



JULIANNA PINELE SANTOS PORTO

**Geometria da Informação:  
Métrica de Fisher**

CAMPINAS  
2013





**Universidade Estadual de Campinas**

Instituto de Matemática, Estatística  
e Computação Científica

**Julianna Pinele Santos Porto**

**Geometria da Informação:  
Métrica de Fisher**

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em matemática aplicada.

**Orientador: João Eloir Strapassoon**

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELA ALUNA JULIANNA PINELE SANTOS PORTO, E ORIENTADA PELO PROF. DR. JOÃO ELOIR STRAPASSOON.

**Assinatura do Orientador**

A handwritten signature in blue ink, reading "João E Strapassoon", is written over a horizontal line.

**Campinas  
2013**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Maria Fabiana Bezerra Muller - CRB 8/6162

P838g Porto, Julianna Pinele Santos, 1990-  
Geometria da informação : métrica de Fisher / Julianna Pinele Santos Porto. –  
Campinas, SP : [s.n.], 2013.

Orientador: João Eloir Strapasson.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de  
Matemática, Estatística e Computação Científica.

1. Geometria da informação. 2. Matriz de informação de Fisher. 3. Distância de  
Rao. 4. Geometria diferencial. 5. Estatística matemática. I. Strapasson, João  
Eloir, 1979-. II. Universidade Estadual de Campinas. Instituto de Matemática,  
Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Information geometry : Fisher's metric

**Palavras-chave em inglês:**

Information geometry

Fisher information metric

Rao distance

Differential geometry

Mathematical statistics

**Área de concentração:** Matemática Aplicada

**Titulação:** Mestra em Matemática Aplicada

**Banca examinadora:**

João Eloir Strapasson [Orientador]

Sandra Augusta Santos

Edson Agustini

**Data de defesa:** 23-08-2013

**Programa de Pós-Graduação:** Matemática Aplicada

**Dissertação de Mestrado defendida em 23 de agosto de 2013 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**

*João E Strapasson*

\_\_\_\_\_  
**Prof.(a). Dr(a). JOÃO ELOIR STRAPASSON**

*Sandra Augusta Santos*

\_\_\_\_\_  
**Prof.(a). Dr(a). SANDRA AUGUSTA SANTOS**

*Edson Agustini*

\_\_\_\_\_  
**Prof.(a). Dr(a). EDSON AGUSTINI**

## Abstract

Information Geometry is an area of mathematics that uses geometric tools in the study of statistical models. In 1945, Rao introduced a Riemannian metric on the space of the probability distributions using the information matrix provided by Ronald Fisher in 1921. With the metric associated with this matrix, we define a distance between two probability distributions (Rao's distance), geodesics, curvatures and other properties. Since then, many authors have been studying this subject, which is associated with various applications, such as: statistical inference, stochastic processes, information theory, and image distortion.

In this work we provide a brief introduction to Differential and Riemannian Geometry and a survey of some results obtained in Information Geometry. We show Rao's distance between some probability distributions, with special attention to the study of such distance in the space of multivariate normal distributions. In this space, since closed forms for the distance and for the geodesic curve are not known yet, we focus on the calculus of bounds for Rao's distance. In some cases, we improve the upper bound provided by Calvo and Oller in 1990.

**Keywords:** Information geometry, Fisher Information Metric, Rao's distance.

## Resumo

A Geometria da Informação é uma área da matemática que utiliza ferramentas geométricas no estudo de modelos estatísticos. Em 1945, Rao introduziu uma métrica Riemanniana no espaço das distribuições de probabilidade usando a matriz de informação, dada por Ronald Fisher em 1921. Com a métrica associada a essa matriz, define-se uma distância entre duas distribuições de probabilidade (distância de Rao), geodésicas, curvaturas e outras propriedades do espaço. Desde então muitos autores veem estudando esse assunto, que está naturalmente ligado a diversas aplicações como, por exemplo, inferência estatística, processos estocásticos, teoria da informação e distorção de imagens.

Neste trabalho damos uma breve introdução à geometria diferencial e Riemanniana e fazemos uma coletânea de alguns resultados obtidos na área de Geometria da Informação. Mostramos a distância de Rao entre algumas distribuições de probabilidade e damos uma atenção especial ao estudo da distância no espaço formado por distribuições Normais Multivariadas. Neste espaço, como ainda não é conhecida uma fórmula fechada para a distância e nem para a curva geodésica, damos ênfase ao cálculo de limitantes para a distância de Rao. Conseguimos melhorar, em alguns casos, o limitante superior dado por Calvo e Oller em 1990.

**Palavras-chave:** Geometria da da Informação, Matriz de Informação de Fisher, Distância de Rao, Limitantes.

# Sumário

Dedicatória	viii
Agradecimentos	ix
Introdução	1
<b>1 Preliminares em Geometria Diferencial</b>	<b>2</b>
1.1 Superfícies regulares . . . . .	2
1.2 Variedade Diferenciável . . . . .	13
1.3 Variedades Riemannianas . . . . .	16
<b>2 Métrica de Fisher</b>	<b>21</b>
2.1 Conceitos Básicos de Estatística . . . . .	21
2.2 A Métrica de Fisher . . . . .	25
2.3 Distância entre Distribuições de Probabilidade . . . . .	31
2.3.1 Distribuições com apenas um parâmetro . . . . .	31
2.3.2 Distribuições Normais Univariadas . . . . .	34
2.3.3 Distribuição Multinomial . . . . .	38
<b>3 Distribuições Normais Multivariadas</b>	<b>40</b>
3.1 Métrica de Fisher . . . . .	40
3.1.1 A matriz de covariância é uma matriz diagonal . . . . .	43
3.1.2 A matriz de covariância é constante. . . . .	46
3.1.3 O vetor das médias é constante. . . . .	46
3.2 Limitantes . . . . .	47
3.2.1 Limitante Inferior . . . . .	47
3.2.2 Limitante Superior . . . . .	49
3.3 Simulações . . . . .	51
<b>A Valores do Parâmetros</b>	<b>64</b>

*À minha família.*

# Agradecimentos

Em primeiro lugar agradeço à minha mãe e ao meu irmão Sandro por sempre me apoiarem em todas as minhas decisões, incentivarem meus estudos e por se fazerem presentes mesmo estando longe.

Agradeço a toda à minha família por sempre torcerem por mim, por me receberem sempre de braços abertos e por compartilharem comigo todas as glórias e dificuldades.

Agradeço meu orientador João Strapasson por todo apoio, paciência e dedicação durante esses dois anos de mestrado.

Ao Professor Aurélio de Oliveira por toda a atenção, paciência e por ser uma pessoa com quem eu sempre pude contar desde quando cheguei no IMECC.

Agradeço aos meus amigos, aos amigos mais antigos que sempre estiveram presentes na minha vida mesmo com a distância e ao amigos que eu fiz aqui, Campello, Karla, João, Larissa e Paula, sem eles superar as dificuldades de um curso do mestrado e de uma nova rotina longe da família não seria possível. Agradeço também a todos os outros amigos e colegas, que não estão presentes a todo tempo mas que fazem o meu dia-a-dia mais leve.

À FAEPEX (Fundo de Apoio ao Ensino, à Pesquisa e à Extensão) pelo apoio financeiro dos meus estudos.

# Introdução

A Geometria da Informação estuda o uso de ferramentas geométricas no estudo de modelos estatísticos (famílias de distribuições de probabilidade). Foi Rao em 1945, [20], o pioneiro nesta área. Ele introduziu uma métrica Riemanniana no espaço das famílias de distribuições de probabilidade parametrizadas usando a matriz de informação, dada por Ronald Fisher em 1921, [7].

Com a métrica associada a essa matriz (métrica de Fisher), podemos calcular a distância entre duas distribuições de probabilidade (também conhecida como distância de Rao ou distância de Fisher), geodésicas, curvaturas e outras propriedades do espaço. Desde então muitos autores veem estudando esse assunto, que está naturalmente ligado a diversas aplicações como, por exemplo, inferência estatística, processos estocásticos, teoria da informação e distorção de imagens.

Amari e Nagaoka deram uma unificação à teoria em [1] e [2] e, simultaneamente, num trabalho isolado, Chentsov, [10], desenvolveu um novo conceito de variedade estatística também usando a métrica de Fisher. Atkinson e Michell, [3], e Burbea, [4], fizeram um estudo detalhado e calcularam a expressão da distância de Rao para algumas famílias de distribuições de probabilidade.

Neste trabalho, fazemos um "apanhado" dos resultados já obtidos nessa área e damos ênfase ao estudo da distância de Rao no espaço das distribuições normais multivariadas. Nesse espaço ainda não se tem uma expressão para a distância de Rao entre duas distribuições no caso geral. No início da década de 90, Calvo e Oller calcularam limitantes para essa distância em [5] e [6].

A base da teoria é a utilização da geometria no espaço formado por distribuições de probabilidade. No Capítulo 1, abordamos alguns conceitos fundamentais de geometria diferencial e geometria Riemanniana tais como: variedades diferenciáveis, espaço tangente, geodésicas e métrica Riemanniana. Esses conceitos são ferramentas essenciais para o desenvolvimento deste trabalho.

No Capítulo 2, damos algumas definições básicas de estatística e também definimos variedade estatística, que nada mais é do que um modelo estatístico munido de uma estrutura de variedade diferenciável. Além disso, damos a definição formal de métrica de Fisher e calculamos a distância de Rao para algumas famílias de distribuições de probabilidade.

Finalmente, no Capítulo 3, nosso estudo se concentra nas famílias formadas pelas distribuições normais multivariadas, onde mostramos como calcular a distância de Rao em alguns casos particulares. Também, calculamos alguns limitantes para essa distância no caso geral e fazemos algumas simulações para analisar o comportamento desses limitantes para famílias de distribuições normais bivariadas.

# Capítulo 1

## Preliminares em Geometria Diferencial

Neste capítulo apresentamos algumas definições básicas de geometria diferencial e Riemanniana. Para se ter uma noção de como é feito o estudo de espaços que não são necessariamente o  $\mathbb{R}^n$ , na Seção 1, falamos sobre superfícies regulares, que são subconjuntos do  $\mathbb{R}^3$ . Na Seção 2, generalizamos o conceito de superfícies, estudando agora um conjunto qualquer, para isso, damos a definição de variedade diferenciável. Na Seção 3, introduzimos um produto interno numa variedade e, a essa variedade, chamamos de variedade Riemanniana. O objetivo deste capítulo é fazer uma breve introdução, logo, os resultados não serão demonstrados. Para um estudo mais aprofundado do assunto e para ver as demonstrações dos teoremas e proposições apresentadas neste capítulo, ver referências [8] e [9].

### 1.1 Superfícies regulares

De forma intuitiva, podemos dizer que uma superfície regular é um subconjunto do  $\mathbb{R}^3$  que, em qualquer ponto, é localmente semelhante ao  $\mathbb{R}^2$ , tal como a superfície da Terra que, embora esférica, parece plana a um observador nela colocado que consegue ver somente até a linha do horizonte.

**Definição 1.1.1.** Um subconjunto  $S$  do  $\mathbb{R}^3$  é uma superfície regular quando, para cada ponto  $p \in S$ , existe uma vizinhança  $V$  de  $p$  em  $\mathbb{R}^3$  e uma aplicação  $\varphi : U \rightarrow V \cap S$  de um aberto  $U \subset \mathbb{R}^2$  sobre  $V \cap S \subset \mathbb{R}^3$  tais que

1.  $\varphi$  é diferenciável. Isto significa que se escrevemos

$$\varphi(x, y) = (\varphi_1(x, y), \varphi_2(x, y), \varphi_3(x, y)), \quad (x, y) \in U$$

as funções  $\varphi_1, \varphi_2, \varphi_3$  tem derivadas parciais contínuas de todas as ordens em  $U$ .

2.  $\varphi$  é um homeomorfismo. Como  $\varphi$  é contínua pela condição 1, basta verificar que  $\varphi$  tem inversa  $\varphi^{-1} : V \cap S \rightarrow U$  que é contínua.
3. Para todo  $q \in U$ , a diferencial  $d\varphi_q : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  é injetiva.

A aplicação  $\varphi$  é chamada de sistema de coordenadas (ou parametrização) em  $p$ . A vizinhança  $V \cap S$  de  $p$  é chamada vizinhança coordenada.

**Observação 1.1.2.** Neste trabalho, dizer que uma aplicação  $f$  é diferenciável significa dizer que  $f$  é  $C^\infty$ .

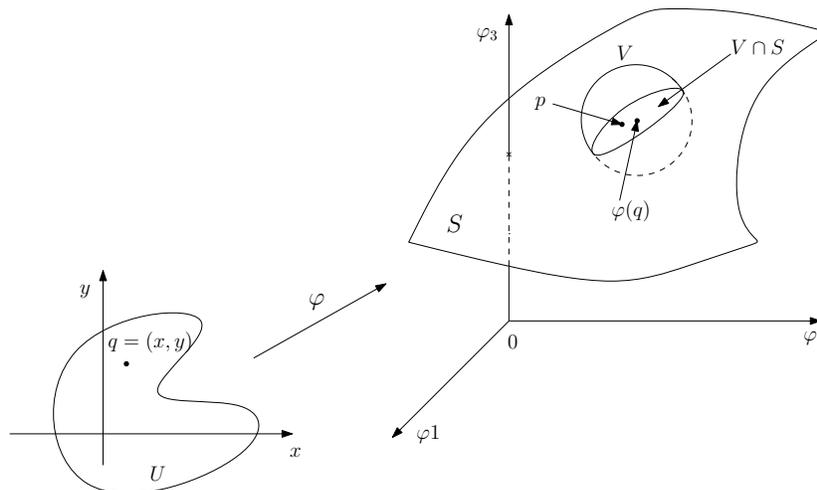


Figura 1.1: Definição de superfície regular.

**Observação 1.1.3.** Em relação à injetividade de  $d\varphi_q : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ , dado  $q \in U$ , lembremos que as seguintes condições são equivalentes:

- $d\varphi_q : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  é injetiva.
- $\frac{\partial \varphi}{\partial x}(q) = \varphi_x(q) = d\varphi_q(e_1)$  e  $\frac{\partial \varphi}{\partial y}(q) = \varphi_y(q) = d\varphi_q(e_2)$  são vetores linearmente independentes.
- A matriz jacobiana de ordem  $3 \times 2$ ,

$$J\varphi_q = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x} & \frac{\partial \varphi_1}{\partial y} \\ \frac{\partial \varphi_2}{\partial x} & \frac{\partial \varphi_2}{\partial y} \\ \frac{\partial \varphi_3}{\partial x} & \frac{\partial \varphi_3}{\partial y} \end{pmatrix} (q),$$

tem posto 2, isto é, algum de seus determinantes menores  $2 \times 2$  é diferente de zero.

Da Definição 1.1.1, temos que a injetividade na condição 2 tem o objetivo de excluir a possibilidades de auto-interseções em superfícies regulares e a condição 3 garante a existência de um "plano tangente" em todos os pontos de  $S$ .

**Exemplo 1.1.4.** Um plano  $\Pi \subset \mathbb{R}^3$  é uma superfície regular. Nesse caso, podemos parametrizar o plano com uma parametrização global, isto é, com uma única parametrização para todo ponto  $p \in \Pi$ . De fato, seja um ponto  $P \in \Pi$  e sejam  $\mathbf{u} = (u_1, u_2, u_3)$  e  $\mathbf{v} = (v_1, v_2, v_3)$  vetores do plano, não nulos, linearmente independentes. Sem perda de generalidade podemos assumir  $\mathbf{u}$  e  $\mathbf{v}$

perpendiculares entre si. Para cada ponto  $Q$  do plano, o seu respectivo vetor de posição  $\mathbf{w}$  é uma combinação linear de  $\mathbf{u}$  e  $\mathbf{v}$ , seja  $\mathbf{w} = x\mathbf{u} + y\mathbf{v}$ , com  $x, y \in \mathbb{R}$ . Portanto

$$Q - P = \mathbf{w} = x\mathbf{u} + y\mathbf{v} \Rightarrow Q = P + x\mathbf{u} + y\mathbf{v}.$$

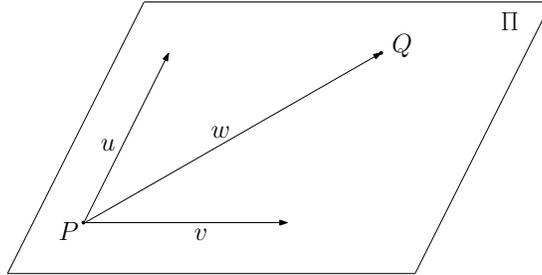


Figura 1.2: Parametrização do plano.

Assim, podemos parametrizar o plano através da seguinte aplicação

$$\begin{aligned} \varphi : \mathbb{R}^2 &\rightarrow \Pi \\ (x, y) &\rightarrow P + x\mathbf{u} + y\mathbf{v} \end{aligned}$$

É fácil ver que  $\varphi$  é diferenciável. Para verificar a condição 2, note que  $\varphi^{-1}$  é dada por

$$\varphi^{-1}(Q) = \left( \frac{\langle Q - P, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}, \frac{\langle Q - P, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \right),$$

onde  $\langle, \rangle$  é o produto interno usual do  $\mathbb{R}^3$ . Além disso, temos que  $\varphi_x(q) = \mathbf{u}$  e  $\varphi_y(q) = \mathbf{v}$ ,  $\forall q \in \mathbb{R}^2$ , e como  $\mathbf{u}$  e  $\mathbf{v}$  são linearmente independentes, pela Observação 1.1.3,  $d\varphi_q$  é injetiva, para todo  $q \in \mathbb{R}^2$ .

Dessa forma as condições de 1 a 3 da Definição 1.1.1 estão satisfeitas.

**Exemplo 1.1.5.** A esfera unitária  $S^2 = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 + z^2 = 1\}$  é uma superfície regular.

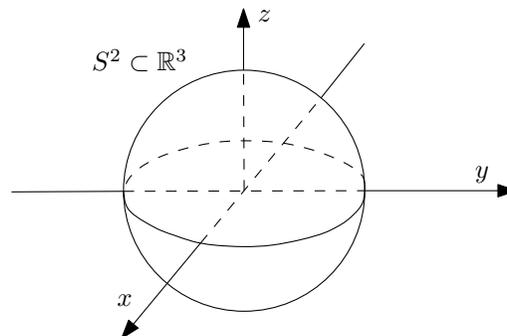


Figura 1.3: Esfera unitária  $S^2$ .

De fato, considere a aplicação  $\varphi_1 : U \rightarrow \mathbb{R}^3$  dada por

$$\varphi_1(x, y) = (x, y, \sqrt{1 - (x^2 + y^2)}),$$

onde  $U = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 < 1\}$ . Observe que  $\varphi_1(U)$  é a parte (aberta) de  $S^2$  acima do plano  $xy$ .

Vamos verificar que  $\varphi_1$  satisfaz as condições da Definição 1.1.1. Como  $x^2 + y^2 < 1$ , a função  $\sqrt{1 - (x^2 + y^2)}$  tem derivadas parciais contínuas de todas as ordens. Portanto,  $\varphi_1$  é diferenciável e a condição 1 é satisfeita. Temos também que

$$\begin{vmatrix} \frac{\partial(x)}{\partial x} & \frac{\partial(x)}{\partial y} \\ \frac{\partial(y)}{\partial x} & \frac{\partial(y)}{\partial y} \end{vmatrix} (x, y) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

ou seja, um dos determinantes menores da matriz jacobiana é diferente de zero e logo a condição 3 é verificada. Para verificar a condição 2 observamos que  $\varphi_1$  é bijetiva e que  $(\varphi_1)^{-1}$  é a restrição da projeção (contínua)  $\pi(x, y, z) = (x, y)$  ao conjunto  $\varphi_1(U)$ . Assim  $(\varphi_1)^{-1}$  é contínua em  $\varphi_1(U)$ .

Agora, vamos cobrir a esfera inteira utilizando parametrizações similares. Definimos  $\varphi_2 : U \rightarrow \mathbb{R}^3$  por

$$\varphi_2(x, y) = (x, y, -\sqrt{1 - (x^2 + y^2)}), (x, y) \in U,$$

de modo análogo a  $\varphi_1$ , temos que  $\varphi_2$  é uma parametrização. Observe que  $\varphi_1(U) \cup \varphi_2(U)$  cobre a esfera menos o equador  $\{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 = 1, z = 0\}$ . Utilizando então os planos  $xz$  e  $zy$ , definimos as seguintes parametrizações

$$\varphi_3(x, z) = (x, \sqrt{1 - (x^2 + z^2)}, z)$$

$$\varphi_4(x, z) = (x, -\sqrt{1 - (x^2 + z^2)}, z)$$

$$\varphi_5(y, z) = (\sqrt{1 - (y^2 + z^2)}, y, z)$$

$$\varphi_6(y, z) = (-\sqrt{1 - (y^2 + z^2)}, y, z)$$

cujas imagens junto com  $\varphi_1$  e  $\varphi_2$ , cobrem  $S^2$  inteiramente. Portanto, mostramos que  $S^2$  é uma superfície regular.

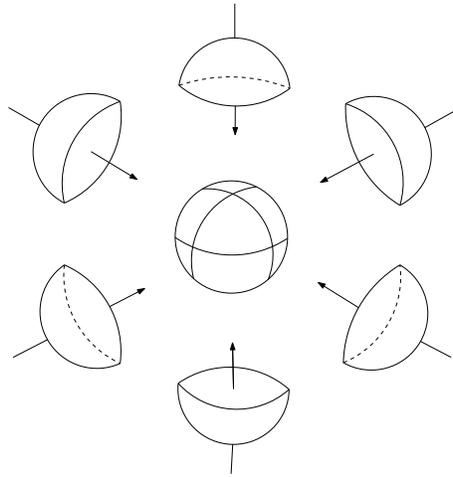


Figura 1.4: Parametrização da esfera.

Podemos também cobrir a esfera com um outro sistemas de coordenadas. Considere a parametrização  $\zeta : U \rightarrow \mathbb{R}^3$  dada por

$$\zeta(\theta, \phi) = (\text{sen } \phi \cos \theta, \text{sen } \phi \text{sen } \theta, \cos \phi),$$

onde  $U = \{(\theta, \phi); 0 < \theta < 2\pi, 0 < \phi < \pi\}$ .

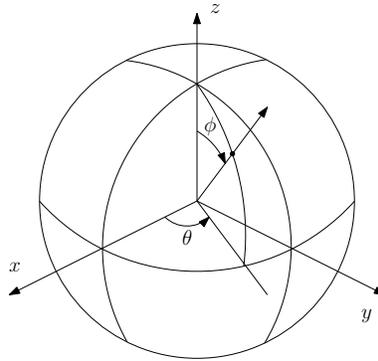


Figura 1.5: Parametrização da esfera.

A imagem de  $\zeta$  não é toda a esfera, mas sim  $\zeta(U) = S^2 \setminus \mathcal{C}$ , onde  $\mathcal{C}$  é o semicírculo  $\mathcal{C} = \{(x, y, z) \in S^2; x \geq 0, y = 0\}$ . A aplicação  $\zeta$  é um homeomorfismo diferenciável de  $U$  na interseção da esfera com o aberto  $V = \{(x, y, x) \in \mathbb{R}^3; x < 0 \text{ ou } y \neq 0\}$ . Além disso, para que os determinantes menores do Jacobiano

$$\begin{vmatrix} \frac{\partial \zeta_1}{\partial \phi} & \frac{\partial \zeta_1}{\partial \theta} \\ \frac{\partial \zeta_2}{\partial \phi} & \frac{\partial \zeta_2}{\partial \theta} \end{vmatrix}(\phi, \theta) = \cos \phi \text{sen } \theta, \quad \begin{vmatrix} \frac{\partial \zeta_2}{\partial \phi} & \frac{\partial \zeta_2}{\partial \theta} \\ \frac{\partial \zeta_3}{\partial \phi} & \frac{\partial \zeta_3}{\partial \theta} \end{vmatrix}(\phi, \theta) = \text{sen}^2 \phi \cos \theta, \quad \begin{vmatrix} \frac{\partial \zeta_1}{\partial \phi} & \frac{\partial \zeta_1}{\partial \theta} \\ \frac{\partial \zeta_3}{\partial \phi} & \frac{\partial \zeta_3}{\partial \theta} \end{vmatrix}(\phi, \theta) = -\text{sen}^2 \phi \text{sen } \theta,$$

se anularem simultaneamente, é necessário que

$$\cos^2 \phi \text{sen}^2 \theta + \text{sen}^4 \phi \cos^2 \theta + \text{sen}^4 \phi \text{sen}^2 \theta = \text{sen}^2 \phi = 0.$$

Isso não ocorre em  $U$ , pois  $\phi \in (0, \pi)$ . Para cobrir toda a esfera, tome a seguinte parametrização  $\bar{\zeta} : U \rightarrow \mathbb{R}^3$  definida por

$$\bar{\zeta} = (\theta, \phi) = (-\cos \phi \cos \theta, -\sin \phi, -\cos \phi \sin \theta),$$

onde  $U$  é o mesmo que definido anteriormente. A imagem de  $\bar{\zeta}$  é dada por  $\zeta(\bar{U}) = S^2 \setminus \bar{C}$ , onde  $\bar{C}$  é o semicírculo  $\bar{C} = \{(x, y, z) \in S^2; x \leq 0, z = 0\}$ . Logo,  $\zeta(U) \cup \bar{\zeta}(U) = S^2$  e, de modo análogo ao feito para  $\zeta$ , provamos que  $\bar{\zeta}$  é uma parametrização de  $S^2$ .

A proposição abaixo mostra que quando um ponto pertence a duas vizinhanças coordenadas, com parâmetros  $(x, y)$  e  $(u, v)$ , por exemplo, é possível passar de um destes pares de coordenadas ao outro através de um difeomorfismo.

**Proposição 1.1.6.** Seja  $p$  um ponto de uma superfície regular  $S$ , e sejam  $\varphi : U \subset \mathbb{R}^2 \rightarrow S$  e  $\psi : V \subset \mathbb{R}^2 \rightarrow S$  duas parametrizações de  $S$ , tais que  $p \in \varphi(U) \cap \psi(V) = W$ . Então a mudança de coordenadas  $h = \varphi^{-1} \circ \psi : \psi^{-1}(W) \rightarrow \varphi^{-1}(W)$  é um difeomorfismo, isto é,  $h$  é diferenciável e tem inversa diferenciável  $h^{-1}$ .

Abaixo definimos o que é uma função diferenciável em uma superfície regular.

**Definição 1.1.7.** Seja  $f : V \subset S \rightarrow \mathbb{R}$  uma função, definida em um subconjunto aberto  $V$  de uma superfície regular  $S$ . Então  $f$  é diferenciável em  $p \in V$  quando, para alguma parametrização  $\varphi : U \subset \mathbb{R}^2 \rightarrow S$ , com  $p \in \varphi(U) \subset V$ , a composição  $f \circ \varphi : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  é diferenciável em  $\varphi^{-1}(p)$ . A função  $f$  é diferenciável em  $V$  quando é diferenciável em todos os pontos de  $V$ .

A definição acima não depende da escolha da parametrização  $\varphi$ , isso é consequência imediata da proposição anterior. O conceito de diferenciabilidade pode ser facilmente estendido a aplicações entre superfícies.

**Definição 1.1.8.** Dizemos que uma aplicação contínua  $f : V_1 \subset S_1 \rightarrow S_2$  de um conjunto aberto  $V_1$  de uma superfície regular  $S_1$  em uma superfície regular  $S_2$  é diferenciável em  $p \in V_1$  quando, dadas parametrizações

$$\varphi : U_1 \subset \mathbb{R}^2 \rightarrow S_1, \quad \phi : U_2 \subset \mathbb{R}^2 \rightarrow S_2$$

com  $p \in \varphi(U_1)$  e  $f(\varphi(U_1)) \subset \phi(U_2)$ , a aplicação

$$\phi^{-1} \circ f \circ \varphi : U_1 \rightarrow U_2$$

é diferenciável em  $q = \varphi^{-1}(p)$ .

Uma característica importante das superfícies regulares é que elas possuem, em cada ponto, uma aproximação linear, que é seu plano tangente.

**Definição 1.1.9.** (Vetor tangente) Seja  $S$  uma superfície regular. Uma curva diferenciável em  $S$  é uma aplicação  $\alpha : (-\epsilon, \epsilon) \rightarrow S$  diferenciável. Um vetor tangente a  $S$  em um ponto  $p \in S$ , é o vetor tangente  $\alpha'(0)$  (derivada de  $\alpha$  no ponto 0) de uma curva diferenciável  $\alpha$  com  $\alpha(0) = p$ .

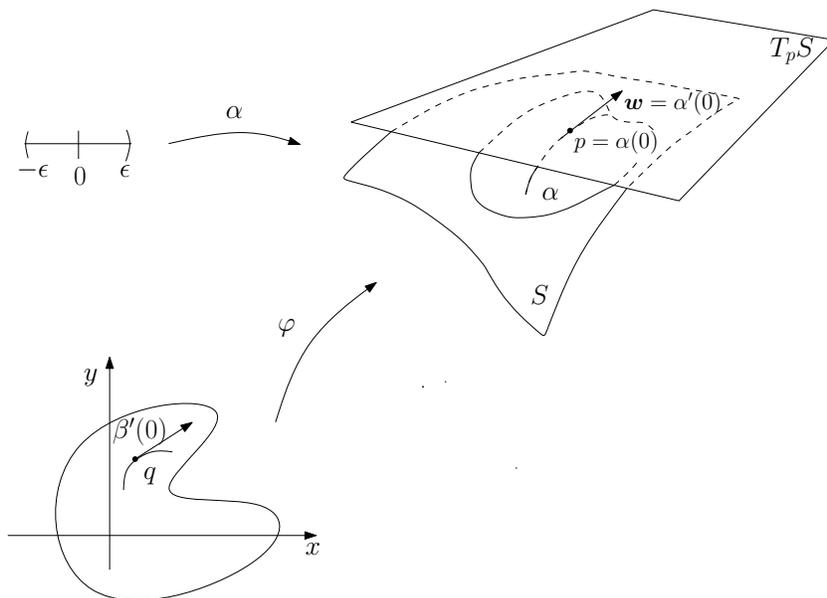


Figura 1.6: Espaço tangente.

**Definição 1.1.10.** (Espaço tangente) O conjunto de todos os vetores tangentes a  $S$  em um ponto  $p$  é chamado de espaço tangente a  $S$  em  $p$  e é denotado por  $T_p S$ .

