Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

6 Chap.6 Maximum Likelihood Methods

6.3 Maximum Likelihood Tests

例のごとく，$X_1, \ldots, X_n$ は pdf として $f(x;\theta)$ を持つ分布からの iid 標本とし，$\hat{\theta}$ を θ の mle とする．両側検定

$$H_0:\ \theta = \theta_0\ v.s.\ H_1:\ \theta \neq \theta_0 \tag{6.3.1}$$

を考える．尤度関数，対数尤度関数は，それぞれ，

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta)$$

$$l(\theta)\ \sum_{i=1}^{n} \log f(X_i;\theta)$$

であった．定理 6.1.1 は，θ_0 が θ の真値であるとすると，漸近的に $L(\theta_0)$ は $L(\theta)$ の最大値を達成するということを主張したものであった．このことから，尤度比

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} \tag{6.3.2}$$

は $\Lambda \leq 1$ であり，H_0 が正しい時，1 に近くなり，H_1 が正しい時，小さな値をとる．尤度比検定はこの性質を利用した検定である．尤度比検定の大前提となる定理をはじめに紹介する．

定理 6.3.1

定理 6.2.2 と同じ正則条件 ((R0)-(R5)) を仮定する．帰無仮説 $H_0:\ \theta = \theta_0$ のもとで，

$$-2log\Lambda \xrightarrow{D} \chi^2(1) \tag{6.3.7}$$

となる．

証明 6.1. $l(\theta)$ を θ_0 まわりで一次のテーラー展開を行い，$\theta = \hat{\theta}$ とすると，

$$l(\hat{\theta}) = l(\theta_0) + (\hat{\theta} - \theta_0)l'(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 l''(\theta_n^*) \tag{6.3.8}$$

となる．但し，ここで θ_n^* は $\hat{\theta}, \theta_0$ の間にある．$\hat{\theta} \xrightarrow{P} \theta_0$ から，$\theta_n^* \xrightarrow{P} \theta_0$ である．これと，$l''(\theta)$ が連続であることで，(6.2.22) 式から，

$$-\frac{1}{n}l''(\theta_n^*) \xrightarrow{P} I(\theta_0) \tag{6.3.9}$$

を得る．系 *6.2.3* から，

$$\frac{1}{\sqrt{n}}l'(\theta_0) = \sqrt{n}(\hat{\theta} - \theta_0)I(\theta_0) + R_n \tag{6.3.10}$$

である．ここで，$R_n \xrightarrow{P} 0$ である．もし，(6.3.9), (6.3.10) を (6.3.8) に代入し，整理すると，

$$-2\log\Lambda = 2(l(\hat{\theta}) - l(\theta_0)) = \{\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)\}^2 + R_n^* \tag{6.3.11}$$

を得る．ここで，$R_n^* \xrightarrow{P} 0$ である．定理 *5.2.4*，定理 *6.2.2* から，右辺第 *1* 項は $\chi^2(1)$ に収束する． □

検定統計量を $\chi_L^2 = -2\log\Lambda$ と定義すると，(6.3.1) の仮説に対して，定理 6.3.1 は判断基準

$$Reject\ H_0\ in\ favor\ of\ H_1\ if\ \chi_L^2 \geq \chi_\alpha^2(1) \tag{6.3.12}$$

を提案している．例を見てみよう．

例 6.1. 正規分布密の度関数の平均に関する尤度比検定 $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ とする．検定問題

$$H_0:\ \theta = \theta_0\ v.s.\ H_1:\ \theta \neq \theta_0$$

を検討せよ．

一旦，尤度比検定を置いといて，尤度に関する二つの検定を紹介する．統計量

$$\chi_W^2 = \left\{ \sqrt{nI(\theta)}(\hat{\theta} - \theta_0) \right\}^2 \tag{6.3.13}$$

を考える．これは $\hat{\theta}$ の漸近分布に基づく自然な検定統計量である．$I(\theta)$ は連続な関数であるため，帰無仮説のもとで，$I(\hat{\theta}) \xrightarrow{P} I(\theta_0)$ となる．つまり，χ_W^2 は帰無仮説のもとで，漸近的に $\chi^2(1)$ に従う．このことから，判断基準

$$Reject\ H_0\ in\ favor\ of\ H_1\ if\ \chi_W^2 \geq \chi_\alpha^2(1) \tag{6.3.13}$$

