

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

Aspectos de Geometria Diferencial em Modelos
Estatísticos

André T. Kaneko
Jorge A. Achcar

Nº 136

RELATÓRIOS TÉCNICOS



São Carlos – SP
Mar./2001

SYSNO	1211493
DATA	/ /
ICMC - SBAB	

Aspectos de Geometria Diferencial em modelos
Estatísticos

André T. Kaneko
Jorge A Achcar

Departamento de Computação
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo-Campus de São Carlos
ICMC, caixa postal 668
13560-970, São Carlos , SP, Brasil

1. Introdução

Neste artigo introduzimos uma conexão entre geometria diferencial e modelos estatísticos. Esses modelos estatísticos são parametrizados por quantidades desconhecidas θ .

O uso de geometria diferencial e sua conexão com modelos estatísticos tem sido estudado por vários autores (ver, por exemplo, Barndorff-Nielsen, 1986).

Em geral, a escolha de uma parametrização é arbitrária e uma transformação de θ pode levar à melhores resultados inferenciais (ver por exemplo, Anscombe, 1964, Sprott, 1973, 1980; ou Achcar e Smith, 1990).

O problema da boa parametrização é objeto de pesquisa intensa entre os estatísticos. Parametrização para melhorar a "normalidade" da função de verossimilhança quando o tamanho da amostra não é grande é um dos objetivos de muita pesquisa estatística (ver por exemplo, Sprott, 1973, 1980). Outra possibilidade é encontrar parametrizações ortogonais (ver Cox e Reid, 1987).

Em muitas situações uma boa parametrização pode ser avaliada a partir de medidas de curvatura usando conceitos de geometria diferencial (ver por exemplo, Bates e Watts, 1992, Kass e Slate, 1992).

Neste artigo, resumizamos alguns conceitos básicos de geometria diferencial e estudamos a conexão existente entre geometria diferencial e modelos estatísticos.

2. Alguns conceitos básicos

2.1 Modelo Estatístico

Considere $(\Omega, \beta(R))$ um espaço mensurável e uma família de probabilidade P_θ indexadas por $\theta \in \Theta \subseteq R^m$, onde Θ é o espaço paramétrico. Vamos considerar que existe uma probabilidade λ definida sobre $(R, \beta(R))$, tal que as probabilidades P_θ sejam absolutamente contínuas com relação a uma medida λ sobre $(R, \beta(R))$:

$$P_\theta \ll \lambda, \forall \theta \in \Theta$$

Assim, para todo $\theta \in \Theta$, existe $f: R \rightarrow [0, \infty)$ mensurável tal que

$$P_\theta(\beta) = \int_B f(x) \lambda dx \quad \forall B \in \beta(R) \quad (1)$$

Desta forma, obtemos uma família de funções densidade $\{f(\cdot); \theta \in \Theta\}$ parametrizadas por θ . Tal família caracteriza o modelo estatístico.

Alguns exemplos :a) Distribuição Normal Univariada

$$\theta = (\mu, \sigma)$$

$$\Theta = \{(\mu, \sigma): \mu \in R, \sigma > 0\} = R \times R_+ \subset R^2$$

$$x > 0$$

$$f(x) = 1/(2\pi)^{1/2} \exp[(-1/2\sigma^2)(x-\mu)^2]$$

b) Distribuição Gama

$$\theta = (\alpha, \beta)$$

$$\Theta = \{(\mu, \beta) : \mu \in \mathbb{R}_+, \beta \in \mathbb{R}_+\} = \mathbb{R}_+ \times \mathbb{R}_+ \subset \mathbb{R}^2$$

$$x > 0$$

$$f(x) = (\beta/\mu)^\beta (x)^{\beta-1} / \Gamma(\beta) \exp(-x \beta/\mu)$$

c) Distribuição Binomial Negativa

$$X(T) = \{0, 1, 2, \dots\} = \mathbb{N}$$

$$f(x) = (r+x-1)! / [x!(r-1)!] p^r q^x$$

$$0 < p < 1$$

$$q = 1 - p$$

$$x = 0, 1, 2, \dots$$

No caso dominado (absolutamente contínuo ou discreto), o modelo estatístico é definido por $S = \{f_\theta(x); \theta \in \Theta \text{ e } x \in \mathbb{R}^n\}$ e para o caso absolutamente contínuo e $S = \{P_\theta(x); \theta \in \Theta \text{ e } x \in \mathbb{N}^m\}$ para o caso discreto. Em geral, tanto no modelo absolutamente contínuo quanto no discreto o espaço paramétrico é contínuo. Em termos de Inferência Estatística, diferentes parâmetros definem diferentes pontos em S . Por exemplo, duas Normais Univariadas com mesmas médias μ , mas com variâncias σ_1 e σ_2 diferentes são representadas pelos pontos (μ, σ_1) e (μ, σ_2) que são diferentes. No entanto, o mesmo não ocorre quanto a x , podemos ter x 's (conjunto de dados) diferentes que conduzem a mesmos parâmetros.

