod_{i=1}^{n} f_i(x_i), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n$$

em que $f_i(x) \geq 0$ e $\int_{-\infty}^{\infty} f_i(x) = 1$, $\forall i$ então $X_1, \ldots, X_n$ são independentes e f_i é a densidade de X_i, $\forall i = 1, \ldots, n$.

2.6 Distribuição Marginal

Seja o vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)$ então a distribuição marginal de X_1, por exemplo, é caracterizada por:

1. Função de distribuição, dada por:

$$F_{X_1}(x_1) = \lim_{x_i \to \infty} F(x_1, x_2, \ldots, x_n) = F(x_1, \infty, \ldots, \infty), \quad \forall i \neq 1.$$

2. Função densidade de probabilidade, dada por:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_2 \ldots \mathrm{d}x_n \tag{22}$$

2.7 Distribuições de Funções de Variáveis e Vetores Aleatórios

Seja $\mathbf{X} = (X_1, \ldots, X_n)$ um vetor aleatório em $(\Omega, \mathbb{A}, P)$ e $Y = g(X_1, \ldots, X_n)$. Para que Y seja uma variável aleatória, a função g deve ser mensurável à Borel, isto é,

$$g^{-1}(B) = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : g(x_1, \ldots, x_n) \in B\} \in \mathcal{B}^n, \quad B \in \mathcal{B}.$$

Toda função que visualizamos é mensurável a Borel.

Para $y \in \mathbb{R}$, seja $B_y = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : g(x_1, \ldots, x_n) \leq y\}$. Então, $g(X_1, \ldots, X_n) \leq y \iff (X_1, \ldots, X_n) \in B_y$. Portanto, a função de distribuição de $Y = g(X_1, \ldots, X_n)$ é dada por

$$F_Y(y) = P(Y \leq y) = P(g(X_1, \ldots, X_n) \leq y) = P((X_1, \ldots, X_n) \in B_y) = P_{X_1, \ldots, X_n}(B_y)$$

Conhecendo a distribuição conjunta de $X_1, \ldots, X_n$, podemos encontrar a distribuição de qualquer função mensurável das X_i's.

Transformações usuais entre variáveis aleatórias são, por exemplo, soma e produto. No primeiro, caso se X e Y tem densidade conjunta dada por $f_{X,Y}(x, y)$ e $Z = X + Y$, então a densidade de Z é dada por

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - t, t)\, \mathrm{d}t = \int_{-\infty}^{\infty} f(t, z - t)\, \mathrm{d}t$$

Por outro lado, se X e Y são independentes, com densidades f_x e f_X, então a densidade de Z, denominada de *convolução*, é definida por

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - t)\, f_Y(t)\, \mathrm{d}t = \int_{-\infty}^{\infty} f_X(t)\, f_Y(z - t)\, \mathrm{d}t$$

Um resultado importante da transformação de variáveis aleatórias refere-se a independência de função de variáveis aleatórias. Nestes casos, se $X_1, \dots, X_n$ são independentes então funções de famílias disjuntas das X_i's também são independentes.

2.8 Método do Jacobiano

O método do jacobiano é uma ferramenta do cálculo para obter distribuições de funções de variáveis aleatórias. Neste método suponhamos que $G_0 \subset \mathbb{R}$ e $G \subset \mathbb{R}$ são regiões abertas e que a função $g : G_0 - \to G$ é bijetora entre G_0 e G. A região G_0 é o suporte da variável ou vetor aleatório não transformado, enquanto que a região G o conjunto de chegada, ou seja, o domínio da variável ou vetor transformado. Assim, temos o seguinte

$$g(x_1, \dots, x_n) = (g_1(x_1, \dots, x_n), \dots, g_n(x_1, \dots, x_n)) = (y_1, \dots, y_n)$$

Como g é bijetora, então existe a função inversa $h = g^{-1}$ em G, tal que

$$x_1 = h_1(y_1, \dots, y_n), \dots, x_n = h_n(y_1, \dots, y_n).$$

Vamos supor que as derivadas parciais

$$\frac{\partial}{\partial y_i} h_1(y_1, \dots, y_n), \quad 1 \leq i, j \leq n$$

existam e sejam contínuas em G. Então, o jacobiano $J(x, y)$ é definido por

$$J(x, y) = \det \begin{bmatrix} \dfrac{\partial}{\partial y_1} h_1(y_1, \dots, y_n) & \dots & \dfrac{\partial}{\partial y_n} h_1(y_1, \dots, y_n) \\ \vdots & & \vdots \\ \dfrac{\partial}{\partial y_1} h_n(y_1, \dots, y_n) & \dots & \dfrac{\partial}{\partial y_n} h_n(y_1, \dots, y_n) \end{bmatrix}$$

Dessa forma, seja f a densidade conjunta das variáveis aleatórias $X_1, \dots, X_n$ definidas em G_0. Sejam $Y_1, \dots, Y_n$ as variáveis transformadas de tal forma que $Y_i = g_i(X_1, \dots, X_n), i = 1, \dots, n$. Então, a densidade conjunta de $Y_1, \dots, Y_n$ é dada por

$$f_{\mathbf{Y}}(y_1, \dots, y_n) = \begin{cases} f(h_1(y_1, \dots, y_n), \dots, h_n(y_1, \dots, y_n)) \, |J(x,y)|, & \boldsymbol{y} \in G \\ 0, & \boldsymbol{y} \notin G. \end{cases}$$

3 Esperança Matemática

Nesta seção discutimos inicialmente uma definição para a integral de Stieltjes que será utilizada posteriormente para uma definição formal e geral da esperança de uma variável aleatória. Na sequência algumas propriedades da esperança são apresentadas, tais como a linearidade e a desigualdade de Jensen. Apresenta-se também a esperança de funções de variáveis aleatórias, sendo que uma atenção especial é dada aos momentos. A esperança de funções de vetores aleatórios é definida, permitindo assim detalhamento da da esperança do produto de variáveis aleatórias e da covariância. Por fim, concluímos o capítulo discutindo dois teoremas de convergência.

3.1 Integral de Stieltjes

Aqui apresento a definição da integral de Stieltjes e algumas de suas propriedades. Interessante observar que começaremos definindo a integral de Riemann-Stieltjes, no entanto o autor apresenta uma caso "trivial" onde a integral de Riemann-Stieltjes não existe, assim define-se a integral de Lebesgue-Stieltjes ou simplesmente Integral de Stieltjes que por ironia ou talvez excesso de generalidades torna-se a integral de Riemann-Stieltjes.

3.1.1 Integral de Riemann-Stieltjes

Seja φ uma função contínua definida em $[a, b]$ e F uma função de distribuição. Sejam $a = x_1 < x_2 < \ldots < x_{n+1} = b$, os pontos x_i formam uma partição e sua soma é $\max_{1 \le i \le n} (x_{i+1} - x_i)$. A integral de Riemann-Stieltjes de φ em $[a, b]$, com relação a F (ou ponderada por F) é o limite de "somas de Riemann" da forma

$$\sum_{i=1}^{n} \varphi(y_i)\ [F(x_{i+1}) - F(x_i)]\,, \tag{23}$$

quando a norma da partição tende a zero, sendo que y_i é um ponto arbitrário de $[x_i, x_{i+1}]$.

