Estimação da média evolucionária de um processo estocástico via reamostragem de séries temporais

Aline Edlaine de Medeiros[1], Eniuce Menezes de Souza[2]

[1] Programa de Pós-graduação em Bioestatística - UEM

[2] Departamento de Estatística - UEM

RESUMO

Neste trabalho pretende-se estimar a incerteza da média evolucionária μ_t de um dado processo estocástico, via reamostragem de uma série temporal, a qual representa uma realização deste processo estocástico. A reamostragem será baseada na análise *wavelet*, em que é possível decompor a série temporal de modo que em cada nível de decomposição os coeficientes *wavelet* possam ser considerandos independentes. Para exemplificar esta metodologia, são apresentados os intervalos de confiança e a estimativa da média μ_t para uma série temporal simulada Y_t. A média estimada se mostrou muito próxima de Y_t, e os intervalos de confiança contém parte substancial dos valores observados em Y_t.

Palavras chave: Séries temporais; *Wavelets*; *Bootstrap*.

1 INTRODUÇÃO

Dado um conjunto arbitrário T, um processo estocástico é uma família $\{X(t), t \in T\}$, tal que, para cada $t \in T$, $X(t)$ é uma variável aleatória. Reunindo uma observação de cada uma das variáveis aleatórias defini-se uma trajetória do processo estocástico ou uma série temporal [5]. Embora existam fenômenos onde mais de uma trajetória do processo estocástico esteja acessível, muitos são os casos em que apenas uma trajetória é conhecida, por exemplo, o número de internações registradas por uma dada doença, o número de acidentes de transito registrados em determinada cidade, entre outros. Em muitas situações considera-se que todas as variáveis aleatórias seguem distribuição normal, e cada observação da série temporal representa a média desta distribuição. Entretanto, esta condição nem sempre é verdadeira. Neste trabalho, propõe-se estimar a incerteza associada à média μ_t sem assumir o pressuposto de normalidade, utilizando técnicas apropriadas de reamostragem.

Para estimar a incerteza da média evolucionária μ_t será utilizada a reamostragem bootstrap baseada em *wavelets*, tal metodologia se faz necessária porque leva em consideração a dependência entre as observações provenientes de uma série temporal. Na literatura, o *bootstrap* baseado em *wavelets* consiste em decompor a série temporal utilizando, por exemplo, a transformada *wavelet* decimada (DWT) para obter os coeficientes *wavelet* considerados independentes em cada nível de decomposição. Então, utiliza-se um método de reamostragem *bootstrap* gerando réplicas dos coeficientes *wavelet*, os quais aplica-se a inversa da DWT (IDWT), para gerar uma réplica da série temporal [2, 8].

A fim de estimar um intervalo de confiança para a média μ_t será empregado o *bootstrap* baseado em *wavelets*, porém, utilizando a transformada wavelet não decimada (NDWT). A NDWT é promissora porque é invariante à translação, além disso, cada nível de decomposição da NDWT possui mais coeficientes que os respectivos níveis da DWT, condição importante para realizar a reamostragem [6]. Em particular, pretende-se estimar um intervalo de confiança para uma série temporal simulada contendo 512 observações, e avaliar a média obtida do *ensemble* de séries temporais simuladas via *bootstrap*.

2 METODOLOGIA

O *bootstrap* baseado na NDWT consiste em desenvolver os seguintes passos:

- Decomposição da série temporal via NDWT;
- Reamostragem dos coeficientes *wavelets*;
- Aplicação da inversa da NDWT (INDWT) aos coeficientes obtidos da reamostragem.

A decomposição da série temporal $Y_t = (y_0, y_1, \cdots, y_{n-1})$ usando a NDWT, consiste em representar Y_t como uma função f em termos da função escala ϕ e da função *wavelet* ψ como segue:

$$f(t) = \sum_{k=0}^{n-1} c_{J_0,k}\phi_{J_0,k}(t) + \sum_{j=J_0}^{J-1}\sum_{k=0}^{n-1} d_{j,k}\psi_{j,k} \tag{1}$$

em que $J - 1 < \log_2 n \leq J$, $j = J_0, \cdots, J-1$ representa os níveis de resolução, e $k = 0, \cdots, n-1$. Os valores $c_{J_0,k}$ e $d_{j,k}$ são chamados de coeficientes suaves (escala) e detalhe (wavelet), respectivamente. Tomando $\phi_{J_0,k}(t) = 2^{J_0/2}\phi(2^{J_0}(t-k))$ e $\psi_{j,k}(t) = 2^{j/2}\psi(2^j(t-k))$, os coeficientes suaves e de detalhes obtidos compreendem a NDWT da série temporal Y_t [4]. A reconstrução da série temporal utilizando a INDWT segue um procedimento semelhante ao descrito para a decomposição conforme apresentado em Nason (1994).

