

ヴァインコピュラに基づく判別分析

北海道大学 安田佑喜
北海道大学 鈴川晶夫

1. はじめに

コピュラとは、多変量同時分布とその周辺分布を接合する関数である。コピュラを用いることにより、多変量同時分布のモデリングを、その周辺分布のモデリングと、確率変数間の従属性構造のモデリングとに分けて考えることができる。主要なコピュラは、大きく分けて、楕円型コピュラと極値コピュラとアルキメデスコピュラに分類され、様々なパラメトリックコピュラが提案されている (Joe 2014)。2次元分布をモデリングするための2変数コピュラを、多変数コピュラに拡張する研究が進んできており、その一つがヴァインコピュラ (vine copula) である (Bedford and Cooke 2002)。

また、近年は、医学・生物学・工学等の分野のみならず、経営学においても顧客判別など、データに基づく判別問題の重要性が増している。判別分析の手法としては、線形判別分析法とロジスティック判別法がよく知られている。また、サポートベクターマシーンも有用な判別手法である。本報告において、ヴァインコピュラを用いて多次元確率分布をモデル化し、そのモデルに基づくベイズ判別ルールについて考える。

2. ヴァインコピュラ

ヴァインコピュラとは同時密度関数の構成に2変数コピュラ密度を用いる方法である。3次元では、 $f(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{i=1}^3 f_i(x_i; \alpha_i) \cdot c_{12}(F_1(x_1; \alpha_1), F_2(x_2; \alpha_2); \theta_{12}) \cdot c_{23}(F_2(x_2; \alpha_2), F_3(x_3; \alpha_3); \theta_{23}) \cdot c_{13|2}(F_{1|2}(x_1|x_2; \alpha_1, \alpha_2, \theta_{12}), F_{3|2}(x_3|x_2; \alpha_2, \alpha_3, \theta_{23}); \theta_{13|2})$ と表現され、2変数コピュラと周辺密度のみによって構成されていることがわかる。ここで、 $\boldsymbol{\alpha}$ は周辺分布のパラメーター、 $\boldsymbol{\theta}$ は2変数コピュラ密度 $c_{st|v}$ のパラメーターである。ヴァインが優れている点は、その構築が非常に柔軟にできるところである。例えば、用いるコピュラの族をそれぞれ変えてみるだけで、モデルの構造は大きく変わる。また一般に多くのコピュラは多次元になるとパラメーターに制約を持ち、扱いが難しくなる。しかし、ヴァインは2変数コピュラのみを用いるのでこの制約を考えなくてもよい。

これらのことから、ヴァインコピュラはその単純な構造にも関わらず、一般のコピュラと比べて、非常に柔軟なモデリングが期待でき、また依存関係を広範囲でとらえることが可能である。

3. ヴァインコピュラを用いた2群判別

ヴァインコピュラを用いて、2つの群の p 変量確率密度関数をモデル化したものを $f_j(\mathbf{x}; \boldsymbol{\alpha}_j, \boldsymbol{\theta}_j)$, $j = 1, 2$ とする。ここに、 $\boldsymbol{\alpha}_j, \boldsymbol{\theta}_j$ は、それぞれ第 j 群における周辺分布とコピュラのパラメータである。このとき、ベイズ判別法の判別関数は $D(\mathbf{x}; \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \log\{f_1(\mathbf{x}; \boldsymbol{\alpha}_1, \boldsymbol{\theta}_1)/f_2(\mathbf{x}; \boldsymbol{\alpha}_2, \boldsymbol{\theta}_2)\}$ により与えられる。パラメータが未知の場合には、最尤法または IFM (inference function for margin) 法 (Joe 1997, Haff 2013) により推定し、推定値を判別関数に差し込むことにより標本判別関数を得る。本報告において、標本判別関数の形状や分布等の性質、誤判別率の評価などについて議論する。

参考文献

- Bedford, T., Cooke, R.M. (2002). Vines—a new graphical model for dependent random variables, *Annals of Statistics*, **30**, 1031-1068
- Haff, I. H. (2013). Parameter estimation for pair-copula constructions, *Bernoulli*, **19**, 462-491.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- Joe, H. (2014). *Dependence Modeling with Copulas*, Chapman and Hall, London.