FILIPE DAS NEVES RIZZO

ABORDAGEM BAYESIANA PARA ESTIMADORES DE ENCOLHIMENTO

LAVRAS - MG

2014

FILIPE DAS NEVES RIZZO

ABORDAGEM BAYESIANA PARA ESTIMADORES DE ENCOLHIMENTO

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

Orientador
Dr. Devanil Jaques de Souza

Coorientador
Dr. Lucas Monteiro Chaves

LAVRAS - MG
2014

Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e Serviços da Biblioteca Universitária da UFLA

Rizzo, Filipe das Neves.
Abordagem Bayesiana para estimadores de encolhimento / Filipe das Neves Rizzo. – Lavras : UFLA, 2014.
69 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2014.
Orientador: Devanil Jaques de Souza.
Bibliografia.

1. Erro quadrático médio. 2. Estimador. 3. Encolhimento. I. Universidade Federal de Lavras. II. Título.

CDD – 519.542

FILIPE DAS NEVES RIZZO

ABORDAGEM BAYESIANA PARA ESTIMADORES DE ENCOLHIMENTO

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

APROVADA em 17 de julho de 2014.

Dra. Maria do Carmo Pacheco de Toledo Costa UFLA

Dra. Thelma Sáfadi UFLA

Dr. Devanil Jaques de Souza
Orientador

LAVRAS - MG
2014

AGRADECIMENTOS

A Deus, meu apoio, refúgio, consolador e razão da minha força para prosseguir até o fim, mesmo com todas as dificuldades encontradas no meio do caminho.

Aos meus pais, Edson e Solange, pela dedicação e incentivo para continuar os estudos.

Ao meu irmão Victor pelo incentivo e principalmente ao meu filho Pedro que, sem perceber, foi responsável por sempre me incentivar a continuar seguindo em frente, mesmo sabendo das dificuldades.

A minha namorada e futura esposa Clarissa Loura por estar sempre ao meu lado me apoiando nos momentos de fraqueza.

Aos colegas de mestrado da Estatística e Experimentação Agropecuária, grandes amigos nos momentos difíceis e de descontração.

Aos professores envolvidos no trabalho, professora Maria do Carmo Pacheco de Toledo Costa e professora Thelma Sáfadi.

Em especial aos Professores Devanil Jaques de Souza e Lucas Monteiro Chaves pelos ensinamentos diários. Através dos senhores me tornei um profissional melhor.

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

Aos professores e funcionários dos Departamento de Ciências Exatas.

RESUMO

O estatístico-matemático Charles Stein, em 1955, em uma publicação denominada "***Inadmissibility of the usual estimator for the mean of a multivariate normal distribution***" (GRUBER, 1998) surpreendeu o mundo da estatística com sua prova de que o estimador de máxima verossimilhança é inadmissível, salvo nos casos unidimensional e bidimensional. Stein mostrou que, caso se admita um estimador viesado, há estimadores com erro quadrático médio inferior ao erro quadrático médio do estimador de máxima verossimilhança. Esses estimadores compõem a classe dos chamados estimadores de encolhimento (*shrinkage*). Esses estimadores têm, em geral, erro quadrático médio menor que os estimadores usuais, como mostraremos no decorrer deste trabalho.

Palavras-chave: *Erro quadrático médio*. Estimador. Encolhimento.

ABSTRACT

The statistician-mathematician Charles Stein, in 1955, in a publication ***"Inadmissibility of the usual estimator for the mean of a multivariate normal distribution"*** (GRUBER, 1998) surprised the world of statistics with its proof that the maximum likelihood estimator is inadmissible, except in the one-dimensional and two-dimensional cases. Stein showed that, in case a biased estimator is admitted, there are estimators with mean square error inferior to the mean square error of the maximum likelihood estimator. These estimators comprise a class denominated shrinkage estimators. These estimators have, in general, mean square error lower than the usual estimators, as will be presented over the course of this work.

Keywords: Mean square error. Estimator. Shrinkage.

LISTA DE FIGURAS

SUMÁRIO

1 INTRODUÇÃO

O estatístico-matemático Charles Stein, em 1955, em uma publicação denominada "***Inadmissibility of the usual estimator for the mean of a multivariate normal distribution***" (GRUBER, 1998) surpreendeu o mundo da estatística com sua prova de que o estimador de máxima verossimilhança é inadmissível, salvo nos casos unidimensional e bidimensional. Stein mostrou que, caso se admita um estimador viesado, há estimadores com erro quadrático médio inferior ao erro quadrático médio do estimador de máxima verossimilhança. Esses estimadores compõem a classe dos chamados estimadores de encolhimento (*shrinkage*).

Como objetivo geral, pretende-se neste estudo verificar a aplicabilidade de alguns dos estimadores de encolhimentos e compará-los, em termos do erro quadrático médio, com os estimadores usuais. Estes estimadores podem ser obtidos por uma abordagem Bayesiana, como pretende-se mostrar no decorrer deste trabalho.

Para os objetivos específicos, pretende-se analisar o estimador viesado e o não viesado da variância através do erro quadrático médio. Comparar o estimador usual com o estimador de encolhimento utilizando do Erro Quadrático Médio como medida base de comparação. Também como objetivo específico, espera-se que a simulação mostre um melhor caminho para a escolha do estimador. Os estimadores de encolhimento, em geral, têm erro quadrático médio menor que os estimadores usuais, como pretende-se mostrar no decorrer deste trabalho e também através da simulação.

Justifica-se a escolha do tema em estudo em razão de sua aplicabilidade em várias áreas do conhecimento onde se destaca a genética. Torna-se, assim, de suma importância aprofundar os conhecimentos em um tema relacionado aos estudos voltados para a estatística e experimentação agropecuária. Esta é uma área que tem como meta qualificar e aprimorar os conhecimentos para que o profissional possa atuar como docente ou pesquisador em instituições de ensino públicas ou privadas. Sendo assim, torna-se relevante buscar temas que contribuam para garantir o alcance da meta almejada.

2 REFERENCIAL TEÓRICO

2.1 Erro Quadrático Médio (EQM)

Definição: O Erro Quadrático Médio (EQM) de um estimador W de um parâmetro θ é a função de θ definida por $E_\theta[(W-\theta)^2]$.

O Erro Quadrático Médio mede a média das diferenças quadráticas entre as estimativas obtidas por W e o parâmetro θ. O EQM tem, pelo menos, duas vantagens em relação a outras medidas de distância: primeiro ele é tratável analiticamente e, segundo, pode ser reescrito como

$$E_\theta[(W-\theta)^2] = Var_\theta[W] + (E_\theta(W-\theta))^2$$

o que permite a sua interpretação como a soma de duas parcelas: a variância e o quadrado do viés.

A primeira parcela, $Var_\theta[W]$, mede a variabilidade do estimador, conhecida como precisão, e a segunda parcela, $(vies_\theta[W])^2$, é denominada de exatidão. Dizemos que W é um estimador não viesado quando $E[W] = \theta$. Podemos destacar também que, para encontrar um estimador com boas propriedades de EQM, precisamos encontrar estimadores que controlem a variância e o viés.

2.2 Inferência Bayesiana

A estatística clássica supõe que o parâmetro de uma distribuição é um número e utiliza uma amostra aleatória para fazer inferência sobre esse número. No contexto de inferência Bayesiana, supõe-se que o parâmetro da distribuição seja uma variável aleatória e que todo o conhecimento sobre essa variável aleatória esteja contido em uma distribuição de probabilidade, chamada de distribuição a priori. Conhecida uma amostra aleatória, o procedimento Bayesiano consiste em obter uma nova distribuição para o parâmetro, a distribuição a posteriori, que incorpore à priori a informação contida na amostra aleatória.

Teorema de Bayes

Considere θ um parâmetro desconhecido da distribuição de uma variável aleatória X. Para obtermos $f_{\Theta|X}(\theta|X)$, função de probabilidade de θ dado X, aplica-se a definição de probabilidade condicional, ou seja,

$$f_{\Theta|X}(\theta|x) = \frac{f_{\Theta,X}(\theta,x)}{f_X(x)}$$

Como $f_{\Theta,X}(\theta,x) = f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)$ segue que

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta)f_{\Theta}(\theta)}{f_X(x)}$$

A função $f_{X|\Theta}(x|\theta)$ é denominada a função de verossimilhança. A função $f_{\Theta}(\theta)$, que será denotada $\pi(\theta)$, representa a distribuição a priori e a função $f_{\Theta|X}(\theta|x)$, que será denotada $\pi(\theta|x)$, representa a distribuição a posteriori. O termo $\frac{1}{f_X(x)}$ é a constante normalizadora e $f_X(x)$ é a marginal. Denotando $f_X(x)$ por $g(x)$, o Teorema de Bayes se expressa por

$$\pi(\theta|x) = \frac{f(x|\theta)\,\pi(\theta)}{g(x)}$$

com

$$g(x) = \int_{-\infty}^{+\infty} f(x|\theta)\,\pi(\theta)d\theta$$

Observe que $g(x)$ só depende dos dados. Em outras palavras, a distribuição a posteriori é proporcional à função de verossimilhança multiplicada pela distribuição a priori.

$$\pi(\theta|x) \propto \underbrace{f(x|\theta)}_{Verossimilhanca} \underbrace{\pi(\theta)}_{Priori} \tag{2.1}$$

A distribuição a priori combinada com a função de verossimilhança nos fornece a distribuição a posteriori $\pi(\theta|x)$ visto em (2.1). A distribuição a posteriori incorpora a informação da amostra. Com a distribuição a posteriori pode-se estimar o parâmetro θ como a média, moda ou mediana da mesma. Neste trabalho vamos utilizar a média da posteriori, denominada de estimador de Bayes, para estimar os parâmetros em estudo.

Na maioria das vezes pode-se prescindir da função $g(x)$ e considerar, simplesmente, que a posteriori é proporcional à verossimilhança multiplicada pela priori, o chamado núcleo da posteriori.

2.3 Estimador de Bayes

Apresentaremos a seguir as definições da função perda, função risco e risco de Bayes, para finalmente apresentarmos o estimador de Bayes.

Definição: ***(Função Perda)*** Considere estimar $\tau(\theta)$. Seja $T = t(X_1,X_2,...,X_n)$ um estimador de $\tau(\theta)$. Define-se como função perda desse estimador, com notação $l(t;\theta)$, qualquer função real que satisfaça:

(i) $l(t;\theta) \geq 0$ para todos os possíveis valores t e todo θ no espaço paramétrico.

(ii) $l(t;\theta) = 0$ para $t = \tau(\theta)$

A função $l(t;\theta)$ representa a perda decorrente de se estimar $\tau(\theta)$ por T, supondo θ o verdadeiro valor do parâmetro.