Sob as condições descritas tal limite existe, é finito e representamos por

$$\int_a^b \varphi(x)\, dF(x) \tag{24}$$

sendo que φ é o integrando e F é o integrador.

A integral de Riemann-Stieltjes sobre a reta é dada por

$$\int_{-\infty}^{+\infty} \varphi(x)\, dF(x) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b \varphi(x)\, dF(x) \tag{25}$$

se o limite existe.

3.1.2 Integral de Lebesgue-Stieltjes

Com intuito de generalizar o generalizável definiremos a integral de Lebesgue-Stieltjes ou simplesmente a integral de Stieltjes.

Seja φ uma função real mensurável no sentido discutido anteriormente. E seja F uma função de variação limitada contínua à direita, isto é, a diferença duas funções monótonas crescentes limitadas e contínua à direita. A seguir apresentamos a definição e algumas propriedades dessa linda integral.

1. Denotamos a integral de Stieltjes por

$$\begin{aligned} \int_a^b \varphi(x)\, dF(x) &= \int_a^b \varphi\, dF \\ \int \varphi\, dF &= \int_{-\infty}^{\infty} \varphi\, dF = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b \varphi\, dF \end{aligned}$$

2. Quando o integrando φ é contínuo em $[a, b]$ a integral de Stieltjes torna-se a integral de Riemann-Stieltjes. Essa propriedade permite trabalharmos efetivamente com a integral de Riemann-Stieltjes.

3. A diferença de F sobre um intervalo é a integral da sua diferencial. Ou seja,

$$\int_a^b dF(x) = F(b) - F(a)$$

4. A integral de Stieltjes é linear no integrando e no integrador.

   Para $\varphi(x) = \alpha f(x) + \beta g(x)$ temos

$$\int \varphi dF = \alpha \int f\, dF + \beta \int g\, dF$$

   Para $H(x) = \alpha F(x) + \beta G(x)$ temos

$$\int \varphi\, dH = \alpha \int \varphi\, dF + \beta \int \varphi\, dG.$$

   Desde que as integrais estejam bem definidas e as somas tenham sentido.

5. A integral de Stieltjes é aditiva, isto é,

$$\int \varphi \, dF = \int_{-\infty}^{a} \varphi \, dF + \int_{a}^{\infty} \varphi \, dF.$$

6. Quando F é a função de distribuição de uma variável aleatória X discreta, a integral de Stieltjes se reduz a uma série.

$$\int \varphi \, dF = \sum_{i} \varphi(x_i) \, p(x_i)$$

em que $p(x_i) = P(X = x_i) > 0$ e $\sum_i p(x_i) = 1$, ou seja, $p(x_i)$ é o salto de F em x_i.

Para regiões de integração com intervalo finito temos

$$\int_{a}^{b} \varphi \, dF = \sum_{i: a \le x_i \le b} \varphi(x_i) \, p(x_i)$$

7. Quando F é a função de distribuição de uma variável aleatória contínua com densidade f então $dF(x) = f(x)dx$. Logo,

$$\begin{aligned} \int_{a}^{b} \varphi \, dF &= \int_{a}^{b} \varphi(x) \, f(x) dx \\ \int \varphi \, dF &= \int_{-\infty}^{\infty} \varphi(x) \, f(x) dx \end{aligned}$$

8. A integral de Stieltjes de φ em um intervalo é definida como a integral sobre a reta do produto de φ com o indicador do intervalo.

Por exemplo, a integral sobre um Boreliano B qualquer é dada por

$$\int_{B} \varphi \, dF = \int \varphi \, I_B \, dF.$$

No caso discreto temos

$$\int_{B} \varphi \, dF = \sum_{i: x_i \in B} \varphi(x_i) \, dF(x_i)$$

3.2 Esperança

Seja X uma variável aleatória qualquer e F sua função de distribuição, então a esperança de X é definida por

$$E(X) = \int_{-\infty}^{\infty} x\, dF(x)$$

quando a integral imprópria de Riemann-Stieltjes esta bem definida.

Dizemos que X é **integrável** se $E(X)$ é finita. Considere

$$\int_{-\infty}^{\infty} x\, dF(x) = \int_{-\infty}^{0} x\, dF(x) + \int_{0}^{\infty} x\, dF(x) = I + II$$

sendo que $I \leq 0$ e $II \geq 0$ e a esperança de X estará bem definida se I ou II for finita. Além disso temos as seguintes situações:

1. Se I e II são finitas, então X é integrável e $E(X)$ existe;
2. Se I é finita e $II = +\infty$, então X é integravel e $E(X) = +\infty$;
3. Se $I = -\infty$ e II é finita, então X é integravel e $E(X) = -\infty$;
4. Se $I = -\infty$ e $II = +\infty$, então X não é integravel e $E(X)$ não existe.

Assim, X é integrável se, e somente se, $\int_{-\infty}^{\infty} |x|\, dF(x) < \infty$.

- Se X é uma variável contínua com densidade $f(x)$, então a esperança de X é

$$E(X) = \int x\, dF(x) = \int x\, f(x) dx.$$

- Se X é uma variável discreta com função de probabilidade $p(x_i)$, então a esperança de X é

$$E(X) = \int x\, dF(x) = \sum_i x_i\, p(x_i).$$

Usando propriedades da integral, mais especificamente integração por partes com $d(x\, F(x)) = xdF(x) + F(x)dx$, podemos mostrar que a esperança de uma variável aleatória pode ser expressa por

$$E(X) = \int_{0}^{\infty} (1 - F(x))dx - \int_{-\infty}^{0} F(x)dx$$

Segue deste resultado que se a variável aleatória X toma somente valores não-negativos, isto é, $X(\omega) \geq 0, \forall \omega \in \Omega$, então $F(x) = 0$ para $x < 0$ implicando que

$$E(X) = \int_0^\infty (1 - F(x))dx.$$

Se X assumir somente valores inteiros não negativos, então

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n)$$

Além disso, como $|X| \geq 0$, decorre deste último resultado que

$$E(|X|) = \int_0^\infty P(|X| > x)dx = \int_0^\infty (1 - F(x))dx + \int_{-\infty}^0 F(x)dx.$$

3.3 Esperança em $(\Omega, \mathbb{A}, P)$

Uma segunda abordagem para definir a esperança utiliza-se conceitos da teoria da medida, portanto é uma definição teórica e bastante abstrata. Os passos para definir a esperança de variáveis aleatórias no caso geral são:

1. Definir a esperança para variáveis aleatórias simples;
2. Definir a esperança de variáveis aleatórias não-negativas como sendo o limite da esperança de variáveis aleatórias simples;
3. Definir a esperança no caso geral reescrevendo uma variável aleatória como a diferença entre a parte positiva e negativa;

Vamos formalizar esses passos. Recomendo ler o livro do Magalhães (Probabilidade e Variáveis Aleatórias) para um tratamento mais detalhado.

Seja $(\Omega, \mathbb{A}, P)$ um espaço de probabilidade. Dizemos que uma variável aleatória $X : \Omega \to \mathbb{R}$ é simples se assume somente um número finito de valores distintos. Neste caso, X pode ser representada por:

$$X = \sum_{i=1}^{n} x_i \, I_{A_i}(\omega), \quad x_i \in \mathbb{R} \; \forall i, n \in \mathbb{N}$$

em que $A_1, \ldots, A_n$ formam uma partição de Ω.

