

O princípio da equivariância: conceitos e aplicações

The Principle of Equivariance: Concepts and Applications

JUVÊNCIO NOBRE^{1,2,a}, CAIO AZEVEDO^{2,b}

¹DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA, UNIVERSIDADE FEDERAL DO CEARÁ, FORTALEZA, BRASIL

²INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, SÃO PAULO, BRASIL

Resumo

Neste trabalho apresentamos uma revisão do princípio da estimação equivariante e algumas de suas aplicações na família de localização-escala e em modelos lineares. Consideramos também o estimador não viciado de variância uniformemente mínima em modelos lineares. Vários exemplos são apresentados para ilustrar o uso destes métodos.

Palavras chave: Estimação equivariante, família de localização-escala, função de perda, modelos lineares, estimador não viciado de variância uniformemente mínima.

Abstract

In this work we present a review under the principle of equivariant estimation and their applications to the location-scale families and some linear models. We also consider the minimum variance unbiased estimation under the linear models framework. We show some examples to illustrate the use of those methods.

Key words: Equivariant estimation, Location-scale families, Loss function, Linear models, Minimum variance unbiased estimator.

^aProfessor assistente. E-mail: juvenio@ime.usp.br

^bAluno de doutorado do curso de estatística. E-mail: cnaber@ime.usp.br

1 Introdução

Inicialmente, considere um modelo estatístico $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ em que \mathcal{X} é o espaço amostral associado a um experimento, \mathbf{X} (vetor aleatório), \mathcal{A} é uma σ -álgebra de subconjuntos de \mathcal{X} e \mathcal{P} é uma família de medidas de probabilidades, \mathbb{P} , no espaço mensurável $(\mathcal{X}, \mathcal{A})$ (fixado). Em geral, assume-se que $\mathcal{P} = \{\mathbb{P}_\theta; \theta \in \Omega\}$ é indexada por um parâmetro (ou vetor de parâmetros) $\theta \in \Omega$, e que existe uma correspondência biunívoca entre Ω e \mathcal{P} (**identificabilidade**), com Ω denominado de espaço paramétrico. O objetivo da inferência estatística consiste em pesquisar sobre a distribuição geradora, isto é, “descobrir” qual distribuição $\mathbb{P}_{\theta_0} \in \mathcal{P}$ gera os dados em questão, ou equivalentemente estimar o valor de θ .

Considere $h : \Omega \rightarrow \mathbb{R}$ uma função mensurável, \mathbf{X} um vetor aleatório, cujo valor em θ tem-se interesse em estimar e $\delta(\mathbf{X}) : \mathcal{X} \rightarrow \mathbb{R}$ um estimador e $d = \delta(\mathbf{x})$ representando uma estimativa de $h(\theta)$. Um critério bastante utilizado para a escolha de estimadores ótimos é tomar um estimador $\delta(\mathbf{X})$ que minimiza o risco $R(\theta, \delta) := \mathbb{E}_\theta[L(\theta, \delta)]$, $\forall \theta \in \Omega$, com $L(\theta, \cdot)$ denotando uma função de perda apropriada. Dada a impossibilidade de se obter tal estimador (Lehmann & Casella 1998, pág. 5), é comum restringir a classe de estimadores, e determinar, dentro desta classe, um estimador que minimiza o risco uniformemente em θ . Desta forma, por exemplo, pode-se obter o ENVVUM (classe dos estimadores não viciados), BLUE (classe dos estimadores lineares), entre outros.

Neste trabalho, temos por objetivo apresentar a classe de estimadores equivariantes, com ênfase nos modelos de localização-escala e lineares, além de considerar a estimação não viciada de variância uniformemente mínima (NVVUM), nesta última classe de modelos. Na Seção 2, fornecemos alguns conceitos e definições requeridas no desenvolvimento do artigo. Na Seção 3, é discutida a estimação equivariante no modelo de escala, enquanto que na Seção 4 é analisado o modelo de localização-escala tanto de forma marginal como conjunta. Na Seção 5, apresentamos alguns resultados básicos sobre a estimação equivariante e NVVUM para modelos lineares.

2 Estrutura matemática do princípio da equivariância

Considere $\mathbf{X} : \mathcal{X} \rightarrow \mathbb{R}$ uma variável aleatória cuja respectiva distribuição pertence a família indexada por θ

$$\mathcal{P} = \{\mathbb{P}_\theta; \theta \in \Omega\} \quad (1)$$

e \mathcal{C} uma classe de funções bijetivas $g : \mathcal{X} \rightarrow \mathcal{X}$.

Definição 1.

- i) Considere $g \in \mathcal{C}$ e \mathbf{X} uma variável aleatória com distribuição $\mathbb{P}_\theta \in \mathcal{P}$. Se $\forall \theta \in \Omega$, a distribuição de $X^* := g(X)$, $\mathbb{P}_{\theta^*} \in \mathcal{P}$, diz-se que o modelo (1) é invariante sob a transformação g .

ii) Se i) vale $\forall g \in \mathcal{C}$, diz-se que o modelo (1) é invariante sobre a classe de transformações \mathcal{C} .

Considere \mathcal{C} uma classe de transformações sob a qual o modelo (1) é invariante. Perceba que \mathcal{C} não é necessariamente um grupo de transformações (fechada por composição e inversão). Definindo $G(\mathcal{C}) := \{g; g = g_1^{\pm 1} \circ \dots \circ g_m^{\pm 1}; g_i \in \mathcal{C}, i = 1, \dots, m\}$ com $g \circ h$ representando a composição das funções g e h , em que os elementos $g_i \in \mathcal{C}$ não são necessariamente distintos, tem-se que o modelo (1) é invariante sob $G(\mathcal{C})$, com $G(\mathcal{C})$ sendo o grupo (**gerado** por \mathcal{C}).

Considere $g \in G(\mathcal{C})$, então $g(X) \sim \mathbb{P}_{\theta^*} \in \mathcal{P}$. Pode-se mostrar que $\theta^* := \bar{g}(\theta) : \Omega \rightarrow \Omega$ é uma transformação bijetiva e que $\bar{G} := \{\bar{g}; g \in G\}$ é um grupo de transformações. Para demonstrar a primeira assertiva, considere $\forall g \in G$ que $X_i \sim \mathbb{P}_{\theta_i} \in \mathcal{P}$ e $g(X_i) \sim \mathbb{P}_{\theta_i^*} \in \mathcal{P}$ ($i = 1, 2$) tal que $\mathbb{P}_{\theta_1^*}(A) = \mathbb{P}_{\theta_2^*}(A)$, $\forall A \in \mathcal{B}(\mathbb{R}) \Leftrightarrow \mathbb{P}_{\theta_1}(g^{-1}(A)) = \mathbb{P}_{\theta_2}(g^{-1}(A))$, $\forall A \in \mathcal{B}(\mathbb{R}) \Leftrightarrow \theta_1 = \theta_2$ e o resultado segue. Para provar a segunda assertiva, mostre que $\overline{(g_1 \circ g_2)} = \overline{(g_1)} \circ \overline{(g_2)}$ e $\overline{(g^{-1})} = \bar{g}^{-1}$, $\forall g_1, g_2, g \in G$ e use o fato de que G é um grupo.

Adicionalmente, segue diretamente da definição de $\bar{g}(\theta)$ que

$$\mathbb{P}_\theta[g(X) \in A] = \mathbb{P}_{\theta^*}[Y \in A] \tag{2}$$

$$\mathbb{E}_\theta[\psi(g(X))] = \mathbb{E}_{\theta^*}[\psi(Y)] \tag{3}$$

para qualquer função ψ , \mathbb{P}_{θ^*} integrável.

Considere o problema de estimar $h(\theta)$ no modelo (1) que é assumido ser invariante sob as transformações $X^* = g(X)$ e $\theta^* = \bar{g}(\theta)$, $g \in G$. Iremos supor também que $\forall \bar{g} \in \bar{G}$, $h(\theta^*)$ dependa de θ somente através de $h(\theta)$, ou seja

$$h(\theta^*) = g^*(h(\theta)) \tag{4}$$

Desta forma pode-se relacionar a estimativa d de $h(\theta)$ com a estimativa d^* de $h(\theta^*)$ da seguinte forma

$$d^* = g^*(d) \tag{5}$$

implicando que o problema de estimar $h(\theta)$ em termos de (X, θ, d) ou $h(\theta^*)$ em termos de (X^*, θ^*, d^*) representam a mesma situação física apenas expressa em um novo sistema de coordenadas. A forma da função a ser estimada tem um papel fundamental nas considerações que serão discutidas adiante.

Exemplo 1. Duas amostras da família de localização.

Considere $\mathbf{X} = (X_1, \dots, X_m)^\top$ e $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, dois vetores aleatórios com respectiva densidade conjunta

$$f(\mathbf{x} - \xi, \mathbf{y} - \eta) = f(x_1 - \xi, \dots, x_m - \xi, y_1 - \eta, \dots, y_n - \eta), \quad \xi, \eta \in \mathbb{R} \tag{6}$$

Este modelo permanece invariante sob as transformações

$$g(\mathbf{X}, \mathbf{Y}) = (\mathbf{X} + a, \mathbf{Y} + b), \quad \bar{g}(\xi, \eta) = (\xi + a, \eta + b) \tag{7}$$

para quaisquer escalares a e b . Considere que o interesse é estimar $h(\xi, \eta) = \Delta := \eta - \xi$. Denotando as variáveis e os parâmetros transformados por $\mathbf{X}^* = \mathbf{X} + a$, $\mathbf{Y}^* = \mathbf{Y} + b$, $\eta^* = \eta + b$ e $\xi^* = \xi + a$ então, tem-se que as transformações em (7) levam Δ em $\Delta^* = \eta^* - \xi^* = \Delta + (b - a)$. Portanto, dada uma estimativa de Δ , digamos d , obtida via modelo (6), tem-se que a estimativa de Δ^* , digamos d^* , no modelo transformado pode ser expressa como $d^* = d + (b - a) = g^*(d)$.

Suponha agora que o interesse é estimar $h(\xi, \eta) = \lambda := \xi^2 + \eta^2$. Considerando as transformações em (7), tem-se que λ é transformado em $\lambda^* = (\xi + a)^2 + (\eta + b)^2 = \lambda + \phi(\xi, \eta, a, b)$ ou seja, λ^* não depende de (η, ξ) somente através de λ . Neste caso o problema de estimar λ , via modelo original, e estimar λ^* , via modelo transformado, não representam a mesma situação.