を利用することができる．χ_W^2, χ_L^2 の間には，

$$\chi_W^2 - \chi_L^2 \xrightarrow{P} 0 \tag{6.3.15}$$

という関係がある．検定問題 (6.3.14) は**ワルド検定** *(Wald-type test)* と呼ばれる．

三つ目の検定は，**スコア検定** *(scores-type test)*，**もしくは，ラオのスコア検定**と呼ばれる検定である．スコアとは，ベクトル

$$\mathbf{S}(\theta) = \left(\frac{\partial \log f(X_1;\theta)}{\partial \theta}, \cdots, \frac{\partial \log f(X_n;\theta)}{\partial \theta} \right)' \tag{6.3.16}$$

の成分である．ちょっと前を思い出すと

$$\frac{1}{\sqrt{n}} l'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta_0)}{\partial \theta} \tag{6.3.17}$$

だった．統計量

$$\chi_R^2 = \left(\frac{l'(\theta_0)}{\sqrt{nI(\theta_0)}} \right)^2 \tag{6.3.18}$$

は，帰無仮説のもとで，

$$\chi_R^2 = \chi_W^2 + R_{0n} \tag{6.3.19}$$

である．ここで，$R_{0n} \xrightarrow{P} 0$ である．よって判断基準は，

$$Reject\ H_0\ in\ favor\ of\ H_1\ if\ \chi_R^2 \geq \chi_\alpha^2(1) \tag{6.3.20}$$

となる．

　これらの三つの統計量のどれを使うべきだろうか．以上の議論から，帰無仮説のもとで，三つとも漸近的には同じ検定であった．検定に対する漸近相対効率と似たような概念を考えることもできるが，三つの検定は同じ漸近効率を持つため，あまり役に立たない．より詳しい議論は Lehmann(1999) の 7 章を参照してください．

¡note¿

6.4 Multiparameter Case: Estimation

　この節では，$\boldsymbol{\theta}$ が p 次元ベクトルの場合を考える．また，$X_1, \ldots, X_n$ を共通の pdf $f(x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$ を持つ分布からの iid 標本とする．未知パラメータ θ がスカラーの場合と同じように，尤度関数と対数尤度関数は，

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \\
l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i; \boldsymbol{\theta}) \qquad (6.4.1)
\end{aligned}
$$

となる．この節では，以前列挙した全ての正則条件を利用する．そのため，「正則条件のもとで」と述べた場合は，「(R0)-(R9) が満たされると仮定して」，と考えてほしい．

定理 6.4.1

$X_1,\ldots,X_n$ を共通の pdf $f(x;\boldsymbol{\theta}),\boldsymbol{\theta}\in\Omega\subset\mathbb{R}^p$ を持つ分布からの iid 標本とする．正則条件を仮定すると，

1. 尤度方程式
$$\frac{\partial}{\partial\boldsymbol{\theta}}l(\boldsymbol{\theta})=\mathbf{0}$$
は，$\hat{\boldsymbol{\theta}}_n\xrightarrow{P}\boldsymbol{\theta}$(一致推定量) なる解 $\hat{\boldsymbol{\theta}}_n$ を持つ．
2. (1) を満たすような列に対して，
$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{D}N_p(\mathbf{0},\boldsymbol{I}^{-1}(\boldsymbol{\theta}))$$
(漸近正規性) が成立する．

証明 6.2. この定理の証明は，より高次な書籍 *(例えば，Lehmann and Casella, 1998)* を参考にしてください．スカラーの場合と同様に，この定理は最尤推定量が一意であることを保証しない．しかし，もし，解の列が一意なら，それらは一致性と漸近正規性を持つ．応用事例では，解の列の一意性を仮定して解析することが多い．

