

一般化平均による統計モデル

A statistical model via generalized mean

数理・推論研究系 江口 真透 (Shinto Eguchi) ,

1. 一般化平均と情報幾何

1930 年に, コルモゴロフと南雲によって独立に発表された一般化平均について考察する. 正の数 x, y に対する一般化平均は

$$(1.1) \quad \text{GM}_\phi(x, y) = \phi((1 - \pi)\phi^{-1}(x) + \pi\phi^{-1}(y))$$

と定まる. ここで $\phi: \mathbb{R} \rightarrow (0, \infty)$ は単調関数とする. 同様に x, y, z の一般化平均を考えると, 定義から $\text{GM}_\phi(x, \text{GM}(y, z)) = \text{GM}_\phi(\text{GM}(x, y), z)$ がいえる. 典型例の算術平均, 幾何平均, 調和平均などが含まれるが, ‘平均’ の持つべき公理から特徴付けられた一般化平均は生成関数 ϕ の関数自由度を持つので多様な平均の考えが展開できる.

統計学のために一般化平均を積極的に援用した方法を提案した. はじめに情報幾何との関連について考察し, つぎに統計モデリングのために幾つかの応用を紹介する. ロジスティック回帰や比例ハザードモデルにおける予測関数を一般化平均を使った準線形モデルを提案した. さらにクラスタリングのエネルギー関数やロス関数の混合のために一般化平均がキーになることが示された.

情報幾何は確率密度の関数空間の上に双対リーマン幾何をベースに豊かな直観を与え, 確率に関連する全ての分野へ幾何的考えを構築している. その中の基本定理として, この関数空間の上でのピタゴラス定理が挙げられる (Amari-Nagaoka, 2007). 確率密度関数 $p(x)$ と $q(x)$ を混合測地線をつなぎ, 一方で, $r(x)$ と $q(x)$ を指数測地線をつないだとき, この 2 つの測地線が $q(x)$ で直交するならば, またその時に限り,

$$(1.2) \quad D_0(p, r) = D_0(p, q) + D_0(q, r)$$

が成立する. ここで D_0 は KL ダンバージェンスとする. この性質から最尤推定と十分統計量の関係, 赤池情報量規準の妥当性などが導かれる. この考察において $r(x)$ と $q(x)$ をつなぐ指数測地線とは

$$(1.3) \quad \text{EG}(q, r) = \exp((1 - \pi) \log q(x) + \pi \log r(x) - \kappa(\pi))$$

と定められる. ここで $\kappa(\pi)$ は正規化定数とする. このように $\text{EG}(q, r)$ は生成関数 $\phi = \exp$ を使って正の数の代わりに密度関数に対する一般化平均と見れる. したがって, 一般の ϕ に

$$(1.4) \quad \text{EG}_\phi(q, r) = \phi((1 - \pi)\phi^{-1}q(x) + \pi\phi^{-1}r(x) - \kappa_\phi(\pi))$$

が定まり, これを一般化指数測地線と呼ぶ (Eguchi-Komori, 2015). 同様な考えから一般化 KL ダイバージェンスを導出すると, ピタゴラス定理が示される. 次の節では一般化平均を直接に統計モデリングに応用することを考察する.

2. 準線形ロジスティック回帰モデル

一般化平均 (1.1) は正の数の平均で在ったが, 実数 x と y に対しては

$$(2.1) \quad \text{RGM}_\phi(x, y) = \phi^{-1}((1 - \pi)\phi(x) + \pi\phi(y))$$

と定める. (1.1) の生成関数 ϕ の代わりに ϕ^{-1} を取られていることに注意する. 統計の応用としては実数の代わりに, 回帰関数, 予測関数, エネルギー関数, ロス関数などの実数値関数の一般化平均を考えることができる. 例えば, ロジスティック回帰において p 変数の説明変数 $X = x$ を与えたとき 2 値反応変数 y の条件付き確率関数を

$$(2.2) \quad p(y|x) = \frac{\exp\{yf(x)\}}{1 + \exp\{f(x)\}} \quad (y = 0, 1)$$

とする. 予測関数を一般化平均によって

$$(2.3) \quad f_\tau(x, \beta, \pi) = \frac{1}{\tau} \log \left(\sum_{k=1}^K \pi_k \exp(\tau \beta_k^\top x_k) \right)$$

とし, 準線形予測関数と呼ぶ. ここで τ は逆温度パラメータ, $x = (x_1, \dots, x_K)$ と $\beta = (\beta_1, \dots, \beta_K)$ は p 変数の同じ K 分割とする. Cf. Omae et al. (2017). 逆温度パラメータ τ を極限 ∞ を取ると $f_\tau(x, \beta, \pi) = \max_{1 \leq k \leq K} \beta_k^\top x_k$ となり, 極限 $-\infty$ を取ると $f_\tau(x, \beta, \pi) = \min_{1 \leq k \leq K} \beta_k^\top x_k$ となる. 極限 0 を取ると線形予測関数に帰着される.

$Y = 0$ のときの X の条件分布 $p(x|Y = 0)$ は正規分布 $N(\mu_0, \Sigma)$ に従っているが, $Y = 1$ のときの X の条件分布 $p(x|Y = 1)$ は混合正規分布 $\sum_{k=1}^K \pi_k^* N(\mu_k, \Sigma)$ に従っていると仮定する. これは $Y = 0$ サンプルは均一な母集団から得られたが $Y = 1$ サンプルは非均一な異質な母集団の混合から得られた状況を考えている. このとき, $\tau = 1$ のとき

$$(2.4) \quad f_\tau(x, \beta, \pi) = \log \frac{p(x|Y = 1)}{p(x|Y = 0)}$$

が成立する. このような仮定のもとでは, 準線形予測関数の判別の最適性が示される.

準線形予測関数の定義において説明変数 x の K 分割が必要であるが, これは教師なし学習によって構成できる. 典型的にはクラスター分析によって K 分割が求まる. またスパース学習を準線形モデル (2.3) をロジスティック回帰 (2.2) に代入したモデルの対数尤度関数に (2.3) のパラメータ β_k の積の L_1 ペナルティを入れたものを考える方法も有力である.

比例ハザードモデル, 分割クラスタリング, 混合ロス関数, モデル平均, メタアナリシス, 半教師学習などについても一般化平均を使うと興味深い展開ができると思われる. 線形モデルから準線形モデルへの拡張や, エネルギー関数やロス関数の混合について, 幾つかの新しい知見が得られた. 広い実用のためには未だ多くの解決すべき問題があるが, 線形モデルを柔軟に結合する統計方法として着実な完成に近い将来になされることが望まれる.

参考文献

Amari, S. I., & Nagaoka, H. (2007). *Methods of information geometry* (Vol. 191). American Mathematical Soc.

Eguchi, S., & Komori, O. (2015). Path connectedness on a space of probability density functions. In *International Conference on Networked Geometric Science of Information* (pp. 615-624). Springer, Cham.

Omae, K., Komori, O., & Eguchi, S. (2017). Quasi-linear score for capturing heterogeneous structure in biomarkers. *BMC bioinformatics*, 18(1), 308.