Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

6 Chap.6 Maximum Likelihood Methods

6.1 Maximum Likelihood Estimation

pdf が $f(x;\theta)$ であり，その形状は未知パラメータ $\theta \in \Omega$ に依存するような確率変数 X を考える．以下の議論は全て連続型確率変数に対してだが，離散型確率変数の場合も同様の結果が得られる．$X_1, X_2, \ldots, X_n$ を X の iid 標本とする．つまり，$X_1, X_2, \ldots, X_n$ は共通の pdf $f(x;\theta), \theta \in \Omega$ を持つ．本節では，未知パラメータ θ はスカラーと仮定する．我々は，尤度関数に基づいて推論を行う．尤度関数とは，$\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ を用いて

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta),\ \theta \in \Omega \tag{6.1.1}$$

で定義される θ の関数である．この関数は，積で定義されているが，積よりも和の方が扱いやすいので，対数変換した，対数尤度関数

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta),\ \theta 1 \Omega \tag{6.1.2}$$

がよく利用される．log は一対一の変換なので，ここで情報の損失はないことに注意!

4 章の時と同様に，我々の点推定値は $L(\theta)$ を最大化する $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ である．この推定量は θ の**最尤推定量** *(maximum likelihood estimator, mle)* と呼ばれる．以下では，θ_0 を θ の真値とする．また，本章で紹介する定理で利用する全ての正則条件を，あらかじめ，ここで紹介しておく．

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．

(多い...!!!)

最初の仮定は，**パラメータ θ が分布を同定する**ということを述べている．2 番目の仮定は，X_i の台は θ に依存しないことを示唆している．これはかなり限定的な仮定であり，テキストで紹介されているいくつかの例では，満たされていない．

定理 6.1.1

$(R0), (R1)$ のもとで，

$$\underset{n\to\infty\theta_0}{P}\left[L(\theta_0, \mathbf{X}) > L(\theta, \mathbf{X})\right] = 1,\ \forall\theta \neq \theta_0 \tag{6.1.3}$$

が成り立つ．

証明 **6.1.** 対数をとると，$L(\theta_0, \mathbf{X}) > L(\theta, \mathbf{X})$ は，

$$\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right] < 0$$

と等しい．和は有限の期待値を持つ iid 標本に対するものであり，さらに，$\phi(x) = -\log(x)$ 狭義凸であることと，大数の法則とイェンセンの不等式から，θ_0 が真のパラメータの時，

$$\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right] \xrightarrow{P} E_{\theta_0}\left[\log\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right] < \log E_{\theta_0}\left[\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right]$$

となるが，

$$E_{\theta_0}\left[\frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right] = \int fracf(x;\theta)f(x;\theta_0)f(x;\theta_0)dx = 1$$

である．つまり，二つ上の最右辺は，$\log 1 = 0$ から，0 となる．よって定理の結果を得る．最後の等式を得るために，$f(x;\theta), f(x;\theta_0)$ が共通の台を持つことが必要であることに注意.

定理 6.1.1 は尤度関数は漸近的に θ_0(の付近) で最大化されることを主張している．そのため，尤度関数を最大化する θ を真値 θ_0 の推定量とすることは自然な考えである．

¡note¿

定義 6.1.1(最尤推定量)

$\hat{\theta} = \hat{\theta}(\mathbf{X})$ が

$$\hat{\theta} = argmax_{\theta}L(\theta, \mathbf{X}) \tag{6.1.4}$$

を満たす時，θ の最尤推定量であるという．

¡note¿

最尤推定量は，方程式

$$\frac{\partial l(\theta)}{\partial \theta} = 0 \tag{6.1.5}$$

の解である．この方程式は，推定方程式 (estimating equation) の一例であり，特に尤度方程式 (likelihood equation) と呼ばれる．

定理 6.1.2

$X_1, \ldots, X_n$ を独立で，共通の pdf $f(x;\theta), \theta \in \Omega$ を持つ (=iid) 標本とする．ある関数 g を用いて，$\eta = g(\theta)$ を興味のあるパラメータとする．また，$\hat{\theta}$ を θ の mle とする．この時，$g(\hat{\theta})$ は $\eta = g(\theta)$ の mle となる．

