Title	Saddlepoint condition on a predictor and its implication (Statistical Information in Inference and Its Related Topics)
Author(s)	大西, 俊郎; 柳本, 武美
Citation	数理解析研究所講究録 (2011), 1758: 80-89
Issue Date	2011-08
URL	http://hdl.handle.net/2433/171320
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
第1758巻 2011年 80-89

Saddlepoint condition on a predictor and its implication

九州大学　大西 俊郎 (Toshio Ohnishi)
Kyushu University
中央大学　柳本 武美 (Takemi Yanagimoto)
Chuo University

1. イントロダクション

予測問題とは，将来の確率変数 y に対する確率密度 $p(y;\,\theta)$ をデータ x に基づく予測分布 $p(y\,|\,x)$ によって推定することである．特に事前分布 $\pi(\theta)$ を仮定するとき Bayes 予測問題と呼ばれる．損失関数としては普通 $\mathrm{D}\big(p(y;\,\theta),\,p(y\,|\,x)\big)$ が用いられるが，本稿では $\mathrm{D}\big(p(y\,|\,x),\,p(y;\,\theta)\big)$ に注目する．ここで $\mathrm{D}\big(p(x),\,q(x)\big)$ は Kullback-Leibler divergence であり，

$$\mathrm{D}\big(p(x),\,q(x)\big)=\mathrm{E}\left[\log\frac{p(x)}{q(x)}\,\middle|\,p(x)\right]$$

によって定義される．伝統的な記法とはいえないが，本稿では確率密度 $p(x)$ の下での $f(x)$ の期待値を $\mathrm{E}\big[f(x)\,\big|\,p(x)\big]$ と書くことにする．前者の損失 $\mathrm{D}\big(p(y;\,\theta),\,p(y\,|\,x)\big)$ は m-divergence，後者の損失 $\mathrm{D}\big(p(y\,|\,x),\,p(y;\,\theta)\big)$ は e-divergence と呼ばれることがある．

予測問題と推定問題との関係との関係を述べる．$\hat{\theta}$ を推定量とし，予測分布を $p(y\,|\,x)=p(y;\,\hat{\theta})$ という形に限定するとき，予測問題は推定問題と等価になる．このタイプの予測分布は plug-in 予測分布と呼ばれる．特に MLE $\hat{\theta}_M$ に基づく plug-in 予測分布 $p(y;\,\hat{\theta}_M)$ を MLE-induced 予測分布と呼ぶことにする．

本稿の主張の 1 つは Deviance Information Criterion (DIC) と呼ばれる Bayes 型モデル選択基準の正当化である．DIC は Spiegelhaleter *et al.* (2002) によって提案された Bayes 型モデル選択基準であり，

$$\mathrm{DIC}=-2\log p(x;\,\hat{\theta})+2p_D \tag{1.1}$$

のように計算される．ただし，$\hat{\theta}$ は適当な推定量であり，事後平均や事後モードなど利用可能なものを使うことになる．また，p_D は model complexity と呼ばれる量で，

$$p_D=2\,\mathrm{E}\left[\log\frac{p(x;\,\hat{\theta})}{p(x;\,\theta)}\,\middle|\,\pi(\theta\,|\,x)\right]$$

によって定義される．

DIC の問題点として次の 2 点を挙げることができる．(1) 推定量を限定していないことは feasibility の向上につながるが，同時に定義の曖昧さを意味する．(2) Model complexity は赤池の情報量基準 AIC におけるパラメータの次元に対応するものであり，負値になると解釈が難しい．$\log p(x;\,\theta)$ が θ について concave かつ $\hat{\theta}$ が事後平均のときは非負となるが保証されるが，一

般には負の値をとり得る．この 2 つの問題点が新しい概念（saddlepoint 予測分布）を導入することによって回避できることを第 2 節および第 3 節で示す．

