2
Conceitos de Teoria da Probabilidade

Neste capítulo, enunciaremos algumas definições e resultados de teoria de probabilidade. A justificativa deste capítulo reside no fato que u objetivo final é estimar momentos de distribuição de probabilidades do resultado de simulações numéricas de um fenômeno físico (no caso simulação de reservatórios de petróleo) a partir das distribuições dos seus parâmetros de entrada. Para conseguirmos investigar o problema em questão, temos que apresentar algumas noções da teoria de probabilidade. Por exemplo, variáveis aleatórias e suas propriedades. Assim, precisamos definir a estrutura do espaço onde iremos trabalhar, pois isto nos permitirá fazer simplificações importantes. Portanto, o principal objetivo deste capítulo é fazer um breve levantamento das noções e resultados tidos como essenciais para a leitura do material. Para realizar este compêndio de resultados foram utilizados os livros: Brexis (2011); Breiman (2009); Durret (2010); Khoshnevisan (2007); Kubrusly (2007, 2010).

2.1
Espaços de Probabilidade

Denotemos Θ o conjunto de todos os possíveis resultados de um experimento. O conjunto Θ é chamado de **espaço de amostras**, onde θ representa um elemento de Θ. Note que Θ pode ser um conjunto finito ou infinito (contável ou não) de elementos.

Intuitivamente, existem relações entre a probabilidade para que um evento ocorra ou para que o mesmo evento não aconteça. Para dar conta destas relações entre eventos, necessitamos introduzir o conjunto chamado de $\sigma-$álgebra.

Definição 2.1 *Uma $\sigma-$álgebra denotada por $\mathcal{F}$ é uma coleção não vazia de subconjuntos de Θ que satisfaz os três axiomas abaixo:*

(1) $\emptyset \in \mathcal{F}$ *e* $\Theta \in \mathcal{F}$

(2) $A \in \mathcal{F}$ *implica* $A^c \in \mathcal{F}$

(3) $A_i \in \mathcal{F}$ *então* $\cup_{i\in I} A_i \in \mathcal{F}$ *e* $\cap_{i\in I} A_i \in \mathcal{F}$,

onde A^c denota o complemento de A em Θ e I é um conjunto de índices contável. Além disso, um elemento $A \in \mathcal{F}$ é chamado de subconjunto $\mathcal{F}$-mensurável de Θ.

Exemplo 2.1 Alguns exemplos de $\sigma-$álgebra são:

(1) $\mathcal{F}_1 = \{\emptyset, \Theta\}$

(2) $\mathcal{F}_2 = 2^\Theta$

(3) $\mathcal{F}_3 = \mathcal{B}(\mathbb{R})$ a coleção de todos os intervalos abertos de $\mathbb{R}$ (espaço conhecido como $\sigma-$álgebra de Borel).

Um **espaço mensurável** é dado por um espaço amostral Θ e uma $\sigma-$álgebra $\mathcal{F}$ (i.e., $(\Theta, \mathcal{F})$ é o espaço no qual vamos colocar uma medida).

Definição 2.2 *Uma função com valor real estendido $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ é uma medida, se este satisfaz as seguintes condições:*

(1) $\mu(\emptyset) = 0$

(2) $\mu(A) \geq 0$ *para cada* $A \in \mathcal{F}$

(3) $\mu(\bigcup_n A_n) = \sum_n \mu_n$ *para cada família contável* $\{A_n\}$ *de conjuntos disjuntos dois a dois em* $\mathcal{F}$ *(i.e.,* $A_n \bigcap A_m = \emptyset$ *quando* $n \neq m$*).*

Dizemos que uma medida finita é uma função de valor real $\mu : \mathcal{F} \to \mathbb{R}$ tal que $\mu(A) < \infty, \forall A \in \mathcal{F}$ ou, equivalentemente, $\mu(\Theta) < \infty$. Em particular, uma medida finita de extrema importância é a medida de probabilidade. P é uma **medida de probabilidade** se P é uma medida finita que satisfaz os seguintes axiomas:

(1) $P : \mathcal{F} \to [0, 1]$

(2) $P(\Theta) = 1$.

