



CARLOS JOSÉ DOS REIS

**TEORIA DOS GRUPOS NA ESTIMAÇÃO
EQUIVARIANTE**

LAVRAS - MG

2016

CARLOS JOSÉ DOS REIS

TEORIA DOS GRUPOS NA ESTIMAÇÃO EQUIVARIANTE

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador

Dr. Lucas Monteiro Chaves

Coorientador

Dr. Devanil Jaques de Souza

LAVRAS - MG
2016

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha
Catalográfica da Biblioteca Universitária da UFLA, com dados informados
pelo(a) próprio(a) autor(a).**

Reis, Carlos José dos.

Teoria dos grupos na estimação equivariante / Carlos José dos
Reis. - Lavras : UFLA, 2016.
200 p. : il.

Dissertação (mestrado acadêmico)-Universidade Federal de
Lavras, 2016.

Orientador: Lucas Monteiro Chaves.

Bibliografia.

1. Teoria da decisão. 2. Teoria da estimação. 3. Função risco. 4.
Ação de grupos. 5. Estimador de Pitman. I. Universidade Federal de
Lavras. II. Título.

CARLOS JOSÉ DOS REIS

TEORIA DOS GRUPOS NA ESTIMAÇÃO EQUIVARIANTE

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 22 de janeiro de 2016.

Dr. Devanil Jaques de Souza

UFLA

Dr. Denismar Alves Nogueira

UNIFAL-MG

Dr. João Domingos Scalon

UFLA

Dr. Lucas Monteiro Chaves
Orientador

**LAVRAS - MG
2016**

A todos que se interessarem pela teoria da estimação equivariante.

DEDICO

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade concedida para realização do mestrado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela concessão da bolsa de estudo, que me deu as condições necessárias para que eu pudesse dedicar-me ao curso.

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária pelos ensinamentos transmitidos para minha formação acadêmica e pessoal. Agradeço o respeito e a harmoniosa convivência nesses dois anos de mestrado.

A todos os funcionários do DEX. Sou muito grato pelo carinho, amizade e pela presteza com que sempre me atenderam. Muito obrigado!

Aos professores Denismar Alves Nogueira, Maria do Carmo Pacheco de Toledo Costa e João Domingos Scalon, que contribuíram de forma valiosa com suas sugestões e comentários para que se chegasse ao documento final desse trabalho.

Aos meus orientadores, Lucas Monteiro Chaves e Devanil Jaques de Souza, deixo aqui registrado o meu eterno carinho. Sinto-me ao final desse trabalho muito feliz e honrado em ter sido orientado por vocês. A disposição de vocês em ensinar e o respeito que dedicam aos seus alunos e orientados são exemplos que desejo praticar. Reservo ao final meu sincero agradecimento pelos ensinamentos e pelo respeito.

Ao professor Luiz Alberto Beijo, expresso também meus agradecimentos, porque mesmo de forma indireta, contribui para a realização deste trabalho, com todos os seus ensinamentos ao longo de minha graduação.

Aos meus alunos do estágio docência, que me propiciavam a cada semana

uma profunda alegria. Digo, com humildade, que o meu aprendizado nas aulas foi imenso por estar com vocês. Em especial, gostaria de agradecer a aluna Joyce Christina da Silva. Sua iniciativa e seu gesto serão sempre lembrados.

Aos meus colegas pelos bons momentos que passamos juntos. Em especial, agradecimentos são devidos aos meus colegas de mestrado, principalmente pela união de nossa turma. Talvez a palavra que melhor descreva o que sentirei a partir desse momento, em relação a vocês, seja saudade.

Gostaria de expressar meu agradecimentos também a Cristiane Alvarenga Gajo, Leandro da Silva Pereira e Danielle Gonçalves de Oliveira Prado pela boa convivência e apoio desde que ingressei no mestrado.

Aos amigos Gilberto Rodrigues Liska, Guido Gustavo Humada Gonzalez, Juliano Bortolini e André Geraldo Nogueira Barbosa o meu muito obrigado. Vocês fizeram com que esses dois anos fossem mais tranquilos em Lavras-MG. A vocês deixo aqui o meu eterno abraço.

A Dais Capucho Afini, por estar ao meu lado nos últimos quatro anos. Você, com seu sorriso, fez com que momentos difíceis parecessem muito mais fáceis.

Aos meus pais, José de Souza Reis e Selma Batista dos Reis que, não somente formam a base desses agradecimentos, mas também formam a base de minha vida. Amo muito vocês!

A Deus, pela oportunidade concedida.

“Quando deixei a Universidade de Melbourne, depois de quatro anos, não tinha treinamento em pesquisa. Acreditava, porém, que aprendera a estudar e a usar a matemática, e que poderia enfrentar qualquer problema que aparecesse.”

Edwin James George Pitman

RESUMO

Um interesse que surge em pesquisas científicas é obter informações acerca da população sob análise, o que pode ser feito utilizando-se uma função densidade de probabilidade (fdp) que a descreva. Por sua vez, uma fdp é caracterizada por quantidades numéricas desconhecidas, denominadas de parâmetros. Assim, o conhecimento dos parâmetros populacionais gera informações sobre a população em que reside um determinado interesse. Existem diferentes métodos para encontrar estimadores e com frequência esses métodos fornecem diferentes estimadores para um mesmo parâmetro populacional. Dessa forma, surge a necessidade de métodos para avaliar esses estimadores, e como a quantidade de estimadores possíveis pode ser muito grande, um procedimento a ser adotado é limitar a classe dos estimadores com base em algumas propriedades dos mesmos. Nesse sentido, o princípio da equivariância pode ser utilizado para essa finalidade. O princípio da equivariância é um princípio que preserva algumas características importantes do modelo, sendo muito utilizado na estatística clássica e intimamente relacionado à importante teoria matemática dos grupos. Nesse contexto, os objetivos gerais no presente trabalho consistiram em: i) apresentar as definições e os resultados teóricos necessários para o desenvolvimento da teoria da estimação equivariante, ii) exemplificar a utilização dos estimadores equivariantes com uma aplicação a dados de precipitação máxima de Piracicaba-SP. No desenvolvimento teórico, foram determinados os estimadores de Pitman para parâmetros de locação e de escala, que são os estimadores com menor risco uniforme na classe dos estimadores equivariantes. A estimação simultânea de parâmetros de locação e de escala em famílias de locação-escala também foi abordada. Demonstrou-se que os estimadores de Pitman para parâmetros de locação e de escala podem ser obtidos como estimadores de Bayes generalizados utilizando-se prioris adequadas. As propriedades de admissibilidade e minimaximalidade dos estimadores de Pitman foram discutidas. Na última seção apresentou-se uma aplicação dos estimadores equivariantes no estudo de precipitações pluviométricas máximas em Piracicaba-SP.

Palavras-chave: Teoria da decisão. Teoria da estimação. Função risco. Ação de grupos. Estimador de Pitman.

ABSTRACT

One concern that arises in scientific research is to gain information about the population under study, which can be made using a probability density function (pdf) that describes it. In turn, a pdf is characterized by unknown numerical quantities, called parameters. Thus, knowledge of population parameters generates information about the population that resides in a particular interest. There are different methods for finding estimators and often these methods provide different estimators for the same population parameter. Thus arises the need for methods to evaluate these estimators and, as the amount of possible estimators can be very large, a procedure to be adopted is to limit the class of estimators based on some properties. In that sense, the principle of equivariance can be used for this purpose. The principle of equivariance is a principle that preserves some important features of the model and is widely used in classical statistics and closely related to important Group Theory in mathematics. In this context, the general objectives of this work consists of: i) present the definitions and theoretical results necessary for the development of the Equivariant Estimation Theory, ii) exemplify the use of equivariant estimators with an application of maximum rainfall data from Piracicaba-SP. In theoretical development, we determined the Pitman estimators for location and scale parameters, which are the estimators with even smaller risk in the class of equivariant estimators. The simultaneous estimation of scale and location parameters on location-scale families was also discussed. It has been shown that the Pitman estimators for location and scale parameters can be obtained as generalized Bayes estimators using appropriate priors. The admissibility and minimax properties of Pitman estimators were discussed. In the last section we have performed an application of equivariant estimators in the study of maximum rainfall in Piracicaba-SP.

Keywords: Decision theory. Estimation theory. Risk function. Group action. Pitman estimator.

LISTA DE FIGURAS

| | | |
|----------|--|-----|
| Figura 1 | Diagrama comutativo da definição de um estimador invariante $\delta(g \cdot \mathbf{X}) = \delta(\mathbf{X})$ | 32 |
| Figura 2 | Diagrama comutativo da definição de um estimador equivariante $\delta(g \cdot \mathbf{X}) = g \cdot (\delta(\mathbf{X}))$ | 33 |
| Figura 3 | Representação da órbita que corta o plano no ponto (x_1, \dots, x_n) . | 44 |
| Figura 4 | Máximos nos blocos 1, 2, 3 e 4. | 155 |
| Figura 5 | Excessos acima de um limiar u | 155 |
| Figura 6 | Gráfico da média dos excessos. | 157 |
| Figura 7 | Função densidade de probabilidade da distribuição Generalizada de Pareto para $\xi = -0,4$ (Beta), $\xi = 0,4$ (Pareto) e $\xi \rightarrow 0$ (Exponencial), com $u = 0$ e $\sigma_u = 30$ | 161 |

LISTA DE TABELAS

| | | |
|----------|---|-----|
| Tabela 1 | Estatísticas descritivas dos dados de precipitação máxima mensal (mm) em Piracicaba-SP, no período de 1917 a 2015. | 169 |
| Tabela 2 | <i>Thresholds</i> escolhidos e estimativas dos parâmetros das distribuições Generalizada de Pareto (GP) e Exponencial _{MV} , obtidas via método da máxima verossimilhança. | 170 |
| Tabela 3 | Resultados dos testes (valor <i>p</i>) de independência (Ljung-Box), de aderência (Kolmogorov-Smirnov) e do teste da razão de verossimilhanças (TRV), referentes ao ajuste das distribuições Generalizada de Pareto (GP) e Exponencial _{MV} aos dados de precipitação máxima mensal em Piracicaba-SP. | 171 |
| Tabela 4 | Estimativas equivariantes dos parâmetros da distribuição Exponencial _E e resultados do teste Kolmogorov-Smirnov (valor <i>p</i>) referentes ao seu ajuste aos dados de precipitação máxima mensal em Piracicaba-SP. | 172 |
| Tabela 5 | Precipitações máximas (mm) esperadas em Piracicaba-SP para os tempos de retorno de 10, 30, 50, 70 e 100 anos, nos meses de outubro a março, obtidas via distribuições Exponencial _{MV} e Exponencial _E | 173 |
| Tabela 6 | Erros quadráticos médios (EQM's) mensais referentes aos níveis máximos esperados de precipitação em Piracicaba-SP, obtidos via distribuições Exponencial _{MV} (Exp _{MV}) e Exponencial _E (Exp _E). | 174 |

LISTA DE ABREVIATURAS

| | |
|--------------|---|
| BLUE | Best linear unbiased estimator |
| Coef. | Coeficiente |
| CV | Coeficiente de variação |
| EQM | Erro quadrático médio |
| fda | Função de distribuição acumulada |
| fdp | Função densidade de probabilidade |
| GP | Generalized Pareto |
| iid | Independente e identicamente distribuída |
| KS | Kolmogorov-Smirnov |
| mm | Milímetro |
| MRE | Minimum risk equivariant |
| POT | Peaks Over Threshold |
| TRV | Teste da razão de verossimilhanças |
| TVE | Teoria de Valores Extremos |
| UMVUE | Uniformly minimum-variance unbiased estimator |

LISTA DE SIGLAS

ESALQ Escola Superior de Agricultura “Luiz de Queiroz”

IBGE Instituto Brasileiro de Geografia e Estatística

LISTA DE SÍMBOLOS

| | |
|-----------------------------|--|
| Θ | Conjunto dos possíveis estados da natureza ou espaço paramétrico |
| \mathcal{A} | Conjunto das ações ou conjunto das estimativas |
| $L(\cdot, \cdot)$ | Função perda |
| α | Decisão assumida em Θ |
| \mathfrak{X}^n | Espaço Euclidiano ou espaço amostral |
| D | Espaço das decisões ou espaço dos estimadores |
| $\delta(\cdot)$ | Função de decisão ou estimador |
| G | Grupo |
| $(G, +)$ | Conjunto não-vazio G munido da operação soma |
| $f_X(x; \theta)$ | Função densidade de probabilidade da variável aleatória X , indexada em θ |
| $\mathcal{R}(\cdot, \cdot)$ | Função risco |
| $E(\cdot)$ | Esperança matemática |
| \sim | Possui distribuição ou é distribuído como |
| $P(\cdot)$ | Probabilidade |
| $\exp(\cdot)$ | Função Exponencial |
| \mathbf{V} | Espaço dos vetores de médias |
| $\text{var}(\cdot)$ | Variância |

| | |
|---------------------------|--|
| χ_n^2 | Variável aleatória Qui-Quadrada com n graus de liberdade |
| $\Gamma(\cdot)$ | Função Gama |
| π | Número Pi |
| $\pi(\cdot)$ | Distribuição a priori |
| $\pi(\cdot \mathbf{x})$ | Distribuição a posteriori |
| $\delta_\pi(\cdot)$ | Estimador de Bayes |
| Θ | Variável aleatória |
| \mathcal{D} | Deviance |

SUMÁRIO

| | | |
|--------|---|-----|
| 1 | INTRODUÇÃO | 17 |
| 2 | ESTIMAÇÃO EQUIVARIANTE | 21 |
| 2.1 | Elementos da teoria da decisão | 21 |
| 2.2 | A teoria da decisão, ação de grupos e o princípio da equivari- ância | 23 |
| 2.3 | A teoria da decisão no contexto estatístico | 25 |
| 2.4 | Equivariância para parâmetros de locação | 38 |
| 2.4.1 | Ação de grupos para parâmetros de locação | 40 |
| 2.4.2 | Estimação equivariante em modelos de regressão linear múl- tipla | 61 |
| 2.5 | Equivariância para parâmetros de escala | 68 |
| 2.5.1 | Ação de grupos para parâmetros de escala | 70 |
| 2.6 | Estimação equivariante em famílias de locação-escala | 91 |
| 3 | ESTIMADORES DE PITMAN NO CONTEXTO BAYESIANO | 106 |
| 4 | ADMISSIBILIDADE E MINIMAXIMALIDADE DE ESTI- MADORES | 124 |
| 5 | ESTUDO DE CASO: APLICAÇÃO A NÍVEIS PLUVIOMÉ- TRICOS MÁXIMOS | 151 |
| 5.1 | Introdução | 151 |
| 5.2 | Material e métodos | 153 |
| 5.2.1 | Dados | 153 |
| 5.2.2 | Composição das amostras de precipitações máximas mensais | 154 |
| 5.2.3 | A escolha do limiar | 155 |
| 5.2.4 | Análise exploratória dos dados | 158 |
| 5.2.5 | Teste de independência de Ljung-Box | 158 |
| 5.2.6 | Distribuição Generalizada de Pareto (GP) | 159 |
| 5.2.7 | Estimação dos parâmetros | 161 |
| 5.2.8 | Teste de aderência de Kolmogorov-Smirnov | 163 |
| 5.2.9 | Teste da razão de verossimilhanças | 164 |
| 5.2.10 | Níveis de retorno | 165 |
| 5.2.11 | Seleção da distribuição com resultados mais acurados | 168 |
| 5.2.12 | Recursos computacionais | 168 |
| 5.3 | Resultados e discussão | 169 |
| 5.4 | Conclusões | 175 |
| 6 | CONSIDERAÇÕES FINAIS | 176 |
| 7 | APÊNDICE | 182 |

1 INTRODUÇÃO

Um interesse que surge por parte de um pesquisador em uma determinada pesquisa científica é obter algum tipo de conhecimento acerca da população que esteja sob análise. Por sua vez, a população, que pode ser definida como o conjunto de todos os elementos sob investigação, é caracterizada por números que são chamados parâmetros, do grego “quase medições”. Nesse sentido, o que se deseja em uma pesquisa científica é obter informações sobre tais parâmetros desconhecidos. Por alguns motivos é impossível ou inviável observar todos os elementos populacionais, e dessa forma, o objetivo de se obter algum tipo de conhecimento da população sob análise ou dos parâmetros que a caracterizam é, na maioria das vezes, alcançado com base na observação de um subconjunto representativo dessa população, denominado amostra.

Os métodos que permitem a obtenção de informações acerca dos parâmetros populacionais, com base em uma amostra, podem ser englobados nos métodos da inferência estatística. Logo, ao traçar especulações sobre o modelo probabilístico populacional utilizando métodos da inferência estatística, o pesquisador objetiva obter aproximações numéricas para seus parâmetros populacionais. Mais especificamente, o procedimento de inferir um valor para um parâmetro populacional na inferência estatística é denominado estimação paramétrica. Por conseguinte, algumas funções da amostra, conhecidas como estatísticas, podem então ser utilizadas para essa finalidade. Essas funções são então denominadas de estimadores.

Alguns métodos já consagrados na literatura para determinar estimadores pontuais de parâmetros populacionais são o método dos momentos, o método da máxima verossimilhança, o método dos mínimos quadrados e o método bayesiano. Naturalmente, diferentes métodos podem fornecer um mesmo estimador, o que

torna a análise geralmente mais simples. Porém isso nem sempre ocorre, ou seja, diferentes métodos geralmente fornecem diferentes estimadores para um mesmo parâmetro populacional, o que exige uma forma de avaliar esses estimadores.

A avaliação de estimadores é parte de um ramo estatístico conhecido como teoria da decisão estatística. Com elementos da teoria da decisão estatística, um critério utilizado para determinar um estimador mais adequado é adotar o estimador que minimiza a função risco, sendo essa função a esperança matemática de uma função perda apropriada. Todavia, nem sempre é possível encontrar tal estimador com base nesse critério, uma vez que a classe de todos os estimadores é muito grande (LEHMANN; CASELLA, 1998, p. 5; CASELLA; BERGER, 2002, p. 334). Uma forma de contornar esse problema é realizar algum tipo de restrição na classe dos estimadores, com base em propriedades que sejam comuns a alguns deles. Essa estratégia permite, de um modo natural, reduzir a classe dos estimadores sujeitos ao estudo, de maneira que as reduções obtidas resultem possivelmente na determinação de estimadores mais adequados de interesse.

Uma maneira de limitar a classe dos estimadores é considerar somente estimadores não viesados. Levando-se em conta somente os estimadores não viesados, o objetivo nessa classe de estimadores é determinar um estimador que possua variância menor ou igual em relação a qualquer outro estimador não viesado. Outra forma que pode ser adotada na restrição aos estimadores é utilizando-se o princípio da equivariância. Com base nesse princípio, Edwin James George Pitman (1897-1993) introduziu em 1939 os melhores estimadores na classe dos estimadores equivariantes, e devido ao seu trabalho, esses ficaram conhecidos posteriormente como estimadores de Pitman (PITMAN, 1939). Pitman os chamou inicialmente de estimadores “adequados”. O termo “equivariante” foi usado primeiramente por Berk (1967) e Wijsman (1967).

Diante do que foi exposto, o presente trabalho têm como objetivos gerais:

- Apresentar a fundamentação teórica necessária para o desenvolvimento da teoria da estimação equivariante.
- Exemplificar a utilização dos estimadores equivariantes com uma aplicação aos dados de precipitação máxima de Piracicaba-SP.

Além dessa breve introdução, o presente trabalho é composto das seções 2, 3, 4 e 5, que são detalhadas a seguir. Nas seções 2, 3 e 4 apresenta-se a fundamentação teórica do trabalho. Nas subseções de 2.1 a 2.3 são apresentados os elementos básicos referentes à teoria da decisão, considerando-se o contexto equivariante e estatístico. Nas subseções 2.4 e 2.5 discute-se a estimação equivariante para parâmetros de locação e de escala, fornecendo-se algumas definições e resultados requeridos para determinar o estimador de Pitman em cada caso. Na subseção 2.6 é abordado o problema de estimação simultânea dos parâmetros de locação e de escala em famílias de locação-escala. Na seção 3 são revisados resultados clássicos da teoria bayesiana, mostrando-se ao final que os estimadores de Pitman, para parâmetros de locação e de escala, podem ser obtidos como estimadores de Bayes generalizados utilizando-se prioris adequadas. Na seção 4 são apresentados alguns resultados sobre a admissibilidade e minimaximalidade de estimadores. Ao final dessa seção abordam-se, de forma geral, as propriedades de admissibilidade e a minimaximalidade dos estimadores de Pitman. Na seção 5 é apresentada uma aplicação dos estimadores equivariantes, referente ao estudo dos níveis pluviométricos máximos em Piracicaba-SP. O objetivo geral nessa seção é comparar a distribuição Exponencial biparamétrica, ajustada com estimadores equivariantes e de máxima verossimilhança, na modelagem dos dados de precipitação máxima nesse município.

O referencial teórico deste trabalho baseia-se principalmente no livro *Theory of Point Estimation*, de Erich Leo Lehmann e George Casella. Uma vez que as demonstrações de alguns resultados nesse livro são apresentadas de forma muito sucinta e a literatura disponível em português é muito restrita em relação à teoria da estimação equivariante, as passagens matemáticas no referencial teórico são desenvolvidas em detalhes, podendo-se afirmar que essa é uma das contribuições deste trabalho. Além disso, outra contribuição deste trabalho foi apresentar uma aplicação envolvendo estimadores equivariantes.

2 ESTIMAÇÃO EQUIVARIANTE

2.1 Elementos da teoria da decisão

A teoria da decisão é uma área interdisciplinar que recebeu contribuições de economistas, matemáticos, filósofos, cientistas sociais e estatísticos. Como o próprio nome sugere, a teoria da decisão está relacionada ao problema de tomada de decisões, em que busca-se compreender como indivíduos ou grupos deveriam ou poderiam tomar decisões. Em um nível teórico, na teoria da decisão busca-se pesquisar as consequências lógicas que podem surgir com a utilização de diferentes regras para a tomada de decisões e com recursos matemáticos busca-se compreender como se combinam decisões racionais (RESNIK, 1987, p. 3).

A teoria da decisão é uma área intimamente relacionada à teoria dos jogos. Primeiramente, o desenvolvimento inicial de ambas teorias ocorreu principalmente na primeira metade do século XX, embora trabalhos em ambas as teorias já tivessem sido desenvolvidos anteriormente. Além disso, os elementos da teoria da decisão são muito semelhantes àqueles da teoria dos jogos. Em particular, na teoria da decisão, um dos “jogadores” é a própria natureza (ao invés de um jogador propriamente dito), e esse “jogador age” por intermédio de uma experiência aleatória. Em vista disso, os elementos básicos fundamentais em um problema da teoria da decisão são:

1. Um conjunto não vazio de números reais Θ , dos possíveis estados da natureza.
2. Um conjunto não vazio \mathcal{A} de números reais de todas as ações (ou decisões) possíveis que podem ser tomadas.
3. Uma função perda $L(\alpha, \theta)$, de valores reais, definida em $\mathcal{A} \times \Theta$.
4. Um subespaço \mathcal{X}^n do espaço euclidiano \mathbb{R}^n .

5. O espaço $D = \{\delta : \mathfrak{X}^n \rightarrow \mathcal{A}\}$, em que as funções δ são denominadas funções de decisão.

Com esses elementos básicos, pode-se observar que, para qualquer decisão α que seja tomada em relação a θ , incorre-se em uma perda $L(\alpha, \theta)$. Para problemas da teoria da decisão, mas em um contexto mais amplo como o da teoria dos jogos, pode-se utilizar a ideia de um jogo entre dois jogadores, fornecida por Ferguson (1967, p. 1), em que um desses jogadores é a natureza. Esse jogo consiste de duas escolhas, em que a escolha θ representa o “verdadeiro estado da natureza” e α denota uma ação a ser realizada por um indivíduo qualquer. Dessa forma, a natureza escolhe um ponto θ em Θ , e o indivíduo, sem ser informado da escolha que a natureza tenha feito, toma uma decisão α em \mathcal{A} . O erro cometido pelo indivíduo surge então naturalmente como consequência dessas duas escolhas, e esse erro, também chamado de perda, pode ser quantificado pela função perda.

Definição 2.1.1 (Função perda): Chama-se *função perda* L , uma função $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$ que associa ao par (α, θ) um número real que indica a perda que o indivíduo tem ao tomar uma decisão α , quando o real estado da natureza é θ .

A função perda é uma função não negativa e a sua escolha, em um determinado problema, pode ser considerada até certo ponto arbitrária. Existem algumas funções perda que são “naturais”, ou pelo menos são mais comuns, como a utilização das funções perda em erro quadrático ou em erro absoluto. Algumas funções perda usuais são:

(i) $L_1(\alpha, \theta) = (\alpha - \theta)^2$.

(ii) $L_2(\alpha, \theta) = |\alpha - \theta|$.

$$\text{(iii)} \quad L_3(\alpha, \theta) = \begin{cases} A, & \text{se } |\alpha - \theta| > \varepsilon \\ 0, & \text{se } |\alpha - \theta| \leq \varepsilon \end{cases}, \text{ em que } A > 0.$$

$$\text{(iv)} \quad L_4(\alpha, \theta) = \rho(\theta) |\alpha - \theta|^r, \text{ para } \rho(\theta) \geq 0 \text{ e } r > 0.$$

$$\text{(v)} \quad L_5(\alpha, \theta) = \begin{cases} 1, & \text{se } |\alpha - \theta| > \varepsilon \\ 0, & \text{caso contrário} \end{cases}.$$

As funções L_1 e L_2 são chamadas função perda em erro quadrático e função perda em erro absoluto, respectivamente. Nota-se que o erro obtido por essas duas funções aumenta à medida que a distância entre α e θ aumenta. A função perda L_3 indica que não ocorre perda alguma se α está dentro de ε unidades de θ , e caso contrário, a perda é uma quantidade A . A função perda L_4 é chamada de função perda em erro quadrático ponderado, sendo essa uma função perda geral que inclui L_1 e L_2 como casos especiais. Por último, a função L_5 é chamada função perda 0-1.

2.2 A teoria da decisão, ação de grupos e o princípio da equivariância

Em muitas situações, as decisões são tomadas em um contexto mais restritivo. Algumas dessas restrições podem ser modeladas considerando algum tipo de simetria. Em estatística, o termo matemático para simetria é invariância (GIRI, 1997, p. 16; VIANA, 2008, p. 160-161). O termo simetria é geralmente utilizado em seu sentido matemático, para denotar uma propriedade que permanece inalterada após uma transformação particular. Na prática, muitos problemas estatísticos possuem essa propriedade.

As simetrias são tratadas na matemática utilizando-se a teoria dos grupos, que foi desenvolvida por matemáticos nos séculos XIX e XX e que se tornou, rapidamente, uma ferramenta importante na organização e no estudo de muitas partes da matemática. De forma simples, um grupo constitui um conjunto de elementos

que podem ser combinados utilizando-se algumas operações, como por exemplo a adição ou a multiplicação. Tais operações devem satisfazer determinadas condições. Uma propriedade importante de um grupo é que, quando dois de seus elementos são combinados, utilizando-se uma determinada operação, o resultado é outro elemento do grupo. Diz-se então que o grupo é fechado sob aquela operação. No Apêndice A é apresentada uma definição mais formal do conceito de grupo. Além disso, algumas referências mais abrangentes sobre a teoria dos grupos, que podem ser consultadas, são Alencar Filho (1985), Souza (2009) e Martin (2010).

O princípio da equivariância está relacionado às ações de um grupo G sobre os quatro elementos da teoria da decisão, isto é, um problema de decisão é equivariante quando ocorrem as ações de um grupo G em Θ , \mathcal{A} , \mathfrak{X}^n e em D , da seguinte forma:

1. $G \times \Theta \rightarrow \Theta$
 $(g, \theta) \mapsto g \cdot \theta$
2. $G \times \mathcal{A} \rightarrow \mathcal{A}$
 $(g, \mathfrak{a}) \mapsto g \cdot \mathfrak{a}$
3. $G \times \mathfrak{X}^n \rightarrow \mathfrak{X}^n$
 $(g, (X_1, \dots, X_n)) \mapsto g \cdot (X_1, \dots, X_n)$
4. $G \times D \rightarrow D$, em que $g \cdot \delta(X_1, \dots, X_n) = \delta(g \cdot (X_1, \dots, X_n))$.
 $(g, \delta) \mapsto g \cdot \delta$

Observação: No presente trabalho será utilizada a notação \mathbf{X} para um vetor de observações (X_1, \dots, X_n) . Como consequência, a ação de um grupo G sobre um vetor (X_1, \dots, X_n) também será denotada por $g \cdot \mathbf{X}$.

Exemplo 2.2.1: Seja $\Theta = \mathcal{A} = \mathbb{R}$, $\mathfrak{X}^n = \mathbb{R}^n$ e $G = (\mathbb{R}, +)$. As ações de G correspondem a soma à direita, dadas por:

1. $G \times \Theta \rightarrow \Theta$, em que $(g, \theta) \mapsto g \cdot \theta = \theta + g$.
2. $G \times \mathcal{A} \rightarrow \mathcal{A}$, em que $(g, a) \mapsto g \cdot a = a + g$.
3. $G \times \mathfrak{X}^n \rightarrow \mathfrak{X}^n$, em que $(g, \mathbf{X}) \mapsto g \cdot \mathbf{X} = (X_1 + g, \dots, X_n + g)$.
4. $G \times D \rightarrow D$, em que $(g, \delta) \mapsto g \cdot \delta(\mathbf{X}) = \delta(X_1 + g, \dots, X_n + g)$.

Exemplo 2.2.2: Seja $\Theta = \mathcal{A} = \mathbb{R}$, $\mathfrak{X}^n = \mathbb{R}^n$ e $G = (\mathbb{R}^+, \cdot)$. As ações de G correspondem a multiplicação à esquerda, definidas por:

1. $G \times \Theta \rightarrow \Theta$, em que $(g, \theta) \mapsto g \cdot \theta = g\theta$.
2. $G \times \mathcal{A} \rightarrow \mathcal{A}$, em que $(g, a) \mapsto g \cdot a = ga$.
3. $G \times \mathfrak{X}^n \rightarrow \mathfrak{X}^n$, em que $(g, \mathbf{X}) \mapsto g \cdot \mathbf{X} = (gX_1, \dots, gX_n)$.
4. $G \times D \rightarrow D$, em que $(g, \delta) \mapsto g \cdot \delta(\mathbf{X}) = \delta(gX_1, \dots, gX_n)$.

Nesse exemplo, a ação de G no conjunto Θ , por exemplo, é representada por $g \cdot \Theta$, correspondendo ao produto usual que é denotado por $g\Theta$. Essa notação também será adotada nos exemplos posteriores em relação ao produto usual, de forma a tornar clara a distinção entre uma ação do grupo, representada por “ \cdot ”, e a operação de produto “ \cdot ”.

2.3 A teoria da decisão no contexto estatístico

A teoria dos jogos e a teoria da decisão estatística foram desenvolvidas por volta do mesmo período. A primeira é parte de um rico legado matemático deixado por John von Neumann. Embora já estivesse presente em trabalhos anteriores, a teoria dos jogos não era um campo unificado até a publicação dos trabalhos de von

Neumann no fim da década de 1920. Em 1927, von Neumann publicou a prova do teorema *minimax* para jogos finitos e esse trabalho é reconhecido atualmente como a fundação da teoria dos jogos. Em 1944, os trabalhos de von Neumann resultaram na publicação do livro *Theory of Games and Economic Behavior*, escrito em parceria com Oskar Morgenstern.

Por sua vez, por volta do mesmo período, a teoria da decisão estatística estava recebendo rigorosa fundamentação teórica em uma série de artigos publicados por Jerzy Neyman e Egon Pearson. Essa teoria já havia sido desenvolvida inicialmente por outros estatísticos, como Karl Pearson e Ronald Aylmer Fisher, mas faltava a precisa formulação matemática, fornecida posteriormente por Neyman e Pearson em seus trabalhos. Abraham Wald foi o primeiro a estabelecer a conexão entre a teoria dos jogos e a teoria estatística de Neyman e Pearson, reconhecendo as vantagens de basear a teoria estatística na teoria dos jogos. Wald foi o responsável por generalizar e simplificar a teoria de Neyman-Pearson, por tratar problemas considerados distintos nessa teoria, como casos especiais da teoria da decisão. Os trabalhos de Wald na década de 1940 culminaram na publicação de seu livro *Statistical Decision Functions* em 1950, ano que também foi marcado por sua morte em um trágico acidente aéreo (FERGUSON, 1967).

Na teoria da decisão estatística, a tomada de decisões é realizada em problemas de decisão que possuem algum tipo de incerteza envolvida. Dessa maneira, o contexto geral da teoria da decisão pode ser particularizado para o contexto da teoria da decisão estatística, em que as decisões tomadas baseam-se em observações de amostras aleatórias. Logo, os elementos da teoria da decisão são redefinidos para o contexto estatístico:

1. O conjunto Θ passa a ser o espaço paramétrico que indexa uma família de distribuições de probabilidade $f_X(x; \theta)$.

2. O conjunto das ações \mathcal{A} é simplesmente o conjunto de estimativas acerca de θ e será representado geralmente pelo conjunto \mathbb{R} dos números reais.
3. A função perda $L(\delta(\mathbf{x}), \theta)$, em que $\delta(\mathbf{x})$ é um valor estimado do parâmetro desconhecido θ ou o valor de alguma função $\omega(\theta)$.
4. O subespaço \mathfrak{X}^n passa a ser o espaço amostral de todas as realizações de amostras independentes e identicamente distribuídas (iid) (X_1, \dots, X_n) de uma variável aleatória X .
5. As funções de decisão passam a ser denominadas estimadores e o espaço $D = \{\delta : \mathfrak{X}^n \rightarrow \mathbb{R}\}$ torna-se o espaço dos estimadores.

Em vista dessas redefinições, o estatístico ou outro cientista deve selecionar um valor para um parâmetro desconhecido θ , que determina a distribuição $f_X(x; \theta)$ da população de interesse. Após a observação de uma amostra aleatória (X_1, \dots, X_n) dessa população, a escolha de um valor para θ se baseará nos valores dessa amostra e esse valor será chamado estimativa de θ . Assim, pode-se proceder da seguinte forma:

1. Para o problema em questão, define-se o conjunto de todos os possíveis valores que θ pode assumir. Como mencionado, tal conjunto de valores de θ é o espaço paramétrico, sendo designado por Θ .
2. Seleciona-se uma função δ da amostra aleatória, isto é, $\delta(X_1, \dots, X_n)$.
3. Uma estimativa para o parâmetro θ ou para uma função de θ , digamos $\omega(\theta)$, é obtida aplicando-se a função δ nos valores observados da amostra.

Exemplo 2.3.1: Seja $f_X(x; \theta, 1)$ uma densidade Normal com média θ e variância

1. Supondo que θ possa assumir qualquer valor real, o espaço paramétrico Θ é dado por:

$$\Theta = \{\theta : -\infty < \theta < \infty\}.$$

Como θ pode assumir qualquer valor real, é natural considerar que uma estimativa razoável de θ também deva assumir qualquer valor real. Essa estimativa pode ser representada por $\hat{\theta}$ ou por $\delta(\mathbf{x})$ com base no estimador δ , quando valores (x_1, \dots, x_n) de uma amostra aleatória (X_1, \dots, X_n) são observados. Pode-se então definir uma estimativa distinta para cada valor de θ em Θ . Dessa forma, o conjunto de estimativas possíveis em relação a θ no processo de estimação é:

$$\mathcal{A} = \{\hat{\theta} : -\infty < \hat{\theta} < \infty\}.$$

Em problemas de estimação pontual é comum que $\Theta = \mathcal{A}$. Neste exemplo, os conjuntos Θ e \mathcal{A} são ambos iguais a \mathbb{R} . Como um estimador δ_1 para θ pode-se utilizar, por exemplo, a média amostral $\delta_1 = \bar{X}$. Nesse caso, com base nos valores observados da amostra aleatória (X_1, \dots, X_n) , uma estimativa para θ é:

$$\hat{\theta} = \delta_1(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Naturalmente, poderiam ser utilizados outros estimadores para θ , como a mediana amostral ou outras funções. Uma vez que para cada problema particular existem muitas funções diferentes que podem ser utilizadas como estimadores, faz-se necessário utilizar ferramentas que permitam avaliar os estimadores, para seleccionar os mais adequados. Nesse sentido, um objetivo na teoria estatística é seleccionar dentre uma certa quantidade de estatísticas, funções de (X_1, \dots, X_n) , aquelas que sejam mais adequadas para serem utilizadas como estimadores dos θ_j , $j = 1, \dots, k$, ou mais geralmente, encontrar estatísticas para serem usadas como

estimadores de certas funções, digamos $\omega_1(\theta_1, \dots, \theta_k), \dots, \omega_k(\theta_1, \dots, \theta_k)$.

Considerando que θ é um parâmetro desconhecido, para cada estimativa $\delta(\mathbf{x})$ de $\omega(\theta)$ comete-se um erro, e nesse caso, alguma medida de avaliação dos erros cometidos é apropriada. Uma maneira de avaliar os erros cometidos é utilizar a função perda. Uma definição formal de função perda para o contexto de estimação é apresentada a seguir.

Definição 2.3.1 (Função perda): Chama-se *função perda*, a função de valores reais denotada por $L(\delta(\mathbf{x}), \theta)$, que satisfaz:

- (i) $L(\delta(\mathbf{x}), \theta) \geq 0$, para todas possíveis estimativas $\delta(\mathbf{x})$ e todo θ em Θ ;
- (ii) $L(\delta(\mathbf{x}), \theta) = 0$, para $\delta(\mathbf{x}) = \omega(\theta)$.

Uma perda $L(\delta(\mathbf{x}), \theta)$ ocorre se uma estimativa para $\omega(\theta)$ é $\delta(\mathbf{x})$, quando o verdadeiro valor do parâmetro é θ . Além disso, as condições anteriores estabelecem que a função perda deve ser não negativa. Segundo Casella e Berger (2002, p. 348-349), a função perda reflete dois fatos que ocorrem num processo de estimação pontual. O primeiro está relacionado a uma pequena perda observada, quando a estimativa $\delta(\mathbf{x})$ está próxima de θ . Nesse caso, assumir que $\delta(\mathbf{x})$ é uma estimativa para θ parece razoável. O segundo fato reflete uma grande perda, ocasionada se $\delta(\mathbf{x})$ está distante de θ .

Para cada problema de estimação pode-se definir uma função perda apropriada para o problema em estudo. As funções perda fornecidas na seção relacionada à teoria da decisão são apresentadas novamente, mas considerando-se o processo de estimação pontual.

(i) $L_1(\delta(\mathbf{x}), \theta) = [\delta(\mathbf{x}) - \omega(\theta)]^2$.

(ii) $L_2(\delta(\mathbf{x}), \theta) = |\delta(\mathbf{x}) - \omega(\theta)|$.

$$(iii) L_3(\delta(\mathbf{x}), \theta) = \begin{cases} A, & \text{se } |\delta(\mathbf{x}) - \omega(\theta)| > \varepsilon \\ 0, & \text{se } |\delta(\mathbf{x}) - \omega(\theta)| \leq \varepsilon \end{cases}, \text{ em que } A > 0.$$

$$(iv) L_4(\delta(\mathbf{x}), \theta) = \rho(\theta) |\delta(\mathbf{x}) - \omega(\theta)|^r, \text{ para } \rho(\theta) \geq 0 \text{ e } r > 0.$$

$$(v) L_5(\delta(\mathbf{x}), \theta) = \begin{cases} 1, & \text{se } |\delta(\mathbf{x}) - \omega(\theta)| > \varepsilon \\ 0, & \text{caso contrário} \end{cases}.$$

Algumas das funções perda apresentadas possuem a propriedade de serem simétricas. Este é o caso das funções perda em erro quadrático (L_1) e em erro absoluto (L_2). A função perda em erro quadrático é provavelmente a mais utilizada no processo de estimação. Uma característica importante dessa função é que ela penaliza pouco se o erro de estimação é pequeno e penaliza muito à medida que o erro torna-se maior. Por sua vez, a penalização para a função perda em erro absoluto cresce linearmente com o erro de estimação. As funções perda L_3 e L_5 são funções que reduzem o efeito para grandes erros de estimação. Isso ocorre pela fixação de uma perda caso o desvio absoluto entre $\delta(\mathbf{x})$ e $\omega(\theta)$ exceda uma quantidade ε pré-determinada. Nessa situação, as funções perda L_3 e L_5 associam uma perda fixa para o erro de estimação cometido, não considerando a magnitude do eventual erro.

Pode-se verificar, pela definição e pelos exemplos, que a função perda possui como um de seus argumentos uma estimativa $\delta(\mathbf{x})$, que é o valor de um estimador δ , que por sua vez é função de uma amostra aleatória. Dessa forma, como $\delta(X_1, \dots, X_n)$, a função perda depende da amostra (X_1, \dots, X_n) . Neste caso, é difícil exigir uma pequena perda (ou erro), por parte de um estimador, para cada amostra que é possível de ser coletada da população. Porém, se alterarmos o objetivo de encontrar um estimador com pequena perda para a escolha de um estimador com pequena perda média, essa dificuldade pode ser superada. Assim, um critério a ser adotado para avaliar a qualidade de um estimador consiste em utilizar

a esperança matemática da função perda. Isso define uma função \mathcal{R} , denominada função risco.

Definição 2.3.2 (Função risco): Para uma conhecida função perda $L(\delta(\mathbf{x}), \theta)$, a função risco \mathcal{R} de um estimador $\delta(X_1, \dots, X_n)$ é definida como:

$$\mathcal{R}(\delta, \theta) = E_{\theta} [L(\delta(\mathbf{X}), \theta)].$$

Dado que o valor de θ é desconhecido, a preferência será por um estimador que possua pequeno valor de $\mathcal{R}(\delta, \theta)$ para todos valores de θ . Logo, independentemente do verdadeiro valor de θ , é desejável encontrar um estimador com pequena perda média. Se o objetivo for, por exemplo, comparar a qualidade de dois estimadores δ_1 e δ_2 , isso poderá ser feito com base em suas funções risco. Nesse caso, se $\mathcal{R}(\delta_1, \theta) < \mathcal{R}(\delta_2, \theta)$, o estimador δ_1 será o estimador preferido. Todavia, essa comparação pode não ser tão clara se as duas funções risco se interceptarem (MOOD; GRAYBILL; BOES, 1974, p. 292, 299; CASELLA; BERGER, 2002, p. 332-333).

Em termos de uma amostra aleatória (X_1, \dots, X_n) , tem-se que a função risco \mathcal{R} é dada por:

$$\mathcal{R}(\delta, \theta) = E_{\theta} [L(\delta(\mathbf{X}), \theta)] = \int \dots \int L(\delta(\mathbf{x}), \theta) \prod_{i=1}^n f_{X_i}(x_i; \theta) dx_i.$$

No contexto em que têm-se ações de grupos, são importantes os estimadores que possuam boas propriedades relativas às ações de um grupo. Nesse sentido, definem-se estimadores invariantes e equivariantes.

Definição 2.3.3 (Estimador invariante): Um estimador pontual δ em \mathfrak{X}^n é invariante em relação às ações de um grupo G em \mathfrak{X}^n se:

$$(g \cdot \delta)(\mathbf{X}) = \delta(g \cdot \mathbf{X}) = \delta(\mathbf{X}).$$

Na Figura 1 é apresentado o diagrama comutativo de um estimador invariante δ . Esse diagrama representa duas possíveis rotas de estimação de \mathfrak{X}^n para \mathcal{A} , dadas por $\mathfrak{X}^n \xrightarrow{g} \mathfrak{X}^n \xrightarrow{\delta} \mathcal{A}$ e $\mathfrak{X}^n \xrightarrow{\delta} \mathcal{A}$, que são equivalentes para um estimador invariante.

