セミパラメトリックコピュラモデル における ダイバージェンス

東京大学　　清 智也
東京大学　松本 和也

パラメトリックな d 次元コピュラ密度関数 c とノンパラメトリックな周辺分布関数 $F_1, \ldots, F_d$ によって

$$p(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \prod_{i=1}^{d} F_i'(x_i)$$

と定義される統計モデルをセミパラメトリックコピュラモデルという。ここで F_i' は F_i の導関数である。

このモデルには座標ごとの変数変換群が作用する。この作用に関する軌道を $[p]$ と表す。大きさ n の標本に対し，多変量順位統計量を R とおき，R の周辺分布 $\bar{p}_n(R)$ を順位尤度と呼ぶ。$\bar{p}_n(R)$ は軌道 $[p]$ のみに依存する。すると，2 つの密度 p, q に対して順位尤度から定まるダイバージェンス

$$D_n([p],[q]) = \frac{1}{n} \sum_{R} \bar{p}_n(R) \log \frac{\bar{p}_n(R)}{\bar{q}_n(R)}$$

を考えることができる。もう一つのダイバージェンスとして

$$\tilde{D}([p],[q]) = \inf_{q_* \in [q]} \mathrm{KL}(p, q_*) = \inf_{q_* \in [q]} \int p(x) \log \frac{p(x)}{q_*(x)} \mathrm{d}x$$

を定義できる。一般に $\mathrm{KL}(p,q) \geq \tilde{D}([p],[q]) \geq D_n([p],[q])$ が成り立つ。

$[0,1]^d$ 上の密度関数 p で，上下に有界（ある定数 $M > 0$ が存在して $M^{-1} \leq p(x) \leq M$）かつ連続なものの全体を $\mathcal{P}_0$ と表す。このとき次の定理が成り立つ。証明は [1] を参照されたい。

定理 1. $p, q \in \mathcal{P}_0$ かつ $[p] \neq [q]$ ならば $\tilde{D}([p],[q]) > 0$ である。

定理 2. $p, q \in \mathcal{P}_0$ ならば $D_n([p],[q]) \to \tilde{D}([p],[q])$ $(n \to \infty)$ である。

また p, q が多変量正規分布の場合，$\tilde{D}([p],[q])$ の計算は有限次元の凸最適化問題に帰着され，$D_n([p],[q])$ の計算は多変量正規分布の象限確率計算に帰着される。後者にはホロノミック勾配法を適用できる。

参考文献

[1] Sei, T. and Matsumoto, K. (2019). Properties of divergence for semiparametric copula models, Technical Repoart METR 2019-11, The University of Tokyo.