O conceito de probabilidade é usado para medir a possibilidade de ocorrência de certos eventos. A atribuição desta probabilidade é baseada em evidência empírica. Em teoria de probabilidade, a Lei dos Grandes Números é a justificativa teórica para esta observação empírica (Xiu, 2010). A seguir, vamos estabelecer algumas propriedades de uma medida de probabilidade.

Proposição 2.1 *Para $A, B \in \Theta$ temos que:*

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- *Se A e B são disjuntos então* $P(A \cup B) = P(A) + P(B)$

- $P(A^c) = 1 - P(A), P(\Theta) = 1, P(\emptyset) = 0.$

Agora, podemos definir o conceito de espaço de probabilidade.

Definição 2.3 *Um **espaço de probabilidade** é uma tripla* $(\Theta, \mathcal{F}, P)$*, onde* Θ *é um espaço amostral,* $\mathcal{F}$ *é uma* $\sigma-$*álgebra de subconjuntos de* Θ *e* P *é uma medida de probabilidade.*

2.2 Funções Mensuráveis

Considerando os espaços $(\Theta_1, \mathcal{F}_{\Theta_1})$ e $(\Theta_2, \mathcal{F}_{\Theta_2})$ espaços mensuráveis, podemos definir o conceito de função mensurável.

Definição 2.4 *Uma função* $f : \Theta_1 \to \Theta_2$ *é mensurável de* $(\Theta_1, \mathcal{F}_{\Theta_1})$ *em* $(\Theta_2, \mathcal{F}_{\Theta_2})$*, se para cada* $A_2 \in \mathcal{F}_{\Theta_2}$

$$f^{-1}(A_2) = \{\theta \in \Theta_1 : f(\theta) \in A_2\} \in \mathcal{F}_{\Theta_1}.$$

Agora, lembrando que a $\sigma-$álgebra de Borel em $\mathbb{R}$, denotada por $\mathcal{B}(\mathbb{R})$, é a $\sigma-$álgebra gerada por todos os subconjuntos abertos em $\mathbb{R}$ na topologia usual, vamos definir o conceito de variável aleatória.

Definição 2.5 *Considere um espaço de probabilidade* $(\Theta, \mathcal{F}, P)$*. Uma variável aleatória* $f : \Theta \to \mathbb{R}$ *é uma função mensurável*[1] *de* $(\Theta, \mathcal{F})$ *em* $(\Theta, \mathcal{B}(\mathbb{R}))$*.*

Portanto, funções mensuráveis em espaços de probabilidade são chamadas de variáveis aleatórias. Por outro lado, se definirmos a função f, mensurável de $(\Theta_1, \mathcal{F}_{\Theta_1})$ em $(\Theta_2, \mathcal{F}_{\Theta_2})$, e $(\Theta, \mathcal{F}_{\Theta_1}, P)$ um espaço de probabilidade, segue que $Q : \mathcal{F}_{\Theta_2} \to [0, 1]$ definida por

$$Q(A_2 \in \mathcal{F}_{\Theta_2}) = P(f_{-1}(A_2)),$$

é uma medida de probabilidade em $(\Theta, \mathcal{F}_{\Theta_2})$. Em particular, chamamos Q da probabilidade induzida por f, ou simplesmente a distribuição de f.

Logo, a **distribuição** de uma variável aleatória X é a medida de probabilidade induzida pela função $X : \Theta \to \mathbb{R}$ no espaço mensurável $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, definida por

$$Q(B) = P(X^{-1}(B)),\ B \in \mathcal{B}.$$

[1]Mensurável no sentido de Borel

Esta definição implica que X é uma variável aleatória se e somente se $X^{-1}((-\infty, x]) \in \mathcal{F},\ \forall x \in \mathbb{R}$ (i.e., X é $\mathcal{F}$-mensurável). Em outras palavras, uma variável aleatória é uma função com valores reais cujo domínio é o espaço de amostras.

Por outro lado, a função de distribuição acumulada (ou simplesmente a função de distribuição) de uma variável aleatória X em um espaço de probabilidade $(\Theta, \mathcal{F}, P)$ é definida por

$$F_X(x) = P(X^{-1}((-\infty, x])) = P(\theta : X(\theta) \leq x) = P(X \leq x).$$

A seguir, a Proposição 2.2, enuncia algumas propriedades da função de distribuição F.

