

Universidade de Brasília

Pedro Assunção Rangel

TEORIA DO VALOR EXTREMO E CÓPULAS

Aplicação de Cópuas em Dados Hidrológicos

Prof^ª. Cira Etheowalda Guevara Otiniano

Brasília
Dezembro de 2017

A teoria de Cópulas é usada para a estimação de distribuições multivariadas a partir das distribuições marginais, assim pelo teorema de Sklar

$$F(X_1, \dots, X_d) = C(F_1(X_1), \dots, F_d(X_d)) \quad (1)$$

uma cópula é uma função d -dimensional representando a função conjunta, com as distribuições marginais sendo aplicadas para restringir o domínio ao espaço $[0, 1]^d$.

Um grupo de Cópulas chamadas de Arquimedianas são frequentemente usadas por terem apenas um parâmetro, α , e as propriedades distributiva, $C(C(u, v), w) = C(u, C(v, w))$, e a simétrica, $C(u, v) = C(v, u)$. As cópulas pertencentes a esse grupo podem ser definidas através da função geradora ψ , como

$$C(u_1, \dots, u_d) = \psi^{[-1]}(\psi(u_1) + \dots + \psi(u_d)), \quad (2)$$

$\psi^{[-1]}$ sendo a função pseudo-inversa de ψ . Com isso, podemos definir as cópulas de Gumbel, Clayton, e Frank, que serão usadas mais a frente, como

$$\text{(Gumbel)} \quad \psi(t) = (-\log t)^\alpha; \quad (3)$$

$$\text{(Clayton)} \quad \psi(t) = \frac{1}{\alpha}(t^{-\alpha} - 1); \quad (4)$$

$$\text{(Frank)} \quad \psi(t) = -\log \left(\frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1} \right). \quad (5)$$

Para a aplicação dessas cópulas, foram usados dados hidrológicos de três usinas do rio Paranapanema, Rosana, Canoas I, e Taquaruçu. Para cada uma das usinas, tem-se a enchente máxima anual para os anos de 1931 a 2009, totalizando 79 observações para as três variáveis.

A figura 1 mostra as séries dos níveis das cheias para cada uma das três usinas, e a figura 2 mostra os diagramas de dispersão para os três pares de variáveis. Nota-se que em todos os casos há um alto grau de correlação linear, especialmente entre as observações da usina de Rosana e de Taquaruçu. Calculando os coeficientes de correlação de Pearson, obtém-se os resultados apresentados na tabela a seguir.

Tabela 1: Coeficientes de correlação de Pearson para os pares de variáveis.

	Rosana	Canoas I	Taquaruçu
Rosana	—	0.8672002	0.9941207
Canoas I	0.8672002	—	0.8896892
Taquaruçu	0.9941207	0.8896892	—

Figura 1: Série temporal dos máximos anuais para cada usina.

