

# ESTIMAÇÃO ROBUSTA DOS PARÂMETROS DA DISTRIBUIÇÃO BETA

Carlos Eduardo Frantz MANCHINI<sup>1</sup>  
Luan PORTELLA<sup>2</sup>  
Fábio Mariano BAYER<sup>3</sup>

- RESUMO: O presente trabalho propõe estimadores robustos para os parâmetros da distribuição beta, úteis para modelar dados contínuos no intervalo  $(0,1)$  na presença de valores discrepantes (*outliers*). Foram considerados estimadores de máxima verossimilhança ponderados (EMVP) com duas propostas de ponderação e um estimador de mínima distância (EMD). Para correção de viés desses estimadores foi utilizado o método *bootstrap* paramétrico adaptado para dados com *outliers*. Os desempenhos dos estimadores propostos foram avaliados por meio de simulações de Monte Carlo, comparativamente aos estimadores usuais pelo método dos momentos ou de máxima verossimilhança. Como medidas específicas de robustez foram adotados o ponto de ruptura e a sensibilidade. A avaliação numérica evidenciou que os EMVP e os EMD possuem menor sensibilidade a *outliers* e maiores pontos de ruptura. Os estimadores corrigidos pelo método *bootstrap* proposto apresentaram menor viés comparativamente às suas versões não corrigidas. Ao final, uma aplicação a dados reais é apresentada e discutida.
- PALAVRAS-CHAVE: *Bootstrap*; estimadores robustos; mínima distância; *outliers*; verossimilhança ponderada.

## 1 Introdução

A distribuição beta (GUPTA e NADARAJAH, 2004) tem sido largamente empregada para modelar dados no intervalo  $(0,1)$ , como taxas e proporções. Esta

---

<sup>1</sup>Universidade Federal de Santa Maria - UFSM, Centro de Ciências Naturais e Exatas, Curso de Estatística, CEP: 97105-900, Santa Maria, RS, Brasil. E-mail: *carlos-efm@hotmail.com*

<sup>2</sup>Universidade Federal de Santa Maria - UFSM, Centro de Tecnologia, Pós-graduação em Engenharia de Produção, CEP: 97105-900, Santa Maria, RS, Brasil. E-mail: *luanportella@gmail.com*

<sup>3</sup>Universidade Federal de Santa Maria - UFSM, Departamento de Estatística e LACESM, CEP: 97105-900, Santa Maria, RS, Brasil. E-mail: *bayer@ufsm.br*

distribuição possui diversas aplicações científicas e profissionais, como em modelos de regressão (FERRARI e CRIBARI-NETO, 2004), em modelos de séries temporais (BAYER et al., 2018b) e no monitoramento de processos industriais (SANTANNA e TEN CATEN, 2012; BAYER et al., 2018a). A estimação dos parâmetros da distribuição beta é usualmente realizada via método dos momentos (EMM) ou de máxima verossimilhança (EMV). Contudo, nem sempre esses estimadores possuem bom desempenho.

Na modelagem de dados, ao assumir um determinado modelo probabilístico, como a distribuição beta, podem existir observações que apontam um padrão diferente, ou não apontam nenhum padrão considerável. Qualquer valor que apresenta-se distante da massa de dados é chamado *outlier* (MARONNA et al., 2006). Na prática, há sempre a chance de ocorrência de *outliers* no conjunto de dados (ROUSSEEUW e LEROY, 2003). Sabe-se que a presença de observações discrepantes nos dados pode fornecer estimativas distorcidas sobre os parâmetros populacionais (CHEN e LIU, 1993). Habitualmente, os *outliers* são descartados e a análise é refeita. Porém, corre-se o risco de ignorar uma observação importante. Ademais, a identificação de *outliers* é um problema de difícil solução. Neste sentido, tem se dado grande atenção a procedimentos inferenciais robustos (HUBER, 2011). Em HU e ZIDEK (1964) é utilizado um método de seleção de pesos na verossimilhança usual fundamentado no critério de máxima entropia de Akaike. AHMED et al. (2005) introduzem estimadores robustos específicos do tipo  $\alpha$  trimédia para estimar o parâmetro de uma distribuição exponencial. Para modelagem de séries temporais, BLASQUES et al. (2016) propõem estimadores robustos baseados em verossimilhança ponderada.

De forma geral, procedimentos estatísticos robustos podem ser considerados de duas formas, a saber: (i) substituindo procedimentos inferenciais usuais por métodos que considerem distribuições com caudas mais pesadas (LANGE et al., 1989; HAHN, 2008); (ii) desenvolvendo estimadores que consideram funções de influência como forma de serem menos afetados por *outliers* (HUBER, 1964; HAMPEL et al., 2011). Funções de influência são as mais eficazes ferramentas heurísticas de robustez para limitar a influência de *outliers* (HUBER, 2011). Neste sentido, o presente trabalho considerará a abordagem (ii) para inferências robustas sobre os parâmetros da distribuição beta.

O presente trabalho propõe estimadores de máxima verossimilhança ponderados (EMVP) (FIELD e SMITH, 1994) e um estimador de mínima distância (EMD) (PARR e SCHUCANY, 1980) para os parâmetros da distribuição beta. Para a função de ponderação do EMVP são consideradas duas propostas, em que uma delas requer um algoritmo iterativo. Para o método de distância mínima é adotada a medida de Kolmogorov-Smirnov (FRANK e MASSEY, 1951). Contudo, estimadores robustos, como os baseados em verossimilhança ponderada, podem apresentar um possível aumento do viés (WANG et al., 2004). Com o intuito de reduzir o viés desses estimadores, principalmente quando a amostra não apresenta *outliers*, consideramos correções *bootstrap* (EFRON e TIBSHIRANI, 1994) dos estimadores usuais e robustos.

A reamostragem *bootstrap* baseia-se em um conjunto de métodos que tem por objetivo criar múltiplas pseudoamostras a partir das observações (*bootstrap* não paramétrico) ou de uma distribuição estimada a partir da amostra (*bootstrap* paramétrico). Contudo, a abordagem não paramétrica apresenta problemas de robustez com a propagação de *outliers* (AMADO et al., 2014). Já a abordagem paramétrica apresenta problemas na estimação dos parâmetros, a qual é afetada pelos *outliers*. Nestes casos, podem ser usados estimadores robustos na reamostragem *bootstrap* paramétrica (DE LA ROSA e FLEURY, 2006; SEKI e YOKOYAMA, 1996). Neste trabalho, consideramos os estimadores robustos EMVP na geração das reamostras do *bootstrap* paramétrico. Os desempenhos dos estimadores considerados são avaliados por meio de simulações de Monte Carlo. Na avaliação numérica, são consideradas medidas específicas de robustez como ponto de ruptura e curva de sensibilidade.

O artigo está organizado da seguinte forma. Na Seção 2 são apresentados a distribuição beta e os estimadores pontuais usuais. Os estimadores robustos são introduzidos na Seção 3. O procedimento para correção de viés dos estimadores é apresentado na Seção 4. A Seção 5 apresenta a avaliação numérica dos estimadores considerados. A Seção 6 apresenta uma aplicação em dados reais, seguida da conclusão.