Proposição 2.2 *Seja F uma função de distribuição arbitrária. As afirmações abaixo são verdadeiras:*

(1) F é não-decrescente

(2) $\lim_{x\to\infty} F(x) = 1,\ \lim_{x\to-\infty} F(x) = 0$

(3) F é contínua pela direita, i.e. $\lim_{y\downarrow x} F(y) = F(x)$.

Demonstração: Ver, e.g., Durret, 2010, p. 10. □

Em particular, quando a função de distribuição $F(x) = P(X \leq x)$ tem a forma

$$F(x) = \int_{-\infty}^{x} f(y)dy,\ x \in \mathbb{R}, \tag{2-1}$$

dizemos que a variável aleatória X tem **função de densidade de probabilidade** ou simplesmente **função de densidade** f_X. Daí,

$$F(b) - F(a) = \int_a^b f_X(x)dx, \text{ onde } a \leq b.$$

Além disso, podemos destacar as seguintes propriedades da função de densidade de probabilidade (ver, e.g., Durret (2010)):

(1) $f_X(x) = F'_X(x)$

(2) $f_X(x) \geq 0$

(3) $\int_{-\infty}^{\infty} f_X(x)dx = 1.$

A seguir, vamos estabelecer quatro exemplos de funções de distribuição usuais: Uniforme; Exponencial; Gaussiana; e Log-Normal.

1. A distribuição Uniforme definida no intervalo $[a, b]$ tem uma densidade dada por:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{se } x \in [a, b], \\ 0, & \text{outro caso.} \end{cases} \tag{2-2}$$

e função de distribuição:

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq a, \\ \frac{x-a}{b-a}, & \text{se } x \in (a, b), \\ 1, & \text{se } x \geq b. \end{cases} \tag{2-3}$$

2. A distribuição Exponencial com taxa λ tem como função de densidade:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{se } x \geq 0, \\ 0, & \text{outro caso.} \end{cases} \tag{2-4}$$

Além disso, a sua função de distribuição é definida por:

$$F_X(x) = \begin{cases} 0, & \text{se } x \leq 0, \\ 1 - e^{-x}, & \text{se } x \geq 0. \end{cases} \tag{2-5}$$

3. A distribuição Gaussiana ou Normal é denotada por $\mathcal{N}(\mu, \sigma^2)$, com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$. A densidade de probabilidade desta distribuição é:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}], \ x \in \mathbb{R}. \tag{2-6}$$

Note que não existe uma expressão fechada para $F_X(x)$.

4. A distribuição Lognormal, denotada por $\ln\mathcal{N}(\mu, \sigma^2)$, com média $\mu \in \mathbb{R}$ e variância $\sigma^2 > 0$ tem densidade de probabilidade dada por:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}], \ x \in \mathbb{R}^+. \tag{2-7}$$

Note que uma distribuição Lognormal é uma distribuição de probabilidade contínua de uma variável cujo logaritmo é normalmente distribuído. Logo, se X é lognormalmente distribuída, então $Y = \log(X)$ tem distribuição Normal. Do mesmo modo, se Y tem distribuição Normal, então $X = \exp(Y)$ tem distribuição Lognormal. Na figura 2.1 encontramos a representação de três tipos diferentes de densidades Lognormal.

