CRISTIANE ALVARENGA GAJO

PROPRIEDADES E ASPECTOS GEOMÉTRICOS DE ESTIMADORES TIPO JAMES-STEIN E DO ESTIMADOR DE HARTIGAN

LAVRAS - MG
2016

CRISTIANE ALVARENGA GAJO

PROPRIEDADES E ASPECTOS GEOMÉTRICOS DE ESTIMADORES TIPO JAMES-STEIN E DO ESTIMADOR DE HARTIGAN

Tese 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 obtenção do título de Doutora.

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).

Gajo, Cristiane Alvarenga.
Propriedades e aspectos geométricos de estimadores tipo James-Stein e do estimador de Hartigan / Cristiane Alvarenga Gajo. – Lavras : UFLA, 2016.
156 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2016.
Orientador(a): Lucas Monteiro Chaves.
Bibliografia.

1. Estimador James-Stein. 2. Normal Multivariada. 3. Geometria. 4. Método Bayes empírico. I. Universidade Federal de Lavras. II. Título.

CRISTIANE ALVARENGA GAJO

PROPRIEDADES E ASPECTOS GEOMÉTRICOS DE ESTIMADORES TIPO JAMES-STEIN E DO ESTIMADOR DE HARTIGAN

Tese 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 obtenção do título de Doutora.

APROVADA em 26 de fevereiro de 2016.

Dra. Carla Regina Guimarães Brighenti	UFSJ
Dr. Devanil Jaques de Souza	UFLA
Dr. Daniel Furtado Ferreira	UFLA
Dr. Denismar Alves Nogueira	UNIFAL
Dra. Maria do Carmo Pacheco de Toledo Costa	UFLA

Dr. Lucas Monteiro Chaves
UFLA
(Orientador)

LAVRAS - MG

2016

Aos meus pais Eloisa e Geraldo (in memorian),

por me ensinarem a amar a vida.

Ao meu amado filho Davi,

por ser o amor em pessoa.

DEDICO.

AGRADECIMENTOS

A Deus, pela presença incessante em minha vida conduzindo meus gestos, palavras e pensamentos propiciando assim sabedoria, entendimento, fé e paz para viver bem cada dia.

Ao Instituto Federal Goiano campus Rio Verde, pela oportunidade de crescimento e aprendizado que me propiciaram.

Aos professores do Departamento de Ciências Exatas da Universidade Federal de Lavras por estarem sempre dispostos a ensinar com sabedoria e humildade.

Ao meu orientador professor Lucas Monteiro Chaves, por sua extrema dedicação, disposição, paciência e bom humor. Foram quatro anos de muito aprendizado e satisfação pelo trabalho realizado, meu sincero agradecimento.

Ao meu coorientador, professor Devanil Jaques de Souza, por sua disposição em colaborar. Sempre atencioso com minhas dificuldades e disposto a me orientar.

Aos meus colegas e amigos de doutorado, pela convivência harmoniosa e colaboração nos estudos. Em especial aos amigos Filipe Rizzo, Leandro, Danielle, Luzia e Carlos pois, partilhamos muito conhecimento na sala do Lucas.

Aos membros da banca do exame de qualificação Daniel Furtado, Carla Brighenti, Maria do Carmo Pacheco e Devanil Souza pelas críticas e sugestões que contribuem para esta tese e aos membros da defesa da tese.

À minha amada mãe, que com seu amor incondicional, cuidou de mim e do meu amado filho Davi. Não há amor mais disponível, cuidadoso e dedicado. Seu apoio e sua presença foram essenciais para tornar este trabalho possível.

Aos meus irmãos Adriano e Giovanni, pelo amor, carinho e incentivo juntamente com minhas cunhadas Fabíula e Letícia pelo grande apoio.

Ao meu namorado Adriano, pela paciência, companheirismo e amor.

RESUMO

O estimador de James-Stein é um estimador de encolhimento viesado que possui risco uniformemente menor que o risco do estimador média amostral para a média da distribuição normal multivariada, salvo nos casos unidimensional ou bidimensional. Interpretou-se com mais argumentos heurísticos e intensificou-se a abordagem geométrica da teoria do estimador de James-Stein. Além disso, propuseram-se novos estimadores de encolhimento tipo James-Stein e utilizou-se a métrica de Mahalanobis para abordar o estimador de James-Stein. Para avaliar o desempenho, em relação ao estimador média amostral, utilizou-se a simulação computacional pelo método Monte Carlo calculando-se o erro quadrático médio. O resultado indicou que o novo estimador apresenta melhor desempenho relativamente ao estimador média amostral.

Palavras-chave: Estimador James-Stein. Normal multivariada. Geometria. Método Bayes empírico.

ABSTRACT

The James-Stein estimator is a biased shrinkage estimator with uniformly smaller risk than the risk of the sample mean estimator for the mean of multivariate normal distribution, except in the one-dimensional or two-dimensional cases. In this work we have used more heuristic arguments and intensified the geometric treatment of the theory of James-Stein estimator. New type James-Stein shrinking estimators are proposed and the Mahalanobis metric used to address the James-Stein estimator. . To evaluate the performance of the estimator proposed, in relation to the sample mean estimator, we used the computer simulation by the Monte Carlo method by calculating the mean square error. The result indicates that the new estimator has better performance relative to the sample mean estimator.

Keywords: James-Stein estimator. Normal multivariate. Geometry. Empirical Bayes method.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

PRIMEIRA PARTE

1 INTRODUÇÃO

Estimadores de encolhimento tornaram-se ferramentas básicas na análise de dados de alta dimensão. Historicamente e conceitualmente, o início do seu desenvolvimento foi à mais de 50 anos quando STEIN (1956) publicou seu clássico artigo, "*Inadmissibility of the usual estimator for the mean of a multivariate normal distribution.*" Este artigo resultou na surpreendente descoberta estatística, em que foi apresentada a prova de que o estimador de máxima verossimilhança para a média de uma distribuição normal multivariada, considerado o melhor estimador não viesado, é inadmissível, salvo nos casos unidimensional e bidimensional.

Quando tratamos de estimadores em geral, buscamos estimadores que possuam a propriedade de serem não viesados para que o erro quadrático médio seja o menor possível. James e Stein (1961) exibiram um estimador com risco uniformemente menor que o do estimador de máxima verossimilhança. Este novo estimador é referido na literatura como o estimador de James-Stein, que é um estimador de encolhimento (*shrinkage estimator*). Este estimador perde a propriedade de ser não viesado, porém ganha no sentido de reduzir a variância e o erro quadrático médio.

Uma das razões do estimador de James-Stein ser tão revolucionário, é que muitos aspectos diferentes de algo que está sendo medido serão considerados para descrever uma única medição. Por exemplo, a incidência de uma doença em diferentes regiões pode ajudar a estimar a incidência de doença em uma pequena área. Surpreendentemente isto pode ser feito e produzir um estimador melhor. A comunidade estatística ficou surpresa com este resultado denominado paradoxo de Stein.

Frequentemente, em aplicações de estatística, quando um modelo de regressão linear é ajustado, algumas das variáveis regressoras são altamente correlacionadas, ou seja, a matriz de covariância do estimador de mínimos quadrados apresentará quasemulticolinearidade e o estimador de mínimos quadrados será muito impreciso. Uma possível solução para este problema, sugerida por Hoerl e Kennard (1970a), foi a formulação do estimador regressão de cumeeira (*ridge*). Este método é um caso particular para se obter estimadores de encolhimento. Esta denominação de estimador de encolhimento vem do fato de que ambos estimadores, James-Stein e regressão de cumeeira, aqui citados, serem obtidos pela multiplicação do estimador de mínimos quadrados por um fator de encolhimento.

O objetivo geral desta pesquisa é interpretar com mais argumentos heurísticos e geométricos a teoria do estimador de James-Stein em que as estimativas de quadrados mínimos são encolhidas em direção ao vetor nulo ou em direção a um vetor qualquer. Para seu desenvolvimento, temos dois objetivos específicos.

O primeiro refere-se à caracterização do estimador de James-Stein, ou seja, definir suas propriedades, ilustrar o estimador baseado na geometria plana, na esperança de esclarecer a ideia de Stein, demonstrar que

estimadores de encolhimento possuem variância mínima, demonstrar que estimador usual é inadmissível para dimensão suficientemente grande e que 3 é a dimensão crítica.

O segundo destina-se a propôr um estimador de James-Stein modificado e analisar o estimador de James-Stein aplicando a transformação de Mahalanobis. Dessa forma, realiza-se simulação computacional pelo método de Monte Carlo para comparar o erro quadrático médio destes estimadores com o estimador de máxima verossimilhança para média de uma distribuição normal multivariada.

Para o desenvolvimento do tema, o referencial teórico dedica-se a apresentar conceitos básicos sobre propriedades dos estimadores, discussão sobre admissibilidade, abordagem bayesiana empírica, estimadores de encolhimento e um caso particular que é o estimador de cumeeira (ridge) e, em seguida, apresenta-se o estimador de James-Stein, o lema de Stein, a justificativa heurística para o coeficiente de encolhimento e em seguida um estudo detalhado desenvolvendo, passo a passo, a interpretação geométrica do estimador de James-Stein e estimadores esfericamente simétricos e apresenta-se o estudo computacional do estimador proposto e a análise do estimador de James-Stein modificado, utilizando a transformação de Mahalanobis.

O desenvolvimento dos resultados apresentados, na literatura, foram expandidos em detalhes como contribuição didática.

Na sequência, tem-se a exposição de dois artigos.

O primeiro, referente a proposta de um novo estimador tipo James-Stein e suas propriedades e a abordagem do estimador de James-Stein utilizando a métrica de Mahalanobis, artigo submetido à revista Brasileira de Biometria.

O segundo artigo refere-se a estimativa do vetor de médias da distribuição normal multivariada limitada a um conjunto convexo fechado e limitado, utilizando o estimador de Hartigan, publicado em abril de 2016 na revista Brasileira de Biometria.

2 REFERENCIAL TEÓRICO

O objetivo central deste estudo é a interpretação geométrica do estimador de James-Stein para uma contribuição didática à teoria. Antes de iniciar as definições formais, enunciar os principais teoremas e lemas relativos a esta teoria, nesta seção, serão abordados os conceitos preliminares e essenciais sobre estimadores.

As citações serão usualmente omitidas, nesta seção, considerando que está amplamente baseada em Mood, Graybill e Boes (1974).

2.1 Propriedades de estimadores pontuais

Em um fenômeno aleatório qualquer, ao estudarmos a estrutura probabilística de quantidades associadas a esse fenômeno, introduzimos o conceito de variável aleatória. No caso unidimensional, em que se quer estudar apenas uma quantidade, é uma função definida do espaço amostral no conjunto dos números reais. No caso de interesse em várias quantidades, a estatística multivariada deve ser utilizada. Na estatística multivariada, observamos realizações de variáveis aleatórias em $\mathbb{R}^p$. A sequência de variáveis aleatórias $\mathbf{X_1}, ..., \mathbf{X_n}$ forma uma amostra aleatória de tamanho n, em que $\mathbf{X_j}$ é uma variável aleatória p-dimensional com função densidade $f(\mathbf{x})$. A realização para a j-ésima observação amostral p-dimensional da variável aleatória populacional $\mathbf{X}$ é $\mathbf{x_j}$ (FERREIRA, 2011).

No processo de encontrar estimadores de parâmetros, surge a dificuldade em escolher entre estimadores obtidos por diferentes métodos. Para a escolha, deve-se fazer um estudo do desempenho e da otimalidade de estimadores segundo alguns critérios.

Ao se estimar um parâmetro real $\boldsymbol{\theta}$ ou uma função do parâmetro $\tau(\boldsymbol{\theta})$ de uma densidade de probabilidade $f(\mathbf{x}; \boldsymbol{\theta})$, por um estimador $\mathbf{T} = \mathbf{t}(\mathbf{X_1}, ..., \mathbf{X_n}) = \hat{\tau}(\boldsymbol{\theta})$, é interessante quantificar uma penalização se o verdadeiro valor $\tau(\boldsymbol{\theta})$ é estimado por $\hat{\tau}(\boldsymbol{\theta})$. Um critério para avaliar esta quantificação é através de funções denominadas funções perda.

As definições abaixo são para o caso univariado.

Definição 2.1 *(Função Perda) Uma função perda l definida em $(\mathbf{t}, \boldsymbol{\theta})$, para o estimador $\hat{\tau}(\boldsymbol{\theta}) = \mathbf{t}\left(\mathbf{X_1}, ..., \mathbf{X_n}\right)$ é uma função d com valor real satisfazendo:*
(i) $l(\mathbf{t}, \boldsymbol{\theta}) \geq \mathbf{0}$ *para todo* $\mathbf{t} = \mathbf{t}(x_1, ..., x_n)$ *e todo* $\boldsymbol{\theta}$*;*
(ii) $l(\mathbf{t}, \boldsymbol{\theta}) = \mathbf{0}$ *para* $\mathbf{t} = \tau(\boldsymbol{\theta})$.

As propriedades são claras, nada se perde se a estimativa está correta e, no caso da estimativa estar incorreta, ocorre uma perda positiva. Como exemplos de funções perda tem-se,

- perda quadrática $l(\mathbf{t}, \boldsymbol{\theta}) = (\mathbf{t} - \tau(\boldsymbol{\theta}))^{\mathbf{2}}$,
- perda em valor absoluto $l(\mathbf{t}, \boldsymbol{\theta}) = |\mathbf{t} - \tau(\boldsymbol{\theta})|$,
- $l(\mathbf{t}, \boldsymbol{\theta}) = \begin{cases} A, & se \quad |\mathbf{t} - \tau(\boldsymbol{\theta})| > \varepsilon \\ 0, & se \quad |\mathbf{t} - \tau(\boldsymbol{\theta})| \leq \varepsilon, \ \text{ em que } \ A > 0 \end{cases}$.

Neste trabalho, será considerada somente a função perda quadrática.

Como as estimativas são obtidas pelo estimador $\mathbf{T} = \mathbf{t}(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}})$, temos que a perda é uma variável aleatória $l(\mathbf{T}, \boldsymbol{\theta}) = \mathbf{l}(\mathbf{t}(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}), \boldsymbol{\theta})$, que corresponde ao valor que se perde ao observar uma amostra aleatória e se fazer a estimativa. Como a perda é aleatória, um aspecto importante é sua esperança, denominada função risco do estimador.

Definição 2.2 *(Função Risco) Para um estimador* $\mathbf{T} = \mathbf{t}(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}})$ *e uma função perda* l *definida em* $(\mathbf{t}, \boldsymbol{\theta})$*, a função risco, denotada por* $\mathscr{R}_{\mathbf{t}}$ *definida em* $(\boldsymbol{\theta})$*, é*

$$\begin{aligned} \mathscr{R}_{\mathbf{t}}(\boldsymbol{\theta}) &= E[l(\mathbf{T}, \boldsymbol{\theta})] \\ &= E[l(\mathbf{t}(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}), \boldsymbol{\theta})] \\ &= \int ... \int l(\mathbf{t}(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}), \boldsymbol{\theta}) \mathbf{\Pi f}(\mathbf{x_i}; \boldsymbol{\theta}) \mathbf{dx_1} ... \mathbf{dx_n}. \end{aligned}$$

Observação: Será utilizada também a notação $\mathscr{R}_{\mathbf{t}}(\boldsymbol{\theta}) = \mathscr{R}(T, \boldsymbol{\theta})$.

Dados dois estimadores $\mathbf{T_1}$ e $\mathbf{T_2}$, as funções risco associadas a esses dois estimadores geralmente possuem gráficos que se interceptam, ou seja, em um determinado intervalo de valores do parâmetro a função risco do primeiro estimador pode ser maior que a função risco do segundo estimador, mas em um outro intervalo essa relação se inverte, conforme ilustrado na Figura 1.

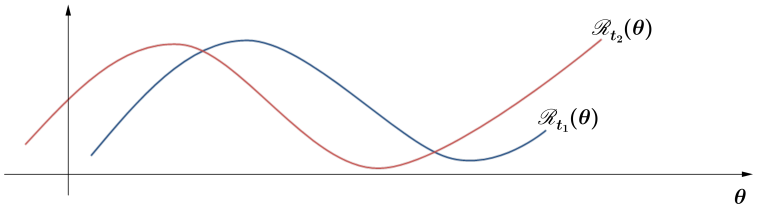

FIGURA 1: Função risco para dois estimadores

Se a opção para medir a penalização de um estimador $\mathbf{T} = \mathbf{t}(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}})$ de $\tau(\boldsymbol{\theta})$ é a função perda quadrática, a função risco será denominada erro quadrático médio (EQM).

Definição 2.3 *(Erro Quadrático Médio) Seja* $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$ *um estimador de* $\tau(\boldsymbol{\theta})$*. Considerando a função perda* $l(\mathbf{t}, \boldsymbol{\theta}) = (\mathbf{t} - \tau(\boldsymbol{\theta}))^{\mathbf{2}}$*, a função risco* $E[l(\mathbf{T}, \theta)] = \mathbf{E}[(\mathbf{T} - \tau(\boldsymbol{\theta}))^{\mathbf{2}}]$ *é denominada erro quadrático médio do estimador* $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$.

Observe que o erro quadrático médio do estimador $\mathbf{T}$ de $\tau(\boldsymbol{\theta})$ pode ser escrito como

$$\begin{aligned} EQM_{\mathbf{t}}(\boldsymbol{\theta}) &= E\left[(\mathbf{T} - \tau\left(\boldsymbol{\theta}\right))^2\right] \\ &= E\left[(\mathbf{T} - \mathbf{E}\left[\mathbf{T}\right] + \mathbf{E}\left[\mathbf{T}\right] - \tau\left(\boldsymbol{\theta}\right))^2\right] \\ &= E\left[(\mathbf{T} - \mathbf{E}\left[\mathbf{T}\right])^2\right] + \left(E\left[\mathbf{T}\right] - \tau\left(\boldsymbol{\theta}\right)\right)^2 \\ &= \mathrm{var}\left[\mathbf{T}\right] + \left(\mathrm{viés}(\mathbf{T})\right)^2. \end{aligned}$$

Portanto, o EQM incorpora dois componentes, um deles medindo a variabilidade do estimador (precisão) e o outro medindo o seu viés (exatidão).

Para o caso em que se tem um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, a função perda quadrática é definida de forma análoga. Se $\mathbf{T} = (\mathbf{t_1}(\mathbf{X_1}, ..., \mathbf{X_n}), ..., \mathbf{t_p}(\mathbf{X_1}, ..., \mathbf{X_n}))$ for o estimador de $\tau(\boldsymbol{\theta})$ então, a função perda depende da norma $\|\mathbf{T} - \tau(\boldsymbol{\theta})\|$, em particular $EQM_{\mathbf{T}}(\tau(\boldsymbol{\theta})) = E\left[\|\mathbf{T} - \tau\left(\boldsymbol{\theta}\right)\|^2\right]$.

Com o objetivo de selecionar bons estimadores, pelo critério de possuírem menor risco, define-se o conceito de estimador admissível.

Definição 2.4 *Para dois estimadores* $\mathbf{T_1} = \mathbf{t_1}(X_1, ..., X_n)$ *e* $\mathbf{T_2} = \mathbf{t_2}(X_1, ..., X_n)$*, o estimador* $\mathbf{T_1}$ *é dito melhor estimador que* $\mathbf{T_2}$*, em relação a uma função perda* l *definida em* $(\mathbf{t}, \boldsymbol{\theta})$*, se* $\mathscr{R}_{\mathbf{t_1}}(\boldsymbol{\theta}) \leq \mathscr{R}_{\mathbf{t_2}}(\boldsymbol{\theta})$ *para todo* $\boldsymbol{\theta}$ *no espaço paramétrico* Θ *e* $\mathscr{R}_{\mathbf{t_1}}(\boldsymbol{\theta}) < \mathscr{R}_{\mathbf{t_2}}(\boldsymbol{\theta})$ *para pelo menos um* $\boldsymbol{\theta}$ *em* Θ*.*

Definição 2.5 *(Estimador Admissível) Um estimador* $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$ *é dito admissível para uma função perda, se não há estimador melhor em relação a esta função perda.*

Uma maneira de remover a dependência da função risco do parâmetro $\boldsymbol{\theta}$ é substituir a função risco pelo seu valor máximo e comparar estimadores, olhando apenas para seus respectivos riscos máximos. Naturalmente, prefere-se o estimador com menor risco máximo. Segue a definição para tal estimador:

Definição 2.6 *(Minimax) Um estimador* $\mathbf{T}^*$ *é definido um estimador minimax se*

$$\sup_{\theta} \mathscr{R}_{\mathbf{t}^*}(\boldsymbol{\theta}) \leq \sup_{\theta} \mathscr{R}_{\mathbf{t}}(\boldsymbol{\theta}), \qquad \textit{para todo estimador } \mathbf{T}.$$

Os estimadores minimax são provenientes de uma abordagem conservadora, isto é, deseja-se controlar o maior valor do risco.

2.2 Discussão sobre admissibilidade

O conceito de admissibilidade parece ser um conceito restritivo, pois é uma propriedade desejável em um estimador, porém não garante que este estimador seja adequado.

O exemplo mais simples é o de considerar para uma população qualquer $f(\mathbf{x};\boldsymbol{\theta})$ o estimador de $\boldsymbol{\theta}$ que desprezando os dados estima $\boldsymbol{\theta}$ por uma constante, $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right) = \theta_{\mathbf{0}}$.

O risco quadrático deste estimador é

$$\mathscr{R}\left(\mathbf{T}, \boldsymbol{\theta}\right) = E_{\theta}\left[\left(\mathbf{T} - \boldsymbol{\theta}\right)^2\right] = E_{\theta}\left[\left(\theta_0 - \boldsymbol{\theta}\right)^2\right] = \left(\theta_0 - \boldsymbol{\theta}\right)^2.$$

Suponha $\mathbf{T}' = \mathbf{t}'\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$ outro estimador com:

$$\mathscr{R}\left(\mathbf{T}', \boldsymbol{\theta}\right) \leq \mathscr{R}\left(\mathbf{T}, \boldsymbol{\theta}\right) = \left(\theta_0 - \boldsymbol{\theta}\right)^2$$

logo,

$$\mathscr{R}\left(\mathbf{T}', \boldsymbol{\theta}\right) = E_{\theta=\theta_0}\left[\left(\mathbf{T}' - \boldsymbol{\theta}\right)^2\right] \leq \mathscr{R}\left(\mathbf{T}, \boldsymbol{\theta}\right) = 0.$$

Assim,

$$\begin{aligned} 0 &= E_{\theta=\theta_0}\left[\left(\mathbf{T}' - \theta_{\mathbf{0}}\right)^2\right] \\ &= E_{\theta=\theta_0}\left[\left(\mathbf{T}' - \mathbf{E}\left[\mathbf{T}'\right] + \mathbf{E}\left[\mathbf{T}'\right] - \theta_{\mathbf{0}}\right)^2\right] \\ &= E_{\theta=\theta_0}\left[\left(\mathbf{T}' - \mathbf{E}\left[\mathbf{T}\right]\right)^2\right] + \left(E\left[\mathbf{T}'\right] - \theta_0\right)^2 \end{aligned}$$

o que implica que:

$$E\left[\mathbf{T}'\right] = \theta_0 \qquad \text{e} \qquad \mathrm{var}\left[\mathbf{T}'\right] = 0.$$

Então concluímos que:

$$\mathbf{T}' = \theta_{\mathbf{0}}.$$

Assim, tem-se um estimador simples e admissível porém inadequado.

A escolha do método para a obtenção de estimadores admissíveis é uma questão pertinente. No caso de perda quadrática se um estimador não viesado é admissível, então possui variância mínima entre todos os estimadores não viesados. Em razão da importância deste conceito, define-se:

Definição 2.7 *(Estimador não viesado com variância uniformemente mínima-UMVUE)*
Se $X_1, ..., X_n$ é uma amostra aleatória de $f(\mathbf{x};\boldsymbol{\theta})$, um estimador $\mathbf{T}^* = \mathbf{t}^*(X_1, ..., X_n)$ de $\tau(\boldsymbol{\theta})$ é definido

ser um estimador não viesado de variância uniformemente mínima de $\tau(\boldsymbol{\theta})$ *se*
(i) $E\left[\mathbf{T}^*\right] = \tau(\boldsymbol{\theta})$*;*
(ii) $\mathrm{var}\left[\mathbf{T}^*\right] \leq \mathrm{var}\left[\mathbf{T}\right]$*, para qualquer outro estimador não viesado* $\mathbf{T} = \mathbf{t}\left(\mathrm{X_1}, ..., \mathrm{X_n}\right)$ *de* $\tau(\boldsymbol{\theta})$.

Seja $X_1, ..., X_n$ uma amostra aleatória de $f(\mathbf{x}; \boldsymbol{\theta})$ em que $\boldsymbol{\theta}$ pertence a Θ, subconjunto dos reais. Considere as condições de regularidade razoáveis se $\mathbf{T} = \mathbf{t}(\mathbf{x_1}, ...\mathbf{x_n})$ é um estimador qualquer de $\tau(\boldsymbol{\theta})$.

Teorema 2.1 *(Desigualdade de Cramér-Rao) Sob as suposições de regularidade e seja* $\tau'(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta}\tau\left(\theta\right)$

$$\mathrm{var}\left[\mathbf{T}\right] \geq \frac{\left[\tau'(\boldsymbol{\theta})\right]^2}{nE\left[\left[\frac{\partial}{\partial\theta}\log f(\mathbf{X}; \boldsymbol{\theta})\right]^2\right]}. \tag{2.1}$$

A igualdade ocorre se, e somente se, existe uma função, $K(\boldsymbol{\theta}, n)$*, tal que*

$$\sum_i \frac{\partial}{\partial\theta}\log f(x_i; \boldsymbol{\theta}) = K(\boldsymbol{\theta}, n)\left[t(x_1, ..., x_n) - \tau(\boldsymbol{\theta})\right].$$

O lado direito da equação (2.1) é chamado de limite inferior de Cramér-Rao para variância de estimadores não viesados de $\tau\boldsymbol{\theta}$.

O teorema 2.1 fornece um limite inferior para a variância dos estimadores não viesados de $\tau\boldsymbol{\theta}$. Se um estimador não viesado tem variância que coincide com o limite inferior de Cramér-Rao, então este estimador é um UMVUE pois satisfaz a condição *(ii)* da definição 2.7.

Para introduzir o conceito de estimador de quadrados mínimos para um vetor de parâmetros, visando a busca do melhor estimador, considere o modelo linear usual $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ com $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{X}$ de posto coluna completo e matriz de covariâncias $\mathrm{cov}[\mathbf{Y}] = \mathbf{E}\left[(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])(\mathbf{Y} - \mathbf{E}[\mathbf{Y}])'\right] = \sigma^{\mathbf{2}}\mathbf{I}$. O estimador de quadrados mínimos para o vetor $\boldsymbol{\beta}$ é dado por $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'Y}$ e $cov[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X'X})^{-\mathbf{1}}$. O teorema de Gauss-Markov garante que $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear não viesado (BLUE).

Teorema 2.2 *(Gauss-Markov) O estimador de quadrados mínimos possui a menor variância total, traço*$\left(cov[\hat{\boldsymbol{\beta}}]\right)$*, entre todos os estimadores lineares não viesados.*

A importância desse teorema se dá principalmente no fato de que nenhuma hipótese distribucional é exigida. O Teorema de Gauss-Markov pode ser generalizado da seguinte forma:

Teorema 2.3 *(Gauss-Markov) Se* $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$*,* $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ *e* $\mathrm{cov}[\mathbf{Y}] = \mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^{\mathbf{2}}\mathbf{V}$*, em que* $\mathbf{V}$ *é uma matriz positiva definida conhecida, então,*

(i) O melhor estimador linear não viesado de $\boldsymbol{\beta}$ é:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X'V^{-1}X}\right)^{-1}\mathbf{X'V^{-1}Y},$$

(ii) A matriz de covariâncias para $\hat{\boldsymbol{\beta}}$ é

$$\mathrm{cov}[\hat{\boldsymbol{\beta}}] = \sigma^2\left(\mathbf{X'V^{-1}X}\right)^{-1},$$

(iii) Um estimador não viesado de σ^2 é

$$\begin{aligned} s^2 &= \frac{\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\mathbf{V^{-1}}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}{n-k-1} \\ &= \frac{\mathbf{y}'\left[\mathbf{V^{-1}}-\mathbf{V^{-1}X}\left(\mathbf{X'V^{-1}X}\right)^{-1}\mathbf{X'V^{-1}}\right]\mathbf{y}}{n-k-1}, \end{aligned}$$

em que $k+1$ é o número de parâmetros ou posto coluna de $\mathbf{X}$, n é o tamanho da amostra e $\mathbf{y}$ é uma realização de $\mathbf{Y}$ de dimensão $n \times 1$.

2.2.1 Contexto Bayesiano

Pode-se utilizar uma abordagem Bayesiana para o problema da otimalidade da função perda utilizando uma distribuição *a priori* para calcular um risco médio. No contexto Bayesiano, a função risco $\mathscr{R}$ pode ser generalizada. Considere que o parâmetro $\boldsymbol{\theta}$ é uma variável aleatória com densidade *a priori* $g(\boldsymbol{\theta})$ definida no espaço paramétrico Θ. Como a função risco é função do parâmetro $\boldsymbol{\theta}$, tal fato a caracteriza como uma variável aleatória e a ideia é tomar a esperança do risco em relação à densidade *a priori*.

Definição 2.8 *(Risco de Bayes) Se $\mathscr{R}_t(\boldsymbol{\theta}) = E\left[l(\mathbf{T},\boldsymbol{\theta})\right]$, a função risco de Bayes do estimador $\mathbf{T}$ em relação a densidade a priori $g(\boldsymbol{\theta})$ é*

$$r_{l,g}(\mathbf{t}) = \int_{\Theta} \mathscr{R}_t(\theta)g(\boldsymbol{\theta})d\theta.$$

Portanto, o risco de Bayes é um número e nos permite sempre comparar dois estimadores, o que gera uma nova definição.

Definição 2.9 *(Estimador de Bayes) Um estimador $\mathbf{T}^* = \mathbf{t}^*(X_1,...,X_n)$ é dito o estimador de Bayes em*

relação a função perda l definida em $(\mathbf{t}, \boldsymbol{\theta})$ e priori $g(\boldsymbol{\theta})$ se

$$r_{l,g}(\mathbf{t}^*) \leq r_{l,g}(\mathbf{t})$$

para todo estimador $\mathbf{T} = \mathbf{t}\,(\mathrm{X}_\mathbf{1}, ..., \mathrm{X}_\mathbf{n})$.

Dada uma densidade *a priori* $g(\boldsymbol{\theta})$, após se observar uma amostra $X_1 = x_1, ..., X_n = x_n$ tem-se, pelo Teorema de Bayes, a densidade *a posteriori* $g(\boldsymbol{\theta}|x_1, ..., x_n)$. O estimador de Bayes em relação à perda quadrática é definido como

Teorema 2.4 *O estimador de Bayes em relação à perda quadrática é dado como a esperança da distribuição a posteriori, isto é,*

$$\mathbf{T}^* = \mathbf{t}^*(x_1, ..., x_n) = \int_\Theta \boldsymbol{\theta} g(\boldsymbol{\theta}|x_1, ..., x_n) d\boldsymbol{\theta}.$$

Tem-se uma interessante relação entre estimadores minimax e estimadores de Bayes.

Teorema 2.5 *Se* $\mathbf{T}^* = \mathbf{t}^*(X_1, ..., X_n)$ *é um estimador de Bayes tal que sua função risco é constante então* $\mathbf{T}^*$ *é um estimador minimax.*

Demonstração:
Sejam $g(\boldsymbol{\theta})$ a densidade *a priori* e o estimador de Bayes $\mathbf{T}^* = \mathbf{t}^*(X_1, ..., X_n)$. Se $\mathscr{R}_{t*} = \text{constante}$ então

$$\begin{aligned}
\sup_{\theta\in\Theta} \mathscr{R}_{t^*}(\boldsymbol{\theta}) &= \sup_{\theta\in\Theta} \text{constante} \\
&= \text{constante} \\
&= \mathscr{R}_{t^*}(\boldsymbol{\theta}) \\
&= \int_\Theta \mathscr{R}_{t^*}(\boldsymbol{\theta}) g(\boldsymbol{\theta})\, d\boldsymbol{\theta} = r_{l,g}(\mathbf{t}^*) \leq r_{l,g}(\mathbf{t}) \\
&= \int_\Theta \mathscr{R}_t(\boldsymbol{\theta}) g(\boldsymbol{\theta})\, d\boldsymbol{\theta} \leq \sup_{\theta\in\Theta} \mathscr{R}_t(\boldsymbol{\theta})
\end{aligned}$$

para um estimador qualquer $\mathbf{T} = \mathbf{t}\,(\mathrm{X}_\mathbf{1}, ..., \mathrm{X}_\mathbf{n})$. ∎

2.2.2 Método Bayesiano Empírico

A precisão das inferências estatísticas que são realizadas utilizando os métodos Bayesianos depende fortemente de um bom conhecimento *a priori*. Quando esta informação, representada pela distribuição *a*

priori, é incompleta ou desconhecida pode ser obtida a partir dos dados. Esse método é chamado de Método Bayesiano Empírico. Vamos contextualizar o método por exemplos.

Exemplo 2.1 *Um exemplo clássico segundo Gruber (1998), considere X uma variável aleatória discreta com distribuição Poisson de parâmetro λ e uma distribuição a priori $g(\lambda)$ própria e totalmente desconhecida. Para estimar o parâmetro λ tomamos a densidade de probabilidade de X, dada por*

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

Aplicando o teorema de Bayes temos que a distribuição a posteriori $g(\lambda|x)$ é igual a

$$g(\lambda|x) = \frac{\frac{e^{-\lambda}\lambda^x}{x!}g(\lambda)}{h(x)}$$

em que

$$h(x) = \int \frac{e^{-\lambda}\lambda^x}{x!}g(\lambda)d\lambda.$$

Observe que $h(x)$ é a distribuição marginal de X, ou seja, só depende dos dados e como $g(\lambda)$ é desconhecido, $h(x)$ também é desconhecida. O estimador de Bayes $\hat{\lambda}_B$ é a média da distribuição a posteriori, logo

$$\begin{aligned}
\hat{\lambda}_B = E[\lambda|x] &= \int \lambda \frac{e^{-\lambda}\lambda^x}{x!h(x)}g(\lambda)d\lambda \\
&= \frac{(x+1)!h(x+1)}{x!h(x)} \underbrace{\int \frac{e^{-\lambda}\lambda^{x+1}}{(x+1)!h(x+1)}g(\lambda)d\lambda}_{1} \\
&= \frac{(x+1)!h\,(x+1)}{x!h\,(x)} \\
&= \frac{(x+1)\,h\,(x+1)}{h\,(x)}.
\end{aligned}$$

A função $h(x)$ pode ser estimada utilizando os dados. Como h tem distribuição discreta e se $x_1, \ldots, x_n$ são observações ocorridas no passado, $h(x)$ e $h(x+1)$ podem ser estimados pela frequência relativa.

$$\hat{h}(x) = \frac{card\,(\{i, x_i = x\})}{n}$$

$$\hat{h}(x+1) = \frac{card\,(\{i, x_i = x+1\})}{n}$$

em que $card(A)$ =número de elementos do conjunto A. Desta forma o estimador de Bayes empírico é

$$\hat{\lambda}_B = \frac{(x+1)\hat{h}(x+1)}{\hat{h}(x)}.$$

Repare que $\hat{\lambda}_B$ não depende da distribuição a priori. Isso só ocorre devido à particularidade da distribuição dos dados ser uma distribuição discreta Poisson, isto é, apesar da distribuição a priori ser totalmente desconhecida conseguimos estimar a média da Poisson pelo método Bayesiano. Esse exemplo ilustra bem em que consiste o método Bayesiano empírico. Nele, os dados são utilizados duas vezes, na função de verossimilhança, neste caso $f(x;\lambda)$ e na estimativa da marginal de x, $h(x)$.

Explicitando numericamente, suponha que o número de acidentes por semana X segue uma distribuição Poisson. Em uma semana, ocorreram 4 acidentes. Suponha que o número de acidentes observados nas últimas 10 semanas foram

5 8 7 4 4 1 4 4 2 5 .

Os estimadores de $h(4)$ e $h(4+1)$ são, respectivamente

$$\hat{h}(4) = \frac{4}{10}$$

$$\hat{h}(4+1) = \frac{2}{10}.$$

Então, para este caso, o Estimador Bayesiano Empírico é

$$\hat{\lambda}_B = \frac{(4+1)\hat{h}(4+1)}{\hat{h}(4)} = \frac{5\frac{2}{10}}{\frac{4}{10}} = \frac{5}{2}.$$

Para o próximo exemplo será utilizado o seguinte teorema

Teorema 2.6 *Se $Z_1, ..., Z_n$ são variáveis aleatórias de uma distribuição normal padrão independentes, então:*
(i) $\bar{Z}$ tem um distribuição normal com média 0 e variância $\frac{1}{n}$,
(ii) $\bar{Z}$ e $\sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2$ são independentes,
(iii) $\sum_{i=1}^{n} \left(Z_i - \bar{Z}\right)^2$ tem distribuição qui-quadrado com $n-1$ graus de liberdade.

Exemplo 2.2 *Segundo Casella (1985), considere uma variável aleatória normal* $\mathbf{X}= (X_1, ..., X_n)$*, tal que*

$$X_i \sim N_n\left(\theta_i, \sigma^2\right) \qquad i = 1, ..., n \qquad \text{com } \sigma^2 \text{ conhecido.}$$

O estimador usual de θ_i é X_i, a observação. Este estimador tem muitas propriedades de otimalidade (BLUE, EMV, Minimax,etc.) mas, podemos melhorá-lo. Pelo método Bayesiano, supomos uma priori

$$\theta_i \sim N\left(\mu, \tau^2\right) \qquad i = 1, ..., n \qquad \textit{com } \tau^2 \textit{ desconhecido.} \tag{2.2}$$

Para uma observação, isto é, uma amostra de tamanho 1, $\boldsymbol{x} = (x_1, ..., x_n)$, a distribuição a posteriori de $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$ é dada por [Apêndice A]

$$g\left(\theta_i | X_i\right) \sim N\left(\frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}X_i, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right).$$

O estimador de Bayes para θ_i é, portanto

$$\begin{aligned}\left(\hat{\theta}_i\right)_B &= \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)\mu + \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)X_i \\ &= \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)\mu + \left(1 - \frac{\sigma^2}{\sigma^2+\tau^2}\right)X_i.\end{aligned} \tag{2.3}$$

Note que $\left(\hat{\theta}_i\right)_B$ é uma média ponderada de μ e X_i. O peso usado depende dos parâmetros da priori que pelo método Bayes empírico podem ser estimados a partir dos dados. Toda a informação dos parâmetros da priori μ e τ^2 está contida na distribuição marginal de X, e outro cálculo padrão [Apêndice A] mostra que a distribuição marginal (preditiva) $f\left(X\right) = (f(X_1), ..., f(X_n))$ é dada por

$$f(X_i) \sim N_n\left(\mu, \sigma^2 + \tau^2\right).$$

Considerando que a amostra $\boldsymbol{x} = (x_1, ..., x_n)$ é também uma amostra da preditiva $f(X)$, podemos construir estimadores de Bayes para os pesos na equação (2.3)*.*

Como $E[f(X_i)] = \mu$ pelo método dos momentos μ pode ser estimado por $\bar{x}$.

Considerando o peso associado a $\boldsymbol{x}$, para estimá-lo, utilizaremos o teorema 2.6.

Logo,

$$\begin{aligned} Z_i &= \frac{X_i - \mu}{\sqrt{\sigma^2+\tau^2}} \\ \frac{\sum_{i=1}^{n} Z_i}{n} &= \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{\sigma^2+\tau^2}} \\ \bar{Z} &= \frac{\bar{X} - \mu}{\sqrt{\sigma^2+\tau^2}} \sim N\left(0, \frac{1}{n}\right)\end{aligned}$$

e consequentemente,

$$\sum_{i=1}^{n}\left(Z_i-\bar{Z}\right)^2=\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sqrt{\sigma^2+\tau^2}}-\frac{\bar{X}-\mu}{\sqrt{\sigma^2+\tau^2}}\right]^2$$
$$=\frac{\sum_{i=1}^{n}\left[X_i-\bar{X}\right]^2}{\sigma^2+\tau^2}\sim\chi^2_{(n-1)}.$$

Reescrevendo a equação acima seja $U=\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ *e* $W=\sum_{i=1}^{n}\left(Z_i-\bar{Z}\right)^2$

$$\frac{1}{U}=\frac{1}{\sigma^2+\tau^2}\frac{1}{W}$$
$$E\left[\frac{1}{U}\right]=\frac{1}{\sigma^2+\tau^2}E\left[\frac{1}{W}\right]. \tag{2.4}$$

Temos que,

$$\frac{1}{W}\sim Gama\left(\boldsymbol{\alpha}=\frac{n-1}{2},\beta=\frac{1}{2}\right),$$

pois, $W\sim\chi^2_{n-1}$ *e a inversa de uma* χ^2_{n-1} *é uma gama. Então,*

$$E\left[\frac{1}{W}\right]=\int_0^\infty\frac{1}{w}\frac{\left(\frac{1}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}w^{\left(\frac{n-1}{2}-1\right)}e^{-\frac{1}{2}w}dw$$
$$=\frac{1}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{2}\right)^{\frac{n-1}{2}}}\int_0^\infty w^{\left(\frac{n-1}{2}-1\right)-1}e^{-\frac{1}{2}w}dw$$
$$=\frac{\Gamma\left(\frac{n-1}{2}-1\right)}{\Gamma\left(\frac{n-1}{2}\right)\left(\frac{1}{2}\right)^{-1}}\int_0^\infty\frac{\left(\frac{1}{2}\right)^{\frac{n-1}{2}-1}}{\Gamma\left(\frac{n-1}{2}-1\right)}w^{\left(\frac{n-1}{2}-1\right)-1}e^{-\frac{1}{2}w}dw$$
$$=\frac{\Gamma\left(\frac{n-1}{2}-1\right)}{2\left(\frac{n-1}{2}-1\right)\Gamma\left(\frac{n-1}{2}-1\right)}$$
$$=\frac{1}{n-3}.$$

Retornando para a equação (2.4)

$$E\left[\frac{1}{U}\right]=\frac{1}{\sigma^2+\tau^2}\frac{1}{n-3}$$
$$E\left[\frac{(n-3)\sigma^2}{\sum_{i=1}^{n}\left[X_i-\bar{X}\right]^2}\right]=\frac{\sigma^2}{\sigma^2+\tau^2}.$$

Substituindo essas estimativas dos parâmetros da priori na equação (2.3)*, obtemos a seguinte expressão*

para o estimador Bayes empírico de θ_i

$$\left(\hat{\theta}_i\right)_E = \left[\frac{(n-3)\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]\bar{X} + \left[1 - \frac{(n-3)\sigma^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\right]X_i.$$

Esse estimador de θ_i *usa informação de todos os* $X_i's$ *quando estima cada* θ_i*, este fato veio a ser conhecido como o paradoxo de Stein. O paradoxo de Stein afirma que estimativas podem ser melhoradas, usando a informação de todas as coordenadas quando estimamos cada coordenada.*

Note que assumimos uma priori como em (2.2) *com média* μ*. Esta escolha resulta em um encolhimento do* $\hat{\boldsymbol{\theta}}_B$ *na direção de* θ*. O estimador de Bayes empírico* $\left(\hat{\theta}_i\right)_E$ *é um bom estimador de* θ_i *pois tem uma propriedade teórica extremamente atraente: em média, ele está sempre mais próximo de* θ_i *do que de* X_i*. Podemos medir o quanto vale a pena um estimador de* θ_i *considerando* $\sum\left(\theta_i - \hat{\theta}_i\right)^2$*. Se* $n \geq 4$ *é verdade que*

$$E\left\{\sum_{i=1}^{n}\left[\theta_i - \left(\hat{\theta}_i\right)_E\right]^2\right\} < E\left\{\sum_{i=1}^{n}\left[\theta_i - X_i\right]^2\right\}, \qquad \textit{para todo } \theta_i. \tag{2.5}$$

Neste sentido $\left(\hat{\theta}_i\right)$*, está sempre mais próximo de* θ_i *do que de* X_i*. Para uma prova mais rigorosa de* (2.5)*, veja Efron e Morris (1973).*

2.2.3 Priori Imprópria e Estimador de Bayes Generalizado

Estimadores de Bayes são definidos a partir de distribuições *a priori* próprias $g\left(\boldsymbol{\theta}\right)$ (SMALL, 2014). Uma função peso $g\left(\boldsymbol{\theta}\right)$ que não define uma distribuição de probabilidade, isto é,

$$\int g\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} = \infty$$

pode ser considerada como *priori* e obter uma distribuição *a posteriori* que define uma distribuição de probabilidade. Neste caso, $g(\boldsymbol{\theta})$, é denominada uma *priori* imprópria e o estimador média da posteriori é denominado estimador de Bayes generalizado em relação a função peso $g(\boldsymbol{\theta})$. Segue um exemplo que utiliza priori imprópria

Exemplo 2.3 *Seja* $X \sim Binomial\,(p, n)$*,* $f(k;p) = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$*. Considere a priori imprópria*

$$g(p) = p^{-1}(1-p)^{-1}, \qquad 0 < p < 1$$

$$\int_0^1 p^{-1}(1-p)^{-1}dp = \infty.$$

$$\begin{aligned}
g(p|X=k) &= \frac{\left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}p^{-1}(1-p)^{-1}}{\int_0^1 \left(\begin{array}{c} n \\ k \end{array}\right) p^k(1-p)^{n-k}p^{-1}(1-p)^{-1}dp} \\
&= \frac{\left(\begin{array}{c} n \\ k \end{array}\right) p^{k-1}(1-p)^{n-k-1}}{\int_0^1 \left(\begin{array}{c} n \\ k \end{array}\right) p^{k-1}(1-p)^{n-k-1}dp} \\
&= \frac{\frac{\Gamma(k+n-k)}{\Gamma(k)\Gamma(n-k)}p^{k-1}(1-p)^{(n-k)-1}}{\frac{\Gamma(k+n-k)}{\Gamma(k)\Gamma(n-k)}\int_0^1 p^{k-1}(1-p)^{n-k-1}dp} \\
&= \frac{Beta(k,n-k)}{\frac{\Gamma(k+n-k)}{\Gamma(k)\Gamma(n-k)}Beta(k,n-k)} \\
&= \frac{\Gamma(k)\,\Gamma(n-k)}{\Gamma(k+n-k)}
\end{aligned}$$

para $X = k$ a posteriori é uma distribuição $Beta(k, n-k)$. O estimador de Bayes generalizado é dado pela média da posteriori

$$t = \frac{k}{k+n-k}$$

substituindo o valor realizado k pela variável X temos

$$T = \frac{X}{n}.$$

Para $X = k = 0$ e $X = k = n$ a posteriori $g(p|k)$ não é própria. Neste caso, temos que $T = t(X) = \frac{X}{n}$ minimiza a perda média esperada a posteriori.

