Detection of Heterogeneous Structure on Gaussian Copula Model Using Power Entropy

Akifumi Notsu Second-year Ph.D. Student, The Graduate University for Advanced Studies

【Abstract】

We have addressed a statistical estimation problem on Gaussian copula model, where a working model is Gaussian copula but an underlying distribution may be largely different from Gaussian copula. Usual estimators, like maximum likelihood estimators (MLEs), do not work well when such the large model misspecification occurs. We have proposed to apply a gamma estimator (GE) associated with projective power cross entropy to this problem. Owing to the property called " emergence ", the gamma estimator works well not only when an underlying distribution is Gaussian copula but also when some kind of large misspecification occurs.

【Gaussian Copula Model and Gamma Estimator】

The density function of Gaussian copula is given by

$$c_{\mathrm{G}}(\boldsymbol{u};P)=\det(P)^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{x}(\boldsymbol{u})^T(P^{-1}-I)\boldsymbol{x}(\boldsymbol{u})\right),\ \boldsymbol{u}\in[0,1]^m,$$

where P is a correlation matrix, $\boldsymbol{x}(\boldsymbol{u})=(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_m))^T$, and $\Phi(x)$ is the distribution function of the standard Gaussian distribution.

Let $\boldsymbol{u}^{(1)},\dots,\boldsymbol{u}^{(n)}$ be a random sample from a copula with the probability density function $c(\boldsymbol{u})$. If we use $c_{\mathrm{G}}(\boldsymbol{u};P)$ as a working model, then the model is called Gaussian copula model. The projective power cross entropy between $c(\boldsymbol{u})$ and $c_{\mathrm{G}}(\boldsymbol{u};P)$ is defined by

$$\begin{aligned}C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P)) &= -\frac{1}{\gamma}(1+\gamma)^{\frac{m\gamma}{2(1+\gamma)}-1}(2\pi)^{-\frac{m\gamma}{2(1+\gamma)}}\det(P)^{-\frac{\gamma}{2(1+\gamma)}}\\ &\quad\times\int_{\mathbb{R}^m}g(\boldsymbol{x})\exp\left(-\frac{\gamma}{2}\boldsymbol{x}^TP^{-1}\boldsymbol{x}\right)d\boldsymbol{x},\quad(1)\end{aligned}$$

where $g(\boldsymbol{x})=c(\boldsymbol{u}(\boldsymbol{x}))\left|\det\frac{\partial\boldsymbol{u}}{\partial\boldsymbol{x}}\right|$. A loss function denoted by $L_\gamma(P)$ associated with the projective power cross entropy is defined by

$$\begin{aligned}L_\gamma(P) &= -\frac{1}{\gamma}(1+\gamma)^{\frac{m\gamma}{2(1+\gamma)}-1}(2\pi)^{-\frac{m\gamma}{2(1+\gamma)}}\det(P)^{-\frac{\gamma}{2(1+\gamma)}}\\ &\quad\times\frac{1}{n}\sum_{i=1}^n\exp\left(-\frac{\gamma}{2}\boldsymbol{x}^{(i)T}P^{-1}\boldsymbol{x}^{(i)}\right),\quad(2)\end{aligned}$$

where $\boldsymbol{x}^{(i)}=\left(\Phi^{-1}\left(u_1^{(i)}\right),\dots,\Phi^{-1}\left(u_m^{(i)}\right)\right)^T,(i=1,\dots,n)$.

- When $L_\gamma(P)$ has one locally minimum solution, then the underlying distribution is estimated to be Gaussian copula and the gamma estimator of P is defined by the minimum solution.
- When $L_\gamma(P)$ has $\ell(\ell\geq 2)$ locally minimum solutions, then the underlying distribution is estimated not to be Gaussian copula but mixture of ℓ Gaussian copulas. The gamma estimators from each Gaussian copula are defined by the locally minimum solutions.

【The Emergence Property of Gamma Estimators】

Note that $E(L_\gamma(P))=C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$. We have considered the following monotonic increase transformation of $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$:

$$\begin{aligned}&-\log(-C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P)))\\ &=-\log\int_{\mathbb{R}^m}g(\boldsymbol{x})\exp\left(-\frac{\gamma}{2}\boldsymbol{x}^TP^{-1}\boldsymbol{x}\right)d\boldsymbol{x}-\left(-\frac{\gamma}{2(1+\gamma)}\log\det P\right)\\ &\quad-\log\left(\frac{1}{\gamma}(1+\gamma)^{\frac{m\gamma}{2(1+\gamma)}-1}(2\pi)^{-\frac{m\gamma}{2(1+\gamma)}}\right).\quad(3)\end{aligned}$$

Note that the first and second terms of (3) are convex functions. That is (3) is difference between two convex functions. Owing to this property $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$ may have some locally minimum solutions.

