Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, Vol. 3, N. 1, 2015.
Trabalho apresentado no XXXV CNMAC, Natal-RN, 2014.

Geometria da Informação: Métrica de Fisher

Julianna Pinele Santos Porto,
Instituto de Matemática, Estatística e Computação Científica,UNICAMP,
13083-859 , Campinas, SP
E-mail: juli_pinele@hotmail.com.

João Eloir Strapasson
Faculdade de Ciências Aplicadas, UNICAMP
13484-350, Limeira, SP
E-mail: joao.strapasson@fca.unicamp.br.

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

Neste trabalho mostramos a distância de Rao no espaço formado por distribuições Normais Multivariadas. Neste espaço, como ainda não é conhecida uma fórmula fechada para a distância e nem para a curva geodésica, damos ênfase ao cálculo de limitantes para a distância de Rao. Conseguimos melhorar, em alguns casos, o limitante superior dado por Calvo e Oller em 1990.

Palavras-chave: *Geometria da Informação, Métrica de Fisher, Limitantes.*

1 Métrica de Fisher

Seja S uma família de distribuições de probabilidade sobre X. Suponha que cada elemento de S, uma distribuição de probabilidade, seja parametrizado por n variáveis reais $(\theta_1, \theta_2, \cdots, \theta_n)$, isto é,

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

para $\Theta \subset \mathbb{R}^n$ e a aplicação injetiva $\boldsymbol{\theta} \mapsto p_{\boldsymbol{\theta}}$. Chamamos S de modelo estatístico de dimensão n.

Dado um modelo estatístico $S = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \mathbb{R}^n\}$ podemos definir uma aplicação injetiva $\varphi : U \to S$, $U \subset \mathbb{R}^n$, dada por $\varphi(\boldsymbol{\theta}) = p_{\boldsymbol{\theta}}$. Dessa forma, dado $p_{\boldsymbol{\theta}} \in S$, o par (U, φ) é um sistema de coordenadas para S em torno de $p_{\boldsymbol{\theta}}$ e dizemos que $(\theta_1, \theta_2, \cdots, \theta_n) = (\theta_i)$ são as coordenadas de $p_{\boldsymbol{\theta}}$ nesse sistema de coordenadas. Seja agora ψ um difeomorfismo injetivo de Θ em $\psi(\Theta)$ tal que ψ e ψ^{-1} sejam C^∞, se usarmos $\boldsymbol{\xi} = \psi(\boldsymbol{\theta})$ como nosso parâmetro em vez de $\boldsymbol{\theta}$, obtemos $S = \{p_{\psi^{-1}(\boldsymbol{\xi})}; \boldsymbol{\xi} \in \psi(\Theta)\}$, a qual é a mesma família de distribuição de probabilidade $S = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$.

Se tomarmos parametrizações que são difeomorfismos C^∞ então podemos considerar S uma variedade diferenciável C^∞, qual chamamos de variedade estatística.

Agora, dada a definição de variedade estatística vamos definir uma métrica nesse espaço. Essa métrica é dada pela matriz de informação de Fisher, definida abaixo:

Definição 1. *(Matriz de informação Fisher) Seja $S = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$ um modelo estatístico de dimensão n. Dado um ponto $\boldsymbol{\theta} \in \Theta$, a matriz de informação Fisher de S em $\boldsymbol{\theta}$ é a matriz*

$G(\boldsymbol{\theta}) = [g_{ij}(\boldsymbol{\theta})]$ de ordem n, com

$$g_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\partial_i \log p(x,\boldsymbol{\theta})\partial_j \log p(x,\boldsymbol{\theta})] = \int \partial_i \log p(x,\boldsymbol{\theta})\partial_j \log p(x,\boldsymbol{\theta}) p(x,\boldsymbol{\theta}) dx,$$

caso a integral exista. Aqui, $\partial_i \stackrel{def}{=} \frac{\partial}{\partial \theta_i}$ e $E_{\boldsymbol{\theta}}$ é a esperança com respeito a distribuição $p_{\boldsymbol{\theta}}$, $E_{\boldsymbol{\theta}}[f] = \int f(x)p(x,\boldsymbol{\theta})dx$. Para $n = 1$, chamamos $G(\boldsymbol{\theta})$ de informação Fisher.

Uma variedade estatística S munida da métrica acima é uma variedade Riemanniana.

