

Chapter 6,

08.15.

Chapter 6.

Estimation

§ 6.1 MME

$m_k := E[X^k]$  kth moment.

$$\hat{m}_k := \bar{x}^k := \frac{1}{n} \sum_{i=1}^n x_i^k : k \geq 2$$

$$\Rightarrow \eta = g(m_1, \dots, m_k) \text{을 추정하는 통계량 } \hat{\eta} = \underline{g(\hat{m}_1, \dots, \hat{m}_k)}$$

Method of moment estimator

e.g. 6.1.1

$$X_1, \dots, X_n : f.s \quad \text{Var}(X_1) = \sigma^2$$

$$\sigma^2 = \text{Var}(X_1) = E[X_1^2] - E[X_1]^2 = m_2 - m_1^2$$

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

e.g. 6.1.2.

$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Poisson}(\lambda) \quad \lambda > 0$

$$E[X_1] = \lambda \quad \text{Var}(X_1) = \lambda$$

$$\hat{\lambda}^{\text{MME}} = \hat{m}_1 = \bar{x} \quad \text{or} \quad \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$\Rightarrow$  MME is not unique.

$$\textcircled{O} \quad \underbrace{M_{r,s} := E[X_1^r Y_1^s]}_{\text{joint-mgf}} \rightarrow \hat{M}_{r,s} := \sum_{i=1}^n X_i^r Y_i^s / n = \overline{X_1^r Y_1^s}$$

e.g. 6.1.3  $(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{\text{iid}}{\sim} ((\mu_1, \mu_2)', (\sigma_1^2, \sigma_2^2)')$

$$\rho = \frac{\text{Cov}(X_1, Y_1)}{\sqrt{\text{Var}(X_1)} \sqrt{\text{Var}(Y_1)}} = \frac{E[X_1 Y_1] - E[X_1]E[Y_1]}{\sqrt{E[X_1^2] - E[X_1]^2} \cdot \sqrt{E[Y_1^2] - E[Y_1]^2}}$$

$$\Rightarrow \hat{\rho}^{\text{MME}} = \frac{\bar{XY} - \bar{X}\bar{Y}}{\sqrt{\bar{X}^2 - (\bar{X})^2} \sqrt{\bar{Y}^2 - (\bar{Y})^2}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Thm 6.1.1.

$$\eta = g(M_1, \dots, M_k) \quad M_k \in \mathbb{R}.$$

$$\Rightarrow \underset{n \rightarrow \infty}{\text{plim}} \hat{\eta}_n^{\text{MME}} = \eta \quad (\text{consistency of } \hat{\eta}_n^{\text{MME}})$$

proof)

$$\hat{M}_i \xrightarrow[n \rightarrow \infty]{P} M_i \quad \forall i=1, \dots, k$$

$$\text{i.e., } (\hat{M}_1, \dots, \hat{M}_k)' \xrightarrow[n \rightarrow \infty]{P} (M_1, \dots, M_k)'$$

$$\text{So, } \hat{\eta}_n^{\text{MME}} = g(\hat{M}_1, \dots, \hat{M}_k) \xrightarrow[n \rightarrow \infty]{P} g(M_1, \dots, M_k) = \eta$$

추정량의 일관성

$$\hat{\eta} \xrightarrow[n \rightarrow \infty]{P} \eta(\theta) \quad \forall \theta \in \Omega$$

e.g. 6.1.4

$$\hat{\sigma}_n^2 \text{MME} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \hat{\sigma}_n^2 \text{MME}$$

We know that  $S_n^2 \xrightarrow[n \rightarrow \infty]{d} \sigma^2$

□

$$\text{So, } \hat{\sigma}_n^2 \text{MME} \xrightarrow[n \rightarrow \infty]{P} \sigma^2$$

Thm 6.1.2.

$$\eta = g(m_1, \dots, m_k), \quad m_1, \dots, m_k \in \mathbb{R}, \quad g \in C^1$$

$$\Rightarrow \sqrt{n}(\hat{\eta}^{\text{MME}} - \eta) \xrightarrow[n \rightarrow \infty]{d} Z_1 \sim N(0, \sigma^2)$$

$$\begin{aligned} \text{where, } \sigma^2 &= \dot{g}(m_1, \dots, m_k)^T \sum g(m_1, \dots, m_k) \\ \sum &= (m_{r+s} - m_r m_s) \end{aligned}$$

proof)

$$Y_i = (X_i, X_i^2, \dots, X_i^k) \quad E[Y_i] = (m_1, \dots, m_k)^T$$

$$\text{By CLT} \quad \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n Y_i - E[Y_i] \right) \xrightarrow[n \rightarrow \infty]{d} W, \quad W \sim N_k(0, \text{Var}(Y_i))$$

$$\text{Let } m := (m_1, \dots, m_k)$$

$$\text{By thm 5.3.3, } \sqrt{n} \left( g \left( \frac{1}{n} \sum_{i=1}^n Y_i - g(m) \right) \right) \xrightarrow[n \rightarrow \infty]{d} Z$$

$$\text{where } Z \stackrel{d}{=} \dot{g}(m)^T W$$

$$\text{Var}(Z) = \dot{g}(m)^T \text{Var}(W) \dot{g}(m), \quad \text{where } \text{Var}(W) = \text{Var}(Y_i)$$

$$(\text{Var}(Y_i))_{rs} = \text{Cov}(X_i^r, X_i^s) = E[X_i^r X_i^s] - E[X_i^r] E[X_i^s]$$

$$= m_{r+s} - m_r m_s$$

$$\therefore Z \sim N(0, \sigma^2) \quad \text{where } \sigma^2 = \dot{g}(m)^T \sum g(m) \quad (\sum)_{ij} = m_{i+j} - m_i m_j \quad \square$$

$$X_1, \dots, X_n \stackrel{iid}{\sim} f(x; \theta)$$

§ 6.2. MLE  
 (Maximum Likelihood Estimation)

$x = (x_1, \dots, x_n)'$ : observation.

$$p_{f_X}(x; \theta) | \Delta x | = \prod_{i=1}^n f(x_i; \theta) | \Delta x_1 | \dots | \Delta x_n |$$

↳ independent of  $\theta$ .

$$\text{i.e., } \max_{\theta \in \Omega} p_{f_X}(x; \theta) | \Delta x | = \max_{\theta \in \Omega} p_{f_X}(x; \theta)$$

$x$  is given.  $\Omega$  is a domain

$$\text{Def (a)} \quad \lambda(\theta; x) = \lambda(\theta) := \prod_{i=1}^n f(x_i; \theta)$$

Likelihood function.

$$\text{(b)} \quad \hat{\theta}_{(x)}^{\text{MLE}} := \underset{\theta \in \Omega}{\operatorname{argmax}} \lambda(\theta; x) \quad \text{"최대가능도 추정값"}$$

$$\hat{\theta}^{\text{MLE}}(x) : \text{최대가능도 추정값}$$

$$\text{(c)} \quad \ell(\theta; x) := \log \lambda(\theta; x) : \text{log likelihood function}$$

$$\text{Rmk } \hat{\theta}^{\text{MLE}} = \underset{\theta}{\operatorname{argmax}} \ell(\theta; x) = \underset{\theta}{\operatorname{argmax}} \lambda(\theta; x)$$

e.g. 6.2. (

$X_1, \dots, X_n \sim \text{Poisson}(\lambda) \quad \lambda \geq 0.$

$\hat{\lambda}^{\text{MLE}}$ ?

$$\text{sol)} \quad L(\lambda; x) = \prod_{i=1}^n \frac{e^\lambda \lambda^{x_i}}{x_i!} = e^{n\lambda} \lambda^{n\bar{x}} / (x_1! \cdots x_n!) \quad \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$$

$$i) \lambda > 0, \quad \ell(\lambda) = -n\lambda + n\bar{x} \log \lambda - \log(x_1! \cdots x_n!)$$

$$\dot{\ell}(\lambda) = -n + \frac{n\bar{x}}{\lambda} = \frac{n(\bar{x} - \lambda)}{\lambda} = 0 \Leftrightarrow \bar{x} = 0.$$

$$\ddot{\ell}(\lambda) = -\frac{n\bar{x}}{\lambda^2}$$

$\left. \begin{array}{l} \bar{x} > 0 \Rightarrow \ddot{\ell}(\lambda) < 0 \\ \Rightarrow \max_{\lambda > 0} \ell(\lambda) = \ell(\bar{x}) \end{array} \right\} \lambda > 0.$

$$\bar{x} = 0 \Rightarrow \ell(\lambda) = -n\lambda - \log(x_1! \cdots x_n!)$$

$$\sup_{\lambda > 0} \ell(\lambda) = \lim_{\lambda \rightarrow 0^+} \ell(\lambda)$$

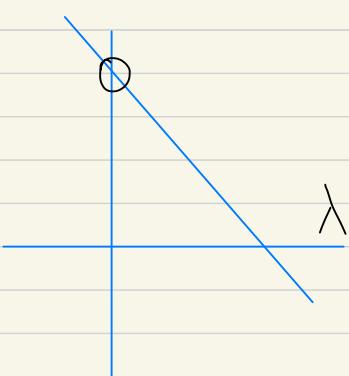
Since  $\ell(\lambda)$  is continuous &  $\lambda \geq 0$ ,  $\sup_{\lambda > 0} \ell(\lambda) = \lim_{\lambda \rightarrow 0^+} \ell(\lambda) = \ell(0) = \ell(\bar{x}) (\bar{x} = 0)$

$$\therefore \sup_{\lambda > 0} L(\lambda; x_1, \dots, x_n) = L(\bar{x}; x_1, \dots, x_n)$$

$$\therefore \max_{\lambda > 0} L(\lambda; x_1, \dots, x_n) = L(\bar{x}; x_1, \dots, x_n)$$

$$\therefore \hat{\lambda}^{\text{MLE}} = \bar{x}$$

□



Thm 6.2.1  $\mathcal{I}_0$ : open interval,  $\ell(\theta)$ : twice differentiable, and its second-order derivative is continuous

$$(\ddot{\ell}(\theta))$$

$$\ddot{\ell}(\theta) < 0, \forall \theta \in \mathcal{I}_0 \text{ and } \dot{\ell}(\hat{\theta}) = 0, \hat{\theta} \in \mathcal{I}_0$$

$$\Rightarrow \max_{\theta \in \mathcal{I}_0} \ell(\theta) = \ell(\hat{\theta})$$

proof)

By Taylor's Theorem,  $\ell(\theta) = \ell(\hat{\theta}) + \dot{\ell}(\hat{\theta})(\hat{\theta} - \theta) + \frac{1}{2}\ddot{\ell}(\theta^*)(\hat{\theta} - \theta)^2$ , for some  $\theta^* \in \mathcal{I}_0$

By the assumption that  $\dot{\ell}(\hat{\theta}) = 0$  &  $\ddot{\ell}(\theta) < 0 \quad \forall \theta \in \mathcal{I}_0$

$$\ell(\theta) = \ell(\hat{\theta}) + \frac{1}{2}\ddot{\ell}(\theta^*)(\hat{\theta} - \theta)^2 < \ell(\hat{\theta}) \quad \forall \theta \in \mathcal{I}_0 \setminus \{\hat{\theta}\}$$

$$\therefore \max_{\theta \in \mathcal{I}_0} \ell(\theta) = \ell(\hat{\theta})$$

□

e.g. 6.2.2.