Definimos o valor esperado de X ou a integral de X por

$$E(X) = \int_\Omega X \, dP = \sum_{i}^{n} x_i \, P(A_i),$$

em que a medida dP é obtida a partir da função de distribuição de X.

Importante salientar que diferentes representações da variável aleatória simples, isto é diferentes x_i's levam ao mesmo valor esperado. Em outras palavras, se $\{A_i, 1 \leq i \leq n\}$ e $\{B_j, 1 \leq j \leq m\}$ são partições de Ω então

$$E(X) = \sum_{i=1}^{n} x_i\, P(A_i) = \sum_{j=1}^{m} y_j\, P(B_j).$$

Qualquer variável aleatória não negativa $X : [0, \infty) \to \Omega$ é o limite de alguma sequência monótona crescente $\{X_n\}$ de variáveis aleatórias simples, ou seja, $X_n(\omega) \uparrow X(\omega)\ \forall \omega \in \Omega$. Em particular seja

$$X_n = \min\left\{n, 2^{-n} \lfloor 2^n\, X \rfloor\right\}$$

em que $\lfloor x \rfloor$ é o maior inteiro menor ou igual a x. Vemos que X_n é uma variável aleatoria finita e simples não-negativa. Além disso, pode-se verificar que $X_n \uparrow X$ quando $n \to \infty$.

Neste caso, define-se a integral de X ou esperança de X por

$$E(X) = \lim_{n \to \infty} E(X_n).$$

Sem perdas de generalidades uma variável aleatória $X : \Omega \to \mathbb{R}$ pode ser representada por

$$X = X^{+} - X^{-}$$

sendo que

$$X^{+}(\omega) = \max\{X(\omega), 0\} \quad \text{e} \quad X^{-}(\omega) = -\min\{X(\omega), 0\}$$

são variáveis aleatórias não negativas.

Se $E(X^{+}) < \infty$ ou $E(X^{-}) < \infty$ então definis-se a integral de X ou valor esperado de X por

$$E(X) = E(X^{+}) - E(X^{-}).$$

A notação comum para representar a esperança de X é

$$E(X) = \int_{\Omega} X(\omega)\, dP \quad \text{ou} \quad E(X) = \int_{\Omega} X(\omega)\, dP(\omega)$$

3.4 Propriedades da Esperança

Listo aqui algumas propriedades interessantes da esperança.

1. Se $X = c$, ou seja, $X(\omega) = c, \forall \omega \in \Omega$. Então $E(X) = c$.

2. Se $X \leq Y$ então $E(X) \leq E(Y)$.

3. Propriedade da linearidade diz que:

$$E(a\,X + b) = a\,E(x) + b, \quad \forall a, b \in \mathbb{R}.$$

e

$$E(aX + bY) = aE(X) + bE(Y), \quad \forall a, b \in \mathbb{R}.$$

4. Desigualdade de Jensen nos diz que: se φ é uma função convexa definida na reta, então

$$E\left[\varphi(X)\right] \geq \varphi\left[E(X)\right].$$

No caso em que φ é côncava temos

$$E\left[\varphi(X)\right] \leq \varphi\left[E(X)\right].$$

Importante ressaltar que para a validade da desigualdade de Jensen, basta que a função φ seja convexa em um intervalo (a, b) tal que $P(a < X < b) = 1$.

Por exemplo, se X é uma variável positiva, então $P(X > 0) = 1$. Seja $\varphi(x) = \frac{1}{x}$ com $(a, b) = (0, \infty)$. Pela desigualdade de Jensen concluímos que:

$$E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)}.$$

Sob a mesma condição pode-se aplicar a desigualdade de Jensen a função côncava $\varphi(x) = \log x$, obtendo:

$$E\left(\log X\right) \leq \log E(X).$$

3.5 Esperança de Funções de Variáveis Aleatórias

Seja X uma variável aleatória, $\varphi(x)$ uma função real mensurável e $Y = \varphi(X)$. Então como vimos na seção anterior, Y é uma variável aleatória cuja esperança é definida por

$$E(Y) = E(\varphi(X)) := \int y\,dF_{\varphi(X)}(y) = \int \varphi(x)\,dF_X(x)$$

Se $\varphi(x) = x^k$ com $k = 1, 2, \ldots$, então

$$E(X^k) = \int x^k\,dF_X(x) = k\left\{\int_0^\infty x^{k-1}\,[1 - F_X(x)]dx - \int_{-\infty}^0 x^{k-1}\,F_X(x)dx\right\}$$

- Se X é uma variável aleatória contínua com densidade $f(x)$, então

$$E(\varphi(X)) = \int \varphi(x)\, f(x) dx$$

- Se X é uma variável aleatória discreta com função de probabilidade $p(x_i)$, então

$$E(\varphi(X)) = \sum_i \varphi(x_i)\, p(x_i)$$

3.6 Momentos

A definição dos momentos é importante para caracterização de uma variável aleatória. A seguir apresentamos os principais momentos de uma variável aleatória X.

- O k-ésimo momento de X em torno de b, para $b \in \mathbb{R}$ é definido por

$$E(X-b)^k, \quad k = 1, 2, \ldots$$

- Naturalmente para $b = 0$ a $E(X^k)$ é o k-ésimo momento em torno de zero ou simplesmente o momento de ordem k de X.

- Se X é integrável, então o k-ésimo momento em torno da média é definido por

$$E\left(X - E(X)\right)^k, \quad k = 1, 2, \ldots$$

  e denominado de k-ésimo momento central de X.

- Para $k = 1$ o momento de X é a esperança e o momento central de X é nulo.

- O segundo momento central de X é chamado de variância e é definido por

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

- Para $t > 0$, o t-ésimo momento absoluto de X definido como $E(|X|^t)$. Os momentos absolutos possuem a seguinte propriedade de monotonia. Se X é uma variável aleatória, então a função

$$f(t) = \left[E(|X|^t)\right]^{1/t}$$

  é não decrescente em t para $t > 0$.

- Se $E(|X|^t)$ é finita para algum $0 < t < \infty$, então $E(|X|^s)$ é finito para todo s tal que $0 < s < t$.

A seguir algumas propriedades importantes da esperança, variância e momentos.

1. Se $X = c$ então $\text{Var}(X) = 0$, isto é, uma constante não tem variação.

2. $Var(aX + b) = a^2 \,\text{Var}(X)$, para todo $a, b \in \mathbb{R}$.

3. Desigualdade básica de Tchebychev: Seja X uma variável aleatória não-negativa ($X \geq 0$). Para todo $\lambda > 0$, temos que

$$P(X \geq \lambda) \leq \frac{1}{\lambda} E(X).$$

   Algumas consequências são:

   (a) Desigualdade clássica de Tchebychev. Se X é integrável, então

$$P(|X - E(X)| \geq \lambda) \leq \frac{\text{Var}(X)}{\lambda^2}, \quad \forall \lambda > 0.$$

   (b) Desigualdade de Markov. Seja X uma variável aleatória qualquer, então para todo $t > 0$, têm-se que

$$P(|X| \geq \lambda) \leq \frac{E(|X|^t)}{\lambda^t}, \quad \forall\, \lambda > 0.$$

   (c) Se $Z \geq 0$ e $E(Z) = 0$, então $P(Z = 0) = 1$. Ou seja, $Z = 0$ quase certamente.