本稿では次のような指数型分布族

$$p(x;\, \eta) = \exp\{\eta x - M(\eta)\}\, a(x) \tag{1.2}$$

を主たる題材とする．ただし，η は正準パラメータ，$M(\eta)$ はキュムラント関数，$a(x)$ は非負の関数である．

指数型分布族の著しい特徴として等式

$$\log \frac{p(x;\, \hat{\eta}_M)}{p(x;\, \eta)} - \mathrm{D}\bigl(p(z;\, \hat{\eta}_M),\, p(z;\, \eta)\bigr) = 0 \quad \text{for any } x \text{ and any } \eta \tag{1.3}$$

を挙げることができる (Kullback, 1959)．第 1 項は尤度原理により大きければ大きいほどよい量であり，第 2 項は損失であるから小さければ小さいほどよい量である．望大項と望小項のバランスを意味する大変興味深い等式であり，本稿のコアをなす．

2. Saddlepoint 予測分布

指数型分布族において成立する等式 (1.3) を少し緩めた条件を予測分布に課すことを考える．

Definition 1. 次の等式を満たす $p(y \mid x)$ を saddlepoint 予測分布という．

$$\mathrm{E}\left[\log \frac{p(x \mid x)}{p(x;\, \theta)} - \mathrm{D}\bigl(p(z \mid x),\, p(z;\, \theta)\bigr)\, \middle|\, p(x;\, \theta)\pi(\theta)\right] = 0. \tag{2.1}$$

この等式条件を saddlepoint condition と呼び，(1.3) を exact な saddlepoint condition と呼ぶことにする．

Saddlepoint condition (2.1) の意味するところを説明する．期待値記号内の第 1 項 $\log\{p(x \mid x)/p(x;\, \theta)\}$ は対数尤度比を一般化した量といえる．Plug-in 予測分布ならば対数尤度比になるからである．したがって，(2.1) は対数尤度比を一般化した量と損失 $\mathrm{D}\bigl(p(z \mid x),\, p(z;\, \theta)\bigr)$ が $p(x;\, \theta)\pi(\theta)$ の下で同じ期待値をもつことを意味する．Saddlepoint 予測分布 $p(y \mid x)$ に対する Bayes risk は $\log\{p(x \mid x)/p(x;\, \theta)\}$ の $p(x;\, \theta)\pi(\theta)$ の下での期待値と一致する訳である．

Saddlepoint 予測分布が満たす等式を導いておく．

Proposition 1. 任意の saddlepoint 予測分布 $p(y \mid x)$ に対して

$$\mathrm{E}\left[\log \frac{p(x \mid x)}{p(y \mid x)} - \mathrm{D}\bigl(p(z \mid x),\, p(z;\, \theta)\bigr) - \mathrm{D}\bigl(p(z;\, \theta),\, p(z \mid x)\bigr)\, \middle|\, p(y;\, \theta)p(x;\, \theta)\pi(\theta)\right] = 0 \tag{2.2}$$

が成り立つ．

Proof. 次の恒等式が基本的である．

$$\log \frac{p(x \mid x)}{p(y \mid x)} = \log \frac{p(x \mid x)}{p(x;\, \theta)} + \log \frac{p(x;\, \theta)}{p(y;\, \theta)} + \log \frac{p(y;\, \theta)}{p(y \mid x)}.$$

Definition 1 により，第 1 項と $\mathrm{D}\big(p(z\,|\,x),\,p(z;\,\theta)\big)$ の $p(x;\,\theta)\pi(\theta)$ の下での期待値は等しい．第 2 項の $p(y;\,\theta)p(x;\,\theta)$ の下での期待値は 0 である．第 3 項の $p(y;\,\theta)$ の下での期待値は定義により $\mathrm{D}\big(p(z;\,\theta),\,p(z\,|\,x)\big)$ となる． □

Saddlepoint 予測分布から簡単に Bayes 型モデル選択基準を導くことができる．赤池の情報量基準 AIC に近い量を導出した Yanagimoto (1994) の手法に従い，$\mathrm{E}\big[-2\log p(y\,|\,x)\ \big|\ p(y;\,\theta)\pi(\theta)\big]$ の不偏推定量として

$$-2\log p(x\,|\,x) + q_D + q'_D \tag{2.3}$$

が得られる．ただし，

$$q_D = 2\,\mathrm{E}\Big[\mathrm{D}\big(p(z\,|\,x),\,p(z;\,\theta)\big)\ \Big|\ \pi(\theta\,|\,x)\Big] \geq 0,$$
$$q'_D = 2\,\mathrm{E}\Big[\mathrm{D}\big(p(z;\,\theta),\,p(z\,|\,x)\big)\ \Big|\ \pi(\theta\,|\,x)\Big] \geq 0$$

である．(1.1) の DIC における model complexity p_D と違って，q_D および q'_D は常に非負である．

次に Definition 1 の saddlepoint 予測分布より強い概念を 2 つ定義する．

Definition 2. (1) 次の等式を満たす予測分布を Bayes の意味の saddlepoint 予測分布という．

$$\mathrm{E}\left[\log\frac{p(x\,|\,x)}{p(x;\,\theta)} - \mathrm{D}\big(p(z\,|\,x),\,p(z;\,\theta)\big)\ \middle|\ \pi(\theta\,|\,x)\right] = 0 \quad \text{for any } x.$$

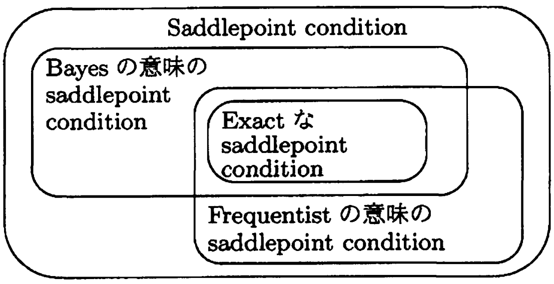
この等式条件を Bayes の意味の saddlepoint condition という．

(2) 次の等式を満たす予測分布を frequentist の意味の saddlepoint 予測分布という．

$$\mathrm{E}\left[\log\frac{p(x\,|\,x)}{p(x;\,\theta)} - \mathrm{D}\big(p(z\,|\,x),\,p(z;\,\theta)\big)\ \middle|\ p(x;\,\theta)\right] = 0 \quad \text{for any } \theta.$$

この等式条件を frequentist の意味の saddlepoint condition という．

4 つの saddlepoint condition の関係を模式図で表すと次のようになる．



Corcuera & Giummole (1999) は e-divergence 損失 $\mathrm{D}\big(p(y\,|\,x),\,p(y;\,\theta)\big)$ の下での Bayes 予測問題の最適解を求めた．解は e-mixture 予測分布と呼ばれるものであり，

$$p_e(y\,|\,x) = \frac{1}{c_e(x)}\exp\Big(\mathrm{E}\big[\log p(y;\,\theta)\ \big|\ \pi(\theta\,|\,x)\big]\Big)$$

のように定義される．ここで $c_e(x)$ は規格化因子である．

この e-mixture 予測分布は Bayes の意味の saddlepoint 予測分布になっている．

Example 1. Yanagimoto & Ohnishi (2009) は e-mixture 予測分布 $p_e(y\,|\,x)$ が

$$\mathrm{E}\left[\log\frac{p_e(x\,|\,x)}{p(x;\,\theta)} - \mathrm{D}\big(p_e(z\,|\,x),\,p(z;\,\theta)\big)\,\middle|\,\pi(\theta\,|\,x)\right] = 0 \quad \text{for any } x \tag{2.4}$$

を満たすことを示した．

Frequentist の意味の saddlepoint 予測分布の例を 2 つ挙げる．

Example 2. $n \geq 3$ とする．n 変量正規分布 $N_n(\mu,\,I)$ において，

$$\mathrm{E}\left[\log\frac{p(x-\hat{\mu}_S)}{p(x-\mu)} - \mathrm{D}\big(p(z-\hat{\mu}_S),\,p(z-\mu)\big)\,\middle|\,p(x-\mu)\right] = 0 \quad \text{for any } \mu \tag{2.5}$$

が成り立つ．ただし，$p(x-\mu)$ は $N_n(\mu,\,I)$ の確率密度，$\hat{\mu}_S = \big\{1-(n-2)\|x\|^{-2}\big\}x$ はいわゆる Stein 推定量 (Stein, 1981) である．つまり，Stein 推定量に基づく plug-in 予測分布 $p(y-\hat{\mu}_S)$ は frequentist の意味の saddlepoint 予測分になっている．

Example 3. $x = (x_1, \ldots, x_n)$ が独立に同一の正規分布 $N(\mu,\,\tau^{-1})$ に従っているとする．μ の MLE $\hat{\mu}_M$ および τ の条件付き MLE $\hat{\tau}_C$ はそれぞれ

$$\hat{\mu}_M = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$
$$\hat{\tau}_C = \frac{n-1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

となる．このとき，$\big(\hat{\mu}_M,\,\hat{\tau}_C\big)$ に基づく plug-in 予測分布 $p(y;\,\hat{\mu}_M,\,\hat{\tau}_C)$ は frequentist の意味の saddlepoint 予測分になる．すなわち，

$$\mathrm{E}\left[\log\frac{p(x;\,\hat{\mu}_M,\,\hat{\tau}_C)}{p(x;\,\mu,\,\tau)} - \mathrm{D}\big(p(z;\,\hat{\mu}_M,\,\hat{\tau}_C),\,p(z;\,\mu,\,\tau)\big)\,\middle|\,p(x;\,\mu,\,\tau)\right] = 0 \quad \text{for any } (\mu,\,\tau) \tag{2.6}$$

が成り立つ．ただし，

$$p(x;\,\mu,\,\tau) = \prod_{i=1}^{n}\sqrt{\frac{\tau}{2\pi}}\exp\left\{-\frac{\tau}{2}(x_i-\mu)^2\right\}$$

である．

Bayes または frequentist の意味の Saddlepoint 予測分布が満たす等式を proposition として述べておく．

Proposition 2. (1) Bayes の意味の saddlepoint 予測分布および frequentist の意味の saddlepoint 予測分布は Proposition 1 の等式 (2.2) を満たす．

(2) 特に frequentist の意味の saddlepoint 予測分布 $p(y\,|\,x)$ は

$$\mathrm{E}\left[\log\frac{p(x\,|\,x)}{p(y\,|\,x)} - \mathrm{D}\big(p(z\,|\,x),\, p(z;\,\theta)\big) - \mathrm{D}\big(p(z;\,\theta),\, p(z\,|\,x)\big)\;\middle|\; p(y;\,\theta)p(x;\,\theta)\right] = 0 \quad \text{for any } \theta$$

を満たす.

指数型分布族の場合には plug-in 予測分布が重要な役割を果たす. Yanagimoto & Ohnishi (2011) の結果を proposition の形で述べておく.

Proposition 3. 標本分布が (1.2) の指数型分布族 $p(x;\,\eta)$ の場合を考える.
(1) Plug-in 予測分布 $p(y;\,\hat{\eta})$ が exact な saddlepoint condition を満たすのは, $\hat{\eta} = \hat{\eta}_M$ のときのみである.
(2) Plug-in 予測分布 $p(y;\,\hat{\eta})$ が Bayes の意味の saddlepoint condition を満たすのは, $\hat{\eta} = \hat{\eta}_M$ or $\hat{\eta}_B$ のときのみである. ただし, $\hat{\eta}_B = \mathrm{E}\big[\eta \mid \pi(\theta\,|\,x)\big]$ である.

3. Pythagorean relationship

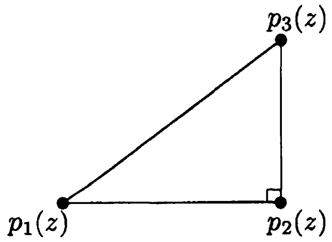
本節では Pythagorean relationship と saddlepoint condition が密接に関係していることを示す. まず, Pythagorean difference を

$$\mathrm{PD}\big(p_1(z),\, p_2(z),\, p_3(z)\big) = \mathrm{D}\big(p_1(z),\, p_3(z)\big) - \mathrm{D}\big(p_1(z),\, p_2(z)\big) - \mathrm{D}\big(p_2(z),\, p_3(z)\big)$$

のように定義する. 特にパラメトリックな 3 つの確率密度に対して,

$$\mathrm{PD}\big(\theta_1,\, \theta_2,\, \theta_3\big) = \mathrm{PD}\big(p(z;\,\theta_1),\, p(z;\,\theta_2),\, p(z;\,\theta_3)\big)$$

と定義する. $\mathrm{E}\big[\mathrm{PD}\big(p_1(z),\, p_2(z),\, p_3(z)\big)\big] = 0$ は平均的な「直角 3 角形」が成り立つことを意味する.



