Variedades Estatísticas e Teoria da Informação

Gustavo Mezzovilla

18 de dezembro de 2022

Sumário

1 *Review* expresso de variedades

Antes de definir nosso principal objeto de estudo, temos que entender os conceitos da Geometria Diferencial. Aqui obviamente só temos o mínimo consolidado de modo que este seja o mais auto contido possível, porém, uma ótima referência para se nortear esta contemplada nas Notas de Aula de Geometria Riemanniana Alexandrino (2022) ou também em Munkres (2018).

1.1 Variedades suaves

Como base, temos que definir os espaços com os quais iremos trabalhar:

Definição 1.1 (Variedade Topológica)**.** Um espaço topológico M é uma *variedade topológica de dimensão n* se as seguintes condições forem satisfeitas:

1. M é Hausdorff;

2. M é segundo contável, ou seja, possui uma base enumerável;

3. M é localmente euclidiano de dimensão n, isto é, todo ponto de M possui uma vizinhança homeomorfa a um aberto de $\mathbb{R}^n$.

No item 3 acima, se U é a vizinhança em questão e $\phi : U \longrightarrow \mathbb{R}^n$ é um homeomorfismo entre U e um aberto $\phi(U) \subseteq \mathbb{R}^n$, chamamos o par (U, ϕ) de *carta coordenada*. No nosso caso, a variedade topológica representa intuitivamente o que vemos como curvas e superfícies em $\mathbb{R}^3$, mas precisaremos muni-la com alguma estrutura diferenciável para podermos trabalhar melhor. Transportaremos o cálculo diferencial de $\mathbb{R}^n$ para a variedade através das cartas coordenadas.

Definição 1.2 (Mudança de Coordenadas)**.** Duas cartas $(U, \varphi : U \longrightarrow \mathbb{R}^n)$, $(V, \psi : V \longrightarrow \mathbb{R}^n)$ de uma variedade topológica são C^∞*-compatíveis* se os *mapas de transição*

$$\varphi \circ \psi^{-1} : \psi(U \cap V) \longrightarrow \varphi(U \cap V), \quad \psi \circ \varphi^{-1} : \varphi(U \cap V) \longrightarrow \psi(U \cap V)$$

são C^∞. Esses também podem ser chamados de *mudanças de coordenadas*.

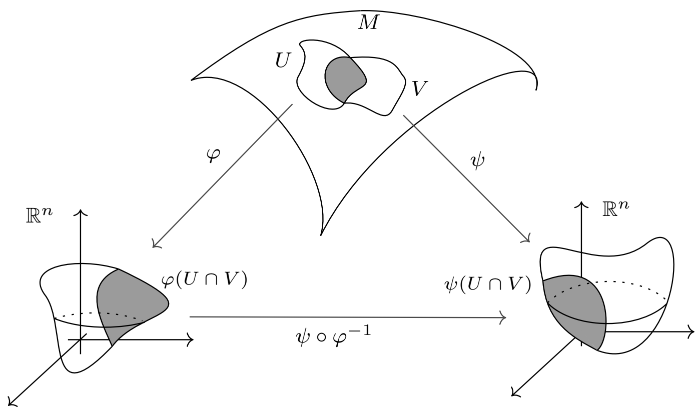

Figura 1: Ilustração de uma mudança de coordenadas

Veja que os mapas de transição acima são homeomorfismos entre abertos de $\mathbb{R}^n$, logo faz sentido verificar se eles são suaves. Passando essa definição ao global:

Definição 1.3 (Atlas)**.** Um *atlas* C^∞ numa variedade topológica M é uma coleção $\{(U_\alpha, \phi_\alpha) \mid \alpha \in \mathbb{A}\}$ de cartas coordenadas duas a duas C^∞-compatíveis que cobrem M, ou seja, $M = \bigcup_{\alpha \in \mathbb{A}} U_\alpha$.

Estamos quase lá! Poderíamos dizer que uma variedade suave é uma variedade topológica que possui um atlas C^∞. O problema é que quando dois atlas diferentes são compatíveis entre si, eles essencialmente carregam a mesma informação. Dessa forma, queremos unir todos os atlas que são "equivalentes". Com essa ideia, não é difícil provar que todo atlas está contido em um único atlas C^∞ *maximal*, ou seja, que não está propriamente contido em nenhum outro atlas. Agora sim estamos prontos!