**Proposição 1.1.11.** Dada uma parametrização  $\varphi$ , o conjunto dos vetores tangentes a  $S$  em  $p = \varphi(q)$  coincide com o subespaço vetorial do  $\mathbb{R}^3$  gerado pelos vetores  $\frac{\partial \varphi}{\partial x}(q)$  e  $\frac{\partial \varphi}{\partial y}(q)$ .

As coordenadas de um vetor  $\mathbf{w} \in T_p S$  na base associada a uma parametrização  $\varphi : U \rightarrow S$  são determinadas do seguinte modo:  $\mathbf{w}$  é um vetor velocidade  $\alpha'(0)$  de alguma curva  $\alpha = \varphi \circ \beta$ , onde  $\beta : (-\epsilon, \epsilon) \rightarrow U$  é dada por  $\beta(t) = (x(t), y(t))$ , com  $\beta(0) = q = \varphi^{-1}(p)$ . Então,

$$\begin{aligned} \alpha'(0) &= \frac{d}{dt}(\varphi \circ \beta)(0) = \frac{d}{dt}\varphi(x(t), y(t))(0) \\ &= \varphi_x(q)x'(0) + \varphi_y(q)y'(0) = \mathbf{w}. \end{aligned}$$

Observamos, do que foi visto acima, que o vetor tangente a uma curva  $\alpha$  em  $p$  depende apenas das derivadas de  $\alpha$  em um sistema de coordenadas. O conjunto  $T_p S$ , com as operações usuais de aplicações, forma um espaço vetorial bi-dimensional e a escolha de uma parametrização  $\varphi : U \rightarrow S$  determina uma base associada  $\left\{ \left( \frac{\partial \varphi}{\partial x}(q), \frac{\partial \varphi}{\partial y}(q) \right) \right\}$  em  $T_p S$ .

Com a definição de espaço tangente, podemos falar na diferencial de uma aplicação diferenciável entre superfícies. Sejam  $S_1$  e  $S_2$  duas superfícies regulares e seja  $f : V \subset S_1 \rightarrow S_2$  uma aplicação diferenciável de um conjunto aberto  $V$  de  $S_1$  em  $S_2$ . Se  $p \in V$ , sabemos que todo vetor tangente  $\mathbf{w} \in T_p S_1$  é o vetor tangente  $\alpha'(0)$  a uma curva diferenciável  $\alpha : (-\epsilon, \epsilon) \rightarrow V$  com  $\alpha(0) = p$ . A curva  $\beta = f \circ \alpha$  é tal que  $\beta(0) = f(p)$ , e portanto,  $\beta'(0)$  é um vetor de  $T_{f(p)} S_2$ .

**Proposição 1.1.12.** Dado  $\mathbf{w} \in T_p S_1$ , o vetor  $\beta'(0)$  não depende da escolha de  $\alpha$ . Além disso, a aplicação  $df_p : T_p S_1 \rightarrow T_{f(p)} S_2$  definida por  $df_p(\mathbf{w}) = \beta'(0)$  é linear.

**Definição 1.1.13.** A aplicação linear  $df_p : T_p S_1 \rightarrow T_{f(p)} S_2$  é chamada de diferencial de  $f$  em  $p \in S_1$ .

Até o momento estudamos superfícies no que diz respeito à sua diferenciabilidade. Suponha agora que exista alguém que habite uma dada superfície. Uma curiosidade que ele pode ter é como medir a distância entre dois pontos da superfície. Em geral, essa distância é diferente da distância do espaço onde a superfície está contida, pois no  $\mathbb{R}^3$ , por exemplo, a reta é a curva que dá o menor caminho entre dois pontos porém nem sempre a reta está contida na superfície. A definição abaixo nos permite estudar algumas propriedades geométricas intrínsecas da superfície como comprimentos, ângulos e áreas.

**Definição 1.1.14.** (Primeira Forma Fundamental) Seja  $S$  uma superfície regular e  $T_p S$  o plano tangente a  $S$  no ponto  $p$ . A forma quadrática  $I_p$  definida por

$$\begin{aligned} I_p : T_p S &\rightarrow \mathbb{R} \\ \mathbf{w} &\rightarrow I_p(\mathbf{w}) = \langle \mathbf{w}, \mathbf{w} \rangle = \|\mathbf{w}\|^2 \geq 0 \end{aligned} \quad (1.1.1)$$

é chamada de primeira forma fundamental de  $S$  em  $p$ .

Ou seja, a cada ponto  $p \in S$  a primeira forma fundamental associa um produto interno  $\langle \cdot, \cdot \rangle_p$  no espaço tangente  $T_p S$  induzido do  $\mathbb{R}^3$ .

Podemos expressar a primeira forma fundamental na base  $\{\varphi_x(q), \varphi_y(q)\}$  associada a uma parametrização  $\varphi(x, y)$  em  $p = \varphi(q)$ . Seja  $\mathbf{w} \in T_p S$  o vetor tangente a uma curva parametrizada  $\alpha(t) = \varphi(x(t), y(t))$ ,  $t \in (-\epsilon, \epsilon)$ , com  $p = \alpha(0) = \varphi(q)$ , onde  $q = (x_0, y_0)$ , obtemos

$$\begin{aligned} I_p(\alpha'(0)) &= \langle \alpha'(0), \alpha'(0) \rangle_p \\ &= \langle \varphi_x(q)x'(0) + \varphi_y(q)y'(0), \varphi_x(q)x'(0) + \varphi_y(q)y'(0) \rangle_p \\ &= \langle \varphi_x(q), \varphi_x(q) \rangle_p (x'(0))^2 + 2\langle \varphi_x(q), \varphi_y(q) \rangle_p x'(0)y'(0) + \langle \varphi_y(q), \varphi_y(q) \rangle_p (y'(0))^2 \\ &= E(q)(x'(0))^2 + 2F(q)x'(0)y'(0) + G(q)(y'(0))^2, \end{aligned} \quad (1.1.2)$$

onde  $x'(0)$ , denota a derivada de  $x$  em relação a  $t = 0$ . Os valores  $E(q) = \langle \varphi_x(q), \varphi_x(q) \rangle_p$ ,  $F(q) = \langle \varphi_x(q), \varphi_y(q) \rangle_p$  e  $G(q) = \langle \varphi_y(q), \varphi_y(q) \rangle_p$  são chamados de coeficientes da primeira forma fundamental na base  $\{\varphi_x(q), \varphi_y(q)\}$  de  $T_p S$ . Fazendo  $p$  variar na vizinhança coordenada correspondente a  $\varphi(x, y)$ , obtemos funções  $E(x, y)$ ,  $F(x, y)$  e  $G(x, y)$  que são diferenciáveis nessa vizinhança. É importante salientar que, embora os coeficientes dependam da parametrização  $\varphi$ , a primeira forma fundamental independe do sistema de coordenadas.

Com a primeira forma fundamental, podemos calcular, por exemplo, o comprimento de arco  $\ell$  de uma curva parametrizada  $\alpha : I \rightarrow S$ , onde  $I$  é um intervalo que possui o 0, em  $S$ . Seja  $p = \alpha(t)$ , temos que

$$\ell(t) = \int_0^t |\alpha'(t)| dt = \int_0^t \sqrt{I_p(\alpha'(t))} dt.$$

Um curva parametrizada diferenciável  $\alpha : I \rightarrow S$  é dita parametrizada pelo comprimento de arco quando  $|\alpha'(t)| = 1$ . Toda curva parametrizada admite uma reparametrização pelo comprimento de arco. A partir de agora assumimos que todas as curvas parametrizadas são parametrizadas pelo comprimento de arco.

A seguir, damos a definição de geodésicas, que são curvas importantes pois tem a propriedade de minimizar, localmente, distâncias sobre as superfícies, como será visto na Proposição 1.1.22. Para isso, é necessário entender a noção de derivada covariante, que é o análogo para superfícies da derivação usual de vetores no plano.

**Definição 1.1.15.** (Campo de vetores tangentes) Seja  $V \subset S$  um conjunto aberto sobre uma superfície regular  $S$ . Um campo de vetores tangentes a  $S$  definido sobre  $V$  é uma aplicação  $\vec{W} : V \rightarrow \mathbb{R}^3$  que associa a cada ponto  $p \in V$  um vetor  $\vec{W}(p) \in T_pS$ . Se  $\varphi : U \subset \mathbb{R}^n \rightarrow V$  é uma parametrização de  $S$  então  $\vec{W}$  pode ser escrito como

$$\vec{W} = a(x, y)\varphi_x + b(x, y)\varphi_y,$$

onde as componentes  $a, b : U \rightarrow \mathbb{R}$  são chamadas de coordenadas do campo na base  $\{\varphi_x, \varphi_y\}$ . O campo de vetores é diferenciável em  $p$  quando, para alguma parametrização  $\varphi$  em  $p$ , as componentes  $a$  e  $b$ , são funções diferenciáveis em  $\varphi^{-1}(p)$ . O campo de vetores  $\vec{W}$  é diferenciável em  $V$  quando é diferenciável para todo  $p \in V$ .

**Definição 1.1.16.** (Derivada covariante) Seja  $\vec{W}$  um campo de vetores em um conjunto aberto  $V \subset S$  e  $p \in V$ . Seja  $\mathbf{v} \in T_pS$ . Considere uma curva parametrizada

$$\alpha : (-\epsilon, \epsilon) \rightarrow V,$$

com  $\alpha(0) = p$  e  $\alpha'(0) = \mathbf{v}$ . Seja  $\vec{W}(t)$ ,  $t \in (-\epsilon, \epsilon)$ , a restrição do campo de vetores à curva  $\alpha$ . O vetor obtido pela projeção de  $(d\vec{W}/dt)(0)$  sobre o plano  $T_pS$  é chamado derivada covariante em  $p$  do campo de vetores  $\vec{W}$  em relação ao vetor  $\mathbf{v}$ . Esta derivada covariante é denotada por  $(D\vec{W}/dt)(0)$ .

A derivada covariante é um conceito da geometria intrínseca e que não depende da escolha da curva  $\alpha$ . De fato, seja

$$\varphi(x(t), y(t)) = \alpha(t)$$

a expressão da curva  $\alpha$  na parametrização  $\varphi(x, y)$  de  $S$  em  $p$ . Podemos escrever o campo de vetores restrito à curva  $\alpha$  como

$$\begin{aligned} \vec{W}(t) &= a(x(t), y(t))\varphi_x + b(x(t), y(t))\varphi_y \\ &= a(t)\varphi_x + b(t)\varphi_y. \end{aligned}$$

Logo, derivando  $\vec{W}$  em relação a  $t$ , temos

$$\frac{d\vec{W}}{dt} = a(\varphi_{xx}x' + \varphi_{xy}y') + b(\varphi_{yx}x' + \varphi_{yy}y') + a'\varphi_x + b'\varphi_y.$$

Como  $D\vec{W}/dt$  é a componente de  $d\vec{W}/dt$  no plano tangente, temos que

$$\begin{aligned} \frac{D\vec{W}}{dt} &= (a' + \Gamma_{11}^1ax' + \Gamma_{12}^1ay' + \Gamma_{12}^1bx' + \Gamma_{22}^1by')\varphi_x \\ &\quad + (b' + \Gamma_{11}^2ax' + \Gamma_{12}^2ay' + \Gamma_{12}^2bx' + \Gamma_{22}^2by')\varphi_y, \end{aligned} \tag{1.1.3}$$

onde os coeficientes  $\Gamma_{ij}^k$ ,  $i, j, k = 1, 2$  são chamados símbolos de Christoffel de  $S$  na parametrização  $\varphi$ . Eles são dados pela solução dos sistemas abaixo

$$\begin{cases} \Gamma_{11}^1 E + \Gamma_{11}^2 F &= \frac{1}{2} E_x \\ \Gamma_{11}^1 F + \Gamma_{11}^2 G &= F_x - \frac{1}{2} E_y, \end{cases}$$

$$\begin{cases} \Gamma_{12}^1 E + \Gamma_{12}^2 F &= \frac{1}{2} E_y \\ \Gamma_{12}^1 F + \Gamma_{12}^2 G &= \frac{1}{2} G_x, \end{cases}$$

$$\begin{cases} \Gamma_{22}^1 E + \Gamma_{22}^2 F &= F_y - \frac{1}{2} G_x \\ \Gamma_{22}^1 F + \Gamma_{22}^2 G &= \frac{1}{2} G_y, \end{cases}$$

onde  $E$ ,  $F$  e  $G$  são os coeficientes da primeira forma fundamental dada uma parametrização  $\varphi$ , com  $EG - F^2 \neq 0$ . Para um melhor entendimento sobre os símbolos de Christoffel ver referência [8] na Seção 4.3.

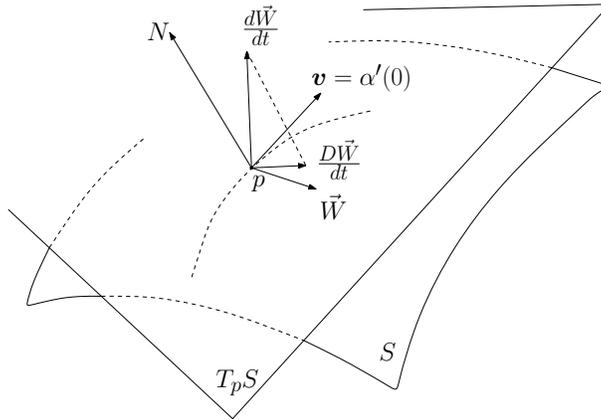


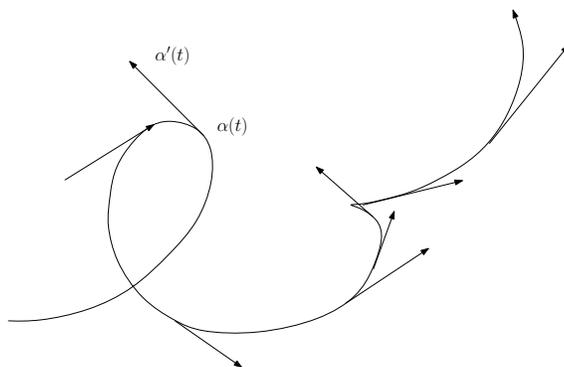
Figura 1.7: Derivada covariante.

Dessa forma, a expressão (1.1.3) mostra que  $D\vec{W}/dt$  depende apenas do vetor  $\mathbf{v} = (x', y')$  e não da curva  $\alpha$ .

A definição de derivada covariante pode ser restrita a um campo de vetores que esteja definido apenas em pontos de uma curva parametrizada.

**Definição 1.1.17.** Seja  $\alpha : I \subset \mathbb{R} \rightarrow S$  uma curva parametrizada em  $S$ . Um campo de vetores  $\vec{W}$  ao longo de  $\alpha$  é uma correspondência que associa a cada  $t \in I$  um vetor  $\vec{W}(t) \in T_{\alpha(t)}S$ . O campo de vetores  $\vec{W}$  é diferenciável em  $t_0 \in I$  quando para alguma parametrização  $\varphi(x, y)$  em  $\alpha(t_0)$  as componentes  $a(t)$  e  $b(t)$  de  $\vec{W}(t) = a(t)\varphi_x + b(t)\varphi_y$  são funções diferenciáveis de  $t$  em  $t_0$ . O campo  $\vec{W}$  é diferenciável em  $I$  quando é diferenciável para todo  $t \in I$ .

**Definição 1.1.18.** Seja  $\vec{W}$  um campo de vetores ao longo de  $\alpha : I \rightarrow S$ . A expressão (1.1.3) de  $(D\vec{W}/dt)$ ,  $t \in I$ , está bem definida e é chamada derivada covariante de  $\vec{W}$  em  $t$ .

Figura 1.8: Campo vetorial ao longo da curva  $\alpha$ .

Para obtermos a derivada covariante de um campo  $\vec{W}$  ao longo de uma curva  $\alpha : I \rightarrow S$  em  $t \in I$  consideramos a derivada usual  $(d\vec{W}/dt)(t)$  de  $\vec{W}$  em  $t$  e projetamos este vetor ortogonalmente sobre o plano tangente  $T_{\alpha(t)}S$ .

**Definição 1.1.19.** (Campo de vetores paralelos) Um campo de vetores  $\vec{W}$  ao longo de uma curva parametrizada  $\alpha : I \rightarrow S$  é chamado paralelo se  $D\vec{W}/dt = 0$  para todo  $t \in I$ .

Tomando como exemplo o caso particular do plano, temos que o campo paralelo ao longo de uma curva parametrizada é o campo constante ao longo da curva, ou seja, o comprimento do vetor e o ângulo que ele faz com uma direção são constantes.

As curvas parametrizadas  $\gamma : I \rightarrow \mathbb{R}^2$  do plano ao longo das quais o campo de vetores tangentes  $\gamma'(t)$  é paralelo são precisamente as retas desse plano. Em uma superfície as curvas que satisfazem uma condição análoga são chamadas geodésicas.

**Definição 1.1.20.** (Geodésica) Uma curva parametrizada, não constante,  $\gamma : I \rightarrow S$  é chamada geodésica em  $t \in I$  quando o seu campo de vetores  $\gamma'(t)$  é paralelo ao longo de  $\gamma$  em  $t$ ; isto é

$$\frac{D\gamma'(t)}{dt} = 0;$$

$\gamma$  é uma geodésica parametrizada quando é geodésica para todo  $t \in I$ .

Vamos determinar as equações satisfeitas por uma geodésica em uma vizinhança coordenada. Seja  $\gamma : I \rightarrow S$  uma curva parametrizada de  $S$  e seja  $\varphi(x, y)$  uma parametrização de  $S$  em uma vizinhança  $V$  de  $\gamma(t_0)$ ,  $t_0 \in I$ . Seja  $J \subset I$  um intervalo aberto contendo  $t_0$  tal que  $\gamma(J) \subset V$ . Seja  $\varphi(x(t), y(t))$ ,  $t \in J$ , a expressão de  $\gamma : J \rightarrow S$  na parametrização  $\varphi$ . Então, o campo de vetores tangentes  $\gamma'(t)$ ,  $t \in J$ , é dado por

$$\vec{W} = x'(t)\varphi_x + y'(t)\varphi_y.$$

Portanto, o fato de  $\vec{W}$  ser paralelo é equivalente ao sistema de equações diferenciais

$$\begin{aligned} x'' + \Gamma_{11}^1(x')^2 + 2\Gamma_{12}^1x'y' + \Gamma_{22}^1(y')^2 &= 0 \\ y'' + \Gamma_{11}^2(x')^2 + 2\Gamma_{12}^2x'y' + \Gamma_{22}^2(y')^2 &= 0 \end{aligned} \tag{1.1.4}$$

obtido a partir da equação (1.1.3) fazendo  $a = x'$  e  $b = y'$ , e igualando a zero os coeficientes de  $\varphi_x$  e  $\varphi_y$ .

Assim, uma curva parametrizada  $\gamma : I \rightarrow S$  é uma geodésica se, e somente se, o sistema (1.1.4) é satisfeito para todo  $J \subset I$  tal que  $\gamma(J)$  esteja contido em uma vizinhança coordenada. O sistema (1.1.4) é conhecido como equações diferenciais das geodésicas de  $S$ .

A proposição abaixo mostra uma consequência importante do fato de que as geodésicas são caracterizadas pelo sistema acima.

**Proposição 1.1.21.** Dado um ponto  $p \in S$  e um vetor  $\mathbf{w} \in T_p S$ ,  $\mathbf{w} \neq 0$ , existe um  $\epsilon > 0$  e uma única geodésica parametrizada  $\gamma : (-\epsilon, \epsilon) \rightarrow S$  tal que  $\gamma(0) = p$  e  $\gamma'(0) = \mathbf{w}$ .

Uma propriedade importante de uma geodésica é o fato de que, localmente, ela minimiza o comprimento de arco.

**Proposição 1.1.22.** Seja  $p$  um ponto e  $W$  uma vizinhança de  $p$  em uma superfície  $S$ . Então existe uma vizinhança  $W \subset S$  de  $p$  tal que se  $\gamma : I \rightarrow W$  é uma geodésica parametrizada com  $\gamma(0) = p$ ,  $\gamma(t_1) = q$ ,  $t_1 \in I$ , e  $\alpha : [0, t_1] \rightarrow S$  é uma curva parametrizada regular (isto é,  $\alpha'(t) \neq 0$  para todo  $t \in I$ ) ligando  $p$  a  $q$ , temos

$$l_\gamma \leq l_\alpha,$$

onde  $l_\alpha$  denota o comprimento da curva  $\alpha$ . Além disso, se  $l_\gamma = l_\alpha$ , então  $\gamma([0, t_1]) = \alpha([0, t_1])$ .

Vale ressaltar que a proposição acima não é global. Se considerarmos uma geodésica com arco suficientemente grande, ela pode não ser o menor caminho entre seus pontos extremos. Como por exemplo na esfera: dois pontos que não são antípodas de uma esfera podem ser conectados por dois meridianos de comprimentos diferentes e apenas o menor deles satisfaz as conclusões da proposição. No entanto, quando uma curva  $\alpha : I \rightarrow S$  regular é o menor caminho entre quaisquer dois de seus pontos, esta curva é necessariamente uma geodésica, como mostra a proposição abaixo:

**Proposição 1.1.23.** Seja  $\alpha : I \rightarrow S$  uma curva parametrizada regular com um parâmetro proporcional ao comprimento de arco. Suponha que o comprimento de  $\alpha$  entre dois pontos quaisquer  $t, \tau \in I$  é menor ou igual ao comprimento de qualquer curva parametrizada regular ligando  $\alpha(t)$  a  $\alpha(\tau)$ . Então  $\alpha$  é uma geodésica.

## 1.2 Variedade Diferenciável

Nesta seção introduzimos o conceito de variedade diferenciável. Dado um conjunto  $M$  qualquer, não necessariamente contido no  $\mathbb{R}^n$ , queremos saber qual conjunto seria o análogo a uma reta no  $\mathbb{R}^n$  e qual a melhor maneira de definir a distância entre dois pontos desse conjunto. A noção de variedade diferenciável estende os métodos de cálculo diferencial a espaços mais gerais que o  $\mathbb{R}^n$ . Aqui, diferenciável sempre significa de classe  $C^\infty$ .

**Definição 1.2.1.** (Variedade diferenciável) Seja  $M$  um conjunto.  $M$  é uma variedade diferenciável de dimensão  $n$  (ou simplesmente, uma variedade), quando existe uma família de aplicações bijetivas  $\{U_\lambda, \varphi_\lambda\}$ , de conjuntos abertos  $U_\lambda \subset \mathbb{R}^n$  em  $M$ ,  $\varphi_\lambda : U_\lambda \rightarrow M$ , satisfazendo as seguintes condições:

1.  $\bigcup_{\lambda} \varphi_{\lambda}(U_{\lambda}) = M$
2. Para cada par  $\lambda, \zeta$  com  $\varphi_{\lambda}(U_{\lambda}) \cap \varphi_{\zeta}(U_{\zeta}) = W \neq \emptyset$ , temos que  $\varphi_{\lambda}^{-1}(W)$  e  $\varphi_{\zeta}^{-1}(W)$  são conjuntos abertos em  $\mathbb{R}^n$  e  $\varphi_{\zeta}^{-1} \circ \varphi_{\lambda}$ ,  $\varphi_{\lambda}^{-1} \circ \varphi_{\zeta}$  são aplicações diferenciáveis.
3. A família  $\{U_{\lambda}, \varphi_{\lambda}\}$  é máxima em relação às condições 1 e 2.

O par  $(U_{\lambda}, \varphi_{\lambda})$  com  $p \in \varphi_{\lambda}(U_{\lambda})$  é chamado uma parametrização (ou sistemas de coordenadas) de  $M$  em  $p$ ;  $\varphi_{\lambda}(U_{\lambda})$  é então chamada uma vizinhança coordenada de  $S$  em  $p$ .

Uma família  $(U_{\lambda}, \varphi_{\lambda})$  satisfazendo as condições 1 e 2 é chamada estrutura diferenciável. Dada uma estrutura diferenciável podemos completá-la com todas as parametrizações que, junto com alguma parametrização da estrutura, satisfaça a condição 2, assim ela irá satisfazer a condição 3. Portanto, com um abuso de linguagem, chamamos de variedade diferenciável um conjunto munido de uma estrutura diferenciável.

**Exemplo 1.2.2.** (Os espaços Euclidianos) Seja o  $\mathbb{R}^n$  e seja a família de sistemas de coordenadas dada por um único sistemas de coordenadas  $(\mathbb{R}^n, \varphi)$ ,  $\varphi = id : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . É fácil ver que  $\mathbb{R}^n$  é uma variedade diferenciável de dimensão  $n$ .

**Exemplo 1.2.3.** (Superfícies regulares) As superfícies regulares definidas na Seção 1 são variedades diferenciáveis.

**Definição 1.2.4.** (Aplicação Diferenciável) Sejam  $M_1$  e  $M_2$  variedades diferenciáveis de dimensões  $n$  e  $m$ , respectivamente. Uma aplicação  $f : M_1 \rightarrow M_2$  é diferenciável em  $p \in M_1$  quando dada uma parametrização  $\phi : V \subset \mathbb{R}^m \rightarrow M_2$  em torno de  $f(p)$ , existe uma parametrização  $\varphi : U \subset \mathbb{R}^n \rightarrow M_1$  em torno de  $p$  tal que  $f(\varphi(U)) \subset \phi(V)$  e a aplicação

$$\phi^{-1} \circ f \circ \varphi : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$$

é diferenciável em  $\varphi^{-1}(p)$ .  $f$  é diferenciável em  $S_1$  quando é diferenciável em todo  $p \in M_1$ .

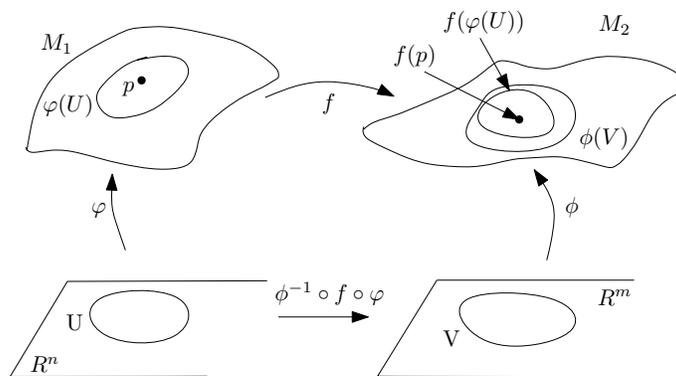


Figura 1.9: Aplicação Diferenciável.

Observe que, pela condição 2, a definição acima não depende das escolhas dos sistemas de coordenadas.

Agora, vamos estender a noção de vetor tangente para variedades diferenciáveis. Numa superfície regular vimos que plano tangente é o conjunto de vetores tangentes em um ponto, sendo um vetor tangente em um ponto definido como a velocidade, em  $\mathbb{R}^3$ , neste ponto, de uma curva na superfície. Como não dispomos do suporte de um espaço ambiente, é necessário buscar uma propriedade característica do vetor tangente que substitua a noção de velocidade.

As seguintes considerações irão motivar a definição que será dada abaixo. Seja  $\alpha : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$  uma curva diferenciável em  $\mathbb{R}^n$ , com  $\alpha(0) = p$ . Escreva

$$\alpha(t) = (x_1(t), \dots, x_n(t)), \quad t \in (-\epsilon, \epsilon), (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Logo  $\alpha'(0) = (x'_1(0), \dots, x'_n(0)) = \mathbf{v} \in \mathbb{R}^n$ . Seja agora  $f$  uma função diferenciável definida em uma vizinhança de  $p$ . Podemos restringir  $f$  à curva  $\alpha$  e escrever a derivada direcional segundo o vetor  $\mathbf{v} \in \mathbb{R}^n$  como

$$\left. \frac{d(f \circ \alpha)}{dt} \right|_{t=0} = \left( \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \right) \Big|_{t=0} = \left( \sum_{i=1}^n x'_i(0) \frac{\partial}{\partial x_i} \right) f.$$

Portanto, a derivada direcional segundo  $\mathbf{v}$  é um operador sobre funções diferenciáveis que depende unicamente de  $\mathbf{v}$ . Esta é a propriedade característica dos vetores tangentes que estávamos buscando.

**Definição 1.2.5.** (Vetor tangente) Seja  $M$  uma variedade diferenciável. Uma curva diferenciável em  $M$  é uma aplicação  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  diferenciável. Suponha que  $\alpha(0) = p$  e seja  $D$  o conjunto de funções em  $M$  que são diferenciáveis em  $p$ . O vetor tangente a  $\alpha$  em  $t = 0$  é definido como sendo a aplicação  $\alpha'(0) : D \rightarrow \mathbb{R}$  dada por

$$\alpha'(0)(f) = \left. \frac{d(f \circ \alpha)}{dt} \right|_{t=0}, \quad f \in D.$$

Um vetor tangente em um ponto  $p \in M$  é o vetor tangente em  $t = 0$  a alguma curva  $\alpha : (-\epsilon, \epsilon) \rightarrow M$  com  $\alpha(0) = p$ .

Se escolhermos uma parametrização  $\varphi : U \rightarrow M$  em torno de um ponto  $p = \varphi(0)$ , podemos expressar a função  $f$  e a curva  $\alpha$  nessa parametrização por

$$(f \circ \varphi)(q) = f(x_1, x_2, \dots, x_n), \quad q = (x_1, x_2, \dots, x_n) \in U.$$

$$(\varphi^{-1} \circ \alpha)(t) = (x_1(t), x_2(t), \dots, x_n(t)),$$

respectivamente. Portanto, restringindo  $f$  a  $\alpha$ , obtemos

$$\begin{aligned} \alpha'(0)(f) &= \left. \frac{d(f \circ \alpha)}{dt} \right|_{t=0} = \left. \frac{d(f(x_1, x_2, \dots, x_n))}{dt} \right|_{t=0} \\ &= \sum_{i=1}^n x'_i(0) \frac{\partial f}{\partial x_i}(0) = \left( \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0 \right) (f). \end{aligned}$$

Dadas coordenadas  $(x_1, \dots, x_n)$  em torno de  $p$ , denotamos por  $(\partial/\partial x_i)_0$  o vetor tangente em  $p$  que aplica a função  $f$  em  $(\partial f/\partial x_i)(0)$ , para  $i = 1, \dots, n$ . Portanto, o vetor  $\alpha'(0)$  pode ser expresso na parametrização  $\varphi$  por

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left( \frac{\partial}{\partial x_i} \right)_0.$$

**Definição 1.2.6.** (Espaço tangente) O conjunto de todos os vetores tangentes a  $M$  em um ponto  $p$  é chamado de espaço tangente a  $M$  em  $p$  e é denotado por  $T_p M$ .

