Universidade Federal do Rio de Janeiro
Instituto de Matemática

Programa de Pós-Graduação em Matemática

Geometria da Informação - O Teorema de Cramér-Rao

Wellington Santiago

Dissertação de Mestrado

UFRJ
Rio de Janeiro - 2017

Dedicado à Andrea Passos

Agradecimentos

Primeiramente agradeço a Deus por ter me dado força e saúde pra cumprir está difícil missão.

Agradeço aos meus pais, pela educação que me deram, que apesar ter sido bem simples, foi eficiente.

Agradeço ao meu orientador Heudson Mirandola pelo seu companheirismo e dedicação, por sua enorme participação neste trabalho , por sua paciência e competência.

Agradeço à Universidade Federal do Rio de Janeiro pela oportunidade de estudar neste instituto de excelência e por me servir com ótimos professores.

Agradeço à Universidade Federal Fluminense. Lá me graduei e fui feliz em conhecer professores que me encentivaram e me inspiraram.

Agradeço a minha esposa Andrea Passos, pelo companheirismo irrestrito, por me dar forças e sempre acreditar em mim.

Agradeço ao apoio dos meus amigos e familiares. Em especial, agradeço ao meu amigo e colega de mestrado Gil Navarro, por sua efetiva ajuda em vários topicos deste trabalho.

<*DIGITE AQUI A CITAÇÂO*>

—<AUTOR> (<NOTA>)

Resumo

A Geometria da Informação é uma área da matemática relativamente nova que faz uso de ferramentas geométricas no estudo de modelos estatísticos. Neste presente trabalho, daremos as condições suficientes para enxergarmos modelos estatísticos como variedades Riemannianas, a saber, a métrica Riemanniana empregada é dada pela matriz de informação de Fisher. Com a noção de conexões afins duais e as α-conexões, investigamos algumas propriedades de modelos estatísticos, principalmente aqueles modelos que são famílias exponenciais.

Na última parte deste trabalho, como aplicação da robusta estrutura geometrica empregada nos modelos estatísticos, provamos um importante teorema de inferência estatística, a saber, o Teorema de Cramér-Rao. Por fim, damos uma condição necessária e suficiente para que modelos estatísticos possuam estimadores eficientes.

Palavras-chave: Geometria da Informação, Matriz de Informação de Fisher, Conexões Duais, α-conexões, Famílias Exponenciais , Teorema de Cramér-Rao.

Abstract

Information Geometry is a relatively new area in mathematics that makes use of geometrical tools in the study of statistical models. In this present work, we will give conditions to see statistical models as Riemannian manifolds, namely, the Riemannian metric used is given by the Fisher information matrix. We eill be able to investigate other properties of statistical models with the notion of dual connections and α-connections.

In the last part of this work, as an application of the robust geometric structure employing statistical models, we prove an important theorem of statistical inference, the Cramér-Rao Theorem. Finally, we provide a necessary and sufficient condition for statistical models to have efficient estimators

Keywords: Information Geometry, Fischer information matrix, Dual connections, α-Connection, Exponential family, Cramér-Rao Theorem.

Sumário

CAPÍTULO 1

Conceitos básicos de probabilidade

O objetivo deste capítulo é fazer uma breve introdução sobre os conceitos básicos da teoria de probabilidade, por esse motivo, tal introdução é feita sem grande rigor matemático e os resultados não são todos demonstrados. Para mais detalhes consultar [8] ou [6]. O propósito principal é introduzir os conceitos de variáveis aleatórias independentes e igualmente distribuidas, esperança e matriz de variância e covariância de vetores aleatórios.

1.1 Espaço de probabilidade

Definição 1.1. *Seja Ω um conjunto qualquer e Σ uma coleção de subconjuntos de Ω. Dizemos que Σ é uma* σ-algebra *quando satisfaz as seguntes condições:*

(i) $\emptyset, \Omega$ *pertencem a* Σ.

(ii) Se A pertence a Σ, então o conjunto complementar A^c pertence a Σ.

(iii) Se (A_n) é uma sequência finita ou enumerável de conjuntos em Σ, então a união $\cup_{n=1}^{\infty} A_n$ pertece a Σ.

O par (Ω, Σ) é chamado de espaço mensurável. *Um conjunto de Σ é chamado de* conjunto Σ-mensurável.

Observe que se (A_n) é uma sequência em Σ, então (A_n^c) também é uma sequência em Σ. Pela lei de Morgan, por *(ii)* e *(iii)* temos que $(\cup_{n=1}^{\infty}(A_n)^c)^c = \cap_{n=1}^{\infty} A_n \in \Sigma$. Isto é, uma σ-algebra é fechada para a intercessão enumerável.

Definição 1.2. *Uma função $\mu : \Sigma \to \mathbb{R} \cup \{\pm\infty\}$ é chamada de* medida *sobre Σ quando satisfaz as seguintes propriedades:*

(i) $\mu(A) \geq 0$ *para todo* $A \in \Sigma$.

(ii) $\mu(\emptyset) = 0$.

(iii) Se (A_n) é uma sequência de conjuntos disjuntos em Σ, então $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.

O trio (Ω, Σ, μ) é chamado de espaço de medida.

Uma função $f : \Omega \to \mathbb{R}$ é dita ser Σ-*mensurável* (ou apenas mensurável) quando para todo $\alpha \in \mathbb{R}$, tem-se que $f^{-1}((\alpha, \infty)) = \{\omega \in \Omega : f(\omega) > \alpha\} \in \Sigma$.

Lema 1.1. *As seguintes afirmações são equivalentes para uma função f de Ω em $\mathbb{R}$:*

(a) Para todo $\alpha \in \mathbb{R}$, o conjunto $A_\alpha = f^{-1}((\alpha,\infty)) \in \Sigma$.

(b) Para todo $\alpha \in \mathbb{R}$, o conjunto $B_\alpha = f^{-1}((-\infty,\alpha]) \in \Sigma$.

(c) Para todo $\alpha \in \mathbb{R}$, o conjunto $C_\alpha = f^{-1}([\alpha,\infty)) \in \Sigma$.

(d) Para todo $\alpha \in \mathbb{R}$, o conjunto $D_\alpha = f^{-1}((-\infty,\alpha)) \in \Sigma$.

(e) Para todo α e $\beta \in \mathbb{R}$ com $\alpha < \beta$ o conjunto $E = f^{-1}((\alpha,\beta)) \in \Sigma$.

Demonstração. Já que A_α e B_α são complementares um do outro, a afirmação (a) é equivalente a afirmação (b). Analogamente, a afirmação (c) é equivalente a afirmação (d). Se vale (a), então $A_{\alpha-1/n} \in \Sigma$, para todo n. Já que

$$C_\alpha = \bigcap_{n=1}^{\infty} A_{\alpha-1/n},$$

segue que $C_\alpha \in \Sigma$. Assim (a) implica (c). Por outro lado, note que

$$A_\alpha = \bigcup_{n=1}^{\infty} C_{\alpha+1/n}.$$

Portanto, (c) implica (a). Agora suponha que vale (a), conseguentemente vale (c). Note que

$$f^{-1}((\alpha,\beta)) = f^{-1}((-\infty,\beta)\cap(\alpha,\infty)) = f^{-1}((-\infty,\beta)\cap f^{-1}((\alpha,\infty)) = D_\beta \cap A_\alpha.$$

Portanto, $f^{-1}((\alpha,\beta) \in \Sigma$, isto é, (a) implica (e). Reciprocamente se vale (e), basta notar que

$$A_\alpha = \bigcup_{n=1}^{\infty} f^{-1}((\alpha,\alpha+n)).$$

logo, (e) implica (a). □

Agora estamos prontos para definirmos os conceitos básicos de probabilidade. Uma *medida de probabilidade* sobre um conjunto Ω ou apenas uma *probabilidade* é uma medida μ com $\mu(\Omega) = 1$. A medida μ neste caso será denotada habitualmente por P. Um *espaço de probabilidade* é um espaço de medida (Ω,Σ,P), com uma medida de probabilidade P. Neste caso os elementos de Σ são chamados de *eventos*. Uma medida de probabilidade atribui um número real não negativo a cada evento, chamado de *probabilidade do evento*. O conjunto Ω é chamado de espaço amostral,Ω é o conjunto que consiste de todos os resultados possíveis de um experimento aleatório. Uma variável aleatória X é qualquer função mensurável $f : \Omega \to \mathbb{R}$.

Existem três classes distintas de variáveis aleatórias: discreta, contínua, e mistura das duas. Vamos discutir apenas os dois primeiros tipos.

Definição 1.3. *Uma variável aleatória é dita ser discreta quando a imagem de X, denotada por χ é da forma $\chi = \{x_1,x_2,\dots,x_n\}$ ou $\chi = \{x_1,x_2,\dots,\}$; em outras palavras, uma variavél aleatória é dita ser discreta quando sua imagem é um conjunto enumerável. A função $P_X(x) = P(X = x) = P(\{\omega \in \Omega\ ;\ X(\omega) = x\})$, $x \in \mathbb{R}$ é chamada de função de probabilidade de X.*

Pela definição segue que $P_X(x) \geq 0$ e que $\sum_{x \in \chi} P_X(x) = 1$.

Dada X uma variável aleatória e A um subconjunto Boreliano de $\mathbb{R}$, denotamos $P(X \in A) = P(X^{-1}(A)) = P(\{\omega \in \Omega\ ;\ X(\omega) \in A\})$.

Definição 1.4. *Uma variável aleatória X é dita ser contínua quando existe uma função não negativa $f_X : \mathbb{R} \to \mathbb{R}$ de modo que $\int_{-\infty}^{\infty} f(x)dx = 1$ e para todo a e $b \in \mathbb{R}$, com $a < b$ tem-se,*

$$P(X \in (a,b)) = \int_a^b f_X(x)dx.$$

A função f_X é chamada de função de distribuição de probabilidade de X.

Note que teoricamente qualquer função $f : \mathbb{R} \to \mathbb{R}$ não negativa, que cumpre a propriedade $\int_{-\infty}^{\infty} f(x)dx = 1$ da origem a uma variável aleatória contínua. Algumas vezes escreveremos $\int f(x)dx$ em vez de $\int_{-\infty}^{\infty} f(x)dx$

Exemplo 1.1. *Dizemos que uma variável aleatória discreta X com $\chi = \{x_0, x_1, \ldots,\}$ tem distribuição de Poison com parâmetro $\lambda \in (0,\infty)$ e escrevemos $X \sim Pois(\lambda)$ quando a função de probabilidade de X é dada por*

$$P_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Note que,

$$\sum_{x \in \chi} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{i=0}^{n}\frac{\lambda^x}{x_i!} = e^{-\lambda}e^{\lambda} = 1.$$

Exemplo 1.2. *Dizemos que uma variável aleatória T tem distribuição exponencial com parâmetros $\beta > 0$ e escrevemos $T \sim \exp(\beta)$ se sua função de distribuição de probabilidade é dada por*

$$f(t;\beta) = \begin{cases} \frac{1}{\beta}e^{-t/\beta}, & \text{se } t \geq 0 \\ \\ 0, & \text{se } t < 0. \end{cases}$$

Claramente $f(t;\beta) \geq 0$, para todo $t \in \mathbb{R}$ e é imediato verificar que $\int_{-\infty}^{\infty} f(t)dt = 1$.

.

1.2 Distribuições bivariadas

Definição 1.5. *Dado X e Y um par de variáveis aleátorias discretas, definimos a função de probabilidade conjunta de X e Y por $f_{X,Y}(x,y) := f(x,y) = P(X = x \text{ e } Y = y)$. Frequentemente escreveremos P(X=x , Y=y) em vez de P(X=x e Y=y).*

Definição 1.6. *Dado X e Y um par de variáveis aleátorias continuais, chamaremos uma função $f(x,y)$ de função de distribuição probabilidade conjunta de X e Y se*

1. *$f(x,y) \geq 0$ para todo (x,y).*

2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) = 1$ e

3. Para todo Boreliano $A \subset \mathbb{R}^2$, tem-se $P((X,Y) \in A) = \int\int_A f(x,y)dxdy$.

Definição 1.7. *Se (X,Y) tem uma função de probabilidade conjunta f_{XY}, definimos a função de probabilidade marginal para X por*

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f(x,y)$$

e a função de probabilidade marginal para Y por

$$f_X(x) = P(X = x) = \sum_x P(X = x, Y = y) = \sum_x f(x,y).$$

Definição 1.8. *No caso de variáveis aleatórias continuais definimos as funções de distribuição de probabilidade marginais de X e Y respectivamente por*

$$f_X(x) = \int f(x,y)dy, \text{ e } f_Y(y) = \int f(x,y)dx.$$

Definição 1.9. *Dizemos que duas variáveis aleatórias X e Y são independentes quando para todos A e B subconjuntos Borelianos de $\mathbb{R}$, vale que*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Teorema 1.2. *Sejam (X,Y) com função de probabilidade conjunta f_{XY}. Então X e Y são independentes se e somente, se $f_{XY}(x,y) = f_X(x)f_Y(y)$ para todo x e y.*

Exemplo 1.3. *Sejam X e Y variáveis aleatórias distribuidas de acordo com a tabela abaixo.*

	$Y = 0$	$Y = 1$	
$X = 0$	$1/4$	$1/4$	$1/2$
$X = 1$	$1/4$	$1/4$	$1/2$
	$1/2$	$1/2$	1

Note que, $f_X(0) = f_X(1) = 1/2$ e $f_Y(0) = f_Y(1) = 1/2$. X e Y são independentes, pois $f_{XY}(0,0) = f_X(0)f_Y(0)$, $f_{XY}(1,0) = f_X(1)f_Y(0)$, $f_{XY}(0,1) = f_X(0)f_Y(1)$ e $f_{XY}(1,1) = f_X(1)f_Y(1)$. Por outro lado, se X e Y distribuidas de acordo com a seguinte tabela.

	$Y = 0$	$Y = 1$	
$X = 0$	$1/2$	0	$1/2$
$X = 1$	0	$1/2$	$1/2$
	$1/2$	$1/2$	1

Então, X e Y não são independentes, pois $f_{XY}(0,1) = 0$ e $f_X(0)f_Y(1) = 1/4$.

Exemplo 1.4. *Sejam X e Y cuja função de distribuição de probabilidade conjunta é dada por*

$$f(x,y) = \begin{cases} 2e^{-(x+2y)}, & \text{se } x > 0 \text{ e } y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

É simples ver que

$$f_X(x) = \begin{cases} -e^{-x}, & \text{se } x > 0 \text{ e } y > 0 \\ 0, & \text{caso contrário.} \end{cases} \quad \text{e} \quad f(x,y) = \begin{cases} 2e^{-2y}, & \text{se } x > 0 \text{ e } y > 0 \\ 0, & \text{caso contrário.} \end{cases}$$

Logo $f_{XY}(x,y) = f_X(x) f_Y(y)$ e portanto X e Y são variáveis aleatórias independentes.

1.3 Distribuição multivariadas e amostras IID

Seja $\mathbf{X} = (X_1, \dots, X_n)$, onde $X_1, \dots, X_n$ são variáveis aleátorias. Vamos chamar $\mathbf{X}$ de *vetor aleatório*. Denotamos a função de distribuição probabilidade conjunta (quando $X_1, \dots, X_n$ são variáveis aleatórias contínuas) ou a função de probabilidade conjunta (quando $X_1, \dots, X_n$ são variáveis aleatórias discretas) apenas por $f(x_1, \dots, x_n)$. Assim, como fizemos no caso bivariado, podemos definir as probabilidades marginais f_{X_k}, definindo, por exemplo, a probabilidade marginal de X_1 pondo

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \dots dx_n.$$

Dizemos que as variáveis aleatórias $X_1, \dots, X_n$ são independentes se, para quaisquer Borelianos $A_1, \dots, A_n$ de $\mathbb{R}$, tem-se

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i).$$

Pelo teorema (1.2), $X_1, \dots, X_n$ são independentes se e somente, se $f(x_1, \dots, x_n) = \prod_{=1}^{n} f_{X_i}(x_i)$. para todo $x_1, \dots, x_n$

Definição 1.10. *Se $X_1, \dots, X_n$ são variáveis aleatórias independentes e cada uma delas tem a mesma função de distribuição de probabilidade marginal (caso contínuo) ou a mesma função de probabilidade marginal (caso discreto) f, dizemos que $X_1, \dots, X_n$ são independentes e igualmente distribuidas (IID) e escrevemos $X_1, \dots, X_n \sim f$. Também chamamos $X_1, \dots, X_n$ de uma amostra aleátoria de tamanho n de f*

No exemplo (1.4), as variáveis aleatórias X e Y não são IID.

1.4 Esperança e Variância de uma Variável Aleatória

Definição 1.11. *A esperança de uma variável aleatória X, quando existir, é assim definida* *(a) X discreta:* $E[X] = \sum_{x \in \chi} x P_X(x)$.

(b) X contínua: $E[X] = \int x f_X(x) dx$.

Algumas vezes denotaremos a esperança de X como μ_X ou apenas por μ.

Teorema 1.3. *Seja $Y = g(X)$. Então se Y tem esperança bem definida*

$$E[Y] = E[g(X)] = \begin{cases} \sum_{x \in \chi} g(x) P_X(x), & \text{se } X \text{ é discreta} \\ \int g(x) f_X(x) dx, & \text{se } X \text{ é contínua.} \end{cases}$$

Propriedades importantes da esperança.

(E_1) Se $X = c$ (isto é, $X(\omega) = c$ para todo $\omega \in \Omega$), então $E[X] = c$.

(E_2) Se $X_1, \ldots, X_n$ são variáveis aleatórias e $a_1, \ldots, a_n$ são constantes, então

$$E\left[\sum_i a_i X_i\right] = \sum_i a_i E[X_i].$$

Definição 1.12. *Seja X uma variável aleatória com esperança μ. A variância de X (denotada por $V[X]$ ou σ_X^2 ou σ^2), quando existir, é assim definida*

$$V[X] = E[(X-\mu)^2] = \begin{cases} \sum_{x \in \chi} (x-\mu)^2 P_X(x), & \text{se } X \text{ é discreta} \\ \int (x-\mu)^2 f_X(x) dx, & \text{se } X \text{ é contínua.} \end{cases} \tag{1.1}$$

Propriedades importantes da variância.

(V_1) Se $X = c$, então $V[X] = 0$.

(V_2) $V(X) = E(X^2) - \mu^2$.

(V_3) Se a e b são constantes então $V(aX+b) = a^2 V(X)$.

(E_a) Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes e $a_1, \ldots, a_n$ são constantes, então

$$V\left[\sum_i a_i X_i\right] = \sum_i a_i^2 V[X_i].$$

Exemplo 1.5. *Dizemos que uma variável aleatória X tem distribuição normal com parâmetros $\mu \in (-\infty, \infty)$, $\sigma^2 \in (0, \infty)$ e escrevemos $X \sim N(\mu, \sigma)$ quando sua função de distribuição de probabilidade é dada por*

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ para todo } -\infty < x < \infty.$$

Evidentemente, $f(x; \mu, \sigma) \geq 0$, para todo x e pode-se provar que:

$$\int f(x; \mu, \sigma) dx = 1, \ \ E[X] = \mu \ \text{ e } \ V[X] = \sigma^2.$$

As distribuições normais desempenham um papel fundamental em probabilidade e estatística. Um dos teoremas mais importantes da probabilidade é o teorema central do limite que garante que a distribuição da média aritmética de variáveis aleátorias IID é aproximadamente uma distribuição normal, veja [8], pag. 77. A figura 1.1 é uma particular curva normal, determinada por uma média μ e uma variância σ^2.

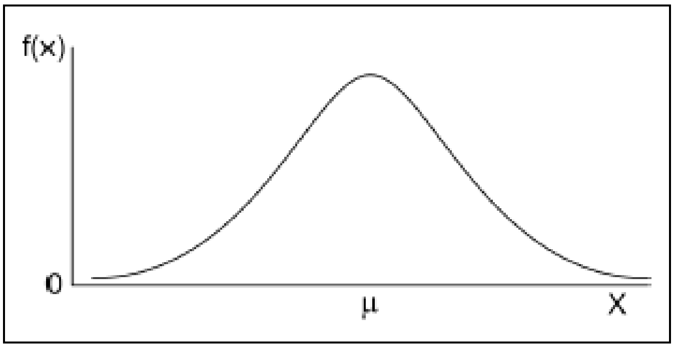

Figura 1.1 Curva Normal

Exemplo 1.6. *Quando $X \sim Pois(\lambda)$ temos*

$$E[X] = \sum_{k=0}^{\infty} \frac{ke^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{k\lambda^(k-1)}{k!} = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{(k)!} = \lambda.$$

De de maneira analoga obtemos que $E[X^2] = \lambda^2 + \lambda$, logo $V[X] = \lambda$.

Para $T \sim \exp(\beta)$, através de integração por partes, obtemos que $E[T] = \beta$ e $V[T] = \beta^2$.

Quando $\mathbf{X}$ é um vetor aleátorio, $\mathbf{X}= (X_1, \ldots, X_n)^T$ definimos a *esperança de* $\mathbf{X}$ como o vetor

$$E[\mathbf{X}] = (E[X_1], \ldots, E[X_n])^T$$

e a *matriz de variância e covariância de* $\mathbf{X}$ por

$$V[\mathbf{X}] = \begin{bmatrix} V[X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_n] \\ Cov[X_2, X_1] & V[X_2] & \cdots & Cov[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & Cov[X_n, X_2] & \cdots & V[X_n] \end{bmatrix}.$$

onde $Cov[X, Y] = E[(X - E[X])(Y - E[Y])]$, chamada de covariância de X e Y.

CAPÍTULO 2

Noções de geometria Riemanniana

Neste capítulo apresentamos alguns conceitos básicos de geometria Riemanniana. Não faremos a demostração de todos os resultados apresentados. Para uma boa compreensão do próximo capitulo, daremos destaque para variedades planas e falaremos sobre conexões afins duais e da geometria Hessiana, conceitos que não são tão comuns nos livros tradicionais de geometria diferencial.

2.1 Variedades diferenciáveis

Definição 2.1. *Uma variedade difenciável de dimensão n é um conjunto M e uma família de aplicações injetivas $x_\alpha : U_\alpha \subset \mathbb{R}^n \to M$ de abertos U_α de $\mathbb{R}^n$ tais que:*

(i) $\bigcup\limits_{\alpha} x_\alpha(U_\alpha) = M$.

(ii) Para todo par α, β, com $x_\alpha(U_\alpha) \cap x_\beta(U_\beta) = W \neq \emptyset$, os conjuntos $x_\alpha^{-1}(W)$ e $x_\beta^{-1}(W)$ são abertos de $\mathbb{R}^n$ e as aplicações $x_\beta^{-1} \circ x_\alpha$ são diferenciáveis (veja Figura 2.1).

