Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Matemática

Aspectos Matemáticos do Filtro de Kalman Discreto

por

Dimas José Gonçalves *

Mestrado em Matemática - Campinas SP

Orientador: Prof. Dr. Raymundo Luiz de Alencar

*Este trabalho contou com apoio financeiro do CNPq.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecário: Maria Júlia Milani Rodrigues – CRB8a / 2116

Gonçalves, Dimas José

G586a Aspectos matemáticos do Filtro de Kalman Discreto / Dimas José Gonçalves -- Campinas, [S.P. :s.n.], 2005.

Orientador : Raymundo Luiz de Alencar

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Kalman, Filtragem de. 2. Teoria da estimativa. 3. Teoria do controle. 4. Hilbert, Espaços de. I. Alencar, Raymundo Luiz de. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Mathematical aspects of the Discrete Kalman's Filter

Palavras-chave em inglês (Keywords): 1. Kalman filtering. 2. Estimation theory. 3. Control theory. 4. Hilbert spaces

Área de concentração: Análise Funcional

Titulação: Mestre em Matemática

Banca examinadora: Prof. Dr. Raymundo Luiz de Alencar (UNICAMP)
Prof. Dr. Sérgio Antonio Tozoni (UNICAMP)
Prof. Dr. Elói Medina Galego (USP)

Data da defesa: 21/02/2005

Aspectos Matemáticos do Filtro de Kalman Discreto

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Dimas José Gonçalves** e aprovada pela comissão julgadora.

Campinas, 21 de fevereiro de 2005.

Raymdo L. de Alencar

Prof. Dr. Raymundo Luiz de Alencar

Orientador

Banca examinadora:

Prof. Dr. Raymundo Luiz Alencar.

Prof. Dr. Sérgio Antonio Tozoni.

Prof. Dr. Elói Medina Galego.

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP como requisito parcial para obtenção do título de **Mestre em Matemática**.

Dissertação de Mestrado defendida em 21 de Fevereiro de 2005 e aprovada pela Banca Examinadora composta pelos Profs. Drs.

Prof (a). Dr (a). RAYMUNDO LUIZ ALENCAR

Prof (a). Dr (a). SÉRGIO ANTONIO TOZONI

Elói Medina Galego

Prof (a). Dr (a). ELÓI MEDINA GALEGO

Agradecimentos

Primeiramente, gostaria de agradecer a pós graduação do IMECC pela estrutura educacional que me foi oferecida e agradecer os professores Sérgio Antonio Tozoni, Elói Medina Galego e Mário Infante pelas correções sugeridas na minha dissertação de mestrado.

Dentre vários amigos que tenho, agradeço Bibiana e Fábio pelas ajudas em vários momentos de que precisei. São grandes amigos que convivem comigo desde a faculdade.

Apenas esse agradecimento seria pouco para demonstrar a gratidão que sinto pelo meu orientador Raymundo e pela sua esposa Cléo. Obrigado por me receberem em sua residência com carinho, pela amizade e pela ajuda nos momentos difíceis.

Na vida nós conhecemos vários tipos de pessoas. Essa família que eu conheci é composta por pessoas especiais. Coloco a família Mustafa nesses agradecimentos de forma alegre, pois é com alegria que eu gosto de lembrar deles. O meu muito obrigado pelo o que vocês foram, são e serão para mim.

Por mais palavras que eu procure, pois mais gestos que eu tente mostrar, por mais... Nada disso vai conseguir demonstrar o quão importante essa pessoa é para mim. Talvez uma frase simples expresse o que essa pessoa significa: eu te amo Tatiana. Muito obrigado pelo amor, pela força, compreensão e alegria que você passa.

O que eu tenho para dizer da minha família , com certeza não caberia em um livro, mas vou tentar caracterizar o que cada um significa: meu pai, símbolo de força e garra; minha mãe, símbolo de amor e carinho; meus írmãos, símbolo de amizade e união. Se hoje tenho saúde, educação, alegria, amor e crença em Deus, é porque vocês me proporcionaram tudo isso. Muito obrigado pelo o que eu sou.

Por fim, dedico essa tese de mestrado ao meu írmão Josmar, que representa para mim a alegria e a "vontade de viver ".

Resumo

Neste trabalho, agrupamos um mínimo de conteúdo matemático para desenvolver uma prova do Teorema do Filtro de Kalman (discreto).

A teoria de integração de Lebesgue e alguns elementos (Teorema da Projeção) da teoria dos espaços de Hilbert são as principais ferramentas matemáticas incluídas, além de um estudo das principais propriedades da esperança condicional.

Abstract

In this work we set a minimum of mathematical background to develop a proof of the (discrete) Kalman Filter Theorem.

The Lebesgue integration theory and some elements (the projection theorem) of the Hilbert space theory are the main mathematical tools included, besides a study of the main properties of the conditional expectation.

Sumário

Plano Geral da Dissertação

Estudar ferramentas e conceitos matemáticos para estabelecer uma forma simplificada do Teorema de Kalman Discreto (solução do problema de estimação recursiva para um sistema discreto).

A ferramenta básica utilizada na demonstração do Teorema de Kalman é o conceito de esperança condicional para estabelecer estimativas de mínima variância. Com este propósito foram estudadas propriedades gerais dos espaços de Hilbert, destacando em particular o Teorema da Projeção, que é essencial para estabelecer certos tipos de aproximações (estimativa pelos mínimos quadrados). Na teoria de integração de Lebesgue, destacamos os principais teoremas de convergência, bem como o teorema de Radon-Nikodým, além do espaço de Hilbert das funções de quadrado integrável.

Finalmente, foi dada uma atenção especial ao estudo da esperança condicional de uma variável aleatória com relação a um vetor aleatório, onde as principais propriedades da esperança condicional foram estudadas.

Introdução

Entre os temas considerados para esta dissertação, optamos por *Filtro de Kalman*, um tema muito estudado pelas seguintes razões: a grande importância do filtro de Kalman por suas aplicações em Engenharia; a necessidade do estudo de ferramentas matemáticas sofisticadas para a compreensão teórica do filtro de Kalman e a possibilidade de envolvimento com conceitos e ferramentas matemáticas que não são, em geral, desenvolvidas em disciplinas de um Mestrado em Matemática.

No Capítulo 1, incluímos pré-requisitos de caráter mais geral, como espaços de Hilbert (Teorema da Projeção), equações normais (matriz de Gram) e elementos da Teoria da Medida e Integração.

No Capítulo 2, desenvolvemos pré-requisitos mais específicos de Probabilidade e Estatística, como por exemplo: variáveis e vetores aleatórios, esperança matemática, variância, covariância e , por fim, esperança condicional de um vetor aleatório baseado em outro vetor aleatório.

No Capítulo 3, utilizando principalmente a esperança condicional, estabelecemos as estimativas de mínima variância, ferramenta básica para a demonstração do Teorema de Kalman, também incluído no terceiro e último capítulo.

Salientamos que muitos dos tópicos abordados na dissertação poderiam ser de maneira natural, ampliados e/ou desenvolvidos com maior profundidade; no entanto, procuramos nos limitar ao mínimo necessário, tendo em vista nosso objetivo.

Achamos que o teorema de Kalman (filtragem) é uma belíssima aplicação da teoria de integração de Lebesgue e da teoria dos espaços de Hilbert na solução de um problema importante da matemática aplicada.

Finalizando, gostaríamos de mencionar que algumas idéias aqui desenvolvidas surgiram

quando o orientador desta dissertação, ministrou a disciplina *Princípios de Análise Funcional Aplicada - FM210* para o programa de mestrado do Instituto Tecnológico de Aeronáutica (ITA). Nossos agradecimentos à pós-graduação do ITA pela aprovação da disciplina e aos alunos Alex, Antônio, Cleverson, Cleyton e Mônica pelo interesse demonstrado.

Capítulo 1

Preliminares

O objetivo desse capítulo é passar para o leitor noções básicas de Espaços de Hilbert, Medida e Integração que serão utilizadas ao longo do trabalho. Na seção 1.1 e 1.2, chamamos a atenção do leitor para o Teorema da Projeção e para as equações normais, que serão resultados essenciais para o desenvolvimento do Capítulo 3. Em relação a seção 1.3, destacamos o Teorema de Radon-Nikodyn, que será utilizado na demonstração da existência da esperança condicional (página 32), e destacamos também o Teorema de Fubini, que será útil para deduzir alguns resultados do Capítulo 2.

1.1 Espaços de Hilbert

Definição 1.1. *Seja V um K-espaço vetorial, onde K=$\mathbb{R}$ ou K=$\mathbb{C}$. Uma norma sobre V é uma aplicação $\| \ \| : V \to \mathbb{R}$ que satisfaz as propriedades:*

a) $\| v \| \geq 0, \forall v \in V$.

b) $\| v \| = 0 \iff v = 0$.

c) $\| \lambda . v \| = |\lambda| . \| v \|, \forall \lambda \in K, \forall v \in V$.

d) $\| u + v \| \leq \| u \| + \| v \|, \forall u, v \in V$.

Nesse caso, dizemos que o par $(V, \| \ \|)$ é um espaço normado.

Definição 1.2. *Seja V um K-espaço vetorial, onde K=$\mathbb{R}$ ou K=$\mathbb{C}$. Um produto interno sobre V é uma função $< \ , \ >: V \times V \to K$ que satisfaz as seguintes propriedades:*

a) $< u+v, w > = < u, w > + < v, w >, \forall u, v, w \in V.$

b) $< \lambda.u, v > = \lambda. < u, v >, \ \forall \lambda \in K, \ \forall u, v \in V.$

c) $< u, v > = \overline{< v, u >}, \forall u, v \in V.$

d) $< u, u > > \ 0, \ se \ u \neq 0.$

Um espaço vetorial V munido de um produto interno $<,>$ é denominado um **espaço pré-hilbertiano** e indicado por $(V, <,>)$. A função que associa cada $v \in V$ ao número $\sqrt{< v, v >}$ é uma norma sobre V, chamada norma induzida pelo produto interno $<,>$.

Propriedades: Seja V um espaço vetorial com produto interno $< \, , >$.

a) Regra do paralelogramo: $\| u+v \|^2 + \| u-v \|^2 = 2(\| u \|^2 + \| v \|^2), \forall u, v \in V.$

b) Desigualdade de Cauchy-Schwarz: $| < u, v > | \leq \| u \| . \| v \|, \forall u, v \in V.$

c) Desigualdade triangular: $\| u+v \| \leq \| u \| + \| v \|, \forall u, v \in V.$

Definição 1.3. *Seja $(x_n)_{n\in\mathbb{N}}$ uma sequência em um espaço normado V. Dizemos que $(x_n)_{n\in\mathbb{N}}$ converge em V, se existe $v \in V$ que satisfaz:*

$$\forall \epsilon > 0, \exists N(\epsilon) \in \mathbb{N} \ tal \ que \ n \geq N(\epsilon) \Rightarrow \| x_n - v \| < \epsilon \ .$$

Nesse caso, dizemos que $(x_n)_{n\in\mathbb{N}}$ converge para v e denotamos $x_n \rightarrow v$ ou $\lim_{n\to\infty} x_n = v$.

Continuidade do produto interno: Sejam $(x_n)_{n\in\mathbb{N}}$ e $(y_n)_{n\in\mathbb{N}}$ sequências em um espaço pré-hilbertiano. Se $x_n \rightarrow x$ e $y_n \rightarrow y$, então $< x_n, y_n > \rightarrow < x, y >$.

Definição 1.4. *Dizemos que dois vetores x, y de um espaço pré-hilbertiano V são ortogonais, se $< x, y > = 0$. Nesse caso, escrevemos $x \perp y$. Sejam $x \in V$ e $A \subseteq V$. Dizemos que x é ortogonal a A, se $x \perp a$, $\forall a \in A$. Nesse caso, escrevemos $x \perp A$. Denotamos por $A^\perp$ o conjunto*

$$A^\perp = \{x \in V : x \perp A\}.$$

Observação: Decorre da continuidade do produto interno o seguinte fato: seja $(x_n)_{n\in\mathbb{N}}$ uma sequência em um espaço pré-hilbertiano V e $y \in V$. Se $x_n \perp y$, $\forall n \in \mathbb{N}$ e $x_n \rightarrow x$, então $x \perp y$. Assim, se A é um subconjunto de um espaço pré-hilbertiano, então $A^\perp$ é um subespaço vetorial fechado em V.

Teorema de Pitágoras: Sejam x, y vetores de um espaço pré-hilbertiano V. Se $x \perp y$, então

$$\| x+y \|^2 = \| x \|^2 + \| y \|^2 \ .$$

Observe que a recíproca do teorema de Pitágoras é verdadeira se V é espaço vetorial sobre o corpo $\mathbb{R}$.

Exemplo: Defina em $\mathbb{C}$ o produto interno canônico, ou seja,

$$< x, y >= x.\bar{y} \ , \ \forall x, y \in \mathbb{C} \ .$$

Se $x = 1$ e $y = i$, então x não é ortogonal a y, mas

$$\| x + y \|^2 = \| x \|^2 + \| y \|^2 \ \ .$$

Portanto, sobre o corpo $\mathbb{C}$ a recíproca do teorema de Pitágoras não é verdadeira.

Definição 1.5. *Seja $(x_n)_{n\in\mathbb{N}}$ uma sequência em um espaço normado. Dizemos que $(x_n)_{n\in\mathbb{N}}$ é sequência de Cauchy, se satisfaz:*

$$\forall \epsilon > 0, \exists N(\epsilon) \in \mathbb{N} \ tal \ que \ m, n \geq N(\epsilon) \Rightarrow \| x_n - x_m \| < \epsilon.$$

Definição 1.6. *Seja A um subconjunto de um espaço normado V. Se toda sequência de Cauchy de elementos de A convergir para um elemento de A, então dizemos que A é completo. Se V é um espaço pré-hilbertiano completo, dizemos que V é um espaço de* ***Hilbert****.*

Propriedades: Seja A um subconjunto de um espaço normado V.

a) Se V é um espaço completo, então A é completo se e somente se A é fechado.

b) Se V tem dimensão finita, então V é completo.

A seguir enunciamos o teorema da projeção, que será um resultado fundamental no desenvolvimento desta dissertação.

Teorema da Projeção: Seja W um subespaço vetorial fechado de um espaço de Hilbert V. Se $v \in V$, então existe um único $w \in W$ que satisfaz as condições equivalentes:

1)$\| v - w \| \leq \| v - u \| \ , \ \forall u \in W \ .$
2)$(v - w) \perp W \ .$

Como consequência imediata do teorema da projeção, temos as propriedades:

Propriedades: Seja H um espaço de Hilbert e $W \subseteq H$ um subespaço fechado.

a) Para cada $x \in H$, existe um único $x_0 \in (x + W)$ de norma mínima. Além disso, $x_0 \in W^{\perp}$.

b) $V = W \oplus W^{\perp}$ e $W = (W^{\perp})^{\perp}$.