## 2 Distribuição beta e estimadores usuais

A distribuição beta é um modelo probabilístico adequado para descrever dados no intervalo  $(0, 1)$  distribuídos de forma contínua (JOHNSON et al., 1995). Seja  $X$  uma variável aleatória com distribuição beta de parâmetros  $\alpha$  e  $\beta$ , isto é  $X \sim \text{Beta}(\alpha, \beta)$ . Sua função densidade de probabilidade é dada por (GUPTA, 2011):

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1,$$

em que  $\alpha > 0$ ,  $\beta > 0$  e  $\Gamma(\cdot)$  é a função gama. A média e a variância de  $X$  são, respectivamente,

$$E(X) = \frac{\alpha}{(\alpha + \beta)},$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

A função distribuição acumulada de  $X$  é (GUPTA e NADARAJAH, 2004):

$$F(x; \alpha, \beta) = \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)}, \quad (1)$$

em que  $B(\cdot, \cdot)$  é a função beta e  $B(x; \cdot, \cdot)$  é a função beta incompleta. Para mais detalhes, ver GUPTA e NADARAJAH (2004).

Dependendo dos valores paramétricos, a densidade beta assume diferentes formas, sendo de grande flexibilidade para modelar proporções. A Figura 1 apresenta a função densidade de probabilidade beta com diferentes valores de  $\alpha$  e  $\beta$ . Nota-se que ela pode assumir formas simétricas, assimétricas, em forma de J, de J invertido, entre outras.

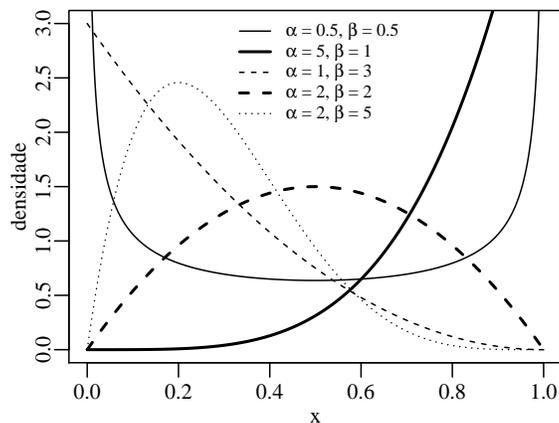


Figura 1 - Função densidade de probabilidade beta para diferentes valores paramétricos.

A ampla aplicabilidade da distribuição beta se deve ao fato de que grande parte dos dados podem ser expressos como taxas, proporções ou índices. Para modelar os dados por meio da distribuição beta é necessário realizar a estimativa dos parâmetros  $\alpha$  e  $\beta$ . Contudo, a qualidade das estimativas depende do método de estimação utilizado. Os métodos usualmente empregados para estimação de parâmetros não se comportam bem na presença de valores aberrantes (MARONNA et al., 2006). Nestas ocasiões, podem ser considerados métodos de estimação robustos resistentes à ocorrências discrepantes.

### 2.1 Estimador pelo método dos momentos (EMM)

Dada uma amostra aleatória  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ , os EMM para os parâmetros  $\alpha$  e  $\beta$  da distribuição beta são dados, respectivamente, por (OWEN, 2008):

$$\bar{\alpha} = \bar{X} \left( \frac{\bar{X}(1 - \bar{X})}{S^2} - 1 \right),$$

$$\bar{\beta} = (1 - \bar{X}) \left( \frac{\bar{X}(1 - \bar{X})}{S^2} - 1 \right),$$

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

## 2.2 Estimador de máxima verossimilhança (EMV)

O método da máxima verossimilhança é o método mais popular para estimação de parâmetros, sendo amplamente explorado na literatura (PAWITAN, 2001). Seja  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  uma amostra observada. A função de verossimilhança para os parâmetros  $\alpha$  e  $\beta$  é definida por:

$$L(\alpha, \beta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \alpha, \beta) = \prod_{i=1}^n \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1} (1 - x_i)^{\beta-1}.$$

Tomando o logaritmo da função de verossimilhança, temos a função de log-verossimilhança, dada por:

$$\ell(\alpha, \beta; \mathbf{x}) = \sum_{i=1}^n \log[f(x_i; \alpha, \beta)] = \sum_{i=1}^n \ell_i(x_i; \alpha, \beta),$$

em que

$$\begin{aligned} \ell_i(x_i; \alpha, \beta) &= \log[\Gamma(\alpha + \beta)] - \log[\Gamma(\alpha)] - \log[\Gamma(\beta)] \\ &\quad + \alpha \log(x_i) + \beta \log(1 - x_i) - \log(x_i) - \log(1 - x_i). \end{aligned}$$

Os estimadores de máxima verossimilhança são os valores  $\hat{\alpha}$  e  $\hat{\beta}$  que maximizam a função de verossimilhança, ou, equivalentemente, a função de log-verossimilhança. Para obtenção dos estimadores, deriva-se a log-verossimilhança em relação a cada um dos parâmetros e iguala-se a zero. Contudo, a solução desse sistema não possui forma fechada. Devido a isso, são considerados métodos numéricos para maximização da função de log-verossimilhança. Maiores detalhes sobre EMV dos parâmetros da distribuição beta podem ser vistos em BECKMAN e JEN (1978).

## 3 Estimadores robustos

Esta seção apresenta os estimadores robustos propostos para os parâmetros  $\alpha$  e  $\beta$  da distribuição beta. Para os EMVP são consideradas duas definições para um vetor de ponderação. O EMD é descrito baseado na distância de Kolmogorov-Smirnov.

### 3.1 Estimador de máxima verossimilhança ponderado (EMVP)

Esses estimadores são uma generalização dos EMV com a inclusão de uma função de ponderação, em que pesos são considerados para reduzir a contaminação causada por *outliers*. Dessa forma, os EMVP são mais resistentes a esses efeitos (FIELD e SMITH, 1994), sendo conhecidos como estimadores robustos. Sob condições usuais de regularidade, os EMVP possuem propriedades assintóticas semelhantes às dos EMV como consistência e normalidade assintótica (WANG et al., 2004).

Dado um vetor de pesos  $\mathbf{w} = (w_1, w_2, \dots, w_n)^\top$ , a função de log-verossimilhança ponderada é dada por:

$$\begin{aligned} \ell_w(\alpha, \beta; \mathbf{x}) &= \sum_{i=1}^n w_i \ell_i(x_i; \alpha, \beta) \\ &= \sum_{i=1}^n w_i \left[ \log[\Gamma(\alpha + \beta)] - \log[\Gamma(\alpha)] - \log[\Gamma(\beta)] \right. \\ &\quad \left. + \alpha \log(x_i) + \beta \log(1 - x_i) - \log(x_i) - \log(1 - x_i) \right]. \end{aligned} \quad (2)$$

Tomando as derivadas de  $\ell_w(\alpha, \beta; \mathbf{x})$  relativamente à  $\alpha$  e à  $\beta$  temos o vetor escore  $\mathbf{U}_w = (U_{w,\alpha}, U_{w,\beta})^\top$ , em que:

$$\begin{aligned} U_{w,\alpha} &= \frac{\partial \ell_w(\alpha, \beta; \mathbf{x})}{\partial \alpha} = \sum_{i=1}^n w_i [\psi(\alpha + \beta)] - \sum_{i=1}^n w_i [\psi(\alpha)] + \sum_{i=1}^n w_i \log(x_i), \\ U_{w,\beta} &= \frac{\partial \ell_w(\alpha, \beta; \mathbf{x})}{\partial \beta} = \sum_{i=1}^n w_i [\psi(\alpha + \beta)] - \sum_{i=1}^n w_i [\psi(\beta)] + \sum_{i=1}^n w_i \log(1 - x_i), \end{aligned}$$

sendo  $\psi(\cdot)$  a função digama, ou seja,  $\psi(u) = \frac{\partial \log \Gamma(u)}{\partial u}$ , para  $u > 0$ .