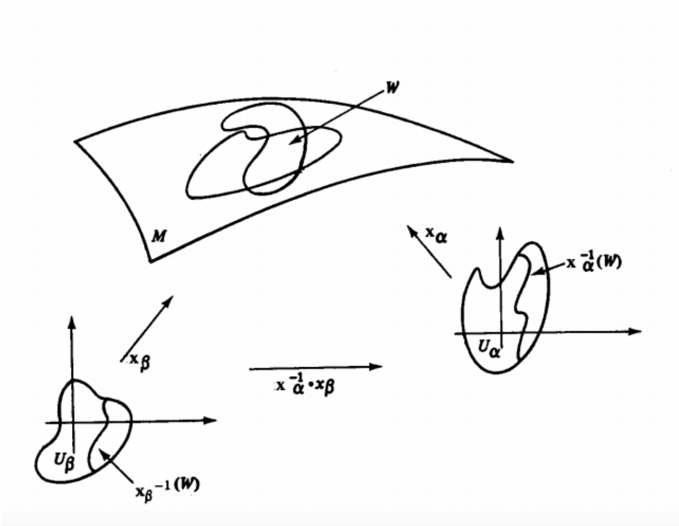

Figura 2.1 Variedades diferenciáveis. Extraída de [5]

(iii) A família $\{(U_\alpha, x_\alpha)\}$ é máximal relativamente às condições (i) e (ii)

O par (U_α, x_α) (ou a aplicação x_α) com $p \in x_\alpha(U_\alpha)$ é chamada uma *parametrização* (ou *sistema de coordenadas*) de M em p; $x_\alpha(U_\alpha)$ é então chamada uma *vizinhança coordenada* em p. Uma família $\{(U_\alpha, x_\alpha)\}$ satisfazendo (i) e (ii) é chamada uma *estrutura diferenciável* em M. Algumas vezes vamos nos referir a um sistema de coordenadas $\mathbf{x}$ como $[x_i]$ ou $([x^i])$. A condição (iii) comparece por razões puramente técnicas. Portanto, diremos que uma variedade diferencíavel é um conjunto munido de uma estrutura diferencíavel.
Observação 1: Uma estrutura diferenciável em um conjunto M induz de maneira natural uma topologia em M. Basta definir que $A \subset M$ é aberto de S se $x_\alpha^{-1}(A \cap x_\alpha(U_\alpha))$ é um aberto de $\mathbb{R}^n$ para todo α. É imediato verificar que M e o conjunto vazio são abertos, que a união de abertos é aberto e que a intersecção finita de abertos é aberto. Observe que a topologia é definida de tal modo que os conjuntos $x_\alpha(U_\alpha)$ são abertos e as aplicações x_α são contínuas.

O espaço euclidiano $\mathbb{R}^n$ e seus subconjuntos abertos, com a estrutura diferenciável dada pela identidade são exemplos triviais de variedades diferenciáveis. De agora em diante, quando indicarmos uma variedade po M^n, o índice superior n indicará a dimensão de M.

Definição 2.2. *Sejam M_1^n e M_2^m variedades diferenciáveis. Uma aplicação $\varphi : M_1 \to M_2$ é diferenciável em $p \in M_1$ se dada uma parametrização $y : V \subset \mathbb{R}^m \to M_2$ em $\varphi(p)$ existe uma parametrização $x : U \subset \mathbb{R}^n \to M_1$ em p tal que $\varphi(x(U)) \subset y(V)$ e a aplicação*

$$y^{-1} \circ \varphi \circ x : U \subset \mathbb{R}^n \to \mathbb{R}^m \tag{2.1}$$

é diferenciável em $x^{-1}(p)$ (veja Figura 2.2). A função φ é diferenciável em um aberto de M_1 se é diferenciável em todos os pontos deste aberto.

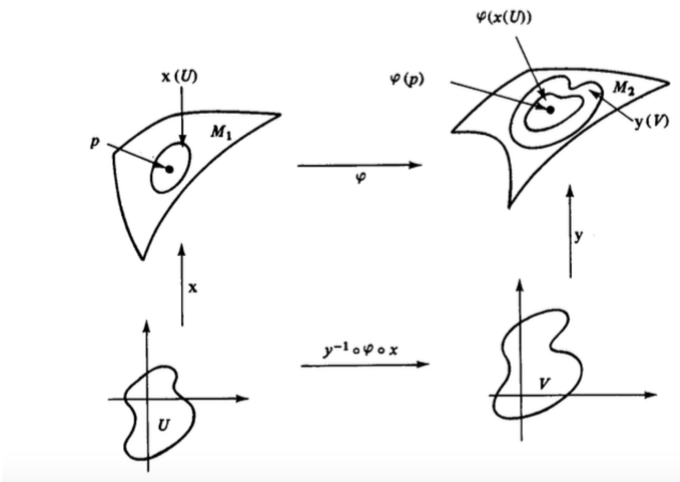

Figura 2.2 Funções diferenciáveis. Extraída de [5]

Decorre da condição (ii) da definição (2.1) que a definição dada é independente da escolha das parametrizações. A aplicação (2.1) é chamada *expressão* de φ nas parametrizações x e

y. Já que $\mathbb{R}$ é uma variedade diferenciável, fica definido o que é uma função $f : M \to \mathbb{R}$ ser diferenciável, o mesmo vale para funções de $\mathbb{R}$ em M. Uma função diferenciável $\alpha : \mathbb{R} \to M$ será chamada de *curva*(diferenciável) em M. Denotaremos por $D(M)$ o conjunto de todas as funções diferenciávei de M em $\mathbb{R}$. Para $f \in D(M)$ a expressão de f na parametrização $[x_i]$ é dada por $f \circ x : U \subset \mathbb{R}^n \to \mathbb{R}$, definimos $\frac{\partial f}{\partial x_i}(p) := \frac{\partial (f \circ x)}{\partial x_i}(q)$, onde $q = x^{-1}(p) \in U \subset \mathbb{R}^n$. Algumas vezes escreveremos $(\frac{\partial}{\partial x_i})_p f$ ou, simplesmente, $(\partial_i)_p f$ no lugar de $\frac{\partial f(p)}{\partial x_i}$.

Definição 2.3. *Sejam S^n uma variedade diferenciável e $\alpha : (-\varepsilon, \varepsilon) \to S$ uma curva diferenciável tal que $\alpha(0) = p \in S$. O vetor tangente à curva α em $t = 0$ é o operador $\alpha'(0) : D(S) \to \mathbb{R}$ dado por*

$$\alpha'(0)f = \frac{d(f \circ \alpha)}{dt}|_{t=0}.$$

Um vetor tangente em p é o vetor em $t = 0$ de alguma curva $\alpha : (-\varepsilon, \varepsilon) \to S$ com $\alpha(0) = p \in S$. O conjunto dos vetores tangentes a S em p será indicado por T_pS.

Sendo $x : U \to S^n$ uma parametrização em $p = x(q)$, podemos exprimir a função f e a curva α nesta parametrização por $(f \circ x)(q) = (f \circ x)(x_1, \ldots, x_n)$, com $q = (x_1, \ldots, x_n) \in U$, e $(x^{-1} \circ \alpha)(t) = (x_1(t), \ldots, x_n(t))$, respectivamente. Portanto, restrigindo f a α,

$$\begin{aligned}
\alpha'(0)f &= \frac{d(f \circ \alpha)}{dt}|_{t=0} = \frac{d(f \circ\ x)(x^{-1} \circ \alpha(t))}{dt}|_{t=0} \\
&= \frac{d}{dt} f \circ\ x(x_1(t), \ldots, x_n(t))|_{t=0} \\
&= \sum_i x_i{}'(0) \frac{\partial}{\partial x_i}(f \circ x)(q) = \sum_i x_i{}'(0) \left(\frac{\partial}{\partial x_i}\right)_p f.
\end{aligned}$$

Em outras palavras, o vetor $\alpha'(0)$ pode ser expresso na parametrização x por

$$\alpha'(0) = \sum_i x_i{}'(0) \left(\frac{\partial}{\partial x_i}\right)_p. \tag{2.2}$$

A equação (2.2) mostra que o vetor tangente a uma curva α em p depende apenas das derivadas de α em um sistema de coordenadas. Decorre também de (2.2) que o conjunto T_pS, com as operações usuais de funções, forma um espaço vetorial de dimensão n, e que a escolha de uma parametrização $x : U \to S$ determina uma base *associada* $B_p = \left\{ \left(\frac{\partial}{\partial x_1}\right)_p, \ldots, \left(\frac{\partial}{\partial x_n}\right)_p \right\}$ em T_pS. O espaço vetorial T_pS é chamado o *espaço tangente* de S em p.
Com a noção de espaço tangente vamos estender a variedades diferenciáveis a noção de diferencial de uma aplicação difenciável.

Proposição 2.1. *Sejam S_1^n e S_2^m variedades diferenciáveis e seja $\varphi : S_1 \to S_2$ uma aplicação diferenciável. Para cada $p \in S_1$ e cada $v \in T_pS_1$, escolha uma curva diferenciável $\alpha : (-\varepsilon, \varepsilon) \to S$ com $\alpha(0) = p \in S$, $\alpha'(0) = v$. Faça $\beta = \varphi \circ \alpha$. A aplicação $d\varphi_p : T_pS_1 \to T_{\varphi(p)}S_2$ dada por $d\varphi_p(v) = \beta'(0)$ é uma aplicação linear que não depende da escolha de α.*

Demonstração. Sejam $x: U \to S_1$ e $y: V \to S_2$ parametrizações em p e $\varphi(p)$, respectivamente. Exprimindo φ nestas parametrização, temos

$$y^{-1} \circ \varphi \circ x(q) = (y_1(x_1, \ldots, x_n), \ldots, y_m(x_1, \ldots, x_n))$$
$$q = (x_1, \ldots, x_n) \in U, \ \ (y_1, \ldots, y_n) \in V.$$

Fazendo o mesmo para α, temos

$$x^{-1} \circ \alpha(t) = (x_1(t), \ldots, x_n(t)).$$

Assim,

$$y^{-1} \circ \beta(t) = (y_1(x_1(t), \ldots, x_n(t)), \ldots, y_m(x_1(t), \ldots, x_n(t))).$$

Decorre daí que a expressão de $\beta'(0)$ na base $\left\{\left(\frac{\partial}{\partial y_i}\right)\right\}$ de $T_{\varphi(p)}S_2$, associada à parametrização y, é dada por

$$\beta'(0) = \left(\sum_{i=1}^{n} \frac{\partial y_1}{\partial x_i} {x_i}'(0), \ldots, \sum_{i=1}^{n} \frac{\partial y_m}{\partial x_i} {x_i}'(0)\right). \tag{2.3}$$

A equação (2.3) mostra que $\beta'(0)$ não depende da escolha de α. Além disso, (2.3) pode ser escrita como

$$\beta'(0) = d\varphi_p(v) = \left(\frac{\partial y_i}{\partial x_j}\right) {x_j}'(0), \quad i = 1, \ldots, m\,; \ \ j = 1, \ldots, n$$

onde $\left(\frac{\partial y_i}{\partial x_j}\right)$ denota uma matriz $m \times n$ e ${x_j}'(0)$ denota uma matriz coluna com n elementos. Portanto, $d\varphi_p(v)$ é uma aplicação linea de T_pS_1 em $T_{\varphi(p)}S_2$ cuja a matriz nas bases associadas às parametrizações x e y é exatamente a matriz $\left(\frac{\partial y_i}{\partial x_j}\right)$. □

Definição 2.4. *Sejam M^m e S^n variedades diferenciáveis. Uma aplicação diferenciável $\varphi: M \to S$ é uma* imersão *se $d\varphi_p: T_pM \to T_{\varphi(p)}S$ é injetiva para todo $p \in M$. Se, além disso, φ é um homeomorfismo sobre $\varphi(M) \subset S$, onde $\varphi(M)$ tem a topologia induzida por S, dizemos que φ é um* mergulho*. Se $M \subset S$ e a inclusão $i: M \hookrightarrow S$ é um imersão, diz-se que M é uma* subvariedade imersa *de S, quando $i: M \hookrightarrow S$ é um mergulho, diz-se que M é uma subvariedade S.*

Definição 2.5. *Um* campo de vetores *X em uma variedade diferenciável S é uma correspondência que a cada ponto $p \in S$ associa um vetor $X(p) \in T_pS$.*

Tomando uma parametrização $x: U \subset \mathbb{R}^n \to S$ é possível escrever

$$X(p) = \sum_i a_i(p) \left(\frac{\partial}{\partial x_i}\right)_p, \tag{2.4}$$

onde a_i são funções de S em $\mathbb{R}$ e $\left\{\left(\frac{\partial}{\partial x_i}\right)_p\right\}$ é a base associada a x, $i = 1, \ldots, n$. Dizemos que X é diferenciável quando as funções a_i são diferenciáveis. Sendo $X = \sum_i a_i \left(\frac{\partial}{\partial x_i}\right)$ e $Y = \sum_i b_j \left(\frac{\partial}{\partial x_j}\right)_p$ campos de vetores diferenciáveis ao longo de S, definimos o campo de vetores $[X, Y]$ por

$$[X, Y] = \sum_{i,j} \left(a_i \frac{\partial b_j}{\partial x_i} - b_j \frac{\partial a_j}{\partial x_i}\right) \frac{\partial}{\partial x_j},$$

tal campo é chamado de campo *colchete*. Claramente o campo colchete é diferenciável. Note que $[\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}] = 0$ para todo $i, j = 1, \ldots, n$. Também pode ser provado que o campo colchete é escrito da seguinte forma $[X,Y] = XY - YX$. A operação colchete possui as seguintes propriedade:

Proposição 2.2. *Se X, Y e Z são campos diferenciáveis em S, a, b são números reais, e f, g são funções diferenciáveis, então*

1. $[X,Y] = -[Y,X]$ *(anticomutativa),*
2. $[aX + bY, Z] = a[X,Z] + b[Y,Z]$ *(linearidade),*
3. $[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0$ *(identidade de Jacobi)*
4. $[fX, gY] = fg[X,Y] + fX(g)Y - gY(f)X$

A prova desta prosição é dada apenas usando a definição do colchete e iremos omiti-la. $[ref..]$. Também temos a noção de campos de vetores ao longo de uma curva.

Definição 2.6. *Um campo de vetores V ao longo de uma curva $c : I \to S$ é uma aplicação que a cada $t \in I$ associa um vetor tangente $V(t) \in T_{c(t)}S$. Diz-se que V é diferenciável se para toda função diferenciável f em S, a função $t \to V(t)f$ é diferenciável em I.*

O campo de vetores $dc\left(\frac{d}{dt}\right)$, indicado por $\frac{dc}{dt}$, é chamado campo *campo velocidade* de c. Observe que um campo de vetores ao longo de uma curva não necessariamente pode ser estendido a um campo de vetores definido em um aberto de S.
Finalizando essa seção, observemos que dados $[x^i]$ e $[y^i]$ dois sistemas de coordenadas para S e $f \in D(S)$ podemos escrever

$$f(p) = f \circ x(q) = f \circ x(x^1(y^1, \ldots, y^n), \ldots, x^n(y^1, \ldots, y^n))$$

usando a regra da cadeia obtemos que

$$\frac{\partial f}{\partial y^j} = \sum_{i=1}^{n} \frac{\partial x^i}{\partial y^j} \frac{\partial f}{\partial x^i}.$$

Olhando a ultima expressão como campo de vetores, temos

$$\frac{\partial}{\partial y^j} = \sum_{i=1}^{n} \frac{\partial x^i}{\partial y^j} \frac{\partial}{\partial x^i}. \tag{2.5}$$

2.2 Métricas Riemannianas

Definição 2.7. *Uma métrica Riemanniana em S é uma aplicacão que associa a cada ponto $p \in S$ um produto interno (isto é, uma forma bilinear simétrica, positiva definida)*

$$g_p = \langle \cdot, \cdot \rangle_p$$

no espaço tangente T_pS que varia diferenciavelmente com p no sentido de que se $x : U \to V$ é uma carta para uma vizinhança coordenada V de S e $B_p = \left\{ \left(\frac{\partial}{\partial x_1}\right)_p, \dots, \left(\frac{\partial}{\partial x_n}\right)_p \right\}$ é a base de T_pS associada a esta carta para cada $p \in V$, então as funções $g_{ij} : V \to R$

$$g_{ij} = \left\langle \left(\frac{\partial}{\partial x_i}\right)_p, \left(\frac{\partial}{\partial x_j}\right)_p \right\rangle_p$$

são diferenciáveis.

É comum deixar de indicar o índice p em $\langle \cdot, \cdot \rangle_p$ sempre que não houver confusão. As funções g_{ij} são chamadas *expressão da métrica Riemanniana* (ou "os g_{ij} da métrica") *no sistema de coordenadas* $x : U \subset \mathbb{R}^n \to S$. Uma variedade diferenciável com uma métrica Riemanniana chama-se uma *variedade Riemanniana*.

Definição 2.8. *Sejam M, S variedades Riemannianas. Um difeomorfismo $f : M \to S$ (isto é, f é uma bijeção diferenciável com inversa diferenciável) é chamado uma isometria se:*

$$\langle u, v \rangle_p = \langle df_p(u), df_p(v) \rangle_{f(p)} \text{ ,para todo } p \in M, u, v \in T_pM \tag{2.6}$$

Definição 2.9. *Sejam M, S variedades Riemannianas. Uma função diferenciável $f : M \to S$ é uma isometria local* em $p \in M$ *se existe uma vizinhança $U \subset M$ de p tal que $f : U \to f(U)$ é um difeomorfismo satisfazendo* (2.6). *Se para todo $p \in M$ f é uma isometria local, dizemos M é localmente isometrica a S.*

Exemplo 2.1. *Métrica Euclidiana. A variedade Riemanniana mais simples é o R^n com a métrica euclidiana $g_{ij} = \langle e_i, e_j \rangle = \delta_{ij}$*

Exemplo 2.2. *Imersões isométricas. Seja $f : M^n \to S^{n+k}$ uma imersão. Se S tem uma estrutura Riemanniana então f induz uma estrutura Riemanniana em M por*

$$\langle u, v \rangle_p := \langle df_p(u), df_p(v) \rangle_{f(p)}, \quad u, v \in T_pM.$$

Como a diferencial df_p é injetiva, $\langle , \rangle_p$ é positivo definido. As demais condições da definição (2.7) *são de simples verificação. A métrica de M é chamada então de métrica induzida, e f, munida das métricas acima definidas, é uma chamada de imersão isométrica.*

Exemplo 2.3. *Espaço Hiperbólico. Considere semi-espaço superior de R^n*

$$H^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \ | x_n > 0\}$$

e introduza em H^n a métrica

$$g_{ij}(x_1, \dots, x_n) = \frac{\delta_{ij}}{x_n^2}.$$

H^n é chamado o o espaço hiperbólico de dimensão *n*.

2.3 Conexões Afins

Denotaremos por $\tau(S)$ o conjunto dos campos de vetores de classe C^∞ em S e por F(S) o anel das função reais de classe C^∞ definidas em S.

Definição 2.10. *Uma conexão afim ∇ em uma variedade deiferenciável S é uma aplicação*

$$\nabla : \tau(S) \times \tau(S) \to \tau(S)$$

denotada por $(X,Y) \overset{\nabla}{\to} \nabla_X Y$ e que satisfaz as seguintes propriedades:

(a) $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z,$

(b) $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z,$

(c) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$

Onde $X,Y,Z \in \tau(S)$ e $f,g \in F(S)$.

Algumas vezes vamos escrever apenas conexão no lugar de conexão afim.

Proposição 2.3. *Seja ∇ uma conexão em uma variedade diferenciável S. Se $X,Y \in \tau(S)$ são campos vetoriais que se expressam em coordenadas locais por*

$$X = \sum_i x_i \partial_i \text{ e } Y = \sum_j y_j \partial_j, \quad \text{onde } \partial_i = \frac{\partial}{\partial x_i}$$

então

$$\nabla_X Y = \sum_{i,j=1}^n x_i y_j \nabla_{\partial_i}\partial_j + \sum_{j=1}^n X(y_j)\partial_j. \tag{2.7}$$

Em particular, $(\nabla_X Y)_p$ depende apenas do valor de X em p e do valor de Y ao longo de uma curva tangente a X_p.

Demonstração. Usando as propriedades de uma conexão, temos

$$\begin{aligned}\nabla_X Y &= \nabla_X\left(\sum_{j=1}^n y_j\partial_j\right) = \sum_{j=1}^n \nabla_X(y_j\partial_j) = \sum_{j=1}^n y_j\nabla_X\partial_j + \sum_{j=1}^n X(y_j)\partial_j \\ &= \sum_{j=1}^n y_j \nabla_{(\sum_{i=1}^n x_i\partial_i)}\partial_j + \sum_{j=1}^n X(y_j)\partial_j = \sum_{i,j=1}^n x_i y_j \nabla_{\partial_i}\partial_j + \sum_{j=1}^n X(y_j)\partial_j.\end{aligned}$$

Em particular

$$(\nabla_X Y)_p = \sum_{i,j=1}^n x_i(p)y_j(p)(\nabla_{\partial_i}\partial_j)_p + \sum_{j=1}^n X(y_j(p))(\partial_j)_p.$$

Os coeficientes $x_1(p),\ldots,x_n(p)$ dependem apenas do valor de $X(p)$ e dos valores de Y ao longo de uma curva passando por p cujo vetor tangente em p é X_p. $\square$

Da equação (2.7), escrevendo os campos vetoriais $\nabla_{\partial_i}\partial_j$ em termos dos campos ∂_k na forma

$$\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij}^k \partial_k,$$

obtemos a seguinte expressão local para o campo $\nabla_X Y$:

$$\nabla_X Y = \sum_{k=1}^{n} \left(\sum_{i,j=1}^{n} x_i y_j \Gamma_{ij}^k + X(y_k) \right) \partial_k. \tag{2.8}$$

Quando S possui um parametrização global $[x_i]$ ($\partial_i = \frac{\partial}{\partial x_i}$), então n^3 funções arbitrariais Γ_{ij}^k diferenciáveis de S em $\mathbb{R}$ definem uma conexão em S. Para ver isto, basta definir a conexão como em (2.8).