O risco de Bayes é

$$\begin{aligned} r_{l,g}(t) &= \int_0^1 \left[\sum_{k=0}^{n} l(p,t(k))f(k,p)\right] g(p)dp \\ &= \int_0^1 \left[\sum_{k=0}^{n} l(p,t(k))f(k,p)g(p)\right] dp \\ &= \int_0^1 \left[\sum_{k=0}^{n} l(p,t(k))\frac{f(k,p)g(p)}{\int_0^1 f(k,p)g(p)dp}\int_0^1 f(k,p)g(p)dp\right] dp \\ &= \int_0^1 \left[\sum_{k=0}^{n} l(p,t(k))g(p|k)\int_0^1 f(k,p)g(p)dp\right] dp \\ &= \sum_{k=0}^{n}\left[\int_0^1 l(p,t(k))g(p|k)dp\right]\int_0^1 f(k,p)g(p)dp. \end{aligned}$$

O estimador de Bayes minimiza a perda quadrática a posteriori.

Para os casos $X = k = 0$ e $X = k = n$ a integral $\int_0^1 f(k,p)g(p)dp$ não está definida. No entanto, vamos mostrar que $T(0)$ e $T(n)$ também minimizam a integral $\int_0^1 \left[\sum_{k=0}^{n} l(p,t(k))f(k,p)\right] g(p)dp$.

Para $X = k = 0$, $T(0) = 0$ tem-se

$$\int_0^1 (p - T(0))^2 p^{-1}(1-p)^{n-1}dp = \int_0^1 \left(p^2\right)p^{-1}(1-p)^{n-1}dp = \int_0^1 p(1-p)^{n-1}dp = \mathbf{B}B(n,2)$$

que é uma integral finita.

Para o caso $X = k = n$, $T(n) = 1$

$$\int_0^1 (p - T(n))^2 p^{n-1}(1-p)^{-1}dp = \int_0^1 (p-1)^2 p^{n-1}(1-p)^{-1}dp = \int_0^1 p^{n-1}(1-p)dp = \mathbf{B}(n,2)$$

que é uma integral finita. Logo, o estimador de Bayes generalizado é $T = t(X) = \frac{X}{n}$, $0 \leq X \leq n$.

Outro conceito útil são os limites de estimadores de Bayes.

Definição 2.10 *Um estimador $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$ é um limite de estimadores de Bayes se existe uma sequência de prioris g_n tal que os estimadores de Bayes relativos a estas priores $\mathbf{T}^*_{\mathbf{g_n}} = \mathbf{t}^*_{\mathbf{g_n}}\left(\mathbf{X}_{\mathbf{1}}, ..., \mathbf{X}_{\mathbf{n}}\right)$ convergem pontualmente para $\mathbf{T} = \mathbf{t}\left(\mathrm{X}_{\mathbf{1}}, ..., \mathrm{X}_{\mathbf{n}}\right)$.*

Para mostrar que o estimador de Bayes generalizado $T = t(X) = \frac{X}{n}$ no exemplo anterior, para o parâmetro p da $Binomial(p, n)$, também é limite de estimadores de Bayes considere uma *priori* com distribuição beta com parâmetros r e s que é própria se $r > 0$, $s > 0$, o estimador de Bayes é dado por

$$\frac{X + r}{n + r + s}.$$

Considere a sequência de *prioris* com distribuição beta e parâmetros $(r, s) = (1, 1)$, $\left(\frac{1}{2}, \frac{1}{2}\right)$, $\left(\frac{1}{3}, \frac{1}{3}\right)$, ...; elas definem a sequência de estimadores de Bayes, logo,

$$\lim_{\substack{r \to 0 \\ s \to 0}} \frac{X + r}{n + r + s} = \frac{X}{n}.$$

Geralmente, estimadores de Bayes generalizados são também limites de estimadores de Bayes. Estes últimos são mais adequados pelo fato de estarem próximos a estimadores de Bayes o que pode não ocorrer com os estimadores de Bayes generalizados.

2.2.4 Admissibilidade dos estimadores de Bayes

Em geral, estimadores de Bayes são admissíveis.

Teorema 2.7 *Suponha que Θ é um intervalo de números reais e $\mathbf{T}^*$ um estimador de Bayes com respeito a priori com função densidade $g(\theta)$ com $g(\boldsymbol{\theta}) > 0$ para todo $\boldsymbol{\theta} \in \Theta$, e $\mathscr{R}$ contínua em $\boldsymbol{\theta}$ para todo estimador $\mathbf{T} = \mathbf{t}(\mathbf{X_1}, ..., \mathbf{X_n})$. Então $\mathbf{T}^*$ é admissível.*

Demonstração:
A prova é por contradição. Suponha que $\mathbf{T}^*$ é inadmissível, logo existe um outro estimador $\mathbf{T}$, tal que

$$\mathscr{R}_t(\boldsymbol{\theta}) \leq \mathscr{R}_{t^*}(\boldsymbol{\theta}), \qquad \text{para todo } \boldsymbol{\theta}$$

e com desigualdade estrita para algum $\boldsymbol{\theta} = \theta_0$. Logo, $\mathscr{R}_{t^*}(\boldsymbol{\theta}) - \mathscr{R}_t(\boldsymbol{\theta}) \geq 0$ e como $\mathscr{R}_t(\boldsymbol{\theta}) \leq \mathscr{R}_{t^*}(\boldsymbol{\theta})$ e $\mathscr{R}$ é uma função contínua de $\boldsymbol{\theta}$, existe um $\varepsilon > 0$ e um intervalo $\theta_0 \pm h$ tal que

$$\mathscr{R}_{t^*}(\boldsymbol{\theta}) - \mathscr{R}_t(\boldsymbol{\theta}) > \varepsilon \qquad \text{para} \qquad \theta_0 - h \leq \theta \leq \theta_0 + h.$$

Então,

$$\begin{aligned}\int_{-\infty}^{\infty} \left[\mathscr{R}_{t^*}(\boldsymbol{\theta}) - \mathscr{R}_t(\boldsymbol{\theta})\right] g(\boldsymbol{\theta}) d\boldsymbol{\theta} &\geq \int_{\theta_0-h}^{\theta_0+h} \left[\mathscr{R}_{t^*}(\boldsymbol{\theta}) - \mathscr{R}_t(\boldsymbol{\theta})\right] g(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &> \varepsilon \int_{\theta_0-h}^{\theta_0+h} g(\boldsymbol{\theta}) d\boldsymbol{\theta} > 0\end{aligned}$$

mas, isto contradiz o fato que $\mathbf{T}^*$ é um estimador de Bayes porque um estimador de Bayes tem a seguinte propriedade

$$r_{l,g}(\mathbf{t}^*) - r_{l,g}(\mathbf{t}) = \int_{-\infty}^{\infty} [\mathscr{R}_{t^*}(\boldsymbol{\theta}) - \mathscr{R}_t(\boldsymbol{\theta})] g(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq 0.$$

■

O teorema identifica uma certa classe de estimadores que são admissíveis, no entanto, existem muitos estimadores admissíveis, em particular um estimador para cada *priori*, e certamente alguns desses estimadores não são adequados. Com este teorema, prova-se que os estimadores admissíveis são estimadores de Bayes ou limites de estimadores de Bayes.

2.3 Estimadores de encolhimento (shrinkage)

Os estimadores não viesados apresentam uma deficiência que geralmente não é explicitada. Seja $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, ..., \hat{\theta}_p\right)$ um estimador não viesado de um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$. Tem - se que

$$\begin{aligned} E\left[||\hat{\boldsymbol{\theta}}||^2\right] &= E\left[\sum_{i=1}^{p} \hat{\theta}_i^2\right] = \sum_{i=1}^{p} E\left[\hat{\theta}_i^2\right] \\ &= \sum_{i=1}^{p} \left(var\left[\hat{\theta}_i\right] + \left(E\left[\hat{\theta}_i\right]\right)^2\right) \\ &= \sum_{i=1}^{p} var\left[\hat{\theta}_i\right] + \sum_{i=1}^{p} \theta_i^2 \\ &= \sum_{i=1}^{p} var\left[\hat{\theta}_i\right] + ||\boldsymbol{\theta}||^2. \end{aligned}$$

Desta forma $||\hat{\boldsymbol{\theta}}||^2$, é um estimador viesado de $||\boldsymbol{\theta}||^2$ com tendência a superestimar o valor $||\boldsymbol{\theta}||^2$. Tal fato é surpreendentemente não intuitivo, pois apesar de $\hat{\theta}_i$ ser bom estimador de β_i, isto é, gera estimativas próximas de θ_i, $||\hat{\boldsymbol{\theta}}||$, tem tendência a ser maior do que $\|\boldsymbol{\theta}\|$. A demonstração anterior deste fato em nada contribui para que se tenha algum tipo de intuição sobre este fenômeno. Uma argumentação geométrica pode contribuir na interpretação de tal fato.

Essa deficiência dos estimadores não viesados foi uma das motivações para obtenção do estimador de James-Stein, um estimador viesado que possui a vantagem de reduzir a variância e consequentemente reduz o erro quadrático médio, e adotar a estratégia de se propor estimadores obtidos por encolhimento de estimadores não viesados. Em geral, os estimadores de encolhimento, possuem Erro Quadrático Médio (EQM) menor, ou ainda, possuem o risco quadrático menor do que os estimadores não viesados, que é enfatizado por James e Stein (1961) em seu clássico teorema, que será exposto neste trabalho.

A seguir são apresentados alguns exemplos de estimadores de encolhimento.

Exemplo 2.4 *(O processo de encolhimento mais simples)*

Seja $\boldsymbol{B}$ *um estimador não viesado do vetor de parâmetros* $\boldsymbol{\beta}$*. A ideia é simplesmente fazer um encolhimento linear do vetor* $\boldsymbol{B}$*, isto é, considerar o estimador* $\boldsymbol{\alpha B}$ *,* $0 < \boldsymbol{\alpha} < 1$*. O erro quadrático médio*

desses estimadores é:

$$\begin{aligned} EQM(\boldsymbol{\alpha}\,\boldsymbol{B}) &= E\left[\|\boldsymbol{\alpha B}-\boldsymbol{\beta}\|^2\right] \\ &= E\left[(\boldsymbol{\alpha B}-\boldsymbol{\alpha\beta}+\boldsymbol{\alpha\beta}-\boldsymbol{\beta})\cdot(\boldsymbol{\alpha B}-\boldsymbol{\alpha\beta}+\boldsymbol{\alpha\beta}-\boldsymbol{\beta})\right] \\ &= E\left[\|\boldsymbol{\alpha}\,\boldsymbol{B}-\boldsymbol{\alpha\beta}\|^2+2\,(\boldsymbol{\alpha}\,\boldsymbol{B}-\boldsymbol{\alpha\beta})\cdot(\boldsymbol{\alpha\beta}-\boldsymbol{\beta})+\|\boldsymbol{\alpha\beta}-\boldsymbol{\beta}\|^2\right] \\ &= E\left[\|\boldsymbol{\alpha B}-\boldsymbol{\alpha\beta}\|^2\right]+2\,(\boldsymbol{\alpha\beta}-\boldsymbol{\beta})\,E\left[(\boldsymbol{\alpha B}-\boldsymbol{\alpha\beta})\right]+E\left[\|\boldsymbol{\alpha\beta}-\boldsymbol{\beta}\|^2\right] \\ &= \boldsymbol{\alpha}^2E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]+2(\boldsymbol{\alpha\beta}-\boldsymbol{\beta})\cdot[\boldsymbol{\alpha}E(\boldsymbol{B})-\boldsymbol{\alpha\beta})]+(\boldsymbol{\alpha}-1)^2\,\|\boldsymbol{\beta}\|^2 \\ &= \boldsymbol{\alpha}^2E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]+(1-\boldsymbol{\alpha})^2\,\|\boldsymbol{\beta}\|^2\,. \end{aligned}$$

Para minimizar o erro quadrático médio, derivamos em relação a $\boldsymbol{\alpha}$

$$\frac{d}{d\boldsymbol{\alpha}}EQM(\boldsymbol{\alpha B})=2\boldsymbol{\alpha}E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]-2(1-\boldsymbol{\alpha})\|\boldsymbol{\beta}\|^2=0$$

logo, o valor ótimo para o encolhimento é

$$\boldsymbol{\alpha}=\frac{\|\boldsymbol{\beta}\|^2}{E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]+\|\boldsymbol{\beta}\|^2}.$$

Para encontrar os valores de $\boldsymbol{\alpha}$ *para os quais* $E\left[\|\boldsymbol{\alpha}\,\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]<E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]$ *têm-se*

$$\begin{aligned} E\left[\|\boldsymbol{\alpha}\,\boldsymbol{B}-\boldsymbol{\beta}\|^2\right] &< E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right] \qquad (2.6)\\ \boldsymbol{\alpha}^2E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]+(1-\boldsymbol{\alpha})^2\,\|\boldsymbol{\beta}\|^2 &< E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right] \\ (E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]+\|\,\boldsymbol{\beta}\|^2)\boldsymbol{\alpha}^2-2\,\|\boldsymbol{\beta}\|^2\,\boldsymbol{\alpha}+\|\boldsymbol{\beta}\|^2-E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right] &< 0. \end{aligned}$$

para facilitar a notação considere $E\left[\|\boldsymbol{B}-\boldsymbol{\beta}\|^2\right]=z$*, assim*

$$(z+\|\,\boldsymbol{\beta}\|^2)\boldsymbol{\alpha}^2-2\,\|\boldsymbol{\beta}\|^2\,\boldsymbol{\alpha}+\|\boldsymbol{\beta}\|^2-z<0$$

As raízes desta equação do segundo grau são

$$\begin{aligned}
\boldsymbol{\alpha} &= \frac{2\left\|\boldsymbol{\beta}\right\|^2 \pm \sqrt{4\left\|\boldsymbol{\beta}\right\|^4 - 4(z+\left\|\boldsymbol{\beta}\right\|^2)(\left\|\boldsymbol{\beta}\right\|^2 - z)}}{2(z+\left\|\boldsymbol{\beta}\right\|^2)} \\
\boldsymbol{\alpha} &= \frac{2\left\|\boldsymbol{\beta}\right\|^2 \pm \sqrt{4\left\|\boldsymbol{\beta}\right\|^4 - 4\left(\left\|\boldsymbol{\beta}\right\|^4 - z^2\right)}}{2(z+\left\|\boldsymbol{\beta}\right\|^2)} \\
\boldsymbol{\alpha} &= \frac{2\left\|\boldsymbol{\beta}\right\|^2 \pm \sqrt{4z^2}}{2(z+\left\|\boldsymbol{\beta}\right\|^2)} \\
\boldsymbol{\alpha} &= \frac{\left\|\boldsymbol{\beta}\right\|^2 \pm z}{z+\left\|\boldsymbol{\beta}\right\|^2} \\
\boldsymbol{\alpha}' = 1 \quad e \quad \boldsymbol{\alpha}'' &= \frac{\left\|\boldsymbol{\beta}\right\|^2 - z}{z+\left\|\boldsymbol{\beta}\right\|^2}
\end{aligned}$$

substituindo $z = E\left[\left\|\boldsymbol{B}-\boldsymbol{\beta}\right\|^2\right]$ *temos que, para* $\boldsymbol{\alpha}$ *no intervalo* $\left(\frac{\left\|\boldsymbol{\beta}\right\|^2 - E\left[\left\|\boldsymbol{B}-\boldsymbol{\beta}\right\|^2\right]}{E\left[\left\|\boldsymbol{B}-\boldsymbol{\beta}\right\|^2\right] + \left\|\boldsymbol{\beta}\right\|^2}, 1\right)$,

$$E\left[\left\|\boldsymbol{\alpha}\,\boldsymbol{B}-\boldsymbol{\beta}\right\|^2\right] < E\left[\left\|\boldsymbol{B}-\boldsymbol{\beta}\right\|^2\right].$$

Os valores adequados de encolhimento dependem de parâmetros populacionais e, portanto, precisam ser estimados.

Exemplo 2.5 *(O estimador da variância)*

Conforme exemplo de Casella e Berger (2010), a variância é estimada de forma não viesada pela variância amostral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

No entanto, o estimador da variância pelo método dos momentos é dado por

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \\
&= \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \\
&= \frac{n-1}{n}S^2.
\end{aligned}$$

Portanto, o estimador pelo método dos momentos é um encolhimento do estimador não viesado. Para o caso em que $X \sim N(\theta, \sigma^2)$*, queremos comparar o* $EQM(\hat{\sigma}^2)$ *com* $EQM(S^2) = var[S^2]$*. Temos que, de*

acordo com Mood, Graybill e Boes (1974)

$$\frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{n-1}$$

e

$$var\left[\frac{(n-1)\,S^2}{\sigma^2}\right] = 2(n-1)$$

logo,

$$\begin{aligned}\left(\frac{n-1}{\sigma^2}\right)^2 var\left[S^2\right] &= 2\,(n-1)\\ var\left[S^2\right] &= \frac{2\sigma^4}{(n-1)}.\end{aligned}$$

$$var\left[\hat{\sigma}^2\right] = var\left[\frac{n-1}{n}S^2\right] = \left(\frac{n-1}{n}\right)^2 var\left[S^2\right] = \frac{2\,(n-1)}{n^2}\sigma^4.$$

Portanto,

$$\begin{aligned}EQM\left(\hat{\sigma}^2\right) &= E\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right]\\ &= var\left[\hat{\sigma}^2-\sigma^2\right] + \left(E\left[\hat{\sigma}^2-\sigma^2\right]\right)^2\\ &= var\left[\hat{\sigma}^2\right] - var\left[\sigma^2\right] + \left(E\left[\hat{\sigma}^2\right] - E\left[\sigma^2\right]\right)^2\\ &= \frac{2(n-1)\sigma^4}{n^2} + \left(E\left[\frac{n-1}{n}S^2\right]-\sigma^2\right)^2\\ &= \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}E\left[S^2\right]-\sigma^2\right)^2\\ &= \frac{2\,(n-1)\,\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2-\sigma^2\right)^2\\ &= \frac{2(n-1)\sigma^4}{n^2} + \left(\left(1-\frac{1}{n}\right)\sigma^2-\sigma^2\right)^2\\ &= \frac{2(n-1)\sigma^4}{n^2} + \frac{1}{n^2}\sigma^4\\ &= \left(\frac{2n-1}{n^2}\right)\sigma^4.\end{aligned}$$

Assim, $E\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right] < E\left[\left(S^2-E\left[S^2\right]\right)^2\right]$*, isto é,*

$$\begin{aligned} EQM(\hat{\sigma}^2) &< EQM(S^2) \\ \left(\frac{2n-1}{n^2}\right)\sigma^4 &< \left(\frac{2}{n-1}\right)\sigma^4 \\ (2n-1)(n-1) &< 2n^2 \\ 1 &< 3n. \end{aligned}$$

Logo $\hat{\sigma}^2$*, que é um estimador de encolhimento, tem EQM menor do que* S^2*. No entanto,* $\hat{\sigma}^2$ *é viesado e irá, em média, subestimar* σ^2*.*

Exemplo 2.6 *(Estimador de Bayes da Binomial)*

Conforme exemplo abordado em Casella e Berger (2010), sejam $X_1, \ldots, X_n$ *independentes e identicamente distribuídas Bernoulli* (p)*. O estimador de máxima verossimilhança de* p *é dado por* $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$*,* $\hat{p}$ *é não viesado e, portanto, o EQM é igual a variância, isto é,*

$$E\left[(\hat{p}-p)^2\right] = \frac{p\,(1-p)}{n}.$$

Seja $Y = \sum_{i=1}^{n} X_i$*, logo* Y *tem distribuição Binomial*(n,p)*. Vamos assumir que a distribuição a priori de* p *é uma Beta*$(\boldsymbol{\alpha},\beta)$*. Assim, aplicando o teorema de Bayes*

$$\begin{aligned} g\,(p|y) &\propto g(y|p)\pi(p) \\ g\,(p|y) &\propto p^y(1-p)^{n-y}p^{\boldsymbol{\alpha}-1}(1-p)^{\beta-1} \\ &\propto p^{y+\boldsymbol{\alpha}-1}(1-p)^{n-y+\beta-1}. \end{aligned}$$

Nestas condições, a distribuição a posteriori $g(p|y)$ *é uma* $Beta(y+\boldsymbol{\alpha}, n-y+\beta)$*.*

$$g\,(p|y) = \frac{1}{B\,(y+\boldsymbol{\alpha}, n-y+\beta)}p^{y+\boldsymbol{\alpha}-1}(1-p)^{n-y+\beta-1}.$$

O estimador de Bayes da Binomial é dado pela média da posteriori. Para a distribuição $Beta(\boldsymbol{\alpha},\beta)$*,*

a média é $\dfrac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta}$*. Portanto,o estimador de Bayes é dado por*

$$\hat{p_B} = \frac{Y+\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n}.$$

Observe que $\hat{p}_B$ *é um estimador de encolhimento em relação ao estimador usual. O EQM de* $\hat{p}_B$ *é*

$$\begin{aligned}
E\left[(\hat{p}_B - p)^2\right] &= Var\left[\hat{p}_B\right] + \left(E\left[\hat{p}_B - p\right]\right)^2 \\
&= Var\left[\frac{Y+\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n}\right] + \left(E\left[\frac{y+\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n}\right] - p\right)^2 \\
&= Var\left[\frac{y}{\boldsymbol{\alpha}+\beta+n} + \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n}\right] + \left(\frac{1}{\boldsymbol{\alpha}+\beta+n}\left(E\left[y\right] + E\left[\boldsymbol{\alpha}\right]\right) - p\right)^2 \\
&= \frac{1}{\left(\boldsymbol{\alpha}+\beta+n\right)^2} Var\left[y\right] + \left(\frac{np+\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n} - p\right)^2 \\
&= \frac{np\left(1-p\right)}{\left(\boldsymbol{\alpha}+\beta+n\right)^2} + \left(\frac{np+\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\beta+n} - p\right)^2.
\end{aligned}$$

Na ausência de boas informações a priori, podemos tentar escolher $\boldsymbol{\alpha}$ *e* β *para tornar o EQM de* $\hat{p}_B$ *constante. Se* $\boldsymbol{\alpha} = \beta = \sqrt{\dfrac{n}{4}}$ *o EQM é constante não dependendo do parâmetro* p*.*

$$\hat{p}_B = \frac{Y + \sqrt{\dfrac{n}{4}}}{n+\sqrt{n}}.$$

Neste caso, o estimador é minimax pelo teorema 2.5.

$$\begin{aligned}
E\left[(\hat{p}_B - p)^2\right] &= \frac{np\left(1-p\right)}{\left(\sqrt{n/4} + \sqrt{n/4} + n\right)^2} + \left(\frac{np + \sqrt{n/4}}{\sqrt{n/4} + \sqrt{n/4} + n} - p\right)^2 \\
&= \frac{n}{4(n+\sqrt{n})^2}.
\end{aligned}$$

Se quisermos escolher entre $\hat{p}_B$ *e* $\hat{p}$ *com base no EQM, para pequenos valores de* n*,* $\hat{p}_B$ *é a melhor opção. Para grandes valores de* n*,* $\hat{p}$ *é a melhor escolha. Mesmo que o EQM não mostre que um estimador é uniformemente melhor que o outro, informações úteis são fornecidas.*

2.3.1 Justificativa Bayesiana para alguns estimadores de encolhimento

Discutiremos a abordagem Bayesiana para alguns dos estimadores de encolhimento estudados anteriormente, segundo resultados discutidos em Gruber (1998). Considere uma variável aleatória $Y =$

$(Y_1, ..., Y_n)$ com distribuição normal e *priori* θ

$$Y \sim N(\theta, \sigma^2) \qquad \sigma^2\text{conhecido}$$
$$\theta \sim N(0, \tau^2).$$

Neste caso, a distribuição a posteriori é [*vide* Apêndice A]

$$\theta|Y \sim N\left(\frac{\tau^2\bar{Y}}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$$

em que o estimador de Bayes de θ é dado pela média da posteriori

$$\hat{\theta}_{Bayes} = \frac{\tau^2}{\sigma^2+\tau^2}\bar{Y} = \left(1 - \frac{\sigma^2}{\sigma^2+\tau^2}\right)\bar{Y}.$$

Este estimador também pode ser obtido através do conceito de mistura. Temos as densidades de probabilidades

$$f_{\bar{Y}}(\bar{y}) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}}\exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\}$$
$$f_\theta(\theta) = \frac{1}{\sqrt{2\pi\tau^2}}\exp\left\{-\frac{1}{2\tau^2}\left(\theta^2\right)\right\}.$$

Tem-se então, a densidade conjunta $f_{\bar{Y},\theta}(\bar{y},\theta) = f_{\bar{Y}}(\bar{y})f_\theta(\theta)$. O EQM é dado pela relação

$$E[(c\bar{Y}-\theta)^2] \quad = \quad E[c^2\bar{Y}^2 - 2c\bar{Y}\theta + \theta^2] \tag{2.7}$$

em que a esperança é tomada em relação a densidade conjunta, isto é,

$$E[c^2\bar{Y}^2 - 2c\bar{Y}\theta + \theta^2] = \int\int (c^2\bar{y}^2 - 2c\bar{y}\theta + \theta^2) f_{\bar{Y}}(\bar{y}) f_\theta(\theta) d\theta d\bar{y}.$$

Calculando a integral termo a termo:

$$\begin{aligned}
E[c^2\bar{y}^2] &= \int\int c^2\bar{y}^2 \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\left(\theta^2\right)\right\} d\theta d\bar{y} \\
&= c^2 \int \bar{y}^2 \left[\int \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\left(\theta^2\right)\right\} d\theta\right] d\bar{y} \\
&= c^2 \int \bar{y}^2 \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n}+\tau^2\right)}} \exp\left\{-\frac{1}{2\left(\frac{\sigma^2}{n}+\tau^2\right)}\bar{y}^2\right\} d\bar{y} \\
&= c^2 E[\bar{y}^2] \\
&= c^2\left[var\left[\bar{y}\right] + \left(E[\bar{y}]\right)^2\right] \\
&= c^2\left(\frac{\sigma^2}{n}+\tau^2\right)
\end{aligned}$$

$$\begin{aligned}
E[\theta\bar{y}] &= \int\int \theta\bar{y} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\left(\theta^2\right)\right\} d\theta d\bar{y} \\
&= \int \theta \left[\int \bar{y} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\} d\bar{y}\right] d\theta \\
&= \int \theta \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\} \left[\int \bar{y} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{y}-\theta)^2\right\} d\bar{y}\right] d\theta \\
&= \int \theta \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\} E[\bar{y}] d\theta \\
&= \int \theta^2 \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}\theta^2\right\} d\theta \\
&= E[\theta^2] \\
&= \left(var[\theta] + \left(E[\theta]\right)^2\right) \\
&= \tau^2.
\end{aligned}$$

Por fim, temos que

$$\begin{aligned}
E[\theta^2] &= var[\theta] + \left(E[\theta]\right)^2 \\
&= \tau^2.
\end{aligned}$$

Substituindo na equação (2.7):

$$\begin{aligned} E[(c\bar{y}-\theta)^2] &= E[c^2\bar{y}^2 - 2c\bar{y}\theta + \theta^2] \\ &= E[c^2\bar{y}^2] - 2cE[\bar{y}\theta] + E[\theta^2] \\ &= c^2\left(\frac{\sigma^2}{n}+\tau^2\right) - 2c\tau^2 + \tau^2. \end{aligned}$$

Derivando em relação a c e igualando a zero, tem-se

$$c_o = \frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}.$$

Portanto, em relação à densidade mistura o estimador de erro quadrático médio mínimo é dado por

$$\hat{\theta}_{mistura} = \frac{\tau^2}{\frac{\sigma^2}{n}+\tau^2}\bar{Y}.$$

Note a diferença em relação ao estimador da média com erro quadrático mínimo

$$\hat{\theta} = \frac{\theta^2}{\frac{\sigma^2}{n}+\theta^2}\bar{Y}.$$

e o estimador de Bayes

$$\hat{\theta}_{Bayes} = \frac{\tau^2}{\sigma^2+\tau^2}\bar{Y}.$$

Se o parâmetro τ^2 d*a priori* é desconhecido este pode ser, pelo método de Bayes empírico, estimado a partir dos dados. Para isso, é necessário calcular a distribuição preditiva. Considerando

$$\bar{Y} \sim N(\theta, \frac{\sigma^2}{n}) \qquad \text{e} \qquad \theta \sim N(0, \tau^2)$$

logo, [Apêndice A]

$$f(\bar{Y}) \sim N\left(0, \frac{\sigma^2}{\sigma^2+\tau^2}\right).$$

Considerando $\bar{Y}$ como uma observação da preditiva, pelo método dos momentos $\left(\frac{\sigma^2}{n}+\tau^2\right)$ pode ser estimado por $\bar{Y}^2$. É razoável supor que $\tau^2 \gg \sigma^2$, significando que a ignorância em relação *a priori* é maior do que em relação a população original. Desta forma o estimador de Bayes fica

$$\hat{\theta}_{Bayes} = \left(\frac{\bar{Y}^2}{\frac{S^2}{n}+\bar{Y}^2}\right)\bar{Y}.$$

Portanto, $\hat{\theta} = \hat{\theta}_{mistura} = \hat{\theta}_{Bayes}$.

2.3.2 Estimadores de Cumeeira

O método de regressão de cumeeira (*ridge*) é um caso particular para se obter estimadores de encolhimento. As estimativas de quadrados mínimos são encolhidas na direção do vetor nulo.

A construção de bons estimadores para os coeficientes de regressão linear múltipla em situações em que se tem quase multicolinearidade, $\det(\mathbf{X'X}) \approx 0$, é um problema que tornou-se um grande desafio para os estatísticos da década de 60 (HOERL; KENNARD, 1970a).

Considere o modelo de regressão linear múltipla

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

sendo $\mathbf{X}$ uma matriz $n \times p$ de variáveis regressoras e de posto p, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ o vetor dos coeficientes de regressão e $\boldsymbol{\varepsilon}$ o vetor de erros normalmente distribuído com média 0 e matriz de covariâncias $\sigma^2\mathbf{I_n}$. Supondo, sem perda de generalidade, que $\mathbf{X'X}$ está na forma de correlação o que implica que $tr(\mathbf{X'X}) = \mathbf{p}$.

O estimador de quadrados mínimos de $\boldsymbol{\beta}$ é dado por

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

A matriz de variâncias e covariâncias de $\hat{\boldsymbol{\beta}}$ é dada por $\mathrm{cov}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X'X})^{-\mathbf{1}}$. Logo, variância total é dada por

$$\begin{aligned}
var_{total}\left[\hat{\boldsymbol{\beta}}\right] &= tr\left(\mathrm{cov}[\hat{\boldsymbol{\beta}}]\right) \\
&= tr\left(\sigma^2(\mathbf{X'X})^{-1}\right) \\
&= \sigma^2 tr(\mathbf{X'X})^{-\mathbf{1}} \\
&= \sigma^2 tr(\mathbf{PP'X'X})^{-\mathbf{1}} \\
&= \sigma^2 tr(\mathbf{P'}\left(\mathbf{X'X}\right)^{-\mathbf{1}}\mathbf{P}) \\
&= \sigma^2 tr\begin{pmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_p} \end{pmatrix} \\
&= \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i}
\end{aligned}$$

em que $\mathbf{P_{p\times p}}$ é uma matriz ortogonal que diagonaliza $\mathbf{X'X}$. A matriz $\mathbf{X'X}$ é simétrica e positiva definida, logo, admite p autovalores $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq ...\lambda_p = \lambda_{min} > 0$. Tem-se então que a variância total será severamente inflada se pelo menos um dos autovalores for muito pequeno. Este é o fenômeno denominado de quase multicolinearidade da matriz $\mathbf{X'X}$. Em problemas de regressão linear com um número grande de covariáveis, é comum que algumas dessas variáveis, ou combinações lineares entre elas, sejam altamente correlacionadas. Tal fato implicará que alguns dos vetores colunas da matriz $\mathbf{X}$ ou combinações lineares dessas colunas, estão próximos sob um ponto de vista geométrico. Tal fato implica que o determinante da matriz quadrada $\mathbf{X'X}$ é próximo de zero. Como o determinante é produto dos autovalores pode-se afirmar que pelo menos um deles é próximo de zero. Neste caso, a variabilidade total do estimador de quadrados mínimos é alta e pode inviabilizar o seu uso. Como o estimador de quadrados mínimos é o melhor entre os não viesados uma possibilidade para se contornar este problema é a utilização de estimadores viesados (COSTA, 2015).

Hoerl e Kennard (1970a e 1970b) propuseram o uso de estimadores obtidos por encolhimento do estimador de quadrados mínimos denominados estimadores de cumeeira (ridge).

O estimador "ridge" é obtido pela simples adição de uma matriz escalar $k\mathbf{I}$ na expressão do estimador de quadrados mínimos,

$$\hat{\boldsymbol{\beta}}(k) = (\mathbf{X'X} + \mathbf{kI})^{-\mathbf{1}}\mathbf{X'Y} \tag{2.8}$$

em que k é uma constante positiva. O estimador "ridge" está relacionado ao estimador de quadrados mínimos $\hat{\boldsymbol{\beta}}$ pela relação

$$\hat{\boldsymbol{\beta}}(k) = \left(\mathbf{I} + k(\mathbf{X'X})^{-1}\right)^{-1}\hat{\boldsymbol{\beta}}.$$

Para provar que para cada k, suficientemente pequeno, iremos usar a propriedade

$$\left\|\hat{\boldsymbol{\beta}}(k)\right\| < \left\|\hat{\boldsymbol{\beta}}\right\|,$$

que nos garante que o estimador de cumeeira é um estimador de encolhimento.

Este estimador é viesado e possui erro quadrático médio dado por

$$EQM\left(\hat{\boldsymbol{\beta}}(k)\right) = \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + k)^2} + k^2\boldsymbol{\beta}'\left(\mathbf{X'X} + \mathbf{kI}\right)^{-2}\boldsymbol{\beta}.$$

O desenvolvimento pode ser verificado com mais detalhes em Costa (2015). Um dos pontos centrais da teoria é obter um valor k que minimiza $EQM\left(\hat{\boldsymbol{\beta}}(k)\right)$. Este valor ótimo depende dos parâmetros populacionais

σ^2 e $\boldsymbol{\beta}$ e vários estimadores na literatura foram propostos para o seu cálculo. Um deles proposto por Lawless e Wang (1976), tem motivação Bayesiana e, em razão deste fato a exposição que será dada a seguir, com detalhes e expansão dos resultados encontrados neste artigo, tem o objetivo de elucidar as passagens mais complicadas do artigo.

Seja $\mathbf{P}$ uma matriz ortogonal que diagonaliza $\mathbf{X}'\mathbf{X}$, isto é,

$$\mathbf{P}\left(\mathbf{X}'\mathbf{X}\right)\mathbf{P}' = \boldsymbol{\Lambda}$$

em que $\boldsymbol{\Lambda}$ é a matriz diagonal formada pelos autovalores λ_i da matriz $\mathbf{X}'\mathbf{X}$. Considere agora a reparametrização $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$. Os parâmetros $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_p)$ são denominados parâmetros canônicos. Com os parâmetros canônicos temos a regressão

$$\mathbf{P}\left(\mathbf{X}'\mathbf{X}\right)\mathbf{P}' = \boldsymbol{\Lambda} \Rightarrow \left(\mathbf{X}\mathbf{P}'\right)'\left(\mathbf{X}\mathbf{P}'\right) = \boldsymbol{\Lambda},$$

Como $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$ segue que $\boldsymbol{\beta} = \mathbf{P}'\boldsymbol{\alpha}$ e fazendo $\mathbf{X}^* = \mathbf{X}\mathbf{P}'$ temos

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\mathbf{P}'\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \mathbf{X}^*\boldsymbol{\alpha} + \boldsymbol{\varepsilon}.$$

O problema de regressão está na forma canônica, cuja vantagem está no fato de que em relação aos parâmetros originais, o erro quadrático não se altera. De fato, se $\hat{\boldsymbol{\beta}}$ é um estimador de $\boldsymbol{\beta}$ tem-se

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \mathbf{P}\hat{\boldsymbol{\beta}} \\
&= \mathbf{P}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{P}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{P}'\mathbf{P}\mathbf{X}'\mathbf{Y} \\
&= (\mathbf{P}\mathbf{X}'\mathbf{X}\mathbf{P}')^{-\mathbf{1}}(\mathbf{X}\mathbf{P}')'\mathbf{Y} \\
&= ((\mathbf{X}\mathbf{P}')'\mathbf{X}\mathbf{P}')^{-\mathbf{1}}(\mathbf{X}\mathbf{P}')'\mathbf{Y} \\
&= ((\mathbf{X}^*)'\mathbf{X}^*)^{-1}(\mathbf{X}^*)'\mathbf{Y} \\
&= \Lambda^{-1}(\mathbf{X}^*)'\mathbf{Y}. \qquad (2.9)
\end{aligned}$$

$\hat{\boldsymbol{\alpha}}$ é o estimador de quadrados mínimos de $\mathbf{Y} = \mathbf{X}^*\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. E ainda

$$\begin{aligned} cov\left[\hat{\boldsymbol{\alpha}}\right] &= cov\left[\mathbf{P}\hat{\boldsymbol{\beta}}\right] \\ &= \mathbf{P}\mathbf{cov}\left[\hat{\boldsymbol{\beta}}\right]\mathbf{P}' \\ &= \sigma^2\mathbf{P}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{P}' \\ &= \sigma^2\boldsymbol{\Lambda}^{-1} \\ &= \sigma^2\begin{pmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_p} \end{pmatrix}. \end{aligned} \tag{2.10}$$

$\hat{\boldsymbol{\beta}}(k)$ é o estimador de cumeeira para a regressão $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ e $\hat{\boldsymbol{\alpha}}(k)$ é o estimador de cumeeira para $\mathbf{Y} = \mathbf{X}^*\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$.

$$\begin{aligned} \hat{\boldsymbol{\beta}}(k) &= \left(\mathbf{I_p} + k(\mathbf{X}'\mathbf{X})^{-1}\right)^{-1}\hat{\boldsymbol{\beta}} \\ \mathbf{P}\hat{\boldsymbol{\beta}}(\mathbf{k}) &= \mathbf{P}\left(\mathbf{I_p} + \mathbf{k}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\right)^{-\mathbf{1}}\hat{\boldsymbol{\beta}} \\ \mathbf{P}\hat{\boldsymbol{\beta}}(\mathbf{k}) &= \mathbf{P}\left(\mathbf{I_p} + \mathbf{k}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\right)^{-\mathbf{1}}\left(\mathbf{P}'\mathbf{P}\right)\hat{\boldsymbol{\beta}} \\ \mathbf{P}\hat{\boldsymbol{\beta}}(\mathbf{k}) &= \left(\mathbf{P}\mathbf{P}' + \mathbf{k}\mathbf{P}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{P}'\right)^{-1}\mathbf{P}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}}(k) &= \left(\mathbf{I} + \mathbf{k}\boldsymbol{\Lambda}^{-\mathbf{1}}\right)^{-1}\hat{\boldsymbol{\alpha}} \\ &= \begin{pmatrix} \frac{\lambda_1}{\lambda_1+k} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2+k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p+k} \end{pmatrix}\begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix} \end{aligned}$$

e em termos das coordenadas

$$\hat{\alpha}_i(k) = \frac{\lambda_i}{\lambda_i + k}\widehat{\alpha}_i, \tag{2.11}$$

isto é, temos um encolhimento do estimador de quadrados mínimos $\widehat{\alpha}_i$. O erro quadrático médio para o

estimador de cada coordenada é

$$\begin{aligned} E\left[(\hat{\alpha}_i(k)-\alpha_i)^2\right] &= E\left[\hat{\alpha}_i^2(k)-2\hat{\alpha}_i(k)\alpha_i+\alpha_i^2\right] \\ &= E\left[\hat{\alpha}_i^2(k)\right]-2\alpha_i E\left[\hat{\alpha}_i(k)\right]+\alpha_i^2 \\ &= \frac{\lambda_i^2}{(\lambda_i+k)^2}E\left[\hat{\alpha}_i^2\right]-2\alpha_i\left(\frac{\lambda_i}{\lambda_i+k}\right)E\left[\hat{\alpha}_i\right]+\alpha_i^2 \\ &= \frac{\lambda_i^2}{(\lambda_i+k)^2}\left(\frac{\sigma^2}{\lambda_i}+\alpha_i^2\right)-2\alpha_i\frac{\lambda_i}{\lambda_i+k}\alpha_i+\alpha_i^2. \end{aligned}$$

Para obter o k que minimiza o erro quadrático médio

$$\begin{aligned} \frac{d}{dk}E\left[(\hat{\alpha}_i-\alpha_i)^2\right] &= \frac{-\lambda_i^2 2(\lambda_i+k)(\sigma^2+\lambda_i\alpha_i^2)}{(\lambda_i+k)^4\lambda_i}+\frac{2\alpha_i^2\lambda_i}{(\lambda_i+k)^2} \\ &= \frac{-2\lambda_i(\sigma^2+\lambda_i\alpha_i^2)}{(\lambda_i+k)^3}+\frac{2\alpha_i^2\lambda_i}{(\lambda_i+k)^2} \end{aligned}$$

igualando a zero

$$\begin{aligned} \frac{\lambda_i(\sigma^2+\lambda_i\alpha_i^2)}{(\lambda_i+k)^3} &= \frac{\alpha_i^2\lambda_i}{(\lambda_i+k)^2} \\ \frac{\lambda_i^2\left(\frac{\sigma^2}{\lambda_i}+\alpha_i^2\right)}{\alpha_i^2\lambda_i} &= \lambda_i+k \\ k &= \frac{\sigma^2}{\alpha_i^2}+\lambda_i-\lambda_i \\ k &= \frac{\sigma^2}{\alpha_i^2}. \end{aligned}$$

Tem-se então um k_{otimo} para cada coordenada. Um k_{otimo} para todas as coordenadas simultaneamente pode ser definido tomando-se a média $\dfrac{\alpha_1^2+...+\alpha_p^2}{p}$.

$$k_{otimo} = \frac{\sigma^2}{\frac{\sum\alpha_i^2}{p}} = \frac{p\sigma^2}{\sum\alpha_i^2}.$$

Tal fato sugere a definição do estimador

$$\hat{k}_{otimo} = \frac{p\hat{\sigma}^2}{\sum\hat{\alpha}_i^2}$$

e utilizando a equação (2.11) obtemos o correspondente estimador de cumeeira

$$\hat{\alpha}_i(\hat{k}_{otimo}) = \left(\frac{\lambda_i}{\lambda_i + \dfrac{p\hat{\sigma}^2}{\sum \hat{\alpha}_i}} \right) \hat{\alpha}_i, \tag{2.12}$$

denominado estimador Hoerl-Kennard-Baldwin (HKB). Uma crítica a este estimador é o fato de se estimar α_i^2 por $\hat{\alpha}_i^2$, pois

$$E\left[\hat{\alpha}_i^2\right] = \text{var}\left[\hat{\alpha}_i\right] + \left(E\left[\hat{\alpha}_i\right]\right)^2 = \frac{\sigma^2}{\lambda_i} + \alpha_i^2 \qquad \text{utilizando (2.10)}$$

e, portanto, com alta probabilidade $\hat{\alpha}_i^2$ superestima α_i^2. Conclui-se, então, que estimar $\frac{\sigma^2}{\alpha_i^2}$ por $\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$ é um procedimento inadequado.

Lawless e Wang (1976) propuseram um estimador para o k_{otimo} substituindo $\frac{1}{p}\sum_i \hat{\alpha}_i^2$ por uma média ponderada. Como o modelo está na forma de correlação, $p = tr(\mathbf{X'X}) = \sum_{\mathbf{i}} \lambda_{\mathbf{i}}$, a média ponderada é dada por $\frac{1}{p}\sum_i \lambda_i\hat{\alpha}_i^2$, obtendo o seguinte estimador

$$\hat{k}_{otimo} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i\hat{\alpha}_i^2},$$

definindo, assim, o estimador dos parâmetros de cumeeira

$$\hat{\alpha}_i(\hat{k}_{otimo}) = \left(\frac{\lambda_i}{\lambda_i + \dfrac{p\hat{\sigma}^2}{\sum \lambda_i\hat{\alpha}_i^2}} \right) \hat{\alpha}_i.$$

A interpretação Bayesiana desse estimador é obtida considerando uma distribuição *a priori* para o vetor de parâmetros $\boldsymbol{\alpha}$,

$$\boldsymbol{\alpha} \sim N_p(\mathbf{0}, \sigma_\alpha^{\mathbf{2}}\mathbf{I_p}).$$

O estimador de Bayes é obtido substituindo $k = \frac{\sigma^2}{\sigma_\alpha^2}$ na expressão (2.11), que é a expressão do estimador de cumeeira,

$$(\hat{\alpha_B})_i = \frac{\lambda_i}{\lambda_i + \frac{\sigma^2}{\sigma_\alpha^2}}\widehat{\alpha_i} \qquad (i = 1, .., p). \tag{2.13}$$

Os parâmetros desconhecidos σ^2 e σ_α^2 devem ser estimados. Uma opção é a de se estimar σ^2 por $\hat{\sigma}^2$ e, quanto ao outro parâmetro desconhecido da priori, σ_α^2, a ideia é utilizar a abordagem de Bayes empírico, estimando

σ_α^2 a partir dos dados observados. A preditiva tem distribuição normal, logo, o estimador Bayesiano empírico para a variância da preditiva é $\frac{1}{p}\sum_1^p \hat{\alpha}_i^2$ e, portanto, superestima σ_α^2. Observe que, como se utiliza $\hat{\sigma}^2$ como estimador de σ^2 e, $\frac{1}{p}\sum_1^p \hat{\alpha}_i^2$ como estimador de σ_α^2, o estimador de Lawless e Wang se reduz ao estimador HKB.

