Journal homepage http://revistas.unitru.edu.pe/index.php/SSMM

SELECCIONES MATEMÁTICAS
Universidad Nacional de Trujillo
ISSN: 2411-1783 (Online)
2019; Vol.06(2): 178 - 188.

Random variable functions used in hydrology.

Funciones de variables aleatórias utilizadas en hidrología.

Cira E. G. Otiniano * and Yuri S. Maluf †

Received, Jul. 17, 2019 *Accepted, Nov. 30, 2019*

How to cite this article:
Otiniano, C., Maluf, Y. *Random variable functions used in hydrology*. Selecciones Matemáticas. 2019; 6(2):178-188. http://dx.doi.org/10.17268/sel.mat.2019.02.04

Abstract
In this work, expressions of the cumulative distribution function of YX, Y/X *and* $X/(X+Y)$ *for continuous dependent random variables with supported on a unbounded and bounded interval are derived. The dependence approach is based on copula functions. Additionally, the methodology is applied to real data on hydrology.*
Keywords. Product, Ratio, Copula, Hydrology.

Resumen
En este trabajo, se derivan expresiones de la función de distribución acumulada de YX, Y/X *y* $X/(X+Y)$ *para variables aleatorias dependientes continuas con soporte en un intervalo ilimitado y limitado. El enfoque de dependencia se basa en funciones cópula. Además, la metodología se aplica a datos reales de hidrología.*
Palabras clave. Producto, Cociente, Copula, Hidrologia.

1. Introduction. The probability distributions of Y/X, YX and $\frac{X}{X+Y}$, where X and Y are random variables(r.v's), commonly arise in statistical inference and in problems in econometrics, engineering, physics and hydrology. Although many results of distributions of products and ratios, based on independent random variables, are available in the literature (for instance see Pham-Gia (2000), Shakil and Kibria (2006), Ali et al. (2007), Nadarajah and Ali (2008), and Idrizi (2014)) little appears to have been done to obtain these distributions for dependent random variables. In all these studies, a particular joint bivariate distribution function is considered. For instance, Nadarajah (2005) ([13]-[15]) include bivariate Gumbel, Lomax and gamma; Nadarajah and Ali (2006) considered Lawrance and Lewis's bivariate exponential; Nadarajah and Dey (2006) considered bivariate t; Gupta and Nadarajah (2006) worked with Mackay's bivariate; Nadarajah and Gupta considered Cherian's bivariate gamma; Nadarajah and Kotz (2006) considred Downton's bivariate exponential; Nadarajah (2007) considered bivariate gamma exponential; Nagar et al. (2009) considered bivariate beta.

Many papers in the literature have been devoted to the tail behavior of the product of dependent random variables. See, for example, Chen et al. (2018), Chen et al. (2016), Yang and Sun (2013), Qu and Chen (2013), and Yang (2011).

Following a different approach Domma and Giordano (2012, 2013) derive the distribution of Y/X introducing the dependence between X and Y through a copula function. They presented an application in the context of household financial fragility through the stress-force model with Frank's copula. However, they employ a specific copula and distributions for random variables to evaluate $P(Y/X < 1)$.

Dolati et al. (2017) presented the cumulative distribution of the linear combination, of the product and of the ratio of continuous dependent random variables in terms of their associated copula. They obtained the distribution of the product and of the ratio for continuous and positive random variables and discussed the effect of dependence

*Departamento de Estatística, UnB, 70910-900 Brasília, DF, Brazil (cira@unb.br)
†Departamento de Estatística, UnB, 70910-900 Brasília, DF, Brazil (yurimaluf@gmail.com),

on stress-strength function. However, quantities that are modeled by random variables are not always positive and continuous. For example, asset returns or exchange rates (in finance) and temperatures (in hydrology) correspond to continuous random variables that assume positive and negative values. In survival analysis, the time until the occurrence of an event of interest (death, cure, disease,marriage, divorce, etc.) can be measured in days, weeks, years, etc. Then, these times are discrete random variables.

Thus, the cumulative distribution of the product and of the ratio of continuous dependent random variables with support on a unbounded and bounded interval is required.

In this work, in Section 2, we first obtained the cumulative distribution of $R = Y/X$, $Z = YX$ for continuous dependent random variables with support in the real line. The Proposition 4.1 of Dolati et al. (2017) is a particular case of this result. Then, the distribution of R and Z for continuous random variables with support on a bounded interval we also obtained. Still in the continuous case, in two propositions, we derive the distribution of $W = \frac{X}{X+Y}$ for random variables with support on a unbounded and bounded interval. For illustrative purposes, in Section 3 , we use real data sets to analized the stress-strength function (for hydrology problems) by calculating the distribution of R. Section 4 concludes the work.

2. Main results. Let X and Y be two random variables (r.v.) with cumulative distribution functions (cdfs) F_X and F_Y. A flexible tool which is able to describe the dependence structure between these variables is the copula function. A bivariate copula is a bivariate distribution function defined on the unit square $[0,1]^2$, with uniformly distributed marginals. According to the Sklar's theorem, any bivariate distribution $F(x,y) = P(X \le x, Y \le y)$ with continuous margins can be written as $F(x,y) = C(F_X(x), F_Y(y))$ or $C(u,v) = F(F_X^{-1}(u), F_Y^{-1}(v))$, where C is a unique copula and $u, v \in [0,1]$. The function F^{-1} is the generalized inverse, $F^{-1} : (0,1) \to R$, defined as

$$F^{-1}(t) = \inf\{x \in R : F(x) \ge t\}. \tag{2.1}$$

The copula is not unique when the marginals are discrete. The same is valid for the survival copula $\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v)$, since it is also a copula function. Theory, properties and applications of the most popular copula families can be found in Nelsen (2006) and Trivedi and Zimmer (2005).