Sob a veracidade de (4), tem-se que os problemas de estimar $h(\theta)$ em termos de (X, θ, d) ou $h(\theta^*)$ em termos de (X^*, θ^*, d^*) são equivalentes; então é razoável que a função de perda seja tal que $L(\theta, d) = L(\theta^*, d^*)$, ou seja, que a função de perda seja **invariante** sob a transformação g [uma caracterização de funções de perda invariantes é dada em Staudte (1971)]. Tal observação conduz à seguinte definição:

Definição 2. Se o modelo estatístico (1) é invariante sob g , a função de perda L satisfaz

$$L(\bar{g}(\theta), g^*(d)) = L(\theta, d) \quad (8)$$

e $h(\theta)$ satisfaz (4), então o problema de estimar $h(\theta)$ com função de perda L é dito ser invariante sobre g .

Em um problema invariante, se δ é um estimador de $h(\theta)$, então existem dois caminhos naturais de se estimar $h(\theta^*)$ (o estimando no modelo transformado), apresentados a seguir.

1. Princípio da equivariância funcional

Se $\delta(\mathbf{X})$ é o estimador de $h(\theta)$, então o estimador de $\phi(h(\theta))$ é dado por $\phi(\delta(\mathbf{X}))$. Fazendo $\phi = g^*$, tem-se que $g^*(\delta(\mathbf{X}))$ é o estimador de $g^*(h(\theta))$, quando $\delta(\mathbf{X})$ for usado para estimar $h(\theta)$.

2. Princípio da invariância formal

Invariância sob as transformações g, \bar{g} e g^* no problema de estimação de $h(\theta)$ significa essencialmente dizer que os problemas de estimar $h(\theta)$ em termos de \mathbf{X}, θ e d^* , e o de estimar $g^*(\theta)$ em termos de \mathbf{X}^*, θ^* e d^* são formalmente o mesmo e, por conseguinte, devem ser tratados da mesma forma. Isto significa que $\delta(\mathbf{X}^*) = \delta(g(\mathbf{X}))$ deve ser usado para estimar $g^*(h(\theta)) = h(\theta^*)$.

É desejável que os dois princípios nos levem ao mesmo estimador, ou seja, que

$$\delta(g(\mathbf{X})) = g^*(\delta(\mathbf{X})). \quad (9)$$

Definição 3. Em um problema de estimação invariante, um estimador $\delta(\mathbf{X})$ é dito ser **equivariante** se ele satisfaz (9), $\forall g \in G$.

Os princípios de equivariância funcional e invariância formal têm sido discutidos por alguns autores, utilizando diferentes denominações. Por exemplo, Casella & Berger (2002, p. 297) denotam por **Princípio de medida equivariante** ao invés de **Princípio da equivariância funcional**. Algumas outras denominações podem ser encontradas em Lehmann & Casella (1998, p. 233). Além disso, alguns autores destacam a diferença entre equivariância, em que as estimativas dos parâmetros se modificam em um determinado sentido quando os dados são transformados, e invariância, na qual as estimativas ficam imutáveis sob transformações. Para detalhes, veja Schervish (1995, p. 344), Borovkov (1998, p. 166), Lehmann & Casella (1998, p. 150) e Casella & Berger (2002, p. 296), por exemplo.

Exemplo 2. Família de localização.

Considere $\mathbf{X} = (X_1, \dots, X_n)^\top$ um vetor aleatório com densidade dada por

$$f(\mathbf{x} - \xi) = f(x_1 - \xi, \dots, x_n - \xi), \quad \xi \in \mathbb{R}$$

O modelo acima é invariante sob as seguintes transformações

$$\mathbf{X}^* = \mathbf{X} + a \text{ e } \xi^* = g^*(\xi) = \xi + a, \quad \forall a \in \mathbb{R}$$

o problema de estimação de ξ é invariante sobre as transformações acima se consideramos funções de perda da forma $L(\xi + a, d + a) = L(\xi, d), \forall a \in \mathbb{R}$, e isto ocorre se e somente se $L(\xi, d) = \rho(d - \xi)$ (Lehmann & Casella 1998, p. 149). Neste caso, um estimador $\delta(\mathbf{X})$ é equivariante (por localização) se e somente se

$$\delta(g(\mathbf{X})) = \delta(\mathbf{X} + a) = \delta(\mathbf{X}) + a = g^*(\delta(\mathbf{X})), \quad \forall a \in \mathbb{R} \tag{10}$$

Exemplo 3. Continuação do exemplo 1.

No exemplo 1, tínhamos $h(\xi, \eta) = \Delta = \eta - \xi$ e $g^*(d) = d + (b - a)$. Considerando uma função de perda invariante sobre as transformações (7), então tem-se um problema de estimação (de Δ) invariante. Neste caso, um estimador $\delta(\mathbf{X}, \mathbf{Y})$ é equivariante se e somente se

$$\delta(\mathbf{X} + a, \mathbf{Y} + b) = \delta(\mathbf{X}, \mathbf{Y}) + (b - a). \tag{11}$$

Se $\delta_1(\mathbf{X})$ e $\delta_2(\mathbf{Y})$ são estimadores equivariantes por localização (da forma (10)) de ξ e η , respectivamente, então $\delta(\mathbf{X}, \mathbf{Y}) = \delta_2(\mathbf{Y}) - \delta_1(\mathbf{X})$ é um estimador equivariante de Δ .

A seguir são obtidas algumas propriedades dos estimadores equivariantes.

Teorema 1. *Considere $\delta(\mathbf{X})$ um estimador equivariante em um problema invariante sob a transformação g , então a função de risco satisfaz*

$$R(\bar{g}(\theta), \delta(g(\mathbf{X}))) = R(\theta, \delta), \quad \forall \theta \in \Omega. \tag{12}$$

Demonstração. Por definição e lembrando (3), segue que

$$\begin{aligned} R(\bar{g}(\theta), \delta) &= \mathbb{E}_{\bar{g}(\theta)}[L(\bar{g}(\theta), \delta(\mathbf{X}))] = \mathbb{E}_\theta[L(\bar{g}(\theta), \delta(g(\mathbf{X}))) \\ &= \mathbb{E}_\theta[L(\bar{g}(\theta), g^*(\delta(\mathbf{X})))] = \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X}))] := R(\theta, \delta) \quad \square \end{aligned}$$

Definição 4. Um grupo de transformações G de Ω é dito ser transitivo se $\forall \theta_1, \theta_2 \in \Omega, \exists g \in G$ tal que $g(\theta_1) = \theta_2$.

O corolário seguinte é útil para generalizar o teorema 1.4 (Lehmann & Casella 1998, p. 150) para o problema de estimação equivariante por localização.

Corolário 1. Sob as suposições do teorema 1 e considerando \overline{G} transitiva sob o espaço paramétrico Ω , então tem-se que a função de risco de qualquer estimador equivariante é constante.

Demonstração. Pelo teorema 1, tem-se que $R(\overline{g}(\theta_1), \delta(g(\mathbf{X}))) = R(\theta_1, \delta)$, $\forall \theta_1 \in \Omega$. Sob a suposição de transitividade de \overline{G} , temos que $\forall \theta_1, \theta_2 \in \Omega, \exists \overline{g}_{12} \in \overline{G}$; $\overline{g}_{12}(\theta_1) = \theta_2$, portanto

$$R(\theta_2, \delta) = R(\overline{g}_{12}(\theta_1), \delta(g_{12}(\mathbf{X}))) = R(\theta_1, \delta), \quad \forall \theta_1, \theta_2 \in \Omega \quad (13)$$

□

Quando o risco de qualquer estimador equivariante é constante, e supondo que exista um estimador equivariante com risco finito, o melhor estimador equivariante δ^* , no sentido de minimizar o risco, denominado EERM (EERM-Estimador Equivariante de Risco Mínimo), é obtido minimizando tal constante. Uma forma de se obter δ^* é encontrar uma função da amostra \mathbf{X} , digamos $(T, W)^\top$, em que T é uma estatística suficiente e W é uma estatística ancilar, ambas para θ . Desta forma, podemos obter δ^* minimizando em δ a seguinte esperança condicional

$$\mathbb{E}_\theta[L(\theta, \delta(\mathbf{X})) \mid W = w] \quad (14)$$

uma vez que

$$\begin{aligned} R(\theta, \delta) &:= \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X}))] = \mathbb{E}_\theta[\mathbb{E}_\theta[L(\theta, \delta(\mathbf{X})) \mid W = w]] \\ &= \int_{\mathbb{R}^k} \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X})) \mid W = w] d\mathbb{P}(w) \\ &\geq \int_{\mathbb{R}^k} \min_{\delta} \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X})) \mid W = w] d\mathbb{P}(w) \\ &= \min_{\delta} \mathbb{E}_\theta[L(\theta, \delta(\mathbf{X})) \mid W = w] \\ &= \min_{\delta} \mathbb{E}_\theta[L(\theta, \delta(T, w)) \mid W = w] \end{aligned} \quad (15)$$

em que k representa a dimensão da estatística ancilar W .

Exemplo 4. Continuação do exemplo 1.

Neste exemplo, tem-se que $\boldsymbol{\theta} = (\xi, \eta)^\top$ e $\overline{g}(\boldsymbol{\theta}) = (\xi + a, \eta + b)$. Esse grupo \overline{G} é claramente transitivo sob $\Omega = \mathbb{R}^2$, dado que $\forall (\xi, \eta)$ e $(\xi^*, \eta^*) \in \mathbb{R}^2, \exists a, b \in \mathbb{R}$ tais que $\xi^* = \xi + a$ e $\eta^* = \eta + b$. Por conseguinte o EERM pode ser obtido através de (15), por exemplo.

Algumas propriedades adicionais, tanto no contexto clássico quanto no Bayesiano, dos estimadores equivariantes, podem ser encontradas em Zacks (1971), Schervish (1995) e Borovkov (1998), por exemplo.

3 Modelos de escala

Nesta seção, vamos aplicar os princípios desenvolvidos na seção anterior para o modelo de escala. Considere que $\mathbf{X} = (X_1, \dots, X_n)^\top$ pertence à família de escala, ou seja, que sua densidade é da forma

$$\frac{1}{\tau^n} f\left(\frac{\mathbf{X}}{\tau}\right) := \frac{1}{\tau^n} f\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}\right), \quad \tau \in (0, \infty) := \mathbb{R}^{++} \quad (16)$$

em que f é uma função conhecida e τ é dito ser **um parâmetro de escala**. O modelo (16) é invariante sob as transformações

$$\mathbf{X}^* = b\mathbf{X}, \quad \tau^* = b\tau, \quad \forall b > 0 \quad (17)$$

Suponha que o interesse seja estimar $h(\tau) = \tau^r$, $r \in \mathbb{N}$. Perceba que (4) é satisfeita uma vez que (17) induz as transformações

$$h(\tau) \rightarrow b^r \tau^r = b^r h(\tau) \quad \text{e} \quad d^* = b^r d$$

a função de perda é invariante sob estas transformações se

$$L(b\tau, b^r d) = L(\tau, d), \quad \forall b > 0 \quad (18)$$

e isto ocorre se e somente se

$$L(\tau, d) = \phi\left(\frac{d}{\tau^r}\right) \quad (19)$$

Para mostrar que (18) implica (19) basta fazer $b = \tau^{-1}$, já o recíproco não é difícil de verificar.