$X_1, \dots, X_n$  iid Bernoulli( $p$ )  $\hat{p}$  MLE?

$$L(p; X_1, \dots, X_n) = \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i}$$

$$\ell(p) = \sum_{i=1}^n X_i \log p + (1-X_i) \log(1-p) \quad 0 < p < 1.$$

$$\dot{\ell}(p) = \frac{1}{p} \sum_{i=1}^n X_i - \frac{1}{1-p} \sum_{i=1}^n (1-X_i) \quad \hat{p} := \sum_{i=1}^n X_i / n$$

$$= \frac{n\hat{p}}{p} - \frac{n(1-\hat{p})}{1-p} = 0 \Leftrightarrow \hat{p}(1-p) = p(1-\hat{p})$$

$$\ddot{\ell}(p) = -\frac{n\hat{p}}{p^2} - \frac{n(1-\hat{p})}{(1-p)^2} \quad p = \hat{p}$$

$$\text{i)} 0 < \hat{p} < 1, \quad \ddot{\ell}(p) < 0 \quad \forall p: 0 < p < 1 \quad \max_{0 < p < 1} \ell(p) = \ell(\hat{p})$$

$$\text{ii)} \hat{p} = 0, \quad \ell(p) = n \log(1-p) \quad \Rightarrow \sup_{0 < p < 1} \ell(p) = \lim_{p \rightarrow 0^+} \ell(p)$$

$$\text{iii)} \hat{p} = 1, \quad \ell(p) = n \log p \quad \Rightarrow \sup_{0 < p < 1} \ell(p) = \lim_{p \rightarrow 1^-} \ell(p)$$

Since  $L$  continuous &  $0 \leq p \leq 1$ ,

$$\sup_{0 \leq p \leq 1} L(p; X_1, \dots, X_n) = L(\hat{p}; X_1, \dots, X_n)$$

$$\therefore \max_{0 \leq p \leq 1} L(p; X_1, \dots, X_n) = L(\hat{p}; X_1, \dots, X_n)$$

$$\therefore \hat{p}^{\text{MLE}} = \bar{X}$$

□

Thm. 6.2.2.

$\mathcal{I}_0$ : open interval,  $\ell(\theta)$ : twice differentiable  
 $\ell(\theta)$ : continuous.

$$\ddot{\ell}(\theta) < 0 \quad \forall \theta \in \mathcal{I}_0, \lim_{\theta \rightarrow \partial(\mathcal{I}_0)} \ell(\theta) = -\infty$$

$$\Rightarrow \exists! \hat{\theta} \in \mathcal{I}_0 \text{ s.t. } \dot{\ell}(\hat{\theta}) = 0$$

proof)

$$\mathcal{I}_0 = (-\infty, +\infty)$$

Since

$$\lim_{\theta \rightarrow -\infty} \ell(\theta) = -\infty, \lim_{\theta \rightarrow \infty} \ell(\theta) = -\infty \quad \text{and } \ell(\theta) \text{: continuous,}$$

$$\ell(\hat{\theta}) = \max_{\theta \in \mathcal{I}_0} \ell(\theta), \quad \hat{\theta} \in \mathcal{I}_0. \quad (*)$$

$$\text{If } \hat{\theta}_1 \neq \hat{\theta}_2 \text{ and } \hat{\theta}_1 = \arg \max_{\theta \in \mathcal{I}_0} \ell(\theta), \quad \hat{\theta}_2 = \arg \max_{\theta \in \mathcal{I}_0} \ell(\theta)$$

Since  $\ell(\theta)$ : concave function,

$$\ell\left(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2\right) > \frac{1}{2}\ell(\hat{\theta}_1) + \frac{1}{2}\ell(\hat{\theta}_2) = \max_{\theta \in \mathcal{I}_0} \ell(\theta)$$

It contradicts to that  $\ell\left(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2\right) > \ell(\hat{\theta}_1) = \max_{\theta \in \mathcal{I}_0} \ell(\theta)$

Thus,  $\hat{\theta}_1 = \hat{\theta}_2$

Since  $\dot{\ell}(\hat{\theta}) = 0$ ,  $\hat{\theta}$  is a unique solution of  $\arg \max_{\theta \in \mathcal{I}_0} \ell(\theta)$   $\square$

e.J. 6.2.3

$X_1, \dots, X_n$  iid  $L(\theta, 1)$  ( $-\infty < \theta < +\infty$ )

$$p.d.f_{X_1}(x_1|\theta) = \frac{e^{-x_1\theta}}{(1+e^{x_1\theta})^2} = \frac{e^{-x_1\theta}}{\left(\frac{e^{x_1\theta}}{e^{x_1\theta}+1}\right)^2} (e^{x_1\theta})^2 = \frac{e^{-x_1\theta}}{\left(\frac{e^{x_1\theta}}{e^{x_1\theta}+1}\right)^2}$$

$-\infty < x_1 < +\infty$ .

$$\begin{aligned} l(\theta) &= \sum_{i=1}^n (-x_i + \theta) - 2 \log(1 + e^{-x_i + \theta}) \\ &= -n\bar{x} + n\theta - 2 \sum_{i=1}^n \log(1 + e^{-x_i + \theta}) \end{aligned}$$

$$l'(\theta) = n - 2 \sum_{i=1}^n e^{-x_i + \theta} / (1 + e^{-x_i + \theta})$$

$$\begin{aligned} l''(\theta) &= 2 \sum_{i=1}^n \frac{e^{-x_i + \theta} (1 + e^{-x_i + \theta}) - e^{-x_i + \theta} \cdot e^{-x_i + \theta}}{(1 + e^{-x_i + \theta})^2} \\ &= -2 \sum_{i=1}^n \frac{e^{-x_i + \theta}}{(1 + e^{-x_i + \theta})^2} \end{aligned}$$

$$\therefore l'(\theta) < 0 \quad \forall \theta \in (-\infty, \infty)$$

$$\lim_{\theta \rightarrow -\infty} l(\theta) = -\infty \quad \lim_{\theta \rightarrow \infty} l(\theta) = -\infty$$

↑

$$\therefore \lim_{\theta \rightarrow \infty} -\log(1 + e^{-x_i + \theta}) < \lim_{\theta \rightarrow \infty} -\log(e^{-x_i + \theta}) = \lim_{\theta \rightarrow \infty} (x_i - \theta) = -\infty$$

$$\therefore l(\hat{\theta}) = n - 2 \sum_{i=1}^n \frac{e^{-x_i + \hat{\theta}}}{(1 + e^{-x_i + \hat{\theta}})^2} = 0$$

□

Thm 6.2.3.

$\exists \hat{\theta}^{\text{MLE}} \in \Omega$  and  $g$ : invertible.  $\eta = g(\theta)$

$$\Rightarrow \eta^{\text{MLE}} = g(\hat{\theta}^{\text{MLE}})$$

$$p\text{df}_X(x_1, \dots, x_n; \theta) = p\text{df}_X(x_1, \dots, x_n; g^{-1}(\eta)) \quad \eta \in g(\Omega)$$

$$\begin{aligned} L(g^{-1}(\eta); x_1, \dots, x_n) &= p\text{df}_X(x_1, \dots, x_n; g^{-1}(\eta)) \\ &= p\text{df}_X(x_1, \dots, x_n; \theta) \end{aligned}$$

$$\max_{\eta \in g(\Omega)} L(g^{-1}(\eta); x_1, \dots, x_n) = \max_{\theta \in \Omega} L(\theta; x_1, \dots, x_n) = L(\hat{\theta}^{\text{MLE}}) \\ = L(g^{-1}(g(\hat{\theta}^{\text{MLE}})))$$

$$\therefore \eta^{\text{MLE}} = g(\hat{\theta}^{\text{MLE}})$$

□

e.g. 6.2.4.

$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\theta) \quad (0 < \theta < +\infty)$

$$p\text{df}(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad I(x > 0)$$

$$\ell(\theta) = -n \log \theta - n \bar{x}/\theta \quad \dot{\ell}(\theta) = -n/\theta + n \bar{x}/\theta^2$$

$$\ddot{\ell}(\theta) = n/\theta^2 - 2n\bar{x}/\theta^3$$

$\ell(\theta)$  is not concave function.

$$\lambda = 1/\theta$$

$$\ell(\lambda) = n \log \lambda - n \bar{x} \lambda \quad \dot{\ell}(\lambda) = n/\lambda - n \bar{x} \quad \ddot{\ell}(\lambda) = -n/\lambda^2$$

$$\lim_{\lambda \rightarrow 0^+} \ell(\lambda) = -\infty \quad \lim_{\lambda \rightarrow \infty} \ell(\lambda) = -\infty$$

$$\dot{\ell}(\lambda) = 0 \Leftrightarrow \lambda = 1/\bar{x}$$

$$\therefore \hat{\lambda}^{\text{MLE}} = 1/\bar{x}$$

$$\therefore \hat{\theta}^{\text{MLE}} = \bar{x}$$

□

# STATISTICS 200822

Thm. 6.2.4.

Exponential family

FACT

①  $\text{supp}(f)$  is open set  
open interval or arbitrary union of ch.

②  $\exists \epsilon > 0$  s.t.  
 $x \in (x-\epsilon, x+\epsilon) \subset \text{supp}(f)$

proof)

Q.  $s+\eta \in N$ ?

$\eta$   $I_\eta$

$\exists \epsilon > 0$  s.t.  
 $(\eta-\epsilon, \eta+\epsilon) \subset I_\eta$   
 $(\because N: \text{open set})$

$|s| < \epsilon$ , i.e.,  $-\epsilon < s < \epsilon$   
 $\Rightarrow \eta - \epsilon < s + \eta < \eta + \epsilon$   
i.e.,  $s + \eta \in (\eta - \epsilon, \eta + \epsilon) \subset I_\eta \subset N$

$$p_{\text{pdf}}(x; \eta) = \exp \{ \eta T(x) - A(\eta) + S(x) \} \quad x \in \mathcal{X}, \eta \in N$$

$X_1, \dots, X_n$ : r.s.

(i)  $\mathcal{X} = \{x : p_{\text{pdf}}(x; \eta) > 0\}$ : 모두에 따른 x

(ii)  $N$ : open in  $\mathbb{R}$

(iii)  $\text{Var}(T(X_1)) > 0$  i.e.  $T(X_1)$  is not constant.