証明 6.2. まず，g は一対一対応の関数とする．興味の対象となる尤度は，$L(g(\theta))$ である．ここで，g は一対一対応なので，

$$max L(g(\theta)) = \max_{\eta = g(\theta)} L(\eta) = \max_{\eta} L(g^{-1}(\eta))$$

である．ここで，$g^{-1}(\eta) = \hat{\theta}$ の時，最大値を達成する．

次に，g が一対一対応でないとする．すべての g の値域の η について，集合 *(=逆像，preimage)*

$$g^{-1}(\eta) = \{\theta : g(\theta) = \eta\}$$

を定義する．尤度関数の最大値は $\theta = \hat{\theta}$ の時達成され，g の領域は Ω であり，これは θ を含む領域である．したがって，$\hat{\theta}$ はこれらの逆像のうちの一つの中に含まれる *(*特に，一つの逆像に含まれている*)*．以上から，$L(\eta)$ を最大化するためには，$g^{-1}(\eta)$ が $\hat{\theta}$ を含むような唯一の逆像となるような η を選び，$\hat{\eta} = g(\hat{\theta})$ とすれば良い．

¡note¿

定理 6.1.3(最尤推定量は一致推定量になりうる)

$X_1, \ldots, X_n$ が正則条件 (R0)-(R2) を満たすとする．また，さらに Ω で $f(x;\theta)$ が θ に関して微分可能であるとする．この時，尤度方程式

$$\frac{\partial}{\partial\theta}L(\theta) = 0$$

$$\frac{\partial}{\partial\theta}l(\theta) = 0$$

は，

$$\hat{\theta}_n \xrightarrow{P} \theta_0$$

なる解 $\hat{\theta}_n$ を持つ．

証明 6.3. θ_0 は Ω の内点である，つまり，ある $a > 0$ を用いて $(\theta_0 - a, \theta_0 + a) \subset \Omega$ と表せる．S_n を

$$S_n = \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 - a; \mathbf{X})\} \cap \{\mathbf{X} : l(\theta_0; \mathbf{X}) > l(\theta_0 + a; \mathbf{X})\}$$

であるような事象と定義する．定理 *6.1.1* から，$P(S_n) \to 1$ である．そのため，事象 S_n に限定して考えることが可能である．しかし，S_n では，$l(\theta)$ は $\theta_0 - a < \hat{\theta}_n < \theta_0 + a, l'(\hat{\theta}_n)$ を満たす局所的な最適解 $\hat{\theta}_n$ を持ちうる．つまり，

$$S_n \subset \left\{\mathbf{X} : |\hat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\hat{\theta}_n(\mathbf{X})) = 0\right\}$$

である．それゆえに，

$$1 = \lim_{n\to\infty} P(S_n) \leq \overline{\lim_{n\to\infty}} P\left[\left\{\mathbf{X} : |\hat{\theta}_n(\mathbf{X}) - \theta_0| < a\right\} \cap \left\{\mathbf{X} : l'(\hat{\theta}_n(\mathbf{X})) = 0\right\}\right] \leq 1$$

が成立する．このことは，解 $\hat{\theta}_n$ の列が $P[|\hat{\theta}_n - \theta_0| < a] \to 1$ であることを示している．

これまでの証明で，論議を生みそうな点は，解の列が a に依存するのではないかという点である．しかし，我々は，θ_0 に”最も近い”ものを次のような手順で選ぶことが可能である．それぞれの n に対して，区間内の全ての解の集合は有界であり，そのため，θ_0 に最も近い解の中に下限 *(infimum)* が存在する．これを最適解とすれば良い．

¡note¿

この定理は，方程式の複数の解について議論している点で曖昧である．しかし，もし，mle が $l'(\theta)=0$ の一意な解であるならば，mle は一致性を持つ．これを以下の系にまとめる．

系 6.1.1

$X_1,\ldots,X_n$ が正則条件 (R0)-(R2) を満たすとする．また，さらに Ω で $f(x;\theta)$ が θ に関して微分可能であるとする．さらに，尤度方程式が一意な解 $\hat{\theta}_n$ を持つとする．この時，$\hat{\theta}_n$ は θ_0 の一致推定量である．

