Chapter 6

08.15.

Chapter 6.

Estimation

§ 6.1 MME

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

$\hat{m}_k := \overline{X^k} := \frac{1}{n}\sum_{i=1}^{k} X_i^k$: k차 표본적률. estimation.

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

Method of moment estimator

e.g. 6.1.1

$X_1, \dots, X_n$: r.s $Var(X_1) = \sigma^2$

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

$\hat{\sigma}^{2\,MME} = \hat{m}_2 - (\hat{m}_1)^2 = \overline{X^2} - (\overline{X})^2$

$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

e.g. 6.1.2.

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

$E[X_1] = \lambda$ $Var(X_1) = \lambda$

$\hat{\lambda}^{MME} = \hat{m}_1 = \overline{X}$ or $\hat{m}_2 - \hat{m}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$\Rightarrow$ MME is not unique.

◎ $m_{r,s} := E[X_1^r Y_1^s] \to \hat{m}_{r,s} := \sum_{i=1}^{n} X_i^r Y_i^s / n$

joint-mgf

$$= \overline{X_1^r Y_1^s}$$

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

$$\rho = \frac{Cov(X_1, Y_1)}{\sqrt{Var(X_1)}\sqrt{Var(Y_1)}} = \frac{E[X_1Y_1] - E[X_1]E[Y_1]}{\sqrt{E[X_1^2] - E[X_1]^2} \cdot \sqrt{E[Y_1^2] - E[Y_1]^2}}$$

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

Thm 6.1.1. $\eta = g(m_1, \dots, m_k) \quad m_k \in \mathbb{R}.$

$$\Rightarrow \operatorname*{plim}_{n\to\infty} \hat{\eta}_n^{MME} = \eta$$ (consistency of $\hat{\eta}_n^{MME}$)

proof) $\hat{m}_i \xrightarrow[n\to\infty]{p} m_i \quad \forall i = 1, \dots, k$

i.e., $(\hat{m}_1, \dots, \hat{m}_k)' \xrightarrow[n\to\infty]{p} (m_1, \dots, m_k)'$

So, $\hat{\eta}_n^{MME} = g(\hat{m}_1, \dots, \hat{m}_k) \xrightarrow[n\to\infty]{p} g(m_1, \dots, m_k) = \eta$

추정량의 일치성

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

e.g. 6.1.4

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

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

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

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

Thm 6.1.2.

$\eta = g(m_1, \ldots, m_k)$, $m_{2k} \in \mathbb{R}$, $g \in C^1$

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

where, $\sigma^2 = \dot{g}(m_1, \ldots, m_k)^t \Sigma \dot{g}(m_1, \ldots, m_k)$

$\Sigma = (m_{r+s} - m_r m_s)$

proof)

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

By CLT $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Y_i - E[Y_1]\right) \xrightarrow[n\to\infty]{d} W, \quad W \sim N_k(0, Var(Y_1))$

Let $m := (m_1, \ldots, m_k)$

By thm 5.3.3, $\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) - g(m)\right) \xrightarrow[n\to\infty]{d} Z_1$

where $Z_1 \stackrel{d}{\equiv} \dot{g}(m)^t W$

$Var(Z_1) = \dot{g}(m)^t Var(W)\dot{g}(m)$, where $Var(W) = Var(Y_1)$

$(Var(Y_1))_{r,s} = Cov(X_1^r, X_1^s) = E[X_1^r X_1^s] - E[X_1^r]E[X_1^s]$

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

$\therefore Z_1 \sim N(0, \sigma^2)$ where $\sigma^2 = \dot{g}(m)^t \Sigma \dot{g}(m) \quad (\Sigma)_{ij} = m_{i+j} - m_i m_j$ $\square$

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

§6.2. MLE

(Maximum Likelihood Estimation)

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

$\mathrm{pdf}_X(x;\theta)|\Delta x| = \prod_{i=1}^{n} f(x_i;\theta) \underline{|\Delta x_1| \cdots |\Delta x_n|}$

↳ independent of θ.

i.e., $\max_{\theta\in\Omega} \mathrm{pdf}_X(x;\theta)|\Delta x| = \max_{\theta\in\Omega} \mathrm{pdf}_X(x;\theta)$

x is given. Ω is a domain

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

Likelihood function.

(b) $\hat{\theta}^{MLE}(x) := \underset{\theta\in\Omega}{\mathrm{argmax}}\, L(\theta; x)$ "최대가능도 추정값"

$\hat{\theta}^{MLE}(X)$: 최대가능도 추정량

(c) $\ell(\theta; x) := \log L(\theta; x)$: log likelihood function

Rmk $\hat{\theta}^{MLE} = \underset{\theta}{\mathrm{argmax}}\, \ell(\theta; x) = \underset{\theta}{\mathrm{argmax}}\, L(\theta; x)$

e.g. 6.2.1

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

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

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

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

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

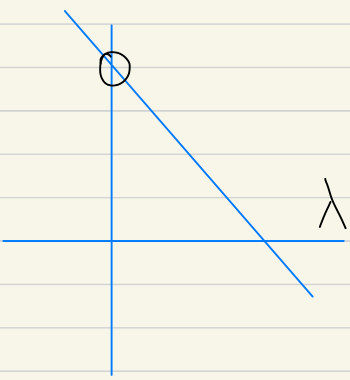
$\ddot{\ell}(\lambda) = -\frac{n\bar{x}}{\lambda^2}$

$\bar{x} > 0 \Rightarrow \ddot{\ell}(\lambda) < 0 \quad \forall \lambda > 0.$

$\Rightarrow \max_{\lambda > 0} \ell(\lambda) = \ell(\bar{x})$

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

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



Since $\ell(\lambda)$: continuous & $\lambda \ge 0$, $\sup_{\lambda > 0} \ell(\lambda) = \lim_{\ell \to 0^+} \ell(\lambda) = \ell(0)$
$= \ell(\bar{x}) \ (\bar{x} = 0)$

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

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

$\therefore \hat{\lambda}^{MLE} = \bar{X}$

Thm 6.2.1

Ω_0: open interval, $\ell(\theta)$: twice differentiable, and its second-order derivative is continuous ($\ddot{\ell}(\theta)$)

$\ddot{\ell}(\theta) < 0, \ \forall \theta \in \Omega_0$ and $\dot{\ell}(\hat{\theta}) = 0, \ \hat{\theta} \in \Omega_0$

$$\Rightarrow \max_{\theta \in \Omega_0} \ell(\theta) = \ell(\hat{\theta})$$

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 $\theta^* \in \Omega_0$

By the assumption that $\dot{\ell}(\hat{\theta}) = 0$ & $\ddot{\ell}(\theta) < 0 \ \forall \theta \ \Omega_0$

$$\ell(\theta) = \ell(\hat{\theta}) + \frac{1}{2}\ddot{\ell}(\theta^*)(\hat{\theta} - \theta)^2 < \ell(\hat{\theta}) \quad \forall \theta \in \Omega_0 \setminus \{\hat{\theta}\}$$

$$\therefore \max_{\theta \in \Omega_0} \ell(\theta) = \ell(\hat{\theta})$$

$\square$

e.g. 6.2.2.

