The R-code for computing the cdf and the df o f the ratio of two correlated Normal rv

LUCIO DE CAPITANI

Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca. <u>lucio.decapitani1@unimib.it</u> ANGIOLA POLLASTRI

Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano-Bicocca. <u>angiola.pollastri@unimib.it</u>

Abstract

The cumulative distribution and the density function of the ratio of two correlated Normal random variables has been proposed and studied by Aroian (1986) and Oksoy and Aroian (1986, 1994) and Kim (2006). For a complete review and further insights on this topic, the interested reader is referred to Galeone (2007). In Galeone and Pollastri (2012a, 2013) the quantiles of this random are used to build confidence intervals for the ratio of two means and for the incremental cost-effectiveness ratio. Pollastri and Tulli (2012) have studied the distribution of absolute ratio of two correlated Normal random variables and used it to define the confidence interval for non-negative ratio indicators. In order to ease the computation of the cumulative distribution function (cdf), the density function (df), and of the quantiles of these ratio random variables, in the present note the R-codes are presented and exemplified.

Keywords: Ratio and absolute ratio of two correlated Normal rv, Cumulative Distribution Function, Density Function, Quantiles, R-code.

1. The distribution of the ratio of two correlated Normal random variables

Let us consider a Bivariate Correlated Normal (BCN) r.v.

$$(Y, X) \sim N(\mu_Y, \mu_X, \sigma_Y^2, \sigma_X^2, \rho).$$

In Oksoy and Aroian (1986) it is shown that the cumulative distribution function (cdf) of the r.v. $W = \frac{Y}{X}$ is given by:

$$F_{W}(w) = L\left(\frac{a - b t_{w}}{\sqrt{1 + t_{w}^{2}}}, -b , \frac{t_{w}}{\sqrt{1 + t_{w}^{2}}}\right) + L\left(\frac{b t_{w} - a}{\sqrt{1 + t_{w}^{2}}}, b , \frac{t_{w}}{\sqrt{1 + t_{w}^{2}}}\right),$$
(1)

where $w \in \Re$, $L(h, k, \rho)$ is the bivariate normal integral according to the indication of Kotz *et al.* (2000), and

$$a = \sqrt{\frac{1}{1 - \rho^2}} \left(\frac{\mu_Y}{\sigma_Y} - \rho \frac{\mu_X}{\sigma_X} \right), \qquad b = \left(\frac{\mu_X}{\sigma_X} \right), \qquad t_w = \sqrt{\frac{1}{1 - \rho^2}} \left(\frac{\sigma_X}{\sigma_Y} w - \rho \right). \tag{2}$$

An alternative expression for $F_W(w)$ involving the V(h,q) function of Nicholson (1943) has been derived in Oksoy and Aroian (1994) and it is given by

$$F_{W}(w) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(t_{w}\right) + 2V \left\{\frac{b \ t_{w} - a}{\sqrt{1 + t_{w}^{2}}}, \frac{b + a \ t_{w}}{\sqrt{1 + t_{w}^{2}}}\right\} - 2V(b, a)$$
(3)

where $V(h,q) = \int_0^h \int_0^y \Phi(x)\Phi(y)dxdy$ and $y = \frac{q}{h}x$.

Starting from expression (3) and remembering the definition of the Owen T function (Owen, 1956)

$$T(h,\lambda) = \frac{1}{2\pi} \arctan \lambda - V(h,\lambda h), \qquad (4)$$

Pollastri and Tulli (2012) derived the following alternative expression for $F_W(w)$:

$$F_{W}(w) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(t_{w}\right) + \frac{1}{\pi} \arctan\left(\frac{b + a t_{w}}{b t_{w} - a}\right) + \frac{1}{2\pi} \operatorname{arctan}\left(\frac{b t_{w} - a}{\sqrt{1 + t_{w}^{2}}}, \frac{b + a t_{w}}{b t_{w} - a}\right) - \frac{1}{\pi} \operatorname{arctan}\left(\frac{a}{b}\right) + 2T\left(b, \frac{a}{b}\right)^{2}.$$
(5)

In Aroian (1986) the expression for the density of W is also given:

$$f_{W}(w) = \frac{\sigma_{2}}{\sigma_{1}} \frac{1}{\sqrt{1 - \rho^{2}}} f(t_{w})$$
(6)

with

$$f(x) = \frac{1}{\pi(1+x^2)} e^{-\frac{1}{2}(a^2+b^2)} \left[1 + \frac{q}{\varphi(q)} \int_0^q \varphi(s) ds \right], \qquad q = \frac{b+a t_w}{\sqrt{1+t_w^2}}, \qquad \varphi(q) = \frac{1}{\sqrt{2\pi}} e^{-\frac{q^2}{2}}, \text{ and } a, b,$$

 t_w defined as in (2).