最適性を示す Pythagorean relationship を proposition の形で述べておく. これは Corcuera & Giummole (1999) によって証明された結果を Yanagimoto & Ohnishi (2009) が Pythagorean difference を使って表現したものである.

Proposition 4. 任意の予測分布 $p(y\,|\,x)$ に対して,

$$\mathrm{E}\Big[\mathrm{PD}\big(p(z\,|\,x),\, p_e(z\,|\,x),\, p(z;\,\theta)\big)\;\Big|\; \pi(\theta\,|\,x)\Big] = 0, \quad \text{for any } x \tag{3.1}$$

が成り立つ．

すべての予測分布の中で最適なので，$p_e(y\,|\,x)$ はもちろん saddlepoint 予測分布の中でも最適である．また，(2.3) で見たとおり saddlepoint 予測分布からシステマティックに Bayes 型モデル選択基準を導くことができる．したがって，$p_e(y\,|\,x)$ に基づくモデル選択基準を構築するのが自然といえる．実際，次のモデル選択基準が得られる．

$$\mathrm{DIC}' = -2\log p_e(x\,|\,x) + q_D + q'_D.$$

指数型分布族において DIC' と DIC は非常に近くなることを示そう．Yanagimoto & Ohnishi (2009) は (1.2) の指数型分布族 $p(x;\,\eta)$ において，

$$p_e(y\,|\,x) = p\big(y;\,\hat{\eta}_B\big)$$

が成り立つことを示した．ただし，$\hat{\eta}_B$ は正準パラメータ η の事後平均である．つまり，正準パラメータの事後平均に基づく plug-in 予測分布が e-mixture 予測分布と一致する．Example 1 と Definition 2 により，推定量として事後平均を採用するとき

$$p_D = q_D$$

である．DIC と DIC' の違いは $p_D = q_D$ と q'_D の違いのみとなる．

次に指数型分布族において MLE-induced 予測分布が plug-in saddlepoint 予測分布の中で最悪であることを示す．実際，次のような最悪性を示す Pythagorean relationship が成り立つ．

Proposition 5. 標本分布は (1.2) の指数型分布族 $p(x;\,\eta)$ とする．任意の plug-in saddlepoint 予測分布 $p(y;\,\hat{\eta})$ に対して，

$$\mathrm{E}\Big[\mathrm{PD}\big(\hat{\eta}_M,\,\hat{\eta},\,\eta\big)\,\Big|\,p(x;\,\eta)\pi(\eta)\Big] = 0$$

が成り立つ．

Proof. Exact な saddlepoint condition (1.3) において $\eta = \hat{\eta}$ を代入する．

$$\log\frac{p(x;\,\hat{\eta}_M)}{p(x;\,\hat{\eta})} = \mathrm{D}\big(p(z;\,\hat{\eta}_M),\,p(z;\,\hat{\eta})\big).$$

(1.3) とこの式を辺ごとに引き算すると

$$\log\frac{p(x;\,\hat{\eta})}{p(x;\,\eta)} = \mathrm{D}\big(p(z;\,\hat{\eta}_M),\,p(z;\,\eta)\big) - \mathrm{D}\big(p(z;\,\hat{\eta}_M),\,p(z;\,\hat{\eta})\big)$$

が得られる．Saddlepoint condition

$$\mathrm{E}\Big[\log\frac{p(x;\,\hat{\eta})}{p(x;\,\eta)} - \mathrm{D}\big(p(z;\,\hat{\eta}),\,p(z;\,\eta)\big)\,\Big|\,p(x;\,\eta)\pi(\eta)\Big] = 0$$