$\Rightarrow$  If  $A(\eta) = \overline{T(x)}$  or  $E_\eta[T(X_1)] = \overline{T(x)}$ : equation w.r.t  $\eta$

has a solution  $\hat{\eta} = \hat{\eta}(x_1, \dots, x_n)$  then  
 $\hat{\eta} = \hat{\eta}^{\text{MLE}}$

$$E[\exp(sT(X_1))] = \int_{-\infty}^{+\infty} \exp(sT(x)) \exp(\eta T(x) - A(\eta) + S(x)) dx$$

$$\stackrel{(5)}{=} \int_{-\infty}^{+\infty} \exp((s+\eta)T(x) - A(s+\eta) + S(x)) dx$$

$$\exp(A(s+\eta) - A(\eta))$$

$\because$  Since  $s+\eta \in N$ ,

$$\exp(s+\eta) T(x) - A(s+\eta) + S(x) : \text{pdf.}$$

$s+\eta \in N, \eta \in N, |s| < \epsilon, \exists \epsilon > 0$ .

The integral over  $(-\infty, +\infty)$

$$c.g.f. T(X_1); \eta(s) = \log \text{pdf}_{T(X_1); \eta}(s) = A(s+\eta) - A(\eta)$$

$$E_\eta[T(X_1)] = \frac{d}{ds} \Big|_{s=0} A(s+\eta) - A(\eta) = \dot{A}(\eta)$$

$$\frac{d^2}{ds^2} \Big|_{s=0} A(s+\eta) - A(\eta) = \ddot{A}(\eta)$$

$$\bar{\ell}(\theta) = \frac{1}{n} \ell(\theta)$$

$$\bar{\ell}(\eta) = \eta \bar{T}(x) - A(\eta) + S(x)$$

$$\bar{\ell}(\eta) = \bar{T}(x) - A(\eta)$$

$$\bar{\ell}'(\eta) = -\bar{A}'(\eta) = -\text{Var}_{\eta}(T(X_1)) < 0 \Rightarrow \bar{\ell}(\eta) \text{ is concave.}$$

$\Rightarrow \hat{\eta}^{\text{MLE}}$ : solution of  $\bar{A}'(\eta) = \bar{T}(x)$  (or  $E_{\eta}[T(X_1)] = \bar{T}(x)$ )

$$\Rightarrow \hat{\eta}^{\text{MLE}} = \hat{\eta}$$

□

Rmk.  $\eta = g(\theta)$ ,  $\theta \in \mathcal{S}$ ,  $g$ : 1-1 corr.

$$p(x|\theta) = \exp \{ g(\theta) T(x) - A(g(\theta)) + S(x) \}$$

By thm 6.2.3  $g(\hat{\theta}^{\text{MLE}}) = \hat{\eta}^{\text{MLE}}$ .

$E_{\theta}[T(X_1)] = E_{\eta}[T(X_1)] = \bar{T}(x)$  : equation of likelihood.

$\hookrightarrow \theta, \eta$ 와 상관없이 확률은 같은데 MLE, MGF는 같다.

e.J. 6.2.5.

(a) Bernoulli( $p$ )  $0 < p < 1$

$$f(x; p) = p^x (1-p)^{1-x}$$

$$= \exp(x \log p + (1-x) \log(1-p))$$

$$= \exp(x \log \frac{p}{1-p} + \log(1-p))$$

$$= \exp(x \log \frac{p}{1-p} - (-\log(1-p)))$$

$$\eta := \log \frac{p}{1-p}$$

$$e^{\eta} = p/(1-p)$$

$$e^{\eta}(1-p) = p$$

$$e^{\eta} = p(1+e^{\eta})$$

$$p = e^{\eta} / (1+e^{\eta})$$

$$\text{Since } \hat{\eta}^{\text{MLE}} = \log \frac{p}{1-p}^{\text{MLE}},$$

$$= \exp(\eta x + \log(\frac{1}{1+e^{\eta}}))$$

$$= \exp(\eta x - \log(1+e^{\eta}))$$

$$\frac{e^{\eta}}{1+e^{\eta}} = \bar{x}$$

$$e^{\eta} = (1+e^{\eta})\bar{x}$$

$$e^{\eta}(1-\bar{x}) = \bar{x}$$

$$\eta = \log \frac{\bar{x}}{1-\bar{x}}$$

$$\hat{p}^{\text{MLE}} = \bar{x}$$

□

(b)

Poisson( $\lambda$ )  $\lambda > 0$ 

$$f(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \exp(-\lambda + x \log \lambda - \log x!)$$

$$= \exp(x \log \lambda - \lambda - \log x!) \quad \eta := \log \lambda$$

$$= \exp(\eta x - e^\eta - \log x!)$$

$$e^\eta = \bar{x} \quad \eta = \log \bar{x}$$

$$\therefore \hat{\lambda}^{\text{MLE}} = \bar{x}$$

(c) Geo( $p$ )  $0 < p < 1$ 

$$f(x|p) = (1-p)^{x-1} p$$

$$= \exp((x-1) \log(1-p) + \log p)$$

$$= \exp(\log(1-p) \cdot x + \log \frac{p}{1-p})$$

$$\frac{1-e^\eta}{e^\eta}$$

$$\eta := \log(1-p) \quad = \exp(\eta x + \log(1-e^\eta) - \eta)$$

$$e^\eta = 1-p$$

$$= \exp(\eta x - (\eta - \log(1-e^\eta)))$$

$$p = 1 - e^\eta$$

$$1 + \frac{e^\eta}{1-e^\eta} = \bar{x}$$

$$\frac{1}{1-e^\eta} = \bar{x} \quad 1 - e^\eta = 1/\bar{x}$$

$$e^\eta = 1 - 1/\bar{x}$$

$$\therefore \hat{p}^{\text{MLE}} = 1/\bar{x} \quad \Omega \quad M = \log(1 - 1/\bar{x})$$

(d)  $\text{Exp}(\theta) \quad \theta > 0$

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0$$

$$= \exp(-\log \theta - \frac{x}{\theta})$$

$$\eta := -\frac{1}{\theta} = \exp(\eta \lambda - (-\log(-\eta)))$$

$$-\frac{1}{\eta} = \bar{x}$$

$$\therefore \hat{\theta}_{MLE} = \bar{x}$$

D

(e)  $\text{Pareto}(1, \theta) \quad \theta > 0$

$$f(x; \theta) = \theta x^{-\theta-1} \quad x > 1$$

$$= \exp(\log \theta - (\theta + 1) \log x)$$

$$= \exp(-\theta \log x - (-\log \theta) - \log x)$$

$$= \exp(\theta(-\log x) - (-\log \theta) - \log x)$$

$$\widehat{f(\log x)} = f \left( \frac{1}{\theta} \right)$$

$$\hat{\theta}_{MLE} = 1 / \widehat{f(\log x)}$$

D

e.g. 6.2.6.

DE( $\theta$ , 1) ( $-\infty < \theta < +\infty$ )

$X_1, \dots, X_n$ : r. s. ( $n=2m+1$ )

$$p.d.f(x; \theta) = \frac{1}{2} e^{-|x-\theta|} \quad -\infty < x < +\infty$$

$$\ell(\theta) = -\sum_{i=1}^n |X_i - \theta| - n \log 2$$

$$X_{(1)} < \dots < X_{(n)} \quad X_{(0)} := -\infty \quad X_{(n+1)} := +\infty$$

$$X_{(r)} \leq \theta < X_{(r+1)} \quad (r=0, \dots, n)$$

$$\ell(\theta) = -\sum_{i=1}^n |X_i - \theta| - n \log 2$$

$$= -\sum_{i=1}^r (\theta - X_{(i)}) - \sum_{i=r+1}^n (X_{(i)} - \theta) - n \log 2$$

$$\therefore \ell(\theta) = (n-2r)\theta + \sum_{i=1}^r X_{(i)} - \sum_{i=r+1}^n X_{(i)} - n \log 2$$

$$r < \frac{n}{2}$$

i)  $(n-2r) > 0$ , i.e.,  $r=0, \dots, m$   $\ell(\theta)$ : increasing  $\theta \in [X_{(r)}, X_{(r+1)}]$

ii)  $(n-2r) < 0$ , i.e.,  $r=m+1, \dots, n$   $\ell(\theta)$ : decreasing  $\theta \in [X_{(r)}, X_{(r+1)}]$

$$\therefore \hat{\theta}^{MLE} = X_{(m+1)}$$

□

E.g. 6.2.7

$$[0, \theta] \quad (0 < \theta < +\infty)$$

$X_1, \dots, X_n$ : r.s.  $\hat{\theta}^{\text{MLE}}$ ?

$$p(\mathbf{x}_1, \dots, \mathbf{x}_n; \theta) = \prod_{i=1}^n \frac{1}{\theta} I(0 \leq x_i \leq \theta)$$

$$= \theta^{-n} I(0 \leq \min_{1 \leq i \leq n} x_i, \max_{1 \leq i \leq n} x_i \leq \theta)$$

$$L(\theta) = \theta^{-n} I(\theta \geq \max_{1 \leq i \leq n} x_i)$$

$$\therefore \hat{\theta}^{\text{MLE}} = X_{(n)}$$

D

E.g. 6.2.8.

$$[\theta-1, \theta+1] \quad (-\infty < \theta < +\infty)$$

$X_1, \dots, X_n$ : r.s.

$$p(\mathbf{x}_1, \dots, \mathbf{x}_n; \theta) = \prod_{i=1}^n \frac{1}{2} I(\theta-1 \leq x_i \leq \theta+1)$$

$$= 2^{-n} I(\theta-1 \leq \min_{1 \leq i \leq n} x_i, \max_{1 \leq i \leq n} x_i \leq \theta+1)$$

$$L(\theta) = 2^{-n} I(\max_{1 \leq i \leq n} x_i - 1 \leq \theta \leq \min_{1 \leq i \leq n} x_i + 1)$$

$$\hat{\theta} := \arg \max_{\theta} L(\theta) \quad \hat{\theta} \in [x_{(n)} - 1, x_{(1)} + 1]$$

$\therefore \hat{\theta}$  is not unique.

$$\therefore X_{(n)} - 1 \leq \hat{\theta}^{\text{MLE}} \leq X_{(1)} + 1$$

D

## 6.3. MLE for multiple parameters.

$X_1, \dots, X_n \stackrel{iid}{\sim} f(x; \theta)$ ,  $\theta = (\theta_1, \dots, \theta_k)^\top \in \mathcal{S}$

Given observation  $(x_1, \dots, x_n)$

likelihood and log likelihood are defined as follows

$$(a) L(\theta; x) = \prod_{i=1}^n f(x_i; \theta), \theta = (\theta_1, \dots, \theta_k) \in \mathcal{S}$$

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

$$(b) L(\hat{\theta}^{MLE}(x); x) := \max_{\theta \in \mathcal{S}} L(\theta; x), \hat{\theta}^{MLE} \in \mathcal{S}$$

$$(c) \hat{\theta}^{MLE} = (\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_k^{MLE})$$

e.g. 6.3.1

$X_1, \dots, X_n \sim D \in \mathcal{D}(\mu, \sigma) (-\infty < \mu < +\infty, \sigma > 0)$

( $n=2m+1$ )  $m \in \mathbb{N}$   $\theta := (\mu, \sigma)$

$\hat{\theta}$  MLE?