Tomando o vetor escore igual a zero e resolvendo o sistema de duas equações e duas incógnitas, têm-se os EMVP. Contudo, assim como para os EMV, não há uma solução em forma fechada. Por esse motivo, a função log-verossimilhança ponderada é maximizada por meio de métodos numéricos.

### 3.1.1 Definição do vetor de pesos

O desempenho dos EMVP é influenciado pelos pesos considerados na função de log-verossimilhança ponderada. Segundo FIELD e SMITH (1994), para obtenção de estimativas robustas é preciso limitar a função escore. A inclusão de um vetor de pesos  $w_i$  multiplicado pelo  $i$ -ésimo componente do vetor escore  $U_w$  possibilita a ponderação de valores extremos de  $x$ . Neste trabalho, são consideradas duas formas de ponderação. Elas têm o intuito de ponderar as observações atípicas, de tal forma que aquelas mais afastadas da massa de dados recebam um peso menor para a estimação.

Uma forma usual para a função peso  $w_i$  é dada por (FIELD e SMITH, 1994):

$$w_i = \begin{cases} F(x_i; \alpha, \beta)/p & , \text{ se } F(x_i; \alpha, \beta) < p \\ 1 & , \text{ se } p \leq F(x_i; \alpha, \beta) \leq 1 - p \\ (1 - F(x_i; \alpha, \beta))/p & , \text{ se } F(x_i; \alpha, \beta) > 1 - p \end{cases} \quad (3)$$

em que  $F$  é a função de distribuição acumulada da distribuição beta dada em (1) e  $p$  pode assumir qualquer valor em  $(0, 1)$ , o qual delimita em  $(1 - 2p)\%$  o intervalo da distribuição para ponderação das observações. Assim como em FIELD e SMITH

(1994), será considerado  $p = 0,01$ .

Na Equação (3),  $\alpha$  e  $\beta$  são substituídos por seus respectivos EMV  $\hat{\alpha}$  e  $\hat{\beta}$ . Valores muito grandes ou muito pequenos de  $F(x_i; \hat{\alpha}, \hat{\beta})$  serão ponderados, minimizando desvios nos resultados. Os EMVP baseados na função peso usual, ou seja, os valores que maximizam (2) atribuindo a função peso acima, são chamados de estimadores de máxima verossimilhança ponderados 1 (EMVP<sub>1</sub>), denotados por  $\hat{\alpha}_p^1$  e  $\hat{\beta}_p^1$ .

Uma segunda proposta para determinação dos pesos para os EMVP é dada por um método iterativo. Os estimadores baseados nessa função de ponderação serão denotados estimadores de máxima verossimilhança ponderados 2 (EMVP<sub>2</sub>),  $\hat{\alpha}_p^2$  e  $\hat{\beta}_p^2$ . O algoritmo a seguir descreve o método iterativo proposto:

1. Determine as estimativas para  $\alpha$  e  $\beta$ ;
2. Determine a função peso como em (3), avaliando  $F(\cdot)$  nas estimativas do item 1;
3. Obtenha EMVP via maximização de (2) com os pesos do item 2;
4. Verifique se a diferença percentual máxima entre as estimativas do item 1 e as estimativas do item 3 é inferior a 1%. Se sim, então os valores do item 3 são os EMVP<sub>2</sub>; senão, retorne ao item 1 e repita o procedimento atualizando as estimativas obtidas através de EMVP para os parâmetros  $\alpha$  e  $\beta$ .

No item 1, atribuem-se inicialmente os EMV usuais.

### 3.2 Estimador de mínima distância (EMD)

A distância mínima é definida como um método de estimação geral desenvolvida por WOLFOWITZ (1953) e MATUSITA (1953). Um método conveniente e aplicável de distância mínima foi descrito por WOLFOWITZ (1957), no qual fornece estimadores com boas propriedades de robustez e consistência. A ideia principal é fazer com que a função de distribuição do modelo assumido avaliada nas estimativas, esteja mais “próxima” da função de distribuição empírica das variáveis observadas.

Considerando  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  uma amostra aleatória de uma família paramétrica com função distribuição  $F_\theta(x)$ ,  $\theta \in \Theta$ , a função de distribuição empírica é dada por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

em que  $I$  é a função indicadora. A minimização da distância entre a função de distribuição empírica  $F_n$  e a função de distribuição estimada  $F_{\hat{\theta}}$ , produz estimadores com boas propriedades de robustez e consistência. Uma medida de distância entre duas distribuições será denotada por  $\delta(\cdot, \cdot)$ . Assim, um estimador de mínima

distância de  $\theta$  é o valor em  $\Theta$  que minimiza  $\delta(F_n, F_{\hat{\theta}})$ . Para este trabalho,  $\theta$  são os parâmetros  $\alpha$  e  $\beta$  da distribuição beta e  $F_{\theta}(x)$  é a função distribuição da beta dada em (1).

O método pode ser aplicado com diversas medidas de distância. Consideramos a distância de Kolmogorov-Smirnov (KS). Conforme WILLIAM (1981), a distância KS é dada por:

$$\delta_{KS}(F_n, F_{\hat{\theta}}) = \sup_x |F_n(x) - F_{\hat{\theta}}(x)| \sqrt{\omega(F_{\hat{\theta}}(x))}.$$

Considerando  $\omega(u) \equiv 1$ , obtemos a estatística usual de Kolmogorov-Smirnov. A distância KS mede a máxima discrepância absoluta entre a função de distribuição empírica da amostra  $F_n(x)$  e a função de distribuição acumulada estimada  $F_{\hat{\theta}}(x)$  da distribuição beta. Os valores de  $\theta$  que minimizam a discrepância das distribuições  $(F_n, F_{\theta})$ , ou seja, os estimadores de mínima distância, chamaremos EMD e serão denotados por  $\hat{\alpha}_{\delta}$  e  $\hat{\beta}_{\delta}$ .

#### 4 Correção *bootstrap* na presença de *outliers*

Os estimadores robustos apresentados anteriormente têm bom desempenho na presença de *outliers*. Contudo, conforme resultados de simulação presentes na Seção 5, pode haver um aumento de viés na ausência de *outliers* para os estimadores propostos, comparativamente aos estimadores usuais não robustos. Por este motivo, métodos de redução de viés podem ser empregados (CORDEIRO e CRIBARI-NETO, 2014). Neste trabalho, são utilizadas correções *bootstrap* (EFRON e TIBSHIRANI, 1994) para a redução de viés dos estimadores.

O método *bootstrap* é um método computacionalmente intensivo que consiste na geração de reamostras a partir da amostra originalmente observada. Essas reamostras podem ser obtidas diretamente a partir das observações (*bootstrap* não paramétrico) ou da distribuição estimada (*bootstrap* paramétrico). Ao utilizar o *bootstrap* não paramétrico, a reamostragem pode ser comprometida quando *outliers* estão presentes no conjunto de dados. Com o efeito de propagação desses valores aberrantes, há o risco de reamostrá-los várias vezes, fazendo com que a proporção de *outliers* na amostra se torne maior que nos dados originais (AMADO et al., 2014). Por outro lado, o *bootstrap* paramétrico gera pseudoamostras de uma função de distribuição conhecida avaliada nas estimativas dos parâmetros a partir dos dados originais. Quando não são utilizados estimadores robustos, a presença de *outliers* entre as observações pode causar distorções na obtenção das estimativas dos parâmetros, fazendo com que as reamostras provenientes do *bootstrap* paramétrico estejam “distantes” da distribuição da amostra original.