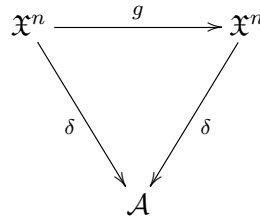


Figura 1 Diagrama comutativo da definição de um estimador invariante $\delta(g \cdot \mathbf{X}) = \delta(\mathbf{X})$.

Exemplo 2.3.2: Seja $\mathbf{X} = (X_1, \dots, X_n)$. Pode-se observar que a amplitude amostral é invariante em relação à ação do grupo $G = (\mathbb{R}, +)$. De fato, dado $c \in \mathbb{R}$, tem-se que:

$$\delta(c \cdot \mathbf{X}) = \delta(X_1 + c, \dots, X_n + c) = Y_n + c - (Y_1 + c) = Y_n - Y_1 = \delta(\mathbf{X}),$$

em que $Y_1 = \min[X_1, \dots, X_n]$ e $Y_n = \max[X_1, \dots, X_n]$.

O termo equivariante é utilizado para estimadores que apresentam a propriedade de modificar suas estimativas em resposta a mudanças que ocorram no conjunto de dados. Essencialmente, a equivariância indica qual é o valor do estimador em qualquer ponto de uma órbita (ver APÊNDICE A), caso seja conhecido um valor particular do estimador em um ponto dessa órbita.

Definição 2.3.4 (Estimador equivariante): Um estimador pontual δ em \mathfrak{X}^n é equivariante em relação às ações de um grupo G em \mathfrak{X}^n se:

$$(g \cdot \delta)(\mathbf{X}) = \delta(g \cdot \mathbf{X}) = g \cdot (\delta(\mathbf{X})).$$

Na Figura 2 é apresentado o diagrama comutativo de um estimador equivariante δ . Nesse diagrama, as duas rotas equivalentes de estimação de \mathfrak{X}^n para \mathcal{A} , para um estimador equivariante, são $\mathfrak{X}^n \xrightarrow{g} \mathfrak{X}^n \xrightarrow{\delta} \mathcal{A}$ e $\mathfrak{X}^n \xrightarrow{\delta} \mathcal{A} \xrightarrow{g} \mathcal{A}$.

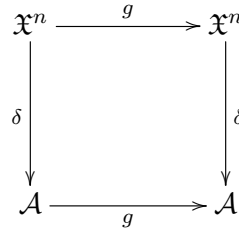


Figura 2 Diagrama comutativo da definição de um estimador equivariante $\delta(g \cdot \mathbf{X}) = g \cdot (\delta(\mathbf{X}))$.

Exemplo 2.3.3: Considere $\mathbf{X} = (X_1, \dots, X_n)$. Pode-se verificar que a média amostral de (X_1, \dots, X_n) é equivariante em relação à ação do grupo $G = (\mathbb{R}, +)$. De fato, dado $c \in \mathbb{R}$ tem-se:

$$\begin{aligned}
 (c \cdot \delta)(\mathbf{X}) &= \delta(c \cdot \mathbf{X}) = \delta(X_1 + c, \dots, X_n + c) \\
 &= \frac{X_1 + c + \dots + X_n + c}{n} \\
 &= \frac{X_1 + \dots + X_n}{n} + \frac{nc}{n} \\
 &= \delta(\mathbf{X}) + c \\
 &= c \cdot (\delta(\mathbf{X})).
 \end{aligned}$$

As definições utilizadas para invariância e equivariância remetem a uma consideração importante. Assim como em Zacks (1971), Schervish (1995), Lehmann e Casella (1998) e Nobre e Azevedo (2006), neste trabalho faz-se a distinção entre invariância e equivariância. Na invariância, a estimativa não se altera à medida que os dados são transformados, enquanto que na equivariância a estimativa se modifica de um modo pré-estabelecido, à medida que os dados são transformados. Em termos de funções, as funções que satisfazem $g(x+c) = g(x)$ são invariantes, ou seja, essas funções possuem a propriedade de seus valores não se alterarem quando seus argumentos mudam. Por sua vez, funções em que $g(x+c) = g(x)+c$ são chamadas equivariantes, ou seja, essas funções têm a propriedade de alterar seus valores de forma pré-determinada quando ocorre variação em x .

Na teoria geral de ações de grupos, o grupo G atua em \mathfrak{X}^n e em Θ , sem que essas ações estejam intrinsecamente relacionadas. No contexto estatístico essas duas ações devem estar relacionadas. Primeiramente, assumamos que uma característica de interesse dos elementos da população possa ser representada por uma variável aleatória X , cuja densidade é $f_X(x; \theta) = f(x; \theta)$. Assumamos também que valores (x_1, \dots, x_n) de uma amostra aleatória (X_1, \dots, X_n) de $f(x; \theta)$ tenham sido observados. Nesse caso, o grupo G atua em \mathfrak{X}^n da forma $g \cdot (X_1, \dots, X_n) = (g \cdot X_1, \dots, g \cdot X_n)$. Pode-se então definir para essa transformação uma nova variável $Y = g \cdot X$. Conseqüentemente, $(Y_1, \dots, Y_n) = (g \cdot X_1, \dots, g \cdot X_n)$. Se X tem distribuição $f(x; \theta)$, para se obter um contexto invariante pelas ações do grupo G é necessário exigir que Y também tenha distribuição na mesma família, isto é, deve existir um $\tilde{\theta} \in \Theta$ tal que Y tenha distribuição $f(y; \tilde{\theta})$. Exigindo-se um pouco mais que isso, é necessário supor que $\tilde{\theta}$ seja também definido por uma ação de G em Θ , ou seja, $\tilde{\theta} = g \cdot \theta$. Dessa forma, essas duas ações estão

relacionadas por:

$$X \sim f(x; \theta) \Leftrightarrow Y = g \cdot X \sim f(y; g \cdot \theta).$$

Esse resultado indica que as ações de um grupo acontecem em primeiro lugar no espaço amostral e, em seguida, no espaço de parâmetros. Além disso, dada a amostra original (X_1, \dots, X_n) e a estrutura matemática associada a essa amostra (espaço paramétrico e modelo subjacente ou conjunto de funções densidade de probabilidade fdp's), o último resultado indica que se um grupo G age sobre a amostra (X_1, \dots, X_n) modificando sua escala de medição, então o problema transformado deve ter a mesma estrutura matemática que o problema original, ou seja, deve estar definido no mesmo espaço paramétrico, para a mesma família paramétrica. Quando isso ocorre, a distribuição subjacente ou a família de distribuições do problema original é dita invariante pela transformação. Essa exigência é sintetizada na definição abaixo.

Definição 2.3.5 (Invariância do modelo): Seja $\mathcal{F} = \{f(x; \theta), \theta \in \Theta\}$ uma família de densidades para X . A família \mathcal{F} é dita *invariante* em relação a um grupo G se, para cada $\theta \in \Theta$ e $g \in G$, existe um único $\tilde{\theta} \in \Theta$ de modo que $Y = g \cdot X$ tenha distribuição $f(y; \tilde{\theta})$ na mesma família de X . O parâmetro $\tilde{\theta}$, que é determinado unicamente por θ e g , é obtido por $g \cdot \theta$.

Dado um evento $B \subset \mathcal{X}^n$, em termos de probabilidade de eventos tem-se, a partir da definição 2.3.5, a seguinte igualdade:

$$P_\theta [Y \in B] = P_{g \cdot \theta} [X \in B],$$

em que o subscrito do lado esquerdo (θ) indica a distribuição de X e não a de Y . Segue ainda, para uma função ψ cuja esperança seja definida, que:

$$E_{\theta} [\psi (Y)] = E_{g,\theta} [\psi (X)].$$

Um exemplo de um modelo probabilístico invariante é apresentado a seguir.

Exemplo 2.3.4: Seja X uma variável aleatória com distribuição Gama(α, β). A fdp de X é dada por:

$$f_X (x; \alpha, \beta) = \frac{1}{\Gamma (\alpha) \beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}},$$

em que $\Gamma (\cdot)$ é a função gama, com $0 < x < \infty$, $\alpha > 0$ e $\beta > 0$. Considere as ações do grupo $G = (\mathbb{R}_+, \cdot)$ de tal forma que $Y = h (X) = kX$, sendo $k > 0$. Como $y = h (x)$, então $h^{-1} (y) = \frac{y}{k}$ e $\frac{d}{dy} h^{-1} (y) = \frac{1}{k}$. Aplicando o teorema da transformação de variáveis, temos que a fdp de Y é:

$$\begin{aligned} f_Y (y) &= \left| \frac{dh^{-1} (y)}{dy} \right| f_X (h^{-1} (y)) \\ &= \left| \frac{1}{k} \right| f_X \left(\frac{y}{k} \right) \\ &= \frac{1}{k} \frac{1}{\Gamma (\alpha) \beta^{\alpha}} \left(\frac{y}{k} \right)^{\alpha-1} e^{-\frac{y}{k\beta}} \\ &= \frac{1}{\Gamma (\alpha) k^{\alpha} \beta^{\alpha}} y^{\alpha-1} e^{-\frac{y}{k\beta}} \\ &= \frac{1}{\Gamma (\alpha) (k\beta)^{\alpha}} y^{\alpha-1} e^{-\frac{y}{k\beta}}. \end{aligned}$$

Logo, a variável $Y \sim \text{Gama} (\alpha, k\beta)$. Portanto, a ação de G nos dados, dada por $k \cdot X = kX$, induz uma ação de G no espaço paramétrico, dada por $k \cdot (\alpha, \beta) = (\alpha, k\beta)$. Como a variável transformada Y preserva a mesma estrutura

matemática que a variável original X (mesma família paramétrica e mesmo espaço paramétrico), essa família é invariante pelas ações do grupo G .

Se um modelo é invariante sob as ações de um grupo G , pode-se exigir que a função perda também seja invariante em relação às ações desse grupo.

Definição 2.3.6 (Função perda invariante): Uma função perda $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$ é dita *invariante* sob as ações de um grupo G se:

$$L(\delta(x), \theta) = L(g \cdot (\delta(x)), g \cdot \theta),$$

para todo $\delta(x) \in \mathcal{A}$, $\theta \in \Theta$ e $g \in G$.

A vantagem de se trabalhar no contexto invariante é que a função risco torna-se mais simples. Se δ é um estimador equivariante, então:

$$\begin{aligned} \mathcal{R}(\delta, \theta) &= E_{\theta}[L(\delta(X), \theta)] \\ &= \int L(\delta(x), \theta) f_X(x; \theta) dx \\ &= \int L(g \cdot (\delta(x)), g \cdot \theta) f_X(x; \theta) dx && \text{(Invariância da função perda } L) \\ &= \int L(\delta(g \cdot x), g \cdot \theta) f_X(x; \theta) dx && \text{(Equivariância do estimador } \delta) \\ &= \int L(\delta(y), g \cdot \theta) f_Y(y; g \cdot \theta) dy && \text{(Invariância do modelo } f) \\ &= E_{g \cdot \theta}[L(\delta(Y), g \cdot \theta)] \\ &= \mathcal{R}(\delta, g \cdot \theta). \end{aligned}$$

Portanto, ao se fixar um estimador equivariante δ , a função risco torna-se constante nas órbitas da ação de G em Θ . Desse modo, se G atua transitivamente em Θ , o risco dos estimadores equivariantes será constante. Dizer que a função risco é constante significa que essa função não depende do parâmetro θ . Esse resultado ficará mais claro nas seções relacionadas a equivariância por locação e

por escala.

Definição 2.3.7 (Estimador equivariante de risco mínimo): Um estimador $\delta : \mathcal{X}^n \rightarrow \mathbb{R}$ de $\omega(\theta)$ é definido como *equivariante de risco mínimo* (MRE, do inglês *minimum risk equivariant*) se:

- (i) δ é equivariante;
- (ii) $\mathcal{R}(\delta) \leq \mathcal{R}(\delta')$, dado qualquer outro estimador equivariante δ' de $\omega(\theta)$.

A importante consequência que surge do fato da função risco ser constante é que, uma vez que o risco de qualquer estimador equivariante independe de θ , o problema de se obter um estimador com menor risco uniforme na classe dos estimadores equivariantes torna-se relativamente mais simples: selecionar entre os estimadores com função risco constante aquele que possua menor risco.

2.4 Equivariância para parâmetros de locação

Nesta subseção apresentam-se algumas definições e resultados requeridos para a determinação do estimador equivariante de menor risco por locação. Inicialmente, antes da abordagem da equivariância por locação, define-se parâmetro de locação.

Definição 2.4.1 (Parâmetro de locação): Seja $\{f(\cdot; \theta), \theta \in \Theta\}$ uma família de densidades indexadas por um parâmetro θ , em que Θ é um subconjunto da reta real. O parâmetro θ é definido como *parâmetro de locação* se, a densidade $f(x; \theta)$ pode ser escrita como uma função de $x - \theta$, isto é, $f(x; \theta) = h(x - \theta)$ para alguma função $h(\cdot)$.

Proposição 2.4.1: θ é um parâmetro de locação se, e somente se, a distribuição de $X - \theta$ não depende de θ .

Prova:

Como θ é um parâmetro de locação, por hipótese $f_X(x; \theta) = h(x - \theta)$. Se $Y = g(X) = X - \theta$, então $g^{-1}(y) = y + \theta$ e $\left| \frac{d}{dy} g^{-1}(y) \right| = 1$. Utilizando o teorema da transformação de variáveis segue que a fdp de $Y = X - \theta$ é:

$$f_Y(y; \theta) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) = 1 f_X(y + \theta) = h(y + \theta - \theta) = h(y).$$

Portanto, a distribuição de $Y = X - \theta$ não depende de θ .

Reciprocamente, seja $Y = X - \theta$ com distribuição dada por uma função $h(y)$. Como $X = g(Y) = Y + \theta$, logo $g^{-1}(x) = x - \theta$ e $\left| \frac{d}{dx} g^{-1}(x) \right| = 1$. Novamente, utilizando o teorema da transformação de variáveis segue que a fdp de $X = Y + \theta$ é:

$$f_X(x; \theta) = \left| \frac{d}{dx} g^{-1}(x) \right| f_Y(g^{-1}(x)) = 1 f_Y(x - \theta) = h(x - \theta).$$

Portanto, a densidade $f_X(x; \theta)$ pode ser escrita como uma função de $x - \theta$, o que demonstra que θ é um parâmetro de locação. ■

Observe que se θ é um parâmetro de locação para a família de densidades $\{f(\cdot; \theta), \theta \in \Theta\}$, a função $h(\cdot)$ da definição é uma fdp dada por $h(\cdot) = f(\cdot; 0)$.

Exemplo 2.4.1: Se $f_X(x; \theta) = \phi_{\theta,1}(x)$, θ é um parâmetro de locação, pois:

$$\phi_{\theta,1}(x) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - \theta)^2 \right] = \phi_{0,1}(x - \theta),$$

isto é, se X é normalmente distribuída com média θ e variância 1, então $X - \theta$ possui distribuição Normal padrão, e portanto, independente de θ .

Exemplo 2.4.2: Se $f_X(x; \theta) = I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x)$, então θ é um parâmetro de localização, pois:

$$f_X(x; \theta) = I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(x) = I_{(-\frac{1}{2}, \frac{1}{2})}(x - \theta),$$

é uma função de $x - \theta$.

2.4.1 Ação de grupos para parâmetros de localização

Um grupo que está naturalmente relacionado a parâmetros de localização é o grupo $G = (\mathbb{R}, +)$. Esse grupo atua em \mathbb{R}^n da forma $c \cdot (X_1, \dots, X_n) = (X_1 + c, \dots, X_n + c)$. A equivariância em relação a essa ação é justificada da seguinte forma: se observações (x_1, \dots, x_n) representam medições de algum tipo, e uma constante c é adicionada em cada uma dessas medições, então um estimador avaliado nas medições $(x_1 + c, \dots, x_n + c)$ deve ajustar o valor estimado $\delta(x_1, \dots, x_n)$ pela adição da mesma constante c , isto é, $\delta(x_1 + c, \dots, x_n + c) = \delta(x_1, \dots, x_n) + c$. Esse fato ocorre com o mais básico de todos estimadores que é a média amostral.

Definição 2.4.2 (Equivariância por localização): Um estimador δ é definido como *equivariante por localização* se:

$$\delta(X_1 + c, \dots, X_n + c) = \delta(X_1, \dots, X_n) + c,$$

para todo (X_1, \dots, X_n) e $(c, \dots, c) \in \mathbb{R}^n$.

Exemplo 2.4.3: Um primeiro exemplo de um estimador equivariante por localização é a média amostral. De fato,

$$\delta(\mathbf{X} + \mathbf{c}) = \frac{\sum_{i=1}^n (X_i + c)}{n} = \frac{\sum_{i=1}^n X_i}{n} + \frac{nc}{n} = \delta(\mathbf{X}) + c.$$

Exemplo 2.4.4: Considerando as estatísticas de ordem do mínimo e do máximo,

$Y_1 = \min[X_1, \dots, X_n]$ e $Y_n = \max[X_1, \dots, X_n]$, segue que a média desses valores $\left(\frac{Y_1 + Y_n}{2}\right)$ é equivariante por locação, pois:

$$\begin{aligned} \delta(\mathbf{X} + \mathbf{c}) &= \frac{\min[X_1 + c, \dots, X_n + c] + \max[X_1 + c, \dots, X_n + c]}{2} \\ &= \frac{\min[X_1, \dots, X_n] + c + \max[X_1, \dots, X_n] + c}{2} \\ &= \frac{\min[X_1, \dots, X_n] + \max[X_1, \dots, X_n]}{2} + \frac{2c}{2} \\ &= \delta(\mathbf{X}) + c. \end{aligned}$$

Por outro lado, existem estimadores que não são equivariantes por locação, como por exemplo, S^2 e $Y_n - Y_1$.

Como mencionado anteriormente, o grupo $G = (\mathbb{R}, +)$ é um grupo natural a ser relacionado a parâmetros de locação. Esse grupo atua, respectivamente, no espaço paramétrico Θ , no espaço amostral \mathfrak{X}^n e no espaço dos estimadores D da forma:

1. $G \times \Theta \rightarrow \Theta$, $(c, \theta) \mapsto c \cdot \theta = \theta + c$.
2. $G \times \mathfrak{X}^n \rightarrow \mathfrak{X}^n$, $(c, \mathbf{x}) \mapsto c \cdot \mathbf{x} = (x_1 + c, \dots, x_n + c)$.
3. $G \times D \rightarrow D$, $(c, \delta) \mapsto (c \cdot \delta)(\mathbf{x}) = \delta(c \cdot \mathbf{x}) = \delta(\mathbf{x} + \mathbf{c}) = \delta(\mathbf{x}) + c$.

Por sua vez, uma função perda L é invariante por locação por essas ações se:

$$\begin{aligned}
L(\delta(x_1, \dots, x_n), \theta) &= L((c \cdot \delta)(x_1, \dots, x_n), c \cdot \theta) \\
&= L(\delta(c \cdot (x_1, \dots, x_n)), c \cdot \theta) \\
&= L(\delta(x_1 + c, \dots, x_n + c), \theta + c) \\
&= L(\delta(x_1, \dots, x_n) + c, \theta + c).
\end{aligned}$$

Um exemplo de uma função perda invariante por locação é a função perda em erro quadrático. De fato, se $L(\delta, \theta) = (\delta - \theta)^2$, então:

$$\begin{aligned}
L(\delta(x_1, \dots, x_n) + c, \theta + c) &= [(\delta(x_1, \dots, x_n) + c) - (\theta + c)]^2 \\
&= (\delta(x_1, \dots, x_n) + c - \theta - c)^2 \\
&= (\delta(x_1, \dots, x_n) - \theta)^2 \\
&= L(\delta(x_1, \dots, x_n), \theta).
\end{aligned}$$

O lema apresentado a seguir fornece um importante resultado na caracterização de todos os estimadores equivariantes por locação.

Lema 2.4.1: Seja δ_0 um estimador equivariante por locação. Uma condição necessária e suficiente para que δ seja equivariante por locação é:

$$\delta(\mathbf{x}) = \delta_0(\mathbf{x}) + u(\mathbf{x}), \quad (2.1)$$

em que $u : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função invariante pela ação do grupo $(\mathbb{R}, +)$, isto é, $u(\mathbf{x} + \mathbf{c}) = u(\mathbf{x})$, para todo $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{c} = (c, \dots, c) \in \mathbb{R}^n$.

Prova:

Se $\delta(\mathbf{x}) = \delta_0(\mathbf{x}) + u(\mathbf{x})$, em que δ_0 é um estimador equivariante conhecido e u é uma função invariante, então:

$$\delta(\mathbf{x} + \mathbf{c}) = \delta_0(\mathbf{x} + \mathbf{c}) + u(\mathbf{x} + \mathbf{c}) = \delta_0(\mathbf{x}) + c + u(\mathbf{x}) = \delta(\mathbf{x}) + c.$$

Portanto, δ é um estimador equivariante por locação.

Reciprocamente, assuma que δ_0 é um estimador equivariante conhecido e seja δ um estimador equivariante arbitrário. Defina: $u = \delta - \delta_0$. Então:

$$u(\mathbf{x} + \mathbf{c}) = \delta(\mathbf{x} + \mathbf{c}) - \delta_0(\mathbf{x} + \mathbf{c}) = \delta(\mathbf{x}) + c - (\delta_0(\mathbf{x}) + c) = \delta(\mathbf{x}) - \delta_0(\mathbf{x}) = u(\mathbf{x}).$$

Portanto, a função u é invariante pela ação do grupo $G = (\mathbb{R}, +)$.

■

Pode-se verificar que esse lema estabelece a conexão entre a equivariância e invariância de funções. Primeiramente, observa-se que um estimador equivariante δ é a combinação entre um estimador equivariante conhecido e uma função invariante u . Além disso, o resultado desse lema indica que a diferença entre dois estimadores equivariantes é uma função invariante de \mathbf{x} . Pelas implicações do lema é possível notar que a caracterização dos estimadores equivariantes também exige uma caracterização das funções invariantes u .

As funções invariantes podem ser expressas em termos de coordenadas mais simples, se ocorre o seguinte contexto: seja G um grupo que atua em um conjunto A . Suponha que exista um subconjunto $B \subset A$ tal que toda órbita da ação de G em A intercepta B em apenas um ponto. Dada uma função invariante $u : A \rightarrow \mathbb{R}$, essa função fica totalmente definida pelos valores que assume sobre os pontos do subconjunto B . Além disso, se u restrita a B for injetiva, os valores de

u distinguem as órbitas. Portanto, as funções invariantes podem ser expressas por coordenadas mais simples que dependem apenas do subconjunto B .

Para o caso específico da ação do grupo $(\mathbb{R}, +)$ em \mathbb{R}^n , $c \cdot \mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$, as órbitas são retas paralelas ao vetor $\vec{1} = (1, \dots, 1)$. Logo, se B é o hiperplano definido por $B = \{(x_1, \dots, x_{n-1}, 0), x_i \in \mathbb{R}\}$, todas as órbitas interceptam esse hiperplano em apenas um ponto.

Proposição 2.4.2: Se $u : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função invariante, então u pode ser expressa como função das diferenças $y_i = x_i - x_n$.

Prova:

Para expressar as funções invariantes em um sistema de coordenadas adequado deve-se observar que as órbitas são retas paralelas ao vetor $\vec{1} = (1, \dots, 1)$ (Figura 3). O hiperplano $x_n = 0$ é tal que toda órbita corta este plano em apenas um ponto. Dada uma órbita pelo ponto (x_1, \dots, x_n) é necessário obter o ponto $\mathbf{y} = (y_1, \dots, y_{n-1}, 0)$ do plano B correspondente a essa órbita.

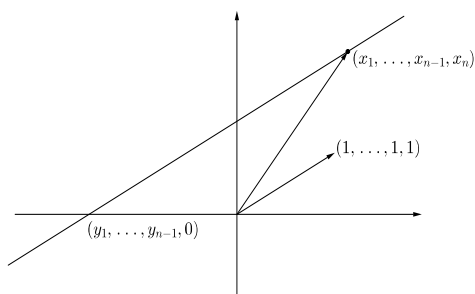


Figura 3 Representação da órbita que corta o plano no ponto (x_1, \dots, x_n) .

A órbita é a reta dada por $\{(y_1, \dots, y_{n-1}, 0) + t(1, \dots, 1); t \in \mathbb{R}\}$. Tem-se então que:

$$\left\{ \begin{array}{l} y_1 + t = x_1 \\ \vdots \\ y_{n-1} + t = x_{n-1} \\ t = x_n \end{array} \right. \Rightarrow \left\{ \begin{array}{l} y_1 = x_1 - x_n \\ \vdots \\ y_{n-1} = x_{n-1} - x_n \\ t = x_n \end{array} \right.$$

$$\mathbf{y} = (y_1 = x_1 - x_n, \dots, y_{n-1} = x_{n-1} - x_n, 0).$$

Dessa forma, toda função invariante u pode ser expressa em termos das variáveis y_i . De fato,

$$\begin{aligned} u(\mathbf{x} + \mathbf{c}) &= u(x_1 + c, \dots, x_{n-1} + c, x_n + c) \\ &= u(x_1 - x_n, \dots, x_{n-1} - x_n, x_n - x_n) \quad (\text{Tomar } c = -x_n) \\ &= u(x_1 - x_n, \dots, x_{n-1} - x_n, 0) \\ &= u(\mathbf{y}), \end{aligned}$$

o que conclui a demonstração. ■

Da hipótese de u ser invariante, isto é, $u(\mathbf{x} + \mathbf{c}) = u(\mathbf{x})$, segue ainda da proposição 2.4.2 que $u(\mathbf{x}) = u(\mathbf{y})$. A expressão dos estimadores equivariantes (2.1), em termos de um estimador equivariante particular mais uma função invariante, permite a obtenção de uma fórmula explícita para o estimador equivariante de menor risco. Esse objetivo pode ser alcançado pela minimização da função risco, que é dada por:

$$\mathcal{R}(\delta, \theta) = E_\theta [L(\delta(\mathbf{X}), \theta)] = \int \dots \int L(\delta(\mathbf{x}), \theta) f_{\mathbf{X}}(\mathbf{x}; \theta) d\mathbf{x}.$$

Com base na equivariância do estimador δ , e na invariância do modelo f e da função perda L , tem-se que:

$$\begin{aligned}
 \mathcal{R}(\delta, \theta) &= E_{\theta} [L(\delta(\mathbf{X}), \theta)] \\
 &= \int \dots \int L(\delta(\mathbf{x}), \theta) f_{\mathbf{X}}(\mathbf{x}; \theta) d\mathbf{x} \\
 &= \int \dots \int L(\delta(\mathbf{x}) + c, \theta + c) f_{\mathbf{X}}(\mathbf{x}; \theta) d\mathbf{x} \quad (\text{Invariância da função } L) \\
 &= \int \dots \int L(\delta(\mathbf{x} + \mathbf{c}), \theta + c) f_{\mathbf{X}}(\mathbf{x}; \theta) d\mathbf{x} \quad (\text{Equivariância de } \delta) \\
 &= \int \dots \int L(\delta(\mathbf{y}), \theta + c) f_{\mathbf{Y}}(\mathbf{y}; \theta + c) d\mathbf{y} \quad (\text{Invariância de } f) \\
 &= E_{\theta+c} [L(\delta(\mathbf{Y}), \theta + c)] \\
 &= \mathcal{R}(\delta, \theta + c).
 \end{aligned}$$

Como a ação de G no espaço paramétrico Θ é transitiva, tomando-se $c = -\theta$ segue que $\mathcal{R}(\delta, \theta) = \mathcal{R}(\delta, \theta - \theta) = \mathcal{R}(\delta, 0)$. Portanto, a função risco é constante em relação a θ e a sua minimização torna-se mais simples, uma vez que essa função depende somente do estimador δ e não mais do parâmetro θ . Dessa maneira,

$$\mathcal{R}(\delta, 0) = E_0 [L(\delta(\mathbf{X}), 0)] = \int \dots \int L(\delta(\mathbf{x}), 0) f_{\mathbf{X}}(\mathbf{x}; 0) d\mathbf{x}.$$

Uma notação mais concisa a ser utilizada para a função risco é:

$$\mathcal{R}(\delta) = \int \dots \int L(\delta(\mathbf{x}), 0) f_{\mathbf{X}}(\mathbf{x}; 0) d\mathbf{x}.$$

Fixando um estimador equivariante δ_0 e observando que qualquer estimador equivariante δ é da forma $\delta = \delta_0 + u$, o problema de minimização fica da forma:

$$\begin{aligned}\min_{\delta} \mathcal{R}(\delta) &= \min_{\delta} \int \dots \int L(\delta(\mathbf{x}), 0) f_{\mathbf{X}}(\mathbf{x}; 0) d\mathbf{x} \\ &= \min_u \int \dots \int L(\delta_0(\mathbf{x}) + u(\mathbf{x}), 0) f_{\mathbf{X}}(\mathbf{x}; 0) d\mathbf{x}.\end{aligned}$$

A ideia para realizar a minimização da função risco \mathcal{R} é utilizar as variáveis $y_i = x_i - x_n$, com $i = 1, \dots, n-1$. Como \mathbf{x} determina totalmente \mathbf{y} , a distribuição conjunta $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}; 0)$ é degenerada, e neste caso,

$$f_{\mathbf{X}}(\mathbf{x}; 0) = f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}; 0) = f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) f_{\mathbf{Y}}(\mathbf{y}; 0).$$

Além disso, como $u(\mathbf{x}) = u(\mathbf{y})$ então:

$$\begin{aligned}\min_{\delta} \mathcal{R}(\delta) &= \min_u \int \dots \int L(\delta_0(\mathbf{x}) + u(\mathbf{y}), 0) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) f_{\mathbf{Y}}(\mathbf{y}; 0) d\mathbf{x} d\mathbf{y} \\ &= \min_u \int \dots \int \left(\int L(\delta_0(\mathbf{x}) + u(\mathbf{y}), 0) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x} \right) f_{\mathbf{Y}}(\mathbf{y}; 0) d\mathbf{y} \\ &= \int \dots \int \left(\min_u \int L(\delta_0(\mathbf{x}) + u(\mathbf{y}), 0) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x} \right) f_{\mathbf{Y}}(\mathbf{y}; 0) d\mathbf{y},\end{aligned}$$

em que pode-se observar que $\int L(\delta_0(\mathbf{x}) + u(\mathbf{y}), 0) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x}$ é uma esperança condicional.

Para o caso em que L é a função perda em erro quadrático $L(\delta, \theta) = (\delta - \theta)^2$, a minimização da esperança condicional é relativamente simples. Desenvolvendo o quadrado, tem-se que $\int (\delta_0(\mathbf{x}) + u(\mathbf{y}))^2 f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x}$ é igual a:

$$\int (\delta_0(\mathbf{x}))^2 f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x} + 2u(\mathbf{y}) \int \delta_0(\mathbf{x}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x} + (u(\mathbf{y}))^2,$$

que corresponde a um polinômio em $u(\mathbf{y})$, que é minimizado quando:

$$u(\mathbf{y}) = - \int \delta_0(\mathbf{x}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}; 0) d\mathbf{x} = -E_0[\delta_0(\mathbf{X})|\mathbf{Y}].$$

Portanto, substituindo $u(\mathbf{y})$ em (2.1), segue que o estimador MRE em relação à função perda em erro quadrático é:

$$\delta(\mathbf{X}) = \delta_0(\mathbf{X}) - E_0[\delta_0(\mathbf{X})|\mathbf{Y}].$$

Para se obter uma forma explícita para o estimador MRE, considere o estimador equivariante $\delta_0(\mathbf{X}) = X_n$. Assim,

$$\delta(\mathbf{X}) = X_n - E_0[X_n|\mathbf{Y}]. \quad (2.2)$$

Para o cálculo da esperança condicional considere também a transformação de coordenadas:

$$(x_1, \dots, x_{n-1}, x_n) \rightarrow (y_1, \dots, y_{n-1}, y_n),$$

dada por:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} = \begin{bmatrix} +1 & 0 & \dots & 0 & -1 \\ 0 & +1 & \dots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & +1 & -1 \\ 0 & 0 & \dots & 0 & +1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix},$$

ou seja,

$$y_i = x_i - x_n; \quad y_n = x_n, \text{ com } i = 1, \dots, n - 1.$$

Disso decorre que,

$$x_i = y_i + y_n; \quad x_n = y_n, \text{ com } i = 1, \dots, n - 1. \quad (2.3)$$

A partir de (2.3) as derivadas parciais dos x 's são facilmente calculadas.

Pode-se verificar que o Jacobiano da transformação é:

$$J = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = 1.$$

Logo, pelo teorema da transformação de variáveis segue que a distribuição conjunta dos Y 's é:

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}; 0) &= |J| f_{X_1, \dots, X_{n-1}, X_n}(x_1(y_1, y_n), \dots, x_{n-1}(y_{n-1}, y_n), x_n(y_n); 0) \\ &= f_{X_1, \dots, X_{n-1}, X_n}(y_1 + y_n, \dots, y_{n-1} + y_n, y_n; 0). \end{aligned}$$

A densidade condicional de X_n dado $\mathbf{Y} = \mathbf{y} = (y_1, \dots, y_{n-1})$ é:

$$\begin{aligned}
f_{X_n|\mathbf{Y}}(x_n|\mathbf{y};0) &= f_{Y_n|Y_1,\dots,Y_{n-1}}(y_n|y_1,\dots,y_{n-1};0) \\
&= \frac{f_{Y_1,\dots,Y_{n-1},Y_n}(y_1,\dots,y_n;0)}{f_{Y_1,\dots,Y_{n-1}}(y_1,\dots,y_{n-1};0)} \\
&= \frac{f_{X_1,\dots,X_{n-1},X_n}(y_1+y_n,\dots,y_{n-1}+y_n,y_n;0)}{\int f_{X_1,\dots,X_{n-1},X_n}(y_1+t,\dots,y_{n-1}+t,t;0)dt},
\end{aligned}$$

em que trocou-se y_n por t para representar a variação da variável $Y_n = X_n$ na integral.

A esperança condicional de X_n dado que $\mathbf{Y} = \mathbf{y} = (y_1, \dots, y_{n-1})$ é:

$$\begin{aligned}
E_0[X_n|\mathbf{Y}] &= E_0[Y_n|\mathbf{Y}] = \int y_n f_{Y_n|Y_1,\dots,Y_{n-1}}(y_n|y_1,\dots,y_{n-1};0) dy_n \\
&= \frac{\int t f_{X_1,\dots,X_{n-1},X_n}(y_1+t,\dots,y_{n-1}+t,t;0) dt}{\int f_{X_1,\dots,X_{n-1},X_n}(y_1+t,\dots,y_{n-1}+t,t;0) dt}.
\end{aligned}$$

Como $y_i = x_i - x_n$ ($i = 1, \dots, n-1$), essa última igualdade pode ser expressa em termos dos x 's como:

$$E_0[X_n|\mathbf{Y}] = \frac{\int t f_{X_1,\dots,X_{n-1},X_n}(x_1-x_n+t,\dots,x_{n-1}-x_n+t,t;0) dt}{\int f_{X_1,\dots,X_{n-1},X_n}(x_1-x_n+t,\dots,x_{n-1}-x_n+t,t;0) dt}.$$

Uma observação neste ponto é importante. Como o vetor \mathbf{y} foi dado, isso implica necessariamente que x_n também foi observado. Assim, como x_n é o valor observado de X_n e t representa a variação da variável $Y_n = X_n$ na integral, a próxima mudança de variáveis justifica-se. Fazendo a mudança de variáveis $u = x_n - t$, tem-se que:

$$\begin{aligned} E_0 [X_n | \mathbf{Y}] &= \frac{\int (x_n - u) f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du}{\int f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du} \\ &= x_n - \frac{\int u f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du}{\int f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du}. \end{aligned}$$

Substituindo $E_0 [X_n | \mathbf{Y}]$ em (2.2) tem-se:

$$\delta (X_1, \dots, X_n) = \frac{\int u f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du}{\int f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) du}.$$

Este estimador é conhecido como estimador de Pitman para parâmetros de locação. Utilizando o fato do parâmetro ser de locação, resulta que:

$$f_{X_1, \dots, X_n} (x_1 - u, \dots, x_n - u; 0) = f_{X_1, \dots, X_n} (x_1, \dots, x_n; u).$$

Renomeando u por θ seque que:

$$\delta (X_1, \dots, X_n) = \frac{\int \theta f_{X_1, \dots, X_n} (x_1, \dots, x_n; \theta) d\theta}{\int f_{X_1, \dots, X_n} (x_1, \dots, x_n; \theta) d\theta}.$$

Para o caso de amostras iid, obtém-se a expressão do estimador de Pitman para um parâmetro de locação θ fornecida em Mood, Graybill e Boes (1974, p. 334).

Teorema 2.4.1: Seja (X_1, \dots, X_n) uma amostra iid de uma densidade $f(\cdot; \theta)$, em que θ é um parâmetro de locação e Θ é a reta real. O estimador

$$\delta (X_1, \dots, X_n) = \frac{\int \theta \prod_{i=1}^n f_{X_i} (X_i; \theta) d\theta}{\int \prod_{i=1}^n f_{X_i} (X_i; \theta) d\theta}, \quad (2.4)$$

é o estimador MRE de θ na classe dos estimadores equivariantes por locação.

Exemplo 2.4.5: Seja (X_1, \dots, X_n) uma amostra aleatória de uma distribuição Normal com média θ e variância 1. O estimador de Pitman de θ é:

$$\begin{aligned}
\frac{\int \theta \prod_{i=1}^n \phi_{\theta,1}(X_i) d\theta}{\int \prod_{i=1}^n \phi_{\theta,1}(X_i) d\theta} &= \frac{\int \theta (1/\sqrt{2\pi})^n \exp \left[-1/2 \sum (X_i - \theta)^2 \right] d\theta}{\int (1/\sqrt{2\pi})^n \exp \left[-1/2 \sum (X_i - \theta)^2 \right] d\theta} \\
&= \frac{\int \theta \exp \left[-1/2 \sum (X_i^2 - 2\theta X_i + \theta^2) \right] d\theta}{\int \exp \left[-1/2 \sum (X_i^2 - 2\theta X_i + \theta^2) \right] d\theta} \\
&= \frac{\int \theta \exp \left[\theta \sum X_i - (n/2) \theta^2 \right] d\theta}{\int \exp \left[\theta \sum X_i - (n/2) \theta^2 \right] d\theta} \\
&= \frac{\int \theta \exp \left[-(n/2) \theta^2 + \theta \sum X_i \right] d\theta}{\int \exp \left[-(n/2) \theta^2 + \theta \sum X_i \right] d\theta} \\
&= \frac{\int \theta \exp \left[-n/2 (\theta^2 - 2\theta/n \sum X_i) \right] d\theta}{\int \exp \left[-n/2 (\theta^2 - 2\theta/n \sum X_i) \right] d\theta} \\
&= \frac{\int \theta \exp \left[-n/2 (\theta^2 - 2\theta\bar{X} + \bar{X}^2 - \bar{X}^2) \right] d\theta}{\int \exp \left[-n/2 (\theta^2 - 2\theta\bar{X} + \bar{X}^2 - \bar{X}^2) \right] d\theta} \\
&= \frac{\exp(n/2\bar{X}^2) \int \theta \exp \left[-n/2 (\theta - \bar{X})^2 \right] d\theta}{\exp(n/2\bar{X}^2) \int \exp \left[-n/2 (\theta - \bar{X})^2 \right] d\theta} \\
&= \frac{\int \theta \exp \left[-n/2 (\theta - \bar{X})^2 \right] d\theta}{\int \exp \left[-n/2 (\theta - \bar{X})^2 \right] d\theta} \\
&= \frac{\int \theta \left[\frac{1}{\sqrt{2\pi}(1/\sqrt{n})} \right] \exp \left\{ -\frac{1}{2} \left[\frac{\theta - \bar{X}}{(1/\sqrt{n})} \right]^2 \right\} d\theta}{\int \left[\frac{1}{\sqrt{2\pi}(1/\sqrt{n})} \right] \exp \left\{ -\frac{1}{2} \left[\frac{\theta - \bar{X}}{(1/\sqrt{n})} \right]^2 \right\} d\theta} \\
&= \bar{X}.
\end{aligned}$$

Observa-se que o último denominador é justamente a integral de uma densidade Normal em θ com média \bar{X} e uma variância $\frac{1}{n}$ e, conseqüentemente, essa integral é unitária. O último numerador é a esperança da mesma densidade Normal e, portanto, o seu valor esperado é \bar{X} .

Neste exemplo, o estimador de Pitman de θ é idêntico ao estimador uniformemente não viesado de variância mínima (UMVUE, do inglês *uniformly minimum-variance unbiased estimator*) de θ , isto é, para o caso Normal, o melhor entre os estimadores equivariantes para a média é também o melhor entre os estimadores não viesados.

Exemplo 2.4.6: Seja (X_1, \dots, X_n) uma amostra aleatória de uma distribuição Uniforme definida sobre o intervalo $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. O estimador de Pitman de θ é:

$$\frac{\int \theta \prod_{i=1}^n I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(X_i) d\theta}{\int \prod_{i=1}^n I_{(\theta - \frac{1}{2}, \theta + \frac{1}{2})}(X_i) d\theta} = \frac{\int \theta \prod_{i=1}^n I_{(X_i - \frac{1}{2}, X_i + \frac{1}{2})}(\theta) d\theta}{\int \prod_{i=1}^n I_{(X_i - \frac{1}{2}, X_i + \frac{1}{2})}(\theta) d\theta} = \frac{\int \theta d\theta}{\int d\theta}.$$

A última igualdade justifica-se pela definição de função indicadora, pois como $\theta \in (X_i - \frac{1}{2}, X_i + \frac{1}{2})$ para todo i , segue que $I_{(X_i - \frac{1}{2}, X_i + \frac{1}{2})}(\theta) = 1$ para todo i . Conseqüentemente, $\prod_{i=1}^n I_{(X_i - \frac{1}{2}, X_i + \frac{1}{2})}(\theta) = 1$.

Considerando o mínimo e o máximo amostral, $Y_1 = \min[X_1, \dots, X_n]$ e $Y_n = \max[X_1, \dots, X_n]$, resulta do fato de $Y_1 \geq \theta - \frac{1}{2}$ e $Y_n \leq \theta + \frac{1}{2}$ que o intervalo de integração para o parâmetro θ é $(Y_n - \frac{1}{2}, Y_1 + \frac{1}{2})$.

Logo,

$$\begin{aligned}
\frac{\int \theta \prod_{i=1}^n I_{(\theta-\frac{1}{2}, \theta+\frac{1}{2})}(X_i) d\theta}{\int \prod_{i=1}^n I_{(\theta-\frac{1}{2}, \theta+\frac{1}{2})}(X_i) d\theta} &= \frac{\int_{Y_n-1/2}^{Y_1+1/2} \theta d\theta}{\int_{Y_n-1/2}^{Y_1+1/2} d\theta} \\
&= \frac{(1/2) \theta^2 \Big|_{Y_n-1/2}^{Y_1+1/2}}{\theta \Big|_{Y_n-1/2}^{Y_1+1/2}} \\
&= \frac{1}{2} \frac{(Y_1 + 1/2)^2 - (Y_n - 1/2)^2}{(Y_1 + 1/2) - (Y_n - 1/2)} \\
&= \frac{1}{2} \frac{[(Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2})] [(Y_1 + \frac{1}{2}) + (Y_n - \frac{1}{2})]}{(Y_1 + \frac{1}{2}) - (Y_n - \frac{1}{2})} \\
&= \frac{Y_1 + 1/2 + Y_n - 1/2}{2} \\
&= \frac{Y_1 + Y_n}{2}.
\end{aligned}$$

Exemplo 2.4.7: Seja (X_1, \dots, X_n) uma amostra aleatória de uma distribuição Gumbel de probabilidade, que possui fdp dada por:

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left[-\frac{x - \mu}{\sigma} - \exp \left(-\frac{x - \mu}{\sigma} \right) \right],$$

em que X é a variável aleatória associada a valores máximos de um período, μ é denominado parâmetro de locação e σ é o parâmetro de escala, sendo essa variável definida em $-\infty < x < +\infty$, com $-\infty < \mu < +\infty$ e $\sigma > 0$.

