Momentos-L na estimação dos parâmetros da distribuição generalizada de valores extremos*

Souza, Mesek Felipe de;
{Discente do curso de Matemática Licenciatura}, {UNIFAL-MG}
mekvemon@gmail.com
Avelar, Fabricio Goecking
{Instituto de Ciências Exatas, Departamento de Estatística }, {UNIFAL-MG}
fabricio@unifal-mg.edu.br
Beijo, Luiz Alberto
{Instituto de Ciências Exatas, Departamento de Estatística }, {UNIFAL-MG}
luiz.beijo@unifal-mg.edu.br

3 de Julho de 2020

1 Introdução

De acordo com Brito e Veiga [1], define-se como eventos extremos aqueles que apresentam um grande desvio em relação a média. Em muitos casos eles provocam grandes impacto como, por exemplo, o lucro ou o prejuízo causados pela alta ou queda extremas dos valores de ações de uma determinada empresa e a perda de vidas humanas ocasionadas por enchentes. Os eventos extremos apresentam ocorrência e frequência rara em relação ao tempo, limitando, dessa forma, a quantidade de informação para seu estudo. A Teoria de Valores Extremos (TVE) possui ferramentas específicas para tratar e estudar os eventos extremos. Assim muitas áreas da ciência, como engenharias, hidrologia, climatologia, economia, entre outras, a aplicam.

Uma distribuição originada da TVE é a Generalizada de Valores Extremos (GEV), que descreve tanto os valores mínimos quanto os valores máximos. A função densidade de probabilidade da distribuição GEV é dado pela expressão

$$f\left(x|\,\mu,\sigma,\xi\right)=\frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\xi}{\xi}\right)}\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

definida em $-\infty\ < x < \mu-\frac{\sigma}{\xi}$ para $\xi<0$ e $\mu-\frac{\sigma}{\xi}<x<\infty$ para $\xi>0$, em que $\mu\in\mathbb{R}$ é o parâmetro de posição, $\sigma>0$ é o de escala e $\xi\neq 0$, de forma. Para utilizar a GEV na modelagem de eventos extremos deve-se conhecer o valor de cada parâmetro, considerando-se a amostra observada. Para isso, busca-se estima-los por meio de um estimador adequado.

O método da máxima verossimilhança é comumente utilizado para a estimação dos parâmetros da distribuição GEV. Esse método consiste em se encontrar os valores dos parâmetros que maximizam a função de verossimilhança. Para isso, deriva-se parcialmente a função de verossimilhança em relação a cada um de seus parâmetros e iguala-se as derivadas parciais a zero. A função de verossimilhança de uma variável aleatória contínua com função densidade $f(x)$ é definida por

$$L(\theta|\mathbf{x})=\prod_{i=1}^{n}f(x_i|\theta),$$

em que $\mathbf{x}=(x_1,x_2,\cdots,x_n)$ é o vetor com os valores observados em uma amostra de tamanho n.

Normalmente, para facilitar os cálculos sem alterar os resultados, maximiza-se a função suporte da verossimilhança, que é igual ao logaritmo natural da função de verossimilhança.

Contudo, de acordo com Smith [4], para valores de $\xi>-0,5$, existem estimadores de máxima verossimilhança e eles são regulares. Quando $-1\leq\xi\leq-0,5$, os estimadores existem mas não atendem

*este trabalho conta com apoio financeiro da CNPq

as condições de regularidade necessárias para a utilização da máxima verossimilhança. No caso em que $\xi < -1$ os estimadores de máxima verossimilhança não existem. Logo, quando $\xi \leq -0,5$, pode ser necessário o uso de outros estimadores ao invés do método da máxima verossimilhança. Uma alternativa é o uso dos método dos momentos-L.

De acordo com Hosking [2], pode-se utilizar o método dos momentos-L para calcular as estimativas dos parâmetros de diversas distribuições, sendo a GEV uma delas. De acordo com Valverde [5], o i-ésimo momento-L populacional de uma variável aleatória X, denotado por λ_i, é dado pela expressão

$$\lambda_i = \int_0^1 Q(x)(P_{i-1} * F(x))f(x)dx$$

em que $f(x), F(x), Q(x)$ são, respectivamente, a função densidade de probabilidade, função de distribuição acumulada e a função quantil de X e $P_i * F(x) = \sum_{k=0}^{i} (-1)^{1-k} \binom{i}{k} \binom{i+k}{k} F(x)^k$, $i = 1, 2, \ldots$

O i-ésimo momento-L amostral é expresso por:

$$l_i = \left[\binom{n}{i}\right]^{-1} \sum_{1 \leq j_1} \sum_{< j_2 <} \cdots \sum_{< j_i \leq n} i^{-1} \sum_{k=0}^{i-1} \binom{i-1}{k} x_{j_{i-k}:n},$$

com $i = 1,\ 2,\ \cdots,\ n$ e uma amostra aleatória $x_1,\ x_2, \cdots, x_n$ de tamanho n da variável aleatória X e $x_{1:n} \leq x_{2:n} \leq \cdots \leq x_{n:n}$ a amostra ordenada.