Definição: ***(Função Risco)*** Para uma dada função perda $l(\cdot;\cdot)$ a função risco de um estimador $T = t(X_1,X_2,...,X_n)$ denotada por $\Re_t(\theta)$, é definida por

$$\begin{aligned}\Re_t(\theta) &= E_\theta[l(t;\theta)] \\ &= E_\theta[l(t(X_1,X_2,...,X_n);\theta)] \\ &= \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} l(t(X_1,X_2,...,X_n);\theta) \prod_{i=1}^{n} f(x_i;\theta)\, dx_1 \ldots dx_n\end{aligned}$$

Definição: ***(Risco de Bayes)*** O risco de Bayes do estimador $T = t(X_1,...,X_n)$ com relação à função perda $l(t;\theta)$ e à distribuição priori acumulada $G(.)$, denotado por $r_{l,G}(t)$, é definido por

$$r_{l,G}(t) = \int_\Theta \Re_t(\theta) dG(\theta) d\theta = \int_\Theta \Re_t(\theta) g(\theta) d\theta$$

Ou, seja, o risco de Bayes de um estimador é o risco médio no espaço paramétrico Θ.

Definição: ***(Estimador de Bayes)*** O estimador de Bayes de $\tau(\theta)$ denotado de $T^*_{l,G} = t^*_{l,G}(X_1,\ldots,X_n)$, em relação à função perda $l(.;.)$ e à distribuição priori acumulada $G(.)$, é definido como sendo o estimador com menor risco de Bayes, isto é, o estimador $t^*_{l,G}$ que satisfaz

$$r_{l,G}(t^*_{l,G}) \leq r_{l,G}(t)$$

para todo estimador $T = t(X_1,\ldots,X_n)$ de $\tau(\theta)$.

Para uma função perda não quadrática, encontrar o estimador de Bayes pode ser complicado. No entanto, para perda quadrática esse problema pode ser relativamente simples. Nesse sentido vamos procurar um estimador $t(X_1,\ldots,X_n)$

que minimize a expressão

$$\int\limits_{\Theta} \Re_t(\theta)g(\theta) = \int\limits_{\Theta} E_\theta\left[[t(X_1,\ldots,X_n) - \tau(\theta)]^2\right]g(\theta)d\theta$$

Agora,

$$\begin{aligned}
&\int\limits_{\Theta} E_\theta\left[[t(X_1,\ldots,X_n) - \tau(\theta)]^2\right]g(\theta)d\theta \\
&= \int\limits_{\Theta}\left\{\int\limits_{\Omega}[t(x_1,\ldots,x_n) - \tau(\theta)]^2 f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta)\prod_{i=1}^{n} dx_i\right\}g(\theta)d\theta \\
&= \int\limits_{\Omega}\left\{\int\limits_{\Theta}[\tau(\theta) - t(x_1,\ldots,x_n)]^2 \frac{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta)g(\theta)}{f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)}d\theta\right\} \\
&.f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\prod_{i=1}^{n} dx_i \\
&= \int\limits_{\Omega}\left\{\int\limits_{\Theta}[\tau(\theta) - t(x_1,\ldots,x_n)]^2 f_{\Theta|X_1=x_1,\ldots,X_n=x_n}(\theta|x_1,\ldots,x_n)d\theta\right\} \\
&.f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\prod_{i=1}^{n} dx_i
\end{aligned}$$

com $\Omega = \mathbb{R}^n$

Desde que o integrando não seja negativo, a integral dupla pode ser minimizada se a expressão dentro das chaves é minimizada para cada $x_1,\ldots,x_n$. Mas, a expressão dentro das chaves é a esperança condicional de $[\tau(\Theta) - t(x_1,\ldots,x_n)]^2$ em relação à distribuição a posteriori de Θ, dado $X_1 = x_1,\ldots,X_n = x_n$, que é minimizada para $t^*(x_1,\ldots,x_n)$ igual à esperança condicional de $\tau(\Theta)$ em relação à mesma. Por isso, o estimador de Bayes de $\tau(\Theta)$ em relação à função perda quadrática é dado por

$$E\left[\tau(\Theta)|X_1 = x_1, \ldots, X_n = x_n\right] = \frac{\int \tau(\Theta)\left[\prod_{i=1}^{n} f(x_i|\theta)\right] g(\theta)d\theta}{\int \left[\prod_{i=1}^{n} f(x_i|\theta)\right] g(\theta)d\theta}$$

Exemplo 1

Suponhamos que $X \sim N(\theta,1)$ e que $\theta \sim N(0,1)$. O objetivo é obter o estimador de Bayes para θ. Para isso calcula-se, utilizando o teorema de Bayes, a distribuição a posteriori

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

$$\begin{aligned}
\pi(\theta|\vec{x}) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}\exp\left\{-\frac{1}{2}\theta^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\theta)^2+\theta^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\left(x_i^2-2x_i\theta+\theta^2\right)+\theta^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[n\theta^2+\theta^2-2\theta\sum_{i=1}^{n}x_i\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2(n+1)-2\theta\sum_{i=1}^{n}x_i\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}(n+1)\left[\theta^2-\frac{2\theta\bar{x}}{\left(1+\frac{1}{n}\right)}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\frac{1}{(n+1)}}\left(\theta-\frac{\bar{x}}{\left(1+\frac{1}{n}\right)}\right)^2\right\}
\end{aligned}$$

A posteriori, neste caso, é $N\left(\frac{\bar{x}}{(1+\frac{1}{n})}, \frac{1}{(n+1)}\right)$. O estimador de Bayes para θ é a esperança da posteriori, ou seja, $\hat{\theta} = \frac{\bar{x}}{1+\frac{1}{n}}$. Repare que neste exemplo não foi necessário o cálculo da marginal $g(.)$. Como queríamos a esperança da posteriori, bastou o núcleo da mesma para obtermos essa informação.

Exemplo 2

Seja $\vec{X} = \{X_1, \ldots, X_n\}$ uma amostra aleatória de uma densidade normal com média desconhecida θ e variância conhecida σ_2^2. Assumindo que $\theta \sim N(\mu, \sigma_1^2)$, com μ e σ_1^2 desconhecidos, o objetivo é construir um estimador Bayesiano empírico para θ. Conforme o teorema de Bayes:

$$\pi(\theta|\vec{x}) = \frac{f(\vec{x}|\theta)\pi(\theta)}{g(\vec{x})}$$

em que

$$f(\vec{x}|\theta) = \frac{1}{\left(2\pi\sigma_2^2\right)^{n/2}} \exp\left\{\frac{1}{2\sigma_2^2} \sum_{i=1}^{n} (x_i - \theta)^2\right\}$$

$$\pi(\theta) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{\frac{(\theta-\mu)^2}{2\sigma_1^2}\right\}$$

$$g(\vec{x}) = \int \underbrace{\frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{\frac{(\theta-\mu)^2}{2\,\sigma_1^2}\right\}}_{\pi(\theta)} \underbrace{\frac{1}{\left(2\,\pi\,\sigma_2^2\right)^{n/2}} \exp\left\{\frac{1}{2\sigma_2^2} \sum_{i=1}^{n} (x_i - \theta)^2\right\} d\theta}_{f(\vec{x}|\theta)}$$

$$f(\vec{x}|\theta)\pi(\theta) = \underbrace{\frac{1}{\left(2\,\pi\,\sigma_2^2\right)^{n/2}} \exp\left\{-\frac{1}{2\sigma_2^2} \sum_{i=1}^{n} (x_i - \theta)^2\right\}}_{f(\vec{x}|\theta)} \underbrace{\frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left\{-\frac{(\theta-\mu)^2}{2\,\sigma_1^2}\right\}}_{\pi(\theta)}$$

Chamando

$$\frac{1}{\sigma_1 \sqrt{2\pi}} \frac{1}{\left(2\,\pi\,\sigma_2^2\right)^{n/2}} = k_1(\vec{x})$$

$$= k_1(\vec{x}) \exp\left\{-\frac{\sum_{i=1}^{n}\left(x_i^2 - 2\,x_i\theta + \theta^2\right)}{2\sigma_2^2} - \frac{\theta^2 - 2\theta\mu + \mu^2}{2\,\sigma_1^2}\right\}$$

$$k_2(\vec{x}) = k_1(\vec{x}) \exp\left\{\frac{-1}{2\sigma_1^2\,\sigma_2^2}\left[\sigma_2^2\mu^2 + \sigma_1^2\sum_{i=1}^{n} x_i^2\right]\right\}$$

$$= k_2(\vec{x}) \exp\left\{\frac{-1}{2\sigma_1^2\,\sigma_2^2}\left[\sigma_2^2\left(\theta^2 - 2\theta\mu\right) + \sigma_1^2\left[n\theta^2 - 2\theta\sum_{i=1}^{n} x_i\right]\right]\right\}$$

$$= k_2\,(\vec{x}) \exp\left\{\frac{-1}{2\sigma_1^2\,\sigma_2^2}\left[\sigma_2^2\left(\theta^2 - 2\theta\mu\right) + \sigma_1^2\left[n\theta^2 - 2\theta\sum_{i=1}^{n} x_i\right]\right]\right\}$$

$$= k_2\,(\vec{x}) \exp\left\{\frac{-1}{2\sigma_1^2\,\sigma_2^2}\left[\sigma_2^2\left(\theta^2 - 2\theta\mu\right) + n\sigma_1^2\left[\theta^2 - 2\theta\,\overline{x}\right]\right]\right\}$$

$$= k_2(\vec{x}) \exp\left\{\frac{-\left(\sigma_2^2 + n\sigma_1^2\right)}{2\sigma_1^2\,\sigma_2^2}\left[\theta^2 - 2\theta\left(\frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right)\right]\right\}$$

$$k\,(\vec{x}) = k_2\,(\vec{x}) \exp\left\{\frac{-\left(\sigma_2^2 + n\sigma_1^2\right)}{2\sigma_1^2\,\sigma_2^2}\left[\theta - \left(\frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right)^2\right]\right\} \tag{2.2}$$

$$= k(\vec{x}) \exp\left\{\frac{-\left(\sigma_2^2 + n\sigma_1^2\right)}{2\sigma_1^2\,\sigma_2^2}\left[\theta^2 - 2\theta\left(\frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right) + \left(\frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right)^2\right]\right\}$$

$$= k\,(\vec{x}) \exp\left\{\frac{-1}{2}\left[\frac{\left(\theta - \frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right)^2}{\frac{\sigma_1^2\,\sigma_2^2}{\sigma_2^2 + n\sigma_1^2}}\right]\right\}$$

então

$$\pi(\theta|\vec{x}) \propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta - \frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}\right)^2}{\frac{\sigma_1^2\,\sigma_2^2}{\sigma_2^2 + n\sigma_1^2}}\right]\right\}$$

Logo, a posteriori é uma normal com média $E\left[\theta|\vec{x}\right] = \frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2}$ e variância $Var\left[\theta|\vec{x}\right] = \frac{\sigma_1^2\,\sigma_2^2}{\sigma_2^2 + n\sigma_1^2}$.

