数理解析研究所講究録
第2047巻 2017年 89-100

尤度法はベイズ法無き里の蝙蝠か

統計数理研究所　　柳本 武美
Takemi Yanagimoto
Institute of Statistical Mathematics

§1. 序：尤度法

尤度法を学んだ後でベイズ法を学ぶと、尤度法はベイズ法の特殊な場合であるとの印象をもつ。最尤推定量が一様事前密度を仮定した下での事後尤度で説明できることは直ぐに理解されることであり、早くから指摘されてきた。しかし、その直感についての議論は多くない。

大西との一連の共同研究 (Yanagimot and Ohnishi, 2009, 2011, 2014, Manuscript) により、Kullback-Leibler separator を損失関数とした予測子の研究とか、Jeffreys 事前分布を仮定した下での事後平均による母数の研究を行っている。これらの共同研究を通じてベイズ法の普遍性と尤度法の限界を改めて考える機会を得た。この考察を纏めることが、が将来の研究の進展に役立つと考えて発表する。

議論を整理するために、先ず尤度法の概要を復習する。話題を絞り焦点をあてるために、観測値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ は指数分布族

$$\mathcal{M} = \{p(\boldsymbol{x}|\theta) = \exp\{n(\bar{t}\theta - M(\theta))\}\, a(\boldsymbol{x}) | \, \theta \in \Theta\} \qquad (1.1)$$

に含まれる確率密度に従う確率変数からの大きさ n の実現値ベクトルとする。密度関数を母数の関数と見なして尤度

$$L(\theta; \boldsymbol{x}) = p(\boldsymbol{x}|\theta)$$

が定義される。尤度法は、尤度だけに依存して推測を行う体系だと定義できる。母数の最尤推定量 $\hat{\theta}_M$ と二つのモデル $\mathcal{M}_i = \{p_i(\boldsymbol{x}|\theta_i)\}$ に対する尤度比 $p_1(\boldsymbol{x}|\hat{\theta}_{1M})/p_2(\boldsymbol{x}|\hat{\theta}_{2M})$ が代表的な推測法である。また、条件付き尤度とか周辺尤度など、最尤推定量の改良がなされている。更には、複数のモデルを比較する AIC も提案されている。

歴史的に見て、尤度法は大きな役割を果たした。モーメント法とか最小二乗法による母数の推定では必ずしもモデルの仮定は明確ではない。実際指数分布族に属する密度 $p(\boldsymbol{x}|\theta)$ が (1.1) の $\mathcal{M}$ に属すると仮定できるときには、最尤推定量は十分統計量の関数 $t(\boldsymbol{x})$ の関数になる。一方で、ガンマ分布の母数推定ではモーメント推定量が提案されることがあった (例えば [JK])。ところがガンマ分布の密度関数

$$p(\boldsymbol{x}|\mu,\tau) = \prod \frac{\tau^\tau}{\Gamma(\tau)} \frac{x_i^{\tau-1}}{\mu^\tau} \exp\left(-\frac{\tau x_i}{\mu}\right)$$

と書けるから、十分統計量は標本平均 $\bar{x}$ と標本幾何平均 $\tilde{x}$ だから、通常のモーメント推定量のように、2次モーメント $\sum x_i^2$ を用いる理由は無い。

一方で、最尤推定量の最適性は驚くほど貧弱である。よく知られた最適性は

1) 弱い正則性条件の下で、漸近的に正規分布に従い、漸近分散は最小である。

2) 指数分布族の平均母数の推定では、UMVUE になる。

二つの最適性において、1) では漸近性、2) では不偏性条件が付けられた下での最適性である。特に後者の不偏性の条件は実際的でない。

§2. 一様事前分布の下での事後モード

Bayes 法では、未知母数 θ が超母集団からの実現値とみなす。その密度を

$$\pi(\theta) \text{ (on } \Theta)$$

と書いて事前分布と呼ぶ。

観測値を固定して考える意味で、事後密度の考え方と相通じる面がある。実際、Θ 上で一様な事前密度

$$\pi(\theta) \equiv 1 \tag{2.1}$$

を仮定すると、尤度と事後密度とは比例する。但し、一様事前密度は、例えば $\Theta = R^p$ の場合、積分して有限でないことが多い。積分が有限でない場合、不適切 (improper) な事前密度と呼ばれる。事前密度が不適切であっても、事後密度が滑らかに定義できれば問題はない。