Definição 1.7. *Sejam W um subespaço de um espaço de Hilbert V e $v \in V$. Se existir $w \in W$ tal que $(v - w) \perp W$, dizemos que w é a projeção ortogonal de v sobre W. Denotaremos essa projeção por $proj_W(v) = w$.*

Proposição 1.8. *Sejam U e V subespaços fechados de um espaço de Hilbert H, tais que $U + V$ é fechado. Se W é o subespaço que satisfaz $W \perp U$ e $W \oplus U = U + V$, então*

$$proj_{U+V}(h) = proj_U(h) + proj_W(h) \ , \ \forall h \in H \ .$$

A demonstrações dos resultados apresentados nessa seção, encontram-se na referência [5].

1.2 Equações Normais

A seguir, introduzimos a matriz de Gram e algumas de suas propriedades que serão utilizadas no texto. Para a facilidade do leitor, provaremos algumas das propriedades.

Definição 1.9. *Sejam $v_1, ..., v_n$ elementos de um espaço pré-hilbertiano. A matriz de ordem n*

$$(< v_i, v_j >)_{i,j} = \begin{bmatrix} < v_1, v_1 > & < v_1, v_2 > & \cdots & < v_1, v_n > \\ < v_2, v_1 > & < v_2, v_2 > & \cdots & < v_2, v_n > \\ \vdots & \vdots & \ddots & \vdots \\ < v_n, v_1 > & < v_n, v_2 > & \cdots & < v_n, v_n > \end{bmatrix}$$

é chamada matriz de Gram de $v_1, ..., v_n$. Denotamos essa matriz por $G(v_1, ..., v_n)$.

Proposição 1.10. *Sejam $v_1, ..., v_n$ elementos de um espaço pré-hilbertiano. Então $det\ G(v_1, ..., v_n) = 0$ se, e somente se, $\{v_1, ..., v_n\}$ é um conjunto linearmente dependente.*

Prova: Se $det\ G(v_1, ..., v_n) = 0$, então as linhas da matriz de Gram formam um conjunto linearmente dependente. Assim, existem escalares não todos nulos $\alpha_1, ..., \alpha_n$ tais que

$$\sum_{i=1}^{n} \alpha_i < v_i, v_j >= 0 \ , \ \forall j \in \{1, ..., n\}$$

Disso, segue que

$$\sum_{j=1}^{n} \bar{\alpha}_j . \left(\sum_{i=1}^{n} \alpha_i < v_i, v_j > \right) = 0 \ .$$

Ou seja, $\| \sum_{i=1}^{n} \alpha_i . v_i \|^2 = 0$. Assim, $\sum_{i=1}^{n} \alpha_i . v_i = 0$ e o conjunto $\{v_1, ..., v_n\}$ é linearmente dependente.

Agora, assuma que $\{v_1, ..., v_n\}$ é linearmente dependente. Podemos supor sem perda de generalidade que v_1 é combinação linear de $v_2, ..., v_n$.Assim, temos que a primeira linha da matriz $G(v_1, ..., v_n)$ é combinação linear das linhas restantes. Logo, $det\ G(v_1, ..., v_n) = 0$. ■

Proposição 1.11. *Sejam $v_1, ..., v_n$ elementos de um espaço de Hilbert H. Denote por V o subespaço gerado por $v_1, ..., v_n$. Se $h \in H$, então $proj_V(h) = \sum_{i=1}^{n} \alpha_i . v_i$, onde $\alpha_1, ..., \alpha_n$ são escalares que satisfazem as equações*

$$G(v_1, ..., v_n)^t . \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} < h, v_1 > \\ < h, v_2 > \\ \vdots \\ < h, v_n > \end{bmatrix} . \qquad (*)$$

Prova: Sejam $\alpha_1, ..., \alpha_n$ escalares tais que

$$proj_V(h) = \sum_{i=1}^{n} \alpha_i . v_i \ .$$

Do Teorema da Projeção seguem as equivalências:

$$proj_V(h) = \sum_{i=1}^{n} \alpha_i . v_i \Longleftrightarrow \left(h - \sum_{i=1}^{n} \alpha_i . v_i \right) \perp V \Longleftrightarrow \left(h - \sum_{i=1}^{n} \alpha_i . v_i \right) \perp v_j \ , \ \forall j \in \{1, ..., n\} \Longleftrightarrow$$

$$\Longleftrightarrow \sum_{i=1}^{n} \alpha_i . < v_i, v_j > = < h, v_j > \ , \ \forall j \in \{1, ..., n\} \Longleftrightarrow$$

$$\Longleftrightarrow G(v_1, ..., v_n)^t . \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} < h, v_1 > \\ < h, v_2 > \\ \vdots \\ < h, v_n > \end{bmatrix} . \qquad ■$$

Observação : As equações em $(*)$ são denominadas **"equações normais "**.

Teorema 1.12. *Seja* $\{v_1, ..., v_n\}$ *um conjunto linearmente independente em um espaço de Hilbert* H. *Fixados escalares* $a_1, ..., a_n$, *o conjunto*

$$T = \{x \in H \ :< x, v_i >= a_i \ , \ \forall i \in \{1, ..., n\}\}$$

possui um único elemento v *de norma mínima. Além disso,* $v = \sum_{i=1}^{n} \alpha_i . v_i$, *onde*

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = (G^t)^{-1} . \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

e $G = G(v_1, ..., v_n)$ *é a matriz de Gram.*

Prova: Denote por V o subespaço gerado por $v_1, ..., v_n$. Primeiro demonstraremos que existe um único $v \in V$ pertencente a T. Pela Proposição 1.10, temos que G é invertível. Assim, existem únicos escalares $\alpha_1, ..., \alpha_n$ tais que

$$G^t . \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} .$$

Logo, $v = \sum_{i=1}^{n} \alpha_i . v_i$ é o único elemento pertencente a $V \cap T$. Provemos que $v + V^{\perp} = T$. De fato, se $u \in V^{\perp}$, então

$$< v + u, v_i >=< v, v_i >= a_i \ , \ \forall i \in \{1, ..., n\} \ .$$

Logo, $v + V^{\perp} \subseteq T$. Reciprocamente, seja $x \in T$. Como H é espaço de Hilbert, temos que $H = V \oplus V^{\perp}$ e portanto existem $w \in V$ e $u \in V^{\perp}$ tais que $w + u = x$. Da expressão

$$a_i =< x, v_i >=< w, v_i > \ , \ \forall i \in \{1, ..., n\}$$

segue que $w \in T \cap V$. Logo, $w = v$ e $T \subseteq (v + V^{\perp})$.

Uma vez que $v + V^{\perp} = T$, temos por propriedades (ítem (a) da página 8) que o elemento de norma mínima de T pertence a $(V^{\perp})^{\perp}$. Como $(V^{\perp})^{\perp} = V$, o elemento de norma mínima de T é v. $\blacksquare$

1.3 Noções de Medida e Integração

Êste parágrafo, com noções da teoria de Integração se faz necessário não somente porque os principais conceitos que serão objetos de estudo (esperança, esperança condicional, variância e outros) dependem da integral de Lebesgue mas também para estabelecer a notação utilizada.

Definição 1.13. *Uma família Λ de subconjuntos de um conjunto Ω é uma σ-álgebra, se satisfaz:*
a) $\emptyset$, $\Omega \in \Lambda$.
b) *Se $B \in \Lambda$, então o complementar $B^c = \Omega - B$ pertence a Λ.*
c) *Se $(B_n)_{n\in\mathbb{N}}$ é uma sequência em Λ, então* $\left(\bigcup_{n=1}^{\infty} B_n\right) \in \Lambda$.
Nesse caso, se $B \in \Lambda$, então dizemos que B é Λ-mensurável.

Seja $\Lambda \neq \emptyset$ uma família de subconjuntos de Ω e A um subconjunto qualquer de Ω. Observe que a interseção de todas as σ-álgebras que contém A é a menor σ-álgebra de subconjuntos de Ω que contém A. Essa menor σ-álgebra é chamada σ-álgebra gerada por A.

Diante desse fato, seja Ω o conjunto dos números reais $\mathbb{R}$. A "álgebra de Borel " é a σ-álgebra $\beta(\mathbb{R})$ gerada por todos intervalos abertos (a, b) de $\mathbb{R}$. Nesse caso, se $B \in \beta(\mathbb{R})$, dizemos que B é um conjunto de Borel. Em geral, temos a seguinte definição:

Definição 1.14. *A menor σ-álgebra de subconjuntos de $\mathbb{R}^n$ que contém todos os conjuntos abertos é chamada σ-álgebra de Borel. Ela será denotada por $\beta(\mathbb{R}^n)$ e qualquer conjunto em $\beta(\mathbb{R}^n)$ será chamado conjunto de Borel.*

Definição 1.15. *Seja Λ uma σ-álgebra de subconjuntos de Ω. Uma função $\mu : \Lambda \to \overline{\mathbb{R}}$,onde $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, é uma medida se satisfaz:*
a)$\mu(\emptyset) = 0$.

b) $\mu(B) \geq 0$, $\forall B \in \Lambda$.

c) *μ é aditiva contável: se $(B_n)_{n\in\mathbb{N}}$ é uma sequência disjunta ($B_n \cap B_m = \emptyset$ se $n \neq m$), então*

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \ .$$

Nesse caso, dizemos que (Ω, Λ, μ) *é um espaço de medida.*

Dizemos ainda, μ *é uma medida* σ*-finita, se existe uma sequência* $(B_n)_{n\in\mathbb{N}}$ *em* Λ *tal que*

$$\mu(B_n) \neq +\infty \ , \forall n \in \mathbb{N} \ \ e \ \ \Omega = \bigcup_{n=1}^{\infty} B_n \ .$$

Um exemplo de medida que será utilizada no decorrer da dissertação é a medida de Lebesgue $\lambda : \beta(\mathbb{R}) \to \overline{\mathbb{R}}$ que possui a seguinte propriedade:

$$\lambda((a,b)) = b - a$$

para todo intervalo aberto (a,b) de $\mathbb{R}$.

As propriedades abaixo serão utilizadas no texto, suas demonstrações podem ser encontradas na página 21 da referência [1].

Propriedades: Seja (Ω, Λ, μ) um espaço de medida.

1) Se $A, B \in \Lambda$ e $A \subseteq B$, então $\mu(A) \leq \mu(B)$. Além disso, se $\mu(A) \neq +\infty$, então

$$\mu(B - A) = \mu(B) - \mu(A) \ .$$

2) Se $(B_n)_{n\in\mathbb{N}}$ é uma sequência crescente em Λ ($B_n \subseteq B_{n+1}$, $\forall n \in \mathbb{N}$), então

$$\lim_{n\to\infty} \mu(B_n) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) \ .$$

3) Se $(B_n)_{n\in\mathbb{N}}$ é uma sequência decrescente em Λ ($B_n \supseteq B_{n+1}$, $\forall n \in \mathbb{N}$) e $\mu(B_1) \neq +\infty$, então

$$\lim_{n\to\infty} \mu(B_n) = \mu\left(\bigcap_{n=1}^{\infty} B_n\right) \ .$$

Definição 1.16. *Seja* Λ *uma* σ*-álgebra de um conjunto* Ω*. Dizemos que uma função* $X : \Omega \to \mathbb{R}$ *é* Λ*-mensurável (ou simplesmente, mensurável) se*

$$[X \leq x] = \{t \in \Omega : X(t) \leq x\}$$

pertence a Λ *para todo* $x \in \mathbb{R}$*. Se* X *é mensurável, onde* $\Omega = \mathbb{R}$ *e* $\Lambda = \beta(\mathbb{R}^n)$*, então dizemos que* X *é Borel-mensurável.*

Sejam X e Y funções mensuráveis e $c \in \mathbb{R}$. Não é difícil ver que as funções

$$c.X, \ \ X + Y, \ \ X.Y, \ \ |X|$$

são mensuráveis.

Seja $f : \Omega \to \mathbb{R}$ uma função simples, ou seja, $Img(f) = \{f(t) : t \in \Omega\}$ é um conjunto finito. Denote $Img(f) = \{a_1, ..., a_n\}$. Para cada $i \in \{1, ..., n\}$ defina $B_i = f^{-1}(a_i)$. Dizemos que

$$\sum_{i=1}^{n} a_i.\chi_{B_i}$$

é a representação padrão de f.

Lembremos que a função característica do conjunto A , $\chi_A : \Omega \to \mathbb{R}$, é definida por

$$\chi_A(x) = 1 \text{ se } x \in A$$
$$\chi_A(x) = 0 \text{ se } x \notin A$$

Deixamos a cargo do leitor a verificação do seguinte fato: se X é mensurável, então existe uma sequência de funções simples $(X_n)_{n\in\mathbb{N}}$ mensuráveis tais que $|X_n| \leq |X_{n+1}| \leq |X|$, $\forall n \in \mathbb{N}$ e $\lim\limits_{n\to\infty} X_n = X$.

Definição 1.17. *Sejam* (Ω, Λ, μ) *um espaço de medida e* f *uma função mensurável não negativa* $(f(t) \geq 0 \ , \ \forall t \in \Omega)$. *Se* f *é simples, então a integral de* f *com relação a* μ *é o número*

$$\int_\Omega f d\mu = \sum_{i=1}^{n} a_i.\mu(B_i) \ ,$$

onde $f = \sum\limits_{i=1}^{n} a_i.\chi_{B_i}$ *é a representação padrão de* f.

Se X *é uma função mensurável não negativa, definimos a integral de* X *com respeito a* μ *por*

$$\int_\Omega X d\mu = sup\{\int_\Omega f d\mu : 0 \leq f \leq X \ e \ f \in MS\} \ ,$$

onde MS *é o conjunto das funções* Λ*-mensuráveis simples.*

Teorema da convergência monótona: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência crescente de funções $(X_n(t) \leq X_{n+1}(t) \ , \ \forall t \in \Omega \ , \ \forall n \in \mathbb{N}\ )$ Λ-mensuráveis não negativas. Se $(X_n)_{n\in\mathbb{N}}$ converge para X, então

$$\int_\Omega X d\mu = \lim_{n\to\infty} \int_\Omega X_n d\mu \ .$$

A demonstração deste teorema pode ser encontrada em [1] na página 31.