O estimador de Pitman para μ é dado por:

$$\begin{aligned}
\delta(X_1, \dots, X_n) &= \frac{\int \mu \prod_{i=1}^n f(X_i; \mu, \sigma)}{\int \prod_{i=1}^n f(X_i; \mu, \sigma)} \\
&= \frac{\int \mu \prod_{i=1}^n \frac{1}{\sigma} \exp\left[-\frac{X_i - \mu}{\sigma} - \exp\left(-\frac{X_i - \mu}{\sigma}\right)\right] d\mu}{\int \prod_{i=1}^n \frac{1}{\sigma} \exp\left[-\frac{X_i - \mu}{\sigma} - \exp\left(-\frac{X_i - \mu}{\sigma}\right)\right] d\mu} \\
&= \frac{\int \mu \left(\frac{1}{\sigma}\right)^n \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma} - \exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu}{\int \left(\frac{1}{\sigma}\right)^n \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma} - \exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu} \\
&= \frac{\int \mu \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right) + \sum_{i=1}^n \left(-\exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu}{\int \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right) + \sum_{i=1}^n \left(-\exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu} \\
&= \frac{\int \mu \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right)\right] \exp\left[\sum_{i=1}^n \left(-\exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu}{\int \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right)\right] \exp\left[\sum_{i=1}^n \left(-\exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] d\mu}.
\end{aligned}$$

Multiplicando $\left\{ \exp\left[\sum_{i=1}^n \left(-\exp\left(-\frac{X_i - \mu}{\sigma}\right)\right)\right] \right\}^{-1}$ no numerador e no denominador, vem que:

$$\delta(X_1, \dots, X_n) = \frac{\int \mu \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right)\right] d\mu}{\int \exp\left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right)\right] d\mu}.$$

Desenvolvendo o somatório $\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma}\right)$ tem-se que:

$$\begin{aligned}
\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma} \right) &= -\frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu) \\
&= -\frac{1}{\sigma} \left(\sum_{i=1}^n X_i - n\mu \right) \\
&= -\frac{n}{\sigma} (\bar{X} - \mu) \\
&= \frac{\mu - \bar{X}}{\sigma/n}.
\end{aligned}$$

Logo,

$$\begin{aligned}
\delta(X_1, \dots, X_n) &= \frac{\int \mu \exp \left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma} \right) \right] d\mu}{\int \exp \left[\sum_{i=1}^n \left(-\frac{X_i - \mu}{\sigma} \right) \right] d\mu} \\
&= \frac{\int \mu \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] d\mu}{\int \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] d\mu} \\
&= \frac{\int \mu \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \exp \left[-\exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu}{\int \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \exp \left[-\exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu} \\
&= \frac{\int \mu \frac{1}{\sigma/n} \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \exp \left[-\exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu}{\int \frac{1}{\sigma/n} \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \exp \left[-\exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu} \\
&= \frac{\int \mu \frac{1}{\sigma/n} \exp \left[\frac{\mu - \bar{X}}{\sigma/n} - \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu}{\int \frac{1}{\sigma/n} \exp \left[\frac{\mu - \bar{X}}{\sigma/n} - \exp \left[\frac{\mu - \bar{X}}{\sigma/n} \right] \right] d\mu}.
\end{aligned}$$

O último denominador é unitário, pois corresponde a integral de uma distribuição Gumbel de mínimos (NAGHETTINI; PINTO, 2007, p. 188), com parâmetro de locação igual a \bar{X} e parâmetro de escala igual a σ/n . O numerador, por sua vez, é a esperança da mesma densidade Gumbel de mínimos. Portanto, o estimador de Pitman do parâmetro de locação μ é:

$$\delta(X_1, \dots, X_n) = \bar{X} - 0,5772 \frac{\sigma}{n}. \quad (2.5)$$

Pode-se observar que o estimador de Pitman é muito semelhante ao estimador obtido pelo método dos momentos para o parâmetro μ da distribuição Gumbel $(\bar{X} - 0,5772\sigma)$. Além disso, pode-se verificar também que, a medida que o tamanho amostral $n \rightarrow \infty$, o estimador de Pitman obtido para μ tende a média amostral.

Exemplo 2.4.8: Considere uma amostra aleatória (X_1, \dots, X_n) de uma distribuição Exponencial de dois parâmetros. A fdp da variável aleatória X é dada por:

$$f_X(x; \theta, \tau) = \frac{1}{\tau} \exp\left(-\frac{x - \theta}{\tau}\right),$$

em que θ e τ são respectivamente os parâmetros de locação e de escala.

Uma vez que $X_i - \theta > 0$, vem que $X_i > \theta$. Em particular, segue para $Y_1 = \min[X_1, \dots, X_n]$ que $Y_1 > \theta$. Dessa maneira, o intervalo de integração de θ é $(-\infty, Y_1)$. Assim, o estimador de Pitman de θ é dado por:

$$\begin{aligned}
\delta(X_1, \dots, X_n) &= \frac{\int_{-\infty}^{Y_1} \theta \prod_{i=1}^n \frac{1}{\tau} \exp\left(-\frac{X_i - \theta}{\tau}\right) d\theta}{\int_{-\infty}^{Y_1} \prod_{i=1}^n \frac{1}{\tau} \exp\left(-\frac{X_i - \theta}{\tau}\right) d\theta} \\
&= \frac{\int_{-\infty}^{Y_1} \theta \left(\frac{1}{\tau}\right)^n \exp\left(\sum_{i=1}^n -\frac{X_i - \theta}{\tau}\right) d\theta}{\int_{-\infty}^{Y_1} \left(\frac{1}{\tau}\right)^n \exp\left(\sum_{i=1}^n -\frac{X_i - \theta}{\tau}\right) d\theta} \\
&= \frac{\int_{-\infty}^{Y_1} \theta \exp\left(-\sum_{i=1}^n \frac{X_i}{\tau}\right) \exp\left(\sum_{i=1}^n \frac{\theta}{\tau}\right) d\theta}{\int_{-\infty}^{Y_1} \exp\left(-\sum_{i=1}^n \frac{X_i}{\tau}\right) \exp\left(\sum_{i=1}^n \frac{\theta}{\tau}\right) d\theta} \\
&= \frac{\int_{-\infty}^{Y_1} \theta \exp\left(\frac{n\theta}{\tau}\right) d\theta}{\int_{-\infty}^{Y_1} \exp\left(\frac{n\theta}{\tau}\right) d\theta}.
\end{aligned}$$

A integral no numerador pode ser resolvida utilizando-se integração por partes. Tomando-se $u = \theta$ e $dv = \exp\left(\frac{n\theta}{\tau}\right) d\theta$, resulta que $du = d\theta$ e $v = \frac{\tau}{n} \exp\left(\frac{n\theta}{\tau}\right)$. Assim:

$$\begin{aligned}
\int_{-\infty}^{Y_1} \theta \exp\left(\frac{n\theta}{\tau}\right) d\theta &= \frac{\tau\theta}{n} \exp\left(\frac{n\theta}{\tau}\right) \Big|_{-\infty}^{Y_1} - \int_{-\infty}^{Y_1} \frac{\tau}{n} \exp\left(\frac{n\theta}{\tau}\right) d\theta \\
&= \frac{\tau Y_1}{n} \exp\left(\frac{nY_1}{\tau}\right) - \frac{\tau^2}{n^2} \exp\left(\frac{nY_1}{\tau}\right).
\end{aligned}$$

Por sua vez, o denominador é:

$$\int_{-\infty}^{Y_1} \exp\left(\frac{n\theta}{\tau}\right) d\theta = \frac{\tau}{n} \exp\left(\frac{n\theta}{\tau}\right) \Big|_{-\infty}^{Y_1} = \frac{\tau}{n} \exp\left(\frac{nY_1}{\tau}\right).$$

Portanto, o estimador de Pitman de θ é dado por:

$$\begin{aligned}\delta(X_1, \dots, X_n) &= \frac{\frac{\tau Y_1}{n} \exp\left(\frac{n Y_1}{\tau}\right) - \frac{\tau^2}{n^2} \exp\left(\frac{n Y_1}{\tau}\right)}{\frac{\tau}{n} \exp\left(\frac{n Y_1}{\tau}\right)} \\ &= \frac{\frac{\tau}{n} \exp\left(\frac{n Y_1}{\tau}\right) [Y_1 - \frac{\tau}{n}]}{\frac{\tau}{n} \exp\left(\frac{n Y_1}{\tau}\right)} \\ &= Y_1 - \frac{\tau}{n}.\end{aligned}$$

Nos exemplos 2.4.7 e 2.4.8 verificou-se que os estimadores MRE para os parâmetros de locação das distribuições Gumbel e Exponencial dependem de seus respectivos parâmetros de escala. Em ambas situações, não existe o estimador MRE para o grupo $G = (\mathbb{R}, +)$ quando o parâmetro de escala é desconhecido. Para esses casos torna-se também necessária a estimação do parâmetro de escala para garantir a existência do estimador MRE de locação.

O estimador de Pitman apresentado em (2.4) é, por construção, equivariante em relação as ações do grupo $G = (\mathbb{R}, +)$. Nesse sentido, no próximo exemplo será demonstrado diretamente que o estimador de Pitman é equivariante por locação.

Exemplo 2.4.9: Seja (X_1, \dots, X_n) uma amostra aleatória de uma fdp $f_X(x; \theta)$.

Se θ é um parâmetro de locação, segue por definição que:

$$f(x + c; \theta) = h(x + c - \theta) = h(x - (\theta - c)) = f(x; \theta - c).$$

Desse modo,

$$\delta(X_1 + c, \dots, X_n + c) = \frac{\int \theta \prod_{i=1}^n f(X_i + c; \theta) d\theta}{\int \prod_{i=1}^n f(X_i + c; \theta) d\theta} = \frac{\int \theta \prod_{i=1}^n f(X_i; \theta - c) d\theta}{\int \prod_{i=1}^n f(X_i; \theta - c) d\theta}.$$

Tomando-se $\theta - c = \theta^*$, segue que $d\theta = d\theta^*$ para θ^* variando de $-\infty$ para ∞ . Logo,

$$\begin{aligned} \delta(X_1 + c, \dots, X_n + c) &= \frac{\int (\theta^* + c) \prod_{i=1}^n f(X_i; \theta^*) d\theta^*}{\int \prod_{i=1}^n f(X_i; \theta^*) d\theta^*} \\ &= \frac{\int \theta^* \prod_{i=1}^n f(X_i; \theta^*) d\theta^*}{\int \prod_{i=1}^n f(X_i; \theta^*) d\theta^*} + c \frac{\int \prod_{i=1}^n f(X_i; \theta^*) d\theta^*}{\int \prod_{i=1}^n f(X_i; \theta^*) d\theta^*} \\ &= \delta(X_1, \dots, X_n) + c, \end{aligned}$$

o que prova que o estimador de Pitman é equivariante por locação.

Proposição 2.4.3: O estimador de Pitman para parâmetros de locação é uma função de estatísticas suficientes.

Prova:

Se $S_1 = s_1(X_1, \dots, X_n), \dots, S_k = s_k(X_1, \dots, X_n)$ é um conjunto de estatísticas suficientes, então pelo critério de fatoração (APÊNDICE C):

$$\prod_{i=1}^n f(x_i; \theta) = g(s_1, \dots, s_n; \theta) h(x_1, \dots, x_n).$$

Então, o estimador de Pitman pode ser escrito como:

$$\begin{aligned} \frac{\int \theta \prod_{i=1}^n f(X_i; \theta) d\theta}{\int \prod_{i=1}^n f(X_i; \theta) d\theta} &= \frac{\int \theta g(S_1, \dots, S_k; \theta) h(X_1, \dots, X_n) d\theta}{\int g(S_1, \dots, S_k; \theta) h(X_1, \dots, X_n) d\theta} \\ &= \frac{\int \theta g(S_1, \dots, S_k; \theta) d\theta}{\int g(S_1, \dots, S_k; \theta) d\theta}. \end{aligned}$$

Portanto, o estimador de Pitman para parâmetros de locação é uma função de estatísticas suficientes.

■

2.4.2 Estimação equivariante em modelos de regressão linear múltipla

Considere o modelo de regressão linear:

$$y = \sum_{i=1}^p x_i \beta_i + \varepsilon,$$

em que $\varepsilon \sim N(0, \sigma^2)$. Para n observações:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1},$$

sendo que $\boldsymbol{\mu} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$.

O espaço dos possíveis vetores de médias $\boldsymbol{\mu}$ é o subespaço de \mathbb{R}^n , definido pela imagem da transformação linear \mathbf{X} . Esse subespaço será denominado V . O espaço de dados \mathfrak{X}^n é o próprio \mathbb{R}^n . O teorema de Gauss-Markov (APÊNDICE C) é um resultado que indica que o melhor estimador linear não viesado (BLUE, do inglês *best linear unbiased estimator*) para a estimação do vetor de parâmetros $\boldsymbol{\beta}$ é:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y},$$

e, portanto, $\widehat{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$. Observe que a matriz $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ é simplesmente o projetor ortogonal de $\mathbf{y} \in \mathbb{R}^n$ no subespaço V e atua como a identidade em V , isto é, $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{v} = \mathbf{v}$, se $\mathbf{v} \in V$.

Considere o problema de se estimar uma combinação linear das compo-

mentos de $\boldsymbol{\mu}$, $\sum_{i=1}^n \lambda_i \mu_i$. Um estimador para a combinação linear das componentes de $\boldsymbol{\mu}$ é:

$$\delta(\mathbf{Y}) = \sum_{i=1}^n \lambda_i \hat{\mu}_i = \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \right)_i.$$

Conseqüentemente, a esperança desse estimador é:

$$\begin{aligned} E[\delta(\mathbf{Y})] &= \sum_{i=1}^n \lambda_i E[\hat{\mu}_i] = \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' E[\mathbf{Y}] \right)_i \\ &= \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\mu} \right)_i \\ &= \sum_{i=1}^n \lambda_i \mu_i, \end{aligned}$$

pois $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\mu} = \boldsymbol{\mu}$, para todo $\boldsymbol{\mu} \in V$. Pelo corolário do teorema de Gauss-Markov (APÊNDICE C), esse estimador é um BLUE.

Vejamos agora o modelo de regressão linear múltipla no contexto da estimação equivariante. O subespaço V munido da operação de soma de vetores é um grupo $G = (V, +)$. Esse grupo atua em si próprio da forma:

$$\begin{aligned} G \times \mathbf{V} &\rightarrow \mathbf{V} \\ (\mathbf{b}, \mathbf{v}) &\mapsto \mathbf{v} + \mathbf{b}, \end{aligned}$$

em que $\mathbf{b} = (b_1, \dots, b_n)$. Essa ação é transitiva.

O grupo G atua no espaço de dados \mathbb{R}^n de forma análoga:

$$\begin{aligned} G \times \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ (\mathbf{b}, \mathbf{Y}) &\mapsto \mathbf{Y} + \mathbf{b} \end{aligned}$$

Tem-se também a ação do grupo G em \mathbb{R} da forma:

$$G \times \mathbb{R} \rightarrow \mathbb{R}$$

$$(\mathbf{b}, a) \mapsto a + \sum_{i=1}^n \lambda_i b_i.$$

Proposição 2.4.4: O estimador de mínimos quadrados $\delta'(\mathbf{Y}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ é equivariante pela ação do grupo $G = (V, +)$.

Prova:

$$(b \cdot \delta')(\mathbf{Y}) = \delta'(b \cdot \mathbf{Y}) = \delta'(\mathbf{Y} + \mathbf{b}) = \delta'(\mathbf{Y}) + \delta'(\mathbf{b}) = \delta'(\mathbf{Y}) + \mathbf{b}.$$

■

Proposição 2.4.5: O estimador $\delta(\mathbf{Y}) = \sum_{i=1}^n \lambda_i \hat{\mu}_i = \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \right)_i$ é equivariante em relação a ação do grupo $G = (V, +)$.

Prova:

$$\begin{aligned} (b \cdot \delta)(\mathbf{Y}) &= \delta(b \cdot \mathbf{Y}) \\ &= \delta(\mathbf{Y} + \mathbf{b}) \\ &= \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} + \mathbf{b}) \right)_i \\ &= \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{b} \right)_i \\ &= \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} + \mathbf{b} \right)_i \\ &= \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \right)_i + \sum_{i=1}^n \lambda_i b_i. \end{aligned}$$

■

Segue do fato de δ ser BLUE que esse estimador é entre todos os estimadores lineares não viesados o de menor variância. Um resultado similar pode ser provado para a classe dos estimadores equivariantes. Vamos provar que o estimador $\delta(\mathbf{Y}) = \sum_{i=1}^n \lambda_i \hat{\mu}_i$ é de fato o estimador de menor risco entre todos os estimadores equivariantes de $\boldsymbol{\mu}$, considerando o modelo de regressão linear múltipla $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Se $\tilde{\delta}$ é um estimador equivariante, segue então do lema 2.4.1 que $\tilde{\delta} = \delta + u$. Neste caso, também é necessária a caracterização da função invariante u . Este resultado é apresentado na próxima proposição.

Proposição 2.4.6: Uma função $u : \mathbb{R}^n \rightarrow \mathbb{R}$ é invariante pela ação do grupo $G = (V, +)$ se, e somente se, u depende de $\mathbf{Y} \in \mathbb{R}^n$ através da projeção ortogonal de \mathbb{R}^n no complemento ortogonal de V , isto é,

$$u(\mathbf{Y}) = u\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y}\right).$$

Prova:

Seja u uma função invariante. Consequentemente:

$$\begin{aligned} u(\mathbf{Y}) &= u\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) \\ &= u\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y} + \hat{\boldsymbol{\mu}}\right) \\ &= u\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y}\right). \quad (u \text{ é invariante}) \end{aligned}$$

Reciprocamente, seja u uma função que dependa de \mathbf{Y} por meio do projetor ortogonal P de \mathbf{Y} no complemento ortogonal do subespaço V . Dessa forma,

$$P(\mathbf{Y}) = \left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \mathbf{Y}.$$

Se $u(\mathbf{Y}) = h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y}\right)$, então para um $\mathbf{b} \in V$ vem que:

$$\begin{aligned} u(\mathbf{Y} + \mathbf{b}) &= h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)(\mathbf{Y} + \mathbf{b})\right) \\ &= h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y} + \left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{b}\right) \\ &= h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y} + \mathbf{b} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{b}\right) \\ &= h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y} + \mathbf{0}\right) \\ &= h\left(\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{Y}\right) \\ &= u(\mathbf{Y}). \end{aligned}$$

Portanto, a função u é invariante pela ação do grupo $G = (V, +)$.

■

Para se obter o estimador MRE para $\boldsymbol{\mu}$ considerando o estimador $\tilde{\delta} = \delta + u$ é necessário minimizar a função risco desse estimador, que é dada por:

$$\mathcal{R}(\tilde{\delta}, \boldsymbol{\mu}) = E_{\boldsymbol{\mu}} \left[\left\| \tilde{\delta}(\mathbf{Y}) - \boldsymbol{\mu} \right\|^2 \right] = E_{\boldsymbol{\mu}} \left[\left\| \delta(\mathbf{Y}) + u(\mathbf{Y}) - \boldsymbol{\mu} \right\|^2 \right].$$

Se \mathbf{Y} tem distribuição Normal multivariada, então o vetor de médias $\boldsymbol{\mu}$ é um vetor de parâmetros de locação. Como a função perda em erro quadrático é invariante e a ação do grupo G no subespaço V é transitiva, segue que a função

risco depende apenas de $\tilde{\delta}$ e não de μ . Dessa forma, $\mathcal{R}(\tilde{\delta}, \mu) = \mathcal{R}(\tilde{\delta}, \mathbf{0})$.

Assim, para minimizar a função risco, basta minimizar:

$$\mathcal{R}(\tilde{\delta}, \mathbf{0}) = E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{Y})\|^2 \right].$$

O valor que minimiza a função risco é obtido pelo condicionamento de sua esperança. Isso será feito condicionando-se em $\mathbf{W} = \left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \mathbf{Y}$. Pela propriedade da esperança condicional tem-se que:

$$E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{Y})\|^2 \right] = E_{\mathbf{0}} \left[E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{W})\|^2 | \mathbf{W} \right] \right],$$

em que a igualdade $u(\mathbf{Y}) = u(\mathbf{W})$ deve-se ao fato de $\left(I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right)$ ser a projeção ortogonal de \mathbf{Y} no complemento ortogonal do subespaço V .

Consequentemente, dado que $\mathbf{w} = (w_1, \dots, w_n)$ foi observado vem que:

$$\begin{aligned} \mathcal{R}(\tilde{\delta}, \mathbf{0}) &= E_{\mathbf{0}} \left[E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{w})\|^2 | \mathbf{w} \right] \right] \\ &= \int E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{w})\|^2 | \mathbf{w} \right] f_{\mathbf{W}}(\mathbf{w}; \mathbf{0}) d\mathbf{w}. \end{aligned}$$

A função risco é minimizada quando $E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{w})\|^2 | \mathbf{w} \right]$ é mínima para cada \mathbf{w} . Derivando a última expressão em relação a $u(\mathbf{w})$ e igualando a derivada a zero, o $\min_u E_{\mathbf{0}} \left[\|\delta(\mathbf{Y}) + u(\mathbf{w})\|^2 | \mathbf{w} \right]$ ocorre para:

$$u(\mathbf{w}) = -E_{\mathbf{0}} [\delta(\mathbf{Y}) | \mathbf{w}].$$

Considerando o estimador equivariante $\delta(\mathbf{Y}) = \sum_{i=1}^n \lambda_i \hat{\mu}_i$, a função risco é minimizada quando $u(\mathbf{w}) = -E_{\mathbf{0}} [\delta(\mathbf{Y}) | \mathbf{w}] = -E_{\mathbf{0}} \left[\sum_{i=1}^n \lambda_i \hat{\mu}_i | \mathbf{w} \right]$.

Tem-se agora que, dado \mathbf{w} no complemento ortogonal de V , os valores de \mathbf{Y} têm que, necessariamente, estar em um plano que passe por \mathbf{w} e seja paralelo ao

subespaço V . Dessa forma, seja $\mathbf{Y} = \mathbf{w} + \mathbf{Z}$, com $\mathbf{Z} \in V$. Como o vetor de médias $\boldsymbol{\mu}$ é igual a zero, a esperança do vetor \mathbf{Y} restrito a qualquer subespaço paralelo a V é:

$$E_0[\mathbf{Y}] = E_0[\mathbf{w} + \mathbf{Z}] = \mathbf{w} + E_0[\mathbf{Z}] = \mathbf{w}.$$

Portanto,

$$\begin{aligned} u(\mathbf{w}) &= -E_0 \left[\sum_{i=1}^n \lambda_i \hat{\mu}_i | \mathbf{w} \right] \\ &= -E_0 \left[\sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \right)_i | \mathbf{w} \right] \\ &= -E_0 \left[\sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' (\mathbf{w} + \mathbf{Z}) \right)_i \right] \\ &= - \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' E_0[\mathbf{w} + \mathbf{Z}] \right)_i \\ &= - \sum_{i=1}^n \lambda_i \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{w} \right)_i \\ &= - \sum_{i=1}^n \lambda_i \left\{ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \left[I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y} \right\}_i \\ &= - \sum_{i=1}^n \lambda_i \left\{ \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right)^2 \right] \mathbf{y} \right\}_i \\ &= - \sum_{i=1}^n \lambda_i \left\{ \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right] \mathbf{y} \right\}_i \\ &= - \sum_{i=1}^n \lambda_i (\mathbf{0y})_i \\ &= 0, \end{aligned}$$

em que $\left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right)^2 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$, pois a matriz $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ é idempotente. Como $\tilde{\delta} = \delta + u$, o estimador MRE de $\boldsymbol{\mu}$ é dado por:

$$\tilde{\delta}(\mathbf{Y}) = \delta(\mathbf{Y}) = \sum_{i=1}^n \lambda_i \hat{\mu}_i.$$

2.5 Equivariância para parâmetros de escala

A estrutura nesta subseção, dedicada à equivariância de parâmetros de escala, será similar à utilizada para equivariância de parâmetros de locação. Nesse sentido, o objetivo principal nesta subseção é determinar o estimador MRE para parâmetros de escala.

Definição 2.5.1 (Parâmetro de escala): Seja $\{f(\cdot; \tau), \tau > 0\}$ uma família de densidades indexadas pelo parâmetro real positivo τ . O parâmetro τ é definido como sendo um *parâmetro de escala* se a densidade $f_X(x; \tau)$ pode ser escrita como $\frac{1}{\tau} h\left(\frac{x}{\tau}\right)$, para alguma função $h(\cdot)$.

Proposição 2.5.1: τ é um parâmetro de escala se, e somente se, a distribuição de $\frac{X}{\tau}$ não depende de τ .

Prova:

Como τ é um parâmetro de escala, segue por hipótese que $f_X(x; \tau) = \frac{1}{\tau} h\left(\frac{x}{\tau}\right)$. Se $Y = g(X) = \frac{X}{\tau}$, então $g^{-1}(y) = \tau y$ e $\left|\frac{d}{dy} g^{-1}(y)\right| = \tau$. Com base no teorema da transformação de variáveis, a distribuição de $Y = \frac{X}{\tau}$ é dada por:

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X(g^{-1}(y)) = \tau f_X(\tau y) = \tau \frac{1}{\tau} h\left(\frac{\tau y}{\tau}\right) = h(y).$$

Portanto, a distribuição de $Y = \frac{X}{\tau}$ não depende de τ .

Reciprocamente, seja $Y = \frac{X}{\tau}$, em que a distribuição de Y é dada por uma função $h(y)$. Como $X = g(Y) = \tau Y$, logo $g^{-1}(x) = \frac{x}{\tau}$ e $\left|\frac{d}{dx} g^{-1}(x)\right| = \frac{1}{\tau}$.

Assim, utilizando novamente o teorema da transformação de variáveis, segue que a distribuição de $X = \tau Y$ é:

$$f_X(x) = \left| \frac{d}{dx} g^{-1}(x) \right| f_Y(g^{-1}(x)) = \frac{1}{\tau} f\left(\frac{x}{\tau}\right) = \frac{1}{\tau} h\left(\frac{x}{\tau}\right).$$

Portanto, a distribuição de $f_X(x; \tau)$ pode ser escrita como $\frac{1}{\tau} h\left(\frac{x}{\tau}\right)$, o que demonstra que τ é um parâmetro de escala.

■

Pode-se observar que, se τ é um parâmetro de escala para a família de densidades $\{f(\cdot; \tau), \tau > 0\}$, então a função $h(\cdot)$ da definição é uma função densidade dada por $h(\cdot) = f(\cdot; 1)$.

Exemplo 2.5.1: Se $f_X(x; \tau) = \frac{1}{\tau} \exp\left(-\frac{x}{\tau}\right) I_{(0, \infty)}(x)$, então τ é um parâmetro de escala, uma vez que $h_X(x) = \exp(-x) I_{(0, \infty)}(x)$ é uma densidade e $f_X(x; \tau) = \frac{1}{\tau} h_X\left(\frac{x}{\tau}\right)$.

Exemplo 2.5.2: Se X possui fdp dada por:

$$f_X(x; \sigma^2) = \phi_{0, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right],$$

então σ é um parâmetro de escala, uma vez que $h_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ é uma densidade e $f_X(x; \sigma^2) = \frac{1}{\sigma} h_X\left(\frac{x}{\sigma}\right)$.

Exemplo 2.5.3: Se $f_X(x; \tau) = \frac{1}{\tau} I_{(0, \tau)}(x) = \frac{1}{\tau} I_{(0, 1)}\left(\frac{x}{\tau}\right)$, então τ é um parâmetro de escala, uma vez que $I_{(0, 1)}(x)$ é uma densidade.

Exemplo 2.5.4: Se $f_X(x; \tau) = \frac{1}{\tau} I_{(\tau, 2\tau)}(x) = \frac{1}{\tau} I_{(1, 2)}\left(\frac{x}{\tau}\right)$, então τ é um parâmetro de escala, uma vez que $I_{(1, 2)}(x)$ é uma densidade.

A definição para parâmetros de escala pode ser generalizada para o caso de densidades conjuntas. Se $\mathbf{X} = (X_1, \dots, X_n)$ tem densidade conjunta dada por $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \tau)$, então τ é um parâmetro de escala se:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \tau) = \frac{1}{\tau^n} f_{X_1, \dots, X_n}\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}; 1\right).$$

2.5.1 Ação de grupos para parâmetros de escala

Em diversas situações é necessário que medições sejam realizadas em diferentes unidades. Alguns exemplos são: altura sendo medida em polegadas ou centímetros, peso sendo medido em gramas ou quilogramas, ou distância sendo mensurada em milhas ou quilômetros. Em todas essas situações comuns em experimentos é razoável exigir que a inferência estatística a ser realizada seja independente das unidades de medida envolvidas. Se o procedimento estatístico é a de estimação pontual, pode-se exigir que o estimador a ser utilizado satisfaça a propriedade de equivariância por escala definida abaixo. A ideia é que um estimador será equivariante por escala se esse estimador não depender essencialmente da escala de medição adotada.

Pode-se considerar o problema de se estimar uma potência do parâmetro de escala τ , isto é, se $\delta(X_1, \dots, X_n)$ é um estimador de τ^r .

Definição 2.5.2 (Equivariância por escala): Um estimador δ de τ^r é *equivariante por escala* se:

$$\delta(cX_1, \dots, cX_n) = c^r \delta(X_1, \dots, X_n),$$

para todo (X_1, \dots, X_n) e todo $c > 0$.

Exemplo 2.5.5: Como exemplo básico de um estimador equivariante por escala,

tem-se o estimador usual da variância no caso de amostras iid, que é dada por:

$$\delta(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

De fato, pode-se observar que a variância amostral satisfaz a definição de equivariância por escala.

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

em que $r = 2$.

Por consequência da relação entre a variância e o desvio padrão, outro estimador equivariante por escala é o desvio padrão amostral:

$$\delta^*(c\mathbf{X}) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (cX_i - c\bar{X}_n)^2} = c \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2} = c\delta^*(\mathbf{X}),$$

em que $r = 1$ para esse estimador.

Se o objetivo é estimar o parâmetro τ^r no contexto de equivariância por escala, um grupo natural a ser considerado é o grupo $G = (\mathbb{R}^+, \cdot)$. Esse grupo atua, respectivamente, no espaço paramétrico Θ , no espaço amostral \mathfrak{X}^n e no espaço dos estimadores D da seguinte forma:

1. $G \times \Theta \rightarrow \Theta$, $(c, \tau) \mapsto c \cdot \tau = c\tau$.
2. $G \times \mathfrak{X}^n \rightarrow \mathfrak{X}^n$, $(c, \mathbf{x}) \mapsto c \cdot \mathbf{x} = (cx_1, \dots, cx_n)$.
3. $G \times D \rightarrow D$, $(c, \delta) \mapsto (c \cdot \delta)(\mathbf{x}) = \delta(c \cdot \mathbf{x}) = \delta(c\mathbf{x}) = c^r \delta(\mathbf{x})$.

Por sua vez, uma função perda é invariante por escala por essas ações se:

$$\begin{aligned}
 L(\delta(x_1, \dots, x_n), \tau) &= L((c \cdot \delta)(x_1, \dots, x_n), c \cdot \tau) \\
 &= L(\delta(c \cdot (x_1, \dots, x_n)), c \cdot \tau) \\
 &= L(\delta(cx_1, \dots, cx_n), c\tau) \\
 &= L(c^r \delta(x_1, \dots, x_n), c\tau).
 \end{aligned}$$

Pode-se verificar que a função perda em erro quadrático $L(\delta, \tau) = (\delta - \tau)^2$ não satisfaz a condição de invariância por escala. Nesse caso, é necessário considerar outras funções perda como $L(\delta, \tau) = \frac{(\delta - \tau^r)^2}{\tau^{2r}}$ ou $L(\delta, \tau) = \frac{|\delta - \tau^r|}{\tau^r}$. Considerando a função perda $L(\delta, \tau) = \frac{(\delta - \tau^r)^2}{\tau^{2r}}$ vem que:

$$\begin{aligned}
 L(c^r \delta(x_1, \dots, x_n), c\tau) &= \frac{[c^r \delta(x_1, \dots, x_n) - (c\tau)^r]^2}{(c\tau)^{2r}} \\
 &= \frac{[c^r (\delta(x_1, \dots, x_n) - \tau^r)]^2}{(c\tau)^{2r}} \\
 &= \frac{c^{2r} (\delta(x_1, \dots, x_n) - \tau^r)^2}{c^{2r} \tau^{2r}} \\
 &= \frac{(\delta(x_1, \dots, x_n) - \tau^r)^2}{\tau^{2r}} \\
 &= L(\delta(x_1, \dots, x_n), \tau),
 \end{aligned}$$

ou seja, a função $L(\delta, \tau)$ assim definida é invariante por escala.

Para se obter o estimador MRE para parâmetros de escala, primeiramente é necessário caracterizar a totalidade dos estimadores equivariantes por escala. Esse é o resultado do próximo lema.

Lema 2.5.1: Seja δ_0 um estimador equivariante de τ^r , em que τ é um parâmetro de escala. Um estimador δ é equivariante por escala se, e somente se, existe

uma função invariante u tal que:

$$\delta(\mathbf{x}) = \frac{\delta_0(\mathbf{x})}{u(\mathbf{x})}. \quad (2.6)$$

Prova:

Se δ_0 é um estimador equivariante por escala conhecido e u é uma função invariante, então:

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

Portanto, δ é um estimador equivariante por escala.

Reciprocamente, se δ é um estimador equivariante e $u(\mathbf{x}) = \frac{\delta_0(\mathbf{x})}{\delta(\mathbf{x})}$, então:

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

Portanto, u é uma função invariante. ■

Vamos obter para a função invariante u um sistema simplificado de coordenadas. As ações do grupo $G = (\mathbb{R}^+, \cdot)$ em relação a \mathbb{R}^n são dadas por:

$$\begin{aligned} G \times \mathbb{R}^n &\rightarrow \mathbb{R}^n \\ (c, \mathbf{x}) &\mapsto c\mathbf{x} \end{aligned}$$

Se $u : \mathbb{R}^n \rightarrow \mathbb{R}$ é invariante pela ação do grupo G , então é possível expressar essa função em um sistema de variáveis mais adequado. As órbitas das ações de G em \mathbb{R}^n são semirretas $\{c\mathbf{x}, c \in \mathbb{R}\}$. Um conjunto de pontos para o qual

as órbitas de G interceptam em apenas um ponto é formado por dois planos paralelos definidos por $x_n = -1$ e $x_n = +1$. Esses planos podem ser adequadamente descritos por:

$$\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, \frac{x_n}{|x_n|} \right).$$

Se $x_n < 0$,

$$\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, -\frac{x_n}{x_n} \right) = \left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, -1 \right).$$

Se $x_n > 0$,

$$\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, \frac{x_n}{x_n} \right) = \left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, 1 \right).$$

Logo, a órbita pelo ponto \mathbf{x} passa pelo ponto $\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, -1 \right)$ ou pelo ponto $\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, 1 \right)$ e, dessa forma, os dois hiperplanos $x_n = -1$ e $x_n = +1$ parametrizam todas as órbitas, com exceção para aquelas que passam por pontos em que $x_n = 0$. Essas órbitas podem ser desconsideradas, pois formam um conjunto de probabilidade nula.

De forma similar ao que foi feito no contexto de equivariância por locação, essa caracterização das funções invariantes permitirá a obtenção de uma expressão para o estimador MRE para parâmetros de escala. Considerando uma função perda invariante $L(\delta(x_1, \dots, x_n), \tau) = L(c^r \delta(x_1, \dots, x_n), c\tau)$, deve-se minimizar a função risco \mathcal{R} , dada por:

$$\mathcal{R}(\delta, \tau) = E_\tau [L(\delta(\mathbf{X}), \tau)] = \int \dots \int L(\delta(\mathbf{x}), \tau) f_{\mathbf{X}}(\mathbf{x}; \tau) d\mathbf{x}.$$

Novamente, utilizando os resultados referentes a equivariância do estimador δ , e da invariância do modelo f e da função perda L , tem-se que:

$$\begin{aligned}
\mathcal{R}(\delta, \tau) &= E_{\tau} [L(\delta(\mathbf{X}), \tau)] \\
&= \int \dots \int L(\delta(\mathbf{x}), \tau) f_{\mathbf{X}}(\mathbf{x}; \tau) d\mathbf{x} \\
&= \int \dots \int L(c^r \delta(\mathbf{x}), c\tau) f_{\mathbf{X}}(\mathbf{x}; \tau) d\mathbf{x} && \text{(Invariância da função } L) \\
&= \int \dots \int L(\delta(c\mathbf{x}), c\tau) f_{\mathbf{X}}(\mathbf{x}; \tau) d\mathbf{x} && \text{(Equivariância de } \delta) \\
&= \int \dots \int L(\delta(\mathbf{y}), c\tau) f_{\mathbf{Y}}(\mathbf{y}; c\tau) d\mathbf{y} && \text{(Invariância de } f) \\
&= E_{c\tau} [L(\delta(\mathbf{Y}), c\tau)] \\
&= \mathcal{R}(\delta, c\tau).
\end{aligned}$$

Como G atua de forma transitiva no espaço paramétrico Θ , se $c = \tau^{-1}$, então $\mathcal{R}(\delta, \tau) = \mathcal{R}(\delta, \tau^{-1}\tau) = \mathcal{R}(\delta, 1)$. Portanto, $\min_{\delta} \mathcal{R}(\delta, \tau) = \min_{\delta} \mathcal{R}(\delta, 1)$, isto é, o problema de minimização está agora apenas relacionado ao estimador δ e não mais depende do parâmetro τ . Portanto, para se obter o estimador MRE basta resolver:

$$\min_{\delta} \mathcal{R}(\delta, 1) = \min_{\delta} \int \dots \int L(\delta(\mathbf{x}), 1) f_{\mathbf{X}}(\mathbf{x}; 1) d\mathbf{x}.$$

Se $\delta_0(\mathbf{x})$ é um estimador equivariante conhecido, todos os outros estimadores equivariantes por escala são da forma $\delta(\mathbf{x}) = \frac{\delta_0(\mathbf{x})}{u(\mathbf{x})}$, em que $u(\mathbf{x}) = u\left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, \frac{x_n}{|x_n|}\right)$. Logo,

$$\min_{\delta} \mathcal{R}(\delta, 1) = \min_u \int \dots \int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{x})}, 1\right) f_{\mathbf{X}}(\mathbf{x}; 1) d\mathbf{x}.$$

Para minimizar a função risco, a ideia é marginalizar em relação aos dois

planos $\left\{z = \left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, \frac{x_n}{|x_n|}\right)\right\}$. Isto é feito utilizando-se a fdp condicional de \mathbf{X} dado $\mathbf{Z} = \mathbf{z} = (z_1, \dots, z_n)$, em que $z_i = \frac{x_i}{x_n}$ para $i = 1, \dots, n-1$ e $z_n = \frac{x_n}{|x_n|}$. Como \mathbf{x} determina totalmente \mathbf{z} , a distribuição conjunta $f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}; 1)$ é degenerada, e neste caso,

$$f_{\mathbf{X}}(\mathbf{x}; 1) = f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z}; 1) = f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) f_{\mathbf{Z}}(\mathbf{z}; 1).$$

Logo,

$$\begin{aligned} \min_{\delta} \mathcal{R}_{\delta}(\delta, 1) &= \min_u \int \dots \int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})}, 1\right) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) f_{\mathbf{Z}}(\mathbf{z}; 1) d\mathbf{x}d\mathbf{z} \\ &= \min_u \int \dots \int \left(\int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})}, 1\right) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} \right) f_{\mathbf{Z}}(\mathbf{z}; 1) d\mathbf{z} \\ &= \int \dots \int \left(\min_u \int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})}, 1\right) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} \right) f_{\mathbf{Z}}(\mathbf{z}; 1) d\mathbf{z}. \end{aligned}$$

A minimização ocorre obtendo-se para cada vetor fixo \mathbf{z} , o valor $u(\mathbf{z})$ que minimiza

$$\int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})}, 1\right) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x}.$$

Com essa função $u(\mathbf{z})$, o estimador equivariante por escala de risco mínimo é:

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

Considere no processo de obtenção do estimador MRE para parâmetros de escala, a função perda quadrática padronizada $L(\delta, \tau) = \frac{(\delta - \tau)^2}{\tau^2}$. Consequentemente,

$$\begin{aligned} \min_u \int L(\delta(\mathbf{x}), 1) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} &= \min_u \int L\left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})}, 1\right) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} \\ &= \min_u \int \left(\frac{\delta_0(\mathbf{x})}{u(\mathbf{z})} - 1\right)^2 f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x}. \end{aligned}$$

Desenvolvendo, resulta que:

$$\min_u \left[\frac{1}{(u(\mathbf{z}))^2} \int (\delta_0(\mathbf{x}))^2 f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} - \frac{2}{u(\mathbf{z})} \int \delta_0(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} + 1 \right].$$

Derivando a última expressão em relação $u(\mathbf{z})$ e igualando a derivada a zero vem que:

$$-\frac{2}{(u(\mathbf{z}))^3} \int (\delta_0(\mathbf{x}))^2 f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} + \frac{2}{(u(\mathbf{z}))^2} \int \delta_0(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x} = 0.$$

Logo,

$$u(\mathbf{z}) = \frac{\int \delta_0(\mathbf{x})^2 f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x}}{\int (\delta_0(\mathbf{x})) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}; 1) d\mathbf{x}} = \frac{E_1[(\delta_0(\mathbf{X}))^2 | \mathbf{Z}]}{E_1[\delta_0(\mathbf{X}) | \mathbf{Z}]}$$

Como $\delta(\mathbf{X}) = \frac{\delta_0(\mathbf{X})}{u(\mathbf{Z})}$, o estimador MRE é dado por:

$$\delta(\mathbf{X}) = \frac{E_1[\delta_0(\mathbf{X}) | \mathbf{Z}]}{E_1[(\delta_0(\mathbf{X}))^2 | \mathbf{Z}]} \delta_0(\mathbf{X}). \quad (2.7)$$

O próximo exemplo ilustra uma situação em que o estimador MRE por escala de σ^2 de uma distribuição Normal com média conhecida pode ser obtido a partir de (2.7), utilizando-se algumas propriedades de estimadores.

Exemplo 2.5.6: Considere o problema de se obter o estimador MRE por escala da variância σ^2 de uma distribuição Normal com média zero. Um estimador equivariante por escala para σ^2 é $\delta_0(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2$.