Sol)  $p_{\mathcal{D}}(x; \theta) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, -\infty < x < +\infty$

$$\ell(\mu, \sigma) = -\sum_{i=1}^n |x_i - \mu|/\sigma - n \log \sigma - n \log 2$$

By e.g. 6.2.6,  $\max_{\mu} \ell(\mu, \sigma) = \ell(\hat{\mu}, \sigma)$

$$\hat{\mu} = x_{(m+1)} \text{ where } x_{(1)} < \dots < x_{(n)}$$

$$\frac{\partial}{\partial \sigma} \ell(\hat{\mu}, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^n |x_i - \hat{\mu}| - \frac{n}{\sigma} > 0 \Leftrightarrow \sigma < \frac{1}{n} \sum_{i=1}^n |x_i - \hat{\mu}|$$

$$\therefore \max_{\sigma} \ell(\hat{\mu}, \sigma) = \max_{\sigma} \ell(\hat{\mu}, \hat{\sigma}) \quad \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{\mu}|$$

$$\therefore \max_{\sigma} \max_{\mu} \ell(\mu, \sigma) = \ell(\hat{\mu}, \hat{\sigma})$$

$$\therefore \hat{\theta}^{\text{MLE}} = \left( x_{(m+1)}, \frac{1}{n} \sum_{i=1}^n |x_i - x_{(m+1)}| \right) \quad \square$$

e.g. 6.3.2

$X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$

$$\Theta := (\theta_1, \theta_2) := (\mu, \sigma^2)$$

$$p(x|\theta) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\ell(\mu, \sigma) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2$$

$$\begin{aligned} \text{Since } \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \end{aligned}$$

$$\max_{\mu} \ell(\mu, \sigma) = \ell(\bar{x}, \sigma) \quad \bar{x} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

$$\frac{d}{d\sigma^2} \ell(\bar{x}, \sigma) = \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} > 0$$

$$\Leftrightarrow \sigma^2 < \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\therefore \max_{\sigma^2} \ell(\bar{x}, \sigma) = \ell(\bar{x}, \hat{\sigma}^2) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\therefore \hat{\theta}^{\text{MLE}} = \left( \frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)'$$

□

Thm 6.3.1

$$\exists \hat{\theta}^{\text{MLE}} \in \mathcal{S} \quad \& \quad \eta = (\eta_1, \eta_2) = (g_1(\theta), g_2(\theta)) \\ = g(\theta)$$

$g$ : bijective

$$\Rightarrow \hat{\eta}_1^{\text{MLE}} = g_1(\hat{\theta}^{\text{MLE}})$$

Proof)

By Thm 6.2.3

$$\hat{\eta}^{\text{MLE}} = g(\hat{\theta}^{\text{MLE}}) = (g_1(\hat{\theta}^{\text{MLE}}), g_2(\hat{\theta}^{\text{MLE}}))$$

By the definition of MLE,

$$\hat{\eta}_1^{\text{MLE}} = g_1(\hat{\theta}^{\text{MLE}})$$

□

e.g. 6.3.3.

(a)

$$N(\mu, \sigma^2)$$

$$\theta := (\mu, \sigma^2)$$

$$\text{sgn}(\mu) = \begin{cases} 1 & \mu > 0 \\ 0 & \mu = 0 \\ -1 & \mu < 0 \end{cases}$$

$$\begin{aligned}\eta &= (g_1(\theta), g_2(\theta), g_3(\theta)) \\ &= (|\mu|, \text{sgn}(\mu), \sigma^2)\end{aligned}$$

Since  $\eta$ : bijective,

$$\eta^{\text{MLE}} = (|\hat{\mu}^{\text{MLE}}|, \text{sgn}(\hat{\mu}^{\text{MLE}}), \hat{\sigma}^2)^{\text{MLE}}$$

$$\therefore \eta_1^{\text{MLE}} = |\hat{\mu}^{\text{MLE}}| = \left| \frac{1}{n} \sum_{i=1}^n x_i \right| \quad D$$

(b)

$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p) \quad 0 \leq p \leq 1$

$$\sigma^2 = p(1-p)$$

$$\eta = (g_1(p), g_2(p))'$$

$$= (p(1-p), \operatorname{sgn}(p-1/2))'$$

Since  $g$ : bijective,

$$\hat{\eta}_1^{\text{MLE}} = \hat{p}^{\text{MLE}}(1 - \hat{p}^{\text{MLE}}) \quad \hat{p}^{\text{MLE}} = \bar{X} \quad \square$$

Thm 6.32 open interval hyper cube  $(a_1, b_1) \times \dots \times (a_k, b_k) = \mathcal{I}_0$

$\ell(\theta)$ : differentiable twice,  $\ell(\theta) \in C^2$ .

$c^T \dot{\ell}(\theta) c < 0 \quad \forall c \neq 0, \quad \theta \in \mathcal{I}_0 \quad \&$

$\dot{\ell}(\hat{\theta}) = 0, \quad \hat{\theta} \in \mathcal{I}_0$

$\Rightarrow \max_{\theta \in \mathcal{I}} \ell(\theta) = \ell(\hat{\theta})$

proof)

$$\begin{aligned} \ell(\theta) &= \ell(\hat{\theta}) + \dot{\ell}(\hat{\theta}) \cdot (\theta - \hat{\theta}) \\ &\quad + \frac{1}{2} (\theta - \hat{\theta})^T \ddot{\ell}(\theta^*) (\theta - \hat{\theta})^T < \ell(\hat{\theta}) \end{aligned}$$

$\therefore \dot{\ell}(\hat{\theta}) = 0 \quad \& \quad \ddot{\ell}(\theta^*)$ : negative definite.

□

Thm 6.3.3.

open interval hypercube  $(a_1, b_1) \times \cdots \times (a_k, b_k) = \mathcal{I}_0$

$\ell(\theta)$ : differentiable twice,  $\ell(\theta) \in C^2$

$$c^T \dot{\ell}(\theta) < 0 \quad \forall c \neq 0, \quad \forall \theta \in \mathcal{I}_0 \text{ & } \lim_{\theta \rightarrow \partial(\mathcal{I}_0)} \ell(\theta) = -\infty$$

$$\Rightarrow \exists! \hat{\theta} \text{ s.t } \dot{\ell}(\hat{\theta}) = 0 \quad \hat{\theta} \in \mathcal{I}_0.$$

Since  $\lim_{\theta \rightarrow \partial(\mathcal{I}_0)} \ell(\theta) = -\infty$ ,  $\exists \hat{\theta} \text{ s.t}$   
 $\hat{\theta} \in \mathcal{I}_0$

$$\ell(\hat{\theta}) = \max_{\theta} \ell(\theta). \quad \hat{\theta} \in \mathcal{I}_0$$

Suppose  $\hat{\theta}_1 < \hat{\theta}_2$  &  $\max_{\theta \in \mathcal{I}_0} \ell(\theta) = \ell(\hat{\theta}_1) = \ell(\hat{\theta}_2)$

Since  $\ell(\theta)$ : concave.

$$\ell\left(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2\right) > \frac{1}{2}\ell(\hat{\theta}_1) + \frac{1}{2}\ell(\hat{\theta}_2)$$

$$= \max_{\theta \in \mathcal{I}_0} \ell(\theta)$$

$$\therefore \hat{\theta}_1 = \hat{\theta}_2$$

$$\therefore \exists! \hat{\theta} \text{ s.t } \dot{\ell}(\hat{\theta}) = 0.$$

$$p\text{df}(x; \eta) = \exp \left\{ \sum_{j=1}^k \eta_j T_j(x) - A(\eta) + S(x) \right\} \quad x \in \mathcal{X},$$

$$\eta = (\eta_1, \dots, \eta_k) \in \mathcal{V}$$

Thm 6.3.4

$$(i) \mathcal{X} = \{x : p\text{df}(x; \eta) > 0\}$$

(ii)  $\mathcal{V} = (a_1, b_1) \times \dots \times (a_k, b_k)$  open interval hyper-cube

$$(iii) \text{Var}(c' T(X) c) > 0 \quad \forall c \in \mathbb{R}^k, c \neq 0$$

$\Rightarrow$  If  $A(\eta) = \overline{T(x)}$  or  $E_\eta[T(X)] = \overline{T(x)}$  has a solution  $\hat{\eta} = \hat{\eta}(x_1, \dots, x_n)$ , then  $\hat{\eta} = \hat{\eta}^{\text{MLE}}$ .

proof)

$$Mgf_{T(X); \eta}(s) = E[\exp(\sum_{j=1}^k s_j T_j(x))]$$

$$T(X) = (T_1(x_1), \dots, T_k(x_1)) \quad s = (s_1, \dots, s_k)$$

Since  $X_i \stackrel{iid}{\sim} f(x; \eta)$ ,  $\prod_i T_i(x_i)$  and  $T_i(x_i) = T_i(x_1)$   $i=1, \dots, k$ .

$$\text{Thus, } Mgf_{T(X); \eta}(s) = \prod_{j=1}^k Mgf_{T_j(x_1); \eta}(s_j)$$

$$= \prod_{j=1}^k Mgf_{T_j(x_1); \eta}(s_j)$$

$$= \prod_{i=1}^k \int_{-\infty}^{+\infty} \exp(s_i T_i(x)) \exp\left(\sum_{j=1}^k \eta_j T_j(x) - A(\eta) + S(x)\right) dx$$

$$= \left\{ \int_{-\infty}^{+\infty} \exp\left(\sum_{j=1}^k \eta_j T_j(x) - A(\eta) + S(x)\right) dx \right\}$$

$$\cdot \int_{-\infty}^{+\infty} \exp\left(\sum_{j=1}^k (s_j + \eta_j) T_j(x) - A(s + \eta) + S(x)\right) dx$$

$$\cdot \exp(A(s + \eta) + S(x)) dx$$

$$= \exp(A(s + \eta) + S(x))$$

$$\therefore s + \eta \in \mathcal{V} \quad \& \quad \int_{-\infty}^{+\infty} p\text{df}(x; \eta) dx = 1$$

$$cgf_{T(X_1); \eta}(s) = A(s+\eta) - A(\eta)$$

$$\begin{aligned} E_\eta[T(X_1)] &= \left( \frac{\partial}{\partial s_1} \Big|_{s=0} A(s+\eta) - A(\eta), \dots, \frac{\partial}{\partial s_k} \Big|_{s=0} A(s+\eta) - A(\eta) \right) \\ &= \dot{A}(\eta) \end{aligned}$$

$$\text{Var}_\eta(T(X_1)) = \ddot{A}(\eta)$$

$$\bar{\ell}(\eta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k \eta_j T_j(x_i) - A(\eta) + \frac{1}{n} \sum_{i=1}^n S(x_i)$$

$$\begin{aligned} \bar{\ell}'(\eta) &= (\bar{T}_1(x), \dots, \bar{T}_k(x))' - \dot{A}(\eta) \\ &= \bar{T}(x) - \dot{A}(\eta) \end{aligned}$$

$$\bar{\ell}''(\eta) = -\ddot{A}(\eta) : \text{negative definite.}$$

$\Rightarrow \hat{\eta}^{\text{MLE}}$ : solution of  $\dot{A}(\eta) = \bar{T}(x)$  or  $E_\eta[T(X_1)] = \bar{T}(x)$

$$\Rightarrow \hat{\eta}^{\text{MLE}} = \hat{\eta}$$

□

Rmk.  $\eta = g(\theta) = (g_1(\theta), \dots, g_k(\theta))$  &  $g$ : bij.

$$pH(x; \theta) = \exp \left( \sum_{j=1}^k g_j(\theta) T_j(x) - A(g(\theta)) + S(x) \right)$$

$$E_\theta[T(X_1)] = \bar{T}(x)$$

$$f(x; \eta) = \exp \left\{ \sum_{j=1}^k \eta_j T_j(x) - A(\eta) + S(x) \right\}$$

e.g. 6.3.4.

