Dependence measures and competing risks models under the generalized Farlie-Gumbel-Morgenstern Copula

Presenter: Jia-Han Shih   Advisor: Takeshi Emura

Graduate Institute of Statistics, National Central University

Abstract: The first part reviews the properties of some selected dependence measures under the generalized Farlie-Gumbel-Morgenstern (FGM) copula. In the second part, we obtain the expression of sub-distribution functions to analyze dependent competing risks under the generalized FGM copula. With the Burr III margins, we show that our expression has a closed form and generalizes the reliability measure previously obtained by Domma and Giordano (2013). We performed maximum likelihood estimation under the proposed competing risks models with a randomized Newton-Raphson algorithm and also conduct simulations to check the correctness of our method. Real data are used for illustration.

1. Introduction

We reviews the properties of some selected dependence measures under the generalized FGM copula of Bairamov and Kotz (2002). We give a few remarks on the relationship among the dependence measures, derive Blest's coefficient, and suggest simplifying the expression of Kochar and Gupta's dependence measure previously obtained by Amini et al. (2011).

The second contribution of this paper is to derive the expressions of sub-distribution functions with competing risks under the generalized FGM copula. With the Burr III margins, we show that our expression is explicitly written and is a generalization of the reliability measure obtained by Domma and Giordano (2013).

2. Copula

By definition, a bivariate copula is a map

$$C:[0,1]^2 \to [0,1]$$

with $C(u,1)=u$ and $C(1,v)=v$.

By Sklar's theorem, one has the representation

$$F(x,y)=C\{F_1(x),F_2(y)\}.$$

The one-parameter FGM copula is

$$C(u,v)=uv\{1+\theta(1-u)(1-v)\},\quad \theta\in[-1,1].$$

The range of Spearman's rho is $[-1/3,1/3]$ and the range of Kendall's tau is $[-2/9,2/9]$.

3. Generalized FGM copula

The generalized FGM copula proposed by Bairamov and Kotz (2002) is

$$C(u,v)=uv\{1+\theta(1-u^p)^q(1-v^p)^q\},\quad p,q\ge 1,$$

where the possible range of θ is

$$-\min\left\{1,\frac{1}{p^{2q}}\left(\frac{1+pq}{q-1}\right)^{2q-2}\right\}\le\theta\le\frac{1}{p^q}\left(\frac{1+pq}{q-1}\right)^{q-1}.$$

3.1. Spearman's rho and Kendall's tau

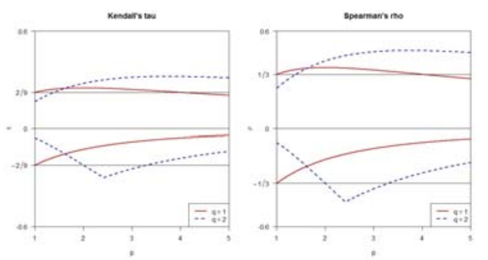
Spearman's rho and Kendall's tau are based on the concept of concordance.

Proposition 1 *Under the generalized FGM copula, Spearman's rho and Kendall's tau are*

$$\rho=12\left\{\frac{q}{2+pq}B\left(\frac{2}{p},q\right)\right\}^2\theta,\ \tau=8\left\{\frac{q}{2+pq}B\left(\frac{2}{p},q\right)\right\}^2\theta,$$

Hence, we have $2\rho=3\tau$ *in the range of* θ.

Proposition 1 is an obvious simplification form the results of Amini et al. (2011).