A densidade preditiva de $\mathbf{Y}$ é dada por

$$\mathbf{Y} \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I} + \sigma_\alpha^{\mathbf{2}}\mathbf{\Lambda})$$

[Apêndice A]. Em relação a densidade da preditiva, a esperança da estatística $\sum_1^p \lambda_i \hat{\alpha}_i^2$ é dada por

$$\begin{aligned}
E\left[\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2\right] &= E\left[\hat{\boldsymbol{\alpha}}'\mathbf{\Lambda}\hat{\boldsymbol{\alpha}}\right] \\
&= E\left[\left(\mathbf{\Lambda}^{-1}(\mathbf{X}^*)'\mathbf{Y}\right)'\mathbf{\Lambda}\left(\mathbf{\Lambda}^{-1}(\mathbf{X}^*)'\mathbf{Y}\right)\right] \text{ (utilizando (2.9))} \\
&= E\left[\mathbf{Y}'\mathbf{X}^*\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{\Lambda}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{X}^{*\prime}\mathbf{Y}\right] \\
&= E\left[\mathbf{Y}'\mathbf{X}^{*\prime}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{X}^*\mathbf{Y}\right] \\
&= E\left[\mathbf{Y}'\left(\mathbf{X}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{X}'\right)\mathbf{Y}\right].
\end{aligned}$$

Utilizando o teorema a seguir, que considera a média da forma quadrática $\mathbf{Y}'\mathbf{A}\mathbf{Y}$,

Teorema 2.8 *Se* $\mathbf{Y}$ *é um vetor aleatório com media* $\boldsymbol{\theta}$ *e matriz de covariâncias* $\mathbf{\Sigma}$ *e se A é uma matriz simétrica de constantes, então*

$$E\left[\mathbf{Y}'\mathbf{A}\mathbf{Y}\right] = tr(\mathbf{A}\mathbf{\Sigma}) + \boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}.$$

Demonstração: (RENCHER;SHAALJE, 2008, p.107).

Obtemos que

$$\begin{aligned}
E\left[\mathbf{Y}'\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\right)\mathbf{Y}\right] &= tr\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'(\sigma^{\mathbf{2}}\mathbf{I}+\sigma_{\alpha}^{\mathbf{2}}\mathbf{X}\mathbf{X}')\right)+\mathbf{0}\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\right)\mathbf{0}\\
&= tr\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\sigma^{\mathbf{2}}\mathbf{I}+\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\sigma_{\alpha}^{\mathbf{2}}\mathbf{X}\mathbf{X}'\right)\\
&= tr\left(\sigma^2\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\right)+tr\left(\sigma_{\alpha}^2\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\mathbf{X}\mathbf{X}'\right)\\
&= \sigma^2 tr\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\right)+\sigma_{\alpha}^2 tr\left(\mathbf{X}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{X}'\mathbf{X}\mathbf{X}'\right)\\
&= \sigma^2 tr\left(\boldsymbol{\Lambda}^{-1}\mathbf{X}\mathbf{X}'\right)+\sigma_{\alpha}^2 tr\left(\mathbf{X}\mathbf{X}'\boldsymbol{\Lambda}^{-\mathbf{1}}\right)\\
&= \sigma^2 tr\left(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}\right)+\sigma_{\alpha}^2 tr\left(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1}\right)\\
&= \sigma^2 tr\left(\mathbf{I}\right)+\sigma_{\alpha}^2 tr\left(\mathbf{I}\right)\\
&= \sigma^2 p+\sigma_{\alpha}^2 p.
\end{aligned}$$

Logo,

$$E\left[\frac{\sum_{i=1}^{p}\lambda_i\widehat{\alpha_i}^2}{p\sigma^2}\right]=1+\frac{\sigma_{\alpha}^2}{\sigma^2}\qquad\Rightarrow\qquad E\left[\frac{\sum_{i=1}^{p}\lambda_i\widehat{\alpha_i}^2}{p\sigma^2}\right]-1=\frac{\sigma_{\alpha}^2}{\sigma^2}.$$

Pelo método dos momentos estima-se $\dfrac{\sigma_{\alpha}^2}{\sigma^2}$ por

$$\frac{\sum_{i=1}^{p}\lambda_i\widehat{\alpha_i}^2}{p\sigma^2}-1.$$

Como é razoável supor que $\sigma_{\alpha}^2 \gg \sigma^2$ o quociente $\frac{\sigma_{\alpha}^2}{\sigma^2}$ é um número grande e pode-se desprezar o termo -1 obtendo-se o estimador de Lawless para o valor do k_{otimo} na regressão de cumeeira dado por

$$k_{otimo}=\frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\lambda_i\hat{\alpha}_i^2}.$$

2.4 O estimador de James-Stein

Stein (1956) em seu clássico artigo, *"Inadmissibility of the usual estimator for the mean of a multivariate normal distribution"*, apresentou a prova de que o estimador de máxima verossimilhança, isto é, a média amostral de uma distribuição normal multivariada é inadmissível. Esse resultado foi um choque para o mundo estatístico com uma série de consequências. Em 1961, James e Stein apresentaram explicitamente um estimador que domina o estimador média amostral. Este novo estimador, um caso particular de estimador de encolhimento, ficou referido na literatura como o estimador de James-Stein.

Foram desenvolvidas várias demonstrações analíticas desse resultado de inadmissibilidade do estimador usual da média. A demonstração original de 1961 foi amplamente simplificada pelo próprio Stein em 1981. Ambas as demonstrações serão apresentadas neste trabalho. Juntamente com a primeira demonstração analítica, Stein apresenta uma justificativa geométrica para obtenção de seu estimador. A partir daí várias justificativas geométricas foram construídas. O foco principal desse trabalho é o de explicitar ao máximo estes aspectos geométricos que serão construídos com o auxílio do software Geogebra (GEOGEBRA, 2001). Um outro aspecto que gerou uma linha de pesquisa com vários resultados relevantes é que o estimador de James-Stein pode ser obtido a partir de uma abordagem Bayesiana, em particular, pelo método Bayesiano empírico.

Segundo James e Stein (1961), para uma população normal p-variada com vetor de médias $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ desconhecido e matriz de variâncias e covariâncias igual a matriz identidade $\mathbf{I}$, logo os X_i são independentes com $E[X_i] = \theta_i$ e $var[X_i] = 1$, variância conhecida. Posteriormente, serão estudadas situações em que X_i possui variância desconhecida. O estimador de $\boldsymbol{\theta}$ é $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, ..., \hat{\theta}_p\right)$. O estimador de máxima verossimilhança, que é também a média amostral, é dado por $\hat{\boldsymbol{\theta}}_0(\mathbf{X}) = \mathbf{X} = (X_1, ..., X_p)$. Se um valor $\mathbf{X}$ é observado um estimador para $\boldsymbol{\theta}$ é:

Definição 2.11 *(Estimador de James-Stein)*

$$\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X}) = \left(1 - \frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X}, \qquad \mathbf{p} \geq \mathbf{3}.$$

O coeficiente de encolhimento depende dos dados. Note que este estimador é da forma

$$\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X}) = w(\mathbf{X})(\mathbf{X}),$$

isto é, é um "estimador de encolhimento linear adaptativo", pelo fato do coeficiente de encolhimento depender de $\mathbf{X}$. Cada vetor de observação $\mathbf{X}$ possui um encolhimento proporcional ao fator de encolhimento

$w(\mathbf{X})$.

Apresentado o estimador de James-Stein, a abordagem do tema se inicia com a descrição do lema de Stein que será necessária para desenvolver o processo de obtenção deste estimador, que será feito posteriormente.

2.5 O lema de Stein e algumas de suas aplicações

Stein (1981) publicou um resultado, hoje denominado lema de Stein, que se tornou uma técnica fundamental para obtenção de estimadores não viesados para o risco, em relação a perda quadrática em populações normais multivariadas. Uma das consequências desse lema é uma considerável simplificação na demonstração original da inadmissibilidade do estimador média amostral para populações normais multivariadas com dimensão maior que 3. O interesse atual pelo lema de Stein é muito mais geral que o problema original, estimação da média da normal, em que foi desenvolvido. Esta seção está amplamente baseada no artigo de Stein (1981), com introdução de alguns exemplos e definições com o intuito de contribuir de forma didática com a teoria.

Seja h uma função real de crescimento lento, isto é,

$$\lim_{y\to\pm\infty} h(y)\exp\left[-\frac{1}{2}(y-\theta)^2\right] = 0.$$

Lema 2.1 *(Stein) Seja $Y \sim N(\theta, 1)$. Então*

$$cov\left[Y, h(Y)\right] = E\left[h(Y)(Y-\theta)\right] = E\left[h'(Y)\right], \tag{2.14}$$

em que todas as integrais são supostas finitas.

Demonstração:

$$\begin{aligned}
\operatorname{cov}\left[Y, h(Y)\right] &= E\left[(Y-\theta)\left(h(Y) - E\left[h(Y)\right]\right)\right] \\
&= E\left[(Y-\theta)\,h(Y)\right] - E\left[(Y-\theta)\,E\left[h(Y)\right]\right] \\
&= E\left[(Y-\theta)\,h(Y)\right] - E\left[(Y-\theta)\right]E\left[h(Y)\right] \\
&= E\left[(Y-\theta)\,h(Y)\right].
\end{aligned}$$

Utilizando integração por partes

$$\begin{aligned} E\left[h(Y)(Y-\theta)\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(y)(y-\theta)\exp\left[-\frac{1}{2}(y-\theta)^2\right]dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(y)\left[\frac{d}{dy}\left(-\exp\left[-\frac{1}{2}(y-\theta)^2\right]\right)\right]dy \\ &= -\frac{1}{\sqrt{2\pi}}h(y)\exp\left[-\frac{1}{2}(y-\theta)^2\right]\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h'(y)\exp\left[-\frac{1}{2}(y-\theta)^2\right]dy \\ &= E\left[h'(Y)\right]. \end{aligned}$$

■

O lema pode ser generalizado para variáveis aleatórias com distribuições normais com variância desconhecida. Se $Y \sim N(\theta, \sigma^2)$ seja $X = \dfrac{Y-\theta}{\sigma} \sim N(0,1)$. Como $Y = \sigma X + \theta$ tem-se

$$h(Y) = h(\sigma X + \theta) = g(X)$$

e,

$$\begin{aligned} E\left[h(Y)\left(\frac{Y-\theta}{\sigma}\right)\right] &= E\left[g(X)X\right] \\ &= E\left[g'(X)\right] \quad \text{(pelo lema)} \\ &= E\left[\sigma h'\left(\sigma X + \theta\right)\right] \\ &= E\left[\sigma h'(Y)\right]. \end{aligned}$$

Portanto,

$$E\left[h(Y)\left(Y-\theta\right)\right] = \sigma^2 E\left[h'(Y)\right].$$

Para ilustrar a aplicação do lema, segue o exemplo.

Exemplo 2.7 *Considere o problema do cálculo do n-ésimo momento central de* $Y \sim N(\theta, \sigma^2)$*, com* n *ímpar.*

Temos

$$\begin{aligned}
E\left[(Y-\theta)^n\right] &= E\left[\underbrace{(Y-\theta)^{n-1}}_{h(Y)}(Y-\theta)\right] \\
&= \sigma^2 E\left[(n-1)(Y-\theta)^{n-2}\right] \\
&= (n-1)\sigma^2 E\left[\underbrace{(Y-\theta)^{n-3}}_{h(Y)}(Y-\theta)\right] \\
&= (n-1)\left(\sigma^2\right)^2 E\left[(n-3)(Y-\theta)^{n-4}\right] \\
&= (n-1)(n-3)\left(\sigma^2\right)^2 E\left[(Y-\theta)^{n-4}\right] \\
&\vdots \\
&= (n-1)(n-3)\ldots(n-(n-2))\left(\sigma^2\right)^{\frac{(n-1)}{2}} E\left[(Y-\theta)^{n-(n-1)}\right] \\
&= (n-1)(n-3)\ldots 2\left(\sigma^2\right)^{\frac{(n-1)}{2}}\left(E\left[Y\right]-E\left[\theta\right]\right) \\
&= 0 \qquad (2.15)
\end{aligned}$$

e de forma equivalente para n par temos

$$\begin{aligned}
E\left[(Y-\theta)^n\right] &= E\left[\underbrace{(Y-\theta)^{n-1}}_{h(Y)}(Y-\theta)\right] \\
&\vdots \\
&= (n-1)(n-3)\ldots(n-(n-3))\left(\sigma^2\right)^{\frac{n-2}{2}} E\left[(Y-\theta)^{n-(n-2)}\right] \\
&= (n-1)(n-3)\ldots 3\left(\sigma^2\right)^{\frac{n-2}{2}}\left(\sigma^2\right) \\
&= (n-1)(n-3)\ldots 3\sigma^n.
\end{aligned}$$

O lema de Stein pode ser generalizado para normais multivariadas, sendo assim serão recordados conceitos do cálculo diferencial de funções de várias variáveis.

Definição 2.12 *Seja $f : \mathbb{R}^p \to \mathbb{R}$ uma função real diferenciável com p variáveis, $f(x_1, ..., x_p)$. O campo*

vetorial

$$\nabla f : \mathbb{R}^p \to \mathbb{R}^p$$
$$\nabla f\left(x_1, ..., x_p\right) = \left(\frac{\partial f}{\partial x_1}\left(x_1, ..., x_p\right), ..., \frac{\partial f}{\partial x_p}\left(x_1, ..., x_p\right)\right),$$

é denominado gradiente da função f.

Tem-se a relação para todo $z \in \mathbb{R}^p$

$$f(x+z) - f(x) = \int_0^1 z \cdot \nabla f(x+tz)dt.$$

Lema 2.2 *(Stein) Se* $\mathbf{X} = (X_1, ..., X_p) \sim N(\boldsymbol{\theta}, \mathbf{I})$ *e* $E\left[\|\nabla f(\mathbf{X})\|\right] < \infty$ *então,*

$$E\left[f(\mathbf{X})\left(\mathbf{X} - \boldsymbol{\theta}\right)\right] = E\left[\nabla f(\mathbf{X})\right]. \tag{2.16}$$

Demonstração: Stein (1981).

Exemplo 2.8 *Considere* $f : \mathbb{R}^3 \to \mathbb{R}$*;* $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$*, logo* $\nabla f(x_1, x_2, x_3) = (2x_1, 2x_2, 2x_3)$*, aplicando o lema*

$$E\left[(x_1^2 + x_2^2 + x_3^2)(\mathbf{X} - \boldsymbol{\theta})\right] = E\left[\nabla f(\mathbf{X})\right] = E\left[2\mathbf{X}\right] = 2\boldsymbol{\theta}.$$

Para se obter importantes resultados utilizando este lema, outros conceitos de cálculo vetorial são necessários.

Definição 2.13 *Uma função contínua* $f : \mathbb{R}^p \to \mathbb{R}$ *é super-harmônica no ponto* $x_0 \in \mathbb{R}^p$ *se, para todo* $r > 0$*, a média de* f *sobre a esfera* $S_r(x_0) = \left\{x : \|x - x_0\|^2 = r^2\right\}$ *com raio* r *centrada em* x_0 *não é maior que* $f(x_0)$*, isto é,*

$$\frac{1}{vol(S_r(x_0))} \int_{S_r(x_0)} f(x)ds \leq f(x_0).$$

A função f *é super-harmônica em* $\mathbb{R}^p$ *se ela é super-harmônica em cada* $x_0 \in \mathbb{R}^p$.

Definição 2.14 *Se* $f : \mathbb{R}^p \to \mathbb{R}$ *possui derivadas segundas contínuas, o laplaciano de* f *é uma função* $\nabla^2 f : \mathbb{R}^p \to \mathbb{R}$ *definida por*

$$\nabla^2 f(x) = \frac{\partial^2 f}{\partial x_1^2}(x) + ... + \frac{\partial^2 f}{\partial x_p^2}(x).$$

Lema 2.3 *f é super-harmônica em $\mathbb{R}^p$ se, e somente se, para todo $x \in \mathbb{R}^p$*

$$\nabla^2 f(x) \leq 0. \tag{2.17}$$

Demonstração: Veja Helms (1969, p. 63).

Definição 2.15 *Uma função $f : \mathbb{R}^p \to \mathbb{R}$ duas vezes diferenciável com derivadas segunda contínuas é harmônica se*

$$\nabla^2 f(x) = 0, \qquad \textit{para todo } x \in \mathbb{R}^p. \tag{2.18}$$

Definição 2.16 *Se $h : \mathbb{R}^p \to \mathbb{R}^p$, $h(x_1, ..., x_p) = (h_1(x_1, ..., x_p), ..., h_p(x_1, ..., x_p))$, o divergente de h é a função $\nabla \cdot h : \mathbb{R}^p \to \mathbb{R}$ definida por*

$$\nabla \cdot h(x_1, ..., x_p) = \frac{\partial h_1}{\partial x_1}(x_1, ..., x_p) + ... + \frac{\partial h_p}{\partial x_p}(x_1, ..., x_p).$$

Notação: $\nabla_i h_i = \dfrac{\partial h_i}{\partial x_i}$ e $\nabla^2 = \sum \nabla_i^2$.

O lema de Stein será utilizado para obter uma estimativa não viesada para o risco do estimador da média. O estimador usual para a média é a própria observação $\mathbf{X}$, se utilizamos uma pertubação desse estimador, dada por $\mathbf{X} + \mathbf{h}(\mathbf{X})$, como um estimador para a média de $N_p(\boldsymbol{\theta}, \mathbf{I})$ tem-se

Teorema 2.9 *Considere o estimador $\mathbf{X} + \mathbf{h}(\mathbf{X})$ para a média $\boldsymbol{\theta}$ tal que $h : \mathbb{R}^p \to \mathbb{R}^p$ é uma função diferenciável tal que*

$$E_{\boldsymbol{\theta}}\left[\sum |\nabla_i h_i(\mathbf{X})|\right] < \infty.$$

Então, para cada $i \in \{1, ..., p\}$,

$$E_\theta\left[(X_i + h_i(\mathbf{X}) - \theta_\mathbf{i})^2\right] = 1 + E_\theta\left[h_i^2(\mathbf{X}) + 2\nabla_i h_i(\mathbf{X})\right]$$

e consequentemente,

$$E_{\boldsymbol{\theta}}\left[\|\mathbf{X} + h(\mathbf{X}) - \boldsymbol{\theta}\|^2\right] = p + E_{\boldsymbol{\theta}}\left[\|h(\mathbf{X})\|^2 + 2\nabla \cdot h(\mathbf{X})\right]. \tag{2.19}$$

Demonstração:

Utilizando o Lema 2 com $f = h_i$, tem-se

$$\begin{aligned} E_\theta\left[(X_i + h_i(\mathbf{X}) - \theta_\mathbf{i})^2\right] &= E_\theta\left[((X_i - \theta_i) + h_i(\mathbf{X}))^2\right] \\ &= E_\theta\left[(X_i - \theta_i)^2 + 2(X_i - \theta_i)h_i(X) + h_i^2(\mathbf{X})\right] \\ &= E_\theta\left[(X_i - \theta_i)^2\right] + 2E_\theta\left[(X_i - \theta_i)h_i(\mathbf{X})\right] + E_\theta\left[h_i^2(\mathbf{X})\right] \\ &= 1 + 2E_\theta\left[\nabla_i h_i(\mathbf{X})\right] + E_\theta\left[h_i^2(\mathbf{X})\right] \\ &= 1 + E_\theta\left[\left(h_i^2(\mathbf{X}) + 2\nabla_i h_i(\mathbf{X})\right)\right]. \end{aligned}$$

Somando em i segue o resultado. ■

Utilizando o método dos momentos, o Teorema 2.9 permite obter um estimador não viesado para o risco do estimador $\mathbf{X} + h(\mathbf{X})$ dado por

$$\hat{\mathscr{R}}_{\mathbf{X}+h(\mathbf{X})}(\boldsymbol{\theta}) = p + \|h(\mathbf{X})\|^2 + 2\nabla h(\mathbf{X}).$$

Este resultado nos remete ao estimador denominado estimador não viesado de Stein para o risco, ou SURE (Stein's unbiased risk estimator) destacado por Efron (2004).

Para definir o SURE, tem-se que, se $h(\mathbf{X}) = f(\mathbf{X}) - \mathbf{X}$ e $f(\mathbf{X})$ é utilizada como estimador da média tem-se

$$\begin{aligned} E_\theta\left[\|\boldsymbol{\theta} - f(\mathbf{X})\|^2\right] &= E_\theta\left[\|\boldsymbol{\theta} - \mathbf{X} + \mathbf{X} - f(\mathbf{X})\|^2\right] \\ &= E_\theta\left[\|\boldsymbol{\theta} - \mathbf{X}\|\right] + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] + 2E_{\boldsymbol{\theta}}\left[(\boldsymbol{\theta} - \mathbf{X})'(\mathbf{X} - f(\mathbf{X}))\right] \\ &= p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] + 2E_\theta\left[(\boldsymbol{\theta} - \mathbf{X})'(\mathbf{X} - \boldsymbol{\theta} + \boldsymbol{\theta} - f(\mathbf{X}))\right] \\ &= p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] + 2E_\theta\left[(\boldsymbol{\theta} - \mathbf{X})'(\mathbf{X} - \boldsymbol{\theta})\right] \\ &\qquad + 2E_\theta\left[(\boldsymbol{\theta} - \mathbf{X})'(\boldsymbol{\theta} - f(\mathbf{X}))\right] \\ &= p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] - 2E_\theta\left[(\mathbf{X} - \boldsymbol{\theta})'(\mathbf{X} - \boldsymbol{\theta})\right] \\ &\qquad - 2E_\theta\left[(\boldsymbol{\theta} - \mathbf{X})' f(\mathbf{X})\right] \end{aligned}$$

$$
\begin{aligned}
&= p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] - 2p\sigma^2 - 2E_\theta\left[(\boldsymbol{\theta} - \mathbf{X})' f(\mathbf{X})\right] \\
&= p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] - 2p\sigma^2 - 2\sum_{i=1}^{p} E_\theta\left[(\boldsymbol{\theta}_i - \mathbf{X_i})\, f_i(\mathbf{X})\right] \\
&= -p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] + 2\sum_{i=1}^{p} \operatorname{cov}\left(\mathbf{X_i}, f_i(\mathbf{X})\right) \\
&= -p\sigma^2 + E_\theta\left[\|\mathbf{X} - f(\mathbf{X})\|^2\right] + 2\sigma^2 \sum_{i=1}^{p} E_\theta\left[\frac{\partial f_i}{\partial x_i}(\mathbf{X})\right] \text{(pelo lema de Stein)}
\end{aligned}
$$

Pelo método dos momentos, um estimador não viesado para o risco do estimador $f(\mathbf{X})$ é

$$\hat{\mathscr{R}}_{\mathbf{X}} + h(\mathbf{X})(\boldsymbol{\theta}) = -p\sigma^2 + \|\mathbf{X} - f(\mathbf{X})\|^2 + 2\sigma^2 \cdot \sum_{i=1}^{p} \frac{\partial f_i(\mathbf{X})}{\partial x_i},$$

denominado SURE.

Retomando o artigo de Stein (1981), tem-se um outro resultado como consequência do lema 2.1

Teorema 2.10 *Seja $f : \mathbb{R}^p \to \mathbb{R}^+$ uma função com gradiente $\nabla f : \mathbb{R}^p \to \mathbb{R}^p$ diferenciável tal que*

$$E_\theta\left[\frac{1}{f(\mathbf{X})} \sum_{i=1}^{p} \left|\nabla_i^2 f(\mathbf{X})\right|\right] < \infty$$

e

$$E_\theta\left[\|\nabla \log f(\mathbf{X})\|^2\right] < \infty.$$

Então,

$$
\begin{aligned}
E_\theta\left[\|\mathbf{X} + \nabla \log f(\mathbf{X}) - \boldsymbol{\theta}\|^2\right] &= p + E_\theta\left[2\frac{\nabla^2 f(\mathbf{X})}{f(\mathbf{X})} - \frac{\|\nabla f(\mathbf{X})\|^2}{f^2(\mathbf{X})}\right] \\
&= p + 4E_\theta\left[\frac{\nabla^2 \sqrt{f(\mathbf{X})}}{\sqrt{f(\mathbf{X})}}\right].
\end{aligned} \tag{2.20}
$$

Demonstração:

Seja $g : \mathbb{R}^p \to \mathbb{R}^p$ definida por

$$\begin{aligned} g\left(x_1, ..., x_p\right) &= \left(g_1\left(x_1, ..., x_p\right), ..., g_p\left(x_1, ..., x_p\right)\right) \\ &= \nabla \log f\left(x_1, ..., x_p\right) \\ &= \left(\frac{\partial}{\partial x_1} \log f\left(x_1, ..., x_p\right), ..., \frac{\partial}{\partial x_p} \log f\left(x_1, ..., x_p\right)\right) \\ &= \left(\frac{1}{f\left(x_1, ..., x_p\right)} \frac{\partial f}{\partial x_1}\left(x_1, ..., x_p\right), ..., \frac{1}{f\left(x_1, ..., x_p\right)} \frac{\partial f}{\partial x_p}\left(x_1, ..., x_p\right)\right) \\ &= \frac{1}{f\left(x_1, ..., x_p\right)} \nabla f\left(x_1, ..., x_p\right). \end{aligned}$$

Então,

$$\begin{aligned} \nabla \cdot g &= \sum_{i=1}^{p} \frac{\partial g_i}{\partial x_i}\left(x_1, ..., x_p\right) \\ &= \sum_{i=1}^{p} \frac{\partial}{\partial x_i}\left(\frac{1}{f\left(x_1, ..., x_p\right)} \frac{\partial f}{\partial x_i}\left(x_1, ..., x_p\right)\right) \\ &= \sum_{i=1}^{p}\left[\frac{\partial}{\partial x_i}\left(\frac{1}{f\left(x_1, ..., x_p\right)}\right) \frac{\partial f}{\partial x_i}\left(x_1, ..., x_p\right) + \frac{1}{f\left(x_1, ..., x_p\right)} \frac{\partial^2 f}{\partial x_i^2}\left(x_1, ..., x_p\right)\right] \\ &= \sum_{i=1}^{p} \frac{-\frac{\partial f}{\partial x_i}\left(x_1, ..., x_p\right) \frac{\partial f}{\partial x_i}\left(x_1, ..., x_p\right)}{\left(f\left(x_1, ..., x_p\right)\right)^2} + \frac{1}{f\left(x_1, ..., x_p\right)} \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i^2}\left(x_1, ..., x_p\right) \\ &= \frac{-\left\|\nabla f\left(x_1, ..., x_p\right)\right\|^2}{\left(f\left(x_1, ..., x_p\right)\right)^2} + \frac{1}{f\left(x_1, ..., x_p\right)} \nabla^2 f\left(x_1, ..., x_p\right) \end{aligned}$$

e, assim segue da Equação (2.19) que

$$\begin{aligned} E_\theta\left[\left\|\mathbf{X} + \nabla \log \mathbf{f}(\mathbf{X}) - \boldsymbol{\theta}\right\|^2\right] &= E_\theta\left[\left\|\mathbf{X} + \mathbf{g}(\mathbf{X}) - \boldsymbol{\theta}\right\|^2\right] \\ &= p + E_\theta\left[\left\|g(\mathbf{X})\right\|^2 + 2\nabla \cdot g(\mathbf{X})\right] \\ &= p + E_\theta\left[\frac{\left\|\nabla f(\mathbf{X})\right\|^2}{f^2(\mathbf{X})} + 2\left(\frac{\nabla^2 f(\mathbf{X})}{f(\mathbf{X})} - \frac{\left\|\nabla f(\mathbf{X})\right\|^2}{f^2(\mathbf{X})}\right)\right] \\ &= p + E_\theta\left[2\frac{\nabla^2 f(\mathbf{X})}{f(\mathbf{X})} - \frac{\left\|\nabla f(\mathbf{X})\right\|^2}{f^2(\mathbf{X})}\right], \end{aligned}$$

que é a primeira forma da Equação (2.20). Observe que

$$\begin{aligned}\nabla^2\sqrt{f} = \nabla\cdot\nabla\sqrt{f} &= \nabla\cdot\frac{\nabla f}{2\sqrt{f}} \\ &= \frac{\nabla^2 f.2\sqrt{f}}{\left(2\sqrt{f}\right)^2} - \frac{\nabla f.f^{-1/2}}{\left(2\sqrt{f}\right)^2} \\ &= \frac{1}{2\sqrt{f}}\nabla^2 f - \frac{1}{4f^{3/2}}\|\nabla f\|^2\end{aligned}$$

logo,

$$4\frac{\nabla^2\sqrt{f}}{\sqrt{f}} = 2\frac{1}{f}\nabla^2 f - \frac{1}{f^2}\|\nabla f\|^2.$$

Portanto,

$$E_\theta\left[\|\mathbf{X}+\nabla\log\mathbf{f}(\mathbf{X})-\boldsymbol{\theta}\|^2\right] = p + 4E_\theta\left[\frac{\nabla^2\sqrt{f}(\mathbf{X})}{\sqrt{f}(\mathbf{X})}\right].$$

■

O seguinte corolário permite obter uma condição suficiente para que $\mathbf{X}+\nabla log f(\mathbf{X})$ seja um estimador minimax de $\boldsymbol{\theta}$.

Corolário 1: Se $f : \mathbb{R}^p \to \mathbb{R}^+$ é duas vezes diferenciável e sua raiz quadrada é superharmônica e as condições do teorema 2.10 são satisfeitas, então $\mathbf{X}+\nabla log f(\mathbf{X})$ é um estimador minimax de $\boldsymbol{\theta}$ com risco satisfazendo

$$\begin{aligned}E_\theta\left[\|\mathbf{X}+\nabla\log f(\mathbf{X})-\boldsymbol{\theta}\|^2\right] &= p + 4E_\theta\left[\frac{\nabla^2\sqrt{f(\mathbf{X})}}{\sqrt{f(\mathbf{X})}}\right] \\ &\leq p = \inf_g\sup_\theta E_\theta\Big[\|\mathbf{X}+g(\mathbf{X})-\boldsymbol{\theta}\|^2\Big].\end{aligned}$$

Demonstração: Stein (1981).

2.5.1 Justificativa heurística para o coeficiente de encolhimento

Os resultados apresentados, nesta seção e na seção subsequente, encontram-se em (GRUBER, 1998). Entretanto, os detalhes das explicações são parte inerente deste trabalho e, de certa forma, uma contribuição para o estudo do tema.

Suponha $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\theta}, \mathbf{I})$ e que apenas um vetor $\mathbf{X}= (X_1, ..., X_p)$ desta população é observado. O estimador de máxima verossimilhança de $\boldsymbol{\theta}$ é o próprio vetor observado $\mathbf{X}$ e X_i é um estimador de θ_i. No

entanto, ao contrário da intuição, $\|\mathbf{X}\|^2 = \sum X_i^2$ não é um bom estimador de $\|\boldsymbol{\theta}\|^2$ pois,

$$E\left[\|\mathbf{X}\|^2\right] = \|\boldsymbol{\theta}\|^2 + p,$$

isto é, $\|\mathbf{X}\|^2 = X_1^2 + ... + X_p^2$ tem tendência a superestimar $\|\boldsymbol{\theta}\|^2 = \theta_1^2 + ... + \theta_p^2$, como visto na seção 2.3.

Como

$$\|\boldsymbol{\theta}\|^2 = E\left[\|\mathbf{X}\|^2\right] - p \tag{2.21}$$

pelo método dos momentos, tem-se que um estimador de $\|\boldsymbol{\theta}\|^2$ é $\|\mathbf{X}\|^2 - p$.

Dividindo ambos os lados da equação (2.21) por p,

$$\frac{\sum_{i=1}^{n} \theta_i^2}{p} = \frac{\sum_{i=1}^{n} X_i^2}{p} - 1$$

tem-se que

$$\frac{1}{p}\sum_{i=1}^{p} X_i^2 \qquad \text{é um estimador de} \qquad 1 + \frac{\sum_{i=1}^{n} \theta_i^2}{p}.$$

Este estimador de fato possui boas propriedades assintóticas como estabelecido no Teorema 2.11.

Teorema 2.11 *Se X_i, $1 \leq i \leq p$ são variáveis aleatórias independentes e $X_i \sim N(\theta_i, 1)$ então, quando $p \to \infty$, $\sum_{i=1}^{p} x_i^2/p$ converge em probabilidade para $1 + \sum_{i=1}^{p} \theta_i^2/p$.*

Um argumento heurístico para a escolha do fator de encolhimento é que entre os estimadores de forma $\hat{\boldsymbol{\theta}} = c\mathbf{X}$, o de menor erro quadrático médio é dado por

$$\hat{\boldsymbol{\theta}} = \left(\frac{\|\boldsymbol{\theta}\|^2}{p + \|\boldsymbol{\theta}\|^2}\right)\mathbf{X},$$

como $\|\boldsymbol{\theta}\|^2 \approx \sum_{i=1}^{p} X_i^2 - p$ tem-se

$$\frac{\|\boldsymbol{\theta}\|^2}{p + \|\boldsymbol{\theta}\|^2} \approx \frac{\sum_{i=1}^{p} X_i^2 - p}{\sum_{i=1}^{p} X_i^2} = 1 - \frac{p}{\sum_{i=1}^{p} X_i^2}.$$

Assim, com este argumento heurístico o estimador de $\boldsymbol{\theta}$ é

$$\hat{\boldsymbol{\theta}} = \left(1 - \frac{p}{\|X\|^2}\right)\mathbf{X}.$$

Esse argumento simples mostra uma versão preliminar para o estimador de Jams-Stein, utilizando p e não $p-2$ no coeficiente de encolhimento.

2.5.2 O estimador de James-Stein como um estimador Bayesiano empírico

Seja uma amostra aleatória $\mathbf{X}= (X_1, \ldots, X_p)$ com distribuição, $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$, e considere uma *priori*, $\theta_i \sim N(0, \sigma^2)$ com variância σ^2 desconhecida, em que foi observado um único valor de $\mathbf{x} = (\mathbf{x_1}, ..., \mathbf{x_p})$, o estimador de Bayes é dado por

$$\hat{\theta}_i = \frac{\sigma^2}{\sigma^2+1} x_i = \left(1 - \frac{1}{\sigma^2+1}\right) x_i$$

o cálculo padrão é descrito no [Apêndice A]. Esse estimador depende do parâmetro desconhecido σ^2, utilizando a abordagem Bayesiana empírica ele pode ser estimado a partir dos dados. A preditiva é dada por

$$f(\mathbf{X}) \sim \mathbf{N}(\mathbf{0}, (\sigma^{\mathbf{2}} + \mathbf{1})\mathbf{I}).$$

Considere Z_i a variável aleatória obtida considerando que $\mathbf{X}$ tem distribuição dada pela preditiva, logo,

$$Z_i = \frac{X_i}{\sqrt{\sigma^2+1}} \Rightarrow Z_i \sim N(0,1).$$

Elevando ambos os membros da igualdade ao quadrado e somando em i temos

$$\sum_{i=1}^{p} Z_i^2 = \frac{\sum\limits_{i=1}^{p} X_i^2}{\left(\sqrt{\sigma^2+1}\right)^2} = \frac{\mathbf{X'X}}{\sigma^2+1}.$$

Como a soma dos quadrados de p normais padrão independentes tem distribuição qui-quadrado com parâmetro p segue que $\mathbf{Y} = \dfrac{\mathbf{X'X}}{\sigma^{\mathbf{2}}+\mathbf{1}}$ tem distribuição $\mathbf{Y} \sim \chi^{\mathbf{2}}(\mathbf{p})$.

Segue a seguinte relação:

$$\mathbf{Y} = \frac{\mathbf{X'X}}{\sigma^{\mathbf{2}}+\mathbf{1}} \;\Leftrightarrow\; \frac{\mathbf{1}}{\mathbf{X'X}} = \frac{\mathbf{1}}{\sigma^{\mathbf{2}}+\mathbf{1}} \frac{\mathbf{1}}{\mathbf{Y}}. \tag{2.22}$$

Aplicando a esperança em ambos os lados da equação

$$E\left[\frac{1}{\mathbf{X'X}}\right] = E\left[\frac{1}{(\sigma^2+1)\mathbf{Y}}\right] = \frac{1}{\sigma^2+1} E\left[\frac{1}{\mathbf{Y}}\right].$$

Para o cálculo de $E\left[\frac{1}{\mathbf{Y}}\right]$, tem-se a relação que a inversa de uma qui-quadrado é um distribuição gama. Portanto, considere que $\frac{1}{\mathbf{Y}} \sim Gama(\alpha = \frac{p}{2}, \beta = \frac{1}{2})$

$$\begin{aligned}
E\left[\frac{1}{\mathbf{Y}}\right] &= \int_0^\infty \frac{1}{y} \underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)} y^{\left(\frac{p}{2}-1\right)} e^{-\frac{1}{2}y} dy}_{f_Y(y)} \\
&= \frac{\Gamma\left(\frac{p}{2}-1\right)}{\Gamma\left(\frac{p}{2}\right)\left(\frac{1}{2}\right)^{-1}} \int_0^\infty \underbrace{\frac{\left(\frac{1}{2}\right)^{\frac{p}{2}-1}}{\Gamma\left(\frac{p}{2}-1\right)} y^{\left(\frac{p}{2}-1\right)-1} e^{-\frac{1}{2}y} dy}_{Gama(\alpha=\frac{p}{2}-1, \beta=\frac{1}{2})} \\
&= \frac{\Gamma\left(\frac{p}{2}-1\right)}{\Gamma\left(\frac{p}{2}\right)\left(\frac{1}{2}\right)^{-1}} \\
&= \frac{\Gamma\left(\frac{p}{2}-1\right)}{2\left(\frac{p}{2}-1\right)\Gamma\left(\frac{p}{2}-1\right)} \\
&= \frac{1}{2\left(\frac{p}{2}-1\right)} = \frac{1}{2\left(\frac{p-2}{2}\right)} = \frac{1}{p-2}
\end{aligned}$$

logo,

$$E\left[\frac{1}{\mathbf{X}'\mathbf{X}}\right] = \frac{1}{\sigma^2+1}\frac{1}{p-2}$$

ou ainda,

$$E\left[\frac{p-2}{\mathbf{X}'\mathbf{X}}\right] = \frac{1}{\sigma^2+1}.$$

Pelo método dos momentos, podemos estimar $\frac{1}{\sigma^2+1}$ por $\frac{p-2}{\mathbf{X}'\mathbf{X}}$. Como

$$\hat{\theta}_i = \frac{\sigma^2}{\sigma^2+1} x_i = \left(1 - \frac{1}{\sigma^2+1}\right) x_i$$

substituindo $\frac{1}{\sigma^2+1}$ por $\frac{p-2}{\mathbf{X}'\mathbf{X}}$, o resultado é o estimador de James-Stein

$$\left(\hat{\theta}_{JS}\right)_i = \left(1 - \frac{p-2}{\mathbf{X}'\mathbf{X}}\right) X_i$$

que na forma vetorial será

$$\hat{\boldsymbol{\theta}}_{JS} = \left(1 - \frac{p-2}{\mathbf{X}'\mathbf{X}}\right) \mathbf{X}.$$

Deve-se resaltar que o estimador de James-Stein só é um estimador de encolhimento para $p > 2$.

2.5.3 O estimador de James-Stein como estimador Bayesiano empírico para o caso de parâmetros de locação

A leitura dos artigos de Brandwein and Strawderman (1990, 2012) possui um conceito intuitivo e apresenta o estimador de James-Stein de forma razoável e convincente através do desenvolvimento simples para estimação linear ótima do vetor de médias em $\mathbb{R}^p$ que leva ao estimador de Stein.

Abandonando a hipótese de normalidade e supondo apenas que $\mathbf{X}$ é um vetor aleatório p-dimensional com $E_\theta[\mathbf{X}] = \boldsymbol{\theta}$ e $\text{cov}[\mathbf{X}] = \sigma^2\mathbf{I}$ com σ^2 conhecido e densidade dada por $f(\mathbf{x} - \boldsymbol{\theta})$ com f desconhecido, isto é, $\boldsymbol{\theta}$ é um parâmetro de locação de uma densidade desconhecida. Considere como *priori* para $\boldsymbol{\theta}$ a convolução da densidade f consigo mesma n vezes, isto é, $f^{*n}(\theta)$. Dessa forma, $\boldsymbol{\theta}$ como variável aleatória pode ser expressa como soma de n variáveis U_i, $i = 1, ..., n$ independentes e com distribuição $f(\mathbf{u})$. Se $U_0 = (\mathbf{X} - \boldsymbol{\theta})$ então sua distribuição também é $f(\mathbf{u})$.

O estimador de Bayes $\boldsymbol{\delta}(\mathbf{X})$ é dado por

$$
\begin{aligned}
\boldsymbol{\delta}(\mathbf{X}) =& E[\boldsymbol{\theta}|\mathbf{X}] \\
=& E[\boldsymbol{\theta}|\mathbf{X} - \boldsymbol{\theta} + \boldsymbol{\theta}] \\
=& E[\boldsymbol{\theta}|U_0 + \theta] \\
=& E\left[\boldsymbol{\theta}|U_0 + \sum_{i=1}^{n} U_i\right] \\
=& E\left[\boldsymbol{\theta}|\sum_{i=0}^{n} U_i\right] \\
=& E\left[\sum_{i=1}^{n} U_i \middle| \sum_{i=0}^{n} U_i\right]
\end{aligned}
$$

como as variáveis U_i, $i = 1, ..., n$ são independentes tem-se que $\sum_{i=1}^{n} U_i = nu_j$, assim

$$\begin{aligned}
\boldsymbol{\delta}(\mathbf{X}) =& nE\left[u_j|\sum_{i=0}^{n} U_i\right] \\
=& \frac{n}{n+1}E\left[(n+1)u_j|\sum_{i=0}^{n} U_i\right] \\
=& \frac{n}{n+1}E\left[\sum_{i=0}^{n} U_i\middle|\sum_{i=0}^{n} U_i\right] \\
=& \frac{n}{n+1}E\left[\mathbf{X}|\,\mathbf{X}\right] \\
=& \frac{n}{n+1}\mathbf{X}
\end{aligned}$$

ou equivalentemente,

$$\boldsymbol{\delta}(\mathbf{X}) = \mathbf{E}\left[\theta|\mathbf{X}\right] = \left(\mathbf{1} - \frac{\mathbf{1}}{\mathbf{n+1}}\right)\mathbf{X}. \tag{2.23}$$

Assumindo que n é desconhecido, podemos estimar a distribuição preditiva de $\mathbf{X}$, que tem a mesma distribuição que $\mathbf{X} - \boldsymbol{\theta} + \boldsymbol{\theta} = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{U_i}$. Em particular,

$$\begin{aligned}
E_\theta\left[\|\mathbf{X}\|^2\right] &= E\left[\left\|\sum_{i=0}^{n} U_i\right\|^2\right] \\
&= \sum_{i=0}^{n} E\left[\|U_i\|^2\right] \\
&= (n+1)p\sigma^2
\end{aligned}$$

pois $E\left[U_i\right] = 0$ e $\mathrm{cov}[U_i] = \sigma^2\mathbf{I}$, $E\left[\|U_i\|^2\right] = p\sigma^2$. Entretanto, n é um parâmetro desconhecido da função densidade d*a priori* de $\boldsymbol{\theta}$ e $n+1$ pode ser estimado pelo método dos momentos por

$$E\left[\frac{\|\mathbf{X}\|^2}{p\sigma^2}\right] = n+1.$$

Substituindo esse estimador por $(n+1)$ na Equação (2.23), temos um estimador empírico de Bayes

$$\boldsymbol{\delta}(\mathbf{X}) = \left(1 - \frac{p\sigma^2}{\|\mathbf{X}\|^2}\right)\mathbf{X}$$

que é novamente o estimador de James-Stein, apenas substituindo $p-2$ por p.

2.5.4 A estimação de James-Stein como um problema de regressão

Stephen M. Stigler no excelente artigo "The 1988 Neyman Memorial Lecture: A Galtonian Perspective on Shrinkage Estimator" de 1990, utilizando ideias que remetem a Galton, obtém uma nova e interessante perspectiva para o estimador de James-Stein a partir de um problema de regressão. Como o artigo é extremamente didático e compreensível, o desenvolvimento desta seção será feita de forma simplificada, porém de forma mais detalhada e com uma contribuição geométrica.

Observando um valor de $\mathbf{X}$=$(X_1, ...X_p)$ de uma normal $N(\boldsymbol{\theta}, \mathbf{I})$, tal que $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, pode-se afirmar que todo o problema de estimação está nos p pares (X_i, θ_i). Como os θ_i são parâmetros desconhecidos não podemos plotar tais pontos, mas para facilitar o raciocínio, suponha que tais pontos estejam plotados da forma

$$X_i = \theta_i + \epsilon \qquad \text{com} \qquad \epsilon \sim N(0, 1)$$

como exposto na Figura (2) que é hipotética, mas reflete com precisão a situação.

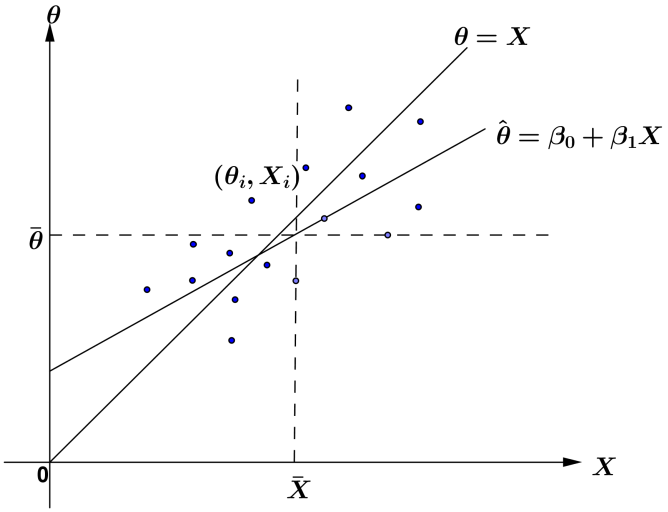

FIGURA 2: Gráfico bivariado hipotético para a regressão de pontos da forma $X_i = \theta_i + \epsilon$.