	$f(x; \theta) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right\} \quad -\infty < x < +\infty$ $\theta = (\mu, \sigma^2)$ $= \exp \left\{ -\frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right\}$ $\eta_2 := \frac{-1}{2\sigma^2} \quad \eta_1 = \frac{\mu}{\sigma^2} \quad \frac{\mu^2}{2\sigma^2} = \frac{\eta_1^2}{4\eta_2^2} \times (-\eta_2)$ $\sigma^2 = -\frac{1}{2\eta_2} \quad \mu = \sigma^2 \eta_1 = -\frac{\eta_1}{2\eta_2} = -\frac{\eta_1^2}{4\eta_2}$ $= \exp \left\{ \eta_1 x + \eta_2 x^2 + \frac{\eta_1^2}{4\eta_2} - \frac{1}{2} \log \left( -\frac{\pi}{\eta_2} \right) \right\}$ $A(\eta) := -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2} \log \left( -\frac{\eta_2}{\pi} \right)$ $\hat{A}(\eta) = (\bar{x}, \bar{x}^2)$ $= \left( -\frac{1}{2} \times \frac{\eta_1}{\eta_2}, +\frac{\eta_1^2}{4\eta_2^2} + \frac{1}{2} \times \frac{\pi}{\eta_2} \times \left( -\frac{1}{\pi} \right) \right)$ $\bar{x} = -\frac{1}{2} \times (-2\hat{\mu}) = \hat{\mu}$ $\bar{x}^2 = (\hat{\mu})^2 - \frac{1}{2} \times (-2\hat{\sigma}^2) = \hat{\sigma}^2 + (\hat{\mu})^2$ $\therefore \hat{\mu}^{\text{MLE}} = \bar{x}, \hat{\sigma}^2 = \bar{x}^2 - \bar{x}$ <span style="float: right;">□</span>
--	---

$$f(x; \theta) = \det(2\pi \Sigma)^{-1/2} \exp(-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu))$$

$$E[X_1] = \mu_1, \quad E[X_2] = \mu_2$$

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)} \sqrt{\text{Var}(X_2)}} \quad \text{Cov}(X_1, X_2) = \sigma_1 \sigma_2 \rho$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \quad \sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix}$$

$$f(x, y; \theta) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-Q}$$

$$Q = \frac{\sigma_1^2 (x - \mu_1)^2}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} - \frac{2\rho \sigma_1 \sigma_2 (x - \mu_1)(y - \mu_2)}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} + \frac{\sigma_2^2 (y - \mu_2)^2}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}$$

$$= \frac{1}{1 - \rho^2} \left\{ \frac{(x - \mu_1)^2}{\sigma_2^2} - 2\rho \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1 \sigma_2} + \frac{(y - \mu_2)^2}{\sigma_1^2} \right\}$$

$$+ \frac{1}{1 - \rho^2} \left\{ \frac{x^2 - 2x\mu_1 + \mu_1^2}{\sigma_2^2} - 2\rho \frac{(xy - y\mu_1 - x\mu_2 + \mu_1\mu_2)}{\sigma_1 \sigma_2} + \frac{y^2 - 2y\mu_2 + \mu_2^2}{\sigma_1^2} \right\}$$

$$= \frac{-2\mu_1 x}{(1 - \rho^2)\sigma_2^2} + \frac{2\rho\mu_2 x}{(1 - \rho^2)\sigma_1\sigma_2} + \frac{-2\mu_2 y}{(1 - \rho^2)\sigma_1^2} + \frac{2\rho\mu_1 y}{(1 - \rho^2)\sigma_1\sigma_2}$$

$$- \frac{2\rho xy}{(1 - \rho^2)\sigma_1\sigma_2} + \frac{x^2}{(1 - \rho^2)\sigma_2^2} + \frac{y^2}{(1 - \rho^2)\sigma_1^2} + \frac{\mu_1^2}{\sigma_2^2} + \frac{\mu_2^2}{\sigma_1^2} - 2\rho \frac{\mu_1\mu_2}{\sigma_1\sigma_2}$$

$$\eta_1 = \frac{\mu_1}{(1 - \rho^2)\sigma_2^2} - \frac{\rho\mu_2}{(1 - \rho^2)\sigma_1\sigma_2}, \quad \eta_2 = \frac{\mu_2}{(1 - \rho^2)\sigma_1^2} - \frac{\rho\mu_1}{(1 - \rho^2)\sigma_1\sigma_2}$$

$$\eta_3 = \frac{\rho}{(1 - \rho^2)\sigma_1\sigma_2}, \quad \eta_4 = -\frac{1}{2(1 - \rho^2)\sigma_2^2}, \quad \eta_5 = -\frac{1}{2(1 - \rho^2)\sigma_1^2}$$

$$p.d.f(x; \eta) = \exp \left\{ \eta_1 x + \eta_2 y + \eta_3 xy + \eta_4 x^2 + \eta_5 y^2 - A(\eta) \right\}$$

$$E[X] - E[X] E[Y] = \rho \sigma_1 \sigma_2$$

$$E_b[X_1] = (\bar{x}, \bar{y}, \bar{xy}, \bar{x^2}, \bar{y^2})$$

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}, \quad (\hat{\mu}_1)^2 + \hat{\sigma}_1^2 = \bar{x^2},$$

$$(\hat{\mu}_2)^2 + \hat{\sigma}_2^2 = \bar{y^2} \quad \hat{\rho} \hat{\sigma}_1 \hat{\sigma}_2 + \hat{\mu}_1 \hat{\mu}_2 = \bar{xy}$$

$$\therefore \hat{C}^{MLE} = \frac{\bar{xy} - \bar{x}\bar{y}}{\sqrt{\bar{x^2} - (\bar{x})^2} \sqrt{\bar{y^2} - (\bar{y})^2}}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

12

20.08.29.

e.J. 6.4. |

(a) Bernoulli,  $0 \leq p \leq 1$

$$\hat{p}^{\text{MLE}} = \bar{X} = \hat{p}^{\text{MME}}$$

$$\Rightarrow \hat{p}^{\text{MLE}} \xrightarrow[n \rightarrow \infty]{P} p \quad \& \quad \sqrt{n}(\hat{p}^{\text{MLE}} - p) \xrightarrow[n \rightarrow \infty]{d} N(0, p(1-p))$$

(b) Poisson( $\lambda$ ),  $\lambda \geq 0$

$$\hat{\lambda}^{\text{MLE}} = \bar{X} = \hat{\lambda}^{\text{MME}}$$

$$\Rightarrow \hat{\lambda}^{\text{MLE}} \xrightarrow[n \rightarrow \infty]{P} \lambda \quad \& \quad \sqrt{n}(\hat{\lambda}^{\text{MLE}} - \lambda) \xrightarrow[n \rightarrow \infty]{d} N(0, \lambda)$$

(c)  $\text{Exp}(\theta)$   $\theta > 0$   $\hat{\theta}^{\text{MLE}} = \bar{X}$

(d)  $N(\mu, \sigma^2)$ ,  $-\infty < \mu < +\infty$   $\sigma^2 > 0$

$$\hat{\mu}^{\text{MLE}} = \bar{X} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

최대가능 추정(ML)

일치성을 가지려면.

$X_1, \dots, X_n$  : r.s.

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

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

If there uniquely exists maximum likelihood estimation  $\hat{\theta}_n$ ,

$$\hat{\theta}_n = \operatorname{argmax}_{\theta \in \mathcal{N}} l_n(\theta) = \operatorname{argmax}_{\theta \in \mathcal{N}} \bar{l}_n(\theta)$$

모두의 확률이 0이하면 대수의 법칙이 있다

$$\begin{aligned} \bar{l}_n(\theta) &\xrightarrow[n \rightarrow \infty]{P_\theta^0} E_{\theta^0} [\log f(X_i; \theta)] \\ \operatorname{argmax}_{\hat{\theta}_n} \bar{l}_n(\theta) &\xrightarrow[n \rightarrow \infty]{P_\theta^0} \operatorname{argmax}_{\theta \in \mathcal{N}} E_{\theta^0} [\log f(X_i; \theta)] = \hat{\theta}^0 \end{aligned}$$

Q1.  $\operatorname{argmax}_{\theta \in \mathcal{N}} E_{\theta^0} [\log f(X_i; \theta)] = \theta^0$  ?

Q2.  $\operatorname{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta^0$  ?