Caso as funções f_θ sejam suficientemente suaves, é possível introduzimos uma estrutura de diferenciabilidade sobre o modelo estatístico, ou seja, sobre a família $\{f_\theta(x); \theta \in \Theta\}$, $\forall \theta \in \Theta$.

Sendo S_x a secção de S num ponto x fixo, $S = \{f_\theta(x); \theta \in \Theta, x \in \Omega\}$, podemos definir uma função $\phi_x: S_x \rightarrow \mathbb{R}^n$ por $\phi[f_\theta(x)] = \theta, \forall \theta \in \Theta$.

Assim na distribuição Normal Univariada, ϕ_x é dada por $\phi = (\mu, \sigma)$; na distribuição Weibull, $\phi = (\alpha, \beta)$, na distribuição Binomial Negativa $\phi = (p)$.

Uma função h é dita ser um homeomorfismo se

a) h é bijetora

b) h e sua inversa h^{-1} são contínuas.

A função $\phi: S \rightarrow O$ é um homeomorfismo onde $O \subset \mathbb{R}^n$ é aberto.

No caso das distribuições estatísticas analisadas, ϕ obedece as condições da definição de homeomorfismo.

Esse homeomorfismo ϕ é chamado função coordenada; ele define a parametrização sob a qual o modelo se encontra.

Podemos definir diversos sistemas de coordenadas para o modelo estatístico. Todos esses diferentes sistemas de coordenadas se referem ao mesmo modelo estatístico. Seja Ψ um outro sistema de coordenadas para o modelo estatístico, se Ψ puder ser expressa como função do sistema anterior de coordenadas, dizemos que foi realizada uma reparametrização do modelo estatístico.

Uma reparametrização bastante utilizada no caso da Normal Univariada é a de parâmetros Naturais:

$$\begin{aligned}\phi &= (\phi_1, \phi_2) \\ \phi_1 &= \mu/\sigma^2 \\ \phi_2 &= 1/(2\sigma^2)\end{aligned}$$

Dependendo de quais questões estamos interessados em resolver, a parametrização usada torna-se fundamental. Como ilustração temos na análise Bayesiana, o interesse em distribuições a posteriori próximas à normal. A pergunta natural que surge é qual reparametrização poderia levar a tal resultado.

2.2. Curvaturas para superfícies

As condições requeridas para a aplicação da teoria geométrica nos modelos estatísticos são

- a) Todas as $p(x, \theta)$ devem ter um suporte comum tal que $p(x, \theta) > 0$ para todos os $x \in X$, onde X é o suporte.
- b) Seja $l(x, \theta) = \ln p(x, \theta)$. Para todo θ fixado, n funções em x

$\partial l(x, \theta) / \partial \theta^i$ são linearmente independentes.

c) Os momentos das variáveis aleatórias $(\partial / \partial \theta^i) l(x, \theta)$ existem até as ordens necessárias.

d) As derivadas parciais $\partial / \partial \theta^i$ e a integração com respeito a medida P pode ser intercambiada como

$$\partial / \partial \theta^i \int f(x, \theta) dP = \int \partial / \partial \theta^i f(x, \theta) dP$$

para quaisquer funções $f(x, \theta)$ a serem tratadas. (2)

2.2.1. Espaço tangente

O espaço tangente T_P no ponto P é o espaço vetorial gerado pelos n vetores tangentes ∂_j à superfície no ponto P . Trata-se de uma linearização local da superfície ao redor do ponto P .

Como os n vetores ∂_j são linearmente independentes (condição de regularidade b), tais vetores formam uma base do espaço tangente T_P . É chamada base natural associada ao sistema de coordenadas θ .

Dessa forma, qualquer vetor tangente $A \in T_P$ pode ser representado como combinação linear de ∂_j :

$$A = \sum_{j=1}^n A^j \partial_j \quad (3)$$

ou na notação de Einstein (notação da Análise Tensorial que omite o sinal de somatório e pressupõe somar sobre os índices repetidos)

$$A = A^j \partial_j$$

Alguns exemplos:

1) Distribuição Normal: A função de verossimilhança para $\theta = (\mu, \sigma)$ é dada por,

$$L(x; \sigma) = \prod_{j=1}^n f(x_j) = \prod_{j=1}^n 1/(2\pi\sigma) \exp[-1/(2\sigma^2)(x_j - \mu)^2]$$

Daí, $l = \ln L(x; \sigma) = -1/2 \ln(\sigma) + \sum_{j=1}^n -(x_j - \mu)^2 / (2\sigma^2) + \text{constante}$

$$l_{\mu} = \sum_{j=1}^n -(x_j - \mu) / (2\sigma^2)$$

$$l_{\sigma} = -1/(2\sigma) + \sum_{j=1}^n (x_j - \mu)^2 / (\sigma^3)$$

2) Distribuição Gama

$$L(x; \mu, \beta) = \prod_{j=1}^n f(x_j) = \prod_{j=1}^n (\beta/\mu)^\beta (x_j)^{\beta-1} / \Gamma(\beta) \exp(-x_j \beta/\mu)$$

$$\text{Daí } l = \ln L(x; p) = \beta/\mu \ln(\beta) - n \ln[\Gamma(\beta)] + (\beta-1) \sum_{j=1}^n \ln(x_j) - \sum_{j=1}^n (x_j \beta/\mu)$$

$$l_{\alpha} = -\beta/\mu + \sum_{j=1}^n x_j \beta/\mu^2$$

$$l_{\beta} = \ln(\beta/\mu) + 1 + \sum_{j=1}^n \ln(x_j) - \partial[\Gamma(\beta)/\partial\beta]/\Gamma(\beta) - \sum_{j=1}^n x_j/\mu$$

