one-step estimator による漸近的に有効な推定量構成

東京大・情報理工　松本 和也
東京大・情報理工　駒木 文保

$\boldsymbol{X} = (X_1, \dots, X_d)$ を d 次元の確率変数とし，X の同時分布関数 $F : \mathbb{R}^d \to [0,1]$ および周辺分布関数 $F_1, \dots, F_d$ は絶対連続であると仮定する．このとき Sklar の定理により，F のコピュラ C が $F_1(x_1), \dots, F_d(x_d)$ の関数として一意に定まる．(C は $[0,1]^d$ 上の確率分布関数)

$$F(\boldsymbol{x}) = C(F_1(x_1), \dots, F_d(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d$$

ここで扱うセミパラメトリックコピュラモデル $\mathcal{P}$ とは，$F_1, \dots, F_d$ を無限次元の局外パラメータとして扱い，コピュラ C が有限次元のパラメータ $\boldsymbol{\theta} \in \mathbb{R}^m$ で表されるモデルをいう．すなわち F_{ac} を $\mathbb{R} \to [0,1]$ の絶対連続な関数の集合として

$$\mathcal{P} = \{F(\boldsymbol{x}; \boldsymbol{\theta}, F_1, \dots, F_d) | \boldsymbol{\theta} \in \Theta, F_1, \dots, F_d \in \mathcal{F}_{\mathrm{ac}}\}, \quad \Theta \subset \mathbb{R}^m$$
$$F(\boldsymbol{x}; \boldsymbol{\theta}, F_1, \dots, F_d) = C(F_1(x_1), \dots, F_d(x_d); \boldsymbol{\theta})$$

と表せるモデルである．ここで，n 個の観測 $\boldsymbol{x}_i = (x_{i1}, \dots, x_{id})$, $(1 \leqslant i \leqslant n)$ からパラメータ $\boldsymbol{\theta}$ を推定する問題を考える．コピュラの性質上，推定量がランク統計量 $\boldsymbol{R}_i^{(n)} = (R_{i1}^{(n)}, \dots, R_{id}^{(n)})^\top$ ($R_{ij}^{(n)} = \sum_{k=1}^n 1_{X_{ik} \leqslant X_{ij}}$) の関数であることがよい性質であるといえる．以後，本稿中ではこのような性質を rank-based と表記する．

rank-based な $\boldsymbol{\theta}$ の推定量 $\hat{\boldsymbol{\theta}}_n$ の構成法として Genest et al.(1995)[1] は

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} = \underset{\theta \in \Theta}{\operatorname{argmax}} \left[\sum_{i=1}^n \log c(\hat{F}_{n1}(X_{i1}) \dots \hat{F}_{nd}(X_{id}); \boldsymbol{\theta}) \right] \tag{1}$$
$$\hat{F}_{nj}(x_{ij}) = \frac{1}{n+1} R_{ij}^{(n)}$$

を提案した．$\hat{\boldsymbol{\theta}}_n^{\mathrm{PLE}}$ は一般には漸近正規性をもつが，漸近有効性を持たない．Segars et al.(2014)[2] は，セミパラメトリック Gaussian コピュラモデルにおける $\boldsymbol{\theta}$ に関する有効スコア関数 $\boldsymbol{s}^{\mathrm{E}}(F_1(X_1), \dots, F_d(X_d), \boldsymbol{\theta})$ を導出した上で，(1) のような漸近正規性をもつ推定量に one-step 更新を行った推定量

$$\hat{\boldsymbol{\theta}}_n^{\mathrm{OSE}} = \hat{\boldsymbol{\theta}}_n^{\mathrm{PLE}} + \frac{1}{n} \sum_{i=1}^n \boldsymbol{\phi}^{\mathrm{E}}(\boldsymbol{x}_i, F_1, \dots F_d, \hat{\boldsymbol{\theta}}_n^{\mathrm{PLE}})$$
$$\boldsymbol{\phi}^{\mathrm{E}} = \mathbb{E}[(s^{\mathrm{E}})^\top \boldsymbol{s}^{\mathrm{E}}] \boldsymbol{s}^{\mathrm{E}}$$

が漸近有効性をもつことを示した．Gaussian コピュラ C_θ は 1 次元標準正規分布の分布関数 Φ，相関行列を $R(\theta)$ とする d 次元正規分布の分布関数 $\Phi_\theta \sim N(\boldsymbol{0}, R(\boldsymbol{\theta}))$ によって

$$C(F_1(x_1), \dots, F_d(x_d); \boldsymbol{\theta}) = \Phi_\theta(\Phi^{-1}(F_1^{-1}(X_1), \dots, \Phi^{-1}(F_d^{-1}(X_d)))$$

とかけるコピュラのことである．
本発表では，Gaussian コピュラモデルにおいて有効スコア関数を推定関数としたときの推定量，すなわち

$$\sum_{i=1}^n \boldsymbol{s}^{\mathrm{E}}(x, \boldsymbol{\theta}, \hat{F}_{n1}(x_{i1}), \dots, \hat{F}_{nd}(x_{id})) = 0$$

の解として得られる推定量 $\hat{\boldsymbol{\theta}}_n^{\mathrm{M}}$ の漸近有効性と，有限の n における数値実験結果について報告する．

参考文献

[1] C. Genest, K. Ghoudi and L.-P. Rivest (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions., *Biometrika* **82**, pp.534-552.

[2] J. Segers, R. van den Akker and B. J. M. Werker (2014) Semiparametric Gaussian copula models: geometry and efficient rank-based estimation, *The Annals of Statistics*, **42(5)**, pp.1911-1940.