2. The distribution of the absolute ratio of two correlated Normal random variables

The cdf of the r.v. $V = \left| \frac{Y}{X} \right|$ can be easily obtained from expressions (1) and (5) using the following realtion:

$$F_{v}(v) = F_{w}(v) - F_{w}(-v) \text{ with } v \ge 0.$$
 (7)

(9)

In particular, starting from expression (5), Pollastri-Tulli (2012) introduced the following formula:

$$F_{V}(v) = \frac{1}{\pi} \left[\arctan(t_{v}) - \arctan(t_{-v}) \right] + \frac{1}{\pi} \left[\arctan\left(\frac{b+a t_{v}}{b t_{v}-a}\right) - \arctan\left(\frac{b+a t_{-v}}{b t_{-v}-a}\right) \right] + 2T \left(\frac{b t_{v}-a}{\sqrt{1+t_{v}^{2}}}, \frac{b+a t_{v}}{b t_{v}-a}\right) + 2T \left(\frac{b t_{-v}-a}{\sqrt{1+t_{-v}^{2}}}, \frac{b+a t_{-v}}{b t_{-v}-a}\right) \right]$$
(8)

Note that, if $t_v t_{-v} > -1$ and $\left(\frac{b+a}{b} \frac{t_v}{t_v - a}\right) \left(\frac{b+a}{b} \frac{t_{-v}}{t_{-v} - a}\right) > -1$, then expression (7) coincides with: $F_V(v) = 2T \left(\frac{b}{\sqrt{1+t_{-v}^2}}, \frac{b+a}{b} \frac{t_{-v}}{t_{-v} - a}\right) - 2T \left(\frac{b}{\sqrt{1+t_v^2}}, \frac{b+a}{b} \frac{t_v}{t_v - a}\right).$

Concerning the density function of V, it can be easily obtained from (6) observing that $f_V(v) = f_W(v) + f_W(-v)$, v > 0.

3. Description of the R-code for computing the cdf and the quantiles of $F_W(w)$ and $F_V(w)$ using the bivariate normal integral

In order to compute the distribution function $F_w(w)$ using expression (1), the R-package mvtnorm is very useful. Indeed, this package contains the function "pmvnorm" which computes the bivariate normal integral according to the indication of Kotz *et al.* (2000). The information regarding the use of this function can be taken from the Help of the mvtnorm package. The R-function for computing $F_w(w)$ using expression (1) is then:

 $F_WI \leftarrow function(w, mu_X, mu_Y, sigma_X, sigma_Y, rho)$ $\begin{cases}
 #definition of the quantities in expression (2) \\
 a <- (1/(1-rho^2))^0.5 * (mu_Y/sigma_Y-rho*mu_X/sigma_X) \\
 b <- mu_X/sigma_X \\
 t_w <- (1/(1-rho^2))^0.5*(sigma_X/sigma_Y*w-rho) \\
 #introduction of auxialiry variables$ $A <- (a-b*t_w)/((1+t_w^2)^0.5) \\
 B <- b \\
 C <- t_w/((1+t_w^2)^0.5) \\
 \end{cases}$

#definition of the two addends in expression (1)
Mcor<-matrix(c(1,C,C,1),2,2)
A1<-pmvnorm(lower=c(A,B),upper=Inf, corr=Mcor)</pre>

A2<-pmvnorm(lower=c(-A,-B),upper=Inf, corr=Mcor)

```
#computation of the cdf
res<-A1+A2
return(res)
}</pre>
```

The arguments required in the above function are named coherently with the notation introduced in the previous sections.

The quantile function can now be defined by numerical inversion. The inversion can be made by means of the uniroot function in the R-package "base". The reader is referred to the official Help of the base R-package for the details on the uniroot function. In the function below, q denotes the desired quantile level.

Analogously, the following functions for computing the cdf $F_V(v)$ and its quantiles can be written.

4. Example of application of the functions F_WI, Quantile_WI, F_VI, and Quantile_VI

Let us consider the Bivariate Normal random variable (X, Y) with $\mu_X = 0.8$, $\mu_Y = 0.4$, $\sigma_X = \sigma_Y = 0.3$, and $\rho = 0.7$.