3.2. Blest's coefficient and symmetrized version

Blest (2000) proposed a rank correlation measure which emphasizes the differences in the top ranks. Blest's coefficient (ν) is defined as

$$\nu=2-12\int_0^1\int_0^1(1-u)^2 v\,dC(u,v).$$

Genest and Plante (2003) suggested a symmetrized version of Blest's coefficient (ξ), defined as

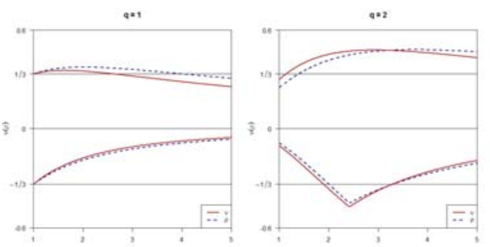
$$\xi=-4+6\int_0^1\int_0^1 uv(4-u-v)\,dC(u,v).$$

Theorem 1 *Under the generalized FGM copula,*

$$\nu=\xi=24\left\{\frac{q}{2+pq}B\left(\frac{2}{p},q\right)\right\}^2\theta$$
$$-24\left\{\frac{q}{2+pq}B\left(\frac{2}{p},q\right)\right\}\left\{\frac{q}{3+pq}B\left(\frac{3}{p},q\right)\right\}\theta.$$

Corollary 1 *The relationship between Blest's coefficient and Spearman's rho is*

$$\nu=2\left\{1-\frac{2+pq}{3+pq}\prod_{i=0}^{q-1}\left(\frac{2+pi}{3+pi}\right)\right\}\rho.$$



3.3. Kochar and Gupta's dependence measure

The dependence measure of Kochar and Gupta (Kochar and Gupta 1987) is based on the concept of quadrant dependence. It is defined as

$$D_k=\int_0^1\int_0^1\{C(u,v)^k-u^k v^k\}\,dC(u,v).$$

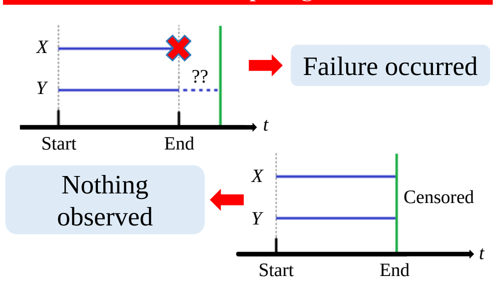
Theorem 2 *Under the generalized FGM copula,*

$$D_k=k\left\{\frac{q}{k+1+pq}B\left(\frac{k+1}{p},q\right)\right\}^2\theta+\sum_{j=1}^{k-1}\binom{k}{j}(k-j)(k+1)$$
$$\times\left\{\frac{q}{k+1+(j+1)pq}B\left(\frac{k+1}{p},(j+1)q\right)\right\}^2\theta^{j+1}.$$

Theorem 2 is a simplification form the results of Amini et al. (2011).

4. Results on competing risks measures

What is Competing risks?



What is sub-distribution (density) function?

$X \le Y$ $T=\min(X,Y)$

$$F(1,t)=P(\Delta=1,T\le t)=\int_0^t f(1,z)\,dz$$

$$f(1,t)=-\partial\bar{F}(x,y)/\partial x|_{x=y=t}$$

Consider the Burr III marginal distributions

$$F_1(x)=(1+x^{-\gamma})^{-\alpha},\ F_2(y)=(1+y^{-\gamma})^{-\beta}.$$

Theorem 3 *We obtain the sub-distribution and sub-density functions under the generalized FGM copula with Burr III margins.* **(Complicate formula omitted)**

Our expression is a generalization of the reliability measure obtained by Domma and Giordano (2013).

5. Maximum likelihood estimation

Given observed data, the log-likelihood function of the competing risks models under the generalized FGM copula with Burr III margins has the form

$$\ell_n(\boldsymbol{\theta})=\sum_{i=1}^n\delta_i\log f(1,T_i)+\sum_{i=1}^n\delta_i'\log f(2,T_i)$$
$$+\sum_{i=1}^n(1-\delta_i-\delta_i')\log\bar{F}(T_i,T_i).$$

Randomized Newton-Raphson algorithm

Step1. Set initial value $(\boldsymbol{\theta}^{(0)})$.

