

# セミパラメトリックモデルにおける 有効スコア関数の近似による推定量の構成

数理情報学専攻 48176224

松本 和也

指導教員 駒木 文保 教授

## 1 はじめに

セミパラメトリックモデルは、興味ある有限次元のパラメータと局外パラメータの2種類のパラメータによって特徴づけられるモデルである。局外パラメータは確率密度関数、条件付き期待値などの無限次元のパラメータとなることがある。

セミパラメトリックモデルでのパラメータ推定では、スコア関数の局外接空間への直交射影として有効スコア関数が定義される。有効スコア関数は漸近有効性をもつ推定量の構成や推定量の漸近分散の下限の評価において重要である。しかし、有効スコア関数を陽に求められるとは限らない。

本研究では、セミパラメトリックモデルにおいて、有効スコア関数を近似することによってパラメータ推定を行う手法を提案した。さらに、提案手法をセミパラメトリックコピュラモデルに適用して、ランク統計量からパラメータ推定を行う方法を提案した。提案手法により得られる推定量がモデルに対しての仮定の下で漸近有効性を持つことを示した。また数値実験によって、有限標本の場合に提案手法による推定量の二乗誤差が既存の推定量より小さくなることを確かめた。

## 2 既存研究

### 2.1 有効スコアを用いたパラメータ推定

セミパラメトリックモデルにおける推定では、有効スコア関数  $s^E$  が重要である。興味あるパラメータ  $\theta$  が与えられたもとの局外パラメータ  $k$  の推定量  $\tilde{k}(\theta)$  がある制約をみたすとき、有効スコア関数に関する方程式

$$\sum_{i=1}^N s^E(x_i, \theta, \tilde{k}(\theta)) = 0$$

の解として得られる推定量  $\hat{\theta}_N$  は漸近有効性をもつ [1]。また、本研究で扱うセミパラメトリックコピュラモデルのうちガウシアンコピュラについて、有効スコアを用いた漸近有効性をもつ推定量が知られている [4]。しかし、有効スコア関数を解析的に求められないモデルが多く存在する。

### 2.2 コピュラモデル

$\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$  を  $d$  次元の確率変数とし、 $\mathbf{X}$  の同時分布関数  $F: \mathbb{R}^d \rightarrow [0, 1]$  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  $F$  に対応する  $[0, 1]^d$  上の確率分布関数  $C$  が一意に定まる:

$$F(\mathbf{x}; F_1, \dots, F_d) = C(F_1(x_1), \dots, F_d(x_d)). \quad (1)$$

この  $C$  をコピュラとよぶ。  $C$  の密度関数を  $c$  とする。

式 (1) は同時分布を周辺分布と変数間の依存関係を表すコピュラ  $C$  に分離している。このようにコピュラは変数間の依存関係を表現する関数で、ファイナンスなどの分野に応用される。

$C$  が有限次元の  $\theta \in \Theta \subset \mathbb{R}^m$  で特徴づけられるとき、

$$F(\mathbf{x}; \theta, F_1, \dots, F_d) = C(F_1(x_1), \dots, F_d(x_d); \theta) \quad (2)$$

とかける。式 (2) で表せる分布の集合  $\{F(\mathbf{x}; \theta, F_1, \dots, F_d)\}$  は、 $\theta, F_1, \dots, F_d$  で特徴づけられるモデルとなる。 $F_1, \dots, F_d$  を無限次元の局外変数とみなすと、 $\theta$  を興味あるパラメータとするセミパラメトリックなコピュラモデルとみることができる。

### 2.3 コピュラモデルにおけるパラメータ推定

$N$  個の実現値  $\mathbf{x}_i = (x_{i1}, \dots, x_{id})$ , ( $1 \leq i \leq N$ ) からパラメータ  $\theta$  を推定する問題を考える。コピュラの性質から、推定量がランク統計量  $\mathbf{R}_{ij} = \sum_{k=1}^N \mathbb{1}\{X_{ik} \leq X_{kj}\}$  の可測関数であることが望ましい条件となる。以後、このような性質を rank-based と表記する。

