

Estudo do intervalo de tempo entre ultrapasses de um limiar de temperatura com o uso de uma distribuição Weibull bivariada

Maria Magdalena Kcala Alvaro ¹, Mário Hissamitsu Tarumoto ², Eliane Regina Rodrigues ³

1 Introdução

Segundo o sistema de classificação climática (SCC) de Köppen e Geiger (1928), o clima de Presidente Prudente é caracterizado como tropical chuvoso, com verões chuvosos e quentes, invernos secos e mais frios, sendo a temperatura mínima média do mês mais quente em torno dos 22°C (Rolim et al., 2007).

Atualmente, as mudanças de temperatura são um dos principais problemas enfrentados por muitos dos grandes centros rurais e urbanos, pois afeta diretamente a saúde das pessoas. Doenças como paludismo (malária), dengue, desnutrição, são sensíveis a mudanças de temperatura. Segundo a Organização Mundial da Saúde, temperaturas extremas do ar contribuem diretamente nas disfunções por doenças cardiovasculares e respiratórias, sobretudo entre as pessoas de idade avançada e crianças. Altas temperaturas provocam também aumento dos níveis de ozônio e de outros contaminantes do ar, que agravam as doenças cardiovasculares e respiratórias. Além disso, temperatura do ar pode afetar a produção agrícola, já que exerce uma influência significativa no processo de germinação das sementes. É considerada como temperatura ótima, aquela na qual a mais alta porcentagem de germinação é obtida em menor espaço de tempo, enquanto temperaturas extremas são pontos nos quais as sementes apresentam menor porcentagem de germinação, Mayer e Poljakoff (1989). Por exemplo, a faixa ótima de temperatura para germinação de espécies de regiões tropicais está entre 20°C e 35°C, segundo Larcher (2006). Desta forma para entender estes problemas gerados por fenômenos naturais, estudos devem ser realizadas constantemente e reformuladas com o intuito de prever a ocorrência de tais fenômenos naturais e proporcionar condições de vida adequada. Por exemplo, empresas privadas podem dimensionar mão de obra ou estoques de aparelhos de climatização, tendo em vista altas temperaturas em determinados períodos do ano, e com respeito à agricultura é interessante, por exemplo, prever o número de dias não favoráveis para germinação de sementes, para evitar o plantio num intervalo de tempo onde a faixa de temperatura não é ótima. Com estas considerações, a construção de modelos que possam emitir alertas ou realizar previsões sobre a ocorrência de ultrapasses de uma certa temperatura limiar aceitável, pode amenizar as consequências causadas por fenômenos naturais relacionados a mudanças de temperatura.

Para este trabalho contamos com a disposição de dados climatológicos da estação Meteorológica da FCT/UNESP do Câmpus de Presidente Prudente (SP). As temperaturas são medidas em graus centígrados (°C), cada hora durante o dia, para os quais podemos associar uma estatística de interesse que represente o comportamento da temperatura naquele dia.

Na abordagem adotada neste trabalho, o interesse é estudar o comportamento da temperatura através do estudo dos tempos entre ocorrências do evento de interesse (ul-

¹PG em Matemática Aplicada - FCT/Unesp. e-mail: *marimagda@hotmail.com*

²Depto de Estatística - FCT/Unesp. e-mail: *mario.tarumoto@unesp.br*

³Instituto de Matematicas - UNAM/México. e-mail: *eliane@matem.unam.mx*

trapasse de um limiar de temperatura fixo). Neste estudo, ao invés de utilizar os dados de temperaturas máximas ou médias diárias, utilizamos as temperaturas mínimas diárias, tendo em vista que o meteorologista suspeita que ao fazer as medições de temperatura ao longo do tempo percebeu que a variação é maior nas temperaturas mínimas do que nas máximas, ou seja, percebe-se aumento das temperaturas mínimas na cidade de Presidente Prudente comparado com anos anteriores. Desta forma, foi fixado a temperatura de 22°C como o limiar aceitável, por ser a temperatura mínima média normal do mês mais quente numa região tropical como a que está localizada a cidade de P. Prudente, ou seja, se a temperatura mínima do dia for maior que 22°C é considerado um dia com temperatura elevada. Na Figura 1 é apresentado o gráfico que pode explicar melhor estas considerações no estudo.

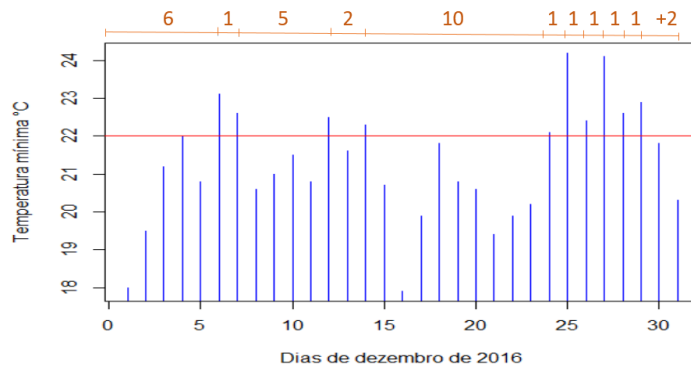


Figura 1: Tempos em dias entre ultrapasses das mínimas temperaturas diárias em relação ao limiar 22°C para a cidade de Presidente Prudente-SP no mês de dezembro de 2016

Desta forma, como os dados de interesse são tempos entre ultrapasses do evento de interesse, estamos considerando um fenômeno cujos estados são interdependentes, e os tempos também são dependentes. Neste trabalho o interesse foi o de construir o modelo através da suposição de que a sequência de tempos entre ultrapasses siga um processo de Markov, ou seja, que cada intervalo de tempo observado depende somente do anterior.

2 Modelo markoviano Weibull via distribuição Weibull bivariada de Marshall-Olkin

Vamos descrever o modelo matemático markoviano considerado neste trabalho. Sejam $S > 0$ e $K > 0$, números reais e inteiro fixados, respectivamente. K é o número de vezes em que a medição da temperatura ultrapassa o limite de interesse fixado, no caso (22°C), durante o intervalo de tempo $[0, S]$.

Sejam d_1, d_2, \dots, d_K os dias em que ocorreram os ultrapasses (excederam o limite de interesse). Seja $D = \{d_1, d_2, \dots, d_K\}$ o conjunto de dias em que ocorreram os ultrapasses durante o tempo de estudo $[0, S]$. Seja $T_i, i = 1, 2, \dots$ o tempo em dias entre i -ésimo e o $(i - 1)$ -ésimo ultrapasse.

Vamos usar a suposição de que os tempos entre ultrapasses $T_i, i = 1, 2, 3, \dots$ são identicamente distribuídos e com distribuição Weibull(α, β), e a suposição de que estes tempos seguem um processo de Markov. Para definir a função de densidade condicional de T_{i+1} dado T_i , vamos definir a função de densidade conjunta de T_i e T_{i+1} .

3 Distribuição Weibull Bivariada de Marshall-Olkin

A distribuição Weibull bivariada de Marshall e Olkin proposta por Marshall e Olkin (1967), é obtida por meio de uma transformação de variáveis da distribuição exponencial bivariada (BVE). Esta distribuição não é absolutamente contínua, ou seja, $P(T_1 = T_2) > 0$. Isto significa que a ocorrência do evento de interesse (morte ou falha) pode ser por causas simultâneas, e esta é uma característica de interesse neste estudo, considerando que os intervalos de tempos entre ultrapasses podem ser iguais entre uma sequência e outro, ou seja, é provável que o intervalo de tempo entre ultrapasse atual e o anterior seja igual ao intervalo de tempo entre ultrapasse atual e o próximo ultrapasse. Denote a distribuição Weibull bivariada como BVW.

A distribuição Weibull bivariada é obtida por meio de uma transformação de variáveis, ou seja se (T_1, T_2) é BVE, então $(T_1^{1/\alpha_1}, T_2^{1/\alpha_2})$ é uma distribuição Weibull bivariada (MOBW), onde a função de sobrevivência é dado por:

$$S(t_1, t_2) = \exp\left\{-\lambda_1 t_1^{\alpha_1} - \lambda_2 t_2^{\alpha_2} - \lambda_{12} \max(t_1^{\alpha_1}, t_2^{\alpha_2})\right\}.$$

Na situação da distribuição Weibull bivariada, $T = \min(T_1, T_2)$ tem distribuição Weibull somente se $\alpha_1 = \alpha_2$. A função de densidade de probabilidade é dada por:

$$f(t_1, t_2) = \begin{cases} (\lambda_1 + \lambda_{12})\lambda_2\alpha_1 t_1^{\alpha_1-1}\alpha_2 t_2^{\alpha_2-1} \exp\left\{-(\lambda_1 + \lambda_{12})t_1^{\alpha_1} - \lambda_2 t_2^{\alpha_2}\right\}, & \text{se } t_1^{\alpha_1} \geq t_2^{\alpha_2} \\ (\lambda_2 + \lambda_{12})\lambda_1\alpha_1 t_1^{\alpha_1-1}\alpha_2 t_2^{\alpha_2-1} \exp\left\{-\lambda_1 t_1^{\alpha_1} - (\lambda_2 + \lambda_{12})t_2^{\alpha_2}\right\}, & \text{se } t_1^{\alpha_1} < t_2^{\alpha_2}. \end{cases}$$

Da expressão anterior considerando a seguinte reparametrização: $\lambda_1 = 1/\beta_1^{\alpha_1}$, $\lambda_2 = 1/\beta_2^{\alpha_2}$ e $\lambda_{12} = \theta$, temos a função de densidade conjunta de (T_{i+1}, T_i) , dada por:

$$f(t_{i+1}, t_i) = \begin{cases} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)\alpha_2 t_{i+1}^{\alpha_2-1} \frac{1}{\beta_1^{\alpha_1}} \alpha_1 t_i^{\alpha_1-1} \exp\left\{-\frac{1}{\beta_1^{\alpha_1}} t_i^{\alpha_1} - \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) t_{i+1}^{\alpha_2}\right\}, & \text{se } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right)\alpha_1 t_i^{\alpha_1-1} \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) t_i^{\alpha_1} - \frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2}\right\}, & \text{se } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases} \quad (1)$$

Assim sendo, de (1) e a função densidade de probabilidade da distribuição Weibull univariado, temos a função de densidade condicional de T_{i+1} dado (T_i) , i.e.

$$f(t_{i+1}|t_i) = \begin{cases} \left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)\alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right) t_{i+1}^{\alpha_2}\right\}, & \text{se } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right) \frac{1}{\beta_2^{\alpha_2}} \alpha_2 t_{i+1}^{\alpha_2-1} \exp\left\{-\frac{1}{\beta_2^{\alpha_2}} t_{i+1}^{\alpha_2} - \theta t_i^{\alpha_1}\right\} \left(\frac{1}{\beta_1^{\alpha_1}}\right)^{-1}, & \text{se } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

4 Estimação de Máxima Verossimilhança

Neste trabalho, dada a característica dos dados, corre-se o risco de ter censura do Tipo I somente no último valor observado. Desta forma, a função de verossimilhança pode ser escrita como:

$$L(\theta; t) = f(t_1) \prod_{i=1}^{n-2} f(t_{i+1}|t_i) P(T_n \geq t_n | T_{n-1} = t_{n-1}). \quad (2)$$

Em que o último termo da verossimilhança, para $t_n^{\alpha_2} < t_{n-1}^{\alpha_1}$ temos uma integral conhecida que é igual à função de sobrevivência da distribuição Weibull, então:

$$P(T_n \geq t_n | T_{n-1} = t_{n-1}) = \begin{cases} \exp\left\{-\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)t_n^{\alpha_2}\right\}, & \text{se } t_n^{\alpha_2} \geq t_{n-1}^{\alpha_1} \\ \beta_1^{\alpha_1}\left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right)\exp\left\{-\theta t_{n-1}^{\alpha_1}\right\}\exp\left\{-\left(\frac{t_n}{\beta_2}\right)^{\alpha_2}\right\}, & \text{se } t_n^{\alpha_2} < t_{n-1}^{\alpha_1}. \end{cases}$$

Para encontrar os valores de $\alpha_1, \beta_1, \alpha_2, \beta_2, \theta$ que maximizam a função dada em (2) é necessário utilizar métodos iterativos. O software R contém pacotes específicos que realizam esse procedimento, desde que fornecidas os logaritmos das funções de verossimilhança. Neste trabalho utilizamos o pacote "Optim" em que se considera o negativo do logaritmo da função de verossimilhança, já que minimiza por padrão. As estimativas obtidas para estes dados são apresentadas na Tabela 1.

Tabela 1: Estimadores de máxima verossimilhança para os parâmetros de interesse do Modelo markoviano via a distribuição bivariada Weibull de Marshall e Olkin.

Parâmetro	α_1	α_2	β_1	β_2	θ
EMV	0,949	0,705	2,531	3,644	$9,999 \times 10^{-21}$
EP(EMV)	0,037	0,022	1,498	0,274	$9,657 \times 10^{-5}$

Observa-se pelos resultados da estimação dos parâmetros de interesse, que o parâmetro de dependência θ tem um valor próximo a zero, o que explica a dependência fraca que existe entre os intervalos de tempo entre ultrapasses do evento de interesse. Em seguida, na Figura 2, são apresentados os gráficos da Função distribuição acumulada estimada pelo modelo e a distribuição acumulada empírica e também o gráfico da Função de sobrevivência estimada e o gráfico de Kaplan-Meier. Observa-se que graficamente os ajustes estão razoáveis. A estimação da Função distribuição acumulada e da função de sobrevivência foram obtidas sob a suposição do modelo markoviano via distribuição bivariada Weibull de Marshall e Olkin.

Outra possibilidade é a estimação do tempo médio esperado para que ocorra o próximo ultrapasse em função do número de dias observado entre o ultrapasse anterior e o ultrapasse atual, isto é, a esperança condicional de T_{i+1} dado $T_i = t_i$. Neste caso os resultados das integrais anteriores são conhecidas e estão dadas pelo valor esperado de T_{i+1} tendo distribuição Weibull. Desta forma, a esperança condicional de T_{i+1} dado $T_i = t_i$ pode ser escrita como.

$$E[T_{i+1} | T_i = t_i] = \begin{cases} \sqrt[\alpha_2]{\left(\frac{1}{\beta_2^{\alpha_2}} + \theta\right)^{-1}} \Gamma[1 + (1/\alpha_2)], & \text{se } t_{i+1}^{\alpha_2} \geq t_i^{\alpha_1} \\ \left(\frac{1}{\beta_1^{\alpha_1}} + \theta\right)\exp\left\{-\theta t_i^{\alpha_1}\right\}\beta_1^{\alpha_1}\beta_2\Gamma[1 + (1/\alpha_2)], & \text{se } t_{i+1}^{\alpha_2} < t_i^{\alpha_1}. \end{cases}$$

conhecendo os parâmetros $\alpha_1, \alpha_2, \beta_1, \beta_2$ e θ pela estimação feita, podemos calcular essas esperanças.

5 Considerações

Este trabalho nasceu a partir de um problema que foi apresentado por um pesquisador, cujo interesse era o de estimar o tempo (em dias) entre ultrapasses de um limiar pré-determinado de temperatura, ou seja, dado que hoje fez muito calor, daqui a quantos

