Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

6 Chap.6 Maximum Likelihood Methods

6.5 Multiparameter Case: Testing

今度は，未知パラメータが多次元ベクトルの時の仮説検定について考えよう．ここでも，$X_1, \dots, X_n$ は，pdf として $f(x;\boldsymbol{\theta})$ $(\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p)$ を持つ分布からの iid 標本とする．この節でも前節同様，正則条件を仮定した際は，(R0)-(R9) が満たされているものとする．仮説検定問題

$$H_0: \ \boldsymbol{\theta} \in \omega \ v.s. \ H_1: \ \boldsymbol{\theta} \in \Omega \cap \omega^c \tag{6.5.1}$$

を考える．ただし，$\omega \subset \Omega$ は $q(0 < q \le p)$ 個の独立な制約 $g_1(\boldsymbol{\theta}) = a_1, \dots, g_q(\boldsymbol{\theta}) = a_q$ で定義される集合である．また，関数 $g_1, \dots, g_q$ は連続的に微分可能であるとする．このことは，ω が $p-q$ 次元空間であることを示している．定理 6.1.1 から，パラメータの真値は尤度関数を最大化するということを考慮すると，検定統計量は，尤度の比

$$\Lambda = \frac{\max_{\boldsymbol{\theta} \in \omega} L(\boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta})} \tag{6.5.2}$$

となると直感的に考えられる．未知パラメータがスカラーの場合と同様に，H_0 が正しいと，Λ は 1 に近づき，H_1 が正しいなら，Λ は小さな値となる．

定理 6.5.1

$X_1, \ldots, X_n$ が pdf として $f(x;\boldsymbol{\theta})$ ($\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^p$) を持つ分布からの iid 標本とする．正則条件を仮定する．$\hat{\boldsymbol{\theta}}_n$ をパラメータ空間が全空間 Ω である際の尤度方程式の解の列とする．また，$\hat{\boldsymbol{\theta}}_{0,n}$ をパラメータ空間が $\omega(p-q\ dimension)$ に縮小された際の尤度方程式の解の列とする．帰無仮説のもとでの尤度と，対立仮説のもとでの尤度の比を

$$\Lambda = \frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} := \frac{L(\hat{\omega})}{L(\hat{\Omega})} \tag{6.5.4}$$

と表すと，尤度比検定の検定統計量 T は

$$T := -2\log\Lambda \xrightarrow{D} \chi^2(q) \tag{6.5.11}$$

となる．つまり，T は漸近的に自由度 q(帰無仮説のもとでパラメータが存在する空間の次元数) のカイ二乗分布に従う．

証明 6.1. 証明は，*Rao(1973)* をみてください．

スカラーの場合と同じように，パラメータが多次元となった場合にも，ワルド検定やスコア検定を行える．興味があったら，Lehmann(1999) を参照.

例 6.1. $X_1, \ldots, X_n \overset{iid}{\to} N(\mu, \sigma^2)$ とし，検定問題

$$H_0:\ \mu = \mu_0\ v.s.\ H_1:\ \mu \neq \mu_0$$

を考える．$\Omega = \{(\mu,\sigma^2):\ -\infty < \mu < \infty, \sigma^2 > 0\}$，$\omega = \{(\mu_0,\sigma^2):\ \sigma^2 > 0\}$ とした時の各最大尤度の尤度比 Λ を求めよ．
¡解答¿

¡note¿

6.6 The EM Algorithm

実データ分析にて，データの一部が欠損していることがよくある (ほんまに)．本節では，そういったケースにおいて最尤推定量を計算する手法である EM アルゴリズムを紹介する．さらっと説明するので，より詳しく知りたい方は，McLachlan とかで検索してみたらいいと思う．

まず，サイズ n のアイテムを考える．ただし，この中で，n_1 個のアイテムは観測され，$n_2 = n - n_1$ 個のアイテムは観測されないとする．観測されたアイテムを $\boldsymbol{X}' = (X_1, X_2, \ldots, X_{n_1})$ として表し，観測されないアイテムを $\boldsymbol{Z}' = (Z_1, Z_2, \ldots, Z_{n_2})$ と表す．X_i は pdf $f(x|\theta), \theta \in \Omega$ の分布からの iid 標本とし，観測されないアイテム Z_j と観測されるアイテム X_i は互いに独立と仮定する．また，$g(x|\theta)$ で $\boldsymbol{X}$ の同時密度を表し，$h(\boldsymbol{x}, \boldsymbol{z}|\theta)$ で観測されたアイテム $\boldsymbol{X}$ と観測されないアイテム $\boldsymbol{Z}$ の同時分布を表し，$k(\boldsymbol{z}|\theta, \boldsymbol{x})$ で，観測されるデータの情報が与えられているもとでの観測されないデータ条件付き分布を表すとする．条件付き pdf の定義から，

$$k(\boldsymbol{z}|\theta, \boldsymbol{x}) = \frac{h(\boldsymbol{x}, \boldsymbol{z}|\theta)}{g(\boldsymbol{x}|\theta)} \tag{6.6.1}$$

である．$L(\theta|\boldsymbol{x}) = g(\boldsymbol{x}|\theta)$ を**観測された尤度関数**と呼び，

$$L^c(\theta|\boldsymbol{x}, \boldsymbol{z}) = h(\boldsymbol{x}, \boldsymbol{z}|\theta) \tag{6.6.2}$$

を**完全尤度関数**と呼ぶ．我々の目的は，完全尤度 $L^c(\theta|\boldsymbol{x}, \boldsymbol{z})$ を用いて，尤度関数 $L(\theta|\boldsymbol{x})$ を最大化することである．

(6.6.1) から，任意の $\theta_0 \in \Omega$ に対して，

$$\begin{aligned}
\log L(\theta|\boldsymbol{x}) &= \int \log L(\theta|\boldsymbol{x}) k(\boldsymbol{z}|\theta_0, \boldsymbol{x}) d\boldsymbol{z} \\
&= \int \log g(\boldsymbol{x}|\theta) k(\boldsymbol{z}|\theta_0, \boldsymbol{x}) d\boldsymbol{z} \\
&= \int [\log h(\boldsymbol{x}, \boldsymbol{z}|\theta) - \log k(\boldsymbol{z}|\theta, \boldsymbol{x})] k(\boldsymbol{z}|\theta_0, \boldsymbol{x}) d\boldsymbol{z} \\
&= \int \log[h(\boldsymbol{x}, \boldsymbol{z}|\theta)] k(\boldsymbol{z}|\theta_0, \boldsymbol{x}) d\boldsymbol{z} - \int \log[k(\boldsymbol{z}|\theta, \boldsymbol{x}] k(\boldsymbol{z}|\theta_0, \boldsymbol{x}) d\boldsymbol{z} \\
&= E_{\boldsymbol{z}|\theta_0, \boldsymbol{x}}[\log L^c(\theta|\boldsymbol{x}, \boldsymbol{Z})|\theta_0, \boldsymbol{x}] - E_{\boldsymbol{z}|\theta_0, \boldsymbol{x}}[\log k(\boldsymbol{Z}|\theta, \boldsymbol{x})|\theta_0, \boldsymbol{x}]
\end{aligned} \tag{6.6.3}$$

という結果が得られる．特に第 1 項を

$$Q(\theta|\theta_0, \boldsymbol{x}) := E_{\boldsymbol{z}|\theta_0, \boldsymbol{x}}[\log L^c(\theta|\boldsymbol{x}, \boldsymbol{Z})|\theta_0, \boldsymbol{x}] \tag{6.6.4}$$

と定義し，これを Q 関数と呼び，これは，EM アルゴリズムの E-step として知られている．θ の最初の推定値を $\hat{\theta}^{(0)}$ とし，$\hat{\theta}^{(1)}$ を $Q(\theta|\theta_0, \boldsymbol{x})$ を最大化する θ の推定値とする．これを繰り返すと，θ の推定値の数列 $\hat{\theta}^{(m)}$ が得られる．ここで，EM アルゴリズムの全体像を先にまとめておく．

Algorithm 6.6.1(EM アルゴリズム)

