エントロピー最大化から得られる客観的総合指数

清 智也 （東京大学・情報理工）

この資料は発表後に差し替えたものです．変更点：「オギュラ」→「Stein 型」．

概 要

本研究では，多次元の確率変数ベクトルを 1 次元の総合指数にまとめる方法を考察する．ただし，総合指数の形は各変数の単調変換の和で与えられるものに限定する．Stein の等式に類似した条件から，一意的な総合指数が得られることを示す．その証明には，最適輸送理論におけるエントロピー汎関数最大化のアイデアが使われる．

キーワード：エントロピー最大化，客観的総合指数，コピュラ，最適輸送，Stein の等式．

1 主結果

d を正の整数とする．$\mathbb{R}^d$ 上の確率測度 μ で次の性質を満たすもの全体を $\mathcal{P}^2 = \mathcal{P}^2(\mathbb{R}^d)$ とおく：各 $1 \le i \le d$ に対し周辺分布 μ_i が絶対連続で，$\int x_i d\mu_i = 0$ かつ $\int x_i^2 d\mu_i < \infty$ が成り立つ．同時分布ではなく周辺分布に対して絶対連続性を課している点に注意しよう．

定義 1. $\mu \in \mathcal{P}^2$ とする．有界な導関数 f' を持つ任意の微分可能関数 $f : \mathbb{R} \to \mathbb{R}$ に対して

$$\int f(x_i) \left(\sum_{j=1}^{d} x_j \right) d\mu = \int f'(x_i) d\mu, \quad i = 1, \ldots, d, \tag{1}$$

が成り立つとき，μ を **Stein 型分布**と呼ぶことにする．

もし $d = 1$ ならば式 (1) は Stein の等式 $\int f(x_1) x_1 d\mu = \int f'(x_1) d\mu$ である [8]．よって 1 次元の Stein 型分布は標準正規分布 $d\mu = (2\pi)^{-1/2} e^{-x_1^2/2} dx_1$ しかない．同様に，μ が独立分布の場合，Stein 型分布は標準正規分布の直積に限られる．我々は独立でない場合に興味がある．

例 1. $W \sim N(0,1)$, $\mathrm{E}[U_i] = 0$, $\mathrm{E}[U_i^2] < \infty$, かつ $(U_1, \ldots, U_{d-1}) \perp\!\!\!\perp W$ を満たす確率変数 $W, U_1, \ldots, U_{d-1}$ を考える．また $U_d = -\sum_{j=1}^{d-1} U_j$ とおく．このとき

$$X_i = \frac{W}{\sqrt{d}} + U_i, \quad i = 1, \ldots, d,$$

とおけば，$X_1, \ldots, X_d$ は Stein 型分布に従う．実際，

$$\mathrm{E}[f(X_i) \sum_j X_j] = \mathrm{E}[f(W/\sqrt{d} + U_i)\sqrt{d}W] = \mathrm{E}[f'(W/\sqrt{d} + U_i)] = \mathrm{E}[f'(X_i)]$$

となる．2 つ目の等号は 1 変量正規分布に対する Stein の等式による．$W, U_1, \ldots, U_{d-1}$ が多変量正規分布にしたがう場合は [7] の Theorem 5 で考えられている．一方，例えば $U_1 = \cdots = U_d = 0$ とすると，分布は対角線上に退化するが，それでも等式 (1) は満たされる．□

分布 μ の写像 T による押し出し $T\sharp\mu$ を，$(T\sharp\mu)(A) = \mu(T^{-1}(A))$ によって定義する．

与えられた $\mu \in \mathcal{P}^2$ に対し，座標ごとの変数変換

$$T(x) = (T_1(x_1), \ldots, T_d(x_d)), \quad x \in \mathbb{R}^d,$$

で，各 $T_i : \mathbb{R} \to \mathbb{R}$ が単調非減少かつ $T\sharp\mu \in \mathcal{P}^2$ となるような T の全体を $\mathcal{T}_{\mathrm{cw}}(\mu)$ と表す[1]．

定義 2. $\mu \in \mathcal{P}^2$ とする．ある変換 $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$ によって $T\sharp\mu$ が Stein 型分布となるとき，T を **Stein 型変換**という．