2.1. Continuous case. Before proceeding to the propositions regarding the distributions of $R = Y/X$, $Z = YX$ and $W = \frac{X}{X+Y}$, we announce the following lemma:

Lemma 1. Let $C(u,v)$ be a copula function associated with vector (X,Y) and continuous margins cdfs F_X and F_Y. Then, $\forall t$ $D_1C\left(u, F_Y\left(tF_X^{-1}(u)\right)\right)$ is well defined $\forall u \in I = [0,1]$ almost surely (a.s.), where $D_1C(u,v)$ is the partial derivative of C with respect to the first argument and $D_1C(u,v) = P(V \le v | U = u)$ with $U = F_X(X)$ and $V = F_Y(Y)$.

The proof of the Lemma 1 has the same arguments as the proof in the Lemma 2.1. of Cherubini et al. (2011). They proved the existence of $D_1(u, F_Y(t - F_X^{-1}(u)))$.

Proposition 2.1. Let X and Y be two continuous random variables on the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let C be a copula function associated with X and Y and marginal distribution functions F_X and F_Y, respectively. Then for all $t \in R$ a.s. the distributions of $R = Y/X$ and $Z = X \cdot Y$ are given by

$$\begin{aligned} F_R(t) = F_X(0) &- \int_0^{F_X(0)} D_1C\left(u, F_Y\left(tF_X^{-1}(u)\right)\right) du \\ &+ \int_{F_X(0)}^1 D_1C\left(u, F_Y\left(tF_X^{-1}(u)\right)\right) du \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} F_Z(t) = F_X(0) &- \int_0^{F_X(0)} D_1C\left(u, F_Y\left(t/F_X^{-1}(u)\right)\right) du \\ &+ \int_{F_X(0)}^1 D_1C\left(u, F_Y\left(t/F_X^{-1}(u)\right)\right) du, \end{aligned} \tag{2.3}$$

respectively.

Demostración: Consider the joint cumulative distribution function of X and R expressed by

$$\begin{aligned} F_{X,R}(s,t) &= P(X \le s, R \le t) \\ &= \int_{-\infty}^s \int_{-\infty}^t f_{X,R}(x,r) dx dr. \end{aligned}$$

With conditioning on X, the cdf $F_{X,R}$ is written as

$$F_{X,R}(s,t) = \int_{-\infty}^{s} P\left(Y/x \le t | X = x\right) dF_X(x).$$

Since $g(y,x) = y/x$ is not a bijection, we divide the function g in two parts, for $x < 0$ and for $x > 0$. Thus, $F_{X,R}$ can be rewritten as

$$F_{X,R}(s,t) = \begin{cases} \displaystyle\int_{-\infty}^{s} P\left(Y > tx | X = x\right) dF_X(x) & \text{if } s < 0 \\ \displaystyle\int_{-\infty}^{0} P\left(Y > tx | X = x\right) dF_X(x) + \int_{0}^{s} P\left(Y \le tx | X = x\right) dF_X(x) & \text{if } s \ge 0. \end{cases} \tag{2.4}$$

To find the marginal distribution of R we apply the limit $s \to \infty$ in eq. (2.4), i.e.

$$\begin{aligned} F_R(t) &= \lim_{s\to\infty} F_{X,R}(s,t) \\ &= \int_{-\infty}^{0} P\left(Y > tx | X = x\right) dF_X(x) + \int_{0}^{\infty} P\left(Y \le tx | X = x\right) dF_X(x) \\ &= \int_{-\infty}^{0} \left[1 - P\left(Y \le tx | X = x\right)\right] dF_X(x) + \int_{0}^{\infty} P\left(Y \le tx | X = x\right) dF_X(x) \\ &= F_X(0) - \int_{-\infty}^{0} P\left(Y \le tx | X = x\right) dF_X(x) + \int_{0}^{\infty} P\left(Y \le tx | X = x\right) dF_X(x). \end{aligned} \tag{2.5}$$

In eq. (2.5), the substitution $u = F_X(x)$ and Lemma 1 give the desired result.

Adopting the same idea, one can establish the distribution of $Z = YX$.

□

Our Proposition 2.1. is a generalization of the Proposition 4.1 given by Dolati et al. (2017), they considered X and Y positive random variables.

When random variables have support on a unbounded interval the expression (2.2) becomes slightly modified. The following propositions highlights these cases.