Exemplo 2. *Uma distribuição normal univariada com média μ e desvio padrão σ, $N(\mu,\sigma)$ é definida por*

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

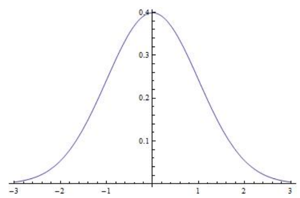

Figura 1: Distribuição normal univariada, $N(0,1)$.

Seja $S = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$ nosso modelo estatístico formado por essas distribuições. Nesse caso o parâmetro $\boldsymbol{\theta}$ é dado por duas variáveis $\boldsymbol{\theta} = (\theta_1,\theta_2) \in \Theta = \{(\mu,\sigma); \mu \in \mathbb{R}, \sigma \in (0,+\infty)\}$. A matriz de informação de Fisher de S em $\boldsymbol{\theta} = (\mu,\sigma)$ é dada por

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

2 Distância entre distribuições

Nesta seção vamos mostrar a distância entre duas distribuições para algumas famílias de distribuições de probabilidade. Encontrar distância geodésicas é um problema muito complexo, pois requer a resolução de equações diferencias de segunda ordem. Em 1945, Rao [8] calculou a distância geodésica entre duas distribuições normais univariadas. Desde então diversos autores veem trabalhando com o problema de encontrar formulas gerias para distribuições geodésica de algumas distribuições, veja [2, 9, 6, 3] .

2.1 Distribuições Normais Univariadas

Para distribuições normais, com parâmetros $\boldsymbol{\theta} = (\mu,\sigma)$, a métrica associada a matriz dada no Exemplo 2 é

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

Agora, dadas duas distribuições normais $N(\mu_1,\sigma_1)$ e $N(\mu_2,\sigma_2)$, queremos saber qual a distância entre elas. Dados dois pontos $P = (\mu_1,\sigma_1)$ e $Q = (\mu_2,\sigma_2)$ do espaço dos parâmetros dado pelo plano superior média×desvio padrão, $\mathbb{H}^2_F$, a distância entre P e Q é dada por

$$d_{\mathbb{H}^2_F}(P,Q) = \min\left\{\int_\gamma \sqrt{\langle \gamma'(t), \gamma'(t)\rangle_{\mathbb{H}^2_F}} dt; \gamma \text{ é curva diferenciável ligando P e Q}\right\}$$

DOI: 10.5540/03.2015.003.01.0263

Observando a matriz da métrica, vemos que é um modelo geométrico hiperbólico o qual pode ser relacionado com o plano de Poincaré, $\mathbb{H}^2 = \{(x,y) \in \mathbb{R}^2; y > 0\}$. A distância de Fisher e Poincaré estão relacionadas por

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

Em [6], temos que a distância é Rao, é dada por

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

Dessa forma, as geodésicas de $\mathbb{H}_F^2$ são as imagens das geodésicas de $\mathbb{H}^2$ por meio da transformação

$$(\mu,\sigma) \to \left(\frac{\mu}{\sqrt{2}},\sigma\right).$$

Essas geodésicas são as semi-retas verticais positivas e as semi-elipses centradas em $\sigma = 0$ com excentricidade $\frac{1}{\sqrt{2}}$.

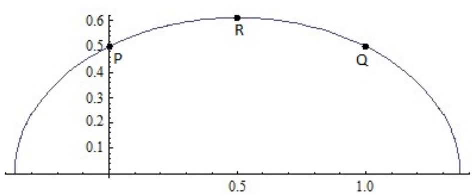

Figura 2: Curva geodésica ligando P e Q no plano $\mu \times \sigma$.

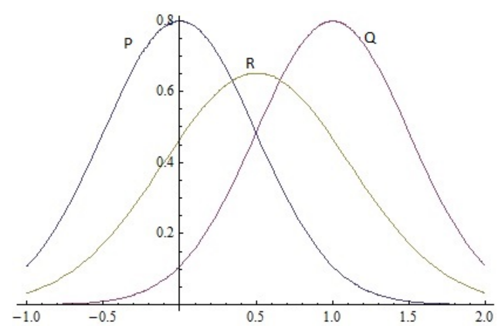

Figura 3: Distribuições normais associadas aos pontos P e Q.

2.2 Distribuições Normais Multivariadas

Uma distribuição normal multivariada é definida por

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

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

Seja $\mathcal{M} = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta = \mathbb{R}^n \times P_n(\mathbb{R})\}$ o modelo estatístico formado por essas distribuições. A equação da métrica da informação de Fisher de $\mathcal{M}$, dada por Skovargaard, ver referência [9] é

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

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

Como ainda não é conhecida uma fórmula fechada pra distância de Rao no caso geral, a seguir descreveremos a distância de Rao de algumas subvariedades $\mathcal{M}$.