系 6.4.1

$X_1,\ldots,X_n$ を共通の pdf $f(x;\boldsymbol{\theta}),\boldsymbol{\theta}\in\Omega\subset\mathbb{R}^p$ を持つ分布からの iid 標本とし，正則条件を仮定する．また，$\hat{\boldsymbol{\theta}}$ を尤度方程式の解の列とする．この時，$\hat{\boldsymbol{\theta}}$ は漸近的に有効な推定量となる，つまり，
$$\sqrt{n}(\hat{\theta}_{n,j}-\theta_j)\xrightarrow{D}N_p(0,[\boldsymbol{I}^{-1}(\boldsymbol{\theta})]_{jj})$$
である．また，$\boldsymbol{g}$ を $\boldsymbol{g}(\boldsymbol{\theta})=(g_1(\boldsymbol{\theta}),\ldots,g_k(\boldsymbol{\theta}))',(1\leq k\leq p)$ とし，$\boldsymbol{B}$ を偏微分を要素として持つ行列
$$\boldsymbol{B}=\left[\frac{\partial g_i}{\partial\theta_j}\right],\ i=1,\ldots,k,j=1,\ldots,p$$
とする．ただし，要素は連続で，$\boldsymbol{\theta}$ の近傍で消滅しない (=0 とならない) とする．$\hat{\boldsymbol{\eta}}=\boldsymbol{g}(\hat{\boldsymbol{\theta}})$ とすると，$\hat{\boldsymbol{\eta}}$ は $\boldsymbol{\eta}=\boldsymbol{g}(\boldsymbol{\theta})$ の最尤推定量である．定理 5.4.6 から，
$$\sqrt{n}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})\xrightarrow{D}N_p(\mathbf{0},\boldsymbol{B}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{B}')\tag{6.4.23}$$
となり，$\sqrt{n}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta})$ の情報行列は，もし，逆行列が存在するならば，
$$\boldsymbol{I}(\boldsymbol{\eta})=[\boldsymbol{B}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\boldsymbol{B}']^{-1}\tag{6.4.24}$$
として与えられる．

例を見てみましょう．

例 6.2. 正規分布の分散の情報量 $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ とする．

1. 情報行列 $\boldsymbol{I}(\mu, \sigma)$ を求めよ．
2. $g(\mu, \sigma) = \sigma^2$ という変換を用いて，σ^2 の情報量が，$1/2\sigma^4$ となることを示せ．

(解答)

6.0　付録：正則条件

正則条件をまとめておく．

6 章で用いる正則条件

($R0$) pdf が全く異なる (distinct)，つまり，$\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x;\theta')$
($R1$) pdf が任意の θ に関して共通の台 (support) を持つ
($R2$) 点 θ_0 が Ω の内点 (interior point) である
($R3$) pdf $f(x;\theta)$ が θ の関数として，二回微分可能
($R4$) 積分 $\int f(x;\theta)dx$ が θ の関数として，二回微分可能
($R5$) pdf $f(x;\theta)$ が θ の関数として三回微分可能であるとし，さらに，任意の θ について，

$$\left|\frac{\partial^3}{\partial\theta^3}\log f(x;\theta)\right| \leq M(x)$$
$$E_{\theta_0}[M(X)] < \infty, \forall\theta_0 - c < \theta < \theta_0 + c, and\ all\ x\ in\ the\ support\ of X$$

なる定数 c，関数 $M(x)$ が存在する．
($R6$) $\theta_0 \in \Omega_0$ なる開部分集合 $\Omega_0 \subset \Omega$ が存在して，任意の $\theta \in \Omega_0$ に対して，pdf $f(x;\theta)$ の全ての 3 階偏微分が存在する
($R7$) 以下の方程式

$$E_\theta\left[\frac{\partial}{\partial\theta_j}\log f(x;\theta)\right] = 0,\ for\ j = 1,\ldots,p$$
$$I_jk(\theta) = E_\theta\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(x;\theta)\right],\ for\ j,k = 1,\ldots,p$$

が成立する (期待値作用素 (=積分) と微分の順序交換が可能である)．($R8$) 任意 $\theta \in \Omega_0$ に対して，$\mathbf{I}(\theta)$ が正定値
($R9$) 以下を満たすような

$$\left|\frac{\partial^3}{\partial\theta_j\theta_k\theta_l}\log f(x;\theta)\right| \leq M_{jkl}(x),\ \forall\theta \in \Omega_0$$
$$E_{\theta_0}[M_{jkl}] < \infty,\ \forall j,k,l \in 1,\ldots,p$$

関数 $M_{jkl}(x)$ が存在する．

(やっぱり多い...!!!)