Corollary 2.2. Let (X, Y) be a vector of random variables defined on the same probability space (B, F, P) where $B = [0, b] \subseteq R^+$, the associated copula C and the continuous marginal distribution functions F_X and F_Y. Then the distribution of $R = Y/X$ is given by

$$F_R(t) = \int_0^1 D_1C\left(u, F_Y(tF_X^{-1}(u))\right) du \quad \text{if} \quad t \le 1 \tag{2.6}$$

and

$$\begin{aligned} F_R(t) = &\int_0^1 D_1C\left(vF_X(b/t), F_Y(tF_X^{-1}(vF_X(b/t)))\right) F_X(b/t)dv \\ &+ \int_{F_X(b/t)}^{1} D_1C\left(u, F_Y(b)\right) du, \quad \text{if} \quad t > 1. \end{aligned} \tag{2.7}$$

Demostración: Consider $f_{X,R}(x,r)$ the joint density of X and R where $0 \le x \le b$ and $0 \le r \le b/x$. Then integrating with respect to x and r and then conditioning on $X = x$, we have that

$$F_{X,R}(s,t) = \int_0^s \int_0^{k(t)} f_{R|X}(r|x) f_X(x) dr dx \tag{2.8}$$

$$= \int_0^s F_{R|X}(k(t)|x) dF_X(x). \tag{2.9}$$

Depending on the values of x and b, the interval of integration with respect to r is limited by $k(t) = \min\{b/x, t\}$. The limit $s \to b$ in eq. (2.8) becomes

$$F_R(t) = \int_0^b P\left(Y \le x \min\{b/x, t\} | X = x\right) dF_X(x). \tag{2.10}$$

Taking into account the cases $t \leq 1$ and $t > 1$ we can rewrite the above equation as

$$F_R(t) = \begin{cases} \int_0^b P\left(Y \leq xt | X = x\right) dF_X(x) & \text{if } t \leq 1, \\ \int_0^{b/t} P\left(Y \leq xt | X = x\right) dF_X(x) + \int_{b/t}^b P\left(Y \leq b | X = x\right) dF_X(x) & \text{if } t > 1. \end{cases}$$

Now, changing the variable $u = F_X(x)$ and using Lemma 1. leads to

$$F_R(t) = \begin{cases} \int_0^1 D_1 C\left(u, F_Y(tF_X^{-1}(u))\right) du & \text{if } t \leq 1, \\ \int_0^{F_X(b/t)} D_1 C\left(u, F_Y(tF_X^{-1}(u))\right) du + \int_{F_X(b/t)}^1 D_1 C\left(u, F_Y(b)\right) du & \text{if } t > 1. \end{cases} \tag{2.11}$$

Making another substitution $v = u\left[F_X(b/t)\right]^{-1}$ in the first term of eq. (2.11) when $t > 1$ gives (2.6) and (2.7). □

Note that when $b \to \infty$ in eq. (2.7), the result is the same that in eq. (2.2) of the Proposition 2.1.

The distribution of $W = \frac{X}{X+Y}$, useful in hydrology, is given in the following two propositions. Here, we denoted the survival copula by $\hat{C}$, defined as

$$\hat{C}(\bar{F}_X(x), \bar{F}_Y(y)) = P\left(X > x, Y > y\right),$$

where

$$\begin{aligned} \hat{C}(1-u, 1-v) &= \bar{C}(u,v) \\ &= 1 - u - v + C(u,v) \\ &= P(U > u, V > v). \end{aligned} \tag{2.12}$$

Proposition 2.3. Let (X, Y) be a vector of random variables defined on the same probability space (R^+, F, P), the associated copula C and the continuous marginal distribution functions F_X and F_Y. Then the distribution of $W = \frac{X}{X+Y}$ is given by

$$F_W(t) = \int_0^1 D_1 \hat{C}\left(1-u, \bar{F}_X\left(\frac{1-t}{t}\bar{F}_Y^{-1}(1-u)\right)\right) du. \tag{2.13}$$

Demostración: Consider the joint cumulative distribution function of X and W given by

$$\begin{aligned} F_{X,W}(s,t) &= P\left(X \leq s, W \leq t\right) \\ &= \int_{-\infty}^s \int_{-\infty}^t f_{X,W}(x,w) dx dw. \end{aligned}$$

With conditioning on X, the cdf $F_{X,W}$ can be written as

$$\begin{aligned} F_{X,W}(s,t) &= \int_{-\infty}^s \int_{-\infty}^t f_{W|X}(w|x) f_X(x) dw dx \\ &= \int_{-\infty}^s F_{W|X}(t|x) f_X(x) dx \\ &= \int_0^s P\left(Y \geq x\frac{1-t}{t} | X = x\right) dF_X(x). \end{aligned} \tag{2.14}$$

Applying the limit $s \to \infty$ in eq. (2.14) gives

$$\begin{aligned} F_W(t) &= \int_0^\infty P\left(Y \geq x\frac{1-t}{t} | X = x\right) dF_X(x) \\ &= \int_0^\infty \left[1 - P\left(Y < x\frac{1-t}{t} | X = x\right)\right] dF_X(x). \end{aligned} \tag{2.15}$$

(2.16)

In eq. (2.16), by changing the variable $u = F_X(x)$ and when applying Lemma 1 we obtain

$$F_W(t) = 1 - \int_0^1 D_1C\left(u, F_Y\left(\frac{1-t}{t}F_X^{-1}(u)\right)\right)du. \tag{2.17}$$

To get (2.13) From (2.17), we use the survival copula. □

By applying the chain rule in eq. (2.13), we obtain its corresponding density function

$$f_W(t) = \int_0^1 c\left(u, F_Y\left(\frac{1-t}{t}F_X^{-1}(u)\right)\right) f_X\left(\frac{1-t}{t}F_X^{-1}(u)\right) F_X^{-1}(u)\frac{1}{t^2}du. \tag{2.18}$$

For random variables with support on a unbounded interval the distribution of W is given in Proposition 2.4.