O procedimento de reamostragem seguirá a ideia clássica do *bootstrap* proposta por Efron (1994), de modo que, em cada nível de decomposição j será realizada um amostra com reposição dos coeficientes $d_{j,k}$. Para ilustrar a metodologia, será simulada a série temporal Y_t, que apresenta um componente sazonal, com forte dependência entre as observações, a qual foi gerada da seguinte forma:

$$Y_t = 1.5W_t + \cos\left(\frac{\pi(t-1)}{180}\right). \tag{2}$$

em que W_t foi gerada a partir de um modelo ARFIMA$(1, d, 0)$ com $\phi_1 = 0.95$ e $d = 2/5$.

Para que as estimativas possam ser realizadas, é necessário um número suficientemente grande de simulações *bootstrap*; neste caso serão geradas $B = 5000$ réplicas de Y_t. A estimativa para μ_t será estimada pontualmente, seguinte o algoritmo usual em *bootstrap*:

$$\overline{\hat{\mu}}_i = \frac{1}{B}\sum_{b=1}^{B} y_{ib}, i \in \{0, 1, \cdots, t\} \tag{3}$$

em que y_{ib} representa valor obtido pela reamostragem para a posição i da série temporal.

A implementação da série temporal Y_t e suas réplicas serão desenvolvidas na linguagem estatística R [7]. Além disso, a função ψ empregada foi a *wavelet* de Daubechies $d4$ [1].

3 RESULTADOS E DISCUSSÕES

Por meio da linguagem estatística R e dos algoritmos descritos na sessão 2 obteve-se 5000 réplicas da série Y_t. A Figura 1 (a) representa a série temporal Y_t (preto) e a média do *ensemble* de réplicas de Y_t (vermelho).

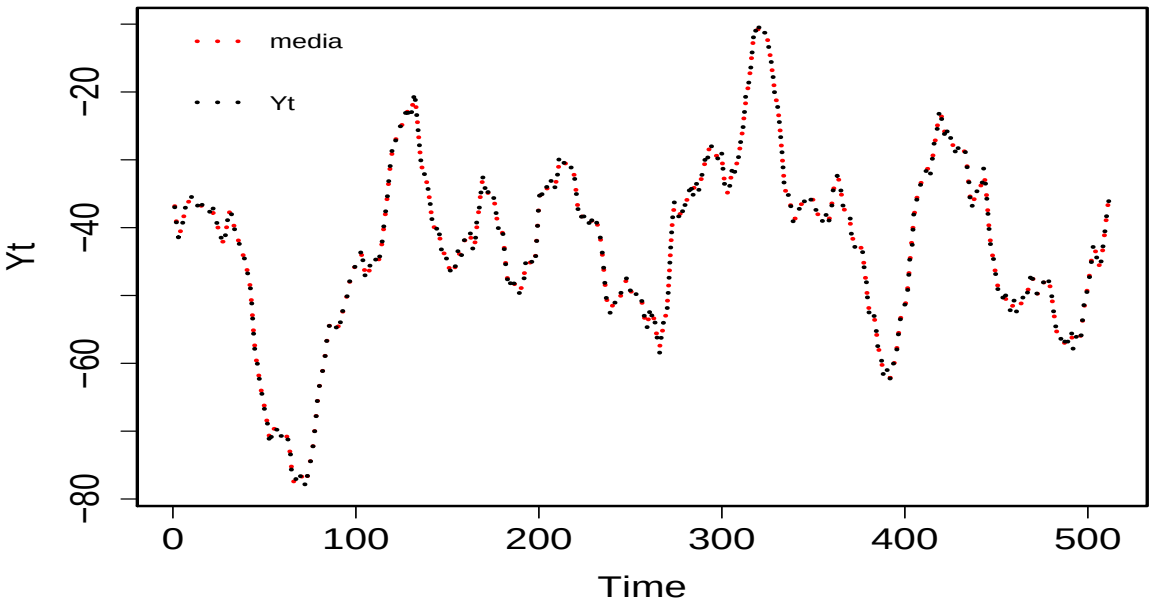

Figura 1: (a) Série temporal Y_t e a média estimada à partir do *ensemble* de réplicas de Y_t.