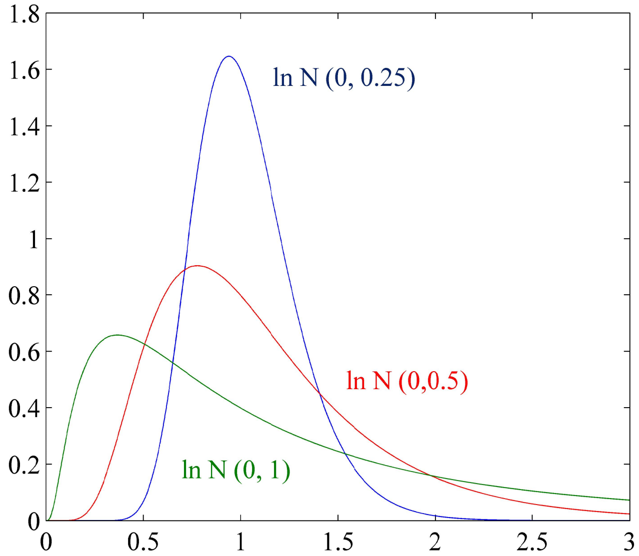

Figura 2.1: Densidade de Probabilidade Lognormal

Encerramos esta seção com alguns resultados para vetores aleatórios. Informalmente, um vetor aleatório é um vetor cujas coordenadas são variáveis aleatórias. A partir da Definição 2.5 podemos definir os vetores aleatórios como funções mensuráveis

$$\boldsymbol{X} : (\Theta, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}^n(\mathbb{R}^n)), \text{ tal que } \boldsymbol{X} = (X_1, \cdots, X_n).$$

Similarmente, podemos estender a definição de função de distribuição de uma variável aleatória para vetores aleatórios. Uma **função de distribuição conjunta** de um vetor aleatório $\boldsymbol{X}$ é:

$$F_{\mathbf{X}}(\boldsymbol{x}) = P\left(\bigcap_{i=1}^{n} \{X_i \leq x_i\}\right) = P(X_1 \leq x_1, \cdots, X_n \leq x_n),$$

onde $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. Daí, podemos verificar que:

(1) $\lim_{x_k \to -\infty} F(\boldsymbol{x}) = 0, \; k = 1, \cdots, n$

(2) $x_k \to F(\boldsymbol{x})$ é não-decrescente, onde $k = 1, \cdots, n$

(3) $x_k \to F(\boldsymbol{x})$ é contínua pela direita, onde $k = 1, \cdots, n$.

Além disso, se o vetor aleatório é dado por

$$\boldsymbol{X}_{|k} = (X_1, \cdots, X_{k-1}, X_k, \cdots, X_n),$$

então

$$\lim_{x_k \to \infty} F(\boldsymbol{x}) = F(\boldsymbol{x}_{|k})$$

é a distribuição conjunta do vetor aleatório em $\mathbb{R}^{n-1}$.

Por outro lado, se a distribuição de um vetor aleatório $\boldsymbol{X}$ tem **função de densidade conjunta** $f_{\boldsymbol{X}}(\boldsymbol{x})$, podemos representar a função de distribuição $F_{\boldsymbol{X}}(\boldsymbol{x})$ como:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{X}}(y_1, \cdots, y_n) dy_1 \cdots dy_n,$$

onde a função de densidade conjunta satisfaz:

(1) $f_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0, \ \forall x \in \mathbb{R}^n$

(2) $f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^n F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n}$

(3) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(y_1, \cdots, y_n) \ dy_1 \cdots dy_n = 1.$

A partir das definições anteriores (distribuição e densidade conjuntas), podemos encontrar expressões para a distribuição e densidade conjuntas de um subconjunto de coordenadas de $\boldsymbol{X}$. Tal redução é conhecida como **marginalização**. Por exemplo, a distribuição conjunta de $\boldsymbol{X}_{|k}$ é

$$F(\boldsymbol{x}_{|k}) = F(x_1, \cdots, x_{k-1}, \infty, \cdots, x_n),$$

donde

$$f(\boldsymbol{x}_{|k}) = \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_k.$$

Em particular, a distribuição marginal de X_i é:

$$F(x_i) = F(\infty, \cdots, x_i, \infty, \cdots, \infty),$$

onde

$$f(x_i) = \int_{\mathbb{R}^{n-1}} f(\boldsymbol{x}) d\boldsymbol{x}_{|i}$$

denota a densidade marginal de X_i.