De fato,

$$\delta_0(cX_1, \dots, cX_n) = \sum_{i=1}^n (cX_i)^2 = c^2 \sum_{i=1}^n X_i^2 = c^2 \delta_0(X_1, \dots, X_n).$$

Além disso, δ_0 é uma estatística suficiente e completa. Considere também a estatística $z(\mathbf{X}) = \mathbf{Z} = \left(\frac{X_1}{X_n}, \dots, \frac{X_{n-1}}{X_n}, \frac{X_n}{|X_n|} \right)$. Uma vez que \mathbf{Z} é uma estatística ancilar (LEHMANN; CASELLA, 1998, p. 170; CASELLA; BERGER, 2002, p. 284), segue pelo teorema de Basu (APÊNDICE C) que $\delta_0(\mathbf{X})$ e \mathbf{Z} são independentes. Logo:

$$E[\delta_0(\mathbf{X}) | \mathbf{Z}] = E[\delta_0(\mathbf{X})] = E\left[\sum_{i=1}^n X_i^2\right] = \sum_{i=1}^n E[X_i^2] = n\sigma^2$$

e

$$E[(\delta_0(\mathbf{X}))^2 | \mathbf{Z}] = E[(\delta_0(\mathbf{X}))^2] = E\left[\left(\sum_{i=1}^n X_i^2\right)^2\right].$$

Tomando $Y = \sum_{i=1}^n X_i^2$, tem-se que $\frac{Y}{\sigma^2} \sim \chi_{(n)}^2$. Dessa forma,

$$E\left[\frac{Y}{\sigma^2}\right] = n$$

e

$$\text{var} \left[\frac{Y}{\sigma^2} \right] = 2n.$$

Esses resultados implicam que $E[Y] = n\sigma^2$ e $\text{var}[Y] = 2n\sigma^4$, respectivamente. Uma vez que $\text{var}[Y] = E[Y^2] - (E[Y])^2$, então:

$$E[Y^2] = \text{var}[Y] + (E[Y])^2 = 2n\sigma^4 + n^2\sigma^4 = (2n + n^2)\sigma^4.$$

Portanto, o estimador MRE do parâmetro σ^2 é:

$$\delta(\mathbf{X}) = \frac{E_1[\delta_0(\mathbf{X})|\mathbf{Z}]}{E_1[(\delta_0(\mathbf{X}))^2|\mathbf{Z}]} \delta_0(\mathbf{X}) = \frac{n}{(n^2 + 2n)} \sum_{i=1}^n X_i^2 = \frac{1}{n+2} \sum_{i=1}^n X_i^2.$$

Para o caso geral nem sempre é possível obter um estimador δ_0 que seja uma estatística suficiente e completa. Para esses casos, seria necessário o cálculo da esperança condicional, o que pode ser muito trabalhoso. A ideia então é utilizar como δ_0 equivariante um estimador que seja o mais simples possível. Uma opção é considerar $\delta_0(X_1, \dots, X_n) = X_n^r$ na expressão em (2.7). Logo,

$$\delta(\mathbf{X}) = \frac{E_1[X_n^r|\mathbf{Z}]}{E_1[X_n^{2r}|\mathbf{Z}]} X_n^r. \quad (2.8)$$

Para o cálculo da esperança condicional, considere a transformação de coordenadas:

$$(x_1, \dots, x_{n-1}, x_n) \rightarrow \left(\frac{x_1}{x_n}, \dots, \frac{x_{n-1}}{x_n}, x_n \right) = (z_1, \dots, z_{n-1}, z_n),$$

dada por:

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{bmatrix} = \begin{bmatrix} 1/x_n & 0 & \dots & 0 & 0 \\ 0 & 1/x_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1/x_n & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix},$$

ou seja,

$$z_i = \frac{x_i}{x_n}; \quad z_n = x_n, \quad \text{com } i = 1, \dots, n-1.$$

Consequentemente,

$$x_i = z_i z_n; \quad x_n = z_n, \quad \text{com } i = 1, \dots, n-1.$$

Segue então que o Jacobiano dessa transformação é:

$$J = \det \begin{bmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_1}{\partial z_2} & \dots & \frac{\partial x_1}{\partial z_n} \\ \frac{\partial x_2}{\partial z_1} & \frac{\partial x_2}{\partial z_2} & \dots & \frac{\partial x_2}{\partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial z_1} & \frac{\partial x_n}{\partial z_2} & \dots & \frac{\partial x_n}{\partial z_n} \end{bmatrix} = \det \begin{bmatrix} z_n & 0 & \dots & z_1 \\ 0 & z_n & \dots & z_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = z_n^{n-1}.$$

Pelo teorema da transformação de variáveis vem que a distribuição conjunta dos Z 's é:

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}; 1) &= |J| f_{X_1, \dots, X_{n-1}, X_n}(x_1(z_1, z_n), \dots, x_{n-1}(z_{n-1}, z_n), x_n(z_n); 1) \\ &= z_n^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 z_n, \dots, z_{n-1} z_n, z_n; 1). \end{aligned}$$

A densidade condicional de X_n dado $\mathbf{Z} = \mathbf{z} = (z_1, \dots, z_{n-1})$ é:

$$\begin{aligned} f_{X_n|\mathbf{Z}}(x_n|\mathbf{z}) &= f_{Z_n|Z_1, \dots, Z_{n-1}}(z_n|z_1, \dots, z_{n-1}) \\ &= \frac{f_{Z_1, \dots, Z_{n-1}, Z_n}(z_1, \dots, z_{n-1}, z_n; 1)}{f_{Z_1, \dots, Z_{n-1}}(z_1, \dots, z_{n-1}; 1)} \\ &= \frac{z_n^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 z_n, \dots, z_{n-1} z_n, z_n; 1)}{\int t^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 t, \dots, z_{n-1} t, t; 1) dt}. \end{aligned}$$

Pode-se agora calcular a esperança condicional de X_n^r dado que $\mathbf{Z} = \mathbf{z} = (z_1, \dots, z_{n-1})$. Assim,

$$\begin{aligned} E_1[X_n^r|\mathbf{Z}] &= E_1[Z_n^r|\mathbf{Z}] = \frac{\int t^r t^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 t, \dots, z_{n-1} t, t; 1) dt}{\int t^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 t, \dots, z_{n-1} t, t; 1) dt} \\ &= \frac{\int t^{n+r-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 t, \dots, z_{n-1} t, t; 1) dt}{\int t^{n-1} f_{X_1, \dots, X_{n-1}, X_n}(z_1 t, \dots, z_{n-1} t, t; 1) dt}. \end{aligned}$$

Com base nas relações $z_i = \frac{x_i}{x_n}$ ($i = 1, \dots, n-1$), a esperança condicional de X_n^r pode ser expressa em termos de x 's como:

$$E_1[X_n^r|\mathbf{Z}] = \frac{\int t^{n+r-1} f_{X_1, \dots, X_{n-1}, X_n}\left(\frac{x_1}{x_n} t, \dots, \frac{x_{n-1}}{x_n} t, t; 1\right) dt}{\int t^{n-1} f_{X_1, \dots, X_{n-1}, X_n}\left(\frac{x_1}{x_n} t, \dots, \frac{x_{n-1}}{x_n} t, t; 1\right) dt}.$$

Fazendo a mudança de variáveis $v = \frac{t}{x_n}$, em que $dt = x_n dv$, vem que:

$$\begin{aligned}
E_1 [X_n^r | \mathbf{Z}] &= \frac{\int (x_n v)^{n+r-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) x_n dv}{\int (x_n v)^{n-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) x_n dv} \\
&= \frac{x_n^{n+r} \int v^{n+r-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv}{x_n^n \int v^{n-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv} \\
&= x_n^r \frac{\int v^{n+r-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv}{\int v^{n-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv}.
\end{aligned}$$

De forma análoga,

$$E_1 [X_n^{2r} | \mathbf{Z}] = x_n^{2r} \frac{\int_0^\infty v^{n+2r-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv}{\int_0^\infty v^{n-1} f_{X_1, \dots, X_{n-1}, X_n} (x_1 v, \dots, x_{n-1} v, x_n v; 1) dv}.$$

Finalmente, o estimador MRE por escala é dada por:

$$\begin{aligned}
\delta (X_1, \dots, X_n) &= \frac{E_1 [X_n^r | \mathbf{Z}]}{E_1 [X_n^{2r} | \mathbf{Z}]} x_n^r \\
&= \frac{x_n^r \int_0^\infty v^{n+r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv}{x_n^{2r} \int_0^\infty v^{n+2r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv} x_n^r \\
&= \frac{x_n^{2r} \int_0^\infty v^{n+r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv}{x_n^{2r} \int_0^\infty v^{n+2r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv} \\
&= \frac{\int_0^\infty v^{n+r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv}{\int_0^\infty v^{n+2r-1} f_{X_1, \dots, X_n} (x_1 v, \dots, x_n v; 1) dv}, \tag{2.9}
\end{aligned}$$

que é conhecido como estimador de Pitman para parâmetros de escala.

Para o caso em que $r = 1$,

$$\delta(X_1, \dots, X_n) = \frac{\int_0^{\infty} v^n f_{X_1, \dots, X_n}(x_1 v, \dots, x_n v; 1) dv}{\int_0^{\infty} v^{n+1} f_{X_1, \dots, X_n}(x_1 v, \dots, x_n v; 1) dv}.$$

Como o parâmetro é de escala, $f(x_1 v, \dots, x_n v; 1) = \frac{1}{v^n} f(x_1, \dots, x_n; \frac{1}{v})$.

Substituindo na expressão anterior tem-se que:

$$\delta(X_1, \dots, X_n) = \frac{\int_0^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \frac{1}{v}) dv}{\int_0^{\infty} v f_{X_1, \dots, X_n}(x_1, \dots, x_n; \frac{1}{v}) dv}.$$

Fazendo $\frac{1}{v} = \tau$, segue que $v = \frac{1}{\tau}$ e $dv = -\frac{1}{\tau^2} d\tau$. Assim, como $v : 0 \rightarrow \infty$, tem-se então que $\tau : \infty \rightarrow 0$. Portanto,

$$\delta(X_1, \dots, X_n) = \frac{\int_0^{\infty} \frac{1}{\tau^2} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \tau) d\tau}{\int_0^{\infty} \frac{1}{\tau^3} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \tau) d\tau}.$$

Se as amostras são iid, obtém-se a expressão do estimador de Pitman para parâmetros de escala fornecida em Mood, Graybill e Boes (1974, p. 337).

Teorema 2.5.1: Seja (X_1, \dots, X_n) uma amostra aleatória de uma densidade $f(\cdot; \tau)$, em que $\tau > 0$ é um parâmetro de escala. Na classe dos estimadores equivalentes por escala, o estimador

$$\delta(X_1, \dots, X_n) = \frac{\int_0^{\infty} \left(\frac{1}{\tau^2}\right) \prod_{i=1}^n f_{X_i}(X_i; \tau) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau^3}\right) \prod_{i=1}^n f_{X_i}(X_i; \tau) d\tau}, \quad (2.10)$$

possui menor risco uniforme para a função perda $L(\delta, \tau) = (\delta - \tau)^2 / \tau^2$.

Exemplo 2.5.7: Seja (X_1, \dots, X_n) uma amostra iid de uma densidade $f_X(x; \tau) = \frac{1}{\tau} I_{(0, \tau)}(x)$. Considere $Y_n = \max[X_1, \dots, X_n]$. Como $Y_n \leq \tau$, o estimador de Pitman para o parâmetro de escala é:

$$\begin{aligned}
\frac{\int_0^\infty (1/\tau^2) \prod_{i=1}^n (1/\tau) I_{(0, \tau)}(X_i) d\tau}{\int_0^\infty (1/\tau^3) \prod_{i=1}^n (1/\tau) I_{(0, \tau)}(X_i) d\tau} &= \frac{\int_{Y_n}^\infty (1/\tau^2) \prod_{i=1}^n (1/\tau) I_{(0, X_i)}(\tau) d\tau}{\int_{Y_n}^\infty (1/\tau^3) \prod_{i=1}^n (1/\tau) I_{(0, X_i)}(\tau) d\tau} \\
&= \frac{\int_{Y_n}^\infty \tau^{-n-2} d\tau}{\int_{Y_n}^\infty \tau^{-n-3} d\tau} \\
&= \frac{\tau^{-(n+2)+1} / [-(n+2)+1] \Big|_{Y_n}^\infty}{\tau^{-(n+3)+1} / [-(n+3)+1] \Big|_{Y_n}^\infty} \\
&= \frac{Y_n^{-(n+2)+1} / [(n+2)-1]}{Y_n^{-(n+3)+1} / [(n+3)-1]} \\
&= \frac{Y_n^{-(n+2)+1} (n+3) - 1}{(n+2) - 1 Y_n^{-(n+3)+1}} \\
&= \frac{n+2}{n+1} \frac{Y_n^{-n-1}}{Y_n^{-n-2}} \\
&= \frac{n+2}{n+1} Y_n.
\end{aligned}$$

Conforme Mood, Graybill e Boes (1974, p. 338), Y_n é uma estatística suficiente e completa, sendo $E[Y_n] = \frac{n}{n+1}\tau$. Então, pelo teorema Lehmann-Schefté (APÊNDICE C), $\frac{n+1}{n}Y_n$ é o UMVUE de τ .

Exemplo 2.5.8: Seja (X_1, \dots, X_n) uma amostra aleatória de uma densidade Exponencial $f_X(x; \tau) = \frac{1}{\tau} \exp\left(-\frac{x}{\tau}\right) I_{(0, \infty)}(x)$. O estimador de Pitman para o parâmetro de escala é:

$$\begin{aligned}
\frac{\int_0^{\infty} \left(\frac{1}{\tau^2}\right) \prod_{i=1}^n f(X_i; \tau) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau^3}\right) \prod_{i=1}^n f(X_i; \tau) d\tau} &= \frac{\int_0^{\infty} \left(\frac{1}{\tau^2}\right) \prod_{i=1}^n \left(\frac{1}{\tau}\right) \exp\left(-\frac{X_i}{\tau}\right) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau^3}\right) \prod_{i=1}^n \left(\frac{1}{\tau}\right) \exp\left(-\frac{X_i}{\tau}\right) d\tau} \\
&= \frac{\int_0^{\infty} \left(\frac{1}{\tau}\right)^{n+2} \exp\left(-\sum_{i=1}^n \frac{X_i}{\tau}\right) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau}\right)^{n+3} \exp\left(-\sum_{i=1}^n \frac{X_i}{\tau}\right) d\tau}. \tag{2.11}
\end{aligned}$$

Seja $\alpha = \frac{\sum X_i}{\tau}$. Isso implica que $\tau = \frac{\sum X_i}{\alpha}$ e $d\tau = -\frac{\sum X_i}{\alpha^2} d\alpha$. Segue também que, como $\tau = \frac{\sum X_i}{\alpha}$ então $\left(\frac{1}{\tau}\right)^{n+2} = \left(\frac{\alpha}{\sum X_i}\right)^{n+2}$. De forma análoga, $\left(\frac{1}{\tau}\right)^{n+3} = \left(\frac{\alpha}{\sum X_i}\right)^{n+3}$.

Substituindo esses resultados em (2.11) resulta que:

$$\begin{aligned}
\frac{\int_0^{\infty} \left(\frac{1}{\tau^2}\right) \prod_{i=1}^n f(X_i; \tau) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau^3}\right) \prod_{i=1}^n f(X_i; \tau) d\tau} &= \frac{\int_0^{\infty} \left(\frac{\alpha}{\sum_{i=1}^n X_i}\right)^{n+2} \exp(-\alpha) \left(-\frac{\sum_{i=1}^n X_i}{\alpha^2}\right) d\alpha}{\int_0^{\infty} \left(\frac{\alpha}{\sum_{i=1}^n X_i}\right)^{n+3} \exp(-\alpha) \left(-\frac{\sum_{i=1}^n X_i}{\alpha^2}\right) d\alpha}. \tag{2.12}
\end{aligned}$$

Desenvolvendo primeiramente o numerador em (2.12) vem que:

$$\begin{aligned}
\int_0^{\infty} \left(\frac{\alpha}{\sum X_i}\right)^{n+2} \exp(-\alpha) \left(-\frac{\sum X_i}{\alpha^2}\right) d\alpha &= -\int_0^{\infty} \frac{\sum X_i}{(\sum X_i)^{n+2}} \frac{\alpha^{n+2}}{\alpha^2} \exp(-\alpha) d\alpha \\
&= -\int_0^{\infty} (\sum X_i)^{-n-1} \alpha^n \exp(-\alpha) d\alpha. \tag{2.13}
\end{aligned}$$

De modo análogo, o denominador em (2.12) é dado por:

$$\int_0^{\infty} \left(\frac{\alpha}{\sum X_i} \right)^{n+3} \exp(-\alpha) \left(-\frac{\sum X_i}{\alpha^2} \right) d\alpha = - \int_0^{\infty} \left(\sum X_i \right)^{-n-2} \alpha^{n+1} \exp(-\alpha) d\alpha. \quad (2.14)$$

Substituindo (2.13) e (2.14) em (2.12), segue que o estimador de Pitman de τ é:

$$\begin{aligned} \frac{\int_0^{\infty} \left(\frac{1}{\tau^2} \right) \prod_{i=1}^n f(X_i; \tau) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau^3} \right) \prod_{i=1}^n f(X_i; \tau) d\tau} &= \frac{\int_0^{\infty} \left(\frac{\alpha}{\sum_{i=1}^n X_i} \right)^{n+2} \exp(-\alpha) \left(-\frac{\sum_{i=1}^n X_i}{\alpha^2} \right) d\alpha}{\int_0^{\infty} \left(\frac{\alpha}{\sum_{i=1}^n X_i} \right)^{n+3} \exp(-\alpha) \left(-\frac{\sum_{i=1}^n X_i}{\alpha^2} \right) d\alpha} \\ &= \frac{-(\sum X_i)^{-n-1} \int_0^{\infty} \alpha^n \exp(-\alpha) d\alpha}{-(\sum X_i)^{-n-2} \int_0^{\infty} \alpha^{n+1} \exp(-\alpha) d\alpha} \\ &= \sum X_i \frac{\int_0^{\infty} \alpha^n \exp(-\alpha) d\alpha}{\int_0^{\infty} \alpha^{n+1} \exp(-\alpha) d\alpha} \\ &= \sum X_i \frac{\Gamma(n+1)}{\Gamma(n+2)} \\ &= \sum X_i \frac{\Gamma(n+1)}{\Gamma[(n+1)+1]} \\ &= \sum X_i \frac{\Gamma(n+1)}{(n+1)\Gamma(n+1)} \\ &= \frac{\sum X_i}{n+1}. \end{aligned}$$

Segundo Mood, Graybill e Boes (1974, p. 338), pode-se provar que o UMVUE de τ é $\frac{\sum X_i}{n}$. Uma vez que $\frac{\sum X_i}{n}$ é equivariante por escala e $\frac{\sum X_i}{n+1}$ é o

estimador MRE do parâmetro de escala τ para a função perda $\frac{(\delta-\tau)^2}{\tau^2}$, o risco de $\frac{\sum X_i}{n+1}$ é uniformemente menor que o risco de $\frac{\sum X_i}{n}$.

Exemplo 2.5.9: Seja (X_1, \dots, X_n) uma amostra aleatória de uma distribuição Weibull, que possui fdp dada por:

$$f_X(x; \tau, \xi) = \frac{\xi}{\tau} \left(\frac{x}{\tau}\right)^{\xi-1} \exp\left[-\left(\frac{x}{\tau}\right)^\xi\right],$$

definida para $x > 0$, sendo $\tau > 0$ e $\xi > 0$ os parâmetros de escala e de forma, respectivamente. Para $\xi = 1$, a distribuição Weibull corresponde à distribuição Exponencial com parâmetro de escala τ .

O estimador de Pitman para o parâmetro τ da distribuição de Weibull é dado por:

$$\begin{aligned} \delta(X_1, \dots, X_n) &= \frac{\int_0^\infty \frac{1}{\tau^2} \prod_{i=1}^n \frac{\xi}{\tau} \left(\frac{X_i}{\tau}\right)^{\xi-1} \exp\left[-\left(\frac{X_i}{\tau}\right)^\xi\right] d\tau}{\int_0^\infty \frac{1}{\tau^3} \prod_{i=1}^n \frac{\xi}{\tau} \left(\frac{X_i}{\tau}\right)^{\xi-1} \exp\left[-\left(\frac{X_i}{\tau}\right)^\xi\right] d\tau} \\ &= \frac{\int_0^\infty \xi^n \left(\frac{1}{\tau}\right)^2 \left(\frac{1}{\tau}\right)^n \left(\frac{1}{\tau}\right)^{n(\xi-1)} \prod_{i=1}^n X_i^{\xi-1} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau}{\int_0^\infty \xi^n \left(\frac{1}{\tau}\right)^3 \left(\frac{1}{\tau}\right)^n \left(\frac{1}{\tau}\right)^{n(\xi-1)} \prod_{i=1}^n X_i^{\xi-1} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau} \\ &= \frac{\int_0^\infty \left(\frac{1}{\tau}\right)^{2+n+n\xi-n} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau}{\int_0^\infty \left(\frac{1}{\tau}\right)^{3+n+n\xi-n} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau} \\ &= \frac{\int_0^\infty \left(\frac{1}{\tau}\right)^{n\xi+2} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau}{\int_0^\infty \left(\frac{1}{\tau}\right)^{n\xi+3} \exp\left[-\sum_{i=1}^n \left(\frac{X_i}{\tau}\right)^\xi\right] d\tau}. \end{aligned}$$

Tomando-se $\lambda = \left(\frac{\sum_{i=1}^n X_i}{\tau} \right)^\xi$ tem-se que:

$$\begin{aligned} \lambda &= \left(\frac{\sum_{i=1}^n X_i}{\tau} \right)^\xi \Rightarrow \ln \lambda = \xi \ln \left(\frac{\sum_{i=1}^n X_i}{\tau} \right) \\ &\Rightarrow \ln \lambda = \xi \ln \sum_{i=1}^n X_i - \xi \ln \tau \\ &\Rightarrow \xi \ln \tau = \xi \ln \sum_{i=1}^n X_i - \ln \lambda \\ &\Rightarrow \ln \tau = \ln \sum_{i=1}^n X_i - \frac{1}{\xi} \ln \lambda \\ &\Rightarrow \ln \tau = \ln \sum_{i=1}^n X_i + \ln \lambda^{-\frac{1}{\xi}} \\ &\Rightarrow \ln \tau = \ln \lambda^{-\frac{1}{\xi}} \sum_{i=1}^n X_i \\ &\Rightarrow \tau = \lambda^{-\frac{1}{\xi}} \sum_{i=1}^n X_i. \end{aligned}$$

Como $\tau = \frac{\sum_{i=1}^n X_i}{\lambda^{\frac{1}{\xi}}}$, segue então que $d\tau = -\frac{1}{\xi} \frac{\sum_{i=1}^n X_i}{\lambda^{\frac{1}{\xi}+1}} d\lambda$. Além disso, segue também que $\left(\frac{1}{\tau}\right)^{n\xi+2} = \left(\frac{\lambda^{\frac{1}{\xi}}}{\sum_{i=1}^n X_i}\right)^{n\xi+2}$ e $\left(\frac{1}{\tau}\right)^{n\xi+3} = \left(\frac{\lambda^{\frac{1}{\xi}}}{\sum_{i=1}^n X_i}\right)^{n\xi+3}$.

Portanto, o estimador de Pitman de τ é dado por:

$$\begin{aligned}
\delta(X_1, \dots, X_n) &= \frac{\int_0^\infty \left(\frac{\lambda^{\frac{1}{\xi}}}{\sum_{i=1}^n X_i} \right)^{n\xi+2} \exp(-\lambda) \left(-\frac{1}{\xi} \frac{\sum_{i=1}^n X_i}{\lambda^{\frac{1}{\xi}+1}} \right) d\lambda}{\int_0^\infty \left(\frac{\lambda^{\frac{1}{\xi}}}{\sum_{i=1}^n X_i} \right)^{n\xi+3} \exp(-\lambda) \left(-\frac{\sum_{i=1}^n X_i}{\lambda^{\frac{1}{\xi}+1}} \right) d\lambda} \\
&= \frac{\int_0^\infty \left(\sum_{i=1}^n X_i \right)^{-n\xi-1} \lambda^{n+\frac{2}{\xi}-\frac{1}{\xi}-1} \exp(-\lambda) d\lambda}{\int_0^\infty \left(\sum_{i=1}^n X_i \right)^{-n\xi-2} \lambda^{n+\frac{3}{\xi}-\frac{1}{\xi}-1} \exp(-\lambda) d\lambda} \\
&= \sum_{i=1}^n X_i \frac{\int_0^\infty \lambda^{(n+\frac{1}{\xi})-1} \exp(-\lambda) d\lambda}{\int_0^\infty \lambda^{(n+\frac{2}{\xi})-1} \exp(-\lambda) d\lambda} \\
&= \frac{\Gamma\left(n + \frac{1}{\xi}\right)}{\Gamma\left(n + \frac{2}{\xi}\right)} \sum_{i=1}^n X_i.
\end{aligned}$$

O estimador MRE para o parâmetro de escala da distribuição Weibull depende do parâmetro de forma ξ e sua existência para o grupo $G = (\mathbb{R}^+, \cdot)$ está condicionada à estimação de ξ . Além disso, da relação entre as distribuições Weibull e Exponencial, tem-se que o estimador MRE para o parâmetro de escala da distribuição Weibull corresponde ao estimador MRE da distribuição Exponencial, quando $\xi = 1$.

No próximo exemplo será demonstrado que o estimador de Pitman (2.10) é equivariante em relação às ações do grupo $G = (\mathbb{R}^+, \cdot)$.

Exemplo 2.5.10: Seja (X_1, \dots, X_n) uma amostra aleatória de uma fdp $f_X(x; \tau)$.

Se τ é um parâmetro de escala, segue por definição que:

$$f(cx; \tau) = \frac{1}{\tau} h\left(\frac{cx}{\tau}\right) = \frac{1}{\tau} h\left(\frac{x}{\tau/c}\right) = \frac{1}{c} \left(\frac{1}{\tau/c}\right) h\left(\frac{x}{\tau/c}\right) = \frac{1}{c} f\left(x; \frac{\tau}{c}\right).$$

Assim,

$$\delta(cX_1, \dots, cX_n) = \frac{\int_0^{\infty} \left(\frac{1}{\tau}\right)^2 \prod_{i=1}^n f(cX_i; \tau) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau}\right)^3 \prod_{i=1}^n f(cX_i; \tau) d\tau} = \frac{\int_0^{\infty} \left(\frac{1}{\tau}\right)^2 \prod_{i=1}^n \frac{1}{c} f\left(X_i; \frac{\tau}{c}\right) d\tau}{\int_0^{\infty} \left(\frac{1}{\tau}\right)^3 \prod_{i=1}^n \frac{1}{c} f\left(X_i; \frac{\tau}{c}\right) d\tau}.$$

Tomando-se $\frac{\tau}{c} = \tau^*$, segue que $d\tau = cd\tau^*$, para τ^* variando de 0 a ∞ .

Desse modo,

$$\begin{aligned} \delta(cX_1, \dots, cX_n) &= \frac{\int_0^{\infty} \left(\frac{1}{c\tau^*}\right)^2 \prod_{i=1}^n \frac{1}{c} f(X_i; \tau^*) cd\tau^*}{\int_0^{\infty} \left(\frac{1}{c\tau^*}\right)^3 \prod_{i=1}^n \frac{1}{c} f(X_i; \tau^*) cd\tau^*} \\ &= \frac{\left(\frac{1}{c}\right)^2 \int_0^{\infty} \left(\frac{1}{\tau^*}\right)^2 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*}{\left(\frac{1}{c}\right)^3 \int_0^{\infty} \left(\frac{1}{\tau^*}\right)^3 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*} \\ &= \frac{c^3 \int_0^{\infty} \left(\frac{1}{\tau^*}\right)^2 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*}{c^2 \int_0^{\infty} \left(\frac{1}{\tau^*}\right)^3 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*} \\ &= c \frac{\int_0^{\infty} \left(\frac{1}{\tau^*}\right)^2 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*}{\int_0^{\infty} \left(\frac{1}{\tau^*}\right)^3 \prod_{i=1}^n f(X_i; \tau^*) d\tau^*} \\ &= c\delta(X_1, \dots, X_n), \end{aligned}$$

o que prova que o estimador de Pitman é equivariante por escala.

Proposição 2.5.2: O estimador de Pitman para parâmetros de escala é uma função de estatísticas suficientes.

Prova:

Se $S_1 = s_1(X_1, \dots, X_n), \dots, S_k = s_k(X_1, \dots, X_n)$ é um conjunto de estatísticas suficientes, então pelo teorema da fatoração:

$$\prod_{i=1}^n f(x_i; \tau) = g(s_1, \dots, s_n; \tau) h(x_1, \dots, x_n).$$

Então, o estimador de Pitman pode ser escrito como:

$$\begin{aligned} \frac{\int_0^{\infty} \frac{1}{\tau^2} \prod_{i=1}^n f(X_i; \tau) d\tau}{\int_0^{\infty} \frac{1}{\tau^3} \prod_{i=1}^n f(X_i; \tau) d\tau} &= \frac{\int_0^{\infty} \frac{1}{\tau^2} g(S_1, \dots, S_k; \tau) h(X_1, \dots, X_n) d\tau}{\int_0^{\infty} \frac{1}{\tau^3} g(S_1, \dots, S_k; \tau) h(X_1, \dots, X_n) d\tau} \\ &= \frac{\int_0^{\infty} \frac{1}{\tau^2} g(S_1, \dots, S_k; \tau) d\tau}{\int_0^{\infty} \frac{1}{\tau^3} g(S_1, \dots, S_k; \tau) d\tau}. \end{aligned}$$

Portanto, o estimador de Pitman para parâmetros de escala é uma função de estatísticas suficientes. ■

2.6 Estimação equivariante em famílias de locação-escala

Nas subseções anteriores foi possível determinar os estimadores MRE na classe dos estimadores equivariantes, considerando separadamente famílias de locação e de escala. Porém, é comum no processo de estimação que a distribuição candidata a modelar os dados seja de locação-escala, e, para esses casos, surge a necessidade de se utilizar métodos que permitam a estimação simultânea dos parâmetros de locação e de escala.

No exemplo relacionado à distribuição Normal (Exemplo 2.4.5), para determinar o estimador MRE da média θ foi necessário supor que a variância era conhecida. Nos exemplos 2.4.7 e 2.4.8, verificou-se que os estimadores MRE para os parâmetros de locação das distribuições Gumbel e Exponencial estavam em função do parâmetro de escala. Nesses exemplos, a condição de existência do estimador MRE para o grupo $G = (\mathbb{R}, +)$ está em função do conhecimento acerca do parâmetro de escala. Esses exemplos ilustram a necessidade da estimação simultânea considerando os parâmetros em famílias de locação-escala. Em 1994, Prabakaran e Chandrasekar (1994) apresentaram resultados sobre a estimação conjunta dos parâmetros em famílias de locação-escala.

Na primeira edição do livro *Theory of point estimation*, Lehmann (1983) já havia discutido o problema de estimação equivariante marginal de parâmetros em famílias de locação-escala. Todavia, Prabakaran e Chandrasekar (1994) desenvolveram alguns resultados requeridos para estimação equivariante simultânea e estenderam os resultados fornecidos por Lehmann (1983). Nesse sentido, o objetivo na presente subseção é apresentar os resultados obtidos por Prabakaran e Chandrasekar (1994). De maneira análoga ao que foi realizado nas subseções relacionadas a equivariância por locação e por escala, são necessários alguns conceitos e resultados requeridos para o desenvolvimento teórico desta subseção. Inicialmente, apresenta-se a definição de uma família locação-escala.

Definição 2.6.1 (Família locação-escala): Seja $h_X(x)$ uma fdp qualquer. Uma família de fdp's \mathcal{F} , indexada pelo vetor de parâmetros (θ, τ) , é chamada de *família de locação-escala* se:

$$f_X(x; \theta, \tau) = \frac{1}{\tau} h_X\left(\frac{x - \theta}{\tau}\right),$$

em que θ e τ são respectivamente os parâmetros de locação e de escala, definidos

para $\theta \in \mathbb{R}$ e $\tau > 0$.

Se $\mathbf{X} = (X_1, \dots, X_n)$ é um vetor de observações pertencente a família de distribuições \mathcal{F} , então \mathbf{X} tem densidade conjunta dada por:

$$f_{\mathbf{X}}(\mathbf{x}; \theta, \tau) = \frac{1}{\tau^n} h_{\mathbf{X}} \left(\frac{x_1 - \theta}{\tau}, \dots, \frac{x_n - \theta}{\tau} \right). \quad (2.15)$$

O grupo a ser considerado para famílias de locação-escala é o grupo $G = \{(a, b), a \in \mathbb{R} \text{ e } b > 0\}$, definido com a operação $(a, b) \cdot (c, d) = (a + bc, bd)$. Esse grupo também pode ser descrito como o grupo de funções lineares munidas da operação de composição de funções, $G = \{f(x) = a + bx, a \in \mathbb{R} \text{ e } b > 0\}$. Observe que esse grupo não é abeliano.

O grupo G atua no espaço de dados da forma $(a, b) \cdot (X_1, \dots, X_n) = (a + bX_1, \dots, a + bX_n)$. Considerando o vetor de parâmetros (θ, τ) , a ação do grupo G no espaço paramétrico ocorre da seguinte forma:

$$(a, b) \cdot (\theta, \tau) = (a + b\theta, b\tau).$$

Definidas as ações do grupo G sobre o espaço de dados e o espaço paramétrico, pode-se agora fornecer a definição de equivariância de locação-escala para um estimador (δ_1, δ_2) de (θ, τ^r) . Se δ_1 é um estimador de θ e δ_2 é um estimador de τ^r , diremos que (δ_1, δ_2) é equivariante pelas ações de G , sendo (δ_1, δ_2) denominado estimador equivariante por locação-escala.

Definição 2.6.2 (Estimador equivariante por locação-escala): Um estimador $\delta = (\delta_1, \delta_2)$ de (θ, τ^r) é *equivariante por locação-escala* se:

$$\begin{aligned} (a, b) \cdot \delta(\mathbf{X}) &= \delta((a, b) \cdot \mathbf{X}) \\ &= (\delta_1(\mathbf{a} + b\mathbf{X}), \delta_2(\mathbf{a} + b\mathbf{X})) \\ &= (a + b\delta_1(\mathbf{X}), b^r \delta_2(\mathbf{X})). \end{aligned} \quad (2.16)$$

Se $G_1 \subset G$ é o subgrupo definido por $G_1 = \{(a, 1), a \in \mathbb{R}\}$, então δ_1 é um estimador equivariante por locação para θ em relação a ação de G_1 . Da mesma forma, se $G_2 \subset G$ é o subgrupo definido por $G_2 = \{(0, b), b > 0\}$, então δ_2 é um estimador equivariante por escala de τ^r . Nesse sentido, pode-se afirmar que δ_1 e δ_2 são estimadores equivariantes marginais por locação e por escala de θ e τ^r , respectivamente, conforme afirmam Prabakaran e Chandrasekar (1994).

Exemplo 2.6.1: Os estimadores usuais média e desvio padrão amostrais satisfazem as condições em (2.16) e, nesse caso, (\bar{X}, S) é equivariante por locação-escala. De fato, se δ_1 é a média amostral e δ_2 é o desvio padrão amostral, então:

$$\begin{aligned} \delta_1(a + bX_1, \dots, a + bX_n) &= \frac{a + bX_1 + \dots + a + bX_n}{n} \\ &= \frac{na}{n} + b \frac{(X_1 + \dots + X_n)}{n} \\ &= a + b\delta_1(X_1, \dots, X_n) \end{aligned}$$

e

$$\begin{aligned} \delta_2(a + bX_1, \dots, a + bX_n) &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a + bX_i - a - b\bar{X})^2} \\ &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n b^2 (X_i - \bar{X})^2} \\ &= \sqrt{b^2 \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \\ &= b \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \\ &= b\delta_2(X_1, \dots, X_n). \end{aligned}$$

Logo, (\bar{X}, S) é equivariante por locação-escala.

Uma vez que o modelo é invariante em relação as ações do grupo, pode-se exigir que a função perda associada ao estimador (δ_1, δ_2) de (θ, τ^r) também o seja. Uma função perda L é invariante sobre as transformações $\mathbf{X}^* = \mathbf{a} + b\mathbf{X}$ se, e somente se:

$$L((\delta_1, \delta_2), (\theta, \tau^r)) = \rho\left(\frac{\delta_1 - \theta}{\tau}, \frac{\delta_2}{\tau^r}\right), \quad (2.17)$$

em que ρ é uma função real não negativa de duas variáveis.

Do fato de grupo de transformações ser transitivo sobre $\mathbb{R} \times \mathbb{R}^+$ e da função perda ser invariante, resulta que a função risco de qualquer estimador equivariante (δ_1, δ_2) é constante nas órbitas, ou seja, não depende de (θ, τ^r) .

Definição 2.6.3 (Função vetorial invariante): Uma função vetorial dada por $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))$ e definida de $\mathbf{u} : \mathbb{R}^n \rightarrow \mathbb{R}^2$, é invariante para um problema de locação-escala se:

$$\mathbf{u}(\mathbf{a} + b\mathbf{x}) = (bu_1(\mathbf{x}), u_2(\mathbf{x})), \mathbf{a} \in \mathbb{R}^n \text{ e } b > 0.$$

De forma análoga às subseções para equivariância por locação e por escala, é necessária uma caracterização dos estimadores equivariantes e das funções invariantes no contexto de locação-escala.

Lema 2.6.1: Um estimador (δ_1, δ_2) é equivariante por locação-escala se, e somente se, para qualquer outro estimador equivariante por locação-escala $(\delta_{01}, \delta_{02})$, existe uma função vetorial $\mathbf{u} = (u_1, u_2)$ tal que:

$$\delta_1(\mathbf{x}) = \delta_{01}(\mathbf{x}) - u_1(\mathbf{x}). \quad (2.18)$$

$$\delta_2(\mathbf{x}) = \frac{\delta_{02}(\mathbf{x})}{u_2(\mathbf{x})}. \quad (2.19)$$

Prova:

Se (δ_1, δ_2) e $(\delta_{01}, \delta_{02})$ são estimadores equivariantes por locação-escala, da relação (2.18) segue que:

$$\begin{aligned} u_1(\mathbf{a} + b\mathbf{x}) &= \delta_{01}(\mathbf{a} + b\mathbf{x}) - \delta_1(\mathbf{a} + b\mathbf{x}) \\ &= a + b\delta_{01}(\mathbf{x}) - a - b\delta_1(\mathbf{x}) \\ &= b(\delta_{01}(\mathbf{x}) - \delta_1(\mathbf{x})) \\ &= bu_1(\mathbf{x}). \end{aligned}$$

Por sua vez, da relação (2.19) tem-se que:

$$u_2(\mathbf{a} + b\mathbf{x}) = \frac{\delta_{02}(\mathbf{a} + b\mathbf{x})}{\delta_2(\mathbf{a} + b\mathbf{x})} = \frac{b^r \delta_{02}(\mathbf{x})}{b^r \delta_2(\mathbf{x})} = \frac{\delta_{02}(\mathbf{x})}{\delta_2(\mathbf{x})} = u_2(\mathbf{x}).$$

Portanto, a função vetorial \mathbf{u} é invariante por locação-escala.

Reciprocamente, seja $(\delta_{01}, \delta_{02})$ um estimador equivariante e \mathbf{u} uma função invariante. Desse modo, com base nas igualdades em (2.18) e em (2.19) tem-se que:

$$\begin{aligned} \delta_1(\mathbf{a} + b\mathbf{x}) &= \delta_{01}(\mathbf{a} + b\mathbf{x}) - u_1(\mathbf{a} + b\mathbf{x}) \\ &= a + b\delta_{01}(\mathbf{x}) - bu_1(\mathbf{x}) \\ &= a + b(\delta_{01}(\mathbf{x}) - u_1(\mathbf{x})) \\ &= a + b\delta_1(\mathbf{x}) \end{aligned}$$

e

$$\delta_2(\mathbf{a} + b\mathbf{x}) = \frac{\delta_{02}(\mathbf{a} + b\mathbf{x})}{u_2(\mathbf{a} + b\mathbf{x})} = \frac{b^r \delta_{02}(\mathbf{x})}{u_2(\mathbf{x})} = b^r \frac{\delta_{02}(\mathbf{x})}{u_2(\mathbf{x})} = b^r \delta_2(\mathbf{x}).$$

Logo, (δ_1, δ_2) é equivariante por locação-escala.

■

Lema 2.6.2: Uma função vetorial $\mathbf{u} = (u_1, u_2)$ é invariante para o problema de locação-escala se, e somente se:

$$u_1(\mathbf{x}) = g(\mathbf{x}) w_1(z_1, \dots, z_{n-1}). \quad (2.20)$$

$$u_2(\mathbf{x}) = w_2(z_1, \dots, z_{n-1}), \quad (2.21)$$

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

Prova:

Pode-se verificar que, g como está definida, é equivariante por escala para τ . Além disso, pode-se observar também que os z_i 's como construídos são invariantes para as transformações $\mathbf{X}^* = \mathbf{a} + b\mathbf{X}$. De fato,

$$z_i(\mathbf{a} + b\mathbf{x}) = \frac{(a + bx_i - a - bx_n)}{g(\mathbf{a} + b\mathbf{x})} = \frac{(bx_i - bx_n)}{bg(\mathbf{x})} = \frac{b(x_i - x_n)}{bg(\mathbf{x})} = z_i(\mathbf{x}).$$

Agora, assuma por hipótese que são válidas as igualdades em (2.20) e (2.21) e que $\mathbf{u} = (u_1, u_2)$. Provemos que \mathbf{u} é invariante por locação-escala. Com efeito, segue que:

$$\begin{aligned}
u_1(\mathbf{a} + b\mathbf{x}) &= g(\mathbf{a} + b\mathbf{x}) w_1 \left[\frac{(a + bx_1 - a - bx_n)}{g(\mathbf{a} + b\mathbf{x})}, \dots, \frac{(a + bx_{n-1} - a - bx_n)}{g(\mathbf{a} + b\mathbf{x})} \right] \\
&= bg(\mathbf{x}) w_1 \left[\frac{b(x_1 - x_n)}{bg(\mathbf{x})}, \dots, \frac{b(x_{n-1} - x_n)}{bg(\mathbf{x})} \right] \\
&= bg(\mathbf{x}) w_1 \left[\frac{x_1 - x_n}{g(\mathbf{x})}, \dots, \frac{x_{n-1} - x_n}{g(\mathbf{x})} \right] \\
&= bg(\mathbf{x}) w_1(z_1, \dots, z_{n-1}) \\
&= bu_1(\mathbf{x})
\end{aligned}$$

e

$$\begin{aligned}
u_2(\mathbf{a} + b\mathbf{x}) &= w_2 \left[\frac{(a + bx_1 - a - bx_n)}{g(\mathbf{a} + b\mathbf{x})}, \dots, \frac{(a + bx_{n-1} - a - bx_n)}{g(\mathbf{a} + b\mathbf{x})} \right] \\
&= w_2 \left[\frac{b(x_1 - x_n)}{bg(\mathbf{x})}, \dots, \frac{b(x_{n-1} - x_n)}{bg(\mathbf{x})} \right] \\
&= w_2(z_1, \dots, z_{n-1}) \\
&= u_2(\mathbf{x}).
\end{aligned}$$

Portanto, $\mathbf{u}(\mathbf{a} + b\mathbf{x}) = (bu_1(\mathbf{x}), u_2(\mathbf{x}))$, isto é, \mathbf{u} é invariante por locação-escala, o que conclui a demonstração.

■

Teorema 2.6.1: Seja $(\delta_{01}, \delta_{02})$ um estimador equivariante de (θ, τ^r) . O estimador (δ_1, δ_2) é equivariante por locação-escala se, e somente se:

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

$$\delta_2(\mathbf{x}) = \frac{\delta_{02}(\mathbf{x})}{w_2(z_1, \dots, z_{n-1})},$$

para funções w_1 e w_2 .

