

A note on estimation of the shape of density level sets of star-shaped distributions

大阪経済大・経済 紙屋英彦

ランダムベクトル $x \in \mathcal{X} := \mathbb{R}^p \setminus \{\mathbf{0}\}$ ($p \geq 2$) は,

$$x \sim h(r(x)) dx$$

と分布するとする. ここで $r: \mathcal{X} \rightarrow \mathbb{R}_{>0}$ は連続で, $r(cx) = cr(x)$ ($x \in \mathcal{X}$, $c \in \mathbb{R}_{>0}$) を満たすとする. また $h: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ は, $0 < c_0 < \infty$ を満たすとする. ただし $c_0 := \int_0^\infty h(r)r^{p-1}dr$ である. 一般性を失うことなく, $c_0 = 1$ となるように $h(\cdot)$ と $r(\cdot)$ が取られているとする.

ここで

$$Z := \{x \in \mathcal{X} : r(x) = 1\} \subset \mathcal{X}$$

と定義する. このとき, 集合 $\mathcal{Z} := \bigcup_{0 < c \leq 1} cZ \subset \mathcal{X} \cup \{\mathbf{0}\} = \mathbb{R}^p$ は原点に関する星型集合となり, Z は $\mathcal{Z}$ の境界となる. ここで $c \in \mathbb{R}_{>0}$ に対して, $cZ = \{x \in \mathcal{X} : r(x) = c\}$ となる. それゆえ, 各 $cZ \subset \mathcal{X}$ ($c \in \mathbb{R}_{>0}$) 上で, 密度 $h(r(x))$ は一定となる: $h(r(x)) = h(c)$ ($x \in cZ$). このことより $h(r(x))dx$ を星型分布という. また Z をこの星型分布の密度の等高線の“形”と呼ぶ.

特に, 正定値行列 Σ に対して $r(x) = \omega_p^{1/p}(\det \Sigma)^{1/(2p)}(x^T \Sigma^{-1} x)^{1/2}$ ($\omega_p := 2\pi^{p/2}/\Gamma(p/2)$ は $\mathbb{S}^{p-1}$ の全体積) と取れば, 星型分布は楕円型分布となり, そのときさらに $h(r) = (2\pi)^{-p/2}(\det \Sigma)^{-1/2} \exp[-r^2/\{2\omega_p^{2/p}(\det \Sigma)^{1/p}\}]$ と取れば, 多変量正規分布 $N_p(\mathbf{0}, \Sigma)$ となる. 一般の星型分布では, $r(\cdot)$ (従って Z) の取り方により分布の歪みを, $h(\cdot)$ の取り方により裾の重さを表すことができる.

ここでは, 密度の等高線の形 Z の推定を考える.

$x_1, \dots, x_n \in \mathcal{X}$ は, 独立に同一の星型分布 $h(r(x))dx$ に従うとする. $x \sim h(r(x))dx$ の方向ベクトル $u := x/\|x\| \in \mathbb{S}^{p-1}$ の分布の du ($\mathbb{S}^{p-1}$ の体積要素) に関する密度 $f(u)$ に対し, $\hat{f}_n(u) = \{C(\eta)/(n\eta^{p-1})\} \sum_{i=1}^n L((1-u^T u_i)/\eta^2)$ (Hall, Watson and Cabrera (1987), Bai, Rao and Zhao (1988)) を $u_i := x_i/\|x_i\| \in \mathbb{S}^{p-1}$ ($i = 1, \dots, n$) に基づくカーネル推定量とする. ここで $\eta = \eta_n > 0$, $C(\eta) := \eta^{p-1} / \int_{\mathbb{S}^{p-1}} L((1-u^T y)/\eta^2) du > 0$ ($y \in \mathbb{S}^{p-1}$) であり, $L: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ はいくつかの条件を満たすとする ($L(s) = \exp(-s)$ や $L(s) = 1(s < 1)$ は条件を満たす). さらに, $n \rightarrow \infty$ のとき, $\eta_n \rightarrow 0$, $n\eta_n^{p-1}/\log n \rightarrow \infty$ とする.

ここで, Z の推定量として,

$$\hat{Z}_n := \left\{ \hat{f}_n(u)^{\frac{1}{p}} u : u \in \mathbb{S}^{p-1} \right\}$$

を提案する. このとき $\hat{Z}_n$ は, ハウスドルフ距離 $\delta_H(\cdot, \cdot)$ に関し, Z の強一致推定量となる: $\delta_H(\hat{Z}_n, Z) \rightarrow 0$ a.s.

当日は, シミュレーション結果, 簡単な場合のパラメトリックな推定との比較, 実データへの適用例も示す.

本報告について詳しくは, Kamiya, H., A note on estimation of the shape of density level sets of star-shaped distributions. *Communications in Statistics—Theory and Methods*, to appear を参照されたい.