修士論文 要旨

Minimum information copulas under fixed Kendall's rank correlation（ケンドールの順位相関係数を固定した下での最小情報コピュラ）

数理情報学専攻 48226213　**助田 一晟**

指導教員　**清 智也 教授**

1　はじめに

一般に d 次元コピュラとは全ての周辺分布が $[0,1]$ 上の一様分布である d 次元の分布関数を指し, 任意の d 次元分布は Sklar の定理に基づきコピュラと周辺分布を用いて表現することができることから, コピュラは異なる変数間の従属性を完全に記述する．なお, 本稿では 2 次元に限定する．

所与の制約条件下で一様コピュラ ($[0,1]^2$ 上の一様密度を持つコピュラ) に Kullback-Leibler 情報量の意味で最も近いコピュラを**最小情報コピュラ**と呼ぶ．コピュラを選択する際, 通常はパラメトリックに分布の形状を仮定することが多い．一方, 最小情報コピュラは既知の情報を制約条件として課しエントロピー最大化により最も自然な分布を決定するという考え方に基づく．

既存研究では確率密度関数に関して線形な制約条件（モーメント条件）のみが考察されていた．しかし, 種々の制約条件を網羅的に指定するためには既存の枠組みの拡張が必要である．特に, 典型的な従属指標の一つであるスピアマンの順位相関は既存研究で扱われている [4] のに対し, 並列に語られることの多いケンドールの順位相関はその非線形性からこの枠組みで扱うことが不可能であった．

本論文では最小情報コピュラの適用範囲の拡張を目的とし, ケンドールの順位相関を固定した下での最小情報コピュラについて理論的な考察を与える．この新しいコピュラは局所従属性により特徴づけられ, また相関が十分に小さい場合に限って一意存在性が保証される．さらに, 当該コピュラが既存研究で典型的に用いられてきた Frank コピュラという分布に一致していることの根拠を与え, Frank コピュラを利用したモデリングに新しい解釈可能性を与える．

2　既存研究

最小情報コピュラは K 本のモーメント制約を課した上での情報量最小化問題の解として与えられる．

$$\text{minimize}\quad \int_0^1\int_0^1 p(x,y)\log p(x,y)dxdy$$

$$\text{s.t.}\quad \int_0^1 p(x,y)dy=1,\ \int_0^1 p(x,y)dx=1$$

$$\int_0^1\int_0^1 h_k(x,y)p(x,y)dxdy=\mu_k,\ 1\le k\le K$$

情報量最小化はエントロピー最大化と等価である．この問題に対しては, 解の存在性と一意性, 解の形が知られている [1]．代表的な例として, スピアマンの順位相関係数を固定した下での最小情報コピュラ（minimum information copula under fixed Spearman's ρ, MICS）が知られる．これは上記の最適化問題において

$$K=1,\ h(x,y)=12(x-1/2)(y-1/2)$$

と設定した場合に相当する．

3　本研究の提案：ケンドールの順位相関係数を固定した下での最小情報コピュラ

本研究では, ケンドールの順位相関係数を制約条件として与えた最小情報コピュラ（minimum information copula under fixed Kendall's τ, MICK）を考察する．まず, チェス盤コピュラと呼ばれる離散近似を施した MICK について結果を述べ, その後連続版の MICK に対する結果を述べる．

3.1　チェス盤（離散版）MICK

定義域 $[0,1]^2$ を $n\times n$ のグリッドで区切り, 区分一様と近似するチェス盤コピュラを正方行列 $\Pi=(\pi_{ij})$ と同一視した上で, 以下の最適化問題を考える:

$$\text{minimize}\quad \sum_{i=1}^n\sum_{j=1}^n \pi_{ij}\log\pi_{ij} \tag{1}$$

$$\text{s.t.}\quad \sum_{j=1}^n \pi_{ij}=\frac{1}{n},\ \sum_{i=1}^n \pi_{ij}=\frac{1}{n} \tag{2}$$

$$1-\mathrm{Tr}(\Xi\Pi\Xi\Pi^{\mathrm{T}})=\tau,\ \Pi=(\pi_{ij}) \tag{3}$$

ただし, $\Xi=\begin{pmatrix}1&0&\dots&0\\2&1&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\2&\dots&2&1\end{pmatrix}$ であり, 制約式はコピュラであるための条件と, チェス盤コピュラに対する