Para estudar o espaço formado por essas distribuições, primeiro vamos analisar alguns casos particulares.

DOI: 10.5540/03.2015.003.01.0263

2.3 A matriz de covariância é uma matriz diagonal

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

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

Se a matriz $\boldsymbol{D} = \boldsymbol{\Sigma}$ com a diagonal formada por σ_i^2's, $i = 1, \cdots, n$, $\mathcal{M}_{\boldsymbol{D}}$ é parametrizado por uma interseção de espaços metade no $\mathbb{R}^{2n}$, com $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \cdots, \mu_n, \sigma_n)$ e a matriz de informação Fisher é:

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

A distância para pontos entre duas distribuições com parâmetros $\boldsymbol{\theta}_1 = (\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \cdots, \mu_{1n}, \sigma_{1n})$ e $\boldsymbol{\theta}_2 = (\mu_{21}, \sigma_{21}, \mu_{22}, \sigma_{22}, \cdots, \mu_{2n}, \sigma_{2n})$ é

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \sqrt{2\sum_{i=1}^{n}\left(\log\frac{\left|\left(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}\right) - \left(\frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i}\right)\right| + \left|\left(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}\right) - \left(\frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i}\right)\right|}{\left|\left(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}\right) - \left(\frac{\mu_{2i}}{\sqrt{2}}, -\sigma_{2i}\right)\right| - \left|\left(\frac{\mu_{1i}}{\sqrt{2}}, \sigma_{1i}\right) - \left(\frac{\mu_{2i}}{\sqrt{2}}, \sigma_{2i}\right)\right|}\right)^2}. \tag{1}$$

2.4 A matriz de covariância é constante.

Seja $\mathcal{M}_{\boldsymbol{\Sigma}} = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta,\ \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \in P_n(\mathbb{R})\ \text{constante}\}$ uma subvariedade formada por distribuições de normais multivariadas que possuem a mesma matriz de covariância $\boldsymbol{\Sigma}$. Nesse espaço podemos relacionar a métrica com a métrica de um espaço Euclidiano. A métrica de Fisher de $\mathcal{M}_{\boldsymbol{\Sigma}}$ é

$$ds^2 = d\boldsymbol{\mu}^t\boldsymbol{\Sigma}^{-1}d\boldsymbol{\mu},$$

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

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

2.5 O vetor das médias é constante.

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

Seja $\mathcal{M}_{\boldsymbol{\mu}} = \{p_{\boldsymbol{\theta}}; \boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta_{\boldsymbol{\mu}}\}$, $\Theta_{\boldsymbol{\mu}} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma});\ \boldsymbol{\mu} = \boldsymbol{\mu}_0 \in \mathbb{R}^n\ \text{constante}\} \subset \Theta$ uma subvariedade formada por distribuições que possuem o mesmo vetor de médias $\boldsymbol{\mu}$. A distância nesse espaço é dada pelo seguinte teorema dado por S. T. Jensen em 1976, ver [2] no Apêndice 1.

Teorema 3. *Considere a família de distribuições normais multivariadas $\mathcal{M}_{\boldsymbol{\mu}}$ com o vetor da média comum $\boldsymbol{\mu}$ mas com diferentes matrizes de covariância $\boldsymbol{\Sigma}$. Dados dois elementos dessa família, parametrizados por $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}, \boldsymbol{\Sigma}_1)$ e $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$, a distância entre dois elementos dessa família é dada por*

$$d_{\boldsymbol{\mu}}^2(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \frac{1}{2}\sum_{i=1}^{n}[log(\lambda_i)]^2,$$

para $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ *autovalores de* $\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2$.

3 Limitantes para a distância

Em 1990, Calvo e Oller, [4], calcularam um limite inferior para a distância de Rao. Eles definiram um mergulho de Θ em $P_{n+1}(\mathbb{R})$ através da aplicação:

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

Uma fórmula para a distância Riemanniana d entre $f(\boldsymbol{\theta}_1)$ e $f(\boldsymbol{\theta}_2)$ é dada pelo teorema abaixo.

Teorema 4. *Seja* $\boldsymbol{S}_i = f(\boldsymbol{\theta}_i) = f(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2$, *dois pontos de* $f(\Theta)$, *então*

$$d^2(\boldsymbol{S}_1, \boldsymbol{S}_2) = \frac{1}{2} \sum_{k=1}^{n+1} [\log(\lambda_i)]^2,$$

para λ_k *autovalor de* $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$.