Definição 1.4 (Variedade Suave)**.** Uma *variedade suave* é uma variedade topológica juntamente com um atlas C^∞ maximal.

Boa parte das propriedades de diferenciabilidade se traduzem localmente para o contexto de variedades suaves através das cartas coordenadas. Por exemplo, podemos definir funções suaves entre variedades. Diremos que uma função $F : N \longrightarrow M$ entre variedades suaves é C^∞ em um ponto $p \in N$ se existem cartas coordenadas (U, ϕ) sobre p e (V, ψ) sobre $F(p)$ tais que $\psi \circ F \circ \phi^{-1}$ é uma função C^∞ em $\phi(p)$. Não é difícil mostrar que, para outras cartas sobre p e $F(p)$, a função induzida entre abertos de $\mathbb{R}^n$ e $\mathbb{R}^m$ ainda é suave no ponto em questão. Assim, nossa definição faz sentido. Nesse caso, F será suave se for C^∞ em todo ponto.

Por fim, um conceito muito importante nessa linha é o de difeomorfismo. Uma função $F : N \longrightarrow M$ entre variedades suaves é um *difeomorfismo* se for suave, bijetora e tiver inversa suave. Duas variedades *difeomorfas* são essencialmente a mesma no quesito da diferenciabilidade.

1.2 O Espaço Tangente

Para definir formas e uma teoria de integração em variedades, é imprescindível definir o espaço tangente e discutir suas propriedades. No caso mais visual de variedades mergulhadas em $\mathbb{R}^2$ ou $\mathbb{R}^3$, intuitivamente vemos o espaço tangente em um ponto p da variedade como sendo o conjunto dos vetores que são tangentes a curvas suaves que passam por p. Porém, como estamos adotando uma definição mais abstrata e geral, não é tão claro qual seria uma boa definição. O que se segue é uma abordagem diferente, que equivale à definição mais visual, nos casos onde ambas se aplicam.

Definimos um *germe* de uma função C^∞ em um ponto p de uma variedade M como a classe de equivalência de funções C^∞ definidas numa vizinhança de p em M, sendo duas funções equivalentes se elas concordam numa vizinhança (possivelmente menor) de p. O conjunto desses germes em p será denotado por $C_p^\infty(M)$. Note que $C_p^\infty(M)$ é um espaço vetorial. Chamaremos de *derivação em um ponto p* numa variedade M um mapa linear $D : C_p^\infty(M) \longrightarrow \mathbb{R}$ que satisfaz a "Regra de Leibniz":

$$D(fg) = (Df)g(p) + f(p)Dg.$$

Definição 1.5 (Espaço Tangente)**.** O *espaço tangente de M em um ponto p* é o espaço vetorial de todas as derivações em p, denotado por T_pM.

Exemplo 1.6. Seja (U, ϕ) uma carta coordenada sobre um ponto p numa variedade suave M. Se $r_1, \dots, r_n$ são as coordenadas padrões em $\mathbb{R}^n$ (i.e., as projeções nos vetores da base canônica), tomamos $x_i = r_i \circ \phi : U \longrightarrow \mathbb{R}$.

Se f é uma função suave numa vizinhança de p, definimos

$$\left.\frac{\partial}{\partial x_i}\right|_p f \coloneqq \left.\frac{\partial}{\partial r_i}\right|_{\phi(p)} (f \circ \phi^{-1}) \in \mathbb{R}\,.$$

A expressão à direita é a derivada parcial usual em relação à i-ésima coordenada. Como a derivada satisfaz a regra do produto, é imediato verificar que cada $\partial/\partial x_i|_p$ é um elemento de T_pM. Mais ainda, é possível mostrar que essas derivações formam uma base para T_pM e, consequentemente, $\dim T_pM = n$, como dizia nossa intuição.