と組み合わせると求める等式を得る． □

(1.3) で見たように，指数型分布族において MLE-induced 予測分布は exact な saddlepoint condition を満たす．この事実は MLE-induced 予測分布の最悪性につながっている訳であるが，次のように解釈してみよう．

最低限の推定が保証されているという意味で指数型分布族は優れた統計モデルである．

Bayes 型モデル選択基準を簡単に導けるという意味で saddlepoint condition は予測分布のよい性質と考えられる．Saddlepoint 予測分布の中で最適なのが e-mixture 予測分布であり，plug-in saddlepoint 予測分布の中で最悪なのが MLE-induced 予測分布である．e-mixture 予測分布の最適性が事前分布に依存するのに対し，MLE-induced 予測分布の最悪性は事前分布に依らない．比喩表現が許されるとすれば，MLE-induced 予測分布は「優秀クラスのビリ」といえる．

本節の最後に Pythagorean relationship から saddlepoint condition を導くことにする．Proposition 4 の最適性を示す Pythagorean relationship (3.1) において置き換え

$$p(z \mid x) \longrightarrow \delta(z-x)$$

を行うと，Eaxmple 1 における (2.4)，すなわち，Bayes の意味の saddlepoint condition が得られる．ここで，$\delta(z-x)$ は x に mass をもつデルタ関数である．

Stein 推定量に関する Example 2 および条件付き MLE に関する Example 3 における frequentist の意味の saddlepoint condition (2.5) および (2.6) は，それぞれ次の Pythagorean relationship と等価である（証明は次の段落）．

$$\mathrm{E}\Big[\mathrm{PD}(\hat{\mu}_M,\, \hat{\mu}_S,\, \mu)\;\Big|\; p(x-\mu)\Big] = 0 \quad \text{for any } \mu,$$

$$\mathrm{E}\Big[\mathrm{PD}\big((\hat{\mu}_M,\, \hat{\tau}_M),\, (\hat{\mu}_M,\, \hat{\tau}_C),\, (\mu,\, \tau)\big)\;\Big|\; p(x;\, \mu,\, \tau)\Big] = 0 \quad \text{for any } (\mu,\, \tau).$$

Saddlepoint condition に関する定義との整合性を考えると，これらは frequentist の意味の Pythagorean relationship と呼ぶべきものである．

上記の等価性は Proposition 5 と同様に証明できる．指数型分布族 (1.2) において一般的な証明を与える．Exact な saddlepoint condition (1.3) を 2 回利用して，任意の plug-in 予測分布 $p(x;\, \check{\eta})$ に対して，

$$\log \frac{p(x;\, \check{\eta})}{p(x;\, \eta)} = \mathrm{D}\big(p(z;\, \hat{\eta}_M),\, p(z;\, \eta)\big) - \mathrm{D}\big(p(z;\, \hat{\eta}_M),\, p(z;\, \check{\eta})\big)$$

を得る．これは frequentist の意味の saddlepoint condition と（frequentist の意味の）Pythagorean relationship が等価であることを示している．

4. Saddlepoing condition の一般化

本節では第 2 節および第 3 節の内容を 2 つのケースに拡張することを考える．具体的には，(1) 損失が α-divergence の場合の saddlepoint condition，および，(2) 推定関数に対する saddlepoint condition を議論する．

まず最初に α-divergence を定義する．

$$\mathrm{D}_\alpha\big(p(y),\, q(y)\big) = \mathrm{E}\left[f_\alpha\left(\frac{q(y)}{p(y)}\right)\;\middle|\; p(y)\right],$$

$$f_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2}\left(1 - z^{\frac{1+\alpha}{2}}\right) & (|\alpha| < 1), \\ z \log z & (\alpha = 1), \\ -\log z & (\alpha = -1). \end{cases}$$

$\alpha=1$ は e-divergence に，$\alpha=-1$ は m-divergence に対応する．

α-divergence を損失とする Bayes 予測問題を考える．すなわち，次の Bayes risk の最小問題

$$\min \mathrm{E}\Big[\mathrm{D}_\alpha\big(p(y;\,\theta),\,p(y\,|\,x)\big)\ \Big|\ p(x;\,\theta)\pi(\theta)\Big]$$