3) Distribuição Binomial Negativa

$$L(x; p) = \prod_{j=1}^n f(x_j) = \prod_{j=1}^n (r+x-1)! / [x!(r-1)!] p^r q^x, \text{ onde } q=1-p$$

$$\text{Daí, } l = \ln L(x; p) = nr \ln(p) + (\sum_{j=1}^n x_j) \ln(1-p) + \text{constante}$$

$$l_p = nr/p - (\sum_{j=1}^n x_j) / (1-p)$$

Notamos que a dimensão do espaço tangente das duas primeiras distribuições é dois e o da terceira distribuição é um. Distribuições com espaço tangente unidimensional não podem ser analisadas pelo cálculo tensorial de superfícies.

2.2.2. Métrica Riemaniana e Informação de Fisher

Se introduzirmos o produto interno $\langle A, B \rangle$ de dois vetores tangentes A, B em T_p , a superfície S é chamada de espaço Riemaniano com o produto interno.

$$\langle A, B \rangle = E[A(x)B(x)] \quad (4)$$

que é uma forma natural de se definir o produto interno.

O produto interno de dois vetores da base ∂_i e ∂_j é dado por

$$g_{ij}(\theta) = \langle \partial_i, \partial_j \rangle = E[\partial_i l(x, \theta), \partial_j l(x, \theta)] \quad (5)$$

Essas n^2 quantidades $g_{ij}(\theta)$, $i, j=1, \dots, n$ compõe um objeto chamado tensor métrico.

Alguns exemplos:

1) Distribuição Normal

$$l_{\mu\mu} = -n/\sigma^2$$

$$l_{\sigma\sigma} = -n/\sigma^2 - 3/(\sigma^4) \sum_{i=1}^n (x_i - \mu)^2$$

$$l_{\mu\sigma} = -2/(\sigma^3) \sum_{i=1}^n (x_i - \mu)$$

Isto é,

$$g_{11} = E\{-q_{\mu\mu}\} = -n/\sigma^2$$

$$g_{22} = E\{-q_{\sigma\sigma}\} = 2n/\sigma^2$$

$$g_{12} = g_{21} = E\{-q_{\mu\sigma}\} = 0$$

Portanto, a matriz de informação de Fisher é dada por

$$g_{11} = -n/\sigma^2$$

$$g_{12}=g_{21}=0$$

$$g_{22}=2n/\sigma^2$$

2) Distribuição Gama

$$l_{\mu\mu} = \beta / -2\beta / \mu \sum_{i=1}^n (x_i)$$

$$l_{\mu\beta} = -1/\mu + \sum_{i=1}^n (x_i) / \mu^2$$

$$l_{\beta\beta} = 1/\beta - \Gamma_{\beta\beta}(\beta) / \Gamma(\beta) + \Gamma_{\beta}^2(\beta) / \Gamma^2(\beta)$$

Isto é,

$$g_{11} = E\{-q_{\mu\mu}\} = \beta / \mu^2$$

$$g_{22} = E\{-q_{\sigma\sigma}\} = 0$$

$$g_{12} = g_{21} = E\{-q_{\mu\sigma}\} = \phi(\beta)$$

onde $\phi(\beta) = [\ln \Gamma(\beta)]_{\beta\beta} - 1/\beta$

Portanto, a matriz de informação de Fisher é dada por

$$g_{11} = \beta / \mu^2$$

$$g_{12} = g_{21} = 0$$

$$g_{22} = \phi(\beta)$$

2.2.3. Conexão Afim

Queremos associar a cada ponto $\theta \in S$, um vetor $A(\theta) \in T_{\theta}$. Essa coleção $A = \{A(\theta) \mid \theta \in S\}$ é chamado campo vetorial.

Um campo vetorial A pode ser representado na forma de componentes como

$$A = A^j(\theta) \partial_j(\theta) \quad (6)$$

Dado um campo vetorial em S , teríamos interesse em compararmos tal campo em dois pontos distintos θ' e θ de S .

Os espaços tangentes T_{θ} e $T_{\theta'}$ são diferentes, não havendo maneira direta de compararmos $A(\theta)$ e $A(\theta')$. Essa diferença ocorre porque as bases $\partial_j(\theta)$ e $\partial_j(\theta')$ são diferentes.

Para compararmos $A(\theta)$ e $A(\theta')$, considere uma aplicação linear $m: T_{\theta+d\theta} \rightarrow T_{\theta}$, entre dois espaços tangentes de pontos próximos por um infinitésimo. Tal aplicação depende de $d\theta$ e se reduz à aplicação identidade quando $d\theta$ tende a zero.