O método dos momentos-L, para uma distribuição com k parâmetros, consiste na obtenção de estimadores para $\theta = (\theta_1, \theta_2, ..., \theta_k)$ de $f\left(x|\,\theta\right)$, resolvendo-se as equações:

$$\begin{cases} l_1 &= \lambda_1 \\ l_2 &= \lambda_2 \\ & \vdots \\ l_k &= \lambda_k \end{cases}$$

Do trabalho de Hosking [2], decorre que uma medida de assimetria, associada com o método dos momentos-L, é $\tau_3 = \dfrac{\lambda_3}{\lambda_2}$.

2 Desenvolvimento

Primeiramente, foram realizados os cálculos teóricos para a obtenção dos estimadores de máxima verossimilhança e de momentos-L dos parâmetros da distribuição GEV. Para a comparação da qualidade das estimativas dos parâmetros da distribuição GEV obtidas pelo método da máxima verossimilhança e pelo método dos momentos-L foram simulados, no *software* computacional R versão 3.6.1 [3] , amostras aleatórias de uma população que possui distribuição GEV. Os dados simulados pertenceram a cenários com parâmetros forma iguais a $\{-0,75; -0,52; -0,46\}$, parâmetros escala iguais a $\{2, 19, 37, 55\}$, parâmetros posição iguais a $\{0, 9; 5; 9; 13\}$ e tamanhos amostrais $\{13, 20, 25, 35, 45, 60, 100, 150, 200\}$. Os dois métodos de estimação foram utilizados na estimação dos parâmetros da distribuição GEV nos dados simulados.

Avaliou-se a precisão e acurácia dos estimadores dos parâmetros da distribuição GEV por meio do erro quadrático médio (EQM) e viés médio relativo (VMR), expressos por

$$EQM = \frac{\sum_{i=1}^{N} \left(\hat{\theta}_i - \theta\right)^2}{N}$$

e

$$VMR = \frac{\sum_{i=1}^{N} \frac{|\hat{\theta}_i - \theta|}{\theta}\,(100)}{N}$$

em que $\hat{\theta}_i$ é o valor obtido para o parâmetro do modelo na i-ésima amostra, θ é o valor real da amostra e N é o total de amostras simuladas.

3 Considerações Finais

Os estimadores de máxima verossimilhança dos parâmetros da distribuição GEV podem ser expressos por

$$\begin{cases} \sum\limits_{i=0}^{n}\left[\dfrac{1}{\hat{\xi}^2}\cdot\ln\left(\hat{w}_i\right)-\dfrac{\hat{w}_i-1}{\hat{w}_i}\right]\cdot\left[1-\left(\hat{w}_i\right)^{\frac{-1}{\hat{\xi}}}\right]-\dfrac{\hat{w}_i-1}{\hat{\xi}\hat{w}_i} & = & 0 \\ \dfrac{1}{\hat{\sigma}}\sum\limits_{i=1}^{n}\dfrac{\left(1+\hat{\xi}\right)-\left[\hat{w}_i\right]^{\frac{-1}{\hat{\xi}}}}{\hat{w}_i} & = & 0 \\ \dfrac{-n}{\hat{\sigma}}+\frac{1}{\hat{\sigma}^2}\sum\limits_{i=1}^{n}\left\{\dfrac{\left(\frac{\hat{\sigma}}{\hat{\xi}}\hat{w}_i-1\right)}{\hat{w}_i}\cdot\left[\left(1+\hat{\xi}\right)-\left[\hat{w}_i\right]^{\frac{-1}{\hat{\xi}}}\right]\right\} & = & 0 \end{cases}$$

em que $\hat{w}_i = 1 + \hat{\xi}\dfrac{(x_i - \hat{\mu})}{\hat{\sigma}}$

Os estimadores de momentos-L dos parâmetros da distribuição GEV podem ser expressos por

$$\begin{cases} l_1 & = \mu + \dfrac{\sigma}{\xi}\left(\Gamma\left(1-\xi\right)-1\right), \\ l_2 & = \dfrac{\sigma}{\xi}\left(2^{\xi}-1\right)\Gamma(1-\xi) \\ t_3 & = \dfrac{2\left(1-3^{\xi}\right)}{1-2^{\xi}}-3 \end{cases}$$

em que $\Gamma(x) = \int\limits_0^{\infty} x^{\alpha-1}e^{-x}dx$ é a função Gama.

Nas figuras 1 e 2 são apresentados os valores dos erros quadráticos médios e vieses médios relativos dos estimadores de máxima verossimilhança e de momentos-L do parâmetro posição avaliados no cenário em que o parâmetro posição é igual a 19, o de escala é 5 e o de forma, -0,46, para todos os tamanhos amostrais avaliados.