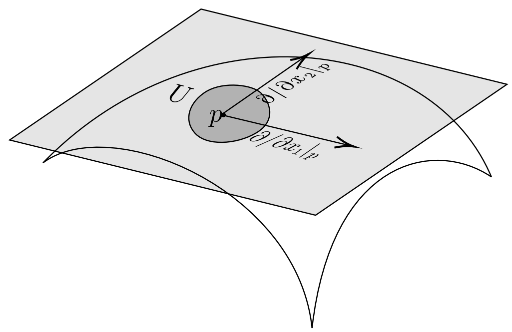

Figura 2: Espaço Tangente

Outro conceito muito importante é o de diferencial de uma função suave. Se $F : N \longrightarrow M$ é uma função C^∞, chamamos de *diferencial* num ponto p o mapa

$$F_* : T_pN \longrightarrow T_{F(p)}M$$

definido do seguinte modo: Se $X_p \in T_pN$, então $F_*(X_p) \in T_pM$ é dado por

$$(F_*(X_p))f = X_p(f \circ F),$$

onde $f \in C^\infty_{F(p)}(M)$. Se for preciso ressaltar o ponto p (mas só se for realmente necessário), denotaremos F_* por $F_{*,p}$.

Teorema 1.7 (Regra da Cadeia)**.** Se $F : N \longrightarrow M$ e $G : M \longrightarrow P$ são mapas suaves entre variedades e $p \in N$, então $(G \circ F)_{*,p} = G_{*,F(p)} \circ F_{*,p}$.

É fácil ver que a diferencial da identidade é a própria identidade no espaço tangente. Assim, utilizando a Regra da Cadeia 1.7, a diferencial possui uma propriedade "funtorial". Com essas observações, não é difícil provar o seguinte resultado bastante natural:

Proposição 1.8. Se $F : N \longrightarrow M$ é um difeomorfismo entre variedades e $p \in N$, então $F_* : T_{F(p)}N \longrightarrow T_pM$ é um isomorfismo de espaços vetoriais.

1.3 Formas diferenciais

Um k-*tensor* num espaço vetorial V é uma função k-linear de $V \times \cdots \times V$ (k vezes) em $\mathbb{R}$. Um k-tensor é *alternado* se, ao trocarmos duas coordenadas de posição, o valor do tensor troca de sinal. Estudaremos k-tensores definidos no espaço tangente a uma variedade. A ideia é que esses tensores dão a noção de volume, assim como o determinante (que é um tensor) nos dá o volume de paralelepípedos em $\mathbb{R}^n$. Denotaremos o conjunto dos k-tensores alternados em V por $A_k(V)$.

Definição 1.9. Uma *k-forma* ω numa variedade M é uma função que associa a cada ponto $p \in M$ um k-tensor alternado $\omega_p \in A_k(T_pM)$.

Será importante trabalhar com 0-formas, e uma 0-forma em M nada mais é do que uma função de M em $\mathbb{R}$ (cada ponto é associado a um 0-tensor, que seria um ponto de $\mathbb{R}$).

Exemplo 1.10 (Exemplo canônico de uma 1-forma)**.** Se (U, ϕ) é uma carta coordenada numa variedade em M e se $x_1, \ldots, x_n$ são suas coordenadas, podemos definir

$$(\mathrm{d}x_n)_p(X_p) = X_p(x_n)$$

para $p \in U$ e $X_p \in T_pM$. É imediato que $(\mathrm{d}x_n)_p$ é uma função linear de T_pM em $\mathbb{R}$, de modo que $\mathrm{d}x_n$ é uma 1-forma em U. Note que $A_1(T_pM)$ é o espaço dual de T_pM e é possível mostrar que $(\mathrm{d}x_1)_p, \ldots, (\mathrm{d}x_n)_p$ formam uma base de $A_1(T_pM)$ dual à base $\partial/\partial x_1|_p, \ldots, \partial/\partial x_n|_p$ de T_pM.

O exemplo anterior pode ser generalizado para k-formas. Denote por $\mathfrak{S}_n$ o conjunto das permutações do conjunto dos índices $1, \ldots, n$. Se $v = (v_1, \ldots, v_n)$ e σ uma permutação, defina $v^\sigma = (v_{\sigma(1)}, \ldots, v_{\sigma(n)})$.

