Copula Analysis for Evangelicals and Bolsonaro votes in 2018 under Bayesian point-of-view

Schultz, J.A.A.[1] and González-López, V.A.[2]

[2]University of Campinas - Department of Statistics. Rua Sérgio Buarque de Holanda 651. Campinas, S.P. Brazil. CEP: 13083-859.
Emails: [1]zeadolfo@gmail.com, [2]veronica@ime.unicamp.br

Abstract

In this article, we tried to use the proportion of Bolsonaro's voter in the second round in each state in Brazil and, one of the well knowing Bolsonaro's basis, the proportion of evangelical. We estimated the parameters using a frequentist and bayesian perspective. Moreover, we compare the using of FBST (Full Bayesian Significance Test) for hipothesis test and compare with the frequentist hipothesis tests using a simple correlation test.

1 Introduction

The study of society's sectors that, due to their high proportion in the whole electoral group, ends deciding countries' elections is always a current topic. Social aspects of the moment and an accurate diagnosis can be identified that explains the victories. One aspect that could help in the construction of such a diagnosis is the identification of the relationships that occur between social groups. We investigate the relationship between the number of voters in the winner of the presidential elections in Brazil (2018) and the percentage of individuals who share the predominant religion. In order to calculate it, copulae are very useful since it gives us the flexibility to calculate dependence measures in a easy way building models.

2 Data and Models

We approach this problem using the concept of copula, where if H is the cumulative distribution function of (X, Y), there is a function C, such that $H(x, y) = C(F_X(x), F_Y(y))$, with $F_X(x) = \lim_{y\to\infty} H(x, y)$ and $F_Y(y) = \lim_{x\to\infty} H(x, y)$. And the function C is the 2-copula of (X, Y). $C(u, v) = \text{Prob}(F_X(X) \leq u, F_Y(Y) \leq v)$, for $u, v \in [0, 1]$, then C is the distribution of the variables $U := F_X(X)$ and $V := F_Y(Y)$. In order to give flexibility to our analysis we investigated 5 types of dependences, (i) Normal copula, (ii) t-student copula, and 3 Archimedean copulas. We say that the dependence between X and Y follows an Archimedean copula, generated by ϕ, if (X, Y) is a pair of continuous random variables with associated 2-copula C given by the form $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))\ \ u, v \in [0, 1]$, with $\phi(\cdot)$ a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty)$, such that $\phi(1) = 0$ and ϕ^{-1} is the pseudo-inverse of ϕ, which is equal to the usual inverse in $t \in [0, \phi(0)]$ and is equal to zero in $t \in [\phi(0), \infty]$, see [9] for details. The values u and v are related to the random variables U and V. Then, the 3 remaining copulas that we consider are (iii) Gumbel's copula ($\phi(t) = \{-\ln(t)\}^\theta, \theta \in [1, \infty)$), (iv) Joe's copula ($\phi(t) = -\ln\{1 - (1 - t)^\theta\}, \theta \in [1, \infty)$) and (v) Frank's copula ($\phi(t) = -\ln\{\frac{e^{-\theta t}-1}{e^{-\theta}-1}\}, \theta \in (-\infty, \infty) \setminus \{0\}$). See [9] for a description of each family.

Now, we will now introduce briefly FBST test.

Let be $\pi(\theta|u, v)$ the posterior and let be an evidence measure in H_0 using the tangent set called T, with the more "probable" points for the set Θ_0.

$$T = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) > t\}, \text{ where } t = \sup_{\Theta_0}\{\pi(\theta|\mathbf{x})\}$$

Let define the evidence measure *evalor* in favor of the set Θ_0 as follow:

$$evalor_{\Theta_0} = \mathbb{P}(\theta \in T^c|\mathbf{x}) = 1 - \mathbb{P}(\theta \in T|\mathbf{x}) = 1 - \int_T \pi(\theta|\mathbf{x})d\theta.$$

3 Model Selection

Dados x é ?, y é ? Gráficos.

In order to proceed with the estimation of the model, the original observations $\{(x_i, y_i)\}_{i=1}^n$ are replaced by their re-scaled marginal ranks to $[0,1]$ (pseudo-observations), $u_i := \frac{|\{j:1\leq j\leq n, x_j\leq x_i\}|}{n}$ and $v_i := \frac{|\{j:1\leq j\leq n, y_j\leq y_i\}|}{n}$, $i = 1, \cdots, n$, where $|A|$ denotes the cardinal of the set A. Through the pseudo-observations $\{(u_i, v_i)\}_{i=1}^n$, for copula families (i)-(v) we construct the likelihood $\prod_{i=1}^n c(u_i, v_i)$, where c is the density of the copula.

X vs Y

U vs V.

In all the cases we use the *Copula R-package*, and the function *fitCopula()*, with arguments *copula* (1) and *method* (2), with (1) 'normalCopula(dim=2)', 'tCopula(dim=2, df = 1)', 'frankCopula(dim=2)', 'gumbelCopula(dim=2)', 'joeCopula(dim=2)', respectively and (2) method= 'mpl'. This package uses a goodness-of-fit test see [8] and we find the following results in the table 1.

explicar que isto tem sido feito com EMV de cada parametro

Table 1: Goodness of fit using copulas

Copula	p value
Frank	0.9206
Gaussian	0.9236
Gumbel	0.5919
t(1)	0.5960
Joe	0.0405

incluir forma da Normal c/ ρ em evidência.

The, unexpected, data symmetry, the lack of extreme values and no right or left tail dependence favored the gaussian copula and the Frank copula. For the inferences, we use the gaussian copula in order to simplify the results and the interpretation since the normal copula uses the correlation as a dependence measure.

4 Bayesian Inferences

mencionar ρ e que a priori é em ρ.

Once decided the likelihood, we need to decide the priors. For this study, we choose 3 distributions as following:

- Beta(1,1), all the support has the same probability, a beta (15),
- beta(15,1), favoring the positive dependence,
- beta(1,15), favoring the negative dependence.

These beta distribution are an extended version for the comun beta, i.e., we extended the support from -1 to 1 instead of 0 to 1.

Then we estimate the posterior function using *HMC* (Hamiltonian Monte-Carlo) and we achieve the following results for the correlation parameter:

Table 2: Summary measures posterior function

Prior	Minimum	1 Quartil	Median	3 Quartil	Maximum	Mean	Std Desv
Beta(1,1)	0.3048	0.6888	0.7438	0.7888	0.8956	0.7311	0.08
Beta(15,1)	0.3290	0.7331	0.7768	0.8125	0.9092	0.7768	0.064
Beta(1,15)	-0.4403	-0.0467	0.0756	0.1987	0.66122	0.0792	0.1754

The posterior mean using the Beta(1,1) is 0.7311 and is very closer to the Pearson's correlation which estimative is 0.7232. The posterior using the Beta(15,1) does not affect very much the estimative and

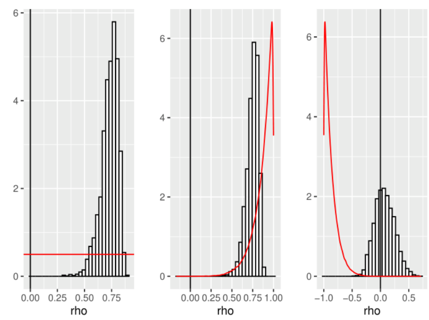

Figure 1: Red curve represents the prior and the black, the posterior

Table 3: e values by prior

Prior	e value
Beta(1,1)	< 0.0001
Beta(15,1)	< 0.0001
Beta(1,15)	0.6427
Frequentist test	< 0.0001

the standard desviation is closer to the prior distribution. However, the posterior Beta(1,15) changes the estimative significantly.

Using the Beta(1,1) (Table 3) we can reach a very closer conclusions to the frenquentist test. Moreover, the Beta(15,1), as expected, reject the no correlation hypothesis. On another hand, the Beta(1,15) does not reject the hypothesis since the mixture between prior and likelihood distributions bring the posterior just in the 0 correlation which demonstrate the model is not very robust.

$H_0: \rho = 0$

5 Conclusion

This analysis showed us, first of all, how correlated is the porcentage of evangelical in each state in Brazil and the proportion of votes in Bolsonaro in 2018. We found the best copula, in terms of goodness-of-fit as well as the interpretability, was the Normal Copula. We decided to use the extended Beta distribution for a prior in order to make the model with a better interpretation. We found the Beta(1,1) had similar results to the frenquentist test and, the prior Beta(1,15) against the likelihood, changed the final results probably because we have only 27 states in Brazil which is a very small data base.

The next works, it is interesting test other kind of hypothesis as well as more complex copulae as a multivariate copula instead of a bivariate copula.

References

[1] Censo 2010. IBGE.

[2] TSE - Divulgação de Resultados de Eleições. TSE. 7 de outubro de 2018.

[3] Madruga, M.R., Esteves, L.G., Wechsler, S., 2001. **On the bayesianity of Pereira-Stern tests** , *Test (Madrid)* , Espanha, Vol. 10, n.2, 291-299.

[4] Pereira, C.A.B., Stern, J.M., 1999. **Evidence Credibility: Full Bayesian SignicanceTest** , *Entropy* ,1, 69-80.

[5] Joe, Harry, 2015, **Dependence Modeling with Copulas**, 2 edition, *Taylor & Francis Group*, Boca Raton.

[6] Congdon, P., 2014. **Applied Bayesian modelling**, 2 edition, John Wiley & Sons, West Sussex.

[7] Vuolo, M., 2015. **Copula Models for Sociology: Measures of Dependence and Probabilities for Joint Distributions** , *Sociological Methods & Research* , 1-45.

[8] Genest, C., Rémillard, B., and Beaudoin, D. (2009). **Goodness-of-fit tests for copulas: A review and a power study**, *Insurance: Mathematics and Economics*, 44, 199-214.

[9] Nelsen, R. (2006). **A Introduction to Copula**, 2 edição, Springer, New York.