Considerando $n = 1$, segue que $\bar{x} = x$ e os "estimadores"para média e variância são, respectivamente

$$E\left[\theta|x\right] = \frac{\sigma_2^2\mu + \sigma_1^2 x}{\sigma_2^2 + \sigma_1^2}$$
$$Var\left[\theta|x\right] = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Podemos observar que os "estimadores"da média e da variância dependem dos parâmetros μ e σ_1^2. No próximo capítulo utilizaremos o método Bayesiano empírico para estimar estes valores.

Exemplo 3

Considere um vetor $X = (X_1, \ldots, X_n)$ em que $X \sim N(\theta, I)$ e $\theta = (\theta_1, \ldots, \theta_n)$ com $\theta_i \sim N(0, a)$. Queremos o estimador de Bayes para o vetor de médias θ. Como a variância a das prioris são desconhecidas tem-se um caso típico de aplicação do estimador Bayesiano empírico. Considere a distribuição de X_i condicionada a θ_i.

$$f(x_i|\theta_i) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \theta_i)^2\right\}$$

com distribuição a priori de θ dada por

$$\pi(\theta_i) = \frac{1}{\sqrt{2\pi a}} \exp\left\{-\frac{1}{2a}{\theta_i}^2\right\}.$$

Aplicando o Teorema de Bayes

$$\pi(\theta_i|x_i) = \frac{f(x_i|\theta_i)\pi(\theta_i)}{g(x_i)}$$

com

$$g(x_i) = \int_{-\infty}^{+\infty} f(x_i|\theta_i)\pi(\theta_i)d\theta_i$$

Calculando $\pi(\theta_i|x_i)g(x_i)$:

$$\begin{aligned}
\pi(\theta_i|x_i)\ g(x_i) &= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x_i-\theta_i)^2\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2a}\theta_i^2\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}(x_i-\theta_i)^2-\frac{\theta_i{}^2}{2a}\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}\left[(x_i-\theta_i)^2+\frac{\theta_i^2}{a}\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}\left(x_i^2-2x_i\theta_i+\theta_i^2+\frac{\theta_i^2}{a}\right)\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}\left[\theta_i^2\left(1+\frac{1}{a}\right)+x_i^2-2x_i\theta_i\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{-\frac{1}{2}\left[\theta_i^2\left(1+\frac{1}{a}\right)-2x_i\theta_i\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{-\frac{1}{2}\left(1+\frac{1}{a}\right)\left[\theta_i^2-\frac{2x_i\theta_i}{\left(1+\frac{1}{a}\right)}\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{-\frac{1}{2}\left(\frac{a+1}{a}\right)\left[\theta_i^2-\frac{2x_i\theta_i}{\left(\frac{a+1}{a}\right)}\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\theta_i^2-2\theta_i x_i\frac{a}{a+1}\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i-x_i\frac{a}{a+1}\right)^2-\left(x_i\frac{a}{a+1}\right)^2\right]\right\}\\
&= \frac{1}{2\pi\sqrt{a}}\exp\left\{-\frac{1}{2}x_i^2\right\}\exp\left\{\frac{1}{2\frac{a}{a+1}}\left(x_i\frac{a}{a+1}\right)^2\right\}\exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i-x_i\frac{a}{a+1}\right)^2\right]\right\}\\
&= \exp\left\{\frac{1}{2\frac{a}{a+1}}\left(x_i\frac{a}{a+1}\right)^2\right\}\frac{1}{\sqrt{2\pi(a+1)}}\times\\
&\quad\times\exp\left\{-\frac{1}{2}x_i^2\right\}\frac{1}{\sqrt{2\pi\frac{a}{a+1}}}\exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i-x_i\frac{a}{a+1}\right)^2\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi\,(a+1)}} \exp\left\{\frac{1}{2\frac{a}{a+1}}\left(-\frac{1}{2}x_i^2 + x_i^2\frac{a^2}{(a+1)^2}\right)\right\} \frac{1}{\sqrt{2\pi\frac{a}{a+1}}} \times \\
&\quad \times exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i - x_i\frac{a}{a+1}\right)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi\,(a+1)}} \exp\left\{-\frac{1}{2}x_i^2 + \frac{1}{2\,(a+1)}x_i^2 a\right\} \frac{1}{\sqrt{2\pi\frac{a}{a+1}}} \exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i - x_i\frac{a}{a+1}\right)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi\,(a+1)}} \exp\left\{-\frac{1}{2}x_i^2\left(1-\frac{a}{a+1}\right)\right\} \frac{1}{\sqrt{2\pi\frac{a}{a+1}}} \exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i - x_i\frac{a}{a+1}\right)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi\,(a+1)}} \exp\left\{-\frac{1}{2}x_i^2\left(\frac{1}{a+1}\right)\right\} \frac{1}{\sqrt{2\pi\frac{a}{a+1}}} \exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i - x_i\frac{a}{a+1}\right)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi\,(a+1)}} \exp\left\{-\frac{1}{2\,(a+1)}x_i^2\right\} \frac{1}{\sqrt{2\pi\frac{a}{a+1}}} \exp\left\{-\frac{1}{2\frac{a}{a+1}}\left[\left(\theta_i - x_i\frac{a}{a+1}\right)^2\right]\right\} \\
&= \frac{1}{2\pi\sqrt{a}} \exp\left\{\frac{\left(\theta_i - \frac{a}{a+1}x_i\right)^2}{-\frac{2a}{a+1}}\right\} \exp\left\{-\frac{1}{2(a+1)}x_i^2\right\} \\
&= \frac{1}{2\pi\sqrt{a}} \exp\left\{-\frac{\left(\theta_i - \frac{a}{a+1}x_i\right)^2}{\frac{2a}{a+1}}\right\} \exp\left\{-\frac{x_i^2}{2\,(a+1)}\right\} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\frac{a}{a+1}}} \exp\left\{-\frac{1}{2}\frac{\left(\theta_i - \frac{a}{a+1}x_i\right)^2}{\frac{a}{a+1}}\right\} \frac{1}{\sqrt{2\pi}\sqrt{a+1}} \exp\left\{-\frac{1}{2}\frac{x_i^2}{a+1}\right\}
\end{aligned}
$$

$$
\Rightarrow \pi\,(\theta_i|x_i) = \frac{1}{\sqrt{2\pi}\sqrt{\frac{a}{a+1}}} \exp\left\{-\frac{1}{2}\frac{\left(\theta_i - \frac{a}{a+1}x_i\right)^2}{\frac{a}{a+1}}\right\}
$$

e

$$
g\,(x_i) = \frac{1}{\sqrt{2\pi}\sqrt{a+1}} \exp\left\{-\frac{1}{2}\frac{x_i^2}{a+1}\right\}
$$

Então, como a posteriori $\pi\,(\theta_i|x_i)$ é $Normal\left(\frac{a}{a+1}x_i, \frac{a}{a+1}\right)$, o estimador de

Bayes para θ_i dado que se conhece x_i é a esperança da posteriori.

$$\hat{\theta}_i = \frac{a}{a+1}x_i = \left(1 - \frac{1}{a+1}\right)x_i \tag{2.3}$$

Então, o estimador de Bayes $\hat{\theta}_i$ é um encolhimento do estimador usual.

2.4 Método Bayesiano empírico

A precisão das inferências estatísticas que são feitas utilizando os métodos Bayesianos, geralmente, depende fortemente de um bom conhecimento a priori. Quando a informação a priori, representada pela distribuição a priori, é incompleta ou desconhecida pode ser obtida a partir dos dados. Esse método é chamado de Método Bayesiano Empírico.

Exemplo 4

Considere X uma variável aleatória discreta com distribuição Poisson de parâmetro μ e uma distribuição a priori $\pi(\mu)$ totalmente desconhecida.

Observada uma amostra $X = x$, a verossimilhança é dada por:

$$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}$$

Aplicando o teorema de Bayes, a posteriori $\pi(\mu|x)$ é igual a

$$\pi(\mu|x) = \frac{\frac{e^{-\mu}\mu^x}{x!}\pi(\mu)}{g(x)}$$

em que a marginal de x é dada por

$$g(x) = \int \frac{e^{-\mu}\mu^x}{x!}\pi(\mu)d\mu$$

O estimador de Bayes para μ é a média da distribuição a posteriori

$$\begin{aligned}
\hat{\mu}_B &= E[\mu|x] = \int \mu \frac{e^{-\mu}\mu^x}{x!g(x)}\pi(\mu)d\mu \\
&= \frac{(x+1)!g(x+1)}{x!g(x)} \underbrace{\int \frac{e^{-\mu}\mu^{x+1}}{(x+1)!g(x+1)}\pi(\mu)d\mu}_{1} \qquad (2.4) \\
&= \frac{(x+1)!g\,(x+1)}{x!g\,(x)}
\end{aligned}$$

Como X tem distribuição discreta, se $x_1, \ldots, x_n$ são observações ocorridas no passado, as prioris $g(x)$ e $g(x+1)$ podem ser estimadas pelas suas frequências relativas.

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

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

Desta forma o estimador de Bayes empírico é dado por

$$\hat{\mu}_B = \frac{(x+1)\hat{g}(x+1)}{\hat{g}(x)}$$

Repare que $\hat{\mu}_B$ não depende dos parâmetros da distribuição a priori. Isso só ocorre devido à particularidade da distribuição dos dados ser uma distribuição discreta que segue uma Poisson. No passo (2.4), onde se completa a integral para que a mesma se transforme numa fdp (função densidade de probabilidade), a distribuição a priori desaparece com a integral.

Observe que mesmo com a distribuição a priori totalmente desconhecida conseguimos estimar a média 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 $f(x|\mu)$, e na estimativa da marginal de x denominada $g(x)$.