Observação: No decorrer da dissertação, usaremos apenas a convergência "pontual "de funções, isto é, $(X_n)_{n\in\mathbb{N}}$ converge para X se

$$\lim_{n\to\infty} X_n(t) = X(t)\ ,\ \forall t \in \Omega\ .$$

Dada uma função mensurável f não negativa, observe que a igualdade $\int_\Omega f d\mu = +\infty$ pode em certos casos aparecer. Assim, para definir a integral de uma função mensurável X qualquer, é necessário fazer algumas considerações: sejam X^+ e X^- as funções mensuráveis definidas por

$$X^+(t) = \text{máx}\{X(t), 0\}\ ,\ \forall t \in \Omega.$$
$$X^-(t) = \text{máx}\{-X(t), 0\}\ ,\ \forall t \in \Omega.$$

Definição 1.18. *Sejam (Ω, Λ, μ) um espaço de medida e X uma função mensurável. Dizemos que X é integrável se*

$$\int_\Omega X^+ d\mu \neq +\infty \quad e \quad \int_\Omega X^- d\mu \neq +\infty.$$

Nesse caso, definimos a integral de X com respeito a μ por

$$\int_\Omega X d\mu = \left(\int_\Omega X^+ d\mu\right) - \left(\int_\Omega X^- d\mu\right)\ .$$

Propriedades: Sejam X e Y funções integráveis e α um escalar real.

1) $X + Y$ é integrável e

$$\int_\Omega X + Y d\mu = \int_\Omega X d\mu + \int_\Omega Y d\mu.$$

2) α.X é integrável e

$$\int_\Omega \alpha.X d\mu = \alpha. \int_\Omega X d\mu.$$

Essas propriedades decorrem facilmente da definição de integral.

Se X é uma função integrável e $B \in \Lambda$, então $X.\chi_B$ é integrável e assim definimos

$$\int_B X d\mu = \int_\Omega X.\chi_B\ d\mu\ .$$

Fixado um espaço de medida (Ω, Λ, μ), dizemos que uma certa proposição vale μ-quase sempre, se existe um subconjunto $N \in \Lambda$ com $\mu(N) = 0$ tal que a proposição vale sobre o complemento N^c. Por exemplo, dizemos que uma sequência de funções $(X_n)_{n\in\mathbb{N}}$ Λ-mensuráveis converge μ -quase sempre para X, se existe conjunto $N \in \Lambda$ de medida nula tal que

$$X(t) = \lim_{n\to\infty} X_n(t) \ , \ \forall t \in N^c \ .$$

Por comodidade, o termo μ-quase sempre será denotado por μ-q.s.

Teorema da convergência dominada de Lebesgue: Seja $(X_n)_{n\in\mathbb{N}}$ uma sequência de funções mensuráveis que converge μ-q.s para uma função X. Se existe uma função integrável Y tal que $|X_n| \leq Y$ para todo $n \in \mathbb{N}$, então X é integrável e

$$\int_\Omega X d\mu = \lim_{n\to\infty} \int_\Omega X_n d\mu \ .$$

A demonstração deste teorema pode ser encontrada em [1] na página 44.

Definição 1.19. *Seja p um número real tal que $1 \leq p < +\infty$. Dizemos que duas funções Λ-mensuráveis f e g são μ-equivalentes, se $f = g$ $\mu-q.s.$ A classe de equivalência determinada por uma função f é indicada por $[f]$ e consiste do conjunto de todas as funções mensuráveis μ-equivalentes a f. O espaço de todas as classes $[f]$ tal que $|f|^p$ é integrável é denotado por $L_p = L_p(\Omega, \Lambda, \mu)$. Definimos a norma $\| \ \|_p$ sobre L_p por*

$$\|[f]\|_p = \left(\int_\Omega |f|^p d\mu \right)^{\frac{1}{p}} \ .$$

É conveniente tratar os elementos de L_p como funções. Assim, devemos nos referir a classe $[f]$ como sendo o "elemento f de L_p ".

Desigualdade de Hölder: Sejam $f \in L_p$ e $g \in L_q$. Se $\dfrac{1}{p} + \dfrac{1}{q} = 1$, então $f.g \in L_1$ e

$$\| \ f.g \ \|_1 \ \leq \| \ f \ \|_p \ . \ \| \ g \ \|_q \ .$$

A demonstração deste teorema pode ser encontrada em [1] na página 56.

Definição 1.20. *Sejam μ e ν medidas sobre uma σ-álgebra Λ. Se todo $B \in \Lambda$, que satisfaz $\mu(B) = 0$, implicar em $\nu(B) = 0$, então dizemos que ν é absolutamente contínua em relação a μ. Nesse caso denotamos $\nu \ll \mu$.*

Propriedades importantes utilizadas no texto, dependem do teorema:

Teorema de Radon-Nikodym: Sejam μ e ν medidas σ-finitas. Se $\nu \ll \mu$, então existe uma função mensurável f não negativa tal que

$$\nu(B) = \int_B f d\mu \ , \ \forall B \in \Lambda \ .$$

A função f é chamada uma derivada de Radon-Nikodym de ν em relação a μ e é denotada por $\dfrac{d\nu}{d\mu}$. Se f e g são duas derivadas de Radon-Nikodym de ν em relação a μ, então $f = g \ \ \mu - q.s.$

A demonstração deste teorema pode ser encontrada em [1] na página 85.

Teorema 1.21. *(Medida Produto) Sejam $(\Omega_1, \Lambda_1, \mu_1), ..., (\Omega_n, \Lambda_n, \mu_n)$ espaços de medidas σ-finitas. Se Λ é a σ-álgebra gerada por todos conjuntos da forma $A_1 \times ... \times A_n$ com $A_i \in \Lambda_i$, então existe uma única medida μ definida sobre Λ tal que*

$$\mu(A_1 \times ... \times A_n) = \mu_1(A_1).\mu_2(A_2). \ ... \ .\mu_n(A_n) \ , \ \forall A_j \in \Lambda_j \ .$$

Denotamos $\mu = \mu_1 \times ... \times \mu_n$ e dizemos que μ é uma medida produto.

A demonstração deste teorema pode ser encontrada em [3] na página 106.

Uma consequência importante deste teorema é a coincidência da medida de Lebesgue λ^n sobre $\beta(\mathbb{R}^n)$ com o completamento da medida produto $\lambda \times ... \times \lambda$ sobre $\beta(\mathbb{R}^n)$, onde λ é a medida de Lebesgue sobre $\beta(\mathbb{R})$.

Teorema de Fubini : Sejam $(\Omega_1, \Lambda_1, \mu_1)$ e $(\Omega_2, \Lambda_2, \mu_2)$ espaços de medida σ-finitos. Se $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ é uma função integrável com respeito a medida produto $\mu_1 \times \mu_2$, então

$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \times \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x,y) d\mu_2(y) \right) d\mu_1(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f(x,y) d\mu_1(x) \right) d\mu_2(y)$$

Aqui $\int_{\Omega_1} f(x,y) d\mu_1(x)$ é definida para μ_2-quase todo y e $\int_{\Omega_2} f(x,y) d\mu_2(y)$ é definida para μ_1-quase todo x.

A demonstração deste teorema pode ser encontrada em [3] na página 104.

Capítulo 2

Noções de Probabilidade e Estatística

O objetivo deste capítulo é introduzir algumas noções básicas da Teoria de Probabilidade e Estatística que permitam conceituar com precisão *Esperança Condicional*. Dentre os conceitos citados neste capítulo, chamamos a atenção do leitor para o conceito de "Esperança Condicional " (seção 2.4) que será uma ferramenta básica para o Capítulo 3, onde será estudado estimador de mínima variância.

2.1 Probabilidade e Função de Distribuição Acumulada

Seja (Ω, Λ, P) um espaço de medida, neste capítulo, o conjunto Ω será chamado de "**espaço amostral** ". Os elementos da σ- álgebra Λ serão denominados "**eventos aleatórios** " e uma função $X : \Omega \to \mathbb{R}$ Λ-mensurável será chamada de " **variável aleatória** ".

Definição 2.1. *Seja (Ω, Λ, P) um espaço de medida. Se $P(\Omega) = 1$, então dizemos que P é uma* ***probabilidade*** *sobre Λ. Nesse caso, dizemos que (Ω, Λ, P) é um espaço de probabilidade.*

Um exemplo simples de espaço de probabilidade (Ω, Λ, P) pode ser:
$\Omega = (0, 1)$ intervalo aberto de $\mathbb{R}$.
$\Lambda = \beta(0, 1) = \{B \cap (0, 1) : B \in \beta(\mathbb{R})\}$ σ-álgebra de Borel em $(0, 1)$.
$P =$ medida de Lebesgue restrita a Λ .

Definição 2.2. *Sejam* (Ω, Λ, P) *um espaço de probabilidade e* $B \in \Lambda$ *um evento aleatório. Se* $A \in \Lambda$ *e* $P(B) \neq 0$ *, então dizemos que o número*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

é a **probabilidade condicional** *de* A *dado* B*. Se* $P(B) = 0$*, definimos* $P(A|B) = 0$*.*

Observe que a função $\overline{P} : \Lambda \to \mathbb{R}$ definida por

$$\overline{P}(A) = P(A|B) \ , \ \forall A \in \Lambda$$

é uma probabilidade se $P(B) \neq 0$. Consequentemente, as propriedades de probabilidade são mantidas, por exemplo:

$$P(A^c|B) = 1 - P(A|B) \ .$$

Observação: Quando não mencionarmos o espaço de probabilidade, estaremos considerando um espaço (Ω, Λ, P), onde P é uma medida de Probabilidade.

Definição 2.3. *Seja* (Ω, Λ, P) *um espaço de probabilidade. Dizemos que* $T \subseteq \Lambda$ *é uma família de eventos independentes, se*

$$P(A_1 \cap ... \cap A_n) = P(A_1). \ ... \ .P(A_n) \ , \ \forall A_1, ..., A_n \in T \ , \ \forall n \in \mathbb{N}.$$

Proposição 2.4. *Seja* (Ω, Λ, P) *um espaço de probabilidade.*

1) *Se* A *e* B *são eventos independentes de* Λ *e* $P(B) \neq 0$*, então*

$$P(A) = P(A|B) \ .$$

2) *Se* $A_1, ..., A_n$ *são eventos independentes, então* $B_1, ..., B_n$ *são eventos independentes, onde* $B_i = A_i$ *ou* $B_i = A_i^c$ *para todo* $i \in \{1, ..., n\}$*.*

Prova: 1) Das igualdades $P(A \cap B) = P(B).P(A|B)$ e $P(A \cap B) = P(B).P(A)$, segue que $P(A) = P(A|B)$.

2)Provemos que $A_1, ..., A_{n-1}, A_n^c$ são eventos independentes: se $i_1, ..., i_k \in \{1, ..., n-1\}$, então por hipótese

$$P(A_{i_1} \cap ... \cap A_{i_k}) = P(A_{i_1}). \ ... \ .P(A_{i_k}) \ .$$

Agora, como $A_{i_1} \cap ... \cap A_{i_k} \cap A_n$ e $A_{i_1} \cap ... \cap A_{i_k} \cap A_n^c$ são eventos disjuntos, cuja união é $A_{i_1} \cap ... \cap A_{i_k}$, temos que

$$P(A_{i_1} \cap ... \cap A_{i_k} \cap A_n^c) + P(A_{i_1} \cap ... \cap A_{i_k} \cap A_n) = P(A_{i_1} \cap ... \cap A_{i_k}) \ .$$

Assim,

$$P(A_{i_1} \cap ... \cap A_{i_k} \cap A_n^c) = P(A_{i_1} \cap ... \cap A_{i_k}) - P(A_{i_1} \cap ... \cap A_{i_k} \cap A_n) =$$

$$= P(A_{i_1}). \ ... \ .P(A_{i_k}).(1 - P(A_n)) = P(A_{i_1}). \ ... \ .P(A_{i_k}).P(A_n^c) \ .$$

Ou seja, $A_1, ..., A_{n-1}, A_n^c$ são eventos independentes. Aplicando sucessivamente esse resultado, teremos a demonstração do ítem 2). ■

Definição 2.5. *A função de distribuição acumulada de uma variável aleatória X, representada por F_X, é definida por*

$$F_X(x) = P([X \leq x]), \forall x \in \mathbb{R} \ .$$

Lembremos que $[X \leq x] = \{t \in \Omega : X(t) \leq x\}$.

Proposição 2.6. *Se X é uma variável aleatória, então a função de distribuição acumulada F_X satisfaz as seguintes propriedades:*

a) *F_X é não-decrescente.*

b) *Se $(x_n)_{n\in\mathbb{N}}$ é uma sequência crescente em $\mathbb{R}$, tal que $\lim\limits_{n\to\infty} x_n = +\infty$, então $\lim\limits_{n\to\infty} F_X(x_n) = 1$.*

c) *Se $(x_n)_{n\in\mathbb{N}}$ é uma sequência decrescente em $\mathbb{R}$, tal que $\lim\limits_{n\to\infty} x_n = -\infty$, então $\lim\limits_{n\to\infty} F_X(x_n) = 0$.*

d) *F_X é contínua à direita .*

e) *Dado $x \in \mathbb{R}$, considere o evento aleatório $[X = x] = \{t \in \Omega : X(t) = x\}$. Se $P([X = x]) = 0$, então F_X é contínua em x.*

Prova: **a)** Se $x, y \in \mathbb{R}$ e $x \leq y$, então $[X \leq x] \subseteq [X \leq y]$. Segue de propriedades ((1) na página 12), que $F_X(x) \leq F_X(y)$. Assim, F_X é não decrescente.

b) Como a sequência de eventos $([X \leq x_n])_{n\in\mathbb{N}}$ é crescente, com relação a inclusão, temos por propriedades ((2) da página 12)

$$\lim_{n\to\infty} F_X(x_n) = P\left(\bigcup_{n=1}^{\infty} [X \leq x_n]\right) \ .$$

Como $\bigcup_{n=1}^{\infty}[X \leq x_n] = \Omega$, temos que

$$\lim_{n\to\infty} F_X(x_n) = P(\Omega) = 1 \ .$$

c) A sequência de eventos $([X \leq x_n])_{n\in\mathbb{N}}$ é decrescente com relação a inclusão e $\bigcap_{n=1}^{\infty}[X \leq x_n] = \phi$. Logo, por propriedades ((3) da página 12) segue que

$$\lim_{n\to\infty} F_X(x_n) = P\left(\bigcap_{n=1}^{\infty}[X \leq x_n]\right) = P(\phi) = 0 \ .$$

d) Dado $x \in \mathbb{R}$, seja $(x_n)_{n\in\mathbb{N}}$ uma sequência decrescente em $\mathbb{R}$, tal que $\lim_{n\to\infty} x_n = x$. Como

$$\bigcap_{n=1}^{\infty}[X \leq x_n] = [X \leq x] \ ,$$

temos que $\lim_{n\to\infty} F_X(x_n) = P([X \leq x]) = F_X(x)$. Logo, F_X é contínua à direita.
e) Seja $(x_n)_{n\in\mathbb{N}}$ uma sequência crescente em $\mathbb{R}$, tal que $\lim_{n\to\infty} x_n = x$ e $x_n \neq x, \forall n \in \mathbb{N}$. Por propriedades ((3) da página 12), temos que

$$\lim_{n\to\infty} F_X(x_n) = P([X < x]) \ ,$$

onde $[X < x] = \{t \in \omega : X(t) < x\}$. Como $[X \leq x] = [X < x] \cup [X = x]$ temos que

$$F_X(x) = P([X < x]) + P([X = x]) = \lim_{n\to\infty} F_X(x_n) \ .$$

Assim F_X é contínua à esquerda de x. Pelo ítem d), F_X é contínua em x. ∎

No caso da σ-álgebra de Borel, a proposição anterior admite uma forma de recíproca:

Proposição 2.7. *Sejam $\Lambda = \beta(\Omega)$ a σ-álgebra de Borel sobre o intervalo aberto $\Omega = (0,1)$ de $\mathbb{R}$ e P a medida de Lebesgue restrita a Λ. Se $F : \mathbb{R} \to \mathbb{R}$ é uma função que satisfaz as propriedades a) , b) , c) e d) da proposição anterior, então existe uma variável aleatória $X : \Omega \to \mathbb{R}$ tal que*

$$F = F_X \ .$$

Prova: Defina $X : \Omega \to \mathbb{R}$ por

$$X(t) = inf\{x \in \mathbb{R} : F(x) \geq t\}, \forall t \in \Omega \ .$$

Uma vez que X é não-decrescente, ela é variável aleatória.
Dado $t \in \Omega$, defina $X(t) = y$. Pela definição de X, tome uma sequência decrescente $(x_n)_{n\in\mathbb{N}}$ tal que

$$F(x_n) \geq t, \forall n \in \mathbb{N} \qquad \text{e} \qquad \lim_{n\to\infty} x_n = y \ .$$

Como F é contínua à direita, temos que $\lim_{n\to\infty} F(x_n) = F(y)$. Logo, $F(y) \geq t$. Ou seja, provamos que $F(X(t)) \geq t$, $\forall t \in \Omega$. Assim, como F é não-decrescente, segue que

$$[X \leq x] = \{t \in \Omega : F(x) \geq t\} = (0, F(x)], \forall x \in \mathbb{R}.$$

Pela definição de P segue

$$F_X(x) = P([X \leq x]) = F(x), \forall x \in \mathbb{R} \ . \qquad \blacksquare$$

2.2 Distribuição de um vetor aleatório

Definição 2.8. *Se $X_1, X_2, ..., X_n$ são variáveis aleatórias definidas sobre um mesmo espaço de probabilidade (Ω, Λ, P), então dizemos que a função $X : \Omega \to \mathbb{R}^n$ dada por $X = (X_1, X_2, ..., X_n)$ é um vetor aleatório.*

A próxima proposição, mostra que a definição de vetor aleatório é coerente com a definição de variável aleatória (definição 1.16).

Proposição 2.9. *Uma função $X : \Omega \to \mathbb{R}^n$ é vetor aleatório se, e somente se, o conjunto*

$$[X \in D] = \{t \in \Omega : X(t) \in D\}$$

é evento aleatório, para todo $D \in \beta(\mathbb{R}^n)$.

Prova: Dado $x \in \mathbb{R}$, temos que $D = (-\infty, x] \times \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}$ pertence a álgebra de Borel $\beta(\mathbb{R}^n)$ e $[X \in D] = [X_1 \leq x]$. Assim, se $[X \in B]$ é evento aleatório para todo $B \in \beta(\mathbb{R}^n)$, então $[X_1 \leq x]$ é evento aleatório, $\forall x \in \mathbb{R}$ .Logo, X_1 é variável aleatória. De modo análogo, prova-se que $X_2, X_3, ..., X_n$ são variáveis aleatórias.
Reciprocamente, suponha que X seja vetor aleatório. É fácil ver que o conjunto

$$\Gamma = \{D : D \subset \mathbb{R}^n \text{ e } [X \in D] \in \Lambda\}$$

é uma σ -álgebra. Uma vez que X é vetor aleatório, temos que

$$(-\infty, a_1] \times ... \times (-\infty, a_n] \in \Gamma, \forall a_1, ..., a_n \in \mathbb{R}.$$

Como $\beta(\mathbb{R}^n)$ também é gerado pelo conjunto

$$\{(-\infty, a_1] \times ... \times (-\infty, a_n] : a_1, ..., a_n \in \mathbb{R}\} ,$$

temos que $\beta(\mathbb{R}^n) \subset \Gamma$. Assim, pela definição de Γ segue o resultado. ■

Definição 2.10. *A distribuição de um vetor aleatório $X : \Omega \to \mathbb{R}^n$ é a probabilidade P_X definida sobre $\beta(\mathbb{R}^n)$ dada por*

$$P_X(B) = P([X \in B]) \ , \ B \in \beta(\mathbb{R}^n) \ .$$

Da proposição anterior, segue que P_X está bem definida. Além disso, como P é uma probabilidade, temos que P_X também é uma probabilidade.

Novos vetores aleatórios podem ser considerados, a partir de um vetor aleatório X dado, conforme proposição abaixo.

Proposição 2.11. *Seja $X : \Omega \to \mathbb{R}^n$ um vetor aleatório. Se $g : \mathbb{R}^n \to \mathbb{R}^m$ é uma função que satisfaz*

$$g^{-1}(D) \in \beta(\mathbb{R}^n), \forall D \in \beta(\mathbb{R}^m) \ ,$$

então $g \circ X$ é vetor aleatório e sua distribuição é dada por

$$P_{g \circ X} = P_X \circ g^{-1} \ .$$

Prova: Se $D \in \beta(\mathbb{R}^m)$, então $[g \circ X \in D] = X^{-1}(g^{-1}(D))$. Assim, da proposição anterior, temos que $g \circ X$ é vetor aleatório. Agora, a distribuição $P_{g \circ X}$ satisfaz :

$$P_{g \circ X}(D) = P(X^{-1}(g^{-1}(D))) = P([X \in g^{-1}(D)]) = P_X(g^{-1}(D)), \forall D \in \beta(\mathbb{R}^m) \ .$$

Logo, $P_{g \circ X} = P_X \circ g^{-1}$. ■

Definição 2.12. *As variáveis aleatórias $X_1, X_2, ..., X_n$ são independentes, se satisfazem*

$$P(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i), \forall B_1, ..., B_n \in \beta(\mathbb{R}) \ ,$$

onde $(X_1 \in B_1, X_2 \in B_2, ..., X_n \in B_n) = \bigcap_{i=1}^{n} [X_i \in B_i]$.

Propriedades: Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes.

1) Se $B_1, ..., B_n \in \beta(\mathbb{R})$, então $[X_1 \in B_1], ..., [X_n \in B_n]$ são eventos aleatórios independentes.

2) Se $g_1, ..., g_n$ são funções Borel-mensuráveis, então $g_1 \circ X_1, ..., g_n \circ X_n$ são variáveis aleatórias independentes.

3) Se $X = (X_1, ..., X_n)$ é um vetor aleatório com distribuição P_X, então

$$P_X(B_1 \times B_2 \times ... \times B_n) = \prod_{i=1}^{n} P_{X_i}(B_i), \forall B_1, ..., B_n \in \beta(\mathbb{R}) \ ,$$

onde P_{X_i} é a distribuição de X_i.

Essas propriedades são consequências direta da definição 2.12 . Por isso omitiremos as suas demonstrações. Observe que a propriedade (3) significa que $X_1, ..., X_n$ são independentes se, e somente se, P_X é a medida produto $P_{X_1} \times P_{X_2} \times ... \times P_{X_n}$.

Definição 2.13. *Sejam (Ω, Λ, P) um espaço de probabilidade e $X = (X_1, ..., X_n)$ um vetor aleatório. Se existe uma função positiva $f : \mathbb{R}^n \to \mathbb{R}$ Borel-mensurável tal que*

$$P_X(B) = \int_B f d\lambda^n \ , \ \forall B \in \beta(\mathbb{R}^n) \ ,$$

então dizemos que f é a densidade de X.

Observe que o Teorema de Radon-Nikodym garante: X possui densidade se, e somente se, $P_X \ll \lambda^n$.

Proposição 2.14. *Seja $X = (Y, Z)$ um vetor aleatório. Se $P_X \ll \lambda^2$, então $P_Y \ll \lambda$ e $P_Z \ll \lambda$.*

Prova: Sejam f a densidade de X e $g : \mathbb{R} \to \mathbb{R}$ a função dada por

$$g(y) = \int_{\mathbb{R}} f(y, z) d\lambda(z) \ .$$

Pelo Teorema de Fubini, temos que g está bem definida e

$$P_Y(B) = P_X(B \times \mathbb{R}) = \int_{B \times \mathbb{R}} f d\lambda^2 = \int f(y, z)\chi_B(y) d\lambda^2 = \int \left(\int f(y, z)\chi_B(y) d\lambda(z) \right) d\lambda(y) =$$

$$= \int g(y)\chi_B(y) d\lambda(y) = \int_B g d\lambda \ , \ \forall B \in \beta(\mathbb{R}) \ .$$

Ou seja, Y possui densidade g e portanto $P_Y \ll \lambda$. De modo análogo, prova-se que a função $h : \mathbb{R} \to \mathbb{R}$ dada por

$$h(z) = \int_{\mathbb{R}} f(y,z)d\lambda(y) \ , \ \forall z \in \mathbb{R}$$

é a função densidade de Z. ∎

Com alguma restrição, a proposição anterior admite uma recíproca:

Proposição 2.15. *Seja $X = (Y, Z)$ um vetor aletório. Suponha que Y e Z são independentes. Se $P_Y \ll \lambda$ e $P_Z \ll \lambda$, então $P_X \ll \lambda^2$.*

Prova: Sejam g e h as densidades de Y e Z respectivamente. Defina a função Borel-mensurável $f : \mathbb{R}^2 \to \mathbb{R}$ por

$$f(y,z) = g(y).h(z) \ , \ \forall (y,z) \in \mathbb{R}^2$$

e a medida $\mu : \beta(\mathbb{R}^2) \to \mathbb{R}$ por

$$\mu(C) = \int_C f d\lambda^2 \ , \ \forall C \in \beta(\mathbb{R}^2) \ .$$

Pelo Teorema de Fubini temos

$$\mu(A \times B) = \left(\int_A g d\lambda \right) . \left(\int_B h d\lambda \right) = P_Y(A).P_Z(B) \ , \ \forall A, B \in \beta(\mathbb{R}) \ .$$

Ou seja, μ é a medida produto $P_Y \times P_Z$. Como Y e Z são variáveis aleatórias independentes, temos que P_X é a medida produto $P_Y \times P_Z$. Pela unicidade da medida produto, temos que $P_X = \mu$. Portanto, f é a densidade de X. ∎

Proposição 2.16. *Seja $X = (Y, Z)$ um vetor aleatório com densidade f. Se existem funções Borel-mensuráveis g e h não negativas, tais que $f(y,z) = g(y).h(z)$ $\lambda - q.s.$ e*

$$\int g d\lambda = \int h d\lambda = 1 \ ,$$

então Y e Z são independentes. Além disso, g e h são as densidades de Y e Z respectivamente.

Prova: Dado $y \in \mathbb{R}$, temos

$$\int_{\mathbb{R}} f(y,z)d\lambda(z) = \int_{\mathbb{R}} g(y).h(z)d\lambda(z) = g(y) \int_{\mathbb{R}} h \ d\lambda = g(y) \ .$$

Assim, pela demonstração da Proposição 2.14, temos que g é a densidade de Y. De modo análogo, prova-se que h é a densidade de Z.

Dados $A, B \in \beta(\mathbb{R})$, temos que

$$P(Y \in A, Z \in B) = P_X(A \times B) = \int_{A \times B} f d\lambda^2 = \left(\int_A g d\lambda \right) . \left(\int_B h d\lambda \right) = P_Y(A).P_Z(B) \ .$$

Assim, Y e Z são variáveis aleatórias independentes. ■

2.3 Esperança, Variância e Covariância

Definição 2.17. *Seja X uma variável aleatória integrável ($X \in L_1(\Omega, \Lambda, P)$). A esperança (média ou valor esperado) de X, denotada por $E(X)$, é o número*

$$E(X) = \int_\Omega X dP \ .$$

Propriedades: Sejam X e Y variáveis aleatórias integráveis e $\alpha \in \mathbb{R}$.

1)$E(\alpha.X) = \alpha.E(X)$.

2)$E(X + Y) = E(X) + E(Y)$.

3)Se $X \leq Y$, então $E(X) \leq E(Y)$.

As demonstrações dessas propriedades são simples, basta usar as propriedades de integral.