$X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p) \qquad \hat{p}^{MLE}?$

$$L(p; x_1, \dots, x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

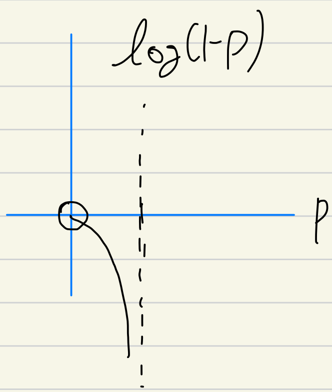
$$\ell(p) = \sum_{i=1}^{n} x_i \log p + (1-x_i)\log(1-p) \qquad 0<p<1.$$

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

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

$$p = \hat{p}$$

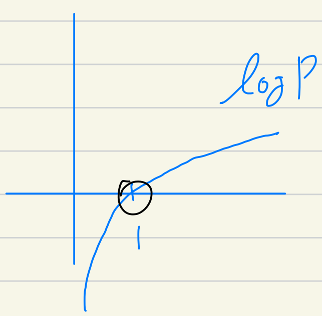
$$\ddot{\ell}(p) = -\frac{n\hat{p}}{p^2} - \frac{n(1-\hat{p})}{(1-p)^2}$$



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

ii) $\hat{p}=0$, $\ell(p) = n\log(1-p)$

$$\Rightarrow \sup_{0<p<1} \ell(p) = \lim_{p \to 0^+} \ell(p)$$



iii) $\hat{p}=1$, $\ell(p) = n\log p$

$$\Rightarrow \sup_{0<p<1} \ell(p) = \lim_{p \to 1^-} \ell(p)$$

Since L: continuous & $0 \le p \le 1$,

$$\sup_{0\le p\le 1} L(p; x_1, \dots, x_n) = L(\hat{p}; x_1, \dots, x_n)$$

$$\therefore \max_{0\le p\le 1} L(p; x_1, \dots, x_n) = L(\hat{p}; x_1, \dots, x_n)$$

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

$\square$

Thm. 6.2.2. Ω_0: open interval, $\ell(\theta)$: twice differentiable

$\ddot{\ell}(\theta)$: continuous.

$\ddot{\ell}(\theta) < 0 \ \forall \theta < \Omega_0$, $\lim_{\theta \to \partial(\Omega_0)} \ell(\theta) = -\infty$.

$\Rightarrow \exists! \ \hat{\theta} \in \Omega_0$ s.t $\dot{\ell}(\hat{\theta}) = 0$

proof) $\Omega_0 = (-\infty, +\infty)$

Since $\lim_{\theta \to -\infty} \ell(\theta) = -\infty$, $\lim_{\theta \to \infty} \ell(\theta) = -\infty$ and $\ell(\theta)$: continuous,

$\ell(\hat{\theta}) = \max_{\theta \in \Omega_0} \ell(\theta)$, $\hat{\theta} \in \Omega_0$. $(*)$

If $\hat{\theta}_1 \neq \hat{\theta}_2$ and $\hat{\theta}_1 = \operatorname{argmax}_{\theta \in \Omega_0} \ell(\theta)$, $\hat{\theta}_2 = \operatorname{argmax}_{\theta \in \Omega_0} \ell(\theta)$

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

$\ell(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2) > \frac{1}{2}\ell(\hat{\theta}_1) + \frac{1}{2}\ell(\hat{\theta}_2) = \max_{\theta \in \Omega_0} \ell(\theta)$

It contradicts to that $\ell(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2) > \ell(\hat{\theta}_1) = \max_{\theta \in \Omega_0} \ell(\theta)$.

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

Since $\dot{\ell}(\hat{\theta}) = 0$, $\hat{\theta}$ is a unique solaition of $\operatorname{argmax}_{\theta \in \Omega_0} \ell(\theta)$ Q.

e.g. 6.2.3

$X_1, \dots, X_n \overset{iid}{\sim} L(\theta, 1) \quad (-\infty < \theta < +\infty)$

$$pdf_{X_1}(x;\theta) = \frac{e^{x-\theta}}{(1+e^{x-\theta})^2} = \frac{\cancel{e^{x-\theta}}}{(e^{-x+\theta}+1)^2 (e^{x-\theta})^{\cancel{2}}} = \frac{e^{-x+\theta}}{(e^{-x+\theta}+1)^2}$$

$-\infty < x < +\infty.$

$$\ell(\theta) = \sum_{i=1}^{n} (-x_i + \theta) - 2\log(1 + e^{-x_i+\theta})$$

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

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

$$\ddot{\ell}(\theta) = -2\sum_{i=1}^{n} \frac{e^{-x_i+\theta}(1+e^{-x_i+\theta}) - e^{-x_i+\theta} \cdot e^{-x_i+\theta}}{(1+e^{-x_i+\theta})^2}$$

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

$\therefore \ddot{\ell}(\theta) < 0 \quad \forall \theta \in (-\infty, \infty)$

$\lim_{\theta \to -\infty} \ell(\theta) = -\infty \qquad \lim_{\theta \to \infty} \ell(\theta) = -\infty$

$\left(\because \lim_{\theta \to \infty} -\log(1+e^{-x_i+\theta}) < \lim_{\theta \to \infty} -\log(e^{-x_i+\theta}) = \lim_{\theta \to \infty} (x_i - \theta) = -\infty \right)$

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

$\square$

Thm 6.2.3.

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

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

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

$L(g^{-1}(\eta); x_1, \dots, x_n) = pdf_X(x_1, \dots, x_n; g^{-1}(\eta))$

$= pdf_X(x_1, \dots, x_n; \theta)$

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

$= L(g^{-1}(g(\hat{\theta}^{MLE})))$

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

$\square$

e.g. 6.2.4.

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

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

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

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

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

$\lambda := 1/\theta$

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

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

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

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

$\therefore \hat{\theta}^{MLE} = \bar{X}$

$\square$

Statistics 200822

Thm. 6.2.4. Exponential family

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

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

(i) $\mathcal{X} = \{x : pdf(x;\eta) > 0\}$: 모수에 의존x

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

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

$\Rightarrow$ If $\dot{A}(\eta) = \overline{T(x)}$ or $E_\eta[T(X_1)] = \overline{T(x)}$: equation w.r.t η has a solution $\hat{\eta} = \hat{\eta}(x_1, \dots, x_n)$ then

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

FACT

① 실수공간에서 open set은 open interval의 arbitrary union이다.