μ の Stein 型変換 T は，存在するならば，μ-a.e. の意味で一意的であることが示される．問題は存在性である．そのために次の概念を導入する．

定義 3. $\mu \in \mathcal{P}^2$ が**共正値** (copositive) であるとは，μ によって定まる量

$$\beta(\mu) = \inf_{T \in \mathcal{T}_{\mathrm{cw}}(\mu)} \frac{\int (\sum_i T_i)^2 d\mu}{\sum_i \int T_i^2 d\mu}$$

が正になることとする．

例えば μ が独立分布の場合は $\int(\sum_i T_i)^2 d\mu = \sum_i \int T_i^2 d\mu$ となり，したがって $\beta(\mu) = 1$ となる．一方，$d = 2$ で $\mu(\{x \mid x_1 + x_2 = 0\}) = 1$ の場合，$\int (x_1 + x_2)^2 d\mu = 0$ なので $\beta(\mu) = 0$ となる．もし μ にコピュラ密度が存在して $[0,1]^d$ 全域で正であり，かつ全ての 2 次元周辺コピュラ密度が 2 乗可積分ならば共正値であることが示される（これは $d = 2$ の場合は [6] で示されている）．特に，退化していない多変量正規分布は共正値である．

次が本稿の主定理である．

定理 1. $\mu \in \mathcal{P}^2$ が共正値ならば，Stein 型変換 $T \in \mathcal{T}_{\mathrm{cw}}(\mu)$ が一意的に存在する．

与えられた同時確率分布に対する座標ごとの単調非減少変換としては，コピュラ変換がよく知られている．コピュラ変換とは $T_i\sharp\mu_i$ が一様分布となるような変換であり，μ_i が atom を持たない限り一意に存在する（Sklar の定理；例えば [4]）．上の定理は，共正値性の条件のもとでは，Stein 型分布がコピュラの代替となり得ることを示唆している．

2 応用：客観的総合指数

定理 1 の応用例を見る．確率変数 $X_1, \ldots, X_d$ が与えられており，その分布 μ は $\mathcal{P}^2$ に属すと仮定する．特に $\mathrm{E}[X_i] = 0, \mathrm{E}[X_i^2] < \infty$ である．これらの確率変数は，例えばある大学の学生の d 個の科目の成績を表す確率変数と思えば良い．本節で考えたいのは，これらの科目の総合点（総合指数）をどう決めるか，という問題である．

Marshall & Olkin の定理 [2] より，共分散行列 $S_{ij} = \int x_i x_j d\mu$ が非退化ならば，

$$\sum_{j=1}^{d} w_i S_{ij} w_j = 1, \quad i = 1, \ldots, d, \tag{2}$$

[1] cw は coordinate-wise の略である．また $\mathcal{T}_{\mathrm{cw}}(\mu)$ は実際には周辺分布 $\mu_1, \ldots, \mu_d$ のみから定まる．

を満たすような正の実数 $w_1, \ldots, w_d$ が一意に存在する．[7] では，この $w_1, \ldots, w_d$ を用いて

$$G = \sum_{i=1}^{d} w_i X_i$$

とおき，**客観的総合指数** (objective general index; **OGI**) と呼んだ．OGI は，式 (2) より全ての X_i と正の相関を持つ．その意味で公平な総合指数となっている．

ところで，$T_i(x_i) = w_i x_i$ とおき，$\nu = T\sharp\mu$ とおくと，式 (2) は次のように書き換えることができる：

$$\int x_i \sum_{j=1}^{d} x_j d\nu = 1.$$

これは Stein 型分布の定義式 (1) で $f(x) = x$ と置いた場合に他ならない．そこで次の定義を考えることは自然である．

定義 4. $(X_1, \ldots, X_d) \sim \mu \in \mathcal{P}^2$ とし，μ には Stein 型変換 T が存在すると仮定する．このとき $G = \sum_{i=1}^{d} T_i(X_i)$ を**関数 OGI** (functional OGI) と呼ぶ．

$Y_i = T_i(X_i)$ とおけば，Stein 型分布の条件より，任意の単調増加関数 f に対して $f(Y_i)$ と $G = \sum_{j=1}^{d} Y_j$ は正の共分散を持つ．この意味で関数 OGI は公平な総合指数になっている．なお上記の関数 OGI の定義は，[7] で発見的に導いた定義と一致する．