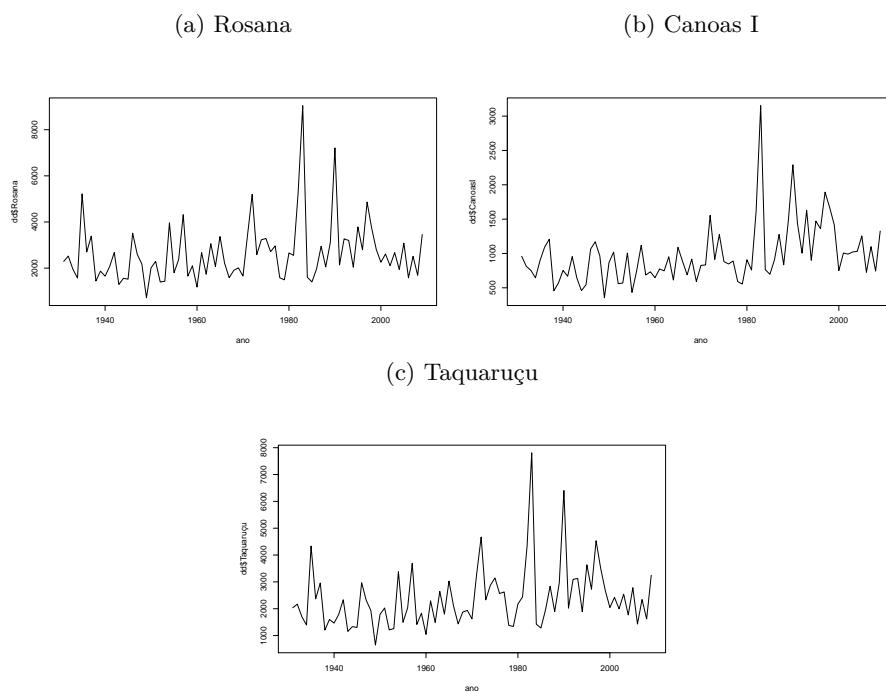
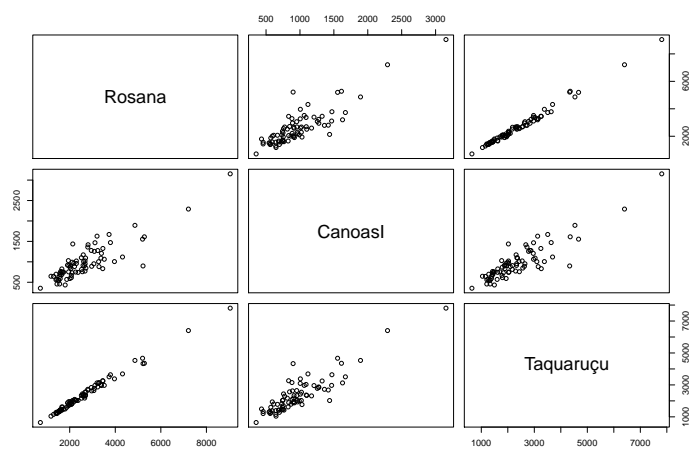


Figura 2: Correlações dos pares das observações das usinas.



Antes de realizar o ajuste das cópulas nessas variáveis, é necessário transformar os valores observados em marginais da cópula. Com o resultado da primeira parte do trabalho, na qual foi concluído que esses dados podem ser ajustados por uma distribuição GEV (*Generalized extreme value distribution*), cada uma das variáveis foi transformada para a acumulada da GEV ajustada, pelos parâmetros apresentados na tabela 2, da forma

$$u_i = G_i(X_i; \hat{\gamma}, \hat{\mu}, \hat{\sigma}), \quad i = 1, 2, 3. \quad (6)$$

Tabela 2: Parâmetros estimados da GEV.

	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$
Rosana	2033.8723	762.1377	0.1453337
Canoas I	781.9971	269.8886	0.1160789
Taquaruçu	1838.0761	701.5323	0.133161

Com as observações transformadas para o espaço $[0, 1]$, foram então ajustadas as três tipos de cópulas, Gumbel, Clayton, e Frank, para cada par. Os valores do parâmetro α estão na tabela 3. Para comparar em cada caso qual cópula foi a melhor ajustada, obtiveram-se os valores dos critérios *AIC* e *BIC*, nas tabelas 4 e 5, respectivamente.

Tabela 3: Estimação de α para cada cópula, por par de variáveis.

	Gumbel	Clayton	Frank
Ros.; Can. I	2.55	1.92	8.53
Ros.; Taq.	11.84	14.89	47.92
Can. I; Taq.	2.84	2.27	9.62

Tabela 4: Valores do *AIC* para cada cópula ajustada, por par de variáveis.