   Deste resultado temos que quando a $Var(X) = 0$ então $E(X - EX)^2 = 0$ e $P((X - EX)^2 = 0) = 1$, logo $P(X = EX) = 1$. Portanto, se $Var(X) = 0$, então X é constante com probabilidade 1 (ou quase certamente).

Em geral queremos predizer um valores de uma variável aleatória X qualquer. Uma pergunta natural que surge é: qual o melhor preditor? Dois casos bem conhecidos: (i) se queremos minimizar o erro quadrático médio, o melhor preditor é a média; (ii) se queremos minimizar o erro absoluto médio, o melhor preditor é a mediana. Segue então a formalização desses conceitos.

Se a variável aleatória X é integrável então $\mu = E(X)$ minimiza $E(X - c)^2$ com $c \in \mathbb{R}$. Ou seja,

$$Var(X) = E(X - \mu)^2 = \min_{c \in \mathbb{R}} E(X - c)^2.$$

Seja m uma mediana da variável aleatória X então m minimiza $E|X-c|$ com $c \in \mathbb{R}$. Ou seja,

$$E|X-m| = \min_{c\in\mathbb{R}} E|X-c|.$$

3.7 Esperanças de funções de vetores aleatórios

Seja $\mathbf{X} = (X_1, \ldots, X_n)$ um vetor aleatório e $\varphi : \mathbb{R}^n \to \mathbb{R}$ uma função mensurável a Borel. Então,

$$E\varphi(\mathbf{X}) := \int_{\mathbb{R}^n} \varphi\, dF_{\mathbf{X}} = \int \cdots \int \varphi(x_1, \ldots, x_n)\, dF_{\mathbf{X}}(x_1, \ldots, x_n). \tag{26}$$

A partir desta definição surgem algumas propriedades e casos especiais, que são apresentados a seguir.

1. **Caso Discreto**: Se X for discreto, tomando valores $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})$ com probabilidade $p(\mathbf{x}_i)$, então

$$E\varphi(\mathbf{X}) = \sum_i \varphi(\mathbf{x}_i)\, p(\mathbf{x}_i).$$

2. **Caso Contínuo**: Se X for contínuo com densidade $f(x_1, \ldots, x_n)$, então

$$E\varphi(\mathbf{X}) = \int \cdots \int \varphi(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n.$$

3. **Linearidade**: A esperança é linear para combinações lineares de variáveis aleatórias, isto é,

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

4. **Independência**: Se X_i, para $i = 1, \ldots, n$ são independentes então a definição (26) se torna:

$$E\varphi(\mathbf{X}) = \int \cdots \int \varphi(x_1, \ldots, x_n)\, dF_{X_1}(x_1), \ldots, dF_{X_n}(x_n).$$

   Como consequência, se X_i, para $i = 1, \ldots, n$ são integráveis, então

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

Desta última igualdade vale ressaltar que se $E(XY) = E(X) \cdot E(Y)$ não necessariamente X e Y são independentes. Assim, podemos pensar em uma medida de dependência **linear** entre duas variáveis aleatórias X e Y como sendo a diferença entre os valores $E(XY)$ e $E(X) \cdot E(Y)$. Tal medida é denominada de covariância entre X e Y. Ela é definida por

$$\mathrm{Cov}(X, Y) = E\left[(X - EX)(Y - EY)\right] = E(XY) - EX\,EY. \tag{27}$$

Para introduzir uma medida que não depende da locação nem da escala da variável aleatória, considera-se as seguintes variáveis aleatórias $\dfrac{X - EX}{\sigma_X}$ e $\dfrac{Y - EY}{\sigma_Y}$. Podemos observar que ambas variáveis tem média 0 e variância 1. Ou seja, não dependem da locação e escala de X e Y. Assim, defini-se o coeficiente de correlação **linear** entre X e Y como sendo

$$\rho_{X,Y} := \frac{\mathrm{Cov}(X, Y)}{\sigma_X\,\sigma_Y} = E\left[\left(\frac{X - EX}{\sigma_X}\right)\left(\frac{Y - EY}{\sigma_Y}\right)\right]. \tag{28}$$

As seguintes propriedades do coeficiente de correlação linear podem ser provadas

- $-1 \leq \rho(X, Y) \leq 1$;
- $\rho(X, Y) = 1$ se, e somente se, $P(Y = aX + b)$ para $a > 0$ e $b \in \mathbb{R}$. Em outras palavras $Y = EY + \dfrac{\sigma_Y}{\sigma_X}(X - EX)$ quase certamente.
- $\rho(X, Y) = -1$ se, e somente se, $P(Y = aX + b)$ para $a < 0$ e $b \in \mathbb{R}$. Em outras palavras $Y = EY - \dfrac{\sigma_Y}{\sigma_X}(X - EX)$ quase certamente.

Por fim, se $X_1, \ldots, X_n$ são variáveis aleatória integráveis, então

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

Se X_i forem independentes então $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$, logo a variância da soma será a soma das variâncias.

3.8 Teoremas de Convergência

Estes dois teoremas mencionados na sala de aula fornecem condições teóricas para que a esperança do limite seja igual o limite das esperanças, isto é,

$$\lim_{n\to\infty} E(X_n) = E\left(\lim_{n\to\infty} X_n\right).$$

Consideramos $X, X_1, X_2, \ldots$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathbb{A}, P)$ e suponhamos que $X_n \uparrow X$ quando $n \to \infty$ ou então

$\lim_{n\to\infty} X_n = X$. Dizemos que a convergência aqui é pontual, ou seja, $X_n(\omega) \to X(\omega)$ para todo $\omega \in \Omega$.

3.8.1 Teorema da Convergência Monótona

Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathbb{A}, P)$. Se $X_n(\omega) \geq 0$ e $X_n(\omega) \uparrow X(\omega)$, $\forall \omega \in \Omega$ quando $n \to \infty$ então $E(X_n) \uparrow E(X)$. Ou seja,

$$\lim_{n\to\infty} E(X_n) = E\left(X\right).$$

3.8.2 Teorema da Convergência Dominada

Sejam $Y, X, X_1, X_2, \ldots$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathbb{A}, P)$. Se Y é integrável, $|X_n| \leq Y$, para todo n e $X_n(\omega) \uparrow X(\omega)$, $\forall \omega \in \Omega$ quando $n \to \infty$ então $E(X_n) \uparrow E(X)$. Ou seja,

$$\lim_{n\to\infty} E(X_n) = E\left(X\right).$$

4 Distribuição e Esperança Condicionais

Iniciamos este capítulo em maio de 2019 discutindo a distribuição condicional de uma variável aleatória dado um evento A qualquer. Em seguida, definimos a distribuição condicional supondo que temos $A_n = [Y = y_n], n = 1, 2, \ldots$ uma partição de Ω formada pelos eventos de uma variável aleátoria discreta Y. Estendemos a definição de distribuição condicional para o caso geral e enunciamos duas definições "formais". Por fim, a esperança condicional é discutida.

