鞍点予測子から見た informative 事前分布の有用性

中央大学　　柳本 武美

1. 序:

Bayes 推論では、可能な限り informative な事前分布を仮定して、より深い解析結果を求めるのが原則である。しかし、頻度法は勿論、近年実際的重要性が増している経験 Bayes 法でも、必ずしも informative 事前分布を仮定する枠組みにはなっていない。経験 Bayes 法の発展を考えれば望ましくない現状と考えられる。そこで鞍点予測子の視点から、事前分布を柔軟に設定できる経験 Bayes 法の枠組みを議論する。記述を簡単にするために標本分布が指数分布族に入ることを仮定する

2. 鞍点予測子：

標本 $x \in R^n$ が密度関数 $p(x;\theta)$ をもつ確率変数の実現値としたとき、標本から推定した密度関数 $p(y|x)$ を予測子と呼ぶ。Kullback [2] は、$\hat{\theta}_M$ を最尤推定量としたとき、任意の x に対して等式

$$\log\{p(x;\,\hat{\theta}_M)/p(x;\,\theta)\} \,=\, \mathrm{D}(p(y;\,\hat{\theta}_M),\,p(y;\,\theta)) \tag{1}$$

この等式 (1) では予測子の概念は明示的でなく、最尤推定量の簡便な性質として扱われてきた。事前密度を $\pi(\theta)$ と書いて、事後分布を $\pi(\theta|\,x)$ と書く。上の条件を弱めて予測子の条件として次の定義を与える

定義 1. *A predictor $p(y\,|\,x)$ is called a saddlepoint predictor, if*

$$E\left\{\log\frac{p(x\,|\,x)}{p(x;\,\theta)} \,-\, \mathrm{D}(p(y\,|\,x),\; p(y;\,\theta))\right|\left. p(x;\,\theta)\pi(\theta)\right\} \,=\, 0. \tag{2}$$

上式の左辺で第１項は望大項であり、第２項は望小項である。従って、(2) 式はバランスを取っていることを示すので、鞍点予測子と呼ぶ。最尤推定量を含めて、Stein 推定量、条件付 MLE 等から構成される予測子がこの条件を満たす ([4])。また、4 節で述べる DIC [3] の変形版が $-\log p(y|\,x)$ の不偏な推定量になるための条件でもある。

3. 最適性：

予測子 $p_e(y|\,x)$ を

$$p_e(y|\,x) \,=\, c(x)\exp[E\{\log p(y;\,\theta)\,|\,\pi(\theta\,|\,x)\}] \tag{3}$$

但し $c(x)$ は正規化定数とする。この予測子は、損失 $\mathrm{D}(p(y\,|\,x),\; p(y;\,\theta))$ とした Bayes リスクを最小にする意味で最適である ([1],[5])。更に、具体的に次の等式が成り立つ。

Proposition 1. *1) The predictor $p_e(y|\,x)$ is the minimizer of the Bayesian risk among saddlepoint predictors.*

2) When the sampling density is in the exponential family, the predictor $p(y;\,\theta_M)$ is the maximizer of the Bayesian risk among plug-in saddlepoint predictors.

Proposition 2. *It holds for a predictor $p(y|\,x)$ that*

$$E\{D(p(y|\,x),\,p(y;\,\theta)) \,-\, D(p(y|\,x),\,p_e(y|\,x)) \,-\, D(p_e(y|\,x),\,p(y;\,\theta))\,|\,\pi(\theta|\,x)\} = 0$$

when the left-hand side makes sense.

4. AIC と DIC：

Proposition 2 から

$$\mathrm{UDC}(\delta) \;=\; -2\log\{p(x;\,\hat{\theta})\} \;+\; 2\mathrm{E}\{\mathrm{D}(p(y;\,\hat{\theta}), p(y;\,\theta)) + \mathrm{D}(p(y;\,\theta), p(y;\,\hat{\theta}))|\,\pi(\theta|x,\delta)\}$$

とおくと母数に依存しないので統計量になり、$-2\log\{p(y;\hat{\theta})\}$ の不偏推定量であることを示している。近似は全く不要である。一方、AIC は

$$\mathrm{AIC} \;=\; -2\log\{p(x;\,\hat{\theta}_M)\} \;+\; 2\cdot\dim(\theta)$$

ただし、$\hat{\theta}_M$ は最尤推定量、$\dim(\theta)$ はベクトル θ の次元である。この量の解釈は色々であるが、$-2\log\{p(y;\hat{\theta}_M)\}$ の近似的な不偏推定量であると見なすことができる。実際、

$$\mathrm{AIC}' \;=\; -2\log p(x;\,\hat{\theta}_M) \;+\; 2\mathrm{D}(p(y;\,\hat{\theta}_M), p(y;\,\theta)) \;+\; 2\mathrm{D}(p(y;\,\theta), p(y;\,\hat{\theta}_M))$$

とおくと、統計量ではないが平均値は一致する。予測子 $p_e(y|\,x)$ は Bayes リスクを最小にするから、AIC$'$ と UDC を比べると後者の方が小さい値を取ることが期待される。AIC には Proposition 2 2) で述べたように、予測子 $p(y;\hat{\theta}_M)$ が Bayes リスクを最大にするという欠点がある。もし事前分布が無情報であれば、$p_e(y;\,\hat{\theta}) \doteq p(y;\,\hat{\theta}_M)$ となるが、それでも Bayes 予測子がよい場合が多い。更に、Bayes 推論のためには、可能な限り informative 事前分布を仮定するのがよいことになる。

一方、DIC (Deviance Information Criteria) は Spiegelhalter ら (2002) によって提案され

$$\mathrm{DIC}(\delta) \;=\; -2\log p(x;\,\check{\theta}) \;+\; 2p_D$$

ただし、$p_D \;=\; 2\mathrm{E}[\log\{p(x;\check{\theta})/p(x;\,\theta)\}|\,\pi(\theta|x,\delta)]$ で定義される。著者らは広く用いられているソフトウェア WinBugs の開発として著名であり、また多くのユーザーがこのソフトを通じて DIC に触れるている。この定義で特に注意されるのは、θ の選び方が特定されないことと、その推定法が特定されないことである。このために、p_D が負になることがあることが知られている。推定量として事後平均に限定して、母数としては自然母数を選ぶことにすれば、$p_D \;=\; \mathrm{E}\{\mathrm{D}(p(y;\,\hat{\theta}), p(y;\,\theta)))|\,\pi(\theta|x,\delta)\}$ となるから DIC は UDC に双対な divergence の違いを除いて一致する。この場合には、p_D が負になることがなくなる。

5. 周辺尤度最大化：

UDC の導出では、専ら母数に対する事前分布の仮定と事後平均の利用により構成される。周辺尤度は陽には使わない。一方、Smoothing, Lasso などの経験 Bayes 法では、事前密度は標本分布の一部の母数にのみしか仮定しないで、他の母数は超母数として扱う。超母数は周辺密度のモードで推定される。超母数が推定されると、事後モードあるいは事後平均で母数が推定される。こうした経験 Bayes 法の枠組みでは informative 事前分布を仮定するインセンティブに欠ける。

文献：[1] Corcuera, J.M. and Giummole F. (1999). *Scand. J. Statist.* 26, 265-279. [2] Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York. [3] Spiegelhalter, D.J. *et al.* (2002). *J. Roy. Statist. Soc.* 64, 583 - 639. [4] Yanagimoto, T. (1994). *Ann. Inst. Statist. Math.*, 46, 29-41. [5] Yanagimoto, T. and Ohnishi, T. (2009). *J. Statist. Plann. Inf.* 139, 3064-3075.