Como alternativa para contornar os possíveis problemas na reamostragem *bootstrap* causados pela presença de *outliers*, consideramos o *bootstrap* paramétrico em que as pseudoamostras são realizações da função distribuição beta avaliada em estimativas robustas. Seja  $\theta = (\alpha, \beta)^T$  o vetor paramétrico da distribuição beta e  $\hat{\theta}$  qualquer estimador de  $\theta$ . A correção *bootstrap* para  $\hat{\theta}$  na presença de *outlier* pode

ser descrita da seguinte forma:

1. Estima-se  $\theta = (\alpha, \beta)^\top$ , a partir da amostra original, por meio de um método robusto. Sugere-se a utilização dos EMVP<sub>2</sub>, ou seja,  $\hat{\alpha}_p^2$  e  $\hat{\beta}_p^2$ .
2. A partir das estimativas robustas obtidas, geram-se  $B$  reamostras *bootstrap*  $\mathbf{x}_b^*$ , com  $n$  observações, a partir da distribuição Beta( $\hat{\alpha}_p^2, \hat{\beta}_p^2$ ), com  $b = 1, \dots, B$ .
3. Para cada reamostra *bootstrap*  $\mathbf{x}_b^*$  obtêm-se as réplicas *bootstrap*  $\hat{\theta}_b^*$  de  $\hat{\theta}$ , obtendo assim  $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$ .
4. Obtém-se uma estimativa *bootstrap* do viés de  $\hat{\theta}$ , da seguinte forma:

$$\hat{B}_{boot}(\hat{\theta}) = \bar{\theta}^* - \hat{\theta},$$

$$\text{em que } \bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \theta^{*b}.$$

5. Obtém-se o estimador corrigido por *bootstrap*:

$$\hat{\theta}^* = \hat{\theta} - \hat{B}_{boot}(\hat{\theta}) = \hat{\theta} - (\bar{\theta}^* - \hat{\theta}) = 2\hat{\theta} - \bar{\theta}^*.$$

Para as reamostras *bootstrap* do item 2 no algoritmo anterior, foi testada a utilização de todos os estimadores descritos anteriormente. Os melhores resultados foram observados quando os estimadores EMVP<sub>2</sub> foram utilizados. Por este motivo, recomenda-se o uso do EMVP<sub>2</sub> na geração das pseudoamostras pelo método *bootstrap* paramétrico.

## 5 Avaliação numérica

Nesta seção, avaliamos numericamente os EMM, EMV, EMVP<sub>1</sub>, EMVP<sub>2</sub> e o EMD, assim como as versões corrigidas por *bootstrap* de cada um desses estimadores. A avaliação foi realizada através de simulações de Monte Carlo. Todas as implementações computacionais foram desenvolvidas em linguagem R (R CORE TEAM, 2019). O número de réplicas de Monte Carlo foi fixado em  $R = 1000$  e o número de reamostras *bootstrap* em  $B = 500$ , com tamanhos amostrais de  $n = 30, 50$  e  $100$ . Nos procedimentos de maximização das funções de log-verossimilhança e log-verossimilhanças ponderadas foi utilizado o método quasi-Newton BFGS (NOCEDAL e WRIGHT, 1999) com primeiras derivadas analíticas. Para minimização da discrepância entre as distâncias que compõem o EMD foi utilizado o método implementado por NELDER e MEAD (1965).

Com o propósito de avaliar os estimadores foram calculados média, viés, viés relativo percentual (VR%), erro padrão e erro quadrático médio (EQM). As Tabelas 1 e 2 apresentam os resultados da avaliação numérica dos estimadores estudados com parâmetros  $\alpha = 2$  e  $\beta = 8$ , para tamanho amostral  $n = 30$ . Os estimadores corrigidos por *bootstrap* são denotados com asterisco (\*). A Tabela 1 apresenta

resultados quando a amostra não possui *outliers* e a Tabela 2 com a presença de um *outlier* igual a 0,95. Cabe destacar que para os valores paramétricos  $\alpha = 2$  e  $\beta = 8$  a média de  $X$  é igual a 0,2.

Tabela 1 - Resultados da simulação de Monte Carlo para avaliação dos estimadores dos parâmetros da distribuição beta com  $\alpha = 2$  e  $\beta = 8$ , baseados em uma amostra de tamanho 30 sem *outliers*.

Estimador	Média	Viés	VR(%)	Erro Padrão	EQM
$\bar{\alpha}$	2,14	0,14	6,99	0,61	0,39
$\hat{\alpha}$	2,19	0,19	9,36	0,57	0,36
$\hat{\alpha}_p^1$	2,31	0,31	15,55	0,64	0,41
$\hat{\alpha}_p^2$	2,37	0,37	18,71	0,69	0,48
$\hat{\alpha}_\delta$	2,26	0,26	12,84	0,75	0,57
-----	-----	-----	-----	-----	-----
$\bar{\alpha}^*$	2,19	0,19	9,98	0,64	0,45
$\hat{\alpha}^*$	2,13	0,13	6,54	0,62	0,40
$\hat{\alpha}_p^{1*}$	1,97	-0,03	-1,38	0,57	0,32
$\hat{\alpha}_p^{2*}$	1,89	-0,10	-5,26	0,54	0,30
$\hat{\alpha}_\delta^*$	2,06	0,06	2,87	0,59	0,36
-----	-----	-----	-----	-----	-----
Estimador	Média	Viés	VR(%)	Erro Padrão	EQM
$\bar{\beta}$	8,69	0,69	8,69	2,67	7,78
$\hat{\beta}$	8,89	0,89	11,17	2,58	7,47
$\hat{\beta}_p^1$	9,44	1,44	18,03	2,87	8,26
$\hat{\beta}_p^2$	9,72	1,72	21,54	3,09	9,55
$\hat{\beta}_\delta$	9,20	1,20	15,04	3,34	11,15
-----	-----	-----	-----	-----	-----
$\bar{\beta}^*$	8,91	0,91	11,31	2,83	8,86
$\hat{\beta}^*$	8,62	0,62	7,72	2,74	7,90
$\hat{\beta}_p^{1*}$	7,90	-0,10	-1,26	2,51	6,33
$\hat{\beta}_p^{2*}$	7,54	-0,45	-5,68	2,40	5,98
$\hat{\beta}_\delta^*$	8,28	0,28	3,50	2,64	7,04

Ao analisar a Tabela 1, percebe-se que os estimadores robustos sem correção *bootstrap* apresentam um desempenho um pouco inferior aos estimadores usuais não corrigidos, em termos de VR, quando não há valor discrepante na amostra. Contudo, após correção *bootstrap* há relevante decréscimo de VR para os EMVP e EMD, ou seja, a média estimada apresenta-se mais próxima dos valores dos parâmetros  $\alpha = 2$  e  $\beta = 8$ . Por exemplo, observa-se que o estimador  $\hat{\beta}_p^1$  possui VR igual a 18,03%, enquanto que sua versão corrigida possui VR = -1,26%.