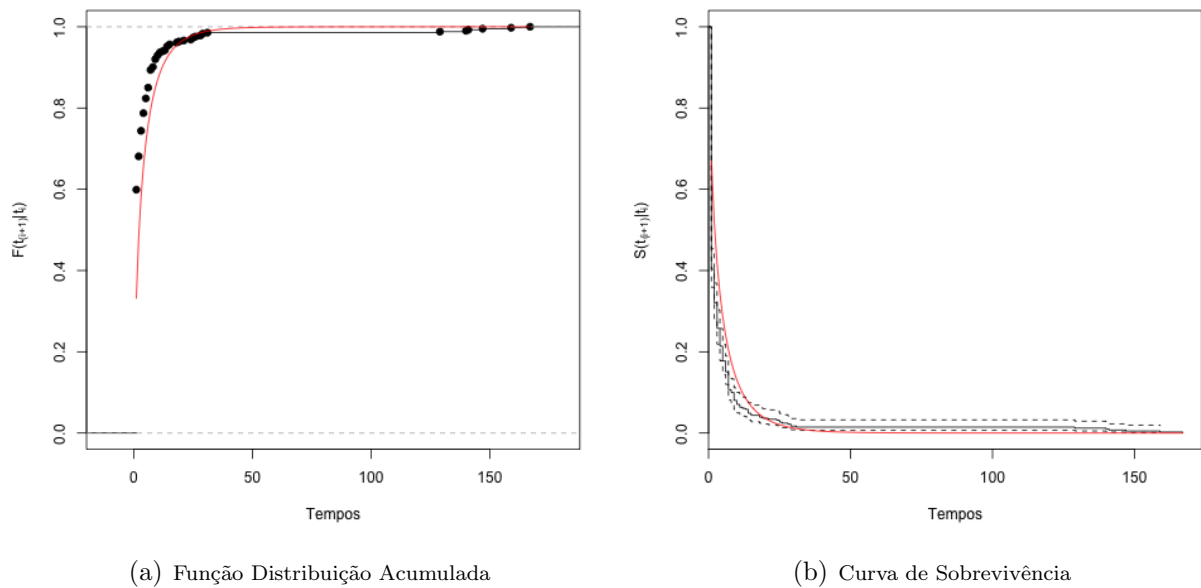


Figura 2: Função de distribuição acumulada estimada *versus* função de distribuição acumulada empírica e Curva de sobrevivência estimada *versus* curva de sobrevivência estimada por Kaplan-Meier

dias espera-se que faça calor novamente. Para este propósito, o objetivo do ponto de vista estatístico, foi sugerir a construção de uma matriz de transição entre os dias de ultrapasse. Por exemplo, dado que teve ultrapasse do limiar de temperatura a 2 dias, qual a probabilidade de ultrapassar após 3 dias. Esta construção foi realizada considerando um modelo com uma estrutura de dependência Markoviana. Para a parte condicional foi considerado o modelo Weibull bivariado de Marshall e Olkin. Pelos resultados obtidos, por meio da avaliação do ajuste do modelo, foi um modelo bastante razoável. Outra parte do estudo, a ser apresentado em outra oportunidade, é considerar a construção do modelo, incorporando modelos bivariados por meio da utilização de cópulas, o que possibilitará a inclusão de diferentes tipos de dependência, podendo abrir as possibilidades do modelo para outras situações.

Agradecimentos

A primeira autora, agradece a CAPES pelo apoio financeiro durante o desenvolvimento do trabalho, que é parte da dissertação de mestrado. O segundo autor, agradece ao Departamento de Estatística da FCT/Unesp, ao Prof. Dr. José Tadeu Garcia Tommaselli, do Departamento de Geografia da FCT/Unesp, por ceder os dados para a pesquisa.

Referencias Bibliográficas

KÖPPEN, W.; GEIGER, R. *Klimate der Erde*. Gotha: Verlag Justus Perthes. 1928. Wall-map 150cmx200cm.

LARCHER. W. Ecofisiologia vegetal. São Carlos: Rima Artes e Textos, 2006. 532p.

MAYER, A. M.; POLJAKOFF-MAYBER, A. A germination od seeds. Londres: Pergamon Press, 1989.

MARSHALL, A. W.; OLKIN, I.A multivariate exponential distribution. *Journal of the American Statistical Association*. Taylor & Francis Group, v. 62, n. 317, p. 30-49, 1967.

RODRIGUES,. E. R.; ACHCAR, J. A. Applications of Discrete-time Markov Chains and Poisson Processes to Air Pollution Modeling and Studies. *Springer*, p. 65-78.

ROLIM, G. S.; CAMARGO, M. B. P.; LANIA, D. G.; MORAES, J. F. L. Classificação climática de Köppen e de Thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de São Paulo. *Bragantia*. Campinas: Instituto Agrônômico de Campinas, p. 711-720, 2007.