Definição 2.11. *As funções suaves* Γ_{ij}^k *definidas pela expressão* (2.8) *são chamadas os símbolos de Christoffel associados à carta particular utilizada*

Proposição 2.4. *Seja S uma variedade diferenciável com uma conexão ∇. Então existe uma única correspondência que associa um campo vetotial V ao longo a curva diferenciável $c : I \to S$ um outro campo vetorial $\frac{DV}{dt}$ ao longo de c, denominado* derivada covariante *de V ao longo de c, tal que:*

1. $\frac{D}{dt}(V+W) = \frac{DV}{dt} + \frac{DW}{dt}$.
2. $\frac{D}{dt}(fV) = \frac{df}{dt}V + f\frac{DV}{dt}$, *onde V é um campo de vetores ao longo de c e f é uma função diferenciável em I.*
3. *Se V é induzido por um campo de vetores $Y \in \tau(S,)$, isto é, $V(t) = Y(c(t))$, então $\frac{DV}{dt} = \nabla_{\frac{dc}{dt}} Y$.*

Demonstração. Suponha que existe uma corresponência satisfazendo $(1),(2)$ e (3). Sejam $x : U \subset \mathbb{R}^n \to S$ um sistema de coordenadas com $c(I) \cap x(U) \neq \emptyset$ e $(x_1(t),\ldots,x_n(t))$ a expressão local de $c(t), t \in I$. Seja $\partial_i = \frac{\partial}{\partial x_i}$, Então podemos expressar o campo V localmente como $V = \sum_j v^j \partial_j$, onde $v^j = v^j(t)$ e $\partial_j = (\partial_j)_{c(t)}$. Por (1) e (2), temos que

$$\frac{DV}{dt} = \sum_j \frac{dv^j}{dt}\partial_j + \sum_j v^j \frac{D\partial_j}{dt}.$$

Por (3) e por (a) da definição (2.10),

$$\frac{D\partial_j}{dt} = \nabla_{\frac{dc}{dt}}\partial_j = \nabla_{\left(\sum_i \frac{dx_i}{dt}\partial_i\right)}\partial_j = \sum_i \frac{dx_i}{dt}\nabla_{\partial_i}\partial_j.$$

Assim,

$$\frac{DV}{dt} = \sum_{j=1}^{n} \frac{dv^j}{dt}\partial_j + \sum_{i,j=1}^{n} \frac{dx_i}{dt} v^j \nabla_{\partial_i}\partial_j. \tag{2.9}$$

A expressão (2.9) mostra que se existe uma correspondência satisfazendo as condições da proposição (2.4), então tal correspondencia é única.

Para determinar a existência de $\frac{DV}{dt}$, dado um sistema de coordenadas (x,U) para uma vizinhança de $c(t)$, defina o campo $\frac{DV}{dt}$ em $x(U)$ pela expressão (2.9). É imediato verificar que um campo definido desta forma satisfaz todas as propriedades do enunciado. □

Ainda na equação (2.9), escrevendo $\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij}^k \partial_k$, e trocando j por k obtemos que

$$\frac{DV}{dt} = \sum_{k=1}^{n} \left\{ \frac{dv^k}{dt} + \sum_{i,j=1}^{n} v^j \frac{dx_i}{dt} \Gamma_{ij}^k \right\} \partial_k.$$

Definição 2.12. *Seja S uma variedade diferenciável com uma conexão ∇. Um campo vetorial V ao longo de uma curva $c : I \to S$ é chamado paralelo (com respeito a conexão ∇) quando $\frac{DV}{dt} = 0$, para todo $t \in I$.*

Proposição 2.5. *Seja S uma variedade diferenciável com uma conexão ∇. Seja $c : I \to S$ uma curva diferenciável em S e v_0 um vetor tangente a S em $c(t_0), t_0 \in I$ (isto é, $v_0 \in T_{(c(t_0))}S$). Então existe um único campo de vetores paralelo $V_{v_o,c}$ ao longo de c, tal que $V_{v_o,c}(t_0) = V_0$, (O campo $V_{v_o,c}$ é chamado o transporte paralelo de $V(t_0)$ ao longo de c)*

Demostração. Ver [5] □

O transporte paralelo $\prod : T_{c(t_0)}S \to T_{c(t)}S$ definida por $\prod(v) = V_{v_o,c}(t)$ é uma aplicação linear. De fato, sejam $v,u \in T_{c(t_0)}S$ e $a,b \in \mathbb{R}$. Considere o campo $W(t) = aV_{v,c}(t) + bV_{u,c}(t)$, note que $\frac{DW}{dt} = a\frac{D}{dt}V_{v,c} + b\frac{D}{dt}V_{u,c} = 0$, isto é, W é um campo paralelo ao longo de c, além disso $W(t_o) = av + bu$. Pela unicidade do campo paralelo temos que $\prod(av+bu) = aV_{v,c}(t) + bV_{u,c}(t) = a\prod(v) + b\prod(u)$

2.4 Conexão Riemanniana

Proposição 2.6. *Seja S uma variedade Riemanniana com uma conexão ∇. Então as seguintes afirmações são equivalentes:*

1. *Para todos os campos paralelos V e W ao longo de qualquer curva diferenciável c em S vale*
$$\langle V,W \rangle \equiv constante.$$

2. *Para todos os campos vetoriais V e W ao longo de qualquer curva diferenciável c em S vale*
$$\frac{d}{dt}\langle V,W \rangle = \left\langle \frac{DV}{dt}, W \right\rangle + \left\langle V, \frac{DW}{dt} \right\rangle.$$

3. *Para todos os campos vetoriais X, Y e Z vale*
$$X\langle Y,Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle$$

Demonstração. Ver [4] □

Definição 2.13. *Seja S uma variedade Riemanniana com uma conexão ∇. Dizemos que a conexão ∇ é compatível com a métrica de S, quando ela satisfaz qualquer uma das condições da proposição anterior.*

Definição 2.14. *Uma Conexão ∇ em uma variedade diferenciável S é dita simétrica quando*

$$\nabla_X Y - \nabla_Y X = [X,Y] \quad \textit{para todo } X,Y \in \tau(S). \tag{2.10}$$

Note que em um sistema de coordenadas (U,x), o fato da conexão ∇ ser simétrica é equivalente a qualquer umas das afirmaçãoes abaixo:

(a) $\nabla_{\partial_i}\partial_j = \nabla_{\partial_j}\partial_i$ para todo $i,j = 1,\dots,n$

(b) $\Gamma_{ij}^k = \Gamma_{ji}^k$

A equivalência $(a) \Leftrightarrow (b)$ é imediata. Supondo ∇ uma conexão simétrica, fazendo na equação (2.10) $X = \partial_i$ e $Y = \partial_j$ e lembrando que $[\partial_i, \partial_j] = 0$ obtemos (a). Reciprocamente, se vale (a), dados $X = \sum_i x_i\partial_i$ e $Y = \sum_j y_j\partial_j$ dois campos de vetores sobre S. Pelas expressão (2.7) temos

$$\nabla_X Y = \sum_{i,j=1}^{n} x_i y_j \nabla_{\partial_i}\partial_j + \sum_{j=1}^{n} X(y_j)\partial_j$$

e

$$\nabla_Y X = \sum_{i,j=1}^{n} x_i y_j \nabla_{\partial_i}\partial_j + \sum_{j=1}^{n} Y(x_j)\partial_j.$$

Assim

$$\nabla_X Y - \nabla_Y X = \sum_{j=1}^{n} \left(X(y_j) - Y(x_j)\right)\partial_j = \sum_{i,j=1}^{n} \left(x_i\partial_i(y_j) - y_i\partial_i(x_j)\right)\partial_j = [X,Y]$$

Lema 2.7 (Fórmula de Koszul)**.** *Seja S uma variedade Riemanniana munida com uma conexão ∇ simétrica e compatível com a métrica de S. Então para todos os campos X,Y e $Z \in \tau(S)$ vale que:*

$$\langle \nabla_Y X, Z\rangle = \frac{1}{2}(X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle - \langle [X,Z],Y\rangle - \langle [Y,Z],X\rangle - \langle [X,Y],Z\rangle). \tag{2.11}$$

Em particular, uma conexão simétrica compatível com a métrica é unicamente determinada pela métrica.

Demonstração. Sejam X,Y e $Z \in \tau(S)$, pelo item (3) da proposição (2.6) temos que:

(i) $X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$,

(ii) $Y\langle Z,X\rangle = \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle$,

(iii) $Z\langle X,Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla_Z Y\rangle$

Somando (i) e (ii), daí, subtraindo (iii), teremos, pela simetria de ∇, que

$$
\begin{aligned}
X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle &= \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle + \langle \nabla_Y Z, X\rangle + \langle Z, \nabla_Y X\rangle \\
&\quad - \langle \nabla_Z X, Y\rangle - \langle X, \nabla_Z Y\rangle \\
&= \langle X, \nabla_Y Z - \nabla_Z Y\rangle + \langle Y, \nabla_X Z - \nabla_Z X\rangle + \langle \nabla_X Y, Z\rangle + \langle Z, \nabla_Y X\rangle \\
&= \langle X, [Y,Z]\rangle + \langle Y, [X,Z]\rangle + \langle \nabla_X Y - \nabla_Y X, Z\rangle + 2\langle Z, \nabla_Y X\rangle \\
&= \langle X, [Y,Z]\rangle + \langle Y, [X,Z]\rangle + \langle [X,Y]\rangle + 2\langle Z, \nabla_Y X\rangle.
\end{aligned}
$$

Daí segue o resultado. □

Teorema 2.8. *(Levi-Civita). Seja (S,g) uma variedade Riemanniana. Então existe uma única conexão ∇ em S que é simétrica e compatível com a métrica g.*

Demonstração. Pelo lema anterior, se tal conexão existe ela é única. Para mostrar a existência, basta definir ∇ por (2.8) e verificar que ∇ está bem definida e que satisfaz as proprieadades desejadas. □

Definição 2.15. *Seja S uma variedade Riemanniana. A conexão dada pelo teorema* (2.8) *é denominada por conexão de Levi-Civita ou, simplesmente, conexão Riemanniana de S.*

Agora veremos como os símbolos de Christoffel de uma conexão Riemanniana podem ser calculados através dos componentes g_{ij} da métrica. Para isto, dado um sistema de coordenadas (U,x), fazendo na equação (2.8) $X = \partial_i, Y = \partial_J$ e $Z = \partial_k$ obtemos

$$\Gamma_{ij,k} = \sum_l \Gamma_{ij}^l g_{lk} = \langle \nabla_{\partial_i} \partial_j, \partial_k\rangle = \frac{1}{2}(\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) \tag{2.12}$$

onde $g_{ij} = \langle \partial_i, \partial_j\rangle$.

Como a matriz $[g_{ij}]$ admite um inversa $[g^{ij}]$, multiplicando a última expressão por $\sum\limits_k g^{km}$ obtemos que

$$\Gamma_{ij}^m = \frac{1}{2}\sum_k (\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij}) g^{km}. \tag{2.13}$$

2.5 Curvatura

Definição 2.16. *O tensor* curvatura *de uma variedade Riemanniana S munida com uma conexão ∇ é uma correspondência R que associa a cada par $X,Y \in \tau(S)$ uma aplicação $R(X,Y)$: $\tau(S) \to \tau(S)$ dada por*

$$R(X,Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z.$$

Proposição 2.9. *A curvatura R de uma variedade Riemanniana goza das seguintes propriedades:*

1. *R é bilinear em $\tau(S)\times\tau(S)$, isto é,*
   (a) $R(fX_1+gX_2,Y_1)=fR(X_1,Y_1)+gR(X_2,Y_1)$,
   (b) $R(X_1,fY_1+gY_2)=fR(X_1,Y_1)+gR(X_1,Y_2)$,

   $f,g\in F(S),\quad X_1,X_2,Y_1,Y_2\in\tau(S)$.

2. *Para todo par $X,Y\in\tau(S)$, o operador curvatura $R(X,Y):\tau(S)\to\tau(S)$ é linear, isto é,*
   (a) $R(X,Y)(Z+W)=R(X,Y)Z+R(X,Y)W$,
   (b) $R(X,Y)fZ=fR(X,Y)Z$,

   $f,g\in F(S),\quad Z,W\in\tau(S)$.

Demonstração. Vamos apenas usar a definição de conexão e as propriedades do campo conchete. A prova de (1a) e (1b) são analogas, portanto, verificaremos apenas (1a).

$$
\begin{aligned}
R(fX_1+gX_2,Y)Z &= \nabla_Y\nabla_{(fX_1+gX_2,Y)}Z-\nabla_{(fX_1+gX_2)}\nabla_YZ+\nabla_{[fX_1+gX_2,Y]}Z\\
&=\nabla_Y(f\nabla_{X_1}Z+g\nabla_{X_2}Z)-f\nabla_{X_1}\nabla_YZ-g\nabla_{X_2}\nabla_YZ+\nabla_{[fX_1,Y]}Z+\nabla_{[gX_2,Y]}Z\\
&=f\nabla_Y\nabla_{X_1}Z+Y(f)\nabla_{X_1}Z+g\nabla_Y\nabla_{X_2}Z+Y(g)\nabla_{X_2}Z-f\nabla_{X_1}\nabla_YZ\\
&-g\nabla_{X_2}\nabla_YZ+\nabla_{(f[X_1,Y]-Y(f)X_1)}Z+\nabla_{(g[X_2,Y]-Y(g)X_2)}Z\\
&=f\nabla_Y\nabla_{X_1}Z+Y(f)\nabla_{X_1}Z+g\nabla_Y\nabla_{X_2}Z+Y(g)\nabla_{X_2}Z-f\nabla_{X_1}\nabla_YZ\\
&-g\nabla_{X_2}\nabla_YZ+f\nabla_{[X_1,Y]}Z-Y(f)\nabla_{X_1}Z+g\nabla_{[X_2,Y]}Z-Y(g)\nabla_{X_2}Z\\
&=f(\nabla_Y\nabla_{X_1}Z-\nabla_{X_1}\nabla_YZ+\nabla_{[X_1,Y]}Z)\\
&+g(\nabla_Y\nabla_{X_2}Z-\nabla_{X_2}\nabla_YZ+\nabla_{[X_2,Y]}Z)\\
&=fR(X_1,Y)+gR(X_2,Y).
\end{aligned}
$$

O item (2a) é de imediata verificação, vejamos a prova de (2b)

$$
\begin{aligned}
R(X,Y)fZ &= \nabla_Y\nabla_XfZ-\nabla_X\nabla_YfZ+\nabla_{[X,Y]}fZ\\
&=\nabla_Y(f\nabla_XZ+X(f)Z)-\nabla_X(f\nabla_YZ+Y(f)Z)+f\nabla_{[X,Y]}Z+[X,Y](f)Z\\
&=f\nabla_Y\nabla_XZ+Y(f)\nabla_XZ+X(f)\nabla_YZ+Y(X(f))Z-f\nabla_X\nabla_YZ\\
&-X(f)\nabla_YZ-Y(f)\nabla_XZ-X(Y(f))Z+f\nabla_{[X,Y]}Z+[X,Y](f)\\
&=f(\nabla_Y\nabla_XZ-\nabla_X\nabla_YZ+\nabla_{[X,Y]}Z)+[X,Y](f)-[XY-YX]f\\
&=f(\nabla_Y\nabla_XZ-\nabla_X\nabla_YZ+\nabla_{[X,Y]}Z)\\
&=fR(X,Y)Z.
\end{aligned}
$$

□

Como $R(X,Y)Z\in\tau(S)$, dado $[x^i]$ um sistema s e coordenadas para S, vamos expressar a curvatura na base $\{(\partial_i)\}$. Escrevemos

$$R(\partial_i,\partial_j)\partial_k=\sum_l R^l_{ijk}\partial_l.$$

Se

$$X = \sum_i x^i \partial_i, \;\; Y = \sum_j y^j \partial_j \;\; Z = \sum_k z^k \partial_k,$$

obteremos pela linearidade de R,

$$R(X,Y)Z = \sum_{i,j,k,l} x^i y^j z^k R^l_{ijk} \partial_l.$$

Por outro lado, da definição de R, vem que

$$\begin{aligned}
R(\partial_i, \partial_j)\partial_k &= \nabla_{\partial_j} \nabla_{\partial_i} \partial_k - \nabla_{\partial_i} \nabla_{\partial_j} \partial_k \\
&= \nabla_{\partial_j} \left(\sum_s \Gamma^s_{ik} \partial_s \right) - \nabla_{\partial_i} \left(\sum_s \Gamma^s_{jk} \partial_s \right) \\
&= \sum_s \nabla_{\partial_j} (\Gamma^s_{ik} \partial_s) - \sum_s \nabla_{\partial_i} \left(\Gamma^s_{jk} \partial_s \right) \\
&= \sum_s \left(\Gamma^s_{ik} \nabla_{\partial_j} \partial_s + \partial_j (\Gamma^s_{ik}) \right) \partial_s - \sum_s \left(\Gamma^s_{jk} \nabla_{\partial_i} \partial_s + \partial_i (\Gamma^s_{jk}) \partial_s \right) \\
&= \sum_s \Gamma^s_{ik} \sum_l \Gamma^l_{js} \partial_l - \sum_s \Gamma^s_{jk} \sum_l \Gamma^l_{is} \partial_l + \sum_s \partial_j (\Gamma^s_{ik}) \partial_s - \sum_s \partial_i (\Gamma^s_{jk}) \partial_s
\end{aligned}$$

trocando o s por l nos últimos dois somatórios, obtemos

$$R(\partial_i, \partial_j)\partial_k = \sum_l \left(\sum_s \Gamma^s_{ik} \Gamma^l_{js} - \sum_s \Gamma^s_{jk} \Gamma^l_{is} + \partial_j (\Gamma^l_{ik}) - \partial_i (\Gamma^l_{jk}) \right) \partial_l$$

e portanto,

$$R^l_{ijk} = \sum_s \Gamma^s_{ik} \Gamma^l_{js} - \sum_s \Gamma^s_{jk} \Gamma^l_{is} + \partial_j (\Gamma^l_{ik}) - \partial_i (\Gamma^l_{jk}). \tag{2.14}$$

Notação: Daqui em diante usaremos a notação de Einstem para somas. Ou seja, em vez de se usar o sinal de somatorio $\sum$ para denotar uma soma, convencionamos que sempre que uma expressão contém um índice como um superescrito e o mesmo índice como subescrito, uma soma é implícita sobre todos os valores que este índice pode tomar. Por exemplo, escrevemos:

$$X = x^i \partial_i, \quad \nabla_{\partial_i} \partial_j = \Gamma^k_{ij} \partial_k, \quad \frac{\partial}{\partial y^j} = \frac{\partial x^i}{\partial y^j} \frac{\partial f}{\partial x^i},$$

no lugar de

$$X = \sum_i x^i \partial_i, \quad \nabla_{\partial_i} \partial_j = \sum_k \Gamma^k_{ij} \partial_k, \quad \frac{\partial}{\partial y^j} = \sum_{i=1}^n \frac{\partial x^i}{\partial y^j} \frac{\partial f}{\partial x^i},$$

respectivamente.

2.6 Variedades planas

No que se segue, vamos trabalhar com variedades diferenciáveis que possuam uma parametrização global.

Definição 2.17. *Sejam S uma variedade diferenciável munida com uma conexão ∇ e X um campo de vetores sobre S. Se para toda curva $c : I \to S$ tem-se que X é paralelo ao longo de c então dizemos que X é um campo paralelo sobre S*

Pela unicidade do campo paralelo (rever teorema (2.5)), se X é um campo paralelo ao longo de S, então o transporte paralelo de X_p ao longo de c $(\prod_c(X_p))$ não depende de c, além disso, vale que $(\prod_c(X_p)) = X_q$. Uma condição necessária e suficiente para que um campo de vetores X seja paralelo é que $\nabla_Y X = 0$ para todo $Y \in \tau(S)$.

Definição 2.18 (Coordenadas planas)**.** *Seja S uma variedade diferenciável munida com uma conexão ∇. Se existe um sistema de coordenadas global $[x^i]$ de S de modo que os campos de vetores $\partial_i = \frac{\partial}{\partial x^i}$ $(i = 1,\dots,n)$ são todos paralelos sobre S, então dizemos que S é plana com respeito a ∇ e que $[x^i]$ é um sistema de coordenadas ∇-afim para S.*

Evidentemente, a condição de que os campos ∂_i sejam todos paralelos é quivalente a qualquer uma das condições abaixo:

(a) $\nabla_{\partial_i}\partial_j = 0$,

(b) $\Gamma_{ij}^k = 0$,

(c) $0 = \Gamma_{ij,k} \overset{def}{=} \langle \nabla_{\partial_i}\partial_j, \partial_k\rangle = \Gamma_{ij}^l g_{lk}$.

Dado alguma conexão ∇, um sistema de coordenadas afim não existe em geral. Sejam $[x^i]$ um sistema de coordenadas afim e $[y_a]$ outro sistema de coordenadas para S (denotamos $\partial^a = \frac{\partial}{\partial y_a}$). Escrevendo ∂^a na base $\{(\partial_i)\}$, temos que $\partial^a = \sum\limits_i \frac{\partial x^i}{\partial y_a}\partial_i$. Daí e da expressão (2.8), uma condição necessária e suficiente para que $\nabla_X \partial^a = 0$ é que $X(\frac{\partial x^i}{\partial y_a}) = 0$, que é o mesmo que dizer que para todo $(i, a = 1,\dots,n)$ os temos $\frac{\partial x^i}{\partial y_a}$ são constantes. Isto é, $[y_a]$ é um sistema de coordenadas afim para S se e somente, se existe uma matriz $n \times n$ A e um vetor $B \in \mathbb{R}^n$ tal que $x = Ay + B$.

Se (S, ∇) admite coordenadas ∇-afins, a derivada covariante não depende da curva, em particular o transporte paralelo também não depende da curva. Note também que pela linearidade do transporte paralelo que um campo $X = x_i\partial_i$ é paralelo se e somente se, os componentes x_i são constantes. Denotando por $[x^i]$ o sistema de coordenadas ∇-afim para S, como $\Gamma_{ij}^k = \Gamma_{ij}^k = 0$ segue que a conexão ∇ é simétrica e, por (2.14), que $R = 0$. Reciprocamente temos o seguinte resultado.

Teorema 2.10. *Seja (S, g, ∇) uma variedade Riemanniana munida com uma conexão afim simétrica ∇ com a propriedade que o tensor curvatura R definido por essa conexão seja identicamente nulo. Então, S é localmente ∇-plana, no seguinte sentido: para todo $p \in S$ existe um aberto $U \subset S$ contendo p que admite coordenadas ∇-afins.*

Demonstração. Ver[4], pag 107.

2.7 Conexões Duais

No próximo capítulo veremos modelos estatísticos como variedades diferenciavéis. Surgirão naturalmente conexões afins sobre essas variedades, que em geral são conexões não-Riemannianas. No entanto, juntamente com a métrica Riemanniana, tais conexões admitem conexões duais, que desempenham um papel fundamental no estudo desses modelos. Mesmo do ponto de vista puramente matemático, a noção de conexões duais é por si só interessante. Nesta seção, estudamos este importante conceito de dualidade.

