Location family における鞍点等式とその含意

データ科学研究系　多次元データ解析グループ
助教　　大西　俊郎

1　イントロダクション

1.1　鞍点等式とは

統計学では，尤度は大きければ大きいほどよい（尤度原理）とされ，損失は小さければ小さいほどいいとされる．このような望大項と望小項がバランスするとき，それを**鞍点等式**と呼ぶことにする．よく知られた例は，指数型分布族 $p(x;\,\theta)$ において最尤推定量 $\hat{\theta}_M$ に対して成立する等式

$$\log\frac{p(x;\,\hat{\theta}_M)}{p(x;\,\theta)}=\mathrm{KL}\big(p(y;\,\hat{\theta}_M),\,p(y;\,\theta)\big)$$

である．ここで，右辺は Kullback-Leibler ダイバージェンスである．

1.2　鞍点等式の例

正規分布 $N_n(\boldsymbol{\mu},\,I)$ における James-Stein 推定量 $\hat{\boldsymbol{\mu}}_S=\{1-(n-2)\|\boldsymbol{x}\|^{-2}\}\boldsymbol{x}$ も鞍点等式

$$\mathrm{E}\left[\log\frac{p(\boldsymbol{x};\,\hat{\boldsymbol{\mu}}_S)}{p(\boldsymbol{x};\,\boldsymbol{\mu})}-\mathrm{KL}\big(p(\boldsymbol{y};\,\hat{\boldsymbol{\mu}}_S),\,p(\boldsymbol{y};\,\boldsymbol{\mu})\big)\right]=0$$

を満たす $(n\geq 3)$．この等式は James-Stein 推定量 $\hat{\boldsymbol{\mu}}_S$ が最尤推定量 $\hat{\boldsymbol{\mu}}_M$ を優越することを意味している．なぜなら，定義により，

$$\log\frac{p(\boldsymbol{x};\,\hat{\boldsymbol{\mu}}_S)}{p(\boldsymbol{x};\,\boldsymbol{\mu})}<\log\frac{p(\boldsymbol{x};\,\hat{\boldsymbol{\mu}}_M)}{p(\boldsymbol{x};\,\boldsymbol{\mu})}$$

であるからである．鞍点等式を満たす推定量の中で最尤推定量は最悪なのである．

その他の例としては，正規分布の分散の条件付き最尤推定量が挙げられる．また，Bayes 予測問題における最適予測分布も鞍点等式を満たすことが知られている (Yanagimoto & Ohnishi, 2009).

1.3　本発表のねらい

$\mathbb{R}^n$ 上の location family $p(\boldsymbol{x}-\boldsymbol{\mu})$ の推定問題を推定関数の観点から論じる．推定関数に対して，(1) 対数尤度比および損失に対応する量を定義し，(2) 尤度推定関数，Stein 型推定関数および Bayes 型推定関数が鞍点等式を満たすことを示す．また，(3) これらの鞍点等式が含意する内容を議論する．

2　Location family における鞍点等式

2.1　対数尤度比および損失の定義

$\boldsymbol{U}(\boldsymbol{x})$ をベクトル場とし，推定関数 $\nabla\log p(\boldsymbol{x}-\boldsymbol{\mu})-\boldsymbol{U}(\boldsymbol{x})$ で位置パラメータ $\boldsymbol{\mu}$ を推定する．ここで，$\nabla=(\dots,\partial/\partial x_i,\dots)^T$ である．ベクトル場 $\boldsymbol{U}_M(x)=\boldsymbol{0}$ は尤度推定関数を与える．対数尤度

比および損失に対応する量をそれぞれ

$$L\big[\boldsymbol{U}(\boldsymbol{x})\big] = \|\nabla \log p(\boldsymbol{x}-\boldsymbol{\mu})\|^2 - \|\boldsymbol{U}(\boldsymbol{x})\|^2$$
$$D\big[\boldsymbol{U}(\boldsymbol{x})\big] = \|\nabla \log p(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{U}(\boldsymbol{x})\|^2$$

のように定義する．尤度推定関数は，‘対数尤度比’$L\big[\boldsymbol{U}(\boldsymbol{x})\big]$ を最大化する推定関数であり，鞍点等式

$$L\big[\boldsymbol{U}_M(\boldsymbol{x})\big] = D\big[\boldsymbol{U}_M(\boldsymbol{x})\big]$$

を満たす．

2.2 Stein 型推定関数が満たす鞍点等式

Coulomb ポテンシャルから導かれるベクトル場 $\boldsymbol{U}_S(\boldsymbol{x}) = \nabla \log \|\boldsymbol{x}\|^{2-n}$ のケースを考える．鞍点等式

$$\mathrm{E}\Big[L\big[\boldsymbol{U}_S(\boldsymbol{x})\big] - D\big[\boldsymbol{U}_S(\boldsymbol{x})\big]\Big] = 0$$

が成立する．これは，Stein 型推定関数が尤度推定関数を優越することを意味している．実際，この等式は，Ohnishi & Yanagimoto (2003) が導出したピタゴラス関係

$$\mathrm{E}\Big[D\big[\boldsymbol{U}_M(\boldsymbol{x})\big] - D\big[\boldsymbol{U}_S(\boldsymbol{x})\big] - \|\boldsymbol{U}_M(\boldsymbol{x}) - \boldsymbol{U}_S(\boldsymbol{x})\|^2\Big] = 0$$

と等価である．

2.3 Bayes 型推定関数が満たす鞍点等式

事前分布 $\pi(\boldsymbol{\mu})$ を仮定し，対応する事後分布 $\pi(\boldsymbol{\mu}\,|\,\boldsymbol{x})$ の下での期待値を $\mathrm{E}_{\mathrm{post}}[\,\cdot\,]$ と書く．ベクトル場 $\boldsymbol{U}_B(\boldsymbol{x}) = \mathrm{E}_{\mathrm{post}}[\,\nabla \log p(\boldsymbol{x}-\boldsymbol{\mu})\,]$ を使った推定関数 $\nabla \log p(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{U}_B(\boldsymbol{x})$ を考える．この推定関数は，鞍点等式

$$\mathrm{E}_{\mathrm{post}}\Big[L\big[\boldsymbol{U}_B(\boldsymbol{x})\big] - D\big[\boldsymbol{U}_B(\boldsymbol{x})\big]\Big] = 0$$

を満たす．また，損失 $D\big[\boldsymbol{U}(\boldsymbol{x})\big]$ の下で最適である．実際，ピタゴラス関係

$$\mathrm{E}_{\mathrm{post}}\Big[D\big[\boldsymbol{U}(\boldsymbol{x})\big] - D\big[\boldsymbol{U}_B(\boldsymbol{x})\big] - \|\boldsymbol{U}(\boldsymbol{x}) - \boldsymbol{U}_B(\boldsymbol{x})\|^2\Big] = 0 \tag{2.1}$$

が成立する．事後平均の意味で鞍点等式を満たす推定関数の中で，Bayes 型推定関数は最適であり，尤度推定関数は最悪である．

続きは当日発表する．

参考文献

Ohnishi, T. & Yanagimoto, T. (2003). Electrostatic views of Stein-type estimation of location vectors. *Journal of the Japan Statistical Society*, **33**, 39–64.

Yanagimoto, T. & Ohnishi, T. (2009). Bayesian prediction of a density function in terms of e-mixture. *Journal of Statistical Planning and Inference*, **139**, 3064–3075.