Prova: Este resultado segue diretamente dos Lemas 2.6.1 e 2.6.2.

Para se obter o estimador MRE na classe D dos estimadores equivariantes de (θ, τ^r) é necessária a especificação de uma função perda. Considere funções perda dadas por formas quadráticas. Seja $L[(\delta_1, \delta_2), (\theta, \tau^r)]$ essa função perda, sendo dada por:

$$a_{11} \left(\frac{\delta_1 - \theta}{\tau} \right)^2 + 2a_{12} \left(\frac{\delta_1 - \theta}{\tau} \right) \left(\frac{\delta_2}{\tau^r} - 1 \right) + a_{22} \left(\frac{\delta_2}{\tau^r} - 1 \right)^2. \quad (2.22)$$

O estimador que possui risco mínimo para essa função perda é denominado Q_A -MRE, em que a matriz $A = (a_{ij})$ é uma matriz 2×2 simétrica positiva definida.

Teorema 2.6.2: Seja $\mathbf{X} = (X_1, \dots, X_n)$ distribuído conforme (2.15). Se existe um estimador equivariante $(\delta_{01}, \delta_{02})$ com risco finito, então para cada $\mathbf{z} = (z_1, \dots, z_{n-1})$, existe uma função vetorial $w(\mathbf{z}) = w^*(\mathbf{z})$ que minimiza:

$$E \left[\left\{ \delta_{01}(\mathbf{X}) - g(\mathbf{X}) w_1(\mathbf{z}), \frac{\delta_{02}(\mathbf{X})}{w_2(\mathbf{z})} \right\} | \mathbf{z} \right],$$

em que E representa a esperança de (θ, τ^r) , sendo calculada para $(\theta = 0, \tau = 1)$.

Então, o estimador MRE de (θ, τ^r) existe e é dado por:

$$\delta_1^*(\mathbf{X}) = \delta_{01}(\mathbf{X}) - g(\mathbf{X}) w_1^*(\mathbf{Z}). \quad (2.23)$$

$$\delta_2^*(\mathbf{X}) = \frac{\delta_{02}(\mathbf{X})}{w_2^*(\mathbf{Z})}. \quad (2.24)$$

Para se obter as expressões para δ_1^* e δ_2^* , considerando a função perda em (2.22), é necessário minimizar a função risco dessa função. Pelo teorema 2.6.2 é possível notar que o problema de minimização da função risco se reduz ao problema de determinar w_1^* e w_2^* que minimizem (2.22).

De fato, considerando as expressões (2.23), (2.24) e que $(\theta = 0, \tau = 1)$, o risco de (2.22) é uma função h de w_1 e w_2 , dada por:

$$h(w_1, w_2) = a_{11}E_{0,1} \left[(\delta_{01} - gw_1)^2 | \mathbf{z} \right] + 2a_{12}E_{0,1} \left[(\delta_{01} - gw_1) \left(\frac{\delta_{02}}{w_2} - 1 \right) | \mathbf{z} \right] \\ + a_{22}E_{0,1} \left[\left(\frac{\delta_{02}}{w_2} - 1 \right)^2 | \mathbf{z} \right].$$

Utilizando cálculo diferencial e a regra de Cramer, obtém-se que:

$$w_1^* = \frac{\begin{bmatrix} a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [g\delta_{01} | \mathbf{z}] - a_{12}^2 E_{0,1} [g\delta_{02} | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] \\ -a_{12}a_{22} \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [g | \mathbf{z}] - E_{0,1} [g\delta_{02} | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] \} \end{bmatrix}}{a_{11}a_{22}E_{0,1} [g^2 | \mathbf{z}] E_{0,1} [\delta_{02}^2 | \mathbf{z}] - a_{12}^2 \{ E_{0,1} [g\delta_{02} | \mathbf{z}] \}^2} \quad (2.25)$$

e

$$\frac{1}{w_2^*} = \frac{\begin{bmatrix} -a_{11}a_{12} \{ E_{0,1} [g^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] - E_{0,1} [g\delta_{02} | \mathbf{z}] E_{0,1} [g\delta_{01} | \mathbf{z}] \} \\ +a_{11}a_{22}E_{0,1} [g^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] - a_{12}^2 E_{0,1} [g\delta_{02} | \mathbf{z}] E_{0,1} [g | \mathbf{z}] \end{bmatrix}}{a_{11}a_{22}E_{0,1} [g^2 | \mathbf{z}] E_{0,1} [\delta_{02}^2 | \mathbf{z}] - a_{12}^2 \{ E_{0,1} [g\delta_{02} | \mathbf{z}] \}^2} \quad (2.26)$$

Para a escolha apropriada da matriz A , o estimador de (θ, τ^r) é dado por (2.23) e (2.24), em que w_1^* e w_2^* são obtidos respectivamente por (2.25) e (2.26). Prabakaran e Chandrasekar (1994) mostram que o estimador MRE por locação-

escala de (θ, τ^r) não depende da escolha da matriz A . Para isso, tomando $g = \delta_{02}$ tem-se que:

$$\begin{aligned}
w_1^* &= \frac{\left[\begin{array}{l} a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] - a_{12}^2 E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] \\ -a_{12}a_{22} \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] - E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] \} \end{array} \right]}{a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02}^2 | \mathbf{z}] - a_{12}^2 \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] - a_{12}^2 E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}]}{a_{11}a_{22} \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2 - a_{12}^2 \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{(a_{11}a_{22} - a_{12}^2) E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}]}{(a_{11}a_{22} - a_{12}^2) \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}]}{\{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}]}{E_{0,1} [\delta_{02}^2 | \mathbf{z}]} \tag{2.27}
\end{aligned}$$

e

$$\begin{aligned}
\frac{1}{w_2^*} &= \frac{\left[\begin{array}{l} -a_{11}a_{12} \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] - E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{01}\delta_{02} | \mathbf{z}] \} \\ +a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] - a_{12}^2 E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] \end{array} \right]}{a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02}^2 | \mathbf{z}] - a_{12}^2 \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{a_{11}a_{22}E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}] - a_{12}^2 E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}]}{a_{11}a_{22} \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2 - a_{12}^2 \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{(a_{11}a_{22} - a_{12}^2) E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}]}{(a_{11}a_{22} - a_{12}^2) \{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{E_{0,1} [\delta_{02}^2 | \mathbf{z}] E_{0,1} [\delta_{02} | \mathbf{z}]}{\{ E_{0,1} [\delta_{02}^2 | \mathbf{z}] \}^2} \\
&= \frac{E_{0,1} [\delta_{02} | \mathbf{z}]}{E_{0,1} [\delta_{02}^2 | \mathbf{z}]} \tag{2.28}
\end{aligned}$$

Substituindo os resultados de (2.27) e (2.28) em (2.23) e (2.24), respectivamente, obtém-se o estimador MRE por locação-escala no processo de estimação de (θ, τ^r) , que independe da escolha da matriz A , sendo dado por:

$$\delta_1^*(\mathbf{X}) = \delta_{01}(\mathbf{X}) - \delta_{02}(\mathbf{X}) \frac{E_{0,1}[\delta_{01}(\mathbf{X}) \delta_{02}(\mathbf{X}) | \mathbf{z}]}{E_{0,1}[\delta_{02}^2(\mathbf{X}) | \mathbf{z}]}. \quad (2.29)$$

$$\delta_2^*(\mathbf{X}) = \delta_{02}(\mathbf{X}) \frac{E_{0,1}[\delta_{02}(\mathbf{X}) | \mathbf{z}]}{E_{0,1}[\delta_{02}^2(\mathbf{X}) | \mathbf{z}]}. \quad (2.30)$$

Exemplo 2.6.2: Seja (X_1, \dots, X_n) uma amostra aleatória de uma densidade Exponencial com parâmetro de locação θ e parâmetro de escala τ . Considere o problema de estimação de (θ, τ) . Seja:

$$(\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X})) = \left(X_{(1)}, \sum_{i=2}^n (X_{(i)} - X_{(1)}) \right).$$

O estimador $(\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X}))$, assim definido, é equivariante por locação-escala. De fato,

$$\delta_{01}(\mathbf{a} + b\mathbf{X}) = a + bX_{(1)} = a + b\delta_{01}(\mathbf{X})$$

e

$$\begin{aligned} \delta_{02}(\mathbf{a} + b\mathbf{X}) &= \sum_{i=2}^n [a + bX_{(i)} - (a + bX_{(1)})] \\ &= \sum_{i=2}^n [bX_{(i)} - bX_{(1)}] \\ &= b \sum_{i=2}^n (X_{(i)} - X_{(1)}) \\ &= b\delta_{02}(\mathbf{X}). \end{aligned}$$

Além disso, o estimador $(\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X}))$ é uma estatística suficiente e completa de (θ, τ) (LEHMANN; CASELLA, 1998, p. 43). Como \mathbf{Z} é uma estatística ancilar, pelo teorema de Basu $(\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X}))$ é independente de \mathbf{Z} . Segue também que os estimadores δ_{01} e δ_{02} são independentes entre si (PRABAKARAN; CHANDRASEKAR, 1994).

Conforme Lehmann e Casella (1998, p. 43), as estatísticas δ_{01} e δ_{02} são distribuídas respectivamente como uma Exponencial $(a, \frac{b}{n})$ e uma $\frac{1}{2}b\chi_{(2n-2)}^2$. Dessa maneira, para $(a = 0, b = 1)$ tem-se que $\delta_{01} \sim \text{Exponencial}(0, \frac{1}{n})$ e $\delta_{02} \sim \frac{1}{2}\chi_{(2n-2)}^2$. Com essas informações é possível calcular as esperanças de δ_{01} e δ_{02} . Para facilidade de cálculo, denote $Y = \delta_{01}$. Assim,

$$\begin{aligned} E_{0,1}[\delta_{01}] &= E_{0,1}[Y] = \int_0^{\infty} \frac{1}{1/n} y \exp\left(-\frac{y-0}{1/n}\right) dy \\ &= \int_0^{\infty} ny \exp(-ny) dy \\ &= -ny \frac{\exp(-ny)}{n} \Big|_0^{\infty} - \int_0^{\infty} -\frac{\exp(-ny)}{n} n dy \\ &= \int_0^{\infty} \exp(-ny) dy \\ &= \frac{1}{n}. \end{aligned}$$

Como $\delta_{02} \sim \frac{1}{2}\chi_{(2n-2)}^2$, resulta que:

$$E_{0,1}[\delta_{02}] = E_{0,1}\left[\frac{1}{2}\chi_{(2n-2)}^2\right] = \frac{1}{2}E_{0,1}\left[\chi_{(2n-2)}^2\right] = \frac{1}{2}(2n-2) = n-1.$$

Uma vez que $\text{var}[\delta_{02}] = E[\delta_{02}^2] - (E[\delta_{02}])^2$, então:

$$\begin{aligned}
E_{0,1} [\delta_{02}^2] &= \text{var}_{0,1} \left[\frac{1}{2} \chi_{(2n-2)}^2 \right] + \left(E_{0,1} \left[\frac{1}{2} \chi_{(2n-2)}^2 \right] \right)^2 \\
&= \frac{1}{4} \text{var}_{0,1} [\chi_{(2n-2)}^2] + \frac{1}{4} \left(E_{0,1} [\chi_{(2n-2)}^2] \right)^2 \\
&= \frac{1}{4} 2(2n-2) + \frac{1}{4} (2n-2)^2 \\
&= (n-1) + (n-1)^2 \\
&= n(n-1).
\end{aligned}$$

Logo, $\delta_1^*(\mathbf{X})$ e $\delta_2^*(\mathbf{X})$ são respectivamente:

$$\begin{aligned}
\delta_1^*(\mathbf{X}) &= \delta_{01}(\mathbf{X}) - \delta_{02}(\mathbf{X}) \frac{E_{0,1} [\delta_{01}(\mathbf{X}) \delta_{02}(\mathbf{X}) | \mathbf{z}]}{E_{0,1} [\delta_{02}^2(\mathbf{X}) | \mathbf{z}]} \\
&= \delta_{01}(\mathbf{X}) - \delta_{02}(\mathbf{X}) \frac{E_{0,1} [\delta_{01}(\mathbf{X}) \delta_{02}(\mathbf{X})]}{E_{0,1} [\delta_{02}^2(\mathbf{X})]} \\
&= \delta_{01}(\mathbf{X}) - \delta_{02}(\mathbf{X}) \frac{E_{0,1} [\delta_{01}(\mathbf{X})] E_{0,1} [\delta_{02}(\mathbf{X})]}{E_{0,1} [\delta_{02}^2(\mathbf{X})]} \\
&= X_{(1)} - \sum_{i=2}^n (X_{(i)} - X_{(1)}) \frac{\frac{1}{n} (n-1)}{n(n-1)} \\
&= X_{(1)} - \frac{1}{n^2} \sum_{i=2}^n (X_{(i)} - X_{(1)}).
\end{aligned}$$

$$\begin{aligned}
\delta_2^*(\mathbf{X}) &= \delta_{02}(\mathbf{X}) \frac{E_{0,1} [\delta_{02}(\mathbf{X}) | \mathbf{z}]}{E_{0,1} [\delta_{02}^2(\mathbf{X}) | \mathbf{z}]} \\
&= \delta_{02}(\mathbf{X}) \frac{E_{0,1} [\delta_{02}(\mathbf{X})]}{E_{0,1} [\delta_{02}^2(\mathbf{X})]} \\
&= \sum_{i=2}^n (X_{(i)} - X_{(1)}) \frac{(n-1)}{n(n-1)} \\
&= \frac{1}{n} \sum_{i=2}^n (X_{(i)} - X_{(1)}).
\end{aligned}$$

Portanto, o estimador MRE (δ_1^*, δ_2^*) de (θ, τ) é:

$$(\delta_1^*, \delta_2^*) = \left(X_{(1)} - \frac{1}{n^2} \sum_{i=2}^n (X_{(i)} - X_{(1)}), \frac{1}{n} \sum_{i=2}^n (X_{(i)} - X_{(1)}) \right).$$

3 ESTIMADORES DE PITMAN NO CONTEXTO BAYESIANO

Em muitos problemas estatísticos verifica-se a ocorrência de simetrias. Uma forma de tratar em uma análise essas simetrias é supor que um grupo, que modela tais simetrias, atua em todo o processo decisório. Como já mencionado, o grupo atua no espaço paramétrico, no espaço amostral e no espaço das estimativas. Nesse contexto, tem-se a teoria da estimação equivariante. Estimadores clássicos nessa teoria, para parâmetros de locação e de escala, são os estimadores de Pitman. A teoria bayesiana também pode ser utilizada no estudo de simetrias, e nesse caso, obtém-se a chamada teoria bayesiana equivariante, que surge com a definição de prioris invariantes. Nesta subseção, pretende-se revisar resultados clássicos da teoria bayesiana e mostrar que os estimadores de Pitman, para parâmetros de locação e de escala, podem ser obtidos como estimadores de Bayes generalizados utilizando-se prioris adequadas. Algumas definições básicas da estimação bayesiana utilizadas nessa seção, como distribuições a priori e a posteriori, risco de Bayes e estimador de Bayes, encontram-se no Apêndice B.

Neste trabalho, assim como a maioria dos textos estatísticos, a teoria da medida será tratada de forma matematicamente imprecisa. Se π é uma densidade de probabilidade em um conjunto Θ , então a medida μ de um subconjunto $A \subset \Theta$ é a probabilidade de ocorrência de A , isto é, $\mu(A) = \int_A \pi(x) dx$, sendo denotada por $P_\pi(A)$. Se π não é integrável, ou seja, $\int_\Theta \pi(x) dx = \infty$, então π é denominada de densidade de probabilidade imprópria e a medida de A é definida da mesma forma:

$$\mu(A) = \int_A \pi(x) dx = P_\pi(A).$$

Como exemplo básico de uma densidade de probabilidade imprópria considere $\pi(x) \equiv 1$ em $\Theta = \mathbb{R}$. Neste caso, se A é um intervalo $[a, b]$, então

$P_\pi(A) = b - a$, sendo denominada a medida de Lebesgue em \mathbb{R} .

Antes de se demonstrar que os estimadores de Pitman podem ser obtidos como estimadores de Bayes generalizados, serão apresentados alguns resultados da estatística bayesiana.

Teorema 3.1: Suponha que Θ tenha distribuição a priori π , e dado $\Theta = \theta$, suponha que X tenha distribuição $f_{X|\Theta}(x|\theta)$. Suponha, em adição, que as seguintes pressuposições para o problema de se estimar $\omega(\Theta)$ com uma função perda $L(\delta, \theta)$ não-negativa:

- (i) Existe um estimador δ_0 com risco finito.
- (ii) Para quase todo x , existe um valor $\delta_\pi(x)$ que minimiza

$$E_\theta [L(\delta(x), \Theta) | X = x]. \quad (3.1)$$

Então, $\delta_\pi(X)$ é um estimador de Bayes.

Prova:

Seja δ_0 um estimador qualquer que possua risco finito. Uma vez que δ_0 possui risco finito e L é uma função não negativa, segue que (3.1) é finito. Consequentemente, como $\delta_\pi(x)$ minimiza (3.1), então:

$$E_\theta [L(\delta_0(x), \Theta) | X = x] \geq E_\theta [L(\delta_\pi(x), \Theta) | X = x], \quad (3.2)$$

para todo estimador δ_0 .

Reescrevendo a expressão (3.2) em termos de integrais tem-se que:

$$\int_{\mathfrak{X}^n} L(\delta_0(x), \theta) f_{X|\Theta}(x|\theta) dx \geq \int_{\mathfrak{X}^n} L(\delta_\pi(x), \theta) f_{X|\Theta}(x|\theta) dx. \quad (3.3)$$

Tomando-se a esperança matemática em ambos os lados de (3.3) e considerando uma priori π , verifica-se que o resultado segue. Neste caso, $r(\delta_0) \geq r(\delta_\pi)$, ou seja, o estimador δ_π possui risco de Bayes menor ou igual em relação a qualquer outro estimador δ_0 . Portanto, δ_π é o estimador de Bayes. ■

Corolário 3.1 Suponha que as hipóteses do teorema 3.1 são válidas.

(i) Se $L(\delta, \theta) = (\delta - \omega(\theta))^2$, então:

$$\delta_\pi(x) = E_\Theta[\omega(\Theta) | x].$$

De forma mais geral, se $L(\delta, \theta) = h(\theta)(\delta - \omega(\theta))^2$, então:

$$\delta_\pi(x) = \frac{\int h(\theta) \omega(\theta) f_{X|\Theta}(x|\theta) \pi(\theta) d\theta}{\int h(\theta) f_{X|\Theta}(x|\theta) \pi(\theta) d\theta} = \frac{E_\Theta[h(\Theta) \omega(\Theta) | x]}{E_\Theta[h(\Theta) | x]}.$$

(ii) Se $L(\delta, \theta) = |\delta - \omega(\theta)|$, então δ_π é qualquer mediana da distribuição condicional de Θ dado x .

(iii) Se

$$L(\delta, \theta) = \begin{cases} 0, & \text{quando } |\delta - \theta| \leq c \\ 1, & \text{quando } |\delta - \theta| > c \end{cases},$$

então, δ_π é o ponto médio no intervalo I de tamanho $2c$ que maximiza $P[\Theta \in I | x]$.

Prova:

Serão apresentadas somente as provas dos casos i) e ii). Pode-se verificar nas condições do teorema 3.1 que a função perda L não foi especificada. Para facilitar as demonstrações dos casos i) e ii), primeiramente o risco de Bayes será desenvolvido. Considerando uma amostra aleatória (X_1, \dots, X_n) , o risco de Bayes de um estimador δ pode ser expresso como:

$$\begin{aligned}
 r(\delta) &= \int_{\Theta} \mathcal{R}(\delta, \theta) \pi(\theta) d\theta \\
 &= \int_{\Theta} E_{\theta} [L(\delta(\mathbf{X}), \omega(\theta))] \pi(\theta) d\theta \\
 &= \int_{\Theta} \left\{ \int_{\mathfrak{X}^n} L(\delta(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \prod_{i=1}^n dx_i \right\} \pi(\theta) d\theta \\
 &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta) d\theta \right\} \prod_{i=1}^n dx_i \\
 &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} f_{\mathbf{X}}(\mathbf{x}) d\theta \right\} \prod_{i=1}^n dx_i \\
 &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\theta \right\} \prod_{i=1}^n dx_i \\
 &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \right\} f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^n dx_i.
 \end{aligned}$$

Uma vez que o integrando é não negativo, a integral dupla é minimizada quando a integral interna é minimizada para cada amostra observada (x_1, \dots, x_n) .

Caso i) Considere primeiramente a função perda em erro quadrático. Assim, para obter o valor mínimo do risco de Bayes basta minimizar a seguinte integral em relação a função perda em erro quadrático.

$$\int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = \int_{\Theta} (\delta(\mathbf{x}) - \omega(\theta))^2 \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Desenvolvendo o quadrado vem que:

$$(\delta(\mathbf{x}))^2 \int_{\Theta} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta - 2\delta(\mathbf{x}) \int_{\Theta} \omega(\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta + \int_{\Theta} (\omega(\theta))^2 \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Derivando-se em relação a $\delta(\mathbf{x})$ e igualando a derivada a zero tem-se:

$$2\delta(\mathbf{x}) \int_{\Theta} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta - 2 \int_{\Theta} \omega(\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = 0.$$

A solução dessa equação é:

$$\delta(\mathbf{x}) = \frac{\int_{\Theta} \omega(\theta) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta}{\int_{\Theta} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta} = E_{\theta}[\omega(\Theta)|\mathbf{x}].$$

Essa solução é o valor mínimo, uma vez que a derivada segunda em relação a $\delta(\mathbf{x})$ é $2 \int_{\Theta} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta > 0$.

Caso ii) Considerando a função perda em erro absoluto, segue que a minimização do risco de Bayes ocorre para a integral,

$$\int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = \int_{\Theta} |\delta(\mathbf{x}) - \omega(\theta)| \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Como em problemas de estimação é comum que $\Theta = \mathbb{R}$, pela definição de função modular segue que a última expressão é equivalente a:

$$\int_{-\infty}^{\delta(\mathbf{x})} -(\delta(\mathbf{x}) - \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta + \int_{\delta(\mathbf{x})}^{+\infty} (\delta(\mathbf{x}) - \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Então, derivando em relação a $\delta(\mathbf{x})$ e igualando a derivada a zero resulta que:

$$-\int_{-\infty}^{\delta(\mathbf{x})} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta + \int_{\delta(\mathbf{x})}^{+\infty} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = 0.$$

Resolvendo, obtém-se que:

$$\int_{-\infty}^{\delta(\mathbf{x})} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta = \int_{\delta(\mathbf{x})}^{+\infty} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Assim, $\delta(\mathbf{x})$ deve ser o valor tal que $P[\Theta \leq \delta(\mathbf{x})] = P[\Theta \geq \delta(\mathbf{x})]$. Como Θ é uma variável contínua, $\delta(\mathbf{x})$ satisfaz $P[\Theta \leq \delta(\mathbf{x})] = P[\Theta \geq \delta(\mathbf{x})] = \frac{1}{2}$. Logo, $\delta(\mathbf{x})$ deve ser a mediana da distribuição de Θ .

■

Corolário 3.2: Considere que $\Theta \sim \pi(\theta)$ e que $X \sim f_{X|\Theta}(x|\theta)$. Se $L(\delta, \theta)$ é uma função perda estritamente convexa em relação a δ , em particular, uma função perda em erro quadrático, e são satisfeitas as seguintes hipóteses:

- (i) Existe um estimador de Bayes δ_π .
- (ii) Os conjuntos de medida nula em relação à medida definida pela marginal:

$$Q(A) = \int_{\Theta} \left(\int_A f_{X|\Theta}(x|\theta) dx \right) \pi(\theta) d\theta,$$

coincidem com os conjuntos de medida nula relativos às medidas definidas por $f_{X|\Theta}(x|\theta) dx$.

Então, o estimador de Bayes δ_π é único.

Prova:

Suponha que δ_1 e δ_2 sejam ambos estimadores de Bayes, e nesse caso, $r(\delta_1) = r(\delta_2)$. Além disso, considere que $\delta = t\delta_1 + (1-t)\delta_2$, para $\delta_1(x) < \delta_2(x)$. Como L é estritamente convexa, então:

$$L(\delta(x), \theta) = L(t\delta_1(x) + (1-t)\delta_2(x), \theta) < tL(\delta_1(x), \theta) + (1-t)L(\delta_2(x), \theta),$$

para $t \in (0, 1)$.

Assim,

$$\begin{aligned} r(\delta) &= r(t\delta_1 + (1-t)\delta_2) \\ &= \int_{\Theta} \mathcal{R}(t\delta_1 + (1-t)\delta_2, \theta) \pi(\theta) d\theta \\ &= \int_{\Theta} E_{\theta}[L(t\delta_1(x) + (1-t)\delta_2(x), \theta)] \pi(\theta) d\theta \\ &= \int_{\Theta} \left(\int_{\mathbb{X}^n} L(t\delta_1(x) + (1-t)\delta_2(x), \theta) f_{X|\Theta}(x|\theta) dx \right) \pi(\theta) d\theta \\ &< \int_{\Theta} \left(\int_{\mathbb{X}^n} tL(\delta_1(x), \theta) + (1-t)L(\delta_2(x), \theta) f_{X|\Theta}(x|\theta) dx \right) \pi(\theta) d\theta \\ &= t \int_{\Theta} E_{\theta}[L(\delta_1(x), \theta)] \pi(\theta) d\theta + (1-t) \int_{\Theta} E_{\theta}[L(\delta_2(x), \theta)] \pi(\theta) d\theta \\ &= t \int_{\Theta} \mathcal{R}(\delta_1, \theta) \pi(\theta) d\theta + (1-t) \int_{\Theta} \mathcal{R}(\delta_2, \theta) \pi(\theta) d\theta \\ &= tr(\delta_1) + (1-t)r(\delta_2) \\ &= tr(\delta_1) + r(\delta_2) - tr(\delta_2) \\ &= r(\delta_2). \end{aligned}$$

Absurdo, pois nenhum estimador pode ter risco de Bayes menor que um estimador de Bayes. Logo, a hipótese de que δ_1 e δ_2 são ambos estimadores de Bayes deve ser removida. Portanto, o estimador de Bayes é único.

■

No próximo exemplo é apresentada uma situação em que não se tem a unicidade do estimador de Bayes.

Exemplo 3.1: Considere que $X \sim \text{Binomial}(n, p) \sim f_X(x|n, p)$ e que $\pi(p)$ é uma distribuição de probabilidade a priori que concentra toda probabilidade em $p = 0$ e $p = 1$, isto é, $\pi(0) = a$ e $\pi(1) = b$, com $\pi(p) = 0$ para $p \neq 0, 1$.

Suponha que $X = (0, 1, \dots, n)$ é observado e que $L(\delta, p) = (\delta - p)^2$. Como X é uma variável aleatória discreta, a função risco é dada por:

$$\begin{aligned} \mathcal{R}(\delta, p) &= E_p[L(\delta(x), p)] \\ &= \int (\delta(x) - p)^2 f_X(x|n, p) dx \\ &= \sum_{x=0}^n (\delta(x) - p)^2 f_X(x|n, p). \end{aligned}$$

Para $x = 0$ e $p = 0$, tem-se em relação a $f_X(x|n, p)$ que:

$$f_X(0|n, 0) = \binom{n}{0} 0^0 (1-0)^{n-0} = 1$$

e

$$f_X(x|n, 0) = \binom{n}{x} 0^x (1-0)^{n-x} = 0, \text{ para } x \neq 0.$$

Para $x = n$ e $p = 1$, tem-se que:

$$f_X(n|n, 1) = \binom{n}{n} 1^n (1-1)^{n-n} = 1$$

e

$$f_X(x|n, 1) = \binom{n}{x} 1^x (1-1)^{n-x} = 0, \text{ para } x \neq n.$$

Dessa forma, o risco de Bayes de δ é dado por:

$$\begin{aligned} r(\delta) &= \int_0^1 \mathcal{R}(\delta, p) \pi(p) dp \\ &= \int_0^1 \sum_{x=0}^n (\delta(x) - p)^2 f_X(x|n, p) \pi(p) dp \\ &= \sum_{x=0}^n \int_0^1 (\delta(x) - p)^2 f_X(x|n, p) \pi(p) dp \\ &= \sum_{x=0}^n \left[(\delta(x) - 0)^2 f_X(x|n, 0) \pi(0) + (\delta(x) - 1)^2 f_X(x|n, 1) \pi(1) \right] \\ &= \sum_{x=0}^n \left[a(\delta(x) - 0)^2 f_X(x|n, 0) + b(\delta(x) - 1)^2 f_X(x|n, 1) \right] \\ &= a(\delta(0) - 0)^2 + b(\delta(n) - 1)^2. \end{aligned}$$

O risco de Bayes é minimizado e vale zero se $\delta_\pi(0) = 0$ e $\delta_\pi(n) = 1$. Portanto, qualquer estimador com esses valores é um estimador de Bayes. A não unicidade do estimador de Bayes decorre do fato da condição *ii*) do teorema não ser satisfeita. A distribuição marginal de X é dada por:

$$f_X(x) = \int_0^1 f_X(x|n, p) \pi(p) dp = a f_X(x|n, 0) + b f_X(x|n, 1).$$

Assim,

$$f_X(x) = a f_X(x|n, 0) + b f_X(x|n, 1) = \begin{cases} a, & \text{se } x = 0 \\ b, & \text{se } x = n \\ 0, & \text{se } x \neq 0, n \end{cases}.$$

Lema 3.1 Se δ_π é o estimador de Bayes de $\omega(\theta)$ em relação a uma priori $\pi(\theta)$ e se $r(\delta_\pi) = E_\theta [(\delta_\pi(\mathbf{X}) - \omega(\Theta))^2]$ é o risco de Bayes, então:

$$r(\delta_\pi) = \int_{\mathfrak{X}^n} \text{var}_\theta [\omega(\Theta) | \mathbf{x}] f_{\mathbf{X}}(\mathbf{x}).$$

Em particular, se a variância da posteriori de $\omega(\Theta) | \mathbf{x}$ é independente de \mathbf{x} , então:

$$r(\delta_\pi) = \text{var}_\theta [\omega(\Theta) | \mathbf{x}].$$

Prova:

$$\begin{aligned}
r(\delta_\pi) &= \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta \\
&= \int_{\Theta} \left(\int_{\mathcal{X}^n} L(\delta_\pi(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) d\mathbf{x} \right) \pi(\theta) d\theta \\
&= \int_{\mathcal{X}^n} \left(\int_{\Theta} L(\delta_\pi(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta) d\theta \right) d\mathbf{x} \\
&= \int_{\mathcal{X}^n} \left(\int_{\Theta} L(\delta_\pi(\mathbf{x}), \omega(\theta)) \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} d\theta \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathcal{X}^n} \left(\int_{\Theta} L(\delta_\pi(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathcal{X}^n} \left(\int_{\Theta} (\delta_\pi(\mathbf{x}) - \omega(\theta))^2 \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathcal{X}^n} \text{var}_\theta[\omega(\Theta)|\mathbf{x}] f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

Se $\text{var}[\omega(\Theta)|\mathbf{x}]$ não depende de \mathbf{x} , o risco de Bayes de δ_π é $\text{var}[\omega(\Theta)|\mathbf{x}]$.

■

Um aspecto interessante dos estimadores de Bayes é que, excetuando-se situações muito particulares, os estimadores de Bayes são sempre estimadores viesados.

Teorema 3.2: Se δ_π é o estimador de Bayes de $\omega(\theta)$ em relação a uma função perda em erro quadrático, então δ_π é um estimador viesado, a menos que:

$$E_\theta \left[(\delta_\pi(X) - \omega(\Theta))^2 \right] = 0, \quad (3.4)$$

em que a esperança é tomada em relação a ambas variáveis X e Θ .

Prova:

Desenvolvendo o quadrado, segue que a $E_\theta \left[(\delta_\pi(X) - \omega(\Theta))^2 \right]$ é dada por:

$$E_\theta \left[(\delta_\pi(X))^2 \right] + E_\theta \left[(\omega(\Theta))^2 \right] - 2E_\theta \left[\delta_\pi(X) \omega(\Theta) \right]. \quad (3.5)$$

Primeiramente, notemos que:

$$\begin{aligned} E_\theta \left[(\delta_\pi(X))^2 \right] &= \int_{\Theta} \int_{\mathfrak{X}^n} (\delta_\pi(x))^2 f_{X|\Theta}(x|\theta) \pi(\theta) dx d\theta \\ &= \int_{\Theta} \int_{\mathfrak{X}^n} (\delta_\pi(x))^2 \frac{f_{X|\Theta}(x|\theta) \pi(\theta)}{f_X(x)} f_X(x) dx d\theta \\ &= \int_{\Theta} \int_{\mathfrak{X}^n} (\delta_\pi(x))^2 \pi_{\Theta|X}(\theta|x) f_X(x) dx d\theta \\ &= \int_{\mathfrak{X}^n} (\delta_\pi(x))^2 \left(\int_{\Theta} \pi_{\Theta|X}(\theta|x) d\theta \right) f_X(x) dx \\ &= \int_{\mathfrak{X}^n} (\delta_\pi(x))^2 f_X(x) dx. \end{aligned}$$

Além disso,

$$\begin{aligned} E_\theta \left[(\omega(\Theta))^2 \right] &= \int_{\Theta} \int_{\mathfrak{X}^n} (\omega(\theta))^2 f_{X|\Theta}(x|\theta) \pi(\theta) dx d\theta \\ &= \int_{\Theta} (\omega(\theta))^2 \left(\int_{\mathfrak{X}^n} f_{X|\Theta}(x|\theta) dx \right) \pi(\theta) d\theta \\ &= \int_{\Theta} (\omega(\theta))^2 \pi(\theta) d\theta \end{aligned}$$

e

$$\begin{aligned}
E_{\theta} [\delta_{\pi}(X) \omega(\Theta)] &= \int_{\Theta} \int_{\mathfrak{X}^n} \delta_{\pi}(x) \omega(\theta) f_{X|\Theta}(x|\theta) \pi(\theta) dx d\theta \\
&= \int_{\mathfrak{X}^n} \delta_{\pi}(x) \left(\int_{\Theta} \omega(\theta) \pi_{\Theta|X}(\theta|x) d\theta \right) f_X(x) dx \\
&= \int_{\mathfrak{X}^n} \delta_{\pi}(x) \left(\int_{\Theta} \omega(\theta) \frac{f_{X|\Theta}(x|\theta) \pi(\theta)}{\int_{\Theta} f_{X|\Theta}(x|\theta) \pi(\theta) d\theta} d\theta \right) f_X(x) dx \\
&= \int_{\mathfrak{X}^n} \delta_{\pi}(x) \delta_{\pi}(x) f_X(x) dx \\
&= \int_{\mathfrak{X}^n} (\delta_{\pi}(x))^2 f_X(x) dx.
\end{aligned}$$

Se $\delta_{\pi}(X)$ é um estimador não viesado de $\omega(\theta)$ então:

$$E_{\theta} [\delta_{\pi}(X)] = \omega(\theta) \Rightarrow \int_{\mathfrak{X}^n} \delta_{\pi}(x) f_{X|\Theta}(x|\theta) dx = \omega(\theta).$$

Desse modo,

$$\begin{aligned}
E_{\theta} [\delta_{\pi}(X) \omega(\Theta)] &= \int_{\Theta} \int_{\mathfrak{X}^n} \delta_{\pi}(x) \omega(\theta) f_{X|\Theta}(x|\theta) \pi(\theta) dx d\theta \\
&= \int_{\Theta} \omega(\theta) \left(\int_{\mathfrak{X}^n} \delta_{\pi}(x) f_{X|\Theta}(x|\theta) dx \right) \pi(\theta) d\theta \\
&= \int_{\Theta} \omega(\theta) \omega(\theta) \pi(\theta) d\theta \\
&= \int_{\Theta} (\omega(\theta))^2 \pi(\theta) d\theta.
\end{aligned}$$

Substituindo esses resultados em (3.5), segue que $E_{\theta} [(\delta_{\pi}(X) - \omega(\Theta))^2]$ é identicamente nula.

■

Uma aplicação do teorema 3.2 é responder uma questão interessante rela-

cionada à média amostral. A estatística $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ é um estimador de Bayes em relação a alguma priori π ? A resposta é em geral não. Seja (X_1, \dots, X_n) uma amostra aleatória com $E[X_i] = \theta$ e $\text{var}[X_i] = \sigma^2$, em que σ^2 não depende de θ . Supondo que Θ é uma variável aleatória com distribuição a priori $\pi(\theta)$ qualquer, o risco de \bar{X} dado θ é:

$$\begin{aligned} E_{\theta} [(\bar{X} - \Theta)^2] &= \int_{\mathfrak{X}^n} \int_{\Theta} (\bar{X} - \theta)^2 f_{X|\Theta}(x|\theta) \pi(\theta) d\theta dx \\ &= \int_{\Theta} \text{var}(\bar{X}) \pi(\theta) d\theta \\ &= \frac{\sigma^2}{n} \\ &\neq 0. \end{aligned}$$

Como \bar{X} não satisfaz a condição (3.4), segue pelo teorema 3.2 que \bar{X} não é um estimador de Bayes. Se $\text{var}[X_i]$ é uma função de θ , digamos $v(\theta)$, então:

$$\begin{aligned} E_{\theta} [(\bar{X} - \Theta)^2] &= \int_{\Theta} \int_{\mathfrak{X}^n} (\bar{X} - \theta)^2 f_{X|\Theta}(x|\theta) \pi(\theta) d\theta dx \\ &= \int_{\Theta} v(\theta) \pi(\theta) d\theta \\ &= 0, \end{aligned}$$

para alguma priori π .

No próximo exemplo é apresentada uma situação em que a média amostral é um estimador de Bayes.

Exemplo 3.2: Seja (X_1, \dots, X_n) uma amostra aleatória Bernoulli(p), com $E[X_i] = p$, $\text{var}[X_i] = p(1-p)$. Considere uma priori π que concentra toda probabilidade em $p = 0$ e $p = 1$, ou seja, $\pi(0) = a$, $\pi(1) = b$ e $\pi(p) = 0$ para $p \in (0, 1)$.

Para o estimador média amostral, $\delta \left(\sum_{i=1}^n X_i \right) = \frac{\sum_{i=1}^n X_i}{n}$, tem-se que:

$$\begin{aligned}
 E_p \left[(\bar{X} - p)^2 \right] &= \int_{\Theta} \int_{\mathfrak{X}^n} (\bar{X} - p)^2 f_X(x|p) \pi(p) dx dp \\
 &= \int_{\Theta} \left(\int_{\mathfrak{X}^n} (\bar{X} - p)^2 f_X(x|p) dx \right) \pi(p) dp \\
 &= \int_0^1 \frac{p(1-p)}{n} \pi(p) dp \\
 &= \frac{0(1-0)}{n} \pi(0) + \frac{1(1-1)}{n} \pi(1) \\
 &= 0.
 \end{aligned}$$

Para a priori π que concentra toda probabilidade em $p = 0$ e $p = 1$, \bar{X} satisfaz a condição (3.4) do teorema 3.2. Logo, \bar{X} é um estimador de Bayes. Além disso, como $\sum_{i=1}^n X_i$ é Binomial(n, p), pelo exemplo 3.1 foi possível ver que:

$$\begin{aligned}
 r(\delta) &= \int_0^1 \mathcal{R}(\delta, p) \pi(p) dp \\
 &= \int_0^1 \sum_{x=0}^n (\delta(x) - p)^2 f_X(x|n, p) \pi(p) dp \\
 &= a(\delta(0) - 0)^2 + b(\delta(n) - 1)^2.
 \end{aligned}$$

Assim, qualquer estimador que satisfaça as condições $\delta(0) = 0$ e $\delta(n) = 1$ é um estimador de Bayes. Consequentemente, em particular, \bar{X} é um estimador de Bayes.

Estimadores de Bayes são definidos a partir de distribuições a priori próprias $\pi(\theta)$. Muitas vezes, no entanto, é comum estender essa definição pela utili-

zação de uma priori imprópria $\pi(\theta)$ que não define uma distribuição de probabilidade, isto é,

$$\int_{\Theta} \pi(\theta) d\theta = \infty.$$

No entanto, pode ocorrer que a distribuição a posteriori, considerando como priori uma função imprópria, defina uma distribuição de probabilidade. Nesses casos, o estimador média da posteriori é denominada estimador de Bayes generalizado em relação a uma priori imprópria peso $\pi(\theta)$.

Os estimadores de Pitman para parâmetros de locação e de escala correspondem a estimadores de Bayes generalizados, com respeito a prioris não informativas específicas para cada caso. Considere inicialmente o estimador de Pitman para parâmetros de locação e a priori imprópria não informativa $\pi(\theta) = 1$, em que $\theta \in \mathbb{R}$. A distribuição a posteriori de Θ é dada por:

$$\begin{aligned} \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) &= \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0)\pi(\theta)}{\int_{\Theta} f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0)\pi(\theta) d\theta} \\ &= \frac{f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0)}{\int_{\Theta} f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0) d\theta}. \end{aligned}$$

Então, o estimador de Bayes generalizado é dado por:

$$\begin{aligned} E_{\theta}[\theta|\mathbf{x}] &= \frac{\int_{\Theta} \theta f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0) d\theta}{\int_{\Theta} f_{\mathbf{X}}(x_1 - \theta, \dots, x_n - \theta|0) d\theta} \\ &= \frac{\int_{\Theta} \theta f_{\mathbf{X}}(x_1, \dots, x_n|\theta) d\theta}{\int_{\Theta} f_{\mathbf{X}}(x_1, \dots, x_n|\theta) d\theta} \\ &= \delta(\mathbf{x}), \end{aligned}$$

ou seja, o estimador de Bayes generalizado para o caso em que a priori imprópria

não informativa é $\pi(\theta) = 1$, corresponde ao estimador de Pitman para parâmetros de locação.

Da mesma forma, o estimador de Pitman para o parâmetro τ^r possui uma interessante interpretação em termos do estimador de Bayes generalizado. Utilizando a expressão (2.9) do estimador de Pitman para parâmetros de escala da subseção 2.5.1 e fazendo a mudança de variáveis $\tau = \frac{1}{v}$, $dv = -\tau^{-2}d\tau$, tem-se que:

$$\begin{aligned}
\delta(\mathbf{X}) &= \frac{\int_0^{\infty} v^{n+r-1} f_{\mathbf{X}}(x_1v, \dots, x_nv|1) dv}{\int_0^{\infty} v^{n+2r-1} f_{\mathbf{X}}(x_1v, \dots, x_nv|1) dv} \\
&= \frac{\int_0^{\infty} \tau^{-n-r+1} f_{\mathbf{X}}\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}|1\right) (-\tau^{-2}) d\tau}{\int_0^{\infty} \tau^{-n-2r+1} f_{\mathbf{X}}\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}|1\right) (-\tau^{-2}) d\tau} \\
&= \frac{\int_0^{\infty} \tau^{-r} [f_{\mathbf{X}}\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}|1\right) / \tau^n] \tau^{-1} d\tau}{\int_0^{\infty} \tau^{-2r} [f_{\mathbf{X}}\left(\frac{x_1}{\tau}, \dots, \frac{x_n}{\tau}|1\right) / \tau^n] \tau^{-1} d\tau} \\
&= \frac{\int_0^{\infty} \tau^{-r} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau}{\int_0^{\infty} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau} \times \left(\frac{\int_0^{\infty} \tau^{-2r} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau}{\int_0^{\infty} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau} \right)^{-1} \\
&= \int_0^{\infty} \tau^{-r} \frac{f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau)}{\int_0^{\infty} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau} d\tau \times \left(\int_0^{\infty} \tau^{-2r} \frac{f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau)}{\int_0^{\infty} f_{\mathbf{X}}(\mathbf{x}|\tau) \pi(\tau) d\tau} d\tau \right)^{-1} \\
&= \frac{E_{\tau}[\tau^{-r} | \mathbf{x}]}{E_{\tau}[\tau^{-2r} | \mathbf{x}]},
\end{aligned}$$

em que $\pi(\tau) = \frac{1}{\tau}$ é a priori imprópria.