	Gumbel	Clayton	Frank
Ros.; Can. I	-91.44311	-63.99471	-82.21746
Ros.; Taq.	-328.46227	-282.43994	-307.61240
Can. I; Taq.	-105.83756	-76.19883	-94.36823

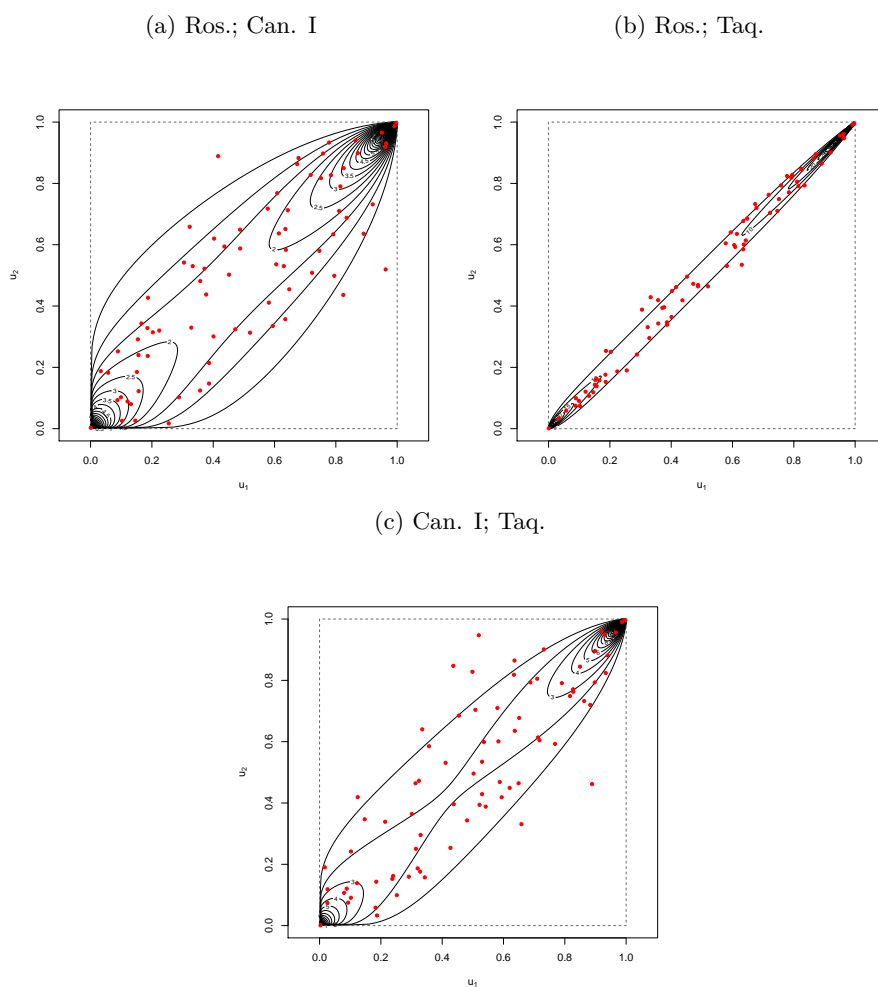
Com os resultados obtidos, nota-se que em cada caso a cópula Gumbel é a mais adequada para modelar os dados bivariados em cópulas, o que é justificado pelos dados representarem valores extremos, e terem cauda pesada.

Os gráficos na figura 3 mostra as curvas de níveis de cada uma das três cópulas ajustadas, junto com os pares de observações.

Tabela 5: Valores do BIC para cada cópula ajustada, por par de variáveis.

	Gumbel	Clayton	Frank
Ros.; Can. I	-89.07366	-61.62526	-79.84801
Ros.; Taq.	-326.09283	-280.07049	-305.24295
Can. I; Taq.	-103.46811	-73.82938	-91.99878

Figura 3: Curvas de níveis para as cópulas de Gumbel ajustadas, com as observações em pontos vermelhos.