Desde que $d\theta$ é pequeno, o versor $\partial'_j = \partial_j(\theta + d\theta) \in T_{\theta+d\theta}$ é aplicado em $m(\partial'_j)$ perto de $\partial_j(\theta)$.

Seja $\Delta\partial_j$ a diferença entre $m(\partial'_j)$ e ∂_j :

$$\Delta\partial_j = m(\partial'_j) - \partial_j \in T_{\theta}$$

negligenciando-se termos de ordem superior a um.

$\Delta\partial_j$ pode ser expresso como

$$\Delta\partial_j = d\theta^i \Gamma_{ij}^k(\theta) \partial_k \quad (7)$$

onde $d\theta^i \Gamma_{ij}^k(\theta)$ são componentes de $\Delta \partial_j \in T_\theta$, são a representação de $\Delta \partial_j$ no espaço tangente T_θ ,

Essas n^3 funções $\Gamma_{ij}^k(\theta)$ são chamadas coeficientes da conexão afim porque m dá uma correspondência afim entre T_θ e $T_{\theta+d\theta}$.

A métrica do tensor de Riemann $g_{ij}(\theta)$ cria uma conexão afim natural, cujos coeficientes são dados por

$$[i j k] = 1/2 [\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}] \quad (8)$$

Associado a esses símbolos de Christoffel temos o tensor de curvatura de Riemann, dado por (Amari 1982)

$$R_{ijkl} = (\partial_i \Gamma_{jk}^s - \partial_j \Gamma_{ik}^s) g_{sm} + (\Gamma_{imn} \Gamma_{jk}^r - \Gamma_{jmn} \Gamma_{ik}^r) \quad (9)$$

e o escalar de Ricci ou curvatura seccional dada por:

$$K = -g^{ik} g^{jm} R_{ijkl} \quad (10)$$

Para casos unidimensionais pode-se usar a curvatura geométrica canônica, dada por

$$K = g_{11} / (1 + (g_1)^{1/2})^{3/2}$$

Alguns Exemplos

1) Distribuição Normal

Para a distribuição Normal, os elementos da inversa de Fisher são dados por

$$\begin{aligned} \det g &= -2n^2/\sigma^4 \\ g^{11} &= g_{22}/\det g = -\sigma^2/n \\ g^{22} &= g_{11}/\det g = -\sigma^2/2n \\ g^{12} &= g^{12} = -g_{12}/\det g = 0 \end{aligned}$$

As derivadas de g_{ij} são dadas por:

$$\begin{aligned} g_{111} &= 0 \\ g_{112} &= 2n/\sigma^3 \\ g_{221} &= 0 \\ g_{222} &= -4n/\sigma^3 \\ g_{211} &= g_{121} = 0 \\ g_{212} &= g_{122} = 0 \end{aligned}$$

Os símbolos de Christoffel são dados por

$$\begin{aligned} \Gamma_{ijk} &= 1/2 (E_i g_{jk} + E_j g_{ik} - E_k g_{ij}) \\ \Gamma_{111} &= 1/2 (E_1 g_{11} + E_1 g_{11} - E_1 g_{11}) = 1/2 (2 \cdot 0 + 0) = 0 \\ \Gamma_{112} &= 1/2 (E_1 g_{12} + E_1 g_{12} - E_2 g_{11}) = 1/2 (0 + 2n/\sigma^3 + 0) = n/\sigma^3 \\ \Gamma_{121} &= \Gamma_{211} = 1/2 (E_1 g_{21} + E_2 g_{11} - E_1 g_{12}) = 1/2 (2 \cdot 0 - 2n/\sigma^3) = -n/\sigma^3 \\ \Gamma_{122} &= \Gamma_{212} = 1/2 (E_1 g_{22} + E_2 g_{12} - E_2 g_{12}) = 1/2 (0 + 0 + 0) = 0 \end{aligned}$$

$$\Gamma_{221}=1/2(E_2g_{21}+E_2g_{21}-E_1g_{22})=1/2(0+0+0)=0$$

$$\Gamma^k_{ij}=g^{kl}\Gamma_{ijl}$$

$$\Gamma^1_{11}=g^{11}\Gamma_{111}+g^{12}\Gamma_{112}=0+0=0$$

$$\Gamma^2_{11}=g^{21}\Gamma_{111}+g^{22}\Gamma_{112}=0+1/2\sigma=1/2\sigma$$

$$\Gamma^2_{12}=\Gamma^2_{21}=g^{21}\Gamma_{121}+g^{22}\Gamma_{122}=0+0=0$$

$$\Gamma^1_{12}=\Gamma^1_{21}=g^{11}\Gamma_{121}+g^{12}\Gamma_{122}=-1/\sigma+0=-1/\sigma$$