However if $c(\boldsymbol{u})$ is a Gaussian copula density function, $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$ has a locally minimum solution. That is we have the following theorem.

Theorem 1 *If $c(\boldsymbol{u})=c_{\mathrm{G}}(\boldsymbol{u};P_0)$, then $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$ has one locally minimum solution P_0.*

Consider $c(\boldsymbol{u})$ is not a Gaussian copula density function, especially

$$c(\boldsymbol{u})=\tau c_{\mathrm{G}}(\boldsymbol{u};P_1)+(1-\tau)c_{\mathrm{G}}(\boldsymbol{u};P_2),\ P_1\neq P_2,\ \tau\in(0,1).$$

In two dimensional case, $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$ is a univariate function of the non diagonal element of P. A sufficient condition that this univariate function has two locally minimum solutions is the following:

Theorem 2 *Suppose $c(\boldsymbol{u})=0.5c_{\mathrm{G}}(\boldsymbol{u};P_1)+0.5c_{\mathrm{G}}(\boldsymbol{u};P_2)$, where*

$$P_1=\begin{pmatrix}1&\rho_0\\ \rho_0&1\end{pmatrix},\quad P_2=\begin{pmatrix}1&-\rho_0\\ -\rho_0&1\end{pmatrix}.$$

If $\rho_0>\sqrt{6-\sqrt{28}}$ and $\gamma=1$ then $C_\gamma(c(\cdot),c_{\mathrm{G}}(\cdot;P))$ has two locally minimum solutions in the interval $(-1,0)$ and $(0,1)$ respectively.

【Simulation Results】

Proceedure: Suppose $c(\boldsymbol{u})=0.5c_{\mathrm{G}}(\boldsymbol{u};P_1)+0.5c_{\mathrm{G}}(\boldsymbol{u};P_2)$, where $P_i=I+\Lambda\odot(a_ia_i^T)$, $\odot$ means Hadamard product, Λ is a matrix such that

$$\Lambda_{ij}=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases},$$

and

$$\begin{aligned}a_1&=(0.92,0.92,0.92,0.92,0.85,0.85,0.85,0.85,0.85,0.85)^T,\\ a_2&=(-0.92,0.92,-0.92,0.92,-0.85,0.85,-0.85,0.85,-0.85,0.85)^T.\end{aligned}$$

We have generated a sample of size 10000 from $c(\boldsymbol{u})$ to calculate the GE and repeated 50 times. If there are two GE's G_1 and G_2 such that

$$||G_1-P_1||<||G_2-P_1||,$$

then G_1 is thought of as an estimator of P_1 and G_2 for P_2. When two GE's are not obtained, we call it failure case.

Result: We have succeeded 42 times out of 50 and calculated the root mean square error (RMSE). Some of the results are tabulated in Table 1. Here (i,j) element of P_k is denoted by $[P_k]_{ij}$. The RMSEs of $[P_1]_{13}$ and $[P_2]_{13}$ are smaller than others because $[P_1]_{13}=[P_2]_{13}$.

Table 1: The RMSEs of the GEs of $[P_1]_{12}$, $[P_1]_{13}$, $[P_2]_{12}$, and $[P_2]_{13}$

	P_1			P_2		
	$[P_1]_{12}$	$[P_1]_{13}$	$[P_1]_{14}$	$[P_2]_{12}$	$[P_2]_{13}$	$[P_2]_{14}$
RMSE	0.137	0.021	0.142	0.145	0.018	0.148

【Discussion】

Our methodology can be applied to the case when an underlying distribution may be mixture of Gaussian copula. It is also possible to apply gamma estimators to the case when the underlying distribution is mixture of t-copula or mixture of Gaussian copula and t-copula. However it is difficult to detect heterogeneous structure when mixture ratio is not same for two components and components are greater than 2. Those are future studies.

Reference

Eguchi, S., Komori, O., and Kato, S. (2011). Projective Power Entropy and Maximum Tsallis Entropy Distributions. *Entropy*, **13**, 1746-1764.

Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *J. Multivariate Analysis*, **99**, 9 2053-2081.

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer.