

## ESTUDO DE FUNÇÕES DE VEROSSIMILHANÇA PARA MODELAGEM DE PROCESSOS DE MARKOV APLICADOS AO ESTUDO DA CONCENTRAÇÃO DE POLUIÇÃO

Mário Hissamitsu Tarumoto<sup>1</sup>

Maicon Aparecido Pinheiro<sup>2</sup>

Eliane Regina Rodrigues<sup>3</sup>

### 1. INTRODUÇÃO

Um dos problemas sérios do mundo atual é a concentração de poluentes em grandes cidades, geradas através de várias fontes, como fábricas, automóveis entre outros. A Organização Mundial da Saúde preconiza que um limite aceitável para que não haja problema de saúde pública é que a concentração de ozônio não ultrapasse 0,11 ppm (partes por milhão). Uma possibilidade de estudo é medir os tempos entre as ocorrências de ultrapassagem deste limite. Observa-se que existe correlação entre esses tempos, desta forma, as técnicas utilizadas para construir a função de verossimilhança nos casos de independência podem não ser adequadas para esta situação.

Existem algumas possibilidades para a construção de modelos desta natureza. Nesta pesquisa, foi construído o modelo através da suposição de que a sequência de tempos segue um processo de Markov, ou seja, que cada tempo observado depende somente do anterior. Para a construção deste modelo, foi necessário o estudo de modelos bivariados aplicáveis à área de análise de sobrevivência.

As funções densidades bivariadas além de serem imprescindíveis ao estudo do comportamento conjunto de duas variáveis aleatórias se mostram uma boa solução para o estudo de variáveis aleatórias interdependentes. A grande dificuldade é conseguir definir tais distribuições densidades bivariadas. No entanto, o teorema de Skalar (1959), garante que uma função de distribuição conjunta pode ser obtida por meio de uma função de cópula aplicada às funções de distribuições marginais. Uma função cópula pode ser definida como uma distribuição multivariada cujas marginais são distribuições uniformes distribuídas no intervalo (0,1). Na literatura, as funções cópulas são classificadas por famílias, sendo algumas delas, a família das cópulas arquimedianas, elípticas, gaussianas, entre outras.

A aplicação deste modelo foi realizada no conjunto de dados de poluição na cidade do México no período de 2003 a 2008.

---

<sup>1</sup> Depto de Estatística – FCT/UNESP

<sup>2</sup> Depto de Estatística/IME-USP

<sup>3</sup> Instituto de Matematicas – UNAM/MÉXICO

## 2.MATERIAIS E MÉTODOS

Em uma situação em que a condição de independência é violada, no entanto, caso a suposição de que a sequência de v.a. siga um processo de Markov, a função de verossimilhança pode ser escrita como:

$$L(\Theta; \mathbf{t}) = f_{T_1}(t_1|\Theta) \left[ \prod_{i=2}^{n-2} f(t_{i+1}|t_i, \Theta) \right] P(T_n \geq t_n | T_{n-1} = t_{n-1}, \Theta)$$

Para a especificação da função de verossimilhança apresentada é necessário uma distribuição de probabilidade a variável aleatória. Nesta forma de construção da função de verossimilhança, para a determinação da distribuição condicional, foi considerada uma distribuição de probabilidade bivariada, construída a partir de cópulas, ou seja, foram obtidas funções densidades conjuntas dos vetores aleatórios  $(T_i, T_{i+1})$ ,  $i = 2, \dots, n - 2$ , utilizando da teoria de cópulas.

Como o conjunto de tempos anteriores sempre irá diferir do conjunto de tempos posteriores por apenas uma observação, é mais do que razoável considerar que  $T_i$  e  $T_{i+1}$  seguem uma mesma distribuição de probabilidade. Ainda para a construção desta cópula, foram consideradas distribuições Weibull para as marginais.

Para a construção deste modelo, considerou-se a cópula de Farlie-Gumbel-Morgenster. Dessa forma, a função distribuição conjunta para  $T_i$  e  $T_{i+1}$  fica dada por:

$$F(t_i, t_{i+1} | \alpha, \beta, \theta) = \left\{ 1 - \exp \left[ - \left( \frac{t_i}{\beta} \right)^\alpha \right] \right\} \left\{ 1 - \exp \left[ - \left( \frac{t_{i+1}}{\beta} \right)^\alpha \right] \right\} \times \\ \times \left\{ 1 + \theta \exp \left[ - \left( \frac{t_i}{\beta} \right)^\alpha - \left( \frac{t_{i+1}}{\beta} \right)^\alpha \right] \right\}$$

A função densidade condicional pode ser escrita como:

$$f(t_{i+1} | t_i, \alpha, \beta, \theta) = \frac{\alpha}{\beta^\alpha} t_{i+1}^{\alpha-1} e^{-\left(\frac{t_{i+1}}{\beta}\right)^\alpha} \times \\ \times \left[ 1 + \theta - 2\theta e^{-\left(\frac{t_i}{\beta}\right)^\alpha} - 2\theta e^{-\left(\frac{t_{i+1}}{\beta}\right)^\alpha} + 4\theta e^{-\left(\frac{t_i}{\beta}\right)^\alpha - \left(\frac{t_{i+1}}{\beta}\right)^\alpha} \right]$$