Observamos, do que foi visto acima, que o vetor tangente a uma curva  $\alpha$  em  $p$  depende apenas das derivadas de  $\alpha$  em um sistema de coordenadas. O conjunto  $T_p M$ , com as operações usuais de funções, forma um espaço vetorial de dimensão  $n$  e a escolha de uma parametrização  $\varphi : U \rightarrow M$  determina uma base associada  $\left\{ \left( \frac{\partial}{\partial x_1} \right)_0, \left( \frac{\partial}{\partial x_2} \right)_0, \dots, \left( \frac{\partial}{\partial x_n} \right)_0 \right\}$  em  $T_p M$ .

Com a noção de espaço tangente, podemos estender às variedades diferenciáveis a noção de diferencial de uma aplicação diferenciável.

**Proposição 1.2.7.** Sejam  $M_1$  e  $M_2$  variedades diferenciáveis de dimensões  $n$  e  $m$ , respectivamente, e seja  $f : M_1 \rightarrow M_2$  uma aplicação diferenciável. Para cada  $p \in M_1$  e cada  $\mathbf{v} \in T_p M_1$ , escolha uma curva diferenciável  $\alpha : (-\epsilon, \epsilon) \rightarrow M_1$  com  $\alpha(0) = p$ ,  $\alpha'(0) = \mathbf{v}$ . Faça  $\beta = f \circ \alpha$ . A aplicação  $df_p : T_p M_1 \rightarrow T_{f(p)} M_2$  dada por  $df_p(\mathbf{v}) = \beta'(0)$  é uma aplicação linear que não depende da escolha de  $\alpha$ .

**Definição 1.2.8.** A aplicação linear  $df_p$  dada pela Proposição 1.2.7 é chamada diferencial de  $f$  em  $p$ .

A seguir introduzimos o conceito de subvariedade.

**Definição 1.2.9.** (Imersão e Mergulho) Sejam  $M$  e  $N$  variedades diferenciáveis de dimensões  $m$  e  $n$ , respectivamente. Uma aplicação diferenciável  $f : M \rightarrow N$  é uma imersão quando  $df_p : T_p M \rightarrow T_p N$  é injetiva para todo  $p \in M$ . Se, além disso,  $f$  é um homeomorfismo de  $M$  sobre o subespaço  $f(M) \subset N$  dizemos que  $f$  é um mergulho.

**Definição 1.2.10.** (Subvariedade) Seja  $N$  uma variedade de dimensão  $n$ . Quando  $M \subset N$  e a inclusão  $i : M \rightarrow N$  é um mergulho, dizemos que  $M$  é uma subvariedade de  $N$ .

## 1.3 Variedades Riemannianas

Assim como fizemos para superfícies, quando definimos a primeira forma fundamental, vamos definir um produto interno numa variedade diferenciável  $M$ . A definição de um produto interno em  $M$  nos permite medir comprimento de vetores tangentes a  $M$ , comprimento de uma curva em  $M$  e também definir geodésicas em  $M$ .

**Definição 1.3.1.** (Variedade Riemanniana) Uma variedade Riemanniana é uma variedade diferenciável  $M$  de dimensão  $n$  que associa para cada ponto  $p \in M$ , um produto interno  $\langle \cdot, \cdot \rangle_p$ , isto é,

uma forma bilinear simétrica definida positiva no espaço tangente  $T_pM$ , que varia diferenciavelmente no seguinte sentido: para alguma parametrização  $\varphi : U \subset \mathbb{R}^n \rightarrow M$  em torno de  $p$ , dado  $\varphi(x_1, x_2, \dots, x_n) = q$ , as funções

$$g_{ij}(x_1, \dots, x_n) = \left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q, \quad i, j = 1, \dots, n, \quad (1.3.1)$$

são diferenciáveis em  $U$ , onde  $\frac{\partial}{\partial x_i}(q) = d\varphi_q \cdot (0, \dots, 1, \dots, 0)$  e o 1 está na  $i$ -ésima posição. A família diferenciável  $\{\langle \cdot, \cdot \rangle_p, p \in M\}$  é chamada de estrutura Riemanniana (ou métrica Riemanniana) sobre  $M$ .

Podemos representar a métrica Riemanniana  $g$ , pela matriz  $G_p = [g_{ij}(p)]$ . Assim, dados  $\mathbf{u} = (u_1, \dots, u_n)$ ,  $\mathbf{v} = (v_1, \dots, v_n) \in T_pM$ , o produto interno associado a essa matriz é dado por

$$\langle \mathbf{u}, \mathbf{v} \rangle_p = \mathbf{u}^T G_p \mathbf{v}.$$

Além disso, o elemento de comprimento  $ds$  na métrica  $G_p$  satisfaz

$$ds^2 = \sum_{i,j=1}^n g_{ij}(p) dx_i dx_j.$$

**Definição 1.3.2.** (Isometria) Sejam  $M$  e  $N$  variedades Riemannianas. Um difeomorfismo  $f : M \rightarrow N$  (isto é,  $f$  é um bijeção diferenciável com inversa diferenciável) é chamado uma isometria quando

$$\langle \mathbf{u}, \mathbf{v} \rangle_p = \langle df_p(\mathbf{u}), df_p(\mathbf{v}) \rangle_{f(p)} \quad (1.3.2)$$

para todo  $p \in S$  e  $\mathbf{u}, \mathbf{v} \in T_pM$ .

Vamos mostrar agora como uma métrica Riemanniana pode ser usada para calcular comprimentos de curvas.

**Definição 1.3.3.** Uma aplicação diferenciável  $\alpha : I \rightarrow M$  de um intervalo aberto  $I \subset \mathbb{R}$  em uma variedade diferenciável  $M$  chama-se curva (parametrizada).

**Definição 1.3.4.** Um campo vetorial  $\vec{V}$  ao longo de uma curva  $\alpha : I \rightarrow M$  é uma aplicação que a cada  $t \in I$  associa um vetor tangente  $\vec{V}(t) \in T_{\alpha(t)}M$ . O campo vetorial  $\vec{V}$  é diferenciável em  $t_0 \in I$  quando, para alguma parametrização  $\varphi : U \subset \mathbb{R}^n \rightarrow S$  em  $\alpha(t_0)$ , as componentes  $v_i : I \rightarrow \mathbb{R}$ ,  $i = 1, 2, \dots, n$  de

$$\vec{V}(t) = \sum_{i=1}^n v_i(t) \frac{\partial}{\partial x_i},$$

na base  $\left\{ \left( \frac{\partial}{\partial x_1} \right)_0, \dots, \left( \frac{\partial}{\partial x_n} \right)_0 \right\}$  são funções diferenciáveis de  $t$  em  $t_0$ . O campo de vetores  $\vec{V}$  é diferenciável em  $I$  quando é diferenciável para todo  $t \in I$ .

Se uma curva  $\alpha$  está restrita a um intervalo fechado  $[a, b] \in I$ , seu comprimento é dado por

$$\ell(\alpha) = \int_a^b \left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle^{\frac{1}{2}} dt.$$

Vamos generalizar o conceito de geodésica dado para superfícies regulares. Para isso é preciso ter a noção do que é uma derivada covariante para variedades Riemannianas. Essa noção é mais complexa que a dada para superfícies regulares e requer um estudo mais aprofundado sobre geometria Riemanniana. A seguir damos apenas a sua definição, para saber mais sobre essa assunto ver referência [9] no Capítulo 2.

Sejam  $\alpha : I \rightarrow M$  uma curva diferenciável e  $\varphi : U \subset \mathbb{R}^n \rightarrow M$  um sistema de coordenadas para  $M$  com  $\alpha(I) \cap \varphi(U) \neq \emptyset$ . Dado  $t \in I$ , a expressão local de  $\alpha(t)$  é dada por  $\alpha(t) = (x_1(t), \dots, x_n(t))$ . Seja  $\vec{V}$  um campo de vetores ao longo da curva  $\alpha$ . Podemos expressar o campo  $\vec{V}$  localmente como

$$\vec{V} = \sum_j v_j X_j,$$

$j = 1, \dots, n$ , onde  $v_j = v_j(t)$  e  $X_j = \frac{\partial}{\partial x_j}(\alpha(t))$ . A derivada covariante  $\frac{D\vec{V}}{dt}$  de  $\vec{V}$  ao longo de  $\alpha$  para variedades Riemannianas no sistema de coordenadas  $(U, \varphi)$  é dada por

$$\frac{D\vec{V}}{dt} = \sum_k \left\{ \frac{dv_k}{dt} + \sum_{i,j} \Gamma_{ij}^k v_j \frac{dx_i}{dt} \right\} X_k.$$

Os coeficientes  $\Gamma_{ij}^k$  são funções diferenciáveis definidas em  $U$  (os símbolos de Christoffel de  $M$  na parametrização  $\varphi$ ). Sendo  $g_{ij} = \langle X_i, X_j \rangle$ , se pensarmos como uma matriz e escrevermos a sua inversa como  $g^{ij}$ , os símbolos de Christoffel são definidos por

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l \left\{ \frac{\partial}{\partial x_j} g_{il} + \frac{\partial}{\partial x_i} g_{jl} - \frac{\partial}{\partial x_l} g_{ij} \right\} g^{kl}.$$

**Definição 1.3.5.** (Geodésica) Uma curva parametrizada  $\gamma : I \rightarrow M$  é uma geodésica em  $t_0 \in I$  quando

$$\frac{D}{dt} \left( \frac{d\gamma}{dt} \right) = 0$$

no ponto  $t_0$ ; quando  $\gamma$  é geodésica em  $t$ , para todo  $t \in I$ , dizemos que  $\gamma$  é uma geodésica.

Vamos agora determinar as equações locais satisfeitas por uma geodésica  $\gamma$  em um sistemas de coordenadas  $(U, \varphi)$  em torno de  $t_0$ . Em  $U$ ,

$$\gamma(t) = (x_1(t), \dots, x_n(t)),$$

$\gamma$  será uma geodésica se, e somente se,

$$0 = \frac{D}{dt} \left( \frac{d\gamma}{dt} \right) = \sum_k \left( \frac{d^2 x_k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx_i}{dt} \frac{dx_j}{dt} \right) \frac{\partial}{\partial x_k}.$$

Logo o sistema de equações diferenciais de 2ª ordem

$$\frac{d^2 x_k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{dx_i}{dt} \frac{dx_j}{dt} = 0, k = 1, \dots, n, \quad (1.3.3)$$

fornece as equações procuradas.

**Definição 1.3.6.** (Variedade Totalmente Geodésica) Uma subvariedade  $M$  de uma variedade Riemanniana  $N$  é dita totalmente geodésica quando toda geodésica de  $M$  é geodésica de  $N$ .

As propriedades de minimização das geodésicas de superfícies regulares dadas pelas proposições 1.1.22 e 1.1.23 também são válidas para variedades Riemannianas, por proposições similares, que podem ser encontradas na referência [9].

**Definição 1.3.7.** (Distância) Dados dois pontos  $p, q \in M$ , a distância (intrínseca)  $d(p, q)$  desses pontos é dada por

$$d(p, q) = \inf \{ \ell(\alpha); \alpha : [p, q] \rightarrow M \text{ é curva diferenciável} \}.$$

**Proposição 1.3.8.** Com a distância  $d$ ,  $M$  é um espaço métrico, isto é, dados  $p, q, r \in M$ .

1.  $d(p, r) \leq d(p, q) + d(q, r)$ ,
2.  $d(p, q) = d(q, p)$ ,
3.  $d(p, q) \geq 0$ , e  $d(p, q) = 0 \Leftrightarrow p = q$ .

**Exemplo 1.3.9.** A métrica Euclidiana em  $\mathbb{R}^2$  é uma métrica Riemanniana, dado  $p = (x, y) \in \mathbb{R}^2$ :

$$g_{ij}(x, y) = \langle e_i, e_j \rangle$$

$$G(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Assim, dados  $\mathbf{u} = (u_1, u_2), \mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ ,

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2.$$

**Exemplo 1.3.10.** (Espaço hiperbólico) O espaço hiperbólico de dimensão  $n + 1$  é o semi-espaço do  $\mathbb{R}^{n+1}$  dado por

$$\mathbb{H}^{n+1} = \{ (x_0, \dots, x_n) \in \mathbb{R}^{n+1}; x_n > 0 \}.$$

Nesse espaço, para cada ponto  $p \in \mathbb{H}^{n+1}$ ,  $p = (x_0, \dots, x_n)$ , associamos o seguinte produto interno:

$$\langle \cdot, \cdot \rangle_p : T_p \mathbb{H}^{n+1} \rightarrow T_p \mathbb{H}^{n+1}$$

dado por

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{x_n^2} \sum_{i=0}^n u_i v_i,$$

onde  $\mathbf{u} = (u_0, \dots, u_n)$  e  $\mathbf{v} = (v_0, \dots, v_n)$ .

Consideremos  $\mathbb{H}^2$ , o plano hiperbólico, ou seja,  $\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2; y > 0\}$  (também conhecido como plano de Poincaré). A métrica Riemanniana nesse espaço é dada por

$$G(x, y) = \frac{1}{y^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

E, dados dois pontos  $\mathbf{x} = (x_1, x_2)$  e  $\mathbf{y} = (y_1, y_2)$  em  $\mathbb{H}^2$  uma expressão analítica para distância é

$$d_{\mathbb{H}}(\mathbf{x}, \mathbf{y}) = \log \frac{|\mathbf{x} - \bar{\mathbf{y}}| + |\mathbf{x} + \mathbf{y}|}{|\mathbf{x} - \bar{\mathbf{y}}| - |\mathbf{x} + \mathbf{y}|}, \quad (1.3.4)$$

onde  $\bar{\mathbf{y}} = (y_1, -y_2)$  e  $|\mathbf{z}| = \sqrt{z_1^2 + z_2^2}$ , para  $\mathbf{z} = (z_1, z_2)$ .

As geodésicas de  $\mathbb{H}^2$  são as semirretas verticais  $\gamma_1 : ]0, \infty[ \rightarrow \mathbb{H}^2$  e as semicircunferências Euclidianas  $\gamma_2 : ]0, \pi[ \rightarrow \mathbb{H}^2$  de centro  $(c, 0)$  e raio  $\rho$  euclidianos dadas por:

$$\gamma_1(r) = (x_0, r) \quad \text{e} \quad \gamma_2(r) = (\rho \cos(r) + c, \rho \sin(r)).$$

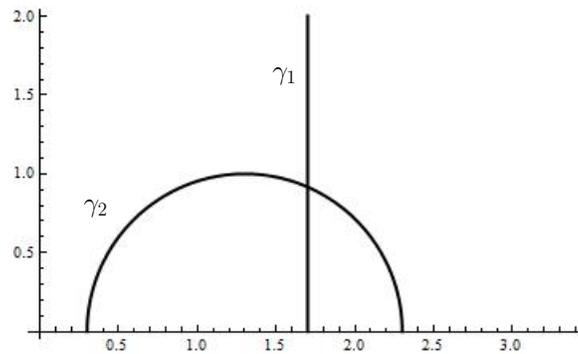


Figura 1.10: Geodésicas de  $\mathbb{H}^2$ .

# Capítulo 2

## Métrica de Fisher

Neste capítulo definimos a métrica de Fisher e calculamos a distância entre algumas distribuições de probabilidade. Como trabalhamos com um espaço formado por distribuições de probabilidade, na Seção 1 damos algumas definições de estatística. Na Seção 2, apresentamos o conceito de variedade estatística e definimos uma métrica nessa variedade, a métrica de Fisher. Na última seção, damos exemplos de como calcular a distância em alguns espaços compostos de distribuições de probabilidade.

### 2.1 Conceitos Básicos de Estatística

A base da estatística é a análise de experimentos aleatórios. Um experimento aleatório é aquele que quando repetido, em iguais condições, pode fornecer resultados diferentes, ou seja, são resultados explicados ao acaso. Para um estudo mais aprofundando sobre esses conceitos estatísticos ver referência [14].

**Definição 2.1.1.** (Espaço Amostral) Um espaço amostral  $\Omega$  de um experimento aleatório é o conjunto de todos os resultados possíveis desse experimento.

**Definição 2.1.2.** (Variável Aleatória) Uma variável aleatória (v.a.)  $X$  é uma função que associa a cada elemento do espaço amostral  $\Omega$  um número real. O espaço de  $X$  é o conjunto de números reais  $\mathcal{X} = \{x; x = X(\omega), \omega \in \Omega\}$ .

1. Se  $\mathcal{X}$  é um conjunto enumerável de pontos da reta,  $X$  é dita uma variável aleatória discreta.
2. Se  $\mathcal{X}$  é um intervalo ou uma coleção de intervalos da reta real,  $X$  é dita variável aleatória contínua.

**Definição 2.1.3.** (Distribuição de Probabilidade) Quando  $X$  é uma v.a. discreta, que assume valores reais  $x_1, x_2, \dots$ , uma distribuição de probabilidade sobre  $X$  é uma função  $p : \mathbb{R} \rightarrow \mathbb{R}$  que satisfaz

$$0 \leq p(x) \leq 1, \forall x \in \mathcal{X} \quad \text{e} \quad \sum_{x \in \mathcal{X}} p(x) = 1. \quad (2.1.1)$$

Quando  $X$  é uma v.a. contínua em  $\mathbb{R}$  então uma distribuição de probabilidade sobre  $X$  é uma função  $p : \mathbb{R} \rightarrow \mathbb{R}$  que satisfaz

$$p(x) \geq 0, \forall x \in X \quad \text{e} \quad \int_{-\infty}^{\infty} p(x)dx = 1. \quad (2.1.2)$$

**Definição 2.1.4.** (Esperança) Seja  $X$  uma v.a. com distribuição de probabilidade  $p$ . Definimos o valor esperado, a esperança matemática, ou a média de  $X$  por

$$E(X) = \sum_{x \in \mathcal{X}} xp(x), \quad (2.1.3)$$

quando  $X$  é uma v.a. discreta e por

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx, \quad (2.1.4)$$

quando  $X$  é uma v.a. contínua.

Generalizando, quando  $f$  é uma função que assume valores sobre  $X$  então

$$E(f(X)) = \sum_{x \in \mathcal{X}} f(x)p(x) \quad \text{e} \quad E(f(X)) = \int_{-\infty}^{\infty} f(x)p(x)dx, \quad (2.1.5)$$

nos casos discreto e contínuo, respectivamente.

**Definição 2.1.5.** (Variância e Desvio Padrão) Chamamos de variância de uma v.a.  $X$  o valor

$$V(X) = \sum_{x \in \mathcal{X}} (x - E(X))^2 p(x), \quad (2.1.6)$$

quando  $X$  é discreta e, quando  $X$  é contínua, o valor

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x)dx. \quad (2.1.7)$$

O desvio padrão de uma v.a.  $X$  é definido por

$$DP(X) = \sqrt{V(X)}. \quad (2.1.8)$$

**Exemplo 2.1.6.** (Distribuição de Bernoulli) Seja  $X$  variável aleatória,  $X$  tem uma distribuição de Bernoulli se  $\mathcal{X} = \{0, 1\}$  e sua distribuição de probabilidade é dada por

$$p(x) = \rho^x (1 - \rho)^{1-x}, \quad (2.1.9)$$

onde  $0 < \rho < 1$ .

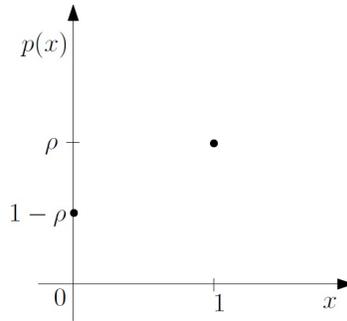


Figura 2.1: Distribuição Bernoulli.

A esperança e a variância de  $X$  são

$$E(X) = \sum_{x \in \mathcal{X}} xp(x) = 0 \cdot p(0) + 1 \cdot p(1) = \rho,$$

$$V(X) = \sum_{x \in \mathcal{X}} (x - E(X))^2 p(x) = (0 - \rho)^2 p(0) + (1 - \rho)^2 p(1)$$

$$= \rho^2(1 - \rho) + (1 - \rho)^2 \rho = \rho(1 - \rho).$$

**Exemplo 2.1.7.** (Distribuição Uniforme) A v.a.  $X$  tem distribuição uniforme no intervalo  $[\alpha, \beta]$  quando sua distribuição de probabilidade é dada por

$$p(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{se } \alpha \leq x \leq \beta \\ 0, & \text{caso contrário} \end{cases}. \quad (2.1.10)$$

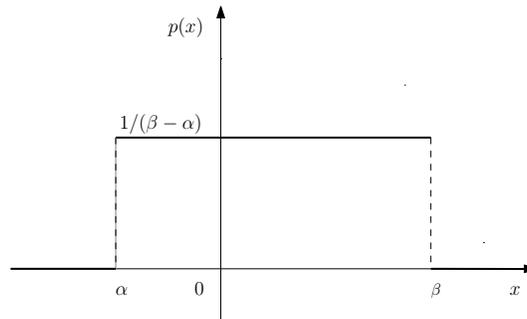


Figura 2.2: Distribuição Uniforme.

Temos que a variância e a esperança de  $X$  são dadas por

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\alpha} xp(x)dx + \int_{\alpha}^{\beta} xp(x)dx + \int_{\beta}^{\infty} xp(x)dx$$

$$= \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha} dx = \frac{\alpha + \beta}{2},$$

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x)dx = \int_{\alpha}^{\beta} \left(x - \frac{\alpha + \beta}{2}\right)^2 \frac{1}{\beta - \alpha} dx = \frac{(\beta - \alpha)^2}{12}.$$

As definições dadas até agora se baseiam no conceito de variáveis aleatórias. Podemos generalizá-las apresentando o conceito de vetores aleatórios.

**Definição 2.1.8.** (Vetores Aleatórios) Um vetor aleatório  $\mathbf{X}$  de dimensão  $n$  é uma aplicação que associa a cada elemento do espaço amostral  $\Omega$  um ponto no espaço  $\mathbb{R}^n$ . Dadas  $n$  variáveis aleatórias  $X_1, X_2, \dots, X_n$ ,

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}. \quad (2.1.11)$$

O espaço de  $\mathbf{X}$  é o conjunto  $\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_n); x_1 = X_1(\omega), \dots, x_n = X_n(\omega), \omega \in \Omega\}$ .

**Definição 2.1.9.** Seja  $\mathbf{X}$  um vetor aleatório, uma distribuição de probabilidade sobre  $\mathbf{X}$  é uma aplicação  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  que satisfaz

$$p(\mathbf{x}) \geq 0 \ (\forall \mathbf{x} \in \mathbf{X}) \quad \text{e} \quad \int_{\mathcal{X}} p(\mathbf{x}) d\mathbf{x} = 1, \quad (2.1.12)$$

onde  $d\mathbf{x}$  é o elemento de volume definido por  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$ .

**Definição 2.1.10.** Seja  $\mathbf{X} \in \mathbb{R}^n$  um vetor aleatório. A esperança e a variância de  $\mathbf{X}$  são os vetores  $E(\mathbf{X})$  e  $V(\mathbf{X})$  dados por

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} \quad \text{e} \quad V(\mathbf{X}) = \begin{pmatrix} V(X_1) \\ V(X_2) \\ \vdots \\ V(X_n) \end{pmatrix}. \quad (2.1.13)$$

Agora, dada uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e o vetor aleatório  $\mathbf{X} \in \mathbb{R}^n$ , com distribuição de probabilidade  $p$ , a esperança de  $f(\mathbf{X})$  é dada por

$$E(f(\mathbf{X})) = \int_{\mathcal{X}} f(\mathbf{X}) p(\mathbf{x}) d\mathbf{x}. \quad (2.1.14)$$

**Definição 2.1.11.** (Matriz de Covariância) A matriz de covariância de um vetor aleatório  $\mathbf{X} \in \mathbb{R}^n$  é uma matriz de ordem  $n$  definida por

$$\Sigma = \begin{pmatrix} E((X_1 - E(X_1))^2) & \cdots & E((X_1 - E(X_1))(X_n - E(X_n))) \\ E((X_2 - E(X_2))(X_1 - E(X_1))) & \cdots & E((X_2 - E(X_2))(X_n - E(X_n))) \\ \vdots & \ddots & \vdots \\ E((X_n - E(X_n))(X_1 - E(X_1))) & \cdots & E((X_n - E(X_n))^2) \end{pmatrix}. \quad (2.1.15)$$

**Exemplo 2.1.12.** (Distribuição Normal Bivariada) Um vetor aleatório  $\mathbf{X} \in \mathbb{R}^2$  tem uma distribuição normal bivariada com média  $\boldsymbol{\mu}$  e covariância  $\Sigma$ ,  $N(\boldsymbol{\mu}, \Sigma)$ , quando a sua distribuição de probabilidade definida por

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi \sqrt{\text{Det}(\Sigma)}} \exp \left( -\frac{(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2} \right) \quad (2.1.16)$$

onde  $\mathbf{x}^t = (x_1, x_2)$ ,  $\boldsymbol{\mu}^t = (\mu_1, \mu_2)$  é o vetor de médias (esperança) e  $\Sigma$  é a matriz de covariância de ordem 2 simétrica definida positiva. Note que a notação  $\mathbf{v}^t$  significa o transposto do vetor  $\mathbf{v}$ . A Figura 2.1.12 ilustra uma distribuição normal bivariada onde  $\boldsymbol{\mu}$  é o vetor nulo de dimensão dois e  $\Sigma$  é a matriz identidade de ordem dois,  $N(\mathbf{0}, I)$ .

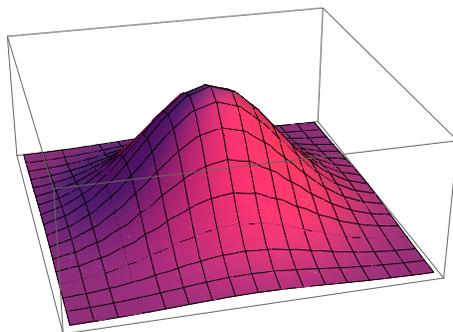


Figura 2.3: Distribuição normal bivariada  $N(\mathbf{0}, I)$ .

## 2.2 A Métrica de Fisher

Em 1945, Rao, ver referência [20], propôs um método para calcular a distância entre distribuições de probabilidade introduzindo uma métrica Riemanniana em termos da chamada matriz de informação de Fisher em uma família paramétrica de distribuições de probabilidade como vemos abaixo.

**Definição 2.2.1.** (Modelo estatístico) Seja  $S$  uma família de distribuições de probabilidade sobre a v. a.  $X$ . Suponha que cada elemento de  $S$ , uma distribuição de probabilidade, seja parametrizado por  $n$  variáveis reais  $(\theta_1, \theta_2, \dots, \theta_n)$ , isto é,

$$S = \{p_\theta = p(x; \theta); \theta = (\theta_1, \theta_2, \dots, \theta_n) \in \Theta\},$$

onde  $\Theta$ , o espaço dos parâmetros (ou espaço paramétrico), é um subconjunto do  $\mathbb{R}^n$  (ou de um espaço isomorfo ao  $\mathbb{R}^n$ ) e a aplicação  $\theta \mapsto p_\theta$  é injetiva. Chamamos  $S$  de modelo estatístico de dimensão  $n$ .

Para que possamos dar continuidade à teoria, vamos assumir algumas condições de regularidade com respeito a um modelo estatístico  $S = \{p_\theta; \theta \in \Theta\}$ . Para fazer livremente uma diferenciação com respeito a um parâmetro, assumimos que  $\Theta$  é um subconjunto aberto do  $\mathbb{R}^n$  e que, para cada valor  $x$  assumido pela v. a.  $X$ , a função  $\theta \mapsto p(x; \theta)$  que vai de  $\Theta$  em  $\mathbb{R}$  é contínua. Assim podemos assumir que expressões do tipo  $\frac{\partial p(x; \theta)}{\partial \theta_i}$  e  $\frac{\partial^2 p(x; \theta)}{\partial \theta_i \partial \theta_j}$  existem. Assumimos também que as ordens de integração e de diferenciação podem ser livremente trocadas, assim podemos fazer contas do tipo

$$\int \frac{\partial p(x; \theta)}{\partial \theta_i} dx = \frac{\partial}{\partial \theta_i} \int p(x; \theta) dx = \frac{\partial}{\partial \theta_i} (1) = 0.$$

**Observação 2.2.2.** Nos casos onde omitimos a região de integração o resultado vale independentemente dos valores que a v. a.  $X$  assume.

Dado um modelo estatístico  $S = \{p_\theta; \theta \in \Theta\}$  a aplicação  $\varphi : \Theta \rightarrow S$  dada por  $\varphi(\theta) = p_\theta$  é uma parametrização para  $S$ . Seja agora  $\psi$  um difeomorfismo injetivo de  $\Theta$  em  $\psi(\Theta)$  tal que  $\psi$  e  $\psi^{-1}$  sejam  $C^\infty$ , se usamos  $\xi = \psi(\theta)$  como nosso parâmetro em vez de  $\theta$ , obtemos  $S = \{p_{\psi^{-1}(\xi)}; \xi \in \psi(\Theta)\}$ . E essa é a mesma família de distribuição de probabilidade  $S = \{p_\theta; \theta \in \Theta\}$ .

Se tomarmos parametrizações que são difeomorfismos  $C^\infty$  então podemos considerar  $S$  uma variedade diferenciável  $C^\infty$ , a qual chamamos de variedade estatística. Ao longo do texto vamos nos referir à distribuição  $p_\theta$  como  $\theta$  e usar frases como "o ponto  $\theta$ " e "o espaço tangente  $T_\theta S$ ".

**Definição 2.2.3.** (Matriz de informação de Fisher) Seja  $S = \{p_\theta; \theta \in \Theta\}$  uma variedade estatística de dimensão  $n$ . Dado um ponto  $\theta \in \Theta$ , a matriz de informação de Fisher de  $S$  em  $\theta$  é a matriz  $G(\theta) = [g_{ij}(\theta)]$  de ordem  $n$ , tal que

$$g_{ij}(\theta) = E_\theta \left( \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) \right) = \int \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) p(x; \theta) dx, \quad (2.2.1)$$

onde  $E_\theta$  é a esperança com respeito à distribuição  $p_\theta$ . Quando  $n = 1$  chamamos o escalar  $G(\theta)$  de informação de Fisher.

Apesar de existirem modelos onde a integral da Equação (2.2.1) diverge, assumimos que  $g_{ij}$  é finito para todo  $\theta$  e, para todo  $i, j$ , que  $g_{ij} : \Theta \rightarrow \mathbb{R}$  é  $C^\infty$ .