Exemplo 5. Funções de perda invariantes por escala.

Exemplos de função de perda que satisfazem (19) são

$$L(\tau, d) = \frac{(d - \tau)^2}{\tau^{2r}} = \left(\frac{d}{\tau^r} - 1\right)^2 \quad \text{e} \quad L(\tau, d) = \frac{|d - \tau^r|}{\tau^r} = \left|\frac{d}{\tau^r} - 1\right| \quad (20)$$

porém, a perda quadrática não é da forma (19).

A seguinte definição segue diretamente de (9).

Definição 5. Um estimador δ de $h(\tau) = \tau^r$ é dito ser equivariante sob as transformações (17), ou **equivariante por escala** se

$$\delta(g(\mathbf{X})) = \delta(b\mathbf{X}) = b^r \delta(\mathbf{X}) = g^*(\delta(\mathbf{X})), \quad \forall b > 0 \quad (21)$$

Exemplo 6. Estimadores equivariantes por escala.

A maioria dos estimadores usuais de τ (parâmetro de escala) são equivariantes por escala, por exemplo, o desvio padrão, o desvio médio, a amplitude e o estimador de MV.

Como o grupo \overline{G} de transformações $\tau^* = b\tau$, $b > 0$, é transitivo sobre $\Omega = \mathbb{R}^{++}$ então, pelo Corolário 1, tem-se que o risco de qualquer estimador equivariante por escala é constante. A seguir, caracterizamos os estimadores equivariantes por escala.

Teorema 2. *Seja \mathbf{X} um vetor aleatório com densidade (16) e seja $\delta_0(\mathbf{X})$ um estimador equivariante por escala de δ^r . Então, se*

$$Z_i := \frac{X_i}{|X_n|} \quad (i = 1, \dots, n) \quad (22)$$

com $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, então uma condição necessária e suficiente para que um estimador δ satisfaça (21) é que exista uma função $w(\mathbf{z})$ tal que

$$\delta(\mathbf{x}) = \frac{\delta_0(\mathbf{x})}{w(\mathbf{z})} \quad (23)$$

Demonstração. Uma condição necessária e suficiente para que δ satisfaça (21) é que seja escrito na forma

$$\delta(\mathbf{x}) = \frac{\delta_0(\mathbf{x})}{u(\mathbf{x})} \quad (24)$$

com $u(\mathbf{x})$ sendo uma função **invariante** por escala, ou seja,

$$u(b\mathbf{x}) = u(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \quad \text{e} \quad \forall b > 0 \quad (25)$$

se $\delta_0(\mathbf{x})$ e u são dados como acima, então tem-se que $\forall \mathbf{x} \in \mathbb{R}^n$ e $\forall b > 0$:

$$\delta(b\mathbf{x}) = \frac{\delta_0(b\mathbf{x})}{u(b\mathbf{x})} = b^r \frac{\delta_0(\mathbf{x})}{u(\mathbf{x})} = b^r \delta(\mathbf{x})$$

satisfazendo (21). Supondo que δ é um estimador equivariante por escala, seja $u(\mathbf{x}) = \delta_0(\mathbf{x})/\delta(\mathbf{x})$, então a função u é invariante por escala. Por conseguinte, uma condição necessária e suficiente para δ seja um estimador equivariante é que ele seja escrito da forma (24). Para terminar a demonstração, vamos mostrar agora que a função u é invariante por escala se e somente se u for função de \mathbf{z} .

Fazendo $b = |x_n|^{-1}$ em (25) tem-se que $u(\mathbf{x}) = u(z_1, \dots, z_n) = w(\mathbf{z})$, $\forall \mathbf{x} \in \mathbb{R}^n$, implicando-nos que $\forall b > 0$, $u(b\mathbf{x}) = w(b\mathbf{z}) = u(bz_1, \dots, bz_n) = u(x_1, \dots, x_n) = u(\mathbf{x})$, o que prova o teorema. \square

Perceba que as componentes de \mathbf{Z} no teorema (2) só estão definidas se $X_n \neq 0$, ou seja, estão bem definidas *q.c.[P]*. Além do mais, tem-se que $w(\mathbf{Z})$ é uma **estatística ancilar** para a família (16) ou, equivalentemente, para τ , pois $w(\mathbf{Z})$ é uma função **invariante por escala**.

Teorema 3. *Seja \mathbf{X} um vetor aleatório com densidade (16) e seja \mathbf{Z} o vetor aleatório cujo componentes são dados em (22). Suponha que a função de perda é da forma (19) e que existe um estimador equivariante por escala, δ_0 , de τ^r com risco finito. Assuma que $\forall \mathbf{Z}$, existe uma função $w(\mathbf{Z}) = w^*(\mathbf{Z})$ que minimiza*

$$\mathbb{E}_{\tau=1} [\phi(\delta_0(\mathbf{X})/w(\mathbf{Z})) \mid \mathbf{Z}] \quad (26)$$

Então, um EERM por escala δ^* de τ^r existe e é dado por

$$\delta^*(\mathbf{X}) = \frac{\delta_0(\mathbf{X})}{w^*(\mathbf{Z})} \tag{27}$$

Demonstração. Seja δ_0 um estimador equivariante por escala de τ^r , então pelo teorema 2 tem-se que uma caracterização dos estimadores equivariantes por escala é

$$\delta(\mathbf{X}) = \frac{\delta_0(\mathbf{X})}{w(\mathbf{Z})}$$

Dada a invariância do problema de estimação de τ^r e a transitividade de \overline{G} , tem-se que o risco de $\delta(\mathbf{X})$ independe de τ . Por conseguinte

$$\begin{aligned} R(\tau^r, \delta) &= \mathbb{E}_\tau[L(\tau^r, \delta)] = \mathbb{E}_\tau[\phi(\delta_0(\mathbf{X})/(w(\mathbf{Z})\tau^r))] = \mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w(\mathbf{Z}))] \\ &= \mathbb{E}_{\tau=1}[\mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w(\mathbf{Z})) \mid \mathbf{Z}]] \\ &= \int_{\mathbb{R}^n} \mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w(\mathbf{Z})) \mid \mathbf{Z}] d\mathbb{P}_1(\mathbf{z}) \\ &\geq \int_{\mathbb{R}^n} \min_z \mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w(\mathbf{Z})) \mid \mathbf{Z}] d\mathbb{P}_1(\mathbf{z}) \\ &= \min_z \mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w(\mathbf{Z})) \mid \mathbf{Z}] = \mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X})/w^*(\mathbf{Z})) \mid \mathbf{Z}] \end{aligned}$$

implicando que o EERM por escala de τ^r é dado por (27). □

Por hipótese, tem-se que $\delta_0(\mathbf{X})$ tem risco finito, ou seja,

$$R(\tau^r, \delta_0) = \mathbb{E}_\tau[\phi(\delta_0(\mathbf{X})/\tau^r)] = \mathbb{E}_1[\phi(\delta_0(\mathbf{X}))] < \infty,$$

implicando que $\mathbb{E}_{\tau=1}[\phi(\delta_0(\mathbf{X}) \mid w(\mathbf{Z}))] < \infty$. Portanto, o procedimento anterior é válido.

Corolário 2. Sob as suposições do Teorema 3, e supondo que $\rho(\nu) = \phi(e^\nu)$ é convexa e não monótona, então existe um EERM por escala para τ^r e ele é único se ρ é estritamente convexa.

Demonstração. Veja Lehmann & Casella (1998, p. 169). □

Corolário 3. Sob as suposições do Teorema 3, se considerarmos

$$\phi\left(\frac{d}{\tau^r}\right) = \frac{(d - \tau^r)^2}{\tau^{2r}} \tag{28}$$

então

$$\delta^*(\mathbf{X}) = \frac{\delta_0(\mathbf{X})\mathbb{E}_1[\delta_0(\mathbf{X}) \mid \mathbf{Z}]}{\mathbb{E}_1[\delta_0^2(\mathbf{X}) \mid \mathbf{Z}]} \tag{29}$$

Demonstração. Basta mostrar que se X é uma variável aleatória positiva com $\mathbb{E}_\theta[X^2] < \infty$, então o valor de c que minimiza $\mathbb{E}[(X/c - 1)^2]$ é $c = \mathbb{E}[X^2]/\mathbb{E}[X]$. □

Corolário 4. *Sob as suposições do Teorema 3, se consideramos*

$$\phi\left(\frac{d}{\tau^r}\right) = \frac{|d - \tau^r|}{\tau^r}, \quad (30)$$

então $\delta^*(\mathbf{X})$ é dado por (27), com $w^*(\mathbf{Z})$ sendo qualquer **mediana-escalar** da distribuição condicional de $\delta_0(\mathbf{X})$ dado \mathbf{Z} com $\tau = 1$, isto é, $w^*(\mathbf{Z})$ satisfaz

$$\mathbb{E}[\mathbf{X}\mathbb{1}(\mathbf{X} \geq w^*(\mathbf{Z})) \mid \mathbf{Z}] = \mathbb{E}[\mathbf{X}\mathbb{1}(\mathbf{X} < w^*(\mathbf{Z})) \mid \mathbf{Z}] \quad (31)$$

em que $\mathbb{1}(x \in A)$ representa a função indicadora de x no conjunto A .

Demonstração. Basta mostrar que se X é uma variável positiva integrável, então o conjunto de valores de c que minimizam $\mathbb{E}|X - c|/|c|$ são os valores de c que satisfazem

$$\int_0^c x d\mathbb{P}(x) = \int_c^\infty x d\mathbb{P}(x). \quad \square$$

Exemplo 7. EERM por escala quando $n = 1$.

Suponha que $n = 1$ e que $X > 0$ q.c. Perceba que X^r satisfaz (21) e que neste caso $Z = 1$. Portanto, tem-se que todos os estimadores equivariantes por escala de τ^r são da forma X^r/w , com $w = w(1)$ sendo uma constante arbitrária. Supondo que X^r tem risco finito, então, pelo teorema 3 tem-se que o EERM por escala de τ^r é dado por X^r/w^* , em que w^* é qualquer constante que minimiza

$$\mathbb{E}_1[\phi(X^r/w)] := \mathbb{E}_{\tau=1}[\phi(X^r/w)] \quad (32)$$

Em particular, se a função de perda é dada por (28), tem-se que o EERM por escala de τ^r é dado por

$$\frac{X^r \mathbb{E}_1[X^r]}{\mathbb{E}_1[X^{2r}]} \quad (33)$$

Quando utilizamos (30) o EERM por escala de τ^r é dado por X^r/w^* com w^r representando qualquer mediana-escalar da distribuição de X^r para $\tau = 1$.