Os pontos (θ_i, X_i) não devem se afastar muito da reta $\theta = X$, uma vez que X_i é um bom estimador não viesado de θ_i. Note também que $\bar{X} = \frac{1}{p}\sum X_i$ é tal que $E\left[\bar{X}\right] = \frac{1}{p}\sum \theta_i = \bar{\theta}$ e $var(\bar{X}) = \frac{1}{p}$ e, portanto, o ponto $\left(\bar{X}, \bar{\theta}\right)$ também deve estar próximo da reta $\theta = X$ (Figura 2).

Do ponto de vista Bayesiano, ou ainda, a aproximação pelo método Bayesiano empírico, deve-se estimar todos os $\theta_i's$ dados todos os $X_i's$, sem suposições sobre distribuição para os $\theta_i's$. Sendo assim, não conhecemos a distribuição condicional de θ dado X e

Temos agora um problema de regressão linear típico, obter a melhor reta para estimar os $\theta_i's$ por um

estimador da forma

$$\hat{\theta}_i = \beta_0 + \beta_1 X_i$$

que minimiza a soma de quadrados, isto é, a função perda quadrática

$$\sum_{i=1}^{p} L\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) = \sum_{i=1}^{p} \left(\theta_i - \hat{\theta}_i\right)^2 = \sum_{i=1}^{p} \left(\theta_i - (\beta_0 + \beta_1 X_i)\right)^2.$$

Sabe-se que esse procedimento tem propriedades estatísticas ótimas no sentido de ser o melhor estimador linear. Obtém-se, então, a regressão de $\boldsymbol{\theta}$ em $\mathbf{X}$ (RENCHER;SCHAALJE, 2008).

$$\hat{\theta}_i = \bar{\boldsymbol{\theta}} + \hat{\boldsymbol{\beta}}(X_i - \bar{\mathbf{X}}) \tag{2.24}$$

$$\hat{\boldsymbol{\beta}} = \frac{\sum \left(X_i - \bar{\mathbf{X}}\right)\left(\theta_i - \bar{\boldsymbol{\theta}}\right)}{\sum \left(X_i - \bar{\mathbf{X}}\right)^2} \tag{2.25}$$

O problema aqui é que os θ_i não são conhecidos. A ideia então é substituir $\bar{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\beta}}$ por estimativas, obtendo então uma estimativa para a reta de regressão de θ em X. O estimador natural de $\bar{\boldsymbol{\theta}}$ é $\bar{\mathbf{X}}$, que possui propriedades ótimas, pois é um UMVUE. Para estimar $\hat{\boldsymbol{\beta}}$ vamos considerar a covariância amostral entre $\mathbf{X}$ e $\boldsymbol{\theta}$

$$\frac{1}{p-1} \sum \left(X_i - \bar{\mathbf{X}}\right)\left(\theta_i - \bar{\boldsymbol{\theta}}\right). \tag{2.26}$$

Os θ_i são constantes desconhecidas, mas supondo que sejam variáveis aleatórias i.i.d., neste caso (2.26) é um estimador não viesado de $\mathrm{cov}[\mathbf{X}, \boldsymbol{\theta}]$. Como $X_i = \theta_i + \epsilon$ com $\epsilon \sim N(0,1)$ e independente de θ_i tem-se

$$var[X_i] = var[\theta_i] + var[\epsilon] = var\left[\theta_i\right] + 1$$

e

$$\begin{aligned}
cov(X_i, \theta_i) &= E\left[\left(\theta_i + \varepsilon\right)\theta_i\right] - E\left[\theta_i + \varepsilon\right] E\left[\theta_i\right] \\
&= E\left[\theta_i^2 + \theta_i \varepsilon\right] - \left(E\left[\theta_i\right]\right)^2 \\
&= E\left[\theta_i^2\right] + E\left[\theta_i \varepsilon\right] - \left(E\left[\theta_i\right]\right)^2 \\
&= E\left[\theta_i^2\right] + E\left[\theta_i\right] E\left[\varepsilon\right] - \left(E\left[\theta_i\right]\right)^2 \\
&= E\left[\theta_i^2\right] - \left(E\left[\theta\right]\right)^2 \\
&= var\left[\theta_i\right] \\
&= var\left[X_i\right] - 1.
\end{aligned}$$

Um estimador não viesado de $var[X_i|\theta_i]$ é

$$\frac{1}{p-1}\sum_{i=1}^{p}\left(X_i-\bar{\mathbf{X}}\right)^2$$

e, portanto, um estimador não viesado de $\mathrm{cov}[\mathbf{X},\boldsymbol{\theta}]$ é

$$\frac{1}{p-1}\sum_{i=1}^{p}\left(X_i-\bar{\mathbf{X}}\right)^2-1.$$

Qualquer que seja a distribuição de θ_i, tem-se

$$\begin{aligned} E\left[\sum\left(X_i-\bar{\mathbf{X}}\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)\right] &= \sum E\left[\left(X_i-\bar{\mathbf{X}}\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)\right] \\ &= (p-1)cov\left[\mathbf{X},\boldsymbol{\theta}\right] \end{aligned}$$

e ainda

$$\begin{aligned} E\left[\sum\left(X_i-\bar{\mathbf{X}}\right)^2-(p-1)\right] &= (p-1)E\left[\frac{1}{p-1}\sum\left(X_i-\bar{\mathbf{X}}\right)^2\right]-(p-1) \\ &= (p-1)\left(var[X_i]\right)-(p-1) \\ &= (p-1)\left(cov[\mathbf{X},\boldsymbol{\theta}]+\mathbf{1}\right)-(p-1) \\ &= (p-1)cov[\mathbf{X},\boldsymbol{\theta}] \end{aligned}$$

portanto, segue que $\sum\left(X_i-\bar{\mathbf{X}}\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)$ e $\sum\left(X_i-\bar{\mathbf{X}}\right)^2-(p-1)$ possuem a mesma esperança $\mathrm{cov}[\mathbf{X},\boldsymbol{\theta}]$. O mesmo ocorre se voltarmos à perspectiva que os θ_i são constantes pois

$$\begin{aligned} E\left[\sum\left(X_i-\bar{\mathbf{X}}\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)\right] &= \sum E\left[\left(X_i-\bar{\mathbf{X}}\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)\right] \\ &= \sum\left[\left(E\left[X_i\right]-E\left[\bar{\mathbf{X}}\right]\right)\left(\theta_i-\bar{\boldsymbol{\theta}}\right)\right] \\ &= \sum\left(\theta_i-\bar{\boldsymbol{\theta}}\right)^2 \end{aligned}$$

substituindo esta igualdade na equação (2.25)

$$\hat{\beta}=\frac{\sum\left(\theta_i-\bar{\theta}\right)^2}{\sum\left(X_i-\bar{\mathbf{X}}\right)^2}.$$

Para estimar $\sum_{i=1}^{p}\left(\theta_i-\bar{\boldsymbol{\theta}}\right)^2$, será utilizada uma demonstração geométrica. Considere o vetor de mé-

dias $\boldsymbol{\theta}$ e o vetor de observações $\mathbf{X}$. Fazendo a projeção desses vetores na reta $\theta = X$ obtemos os vetores $\bar{\boldsymbol{\theta}}$ e $\bar{\mathbf{X}}$, (Figura 3).

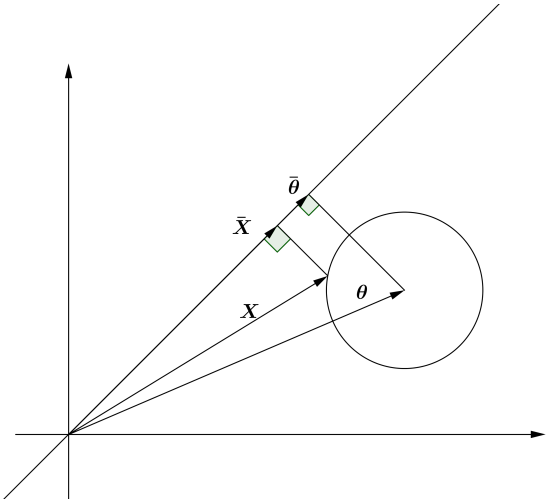

FIGURA 3: Projeção dos vetores θ e X na reta $\theta = X$.

Por Pitágoras,

$$\|\mathbf{X}\|^2 = \left\|\bar{\mathbf{X}}\right\|^2 + \left\|\mathbf{X} - \bar{\mathbf{X}}\right\|^2$$

e, portanto,

$$\begin{aligned}
E\left[\|\mathbf{X}\|^2\right] &= E\left[\left\|\overline{\mathbf{X}}\right\|^2\right] + E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right] \\
\sum_{i=1}^{p} E\left[(X_i)^2\right] &= E\left[p\left(\overline{\mathbf{X}}\right)^2\right] + E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right] \\
\sum_{i=1}^{p}\left((\theta_i)^2 + 1\right) &= p\left[var\left[\overline{\mathbf{X}}\right] + \left(E\left[\overline{\mathbf{X}}\right]\right)^2\right] + E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right] \\
\|\theta\|^2 + p &= p\left[\frac{1}{p} + \left(\overline{\boldsymbol{\theta}}\right)^2\right] + E\left[\left\|\boldsymbol{X} - \overline{\mathbf{X}}\right\|^2\right] \\
\|\boldsymbol{\theta}\|^2 - p\left(\overline{\boldsymbol{\theta}}\right)^2 + p - 1 &= E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right] \\
\|\boldsymbol{\theta}\|^2 - \left\|\overline{\boldsymbol{\theta}}\right\|^2 + p - 1 &= E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right].
\end{aligned}$$

Aplicando Pitágoras novamente, tem-se $\|\boldsymbol{\theta}\|^2 = \left\|\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\right\|^2 + \left\|\overline{\boldsymbol{\theta}}\right\|^2$ substituindo na equação acima temos

$$\left\|\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\right\|^2 + p - 1 = E\left[\left\|\mathbf{X} - \overline{\mathbf{X}}\right\|^2\right].$$

Isto sugere estimar o parâmetro $\boldsymbol{\beta}$ por

$$\frac{\sum \left(X_i - \bar{\mathbf{X}}\right)^2 - (p-1)}{\sum \left(X_i - \bar{\mathbf{X}}\right)^2} = 1 - \frac{p-1}{\sum \left(X_i - \bar{\mathbf{X}}\right)^2}$$

e, portanto, a reta de quadrados mínimos estimada dado por (2.24) é

$$\hat{\theta}_i = \bar{X} + \left(1 - \frac{p-1}{\sum\limits_{i=1}^{p} \left(X_i - \bar{X}\right)^2}\right)\left(X_i - \bar{X}\right).$$

Este é justamente o estimador de Efron-Morris introduzido pelo artigo Efron e Morris (1975) com $c = p-1$. Esse valor não é a melhor escolha, mas possui risco uniformemente menor que o estimador se $p > 5$. O estimador de James-Stein pode ser deduzido de forma similar considerando uma reta de regressão pela origem $\bar{\theta}_i = bX_i$ obtendo-se

$$\hat{\theta}_i^{JS} = \left(1 - \frac{p}{\|\mathbf{X}\|^2}\right) X_i.$$

Em síntese, a abordagem de Stigler (1990) afirma que a ideia da admissibilidade de $\mathbf{T_0}(\mathbf{X}) = \mathbf{X}$ ocorre pelo fato deste estimador ser obtido pela regressão de X em θ, a "regressão não usual". O estimador de James-Stein é obtido como uma aproximação da "regressão usual" de θ em X. Observe que para $p = 1$ ou $p = 2$ ambas as regressões coincidem e tem-se aí o motivo de $\mathbf{T_0}(\mathbf{X})$ ser admissível para essas dimensões.

2.5.5 O estimador de James-Stein versus o estimador média amostral

O estimador usual, que é o estimador de máxima verossimilhança para a média, ou ainda, a média amostral da distribuição normal multivariada, é inadmissível para dimensão suficientemente grande e, segundo James-Stein (1961), 3 é a dimensão crítica. Será analisada a admissibilidade do estimador média amostral para os casos de dimensões $p = 1$ e $p = 2$.

1) Caso $p = 1$: A admissibilidade de $\overline{\mathbf{X}}$ será demonstrada por absurdo utilizando o método de limite de Bayes segundo Lehmann e Casella (1998, p. 325). O lema enunciado a seguir, será utilizado na demonstração de tal fato.

Lema 2.4 *Se $\hat{\boldsymbol{\theta}}^*$ é um estimador de Bayes de $\tau(\boldsymbol{\theta})$, para todo $\theta \in \Theta$, em relação a função perda $l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ e priori $g(\boldsymbol{\theta})$ e se*

$$r_{l,g}(\hat{\boldsymbol{\theta}}^*) = E\left[\hat{\boldsymbol{\theta}}^*(\mathbf{X}) - \tau\left(\boldsymbol{\Theta}\right)\right]^2$$

é o risco de Bayes, então

$$r_{l,g}(\hat{\boldsymbol{\theta}}^*) = \int \text{var}\left[\tau\left(\Theta\right)|\mathbf{x}\right] d\mathbf{x}$$

se em particular, a variância a posteriori de $\tau\left(\Theta\right)|\mathbf{x}$ *é independente de* $\mathbf{x}$*, então*

$$r_{l,g}(\hat{\boldsymbol{\theta}}^*) = \text{var}\left[\tau\left(\Theta\right)|\mathbf{x}\right].$$

Demonstração: (LEHMANN; CASELLA, 1998, p. 317)

Sem perda de generalidade, suponha que a variância populacional é unitária, isto é, $\overline{\mathbf{X}} \sim N\left(\theta, \frac{1}{n}\right)$ e que $\bar{\mathbf{X}}$ não é admissível. Como $\mathscr{R}\left(\bar{\mathbf{X}}, \boldsymbol{\theta}\right) = \frac{1}{n}$ existe $\hat{\boldsymbol{\theta}}^*$ tal que

$$\mathscr{R}\left(\hat{\boldsymbol{\theta}}^*, \boldsymbol{\theta}\right) \leq \frac{1}{n} \qquad \text{para todo } \boldsymbol{\theta} \text{ com desigualdade estrita para algum } \boldsymbol{\theta}'.$$

Pela continuidade da função risco existem $\theta_0 < \theta_1$ e $\epsilon > 0$ tal que

$$\mathscr{R}\left(\hat{\boldsymbol{\theta}}^*, \boldsymbol{\theta}\right) < \frac{1}{n} - \varepsilon \qquad \text{para todo } \theta_0 < \boldsymbol{\theta} < \theta_1. \tag{2.27}$$

Considere uma *priori* com a distribuição $\theta \sim N(0, \tau^2)$. Seja r_g^* o risco de Bayes de $\hat{\boldsymbol{\theta}}^*$ em relação a esta *priori*. O risco de Bayes do estimador de Bayes, em relação a esta *priori*, é dado pela variância da posteriori conforme o lema 2.4. Utilizando o cálculo padrão da posteriori descrito em [Apêndice A] para determinar sua variância tem-se

$$r_{l,g}(\hat{\boldsymbol{\theta}}) = \frac{\tau^2}{1 + n\tau^2}.$$

Por outro lado,

$$\begin{aligned} r_{l,g}(\hat{\boldsymbol{\theta}}^*) &= \int_{-\infty}^{\infty} \mathscr{R}(\hat{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{-\infty}^{\infty} \mathscr{R}(\hat{\boldsymbol{\theta}}^*, \boldsymbol{\theta}) \frac{1}{\sqrt{2\pi\tau}} e^{-\theta^2 \big/ 2\tau^2} d\boldsymbol{\theta} \end{aligned}$$

Logo, para definir a razão a seguir, utilizando a desigualdade dada em (2.27), tem-se

$$\begin{aligned}\frac{\frac{1}{n}-r_{l,g}(\hat{\boldsymbol{\theta}}^*)}{\frac{1}{n}-r_{l,g}(\widehat{\boldsymbol{\theta}})} &= \frac{\frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty}\left[\frac{1}{n}-\mathscr{R}\left(\hat{\boldsymbol{\theta}}^*,\boldsymbol{\theta}\right)\right]e^{-\theta^2/2\tau^2}d\theta}{\frac{1}{n}-\frac{\tau^2}{1+n\tau^2}} \\ &\geq \frac{\frac{1}{\sqrt{2\pi\tau}}\int_{-\infty}^{\infty}\varepsilon e^{-\boldsymbol{\theta}^2/2\tau^2}d\boldsymbol{\theta}}{\frac{1}{n}-\frac{\tau^2}{1+n\tau^2}} \\ &= \frac{n\left(1+n\tau^2\right)\varepsilon}{\sqrt{2\pi\tau}}\int_{\theta_0}^{\theta_1}e^{-\frac{\boldsymbol{\theta}^2}{2\tau^2}}d\boldsymbol{\theta}.\end{aligned}$$

Quando $\tau^2 \to \infty$ o integrando, $e^{-\frac{\theta^2}{2\tau^2}} \to 1$ e $\int_{\theta_0}^{\theta_1} e^{-\frac{\theta^2}{2\tau^2}}d\theta \to \theta_1 - \theta_0$, portanto

$$\text{quando } \tau^2 \to \infty, \qquad \frac{1/n - r^*_{l,g}}{1/n - r_{l,g}} \to \infty.$$

Então, existe um τ_0 tal que $r_{\tau_0^*} < r_{\tau_0}$, absurdo pois r_{τ_0} é o menor risco de Bayes possível quando $\tau = \tau_0$.

2) Caso $p = 2$: No sentido de completude do texto, será apresentado um resultado clássico, o método de Blyth, que tem interesse em si próprio para outros resultados relativos à admissibilidade de estimadores.

De acordo com Lehmann e Casella (1998, p. 379), para garantir que a função risco seja contínua, suposições devem ser feitas sobre a função perda e a função densidade para afirmar a continuidade do risco. O seguinte teorema é enunciado sem provas, e está baseado em um conjunto de premissas muitas vezes satisfeitas na prática.

Teorema 2.12 *Considere a estimação de $\boldsymbol{\theta}$ com função perda $l(\boldsymbol{\delta},\boldsymbol{\theta})$, em que $\mathbf{X} \sim \mathbf{f}(\mathbf{x}|\boldsymbol{\theta})$ tem razão de verossimilhança monótona e é contínua em $\boldsymbol{\theta}$ para cada x. Se a função perda $l(\boldsymbol{\delta},\boldsymbol{\theta})$ satisfaz:*

(a)$l(\boldsymbol{\delta},\boldsymbol{\theta})$ é contínua em $\boldsymbol{\theta}$ para cada $\boldsymbol{\delta}$,

(b)l é decrescente em $\boldsymbol{\delta}$ para $\boldsymbol{\delta} < \boldsymbol{\theta}$ e crescente em $\boldsymbol{\delta}$ para $\boldsymbol{\delta} > \boldsymbol{\theta}$,

(c) Existem funções a e b que são limitadas e todo subconjunto limitado do espaço de parâmetros, tal que para todo $\boldsymbol{\delta}$

$$l(\theta,\boldsymbol{\delta}) \leq a\left(\boldsymbol{\theta}',\boldsymbol{\theta}\right)l\left(\theta',\boldsymbol{\delta}\right) + b\left(\boldsymbol{\theta}',\boldsymbol{\theta}\right)$$

então, os estimadores com valores finitos, função risco contínua $\mathscr{R}(\boldsymbol{\delta},\boldsymbol{\theta}) = E[l(\boldsymbol{\delta},\boldsymbol{\theta})]$ formam uma classe completa.

Para considerar estimadores que possuem um valor finito para o risco contínuo, os estimadores devem satisfazer as condições do teorema 2.12. Restrição ao risco contínuo, permite-nos utilizar o método para provar a

admissibilidade. O seguinte teorema estende a admissibilidade de estimadores de Bayes para seqüências de estimadores de Bayes.

Teorema 2.13 *(Método de Blyth) Suponha o espaço paramétrico $\Theta \subset \mathbb{R}^p$ aberto, e estimadores com função risco contínua formam uma classe completa. Se $\mathscr{R}(\boldsymbol{\delta}, \boldsymbol{\theta})$ é contínuo para todo estimador $\boldsymbol{\delta}$, e seja $\{g_n\}$ uma sequência de prioris definidas em Θ e $\boldsymbol{\delta}_{g_n}$ seus respectivos estimadores de Bayes, tais que*

(a) $r_{l,g_n}(\boldsymbol{\delta}) < \infty$ para todo n,

(b) para algum conjunto aberto não vazio $\Theta_0 \subset \Theta$, existe uma constante $B > 0$ e N tal que

$$\int_{\Theta_0} g_n(\boldsymbol{\theta}) d\theta \geq B, \qquad \text{para todo } n \geq N$$

(c) $r_{l,g_n}(\boldsymbol{\delta}) - r_{l,g_n}(\boldsymbol{\delta}_{g_n}) \to 0$ quando $n \to \infty$.

Então, $\boldsymbol{\delta}$ é um estimador admissível.

Demonstração:

Suponha que $\boldsymbol{\delta}$ é inadimissível, então existe um $\boldsymbol{\delta}'$ tal que $\mathscr{R}(\boldsymbol{\delta}', \boldsymbol{\theta}) \leq \mathscr{R}(\boldsymbol{\delta}, \boldsymbol{\theta})$, com desigualdade estrita para algum θ. Pela continuidade da função risco, isto implica que existe um conjunto Θ_0 e $\epsilon > 0$ tal que $\mathscr{R}(\boldsymbol{\delta}, \boldsymbol{\theta}) - \mathscr{R}(\boldsymbol{\delta}', \boldsymbol{\theta}) > \epsilon$ para $\theta \subset \Theta_0$. Logo para todo $n \geq N$,

$$r_{l,g_n}(\boldsymbol{\delta}) - r_{l,g_n}(\boldsymbol{\delta}') > \epsilon \int_{\Theta_0} g_n(\theta) d\theta \geq B$$

e, portanto, o item (c) não se verifica. ■

O teorema 2.13 mostra que uma das condições suficientes para um estimador ser admissível é que o seu risco de Bayes é aproximável por uma sequência de riscos de Bayes de estimadores de Bayes. Seria conveniente se fosse possível substituir os riscos pelos próprios estimadores. Como este não é o caso, pode-se concluir desse fato que a média amostral da normal de três ou mais dimensões não é admissível embora seja o limite de estimadores de Bayes. Esta demonstração é um tanto mais avançada, sendo o método apenas descrito para dar ao leitor uma ideia destes desenvolvimentos e para servir como uma introdução à literatura. (LEHMANN; CASELLA, 1998, p. 382)

O método obtém para $p = 2$, que a média amostral da normal multivariada é admissível. Porém, o método falha para $p \geq 3$ como de fato deveria ocorrer, pois para $p \geq 3$ a média amostral, no caso normal, não é admissível para $p \geq 3$.

Recordando que a variável aleatória $\mathbf{X} = (X_1, ...X_p)$ de uma normal $N(\boldsymbol{\theta}, \mathbf{I})$, tal que $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$

e considerando a perda quadrática, enunciamos o teorema de James-Stein.

Teorema 2.14 *(James,Stein 1961) O estimador de James-Stein tem risco menor que o estimador média amostral* $\hat{\boldsymbol{\theta}}_0(\mathbf{X}) = \mathbf{X}$ *para* $p \geq 3$.

Demonstração:

O risco quadrático médio do estimador $\hat{\boldsymbol{\theta}}_0(\mathbf{X}) = \mathbf{X}$ é dado por

$$\mathscr{R}\left(\widehat{\theta}_0, \theta\right) = \mathrm{E}\left[\|\mathbf{X} - \boldsymbol{\theta}\|^2\right] = \mathrm{E}\left[\sum_{\mathrm{i}=1}^{\mathrm{p}} (\mathrm{X_i} - \theta_\mathrm{i})^2\right] = \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{E}\left[(\mathrm{X_i} - \theta_\mathrm{i})^2\right] = \sum_{\mathrm{i}=1}^{\mathrm{p}} 1 = \mathrm{p}.$$

O risco quadrático médio do estimador James-Stein é dado por

$$\begin{aligned}
\mathscr{R}\left(\hat{\boldsymbol{\theta}}_{JS}, \boldsymbol{\theta}\right) &= E\left[\left(\hat{\boldsymbol{\theta}}_{JS} - \boldsymbol{\theta}\right)' \left(\hat{\boldsymbol{\theta}}_{JS} - \boldsymbol{\theta}\right)\right] \\
&= E\left[\left(\left(1 - \frac{(p-2)}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X} - \boldsymbol{\theta}\right)' \left(\left(1 - \frac{(p-2)}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X} - \boldsymbol{\theta}\right)\right] \\
&= E\left[\left((\mathbf{X} - \boldsymbol{\theta}) - \frac{(\mathbf{p}-\mathbf{2})\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right)' \left((\mathbf{X} - \boldsymbol{\theta}) - \frac{(\mathbf{p}-\mathbf{2})\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right)\right] \\
&= E\left[(\mathbf{X} - \boldsymbol{\theta})'(\mathbf{X} - \boldsymbol{\theta}) - \mathbf{2}(\mathbf{p}-\mathbf{2})\frac{(\mathbf{X} - \boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}} + \frac{(\mathbf{p}-\mathbf{2})^\mathbf{2}\mathbf{X}'\mathbf{X}}{(\mathbf{X}'\mathbf{X})^\mathbf{2}}\right] \\
&= E\left[(\mathbf{X} - \boldsymbol{\theta})'(\mathbf{X} - \boldsymbol{\theta})\right] - 2(p-2)E\left[\frac{(\mathbf{X} - \boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right] + (p-2)^2 E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right] \\
&= p + (p-2)\left(-2E\left[\frac{(\mathbf{X} - \boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right] + (p-2)E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right]\right).
\end{aligned}$$

Demonstrando que $\mathscr{R}\left(\hat{\boldsymbol{\theta}}_{JS}, \boldsymbol{\theta}\right) < p$ que segue da seguinte igualdade

$$E\left[\frac{(\mathbf{X} - \boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right] = (p-2)E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right].$$

Verificando esta igualdade, tem-se

$$\begin{aligned}
E\left[\frac{(\mathbf{X} - \boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right] &= \sum_{i=1}^{p} E\left[\frac{(X_i - \theta_i)X_i}{\sum_{i=1}^{p} X_i^2}\right] \\
&= \sum\nolimits_{i=1}^{p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{(X_i - \theta_i)X_i e^{-\frac{1}{2}\sum_{i=1}^{p}(X_i - \theta_i)^2}}{\sum_{i=1}^{p} X_i^2} dX_1 \ldots dX_p.
\end{aligned}$$

Calculando a integral de forma iterativa

$$\int_{-\infty}^{\infty} \frac{(X_i-\theta_i)X_i e^{-\frac{1}{2}(X_i-\theta_i)^2}}{\sum_{i=1}^{p} X_i^2} dX_i$$

e integrando por partes tem-se

$$du = (X_i-\theta_i)e^{-\frac{1}{2}(X_i-\theta_i)^2}dX_i \qquad \Rightarrow \qquad u = -e^{-\frac{1}{2}(X_i-\theta_i)^2}$$

$$v = \frac{X_i}{\sum_{i=1}^{p} X_i^2} \qquad \Rightarrow \qquad dv = \frac{\sum_{i=1}^{p} X_i^2 - X_i \frac{d}{dX_i}(\sum_{i=1}^{p} X_i^2)}{(\sum_{i=1}^{p} X_i^2)^2} = \frac{\sum_{i=1}^{p} X_i^2 - 2X_i^2}{(\sum_{i=1}^{p} X_i^2)^2}$$

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{(X_i-\theta_i)X_i e^{-\frac{1}{2}(X_i-\theta_i)^2}}{\sum_{i=1}^{p} X_i^2} dX_i &= -\frac{X_i}{\sum_{i=1}^{p} X_i^2} e^{-\frac{1}{2}(X_i-\theta_i)^2}\Big|_{-\infty}^{\infty} \\
&\quad + \int_{-\infty}^{\infty} \left(\frac{\sum_{i=1}^{p} X_i^2 - 2X_i^2}{(\sum_{i=1}^{p} X_i^2)^2} \right) e^{-\frac{1}{2}(X_i-\theta_i)^2} dX_i \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sum_{i=1}^{p} X_i^2} - \frac{2X_i^2}{(\sum_{i=1}^{p} X_i^2)^2} \right) e^{-\frac{1}{2}(X_i-\theta_i)^2} dX_i
\end{aligned}$$

logo

$$\begin{aligned}
E\left[\frac{(\mathbf{X}-\boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right] &= \sum_{i=1}^{p} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\frac{1}{\mathbf{X}'\mathbf{X}} - \frac{2x_i^2}{(\mathbf{X}'\mathbf{X})^2} \right) e^{-\frac{1}{2}(x_i-\boldsymbol{\theta}_i)^2} dx_1 ... dx_p \\
&= p \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\frac{1}{\mathbf{X}'\mathbf{X}} \right) e^{-\frac{1}{2}(x_i-\boldsymbol{\theta}_i)^2} dx_1 ... dx_p \\
&\quad - \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\frac{2\sum_{i=1}^{p} x_i^2}{(\mathbf{X}'\mathbf{X})^2} \right) e^{-\frac{1}{2}(x_i-\boldsymbol{\theta}_i)^2} dx_1 ... dx_p \\
&= p \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\frac{1}{\mathbf{X}'\mathbf{X}} \right) e^{-\frac{1}{2}(x_i-\boldsymbol{\theta}_i)^2} dx_1 ... dx_p \\
&\quad - \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \left(\frac{2}{\mathbf{X}'\mathbf{X}} \right) e^{-\frac{1}{2}(x_i-\boldsymbol{\theta}_i)^2} dx_1 ... dx_p \\
&= (p-2)E\left[\frac{1}{\mathbf{X}'\mathbf{X}}\right].
\end{aligned}$$

Assim,

$$\begin{aligned}
\mathscr{R}\left(\hat{\boldsymbol{\theta}}_{JS},\boldsymbol{\theta}\right) &= p+(p-2)\left(-2E\left[\frac{(\mathbf{X}-\boldsymbol{\theta})'\mathbf{X}}{\mathbf{X}'\mathbf{X}}\right]+(p-2)E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right]\right)\\
&= p+(p-2)\left(-2(p-2)E\left[\frac{1}{\mathbf{X}'\mathbf{X}}\right]+(p-2)E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right]\right)\\
&= p-(p-2)^2E\left[\frac{1}{(\mathbf{X}'\mathbf{X})}\right]\\
&< p.
\end{aligned}$$

■

No teorema seguinte, será demonstrado que o coeficiente de encolhimento utilizando $p-2$ é ótimo. Esse é um importante resultado da teoria do estimador de James-Stein abordado no artigo de Brandwein e Strawderman (1990). Considere estimadores da forma

$$\boldsymbol{\delta}_a(\mathbf{X})=\left(\mathbf{1}-\frac{\mathbf{a}}{\|\mathbf{X}\|^{\mathbf{2}}}\right)\mathbf{X}$$

observe que para $a=0$ tem-se o estimador usual $\boldsymbol{\delta}_0(\mathbf{X})=\mathbf{X}$.

Teorema 2.15 ***(a)*** *O estimador $\boldsymbol{\delta}_a(\mathbf{X})$ domina $\boldsymbol{\delta}_0(\mathbf{X})=\mathbf{X}$ para $0<a<2(p-2)$ e $p\geq 3$. O estimador $\boldsymbol{\delta}_{p-2}(\mathbf{X})=\left(\mathbf{1}-\frac{\mathbf{p-2}}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X}$ tem o risco uniformemente menor que qualquer outro estimador nesta classe.*
(b) *O risco de $\boldsymbol{\delta}_{p-2}(\mathbf{X})$ para $\boldsymbol{\theta}=0$ é igual a 2 para toda dimensão $p\geq 3$.*

Demonstração:

(a)

$$\begin{aligned}
\mathscr{R}\left(\boldsymbol{\delta}_a,\boldsymbol{\theta}\right) &= E\left[\|\boldsymbol{\delta}_a-\boldsymbol{\theta}\|^2\right]\\
&= E\left[\left\|\left(1-\frac{a}{\|\mathbf{X}\|^2}\right)\mathbf{X}-\boldsymbol{\theta}\right\|^2\right]\\
&= E\left\|(\mathbf{X}-\boldsymbol{\theta})-\frac{\mathbf{aX}}{\|\mathbf{X}\|^{\mathbf{2}}}\right\|^2\\
&= E\|(\mathbf{X}-\boldsymbol{\theta})\|^{`2}+a^2E\frac{1}{\|\mathbf{X}\|^2}-2aE\frac{\|\mathbf{X}(\mathbf{X}-\boldsymbol{\theta})\|}{\|\mathbf{X}\|^2}
\end{aligned}\tag{2.28}$$

$$
\begin{aligned}
&= p + a^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] - 2a\sum_{i=1}^{p} E\frac{X_i(X_i-\theta_i)}{\sum_{j=1}^{p} X_j^2} \\
&= p + a^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] - 2a\sum_{i=1}^{p} E\left[\frac{d}{dx_i}\left(\frac{X_i}{\sum_{j=1}^{p} X_j^2}\right)\right] \text{ pelo lema 2} \\
&= p + a^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] - 2a\sum_{i=1}^{p} E\left[\frac{\sum_{j=1}^{p} X_j^2 - 2X_i^2}{\left(\sum_{j=1}^{p} X_j^2\right)^2}\right] \\
&= p + a^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] - 2aE\left[\frac{p\|\mathbf{X}\|^2 - 2\|\mathbf{X}\|^2}{\left(\|\mathbf{X}\|^2\right)^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= p + a^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] - 2a(p-2)E\left[\frac{1}{\|\mathbf{X}\|^2}\right] \\
&= p + \left[a^2 - 2a(p-2)\right] E\left[\frac{1}{\|\mathbf{X}\|^2}\right].
\end{aligned}
$$

$$a^2 - 2a(p-2) < 0 \Leftrightarrow 0 < a < 2(p-2)$$

o polinômio $a^2 - 2a(p-2)$ assume o mínimo para $a = p-2$. Logo,

$$\mathscr{R}\left(\boldsymbol{\delta}_a, \boldsymbol{\theta}\right) < \mathscr{R}\left(\boldsymbol{\delta}_0, \boldsymbol{\theta}\right) = p, \qquad \text{para } 0 < a < 2(p-2).$$

(b) Para $\boldsymbol{\theta} = \mathbf{0}$, observando que $\mathbf{X} = (X_1, ...X_p)$ e $\mathbf{X} \sim \mathbf{N}\left(\boldsymbol{\theta}, \mathbf{I}\right)$ com $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, $\|\mathbf{X}\|^2$ tem uma distribuição qui-quadrado com p graus de liberdade, logo

$$
\begin{aligned}
\mathscr{R}\left(\boldsymbol{\delta}_{p-2}, \mathbf{0}\right) &= E\left[\left\|\boldsymbol{\delta}_{p-2} - \mathbf{0}\right\|^2\right] \\
&= E\left[\left\|\left(1 - \frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X}\right\|^2\right] \\
&= E\left[\|\mathbf{X}\|^2\right] - E\left[2\frac{p-2}{\|\mathbf{X}\|^2}\|\mathbf{X}\|^2\right] - (p-2)^2 E\left[\frac{1}{\|\mathbf{X}\|^2}\right] \\
&= p - \frac{(p-2)^2}{p-2} \\
&= 2
\end{aligned}
$$

■

2.5.6 O estimador de James-Stein para o caso de variância conhecida

Como proposto por Lemehmann e Casella (1998, p. 368), o estimador de James-Stein deve ser estudado considerando $\mathbf{X} \sim N\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$ com σ^2 conhecido. Ainda nesta situação, o estimador de James-Stein domina o estimador média amostral.

Se $\mathbf{X} = (X_1, ..., X_p) \sim N\left(\boldsymbol{\theta}, \mathbf{I}\right)$ então, o estimador $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{X}$, isto é, uma única observação, não é um estimador admissível para a média $\boldsymbol{\theta}$.

Suponha agora uma população $N\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$, σ^2 conhecido.

$$
\mathbf{X} \sim N\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right) \qquad \mathbf{Y} = \frac{\mathbf{X}}{\sigma} \sim N\left(\frac{\boldsymbol{\theta}}{\sigma}, \mathbf{I}\right).
$$

O estimador $\boldsymbol{\delta}(\mathbf{Y}) = \mathbf{Y}$ não é admissível para $\frac{\boldsymbol{\theta}}{\sigma}$, logo existe $\boldsymbol{\delta}^*(\mathbf{Y})$ que domina $\boldsymbol{\delta}(\mathbf{Y}) = \mathbf{Y}$, isto é,

$$
\mathscr{R}\left(\boldsymbol{\delta}^*, \boldsymbol{\theta}\right) \leq \mathscr{R}\left(\boldsymbol{\delta}, \boldsymbol{\theta}\right)
$$

$$
\int ... \int \left\|\boldsymbol{\delta}^*(y) - \frac{\boldsymbol{\theta}}{\sigma}\right\|^2 \phi_{\frac{\boldsymbol{\theta}}{\sigma}, \mathbf{I}}(y) d\mathbf{y} \leq \int ... \int \left\|y - \frac{\boldsymbol{\theta}}{\sigma}\right\|^2 \phi_{\frac{\boldsymbol{\theta}}{\sigma}, \mathbf{I}}(y) d\mathbf{y}.
$$

Fazendo a mudança de coordenadas $\mathbf{X} = \sigma\mathbf{Y}$

$$\int ... \int \left\| \boldsymbol{\delta}^* \left(\frac{x}{\sigma}\right) - \frac{\boldsymbol{\theta}}{\sigma} \right\|^2 \phi_{\boldsymbol{\theta},\sigma^2\mathbf{I}}(x) d\mathbf{x} \leq \int ... \int \left\| \frac{x}{\sigma} - \frac{\boldsymbol{\theta}}{\sigma} \right\|^2 \phi_{\boldsymbol{\theta},\sigma^2\mathbf{I}}(x) d\mathbf{x}$$

$$\frac{1}{\sigma^2} \int ... \int \left\| \sigma\boldsymbol{\delta}^* \left(\frac{x}{\sigma}\right) - \boldsymbol{\theta} \right\|^2 \phi_{\boldsymbol{\theta},\sigma^2\mathbf{I}}(x) d\mathbf{x} \leq \frac{1}{\sigma^2} \int ... \int \|x - \boldsymbol{\theta}\|^2 \phi_{\boldsymbol{\theta},\sigma^2\mathbf{I}}(x) d\mathbf{x}$$

$$\mathscr{R}\left(\sigma\boldsymbol{\delta}^*\left(\frac{\mathbf{X}}{\sigma}\right), \boldsymbol{\theta}\right) \leq \mathscr{R}\left(\boldsymbol{\delta}, \boldsymbol{\theta}\right)$$

o que significa que o estimador $\sigma\boldsymbol{\delta}^*\left(\frac{\mathbf{X}}{\sigma}\right)$ domina o estimador $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{X}$ para uma população $N_p\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$.

Particularizando o resultado para a média, $\bar{\mathbf{X}} \sim N\left(\boldsymbol{\mu}, \frac{\sigma^2}{n}\mathbf{I}\right)$, o estimador $\boldsymbol{\delta}(\bar{\mathbf{X}}) = \bar{\mathbf{X}}$ não é admissível como estimador de $\boldsymbol{\theta}$.

Por essa resultado, podemos obter a expressão do estimador de James-Stein no caso geral

$$\mathbf{X} \sim N\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$$

$$\mathbf{X} \sim N\left(\boldsymbol{\theta}, \frac{\sigma^2}{n}\mathbf{I}\right)$$

utilizando o estimador $\sigma\boldsymbol{\delta}^*\left(\frac{\mathbf{X}}{\sigma}\right)$ que tem menor risco, temos

$$\begin{aligned}
\frac{\sigma}{\sqrt{n}}\boldsymbol{\delta}_{JS}\left(\frac{\mathbf{X}}{\sigma/\sqrt{n}}\right) &= \frac{\sigma}{\sqrt{n}}\left(1 - \frac{p-2}{\left\|\frac{\mathbf{X}}{\sigma/\sqrt{n}}\right\|^2}\right)\frac{\mathbf{X}}{\sigma/\sqrt{n}} \\
&= \frac{\sigma}{\sqrt{n}}\left(1 - \frac{p-2}{\|\mathbf{X}\|^2}\left(\sigma/\sqrt{n}\right)^2\right)\frac{\mathbf{X}}{\sigma/\sqrt{n}} \\
&= \left(1 - \frac{\sigma^2}{n}\frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X}
\end{aligned}$$

Se $n = 1$ o estimador de James-Stein para $N\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$ é

$$\hat{\boldsymbol{\theta}}_{JS}\left(\mathbf{X}\right) = \left(1 - \sigma^2\frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X}.$$

2.6 Generalizações do estimador de James-Stein

O estimador de James-Stein é obtido como um encolhimento do vetor de dados $\mathbf{X}$ na direção da origem

$$\hat{\boldsymbol{\theta}}_{JS}\left(\mathbf{X}\right) = \left(1 - \frac{p-2}{\mathbf{X}'\mathbf{X}}\right)\mathbf{X}.$$

Suponha que se tenha algum tipo de informação que sugira que um determinado vetor $\boldsymbol{\mu}$ seja uma boa estimativa do vetor de médias $\boldsymbol{\theta}$. Neste caso, é possível se definir uma generalização do estimador de James-Stein fazendo-se encolhimento na direção do vetor $\boldsymbol{\mu}$, tal que $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)$ uma suposição inicial qualquer para $\boldsymbol{\theta}$. O estimador fica da seguinte forma, como representado pela Figura 4

$$\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X}) = \boldsymbol{\mu} + \left(1 - \frac{p-2}{(\mathbf{X}-\boldsymbol{\mu})'(\mathbf{X}-\boldsymbol{\mu})}\right)(\mathbf{X}-\boldsymbol{\mu})$$

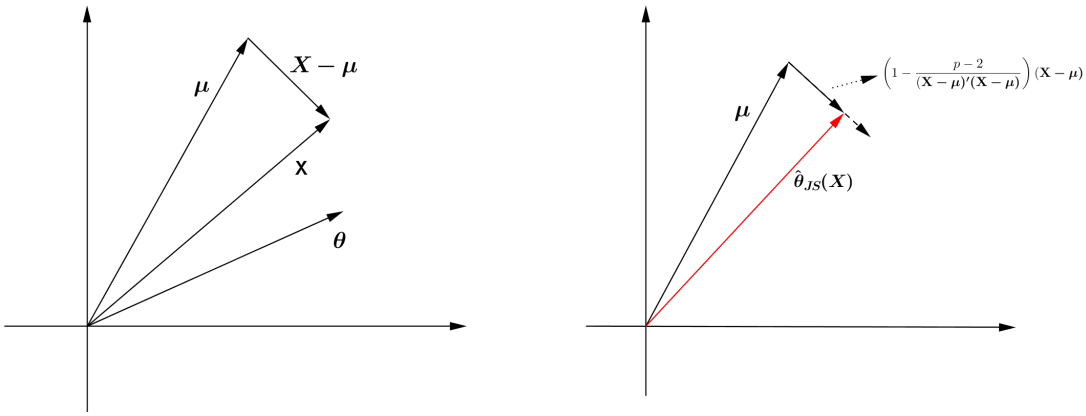

FIGURA 4: Geometria do estimador de James-Stein com encolhimento na direção do vetor $\boldsymbol{\mu}$.

Esse estimador domina $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{X}$ qualquer que seja o vetor $\boldsymbol{\mu}$ pois

$$\begin{aligned}\mathscr{R}\left(\hat{\boldsymbol{\theta}}_{JS}, \boldsymbol{\theta}\right) &= E_\theta\left[\left(\hat{\boldsymbol{\theta}}_{JS} - \boldsymbol{\theta}\right)^2\right] \\ &\leq p - \frac{(p-2)^2}{p - 2 + \sum_{i=1}^{p}(\theta_i - \mu_i)^2}.\end{aligned}$$

Se $\boldsymbol{\theta} = \boldsymbol{\mu}$ então o risco é 2 e, portanto, se p é muito maior que 2 o risco deste estimador é bastante inferior ao risco de $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{X}$ (EFRON; MORRIS, 1975).

Uma forma de se ter uma boa estimativa de $\boldsymbol{\theta}$ é utilizar a média amostral $\bar{\mathbf{X}} = \frac{\sum X_i}{p}$. Substituindo $\boldsymbol{\mu} = \frac{\sum \mu_i}{p}$ por $\bar{\mathbf{X}}$, encolhendo todo X_i na direção de $\mathbf{X}$, uma ideia sugerida por Lindley (1962) como citado em Efron e Morris (1975). Definindo o estimador e representado geometricamente na Figura 5

$$\hat{\boldsymbol{\theta}}_{EM}(\mathbf{X}) = \bar{\mathbf{X}} + \left(1 - \frac{p-3}{(\mathbf{X}-\bar{\mathbf{X}})'(\mathbf{X}-\bar{\mathbf{X}})}\right)\left(\mathbf{X}-\bar{\mathbf{X}}\right). \tag{2.29}$$

em que $\bar{\mathbf{X}} = \left(\bar{\mathbf{X}}, ..., \bar{\mathbf{X}}\right)$. Neste caso, $p-3$ é mais apropriado que $p-2$ uma vez que os dados já foram utilizados na estimação de $\boldsymbol{\theta}$. Esse estimador também domina $\boldsymbol{\delta}(\mathbf{X}) = \mathbf{X}$, pois

$$\begin{aligned}\mathscr{R}\left(\hat{\boldsymbol{\theta}}_{JS},\boldsymbol{\theta}\right) &= E_\theta\left[\left(\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X})-\boldsymbol{\theta}\right)^2\right] \\ &\leq p-\frac{(p-3)^2}{p-3+\sum_{i=1}^{p}\left(\theta_i-\bar{\boldsymbol{\theta}}\right)^2}\end{aligned}$$

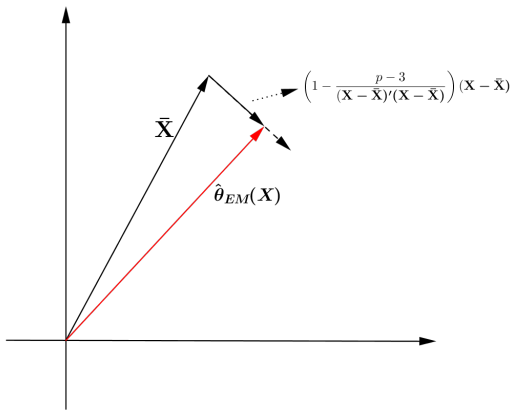

FIGURA 5: Geometria do estimador de James-Stein com encolhimento na direção do vetor $\bar{X}$.