Sejam S uma variedade diferenciável, $g = \langle \cdot, \cdot \rangle$ uma métrica Riemanniana sobre S e ∇ uma conexão afim sobre S. A conexão dual ∇^* de (S, g, ∇) é definida por

$$Z\langle X,Y\rangle = \langle \nabla_Z X, Y\rangle + \langle X, \nabla^*_Z Y\rangle, \tag{2.15}$$

para todo campo de vetores $X,Y,Z \in \tau(S)$. Verifica-se facilmente que ∇^* define, de fato, uma conexão afim sobre S. Dizemos que ∇ e ∇^* são conexões duais ou conexões conjugadas com respeito a métrica g. Note que da equação acima, dadas uma métrica g e uma conexão ∇ sobre S, existe uma única conexão dual ∇^* de ∇. A consequência disso é que $(\nabla^*)^* = \nabla$. Vamos denotar este trio (g, ∇, ∇^*) por estrutura dual sobre S. Em termos de coordenadas, podemos definir ∇^* como a conexão afim que satisfaz:

$$\partial_k g_{ij} = \Gamma_{ki,j} + \Gamma^*_{kj,i}. \tag{2.16}$$

Vimos que uma conexão é plana se existe um sistema de coordenadas afins $[\theta^i]$, ou seja, $\nabla_{\partial_i}\partial_j = 0$, para todo i, j.

Definição 2.19. *Sejam S uma variedade diferenciável e (g, ∇, ∇^*) uma estrutura dual sobre S. Se (S, ∇) e (S, ∇^*) são planas diremos que S é dualmente plana.*

Provaremos a seguir que se uma conexão ∇ possui tensor curvatura $R = 0$ e ∇ e ∇^* são ambas conexões simétricas então a estrutura (g, ∇, ∇^*) é dualmente plana e, além disso, existem sistemas de coordenadas afins $[\theta^i]$ e $[\eta^j]$ de ∇ e ∇^*, respectivamente, relacionando-se com a métrica g da seguinte forma:

$$g^{ij} = \frac{\partial \theta^i}{\partial \eta^j} \quad \text{e} \quad g_{ij} = \frac{\partial \eta^i}{\partial \theta^j}.$$

Primeiramente, verificamos o seguinte resultado:

Proposição 2.11. *Assuma que ∇ e ∇^* são conexões simétricas então a conexão $\tilde{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$ é Levi-Civita.*

Demonstração. É simples verificar que $\tilde{\nabla} = \frac{1}{2}(\nabla + \nabla^*)$ é uma conexão compatível com a métrica. Para isto, basta somar as duas equações abaixo

$$\frac{1}{2} Z\langle X,Y\rangle = \frac{1}{2}\left(\langle \nabla_Z X, Y\rangle + \langle X, \nabla^*_Z Y\rangle\right)$$

e

$$\frac{1}{2}Z\langle Y,X\rangle = \frac{1}{2}\left(\langle \nabla_Z Y, X\rangle + \langle Y, \nabla^*_Z X\rangle\right),$$

as quais são válidas pela equação (2.15). Agora, assuma que ∇ e ∇^* são simétricas. Temos que

$$\begin{aligned}\tilde{\nabla}_X Y - \tilde{\nabla}_Y X &= \frac{1}{2}(\nabla_X Y + \nabla^*_X Y) - \frac{1}{2}(\nabla_Y X + \nabla^*_Y X)\\ &= \frac{1}{2}(\nabla_X Y - \nabla_Y X) + \frac{1}{2}(\nabla^*_X Y - \nabla^*_Y X)\\ &= \frac{1}{2}[X,Y] + \frac{1}{2}[X,Y].\end{aligned}$$

Assim, segue-se que $\tilde{\nabla}$ é simétrica. Conclui-se então que $\tilde{\nabla}$ é uma conexão Riemanniana.

□

Uma propriedade importante que vale apenas para conexões Riemannianas é a seguinte: Dados campos paralelos $X(t)$ e $Y(t)$ ao longo de uma curva diferenciável em S, o produto $\langle X(t),Y(t)\rangle$ é constante em t. Isso falha em conexões não-Riemannianas. No entanto, o seguinte vale: Sejam $\gamma : I \to S$ uma curva diferenciável em S, X e Y campos de vetores ao longo de γ e $\frac{DX}{dt}$ e $\frac{D^*Y}{dt}$, respectivamente, as derivadas covariantes de X com respeito a ∇ e de Y com respeito a ∇^*. Usando a equação (2.15) temos que

$$\frac{d}{dt}\langle X(t),Y(t)\rangle = \langle \frac{DX}{dt}, Y(t)\rangle + \langle X(t), \frac{D^*Y}{dt}\rangle.$$

Assim, se X é paralelo com respeito a ∇ e Y é paralelo com respeito a ∇^*, segue dá última equação que $\langle X(t),Y(t)\rangle$ é constante. Em particular temos o seguinte resultado:

Proposição 2.12. *Sejam $\prod_\gamma$, $\prod^*_\gamma : T_p(S) \to T_q(S)$, onde p e q são pontos de γ e $\prod_\gamma$ e $\prod^*_\gamma$ denotam, respectivamente, os transportes paralelos ao longo de γ com respeito a ∇ e ∇^*, então para todo $X,Y \in T_pS$ vale que*

$$\left\langle \prod(X), \prod^*(Y)\right\rangle_q = \langle X,Y\rangle_p.$$

A expressão acima nos diz que se conhecermos o transporte paralelo $\prod$ com respeito a ∇, então também conheceremos o transporte paralelo $\prod^*$ com respeito a ∇^*. Em particular, se $\prod$ não depende da curva, o mesmo vale para $\prod^*$. No caso em que (S,∇) é plana obtemos que $\prod$ e $\prod^*$ independem da curva.

Proposição 2.13. *Denotando R e R^* respectivamente a curvatura com respeito ∇ e ∇^*, então para todo campo de vetores $X,Y,Z,W \in T(S)$ temos que $\langle R^*(X,Y)Z,W\rangle = -\langle R(X,Y)W,Z\rangle$. Em particular $R = 0$ se, e somente se, $R^* = 0$.*

Demonstração.

$$\begin{aligned}
\langle R^*(X,Y)Z,W\rangle &= \langle \nabla^*_Y\nabla^*_X Z,W\rangle - \langle \nabla^*_X\nabla^*_Y Z,W\rangle + \langle \nabla^*_{[X,Y]}Z,W\rangle \\
&= Y\langle \nabla^*_X Z,W\rangle - \langle \nabla^*_X Z,\nabla_Y W\rangle - (X\langle \nabla^*_Y Z,W\rangle - \langle \nabla^*_Y Z,\nabla_X W\rangle) + \langle \nabla^*_{[X,Y]}Z,W\rangle \\
&= YX\langle Z,W\rangle - Y\langle Z,\nabla_X W\rangle - \langle \nabla^*_X Z,\nabla_Y W\rangle - XY\langle Z,W\rangle + X\langle Z,\nabla_Y W\rangle \\
&\quad + \langle \nabla^*_Y Z,\nabla_X W\rangle + \langle \nabla^*_{[X,Y]}Z,W\rangle \\
&= -(Y\langle Z,\nabla_X W\rangle - \langle \nabla^*_Y Z,\nabla_X W\rangle) + (X\langle Z,\nabla_Y W\rangle - \langle \nabla^*_X Z,\nabla_Y W\rangle) \\
&\quad -([X,Y]\langle Z,W\rangle - \langle \nabla^*_{[X,Y]}Z,W\rangle) \\
&= -\langle \nabla_Y\nabla_X W,Z\rangle + \langle \nabla_X\nabla_Y W,Z\rangle - \langle \nabla_{[X,Y]}W,Z\rangle \\
&= -\langle R(X,Y)W,Z\rangle.
\end{aligned}$$

E, por fim, provamos o seguinte resultado:

Teorema 2.14. *Assuma o tensor curvatura R de (S,g,∇) seja identicamente nulo e que ∇ e ∇^* são ambas conexões simétricas. Então, (S,g,∇,∇^*) é dualmente plana. Além disso, localmente, valem as seguintes itens:*

(i) Existem sistemas de coordenadas afins $\theta=[\theta^i]$ e $\eta=[\eta_j]$ de ∇ e ∇^*, respectivamente, relacionando-se com a métrica g da seguinte forma:

$$g^{ij}=\frac{\partial\theta^i}{\partial\eta_j} \quad \text{e} \quad g_{ij}=\frac{\partial\eta_i}{\partial\theta^j}.$$

(ii) Existem funções convexas $\psi(\theta)$, $\varphi(\eta)$, que satisfazem a Transformada de Legendre:

$$\frac{\partial\psi(\theta)}{\partial\theta^i}=\eta_i \quad e \quad \frac{\partial\varphi(\eta)}{\partial\eta_i}=\theta^i.$$

Em particular, $\mathrm{Hess}_\psi = g_{ij}$ e $\mathrm{Hess}_\varphi = g^{ij}$.

Demonstração. Pela proposição (2.13), de $R=0$ tem-se $R^*=0$. Como ∇ é simétrica, segue do Teorema (2.10) que S é plana com respeito a conexão ∇^*. No entanto, o sistema de coordenadas ∇^*-afim que procuramos deve satisfazer os itens (i) e (ii).

Primeiramente, provaremos (i). Dado um sistema de coordenadas $[\theta^i]$ sobre $(S,g=\langle\cdot,\cdot\rangle,\nabla)$, consideremos o referencial dual $\{\partial^i=g^{ik}\partial_k\}$, onde $\{\partial_i=\frac{\partial}{\partial\theta^i}\}$ denota o referencial dos vetores coordenados e (g^{ij}) a matriz inversa de (g_{ij}). Em outras palavras, ∂^j é definido por $\left\langle\partial_i,\partial^j\right\rangle=\delta_i^j$, para todo i,j. Assim, temos o seguinte:

$$\begin{aligned}
\left\langle\nabla^*_{\partial^i}\partial^j,\partial_k\right\rangle &= \partial^i\left\langle\partial^j,\partial_k\right\rangle - \left\langle\partial^j,\nabla_{\partial^i}\partial_k\right\rangle = \partial^i(\delta_k^j) - \left\langle\partial^j,g^{il}\nabla_{\partial_l}\partial_k\right\rangle \\
&= -\left\langle\partial^j,g^{il}\Gamma^r_{lk}\partial_r\right\rangle = -g^{il}\Gamma^r_{lk}\left\langle\partial^j,\partial_r\right\rangle = -g^{il}\Gamma^r_{lk}\delta^j_r = -g^{il}\Gamma^j_{lk}. \qquad (2.17)
\end{aligned}$$

Assumindo que $[\theta^i]$ é um sistema de coordenadas afins em (S,g,∇) temos que $\Gamma^k_{ij}=0$, donde $\nabla^*_{\partial^i}\partial^j=0$, para todos i,j. Por hipotese, ∇^* simétrica, logo

$$[\partial^i,\partial^j]=\nabla^*_{\partial^i}\partial^j-\nabla^*_{\partial^j}\partial^i=0.$$

Assim, pelo Teorema de Frobenius, localmente, existe um sistema de coordenadas $[\eta_j]$ tal que $\partial^j = \frac{\partial}{\partial \eta_j}$. Da equação (2.17) e do fato que $[\theta^i]$ é ∇-afim, segue-se que $[\eta_j]$ é ∇^*-afim.

Como $\frac{\partial}{\partial \theta^i}$ e $\frac{\partial}{\partial \eta_j}$ são campos coordenados em S, temos da equação 2.5 (mudança de variável), que

$$g^{ij}\partial_i = \partial^j = \frac{\partial}{\partial \eta_j} = \sum_i \frac{\partial \theta^i}{\partial \eta_j}\frac{\partial}{\partial \theta^i} = \frac{\partial \theta^i}{\partial \eta_j}\partial_i$$
$$g_{ij}\partial^i = \partial_j = \frac{\partial}{\partial \theta^j} = \sum_i \frac{\partial \eta_i}{\partial \theta^j}\frac{\partial}{\partial \eta_i} = \frac{\partial \eta_i}{\partial \theta^j}\partial^i.$$

Assim, temos $\frac{\partial \theta^i}{\partial \eta_j} = g^{ij}$ e $\frac{\partial \eta_i}{\partial \theta^j} = g_{ij}$. Note ainda que a inversa da matriz $g_{ij}(\theta)$ é a matriz $g_{ij}(\eta) = \langle \partial^i, \partial^j \rangle = \delta_i^j$. □

O item (i) do Teorema (2.14) pode ser reescrito da seguinte maneira: Seja (g, ∇, ∇^*) uma estrutura dual para S, tal que as conexões ∇ e ∇^* são simétricas, então S é plana com respeito a ∇ se e somente se, S é plana com respeito a ∇^*. Além disso, se $[\theta^i]$ é um sistema de coordenadas ∇-afim para S, então os campos de vetores definidos por $\{\partial^i = g^{ik}\partial_k\}$ são campos coordenados de um sistema de coordenadas ∇^*-afim para S.

Agora, provaremos (ii). Considere as 1-formas $w = \sum_i \eta_i d\theta^i$ e $v = \sum_i \theta^i d\eta_i$. Usando que $\frac{\partial \eta_i}{\partial \theta^j} = g_{ij} = g_{ji}$, temos que a diferencial dw satisfaz:

$$\begin{aligned} dw &= \sum_i d\eta_i \wedge d\theta^i = \sum_{i,j} \frac{\partial \eta_i}{\partial \theta^j} d\theta^j \wedge d\theta^i = \sum_{i,j} g_{ij}\, d\theta^j \wedge d\theta^i \\ &= \sum_{i,j} g_{ji}\, d\theta^i \wedge d\theta^j = -\sum_{i,j} g_{ij}\, d\theta^j \wedge d\theta^i = -dw, \end{aligned}$$

donde $dw = 0$. De maneira análoga, $dv = 0$. Como 1-formas fechadas são localmente exatas, segue-se que, localmente, $w = d\psi$ e $v = d\varphi$, para funções ψ e φ, definidas sobre um aberto de $\mathbb{R}^n$. Tem-se que $\frac{\partial \psi}{\partial \theta^i} = \eta_i$ e $\frac{\partial \varphi}{\partial \eta_i} = \theta^i$; donde, por (i), segue-se que $\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} = \frac{\partial \eta_i}{\partial \theta^j} = g_{ij}$ e $\frac{\partial^2 \varphi}{\partial \eta_i \partial \eta_j} = \frac{\partial \theta^i}{\partial \eta_j} = g^{ij}$. Como (g_{ij}) e (g^{ij}) são matrizes positiva-definidas, segue-se que as funções ψ e φ são convexas. Item (ii) está provado. □

2.8 Subvariedades Autoparalelas

Definição 2.20. *Seja S uma variedade Riemanniana munida com uma conexão $\bar{\nabla}$. Uma subvariedade M de S é dita ser* autoparalela ou totalmente geodésica *quando para quaisquer dois campos $X, Y \in \tau(M)$ vale que $\bar{\nabla}_X Y \in \tau(M)$.*

Dada qualquer subvariedade M em S, pode-se definir naturalmente uma conexão sobre M, simplesmente projetando-se ortogonalmente a conexão de S sobre M, ou seja, $\nabla_X Y = (\bar{\nabla}_X Y)^{\top}$, com $X, Y \in \tau(M)$. Verifica-se facilmente ∇ define uma conexão sobre M. Além disso, define-se a *segunda forma fundamental* de M em (S, g, ∇) por

$$\alpha(X,Y) = (\bar{\nabla}_X Y)^{\perp},$$

onde $(\cdot)^{\perp}$ denota a projeção ortogonal sobre o fibrado normal de M em S. Assim, em outras palavras, M é autoparalela se, e somente se, a segunda forma fundamental α de M em S é identicamente nula. Um conjunto aberto $A \subset S$ é um exemplo trivial de subvariedade autoparalela. Como M herda a conexão de S, podemos naturalmente definir o transporte paralelo $\prod : T_{c(t_0)}M \to T_{c(t)}M$ sobre M ao longo de $c : [a,b] \to M$. Por unicidade, este transporte concide com o transporte $\prod : T_{c(t_0)}S \to T_{c(t)}S$ sobre S restrito ao espaço tangente de M. Isto é,

$$\prod_c M = \prod_c |_{T_{c(t_0)}M}. \tag{2.18}$$

Sejam M^m uma subvariedade de S^n , $[x^i]$ e $[y^a]$ sistemas de coordenadas de S e M, respectivamente, e $\partial_i = \frac{\partial}{\partial x^i}$, $\partial_a = \frac{\partial}{\partial y^a}$. sejam ∇ uma conexão sobre S e Γ^k_{ij} os símbolos de Christofel de ∇ com respeito a $[x^i]$. Sendo $X = x^a\partial_a$ e $Y = y^b\partial_b \in \tau(M)$, temos

$$\nabla_X Y = X(y^b)\partial_b + x^a y^b \nabla_{\partial_a}\partial_b. \tag{2.19}$$

Portanto, M é autoparalela se e somente, se $\nabla_{\partial_a}\partial_b \in \tau(M)$. Que por sua vez, é equivalente a existirem m^3 funções Γ^c_{ab} diferenciáveis de M em $\mathbb{R}$ de modo que

$$\nabla_{\partial_a}\partial_b = \Gamma^c_{ab}\partial_c. \tag{2.20}$$

Essas funções Γ^c_{ab} são os coeficientes de ∇ com respeito ao sistema de coordenadas $[y^a]$. Usando a identidade $\partial_a = \partial_a x^i \partial_i$, podemos reescrever (2.20) da seguinte forma

$$\Gamma^c_{ab}\partial_c x^k = (\partial_a x^i)(\partial_b x^j)\Gamma^k_{ij} + \partial_a\partial_b x^k. \tag{2.21}$$

Note que, se existem $E_1,\ldots,E_m$ $(m = dimM)$ campo de vetores linearmente independentes sobre M que são paralelos com respeito a conexão de S, então M é subvariedade autoparalela. De fato, para $X = x^a E_a$ e $Y = y^b E_b \in \tau(M)$ obtemos

$$\nabla_X Y = X(y^b)E_b + y^b \nabla_X E_b = X(y^b)E_b \in \tau(M).$$

Quando M é uma subvariedade autoparalela de S, a conexão de S é ainda uma conexão em M. Portanto, Se $R_S = 0$ (curvatura de S) então $R_M = 0$. Em particular se S é plana (com respeito ∇), tem-se que ∇ é simétrica e $R_M = 0$. Pelo teorema (2.10) M é localmente plana. Sem perda de generalidade, suponha que $[x^i]$ e $[y^a]$ na equação (2.21) são os sistemas de coordenadas afim de S e M, respectivamente. Portanto, se M é autoparalela $(\Gamma^k_{ij} = 0$ e $\Gamma^C_{ab} = 0)$ vale que $\partial_a\partial_b x^k = 0$ para todo $a,b = 1,\ldots,m$ e $k = 1,\ldots,n$, . Isto é, existe uma matriz $n \times m$ A e um vetor $B \in \mathbb{R}^n$ de modo que

$$x = Ay + B. \tag{2.22}$$

CAPÍTULO 3

A estrutura geométrica de modelos estatísticos

No presente capítulo, veremos certos modelos estatísticos como variedades Riemannianas. Com uso de tal estrutura geométrica, faremos a prova do famoso teorema de Cramér-Rao e em seguida daremos uma condição necessária e suficiente para que modelos estatísticos possuam estimadores eficientes.

3.1 Modelos Estatísticos

Definição 3.1. *Uma distribuição de probabilidade sobre um conjunto χ é uma função $p:\chi\to\mathbb{R}$ tal que*

(i) $p(x)\geq 0$, *para todo* $x\in\chi$;

(ii) $\sum_{x\in X} p(x)=1$, *quando χ é um conjunto enumerável e*

(iii) $\int p(x)dx=1$, *quando χ é um aberto de $\mathbb{R}^n$.*

Para n ≥ 2, o símbolo da integral em (iii) denota uma integral múltipla. Quando χ for um conjunto discreto, frequentemente escreveremos $\int p(x)dx$ para denotar $\sum p(x)$.

Definição 3.2. *Seja S uma família de distribuições de probabilidades sobre um conjunto χ. Suponha que cada ponto de S seja parametrizado por n valores reais $\xi=(\xi^1,\ldots,\xi^n)$ num subconjunto $\Theta\subset\mathbb{R}^n$, isto é,*

$$S=\left\{P_\xi=p(x;\xi)\mid\ \xi=(\xi^1,\ldots,\xi^n)\in\Theta\subset\mathbb{R}^n\right\}. \tag{3.1}$$

E, além disso, que a função $\xi:\Theta\to S$ seja injetiva. Chamamos S de um modelo estatístico *sobre χ, de dimensão n.*

Vamos denotar S como $S=\{p_\xi\}$ e por $p_\xi=p$ um ponto de S.

Definição 3.3. *Seja S um modelo estatístico sobre χ de dimensão n. Diremos que S é um modelo estatístico* regular *se satisfaz os itens abaixo:*

(1) $\Theta\subset\mathbb{R}^n$ *é aberto.*

(2) Fixado $x\in\chi$, as funções $\xi\in\Theta\mapsto p(x,\xi)$ são suaves.

(3) Podemos trocar livremente a ordem de integração e derivação, isto é,

$$\frac{\partial}{\partial \xi^i}\int p(x;\xi)dx = \int \frac{\partial p(x;\xi)}{\partial \xi^i}dx. \tag{3.2}$$

Para todo $p \in S$ e $i = (1,2,..,n)$.

(4) Para todo $p \in S$, o conjunto $Z_+ := \{x \in \chi \mid p(x;\xi) > 0\}$ independe de ξ.

(5) Para cada $p \in S$ as funções $\frac{\partial}{\partial \xi^i}p(x;\xi)$, com $i = 1,2,...,n$, são linearmente independentes (LI) como funções de x.

Segue-se, dos itens acima, a aplicação $\xi \in \Theta \mapsto p_\xi \in S$ define um sistema de coordenadas em S, fazendo de S uma variedade globalmente parametrizada, e o referencial $\{\partial_i = \frac{\partial p}{\partial \xi^i}\}$ são os campos coordenados desta parametrização. Abaixo faremos algumas outras obsevações sobre a Definição 3.3 acima.

1: Verificar a equação (3.2) é equivalente verificar que:

$$\int \frac{\partial p(x;\xi)}{\partial \xi^i}dx = 0. \tag{3.3}$$

Já que $\int p(x;\xi)dx = 1$.

2: Sendo S um modelo regular podemos redefinir $\chi = Z_+$. Dessa forma, tem-se $p(x) > 0$ para todo $x \in \chi$ e $\xi \in \Theta$.

3: A condição das funções $\frac{\partial}{\partial \xi^i}p(x;\xi)$, com $i = 1,2,...,n$, serem LI (como funções de x) é equivalente a condição das funções $\frac{\partial}{\partial \xi^i}\ln(p(x;\xi))$ com $i = 1,2,...,n$ serem LI.