O próximo passo foi encontrar os níveis das enchentes máximas anuais para um tempo de retorno fixo. Como nesse caso, em vez de apenas uma variável, temos pelo menos duas, fixando o tempo de retorno, o podemos encontrar um valor fixo na cópula de sobrevivência, porém dessa forma haverá uma curva inteira podendo corresponder ao valor desejado. Por exemplo, nessa aplicação, o valor fixo do tempo de retorno foi $T = 100$ anos, e como as observações correspondem ao máximo anual, queremos uma enchente que daria um valor maior do que um limiar para duas usinas no mesmo ano. Como o tempo de retorno é definido por

$$T(x) = \frac{1}{1 - F(x)} = \frac{1}{\bar{F}(x)} \Rightarrow T(u, v) = \frac{1}{\bar{C}(u, v)}, \quad (7)$$

isso corresponde a procurar na cópula de sobrevivência quais pares correspondem ao valor de 0.01, porém são infinitos valores, todos presentes numa mesma curva, definida pelo corte do gráfico nessa altura do eixo z .

Para facilitar a busca de pelo menos um valor, foi adicionada a restrição de que na cópula, os limiares devem ser os mesmos. Consequentemente, a busca foi restrita à diagonal $u = v$, que vai de 0 a 1 na cópula de sobrevivência, e procuramos o valor que retorna a probabilidade de 0.01.

Os valores obtidos dessa forma estão apresentados na tabela 6, e ilustrados na figura 4.

Tabela 6: Valores dos limiares encontrados.

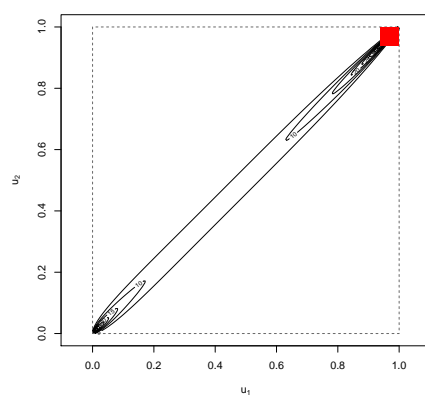
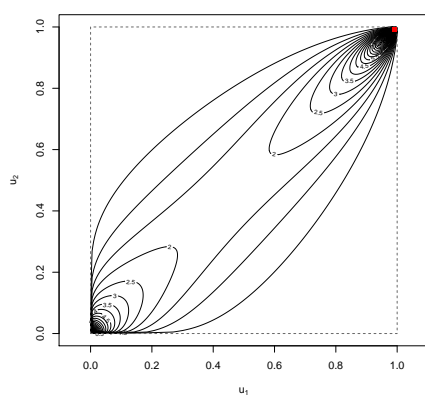
Ros.; Can. I	Ros.; Taq.	Can. I; Taq.
0.9855129	0.9381373	0.9605134

Com esses resultados, foi possível obter a vazão para outros tempos de retorno, a partir da distribuição GEV, e os limiares obtidos das cópulas da mesma forma.

Figura 4: Curvas de níveis para as cópulas de Gumbel ajustadas, com os níveis encontrados indicados pelos quadrados vermelhos.

(a) Ros.; Can. I

(b) Ros.; Taq.



(c) Can. I; Taq.

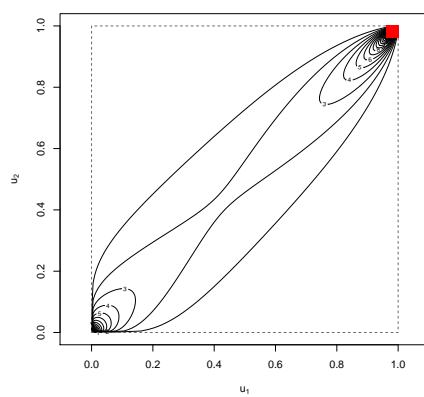


Tabela 7: Limiares por tempo de retorno para a cópula Rosana-Canoas I.