Definição 1.11. Se α é um k-tensor alternado e β é um ℓ-tensor alternado em V, podemos definir o seu *produto exterior*, que é o $(k+\ell)$-tensor alternado $\alpha \wedge \beta$ dado por

$$(\alpha \wedge \beta)(u, v) = \frac{1}{k!\ell!} \sum_{\sigma \in \mathfrak{S}_{k+\ell}} \operatorname{sgn}(\sigma)\, \alpha(u^\sigma)\beta(v^\sigma),$$

onde σ percorre todas as permutações dos números de 1 até $k + \ell$ e $\operatorname{sgn}(\sigma)$ denota o *sinal* da permutação σ.

É importante ressaltar que o produto exterior é associativo e anticomutativo. Nesse sentido, se ω é uma k-forma e τ é uma ℓ-forma em M, então seu produto exterior é definido ponto a ponto:

$$(\omega \wedge \tau)_p = \omega_p \wedge \tau_p.$$

Um resultado interessante é que, sobre uma carta coordenada (U, ϕ) com componentes $x_1, \ldots, x^n$, uma base para $A_k(T_pM)$ é dada pelos tensores

$$(\mathrm{d}x_{i_1})_p \wedge \cdots \wedge (\mathrm{d}x_{i_k})_p, \quad 1 \leqslant i_1 < \cdots < i_k \leqslant n.$$

Assim, toda k-forma ω em U se escreve como

$$\omega = \sum a_{i_1 \cdots i_k} \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k},$$

onde $a_{i_1 \cdots i_k}$ são funções em U. Com base nessa propriedade, diremos que uma k-forma ω em M é *suave*[1] ou C^∞ se em toda carta (U, ϕ) os coeficientes $a_{i_1 \cdots i_k}$ acima forem funções C^∞ em U. Denotaremos o conjunto de todas as k-formas C^∞ em M por $\Omega^k(M)$. Vale notar que $\Omega^0(M)$ é exatamente o conjunto das funções suaves de M em $\mathbb{R}$. Também podemos considerar $\Omega^*(M)$ como sendo a soma direta dos espaços $\Omega^k(M)$ para $k \geqslant 0$.

Agora vamos definir duas operações muito importantes no estudo de formas: o *pullback* e a derivada exterior. Se $T : V \longrightarrow W$ é uma transformação linear entre espaços vetoriais, temos o *pullback* induzido $T^* : A_k(W) \longrightarrow A_k(V)$ dado por

$$(T^*\alpha)(v_1, \ldots, v_k) = \alpha(T(v_1), \ldots, T(v_k))$$

para $\alpha \in A_k(W)$ e $v_1, \ldots, v_k \in V$. No contexto de espaços tangentes, se $F : N \longrightarrow M$ é um mapa C^∞ entre variedades, para cada ponto $p \in N$ podemos tomar o *pullback* da diferencial $F_{*,p}$:

$$(F_{*,p})^* : A_k(T_{F(p)}M) \longrightarrow A_k(T_pN)$$

Simplificaremos essa notação para F^* apenas. Assim, se ω é uma k-forma em M, o seu *pullback* é a k-forma $F^*\omega$ em N dada por $(F^*\omega)_p = F^*(\omega_{F(p)})$ para todo $p \in N$.

O *pullback* de formas possui algumas propriedades importantes. Se ω é uma k-forma suave, então $F^*\omega$ também é suave. Além disso, visto como função de $\Omega^k(M)$ em $\Omega^k(N)$, o *pullback* é linear. Outra propriedade importante é que ele "comuta" com o produto exterior, ou seja,

$$F^*(\omega \wedge \tau) = F^*(\omega) \wedge F^*(\tau).$$

Por fim, temos mais uma propriedade funtorial importante: $(G \circ F)^* = F^* \circ G^*$, para mapas suaves $F : N \longrightarrow M$ e $G : M \longrightarrow P$ entre variedades.