$\hat{\theta}^{(m)}$ を m 番目のステップでの推定値とする．$m+1$ 番目の step での推定値を計算するために，

1. Expectation step:

$$Q(\theta|\hat{\theta}^{(m)}, \boldsymbol{x}) = E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}[\log|^{c}(\theta|\boldsymbol{x}, \boldsymbol{Z})|\hat{\theta}^{(m)}, \boldsymbol{x}] \tag{6.6.5}$$

を計算する．

2.

$$\hat{\theta}^{(m+1)} = argmaxQ(\theta|\hat{\theta}^{(m)}, \boldsymbol{x}) \tag{6.6.6}$$

なる θ を $\hat{\theta}^{(m+1)}$ とする．

強い仮定のもとで，$m \to \infty$ の時，$\hat{\theta}^{(m)}$ **は最尤推定量に確率収束する**ことが知られています．このことは，ここでは示さないので，調べてててください．

定理 6.6.1

アルゴリズム 6.6.1 で定義された，推定値 $\hat{\theta}^{(m)}$ の数列は

$$L(\hat{\theta}^{(m+1)}|\boldsymbol{x}) \geq L(\hat{\theta}^{(m)}|\boldsymbol{x}) \tag{6.6.7}$$

を満たす．

証明 6.2. $\hat{\theta}^{(m+1)}$ は $Q(\theta|\hat{\theta}^{m}, \boldsymbol{x})$ を最大化するため，

$$Q(\hat{\theta}^{(m+1)}|\hat{\theta}^{(m)}, \boldsymbol{x}) \geq Q(\hat{\theta}^{(m)}|\hat{\theta}^{(m)}, \boldsymbol{x})$$

であり，これは，

$$E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}[\log L^{c}(\hat{\theta}^{(m+1)}|\boldsymbol{x}, \boldsymbol{Z})] \geq E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}[\log L^{c}(\hat{\theta}^{(m)}|\boldsymbol{x}, \boldsymbol{Z})] \tag{6.6.8}$$

と等しい．ただしここで，期待値は，$k(\boldsymbol{z}|\hat{\theta}^{(m)}, \boldsymbol{x})$ のもとでとられている．(6.6.3) から，

$$E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}[\log k(\boldsymbol{Z}|\hat{\theta}^{(m+1)}, \boldsymbol{x})] \leq E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}[\log k(\boldsymbol{Z}|\hat{\theta}^{(m)}, \boldsymbol{x})] \tag{6.6.9}$$

となることを示せば，証明完了である．イェンセンの不等式から，

$$\begin{aligned}
E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}\left\{\log\left[\frac{k(\boldsymbol{Z}|\hat{\theta}^{(m+1)}, \boldsymbol{x})}{k(\boldsymbol{Z}|\hat{\theta}^{(m)}, \boldsymbol{x})}\right]\right\} &\leq \log E_{\boldsymbol{z}|\hat{\theta}^{(m)},\boldsymbol{x}}\left[\frac{k(\boldsymbol{Z}|\hat{\theta}^{(m+1)}, \boldsymbol{x})}{k(\boldsymbol{Z}|\hat{\theta}^{(m)}, \boldsymbol{x})}\right] \\
&= \log\int\frac{k(\boldsymbol{z}|\hat{\theta}^{(m+1)}, \boldsymbol{x})}{k(\boldsymbol{z}|\hat{\theta}^{(m)}, \boldsymbol{x})}k(\boldsymbol{z}|\hat{\theta}^{(m)}, \boldsymbol{x})d\boldsymbol{z} \\
&= \log(1) \\
&= 0
\end{aligned} \tag{6.6.10}$$

よって，(6.6.9) が成立する．以上から，定理の結果を得る． □

例 **6.2.** $Y_1 \sim N(\mu_1, \sigma_1^2), Y_2 \sim N(\mu_2, \sigma_2^2), W \sim Bin(1, \epsilon), X = (1-W)Y_1 + WY_2$ とする．ここで，W は Y_1, Y_2 と独立である．また，$\boldsymbol{\theta}' = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \epsilon)$ とする．確率変数 X の pdf は

$$f(x) = (1-\epsilon)f_1(x) + \epsilon f_2(x),\ -\infty < x < \infty \tag{6.6.17}$$