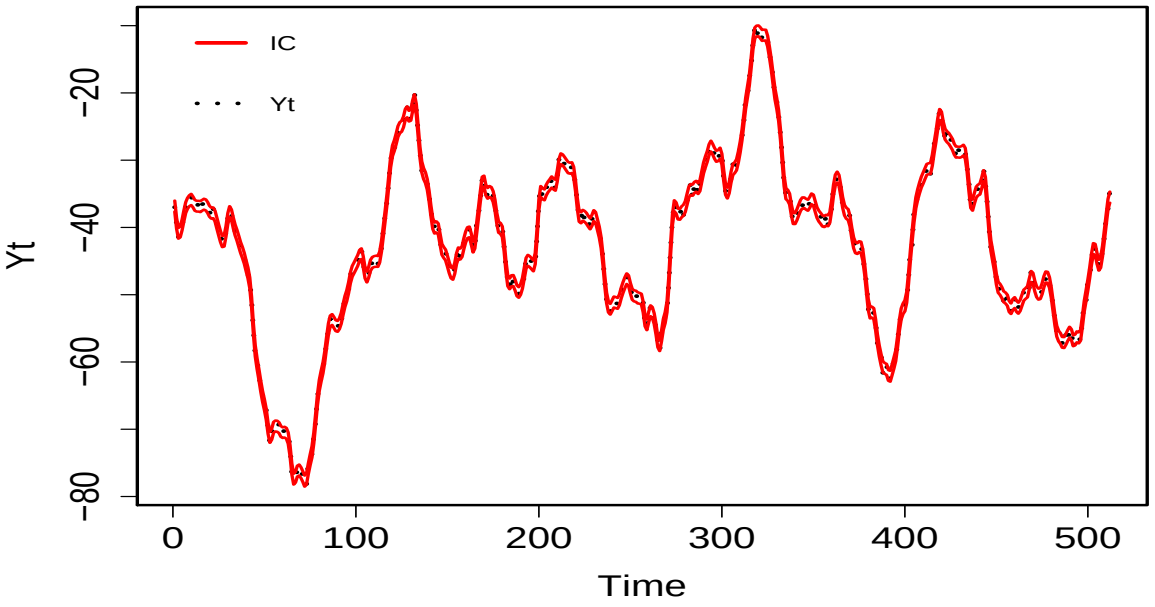

Figura 2: Série temporal Y_t e o intervalo de confiança estimado à partir do *ensemble* de réplicas de Y_t.

A Figura 2 representa a série temporal Y_t (preto) e o intervalo de confiança gerado a partir do *ensemble* de réplicas de Y_t (vermelho).

Pode-se observar que o valor estimado para μ_t e a série temporal observada são graficamente muito próximos. Além disso, o intervalo de confiança possui raras extrapolações, e contempla os diversos comportamentos ao longo da série temporal mesmo em locais de picos.

Ao determinar uma estimativa para a média μ_t via análise *wavelet*, torna-se possível estimar a incerteza dos modelos de regressão baseados em *wavelet*, uma vez que cada modelo é também um representante da média do processo estocástico.

4 CONCLUSÃO

A metodologia empregada neste trabalho se mostra uma alternativa para a reamostragem de séries temporais, assim como fornece uma estimativa para a média evolucionária de um processo estocástico sem assumir a pressuposição de normalidade. Embora, neste resumo apenas uma série temporal tenha sido utilizada para descrever a metodologia, esta análise faz parte de um projeto maior em que foram simuladas um total de 24 séries temporais que variam em tamanho e comportamento, e evidenciam os resultados descritos aqui. Para todas as séries, foram avaliados outros indicadores estatísticos, tais como erro padrão e viés, o estudo sugere que a medida em que o tamanho da amostra é elevado o viés se aproxima de zero, e o erro padrão se torna constante.

5 AGRADECIMENTOS

Este trabalho teve apoio financeiro da CAPES, mediante a concessão de bolsa de mestrado à primeira autora.

Referências

[1] Daubechies, I. *Ten Lectures on Wavelets —— 1. The What, Why, and How of Wavelets*, vol. 10.1137/1.9781611970104. 1992.

[2] Golia, S. *Evaluating the GPH estimator via bootstrap technique.* In *Compstat* (2002), Springer, pp. 343–348.

[3] Efron, B., and Tibshirani, R. J. *An introduction to the bootstrap.* CRC press, 1994.

[4] Kang, M., and Vidakovic, B. *MEDL and MEDLA: Methods for assessment of scaling by medians of log-squared nondecimated wavelet coefficients. arXiv preprint arXiv:1703.04180* (2017).

[5] Morettin, P. A., and Toloi, C. *Análise de séries temporais.* Blucher, 2006.

[6] Nason, G. *Wavelet methods in statistics with R.* Springer Science & Business Media, 2010.

[7] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2016.

[8] Yi, J. S., Jung, Y. Y., Jacko, J. A., Sainfort, F., and Vidakovic, B. *Two-step wave-trapping: Simulating non-stationary acceleration data in the mobile computing context.* Tech. rep., Georgia Institute of Technology, 2005.