Exemplo 8. Distribuição $U(0, \tau)$.

Considere que X_1, \dots, X_n são variáveis aleatórias i.i.d. com distribuição $U(0, \tau)$, $\tau > 0$ e que o interesse é estimar τ . Um estimador equivariante por escala para τ é $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Além disso tem-se que $X_{(n)}$ é uma estatística suficiente e completa para τ . Dado que $\mathbf{Z} = (X_1/X_n, \dots, X_{n-1}/X_n, 1)$ é uma estatística ancilar, então, pelo teorema de Basu, tem-se que $X_{(n)}$ e \mathbf{Z} são independentes. Considerando a função de perda (28), tem-se que o EERM por escala para τ é dado por

$$\delta(\mathbf{X}) = \frac{X_{(n)} \mathbb{E}_1[X_{(n)}]}{\mathbb{E}_1[X_{(n)}^2]} = \frac{n+2}{n+1} X_{(n)}$$

que não coincide com o ENVVUM de τ , dado por $[(n+1)/n]X_{(n)}$, que é um estimador equivariante por escala.

Exemplo 9. Estimador equivariante para a variância duma distribuição normal com média conhecida.

Considere $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ e admita que o interesse é estimar σ^2 . Um estimador equivariante por escala para σ^2 é $\delta_0(\mathbf{X}) = \sum_{i=1}^n X_i^2$ que também é uma estatística suficiente e completa. Então, pelo teorema de Basu (vide, por exemplo, Lehmann & Casella 1998, p. 42), conclui-se que $\delta_0(\mathbf{X})$ e \mathbf{Z} são independentes, pois \mathbf{Z} é uma estatística ancilar. Desta forma, considerando a função de perda (28), temos que um ERRM por escala de σ^2 é

$$\delta(\mathbf{X}) = \frac{\delta_0(\mathbf{X})\mathbb{E}_1[\delta_0(\mathbf{X})]}{\mathbb{E}_1[\delta_0^2(\mathbf{X})]} \tag{34}$$

$$= \frac{1}{n+2} \sum_{i=1}^n X_i^2 \tag{35}$$

pois $\mathbb{E}_1[\delta_0(\mathbf{X})] = n$ e $\mathbb{E}_1[\delta_0^2(\mathbf{X})] = n(n+2)$. Neste caso, o ENVVUM de σ^2 é dado por $\sum_{i=1}^n X_i^2/n$.

Teorema 4. Sob as suposições do Teorema 3 com função de perda (28), o EERM por escala para τ é dado por

$$\delta^*(\mathbf{X}) = \frac{\int_0^\infty \tau^{-(n+2)} f(x_1/\tau, \dots, x_n/\tau) d\tau}{\int_0^\infty \tau^{-(n+3)} f(x_1/\tau, \dots, x_n/\tau) d\tau} \tag{36}$$

e, nesta forma, é chamado de estimador de Pitman de τ .

Demonstração. Veja Schervish (1995, p. 352). □

Lehmann & Casella (1998, p. 170) mostram a expressão do estimador de Pitman para τ^r .

Exemplo 10. Distribuição $\exp(\lambda^{-1})$.

Considere que X_1, \dots, X_n são variáveis aleatórias i.i.d. tais que

$$f(x, \lambda) = \lambda^{-1} e^{-x/\lambda} \mathbf{1}_{\mathbb{R}^+}(x), \quad \forall \lambda > 0$$

Neste caso, o estimador de Pitman de λ é dado por

$$\delta^*(\mathbf{X}) = \frac{\int_0^\infty (1/\lambda^{n+2}) e^{-\sum_{i=1}^n X_i/\lambda} d\lambda}{\int_0^\infty (1/\lambda^{n+3}) e^{-\sum_{i=1}^n X_i/\lambda} d\lambda}$$

Chamando $\alpha = \sum_{i=1}^n X_i/\lambda$, tem-se

$$\begin{aligned} \delta^*(\mathbf{X}) &= \frac{\int_0^\infty (\alpha/\sum_{i=1}^n X_i)^{n+2} e^{-\alpha} (\sum_{i=1}^n X_i/\alpha^2) d\alpha}{\int_0^\infty (\alpha/\sum_{i=1}^n X_i)^{n+3} e^{-\alpha} (\sum_{i=1}^n X_i/\alpha^2) d\alpha} = \sum_{i=1}^n X_i \frac{\int_0^\infty \alpha^n e^{-\alpha} d\alpha}{\int_0^\infty \alpha^{n+1} e^{-\alpha} d\alpha} \\ &= \sum_{i=1}^n X_i \frac{\Gamma(n+1)}{\Gamma(n+2)} = \frac{1}{n+1} \sum_{i=1}^n X_i \end{aligned}$$

que não coincide com o ENVVUM, que é dado por \bar{X} , que também é um estimador equivariante por escala; porém, o EERM possui risco uniformemente menor do que o ENVVUM para a função de perda (28).

Na próxima seção consideraremos o processo de estimação equivariante nos modelos de localização-escala.

4 Modelos de localização-escala

Nesta seção estudaremos o processo de construção dos estimadores equivariantes dos parâmetros de localização-escala considerando ambos os parâmetros desconhecidos. Salientamos que estaremos focados em apenas um dos parâmetros por vez.

Primeiramente, vamos introduzir a família de localização-escala. Consideramos que a densidade do vetor aleatório $\mathbf{X} = (X_1, \dots, X_n)^\top$ é

$$\frac{1}{\tau^n} f\left(\frac{x_1 - \xi}{\tau}, \dots, \frac{x_n - \xi}{\tau}\right) \quad (37)$$

em que o vetor de parâmetros $\theta = (\xi, \tau)^\top$ é desconhecido. Este problema permanece invariante sob as transformações,

$$\mathbf{X}_i^* = a + b\mathbf{X}_i, \quad \xi^* = a + b\xi, \quad \tau^* = b\tau, \quad i = 1, \dots, n \quad (b > 0) \quad (38)$$

Nas próximas duas seções apresentaremos o procedimento para a obtenção dos estimadores dos parâmetros de escala e de localização, respectivamente. Note que este grupo de transformações é transitivo, o que, com a escolha duma função de perda adequada, torna o risco constante com relação ao parâmetro de interesse.

4.1 Parâmetro de interesse é o de escala

Na Seção 3, os estimadores equivariantes por escala foram caracterizados como a razão entre um estimador equivariante por escala (função dum vetor aleatório pertencente à família de escala) e uma função dum vetor de estatísticas ancilares para τ . O desenvolvimento no presente caso é basicamente uma extensão daquele primeiro. Sob o grupo de transformações definido em (38), um estimador de τ^r será equivariante por escala se,

$$\delta(a + b\mathbf{X}) = b^r \delta(\mathbf{X}) \quad (39)$$

Sendo assim, temos que a classe dos estimadores equivariantes por escala pode ser descrita como

$$\delta(\mathbf{X}) = \frac{\delta_0(\mathbf{Y})}{w(\mathbf{Z})}$$

em que δ_0 um estimador equivariante por escala como em (39), $i = 1, \dots, n - 1$, $\mathbf{Y} = (Y_1, \dots, Y_{n-1})^\top$, $Y_i = X_i - X_n$, $\mathbf{Z} = (Z_1, \dots, Z_{n-1})^\top$

$$Z_i = \frac{Y_i}{|Y_{n-1}|}, \quad i = 1, \dots, n - 2, \quad Z_{n-1} = \frac{Y_{n-1}}{|Y_{n-1}|} \tag{40}$$

Além disso, a densidade de \mathbf{Y} possui estrutura da família de escala e \mathbf{Z} é uma estatística ancilar para $\boldsymbol{\theta}$ (Lehmann & Casella 1998, p. 168).

Segue então, do teorema 3, que o EERM para τ^r é dado por

$$\delta(\mathbf{X}) = \frac{\delta_0(\mathbf{Y})}{w^*(\mathbf{z})} \tag{41}$$

em que $w^*(\mathbf{z})$ é um número que minimiza (o risco)

$$E_{\tau=1}[\phi(\delta_0(\mathbf{Y})/w(\mathbf{z})) \mid \mathbf{Z} = \mathbf{z}] \tag{42}$$

Exemplo 11. EERM para a variância duma distribuição normal com média desconhecida.

Considere X_1, \dots, X_n uma amostra aleatória duma distribuição $N(\xi, \tau^2)$. Temos que $\mathbf{T} = (\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)^\top$ é uma estatística suficiente e completa para $\boldsymbol{\theta}$ e \mathbf{Z} é ancilar. Logo, pelo teorema de Basu, \mathbf{T} e \mathbf{Z} são independentes e, portanto, $\delta_0(\mathbf{X}) = \sum_{i=1}^n (X_i - \bar{X})^2$ e \mathbf{Z} também o são. Além disso, δ_0 é um estimador equivariante por escala [(39), com $r = 2$]. Portanto, considerando a função de perda $\phi(d/\tau^2) = [(d - \tau^2)^2]/\tau^4$, temos que,

$$w^*(\mathbf{z}) = \frac{\mathbb{E}_1[\delta_0^2(\mathbf{X})|\mathbf{Z}]}{\mathbb{E}_1[\delta_0(\mathbf{X})|\mathbf{Z}]} = \frac{\mathbb{E}_1[\delta_0^2(\mathbf{X})]}{\mathbb{E}_1[\delta_0(\mathbf{X})]} = \frac{(\frac{n-1}{2} + 1) (\frac{n-1}{2}) 4}{n - 1} = n + 1$$

pois $\delta_0(\mathbf{X})_{\tau=1} = \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2$. Portanto, o EERM de τ será

$$\delta(\mathbf{X}) = \frac{1}{n + 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Exemplo 12. Distribuição uniforme

Sejam X_1, \dots, X_n uma amostra aleatória duma distribuição $U(\xi - \frac{\tau}{2}, \xi + \frac{\tau}{2})$ e considere o problema de estimar τ com função de perda igual à do exemplo 11 [com $r = 1$]. Temos que $\mathbf{T} = (X_{(1)}, X_{(n)})^\top$ é suficiente e completa e, pelo teorema de Basu, é independente de \mathbf{Z} (Lehmann & Casella 1998). Além disso, $\delta_0 = X_{(n)} - X_{(1)}$ é um estimador equivariante por escala [sob (38)] para τ e também, é independente de \mathbf{Z} . Como $[X_{(n)} - X_{(1)}] \sim \beta(n - 1, 2)$ se $\xi = 0$ e $\tau = 1$, temos que,

$$w^*(\mathbf{Z}) = \frac{\mathbb{E}_1[\delta_0^2(\mathbf{X})]}{\mathbb{E}_1[\delta_0(\mathbf{X})]} = \frac{\frac{n(n-1)}{(n+2)(n+1)}}{\frac{n-1}{n+1}} = \frac{n}{n + 2}$$

Dessa forma, o EERM de τ será

$$\delta^*(\mathbf{X}) = \frac{n + 2}{n} (X_{(n)} - X_{(1)})$$

4.2 Parâmetro de interesse é o de localização

Tal como na seção anterior, o desenvolvimento aqui apresentado constitui, essencialmente, uma extensão daquele apresentado na Seção 1, cap. 3 de Lehmann & Casella (1998). As transformações definidas em (38), relacionadas aos espaços amostral e paramétrico, permanecem as mesmas. Contudo, a transformação relacionada ao estimador deve ser

$$\delta(a + b\mathbf{X}) = a + b\delta(\mathbf{X}) \quad (43)$$

Uma função de perda é invariante sob essas transformações se e somente se for da forma

$$L(\xi, \tau, d) = \rho\left(\frac{d - \xi}{\tau}\right) \quad (44)$$

Pela transitividade do grupo de transformações (38), a função de risco será constante [Seção 2].