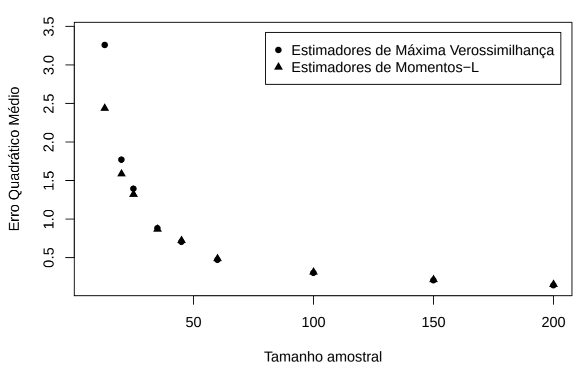

Figura 1: Erro quadrático médio dos estimadores de máxima verossimilhança e de momentos-L do parâmetro posição da distribuição Generalizada de Valores Extremos obtidos no cenário simulado em que $\mu = 19,\ \sigma = 5,\ \xi = -0,46$ e tamanhos amostrais iguais a $13,\ 20,\ 25,\ 35,\ 45,\ 60,\ 100,\ 150$ e 200.

A partir da Figura 1 é possível observar que o estimador de momentos-L do parâmetro posição possui menor EQM do que o estimador de máxima verossimilhança quando o tamanho amostral é menor que 60. A partir desse tamanho amostral, os dois estimadores possuem valores equivalentes de EQM. O EQM dos dois estimadores diminuem à medida que o tamanho amostral aumenta.

Na Figura 2 observa-se que o estimador de momentos-L do parâmetro posição possui menor VMR do que o estimador de máxima verossimilhança quando o tamanho amostral é menor que 60 e a partir do tamanho amostral 60, os dois estimadores possuem valores equivalentes de VMR. A medida que o tamanho amostral aumenta o VMR dos dois estimadores diminuem, aproximando-se de zero.

O comportamento apresentado para o EQM e o VMR nos demais cenários foram semelhantes e, dessa forma, os gráficos nos quais eles são apresentados não serão colocados neste trabalho.

Assim conclui-se que o método dos momentos-L apresenta maior precisão e acurácia que o método da máxima verossimilhança na estimação dos parâmetros da distribuição GEV para tamanhos amostrais

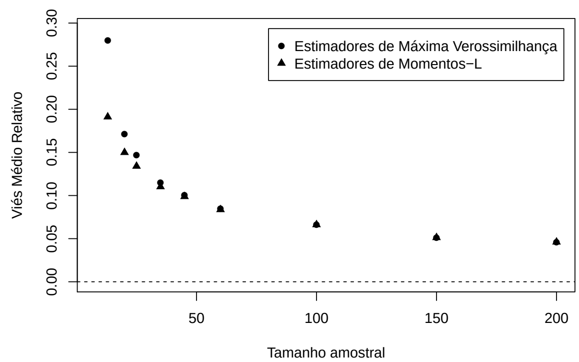

Figura 2: Viés médio relativo dos estimadores de máxima verossimilhança e de momentos-L do parâmetro posição da distribuição GEV obtidos no cenário simulado em que $\mu = 19,\ \sigma = 5,\ \xi = -0,46$ e tamanhos amostrais iguais a 13, 20, 25, 35, 45, 60, 100, 150 e 200.

menores ou iguais que 60. Também observa-se que os dois métodos possuem precisão e acurácia semelhantes na estimação dos parâmetros da distribuição GEV para tamanhos amostrais maiores que 60. Por fim, conclui-se que os dois métodos são assintoticamente não enviesados na estimação dos parâmetros da distribuição GEV.

Referências

[1] BRITO, A. L.; VEIGA, J. A. P. Um estudo observacional sobre a frequência, intensidade e climatologia de eventos extremos de chuva na Amazônia. *Ciência e Natura*, v. 37, 163-169, 2015.

[2] HOSKING, J. R. M.; WALLIS, J. R. *Regional frequency analysis:* an approach based on L-moments. Cambridge: Cambridge University Press, 1997.

[3] R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing, 2019. Disponível em: <http://www.R-project.org>. Acesso em: 01 abr. 2019.

[4] SMITH, R. L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, v. 72, n. 1, p. 67-90, 1985.

[5] VALVERDE, A. E. L.; et al. Momentos-L: teoria e aplicação em hidrologia. *Revista Árvore*, v. 28, n. 6, p. 927-933, 2004.

Descrição do video:

Primeiramente, o vídeo apresenta um pouco da teoria de eventos extremos, com foco na distribuição generalizada de valores extremos e nos estimadores de máxima verossimilhança e de momentos-L dos seus parâmetros. Após essa introdução, foram apresentados os aspectos envolvidos na simulação de dados e na avaliação dos dois estimadores na estimação dos parâmetros da distribuição generalizada de valores extremos. Por fim, foram apresentados os resultados obtidos na avaliação da acurácia e da precisão dos dois estimadores dentro dos cenários estudados.