Assim, temos que a distância dada pelo teorema acima é limitante inferior para a distância de Rao do espaço $\mathcal{M}$.

Em um outro artigo, [5], Calvo e Oller calcularam um limite superior para a distância de Rao. Eles definiram uma distância para o subconjunto $\Theta_{\alpha\boldsymbol{\Sigma}} \subset \Theta$,

$$\Theta_{\alpha\boldsymbol{\Sigma}} = \{\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \Theta; \boldsymbol{\Sigma} = \alpha\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_0 \in P_n(\mathbb{R}), \alpha \in \mathbb{R}_+^*\}.$$

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

$$d_{\alpha\boldsymbol{\Sigma}}^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = 2 \operatorname{arccosh}\left(\frac{\sqrt{\alpha}}{2} + \frac{1}{2\sqrt{\alpha}} + \frac{1}{4\sqrt{\alpha}}\boldsymbol{\delta}^t\boldsymbol{\delta}\right) + \frac{n-1}{2}\log^2 \alpha,$$

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

Com essa distância eles determinam um limite superior para distância de Rao, d_R, da variedade $\mathcal{M}$ no caso geral. Dados pontos $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\boldsymbol{\theta}_\alpha = (\boldsymbol{\mu}_2, \alpha\boldsymbol{\Sigma}_1)$ e $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, temos que

$$d_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \leq d_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_\alpha) + d_R(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_2) \leq d_{\alpha\boldsymbol{\Sigma}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_\alpha) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_2),$$

ou seja, $LS_\alpha = d_{\alpha\boldsymbol{\Sigma}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_\alpha) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_2)$ é um limitante superior para a distância de Rao entre $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$.

Podemos melhorar esse limitante escolhendo um α adequado. Em [5], Calvo e Oller determinam uma expressão analítica do escalar α que minimiza a distância entre $\boldsymbol{\theta}_\alpha$ e $\boldsymbol{\theta}_2$, $\alpha = \|\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-\frac{1}{2}}\|^{\frac{1}{n}}$. O mínimo para LS_α pode ser calculado através de métodos numéricos, isto é, podemos encontrar um α_0 tal que $\alpha_0 = \min_\alpha\{d_{\alpha\boldsymbol{\Sigma}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_\alpha) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_2)\}$.

Calculamos um outro limitante superior para a distância de Rao usando a distância do espaço onde a matriz de covariância $\boldsymbol{\Sigma}$ é diagonal. Dados pontos $\boldsymbol{\theta}_1 = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ e $\boldsymbol{\theta}_2 = (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, temos que

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = d((\mathbf{0}, \boldsymbol{I}); (\boldsymbol{\Sigma}_1^{-(1/2)}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_1^{-(1/2)}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-(1/2)})),$$

Fazendo $\boldsymbol{\theta}_0 = (\mathbf{0}, \boldsymbol{I}_n)$, $\bar{\boldsymbol{\theta}}_2 = (\bar{\boldsymbol{\mu}}_2, \bar{\boldsymbol{\Sigma}}_2) = (\boldsymbol{\Sigma}_1^{-(1/2)}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_1^{-(1/2)}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-(1/2)})$ e $\boldsymbol{\theta}_{\boldsymbol{\Lambda}} = (\bar{\boldsymbol{\mu}}_2, \boldsymbol{\Lambda})$, onde $\boldsymbol{\Lambda}$ é a projeção de $\bar{\boldsymbol{\theta}}_2$ na sua diagonal, temos

$$d_R(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = d(\boldsymbol{\theta}_0, \bar{\boldsymbol{\theta}}_2) \leq d_R(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\boldsymbol{\Lambda}}) + d_R(\boldsymbol{\theta}_{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\theta}}_2) \leq d_{\boldsymbol{D}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\boldsymbol{\Lambda}}) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\theta}}_2),$$

e $d_{\boldsymbol{D}}$ é a distância dada na Equação 1. Assim encontramos outro limitante para a distância de Rao dado por $LS_{\boldsymbol{\Lambda}} = d_{\boldsymbol{D}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\boldsymbol{\Lambda}}) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\theta}}_2)$.