2.7 O paradoxo de Stein

O método de estimação utilizando o estimador de James-Stein foi usado por autores como Efron e Morris (1977) que utilizaram as médias observadas para estimar quantidades não observadas e evidencia o paradoxo de Stein. De forma a contribuir com a compreensão da ideia proposta nestes artigos, os resultados foram expandidos.

Efron e Morris (1975, 1977), propuseram a estimação das médias de rebatidas observadas durante as primeiras semanas de jogos da liga de baseball nos EUA. Consideraram a média das rebatidas dos 18 maiores jogadores através de suas 45 tentativas de rebatidas até 26 de abril de 1970, isto representa cerca de um décimo de uma temporada completa. O problema era prever a média de rebatidas de cada jogador durante o restante da temporada usando apenas os dados da primeira coluna da Tabela 1. Essa amostra foi escolhida porque queriam entre 30 e 50 rebatidas para assegurar uma aproximação satisfatória da binomial pela distribuição normal. Para testar o método, incluíram um jogador extremamente bom e diferenciado, Clemente, criando uma situação favorável para o estimador de James-Stein. Efron e Morris (1975) assumiram que a média observada de rebatidas dos jogadores é uma boa aproximação para a verdadeira habilidade de rebater dos jogadores, não observada, para o restante da temporada.

Seja $\boldsymbol{\theta} = (\theta_1, ..., \theta_{18})$ o vetor da verdadeira habilidade de rebater e seja $\mathbf{X} = (X_1, ..., X_{18})$ o vetor

das observações feitas em 26 de abril de 1970. Efron e Morris (1975) fizeram várias suposições razoáveis para simplificar a aplicação do estimador de James-Stein. Cada rebatida foi modelada como uma variável aleatória que segue o modelo Bernoulli onde a probabilidade do jogador *i* obter um sucesso é p_i. Seja $\mathbf{X}$ o número total de sucessos obtidos na realização de 45, turnos de rebatidas independentes, $\mathbf{X}$ segue o modelo binomial, $X_i \sim B(45, p_i)$. Utilizando o teorema central do limite, a aproximação da binomial pela normal para cada x_i é feita através da variável

$$\begin{aligned} Y_i &\sim N(45p_i, 45p_i(1-p_i)) \\ \frac{Y_i}{45} &\sim N\left(p_i, \frac{1}{45}p_i(1-p_i)\right). \end{aligned}$$

Como a variância de X_i depende da média, a transformação $arcsen$ é utilizada para estabilizar a variância da distribuição binomial. Temos que $var\left[Y_i\right] = 45p_i(1-p_i)$, aplicaremos uma transformação $f(Y_i)$ que tem uma variância constante, pois o estimador de James-Stein requer variâncias iguais. Pela expansão de primeira ordem da série de Taylor

$$\begin{aligned} f(Y) &\approx f\left(p\right) + (Y-p)f'(p) \\ \left[f(Y) - f\left(p\right)\right]^2 &\approx (Y-p)^2\big(f'(p)\big)^2. \end{aligned}$$

Aplicando a esperança em ambos os lados da equação, tem-se o método delta

$$\begin{aligned} var\left[f(Y)\right] &\approx var\left[Y\right]\left[f'(p)\right]^2 \\ &\approx g(p)\big[f'(p)\big]^2. \end{aligned}$$

Seja,

$$f(p) = \int \frac{1}{[g(p)]^{1/2}} dp,$$

e

$$g(p) = \frac{1}{45}p(1-p),$$

logo,

$$f(p) = \int \frac{1}{\left[\frac{1}{45}p(1-p)\right]^{1/2}}dp$$
$$= \sqrt{45}\int \frac{1}{\sqrt{p-p^2}}dp$$

realizando o completamento de quadrado sob o termo dentro da raiz

$$f(p) = \sqrt{45}\int \frac{1}{\sqrt{\frac{1}{4} - \left(p-\frac{1}{2}\right)^2}dp}$$

fazendo a substituição $u = p - \frac{1}{2}$, $du = dp$, de modo que

$$f(p) = \sqrt{45}\int \frac{1}{\sqrt{\left(\frac{1}{2}\right)^2 - u^2}}du$$

substituindo $u = 1/2 \ \ sen\gamma$, obtendo $du = 1/2 \ \ cos\gamma d\gamma$ e $\sqrt{\left(\frac{1}{2}\right)^2 - u^2} = \frac{1}{2}\cos\gamma$, de forma que

$$\sqrt{45}\int \frac{1}{\sqrt{\left(\frac{1}{2}\right)^2 - u^2}}du = \int \frac{1}{1/2\cos\gamma}1/2\cos\gamma d\gamma$$
$$= \sqrt{45}\gamma$$
$$= \sqrt{45}\text{arcsen(2u)}$$
$$= \sqrt{45}\text{arcsen(2p} - 1).$$

Podemos nos certificar que esta transformação nos fornece variância igual a 1, ou seja,

$$var\left[f(Y)\right] \approx g(p)\left[\frac{\partial}{\partial p}\int \frac{1}{[g(p)]^{1/2}}dp\right]^2$$
$$= \frac{g(p)}{g(p)}\left[\frac{\partial}{\partial p}\int [g(p)]^{1/2}\frac{1}{[g(p)]^{1/2}}dp\right]^2$$
$$= \left[\frac{\partial}{\partial p}\int dp\right]^2$$
$$= \left[\frac{\partial}{\partial p}(p+c)\right]^2$$
$$= 1.$$

Como $\mathbf{Y} = \mathbf{f}(\mathbf{X})$ definimos

$$y_i = \sqrt{45}\,\text{arcsen}(2x_i - 1). \tag{2.30}$$

Como x_i é uma estimativa de p_i a média θ_i de Y_i é dada aproximadamente por $\theta_i = f(p_i)$. Portanto, $Y_i \sim N(\theta_i, 1)$, $\theta_i = \sqrt{45}\,\text{arcsen}(2p_i - 1)$. O problema é estimar $\boldsymbol{\theta}$ a partir de uma observação $\mathbf{Y}$ aproximadamente distribuída por $N((\boldsymbol{\theta}), \mathbf{I})$.

Efron e Morris (1975) implementaram o estimador de James-Stein, encolhendo na direção $\bar{\mathbf{Y}} = \left(\frac{1}{18}\sum_{i=1}^{18} y_i\right)\mathbf{1}_{18}$, definido anteriormente como estimador de Efron-Morris dado na equação (2.29),

$$\widehat{\boldsymbol{\theta}}_{JS}(Y) = \overline{\mathbf{y}} + \left(1 - \frac{(18-3)}{\sum_{i=1}^{18}(y_i - \overline{\mathbf{y}})^2}\right)(\mathbf{y} - \overline{\mathbf{y}}).$$

As estimativas obtidas $\hat{\boldsymbol{\theta}}$ podem ser retransformadas para as médias de rebatidas usando o inverso da transformação de (2.30), ou seja, $p_i = \dfrac{sen\left(y_i/\sqrt{45}\right)}{2} + 1$.

A estimação da verdadeira habilidade do jogador de rebater em termos da razão das médias de acerto observadas é dada por $x_i = \hat{p}_i^{EMV}$. O processo essencial de Stein é encolher todas as médias individuais para a média geral $\bar{\mathbf{y}}$. O procedimento de James-Stein faz uma suposição preliminar que todas as médias não observadas estão próximas da média geral $\bar{\mathbf{y}}$. O paradoxo de Stein consiste no fato que seus estimadores resultam em uma estimação melhor que a simples média individual de cada jogador, como pode-se verificar na Tabela 2.7 e de forma mais didática na Figura 6 obtida em (JAMES-STEIN...,2014).

Analisando a Figura 6, o estimador de James-Stein está mais próximo da média para a maioria dos jogadores. O paradoxo de Stein vem do seguinte fato: Clemente, que está no topo da Tabela 2.7 e é considerado um jogador excepcional devido a sua performance nos anos anteriores, é uma realização independente de Munson, na parte inferior da Tabela 2.7. Por que o bom desempenho de Clemente aumenta a predição para Munson? O estimador de James-Stein acrescenta o fator $\|\mathbf{X}\|$ no seu coeficiente de encolhimento que engloba todos as médias, assim os dados de Clemente afetam a estimativa de Munson e cada um dos demais jogadores. Comparando os dados apresentados na Tabela 2.7 pode-se verificar que $p_i = \hat{p}_i^{JS}$ se aproxima mais da média p_i, para a maioria dos jogadores, que o estimador de máxima verossimilhança $x_i = \hat{p}_i^{EMV}$.

Um método de comparar ambas as técnicas consiste em encontrar o erro quadrático total de cada estimador como exposto na Figura 7. As médias observadas y tem um erro quadrático total $\left\|\hat{\theta}^{EMV} - \theta\right\|^2 = 0,0777$ enquanto que o erro quadrático total dos estimadores de James-Stein é $\left\|\hat{\theta}^{JS} - \theta\right\|^2 = 0,022$. Segundo este critério de comparação, o método de James-Stein é 3,5 vezes mais preciso. Ainda com foco no

TABELA 1: Dados dos 18 maiores jogadores da liga de baseball do início da temporada de 1970 e valores transformados y_i e θ_i.

Jogadores	$x_i = \hat{p}_i^{EMV}$ estimador de máxima verossimilhança	p_i média das rebatidas para o restante da temporada	$p_i = \hat{p}_i^{JS}$ retransformação do estimador de James-Stein	y_i	θ_i
Clemente, Roberto	18/45= 0,400	0,346	0,290	-1,35	-2,10
Robinson, Frank	17/45=0,378	0,298	0,286	-1,66	-2,79
Howard, Frank	16/45=0,356	0,276	0,281	-1,97	-3,11
Johnstone, Jay	15/45=0,333	0,222	0,277	-2,28	-3,96
Berry, Ken	14/45=0,311	0,273	0,273	-2,60	-3,17
Spencer, Jim	14/45=0,311	0,270	0.273	-2,60	-3,20
Kessinger, Don	13/45=0,289	0,263	0,268	-2,92	-3,32
Alvarado, Luis	12/45=0,267	0,210	0,264	-3,26	-4,15
Santo, Ron	11/45=0,244	0,269	0,259	-3,60	-3,23
Swaboda, Ron	11/45=0,244	0,230	0,259	-3,60	-3,83
Unser, Del	10/45=0,222	0,264	0,254	-3,95	-3,30
Williams, Billy	10/45=0,222	0,256	0,254	-3,95	-3,43
Scott, George	10/45=0,222	0,303	0,254	-3,95	-2,71
Petrocelli, Rico	10/45=0,222	0,264	0,254	-3,95	-3,30
Rodriguez, Ellie	10/45=0,222	0,226	0,254	-3,95	-3,89
Campaneris, Bert	9/45=0,200	0,285	0,249	-4,32	-2,98
Munson, Thurman	8/45=0,178	0,316	0,244	-4,70	-2,53
Alvis, Max	7/45=0,156	0,200	0,239	-5,10	-4,32
Média Geral	0,265	0,265	0,265		

FONTE: Adaptada Efron e Morris (1975).

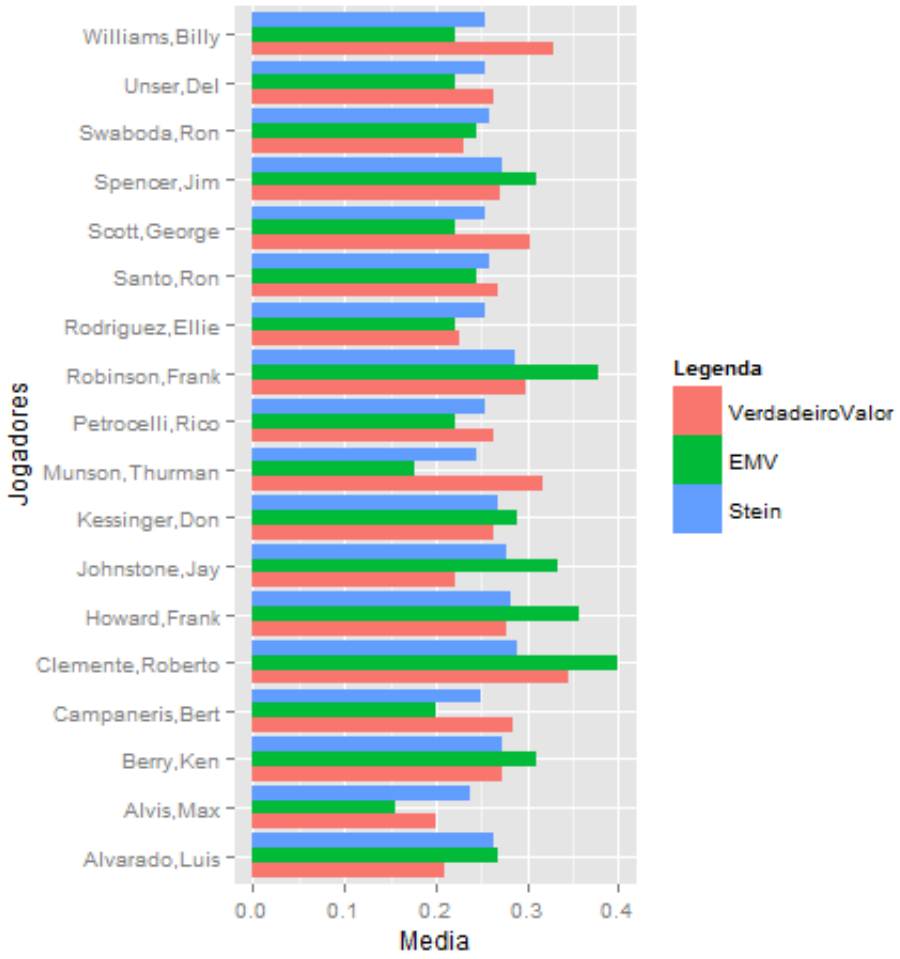

FIGURA 6: Comparação das médias estimadas para a habilidade de rebater dos jogadores de baseball.

erro quadrático médio é apresentada a Figura 7 que calcula o o erro quadrático médio de cada jogador para o estimador de máxima verossimilhança e o estimador de James-Stein (JAMES-STEIN..., 2014). O erro quadrático médio do estimador de máxima verossimilhança é menor apenas para os jogadores Swaboda, Rodrigues e Clemente.

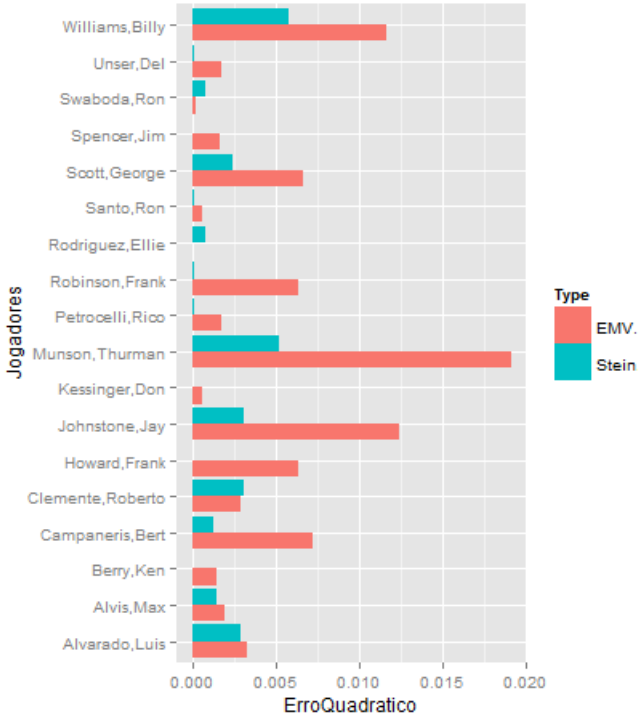

FIGURA 7: Comparação do erro quadrático médio entre o estimador de máxima verossimilhança e o estimador de James-Stein.

2.8 A geometria do estimador de James-Stein

Nesta seção, serão apresentadas a representação geométrica e interpretação da ideia inicial de Stein para introduzir este estimador de encolhimento.

2.8.1 O argumento geométrico original de Stein

A quantidade de encolhimento permitida para garantir que o estimador de encolhimento preserve a dominação do estimador média amostral, é livre de suposição de normalidade. Brandwein e Strawderman (1990) enfatizam a escolha da quantidade de encolhimento, considerando estimadores esfericamente simétricos, em vez das demonstrações sobre os ganhos adquiridos com o encolhimento.

O desenvolvimento do argumento geométrico é devido a Stein (1962) e explicitado com detalhes em Brandwein e Strawderman (1990). Seja $\mathbf{X}= (X_1, ..., X_p)$ um vetor aleatório p-dimensional com vetor de médias $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$ com componentes independentes e variâncias supostas conhecidas, σ^2. Como $E[(\mathbf{X} - \boldsymbol{\theta})] = \mathbf{0}$, tem-se que $E\left[(\mathbf{X} - \boldsymbol{\theta})' \boldsymbol{\theta}\right] = 0$ e, portanto, em média pode-se afirmar que os vetores $\mathbf{X} - \boldsymbol{\theta}$ e $\boldsymbol{\theta}$ são ortogonais. Considere uma realização $\mathbf{X}$ desse vetor aleatório. Pode-se esperar que os vetores $\mathbf{X} - \boldsymbol{\theta}$ e $\boldsymbol{\theta}$ sejam quase ortogonais. Como $E\left[\|\mathbf{X}\|^2\right]$ superestima $\|\boldsymbol{\theta}\|^2$, ou seja, $E\left[\|X\|^2\right] = p\sigma^2 + \|\theta\|^2$ é razoável supor que para algum valor de a próximo de 1 o vetor $a\mathbf{X}$ esteja mais próximo do vetor $\boldsymbol{\theta}$ que o próprio vetor $\mathbf{X}$. A ideia é projetar o vetor $\boldsymbol{\theta}$ na direção de $\mathbf{X}$ e o vetor resultante desta projeção, ou algo próximo disto, pode ser um estimador melhor. Esta projeção depende de $\boldsymbol{\theta}$ e, portanto, não é válida como estimador. Denotando essa projeção por $(1 - a)\mathbf{X}$ o problema será aproximar o valor a.

Supondo a ortogonalidade entre $\mathbf{X} - \boldsymbol{\theta}$ e $\boldsymbol{\theta}$ e assumindo que $\|\mathbf{X}\|^2 = E\left[\|\mathbf{X}\|^2\right]$, ou seja, $\|\mathbf{X}\|^2 = p\sigma^2 + E\left[\|\theta\|^2\right]$ e similarmente supondo que $\|\mathbf{X} - \boldsymbol{\theta}\|^2 = E\left[\|\mathbf{X} - \boldsymbol{\theta}\|^2\right] = p\sigma^2$ aplica-se o teorema de Pitágoras, conforme Figura 8 (b).

Considere o triângulo retângulo ABC, então

$$\begin{aligned} \|\mathbf{Y}\|^2 &= \|\mathbf{X} - \boldsymbol{\theta}\|^2 - a^2\|\mathbf{X}\|^2 \\ &\simeq p\sigma^2 - a^2\|\mathbf{X}\|^2. \end{aligned}$$

Considerando o triângulo retângulo OAB, temos que $\|\boldsymbol{\theta}\|^2 = \|\mathbf{X}\|^2 - \|\mathbf{X} - \boldsymbol{\theta}\|^2$. Agora considere o

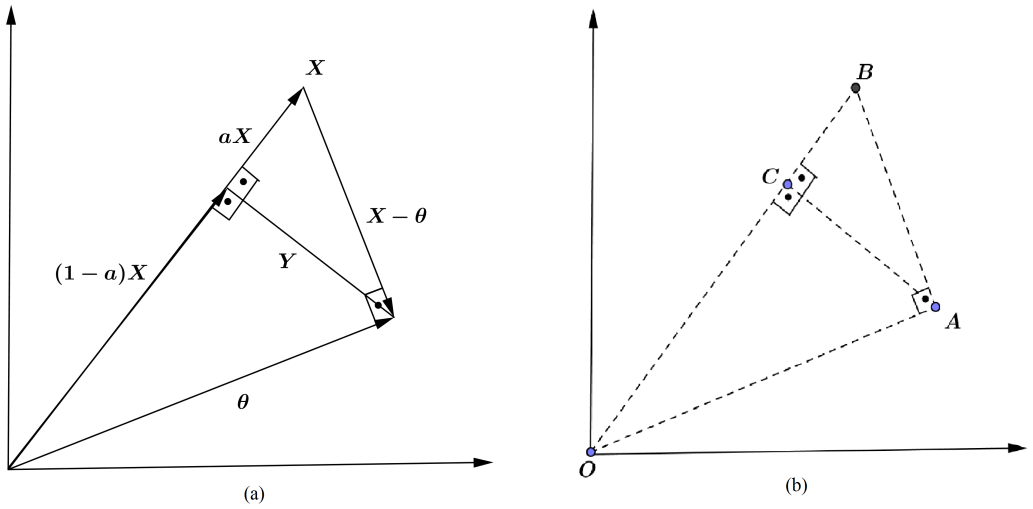

FIGURA 8: Projeção do vetor de parâmetros $\boldsymbol{\theta}$ no vetor de dados $\mathbf{X}$.

triângulo OAC

$$\begin{aligned}
\|\mathbf{Y}\|^2 &= \|\boldsymbol{\theta}\|^2 - (1-a)^2\|\mathbf{X}\|^2 \\
&= \|\mathbf{X}\|^2 - \|\mathbf{X}-\boldsymbol{\theta}\|^2 - (1-a)^2\|\mathbf{X}\|^2 \\
&\simeq \|\mathbf{X}\|^2 - p\sigma^2 - (1-a)^2\|\mathbf{X}\|^2.
\end{aligned}$$

Igualando essas duas equações temos:

$$\begin{aligned}
p\sigma^2 - a^2\|\mathbf{X}\|^2 &= \|\mathbf{X}\|^2 - p\sigma^2 - (1-a)^2\|\mathbf{X}\|^2 \\
(1-2a)\|\mathbf{X}\|^2 &= \|\mathbf{X}\|^2 - 2p\sigma^2 \\
2a\|\mathbf{X}\|^2 &= 2p\sigma^2 \\
a &\simeq \frac{p\sigma^2}{\|\mathbf{X}\|^2}
\end{aligned}$$

assim o estimador de $\boldsymbol{\theta}$ sugerido por este argumento geométrico é

$$\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X}) = (1-a)\,\mathbf{X} = \left(\mathbf{1} - \frac{\mathbf{p}\boldsymbol{\sigma}^{\mathbf{2}}}{\|\mathbf{X}\|^{\mathbf{2}}}\right)\mathbf{X}.$$

Observe que o argumento acima não depende da normalidade de $\mathbf{X}$ e, é válido mesmo que $\boldsymbol{\theta}$ seja um parâmetro de locação. O argumento geométrico funciona para qualquer σ^2. Este argumento sugere a possibilidade de melhoria do vetor não viesado $\mathbf{X}$ através do encolhimento em direção a origem de forma geral.

2.9 Estimadores esfericamente simétricos

Brown e Zao (2012) apresentam uma elaborada justificativa geométrica para o estimador de James-Stein, como será detalhadamente descrito nesta seção. A ideia é justificar geometricamente o teorema de Stein, ou seja, o estimador média amostral da normal multivariada é admissível se, e somente se, $p \geq 2$.

Definição 2.17 *Estimadores esfericamente simétricos são aqueles que satisfazem*

$$\boldsymbol{\delta}(\mathbf{X}) = \phi(\|\mathbf{X}\|)\,\mathbf{X}$$

para alguma função escalar ϕ.

Note que se $\phi < 1$ o estimador $\boldsymbol{\delta}$ é um estimador de encolhimento simples uma vez que, o encolhimento está na direção do próprio $\mathbf{X}$. Uma das grandes vantagens de se considerar estimadores dessa forma é possibilitar o estudo da variável multidimensional em um sistema de coordenadas bidimensional. O estimador de James-Stein é um estimador esfericamente simétrico em que ϕ é uma função escalar. A justificativa é clara, pois se $\mathbf{X}$ está sobre uma esfera de raio r então, a imagem de $\mathbf{X}$ pelo estimador $\boldsymbol{\delta}$ estará em uma esfera de raio $\phi(r)r$, pois $\|\boldsymbol{\delta}(\mathbf{X})\| = \phi(\|\mathbf{X}\|)\,\|\mathbf{X}\| = \phi(r)r$. Seja $E[\mathbf{X}] = \boldsymbol{\theta}$ e $\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$ o vetor unitário na direção de $\boldsymbol{\theta}$. Dessa forma, a projeção ortogonal do vetor $\mathbf{X}$ pertencente ao espaço $\mathbb{R}^p$, na direção do vetor $\boldsymbol{\theta}$ pode ser escrita de maneira única na forma $P_{\boldsymbol{\theta}}\mathbf{X} = \mathbf{X} \cdot \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$, conforme Figura 9. Definimos, então, a variável aleatória unidimensional, $X_1 = \mathbf{X} \cdot \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$, ou seja, um escalar pois, representa o produto interno entre dois vetores. Um outro vetor aleatório definido por $\mathbf{X_{(2)}} = \mathbf{X} - \mathbf{P_{\boldsymbol{\theta}}X}$.

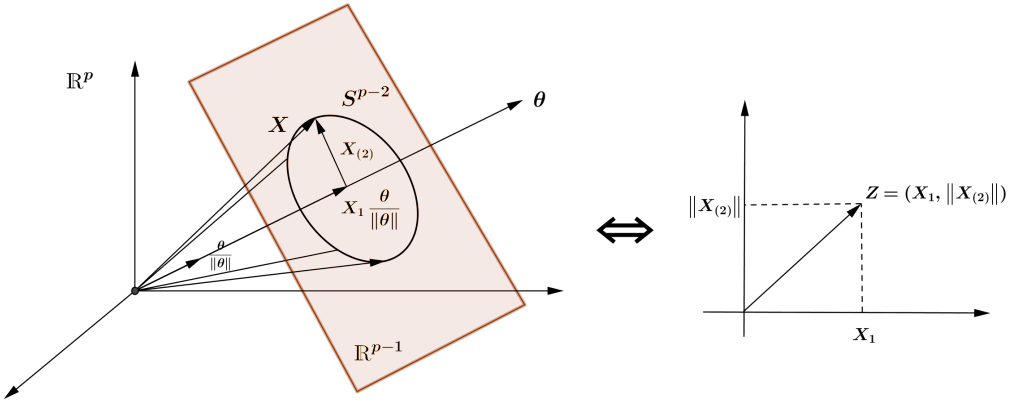

FIGURA 9: Representação bidimensional para estimadores com simetria esférica.

O vetor $\mathbf{X_{(2)}}$ pertence ao subespaço vetorial de dimensão $p-1$ perpendicular ao vetor $\boldsymbol{\theta}$, ou seja, um hiperplano de dimensão $p-1$. Temos que a esfera S^{p-2} contida neste hiperplano, está centrada na extremidade do vetor $P_{\boldsymbol{\theta}}\mathbf{X} = \mathbf{X_1}\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$ e tem raio $\left\|\mathbf{X_{(2)}}\right\|$. Podemos representar cada esfera S^{p-2} como um

ponto em $\mathbb{R}^2$. Para definirmos as coordenadas em $\mathbb{R}^2$, em um dos eixos representaremos $\left\|\mathbf{X}_{(\mathbf{2})}\right\|$ e no outro eixo representaremos a coordenada X_1. Analisando a norma de $\mathbf{X}$,

$$\|\mathbf{X}\|^2 = \|X_1\|^2 + \left\|\mathbf{X}_{(\mathbf{2})}\right\|^2$$

e a quantidade de encolhimento do vetor $\mathbf{X}$ é dada por

$$\phi(\|\mathbf{X}\|) = \phi\left(\sqrt{X_1^2 + \left\|\mathbf{X}_{(\mathbf{2})}\right\|^2}\right)$$

e, portanto, não depende da direção de $\mathbf{X}_{(\mathbf{2})}$ mas, apenas de sua norma. Em razão desse fato, definiremos a coordenada do segundo eixo por $\left\|\mathbf{X}_{(\mathbf{2})}\right\| = R$.

Um vetor aleatório $\mathbf{Z}$ neste espaço $\mathbb{R}^2$ é dado por $\mathbf{Z} = (X_1, R)$. Observe que cada vetor $\mathbf{Z}$ em $\mathbb{R}^2$ corresponde a uma esfera S^{p-2} no espaço afim p-dimensional perpendicular í Â $X_1 \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$, conforme Figura 9.

A distribuição das coordenadas do vetor aleatório $\mathbf{Z} = (X_1, R)$ são definidas a partir de $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \mathbf{I})$ então, como X_1 é uma variável univariada $X_1 \sim N(\|\boldsymbol{\theta}\|, 1)$ e como o vetor $\mathbf{X}_{(\mathbf{2})}$ é uma projeção ortogonal no subespaço perpendicular ao vetor de médias, este vetor é normal $\mathbf{X}_{(\mathbf{2})} \sim \mathbf{N_{p-1}}(\mathbf{0}, \mathbf{I})$, portanto, $R^2 = \left\|\mathbf{X}_{(\mathbf{2})}\right\|^2$ é soma de quadrados de normais padrão independentes, logo $R^2 \sim \chi^2_{p-1}$. Segue da normalidade e da ortogonalidade que X_1 e R são independentes.

O estimador com simetria esférica $\boldsymbol{\delta}(\mathbf{X}) = \phi(\|\mathbf{X}\|)\mathbf{X}$ definido em $\mathbb{R}^p$, pode ser definido no $\mathbb{R}^2$ como o estimador $\boldsymbol{\delta}'(\mathbf{Z}) = \phi(\|\mathbf{Z}\|)\mathbf{Z}$.

A relação fundamental entre $\mathbf{X} = \left(X_1, \mathbf{X}_{(\mathbf{2})}\right)$ e $\mathbf{Z} = (X_1, R)$ é que a função perda quadrática dos seus estimadores são iguais. Observando que nesse novo sistema de coordenadas o vetor de médias $\boldsymbol{\theta}$ tem coordenadas $\boldsymbol{\theta} = (\|\boldsymbol{\theta}\|, 0)$, tem-se

$$\begin{aligned}
\|\boldsymbol{\delta}(\mathbf{X}) - \boldsymbol{\theta}\|^2 &= \|\phi(\|\mathbf{X}\|)\mathbf{X} - \boldsymbol{\theta}\|^2 \\
&= \left\|\left(\phi(\|\mathbf{X}\|)X_1, \phi(\|\mathbf{X}\|)\mathbf{X}_{(\mathbf{2})}\right) - (\|\boldsymbol{\theta}\|, 0)\right\|^2 \\
&= \|\phi(\|\mathbf{X}\|)X_1 - \|\boldsymbol{\theta}\|\|^2 + \left\|\phi(\|\mathbf{X}\|)\mathbf{X}_{(\mathbf{2})}\right\|^2 \\
&= \|\phi(\|\mathbf{Z}\|)X_1 - \|\boldsymbol{\theta}\|\|^2 + \phi(\|\mathbf{Z}\|)\left\|\mathbf{X}_{(\mathbf{2})}\right\|^2 \\
&= \|\phi(\|\mathbf{Z}\|)X_1 - \|\boldsymbol{\theta}\|\|^2 + \phi(\|\mathbf{Z}\|)R^2 \\
&= \left\|\boldsymbol{\delta}'(\mathbf{Z}) - \|\boldsymbol{\theta}\|\right\|^2.
\end{aligned}$$

A função risco desses estimadores é dada por

$$\mathscr{R}\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)=E_{\boldsymbol{\theta}}\left[\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\right]=\int\cdots\int_{\mathbb{R}^{p}}\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\prod f(x_{i},\theta_{i})dx_{i}.$$

A ideia é aplicar o Teorema de Fubinni para cada $\mathbf{Z}$ em $\mathbb{R}^2$. Fixado $\mathbf{Z}\in\mathbb{R}^{\mathbf{2}}$ considere

$$A_{\mathbf{Z}}=\left\{\mathbf{X}=\left(X_{1},\mathbf{X}_{(\mathbf{2})}\right),\ \text{ tal que }\mathbf{Z}=(X_{1},R)\right\},$$

onde $R=\left\|\mathbf{X}_{(\mathbf{2})}\right\|$. Vimos que $A_{\mathbf{Z}}$ é uma esfera em um subespaço afim. Sobre $A_{\mathbf{Z}}$, $\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}$ é uma constante e, portanto,

$$\int_{A_{\mathbf{Z}}}\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\prod f(x_{i},\theta_{i})dx_{i}=\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\int_{A_{\mathbf{Z}}}\prod f(x_{i},\theta_{i})dx_{i}.$$

Definindo

$$g(x_{1},r)=\int_{A_{\mathbf{Z}}}\prod f(x_{i},\theta_{i})dx_{i}$$

tem-se

$$\begin{aligned}\mathscr{R}\left(\boldsymbol{\delta},\boldsymbol{\theta}\right) &= \int\cdots\int_{\mathbb{R}^{p}}\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\prod f(x_{i},\theta_{i})dx_{i}\\ &= \int_{\mathbf{Z}}\int_{A_{\mathbf{Z}}}\left\|\boldsymbol{\delta}(\mathbf{X})-\boldsymbol{\theta}\right\|^{2}\prod f(x_{i},\theta_{i})dx_{i}\\ &= \int_{\mathbf{Z}}\left\|\boldsymbol{\delta}'(\mathbf{Z})-\boldsymbol{\theta}\right\|^{2}g(x_{1},r)dx_{1}dr\\ &= \mathscr{R}\left(\boldsymbol{\delta}',\boldsymbol{\theta}\right).\end{aligned}$$

Como os dois estimadores possuem a mesma função risco, o problema de admissibilidade de $\boldsymbol{\delta}_0(\mathbf{X})=\mathbf{X}$ em relação aos estimadores simétricos esféricos $\boldsymbol{\delta}(\mathbf{X})=\phi\left\|\mathbf{X}\right\|\mathbf{X}$ pode ser estudado em duas dimensões assim como a admissibilidade do estimador $\boldsymbol{\delta}_0'(\mathbf{Z})=\mathbf{Z}$ em relação aos estimadores simétricos esféricos $\boldsymbol{\delta}'(\mathbf{Z})=\phi\left\|\mathbf{Z}\right\|\mathbf{Z}$. Tem-se que

$$E\left[\mathbf{Z}\right]=\left(E[X_{1}],E[R]\right)=\left(\left\|\boldsymbol{\theta}\right\|,E[R]\right).$$

Como $R^2 \sim \chi^2_{p-1}$ logo $E\left[R^2\right] = p-1$. Desse fato prova-se que

$$\begin{aligned} E[R] &= \int_0^\infty \sqrt{x} \frac{\left(\frac{1}{2}\right)^{\frac{p-1}{2}} (x)^{\frac{p-1}{2}-1} e^{-\frac{1}{2}x}}{\Gamma\left(\frac{p-1}{2}\right)} dx \\ &= \frac{\Gamma\left(\frac{p}{2}\right)\left(\frac{1}{2}\right)^{-\frac{1}{2}}}{\Gamma\left(\frac{p-1}{2}\right)} \int_0^\infty \frac{\left(\frac{1}{2}\right)^{\frac{p}{2}} (x)^{\frac{p}{2}-1} e^{-\frac{1}{2}x}}{\Gamma\left(\frac{p}{2}\right)} dx \\ &= \frac{\sqrt{2}\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)}. \end{aligned}$$

Temos que $X_i \sim N(0,1)$ logo, $X_i^2 \sim \chi^2_1$. Aplicando o teorema central do limite

$$\begin{aligned} \frac{\frac{1}{p-1}\sum_{i=1}^{p-1} X_i^2 - 1}{2/\sqrt{p-1}} &\sim N(0,1) \\ \frac{1}{p-1}\sum_{i=1}^{p-1} X_i^2 - 1 &\sim N\left(0, \frac{4}{p-1}\right) \\ \frac{1}{p-1}\sum_{i=1}^{p-1} X_i^2 &\sim N\left(1, \frac{4}{p-1}\right) \\ \sum_{i=1}^{p-1} X_i^2 &\sim N\left(p-1, 4(p-1)\right) \end{aligned}$$

Assintoticamente, temos que $R^2 = \sum_{i=1}^{p-1} X_i^2 \sim N\left(p-1, 4(p-1)\right)$. Sabendo que $R = \sqrt{R^2}$, para determinar $E[R]$ utilizaremos o delta método

$$\begin{aligned} E\left[g(\mathbf{X})\right] &\approx g\left[E(\mathbf{X})\right] \\ E\left[\sqrt{R^2}\right] &\approx \sqrt{E\left[R^2\right]} = \sqrt{p-1}. \end{aligned}$$

Quando p cresce esta aproximação fica cada vez melhor, portanto uma observação típica do vetor aleatório $\mathbf{Z}$ deve estar próxima do ponto $\left(\|\boldsymbol{\theta}\|, \sqrt{p-1}\right)$, conforme Figura 10.

Como se quer minimizar a distância da estimativa ao ponto $(\|\boldsymbol{\theta}\|, 0)$ é razoável que se realize um encolhimento na direção de $\mathbf{Z}$ até um vetor $\boldsymbol{\delta}'(\mathbf{Z})$ com distância mínima ao vetor $(\|\boldsymbol{\theta}\|, 0)$, Figura 11 . Stein considerou que o argumento geométrico era tão claro que poderia servir como base para a demonstração da inadmissibilidade de $\boldsymbol{\delta}'_0(\mathbf{Z}) = \mathbf{Z}$. O argumento é baseado no fato que se p é suficientemente grande o valor

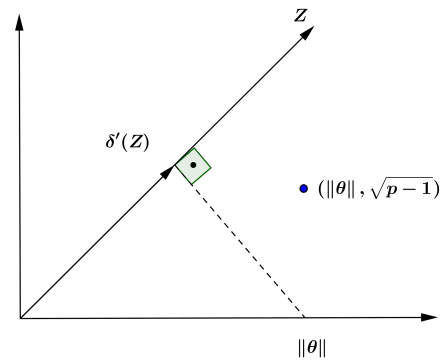

FIGURA 10: O estimador $\boldsymbol{\delta}'(\mathbf{Z})$ como encolhimento do vetor $\mathbf{Z}$.

observado $\mathbf{Z}$ deve estar próximo ao vetor $\left(\|\boldsymbol{\theta}\|, \sqrt{p-1}\right)$ e satisfaz $\|\mathbf{Z}\|^2 > \|\boldsymbol{\theta}\|^2$.

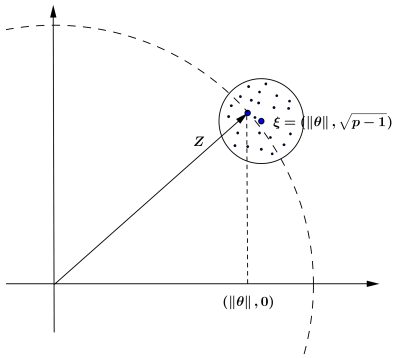

FIGURA 11: Uma observação típica de $\mathbf{Z}$ que satisfaz $\|\mathbf{Z}\| > \|\theta\|$.

Mas, como $\mathbf{Z}$ está próximo de $\xi = \left(\|\boldsymbol{\theta}\|, \sqrt{p-1}\right)$

$$\|\mathbf{Z}\|^2 \simeq \|\theta\|^2 + p - 1.$$

Desta forma, pode-se afirmar que $\|\mathbf{Z}\|^2 = \|\boldsymbol{\theta}\|^2 + p + \vartheta(\sqrt{p})$ em que ϑ é função de aproximação próxima de zero. Segue que

$$\begin{aligned} \|\theta\| &= \sqrt{\|\mathbf{Z}\|^2 - p - \vartheta(\sqrt{p})} \\ &= \|\mathbf{Z}\| \sqrt{1 - \frac{p + \vartheta(\sqrt{p})}{\|\mathbf{Z}\|^2}}. \end{aligned}$$

Utilizando semelhança entre os triângulos retângulos Δ_{OAB} e Δ_{OAC} da Figura 12, o fator de encolhimento é então sugerido por

$$\begin{aligned}
\frac{\|\overline{OB}\|}{\|\overline{OA}\|} &= \frac{\|\overline{OA}\|}{\|\overline{OC}\|} \\
\frac{\|\mathbf{Z}\|}{\|\theta\|} &= \frac{\|\theta\|}{\alpha\,\|\mathbf{Z}\|} \\
\alpha &= \frac{\|\theta\|^2}{\|\mathbf{Z}\|^2} \\
\alpha &= \frac{\left(\|\mathbf{Z}\|\sqrt{1-\dfrac{p+\upsilon(\sqrt{p})}{\|\mathbf{Z}\|^2}}\right)^2}{\|\mathbf{Z}\|^2} \\
\alpha &= \left(1-\frac{p+\upsilon(\sqrt{p})}{\|\mathbf{Z}\|^2}\right).
\end{aligned}$$

Stein então, propõe estimadores da forma

$$\boldsymbol{\delta}_p(\mathbf{X}) = \left(\mathbf{1} - \frac{\mathbf{p}}{\|\mathbf{X}\|^{\mathbf{2}}}\right)\mathbf{X}.$$

O argumento acima não é suficiente para a demonstração da inadmissibilidade pois, tem-se que fixar $\|\boldsymbol{\theta}\|$ e fazer p crescer. O argumento teria que ser uniforme em $\boldsymbol{\theta}$ e p, isto é, ocorrer simultaneamente $\|\boldsymbol{\theta}\| \to \infty$ e $\|p\| \to \infty$. De fato, o argumento acima faz com que $P_\theta\left(\|\mathbf{X}\| \geq \|\boldsymbol{\theta}\|\right)$ tende a 1 quando p é grande como mencionado, mas com $\boldsymbol{\theta}$ fixado. Portanto, um argumento geométrico mais elaborado é necessário para demonstrações heurísticas da inadmissibilidade de $\boldsymbol{\delta}_0(\mathbf{X}) = \mathbf{X}$, complementando os argumentos geométricos de Stein.

Para quantificar a quantidade de encolhimento necessária, vamos supor que $\mathbf{Z} = \xi$. Tal fato não é arbitrário, pois espera-se que $\mathbf{Z}$ esteja próximo de ξ. Aplicando o teorema de Pitágoras, conforme Figura 12

Portanto, considerando Δ_{OAB} temos que $\|\xi\|^2 = \|\boldsymbol{\theta}\|^2 + p - 1$ e utilizando semelhança de triângulos

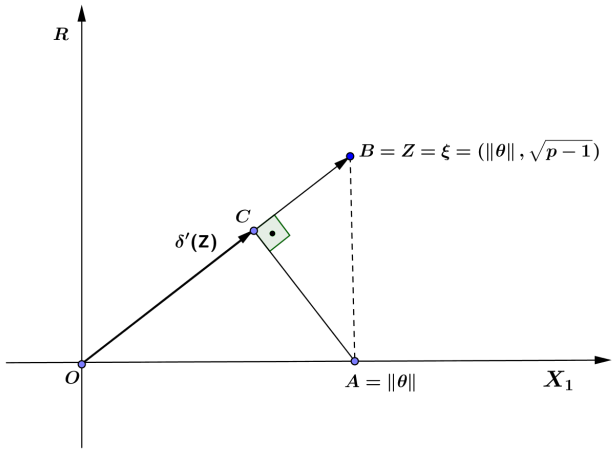

FIGURA 12: Geometria do estimador ótimo $\boldsymbol{\delta}'(\mathbf{Z})$.

para Δ_{OAB} e Δ_{OAC} temos

$$\begin{aligned}
\frac{\left\|\overline{OB}\right\|}{\left\|\overline{OA}\right\|} &= \frac{\left\|\overline{OA}\right\|}{\left\|\overline{OC}\right\|} \\
\frac{\|\xi\|}{\|\theta\|} &= \frac{\|\boldsymbol{\theta}\|}{\left\|\boldsymbol{\delta}'(\xi)\right\|} \\
\left\|\delta'(\xi)\right\| &= \frac{\|\theta\|^2}{\|\xi\|} \\
\left\|\delta'(\xi)\right\| &= \frac{\|\xi\|^2 - p + 1}{\|\xi\|} \\
\left\|\delta'(\xi)\right\| &= \frac{\|\xi\|^2 \left(1 - \frac{p-1}{\|\xi\|^2}\right)}{\|\xi\|} \\
\left\|\delta'(\xi)\right\| &= \left(1 - \frac{p-1}{\|\xi\|^2}\right) \|\xi\|
\end{aligned}$$

assim,

$$\delta'(\xi) = \left(1 - \frac{p-1}{\|\xi\|^2}\right) \xi.$$

Substituindo ξ por uma observação típica $\mathbf{Z}$ tem-se

$$\boldsymbol{\delta}'(\mathbf{Z}) = \left(1 - \frac{p-1}{\|\mathbf{Z}\|^2}\right) \mathbf{Z}$$

que em termos de variável original $\mathbf{X}$ nos dá o estimador

$$\boldsymbol{\delta}_{p-1}(\mathbf{X}) = \boldsymbol{\delta}(\mathbf{X}) = \left(1 - \frac{p-1}{\|\mathbf{X}\|^2}\right) \mathbf{X}.$$

Como neste argumento não há necessidade de $p \to \infty$, ele sugere a inadmissibilidade de $\boldsymbol{\delta}_0(\mathbf{X}) = \mathbf{X}$. Mas, aqui novamente se tem um erro, pois o argumento é também válido para $p = 2$ e $\boldsymbol{\delta}_1(\mathbf{X})$ não domina $\boldsymbol{\delta}_0(\mathbf{X}) = \mathbf{X}$. Um pouco mais de geometria, é necessário. O que se tem é que, com alta probabilidade, $\mathbf{Z}$ está próximo de ξ. Na tentativa de se modelar esta variabilidade em torno de ξ, vamos considerar 2 valores equiprováveis para $\mathbf{Z}$,

$$\xi_{\pm} = \left(\|\boldsymbol{\theta}\| \pm 1, \sqrt{p-1} \right).$$

Observe que $X_1 \sim N\left(\|\boldsymbol{\theta}\|, 1\right)$ e, portanto simétrica em relação a $\|\boldsymbol{\theta}\|$. A variância de X_1 é igual a 1, e a variância de R, isto é, a variância do vetor aleatório $\mathbf{Z}$, na direção ortogonal a $\boldsymbol{\theta}$, não reflete bem a variação do vetor aleatório $\mathbf{X}$, em razão disto será desprezada. Como vamos considerar um encolhimento na direção do vetor $\mathbf{Z}$, fica também enfatizado que a variação na direção ortogonal a $\boldsymbol{\theta}$ parece ser de menor importância. Considere, então, o estimador obtido por encolhimento da forma

$$\delta_C(\mathbf{Z}) = \left(\mathbf{1} - \frac{\mathbf{C}}{\|\mathbf{Z}\|^{\mathbf{2}}} \right) \mathbf{Z}.$$

Nos pontos ξ_+ e ξ_- as estimativas $\delta_C(\xi_+)$ e $\delta_C(\xi_-)$ são

$$\begin{aligned} \delta_C(\xi_{\pm}) &= \left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right) \xi_{\pm} \\ &= \left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right) \left(\|\boldsymbol{\theta}\| \pm 1, \sqrt{p-1} \right) \end{aligned} \tag{2.31}$$

e definem perdas L_+ e L_-

$$\begin{aligned} L_{\pm} &= \|\delta_C(\xi_{\pm}) - (\|\boldsymbol{\theta}\|, 0)\|^2 \\ &= \left\| \left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right) \xi_{\pm} - (\|\boldsymbol{\theta}\|, 0) \right\|^2 \\ &= \left(\left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right) (\|\boldsymbol{\theta}\| \pm 1) - \|\boldsymbol{\theta}\| \right)^2 + \left(\left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right) \left(\sqrt{p-1} - 0 \right) \right)^2 \\ &= \left[\pm 1 - \frac{C}{\|\xi_{\pm}\|^2} (\|\boldsymbol{\theta}\| \pm 1) \right]^2 + \left(1 - \frac{C}{\|\xi_{\pm}\|^2} \right)^2 (p-1) \\ &= L_{\pm}^{(1)} + L_{\pm}^{(2)}. \end{aligned}$$

Vamos considerar o risco condicional, dado que $\mathbf{Z} = \xi_+$ ou $\mathbf{Z} = \xi_-$, conforme Figura 13, isto é, a perda

média (risco) para estes dois valores de $\mathbf{Z}$ é dada por

$$\mathscr{R}\left(\delta_C, \boldsymbol{\theta}\right) = \frac{1}{2}\left[\left\|\delta_C(\xi_+) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2 + \left\|\delta_C(\xi_-) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2\right].$$

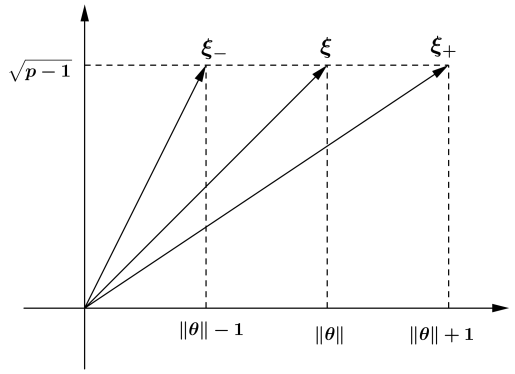

FIGURA 13: Representação de $\mathbf{Z} = \xi_+$ e $\mathbf{Z} = \xi_-$.