Além disso:

$$P(T_n \geq t_n | T_{n-1} = t_{n-1}, \Theta) \\ = e^{-\left(\frac{t_n}{\beta}\right)^\alpha} \left[ 1 + \theta - 2\theta e^{-\left(\frac{t_{n-1}}{\beta}\right)^\alpha} - \theta e^{-\left(\frac{t_n}{\beta}\right)^\alpha} + 2\theta e^{-\left(\frac{t_{n-1}}{\beta}\right)^\alpha - \left(\frac{t_n}{\beta}\right)^\alpha} \right]$$

E portanto, a função de verossimilhança pode ser escrito como:

$$L(\alpha, \beta, \theta; \mathbf{t}) = \left\{ \frac{\alpha}{\beta^\alpha} t_1^{\alpha-1} e^{-\left(\frac{t_1}{\beta}\right)^\alpha} \right\} \times \left\{ e^{-\left(\frac{t_n}{\beta}\right)^\alpha} \right\} \times \\ \times \left[ 1 + \theta - 2\theta e^{-\left(\frac{t_{n-1}}{\beta}\right)^\alpha} - \theta e^{-\left(\frac{t_n}{\beta}\right)^\alpha} + 2e^{-\left(\frac{t_{n-1}}{\beta}\right)^\alpha} - \left(\frac{t_n}{\beta}\right)^\alpha \right] \times \\ \times \left\{ \prod_{i=1}^{n-2} \frac{\alpha}{\beta^\alpha} t_{i+1}^{\alpha-1} e^{-\left(\frac{t_{i+1}}{\beta}\right)^\alpha} \left[ 1 + \theta - 2\theta e^{-\left(\frac{t_i}{\beta}\right)^\alpha} - 2\theta e^{-\left(\frac{t_{i+1}}{\beta}\right)^\alpha} + 4\theta e^{-\left(\frac{t_i}{\beta}\right)^\alpha} - \left(\frac{t_{i+1}}{\beta}\right)^\alpha \right] \right\}$$

Com base nos valores dos parâmetros estimados, é possível obter a esperança condicional dos tempos, assim, de acordo com a modelagem proposta, obteve-se a seguinte forma fechada para a esperança condicional de  $T_{i+1}$  dado que  $T_i = t_i$ :

$$E(T_{i+1}|T_i = t_i) = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \left[ 1 + \theta \left(1 - 2^{-\frac{1}{\alpha}}\right) \left(1 - 2e^{-\left(\frac{t_i}{\beta}\right)^\alpha}\right) \right].$$

### 3. RESULTADOS

Nesse trabalho a v.a. de interesse é o tempo entre ultrapasses de um limiar das concentrações de ozônio na região Nordeste da Cidade do México. O limiar considerado foi de 0,11 ppm. Esse conjunto de dados consiste em 717 tempos entre ultrapasses vinculados às máximas diárias observadas no período que se estende desde o primeiro dia do ano de 2003 ao último dia do ano de 2008. Os estimadores de máxima verossimilhança de  $\alpha$ ,  $\beta$  e  $\theta$  foram obtidas através de métodos numéricos, através do *proc IML* do SAS e a função *nlpqn*.

**Tabela 1** – Estimativa de Máxima Verossimilhança para os parâmetros de interesse.

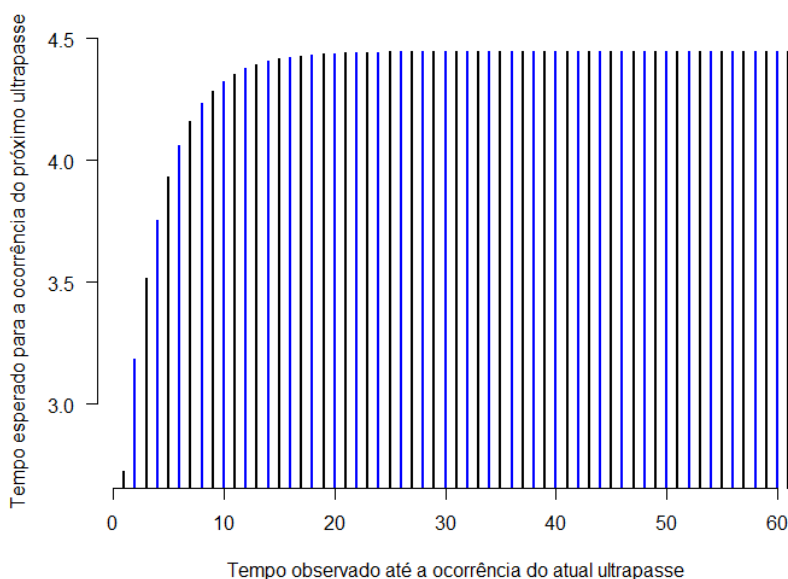
Parâmetro	$\alpha$	$\beta$	$\theta$
Estimativa de Máxima Verossimilhança	0,9553035	3,1828128	0,7143529

Desta forma, os tempos médios esperados para que ocorra o próximo ultrapasse em função da quantidade de dias entre o penúltimo e o atual ultrapasse são apresentados na Tabela 2.