Corollary 2.4. Let (X, Y) be a vector of random variables defined on the same probability space $([0, b], F, P)$ where $b \in R^+$, the associated copula C and the continuous marginal distribution functions F_X and F_Y. Then the distribution of $W = \frac{X}{X+Y}$ is given by

$$F_W(t) = \begin{cases} F_X(b\frac{1-t}{t}) - \int_0^{F_X(b\frac{1-t}{t})} D_1C\left(u, F_Y(\frac{1-t}{t}F_X^{-1}(u))\right)du & \text{if } t \le 1/2, \\ 1 - \int_0^1 D_1C\left(u, F_Y(\frac{1-t}{t}F_X^{-1}(u))\right)du & \text{if } t > 1/2. \end{cases} \tag{2.19}$$

Demostración: Let $f_{X,W}(x, w)$ be the joint density of X and W where $0 \le x \le b$ and $\frac{x}{x+b} \le w \le 1$. Integrating with respect to x and w and then conditioning on $X = x$, we have

$$\begin{aligned} F_{X,W}(s,t) &= \int_0^s \int_{x/(x+b)}^{k(t)} f_{W|X}(w|x)f_X(x)dwdx \\ &= \int_0^s F_{W|X}(k(t)|x)dF_X(x). \end{aligned} \tag{2.20}$$

Depending on the values of x and b, the interval of integration with respect to w is limited by $k(t) = \max\left\{\frac{x}{x+b}, t\right\}$. Applying the limit $s \to b$ in eq. (2.20) gives

$$\begin{aligned} F_W(t) &= \int_0^b P\left(W \le k(t)|X = x\right)dF_X(x) - \int_0^b P\left(W \le \frac{x}{x+b}|X = x\right)dF_X(x) \\ &= \int_0^b P\left(Y \ge x\frac{1-k(t)}{k(t)}|X = x\right)dF_X(x). \end{aligned}$$

Taking into account the cases $t \le 1/2$ and $t > 1/2$ leads to

$$F_W(t) = \begin{cases} \int_0^{b\frac{t}{1-t}} 1 - P\left(Y \le x\frac{1-t}{t}|X = x\right)dF_X(x) & \text{if } t \le 1/2, \\ \int_0^b 1 - P\left(Y \le x\frac{1-t}{t}|X = x\right)dF_X(x) & \text{if } t > 1/2. \end{cases}$$

Now, changing the variable $u = F_X(x)$ and using Lemma 1, we have

$$F_W(t) = \begin{cases} F_X(b\frac{1-t}{t}) - \int_0^{F_X(b\frac{1-t}{t})} D_1C\left(u, F_Y(\frac{1-t}{t}F_X^{-1}(u))\right)du & \text{if } t \le 1/2, \\ 1 - \int_0^1 D_1C\left(u, F_Y(\frac{1-t}{t}F_X^{-1}(u))\right)du & \text{if } t > 1/2. \end{cases}$$

Changing again $v = \left[F_X(b\frac{1-t}{t})\right]^{-1} u$ yields eq. (2.19).
□

The expressions of the cumulative distribution function of R, Z and W, given in equations (2.2),(2.3) and (2.13), respectively, are formulas that involve only one definite integral of the derivative of copula function. On the other hand, it is well known that the general expressions of the probability density functions of the product and ratio of two continuous random variables are expressed by (see Mood et al. (1974))

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx \tag{2.21}$$

and

$$f_R(r) = \int_{-\infty}^{+\infty} |x| f_{X,Y}(x, rx) dx. \tag{2.22}$$

Using the expressions (2.21) and (2.22) directly along with

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y),$$

where $c(\cdot,\cdot)$ is the copula density, one can obtain

$$f_Z(t) = \int_0^1 c\left(u, F_Y\left(t/F_X^{-1}(u)\right)\right) f_Y(t/F_X^{-1}(u)) \left(F_X^{-1}(u)\right)^{-1} du \tag{2.23}$$

and

$$f_R(t) = \int_0^1 c\left(u, F_Y\left(tF_X^{-1}(u)\right)\right) f_Y(tF_X^{-1}(u)) F_X^{-1}(u) du. \tag{2.24}$$

Applying the chain rule in eq. (2.2) and eq. (2.3), we also obtain the densities functions (2.23) and (2.24).

Now, integrating eq. (2.24) results in the expression of F_R that involves double integral, which is computationally more complex than eq. (2.2). However, with some algebraic manipulations, it is also possible to obtain eq. (2.2). The alternative proof of Proposition 2.1., involving the above density expression, can be found in Appendix. The product case, Z, is analogous.

2.2. Particular Cases. There are several types of dependence structures induced by copula functions. Here, we used a few of them to illustrate the expressions of $R = Y/X$ based on Proposition 2.1.. In the first case we consider the simplest one, the independence copula, and the remaining ones are usual copula functions.