Aplicação:

O número de acidentes por semana 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 $g(4)$ e $g(4+1)$ são, respectivamente

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

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

Então, para este caso, o estimador Bayesiano empírico é

$$\hat{\mu}_B = \frac{(4+1)\hat{g}(4+1)}{\hat{g}(4)} = \frac{5\frac{2}{10}}{\frac{4}{10}} = \frac{5}{2}$$

Continuação do Exemplo 2

Para um melhor entendimento do método Bayesiano empírico será apresentada a continuação do exemplo 2. Suponha p conjuntos de dados independentes, cada um com n dados.

$$\begin{matrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1p} & \cdots & x_{np} \end{matrix}$$

Para o modelo de efeito aleatório a tabela ANOVA é

	SQ	E
dentro	B	$(p-1)(n\sigma_1^2+\sigma_2^2)$
entre	W	$p(n-1)\sigma_2^2$

E - Esperança

SQ - Soma de quadrados

Pelo método ANOVA o estimador da variância da priori é

$$\hat{\sigma}_1^2 = \frac{B}{n(n-1)} - \frac{W}{np(n-1)} - \frac{\sigma_2^2}{n}$$

e o estimador para média da priori é

$$\hat{\mu} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n} x_{ij}}{np}$$

Utilizando estes estimadores, o estimador Bayesiano empírico é dado por

$$E\left[\theta|\vec{x}\right] = \frac{\sigma_2^2\mu + n\sigma_1^2\,\overline{x}}{\sigma_2^2 + n\sigma_1^2} = \hat{\theta} = \frac{\frac{\hat{\mu}}{\sigma_1^2} + \frac{n\bar{x}}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{n}{\sigma_2^2}}$$

Este estimador tem uma interpretação interessante. Reescrito como

$$\hat{\theta} = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{n}{\sigma_2^2}}\hat{\mu} + \frac{\frac{n}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\frac{\sigma_2^2}{n}}}\bar{x}$$

ele é uma combinação linear da média da priori e da média amostral. Se $n \to 0$ então $\hat{\theta} \to \hat{\mu}$ e se $n \to \infty$ então $\hat{\theta} \to \bar{x}$

3 ESTIMADORES DE ENCOLHIMENTO

Os estimadores não viesados apresentam uma deficiência que geralmente não é explicitada. Seja $\hat{\beta} = \left(\hat{\beta}_1,...,\hat{\beta}_k\right)$ um estimador não viesado de um vetor de parâmetros $\beta = (\beta_1,...,\beta_k)$. Tem-se que

$$\begin{aligned} E\left[||\hat{\beta}||^2\right] &= E\left[\sum_{i=1}^{k}\hat{\beta}_i^2\right] = \sum_{i=1}^{k} E\left[\hat{\beta}_i^2\right] \\ &= \sum_{i=1}^{k}\left(Var\left[\hat{\beta}_i\right] + \left(E\left[\hat{\beta}_i\right]\right)^2\right) \\ &= \sum_{i=1}^{k} Var\left[\hat{\beta}_i\right] + ||\beta||^2 \end{aligned}$$

Observe, então, que, apesar de $\hat{\beta}$ ser um estimador **não viesado** de β, $||\hat{\beta}||^2$ é um estimador **viesado** de $||\beta||^2$, com tendência a superestimar o valor $||\beta||^2$. Esta deficiência dos estimadores não viesados foi uma das motivações para a obtenção do estimador de James-Stein e também para se adotar a estratégia de propor estimadores obtidos por encolhimento de estimadores não viesados.

Os estimadores de encolhimento, apesar de viesados, estão sendo estudados em várias áreas da estatística aplicada. Em geral, esses estimadores têm Erro Quadrático Médio (EQM) menor do que os estimadores não viesados. Vamos destacar alguns casos clássicos de encolhimento em que isso acontece.

3.1 O estimador usual da Variância

Sejam $X_1,X_2,...,X_n$ cópias independentes de uma mesma $N(\mu,\sigma^2)$. A estatística

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

é um estimador não viesado de σ^2 pois

$$\begin{aligned}
E\left[S^2\right] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n\left(\bar{X}-\mu\right)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left[(X_i-\mu)^2\right] - nE\left[\left(\bar{X}-\mu\right)^2\right]\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2 - nVar\left[\bar{X}\right]\right] \\
&= \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = \sigma^2
\end{aligned}$$

e seu erro quadrático médio é dado por:

$$E\left[\left(S^2-\sigma^2\right)^2\right] = Var\left[S^2\right]$$

Sabemos que:

a) Se $Y \sim \chi^2_n$, então $E\left[Y\right] = 2n$. (MOOD; BOES; GRAYBILL, 1974, p. 542-543)

b) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. (MOOD; BOES; GRAYBILL, 1974, p. 245)

Então,

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

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

O estimador para σ^2, tanto pelo método dos momentos, quanto pelo método da máxima verossimilhança, é dado por

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Observe que este estimador, além de ser um estimador viesado de σ_2, já que

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

também é um encolhimento do estimador não viesado S^2 pois,

$$\hat{\sigma}^2 = \frac{n-1}{n}S^2 = \left(1 - \frac{1}{n}\right)S^2$$

Vamos calcular a variância de $\hat{\sigma}^2$ para nos auxiliar no cálculo dos EQMs.

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

Mostraremos a seguir que o

$$E\left[\left(S^2 - \sigma^2\right)^2\right] > E\left[\left(\hat{\sigma}^2 - \sigma^2\right)^2\right]$$

ou seja, o EQM do estimador viesado $\hat{\sigma}^2$ é menor que o EQM do estimador não viesado S^2.

$$\begin{aligned} E\left[\left(\hat{\sigma}^2 - \sigma_2\right)^2\right] &= Var\left[\hat{\sigma}^2 - \sigma_2\right] + \left(E\left[\hat{\sigma}^2 - \sigma^2\right]\right)^2 \\ &= \frac{2\,(n-1)\,\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\ &= \left(\frac{2n-1}{n^2}\right)\sigma^4 \end{aligned} \tag{3.1}$$

Como S^2 é um estimador não viesado, então podemos escrever

$$\begin{aligned} E\left[\left(S^2 - \sigma^2\right)^2\right] &= Var[S^2] \\ &= \left(\tfrac{2}{n-1}\right)\sigma^4 \end{aligned} \tag{3.2}$$

Desta forma, com as equações (3.1) e (3.2) podemos concluir que

$$E\left[\left(\hat{\sigma}^2 - \sigma^2\right)^2\right] = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = E\left[\left(S^2 - \sigma^2\right)^2\right] \tag{3.3}$$

Para mostrar que a desigualdade (3.3) é verdadeira basta observar que

$$\begin{aligned} &\tfrac{2n-1}{n^2} < \tfrac{2}{n-1} \\ &\Leftrightarrow (2n-1)\,(n-1) < 2n^2 \\ &\Leftrightarrow 2n^2 - 2n - n + 1 < 2n^2 \\ &\Leftrightarrow 2n^2 + 1 < 2n^2 + 3n \\ &\Leftrightarrow 1 < 3n \end{aligned}$$

Esta desigualdade sempre é valida já que n representa o tamanho da amostra, ou seja, $n > 1$. Portanto, podemos concluir que $\hat{\sigma}^2$ tem menor EQM do que S^2.

3.2 Estimador com Erro Quadrático Médio Mínimo (EQMM)

Considere o modelo

$$Y_i = \mu + \varepsilon_i$$

em que $i = 1,2,3,...,n$. Y_i são observações independentes, μ representa o parâmetro a ser estimado e ε_i são os erros independentes associados ao modelo. Assumindo que $\varepsilon_i \sim N(0,\sigma^2)$, o estimador de quadrados mínimos para μ é dado por

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}.$$

Então, $\hat{\mu}$ é um estimador não viesado de μ. Queremos determinar uma constante c tal que o estimador $c\bar{Y}$ tenha erro quadrático médio mínimo. Para tanto considere a função H dada por

$$\begin{aligned}
H(c) &= E\left[\left(c\bar{Y} - \mu\right)^2\right] \\
&= E\left[c^2\bar{Y}^2 - 2c\bar{Y}\mu + \mu^2\right] \\
&= c^2E\left[\bar{Y}^2\right] - 2c\mu E\left[\bar{Y}\right] + \mu^2 \\
&= c^2\left(Var\left[\bar{Y}\right] + \left(E\left[\bar{Y}\right]\right)^2\right) - 2c\mu^2 + \mu^2 \\
&= c^2\left(\frac{\sigma^2}{n} + \mu^2\right) - 2c\mu^2 + \mu^2
\end{aligned}$$

Utilizando a técnica usual da diferenciação e igualando $H^{'}(c)$ a zero temos condições de encontrar o valor de c que minimiza ou maximiza o erro quadrático médio do estimador $c\bar{Y}$. Neste caso calcularemos a segunda derivada de H para mostrar que c é mínimo.

$$H^{'}(c) = 2c\left(\frac{\sigma^2}{n} + \mu^2\right) - 2\mu^2 = 0$$

Efetuando as contas teremos

$$2c_o\left(\frac{\sigma^2}{n} + \mu^2\right) - 2\mu^2 = 0$$

$$c_o = \frac{\mu^2}{\left(\frac{\sigma^2}{n} + \mu^2\right)} \tag{3.4}$$

A derivada segunda de H é

$$H''(c_o) = 2\left(\frac{\sigma^2}{n} + \mu^2\right) > 0$$

Logo, c_o é ponto de mínimo

Repare que o valor de c que minimiza a função H depende dos parâmetros populacionais. Substituindo-se μ e σ^2 pelos estimadores usuais, não viesados, $\bar{Y}$ e s^2 tem-se este estimador.

$$\hat{\mu}_o = \underbrace{\frac{\bar{Y}^2}{\frac{s^2}{n} + \bar{Y}^2}}_{\hat{c}_o} \bar{Y} \tag{3.5}$$

O estimador (3.5) pode ser obtido com uma abordagem Bayesiana como será visto no capítulo seguinte.