Vejamos agora o que é a derivada exterior. Começamos com o conceito para uma 0-forma. Se $f \in \Omega^0(M)$, sua *diferencial* é a 1-forma $\mathrm{d}f$ em M dada por

$$(\mathrm{d}f)_p(X_p) = X_p(f)$$

para $p \in M$ e $X_p \in T_pM$. Note que a definição que demos para $\mathrm{d}x_n$ é exatamente a diferencial da função coordenada x_n. Com essa definição em mãos, é possível mostrar que existe uma única função linear $\mathrm{d} : \Omega^*(M) \longrightarrow \Omega^*(M)$ satisfazendo:

1. Se ω é uma k-forma, então $\mathrm{d}\omega$ é uma $(k+1)$-forma.

2. Se ω é uma k-forma e τ uma ℓ-forma, então[2]

$$\mathrm{d}(\omega \wedge \tau) = (\mathrm{d}\omega) \wedge \tau + (-1)^k \omega \wedge (\mathrm{d}\tau).$$

[1] Um jeito natural de definir a suavidade de uma k-forma é através do *fibrado cotangente* da variedade, mas não entraremos nos detalhes dessa abordagem.

[2] Essa é quase uma "regra de Leibniz" com um fator de sinal, que é herdado da anticomutatividade do produto exterior.

3. $\mathrm{d}^2 = 0$.

4. Se f é uma 0-forma, então $\mathrm{d}f$ é a diferencial de f como definida acima.

Se (U, ϕ) é uma carta coordenada em M com funções coordenadas $x_1, \ldots, x^n$ e ω é uma k-forma em U, pode-se mostrar que

$$\mathrm{d}\omega = \sum_{i_1 \cdots i_k} \sum_j \frac{\partial a_{i_1 \cdots i_k}}{\partial x^j} \mathrm{d}x^j \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k},$$

onde $a_{i_1 \cdots i_k}$ são os coeficientes de ω escrita na "base" $\mathrm{d}x_1, \ldots, \mathrm{d}x^n$. Outra propriedade digna de nota é que a derivada exterior "comuta" com o *pullback*: se $F : N \longrightarrow M$ é um mapa suave entre variedades, então

$$\mathrm{d}(F^*\omega) = F^*\mathrm{d}\omega \tag{1}$$

para qualquer $\omega \in \Omega^k(M)$.

1.4 Integração em variedades

Terminamos essa seção com uma breve descrição de como utilizar formas para integrar em variedades. Começamos com a integral em $\mathbb{R}^n$. Se $\omega = f(x)\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x^n$ é uma n-forma C^∞ definida num aberto $U \subseteq \mathbb{R}^n$, sua integral sobre U é definida pela Integral de Riemann de f:

$$\int_U \omega = \int_U f(x)\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x^n := \int_U f(x)\,\mathrm{d}x_1 \cdots \mathrm{d}x^n,$$

se a integral existir. A priori, parece que não estamos fazendo nada de novo, mas a linguagem de formas se comporta muito bem com a mudança de coordenadas. Se $T : \mathbb{R}^n \supseteq V \longrightarrow U \subseteq \mathbb{R}^n$ é um difeomorfismo que preserva orientação, então

$$\int_V T^*\omega = \int_U \omega.$$

É muito importante que o difeomorfismo anterior preserve orientação. Por causa disso, só conseguiremos integrar em variedades orientáveis[3].

Finalmente, seja M uma variedade orientada de dimensão n. Tome $\omega \in \Omega^n(M)$ de suporte compacto. Se o suporte de ω está contido numa carta coordenada (U, ϕ) do atlas orientado, definimos

$$\int_U \omega := \int_{\phi(U)} (\phi^{-1})^*\omega,$$

que é a integral de uma forma em $\mathbb{R}^n$. Mostra-se que isso está bem-definido e não depende da carta escolhida. Caso não consigamos encontrar a carta U, ainda conseguimos definir a integral de ω para todo M através de uma partição da unidade.

[3]Uma variedade é *orientável* se possui um atlas tal que a função de transição entre duas cartas quaisquer sempre preserva orientação, ou seja, tem determinante Jacobiano positivo.