Thm 6.4.1

<p>KL-divergence</p> <p>(R<sub>0</sub>) (identifiable) <math>f(\cdot; \theta) = f(\cdot; \theta^*) \Rightarrow \theta = \theta^*</math></p> <p>(R<sub>1</sub>) (common Support) <math>\{x : f(x; \theta) &gt; 0\} \subset \mathbb{R}</math>에 의존 x.</p> <p><math>KL(\theta, \theta^*) := -E_{\theta^*} [\log \frac{f(x; \theta)}{f(x; \theta^*)}]</math></p> <p>(i) <math>KL(\theta, \theta^*) \geq 0</math>, <math>\forall \theta \in \mathbb{R}</math></p> <p>(ii) <math>KL(\theta, \theta^*) = 0 \Leftrightarrow \theta = \theta^*</math></p>	$X \sim f(x; \theta) \quad \theta \in \mathbb{R}$ <p><math>t \in \mathbb{R}, \quad -\log t \geq -t + 1</math></p>
<p>proof)</p> <p><math>-\log \frac{f(x; \theta)}{f(x; \theta^*)} \geq -\frac{f(x; \theta)}{f(x; \theta^*)} + 1</math></p> <p><math>KL(\theta, \theta^*) = -E_{\theta^*} [-\log \frac{f(x; \theta)}{f(x; \theta^*)}] \geq E_{\theta^*} [-\frac{f(x; \theta)}{f(x; \theta^*)} + 1]</math></p> <p><math>E_{\theta^*} [-\frac{f(x; \theta)}{f(x; \theta^*)} + 1] = - \int_{-\infty}^{+\infty} \frac{f(x; \theta)}{f(x; \theta^*)} f(x; \theta^*) dx + 1</math></p> <p><math>= -1 + 1 = 0</math></p> <p><math>\therefore KL(\theta, \theta^*) \geq 0</math>. equality holds when <math>t=1</math>. i.e., <math>\frac{f(x; \theta)}{f(x; \theta^*)} = 1</math>.</p> <p>By identifiability, <math>f(\cdot; \theta) = f(\cdot; \theta^*) \Rightarrow \theta = \theta^*</math></p>	<p><math>\therefore \text{Monotonicity of expectation.}</math></p> <p><math>\therefore KL(\theta, \theta^*) = 0 \text{ if and only if } \theta = \theta^*</math> <math>\square</math></p>

To answer Q<sub>1</sub>,

$$\begin{aligned} \underset{\theta \in \mathcal{R}}{\operatorname{argmax}} E_{\theta^0} [\log f(x_1; \theta)] &= \underset{\theta \in \mathcal{R}}{\operatorname{argmin}} E_{\theta^0} [-\log f(x_1; \theta)] \\ &= \underset{\theta \in \mathcal{R}}{\operatorname{argmin}} E_{\theta^0} \left[ -\log \frac{f(x_1; \theta)}{f(x_1; \theta^0)} \right] \end{aligned}$$

$$\begin{array}{l} \underset{\theta \in \mathcal{R}}{\operatorname{argmin}} KL(\theta, \theta^0) = 0 \\ \Leftrightarrow \theta = \theta^0. \end{array} \quad \begin{array}{l} = \underset{\theta \in \mathcal{R}}{\operatorname{argmin}} KL(\theta, \theta^0) \\ = \theta^0 \end{array}$$

To answer Q<sub>2</sub>,

We assume

i) uniform convergence in probability

$$\sup_{|\theta - \theta^0| \leq K} |\bar{l}_n(\theta) - E_{\theta^0} [\log f(x_1; \theta)]| \xrightarrow{n \rightarrow \infty} 0, \quad \forall K > 0.$$

ii) continuity of expectation of log likelihood

$E_{\theta} [\log f(x; \theta)]$ : continuous w.r.t.  $\theta$ .

iii) Concave of log likelihood

$$C^t \bar{l}_n(\theta) < 0 \quad \forall C \neq 0, \quad \forall \theta \in \mathcal{R} \quad \lim_{\theta \rightarrow \partial(\mathcal{R})} \bar{l}_n(\theta) = -\infty.$$

Theorem 6.4.2.  $f(\cdot, \theta) = f(\cdot, \theta^0) \Rightarrow \theta = \theta^0, \quad \exists \{x: f(x; \theta) > 0\}: \text{independent of } \theta,$

$$C^t \bar{l}_n(\theta) < 0 \quad \forall C \neq 0, \quad \forall \theta \in \mathcal{R} \quad \lim_{\theta \rightarrow \partial(\mathcal{R})} \bar{l}_n(\theta) = -\infty$$

$\forall \theta^0 \in \mathcal{R}, \quad \exists E_{\theta^0} [\log f(x_1; \theta)]$  & continuous w.r.t.  $\theta \in \mathcal{R}$

$$\Rightarrow \exists! \hat{\theta}_n^{\text{MLE}} \text{ s.t. } \dot{l}_n(\hat{\theta}_n^{\text{MLE}}) = 0 \quad \& \quad \hat{\theta}_n^{\text{MLE}} \xrightarrow{n \rightarrow \infty} \theta, \quad \forall \theta \in \mathcal{R}$$

Asymptotic distribution  
of maximum likelihood

Let  $\hat{\theta}_n$ : MLE.

$$\bar{l}_n(\theta) := \frac{1}{n} l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(x_i; \theta)$$

$$0 = \bar{l}'_n(\hat{\theta}_n) \approx \bar{l}'_n(\theta) + \bar{l}''_n(\theta)(\hat{\theta}_n - \theta)$$

$$\therefore \sqrt{n}(\hat{\theta}_n - \theta) \approx -(\bar{l}''_n(\theta))^{-1} \sqrt{n} \bar{l}'_n(\theta)$$

By the law of large number,

$$-\bar{l}''_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left[ -\frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) \right] \xrightarrow[n \rightarrow \infty]{P_\theta} E_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) \right]$$

$$\text{Let } I(\theta) := E_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) \right]$$

$$\sqrt{n}(\hat{\theta}_n - \theta) \approx (I(\theta))^{-1} \sqrt{n} \bar{l}'_n(\theta)$$

Since  $\frac{\partial}{\partial \theta} \log f(x_i; \theta)$  : iid,

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i; \theta) - E_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right] \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right])$$

$$E_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right] = \int_{-\infty}^{+\infty} \frac{\frac{\partial}{\partial \theta} f(x; \theta)}{f(x; \theta)} f(x; \theta) dx$$

$$\text{under "some" condition.} \quad \hookrightarrow \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} f(x; \theta) dx = 0$$

$$\therefore \sqrt{n} \cdot \bar{l}'_n(\theta) \xrightarrow[n \rightarrow \infty]{d} N(0, \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right])$$

$$\therefore \sqrt{n}(\hat{\theta}_n - \theta) \approx (I(\theta))^{-1} \sqrt{n} \bar{l}'_n(\theta) \xrightarrow[n \rightarrow \infty]{d} N(0, [I(\theta)]^2 \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right])$$

"Asymptotic Normality"

최대가능도 추정량의 점근 정규성을 위한 기본 조건:

(R0) (식별 가능성)  $f(\cdot; \theta) = f(\cdot; \theta^0) \Rightarrow \theta = \theta^0$

(R1) (공통의 토대)  $\{x : f(x; \theta) > 0\} = \emptyset$ 로서  $\theta \in \Omega$ 에 의존하지 않는다.

(R2) (열린 모수공간) 모수공간  $\Omega$ 은  $k$  차원 공간  $R^k$ 에서의 열린집합이다.

(R3) (미분가능한 로그가능도) 모든 관측 결과  $(x_1, \dots, x_k)^t, x_i \in \mathcal{X} (i=1, \dots, n)$ 에 대하여 로그가능도함수  $l_n(\theta) = \sum_{i=1}^n \log f(x_i; \theta)$ 의 일차 및 이차 편도함수<sup>99)</sup>  $\dot{l}_n(\theta), \ddot{l}_n(\theta)$ 가 존재하며 이들은 모두 연속함수이다.

(R4) (적분 또는 합과 미분의 순서 교환 가능) 랜덤표본  $X = (X_1, \dots, X_n)^t$ 의 함수  $u(X)$ 의 기댓값  $E_\theta[u(X)]$ 이 존재할 때, 모수  $\theta$ 에 대한 미분<sup>100)</sup>을  $\partial_\theta, \partial_\theta^2$ 으로 나타내면

$$\partial_\theta^r E_\theta[u(X)] = \begin{cases} \int_{R^n} u(x) \partial_\theta^r (pdf(x; \theta)) dx & (\text{연속형인 경우}) \\ \sum_{x \in R^n} u(x) \partial_\theta^r (pdf(x; \theta)) & (\text{이산형인 경우}) \end{cases} \quad (r=1,2)$$

$$\text{여기에서 } pdf(x; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

(R5) (정보량의 존재) 모든  $\theta \in \Omega$ 에 대하여

$$I(\theta) = \text{Var}_\theta \left( \frac{\partial}{\partial \theta} \log f(X_1; \theta) \right)$$

가 실수 또는 실수 행렬로 잘 정의될 수 있으며 그 역수 또는 역행렬이 존재한다.

Thm 6.4.3.

If (R0)~(R5) holds

$$(a) E_\theta \left[ \frac{\partial}{\partial \theta} \log f(X_1; \theta) \right] = 0$$

$$(b) I(\theta) = \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(X_1; \theta) \right] = E_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f(X_1; \theta) \right]$$

Proof)

$$\int_{\mathcal{X}} f(x; \theta) dx = 1 \quad i.e., \int_{\mathcal{X}} \exp(\log f(x; \theta)) dx = 1$$

$$\Rightarrow 0 = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} \exp(\log f(x; \theta)) dx$$

$$= \int_{\mathcal{X}} \frac{\partial}{\partial \theta} \exp(\log f(x; \theta)) dx$$

$$= \int_{\mathcal{X}} \exp(\log f(x; \theta)) \frac{\partial}{\partial \theta} \log f(x; \theta) dx = E_\theta \left[ \frac{\partial}{\partial \theta} \log f(x; \theta) \right]$$

$$\Rightarrow 0 = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} \exp(\log f(x; \theta)) \frac{\partial}{\partial \theta} \log f(x; \theta) dx$$

$$= \int_{\mathcal{X}} \left[ \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) + \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right] \exp(\log f(x; \theta)) dx$$

$$= E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f(x_1; \theta) + \left( \frac{\partial}{\partial \theta} \log f(x_1; \theta) \right)^2 \right]$$

$$\therefore E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log f(x_1; \theta) \right)^2 \right] = -E_\theta \left[ \frac{\partial^2}{\partial \theta^2} \log f(x_1; \theta) \right]$$

$$\therefore \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_1; \theta) \right] = E_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f(x_1; \theta) \right]$$

□

## Thm 6.4.4

(R6) (일치성) 랜덤표본  $X_1, \dots, X_n$ 을 이용한  $\theta$ 의 최대가능도 추정량  $\hat{\theta}_n^{MLE}$  가 가능도방정식

$$l'_n(\theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(X_i; \theta) = 0$$

의 단 하나뿐인 근이고 일치성을 갖는다. 즉

$$\hat{\theta}_n^{MLE} \xrightarrow[n \rightarrow \infty]{P_\theta} \theta$$

(R7) (잉여항의 크기) 로그가능도함수의 삼차 편도함수<sup>101)</sup>가 연속함수로 존재하며

$$\max_{\theta \in \Omega} \| \partial_\theta^3 \log f(X_1; \theta) \| \leq M(X_1), E_\theta M(X_1) < +\infty$$

을 만족하는  $M(X_1)$ 이 존재한다.

이러한 조건들이 만족된다면 최대가능도 추정량은 다음의 극한분포를 갖는다.