¡note¿

6.2 Rao-Cramér Lower Bound and Efficiency

Rao-Cramér の不等式 (Rao-Cramér lower bound) は，不偏推定量の分散の下界を与える不等式である．正則条件のもと，最尤推定量の分散は，漸近的にこの下界を達成するということを示す．

この節では，(R0)-(R2) に加えて，(R3),(R4) も成立すると仮定する．正則条件 (R1)-(R4) は，パラメータは，$f(x;\theta)>0$ なる区間の端点ではないということと，θ に関して微分と積分の順序交換が可能であるということを意味している．以下では，連続の場合の導出をするが，離散の場合も同様の論理で導出が可能である．まず，pdf の全区間での積分が 1 となること

$$1=\int_{-\infty}^{\infty} f(x;\theta)dx$$

を利用する．両辺を θ について偏微分すると，

$$0=\int_{-\infty}^{\infty} \frac{\partial f(x;\theta)}{\partial\theta}dx$$

となる．これは，

$$0=\int_{-\infty}^{\infty} \frac{\partial f(x;\theta)/\partial\theta}{f(x;\theta)}f(x;\theta)dx$$

と等しい．またさらに，これは

$$0=\int_{-\infty}^{\infty}\frac{\partial\log f(x;\theta)}{\partial\theta}f(x;\theta)dx \tag{6.2.1}$$

と等しい．(6.2.1) を期待値の形で書き直すと，

$$E\left[\frac{\partial\log f(X;\theta)}{\partial\theta}\right]=0 \tag{6.2.2}$$

とかける．これは，確率変数 $\frac{\partial\log f(X;\theta)}{\partial\theta}$ の平均が 0 であることを表している．もし，(6.2.1) をもう一回偏微分すると，

$$0=\int_{-\infty}^{\infty}\frac{\partial^2\log f(x;\theta)}{\partial\theta^2}f(x;\theta)dx+\int_{-\infty}^{\infty}\frac{\partial\log f(x;\theta)}{\partial\theta}\frac{\partial\log f(x;\theta)}{\partial\theta}f(x;\theta)dx \tag{6.2.3}$$

右辺第二項は，**フィッシャーの情報量** *(Fisher information)* と呼ばれる期待値として書き直せる．つまり，フィッシャー情報量 $I(\theta)$ は，

$$I(\theta)=\int_{-\infty}^{\infty}\frac{\partial\log f(x;\theta)}{\partial\theta}\frac{\partial\log f(x;\theta)}{\partial\theta}f(x;\theta)dx=E\left[\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right)^2\right] \tag{6.2.4}$$

で定義される．式 (6.2.3) から，$I(\theta)$ は

$$I(\theta)=-\int_{-\infty}^{\infty}\frac{\partial^2\log f(x;\theta)}{\partial\theta^2}f(x;\theta)dx=-E\left[\frac{\partial^2\log f(X;\theta)}{\partial\theta^2}\right] \tag{6.2.5}$$

とも計算できる．(6.2.2) から，フィッシャー情報量は，

$$I(\theta)=E\left[\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right)^2\right]-0^2=V\left(\frac{\partial\log f(X;\theta)}{\partial\theta}\right) \tag{6.2.6}$$

つまり，確率変数 $\frac{\partial\log f(X;\theta)}{\partial\theta}$ の分散としても表酢ことが可能である．一般的に，(6.2.4) より，(6.2.5) の方が計算しやすい．

Remark 6.1. 情報量は，

$$\left[\frac{\log f(x;\theta)}{\partial\theta}\right]^2 \ \ or \ -\frac{\partial^2\log f(x;\theta)}{\partial\theta^2}$$