Vejamos alguns exemplos:

Exemplo 3.1. *(Modelo Normal) Sendo $\chi = \mathbb{R}$, $\xi = (\mu,\sigma)$, $\Theta = \{(\mu,\sigma) \in \mathbb{R}^2 | \sigma > 0\}$ e $p_\xi = p(x;\mu,\sigma)$ dado por*

$$p_\xi = p(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.4}$$

Pode-se mostrar que $\int p(x;\xi)dx = 1$ para todo $\xi \in \Theta$. É claro que $p(x;\mu,\sigma) > 0$ para todo $x \in \chi$. Suponha que $p(x;\mu_1,\sigma_1)$=$p(x;\mu_2,\sigma_2)$ para todo $x \in \chi$. Temos que

$$\ln\sigma_2 - \ln\sigma_1 = \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}\text{ , para todo } x \in \chi$$

Escolhendo $x_0 = \mu_1$ e $x_1 = \mu_2$, obtemos:

$$\left(\frac{\mu_2-\mu_1}{\sigma_1}\right)^2 + \left(\frac{\mu_2-\mu_1}{\sigma_2}\right)^2 = 0$$

Segue-se que $\mu_1 = \mu_2$ e $\sigma_1 = \sigma_2$, *donde mostramos que a função* $\xi : \Theta \to S$ *é injetiva. Portanto* $S = \{p_\xi\}$ *é um modelo estatístico sobre* $\mathbb{R}$ *de dimensão 2. Agora, vamos mostrar que* S *é um modelo regular. Para isto, observe que*

$$\frac{\partial p(x;\mu,\sigma)}{\partial \mu} = \frac{(x-\mu)}{\sqrt{2\pi}\sigma^3} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.5}$$

$$\frac{\partial p(x;\mu,\sigma)}{\partial \sigma} = -\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} + \frac{(x-\mu)^2}{\sqrt{2\pi}\sigma^4} e^{-\frac{(x-\mu)2}{2\sigma^2}}. \tag{3.6}$$

O fato de as funções $\{1, x, x^2\}$ *serem LI implica que* $\dfrac{\partial p(x;\mu,\sigma)}{\partial \mu}$ *e* $\dfrac{\partial p(x;\mu,\sigma)}{\partial \sigma}$ *são LI. Além disso,*

$$\int \frac{\partial p(x;\mu,\sigma)}{\partial \mu} dx = \frac{1}{\sigma^2} E(X-\mu) = \frac{1}{\sigma^2}(E(X)-\mu) = \frac{1}{\sigma^2} 0 = 0.$$

$$\int \frac{\partial p(x;\mu,\sigma)}{\partial \sigma} dx = -\frac{1}{\sigma} \int p(x;\mu,\sigma) dx + \frac{1}{\sigma^3} V(X) = -\frac{1}{\sigma} + \frac{1}{\sigma} = 0.$$

Vemos assim que S *é um modelo regular.*

Exemplo 3.2. *(Distribuições exponenciais) Sendo* $\chi = \mathbb{R}$, $\xi = \beta$, $\Theta = \{\beta \in \mathbb{R} | \ \beta > 0\}$ e $p_\xi = p(x;\beta)$ *dado por*

$$p_\xi = \frac{1}{\beta} e^{-x/\beta}. \tag{3.7}$$

É simples ver que $S = \{p_\xi\}$ *para* p_ξ *definido como na equação anterior é um modelo estatístico de dimensão 1 sobre o intervalo* $(0, \infty)$. *Como*

$$\begin{aligned}\int \frac{\partial p(x;\beta)}{\partial \beta} dx &= -\frac{1}{\beta} \int p(x;\beta) dx + \frac{1}{\beta^2} \int x p(x;\beta) dx \\ &= -\frac{1}{\beta} + \frac{1}{\beta^2} E(X) = -\frac{1}{\beta} + \frac{1}{\beta} = 0.\end{aligned}$$

Mostrando assim que S *é modelo regular.*

Exemplo 3.3. *(Distribuições de Poisson) Sendo* $\chi = \{0, 1, 2, 3, ...\}$, $\Theta = \{\xi \in \mathbb{R} \mid \xi > 0\}$ e $p_\xi = p(x;\xi)$ *dado por*

$$p_\xi = e^{-\xi} \frac{\xi^x}{x!}. \tag{3.8}$$

Sabemos que $\displaystyle\sum_{x=0}^{\infty} \frac{\xi^x}{x!}$ *é a série de potência de* e^ξ, *como em uma série de potência podemos derivar termo a termo, segue a validade da condição* (3) *da definição* (3.3).
Todas as outras condições são de imediata verificação, logo $S = \{p_\xi\}$ *com* p_ξ *definido em* (3.8) *é um modelo estatístico regular de dimensão 1.*

Exemplo 3.4. *(Modelo finito) Seja χ um conjunto finito, onde*

$$\chi = (x_0,\ldots,x_n),\ \Theta = \{(\xi^1,\ldots,\xi_n) \in \mathbb{R}^n |\ \xi^i > 0,\ \forall i = 1,\ldots,n \ e\ \sum_{i=1}^{n} \xi^i < 1\} \text{ e}$$

$$p(x_i;\xi) = \begin{cases} \xi^i, & \text{se } 1 \le i \le n \\ 1 - \sum_{j=1}^{n} \xi^j, & \text{se } i = 0. \end{cases} \tag{3.9}$$

Suas derivadas parcias são

$$\frac{\partial p(x_i;\xi)}{\partial \xi^l} = \begin{cases} 0, & \text{se } i \neq l \text{ e } i \neq 0 \\ 1, & \text{se } l = i \\ -1, & \text{se } i = 0. \end{cases} \tag{3.10}$$

Logo $\sum_{j=1}^{n} \frac{\partial p(x_i;\xi)}{\partial \xi^l} = 0$ *para todo* $l = 1,\ldots,n$ *verificando a condição (3) da definição 4.*

Agora suponha $a_1,\ldots,a_n$ *números reais tais que* $\sum_{l=1}^{n} a_l \frac{\partial p(x;\xi)}{\partial \xi^l} = 0$ *para todo* $x \in \chi$*, fazendo* $x = x_1$*, na última equação obtemos que* $a_1 = 0$*. Analogamente mostramos* $a_2 = a_3 = \ldots = a_n = 0$ *verificando assim a condição* (5) *da definição* (3.3)*. As outras condições são imediatas e portanto temos aqui um modelo estatístico regular de dimensão* $n-1$*. Chamaremos este modelo de* modelo finito *e o denotaremos por* $\wp(\chi)$*. As coordenadas* $\xi = [\xi^i]$ *serão chamados de parametros naturais de* $\wp(\chi)$

Vale observar que apesar da definição de modelo estatístico regular ter bastante restrições, esta não é tão excludente como parece, isto é, os modelos estatístico tradicionais na maioria das vezes são modelos regulares. No entanto, segue abaixo um exemplo de modelo estatístico que não é modelo regular.

Exemplo 3.5. *(Distribuições uniformes) Sendo* $\chi = \mathbb{R}$ *,* $\xi = (\alpha,\beta)$*,* $\Theta = \{(\alpha,\beta) \in \mathbb{R}^2 \mid \beta > \alpha\}$ e $p_\xi = p(x;\alpha,\beta)$ *dado por*

$$p_\xi = \begin{cases} \frac{1}{\beta-\alpha}, & \text{se } \alpha < x < \beta \\ 0, & \text{caso contrário.} \end{cases} \tag{3.11}$$

Considere que $p(x;\alpha_1,\beta_1) = p(x;\alpha_2,\beta_2)$ *para todo* $x \in \chi$*. Suponha por contradição que* $(\alpha_1,\beta_1) \neq (\alpha_2,\beta_2)$*, sem perda de generalidade podemos supor que* $\alpha_1 < \alpha_2$*.*
Seja x_0 *entre* α_1*e* α_2 *tal que* $x_0 \in (\alpha_1,\beta_1)$*; aplicando* x_0 *na última igualdade vem que:*

$$\frac{1}{\beta_1 - \alpha_1} = 0.$$

Absurdo! Logo $(\alpha_1,\beta_1) = (\alpha_2,\beta_2)$*, a função* $\xi : \Theta \to S$ *é injetiva e* $S = \{p_\xi\}$*, onde* P_ξ *é definido na equação (*3.11) é um modelo estatístico sobre X, de dimensão 2. Mas S não é modelo regular pois o supp(p) varia com α *e* β. Este fato já é suficiente para que S não seja regular, mas além disso, é simples ver que as derivadas parcias de $p(x,\alpha,\beta)$ com relação a α *e* β não são LI.

Teorema 3.1. *Seja $S = \{p_\xi\}$ um modelo estatístico regular sobre χ de dimensão n. Então S é uma variedade diferenciável de dimensão n, de Hausdorff e com base enumerável.*

Demonstração. Da hipótese segue a existência de uma função bijetiva $\varphi : \Theta \to S$, e tal que $\Theta \subset \mathbb{R}^n$ é aberto. Seja o conjunto:

$$M = \{f_\alpha : \Theta \to U_\alpha \subset \mathbb{R}^n \mid U_\alpha \text{ é aberto e } f_\alpha \text{ é difeomorfismo } C^\infty\}.$$

Defina $y_\alpha := \varphi \circ f_\alpha^{-1} : U_\alpha \to S$, é simples ver que a família $A = \{(U_\alpha, y_\alpha)\}$ é uma estrutura diferenciável em S e portanto S é uma variedade diferenciável de dimensão n. A afirmação de que esta variedade é de Hausdorff e com base enumerável decorre do fato de S possuir uma parametrização global. □

Teorema 3.2. *Seja S um modelo estatístico regular sobre o conjunto $\chi = (x_0, \ldots, x_n)$. Então S é uma subvariedade imersa do modelo finito $\wp(\chi)$.*

Demonstração. Note que $\wp(\chi)$ nada mais é do que o conjunto de todas as distribuições de probabilidade (estritamente positivas) sobre o conjunto $\chi = (x_0, \ldots, x_n)$ e que é parametrizado por $\varphi : \Theta \to \wp(\chi)$ onde $\varphi(\xi) = p(x;\xi)$ ($p(x;\xi)$ e Θ foram definidos no exemplo (3.4)). Assim, considere

$$S = \left\{q_\eta = q(x;\eta) \mid \eta = (\eta^1, \ldots, \eta^k) \in U \subset \mathbb{R}^k\right\},$$

um modelo estatístico regular sobre χ. Naturalmente, $S \subset \wp(\chi)$ e $k \leq n$. Assim, para provarmos que S é uma subvariedade de $\wp(\chi)$ temos que mostrar que os k vetores $B_l = (\frac{\partial \xi^1}{\partial \eta^l}, \ldots, \frac{\partial \xi^n}{\partial \eta^l})$, com $l = 1, \ldots, k$, são linearmente independentes. Escrevemos

$$q(x_i;\eta) = p(x_i;\xi(\eta)) = \begin{cases} \xi^i(\eta^1, \ldots, \eta^k), & \text{se } 1 \leq i \leq n \\ 1 - \sum_{j=1}^{n} \xi^j(\eta^1, \ldots, \eta^k), & \text{se } i = 0. \end{cases}$$

Assim

$$\frac{\partial}{\partial \eta^l} q(x_i;\eta) = \left(\frac{\partial \xi^1}{\partial \eta^l}, \ldots, \frac{\partial \xi^n}{\partial \eta^l}, -\sum_{j=1}^{n} \frac{\partial \xi^j}{\partial \eta^l}\right).$$

Como, por definição de modelos regulares, as k funções $\frac{\partial}{\partial \eta^l} q(x_i;\eta)$ são linearmente independentes, segue-se que os vetores B_l são linearmente independentes.

□

3.2 A Métrica de Fisher

Definição 3.4. *Seja $S = \{p_\xi\}$ um modelo estatístico regular. Definimos a matriz de Informação de Fisher de S em p_ξ, como sendo a matriz $G(\xi) = [g_{ij}(\xi)]$ definida por*

$$g_{ij}(\xi) = E_\xi[\partial_i \ell_\xi \partial_j \ell_\xi] = \int \partial_i \ell(x;\xi) \partial_j \ell(x;\xi) p(x;\xi) dx. \tag{3.12}$$

Aqui denotamos $\partial_i = \dfrac{\partial}{\partial \xi^i}$ e $\ell_\xi = \ell(x;\xi) = \ln p(x;\xi)$.

Algumas vezes escreveremos $E_\xi[.]$ para denotar $E_p[.]$. Embora existam modelos nos quais a integral da definição acima diverge, neste trabalho, vamos considerar apenas modelos nos quais a matriz (g_{ij}) é finita para todo ξ e, além disso, a função $g_{ij} : \Theta \to \mathbb{R}$ é de classe C^∞. Veremos que a matriz de informação de Fisher define uma métrica Riemanniana sobre modelos estatísticos. Vejamos alguns exemplos.

Exemplo 3.6. *Considere S modelo das distribuições normais visto no exemplo* (3.1). *Vamos calcular sua matriz de informação de Fisher.*

$$\ell(x;\mu,\sigma) = \ln p(x;\mu,\sigma) = -\ln(\sqrt{2}\pi) - \ln\sigma - \frac{(x-\mu)^2}{2\sigma^2}.$$

Logo, os coeficientes da matriz de informação de Fisher são dados por

$$\begin{aligned} g_{11}(\mu,\sigma) &= \int \frac{\partial \ell(x;\mu,\sigma)}{\partial \mu}\frac{\partial \ell(x;\mu,\sigma)}{\partial \mu} p(x;\mu,\sigma)dx \\ &= \int \frac{(x-\mu)}{\sigma^2}\frac{(x-\mu)}{\sigma^2} p(x;\mu,\sigma)dx = \frac{1}{\sigma^4}V(X) = \frac{1}{\sigma^2} \end{aligned}$$

$$\begin{aligned} g_{12}(\mu,\sigma) = g_{21}(\mu,\sigma) &= \int \frac{\partial \ell(x;\mu,\sigma)}{\partial \mu}\frac{\partial \ell(x;\mu,\sigma)}{\partial \sigma} p(x;\mu,\sigma)dx \\ &= \int \frac{(x-\mu)}{\sigma^2}\left(\frac{-1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}\right) p(x;\mu,\sigma)dx \\ &= \frac{-1}{\sigma^3}E[X-\mu] + \frac{1}{\sigma^5}E[(X-\mu)^3] = 0+0=0 \end{aligned}$$

$$\begin{aligned} g_{22}(\mu,\sigma) &= \int \frac{\partial \ell(x;\mu.\sigma)}{\partial \sigma}\frac{\partial \ell(x;\mu.\sigma)}{\partial \sigma} p(x;\mu.\sigma)dx \\ &= \int \left(-\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}\right)^2 p(x;\mu.\sigma)dx \\ &= \frac{1}{\sigma^2}\int p(x;\mu,\sigma) + -\frac{2}{\sigma^4}V[X] + \frac{1}{\sigma^6}E[(x-\mu)^4] \\ &= \frac{1}{\sigma^2} - \frac{2}{\sigma^2} + \frac{3\sigma^4}{\sigma^6} = \frac{2}{\sigma^2} \end{aligned}$$

Logo,

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

OBS. *Vemos que os coeficientes da métrica G são* $E = 1/\sigma^2$, $F = 0$ e $G = 2/\sigma^2$. *Sendo assim, pode-se provar que* (S,G) *é uma superfície completa, simplesmente conexa, e de curvatura Gaussiana constante* $-1/2$. *Sendo assim* (S,G) *é isométrica ao plano hiperbólico de curvatura Gaussiana constante* $-1/2$.

Exemplo 3.7. *Seja $\wp(\chi)$, com $\chi = \{0,1,\dots,n\}$, o modelo finito, visto na seção anterior. Temos*

$$\ell_\xi(i) = \ln p_\xi(i) = \begin{cases} \ln \xi^i, & \text{se } 1 \le i \le n \\ \ln(1 - \sum_{j=1}^n \xi^j), & \text{se } i = 0 \end{cases}$$

$$\frac{\partial \ell_\xi(i)}{\partial \xi^l} = \frac{\partial \ln p_\xi(i)}{\partial \xi^l} = \begin{cases} 0, & \text{se } i \neq l \text{ e } i \neq 0 \\ 1/\xi^l, & \text{se } i = l \\ -1/(1 - \sum_{j=1}^n \xi^j), & \text{se } i = 0 \end{cases}$$

e os coeficientes da matriz de informação de Fisher são dados por

$$g_{lk}(\xi) = \sum_{i=0}^n \frac{\partial \ell_\xi(i)}{\partial \xi^l}\frac{\partial \ell_\xi(i)}{\partial \xi^k} p_\xi(x_i) = \begin{cases} 1/(1 - \sum_{j=1}^n \xi^j) + 1/\xi^l, & \text{se } l = k \\ 1/(1 - \sum_{j=1}^n \xi^j), & \text{se } l \neq k. \end{cases}$$

Mostraremos que $\wp(\chi)$, com a métrica de Fisher, possui curvatura constante positiva. De fato, considere $q_\xi(x) = 2\sqrt{p_\xi(x)}$, com $x \in \chi$, e $q_\xi := [q_\xi(0),\dots,q_\xi(n)]$. Temos que $q_\xi \in \mathbb{R}^{n+1}$ satisfaz $q_\xi(i) > 0$, para todo i, e $\sum_{j=0}^n q_\xi(j)^2 = 4$, donde q_ξ pertence à esfera $S_2^n \subset \mathbb{R}^{n+1}$ centrada na origem e raio 2. Note que a matriz de informação de Fisher $g = (g_{ij})$ de $S = \{p_\xi\}$ também se escreve-se como

$$g_{ij} = \sum_{i=0}^n \frac{\partial(2\sqrt{p_\xi})}{\partial \xi^i}\frac{\partial(2\sqrt{p_\xi})}{\partial \xi^j} = \frac{\partial q_\xi}{\partial \xi^i} \cdot \frac{\partial q_\xi}{\partial \xi^j},$$

onde " $\cdot$ " denota o produto escalar de $\mathbb{R}^{n+1}$. Segue-se que $p \in S \mapsto q = 2\sqrt{p} \in S_2^n$ é uma imersão isométrica, donde (S,g) possui curvatura secctional constante positiva.

Exemplo 3.8. *Considere S o modelo formado pelas distribuições exponenciais, $S = \{p(x;\beta) = \frac{1}{\beta}e^{-x/\beta}\}$. Neste caso, sabemos que S é um modelo estatístico de dimensão 1. Além disso, usando integração por partes, vemos facilmente que a média $E_\beta(X) = \beta$ e a variância $V_\beta(X) = \beta^2$. Usando que*

$$\ell(x;\beta) = \ln p(x;\beta) = -\ln\beta - \frac{x}{\beta},$$

segue-se que a Informação de Fisher é dada por

$$I(\beta) = \int \frac{\partial \ell(x;\beta)}{\partial \beta}\frac{\partial \ell(x;\beta)}{\partial \beta} p(x;\beta)dx = \int \frac{(x-\beta)^2}{\beta^4} p(x;\beta)dx = \frac{V(X)}{\beta^4} = \frac{1}{\beta^2}.$$

Note que $E_\beta(X) = \beta$ e a variância $V(X) = \beta^2 = 1/I(\beta)$. Veremos adiante que isso é um caso particular do Teorema de Cramer-Rao.

Lema 3.3. *A expressão dos coeficientes da matriz de informação de Fisher também podem ser expressos da seguinte forma;*

$$g_{ij}(\xi) = -E_\xi[\partial_i \partial_j \ell_\xi]. \tag{3.13}$$

Demonstração. Primeiro observe que

$$E_\xi[\partial_j \ell_\xi] = 0, \tag{3.14}$$

pois,

$$\begin{aligned} E_\xi[\partial_j \ell_\xi] = \int \frac{\partial \ln p(x;\xi)}{\partial \xi^j} p(x;\xi))dx &= \int \frac{1}{p(x;\xi)} \frac{\partial p(x;\xi)}{\partial \xi^j} p(x;\xi)dx \\ &= \frac{\partial}{\partial \xi^j} \int p(x;\xi)dx \\ &= \frac{\partial}{\partial \xi^j} 1 = 0. \end{aligned}$$

Agora derivando a equação (3.14) em relação ξ^i, obtemos

$$\begin{aligned} 0 &= \frac{\partial}{\partial \xi^i} \int \frac{\partial \ln p(x;\xi)}{\partial \xi^j} p(x;\xi))dx \\ &= \int \frac{\partial^2 \ln p(x;\xi)}{\partial \xi^i \partial \xi^j} p(x;\xi))dx + \int \frac{\partial \ln p(x;\xi)}{\partial \xi^j} \frac{\partial p(x;\xi)}{\partial \xi^i} dx \\ &= E_\xi[\partial_i \partial_j \ell_\xi] + \int \frac{\partial \ln p(x;\xi)}{\partial \xi^j} \frac{\partial p(x;\xi)}{\partial \xi^i} \frac{p(x;\xi)}{p(x;\xi)} dx \\ &= E_\xi[\partial_i \partial_j \ell_\xi] + \int \frac{\partial \ln p(x;\xi)}{\partial \xi^j} \frac{\partial \ln p(x;\xi)}{\partial \xi^i} p(x;\xi)dx \\ &= E_\xi[\partial_i \partial_j \ell_\xi] + g_{ij}(\xi). \end{aligned}$$

□

Teorema 3.4. *Seja $S = \{p_\xi\}$ um modelo estatístico regular. Então a matriz de informação de Fisher define uma métrica Riemanniana sobre S.*

Demonstração. Fixemos $p = p(\cdot;\xi) \in S$ e $X, Y \in T_pS$. Escrevamos X e Y em termos da base coordenada, $X = \sum_{i=1}^n X^i \partial_i$ e $Y = \sum_{i=1}^n Y^i \partial_i$, onde $\partial_i = \frac{\partial}{\partial \xi^i}$. Usando que $\partial_i \ell = \frac{1}{p} \partial_i(p)$, temos o seguinte:

$$\begin{aligned} \sum_{i,j} X^i Y^j g_{ij}(\xi) &= \int \sum_{i,j} X^i Y^j \partial_i \ell(x;\xi)\, \partial_j \ell(x;\xi) p(x;\xi)dx \\ &= \int \sum_i \frac{X^i \partial_i p(x;\xi)}{p(x;\xi)} \sum_j \frac{Y^j \partial_j p(x;\xi)}{p(x;\xi)} p(x;\xi)dx \\ &= \int \frac{X p(x;\xi)}{p(x;\xi)} \frac{Y p(x;\xi)}{p(x;\xi)} p(x;\xi)\, dx. \end{aligned}$$

Da ultima igualdade, segue-se que a matriz de informação de Fisher determina o seguinte 2-tensor sobre S

$$g(p)(X,Y) = \int \frac{X p(x;\xi)}{p(x;\xi)} \frac{Y p(x;\xi)}{p(x;\xi)} p(x;\xi)\, dx.$$

Segue-se diretamente que $g(p)$ está bem definida sobre T_pS, é simétrica e $g(p)(X,X) \geq 0$, valendo zero se, e somente se $X(p) = \sum X^i \partial_i(p) = 0$. Como $\partial_i(p)$ são linearmente independentes (como função de x), segue-se que $X = 0$. Donde g é uma métrica Riemanniana sobre S.

