https://eventos.utfpr.edu.br//sicite/sicite2019

Estimação dos parâmetros da distribuição alfa estável

Estimating the parameters of the alpha stable distribution

Bruna Alves da Silva
bsilva.2016@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Roberto Molina de Souza
rmolinasouza@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

RESUMO

Este trabalho tem como objetivo aplicar métodos bayesianos de estimação de parâmetros e de modelos para resolver problemas utilizando a classe das distribuições alfa estáveis. Essa distribuição é relevante por apresentar mais alternativas para análise de dados quando comparada a distribuição normal, amplamente utilizada na estatística. No presente trabalho, estudou-se as formas de estimação dos parâmetros da distribuição alfa estável, por meio de métodos de simulação. Na literatura, os métodos utilizados são os de Monte Carlo via Cadeia de Markov (MCMC), que permitem obter amostras muito próximas de uma distribuição alvo qualquer em amostras multidimensionais, especialmente quando o número de dimensões é alto. Para a simulação e comparação numérica utilizou-se o software JAGS a partir da plataforma R.

PALAVRAS-CHAVE: JAGS. Alfa Estável. MCMC.

Recebido: 16 ago. 2019.
Aprovado: 16 set. 2019.

ABSTRACT

This work aims to apply Bayesian methods of parameter estimation and models to solve problems using the class of alpha stable distributions. This distribution is relevant because it presents more alternatives for data analysis when compared to the normal distribution, widely used in Statistics. In the present work, the estimation methods of the parameters of the alpha stable distribution were studied, using simulation methods and numerical studies. In the literature, the methods used are Markov Chain Monte Carlo (MCMC), which allow to obtain samples very close to any target distribution in multidimensional samples, especially when the number of dimensions is high. For the simulation and numerical comparison he used the JAGS software from the R platform.

KEYWORDS: JAGS. Alpha Stable. MCMC.

INTRODUÇÃO

As distribuições estáveis são importantes generalizações de diversas distribuições de probabilidade, como a distribuição normal e Cauchy. Elas permitem, a partir de seus parâmetros, que se modele a assimetria e as caudas de uma variável aleatória. Embora bastante flexível, as distribuições estáveis tem pouca aplicação na literatura pois não apresentam uma função densidade fechada, o que facilitaria a estimação de seus parâmetros pelos métodos convencionais, como estimadores de máxima verossimilhança, por exemplo.

Este trabalho tem como objetivo aplicar métodos Bayesianos de estimação de parâmetros e de modelos para resolver problemas utilizando a classe das distribuições alfa estáveis. Objetiva-se, estudar as formas de estimação dos parâmetros da distribuição alfa estável, por meio de métodos de simulação e estudos numéricos para, na sequência, ser aplicada na resolução de alguns problemas reais, como Mandelbrot (1963) Fama (1965) que tentaram utilizar distribuições estáveis, ambos para modelar mudanças no preço do mercado de ações. Pretende-se entender os métodos de estimação proposto pela literatura para estimar os parâmetros da distribuição alfa estável.

O problema é relevante por se tratar do estudo de uma distribuição de probabilidade pouco explorada na literatura devido a dificuldade da estimação de seus parâmetros. Logo, explorar métodos que possam facilitar esta estimação é de interesse científico.

REFERENCIAL TEÓRICO

DuMouchel (1973) e DuMouchel (1975) forneceu alguns resultados relativos a distribuições estáveis na inferência estatística. A análise estatística geral baseada em distribuições estáveis foi dificultada pela inexistência de uma forma simples de representação de densidade de probabilidade, implicando que a função de similaridade das distribuições estáveis não pode ser representada em uma "forma analítica utilizável".

Buckle (1995) apresenta uma representação para a função densidade de probabilidade por meio da derivação de um algoritmo que permite fazer a inferência Bayesiana paramétrica e preditiva para distribuições estáveis. Para tanto, utiliza o amostrador de Gibbs e de Metropolis Hastings.

Os resultados apresentados por Buckle (1995) permitem não apenas estimar os valores dos parâmetros, mas também fazer inferências mais gerais. Ele mostra como os métodos de simulação de Monte Carlo via Cadeia de Markov (MCMC) podem ser usados para gerar amostras Bayesianas a partir dos parâmetros de uma distribuição estável.

Segundo Buckle (1995), o algoritmo apresentado necessita de melhorias, como o aumento na velocidade dos esquemas de amostragem univariados, ajustes na reparametrização, redução na dependência inter-cadeias por meio de alguma outra parametrização e talvez a amostragem de mais de um parâmetro por vez seja mais interessante. O sucesso de um trabalho futuro está ligado ao avanço da teoria MCMC.

Lombardi (2007) introduz uma nova abordagem para a construção da densidade a posteriori da distribuição alfa estável que evitam o uso do vetor auxiliar. Para colocá-lo em bases mais formais, procura-se uma expressão computacionalmente viável para a função de verossimilhança, a fim de ser capaz de produzir amostras a partir da distribuição a posteriori dos parâmetros de acordo com o teorema de Bayes.