Let C be the product copula, $C(u, v) = uv$. Then the distribution of R is expressed by

$$F_R(t) = \int_{-\infty}^{0} \bar{F}_Y(tx) dF_X(x) + \int_0^{\infty} F_Y(tx) dF_X(x).$$

(1.) The upper bound of Fréchet-Hoeffding, $C(u, v) = \min\{u, v\}$, yields

$$\begin{aligned} F_R(t) &= F_X(0) - \int_0^{F_X(0)} I_{\left\{u: \frac{F_Y^{-1}(u)}{F_X^{-1}(u)} < t\right\}}(u) du \\ &+ \int_{F_X(0)}^{1} I_{\left\{u: \frac{F_Y^{-1}(u)}{F_X^{-1}(u)} < t\right\}}(u) du \\ &= F_X(0) - \sup\left\{u \in (0, F_X(0)) : \frac{F_Y^{-1}(u)}{F_X^{-1}(u)} < t\right\} \\ &+ \sup\left\{u \in (F_X(0), 1) : \frac{F_Y^{-1}(u)}{F_X^{-1}(u)} < t\right\}. \end{aligned}$$

(2.) The Gaussian copula defined by

$$C_\rho(u, v) = \int_0^u \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(t)}{\sqrt{1-\rho^2}}\right) dt,$$

where Φ is the standard Gaussian cumulative distribution function, provides

$$\begin{aligned} F_R(t) = F_X(0) &- \int_0^{F_X(0)} \Phi\left(\frac{\Phi^{-1}\left(F_Y(tF_X^{-1}(u))\right) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) du \\ &+ \int_{F_X(0)}^{1} \Phi\left(\frac{\Phi^{-1}\left(F_Y(tF_X^{-1}(u))\right) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) du. \end{aligned}$$

TABLE 3.1
Descriptive statistics - annual peak river flow per season in $l/s/km^2$.

	Mean	Median	Maximum	Minimum	Std Dev	Skewness
Rainy	2314	2121	7210	658	1038.5	1.7
Dry	1814	1503	9044	479	1263.08	3.04

(3.) Let $C(u,v)$ be a Marshall-Olkin copula defined by

$$C(u,v) = vu^{1-n}I_{[u^m \geq v^n]} + uv^{1-n}I_{[u^m < v^n]},$$

where $m,n \in N$. Then

$$\begin{aligned} F_R(t) = F_X(0) - \int_0^{F_X(0)} & (1-m)u^{-m}F_Y\left(tF_X^{-1}(u)\right)I_{\left[u^m \geq F_Y^n\left(tF_X^{-1}(u)\right)\right]} \\ & + F_Y^{1-n}(tF_X^{-1}(u))I_{\left[u^m < F_Y^n\left(tF_X^{-1}(u)\right)\right]}du \\ & + \int_{F_X(0)}^1 (1-m)u^{-m}F_Y\left(tF_X^{-1}(u)\right)I_{\left[u^m \geq F_Y^n\left(tF_X^{-1}(u)\right)\right]} \\ & + F_Y^{1-n}\left(tF_X^{-1}(u)\right)I_{\left[u^m < F_Y^n\left(tF_X^{-1}(u)\right)\right]}du. \end{aligned}$$

(4.) With $C(u,v)$ being an Archimedean copula and ψ its generator function, defined by

$$C(u,v) = \psi^{-1}\left(\psi(u) + \psi(v)\right),$$

result

$$\begin{aligned} F_R(t) = F_X(0) - & \int_0^{F_X(0)} \frac{\psi'(u)}{\psi'\left(\psi^{-1}\left(\psi(u) + \psi\left(F_Y(tF_X^{-1}(u))\right)\right)\right)}du \\ & + \int_{F_X(0)}^1 \frac{\psi'(u)}{\psi'\left(\psi^{-1}\left(\psi(u) + \psi\left(F_Y(tF_X^{-1}(u))\right)\right)\right)}du, \end{aligned}$$

where ψ' is the first derivative of the function ψ.

3. Application. In this section, we illustrate an application using the previous results. In hydrology research, one may be interested in modeling the relation between maximum river flows in two different seasons, namely dry and rainy. This relation can be analyzed through the distribution of $R = Y/X$ and $W = X/(X+Y)$, where Y and X are the maximum river flows in the dry and rainy seasons, respectively. In this way, our goal is to assess the probability of the following events: $R > 1$ or equivalently $W < 1/2$. These events represent the risk of maximum river flow in the dry season exceeding the maximum of the rainy season. The dataset corresponds to the Paranapanema watershed in the south of Brazil. In particular, the measurements are obtained by hydrometric stations, located at the Rosana hydroelectric plant and are published by the Brazilian Power System Operator (ONS), http://ons.org.br/. The river flow is characterized by two main seasons: rainy (November to April) and dry (May to October) and it is measured in total liters per second per square kilometer (l/s/km^2).

The data regard the maximum river flow computed for each month, ranging from 1937 to 2008. The descriptive statistics of this dataset are shown in Table 1.

In order to obtain the distributions of R, we computed the results of (2.2). First, we estimated the parameters of the appropriate copula for the data and then we calculated R via numerical integration. Copula parameter estimations by maximum likelihood were obtained using the inference Function for Margins (IFM) method, described by Harry (1997). This method consists of estimating the marginal parameters by maximum likelihood in a first step and then estimating the association parameter in a second step. All numerical issues treated in this section were computed via R programming language.

The data fit distribution of the maximum flow in the rainy season and the data fit distribution of the maximum flow in the dry season are the marginal distributions of the bivariate model. Here, these marginal distributions are Generalized Extreme Value (GEV) distributions.