Para um valor fixo de τ , seja

$$g_\tau = \frac{1}{\tau^n} f\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}\right)$$

de tal modo que (37) se torne

$$g_\tau(x_1 - \xi, \dots, x_n - \xi) \quad (45)$$

O lema 4.1 fornece um modo de obter EERM de ξ em certas situações.

Lema 1. *Suponha que para a família de localização (45) e função de perda (44), exista um EERM, digamos δ^* , considerando τ conhecido e que*

- i) δ^* não é função de τ , e
- ii) δ^* satisfaz (43).

Então δ^* é o EERM de ξ satisfazendo (43).

Demonstração. Como δ^* minimiza o risco, qualquer outro estimador terá risco maior ou igual a ele. Como isso vale $\forall \tau$, então o resultado segue. \square

Exemplo 13. EERM para a média da distribuição normal.

Sejam X_1, \dots, X_n uma amostra aleatória duma distribuição $N(\xi, \tau^2)$ com ambos os parâmetros desconhecidos. Lehmann & Casella (1998) encontram $\delta^* = \bar{X}$ como EERM de ξ com a variância conhecida, para qualquer função de perda (44) (convexa e par; note que, no caso considerado por esses autores, τ é uma constante conhecida). Além disso, δ^* satisfaz as suposições do Lema 1, pois não é função de τ e $\delta^*(a + b\mathbf{X}) = a + b\bar{X} = a + b\delta^*(\mathbf{X})$. Logo δ^* é EERM de ξ também neste caso.

Exemplo 14. Parâmetro de localização da distribuição uniforme.

Seja X_1, \dots, X_n uma amostra aleatória de $U\left(\xi - \frac{\tau}{2}, \xi + \frac{\tau}{2}\right)$ com ambos os parâmetros desconhecidos e considere uma função de perda da forma $L(\xi, \tau, d) = [(d - \xi)^2]/(\tau^2)$. Lehmann & Casella (1998) demonstram que $\delta^* = (X_{(1)} + X_{(n)})/2$ é o EERM de ξ quando τ é conhecido. Pelas mesmas justificativas apresentadas no exemplo 14, temos que δ^* é o EERM de ξ também neste caso.

Entretanto, alguns estimadores não satisfazem as condições do lema 1, como aqueles apresentados em Lehmann & Casella (1998, p. 153 e 155). Sendo assim, é necessário o desenvolvimento de alguns EERM que não dependam das referidas suposições.

Teorema 5. *Seja δ_0 qualquer estimador de ξ satisfazendo (43) e δ_1 qualquer estimador de τ tomando valores positivos e satisfazendo*

$$\delta_1(a + b\mathbf{x}) = b \delta_1(\mathbf{x}), \quad \forall b > 0, \quad \forall a \tag{46}$$

Então, δ satisfaz (43) se e somente se for da forma

$$\delta(\mathbf{x}) = \delta_0(\mathbf{x}) - w(\mathbf{z})\delta_1(\mathbf{x}) \tag{47}$$

em que \mathbf{z} é dado por (40).

Demonstração. Primeiramente, pelo lema 1.6 (Lehmann & Casella 1998, p. 150), temos que δ satisfaz (43) se e somente se for da forma,

$$\delta(\mathbf{x}) = \delta_0(\mathbf{x}) - u(\mathbf{x})\delta_1(\mathbf{x}) \tag{48}$$

em que

$$u(a + b\mathbf{x}) = u(\mathbf{x}), \quad \forall b > 0 \text{ e } \forall a \tag{49}$$

Suficiência. Considere que $\delta(\mathbf{x}) = \delta_0(\mathbf{x}) - u(\mathbf{x})\delta_1(\mathbf{x})$ e $u(a + b\mathbf{x}) = u(\mathbf{x})$. Dessa forma, temos que

$$\begin{aligned} \delta(a + b\mathbf{x}) &= \delta_0(a + b\mathbf{x}) - u(a + b\mathbf{x})\delta_1(a + b\mathbf{x}) \\ &= a + b \delta_0(\mathbf{x}) - b u(\mathbf{x})\delta_1(\mathbf{x}) \\ &= a + b [\delta(\mathbf{x}) - u(\mathbf{x})\delta_1(\mathbf{x})] = a + b \delta(\mathbf{x}) \end{aligned}$$

Necessidade. Considere que $\delta(a + b\mathbf{X}) = a + b\delta(\mathbf{X})$ e defina $u(\mathbf{x}) = (\delta(\mathbf{x}) - \delta_0(\mathbf{x}))/(\delta_1(\mathbf{x}))$. Portanto,

$$\begin{aligned} u(a + b\mathbf{x}) &= \frac{\delta(a + b\mathbf{x}) - \delta_0(a + b\mathbf{x})}{\delta_1(a + b\mathbf{x})} = \frac{a + b\delta(\mathbf{x}) - a - b\delta_0(\mathbf{x})}{b\delta_1(\mathbf{x})} \\ &= \frac{\delta(\mathbf{x}) - \delta_0(\mathbf{x})}{\delta_1(\mathbf{x})} = u(\mathbf{x}) \quad \square \end{aligned}$$

Logo, (48) e (49) são válidos. O fato de (48) ser válido se e somente se u depender de \mathbf{x} através de \mathbf{z} decorre do lema 1.7 (Lehmann & Casella 1998, p. 151) e do teorema 2.

Por outro lado, um argumento semelhante ao teorema 1.10 (Lehmann & Casella 1998, p. 151) mostra que o EERM de ξ é,

$$\delta(\mathbf{X}) = \delta_0(\mathbf{X}) - w^*(\mathbf{z})\delta_1(\mathbf{X}) \quad (50)$$

em que $w^*(\mathbf{z})$ é um número que minimiza

$$\mathbb{E}_{\xi=0, \tau=1} [\rho(\delta_0(\mathbf{X}) - w^*(\mathbf{z})\delta_1(\mathbf{X})) | \mathbf{z}] = \mathbb{E}_{0,1} [\rho(\delta_0(\mathbf{X}) - w^*(\mathbf{z})\delta_1(\mathbf{X})) | \mathbf{z}]$$

Em particular, se

$$\rho\left(\frac{d-\xi}{\tau}\right) = \frac{(d-\xi)^2}{\tau^2} \quad (51)$$

não é difícil ver que

$$w^*(\mathbf{z}) = \frac{\mathbb{E}_{0,1} [\delta_1(\mathbf{X})\delta_0(\mathbf{X})|\mathbf{Z}]}{\mathbb{E}_{0,1} [\delta_1^2(\mathbf{X})|\mathbf{Z}]} \quad (52)$$

Exemplo 15. Exponencial deslocada Sejam X_1, \dots, X_n uma amostra aleatória de uma distribuição $E(\xi, \tau)$ cuja densidade é dada por

$$f_{X_i}(x_i) = \tau^{-1} e^{-(x-\xi)} \mathbf{1}_{[\xi, \infty)}(x), \quad \xi \in \mathbb{R}, \quad \tau > 0$$

com ambos os parâmetros desconhecidos. Considere $\delta_0(\mathbf{X}) = X_{(1)}$ e $\delta_1(\mathbf{X}) = \sum_{i=1}^n [X_i - X_{(1)}]$. Como $\boldsymbol{\delta} = (\delta_0, \delta_1)^\top$ é uma estatística suficiente e completa para $\boldsymbol{\theta}$, ela é independente de \mathbf{Z} pelo teorema de Basu e, além do mais, é possível demonstrar que são independentes entre si (Lehmann & Casella 1998, p. 43). Por outro lado, $\delta_0(a + b\mathbf{X}) = a + bX_{(1)} = a + b\delta_0$ e $\delta_1(a + b\mathbf{X}) = b \sum_{i=1}^n [X_i - X_{(1)}] = b\delta_1(\mathbf{X})$. Sendo assim, o teorema 5 pode ser aplicado, o que, unido à função de perda (51) e com o fato de que $\delta_0(\mathbf{X})_{\tau=1} \sim E(0, 1/n)$ e $\delta_1(\mathbf{X})_{\tau=1} \sim \Gamma(n-1, 1)$ (Lehmann & Casella 1998), leva a

$$w^*(\mathbf{z}) = \frac{\mathbb{E}_{0,1} [\delta_0(\mathbf{X})\delta_1(\mathbf{X}) | \mathbf{Z}]}{\mathbb{E}_{0,1} [\delta_1^2(\mathbf{X}) | \mathbf{Z}]} = \frac{\mathbb{E}_{0,1} [\delta_0(\mathbf{X})] \mathbb{E}_{0,1} [\delta_1(\mathbf{X})]}{\mathbb{E}_{0,1} [\delta_1^2(\mathbf{X})]} = \frac{\frac{1}{n} \frac{\Gamma(n)}{\Gamma(n-1)}}{\frac{\Gamma(n+1)}{\Gamma(n-1)}} = \frac{1}{n^2}$$

Logo, o EERM de ξ é

$$\delta^*(\mathbf{X}) = X_{(1)} - \frac{1}{n^2} \sum_{i=1}^m [X_i - X_{(1)}]$$

4.3 Estimação simultânea

Nas subseções anteriores, mostramos como encontrar EERM para os parâmetros do modelo de localização-escala de forma **marginal**. Nesta subseção, apresentamos os resultados obtidos por Prabakaran & Chandrasekar (1994), que estimam

de forma conjunta os parâmetros $(\xi, \tau^r)^\top$ no modelo de localização-escala (37). Inicialmente, são dadas algumas definições e conceitos necessários para o desenvolvimento desta subseção.