を考える．

Corcuera & Giummole (1999) は次の式で定義される α-mixture 予測分布

$$p_\alpha(\boldsymbol{y}\,|\,\boldsymbol{x})=\frac{1}{c_\alpha(x)}\left(\mathrm{E}\left[\{p(\boldsymbol{y};\,\theta)\}^{\frac{1-\alpha}{2}}\,\Big|\,\pi(\theta\,|\,x)\right]\right)^{\frac{2}{1-\alpha}}$$

が最適解であることを示した．ただし，$c_\alpha(x)$ は規格化因子である．

実際，Yanagimoto & Ohnishi (2009) が指摘しているとおり，最適性を示す modified Pythagorean relationship が成り立つ．すなわち，任意の $p(y\,|\,x)$ に対して，

$$\begin{aligned}\mathrm{E}\Big[&\mathrm{D}_\alpha\big(p(z;\,\theta),\,p(z\,|\,x)\big)-\mathrm{D}_\alpha\big(p(z;\,\theta),\,p_\alpha(z\,|\,x)\big)\\&-\{c_\alpha(x)\}^{\frac{1-\alpha}{2}}\mathrm{D}_\alpha\big(p_\alpha(z\,|\,x),\,p(z\,|\,x)\big)\ \Big|\ \pi(\theta\,|\,x)\Big]=0,\quad\forall x\end{aligned}$$

が成り立つ．この modified Pythagorean relationship において形式的な置き換え

$$\{p(z\,|\,x)\}^{\frac{1+\alpha}{2}}\longrightarrow\frac{\delta(z-x)}{\{p_\alpha(\boldsymbol{x}\,|\,\boldsymbol{x})\}^{\frac{1-\alpha}{2}}}$$

を行うと，次の example を得る．

Example 4. $-1<\alpha\le 1$ とする．Yanagimoto & Ohnishi (2009) は $p_\alpha(y\,|\,x)$ が Bayes の意味の saddlepoint condition

$$\mathrm{E}\left[f_{-\alpha}\left(\frac{p(x;\,\theta)}{p_\alpha(x\,|\,x)}\right)-\mathrm{D}_\alpha\big(p(z;\,\theta),\,p_\alpha(z\,|\,x)\big)\ \Big|\ \pi(\theta\,|\,x)\right]=0\quad\text{for any } x$$

を満たすことを示した．$\alpha=1$ は Example 1 に対応する．興味深いことに，$\alpha=-1$ のケース，すなわち，通常使われる m-mixture 予測分布は Bayes の意味の saddlepoint condition を満たさない．

次に location family における推定問題を考える．

推定問題　$\mathbb{R}^n$ 上の location family $p(x-\theta)$ において，適当な関数 $u(x)$ を選び，推定関数

$$\nabla\log\frac{p(x-\theta)}{u(x)}$$

によって θ を推定せよ．ただし，$n\ge 3$ であり，∇ は x に関する微分を意味する．

天下り的ではあるが，対数尤度比に対応する量を

$$L=\frac{1}{2}\|\nabla\log p(x-\theta)\|^2-\frac{1}{2}\|\nabla\log u(x)\|^2=\frac{1}{2}\|\nabla\log p(x-\theta)\|^2-\frac{1}{2}\|\nabla\log p(x-\hat{\theta})\|^2$$

で定義し，$L(\theta,\hat{\theta}\,|\,x)$ と書く．ここで，$\hat{\theta}$ は推定関数 $\nabla\log\{p(x-\theta)/u(x)\}$ から得られる推定量とする．定義により，

$$\nabla\log u(x)=\nabla\log p(x-\hat{\theta})$$

であることに注意されたい．左辺の $\nabla\log p(x-\hat{\theta})$ は $\nabla\log p(x-\theta)$ において $\theta=\hat{\theta}$ を代入したものである．L が最大化されるのは $u(x)$ が定数関数のときであり，

$$\nabla\log p(x-\theta)=-\frac{\partial}{\partial\theta}\log p(x-\theta)$$

であるから，このとき推定関数は尤度推定関数と符号を除いて一致する．また，正規分布のときには対数尤度比と一致する．したがって，対数尤度比の拡張概念と考えることができる．