セミパラメトリックコピュラモデルにおける rank-based な  $\theta$  の推定量  $\hat{\theta}_N$  の構成法として、周辺分布に経験分布をプラグインして  $\theta$  を最尤推定する方法 [3]

$$\tilde{F}_{Nj}(x) = \frac{1}{N+1} \sum_{i=1}^N \mathbb{1}\{x_{ji} \leq x\},$$

$$\hat{\theta}_N^{\text{PLE}} = \operatorname{argmax}_{\theta \in \Theta} \left[ \sum_{i=1}^N \log c(\tilde{F}_{N1}(x_{i1}), \dots, \tilde{F}_{Nd}(x_{id}); \theta) \right] \quad (3)$$

などが知られている。しかし、これらの推定量は漸近有効性をもたない。逆に、既知の漸近有効性をもつ推定量 [2] は rank-based ではない。

コピュラモデルの1クラスであるガウシアンコピュラモデルについては、[4] が有効スコアを導出した。さらに有効スコアと式 (3) の推定量  $\hat{\theta}_N^{\text{PLE}}$  を用いて rank-based かつ漸近有効性をもつ推定量を構成した。 $\mathbf{x}$  にお

ける有効スコアの値を

$$\alpha(x) = s^E(x, \hat{\theta}_N^{\text{PLE}}, \tilde{F}_{N1}, \dots, \tilde{F}_{Nd})$$

と表すと, [4] による推定量は次のように書ける:

$$\hat{\theta}_N = \hat{\theta}_N^{\text{PLE}} + \frac{1}{N} E_{\hat{\theta}_N^{\text{PLE}}} \left[ \alpha(x) \alpha(x)^\top \right]^{-1} \sum_{i=1}^N \alpha(x_i). \quad (4)$$

$E_{\hat{\theta}_N^{\text{PLE}}}$  はパラメータ  $\hat{\theta}_N^{\text{PLE}}, F_1, \dots, F_d$  での期待値である.

ガウシアンコピュラ以外では rank-based かつ漸近有効性をもつ推定量は知られていない. この理由として, 有効スコア  $s^E$  が陽に求められていないこと,  $s^E$  を得られたとしても式 (4) の形の推定量が漸近有効性をもつか明らかでないことが挙げられる. 一般のセミパラメトリックコピュラモデルにおいて, rank-based かつ漸近有効性をもつ推定量が構成可能かは未解決である.

### 3 提案手法

#### 3.1 提案手法の概要

有効スコア関数  $s^E$  は, スコア関数  $s$  をある無限次元空間  $\Lambda$  へ直交射影することで得られる. スコア関数  $s$  は確率密度関数の  $\theta$  についての偏微分である. 本研究では, 直交射影をする空間を有限次元とする近似により有効スコア関数を近似し, 近似した有効スコアを真の有効スコアの代わりに用いる推定方法を提案した.

#### 3.2 提案手法のコピュラモデルへの適用

$\mathcal{L}_2^0 := \left\{ e : [0, 1] \rightarrow \mathbb{R} \mid \int_0^1 e(x) dx = 0, \int_0^1 e(x)^2 dx < \infty \right\}$  とする. コピュラモデルにおいては,  $e \in \mathcal{L}_2^0$  が  $\Lambda$  の元と 1 対 1 に対応する. 提案手法では,  $\mathcal{L}_2^0$  を近似する有限次元空間  $\mathcal{S}_{J_N} \subset \mathcal{L}_2^0$  をとり,  $e \in \mathcal{S}_{J_N}$  に対応する  $\Lambda$  の元の集合を  $\Lambda_N$  とする. 有効スコア関数の近似  $\hat{s}^E$  を,  $s$  の有限次元空間  $\Lambda_N$  への直交射影により定義する.  $x$  における有効スコアの推定量の値を

$$\hat{\alpha}(x) = \hat{s}^E(x, \hat{\theta}_N^{\text{PLE}}, \tilde{F}_{N1}, \dots, \tilde{F}_{Nd})$$