Considere o grupo de transformações

$$\mathbf{X}^* = a + b\mathbf{X} \tag{53}$$

que induzem à transformação $\boldsymbol{\theta}^* = (\xi^*, \tau^{*r})^\top = (a + b\xi, b\tau)^\top$, sob a qual o modelo (37) permanece invariante. Se $\mathbf{d} = (d_1, d_2)^\top$ é uma estimativa de $(\xi, \tau^r)^\top$ obtida no modelo original, então podemos relacionar a mesma com a estimativa de $(\xi^*, \tau^{*r})^\top$ obtida no modelo transformado, da forma $\bar{\mathbf{g}}(\mathbf{d}) = \mathbf{d}^* = (a + b d_1, b^r d_2)^\top$, que é uma função apenas de \mathbf{d} . Pode-se mostrar (Prabakaran & Chandrasekar 1994) que uma função de perda é invariante sobre as transformações acima se e somente se

$$L(\boldsymbol{\theta}, \mathbf{d}) = \rho \left(\frac{d_1 - \xi}{\tau}, \frac{d_2}{\tau^r} \right) \tag{54}$$

e que o grupo \bar{G} é transitivo sobre $\Omega = \mathbb{R} \times \mathbb{R}^+$. Portanto, o risco de qualquer estimador equivariante, calculado sob funções de perda da forma (54), não depende de $\boldsymbol{\theta}$.

Definição 6. Um estimador $(\delta_1, \delta_2)^\top$ de $(\xi, \tau^r)^\top$ é dito ser equivariante por **localização-escala** se δ_1 e δ_2 são estimadores marginalmente equivariantes por localização-escala, respectivamente, para ξ e τ^r , ou seja, se δ_1 satisfaz (43) e δ_2 satisfaz (39).

Definição 7. Uma função vetorial $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))^\top$ é dita ser invariante para o problema de localização-escala se

$$\mathbf{u}(a + b\mathbf{x}) = (bu_1(\mathbf{x}), u_2(\mathbf{x}))^\top, \quad \forall a \in \mathbb{R} \quad \text{e} \quad \forall b > 0 \tag{55}$$

A seguir, mostramos alguns resultados importantes acerca da caracterização dos estimadores equivariantes por localização-escala de $(\xi, \tau^r)^\top$.

Lema 2. Um estimador $(\delta_1(\mathbf{x}), \delta_2(\mathbf{x}))^\top$ é equivariante por localização-escala para (ξ, τ^r) se e somente se, para todo estimador equivariante $(\delta_{01}(\mathbf{x}), \delta_{02}(\mathbf{x}))^\top$, existir uma função invariante por localização-escala \mathbf{u} tal que

$$\delta_1(\mathbf{x}) = \delta_{01}(\mathbf{x}) - u_1(\mathbf{x}) \tag{56}$$

$$\delta_2(\mathbf{x}) = \delta_{02}(\mathbf{x})/u_2(\mathbf{x}) \tag{57}$$

O seguinte lema fornece uma caracterização das funções invariantes por localização-escala.

Lema 3. Uma função $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))^\top$ é invariante para o problema de localização-escala se e somente se

$$u_1(\mathbf{x}) = g(\mathbf{x})w_1(z_1, \dots, z_{n-1}) \tag{58}$$

$$u_2(\mathbf{x}) = w_2(z_1, \dots, z_{n-1}) \tag{59}$$

para alguma função positiva g tal que $g(a + b\mathbf{x}) = bg(\mathbf{x})$ e $z_i = (x_i - x_n)/g(\mathbf{x})$ ($i = 1, \dots, n - 1$).

Perceba que se fizermos $g(\mathbf{x}) = |x_{n-1} - x_n|$, obtemos a mesma caracterização obtida nas Seções 4.1 e 4.2 de forma marginal. Pode-se observar também que g , como definida acima, é um estimador equivariante por escala de τ .

Teorema 6. *Seja $(\delta_{01}(\mathbf{x}), \delta_{02}(\mathbf{x}))^\top$ um estimador equivariante de $(\xi, \tau^r)^\top$. Então uma condição necessária e suficiente para que $(\delta_1(\mathbf{x}), \delta_2(\mathbf{x}))^\top$ seja um estimador equivariante por localização-escala é que ele seja da forma*

$$\delta_1(\mathbf{x}) = \delta_{01}(\mathbf{x}) - g(\mathbf{x})w_1(z_1, \dots, z_{n-1}) \quad (60)$$

$$\delta_2(\mathbf{x}) = \delta_{02}(\mathbf{x})/w_1(z_1, \dots, z_{n-1}) \quad (61)$$

para algumas funções w_1 e w_2 .

Demonstração. É uma consequência imediata dos lemas 2 e 3. □

Considere D a classe de todos os estimadores equivariantes para $(\xi, \tau^r)^\top$ que tenham risco finito. Em particular, quando a função de perda é da forma

$$L(\boldsymbol{\theta}, \mathbf{d}) = a_{11} \left(\frac{d_1 - \xi}{\tau} \right)^2 + 2a_{12} \left(\frac{d_1 - \xi}{\tau} \right) \left(\frac{d_2}{\tau^r} - 1 \right) + a_{22} \left(\frac{d_2}{\tau^r} - 1 \right)^2 \quad (62)$$

então um estimador $\boldsymbol{\delta}^* \in D$ que minimiza risco sobre esta função de perda é denominado Q_A -EERM (Prabakaran & Chandrasekar 1994), com $\mathbf{A} = (a_{ij})_{1 \leq i, j \leq 2}$ representando uma matriz 2×2 simétrica positiva definida. A função de perda (62) é dita ser **quadrática**, conforme definido em Zacks (1971, p. 102).

Teorema 7. *Considere \mathbf{X} um vetor aleatório com densidade (37). Se*

i) $L(\xi, \tau, d_1, d_2) = \rho \left(\frac{d_1 - \xi}{\tau}, \frac{d_2}{\tau^r} \right);$

ii) *existir um estimador equivariante $\boldsymbol{\delta}_0 = (\delta_{01}, \delta_{02})^\top$ com risco finito;*

iii) *para cada $\mathbf{z} = (z_1, \dots, z_{n-1})^\top$, existir uma função vetorial $\mathbf{w}^*(\mathbf{z})$ que minimize $\mathbb{E}[\rho(\delta_{01}(\mathbf{X}) - g(\mathbf{X})w_1(\mathbf{z}), \delta_{02}(\mathbf{X})/w_2(\mathbf{z}))/\mathbf{z}]$, com o operador esperança sendo calculado quando $\boldsymbol{\theta} = (0, 1)^\top$.*

Então um EERM $\boldsymbol{\delta}^ = (\delta_1^*, \delta_2^*)^\top$ existe e é dado por*

$$\delta_1^*(\mathbf{X}) = \delta_{01}(\mathbf{X}) - g(\mathbf{X})w_1^*(\mathbf{z}) \quad (63)$$

$$\delta_2^*(\mathbf{X}) = \delta_{02}(\mathbf{X})/w_2^*(\mathbf{z}) \quad (64)$$

Demonstração. Análoga à demonstração do teorema 3. □

A abordagem acima é bem geral e fornece uma estimação simultânea de ξ e τ^r . Escolhendo a função de perda de forma apropriada, pode-se obter os mesmos EERM para ξ e τ^r obtidos (de forma marginal) nas subseções anteriores. Para isto, basta escolher de forma conveniente a matriz \mathbf{A} .

Considerando que a função de perda seja dada por (62), Prabakaran & Chandrasekar (1994, eq. 3.6 e 3.7) obtêm expressões explícitas dos EERM de $(\xi, \tau^r)^\top$

que dependem da particular escolha da matriz \mathbf{A} . Os autores mostram também que quando o interesse é estimar $\boldsymbol{\theta} = (\xi, \tau)^\top$, o EERM por localização-escala é independente da escolha da matriz \mathbf{A} e é dado por $\boldsymbol{\delta}^*(\mathbf{X}) = (\delta_1^*(\mathbf{X}), \delta_2^*(\mathbf{X}))^\top$ com

$$\delta_1^*(\mathbf{X}) = \delta_{01}(\mathbf{X}) - \delta_{02}(\mathbf{X}) \frac{\mathbb{E}[\delta_{01}(\mathbf{X})\delta_{02}(\mathbf{X}) | \mathbf{Z}]}{\mathbb{E}[\delta_{02}^2(\mathbf{X}) | \mathbf{Z}]} \tag{65}$$

$$\delta_2^*(\mathbf{X}) = \frac{\delta_{02}(\mathbf{X})\mathbb{E}[\delta_{02}(\mathbf{X}) | \mathbf{Z}]}{\mathbb{E}[\delta_{02}^2(\mathbf{X}) | \mathbf{Z}]} \tag{66}$$

que coincidem com os EERM marginais de ξ e τ , dados por (50) e (41), sob as funções de perda quadráticas (51) e (28), respectivamente.

Além disso, Prabakaran & Chandrasekar (1994) concluem que sob funções de perda quadráticas, os EERM de $(\xi, \tau^r)^\top, r > 1$, não coincidem com os EERM de ξ e τ^r obtidos de forma marginal, e que esta diferença pode ser atribuída ao fato de que $\boldsymbol{\theta} = (\xi, \tau)^\top$ é o parâmetro natural enquanto que $(\xi, \tau^r)^\top$ é uma função paramétrica de $\boldsymbol{\theta}$. Entre outras propriedades, Prabakaran & Chandrasekar (1994) caracterizam o Q_A -EERM (caracterização semelhante à do ENVVUM) de $\boldsymbol{\theta}$ e mostram que se ele existe, então é único q.c. (tais propriedades podem ser aplicadas nos resultados obtidos nas subseções anteriores quando se tem interesse em estimar marginalmente ξ ou τ sob funções de perda quadráticas das formas (51) e (28)).