$$\Gamma^2_{22}=g^{21}\Gamma_{221}+g^{22}\Gamma_{222}=0+0=0$$

O tensor de Riemann é dado por,

$$R_{1212}=1/\sigma^4,$$

e a curvatura seccional é dada por,

$$K=-g^{11}g^{22}R_{1212}=-1/2$$

2) Distribuição Gama

Neste caso, a matriz de informação de Fisher é dada por

$$g_{11}=\beta/\mu^2$$

$$g_{12}=g_{21}=0$$

$$g_{22}=\phi(\beta)$$

Os elementos da inversa de Fisher são dados por

$$\det g=\beta\phi(\beta)/\mu^2$$

$$g^{11}=g_{22}/\det g=\mu^2/\beta$$

$$g^{22}=g_{11}/\det g=1/\phi(\beta)$$

$$g^{12}=g^{21}=-g_{12}/\det g=0$$

As derivadas de g_{ij} são dadas por:

$$g_{111}=-2\beta/\mu^3$$

$$g_{112}=1/\mu^2$$

$$g_{221}=0$$

$$g_{222}=\partial\phi(\beta)/\partial\beta=\phi'(\beta)$$

$$g_{211}=g_{121}=0$$

$$g_{212}=g_{122}=0$$

Os símbolos de Christoffel são dados por

$$\Gamma_{ijk}=1/2(E_ig_{jk}+E_jg_{ik}-E_kg_{ij})$$

$$\Gamma_{111}=1/2(E_1g_{11}+E_1g_{11}-E_1g_{11})=-\beta/\mu^3$$

$$\Gamma_{112}=1/2(E_1g_{12}+E_1g_{12}-E_2g_{11})=-1/(2\mu^2)$$

$$\Gamma_{121}=\Gamma_{211}=1/2(E_1g_{21}+E_2g_{11}-E_1g_{12})=1/2(2\cdot 0-2n/\sigma^3)=1/2\phi'(\beta)$$

$$\Gamma_{122}=\Gamma_{212}=1/2(E_1g_{22}+E_2g_{12}-E_2g_{12})=0$$

$$\Gamma_{221}=1/2(E_2g_{21}+E_2g_{21}-E_1g_{22})=0$$

$$\Gamma^k_{ij}=g^{kl}\Gamma_{ijl}$$

$$\begin{aligned}\Gamma^1_{11} &= g^{11}\Gamma_{111} + g^{12}\Gamma_{112} = -1/\mu \\ \Gamma^2_{11} &= g^{21}\Gamma_{111} + g^{22}\Gamma_{112} = -1/(2\mu^2\phi(\beta)) \\ \Gamma^2_{12} &= \Gamma^2_{21} = g^{21}\Gamma_{121} + g^{22}\Gamma_{122} = 0 \\ \Gamma^1_{12} &= \Gamma^1_{21} = g^{11}\Gamma_{121} + g^{12}\Gamma_{122} = 1/(2\beta) \\ \Gamma^2_{22} &= g^{21}\Gamma_{221} + g^{22}\Gamma_{222} = 1/2 \phi'(\beta)/\phi(\beta)\end{aligned}$$

O tensor de Riemann

$$R_{1212} = -[\phi(\beta) + \beta\phi'(\beta)]/[4\mu^2\beta\phi(\beta)],$$

e a curvatura seccional

$$K = -R_{1212} g^{11} g^{22} = 1/4 [\phi(\beta) + \beta\phi'(\beta)]/[\beta^2\phi(\beta)]$$

2.3. Algumas aplicações numéricas

2.3.1 Conexão entre aproximação de Laplace e tensores associados

1. Distribuição Binomial

Se a verossimilhança é parametrizada em termos da probabilidade de sucesso θ , temos:

$$L(\theta | x) = \theta^x (1-\theta)^{n-x}$$

Considerar uma priori de Jeffreys (ver por exemplo, Box e Tiao, 1973) dada por,

$$\pi(\theta) \propto \theta^{-1/2} (1-\theta)^{-1/2}$$

A média exata a posteriori é dada por

$$E(\theta | x) = (x+1/2)/(n+1)$$

Uma aproximação de Laplace na parametrização original é dada por

$$(n-1)^{n+1/2} (x+1/2)^{x+1} / [n^{n+3/2} (x-1/2)^x]$$

Uma possível reparametrização para θ é dada por $\arcsen(\theta^{1/2})$. A verossimilhança nesta reparametrização é dada por

$$L(\phi | x) = (\sin^2(\phi))^x (\cos^2(\phi))^{n-x}$$

Neste caso a priori de Jeffreys é dada por $\pi(\phi) \propto$ constante, assim a aproximação de Laplace para a média a posteriori é dada por:

$$n^{n+1/2} (x+1)^{x+1} / [(n+1)^{n+3/2} x^x]$$

Na Figura 1, tem-se um gráfico dos erros em porcentagem de $E(\theta | x)$ versus x na parametrização θ com $n=18$ e na figura 2 tem-se um gráfico dos erros em porcentagem de $E(\theta | x)$ na parametrização ϕ para $n=18$.