We want to plot the cdf, the df and the quantile functions of the random variables $W = \frac{Y}{V}$ and

$$V = \left| \frac{Y}{X} \right|.$$

Below, we report the R-code for obtaining these graphs and their results (the comments to the code are reported in dark-gray).

library(mvtnorm)

#specifying the parameters of the Bivariate Normal Distribution mu X <- 0.8 <- 0.4 mu Y $sigma_X < sigma_Y < 0.3$ rho <- 0.7 #Graph of the cdfs #defining the grid of values or which the cdfs are evaluated step <- 0.01 lower <- -1.5 upper <-3x<-seq(lower,upper,step) #computing the cdf of W over the grid x YI<-sapply(x, function(w) F_WI(w,mu_X,mu_Y,sigma_X,sigma_Y,rho)) #computing the cdf of V over the grid x Yassl<-sapply(x[which(x>0)], function(v) F VI(v,mu X,mu Y,sigma X,sigma Y,rho)) #plotting the results plot(x, YI,type="l",col="green", main="CDF") points(x[which(x>0)], Yassl,type="l",col="red") #Graph of the quantile functions #defining the grid of values over which the guantile functions are evaluated step <- 0.01 lower <- 0.01 upper <- 0.99 p<-seq(lower,upper,step)</pre> #computing the quantile function of W over the grid p Ql<-sapply(p, function(w) Quantile Wl(w,mu X,mu Y,sigma X,sigma Y,rho) \$root) #computing the cdf of V over the grid p Qassl<-sapply(p, function(v) Quantile_VI(v,mu_X,mu_Y,sigma_X,sigma_Y,rho) \$root) #plotting the results plot(p, QI,type="l",col="green", main="Quantile functions") points(p, Qassl,type="l",col="red") #Graph of the density functions #defining the grid of values over which the density functions are evaluated Xden <-c((x[2:length(x)]+x[1:(length(x)-1)])/2)xx < -x[which(x>0)]Xdenass<-c((xx[2:length(xx)]+xx[1:(length(xx)-1)])/2) #approximating the density of W Ylden<-c(Yl[2:length(Yl)]-Yl[1:(length(Yl)-1)])/step #approximating the density of V YassIden<-c(YassI[2:length(YassI)]-YassI[1:(length(YassI)-1)])/step #plotting the results plot(Xden,Ylden, type="l", col="green", main="density functions") points(Xdenass, YassIden,type="l", col="red")

The graphs obtained from the above example are given below.









5. Description of the R-code for computing the cdf and the quantiles of $F_W(w)$ and $F_V(w)$ using the Owen *T* function along with the Pollastri and Tulli (2012) representation.

In order to compute the distribution function $F_w(w)$ as in expression (5), the R-package Sn is very useful. Indeed, this package contains the function "T.Owen" which computes the Owen *T*. The information regarding the use of this function can be taken from the Help of the Sn package. Also this approach to compute $F_w(w)$ turns out to be very efficient even if it requires to treat the scenario $(\mu_x, \mu_y) = (0,0)$ as a special case. In particular, when $(\mu_x, \mu_y) = (0,0)$ some of the quantities in expression (5) are not defined since they coincides with $\frac{0}{0}$. The case $(\mu_x, \mu_y) = (0,0)$ is then treated separately by approximating the null expectations with a very small positive number. The R-function for computing $F_w(w)$ using expression (5) is then:

```
F_W<-function(w,mu_X,mu_Y,sigma_X,sigma_Y,rho)
{
    #if (mu_X, mu_Y)=(0,0) choose a very small value for mu_X and mu_Y
    if(mu_X==0 & mu_Y==0){
        mu_X<-0.0000000001
        mu_Y<-0.0000000001
        }
    #definition of the quantities in expression (2)
        a <- (1/(1-rho^2))^0.5*(mu_Y/sigma_Y-rho*mu_X/sigma_X)
        b <- mu_X/sigma_X
        t_w <- (1/(1-rho^2))^0.5*(sigma_X/sigma_Y*w-rho)
```

```
#first addend in expression (5)
A <- 0.5+1/pi^*atan(t_w)
#second addend in expression (5)
B <- 1/pi*atan((b+a*t_w)/(b*t_w-a))
#auxiliary variables
      <- (b*t_w-a)/((1+t_w^2)^0.5)
h1
      <- (b+a*t_w)/(b*t_w-a)
h2
#third and fourth addend in expression (5)
C <- -2*T.Owen(h1, h2, jmax=170, cut.point=10)
D <- -1/pi^* atan(a/b)+2^*T.Owen(b, a/b, jmax=170, cut.point=10)
      <- A+B+C+D
res
return(res)
}
```

The quantile function can now be derived by numerical inversion using, again, the uniroot function.

Finally, the following functions for computing the cdf $F_V(v)$ and its quantiles can be written.

6. Example of application of the functions F_W, Quantile_W, F_V, and Quantile_V

}

Let us consider the Bivariate Normal random variable (X,Y) with $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$, and $\rho = 0$.