Kendall の順位相関係数 [2] を定数 τ に固定する条件である．Tr は行列のトレースを表す．

この問題の最適解をチェス盤 MICK と名付ける．このコピュラはオッズ比に類する局所従属性が一定であると特徴づけられる．

定理 1. Π をチェス盤 MICK とする．任意の (i,j) の組 $(i,j=1,\ldots,n-1)$ に対し，以下の値が一定．

$$\frac{1}{\pi_{i,j}+\pi_{i+1,j}+\pi_{i+1,j}+\pi_{i+1,j+1}}\log\frac{\pi_{i,j}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}$$

さらに，このコピュラは従属性が十分に小さいケースでは一意存在性が保証される．

定理 2. 最適化問題の制約式 (3) に対応するラグランジュ未定乗数を λ と表す．$\lambda<2$ となる時，最適解は一意に存在する．

3.2 連続版 MICK

次に連続版の MICK を考察する．以下の最適化問題の最適解を連続版 MICK と名付ける．

$$\text{minimize}\quad \int_0^1\int_0^1 p(x,y)\log p(x,y)\mathrm{d}x\mathrm{d}y,$$

$$\text{s.t.}\quad \int_0^1 p(x,y)\mathrm{d}x=1,\ \int_0^1 p(x,y)\mathrm{d}y=1,$$

$$0\le p(x,y),$$

$$\int_0^1\int_0^1\int_0^1\int_0^1 \mathrm{d}x\mathrm{d}y\mathrm{d}\tilde{x}\mathrm{d}\tilde{y}\ \mathrm{sgn}(x-\tilde{x})\mathrm{sgn}(y-\tilde{y})p(x,y)p(\tilde{x},\tilde{y})=\tau.$$

連続版 MICK は Frank コピュラと一致することが理論的，数値的に示唆される（命題 1，図 1）．Frank コピュラは裾従属性（$(u,v)=(0,0)$ や $(u,v)=(1,1)$ 付近における従属性）を持たず（漸近独立），そのようなデータによく適合するとして典型的に用いられるコピュラである．

定義 3 (Frank コピュラ)**.** 以下の密度関数を持つ 1 パラメタコピュラを Frank コピュラと呼ぶ．

$$c_\theta^{Frank}(u,v)=\frac{\theta(1-e^{-\theta})e^{-\theta(u+v)}}{\{1-e^{-\theta}-(1-e^{-\theta u})(1-e^{-\theta v})\}^2}.$$

命題 1. 連続版 MICK と Frank コピュラは，密度関数 $p(x,y)$ がともに以下の偏微分方程式を満たす．

$$\frac{1}{p(x,y)}\frac{\partial}{\partial x}\frac{\partial}{\partial y}\log p(x,y)=\zeta,\zeta:\text{定数}\qquad(4)$$

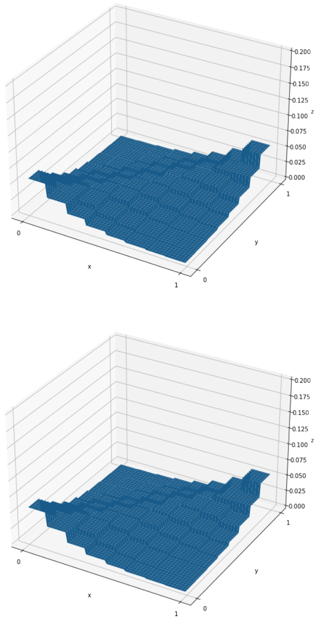

図 1.（上）$\tau=0.5$ と設定した時のチェス盤 MICK，（下）Frank コピュラの離散的なプロット

(4) 式の左辺は確率密度関数 $p(x,y)$ と，その分布が持つ interaction [3] の比となっており，局所従属性を規定する統計学的に重要な量であると考えられる．

連続版 MICK と Frank コピュラの厳密な一致を示すための方法の一つが，偏微分方程式 (4) の解の一意性を示すことであるが，この方程式は非線形非斉次であるため扱いが困難であり，今後の課題となっている．

4 結論と考察

本研究の結果により，Frank コピュラを選択し利用することは，エントロピー最大化原理に則った下で Kendall の順位相関係数の値という情報のみを利用しながらコピュラ，すなわち従属性を決定することに相当するという解釈が可能になった．

参考文献

[1] T. Bedford and K. J. Wilson. On the construction of minimum information bivariate copula families. *Annals of the Institute of Statistical Mathematics*, 66:703–723, 2014.

[2] V. Durrleman, A. Nikeghbali, and T. Roncalli. Copulas approximation and new families. *Available at SSRN 1032547*, 2000.

[3] D. Kurowicka and W. T. van Horssen. On an interaction function for copulas. *Journal of Multivariate Analysis*, 138:127–142, 2015.

[4] A. Meeuwissen and T. Bedford. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *Journal of Statistical Computation and Simulation*, 57(1-4):143–174, 1997.