Uma Aplicação de Cópulas em Hidrologia

Erique Pereira Neto *
Universidade de Brasília

Geiziane Oliveira †
Universidade de Brasília

Cira Etheowalda Guevara Otiniano ‡
Universidade de Brasília

30 de abril de 2018

Resumo

Neste trabalho utiliza-se a teoria de cópulas para modelar dados de máximos anuais de vazões para fins de estimativas do tempo de retorno de vazão de rios para dados de represas brasileiras. Utilizou-se cópulas arquimedianas para modelar a distribuição conjunta bivariada de máximos de vazões anuais das represas de Taquaruçu e Capivara localizadas no Rio Paranapanema. As cópulas foram estimadas utilizando como marginais a distribuição Gev, a cópula com melhor ajuste foi a Gumbel.

Palavras Chaves: Cópulas, hidrologia, tempo de retorno bivariado

1 Introdução

A teoria de cópulas vem sendo apresentada na literatura como uma ferramenta que permite construir uma função de distribuição multivariada das variáveis de interesse. E essa propriedade ganhou muita importância, pois permite descrever estrutura de dependência entre variáveis aleatórias por meio da função de distribuição conjunta. Através da copula pode-se separar o comportamento marginal das variáveis aleatórias de sua estrutura de dependência.

Um dos problemas importantes em hidrologia é a previsão do tempo de retorno de um fenômeno extremal, que pode ser um desastre natural. Para tal, quando os dados não apresenta correlação serial as distribuições extremal são as mais utilizadas. Nesse trabalho, será utilizado as famílias de cópulas arquimedianas para modelar o tempo de retorno bivariado para máximo anuais de vazão.

2 Metologia e Análise

2.1 Cópulas

Cópula é uma função de distribuição n-variada $C : [0,1]^n \to [0,1]$ não decrescente em cada componente, contínua a direita em cada componente, com

*UnB, E-mail: epnetos300@outlook.com
†UnB, E-mail: geizianeos@outlook.com
‡UnB, E-mail: cira@unb.br

marginais uniforme [0,1]. De modo geral, pode-se dizer que uma cópula é uma função de distribuição n-variada restrita ao retângulo $[0,1]^n$, cujas as marginais possui distribuição uniforme. Para o caso bivariado(n=2), tem-se que $C:[0,1]^2 \to [0,1]$ função de distribuição 2-variada com as seguintes propriedades:
1. Para todo u, v em $[0,1]$,

$$C(u,0) = C(0,v) = 0$$

e

$$C(u,1) = u \text{ e } C(1,v) = v \; ;$$

2. Para todo u_1, u_2, v_1, v_2 em $[0,1]$ tal que $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0.$$

Teorema de Sklar (Sklar; 1959): Seja H uma função de distribuição bivariada com marginais F e G, então existe uma cópula C, tal que

$$H(x,y) = C(F(x),G(y)). \tag{1}$$

A cópula é única se as marginais forem contínuas. A

2.2 Famílias de Cópulas Arquimedianas

Uma cópula Arquimediana é uma função $C:[0,1]^2 \longrightarrow [0,1]$, dada por $C(u,v) = \varphi^{[-1]}\left(\varphi(u)+\varphi(v)\right)$ em que φ uma função contínua convexa estritamente decrescente de $[0,1] \longrightarrow [0,\infty]$ tal que $\varphi(1) = 0$ e, em que $\varphi[-1]$ denota a "pseudo-inversa"de φ, isto é

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{se} \quad t \in [0,\varphi(0)) \\ 0 & \text{se} \quad t \geq \varphi(0) \end{cases} \tag{2}$$

Quando $\varphi(0) = \infty$, φ é dito ser estrito e $\varphi^{[-1]} \equiv \varphi^{-1}$. φ é chamado de gerador.

A tabela abaixo apresenta as principais cópulas arquimedianas e suas funções geradoras.

Tabela 1: Principais Cópulas arquimedianas

Copula	$C_\theta(u,v)$	$\varphi_\theta(t)$	$\theta \in$
Clayton	$[u^{-\theta}+v^{-\theta}-1]^{1/\theta}$	$\frac{t^{-\theta}-1}{\theta}$	$[-1,\infty]\diagdown\{0\}$
Gumbel	$\exp(-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{1/\theta}$	$(\ln t)^{-\theta}$	$[1,\infty)$
Frank	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\diagdown\{0\}$

2.3 Métodos de Estimação de Cópulas

Existe alguns métodos para estimar cópulas, paramétricos e não-paramétricos, dentre os métodos paramétricos os mais comuns são métodos de máxima verossimilhança exato(MLE), método de verossimilhança em dois passos. No primeiro, o parâmetro da cópula é estimado junto com suas marginais em um único passo, no segundo, as marginais são estimadas separadamente e depois em um segundo passo junta-se a estimação das cópulas.

2.4 Método de Máxima Verossimilhança

Considere o vetor aleatório bivariado $(X, Y) \sim H$, tal que $(F(x), G(y)) \sim C(F^{-1}(u), G^{-1}(v))$, com $u = F(x)$ e $v = G(y)$.

Pelo Teorema de Sklar, a distribuição conjunta das variáveis aleatórias são descritas pela cópula :

$$H(x, y) = C_\theta(F(x), G(y))$$

.

Temos $F(x) = F(x; \Phi_1)$ e $G(y) = G(y; \Phi_2)$, portanto, tem-se no mínimo 3 parâmetros, e o objetivo é estimar (Φ_1, Φ_2, θ), sendo o vetor de parâmetros referente à cópula e marginais.

$$\hat{\theta} = \mathbf{argmax}_{\theta \in \Theta} \sum_{i=1}^{n} log(c(x_i, y_i; \Phi_1, \Phi_2, \theta))$$

Com $c(x_i, y_i; \Phi_1, \Phi_2, \theta)$ sendo a densidade da cópula, ou seja, da função de distribuição conjunta.

2.5 Método Inferência pelas Marginais

O método de Inferência pelas Marginais foi proposto por Joe (1997), e ressalta o próprio conceito de cópulas, portanto, o vetor de parâmetros (Φ_1, Φ_2, θ) é separado em duas partes, parâmetros específicos para as marginais (Φ_1, Φ_2) e θ parâmetro comum para a estrutura de dependência. Portanto, no primeiro passo estima-se os parâmetros (Φ_1, Φ_2, θ) pelo método de máxima verossimilhança e em seguida substitui-se na Equação 4, para estimar θ.

$$\hat{\theta} = \mathbf{argmax}_\theta l(\theta)$$

$$l(\theta, \hat{\Phi}_1, \hat{\Phi}_2) = \sum_{i=1}^{n} log[c(F(x_i, \hat{\Phi}_1), G(y_i, \hat{\Phi}_2); \theta)] + \sum_{i=1}^{n} log[f(x_i, \hat{\Phi}_1 g(y_i, \hat{\Phi}_2)].$$

3 Aplicação em dados de Hidrologia

Em hidrologia um dos principais objetivos é a estimativa do tempo de retorno. O tempo de retorno é definido como sendo o tempo médio entre duas ocorrências sucessivas de um evento recorrente, geralmente eventos extremos e que causam enorme impacto.

Neste trabalho, o objetivo é estimar o tempo de retorno bivariado, através da modelagem de cópulas, de máximo de vazões anuais das represas de Taquaruçu e Capivara localizadas no Rio Paranapanema.

3.1 Tempo de retorno bivariado

Considerando o par de variáveis aleatórias (X, Y), com marginais F_X e G_Y respectivamente. O tempo de retorno bivariado é dado por:

$$T = \frac{1}{P(X > x, Y > y)} \tag{3}$$

A probabilidade $P(X > x, Y > y)$ pode ser do tipo OU ou E, ou seja, pode ocorrer um evento extremo se X ou Y forem maior que um certo limiar, ou se ambos são maiores que um respectivo limiar, (Salvadori e De Michele, 2011). Neste trabalho, estamos interessados no primeiro caso, que significa para os dados em análise que o máximo anual de vazão da represa de Taquaruçu ou Capivara excedem um determinado limiar. Usando cópulas, tempos que o tempo de retorno bivariado pode ser estimado por:

$$T = \frac{1}{1 - C(u, v)} \tag{4}$$

Para saber se um evento é extremo, definimos regiões críticas. São elas:

1. $R_t^< = \{(u, v) \in I^2 : C(u, v) < t\}$, $0 < t \leq 1$. Chamada de região subcrítica.

2. $L_t = \{(u, v) \in I^2 : C(u, v) = t\}$, $0 < t \leq 1$. Chamada de região crítica.

3. $R_t^> = \{(u, v) \in I^2 : C(u, v) > t\}$, $0 < t \leq 1$. Chamada de região supercrítica.

A região onde ocorre os eventos extremos de interesse neste trabalho pertencem à região $R_t^>$. E assim, o tempo de retorno é dado por

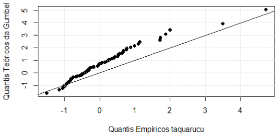
$$T = \frac{1}{P(R_t^>)} = \frac{1}{1 - P(C(u, v) \leq t)} = \frac{1}{1 - K_C(t)} \tag{5}$$

Em que $K_C(t) = P(C(u, v) \leq t)$ é a função de Kendall. Nas cópulas arquimedianas temos $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$.

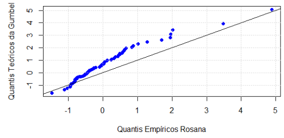
3.2 Resultado de análise dos dados de vazões

As séries de dados usadas neste trabalho são os máximos anuais de vazão das represas de Taquaruçu e Capivara localizadas no rio Paranapanema no período de 1931 a 2009.

Com base nos QQ-Plots foi ajustada uma distribuição GEV, usando estimação por máxima verossimilhança.



(a) QQ-Plot dos quantis empíricos padronizados das vazões anuais da represa de Taquaruçu x quantis teóricos da Gumbel



(b) QQ-Plot dos quantis empíricos padronizados das vazões anuais da represa de Rosana x quantis teóricos da Gumbel

Tabela 2: Estimativas via método de máxima verossimilhança das Marginais GEV dos máximos anuais de vazões conjunto das represas de Taquaruçu e Capivara

Estimativas parâmetros Mle			
Represa	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\mu}$
Taquaraçu	0.13	701.54	1838.08
Capivara	0.13	659.08	1742.13

Usando método de Inferência pelas Marginais, os parâmetros das cópulas estimados estão na tabela 4.

Tabela 3: Estimativas para as cópulas arquimedianas com marginais GEV

Cópula	Parâmetros	Estimativas	L	BIC
Gumbel	$\hat{\theta}$	36.60	1001.45	-1998.54
Frank	$\hat{\theta}$	36.68	1117.733	-2231.09
Clayton	$\hat{\theta}$	51.47	1061.47	-2118.57

Nota-se o melhor ajuste ocorre com a cópula de Gumbel, pois o BIC é menor e Log verossimilhança maior.

A tabela a seguir mostra os valores limiares de dos máximos anuais de vazões das duas represas e o tempo de retorno bivariado,considerando a cópula de Gumbel.

Tabela 4: Estimativas do tempo de retorno em anos e do limiar bivariado em m^3/s dado um valor crítico p.

X(Máx vazões Capivara)	Y(Máx vazões Taquaruçu)	p	$K_C(p)$	TR (anos)
3679,846	3661,438	0,90	0,9217251	12,77
4395,002	4374,794	0,95	0,9618897	26,23
6291,614	6266,631	0,99	0,9927080	137,13
9787,575	9753,790	0,999	0,9993004	1429,48
14531,061	14485,332	0,9999	0,9999319	14681,77

4 Conclusão

Neste trabalho foi explorado a aplicação da teoria de cópulas no estudo de dados de hidrologia. Obteve-se estimativas do tempo de retorno bivariado, em que as variáveis marginais eram os máximos anuais de vazões das represas de Taquaruçu e Capivara localizadas no Rio Paranapanema, estimando-se distribuições marginais GEV. Essas duas variáveis possuem uma forte correlação, apresentando um valor de 0.978 para o tau de Kendall. Foram estimados os parâmetros das cópulas arquimedianas de Gumbel, Clayton e Frank usando o método de Inferência por Marginais, sendo melhor ajustado a cópula de Gumbel que resultou em um menor BIC. Estimado a cópula, foram encontrados os tempos de retornos bivariado estimados através da função de Kendall.

5 Referências Bibliográficas

[1] Nelsen, Roger *An Introduction to Copulas* , Springer, 2011

[2] P, Embrechts; C, Kuppelberg; T, Mikosch. *Modelling Extremal Events for Insurance and Finance.*Springer-Verlag, NewYork, 1997.

[3] G. Salvadori, C De. Michele e F.Duarte *On the return and design in a multivariate framework.* Hidrology and earth System Sciences, 2011.

[4] G. Salvadori, C De. Michele e F.Duarte *Frequency analysis via copulas:Theorical aspects and applications to hidrological events* . Water Resour, 2011.