-q,

We are interested in plotting the cdf, the df and the quantile functions of the random variables $W = \frac{Y}{X}$ and $V = \left|\frac{Y}{X}\right|$. Note that X and Y are two independent standard Normal and, consequently, W follows the Cauchy distribution. Moreover, note that the Cauchy distribution is symmetric around 0 and, consequently V is a "half-Cauchy" random variable whose density is given by twice

the density of a Cauchy random variable over the positive real numbers and zero otherwise. This particular example has been chosen in order to evaluate the impact of the approximation of the values of μ_x and μ_y employed in the functions F_W and F_V to avoid the case $\mu_x = \mu_y = 0$.

Below, we report the R-code for obtaining the graphs. A comparison with the Cauchy distribution is also made in order to show that the approximation of the values of μ_x and μ_y employed to avoid the case $\mu_x = \mu_y = 0$ has a negligible effect.

#loading the package sn library(sn) #specifying the parameters of the Bivariate Normal Distribution mu X <- 0 mu Y <- 0 sigma_X <- sigma_Y <- 1 rho <- 0 #Graph of the cdfs #defining the grid of values or which the cdf is evaluated <- 0.01 step lower <- -1.5 upper <-1.5x<-seq(lower,upper,step) #computing the cdf of W over the grid x Y<-sapply(x, function(w) F_W(w,mu_X,mu_Y,sigma_X,sigma_Y,rho)) #computing the cdf of V over the grid x Yass<-sapply(x[which(x>0)], function(v) F_V(v,mu_X,mu_Y,sigma_X,sigma_Y,rho)) #plotting the results plot(x, Y,type="l",col="green", main="CDF") points(x[which(x>0)], Yass,type="l",col="red") #Graph of the quantile functions #defining the grid of values over which the quantile functions are evaluated step <- 0.01 lower <- 0.01 upper <- 0.99 p<-seq(lower,upper,step) #computing the quantile function of W over the grid p Q<-sapply(p, function(w) Quantile_W(w,mu_X,mu_Y,sigma_X,sigma_Y,rho) \$root) #computing the cdf of V over the grid x Qass<-sapply(p, function(v) Quantile_V(v,mu_X,mu_Y,sigma_X,sigma_Y,rho) \$root) #plotting the results plot(p, Q,type="l",col="green", main="Quantile functions") points(p, Qass,type="l",col="red") #Graph of the density functions #defining the grid of values over which the density functions are evaluated Xden < -c((x[2:length(x)]+x[1:(length(x)-1)])/2)xx < -x[which(x>0)]X denass < -c((xx[2:length(xx)]+xx[1:(length(xx)-1)])/2)#approximating the density of W Yden<-c(Y[2:length(Y)]-Y[1:(length(Y)-1)])/step #approximating the density of V Yassden<-c(Yass[2:length(Yass)]-Yass[1:(length(Yass)-1)])/step #plotting the results plot(Xden,Yden, type="l", col="green", main="density functions")

points(Xdenass, Yassden,type="l", col="red") #plotting the density functions of the Cauchy and "half-Cauchy" random variables in order to evaluate the effect of the approximation of mu_X and mu_Y s <-seq(1,length(Xden),10) points(Xden[s],dcauchy(Xden[s]),type="p", col="green") s<-seq(1,length(Xdenass),10) points(Xdenass[s],2*dcauchy(Xdenass[s]),type="p",col="red")</pre>

The graphs obtained from the above example are given below. In the plot of the density functions, the "circles" represent the theoretical values of the density of the Cauchy and "half-Cauchy" distribution. This graph show that the effect of the approximation of μ_x and μ_y is negligible.



density functions



Quantile functions



7 Description of the R-code for computing the density functions of W and V.

In order to compute the density functions $f_W(w)$ and $f_V(w)$ the expressions (6) and (9) can by adopted. The resulting functions are:

```
#Density function of W
f W<-function(w,mu X,mu Y,sigma X,sigma Y,rho){
    a<-(1/(1-rho^2))^0.5*(mu_Y/sigma_Y-rho*mu_X/sigma_X)
    b<-mu X/sigma X
    t w<-(1/(1-rho^2))^0.5*(sigma X/sigma Y*w-rho)
    q<-(b+a*t_w)/(1+t_w^2)^0.5
    phi<-dnorm(q)
    Phi<-pnorm(q)-0.5
    #density computation
    f<-1/(pi*(1+t_w^2))*exp(-0.5*(a^2+b^2))*(1+q/phi*Phi)
    den<-(sigma_X/sigma_Y)*(1/(1-rho^2)^0.5)*f
return(den)
}
#Density function of V
f V<-function(v,mu X,mu Y,sigma X,sigma Y,rho){
     if (v < 0) den < -0 else
      A<-f W(v,mu X,mu Y,sigma X,sigma Y,rho)
     B<-f_W(v,mu_X,mu_Y,sigma_X,sigma_Y,rho)
     den<-A+B
return(den)
}
```

8. Example of application of the functions f_W and f_V.

Let us consider the Bivariate Normal random variable (X, Y) with $\mu_X = 0.8$, $\mu_Y = 0.4$, $\sigma_X = \sigma_Y = 0.3$, and $\rho = 0.7$.