Agora, consideremos um conjunto de variáveis aleatórias X_i, onde I é um conjunto de índices contável, definidas em um espaço de probabilidade $(\Theta, \mathcal{F}, P)$. Dizemos que estas variáveis aleatórias são independentes se e somente se

$$P(X_i \leq x_i, i \in J) = \prod_{i \in J} P(X_i \leq x_i), \ \forall J \subset I, x_i \in \mathbb{R}.$$

Para um conjunto de índices $I = \{1, \cdots, n\}$, podemos interpretar a coleção de variáveis aleatórias como um vetor aleatório $\boldsymbol{X}$ em $\mathbb{R}^n$. Além disso, como as variáveis aleatórias X_i são independentes, segue que as funções de distribuição conjunta $F_{\boldsymbol{X}}(x_1, \cdots, x_n)$ e de densidade de probabilidade $f_{\boldsymbol{X}}(x_1, \cdots, x_n)$ podem ser fatoradas, respectivamente, por:

$$F_{\boldsymbol{X}}(x_1, \cdots, x_n) = \prod_{i=1}^{n} F_X(x_i),$$

$$f_{\boldsymbol{X}}(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_X(x_i).$$

2.3 Operador Integração

Precisamos estabelecer as noções básicas do operador integração para poder definir na Seção 2.4 o espaço das funções quadraticamente integráveis. Para isso, considere $X : (\Theta, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$. A integral de X com respeito a P sobre um evento $A \in \mathcal{F}$ é:

$$\int_{\Theta} I_A(\theta) X(\theta) dP(\theta) = \int_A X(\theta) dP(\theta), \tag{2-8}$$

onde I_A representa a função indicadora de A, isto é:

$$I_A = \begin{cases} 1 & \text{se } \theta \in A \\ 0 & \text{se } \theta \notin A \end{cases}$$

Se a integral da eq. (2-8) existe e é finita, então X é chamada de P-**integrável sobre A**. Consequentemente, um vetor aleatório $\boldsymbol{X}$ é P-integrável sobre A, se cada um de seus componentes é P-integrável sobre A.

Proposição 2.3 *Agora considere X, Y duas variáveis aleatórias em $(\Theta, \mathcal{F}, P)$. As afirmações abaixo são verdadeiras:*

(1) Para $A \in \mathcal{F}$:

$$\int_A |X| \, dP < \infty \Rightarrow \int_A X dP \text{ é finito.}$$

(2) Se X, Y são P-integráveis sobre $A \in \mathcal{F}$, a variável aleatória $aX + bY$ é P-integrável sobre A e:

$$\int_A (aX + bY) dP = a \int_A X dP + b \int_A Y dP.$$

(3) Se X é P-integrável em Θ e $\{A_i \in \Theta\}$ é uma partição de Θ, então:

$$\int_{\cup_i A_i} XdP = \sum_i \int_{A_i} XdP.$$

(4) Se X é P-integrável em Θ, então X é finito a.s. (almost surely) [2].

(5) Se $X \geq 0$ a.s., então

$$\int_A XdP \geq 0.$$

(6) Se $Y \leq X$ a.s., então

$$\int_A YdP \leq \int_A YdP.$$

Demonstração: Ver, e.g., Durret, 2010, p. 19. □

Em particular, se $A = \Theta$ para o domínio de integração temos que o operador esperança, notação $\mu_X = E[\cdot]$, é definido pela expressão abaixo:

$$E[X] = \int_\Theta X(\theta)dP(\theta). \tag{2-9}$$

Proposição 2.4 *Sejam X e Y duas variáveis aleatórias de valor real definidas no espaço de probabilidade $(\Theta, \mathcal{F}, P)$. Se X e Y são P-integráveis sobre Θ, então:*

(1) $E[aX + bY] = aE[X] + bE[Y]$ (linearidade da esperança)

(2) Se $X \geq 0$ a.s., então $E[X] \geq 0$

(3) Se $X \leq Y$ a.s., então $E[X] \leq E[Y]$

(4) $|E[X]| \leq E[|X|]$.

Demonstração: Ver, e.g., Durret, 2010, p.27. □

Agora, considere X uma variável aleatória definida no espaço de probabilidade $(\Theta, \mathcal{F}, P)$, e seja $g : (\mathbb{R}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ uma função mensurável. Daí, $Y = g \circ X$ [3], é uma variável aleatória definida em $(\Theta, \mathcal{F}, P)$ e o valor esperado de Y é dado por:

$$\begin{aligned} E[Y] &= \int_\Theta Y(\theta)dP(\theta) = \int_\Theta g(X(\theta))dP(\theta) \\ &= \int_\mathbb{R} g(x)dQ(x) = \int_\mathbb{R} g(x)dF_X(x) = \int_\mathbb{R} g(x)f_X(x)dx. \end{aligned} \tag{2-10}$$

[2]Almost surely significa que a probabilidade de esse evento aconteçer é igual a 1.