一様密度の困難は、θ に対する一様密度と $\eta = g(\theta)$ に対する一様密度が同等でない事実である。事前密度 (2.1) と $\pi(\eta) \equiv 1$ は同等ではない。変数変換に伴う Jacobian が無視されているからである。一方最尤推定量では元々事前密度を仮定しないから、そうした困難は生じない。

また、一様事前密度は常識的に見て不自然なことも多い。実際２項分布 $\mathrm{Bi}(n,p)$ では $\theta = \log\{p/(1-p)\}$ に対して $(-\infty,\infty)$ 上の一様密度が仮定されることがあるが、滑らかな事後密度が導かれない。そして、その一様密度から導かれる事後平均が最尤推定量と対応している。

確かに、一様事前密度を仮定することにより、最尤推定量を説明できる。しかしその説明は簡単ではあるが無理があり、より一層の深い議論は行えない。

§3. Jeffreys 事前分布の下での事後平均

Jeffreys 事前分布は標本密度が指数分布族 (1.1) に従うときには

$$\pi_J(\theta) \propto \sqrt{M''(\theta)}. \tag{3.1}$$

で定義される。より一般的には行列式を用いて、$M''(\theta)$ を

$$|\mathrm{E}\{\partial\theta^2 \log L(\theta;\boldsymbol{x})/\partial^2\theta;\, p(\boldsymbol{x}|\theta)\}|$$

と置き換える。この事前密度は変数変換を行っても同等な事前密度が得られる。むしろ変数変換に関して不変な事前密度として導入されてた (Jeffreys, 1961)。

共役事前密度は適当に m, δ を選んで、次のように表される。

$$\pi(\theta; m, \delta) = \exp\{\delta\big(m\theta - M(\theta) - N(m)\big)\}\, b(\theta)k(m, \delta) \qquad (3.2)$$

但し、$N(t)$ は凸関数 $M(\theta)$ に共役な凸関数である。二つの確率密度 $p_1(\boldsymbol{y})$ and $p_2(\boldsymbol{y})$ の Kullback-Leibler divergence, $\mathrm{E}\{\log\big(p_1(\boldsymbol{y})/p_2(\boldsymbol{y})\big); p_1(\boldsymbol{y})\}$, を $\mathrm{D}(p_1(\boldsymbol{y}), p_2(\boldsymbol{y}))$ と書く。この記法を使うと (3.2) は $c = N'(m)$ とおいて、次のように書き直すことができる。

$$\pi(\theta; m, \delta) = \exp\{-\delta \mathrm{D}(p(\boldsymbol{y}|c), p(\boldsymbol{y}|\theta))\}\, b(\theta)k(m, \delta). \qquad (3.3)$$

この事前密度を特定するためには、δ と m を選ぶと共に $b(\theta)$ をどのように選ぶ必要がある。この問題は、m の選択が難しいときには $\delta = 0$ と置くことにより困難が解消できるので特に重要になる。この場合 $b(\theta)$ は無情報事前密度と呼ばれる。Jeffreys 事前密度は一つの提案である。

事後平均で母数を推定することにより、一つの最適性が導かれる。勝手な $\check{\theta}$ に対して次の不等式が成り立つ。

$$\mathrm{E}\{(\hat{\theta} - \theta)^2; \pi(\theta|\boldsymbol{x})\} \leq \mathrm{E}\{(\check{\theta} - \theta)^2; \pi(\theta|\boldsymbol{x})\}.$$

Jeffreys 事前密度の下での事後平均により母数を推定すると、別の魅力的な最適性が導かれる。推定値を plug-in した予測子は e-最適な予測子である。

Yanagimoto and Ohnishi (Manuscript) は次の関係式を得て、同時に Jeffreys 事前密度を特徴付けた。

命題. $M(\theta)$ と Fisher 情報行列 $I(\theta)$ についての適当な正則条件の下に、事前確率密度

$$\pi(\theta; \theta_0, n) \propto \exp\{-n\mathrm{D}(\theta_0, \theta)\} b(\theta).$$