4.1 Distribuição condicional de X dada Y discreta

Seja X uma variável aleatória definida em $(\Omega, \mathbb{A}, P)$ e seja A um evento aleatório tal que $P(A) > 0$. A distribuição condicional de X dado o evento A é definida por

$$P(X \in B \mid A) = \frac{P\left([X \in B] \cap A\right)}{P(A)}, \quad B \in \mathcal{B}$$

Está é de fato uma probabilidade nos borelianos, pois os axiomas podem ser verificados. Dessa forma, a função de distribuição e a esperança condicional de X dado A são definidas, respectivamente, por

$$F_X(x \mid A) = P(X \leq x \mid A) = \frac{P\left([X \in B] \cap A\right)}{P(A)},$$

e

$$E(X \mid A) = \int x \, dF_X(x \mid A).$$

Se os eventos aleatórios $A_1, A_2, \ldots$ formam uma partição de Ω, então pelo Teorema de Probabilidade Total, temos

$$P(X \in B) = \sum_n P(A_n) \, P(X \in B \mid A_n), \qquad \forall B \in \mathcal{B},$$

Usando este resutado a função de distribuição e esperança de X podem ser expressas por

$$F_X(x) = P(X \leq x) = \sum_n P(A_n) \, P(X \leq x \mid A_n) = \sum_n P(A_n) \, F_X(x \mid A_n), \quad \forall x$$

e

$$E(X) = \int x \, dF_x(x) = \int x \, d\left(\sum_n P(A_n) \, F_X(x \mid A_n)\right) = \sum_n P(A_n) \, E(X \mid A_n).$$

Consideramos agora o caso em que a partição é gerada por uma variável aleatória discreta Y definida no mesma espaço de probabilidade $(\Omega, \mathbb{A}, P)$ tomando valores $y_1, y_2, \ldots$. Os eventos $A_n = [Y = y_n]$ formam uma partição de Ω, então a distribuição condicional de X dado que $Y = y_n$ é dada por

$$P(X \in B \mid Y = y_n) = P(X \in B \mid A_n), \quad B \in \mathcal{B}.$$

De forma que a distribuição marginal de X pode ser caracterizada a partir da distribuição condicional a partir das seguintes relações:

$$P(X \in B) = \int P(X \in B \mid Y = y) dF_Y(y) = \sum_n P(Y = y_n)\, P(X \in B \mid Y = y_n), \quad B \in \mathcal{B},$$

$$F_X(x) = \int F_X(x \mid Y = y) dF_Y(y) = \sum_n P(Y = y_n)\, F_(x \mid Y = y_n), \quad x \in \mathbb{R},$$

e

$$E(X) = \int E(X \mid Y = y) dF_Y(y) = \sum_n P(Y = y_n)\, E(X \mid Y = y_n).$$

Observe que $E(X \mid Y = y)$ é o valor particular da variável aleatória $\varphi(Y) = E(X \mid Y)$ quando esta assume o valor $Y = y$. Ou seja,

$$E(X) = E\left[\varphi(Y)\right] = E\left[E(X \mid Y)\right].$$

Alguns exemplos do uso dessas definições foram apresentados. Entre eles estão:

1. Se as variávies aleatórias X e Y são independentes com $X \sim \text{Poisson}(\lambda_1)$ e $Y \sim \text{Poisson}(\lambda_2)$, então qual a distribuição de $X \mid X + Y$.

2. Seja $Y \sim \text{Poisson}(\lambda)$ e $X \mid Y = y \sim \text{Binomial}(n, p)$, então qual o valor esperado de X.

4.2 Distribuição condicional de X dada Y caso geral

Nesta seção duas definições que caracterizam a distribuiçõa de X dada uma variável aleatória Y (independente da sua natureza) são apresentadas. Na sequência, três casos de "fáceis" soluções serão mencionados. Dois princípios que auxiliam na determinação da distribuição condicional são explicados brevemente.

- **Definição I:**

Sejam X e Y variáveis aleatórias definida no mesmo espaço de probabilidade $(\Omega, \mathbb{A}, P)$. Uma função $P(X \in B \mid Y = y)$, definida para B borelianao e $y \in \mathbb{R}$ será chamada uma distribuição condicional para X dada Y se

1. para todo $y \in \mathbb{R}$ fixo, $P(X \in B \mid Y = y)$ define uma probabilidade em $(\Omega, \mathbb{A}, P)$.
2. para todo $B \in \mathcal{B}$ fixo, $P(X \in B \mid Y = y)$ é uma função mensurável de y;
3. para todo $(x, y) \in \mathbb{R}^2$

$$\int_{-\infty}^{y} P(X \leq x \mid Y = t) dF_Y(t) = P(X \leq x, Y \leq y).$$

ou seja,

$$\int_{-\infty}^{y} F_X(x \mid Y = t) dF_Y(t) = F_{X,Y}(x, y).$$

- **Definição II:**

Sejam $y \in \mathbb{R}$ e $B \in \mathcal{B}$, a probabilidade condicional de que $X \in B$ dado que $Y = y$ é definida por

$$P(X \in B \mid Y = y) = \lim_{\Delta y \to 0} P(X \in B \mid Y \in I),$$

em que I representa um intervalo de comprimento Δy contendo y.

Os casos de "fáceis" soluções são

1. Quando a variável aleatória Y é discreta.

   Neste caso já definidmos a distribuição condicional para uma sequeência finita ou enumerável de valores $y_1, y_2, \ldots$ tais que $P(Y \leq y_n) \geq 0$ e admitivos a definição arbitrária para o conjunto dos outros valores de y com probabilidade zero.

2. Quando X e Y são variávies aleatórias independentes.

   Intuitivamente, a distribuição condicional de X dado que $Y = y$ vai depender de y. Logo, o candidato é

$$P(X \in B \mid Y = y) = P(X \in B), \qquad B \in \mathcal{B}, y \in \mathbb{R},$$

3. Quando X e Y possuem densidade conjunta $f(x, y)$.

   Aqui o candidato para densidade condicional será

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}, \quad x \in \mathbb{R}.$$

Princípio da preservação de chances relativas: Os resultados possíveis de uma variável aleatória condicionada a ocorrência de um evento A mantêm as mesmas chances relativas que tinham antes da realização do experimento. Em outras palavras, dado que $Y = y$, os valores possíveis de X mantêm as mesmas chances relativas de antes do experimento. Ou seja, um valor possível de X é um x tal que (x, y) era um valor possível de (X, Y) antes do experimento, e o princípio diz que estes pontos x mantêm na distribuição condicional, as mesmas chances relativas que os pontos (x, y) tinham na distribuição conjunta de X e Y.

Para X e Y conjuntamente contínuas com densidade f(x,y) este princípio pode ser expresso da seguinte forma

$$\frac{f(x_1 \mid y)}{f(x_2 \mid y)} = \frac{f(x_1, y)}{f(x_2, y)}.$$

Princípio da substituição para distribuição condicional: Sejam X e Y variáveis aleatórias definida em $(\Omega, \mathbb{A}, P)$ e seja $\varphi(x, y)$ uma função mensurável. Se a distribuição condicional de X dada Y é dada por

$$P(X \in B \mid Y = y),$$

então a distribuição condicional de $\varphi(X, Y)$ dada Y é

$$P(\varphi(X, Y) \in B \mid Y = y) = P(\varphi(X, y) \in B \mid Y = y) = P(X \in \{x : \varphi(x, y) \in B\} \mid Y = y)$$

com $B \in \mathcal{B}, y \in \mathbb{R}$. Simbolicamente temos

$$\varphi(X, Y) \mid Y = y \sim \varphi\left(X \mid Y = y =, y\right).$$

Este princípio fornece condições mais práticas para se obter a distribuição condicional de $\varphi(X, Y)$ dado que $Y = y$. Para isso, basta substituir Y pelo valor y e X pela variável aleatória condicional.