[3]$\circ$ representa composição.

A eq. (2-10) nos da uma expressão da esperança de Y que envolve a função de densidade de probabilidade de X. Nesta dissertação, faremos uso de tal expressão do operador esperança em termos de funções de densidade.

Por outro lado, a variância da variável aleatória X pode ser definida por meio de o valor esperado, isto é:

$$\sigma_X^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx. \qquad (2\text{-}11)$$

Em geral, X é uma variável aleatória de valor real definida em um espaço de probabilidade $(\Theta, \mathcal{F}, P)$, e $Y = g(X) = X^r$ para $r \geq 1$. Se a função $g(x)$ é contínua, então é Borel-mensurável e, assim, $Y = X^r$ é uma variável aleatória. O valor esperado de Y é chamado de r-ésimo momento da variável aleatória X, notação $m_r(X)$, definido como:

$$m_r(X) = E[X^r] = \int_{-\infty}^{\infty} x^r f_X(x) dx.$$

Similarmente, podemos definir os momentos para os vetores aleatórios. Por exemplo, o valor esperado de um vetor aleatório $\boldsymbol{X}$ é dado por:

$$\mu_{\boldsymbol{X}} = E[\boldsymbol{X}] = (E[X_1], \cdots, E[X_n]).$$

E a matriz de covariância de $\boldsymbol{X}$ é definida como:

$$\boldsymbol{C}_{\boldsymbol{X}} = (cov(X_i, X_j))_{i,j=1}^n,$$

onde $cov(X_i, X_j) = E[(X_i - \mu_{X_i})(X_j - \mu_{X_j})] = E[X_i X_j] - \mu_{X_i}\mu_{X_j}$, é a covariância de X_i e X_j. Note que $cov(X_i, X_i) = \sigma_{X_i}^2$.

2.4
Espaço L^2

Nesta seção vamos investigar algumas propriedades do espaço L^2 que nos serão úteis ao longo desta dissertação. Primeiro, consideremos o espaço de probabilidade $(\Theta, \mathcal{F}, P)$. Para $q \geq 1$, denotamos $L^q(\Theta, \mathcal{F}, P)$, a coleção de variáveis aleatórias de valor real X definidas em $(\Theta, \mathcal{F}, P)$ tal que:

$$E\left[|X|^q\right] \leq \infty.$$

Em particular, estamos interessados no caso quando $q = 2$. Neste caso, $L^q = L^2$ é o espaço das funções quadraticamente integráveis. O espaço L^2 representa o conjunto de todas as variáveis aleatórias X tal que o segundo momento é finito, ou seja,

$$E[|X|^2] = \int_\Theta |X|^2 dP < \infty.$$

As seguintes proposições estabelecem algumas propriedades do espaço L^2.

Proposição 2.5 *O espaço L^2 é um espaço linear.*

Demonstração:

Para provar que L^2 é um espaço linear precisamos verificar que o espaço satisfaz as propriedades de soma vetorial e multiplicação por escalar, segundo a Definição 9.1 do Apêndice.

1. Multiplicação escalar:
   Seja $X \in L^2$ e $\alpha \in \mathbb{R}$. Note que $E[(\alpha X)^2] = \alpha^2 E[X^2] \leq \infty$. Logo, $\alpha X \in L^2$.