A partir da Tabela 2, em que está presente um *outlier* na amostra aleatória, é perceptível que os estimadores robustos fornecem estimativas menos viesadas para os parâmetros  $\alpha$  e  $\beta$  do que os estimadores usuais. Os VR dos estimadores usuais na presença de *outlier* são destacadamente grandes. Por exemplo, os estimadores usuais  $\bar{\beta}$  e  $\hat{\beta}$  na presença de *outlier* apresentam, respectivamente, VR

Tabela 2 - Resultados da simulação de Monte Carlo para avaliação dos estimadores dos parâmetros da distribuição beta com  $\alpha = 2$  e  $\beta = 8$ , baseados em uma amostra de tamanho 30 com *outlier*.

Estimador	Média	Viés	VR(%)	Erro Padrão	EQM
$\bar{\alpha}$	0,98	-1,02	-51,19	0,18	1,08
$\hat{\alpha}$	1,17	-0,83	-41,56	0,16	0,71
$\hat{\alpha}_p^1$	2,21	0,21	10,51	0,57	0,32
$\hat{\alpha}_p^2$	2,36	0,36	17,89	0,67	0,45
$\hat{\alpha}_\delta$	2,10	0,11	5,44	0,66	0,44
-----					
$\bar{\alpha}^*$	2,18	0,18	9,22	0,62	0,42
$\hat{\alpha}^*$	2,12	0,12	5,85	0,60	0,37
$\hat{\alpha}_p^{1*}$	1,96	-0,04	-2,02	0,55	0,30
$\hat{\alpha}_p^{2*}$	1,88	-0,12	-5,88	0,52	0,29
$\hat{\alpha}_\delta^*$	2,05	0,04	2,24	0,58	0,33
Estimador	Média	Viés	VR(%)	Erro Padrão	EQM
$\bar{\beta}$	3,38	-4,62	-57,70	0,42	21,49
$\hat{\beta}$	3,53	-4,47	-55,91	0,37	20,15
$\hat{\beta}_p^1$	9,06	1,06	13,28	2,59	6,70
$\hat{\beta}_p^2$	9,74	1,74	21,76	3,05	9,31
$\hat{\beta}_\delta$	8,26	0,26	3,32	2,93	8,57
-----					
$\bar{\beta}^*$	8,92	0,92	11,45	2,80	8,86
$\hat{\beta}^*$	8,63	0,63	7,88	2,71	7,74
$\hat{\beta}_p^{1*}$	7,91	-0,09	-1,13	2,48	6,17
$\hat{\beta}_p^{2*}$	7,55	-0,45	-5,57	2,37	5,83
$\hat{\beta}_\delta^*$	8,30	0,30	3,72	2,61	6,88

de  $-57,70\%$  e  $-55,91\%$ , enquanto que o estimador  $\hat{\beta}_\delta$  possui VR igual a  $3,32\%$ . Adicionalmente, percebe-se que a correção *bootstrap* considerada também foi capaz de melhorar os estimadores usuais na presença de *outliers*. Ou seja, a correção *bootstrap* proposta, além de diminuir o viés dos estimadores robustos, tornou EMM e EMV mais robustos a *outliers*. Isso se deve ao fato das reamostras serem geradas da distribuição beta avaliada em bons estimadores robustos, diminuindo consideravelmente a interferência de *outliers*.

Contudo, a melhor forma de avaliar a robustez de estimadores é por meio de medidas específicas como ponto de ruptura e curva de sensibilidade (MARONNA et al., 2006). Essa avaliação é feita a seguir.

### 5.1 Ponto de ruptura

O ponto de ruptura é uma das medidas mais utilizadas na avaliação de estimadores quando se trata de robustez. O ponto de ruptura de um estimador é dado por  $\frac{m}{n}$ , para uma amostra com  $n$  elementos, em que  $m$  é o maior número de

pontos que podem ser substituídos por valores arbitrários sem que os estimadores dos parâmetros deixem de ser limitados (HAMPEL, 1971). Assim, quanto maior for o ponto de ruptura, mais robusto a *outliers* o estimador será.

Utilizando o viés relativo, têm-se os gráficos dos pontos de ruptura dos estimadores sem correção *bootstrap* na Figura 2 e com correção na Figura 3. Como em casos anteriores,  $\alpha = 2$ ,  $\beta = 8$  e  $n = 30, 50$ . O valor dos *outliers* foi definido 0,95 por estar distante da média igual a 0,2. Pela Figura 2, observa-se que o viés relativo dos estimadores usuais já na presença de apenas um *outlier* mostra-se distante do valor ideal zero. O crescimento da quantidade de *outliers* naturalmente acompanha o crescimento da magnitude do viés nos estimadores. Contudo, é perceptível que os estimadores robustos apresentam melhores resultados. Para a amostra de tamanho 30, os EMVP suportam uma contaminação de 3 a 4 *outliers* no conjunto de dados, até se igualar ao EMV. O mesmo ocorre com  $n = 50$ ; o ponto de ruptura permanece em torno de 10% do tamanho da amostra ( $m = 5$ ). Já o comportamento dos EMD mostram inicialmente VR baixo seguido de seu aumento gradual de forma mais constante.

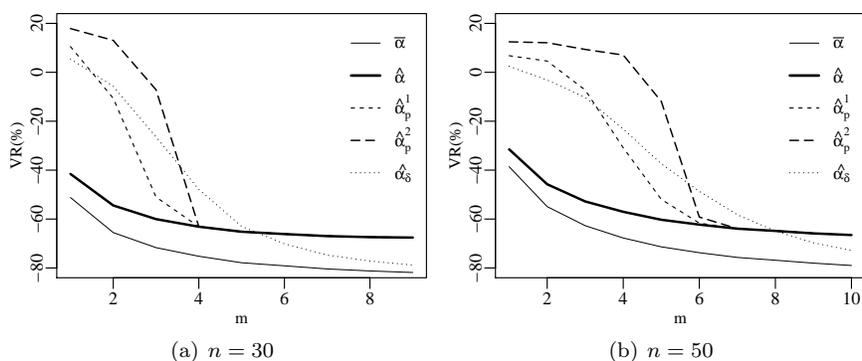


Figura 2 - Ponto de ruptura considerando viés relativo dos estimadores não corrigidos de  $\alpha = 2$  para tamanhos amostrais 30 e 50.

Simulações com tamanhos amostrais inferiores a  $n = 30$ , não puderam ser consideradas, devido ao grande número de *outliers* introduzidos comparativamente ao tamanho da amostra. Já em tamanhos amostrais maiores, o ponto de ruptura também aumentou. Por exemplo, para  $n = 100$ , o EMVP suporta de 9 a 10 *outliers*, caracterizando um ponto de ruptura de aproximadamente 10% do tamanho amostral.

Apesar de os estimadores robustos apresentarem melhores resultados (Figura 2), os mesmos ainda possuem notável viés. A Figura 3 apresenta os resultados da avaliação do ponto de ruptura dos estimadores corrigidos por *bootstrap*. Note que os pontos de ruptura permanecem os mesmos, porém há um expressivo decréscimo no viés. EMM e EMV com a correção apresentam pontos de ruptura semelhantes



observação dentro da amostra. A função de sensibilidade é dada por:

$$CS(x_1; x_2, \dots, x_n, T_n) = [T_n(x_1, x_2, \dots, x_n) - T_{n-1}(x_2, \dots, x_n)],$$

sendo  $(x_1, x_2, \dots, x_{n-1}, x_n)$  uma amostra com  $n$  elementos e  $T$  o estimador avaliado. É importante observar que  $x_1$  será um valor a ser arbitrado, variando de pouco a muito influente.