1.5 Campos e Fibrados vetoriais

Se voltarmos as aulas de cálculo e considerar um aberto $U \subset \mathbb{R}^n$, os campos suaves eram aplicações com a seguinte cara:

$$\begin{aligned} F: \ U &\longrightarrow U \times \mathbb{R}^n \\ x &\longmapsto (x, f_1(x), \ldots, f_n(x)) \end{aligned}$$

de modo que a aplicação $(f_1, \ldots, f_n) : U \longrightarrow \mathbb{R}^n$ era suave. Podemos reinterpretar essa noção através da projeção canônica da primeira coordenada Π_1, onde os campos suaves são as aplicações suaves $F : U \longrightarrow U \times \mathbb{R}^n$ tal que $\Pi_1 \circ F = I_{\mathbb{R}^n}$. Nesse contexto, o aberto U é o nosso *espaço de configurações* e o produto cartesiano $U \times \mathbb{R}^n$ o nosso *espaço de fases*.

Definição 1.12. Sejam E e M variedades, $\pi : E \to M$ uma submersão. Suponha que existe uma cobertura $(U_\alpha)_\alpha$ de M e difeomorfismos $\phi_\alpha : U_\alpha \times \mathbb{R}^n \longrightarrow \pi^{-1}(U_\alpha)$ tais que:

(i) $\pi \circ \phi_\alpha = I_M$.

(ii) Se U_α e U_β não são disjuntos, então

$$\phi_\beta^{-1} \circ \phi_\alpha(p, v) = (p, \theta_{\beta,\alpha}(p)v) \qquad \begin{pmatrix} p \in U_\alpha \cap U_\beta \\ v \in \mathbb{R}^n \end{pmatrix}$$

onde $\theta_{\alpha,\beta} : U_\alpha \cap U_\beta \longrightarrow \mathrm{GL}_n(\mathbb{R})$ é suave.

(iii) (ϕ_α, U_α) é máximo em relação aos itens acima.

A tripla (E, M, π) é chamada de *fibrado vetorial* de posto n e projeção π. Para cada $p \in M$ o espaço $E_p := \pi^{-1}(p)$ é chamado *fibra* sobre p e herda naturalmente uma estrutura de espaço vetorial.

O espaço tangente já nos fornece dois exemplos importantes de fibrados: O *fibrado tangete* $TM := \bigcup_{p \in M} T_pM$ e o *cotangente* $TM^* := \bigcup_{p \in M} (T_pM)^*$ onde cada $(T_pM)^*$ é o espaço dual de T_pM.

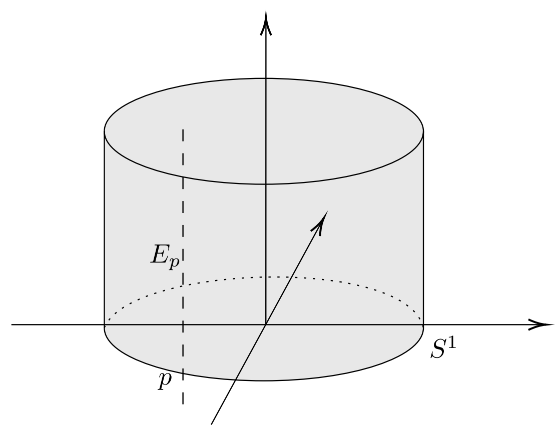

Figura 3: Fibrado trivial de S^1

Definição 1.13. Uma *seção* de um fibrado vetorial (E, M, π) é uma aplicação $\xi : M \longrightarrow E$ tal que $\pi \circ \xi = I_M$. Em particular, um *campo vetorial* F é uma seção do fibrado tangente TM. O conjunto dos campos vetoriais de M é denotado por $\mathfrak{X}(M)$ e é um módulo.

1.6 Métricas Riemannianas

Definição 1.14. Uma *métrica Riemanniana* de uma variedade M é uma seção suave do fibrado dos 2-tensores simétricos positivos definidos em TM. Em outras palavras é uma aplicação que associa a cada ponto p da variedade M um produto interno g_p de T_pM tal que $\mathrm{g}_{i,j} := \mathrm{g}\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right)$ é suave. Uma variedade M com uma métrica g é chamada então de variedade Riemanniana.