3 最適化問題への書き換え

定理 1 はエントロピー最大化に関係しており，さらに最適輸送理論（例えば [5] や [9]）の観点からも綺麗な構造を持っている．これらを説明しよう．

負の周辺エントロピーに「対角方向の 2 次モーメント」を加えたエネルギー汎関数を

$$\mathcal{E}(\mu) = \sum_{i=1}^{d} \int \frac{d\mu_i}{dx_i} \log\left(\frac{d\mu_i}{dx_i}\right) dx_i + \frac{1}{2} \int \left(\sum_{i=1}^{d} x_i\right)^2 d\mu, \quad \mu \in \mathcal{P}^2,$$

と定義する．条件 $\int x_i^2 d\mu_i < \infty$ と KL ダイバージェンスの性質から $\mathcal{E}(\mu) > -\infty$ が言える．そして，与えられた μ に対して次の最小化問題を考える：

$$\text{Minimize} \quad \mathcal{E}(T\sharp\mu) \quad \text{subject to} \quad T \in \mathcal{T}_{\mathrm{cw}}(\mu). \tag{3}$$

この最小化問題は,「対角方向の 2 次モーメント」を一定値に固定した下で，$T\sharp\mu$ のエントロピーを最大化する問題と本質的には等価である．

式 (3) は狭義凸関数を目的関数とする凸最適化問題になっており，最小解が存在するならばそれは一意である．この凸性は最適輸送理論では (generalized) displacement convexity と呼ばれるものである [3, 1]．ところが $d \geq 2$ のとき $\mathcal{E}$ は定義域 $\mathcal{P}^2$ 全体では下に有界でないため，制約が加わった式 (3) に最適解が存在するかどうかは自明ではない．

さて式 (3) の停留条件を考えることにより，次の定理を得る．

定理 2. $\mu \in \mathcal{P}^2$ とし，$T \in \mathcal{T}_{\mathrm{cw}}(\mu)$ とする．このとき次の 2 つの条件は同値である．

(i) $T\sharp\mu$ は Stein 型である．
(ii) $T\sharp\mu$ は $\mathrm{dom}\,\mathcal{E} = \{\nu \mid \mathcal{E}(\nu) < \infty\}$ に属し，かつ式 (3) の最適解である．

定理 2 は，変数変換の微分可能性などの議論を除けば，変分法によって比較的容易に導かれる．非自明なのは，μ の共正値性のもとで式 (3) の最小化問題には解が存在することであり，これを示すことによって定理 1 が証明される．

以上のように問題を確率分布の空間上の最適化問題に書き換える利点の一つは，確率分布の弱収束に関する諸定理を使える点である．実際，定理 1 の証明でもこの利点が使われる．

4 今後の課題

本稿では Stein 型分布の応用や Stein 型変換の存在性について議論したが，その統計モデルについては扱わなかった．例 1 の Stein 型分布は，密度の形が陽に得られる，現状ではほぼ唯一のクラスであり，統計モデルの候補になり得る．

また，裾従属性を持つコピュラについては，それが共正値かどうかがまだ不明である．

最後に Stein 型分布に関する予想を述べて本稿を締めくくる．

予想 1. 任意の Stein 型分布に対し，その周辺分布のサポートは $\mathbb{R}$ 全域である．

参考文献

[1] Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient Flows – in Metric Spaces and in the Space of Probability Measures*, Birkhäuser.

[2] Marshall, A. W., Olkin, I., (1968). Scaling of matrices to achieve specified row and column sums. *Numer. Math.*, 12, 83–90.

[3] McCann, R. J. (1997). A convexity principle for interacting gases, *Adv. Math.*, **128** (1), 153–179.

[4] Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer.

[5] Rachev, S. T. and Rüschendorf, L. (1998). *Mass Transportation Problems I: Theory*, Springer-Verlag.

[6] Rényi (1959), On measures of dependence, *Acta Math.*, **10**, 441–451.

[7] Sei, T. (2016). An objective general index for multivariate ordered data, *J. Multivariate Anal.*, 147, 247–264.

[8] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, Vol. 2, 583–602.

[9] Villani, C. (2003). *Topics in Optimal Transportation*, American Mathematical Society.