のいずれかの重みを $f(x;\theta)$ とした時の重み付き平均である．つまり，これらの偏微分が，平均的に大きければ大きいほど，我々は θ に関するより多くの情報を得ることができる．もし，これらが *0* と等しいならば，θ に関する情報は *0* である．関数

$$\frac{\partial\log f(x;\theta)}{\partial\theta}$$

はスコア関数 (score function) と呼ばれる．これが，*mle* の推定方程式の元となることを思い出すと，$mle\hat{\theta}$ は

$$\sum_{i=1}^{n}\frac{\partial\log f(x_i;\theta)}{\partial\theta}=0$$

の解となる．

例 6.1. ベルヌーイ分布に従う確率変数のフィッシャー情報量

標本数が 1 ならば，(6.2.6) から，フィッシャー情報量は確率変数 $frac\partial \log f(X_1;\theta)\partial\theta$ の分散である．では，サンプルサイズが n となるとどうなるのだろうか．$X_1,\dots,X_n$ を pdf として $f(x;\theta)$ を持つ分布からの無作為標本とする．尤度 $L(\theta)$ は無作為標本の pdf であり，分散が標本のフィッシャー情報量であるような確率変数は，

$$\frac{\partial \log f(\theta,\mathbf{X})}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial \log f(x_i;\theta)}{\partial\theta}$$

で与えられる．$X_1,\dots,X_n$ は無作為標本なので，iid であり，共通な分散 $I(\theta)$ をもつ．そのため，標本の持つ情報は，

$$Var\left(\frac{\partial \log f(\theta,\mathbf{X})}{\partial\theta}\right) = nI(\theta) \tag{6.2.11}$$

となる．ゆえに，サンプルサイズ n の標本の情報量は，サンプルサイズ 1 の標本の情報量の n 倍である．

定理 6.2.1(Rao-Cramér Lower Bound)

$X_1,\dots,X_n$ を $f(x;\theta)$ をpdf として持つ分布からの無作為標本とする．正則条件 (R0)-(R4) を仮定する．$Y=u(X_1,\dots,X_n)$ を平均が $E[Y]=E[u(X_1,\dots,X_n)]=k(\theta)$ であるような統計量とする．この時，

$$Var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)} \tag{6.2.12}$$

である．

証明 6.4. Y の平均を

$$k(\theta)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}u(x_1,\dots,x_n)f(x_1,\theta)\cdots f(x_n,\theta)dx_1\cdots dx_n$$

と表す．両辺を θ に関して微分すると，

$$\begin{aligned}
k'(\theta) = & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \left[\sum_{i=1}^{n} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial \theta} \right] \\
& \times f(x_1, \theta) \cdots f(x_n, \theta) dx_1 \cdots dx_n \\
= & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \left[\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial \theta} \right] \\
& \times f(x_1, \theta) \cdots f(x_n, \theta) dx_1 \cdots dx_n
\end{aligned} \tag{6.2.13}$$

ここで，確率変数 Z を $Z = \sum_{i=1}^{n} [\partial \log f(X_i;\theta)/\partial\theta]$ と定義する．(6.2.2), (6.2.11) から，$E(Z) = 0, Var(Z) = nI(\theta)$ である．(6.2.13) を $k'(\theta) = E(YZ)$ と考えると，

$$k'(\theta) = E(YZ) = E(Y)E(Z) + \rho\sigma_Y \sqrt{nI(\theta)}$$

となる．ここで，ρ は Y, Z の相関係数である．$E(Z) = 0$ から，

$$\rho = \frac{k'(\theta)}{\sigma_Y \sqrt{nI(\theta)}}$$

であり，$\rho^2 \leq 1$ であることを利用すると，

$$\frac{[k'(\theta)]^2}{\sigma_Y^2 nI(\theta)} = \frac{[k'(\theta)]^2}{Var(Y)^2 nI(\theta)} \leq 1$$

であり，定理の結果を得る． □

系 6.2.1

定理 6.2.2 の仮定のもと，もし Y が θ の不偏推定量ならば，$k(\theta) = \theta$ であり，この時，クラメールラオの不等式は，

$$Var(Y) \geq \frac{1}{nI(\theta)}$$

となる．

¡note¿

定義 6.2.1(有効推定量)

Y は θ の点推定にて不偏推定量とする．Y の分散が，Rao-Cramér の下限を達成するとき，統計量 Y を θ の有効推定量 (efficient estimator) であるという．