3.3 O Estimador de James - Stein como um estimador Bayesiano empírico

No exemplo 3 com $X = \{X_1, \ldots, X_n\}$, $X_i \sim N(\theta, I)$ e $\theta_i \sim N(0, a)$ o estimador de Bayes (2.3) foi expresso por

$$\hat{\theta}_i = \frac{a}{a+1} x_i = \left(1 - \frac{1}{a+1}\right) x_i.$$

O estimador $\hat{\theta}_i$ depende do parâmetro a das prioris, que é desconhecido, por isso vamos estimá-lo. Suponhamos que os dados observados são provenientes de uma população com densidade $g(x_i)$ que, como mostrado no exemplo 3, é $N(0,a+1)$. Considere Z_i a variável aleatória dada por

$$Z_i = \frac{X_i}{\sqrt{a+1}} \Rightarrow Z_i \sim N(0,1)$$

Assim, se elevarmos Z ao quadrado e somarmos teremos

$$\sum_{i=1}^{p} Z_i^2 = \frac{\sum_{i=1}^{p} X_i^2}{\left(\sqrt{a+1}\right)^2} = \frac{X'X}{a+1}$$

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

Segue a seguinte relação:

$$Y = \frac{X'X}{a+1} \Leftrightarrow \frac{1}{X'X} = \frac{1}{Y}\frac{1}{a+1} \tag{3.6}$$

Aplicando a esperança em ambos os lados da equação (3.6) segue que

$$\begin{aligned} E\left[\frac{1}{X'X}\right] &= E\left[\frac{1}{(a+1)Y}\right] \\ &= \frac{1}{a+1} E\left[\frac{1}{Y}\right] \end{aligned}$$

Como $Y \sim Gama(\frac{p}{2}, \frac{1}{2})$ segue que

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

segue que,

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

e portanto

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

Portanto, pelo métodos dos momentos podemos estimar $\frac{1}{a+1}$ por $\frac{p-2}{X'X}$. Vimos em (2.3) que o estimador de Bayes é

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

então, substituindo $\frac{1}{a+1}$ por $\frac{p-2}{X'X}$, o resultado é o estimador de James-Stein explicitado a seguir.

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

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

4 ABORDAGEM BAYESIANA EM ALGUNS ESTIMADORES DE ENCOLHIMENTO

Neste capítulo vamos discutir a abordagem Bayesiana para alguns dos estimadores de encolhimento que estudamos anteriormente.

4.1 Estimador com erro quadrático médio mínimo, sua aproximação pelo estimador Bayesiano empírico

Neste tópico vamos discutir a abordagem Bayesiana para o estimador com erro quadrático médio mínimo. Vamos assumir que o parâmetro $\mu \sim N(0,\tau^2)$. Neste caso, diferente da sessão (3.2) onde a análise foi feita com uma abordagem frequentista, o parâmetro μ tem distribuição de probabilidade. O EQM é dado pela relação

$$E[(c\bar{y}-\mu)^2] = E[c^2\bar{y}^2 - 2c\bar{y}\mu + \mu^2] \tag{4.1}$$

Para calcular o EQM vamos calcular as esperanças separadamente. Para isso precisamos das funções densidades de probabilidade de μ e $\bar{y}$.

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

$$f_{\mu} = \frac{1}{\sqrt{2\pi\tau^2}}\exp\left\{-\frac{1}{2\tau^2}\left(\mu^2\right)\right\}$$

Neste caso

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

Calculemos cada integral separadamente:

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

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

Por fim, temos que

$$
\begin{aligned}
E[\mu^2] &= Var[\mu] + (E[\mu])^2 \\
&= \tau^2
\end{aligned}
$$

Substituindo na equação (4.1):

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

Derivando a função $H(c) = E\left[(c\bar{y} - \mu)^2\right]$ em relação a c:

$$
\frac{dH(c)}{dc} = 2c\frac{\sigma^2}{n} + 2c\tau^2 - 2c\tau^2
$$

Igualando a zero, para determinarmos os pontos críticos, neste caso mínimo já que $H''(c_o) > 0$.

$$
\begin{aligned}
2c\frac{\sigma^2}{n} + 2c\tau^2 - 2c\tau^2 &= 0 \\
c_o &= \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}
\end{aligned}
$$

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

$$
\hat{\mu} = \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}\bar{y}
$$

Note a diferença em relação ao exemplo 3 em que

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

Observe que c_o depende de parâmetros populacionais, então vamos ter que estimá-los. Primeiro, vamos reescrever c_o

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

Vamos estimar $\frac{\sigma^2}{\tau^2}$ com $\frac{s^2}{\bar{Y}^2}$. Então, o estimador de μ_1 é dado por

$$\hat{\mu}_1 = \frac{n}{\frac{s^2}{\bar{Y}^2} + n}\bar{Y}$$

.

O exemplo 2 nos mostrou que se uma variável aleatória X tem distribuição $X \sim N\left(\theta,\sigma_2^2\right)$ e θ e tem uma priori $\theta \sim N\left(\mu_0,\sigma_1^2\right)$, então a posteriori, aplicando o teorema de Bayes, é

$$\pi(\theta|\vec{x}) \propto \exp\left\{-\frac{1}{2}\left[\frac{\left(\theta - \frac{\sigma_2^2\mu_0+n\sigma_1^2\,\bar{x}}{\sigma_2^2+n\sigma_1^2}\right)^2}{\frac{\sigma_1^2\,\sigma_2^2}{\sigma_2^2+n\sigma_1^2}}\right]\right\}$$

Ou seja, a posteriori é uma normal com média $\frac{\sigma_2^2\mu_0+n\sigma_1^2\,\bar{x}}{\sigma_2^2+n\sigma_1^2}$ e variância $\frac{\sigma_1^2\,\sigma_2^2}{\sigma_2^2+n\sigma_1^2}$

Em particular, temos que $X = Y \sim N\left(\mu,\sigma^2\right)$ e $\mu \sim N\left(0,\tau^2\right)$ então para $\mu_0 = 0$ e $n = 1$ podemos escrever a posteriori da seguinte forma

$$\pi\left(\mu|y\right) \propto \exp\left\{-\frac{1}{2\frac{\tau^2\sigma^2}{\sigma^2+\tau^2}}\left(\mu - \frac{\tau^2\bar{Y}}{\sigma^2 + \tau^2}\right)^2\right\} \tag{4.2}$$

Utilizando o estimador de Bayes, a média da posteriori (4.2), segue que

$$\hat{\mu}_2 = \frac{\tau^2 \bar{Y}}{\sigma^2 + \tau^2}$$

Dividindo numerador e denominador por τ^2 e se estimarmos $\frac{\sigma^2}{\tau^2}$ com $\frac{s^2}{\bar{Y}^2}$ vamos concluir que

$$\hat{\mu}_1 = \hat{\mu}_2$$

Portanto, estimador de Bayes é o mesmo que o de menor erro quadrático médio entre os estimadores da forma $c\bar{y}$ em relação à preditiva.

4.2 EQM do estimador de Bayes da Binomial

Sejam $X_1, \ldots, X_n$ iid Bernoulli (p). O EMV $\hat{p}$ de p é dado por $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$ já que a função de verossimilhança $L(p|x) = p^y(1-p)^{n-y}$ em que $y = \sum_{i=1}^{n} x_i$. Aplicando o log em ambos os lados da equação acima temos

$$\log L(p|x) = y \log p + (n-y) \log(1-p)$$

Derivando em relação ao parâmetro p e igualando a zero vamos obter o estimador de máxima verossimilhança:

$$\frac{\partial \log L(p|x)}{\partial p} = \frac{y}{p} - \frac{(n-y)}{1-p} = 0$$

O valor de p que maximiza a função de verossimilhança é

$$\frac{y}{\hat{p}} = \frac{(n-y)}{1-\hat{p}}$$

$$(1-\hat{p})y = (n-y)\,\hat{p}$$

$$\hat{p} = \frac{y}{n} = \bar{X} \tag{4.3}$$

Cálculo do EQM de $\hat{p}$

$$\begin{aligned} E\left[(\hat{p} - p)^2\right] &= Var\left[\hat{p}\right] + \left(E\left[\hat{p} - p\right]\right)^2 \\ &= Var\left[\hat{p}\right] \\ &= Var[\bar{X}] \end{aligned}$$

Vale lembrar que

$$\begin{aligned} E\left[\bar{X}\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] \\ &= \frac{1}{n} np = p \\ Var[\bar{X}] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} p\left(1 - p\right) \\ &= \frac{1}{n^2} np\left(1 - p\right) \\ Var[\bar{X}] &= \frac{p\left(1 - p\right)}{n} \end{aligned}$$

OBS: Como a soma da Bernoulli é uma Binomial sabemos que $\sum\limits_{i=1}^{n} E\left[X_i\right] = np$ e $\sum\limits_{i=1}^{n} Var\left[X_i\right] = np\left(1 - p\right)$

Agora já temos ferramentas suficientes para calcular o EQM do $\hat{p}$ que é dado por

$$\begin{aligned} E\left[(\hat{p}-p)^2\right] &= Var\left[\hat{p}\right] + \left(E\left[(\hat{p}-p)\right]\right)^2 \\ &= Var\left[\bar{X}\right] + \left(E\left[\bar{X}\right]-p\right)^2 \\ &= Var\left[\bar{X}\right] \\ &= \frac{p(p-1)}{n} \end{aligned}$$

Vamos calcular estimador de Bayes da Binomial. Como $Y = \sum_{i=1}^{n} X_i$ é Binomial (n,p) é conveniente assumir que a distribuição a *priori* de p seja uma Beta (α,β). Assim, aplicando o teorema de Bayes

$$f(p|y) \propto \underbrace{p^y(1-p)^{n-y}}_{f(y|p)}\underbrace{p^{\alpha-1}(1-p)^{\beta-1}}_{\pi(p)} \propto p^{y+\alpha-1}(1-p)^{n-y+\beta-1}$$

Nestas condições a distribuição a posteriori $f(p|y)$ é uma $Beta(y+\alpha, n-y+\beta)$. Assim $f(p|y)$ é descrita da seguinte forma

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

O estimador de Bayes da Binomial é dado pela média da posteriori. Se $X \sim Beta(\alpha,\beta)$, $E[X] = \frac{\alpha}{\alpha+\beta}$. Então, neste caso

$$\hat{p}_B = \frac{y+\alpha}{\alpha+\beta+n}$$

Agora podemos calcular o EQM de $\hat{p}_B$

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

Vamos escolher α e β de tal forma que torne o EQM de $\hat{p}_B$ constante para futuras comparações. Adotaremos $\alpha = \beta = \sqrt{\frac{n}{4}}$. Substituindo os valores $\alpha = \beta = \sqrt{\frac{n}{4}}$ em $\hat{p}_B$ e $\hat{p}_B$ em $E\left[(\hat{p}_B - p)^2\right]$ temos que

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

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

e o erro quadrático médio de $\hat{p}_B$ pode ser expresso por

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

Se quisermos escolher entre $\hat{p}_B$ e $\hat{p}$ baseado no EQM a figura Figura 4 é útil. Para pequenos valores de n, $\hat{p}_B$ é a melhor escolha, a menos que exista uma forte crença de que p está próximo de 0 ou 1. Já para valores maiores de n, $\hat{p}$ é a melhor escolha, a menos que exista uma grande crença de p estar próximo de $1/2$. Podemos destacar também que mesmo que o EQM não mostre que um estimador é sempre melhor do que o outro, essas informações fornecidas são úteis. Informações estas que combinadas com o conhecimento do problema, por parte do pesquisador, podem nos auxiliar na escolha de um melhor estimador para determinada situação (CASELLA; BERGER, 2010).