Escrevendo em coordenadas, podemos considerar

$$\mathrm{g} \coloneqq \sum_{i,j} \mathrm{g}_{ij}\, \mathrm{d}x_i \otimes \mathrm{d}x_j.$$

Utilizando partições da unidade, sempre é possível considerar uma métrica riemanniana em qualquer variedade, induzida pela métrica euclidiana em cada carta coordenada. Isso é uma particularidade da métrica riemanniana em relação a outras estruturas geométricas (e.g. Lorentziana, simplética).

2 Variedades Estatísticas

2.1 Objeto de estudo

Considere uma família M de distribuições paramétricas sobre um espaço Ω, e espaço de parâmetros $U \subset \mathbb{R}^n$ é um aberto. Então M pode ser descrito como

$$M = \{p(\,\cdot\,|\theta) \mid \theta \in U\}.$$

Para que as coisas façam sentido, é necessário que o mapa $\theta \longmapsto p(\,\cdot\,|\theta)$ seja uma injeção. Além disso, como o contexto é geométrico, cada $p(x|\,\cdot\,)$ precisa ser uma aplicação suave.

O mapa $\phi(p(\,\cdot\,|\theta)) \coloneqq \theta$ é uma carta coordenada global. Além disso, qualquer difeomorfismo suave ψ entre abertos U e V induz uma reparametrização de M.

2.2 A matriz de Informação de Fisher

Agora que M pode ser vista como uma variedade, introduzimos uma noção de métrica. Como cada ponto de M é uma distribuição, a a referida métrica calculará a distância entre duas distribuições.

Se $\ell(\theta) \coloneqq \log p(x|\theta)$, considere

$$\mathrm{g}_{ij}(\theta) \coloneqq \mathbb{E}_\theta \left[\frac{\partial \ell(\theta)}{\partial \theta_i} \frac{\partial \ell(\theta)}{\partial \theta_j} \right]$$

onde $\mathbb{E}_\theta[Y(\theta)] = \int_\Omega Y(\theta) p(x|\theta)\,\mathrm{d}x$. Facilitando a notação, escreva $\partial_i \ell(\theta)$ para a derivada partial em relação a i-ésima coordenada do parâmetro θ. Estes elementos g_{ij} dão luz a matriz de informação de Fisher $\mathcal{I}(\theta)$.

Proposição 2.1. $\mathcal{I}(\theta)$ é uma matriz simétrica, positiva definida.

Demonstração. Trocando a ordem de integração pela derivação, nota-se que $\mathrm{g}_{ij}(\theta) = -\mathbb{E}_\theta[\partial_i^2 \ell_w]$ e portanto, $\mathrm{g}_{ij} = \mathrm{g}_{ji}$, i.e. é simétrica. Note que os vetores da forma $\partial_i \ell(\theta)$ são linearmente independentes. Se $c \in \mathbb{R}^n$ é um vetor qualquer, note que

$$\begin{aligned}
c^{\mathsf{t}} \mathrm{g}_{ij}(\theta) c &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \mathbb{E}_\theta \left[\partial_i \ell(\theta) \partial_j \ell(\theta) \right] \\
&= \mathbb{E}_\theta \left[\sum_{i=1}^{n} c_i \partial_i \ell_\xi \sum_{j=1}^{n} \partial_j \ell(\theta) \right] \\
&= \mathbb{E}_\theta \left[\left(\sum_{i=1}^{n} c_i \partial_i \ell(\theta) \right)^2 \right] \\
&= \int_\Omega \left\{ \sum_{i=1}^{n} c_i \partial_i \ell(\theta) \right\}^2 p(x|\theta)\,\mathrm{d}x \geqslant 0
\end{aligned}$$

Portanto, $\mathcal{I}(\theta)$ é positiva definida. $\square$

Daí, temos uma métrica Riemmaniana em M e para cada parâmetro θ, o produto interno do espaço tangente $T_{p(\,\cdot\,|\theta)}M$ é dado por $\mathrm{g}_\theta \coloneqq \sum_{i,j} \mathrm{g}_{ij}(\theta)\,\mathrm{d}x_i \otimes \mathrm{d}x_j$.