Tempo	$u = v$	Rosana	Canoas I
100	0.9855	6483	2255
200	0.9927	7512	2573
300	0.9952	8167	2773
400	0.9963	8640	2916
500	0.9971	9049	3038
600	0.9976	9379	3136
700	0.9979	9674	3224
800	0.9982	9910	3294
900	0.9984	10118	3355
1000	0.9986	10363	3427

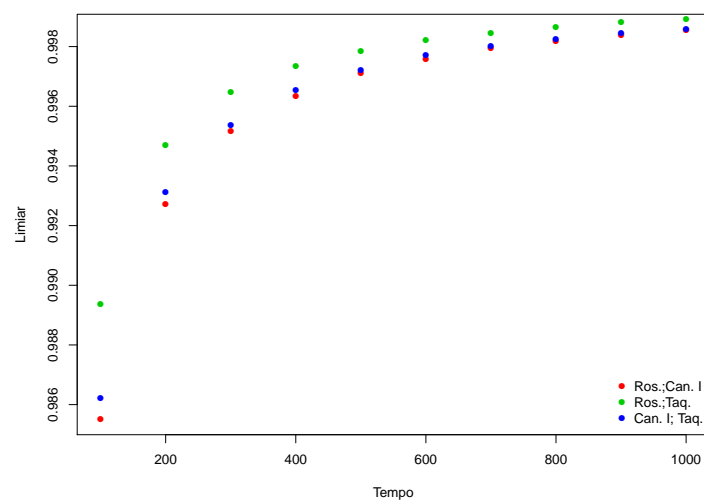
Tabela 8: Limiares por tempo de retorno para a cópula Rosana-Taquaruçu.

Tempo	$u = v$	Rosana	Taquaruçu
100	0.9894	6932	6212
200	0.9947	8021	7155
300	0.9965	8700	7741
400	0.9973	9203	8172
500	0.9979	9605	8516
600	0.9982	corr 9958	8817
700	0.9985	10236	9054
800	0.9987	10520	9295
900	0.9988	10749	9489
1000	0.9989	10968	9675

Tabela 9: Limiares por tempo de retorno para a cópula Canoas I-Taquaruçu.

Tempo	$u = v$	Canoas I	Taquaruçu
100	0.9862	2277	5882
200	0.9931	2600	6790
300	0.9954	2794	7342
400	0.9965	2944	7771
500	0.9972	3059	8100
600	0.9977	3167	8409
700	0.9980	3247	8642
800	0.9983	3317	8843
900	0.9985	3385	9041
1000	0.9986	3441	9205

Figura 5: Limiares por tempo de retorno.



Referências

- [1] Roger B. Nelsen (2006). *An Introduction to Copulas*. Springer, 2 ed.
- [2] Davide Luciano De Luca & Daniela Biondi (2017). “Bivariate Return Period for Design Hyetograph and Relationship with T-Year Design Flood Peak”. *Water*, 9.673. DOI [10.3390/w9090673](https://doi.org/10.3390/w9090673).
- [3] R Core Team (2017). “R: A language and environment for statistical computing”. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [4] Eric Gilleland, & Richard W. Katz (2016). “extRemes 2.0: An Extreme Value Analysis Package in R”. *Journal of Statistical Software*, 72.8, 1-39. DOI [10.18637/jss.v072.i08](https://doi.org/10.18637/jss.v072.i08).
- [5] Jun Yan (2007). “Enjoy the Joy of Copulas: With a Package copula”. *Journal of Statistical Software*, 21.4, 1-21. URL <http://www.jstatsoft.org/v21/i04/>.