4.3 Esperança Condicional

Sejam X e Y variáveis aleatórias definida em $(\Omega, \mathbb{A}, P)$. A esperança condicional de X dado que $Y = y$, é a esperança condicional da **distribuição** condicional de X dado que $Y = y$, ou seja

$$E(X \mid Y = y) = \int x \, dF(x \mid Y = y)$$

Definindo a variável aleatória $\varphi(Y) = E(X \mid Y)$ é denominada de esperança condicional de X dada Y. A principal propriedade desta variável aleatória é

$$E\left[\varphi(Y)\right] = E\left[E(X \mid Y\right] = E(X).$$

De forma análoga,

$$E(X) = \int E(X \mid Y = y)\, dF_Y(y).$$

O princípio da substituição para esperança condicional afirma que se $\varphi(X, Y)$ é integrável, então

$$E(\varphi(X, Y) \mid Y = y) = E(\varphi(X, y) \mid Y = y) = \int \varphi(x, y)\, dF_X(x \mid Y = y).$$

Decorre deste princípio que se X é integrável e $Xg(y)$ também é, então

$$E(Xg(Y) \mid Y = y) = g(y)E(X \mid Y = y).$$

Portanto, temos a seguinte igualdade de distribuições

$$E(Xg(Y) \mid Y) \sim g(Y)E(X \mid Y).$$

Na última aula desta seção provamos a desigualdade de Cauchy-Schwarz para variáveis aleatórias. Nesta desigualdade temos que

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

Desta desigualdade concluímos que XY será integravel se o segundo momento de X e Y existir. Em outas palavras, se X e Y tiverem variâncias finitas.

Finalizando a seção determinamos a distribuição condicional do evento aleatória A dada uma variável aleatória Y. Observando que $A = [I_A = 1]$, assim $P(A) = P(I_A = 1)$ de modo que

$$P(A \mid Y = y) = P(I_A = 1 \mid Y = y).$$

Como I_A assume somente dois valores, pelo princípio de preservação de chances relativas temos que

$$P(I_A = 1 \mid Y = y) + P(I_A = 0 \mid Y = y) = 1$$

Portanto, a esperança condicional de I_A dado Y é dada por

$$E(I_A \mid Y = y) = 1 \cdot P(I_A = 1 \mid Y = y) + 0 \cdot P(I_A = 0 \mid Y = y) = P(I_A = 1 \mid Y = y).$$

Dessa forma, temos a seguinte variável aleatória $\varphi(Y) = E(I_A \mid Y) = P(A \mid Y)$. Calculando a esperança dessa variável aleatória concluímos que

$$E(I_A) = P(A) = E(E(I_A \mid Y)) = E(P(A \mid Y)).$$

ou seja, a probabilidade de um evento A é a esperança de sua probabilidade condicional dada uma variável aleatória Y qualquer.

5 A Lei dos Grandes números

Esta é uma seção altamente teórica sob o ponto de vista matemático. Inicia-se com uma discussão intuitiva da Lei dos Grandes Número. Destaca-se a definição de convergência “em probabilidade” e “quase certa”. A partir deste conceito apresenta-se a Lei Fraca de Tchebychev, a Lei dos Grandes Números de Bernoulli e a Lei Fraca de Khintchin. O capítulo segue definindo o que é sequencias de eventos e enunciando e provando o Lema de Borel-Cantelli. Na sequência apresenta-se três importantes teoremas, sendo eles: Recíproca para a Lei Forte de Kolmogorov, Primeira Lei Forte de Kolmogorov, Lei Forte de Kolmogorov. Aqui não apresento resultado das provas, apenas exponho as definições.

5.1 Definições Importantes

Suponha que estamos interessados em estudar uma variável aleatória X que descreve uma determinada característica sobre um fenômeno aleatório qualquer. Para estudar o comportamento da variável aleatória X conduzimos um experimento e observamos seu valor. Realizando um número n de vezes o mesmo experimento de forma independente observamos n vezes o comportamento da variável aleatória X. Sob essas condições a Lei dos Grandes Número afirma que a média aritmética dos n valores observados é aproximadamente igual ao valor esperado de X quando n tende ao infinito.

Em uma linguagem formal e probabilística se $X_1, X_2, \ldots$ são variáveis aleatórias independentes, integráveis e identicamente distribuídas então

$$\frac{X_1 + \ldots + X_n}{n} \to E(X_1).$$

A expressão acima indica que as variáveis aleatórias convergem para determinado número. O ponto agora é determinar qual o tipo de convergência. As definições de convergência estão apresentadas a seguir:

- **Convergência em Probabilidade:** Para todo $\varepsilon > 0$, dizemos que Y_n converge para Y em probabilidade se

$$P(|Y_n - Y| \geq \varepsilon) \to 0 \quad \text{quando} \quad n \to \infty.$$

  Este tipo de convergência é denotada por: $Y_n \xrightarrow{P} Y$.

- **Convergência quase certa:** Dizemos que Y_n converge para Y quase certamente se

$$P(Y_n \to Y \,\text{quando}\, n \to \infty) = 1.$$

Analogamente se o evento

$$\{\omega : Y_n(\omega) \to Y(\omega)\}$$

tem probabilidade 1, então Y_n converge para Y quase certamente.

Este tipo de convergência é denotada por: $Y_n \xrightarrow{q.c.} Y$.

Importante mencionar que convergência em probabilidade é mais fraca que a convergência quase certa. Além disso, a convergência quase certa implica em convergência em probabilidade, sendo que a recíproca não é válida. Com a definição desses tipos de convergência estamos aptos a definir as Lei Fraca e Forte dos grandes números.

Sejam $X_1, X_2, \ldots$ variáveis aleatórias integráveis em $(\Omega, \mathbb{A}, P)$. Considere $S_1, S_2, \ldots$ as somas parciais, isto é, $S_n = X_1 + \ldots + X_n$.

Dizemos que $X_1, X_2, \ldots$ satisfazem a Lei Fraca dos Grandes Números se

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

ou, equivalentemente, se

$$P\left(\left|\frac{X_1 + \ldots X_n - (EX_1 + \ldots + EX_n)}{n}\right| \geq \varepsilon\right) \to 0, \quad \forall \varepsilon > 0.$$

Dizemos que $X_1, X_2, \ldots$ satisfazem a Lei Forte dos Grandes Números se

$$\frac{S_n - ES_n}{n} \xrightarrow{q.c.} 0.$$

ou, equivalentemente, se

$$\frac{(X_1 - EX_1) + \ldots + (X_n - EX_n)}{n} \xrightarrow{q.c.} 0.$$

Para finalizar esta seção apresento a definição de três Teoremas relacionados a convergência em probabilidade.

- **Lei Fraca de Tchebychev:** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes 2 a 2 com variância finitas e uniformemente limitadas. Então,$X_1, X_2, \ldots$ satisfazem a Lei Fraca dos Grandes Números, isto é,

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

  Importante observar que a suposição de independência pode ser alterada para a suposição de que as X_n sejam não correlacionadas.