② O . open set in $\mathbb{R}$, $x \in O$

$\Rightarrow \exists \epsilon_0 > 0$ s.t $x \in (x-\epsilon_0, x+\epsilon_0) \in O$

proof)

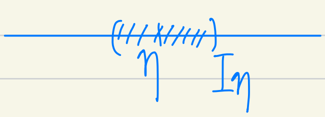
$E[\exp(sT(X_1))] = \int_{-\infty}^{+\infty} \exp(sT(x)) \exp(\eta T(x) - A(\eta) + S(x))dx$

$mgf_{T(X_1);\eta}(s) = \int_{-\infty}^{+\infty} \exp((s+\eta)T(x) - A(s+\eta) + S(x))dx \ \exp(A(s+\eta) - A(\eta))$

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

$\because$ Since $s+\eta \in N$, $\exp\{(s+\eta)T(x) - A(s+\eta) + S(x)\}$: pdf. The integral over $(-\infty, +\infty)$ is 1.

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



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

$|s| < \epsilon$, i.e., $-\epsilon < s < \epsilon$

$\Rightarrow \eta - \epsilon < s+\eta < \eta + \epsilon$

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

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

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

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

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

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

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

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

$\ddot{\bar{\ell}}(\eta) = -\ddot{A}(\eta) = -Var_\eta(T(X_1)) < 0 \Rightarrow \bar{\ell}(\eta)$: concave.

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

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

Rmk. $\eta = g(\theta)$, $\theta \in \Omega$, g: 1-1 corr.

$\text{pdf}(x;\theta) = \exp\{g(\theta)T(x) - A(g(\theta)) + S(x)\}$

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

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

↳ θ, η와 상관없이 분포는 같으므로 mgf, cgf 모두 같다.

e.g. 6.2.5. (a) Bernoulli(p) $0<p<1$

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

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

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

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

$\eta := \log\frac{p}{1-p}$ $\quad e^\eta = p/(1-p)$

$e^\eta(1-p) = p \quad e^\eta = p(1+e^\eta) \quad p = e^\eta/(1+e^\eta)$

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

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

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

$\hat{p}^{MLE} = \bar{X}$ □

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

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

(b) Poisson(λ) $\lambda > 0$

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

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

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

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

$\dot{A}(\eta) = \overline{T(x)}$

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

$\therefore \hat{\lambda}^{MLE} = \bar{X}$

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

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

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

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

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

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

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

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

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

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

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

$\therefore \hat{p}^{MLE} = 1/\bar{X}$ $\square$ $\eta = \log(1-1/\bar{x})$

(d) $\mathrm{Exp}(\theta)$ $\theta > 0$.

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

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

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

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

$$\therefore \hat{\theta}^{MLE} = \bar{X}$$

$\square$

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

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

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

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

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

$$+\overline{\log x} = +\frac{1}{\theta}$$

$$\hat{\theta}^{MLE} = 1/\overline{\log x}$$

$\square$

e.g. 6.2.6.

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

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

$$pdf(x;\theta) = \frac{1}{2}e^{-|x-\theta|} \qquad -\infty<x<+\infty$$

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

$x_{(1)}<\cdots<x_{(n)} \qquad x_{(0)}:=-\infty \qquad x_{(n+1)}:=+\infty$

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

$$\ell(\theta) = -\sum_{i=1}^{n}|x_i-\theta| - n\log 2$$
$$= -\sum_{i=1}^{r}(\theta-x_{(i)}) - \sum_{i=r+1}^{n}(x_{(i)}-\theta) - n\log 2$$

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

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

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

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

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

e.g. 6.2.7

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

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

$$pdf(x_1,\dots,x_n;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(0\le x_i \le \theta)$$

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

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

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

e.g. 6.2.8.

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

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

$$pdf(x_1,\dots,x_n;\theta) = \prod_{i=1}^{n} \frac{1}{2} I(\theta-1\le x_i \le \theta+1)$$

$$= 2^{-n} I(\theta-1 \le \min x_i , \max x_i \le \theta+1)$$

$$L(\theta) = 2^{-n} I(\max x_i - 1 \le \theta \le \min x_i + 1)$$

$\hat{\theta} := \underset{\theta}{\arg\max}\, L(\theta)$ $\qquad$ $\hat{\theta} \in [x_{(n)}-1, x_{(1)}-1]$

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

$\therefore$ $X_{(n)}-1 \le \hat{\theta}^{MLE} \le X_{(1)}-1$ $\square$

6.3. MLE for multiple parameters.

$X_1, \dots, X_n \overset{iid}{\sim} f(x;\theta)$, $\theta = (\theta_1, \dots, \theta_k)' \in \Omega$

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

likelihood and log likelihood are defined as follows

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

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

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

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

e.g. 6.3.1

$X_1, \dots, X_n \overset{iid}{\sim} DE(\mu, \sigma)$ $(-\infty < \mu < +\infty, \sigma > 0)$

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

$\hat{\theta}^{MLE}$?

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

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

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

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

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

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

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

$$\therefore \hat{\theta}^{MLE} = \left(X_{(m+1)}, \frac{1}{n}\sum_{i=1}^{n}|X_i - X_{(m+1)}|\right)$$

$\square$

e.g. 6.3.2

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

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

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

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

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

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

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

$$\frac{d}{d\theta_2}\ell(\hat{\mu}, \theta_2) = \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{n}{2\theta_2} > 0$$

$$\Leftrightarrow \theta_2 < \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$\therefore \max_{\theta_2} \ell(\hat{\mu}, \theta_2) = \ell(\hat{\mu}, \hat{\sigma}^2)$ $\quad \hat{\sigma}_2 = \sum_{i=1}^{n}(x_i-\bar{x})^2$

$$\therefore \hat{\theta}^{MLE} = \left(\frac{1}{n}\sum_{i=1}^{n}X_i,\ \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)$$

$\square$

Thm 6.3.1	$\exists\ \hat{\theta}^{MLE} \in \Omega$ & $\eta = (\eta_1, \eta_2) = (g_1(\theta), g_2(\theta))$ $= g(\theta)$ g: bijective $\Rightarrow \hat{\eta}_1^{MLE} = g_1(\hat{\theta}^{MLE})$
proof)	By thm 6.2.3 $\hat{\eta}^{MLE} = g(\hat{\theta}^{MLE}) = (g_1(\hat{\theta}^{MLE}), g_2(\hat{\theta}^{MLE}))$ By the definition of MLE, $\hat{\eta}_1^{MLE} = g_1(\hat{\theta}^{MLE})$ $\square$

e.g. 6.3.3.