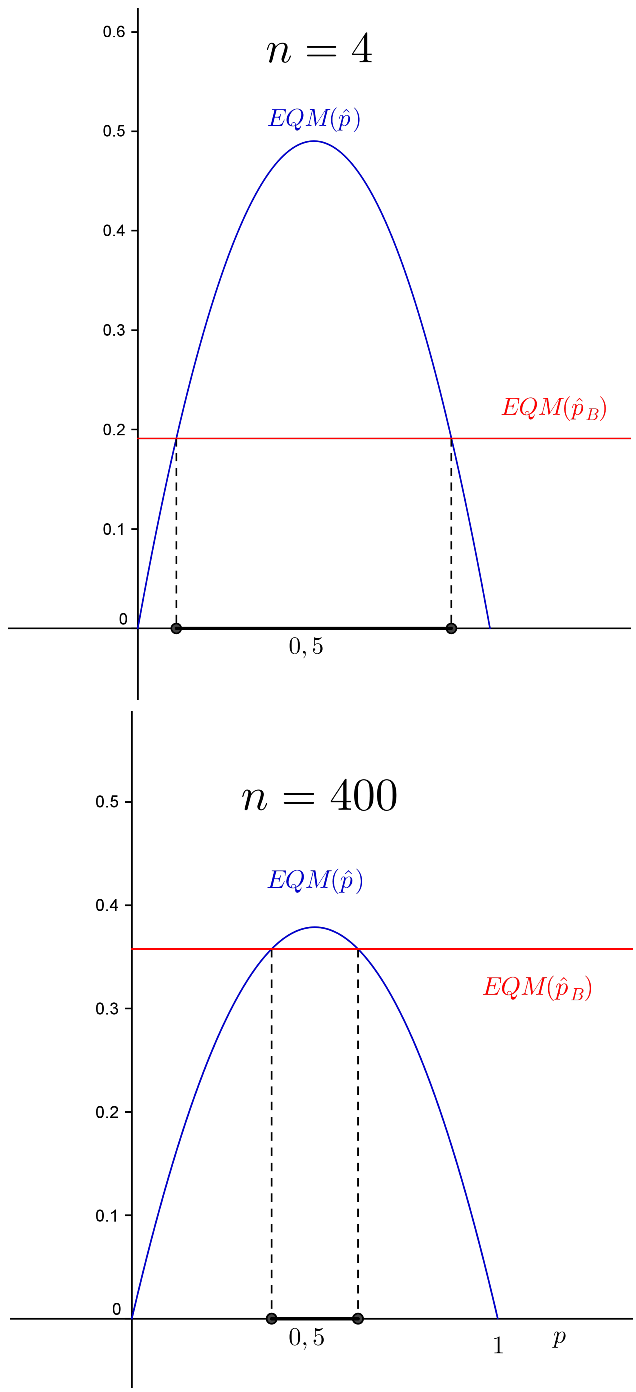

Figura 1 Comparação dos EQMs de $\hat{p}$ (azul) e $\hat{p}_B$ (vermelho).

5 A SIMULAÇÃO

Na seção 3.2 foi discutido o estimador com erro quadrático médio mínimo. Esse "estimador", na sua forma não operacional, é expresso por:

$$\hat{\mu} = c_o \bar{Y} = \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2} \bar{Y} \tag{5.1}$$

Como a expressão (5.1) depende dos parâmetros populacionais σ^2 e μ, o caminho para transformá-lo em um legítimo estimador é substituir esses parâmetros por seus estimadores usuais.

$$\hat{\mu}_o = \underbrace{\frac{\bar{Y}^2}{\frac{s^2}{n} + \bar{Y}^2}}_{\hat{c}_o} \bar{Y} \tag{5.2}$$

Por simulação estudamos o erro quadrático médio do estimador (5.2). Para isso foram geradas 10.000 amostras de tamanhos 20, 40 e 100, provenientes das distribuições uniforme, normal e exponencial dupla, com variâncias 4, 16 e 36 conhecidas. Os Erros Quadráticos Médios observados nas simulações estão mostrados nas figuras 2 , 3 e 4 a seguir.

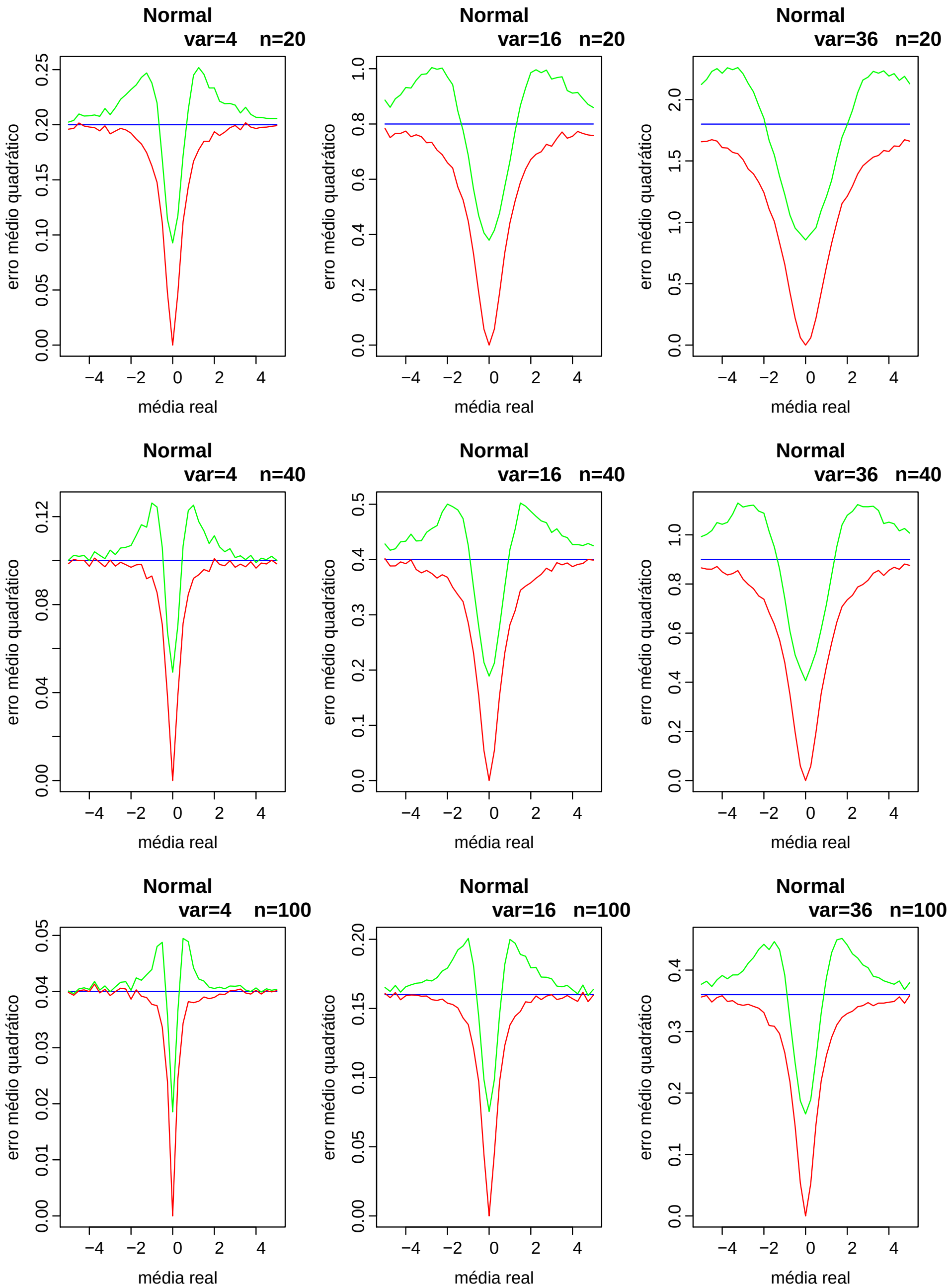

Figura 2 Erros quadráticos médios dos estimadores $\hat{\mu}$ (vermelho), $\hat{\mu}_o$ (verde) e $\bar{Y}$ (azul), para amostras de tamanhos 20, 40 e 100 e variâncias 4, 16 e 36 para uma população com distribuição normal.