¡note¿

定義 6.2.2(有効性)

パラメータの不偏推定量の実際の分散と，Rao-Cramér の下限の比を推定量の有効性 (efficiency) という．

例 6.2. ポアソン分布の標本平均が有効推定量であることを示す

定理 6.2.2(最尤推定量の漸近正規性)

$X_1, \ldots, X_n$ を，正則条件 (R0)-(R5) を満たすような pdf $f(x;\theta_0), \theta_0 \in \Omega$ を持つ分布からの iid 標本とする．さらに，フィッシャー情報量が，$0 < I(\theta_0) < \infty$ を満たすとする．この時，尤度方程式をの解の列は

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta)}\right) \tag{6.2.18}$$

を満たす．

証明 6.5. 対数尤度関数の一次微分 $l'(\theta)$ を θ_0 の周りで *2* 次までのテーラー展開を行い，$\theta = \hat{\theta}_n$ とすると，

$$l'(\hat{\theta}_n) = l'(\theta_0) + (\hat{\theta}_n - \theta_0)l''(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 l'''(\theta_n^*) \tag{6.2.19}$$

となる．ただしここで，θ_n^* は $\theta_0, \hat{\theta}_n$ の間にある．また，$l'(\hat{\theta}_n) = 0$ である．よって，整理すると，

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{n^{-1/2} l'(\theta_0)}{-n^{-1} l''(\theta_0) - (2n)^{-1}(\hat{\theta}_n - \theta_0) l'''(\theta_n^*)} \tag{6.2.20}$$

を得る *((6.2.19)* と *(6.2.20)* が等しいことを確かめてみましょう*)*．中心極限定理から，

$$\frac{1}{\sqrt{n}} l'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta_0)}{\partial \theta} \xrightarrow{D} N(0, I(\theta_0)) \tag{6.2.21}$$

である．ただし，$Var(l'(\theta_0)) = (1/n)\sum_{i=1}^{n} Var(\partial \log f(X_i;\theta_0)/\partial\theta) = I(\theta_0) < \infty$ を利用した．また，大数の法則から，

$$-\frac{1}{n} l''(\theta_0) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(x_i;\theta_0)}{\partial \theta^2} \xrightarrow{P} I(\theta_0) \tag{6.2.22}$$

である．よって，あとは，(6.2.20) の分母の第 *2* 項が *0* に確率収束することを示せばよい．$n^{-1} l'''(\theta_n^*)$ が確率有界であるなら，定理 5.2.7 が使えて，$\hat{\theta}_n - \theta_0 \xrightarrow{P} 0$ である．よって以下では，$n^{-1} l'''(\theta_n^*)$ が確率有界であることを示す．c_0 を *(R5)* の条件を満たすような定数とする．$|\hat{\theta}_n - \theta_0| < c_0$ ならば，$|\theta_n^* - \theta_0| < c_0$ であり，*(R5)* から，

$$\left| -\frac{1}{n} l'''(\theta_n^*) \right| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\partial^3 \log f(X_i;\theta)}{\partial \theta^3} \right| \leq \frac{1}{n} \sum_{i=1}^{n} M(X_i) \tag{6.2.23}$$

が成立する．*(R5)* から，$E_{\theta_0}[M(X)] < \infty$ であり，大数の法則から，$(1/n)\sum_{i=1}^{n} M(X_i) \xrightarrow{P} E_{\theta_0}[M(X)]$ となる．境界として $1 + E_{\theta_0}[M(X)]$ を選べばよい．$\epsilon > 0$ が与えられているとする．

$$n \geq N_1 \Rightarrow P[|\hat{\theta}_n - \theta_0| < c_0] \geq 1 - \frac{\epsilon}{2} \tag{6.2.24}$$

$$n \geq N_2 \Rightarrow P\left[\left| \frac{1}{n} \sum_{i=1}^{n} M(X_i) - E_{\theta_0}[M(X)] \right| < 1 \right] \geq 1 - \frac{\epsilon}{2} \tag{6.2.25}$$