The Generalized Extreme Value (GEV) distribution, F, is the limit distribution of the normalized maximum (or minimum) of a sequence of independent and identically distributed (i.i.d.) random variables. That is, if $\{X_n\}$ is such sequence and there are sequences of real numbers $\{c_n\}$, $c_n > 0$ and $\{d_n\}$ such that

$$\lim_{n\to\infty} P\left(\frac{\max\{X_1, X_2, ..., X_n\} - d_n}{c_n} \leq x\right) = F(x),$$

then the GEV distribution is defined by

$$F(x;\mu,\sigma,\xi)=\begin{cases}\exp\left\{-\left[1+\xi(\frac{x-\mu}{\sigma})\right]^{-1/\xi}\right\} & \text{if } \xi\neq 0,\\ \exp\left\{-\exp[-(x-\mu)/\sigma]\right\} & \text{if } \xi=0,\end{cases} \tag{3.1}$$

where $\mu \in R$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in R$ is the shape parameter. The last parameter determines the tail behavior of the distribution. Setting the shape parameter as $\xi = 0$, $\xi > 0$ and $\xi < 0$ defines the Gumbel, Fréchet and Weibull distribution families, respectively.

Since the data used in this work correspond to the monthly maximum flows of a river, we adjusted the subsample of block maximums of size 6 by a GEV distribution. initially, we applied the Ljung-Box test, Ljung and Box (1978), to verify the null hypothesis of serial independence of the subsample of block maximums. The test statistics did not reject the null hypothesis of either series. The test statistics are summarized in Table 3.2. Additionally, the autocorrelation function and partial autocorrelation function with 0.95 confidence bounds are displayed in Fig 3.1.

TABLE 3.2
Ljung-Box test - peak river flow series.

Series	Ljung-Box Test	d.f.	p-Value
Rainy	2.1103	1	0.1463
Dry	1.0499	1	0.3055

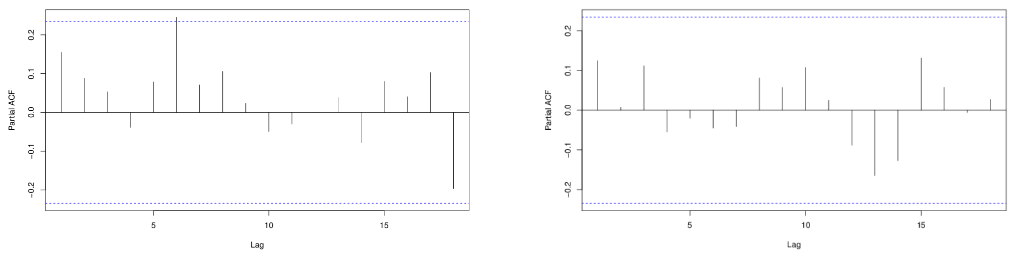

FIGURE 3.1. *Partial autocorrelation function of subsamples of block maximums of size 6 in the rainy season (left) and the dry season (right) with* 0.95% *confidence intervals*

The maximization of GEV log-likelihood leads to the estimate $(\hat{\mu},\ \hat{\sigma},\ \hat{\xi}) =$ $(1863.61, 760.17, 0.005)$ for rainy season and $(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (1328.98, 552.95, 0.006)$ for dry season with log-likelihood values of -640.2577 and -554.31 respectively. It is noteworthy that when $\xi > -0.5$ the maximum likelihood estimators are regular, that is, the asymptotics properties of the MLE hold, Smith (1985). Incorporating estimate and standard error, the approximate 95% confidence intervals of the parameter ξ are $[-0.083, 0.183]$ and $[-0.075, 0.212]$ for rainy and dry respectively. Thus, the estimate of ψ can be considered zero in both cases.

The corresponding GEV family with $\xi = 0$ is the Gumbel distribution. The MLE, in the Gumbel case, leads to log-likelihood values of -640.4542 and -554.6957 for rainy and dry season. The likelihood ratio test statistics are $D_{rainy} = 2\,[-640.2577 - (-640.4542)] = 0.393$ and $D_{dry} = 2[-554.3125]$- $2[(-554.6957) = 0.7664$. The values of these statistics are small when we consider the χ^2_1 distribution, which suggests that the Gumbel distribution is adequate for these data.

Using the Gumbel distribution, the estimation results, standar errors and p-values are given in Table 3.3.

TABLE 3.3
Gumbel parameters estimates. The symbol *= 10^{-3}.

	Estimate	Std. Error	p-Value
		Rainy	
μ	1876.05	89.44	0.00*
σ	751.47	66.87	0.00*
		Dry	
μ	1365.07	73.36	0.00*
σ	590.81	56.4	0.00*

In order to check the goodness-of-fit, we employed the P-P plot in Figure 3.2 as suggested by Stuart (2001), for GEV family distributions. These curves indicate that assessed models are in agreement with the empirical distribution.

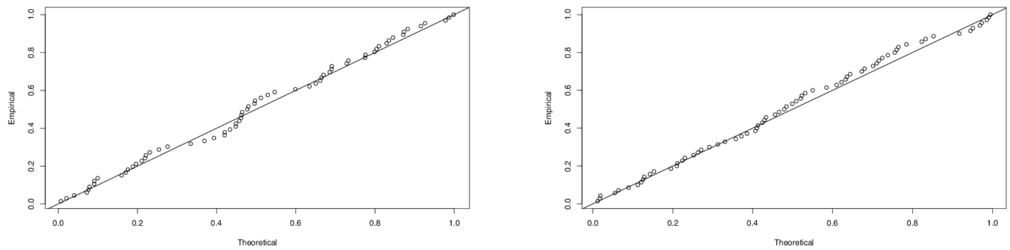

FIGURE 3.2. *Gumbel P-P Plot. Rainy (first) and dry (second).*

Note that the marginal distributions have support on the real line, so Proposition 2.1 is required for calculate the distribution of $R = Y/X$.