Como o risco é uma soma de quadrados aplicando o teorema de Pitágoras, podemos decompor o risco na direção de X_1 e de R que denominaremos $R^{(1)}_{|\xi_\pm}$ e $R^{(2)}_{|\xi_\pm}$, então,

$$\begin{aligned}\mathscr{R}\left(\delta_C, \boldsymbol{\theta}\right) &= \frac{1}{2}\left(L^{(1)}_+ + L^{(1)}_-\right) + \frac{1}{2}\left(L^{(2)}_+ + L^{(2)}_-\right) \\ &= R^{(1)}_{|\xi_\pm} + R^{(}_{|\xi_\pm}2)\end{aligned}$$

Vamos comparar com o risco de δ_0 para a mesma condição $Z = \xi_+$ ou $Z = \xi_-$

$$\begin{aligned}\mathscr{R}\left(\delta_0, \boldsymbol{\theta}\right) &= \frac{1}{2}\left[\left\|\delta_0(\xi_+) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2 + \left\|\delta_0(\xi_-) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2\right] \\ &= \frac{1}{2}\left[\left\|\left(\|\boldsymbol{\theta}\| + 1, \sqrt{p-1}\right) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2 + \left\|\left(\|\boldsymbol{\theta}\| - 1, \sqrt{p-1}\right) - \left(\|\boldsymbol{\theta}\|, 0\right)\right\|^2\right] \\ &= \frac{1}{2}\left[\left\|\left(1, \sqrt{p-1}\right)\right\|^2 + \left\|\left(-1, \sqrt{p-1}\right)\right\|^2\right] \\ &= \frac{1}{2}\left[1^2 + \left(\sqrt{p-1}\right)^2 + (-1)^2 + \left(\sqrt{p-1}\right)^2\right] \\ &= 1 + (p-1).\end{aligned}$$

Para o estimador $\boldsymbol{\delta}_0$, tem-se $\mathscr{R}^{(1)}_{|\xi_\pm} = 1$ e $\mathscr{R}^{(2)}_{|\xi_\pm} = p - 1$.

A diferença dos riscos dos estimadores $\boldsymbol{\delta}_0(\mathbf{X}) = \mathbf{X}$ e δ_C, dado pela equação (2.31), na direção de X_1 e de R respectivamente são

$$
\begin{aligned}
\mathscr{R}(,\boldsymbol{\delta}_0,\boldsymbol{\theta}) - \mathscr{R}(\boldsymbol{\delta}_c,\boldsymbol{\theta}) &= 1 - \mathscr{R}^{(1)}_{|\xi_\pm} \\
&= 1 - \frac{1}{2}\left(L_+^{(1)} + L_-^{(1)}\right) \\
&= 1 - \frac{1}{2}\left[\left(1 - \frac{C\left(\|\boldsymbol{\theta}\|+1\right)}{\|\xi_+\|^2}\right)^2 + \left(-1 - \frac{C\left(\|\boldsymbol{\theta}\|-1\right)}{\|\xi_-\|^2}\right)^2\right] \\
&= 1 - \frac{1}{2}\Bigg[1 - \frac{2C\left(\|\boldsymbol{\theta}\|+1\right)}{\|\xi_+\|^2} + \frac{C^2(\|\boldsymbol{\theta}\|+1)^2}{\|\xi_+\|^4} + 1 + \frac{2C\left(\|\boldsymbol{\theta}\|-1\right)}{\|\xi_-\|^2} + \\
&\qquad \frac{C^2(\|\boldsymbol{\theta}\|-1)^2}{\|\xi_-\|^4}\Bigg] \\
&= \frac{1}{2}\left(\frac{2C\left(\|\boldsymbol{\theta}\|+1\right)}{\|\xi_+\|^2} - \frac{C^2(\|\boldsymbol{\theta}\|+1)^2}{\|\xi_+\|^4} - \frac{2C\left(\|\boldsymbol{\theta}\|-1\right)}{\|\xi_-\|^2} - \frac{C^2(\|\boldsymbol{\theta}\|-1)^2}{\|\xi_-\|^4}\right)
\end{aligned}
$$

e

$$
\begin{aligned}
\mathscr{R}(\boldsymbol{\delta}_0,\boldsymbol{\theta}) - \mathscr{R}(,\delta_c,\boldsymbol{\theta}) &= (p-1) - \mathscr{R}^{(2)}_{|\xi_\pm} \\
&= (p-1) - \frac{1}{2}\left(L_+^{(2)} + L_-^{(2)}\right) \\
&= (p-1) - \frac{1}{2}\left[\left(1 - \frac{C}{\|\xi_+\|^2}\right)^2 (p-1) + \left(1 - \frac{C}{\|\xi_-\|^2}\right)^2 (p-1)\right] \\
&= (p-1) - \frac{1}{2}\Bigg[p - 1 - \frac{2C\,(p-1)}{\|\xi_+\|^2} + \frac{C^2\,(p-1)}{\|\xi_+\|^4} + \\
&\qquad p - 1 - \frac{2C\,(p-1)}{\|\xi_-\|^2} + \frac{C^2\,(p-1)}{\|\xi_-\|^4}\Bigg] \\
&= \frac{1}{2}(p-1)\left[\left(\frac{2C}{\|\xi_+\|^2} + \frac{2C}{\|\xi_-\|^2}\right) - \left(\frac{C^2}{\|\xi_+\|^4} + \frac{C^2}{\|\xi_-\|^4}\right)\right]
\end{aligned}
$$

A diferença entre os riscos destes estimadores é obtida pela soma da diferença entre os riscos na duas direções, que reordenada é dada por

$$
\begin{aligned}
\Delta_{|\xi_\pm} &= p - \mathscr{R}_{|\xi_\pm} \\
&= C\,\|\boldsymbol{\theta}\|\left(\frac{1}{\|\xi_+\|^2} - \frac{1}{\|\xi_-\|^2}\right) + Cp\left(\frac{1}{\|\xi_+\|^2} + \frac{1}{\|\xi_-\|^2}\right) \\
&\quad -\frac{1}{2}C^2\left(\frac{(\|\boldsymbol{\theta}\|+1)^2 + p - 1}{\|\xi_+\|^4} + \frac{(\|\boldsymbol{\theta}\|-1)^2 + p - 1}{\|\xi_-\|^4}\right) \\
&= C\,\|\boldsymbol{\theta}\|\left(\frac{1}{\|\xi_+\|^2} - \frac{1}{\|\xi_-\|^2}\right) + \left(Cp - \frac{1}{2}C^2\right)\left(\frac{1}{\|\xi_+\|^2} + \frac{1}{\|\xi_-\|^2}\right)
\end{aligned}
$$

lembrando que $\|\xi_\pm\|^2 = (\|\boldsymbol{\theta}\| \pm 1)^2 + p - 1$.

Observe que se $\|\xi_+\| = \|\xi_-\|$ a diferença entre os riscos seria positiva pra qualquer $\left(Cp - \frac{1}{2}C^2\right) > 0$, ou seja, $0 < C < 2p$. Em particular, é positiva para $p \geq 2$. o que sabemos não ser verdade para a diferença entre os riscos não condicionais. Mas, de fato $\|\xi_+\| > \|\xi_-\|$ portanto, $\frac{1}{\|\xi_+\|^2} - \frac{1}{\|\xi_-\|^2} < 0$.

$$\begin{aligned}
\frac{1}{\|\xi_+\|^2} - \frac{1}{\|\xi_-\|^2} &= \frac{\|\xi_-\|^2 - \|\xi_+\|^2}{\|\xi_+\|^2\|\xi_-\|^2} \\
&= \frac{(\|\boldsymbol{\theta}\| - 1)^2 + p - 1 - \left[(\|\boldsymbol{\theta}\| + 1)^2 + p - 1\right]}{\|\xi_+\|^2\|\xi_-\|^2} \\
&= -4\frac{\|\boldsymbol{\theta}\|}{\|\xi_+\|^2\|\xi_-\|^2} \\
\frac{1}{\|\xi_+\|^2} + \frac{1}{\|\xi_-\|^2} &= \frac{\|\xi_-\|^2 + \|\xi_+\|^2}{\|\xi_+\|^2\|\xi_-\|^2} \\
&= \frac{(\|\boldsymbol{\theta}\| - 1)^2 + p - 1 + \left[(\|\boldsymbol{\theta}\| + 1)^2 + p - 1\right]}{\|\xi_+\|^2\|\xi_-\|^2} \\
&= \frac{\|\boldsymbol{\theta}\|^2 - 2\|\boldsymbol{\theta}\| + 1 + p - 1 + \|\boldsymbol{\theta}\|^2 + 2\|\boldsymbol{\theta}\| + 1 + p - 1}{\|\xi_+\|^2\|\xi_-\|^2} \\
&= 2\frac{\|\boldsymbol{\theta}\|^2 + p}{\|\xi_+\|^2\|\xi_-\|^2}
\end{aligned}$$

Portanto, a diferença no risco condicional para $p \geq 2$ é

$$\begin{aligned}
\Delta_{|\xi_\pm} &= p - \mathscr{R}_{|\xi_\pm} \\
&= C\|\boldsymbol{\theta}\|\left(\frac{1}{\|\xi_+\|^2} - \frac{1}{\|\xi_-\|^2}\right) + \left(Cp - \frac{1}{2}C^2\right)\left(\frac{1}{\|\xi_+\|^2} + \frac{1}{\|\xi_-\|^2}\right) \\
&= C\|\boldsymbol{\theta}\|\left(-4\frac{\|\boldsymbol{\theta}\|}{\|\xi_+\|^2\|\xi_-\|^2}\right) + \left(Cp - \frac{1}{2}C^2\right)\left(2\frac{\|\boldsymbol{\theta}\|^2 + p}{\|\xi_+\|^2\|\xi_-\|^2}\right) \\
&= C\left(-4\frac{\|\boldsymbol{\theta}\|^2}{\|\xi_+\|^2\|\xi_-\|^2}\right) + \left(Cp - \frac{C^2}{2}\right)\left(2\frac{\left(\|\boldsymbol{\theta}\|^2 + p\right)}{\|\xi_+\|^2\|\xi_-\|^2}\right) \\
&= \frac{2}{\|\xi_+\|^2\|\xi_-\|^2}\left[\left(C(p-2) - \frac{C^2}{2}\right)\|\boldsymbol{\theta}\|^2 + \left(Cp - \frac{C^2}{2}\right)p\right] \\
&= \frac{2}{\|\xi_+\|^2\|\xi_-\|^2}\left[\left(C(p-2) - \frac{C^2}{2}\right)\|\boldsymbol{\theta}\|^2 + \left(Cp - 2C + 2C - \frac{C^2}{2}\right)p\right] \\
&> \frac{2\left(\|\boldsymbol{\theta}\|^2 + p\right)}{\|\xi_+\|^2\|\xi_-\|^2}\left(C(p-2) - \frac{C^2}{2}\right).
\end{aligned}$$

Para $p = 2$, a diferença entre os riscos condicionais é

$$\begin{aligned}\Delta_{|\xi_\pm} &= \frac{2}{\|\xi_+\|^2\|\xi_-\|^2}\left[\left(-\frac{C^2}{2}\right)\|\boldsymbol{\theta}\|^2 + \left(2C - \frac{C^2}{2}\right)2\right] \\ &= \frac{2}{\|\xi_+\|^2\|\xi_-\|^2}\left[\left(-\frac{C^2}{2}\right)\|\boldsymbol{\theta}\|^2 + 4C - C^2\right] \\ &= \frac{2}{\|\xi_+\|^2\|\xi_-\|^2}\left[\left(-\frac{C^2}{2}\right)(\|\boldsymbol{\theta}\| + 2) + 4C\right] \qquad (2.32)\end{aligned}$$

que pode não ser positivo. Para $p \geq 3$, a diferença entre os riscos condicionais será positiva se

$$\begin{aligned} C(p-2) - \frac{C^2}{2} &> 0 \\ (p-2) - \frac{C}{2} &> 0 \\ 0 < C &< 2(p-2). \end{aligned}$$

O valor máximo da diferença entre os riscos condicionais ocorrerá se $C = p - 2$, Figura 14. Em particular

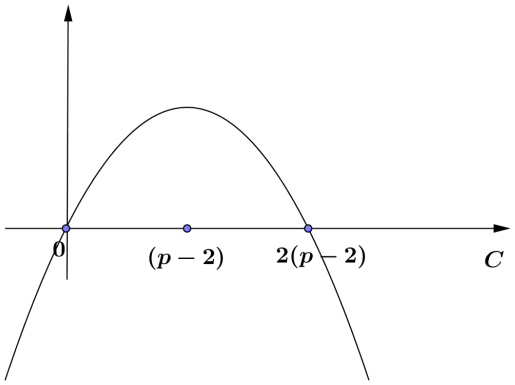

FIGURA 14: Representação gráfica da equação $C(p-2) - \frac{C^2}{2} = 0$, cujo valor máximo expressa a cota superior para a diferença entre os riscos condicionais com $p \geq 3$.

a diferença é positiva para $p > 3$ e $C = p - 1$, valor motivado pelo argumento geométrico apresentado na Figura 12. Por outro lado, a melhor escolha para a constante, em (2.32), é o valor ligeiramente menor $C = p - 2$. A melhoria é menor e isto pode ser considerado como uma penalidade necessária devido à aleatoriedade de $\mathbf{X}$. Se $\|\boldsymbol{\theta}\|^2$ não for muito próximo de zero tem-se

$$\|\xi\|^2 = \|\boldsymbol{\theta}\|^2 + p - 1 \approx \|\boldsymbol{\theta}\|^2 + p \approx \|\xi_+\|^2 \approx \|\xi_-\|^2.$$

Utilizando estas aproximações a diferença dos riscos condicionais de fato não depende do $\xi_\pm$

e tem-se

$$\begin{aligned}\Delta_{|\xi_\pm} &> \frac{2\left(\|\boldsymbol{\theta}\|^2+p\right)}{\|\xi_+\|^2\|\xi_-\|^2}\left(C(p-2)-\frac{C^2}{2}\right)\\ &\approx \frac{2\left(\|\boldsymbol{\theta}\|^2+p\right)}{\left(\|\boldsymbol{\theta}\|^2+p\right)\left(\|\boldsymbol{\theta}\|^2+p\right)}\left(C(p-2)-\frac{C^2}{2}\right)\\ &= \frac{2}{\left(\|\boldsymbol{\theta}\|^2+p\right)}\left(C(p-2)-\frac{C^2}{2}\right)\end{aligned}$$

como a diferença entre os riscos condicionais não depende de ξ, temos uma boa aproximação para a diferença entre os riscos não condicionais

$$\begin{aligned}\Delta &= R(\boldsymbol{\theta},\boldsymbol{\delta}_0)-R(\boldsymbol{\theta},\boldsymbol{\delta}_C)\\ &\approx \frac{2}{\|\boldsymbol{\theta}\|^2+p}\left(C(p-2)-\frac{C^2}{2}\right)>0.\end{aligned}$$

Portanto, toda esta construção geométrica é uma motivação heurística para a não admissibilidade do estimador $\boldsymbol{\delta}_0$ para $p \geq 3$. Esta aproximação melhora quanto maior for o valor da $\|\theta\|$, uma vez que se tem uma maior liberdade para se escolher ξ_+ e ξ_- próximo de ξ.

2.10 Estimador de James-Stein com encolhimento na direção de um vetor arbitrário

Seja $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)$ um vetor arbitrário que pode ser considerado como um*a priori* para o verdadeiro vetor de médias $\boldsymbol{\theta}$, tem-se, então, uma versão do estimador de James-Stein dado por

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu} + \left(1 - \frac{p-2}{\|\mathbf{X}-\boldsymbol{\mu}\|^2}\right)(\mathbf{X}-\boldsymbol{\mu}).$$

Esses resultados são dos trabalhos de Stein (1956), James e Stein (1961) e Lindley (1962). A interpretação geométrica desse estimador está exposta na Figura 15.

Uma observação trivial é que ao se fazer um encolhimento da forma αv uma esfera de raio r centrada em $\boldsymbol{\theta}$ se transforma em uma esfera de raio αr centrada em $\alpha\theta$ como na Figura 16.

A demonstração é trivial pois

$$\|\alpha\mathbf{X} - \alpha\boldsymbol{\theta}\|^2 = |\alpha|^2\|\mathbf{X}-\boldsymbol{\theta}\|^2 = |\alpha|^2 r^2 = (\alpha r)^2$$

De fato, esta propriedade é válida quando o encolhimento se dá na direção de um vetor qualquer $\boldsymbol{\mu}$.

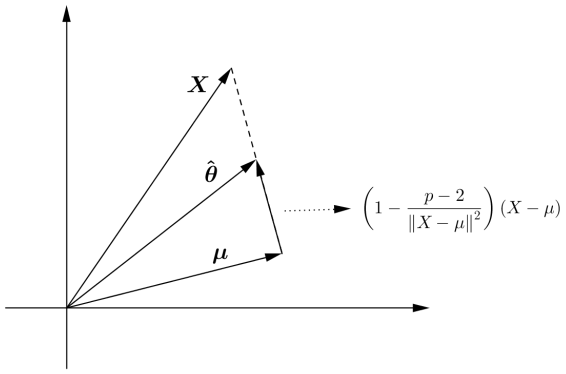

FIGURA 15: Encolhimento na direção de um vetor arbitrário $\boldsymbol{\mu}$.

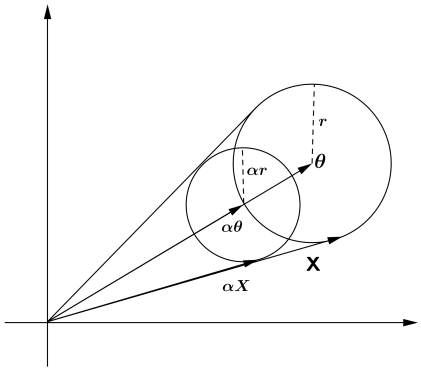

FIGURA 16: Encolhimento na direção da origem.

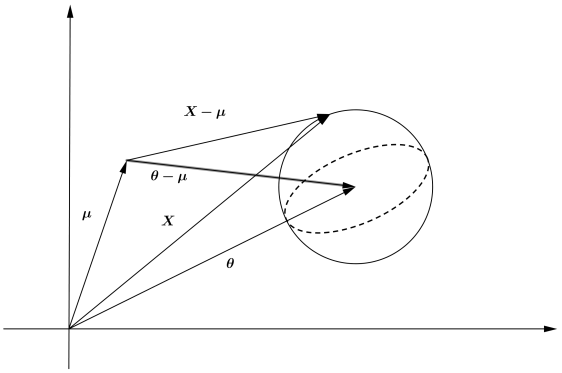

FIGURA 17: A geometria do encolhimento na direção do vetor $\boldsymbol{\mu}$.

De acordo com as Figuras 17 e 18

$$\boldsymbol{\mu} + \alpha\left(\boldsymbol{\theta} - \boldsymbol{\mu}\right) = \left(1 - \alpha\right)\boldsymbol{\mu} + \alpha\boldsymbol{\theta}$$

$$r^2 = \|\mathbf{X} - \boldsymbol{\theta}\|^2 = \|(\mathbf{X} - \boldsymbol{\mu}) - (\boldsymbol{\theta} + \boldsymbol{\mu})\|^2.$$

Logo, o encolhimento na direção de $\boldsymbol{\mu}$ satisfaz

$$\begin{aligned}\|\alpha\left(\mathbf{X} - \boldsymbol{\mu}\right) - \alpha\left(\boldsymbol{\theta} - \boldsymbol{\mu}\right)\|^2 &= (\alpha)^2\|(\mathbf{X} - \boldsymbol{\mu}) - (\boldsymbol{\theta} - \boldsymbol{\mu})\|^2 \\ &= (\alpha)^2\|\mathbf{X} - \boldsymbol{\theta}\|^2 \\ &= (\alpha)^2 r^2 \\ &= (\alpha r)^2\end{aligned}$$

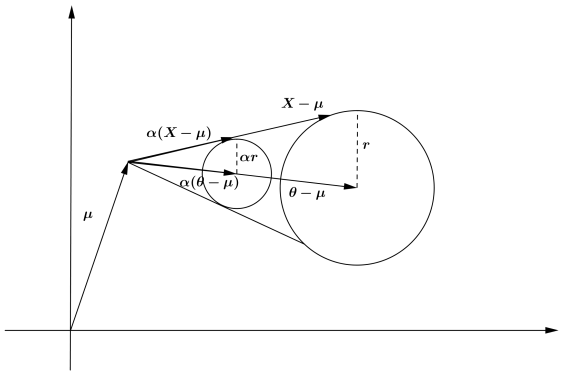

FIGURA 18: Encolhimento preserva a circunferência.

2.10.1 Estimador de James-Stein com encolhimento na direção do vetor $\vec{\mathbf{1}} = (1, ..., 1)$

Se $\mathbf{X} = (X_1, ..., X_p)$ seja $\bar{\mathbf{X}} = \left(\frac{\sum_{i=1}^{p} X_i}{p}, ..., \frac{\sum_{i=1}^{p} X_i}{p}\right)$

$$\hat{\boldsymbol{\theta}} = \bar{\mathbf{X}} + \left(1 - \frac{p-3}{\left\|\mathbf{X} - \bar{\mathbf{X}}\right\|^2}\right)\left(\mathbf{X} - \bar{\mathbf{X}}\right) \tag{2.33}$$

sugerido por Lindley (1962). Veja a Figura 19

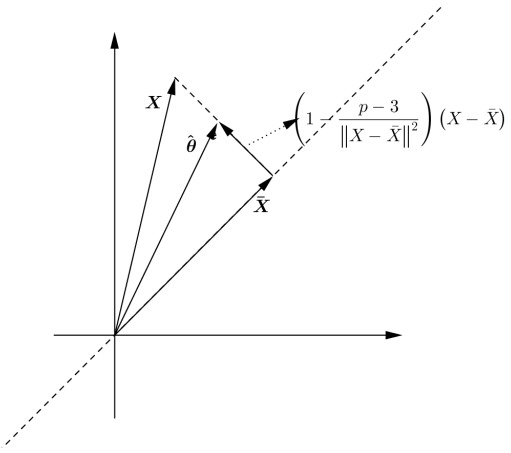

FIGURA 19: Encolhimento na direção do vetor $\bar{X}$.

2.11 Justificativas heurísticas para o fator de encolhimento

Ao longo do tempo, várias justificativas heurísticas, utilizando a geometria foram desenvolvidas no sentido de se ter uma compreensão intuitiva do estimador de James e Stein e de outros estimadores de encolhimento. De fato, o próprio Stein se utilizou desse recurso quando apresentou, pela primeira vez, seus resultados. Nesta seção, algumas novas justificativas são apresentadas sempre no sentido de ressaltar a compreensão intuitiva do fenômeno do melhor desempenho de estimadores obtidos por encolhimento.

Justificativa 1: Abordaremos o fato que levou Stein a desenvolver sua teoria. Como abordado na seção 3, se $\mathbf{X}$ é uma variavel aleatória p-dimensional com $E[\mathbf{X}] = \boldsymbol{\theta}$ e $\mathrm{cov}[\mathbf{X}] = \sigma^{\mathbf{2}}\mathbf{I}$ então, $E\left[\|\mathbf{X}\|^2\right] = \|\boldsymbol{\theta}\|^2 + p\sigma^2$. Tem-se, então, que, apesar de cada componente X_i estimar θ_i de forma não viesada, isto é, pode-se supor que as estimativas $X_i = x_i$ estejam próximas das componentes θ_i para todo i, $\sum_{i=1}^{p} X_i^2$ não é uma boa estimativa de $\|\boldsymbol{\theta}\|^2$, de fato $\sum_{i=1}^{p} X_i^2$ tem tendência a superestimar $\|\boldsymbol{\theta}\|^2$.

Para o caso unidimensional, pode-se argumentar o seguinte: vamos supor por simplicidade que a densidade $f_X(x)$ da variável aleatória X seja simétrica. Os pontos $x_1 = \theta - h$ e $x_2 = \theta + h$ tal que $(\theta - h)^2 = x_1^2 << x_2^2 = (\theta + h)^2$ como pode ser claramente visto no gráfico da função quadrática $g(x) = x^2$ (Figura 20).

Fixado h, a média do valor observado entre estes dois pontos é dada pela média ponderada

$$\begin{aligned} f_X(\theta+h)(\theta+h)^2 + f_X(\theta-h)(\theta-h)^2 &= f_X(\theta+h)\left[(\theta+h)^2 + (\theta-h)^2\right] \\ &= f_X(\theta+h)\left[2\theta^2 + 2h^2\right]. \end{aligned}$$

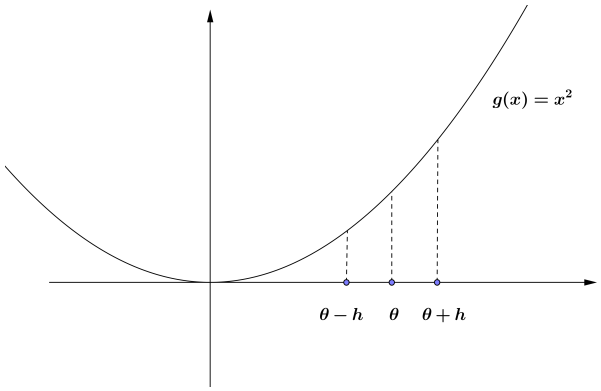

FIGURA 20: Gráfico da função quadrática $g(x) = x^2$.

Pode-se agora calcular o valor esperado de X^2 da forma

$$\begin{aligned} E\left[X^2\right] &= 2\int_0^\infty f_X(\theta+h)\left[\theta^2+h^2\right]dh \\ &= 2\int_\theta^\infty f_X(u)\left[\theta^2+(u-\theta)^2\right]du \qquad (u=\theta+h) \\ &= 2\theta^2\int_\theta^\infty f_X(u)du + 2\int_\theta^\infty f_X(u)(u-\theta)^2du \\ &= \theta^2 + 2\frac{1}{2}\int_{-\infty}^\infty f_X(u)(u-\theta)^2du \\ &= \theta^2+\sigma^2. \end{aligned}$$

Esta é a ideia probabilística para justificar, de forma intuitiva, o fato de X^2 superestimar θ.

Justificativa 2: Considerando o caso $p \geq 2$, uma justificativa probabilística para o caso normal é a seguinte. Se $\mathbf{X} \sim \mathbf{N_p}\left(\boldsymbol{\theta}, \sigma^2\mathbf{I}\right)$, considere em $\mathbb{R}^p$ uma base ortonormal em que um dos vetores desta base é $\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$. Por uma transformação ortogonal B, é possível transformar esta base na base canônica $e_1, e_2, ..., e_p$ tal que o vetor $\frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$ é levado no vetor e_1. Se $\mathbf{Y} = (Y_1, ..., Y_p) = B\mathbf{X}$ então $\mathbf{Y} \sim \mathbf{N_p}\left(\|\boldsymbol{\theta}\|\,\mathbf{e_1}, \sigma^2\mathbf{I}\right)$ e $\|\mathbf{Y}\|^2 = \|\mathbf{X}\|^2$. Tem-se marginais, $Y_1 \sim N\left(\|\boldsymbol{\theta}\|, \sigma^2\right)$ e $Y_i \sim N\left(0, \sigma^2\right)$, $i = 2, ..., p$.

Portanto, como $\frac{Y_1}{\sigma} \sim N\left(\frac{\|\boldsymbol{\theta}\|}{\sigma}, 1\right)$, $\left(\frac{Y_1}{\sigma}\right)^2$ tem distribuição qui-quadrado não central com 1 grau de

liberdade e parâmetro de não centralidade $\frac{\|\boldsymbol{\theta}\|^2}{\sigma^2}$. Neste caso

$$\begin{aligned} E\left[\|\mathbf{X}\|^2\right] &= E\left[\|\mathbf{Y}\|^2\right] \\ &= E\left[Y_1^2\right] + E\left[Y_2^2\right] + ... + E\left[Y_p^2\right] \\ &= \sigma^2 E\left[\left(\frac{Y_1}{\sigma}\right)^2\right] + \sigma^2 + ... + \sigma^2 \\ &= \sigma^2\left[1 + \frac{\|\boldsymbol{\theta}\|^2}{\sigma^2}\right] + (p-1)\,\sigma^2 \\ &= \|\boldsymbol{\theta}\|^2 + p\sigma^2. \end{aligned}$$

A ideia é utilizar tal fato para se ter uma informação para a quantidade de encolhimento necessário. Seja $0 < \alpha < 1$ e se quer saber o valor de α tal que $E\left[\|\alpha\mathbf{X}\|\right]$ seja igual a $\|\boldsymbol{\theta}\|^2$. Tem-se

$$\begin{aligned} \|\boldsymbol{\theta}\|^2 &= E\left[\|\alpha\mathbf{X}\|^2\right] \\ \|\boldsymbol{\theta}\|^2 &= \alpha^2 E\left[\|\mathbf{X}\|^2\right] \\ \|\boldsymbol{\theta}\|^2 &= \alpha^2\left(\|\boldsymbol{\theta}\|^2 + p\sigma^2\right) \\ \alpha^2 &= \frac{\|\boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2 + p\sigma^2} \\ \alpha &= \sqrt{\frac{\|\boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2 + p\sigma^2}} \end{aligned}$$

utilizando a aproximação $a = \sqrt{b} \simeq b$ quando $0 \leq b < 1$, tem-se

$$\begin{aligned} \alpha &\simeq \frac{\|\boldsymbol{\theta}\|^2}{\|\boldsymbol{\theta}\|^2 + p\sigma^2} \\ &= \frac{\|\boldsymbol{\theta}\|^2 + p\sigma^2 - p\sigma^2}{\|\boldsymbol{\theta}\|^2 + p\sigma^2} \\ &= 1 - \frac{p\sigma^2}{\|\boldsymbol{\theta}\|^2 + p\sigma^2}. \end{aligned}$$

Utilizando $\|\mathbf{X}\|^2$ como estimador de $\|\boldsymbol{\theta}\|^2 + p\sigma^2$ tem-se que

$$\alpha \approx \left(1 - \frac{p\sigma^2}{\|\mathbf{X}\|^2}\right).$$

Este valor é semelhante ao coeficiente do estimador de James-Stein.

Justificativa 3: Dado um vetor de médias $\boldsymbol{\theta}$ e uma "nuvem" de pontos equidistantes de $\boldsymbol{\theta}$ representados por uma esfera conforme a Figura 21 . Uma realização $\mathbf{X} = \mathbf{x}$ ocorre com igual probabilidade em qualquer ponto desta esfera.

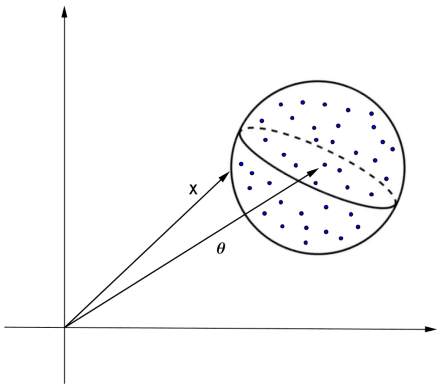

FIGURA 21: Nuvem de dados ao redor do vetor de médias $\boldsymbol{\theta}$.

Os pontos amostrais $\mathbf{x}$ que possuem norma menor ou igual a $\|\boldsymbol{\theta}\|$ formam uma calota esférica que possui área menor que a área do seu complementar na esfera. Isto pode ser visualizado construindo uma nova esfera com raio $\|\boldsymbol{\theta}\|$ e fazendo-se a interseção dessas duas esferas, com raios $\|\boldsymbol{\theta}\|$ e $R < \|\boldsymbol{\theta}\|$, que gera uma calota esférica, como na Figura 22.

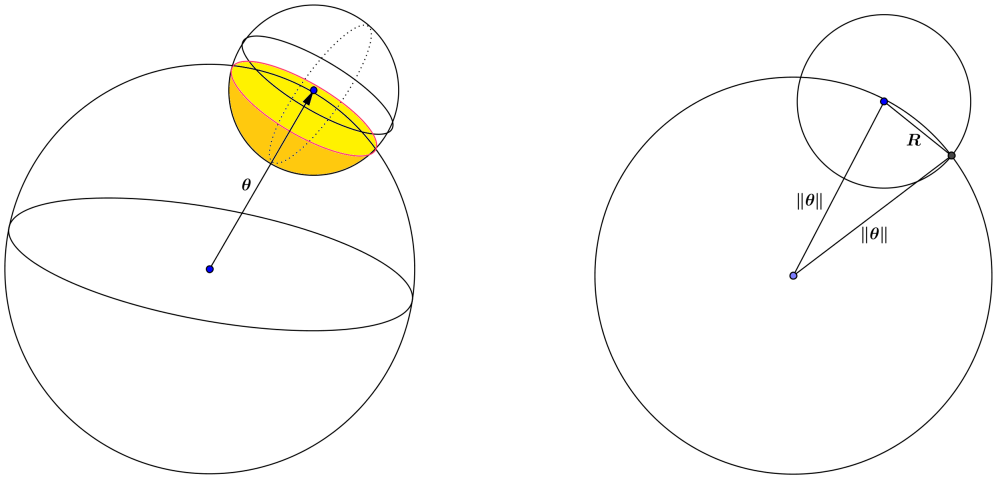

FIGURA 22: Interseção entre as esferas.

Como esses pontos são equiprováveis, é mais verossímil observar $\mathbf{x}$ com $\|\mathbf{x}\| > \|\boldsymbol{\theta}\|$ do que $\|\mathbf{x}\| < \|\boldsymbol{\theta}\|$. Este fato justifica heuristicamente que $E\left[\|\mathbf{X}\|^2\right] > \|\boldsymbol{\theta}\|^2$. Vejamos a consequência dessa construção. A geometria nos auxilia a visualizar que valor observado $\mathbf{X} = \mathbf{x}$ como estimador de $\boldsymbol{\theta}$ pode ser melhorado utilizando o processo de encolhimento. Considere a esfera, isto é, a "nuvem de pontos" em uma região que

não contém a origem. Se encolhermos esses pontos, quais pontos da esfera traríamos para mais perto do centro da esfera? A resposta é que mais da metade dos pontos da esfera ficarão mais próximos do centro. Visualizando em duas dimensões, podemos observar que os pontos que estiverem pontilhados na esfera irão se afastar com o encolhimento na direção da origem, mas os pontos que estão além da linha pontilhada irão se aproximar (HARRIS, 2012).

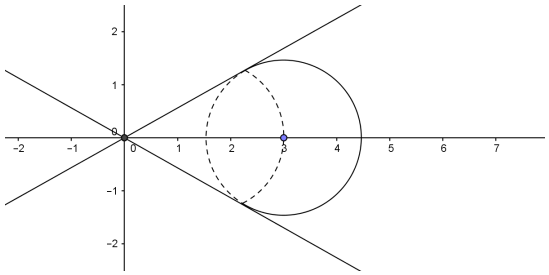

FIGURA 23: Encolhimento para $\|\boldsymbol{\theta}\| = 3$.

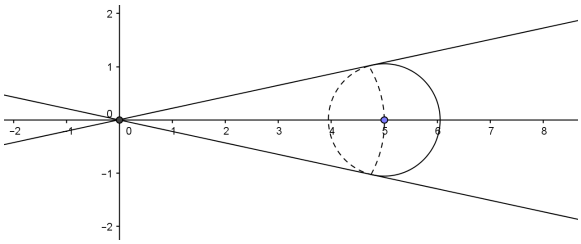

FIGURA 24: Encolhimento para $\|\boldsymbol{\theta}\| = 5$.

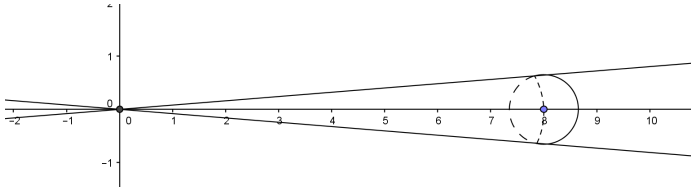

FIGURA 25: Encolhimento para $\|\boldsymbol{\theta}\| = 8$.

Para calcularmos qual a proporcionalidade entre as áreas das calotas resultantes da interseção das duas esferas, utilizaremos geometria plana. Tem-se que área da calota esférica=$2\pi Rh$, conforme Figura 26. Estabelecendo a seguinte notação área da calota=AC, área total da esfera=ATE e área complementar da calota=ACC.

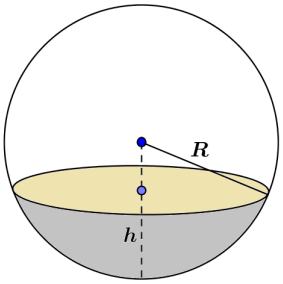

FIGURA 26: Calota resultante da interseção entre as esferas.

$$\begin{aligned}\frac{\text{AC}}{\text{ATE}} &= \frac{\text{AC}}{\text{AC}+\text{ACC}} \\ &= \frac{\dfrac{\text{AC}}{\text{ACC}}}{\dfrac{\text{AC}}{\text{ACC}}+1} \\ &= \frac{\alpha}{\alpha+1} \qquad \text{sendo} \qquad \alpha = \frac{\text{AC}}{\text{ACC}}.\end{aligned} \tag{2.34}$$

Como para levarmos em conta o volume da nuvem de pontos vamos propor como coeficiente de encolhimento o quadrado da razão entre as áreas da calota e do seu complementar. Logo, temos que

$$\frac{\text{AC}}{\text{ATE}} = \frac{2\pi Rh}{4\pi R^2} = \frac{h}{2R} \tag{2.35}$$

Igualando (2.34) e (2.35)

$$\begin{aligned}\frac{h}{2R} &= \frac{\alpha}{\alpha+1} \\ h\,(\alpha+1) &= 2R\alpha \\ \alpha\,(2R-h) &= h \\ \alpha &= \frac{h}{(2R-h)}\end{aligned} \tag{2.36}$$

Considerando a calota destacada na Figura 26, temos a representação da seção da calota esférica na Figura 27.

Para o triângulo ABC, em destaque na Figura 27, temos

$$\|\boldsymbol{\theta}\| - l = \|\boldsymbol{\theta}\|\cos\gamma \tag{2.37}$$

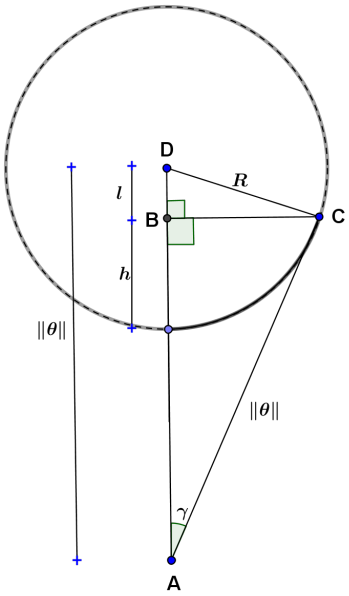

FIGURA 27: Representação geométrica da seção da calota esférica.

Para determinarmos o $cos(\gamma)$, consideremos o triângulo isósceles ADC, conforme Figura 28. Logo,

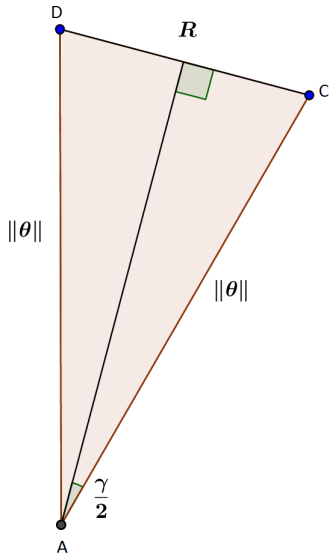

FIGURA 28: Representação do triângulo isósceles.

$$\begin{aligned} \frac{R}{2} &= sen\left(\frac{\gamma}{2}\right)\|\boldsymbol{\theta}\| \\ sen\left(\frac{\gamma}{2}\right) &= \frac{R}{2\|\boldsymbol{\theta}\|}. \end{aligned}$$

Aplicando identidades trigonométricas temos que

$$\begin{aligned} sen\left(\gamma\right) &= 2sen\left(\frac{\gamma}{2}\right)\cos\left(\frac{\gamma}{2}\right) \\ &= 2\frac{R}{2\left\|\boldsymbol{\theta}\right\|}\sqrt{1-sen^2\left(\frac{\gamma}{2}\right)} \\ &= \frac{R}{\left\|\boldsymbol{\theta}\right\|}\sqrt{1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}} \end{aligned}$$

e, ainda,

$$\begin{aligned} \cos^2\left(\gamma\right) &= 1-sen^2\left(\gamma\right) \\ &= 1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right). \end{aligned}$$

Retornando para a Equação (2.37)

$$\begin{aligned} \left\|\boldsymbol{\theta}\right\| - l &= \left\|\boldsymbol{\theta}\right\|\cos\gamma \\ \left\|\boldsymbol{\theta}\right\| - l &= \left\|\boldsymbol{\theta}\right\|\sqrt{1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right)} \\ l &= \left\|\boldsymbol{\theta}\right\|\left(1-\sqrt{1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right)}\right). \end{aligned}$$

Sabendo que

$$\begin{aligned} h &= R-l \\ h &= R-\left\|\boldsymbol{\theta}\right\|+\left\|\boldsymbol{\theta}\right\|\sqrt{1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right)} \end{aligned}$$

e de acordo com a Equação (2.36)

$$\begin{aligned} \alpha &= \frac{h}{(2R-h)} \\ &= \frac{R-\left\|\boldsymbol{\theta}\right\|+\left\|\boldsymbol{\theta}\right\|\sqrt{1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right)}}{2R-R+\left\|\boldsymbol{\theta}\right\|-\left\|\boldsymbol{\theta}\right\|\sqrt{1-\frac{R^2}{\left\|\boldsymbol{\theta}\right\|^2}\left(1-\frac{R^2}{4\left\|\boldsymbol{\theta}\right\|^2}\right)}} \end{aligned}$$

utilizando aproximação por série de Taylor para a função $f(x) = \sqrt{(1+x)}$ temos que $f(x) = f(0) +$

$f'(0)x = 1 + \frac{1}{2}x$ segue que

$$\begin{aligned}\alpha &= \frac{R - \|\boldsymbol{\theta}\| + \|\boldsymbol{\theta}\|\left(1 - \frac{1}{2}\frac{R^2}{\|\boldsymbol{\theta}\|^2}\right)}{R + \|\boldsymbol{\theta}\| - \|\boldsymbol{\theta}\|\left(1 - \frac{1}{2}\frac{R^2}{\|\boldsymbol{\theta}\|^2}\right)} \\ &= \frac{R - \frac{R^2}{2\|\boldsymbol{\theta}\|}}{R + \frac{R^2}{2\|\boldsymbol{\theta}\|}}.\end{aligned}$$

Para associar o estimador de James-Stein, cujo fator de encolhimento é do tipo $(1 - \beta)$, com a proporção α temos que

$$\begin{aligned}(1-\beta) &= \alpha \\ \beta &= 1 - \alpha \\ &= 1 - \frac{R - \frac{R^2}{2\|\boldsymbol{\theta}\|}}{R + \frac{R^2}{2\|\boldsymbol{\theta}\|}} \\ &= \frac{R + \frac{R^2}{2\|\boldsymbol{\theta}\|} - R + \frac{R^2}{2\|\boldsymbol{\theta}\|}}{R + \frac{R^2}{2\|\boldsymbol{\theta}\|}} \\ &= \frac{\frac{R^2}{\|\boldsymbol{\theta}\|}}{R + \frac{R^2}{2\|\boldsymbol{\theta}\|}} \qquad \text{considere } \frac{R^2}{2\,\|\boldsymbol{\theta}\|} \text{ desprezível} \\ &\approx \frac{\frac{R^2}{\|\boldsymbol{\theta}\|}}{R} \\ &= \frac{R}{\|\boldsymbol{\theta}\|}.\end{aligned}$$

Comparando o fator de encolhimento do estimador de James-Stein

$$\hat{\boldsymbol{\theta}}_{JS} = \left(1 - \frac{\sigma^2}{n}\frac{p-2}{\|\mathbf{X}\|^2}\right)\mathbf{X},$$

através da dedução heurística, encontramos um fator de encolhimento

$$\alpha = (1 - \beta) = \left(1 - \frac{R}{\|\boldsymbol{\theta}\|}\right).$$

3 CONCLUSÃO

- A abordagem geométrica do estimador de James-Stein é uma ferramenta útil na compreensão de suas propriedades analíticas. Na análise do estimador de James-Stein, como um caso particular de estimadores obtidos por encolhimento(shrinkage) como, por exemplo, estimadores esfericamente simétricos, descreveu-se detalhadamente seus aspectos analíticos e geométricos.