Também podemos melhorar o limitante $LS_{\boldsymbol{\Lambda}}$ através de métodos numéricos, observe que ele é obtido projetando $\boldsymbol{\theta}_2$ no espaço $\Theta_{\boldsymbol{D}} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}); \boldsymbol{\Sigma} = \boldsymbol{D} \text{ diagonal}\}$. Logo, basta encontrar uma matriz diagonal $\boldsymbol{\Lambda}_0$ tal que $\boldsymbol{\Lambda}_0 = \min_{\boldsymbol{\Lambda}}\{d_{\boldsymbol{D}}(\boldsymbol{\theta}_0, \boldsymbol{\theta}_{\boldsymbol{\Lambda}}) + d_{\boldsymbol{\mu}}(\boldsymbol{\theta}_{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\theta}}_2)\}$.

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

Fizemos algumas simulações e vimos que, em alguns casos, o nosso limitante ($LS_{\boldsymbol{\Lambda}}$) é melhor que o limitante dado por Calvo e Oller (LS_α). As figuras abaixo mostram os valores dos limitantes para a distância de Rao entre o ponto $\boldsymbol{\theta} = (\mathbf{0}, \boldsymbol{I}_2)$ e um ponto $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Na Figura 4, tomamos $\boldsymbol{\theta}$, tais que os autovalores da matriz $\boldsymbol{\Sigma}$ assumem valores próximos de 1 e na Figura 5 os autovalores de $\boldsymbol{\Sigma}$ assumem valores grandes.

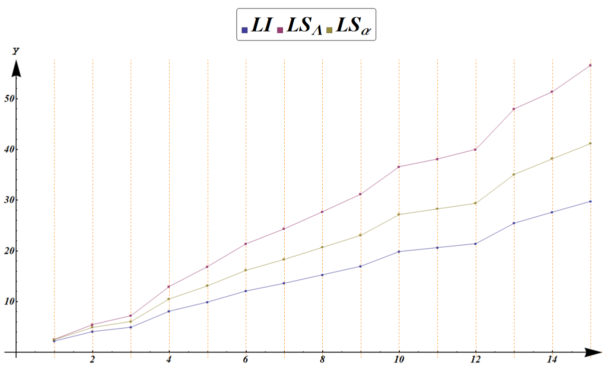

Figura 4: Gráfico dos valores dos limitantes LI, $LS_{\boldsymbol{\Lambda}}$ e LS_α.

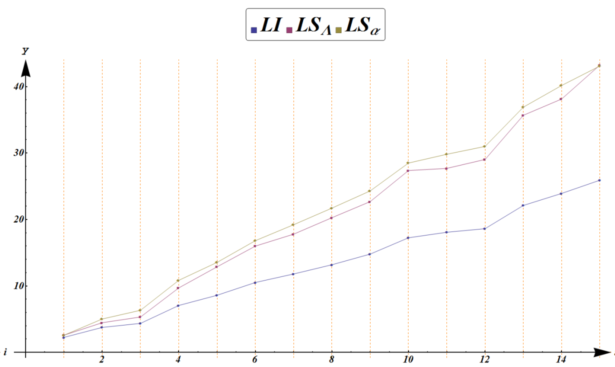

Figura 5: Gráfico dos valores dos limitantes LI, $LS_{\boldsymbol{\Lambda}}$ e LS_α.

Referências

[1] Amari, S. e Nagaoka, H. *Methods of Information Geometry*, Translations of Mathematical Monographs, Vol.191, Am. Math. Soc., 2000.

[2] Atkinson, C. e Mitchell, A. F. S., *Rao's Distance* Measure, Samkhyã- The Indian Journal of Statistics, 43:345-365, 1981.

[3] Burbea, J., *Informative geometry of probability spaces*, Expositiones Mathematica 4, 347-378, 1986.

[4] Calvo, M. e Oller, J. M., *A distance between multivariate normal distributions based in an embedding into the Siegel group*, Journal of Multivariate Analysis 35.2, 223-242, 1990.

[5] Calvo, M., e Oller, J. M., *An explicit solution of information geodesic equations for the multivariate normal model*, Statistics and Decisions 9, 119-138, 1991.

[6] Costa, S. I., Santos, S. A., e Strapasson, J. E., *Fisher information matrix and hyperbolic geometry*, In Information Theory Workshop, 2005 IEEE (pp. 3-pp), IEEE, 2005.

[7] Costa, S. I. R., Santos, S. A., e Strapasson, J. E., *Fisher information distance: a geometrical reading*, arXiv preprint arXiv:1210.2354, 2012.

[8] Rao, C. R., *Information and the accuracy attainable in the estimation of statistical parameters*, Bulletin of the Calcutta Math. Soc. 37:81-91, 1945.

[9] Skovgaard, L. T., *A Riemannian geometry of the multivariate normal model*, Scand, J. Statist., 11:211-223, 1984.