A métrica acima definida é chamada de métrica de informação de Fisher, ou, simplesmente, métrica de Fisher.

3.3 As α-Conexões

Como vimos no capítulo 1, dada M uma variedade Riemanniana, existe uma única conexão ∇ em M que é simétrica e compátivel com a métrica de M , chamada de conexão de Levi-Civita. Nesta seção, para S um modelo estatístico munido com a métrica de Ficher, vamos introduzir uma família de conexões simétricas. Evidentemente, pelo que mencionamos, apenas uma conexão dessa família será compatível com a métrica de Fisher. Tais conexões são fundamentais para o estudo dos modelos estatísticos do ponto de vista geométrico.

Seja $S = \{p_\xi\}$ um modelo estatístico regular. Para cada $\alpha \in \mathbb{R}$, consideremos as funcões $\Gamma^{(\alpha)}_{ij,k} : \Theta \to \mathbb{R}$ definidas por:

$$\Gamma^{(\alpha)}_{ij,k}(\xi) = E_\xi[(\partial_i \partial_j \ell_\xi + \frac{1-\alpha}{2} \partial_i \ell_\xi \partial_j \ell_\xi) \partial_k \ell_\xi]. \tag{3.15}$$

As n^3 funções $\Gamma^{(\alpha)}_{ij,k}(\xi)$ definem uma conexão $\nabla^{(\alpha)}$ sobre S dada por

$$\langle \nabla^{(\alpha)}_{\partial_i} \partial_j, \partial_k \rangle = \Gamma^{(\alpha)}_{ij,k}(\xi), \tag{3.16}$$

onde $g = \langle \cdot, \cdot \rangle$ é a métrica de Fisher. Vamos chamar a conexão $\nabla^{(\alpha)}$ de *α-conexão*. Lembrando que para cada $x \in \chi$ as funções $p(x, \xi)$ são de classe C^∞, segue do teorema de Clairaut-Schwarz que $\partial_i \partial_j \ell_\xi = \partial_j \partial_i \ell_\xi$ e portanto as funções $\Gamma^{(\alpha)}_{ij,k}$ são simétricas nos dois primeiros índices. Consequência imediata disso é que as α-conexões são conexões simétricas.

Quando S for plana com respeito $\nabla^{(\alpha)}$ diremos que S é α-plana e que $[\xi]$ é um sistena de cordenadas α- afim. Nos casos especiais: $\alpha = 1$, diremos que S é e-plana e, para $\alpha = -1$, que S é m-plana.

Teorema 3.5. *A 0-conexão é a conexão Riemanniana com respeito à métrica de Fisher.*

Demonstração. Como a conexão $\nabla^{(0)}$ é simétrica, basta mostrar que $\nabla^{(0)}$ é compatível com a métrica de Fisher.

Derivando a equação (3.12) em relação ξ^k temos

$$\begin{aligned}
\partial_k g_{ij} &= E_\xi[(\partial_k\partial_i\ell_\xi)\partial_j\ell_\xi] + E_\xi[\partial_i\ell_\xi(\partial_k\partial_j\ell_\xi)] + E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] \\
&= E_\xi[(\partial_k\partial_i\ell_\xi)\partial_j\ell_\xi] + E_\xi[\partial_i\ell_\xi(\partial_k\partial_j\ell_\xi)] + \frac{1}{2}E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] + \frac{1}{2}E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] \\
&= E_\xi[(\partial_k\partial_i\ell_\xi)\partial_j\ell_\xi] + \frac{1}{2}E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] + E_\xi[\partial_i\ell_\xi(\partial_k\partial_j\ell_\xi)] + \frac{1}{2}E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] \\
&= \Gamma^{(0)}_{ki,j} + \Gamma^{(0)}_{kj,i}.
\end{aligned}$$

Mostrando assim que a 0-conexão é compatível com a métrica de Fischer. $\square$

Teorema 3.6. *Seja $S = \{p_\xi\}$ um modelo estatístico regular e $g = \langle\cdot,\cdot\rangle$ a métrica de Fisher. Então as conexões $\nabla^{(\alpha)}$ e $\nabla^{(-\alpha)}$ são duais.*

Demonstração. Em termos de coordenadas,

$$\begin{aligned}
\Gamma^{(\alpha)}_{ki,j} + \Gamma^{(-\alpha)}_{kj,i} &= E_\xi[(\partial_k\partial_i\ell_\xi + \frac{1-\alpha}{2}\partial_k\ell_\xi\partial_i\ell_\xi)\partial_j\ell\xi] + E_\xi[(\partial_k\partial_j\ell_\xi + \frac{1+\alpha}{2}\partial_k\ell_\xi\partial_j\ell_\xi)\partial_i\ell\xi] \\
&= E_\xi[(\partial_k\partial_i\ell_\xi)\partial_j\ell_\xi] + E_\xi[\partial_i\ell_\xi(\partial_k\partial_j\ell_\xi)] + E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi\partial_k\ell_\xi] \\
&= \partial_k E_\xi[\partial_i\ell_\xi\partial_j\ell_\xi] = \partial_k g_{ij}.
\end{aligned}$$

Segue de (2.16) que $\nabla^{(-\alpha)}$ é a conexão dual de $\nabla^{(\alpha)}$. $\square$

A partir deste último resultado, juntamente com a Teorema (2.14), temos que um modelo estatístico é α-plana se e somente se, é $(-\alpha)$-plana.

Para exemplificar, vamos calcular os símbolos de Christoffel $\Gamma^{(\alpha)}_{ij,k}$ e o α-tensor curvatura $R^{(\alpha)}$ para o modelo normal. Antes vejamos três identidades que nos ajudarão nas próximas contas.

Lema 3.7. *Se $p = p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, então*

(a) $\displaystyle\int (x-\mu)^4 p\, dx = 3\sigma^4,$

(b) $\displaystyle\int (x-\mu)^6 p\, dx = 15\sigma^6,$

(c) $\displaystyle\int (x-\mu)^3 p\, dx = 0.$

Demonstração. Basta usar integração por partes.

Para (a), considere

$$z = (x-\mu)^3 \quad \text{e} \quad dv = (x-\mu)p\, dx,$$

daí

$$dz = 3(x-\mu)^2 dx \quad \text{e} \quad v = -\sigma^2 p.$$

Já que $zv\big|_{-\infty}^{\infty} = 0$, temos que

$$\int (x-\mu)^4 p \text{ dx} = 3\sigma^2 \int (x-\mu)^2 p \text{ dx} = 3\sigma^2 V[X] = 3\sigma^4.$$

(b), seja

$$y = (x-\mu)^5 \quad \text{e} \quad dv = (x-\mu)p\, dx,$$

então

$$dy = 5(x-\mu)^4 dx \quad \text{e} \quad v = -\sigma^2 p.$$

Como $yv\big|_{-\infty}^{\infty} = 0$, temos que

$$\int (x-\mu)^6 p \text{ dx} = 5\sigma^2 \int (x-\mu)^4 p \text{ dx} = 5\sigma^2 3\sigma^4 = 15\sigma^6.$$

(c), tome

$$y = (x-\mu)^2 \quad \text{e} \quad dv = (x-\mu)p\, dx,$$

logo

$$dy = 2(x-\mu)dx \quad \text{e} \quad v = -\sigma^2 p.$$

Como $yv\big|_{-\infty}^{\infty} = 0$, temos que

$$\int (x-\mu)^3 p\, dx = 2\sigma^2 \int (x-\mu) p\, dx = 2\sigma^2 E[X-\mu] = 0.$$

□

Proposição 3.8. *As funções $\Gamma_{ij,k}^{(\alpha)}$ são dadas por*

$$\begin{aligned} \Gamma_{11,2}^{(\alpha)} &= \frac{1-\alpha}{\sigma^3}, \\ \Gamma_{12,1}^{(\alpha)} = \Gamma_{21,1}^{(\alpha)} &= -\frac{1+\alpha}{\sigma^3}, \\ \Gamma_{22,2}^{(\alpha)} &= -\frac{2+\alpha}{\sigma^3} \end{aligned}$$

Enquanto os outros componentes são zero.

Demonstração. Por definição

$$\Gamma_{ij,k}^{(\alpha)} = \int \left[\frac{\partial^2 \ln p}{\partial \xi^i \partial \xi^j} \frac{\partial \ln p}{\partial \xi^k} + \frac{1-\alpha}{2} \frac{\partial \ln p}{\partial \xi^i} \frac{\partial \ln p}{\partial \xi^j} \frac{\partial \ln p}{\partial \xi^k} \right] p\, dx$$

e

$$\ln p(x;\mu,\sigma) = -\left(\ln \sqrt{2\pi} + \ln \sigma + \frac{(x-\mu)^2}{2\sigma^2} \right).$$

Calculando as derivadas parciais, temos

$$\frac{\partial \ln p}{\partial \mu} = \frac{x-\mu}{\sigma^2}\,, \qquad \frac{\partial^2 \ln p}{\partial^2 \mu} = \frac{-1}{\sigma^2}\,,$$

$$\frac{\partial \ln p}{\partial \sigma} = \frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\,, \quad \frac{\partial^2 \ln p}{\partial^2 \sigma} = \frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}$$

e

$$\frac{\partial^2 \ln p}{\partial \mu \partial \sigma} = -2\frac{(x-\mu)}{\sigma^3}.$$

$$\begin{aligned}
\Gamma_{11,2}^{(\alpha)} &= \int \left[\frac{-1}{\sigma^2}\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right) + \frac{1-\alpha}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right)\right] p \text{ dx} \\
&= \int \frac{1}{\sigma^3} p \text{ dx} - \frac{1}{\sigma^5}\int (x-\mu)^2 p \text{ dx} + \frac{1-\alpha}{2}\left(\frac{1}{\sigma^7}\int (x-\mu)^4 p \text{ dx} - \frac{1}{\sigma^5}\int (x-\mu)^2 p \text{ dx}\right) \\
&= \frac{1}{\sigma^3} - \frac{1}{\sigma^5} V[X] + \frac{1-\alpha}{2}\left(\frac{1}{\sigma^7} 3\sigma^4 - \frac{1}{\sigma^5} V[X]\right) \\
&= 0 + \frac{1-\alpha}{2}\frac{2}{\sigma^3} = \frac{1-\alpha}{\sigma^3}.
\end{aligned}$$

$$\begin{aligned}
\Gamma_{12,1}^{(\alpha)} &= \int \left[\left(-2\frac{(x-\mu)}{\sigma^3}\right)\left(\frac{x-\mu}{\sigma^2}\right) + \frac{1-\alpha}{2}\left(\frac{x-\mu}{\sigma^2}\right)\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right)\left(\frac{x-\mu}{\sigma^2}\right)\right] p \text{ dx} \\
&= -\frac{2}{\sigma^5}\int (x-\mu)^2 p\, dx + \frac{1-\alpha}{2}\int \left(\frac{x-\mu}{\sigma^2}\right)^2\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right) p\, dx \\
&= -\frac{2}{\sigma^5} V[X] + \frac{1-\alpha}{\sigma^3} = \\
&= -\frac{1+\alpha}{\sigma^3}.
\end{aligned}$$

$$\begin{aligned}
\Gamma_{22,2}^{(\alpha)} &= \int \left[\left(\frac{1}{\sigma^2} - 3\frac{(x-\mu)^2}{\sigma^4}\right)\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right) + \frac{1-\alpha}{2}\left(\frac{(x-\mu)^2}{\sigma^3} - \frac{1}{\sigma}\right)^3\right] p \text{ dx} \\
&= \frac{1}{\sigma^5}\int (x-\mu)^2 p\, dx - \frac{1}{\sigma^3}\int p\, dx - \frac{3}{\sigma^7}\int (x-\mu)^4 p\, dx + \frac{3}{\sigma^5}\int (x-\mu)^2 p\, dx \\
&+ \frac{1-\alpha}{2}\left(\frac{1}{\sigma^9}\int (x-\mu)^6 p\, dx - \frac{3}{\sigma^7}\int (x-\mu)^4 p\, dx + \frac{3}{\sigma^5}\int (x-\mu)^2 p\, dx - \frac{1}{\sigma^3}\int p\, dx\right) \\
&= \frac{1}{\sigma^3} - \frac{1}{\sigma^3} - \frac{3}{\sigma^7} 3\sigma^4 + \frac{3}{\sigma^3} + \frac{1-\alpha}{2}\left(\frac{1}{\sigma^9} 15\sigma^6 - \frac{3}{\sigma^7} 3\sigma^4 + \frac{3}{\sigma^5}\sigma^2 - \frac{1}{\sigma^3}\right) \\
&= -\frac{2+4\alpha}{\sigma^3}.
\end{aligned}$$

□

Como

$$\sum_{l=1}^{2} \Gamma_{ij}^{l(\alpha)} g_{lm} = \langle \nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_m \rangle = \Gamma_{ij,m}^{(\alpha)}, \quad (m = 1, 2.).$$

Então

$$\Gamma_{ij}^{k(\alpha)} = \sum_{m=1}^{2} \Gamma_{ij,m}^{(\alpha)} g^{mk}.$$

Lembrando que

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

obtemos o seguinte resultado:

Corolário 3.9. *Os símbolos de Christoffel $\Gamma_{ij}^{k(\alpha)}$ da $\nabla^{(\alpha)}$ são dados por*

$$\Gamma_{12}^{1(\alpha)} = \Gamma_{21}^{1(\alpha)} = -\frac{1+\alpha}{\sigma},$$
$$\Gamma_{11}^{2(\alpha)} = \frac{1-\alpha}{2\sigma},$$
$$\Gamma_{22}^{2(\alpha)} = -\frac{2\alpha+1}{\sigma}.$$

Enquanto os outros componentes são zero.

Corolário 3.10. *Os coeficientes $R_{ijk}^{l(\alpha)}$ do tensor curvatura $R^{(\alpha)}$ são dados por*

$$R_{121}^{2(\alpha)} = -R_{211}^{2(\alpha)} = \frac{\alpha^2 - 1}{2\sigma^2}.$$

Enquanto os outros componentes são zero. Em particular, $R^{(\alpha)}$ é identicamente nulo se e somente se, $\alpha = 1$ ou $\alpha = -1$, equivalentemente (S, ∇^{α}) é plana se e somentese, $\alpha = 1$ ou $\alpha = -1$. Neste caso, podemos dizer que o modelo normal é e-plana e m-plana (mostraremos este fato de um modo muito mais simples na próxima seção).

A demostração é feita por um cálculo direto, apenas usando a expressão dos coeficientes do tensor curvatura e o Corolário (3.9). Mais adiante veremos que muitos modelos estatísticos são variedades dualmente planas.

3.4 Famílias Exponenciais e Misturas

Dentre os modelos estatísticos, duas classes de destaque são as Famílias exponenciais e as misturas. Veremos que as misturas são, naturalmente, *m*-planas, enquanto que as famílias exponenciais são, naturalmente, *e*-planas. Conclui-se que ambas são dualmente *e* e *m* planas. Muito dos modelos exemplificados neste trabalho são famílias exponenciais.

Definição 3.5 (Famílias Exponenciais)**.** *Sejam* $C, F_1, \ldots, F_n : \chi \to \mathbb{R}$ *funções tais que* $1, F_1, \ldots, F_n$ *são linearmente independentes. Seja*

$$\Theta = \{\theta = [\theta^1, \ldots, \theta^n] \mid \psi(\theta) := \ln \int e^{C(x)+\sum_i \theta^i F_i(x)} dx \quad \text{é finito}\}.$$

Assim,

$$S = \{p_\theta(x) = p(x;\theta) = e^{C(x)+\sum_i \theta^i F_i(x) - \psi(\theta)} \mid \theta = (\theta^1, \ldots, \theta^n) \in \Theta \subset \mathbb{R}^n\}$$

é um modelo estatístico regular sobre χ *de dimensão* n*. Dizemos que* S *é uma família exponencial. Neste caso, chamamos* $\theta = [\theta^i]$ *de parametros canônicos de* S*.*

Além disso, como $\ell_\theta = \ln p_\theta = C(x) + \sum_i \theta^i F_i(x) - \psi(\theta)$, temos

$$0 = E[\frac{\partial}{\partial \theta^j} \ln p_\theta] = \int p_\theta(x;\theta)(F_j(x) - \frac{\partial \psi(\theta)}{\partial \theta^j})dx = E_\theta[F_j(X)] - \frac{\partial \psi(\theta)}{\partial \theta^j}.$$

Assim,

$$E_\theta[F_j(X)] = \partial_j \psi(\theta). \tag{3.17}$$

Além disso, $\frac{\partial^2}{\partial \theta^i \partial \theta^j} \ln p_\theta = -\frac{\partial^2}{\partial \theta^i \partial \theta^j} \psi(\theta)$. Logo, a métrica de Fisher satisfaz

$$g_{ij}(\theta) = -E_\theta[\partial_i \partial_j \ell_\theta] = \partial_i \partial_j \psi(\theta). \tag{3.18}$$

Em particular, ψ é uma função convexa.

É sabido que muitos modelos estatísticos tradicionias são famílias exponenciais. Abaixo vamos mostrar que todos os exemplos de modelos estatístico regulares da seção (3.1) são famílias exponenciais.

Exemplo 3.9. *Dada uma distribuição de probabilidade do modelo normal, podemos reescrevê-la da seguinte forma*

$$p(x;\mu,\sigma) = e^{\left(-\ln\sqrt{2\pi}\sigma - \frac{(x-\mu)^2}{2\sigma^2}\right)} = e^{\left(x\frac{\mu}{\sigma^2} - x^2\frac{1}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln\sqrt{2\pi}\sigma\right)}. \tag{3.19}$$

Tomando

$$\begin{aligned} &\theta^1 = \frac{\mu}{\sigma^2},\ \theta^2 = -\frac{1}{2\sigma^2}, \\ &F_1(x) = x,\ E_2(x) = x^2,\ C(x) = 0,\ e \\ &\psi(\theta^1,\theta^2) = \frac{\mu^2}{2\sigma^2} + \ln\sqrt{2\pi}\sigma = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\ln(\frac{-\pi}{\theta^2}). \end{aligned}$$

Obtemos que

$$p(x;\mu,\sigma) = q(x;\theta^1,\theta^2) = e^{C(x)+\sum_{n=1}^2 F_i(x)\theta^i + \psi(\theta^1,\theta^2)}.$$

Exemplo 3.10. *Quando S é um modelo formado pelas distribuições exponenciais, escrevemos*

$$p(x;\beta) = \frac{1}{\beta}e^{-\frac{x}{\beta}} = e^{-ln\beta - \frac{x}{\beta}}.$$

Tomando

$$F_1(x) = x, \quad \theta = -\frac{1}{\beta}, \quad C(x) \equiv 0 \quad e \quad \psi(\theta) = \ln(\beta) = \ln(-1/\theta).$$

Obtemos que

$$p(x;\beta) = q(x;\theta) = e^{F_1(x)\theta - \psi(\theta)}.$$

Exemplo 3.11. *Para S um modelo de Poisson, temos*

$$p(x;\xi) = e^{-\xi}\frac{\xi^x}{x!} = e^{-lnx! + xln\xi - \xi}.$$

Tomando

$$F_1(x) = x, \quad \theta = ln\xi, \quad C(x) = -lnx \quad e \quad \psi(\theta) = e^\theta$$

segue-se que

$$p(x;\xi) = q(x;\theta) = e^{C(x) + F_1(x)\theta - \psi(\theta)}.$$

Exemplo 3.12. *Considere o modelo finito $\wp(\chi)$, com $\chi = \{0, 1, \ldots, n\}$. Escreva $F_i(j) = \delta_{ij}$, e*

$$\theta^i = \ln\left(\frac{p(x_i)}{p(x_0)}\right) = \ln\left(\frac{\xi^i}{1 - \sum_{j=1}^n \xi^j}\right), \qquad (i = 1, \ldots, n)$$

$$C(x) \equiv 0 \quad e \quad \psi(\theta) = -\ln p(x_0) = \ln\left(1 + \sum_{i=1}^n e^{\theta^i}\right).$$

Segue-se que $\xi = \xi(\theta)$ e $p_\xi(x) = e^{C(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)}$, para todo $x \in \chi$. Concluímos que S é uma família exponencial.

Proposição 3.11. *Uma família exponencial S é uma variedade dualmente plana. Mais especificamente, se $S = \{p(\cdot\,;\theta)\}$ é da forma*

$$p(x;\theta) = e^{C(x) + \theta^i F_i(x) - \psi(\theta)},$$

então as coordenadas $\theta = [\theta^i]$ são e-afins e as coordenas $\eta = [E[F_j]] = \nabla\psi$ são m-afins. Além disso, a métrica de Fisher $g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}$ e as coordenadas θ e η são duais, i.e., satisfazem as seguintes relações:

$$\frac{\partial\eta_i}{\partial\theta^j} = g_{ij}(\theta) = g^{ij}(\eta).$$

Demonstração. Vejamos que os parametros canônicos $\theta = [\theta^i]$ formam um sistema de coordenadas e-afim. De fato,

$$\ell(x;\theta) = \ln p(x;\theta) = C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta).$$

Logo,

$$\partial_i \ell(x;\theta) = F_i(x) - \partial_i \psi(\theta) \tag{3.20}$$

e

$$\partial_i \partial_j \ell(x;\theta) = -\partial_i \partial_j \psi(\theta). \tag{3.21}$$

Daí, por (3.14), temos que a métrica de Fisher é dada por $g_{ij}(\theta) = -E[\partial_i \partial_j(\ell)] = \partial_i \partial_j \psi$ e a 1-conexão $\Gamma^{(e)}_{ij,k} = E_p[\partial_i \partial_j(\ell)\partial_k(\ell)] = -\partial_i \partial_j \psi(\theta) E_\theta[\partial_k l_\theta] = 0$. Assim, θ é um sistema de coordenadas e-afim para S.

Agora, considere

$$\eta_j := E_\theta[F_j] = \frac{\partial}{\partial \theta^j}\psi(\theta). \tag{3.22}$$

A última igualdade segue-se de (3.20). Derivando em relação a θ^i, obtemos

$$\frac{\partial \eta_j}{\partial \theta^i} = \frac{\partial^2}{\partial \theta^i \partial \theta^j}\psi(\theta) = g_{ij}(\theta). \tag{3.23}$$

Observe que η define um sistema de coordenas, visto que g_{ij} é uma matriz positiva-definida, logo não-singular.