一方，損失に対応する量を

$$\mathrm{D}=\frac{1}{2}\left\|\nabla\log\frac{p(x-\theta)}{u(x)}\right\|^2=\frac{1}{2}\left\|\nabla\log\frac{p(x-\theta)}{p(x-\hat{\theta})}\right\|^2$$

で定義し，$\mathrm{D}(\theta,\hat{\theta}\,|\,x)$ と書く．これは推定関数の評価基準のうち trace 基準 (Godambe & Kale, 1991) と密接に関連しており，Ohnishi & Yanagimoto (2003) によって推定関数に対する損失として採用されている．

Ohnishi & Yanagimoto (2003) は推定関数に関する Pythagorean relationship を導いた．具体的には，$u(x)=\|x\|^{2-n}$ としたときの推定関数を Stein 型推定関数と呼び，Stein 型推定関数がどのように尤度推定関数を優越するかを示した．この等式は今の記法で書けば，

$$\mathrm{E}\Big[\mathrm{D}(\theta,\hat{\theta}_M\,|\,x)-\mathrm{D}(\theta,\hat{\theta}_S\,|\,x)-\mathrm{D}(\hat{\theta}_S,\hat{\theta}_M\,|\,x)\ \Big|\ p(x-\theta)\Big]=0\quad\text{for any }\theta \tag{4.1}$$

となる．ただし，$\hat{\theta}_S$ は Stein 型推定関数に対する推定量である．

Exact な saddlepoint condition と frequentist の意味の saddlepoint condition を導く．

Example 5. (1) 尤度推定関数は

$$L(\theta,\hat{\theta}_M\,|\,\boldsymbol{x})-\mathrm{D}(\theta,\hat{\theta}_M\,|\,\boldsymbol{x})=0\quad\text{for any }x\text{ and any }\theta$$

を満たす．これは $L(\theta,\hat{\theta}\,|\,x)$ および $\mathrm{D}(\theta,\hat{\theta}\,|\,x)$ の定義から明らかである．
(2) Stein 型推定関数は

$$\mathrm{E}\Big[L(\theta,\hat{\theta}_S\,|\,x)-\mathrm{D}(\theta,\hat{\theta}_S\,|\,x)\ \Big|\ p(x-\theta)\Big]=0\quad\text{for any }\theta$$

を満たす．次の等式

$$L(\theta,\hat{\theta}_S\,|\,x)=\mathrm{D}(\theta,\hat{\theta}_M\,|\,x)-\mathrm{D}(\hat{\theta}_S,\hat{\theta}_M\,|\,x)$$

に注意すれば，Pythagorean relationship (4.1) と等価であることが分かる．

Location family は指数型分布族と同様に最低限の推定が保証されているという意味で優れたモデルであるといえる．

References

Corcuera, J. M. and Giummole, F. (1999). A generalized Bayes rule for prediction. *Scand. J. Statist.* **26**, 265–279.

Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York.

Godambe, V.P. and Kale, B.K. (1991). Estimating functions: an overview, *Estimating Functions* (V. P. Godambe, ed.), 3–20, Clarendon Press, Oxford.

Ohnishi, T. and Yanagimoto, T. (2003). Electrostatic views of Stein-type estimation of location vectors. *J. Japan Statist. Soc.* **33**, 39–64.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussions). *J. Roy. Statist. Soc.*, **64**, 583–639.

Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135–1151.

Yanagimoto, T. (1994). The Kullback-Leibler risk of the Stein estimator and the conditional MLE. *Ann. Inst. Statist. Math.*, **46**, 29-41.

Yanagimoto, T. and Ohnishi, T. (2009). Bayesian prediction of a density function in terms of e-mixture. *J. Statist. Plann. Inf.*, **139**, 3064–3075.

Yanagimoto, T. and Ohnishi, T. (2011). Saddlepoint condition on a predictor to reconfirm the need for the assumption of a prior distribution. *J. Statist. Plann. Inf.*, **141**, 1990–2000.