と表すと, 提案手法による推定量は次のように書ける:

$$\hat{\theta}_N = \hat{\theta}_N^{\text{PLE}} + \frac{1}{N} \left[ \frac{1}{N} \sum_{i=1}^N \hat{\alpha}(x_i) \hat{\alpha}(x_i)^\top \right]^{-1} \sum_{i=1}^N \hat{\alpha}(x_i). \quad (5)$$

### 4 数値実験

有効スコアが既知のガウシアンコピュラモデルと, 有効スコアが未知の Clayton コピュラモデルの 2 種類について数値実験を行った. ある真の分布から  $N$  個の標本を発生させ, [3] の方法 (式 (3)) と提案手法 (式 (5)) により推定量を求める実験を  $k = 8000$  回行った. 各実験の推定量を  $\hat{\theta}_N^{(1)}, \dots, \hat{\theta}_N^{(k)}$  として以下を比較した:

$$\text{Mean} = \frac{\sum_{i=1}^N \hat{\theta}_N^{(i)}}{N} \quad \text{Bias} = \text{Mean} - \theta_0$$

$$\text{Var} = \frac{\sum_{i=1}^N (\hat{\theta}_N^{(i)} - \text{Mean})^2}{N} \quad \text{MSE} = \frac{\sum_{i=1}^N (\hat{\theta}_N^{(i)} - \theta_0)^2}{N}$$

図 1. に示すように特に変数間の相関が大きいとき, 提案手法の推定量が優れた結果を示した.

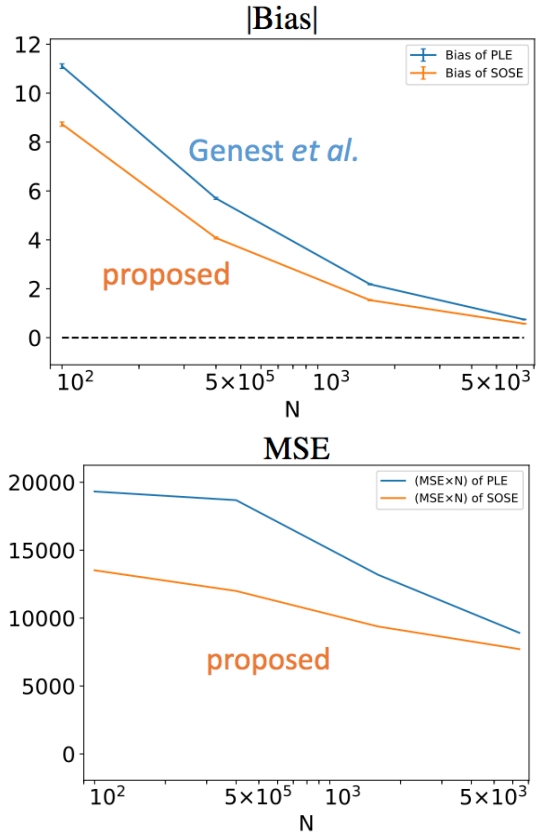


図 1. Clayton モデル,  $\theta_0 = (55)$ . Bias · MSE の標本数依存性.

### 5 まとめ

本研究では, セミパラメトリックモデルのパラメータ推定において, 近似した有効スコア関数を用いる手法を提案した. 提案手法をセミパラメトリックコピュラモデルに適用し, 従来より広いクラスに対して rank-based かつ漸近有効性を持つ推定量を提案した. 数値実験により変数間の相関が大きい場合に提案手法による推定量の二乗誤差が小さくなることを確認した.

### 参考文献

- [1] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore, 1993.
- [2] X. Chen, Y. Fan, and V. Tsyrennikov. Efficient Estimation of Semiparametric Multivariate Copula Models. *J. Am. Stat. Assoc.*, 101: 1228–1240, 2006.
- [3] C. Genest, K. Ghoudi, and L.-P. Rivest. A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. *Biometrika*, 82: 543–552, 1995.
- [4] J. Segers, R. van den Akker, and B. J. M. Werker. Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann. Stat.*, 42: 1911–1940, 2014.