(a)

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

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

$$\eta = (g_1(\theta), g_2(\theta), g_3(\theta))$$
$$= (|\mu|, \mathrm{sgn}(\mu), \sigma^2)$$

Since g: bijective,

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

$$\therefore \hat{\eta}_1^{MLE} = |\hat{\mu}^{MLE}| = \left|\frac{1}{n}\sum_{i=1}^{n} X_i\right|$$

$\square$

(b) $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p) \quad 0 \le p \le 1$

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

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

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

Since g: bijective,

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

Thm 6.3.2

open interval hyper cube $(a_1, b_1) \times \cdots \times (a_k, b_k) = \Omega_0$

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

$$c^t \ddot{\ell}(\theta) c < 0 \quad \forall c \neq 0, \ \forall \theta \in \Omega_0 \ \&$$

$$\dot{\ell}(\hat{\theta}) = 0, \ \hat{\theta} \in \Omega_0$$

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

proof)

$$\ell(\theta) = \ell(\hat{\theta}) + \dot{\ell}(\hat{\theta}) \cdot (\theta - \hat{\theta}) + \frac{1}{2}(\theta - \hat{\theta})^t \ddot{\ell}(\theta^*)(\theta - \hat{\theta})^t < \ell(\hat{\theta})$$

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

$\square$

Thm 6.3.3. open interval hypercube $(a_1, b_1) \times \cdots \times (a_k, b_k) = \Omega_0$

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

$$c^t \ddot{l}(\theta) c < 0 \quad \forall c \neq 0, \ \forall \theta \in \Omega_0 \ \& \ \lim_{\theta \to \partial(\Omega_0)} l(\theta) = -\infty$$

$$\Rightarrow \exists! \hat{\theta} \text{ s.t } \dot{l}(\hat{\theta}) = 0 \quad \hat{\theta} \in \Omega_0.$$

Since $\lim_{\theta \to \partial(\Omega_0)} l(\theta) = -\infty$, $\exists \hat{\theta}$ s.t

$$l(\hat{\theta}) = \max_{\theta} l(\theta) \quad \hat{\theta} \in \Omega_0$$

Suppose $\hat{\theta}_1 < \hat{\theta}_2$ & $\max_{\theta \in \Omega_0} l(\theta) = l(\hat{\theta}_1) = l(\hat{\theta}_2)$

Since $l(\theta)$: concave.

$$l\left(\frac{1}{2}\hat{\theta}_1 + \frac{1}{2}\hat{\theta}_2\right) > \frac{1}{2} l(\theta_1) + \frac{1}{2} l(\theta_2)$$

$$= \max_{\theta \in \Omega_0} l(\theta).$$

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

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

$$pdf(x;\eta) = \exp\left\{\sum_{j=1}^{k} \eta_j T_j(x) - A(\eta) + S(x)\right\} \quad x \in \mathcal{X},$$
$$\eta = (\eta_1, \dots, \eta_k) \in N$$

Thm 6.3.4

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

(ii) $N = (a_1, b_1) \times \cdots \times (a_k, b_k)$ open interval hyper-cube

(iii) $Var(c' T(X_1) c) > 0 \quad \forall c \in \mathbb{R}^k, \ c \neq 0$

$\Rightarrow$ If $\dot{A}(\eta) = \overline{T(x)}$ or $E_\eta[T(X_1)] = \overline{T(x)}$ has a solution $\hat{\eta} = \hat{\eta}(x_1, \dots, x_n)$, then $\hat{\eta} = \hat{\eta}^{MLE}$.

proof) $mgf_{T(X);\eta}(s) = E[\exp(\sum_{j=1}^{k} s_j T_j(X_1))]$

$T(X_1) = (T_1(X_1), \dots, T_k(X_1))'$ $\quad s = (s_1, \dots, s_k)'$

Since $X_i \overset{iid}{\sim} f(x;\eta)$, $\perp\!\!\!\perp_i T_i(X_i)$ and $T_i(X_i) = T_i(X_1)$ $i = 1, \dots, k$.

Thus, $mgf_{T(X_1)}(s) = \prod_{j=1}^{k} mgf_{T_j(X_1)}(s_j)$

$$= \prod_{j=1}^{k} mgf_{T_j(X_1)}(s_j)$$

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

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

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

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

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

$\because s + \eta \in N$ & $\int_{-\infty}^{+\infty} pdf(x;\eta)\, dx = 1$

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

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

$$Var_\eta(T(X_1)) = \ddot{A}(\eta)$$

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

$$\dot{\bar{\ell}}(\eta) = (\overline{T_1(x)}, \ldots, \overline{T_k(x)})' - \dot{A}(\eta)$$
$$= \overline{T(x)} - \dot{A}(\eta)$$

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

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

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

$\square$

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

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

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

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

e.g. 6.3.4.

$$f(x;\theta) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \quad -\infty < x < +\infty$$

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

$$= \exp\left\{-\frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\}$$

$$\eta_2 := \frac{-1}{2\sigma^2} \quad \eta_1 = \frac{\mu}{\sigma^2} \qquad \frac{\mu^2}{2\sigma^2} = \frac{\eta_1^2}{4\eta_2^2} \times (-\eta_2)$$

$$\sigma^2 = -\frac{1}{2\eta_2} \quad \mu = \sigma^2\eta_1 = -\frac{\eta_1}{2\eta_2} \qquad = -\frac{\eta_1^2}{4\eta_2}$$

$$= \exp\left\{\eta_1 x + \eta_2 x^2 + \frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log\left(-\frac{\pi}{\eta_2}\right)\right\}$$

$$A(\eta) := -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log\left(-\frac{\eta_2}{\pi}\right)$$

$$\dot{A}(\eta) = (\overline{x}, \overline{x^2})$$

$$= \left(-\frac{1}{2}\times\frac{\eta_1}{\eta_2}, \quad +\frac{\eta_1^2}{4\eta_2^2} + \frac{1}{2}\times\frac{\pi}{\eta_2}\times\left(-\frac{1}{\pi}\right)\right)$$

$$\overline{x} = -\frac{1}{2}\times(-2\hat{\mu}) = \hat{\mu}$$

$$\overline{x^2} = (\hat{\mu})^2 - \frac{1}{2}\times(-2\hat{\sigma}^2) = \hat{\sigma}^2 + (\hat{\mu})^2$$

$$\therefore \hat{\mu}^{MLE} = \overline{X}, \quad \hat{\sigma}^{2\,MLE} = \overline{X^2} - \overline{X}$$

$\square$

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

$E[X_1$ $\quad (x\ y)\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}} \qquad Cov(X_1, X_2) = \sigma_1\sigma_2\rho \qquad ax^2 + (b+c)xy + dy^2$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2$$

$$= \sigma_1^2\sigma_2^2(1-\rho^2)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

$$(\hat{\mu}_2)^2 + \sigma_2^2 = \overline{y^2} \qquad \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2 + \mu_1\mu_2 = \overline{xy}$$

$$\therefore \hat{\rho}^{MLE} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sqrt{\overline{X^2} - (\bar{X})^2}\sqrt{\overline{Y^2} - (\bar{Y})^2}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

$\square$

20.08.29.

e.g. 6.4.1

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

$$\hat{p}^{MLE} = \overline{X} = \hat{p}^{MME}$$

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

(b) Poisson(λ), $\lambda \ge 0$

$$\hat{\lambda}^{MLE} = \overline{X} = \hat{\lambda}^{MME}$$

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

(c) Exp(θ) $\theta > 0$ $\quad \hat{\theta}^{MLE} = \overline{X}$

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

$$\hat{\mu}^{MLE} = \overline{X} \qquad \hat{\sigma}^{2\,MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

최대가능도 추정량이 일치성을 가지려면?