Apêndice: Código em R

```
library("copula")
library("scatterplot3d")
library("extRemes")

5 dd <- read.csv("usinas.csv")
  ano <- 1931:2009

  plot(ano, dd$Rosana, type = "l")
  plot(ano, dd$CanoasI, type = "l")
10 plot(ano, dd$Taquaruçu, type = "l")

  plot(dd)
  cor(dd)

15 gev1 <- fevd(dd[, 1])$results$par
  u1 <- pevd(dd[, 1], loc = gev1[1], scale = gev1[2], shape = gev1[3])

  gev2 <- fevd(dd[, 2])$results$par
  u2 <- pevd(dd[, 2], loc = gev2[1], scale = gev2[2], shape = gev2[3])
20
  gev3 <- fevd(dd[, 3])$results$par
  u3 <- pevd(dd[, 3], loc = gev3[1], scale = gev3[2], shape = gev3[3])

  cop1 <- archmCopula("gumbel", dim = 2)
25 cop2 <- archmCopula("clayton", dim = 2)
  cop3 <- archmCopula("frank", dim = 2)

  alpha <- matrix(0, 3, 3)
  ll <- matrix(0, 3, 3)
30
  dd1 <- matrix(c(u1, u2), ncol = 2)

  ajus11 <- fitCopula(cop1, dd1, method = "mpl")
  alpha[1, 1] <- ajus11@estimate
35 ll[1, 1] <- ajus11@loglik

  ajus12 <- fitCopula(cop2, dd1, method = "mpl")
  alpha[1, 2] <- ajus12@estimate
  ll[1, 2] <- ajus12@loglik
40
  ajus13 <- fitCopula(cop3, dd1, method = "mpl")
  alpha[1, 3] <- ajus13@estimate
  ll[1, 3] <- ajus13@loglik

45 dd2 <- matrix(c(u1, u3), ncol = 2)

  ajus21 <- fitCopula(cop1, dd2, method = "mpl")
  alpha[2, 1] <- ajus21@estimate
  ll[2, 1] <- ajus21@loglik
50
  ajus22 <- fitCopula(cop2, dd2, method = "mpl")
  alpha[2, 2] <- ajus22@estimate
  ll[2, 2] <- ajus22@loglik

55 ajus23 <- fitCopula(cop3, dd2, method = "mpl")
```

```

alpha[2, 3] <- ajus23@estimate
ll[2, 3]    <- ajus23@loglik

dd3        <- matrix(c(u2, u3), ncol = 2)

60  ajus31    <- fitCopula(cop1, dd3, method = "mpl")
alpha[3, 1] <- ajus31@estimate
ll[3, 1]    <- ajus31@loglik

65  ajus32    <- fitCopula(cop2, dd3, method = "mpl")
alpha[3, 2] <- ajus32@estimate
ll[3, 2]    <- ajus32@loglik

ajus33      <- fitCopula(cop3, dd3, method = "mpl")
70  alpha[3, 3] <- ajus33@estimate
ll[3, 3]    <- ajus33@loglik

aic <- 2 - 2 * ll
bic <- log(79) - 2 * ll

75  contour(ajus11@copula, dCopula, n = 200, nlevels = 160, asp = 1)
points(u1, u2, pch = 20, col = 2)

contour(ajus21@copula, dCopula, n = 200, nlevels = 160, asp = 1)
80  points(u1, u3, pch = 20, col = 2)

contour(ajus31@copula, dCopula, n = 200, nlevels = 160, asp = 1)
points(u2, u3, pch = 20, col = 2)

85  persp(ajus11@copula, dCopula)
persp(ajus21@copula, dCopula)
persp(ajus31@copula, dCopula)

fd <- function(x, alpha, cop0) {
90  pr <- pCopula(c(1 - x, 1 - x), rotCopula(cop0))
    return(abs(alpha - pr))
}

limiar <- function(alpha, cop0) {
95  return(optimize(fd, interval = c(0, 1), alpha = alpha, cop0 = cop0)
    $minimum)
}

lim1 <- limiar(.01, ajus11@copula)
lim2 <- limiar(.01, ajus12@copula)
100 lim3 <- limiar(.01, ajus13@copula)

contour(ajus11@copula, dCopula, n = 200, nlevels = 160, asp = 1)
rect(lim1, lim1, 1, 1, col = 2, border = NA)

105 contour(ajus21@copula, dCopula, n = 200, nlevels = 160, asp = 1)
rect(lim2, lim2, 1, 1, col = 2, border = NA)

contour(ajus31@copula, dCopula, n = 200, nlevels = 160, asp = 1)
rect(lim3, lim3, 1, 1, col = 2, border = NA)

```