Vamos concluir mostrando que η é um sistema de coordenas m-afim. Escrevendo $\partial_i = \frac{\partial}{\partial \theta^i}$, como $\frac{\partial}{\partial \eta_k} = \frac{\partial \theta^i}{\partial \eta_k}\frac{\partial}{\partial \theta^i} = g^{ij}(\theta)\partial_i$, segue-se que $\partial^i := g^{ij}\partial_j = \frac{\partial}{\partial \eta_j}$. Assim,

$$\left\langle \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j} \right\rangle = \langle \partial_i, \partial^j \rangle = \left\langle \partial_i, g^{jk}\partial_k \right\rangle = g^{jk} g_{ik} = \delta_i^j.$$

Portanto, para todo $Z \in T_pS$,

$$0 = Z(\delta_i^j) = Z\langle \partial_i, \partial^j \rangle = \left\langle \nabla^{(e)}_Z \partial_i, \partial^j \right\rangle + \left\langle \partial_i, \nabla^{(m)}_Z \partial^j \right\rangle = \left\langle \partial_i, \nabla^{(m)}_Z \partial^j \right\rangle.$$

Segue-se que $\partial^j = \frac{\partial}{\partial \eta_j}$ são campos m-paralelos, donde o sistema de coordenadas η é m-afim. □

Observe que a demostração da Proposição (3.11) nos diz como obter um sistema de coordenadas m-afim para uma família exponencial. Por exemplo:

(i) para o modelo normal, $(\eta_1, \eta_2) = (E[X], E[X^2]) = (\mu, \mu^2 + \sigma^2)$,

(ii) para o modelo finito $\wp(\chi)$, $\eta_i = E[F_i(X)] = \xi^i$ e

(iii) para o modelo de Poisson, $\eta = E[X] = \xi$.

Note ainda que depois de concluirmos que $\partial^k = \frac{\partial}{\partial \eta_k} = g^{ki}(\theta)\frac{\partial}{\partial \theta^i}$, não era preciso mostrar que o sistema de cordenadas η é m-afim, pois este argumento é dado na demostração do Teorema (2.14).

Proposição 3.12. *Sejam $S^n = \{p_\xi = p(\cdot;\xi)\}$ e $M^m = \{q_u = q(\cdot;u)\}$ dois modelos estatístico sobre χ. Assuma que S é uma família esponencial e M seja uma subvariedade isometricamente imersa em S. Então, M e-autoparalela em S se, e somente se, M é uma família exponencial.*

Demonstração. Seja $f : M \to S$ uma imersão isométrica injetiva de $M^m = \{q_u\}$ sobre a família exponencial $S = \{p_\xi\}$. Escreva $p_\xi(x) = e^{C(x)+\sum_i \xi^i F_i(x)-\psi(\xi)}\}$, onde $1, F_1(x), \ldots, F_n(x)$ são funções linearmente independentes. Segue-se que $\bar{M} = \{\bar{q}_u = f(q(x;u)) = p(x;\xi(u))\}$ é um modelo estatístico.

Primeiro, assuma que M é uma família exponencial e escreva $q_u(x) = e^{D(x)+u^a G_a(x)-\varphi(u)}$, sendo $1, G_1, \ldots, G_m$ funções linearmente independentes. Assim, $M' = \{\bar{q}_u = f \circ q_u\}$ é uma família exponencial da forma

$$\ln \bar{q}_u(x) = D(x) + \sum_a u^a G_a(x) - \varphi(u) = C(x) + \sum_i \xi^i(u) F_i(x) - \psi(\xi(u)).$$

Derivando-se ambos lados duas vezes em relação a u^a e u^b, temos

$$-\partial_a \partial_b \varphi(u) = \sum_i \frac{\partial^2 \xi^i}{\partial u^a u^b} F_i(x) - \partial_a \partial_b \psi(\xi(u)).$$

Temos, $\sum_i \frac{\partial^2 \xi^i}{\partial u^a u^b} F_i(x) + \partial_a \partial_b(\varphi(u) - \psi(\xi(u))) = 0$. Usando que $1, F_1, \ldots, F_n$ são l.i., segue-se que $\frac{\partial^2 \xi^i}{\partial u^a u^b} = \partial_a \partial_b(\varphi(u) - \psi(\xi(u))) = 0$, para quaisquer $i = 1, \ldots, n$ e $a = 1, \ldots, m$. Assim,

$$\xi^i(u) = \sum_a A^i_a u^a + b^i \quad \text{e} \quad \psi(\xi(u)) = cu + d,$$

com A^i_a, b^i, c, d constantes. Os campos coordenados de $M = \{q_u\}$ satisfazem

$$\partial_a \bar{q}_u = f_* \partial_a q_u = \sum_i \frac{\partial p_\xi}{\partial \xi^i} \frac{\partial \xi^i}{\partial u^a} = \sum_i \frac{\partial \xi^i}{\partial u^a} \partial_i p_{\xi(u)} = \sum_i A^i_a \partial_i p_\xi.$$

Como S é e-plana, segue-se que os vetores coordenados ∂_i são e-paralelos. Assim,

$$\left\langle \nabla^{(e)}_{f_*\partial_a} f_* \partial_b, \partial_k \right\rangle = \sum_{i,j} A^a_i A^b_j \left\langle \nabla^{(e)}_{\partial_i} \partial_j, \partial_k \right\rangle = 0,$$

para todos $a, b = 1, \ldots, n$ e $k = 1, \ldots, m$. Segue-se que M é e-autoparalela.

Agora, assuma que M é e-autoparalela. Então, a segunda forma fundamental

$$B(\partial_a, \partial_b) = (\nabla^{(e)}_{f_*\partial_a} f_* \partial_b)^N = 0.$$

Logo, dada uma curva $\alpha(t) \in M$ ligando dois pontos e um e-transporte paralelo $V(t)$ ao longo de α, tem-se que $f_*(V(t))$ é um e-transporte de S ao longo de $\beta(t) = f(\alpha(t))$. Como S é

e-plana, o e-transporte paralelo não depende da curva ligando os pontos extremos. Assim, o e-transporte paralelo sobre M também não depende da curva. Segue-se que M possui e-curvatura $R_M^{(e)} = 0$. Como a conexão $\nabla^{(e)}$ é simétrica, tem-se que M admite coordenadas e-afins $u = [u^a]$. Usando que ∂_i são e-paralelos, $f_*(\partial_a) = \sum_i \frac{\partial \xi^i}{\partial u^a}\partial_i$, e $\nabla^{(e)}_{f_*\partial_a} f_*\partial_b = \sum_c \Gamma^{(e)c}_{ab} f_*\partial_c + B(\partial_a, \partial_b) = 0$, segue-se que

$$0 \;=\; \nabla^{(e)}_{f_*\partial_a} f_*\partial_b = \nabla^{(e)}_{f_*\partial_a}(\sum_i \frac{\partial \xi^i}{\partial u^b}\partial_i) = \sum_i \frac{\partial^2 \xi^i}{\partial u^a \partial u^b}\partial_i + \frac{\partial \xi}{\partial u^b}\nabla^{(e)}_{f_*\partial u^a}\partial_i = \sum_i \frac{\partial^2 \xi^i}{\partial u^a \partial u^b}\partial_i.$$

Logo, $\frac{\partial^2 \xi^i}{\partial u^a \partial u^b} = 0$, para quaisquer i, a, b. Temos que $\xi^i(u) = \sum_a A^i_a u^a + b^i$, sendo A^i_a, b^i constantes. E como S é uma família exponencial, segue-se que

$$\begin{aligned} \ln \bar{q}_u(x) &= p(x;\xi(u)) = C(x) + \sum_i \xi^i(u) F_i(x) - \psi(\xi(u)) \\ &= C(x) + b^i F_i(x) + \sum_a (\sum_i A^i_a F_i(x)) u^a - \psi(\xi(u)). \end{aligned}$$

Donde, $M' = \{\bar{q}_u\}$ é uma família exponencial. Isto implica que $M = \{q_u\}$ é também uma família exponencial.

Definição 3.6 (Misturas). *Seja $S = \{P_\theta = p(x;\theta) \mid \xi = (\theta^1, \ldots, \theta^n) \in \Theta \subset \mathbb{R}^n\}$ um modelo estatístico regular sobre χ de dimensão n tal que existam funções $C, F_i : \chi \to \mathbb{R}$ com $i = 1.., n$ de modo que cada ponto de S é escrito na forma*

$$p(x;\theta) = C(x) + \sum_{i=1}^{n} \theta^i F_i(x). \tag{3.24}$$

Dizemos que S é uma família mistura. Neste caso chamamos $[\theta^i]$ de parametros mistura de S.

Exemplo 3.13. *O modelo finito $\wp(\chi)$ é uma família mistura. Para verificar isto, basta tomar*

$$F_j(x) = \begin{cases} 1 & \text{se } x = x_j \\ 0 & \text{se } x \neq x_j \end{cases}$$

$j = 0, 1, \ldots, n$ e ver que

$$p(x;\xi) = F_0(x) + \sum_{i=1}^{n} \xi^i (F_i(x) - F_0(x)).$$

Dado $n+1$ funções de distribuição de probabilidade $\{p_0, \ldots, p_n\}$ sobre um conjunto χ, sempre podemos construir uma família mistura S de dimensão n, pela seguinte maneira

$$\begin{aligned} p(x;\theta) &= \sum_{i=1}^{n} \theta^i p_i(x) + \left(1 - \sum_{i=1}^{n} \theta^i\right) p_0(x) \\ &= p_0(x) + \sum_{i=1}^{n} \theta^i \left(p_i(x) - p_0(x)\right). \end{aligned}$$

Onde $\theta^i > 0$ para todo $i = 1, \ldots, n$ e $\sum_{i=1}^{n} \theta^i < 1$. É claro que $p(x,\theta) \geq 0$ e que $\int p(x,\theta)dx = 1$.

Proposição 3.13. *Uma família mistura é m-plana*

Demonstração. Basta mostrar que $\Gamma_{ij,k}^{(-1)} = 0$. De fato, da definição de família mistura, segue que

$$\ell(x,\theta) = ln\, p(x,\theta) = ln\left(C(x) + \sum_{i=1}^{n} \theta^i F_i(x)\right).$$

Logo,

$$\partial_i \ell(x,\theta) \partial_j \ell(x,\theta) = \frac{F_i(x)F_j(x)}{p(x,\theta)^2} \quad e \quad \partial_i \partial_j \ell(x,\theta) = -\frac{F_i(x)F_j(x)}{p(x,\theta)^2}.$$

Assim temos que $\Gamma_{ij,k}^{(-1)} = 0$.

3.5 O Teorema de Cramér-Rao

Até aqui tratamos sobre modelos estatístico e de sua estrutura geométrica. Nesta seção final, falaremos do Teorema de Cramer-Rao, que dá uma cota inferior para a matriz de covariância de estimadores não-enviesados. Daremos uma prova geométrica deste teorema. Daremos também uma condição necessária e suficiente para que modelos estatísticos possuam estimadores eficientes, i.e., estimadores não-enviesados cuja variância atinge esta tal cota inferior.

Seja X uma variavél aleatória e suponha que sua função de distribuição de probabilidade pertença a um modelo estatístico resular $S = \{p_\xi\}$ de dimensão n. Consideremos o problema de estimar o parâmetro ξ (desconhecido por natureza). Definimos o estimador para o parâmetro ξ como sendo uma função $\widehat{\xi} : \chi \to \mathbb{R}^n$. Além disso, quando

$$E_\xi[\widehat{\xi}(X)] = \xi \quad \text{para todo} \quad \xi \in \Theta$$

diremos que $\widehat{\xi}$ é um *estimador não-enviesado*. Do contrário, diremos que $\hat{\xi}$ é um estimador enviesado. Por exemplo, considerando $S = \{P_{\mu,\sigma}(x) = p(x;\mu,\sigma)\}$ o modelo normal com parâmetros (μ,σ), o estimador $\widehat{\xi} : \chi \to \mathbb{R}^2$, definido por $\widehat{\xi}(x) = (x,x^2)$, é um estimador enviesado, pois $E_{P_{\mu,\sigma}}[\widehat{\xi}(x)] = E_{P_{\mu,\sigma}}[(x,x^2)] = (\mu,\mu^2+\sigma^2)$. Contudo, se supormos σ conhecido, o estimador $\widehat{\mu}(x) = x$ é não-enviesado. Para um estimador não-enviesado a matriz de covariância é dada por $V_\xi[\widehat{\xi}] = [v_\xi^{ij}]$, onde

$$v_\xi^{ij} = E_\xi[(\widehat{\xi}^i(X) - \xi^i)(\widehat{\xi}^j(X) - \xi^j)].$$

Teorema 3.14. *(Desigualdade de Cramér-Rao) Seja $S = \{p_\xi = p(\cdot\,;\xi)\}$ um modelo estatístico regular de dimensão n e $\widehat{\xi}$ um estimador não-enviesado do parametro ξ. Então, a matriz de variância e covariância $V_\xi[\widehat{\xi}]$ de $\widehat{\xi}$ satisfaz a desigualdade*

$$V_\xi[\widehat{\xi}] \succeq G(\xi)^{-1}, \tag{3.25}$$

onde $G(\xi)^{-1}$ é a inversa da matriz de informação de Fisher de S. A desigualdade acima é no sentido de matrizes, ou seja, a matriz $V_\xi[\widehat{\xi}] - G(\xi)^{-1}$ é positiva semidefinida. Além disso, a igualdade ocorre se, e somente se, S é uma família exponencial da forma:

$$p(x;\theta) = e^{C(x)+\theta_i\hat{\xi}^i - \psi(\theta)},$$

onde $\theta = \theta(\xi)$ é uma reparametrização de S. Aqui, $C = C(x), \psi = \psi(\theta)$ são funções reais.

Um estimador não-enviesado $\widehat{\xi}$ que alcança a igualdade $V_\xi[\widehat{\xi}] = G(\xi)^{-1}$ é chamado de estimador *eficiente*. A primeira parte do Teorema 3.25 diz que um estimador eficiente é o melhor estimador dentre os estimadores os não-enviesados, no sentido de possuir a menor variância possível. No entanto, um estimador eficiente pode não existir. A segunda parte do Teorema 3.25 diz que a existência de estimadores eficientes implica em fortes restrições tanto no modelo estatístico quanto na sua parametrização. Antes de iniciarmos a prova deste teorema (que na verdade será parte da demostração deste) vamos mostrar que famílias exponenciais admitem naturalmente estimadores eficientes. Considere $S = \{p_\theta\}$ uma família exponencial da forma

$$p(x;\theta) = e^{C(x)+\theta^i F_i(x)-\psi(\theta)}.$$

Segue-se da Proposição 3.11 que as coordenadas $\theta = [\theta^j]$ são e-afins e as coordenadas $\eta = [E_\theta[F_i]]$ são m-afins. Assim, a aplicação $\hat{\eta}(x) := [F_i(x)]$ é um estimador não-enviesado para η. Usando que $\partial_i(\ell) = F_i - \psi_i = F_i - E[F_i]$, segue-se que a matriz de covariância

$$V[F]_{ij} = E_\theta[(F_i - E[F_i])(F_j - E[F_j])] = E[\partial_i(\ell)\partial_j(\ell)] = g_{ij}(\theta) = g^{ij}(\eta).$$

Concluímos que $\hat{\eta}(x)$ é um estimador eficiente para η.

Provaremos agora o Teorema 3.14.

Demonstração. A condição de $\widehat{\xi} : \chi \to \mathbb{R}^n$ ser um estimador não-enviesado dá-nos que

$$\int (\widehat{\xi}^j(x) - \xi^j)p(x;\xi)dx = 0.$$

Assim, derivando em relação a ξ^i, tem-se

$$\int (\widehat{\xi}^j(x) - \xi^j)\partial_i p(x;\xi)dx = \delta_i^j = g_{ik}g^{kj}. \tag{3.26}$$

Multiplicando g^{li} em ambos lados, e somando em i, dá-nos que

$$\int (\widehat{\xi}^j(x) - \xi^j)\partial^l dx = g^{li}g_{ik}g^{kj} = g^{lj}, \tag{3.27}$$

onde $\partial^l = g^{li}\partial_i p(\cdot;\xi)$.

Fixe $v = (v_1,\dots,v_n) \in \mathbb{R}^n$, $v \neq 0$, arbitrário e consideremos $v^j = v_i g^{ij}$, $X = v_i\partial^i = v^j\partial_j$, $A = \xi^j v_j$ e $\widehat{A} = \widehat{\xi}^j v_j$. Assim, a métrica g de S satisfaz:

$$g(X,X) = g(v_i g^{ik}\partial_k, v_j g^{jl}\partial_l) = v_i g^{ik} v_j g^{jl} g_{kl} = v_i v_j g^{ij}. \tag{3.28}$$

Por (3.27), tem-se

$$g(X,X) = v_j v_l g^{jl} = \int (\widehat{A}(x) - A)X(p)dx. \tag{3.29}$$

Usando que $g(X,Y) = \int \frac{X(p)}{p}\frac{Y(p)}{p}p$, para todo $X,Y \in T_pS$, temos que

$$\begin{aligned} g(X,X) &= \int (\widehat{A}(x) - A)X(p)dx = \int (\widehat{A}(x) - A)p^{\frac{1}{2}}\frac{X(p)}{p}p^{\frac{1}{2}}dx && (3.30)\\ &\leq [\int (\widehat{A}(x) - A)^2 p\,dx]^{1/2}[\int (\frac{X(p)}{p})^2 p\,dx]^{1/2} && (3.31)\\ &= V[\hat{A}]^{1/2}g(X,X)^{1/2}. \end{aligned}$$

Donde, $g(X,X)^{1/2} \leq V[\hat{A}]^{1/2}$. Logo, por (3.28), temos $v_i v_j g^{ij} = g(X,X) \leq V[\hat{A}] = v_i v_j v^{ij}[\hat{\xi}]$. Assim, como v é arbitrário, a matriz $(v^{ij}[\hat{\xi}] - g^{ij})$ é positiva semidefinida. Aqui, $v^{ij}[\hat{\xi}] = cov[\hat{\xi}^i, \hat{\xi}^j] = E[(\hat{\xi}^i - \xi^i)(\hat{\xi}^j - \xi^j)]$ denota a matriz de covariância de ξ. A primeira parte do teorema está provada.

Suponhamos que ocorra a igualdade em (3.25). Então vale a igualdade em (3.31), o que implica (usando o caso da igualdade na desigualdade de Cauchy-Schwarz) que $(\hat{A} - A)p^{1/2} = \lambda \frac{X(p)}{p} p^{1/2} = \lambda \frac{X(p)}{p^{1/2}}$, para algum $\lambda = \lambda(\xi)$. Assim, por (3.30),

$$\begin{aligned} g(X,X) &= \int (\hat{A}(x) - A)p^{1/2} \frac{X(p)}{p^{1/2}} dx = \int \lambda \frac{X(p)}{p^{1/2}} \frac{X(p)}{p^{1/2}} dx \\ &= \lambda \int \frac{X(p)}{p} \frac{X(p)}{p} p\, dx = \lambda\, g(X,X), \end{aligned}$$

donde $\lambda = 1$. Segue-se então que $X(\ell) = \frac{X(p)}{p} = \hat{A} - A$, onde $\ell = \ln p$. Pela arbitrariedade de v, temos

$$\partial^i(\ell) = \hat{\xi}^i - \xi^i, \tag{3.32}$$

donde

$$\begin{aligned} &\partial_k(\ell) = g_{ks}\partial^s(\ell) = g_{ks}(\hat{\xi}^s - \xi^s) \quad \text{e} \\ &\partial_i\partial_j(\ell) = (\partial_i g_{jm})(\hat{\xi}^m - \xi^m) + g_{jm}(-\delta_i^m) = (\partial_i g_{jm})(\hat{\xi}^m - \xi^m) - g_{ij}. \end{aligned} \tag{3.33}$$

Afirmamos que as coordenadas ξ são m-planas. De fato,

$$\begin{aligned} [\partial_i\partial_j(\ell) + \partial_i(\ell)\partial_j(\ell)]\partial_k(\ell) &= (\partial_i g_{jm})g_{ks}(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s) - g_{ij}g_{ks}(\hat{\xi}^s - \xi^s) \\ &+ g_{il}\, g_{jm}\, g_{ks}(\hat{\xi}^l - \xi^l)(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s). \end{aligned} \tag{3.34}$$

Agora, como ξ é um estimador eficiente, temos

$$E_\xi[(\partial_i g_{jm})g_{ks}(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)] = (\partial_i g_{jm})g_{ks}v^{ms}[\hat{\xi}] = (\partial_i g_{jm})g_{ks}g^{ms} = \partial_i g_{jk} \tag{3.35}$$

$$E_\xi[g_{ij}g_{ks}(\hat{\xi}^s - \xi^s)] = g_{ij}g_{ks}E[\hat{\xi}^s - \xi^s] = 0. \tag{3.36}$$

Usando (3.33), temos $g_{jk} = E_\xi[\partial_j(\ell)\partial_k(\ell)] = g_{jm}g_{ks}E_\xi[(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)] = g_{jm}g_{ks}v^{ms}$. Como $\hat{\xi}$ é um estimador eficiente, $g_{ks}v^{ms} = g_{ks}g^{ms} = \delta_k^m$, donde

$$\begin{aligned} \partial_i g_{jk} &= (\partial_i g_{jm})g_{ks}v^{ms} + g_{jm}(\partial_i g_{ks})v^{ms} + g_{jm}g_{ks}(\partial_i v^{ms}) \\ &= \partial_i g_{jk} + \partial_i g_{kj} + g_{jm}g_{ks}(\partial_i v^{ms}) \\ &= 2\partial_i g_{jk} + g_{jm}g_{ks}\,\partial_i \int (\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)p\, dx \\ &= 2\partial_i g_{jk} - g_{jm}g_{ks} \int \delta_i^m(\hat{\xi}^s - \xi^s)p\, dx - g_{jm}g_{ks} \int (\hat{\xi}^m - \xi^m)\delta_i^s\, p\, dx \\ &\quad + g_{jm}g_{ks} \int (\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)\partial_i(\ell)p\, dx \\ &= 2\partial_i g_{jk} + g_{jm}g_{ks} \int (\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)g_{il}(\hat{\xi}^l - \xi^l)\, p\, dx. \end{aligned}$$

Assim,

$$\partial_i g_{jk} = -g_{il}g_{jm}g_{ks}E_\xi[(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)(\hat{\xi}^l - \xi^l)]. \tag{3.37}$$

Usando as equações (3.34) a (3.37), segue-se que

$$\begin{aligned}
\Gamma^{(m)}_{ij,k} &= E_\xi[(\partial_i\partial_j(\ell) + \partial_i(\ell)\partial_j(\ell))\partial_k(\ell)] \\
&= \partial_i g_{jk} + g_{il}g_{jm}g_{ks}E_\xi[(\hat{\xi}^l - \xi^l)(\hat{\xi}^m - \xi^m)(\hat{\xi}^s - \xi^s)] \\
&= \partial_i g_{jk} - \partial_i g_{jk} = 0,
\end{aligned}$$

Portanto, as coordenadas são m-planas.