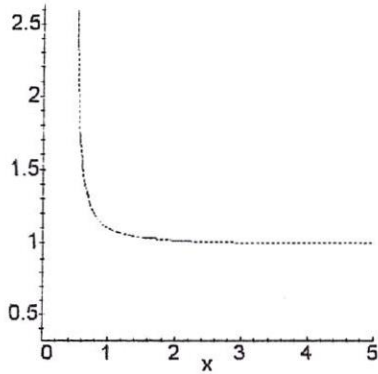


Figura 1: Erros percentuais de $E(\theta | x)$ para a parametrização θ

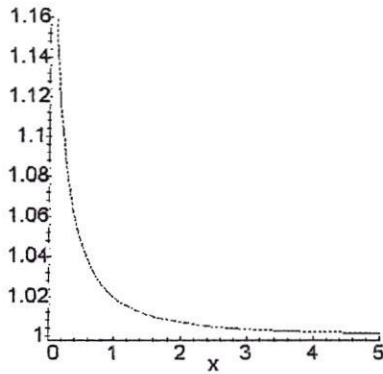


Figura 2: Erros percentuais de $E(\theta | x)$ para a reparametrização ϕ

Notamos que para $n=18$, a parametrização ϕ é mais satisfatória pois os erros percentuais são menores nesta parametrização.

A seguir, calculamos os símbolos de Cristofell de segunda espécie. Como temos uma distribuição uniparamétrica, a imergimos em uma bivariada, onde a segunda distribuição tem parâmetro distinto da binomial.

O símbolo de Cristofell de interesse é dado por:

$$(ijk) = \Gamma_{ij}^k = g^{kl} \Gamma_{ijl}$$

$$\Gamma_{ijk} = 1/2(E_i(g_{jk}) + E_j(g_{ik}) - E_k(g_{ij}))$$

Para as parametrizações analisadas temos:

$$(111)_{\theta} = [-2x + 6x\theta - 6x\theta^2 + 1 - 3\theta + 3\theta^2 - 2\theta^3 + 2\theta^3n] / [\theta(\theta-1)(2x-4x\theta-1+2\theta-2\theta^2+2\theta^2n)]$$

$$(111)_{\phi} = [-\sin^4(\phi) + 2\sin^4(\phi)n - 2\sin^4(\phi)x - 2\cos^4(\phi)x + \cos^4(\phi)] / [\sin(\phi)\cos(\phi)(2\sin^2(\phi)n - 2\sin^2(\phi)x - \sin^2(\phi) + 2\cos^2(\phi)x - \cos^2(\phi))]$$

Simulação com $\theta=0,3$ (ver figura 3) mostra $(111)_{\phi}$ complexo, mas suas componentes tanto real como imaginária se mostram menores que $(111)_{\theta}$, indicando se

tratar de uma melhor reparametrização. Tal resultado se mostra coerente com os obtidos pela aproximação de Laplace.

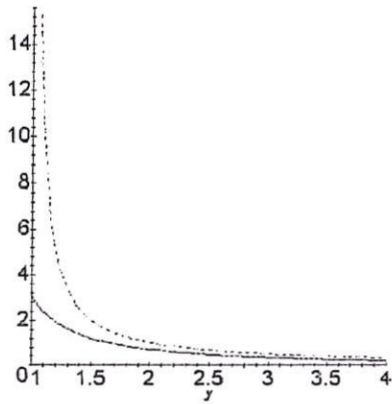


Figura 3: $\text{Re}(111)_\theta$ e $\text{Re}(111)_\phi$ com $\theta=0,3$

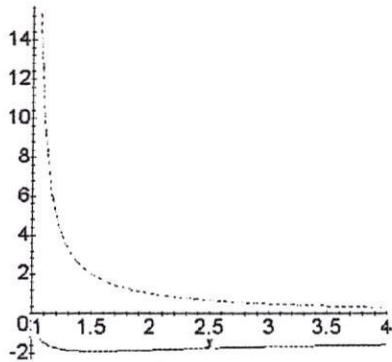


Figura 3.3: $\text{Re}(111)_\theta$ e $\text{Im}(111)_\phi$ com $\theta=0,3$

O cálculo das curvaturas geométricas usuais para o caso unidimensional, fornecem bons resultados, como observamos na figura 5:

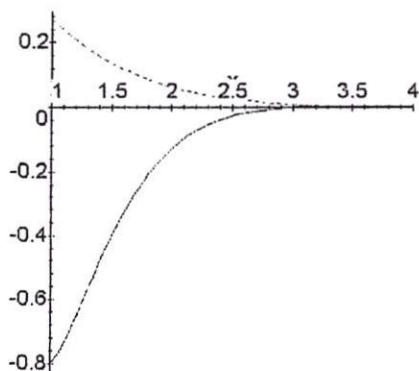


Figura 5. Curvaturas geométricas usuais para o caso Binomial nas parametrizações θ e ϕ

Nas figuras 3,4,5 a curva pontilhada é referente à θ e a linha contínua é referente à ϕ . Concluímos que a parametrização ϕ é melhor.