The Gumbel expected value and variance are given by $\mu + e\sigma$ and $\frac{\sigma^2\pi^2}{6}$ respectively. Due to invariant properties of MLE, estimated expected value and variance are 3918.757 and 928906 in the rainy season and 2971.06 and 365531.1 in the dry season. As we expected, the mean of rainy season is greater than that of the dry season. However, the rainy season has larger variability.

To check the existence of dependence between rainy and dry season data, we applied Kendall's tau test. The test statistic rejects the null hypothesis of independence, given by the resulting p-value. After estimating the distribution of X and Y and testing the dependence between them, the next step is to model the dependence structure. Based on Clarke's test, suggested by Clarke(2007), we selected the Clayton copula. The test compares two competing models and selects which one is preferred. This procedure is done for all pair combinations in the list of possible models. If a model is preferred over another the favored model receives score $+1$ and in similar way -1 for another model. No score is computed if Clarke's test cannot distinguish the favored model. This selection method is available in the CDVine R package; for more details see Brechmann and Schepsmeier (2013). The test result is described in Table 3.4.

TABLE 3.4
Clarke copula selection and Kendall's tau independence test.

Gaussian	Student-t	Clayton	Gumbel	Frank	Joe
0	2	3	-2	2	-5
Kendall's tau test	Statistic	p-value			
	2.85	0.004			

The parameter of the Clayton copula, $\theta \in (0, \infty)$, was also estimated by MLE via the IFM method, using the generalized inverse function of the Gumbel marginal distributions given in Table 3.3. Table 3.5 gives the results of copula parameter estimation. The parameter θ was statistically significant with p-value 0.0049. Figure 3.3 combines scatterplot with contour lines of estimated Clayton copula with corresponding Gumbel marginals.

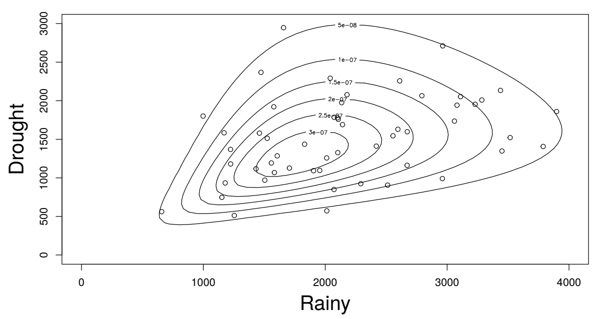

FIGURE 3.3. *Scatterplot of seasonal peak river flows and contour lines of estimated Clayton density copula.*

Through the relation $\tau = \frac{\theta}{\theta+2}$, we obtain $\tau = 0.22$, where τ is the Kendall correlation. As the raw data

TABLE 3.5
Clayton copula parameter estimate.

	θ	τ-Kendall	λ_L
Estimate	0.5614	0.22	0.29

TABLE 3.6
Computed moments of R and W.

	$E[\cdot]$	$V(\cdot)$	$P(Y > X)$
R	0.8082	0.422	0.2418

suggest, there is a positive correlation between seasons. Furthermore, the Clayton copula exhibits a lower tail dependence, which indicates a strict relation between low peaks. The coefficient is given by the relation $\lambda_L = 2^{-1/\theta}$ and yields an estimated measure of lower tail dependence.

Up to this point, we obtained the marginal distributions and the copula function, so using Propositions 2.1. and 2.3., we are able to establish the distribution of R. We plotted the probability density functions of R in Fig 3.4. Moreover, we added a kernel density estimation in these graphs. The agreement is quite good, since they behave in a similar way.

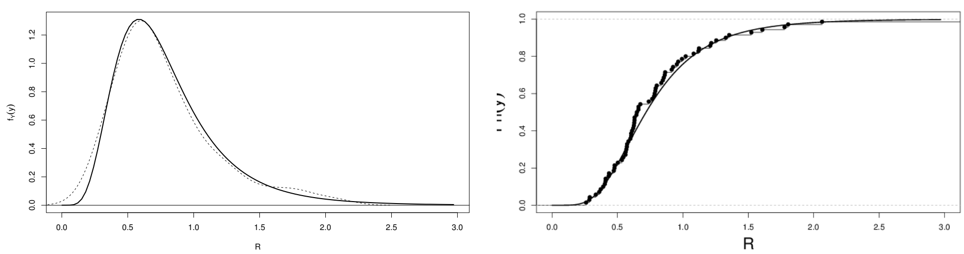

FIGURE 3.4. *Left: Densities of R, fitted (solid) and empirical (dashed); Rright: Distributions of R: fitted (solid) and empirical (dotted)..*

The empirical distribution as well as the estimated cumulative distribution function for both R and W random variables are shown in Figure **??**. Via numerical integration of equations (2.24) and (2.18), we computed the expectation and variance of R and W based on estimated densities. These values are exhibited in Table 3.6. On the other hand, Table 3.7 provides the same estimates, but obtained directly from the observed data. The last column of Table 3.7 refers to the empirical probability, $\tilde{p}$, of annual peak of river flow in the rainy season being lower than in the dry season.

Now, we are ready to calculate the risk of the peak river flow in the dry season being higher than in the rainy season in a certain year. This risk corresponds to the value $P(R > 1)$. The estimated cumulative function and the empirical cumulative function are close.