Exemplo 2.2 (Modelo Gaussiano)**.** Considere $\mathcal{N}$ a variedade *normal* cuja os parâmetros são da forma (μ, σ^2) e a distribuição é esperada:

$$p(x|\mu,\sigma^2) \coloneqq \frac{1}{\sqrt{2\pi\sigma}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2} \Big).$$

Note que o espaço dos parâmetros é um semi-plano da forma $\mathbb{R} \times \mathbb{R}_{>0}$ e portanto:

$$\ell(\mu, \sigma^2) = -\frac{1}{2}\log(2\pi\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$$

de modo que $\mathrm{g}_{ij}(\mu, \sigma^2) = \sigma^{-2}\delta_{ij}$ com $i, j \in \{1, 2\}$.

2.3 Uma abordagem utilizando a divergência KL

Essa métrica pode ser obtida através da noção de entropia cruzada. Se θ e η são parâmetros, a entropia relativa pode ser definida como

$$D_{\mathrm{KL}}(\eta \| \theta) := -\int_\Omega p(x|\eta) \log \frac{p(x|\theta)}{p(x|\eta)}\, \mathrm{d}x$$

que representa a informação perdida caso a $p(\,\cdot\,|\eta)$ for utilizada ao invés de $p(\,\cdot\,|\theta)$. É sabido que essa função em sí não é simétrica, mas como a ideia é uma noção local, faz sentido avaliar caso η e θ forem próximos o suficiente.

Dessa forma, podemos utilizar a expansão de Taylor ao redor de θ e considerar que η é uma perturbação infinitesimal de θ. Nota que a entropia relativa é estritamente maior que 0, visto que a igualdade ocorre somente quanto $\eta = \theta$. Utilizando o polinômio de grau 2, os dois primeiros termos tendem a 0 e portanto:

$$\lim_{\eta \to \theta} D_{\mathrm{KL}}(\eta \| \theta) = \frac{1}{2} \sum_{i,j} \frac{\partial^2 D_{\mathrm{KL}}(\eta \| \theta)}{\partial \eta_i \partial \eta_j}\Big|_{\eta=\theta} \mathrm{d}\theta_i \mathrm{d}\theta_j .$$

Avaliando o termo da derivada, temos:

$$\begin{aligned}
\frac{\partial^2 D_{\mathrm{KL}}(\eta \| \theta)}{\partial \eta_i \partial \eta_j} &= \frac{\partial}{\partial \eta_i} \int_\Omega \left[\log \frac{p(x|\eta)}{p(x|\theta)} + 1\right] \frac{\partial p(x|\eta)}{\partial \eta_j}\, \mathrm{d}x \\
&= \int_\Omega \Bigg[\frac{\partial \log p(x|\eta)}{\partial \eta_i} \frac{\partial p(x|\eta)}{\partial \theta_j} \\
&\qquad + \left[\log \frac{p(x|\eta)}{p(x|\theta)} + 1\right] \frac{\partial^2 p(x|\eta)}{\partial \eta_j \partial \theta_i} \Bigg] \mathrm{d}x
\end{aligned}$$

Fazendo as contas necessárias, temos

$$\mathrm{g}(\theta) := \lim_{\eta \to \theta} D_{\mathrm{KL}}(\eta \| \theta) = \frac{1}{2} \sum_{i,j} \int_\Omega \partial_i \ell(\theta) \partial_j \ell(\theta)\, \mathrm{d}x = \frac{1}{2} \mathbb{E}_\theta[\partial_i \ell(\theta) \partial_j \ell(\theta)]$$

i.e., a menos de uma constante, a matriz de Fisher mais uma vez.

***Observação* 2.3.** Qualquer função de divergência f traria a menos de uma constante a mesma métrica. Essa é uma propriedade que garante a unicidade da métrica riemanniana das variedades estatísticas. A premissa chave do argumento é a monotonicidade da informação. Este é um princípio sobre como qualquer medida de divergência, ou diferença, entre as distribuições de probabilidade deve se comportar sob o "aumento" de nosso conhecimento.

3 Aplicações

3.1 *Prioris* impróprios

Referências

Alexandrino, M. M. (2022). Introdução à Geometria Riemanniana. https://www.ime.usp.br/~malex/arquivos/listas2022/2022-Geometria-Riemanniana-alexandrino.pdf.

Munkres, J. R. (2018). *Analysis on Manifolds*. CRC Press.