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \xrightarrow[n \rightarrow \infty]{d} N(0, [I(\theta)]^{-1})$$

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \xrightarrow[n \rightarrow \infty]{d} N(0, (I(\theta))^{-1})$$

Proof)

Step 1.

$$0 = \bar{l}'_n(\hat{\theta}_n) = \bar{l}'_n(\theta) + \bar{l}''_n(\theta)(\hat{\theta}_n - \theta) + \frac{1}{2} \bar{l}'''_n(\theta^*)(\hat{\theta}_n - \theta)^2$$

for some  $\theta^* \in |\hat{\theta}_n - \theta| \leq |\hat{\theta}_n - \theta|$  by Taylor's thm.

$$\text{Let } R_n := \frac{1}{2} \bar{l}'''_n(\theta^*)(\hat{\theta}_n - \theta)^2$$

Step 2

$$\text{WTS} \quad R_n = \frac{1}{2} \overline{\ell_n^{(3)}}(\theta_n^*) (\hat{\theta}_n - \theta)^2 \xrightarrow[n \rightarrow \infty]{P_\theta} 0.$$

 $\forall \epsilon, K > 0$ 

$$P_\theta(|R_n| \geq \epsilon) = P\left(|\overline{\ell_n^{(3)}}(\theta_n^*)| \geq K, |R_n| \geq \epsilon\right) + P\left(|\overline{\ell_n^{(3)}}(\theta_n^*)| \leq K, |R_n| \geq \epsilon\right)$$

Since  $|\overline{\ell_n^{(3)}}(\theta_n^*)| \leq K \Rightarrow \epsilon \leq |R_n| = \frac{1}{2} |\overline{\ell_n^{(3)}}(\theta_n^*)| \cdot |\hat{\theta}_n - \theta|^2$

$$\because |\hat{\theta}_n - \theta| \ll 1 \quad \begin{cases} \leq \frac{1}{2} |\hat{\theta}_n - \theta|^2 K \\ \leq \frac{1}{2} |\hat{\theta}_n - \theta| K \end{cases}$$

$$P_\theta(|R_n| \geq \epsilon) \leq P\left(|\overline{\ell_n^{(3)}}(\theta_n^*)| \geq K\right) + P_\theta(|\hat{\theta}_n - \theta| \geq 2\epsilon/K)$$

$$\begin{aligned} P_\theta(|\overline{\ell_n^{(3)}}(\theta_n^*)| \geq K) &= P_\theta\left(\left|\frac{1}{n} \sum_{i=1}^n \partial_\theta^3 \log f(x_i; \theta)\right| \geq K\right) \\ &\leq P_\theta\left(\left|\frac{1}{n} \sum_{i=1}^n \partial_\theta^3 \log f(x_i; \theta)\right| \geq K\right) \\ &\leq P_\theta\left(\frac{1}{n} \sum_{i=1}^n \max_{\theta \in \Omega} |\partial_\theta^3 \log f(x_i; \theta)| \geq K\right) \end{aligned}$$

$$\begin{aligned} \because \text{Markov's Inequality} &\leq P_\theta\left(\frac{1}{n} \sum_{i=1}^n M(X_i) \geq K\right) \\ &\leq E_\theta\left[\frac{1}{n} \sum_{i=1}^n M(X_i)\right] / K \\ &= E_\theta[M(X_1)] / K \end{aligned}$$

$$P_\theta(|R_n| \geq \epsilon) = P_\theta(|\bar{L}_n^{(1)}(\theta^*)| > K) + P_\theta(|\hat{\theta}_n - \theta| \geq \frac{2\epsilon}{K})$$

$$\leq E_\theta[M(X_1)]/K + P_\theta(|\hat{\theta}_n - \theta| \geq \frac{2\epsilon}{K})$$

Since  $\underset{n \rightarrow \infty}{plim} \hat{\theta}_n = \theta$ ,  $\lim_{n \rightarrow \infty} P_\theta(|R_n| \geq \epsilon) = E_\theta[M(X_1)]/K$

AS  $K \rightarrow \infty$ ,  $\lim_{n \rightarrow \infty} P_\theta(|R_n| \geq \epsilon) = 0$ .

$$\therefore R_n \xrightarrow[n \rightarrow \infty]{P_\theta} 0$$

Step 3.

$$0 = \bar{L}_n(\theta) + \left( \bar{L}_n(\theta) + \frac{1}{2} \bar{L}_n^{(1)}(\theta^*) (\hat{\theta}_n - \theta) \right) (\hat{\theta}_n - \theta)$$

$$i) \quad \sqrt{n}(\hat{\theta}_n - \theta) = \left( -\bar{L}_n(\theta) - \frac{1}{2} \bar{L}_n^{(1)}(\theta^*) (\hat{\theta}_n - \theta) \right) \sqrt{n} \dot{L}_n(\theta)$$

$$\text{by the law of large number} \quad -\bar{L}_n(\theta) = \sum_{i=1}^n \left( -\frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) \right) \xrightarrow[n \rightarrow \infty]{P_\theta} E_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) \right] \stackrel{\text{II}}{=} I(\theta)$$

$$ii) \quad R_n \xrightarrow[n \rightarrow \infty]{P_\theta} 0, \quad -\frac{1}{2} \bar{L}_n^{(1)}(\theta^*) (\hat{\theta}_n - \theta) =: r_n \xrightarrow[n \rightarrow \infty]{P_\theta} 0.$$

$$iii) \quad \sqrt{n} \dot{L}_n(\theta) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i; \theta) - E_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right] \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \text{Var}_\theta \left[ \frac{\partial}{\partial \theta} \log f(x_i; \theta) \right]) = I(\theta)$$

By i) ~ iii) & Slutsky's thm,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} (I(\theta))^{-1} N(0, I(\theta)) \stackrel{d}{=} N(0, (I(\theta))^{-1})$$

□

20.09.05

Lemma 1.

Let  $V$  be a finite dimensional inner product vector space,  $W \subseteq V$ .

$\Rightarrow \forall y \in V, \exists! w \in W, u \in W^\perp$  s.t.  $y = w + u$ .

Proof) Let  $\mathcal{B} := \{v_1, \dots, v_k\}$  an orthonormal basis for  $W$ .

$$w := \sum_{i=1}^k \langle y, v_i \rangle v_i, \quad u := y - w$$

By the definition,  $w \in W$ .

We want to show  $u \in W^\perp$ , i.e.,  $\langle u, v_j \rangle = 0 \quad \forall j = 1, \dots, k$ .

$$\langle u, v_j \rangle = \langle y - \sum_{i=1}^k \langle y, v_i \rangle v_i, v_j \rangle$$

$$= \langle y, v_j \rangle - \sum_{i=1}^k \langle y, v_i \rangle \langle v_i, v_j \rangle$$

$$= \langle y, v_j \rangle - \sum_{i=1}^k \langle y, v_i \rangle \delta_{ij}$$

$$= 0.$$

Thus  $u \in W^\perp$ .

To show the uniqueness of  $(w, u)$ ,

Let  $y = w + u = w' + u'$   $w' \in W, u' \in W^\perp$ .

$$(w - w') = (u' - u) \quad \& \quad w - w' \in W, u' - u \in W^\perp$$

Since  $W \cap W^\perp = \{0\}$ ,  $w - w' = 0, u' - u = 0$ .

i.e.,  $w = w', u = u'$ .

$\therefore \forall y \in V, \exists! w \in W, u \in W^\perp$  s.t.  $y = w + u$

D

Corollary.

$$\forall y \in V, \exists ! w \in W \text{ s.t } \|y - x\| \geq \|y - w\| \quad \forall x \in W.$$

proof).

By the lemma 1,  $\exists ! w \in W, u \in W^\perp$  s.t  $y = w + u$

$$\begin{aligned} \|y - x\|^2 &= \|w + u - x\|^2 = \|(w - x) + u\|^2 \\ &= \|w - x\|^2 + \|u\|^2 \quad (\because w - x \perp u) \\ &\geq \|u\|^2 = \|y - w\|^2 \end{aligned}$$

$$\text{Suppose } \|y - x\| = \|y - w\|$$

$$\text{Then } \|w - x\|^2 + \|u\|^2 = \|u\|^2, \text{ i.e., } \|w - x\|^2 = 0.$$

$$\text{Thus } x = w.$$

□

Lemma 2.

$$A \in M_{m \times n}(F) \quad x \in F^n \quad y \in F^m$$

$$\langle Ax, y \rangle_m = \langle x, A^*y \rangle_n. \quad \text{where } A^* := (A)^t$$

proof)

$$\langle Ax, y \rangle_m = y^* A x = (A^* y)^* x = \langle x, A^* y \rangle_n.$$

□

Lemma 3.

$$\text{Let } A \in M_{m \times n}(F) \quad \text{Then } \text{rank}(A^* A) = \text{rank } A.$$

proof)

$$\boxed{\text{ETS}} \quad \ker A^* A = \text{rank } A \quad (\because n = \text{rank } A + \text{nullity of } A)$$

Suppose  $x \in \ker A^* A$ .

$$\langle x, A^* A x \rangle = \langle Ax, Ax \rangle_m \text{ by the lemma 2.}$$

Since  $A^* A x = 0$ ,  $\langle Ax, Ax \rangle = 0$ . i.e.,  $Ax = 0$ . That is  $x \in \ker A$ .

Clearly  $Ax = 0$  implies  $A^* A x = 0$ .

$$\therefore \text{rank } A^* A = \text{rank } A.$$

□

Thm.

Let  $A \in M_{m \times n}(F)$   $y \in F^m$ .  $\exists! x_0 \in F^n$  s.t.  $(A^*A)x = A^*y$   
&  $\|Ax_0 - y\| \leq \|Ax - y\| \quad \forall x \in F^n$ .

proof)

By the corollary to the lemma 1,  $\exists! x_0$  s.t  $\|Ax_0 - y\| \leq \|Ax - y\|$ ,  
for all  $x \in F^n$ . Since  $Ax_0 - y \perp Ax \quad \forall x \in F^n$ ,

$$\langle Ax, Ax_0 - y \rangle_m = \langle x, A^*(Ax_0 - y) \rangle_n = 0 \quad \text{by the lemma 2.}$$

$$\text{Thus } A^*(Ax_0 - y) = 0. \text{ i.e., } A^*Ax_0 = A^*y.$$

By the lemma 3,  $\text{rank } A^*A = \text{rank } A$ .

Suppose  $\text{rank } A = n$ , i.e., column full rank.

Then  $A^*A$  invertible.

$$\therefore x_0 = (A^*A)^{-1}A^*y.$$

D

## 09.05. Statistics.

Least Square

Estimation.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + e_i$$

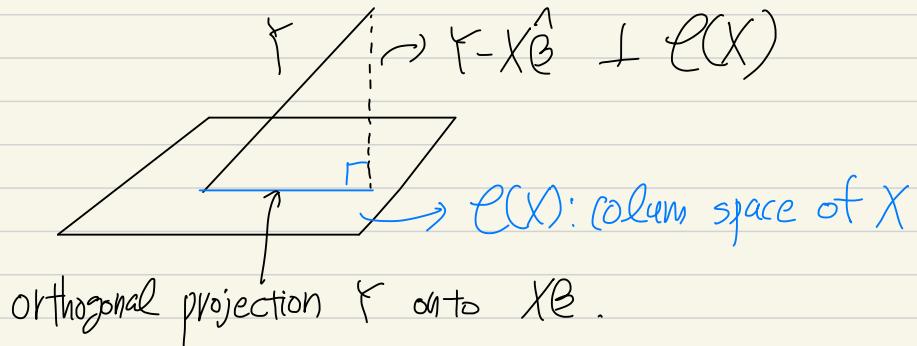
$$X = \begin{bmatrix} 1 & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ 1 & \dots & X_{np} \end{bmatrix} \in M_{n \times (p+1)}(\mathbb{R}) \quad Y = (Y_1, \dots, Y_n)' \quad E = (e_1, \dots, e_n)' \quad \beta = (\beta_0, \dots, \beta_p)'$$

$$E[e] = 0, \quad \text{Var}(e) = \sigma^2 I_n, \quad 0 < \sigma^2 < \infty.$$

rank  $X = p+1$ , i.e., full column rank.