4. Conclusion. In this paper, expressions of the cumulative distribution function of the product and ratio for two continuous and discrete dependent random variables are obtained. We considered random variables with support on a unbounded and bounded interval. An application of the cumulative distribution function of the product and ratio to compute the stress-strength function to real data is added. For future work we are considering the inclusion of the hypothesis of the temporal dependence of real data to obtain expressions of the cumulative distribution function of the product and ratio for two dependent random variables.

5. Acknowledgements. We thank the anonymous referees for very useful remarks to improve the earlier version. Yuri S. Maluf was partially supported by CNPq-Brazil.

TABLE 3.7
Descriptive statistics of observed ratios.

	Sample Mean	Sample Var	$\tilde{p}(y > x)$
r	0.8321	0.3578	0.2142

ORCID and License

Cira Otiniano http://orcid.org/0000-0002-5619-0478,
Yuri Maluf https://orcid.org/0000-0001-5063-0113.

References

[1] Ali, M. M., Pal, M., Woo,J. *On the ratio of inverted gamma variates*, Austrian Journal of Statistics. 2007; 36(2):153-159.
[2] Brechmann, E. C., Schepsmeier, U. *Modeling dependence with c- and d-vine copulas: The R package CDVine*, Journal of Statistical Software. 2013: 52(3):1-27.
[3] Cherubini, U., Mulinacci, S. and Romagnoli, S. *On the distribution of the (un)bounded sum of random variables*, Insurance: Mathematics and Economics, 2011; 48:56-63.
[4] Clarke, K. A., *A simple distribution-free test for nonnested model selection*, Political Analysis, 2007; 15:347-363.
[5] Dolati,V., Roozegar,R., Ahmadi, N., Shishebor, Z., *The effect of dependence on distribution of the functions of random variables*, Communications in Statistics - Theory and Methods, 2007; 46:10704-10717.
[6] Domma, F., Giordano, S., *A stress-strength model with dependent variables to measure household financial fragility*, Statistical Methods & Applications, 2012; 21:375-389.
[7] Domma, F., Giordano, S., *A copula-based approach to account for dependence in stress-strength models*, Statistical Papers, 2013; 54:807-826.
[8] Gupta, A. K., Nadarajah, S., *Sums, products and ratios for McKay's bivariate gamma distributions*, Mathematical and Computer Modelling, 2006; 43:185-193.
[9] Harry, J., *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC, 1997.
[10] Idrizi, L., *On the product and ratio of Pareto and Kumaraswamy random variable*, Mathematical Theory and Modeling, 2014; 4(3):136-146.
[11] Ljung, G. M., Box, G. E. P., *On a measure of lack of fit in time series models*, Biometrika, 1978; 65:297-303.
[12] Mood, V., Graybill,F., Boes, D., *Introduction to the Theory of Statistics*. McGraw Hill Inc, 1974.
[13] Nadarajah, S., Ali, M. M., *Pareto random variables for hydrological modeling*, Water Resources Management, 2008; 22:1381-1393.
[14] Nadarajah, S., *Sums, products and ratios for the bivariate Gumbel distribution*, Mathematical and Computer Modelling, 2005; 42:499-518.
[15] Nadarajah, S., *Sums, products and ratios for the bivariate Lomax distribution*, Computational Statistics & Data Analysis, 2005; 49:109-129.
[16] Nadarajah, S., *Products and ratios for a bivariate gamma distribution*, Applied Math ematics and Computation, 2005; 171:581-595.
[17] Nadarajah, S., Ali, M. M., The distribution of sums, products and ratios for Lawrance and Lewis's bivariate exponential random variables, Computational Statistics & Data Analysis, 2006; 50:3449-3463.
[18] Nadarajah, S., Dey, D. K., *On the product and ratio of t random variables*, Applied Mathematics Letters, 2006; 19:45-55.
[19] Nadarajah, S., Gupta, A. K., *Cherian's bivariate gamma distribution as a model for drought data*, Agrociencia, 2006; 40:483-490.
[20] Nadarajah, S., Kotz, S. , *Sums, products, and ratios for downton's bivariate exponential distribution*, Stochastic Environmental Research & Risk Assessment, 2006; 20:164-170.
[21] Nadarajah, S., *The bivariate gamma exponential distribution with application to drought data*, Journal of Applied Mathematics and Computing, 2007; 24:221-230.
[22] Nagar,D. A., Orozco-Castaneda, J. M., Gupta, A. K., *Product and quotient of correlated beta variables*, Applied Mathematics Letters, 2009; 22:105-109.
[23] Nelsen, R. B., *An Introduction to Copulas*. Springer, 2006.
[24] Pham-Gia, T., *Distributions of the ratios of independent beta variables and applications*, Communications in Statistics - Theory and Methods, 2000; 29(12):2693-2715.
[25] Shakil, M., Golam Kibria, B. M., *Exact distribution of the ratio of gamma and Rayleigh random variables*, Pakistan Journal of Statistics and Operation Research, 2006; 2(2):87-98.
[26] Smith,R. L. , *Maximum likelihood estimation in a class of non-regular cases*, Biometrika, 1985; 72:67-90.
[27] Stuart, C., *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
[28] Trivedi,P. K., Zimmer, D. M., *Copulas Modeling: An Introduction for Practitioners*. Foundations and Trends in Econometrics, 2005.