ORF 505: Statistical Analysis of Financial Data

Tuesday Precept 3: Dependence and Multivariate Data

Sep 24, 2024

Lecturer: Qishuo Yin *Scribe: Qishuo Yin*

1 Review: Copulas

1.1 Motivation

Recall from Section 1.3 of [1]:

If X is a random variable with continuous c.d.f. $x \mapsto F_X(x)$ then the random variable $F_X(X)$ is uniformly distributed on $[0, 1]$.

1.2 Definition

A copula is **the joint distribution of uniformly distributed random variables**. If $U_1, \cdots, U_k$ are $U(0, 1)$ then

$$C(u_1, \cdots, u_k) = \mathbb{P}\{U_1 \leq u_1, \cdots, U_k \leq u_k\}$$

is a **copula**. C is a function from $[0, 1] \times \cdots \times [0, 1]$ into $[0, 1]$

- $C(u_1, \cdots, u_k)$ is non-decreasing in each variable u_i.
- $C(1, \cdots, 1, u_i, 1, \cdots, 1) = u_i$ for all i.

1.3 Examples

- **Independent copula**: When X and Y are independent,

$$C_{ind}(u_1, \cdots, u_k) = u_1 \cdots u_k$$

- **Gaussian copula**: The family of Gaussian copulas depend upon each other as jointly Gaussian random variables with Pearson correlation coefficient $\rho \in (0, 1)$. The marginal distributions can or cannot be Gaussian.

$$C_{Gauss,\rho}(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} e^{-[s^2 - 2\rho st + t^2]/2(1-\rho^2)} ds\, dt$$

- **Gumbel(or Logistic) copula ($0 < \beta \leq 1$)**

$$C_{Gumbel,\beta}(u_1, u_2) = e^{-[(-\log u_1)^{1/\beta} + (-\log u_2)^{1/\beta}]}$$

2 Problem 3.7 from Textbook

Priblem 3.7. This problem is based on the data contained in the data set `UTILITIES` included in the library texttttRsafd. It is a matrix with two columns, each row corresponding to a given day. The first column gives the log of the weekly return on an index based on Southern Electric stock value and capitalization, (well call that variable X), and the second column gives, on the same day, the same quantity for Duke Energy (well call that variable Y), another large utility company.

```
X <- UTILITIES[,1]
Y <- UTILITIES[,2]
```

Listing 1: load dataset

1. Compute the means and the standard deviations of X and Y , and compute their correlation coefficients.

```
muX <- mean(X)
muY <- mean(Y)
sX <- sd(X)
sY <- sd(Y)
rho <- cor(X,Y)

tau <- cor.test(X,Y,method="k")
rhoSpearman <- cor.test(X,Y,method="s", exact=FALSE)
```

Listing 2: section 3.7.1

2. We first assume that X and Y are samples from a jointly Gaussian distribution with parameters computed in question 1. Compute the q-percentile with $q = 2\%$ of a the variables $X + Y$ and $X - Y$.

```
varXplusY = sX^2 + sY^2 + 2*rho*sX*sY
qnorm(0.02, mean = muX + muY, sd = sqrt(varXplusY))

varXminusY = sX^2 + sY^2 - 2*rho*sX*sY
qnorm(0.02, mean = muX - muY, sd = sqrt(varXminusY))

par(mfrow=c(2, 2))
shape.plot(X, tail="upper")
title("Positive X")
shape.plot(X, tail="lower")
title("Negative X")
shape.plot(Y, tail="upper")
title("Psotive Y")
shape.plot(Y, tail="lower")
title("Negative Y")
```

Listing 3: section 3.7.2 codes

3. Fit a generalized Pareto distribution (GPD) to X and Y separately, and fit a copula of the Gumbel family to the empirical copula of the data.

```
Xest <- fit.gpd(X, lower=-0.02, upper=0.03)
```

```r
Yest <- fit.gpd(Y, lower=-0.05, upper=0.025)

par(mfrow=c(2, 2))
tailplot(Xest, tail="lower")
title("Lower X")
tailplot(Xest, tail="upper")
title("Upper X")
tailplot(Yest, tail="lower")
title("Lower Y")
tailplot(Yest, tail="upper")

X1 <- pgpd(Xest, X)
Y1 <- pgpd(Yest, Y)
plot(X1, Y1)
empC <- empirical.copula(X1, Y1)
est.Cop <- fit.copula(empC, "gumbel")
print(est.Cop)
```

Listing 4: section 3.7.3 codes

References

[1] R. Carmona. *Statistical analysis of financial data in R*, volume 2. Springer, 2014.