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \|Y - X\beta\|^2 = (X'X)^{-1}X'Y$$

(\*) By the previous thm,  $\exists! \hat{\beta} \in \mathbb{R}^{p+1}$  s.t.  $\|Y - X\hat{\beta}\| \leq \|Y - X\beta\| \quad \forall \beta \in \mathbb{R}^{p+1}$



We know that the solution of normal equation is

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \|Y - X\beta\|^2.$$

$$X'X\hat{\beta} = X'Y \quad (*) \text{ Since } X: \text{full rank and}$$

$$\hat{\beta} = (X'X)^{-1}X'Y \quad \text{rank } X'X = \text{rank } X \text{ by the lemma 2,}$$

$$\Rightarrow X\hat{\beta} = \underbrace{X(X'X)^{-1}X'Y}_{\text{projection map onto } C(X)}.$$

Thm 6.5.1

$$(a) \Pi := X(X'X)^{-1}X$$

$$\Pi' = \Pi, \quad \Pi'(I - \Pi) = 0, \quad \Pi X = X.$$

$$(b) \|F - X\beta\|^2 = \|\Pi(F - X\beta)\|^2 + \|(I - \Pi)F\|^2$$

$$\hat{X}\hat{\beta}^{\text{LSE}} = \Pi F, \quad \hat{\beta}^{\text{LSE}} = (X'X)^{-1}X'F.$$

proof)

$$(a) \Pi' = (X(X'X)^{-1}X')' = X((X'X)^{-1})'X' = X(X'X)^{-1}X = \Pi.$$

$$\begin{aligned} & (X(X'X)^{-1}X')' (I - X(X'X)^{-1}X) \\ &= X(X'X)^{-1}X' - X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= X(X'X)^{-1}X' - X(X'X)^{-1}X' = 0. \end{aligned}$$

or  $\Pi' = \Pi$ : projection onto  $\ell(X)$

$(I - \Pi)$ : projection onto  $\ell(X)^\perp$ .

$\therefore$  composition of two (linear) map is zero-map.

$$\Pi X = X(X'X)^{-1}X'X = X$$

$$\begin{aligned}
 (b) \|\mathbf{Y} - \mathbf{X}\beta\|^2 &= \|\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta) + (\mathbf{I} - \mathbf{\Pi})(\mathbf{Y} - \mathbf{X}\beta)\|^2 \\
 &= \|\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta) + (\mathbf{I} - \mathbf{\Pi})\mathbf{Y}\|^2 \\
 &= \|\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta)\|^2 + \|(\mathbf{I} - \mathbf{\Pi})\mathbf{Y}\|^2 \\
 &\quad + 2 \underbrace{\|(\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta))'(\mathbf{I} - \mathbf{\Pi})\mathbf{Y}\|^2}_{=0} \\
 &= (\mathbf{Y} - \mathbf{X}\beta)' \mathbf{\Pi}' (\mathbf{I} - \mathbf{\Pi}) \mathbf{Y} \\
 &= \|\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta)\|^2 + \|(\mathbf{I} - \mathbf{\Pi})\mathbf{Y}\|^2
 \end{aligned}$$

$$\|\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta)\|^2 = 0 \Rightarrow \text{minimum of } \|\mathbf{Y} - \mathbf{X}\beta\|^2$$

Since  $\|\mathbf{X}\| = 0 \Leftrightarrow \mathbf{X} = \mathbf{0}$ ,  $\mathbf{\Pi}(\mathbf{Y} - \mathbf{X}\beta) = \mathbf{0}$

$$\therefore \mathbf{X}\beta^{\text{LSE}} = \mathbf{\Pi}\mathbf{Y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Y}$$

□

Thm 6.5.2.

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e} \quad \mathbf{X} \in \mathbb{M}_{n \times (p+1)}(\mathbb{R})$$

$$E[\mathbf{e}] = \mathbf{0}, \quad \text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}_n, \quad \beta \in \mathbb{R}^{p+1}, \quad \text{rank } \mathbf{X} = p+1.$$

$$\hat{\sigma}^2 = \|\mathbf{Y} - \mathbf{X}\hat{\beta}^{\text{LSE}}\|^2 / (n-p-1)$$

(a)  $E[\hat{\beta}^{\text{LSE}}] = \beta \quad \text{Var}(\hat{\beta}^{\text{LSE}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$

(b)  $E[\hat{\sigma}^2] = \sigma^2$

(c) If  $\mathbf{e} \sim N_n(0, \sigma^2 \mathbf{I}_n)$   $\Rightarrow \hat{\beta}^{\text{LSE}} \sim N(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$   
 $(n-p-1)\hat{\sigma}^2 \sim \chi^2_{n-p-1}$

proof)

(a)  $\hat{\beta}^{\text{LSE}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$

$$\begin{aligned} E[\hat{\beta}^{\text{LSE}}] &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' E[\mathbf{Y}], \quad E[\mathbf{Y}] = E[\mathbf{X}\beta + \mathbf{e}] \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X}\beta \\ &= \beta. \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{\beta}^{\text{LSE}}) &= \text{Var}((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

$$(b) \quad \text{Var}(\hat{\sigma}^2) = \sigma^2$$

$$\hat{\sigma}^2 = \| \mathbf{Y} - \mathbf{X}\hat{\beta}_{\text{LSE}} \|^2 / (n-p-1)$$

$$\| \mathbf{Y} - \mathbf{X}\hat{\beta}_{\text{LSE}} \|^2 = \| \mathbf{Y} - \mathbf{X}\hat{\beta}_{\text{LSE}} \|^2 + \| (\mathbf{I} - \mathbf{P})\mathbf{Y} \|^2$$

$$= \| (\mathbf{I} - \mathbf{P})\mathbf{Y} \|^2 \quad \mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$$

$$= \| (\mathbf{I} - \mathbf{P})\mathbf{e} \|^2$$

$$= \mathbf{e}' (\mathbf{I} - \mathbf{P})' (\mathbf{I} - \mathbf{P}) \mathbf{e}$$

$$= \mathbf{e}' (\mathbf{I} - \mathbf{P}) \mathbf{e} \in \mathbb{R}.$$

$$= \text{Tr}(\mathbf{e}' (\mathbf{I} - \mathbf{P}) \mathbf{e})$$

$$= \text{Tr}((\mathbf{I} - \mathbf{P}) \mathbf{e} \mathbf{e}' )$$

$$(n-p-1) E[\hat{\sigma}^2] = E[\mathbf{e}' (\mathbf{I} - \mathbf{P}) \mathbf{e}]$$

$$= E[\text{Tr}((\mathbf{I} - \mathbf{P}) \mathbf{e} \mathbf{e}')]$$

$$= \text{Tr}(E[(\mathbf{I} - \mathbf{P}) \mathbf{e} \mathbf{e}'])$$

$$= \text{Tr}((\mathbf{I} - \mathbf{P}) E[\mathbf{e} \mathbf{e}'])$$

$$= \text{Tr}((\mathbf{I} - \mathbf{P}) \text{Var}(\mathbf{e}))$$

$$= \text{Tr}(\sigma^2 (\mathbf{I} - \mathbf{P}))$$

$$= \sigma^2 \text{Tr}(\mathbf{I} - \mathbf{P})$$

$$= \sigma^2 (n - \text{Tr}(\mathbf{X}(\mathbf{X}'\mathbf{X}')^\dagger))$$

$$= \sigma^2 (n - \text{Tr}(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^\dagger))$$

$$= \sigma^2 (n - \text{Tr} \mathbf{I}_{(p+1)}) = \sigma^2 (n-p-1)$$

$$\therefore E[\hat{\sigma}^2] = \sigma^2$$

$$\begin{aligned} \text{Var}(\mathbf{e}) &= \mathbf{E}[(\mathbf{e} - \mathbf{0})(\mathbf{e} - \mathbf{0})'] \\ &= \mathbf{E}[\mathbf{e} \mathbf{e}'] \end{aligned}$$

(c)  $\epsilon \sim N(0, \sigma^2 I_n)$

$\Rightarrow \epsilon \sim N(X\beta, \sigma^2 I_n)$  &  $\hat{\beta} \perp\!\!\!\perp \hat{\epsilon}$

$$\begin{aligned}\hat{\beta}_{LS} &= (X'X)^{-1} X' \epsilon \sim N((X'X)^{-1} X' X \beta, \sigma^2 (X'X)^{-1}) \\ &\stackrel{d}{=} N(\beta, \sigma^2 (X'X)^{-1})\end{aligned}$$

$$(n-p-1)\hat{\sigma}^2/\sigma^2 = (n-p-1)(Y - X\hat{\beta}_{LS})'(Y - X\hat{\beta}_{LS})/\sigma^2$$

$$= (n-p-1)(Y - X\beta)'(I - \Pi)(Y - X\beta)/\sigma^2$$

$$= (n-p-1) \frac{(Y - X\beta)'}{\sigma} (I - \Pi) \frac{(Y - X\beta)}{\sigma}$$

Since  $(Y - X\beta)/\sigma \sim N(0, I_n)$ ,

$$(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-p-1) \stackrel{d}{=} \chi^2(n-p-1)$$

$$\text{Cov}((X'X)^{-1} X' Y, (I - \Pi) Y) = (X'X)^{-1} X' \text{Cov}(Y, Y) (I - \Pi)'$$

$$= \sigma^2 (X'X)^{-1} X' (I - \Pi)$$

$$= \sigma^2 \left\{ (X'X)^{-1} X' - (X'X)^{-1} X' \cancel{X' X (X'X)^{-1} X'} \right\}$$

$$= \sigma^2 \text{O} = \text{O}.$$

$\therefore \hat{\beta} = (X'X)^{-1} X' Y \perp\!\!\!\perp (I - \Pi) Y$

$$\begin{aligned}\therefore \hat{\beta} \perp\!\!\!\perp \hat{\epsilon} &= (Y - X\hat{\beta})' (Y - X\hat{\beta}) = ((I - \Pi) Y)' ((I - \Pi) Y) \\ &= g((I - \Pi) Y)\end{aligned}$$

Q

$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$

$Y - X\hat{\beta}$