- A interpretação heurística do estimador de James-Stein é abordada geometricamente e busca uma interpretação para o fator de encolhimento que heuristicamente se assemelha ao fator de encolhimento do estimador de James-Stein.

REFERÊNCIAS

BRANDWEIN,A.C.; STRAWDERMAN,W.E. Stein Estimation: The Spherically Symmetric Case. **Statistical Science.** v. 5, n. 3, p. 356-369, 1990.

BRANDWEIN,A.C.; STRAWDERMAN,W.E. Stein Estimation for Spherically Symmetric Distributions: Reent Developments.**Statistical Science.** v. 27, n. 1, p. 11-23, 2012.

BROWN,L.D.; ZHAO,L.H. A geometrical explanation of Stein shirinkage. **Statistical Science.** v. 27, n. 1, p. 24-30, 2012.

CASELLA, G. An Introduction to Empirical Bayes Data Analysis. **The American Statistician**, v.39, n.2,p. 83-87, 1985.

CASELLA,G.; BERGER,R.L. **Inferência Estatística**. São Paulo: Cengage Learning, 2010. 588 p.

COSTA, L.A. **Novo estimador de cumeeeira de Rao com aplicações em seleção genômica**. 2015. 126 p. Tese (Doutorado em Estatística e Experimentação Agropecuária)- Universidade Federal de Lavras, Lavras. 2015.

EFRON, B.; MORRIS, C. Stein's Estimation Rule and Its Competitors-An Empirical Bayes Approaximation. **Journal of the American Statistical Association**, v. 68, p. 117-130, 1973.

EFRON, B.; MORRIS, C. Data analysis using Stein's estimator and its generalizations. **Journal of the American Statistical Association**, v. 70, p. 311-319, 1975.

EFRON, B.; MORRIS,C. Stein's Paradox in Statistics. **Scientific American**, v.5, p.119-127, 1977.

ELERHS,R.S. **Introdução a Inferência Bayesiana**, 2014. Disponível em: <http://icmc.usp.br/ ehlers/bayes/>. Acesso em: 10 de fevereiro de 2014.

FERREIRA,D.F. **Estatística Multivariada**. Lavras: Ed. UFLA, 2011. 676 p.

GEOGEBRA. **Organization International GeoGebra Institute, Linz, Austria, 2001**. Disponível em: <http://http://www.geogebra.org>. Acesso em: 03 de junho de 2013.

GRUBER,M.H.J. **Improving Efficiency by Shririnkage**. New York:Marcell Dekker, 1998. 632 p.

HARRIS, N. **Visualizing the James-Stein estimator**. 2012. Disponível em: <http://www.naftaliharris.com/blog/steinviz/>. Acesso em: 13 março de 2013.

HELMS, L. **Introduction a potencial theory**. New York: Wiley, 1969.

HOERL, A.E.; KENNARD,R.W. Ridge regression: biased estimation for nonorthogonal problems. **Technometrics**, v.12, p.55-67, 1970a.

HOERL, A.E.; KENNARD,R.W. Ridge regression, applications to nonorthogonal problems. **Technometrics**, v.12, p.69-82, 1970b.

HOERL, A.E.; KENNARD,R.W., and BALDWIN,K.F. Ridge regression: some simulations. **Communications in Statisticas: Theory and Methods**, v.4, n.2, p.105-123, 1975.

JAMES, W.; STEIN, C. Estimation with quadratic loss.**Proceedings of the Fourth Berkeley Symposium on Mathematics and Statistics. Berkeley**:University of California Press 1, 1961, v.1, p.361-380.

JAMES-STEIN ESTIMATOR. Disponível em: <http://radhakrisna.typedad.com/james-stein-estimator .pdf>. Acesso em: 23 de outubro de 2014.

LAWLESS, J.F.; WANG, P. A simulation study of ridge and other regression estimators. **Communications in Statistics- Theory e Methods**, v.5, n.4, p.307-323, 1976.

LEHMANN, E.L.; CASELLA, G. **Theory of Point Estimation**. New York: Springer-Verlag, 1998.

LINDLEY, D.V. Discussion of "Confidence sets for the mean of a multivariate normal distribution,"by C.M. Stein. **Journal Royal Statistical Society**, v.24, p.285-287, 1962.

MOOD. A.M.; GRAYBILL, F.A.; BOES. B.C. **Introduction to the Theory of Statistics**. New York: McGraw Hill, 1974.

RENCHER,A.C.; SCHAALJE,G.B. **Linear models in statistics**. New Jersey: J. Wiley & Sons, 2008. 672 p.

SMALL, D. **Statistics 550 Notes 17**, 2014. Disponível em:<www-stat.wharton.upenn.edu/ dsmall/stat550 >. Acesso em: 12 de janeiro de 2015.

STEIN, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. **In Proc. Third Berkeley Sympos. Math. Statist. Probab.**, Univ. California Press, Berkeley, p. $1954 - 1955$, 1956.

STEIN, C. Confidence sets for the mean of a multivariate normal distribution. **Journal Royal Statistical Society**, v.24, p.285-287, 1962.

STEIN, C. Estimation of the mean of a multivariate normal distribution. **In Proceedings of the Prague Symposium on Asymptotic Statistics (Charles Univ., Prague)**, v. II, p. 345-381, 1973.

STEIN, C. Estimation of the mean of a multivariate normal distribution. **The Annals of Statistics**, v.9, n.6, p.1135-1151, 1981.

STIGLER, S. M. The 1988 Neyman memorial lecture: A Galtonian perspective on shrinkage estimators. **Statistical Science**, v.5, p.$147 - 155$, 1990.

APÊNDICES

APÊNDICE A - Teorema de Bayes e casos para a normal

Considere uma quantidade de interesse desconhecida θ . A informação de que dispomos sobre θ, resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ. A distribuição amostral $p(x|\theta)$ define esta relação. A ideia de que após observar $X = x$, a quantidade de informação sobre θ aumenta é bastante intuitiva e, o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação (EHLERS, 2014).

$$p(\theta|x) = \frac{p(\theta; x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta; x)d\theta}. \tag{3.1}$$

Note que $\frac{1}{p(x)}$, que não depende de θ, funciona como uma constante normalizadora de $p(x|\theta)$. Para um valor fixo de x, a função $l(\theta; x) = p(x|\theta)$ fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de θ enquanto $p(\theta)$ é chamada distribuição *a priori* de θ. Essas duas fontes de informação, priori e verossimilhança, são combinadas levando à distribuição *a posteriori* de θ, $p(\theta|x)$. Assim, a forma usual do teorema de Bayes é

$$p(\theta|x) \propto l(\theta; x)p(\theta).$$

Note que, ao omitir o termo $p(x)$, a igualdade em (3.1) foi substituída por uma proporcionalidade. Essa forma simplificada do teorema de Bayes será útil em problemas que envolvam estimação de parâmetros já que o denominador é apenas uma constante normalizadora. A constante normalizadora d*a posteriori* pode ser facilmente recuperada pois $p(\theta|x) = kp(x|\theta)p(\theta)$ onde

$$k^{-1} = \int p(x|\theta)p(\theta)d\theta = E_\theta[p(X|\theta)] = p(x)$$

é chamada distribuição preditiva.

$\theta|x) = \frac{k(x)h(x,\theta)}{g(x)}$

1) **Caso Normal com σ^2 conhecido**:

a) Considere a observação de uma variável aleatória $Y = (Y_1, ..., Y_n)$ com distribuição $Y \sim N\left(\theta, \sigma^2\right)$ com σ^2 conhecido e admita-se *a priori* $\theta \sim N\left(0, \tau^2\right)$. Para determinar a distribuição *a posteriori* pelo teorema de Bayes, temos que

$$p(\theta|y) \propto l(y|\theta)p(\theta).$$

A função de densidade de y é dada por

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\theta)^2\right\}.$$

e sua função de verossimilhança

$$l(y|\theta) = \prod_{i=1}^{n} f(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}.$$

A densidade da priori é dada por

$$p(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\theta)^2\right\}.$$

Portanto,

$$\begin{aligned}
p(\theta|y) &\propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\} \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(\theta)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2 - \frac{1}{2\tau^2}\theta^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{\sum_{i=1}^{n}(y_i-\theta)^2}{\sigma^2} + \frac{\theta^2}{\tau^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}y_i^2 - 2\theta\sum_{i=1}^{n}y_i + n\theta^2\right) + \frac{\theta^2}{\tau^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right) - \frac{2\theta\sum_{i=1}^{n}y_i}{\sigma^2}\right]\right\} \exp\left\{-\frac{\sum_{i=1}^{n}y_i^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)\left[\theta^2 - \frac{2\theta\sum_{i=1}^{n}y_i}{\sigma^2\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)}\right]\right\}
\end{aligned}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)\left[\left(\theta-\frac{\sum_{i=1}^{n} y_i}{\sigma^2\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)}\right)^2-\left(\frac{\sum_{i=1}^{n} y_i}{\sigma^2\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)}\right)^2\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)\left[\left(\theta-\frac{\sum_{i=1}^{n} y_i \Big/ n}{\frac{\sigma^2}{n}\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)}\right)^2\right]\right\}\exp\left\{\frac{1}{2}\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)\left(\frac{\sum_{i=1}^{n} y_i}{\sigma^2\left(\frac{\sigma^2+n\tau^2}{\sigma^2\tau^2}\right)}\right)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\left(\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right)}\left[\left(\theta-\frac{\tau^2}{\left(\frac{\sigma^2}{n}+\tau^2\right)}\bar{y}\right)^2\right]\right\}.$$

Logo,

$$\theta|y \sim N\left(\frac{\tau^2}{\left(\frac{\sigma^2}{n}+\tau^2\right)}\bar{y}, \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right).$$

b) Considere a observação de uma variável aleatória $Y=(Y_1,...,Y_n)$ tal que $Y|\alpha \sim N(\alpha,\sigma^2)$ com σ^2 conhecido e admita-se uma priori $\alpha \sim N(0,\sigma_\alpha^2)$. Observando que Y depende de α, para determinar a distribuição de Y, utiliza-se o conceito de mistura, que é dado por:

$$\begin{aligned}
f(y) &= \int f(y|\alpha)g(\alpha)d\alpha \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}}exp\left\{-\frac{1}{2\sigma^2}(y-\alpha)^2\right\}\frac{1}{\sqrt{2\pi\sigma_\alpha^2}}exp\left\{-\frac{1}{2\sigma_\alpha}\alpha^2\right\}d\alpha \\
&= \frac{1}{\sqrt{2\pi}\sigma\sigma_\alpha}\int \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2\sigma}\left(y^2-2y\alpha+\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\alpha^2\right\}d\alpha \\
&= \frac{1}{\sqrt{2\pi}\sigma\sigma_\alpha}\int \frac{1}{\sqrt{2\pi}}exp\left\{\alpha^2\left[\left(-\frac{1}{2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)\right]+\frac{2y\alpha}{2\sigma^2}-\frac{1y^2}{2\sigma^2}\right\}d\alpha \\
&= \frac{1}{\sqrt{2\pi}\sigma\sigma_\alpha}\exp\left\{-\frac{1y^2}{2\sigma^2}\right\}\int \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{1}{\left(\frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha}\right)}\left[\alpha^2-\frac{2y^2\sigma_\alpha^2}{(\sigma^2+\sigma_\alpha)}\right]\right\}d\alpha \\
&= \frac{1}{\sqrt{2\pi}\sigma\sigma_\alpha}\exp\left\{-\frac{1y^2}{2\sigma^2}\right\}\int \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{1}{\left(\frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha}\right)}\right. \\
&\qquad \left.\left\{\left[\alpha-\frac{y^2\sigma_\alpha^2}{(\sigma^2+\sigma_\alpha)}\right]^2-\left(\frac{y^2\sigma_\alpha^2}{(\sigma^2+\sigma_\alpha)}\right)\right\}\right\}d\alpha
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma\sigma_\alpha}\exp\left\{-\frac{1y^2}{2\sigma^2}+\frac{y^2\sigma_\alpha^4\left(\sigma^2+\sigma_\alpha\right)}{\left(\sigma^2+\sigma_\alpha\right)2\sigma^2\sigma_\alpha^2}\right\}\sqrt{\frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha}} \\
&\qquad \int\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{1}{\sqrt{\frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha}}}\left[\alpha-\frac{y\alpha_\alpha^2}{\sigma^2+\sigma_\alpha}\right]\right\}d\alpha \\
&= \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_\alpha\right)}}\exp\left\{\left[-\frac{1}{2\sigma^2}+\frac{\sigma_\alpha^2}{\left(\sigma^2+\sigma_\alpha\right)}\right]y^2\right\} \\
&= \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_\alpha\right)}}\exp\left\{\left[\frac{-\sigma_\alpha^2+\sigma^2+\sigma_\alpha^2}{2\sigma^2\left(\sigma^2+\sigma_\alpha\right)}\right]y^2\right\} \\
&= \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_\alpha\right)}}\exp\left\{\left[-\frac{1}{2}\frac{1}{\left(\sigma^2+\sigma_\alpha\right)}\right]y^2\right\} \\
&= \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_\alpha\right)}}\exp\left\{\left[-\frac{1}{2}\frac{1}{\left(\sigma^2+\sigma_\alpha\right)}\right]y^2\right\}
\end{aligned}
$$

Desta forma, podemos notar que $Y \sim N(0, \sigma^2+\sigma_\alpha^2)$. para o cálculo d*a posteriori tem*os que

$$
\begin{aligned}
f(\alpha|y) &\propto f(y|\alpha)g(\alpha) \\
&\propto \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\alpha)^2\right\}\frac{1}{\sqrt{2\pi\sigma_\alpha^2}}\exp\left\{-\frac{1}{2\sigma_\alpha^2}\alpha^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)\alpha^2+\frac{2y\alpha}{2\sigma^2}-\frac{1}{2\sigma^2}y^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)\left[\alpha^2-\frac{2\alpha y}{\sigma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)\left[\alpha-\frac{y}{\sigma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\alpha^2}\right)}\right]^2-\left(\frac{y\sigma_\alpha^2}{(\sigma^2+\sigma_\alpha^2)}\right)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\left(\frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right)}\left[\alpha-\frac{y\sigma_\alpha^2}{(\sigma^2+\sigma_\alpha^2)}\right]^2\right\}.
\end{aligned}
$$

Assim, concluímos que *a posteriori*

$$
\alpha|Y \sim N\left(\frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}y, \frac{\sigma^2\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}\right).
$$

2) **Caso normal com σ^2 desconhecido**: Considere a observação de uma variável aleatória $Y = (Y_1, ..., Y_n)$ com distribuição $Y|\alpha \sim N(\alpha, 1)$ e admita-se *a priori* $\alpha \sim N(0, \sigma^2)$ com σ^2 desconhecido.

Para determinar a distribuição *a posteriori* pelo teorema de Bayes temos que

$$\begin{aligned} g(\alpha|y) &\propto l(\alpha;y)g(\alpha) \\ &\propto \frac{1}{\sqrt{2\pi}}exp\left\{-\frac{1}{2}\left(\alpha-y\right)^2\right\}\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}}exp\left\{-\frac{1}{2\sigma^2}\alpha^2\right\} \\ &\propto exp\left\{-\frac{1}{2}(\alpha^2-2\alpha y+y^2)-\frac{1}{2\sigma^2}\alpha^2\right\} \\ &\propto exp\left\{-\frac{1}{2}\left(\alpha^2-2\alpha y+y^2+\frac{\alpha^2}{\sigma^2}\right)\right\} \\ &\propto exp\left\{-\frac{1}{2}\left(\alpha^2\left(1+\frac{1}{\sigma^2}\right)-2\alpha y+y^2\right)\right\} \\ &\propto exp\left\{-\frac{(\sigma^2+1)}{2\sigma^2}\left(\alpha^2-2\alpha y\left(\frac{\sigma^2}{\sigma^2+1}\right)\right)\right\} \end{aligned}$$

Assim, *a posteriori* tem distribuição normal

$$\alpha|Y \sim N\left(\frac{\sigma^2}{\sigma^2+1}y, \frac{\sigma^2}{\sigma^2+1}\right).$$

Para o caso multivariado, considerando

$$y_i|\alpha_i \sim N(\alpha_i, 1) \qquad \text{e} \qquad \alpha_i \sim N(0,\sigma^2) \qquad i=1,2,...,n.$$

Considerando $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)'$ e $y = (y_1, y_2, ..., y_n)'$, usando a notação para a distribuição normal n-dimensional,

$$Y|\alpha \sim N(\alpha, I) \qquad \text{e} \qquad \alpha \sim N(0, \sigma^2 I)$$

onde I é a matriz identidade de ordem *n x n*. A regra de Bayes nos fornece a distribuição *a posteriori*

$$\begin{aligned}
g(\alpha|y) &\propto l(\alpha;y)g(\alpha) \\
&\propto \left(\frac{1}{\sqrt{2\pi I}}\right)^2 exp\left\{-\frac{1}{2I}(y-\alpha)^2\right\}\frac{1}{\sqrt{2\pi\sigma^2 I}}exp\left\{-\frac{1}{2\sigma^2 I}\alpha^2\right\} \\
&\propto exp\left\{-\frac{1}{2I}(y^2-2y\alpha+\alpha^2)-\frac{1}{2\sigma^2 I}\alpha^2\right\} \\
&\propto exp\left\{-\frac{1}{2I}\left(\alpha^2-2\alpha y+y^2+\frac{\alpha^2}{\sigma^2 I}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2}\left(\alpha^2\left(\frac{1}{I}+\frac{1}{\sigma^2 I}\right)-2y\alpha+y^2\right)\right\} \\
&\propto exp\left\{\frac{-\left(1+\frac{1}{\sigma^2}\right)\frac{1}{I}}{2}\left(\alpha^2-2y\alpha\frac{1}{\left(1+\frac{1}{\sigma^2}\right)\frac{1}{I}}\right)\right\} \\
&\propto exp\left\{-\frac{1}{2\frac{\sigma^2 I}{(\sigma^2+1)}}\left(\alpha-y\frac{\sigma^2 I}{(\sigma^2+1)}\right)^2\right\}
\end{aligned}$$

logo,

$$\alpha|Y \sim N\left(\frac{\sigma^2}{\sigma^2+1}y, \frac{\sigma^2}{\sigma^2+1}I\right).$$

APÊNDICE B - Cálculo da Preditiva Multivariada

Dado o modelo

$$Y_{n\times 1} = X_{n\times p}\alpha_{p\times 1} + \varepsilon_{n\times 1}$$

A distribuição da marginal de Y dado α é dada por

$$Y|\alpha \sim N\left(X\alpha, A^{-1}\right)$$

atribuímos uma informação *a priori* sobre o parâmetro α tal que

$$\alpha \sim N\left(0, B^{-1}\right) \tag{3.2}$$

Para calcularmos a preditiva, devemos calcular a seguinte integral da função densidade da marginal de $Y|\theta$ e a função densidade *a priori* e integrarmos em relação a α, ou seja,

$$\begin{aligned} f(Y) &= \int f(Y|\theta)g(\alpha)d\alpha \\ &= \int \frac{1}{(2\pi)^{p/2}|A^{-1}|^{1/2}} \exp\left(-\frac{1}{2}(Y-X\alpha)'A\,(Y-X\alpha)\right)\frac{1}{(2\pi)^{p/2}|B^{-1}|^{1/2}} \exp\left(-\frac{1}{2}\alpha' B\alpha\right) d\alpha \\ &= \frac{1}{(2\pi)^{p}|A^{-1}|^{\frac{1}{2}}|B^{-1}|^{\frac{1}{2}}} \int \exp\left((Y-X\alpha)'A\,(Y-X\alpha)+\alpha' B\alpha\right) d\alpha \end{aligned}$$

queremos o completamento de quadrados em α para o seguinte termo:

$$\begin{aligned} (Y-X\alpha)'\,A\,(Y-X\alpha)+\alpha' B\alpha &= Y'AY - 2Y'AX\alpha + (X\alpha)'\,A\,(X\alpha) + \alpha' B\alpha \\ &= Y'AY - 2Y'AX\alpha + \alpha'\left(X'AX+B\right)\alpha \end{aligned}$$

ou seja, queremos determinar W tal que:

$$\begin{aligned} (\alpha-W)'\left(X'AX+B\right)(\alpha-W) &= \ldots - 2Y'AX\alpha + \alpha'\left(X'AX+B\right)\alpha \\ &= \alpha'\left(X'AX+B\right)\alpha - 2W'\left(X'AX+B\right)\alpha \\ &\quad + W'\left(X'AX+B\right)W \end{aligned}$$

Então, $W = (X'AX + B)^{-1}X'AY$. Observe que : $Y'AX\alpha = W'(X'AX + B)\alpha$.

$$\begin{aligned}
(Y - X\alpha)' A (Y - X\alpha) + \alpha' B\alpha &= Y'AY + \alpha' \left(X'AX + B\right)\alpha - 2W' \left(X'AX + B\right)\alpha \\
&= Y'AY + \alpha' \left(X'AX + B\right)\alpha - 2W' \left(X'AX + B\right)\alpha \\
&\quad + W' \left(X'AX + B\right) W - W' \left(X'AX + B\right) W \\
&= (\alpha - W)' \left(X'AX + B\right)(\alpha - W) + Y'AY \\
&\quad - W' \left(X'AX + B\right) W
\end{aligned}$$

logo, retornando a preditiva, temos que:

$$\begin{aligned}
f(Y) &\propto \exp\left[Y'AY - W' \left(X'AX + B\right) W\right] \\
&\propto \exp\left[Y'AY - Y'AX\left(X'AX + B\right)^{-1} \left(X'AX + B\right) \left(X'AX + B\right)^{-1} X'AY\right] \\
&\propto \exp\left[Y'AY - Y'AX\left(X'AX + B\right)^{-1} X'AY\right] \\
&\propto \exp\left[Y' \left(A - AX\left(X'AX + B\right)^{-1} X'A\right) Y\right].
\end{aligned}$$

e *a posteriori* é

$$\begin{aligned}
f(\alpha|Y) &\propto \exp\left[(\alpha - W)' \left(X'AX + B\right)(\alpha - W) + Y'AY - W' \left(X'AX + B\right) W\right] \\
f(\alpha|Y) &\propto \exp\left[(\alpha - W)' \left(X'AX + B\right)(\alpha - W)\right]
\end{aligned}$$

Portanto, *a posteriori* é uma normal com média $W = (X'AX + B)^{-1}X'AY$ que é o estimador de Bayes do vetor de parâmetros α.

SEGUNDA PARTE

ARTIGO 1: Estimadores tipo James-Stein e suas propriedades via simulação computacional

Artigo redigido conforme as normas da Revista Brasileira de Biometria - artigo submetido em maio de 2016.

ESTIMADORES TIPO JAMES-STEIN E SUAS PROPRIEDADES VIA SIMULAÇÃO COMPUTACIONAL

Cristiane Alvarenga GAJO[1]
Lucas Monteiro CHAVES[2]
Devanil Jaques de SOUZA[3]

- RESUMO: O estimador de James-Stein para a média de uma normal multivariada independente com variâncias iguais, é obtido por encolhimento adaptativo do estimador média amostral. Esse estimador domina o estimador média para as dimensões maiores ou iguais a três. Neste trabalho, são apresentadas variações deste estimador para o caso normal independente com variâncias distintas e para o caso geral, em que o estimador é obtido utilizando-se a métrica de Mahalanobis. Justificativas geométricas e um estudo do comportamento desses estimadores por simulação computacional para a dimensão três são apresentadas, utilizando o erro quadrático médio como medida de qualidade.
- PALAVRAS-CHAVE: Estimador de James-Stein; Normal Multivariada; Métrica de Mahalanobis.

1 Introdução

Stein (1956) em seu artigo "*Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*" apresentou uma prova de que o estimador de máxima verossimilhança de uma população normal multivariada, isto é, a média amostral, é inadmissível. Esse resultado foi um choque para o mundo estatístico com consequências em várias áreas. Em James e Stein (1961), é apresentado explicitamente um estimador que domina a média amostral. Esse estimador passou a ser denominado estimador de James-Stein e é dado por

[1]Universidade Federal de Lavras, Departamento de Ciências Exatas, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *cristianegajo@yahoo.com.br*
[2]Universidade Federal de Lavras, Departamento de Ciências Exatas - DEX, CEP: 37200-000, Lavras, Minas Gerais, Brasil. E-mail: *lucas@dex.ufla.br*
[3]Universidade Federal de Lavras, Departamento de Ciências Exatas - DEX, CEP: 37200-000, Lavras, Minas Gerais, Brasil. E-mail: *devanil@dex.ufla.br*

$$\widehat{\boldsymbol{\theta}}_{JS}\left(\mathbf{X}\right)=\left(1-\sigma^{2}\frac{\mathrm{p}-2}{\left\|\mathbf{X}\right\|^{2}}\right)\mathbf{X},\qquad \mathrm{p}\geq 3.$$

em que $\mathbf{X}$ tem distribuição normal independente multivariada de dimensão $\mathrm{p} \geq 3$ e a variância σ^2 conhecida. O estudo do estimador de James-Stein apresenta vários aspectos como: relações com teoria Bayesiana empírica, propriedades de funções harmônicas e uma rica abordagem geométrica. Uma referência abrangendo estes vários aspectos é Gajo (2016), em que se abordam as referências básicas da teoria. Neste trabalho, modificações do estimador de James-Stein são justificadas geometricamente e estudados por simulação computacional.

2 Uma justificativa geométrica para estimadores de encolhimento

O estimador de James-Stein é um caso particular de estimador obtido por encolhimento (shrinkage). De fato, é um estimador de encolhimento adaptativo, pois o fator de encolhimento $\sigma^2\frac{(p-2)}{\|\mathbf{X}\|^2}$ depende do valor observado $\mathbf{X}$. Uma das motivações de Stein, para obtenção deste estimador, é o fato que um estimador não viesado $\hat{\boldsymbol{\theta}}=\left(\hat{\theta}_1,...,\hat{\theta}_k\right)$ do vetor de parâmetros $\boldsymbol{\theta}=(\theta_1,...,\theta_k)$ apresenta a seguinte deficiência, nem sempre observada.

$$\begin{aligned}
E\left[||\hat{\boldsymbol{\theta}}||^2\right] &= E\left[\sum_{i=1}^{p}\hat{\theta}_i^2\right]=\sum_{i=1}^{p}E\left[\hat{\theta}_i^2\right]\\
&=\sum_{i=1}^{p}\left(var\left[\hat{\theta}_i\right]+\left(E\left[\hat{\theta}_i\right]\right)^2\right)\\
&=\sum_{i=1}^{p}var\left[\hat{\theta}_i\right]+\sum_{i=1}^{p}\theta_i^2\\
&=\sum_{i=1}^{p}var\left[\hat{\theta}_i\right]+||\boldsymbol{\theta}||^2.
\end{aligned}$$

Portanto, apesar de $\hat{\theta}_i$ estimar de forma não viesada θ_i, $||\hat{\boldsymbol{\theta}}||^2$, superestima $||\boldsymbol{\theta}||^2$. Surge desse fato a ideia de encolhimento. Stein utilizou argumentos geométricos para justificar a construção do seu estimador. Neste trabalho, uma nova justificativa geométrica é apresentada.

Considere a "nuvem de pontos" obtida por uma amostra $\{\mathrm{x}_1,...,\mathrm{x}_\mathrm{n}\}$ com n grande, conforme Figura 1.

Como $\mathbf{X}\sim\mathbf{N}(\boldsymbol{\theta},\sigma^{\mathbf{2}}\mathbf{I})$, esta nuvem possui, com alta probabilidade, simetria esférica com o vetor de médias $\boldsymbol{\theta}$ no centro da nuvem. Considere os pontos desta nuvem que estão próximos à superfície de uma esfera centrada em $\boldsymbol{\theta}$ com raio igual ao desvio padrão σ. Esses pontos são aproximadamente equiprováveis em razão da

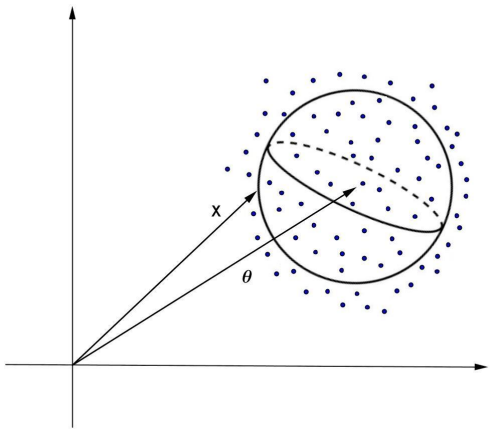

Figura 1 - Nuvem de dados ao redor do vetor de médias $\boldsymbol{\theta}$.

independência das componentes do vetor aleatório $\mathbf{X}$. Considere uma outra esfera centrada na origem e de raio $||\hat{\boldsymbol{\theta}}||$. A interseção dessas duas esferas gera na esfera centrada em $||\hat{\boldsymbol{\theta}}||$ duas calotas esféricas, conforme Figura 2. A calota definida pelos vetores x tais que $||x|| \leq ||\boldsymbol{\theta}||$ tem evidentemente área menor que a calota definida pelos vetores x com $||x|| \geq ||\boldsymbol{\theta}||$. Portanto, é intuitivo que ao se amostrar um vetor x sobre a esfera de centro $\boldsymbol{\theta}$, como todos os pontos sobre esta esfera são equiprováveis, a probabilidade de se observar um vetor com $||x|| > ||\boldsymbol{\theta}||$ é proporcional a área da calota, o que significa que é mais provável se observar um vetor de comprimento maior do que $\|\boldsymbol{\theta}\|$. Desta forma, temos uma justificativa heurística para o fato que $E\left[\|\mathbf{X}\|^2\right] = ||\boldsymbol{\theta}|| + p\sigma^2 > \|\boldsymbol{\theta}\|^2$. A ideia do encolhimento é agora justificada pois, ao se realizar o encolhimento podemos obter duas calotas esféricas $||x|| \geq ||\boldsymbol{\theta}||$ e $||x|| \leq ||\boldsymbol{\theta}||$ com áreas iguais.

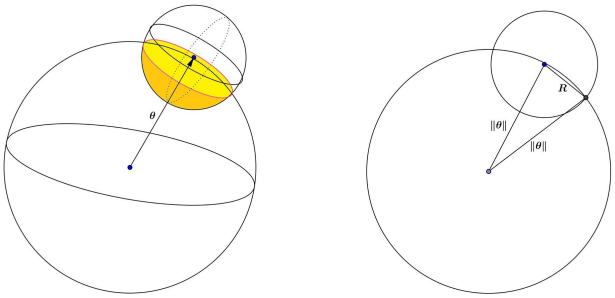

Figura 2 - Interseção entre as esferas.

3 Estimadores tipo James-Stein

Uma questão pertinente é o porquê do estimador de James-Stein não dominar a média amostral para o caso geral $\mathbf{X} \sim \mathbf{N_p}\left(\boldsymbol{\theta}, \sigma^{\mathbf{2}}\mathbf{B}\right)$. Certamente, isto ocorre em razão da não simetria esférica que ocorre pela presença de covariâncias. Tem-se, então, simetria elíptica que talvez não seja muito adequada para o encolhimento na direção da origem. O encolhimento de elipses na direção da origem são elipses Figura 3, pois

$$\frac{(x-a)^2}{c_1^2} + \frac{(y-b)^2}{c_2^2} = 1$$
$$\frac{(\alpha x - \alpha a)^2}{c_1^2} + \frac{(\alpha y - \alpha b)^2}{c_2^2} = \alpha^2$$
$$\frac{(\alpha x - \alpha a)^2}{(\alpha c_1)^2} + \frac{(\alpha y - \alpha b)^2}{(\alpha c_2)^2} = 1.$$

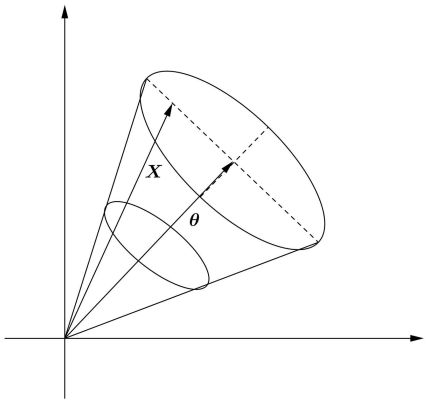

Figura 3 - Simetria elíŋptica para encolhimento.

Outras opções de direção para o encolhimento seriam as direções dos autovetores, que respeitam a simetria elíptica. Qual dessas direções seria a mais adequada para o encolhimento, visando diminuir o erro quadrático médio? Para identificar a melhor direção, será utilizada simulação computacional. Uma possibilidade de encolhimento é a de se ter encolhimento com taxas diferentes para cada coordenada.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} \quad \text{e} \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{\alpha} \otimes \mathbf{X} = \begin{pmatrix} \alpha_1 X_1 \\ \vdots \\ \alpha_p X_p \end{pmatrix}$$

$$\hat{\boldsymbol{\theta}}_{JS} = \left(1 - \frac{p-2}{\|\mathbf{X}\|^2}\right) \begin{pmatrix} \alpha_1 X_1 \\ \vdots \\ \alpha_p X_p \end{pmatrix}$$

em que os α_i devem depender das variâncias em cada direção principal que valem $\frac{\sigma^2}{\lambda_i}$, com λ_i autovalor. Os vetores canônicos e_i não são direções principais a não ser que

$$\text{cov}[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2 \end{pmatrix},$$

pois, neste caso, a elipse possui como eixos de simetria os eixos coordenados. Estas considerações geométricas levam à definição de um estimador tipo James-Stein modificado descrito na subseção seguinte.

3.1 Estimador tipo James-Stein para o caso normal independente com variâncias distintas

Considere

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \sim N_p \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix} \right).$$

O estimador James-Stein modificado para variâncias distintas, no caso $p = 3$, é

$$\hat{\theta}_{JS \text{ mod}}\ (\mathbf{X}) = \begin{pmatrix} \left(1 - \sigma_1^2 \frac{1}{\|\mathbf{X}\|^2}\right) X_1 \\ \left(1 - \sigma_2^2 \frac{1}{\|\mathbf{X}\|^2}\right) X_2 \\ \left(1 - \sigma_3^2 \frac{1}{\|\mathbf{X}\|^2}\right) X_3 \end{pmatrix}.$$

Esse estimador tem valores diferenciados para o coeficiente de encolhimento em cada coordenada. Assim, o estimador de encolhimento terá direção diferente do vetor $\mathbf{X}$, isto é, a direção do encolhimento não é radial em relação à origem.

Para exemplificar, suponha o estimador em $\mathbb{R}^2$, $\mathbf{X} = (X_1, X_2)$, considerando $\sigma_1^2 = 1$ e $\sigma_2^2 = 2$. Tem-se que

$$\hat{\theta}_{JS \text{ mod}}\ (\mathbf{X}) = \left(\underbrace{\left(1 - \frac{1}{X_1^2 + X_2^2}\right)}_{encolhe\,menos} X_1, \underbrace{\left(1 - \frac{2}{X_1^2 + X_2^2}\right)}_{encolhe\,mais} X_2 \right).$$

Ocorre um encolhimento maior na coordenada de maior variabilidade, o que diminui a variabilidade do estimador nesta direção, conforme Figura 4.

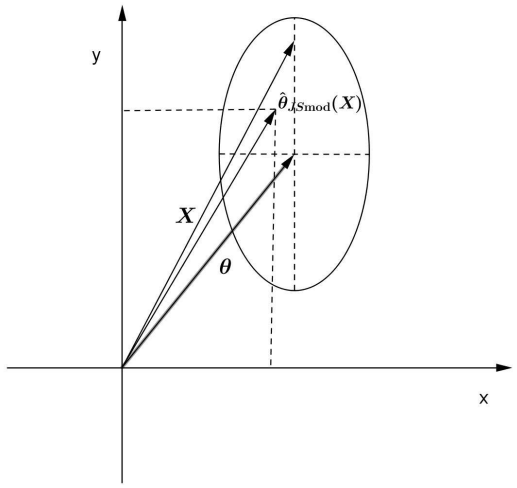

Figura 4 - $\hat{\theta}_{JS\,\text{mod}}$ $(\mathbf{X})$ com simetria elíptica no $\mathbb{R}^2$.

3.2 Estimador tipo James-Stein para o caso geral

Considere a situação geral $\mathbf{X} = (X_1, X_2, ..., X_2) \sim \mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$ em que em que $\boldsymbol{\Sigma}$ é uma matriz conhecida, simétrica e positiva definida (ANDERSON,2003). Neste caso, os eixos principais da elipse não são mais paralelos aos eixos coordenados.

A ideia é fazer uma transformação linear na variável $\mathbf{X}$ tal que a nova variável tenha distribuição normal independente. Para isto, é necessário o processo de diagonalização da matriz $\boldsymbol{\Sigma}$.

De forma geral, este processo pode ser entendido conforme se segue. A decomposição espectral da matriz $\boldsymbol{\Sigma}$ dada por $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$, em que $\mathbf{P}$ é a matriz composta pelos autovetores de $\boldsymbol{\Sigma}$ em suas colunas e pelos autovalores λ_i $\boldsymbol{\Lambda}$ de$\boldsymbol{\Sigma}$ (FERREIRA, 2011). A inversa de $\boldsymbol{\Sigma}$ é dada por $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}'$ e esta expressão pode ser generalizada para obter qualquer potência com expoente real da matriz $\boldsymbol{\Sigma}$, particularmente para o expoente 1/2

$$\boldsymbol{\Sigma}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'.$$

Considere o vetor $\mathbf{Y}$=$(Y_1, Y_2, ..., Y_p)$ obtido pela transformação linear de Mahalanobis (MARDIA; KENT; BIBBY, 2003).

$$\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}\mathbf{X}$$

Como $\text{cov}[\mathbf{X}] = \boldsymbol{\Sigma}$, tem-se que

$$\begin{aligned}\text{cov}[\mathbf{Y}] &= \text{cov}[\boldsymbol{\Sigma}^{-1/2}\mathbf{X}] \\ &= \boldsymbol{\Sigma}^{-1/2}\text{cov}[\mathbf{X}]\boldsymbol{\Sigma}^{-1/2} \\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} \\ &= \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} \\ &= \mathbf{I}.\end{aligned}$$

O vetor aleatório $\mathbf{Y}$ tem distribuição normal multivariada e o estimador usual de James-Stein é dado por

$$\begin{aligned}\hat{\theta}_{JS}(\mathbf{Y}) &= \left(1 - \frac{p-2}{\mathbf{Y}'\mathbf{Y}}\right)\mathbf{Y} \\ &= \left(1 - \frac{p-2}{\left(\boldsymbol{\Sigma}^{-1/2}\mathbf{X}\right)'\left(\boldsymbol{\Sigma}^{-1/2}\mathbf{X}\right)}\right)\boldsymbol{\Sigma}^{-1/2}\mathbf{X} \\ &= \left(1 - \frac{p-2}{\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}}\right)\boldsymbol{\Sigma}^{-1/2}\mathbf{X}.\end{aligned}$$

Como $E[\mathbf{Y}] = \boldsymbol{\Sigma}^{-1/2}E[\mathbf{X}] = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}$ e a matriz $\mathbf{P}$ é ortogonal, o vetor $\mathbf{X}$ pode ser recuperado pela transformação $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Y}$. Obtém-se então em relação à variável $\mathbf{X}$ um estimador tipo James-Stein definido por

$$\begin{aligned}\hat{\theta}_{JSmah}(\mathbf{X}) &= \boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{\theta}}_{JS}(\mathbf{Y}) \\ &= \left(1 - \frac{p-2}{\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}}\right)\mathbf{X}.\end{aligned}$$

A seção seguinte apresenta um estudo computacional para este estimador no caso particular em que se tem que a variável aleatória $\mathbf{X}$ é uma normal multivariada equicorrelacionada e, portanto, a matriz $\boldsymbol{\Sigma}$ expressa por $\boldsymbol{\Sigma} = \mathbf{V}^{\frac{1}{2}}\boldsymbol{\rho}\mathbf{V}^{\frac{1}{2}}$, e $\boldsymbol{\rho}$ é a matriz de correlação populacional dada por

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix},$$

com a matriz $\mathbf{V}^{\frac{1}{2}} = diag(\sqrt{\sigma_i^2})$.

3.3 Estudo computacional

Por simulação Monte Carlo, são avaliados os estimadores propostos $\hat{\theta}_{JSmod}$ e $\hat{\theta}_{JSmah}$ para $p = 3$, comparando-se o desempenho dos mesmos com o estimador

média amostral. A avaliação dos estimadores se deu por meio do erro quadrático médio, que será estimado por

$$EQM(\hat{\theta}) = \frac{1}{m}\sum_{j=1}^{m}\left\|\hat{\theta}_j - \theta\right\|^2 = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{p}\left(\hat{\theta}_{ji} - \theta_i\right)^2,$$

em que $\hat{\boldsymbol{\theta}}_j$ é a estimativa de $\boldsymbol{\theta}$ obtida na j-ésima amosta e m é o número total de amostras. Para as simulações, considerou-se apenas o caso $p = 3$ e as rotinas foram desenvolvidas na linguagem R.

Para a simulação, gerou-se uma normal multivariada independente de dimensão $p = 3$, e o vetor de médias variando em um reticulado contido no cubo $(-10, 10) \times (-10, 10) \times (-10, 10)$, totalizando 9.261 pontos no reticulado. Os estimadores são avaliados para apenas uma observação da normal multivariada tridimensional. Para o cálculo do EQM, o processo é repetido 100, 1000 e 10000 vezes, conforme Tabelas 2, 3, 4, 5 e 6. Contabilizou-se a porcentagem de vezes que a estimativa do erro quadrático médio dos estimadores de James-Stein modificado foi menor do que a estimativa do erro quadrático médio correspondente do estimador média usual, que no caso é o próprio valor observado da normal uma vez que a amostra é de tamanho unitário. As rotinas necessárias para a implementação e avaliação dos testes foram realizadas, utilizando o programa R em sua versão 3.0.2 (R CORE TEAM, 2014).

Na Tabela 1, com o objetivo de avaliar o comportamento da estimativa do erro quadrático em relação ao número de repetições, a comparação se faz entre o estimador de James-Stein usual e a média amostral. Como era de se esperar, a partir do número de repetições 1000, praticamente em 100% dos casos o estimador de James-Stein possui estimativa de erro quadrático médio menor do que a média amostral. Portanto, a escolha dos números de repetições 100, 1000 e 10000 é adequada para a comparação.

Tabela 1 - Comparação do erro quadrático médio dos estimadores James-Stein $\hat{\theta}_{JS}(\mathbf{X})$ e o estimador usual $\mathbf{X}$, para $p = 3$.

Número de repetições	Variâncias	$EQM(\hat{\boldsymbol{\theta}}_{JS}(\mathbf{X})) < EQM(\mathbf{X})$ (%)
100	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	76,58
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	92,60
1.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	97,70
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,74
10.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	100
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,97

Na Tabela 2 é comparado o estimador James-Stein modificado com variâncias $\sigma_1^2 = 2, \sigma_2^2 = 4$ e $\sigma_3^2 = 9$, e também com variâncias $\sigma_1^2 = 1, \sigma_2^2 = 2$ e $\sigma_3^2 = 4$. Constata-se, então, que de forma aparentemente independente do número de repetições e dos valores das variâncias, o estimador de James-Stein modificado

domina o estimador média em torno de 72% das vezes. Uma questão interessante, a qual não foi abordada, seria detectar no reticulado cúbico, qual a região em que o comportamento do estimador modificado tem desempenho ainda melhor, e analisar se essa região está relacionada à geometria elíptica da normal.

Tabela 2 - Comparação do erro quadrático médio dos estimadores James-Stein modificado $\hat{\theta}_{JS\,\mathrm{mod}}\ (\mathbf{X})$ e o estimador usual $\mathbf{X}$, para $p = 3$.

Número de repetições	Variâncias	$EQM\left(\hat{\boldsymbol{\theta}}_{JS\,\mathrm{mod}}\ (\mathbf{X})\right) < EQM(\mathbf{X})$ (%)
100	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	75,58
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	70,79
1.000	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	75,20
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	71,13
10.000	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	75,42
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	71,18

Nas tabelas 3, 4, 5 e 6, os resultados obtidos para o estimador $\hat{\theta}_{JSmah}$ são apresentados para o caso em que a normal multivariada é equicorrelacionada para valores do coeficiente de correlação $\rho = 0, 0.2, 0.5$ e 0.9.

Em todas as tabelas e para todos os valores de correlação, para 1.000 ou 10.000 repetições, obtém-se que o estimador $\hat{\theta}_{JSmah}$ domina o estimador média amostral em uma proporção superior à 91%.

Tabela 3 - Comparação do erro quadrático médio do estimador obtido pela transformação de Mahalanobis $\hat{\theta}_{JS\underline{\mathrm{m}}\mathrm{ah}}\,(\mathbf{X})$ com correlação $\rho = 0$ (sem correlação entre os pares de variáveis) e o estimador usual $\mathbf{X}$, para $p = 3$ e tamanho da amostra 1.