Assim, o elemento infinitesimal  $ds$  da métrica dada pela matriz de informação de Fisher satisfaz

$$ds^2 = \sum_{i,j=1}^n g_{ij}(\theta) d\theta_i d\theta_j. \quad (2.2.2)$$

**Proposição 2.2.4.** A matriz  $G(\theta)$ , como definida acima, é simétrica semi-definida positiva.

**Demonstração:**  $G(\theta)$  é simétrica por definição. Além disso dado um vetor  $c^t = (c_1, \dots, c_n) \in \mathbb{R}^n$ ,

$$\begin{aligned} c^t G(\theta) c &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j g_{ij} = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \int \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) p(x; \theta) dx = \\ &= \int \sum_{i=1}^n \sum_{j=1}^n c_i c_j \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) p(x; \theta) dx = \int \left[ \sum_{i=1}^n c_i \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \right]^2 p(x; \theta) dx \geq 0. \end{aligned}$$

■

Nesse trabalho vamos assumir que  $G(\theta)$  é definida positiva.

A variedade estatística  $S$  munida da métrica dada pela matriz  $G(\theta)$  é uma variedade Riemanniana.

**Lema 2.2.5.** Cada entrada da matriz de informação de Fisher  $g_{ij}(\theta)$  pode ser escrita como

$$g_{ij}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\log p(x; \theta)) \right). \quad (2.2.3)$$

**Demonstração:** Observe que

$$0 = \int \frac{\partial}{\partial \theta_j} (p(x; \theta)) dx = \int p(x; \theta) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) dx.$$

Agora, derivando os dois lados em relação a  $\theta_i$ , temos

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta_i} \int p(x; \theta) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) dx \\ &= \int \frac{\partial}{\partial \theta_i} \left( p(x; \theta) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) \right) dx \\ &= \int \left( \frac{\partial}{\partial \theta_i} (p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) + p(x; \theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\log p(x; \theta)) \right) dx \\ &= \int p(x; \theta) \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) dx + \int p(x; \theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\log p(x; \theta)) dx \\ &= E \left( \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) \right) + E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\log p(x; \theta)) \right). \\ \therefore g_{ij}(\theta) &= E \left( \frac{\partial}{\partial \theta_i} (\log p(x; \theta)) \frac{\partial}{\partial \theta_j} (\log p(x; \theta)) \right) = -E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\log p(x; \theta)) \right). \end{aligned}$$

■

**Exemplo 2.2.6.** (Distribuição de Poisson) Uma distribuição de Poisson com parâmetro  $\theta$  é dada por

$$p(x; \theta) = \frac{e^{-\theta} \theta^x}{x!}, \quad (2.2.4)$$

onde  $x = 0, 1, \dots$ , e  $\theta > 0$

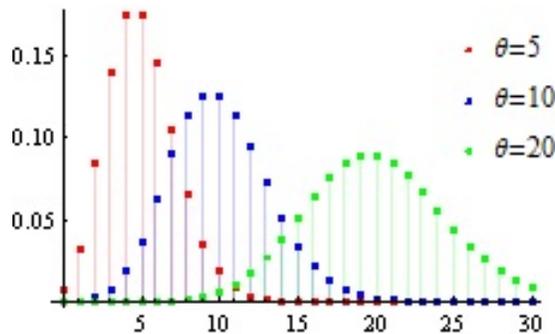


Figura 2.4: Distribuição de Poisson.

Sendo assim, seja  $S = \{p_\theta; \theta \in \mathbb{R}^+\}$  o espaço formado por distribuições de Poisson, temos que

a informação de Fisher de  $S$  é dada por

$$\begin{aligned} g_{11}(\theta) &= E_{\theta} \left[ \left( \frac{\partial(\log p(x; \theta))}{\partial \theta} \right)^2 \right] = \sum_{x=0}^{\infty} \left( \frac{\partial(\log p(x; \theta))}{\partial \theta} \right)^2 p(x; \theta) = \\ &= \sum_{x=0}^{\infty} \left( \frac{x}{\theta} - 1 \right)^2 \frac{e^{-\theta} \theta^x}{x!} = \frac{1}{\theta}. \end{aligned}$$

Agora, vamos dar a definição de família de distribuição exponencial, ver referência [19]. Essas distribuições merecem destaques pois para elas existe uma outra forma de calcular a matriz de Fisher.

**Definição 2.2.7.** (Família exponencial) Seja  $S = \{p_{\theta}; \theta \in \Theta\}$  um modelo estatístico de dimensão  $n$ . Se cada elemento de  $S$  pode ser escrito da forma

$$p(x; \theta) = \exp \left( \sum_{i=1}^n \theta_i t_i(x) - F(\theta) + k(x) \right),$$

$k(x) \geq 0, \forall x$ , então dizemos que  $S$  é uma família exponencial e que os  $\theta_i$ 's são os parâmetros naturais.

Muitas famílias de distribuições são exponenciais, como por exemplo, as distribuições Gaussianas, Poisson, Bernoulli, Multinomial, Laplaciana, Gama, Beta entre outras. Elas são caracterizadas por suas funções diferenciáveis  $F$  chamadas de *log-normalizer*. Podemos obter  $F$  da seguinte maneira: da definição de distribuição de probabilidade, temos que

$$\begin{aligned} 1 &= \int p(x; \theta) dx \\ &= \int \exp \left( \sum_{i=1}^n \theta_i t_i(x) - F(\theta) + k(x) \right) dx \\ &= \exp(-F(\theta)) \int \exp \left( \sum_{i=1}^n \theta_i t_i(x) + k(x) \right) dx, \end{aligned}$$

e logo

$$F(\theta) = \log \int \exp(\langle t(x), \theta \rangle + k(x)) dx. \quad (2.2.5)$$

A proposição abaixo relaciona a matriz de Fisher de uma dada distribuição com a sua função *log-normalizer*.

**Proposição 2.2.8.** Dado um modelo estatístico  $S = \{p_{\theta}; \theta \in \Theta\}$ , a matriz de informação de Fisher de  $S$  é a matriz Hessiana da *log-normalizer*.

**Demonstração:** Seja  $G(\theta) = [g_{ij}(\theta)]$  a matriz de informação de Fisher de  $S$  em  $\theta$ , pelo Lema 2.2.5, temos que

$$g_{ij}(\theta) = -E \left( \frac{\partial^2(\log p(x; \theta))}{\partial \theta_i \partial \theta_j} \right).$$

Além disso, observe que

$$\begin{aligned} p(x; \theta) &= \exp(\langle t(x), \theta \rangle - F(\theta) + k(x)) \\ \log p(x; \theta) &= \langle t(x), \theta \rangle - F(\theta) + k(x) \\ \frac{\partial(\log p(x; \theta))}{\partial \theta_j} &= t(x) - \frac{\partial F(\theta)}{\partial \theta_j} \\ \frac{\partial^2(\log p(x; \theta))}{\partial \theta_i \partial \theta_j} &= -\frac{\partial^2 F(\theta)}{\partial \theta_i \partial \theta_j}. \end{aligned}$$

Dessa forma,

$$\begin{aligned} g_{ij}(\theta) &= -E \left( \frac{\partial^2(\log p(x; \theta))}{\partial \theta_i \partial \theta_j} \right) \\ &= \int -\frac{\partial^2(\log p(x; \theta))}{\partial \theta_i \partial \theta_j} p(x; \theta) dx \\ &= \int -\left( -\frac{\partial^2 F(\theta)}{\partial \theta_i \partial \theta_j} \right) p(x; \theta) dx \\ &= \frac{\partial^2 F(\theta)}{\partial \theta_i \partial \theta_j} \int p(x; \theta) dx \\ &= \frac{\partial^2 F(\theta)}{\partial \theta_i \partial \theta_j}, \end{aligned}$$

onde a última igualdade vale pela Definição 2.1.3. Portanto

$$G(\theta) = \nabla^2 F(\theta). \quad (2.2.6)$$

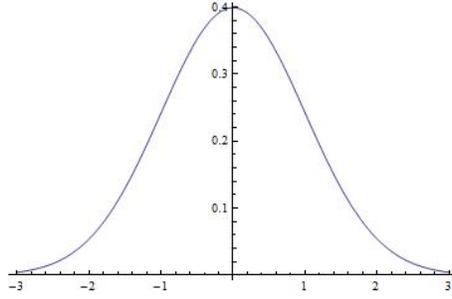
**Exemplo 2.2.9.** (Distribuição normal) Uma v.a.  $X$  tem distribuição normal univariada com média  $\mu$  e desvio padrão  $\sigma$ ,  $N(\mu, \sigma)$ , se a sua distribuição de probabilidade é definida por

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right), \quad (2.2.7)$$

onde  $\mu \in \mathbb{R}$  e  $\sigma \in (0, \infty)$ .

Seja  $S = \{p(x; \theta)\}$  o modelo estatístico formado por essas distribuições. Nesse caso o parâmetro  $\theta$  é dado por duas variáveis  $\theta = (\mu, \sigma) \in \mathbb{R} \times (0, +\infty)$ . A matriz de informação de Fisher de  $S$  em  $\theta = (\mu, \sigma)$  é dada por

$$\begin{aligned} g_{11}(\theta) &= \int_{-\infty}^{\infty} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \mu} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \mu} p(x; \mu, \sigma) dx \\ &= \int_{-\infty}^{\infty} \left(\frac{x - \mu}{\sigma^2}\right) \left(\frac{x - \mu}{\sigma^2}\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma^2} \end{aligned}$$


 Figura 2.5: Distribuição normal univariada,  $N(0,1)$ .

$$\begin{aligned}
 g_{12}(\theta) &= \int_{-\infty}^{\infty} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \mu} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \sigma} p(x; \mu, \sigma) dx \\
 &= \int_{-\infty}^{\infty} \frac{x - \mu}{\sigma^2} \left( \frac{(x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 g_{22}(\theta) &= \int_{-\infty}^{\infty} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \sigma} \frac{\partial \log(p(x; \mu, \sigma))}{\partial \sigma} p(x; \mu, \sigma) dx \\
 &= \int_{-\infty}^{\infty} \left( \frac{(x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} \right)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\
 &= \frac{2}{\sigma^2}.
 \end{aligned}$$

Portanto

$$G(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \quad (2.2.8)$$

A família das distribuições normais univariadas é uma família exponencial. Observe que podemos reescrever a Equação (2.2.7) da seguinte forma:

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right), \quad (2.2.9)$$

onde  $k(x) = 0$ ,  $t(x) = (t_1(x), t_2(x)) = (x, x^2)$ ,  $\vartheta = (\vartheta_1, \vartheta_2) = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$  e a *log-normalizer* é

$$F(\vartheta) = -\frac{\vartheta_1^2}{4\vartheta_2} + \frac{1}{2} \log\left(-\frac{\pi}{\vartheta_2}\right)$$

Logo, pela Proposição 2.2.8, temos que a matriz de informação de Fisher, com esses parâmetros, é dada por:

$$G(\vartheta) = \begin{pmatrix} -\frac{1}{2\vartheta_2} & \frac{\vartheta_1}{\vartheta_2^2} \\ \frac{\vartheta_1}{\vartheta_2^2} & \frac{1}{2} \left(\frac{\vartheta_1}{\vartheta_2}\right)^2 + \frac{1}{2\vartheta_2^2} \end{pmatrix}.$$

## 2.3 Distância entre Distribuições de Probabilidade

Nesta seção relacionamos a métrica de Fisher e a distância entre algumas distribuições de probabilidade.

Seja

$$\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$$

uma curva ligando dois pontos  $\theta_1$  e  $\theta_2$  em  $\Theta$ . Suponha que  $\alpha(t_1) = \theta_1$  e  $\alpha(t_2) = \theta_2$ , e logo, pela métrica de Fisher dada pela Equação (2.2.2) temos que o comprimento de arco da curva ao longo  $\alpha$  entre  $\theta_1$  e  $\theta_2$  é dado por

$$\ell(\alpha) = \int_{t_1}^{t_2} \left( \langle \alpha'(t), \alpha'(t) \rangle_{\alpha(t)} \right)^{\frac{1}{2}} dt = \int_{t_1}^{t_2} \left( \sum_{i,j=1}^n g_{ij}(\theta) \frac{d\alpha_i}{dt} \frac{d\alpha_j}{dt} \right)^{\frac{1}{2}} dt. \quad (2.3.1)$$

Vimos, no capítulo anterior, que a curva que minimiza esse comprimento é chamada curva geodésica e ela é dada pela solução do sistema de equações diferenciais

$$\frac{d^2\alpha_k}{dt^2} + \sum_{i,j=1}^n \Gamma_{ij}^k \frac{d\alpha_i}{dt} \frac{d\alpha_j}{dt} = 0, k = 1, \dots, n, \quad (2.3.2)$$

onde

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l \left\{ \frac{\partial}{\partial \alpha_j} g_{il} + \frac{\partial}{\partial \alpha_i} g_{jl} - \frac{\partial}{\partial \alpha_l} g_{ij} \right\} g^{kl}. \quad (2.3.3)$$

são os símbolos de Christoffel.

A distância geodésica entre  $\theta_1$  e  $\theta_2$  é a distância proposta por Rao entre duas distribuições de probabilidade parametrizadas por  $\theta_1$  e  $\theta_2$ . Chamamos essa distância de distância de Rao.

É muito difícil o cálculo da distância de Rao para grande parte das distribuições de probabilidade, pois envolve a solução de equações de diferenciais de 2ª ordem. Em alguns casos podemos simplificar o cálculo dessa distância relacionando a métrica do espaço com a de um espaço já conhecido. Por exemplo, podemos relacionar com a métrica do espaço Euclidiano, hiperbólico ou esférico. Atkinson e Mitchell em [3] e Burbea em [4] calcularam a distância de Rao para alguns espaços de distribuições conhecidos. A seguir, mostramos o cálculo dessa distância de Rao em alguns espaços.

### 2.3.1 Distribuições com apenas um parâmetro

Para distribuições de probabilidade com apenas um parâmetro é possível transformar a métrica do espaço na métrica de um espaço Euclidiano.

Seja uma distribuição de probabilidade com apenas um parâmetro  $\theta$ , temos que a métrica desse espaço é dada por

$$ds^2 = g_{11}(\theta)(d\theta)^2. \quad (2.3.4)$$

Considere a seguinte transformação isométrica

$$\theta = \theta(\lambda). \quad (2.3.5)$$

Em termos desse novo parâmetro, temos que

$$ds^2 = g_{11}(\lambda)(d\lambda)^2, \quad (2.3.6)$$

onde  $g_{11}(\lambda)$  é a informação de Fisher para o parâmetro  $\lambda$ . Como  $\theta$  é uma isometria, temos que as métricas dadas em (2.3.4) e (2.3.6) são iguais e, portanto,

$$g_{11}(\lambda) = g_{11}(\theta) \left( \frac{d\theta}{d\lambda} \right)^2.$$

Se a transformação em (2.3.5) é tal que

$$g_{11}(\lambda) = 1,$$

por (2.3.6), temos que

$$ds^2 = (d\lambda)^2,$$

que é a métrica do espaço Euclidiano. Dessa forma, a distância entre duas distribuições com parâmetros  $\lambda_1$  e  $\lambda_2$  é dada por

$$d(\lambda_1, \lambda_2) = |\lambda_1 - \lambda_2|.$$

Logo,

$$\lambda = \int \pm \sqrt{g_{11}(\theta)} d\theta + k,$$

onde  $k$  é uma constante arbitrária. Portanto segue que, em termos dos valores  $\theta_1$  e  $\theta_2$  de  $\theta$ ,

$$d(\theta_1, \theta_2) = \left| \int_{\theta_1}^{\theta_2} \sqrt{g_{11}(\theta)} d\theta \right|, \quad (2.3.7)$$

que não depende nem da constante arbitrária  $k$  e nem do sinal positivo ou negativo da raiz quadrada de  $g_{11}(\theta)$ .

**Exemplo 2.3.1.** Seja  $S$  a variedade estatística formada pelas distribuições de Poisson definidas no Exemplo 2.2.6. Vimos que a informação de Fisher de  $S$  é

$$g_{11}(\theta) = \frac{1}{\theta}.$$

Logo, dadas duas distribuições  $p(x; \theta_1)$  e  $p(x; \theta_2)$  em  $S$ , temos, pela Equação (2.3.7), que a distância entre essas distribuições é dada por

$$d(\theta_1, \theta_2) = \left| \int_{\theta_1}^{\theta_2} \sqrt{\frac{1}{\theta}} d\theta \right| = 2 \left| \sqrt{\theta_1} - \sqrt{\theta_2} \right|. \quad (2.3.8)$$

**Exemplo 2.3.2.** (Distribuição Binomial) Uma distribuição Binomial com parâmetros  $n$  e  $\rho$  é dada por

$$p(x; n, \rho) = \binom{n}{x} \rho^x (1 - \rho)^{n-x}, \quad x = 0, 1, \dots, n, \quad (2.3.9)$$

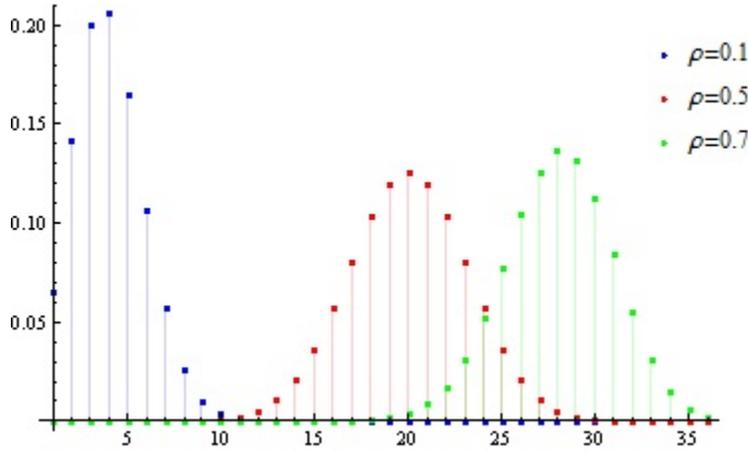


Figura 2.6: Distribuição Binomial.

onde  $n \geq 1$  e  $0 < \rho < 1$ .

Supondo o parâmetro  $n$  conhecido, a distribuição torna-se uma distribuição uniparamétrica, seja  $S = \{p_\rho; \rho \in (0, 1)\}$  o espaço formado por essas distribuições, temos que a informação de Fisher é

$$g_{11}(\rho) = E_\rho \left[ \left( \frac{\partial \log p(x; \rho)}{\partial \rho} \right)^2 \right] = \sum_{x=0}^n \left( \frac{x - \rho n}{\rho(1 - \rho)} \right)^2 p(x; \rho) = \frac{n}{\rho(1 - \rho)}. \quad (2.3.10)$$

Dessa forma, dadas duas distribuições binomiais  $p(x; \rho_1)$  e  $p(x; \rho_2)$  em  $S$ , temos, pela Equação (2.3.7), que a distância entre essas distribuições é dada por

$$d(\rho_1, \rho_2) = \left| \int_{\rho_1}^{\rho_2} \sqrt{\frac{n}{\rho(1 - \rho)}} d\rho \right| = 2\sqrt{n} |\arcsen(\sqrt{\rho_1}) - \arcsen(\sqrt{\rho_2})|. \quad (2.3.11)$$

**Exemplo 2.3.3.** (Distribuição exponencial) Uma distribuição exponencial com parâmetro  $\lambda$  é dada por

$$p(x; \theta) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}, \quad (2.3.12)$$

com  $\lambda > 0$ .

Sendo assim, seja  $S = \{p_\lambda; \lambda \in \mathbb{R}^+\}$  o espaço formado por distribuições exponenciais, a informação de Fisher é dada por

$$g_{11}(\theta) = E_\lambda \left[ \left( \frac{\partial \log p(x; \lambda)}{\partial \lambda} \right)^2 \right] = \int_{-\infty}^{\infty} \left( \frac{\partial \log p(x; \lambda)}{\partial \lambda} \right)^2 p(x; \lambda) dx = \quad (2.3.13)$$

$$= \int_0^{\infty} \left( \frac{1}{\lambda} - x \right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}, \quad (2.3.14)$$

e, dadas duas distribuições exponenciais  $p(x; \lambda_1)$  e  $p(x; \lambda_2)$ , a distância entre elas é

$$d(\lambda_1, \lambda_2) = \left| \int_{\lambda_1}^{\lambda_2} \sqrt{\frac{1}{\lambda^2}} d\lambda \right| = 2 \left| \log \left( \frac{\lambda_1}{\lambda_2} \right) \right|. \quad (2.3.15)$$

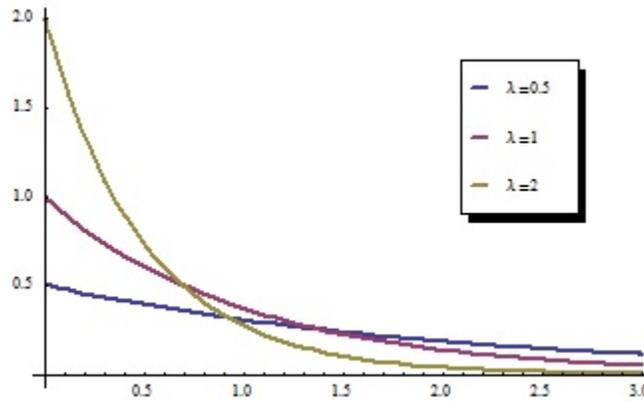


Figura 2.7: Distribuição exponencial.

### 2.3.2 Distribuições Normais Univariadas

Nesse caso, podemos relacionar a métrica do espaço com a métrica do espaço hiperbólico.

Vimos, no Exemplo 2.2.9, que a matriz de Fisher do espaço formado pelas distribuições normais univariadas com parâmetros  $\mu$  e  $\sigma$ ,  $N(\mu, \sigma)$ , é

$$G(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

Logo a expressão da métrica é dada por

$$ds^2 = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}. \tag{2.3.16}$$

Agora, dadas duas distribuições normais  $N(\mu_1, \sigma_1)$  e  $N(\mu_2, \sigma_2)$ , queremos saber qual a distância entre elas.

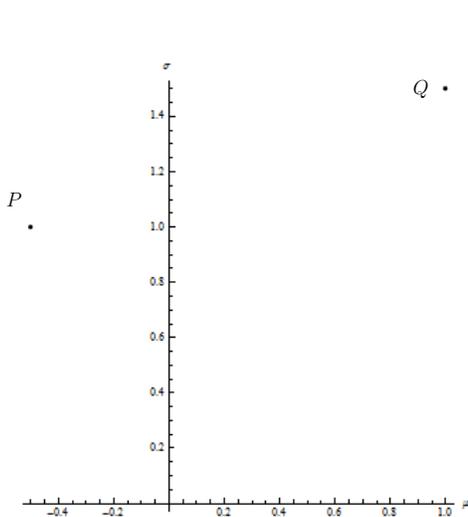


Figura 2.8: Plano  $\mu \times \sigma$ .

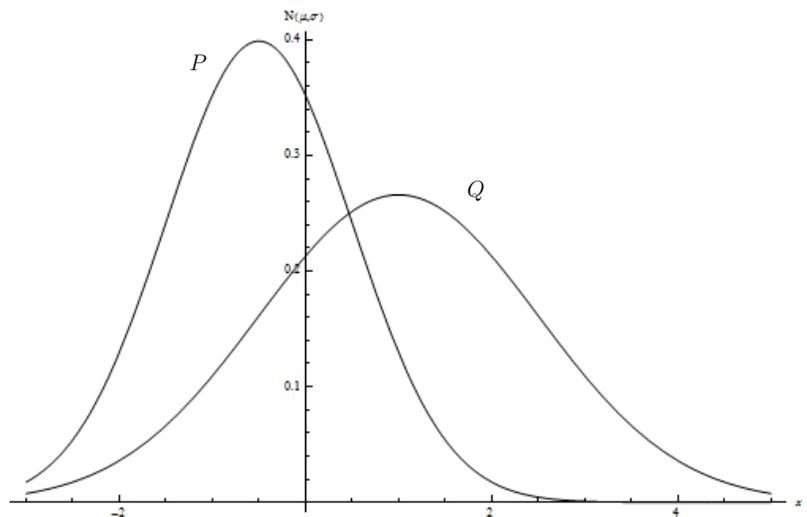


Figura 2.9: Distribuições  $N(-0.5, 1)$  e  $N(1, 1.5)$ .

As Figuras 2.8 e 2.9 relacionam os pontos  $P = (-0.5, 1)$  e  $Q = (1, 1.5)$  do espaço dos parâmetros dado pelo semiplano superior média  $\times$  desvio padrão,  $\mathbb{H}_F^2$ , com o gráfico das respectivas distribuições normais.

Observando a métrica de  $\mathbb{H}_F^2$ , vemos que é um modelo geométrico hiperbólico o qual pode ser relacionado com o plano de Poincaré,  $\mathbb{H}^2$ , visto no Exemplo 1.3.10. Esta relação é dada pela transformação

$$\begin{aligned} f : \mathbb{H}_F^2 &\rightarrow \mathbb{H}^2 \\ (\mu, \sigma) &\mapsto \left( \frac{\mu}{\sqrt{2}}, \sigma \right). \end{aligned} \quad (2.3.17)$$

Logo as distâncias de Fisher e Poincaré estão relacionadas por

$$d_{\mathbb{H}_F^2}((\mu_1, \sigma_1); (\mu_2, \sigma_2)) = \sqrt{2} d_{\mathbb{H}^2} \left( \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right); \left( \frac{\mu_2}{\sqrt{2}}, \sigma_2 \right) \right),$$

e pela Equação (1.3.4), temos que uma expressão analítica para  $d_{\mathbb{H}_F^2}$  é

$$d_{\mathbb{H}_F^2}((\mu_1, \sigma_1); (\mu_2, \sigma_2)) = \sqrt{2} \log \frac{\left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, -\sigma_2 \right) \right| + \left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, \sigma_2 \right) \right|}{\left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, -\sigma_2 \right) \right| - \left| \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2}}, \sigma_2 \right) \right|}. \quad (2.3.18)$$

As geodésicas de  $\mathbb{H}_F^2$  são as imagens inversas, por meio da transformação  $f$ , das geodésicas de  $\mathbb{H}^2$ . Essas geodésicas são as semirretas verticais positivas  $\gamma_1 : ]0, \infty[ \rightarrow \mathbb{H}_F^2$  e as semi-elipses  $\gamma_2 : ]0, \pi[ \rightarrow \mathbb{H}_F^2$  centradas em  $\sigma = 0$  com excentricidade  $\frac{1}{\sqrt{2}}$  dadas por

$$\gamma_1(r) = (\sqrt{2}\mu_0, r) \quad \text{e} \quad \gamma_2(r) = (\sqrt{2}(\rho \cos(r) + c), \rho \sin(r)). \quad (2.3.19)$$

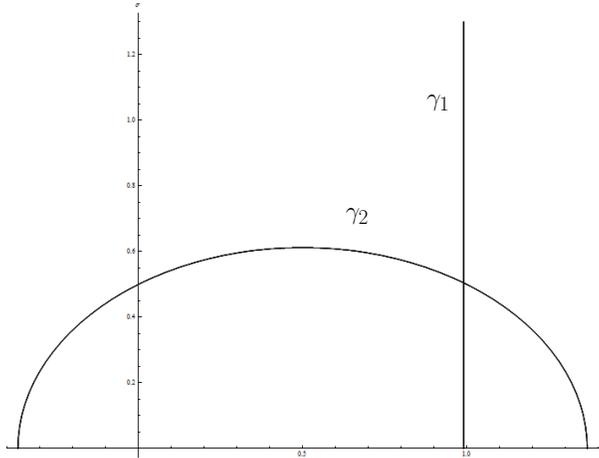


Figura 2.10: Geodésicas de  $\mathbb{H}_F^2$ .

A Figura 2.11 mostra a curva geodésica que liga os pontos  $P = (-0.5, 1)$  e  $Q = (1, 1.5)$  no plano  $\mathbb{H}_F^2$  e a Figura 2.12 mostra o gráfico de algumas distribuições que estão relacionadas com pontos dessa curva.

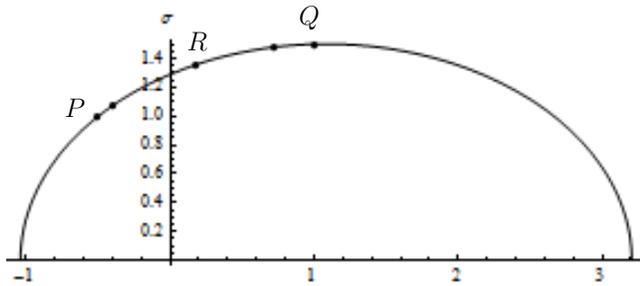


Figura 2.11: Curva geodésica ligando  $P$  e  $Q$ .

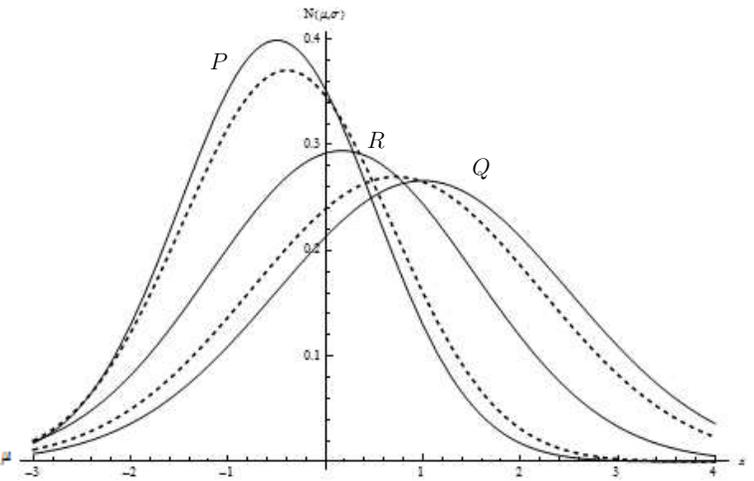


Figura 2.12: Gráfico das distribuições normais.

Considerando apenas a subvariedade  $\hat{S}$  de  $S$  formada pelas distribuições normais com média constante,  $\hat{S} = \{p(x; \mu, \sigma); \mu = \mu_0 \text{ constante}, \sigma \in (0, \infty)\}$ , temos que a distância entre duas distribuições dessa subvariedade parametrizadas por  $(\mu_0, \sigma_1)$  e  $(\mu_0, \sigma_2)$  é

$$d_{\mathbb{H}_F^2}((\mu_0, \sigma_1); (\mu_0, \sigma_2)) = \sqrt{2} \log \left( \frac{\sigma_2}{\sigma_1} \right).$$

A subvariedade  $\hat{S}$  é totalmente geodésica.