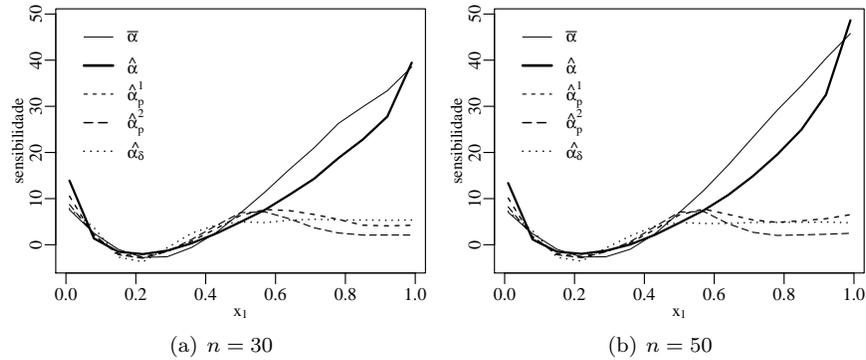


Figura 5 - Curva de sensibilidade dos estimadores não corrigidos de  $\alpha = 2$  para tamanhos amostrais 30 e 50.

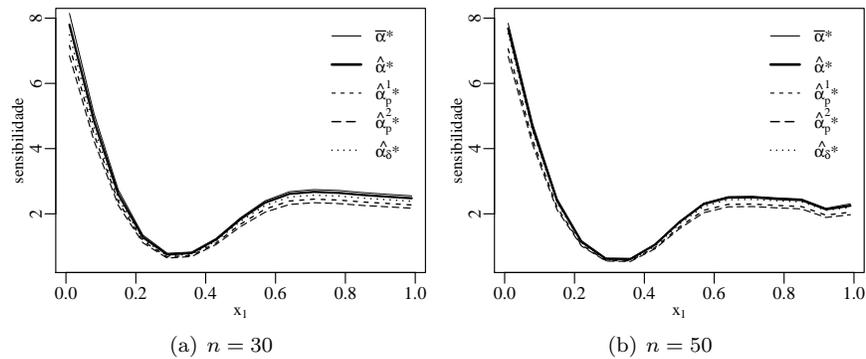


Figura 6 - Curva de sensibilidade dos estimadores corrigidos via *bootstrap* de  $\alpha = 2$  para tamanhos amostrais 30 e 50.

A Figura 5 apresenta as curvas de sensibilidade dos EMM, EMV, EMVP<sub>1</sub>, EMVP<sub>2</sub> e EMD não corrigidos. São consideradas amostras com  $n = 30$  e  $n = 50$ , ambas oriundas da distribuição beta com  $\alpha = 2$  e  $\beta = 8$ . O elemento da amostra  $x_1$  (*outlier*) varia no intervalo  $(0, 1)$ . O mesmo experimento foi realizado considerando

os estimadores corrigidos, apresentados na Figura 6.

Percebe-se na Figura 5 que os estimadores usuais se tornam mais sensíveis à medida que o valor  $x_1$  se afasta da média da distribuição beta (0, 2 neste caso). Por outro lado, os estimadores robustos e os corrigidos, mostraram-se mais próximos de zero e, conseqüentemente, menos sensíveis a valores discrepantes. Além disso, observa-se na Figura 6, por meio da correção de viés realizada, o decréscimo da sensibilidade a *outliers* nas estimativas clássicas (EMM e EMV) corrigidas.

A Figura 7 mostra o efeito da correção *bootstrap* proposta nos estimadores com base em sua sensibilidade a valores discrepantes, para  $n = 50$ . Pode-se verificar que a correção para os estimadores robustos provocou menos oscilações na sensibilidade e a manteve baixa. Para EMM e EMV não corrigidos, nota-se uma expressiva sensibilidade, que após correção apresenta-se próxima de zero. Portanto, a correção *bootstrap* proposta, além de melhorar o viés, tornou os estimadores usuais mais resistentes a *outliers*.

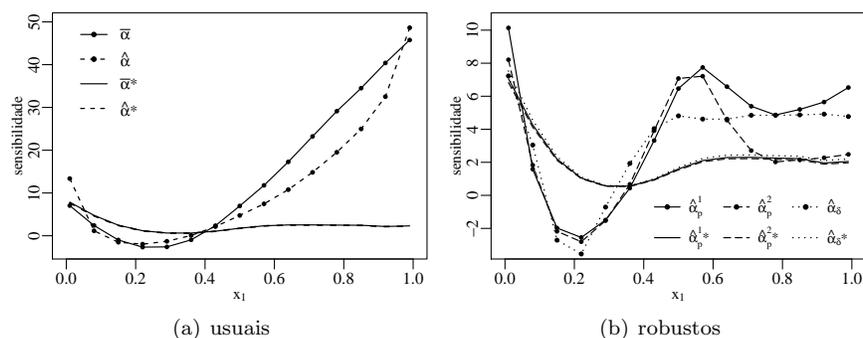


Figura 7 - Sensibilidade dos estimadores antes e depois da correção *bootstrap*.

Os resultados do estudo de simulação de Monte Carlo, de forma geral, indicam que o  $EMVP_2$  corrigido por *bootstrap* apresenta bons resultados tanto em medidas habituais para avaliação de estimadores quanto em medidas específicas de robustez. Além disso, o melhoramento inferencial alcançado via *bootstrap* proposto foi baseado em reamostragens baseadas no  $EMVP_2$ . Portanto, destaca-se a versão corrigida do  $EMVP_2$  ( $\hat{\alpha}_p^{2*}$ ,  $\hat{\beta}_p^{2*}$ ) como o melhor estimador robusto para os parâmetros da distribuição beta.

## 6 Aplicação a dados reais

Foram considerados dados referentes às informações médicas de vários atletas do *Australian Institute of Sports* (AIS). Os dados estão disponíveis no pacote `sn` (AZZALINI, 2017) do R (R CORE TEAM, 2019). Dentre as variáveis apresentadas, optou-se por modelar os dados sobre os hematócritos em 102 homens. A medida de hematócritos representa a porcentagem do volume ocupado pelos glóbulos vermelhos

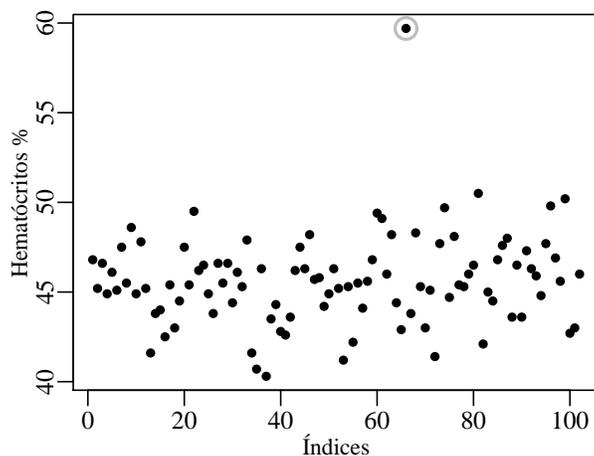


Figura 8 - Dados do AIS da porcentagem de hematócritos em 102 atletas masculinos.

ou hemácias dentre o volume total do sangue. Logo, sendo uma proporção, se torna adequado modelar estes dados com a distribuição beta.