を定義すると仮定する。このとき漸近的な関係式

$$\mathrm{E}\big\{\theta\,;\, \pi(\theta; \theta_0, n)\big\} = \theta_0 + O(n^{-2})$$

をすべての $\theta_0 \in \Theta$ で満たすための必要十分条件は、$b(\theta)$ が偏微分方程式

$$\nabla \left\{ \log \frac{b(\theta)}{\sqrt{\det I(\theta)}} \right\} = 0 \tag{3.4}$$

を満たすことである。Jeffreys 事前密度はこの偏微分方程式を満たす唯一の解である。

この命題は最尤推定量とベイズ推定量の間の深い関係を示している。有効性など最尤推定量の漸近的性質は、そのまま Jeffreys 事前密度の下での事後平均 $\hat{\theta}$ が満たす性質になる。また、尤度比検定統計量も同様である。最尤推定量が満たして、$\hat{\theta}$ が満たさない性質は、指数分布族の平均母数の UMVUE であることと、修正尤度法だけである。この点も次節でその限界を指摘する。

予測子の視点から見る。一つの最適な予測子 $p_e(\boldsymbol{y} \,|\, \boldsymbol{x})$ は、勝手な予測子 $p(\boldsymbol{y} \,|\, \boldsymbol{x})$ に対して

$$E\big\{\mathrm{PD}\big(p(\boldsymbol{y} \,|\, \boldsymbol{x}),\ p_e(\boldsymbol{y} \,|\, \boldsymbol{x}),\ p(\boldsymbol{y};\, \theta)\big) \,\big|\, \pi(\theta \,|\, x)\big\} = 0$$

を満たす。その予測子は、plug-in 予測子 $p(\boldsymbol{y}|\hat{\theta})$ と一致する。しかしが、最尤推定量の plug-in 予測子にはこれと言った最適性はない。

更に詳しく調べるために、次の等式を満たす予測子の族を考える。

$$\mathrm{E}\{\log\{p(\boldsymbol{x}|\boldsymbol{x})/p(\boldsymbol{x}|\theta)\} \ - \ \mathrm{D}(p(\boldsymbol{y}|\boldsymbol{x}), p(\boldsymbol{y}|\theta)); \pi(\theta|\boldsymbol{x})\} \ = \ , 0. \tag{3.5}$$

最尤推定量を plug-in した予測子 $p(\boldsymbol{y}|\hat{\theta}_M)$ は上の等式を事後平均を取らないでも成立する。その等式は、最尤推定量と Kullback-Leibler separator の双方にとって好ましい性質とされている。しかし、同時に $p(\boldsymbol{y}|\hat{\theta}_M)$ は上の族の中でベイズリスクを最大化する (Yanagimoto and Ohnishi, 2011). ベイズリスクを最小にするのはベイズ予測子は $p(\boldsymbol{y}|\hat{\theta})$ で表される。言い換えると、漸近的には二つの推定量は、$O\big(n^{-2}\big)$ まで同等であるが、plug-in 予測子を評価すると Jeffreys 事前密度の下では Kullback-Leibler 損失のベイズリスクの最大化させる予測子と最小化させる予測子に分かれる。

近年統計的推測における予測子の役割が広く認められている中ではこの違いは大きい。例えば AIC は観察 plug-in 予測子 $p(\boldsymbol{x}|\hat{\theta}_M)$ に基づいているが、$p(\boldsymbol{x}|\hat{\theta})$ に基づいて DIC を補正した方が望ましいことが予想される。

上の定理がもたらす興味深くある意味で皮肉な側面は、最尤推定量も Jeffreys 事前密度 (従ってそれから導かれる推定量) も批判の対象である事実である。しかもその批判は故のない批判ではなくて、ごく真っ当で十分納得できる。実際、最尤推定量は Neyman-Scott 推定量とか Stein 推定量が示すように、母数の次元が高いときを始め多くの欠点があることが受け入れられている。修正尤度法は欠点を避けたり小さくするための工夫と理解できる。Jeffreys 事前密度も同様に批判の対象である。Reference 事前密度の方が受け入れられているように思われる。左右 Haar 測度に関しては Robert (2001, Chapter 9.3) が参照できる。