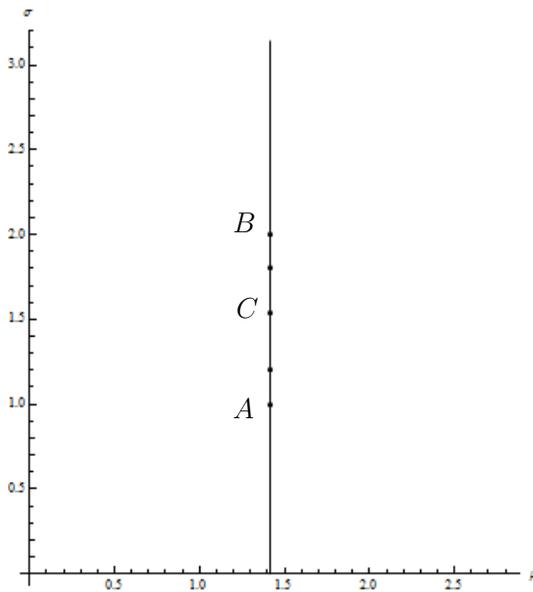


Figura 2.13: Reta ligando  $A$  e  $B$ .

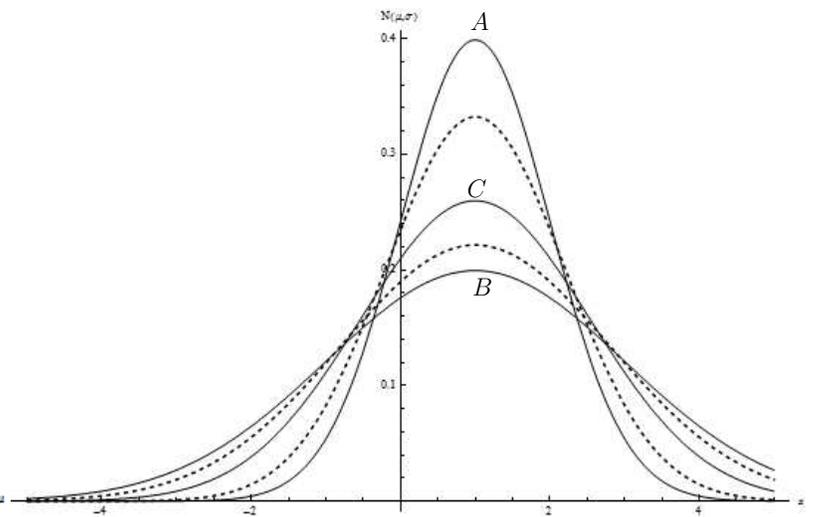


Figura 2.14: Gráfico das distribuições normais.

Considerando agora a subvariedade  $\bar{S}$  de  $S$  formada pelas distribuições normais desvio padrão constante,  $\bar{S} = \{p(x; \mu, \sigma); \sigma = \sigma_0 \text{ constante}, \mu \in \mathbb{R}\}$ , temos que a distância entre duas distribuições

desse espaço parametrizadas por  $(\mu_1, \sigma_0)$  e  $(\mu_2, \sigma_0)$  é

$$d_{\mathbb{H}_F^2}((\mu_1, \sigma_0); (\mu_2, \sigma_0)) = \frac{|\mu_1 - \mu_2|}{\sigma_0}.$$

Diferente de  $\hat{S}$ , a subvariedade  $\bar{S}$ , não é totalmente geodésica.

De fato, dados dois pontos  $M = (\mu_1, \sigma)$  e  $N = (\mu_2, \sigma)$ , temos que

$$d((\mu_1, \sigma), (\mu_2, \sigma)) = \sqrt{2} \log \left( \frac{(\mu_1 - \mu_2)^2 + 4\sigma^2 + |\mu_1 - \mu_2| \sqrt{(\mu_1 - \mu_2)^2 + 8\sigma^2}}{4\sigma^2} \right) < \frac{|\mu_1 - \mu_2|}{\sigma}.$$

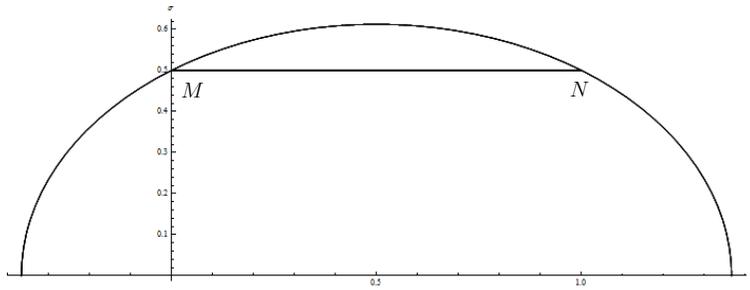


Figura 2.15: Distância não geodésica.

Podemos também expressar a distância em termos dos parâmetros naturais, como feito em [12]. Como visto no Exemplo 2.2.9, a relação entre os parâmetros média e desvio padrão com os parâmetros naturais é

$$(\theta_1, \theta_2) = \left( \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right).$$

Dessa forma, a distância de Rao nesses parâmetros é

$$d((\theta_{11}, \theta_{12}), (\theta_{21}, \theta_{22})) = d \left( \left( \frac{-\theta_{11}}{2\theta_{12}}, \frac{1}{\sqrt{-2\theta_{12}}} \right), \left( \frac{-\theta_{21}}{2\theta_{22}}, \frac{1}{\sqrt{-2\theta_{22}}} \right) \right) = \sqrt{2} \log \left( \frac{\sqrt{4 \left( \frac{1}{\sqrt{-\theta_{12}}} + \frac{1}{\sqrt{-\theta_{22}}} \right)^2 + \left( \frac{\theta_{11}}{\theta_{12}} - \frac{\theta_{21}}{\theta_{22}} \right)^2} + \sqrt{4 \left( \frac{1}{\sqrt{-\theta_{12}}} - \frac{1}{\sqrt{-\theta_{22}}} \right)^2 + \left( \frac{\theta_{11}}{\theta_{12}} - \frac{\theta_{21}}{\theta_{22}} \right)^2}}{\sqrt{4 \left( \frac{1}{\sqrt{-\theta_{12}}} + \frac{1}{\sqrt{-\theta_{22}}} \right)^2 + \left( \frac{\theta_{11}}{\theta_{12}} - \frac{\theta_{21}}{\theta_{22}} \right)^2} - \sqrt{4 \left( \frac{1}{\sqrt{-\theta_{12}}} - \frac{1}{\sqrt{-\theta_{22}}} \right)^2 + \left( \frac{\theta_{11}}{\theta_{12}} - \frac{\theta_{21}}{\theta_{22}} \right)^2}} \right). \quad (2.3.20)$$

As figuras abaixo ilustram as geodésicas no plano  $\theta_1 \times \theta_2$ .

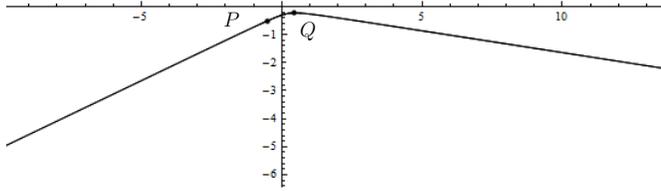


Figura 2.16: Curva geodésica ligando os pontos P e Q ilustrados na Fig. 2.11 no plano  $\theta_1 \times \theta_2$ .

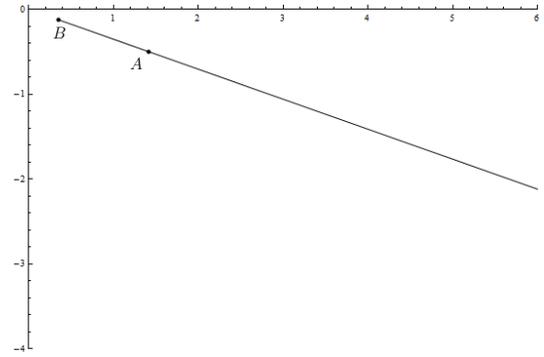


Figura 2.17: Curva geodésica ligando os pontos A e B ilustrados na Fig. 2.13 no plano  $\theta_1 \times \theta_2$ .

### 2.3.3 Distribuição Multinomial

Para obter a distância entre duas distribuições multinomiais reduzimos a métrica do espaço à métrica da geometria esférica.

Uma distribuição de probabilidade multinomial com índice  $n$  e parâmetros  $\theta_i$ , com  $i = 1, 2, \dots, k$ , denotada por  $M_k(n, \theta_1, \dots, \theta_k)$ , é dada por

$$p(x; \theta_1, \dots, \theta_k) = \frac{n!}{x_1! \dots x_k!} \theta_1^{x_1} \dots \theta_k^{x_k}, \quad x_i = 0, \dots, n, \quad \sum_{i=1}^k x_k = n,$$

onde  $0 < \theta_i < 1$  e  $\sum_{i=1}^k \theta_i = 1$ . Observe que a distribuição binomial é um caso particular da multinomial quando  $k = 2$ .

Podemos expressar a métrica em termos dos seguintes parâmetros

$$\begin{aligned} \beta_1 &= \frac{\theta_1}{\theta_k} \\ \beta_2 &= \frac{\theta_2}{\theta_1 + \theta_k} \\ &\vdots \\ \beta_{k-1} &= \frac{\theta_{k-1}}{\theta_1 + \theta_2 + \dots + \theta_{k-2} + \theta_k}. \end{aligned} \tag{2.3.21}$$

Assim

$$\begin{aligned} ds^2 &= \frac{n(d\beta_1)^2}{\beta_1(1 + \beta_1)^2(1 + \beta_2) \dots (1 + \beta_{k-1})} \\ &+ \frac{n(d\beta_2)^2}{\beta_2(1 + \beta_2)^2(1 + \beta_3) \dots (1 + \beta_{k-1})} + \dots + \frac{n(d\beta_{k-1})^2}{\beta_{k-1}(1 + \beta_{k-1})^2}. \end{aligned} \tag{2.3.22}$$

Agora, fazendo a transformação

$$\beta_j = \tan^2 \alpha_j, \quad j = 1, 2, \dots, k-1,$$

a métrica em (2.3.22) é escrita da seguinte forma

$$\frac{ds^2}{4n} = \prod_{i=2}^{k-1} \cos^2 \alpha_i (d\alpha_1)^2 + \prod_{i=3}^{k-1} \cos^2 \alpha_i (d\alpha_2)^2 + \dots + \cos^2 \alpha_{k-1} (d\alpha_{k-2})^2 + (d\alpha_{k-1})^2. \quad (2.3.23)$$

A Equação acima é a métrica ao longo de uma esfera de dimensão  $k$  em coordenadas esféricas, de raio 1 e ângulos  $\alpha_1, \alpha_2, \dots, \alpha_{k-1}$ . De fato, considere a transformação

$$\begin{aligned} \eta_1 &= \prod_{i=1}^{k-1} \cos \alpha_i \\ \eta_j &= \text{sen } \alpha_{j-1} \prod_{i=j}^{k-1} \cos \alpha_i, \quad j = 2, \dots, k-1, \\ \eta_k &= \text{sen } \alpha_{k-1}, \end{aligned} \quad (2.3.24)$$

logo, a métrica em (2.3.23) se reduz a

$$\frac{ds^2}{4n} = \sum_{i=1}^k (d\eta_i)^2 \quad (2.3.25)$$

sujeito a

$$\sum_{i=1}^k \eta_i^2 = 1. \quad (2.3.26)$$

Considere agora a distância entre duas distribuições multinomiais com índice  $n$  e parâmetros  $\theta_i^t = (\theta_{1i}, \theta_{2i}, \dots, \theta_{ki})$ , onde  $\sum_{j=1}^k \theta_{ij} = 1$  e  $i = 1, 2$ . Em termos dos parâmetros  $\eta_i^t = (\eta_{1i}, \eta_{2i}, \dots, \eta_{ki})$ ,  $i = 1, 2$ , correspondentes, segue de (2.3.25), (2.3.26) e da geometria da esfera, que a distância entre essas distribuições é dada por

$$d(\eta_1, \eta_2) = 2\sqrt{n} \arccos(\eta_1^t \eta_2).$$

Portanto, segue que a distância, em termos dos parâmetros originais  $\theta$  é

$$d(\theta_1, \theta_2) = 2\sqrt{n} \arccos \left( \sum_{j=1}^k \sqrt{\theta_{j1} \theta_{j2}} \right). \quad (2.3.27)$$

# Capítulo 3

## Distribuições Normais Multivariadas

Nesse capítulo vamos estudar a métrica de Fisher no espaço formado pelas distribuições normais multivariadas. Nesse espaço ainda não se conhece uma fórmula explícita para a distância de Rao no caso geral.

### 3.1 Métrica de Fisher

Uma distribuição normal multivariada é definida por

$$p(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{(2\pi)^{-\left(\frac{n}{2}\right)}}{\sqrt{\text{Det}(\Sigma)}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2}\right), \quad (3.1.1)$$

onde  $\mathbf{x}^t = (x_1, \dots, x_n)$ ,  $\boldsymbol{\mu}^t = (\mu_1, \dots, \mu_n)$  é o vetor de médias e  $\Sigma$  é a matriz de covariância de ordem  $n$  simétrica definida positiva.

Observe que a distribuição dada no Exemplo 2.1.12 é uma distribuição normal multivariada com  $n = 2$ .

Vamos introduzir algumas notações, sejam

- $M_n(\mathbb{R})$  o espaço das matrizes reais de ordem  $n$ ;
- $S_n(\mathbb{R})$  o subespaço das matrizes simétricas de  $M_n(\mathbb{R})$ ;
- $GL_n(\mathbb{R})$  o grupo das matrizes não-singulares sobre  $M_n(\mathbb{R})$ ;
- $P_n(\mathbb{R})$  o subconjunto das matrizes simétricas positivas definidas em  $GL_n(\mathbb{R})$ .

O produto interno e a norma em  $M_n(\mathbb{R})$  são dados por

$$\langle A, B \rangle = \text{tr}(AB^t) \quad \text{e} \quad \|A\| = \sqrt{\langle A, A \rangle}, \quad A, B \in M_n(\mathbb{R}).$$

Seja  $\mathcal{M} = \{p_\theta; \theta \in \Theta = \mathbb{R}^n \times P_n(\mathbb{R})\}$  o modelo estatístico formado por essas distribuições. Observe que  $\Theta$  é um espaço de dimensão  $\left(n + \frac{n(n+1)}{2}\right)$ .

A equação da métrica da informação de Fisher de  $\mathcal{M}$ , dada por Skovgaard, ver referência [22] é

$$ds^2 = d\boldsymbol{\mu}^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(\Sigma^{-1} d\Sigma)^2], \quad (3.1.2)$$

onde  $d\boldsymbol{\mu}^t = (d\mu_1, \dots, d\mu_n) \in \mathbb{R}^n$  e  $d\Sigma = [d\sigma_{ij}] \in P_n(\mathbb{R})$ .

Note que, para todo  $(\mathbf{c}, Q) \in \mathbb{R}^n \times GL_n(\mathbb{R})$ , a aplicação

$$(\boldsymbol{\mu}, \Sigma) \mapsto (\bar{\boldsymbol{\mu}}, \bar{\Sigma}), \quad (3.1.3)$$

onde  $\bar{\Sigma} = Q\Sigma Q^t$  e  $\bar{\boldsymbol{\mu}} = Q\boldsymbol{\mu} + \mathbf{c}$  estabelece um homeomorfismo de  $P_n(\mathbb{R})$  em  $P_n(\mathbb{R})$ . Esse homeomorfismo é também uma isometria com respeito à métrica dada em (3.1.2), de fato:

$$\begin{aligned} ds^2 &= (d\bar{\boldsymbol{\mu}})^t \bar{\Sigma}^{-1} d\bar{\boldsymbol{\mu}} + \frac{1}{2} \text{tr}[(\bar{\Sigma}^{-1} d\bar{\Sigma})^2] \\ &= (Qd\boldsymbol{\mu})^t (Q\Sigma Q^t)^{-1} (Qd\boldsymbol{\mu}) + \frac{1}{2} \text{tr}\{[(Q\Sigma Q^t)^{-1} (Qd\Sigma Q^t)]^2\} \\ &= (d\boldsymbol{\mu})^t Q^t Q^{-t} \Sigma^{-1} Q^{-1} Q d\boldsymbol{\mu} + \frac{1}{2} \text{tr}\{[Q^{-t} \Sigma^{-1} Q^{-1} Q d\Sigma Q^t]^2\} \\ &= (d\boldsymbol{\mu})^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[Q^{-t} (\Sigma^{-1} d\Sigma)^2 Q^t] \\ &= (d\boldsymbol{\mu})^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(\Sigma^{-1} d\Sigma)^2 Q^t Q^{-t}] \\ &= (d\boldsymbol{\mu})^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(\Sigma^{-1} d\Sigma)^2]. \end{aligned}$$

Consequentemente, a distância de Rao entre  $\theta_1 = (\boldsymbol{\mu}_1, \Sigma_1)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \Sigma_2)$  em  $\theta$  satisfaz

$$d(\theta_1, \theta_2) = d((\boldsymbol{\mu}_1, \Sigma_1); (\boldsymbol{\mu}_2, \Sigma_2)) = d((Q\boldsymbol{\mu}_1 + \mathbf{c}, Q\Sigma_1 Q^t); (Q\boldsymbol{\mu}_2 + \mathbf{c}, Q\Sigma_2 Q^t)),$$

para todo  $(\mathbf{c}, Q) \in \mathbb{R}^n \times GL_n(\mathbb{R})$ . Em particular, tomando  $Q = \Sigma_1^{-(1/2)}$  e  $\mathbf{c} = -\Sigma_1^{-(1/2)} \boldsymbol{\mu}_1$  (ver Observação 3.1.1) a distância de Rao admite a forma

$$d(\theta_1, \theta_2) = d((\mathbf{0}, I_n); (\Sigma_1^{-(1/2)}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \Sigma_1^{-(1/2)} \Sigma_2 \Sigma_1^{-(1/2)})). \quad (3.1.4)$$

onde  $\mathbf{0}$  é o vetor nulo de dimensão  $n$  e  $I_n$  é a matriz identidade de ordem  $n$ .

**Observação 3.1.1.** A raiz quadrada de uma matriz  $A$ ,  $A^{(1/2)}$  ou  $\sqrt{A}$ , é uma matriz  $X$  tal que  $XX = A$ . Se  $A$  é uma matriz simétrica definida positiva então  $A$  pode ser escrita como  $A = BDB^t$ , onde  $B$  é uma matriz ortogonal cujas colunas são os autovetores da matriz  $A$  e  $D$  é uma matriz diagonal formada pelos autovalores de  $A$ . Tome

$$X = BD^{(1/2)}B^t,$$

temos que

$$XX = BD^{(1/2)}B^t BD^{(1/2)}B^t = BD^{(1/2)}D^{(1/2)}B^t = BDB^t = A.$$

**Exemplo 3.1.2.** Dadas duas distribuições normais  $N_1(\boldsymbol{\mu}_1, \Sigma_1)$  e  $N_2(\boldsymbol{\mu}_2, \Sigma_2)$  parametrizadas por  $\theta_1$  e  $\theta_2$ , respectivamente, onde

$$\theta_1 = (\boldsymbol{\mu}_1, \Sigma_1) = \left( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.5 & 0.25 \\ 0.25 & 1.5 \end{pmatrix} \right),$$

$$\theta_2 = (\boldsymbol{\mu}_2, \Sigma_2) = \left( \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0.7 & 0.05 \\ 0.05 & 1 \end{pmatrix} \right).$$

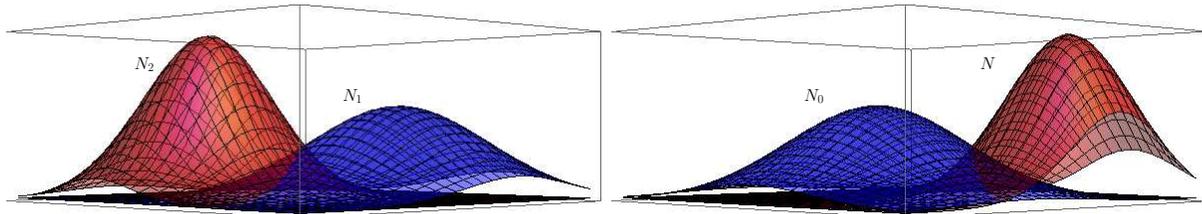
Temos que a distância entre  $N_1(\boldsymbol{\mu}_1, \Sigma_1)$  e  $N_2(\boldsymbol{\mu}_2, \Sigma_2)$  é igual à distância entre  $N_0(\mathbf{0}, I_2)$  e  $N(\bar{\boldsymbol{\mu}}, \bar{\Sigma})$  onde

$$\bar{\boldsymbol{\mu}} = -\Sigma_1^{-1/2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = \begin{pmatrix} 1.51 \\ 1.51 \end{pmatrix},$$

$$\bar{\Sigma} = \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} = \begin{pmatrix} 0.48 & -0.06 \\ -0.06 & 0.68 \end{pmatrix}.$$

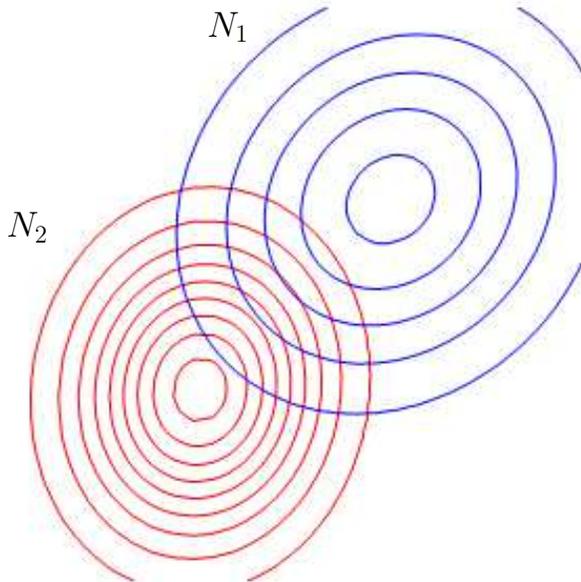
Ou seja, para  $\theta_0 = (\mathbf{0}, I_2)$  e  $\bar{\theta} = (\bar{\boldsymbol{\mu}}, \bar{\Sigma})$ ,

$$d(\theta_1, \theta_2) = d(\theta_0, \bar{\theta}).$$

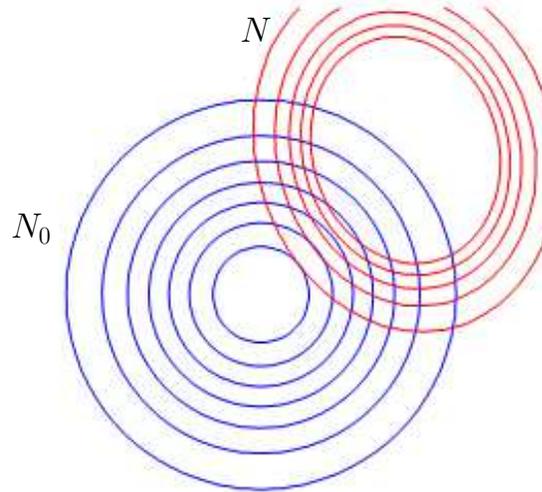


(a) Gráficos de  $N_1$  e  $N_2$ .

(b) Gráficos de  $N_0$  e  $N$ .



(c) Curvas de níveis de  $N_1$  e  $N_2$ .



(d) Curvas de níveis de  $N_0$  e  $N$ .

Seja  $\gamma : I \rightarrow \Theta$ ,  $I \subset \mathbb{R}$  um intervalo, uma curva diferenciável, dada por  $\gamma(t) = (\boldsymbol{\mu}(t), \Sigma(t))$ . A curva  $\gamma$  é uma curva geodésica de  $\mathcal{M}$  se suas funções coordenadas satisfazem as seguintes equações

$$\begin{aligned} \frac{d^2 \boldsymbol{\mu}}{dt^2} - \left( \frac{d\Sigma}{dt} \right) \Sigma^{-1} \left( \frac{d\boldsymbol{\mu}}{dt} \right) &= 0 \\ \frac{d^2 \Sigma}{dt^2} + \left( \frac{d\boldsymbol{\mu}}{dt} \right) \left( \frac{d\boldsymbol{\mu}}{dt} \right)^t - \left( \frac{d\Sigma}{dt} \right) \Sigma^{-1} \left( \frac{d\Sigma}{dt} \right) &= 0. \end{aligned} \quad (3.1.5)$$

Essas equações são obtidas calculando os símbolos de Christoffel de  $\mathcal{M}$  na parametrização  $\Theta$  e substituindo na Equação (1.3.3), ver referências [18] e [22].

Como ainda não é conhecida uma fórmula fechada para a distância de Rao no caso geral, muitos autores estudaram esse espaço, ver referências [15], [16] e [21]. A seguir descrevemos a distância de Rao e as geodésicas de algumas subvariedades de  $\mathcal{M}$ .

### 3.1.1 A matriz de covariância é uma matriz diagonal

Seja  $\mathcal{M}_D = \{p_\theta; \theta = (\boldsymbol{\mu}, D) \in \Theta_D\}$ ;  $\Theta_D = \{(\boldsymbol{\mu}, D), D \text{ é diagonal}\} \subset \Theta$ , uma subvariedade de  $\mathcal{M}$  formada pelas distribuições cuja matriz de covariância é uma matriz diagonal  $D$ . A métrica de Fisher de  $\mathcal{M}_D$  é

$$\begin{aligned} ds^2 &= d\boldsymbol{\mu}^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(\Sigma^{-1} d\Sigma)^2] = d\boldsymbol{\mu}^t D^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(D^{-1} dD)^2] \\ &= \sum_{i=1}^n \frac{(d\mu_i)^2}{d_{ii}} + \frac{1}{2} \text{tr}[D^{-2} dD^2] = \sum_{i=1}^n \frac{(d\mu_i)^2}{d_{ii}} + \frac{1}{2} \sum_{i=1}^n \frac{(dd_{ii})^2}{d_{ii}^2}. \end{aligned} \quad (3.1.6)$$

Se  $D$  é uma matriz de covariância dada por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}, \quad (3.1.7)$$

a métrica de Fisher de  $\mathcal{M}_D$  é

$$\begin{aligned} ds^2 &= \sum_{i=1}^n \frac{(d\mu_i)^2}{d_{ii}} + \frac{1}{2} \sum_{i=1}^n \frac{(dd_{ii})^2}{d_{ii}^2} = \sum_{i=1}^n \frac{(d\mu_i)^2}{\sigma_i^2} + \frac{1}{2} \sum_{i=1}^n \frac{(2\sigma_i d\sigma_i)^2}{\sigma_i^4} \\ &= \sum_{i=1}^n \frac{(d\mu_i)^2}{\sigma_i^2} + 2 \sum_{i=1}^n \left( \frac{d\sigma_i}{\sigma_i} \right)^2 = \sum_{i=1}^n \left( \frac{(d\mu_i)^2 + 2(d\sigma_i)^2}{\sigma_i^2} \right). \end{aligned} \quad (3.1.8)$$

A matriz de informação de Fisher associada a essa métrica, ver referência [11], é

$$\begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{2}{\sigma_1^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n^2} & 0 \\ 0 & 0 & \cdots & 0 & \frac{2}{\sigma_n^2} \end{pmatrix}. \quad (3.1.9)$$

Observe que  $\Theta_D$  é um espaço de dimensão  $2n$  e pode ser identificado como o espaço  $\mathbb{H}_F^{2n} = (\mathbb{H}_F^2)^n$ , assim podemos escrever o parâmetro  $\theta$  como  $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \dots, \mu_n, \sigma_n)$ .

A métrica de  $\mathbb{H}_F^{2n}$  está relacionada com a métrica produto do espaço  $\mathbb{H}^{2n}$ . Essa relação é dada pela transformação

$$\begin{aligned} \phi : \mathbb{H}_F^{2n} &\rightarrow \mathbb{H}^{2n} \\ (\mu_1, \sigma_1, \dots, \mu_n, \sigma_n) &\mapsto \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1, \dots, \frac{\mu_n}{\sqrt{2}}, \sigma_n \right). \end{aligned} \quad (3.1.10)$$

Dados  $\theta_1 = (\mu_{11}, \sigma_{11}, \dots, \mu_{1n}, \sigma_{1n})$  e  $\theta_2 = (\mu_{21}, \sigma_{21}, \dots, \mu_{2n}, \sigma_{2n})$ , a distância entre duas distribuições de probabilidade com esses parâmetros é dada por

$$\begin{aligned} d_D(\theta_1, \theta_2) &= d_D((\mu_{11}, \sigma_{11}, \dots, \mu_{1n}, \sigma_{1n}), (\mu_{21}, \sigma_{21}, \dots, \mu_{2n}, \sigma_{2n})) \\ &= \sqrt{2} d_{\mathbb{H}^{2n}} \left( \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1, \dots, \frac{\mu_n}{\sqrt{2}}, \sigma_n \right), \left( \frac{\mu_1}{\sqrt{2}}, \sigma_1, \dots, \frac{\mu_n}{\sqrt{2}}, \sigma_n \right) \right) \\ &= \sqrt{\sum_{i=1}^n 2 \left( d_{\mathbb{H}^2} \left( \left( \frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i} \right), \left( \frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i} \right) \right) \right)^2}, \end{aligned}$$

isto é,

$$d_D(\theta_1, \theta_2) = \sqrt{2 \sum_{i=1}^n \left( \log \frac{\left| \left( \frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i} \right) - \left( \frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i} \right) \right| + \left| \left( \frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i} \right) - \left( \frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i} \right) \right|}{\left| \left( \frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i} \right) - \left( \frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i} \right) \right| - \left| \left( \frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i} \right) - \left( \frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i} \right) \right|} \right)^2}. \quad (3.1.11)$$

Nesse espaço, uma curva diferenciável

$$\begin{aligned} \alpha : I &\rightarrow \mathbb{H}_F^{2n} \\ t &\mapsto \alpha(t) = (\alpha_1(t), \dots, \alpha_n(t)) \end{aligned}$$

é uma geodésica em  $\mathbb{H}_F^{2n}$  se, e somente se,  $\alpha_i, \forall i$ , é uma geodésica em  $\mathbb{H}_F$ . Ou seja, os  $\alpha_i$ 's são as semirretas verticais e as semi-elipses de excentricidade  $\frac{1}{\sqrt{2}}$ .