Teorema 2.18. *Sejam $g : \mathbb{R}^n \to \mathbb{R}$ uma função Borel-mensurável e $X : \Omega \to \mathbb{R}^n$ um vetor aleatório. Se $g \circ X \in L_1(\Omega, \Lambda, P)$, então $g \in L_1(\mathbb{R}^n, \beta(\mathbb{R}^n), P_X)$. Além disso,*

$$E(g \circ X) = \int_{\mathbb{R}^n} g dP_X \ .$$

Prova: Seja φ uma função Borel-mensurável simples, tal que

$$0 \leq \varphi \leq g^+ \ .$$

Como $g^+ \circ X = (g \circ X)^+$, temos que

$$0 \leq \varphi \circ X \leq (g \circ X)^+ \ . \qquad (*)$$

Se $\sum_{i=1}^{n} a_i.\chi_{A_i}$ é a representação padrão de φ, então

$$\varphi \circ X = \sum_{i=1}^{n} a_i.\chi_{B_i} \ ,$$

onde $B_i = [X \in A_i]$ para todo $i \in \{1, ..., n\}$. Portanto

$$\int_\Omega \varphi \circ X dP = \sum_{i=1}^{n} a_i.P(B_i) = \sum_{i=1}^{n} a_i.P_X(A_i) = \int_{\mathbb{R}^n} \varphi dP_X \ .$$

Logo, segue de (*) que

$$0 \leq \int_{\mathbb{R}^n} g^+ dP_X \leq \int_\Omega (g \circ X)^+ dP < +\infty \ .$$

De modo análogo, mostra-se que

$$\int_{\mathbb{R}^n} g^- dP_X < +\infty \ .$$

Portanto, $g \in L_1(\mathbb{R}^n, \beta(\mathbb{R}^n), P_X)$. Agora, suponha que g é função simples com representação padrão

$$\sum_{i=1}^{n} d_i.\chi_{D_i} \ .$$

Uma vez que $g \circ X = \sum_{i=1}^{n} d_i.\chi_{C_i}$, onde $C_i = [X \in D_i]$, temos

$$E(g \circ X) = \sum_{i=1}^{n} d_i.P(C_i) = \sum_{i=1}^{n} d_i.P_X(D_i) = \int_{\mathbb{R}^n} g dP_X \ .$$

Ou seja, o teorema é válido para g simples. Agora, se g é uma função qualquer, seja $(g_n)_{n\in\mathbb{N}}$ uma sequência de funções Borel-mensuráveis simples tais que

$$\lim_{n\to\infty} g_n = g \ \ e \ \ |g_n| \leq |g|, \forall n \in \mathbb{N}.$$

Temos que $(g_n \circ X)_{n\in\mathbb{N}}$ é uma sequência de variáveis aleatórias que satisfaz

$$\lim_{n\to\infty} g_n \circ X = g \circ X \ \ e \ \ |g_n \circ X| \leq |g \circ X| \ , \ \forall n \in \mathbb{N}.$$

Pelo Teorema da convergência dominada de Lebesgue

$$\lim_{n\to\infty} \int_{\mathbb{R}^n} g_n dP_X = \int_{\mathbb{R}^n} g dP_X$$

e

$$\lim_{n\to\infty}\int_\Omega (g_n \circ X)dP = \int_\Omega g \circ X dP \ .$$

Como o teorema é válido para funções simples, segue que

$$\int_\Omega (g_n \circ X)dP = \int_{\mathbb{R}^n} g_n dP_X, \forall n \in \mathbb{N}$$

e portanto

$$E(g \circ X) = \int_\Omega g \circ X dP = \int_{\mathbb{R}^n} g dP_X \ . \qquad \blacksquare$$

Proposição 2.19. *Se* $X \in L_1(\Omega, \Lambda, P)$*, então*

$$E(X) = \int_{\mathbb{R}} t \ dP_X(t)$$

Prova: Basta aplicar o teorema anterior a função $g : \mathbb{R} \to \mathbb{R}$ definida por

$$g(t) = t \ , \ \forall t \in \mathbb{R} \ . \qquad \blacksquare$$

Antes de definir a variância de uma variável aleatória $X \in L_2(\Omega, \Lambda, P)$, será necessário garantir a existência da esperança $E(X)$. Para isso precisamos da seguinte proposição :

Proposição 2.20. *Seja* X *uma variável aleatória. Se* $X \in L_p(\Omega, \Lambda, P)$ *e* $1 \leq q \leq p$*, então* $X \in L_q(\Omega, \Lambda, P)$*.*

Prova: Como P é uma probabilidade, temos que a função constante que assume o valor "1 "é integrável. Assim, $1 + |X|^p$ é integrável. Da desigualdade

$$|X(t)|^q \ < \ 1 + |X(t)|^p \ , \ \forall t \in \Omega$$

segue que $X \in L_q(\Omega, \Lambda, P)$. $\blacksquare$

Definição 2.21. *Seja* $X \in L_2(\Omega, \Lambda, P)$ *uma variável aleatória. A variância de* X*, denotada por* $var(X)$*, é a esperança*

$$var(X) = E([X - E(X)]^2) \ .$$

Dizemos ainda que $\sigma(X) = (var(X))^{\frac{1}{2}}$ *é o desvio padrão de* X*. Se* $Y \in L_2(\Omega, \Lambda, P)$ *é uma variável aleatória, então definimos a covariância de* X *e* Y *por*

$$Cov(X, Y) = E([X - E(X)].[Y - E(Y)]) \ .$$

A desigualdade de Hölder garante que $(X - E(X)).(Y - E(Y))$ é integrável. Deste modo, a covariância está bem definida.

Propriedades: Sejam $X, Y \in L_2(\Omega, \Lambda, P)$ variáveis aleatórias e $\alpha \in \mathbb{R}$ um número Real.

1) $var(X) = Cov(X, X)$.

2) $var(X) = E(X^2) - (E(X))^2$.

3) $var(\alpha.X) = \alpha^2.var(X)$.

4) $Cov(X, Y) = E(X.Y) - E(X).E(Y)$.

5) $var(X + Y) = var(X) + 2Cov(X, Y) + var(Y)$.

6) $|Cov(X, Y)| \leq \sigma(X).\sigma(Y)$.

Prova: Essas propriedades são fáceis de serem demonstradas. Provaremos apenas àquela mencionada no ítem (6):

$$|Cov(X, Y)| = \left| \int_{\Omega} [X - E(X)].[Y - E(Y)] dP \right| \leq \int_{\Omega} |X - E(X)|.|Y - E(Y)| dP \ .$$

Da desigualdade de Hölder segue:

$$|Cov(X, Y)| \leq \left(\int_{\Omega} |X - E(X)|^2 dP \right)^{\frac{1}{2}} . \left(\int_{\Omega} |Y - E(Y)|^2 dP \right)^{\frac{1}{2}} = \sigma(X).\sigma(Y) \ . \qquad \blacksquare$$

Definição 2.22. *Sejam X e Y variáveis aleatórias pertencentes a $L_2(\Omega, \Lambda, P)$. Se $Cov(X, Y) = 0$, dizemos que X e Y são não correlacionadas.*

Observe que, perante as propriedades anteriores, as condições abaixo são equivalentes:

1) X e Y são não correlacionadas.

2) $E(X.Y) = E(X).E(Y)$.

3) $var(X + Y) = var(X) + var(Y)$.

Proposição 2.23. *Se $X, Y \in L_2(\Omega, \Lambda, P)$ são variáveis aleatórias independentes, então X e Y são não correlacionadas.*

Prova: Seja $g : \mathbb{R}^2 \to \mathbb{R}$ a função Borel-mensurável dada por

$$g(x, y) = x.y \qquad .$$

Temos que

$$E(X).E(Y) = \left(\int_{\Omega} X dP \right) \left(\int_{\Omega} Y dP \right) .$$

Pela Proposição 2.19 e Teorema de Fubini segue

$$E(X).E(Y) = \left(\int_{\mathbb{R}} x \; dP_X(x) \right) . \left(\int_{\mathbb{R}} y \; dP_Y(y) \right) = \int_{\mathbb{R}^2} g d\mu \; ,$$

onde μ é a medida produto $P_X \times P_Y$. Denotando por Z o vetor aleatório (X, Y), temos que $P_Z = \mu$, pois X e Y são independentes. Portanto, do Teorema 2.18 segue:

$$E(X).E(Y) = \int_{\mathbb{R}^2} g dP_Z = \int_{\Omega} g \circ Z dP = \int_{\Omega} X.Y dP = E(X.Y) \; .$$

Logo, X e Y são não correlacionadas. ■

Definição 2.24. *Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias pertencentes a $L_1(\Omega, \Lambda, P)$. A esperança do vetor aleatório $X = (X_1, X_2, ..., X_n)$, denotada por $E(X)$, é o vetor do $\mathbb{R}^n$ definido por*

$$E(X) = (E(X_1), E(X_2), ..., E(X_n)) \; .$$

No que segue, usaremos a notação matricial para X e $E(X)$:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \; e \; E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix} .$$

Caso contrário, ficará claro no contexto quando não estamos usando notação matricial. Por comodidade, dizemos que X é um **n-vetor aleatório** se $X_i \in L_2(\Omega, \Lambda, P)$. Observe que estamos diferenciando o termo **vetor aleatório** de **n-vetor aleatório**.

Definição 2.25. *Sejam X e Y n-vetor e m-vetor aleatórios respectivamente. A covariância de X e Y , denotada por $Cov(X, Y)$, é a matriz definida por*

$$Cov(X, Y) = E([X - E(X)][Y - E(Y)]^t) \; .$$

Observe que o produto acima é o produto de matrizes e que $Cov(X,Y)$ é uma matriz pertencente a $M_{n\times m}(\mathbb{R})$. Se

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad e \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$$

então a entrada (i,j) da matriz $Cov(X,Y)$ é dada por $Cov(X_i,Y_j)$.

Nesse sentido, podemos estender a definição de não correlacionados para vetores aleatórios. Dizemos que X e Y são *não correlacionados* se a matriz $Cov(X,Y)$ é nula.

Propriedades: Sejam X e Y n-vetor e m-vetor aleatórios respectivamente.

1) Se $A \in M_{k\times n}(\mathbb{R})$, então $E(A.X) = A.E(X)$.

2) Se $A \in M_{k\times n}(\mathbb{R})$ e $B \in M_{l\times m}(\mathbb{R})$, então $Cov(A.X, B.Y) = A.Cov(X,Y).B^t$.

3) $Cov(X,Y) = E(X.Y^t) - E(X).E(Y^t)$.

4) Se $E(X) = 0$ ou $E(Y) = 0$, então $Cov(X,Y) = E(X.Y^t)$.

Prova:

1) Se $A = (a_{i,j})_{i,j}$, então

$$E\left(\sum_{j=1}^{n} a_{i,j}.X_j\right) = \sum_{j=1}^{n} a_{i,j}E(X_j)\ ,\ \forall i \in \{1,...,k\}.$$

Logo, $E(A.X) = A.E(X)$.

2) $Cov(A.X,B.Y) = E([A.X - E(A.X)].[B.Y - E(B.Y)]^t) = E(A[X - E(X)].[Y - E(Y)]^t B^t) =$

$$= A.E([X - E(X)].[Y - E(Y)]^t)B^t = A.Cov(X,Y).B^t\ .$$

3) $Cov(X,Y) = E([X - E(X)].[Y - E(Y)]^t) = E(X.Y^t - X.E(Y)^t - E(X).Y^t + E(X).E(Y)^t) =$

$$= E(X.Y^t) - E(X).E(Y)^t - E(X).E(Y^t) + E(X).E(Y)^t = E(X.Y^t) - E(X).E(Y^t)\ .$$

4) Consequência direta da anterior. ■

Seja $H = L_2(\Omega, \Lambda, P) \times ... \times L_2(\Omega, \Lambda, P)$ o espaço vetorial formado por todos **n-vetores aleatórios**. Defina em H o produto interno $< \, , \, >: H \times H \to \mathbb{R}$ dado por:

$$< X, Y >= \sum_{i=1}^{n} E(X_i.Y_i) \ ,$$

onde $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ e $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$. Uma vez que $L_2(\Omega, \Lambda, P)$ é espaço de Hilbert com o produto interno $(\ | \ )$ definido por $(f|g) = E(f.g)$, temos que H é espaço de Hilbert com o produto interno acima. Algumas das propriedades do espaço H que serão utilizadas no próximo capítulo, são as seguintes:

Propriedades: Sejam $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ e $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ elementos de H.

1) $< X, Y >= E(X^t.Y)$.

2) $< X, Y >= tr \ E(X.Y^t)$, onde $tr : M_n(\mathbb{R}) \to \mathbb{R}$ é a função traço (soma dos elementos da diagonal da matriz).

3) $\|X\|^2 = \sum_{i=1}^{n} \|X_i\|_2^2$, onde $\| \ \|_2$ é a norma induzida pelo produto interno $(\ | \ )$ em $L_2(\Omega, \Lambda, P)$.

4) $||X||^2 = tr \ G$, onde $G = G(X_1, ..., X_n)$ é a matriz de Gram relativa ao produto interno $(\ | \ )$.

Entre as propriedades citadas, chamamos a atenção para a propriedade (3). Ela será de grande importância na seção de Estimador de mínima variança.

$$\|X\|^2 =< X, X >= \sum_{i=1}^{n} E(X_i^2) = \sum_{i=1}^{n} (X_i|X_i) = \sum_{i=1}^{n} \|X_i\|_2^2 \ .$$

2.4 Esperança Condicional

Sejam Λ e β σ-álgebras de um espaço amostral Ω. Se $\beta \subseteq \Lambda$, então dizemos que β é sub-σ-álgebra de Λ.

Proposição 2.26. *Sejam (Ω, Λ, μ) um espaço de medida e β uma sub-σ-álgebra de Λ. Denote por ν a medida μ restrita a β. Se X é função β-mensurável não negativa, então*

$$\int_\Omega X d\nu \ = \ \int_\Omega X d\mu \ .$$

Prova: Se φ é função β-mensurável simples, não negativa e com representação padrão $\sum_{i=1}^{n} a_i.\chi_{A_i}$, então

$$\int_\Omega \varphi d\nu \ = \ \sum_{i=1}^{n} a_i.\nu(A_i) \ = \ \sum_{i=1}^{n} a_i.\mu(A_i) = \int_\Omega \varphi d\mu \ ,$$

ou seja, o resultado é válido para funções simples. Agora, seja $(\varphi_n)_{n\in\mathbb{N}}$ uma sequência crescente de funções β-mensuráveis simples e não negativas tais que :

$$\lim_{n\to\infty} \varphi_n = X \ .$$

Pelo Teorema da Convergência Monótona, temos

$$\int_\Omega X d\nu \ = \ \lim_{n\to\infty} \int_\Omega \varphi_n d\nu \ = \ \lim_{n\to\infty} \int_\Omega \varphi_n d\mu \ = \ \int_\Omega X d\mu \quad . \qquad \blacksquare$$

Definição 2.27. *Sejam $X \in L_1(\Omega, \Lambda, P)$, onde P é uma probabilidade, e β uma sub-σ-álgebra de Λ. A* **esperança condicional** *de X em relação a β é uma função $Y : \Omega \to \mathbb{R}$, β-mensurável que satisfaz*

$$\int_B Y dP = \int_B X dP \ , \ \forall B \in \beta \ .$$

Nesse caso, denotamos $Y = E(X|\beta)$.

Observações: (a) Se Y e Z são as esperanças condicionais de X em relação a β , então $Y = Z$ P-q.s.

(b) $E(X|\beta)$ sempre existe, de fato: defina as medidas μ^+ e μ^- sobre a σ-ágebra β por

$$\mu^+(B) = \int_B X^+ dP \ , \ \forall B \in \beta \ ,$$

$$\mu^-(B) = \int_B X^- dP \ , \ \forall B \in \beta \ .$$

Denote por $\widehat{P}$ a probabilidade P restrita a β. Pela definição de μ^+, temos que $\mu^+ \ll \widehat{P}$. Assim, pelo teorema de Radon-Nikodym, existe uma função β-mensurável f não negativa tal que

$$\mu^+(B) = \int_B f d\widehat{P} \ , \ \forall B \in \beta \ .$$

Pela proposição anterior, temos que:

$$\mu^+(B) = \int_B f dP \ , \ \forall B \in \beta \ .$$

De maneira análoga, prova-se que existe uma função β-mensurável g tal que

$$\mu^-(B) = \int_B g dP \ , \ \forall B \in \beta \ .$$

Portanto, a esperança condicional de X é a função β-mensurável

$$E(X|\beta) = f - g \ . \qquad \blacksquare$$

A seguir listamos as principais propriedades da esperança condicional.