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

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

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

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

$$\hat{\theta}_n = \underset{\theta \in \Omega}{\operatorname{argmax}}\, l_n(\theta) = \underset{\theta \in \Omega}{\operatorname{argmax}}\, \bar{l}_n(\theta)$$

모수의 참값이 θ^0라하면   대수의 법칙에 의해

$$\bar{l}_n(\theta) \xrightarrow[n \to \infty]{P_{\theta^0}} E_{\theta^0}[\log f(X_1 ; \theta)]$$

argmax $\Downarrow$ $\qquad\qquad$ $\Downarrow$

$$\hat{\theta}_n \xrightarrow[n \to \infty]{P_{\theta^0}} \underset{\theta \in \Omega}{\operatorname{argmax}}\, E_{\theta^0}[\log f(X_1 ; \theta)] \overset{?}{=} \theta^0$$

Q1. $\underset{\theta \in \Omega}{\operatorname{argmax}}\, E_{\theta^0}[\log f(X_1 ; \theta)] = \theta^0$?

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

Thm 6.4.1

KL-divergence

$X \sim f(x;\theta) \quad \theta \in \Omega$

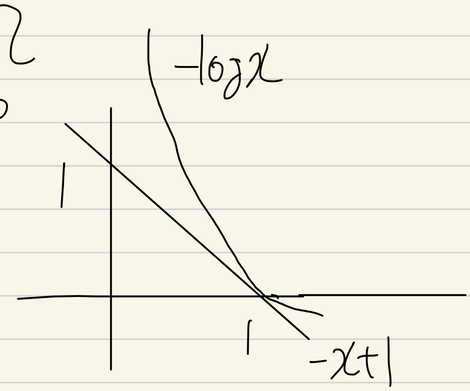
(R0) (identifiable) $f(\cdot;\theta) = f(\cdot;\theta^0) \Rightarrow \theta = \theta^0$

(R1) (Common Support) $\{x : f(x;\theta) > 0\}$ $\theta \in \Omega$에 의존X.

$$KL(\theta, \theta^0) := -E_{\theta^0}\left[\log \frac{f(X;\theta)}{f(X;\theta^0)}\right]$$

(i) $KL(\theta, \theta^0) \geq 0, \quad \forall \theta \in \Omega$

(ii) $KL(\theta, \theta^0) = 0 \iff \theta = \theta^0$



proof) $\forall t \in \mathbb{R}, \quad -\log t \geq -t+1$

$$-\log \frac{f(X;\theta)}{f(X;\theta^0)} \geq -\frac{f(X;\theta)}{f(X;\theta^0)} + 1$$

∵ monotonicity of expectation.

$$KL(\theta, \theta^0) = -E_{\theta^0}\left[-\log \frac{f(X;\theta)}{f(X;\theta^0)}\right] \geq E_\theta\left[-\frac{f(X;\theta)}{f(X;\theta^0)} + 1\right]$$

$$E_\theta\left[-\frac{f(X;\theta)}{f(X;\theta^0)} + 1\right] = -\int_{-\infty}^{+\infty} \frac{f(x;\theta)}{\cancel{f(x;\theta^0)}} \cancel{f(x;\theta^0)}\, dx + 1$$

$$= -1 + 1 = 0$$

$\therefore KL(\theta, \theta^0) \geq 0$. equality holds when $t=1$. i.e., $\frac{f(X;\theta)}{f(X;\theta^0)} = 1$.

By identifiability, $f(\cdot, \theta) = f(\cdot, \theta^0) \Rightarrow \theta = \theta^0$

$\therefore KL(\theta, \theta^0) = 0$ if and only if $\theta = \theta^0$ $\square$

To answer Q_1,

$$\underset{\theta\in\Omega}{\operatorname{argmax}} E_{\theta^0}[\log f(X_1;\theta)] = \underset{\theta\in\Omega}{\operatorname{argmin}} E_{\theta^0}[-\log f(X_1;\theta)]$$

$$= \underset{\theta\in\Omega}{\operatorname{argmin}} E_{\theta^0}\left[-\log \frac{f(X_1;\theta)}{f(X_1;\theta^0)}\right]$$

$$= \underset{\theta\in\Omega}{\operatorname{argmin}} KL(\theta,\theta^0)$$

$\because \min_{\theta\in\Omega} KL(\theta,\theta^0) = 0$

$\Leftrightarrow \theta = \theta^0.$

$$= \theta^0$$

To answer Q_2,

We assume

i) uniform convergence in probability

$$\sup_{|\theta-\theta^0|\le K} \left| \bar{l}_n(\theta) - E_{\theta^0}[\log f(X_1;\theta)] \right| \xrightarrow[n\to\infty]{P_{\theta^0}} 0, \quad \forall K>0.$$

ii) Continuity of Expectation of log likelihood

$E_{\theta^0}[\log f(X;\theta)]$: continuous w.r.t θ.

FACT) iii) implies i)

iii) Concave of log likelihood

$c^t \ddot{l}_n(\theta) c < 0 \quad \forall c\neq 0,\ \forall \theta\in\Omega \quad \lim_{\theta\to\partial(\Omega)} l_n(\theta) = -\infty.$

Thm 6.4.2. $f(\cdot,\theta) = f(\cdot,\theta^0) \Rightarrow \theta=\theta^0$, $\mathcal{X} = \{x : f(x;\theta)>0\}$: independent of θ,

$c^t \ddot{l}_n(\theta) c < 0 \quad \forall c\neq 0,\ \forall \theta\in\Omega \quad \lim_{\theta\to\partial(\Omega)} l_n(\theta) = -\infty$

$\forall \theta^0\in\Omega$, $\exists E_{\theta^0}[\log f(X_1;\theta)]$ & continuous w.r.t $\theta\in\Omega$

$$\Rightarrow \exists!\ \hat{\theta}_n^{MLE} \text{ s.t } \dot{l}_n(\hat{\theta}_n^{MLE}) = 0 \ \&\ \hat{\theta}_n^{MLE} \xrightarrow[n\to\infty]{P_\theta} \theta, \quad \forall\theta\in\Omega$$

Asymptotic distribution of maximum likelihood

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

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

$$0 = \dot{\bar{\ell}}_n(\hat{\theta}_n) \simeq \dot{\bar{\ell}}_n(\theta) + \ddot{\bar{\ell}}_n(\theta)(\hat{\theta}_n - \theta)$$

$$\therefore \sqrt{n}(\hat{\theta}_n - \theta) \simeq -\left(\ddot{\bar{\ell}}_n(\theta)\right)^{-1}\sqrt{n}\,\dot{\bar{\ell}}_n(\theta)$$

By the law of large number,

$$-\ddot{\bar{\ell}}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta)\right] \xrightarrow[n\to\infty]{P_\theta} E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta)\right]$$

Let $I(\theta) := E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta)\right]$

$$\sqrt{n}(\hat{\theta}_n - \theta) \simeq \left(I(\theta)\right)^{-1}\sqrt{n}\,\dot{\bar{\ell}}_n(\theta)$$

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

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\theta) - E_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]\right) \xrightarrow[n\to\infty]{d} N\left(0, Var_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]\right)$$

$$E_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right] = \int_{-\infty}^{+\infty}\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)\,dx$$

under "some" condition.