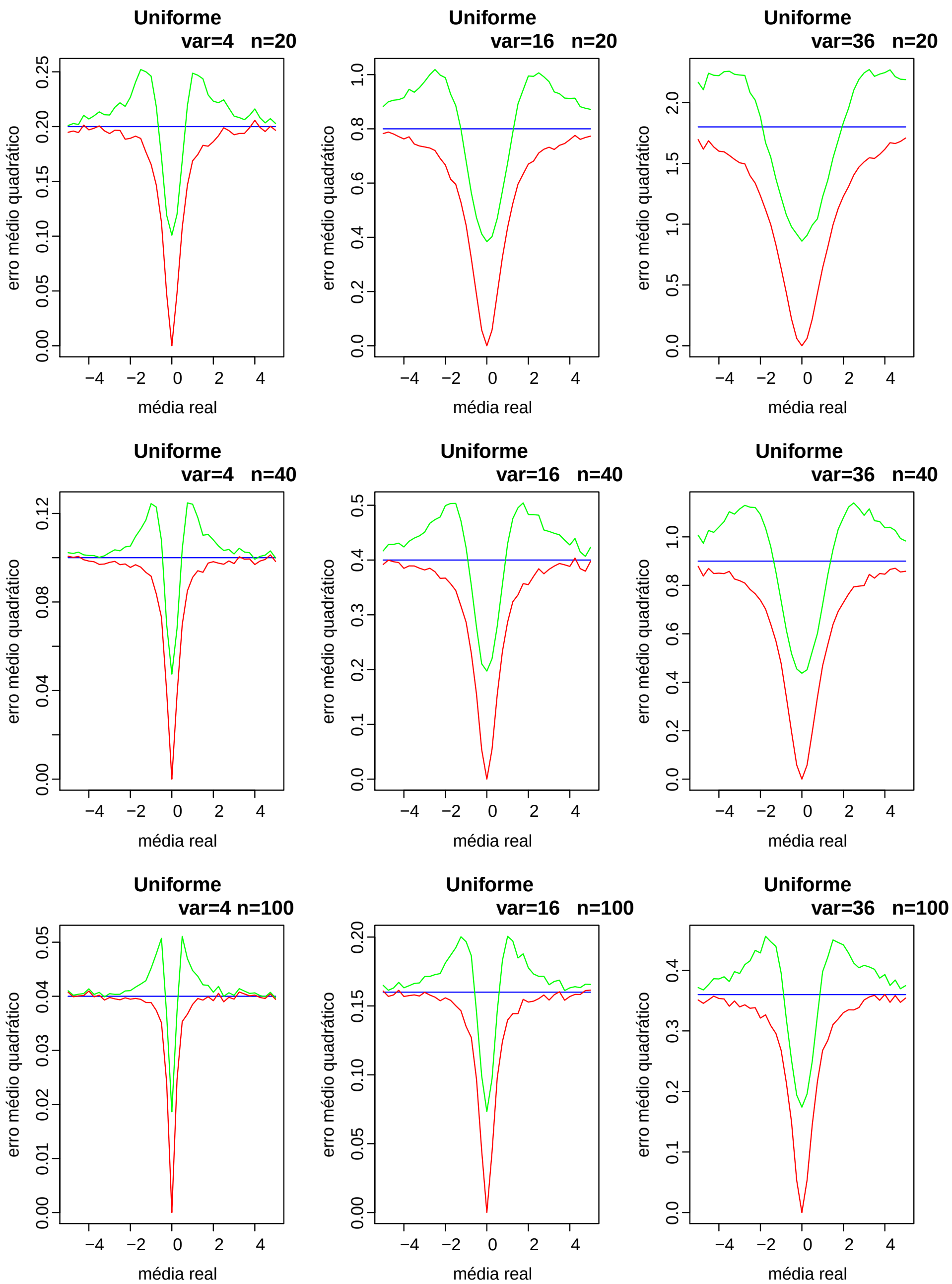

Figura 3 Erros quadráticos médios dos estimadores $\hat{\mu}$ (vermelho), $\hat{\mu}_o$ (verde) e $\bar{Y}$ (azul), para amostras de tamanhos 20, 40 e 100 e variâncias 4, 16 e 36 para uma população com distribuição uniforme.

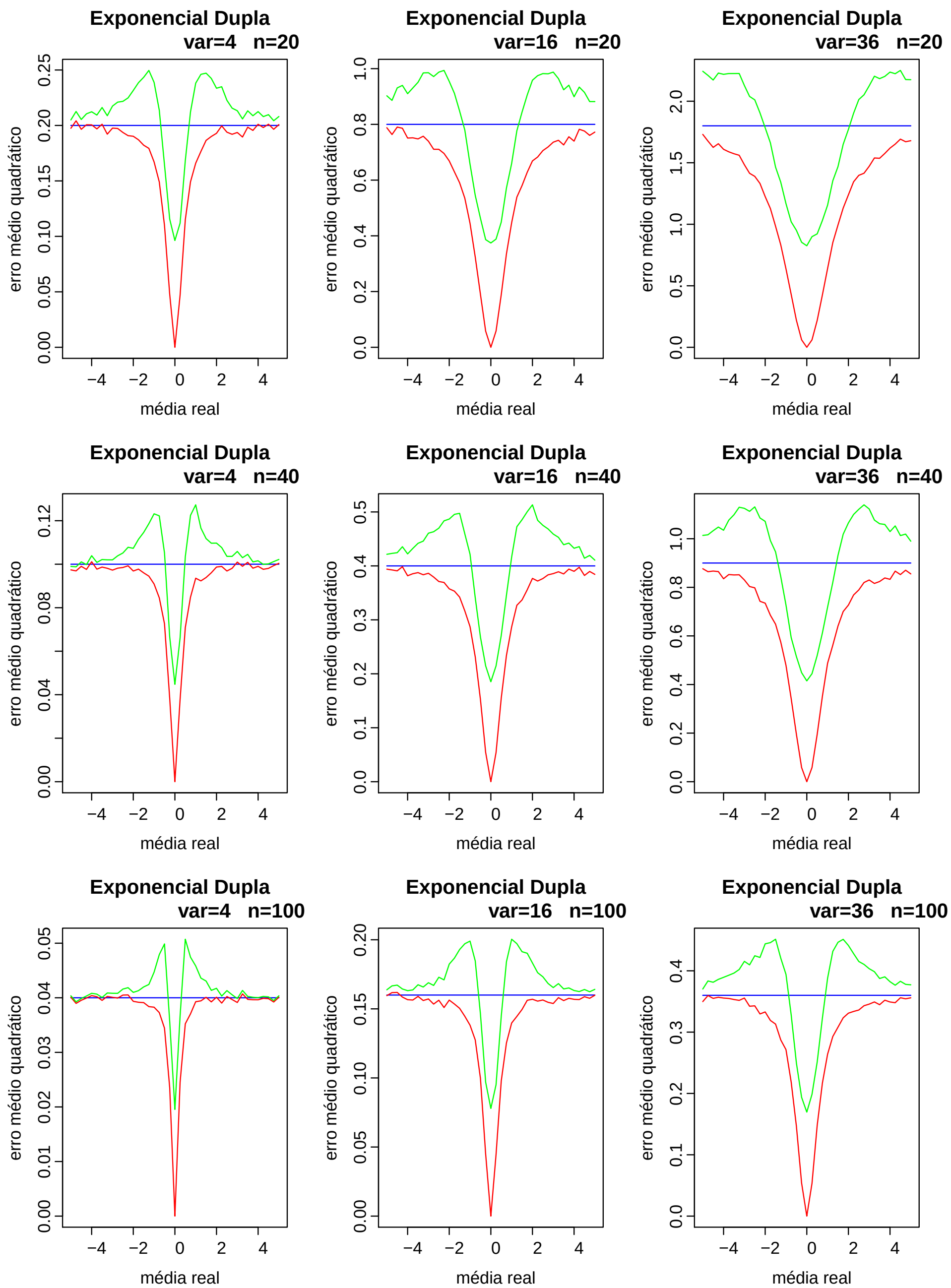

Figura 4 Erros quadráticos médios dos estimadores $\hat{\mu}$ (vermelho), $\hat{\mu}_o$ (verde) e $\bar{Y}$ (azul), para amostras de tamanhos 20, 40 e 100 e variâncias 4, 16 e 36 para uma população com distribuição exponencial dupla.

6 CONSIDERAÇÕES FINAIS

As figuras 2, 3 e 4 mostram que os comportamentos dos erros quadráticos médios dos estimadores $\hat{\mu}$, $\hat{\mu}_o$ e $\bar{Y}$, praticamente, não dependem da distribuição da população original. Os gráficos mostraram que o erro quadrático médio do "estimador"não operacional $\hat{\mu}$ se comporta conforme o esperado, ou seja, sistematicamente menor do que o erro quadrático médio do estimador usual $\bar{Y}$. O que não se esperava é o resultado obtido para o erro quadrático médio do estimador $\hat{\mu}_o$ que só é menor que o erro quadrático médio de $\bar{Y}$ para valores da esperança populacional em um intervalo centrado na média zero. Quando os parâmetros populacionais presentes na expressão $\hat{\mu}$ foram substituídos pelos seus estimadores usuais esperava-se que o estimador $\hat{\mu}_o$ assim obtido mantivesse seu comportamento, ou seja, erro quadrático médio inferior ao erro quadrático médio de $\bar{Y}$, tal fato não foi observado no estudo das simulações. As figuras 2, 3 e 4 mostram que esse comportamento só acontece para valores de média populacional em um intervalo centrado no zero. As figuras mostram ainda que esse intervalo é crescente com a variância populacional e decrescente com o tamanho da amostra, porém fora desse intervalo o erro quadrático médio de $\hat{\mu}_o$ é, sistematicamente, maior que o erro quadrático médio do estimador usual. Este fato, inesperado, abre portas para futuras investigações.

REFERÊNCIAS

CASELLA, G.; BERGER, R. L. **Inferência estatística**. 2. ed. São Paulo: Cengage Learning, 2010. 573 p.

GRUBER, M. H. J. **Improving efficiency by shrinkage**: the James-Stein and ridge regression estimators. New York: CRC, 1998. v. 1, 648 p.

MOOD, A. M.; BOES, F. A.; GRAYBILL, F. A. **Introduction to the theory of statistics**. 3rded. Columbus: McGraw-Hill, 1974. 564 p.

APÊNDICE

APÊNDICE A - Rotina para geração do gráfico dos erros quadráticos médios para uma amostra com população com distribuição normal, uniforme e Exponencial dupla.

1. Rotina para geração do gráfico dos erros quadráticos médios para uma amostra com população com distribuição normal.

```
simulação Normal
 VAR = c(4,16,36)
 nvar = length(VAR) # As variâncias
 MU = seq(-5,5,0.25) # As médias
 nmu = length(MU)
 N = 1000 # O número de amostras para cada média
O caso da distribuição NORMAL(mu,var)
MN = matrix(0,nvar*nmu,4)
colnames(MN) = c("Variância real","Média real","Erro do
              encolhimento", "Erro do encolhimento estimado")
k =
for (i in 1:nvar){
for (j in 1:nmu){
k = k + 1
MN[k,1] = VAR[i]
MN[k,2] = MU[j]}}
n = 20 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
```

```
B = matrix(0,N,4)
vari = MN[i,1]
mu = MN[i,2]
for (j in 1:N){
amostra = rnorm(n,mu,sqrt(vari))
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MN[i,2]^2/(vari/n+MN[i,2]^2))*B[j,1]
erro1 = erro1 + (MN[i,2]-B[j,3])^2
erro2 = erro2 + (MN[i,2]-B[j,4])^2   }
MN[i,4] = erro1/N
MN[i,3] = erro2/N
MN[1:5,] }
par(mfrow=c(3,3))
plot(c(MU[1],MU[nmu]),c(0,max(MN[1:nmu,4])),type="n",
                    main="Normal    var=4    n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MN[1:nmu,3],col="red")
lines(MU,MN[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(1+nmu):(2*nmu),4])),type="n",
main="Normal
var=16   n=20
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
```

```
lines(MU,MN[(1+nmu):(2*nmu),3],col="red")
lines(MU,MN[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(2+nmu):(3*nmu),4])),type="n",
main="Normal
var=36   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MN[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MN[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n = 20

n = 40 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MN[i,1]
mu = MN[i,2]
for (j in 1:N){
amostra = rnorm(n,mu,sqrt(vari))
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MN[i,2]^2/(vari/n+MN[i,2]^2))*B[j,1]
erro1 = erro1 + (MN[i,2]-B[j,3])^2
erro2 = erro2 + (MN[i,2]-B[j,4])^2     }
```

```
MN[i,4] = erro1/N
MN[i,3] = erro2/N
MN[1:5,] }
plot(c(MU[1],MU[nmu]),c(0,max(MN[1:nmu,4])),type="n",
                      main="Normal     var=4    n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MN[1:nmu,3],col="red")
lines(MU,MN[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(1+nmu):(2*nmu),4])),type="n",
main="Normal
var=16   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MN[(1+nmu):(2*nmu),3],col="red")
lines(MU,MN[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(2+nmu):(3*nmu),4])),type="n",
main="Normal
var=36   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MN[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MN[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=40

n = 100 # O tamanho da cada amostra
```

```
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MN[i,1]
mu = MN[i,2]
for (j in 1:N){
amostra = rnorm(n,mu,sqrt(vari))
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MN[i,2]^2/(vari/n+MN[i,2]^2))*B[j,1]
erro1 = erro1 + (MN[i,2]-B[j,3])^2
erro2 = erro2 + (MN[i,2]-B[j,4])^2    }
MN[i,4] = erro1/N
MN[i,3] = erro2/N
MN[1:5,] }
plot(c(MU[1],MU[nmu]),c(0,max(MN[1:nmu,4])),type="n",
                       main="Normal    var=4    n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MN[1:nmu,3],col="red")
lines(MU,MN[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(1+nmu):(2*nmu),4])),type="n",
main="Normal
var=16   n=100"
```

```
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MN[(1+nmu):(2*nmu),3],col="red")
lines(MU,MN[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MN[(2+nmu):(3*nmu),4])),type="n",
main="Normal
var=36   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MN[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MN[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=100
```