Conforme KEEN (1998), homens saudáveis apresentam hematócritos entre 42% e 53%. Aproximadamente 8% dos homens possuem taxas menores que 42% e apenas 2% apresentam mais de 54% de hematócritos no volume total de sangue (ISHANI et al., 2006). Neste conjunto de dados, representado na Figura 8, é possível perceber que a maioria das observações se encontra na faixa esperada, porém podemos observar claramente um *outlier* próximo ao 60%. A escolha por esses dados se deu justamente pela presença deste evidente valor discrepante, objetivando demonstrar a capacidade dos estimadores robustos em diminuir, ou anular, a influência causada pelo *outlier* na estimação dos parâmetros da distribuição beta.

A análise dos estimadores nesses dados reais é feita a partir da comparação entre as estimativas: (i) primeiramente com *outlier* naturalmente presente e, (ii) posteriormente, eliminando-o. A maior proximidade das estimativas dos parâmetros entre os dois casos evidencia a robustez perante valores discrepantes dos estimadores considerados. A Tabela 3 apresenta as estimativas de  $\alpha$  e  $\beta$  baseadas em todos os estimadores considerados neste trabalho, antes e depois de eliminar o *outlier*. Percebemos que ao não considerar a correção *bootstrap* (nas duas primeiras colunas), os EMV apresentam uma expressiva diferença nas estimativas de  $\alpha$  e  $\beta$  na presença ou ausência de *outlier*. Ao considerar a amostra com o *outlier*, obtém-se  $\hat{\alpha} = 171,2504$ , enquanto que ao remover o *outlier* o valor de  $\hat{\alpha}$  passa a ser igual a 241,2483, resultando em uma mudança percentual de 40,9%.

A aplicação neste conjunto de dados confirmou o bom desempenho dos estimadores propostos. É possível verificar a proximidade nas estimativas robustas de  $\alpha$  e  $\beta$ . Por exemplo, a estimativa de EMVP<sub>2</sub> que apresentou  $\hat{\alpha}_p^2 = 243,6810$

quando a amostra continha o *outlier* e um  $\hat{\alpha}_p^2 = 244,3159$  ao extraí-lo, totalizando alteração de apenas 0,26%. Além disso, é possível constatar que após a correção *bootstrap* proposta todos estimadores apresentam estimativas mais resistentes à presença do valor aberrante. Destacam-se os estimadores usuais corrigidos, os quais obtiveram relevante diferença em comparação a quando a correção não é realizada. Como exemplo, apontamos na Tabela 3 o EMV obtido na presença de *outlier* com  $\hat{\alpha}^* = 239,0538$ . Após a exclusão do *outlier*, a estimativa apresenta valor de 239,2735, confirmando a robustez.

Tabela 3 - Estimativas sem correção e com correção *bootstrap* dos parâmetros da distribuição beta aplicada aos dados de porcentagem de hematócritos.

	Sem correção <i>bootstrap</i>		Com correção <i>bootstrap</i>		
	Com <i>outlier</i>	Sem <i>outlier</i>	Com <i>outlier</i>	Sem <i>outlier</i>	
$\bar{\alpha}$	171.2299	241.2474	$\bar{\alpha}^*$	239.0808	239.3024
$\bar{\beta}$	203.8630	288.8398	$\bar{\beta}^*$	285.9951	286.2117
$\hat{\alpha}$	171.2504	241.2483	$\hat{\alpha}^*$	239.0538	239.2735
$\hat{\beta}$	203.8572	288.8407	$\hat{\beta}^*$	285.9638	286.1779
$\hat{\alpha}_p^1$	240.7628	244.2963	$\hat{\alpha}_p^{1*}$	221.4283	221.6989
$\hat{\beta}_p^1$	288.2590	292.3428	$\hat{\beta}_p^{1*}$	264.8526	265.1645
$\hat{\alpha}_p^2$	243.6810	244.3159	$\hat{\alpha}_p^{2*}$	216.9852	217.0616
$\hat{\beta}_p^2$	291.6092	292.3384	$\hat{\beta}_p^{2*}$	259.5326	259.6081
$\hat{\alpha}_\delta$	257.9978	262.2636	$\hat{\alpha}_\delta^*$	234.2959	235.0945
$\hat{\beta}_\delta$	307.9539	313.4072	$\hat{\beta}_\delta^*$	280.2310	281.1710

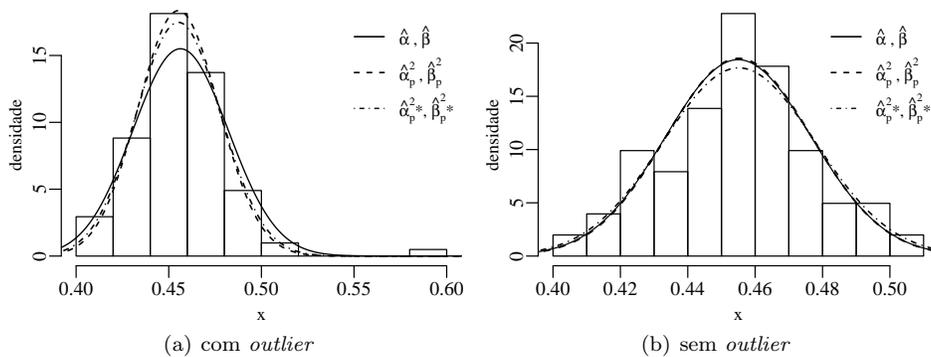


Figura 9 - Histograma e densidades estimadas por método usual e robustos a partir dos dados de hematócritos.

A Figura 9 apresenta os histogramas de frequência relativa dos dados reais e as densidades estimadas obtidas pelo método usual de máxima verossimilhança e pelos estimadores robustos propostos, EMVP<sub>2</sub> e EMVP<sub>2</sub> corrigido. As estimativas dos parâmetros utilizados se encontram na Tabela 3. Na Figura 9(a), observa-se que as funções densidades baseadas nos estimadores robustos apresentam melhor adequação aos dados com *outlier*. As estimativas  $\hat{\alpha}_p^2, \hat{\beta}_p^2$  e  $\hat{\alpha}_p^{2*}, \hat{\beta}_p^{2*}$  são menos influentes ao *outlier* em relação às estimativas de  $\hat{\alpha}, \hat{\beta}$  obtidas pelo estimador usual EMV. A Figura 9(b) evidencia similaridade entre as densidades estimadas, mostrando que os estimadores propostos apresentam comportamento adequado mesmo em dados não contaminados por valores discrepantes.

## Conclusões

Este trabalho apresenta abordagens de estimação robusta para os parâmetros da distribuição beta. Os estimadores pontuais usuais pelos métodos dos momentos e de máxima verossimilhança, quando os dados estão contaminados por valores discrepantes, não apresentam bom desempenho. São propostas duas versões de estimadores de máxima verossimilhança ponderados e uma baseada na distância mínima entre funções de distribuição empírica e teórica. A avaliação numérica, por meio de simulações de Monte Carlo, evidenciou superioridade dos estimadores robustos quando há presença de *outliers* na amostra, os quais se mostraram menos sensíveis e com ponto de ruptura maior. Neste mesmo contexto, é introduzida uma correção de viés via método *bootstrap* paramétrico baseado nas estimativas robustas dadas pelo EMVP<sub>2</sub>. Os estimadores corrigidos pelo método *bootstrap* robusto apresentaram melhoria/redução em termos de viés, viés relativo e erro quadrático médio. Por fim, a aplicação à dados reais confirmou a adequação e a aplicabilidade dos estimadores robustos considerados.

## Agradecimentos

Os autores agradecem à FAPERGS, à CAPES e ao CNPq pelo apoio financeiro e aos revisores e editores pelas sugestões.

