擬球スコアとその周辺

金森敬文 (名大)

概要

スコアとは確率分布に対する損失関数であり，統計的推論のさまざまな場面で用いられている．本発表では Hölder 不等式から導出されるスコアである擬球スコアを用いた統計的推論に関する最近の研究結果を報告する．なお本研究は藤澤洋徳氏 (統数研)，竹之内高志氏 (はこだて未来大) との共同研究である．

スコアによる統計的推論

統計的推論においてデータから確率分布を推定するとき，最尤法など score (scoring rule [3]) に基づく方法がよく用いられる．標本空間 $\mathcal{Z}$ 上の非負値関数の集合を $\mathcal{M}$ とするとき，次の条件を満たす $S : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ を score という [1]．

1. $f, g \in \mathcal{M}$ に対して $S(f, g) \geq S(f, f)$.
2. $\mathcal{Z}$ 上の確率密度 p, q に対して $S(p, q) = S(p, p)$ なら $p = q$.
3. p が確率密度のとき，$S(p, f)$, $f \in \mathcal{M}$ は $\mathbb{E}_{Z\sim p}[\ell(Z, f)]$ と表せる．

定義より $D(f, g) = S(f, g) - S(f, f)$ とすると $D(f, g) \geq 0$ が成り立つ．$D(f, g)$ を divergence という．最尤推定では $S(p, q) = -\mathbb{E}_{Z\sim p}[\log q(Z)]$ を用い，対応する divergence は Kullback-Leibler (KL) divergence とよばれる．

データ $z_1, \ldots, z_n \sim_{\text{i.i.d.}} p$ が観測されたとき，スコア $S(p, q)$ の p にデータの経験分布 $\widetilde{p}$ を代入すると

$$S(\widetilde{p}, q) = \frac{1}{n}\sum_{i=1}^{n} \ell(z_i, q)$$

となる．大数の法則から $S(\widetilde{p}, q)$ の値は $S(p, q)$ に近いと考えられるので，$S(\widetilde{p}, q)$ を小さくするような q はデータの分布 p に近いと期待される．このようにしてスコアを用いて統計的推論を行うことができる．

本発表では特に次の 2 つの score に着目する．記号 $\langle f\rangle$ は $\mathcal{Z}$ 上の測度に関する積分 $\int_{\mathcal{Z}} f d\mu$ (離散空間なら和) を意味する．

例 1 (密度冪スコア; Density-Power score [1])．パラメータ $\gamma > 0$ に対して

$$S_{\mathrm{DP}}(f, g) = \gamma\langle g^{1+\gamma}\rangle - (1 + \gamma)\langle f g^{\gamma}\rangle$$

とする．$\mathcal{M}$ 上の関数 $f \mapsto \langle f^{1+\gamma}\rangle$ の凸性から $S_{\mathrm{DP}}(f, g) \geq S_{\mathrm{DP}}(f, f)$ となる．等号が成立するとき $f = g$ が成り立つ．

例 2 (擬球スコア; Pseudo-Spherical score [2, 4])．パラメータ $\gamma > 0$ に対して

$$S_{\mathrm{PS}}(f, g) = -\frac{\langle f g^{\gamma}\rangle}{\langle g^{1+\gamma}\rangle^{\gamma/(1+\gamma)}}$$

とする．*Hölder* 不等式から $S_{\mathrm{PS}}(f, g) \geq S_{\mathrm{PS}}(f, f)$ となることが分かる．等号が成立するとき f と g は 1 次従属である．

[1] 厳密には，S の定義域は $\mathcal{M} \times \mathcal{M}$ を適切に制限した集合となる．また a.e. などは省略する．

上記の 2 つの score を適当にスケーリングすると，対応するダイバージェンスは $\gamma \to 0$ の極限で KL-divergence に一致する．また等式

$$\min_{c>0} S_{\mathrm{DP}}(p, cp_\theta) = -\{-S_{\mathrm{PS}}(p, p_\theta)\}^{1+\gamma}$$

が成り立つ．

本発表では，擬球スコアや密度冪スコアを用いた統計的推論について考察する．既存の推定法のロバスト性や計算効率を改善するために，上記のスコアを非正規化モデルと組み合わせて用いる統計的手法を提案し，その統計的性質について調べる．詳細は [5, 6] を参照のこと．

References

[1] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

[2] H. Fujisawa and S. Eguchi. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.*, 99(9):2053–2081, 2008.

[3] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378, 2007.

[4] I. J. Good. Comment on "measuring information and uncertainty," by R. J. Buehler. In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*, page 337339, Toronto: Holt, Rinehart and Winston, 1971.

[5] T. Kanamori and H. Fujisawa. Robust estimation under heavy contamination using unnormalized models. *Biometrika*, in press.

[6] T. Takenouchi and T. Kanamori. Empirical localization of homogeneous divergences on discrete sample spaces. NIPS 2015, to appear.