問題は欠点があるかどうかではない。欠点が小さいときにだけ適用して、欠点が大きいときにはその欠点が小さくなるように対策を施せば良いからである。だから、欠点が見込まれるときの対策を議論する。この議論に入ると、尤度推論の限界が明瞭になる。

§4. ベイズ法の多様性と coherency

最尤推定量 $\hat{\theta}_M$ と Jeffreys 事前密度の下での事後平均 $\hat{\theta}$ に欠点があるときにどのように対処できるだろうか。最も簡単な例として正規分布 $\mathrm{N}(\mu, \sigma^2)$ の場合を考える。この場合は両者は同じ推定量 $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \sum(x_i - \bar{x})^2/n$ となる。しかし、この推定量は欠点がある。その欠点のために Neyman-Scott 問題が生じる。この欠点を排除するために提案されるのが修正尤度法である。正規密度を

$$
\begin{aligned}
p_N(\boldsymbol{x}|\mu, \sigma^2) &= \prod \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2\sigma^2}(x_i - \mu)^2 \\
&= \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp -\frac{n}{2\sigma^2}(\bar{x} - \mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^{2(n-1)}}\sqrt{n}} \exp -\frac{1}{2\sigma^2}\sum(x_i - \bar{x})^2
\end{aligned}
$$

と分解して、分散の推定を右辺の第２項を最大化させる。第２項は周辺尤度とも見られるし、$\bar{x}$ を与えた時の条件付き尤度とも見られる。つまり、通常の尤度ではなくて修正尤度である。。結果として得られる推定量は $\hat{\sigma}_C^2 = \sum x_i - \bar{x})^2/(n-1)$ である。

ベイズ法での対応の一つは、$\tau = 1/\sigma^2$ とおくと (μ, τ) が自然母数であることに注意して、いて reference 事前密度として $\pi_R(\mu, \tau) \propto 1/\tau$ を仮定する。τ の事後平均を求めると $\hat{\tau} = 1/\hat{\sigma}_C^2$ となる。自然母数の事後平均であったから、$\bar{x}, \hat{\tau}$ を plug-in して得られる予測子は最適になる。多くの研究者は reference 事前密度の方が尤もらしいと考えているから、事前密度を変えるだけである。

事前密度を変えることは極めて広範囲の中から選ぶことである。実際には、事前密度を選択することは必ずしも容易ではない。しかし、選択の範囲は広い。Neyman-Scott 問題は reference 事前密度を仮定して、Stein 問題もクーロン・ポテンシャルに基づいた事前密度を仮定した下での事後平均を用いて解消される。事前密度の選択は多様である。無情報事前密度に限らないで、(3.2) のような共役事前密度を利用することも可能である。更には、より一般的な事前密度が状況に応じて選ぶことができる。

事前密度の選択が自由であること特徴は、母数の次元が高いときにより明瞭になる。一つの実用的な例として、平滑化問題を考える。正規モデル $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, (1/\tau)\mathrm{I})$ を仮定して、μ_i が i に伴い変化すると仮定する。問題はその変化の傾向をデータ $\boldsymbol{x}$ から推測することである。尤度法では回帰モデルとして定式化する。適当な関数系 $\boldsymbol{\mu}(\boldsymbol{\theta})$ を仮定して, 回帰モデルを

$$\boldsymbol{x} = \boldsymbol{\mu}(\boldsymbol{\theta}) + \boldsymbol{e}, \quad \boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0}, (1/\tau)\mathrm{I})$$

とする。多項式回帰モデルでは $\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \sum_{j=0}^{p} \boldsymbol{\theta}_{j+1} i^j$ とする。その上で母数は最尤推定量で推定される。母数の次元が高いと最尤推定量の性能は全く保証されない。それより重要な点は、関数系 $\boldsymbol{\mu}(\boldsymbol{\theta})$ を見つけることは難しい事実である。