Para a priori imprópria $\pi(\tau) = \frac{1}{\tau}$ e a função perda quadrática padroni-

zada $L(\delta, \tau) = \frac{(\delta - \tau^r)^2}{\tau^{2r}}$, o estimador de Bayes que minimiza o risco de Bayes a posteriori é dado por:

$$\begin{aligned}
 r(\delta_\pi) &= \int_0^\infty \mathcal{R}(\delta_\pi, \tau) \pi(\tau) d\tau \\
 &= \int_0^\infty \frac{(\delta_\pi(\mathbf{x}) - \tau^r)^2}{\tau^{2r}} \pi(\tau) d\tau \\
 &= \int_0^\infty \frac{(\delta_\pi(\mathbf{x}))^2 - 2\delta_\pi(\mathbf{x})\tau^r + \tau^{2r}}{\tau^{2r}} \pi(\tau) d\tau \\
 &= (\delta_\pi(\mathbf{x}))^2 \int_0^\infty \frac{1}{\tau^{2r}} \pi(\tau) d\tau - 2\delta_\pi(\mathbf{x}) \int_0^\infty \frac{1}{\tau^r} \pi(\tau) d\tau + \int_0^\infty \pi(\tau) d\tau.
 \end{aligned}$$

Derivando em relação a $\delta_\pi(\mathbf{x})$ e igualando a derivada a zero, o risco de Bayes é minimizado para:

$$\delta_\pi(\mathbf{x}) = \frac{\int \frac{1}{\tau^r} f(\tau | \mathbf{x}) d\tau}{\int \frac{1}{\tau^{2r}} f(\tau | \mathbf{x}) d\tau} = \frac{E[\tau^{-r} | \mathbf{x}]}{E[\tau^{-2r} | \mathbf{x}]},$$

uma vez que a derivada segunda é $2 \int_0^\infty \frac{1}{\tau^{2r}} \pi(\tau) d\tau > 0$.

Portanto, o estimador de Bayes generalizado para a priori imprópria $\pi(\tau) = \frac{1}{\tau}$ é igual ao estimador de Pitman para parâmetros de escala.

4 ADMISSIBILIDADE E MINIMAXIMALIDADE DE ESTIMADORES

Definição 4.1 (Estimador admissível): Um estimador δ_A de um parâmetro θ é denominado *admissível* se seu risco é menor ou igual ao risco de qualquer outro estimador δ , com a desigualdade estrita para pelo menos um valor do parâmetro θ .

$$\mathcal{R}(\delta_A, \theta) \leq \mathcal{R}(\delta, \theta), \text{ para algum todo } \theta \in \Theta$$

e

$$\mathcal{R}(\delta_A, \theta_0) < \mathcal{R}(\delta, \theta_0), \text{ para algum } \theta_0 \in \Theta.$$

A propriedade de admissibilidade de um estimador é altamente desejável, pois se um estimador possui essa propriedade, não é possível encontrar outro estimador que possua risco menor ou igual, dado qualquer valor do parâmetro. Os estimadores UMVUE são os estimadores admissíveis quando se restringe o conjunto de todos os estimadores apenas aos estimadores não viesados.

A dificuldade no julgamento de um estimador com base nessa propriedade é que a integração na função risco normalmente é uma função de θ , e, nesse caso, raramente existe um estimador com menor risco uniforme para todos os possíveis valores de θ . Logo, é necessário introduzir algum critério complementar para que a função risco possa ser utilizada como uma ferramenta para determinar se um estimador é bom. Uma vez que é improvável que o gráfico da função $\mathcal{R}(\delta, \theta)$ para uma estimativa esteja abaixo dos gráficos de todas as outras estimativas, um procedimento a ser adotado é observar somente o valor máximo no gráfico da função risco e procurar um estimador que forneça uma estimativa que seja a menor de tais valores. Se existe um estimador com essa característica, esse estimador é

chamado *minimax*.

Definição 4.2 (Estimador *minimax*): Um estimador δ_M de um parâmetro θ é denominado *minimax* se minimiza o risco máximo entre todos os estimadores do parâmetro θ , isto é,

$$\sup_{\theta} \mathcal{R}(\delta_M, \theta) = \inf_{\delta} \sup_{\theta} \mathcal{R}(\delta, \theta).$$

Os estimadores *minimax* são utilizados quando se adota no processo de estimação uma abordagem conservadora, em que se deseja manter sob controle o valor máximo do risco para uma determinada decisão que deve ser tomada. Como quase todo problema de minimização em matemática, obter estimadores *minimax* é um problema difícil e resolvido apenas para algumas famílias de distribuições.

Antes de se abordar o problema de admissibilidade e minimaximalidade dos estimadores de Pitman, serão apresentados alguns resultados essenciais para o cálculo de estimadores admissíveis e *minimax*. Apesar de ambos as definições serem essencialmente conceitos frequentistas, alguns resultados bayesianos são mais adequados para o cálculo de estimadores com essas propriedades.

A propriedade de admissibilidade é certamente desejável para um estimador, mas é uma propriedade em um certo sentido fraca, pois existem estimadores admissíveis que são evidentemente inadequados para qualquer aplicação. Um exemplo com essa característica é apresentado.

Exemplo 4.1: Considere um estimador que despreze toda informação que se possa obter utilizando os dados amostrais. Para uma amostra aleatória (X_1, \dots, X_n) , considere que θ é estimado por 5, isto é, $\delta(X_1, \dots, X_n) = 5$.

O risco quadrático desse estimador é dado por:

$$\mathcal{R}(\delta, \theta) = E_{\theta} [(\delta(\mathbf{X}) - \theta)^2] = E_{\theta} [(5 - \theta)^2] = (5 - \theta)^2.$$

Suponha que δ' é outro estimador, de tal forma que:

$$\mathcal{R}(\delta', \theta) \leq \mathcal{R}(\delta, \theta) = (5 - \theta)^2, \text{ para todo } \theta \in \Theta.$$

Como a desigualdade deve ser satisfeita para todo $\theta \in \Theta$, sem perda de generalidade, suponha que $\theta = 5$. Dessa forma, para o estimador δ vem que:

$$\mathcal{R}(\delta, \theta) = E_{\theta=5} [(\delta(\mathbf{X}) - 5)^2] = E_{\theta=5} [(5 - 5)^2] = 0.$$

Como $\mathcal{R}(\delta', \theta) \leq \mathcal{R}(\delta, \theta)$ para todo $\theta \in \Theta$, isso implica necessariamente que $\mathcal{R}(\delta', \theta) = 0$ em $\theta = 5$. Então, com base no erro quadrático médio do estimador δ' resulta que:

$$\begin{aligned} \mathcal{R}(\delta', \theta) &= E_{\theta=5} [(\delta'(\mathbf{X}) - 5)^2] \\ &= E_{\theta=5} \left[\left(\delta'(\mathbf{X}) - E_{\theta=5} [\delta'(\mathbf{X})] \right)^2 \right] + \left(E_{\theta=5} [\delta'(\mathbf{X})] - 5 \right)^2 \\ &= \text{var}_{\theta=5} [\delta'(\mathbf{X})] + \left(E_{\theta=5} [\delta'(\mathbf{X})] - 5 \right)^2 \\ &= 0. \end{aligned}$$

Para que $\mathcal{R}(\delta', \theta) = 0$, necessariamente tem-se que $\text{var}_{\theta=5} [\delta'(\mathbf{X})] = 0$ e $E_{\theta=5} [\delta'(\mathbf{X})] = 5$. Logo, $\delta'(\mathbf{X}) \equiv 5$, e com isso, conclui-se que $\delta' = \delta$. Portanto, o estimador $\delta(X_1, \dots, X_n) = 5$ é admissível. Esse exemplo ilustra

que, apesar de ser um pobre estimador para outros valores de θ , o estimador δ não pode ser superado em risco quadrático em $\theta = 5$, o que demonstra uma certa limitação da função risco quando se considera a propriedade de admissibilidade de um estimador.

Um outro exemplo interessante é apresentado a seguir e encontra-se em Makani (1977). Esse exemplo ilustra que, dada duas variáveis independentes X_1 e X_2 com médias θ_1 e θ_2 , respectivamente, existe um estimador não trivial $\delta(X_1, X_2)$, não dependente de X_1 , que é admissível para estimar θ_1 .

Exemplo 4.2: Considere variáveis independentes normais $X_1 \sim N(\theta_1, 1)$ e $X_2 \sim N(\theta_2, 1)$. Deseja-se estimar θ_1 com base em uma amostra conjunta (X_1, X_2) . Considere a função perda em erro quadrático para a amostra bivariada, dada por:

$$L(\delta(X_1, X_2), (\theta_1, \theta_2)) = (\delta(X_1, X_2) - \theta_1)^2.$$

Seja $\delta(X_1, X_2) = \text{sinal}(X_2)$. Vamos mostrar que apesar de X_2 não conter nenhuma informação sobre θ_1 , δ é admissível para estimar esse parâmetro. A função risco desse estimador é dada por:

$$\begin{aligned} \mathcal{R}(\delta, (\theta_1, \theta_2)) &= E_{(\theta_1, \theta_2)} [L(\delta(X_1, X_2), (\theta_1, \theta_2))] \\ &= E_{(\theta_1, \theta_2)} \left[(\text{sinal}(X_2) - \theta_1)^2 \right] \\ &= E_{(\theta_1, \theta_2)} \left[(\text{sinal}(X_2))^2 \right] - 2\theta_1 E_{(\theta_1, \theta_2)} [\text{sinal}(X_2)] + \theta_1^2 \\ &= 1 - 2\theta_1 E_{(\theta_1, \theta_2)} [\text{sinal}(X_2)] + \theta_1^2, \end{aligned} \quad (4.1)$$

em que $E_{(\theta_1, \theta_2)} \left[(\text{sinal}(X_2))^2 \right] = 1$, pois $\text{sinal}(X_2) = \pm 1$.

Denotando a função de distribuição acumulada (fda) da distribuição Normal pelo símbolo $\Phi(x)$, resulta que:

$$\begin{aligned} E_{(\theta_1, \theta_2)} [\text{sinal}(X_2)] &= \Phi_{\theta_2, 1}(0)(-1) + (1 - \Phi_{\theta_2, 1}(0))(1) \\ &= 1 - 2\Phi_{\theta_2, 1}(0). \end{aligned} \quad (4.2)$$

Portanto, substituindo (4.2) em (4.1) vem que:

$$\mathcal{R}(\delta, (\theta_1, \theta_2)) = 1 - 2\theta_1 + 4\theta_1\Phi_{\theta_2, 1}(0) + \theta_1^2.$$

Para $\theta_1 = 1$ e $\theta_1 = -1$, tem-se respectivamente que:

$$\mathcal{R}(\delta, (1, \theta_2)) = 1 - 2 + 4\Phi_{\theta_2, 1}(0) + 1 = 4\Phi_{\theta_2, 1}(0). \quad (4.3)$$

$$\mathcal{R}(\delta, (-1, \theta_2)) = 1 + 2 - 4\Phi_{\theta_2, 1}(0) + 1 = 4 - 4\Phi_{\theta_2, 1}(0). \quad (4.4)$$

Suponha que exista outro estimador δ' tal que:

$$\mathcal{R}(\delta', (\pm 1, \theta_2)) \leq \mathcal{R}(\delta, (\pm 1, \theta_2)). \quad (4.5)$$

Assim, pela definição da função risco, segue para o estimador δ' :

$$\begin{aligned} \mathcal{R}(\delta', (1, \theta_2)) &= E_{(1, \theta_2)} \left[\left(\delta'(X_1, X_2) - 1 \right)^2 \right] \\ &= E_{(1, \theta_2)} \left[\left(\delta'(X_1, X_2) \right)^2 \right] - 2E_{(1, \theta_2)} \left[\delta'(X_1, X_2) \right] + 1 \end{aligned}$$

e

$$\begin{aligned}\mathcal{R}(\delta', (-1, \theta_2)) &= E_{(-1, \theta_2)} \left[\left(\delta'(X_1, X_2) - (-1) \right)^2 \right] \\ &= E_{(-1, \theta_2)} \left[\left(\delta'(X_1, X_2) \right)^2 \right] + 2E_{(-1, \theta_2)} \left[\delta'(X_1, X_2) \right] + 1.\end{aligned}$$

Consequentemente, em função da relação em (4.5), para $\mathcal{R}(\delta, (1, \theta_2))$ e $\mathcal{R}(\delta, (-1, \theta_2))$ dados respectivamente em (4.3) e (4.4), segue que:

$$E_{(1, \theta_2)} \left[\left(\delta'(X_1, X_2) \right)^2 \right] - 2E_{(1, \theta_2)} \left[\delta'(X_1, X_2) \right] + 1 \leq 4\Phi_{(\theta_2, 1)}(0)$$

e

$$E_{(-1, \theta_2)} \left[\left(\delta'(X_1, X_2) \right)^2 \right] + 2E_{(-1, \theta_2)} \left[\delta'(X_1, X_2) \right] + 1 \leq 4 - 4\Phi_{\theta_2, 1}(0).$$

Para completar o exercício, considere dois casos:

Caso i) Considere $\theta_1 = 1$ e seja a função:

$$r(x_2) = \int_{-\infty}^{+\infty} \left(\delta'(x_1, x_2) - 1 \right)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1 - 1)^2\right) dx_1 \geq 0.$$

Assim, segue que o $\mathcal{R}(\delta', (1, \theta_2))$ de δ' é dado por:

$$\begin{aligned}
& \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\delta'(x_1, x_2) - 1)^2 \frac{\exp\left(-\frac{(x_1-1)^2}{2}\right)}{\sqrt{2\pi}} \frac{\exp\left(-\frac{(x_2-\theta_2)^2}{2}\right)}{\sqrt{2\pi}} dx_1 dx_2 \\
&= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} (\delta'(x_1, x_2) - 1)^2 \frac{\exp\left(-\frac{(x_1-1)^2}{2}\right)}{\sqrt{2\pi}} dx_1 \right] \frac{\exp\left(-\frac{(x_2-\theta_2)^2}{2}\right)}{\sqrt{2\pi}} dx_2 \\
&= \int_{-\infty}^{+\infty} r(x_2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2 \\
&= \int_{-\infty}^0 r(x_2) \frac{\exp\left(-\frac{(x_2-\theta_2)^2}{2}\right)}{\sqrt{2\pi}} dx_2 + \int_0^{+\infty} r(x_2) \frac{\exp\left(-\frac{(x_2-\theta_2)^2}{2}\right)}{\sqrt{2\pi}} dx_2.
\end{aligned}$$

Como $\mathcal{R}(\delta, (1, \theta_2)) = 4\Phi_{\theta_2, 1}(0)$, resulta que o risco do estimador δ para $x_2 > 0$ é:

$$\begin{aligned}
& \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\text{sinal}(x_2) - 1)^2 \frac{\exp\left(-\frac{1}{2}(x_1 - 1)^2\right)}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_1 dx_2 \\
&= \int_{-\infty}^0 \int_{-\infty}^{+\infty} \frac{4}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1 - 1)^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_1 dx_2 \\
&= \int_{-\infty}^0 \left(4 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1 - 1)^2\right) dx_1 \right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2 \\
&= 4 \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2,
\end{aligned}$$

uma vez que $(\delta(x_1, x_2) - 1)^2 = (\text{sinal}(x_2) - 1)^2 = 0$ para $x_2 > 0$.

Utilizando a suposição de que δ' domina δ , isto é, $\mathcal{R}(\delta', (1, \theta_2)) \leq$

$\mathcal{R}(\delta, (1, \theta_2))$, tem-se que $\mathcal{R}(\delta', (1, \theta_2)) - \mathcal{R}(\delta, (1, \theta_2)) \leq 0$. Logo,

$$\begin{aligned}
0 &\geq \mathcal{R}(\delta', (1, \theta_2)) - \mathcal{R}(\delta, (1, \theta_2)) \\
&= \int_{-\infty}^0 r(x_2) \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_2 + \int_0^{+\infty} r(x_2) \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_2 \\
&\quad - 4 \int_{-\infty}^0 \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_2 \\
&\geq \int_0^{+\infty} r(x_2) \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_2 - 4 \int_{-\infty}^0 \frac{\exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right)}{\sqrt{2\pi}} dx_2 \\
&= \frac{\int_0^{+\infty} r(x_2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2}{\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2} - 4.
\end{aligned}$$

Conforme Makani (1977), pelas propriedades da família exponencial tem-se que:

$$\lim_{\theta_2 \rightarrow \infty} \frac{\int_0^{+\infty} r(x_2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2}{\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_2 - \theta_2)^2\right) dx_2} \rightarrow \infty.$$

Absurdo, a menos que $r(x_2) = 0$ para quase todo ponto $x_2 > 0$. Assim,

$$r(x_2) = \int_{-\infty}^{+\infty} \left(\delta'(x_1, x_2) - 1\right)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_1 - 1)^2\right) dx_1 = 0.$$

Essa igualdade somente é possível se $\left(\delta'(x_1, x_2) - 1\right) = 0$. Portanto,

$$\delta' (X_1, X_2) = +1.$$

Caso ii) Com base na simetria da distribuição Normal, concluí-se que $r(x_2) = 0$ para $\theta_1 = -1$ e $x_2 \leq 0$. De modo análogo, essa igualdade somente é possível se $(\delta'(x_1, x_2) - (-1)) = 0$. Portanto, $\delta'(X_1, X_2) = -1$.

Como consequência destes dois casos, $\delta'(X_1, X_2) = \pm 1$. Isso implica que $\delta' = \delta$, e neste caso, o estimador $\delta(X_1, X_2) = \text{sinal}(X_2)$ é admissível. Neste exemplo foi mostrado que, dados dois problemas de estimação independentes, um estimador baseado em dados de somente um problema pode ser admissível como estimador do parâmetro do outro problema.

Lema 4.1: Seja δ um estimador de Bayes (respectivamente, UMVUE, *minimax* e admissível) de $\omega(\theta)$ em relação a função perda em erro quadrático. Então, $a\delta + b$ é um estimador de Bayes (respectivamente, UMVUE, *minimax* e admissível) para $a\omega(\theta) + b$.

Prova:

A prova segue imediatamente a partir da definição da função risco. De fato,

$$\begin{aligned} \mathcal{R}(a\delta + b, a\omega(\theta) + b) &= E_{\theta} [L(a\delta + b, a\omega(\theta) + b)] \\ &= E_{\theta} \left[[a\delta + b - (a\omega(\theta) + b)]^2 \right] \\ &= E_{\theta} \left[(a\delta - a\omega(\theta))^2 \right] \\ &= a^2 E_{\theta} \left[(\delta - \omega(\theta))^2 \right] \\ &= a^2 E_{\theta} [L(\delta, \omega(\theta))] \\ &= a^2 \mathcal{R}(\delta, \omega(\theta)). \end{aligned} \tag{4.6}$$

(i) $a\delta_\pi + b$ é um estimador de Bayes para $a\omega(\theta) + b$.

Seja δ_π estimador de Bayes de $\omega(\theta)$. A função risco do estimador $a\delta_\pi + b$ de $a\omega(\theta) + b$ é dada por:

$$\begin{aligned}\mathcal{R}(a\delta_\pi + b, a\omega(\theta) + b) &= E_\theta [L(a\delta_\pi + b, a\omega(\theta) + b)] \\ &= \int_{\mathfrak{X}^n} (a\delta_\pi + b - (a\omega(\theta) + b))^2 f_{X|\Theta}(x|\theta) dx \\ &= \int_{\mathfrak{X}^n} a^2 (\delta_\pi - \omega(\theta))^2 f_{X|\Theta}(x|\theta) dx \\ &= a^2 E_\theta [L(\delta_\pi, \omega(\theta))] \\ &= a^2 \mathcal{R}(\delta_\pi, \omega(\theta)).\end{aligned}$$

Logo, o risco de Bayes desse estimador é:

$$\begin{aligned}r(a\delta_\pi + b) &= \int_{\Theta} \mathcal{R}(a\delta_\pi + b, a\omega(\theta) + b) \pi(\theta) d\theta \\ &= a^2 \int_{\Theta} \mathcal{R}(\delta_\pi, \omega(\theta)) \pi(\theta) d\theta \\ &= a^2 r(\delta_\pi).\end{aligned}$$

Seja δ' um estimador qualquer de $a\omega(\theta) + b$. Logo,

$$\begin{aligned}
r(\delta') &= \int_{\Theta} \mathcal{R}(\delta', a\omega(\theta) + b) \pi(\theta) d\theta \\
&= \int_{\Theta} \left[\int_{\mathfrak{X}^n} (\delta' - (a\omega(\theta) + b))^2 f_{X|\Theta}(x|\theta) dx \right] \pi(\theta) d\theta \\
&= \int_{\Theta} \left\{ \int_{\mathfrak{X}^n} \left[a \left(\frac{\delta' - b}{a} - \omega(\theta) \right) \right]^2 f_{X|\Theta}(x|\theta) dx \right\} \pi(\theta) d\theta \\
&= a^2 \int_{\Theta} \left[\int_{\mathfrak{X}^n} \left(\frac{\delta' - b}{a} - \omega(\theta) \right)^2 f_{X|\Theta}(x|\theta) dx \right] \pi(\theta) d\theta \\
&= a^2 \int_{\Theta} \mathcal{R}\left(\frac{\delta' - b}{a}, \omega(\theta)\right) \pi(\theta) d\theta \\
&= a^2 r\left(\frac{\delta' - b}{a}\right) \\
&\geq a^2 r(\delta_\pi) \\
&= r(a\delta_\pi + b).
\end{aligned}$$

Logo, o estimador $a\delta_\pi + b$ possui risco de Bayes menor ou igual em relação a qualquer outro estimador de $a\omega(\theta) + b$. Portanto, $a\delta_\pi + b$ é um estimador de Bayes para $a\omega(\theta) + b$. A desigualdade na demonstração segue da hipótese de δ_π ser o estimador de Bayes de $\omega(\theta)$, e por isso, o risco de Bayes de $\frac{\delta' - b}{a}$ é maior ou igual ao risco de Bayes de δ_π .

(ii) O estimador $a\delta + b$ é um UMVUE para $a\omega(\theta) + b$.

Para provar que $a\delta + b$ é um UMVUE para $a\omega(\theta) + b$ é necessário mostrar que esse estimador é não viesado e que possui variância uniforme mínima entre todos os estimadores de $a\omega(\theta) + b$. Primeiramente, como δ é um estimador não viesado de $\omega(\theta)$, segue que $a\delta + b$ é um estimador não viesado de $a\omega(\theta) + b$. De

fato,

$$E_{\theta} [a\delta + b] = aE_{\theta} [\delta] + b = a\omega(\theta) + b.$$

Segue também para o estimador $a\delta + b$ que:

$$\text{var}_{\theta} [a\delta + b] = a^2 \text{var}_{\theta} [\delta].$$

Se δ' é um estimador não viesado qualquer de $a\omega(\theta) + b$, então $\frac{\delta' - b}{a}$ é um estimador não viesado de $\omega(\theta)$. De fato,

$$E_{\theta} \left[\frac{\delta' - b}{a} \right] = \frac{1}{a} \left(E_{\theta} [\delta'] - b \right) = \frac{1}{a} (a\omega(\theta) + b - b) = \omega(\theta).$$

Como $\frac{\delta' - b}{a}$ é um estimador não viesado de $\omega(\theta)$ e δ é um UMVUE para $\omega(\theta)$, resulta que $\text{var}_{\theta} \left[\frac{\delta' - b}{a} \right] \geq \text{var}_{\theta} [\delta]$. Consequentemente,

$$\text{var}_{\theta} \left[\frac{\delta' - b}{a} \right] = \frac{1}{a^2} \text{var}_{\theta} [\delta'] \geq \text{var}_{\theta} [\delta] \Rightarrow \text{var}_{\theta} [\delta'] \geq a^2 \text{var}_{\theta} [\delta] = \text{var}_{\theta} [a\delta + b].$$

Logo, $a\delta + b$ possui variância menor ou igual em relação a qualquer outro estimador de $a\omega(\theta) + b$. Portanto, $a\delta + b$ é um UMVUE para $a\delta + b$.

(iii) $a\delta + b$ é um estimador *minimax* para $a\omega(\theta) + b$.

Se δ é *minimax* para $\omega(\theta)$, então para qualquer outro estimador δ' de $\omega(\theta)$, segue que:

$$\sup_{\theta} \mathcal{R}(\delta, \omega(\theta)) \leq \sup_{\theta} \mathcal{R}(\delta', \omega(\theta)).$$

Seja δ' um estimador qualquer de $a\omega(\theta) + b$. Isso implica que $\frac{\delta' - b}{a}$ é um estimador de $\omega(\theta)$. Assim,

$$\begin{aligned}
 \sup_{\theta} \mathcal{R}(\delta', a\omega(\theta) + b) &= \sup_{\theta} \int_{\mathfrak{X}^n} L(\delta', a\omega(\theta) + b) f_X(x; \theta) dx \\
 &= \sup_{\theta} \int_{\mathfrak{X}^n} (\delta' - (a\omega(\theta) + b))^2 f_X(x; \theta) dx \\
 &= \sup_{\theta} \int_{\mathfrak{X}^n} \left[a \left(\frac{\delta' - b}{a} - \omega(\theta) \right) \right]^2 f_X(x; \theta) dx \\
 &= a^2 \sup_{\theta} \int_{\mathfrak{X}^n} \left(\frac{\delta' - b}{a} - \omega(\theta) \right)^2 f_X(x; \theta) dx \\
 &= a^2 \sup_{\theta} \mathcal{R} \left(\frac{\delta' - b}{a}, \omega(\theta) \right) \\
 &\geq a^2 \sup_{\theta} \mathcal{R}(\delta, \omega(\theta)) \quad (\delta \text{ é minimax para } \omega(\theta)) \\
 &= \sup_{\theta} \mathcal{R}(a\delta + b, a\omega(\theta) + b). \quad (\text{Usar (4.6)})
 \end{aligned}$$

Como o risco máximo de $a\delta + b$ é menor ou igual ao risco máximo de δ' , conclui-se que $a\delta + b$ é um estimador *minimax* para $a\omega(\theta) + b$

(iv) $a\delta + b$ é um estimador admissível para $a\omega(\theta) + b$.

Seja δ' um estimador de $a\omega(\theta) + b$. Então, $\frac{\delta' - b}{a}$ é um estimador de $\omega(\theta)$. Assim,

$$\begin{aligned}
\mathcal{R}(\delta', a\omega(\theta) + b) &= \int_{\mathfrak{X}^n} L(\delta', a\omega(\theta) + b) f_X(x; \theta) dx \\
&= \int_{\mathfrak{X}^n} (\delta' - (a\omega(\theta) + b))^2 f_X(x; \theta) dx \\
&= \int_{\mathfrak{X}^n} \left[a \left(\frac{\delta' - b}{a} - \omega(\theta) \right) \right]^2 f_X(x; \theta) dx \\
&= a^2 \int_{\mathfrak{X}^n} \left(\frac{\delta' - b}{a} - \omega(\theta) \right)^2 f_X(x; \theta) dx \\
&= a^2 \mathcal{R} \left(\frac{\delta' - b}{a}, \omega(\theta) \right) \\
&\geq a^2 \mathcal{R}(\delta, \omega(\theta)) \quad (\delta \text{ é admissível para } \omega(\theta)) \\
&= \mathcal{R}(a\delta + b, a\omega(\theta) + b). \quad (\text{Usar (4.6)})
\end{aligned}$$

Uma vez que o risco de $a\delta + b$ é menor ou igual ao risco de δ' , segue que $a\delta + b$ é um estimador admissível para $a\omega(\theta) + b$.

■

Definição 4.3 (Priori menos favorável): Uma distribuição a priori π é dita *menos favorável* se $r_\pi \geq r_{\pi'}$, para todas distribuições a priori π' .

Portanto, quando o estatístico adota como priori a distribuição menos favorável ele incorre num maior risco médio (risco de Bayes). O próximo teorema apresenta uma condição simples para um estimador de Bayes δ_π ser *minimax*.

Teorema 4.1: Suponha que π é uma distribuição a priori para Θ tal que:

$$r(\delta_\pi) = \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta = \sup_{\theta} \mathcal{R}(\delta_\pi, \theta).$$

Então:

- (i) δ_π é *minimax*.
- (ii) Se δ_π é o único estimador de Bayes em relação à priori π , então δ_π é o único estimador *minimax*.
- (iii) π é uma priori menos favorável.

Prova:

- (i) Seja δ um estimador de θ . Como δ_π possui menor risco de Bayes, então:

$$\sup_{\theta} \mathcal{R}(\delta, \theta) d\theta \geq \int_{\Theta} \mathcal{R}(\delta, \theta) \pi(\theta) d\theta \geq \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta = \sup_{\theta} \mathcal{R}(\delta_\pi, \theta).$$

Portanto, δ_π é *minimax*.

- (ii) A prova segue pela troca \geq por $>$ na segunda igualdade da prova de i).

- (iii) Seja π' outra priori de Θ . Então,

$$r(\delta_{\pi'}) = \int_{\Theta} \mathcal{R}(\delta_{\pi'}, \theta) \pi'(\theta) d\theta \leq \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi'(\theta) d\theta \leq \sup_{\theta} \mathcal{R}(\delta_\pi, \theta) = r(\delta_\pi).$$

Portanto, π é uma priori menos favorável.

■

Um caso particular em relação ao teorema 4.1 ocorre quando a função risco é constante, ou mais geralmente, quando π atribui probabilidade 1 para o conjunto em que a função de risco assume o seu valor máximo.

Corolário 4.1: Se um estimador de Bayes δ_π têm risco constante, então esse estimador é *minimax*.

Prova:

Como o estimador δ_π possui risco constante, o risco de Bayes desse estimador é dado por:

$$r(\delta_\pi) = \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta = \sup_{\theta} \mathcal{R}(\delta_\pi, \theta).$$

Dado outro estimador δ' , o seu risco de Bayes é:

$$r(\delta') = \int_{\Theta} \mathcal{R}(\delta', \theta) \pi(\theta) d\theta \leq \sup_{\theta} \mathcal{R}(\delta', \theta).$$

Assim,

$$r(\delta_\pi) = \sup_{\theta} \mathcal{R}(\delta_\pi, \theta) \leq r(\delta') \leq \sup_{\theta} \mathcal{R}(\delta', \theta).$$

Uma vez que $\sup_{\theta} \mathcal{R}(\delta_\pi, \theta) \leq \sup_{\theta} \mathcal{R}(\delta', \theta)$, conclui-se que o estimador δ_π é *minimax*. ■

No próximo exemplo apresenta-se uma aplicação do corolário 4.1.

Exemplo 4.3: Seja (X_1, \dots, X_n) uma amostra iid Bernoulli(p). O risco de $\hat{p} = \bar{X}$, considerando a função perda em erro quadrático é:

$$E_p \left[(\hat{p} - p)^2 \right] = \text{var}_p (\bar{X}) = \frac{p(1-p)}{n}.$$

Seja $Y = \sum_{i=1}^n X_i$. Neste caso, $Y \sim \text{Binomial}(n, p)$. A família de distribuições Beta de parâmetros α e β é uma família de possíveis distribuições a priori de p . A distribuição conjunta de Y e p é:

$$\begin{aligned} f_{Y,p}(y, p) &= \left[\binom{n}{y} p^y (1-p)^{n-y} \right] \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \right] \\ &= \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}. \end{aligned}$$

A fdp marginal de Y é:

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{Y,p}(y, p) dp \\ &= \int_0^1 \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1} dp \\ &= \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y + \alpha)\Gamma(n - y + \beta)}{\Gamma(n + \alpha + \beta)}, \end{aligned}$$

em que a última igualdade é obtida multiplicando-se a integral por um, da forma:

$$\frac{\Gamma(y + \alpha)\Gamma(n - y + \beta)}{\Gamma(n + \alpha + \beta)} \frac{\Gamma(n + \alpha + \beta)}{\Gamma(y + \alpha)\Gamma(n - y + \beta)}.$$

A fdp marginal de Y resultante é conhecida como Beta-Binomial. A distribuição a posteriori de p , dado $Y = y$, é:

$$f_{p|Y}(p|y) = \frac{f_{Y,p}(y,p)}{f_Y(y)} = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(y + \alpha) \Gamma(n - y + \beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}.$$

Note que a distribuição a posteriori de p , dado $Y = y$, é uma densidade Beta $(y + \alpha, n - y + \beta)$. Uma estimativa para p é a média da distribuição a posteriori, que fornece como estimador de Bayes de p :

$$\hat{p}_B = \frac{y + \alpha}{n + \alpha + \beta}.$$

O risco desse estimador de Bayes é:

$$\begin{aligned} E_p \left[(\hat{p}_B - p)^2 \right] &= \text{var}_p [\hat{p}_B] + (\text{viés } \hat{p}_B)^2. \\ &= \text{var}_p \left[\frac{Y + \alpha}{n + \alpha + \beta} \right] + \left(E_p \left[\frac{Y + \alpha}{n + \alpha + \beta} \right] - p \right)^2 \\ &= \left(\frac{1}{n + \alpha + \beta} \right)^2 \text{var}_p [Y] + \left(\frac{1}{n + \alpha + \beta} E_p [Y + \alpha] - p \right)^2 \\ &= \frac{np(1-p)}{(n + \alpha + \beta)^2} + \left(\frac{np + \alpha}{n + \alpha + \beta} - p \right)^2. \end{aligned}$$

Na ausência de informações a priori sobre p , podemos escolher α e β para tornar a função risco constante. Assim, seja $\delta_{A,B}^*(Y_1, \dots, Y_n) = AY + B = \frac{Y + \alpha}{n + \alpha + \beta}$. Então:

$$\begin{aligned}
\mathcal{R}(\delta_{A,B}^*, p) &= E_p \left[(AY + B - p)^2 \right] \\
&= E_p \left[[A(Y - np) + B - p + nAp]^2 \right] \quad (\text{Somar e subtrair } nAp) \\
&= A^2 E_p \left[(Y - E_p[Y])^2 \right] + 0 + (B - p + nAp)^2 \\
&= A^2 \text{var}_p[Y] + (B - p + nAp)^2 \\
&= A^2 np(1 - p) + (B - p + nAp)^2 \\
&= nA^2 p - nA^2 p^2 + B^2 - 2B(p - nAp) + (p - nAp)^2 \\
&= nA^2 p - nA^2 p^2 + B^2 - 2Bp + 2nABp + p^2 - 2nAp^2 + (nAp)^2 \\
&= p^2 \left[(nA)^2 - 2nA + 1 - nA^2 \right] + p \left[nA^2 + 2nAB - 2B \right] + B^2 \\
&= p^2 \left[(nA - 1)^2 - nA^2 \right] + p \left[nA^2 + 2(nA - 1)B \right] + B^2.
\end{aligned}$$

Pode-se verificar que a função risco será constante quando $(nA - 1)^2 - nA^2 = 0$ e $nA^2 + 2(nA - 1)B = 0$. Segue então que:

$$\begin{aligned}
(nA - 1)^2 - nA^2 = 0 &\Rightarrow (nA - 1)^2 = nA^2 \\
&\Rightarrow nA - 1 = \pm \sqrt{n}A \\
&\Rightarrow nA \pm \sqrt{n}A = 1 \\
&\Rightarrow A = \frac{1}{n \pm \sqrt{n}} \\
&\Rightarrow A = \frac{1}{\sqrt{n}(\sqrt{n} \pm 1)}.
\end{aligned}$$

De forma análoga, $nA^2 + 2(nA - 1)B = 0$ se $B = \frac{-nA^2}{2(nA - 1)}$. Para $A = \frac{1}{\sqrt{n}(\sqrt{n} + 1)}$ tem-se então que:

$$\begin{aligned}
B &= \frac{-nA^2}{2(nA-1)} = \frac{-n \left[\frac{1}{\sqrt{n}(\sqrt{n+1})} \right]^2}{2 \left[n \frac{1}{\sqrt{n}(\sqrt{n+1})} - 1 \right]} \\
&= \frac{-\frac{1}{(\sqrt{n+1})^2}}{2 \left[\frac{\sqrt{n}}{(\sqrt{n+1})} - 1 \right]} \\
&= \frac{-\frac{1}{(\sqrt{n+1})^2}}{2 \left[\frac{\sqrt{n}-\sqrt{n-1}}{(\sqrt{n+1})} \right]} \\
&= \frac{-\frac{1}{(\sqrt{n+1})^2}}{-\frac{2}{(\sqrt{n+1})}} \\
&= \frac{1}{2(\sqrt{n+1})}.
\end{aligned}$$

Finalmente, resolvendo para α e β , obtêm-se que $\alpha = \beta = \sqrt{\frac{n}{4}}$. Portanto,

$$\hat{p}_B = \frac{Y + \sqrt{\frac{n}{4}}}{n + \sqrt{\frac{n}{4}} + \sqrt{\frac{n}{4}}} = \frac{Y + \sqrt{\frac{n}{4}}}{n + \sqrt{n}},$$

é o estimador de Bayes com risco constante para p , e conseqüentemente, pelo corolário 4.1 esse estimador é *minimax*.

Teorema 4.2: Se o estimador de Bayes é único, então esse estimador é admissível.

Prova:

Seja δ_π o único estimador de Bayes em relação a uma priori π . Suponha agora que δ_π é dominado por δ' . Nesse caso, $\mathcal{R}(\delta', \theta) \leq \mathcal{R}(\delta_\pi, \theta)$ e $\mathcal{R}(\delta', \theta_0) < \mathcal{R}(\delta_\pi, \theta_0)$ para algum $\theta_0 \in \Theta$. Assim,

$$r(\delta_\pi) = \int_{\Theta} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta \geq \int_{\Theta} \mathcal{R}(\delta', \theta) \pi(\theta) d\theta = r(\delta').$$

Absurdo, pois a unicidade do estimador de Bayes δ_π assegura que não existe outro estimador de θ que possua menor risco de Bayes, para a priori π . Logo, dado um estimador qualquer δ' , $\mathcal{R}(\delta_\pi, \theta) \leq \mathcal{R}(\delta', \theta)$, e portanto, δ_π é admissível.

■

No resultado anterior, a necessidade da unicidade do estimador de Bayes pode ser substituída por outras condições.

Teorema 4.3: Para um possível vetor de valores de parâmetros θ , suponha que δ_π é um estimador de Bayes com risco finito em relação a uma priori π positiva para todo θ , e que a função risco de todo estimador δ é uma função contínua de θ . Então, δ_π é admissível.

Prova:

Se δ_π não é admissível, existe um estimador δ tal que $\mathcal{R}(\delta, \theta) \leq \mathcal{R}(\delta_\pi, \theta)$, para todo θ e $\mathcal{R}(\delta, \theta) < \mathcal{R}(\delta_\pi, \theta)$, para algum θ . Segue então por continuidade da função risco que $\mathcal{R}(\delta, \theta) < \mathcal{R}(\delta_\pi, \theta)$ para todo θ em algum subconjunto aberto Θ_0 do espaço paramétrico. Consequentemente:

$$r(\delta_\pi) = \int_{\Theta_0} \mathcal{R}(\delta_\pi, \theta) \pi(\theta) d\theta > \int_{\Theta_0} \mathcal{R}(\delta, \theta) \pi(\theta) d\theta,$$

o que contradiz o fato de δ_π ser um estimador de Bayes. Portanto, δ_π é admissível.

■

Teorema 4.4: Seja X uma variável aleatória com média θ e variância σ^2 . Então, $aX + b$ é um estimador de θ inadmissível para a função perda em erro quadrático, sempre que:

- (i) $a > 1$, ou
- (ii) $a < 0$, ou
- (iii) $a = 1$ e $b \neq 0$.

Prova:

O risco de $aX + b$ é:

$$\begin{aligned}
 \rho(a, b) &= E_{\theta} \left[(aX + b - \theta)^2 \right] \\
 &= \text{var}_{\theta} [aX + b] + [\text{viés} (aX + b)]^2 \\
 &= \text{var}_{\theta} [aX + b] + (E_{\theta} [aX + b] - \theta)^2 \\
 &= a^2 \text{var}_{\theta} [X] + (aE_{\theta} [X] + b - \theta)^2 \\
 &= a^2 \sigma^2 + (a\theta + b - \theta)^2 \\
 &= a^2 \sigma^2 + [(a - 1)\theta + b]^2.
 \end{aligned} \tag{4.7}$$

(i) Se $a > 1$, então:

$$\rho(a, b) \geq a^2 \sigma^2 > \sigma^2 = E_{\theta} \left[(X - E_{\theta} [X])^2 \right] = E_{\theta} \left[(1X + 0 - \theta)^2 \right] = \rho(1, 0).$$

Logo, $aX + b$ é dominado por X .

(ii) Se $a < 0$, então $(a - 1)^2 > 1$. Consequentemente:

$$\begin{aligned}
\rho(a, b) &\geq [(a-1)\theta + b]^2 \\
&= (a-1)^2 \left(\theta + \frac{b}{a-1} \right)^2 \\
&> \left(\theta + \frac{b}{a-1} \right)^2 \\
&= \left(-\theta - \frac{b}{a-1} \right)^2 \\
&= \rho \left(0, -\frac{b}{a-1} \right). \quad (\text{Usar (4.7)})
\end{aligned}$$

Portanto, $aX + b$ é dominado pelo estimador constante $\delta \equiv -\frac{b}{a-1}$.

(iii) Se $a = 1$ e $b \neq 0$, então $aX + b = X + b$ é dominado por X . Isso ocorre em razão da $E_\theta[X] = \theta$ e do viés de b não depender de θ . Nessas condições, $X + b$ possui risco maior que X . De fato,

$$\begin{aligned}
\rho(1, b) &= E_\theta \left[(X + b - \theta)^2 \right] \\
&= \text{var}_\theta [X + b] + [\text{viés}(X + b)]^2 \\
&= \text{var}_\theta [X + b] + (E_\theta [X + b] - \theta)^2 \\
&= \text{var}_\theta [X] + (E_\theta [X] + b - \theta)^2 \\
&= \text{var}_\theta [X] + (\theta + b - \theta)^2 \\
&> \text{var}_\theta [X] \\
&= E_\theta \left[(X - E_\theta [X])^2 \right] \\
&= E_\theta \left[(X - \theta)^2 \right] \\
&= \rho(1, 0).
\end{aligned}$$

Portanto, para a função perda em erro quadrático e considerando os três casos, $aX + b$ é um estimador inadmissível de θ .



Como \bar{X} é o estimador de Pitman para estimar a média da distribuição Normal quando a variância é conhecida, no próximo exemplo é apresentado uma prova, que também se encontra em Lehmann e Casella (1998, p. 324-325), da admissibilidade de \bar{X} .

Exemplo 4.4 (Admissibilidade de \bar{X}): A admissibilidade de \bar{X} para estimar a média da distribuição Normal não é só de grande interesse em si só, mas também pode ser considerada como o ponto de partida de muitas outras investigações de admissibilidade.