Exemplo 16. Distribuição $E(\xi, \tau)$. Suponha que $X_1, \dots, X_n \stackrel{\text{a.a.}}{\sim} E(\xi, \tau)$ e que se tem interesse em estimar obter o EERM de $\boldsymbol{\theta} = (\xi, \tau)^\top$, considerando a função de perda (62). Como foi discutido anteriormente, tem-se que o EERM por localização-escala de $\boldsymbol{\theta}$ é independente da particular escolha da matriz \mathbf{A} e é dado por $\boldsymbol{\delta}^*(\mathbf{X}) = (\delta_1^*(\mathbf{X}), \delta_2^*(\mathbf{X}))^\top$, com $\delta_1^*(\mathbf{X}) = X_{(1)}$ sendo o EERM marginal por localização de ξ sob a função de perda (51) e $\delta_2^*(\mathbf{X}) = \sum_{i=1}^n (X_{(i)} - X_{(1)})$ sendo o EERM marginal por escala de τ sob a função de perda (28), como foi mostrado anteriormente. Dado que o EERM por localização-escala existe, conclui-se também que ele é único quase-certamente $[\mathcal{P}]$.

5 Aplicação em modelos lineares e ENVVUM

Os modelos de regressão constituem uma das mais importantes ferramentas de análise estatística. Nesta seção, apresentaremos alguns resultados de estimação equivariante e não-viciada de variância uniformemente mínima, aplicados à classe de modelos de regressão normais lineares. Existe uma vasta literatura sobre esses modelos, entre as quais destacamos Scheffé (1959), Seber (1977), Searle (1987), entre outros.

Antes de abordar propriamente os processos de estimação, vamos definir o chamado Modelo Linear Geral (normal) (Searle 1987), qual seja,

$$X_i \sim N(\xi_i, \sigma^2), \quad i = 1, \dots, n \tag{67}$$

em que os X_i são independentes e $\xi_1, \dots, \xi_n \in \prod_{\Omega}$ que é um sub-espço linear de dimensão s de $E_n (s < n)$.

Para evitar problemas de não-identificabilidade (Searle 1987) e para facilitar o processo de obtenção de estatísticas suficientes e completas para a estimação NVVUM (Lehmann & Casella 1998), é conveniente reduzir este modelo à forma canônica através da transformação ortogonal

$$\mathbf{Y} = \mathbf{XC} \quad (68)$$

que leva a

$$\eta = \mathbb{E}(\mathbf{Y}) = \boldsymbol{\xi}\mathbf{C}$$

em que $\eta = (\eta_1, \dots, \eta_n)^\top$ e $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^\top$. Note que a transformação (68) é 1 a 1 e, além disso, o Jacobiano é igual a 1. Segue daí, devido às propriedades da distribuição normal multivariada (Mardia et al. 1979), que $\mathbf{Y} \sim N(\eta, \sigma^2\mathbf{I}_n)$, pois

$$\text{Cov}(\mathbf{Y}) = \sigma^2\mathbf{C}\mathbf{I}\mathbf{C}^\top = \sigma^2\mathbf{C}\mathbf{C}^\top = \sigma^2\mathbf{I}_n$$

notando que \mathbf{C} é ortogonal, com \mathbf{I}_n representando a matriz identidade de ordem n . Denotando por \mathbf{c}_i^\top a i -ésima coluna de \mathbf{C} , é desejável escolher \mathbf{c}_i de tal modo que $\mathbf{c}_1, \dots, \mathbf{c}_s$ gerem \prod_{Ω} [para garantir a identificabilidade]. Então,

$$\boldsymbol{\xi} \in \prod_{\Omega} \iff \boldsymbol{\xi} \text{ for ortogonal às } n - s \text{ colunas restantes de } \mathbf{C}$$

Como $\eta = \boldsymbol{\xi}\mathbf{C} = [\boldsymbol{\xi}_1\mathbf{C}_1 \quad \boldsymbol{\xi}_2\mathbf{C}_2] = [\boldsymbol{\xi}_1\mathbf{C}_1 \quad \mathbf{0}]$, segue que,

$$\boldsymbol{\xi} \in \prod_{\Omega} \iff \eta_{s+1} = \dots = \eta_n = 0$$

Em termos dos Y , temos que,

$$Y_i = \begin{cases} N(\eta_i, \sigma^2), & i = 1, \dots, s; \\ N(0, \sigma^2), & i = s + 1, \dots, n. \end{cases}$$

Note que, $\boldsymbol{\xi}$ varia em \prod_{Ω} e η_1, \dots, η_s varia irrestritamente sobre E_s com $\eta_{s+1} = \dots = \eta_n = 0$.

Nesta representação $\mathbf{T} = (Y_1, \dots, Y_s, S^2)^\top$, $S^2 = \sum_{j=s+1}^n Y_j^2$, é uma estatística suficiente e completa para $(\eta_1, \dots, \eta_s, \sigma^2)^\top$ (Lehmann & Casella 1998). O teorema a seguir apresenta um modo de obter os EERM e ENVVUM dos parâmetros de interesse.

Teorema 8.

i) Os ENVVUM de $\sum_{i=1}^s \lambda_i \eta_i$ (λ são constantes conhecidas) e σ^2 são $\sum_{i=1}^s \lambda_i Y_i$ e $S^2/(n-2)$, respectivamente.

ii) Sob as transformações

$$\begin{aligned} Y_i^* &= Y_i + a_i \quad (i = 1, \dots, s) & Y_i^* &= Y_i \quad (i = s + 1, \dots, n) \\ \eta_i^* &= \eta_i + a_i \quad (i = 1, \dots, s) & d^* &= d + \sum_{i=1}^n a_i \lambda_i \end{aligned}$$

com função de perda $L(\eta, d) = \rho(d - \sum_{i=1}^s \lambda_i \eta_i)$, em que ρ é convexa e par, o ENVVUM $\sum_{i=1}^s \lambda_i Y_i$ também é o EERM de $\sum_{i=1}^s \lambda_i \eta_i$.

iii) Sob a função de perda $(d - \sigma^2)^2 / \sigma^4$, o EERM de σ^2 é $S^2 / (n - s + 2)$.

Demonstração.

1. Basta observar que os estimadores propostos são não-viesados e, além disso, funções de estatísticas suficientes e completas, no caso $(Y_1, \dots, Y_s, S^2)^\top$.
2. Note que o grupo de transformações é transitivo e a função de perda é invariante por localização e portanto, o risco é constante. Por outro lado, note, denotando $\delta(\mathbf{Y}) = \sum_{i=1}^s \lambda_i Y_i$, que

$$\begin{aligned} \delta(Y_1 + a_1, Y_2 + a_2, \dots, Y_s + a_s, Y_{s+1}, \dots, Y_n) &= \\ \sum_{i=1}^s \lambda_i (Y_i + a_i) &= \sum_{i=1}^s \lambda_i Y_i + \sum_{i=1}^s \lambda_i a_i = \delta(\mathbf{Y}) + \sum_{i=1}^s \lambda_i a_i \end{aligned}$$

Ou seja, o ENNVUM também é equivariante por localização. Pelo teorema de Rao-Blackwell (Lehmann & Casella 1998), temos que, para qualquer função de perda convexa, o risco de $\delta(\mathbf{Y})$ é menor ou igual ao de qualquer outro estimador. Logo, $\delta(\mathbf{Y})$ é o EEMR de $\sum_{i=1}^s \lambda_i \eta_i$.

3. Segue essencialmente do exemplo 11. □

É conveniente expressar os estimadores desenvolvidos em termos das variáveis originais \mathbf{X} ao invés de \mathbf{Y} (lembre-se de que a transformação é 1 a 1). Para tal, vamos introduzir o seguinte conceito.

Tome $\boldsymbol{\xi} \in \prod_{\Omega}$, então os estimadores de mínimos quadrados destes são $(\hat{\xi}_1, \dots, \hat{\xi}_n)$ que minimizam $\sum_{i=1}^n (X_i - \xi_i)^2$, sujeito à condição $\boldsymbol{\xi} \in \prod_{\Omega}$.

Teorema 9. Sob o modelo (67), o ENVVUM de $\sum_{i=1}^n \gamma_i \xi_i$ é $\sum_{i=1}^n \gamma_i \hat{\xi}_i$, $\forall \gamma_i \in \mathbb{R}$ conhecido.

Demonstração. Pelo teorema 8 e a completividade de \mathbf{T} é suficiente mostrar que $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ é uma função linear de Y_1, \dots, Y_s e que é não-viesado para $\sum_{i=1}^n \gamma_i \xi_i$.

Note que

$$\begin{aligned} \sum_{i=1}^n (X_i - \xi_i)^2 &= \sum_{i=1}^n (Y_i - \mathbb{E}(Y_i))^2 \\ &= \sum_{i=1}^s (Y_i - \eta_i)^2 + \sum_{i=s+1}^n Y_i^2 \end{aligned} \tag{69}$$

O lado direito de (69) é minimizado por $\hat{\eta}_i = Y_i$, $i = 1, \dots, s$, enquanto que o lado esquerdo, por $\hat{\xi}_1, \dots, \hat{\xi}_n$. Assim, temos que $(\eta = \boldsymbol{\xi} \mathbf{C})$

$$\hat{\eta} = \hat{\boldsymbol{\xi}} \mathbf{C} \Rightarrow (Y_1 \dots Y_s 0 \dots 0) = (\hat{\xi}_1 \dots \hat{\xi}_n) \mathbf{C} \Rightarrow \hat{\boldsymbol{\xi}} = (Y_1 \dots Y_s 0 \dots 0) \mathbf{C}^{-1} \tag{70}$$

Como $\mathbb{E}(\hat{\boldsymbol{\xi}}) = \boldsymbol{\eta} \mathbf{C}^{-1} = \boldsymbol{\xi} \mathbf{C} \mathbf{C}^{-1} = \boldsymbol{\xi}$, ou seja, é um estimador não-viesado, e de (70), vemos que são funções lineares do vetor \mathbf{Y} . □

Agora, vamos reinterpretar as considerações sobre equivariância em termos das variáveis originais. Antes, precisamos definir o grupo de transformações que deixam o problema invariante. As transformações conduzidas nos Y (teorema 8) em termos das variáveis X_1, \dots, X_n , tornam-se

$$\begin{aligned} \mathbf{X}^* &= \mathbf{X} + \mathbf{b} & \mathbf{b} &= (b_1, \dots, b_n)^\top \in \prod_{\Omega} \\ \boldsymbol{\xi}^* &= \boldsymbol{\xi} + \mathbf{b} \in \prod_{\Omega} & d^* &= d + \sum_{i=1}^n b_i \gamma_i \end{aligned} \quad (71)$$

Podemos então estender o teorema 8 para o seguinte corolário.

Corolário 5. *Sob as transformações (71), $\sum_{i=1}^n \gamma_i \widehat{\xi}_i$ é o EERM de $\sum_{i=1}^n \gamma_i \xi_i$ com função de perda $\rho(d - \sum_{i=1}^n \gamma_i \xi_i)$ convexa e par.*

Demonstração. Notando que de (70) $\widehat{\xi}_j = \sum_{i=1}^s c_{ij} Y_i$, $j = 1, \dots, n$, então $\sum_{j=1}^n \gamma_j \widehat{\xi}_j = \sum_{i=1}^s c_i^* Y_i$, com $c_i^* = \sum_{j=1}^n \gamma_j c_{ij}$ e o resultado segue do teorema 8(i). \square

Para obter o ENNVUM e o EERM de σ^2 em termos do vetor \mathbf{X} , é necessário apenas expressar S^2 em função desse vetor. Note que, da minimização de (69), temos

$$\sum_{i=1}^n (X_i - \widehat{\xi}_i)^2 = \sum_{i=s+1}^n Y_i^2 = S^2 \quad (72)$$

Logo, o ENNVUM e o EERM de σ^2 são, respectivamente teorema 8 (iii),

$$\frac{\sum_{i=1}^n (X_i - \widehat{\xi}_i)^2}{n-s} \quad \text{e} \quad \frac{\sum_{i=1}^n (X_i - \widehat{\xi}_i)^2}{n-s+2}$$

Vamos agora ilustrar os resultados apresentados.

Exemplo 17. Anova com 1 fator.

Suponha que $X_{ij} \sim N(\xi_i, \sigma^2)$, $i = 1, \dots, s$; $j = 1, \dots, n_i$ e que sejam independentes. Do corolário 5 temos que, para encontrar os ENNVUM ou EERM de combinações lineares de ξ , basta encontrar os estimadores de mínimos quadrados. Estes, por sua vez, são os valores $\widehat{\xi}$ que minimizam,

$$\sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - \xi_i)^2 = \sum_{i=1}^s \left\{ \sum_{j=1}^{n_i} (X_{ij} - X_i)^2 + n_i (X_i - \xi_i)^2 \right\}$$

que resulta em

$$\widehat{\xi}_i = X_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

Além disso, de (72), temos que o ENVVUM de σ^2 é

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^s n_i - s} \sum_{i=1}^s \sum_{j=1}^{n_i} (X_{ij} - X_{i.})^2$$

Exemplo 18. Anova com 2 fatores.

Considere $X_{ijk} \sim N(\xi_{ij}, \sigma^2)$, $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, m$. Uma reparametrização usual para este modelo é

$$\xi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

com as seguintes restrições de identificabilidade

$$\sum_{i=1}^I \alpha_i = \sum_{j=1}^J \beta_j = \sum_{i=1}^I \gamma_{ij} = \sum_{j=1}^J \gamma_{ij} = 0 \quad (73)$$

Usando as restrições (73) temos que (o ponto representa a média calculada no índice de interesse)

$$\xi_{.} = \mu, \quad \xi_{i.} = \mu + \alpha_i, \quad \xi_{.j} = \mu + \beta_j$$

Então,

$$\mu = \xi_{.}, \alpha_i = \xi_{i.} - \xi_{.}, \beta_j = \xi_{.j} - \xi_{.}, \gamma_{ij} = \xi_{ij} - \mu - \alpha_i - \beta_j$$

ou ainda, $\gamma_{ij} = (\xi_{ij} - \xi_{.}) - [(\xi_{i.} - \xi_{.}) + (\xi_{.j} - \xi_{.})]$. Note que α_i é o efeito médio do nível i do primeiro fator, β_j é o efeito médio do nível j do segundo fator e γ_{ij} é a diferença entre o efeito conjunto dos dois fatores e a soma dos efeitos dos fatores separados de cada um (chamado de interação).

Os ENVVUM desses parâmetros (efeitos) seguem-se imediatamente do teorema 8 e do exemplo 17. Essencialmente, os ENVVUM são obtidos calculando-se os estimadores de mínimos quadrados do vetor $\boldsymbol{\xi}$, que neste caso são (denotando pelo ponto a média calculada num determinado índice),

$$\hat{\mu} = X_{...}, \quad \hat{\alpha}_i = X_{i..} - X_{...}, \quad \hat{\beta}_j = X_{.j.} - X_{...}, \quad \hat{\gamma}_{ij} = X_{ij.} - X_{i..} - X_{.j.} + X_{...}$$

Analogamente, o ENVVUM de σ^2 é

$$\frac{1}{IJ(m-1)} \sum_{k=1}^m \sum_{j=1}^J \sum_{i=1}^I (X_{ijk} - X_{ij.})^2$$

Note ainda que, do corolário 5, os EMQ (estimadores de mínimos quadrados) de $\hat{\boldsymbol{\xi}}$ são também os EERM, sob uma perda convexa, par e invariante por localização.

Estes resultados podem ser generalizados para experimentos fatoriais, ou seja, experimentos que envolvem um número geral de fatores.

Podemos ainda considerar situações em que restringimos o interesse em estimadores não-viciados e funções de perda quadrática mas, por outro lado, desconsiderando a normalidade e a independência.

Suponha que consideramos de (67) somente suposições a respeito dos dois primeiros momentos,

$$\mathbb{E}(X_i) = \xi_i, \quad \xi \in \prod_{\Omega}, \quad \text{Var}(X_i) = \sigma^2, \quad \text{Cov}(X_i, X_j) = 0, \quad i \neq j \quad (74)$$

sem considerar as suposições de independência ou normalidade.

Teorema 10 (Gauss). Para os estimadores de Mínimos Quadrados

Sob as suposições (74), $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ do teorema 9 é ENVVUM, entre todos os estimadores lineares, de $\sum_{i=1}^n \gamma_i \xi_i$.

Demonstração. Este estimador também é não-viesado, nas referidas condições. Seja $\sum_{i=1}^n c_i X_i$ qualquer outro estimador linear não-viesado de $\sum_{i=1}^n \gamma_i \xi_i$. Como $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ é o ENVVUM no caso normal e a variância de funções lineares dos X_i dependem somente do primeiro e segundo momentos, segue que $\text{Var} \left\{ \sum_{i=1}^n \gamma_i \hat{\xi}_i \right\} \leq \text{Var} \left\{ \sum_{i=1}^n c_i \hat{\xi}_i \right\}$. Então, $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ é o ENNVUM entre todos os estimadores lineares não-viesados. \square

Corolário 6. *Sob as suposições (74) e com perda quadrática, $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ é o EERM com respeito as transformações (71) entre todos os estimadores equivariantes lineares de $\sum_{i=1}^n \gamma_i \xi_i$.*

Demonstração. Este resultado segue do lema 1.23 (Lehmann & Casella 1998, p. 157), dado que $\sum_{i=1}^n \gamma_i \hat{\xi}_i$ é o ENVVUM (entre os estimadores lineares) e além disso é equivariante. \square

Para finalizar, gostaríamos de salientar que os resultados apresentados nesta seção podem ser estendidos para Modelos Lineares Mistos, como no teorema 4.14, Lehmann & Casella (1998, p. 185); veja também Harville (1976).

6 Conclusões e comentários adicionais

Verificamos que, em sua essência, os estimadores equivariantes podem ser construídos a partir dum estimador equivariante qualquer e dum estatística ancilar. Além disso, se este estimador equivariante escolhido for função dum estatística suficiente e completa, ele será independente da estatística ancilar em questão e isso facilita a obtenção do EERM.

Desde que se restrinja aos estimadores lineares, a estimação NVVUM, no contexto de modelos lineares, não fica comprometida sem a suposição de normalidade e, além disso, estes estimadores podem ser obtidos em várias situações, inclusive para os efeitos aleatórios em modelos mistos (Harville 1976).

Além das situações apresentadas neste trabalho, família de localização-escala e alguns modelos lineares, podemos citar Zacks (1971), Schervish (1995) e Lehmann & Casella (1998) que discutem estimação equivariante no contexto bayesiano; Borovkov (1998) e Lehmann & Romano (2005) definem os **testes de hipóteses invariantes** e apresenta algumas propriedades destes testes; (Khuri et al. 1998) fazem uso da teoria de testes invariantes para definir testes invariantes uniformemente mais poderosos em modelos mistos, tanto para as componentes de variância, como para os efeitos fixos (dado a inexistência de testes UMP na maioria das situações nessa classe de modelos); e Alexander & Chandrasekar (1999) que, dentro do contexto de Análise de Sobrevivência (amostra com censura), discutem o problema de estimação equivariante dos parâmetros do modelo exponencial.

Agradecimentos

Este trabalho foi apresentado na disciplina MAE 5834 - Estatística Avançada I (2004) no IME-USP. Os autores gostariam de agradecer à Profa. Dra. Silvia Ferrari (IME/USP) que revisou paciente e cuidadosamente todo o manuscrito e nos concedeu imprescindíveis sugestões e ao colega de doutorado Raydonal Ospina por sugerir a submissão do referido trabalho e aos dois árbitros pelas valiosas sugestões para a melhoria do nosso trabalho. Gostaríamos também de agradecer ao CNPq pelo suporte financeiro ao curso de Doutorado.

Recibido: agosto de 2006

Aceptado: septiembre de 2006

Referências

- Alexander, T. L. & Chandrasekar, B. (1999), 'Equivariant Estimation for the Parameters of an Exponential Model Based on Censored Sampling', *Biometrical Journal* **41**, 471–481.
- Borovkov, A. A. (1998), *Mathematics Statistics*, Gordon and Breach Science Publishers, Moscow.
- Casella, G. & Berger, R. L. (2002), *Statistical Inference*, 2nd edn, Duxbury Advanced Series, New York.
- Harville, D. A. (1976), 'Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects', *The Annals of Statistics* **4**, 384–395.
- Khuri, A. I., Mathew, T. & Sinha, B. K. (1998), *Statistical Tests for Mixed Linear Models*, John Wiley & Sons, New York.
- Lehmann, E. L. & Casella, G. (1998), *Theory of Point Estimation*, 2nd edn, Springer-Verlag, New York.

- Lehmann, E. L. & Romano, J. P. (2005), *Testing Statistical Hypothesis*, 3rd edn, Springer-Verlag, New York.
- Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press, London.
- Prabakaran, T. & Chandrasekar, B. (1994), 'Simultaneous Equivariant Estimation for Location-Scales Models', *Journal of Statistical Planning and Inference* **40**, 51–59.
- Scheffé, H. (1959), *The Analysis of Variance*, Wiley, New York.
- Schervish, M. J. (1995), *Theory of Statistics*, Springer-Verlag, New York.
- Searle, S. R. (1987), *Linear Models for Unbalanced Data*, Wiley, New York.
- Seber, G. A. F. (1977), *Linear Regression Analysis*, Wiley, New York.
- Staudte, R. G. (1971), 'A Characterization of Invariant Loss Functions', *The Annals of Mathematical Statistics* **42**, 1322–1327.
- Zacks, S. (1971), *The Theory of Statistical Inference*, John Wiley, New York.