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

$$\therefore \sqrt{n}\cdot\dot{\bar{\ell}}_n(\theta) \xrightarrow[n\to\infty]{d} N\left(0, Var_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]\right)$$

$$\therefore \sqrt{n}(\hat{\theta}_n - \theta) \simeq \left(I(\theta)\right)^{-1}\sqrt{n}\,\dot{\bar{\ell}}_n(\theta) \xrightarrow[n\to\infty]{d} N\left(0, [I(\theta)]^{-2}\,Var_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]\right)$$

"Asymptotic Normality"

최대가능도 추정량의 점근 정규성을 위한 기본 조건:

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

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

($R2$) (열린 모수공간) 모수공간 Ω는 k 차원 공간 R^k에서의 열린집합이다.

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

($R4$) (적분 또는 합과 미분의 순서 교환 가능) 랜덤표본 $\mathrm{X}=(\mathrm{X}_1,\cdots,\mathrm{X}_n)^t$의 함수 $u(\mathrm{X})$의 기댓값 $\mathrm{E}_\theta[u(\mathrm{X})]$이 존재할 때, 모수 θ에 대한 미분[100]을 ∂_θ, ∂_θ^2으로 나타내면

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

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

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

$$I(\theta)=\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(\mathrm{X}_1;\theta)\right)$$

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

Thm 6.4.3. If $(R_0)\sim(R_5)$ holds

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

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

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

$$\Rightarrow\ 0=\frac{\partial}{\partial\theta}\int_{\mathscr{X}}\exp(\log f(x;\theta))\,dx$$
$$=\int_{\mathscr{X}}\frac{\partial}{\partial\theta}\exp(\log f(x;\theta))\,dx$$
$$=\int_{\mathscr{X}}\exp(\log f(x;\theta))\frac{\partial}{\partial\theta}\log f(x;\theta)\,dx=E_\theta\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]$$

$$\Rightarrow\ 0=\frac{\partial}{\partial\theta}\int_{\mathscr{X}}\exp(\log f(x;\theta))\frac{\partial}{\partial\theta}\log f(x;\theta)\,dx$$
$$=\int_{\mathscr{X}}\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)+\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]\exp(\log f(x;\theta))\,dx$$
$$=E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta)+\left(\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right)^2\right]$$

$\therefore\ E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right)^2\right]=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta)\right]$

$\therefore\ Var_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]=E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f(X_1;\theta)\right]$ □

Thm 6.4.4

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

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

의 단 하나뿐인 근이고 일치성을 갖는다. 즉

$$\hat{\theta}_n^{MLE} \underset{n\to\infty}{\overset{P_\theta}{\longrightarrow}} \theta$$

($R7$) (잉여항의 크기) 로그가능도함수의 삼차 편도함수[101]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.

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N(0, [I(\theta)]^{-1})$$

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \underset{n\to\infty}{\overset{d}{\longrightarrow}} N(0, (I(\theta))^{-1})$$

proof) step1.

$$0 = \overline{\dot{\ell}_n(\hat{\theta}_n)} = \overline{\dot{\ell}_n(\theta)} + \overline{\ddot{\ell}_n(\theta)}(\hat{\theta}_n - \theta) + \frac{1}{2}\overline{\ell_n^{(3)}}(\theta_n^*)(\hat{\theta}_n - \theta)^2$$

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

Let $R_n := \frac{1}{2}\overline{\ell_n^{(3)}}(\theta_n^*)(\hat{\theta}_n - \theta)^2$

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

$\forall \epsilon, K > 0$

$$P_\theta(|R_n| \ge \epsilon) = P(|\overline{\ell_n^{(3)}}(\theta_n^*)| > K, |R_n| \ge \epsilon)$$
$$+ P(|\overline{\ell_n^{(3)}}(\theta_n^*)| \le K, |R_n| \ge \epsilon)$$

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

$\because |\hat{\theta}_n - \theta| << 1$

$$\le \frac{1}{2}|\hat{\theta}_n - \theta|^2 K$$
$$\le \frac{1}{2}|\hat{\theta}_n - \theta| K$$

$$P_\theta(|R_n| \ge \epsilon) \le P_\theta(|\overline{\ell_n^{(3)}}(\theta_n^*)| > K) + P_\theta(|\hat{\theta}_n - \theta| \ge 2\epsilon/K)$$

$$P_\theta(|\overline{\ell_n^{(3)}}(\theta_n^*)| > K) = P_\theta\left(\left|\frac{1}{n}\sum_{i=1}^n \partial_\theta^3 \log f(X_i;\theta)\right| > K\right)$$
$$\le P_\theta\left(\frac{1}{n}\sum_{i=1}^n |\partial_\theta^3 \log f(X_i;\theta)| > K\right)$$
$$\le P_\theta\left(\frac{1}{n}\sum_{i=1}^n \max_{\theta\in\Omega} |\partial_\theta^3 \log f(X_i;\theta)| > K\right)$$
$$\le P_\theta\left(\frac{1}{n}\sum_{i=1}^n M(X_i) > K\right)$$

$\because$ Markov's Inequality

$$\le E_\theta\left[\frac{1}{n}\sum_{i=1}^n M(X_i)\right]/K$$
$$= E_\theta[M(X_1)]/K$$

$$P_\theta(|R_n| \ge \epsilon) = P_\theta\left(|\overline{\ell_n^{(3)}}(\theta_n^*)| > K\right) + P_\theta\left(|\hat{\theta}_n - \theta| \ge \frac{2\epsilon}{K}\right)$$

$$\le E_\theta[M(X_1)]/K + P_\theta\left(|\hat{\theta}_n - \theta| \ge \frac{2\epsilon}{K}\right)$$

Since $\operatorname{plim}_{n\to\infty} \hat{\theta}_n = \theta$, $\lim_{n\to\infty} P_\theta(|R_n| \ge \epsilon) = E_\theta[M(X_1)]/K$

