最小情報従属モデル

統計数理研究所　矢野　恵佑
(共同研究者：東京大学　清　智也)

最小情報従属モデルは, Sei and Yano (2024) により提案された, 混合ドメインデータにおける高次従属性を扱うための統計モデルである. 本発表では, その基礎から応用までを紹介する.

最小情報従属モデル

最小情報従属モデル (Mindemo; **M**inimum **in**formation **de**pendence **mo**del)は, Sei and Yano (2024) によって提案された混合ドメインデータに対する同時分布モデルである. $i = 1, \ldots, d$ について $(\mathcal{X}_i, \mathcal{F}(\mathcal{X}_i), \mathrm{d}x_i)$ を測度空間とし, それらの直積空間を $\mathcal{X} = \prod_{i=1}^d \mathcal{X}_i$, $\mathrm{d}x = \prod_{i=1}^d \mathrm{d}x_i$ と表す. 添字 i に対して、$-i$ は第 i 座標を取り除くことを表す記法として用いる.

次の統計モデル (1) を最小情報従属モデルと呼ぶ：

$$p(x;\theta,\nu) = \exp\left(\theta^\top h(x) - \sum_{i=1}^d a_i(x_i;\theta,\nu) - \psi(\theta,\nu)\right)\prod_{i=1}^d r_i(x_i;\nu). \tag{1}$$

ここで, $\theta \in \mathbb{R}^K$ は従属構造を表す K 次元パラメータであり, $h : \mathcal{X} \to \mathbb{R}^K$ は所与の関数である. また, $r_1(x_1;\nu), \ldots, r_d(x_d;\nu)$ は $\mathcal{X}_1, \ldots, \mathcal{X}_d$ 上の周辺密度の統計モデルであり, ν はそれらの周辺密度を特徴づけるパラメータを表す. 関数 $a_i(x_i;\theta,\nu)$ および $\psi(\theta,\nu)$ は、次の制約条件によって同時に定められる：

$$\int p(x;\theta,\nu)\mathrm{d}x_{-i} = r_i(x_i;\nu), \quad i = 1, \ldots, d, \text{ and } \int \sum_{i=1}^d a_i(x_i;\theta,\nu)p(x;\theta,\nu)\mathrm{d}x = 0. \tag{2}$$

なお, $\theta = 0$ のときは独立モデル $\prod_{i=1}^d r_i(x_i;\nu)$ に帰着される.

このモデルには以下の重要な性質がある： (a) 連続値, カテゴリ値, 多様体上の値など, さまざまなデータ型を自然に扱うことができる. (b) 高次相互作用や負の相互作用など, さまざまな種類の従属構造を扱うことができる.

本講演では, Mindemoの基礎から幅広い活用に向けた工夫・活用例について報告する.

参考文献

1. T. Sei and K. Yano (2024). Minimum information dependence modeling. *Bernoulli*, 30(4), pp.2623-2643.
2. T. Sei and K. Yano (2026). Estimation of bivariate minimum information dependence models using the density-ratio framework. *Annals of Institute of Statistical Mathematics*. `https://doi.org/10.1007/s10463-026-00995-8`
3. `https://github.com/kyanostat/min-info/`
4. `https://github.com/tanaken-basis/mindemo3d`
5. `https://huggingface.co/spaces/stardust-coder/MinDeMoDeMo`