Propriedades: Sejam $X, Y \in L_1(\Omega, \Lambda, P)$ e β uma sub-σ-álgebra de Λ.

1) Se $a, b \in \mathbb{R}$, então $E(a.X + b.Y|\beta) = a.E(X|\beta) + b.E(Y|\beta)$.

2) Se X é β-mensurável, então $E(X|\beta) = X$.

3) Se $X \geq 0$, então $E(X|\beta) \geq 0$ P-q.s.

4) Se $X \leq Y$, então $E(X|\beta) \leq E(Y|\beta)$ P-q.s.

5) Se γ é sub-σ-álgebra de β, então $E(X|\gamma) = E(E(X|\beta)|\gamma)$ P-q.s.

Prova: 1) Temos que:

$$\int_B a.E(X|\beta) + b.E(Y|\beta) dP = a \int_B E(X|\beta) dP + b \int_B E(Y|\beta) dP =$$

$$= a \int_B X dP + b \int_B Y dP = \int_B aX + bY dP \ , \ \forall B \in \beta \ .$$

Logo, segue o resultado.

2) A verificação é imediata.

3) Defina o conjunto β-mensurável $N = \{t \in \Omega : E(X|\beta)(t) < 0\}$. Como $-E(X|\beta).\chi_N \geq 0$, temos que

$$\int_N E(X|\beta)dP \leq 0 \ .$$

Mas,

$$\int_N E(X|\beta)dP = \int_N XdP \ e \ \int_N XdP \geq 0 \ .$$

Logo,

$$\int_N E(X|\beta)dP = 0 \ .$$

Assim, de $-E(X|\beta).\chi_N \geq 0$, temos que $-E(X|\beta).\chi_N = 0$ P-q.s. Logo, $P(N) = 0$. ou seja, $E(X|\beta) \geq 0$ a menos do conjunto de medida nula N.

4) Se $X \leq Y$, então $0 \leq (Y - X)$. Assim, pela propriedade (3) temos

$$E(Y - X|\beta) \geq 0 \ \text{ P-q.s .}$$

Como $E(Y - X|\beta) = E(Y|\beta) - E(X|\beta)$, segue o resultado.

5) Para todo $B \in \gamma$ vale:

$$\int_B E(E(X|\beta)|\gamma)dP = \int_B E(X|\beta)dP = \int_B XdP \ .$$

Logo, $E(E(X|\beta)|\gamma) = E(X|\gamma)$ P-q.s. ■

Proposição 2.28. *Sejam $X, Y \in L_2(\Omega, \Lambda, P)$ e β uma sub-σ-álgebra de Λ. Se Y é β-mensurável, então $E(X.Y|\beta) = Y.E(X|\beta)$.*

Prova: Seja φ uma função β-mensurável simples com representação padrão $\sum_{i=1}^{m} a_i.\chi_{B_i}$. Para todo $B \in \beta$ temos

$$\int_B \varphi.E(X|\beta)dP = \sum_{i=1}^{m} a_i. \int_B E(X|\beta).\chi_{B_i}dP = \sum_{i=1}^{m} a_i. \int_{B\cap B_i} E(X|\beta)dP =$$

$$= \sum_{i=1}^{m} a_i . \int_{B \cap B_i} X dP = \sum_{i=1}^{m} a_i . \int_B X . \chi_{B_i} dP = \int_B \varphi . X dP \ .$$

Portanto, $E(\varphi.X|\beta) = \varphi.E(X|\beta)$.

Como Y é mensurável, existe uma sequência de funções $(\varphi_n)_{n\in\mathbb{N}}$ β-mensuráveis simples que satisfazem:

a) $|\varphi_n(x)| \leq |Y(x)|$, $\forall n \in \mathbb{N}$ e $\forall x \in \Omega$.

b) $\lim\limits_{n\to\infty} \varphi_n = Y$.

Pelo Teorema da Convergência Dominada de Lebesgue temos

$$\lim_{n\to\infty} \int_B \varphi_n X dP = \int_B Y.X dP \ , \ \forall B \in \beta.$$

Por outro lado,

$$\lim_{n\to\infty} \int_B \varphi_n X dP = \lim_{n\to\infty} \int_B E(\varphi_n X|\beta) dP = \lim_{n\to\infty} \int_B \varphi_n . E(X|\beta) dP \ , \ \forall B \in \beta.$$

Assim, pelo Teorema da Convergência Dominada de Lebesgue temos

$$\lim_{n\to\infty} \int_B \varphi_n X dP = \int_B Y.E(X|\beta) dP, \forall B \in \beta \ .$$

Logo

$$\int_B Y.E(X|\beta) dP = \int_B Y.X dP, \forall B \in \beta. \qquad \blacksquare$$

Observação: seja Ω um conjunto não vazio e $Y : \Omega \to \mathbb{R}^n$ uma função qualquer. É fácil ver que a família $\beta(Y)$, dada por $\beta(Y) = \{Y^{-1}(B) : B \in \beta(\mathbb{R}^n)\}$, define uma σ-álgebra de subconjuntos de Ω. Por outro lado, também é fácil ver que se Λ é uma σ-álgebra de subconjuntos de Ω tal que Y é Λ-mensurável, então $\beta(Y) \subseteq \Lambda$, ou seja, $\beta(Y)$ é a menor σ-álgebra sobre Ω que torna Y mensurável.

Definição 2.29. *Seja $X \in L_1(\Omega, \Lambda, P)$ uma variável aleatória. A esperança condicional de X, dado um vetor aleatório $Y : \Omega \to \mathbb{R}^n$, é a variável aleatória $E(X|\beta(Y))$.*

Notação: $E(X|Y) = E(X|\beta(Y))$.

Teorema 2.30. *Seja $X \in L_1(\Omega, \Lambda, P)$ uma variável aleatória. Se $Y : \Omega \to \mathbb{R}^n$ é um vetor aleatório, então existe uma função Borel-mensurável $g : \mathbb{R}^n \to \mathbb{R}$ tal que:*

$$E(X|Y) = g \circ Y \ \ P - q.s$$

Observação: omitiremos a demonstração deste teorema, mas ela é uma consequência direta do seguinte resultado sobre mensurabilidade: Sejam Ω, W conjuntos não vazios, (W, Σ) um espaço mensurável e $Y : \Omega \to W$ uma aplicação. Então, uma função $X : \Omega \to \mathbb{R}$ é $\beta(Y)$-mensurável se, e somente se, $X = g \circ Y$ para alguma função g Σ -mensurável sobre W. (para os detalhes veja [3] na página 98)

Proposição 2.31. *Sejam $X \in L_2(\Omega, \Lambda, P)$ e $Y : \Omega \to \mathbb{R}^n$ uma variável e um vetor aleatórios respectivamente. Se $g : \mathbb{R}^n \to \mathbb{R}$ é uma função Borel-mensurável tal que $g \circ Y \in L_2(\Omega, \Lambda, P)$, então*

$$(X - E(X|Y)) \perp (E(X|Y) - g \circ Y) \ .$$

Antes de demonstrar a proposição, enunciaremos um resultado:

Desigualdade Condicional de Jensen: seja $f \in L_1(\Omega, \Lambda, P)$ uma variável aleatória cuja imagem está contida em um conjunto aberto e convexo C. Seja $h : C \to \mathbb{R}$ uma função convexa:

$$h(t.x + (1-t).y) \leq t.h(x) + (1-t).h(y), \quad \forall x, y \in C \ \ e \ \ 0 \leq t \leq 1 \ .$$

Se $h \circ f$ é integrável e β é sub-σ-álgebra de Λ, então

$$E(h \circ f|\beta) \geq h(E(f|\beta)) \quad .$$

A demonstração desse resultado pode ser encontrada em [3] na página 274.

Prova: Pois bem, a função $h : \mathbb{R} \to \mathbb{R}$ definida por $h(t) = t^2$ é convexa e $h \circ X = X^2$ é integrável. Logo, pela desigualdade de Jensen temos

$$E(X^2|Y) \geq [E(X|Y)]^2$$

Ou seja, $E(X|Y) \in L_2$. Da definição de esperança condicional e da Proposição 2.28, seguem as igualdades:

$$(1) \qquad \int_\Omega X.E(X|Y)dP = \int_\Omega E(X.E(X|Y)|Y)dP = \int_\Omega E(X|Y).E(X|Y)dP$$

$$(2) \qquad \int_\Omega X.g\circ Y dP = \int_\Omega E(X.g\circ Y|Y)dP = \int_\Omega g\circ Y.E(X|Y)dP \ .$$

Agora, das propriedades de produto interno e das expressões (1) e (2) temos

$$< X - E(X|Y)\ ,\ E(X|Y) - g\circ Y >= \int_\Omega X.E(X|Y)dP - \int_\Omega X.g\circ Y dP +$$

$$- \int_\Omega E(X|Y)^2 dP + \int_\Omega E(X|Y).g\circ Y dP = 0 \ .$$

Portanto $(X - E(X|Y)) \perp (E(X|Y) - g\circ Y)$. ■

Capítulo 3

Estimadores e o Teorema de Kalman

Nesse capítulo trataremos do tema "Estimador ". Com o conceito de Esperança Condicional, descrito na seção 2.4, obtemos a relação que caracteriza o Estimador de mínima variância (Teorema 3.6).

Com relação a importância da seção 3.3, a partir dela obteremos ferramentas necessárias para que o Teorema de Kalman, por fim, seja demonstrado.

3.1 Estimativa pelos mínimos quadrados

Sejam (x_1, y_1) , (x_2, y_2) , ... , (x_n, y_n) pontos do $\mathbb{R}^2$. Estamos interessados em determinar uma linha reta que "melhor se adapta " aos dados acima, segundo um critério pré estabelecido. Devemos procurar uma função $f : \mathbb{R} \to \mathbb{R}$ da forma $f(x) = a + bx$. Dada uma função f desta forma, obtemos os pontos

$$(x_1, f(x_1)) \ , \ ... \ , \ (x_n, f(x_n)) \ .$$

Obviamente, não temos necessariamente as igualdades

$$f(x_i) = y_i \ , \ \forall i \in \{1, ..., n\} \ .$$

Agora, para cada i, temos um erro e_i definido por $e_i = y_i - f(x_i)$. O critério que será utilizado para determinar a "melhor reta " é dado por: minimizar a soma dos erros ao quadrado $\sum_{i=1}^{n} {e_i}^2$

, isto é, encontrar uma função f que minimiza a soma

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 \ .$$

Pois bem, sejam

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad , \quad M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{e} \quad Z = \begin{bmatrix} a \\ b \end{bmatrix}$$

matrizes e $< \ , \ >$ o produto interno Euclidiano no $\mathbb{R}^n$. Uma vez que

$$\|Y - M.Z\|^2 =< Y - M.Z, Y - M.Z >= \sum_{i=1}^{n}(y_i - f(x_i))^2 \ ,$$

temos que determinar Z que minimiza

$$\|Y - M.Z\|^2 \ .$$

Como $\mathbb{R}^n$ é um espaço de Hilbert e $W = \{M.K : K \in M_{2\times 1}(\mathbb{R})\}$ é um subespaço vetorial fechado de $\mathbb{R}^n$, temos pelo Teorema da Projeção, que o vetor $Y - M.Z$ de norma mínima, é o vetor que satisfaz

$$(Y - M.Z) \perp W \ .$$

Diante dessa condição, podemos caracterizar a matriz Z desejada conforme abaixo

$$(Y - M.Z) \perp W \iff \ < Y - M.Z \ , \ M.K >= 0 \ , \ \forall K \in M_{2\times 1}(\mathbb{R}) \iff$$

$$\iff 0 = (M.K)^t.(Y - M.Z) = K^t.(M^t.Y - M^t.M.Z) \ , \ \forall K \in M_{2\times 1}(\mathbb{R}) \iff$$

$$\iff M^t.Y - M^t.M.Z = 0 \ .$$

Assim, a matriz Z desejada é a solução do sistema:

$$(M^t.M).Z = M^t.Y \ .$$

Denotando por M_i a i-ésima coluna de M, temos que $(M^t.M)$ é a matriz de Gram $G(M_1, M_2)$ com relação ao produto interno $< \ , \ >$ em $\mathbb{R}^n$. Como $\{M_1, M_2\}$ é um conjunto linearmente independente, temos pela Proposição 1.10 que $M^t.M$ é invertível. Logo,

$$Z = (M^t.M)^{-1}.M^t.Y \ .$$

A linha $f(x) = a + b.x$ que minimiza

$$\sum_{i=1}^{n} (y_i - f(x_i))^2$$

é chamada *aproximação linear pelos mínimos quadrados.*

Teorema 3.1. *(Estimativa pelos mínimos quadrados): Sejam $y \in M_{m\times 1}(\mathbb{R})$ um vetor e $W \in M_{m\times n}(\mathbb{R})$ uma matriz, cujas colunas formam um conjunto linearmente independente e $n \leq m$. Então existe um único vetor $\widehat{x}$ tal que*

$$\|y - W.\widehat{x}\| \leq \|y - W.x\| \ , \ \forall x \in M_{n\times 1}(\mathbb{R})$$

Além disso, $\widehat{x} = (W^t.W)^{-1}.W^t.y$.

Prova: O subconjunto $H = \{W.x : x \in M_{n\times 1}(\mathbb{R})\}$ é um subespaço vetorial fechado de $\mathbb{R}^m$, pois tem dimensão finita. Como $\mathbb{R}^m$ é espaço de Hilbert, temos pelo Teorema da Projeção, que existe um único $Z \in H$ que minimiza $\|y - Z\|$. Denotando $Z = W.\widehat{x}$, o Teorema da Projeção garante que

$$(y - W.\widehat{x}) \perp H \ .$$

Portanto, para todo $x \in M_{n\times 1}(\mathbb{R})$, temos

$$0 = < y - W.\widehat{x}, W.x > = (W.x)^t.(y - W.\widehat{x}) = x^t.(W^t.y - W^t.W.\widehat{x}) \ .$$

Logo,

$$W^t.y - W^t.W.\widehat{x} = 0 \ .$$

Denotando por W_i a i-ésima coluna da matriz W, temos que $W^t.W$ é a matriz de Gram $G(W_1, ..., W_n)$. Como $\{W_1, ..., W_n\}$ é um conjunto linearmente independente, temos pela Proposição 1.10 que $W^t.W$ é invertível. Logo, de $W^t.y = W^t.W.\widehat{x}$, segue que

$$\widehat{x} = (W^t.W)^{-1}.W^t.y \ . \qquad \blacksquare$$

Observe que a aproximação linear pelos mínimos quadrados é um caso particular desse teorema.

3.2 Estimador de mínima variância

Nessa seção (Ω, Λ, P) é um espaço de probabilidade, $Y : \Omega \to \mathbb{R}^n$ um vetor aleatório e $X : \Omega \to \mathbb{R}$ uma variável aleatória.

Definição 3.2. *Um* **estimador** $\widehat{X}$ *de* X*, baseado em* Y*, é uma variável aleatória da forma*

$$\widehat{X} = g \circ Y$$

para alguma função $g : \mathbb{R}^n \to \mathbb{R}$ *Borel-mensurável. Se* $y = Y(t)$ *é uma "medida " de* Y*, então dizemos que* $\widehat{x} = g(y)$ *é uma* **estimativa** *de* X *baseado em* y*, isto é,* $\widehat{x} = \widehat{X}(t)$ *.*

Claramente o estimador $\widehat{X}$ depende da escolha da função g. No entanto, estaremos interessados em "estimadores especiais" no decorrer da dissertação. Pois bem, suponha $X \in L_2(\Omega, \Lambda, P)$ e seja $\| \ \|_2$ a norma sobre L_2 induzida pelo produto interno $(\ | \ )$ definido na página 31, ou seja,

$$\| X \|_2 = (E(X^2))^{\frac{1}{2}} = \left(\int_\Omega |X|^2 dP \right)^{\frac{1}{2}} .$$

Definição 3.3. *Um estimador* $\widehat{X}$ *de* X*, baseado em* Y*, é chamado* ***estimador de mínima variância*** *se* $\widehat{X} \in L_2$ *e*

$$\| \widehat{X} - X \|_2 \leq \| h \circ Y - X \|_2$$

para toda função h *Borel-mensurável que satisfaz* $h \circ Y \in L_2$*.*

Nesse caso, se $y = Y(t)$ é uma medida de Y, então dizemos que $\widehat{x} = \widehat{X}(t)$ é uma estimativa de mínima variância de X baseado em y.

Teorema 3.4. *Se* $M(Y)$ *é o conjunto de todas funções* $h \circ Y$ *tal que* h *é Borel-mensurável e* $h \circ Y \in L_2$*, então* $M(Y)$ *é um subespaço fechado de* L_2*.*

Prova: A demonstração que $M(Y)$ é subespaço de L_2 é imediata. Provemos que $M(Y)$ é fechado: seja $(g_n)_{n\in\mathbb{N}}$ uma sequência de funções Borel-mensuráveis tal que $g_n \circ Y \in M(Y)$ para todo $n \in \mathbb{N}$ e $g_n \circ Y \to Z$ na norma de L_2. Da Proposição 2.31 e Teorema de Pitágoras, temos que

$$\| Z-g_n\circ Y \|_2^2 = \| Z-E(Z|Y)+E(Z|Y)-g_n\circ Y \|_2^2 = \| Z-E(Z|Y) \|_2^2 + \| E(Z|Y)-g_n\circ Y \|_2^2 \Rightarrow$$

$$\Rightarrow \| Z - E(Z|Y) \|_2^2 + \lim_{n\to\infty} \| E(Z|Y) - g_n \circ Y \|_2^2 = 0 \Rightarrow$$

$$\Rightarrow \| Z - E(Z|Y) \|_2^2 = 0 \Rightarrow Z = E(Z|Y) \ .$$

Pelo Teorema 2.30 existe g Borel-mensurável tal que $Z = g \circ Y$. Portanto, $M(Y)$ é fechado. ■

Corolário 3.5. *O estimador de mínima variância $\widehat{X}$ de X é a projeção ortogonal de X sobre o subespaço $M(Y)$. Além disso, temos que $(\widehat{X} - X) \perp M(Y)$.*

Prova: Como L_2 é espaço de Hilbert e $M(Y)$ é um subespaço fechado de L_2, temos pelo Teorema da Projeção , que o estimador $\widehat{X}$ satisfaz

$$(\widehat{X} - X) \perp M(Y) \ .$$

Ou seja, $\widehat{X}$ é a projeção ortogonal de X sobre $M(Y)$. ■

Teorema 3.6. *A variável aleatória $E(X|Y)$ é o estimador de mínima variância de X baseado em Y.*

Prova: Seja $g : \mathbb{R}^n \to \mathbb{R}$ uma função Borel-mensurável tal que $g \circ Y \in M(Y)$. Então pelo Teorema de Pitágoras e Proposição 2.31 temos que

$$\| X - g \circ Y \|_2^2 = \| X - E(X|Y) + E(X|Y) - g \circ Y \|_2^2 =$$

$$= \| X - E(X|Y) \|_2^2 + \| E(X|Y) - g \circ Y \|_2^2 \geq \| X - E(X|Y) \|_2^2 \ .$$

Logo $E(X|Y)$ é o estimador de mínima variância. ■

Definição 3.7. *Um estimador $\widehat{X}$ de X ,baseado em Y, é chamado* ***não-tendencioso*** *se*

$$E(\widehat{X}) = E(X) \ .$$

Observação: Se $E(\widehat{X}|Y) = X$, então $\widehat{X}$ é um estimador não tendencioso de X, de fato, pela definição de esperança condicional, em particular temos para o conjunto Ω

$$E(X) = E(E(\widehat{X}|Y)) = \int_\Omega E(\widehat{X}|Y)dP = \int_\Omega \widehat{X}dP = E(\widehat{X}) \ .$$

Observamos que em geral, a condição $E(\widehat{X}|Y) = X$ aparece como a definição de estimador não tendencioso. Um exemplo de estimador não tendencioso é o de mínima variância. De fato,

$$E(E(X|Y)) = \int_\Omega E(X|Y)dP = \int_\Omega XdP = E(X) \ .$$

3.3 Estimador linear de mínima variância

Para facilitar a compreensão da demonstração do teorema de Kalman no próximo parágrafo, parte de sua demonstração foi antecipada para alguns dos teoremas deste parágrafo.

Sejam X , Y n-vetor e m-vetor aleatórios respectivamente indicados por

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad e \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} .$$

Como foi visto na página 31, o espaço $H = L_2(\Omega, \Lambda, P) \times ... \times L_2(\Omega, \Lambda, P)$ possui uma norma $\| \ \|$ induzida pela norma $\| \ \|_2$ de $L_2(\Omega, \Lambda, P)$, a saber,

$$\| X \| = \left(\sum_{i=1}^{n} \| X_i \|_2^2 \right)^{\frac{1}{2}} = \left(\sum_{i=1}^{n} E(X_i^2) \right)^{\frac{1}{2}} .$$

Definição 3.8. *O estimador linear de mínima variância de X, baseado em Y, é o vetor aleatório $\widehat{X} = (\widehat{X}_1, ..., \widehat{X}_n) \in H$ que satisfaz:*

a) $\widehat{X} = K.Y$, *onde $K \in M_{n\times m}(\mathbb{R})$ é uma matriz constante (estimador linear).*

b) $\| \widehat{X} - X \|$ *é mínimo com respeito ao ítem* **a)**, *ou seja,*

$$\| \widehat{X} - X \| \leq \| T.Y - X \| \ , \ \forall T \in M_{n\times m}(\mathbb{R}) \ .$$

O próximo teorema estabelece uma forma explícita para a matriz K.

Teorema 3.9. *Se $E(Y.Y^t)$ é invertível, então o estimador linear de mínima variância $\widehat{X}$ de X, baseado em Y, é dado por*

$$\widehat{X} = E(X.Y^t).E(Y.Y^t)^{-1}.Y \ .$$

Além disso, se X e Y possuem esperanças nulas, então a matriz covariância do erro $\widehat{X} - X$ é dada por

$$E([\widehat{X} - X].[\widehat{X} - X]^t) = E(X.X^t) - E(X.Y^t).E(Y.Y^t)^{-1}.E(Y.X^t) \ .$$

Prova: Sejam $K \in M_{n\times m}(\mathbb{R})$ e $K_i^t = \begin{bmatrix} k_{i1} & k_{i2} & \ldots & k_{im} \end{bmatrix}$ a linha i da matriz K. Se

$$\widehat{X} = \begin{bmatrix} \widehat{X}_1 \\ \vdots \\ \widehat{X}_n \end{bmatrix} = K.Y \ ,$$

então $\widehat{X}_i = K_i^t.Y$, isto é, $\widehat{X}_i$ pertence ao subespaço gerado $V = [Y_1, ..., Y_m]$. Assim, da igualdade

$$\|K.Y - X\|^2 = \sum_{i=1}^{m} \|K_i^t.Y - X_i\|_2^2 \ ,$$

segue que $\widehat{X}_i = proj_V(X_i)$. Da Proposição 1.11, temos que $\widehat{X}_i = K_i^t.Y$, onde

$$G(Y_1, ..., Y_m).K_i = \begin{bmatrix} < X_i, Y_1 > \\ < X_i, Y_2 > \\ \vdots \\ < X_i, Y_m > \end{bmatrix} = \begin{bmatrix} E(X_i.Y_1) \\ E(X_i.Y_2) \\ \vdots \\ E(X_i.Y_m) \end{bmatrix} \ .$$

Como a matriz de Gram é dada por $G(Y_1, ..., Y_m) = E(Y.Y^t)$, temos que $E(Y.Y^t).K^t = E(Y.X^t)$ e portanto

$$\widehat{X} = E(X.Y^t).E(Y.Y^t)^{-1}.Y \quad .$$

Por substituição direta, obtemos a expressão

$$E([\widehat{X} - X].[\widehat{X} - X]^t) = E(X.X^t) - E(X.Y^t).E(Y.Y^t)^{-1}.E(Y.X^t) \quad .$$

Por outro lado, se $E(X)$ e $E(Y)$ são nulos, então

$$Cov(\widehat{X} - X, \widehat{X} - X) = E([\widehat{X} - X].[\widehat{X} - X]^t)$$

e portanto segue o resultado. $\blacksquare$

Observação: a hipótese $E(Y.Y^t)$ ser invertível do teorema anterior poderia ser evitada, mas nêsse caso deveria-se introduzir o conceito de pseudo-inverso de uma matriz.

Corolário 3.10. *Suponha que*

$$Y = W.X + \varepsilon \ ,$$

onde Y e ε são m-vetores aleatórios, X é n-vetor aleatório e $W \in M_{m\times n}(\mathbb{R})$ é uma matriz constante. Sejam F e D matrizes definidas por

$$E(X.X^t) = F$$
$$E(\varepsilon.\varepsilon^t) = D$$

Se $E(\varepsilon.X^t) = 0$ e $(W.F.W^t + D)$ é invertível, então o estimador linear de mínima variância $\widehat{X}$ de X, baseado em Y, é dado por

$$\widehat{X} = F.W^t.(W.F.W^t + D)^{-1}.Y \ .$$

Se X e Y possuem esperanças nulas, então a covariância do erro $\widehat{X} - X$ é dada por

$$E([\widehat{X} - X].[\widehat{X} - X]^t) = F - F.W^t.(W.F.W^t + D)^{-1}.W.F \ .$$

Prova: Computando $E(Y.Y^t) = W.F.W^t + D$ e $E(X.Y^t) = F.W^t$, temos pelo teorema anterior

$$\widehat{X} = F.W^t.(W.F.W^t + D)^{-1}.Y$$

e

$$E([\widehat{X} - X].[\widehat{X} - X]^t) = F - F.W^t.(W.F.W^t + D)^{-1}.W.F \ .$$

Como visto antes, se X e Y possuem esperanças nulas, então

$$Cov(\widehat{X} - X, \widehat{X} - X) = E([\widehat{X} - X].[\widehat{X} - X]^t). \qquad \blacksquare$$

Teorema 3.11. *Sejam Y e X m-vetor e n-vetor aleatórios respectivamente. Suponha que $E(Y.Y^t)$ é invertível e denote por $\widehat{X}$ o estimador linear de mínima variância de X baseado em Y.*

a) *Se $T \in M_{p\times n}(\mathbb{R})$ é uma matriz constante, então $T.\widehat{X}$ é o estimador linear de mínima variância de $T.X$ baseado em Y.*

b) *Se $P \in M_n(\mathbb{R})$ é uma matriz positiva definida, então o estimador $\widehat{X}$ é o estimador linear Z que minimiza*

$$E([Z - X]^t.P.[Z - X])$$

Prova: a) Do Teorema 3.9, temos que o estimador linear $\widehat{T.X}$ de $T.X$ é dado por

$$\widehat{T.X} = E(T.X.Y^t).E(Y.Y^t)^{-1} = T.E(X.Y^t).E(Y.Y^t)^{-1} = T.\widehat{X}$$

b) Como P é uma matriz positiva definida, existe $A \in M_n(\mathbb{R})$ tal que $A^t.A = P$. Da igualdade

$$E([Z - X]^t.P.[Z - X]) = E([A.Z - A.X]^t.[A.Z - A.X]) = \|A.Z - A.X\|^2$$

e do ítem a), temos que $\widehat{X}$ é o estimador linear Z que minimiza a expressão acima. ■

Teorema 3.12. *Sejam $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ um n-vetor aleatório e V o subespaço gerado por $Y_1, ..., Y_n$.*

Dado um subespaço fechado U de $L^2(\Omega, \Lambda, P)$, defina $\tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_n \end{bmatrix}$, de modo que

$$\tilde{Y}_i = Y_i - proj_U(Y_i) \ , \ \forall i \in \{1, ..., n\} \ .$$

Suponha que $U + V$ é fechado. Se X é variável aleatória, então a sua projeção sobre $U + V$ é

$$\widehat{X} = proj_U(X) + E(X.\tilde{Y}^t).E(\tilde{Y}.\tilde{Y}^t)^{-1}.\tilde{Y} \ .$$

Prova: Vamos assumir que $A = \{\tilde{Y}_1, ..., \tilde{Y}_n\}$ é um conjunto linearmente independente , pois usaremos o fato que $E(\tilde{Y}.\tilde{Y}^t)^{-1}$ existe. Caso contrário, escolhemos um subconjunto $\{\tilde{Y}_{i_1}, ..., \tilde{Y}_{i_k}\}$ de A, de modo que ele seja uma base para o subespaço gerado $[\tilde{Y}_1, ..., \tilde{Y}_n]$ e definimos

$$\tilde{Y} = \begin{bmatrix} \tilde{Y}_{i_1} \\ \vdots \\ \tilde{Y}_{i_k} \end{bmatrix}$$

Suponhamos então que $\{\tilde{Y}_1, ..., \tilde{Y}_n\}$ seja um conjunto linearmente independente. Pois bem, seja T o subespaço gerado pelas coordenadas de $\tilde{Y}$. Observando que $proj_T(X)$ é o estimador linear de mínima variância de X baseado em $\tilde{Y}$, em consequência do Teorema 3.9 temos a igualdade:

$$proj_T(X) = E(X.\tilde{Y}^t).E(\tilde{Y}.\tilde{Y}^t)^{-1}.\tilde{Y} \ .$$

Uma vez que $T \perp U$ e $T \oplus U = V + U$, temos pela Proposição 1.8 a igualdade

$$\tilde{X} = proj_U(X) + proj_T(X) = proj_U(X) + E(X.\tilde{Y}^t).E(\tilde{Y}.\tilde{Y}^t)^{-1}.\tilde{Y} \ . \qquad \blacksquare$$

Observação: o teorema anterior será aplicado com as seguintes hipóteses. Suponhamos que num primeiro passo, obtemos X' um estimador linear de mínima variância de $X = (X_1, ..., X_n)$ baseado no vetor $u = (u_1, ..., u_k)$. Num segundo passo, dado o vetor $Y = (Y_1, ..., Y_m)$, procuramos um estimador X'' de X baseado no vetor

$$(u_1, ..., u_k, Y_1, ..., Y_m) \ .$$

Nesse sentido, dizemos que X'' é um estimador baseado nos dados "antigos (u) e novos (Y) " .