Para provar admissibilidade de \bar{X} será utilizado o método do limite de estimadores de Bayes. Um estimador $\delta(X_1, \dots, X_n)$ é um limite de estimadores de Bayes se existe uma sequência de prioris π_n tal que os estimadores de Bayes relativos a essas prioris $\delta_{\pi_n}(X_1, \dots, X_n)$ convergem pontualmente para δ . Suponha que \bar{X} não é admissível, e sem perda de generalidade, suponha que $\sigma = 1$. Então, existe δ^* tal que:

$$\mathcal{R}(\delta^*, \theta) \leq \frac{1}{n}, \text{ para todo } \theta \in \Theta.$$

$$\mathcal{R}(\delta^*, \theta_0) < \frac{1}{n}, \text{ para algum } \theta_0 \in \Theta.$$

Como $\mathcal{R}(\delta, \theta)$ é uma função contínua de θ para qualquer δ , existe um $\varepsilon > 0$ e um $\theta_1 < \theta_0 < \theta_2$ tal que:

$$\mathcal{R}(\delta^*, \theta) \leq \frac{1}{n} - \varepsilon, \text{ para todo } \theta_1 < \theta < \theta_2.$$

Considere uma sequência de prioris normais $\pi_\tau \sim N(0, \tau^2)$. Seja $r_\tau = \frac{\tau^2}{1+n\tau^2}$ o risco de Bayes do estimador \bar{X} relativo a priori π_τ e r_τ^* o risco de Bayes

de δ^* . Assim,

$$\begin{aligned} \frac{\frac{1}{n} - r_{\tau}^*}{\frac{1}{n} - r_{\tau}} &= \frac{\frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} \left[\frac{1}{n} - \mathcal{R}(\delta^*, \theta) \right] \exp\left(-\frac{\theta^2}{2\tau^2}\right) d\theta}{\frac{1}{n} - \frac{\tau^2}{1+n\tau^2}} \\ &\geq \frac{n(1+n\tau^2)\varepsilon}{\tau\sqrt{2\pi}} \int_{\theta_0}^{\theta_1} \exp\left(-\frac{\theta^2}{2\tau^2}\right) d\theta. \end{aligned}$$

O integrando converge monoticamente para 1 quando $\tau \rightarrow \infty$. Pelo teorema da convergência monótona de Lebesgue, a integral converge para $\theta_1 - \theta_0$, e, conseqüentemente, como $\tau^2 \rightarrow \infty$,

$$\frac{\frac{1}{n} - r_{\tau}^*}{\frac{1}{n} - r_{\tau}} \rightarrow \infty.$$

Então, existe τ_0 tal que $r_{\tau_0}^* < r_{\tau_0}$, o que contradiz o fato de r_{τ_0} ser o risco de Bayes para π_{τ_0} . Isto completa a prova.

A seguir, são apresentados dois lemas que relacionam as propriedades de admissibilidade e minimaximalidade de estimadores.

Lema 4.2: Se um estimador δ é admissível e possui risco constante, então é *minimax*.

Prova:

Suponha que δ não seja *minimax*, ou seja, existe um outro estimador δ' tal que:

$$\mathcal{R}(\delta', \theta) \leq \sup_{\theta} \mathcal{R}(\delta', \theta) \leq \sup_{\theta} \mathcal{R}(\delta, \theta) = \mathcal{R}(\delta, \theta).$$

Absurdo, pois δ é, por hipótese, um estimador admissível. Portanto, δ é *minimax*.

■

Lema 4.3: Se um estimador δ é o único estimador *minimax*, então é admissível.

Prova:

Suponha que exista um estimador δ' que domine δ , isto é, $\mathcal{R}(\delta', \theta) \leq \mathcal{R}(\delta, \theta)$. Isso implica que $\sup \mathcal{R}(\delta', \theta) \leq \sup \mathcal{R}(\delta, \theta)$, e nesse caso, δ' também seria *minimax*. Absurdo, pois, por hipótese, δ é o único estimador *minimax*. Portanto, δ é admissível.

■

A seguir, são apresentados dois teoremas, que se encontram em Lehmann e Casella (1998, p. 340-342), que relacionam o estimador de Pitman com as propriedades de admissibilidade e minimaximalidade de estimadores.

Teorema 4.5: Se o estimador de Pitman para parâmetros de locação possui variância finita, então é *minimax* para função perda em erro quadrático.

Teorema 4.6: Se existe um estimador equivariante δ_0 para o parâmetro de locação θ com $E_0 \left[(\delta_0(\mathbf{X}))^3 \right] < \infty$, então o estimador de Pitman é admissível para a função perda em erro quadrático.

Lehmann e Casella (1998, p. 342) afirmam que, os estimadores MRE tipicamente não serão admissíveis, exceto em situações simplistas, mas eles tem melhores chances de serem *minimax*. Um tratamento maior sobre a relação entre

o estimador de Pitman para parâmetros de locação e as propriedades de admissibilidade e minimaximalidade pode ser visto em Zacks (1971). Zacks (1971, p. 325-331) prova que o estimador de Pitman para parâmetros de locação é um estimador *minimax*, para a função risco considerando a função perda em erro quadrático. Além disso, Zacks (1971, p. 387-395) demonstra que, se a distribuição de \mathbf{X} depende de no máximo dois parâmetros de locação (quando θ é um vetor de parâmetros bidimensional), o estimador de Pitman é admissível. Segundo esse autor, se a dimensão do vetor de observações \mathbf{X} é pelo menos três, o estimador de Pitman será geralmente inadmissível.

5 ESTUDO DE CASO: APLICAÇÃO A NÍVEIS PLUVIOMÉTRICOS MÁXIMOS

5.1 Introdução

A ocorrência de eventos climáticos extremos severos tem chamado a atenção da sociedade e da comunidade científica, principalmente em função de perdas de vidas humanas e animais, e de ordem econômica. Como eventos climáticos extremos, é possível citar fenômenos naturais tais como chuvas intensas, secas prolongadas, vendavais, furacões, marés meteorológicas e temperaturas mínimas ou máximas. O interesse na análise em um dos referidos eventos centra-se tanto na intensidade de sua manifestação quanto na sua duração prolongada. Por apresentarem baixa frequência relativa e elevado impacto, a previsão desses eventos torna-se difícil.

Uma variável climática de grande interesse na análise de seu comportamento é a precipitação pluviométrica, uma vez que diversas atividades econômicas e processos ambientais são altamente dependentes dessa variável. Elevados regimes de precipitação podem ocasionar graves efeitos, como enchentes, deslizamentos de terras, atrasos em colheitas, entre outros. Por outro lado, o déficit de precipitação em larga escala, muitas vezes, causam efeitos adversos sobre atividades como agricultura, silvicultura, produção hidrelétrica, ecossistemas alagados (manguezais) e a vida selvagem em geral (MARENGO et al., 2010).

A ocorrência de grandes desvios de um estado climático moderado tem demonstrado a vulnerabilidade do Brasil, principalmente em relação à precipitação pluviométrica. O baixo índice de chuvas durante o verão e o outono de 2001 culminou com uma redução significativa da vazão dos rios brasileiros, afetando a produção de energia hidrelétrica. Esse cenário impeliu o governo a tomar medidas visando à conservação de energia durante os anos de 2001 e 2002, para impedir a

interrupção total do fornecimento (blackout).

Anomalias no regime de precipitação no sudoeste da amazônia foram presenciadas em 2005, culminando em uma das secas mais severas dos últimos cem anos na região. A navegação em rios como o Amazonas, o Solimões e o rio Madeira foi prejudicada, afetando a população ribeirinha. Países como o Brasil, a Bolívia, o Peru e a Colômbia declararam estado de calamidade pública em setembro de 2005. Devido à baixa umidade do ar, estima-se que o número de incêndios florestais em 2005 foi cerca de 300% maior do que em 2004. Como consequência, a fumaça originada dos incêndios afetou o tráfego aéreo, escolas e empresas foram fechadas, e muitas pessoas foram atendidas nos hospitais devido à inalação da fumaça (MARENGO et al. 2008).

Vários municípios dos estados de Santa Catarina, do Rio de Janeiro e de Minas Gerais vieram a decretar estado de calamidade devido a fortes chuvas ocorridas nos meses de novembro e dezembro de 2008. Essas chuvas fortes causaram destruição total de muitas casas, deixando milhares de pessoas desabrigadas e centenas de mortos. Os prejuízos causados por esse evento pluvial extremo no sul do país, pelas subseqüentes enchentes e deslizamentos, foram estimados em 350 milhões de dólares, principalmente pelo fechamento do Porto de Paranaguá (MARENGO et al., 2010).

As fortes chuvas ocorridas em janeiro de 2011, na região serrana do estado do Rio de Janeiro, ocasionaram inundações e deslizamentos de terra, deixando aproximadamente 14 mil pessoas desabrigadas, além de ocasionar a morte de pessoas e animais. Os anos de 2014 e 2015 foram marcados pelo baixo regime de chuvas em quase todo país. Nesses dois anos, o verão foi anormalmente seco e quente. Os níveis dos reservatórios de algumas usinas hidrelétricas atingiram níveis mínimos críticos, sendo necessário, em alguns casos, a utilização do

chamado “volume morto”, para atender a demanda exigida. Um fato ainda mais preocupante nesse período foi a carência de água potável para atender o consumo humano. Em alguns municípios foi necessária a determinação de períodos em que a população teria acesso à água. O fim da primavera de 2015 e o início do verão de 2016 na região sul do Brasil têm sido marcados por chuvas e enchentes intensas, o que tem gerado prejuízos econômicos. A ocorrência desses eventos atípicos no sul do Brasil tem sido associado ao fenômeno *El Niño*, que corresponde a um superaquecimento das águas do oceano Pacífico.

Diante da necessidade do planejamento e cálculo de riscos de ocorrência de precipitações máximas raras para, por exemplo, atividades como a agricultura em Piracicaba-SP, objetivou-se no presente trabalho: *i*) verificar o ajuste das distribuições Generalizada de Pareto e Exponencial aos dados de precipitação máxima em Piracicaba-SP, nos meses de outubro a março, *ii*) calcular as precipitações máximas esperadas para os tempos de retorno de 10, 30, 50, 70 e 100 anos nesses meses, e *iii*) selecionar para cada mês, a distribuição que forneça resultados de maior acurácia.

5.2 Material e métodos

5.2.1 Dados

Os dados de precipitação pluviométrica (mm) foram obtidos a partir dos registros históricos da Estação Convencional do Posto Agrometeorológico da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP, 2015). Os registros históricos compreendem o período de 1917 a 2015, totalizando 99 anos de observações. O clima de Piracicaba, conforme a classificação de Köppen, é do tipo Cwa: tropical de altitude, com chuvas de verão e seca no inverno, sendo os meses de junho, julho e agosto mais secos. A temperatura média do mês mais quente

é superior a 22°C e a do mês mais frio não é inferior a 16°C (GHIBERTO; MORAES, 2011).

O município de Piracicaba situa-se entre as coordenadas geográficas de 22°42'30" de latitude Sul e de 47°38'01" de longitude oeste, com uma altitude média de 554 metros. A área total do município é de 1378 km² e a população é estimada em 364.571 habitantes, sendo que aproximadamente 97% encontram-se na área urbana (IBGE, 2010).

5.2.2 Composição das amostras de precipitações máximas mensais

A Teoria de Valores Extremos (TVE) constitui-se de duas metodologias para seleção de valores extremos. Na primeira, um valor extremo, máximo ou mínimo, é definido como a maior ou a menor observação em um determinado bloco. Em análises de valores extremos é comum considerar como blocos, períodos de tempo de interesse, como por exemplo, dias, meses ou anos. Nesse caso, de cada período é extraído o valor máximo (ou mínimo) que constituíra a amostra de valores extremos. Esse método é conhecido como método dos blocos máximos (ou mínimos). Na segunda abordagem da TVE, uma observação é considerada um valor extremo máximo (ou mínimo), quando excede um determinado limiar suficientemente alto (ou baixo), conhecido como *threshold*. Essa técnica de obtenção de valores extremos é conhecida como método dos picos acima de um limiar (POT, do inglês *Peaks Over Threshold*). As observações que excedem um determinado *threshold* u são chamadas de excessos ou excedências.

Nas Figuras 4 e 5 são ilustradas essas duas definições de valor extremo. Na Figura 4, as observações X_2 , X_5 , X_7 e X_{11} representam os máximos de quatro blocos, sendo que cada bloco possui três observações. Na Figura 5, as observações X_1 , X_2 , X_5 , X_7 , X_8 , X_9 e X_{11} excedem o *threshold* u , e dessa forma, constituem

o conjunto de valores extremos.

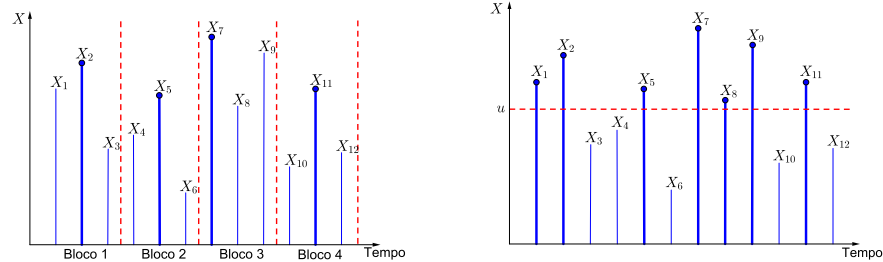


Figura 4 Máximos nos blocos 1, 2, 3 e Figura 5 Excessos acima de um limiar u .

No presente trabalho, foi utilizada a metodologia POT para a composição das amostras de precipitações máximas em Piracicaba-SP, nos meses de outubro a março. Esses meses foram escolhidos em função de apresentarem as maiores precipitações ao longo do ano na região de estudo. A forma como cada conjunto de precipitações máximas mensais foi constituído é abordada a seguir, após algumas considerações acerca da escolha do *threshold* u .

5.2.3 A escolha do limiar

Conforme Coles (2001, p. 79), a escolha de um limiar suficientemente alto nem sempre é tão simples. A escolha de um *threshold* u está sujeita ao que é denominado *trade-off* entre variância e viés. Um *threshold* baixo implica um maior número de observações. Neste caso, são introduzidas no conjunto de dados algumas observações do centro da distribuição. Como consequência do maior número de observações, os estimadores possuirão menor variância, porém com maior viés. Por outro lado, a escolha de um *threshold* alto reduz o viés dos estimadores, porém os tornam mais imprecisos (maior variância), devido à utilização de uma menor quantidade de observações para sua estimação.

Para seleção do limiar em cada um dos meses do estudo foi utilizado o método proposto por Davison e Smith (1990), que consiste na escolha de um limiar por meio da análise do gráfico da média dos excessos. Para a determinação visual do valor de u , foi utilizado o gráfico de vida média residual, que é construído da seguinte forma:

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{\max} \right\},$$

em que (x_1, \dots, x_{n_u}) são as n_u observações que excedem u .

Conforme Mendes (2004, p. 87), deve-se procurar aquele(s) valor(es) de u de tal maneira que acima dele a função média dos excessos se torne mais ou menos linear. Em adição a esse primeiro critério, um segundo procedimento a ser adotado para a seleção de um limiar é verificar se as pressuposições de independência das observações e de ajuste do modelo aos dados são satisfeitas. Em alguns casos, um valor candidato a *threshold* pode fornecer um conjunto de observações extremas que não satisfaz uma dessas pressuposições de análise. Na Figura 6 é apresentado o gráfico da média dos excessos considerando a variável climática precipitação máxima.

Pode-se verificar na Figura 6 que as médias dos excessos apresentam decréscimo até o nível pluviométrico de 45mm. A mudança na inclinação e a linearidade entre 45mm e 57mm indicam um conjunto de possíveis valores candidatos a *threshold*. Como mencionado anteriormente, após escolher-se um valor como *threshold*, deve-se verificar se o conjunto de valores extremos selecionados a partir desse valor satisfaz as pressuposições de análise. O comportamento irregular após o nível pluviométrico de 57mm deve-se ao pequeno número de excessos acima desse nível.

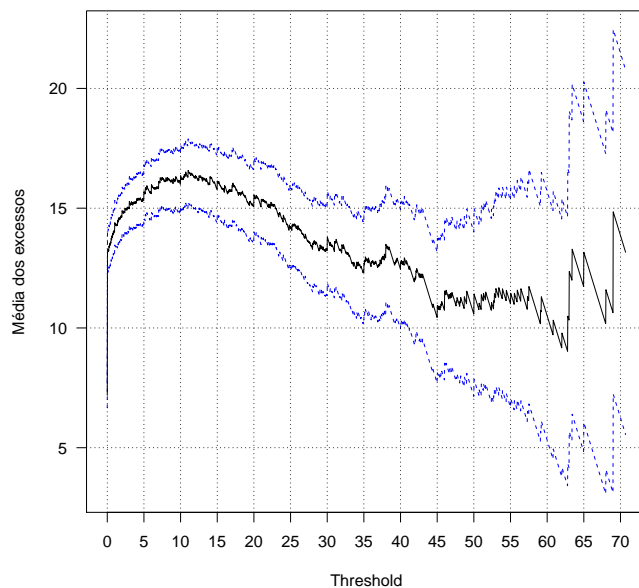


Figura 6 Gráfico da média dos excessos.

Após essas considerações acerca da escolha do *threshold*, pode-se proceder aos detalhes sobre a composição do conjunto de precipitações máximas para cada um dos meses do estudo. Para exemplificar, será utilizado o mês de janeiro, sendo análoga à descrição para os demais meses. Para utilizar a metodologia POT, o procedimento inicial foi formar o vetor com nk observações $\{x_i\}_i^{nk}$ para o mês de janeiro. Uma vez que o mês de janeiro possui 31 dias e o período de estudo é de 1917 a 2015, as quantidades n e k representam respectivamente a quantidade de precipitações diárias observadas desses mês ($n = 31$) e a quantidade de anos utilizados no estudo. O segundo procedimento foi selecionar o *threshold* u , com base no gráfico da média dos excessos. Assim, a amostra de precipitações máximas do mês de janeiro foi constituída, pela seleção dos n_u excessos acima do limiar u .

Os procedimentos descritos para a composição do conjunto de precipita-

ções máximas em janeiro foi aplicado de forma análoga nos demais meses. Logo, foram selecionados ao todo seis *thresholds*. A quantidade de anos com registros de precipitação diária variou entre os meses de outubro a março, devido à indisponibilidade de observações em alguns anos do estudo.

5.2.4 Análise exploratória dos dados

A análise exploratória dos dados baseou-se na estimação das medidas de tendência central (média e mediana), nas medidas de dispersão (variância, desvio padrão e coeficiente de variação (CV)) e nos coeficientes de assimetria e de curtose.

5.2.5 Teste de independência de Ljung-Box

A independência entre as observações em cada mês foi avaliada utilizando-se o teste proposto por Ljung e Box (1978). Conforme Ljung e Box (1978), o teste de Ljung-Box, também conhecido como teste de Portmanteau, é um teste que verifica se alguns grupos de autocorrelação de uma série temporal são diferentes de zero. Esse teste verifica a aleatoriedade total baseando-se no número de desvios, ao invés de testar a aleatoriedade de cada desvio distinto. Se representarmos o coeficiente de autocorrelação por r , temos que o par de hipóteses desse teste consiste em:

$$\begin{cases} H_0 : r_1 = r_2 = \dots = r_k = 0 \\ H_1 : \text{pelo menos um } r_k \text{ não é nulo.} \end{cases}$$

A estatística do teste de Ljung-Box é dada por:

$$Q = n(n+2) \sum_{k=1}^S \frac{r_k^2}{(n-k)}, \quad (5.1)$$

em que n é o número de observações, S é o número de coeficientes para se testar a autocorrelação e r_k é o coeficiente de autocorrelação para o desvio k .

A estatística em (5.1) distribuí-se como uma $\chi^2_{(S)}$ com S graus de liberdade. Escolhendo um nível de significância α , rejeita-se a hipótese H_0 de que todos os coeficientes de autocorrelação são iguais a zero se Q for superior ao valor crítico $\chi^2_{\alpha, S}$ de uma distribuição Qui-Quadrado.

5.2.6 Distribuição Generalizada de Pareto (GP)

Seja (X_1, \dots, X_n) uma amostra iid de uma variável aleatória X , que possui fda F_X . Para um determinado *threshold* u , um conjunto de n_u excedências será obtido a partir da amostra original. A nova amostra será construída através da transformação:

$$Y = X - u | X > u .$$

Se x_{\max} é a maior das observações de (X_1, \dots, X_n) , a fda dos excessos acima do nível u é caracterizada por:

$$H(y) = P(X - u \leq y | X > u) = \frac{P(u < X \leq y + u)}{P(X > u)} = \frac{F_X(y + u) - F_X(u)}{1 - F_X(u)},$$

em que $y > 0$ e $u < x_{\max}$, sendo H chamada de fda dos excessos.

Balkema e de Haan (1974) e Pickands (1975) demonstraram que a distribuição assintótica dos valores de uma variável aleatória que excedem um limiar pode ser aproximada por meio da distribuição Generalizada de Pareto (GP, do inglês *Generalized Pareto*). A fda da GP é dada por:

$$H(y; \sigma_u, \xi) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \text{para } y \geq 0 \text{ e } \xi > 0, \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \text{para } y \geq 0 \text{ e } \xi \rightarrow 0, \\ 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \text{para } 0 \leq y \leq -\frac{\sigma_u}{\xi} \text{ e } \xi < 0, \end{cases}$$

Em termos de X , a fda da GP pode ser reescrita como:

$$H(x; u, \sigma_u, \xi) = \begin{cases} 1 - \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \text{para } x \geq u \text{ e } \xi > 0, \\ 1 - \exp\left(-\frac{x-u}{\sigma_u}\right), & \text{para } x \geq u \text{ e } \xi \rightarrow 0, \\ 1 - \left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \text{para } u \leq x \leq u - \frac{\sigma_u}{\xi} \text{ e } \xi < 0, \end{cases}$$

em que u , σ_u e ξ denotam respectivamente os parâmetros de localização, escala e forma, com $-\infty < u < \infty$, $\sigma_u > 0$ e $\xi \in \mathbb{R} - \{0\}$. Para $\xi \rightarrow 0$, $\xi > 0$ e $\xi < 0$, a fda da GP reduz-se à classe de distribuições do tipo I ou Exponencial, do tipo II ou Pareto ou do Tipo III ou Beta, respectivamente.

Na Figura 7 são apresentados os três casos da distribuição GP, com destaque ao caso em que $\xi \rightarrow 0$ (Exponencial). Nessa figura pode-se observar que o comportamento das caudas é influenciado pelo parâmetro ξ . Esse parâmetro está relacionado com o peso da cauda da distribuição, e por essa razão, também é denominado índice de cauda. Quando $\xi < 0$, as distribuições se caracterizam por terem caudas leves e com suporte superior finito. Quando $\xi > 0$, as distribuições apresentam cauda pesada, possuindo limite superior do suporte infinito. A família Exponencial ($\xi \rightarrow 0$) corresponde ao caso intermediário, isto é, essa família possui caudas mais leves quando comparada com a família Pareto ($\xi > 0$), mas possui caudas mais pesadas em relação a família Beta ($\xi < 0$) (Figura 7). Essas consi-

derações acerca do parâmetro de forma ξ resultam em diferentes aplicações das três famílias de distribuições de valores extremos, em função de seus diferentes comportamentos na direção do extremo.

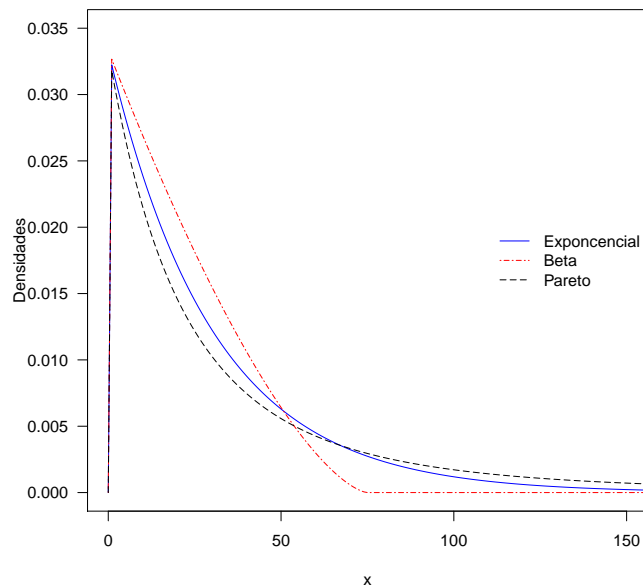


Figura 7 Função densidade de probabilidade da distribuição Generalizada de Pareto para $\xi = -0,4$ (Beta), $\xi = 0,4$ (Pareto) e $\xi \rightarrow 0$ (Exponencial), com $u = 0$ e $\sigma_u = 30$.

5.2.7 Estimação dos parâmetros

Os parâmetros das distribuições GP e Exponencial foram estimados utilizando-se método da máxima verossimilhança. Se (x_1, \dots, x_{n_u}) são os n_u excessos em relação a um *threshold* u , para $\xi \neq 0$, a log-verossimilhança é dada por:

$$l(u, \sigma_u, \xi | \mathbf{x}) = -n_u \log \sigma_u - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n_u} \log \left(1 + \xi \frac{x_i - u}{\sigma_u}\right),$$

em que $\left(1 + \xi \frac{x_i - u}{\sigma_u}\right) > 0$, para $i = 1, \dots, n_u$. Caso contrário, $l(u, \sigma_u, \xi | \mathbf{x}) = -\infty$. Para o caso em que $\xi \rightarrow 0$ (Exponencial), a função log-verossimilhança é dada por:

$$l(u, \sigma_u | \mathbf{x}) = -n_u \log \sigma_u - \sum_{i=1}^{n_u} \left(\frac{x_i - u}{\sigma_u}\right).$$

Para a obtenção das estimativas dos parâmetros, a maximização analítica da função de log-verossimilhança não é possível, o que exige a utilização de métodos numéricos (COLES, 2001, p. 80-81). Em particular, a fim de comparação, também será considerado o estimador MRE de locação-escala obtido na subeção 2.6, para a estimação dos parâmetros de locação e de escala da distribuição Exponencial. Esses estimadores são novamente apresentados:

$$\hat{u} = X_{(1)} - \frac{1}{n_u^2} \sum_{i=2}^{n_u} (X_{(i)} - X_{(1)}).$$

$$\hat{\sigma}_u = \frac{1}{n_u} \sum_{i=2}^{n_u} (X_{(i)} - X_{(1)}).$$

Para distinguir a distribuição Exponencial ajustada com estimativas obtidas via método da máxima verossimilhança e com estimativas equivariantes, será adotada a notação Exponencial_{MV} e Exponencial_E, respectivamente.

5.2.8 Teste de aderência de Kolmogorov-Smirnov

O ajuste das distribuições aos dados em cada mês foi avaliado utilizando-se o teste de Kolmogorov-Smirnov (KS). O teste de KS é um teste de aderência não paramétrico. A estatística desse teste, denotada por D , é usada para testar se uma amostra aleatória originou-se de uma população com uma função distribuição especificada. A estatística D baseia-se na diferença máxima das funções de probabilidades acumuladas, empírica e teórica (SIEGEL; CASTELLAN JÚNIOR, 2006).

Inicialmente, na hipótese nula é especificada alguma fda $F_X(x)$, sob a qual reside a hipótese de ter originado o conjunto de observações amostrais. O teste de KS adapta a fda hipotética $F_X(x)$ à fda empírica $S_n(x)$, determinando se a diferença máxima entre essas funções é significativa. Nesse sentido, uma amostra (x_1, \dots, x_n) é retirada de uma população, que não é conhecida. O confronto entre $S_n(x)$ e $F_X(x)$ é estabelecido para verificar se é razoável estudar os dados amostrais utilizando $F_X(x)$, admitida como a verdadeira fda da amostra casualizada.

Para a realização do teste de KS seguem-se os seguintes passos:

- (i) Colocar as observações em ordem ascendente.
- (ii) Calcular as frequências acumuladas em cada classe e estimar os parâmetros de $F_X(x)$.
- (iii) Obter as probabilidades acumuladas da distribuição teórica $F_X(x_{(i)})$ e da distribuição empírica $S_n(x_{(i)})$ em cada classe. Considerando a fda empírica $S_n(x)$, para uma determinada classe de frequência, $S_n(x) = fa/(n + 1)$, em que fa é a frequência acumulada da classe.
- (iv) Calcular a estatística D , que é dada por:

$$D = \max |F_X(x_{(i)}) - S_n(x)|.$$

- (v) A hipótese H_0 , de que os dados provêm de uma população com distribuição teórica $F_X(x)$, é testada pela comparação de D com o quantil superior da distribuição da estatística de D unilateral ou pela comparação do valor p com o nível de significância adotado (α). Se ao nível de significância α estabelecido, o valor observado de $D > D_{n,\alpha}$, rejeita-se H_0 . A hipótese nula também é rejeitada quando o valor p é inferior ao nível de significância α .

5.2.9 Teste da razão de verossimilhanças

Uma das dificuldades iniciais para a aplicação das três formas assintóticas de valores extremos é escolher, dentre as três famílias possíveis, a mais adequada para descrever um determinado conjunto de dados. A distribuição Exponencial é uma opção atrativa, por ser um modelo mais simples e por possuir índice de cauda nulo ($\xi = 0$). Essa distribuição corresponde a um modelo de transição entre a família de distribuições com $\xi < 0$, que possui limite superior do suporte finito, e a família de distribuições com $\xi > 0$, que possui limite superior de suporte infinito. Por essas razões, é interessante considerar a hipótese do modelo para os dados ser Exponencial contra a hipótese alternativa de não ser, o que pode ser expresso pelo seguinte par de hipóteses:

$$\begin{cases} H_0 : \xi = 0 \\ H_1 : \xi \neq 0 \end{cases}. \quad (5.2)$$

Para testar as hipóteses definidas em (5.2), o teste da razão de verossimilhanças (TRV) pode ser aplicado. Se \mathcal{M}_0 e \mathcal{M}_1 são dois modelos encaixados ou

aninhados ($\mathcal{M}_0 \subset \mathcal{M}_1$), a estatística do TRV é dada pela *deviance*, que é definida como:

$$\mathcal{D} = 2 \{l_1(\mathcal{M}_1) - l_0(\mathcal{M}_0)\}, \quad (5.3)$$

em que $l_0(\mathcal{M}_0)$ e $l_1(\mathcal{M}_1)$ são os valores do logaritmo da função de verossimilhança de \mathcal{M}_0 e \mathcal{M}_1 , respectivamente. A estatística \mathcal{D} tem distribuição assintótica $\chi^2_{(1)}$ com um grau de liberdade.

Neste trabalho, o TRV foi utilizado para avaliar se cada conjunto de observações mensais poderia ser descrito pelo modelo mais simples (Exponencial_{MV}), em detrimento do modelo mais geral (GP). Suponha que (x_1, \dots, x_{n_u}) sejam os n_u excessos para um *threshold* u . Se $l_1(u, \sigma_u, \xi | \mathbf{x})$ e $l_0(u, \sigma_u | \mathbf{x})$ são respectivamente os logaritmos das funções de verossimilhança das distribuições GP e Exponencial_{MV}, a estatística do TRV pode ser expressa como:

$$\mathcal{D} = 2 \left\{ l_1(\hat{u}, \hat{\sigma}_u, \hat{\xi} | \mathbf{x}) - l_0(\hat{u}, \hat{\sigma}_u | \mathbf{x}) \right\},$$

Para testar a hipótese $H_0 : \xi = 0$ contra $H_1 : \xi \neq 0$, compara-se o valor da estatística \mathcal{D} , calculada a partir de cada amostra mensal, com o quantil superior da distribuição $\chi^2_{(1)}$ ao nível de significância α . Se $\mathcal{D} \geq \chi^2_{\alpha,1}$, a hipótese H_0 é rejeitada.

5.2.10 Níveis de retorno

Seja (x_1, \dots, x_{n_u}) um conjunto de observações máximas. A probabilidade que ocorra um evento extremo superior a x é calculada pelo complementar da fda. Considerando inicialmente a distribuição GP, segue que:

$$P[X > x | X > u] = 1 - \left[1 - \left[1 + \xi \left(\frac{x - u}{\sigma_u} \right) \right]^{-\frac{1}{\xi}} \right] = \left[1 + \xi \left(\frac{x - u}{\sigma_u} \right) \right]^{-\frac{1}{\xi}}.$$

Disso segue que:

$$P[X > x] = \varsigma_u \left[1 + \xi \left(\frac{x - u}{\sigma_u} \right) \right]^{-\frac{1}{\xi}},$$

em que $\varsigma_u = P[X > u]$.

Conforme Coles (2001, p. 81), um nível x_m , que é excedido em média uma vez a cada m observações, é a solução de:

$$\varsigma_u \left[1 + \xi \left(\frac{x_m - u}{\sigma_u} \right) \right]^{-\frac{1}{\xi}} = \frac{1}{m}.$$

Reorganizando, resulta que:

$$x_m = u + \frac{\sigma_u}{\xi} \left[(m\varsigma_u)^\xi - 1 \right],$$

provando que m é suficientemente grande para assegurar que $x_m > u$.

A probabilidade que ocorra um evento extremo superior a x também pode ser calculado pelo complementar da fda da distribuição Exponencial ($\xi \rightarrow 0$), que é dado por:

$$P[X > x | X > u] = 1 - \left[1 - \exp \left(-\frac{x - u}{\sigma_u} \right) \right] = \exp \left(-\frac{x - u}{\sigma_u} \right).$$

Utilizando o mesmo raciocínio que foi realizado para a distribuição GP, tem-se que a função de quantis da distribuição Exponencial é dada por:

$$x_m = u + \sigma_u \log(m\varsigma_u),$$

novamente para m suficientemente grande.

Por construção, x_m é o nível de retorno da m -observação. Para apresentação das informações, é mais conveniente fornecer os níveis de retorno em uma escala anual. Desse modo, um nível de retorno anual z_T é o nível esperado para ser excedido uma vez a cada T anos (em média). Se existem n_y observações por ano, isso corresponde ao nível de retorno da m -observação, sendo $m = T \times n_y$.

Consequentemente, conhecidas as estimativas dos parâmetros, os níveis de retorno máximos para uma variável X , em um tempo médio T (em anos), podem ser calculados pelas distribuições GP e exponencial respectivamente por:

$$z_T = \hat{u} + \frac{\hat{\sigma}_u}{\hat{\xi}} \left[(Tn_y\hat{\varsigma}_u)^{\hat{\xi}} - 1 \right].$$

$$z_T = \hat{u} + \hat{\sigma}_u \log(Tn_y\hat{\varsigma}_u).$$

Pode-se verificar que também é necessária a estimação de ς_u . Como ς_u é a probabilidade de uma observação exceder o *threshold* u , um estimador natural é a proporção de pontos que excedem u , dada por:

$$\hat{\varsigma}_u = \frac{n_u}{n},$$

em que n é o tamanho da amostra original.

5.2.11 Seleção da distribuição com resultados mais acurados

Para analisar a distribuição que fornece precipitações máximas esperadas mais acuradas em cada mês, adotou-se o critério do menor erro quadrático médio (EQM), que é dado por:

$$EQM = \sum_{i=1}^n \frac{(\widehat{Pe}_i - Po_i)^2}{n}, \quad (5.4)$$

em que \widehat{Pe}_i é a precipitação máxima esperada no i -ésimo tempo de retorno, Po_i é a maior precipitação máxima observada no intervalo de tempo $(t ; t+T)$, sendo T o tempo de retorno (em anos) e n o número de valores de precipitação máxima comparados.

Para o cálculo do EQM, as observações de precipitação máxima em cada mês foram divididas em dois conjuntos. O primeiro de 1917-1986 foi utilizado para estimar os parâmetros e calcular as precipitações máximas prováveis para os tempos de retorno de 5, 10, 15, 20 e 25 anos. Do segundo conjunto, de 1987-2011, extraiu-se a maior precipitação máxima observada para os seguintes intervalos de tempo: 5 anos (1987 ; 1991), 10 anos (1987 ; 1996), 15 anos (1987 ; 2001), 20 anos (1987 ; 2006) e 25 anos (1987 ; 2011), para cada mês.

5.2.12 Recursos computacionais

Todas as análises estatísticas foram realizadas no software R Core Team (2015). Os pacotes *evd* (STEPHENSON, 2002) e *fBasics* (WUERTZ; SETZ; CHALABI, 2014) da biblioteca do R foram utilizados para o estudo dos dados. Em particular, o pacote *evd* foi utilizado para a análise dos dados, por possuir funções específicas na análise de valores extremos.

5.3 Resultados e discussão

Na Tabela 1 são apresentadas algumas estatísticas descritivas dos dados de precipitação máxima (mm) dos meses de outubro a março em Piracicaba-SP, relativos ao período de 1917 a 2015. Pode-se observar que o valor do CV nos períodos mensais analisados variou de 17,38 a 24,58. Desse modo, a variabilidade dos dados pode ser classificada como média ($12\% < CV < 60\%$), conforme a classificação dos limites do CV proposta por Warrick e Nielsen (1980). Pode-se verificar também que, em média, a precipitação máxima em Piracicaba-SP nos meses estudados variou de 48,89mm (outubro) a 71,24mm (dezembro). Nota-se também que as medianas dos dados em todos os meses são inferiores às médias amostrais observadas, o que sugere que as distribuições empíricas mensais são assimétricas à direita. Esse fato é reforçado com base nos valores positivos do coeficiente de assimetria. O grau de achatamento das distribuições empíricas, em todos os meses, foi classificada como platicúrtica, conforme os resultado do coeficiente de curtose (< 3).

Tabela 1 Estatísticas descritivas dos dados de precipitação máxima mensal (mm) em Piracicaba-SP, no período de 1917 a 2015.

| Mês | Mediana | Média | Variância | Desvio Padrão | Coef. de variação (%) | Coef. de assimetria | Coef. de curtose |
|-----------|---------|-------|-----------|---------------|-----------------------|---------------------|------------------|
| Janeiro | 65,00 | 69,53 | 146,10 | 12,09 | 17,38 | 1,25 | 0,46 |
| Fevereiro | 64,55 | 71,18 | 306,12 | 17,50 | 24,58 | 1,07 | 0,08 |
| Março | 55,40 | 58,64 | 175,54 | 13,24 | 22,59 | 1,50 | 2,77 |
| Outubro | 46,80 | 48,89 | 80,77 | 8,99 | 18,38 | 1,04 | 0,34 |
| Novembro | 61,50 | 67,79 | 211,98 | 14,56 | 21,47 | 1,03 | $< 0,01$ |
| Dezembro | 66,45 | 71,24 | 178,14 | 13,35 | 18,73 | 1,48 | 1,64 |

Coef.: Coeficiente

Na Tabela 2 são apresentados os *thresholds* escolhidos e as estimativas dos

parâmetros das distribuições GP e Exponencial_{MV}, obtidas utilizando-se o método da máxima verossimilhança. Analisando-se as estimativas do parâmetro de forma da distribuição GP, observa-se que suas estimativas pontuais são todas inferiores a zero, o que corresponde à distribuição Beta. Por possuir cauda superior com limite finito (Figura 7), essa distribuição é apropriada para modelar fenômenos ambientais, como precipitação pluviométrica máxima. Ainda na Tabela 2, pode-se verificar que o tamanho das amostras variou de 18 a 53 observações.

Tabela 2 *Thresholds* escolhidos e estimativas dos parâmetros das distribuições Generalizada de Pareto (GP) e Exponencial_{MV}, obtidas via método da máxima verossimilhança.

| Mês | <i>Threshold</i> (\hat{u}) | n | GP | | Exponencial _{MV} |
|-----------|--------------------------------|-----|------------------|-------------|---------------------------|
| | | | $\hat{\sigma}_u$ | $\hat{\xi}$ | $\hat{\sigma}_u$ |
| Janeiro | 57,00 | 34 | 13,62 | -0,08 | 12,53 |
| Fevereiro | 53,00 | 36 | 20,17 | -0,11 | 18,18 |
| Março | 43,00 | 53 | 18,18 | -0,16 | 15,64 |
| Outubro | 39,00 | 36 | 11,83 | -0,19 | 9,87 |
| Novembro | 52,00 | 18 | 18,86 | -0,19 | 15,79 |
| Dezembro | 57,00 | 40 | 15,62 | -0,10 | 14,24 |

Pela análise dos resultados do teste de Ljung-Box (Tabela 3), pode-se verificar que as observações em todos os meses estudados são independentes (valor $p > 0,05$). Sartori et al. (2010) não obtiveram resultados semelhantes no município de Botucatu-SP. Ao analisar séries históricas da Fazenda Lageado da UNESP de Botucatu-SP (1988-2007), referentes à precipitação pluviométrica, umidade relativa do ar e temperatura do ar, os autores verificaram que as observações das três variáveis climáticas estudadas apresentaram forte dependência temporal.

Os resultados do teste de KS (Tabela 3) indicam, para todos os meses sob estudo, que as distribuições GP e Exponencial_{MV} ajustaram-se aos dados (valor $p > 0,05$). Como a distribuição Exponencial é um caso especial da dis-

Tabela 3 Resultados dos testes (valor p) de independência (Ljung-Box), de aderência (Kolmogorov-Smirnov) e do teste da razão de verossimilhanças (TRV), referentes ao ajuste das distribuições Generalizada de Pareto (GP) e Exponencial $_{MV}$ aos dados de precipitação máxima mensal em Piracicaba-SP.

| Mês | Ljung-Box | Kolmogorov-Smirnov | | TRV |
|-----------|-----------|--------------------|---------------------|--------|
| | | GP | Exponencial $_{MV}$ | |
| Janeiro | 0,3510 | 0,5303 | 0,6426 | 0,7017 |
| Fevereiro | 0,1101 | 0,4109 | 0,5373 | 0,6516 |
| Março | 0,6397 | 0,6383 | 0,6233 | 0,2154 |
| Outubro | 0,1607 | 0,3899 | 0,5436 | 0,3858 |
| Novembro | 0,1485 | 0,5271 | 0,7249 | 0,5417 |
| Dezembro | 0,8874 | 0,5617 | 0,7071 | 0,5852 |

tribuição GP (modelos aninhados), verificou-se também o ajuste da distribuição Exponencial $_{MV}$ comparando-a com o modelo mais completo (GP), utilizando-se o TRV. Em todos os meses não houve evidências estatísticas para a rejeição da hipótese $H_0 : \xi = 0$ (valor $p > 0,05$). Neste caso, a distribuição Exponencial $_{MV}$ é o modelo mais adequado para modelar os dados de precipitação máxima considerados. Como a distribuição Exponencial $_{MV}$ mostrou-se mais adequada, procedeu-se também ao ajuste da distribuição Exponencial $_E$.

Na Tabela 4 estão apresentadas as estimativas equivariantes dos parâmetros de locação e de escala da distribuição Exponencial $_E$ e os resultados do teste de KS, referentes ao seu ajuste aos dados de precipitação máxima. Conforme os resultados do teste de KS, a distribuição Exponencial $_E$ ajustou-se aos dados em todos os períodos mensais considerados (valor $p > 0,05$). No município de Piracicaba-SP, verificou-se uma certa carência de estudos de valores extremos utilizando-se a metodologia POT e as distribuições que originam-se dessa metodologia. Os estudos observados na literatura utilizam principalmente as distribuições de valores extremos que originam-se do método dos blocos máximos, como as distribuições Gum-

bel e Generalizada de Valores Extremos (BAUTISTA; ZOCCHI; ANGELOCCI, 2004; SANSIGOLO, 2008).