Considere agora a subvariedade  $\mathcal{M}_{D_\sigma} = \{p_\theta; \theta = (\boldsymbol{\mu}, \sigma^2 I_n) \in \Theta_{D_\sigma}\}$ ,  $\Theta_{D_\sigma} = \{(\boldsymbol{\mu}, \Sigma); \Sigma = \sigma^2 I\} \subset \Theta_D$ , nesse caso a matriz  $D$  é uma matriz de covariância múltipla da identidade,

$$\Sigma = \sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}, \quad (3.1.12)$$

e a métrica de Fisher é

$$\begin{aligned} ds^2 &= \sum_{i=1}^n \frac{(d\mu_i)^2}{d_{ii}} + \frac{1}{2} \sum_{i=1}^n \frac{(dd_{ii})^2}{d_{ii}^2} = \sum_{i=1}^n \frac{(d\mu_i)^2}{\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(2\sigma d\sigma)^2}{\sigma^4} \\ &= \sum_{i=1}^n \frac{(d\mu_i)^2}{\sigma^2} + 2n \left( \frac{d\sigma}{\sigma} \right)^2. \end{aligned} \quad (3.1.13)$$

A matriz de informação de Fisher associada a métrica dada em (3.1.13) é

$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sigma^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma^2} & 0 \\ 0 & 0 & \cdots & 0 & \frac{2n}{\sigma^2} \end{pmatrix}, \quad (3.1.14)$$

$\Theta$  pode ser identificado como um espaço de dimensão  $n+1$ , e  $\theta$  pode ser dado por  $\theta = (\mu_1, \mu_2, \dots, \mu_n, \sigma)$ . Analogamente ao espaço das distribuições normais univariadas,  $\mathbb{H}_F^{n+1}$  está relacionado com o espaço hiperbólico de dimensão  $n+1$ ,  $\mathbb{H}^{n+1}$ , cuja matriz da métrica é dada por

$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sigma^2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma^2} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\sigma^2} \end{pmatrix}. \quad (3.1.15)$$

Esta relação é dada pela transformação

$$\begin{aligned} \phi : \mathbb{H}_F^{n+1} &\rightarrow \mathbb{H}^{n+1} \\ (\mu_1, \mu_2, \dots, \mu_n, \sigma) &\mapsto \left( \frac{\mu_1}{\sqrt{2n}}, \frac{\mu_2}{\sqrt{2n}}, \dots, \frac{\mu_n}{\sqrt{2n}}, \sigma \right). \end{aligned} \quad (3.1.16)$$

Dessa forma dados  $\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, \dots, \mu_{1n})$  e  $\boldsymbol{\mu}_2 = (\mu_{21}, \mu_{22}, \dots, \mu_{2n})$ , a distância entre duas distribuições de probabilidade com parâmetros  $\theta_1 = (\boldsymbol{\mu}_1, \sigma_1)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \sigma_2)$  é

$$d_{D_\sigma}(\theta_1, \theta_2) = \sqrt{2n} \log \frac{\left| \left( \frac{\mu_1}{\sqrt{2n}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2n}}, -\sigma_2 \right) \right| + \left| \left( \frac{\mu_1}{\sqrt{2n}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2n}}, \sigma_2 \right) \right|}{\left| \left( \frac{\mu_1}{\sqrt{2n}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2n}}, -\sigma_2 \right) \right| - \left| \left( \frac{\mu_1}{\sqrt{2n}}, \sigma_1 \right) - \left( \frac{\mu_2}{\sqrt{2n}}, \sigma_2 \right) \right|}. \quad (3.1.17)$$

Novamente, de maneira análoga ao caso das distribuições normais univariadas, as geodésicas de  $\mathbb{H}^{n+1}$  são as semirretas verticais perpendiculares ao hiperplano  $\sigma = 0$  e as semi-elipses com excentricidade  $\frac{1}{\sqrt{2n}}$ .

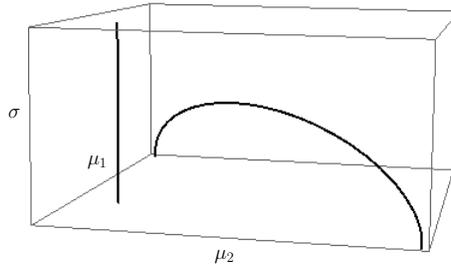


Figura 3.2: Geodésicas de  $\mathbb{H}_F^3$ .

### 3.1.2 A matriz de covariância é constante.

Seja  $\mathcal{M}_\Sigma = \{p_\theta; \theta = (\boldsymbol{\mu}, \Sigma) \in \Theta, \Sigma = \Sigma_0 \in P_n(\mathbb{R}) \text{ constante}\}$  uma subvariedade, de dimensão  $n$ , formada por distribuições normais multivariadas que possuem a mesma matriz de covariância  $\Sigma$ . Nesse espaço podemos relacionar a métrica com a métrica de um espaço Euclidiano. A métrica de Fisher de  $\mathcal{M}_\Sigma$  é

$$ds^2 = d\boldsymbol{\mu}^t \Sigma^{-1} d\boldsymbol{\mu}. \quad (3.1.18)$$

Seja agora uma matriz  $P$  de ordem  $n$  tal que

$$P^t \Sigma^{-1} P = I_n,$$

onde  $I_n$  é a matriz identidade de ordem  $n$ . A matriz  $P$  existe pois  $\Sigma$  é uma matriz simétrica definida positiva. Seja agora um vetor  $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_n)$  tal que

$$\boldsymbol{\mu} = P\boldsymbol{\nu} \Rightarrow d\boldsymbol{\mu} = Pd\boldsymbol{\nu}.$$

Sendo assim,

$$ds^2 = d\boldsymbol{\nu}^t P^t \Sigma^{-1} P d\boldsymbol{\nu} = d\boldsymbol{\nu}^t d\boldsymbol{\nu}.$$

Ou seja, a métrica coincide com a métrica do espaço Euclidiano. Logo, para valores  $\boldsymbol{\nu}_1$  e  $\boldsymbol{\nu}_2$  de  $\boldsymbol{\nu}$

$$d(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \sqrt{(\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2)^t (\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2)}.$$

Portanto segue que a distância entre duas distribuições normais multivariadas parametrizadas por  $\theta_1 = (\boldsymbol{\mu}_1, \Sigma)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \Sigma)$  é

$$d_\Sigma(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}. \quad (3.1.19)$$

A distância dada pela Equação (3.1.19) é igual à distância dada por Mahalanobis, ver referência [17], um dos pioneiros no estudo sobre distâncias entre distribuições de probabilidade.

### 3.1.3 O vetor das médias é constante.

Até agora apresentamos a distância de Rao em algumas subvariedades de  $\mathcal{M}$ . Acontece que nenhuma das subvariedades citadas são subvariedades totalmente geodésicas. A seguir, apresentamos uma subvariedade totalmente geodésica de  $\mathcal{M}$ .

Seja  $\mathcal{M}_\boldsymbol{\mu} = \{p_\theta; \theta = (\boldsymbol{\mu}, \Sigma) \in \Theta_\boldsymbol{\mu}\}$ ,  $\Theta_\boldsymbol{\mu} = \{(\boldsymbol{\mu}, \Sigma); \boldsymbol{\mu} = \boldsymbol{\mu}_0 \in \mathbb{R}^n \text{ constante}\} \subset \Theta$  uma subvariedade, de dimensão  $\frac{n(n+1)}{2}$ , formada por distribuições que possuem o mesmo vetor de médias  $\boldsymbol{\mu}$ . A expressão da distância de Rao desse espaço foi obtida por diversos autores: Skovgaard [22], Moakher [18], Burbea [4], Forstner e Moonen [13] entre outros. Apresentamos aqui o teorema dado por S. T. Jensen em 1976, no qual ele determina a distância da subvariedade  $\mathcal{M}_\boldsymbol{\mu}$ . A demonstração desse teorema pode ser encontrada em [3] no Apêndice 1.

**Teorema 3.1.3.** Considere a família de distribuições normais multivariadas  $\mathcal{M}_\mu$  com o vetor da média comum  $\mu$  mas com diferentes matrizes de covariância  $\Sigma$ . Dados dois elementos dessa família, parametrizados por  $\theta_1 = (\mu, \Sigma_1)$  e  $\theta_2 = (\mu, \Sigma_2)$ , a distância entre dois elementos dessa família é dada por

$$d_\mu^2(\Sigma_1, \Sigma_2) = \frac{1}{2} \sum_{i=1}^n [\log(\lambda_i)]^2, \quad (3.1.20)$$

onde  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  são os autovalores de  $\Sigma_1^{-1}\Sigma_2$ .

Observe que as equações que determinam as geodésicas de  $\Theta$ , dadas em (3.1.5), quando restritas à  $\Theta_\mu$  se reduzem a

$$\frac{d^2\Sigma}{dt^2} - \left(\frac{d\Sigma}{dt}\right) \Sigma^{-1} \left(\frac{d\Sigma}{dt}\right) = 0. \quad (3.1.21)$$

A curva  $\gamma(t) = (\mu(t), \Sigma(t))$  que satisfaz a equação acima ligando dois pontos  $\theta_1 = (\mu, \Sigma_1)$  e  $\theta_2 = (\mu, \Sigma_2)$  em  $\Theta_\mu$  com  $\gamma(t_1) = (\mu, \Sigma_1)$  e  $\gamma(t_2) = (\mu, \Sigma_2)$  é dada por

$$\gamma(t) = (\mu, \Sigma(t_1)^{1/2} \exp((t - t_1)B)\Sigma(t_1)^{1/2}), \quad (3.1.22)$$

$\forall t \in [t_1, t_2]$ , onde  $B \in S_n(\mathbb{R})$  é uma constante de integração. Para mais detalhes ver referências [4], [18] e [22].

## 3.2 Limitantes

Nessa seção vamos calcular alguns limitantes para a distância de Rao para distribuições normais multivariadas.

### 3.2.1 Limitante Inferior

Em 1990, Calvo e Oller, ver referência [5], mostraram um mergulho do espaço de parâmetros da variedade estatística formada pelas distribuições de probabilidades normais multivariadas na variedade formada pelas matrizes simétricas positivas definidas. Esse mergulho permite obter um limitante inferior para a distância de Rao. As demonstrações dessa seção encontram-se em [5].

Antes de definir o mergulho, seja o seguinte lema

**Lema 3.2.1.** Toda matriz  $S \in P_{n+1}(\mathbb{R})$  pode ser escrita da seguinte forma

$$\begin{pmatrix} \Sigma + \beta\mu\mu^t & \beta\mu \\ \beta\mu^t & \beta \end{pmatrix} \quad (3.2.1)$$

onde  $\beta \in \mathbb{R}_*^+$ ,  $\mu \in \mathbb{R}^n$  e  $\Sigma \in P_n(\mathbb{R})$ . Reciprocamente, toda matriz da forma acima é simétrica definida positiva. Também, para todo  $S \in P_{n+1}(\mathbb{R})$  temos que

$$ds^2 = \frac{1}{2} \left(\frac{d\beta}{\beta}\right)^2 + \beta(d\mu)^t \Sigma^{-1} d\mu + \frac{1}{2} \text{tr}[(\Sigma^{-1}d\Sigma)^2]. \quad (3.2.2)$$

A seguinte aplicação é um mergulho de  $\Theta$  em  $P_{n+1}(\mathbb{R})$

$$\begin{aligned} f : \Theta &\rightarrow P_{n+1}(\mathbb{R}) \\ (\boldsymbol{\mu}, \Sigma) &\mapsto \begin{pmatrix} \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^t & \boldsymbol{\mu} \\ \boldsymbol{\mu}^t & 1 \end{pmatrix}. \end{aligned} \quad (3.2.3)$$

Pelo Lema 3.2.1 a aplicação  $f$  está bem definida. O teorema abaixo estabelece algumas propriedades geométricas entre  $\Theta$  e  $f(\Theta)$ .

**Teorema 3.2.2.** Dada a aplicação  $f$  definida em (3.2.3), temos que

1.  $f$  é um difeomorfismo de  $\Theta$  sobre  $f(\Theta)$ .
2.  $f(\Theta)$  é uma subvariedade de  $P_{n+1}(\mathbb{R})$  de dimensão  $\frac{(n+1)(n+2)}{2} - 1$ .
3. A expressão da métrica induzida da subvariedade  $f(\Theta)$  pode ser expressa por

$$ds^2 = d\boldsymbol{\mu}^t \Sigma^{-1} d\boldsymbol{\mu} + \frac{1}{2} \text{tr}[(\Sigma^{-1} d\Sigma)^2].$$

4.  $\Theta$  é isométrico a  $f(\Theta)$ .
5.  $f(\Theta)$  é uma subvariedade não geodésica de  $P_{n+1}(\mathbb{R})$ .
6. Se  $\Theta_{\boldsymbol{\mu}} = \{(\boldsymbol{\mu}, \Sigma) \in \Theta : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ constante}\}$  então  $f(\Theta_{\boldsymbol{\mu}})$  é uma subvariedade geodésica de  $P_{n+1}(\mathbb{R})$  de dimensão  $\frac{n(n+1)}{2}$ .

**Observação 3.2.3.** O  $\Theta_{\boldsymbol{\mu}}$  é o espaço de parâmetros da subvariedade  $\mathcal{M}_{\boldsymbol{\mu}}$  definida na Seção 3.1.3.

**Corolário 3.2.4.** Sejam  $\theta_1$  e  $\theta_2$  dois pontos de  $\Theta$ . Se  $d_R$  é a distância de Rao entre eles e  $d$  é a distância Riemanniana entre  $f(\theta_1)$  e  $f(\theta_2)$  em  $P_{n+1}(\mathbb{R})$ , então  $d_{\boldsymbol{\mu}} \geq d$ . Se  $\theta_1, \theta_2 \in \Theta_{\boldsymbol{\mu}}$  então  $d = d_R = d_{\boldsymbol{\mu}}$ , onde  $d_{\boldsymbol{\mu}}$  é a distância dada na Equação (3.1.20).

Uma fórmula para a distância Riemanniana  $d$  entre  $f(\theta_1)$  e  $f(\theta_2)$  é dada pelo teorema abaixo.

**Teorema 3.2.5.** Seja  $S_i = f(\theta_i) = f(\boldsymbol{\mu}_i, \Sigma_i)$ ,  $i = 1, 2$ , dois pontos de  $f(\Theta)$ , então

$$d^2(S_1, S_2) = \frac{1}{2} \sum_{k=1}^{n+1} [\log(\lambda_k)]^2, \quad (3.2.4)$$

onde  $\lambda_k$  são os autovalores de  $S_1^{-1}S_2$ .

Assim, temos que a distância dada pelo teorema acima é limitante inferior (que denotamos por  $LI$ ) para a distância de Rao da variedade  $\mathcal{M}$ .

### 3.2.2 Limitante Superior

Em um outro artigo, ver referência [6], Calvo e Oller calcularam um limitante superior para a distância de Rao.

Eles conseguiram resolver o sistema de equações diferenciais, dado na Equação (1.3.3), que determinam as geodésicas de  $\Theta$ . Dessa forma eles calcularam uma expressão explícita para as geodésicas do espaço formado pelas distribuições normais multivariadas. Porém, dados dois pontos em  $\Theta$  eles não determinaram qual a curva geodésica passando por esses pontos que minimiza a distância entre eles, ou seja, não calcularam a distância de Rao entre esses pontos. Neste caso é necessário resolver um sistema de equações diferenciais de segunda ordem com condições de contorno.

Fazendo algumas restrições nesse sistema, Calvo e Oller calcularam a distância de Rao entre os pontos do subconjunto  $\Theta_{\alpha\Sigma} \subset \Theta$ ,

$$\Theta_{\alpha\Sigma} = \{\theta = (\boldsymbol{\mu}, \Sigma) \in \Theta; \Sigma = \alpha\Sigma_0, \Sigma_0 \in P_n(\mathbb{R}), \alpha \in \mathbb{R}_+^*\}. \quad (3.2.5)$$

Dados dois pontos  $\theta_1 = (\boldsymbol{\mu}_1, \Sigma)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \alpha\Sigma)$ , obtiveram

$$d_{\alpha\Sigma}^2(\theta_1, \theta_2) = 2 \operatorname{arccosh} \left( \frac{\sqrt{\alpha}}{2} + \frac{1}{2\sqrt{\alpha}} + \frac{1}{4\sqrt{\alpha}} \delta^t \delta \right) + \frac{n-1}{2} \log^2 \alpha, \quad (3.2.6)$$

onde  $\delta = \Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ .

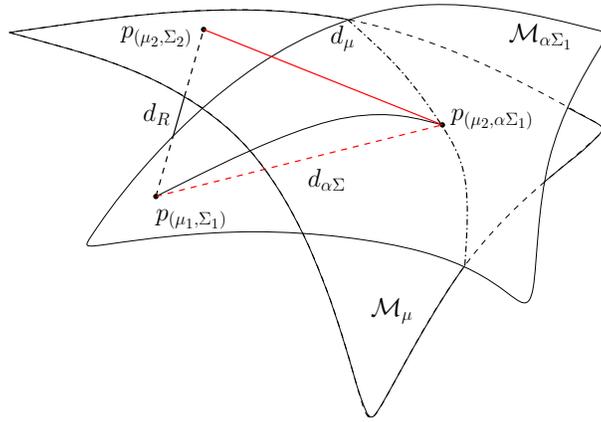
**Observação 3.2.6.** A distância  $d_{\alpha\Sigma}$  é a distância de Rao para pontos de  $\Theta_{\alpha\Sigma}$  mas não é uma distância de Rao restrita à subvariedade  $\Theta_{\alpha\Sigma}$ , ou seja, a curva geodésica que liga os pontos de  $\Theta_{\alpha\Sigma}$  percorre pontos que não necessariamente estão em  $\Theta_{\alpha\Sigma}$ .

Com essa distância eles determinam um limitante superior para distância de Rao,  $d_R$ , da variedade  $\mathcal{M}$  no caso geral. Dados pontos  $\theta_1 = (\boldsymbol{\mu}_1, \Sigma_1)$ ,  $\theta_\alpha = (\boldsymbol{\mu}_2, \alpha\Sigma_1)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \Sigma_2)$ , temos que

$$d_R(\theta_1, \theta_2) \leq d_R(\theta_1, \theta_\alpha) + d_R(\theta_\alpha, \theta_2) \leq d_{\alpha\Sigma}(\theta_1, \theta_\alpha) + d_\mu(\theta_\alpha, \theta_2), \quad (3.2.7)$$

ou seja,  $LS_\alpha = d_{\alpha\Sigma}(\theta_1, \theta_\alpha) + d_\mu(\theta_\alpha, \theta_2)$  é um limitante superior para a distância de Rao entre  $\theta_1$  e  $\theta_2$ .

Observe a Figura 3.3, dadas duas distribuições de probabilidade  $p(\boldsymbol{\mu}_1, \Sigma_1)$  e  $p(\boldsymbol{\mu}_2, \Sigma_2)$ , tome uma distribuição  $p(\boldsymbol{\mu}_2, \alpha\Sigma_1)$  tal que  $p(\boldsymbol{\mu}_2, \alpha\Sigma_1) \in \mathcal{M}_{\alpha\Sigma_1} \cap \mathcal{M}_\mu$ . Dessa forma, somando o valor da distância  $d_{\alpha\Sigma}$  entre  $p(\boldsymbol{\mu}_2, \Sigma_2)$  e  $p(\boldsymbol{\mu}_2, \alpha\Sigma_1)$ , que é a distância de Rao para pontos do conjunto  $\mathcal{M}_{\alpha\Sigma_1}$ , com o valor da distância  $d_\mu$  entre  $p(\boldsymbol{\mu}_2, \Sigma_2)$  e  $p(\boldsymbol{\mu}_2, \alpha\Sigma_1)$ , que é a distância de Rao restrita ao conjunto  $\mathcal{M}_\mu$  e, usando a desigualdade triangular, obtemos o limitante superior  $LS_\alpha$  (ilustrado em vermelho) para a distância de Rao,  $d_R$ .


 Figura 3.3: Limitante  $LS_\alpha$ .

Podemos melhorar esse limitante escolhendo um  $\alpha$  adequado. Em [6], Calvo e Oller determinam uma expressão analítica do escalar  $\alpha$  que minimiza a distância entre  $\theta_\alpha$  e  $\theta_2$

$$\alpha = \|\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}\|^{\frac{1}{n}}. \quad (3.2.8)$$

O mínimo para  $LS_\alpha$  pode ser calculado através de métodos numéricos, isto é, podemos encontrar um  $\alpha_0$  tal que

$$\alpha_0 = \min_{\alpha} \{d_{\alpha\Sigma}(\theta_1, \theta_\alpha) + d_\mu(\theta_\alpha, \theta_2)\}. \quad (3.2.9)$$

Com base na ideia de calcular um limitante superior usando uma desigualdade triangular, calculamos um outro limitante superior para a distância de Rao. Esse limitante é obtido através da distância do espaço onde a matriz de covariância  $\Sigma$  é diagonal. Dados pontos  $\theta_1 = (\boldsymbol{\mu}_1, \Sigma_1)$  e  $\theta_2 = (\boldsymbol{\mu}_2, \Sigma_2)$ , temos pela Equação (3.1.4) que

$$d(\theta_1, \theta_2) = d((0, I); (\Sigma_1^{-(1/2)}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \Sigma_1^{-(1/2)}\Sigma_2\Sigma_1^{-(1/2)})).$$

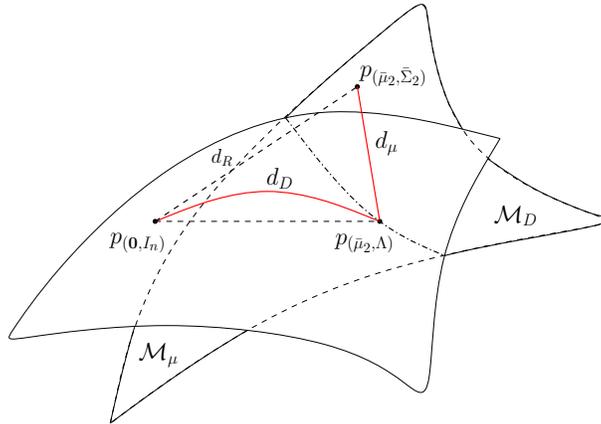
Fazendo  $\theta_0 = (\mathbf{0}, I_n)$ ,  $\bar{\theta}_2 = (\bar{\boldsymbol{\mu}}_2, \bar{\Sigma}_2) = (\Sigma_1^{-(1/2)}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \Sigma_1^{-(1/2)}\Sigma_2\Sigma_1^{-(1/2)})$  e  $\theta_\Lambda = (\bar{\boldsymbol{\mu}}_2, \Lambda)$ , onde  $\Lambda$  é a projeção de  $\bar{\theta}_2$  na sua diagonal, temos

$$d_R(\theta_1, \theta_2) = d(\theta_0, \bar{\theta}_2) \leq d_R(\theta_0, \theta_\Lambda) + d_R(\theta_\Lambda, \bar{\theta}_2) \leq d_D(\theta_0, \theta_\Lambda) + d_\mu(\theta_\Lambda, \bar{\theta}_2), \quad (3.2.10)$$

onde  $d_D$  é a distância dada na Equação (3.1.11).

Ilustramos esse limitante na Figura 3.4. Dadas duas distribuições de probabilidade  $p(\mathbf{0}, I_n)$  e  $p(\bar{\boldsymbol{\mu}}_2, \bar{\Sigma}_2)$ , tome uma distribuição  $p(\bar{\boldsymbol{\mu}}_2, \Lambda)$  tal que  $p(\bar{\boldsymbol{\mu}}_2, \Lambda) \in \mathcal{M}_D \cap \mathcal{M}_\mu$ . Dessa forma, somando o valor da distância  $d_D$  entre  $p(\bar{\boldsymbol{\mu}}_2, \Lambda)$ , que é a distância de Rao restrita ao conjunto  $\mathcal{M}_D$ , com o valor da distância  $d_\mu$  entre  $p(\bar{\boldsymbol{\mu}}_2, \bar{\Sigma}_2)$  e  $p(\bar{\boldsymbol{\mu}}_2, \Lambda)$ , que é a distância de Rao restrita ao conjunto  $\mathcal{M}_\mu$  e, usando a desigualdade triangular, obtemos o limitante superior  $LS_\Lambda$  (ilustrado em vermelho) para a distância de Rao,  $d_R$ .

Note que, diferente de  $d_{\alpha\Sigma}$ ,  $d_D$  é a distância de Rao restrita a uma subvariedade.


 Figura 3.4: Limitante  $LS_\Lambda$ .

Também podemos melhorar o limitante  $LS_\Lambda = d_D(\theta_0, \theta_\Lambda) + d_\mu(\theta_\Lambda, \bar{\theta}_2)$  através de métodos numéricos, observe que ele é obtido projetando  $\theta_2$  no espaço  $\Theta_D = \{(\boldsymbol{\mu}, \Sigma); \Sigma = D \text{ diagonal}\}$ . Logo, basta encontrar uma matriz diagonal  $\Lambda_0$  tal que

$$\Lambda_0 = \min_{\Lambda} \{d_D(\theta_0, \theta_\Lambda) + d_\mu(\theta_\Lambda, \bar{\theta}_2)\}. \quad (3.2.11)$$

Observe que neste caso estamos minimizando uma função de  $n$  variáveis.

Assim, munidos desses limitantes, mesmo sem uma expressão explícita para o cálculo da distância de Rao em  $\mathcal{M}$ , dados dois pontos nesse espaço podemos determinar o intervalo no qual o valor da distância entre esses dois pontos se encontra.

### 3.3 Simulações

Nessa seção fazemos algumas comparações entre os limitantes obtidos na seção anterior no espaço das distribuições normais bivariadas.

Fixamos o ponto  $\theta_0 = (\mathbf{0}, I_2)$  e calculamos os limitantes para a distância entre  $\theta_0$  e um ponto  $\theta = (\boldsymbol{\mu}, \Sigma)$ , podemos fazer isso por causa da relação dada na Equação (3.1.4).

As comparações foram divididas em casos. Em cada caso, tomamos pontos  $\theta = (\boldsymbol{\mu}, \Sigma)$  nos quais variamos a distância entre vetor média  $\boldsymbol{\mu}$  e o vetor  $\mathbf{0}$  e o crescimento dos autovalores da matriz  $\Sigma$ :

- *Caso 1:* A distância entre o vetor média  $\boldsymbol{\mu}$  e o vetor  $\mathbf{0}$  é pequena ( $\boldsymbol{\mu}$  está próximo da origem), os autovalores de  $\Sigma$  são maiores do que 1 e seus valores crescem.
- *Caso 2:* O vetor média próximo da origem, os autovalores de  $\Sigma$  são menores do que 1 e seus valores decrescem.
- *Caso 3:* A distância entre o vetor média  $\boldsymbol{\mu}$  e o vetor  $\mathbf{0}$  aumenta ( $\boldsymbol{\mu}$  se afasta da origem) e os autovalores da matriz  $\Sigma$  estão próximos de 1.

- *Caso 4*: O vetor média  $\boldsymbol{\mu}$  se afasta da origem, os autovalores de  $\Sigma$  são maiores do que 1 e seus valores crescem.
- *Caso 5*: O vetor média  $\boldsymbol{\mu}$  se afasta da origem, os autovalores de  $\Sigma$  são menores do que 1 e seus valores decrescem.

Os valores dos parâmetros  $\theta = (\boldsymbol{\mu}, \Sigma)$  utilizados em cada caso encontram-se no Apêndice A. Analisamos os limitantes  $LI$ ,  $LS_\alpha$  e  $LS_\Lambda$ , dados na seção anterior, com e sem processos de minimização.

### Limitantes $LI$ , $LS_\alpha$ e $LS_\Lambda$ sem processos de minimização

Nas simulações abaixo para cada distância entre  $\theta_0$  e  $\theta$  plotamos os valores do limitante inferior dado por Calvo e Oller na Equação (3.2.4),  $LI$ , do limitante superior de Calvo e Oller dado na Equação (3.2.7),  $LS_\alpha$ , tomando o  $\alpha$  analítico dado em (3.2.8), e do limitante superior dado na Equação (3.2.10),  $LS_\Lambda$ , que é obtido projetando  $\Sigma$  na sua diagonal.

Observando as Figuras 3.5 e 3.6 e as Tabelas 3.1 e 3.2, que ilustram as simulações dos Casos 1 e 2 (casos onde o vetor das médias está próximo da origem), respectivamente, vemos que os limitantes estão bem próximos e que os valores do limitante superior dado pela projeção,  $LS_\Lambda$ , são um pouco maiores que os valores do limitante superior dado por Calvo e Oller.

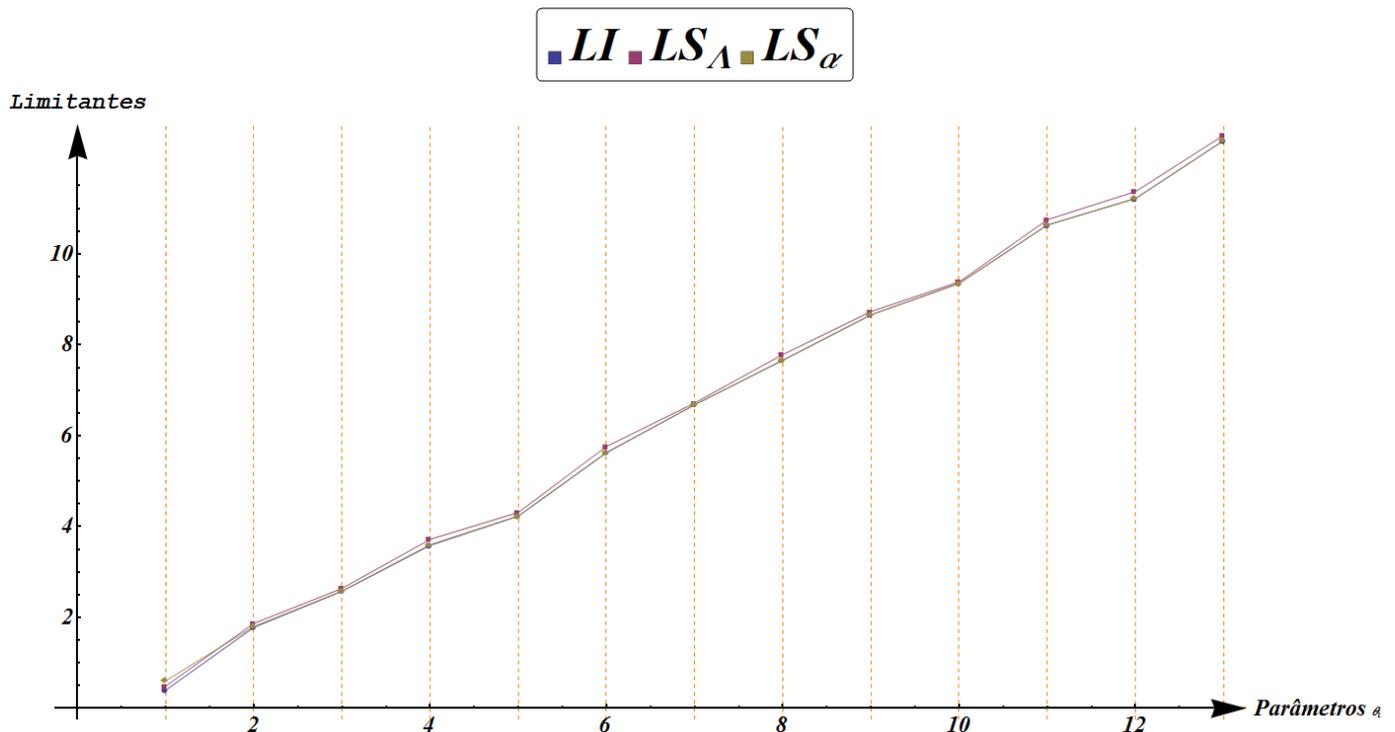


Figura 3.5: Gráfico dos valores dos limitantes  $LI$ ,  $LS_\Lambda$  e  $LS_\alpha$  para cada ponto  $\theta = \theta_i$  no Caso 1.