Sejam $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}$, V o subespaço gerado por $Y_1, ..., Y_m$ e U o subespaço gerado por $u_1, ..., u_k$.

Como o estimador $X'' = \begin{bmatrix} X''_1 \\ \vdots \\ X''_n \end{bmatrix}$ é tal que X''_i é a projeção de X_i sobre $U + V$, temos pelo último teorema:

$$X''_i = X'_i + E(X_i.\tilde{Y}^t).E(\tilde{Y}.\tilde{Y}^t)^{-1}.\tilde{Y} \ ,$$

onde a i-ésima coordenada de $\tilde{Y}$ é $\tilde{Y}_i = Y_i - proj_U(Y_i)$. Ou seja,

$$X'' = X' + E(X.\tilde{Y}^t).E(\tilde{Y}.\tilde{Y}^t)^{-1}.\tilde{Y} \ .$$

Observe que o vetor Y menos o seu estimador linear, baseado em u, resulta no vetor $\tilde{Y}$.

Conforme notação introduzida na observação acima, temos um próximo teorema na forma:

Teorema 3.13. *Seja $R = E([X - X'].[X - X']^t)$ e suponha que Y (dados novos) tenha a forma*

$$Y = W.X + \varepsilon \ ,$$

onde ε é um vetor aleatório de esperança nula não correlacionado com u e X. Então

$$X'' = X' + R.W^t.(W.R.W^t + Q)^{-1}.(Y - W.X') \ ,$$

sendo $Q = E(\varepsilon.\varepsilon^t)$. Além disso,

$$E([X - X''].[X - X'']^t) = R - R.W^t.(W.R.W^t + Q)^{-1}.W.R \ .$$

Prova: Das propriedades de Covariância, temos que

$$Cov(\varepsilon, u) = E(\varepsilon.u^t) - E(\varepsilon).E(u^t) = E(\varepsilon.u^t) \ .$$

Como ε e u são não correlacionados, temos que $E(\varepsilon.u^t) = 0$. Essa condição nos diz que $\varepsilon_i \perp U$ para todo i. Assim, denotando por W_i a i-ésima linha de W, temos

$$proj_U(Y_i) = proj_U(W_i.X) + proj_U(\varepsilon_i) = proj_U(W_i.X) \ ,$$

ou seja, o estimador linear de mínima variância de Y, baseado em u, é o estimador linear de $W.X$ baseado em u. Assim, pelo Teorema 3.11 e pela definição de $\tilde{Y}$ segue que

$$\tilde{Y} = Y - W.X' \ .$$

Uma vez que $E(X'.\tilde{Y}^t) = 0$, a última observação nos diz que

$$X'' \ = \ X' + E([X - X'].\tilde{Y}^t).[E(\tilde{Y}.\tilde{Y}^t)]^{-1}.\tilde{Y} \ .$$

Denotando por e o erro $X - X'$, segue do Teorema 3.9

$$X'' = X' + \widehat{e} \ ,$$

onde $\widehat{e}$ é o estimador de e baseado em $\tilde{Y}$. Uma vez que

$$\tilde{Y} \ = \ Y - W.X' \ = \ W.X + \varepsilon - W.X' \ = W.e + \varepsilon \ ,$$

como $E(\varepsilon.e^t) = 0$ e $R = E([X - X'].[X - X']^t)$, o Corolário 3.10 implica em

$$X'' = X' + R.W^t.(W.R.W^t + Q)^{-1}.(Y - W.X')$$

e

$$E([X - X''].[X - X'']^t) = E([e - \widehat{e}].[e - \widehat{e}]^t) = R - R.W^t.(W.R.W^t + Q)^{-1}.W.R \ . \qquad \blacksquare$$

3.4 Teorema de Kalman

Seja $(u(k))_{k\in\mathbb{N}}$ uma sequência de n-vetores aleatórios que satisfazem :

$$E(u(k)) = 0 \ , \ \forall k \in \mathbb{N} \ ,$$
$$E(u(k).u(l)^t) = Q(k).\delta_{kl} \ , \ \forall k, l \in \mathbb{N} \ ,$$

onde δ_{kl} é o símbolo de Kronecker.

Defina a sequência $(x(k))_{k\in\mathbb{N}}$ pela equação :

$$x(k+1) = \Phi(k).x(k) + u(k) \ ,$$

sendo $\Phi(k) \in M_n(\mathbb{R})$ uma matriz constante para cada k e $x(0)$ um vetor aleatório inicial com esperança nula.

Dada uma sequência de vetores aleatórios $(v(k))_{k\in\mathbb{N}}$, vamos encontrar um método para estimar $x(k)$, baseado nos dados $v(0), v(1), ..., v(k-1)$. Para isso, vamos fazer algumas considerações:

$$v(k) = M(k).x(k) + w(k) \ , \ \forall k \in \mathbb{N} \ ,$$

onde $M(k) \in M_{m\times n}(\mathbb{R})$ é matriz constante para cada $k \in \mathbb{N}$ e $w(k)$ é um n-vetor aleatório que satisfaz:

$$E(w(k)) = 0 \ , \ \forall k \in \mathbb{N}$$
$$E(w(k).w(l)^t) = R(k).\delta_{kl} \ , \ \forall k, l \in \mathbb{N} \ ,$$

onde $R(k)$ é positiva definida. Assuma que $x(0), u(k)$ e $w(l)$ são dois a dois não correlacionados para todo $k, l \in \mathbb{N}$.

No que segue, denotaremos o estimador linear de mínima variância de $x(k)$, baseado em $v(1),...,v(j)$, por $\widehat{x}(k|j)$. Assim, se $V(j)$ denota o subespaço gerado pelas coordenadas de $v(1),...,v(j)$, então a i-ésima coordenada de $\widehat{x}(k|j)$ é a projeção da i-ésima coordenada de $x(k)$ sobre o subespaço $V(j)$.

Teorema de Kalman (Solução para o problema de estimação recursiva)
O estimador linear $\widehat{x}(k+1|k)$ é gerado recursivamente de acordo com a equação

$$(1) \quad \widehat{x}(k+1|k) = \Phi(k).P(k).M(k)^t[M(k).P(k).M(k)^t + R(k)]^{-1}.[v(k) - M(k).\widehat{x}(k|k-1)]+$$
$$+\Phi(k).\widehat{x}(k|k-1) \quad ,$$

onde $P(k) = E([x(k) - \widehat{x}(k|k-1)].[x(k) - \widehat{x}(k|k-1)]^t)$ é gerado recursivamente por

$$(2) P(k+1) = \Phi(k).P(k).\{Id - M(k)^t.[M(k).P(k).M(k)^t + R(k)]^{-1}.M(k).P(k)\}.\Phi(k)^t + Q(k).$$

As condições iniciais para essas equações são o estimador inicial $\widehat{x}(0|-1) = \widehat{x}(0)$ e $P(0)$.

Prova: Tomando como exemplo as duas expressões abaixo :

$$x(2) = \Phi(1).\Phi(0).\boldsymbol{x}(0) + \Phi(1).\boldsymbol{u}(0) + \boldsymbol{u}(1) \quad ,$$

$$x(3) = \Phi(2).\Phi(1).\Phi(0).\boldsymbol{x}(0) + \Phi(2).\Phi(1).\boldsymbol{u}(0) + \Phi(2)\boldsymbol{u}(1) + u(2) \ ,$$

é fácil ver que $x(i)$ é combinação de $u(i-1), u(i-2), ..., u(0)$ e $x(0)$, ou seja, fixado $i \in \mathbb{N}$, existem matrizes $A_0, A_1, ..., A_i$ de $M_n(\mathbb{R})$ tais que

$$x(i) = A_i.x(0) + \sum_{j=0}^{i-1} A_j.u(j) \ .$$

Por propriedades de covariância, se $k \geq i$, então

$$Cov(x(i), u(k)) = Cov(A_i.x(0), u(k)) + \sum_{j=0}^{i-1} Cov(A_j.u(j), u(k)) =$$

$$= A_i.Cov(x(0), u(k)) + \sum_{j=0}^{i-1} A_j.E(u(j).u(k)^t) = 0 \ ,$$

ou seja, se $k \geq i$, então $x(i)$ e $u(k)$ são não correlacionados. Agora, como $v(i) = M(i).x(i) + w(i)$, temos que

$$Cov(v(i), u(k)) = M(i).Cov(x(i), u(k)) + Cov(w(i), u(k)) = 0 \ , \ se \ k \geq i \ ,$$

ou seja, $v(i)$ e $u(k)$ são não correlacionados. Essa condição nos diz: cada cordenada de $u(k)$ é ortogonal ao subespaço $V(i)$ se $k \geq i$. Assim, o estimador linear de mínima variância de $x(k+1)$, baseado em $v(1), ..., v(k)$, é o estimador linear de mínima variância de $\Phi(k).x(k)$ baseado em $v(1), ..., v(k)$.

Pelo Teorema 3.11, temos

$$\widehat{x}(k+1|k) = \Phi(k).\widehat{x}(k|k) \ .$$

Supondo que $\widehat{x}(k|k-1)$ e $P(k)$ foram computados, temos pelo Teorema 3.13:

$$\widehat{x}(k|k) = \widehat{x}(k|k-1) + P(k).M(k)^t[M(k).P(k).M(k)^t + R(k)]^{-1}.[v(k) - M(k).\widehat{x}(k|k-1)].$$

Observe que ao aplicarmos o Teorema 3.13, estamos considerando $v(1), ..., v(k-1)$ o conjunto de dados "antigos "e $v(k)$ o vetor de dados "novos".

Das duas expressões acima, segue a equação (1) do teorema. Agora, deduziremos $P(k)$ recursivamente: uma vez que

$$x(k+1) - \widehat{x}(k+1|k) = \Phi(k).[x(k) - \widehat{x}(k|k)] + u(k) \ ,$$

segue que

$$P(k+1) = \Phi(k).E([x(k) - \widehat{x}(k|k)].[x(k) - \widehat{x}(k|k)]^t).\Phi(k)^t + Q(k) \ .$$

Pelo Teorema 3.13, temos

$$E([x(k) - \widehat{x}(k|k)].[x(k) - \widehat{x}(k|k)]^t) = P(k) - P(k).M(k)^t.[M(k).P(k).M(k)^t + R(k)]^{-1}.M(k).P(k) \ .$$

Das duas últimas expressões segue a equação (2) do teorema. ∎

Apêndice A

Algumas Aplicações Práticas do Filtro de Kalman

Em 1960, Rudolf Emil Kalman publicou o seu famoso artigo descrevendo uma solução recursiva para o problema da filtragem linear de dados discretos. Desde então, devido principalmente aos avanços da computação digital, o Filtro de Kalman tem sido uma ferramenta de muita pesquisa e desenvolvimento. A filtragem de Kalman vem sendo aplicada em áreas tão diversas quanto a aeroespacial, navegação marítima, instrumentação de usinas nucleares, modelamento demográfico, astronomia, economia e indústrias em geral.

A primeira aplicação prática para o Filtro de Kalman foi encontrada por Stanley F. Shimidt que trabalhava no projeto Apollo da NASA cujo objetivo era levar uma nave á Lua e traze-la de volta a Terra. No momento ele tinha problemas na estimação de trajetórias e controle. Shimidt trabalhou no que seria a primeira implementação completa do Filtro de Kalman e tornou o mesmo parte integrante do sistema de controle da Apollo. Também por influência de Shimidt, o Filtro de Kalman foi incluído no sistema de navegação do cargueiro aéreo C5A. Neste caso, o Filtro de Kalman resolveu o problema da fusão sensorial, quando combinou dados de radar com aqueles provenientes de sensores inerciais para estimar a trajetória do avião. Desde então o Filtro de Kalman vem sendo parte integrante da maioria dos sistemas *onboard* de estimação de trajetória e controle em aeronaves.

O Filtro de Kalman é considerado por muitos o grande avanço da teoria de estimação do século vinte. Muitas realizações desde sua introdução talvez não fossem possíveis sem ele. As principais aplicações da filtragem de Kalman estão nos sistemas de controle modernos e na navegação e rastreamento de todos os tipos de veículos.

Mas o que é o Filtro de Kalman? Teoricamente, ele é um estimador para aquilo que é chamado o "problema Gaussiano-linear-quadrático", que é o problema da estimação dos estados instantâneos de um sistema linear dinâmico perturbado por ruído Gaussiano branco usando-se medições linearmente relacionadas aos estados e também corrompidas por ruído branco. Na prática, o Filtro de Kalman é um conjunto de equações matemáticas que provê uma solução computacional eficiente para o método dos mínimos quadrados. O filtro é muito poderoso pois permite a estimação dos estados passados, presentes e futuros de um sistema sendo que para isso não é necessário um conhecimento preciso de seu modelo.

Referências Bibliográficas

[1] Bartle, R.G. *Elements of Integration*, John Wiley Sons Iwc, 1966.

[2] Bussab, W.O. e Morettin, P.A. *Estatística Básica*, Atual Editora, 1985.

[3] Dudley, R.M. *Real Analysis and Probability*, Wadsworth and Brooks / Cole Advanced Books-California,1989.

[4] Honig,C.S. *Análise Funcional e Aplicações ,vol. 1*, IME-USP, São Paulo, 1970.

[5] Luenberger, D.G. *Optimization by Vector Space Methods*, John Wiley,1969.

[6] Pereira, G. A. S.- *Filtro de Kalman: Teoria e Aplicações*, Centro de Pesquisa e Desenvolvimento em Engenharia Elétrica-U.F.M.G