Pelo Teorema 2.14, existem coordenadas e-afins $\theta = [\theta_i]$ e funções convexas $\psi = \psi(\theta)$ e $\varphi = \varphi(\xi)$ satisfazendo $\partial^i = \frac{\partial}{\partial \theta_i}$, $\nabla\psi(\theta) = \xi$, $\nabla\varphi(\xi) = \theta$, $\mathrm{Hess}_\psi(\theta) = G$ e $\mathrm{Hess}_\varphi(\xi) = G^{-1}$, onde $G = (g_{ij}(\xi))$ denota a matriz de informação de Fisher. Considere a seguinte função:

$$q(x;\theta) = e^{\theta_i\hat{\xi}^i - \psi(\theta)}.$$

Segue-se que $\partial^i(\ln q) = \frac{\partial \ln q}{\partial \theta_i} = \hat{\xi}^i - \psi_i(\theta) = \hat{\xi}^i - \xi^i = \partial^i(\ell) = \partial^i(\ln p)$. Assim,

$$\ell(x;\xi) = \ln q(x;\theta) + C(x) = C(x) + \theta_i\hat{\xi}^i - \psi(\theta).$$

Segue-se que p é uma família exponencial. O teorema está provado. □

3.6 Observações Independentes

Seja $S = \{P_\xi\}$ um modelo estatístico sobre χ. É comum na inferência estatística definir um estimador para o parâmetro ξ como uma função definida sobre observações $x_1, \ldots, x_N$, com $x_j \in \chi$, independentes e igualmente distribuidas (iid). Provaremos o seguinte:

Teorema 3.15. *Usando a notação acima, considere $\widehat{\xi}_N : \chi^N = \chi \times \ldots \times \chi \to \mathbb{R}^n$ um estimador não-enviesado do parâmetro ξ com respeito a distribuição conjunta*

$$p_N(x;\xi) = p_\xi(x_1)\ldots p_\xi(x_N),$$

com $x = (x_1, \ldots, x_N)$, i.e., $E_{p_N}[\widehat{\xi}_N] = \xi$. A desigualdade de Cramer-Rao toma a seguinte forma:

$$V_\xi[\widehat{\xi}_N] \succeq \frac{1}{N}G(\xi)^{-1}. \tag{3.38}$$

Onde $V_\xi[\widehat{\xi}_N]$ é a matriz de covariância em relação a distribuição $p_N(x;\xi)$ e $G(\xi)^{-1}$ é a inversa da matriz de informação de Fisher de $S = \{p_\xi\}$. Além disso, a igualdade ocorre se, e somente, se S é uma família exponencial.

De fato, primeiro observe que $S_N = \{p_N(x;\xi)\}$ é um modelo estatístico sobre χ^N. Supondo que $\widehat{\xi}_N$ é não-enviesado então, pela desigualdade de Cramér-Rao (Teorema 3.14), vale que

$$V_\xi[\widehat{\xi}_N] \succeq G_N(\xi)^{-1}, \tag{3.39}$$

onde $G_N(\xi)^{-1}$ é a inversa da matriz de informação de Fisher de S_N. Como $x_1,\ldots,x_n \in \chi$ são observações iid, $p_N(x;\xi) = p_\xi(x_1)\ldots p_\xi(x_N)$, segue-se que

$$\ell_N(x,\xi) := \ln p_N(x;\xi) = \sum_{t=1}^{N} \ln p(x_t;\xi). \tag{3.40}$$

Assim, pelo Lema 3.3,

$$\begin{aligned} g_{ij}^N(\xi) &= -E[\partial_i\partial_j \ell_N(x;\xi)] = -E[\partial_i\partial_j \sum_{t=1}^{N} \ln p(x_t;\xi)] = -\sum_{t=1}^{N} E[\partial_i\partial_j \ln p(x_t;\xi)] \\ &= N g_{ij}(\xi), \end{aligned} \tag{3.41}$$

isto é, $G_N(\xi) = NG(\xi)$, em particular, $G_N(\xi)^{-1} = \frac{1}{N}G(\xi)^{-1}$. Por (3.39), segue que

$$V_\xi[\widehat{\xi}_N] \succeq \frac{1}{N}G(\xi)^{-1}.$$

A primeira parte do teorema está provada.

Agora, vamos assumir que a igualdade acima ocorre, para todo ξ. Do Teorema 3.14 segue-se que $S_N = \{p_N(\cdot\,;\xi)\}$ é uma família exponencial da forma:

$$\ln p_N(x;\theta) = C(x) + \theta_i \hat{\xi}_N^i(x) - \psi(\theta), \tag{3.42}$$

com $x = [x_t] \in \chi^N$ e $\xi = E_{p_N}[\hat{\xi}_N(x)]$ sendo coordenadas m-planas em S_N. Aqui, $\theta = [\theta_i]$ são coordenadas e-planas, satisfazendo $\psi_i(\theta) = E_{p_N}[\hat{\xi}_N^i] = \xi^i$, para todo i. Vamos provar que S é também uma família exponencial. De fato, usando (3.40) e (3.42), segue-se que

$$\sum_t \frac{\partial}{\partial\theta^i} \ln p(x_t;\xi) = \frac{\partial}{\partial\theta^i} \ln p_N(x;\theta) = \hat{\xi}_N^i(x) - \psi_i(\theta).$$

Integrando ambos lados sobre $\bar{x} = [x_2,\ldots,x_N] \in \chi^{N-1}$ e como

$$\int_{\chi^{N-1}} \frac{\partial}{\partial\theta^i} \ln p(x_t,\xi) p_{N-1}(\bar{x}) d\bar{x} = \int_{\chi^{N-2}} \int_{\chi} \frac{\partial}{\partial\theta^i} \ln p(x_t,\xi) p(x_t) dx_t d\tilde{x} = 0,$$

para todo $t = 2,\ldots,N$, onde $\tilde{x} = [x_2,\ldots,x_{t-1},x_{t+1},\ldots,x_N]$, temos

$$\frac{\partial}{\partial\theta^i} \ln p(x_1;\xi) = \hat{\xi}^i(x_1) - \psi_i(\theta) = \hat{\xi}^i(x_1) - \xi^i,$$

onde $\hat{\xi}(x_1) = \int_{\chi^{N-1}} \hat{\xi}_N^i(x) d\bar{x}$, com $x = [x_1,\ldots,x_N] = [x_1,\bar{x}] \in \chi^N$. Assim, $E_p[\hat{\xi}^i(x_1)] = \xi^i$, para todo i. Isso implica que $\hat{\xi}(x_1) = [\hat{\xi}^i(x_1)]$, com $x_1 \in \chi$, é um estimador não-enviesado de ξ em S. Assim, usando a desigualdade de Cramer-Rao,

$$V_p^{ij}[\hat{\xi}(x_1)] \succeq g^{ij}(\xi)$$

Com os mesmos argumentos acima, com x_t ao invés de x_1, podemos definir os estimadores $\hat{\xi}(x_t)$, com $t \in \{1,\ldots,N\}$. Por outro lado, a matriz de variância $V_\xi[\hat{\xi}_N]$ é dada por:

$$\begin{aligned}
V_\xi^{ij}[\hat{\xi}_N] &= E_{p_N}[(\hat{\xi}_N^i - \xi^i)(\hat{\xi}_N^j - \xi^j)] = E_{p_N}[(\frac{\partial}{\partial\theta^i}\ln p_N(x;\theta))(\frac{\partial}{\partial\theta^j}\ln p_N(x;\theta))] \\
&= \sum_{s,t} E_{p_N}[(\frac{\partial}{\partial\theta^i}\ln p_N(x_s;\theta))(\frac{\partial}{\partial\theta^j}\ln p_N(x_t;\theta))] \\
&= \sum_t E_p[(\frac{\partial}{\partial\theta^i}\ln p_N(x_t;\theta))(\frac{\partial}{\partial\theta^j}\ln p_N(x_t;\theta))] \qquad (3.43) \\
&= \sum_t E_p[(\hat{\xi}^i(x_t) - \xi^i)(\hat{\xi}^j(x_t) - \xi^j)] \\
&= NV_p^{ij}[\hat{\xi}(x_1)].
\end{aligned}$$

A igualdade (3.43) é obtida a partir da independência de $x_1,\ldots,x_N$.

Fixemos $(v_1,\ldots,v_N) \in \mathbb{R}^N$ arbitrário e considere $X = v_i\partial^i \in T_pS$ e $X_N = v_i\partial_N^i \in T_{p_N}S_N$. Tal como fizemos em (3.41), temos que

$$\begin{aligned}
v_iv_jg_N^{ij}(\xi) &= g_N(X_N,X_N) = E_{p_N(\cdot;\xi)}[X(\ln p_N(x;\xi))X(\ln p_N(x;\xi))] \\
&= \sum_{s,t} E_{p_\xi}[X(\ln p_N(x_t;\xi))X(\ln p_N(x_s;\xi))] \\
&= \sum_t E_{p_\xi}[X(\ln p_N(x_t;\xi))X(\ln p_N(x_t;\xi))] \\
&= NE_{p_\xi}[X(\ln p_\xi(x_1))X(\ln p_\xi(x_1))] = Ng(X,X) = Nv_iv_jg^{ij}(\xi).
\end{aligned}$$

Usando que $\hat{\xi}_N(x)$ é um estimador eficiente, segue-se que

$$Nv_iv_jV_p^{ij}[\hat{\xi}(x_1)] = v_iv_jV_\xi^{ij}[\hat{\xi}_N] = v_iv_jg_N^{ij}(\xi) = g_N(X,X) = Ng(X,X) = Nv_iv_jg^{ij}(\xi).$$

Logo, $V_p^{ij}[\hat{\xi}(x_1)] = g^{ij}(\xi)$, donde $\hat{\xi}$ é um estimador eficiente. Portanto, do Teorema 3.14, segue-se que $S = \{p_\xi\}$ é uma família exponencial. Reciprocamente, é claro que se S é uma família exponencial, digamos da forma

$$p(z,\xi) = e^{C(z)+\xi^i\hat{\xi}_i(z)-\psi(\xi)},$$

com $z \in \chi$, então $S_N = \{p_N(x;\xi)\}$ é uma família exponencial, da forma

$$p_N(x,\xi) = e^{\bar{C}(x)+\xi^i(\hat{\xi}_N)_i(x)-N\psi(\theta)},$$

com $x = (x_1,\ldots,x_N) \in \chi^N$, $\bar{C}(x) = C(x_1)+\ldots+C(x_N)$ e $\hat{\xi}_N(x) = \hat{\xi}(x_1)+\ldots+\hat{\xi}(x_N)$. E como $\hat{\xi}(z)$ é um estimador eficiente de ξ em S, segue-se diretamente que $\hat{\xi}_N(x)$ é um estimador eficiente em S_N. O teorema está provado. □

3.7 A geometria de $\wp(\chi)$, com χ finito.

Nesta ultima seção, apresentaremos fatos geométricos importantes sobre a geometria de $\wp(\chi)$. Como já vimos, o espaço das distribuições de probabilidades sobre χ, $\wp(\chi)$, com χ finito, é um modelo estatístico. A dificuldade de se estudar $\wp(\chi)$ no caso em que χ é infinito é que, neste caso, $\wp(\chi)$ passa a ser uma variedade diferenciável de dimensão infinita. Como não tratamos desse caso nos preliminares, considemos apenas o modelo finito $\wp(\chi)$, onde $\chi = \{0,\ldots,n\}$. Para simplificar a notação, iremos denotar $\wp(\chi)$ apenas por $\wp$.

Como já vimos, $\wp$ além de ser uma família exponencial também é um família mistura. Neste ultimo caso, o sistema de coordenadas (m-afim) para $\wp$ é dada pelos parâmetros $\xi = [\xi^i]$, com $\xi^i = p_i = p(i)$, com $i = 1,\ldots,n$, donde $p_0 = p(0) = 1 - \sum_k \xi^k$. Assim,

$$p_\xi = (1 - \sum_{i=1}^{n} \xi^i)\delta_0 + \sum_{i=1}^{n} \xi^i \delta_i,$$

onde $\delta_i(j) = \delta_{ij}$. Para simplificar a notação, denote $\wp(\chi)$ simplesmente por $\wp$.

Seja V_0 o espaço vetorial definido por

$$V_0 = \{A \in \mathbb{R}^\chi \mid \sum_{x\in\chi} A(x) = 0\},$$

onde $\mathbb{R}^\chi := \{A : \chi \to \mathbb{R}\}$. É simples verificar que V_0 tem dimensão n. Portanto, fixado $p_\xi \in \wp$, e usando que $\sum_x \partial_i p_\xi = 0$, é imediato verificar que a aplicação definida por $X \in T_p\wp \mapsto X^{(m)} := X(p_\xi)$ é um isomorfismo linear entre $T_p\wp$ e V_0. Chamaremos $X^{(m)}$ de m-representação de X e escrevemos

$$T_p^{(m)}\wp := \{X^{(m)} = X(p) \mid X \in T_p\wp\} = V_0.$$

Observamos que $T_p^{(m)}\wp = T_q^{(m)}\wp = V_0$, para todo $p,q \in \wp$ e que $(\partial_i)_\xi^{(m)} = \partial_i p_\xi = \delta_i - \delta_0 \in V_0$.

Considere $\Pi_{p,q}^{(m)}$ o transporte paralelo de $T_p\wp$ para $T_q\wp$, com respeito a m-conexão $\nabla^{(m)}$. Temos que

$$\Pi_{p,q}^{(m)}(X) = X' \Leftrightarrow X'^{(m)} = X^{(m)}.$$

De fato, se $V(t) = V^i(t)\partial_i$, $t \in [0,1]$, é um transporte paralelo ao longo de uma curva qualquer ligando p a q, então, usando que $\Gamma_{ij,k}^{(m)} = 0$, segue-se que V^i é constante, donde, $V(t) = V^i\partial_i$. Como $\partial_i^{(m)} = (\delta_i - \delta_0)$, temos $V^{(m)}(t) = V^i(\delta_i - \delta_0) = X^{(m)}$, para todo t. Logo, $X^{(m)} = X'^{(m)}$.

Faremos outra identificação de $T_p\wp$. Considere V_1 o espaço vetorial definido por

$$V_1 = \{A \in \mathbb{R}^\chi \mid \sum_{x\in\chi} A(x)p(x) = E_p[A] = 0\} = \{B - E_p[B] \mid B \in \mathbb{R}^\chi\}.$$

É claro que se $E_p[A] = 0$ então $A = A - E_p[A]$. Por outro lado, $E_p[A - E_p[A]] = 0$, o que justifica a igualdade dos conjuntos acima. Além disso, vê-se facilmente que V_1 é n-dimensional.

Observe que a aplicação linear

$$X \in T_p\wp \mapsto X^{(e)} = X(\ln p_\xi) = \frac{1}{p}X(p_\xi) = \frac{1}{p}X^{(m)} \in V_1$$

é um isomorfismo. Chamaremos $X^{(e)}$ de e-representação de X e $T_p^{(e)}\wp = \{X^{(e)} \mid X \in T_p\wp\}$. Em particular, $X^{(m)} \mapsto X^{(e)} = \frac{1}{p}X^{(m)}$ é um isomorfismo linear entre $T_p^{(m)}\wp = V_0$ e $T_p^{(e)}\wp = V_1$. A relação entre a métrica de Fischer e a e-representação é dada por

$$\langle X, Y\rangle_p = \int \frac{X(p)}{p}\frac{Y(p)}{p}\, p\, dx = E_p[X^{(e)}Y^{(e)}] \tag{3.44}$$

onde $\langle , \rangle$ é a métrica de Fisher.

Ao contrário do que acontece com $T_p^{(m)}\wp$, o espaço vetorial $T_p^{(e)}\wp$ depende de p, isto é, um elemento $A \in T_p^{(e)}\wp$ pode não pertencer a $T_q^{(e)}\wp$, com $q \neq p$. No entanto, $A' = A - E_q[A]$ pertence a $T_q^{(e)}(\wp)$, visto que $E_q[A'] = 0$. Segue-se que $A \in T_p^{(e)}\wp \mapsto A - E_q[A] \in T_q^{(e)}\wp$ é um isomorfismo linear entre $T_p^{(e)}(\wp)$ e $T_q^{(e)}(\wp)$.

Dados $X : p \mapsto X_p$ um campo de vetores arbitrário sobre $\wp$ e ∂_i os campos de vetores do sistema de coordenadas $[\xi^i]$, segue da definição da e-conexão (equação (3.16) para $\alpha = 1$) que

$$\langle \nabla^{(e)}_{\partial_i} X, \partial_k\rangle_p = E_p[\partial_i X(\ell_\xi)\partial_k \ell_\xi] = E_p[\partial_i X_p^{(e)} \partial_k^{(e)}]. \tag{3.45}$$

Vamos provar que o e-transporte paralelo, $\Pi^{(e)}_{p,q}(X) = X'$ é dado por

$$X'^{(e)} = X^{(e)} - E_q[X^{(e)}].$$

De fato, fixado p, considere o vetor $X \in T_p\wp$. Seja $A \in R^\chi$ tal que a e-representação de X seja dada por $X_p^{(e)} = A - E_p[A]$. Considere Y o campo de vetores em $\wp$, dado por $Y_q^{(e)} = A - E_q[A]$, para todo $q \in \wp$. Tomando a derivada em relação a ξ^i, temos que $\partial_i Y^{(e)} = -\partial_i E_p[A]$, não depende de x. Por (3.45), segue-se que

$$\left\langle \nabla^{(e)}_{\partial_i} Y, \partial_k \right\rangle_q = E_p[\partial_i Y_q^{(e)} \partial_k^{(e)}] = \partial_i Y_q^{(e)} E_q[\partial_k^{(e)}] = 0,$$

donde Y é um campo e-paralelo. Como $Y_p = X_p$, pela unicidade dos campos paralelos, segue-se que $X' = Y_q$. Assim, $X'^{(e)} = Y_q^{(e)} = A - E_q[A] = A - E_p[A] - (E_q[A] - E_p[A]) = X_p^{(e)} - E_q[X_p^{(e)}]$, como queriamos.

Definição 3.7. *Dada uma variedade Riemanniana (S, g) e uma função diferenciável $f : S \to \mathbb{R}$, defina o gradiente $(\nabla f)_p \in T_pS$ dado por $\langle (\nabla f)_p, v\rangle = v(f)$, para todo vetor $v \in T_pS$.*

Da definição acima, dada uma curva $\alpha = \alpha(t)$ em S, com $\alpha(0) = p$ e $|\alpha'(0)| = 1$, segue-se que $\|\nabla f\| \geq \langle (\nabla f)_p, \alpha'(0) \rangle = \frac{d}{dt}|_{t=0}(f \circ \alpha)(t)$. Logo, assim como no caso Euclideano, o vetor gradiente de f aponta na direção onde f mais cresce.

Agora, fixemos uma função $A : \chi \to \mathbb{R}$ e considere a função $E[A] : p \in \wp \mapsto E_p[A]$. Como observado, o vetor gradiente $\nabla E[A]$ aponta na direção em que a média $E[A]$ mais cresce. Assim, considerando g como sendo a métrica de Fisher, a norma $\|\nabla E[A]\|_p^2 = g(\nabla E[A], \nabla E[A])_p$, daria uma forma natural de se medir a dispersão de A. No entanto, provaremos que esta medida de dispersão coincide com a variância $V_p[A] = E_p[(A - E_p[A])^2]$, para todo $p \in \wp$.

Teorema 3.16. *Dada uma função $A : \chi \to \mathbb{R}$, a função $E[A] : p \in \chi \mapsto E_p[A]$ satisfaz*

$$\|\nabla E[A]\|_p^2 = V_p[A],$$

para todo $p \in \wp$. Aqui, denotamos $\|X\|^2 = g(X, X)$, sendo g a métrica de Fisher de $\wp$.

Demonstração. Dado $X \in T_p\wp$ temos que

$$\begin{aligned} g(X, \nabla E[A])_p &= X(E[A]) = X(\sum_{x \in X} p(x,\xi)A(x)) = \sum_{x \in X} X(p(x,\xi))A(x) \\ &= \sum_{x \in X} X^{(m)}A(x) = E_p[X^{(e)}A] = E_p[X^{(e)}(A - E_p[A])]. \end{aligned} \tag{3.46}$$

Para a última igualdade, usamos que $E_p[X^{(e)}E_p[A]] = E_p[A]E_p[X^{(e)}] = 0$.

Como $(A - E_p[A]) \in V_1 = T_p^{(e)}\wp$, existe $Y_p \in T_p\wp$ tal que $Y_p^{(e)} = A - E_p[A]$. Assim, de (3.46), segue-se que $g(X, E[A])_p = E_p[X^{(e)}Y^{(e)}] = g(X,Y)_p$, para todo $X \in T_pS$. Logo,

$$(\nabla E[A])_p^{(e)} = A - E_p[A], \tag{3.47}$$

para todo $p \in \wp$. Portanto, $\|\nabla E[A]\|_p^2 = E_p[(A - E_p[A])^2] = V_p[A]$. □

Seja $S = \{p_\xi\}$ um modelo estatístico e $A \in \mathbb{R}^\chi$. O gradiente da restrição $E[A]|_S : S \to \mathbb{R}$, no ponto p, é dada pela projeção ortogonal $\nabla(E[A]|_S) = (\nabla E[A])^T$ sobre o plano tangente T_pS. Do Teorema 3.16, segue-se

Corolário 3.17. *Seja S uma subvariedade de $\wp$, a restrição $E[A]|_S : S \to \mathbb{R}$ satisfaz*

$$\|\nabla E[A]|_S\|_p^2 \leq V_p[A],$$

e vale a igualdade se e somente, se $A - E_p[A] \in T_p^{(e)}S = \{X^{(e)} \mid X \in T_p(S)\}$.

Referências Bibliográficas

[1] AMARI, Shun-Ichi. *Information geometry and its Applications*, Applied Mathematics Sciences, Springer, 2016.

[2] AMARI, Shun-Ichi e NAGAOKA, Hiroshi. *Methods of Information Geometry*. Translations of Mathematical Monographs, Vol.191, Am. Math. Soc., 2000.

[3] ARWINI, Khadiga A. e DODSON, Christopher T.J.. *Information Geometry - Near Randomness and Near Independence*. Lecture Notes in Mathematics, 2008.

[4] BIEZUNER, Rodney Josué. Notas de aula: Geometria Riemanniana. Disponível online.

[5] do CARMO, Manfredo Perdigão. Geometria Riemanniana. 2011, 5^a edição, Projeto Euclides, 2011.

[6] CALIN, Ovidiu e UDRISTE, Constantin. Geometric Modeling in Probability and Statistics, Springer, 2014.

[7] SHERN, S.S., SHEN, H. W. e LAM, K. S., *Lectures on Differential Geometry*. Series on University Mathematics, Vol. 1, 1999.

[8] WASSERMAN, Larry. *All of Statistics: A Concise Couse in Statistical Inference*. Springer Text in Statistics, Springer 2004.