MANCHINI, C. E. F.; PORTELLA, L.; BAYER, F. M. Robust estimation of beta distribution parameters. *Rev. Bras. Biom.*, Lavras, v.37, n.3, p.350-371, 2019.

- **ABSTRACT:** *This paper proposes robust estimators for parameters of the beta distribution, useful for modelling continuous data restricted to the interval (0,1) in the presence of outliers. Weighted maximum likelihood estimators (WMLE) with two weighting proposals and a minimum distance estimator (MDE) were considered. The bias correction of these estimators was performed using the parametric bootstrap method adapted for data with outliers. The performances of the proposed estimators were evaluated by means of Monte Carlo simulations and their results compared to the performance of the usual estimators by the moment and maximum likelihood methods. The breakdown point and the sensitivity curve were adopted as specific measures of robustness. The numerical evaluation pointed out that the WMLE and MDE have lower sensitivity to outliers and greater breakdown points. The estimators corrected by the proposed bootstrap method had lower bias compared to their versions uncorrected. Moreover, an application to real data is also presented and discussed.*
- **KEYWORDS:** *Bootstrap; robust estimators; minimum distance; outliers; weighted likelihood.*

## Referências

- AMADO, C.; BIANCO, A. M.; BOENTE, G. L.; PIRES, A. M. Robust bootstrap: an alternative to bootstrapping robust estimators. *REVSTAT Statistical Journal*, v.12, n.2, p.169-197, 2014.
- AHMED, E. S.; VOLODIN, A. I.; HUSSEIN, A. A. Robust weighted likelihood estimation of exponential parameters. *IEEE Transactions on Reliability*, v.54, n.3, p.389-395, 2005.
- AZZALINI, A. *The R package sn: The Skew-Normal and Skew-t distributions.* (version 1.5-0). Università di Padova, Italia, 2017.
- BAYER, F.; TONDOLO, C. M.; MÜLLER, F. Beta regression control chart for monitoring fractions and proportions. *Computer and Industrial Engineering*, v.119, p.416-426, 2018a.
- BAYER, F. M.; CINTRA, R. J.; CRIBARI-NETO, F. Beta seasonal autoregressive moving average models. *Journal of Statistical Computation and Simulation*, v.88, n.15, p.2961-2981, 2018b.
- BECKMAN, R. J.; JEN, G. L. T. Maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, v.7, n.3-4, p.253-258, 1978.
- BLASQUES, F.; KOOPMAN, S. J.; MALLE, M.; ZHANG, Z. Weighted maximum likelihood for dynamic factor analysis and forecasting with mixed frequency data. *Journal of Econometrics*, v.196, n.2, p.405-417, 2016.

- CHEN, C.; LIU, L.-M. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, v.88, n.421, p.284-297, 1993.
- CORDEIRO, G. M.; CRIBARI-NETO, F. *An introduction to Bartlett correction and bias reduction*. Heidelberg: Springer, 2014.
- DE LA ROSA, J. I.; FLEURY, G. A. Bootstrap methods for a measurement estimation problem. *IEEE Transactions on Instrumentation and Measurement*, v.55, n.3, p.820-827, 2006.
- EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. New York: Chapman and Hall/CRC, 1994.
- FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v.31, n.7, p.799-815, 2004.
- FIELD, C.; SMITH, B. Robust estimation: A weighted maximum likelihood approach. *International Statistical Review*, v.62, n.3, p.405-424, 1994.
- FRANK, J.; MASSEY, J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, v.46, n.253, p.68-78, 1951.
- GUPTA, A. K. Beta distribution. In: *International Encyclopedia of Statistical Science*. New York: Springer, p.144-145, 2011.
- GUPTA, A. K.; NADARAJAH, S. *Handbook of Beta distribution and its applications*. Boca Raton: CRC Press, 2004.
- HAHN, E. D. Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research*, v.188, n.2, p.450-459, 2008.
- HAMPEL, F. R. A general qualitative definition of robustness. *Annals of Mathematical Statistics*, v.42, n.6, p.1887-1896, 1971.
- HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEEUW, P. J.; STAHEL, W. A. *Robust statistics: The approach based on influence functions*. New York: John Wiley & Sons, 2011.
- HU, F.; ZIDEK, J. V. The weighted likelihood. *Canadian Journal of Statistics*, v.30, n.3, p.347-371, 2002.
- HUBER, P. J. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, v.35, n.1, p.73-101, 1964.
- HUBER, P. J. Robust statistics. In: *International encyclopedia of statistical science*. New York: Springer, p.1248-1251, 2011.

- ISHANI, A.; GRANDITS, G.; GRIMM, R.; SVENDSEN, K.; COLLINS, A.; PRINEAS, R.; NEATON, J.; GRP, M. R. Association of single measurements of dipstick proteinuria, estimated glomerular filtration rate, and hematocrit with 25-year incidence of end-stage renal disease in the multiple risk factor intervention trial. *Journal of the American Society of Nephrology*, v.17, n.5, p.1444-1452, 2006.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. *Continuous univariate distributions*. v.2. New York: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, 1995.
- KEEN, M. L. Hemoglobin and hematocrit: An analysis of clinical accuracy. Case study of the anemic patient. *Nephrology Nursing Journal*, v.25, n.1, p.83, 1998.
- LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, v.84, n.408, p.881-896, 1989.
- MARONNA, R.; MARTIN, D.; YOHAI, V. *Robust Statistics*. Chichester: John Wiley & Sons, 2006.
- MATUSITA, K. On the estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, v.5, n.1, p.59-65, 1953.
- NELDER, J. A.; MEAD, R. A simplex method for function minimization. *The Computer Journal*, v.7, n.4, p.308-313, 1965.
- NOCEDAL, J.; WRIGHT, S. *Numerical optimization*. New York: Springer, 1999.
- OWEN, C. E. B. *Parameter estimation for the beta distribution*. Brigham Young University, 2008.
- PARR, W. C.; SCHUCANY, W. R. Minimum distance and robust estimation. *Journal of the American Statistical Association*, v.75, n.371, p.616-624, 1980.
- PAWITAN, Y. *All likelihood: Statistical modelling and inference using likelihood*. Oxford: Oxford University Press, 2001.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.  
URL <https://www.R-project.org/>, 2019.
- ROUSSEEUW, P. J.; LEROY, A. M. *Robust regression and outlier detection*. New York: John Wiley & Sons, 2003.
- SANTANNA, A. O.; TEN CATEN, C. S. Beta control charts for monitoring fraction data. *Expert Systems with Applications*, v.39, n.11, p.10236-10243, 2012.
- SEKI, T.; YOKOYAMA, S. Robust parameter-estimation using the bootstrap method for the 2-parameter Weibull distribution. *IEEE Transactions on Reliability*, v.45, n.1, p.34-41, 1996.

- WANG, X.; VAN EEDEN, C.; ZIDEK, J. V. Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference*, v.119, n.1, p.37-54, 2004.
- WILLIAM, C. P. Minimum distance estimation:a bibliography. *Communications in Statistics - Theory and Methods*, v.10, n.12, p.1205-1224, 1981.
- WOLFOWITZ, J. Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics*, v.5, n.1, p 9-23, 1953.
- WOLFOWITZ, J. The minimum distance method. *The Annals of Mathematical Statistics*, v.28, n.1, p.75-88, 1957.

Recebido em 14.09.2018.

Aprovado após revisão em 13.02.2019.