Em seu trabalho, Lombardi (2007) estudou a representação das distribuições estáveis em séries e não em integrais. Ele ainda relata que a abordagem utilizada é duas vezes mais rápida do que o amostrador de Gibbs proposto por Buckle (1995). Já no contexto das aplicações, Achcar e Oliveira (2015) apresentam o uso da distribuição alfa estável no contexto de modelos de regressão não linear, mais especificamente o modelo de crescimento de Gompertz.

Achcar, Achcar e Martinez (2013) utilizam dois exemplos para mostrar a utilidade dos aspectos computacionais: um está relacionado a um modelo de regressão linear padrão com uma variável explicativa e o outro está relacionado a um conjunto de dados simulados, assumindo um experimento 2^3. Em ambos os exemplos, o uso de técnicas de aumento de dados é a chave para obter um bom desempenho para o método de simulação MCMC para aplicações que usam distribuições estáveis. Esses resultados podem ser de grande interesse para pesquisadores e profissionais, quando lidam com dados não-gaussianos, como nas aplicações apresentadas por eles.

Casarin (2004) propõe um modelo de mistura de distribuições estável que possibilita a captura da multimodalidade dessas distribuições e, além disso, fornece ferramentas inferenciais para a distribuição estável. Casarin (2004) verificou a eficiência do amostrador de Gibbis simulando a mistura estável de dois componentes em dados sintéticos, uma abordagem geral e que funcionou bem, porém, de acordo com o autor, é necessária uma análise mais profunda em dados sintéticos e dados reais.

Logo, o estudo das distribuições alfa estáveis vem sendo realizado por diversos pesquisadores e, basicamente, avançando na medida que os métodos computacionais para a estimação de parâmetros, principalmente bayesianos, evoluem também.

METODOLOGIA

Uma distribuição alfa estável apresenta quatro parâmetros que podem ser denotados α, β, δ e γ (Oral; Erdemir 2012). Em que $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $0 \leq \delta$ e $-\infty < \gamma < \infty$. O parâmetro α é chamado de expoente característico ou índice de cauda (ele é responsável pela espessura das caudas, em que quanto maior o α, menos espessas serão as caudas). Quando $\alpha<2$ momentos absolutos de ordem igual e maior que α não existem enquanto os de ordem inferior a α existem. δ e γ são os parâmetros de locação e escala respectivamente. β é o parâmetro de assimetria ou curtose. Quando β é positivo, a distribuição é inclinada para a direita. Quando β é negativo, a distribuição é inclinada para a esquerda se β é zero, a distribuição é simétrica em relação ao parâmetro de locação δ.

A distribuição normal é um caso particular da distribuição alfa estável, quando $\alpha=2$ e a distribuição de Cauchy é outro caso particular da distribuição alfa estável quando α=1 e β=0. Existe ainda, a distribuição de Levy, que particulariza a distribuição alfa estável para α=0,5 e β=1.

A função utilizada para a estimação dos dados será a de densidade de probabilidade bivariada de z e y, condicional em α e β, $f\ :(-\infty,0)\times(-\frac{1}{2},l_{\alpha,\beta})\cup(0,\infty)\times(l_{\alpha,\beta},\frac{1}{2})\rightarrow(0,\infty)$ é dada por:

$$f(z,y|\alpha,\beta)=\frac{\alpha}{|\alpha-1|}exp\left\{-\left|\frac{z}{t_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\right\}\left|\frac{z}{t_{\alpha,\beta}(y)}\right|^{\alpha/(\alpha-1)}\frac{1}{|z|}, \quad (1)$$

em que

$$t_{\alpha,\beta}(y)=\left(\frac{sen[\pi\alpha y+\eta_{\alpha,\beta}]}{cos\pi y}\right)\left(\frac{cos\pi y}{cos[\pi(\alpha-1)y+\eta_{\alpha,\beta}]}\right)^{(\alpha-1)/\alpha} \quad (2)$$

e $\alpha\in(0,1)\cup(1,2],\beta\in[-1,1],z\in(-\infty,\infty)$, e $y\in(-1/2,1/2)$ com $\eta_{\alpha,\beta}=\beta\ \min\ (\alpha,2-\alpha)\pi/2$ e $l_{\alpha,\beta}=-\eta_{\alpha,\beta}/\pi\alpha$. Então $f(z,y|\alpha,\beta)$ é uma densidade de probabilidade bivariada adequada para a distribuição de (z,y), e a distribuição marginal de z é $S_\alpha(\beta,0,1)$.