ただしここで，$f_j(x) = \sigma_j^{-1}\phi[(x - mu_j)/\sigma_j]$ である．この pdf を持つ分布からの観測された標本を $\boldsymbol{X}' = (X_1, X_2, \ldots, X_n)$ と表す．この時，対数尤度関数は，

$$l(\boldsymbol{\theta}|x) = \sum_{i=1}^{n} \log[(1-\epsilon)f_1(x_i) + \epsilon f_2(x_i)] \tag{6.6.18}$$

となる．この問題において，観測されないデータは，どの分布に所属するかを表す確率変数である．$i = 1, \ldots, n$ にたいして，

$$W_i = \begin{cases} 0 & if\ X_i\ has\ pdf\ f_1(x) \\ 1 & if\ X_i\ hat\ pdf\ f_2(x) \end{cases}$$

とする．これらの確率変数は，成功確率 ϵ のベルヌーイ分布からの無作為標本 *(=iid* 標本*)* である．完全尤度は，

$$L^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{w}) = \prod_{W_i=0} f_1(x_i) \prod_{W_i=1 f_2(x_i)}$$

であり，完全尤度の対数をとると，

$$\begin{aligned} l^c(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{w}) &= \sum_{W_i=0} \log f_1(x_i) + \sum_{W_i=1} \log f_2(x_i) \\ &= \sum_{i=1}^{n} [(1-w_i)\log f_1(x_i) + w_i \log f_2(x_i)] \end{aligned} \tag{6.6.19}$$

となる．$Estep$ のために，$\boldsymbol{x}, \boldsymbol{\theta}_0$ を与えたもとでの W_i の条件付き期待値，

$$E_{\boldsymbol{\theta}_0}[W_i|\boldsymbol{\theta}_0, \boldsymbol{x}] = P[W_i = 1|\boldsymbol{\theta}_0, \boldsymbol{x}]$$

が必要である．この期待値の推定値として，x_i が分布 $f_2(x)$ からの標本であるとした時の尤度，

$$\gamma_i = \frac{\hat{\epsilon} f_{2,0}(x_i)}{(1-\hat{\epsilon})f_{1,0}(x_i) + \hat{\epsilon} f_{2,0}(x_i)} \tag{6.6.20}$$

を利用可能である．ただし，添字の 0 はパラメータ $\boldsymbol{\theta}_0$ とした時の分布が利用されたことを示している．(6.6.19) の w_i を γ_i に置き換えると，$Mstep$ では，

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{x}) = \sum_{i=1}^{n} [(1-\gamma_i)\log f_1(x_i) + \gamma_i \log f_2(x_i)] \tag{6.6.21}$$

を最大化すれば良いということがわかる．この最大化は，各パラメータに関して，$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0,\boldsymbol{x})$ を微分した際の停留点を更新式として利用すれば達成される．色々省くが，それぞれの更新式は，

$$\begin{aligned}
\hat{\mu}_1 &= \frac{\sum_{i=1}^n (1-\gamma_i)x_i}{\sum_{i=1}^n (1-\gamma_i)} \\
\hat{\sigma}_1^2 &= \frac{\sum_{i=1}^n (1-\gamma_i)(x_i-\hat{\mu}_1)^2}{\sum_{i=1}^n (1-\gamma_i)} \\
\hat{\mu}_2 &= \frac{\sum_{i=1}^n \gamma_i x_i}{\sum_{i=1}^n \gamma_i} \\
\hat{\sigma}_2^2 &= \frac{\sum_{i=1}^n \gamma_i(x_i-\hat{mu}_2)^2}{\sum_{i=1}^n \gamma_i} \\
\hat{\epsilon} &= \frac{1}{n}\sum_{i=1}^n \gamma_i
\end{aligned}$$

となる．これを収束するまで更新し続ける．

例 6.3. 上の更新式を導出してみましょう．
¡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

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

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

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

関数 $M_{jkl}(x)$ が存在する．

(いつみても多い...!!!)