We are interested in plotting the df of the random variables $W = \frac{Y}{X}$ and $V = \left|\frac{Y}{X}\right|$.

Below, we report the R-code for obtaining these graphs. The graph are not included in the note since they are identical to those presented in Section 4.

```
#specifying the parameters of the Bivariate Normal Distribution
mu_X
         <- 0.8
         <- 0.4
mu Y
sigma_X <- sigma_Y <- 0.3
        <- 0.7
rho
#Graph of the dfs
#defining the grid of values or which the cdfs are evaluated
step
     <- 0.01
lower <- -1.5
upper <-3
x<-seq(lower,upper,step)
#computing the cdf of W over the grid x
Y<-sapply(x, function(w) f W(w,mu X,mu Y,sigma X,sigma Y,rho))
```

9. Conclusions

The functions described in this short note allow to evaluate the cdf, the df, and the quantiles of the ratio (W) and of the absolute ratio (V) of two correlated Normal random variables. As shown in the examples of Section 4 and Section 6, the functions for the cdf-computation can also be used to evaluate the density of W and V by approximating the derivative of the cdf with a proper difference quotient. The results obtained with this procedure are substantially identical to those obtained using the functions f_W and f_V for the density of W and V provided in Section 7 and exemplified in Section 8. Two alternative versions of the functions for computing the cdf and the quantiles are presented. The two versions stem from two different representations of the cdf of W and V. The first representation is due to Oksoy and Aroian (1986) and involves the bivariate normal integral. The functions based on this representation are: F_WI, QUANTILE_WI (computation of cdf and quantiles of W), F_VI, QUANTILE_VI (computation of cdf and quantiles of V). The second representation is due to Pollastri and Tulli (2012) and involves the arctangent function and the Owen T function (see Owen, 1956). The functions based on this representation are: F_W , QUANTILE_W (computation of cdf and quantiles of W), F_V, QUANTILE_V (computation of cdf and quantiles of V). Their implementation has been greatly facilitated by the existence of the functions pmvnorm and T.Owen contained in the R-packages mvtnorm and sn, respectively. Obviously, the results related to the two different representations are identical and two approaches resulted comparable form the computational point of view. In order to ease the usage of the functions here presented, the R-code is available upon request from the authors.

REFERENCES

AROIAN, L.A., (1986). The distribution of the quotient of two correlated random variables. *Proceedings of the Am. Stat. Ass. Business and Economic Section.*

GALEONE, C., (2007). On the ratio of two Normal r.v., jointly distributed as a Bivariate Normal. *PhD Thesis, Università di Milano-Bicocca*.

GALEONE, C., POLLASTRI, A., (2012). Confidence intervals for the ratio of two means using the distribution of the quotient of two normals. Forthcoming on *Statistics in Transition*.

GALEONE, C., POLLASTRI, A., (2013). Cost-effectiveness Ratio for comparing social and health policies. Forthcoming on *Silesian Statistical Review*.

KIM, H.J., (2006). On the Ratio of Two Folded Normal Distributions. *Communications in Statistics – Theory and Methods*, **35**, 965-977.

KOTZ, S., BALAKRISHNAN, N., JOHNSON, N.L., (2000). *Continuous Multivariate Distributions*. Wiley, New York.

OKSOY, D., AROIAN, L.A., (1986). Computational Techniques and examples of the density and the distribution of the quotient of two correlated normal variables. *Proceeding of the American Statistical Association Business an Economic Section 660-663*.

OKSOY D., AROIAN L.A. (1994) The quotient of two correlated normal variables with applications, *Communications in Statistics – Simulation and Computation*, **23** (1): 223-241.

OWEN, D.B., (1956): Tables for computing Bivariate Normal Probabilities. Annals of Mathematical Statistics, 27, 1075-1090.

POLLASTRI, A., TULLI, V., (2011). The distribution of the absolute value of the ratio of two correlated Normals Working paper n. 223 Dipartimento di Statistica e Metodi Quantitativi Università Milano-Bicocca.