RESULTADOS DAS SIMULAÇÕES

Os resultados obtidos com as simulações, para uma amostra aleatória de tamanho n = 30 e delta e beta nulos, estão apresentados nos quadros 1, 2, 3, 4, 5 e 6, em que os quadros 1, 3 e 5 representam os valores anteriores a estimação e os quadros 2, 4 e 6 os valores estimados.

As demais Simulações, para beta variando entre -1 e 1 e amostras de tamanho 50 e 100, estão sendo trabalhadas.

Quadro 1 – valores iniciais

Alfa	Beta	Gama	Delta
1,8	0,0	1,0	0,0
1,6	0,0	1,0	0,0
1,4	0,0	1,0	0,0
1,2	0,0	1,0	0,0

Fonte: autor

Quadro 2 – valores estimados

Alfa	Beta	Gama	Delta
1,89	-0,04	0,89	0,30
1,83	-0,10	0,947	0,32
1,67	-0,02	0,95	0,30
1,49	0,01	0,939	0,29

Fonte: autor

Quadro 3 – valores iniciais

Alfa	Beta	Gama	Delta
1,8	0,0	1,5	0,0
1,6	0,0	1,5	0,0
1,4	0,0	1,5	0,0
1,2	0,0	1,5	0,0

Fonte: autor

Quadro 4 – valores estimados

Alfa	Beta	Gama	Delta
1,88	-0,05	1,34	0,44
1,82	-0,06	1,40	0,46
1,66	-0,02	1,37	0,45
1,49	0,01	1,40	0,46

Fonte: autor

Quadro 5 – valores iniciais

Alfa	Beta	Gama	Delta
1,8	0,0	0,5	0,0
1,6	0,0	0,5	0,0
1,4	0,0	0,5	0,0
1,2	0,0	0,5	0,0

Fonte: autor

Quadro 6 – valores estimados

Alfa	Beta	Gama	Delta
1,89	-0,02	0,45	0,15
1,83	-0,08	0,48	0,15
1,68	-0,02	0,47	0,15
1,50	-0,02	0,49	0,15

Fonte: autor

CONCLUSÃO

Pode-se concluir que a distribuição alfa estável é de grande importância para o campo científico, pois é uma generalização de importantes distribuições, como a normal e a de Cauchy.

Na literatura, os métodos utilizados são os de Monte Carlo via Cadeia de Markov, que permitem obter amostras muito próximas de uma distribuição alvo qualquer em amostras multidimensionais, especialmente quando o número de dimensões é alto.

A distribuição alfa estável é pouco explorada na literatura e, nos estudos já realizados sobre ela, de acordo com autores, necessita de melhorias, tanto no algoritmo, como no avanço da teoria.

Para trabalhos futuros, pretende-se concluir as simulações para amostras maiores, de tamanho 50 e 100, além disso variar os valores de beta. Por fim, pretende-se aplicar a distribuição alfa estável em um problema.

REFERÊNCIAS

ACHCAR, Jorge A.; ACHCAR, Angela; MARTINEZ, Edson Zangiacomi. Robust linear regression models: Use of a stable distribution for the response data. Open Journal of Statistics, v. 3, p. 409–416, 2013.

ACHCAR, Jorge Alberto; OLIVEIRA, Breno Raphael Gomes de. Gompertz growth curves assuming stable distributions: an application to intrauterine growth for preterm infants. Revista Bras. Biom., v. 33, n. 2, p. 170–183, 2015.

BUCKLE, D. J. Bayesian inference for stable distributions. Journal of the American Statistical Association, v. 90, n. 430, p. 605–613, 1995.

CASARIN, Roberto. Bayesian inference for mixtures of stable distributions. 2004.

DUMOUCHEL, W. H. Stable distributions in statistical inference: 1. symmetric stable distributions compared to other long-tailed distribution. Journal of the American Statistical Association, v. 68, p. 469–477, 1973.

. Stable distributions in statistical inference: 2. information from stably distributed samples. Journal of the American Statistical Association, v. 70, p. 368–393, 1975.

FAMA, E. F. The behaviour of stock market prices. Journal of Business, v. 38, p. 35–105, 1965.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Biometrika, v. 57, n. 1, p. 97–109, 1970.

LOMBARDI, Marco J. Bayesian inference for alpha stable distributions: A random walk mcmc approach. Computational Statistics & Data Analysis, v. 51, n. 5, p. 2688–2700, 2007.

MANDELBROT, B. The variation of certain speculative prices. Journal of Business, v. 36, p. 394–419, 1963.

METROPOLIS, Nicholas et al. Equation of state calculations by fast computing machines. The Journal of Chemical Physics, v. 21, n. 6, p. 1087–1092, 1953.

ORAL, Ece; ERDEMIR, Cenap. A bayesian estimation of stable distributions. Journal of Statistical and Econometric Methods, v. 1, n. 3, p. 39–52, 2012.

AGRADECIMENTOS

Agradecemos a Fundação Araucária-AF pelo apoio financeiro para realização deste trabalho.