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

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

Step 3. $0 = \bar{\dot{\ell}}_n(\theta) + \left(\bar{\ddot{\ell}}_n(\theta) + \frac{1}{2} \overline{\ell_n^{(3)}}(\theta_n^*)(\hat{\theta}_n - \theta) \right)(\hat{\theta}_n - \theta)$

i) $\sqrt{n}(\hat{\theta}_n - \theta) = \left(-\bar{\ddot{\ell}}_n(\theta) - \frac{1}{2} \overline{\ell_n^{(3)}}(\theta_n^*)(\hat{\theta}_n - \theta) \right) \sqrt{n}\, \bar{\dot{\ell}}_n(\theta)$

by the law of large number $-\bar{\ddot{\ell}}_n(\theta) = \sum_{i=1}^{n} \left(-\frac{\partial^2}{\partial\theta^2} \log f(X_i;\theta)\right) \xrightarrow[n\to\infty]{P_\theta} \underset{\displaystyle I(\theta)}{\underset{\shortparallel}{E_\theta}}\left[-\frac{\partial^2}{\partial\theta^2} \log f(X_i;\theta)\right]$

ii) $R_n \xrightarrow[n\to\infty]{P_\theta} 0$, $-\frac{1}{2} \overline{\ell_n^{(3)}}(\theta_n^*)(\hat{\theta}_n - \theta) =: r_n \xrightarrow[n\to\infty]{P_\theta} 0$.

iii) $\sqrt{n}\, \bar{\dot{\ell}}_n(\theta) = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i;\theta) - \overbrace{E_\theta\left[\frac{\partial}{\partial\theta} \log f(X_1;\theta)\right]}^{=0} \right)$

$$\xrightarrow[n\to\infty]{d} N\left(0, \operatorname{Var}_\theta\left[\frac{\partial}{\partial\theta} \log f(X_1;\theta)\right] = I(\theta)\right)$$

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

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

$\square$

20.09.05

Lemma 1. Let V be an finite dimensional inner product vector space, $W \le V$.

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

proof) Let $\beta := \{v_1, \ldots, v_k\}$ an orthonormal basis for W.

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

By the definition, $w \in W$.

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

$$\begin{aligned} \langle u, v_j \rangle &= \langle y - \sum_{i=1}^{k} \langle y, v_i \rangle v_i , v_j \rangle \\ &= \langle y, v_j \rangle - \sum_{i=1}^{k} \langle y, v_i \rangle \langle v_i, v_j \rangle \\ &= \langle y, v_j \rangle - \sum_{i=1}^{k} \langle y, v_i \rangle \delta_{ij} \\ &= 0. \end{aligned}$$

Thus $u \in W^{\perp}$.

To show the uniqueness of w, u,

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

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

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

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

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

$\square$

Corollary. $\forall y \in V$, $\exists!\, w \in W$ s.t $\|y - x\| \geq \|y - w\|$ $\forall x \in W$.

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

$$\|y-x\|^2 = \|w+u-x\|^2 = \|w-x+u\|^2 \quad (\because w-x \perp u)$$
$$= \|w-x\|^2 + \|u\|^2$$
$$\geq \|u\|^2 = \|y-w\|$$

Suppose $\|y-x\| = \|y-w\|$

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

Thus $x = w$. $\square$

Lemma 2. $A \in M_{m\times n}(F)$ $\quad x \in F^n$ $\quad y \in F^m$

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

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

Lemma 3. Let $A \in M_{m\times n}(F)$ Then $\operatorname{rank}(A^*A) = \operatorname{rank} A$.

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

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

$\langle x, A^*Ax\rangle_n = \langle Ax, Ax\rangle_m$ by the lemma 2.

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

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

$\therefore \operatorname{rank} A^*A = \operatorname{rank} A$. $\square$

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

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 $Ax_0 - y \perp Ax$ $\forall x \in F^n$,

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

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

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

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

Then A^*A: invertible.

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

$\square$

09.05. Statistics.

Least Square Estimation.

$Y_i = x_{i0}\beta_0 + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + e_i$

$X = \begin{bmatrix} x_{10} & \cdots & x_{1p} \\ \vdots & \ddots & \\ x_{n0} & \cdots & x_{np} \end{bmatrix} \in M_{n\times(p+1)}(\mathbb{R})$ $Y = (Y_1, \dots, Y_n)'$ $e := (e_1, \dots, e_n)$

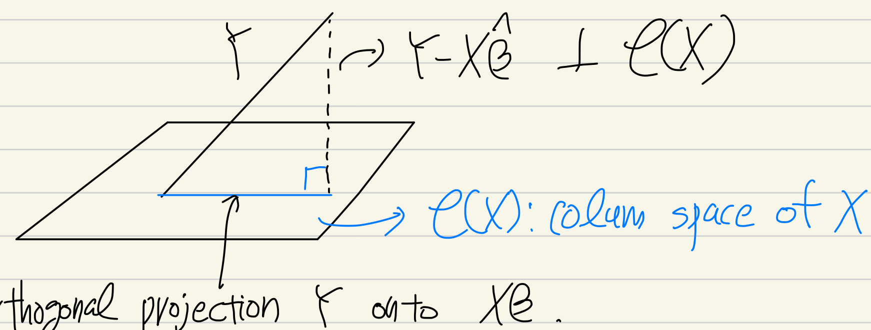
$\beta := (\beta_0, \dots, \beta_p)$

$E[e] = 0$, $Var(e) = \sigma^2 I_n$. $0 < \sigma^2 < +\infty$.

$\operatorname{rank} X = p+1$, i.e., full column rank.

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

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



We know that the solution of normal equation is

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

$X'X\beta = X'Y$

$\hat{\beta} = (X'X)^{-1}X'Y$ ($\because$ Since X: full rank and $\operatorname{rank} X'X = \operatorname{rank} X$ by the lemma 2, $X'X$: invertible.

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

Thm 6.5.1

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

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

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

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

proof)

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

$(X(X'X)^{-1}X')'(I - X(X'X)^{-1}X')$

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

$= X(X'X)^{-1}X' - X(X'X)^{-1}X' = 0.$

or $\Pi' = \Pi$: projection onto $\mathcal{C}(X)$

$(I-\Pi)$: projection onto $\mathcal{C}(X)^{\perp}$.

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

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

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

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

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

$$+ 2\|(\Pi(Y - X\beta))'(I - \Pi)Y\|^2$$

$$= (Y - X\beta)'\underbrace{\Pi'(I - \Pi)}_{=0}Y$$

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

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

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

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

$\square$

Thm 6.5.2.

$Y = X\beta + e \qquad X \in M_{n\times(p+1)}(\mathbb{R})$

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

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

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

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

(c) If $e \sim N_n(0, \sigma^2 I_n) \Rightarrow \hat{\beta}^{LSE} \sim N(\beta, \sigma^2 (X'X)^{-1})$

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

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

$E[\hat{\beta}^{LSE}] = (X'X)^{-1} X' E[Y]$, $\quad E[Y] = E[X\beta + e]$

$= E[e] + E[X\beta]$

$= X\beta$

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

$= \beta$.