$\theta$	$LI$	$LS_{\Lambda}$	$LS_{\alpha}$
$\theta_1$	0.400226	0.489838	0.605976
$\theta_2$	1.78151	1.87014	1.80135
$\theta_2$	2.5759	2.6387	2.58211
$\theta_4$	3.58633	3.71852	3.6031
$\theta_5$	4.22943	4.31242	4.23506
$\theta_6$	5.62373	5.7628	5.63439
$\theta_7$	6.68029	6.7187	6.68105
$\theta_8$	7.65805	7.78645	7.66447
$\theta_9$	8.65806	8.73009	8.65992
$\theta_{10}$	9.35049	9.38464	9.35089
$\theta_{11}$	10.6338	10.7512	10.6376
$\theta_{12}$	11.2169	11.371	11.2228
$\theta_{13}$	12.492	12.6095	12.4952

Tabela 3.1: Valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 1.

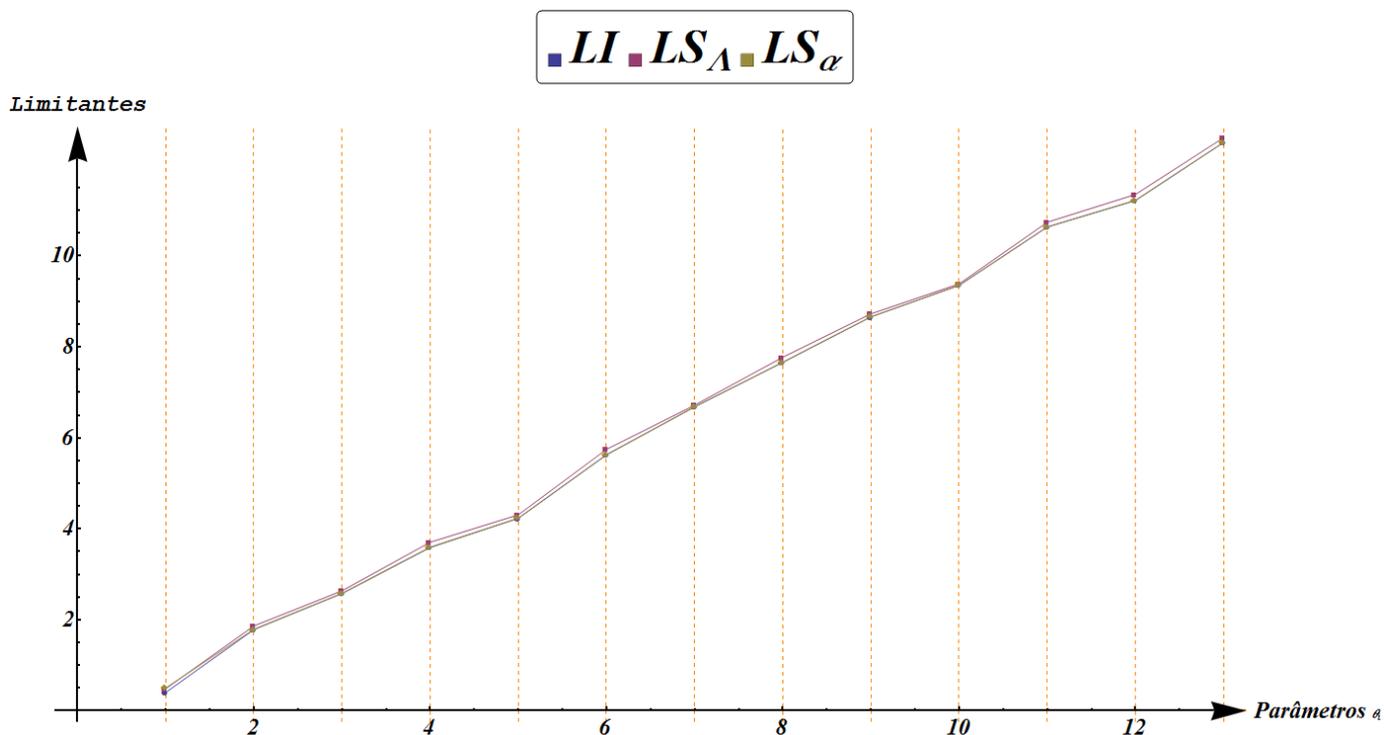


Figura 3.6: Gráfico dos valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 2.

$\theta$	$LI$	$LS_{\Lambda}$	$LS_{\alpha}$
$\theta_1$	0.402193	0.484152	0.501608
$\theta_2$	1.78322	1.86437	1.79238
$\theta_3$	2.5774	2.6364	2.58143
$\theta_4$	3.588	3.70435	3.59847
$\theta_5$	4.23095	4.30747	4.23506
$\theta_6$	5.62523	5.74702	5.63315
$\theta_7$	6.68185	6.71883	6.68251
$\theta_8$	7.65958	7.77319	7.66477
$\theta_9$	8.65966	8.72682	8.66121
$\theta_{10}$	9.35207	9.3851	9.35242
$\theta_{11}$	10.6354	10.7404	10.6386
$\theta_{12}$	11.2185	11.3518	11.2235
$\theta_{13}$	12.4936	12.5987	12.4964

Tabela 3.2: Valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 2.

No Caso 3, ver Figura 3.7 e Tabela 3.3, vemos que o limitante  $LS_{\alpha}$  é muito melhor que o limitante  $LS_{\Lambda}$  e que, quanto mais afastamos o vetor média da origem, maior fica a distância entre os valores dos limitantes superiores e o valor do limitante inferior.

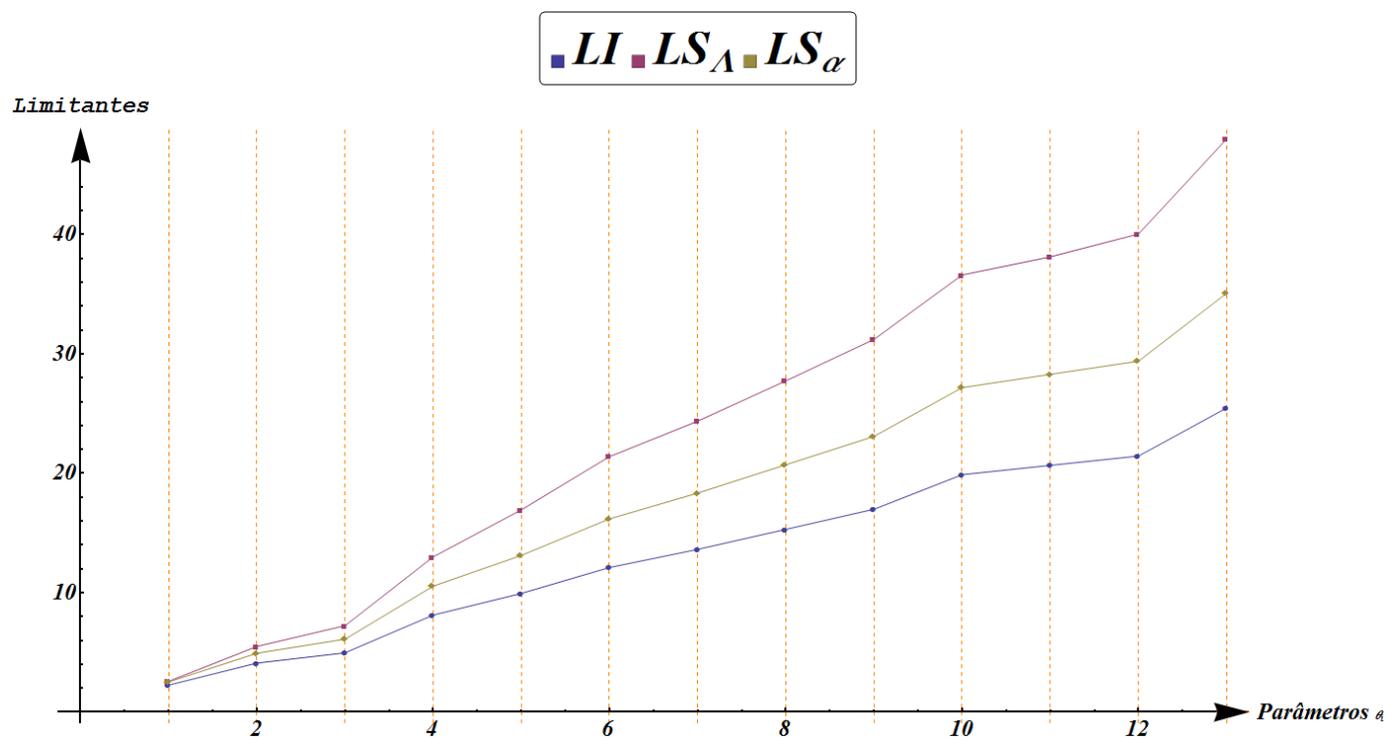


Figura 3.7: Gráfico dos valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 3.

$\theta$	$LI$	$LS_{\Lambda}$	$LS_{\alpha}$
$\theta_1$	2.25382	2.57825	2.52873
$\theta_2$	4.10213	5.49074	4.92347
$\theta_3$	4.95886	7.23582	6.10228
$\theta_4$	8.12118	12.9987	10.5456
$\theta_5$	9.93721	16.9382	13.1228
$\theta_6$	12.1038	21.3969	16.1812
$\theta_7$	13.6278	24.3631	18.3413
$\theta_8$	15.3045	27.7431	20.7122
$\theta_9$	16.9899	31.2175	23.093
$\theta_{10}$	19.8782	36.6178	27.1722
$\theta_{11}$	20.6624	38.1272	28.2911
$\theta_{12}$	21.4449	40.0392	29.3907
$\theta_{13}$	25.4678	48.0029	35.0789

Tabela 3.3: Valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 3.

Já no Caso 4, representado pela Figura 3.8 e pela Tabela 3.4, que é o caso onde o vetor da média se afasta da origem e os autovalores de  $\Sigma$  assumem valores cada vez maiores, observamos que o limitante  $LS_{\Lambda}$  é melhor que o limitante  $LS_{\alpha}$ . Ou seja, para este caso conseguimos melhorar o limitante dado por Calvo e Oller.

$\theta$	$LI$	$LS_{\Lambda}$	$LS_{\alpha}$
$\theta_1$	2.25376	2.58657	2.572
$\theta_2$	3.75615	4.46553	5.02165
$\theta_3$	4.36503	5.32243	6.33755
$\theta_4$	7.05948	9.70361	10.844
$\theta_5$	8.59487	12.8752	13.5424
$\theta_6$	10.5157	15.9907	16.8112
$\theta_7$	11.7988	17.7895	19.2148
$\theta_8$	13.1852	20.2647	21.7087
$\theta_9$	14.7915	22.6902	24.3017
$\theta_{10}$	17.2455	27.3642	28.4905
$\theta_{11}$	18.0726	27.6868	29.8133
$\theta_{12}$	18.6169	29.0315	30.9896
$\theta_{13}$	22.1076	35.676	36.9006

Tabela 3.4: Valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 4.

No Caso 5, ver Figura 3.9 e Tabela 3.5, de maneira análoga ao Caso 3, vemos que o limitante  $LS_{\alpha}$  é muito melhor que o limitante  $LS_{\Lambda}$ .

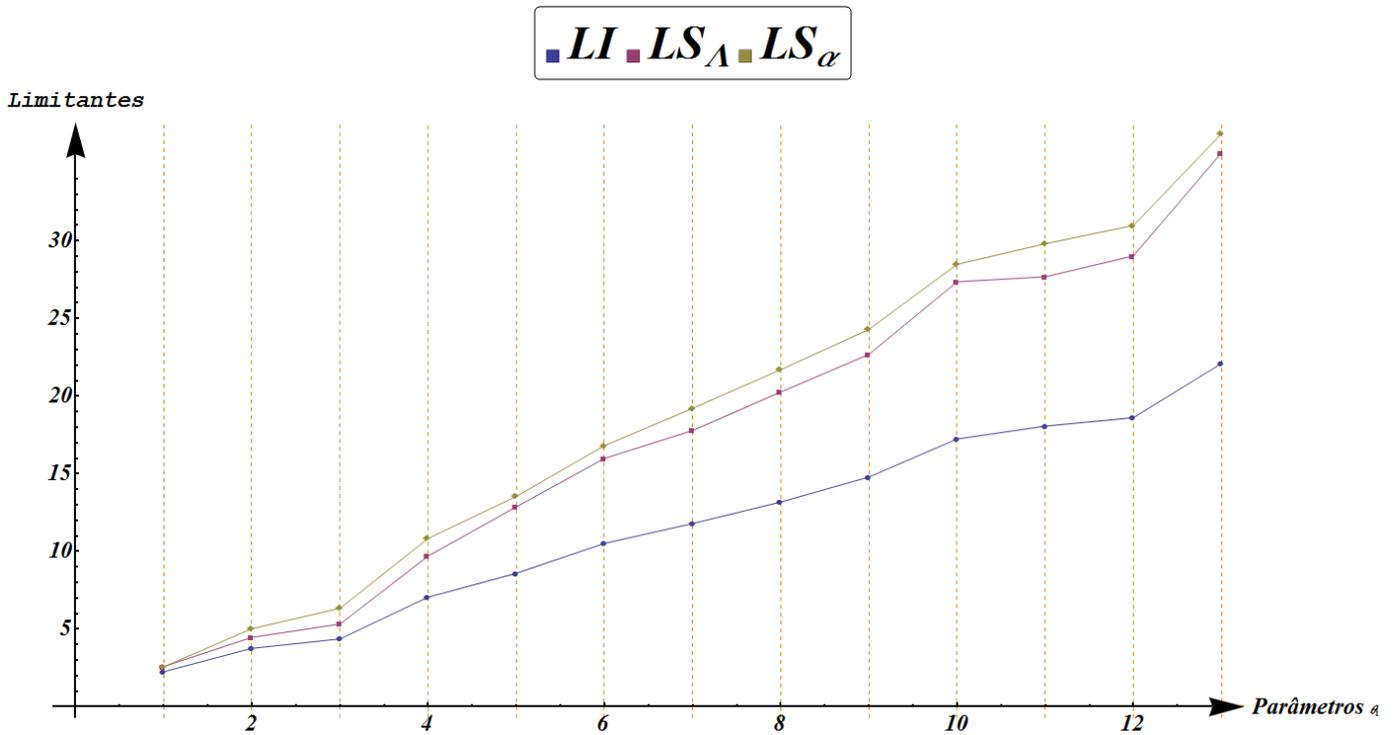


Figura 3.8: Gráfico dos valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 4.

$\theta$	$LI$	$LS_{\Lambda}$	$LS_{\alpha}$
$\theta_1$	2.38534	2.8892	2.89188
$\theta_2$	5.17697	7.24366	6.50347
$\theta_3$	6.68414	9.87698	8.388
$\theta_4$	10.4606	16.752	13.7371
$\theta_5$	12.6911	21.3032	16.8599
$\theta_6$	15.7577	27.1896	21.1508
$\theta_7$	17.9876	31.139	24.2202
$\theta_8$	20.3661	35.574	27.4605
$\theta_9$	22.6094	40.0009	30.7073
$\theta_{10}$	25.8903	46.0573	35.375
$\theta_{11}$	27.507	48.9104	37.6395
$\theta_{12}$	28.7822	51.4403	39.2583
$\theta_{13}$	32.0143	60.659	46.04

Tabela 3.5: Valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 5.

Assim, podemos concluir que para determinar um intervalo para distância de Rao para um conjuntos de pontos formados por vetores da média  $\boldsymbol{\mu}$  afastados da origem e com a matriz  $\Sigma$  de covariância com autovalores grandes é melhor utilizar o limitante  $LS_{\Lambda}$ . Nos outros casos é melhor

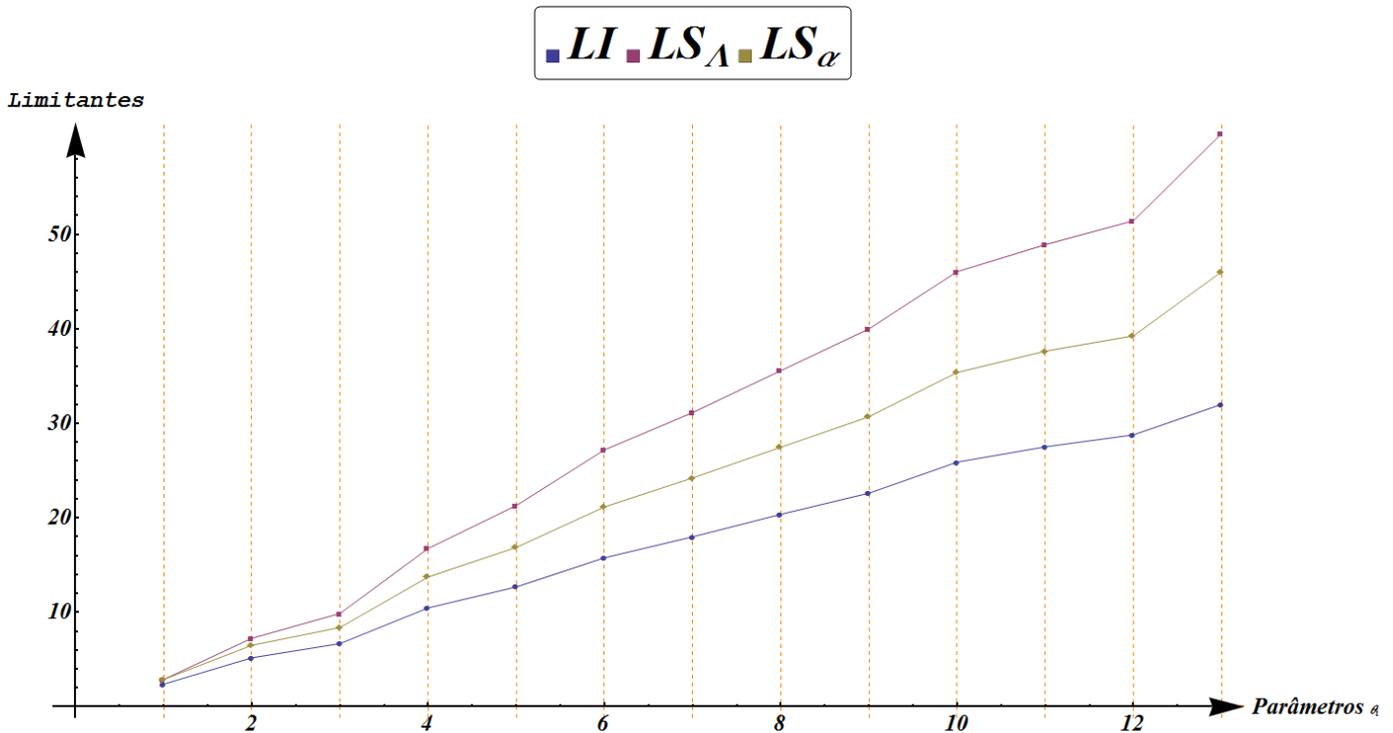


Figura 3.9: Gráfico dos valores dos limitantes  $LI$ ,  $LS_{\Lambda}$  e  $LS_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 5.

utilizar  $LS_{\alpha}$ .

### Limitantes $LI$ , $LS_{\alpha}$ e $LS_{\Lambda}$ com processos de minimização

Vamos agora ver o que acontece com os limitantes superiores quando utilizamos métodos numéricos para minimizá-los. Para cada distância entre  $\theta_0$  e  $\theta$  plotamos os valores de  $LI$ , do limitante superior de Calvo e Oller dado na Equação (3.2.7) tomando o  $\alpha$  que minimiza  $LS_{\alpha}$  através de métodos numéricos (ver Equação (3.2.9)), que denotamos por  $LSM_{\alpha}$ , e do limitante superior dado na Equação (3.2.10) usando uma matriz diagonal  $\Lambda$  que minimiza  $LS_{\Lambda}$  através de métodos numéricos (ver Equação (3.2.11)), que denotamos por  $LSM_{\Lambda}$ .

Para uma melhor visualização das simulações, os gráficos ilustram os valores de  $LSM_{\alpha}$  e  $LSM_{\Lambda}$  em relação a  $LI$ , isto é, ilustram os valores  $(LSM_{\alpha} - LI)/LI$  e  $(LSM_{\Lambda} - LI)/LI$ . As tabelas continuam representando os valores reais dos limitantes.

Observamos, pela Figuras 3.10 e 3.11 e pelas Tabelas 3.6 e 3.7, que nos Casos 1 e 2, o limitante  $LSM_{\Lambda}$  é melhor que o limitante  $LSM_{\alpha}$ .

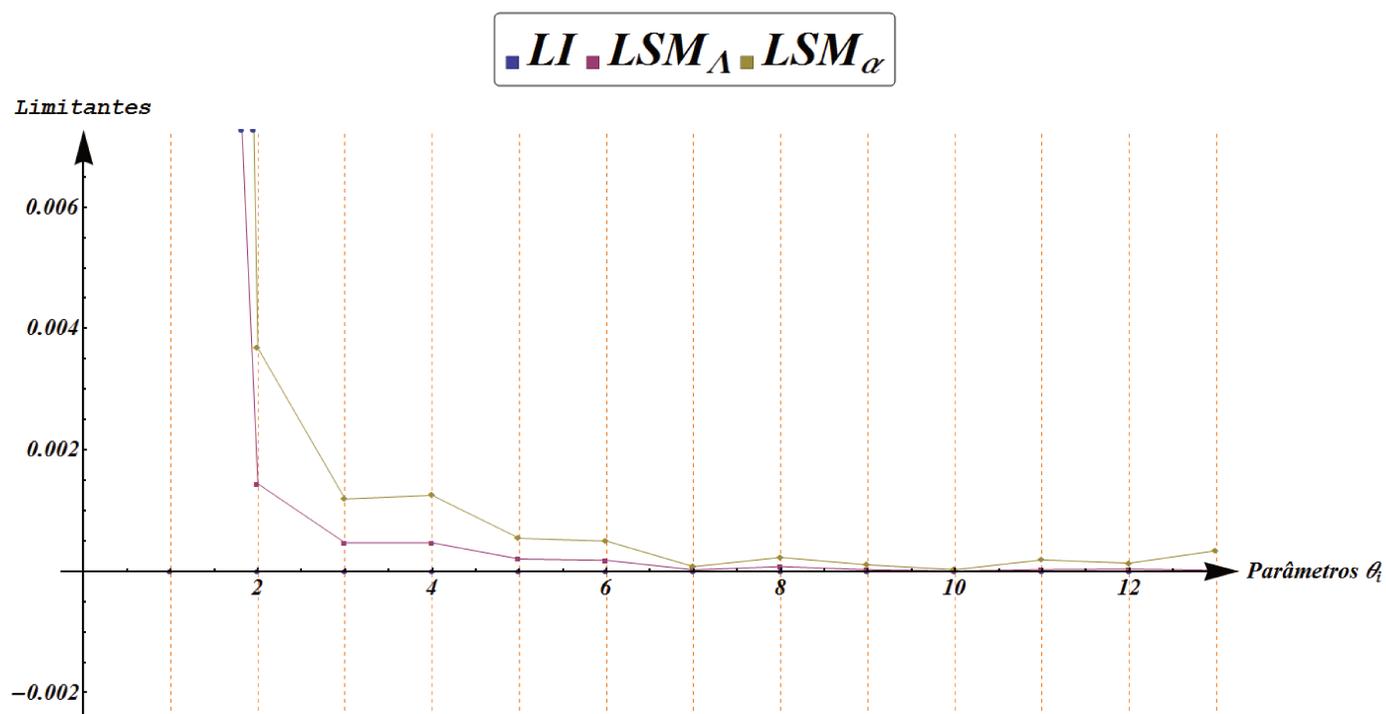


Figura 3.10: Gráfico dos valores relativos dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 1.

$\theta$	$LI$	$LSM_{\Lambda}$	$LSM_{\alpha}$
$\theta_1$	0.400226	0.413578	0.432752
$\theta_2$	1.78151	1.78408	1.78806
$\theta_3$	2.5759	2.5771	2.57897
$\theta_4$	3.58633	3.588	3.59081
$\theta_5$	4.22943	4.23028	4.23172
$\theta_6$	5.62373	5.62474	5.62651
$\theta_7$	6.68029	6.68045	6.6808
$\theta_8$	7.65805	7.65864	7.65979
$\theta_9$	8.65806	8.65828	8.65897
$\theta_{10}$	9.35049	9.35056	9.35075
$\theta_{11}$	10.6338	10.6341	10.6358
$\theta_{12}$	11.2169	11.2173	11.2184
$\theta_{13}$	12.492	12.4923	12.4963

Tabela 3.6: Valores dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 1.

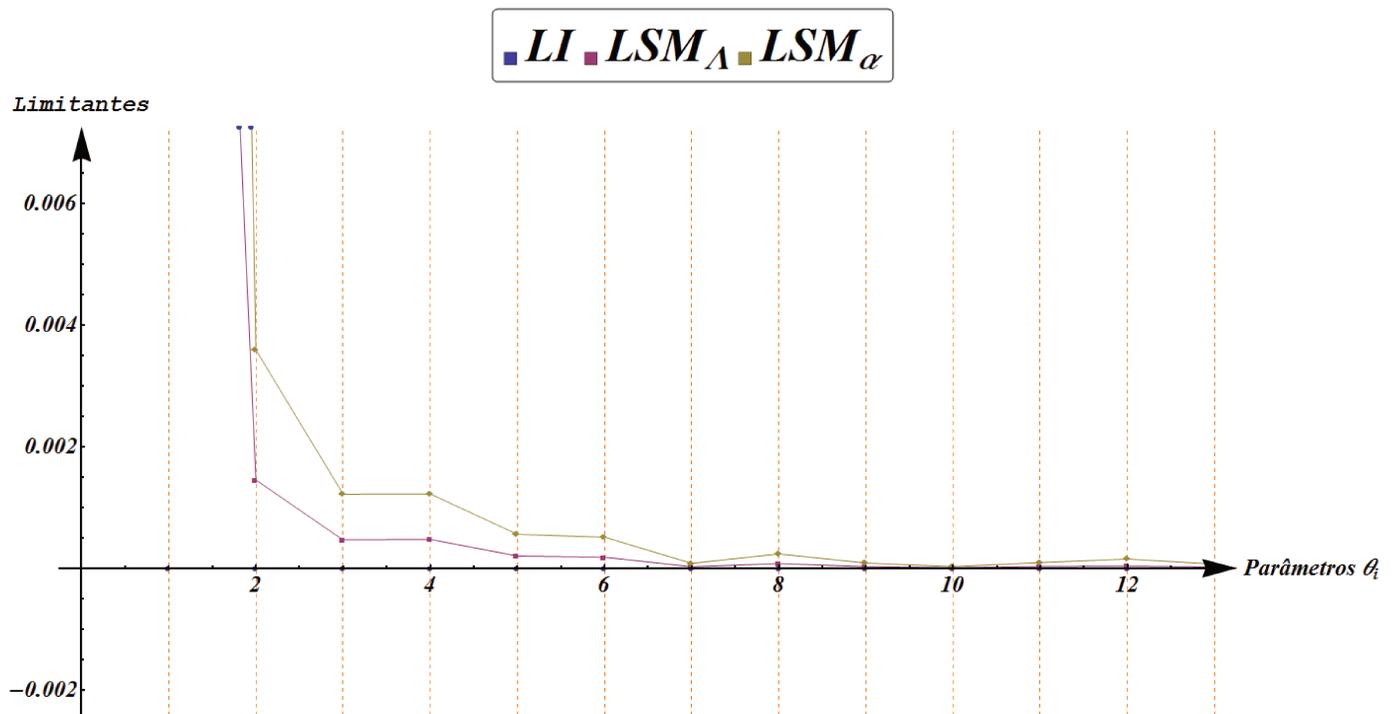


Figura 3.11: Gráfico dos valores relativos dos limitantes  $LI$ ,  $LMS_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 2.

$\theta$	$LI$	$LSM_{\Lambda}$	$LSM_{\alpha}$
$\theta_1$	0.402193	0.415582	0.433906
$\theta_2$	1.78322	1.7858	1.78962
$\theta_3$	2.5774	2.5786	2.58054
$\theta_4$	3.588	3.58969	3.59238
$\theta_5$	4.23095	4.23179	4.2333
$\theta_6$	5.62523	5.62624	5.62808
$\theta_7$	6.68185	6.68201	6.68238
$\theta_8$	7.65958	7.66016	7.66137
$\theta_9$	8.65966	8.65989	8.66039
$\theta_{10}$	9.35207	9.35215	9.35232
$\theta_{11}$	10.6354	10.6357	10.6364
$\theta_{12}$	11.2185	11.2189	11.2201
$\theta_{13}$	12.4936	12.4939	12.4945

Tabela 3.7: Valores dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 2.

Nos Casos 3, 5 e 4, que são casos cujo vetor média se afasta da origem, o limitante  $LSM_{\alpha}$  é melhor que o limitante  $LSM_{\Lambda}$ . As Figuras 3.12, 3.13 e 3.14, e as Tabelas 3.8, 3.9 e 3.10 ilustram os Casos 3, 4 e 5, respectivamente.