2. Soma vetorial:
   Sejam X, $Y \in L^2$. Como:

   $$E[(X+Y)^2] = E[(X)^2] + E[(Y)^2] + 2E[(XY)^2] \leq \infty,$$

   temos que $X + Y \in L^2$.
   Note que para provar a última desigualdade, temos que verificar que $E[(XY)^2]$ é finito. Considere $X \neq Y$, então o polinômio:

   $$p(\alpha) = E[(X + \alpha Y)^2] = E[(X)^2] + 2\alpha E[\alpha XY] + \alpha^2 E[(Y)^2],$$

   não têm raiz real. Daí, $E[(XY)]^2 - E[(X)^2] \cdot E[(Y)^2] \leq 0$ ou seja,

   $$\left|E[(XY)^2]\right| \leq E[X^2]^{\frac{1}{2}}[Y^2]^{\frac{1}{2}}$$

   para cada $X, Y \in L^2$. Logo, $E[(XY)^2] < \infty$.

$\square$

Proposição 2.6 *O valor esperado $E[XY]$ define um produto interno em L^2.*

Demonstração: Consideremos o produto interno $\langle X; Y \rangle$ em L^2 como $E[XY]$. Vamos a provar que $E[XY]$ satisfaz as condições da Definição 9.2 (do Apêndice):

(a) $E[\alpha XY] = \alpha E[XY]$, pois a esperança é homogênea, onde $X, Y \in L^2, \alpha \in \mathbb{R}$.

(b) $E[(\alpha X + Y)Z] = \alpha E[XZ] + E[YZ]$, pois a esperança é linear, onde $X, Y, Z \in L^2$.

(c) $E[\alpha XY] = \alpha E[YX],\ \ X, Y \in L^2$ (Simetria).

(d) $E[\alpha X^2] \geq 0$, dado que $X^2 > 0$ a.s., com $X \in L^2$ (Não-negativa).

(e) $E[\alpha X^2] = 0$ solo se $X = 0$ a.s., com $X \in L^2$ (Positividade).

Portanto, $E[XY]$ define um produto interno em L^2. $\square$

Daí, podemos denotar a norma associada a L^2 como $\|X\| = E[X^2]^{1/2}$.

Agora, considere $d : L^2 \times L^2 \to [0, \infty)$ definida por $d\,(X, Y) = \|X - Y\|$. Então, podemos ver que d é a métrica induzida pelo produto interno em L^2 :

(a) $d(X, Y) = \|X - Y\| = E[(X - Y)^2]^{1/2} = 0 \Leftrightarrow X = Y$ a.s.

(b) $d(X, Y) = \|X - Y\| = E[(X - Y)^2]^{1/2} = E[(Y - X)^2]^{1/2} = \|Y - X\| = d(Y, X),\ \ X, Y \in L^2$.

(c) $d(X, Z) = \|X - Z\| = E[(X - Z)^2]^{1/2} = E[(X + Y + Y - Z)^2]^{1/2} \leq E[(X - Y)^2]^{1/2} + E[(Y - Z)^2]^{1/2} = \|X - Y\| + \|Y - Z\| = d(X, Y) + d(Y, Z),\ \ \forall X, Y, Z \in L^2$.

Considere X uma variável aleatória de valor real e $\{X\}_{n\geq 1}$ uma sequência de variáveis aleatórias definidas num espaço de probabilidade $(\Theta, \mathcal{F}, P)$. A convergência da sequência $\{X_n\}$ depende da medida. Neste trabalho adotaremos a convergência em L^p, i.e.

$$X_n \to X \text{ se } \lim_{n\to\infty} E[|X_n - X|^p] = 0.$$

Particularmente, será usada a convergência em média quadrática, i.e. convergência em L^2.

Com todos os resultados obtidos nesta seção podemos fazer algumas observações. Temos L^2 é um espaço com álgebra e topologia juntas. Isto porque da Proposição 2.5 o espaço L^2 é uma estrutura algébrica que nos permite efetuar operações (soma vetorial e multiplicação escalar).

Além disso, temos uma estrutura topológica dada pela norma $\|\cdot\|$ que nos dá a noção de "distância" por meio da métrica d. Daí, quando equipamos esse espaço linear com um produto interno (como foi visto na Proposição 2.6) podemos obter o conceito de ortogonalidade, que terá um importante papel no nosso estudo para propagar incertezas, e que será visto com maior detalhe no próximo capítulo.

Portanto, L^2 equipado com o produto interno $\langle \cdot ; \cdot \rangle$ é um espaço de Hilbert (Kubrusly, 2010), isto é, um espaço produto interno completo.