

ALGUNS MODELOS DE CÓPULAS EM DADOS DE SOBREVIVÊNCIA BIVARIADOS.

Alex Fujisaki, Mário Hissamitsu Tarumoto – Estatística - Departamento de Matemática, Estatística e Computação – Faculdade de Ciências e Tecnologia – Campus de Presidente Prudente.

Os eventos ocorrem na natureza de maneira aleatória, entretanto, podemos encontrar algum padrão ou regularidade nos acontecimentos do dia a dia. A Estatística busca prever ou representar os fatos da melhor maneira possível, para tanto podemos ajustar algum modelo probabilístico para os dados obtidos.

Do ponto de vista da modelagem de dados, é comum a suposição de independência entre as variáveis ou componentes envolvidos no estudo, que em geral é matematicamente conveniente, mas muitas vezes inadequado para se representar a realidade. Na prática, em muitas aplicações, é comum a existência de algum tipo de associação entre as variáveis. Este tipo de associação pode ser completamente descrito pela função de distribuição conjunta, pois ela é capaz de representar a relação de dependência entre as variáveis.

O enfoque será dado a observações que envolvem tempos de falha, portanto definições e distribuições relacionadas à Análise de Sobrevida serão utilizadas. Em particular o estudo será feito para o caso bivariado. Diversos enfoques têm sido considerados quando se lida com dados de sobrevida correlacionados. Oakes (1989) considera que os tempos de falhas marginais são considerados condicionalmente independentes, dado que uma variável de fragilidade é conhecida (uma variável aleatória não observável que estabelece a relação de dependência entre os tempos de falha).

Alternativamente, temos uma maneira consideravelmente simples para estabelecer a estrutura de associação. Pela técnica de Cópias pode-se modelar a relação de dependência entre as variáveis por um parâmetro de associação considerando que os tempos de falha assumem distribuições de probabilidade marginais independentes, ou seja, cópias estabelecem um meio de relacionar funções de distribuição multivariadas com suas funções de distribuição marginais (NÚÑEZ, 2005).

Se assumirmos que $F_j(X_j) \sim U(0,1)$, a distribuição conjunta C_α das distribuições marginais $F_1(X_1), F_2(X_2), \dots, F_p(X_p)$ é chamada de cópia do vetor aleatório (X_1, X_2, \dots, X_p) , indexada por um parâmetro de dependência α . Dessa maneira, se considerarmos $U_j = F_j(X_j)$.

$$\begin{aligned} F(x_1, x_2, \dots, x_p) &= P[U_1 \leq F_1(X_1), U_2 \leq F_2(X_2), \dots, U_p \leq F_p(X_p)] \\ &= C_\alpha[F_1(X_1), F_2(X_2), \dots, F_p(X_p)], \end{aligned}$$

que é uma distribuição multivariada, definida no cubo $[0,1]^p$ com marginais uniformemente distribuídas.

Dentre a variedade de família de cópias podem ser destacadas as classes de cópias de Marshall-Olkin, elípticas e Arquimedianas (EMBRECHTS et al., 2003). A classe de cópias Arquimedianas (enfoque deste trabalho) está relacionada às distribuições de sobrevida bivariadas.

Suponha $\psi: [0, \infty) \rightarrow [0, 1]$, uma função convexa estritamente decrescente com $\psi(0) = 1$, $\psi' < 0$ e $\psi'' > 0$. A cópia Arquimediana é expressa por:

$$C_{\alpha}(u_1, u_2) = \psi(\psi^{-1}(u_1) + \psi^{-1}(u_2)).$$

Para todo $(u_1, u_2) \in [0,1]^2$, sendo α definido como parâmetro de associação. Dentre a ampla classe citaremos três famílias (Frank, Clayton e Estável-Positiva) que possuem um gerador específico ψ^{-1} com um único parâmetro de associação α .

A partir de duas distribuições marginais Weibull, estabelecemos uma distribuição de sobrevivência bivariada utilizando a família Estável-Positiva. A cópula Estável-Positiva (Gumbel-Hougaard) é gerada por:

$$\psi^{-1}(u) = [-\ln(u)]^{\alpha}, \text{ com parâmetro } \alpha \geq 1.$$

Realizando as transformações necessárias encontramos:

$$\psi(u) = \exp\{-u^{\frac{1}{\alpha}}\},$$

então, por substituição na Cópula Arquimediana, a função pode ser escrita como:

$$C_{GH}(u_1, u_2; \alpha) = \exp\{ - [(-\ln(u_1))^{\alpha} + (-\ln(u_2))^{\alpha}]^{\frac{1}{\alpha}} \}.$$

Podemos agora encontrar a função de sobrevivência conjunta dada por:

$$S(t_1, t_2; \theta_1, \theta_2; \alpha) = C(S_1(t_1, \theta_1), S_2(t_2, \theta_2); \alpha),$$

com $\theta_j, j = 1, 2$ sendo parâmetros das distribuições marginais e α o parâmetro de associação entre as variáveis aleatórias.

Para a estimação dos parâmetros, o método de máxima verossimilhança é amplamente utilizado. A função de verossimilhança com o vetor de parâmetros $\theta = (\theta_1, \theta_2; \alpha)$ é dada por:

$$L(\theta/y) = \prod_{i=1}^n f(t_{i1}, t_{i2}; \theta_1, \theta_2),$$

de forma que:

$$f(t_1, t_2; \theta_1, \theta_2) = \frac{\partial^2}{\partial t_1 \partial t_2} S(t_1, t_2; \theta_1, \theta_2).$$

A distribuição de sobrevivência conjunta obtida empregando-se a família Estável-Positiva, e considerando duas marginais com distribuições Weibull pode ser expressa por:

$$S(t_1, t_2) = \exp\{ - [(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2}]^{\frac{1}{\alpha}} \}.$$

A partir desta, a função densidade de probabilidade pode ser obtida. Esta é dada por:

$$f(t_1, t_2) = \frac{ \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} (\lambda_2 t_2)^{\alpha\beta_2} \beta_2 (\lambda_1 t_1)^{\alpha\beta_1} \beta_1 \exp\left\{ - \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} \right\} }{ t_2 \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^2 t_1 } +$$

$$\begin{aligned}
& + \frac{\left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} (\lambda_1 t_1)^{\alpha\beta_1} \beta_1 \exp \left\{ - \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} \right\} (\lambda_2 t_2)^{\alpha\beta_2} \alpha \beta_2}{t_1 \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^2 t_2} + \\
& + \frac{\left(\left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} \right)^2 (\lambda_1 t_1)^{\alpha\beta_1} \beta_1 (\lambda_2 t_2)^{\alpha\beta_2} \beta_2 \exp \left\{ - \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^{\frac{1}{\alpha}} \right\}}{t_1 \left[(\lambda_1 t_1)^{\alpha\beta_1} + (\lambda_2 t_2)^{\alpha\beta_2} \right]^2 t_2}.
\end{aligned}$$

Assim a função de verossimilhança é expressa por:

$$L(\theta/y) = \prod_{i=1}^n f(t_{i1}, t_{i2}; \theta_1, \theta_2).$$

O ponto de máximo da função de verossimilhança (em relação aos parâmetros) para esta função será a mesma se aplicamos logaritmos. Então:

$$L(\theta/y) = \prod_{i=1}^n f(t_{i1}, t_{i2}; \theta_1, \theta_2) = \log \prod_{i=1}^n f(t_{i1}, t_{i2}; \theta_1, \theta_2) = \sum_{i=1}^n \log f(t_{i1}, t_{i2}; \theta_1, \theta_2).$$

Como ilustração, este modelo foi aplicado a um conjunto de dados bivariados independentes, para dados gerados, considerando duas variáveis aleatórias marginais com distribuição Weibull. O mesmo modelo foi aplicado em um conjunto de dados bivariado dependente, tendo distribuição exponencial de Block-Basu, gerados através do método da rejeição. Em ambos, os parâmetros foram gerados com um vício pequeno na situação em que as amostras são independentes, ou seja, quando o parâmetro que mede a dependência está bem próximo do valor 1 (situação de independência). Na situação em que existe dependência entre as duas variáveis aleatórias, observa-se elas afetam os parâmetros de forma e escala das distribuições marginais, o que é natural, considerando que este parâmetro de associação se encontra como o expoente da função. Estes resultados demonstram que o modelo encontrado poderá ser utilizado em diversos aspectos, podendo inclusive ser utilizado para estimar o grau de dependência entre as variáveis aleatórias em estudo.

Referências Bibliográficas

EMBRECHTS, P., LINSKOG, F., MCNIEL, A. *Modelling dependence with copulas and applications to risk management*. <http://www.math.ethz.ch/~baltes/fip/papers.html>, 2003.

NÚÑEZ, J.S. *Modelagem bayesiana para dados de sobrevivência bivariados através de cópulas*. Tese apresentada ao Instituto de Matemática e Estatística da universidade de São Paulo para obtenção do grau de doutor em Estatística, São Paulo, 2005.

OAKES, D. *Bivariate survival models induced by frailties*. **Journal of the American Statistical Association**, v.84, p.487-493, 1989.

Bolsa: FAPESP.