Step2. Repeat the following iterations:

$$\boldsymbol{\theta}^{(k+1)}=\boldsymbol{\theta}^{(k)}-\left\{\frac{\partial^2}{\partial\boldsymbol{\theta}^2}\ell_n(\boldsymbol{\theta}^{(k)})\right\}^{-1}\frac{\partial}{\partial\boldsymbol{\theta}}\ell_n(\boldsymbol{\theta}^{(k)}).$$

- If $|\boldsymbol{\theta}^{(k+1)}-\boldsymbol{\theta}^{(k)}|<10^{-5}$, then stop and the MLE is $\boldsymbol{\theta}^{(k+1)}$.
- If $|\boldsymbol{\theta}^{(k+1)}-\boldsymbol{\theta}^{(k)}|>10$ or $f(\Delta,T_i)=0$, $\bar{F}(T_i,T_i)=0$, for some i and $\Delta=1,2$, then return to Step2 with the initial value replaced by $(\exp(\mathbf{u}')\times\boldsymbol{\theta}^{(0)})$.

Copula parameters are unidentifiable!

6. Simulations

Simulation results are based on 1,000 repetitions with given copula parameters $\theta=0.7$, $p=3$, $q=2$.

True	Prop.	n	$E(\hat\alpha)$	$E(\hat\beta)$	$E(\hat\gamma)$	AI	AR
$\alpha=3$,	$X=29\%$,	100	3.065	2.033	7.076	8.3	1.1
$\beta=2$,	$Y=51\%$,	200	3.021	2.014	7.057	8.6	1.3
$\gamma=7$	Cen $=20\%$	300	3.015	2.006	7.045	8.8	1.5
	$X=21\%$,	100	3.089	2.048	7.127	7.3	0.5
	$Y=38\%$,	200	3.032	2.018	7.093	7.2	0.5
	Cen $=41\%$	300	3.027	2.011	7.055	7.2	0.5

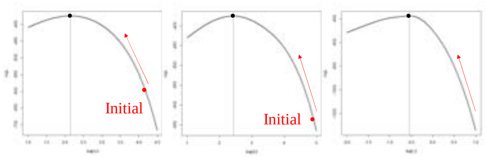
True	Prop.	n	$MSE(\hat\alpha)$	$MSE(\hat\beta)$	$MSE(\hat\gamma)$
$\alpha=3$,	$X=29\%$,	100	0.155	0.050	0.253
$\beta=2$,	$Y=51\%$,	200	0.065	0.024	0.138
$\gamma=7$	Cen $=20\%$	300	0.040	0.015	0.088
	$X=21\%$,	100	0.208	0.068	0.401
	$Y=38\%$,	200	0.098	0.031	0.180
	Cen $=41\%$	300	0.058	0.020	0.122

7. Real data illustration

We use the CVD risk data which is available in the website of Statistical Software Information, University of Massachusetts Amherst.

	The number of observed events (rates)		
Sample size	CVD	Other cause	Censoring
65	32 (49%)	24 (37%)	9 (14%)

Given copula parameters $\theta=0.7$, $p=3$, $q=2$, we obtain the MLE $\hat\alpha=8.576$, $\hat\beta=11.299$, $\hat\gamma=0.579$.



8. Conclusions

- New results (Theorem 1 – 3, Corollary 1).
- The MLE is obtained by using a randomized Newton-Raphson algorithm.

References: [1] Amini M, Jabbari H, Mohtashami Borzadaran GR (2011) Commun Stat Simul Comput. 40: 1192-1205. [2] Bairamov I, Kotz S (2002) Metrika. 56: 55-72. [3] Blest DC (2000) Aust N Z J Stat. 42: 101-111. [4] Domma F, Giordano S (2013) Stat Pap. 54: 807-826. [5] Genest C, Plante JF (2003) Can J Stat. 31: 35-52. [6] Kochar SC, Gupta RP (1987) Biometrika. 74: 664-666.