平滑化法を階差2の行列 D_2 を $\boldsymbol{\mu}' D_2 \boldsymbol{\mu} = \sum(\mu_{i+2} - 2\mu_{i+1} + \mu_i)^2$ を使った場合で考える。事前密度を D_2^- を D_2 の (Moore-Penrose の一般化) 逆行列、$\boldsymbol{e}_1' = (1, \ldots\ldots, 1)$, $\boldsymbol{e}_2' = (\ldots, i - n/2, \ldots)$ として

$$\boldsymbol{\mu}(\boldsymbol{\theta}) \sim \mathrm{N}(\theta_1 \boldsymbol{e}_1 + \theta_2 \boldsymbol{e}_2, (1/\theta_3) D_2^-) \tag{4.1}$$

で与えられる。これから母数が、θ_3 については細かい議論が必要になるが、推定できる。必要であれば、階差の次数を変えても良い。指摘しておきたい点は、階差行列は多項式回帰モデルのように格別の工夫もなく仮定できることである。また、隣り合った平均値が近いことは平滑化ではごく自然な仮定であり、Wiener 過程とかスプライン関数などと関連がある理論的な裏付けのある仮定である。一方で、多項式回帰モデルは大抵のデータには適合が悪いことが知られている。更には、良い適合が見込まれる関数系 $\boldsymbol{\mu}(\boldsymbol{\theta})$ を見つけることには匠の技が求められる。しかも、今日解析が求められるデータは大規模で複雑化しているから、従来の匠の技では対応不可能である。

尤度法の本質的な限界は、モデルが制約される点にある。また、モデルの適合が疑われるときの対応が限られている。

§5. 尤度法に長所はあるのか

改めて尤度法見直すと

1) 事前密度仮定しなくても良い

2) 指数分布の平均母数の推定量が UMVUE である。

の2点が残された長所である。第1点は推測における根源に関わる問題である。不要な仮定は排除して可能な限り観察から推測することは経験科学根源である。だから、事前密度のような排除が可能でデータに基づかない仮定を設けることに抵抗があるのも納得できる。一方で、事前密度を仮定しなければ客観性が担保されるかを議論すると、話が変わる。要するに、本質的な違いは無い。回帰モデルのような統計モデルは客観的も受けられる訳ではない。また、データから自然に導かれる訳でもない。

解析者の主観的判断で仮定したモデルである。この事実は広く受け入れられている。

科学哲学の世界では広く受け入れられているが、しばしば看過される事実に観測値データ $\boldsymbol{x}$ が関係者の主観・経験・関心に影響された結果である事実である。いわゆる観察の「理論依存 (負荷) 性」である（例えば内井 (1995)）。統計解析が受け入れられている典型的な分野である、比較対照臨床試験を例に考える。試験は、プロトコールに定められた手順により実施される。そのプロトコールは専門家グループにより作成されるが、専門家とは将に既存の知識に囚われた人でであり、その知識も不十分である。実施に携わるスタッフも同様である。データは比較が公平になされるように努力がなされる一方で、関係者の主観・経験・関心が深く入り組んだ結果でもある。臨床試験に疎い人にとっては、臨床試験におけるデータ・マネージャーの大きさには気付きにくいと思われる。プロトコールに忠実に試験が実施されるように、データのを正確に扱う職務を担当する。真面目な人であれば直ぐにでも携わることができるとも考えられるが、実際はそうではない。専門的な研修を経たデータマネージャーが参画した試験の方が、信頼性の高いデータが得られる。「先入観を排除して虚心坦懐に観察する」としたイメージとは真逆である。

確かにデータに含まれる情報を事前密度の仮定によりねじ曲げられる怖れは否定できない。しかし、仮定しないとより客観的な推論になるかというと何の保証もない。臨床試験データの収集と解析の分野でベイズ法の導入が試みられている現状は、おかしな試みと言うより目的に即してデータに情報を作り込んでそこから情報を縛り出すための有効な方法の開発と捉えることができる。従来、事前密度はデータが収集された後の事後解析の一部であった。今日では科学研究は計画された下にデータが収集され、その計画の一部として統計解析手法が予め設定される。臨床試験の場合には登録制度が普及しているので、事前密度の事前登録の推奨が提案されつつある (例えば、Ogura and Yanagimoto, 2016)。