Tabela 4 Estimativas equivariantes dos parâmetros da distribuição Exponencial $_E$ e resultados do teste Kolmogorov-Smirnov (valor p) referentes ao seu ajuste aos dados de precipitação máxima mensal em Piracicaba-SP.

| Mês | Estimativas | | Kolmogorov-Smirnov |
|-----------|-------------|------------------|--------------------|
| | \hat{u} | $\hat{\sigma}_u$ | |
| Janeiro | 56,94 | 12,23 | 0,7170 |
| Fevereiro | 52,70 | 17,98 | 0,6448 |
| Março | 43,11 | 15,24 | 0,6769 |
| Outubro | 38,93 | 9,69 | 0,6085 |
| Novembro | 51,55 | 15,39 | 0,8416 |
| Dezembro | 56,85 | 14,05 | 0,7763 |

Na Tabela 5 estão apresentadas as precipitações máximas (mm) esperadas em Piracicaba-SP para os tempos de retorno de 10, 30, 50, 70 e 100 anos, nos meses de outubro a março, obtidas via distribuições Exponencial $_{MV}$ e Exponencial $_E$. Uma interpretação dessas informações pode ser feita do seguinte modo: escolhendo o mês de março e o tempo de retorno de 30 anos, conforme o resultado obtido utilizando-se a distribuição Exponencial $_{MV}$, espera-se que em um tempo médio de 30 anos, que em pelo menos um dia desse mês ocorra uma precipitação máxima igual ou superior a 96,19mm.

Pode-se verificar também que as maiores precipitações máximas mensais são esperadas no mês de fevereiro, em todos os tempos de retorno (Tabela 5). Os resultados obtidos por Sansigolo (2008) corroboram com os resultados obtidos pelas distribuições Exponencial $_{MV}$ e Exponencial $_E$ no mês de fevereiro. Ao utilizar a distribuição Gumbel na modelagem das precipitações máximas anuais de Piracicaba-SP, no período de 1917 a 2004, Sansigolo (2008) verificou, para os tempos de retorno de 50 e 100 anos, que são esperadas ocorrências de precipitações

Tabela 5 Precipitações máximas (mm) esperadas em Piracicaba-SP para os tempos de retorno de 10, 30, 50, 70 e 100 anos, nos meses de outubro a março, obtidas via distribuições Exponencial_{MV} e Exponencial_E.

| Mês | Distribuição | Tempos de retorno | | | | |
|-----------|---------------------------|-------------------|--------|--------|--------|--------|
| | | 10 | 30 | 50 | 70 | 100 |
| Janeiro | Exponencial _{MV} | 85,86 | 99,63 | 106,03 | 110,24 | 114,71 |
| | Exponencial _E | 85,11 | 98,54 | 104,79 | 108,91 | 113,27 |
| Fevereiro | Exponencial _{MV} | 94,86 | 114,84 | 124,12 | 130,24 | 136,72 |
| | Exponencial _E | 94,10 | 113,86 | 123,04 | 129,09 | 135,50 |
| Março | Exponencial _{MV} | 79,01 | 96,19 | 104,18 | 109,44 | 115,02 |
| | Exponencial _E | 78,20 | 94,95 | 102,73 | 107,86 | 113,29 |
| Outubro | Exponencial _{MV} | 61,76 | 72,62 | 77,67 | 81,00 | 84,53 |
| | Exponencial _E | 61,23 | 71,88 | 76,82 | 80,08 | 83,54 |
| Novembro | Exponencial _{MV} | 88,36 | 105,70 | 113,77 | 119,08 | 124,71 |
| | Exponencial _E | 86,98 | 103,89 | 111,75 | 116,92 | 122,41 |
| Dezembro | Exponencial _{MV} | 89,80 | 105,45 | 112,73 | 117,52 | 122,60 |
| | Exponencial _E | 89,19 | 104,62 | 111,79 | 116,52 | 121,53 |

máximas iguais ou superiores a $119 \pm 9\text{mm}$ e $129 \pm 10\text{mm}$, respectivamente.

Pode-se observar que já a partir do tempo de retorno de 10 anos é esperada a ocorrência de precipitações superiores a 85mm nos meses de janeiro, fevereiro, novembro e dezembro. A ocorrência de precipitações pluviométricas dessa magnitude em um curto período de tempo podem ocasionar enchentes e inundações em Piracicaba-SP. A ocorrência desses fenômenos pode ser associada também a fatores relacionados à intervenção humana sobre o meio ambiente, como desmatamentos de encostas, ocupação de planícies de inundação, ao assoreamento dos rios e a impermeabilização urbana (MARCELINHO, 2008). Uma vez que não é possível evitar a ocorrência de eventos extremos de precipitação, a tomada de ações de correção e de prevenção, por órgãos responsáveis de Piracicaba-SP, podem minimizar os eventuais danos e prejuízos que venham a ser causados por inundações na área urbana desse município.

Na Tabela 6 são apresentados os resultados dos EQM's mensais para as distribuições Exponencial_{MV} e Exponencial_E. Primeiramente, foi necessário verificar se as observações mensais do período utilizado para estimação dos parâmetros e cálculo das precipitações máximas esperadas (1917-1986) satisfazem as pressuposições de análise. Pode-se observar que a pressuposição de independência das observações em cada mês foi satisfeita, conforme os resultados do teste de Ljung-Box (valor $p > 0,05$). Além disso, os resultados do teste de KS (valor $p > 0,05$) indicaram que as distribuições Exponencial_{MV} e Exponencial_E ajustaram-se aos dados em todos os meses. Como as pressuposições de independência e ajuste das distribuições foi satisfeita, pode-se proceder ao cálculo dos EQM's mensais.

Tabela 6 Erros quadráticos médios (EQM's) mensais referentes aos níveis máximos esperados de precipitação em Piracicaba-SP, obtidos via distribuições Exponencial_{MV} (Exp_{MV}) e Exponencial_E (Exp_E).

| Mês | Ljung-Box | Kolmogorov-Smirnov | | EQM | |
|-----------|-----------|--------------------|------------------|-------------------|------------------|
| | | Exp _{MV} | Exp _E | Exp _{MV} | Exp _E |
| Janeiro | 0,2575 | 0,6128 | 0,7253 | 356,71 | 391,52 |
| Fevereiro | 0,9667 | 0,8612 | 0,9003 | 26,86 | 35,19 |
| Março | 0,6057 | 0,5126 | 0,5474 | 222,24 | 239,23 |
| Outubro | 0,6064 | 0,6533 | 0,7308 | 110,08 | 117,16 |
| Novembro | 0,1863 | 0,7016 | 0,5260 | 428,97 | 260,68 |
| Dezembro | 0,7492 | 0,6751 | 0,7860 | 572,36 | 535,61 |

Pode-se observar que a distribuição Exponencial_{MV} forneceu menores EQM's nos meses de janeiro, fevereiro, março e outubro, quando comparada à distribuição Exponencial_E. Em termos práticos, esses resultados indicam que a distribuição Exponencial_{MV} possui maior acurácia na modelagem dos níveis máximo de precipitação nesses meses. Nesses meses, as estimativas obtidas via estimadores de máxima verossimilhança para os parâmetros da distribuição Exponencial são preferidos em relação às estimativas obtidas via estimadores equivariantes.

Nos demais meses (novembro e dezembro), o melhor desempenho foi alcançado pela distribuição Exponencial $_E$, sendo análoga a interpretação dos resultados à descrição realizada para a distribuição Exponencial $_{MV}$, isto é, nesses meses os estimadores equivariantes são preferidos para o ajuste da distribuição Exponencial em detrimento dos estimadores de máxima verossimilhança.

Mesmo não apresentando melhor desempenho em quatro meses estudados, um fato importante é que os estimadores equivariantes dos parâmetros da distribuição Exponencial fornecem mais uma opção para a modelagem de dados utilizando essa distribuição. Outros cenários de análise podem ser propostos para a distribuição Exponencial com estimadores equivariantes, envolvendo aplicações com outras variáveis. Além disso, outros critérios de avaliação dos estimadores podem ser adotados, em detrimento da utilização do EQM.

5.4 Conclusões

- (i) Em todos os períodos mensais analisados verificou-se o ajuste das distribuições GP, Exponencial $_{MV}$ e Exponencial $_E$ aos dados.
- (ii) A distribuição Exponencial $_{MV}$ mostrou-se mais adequada para modelar os dados em todos os períodos mensais, quando comparada à distribuição GP.
- (iii) Os maiores níveis de precipitação máxima em Piracicaba-SP são esperados para o mês de fevereiro.
- (iv) A distribuição Exponencial ajustada com estimativas de máxima verossimilhança apresentou maior acurácia nos meses de janeiro, fevereiro, março e outubro, enquanto a distribuição Exponencial ajustada com estimativas equivariantes apresentou melhor desempenho nos meses de novembro e dezembro.

6 CONSIDERAÇÕES FINAIS

Uma das justificativas para a abordagem do tema estimação equivariante neste trabalho foi a carência observada na literatura em português referente a estudos tratando desse importante tema da inferência estatística clássica. Esse assunto já vem sendo discutido há um longo período na literatura em geral. A teoria da estimação equivariante, para parâmetros de locação e de escala, iniciou-se com o trabalho de Edwin James George Pitman em 1939 e, a partir da década de 1950, uma discussão mais geral sobre esse assunto passou a receber contribuições de vários autores. Devido à carência de trabalhos sobre o assunto na literatura estatística brasileira, objetivou-se neste trabalho estudar aspectos da simetria no processo de estimação, considerando as ações de um grupo sobre o espaço paramétrico, o espaço amostral e o espaço das estimativas. Como a proposta inicial era oferecer um material que fosse o mais completo possível, a fundamentação teórica necessária para a determinação das fórmulas explícitas dos estimadores de Pitman, que são os estimadores com menor risco uniforme na classe dos estimadores equivariantes por locação e por escala, foi feita de forma mais detalhada possível. Além disso, para o enriquecimento do material foram apresentados exemplos de estimadores de Pitman para parâmetros de locação e de escala de algumas funções densidades muito conhecidas e utilizadas na literatura estatística. Espera-se que este estudo possa servir como uma referência a todos aqueles que se interessarem pela teoria da estimação equivariante.

As próximas etapas neste trabalho de pesquisa serão:

- (i) Estudar as propriedades de admissibilidade e minimaximalidade dos estimadores de Pitman.
- (ii) Estender a teoria da estimação equivariante para o contexto da inferência baye-

siana, pela definição de prioris invariantes.

REFERÊNCIAS

- ALENCAR FILHO, E. **Teoria dos Grupos**. São Paulo: Edgard Blücher. 1985. 280 p.
- BALKEMA, A. A.; DE HAAN, L. Residual life time at great age. **Ann. Probability**, n. 2, p. 792-804, 1974.
- BAUTISTA, E.A.L.; ZOCCHI, S.S.; ANGELOCCI, L.R. A distribuição Generalizada de Valores Extremos aplicada ao ajuste dos dados de velocidade máxima do vento em Piracicaba, São Paulo, Brasil. **Rev. Mat. Estat.**, São Paulo, v. 22, n. 1, p. 95-111, 2004.
- BERK, R. H. A special group structure and equivariant estimation. **Ann. Math. Statist.**, v. 38, p. 1436-1445, 1967.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. 2nd ed. Belmont, CA: Duxbury Thomson Learning, 2002. 660 p.
- COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. London: Springer, 2001. 228 p.
- DAVISON, A. C.; SMITH, R. L., Models for exceedances over high thresholds. **Journal of the Royal Statistical Society B**, n. 52, p. 393-442, 1990.
- ESALQ/USP, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. **Base de Dados Meteorológicos**. Disponível em: <<http://www.esalq.usp.br/departamentos/leb/posto.html>>. Acesso em: 05 jan. 2016.
- FERGUSON, T. S. **Mathematical statistics: a decision theoretic approach**. San Diego, CA: Academic Press, 1967. 408 p.
- GIRI, N. C. **Group invariance in statistical inference**. Singapore: World Scientific, 1997. 176 p.
- GHIBERTO, P. J.; MORAES, S. O. Comparação de métodos de determinação da Condutividade hidráulica em um latossolo Vermelho-amarelo. **R. Bras. Ci. Solo**,

Viçosa, v. 35, p. 1177-1188, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Banco de dados Cidades@**. 2010. Disponível em:

<<http://www.ibge.gov.br/cidadesat/xtras/temas.php?codmun=353870&idtema=16&search=sao-paulo|piracicaba|sintese-das-informacoes>>. Acesso em: 07 jan. 2016.

LEHMANN, E. L. **Theory of point estimation**. New York: John Wiley and Sons, 1983. 506 p.

LEHMANN, E. L.; CASELLA, G. **Theory of point estimation**. 2nd ed. New York: Springer-Verlag, 1998. 590 p.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, p. 297-303, 1978.

MAKANI, S. M. A paradox in admissibility. **The Annals of Statistics**, v. 5, n. 3, p. 544-546, 1977.

MARCELINHO, E. V. **Desastres naturais e geotecnologias: conceitos básicos**. Santa Maria: Instituto Nacional de Pesquisas Espaciais - INPE/CRS, 2008.

MARENGO, J. A.; NOBRE, C. A.; TOMASELLA, J.; CARDOSO, M. F. OYAMA, M. D. Hydro-climatic and ecological behaviour of the drought of Amazonia in 2005. **Phil. Trans. R. Soc. B**, v. 363, p. 1773-1778, 2008.

MARENGO, J. A. Mudanças climáticas, condições meteorológicas extremas e eventos climáticos no Brasil. In: MARENGO, J. A.; SCHAEFFER, P.; PINTO, H. S.; ZEE, D. M. W.; LUCENA, A. F. P.; SZKLO, A. **Mudanças climáticas e eventos extremos no Brasil**. Rio de Janeiro: Fundação Brasileira para o Desenvolvimento Sustentável (FBDS) e Lloyd's, 2010. p. 4-19. Disponível em: <<http://www.fbds.org.br/fbds/IMG/pdf/doc-504.pdf>>. Acesso em: 20 dez. 2015.

MARTIN, P. A. **Grupos, corpos e teoria de Galois**. São Paulo: Livraria da Física. 2010. 430 p.

MENDES, B. V. M. **Introdução à análise de eventos extremos**. São Paulo:

E-papers Editora, 2004. 232 p.

MOOD, A. M.; GRAYBILL, F. A. ; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. New York: McGraw-Hill, 1974. 564 p.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia Estatística**. Belo horizonte: CRPM, 2007. 552 p.

NOBRE, J.; AZEVEDO, C. O princípio da equivariância: conceitos e aplicações. **Revista Colombiana de Estadística**, v. 29, n. 2. p. 195-220, 2006.

PICKANDS, J. III. Statistical inference using extreme order statistics. **Annals of Statistics**, n. 3, p.119-131, 1975.

PITMAN, E. J. G. The estimation of the location and scale parameters of a continuous population of any given form. **Biometrika**, v. 30, p. 391-421, 1939.

PRABAKARAN, T.; CHANDRASEKAR, B. Simultaneous equivariant estimation for location-scales models. **Journal of Statistical Planning and Inference**, v. 40, p. 51-59, 1994.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria, 2015. Disponível em:
<<http://www.r-project.org>>. Acesso em: 15 out. 2015.

RESNIK, M. D. **Choices: An introduction to decision theory**. Minneapolis: University of Minnesota Press, 1987. 222 p.

SANSIGOLO, C. A. Distribuições de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006). **Revista Brasileira de Meteorologia**, São José dos Campos, v. 23, n. 3, p. 341-346, 2008.

SARTORI, A. A. C.; SILVA, A. F.; RAMOS, C. M. C.; ZIMBACK, C. R. L. Variabilidade temporal e mapeamento dos dados climáticos de Botucatu-SP. **Irriga**, Botucatu, v. 15, n. 2, p. 131-139, 2010.

SCHERVISH, M. J. **Theory of Statistics**. New York: Springer-Verlag, 1995. 702 p.

SIEGEL, S.; CASTELLAN JÚNIOR, N. J. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006. 448 p.

SOUZA, D. J. **O uso de grupos de permutação no estudo estatístico de sequências simbólicas**. 2009. 194 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2009.

STEPHENSON, A. G. evd: Extreme Value Distributions. **R News**, v. 2, n. 2, p. 31-32. 2002. Disponível em:
<<http://CRAN.R-project.org/doc/Rnews/>>. Acesso em: 22 dez. 2015.

VIANA, M. A. G. **Symmetry studies: an introductory to the analysis of structured data in applications**. New York: Cambridge University Press, 2008. 235 p.

ZACKS, S. **The theory of statistical inference**. New York: John Wiley & Sons, 1971. 609 p.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of physical properties in the field. In: HILLEL, D. **Applications of soil physics**. New York: Academic Press, 1980. p. 319-344.

WIJSMAN, R. A. Cross-sections of orbits and their application to densities of maximal invariants. **Proc. 5th Berkeley Symp. Math. Statist. Prob.**, v. 1, p. 389-400, 1967.

WUERTZ, D.; SETZ, T.; CHALABI, Y. **Rmetrics - Markets and Basic Statistics**. 2014. Disponível em:
<<http://CRAN.R-project.org/package=fBasics>>. Acesso em: 22 dez. 2015.

7 APÊNDICE

Os conceitos e resultados deste Apêndice também podem ser encontrados em Mood, Graybill e Boes (1974), Alencar Filho (1985), Casella e Berger (2002) e Martin (2010). Neste caso, as citações dessas referências serão usualmente omitidas.

A - NOÇÕES GERAIS SOBRE GRUPOS

Definição (Grupo): Um conjunto não vazio G munido de uma operação binária “ \cdot ” é um *grupo* se essa operação satisfaz as seguintes propriedades:

- (i) **Associatividade:** $(a \cdot b) \cdot c = a \cdot (b \cdot c)$, quaisquer que sejam $a, b, c \in G$;
- (ii) **Existência de elemento neutro:** existe um elemento $e \in G$ tal que $a \cdot e = e \cdot a = a$, qualquer que seja $a \in G$;
- (iii) **Existência de simétricos:** para todo $a \in G$ existe um elemento $a' \in G$ tal que $a \cdot a' = a' \cdot a = e$.

Além disso, se G é um grupo e a operação “ \cdot ” é comutativa, ou seja, $a \cdot b = b \cdot a$ para quaisquer $a, b \in G$, o grupo é denominado *grupo comutativo* ou *abeliano*, nome em referência ao matemático norueguês do século XIX Niel Henrik Abel.

Grupos de permutação e o grupo simétrico S_n

Na teoria dos grupos chama-se *permutação* a bijeção de um conjunto nele mesmo. Se A é um conjunto não vazio, pode-se denotar por $S(A)$ o conjunto das permutações do conjunto A . Nesse contexto, a composição de aplicações é uma operação sobre $S(A)$. De fato, se f e g são permutações em A , isto é, $f : A \rightarrow A$ e $g : A \rightarrow A$ são bijeções, então a composição $g \circ f : A \rightarrow A$ também é uma bijeção.

Teorema: O conjunto de todas as permutações de um conjunto A , $S(A)$, munido da operação composição de funções (\circ) é um grupo.

Prova:

Sejam $f : A \rightarrow A$, $g : A \rightarrow A$ e $h : A \rightarrow A$ três permutações quaisquer de A . É possível verificar que:

(i) Associatividade:

$$\begin{aligned} f \circ (g \circ h)(x) &= f[(g \circ h)(x)] \\ &= f[g(h(x))] \\ &= [(f \circ g)(h(x))] \\ &= (f \circ g) \circ h(x). \end{aligned}$$

(ii) Existência de elemento neutro: Dado que $A \neq \emptyset$, chama-se aplicação idêntica de A a aplicação $i_A : A \rightarrow A$, dada pela lei de formação $i_A(x) = x$, para todo $x \in A$. Temos dessa forma que $i_A : A \rightarrow A$ é a permutação idêntica de A . De fato,

$$(i_A \circ f)(x) = i_A(f(x)) = f(x),$$

e, da mesma forma,

$$(f \circ i_A)(x) = f(i_A(x)) = f(x).$$

(iii) Existência de simétricos: Se f é uma permutação de A , então o mesmo acontece com f^{-1} (aplicação inversa de f), que também é uma bijeção,

sendo f^{-1} o elemento inverso de f para a composição de aplicações, pois $f \circ f^{-1} = f^{-1} \circ f = i_A$. De fato, como $f : A \rightarrow A$ é bijetora, decorre também que f^{-1} é uma aplicação de A em A . Além disso, são válidas as igualdades $f(x) = y$ e $f^{-1}(y) = x$. Consequentemente,

$$f(f^{-1}(y)) = y \text{ e } f^{-1}(f(x)) = x,$$

isto é,

$$(f \circ f^{-1})(y) = y \text{ e } (f^{-1} \circ f)(x) = x.$$

Disso decorre que $f \circ f^{-1} = i_A$ e $f^{-1} \circ f = i_A$. Portanto, $(S(A), \circ)$ é um grupo.

Um importante grupo de permutações é aquele em que $A = \{1, 2, \dots, n\}$, com $n \geq 1$. Para esse caso, usa-se a notação S_n , em vez da notação genérica $(S(A))$, para indicar o conjunto das permutações sobre A . O grupo (S_n, \circ) tem um nome especial: grupo simétrico de grau n . Utilizando-se análise combinatória, pode-se verificar que esse grupo possui ordem $n!$, ou seja, $n!$ é o número de permutações que podem ser construídas com n elementos. Essas permutações podem ser colocadas em correspondência biunívoca com os elementos de S_n .

Se $A = \{1, 2, \dots, n\}$, uma permutação f de A pode ser representada pela notação funcional:

$$f = \{(1, f(1)), (2, f(2)), \dots, (n, f(n))\},$$

ou pela usual notação de duas linhas:

$$f = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ f(1) & f(2) & f(3) & \dots & f(n) \end{pmatrix}.$$

Com essas notações estabelecidas, a permutação idêntica de A é escrita como:

$$I_A = \{(1, 1), (2, 2), (3, 3) \dots, (n, n)\},$$

ou

$$f = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ 1 & 2 & 3 & \dots & n \end{pmatrix}.$$

A permutação inversa de uma permutação $f : A \rightarrow A$ é a função bijetora $f^{-1} : A \rightarrow A$ definida como:

$$f^{-1} = \{(y, x) \mid f(x) = y\}.$$

Se o conjunto $A = \{1, 2, \dots, n\}$, então:

$$\begin{aligned} f^{-1} &= \{(f(1), 1), (f(2), 2), \dots, (f(n), n)\} = \\ &= \{(1, f^{-1}(1)), (2, f^{-1}(2)), \dots, (n, f^{-1}(n))\}, \end{aligned}$$

ou seja:

$$\begin{aligned}
 f^{-1} &= \begin{pmatrix} f(1) & f(2) & f(3) & \dots & f(n) \\ 1 & 2 & 3 & \dots & n \end{pmatrix} = \\
 &= \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ f^{-1}(1) & f^{-1}(2) & f^{-1}(3) & \dots & f^{-1}(n) \end{pmatrix}.
 \end{aligned}$$

Sejam f e g permutações de A dadas por:

$$f = \begin{pmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{pmatrix} \text{ e } g = \begin{pmatrix} 1 & 2 & \dots & n \\ j_1 & j_2 & \dots & j_n \end{pmatrix}.$$

A composição dessas permutações é feita da seguinte maneira:

$$\begin{aligned}
 g \circ f &= \begin{pmatrix} 1 & \dots & i_r & \dots & n \\ j_1 & \dots & j_{i_r} & \dots & j_n \end{pmatrix} \circ \begin{pmatrix} 1 & \dots & r & \dots & n \\ i_1 & \dots & i_r & \dots & i_n \end{pmatrix} \\
 &= \begin{pmatrix} \dots & r & \dots \\ \dots & j_{i_r} & \dots \end{pmatrix},
 \end{aligned}$$

pois $(g \circ f)(r) = g(f(r)) = g(i_r) = j_{i_r}$.

Exemplo: Considere o grupo S_4 . Assim,

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Pode-se observar, por exemplo, que a imagem de 2 pela composta se obtém da seguinte maneira $2 \mapsto 1 \mapsto 2$.

Uma forma de dispor os elementos de um grupo finito e verificar as suas propriedades é utilizando uma tábua de dupla entrada.

Definição (Tábua de um grupo finito): Seja $G = \{a_1, a_2, \dots, a_n\}$ um grupo finito munido pela operação “ \cdot ”. Uma tábua do grupo (G, \cdot) é uma tábua de dupla entrada que indica o resultado correspondente a operação para cada par (a_i, a_j) de elementos de G .

| | | | | | | |
|----------|----------|----------|----------|----------|----------|----------|
| \cdot | a_1 | a_2 | \dots | a_j | \dots | a_n |
| a_1 | a_{11} | a_{12} | \dots | a_{1j} | \dots | a_{1n} |
| a_2 | a_{21} | a_{22} | \dots | a_{2j} | \dots | a_{2n} |
| \vdots | \vdots | \vdots | \ddots | \vdots | \ddots | \vdots |
| a_i | a_{i1} | a_{i2} | \dots | a_{ij} | \dots | a_{in} |
| \vdots | \vdots | \vdots | \ddots | \vdots | \ddots | \vdots |
| a_n | a_{n1} | a_{n2} | \dots | a_{nj} | \dots | a_{nn} |

As tábuas dos grupos S_2 e S_3 são apresentadas no exemplo a seguir.

Exemplo: Considere os grupos S_2 e S_3 . A construção das tábuas para esses grupos é muito laboriosa, envolvendo muitos cálculos. Por essa razão, para cada grupo será apresentada uma composição somente.

Primeiramente, vamos considerar a tábua de S_2 . Seja:

$$S_2 = \left\{ f_0 = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, f_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\}.$$

Segue então que, por exemplo, $f_1 \circ f_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} = f_0$. Logo, a tábua de S_2 é dada por:

| | | |
|---------|-------|-------|
| \circ | f_0 | f_1 |
| f_0 | f_0 | f_1 |
| f_1 | f_1 | f_0 |

A tábua de S_3 pode ser obtida de forma análoga. De fato, seja S_3 definido da forma:

$$S_3 = \left\{ \begin{array}{l} f_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, f_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, f_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \\ f_3 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, f_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, f_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \end{array} \right\}.$$

Pode-se observar, por exemplo, como se obtém $f_1 \circ f_5$. Desse modo:

$$f_1 \circ f_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = f_4.$$

As demais composições podem ser obtidas da mesma forma. Ao se realizar as demais composições e dispondo-se esses resultados numa tábua, tem-se a tábua de S_3 :

| | | | | | | |
|---------|-------|-------|-------|-------|-------|-------|
| \circ | f_0 | f_1 | f_2 | f_3 | f_4 | f_5 |
| f_0 | f_0 | f_1 | f_2 | f_3 | f_4 | f_5 |
| f_1 | f_1 | f_2 | f_0 | f_5 | f_3 | f_4 |
| f_2 | f_2 | f_0 | f_1 | f_4 | f_5 | f_3 |
| f_3 | f_3 | f_4 | f_5 | f_0 | f_1 | f_2 |
| f_4 | f_4 | f_5 | f_3 | f_2 | f_0 | f_1 |
| f_5 | f_5 | f_3 | f_4 | f_1 | f_2 | f_0 |

Pode-se verificar que o grupo (S_3, \circ) não é abeliano. De fato, considerando as funções f_1 e f_3 , podemos notar que:

$$f_1 \circ f_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = f_5$$

$$f_3 \circ f_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = f_4$$

Portanto, $f_1 \circ f_3 \neq f_3 \circ f_1$, ou seja, o grupo (S_3, \circ) não é abeliano.

Ação de grupos, órbita e transitividade

O grupo $S(X)$ das permutações de um conjunto X possui um importante papel na teoria dos grupos. Esse grupo está associado ao conjunto X , o qual tem seus elementos “transformados” pelos elementos de $S(X)$. A generalização da relação entre o grupo $S(X)$ e o conjunto X pode ser feita mediante o conceito de ação de um grupo, que indica como um grupo G atua em um conjunto X .

Definição (Ação de um grupo): Seja G um grupo e X um conjunto não vazio.

Uma *ação* de um grupo G em um conjunto X é uma função $G \times X \rightarrow X$, com $(g, x) \mapsto g \cdot x$, tal que as seguintes condições são satisfeitas:

- (i) $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$, para todo $g_1, g_2 \in G$ e $x \in X$,
- (ii) $e_G \cdot x = x$, para todo $x \in X$, em que e_G é a identidade de G .

A ação de G em X é representada por $g \cdot x$ para distinguir do produto usual gx . Dizemos que G age sobre X e que X é um G -conjunto. A definição de ação de um grupo fornece uma interessante ideia sobre a interação de um grupo e um conjunto não vazio X qualquer. Essa definição fornece a noção de que grupos operam ou agem em determinados conjuntos, alterando os seus elementos de lugar. Nesse sentido, podemos notar que a função identidade $e : X \rightarrow X$ opera fixando cada elemento de X , isto é, $e \cdot x = x$.

Definição (Órbita de um grupo): chama-se *órbita* de um grupo G , que atua em um conjunto X , o subconjunto de X tal que:

$$o(x) = \{g \cdot x : g \in G, x \in X\}.$$

Outra forma de definir órbita de um grupo é utilizar a definição de equivalência. Para um grupo G de permutações de X , dois pontos $x_1, x_2 \in X$ são *equivalentes* se existe um $g \in G$ tal que $g \cdot x_1 = x_2$. A totalidade de pontos equivalentes para um dado ponto, e conseqüentemente para qualquer outro, é chamado órbita de G .

A ação de um grupo G sobre X é *transitiva* se existe somente uma órbita. Nesse caso, uma ação de um grupo G em um conjunto X é dita transitiva quando uma das condições equivalentes i), ii) e iii) ocorre:

- (i) $\exists x \in X$ tal que $o(x) = X$.
- (ii) Para todo $x, y \in X, \exists g \in G$ tal que $x = g \cdot y$.
- (iii) Para todo $x \in X, o(x) = X$.

B - ABORDAGEM BAYESIANA NA ESTIMAÇÃO

O teorema de Bayes representa um importante resultado na estatística matemática. Suas origens estão relacionadas a publicação do artigo *An essay towards solving a problem in the doctrine of chances* (Ensaio buscando resolver um problema na doutrina das probabilidades) de Thomas Bayes (?-1761). Pouco se conhece sobre a vida de Thomas Bayes. Sabe-se que ele foi um reverendo presbiteriano que viveu na Inglaterra do século XVIII. Um ponto interessante sobre o teorema de Bayes é que esse resultado somente foi conhecido pela comunidade científica inglesa de seu tempo graças à Richard Price (1723-1791). Richard Price era filósofo e amigo de Thomas Bayes. Em 1763, dois anos após a morte de Bayes, Price encontrou entre os papéis do reverendo o artigo *An essay towards solving a problem in the doctrine of chances*. O artigo continha a demonstração do famoso teorema. Price foi o responsável por apresentar o artigo a Royal Society. Mas coube ao matemático francês Pierre-Simon de Laplace (1749-1827) a divulgação do teorema de Bayes para a comunidade científica em geral. Isso ocorreu porque após a publicação do artigo de Bayes pela Royal Society, o artigo caiu no esquecimento e coube a Laplace o seu resgate em 1812, em seu livro *Théorie Analytique des Probabilités*.

Considere uma partição finita A_1, A_2, \dots, A_n do espaço amostral \mathfrak{X}^n , em que $P(A_i) > 0$, $A_i \cap A_j = \emptyset$ ($i \neq j$) e $\bigcup_i A_i = \mathfrak{X}^n$. Considere também um evento B qualquer, tal que $P(B) > 0$ e tal que esse evento possa ser decomposto na união de conjuntos disjuntos A_i 's, isto é, $B = \bigcup_i (A_i \cap B)$. Com esses elementos, o teorema de Bayes pode ser enunciado da seguinte forma:

$$P(A_i | B) = \frac{P(B | A_i) P(A_i)}{P(B)} = \frac{P(B | A_i) P(A_i)}{\sum_{i=1}^n P(B | A_i) P(A_i)}, \quad (7.1)$$

em que:

$P(A_i | B)$ são probabilidades a posteriori;

$P(A_i)$ são probabilidades a priori;

$P(B | A_i)$ são probabilidades condicionais.

A importância do teorema de Bayes fundamenta-se no fato desse teorema conectar a inferência racional, ou seja, a probabilidade a posteriori (lado esquerdo do teorema), à subjetividade ou crença do pesquisador (probabilidade a priori) mais as informações empíricas (probabilidades condicionais), ambas do lado direito. Dito de outra forma, o teorema de Bayes é um dos poucos teoremas matemáticos que associa, pela probabilidade a posteriori, a crença anterior do pesquisador mais as informações que se originam com a observação dos dados. Além disso, o teorema de Bayes é o fundamento de todo um ramo da estatística conhecido como estatística bayesiana. Nesse ponto, é necessário abordar de forma bem geral a diferença básica que distingue a filosofia da inferência bayesiana em relação à inferência clássica, com respeito a um parâmetro θ .

Na análise estatística clássica, assume-se que θ é uma quantidade que indexa uma família de distribuições $\{f(\cdot; \theta)\}$, sendo essa quantidade um valor fixo e desconhecido. Busca-se dessa forma obter informações sobre o parâmetro populacional, visando-se obter generalizações para a população à partir de uma amostra representativa da mesma. Todavia, em diversas situações é comum que o pesquisador possua alguma informação sobre θ . Nesse caso, associada a pressuposição de que uma amostra aleatória (X_1, \dots, X_n) de uma característica de interesse provém de uma população com distribuição $f_X(x; \theta)$, tem-se também que o parâmetro θ é um valor de uma variável aleatória, digamos Θ . Suponha que o interesse seja estimar alguma função de θ , ou seja, $\omega(\theta)$. Como Θ é uma variável aleatória, ela possui uma distribuição. Dessa maneira, sejam $G(\cdot) = G_\Theta(\cdot)$ e $g(\cdot) = g_\Theta(\cdot)$

respectivamente a fda e a fdp de Θ . Assumi-se nesse caso que ambas funções não possuem parâmetros desconhecidos.

Como mencionado, na abordagem bayesiana θ representa o valor da variável Θ , cuja variabilidade pode ser descrita pela distribuição $g_{\Theta}(\cdot)$. Essa distribuição é subjetiva, isto é, baseia-se na crença (ou experiência) do pesquisador. A formulação da distribuição $g_{\Theta}(\cdot)$ ocorre antes da observação dos dados e por essa razão, ela é conhecida como *distribuição a priori*. Quando uma amostra é coletada de uma população indexada por θ , a distribuição a priori é atualizada com as informações adquiridas da amostra. A distribuição a priori atualizada é, por sua vez, chamada de *distribuição a posteriori*.

Uma observação sobre a notação: Na estatística bayesiana é comum que as distribuições a priori e a posteriori sejam denotadas respectivamente por $\pi(\cdot)$ e $\pi(\cdot | \mathbf{x})$. Nesse caso, a partir desse ponto será adotada a notação mais comum na literatura. Além disso, na inferência clássica é comum a utilização da notação $f_X(x; \theta)$ para indicar a densidade de uma variável aleatória X , para cada $\theta \in \Theta$. Entretanto, para indicar que o parâmetro θ é o valor de uma variável aleatória Θ , será adotada a notação $f_{X|\Theta}(x|\theta)$, que corresponde a densidade condicional de X dado $\Theta = \theta$.

A atualização da priori ocorre pela aplicação direta do teorema de Bayes (7.1), o que justifica o nome estatística bayesiana. Suponha que uma amostra (x_1, \dots, x_n) tenha sido observada. Considerando uma fdp f qualquer indexada por θ e uma priori $\pi(\theta)$ estabelecida pelo pesquisador, o teorema de Bayes pode ser expresso em termos de densidades, o que é o resultado da definição a seguir.

Definição (Distribuição a posteriori): Se $\pi(\theta)$ é a distribuição a priori de Θ , chama-se *distribuição a posteriori*, a densidade condicional de Θ dada a

amostra \mathbf{x} , que pode ser expressa por:

$$\begin{aligned}\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) &= \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta)\pi(\theta)d\theta}, \quad \theta \in \Theta.\end{aligned}\quad (7.2)$$

Quando a amostra é iid, a posteriori é dada por:

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{\left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i|\theta) \right] \pi(\theta)}{\int_{\Theta} \left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i|\theta) \right] \pi(\theta) d\theta}.$$

A distribuição a posteriori representa a distribuição de Θ , após uma amostra $\{x_1, \dots, x_n\}$ ter sido observada. Uma característica importante e que difere uma análise bayesiana de uma análise estatística clássica é que, a distribuição a posteriori é utilizada para fazer inferências acerca de θ , que é um escalar ou vetor aleatório. Nesse sentido, a atitude inicial do pesquisador diante de um problema, caracterizado por seu conhecimento prévio fornecido em $\pi(\theta)$, modifica-se em consideração as informações contidas na amostra observada. Essa nova atitude do pesquisador diante do problema sob investigação é traduzida pela distribuição a posteriori $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$.

Se o objetivo reside em estimar θ , a média da distribuição a posteriori pode ser empregada como uma estimativa. Em geral, pode-se estimar $\omega(\theta)$ utilizando a média de $\omega(\Theta)$, após a amostra ter sido observada, ou seja, $E[\omega(\Theta)|x_1, \dots, x_n]$.

Definição (Estimador de Bayes a posteriori): O *estimador de Bayes* de $\omega(\theta)$ em relação a priori $\pi(\theta)$ é definido como:

$$\begin{aligned}
E[\omega(\Theta) | X_i = x_i] &= \int_{\Theta} \omega(\theta) \pi_{\Theta|\mathbf{X}}(\theta | \mathbf{x}) d\theta \\
&= \int_{\Theta} \omega(\theta) \frac{\left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i | \theta) \right] \pi(\theta)}{\int \left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i | \theta) \right] \pi(\theta) d\theta} d\theta \\
&= \frac{\int_{\Theta} \omega(\theta) \left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i | \theta) \right] \pi(\theta) d\theta}{\int_{\Theta} \left[\prod_{i=1}^n f_{\mathbf{X}|\Theta}(x_i | \theta) \right] \pi(\theta) d\theta}.
\end{aligned}$$

As informações adicionais adquiridas com a distribuição a priori podem ser utilizadas em associação com as funções perda e risco para selecionar um estimador ótimo. Mas antes é necessário definir o risco de Bayes de um estimador.

Definição (Risco de Bayes): Seja (X_1, \dots, X_n) uma amostra aleatória de uma densidade $f_{\mathbf{X}|\Theta}(\mathbf{x}|\Theta)$. Chama-se *risco de Bayes* de um estimador $\delta(\mathbf{X})$, em relação a função perda $L(\cdot; \cdot)$ e distribuição a priori $\pi(\cdot)$, a função dada por:

$$r(\delta) = r_{L,\pi}(\delta) = \int_{\Theta} \mathcal{R}(\delta, \theta) \pi(\theta) d\theta.$$

O risco de Bayes de um estimador é um número real útil para comparação e seleção de um estimador. De forma análoga a comparação de estimadores com base na função risco clássica, ao comparar-se dois estimadores será adotado aquele que possua menor risco de Bayes. Uma diferença importante é que a adoção do risco de Bayes, com uma função perda e uma priori adequada, permite, de fato, a obtenção de um estimador com risco mínimo.

Definição (Estimador de Bayes): Seja δ' um estimador qualquer de $\omega(\theta)$. O estimador δ_π é o *estimador de Bayes* de $\omega(\theta)$ se:

$$r(\delta_\pi) \leq r(\delta'),$$

dado qualquer outro estimador δ' de $\omega(\theta)$.

O estimador de Bayes a posteriori foi definido sem menção de uma função perda. Porém, na definição do risco de Bayes é exigido a especificação de uma função perda, para o seu cálculo utilizando a função risco. Desse modo, o risco de Bayes de um estimador δ é dado por:

$$\begin{aligned} r(\delta) &= \int_{\Theta} \mathcal{R}(\delta, \theta) \pi(\theta) d\theta \\ &= \int_{\Theta} E_{\theta} [L(\delta(\mathbf{X}), \omega(\theta))] \pi(\theta) d\theta \\ &= \int_{\Theta} \left\{ \int_{\mathfrak{X}^n} L(\delta(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \prod_{i=1}^n dx_i \right\} \pi(\theta) d\theta \\ &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta) d\theta \right\} \prod_{i=1}^n dx_i \\ &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \frac{f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} f_{\mathbf{X}}(\mathbf{x}) d\theta \right\} \prod_{i=1}^n dx_i \\ &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\theta \right\} \prod_{i=1}^n dx_i \\ &= \int_{\mathfrak{X}^n} \left\{ \int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \right\} f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^n dx_i. \end{aligned}$$

Nesse caso, o estimador de Bayes é o estimador que minimiza essa expressão. A minimização é obtida minimizando-se a integral interna:

$$\int_{\Theta} L(\delta(\mathbf{x}), \omega(\theta)) \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Essa última integral é denominada algumas vezes risco a posteriori e, portanto, o estimador de Bayes é o estimador que minimiza o risco a posteriori.

C - CONCEITOS E RESULTADOS MATEMÁTICOS

O objetivo nessa parte do Apêndice é oferecer ao leitor alguns conceitos e resultados matemáticos que foram utilizados ao longo do trabalho.

Definição (Estatística suficiente): Uma estatística $T(\mathbf{X})$ é uma *estatística suficiente* para θ se, e somente se, a distribuição condicional da amostra \mathbf{X} dado o valor de $T(\mathbf{X})$, não depender de θ , para qualquer valor de $T(\mathbf{X})$.

Teorema da fatoração: Seja $f(\mathbf{x}; \theta)$ a fdp conjunta de uma amostra \mathbf{X} . Uma estatística $T(\mathbf{X})$ é a estatística suficiente para θ se, e somente se, existirem funções $g(t; \theta)$ e $h(\mathbf{x})$ de modo que, para todos os pontos amostrais \mathbf{x} e todos os pontos de parâmetro θ ,

$$f(\mathbf{x}; \theta) = g(T(\mathbf{x}); \theta) h(\mathbf{x})$$

Definição (Estatística suficiente mínima): Seja $f(\mathbf{x}; \theta)$ a fdp de uma amostra \mathbf{X} . Suponhamos que exista uma função $T(\mathbf{x})$ de modo que, para cada dois pontos amostrais \mathbf{x} e \mathbf{y} , a razão $\frac{f(\mathbf{x}; \theta)}{f(\mathbf{y}; \theta)}$ seja constante como uma função de θ se, e somente se, $T(\mathbf{x}) = T(\mathbf{y})$. Então, $T(\mathbf{X})$ é uma *estatística suficiente mínima* para θ .

Definição (Estatística ancilar): Uma estatística $T(\mathbf{X})$ cuja distribuição não depende do parâmetro θ é chamada *estatística ancilar*.

Definição (Estatística completa): Seja $f(t; \theta)$ uma família de fdp's para uma estatística $T(\mathbf{X})$. A família de distribuições de probabilidade é chamada completa se $E_\theta[g(T)] = 0$ para todo θ , o que implica que $P_\theta[g(T) = 0] = 1$ para todo θ . A estatística $T(\mathbf{X})$ é chamada de uma *estatística completa* se, e somente se, sua família de densidade é completa.

Teorema de Basu: Se $T(\mathbf{X})$ é uma estatística suficiente mínima e completa, então $T(\mathbf{X})$ é independente de toda estatística ancilar.

Teorema Lehmann-Schefté: Seja (X_1, \dots, X_n) uma amostra aleatória da densidade $f(x; \theta)$. Se $S = s(X_1, \dots, X_n)$ é uma estatística suficiente e completa, e, se $T = t(S)$, uma função de S , é um estimador não viesado de $\omega(\theta)$, então T é um UMVUE de $\omega(\theta)$.

Teorema de Gauss-Markov: Se $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ e $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$, os estimadores de mínimos quadrados $\hat{\beta}_j$, $j = 0, \dots, k$, tem variância mínima entre todos os estimadores lineares não viesados.

Desse teorema segue o seguinte corolário, que é uma extensão para as combinação linear dos $\hat{\beta}$'s.

Corolário: Se $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ e $\text{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$, o melhor estimador linear não viesado de $a'\boldsymbol{\beta}$ é $a'\hat{\boldsymbol{\beta}}$, em que $\hat{\boldsymbol{\beta}}$ é o estimador de mínimos quadrados $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.