となるような整数 N_1, N_2 を選ぶ．この時，(6.2.23) – (6.2.25) から，

$$n \geq \max\{N_1, N_2\} \Rightarrow P\left[\left| \frac{1}{n} l'''(\theta_n^*) \right| \leq 1 + E_{\theta_0}[M(X)] \right] \geq 1 - \frac{\epsilon}{2}$$

となり，このことから，$n^{-1} l'''(\theta_n^*)$ は確率有界であることがわかる．よって，定理の結果を得る． □

以下では，定義 6.2.1,6.2.2 を漸近的な場合を考慮し，一般化する．

定義 6.2.3

$X_1,\ldots,X_n$ が pdf が $f(x;\theta)$ であるような分布からの無作為標本 (=iid) であるとする．$\hat{\theta}_{1n}=\hat{\theta}_{1n}(X_1,\ldots,X_n)$ を $\sqrt{n}(\hat{\theta}_{1n}-\theta_0)\xrightarrow{D}N(0,\sigma^2_{\hat{\theta}_{1n}})$ となるような θ_0 の推定量とする．この時，

1. $\hat{\theta}_{1n}$ の漸近効率は
$$e(\hat{\theta}_{1n})=\frac{1/I(\theta_0)}{\sigma^2_{\hat{\theta}_{1n}}} \tag{6.2.26}$$
と定義される．
2. 推定量 $\hat{\theta}_{1n}$ は漸近有効であるとは，(6.2.26) の比率が 1 となることをいう．
3. $\hat{\theta}_{2n}$ を $\sqrt{n}(\hat{\theta}_{2n}-\theta_0)\xrightarrow{D}N(0,\sigma^2_{\hat{\theta}_{2n}})$ となるような別の推定量とする．この時，$\hat{\theta}_{2n}$ に対する $\hat{\theta}_{1n}$ の漸近相対効率 (asymptotic relative efficiency, ARE) は，
$$e(\hat{\theta}_{1n},\hat{\theta}_{2n})=\frac{\sigma^2_{\hat{\theta}_{2n}}}{\sigma^2_{\hat{\theta}_{1n}}} \tag{6.2.27}$$
で与えられる．

したがって，定理 6.2.2 から，**正則条件のもとでは，最尤推定量は漸近有効推定量となることがわかる．**また，もし，二つの推定量が漸近的に正規性を持ち，漸近的な平均が同じならば，漸近分散が小さい方が良い推定量として選ばれることが直感的にわかるだろう．

¡note¿

系 6.2.2

定理 6.2.2 の仮定のもと，$g(x)$ を x に関して連続で，θ_0 において微分可能かつ，$g'(\theta) \neq 0$ とする．この時，

$$\sqrt{n}g(\hat{\theta}_n) - g(\theta_0) \xrightarrow{D} N\left(0, \frac{g'(\theta_0)^2}{I(\theta_0)}\right) 6.2.30) \tag{()}$$

この系の証明は，Δ 法 (定理 5.2.9)，定理 6.2.2 から直ちに導かれる．

系 6.2.3

定理 6.2.2 の仮定のもと，

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{I(\theta_0)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta_0)}{\partial \theta} + R_n \tag{6.2.31}$$

ただし，$R_n \xrightarrow{P} 0$，つまり，標本数が大きくなるにつれて，$|R_n - 0| > \epsilon$ である確率は 0 となる．

証明は，(6.2.20) を並べ替え，定理 6.2.2 の証明を利用する ((6.2.20) – (6.2.22) あたりを)．
¡note¿

mle を求める際に，mle が存在することはわかっているが，$l'(\theta) = 0$ の解が解析的に求まらないケースはよくある．そのような場合には，数値的計算を行う．数値計算法としてはニュートン法が有名である．詳しくは，ここでは書かないが，

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \frac{l'(hat\theta^{(0)})}{l''(hat\theta^{(0)})} \tag{6.2.32}$$

のように，二階微分を利用した勾配法である．特に $hat\theta^{(1)}$ は *one-step estimator* と呼ばれており，初期値 $hat\theta^{(0)}$ が θ の一致推定量ならば，$hat\theta^{(1)}$ は，mle と同じ漸近分布を持つ．つまり，one-step estimator は θ の漸近有効推定量である．