この節に挙げた 2) の長所は、その簡明さと直感的な訴求性にも拘わらず、その内容は貧弱である。先ず、平均母数の不偏な推定量に限る根拠が明瞭でない。確かに偏りがある推定量は望ましくないが、推定量の評価はリスクで評価することが原則である。そのためには妥当な損失を選択する必要がある。平均母数についても、位置母数でなければ、2乗損失が広く受け入れられるとは限らない。近年では、Kullback-Leibler 損失のような予測子に関連した損失がより受け入れられている。

上の議論はおいて平均母数の2乗損失を仮定した議論を、正規モデル $N(\mu, \sigma^2)$ を例に考える。平均母数は $(\mu, \mu^2+\sigma^2)$ となる。その UMVUE は $(\bar{x}, \sum x_i^2)$ であり、最尤推定量でもある。改めて見直すと、μ についてはともかく $\mu^2+\sigma^2$ の不偏な推定量に議論を限ることの妥当性は説得的ではない。実際この UMVUE は前節でも指摘したように、批判の対象である。その結果、修正尤度法とか reference 事前密度を仮定した下での事後平均が勧められる。両者は一致して σ^2 の UMVUE でもある。

以上の考察を要するに、「尤度法はベイズ法無き里の蝙蝠」であるとの結論になる。この言明はやや性急すぎるかも知れない。しかし一歩譲っても、殆ど成り立つ言明と思われる上に、更に議論を進めるに値する言明である。

補注：

本文でも述べたように、本稿の骨子は専ら大西教授 (九州大学) との共同研究による成果である。しかし、本稿で展開した推測に関する論考は専ら筆者によることから単著とした。

文献

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, 62, 547-554.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd Inter. Symp. on Information Theory.* (eds. Petroc, B. N. and Csak, F.), 267-281, Akademia Kiado, Budapest.

Altham, P.M.E. (1969). Exact Bayesian analysis of a 2×2 contingency table, and Fisher's "exact" significance test. *Journal of the Royal Statistical Society: Series B*, 31, 261-269.

Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference. Journal of the Royal Statistical Society. Series B, **41**, 113-147.

Berger, J.O., Bernardo, J.M. and Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics.* 37, 905-938.

Bernardo, J.M. and Smith, A.F.M., (2000). *Bayesian Theory.* Wiley, Chichester.

Corcuera, J. M. and Giummole F. (1999). A generalized Bayes rule for prediction. *Scandinavian Journal of Statistics*, 26, 265-279.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.

Jeffreys, H. (1961). *Theory of Probability (third edition)*, Oxford University Press, Oxford.

Ogura, T. and Yanagimoto, T. (2016). Improving and extending the McNemar test using the Bayesian method. *Statistics in Medicine*, 35. 2455-2466.

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1-32.

Robert, C.P. (2001). *The Bayesian Choice* Second ed. Springer, New York.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A.

(2002). Bayesian measures of model complexity and fit (with discussions). *Journal of the Royal Statistical Society, Series B*, 64, 583 - 639.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society, Series B*, 76, 485-493.

Yanagimoto, T. and Ohnishi, T. (2009). Bayesian prediction of a density function in terms of *e*-mixture. *Journal of Statistical Planning and Inference*, 39, 3064-3075.

Yanagimoto, T. and Ohnishi, T. (2011). Saddlepoint condition on a predictor to reconfirm the need for the assumption of a prior distribution. *Journal of Statistical Planning and Inference*, 41, 1990-2000.

Yanagimoto, T. and Ohnishi, T. (2014). Permissible boundary prior function as a virtually proper prior density. *Annals of the Institute of Statistical Mathematics*, 66, 789-809.

Yanagimoto T. and Ohnishi T. Conjugate analysis based on the plug-in optimal predictor, Manuscript.

内井 惣七 (1995). 科学哲学入門, 世界思想社