- **Lei dos Grandes Números de Bernoulli:** Considere uma sequência de ensaios binomiais independentes com a mesma probabilidade p de sucesso. Se S_n é o número de sucessos nos n primeiros ensaios, então

$$\frac{S_n}{n} \xrightarrow{P} p.$$

- **Lei Fraca de Khintchin:** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, com média comum μ, então
$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

5.2 O Lema de Borel-Cantelli

O Lema de Borel-Cantelli é uma importante ferramenta utilizada para provar a Lei Forte dos grandes números.

Sejam $A_1, A_2, \ldots$ uma sequência de eventos aleatórios em $(\Omega, \mathbb{A}, P)$, ou seja, $A_n \in \mathbb{A}\, \forall n$. Define-se o limite superior da sequência como

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \tag{29}$$

e o limite inferior é, por definição

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \tag{30}$$

O evento $\limsup A_n$ pode ser interpretado como o evento "ocorrência de um número infinito dos A_n". A notação utilizada é

$$\limsup A_n = [A_n \text{ infinitas vezes}].$$

O evento $\liminf A_n$ pode ser interpretado como o evento "ocorrência de A_n, para todo n suficientemente grande".

O Lema de **Borel-Cantelli** afirma o seguinte:

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

(b) Se $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$ e os A_n são independentes, então $P(A_n \text{ infinitas vezes}) = 1$.

Importante mencionar as seguintes igualdades entre eventos:

$$[A_n \text{ infinitas vezes}]^c = [A_n \text{ finitas vezes}].$$

$$[A_n \text{ finitas vezes}] \Rightarrow [A_n^c \text{ infinitas vezes}],$$

ou seja,

$$[A_n^c \text{ infinitas vezes}] \subset [A_n \text{ finitas vezes}].$$

O Lema de Borel-Cantelli é utilizado como uma ferramenta que estabelece condições suficientes para convergência quase certa. Temos a seguinte relação

$$X_n \xrightarrow{q.c.} X \iff P(|X_n - X| > \varepsilon \text{ infinitas vezes}) = 0 \quad \forall \varepsilon > 0.$$

Em particular, se $p_n(\varepsilon) = P(|X_n - X| > \varepsilon)$ satisfaz $\sum_{n=1}^{\infty} p_n(\varepsilon) < \infty$ para todo $\varepsilon > 0$, então $X_n \xrightarrow{q.c.} X$.

5.3 A Lei Forte

Finalizando esta seção enunciaremos duas versões da Lei Forte de Kolmogorov que é conhecida na dia a dia como a Lei Forte dos Grandes Números. A prova dos Teoremas são altamente técnicas e estão apresentadas no livro do Barry James.

- **Primeira Lei Forte de Kolmogorov:** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes e integráveis, e suponha que

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

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

  $$\frac{X_1 + \cdots + X_n}{n} - \frac{EX_1 + \cdots EX_N}{n} \xrightarrow{q.c.} 0.$$

- **A Lei Forte de Kolmogorov:** Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas e integráveis, tal que $EX_n = \mu$. Então

  $$\frac{X_1 + \cdots X_n}{n} \xrightarrow{q.c.} \mu.$$

6 Funções que Caracterizam uma Distribuição

Esta seção é dedicada para definição de algumas funções importantes que servem para caracterizar uma distribuição de probabilidade.

6.1 Funções Características

As funções características são úteis para estudo de convergência em distribuição e possuem propriedades atraentes sob o ponto de vista matemático.

Seja X e Y variáveis aleatórias definidas em $(\Omega, \mathbb{A}, P)$, então $Z = X + iY$ é denominada de variável aleatória complexa, em que i representa o número imaginário $\sqrt{-1}$.

A função característica de um variável aleatória X é definida por

$$\varphi_X(t) = E(e^{itX}) = \int e^{itx} dF_X(x), \quad t \in \mathbb{R}. \tag{31}$$

Note que $\varphi : \mathbb{R} \to \mathbb{C}$.

A formula de Euler afirma que $e^{iX} = \cos(X) + i\,\mathrm{sen}(X)$. Dessa forma a função característica de X definida em (31) pode ser expressa por

$$\varphi_X(t) = E\left[\cos(tX)\right] + iE\left[\mathrm{sen}(tX)\right] = \int \cos(tx)\, dF_X(x) + i \int \mathrm{sen}(tx) dF_X(x). \tag{32}$$

Na sequência apresentamos algumas propriedades importantes da função característica.

(i) A função característica é limita por 1, isto é, $|\varphi_X(t)| \leq 1, \forall\, t \in \mathbb{R}$.

(ii) A função característica assume valor 1 para $t = 0$, isto é, $\varphi_X(0) = 1$.

(iii) $\overline{\varphi_X(t)} = \varphi_X(-t)$, em que $\bar{c}$ é o complexo conjugado de c. Por exemplo, se $c = x + iy$, então $\bar{c} = x - iy$

(iv) φ_X é uniformemente limitada na reta.

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

(vi) A função característica de uma variável aleatória X determina (caracteriza) a função de distribuição de X, isto é, φ_X determina F_X. Este resultado é conhecido como o Teorema da Unicidade e pela fórmula da inversão temos

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

(vii) A variável aleatória X tem distribuição simétrica em torno de zero se, e somente se, $\varphi_X(t)$ é uma função real para todo t.

(viii) Se $Y = a\,X + b$, então $\varphi_Y(t) = e^{itb}\,\varphi_X(at)$.

(ix) Se $E|X|^n < \infty$, então φ_X possui n derivadas contínuas e

$$\varphi_X^{(k)}(t) = \int (it)^k\, e^{itx}\, dF_X(x), \quad k = 1, 2, \ldots, n.$$

Em particular,

$$\varphi_X^{(k)}(0) = i^k\, EX^k, \quad k = 1, 2, \ldots, n.$$

de modo que a função característica pode determinar os momentos da variável aleatória X.

É possível estender a definição de funções características para o caso de vetores aleatórios. Seja $\mathbf{X} = (X_1, \ldots, X_k)$ um vetor aleatório k-dimensional. Então a função característica de $\mathbf{X}$ é definida por

$$\varphi_{\mathbf{X}}(\boldsymbol{t}) = E \exp\left(i \sum_{j=1}^{k} t_j X_j\right) = E e^{i\boldsymbol{t}\cdot\mathbf{X}} \tag{33}$$

em que $\boldsymbol{t} \cdot \mathbf{X}$ denota o produto interno entre $\boldsymbol{t}$ e $\mathbf{X}$. Note que $\varphi : \mathbb{R}^k \to \mathbb{C}$.

As propriedades da função característica multivariada são análogas da função característica de uma variável aleatória. Destacamos duas propriedades úteis:

- Para obter a função característica de qualquer distribuição marginal, basta tomar os outros termos iguais a zero. Por exemplo, para as variáveis aleatórias X, Y e Z, temos que

$$\varphi_{X,Y}(x, y) = \varphi_{X,Y,Z}(x, y, 0), \quad (x, y) \in \mathbb{R}^2.$$

7 Convergência em Distribuição

Nesta seção discutimos a convergência em distribuição e várias ferramentas que são úteis para resolução de exercícios para determinar a convergência de distribuição de uma sequência de variáveis aleatórias.