2. Rotina para geração do gráfico dos erros quadráticos médios para uma amostra com população com distribuição uniforme.

```
#  simulação   Uniforme
VAR = c(4,16,36)
nvar = length(VAR) # As variâncias
MU = seq(-5,5,0.25) # As médias
nmu = length(MU)
N = 1000 # O número de amostras para cada média
# O caso da distribuição UNIFORME com média mu variância vari
MUn = matrix(0,nvar*nmu,4)
colnames(MUn) = c("Variância real","Média real","Erro do
                encolhimento", "Erro do encolhimento estimado")
k = 0
for (i in 1:nvar){
```

```
for (j in 1:nmu){
k = k + 1
MUn[k,1] = VAR[i]
MUn[k,2] = MU[j] }}

n = 20 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MUn[i,1]
mu = MUn[i,2]
A = sqrt(12*vari) # A amplitude
for (j in 1:N){
amostra = runif(n,mu-A/2,mu+A/2)
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MUn[i,2]^2/(vari/n+MUn[i,2]^2))*B[j,1]
erro1 = erro1 + (MUn[i,2]-B[j,3])^2
erro2 = erro2 + (MUn[i,2]-B[j,4])^2    }
MUn[i,4] = erro1/N
MUn[i,3] = erro2/N
MUn[1:5,] }
par(mfrow=c(3,3))
plot(c(MU[1],MU[nmu]),c(0,max(MUn[1:nmu,4])),type="n",
```

```
                 main="Uniforme    var=4   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MUn[1:nmu,3],col="red")
lines(MU,MUn[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(1+nmu):(2*nmu),4])),type="n",
main="Uniforme
var=16   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+nmu):(2*nmu),3],col="red")
lines(MU,MUn[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(2+nmu):(3*nmu),4])),type="n",
main="Uniforme var=36   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MUn[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=20

n = 40 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MUn[i,1]
```

```
mu = MUn[i,2]
A = sqrt(12*vari) # A amplitude
for (j in 1:N){
amostra = runif(n,mu-A/2,mu+A/2)
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MUn[i,2]^2/(vari/n+MUn[i,2]^2))*B[j,1]
erro1 = erro1 + (MUn[i,2]-B[j,3])^2
erro2 = erro2 + (MUn[i,2]-B[j,4])^2  }
MUn[i,4] = erro1/N
MUn[i,3] = erro2/N
MUn[1:5,] }
plot(c(MU[1],MU[nmu]),c(0,max(MUn[1:nmu,4])),type="n",
                   main="Uniforme    var=4   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MUn[1:nmu,3],col="red")
lines(MU,MUn[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(1+nmu):(2*nmu),4])),type="n",
main="Uniforme
var=16   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+nmu):(2*nmu),3],col="red")
lines(MU,MUn[(1+nmu):(2*nmu),4],col="green")
```

```
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(2+nmu):(3*nmu),4])),type="n",
main="Uniforme  var=36   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MUn[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=40

n = 100 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MUn[i,1]
mu = MUn[i,2]
A = sqrt(12*vari) # A amplitude
for (j in 1:N){
amostra = runif(n,mu-A/2,mu+A/2)
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MUn[i,2]^2/(vari/n+MUn[i,2]^2))*B[j,1]
erro1 = erro1 + (MUn[i,2]-B[j,3])^2
erro2 = erro2 + (MUn[i,2]-B[j,4])^2 }
MUn[i,4] = erro1/N
MUn[i,3] = erro2/N
```

```
MUn[1:5,] }
plot(c(MU[1],MU[nmu]),c(0,max(MUn[1:nmu,4])),type="n",
                  main="Uniforme    var=4 n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MUn[1:nmu,3],col="red")
lines(MU,MUn[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(1+nmu):(2*nmu),4])),type="n",
main="Uniforme
  var=16   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+nmu):(2*nmu),3],col="red")
lines(MU,MUn[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MUn[(2+nmu):(3*nmu),4])),type="n",
main="Uniforme
  var=36   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MUn[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MUn[(1+2*nmu):(3*nmu),4],col="green")
```

3. Rotina para geração do gráfico dos erros quadráticos médios para uma amostra com população com distribuição exponencial dupla.

```
#  simulação Exponencial Dupla
VAR = c(4,16,36)
nvar = length(VAR) # As variâncias
```

```
MU = seq(-5,5,0.25) # As médias
nmu = length(MU)
N = 1000 # O número de amostras para cada média
# O caso da distribuição NORMAL(mu,var)
MEd = matrix(0,nvar∗nmu,4)
colnames(MEd) = c("Variância real","Média real","Erro do
                encolhimento", "Erro do encolhimento estimado")
k = 0
for (i in 1:nvar){
for (j in 1:nmu){
k = k + 1
MEd[k,1] = VAR[i]
MEd[k,2] = MU[j] } }

n = 20 # O tamanho da cada amostra
for (i in 1:(nmu∗nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MEd[i,1]
beta = sqrt(vari/2)
mu = MEd[i,2]
for (j in 1:N){
mult = 2∗floor(2∗runif(n)) - 1
amostra = rexp(n,1/beta)∗mult + mu
B[j,1] = mean(amostra)
```

```
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MEd[i,2]^2/(vari/n+MEd[i,2]^2))*B[j,1]
erro1 = erro1 + (MEd[i,2]-B[j,3])^2
erro2 = erro2 + (MEd[i,2]-B[j,4])^2  }
MEd[i,4] = erro1/N
MEd[i,3] = erro2/N }
par(mfrow=c(3,3))
plot(c(MU[1],MU[nmu]),c(0,max(MEd[1:nmu,4])),type="n",
main="Exponencial Dupla
var=4   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MEd[1:nmu,3],col="red")
lines(MU,MEd[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(1+nmu):(2*nmu),4])),type="n",
main="Exponencial Dupla
var=16   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MEd[(1+nmu):(2*nmu),3],col="red")
lines(MU,MEd[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(2+nmu):(3*nmu),4])),type="n",
main="Exponencial Dupla
var=36   n=20"
   ,xlab = "média real", ylab = "erro médio quadrático")
```

```
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MEd[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MEd[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=20

n = 40 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
vari = MEd[i,1]
beta = sqrt(vari/2)
mu = MEd[i,2]
for (j in 1:N){
mult = 2*floor(2*runif(n)) - 1
amostra = rexp(n,1/beta)*mult + mu
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MEd[i,2]^2/(vari/n+MEd[i,2]^2))*B[j,1]
erro1 = erro1 + (MEd[i,2]-B[j,3])^2
erro2 = erro2 + (MEd[i,2]-B[j,4])^2  }
MEd[i,4] = erro1/N
MEd[i,3] = erro2/N   }
plot(c(MU[1],MU[nmu]),c(0,max(MEd[1:nmu,4])),type="n",
main="Exponencial Dupla
```

```
var=4   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MEd[1:nmu,3],col="red")
lines(MU,MEd[1:nmu,4],col="green"
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(1+nmu):(2*nmu),4])),type="n",
main="Exponencial Dupla
var=16   n=40"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
lines(MU,MEd[(1+nmu):(2*nmu),3],col="red")
lines(MU,MEd[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(2+nmu):(3*nmu),4])),type="n",
main="Exponencial Dupla
var=36   n=40"
 ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MEd[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MEd[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=40

n = 100 # O tamanho da cada amostra
for (i in 1:(nmu*nvar)){
erro1 = 0
erro2 = 0
B = matrix(0,N,4)
```

```
vari = MEd[i,1]
beta = sqrt(vari/2)
mu = MEd[i,2]
for (j in 1:N){
mult = 2*floor(2*runif(n)) - 1
amostra = rexp(n,1/beta)*mult + mu
B[j,1] = mean(amostra)
B[j,2] = var(amostra)
B[j,3] = B[j,1]^3/(B[j,2]/n+B[j,1]^2)
B[j,4] = (MEd[i,2]^2/(vari/n+MEd[i,2]^2))*B[j,1]
erro1 = erro1 + (MEd[i,2]-B[j,3])^2
erro2 = erro2 + (MEd[i,2]-B[j,4])^2  }
MEd[i,4] = erro1/N
MEd[i,3] = erro2/N }
plot(c(MU[1],MU[nmu]),c(0,max(MEd[1:nmu,4])),type="n",
main="Exponencial Dupla
var=4   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[1]/n,nmu),type="l",col="blue")
lines(MU,MEd[1:nmu,3],col="red")
lines(MU,MEd[1:nmu,4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(1+nmu):(2*nmu),4])),type="n",
main="Exponencial Dupla
var=16   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[2]/n,nmu),type="l",col="blue")
```

```
lines(MU,MEd[(1+nmu):(2*nmu),3],col="red")
lines(MU,MEd[(1+nmu):(2*nmu),4],col="green")
plot(c(MU[1],MU[nmu]),c(0,max(MEd[(2+nmu):(3*nmu),4])),type="n",
main="Exponencial Dupla
var=36   n=100"
   ,xlab = "média real", ylab = "erro médio quadrático")
lines(MU,rep(VAR[3]/n,nmu),type="l",col="blue")
lines(MU,MEd[(1+2*nmu):(3*nmu),3],col="red")
lines(MU,MEd[(1+2*nmu):(3*nmu),4],col="green")
# Fim para n=100
```