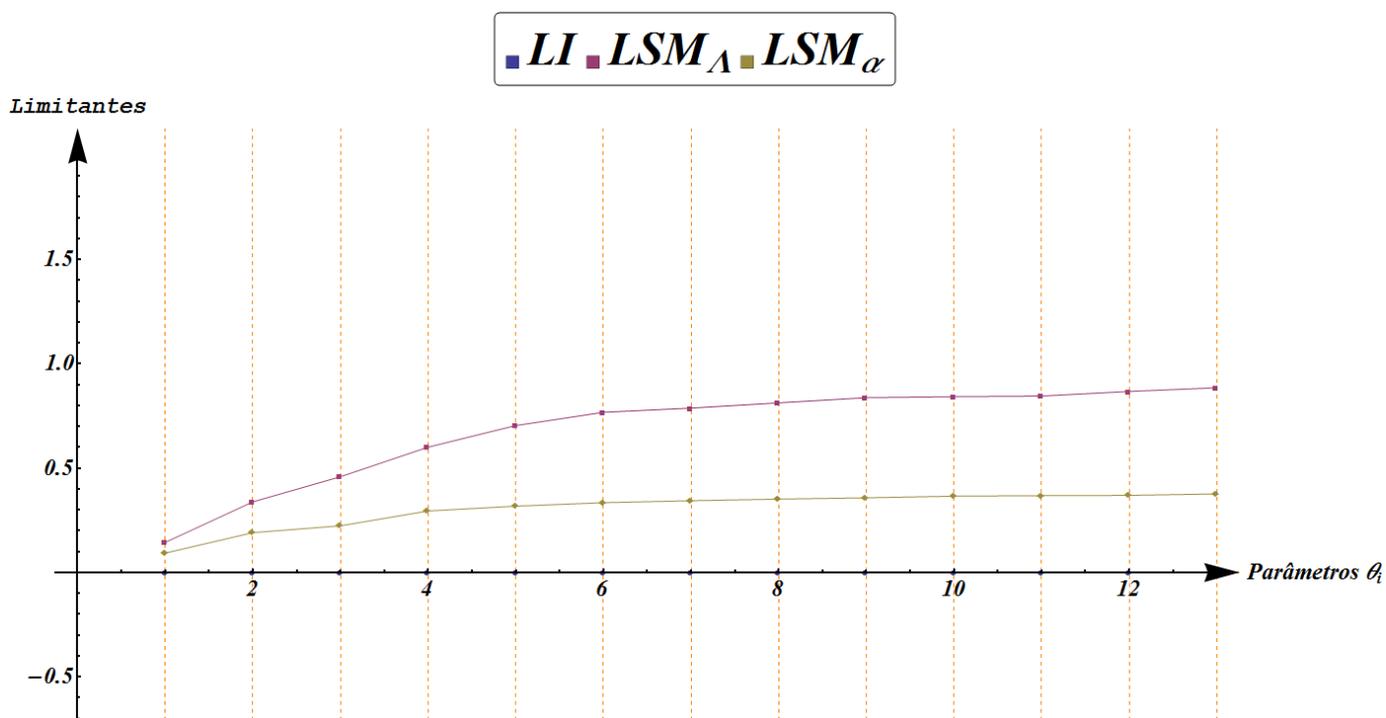


Figura 3.12: Gráfico dos valores relativos dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 3.

$\theta$	$LI$	$LSM_{\Lambda}$	$LSM_{\alpha}$
$\theta_1$	2.25382	2.57736	2.46111
$\theta_2$	4.10213	5.48796	4.88262
$\theta_3$	4.95886	7.23224	6.06533
$\theta_4$	8.12118	12.9939	10.512
$\theta_5$	9.93721	16.933	13.0894
$\theta_6$	12.1038	21.3917	16.148
$\theta_7$	13.6278	24.3578	18.3081
$\theta_8$	15.3045	27.7378	20.6791
$\theta_9$	16.9899	31.2122	23.06
$\theta_{10}$	19.8782	36.6125	27.1392
$\theta_{11}$	20.6624	38.1219	28.2582
$\theta_{12}$	21.4449	40.0339	29.3577
$\theta_{13}$	25.4678	47.9976	35.0459

Tabela 3.8: Valores dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 3.

Nos Casos 3, 5 e 4, que são casos cujo vetor média se afasta da origem, o limitante  $LSM_{\alpha}$  é melhor que o limitante  $LSM_{\Lambda}$ . A Figuras 3.13 e a Tabela 3.9 ilustram o Caso 4.

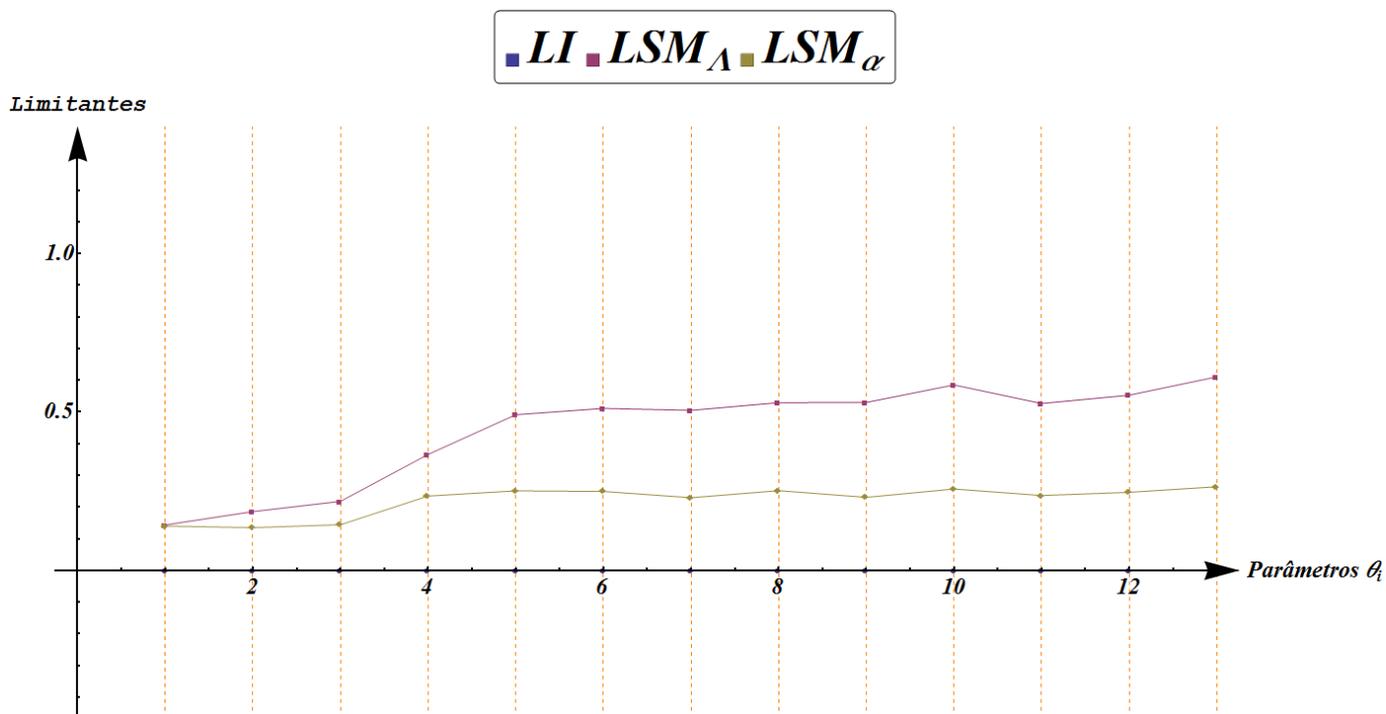


Figura 3.13: Gráfico dos valores relativos dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 4.

$\theta$	$LI$	$LSM_{\Lambda}$	$LSM_{\alpha}$
$\theta_1$	2.25376	2.57648	2.56882
$\theta_2$	3.75615	4.45301	4.26226
$\theta_3$	4.36503	5.31573	4.99403
$\theta_4$	7.05948	9.64217	8.71404
$\theta_5$	8.59487	12.8181	10.752
$\theta_6$	10.5157	15.892	13.1333
$\theta_7$	11.7988	17.7575	14.5031
$\theta_8$	13.1852	20.1639	16.5018
$\theta_9$	14.7915	22.6281	18.2008
$\theta_{10}$	17.2455	27.3321	21.6744
$\theta_{11}$	18.0726	27.5873	22.3422
$\theta_{12}$	18.6169	28.9045	23.2011
$\theta_{13}$	22.1076	35.603	27.9405

Tabela 3.9: Valores dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 4.

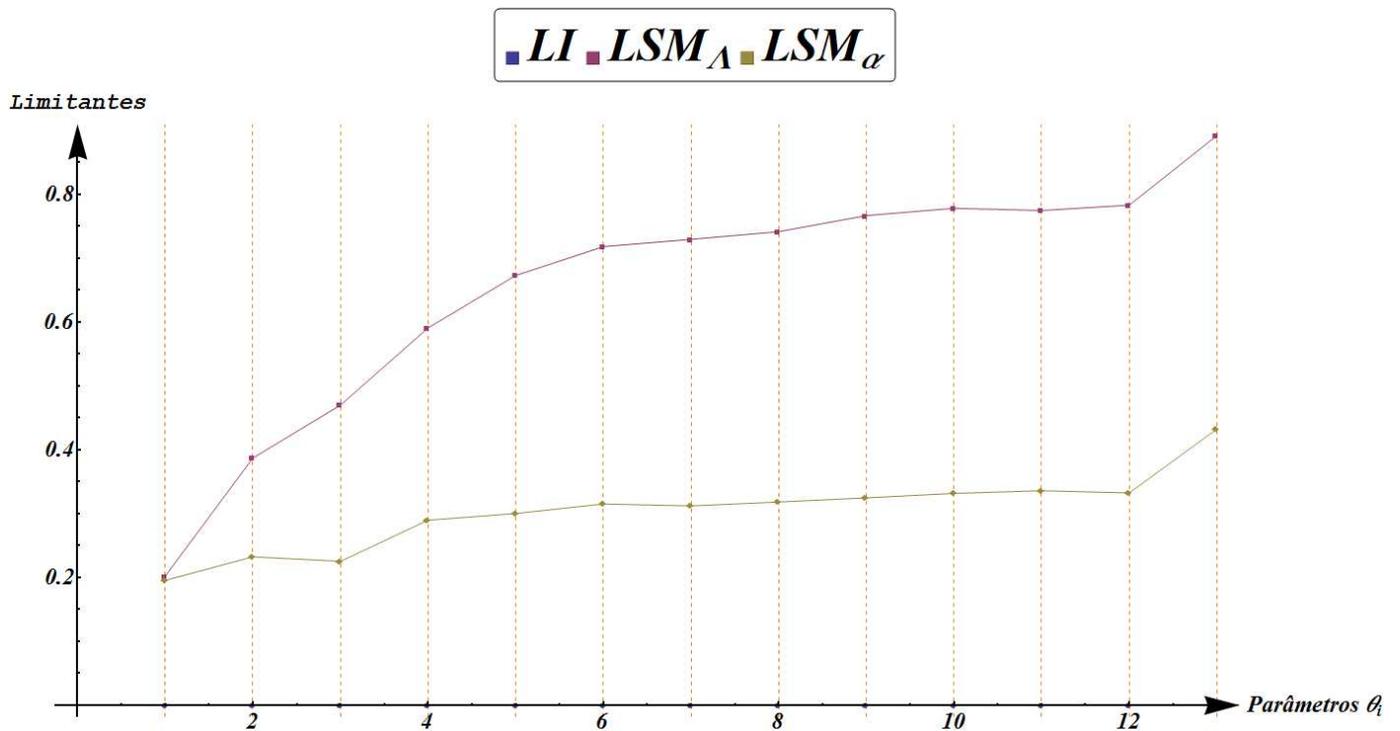


Figura 3.14: Gráfico dos valores relativos dos limitantes  $LI$ ,  $LMS_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 5.

$\theta$	$LI$	$LSM_{\Lambda}$	$LSM_{\alpha}$
$\theta_1$	2.38534	2.86747	2.85163
$\theta_2$	5.17697	7.17978	6.37871
$\theta_3$	6.68414	9.82359	8.18886
$\theta_4$	10.4606	16.6378	13.49
$\theta_5$	12.6911	21.2271	16.4999
$\theta_6$	15.7577	27.0676	20.7215
$\theta_7$	17.9876	31.102	23.5975
$\theta_8$	20.3661	35.4601	26.8445
$\theta_9$	22.6094	39.9337	29.9455
$\theta_{10}$	25.8903	46.0243	34.4741
$\theta_{11}$	27.507	48.8052	36.741
$\theta_{12}$	28.7822	51.3065	38.3458
$\theta_{13}$	32.0143	60.5537	45.8521

Tabela 3.10: Valores dos limitantes  $LI$ ,  $LSM_{\Lambda}$  e  $LSM_{\alpha}$  para cada ponto  $\theta = \theta_i$  no Caso 5.

Portanto para determinar um intervalo para a distância de Rao para um conjuntos de pontos formados por vetores da média  $\mu$  que estão próximos da origem e com a matriz  $\Sigma$  de covariância

com autovalores com valores não próximos de 1 é melhor utilizar o limitante  $LSM_\Lambda$ . Nos outros casos é melhor utilizar  $LSM_\alpha$ .

Em todos os casos, quando fazemos uma comparação entre os limitantes obtidos com e sem processos de minimização, ou seja, entre os limitantes  $LS_\alpha$  e  $LSM_\alpha$  e entre os limitantes  $LS_\Lambda$  e  $LSM_\Lambda$ , os limitantes com processos de minimização tiveram valores menores.

O software utilizado para fazer as simulações foi Wolfram Mathematica 9.0 e as funções de minimização para minimizar os limitantes  $LS_\alpha$  e  $LS_\Lambda$  foram funções do mesmo.

Observe que uma das distâncias que determinam o limitante  $LS_\Lambda$  é a distância restrita à subvariedade  $\Theta_D$ . O limitante  $LS_\Lambda$  pode ser melhorado se, por exemplo, usando procedimentos análogos ao de Calvo e Oller em [6], encontrarmos a distância entre pontos do conjunto  $\Theta_D$ .

# Apêndice A

## Valores do Parâmetros

Nesse apêndice mostramos quais foram os valores dos parâmetros  $\theta = (\mu, \Sigma)$  utilizados, em cada caso, nas simulações.

No Caso 1, os parâmetros  $\theta$  são tais que distância entre o vetor média  $\mu$  e o vetor  $\mathbf{0}$  é pequena ( $\mu$  está próximo da origem), os autovalores de  $\Sigma$  são maiores do que 1 e seus valores crescem. Os valores dos parâmetros  $\theta$  usados estão na Tabela A.1.

$\theta$	$\mu$	$\Sigma$
$\theta_1$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 1.10953 & -0.134778 \\ -0.134778 & 1.71553 \end{pmatrix}$
$\theta_2$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 4.84764 & -0.510336 \\ -0.510336 & 7.14226 \end{pmatrix}$
$\theta_3$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 15.0462 & 0.806651 \\ 0.806651 & 11.4193 \end{pmatrix}$
$\theta_4$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 27.1749 & -4.51166 \\ -4.51166 & 47.4607 \end{pmatrix}$
$\theta_5$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 82.0715 & 5.50669 \\ 5.50669 & 57.3118 \end{pmatrix}$
$\theta_6$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 370.053 & 36.2547 \\ 36.2547 & 207.041 \end{pmatrix}$
$\theta_7$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 867.128 & 30.0952 \\ 30.0952 & 731.811 \end{pmatrix}$
$\theta_8$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 2776.31 & 257.033 \\ 257.033 & 1620.62 \end{pmatrix}$
$\theta_9$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 4927.97 & -401.667 \\ -401.667 & 6733.98 \end{pmatrix}$
$\theta_{10}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 10669.5 & -387.018 \\ -387.018 & 12409.6 \end{pmatrix}$

$\theta_{11}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 32505.9 & -4627.01 \\ -4627.01 & 53310.3 \end{pmatrix}$
$\theta_{12}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 102613 & 10712.3 \\ 10712.3 & 54447.5 \end{pmatrix}$
$\theta_{13}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 208551 & -29686.6 \\ -29686.6 & 342030 \end{pmatrix}$

Tabela A.1: Valores dos parâmetros  $\theta$  utilizados no Caso 1.

No Caso 2, os parâmetros  $\theta$  são tais que o vetor média próximo da origem, os autovalores de  $\Sigma$  são menores do que 1 e seus valores decrescem. Seus valores são mostrados na Tabela A.2.

$\theta$	$\mu$	$\Sigma$
$\theta_1$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.909967 & 0.0714901 \\ 0.0714901 & 0.588526 \end{pmatrix}$
$\theta_2$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.20785 & 0.0148515 \\ 0.0148515 & 0.141073 \end{pmatrix}$
$\theta_3$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.0667144 & -0.00471265 \\ -0.00471265 & 0.0879039 \end{pmatrix}$
$\theta_4$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.0373887 & 0.0035542 \\ 0.0035542 & 0.0214079 \end{pmatrix}$
$\theta_5$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.0122636 & -0.00117832 \\ -0.00117832 & 0.0175616 \end{pmatrix}$
$\theta_6$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.00274948 & -0.000481459 \\ -0.000481459 & 0.00491426 \end{pmatrix}$
$\theta_7$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.00115488 & -0.0000474936 \\ -0.0000474936 & 0.00136843 \end{pmatrix}$
$\theta_8$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.000365558 & -0.000057978 \\ -0.000057978 & 0.000626244 \end{pmatrix}$
$\theta_9$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.000203915 & 0.0000121631 \\ 0.0000121631 & 0.000149226 \end{pmatrix}$
$\theta_{10}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.0000938314 & 2.92631 \times 10^{-6} \\ 2.92631 \times 10^{-6} & 0.0000806738 \end{pmatrix}$
$\theta_{11}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 0.0000311485 & 2.7035 \times 10^{-6} \\ 2.7035 \times 10^{-6} & 0.0000189927 \end{pmatrix}$
$\theta_{12}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 9.9497 \times 10^{-6} & -1.95756 \times 10^{-6} \\ -1.95756 \times 10^{-6} & 0.0000187515 \end{pmatrix}$
$\theta_{13}$	$\begin{pmatrix} 0.00777961 \\ 0.0556266 \end{pmatrix}$	$\begin{pmatrix} 4.85498 \times 10^{-6} & 4.21389 \times 10^{-7} \\ 4.21389 \times 10^{-7} & 2.96029 \times 10^{-6} \end{pmatrix}$

Tabela A.2: Valores dos parâmetros  $\theta$  utilizados no Caso 2.

Já no Caso 3, os parâmetros  $\theta$  foram tais que distância entre o vetor média  $\boldsymbol{\mu}$  e o vetor  $\mathbf{0}$  aumenta ( $\boldsymbol{\mu}$  se afasta da origem) e os autovalores da matriz  $\Sigma$  estão próximos de 1. Seus valores são mostrados na Tabela A.3.

$\theta$	$\boldsymbol{\mu}$	$\Sigma$
$\theta_1$	$\begin{pmatrix} 2.2427 \\ 1.6955 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_2$	$\begin{pmatrix} 6.7369 \\ 3.91334 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_3$	$\begin{pmatrix} 8.63661 \\ 8.41504 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_4$	$\begin{pmatrix} 24.1762 \\ 53.6762 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_5$	$\begin{pmatrix} 123.85 \\ 78.2736 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_6$	$\begin{pmatrix} 269.447 \\ 337.748 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_7$	$\begin{pmatrix} 767.156 \\ 520.948 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_8$	$\begin{pmatrix} 1748.77 \\ 1240.86 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_9$	$\begin{pmatrix} 3514.39 \\ 3522.44 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_{10}$	$\begin{pmatrix} 9765.78 \\ 18645.9 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_{11}$	$\begin{pmatrix} 28060.3 \\ 13783.4 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_{12}$	$\begin{pmatrix} 27429.1 \\ 37072.8 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$
$\theta_{13}$	$\begin{pmatrix} 188943 \\ 288139 \end{pmatrix}$	$\begin{pmatrix} 1.07956 & 0.00568373 \\ 0.00568373 & 1.05401 \end{pmatrix}$

Tabela A.3: Valores dos parâmetros  $\theta$  utilizados no Caso 3.

No Caso 4, os parâmetros  $\theta$  são tais que o vetor média  $\boldsymbol{\mu}$  se afasta da origem, os autovalores de  $\Sigma$  são maiores do que 1 e seus valores crescem. Os valores de  $\theta$  estão na Tabela A.4.

$\theta$	$\mu$	$\Sigma$
$\theta_1$	$\begin{pmatrix} 2.2427 \\ 1.6955 \end{pmatrix}$	$\begin{pmatrix} 1.10953 & -0.134778 \\ -0.134778 & 1.71553 \end{pmatrix}$
$\theta_2$	$\begin{pmatrix} 6.7369 \\ 3.91334 \end{pmatrix}$	$\begin{pmatrix} 4.84764 & -0.510336 \\ -0.510336 & 7.14226 \end{pmatrix}$
$\theta_3$	$\begin{pmatrix} 8.63661 \\ 8.41504 \end{pmatrix}$	$\begin{pmatrix} 15.0462 & 0.806651 \\ 0.806651 & 11.4193 \end{pmatrix}$
$\theta_4$	$\begin{pmatrix} 24.1762 \\ 53.6762 \end{pmatrix}$	$\begin{pmatrix} 27.1749 & -4.51166 \\ -4.51166 & 47.4607 \end{pmatrix}$
$\theta_5$	$\begin{pmatrix} 123.85 \\ 78.2736 \end{pmatrix}$	$\begin{pmatrix} 82.0715 & 5.50669 \\ 5.50669 & 57.3118 \end{pmatrix}$
$\theta_6$	$\begin{pmatrix} 269.447 \\ 337.748 \end{pmatrix}$	$\begin{pmatrix} 370.053 & 36.2547 \\ 36.2547 & 207.041 \end{pmatrix}$
$\theta_7$	$\begin{pmatrix} 767.156 \\ 520.948 \end{pmatrix}$	$\begin{pmatrix} 867.128 & 30.0952 \\ 30.0952 & 731.811 \end{pmatrix}$
$\theta_8$	$\begin{pmatrix} 1748.77 \\ 1240.86 \end{pmatrix}$	$\begin{pmatrix} 2776.31 & 257.033 \\ 257.033 & 1620.62 \end{pmatrix}$
$\theta_9$	$\begin{pmatrix} 3514.39 \\ 3522.44 \end{pmatrix}$	$\begin{pmatrix} 4927.97 & -401.667 \\ -401.667 & 6733.98 \end{pmatrix}$
$\theta_{10}$	$\begin{pmatrix} 9765.78 \\ 18645.9 \end{pmatrix}$	$\begin{pmatrix} 10669.5 & -387.018 \\ -387.018 & 12409.6 \end{pmatrix}$
$\theta_{11}$	$\begin{pmatrix} 28060.3 \\ 13783.4 \end{pmatrix}$	$\begin{pmatrix} 32505.9 & -4627.01 \\ -4627.01 & 53310.3 \end{pmatrix}$
$\theta_{12}$	$\begin{pmatrix} 27429.1 \\ 37072.8 \end{pmatrix}$	$\begin{pmatrix} 102613 & 10712.3 \\ 10712.3 & 54447.5 \end{pmatrix}$
$\theta_{13}$	$\begin{pmatrix} 188943 \\ 288139 \end{pmatrix}$	$\begin{pmatrix} 208551 & -29686.6 \\ -29686.6 & 342030 \end{pmatrix}$

Tabela A.4: Valores dos parâmetros  $\theta$  utilizados no Caso 4.

Finalmente, no Caso 5 os parâmetros  $\theta$  são tais que o vetor média  $\mu$  se afasta da origem, os autovalores de  $\Sigma$  são menores do que 1 e seus valores decrescem. Seus valores são mostrados na Tabela A.5.

$\theta$	$\mu$	$\Sigma$
$\theta_1$	$\begin{pmatrix} 2.2427 \\ 1.6955 \end{pmatrix}$	$\begin{pmatrix} 0.909967 & 0.0714901 \\ 0.0714901 & 0.588526 \end{pmatrix}$
$\theta_2$	$\begin{pmatrix} 6.7369 \\ 3.91334 \end{pmatrix}$	$\begin{pmatrix} 0.20785 & 0.0148515 \\ 0.0148515 & 0.141073 \end{pmatrix}$
$\theta_3$	$\begin{pmatrix} 8.63661 \\ 8.41504 \end{pmatrix}$	$\begin{pmatrix} 0.0667144 & -0.00471265 \\ -0.00471265 & 0.0879039 \end{pmatrix}$
$\theta_4$	$\begin{pmatrix} 24.1762 \\ 53.6762 \end{pmatrix}$	$\begin{pmatrix} 0.0373887 & 0.0035542 \\ 0.0035542 & 0.0214079 \end{pmatrix}$
$\theta_5$	$\begin{pmatrix} 123.85 \\ 78.2736 \end{pmatrix}$	$\begin{pmatrix} 0.0122636 & -0.00117832 \\ -0.00117832 & 0.0175616 \end{pmatrix}$
$\theta_6$	$\begin{pmatrix} 269.447 \\ 337.748 \end{pmatrix}$	$\begin{pmatrix} 0.00274948 & -0.000481459 \\ -0.000481459 & 0.00491426 \end{pmatrix}$
$\theta_7$	$\begin{pmatrix} 767.156 \\ 520.948 \end{pmatrix}$	$\begin{pmatrix} 0.00115488 & -0.0000474936 \\ -0.0000474936 & 0.00136843 \end{pmatrix}$
$\theta_8$	$\begin{pmatrix} 1748.77 \\ 1240.86 \end{pmatrix}$	$\begin{pmatrix} 0.000365558 & -0.000057978 \\ -0.000057978 & 0.000626244 \end{pmatrix}$
$\theta_9$	$\begin{pmatrix} 3514.39 \\ 3522.44 \end{pmatrix}$	$\begin{pmatrix} 0.000203915 & 0.0000121631 \\ 0.0000121631 & 0.000149226 \end{pmatrix}$
$\theta_{10}$	$\begin{pmatrix} 9765.78 \\ 18645.9 \end{pmatrix}$	$\begin{pmatrix} 0.0000938314 & 2.92631 \times 10^{-6} \\ 2.92631 \times 10^{-6} & 0.0000806738 \end{pmatrix}$
$\theta_{11}$	$\begin{pmatrix} 28060.3 \\ 13783.4 \end{pmatrix}$	$\begin{pmatrix} 0.0000311485 & 2.7035 \times 10^{-6} \\ 2.7035 \times 10^{-6} & 0.0000189927 \end{pmatrix}$
$\theta_{12}$	$\begin{pmatrix} 27429.1 \\ 37072.8 \end{pmatrix}$	$\begin{pmatrix} 9.9497 \times 10^{-6} & -1.95756 \times 10^{-6} \\ -1.95756 \times 10^{-6} & 0.0000187515 \end{pmatrix}$
$\theta_{13}$	$\begin{pmatrix} 188943 \\ 288139 \end{pmatrix}$	$\begin{pmatrix} 4.85498 \times 10^{-6} & 4.21389 \times 10^{-7} \\ 4.21389 \times 10^{-7} & 2.96029 \times 10^{-6} \end{pmatrix}$

Tabela A.5: Valores dos parâmetros  $\theta$  utilizados no Caso 5.

# Referências Bibliográficas

- [1] Amari, S. e Nagaoka, H., *Differential Geometrical Methods in Statistics*, Lecture Notes in Statistics, 28, Springer-Verlag, Heidelberg, 1986.
- [2] Amari, S. e Nagaoka, H. *Methods of Information Geometry*, Translations of Mathematical Monographs, Vol.191, Am. Math. Soc., 2000.
- [3] Atkinson, C. e Mitchell, A. F. S., *Rao's Distance Measure*, Samkhyā- The Indian Journal of Statistics, 43:345-365, 1981.
- [4] Burbea, J., *Informative geometry of probability spaces*, Expositiones Mathematica 4, 347-378, 1986.
- [5] Calvo, M. e Oller, J. M., *A distance between multivariate normal distributions based in an embedding into the Siegel group*, Journal of Multivariate Analysis 35.2, 223-242, 1990.
- [6] Calvo, M., e Oller, J. M., *An explicit solution of information geodesic equations for the multivariate normal model*, Statistics and Decisions 9, 119-138, 1991.
- [7] Fisher, R. A., *On the mathematical foundations of theoretical statistics*, Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character, 222, 309-368, 1921.
- [8] Carmo, M. P., *Geometria diferencial de curvas e superfícies*, Sociedade Brasileira de Matemática, Rio de Janeiro, 2005.
- [9] Carmo, M. P. *Geometria Riemanniana*, IMPA, Rio de Janeiro, 2008.
- [10] Chentsov, N. N., *Statistical decision rules and optimal inference*, Vol. 53, AMS Bookstore, 1982.
- [11] Costa, S. I., Santos, S. A., e Strapasson, J. E., *Fisher information matrix and hyperbolic geometry*, In Information Theory Workshop, 34-36, IEEE, 2005.
- [12] Costa, S. I. R., Santos, S. A., e Strapasson, J. E., *Fisher information distance: a geometrical reading*, arXiv preprint arXiv:1210.2354, 2012.
- [13] Förstner, W. e Moonen, B., *A metric for covariance matrices*, Quo vadis geodesia, 113-128, 1999.

- [14] Hogg, R. V., e Craig, A., *Introduction to mathematical statistics*, 1994.
- [15] Lenglet, C., Rousson, M., Deriche, R. e Faugeras, O., *Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing*, Journal of Mathematical Imaging and Vision, 25(3), 423-444, 2006.
- [16] Lovrić, M., Min-Oo, M. e Ruh, E. A., *Multivariate normal distributions parametrized as a Riemannian symmetric space*, Journal of Multivariate Analysis, 74(1), 36-48, 2000.
- [17] Mahalanobis, P. C., *On the generalized distance in statistics*, Proceedings of the National Institute of Sciences (Calcutta) 2, 49-55, 1936.
- [18] Moakher, M., *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*, SIAM Journal on Matrix Analysis and Applications 26(3), 735-747, 2005.
- [19] Nielsen, F., e Garcia, V., *Statistical exponential families: A digest with flash cards*, arXiv preprint arXiv:0911.4863, (2009).
- [20] Rao, C. R., *Information and the accuracy attainable in the estimation of statistical parameters*, Bulletin of the Calcutta Math. Soc. 37:81-91, 1945.
- [21] Siegel, C. L., *Symplectic geometry*, American Journal of Mathematics 65(1), 1-86, 1943.
- [22] Skovgaard, L. T., *A Riemannian geometry of the multivariate normal model*, Scand. J. Statist., 11:211-223, 1984.