Um Exemplo bidimensional

2.Componentes de Duas Variâncias

Considere o modelo de efeitos aleatórios $x_{jk} = \theta + e_j + e_{jk}$, $j=1, \dots, J$, $k=1, \dots, K$, onde $E(e_j) = E(e_{jk}) = 0$, $\text{var}(e_j) = \sigma^2_2$ e $\text{var}(e_{jk}) = \sigma^2_1$. Assumir uma priori não informativa para θ, σ^2_1 e σ^2_2 dada por

$$\pi(\theta, \sigma^2_1, \sigma^2_2) \propto \sigma^{-2}_1 (\sigma^2_1 + K\sigma^2_2)^{-1}$$

(Ver, por exemplo, Box e Tiao(1973)). A densidade a posteriori é dada por

$$\pi(\theta, \sigma^2_1, \sigma^2_2) \propto (\sigma^2_1)^{-(v_1/2+1)} (\sigma^2_1 + K\sigma^2_2)^{-(v_2/2+1)} \exp[-1/2(v_1 m_1 / \sigma^2_1 + v_2 m_2 / (\sigma^2_1 + K\sigma^2_2))]$$

onde $v_1 = J(K-1)$, $v_2 = J-1$, $v_1 m_1 = S_1$, $v_2 m_2 = S_2$, $S_1 = \sum_j \sum_k (x_{jk} - x_{j.})^2$, $S_2 = K \sum_j (x_{j.} - x_{..})^2$.

Em Achcar e Smith (1990), temos uma comparação dos erros nas aproximações de Laplace para várias funções de interesse:

	Exata	(σ^2_1, σ^2_2)	$(\ln(\sigma^2_1), K^{-1} \ln(\sigma^2_1 + K\sigma^2_2))$
$E(\sigma^2_{12} / \sigma^2_1 x)$	0	19,80	2,53
$E(\sigma^2_{12} x)$	0	20,59	2,54
$E(\sigma^2_1 x)$	0	0,27	0,27
$E(\sigma^2_2 x)$	0	510,73	0,20

Tabela 1. Erros nas aproximações de Laplace para caso de Duas Variâncias ($K=5$; $S_1=58830,00$; $v_1=24$; $m_1=2451,25$; $S_2=56357,00$; $v_2=5$; $m_2=11271,50$)

Conclui-se que a reparametrização considerada reduz de forma acentuada o erro. Neste caso, os tensores são dados por:

a) Parametrização canônica (σ^2_1, σ^2_2) :

$$R_{1212} = 0,3435266989 \cdot 10^{-25}$$

b) Reparametrização $(\ln(\sigma^2_1), K^{-1} \ln(\sigma^2_1 + K\sigma^2_2))$

$$R_{1212} = 0$$

	(σ^2_1, σ^2_2)	$(\ln(\sigma^2_1), K^{-1} \ln(\sigma^2_1 + K\sigma^2_2))$
K	$.1390304059 \cdot 10^{-7}$	0

Tabela 2. escalar de Ricci para o caso de componentes de duas Variâncias

Um fato notável ocorre: as expressões matemáticas dos símbolos de Christoffel da reparametrização são muito simplificados e não dependem dos parâmetros; podemos supor que esse fato seja o principal responsável pela alta precisão obtida pela melhor reparametrização.

A tabela 2 mostra os valores do escalar de Ricci nas duas parametrizações. Mais uma vez o menor escalar de Ricci se apresenta na melhor parametrização dada por $(\ln(\sigma^2_1), K^{-1} \ln(\sigma^2_1 + K\sigma^2_2))$

3. Conexão entre o tensor de Riemann e dados censurados Distribuição Weibull e Gumbel

Considerar o conjunto de dados de sobrevivência da tabela 3:

[39,4;45,3;49,2;49,4;51,3;52,0;53,2;53,2;54,9;55,5;57,1;57,2;57,5;59,2;61,0;62,4;63,8;64,3;67,3;67,7]

Tabela 3. Dados de sobrevivência de $n=20$, cabos elétricos submetidos a um teste industrial (valores de stress em KV/mm)

Para os dados da tabela 3, ajustamos uma distribuição Weibull com dois parâmetros. O logaritmo da função de verossimilhança é dado por,

$$l(t, \alpha, \beta) = r \ln(\beta) - r\beta \ln(\alpha) + (\beta - 1) \sum_{i \in D} \ln(t_i) - \sum_{i=1}^n (t_i / \alpha)^\beta \quad (2.3.1)$$

onde n é o tamanho amostral, r é o número de falhas observadas e D é o conjunto com falhas observadas

Os dados foram censurados (censura tipo 1) sucessivamente a começar do último. Calculamos os estimadores de máxima verossimilhança para os parâmetros da distribuição Weibull por Newton-Raphson. Os resultados estão na tabela 4.