**Tabela 2** – Estimativa do tempo médio necessário para a ocorrência do próximo ultrapasse em função do número de dias entre o penúltimo e o atual ultrapasse.

$T_i = t_i$	1	2	3	4	5	6	7
$E(T_{i+1}/T_i = t_i)$	2,725906	3,185280	3,515308	3,755794	3,932420	4,062837	4,159514
$T_i = t_i$	8	9	10	11	12	13	...
$E(T_{i+1}/T_i = t_i)$	4,231402	4,284995	4,325034	4,355003	4,377470	4,394339	...
$T_i = t_i$	56	57	58	59	60	61	...
$E(T_{i+1}/T_i = t_i)$	4,446045	4,446045	4,446045	4,446045	4,446045	4,446045	...

Para uma melhor compreensão do que está acontecendo, segue a Figura 6.



**Figura 1** – Tempo médio esperado para o próximo ultrapasse versus tempo observado até a ocorrência do atual ultrapasse.

A matriz de transição deste modelo é dada por:

$$\begin{matrix}
 & \begin{matrix} 1 & 2 & 3 & \dots & 60 & 61 & \dots \end{matrix} \\
 \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 61 \\ \vdots \end{matrix} & \left[ \begin{array}{ccccccc}
 0,3448124 & 0,206469 & 0,1341651 & \dots & 5,686001e-08 & 1,051958e-08 & \dots \\
 0,2893568 & 0,1936011 & 0,1373707 & \dots & 1,3694749e-08 & 1,470928e-08 & \dots \\
 0,2495158 & 0,1843565 & 0,1396736 & \dots & 1,915288e-08 & 1,771928e-08 & \dots \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
 0,1371572 & 0,158285 & 0,1461685 & \dots & 3,412535e-08 & 2,620803e-08 & \dots \\
 \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \ddots
 \end{array} \right]
 \end{matrix}$$

Na matriz acima, o elemento posicionado na intersecção entre a primeira linha e a primeira coluna indica a probabilidade de que o próximo ultrapasse ocorra amanhã dado que o tempo entre ultrapasses anterior foi igual a 1. De modo análogo, é feita a interpretação dos demais elementos.

#### 4. DISCUSSÕES E CONCLUSÕES

A dependência existente entre fenômenos aleatórios ou estados de um mesmo fenômeno aleatório consiste em uma informação extremamente importante, que se levada em consideração quando em uma modelagem estatística permite a obtenção de estimativas mais precisas e fiéis à realidade. Por meio desta abordagem, foi possível

obter importantes resultados como o tempo médio para a ocorrência do próximo ultrapasse em função do tempo entre ultrapasses anterior, bem como a probabilidade de que o próximo ultrapasse ocorra daqui à  $k$  dias dado o valor do tempo entre ultrapasses anterior. Para o exemplo em estudo, em que se considerou a região Nordeste da Cidade do México no período que se estende desde 01 de janeiro de 2003 a 31 de dezembro de 2008, as estimativas obtidas se mostraram boas quando comparadas com o observado empiricamente.

## 5. BIBLIOGRAFIA

ACHCAR, J. A.; RODRIGUES, E. R.; TZINTZUN, G. Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City. **EnvironMetrics**, London, v. 22, p. 1-12, 2011.

ACHCAR, J. A.; SANTOS, C. A. A Bayesian analysis for bivariate Weibull distributions derived from copula functions in the presence of covariates and censored data. **Advances and Applications in Statistics**, v. 15, p. 01-25, 2010.

ACHCAR, J. A.; RODRIGUES, E. R.; Ortis-Rodrigues. Estimating the number of ozone peaks in Mexico city using a non-homogeneous Poisson model and a Metropolis-Hastings algorithm. **International Journal of Pure and Applied Mathematics**, v. 53, p. 1-20, 2009.

BALAKRISHNAN, N.; LAI, C-D. **Continuous bivariate distributions**, 2. ed. New York: Springer, 2009.

EMBRECHTS, P.; LINSKOG, F.; MCNIEL, A. **Modelling dependence with copulas and applications to risk management**. Department of Mathematik, EHT Zentrum, Zürich, 2001.

SKLAR, A. Random variables, distribution functions, and copulas - a personal look backward and forward. In: DISTRIBUTIONS WITH FIXED MARGINALS AND RELATED TOPICS. Hayward: Institute of Mathematical, 1996, p. 1-14.