Sejam X, $X_1, X_2, \ldots$ variáveis aleatórias com, respectivamente função de distribuição $F, F_1, F_2, \ldots$. Dizemos que X_n converge em distribuição para X, quando $n \to \infty$, se

$$F_n(x) \to F(x)$$

para todo ponto de continuidade de F. A notação utilizada é $F_n \xrightarrow{D} F$.

Importante de mencionar que este tipo de convergência é o mais fraco no sentido de que a convergência quase certa e em probabilidade implicam a convergência em distribuição. Também vale notar que a convergência em distribuição é definida em termos da função de distribuição o que implicitamente significa que as variáveis aleatórias não precisam estar definidas no mesmo espaço de probabilidade.

Os teoremas enunciado a seguir forneceram suporte para avaliar a convergência em distribuição com base na função característica.

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

$$\int g(x)dF_n(x) \xrightarrow{n\to\infty} \int g(x)dF(x)$$

  para toda função $g : \mathbb{R} \to \mathbb{R}$ contínua e limitada. Em outras palavras se $X_n \xrightarrow{D} X$, então

$$Eg(X_n) \xrightarrow{n\to\infty} Eg(X).$$

  Em particular como as funções $\operatorname{sen}(tx)$ e $\cos(tx)$ são limitadas temos que

$$\varphi_{X_n}(t) \xrightarrow{n\to\infty} \varphi_X(t), \quad \forall t \in \mathbb{R}.$$

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

  1. existe uma função de distribuição F tal que $F_n \xrightarrow{n\to\infty} F$;
  2. φ é a função característica de F.

- **Teorema de Cramér-Wold:** Sejam $\mathbf{X}_n = (X_{n1}, \ldots, X_{nk})$ e $\mathbf{X} = (X_{01}, \ldots, X_{0k})$ vetores aleatórios k-dimensionais. Dizemos que $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$ se, e somente se,

$$\sum_{j=1}^{k} t_j X_{nj} \xrightarrow{D} \sum_{j=1}^{k} t_j X_{0j} \quad \text{quando } n \to \infty$$

para todo $(t_1, \ldots, t_k) \in \mathbb{R}$.

Para os casos em que a variável aleatória é discreta ou contínua existem definições particulares para avaliar a convergência em distribuição.

Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias discretas que assumem somente valores inteiros não negativos, isto é, $k = 0, 1, 2, \ldots$ e sejam $p(k), p_1(k), p_2(k), \ldots$ as respectivas funções de probabilidade. Então, $X_n \xrightarrow{D} X$ se, e somente se,

$$p_n(k) \to p(k) \quad \text{quando } n \to \infty \quad \forall\, k = 0, 1, 2, \ldots.$$

Para o caso de variáveis aleatórias absolutamente contínuas o Teorema de Scheffé estabelece condições alternativas para convergência de distribuição. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias com funções de densidade $f, f_1, f_2, \ldots$, respectivamente. Então $X_n \xrightarrow{D} X$ se

$$f_n(x) \to f(x) \quad \text{quando } n \to \infty$$

para todo ponto x de medida de Lebesgue não nula.

Ressaltamos algumas proposições importantes. Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias, $g : \mathbb{R} \to \mathbb{R}$ e c uma constante. Então,

1. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$;
2. $X_n \xrightarrow{D} c \Rightarrow X_n \xrightarrow{P} c$;
3. $X_n \xrightarrow{q.c.} X \Rightarrow g(X_n) \xrightarrow{q.c.} g(X)$;
4. $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$;
5. $X_n \xrightarrow{D} X \Rightarrow g(X_n) \xrightarrow{D} g(X)$.

O Teorema de Slutsky estabelece condições para convergência de soma e produto de variáveis aleatórias. Sejam $X, X_1, X_2, \ldots$ e $Y, Y_1, Y_2, \ldots$ variáveis aleatórias tais que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, onde c é uma constante. Então,

1. $X_n + Y_n \xrightarrow{D} X + c$;
2. $X_n Y_n \xrightarrow{D} cX$.

Finalizando essa seção vamos enunciar um resultado utilizado frequentemente em inferência estatística, o bem conhecido Método Delta.

Sejam $Y_1, Y_2, \ldots$ variáveis aleatórias tais que $\sqrt{n}\,(Y_n - \mu) \xrightarrow{D} N(0, \sigma^2)$. Se $g(y)$ é uma função derivável no ponto μ, então $\sqrt{n}\,(g(Y_n) - g(\mu)) \xrightarrow{D} N\left(0, \sigma^2\ (g'(\mu))^2\right)$.

8 O Teorema Central do Limite

A última seção desta nota de aula enuncia três diferentes versões do Teorema Central do Limite. Este teorema apresenta condições suficientes para garantir a convergência em distribuição das somas parciais normalizadas de variáveis aleatórias. Ou seja, estamos interessados em determinar as condições sob as quais

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} \xrightarrow{D} N(0,1)$$

em que $S_n = X_1 + \cdots + X_n$.

(i) Variáveis Aleatórias Independentes e Identicamente Distribuídas

Este é o caso mais simples do Teorema Central do Limite.

Sejam X_1, X_2 variáveis aleatórias independentes e identicamente distribuídas, com média μ e variância $\sigma^2 > 0$. Então,

$$\frac{S_n - ES_n}{\sqrt{\mathrm{Var}S_n}} \xrightarrow{D} N(0,1), \quad \text{ou seja,} \quad \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1).$$

(ii) Teorema Central do Limite de Liapunov

Neste caso, a hipótese de que as variáveis aleatórias devem ter distribuições iguais é relaxada.

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes tais que $EX_n = \mu_n$ e $\mathrm{Var}X_n = \sigma_n^2 < \infty$ com pelo menos um $\sigma_n^2 > 0$. Seja $s_n^2 = \sigma_1^2 + \cdots \sigma_n^2$. Se existir $\delta > 0$ tal que

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - \mu_k|^{2+\delta} \xrightarrow{n \to \infty} 0,$$

então

$$\frac{S_n - ES_n}{s_n} \xrightarrow{D} N(0,1).$$

(iii) Teorema Central do Limite de Lindeberg

Esta versão abrange as duas outras versões.

Sejam $X_1, X_2, \dots$ variáveis aleatórias independentes e $F_n = F_{X_n}$ as respectivas funções de distribuições tais que $EX_n = \mu_n$ e $\mathrm{Var}X_n = \sigma_n^2 < \infty$ com pelo menos um $\sigma_n^2 > 0$. Seja $s_n^2 = \sigma_1^2 + \cdots \sigma_n^2$ Se a seguinte condição

$$\forall\, \varepsilon > 0, \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| > \varepsilon s_n} (x - \mu_k)^2 \, dF_k(x) = 0$$

denominada de condição de Lindeberg estiver satisfeita, então

$$\frac{S_n - ES_n}{s_n} \xrightarrow{D} N(0,1).$$

A condição de Lindeberg também pode ser escrita da seguinte forma

$$\forall\, \varepsilon > 0, \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k| \leq \varepsilon s_n} (x-\mu_k)^2 \, dF_k(x) \to 1 \text{ quando } n \to \infty.$$

Vale mencionar que a condição de Lindeberg implica que

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \to 0 \text{ quando } n \to \infty.$$