Número de censuras	β^+	α^+
0	9,1410	59,1245
1	8,6368	59,3190
2	8,1393	59,5754
3	7,7378	59,9238
4	7,3393	60,3533
5	6,9741	60,8825
6	6,6402	61,5246
7	6,3510	62,2909
8	6,1030	63,1917
9	5,8413	64,2769
10	5,5544	65,6176
11	5,3001	67,2268
12	5,0364	69,2455
13	4,8081	71,7057
14	4,5295	75,0517

Tabela 4. Parâmetros da Weibull ajustada aos dados da tabela (3)

Calculamos, a partir desses valores o tensor de Riemann e do escalar de Ricci:

Número de censuras	R_{1212}	K
0	-1327478336 e8	-.5931666930e-9
1	-4018560.208e7	-.2183482878e-8
2	-1236043.169e7	-.8081946598e-8
3	-472914.4965	-.2428765857e-7
4	-182382.3142	-.7446953960e-7
5	-75967.57233	-.2173875275e-6
6	-34048.11220	-.6106008265e-6
7	-16933.93620	-.1604671907e-5
8	-9284.570148	-.4010053465e-5
9	-4996.745160	-.1127834733e-4
10	-2608.037973	-.3876128425e-4
11	-1524.669613	-.1477863129e-3
12	-967.1168818	-.8574037853e-3
13	-863.5189383	-.1103966913e-1
14	2041.789140	4.910320893

Tabela 5. Tensores de Riemann e escalar de Ricci para o caso de dados com censuras na Weibull

É usual o estatístico trabalhar com tempos na escala logarítmica. Trata-se da parametrização Gumbel ou de valor extremo:

$$x = \ln(t)$$

$$u = \ln(\alpha)$$

$$b = 1/\beta$$

$$w_j = (x_j - u)/b$$

$$f(x_j; u, b) = 1/b \exp[(w_j) - \exp(w_j)]$$

Na reparametrização Gumbel temos os seguintes valores do tensor de Riemann e do escalar de Ricci

Número de censuras	R_{1212}	K
0	.1022650525e15	-.7796923450e-7
1	.2124366788e14	-.1635238900e-6
2	.4420286415e13	-.3429564385e-6
3	.1212664106e13	-.6329923715e-6
4	.3300531935e12	-.1167264845e-5
5	.9790905860e11	-.2058407487e-5
6	.3151488948e11	-.3480175755e-5
7	.1151076671e11	-.5537005850e-5
8	4731173678e10	-.8338452483e-5
9	1828756058e10	-.1282249444e-4
10	636815235.8e9	-.2042961922e-4
11	.2601258878e-3	-.1976520866e-3
12	86680759.28e8	-.4848218800e-4
13	34125372.80e8	-.7227065603e-4
14	10847722.89e8	-.1154189020e-3

Tabela 6: Tensores de Riemann e escalar de Ricci para distribuição Gumbel com o número crescente de censuras

Comparando-se os resultados das tabelas 5 e 6, concluímos que a parametrização Gumbel apresenta menores valores de curvatura do que a parametrização Weibull.

Outro fato interessante é que a curvatura em ambos os casos aumenta com o aumento da censura, apesar desse aumento não se apresentar de forma monótona.

3. Conclusões

A medida de curvatura escalar de Ricci mostra-se bastante adequada e coerente com os resultados da aproximação de Laplace: parametrizações que levam a melhores aproximações de Laplace têm curvaturas menores.

O escalar de Ricci mostra-se sensível ao número de censuras crescentes numa amostra. Amostras com mais censuras geram curvaturas maiores. No caso específico de dados analisados que se adequaram aos modelos Weibull e a Gumbell, a Gumbell mostrou curvaturas menores.

4. Referências

Referências

Achcar, J. A. e Smith, A. F.M. (1990). Aspects of Reparametrization in Approximative Bayesian Inference, Geisser, S. Hodges, J.S., Press, S.J. Zellner, A. (Editors), Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard, Elsevier Science Publisher, BV (North-Holland)

Amari, S.I. (1982). Differential Geometry of Curved Exponential families- Curvatures and Information Loss, Ann. Statist., 10, 375-385.

Anscombe, F.J. (1964). Normal Likelihood Functions, Annals of the Institute of Statistical Mathematics, 16, 1-19.

Bates, D.M. e Watts, D.G. (1980). Relative curvature measures of nonlinearity, J. Roy. Statist. Soc. B40, 1-25.

Bates, D.M. e Watts, D.G. (1988). Nonlinear Regression Analysis and Its Applications, Wiley, New York.

Barndorff-Nielsen, O. E. (1986). Likelihood and observed geometries. Ann. Statist. 14, 856-873.

Box, M.J. (1971). Bias in Nonlinear Estimation, Journal of the Royal Statistical Society, Series B, 33, 171-201.

Box, G.E.P. e G.C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley.

Cox, D.R. e N. Reid (1987). Parameter Orthogonality and Approximative Conditional Inference, J. R. Statist. Soc., B, Vol. 49, N 1, 1-39.

Hills, S. (1989). The Parameterisation of Statistical Models, Ph.D. Thesis, University of Nottingham.

Kass, R.E., Slate, E.H.(1992). Reparameterization of Statistical Models, Tese de Doutorado, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.

Ratkowsky, D. A. (1990). Handbook of Nonlinear Regression Models, Marcel Dekker, New York.

Seber, G.A .F. e Wild, C.J.(1989).Nonlinear Regression, John Wiley & Sons, New York.

Sprott, D.A.(1973).Normal likelihoods and their relation to large sample theory of estimation,Biometrika,60,457-465.

Sprott, D.A (1980). Maximum Likelihood in Small Samples: Estimation in the Presense of Nuisance Parameters, Biometrika, 67, 515-523.