Número de repetições	Variâncias	$EQM(\hat{\boldsymbol{\theta}}_{JS\underline{\mathrm{m}}\mathrm{ah}}\,(\mathbf{X})) < EQM(\mathbf{X})$ (%)
100	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	78,01
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	92,79
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	90,70
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	85,70
1.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	97,85
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,75
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	93,97
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	91,37
10.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	99,99
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,95
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	94,26
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	92,25

Uma observação interessante é que o efeito da correlação sobre o desempenho

dos estimadores não é muito relevante. Este fato é um tanto quanto inesperado, uma vez que, a correlação alta implica que a simetria elíptica da normal multivariada tem preponderância em um dos eixos principais.

Tabela 4 - Comparação do erro quadrático médio do estimador obtido pela transformação de Mahalanobis $\hat{\theta}_{JSmah}(\mathbf{X})$, considerando a correlação entre os pares de variáveis de $\rho = 0,2$, e o estimador usual $\mathbf{X}$, para $p = 3$ e tamanho de amostra 1.

Número de repetições	Variâncias	$EQM(\hat{\boldsymbol{\theta}}_{JSmah}(\mathbf{X})) < EQM(\mathbf{X})$ (%)
100	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	78,51
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	92,54
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	91,23
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	86,36
1.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	97,53
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,83
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	94,08
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	91,72
10.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	99,98
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,92
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	94,42
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	92,26

Tabela 5 - Comparação do erro quadrático médio do estimador obtido pela transformação de Mahalanobis $\hat{\theta}_{JSmah}(\mathbf{X})$, considerando a correlação entre os pares de variáveis de $\rho = 0,5$, e o estimador usual $\mathbf{X}$, para $p = 3$ e tamanho da amostra 1.

Número de repetições	Variâncias	$EQM(\hat{\boldsymbol{\theta}}_{JSmah}(\mathbf{X})) < EQM(\mathbf{X})$ (%)
100	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	80,40
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	92,80
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	92,29
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	87,81
1.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	96,65
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,28
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	94,57
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	92,46
10.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	99,44
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,95
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	94,88
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	92,96

Uma observação interessante é que o efeito da correlação sobre o desempenho dos estimadores não é muito relevante. Este fato é um tanto quanto inesperado uma vez que a correlação alta implica que a simetria elíptica da normal multivariada tem preponderância em um dos eixos principais.

Tabela 6 - Comparação do erro quadrático médio do estimador obtido pela transformação de Mahalanobis $\hat{\theta}_{JS\underline{m}\underline{a}h}(\mathbf{X})$, considerando a correlação entre os pares de variáveis de $\rho = 0,9$, e o estimador usual $\mathbf{X}$, para $p = 3$ e tamanho da amostra 1.

Número de repetições	Variâncias	$EQM(\hat{\boldsymbol{\theta}}_{JS\underline{m}\underline{a}h}) < EQM(\mathbf{X})(\%)$
100	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	86,63
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	94,82
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	93,95
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	90,69
1.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	98,19
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,51
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	95,78
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	94,52
10.000	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 2$	99,26
	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$	99,94
	$\sigma_1^2 = 2; \sigma_2^2 = 4; \sigma_3^2 = 9$	96,02
	$\sigma_1^2 = 1; \sigma_2^2 = 2; \sigma_3^2 = 4$	94,99

4 Conclusão

A abordagem geométrica permite uma justificação natural e intuitiva aos estimadores tipo James-Stein estudados. A questão do encolhimento na direção dos eixos principais é abordada e apresenta-se promissora para futuros estudos. O desempenho desses estimadores para os casos independentes e com variâncias distintas, bem como no caso equicorrelacionado, é superior ao desempenho do estimador média para a amplitude do vetor de médias analisado e, portanto, apresenta-se como uma opção viável em possíveis aplicações.

Referências

ANDERSON, T.W.*An Introduction to Multivariate Statistical Analysis*. New Jersey: John Wiley & Sons, 2003. 667 p.

BRANDWEIN, A.C; STRAWDERMAN, W. E. *Stein Estimation: The Spherically Symetric Case*. Statistical Science. v.5, n.3, p. 356-369, 1990.

BRANDWEIN, A.C; STRAWDERMAN, W. E. *Stein Estimation for Spherically Symetric Distributions: Recent Developmentse*. Statistical Science. v.27, n.1, p. 11-23, 2012.

BROWN, L.D; ZHAO, L. H.*A Geometrical Explanation of Stein Shrinkage*. Statistical Science. v.27, n.1, p. 24-30, 2012.

FERREIRA,D.F. *Estatística Multivariada*. Lavras: Ed. UFLA, 2011. 676 p.

GAJO, C. A. *Propriedades e Aspectos Geométricos de Estimadores Tipo James-Stein e do Estimador de Hartigan*. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.

GRUBER, M. H. J.*Improving Efficiency by Shrinkage*. New York: Marcell Dekker, 1998. 632 p.

JAMES, W.; STEIN, C. *Estimation with quadratic loss*. Proc. Fourth Berkeley Symp. Math. Statist. Probab., Univ. California Press, v.1, p. 361-380, 1961.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. *Multivariate Analysis*. San Diego:Academic Press, 2003.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>. Acesso em: 1 de agosto de 2015.

Recebido em 0x.0x.20xx.

Aprovado após revisão em 0x.0x.20xx.

ARTIGO 2: Estimating bounded mean vector in multivariate normal: the geometry of Hartigan estimator

Artigo redigido conforme as normas da Revista Brasileira de Biometria - artigo publicado em abril de 2016.

ESTIMATING BOUNDED MEAN VECTOR IN MULTIVARIATE NORMAL: THE GEOMETRY OF HARTIGAN ESTIMATOR

Cristiane Alvarenga GAJO[1]
Leandro da Silva PEREIRA[2]
Lucas Monteiro CHAVES[3]
Devanil Jaques SOUZA[4]

▪ *ABSTRACT: The problem on estimating a multivariate normal mean $N_p(\theta, I)$ when it is supposed that the mean vector is bounded arouses practical and theoretical interest since under this hypothesis it's possible obtain estimators which dominate the sample mean estimator in relation to square loss. Generalizing previous results obtained for univariate normal, J.A. Hartigan obtained for multivariate normal with independent components a Bayes estimator defined on a bounded closed convex set, with non-empty interior, which dominate the sample mean estimator. In this work, this result is presented in detail for the case where the convex set is a sphere centered at origin and a geometrical interpretation, useful to understand the phenomenon, is presented. Others estimators based on Gatsonis et al. work are proposed and the risk of these estimators are compared through simulation for the cases $p = 2$ and $p = 3$.*

▪ *KEYWORDS: Multivariate Normal; Convex Sets; Uniform Priors; Bayes Estimator.*

1 Introduction

After Stein's result (1956) which obtained a shrinkage estimator for the mean vector of a p-variate normal, $\mathrm{N_p}(\theta, \mathrm{I})$, which dominates the usual estimator $\delta(\mathrm{X}) = \mathrm{X}$, in relation to squared risk, there was an intense search for other estimators which dominate $\delta(\mathrm{X})$. In this line of thought, it arises the problem of, supposing the mean vector θ is restricted to a limited set $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{p}}$, , obtain estimator which dominate

[1]University Federal of Lavras, Department of the Exacts Sciences, Postal Box 37, CEP: 372000-000, Lavras, MG, Brazil. E-mail: *cristianegajo@yahoo.com.br*
[2]University Federal of Lavras, Department of the Exacts Sciences, Postal Box 37, CEP: 37200-000, Lavras, MG,Brazil. E-mail: *lespleandro2@hotmail.com*

$\delta(\mathrm{X})$ when to the risk function is restricted to $\theta \in \mathrm{C}$. Generalizing previous results from Casella-Strawderman (1981), Gatsonis et al. (1987), Hartigan (2004) obtained a generalized Bayes estimator from a uniform priori in a bounded closed convex set, with non-empty interior C and smooth enough border that has squared risk smaller or equal of $\delta(\mathrm{X})$ for any $\theta \in \mathrm{C}$. Since the result is stated for general convex sets it is necessary the use of technical results as measure theory and the fact that the paper is too succinct, makes it comprehension difficult, both for the mathematical steps and for a intuitive justification of the result.

In this paper, Hartingan's result is redone in details, supposing C a sphere centered at origin and emphasizing the geometrical aspect. Others estimators based on Gatsonis et al. work are proposed and a computational simulation study related to risk reduction is performed. It is also proved that for origin-centered hypercube Bayes estimator also dominates the maximum likelihood estimator. The used notation is as close as possible to the papers Gatsonis et al. (1987) and Hartigan (2004).

2 Bayes estimator related to uniform priori dominates mean estimator

Consider the random vector $\mathbf{X} = (X_1, \ldots, X_{\mathrm{p}}) \sim \mathrm{N_p}(\theta, \mathbf{I})$, $\theta = (\theta_1, \ldots, \theta_{\mathrm{p}}) \in \mathrm{C} \subseteq \mathbb{R}^{\mathrm{p}}$ with C as a ball centered at origin and radius m. Probability density of $\mathbf{X}$ can be written as

$$f_{\mathbf{X}}(\mathbf{x};\theta) = \frac{1}{(2\pi)^{\frac{\mathrm{p}}{2}}} e^{-\frac{1}{2}\|\mathbf{x}-\theta\|^2} = \prod_{\mathrm{i}=1}^{\mathrm{p}} \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}(x_{\mathrm{i}}-\theta_{\mathrm{i}})^2} = \prod_{\mathrm{i}=1}^{\mathrm{p}} \phi(x_{\mathrm{i}} - \theta_{\mathrm{i}})$$

in which $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Is used, with certain abuse of the notation, $\phi(\mathbf{x}-\theta) = f_{\mathbf{X}}(x_1, \ldots x_{\mathrm{p}};\theta) = \prod_{\mathrm{i}=1}^{\mathrm{p}} \phi(x_{\mathrm{i}} - \theta_{\mathrm{i}})$.

Considering a uniform priori on C given by $\pi(\theta) = \frac{1}{vol(\mathrm{C})}$ it has a posteriori $\pi(\theta|\,\mathbf{x}) = \frac{\phi(\mathbf{x}-\theta)\frac{1}{vol(\mathrm{C})}}{\int_{\mathrm{C}} \phi(\mathbf{x}-\theta)\frac{1}{vol(\mathrm{C})}\mathrm{d}\theta} = \frac{\phi(\mathbf{x}-\theta)}{\int_{\mathrm{C}} \phi(\mathbf{x}-\theta)\mathrm{d}\theta}$.

Bayes estimator $\mathbf{T}(\mathbf{x}) = (\mathrm{T}_1(\mathbf{x}), \ldots, \mathrm{T}_{\mathrm{p}}(\mathbf{x}))$is given by the posteriori mean, e.g.

$$T_j(\mathbf{x}) \quad = \quad \frac{\int_C \theta_j \phi(\mathbf{x}-\theta)\, d\theta}{\int_C \phi(\mathbf{x}-\theta)\, d\theta}.$$

For simplicity is used the notation $\mathbf{T}(\mathbf{x}) = \frac{\int_{\mathrm{C}} \theta\, \phi(\mathbf{x}-\theta)\, \mathrm{d}\theta}{\int_{\mathrm{C}} \phi(\mathbf{x}-\theta)\, \mathrm{d}\theta}$. Squared risk of this

estimator is given by

$$\mathrm{R}_{\mathbf{T}}(\theta) = \mathrm{E}_{\theta}\left[\sum_{\mathrm{i}=1}^{\mathrm{p}} \left(\mathrm{T}_{\mathrm{i}}(\mathbf{x}) - \theta_{\mathrm{i}}\right)^2\right] = \sum_{i=1}^{p} \int_{\mathbb{R}^p} (T_i(\mathbf{x}) - \theta_{\mathbf{i}})^2 \phi(\mathbf{x} - \theta) d\mathbf{x}.$$

For the estimator $\delta(\mathrm{X}) = \mathrm{X}$, not considering any restriction about mean vector θ, the risk is

$$\mathrm{R}_{\delta}(\theta) = \mathrm{E}_{\theta}\left[\|\mathrm{X} - \theta\|^2\right] = \mathrm{E}_{\theta}\left[\sum_{\mathrm{i}=1}^{\mathrm{p}} \left(\mathrm{X}_{\mathrm{i}} - \theta_{\mathrm{i}}\right)^2\right] = \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{E}_{\theta}\left[\left(\mathrm{X}_{\mathrm{i}} - \theta_{\mathrm{i}}\right)^2\right] = \sum_{\mathrm{i}=1}^{\mathrm{p}} 1 = \mathrm{p}.$$

The boundary $\partial\mathrm{C}$ of the p-dimensional ball C is a $(p-1)$-dimensional sphere. The unit normal vector to $\partial\mathrm{C}$ in a point $\theta \in \partial\mathrm{C}$ is given by $\frac{\theta}{\|\theta\|} = \eta(\theta)$.

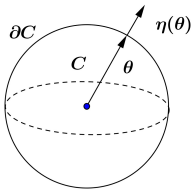

Figure 1 - Sphere unit normal field.

If $\mathrm{p}(\mathbf{x}) = \mathrm{p}(\mathrm{x}_1, \ldots \mathrm{x}_{\mathrm{p}}) = \int_{\mathrm{C}} \phi(\mathbf{x} - \theta)\, \mathrm{d}\theta$, then

$$\begin{aligned}
\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x}) &= \frac{\partial}{\partial x_{\mathrm{i}}} \int_{\mathrm{C}} \phi(\mathbf{x} - \theta)\, \mathrm{d}\theta = \int_{\mathrm{C}} \frac{\partial}{\partial x_{\mathrm{i}}} \phi(\mathbf{x} - \theta)\, \mathrm{d}\theta \\
&= \int_{\mathrm{C}} \frac{\partial}{\partial x_{\mathrm{i}}} \frac{1}{(2\pi)^{\frac{\mathrm{p}}{2}}} e^{-\frac{1}{2}\sum (x_{\mathrm{i}} - \theta_{\mathrm{i}})^2} \mathrm{d}\theta = \int_{\mathrm{C}} -\phi(\mathbf{x} - \theta)(x_{\mathrm{i}} - \theta_{\mathrm{i}})\, \mathrm{d}\theta \\
&= -x_{\mathrm{i}} \int_{\mathrm{C}} \phi(\mathbf{x} - \theta)\, \mathrm{d}\theta + \int_{\mathrm{C}} \theta_{\mathrm{i}} \phi(\mathbf{x} - \theta)\, \mathrm{d}\theta = -x_{\mathrm{i}}\, \mathrm{p}(\mathbf{x}) + \mathrm{T}_{\mathrm{i}}(\mathbf{x})\, \mathrm{p}(\mathbf{x}),
\end{aligned}$$

and therefore $\mathrm{T}_{\mathrm{i}}(\mathbf{x}) = x_{\mathrm{i}} + \frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})$.

Comparing risks for a vector ξ in the interior of C

$$
\begin{aligned}
\mathrm{R}_{\mathbf{T}}(\xi)-\mathrm{R}_{\delta}(\xi) &= \mathrm{R}_{\mathbf{T}}(\xi)-p \\
&= \mathrm{E}_{\xi}\left[\|\mathbf{T}-\xi\|^{2}\right]-\mathrm{E}_{\xi}\left[\|\mathbf{X}-\xi\|^{2}\right] \\
&= \mathrm{E}_{\xi}\left[\|\mathbf{T}-\xi\|^{2}-\|\mathrm{X}-\xi\|^{2}\right] \\
&= \int_{\mathbb{R}^{\mathrm{p}}}\left(\sum_{\mathrm{i}=1}^{\mathrm{p}}\left(\mathrm{T_i}(\mathbf{x})-\xi_{\mathrm{i}}\right)^{2}-\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)^{2}\right)\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^{\mathrm{p}}}\left[\sum_{\mathrm{i}=1}^{\mathrm{p}}\left(\mathrm{x_i}+\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})-\xi_{\mathrm{i}}\right)^{2}-\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)^{2}\right]\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^{\mathrm{p}}}\sum_{\mathrm{i}=1}^{\mathrm{p}}\left[\left(\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\right)^{2}+2\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)+\right. \\
&\quad \left.\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)^{2}-\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)^{2}\right]\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x} \\
&= \int_{\mathbb{R}^{\mathrm{p}}}\sum_{\mathrm{i}=1}^{\mathrm{p}}\left[\left(\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\right)^{2}+2\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\right]\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x} \\
&= \sum_{\mathrm{i}=1}^{\mathrm{p}}\left[\int_{\mathbb{R}^{\mathrm{p}}}\left(\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\right)^{2}\phi(\mathbf{x}-\xi)\,dx\right. \\
&\quad \left.+2\int_{\mathbb{R}^{\mathrm{p}}}\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x}\right].
\end{aligned}
$$

For calculate this integral observing that

$$
\frac{\partial \phi}{\partial x_{\mathrm{i}}}(\mathbf{x}-\xi)=-\phi(\mathbf{x}-\xi)\left(x_{\mathrm{i}}-\xi_{i}\right), \tag{1}
$$

thus, using (1)

$$
\begin{aligned}
\int_{\mathbb{R}^{\mathrm{p}}}\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\,\phi(\mathbf{x}-\xi)\,\mathrm{d}\mathbf{x} &= \int_{\mathbb{R}^{\mathrm{p}}}\left(\mathrm{x_i}-\xi_{\mathrm{i}}\right)\phi(\mathbf{x}-\xi)\,\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^{\mathrm{p}}}\left(\frac{\partial}{\partial x_{\mathrm{i}}}\phi(\mathbf{x}-\xi)\,\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}}(\mathbf{x})\right)\mathrm{d}\mathbf{x}
\end{aligned}
$$

using integration by parts

$$
\begin{aligned}
u=\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial \mathrm{p}(\mathbf{x})}{\partial x_{\mathrm{i}}} \quad &\Rightarrow \quad du=\frac{-\frac{\partial \mathrm{p}(\mathbf{x})}{\partial x_{\mathrm{i}}}}{\mathrm{p}^{2}(\mathbf{x})}\frac{\partial \mathrm{p}(\mathbf{x})}{\partial x_{\mathrm{i}}}+\frac{1}{\mathrm{p}(\mathbf{x})}\frac{\partial^{2}\mathrm{p}(\mathbf{x})}{\partial x_{i}^{2}} \\
dv=\frac{\partial \phi}{\partial x_{\mathrm{i}}}(\mathbf{x}-\xi) \quad &\Rightarrow \quad v=\phi(\mathbf{x}-\xi)
\end{aligned}
$$

$$\int_{\mathbb{R}^{\mathrm{p}}} (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x} = -\left[\phi(\mathbf{x} - \xi) \left(\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right) \Big|_{-\infty}^{\infty} \right.$$
$$\left. - \int_{\mathbb{R}^{\mathrm{p}}} \left(\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial^2 \mathrm{p}}{\partial {x_{\mathrm{i}}}^2} (\mathbf{x}) - \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \frac{\frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x})}{\mathrm{p}(\mathbf{x})^2} \right) \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x} \right]$$
$$= \int_{\mathbb{R}^{\mathrm{p}}} \left[\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial^2 \mathrm{p}}{\partial {x_{\mathrm{i}}}^2} (\mathbf{x}) - \left(\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right)^2 \right] \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x}.$$

From this identity follows that,

$$\int_{\mathbb{R}^{\mathrm{p}}} \left(\frac{1}{\mathrm{p}} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right)^2 \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^{\mathrm{p}}} \left[\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial^2 \mathrm{p}}{\partial {x_{\mathrm{i}}}^2} (\mathbf{x}) \phi(\mathbf{x} - \xi) \right.$$
$$\left. - \frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \phi(\mathbf{x} - \xi) (\mathrm{x_i} - \xi_{\mathrm{i}}) \right] \mathrm{d}\mathbf{x}$$
$$= \int_{\mathbb{R}^{\mathrm{p}}} \frac{1}{\mathrm{p}(\mathbf{x})} \left[\frac{\partial^2 \mathrm{p}}{\partial {x_{\mathrm{i}}}^2} (\mathbf{x}) + (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right] \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x}$$

Replacing this equality in the expression $\mathrm{R}_{\mathbf{T}}(\xi) - \mathrm{R}_{\delta}(\xi)$ we have

$$\mathrm{R}_{\mathbf{T}}(\xi) - \mathrm{R}_{\delta}(\xi) = \int_{\mathbb{R}^{\mathrm{p}}} \sum_{\mathrm{i}=1}^{\mathrm{p}} \left[\left(\frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right)^2 + 2 (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right] \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x}$$
$$= \int_{\mathbb{R}^{\mathrm{p}}} \sum_{\mathrm{i}=1}^{\mathrm{p}} \left[\frac{1}{\mathrm{p}(\mathbf{x})} \left(\frac{\partial^2 \mathrm{p}}{\partial x_i^2} (\mathbf{x}) - (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right) \phi(\mathbf{x} - \xi) \right.$$
$$\left. + 2 (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{1}{\mathrm{p}(\mathbf{x})} \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \phi(\mathbf{x} - \xi) \right] \mathrm{d}\mathbf{x}$$
$$= \int_{\mathbb{R}^{\mathrm{p}}} \sum_{\mathrm{i}=1}^{\mathrm{p}} \frac{1}{\mathrm{p}(\mathbf{x})} \left[\frac{\partial^2 \mathrm{p}}{\partial x_i^2} (\mathbf{x}) + (\mathrm{x_i} - \xi_{\mathrm{i}}) \frac{\partial \mathrm{p}}{\partial x_{\mathrm{i}}} (\mathbf{x}) \right] \phi(\mathbf{x} - \xi) \, \mathrm{d}\mathbf{x}.$$

Note that $\mathrm{p}(\mathbf{x}) = \int_{\mathrm{C}} \phi(\mathbf{x} - \theta) \mathrm{d}\theta$, $\frac{\partial \mathrm{p}}{\partial x_i} (\mathbf{x}) = -\int_{C} \phi(\mathbf{x} - \theta) (x_i - \theta_i) d\theta$,

$$\frac{\partial^2 \mathrm{p}}{\partial x_i^2} (\mathbf{x}) = -\int_{C} \left[-\phi(\mathbf{x} - \theta) (x_i - \theta_i)^2 + \phi(\mathbf{x} - \theta) \right] d\theta$$
$$= \int_{C} \phi(\mathbf{x} - \theta) \left[(x_i - \theta_i)^2 - 1 \right] d\theta$$

hence

$$\frac{\partial^2 \mathrm{p}}{\partial x_i^2}(\mathbf{x}) + (x_i - \xi_i)\frac{\partial p}{\partial x_i}(\mathbf{x}) = \int_C \phi(\mathbf{x}-\theta)\left[(x_i-\theta_i)^2 - 1\right]d\theta - (x_i-\xi_i)\int_C \phi(\mathbf{x}-\theta)(x_i-\theta_i)d\theta$$

$$= \int_C \left\{\left[(x_i-\theta_i)^2 - 1\right]\phi(\mathbf{x}-\theta) - (x_i-\xi_i)(x_i-\theta_i)\phi(\mathbf{x}-\theta)\right\}d\theta. \quad (2)$$

Using the following equality

$$\begin{aligned}\frac{\partial \phi}{\partial \theta_i}(\mathbf{x}-\theta) &= -\phi(\mathbf{x}-\theta)(x_i-\theta_i)\\ \frac{\partial^2 \phi}{\partial \theta_i^2}(\mathbf{x}-\theta) &= \frac{\partial \phi}{\partial \theta_i}(\mathbf{x}-\theta)(x_i-\theta_i) - \phi(\mathbf{x}-\theta)\\ &= \phi(\mathbf{x}-\theta)(x_i-\theta_i)^2 - \phi(\mathbf{x}-\theta) = \left[(x_i-\theta_i)^2-1\right]\phi(\mathbf{x}-\theta),\end{aligned}$$

replacing in equation (2) and sum in i we have

$$\sum_i \left[\frac{\partial^2 p}{\partial x_i^2}(\mathbf{x}) + (x_i-\xi_i)\frac{\partial p}{\partial x_i}(\mathbf{x})\right] = \int_C \sum_i \left[\frac{\partial^2 \phi}{\partial \theta_i^2}(\mathbf{x}-\theta) + (x_i-\xi_i)\frac{\partial \phi}{\partial \theta_i}(\mathbf{x}-\theta)\right]d\theta.$$

If $g(x_1,...,x_p) = (g_1(x_1,...,x_p),...,g_p(x_1,...,x_p))$ is a vector field, then the divergence theorem ensure us that the volume integral of the divergent of the field in a ball is equal to surface integral given by the inner product of the field with unit normal field

$$\int_C \nabla \cdot g dv = \int_{\partial C} (g \cdot \eta) ds.$$

Considering the vector field

$$\left(\frac{\partial \phi}{\partial \theta_1}(\mathbf{x}-\theta) - (x_1-\xi_1)\phi(\mathbf{x}-\theta), ..., \frac{\partial \phi}{\partial \theta_p}(\mathbf{x}-\theta) - (x_p-\xi_p)\phi(\mathbf{x}-\theta)\right) = \mathsf{grad}\phi(\mathbf{x}-\theta) - (\mathbf{x}-\xi)\phi(\mathbf{x}-\theta),$$

in which $\mathtt{grad}\phi(\mathbf{x}-\theta)$ is the gradient field of the function $\phi(\mathbf{x}-\theta)$. Therefore $\mathtt{grad}\phi(\mathbf{x}-\theta) = (\mathbf{x}-\theta)\phi(\mathbf{x}-\theta)$ and the gradient field can be expressed as

$$\mathsf{grad}\phi(\mathbf{x}-\theta) - (\mathbf{x}-\xi)\phi(\mathbf{x}-\theta) = (\xi-\theta)\phi(\mathbf{x}-\theta).$$

Applying the divergence theorem we have equality

$$\int_C \nabla \cdot (\xi-\theta)\phi(\mathbf{x}-\theta)dv = \int_{S^{p-1}} (\xi-\theta)\phi(\mathbf{x}-\theta)\cdot\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)ds$$

in which $\eta(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ is the unit normal field to sphere S^{p-1}. Now it is possible to have an interesting geometrical interpretation as described on Figure 2. The angle between the vector $\xi - \theta$ and the unit normal vector, for any ξ in the interior of the ball and any vector θ in the border S^{p-1}, is bigger than 90^0, so the inner product of these vectors is always negative that is, $(\xi - \theta) \cdot \left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) < 0$. So

$$\mathrm{R}_{\mathbf{T}}(\xi) - \mathrm{R}_{\delta}(\xi) = \sum_{i=1}^{p} \left[\int_{\mathbb{R}^p} \frac{1}{p(\mathbf{x})} \left[\int_{S^{p-1}} (\xi - \theta)\phi(\mathbf{x} - \theta) \cdot \left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right) ds \right] \phi(\mathbf{x} - \xi) d\mathbf{x} \right] < 0$$

and hence $\mathbf{T}$ estimator dominates mean estimator δ.

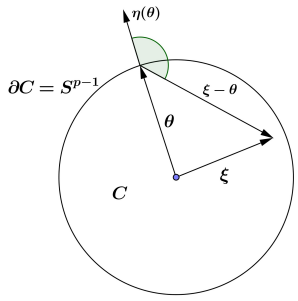

Figure 2 - The angle between $\eta(\theta)$ and $\xi - \theta$.

2.1 Some new results

Comparison between Bayes estimator $\mathbf{T}$ and the usual estimator $\delta(X) = X$ is not really fair, because δ estimator does not consider that the mean vector parameter is bounded. A more suitable estimator to compare is the maximum likelihood estimator in relation to parametric restriction.

$$\delta_m^{ML}(\mathbf{x}) = \begin{cases} m\dfrac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| \geq m \\ \mathbf{x}, & \|\mathbf{x}\| < m. \end{cases}$$

A Hartigan-kind integral formula for the difference between the risks of these estimators would be excessively complex. However, an interesting observation is that the comparison can be obtained through the analysis of unidimensional case developed by Gatsonis et al. (1987). They also considered an uniform priori in the interval $[-m, m]$, getting the Bayes estimator

$$\delta_m(x) = \frac{\int_{-m}^{m} \theta\phi(x - \theta)\, d\theta}{\int_{-m}^{m} \phi(x - \theta)\, d\theta}.$$

δ_m dominates the usual δ estimator, and dominates the maximum likelihood unit estimator δ_m^{ML} for θ in an interval close to $\left[-\frac{3}{4}m,\ \frac{3}{4}m\right]$.

To use the unidimensional result consider the case in which the parametric restriction defines a hypercube centered at origin, i.e. $\theta = (\theta_1, ..., \theta_p)$ with $|\theta_i| \leq m$ $i = 1, ..., p$. Observe that for practical applications this restriction can be more natural than $\|\theta\| \leq m$. In this case, Bayes estimator $\mathbf{T}$ has components in the form

$$T_i(x) = \frac{\int\limits_C \theta_i \phi(x-\theta)\, d\theta}{\int\limits_C \phi(x-\theta)\, d\theta} = \frac{\int\limits_C \theta_i \prod\limits_{j=1}^{p} \phi(x_j - \theta_j) d\theta}{\int\limits_C \prod\limits_{j=1}^{p} \phi(x_j - \theta_j) d\theta}$$

$$= \frac{\int_{-m}^{m} \theta_i \phi(x_i - \theta_i) d\theta_i \int_{-m}^{m} ... \int_{-m}^{m} \prod\limits_{\substack{=1 \\ \neq i}}^{p} \phi(x_j - \theta_j) d\theta}{\int_{-m}^{m} \phi(x_i - \theta_i) d\theta_i \int_{-m}^{m} ... \int_{-m}^{m} \prod\limits_{\substack{=1 \\ \neq i}}^{p} \phi(x_j - \theta_j) d\theta}$$

$$= \frac{\int_{-m}^{m} \theta_i \phi(x_i - \theta_i) d\theta_i}{\int_{-m}^{m} \phi(x_i - \theta_i) d\theta_i} = \delta_m(x_i).$$

Then follows that Hartigan estimator for the hypercube is expressed in terms of the unidimensional estimator obtained by Gatsonis et al. (1987) as $T(\mathbf{x}) = (\delta_{\mathbf{m}}(\mathbf{x_1}), ..., \delta_{\mathbf{m}}(\mathbf{x_p}))$. this estimator will be denoted by $\delta_m(\mathbf{x})$. In this case the difference is

$$\begin{aligned} \mathrm{R}_{\mathbf{T}}(\xi) - \mathrm{R}_{\delta_m^{ML}}(\xi) &= E\left[\|T(x) - \xi\|^2\right] - E\left[\left\|\delta_m^{ML}(x) - \xi\right\|^2\right] \\ &= \sum_{i=1}^{p} E\left[|T_i(x) - \xi_i|^2\right] - E\left[\left|\delta_m^{ML}(x_i) - \xi_i\right|^2\right] \\ &= \sum_{i=1}^{p} E\left[|\delta_m(x_i) - \xi_i|^2\right] - E\left[\left|\delta_m^{ML}(x_i) - \xi_i\right|^2\right]. \end{aligned}$$

By the result of Galtonis et al. (1987) for $-\frac{3}{4}m \leq x_i \leq \frac{3}{4}m$, we have $\mathrm{R}_{\mathbf{T}}(\xi) - \mathrm{R}_{\delta_m^{ML}}(\xi) \leq 0$ and therefore $\mathbf{T}$ dominates δ_m^{ML}. The case where the restriction is a sphere will be studied through computational simulation.

2.2 A study through computational simulation

Aiming to relate Hartigan's result (2004) with the estimators studied in Casella and Strawderman (1981) e Gatsonis et al (1987) a computational simulation in

dimension 2 was performed. In these papers, the estimators

$$\begin{aligned} \delta_m^{ML} \quad &, \quad \text{Maximum likelihood restricted estimator} \\ \delta_m(x) = x + \frac{g'(x)}{g(x)} \quad &, \quad g(x) = \Phi(m-x) + \Phi(m+x) - 1 \\ \delta_m^0(x) = m\ tanh(mx) & \end{aligned}$$

Φ is the distribution function of the standard normal and the mean limited in interval $(-m, m)$.
The results was:

- δ_m^0 is minimax for $0 \leq m \leq m_0 \simeq 1,056742$ and dominates δ_m^{ML} for $m < 1$.
- δ_m dominates usual mean estimator, e.g, it has risk smaller than 1.
- δ_m dominates δ_m^{ML} in the approximated interval $\left(-\frac{3}{4}m, \frac{3}{4}m\right)$.

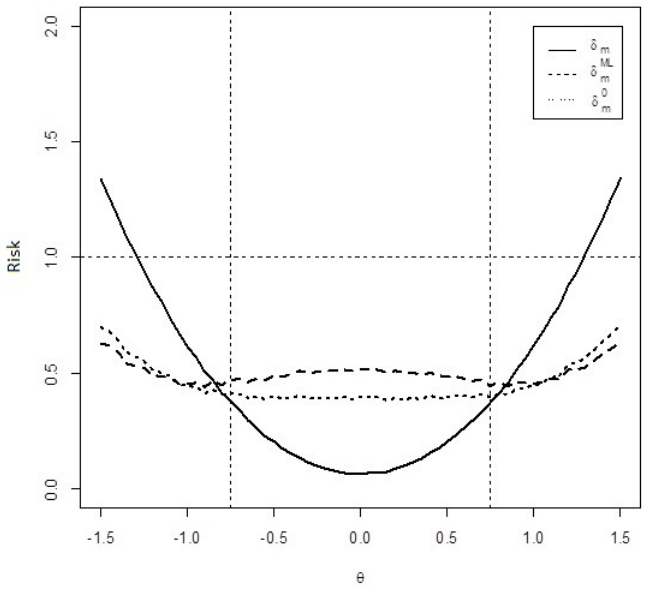

Figure 3 - Estimator risk for dimension 1 and m=1

For a multivariate normal mean case with the vector mean limited to a sphere in $\mathbb{R}^p$ with radius m these results can be generalized:

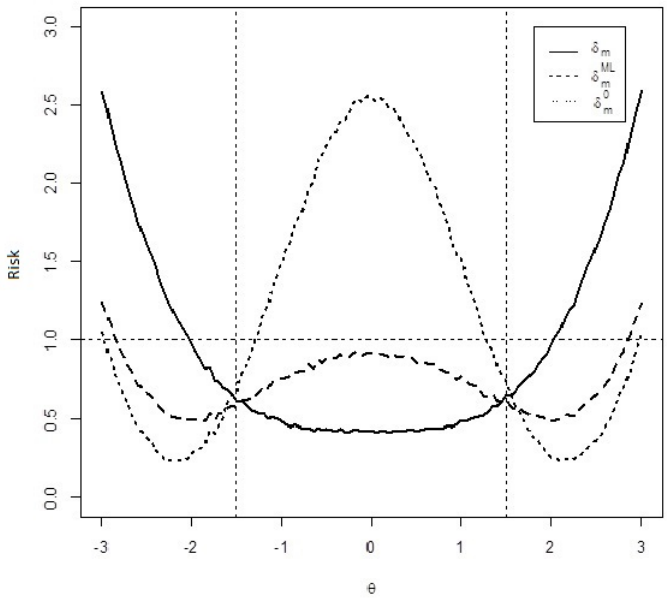

Figure 4 - Estimator risk for dimension 1 and m=2

- Hartigan **T** estimator generalizes estimator δ_m.
- Estimator δ_m^0 will be generalized in the form

$$\delta_m^0(\mathbf{x}) = (m\,tanh(mx_1), ..., m\,tanh(mx_p))\,.$$

- δ_m can be also generalized as $\delta_m(\mathbf{x}) = (\delta_m(x_1), ..., \delta_m(x_p))$.

Observe that estimators $\delta_m^0(\mathbf{x})$ and $\delta_m(\mathbf{x})$ are in fact generating estimatives on a p-dimensional m sided hypercube,$\{(x_1, ..., x_p) \in \mathbb{R}^p,\ -m \leq x_i \leq m,\ i = 1, ..., p.$ However these estimatives more concentrated in the m radius sphere inside the hipercube. Therefore it makes sense to compare $\delta_m^0(\mathbf{x})$ and $\delta_m(\mathbf{x})$ with the estimators δ_m^{ML} and **T** which generate only estimatives in the sphere. Since there isn't an explicit formula for the risks of these estimators the comparation were obtained through simulation.

Simulation was carried out at is software R environment. A reticulate was built in the sphere for values $\theta = (\theta_1, \theta_2)$. For each value of θ a sample of the two dimensional normal $N_2\,(\theta, I)$) is generated. Then, it is obtained a discret version of the risk graphics. For the analisy, bidimensional sections of these graphics in the trdimensional space were plotted Figuras (6, 7 and 8).

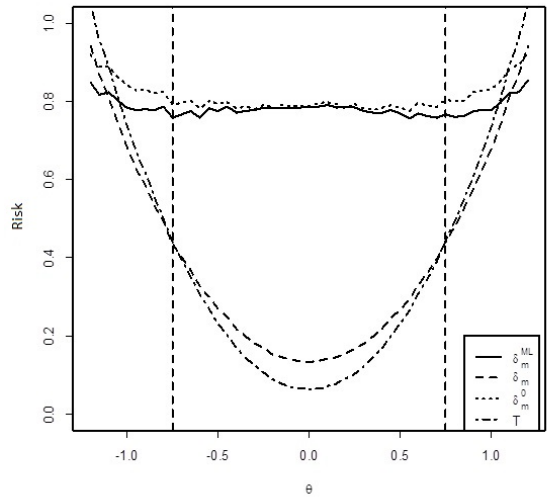

Figure 5 - Estimator risk for dimension 2 and m= 1.

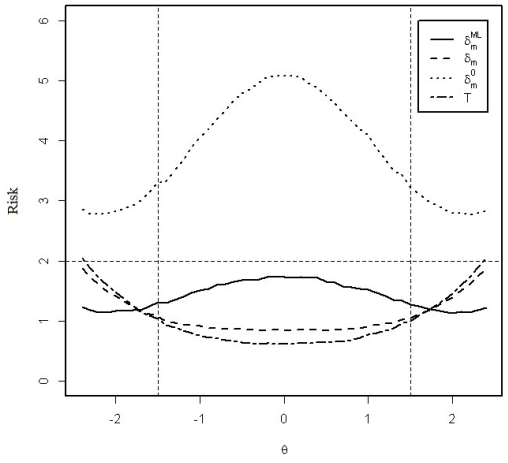

Figure 6 - Estimator risk for dimension 2 and m= 2.

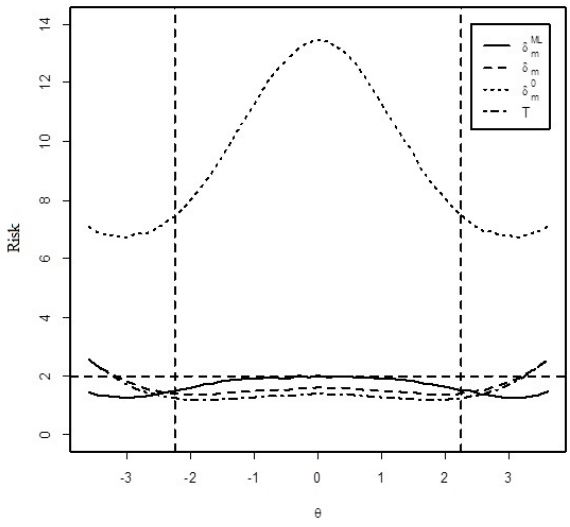

Figure 7 - Estimator risk for dimension 2 and m= 3.

Through graphic analysis we obtain:

- Unidimensional minimax estimator δ_m^0 when generalized to dimension 2 completely loses its relative advantages to other estimators.

- Hartigan estimator $\mathbf{T}(\mathbf{x})$, dominates $\delta_m((\mathbf{x}))$ for the sphere of radius $m = 3$. For radius 2 and 1 and θ close to the border of the sphere an inversion happens with $\delta_m(\mathbf{x})$ dominating $\mathbf{T}(\mathbf{x})$.

- Hartigan estimator $\mathbf{T}(\mathbf{x})$ dominates the maximum likelihood estimator $\delta_m^{ML}(\mathbf{x})$ except when θ is close to the sphere's border. This fact was intuitively expected since, $\delta_m^{ML}(\mathbf{x})$ tends to produce estimates on the border.

3 Conclusion

The mathematical theory of Hartigan's estimator is accessible and for the sphere has a geometrycal meaning. The properties of unidimensional estimators for the bounded mean changes when generalized to the two dimensional space.

References

CASELLA, G.; STRAWDERMAN,W. Estimating a bounded normal mean. *The Annals Statistics*, v.9, p.870-878, 1981.

GATSONIS, C.; MACGIBBON,B.; STRAWDERMAN,W. On the estimation of a restricted normal mean. *Statistics and Probability Letters*, v.6, p.21-30, 1987.

HARTIGAN, J.A. Uniform prioris on convex sets improve risk. *Statistics and Probability Letters*, v.67, p.285-288, 2004.

APÊNDICE C - Simulação Computacional para o Artigo 1

```
# Simulação para normais tridimensionais com o objetivo de comparar o EQM
# do estimador média amostral com o estimador de James-Stein modificado
# m1: média 1
# m2: média 2
# m3: média 3
# s21: variância 1
# s22: variância 2
# s23: variância 3
# ss: tamanho da amostra

inicio=Sys.time()
NormMult3 <- function(m1,m2,m3,S21,S22,S23,ss){
x = as.vector(1:9)
x[1] <- rnorm(1,m1,sqrt(S21/ss)) #  First component mean
x[2] <- rnorm(1,m2,sqrt(S22/ss)) #  Second component mean
x[3] <- rnorm(1,m3,sqrt(S23/ss)) #  Third component mean
x[4] <- x[1]^2+x[2]^2+x[3]^2 #  The squared norm of the vector of means
x[5] <- (1-S21/ss/(x[4]))*x[1]     # Shrinks the first component
x[6] <- (1-S22/ss/(x[4]))*x[2] # Shrinks the second component
x[7] <- (1-S23/ss/(x[4]))*x[3] # Shrinks the third component
x[8] <- (x[1]-m1)^2+(x[2]-m2)^2+(x[3]-m3)^2
x[9] <- (x[5]-m1)^2+(x[6]-m2)^2+(x[7]-m3)^2
x
list(eqmx = x[8], eqmshr = x[9])
}
S21 <- 10
S22 <- 10
S23 <- 10
ss <- 1
mu1 = (-10:10)
```

```
mu2 = (-10:10)
mu3 = (-10:10)

NL = length(mu1)*length(mu2)*length(mu3)
resu <- matrix(0,NL,6)
colnames(resu) <- c("média 1","média 2","média 3","EQMX","EQMShr",
"Indicador EQMShr<EQM")
L <- 0
nrep <- 100000
for (i in 1:length(mu1)){
    for (j in 1:length(mu2)){
       for (k in 1:length(mu3)){
          L <- L + 1
           resu[L,1] <- mu1[i]
            resu[L,2] <- mu2[j]
            resu[L,3] <- mu3[k]
            for (z in 1:nrep){
                res <- NormMult3(mu1[i],mu2[j],mu3[k],S21,S22,S23,ss)
                resu[L,4] <- resu[L,4] + res$eqmx/nrep
                resu[L,5] <- resu[L,5] + res$eqmshr/nrep
            }
           if(resu[L,5]< resu[L,4]) resu[L,6] <- 1
        }
     }
}
sum(resu[,6])/NL*100
min(resu[,4])
mean(resu[,4])
max(resu[,4])
min(resu[,5])
mean(resu[,5])
max(resu[,5])
```

APÊNDICE D - Simulação Computacional para o artigo 1- Transformação de Mahalanobis

```
# Simulação para normais tridimensionais com o objetivo de
# comparar o EQM do  estimador obtido com a transformação
# de Mahalanobis com o estimador média amostral
# p:  valor atribuído para a correlação entre pares de variáveis

# Função que calcula potência de matrizes
PotMat <- function(alpha, A)
{
  av <- eigen(A)
  kern <- av$values^alpha
  res <- av$vectors%*%diag(kern)%*%t(av$vectors)
  return(res)
}

#  Gerar a matriz sigma = M
p <-0.5
rho <- matrix( c(1,p,p,p,1,p,p,p,1), 3,3)
S21 <- 10
S22 <- 10
S23 <- 10
V <- matrix(c(S21,0,0,0,S22,0,0,0,S23),3,3)
Vroot <- PotMat(1/2,V)
M <- Vroot%*%rho%*%Vroot
inM <- solve(M)

# Função para comparar o EQM dos estimadores
NormMult3 <- function(m1,m2,m3,S21,S22,S23,ss){
  x = as.vector(1:9)
  x[1] <- rnorm(1,m1,sqrt(S21/ss))
  x[2] <- rnorm(1,m2,sqrt(S22/ss))
```

```
  x[3] <- rnorm(1,m3,sqrt(S23/ss))
  y <- x[1:3]
  x[4] <- t(y) %*% inM %*% y
  x[5] <- (1-1/(x[4]))* x[1]
  x[6] <- (1-1/(x[4]))* x[2]
  x[7] <- (1-1/(x[4]))* x[3]
  x[8] <- (x[1]-m1)^2+(x[2]-m2)^2+(x[3]-m3)^2
  x[9] <- (x[5]-m1)^2+(x[6]-m2)^2+(x[7]-m3)^2
  x
  list(eqmx = x[8], eqmshr = x[9])
}

ss <- 1
mu1 = m1 = (-10:10)
mu2 = m2 = (-10:10)
mu3 = m3 = (-10:10)
NL = length(mu1)*length(mu2)*length(mu3)
resu <- matrix(0,NL,6)
colnames(resu) <- c("média 1","média 2","média
 3","EQMX","EQMShr","Indicador EQMShr<EQM")
L <- 0
nrep <- 100000
for (i in 1:length(mu1)){
  for (j in 1:length(mu2)){
    for (k in 1:length(mu3)){
      L <- L + 1
      resu[L,1] <- mu1[i]
      resu[L,2] <- mu2[j]
      resu[L,3] <- mu3[k]
      for (z in 1:nrep){
        res <- NormMult3(mu1[i],mu2[j],mu3[k],S21,S22,S23,ss)
        resu[L,4] <- resu[L,4] + res$eqmx/nrep
```

```
        resu[L,5] <- resu[L,5] + res$eqmshr/nrep
        }
      if(resu[L,5]< resu[L,4]) resu[L,6] <- 1
    }
  }
}

sum(resu[,6])/NL*100
min(resu[,4])
mean(resu[,4])
max(resu[,4])
min(resu[,5])
mean(resu[,5])
max(resu[,5])
```