$Var(\hat{\beta}^{LSE}) = Var((X'X)^{-1} X' Y) = (X'X)^{-1} X' \, Var(Y) \, X (X'X)^{-1}$

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

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

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

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

$\|Y - X\hat{\beta}^{LSE}\|^2 = \|\Pi(Y - X\hat{\beta}^{LSE})\|^2 + \|(I-\Pi)Y\|^2$

$= \|(I-\Pi)Y\|^2 \qquad Y = X\beta + e$

$= \|(I-\Pi)e\|^2$

$= e'(I-\Pi)'(I-\Pi)e$

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

$= tr(e'(I-\Pi)e)$

$= tr((I-\Pi)ee')$

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

$= E[tr((I-\Pi)ee')]$

$= tr(E[(I-\Pi)ee'])$

$= tr((I-\Pi)E[ee']) \quad \because Var(e) = E[(e-0)(e-0)'] = E[ee']$

$= tr((I-\Pi)Var(e))$

$= tr(\sigma^2(I-\Pi))$

$= \sigma^2 tr(I-\Pi)$

$= \sigma^2(n - tr(X(X'X)^{-1}X'))$

$= \sigma^2(n - tr(X'X(X'X)^{-1}))$

$= \sigma^2(n - tr\, I_{(p+1)}) = \sigma^2(n-p-1).$

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

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

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

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

$\stackrel{d}{\equiv} N(\beta, \sigma^2(X'X)^{-1})$

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

$(I-\Pi)$: projection onto $\mathcal{C}(X)^\perp$

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

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

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

$(n-p-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(\mathrm{tr}(I-\Pi)) \stackrel{d}{\equiv} \chi^2(n-p-1)$

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

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

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

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

$Y - X\hat{\beta}$

$= \sigma^2 \, 0 = 0.$

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

$\therefore \hat{\beta} \perp\!\!\!\perp \hat{\sigma}^2 = (Y-X\hat{\beta})'(Y-X\hat{\beta}) = ((I-\Pi)Y)'((I-\Pi)Y)$

$= g((I-\Pi)Y)$

$\square$

orthogonalization of explanatory variables

$X \in M_{n\times(p+1)}(\mathbb{R})$, $\operatorname{rank} X = p+1$ $\qquad p_0 + p_1 = p+1$

$X = (X_0, X_1)$, $X_0 \in M_{n\times p_0}(\mathbb{R})$, $X_1 \in M_{n\times p_1}(\mathbb{R})$

$\Pi_0 = X_0(X_0'X_0)^{-1}X_0'$

$$\Rightarrow X\beta = (X_0, X_1)\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} = X_0\beta_0 + X_1\beta_1$$

$$= X_0\beta_0 + \Pi_0 X_1\beta_1 + (I-\Pi_0)X_1\beta_1$$

$$= X_0(\beta_0 + (X_0'X_0)^{-1}X_0'X_1\beta_1) + \underbrace{(I-\Pi_0)X_1}_{=:X_{1|0}}\beta_1$$

$$= \underbrace{X_0\gamma_0}_{\in \mathcal{C}(X_0)} + \underbrace{X_{1|0}\beta_1}_{\in \mathcal{C}(X_0)^{\perp}}$$

$\Pi_{0,1} = X(X'X)^{-1}X'$, $\Pi_0 = X_0(X_0'X_0)^{-1}X_0'$, $\Pi_{1|0} = X_{1|0}(X_{1|0}'X_{1|0})^{-1}X_{1|0}'$

$\Rightarrow \Pi_{0,1} = \Pi_0 + \Pi_{1|0} \qquad \Pi_0'\Pi_{1|0} = 0$

(a) $\mathcal{C}(X_0, X_1) = \mathcal{C}(X_0, X_{1|0})$

(b) $\mathcal{C}(X_0, X_1) = \mathcal{C}(X_0) \oplus \mathcal{C}(X_0)^{\perp}$, $\quad \mathcal{C}(X_0)^{\perp} = \mathcal{C}(X_{1|0})$

(c) $\Pi_{0,1} = \Pi_0 + \Pi_{1|0}$, $\quad \Pi_0\Pi_{1|0} = 0$

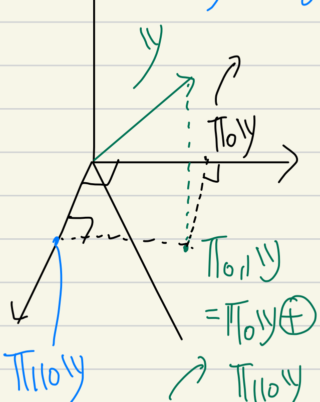
proof) (a) $\mathcal{C}(X_0, X_1) = \{X_0\beta_0 + X_1\beta_1 : \beta_0 \in \mathbb{R}^{p_0} + \beta_1 \in \mathbb{R}^{p_1}\}$

$= \{X_0\gamma_0 + X_{1|0}\beta_1 : \gamma_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}\}$

Π_0 is the orthogonal projection on $\mathcal{C}(X_0)$ along with $\mathcal{C}(X_0)^\perp = \mathcal{C}(X_{1|0})$

$= \mathcal{C}(X_0, X_{1|0})$

(b) Since Π_0 is projection onto $\mathcal{C}(X_0)$, clearly $\mathcal{C}(\Pi_0)$ is subspace of $\mathcal{C}(X_0)$.



Conversely, let $y \in \mathcal{C}(X_0)$. $\exists\, y^\perp \in \mathcal{C}(X_0)^\perp$

For $y' = y + y^\perp \in \mathbb{R}^n$

$\Pi_0 y' = \Pi_0(y + y^\perp) = \Pi_0 y + \cancelto{0}{\Pi_0 y^\perp} = y$

$\Pi_{1,0}$: orthogonal projection on $\mathcal{C}(X_0, X_1)$ along with $\mathcal{C}(X_0, X_1)^\perp$

So that $y \in \mathcal{C}(\Pi_0)$. $\therefore \mathcal{C}(X_0) = \mathcal{C}(\Pi_0)$

Thus, $\langle \cdot, \cdot \rangle$: dot product on $\mathbb{R}^n$

$\mathcal{C}(X_0)^\perp \cap \mathcal{C}(X_0, X_1) = \{y \in \mathcal{C}(X_0, X_1) : \langle y_0, y \rangle = 0 \quad \forall y_0 \in \mathcal{C}(X_0)\}$

Since $y \perp \mathcal{C}(X_0)$,
$\langle X_0\beta_0, X_0\beta_0 + X_1\beta_1 \rangle = 0$
$\forall \beta_0 \in \mathbb{R}^{p_0}$
i.e., $(X_0\beta_0 + X_1\beta_1)' X_0\beta_0 = 0$
$\forall \beta_0 \in \mathbb{R}^{p_0}$
$\therefore (X_0\beta_0 + X_1\beta_1)' X_0$: zero map.

$= \{X_0\beta_0 + X_1\beta_1 \in \mathcal{C}(X_0, X_1) : (X_0\beta_0 + X_1\beta_1)' X_0 = 0, \ \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}\}$

$\because \mathcal{C}(X_0) = \mathcal{C}(\Pi_0)$

$= \{X_0\beta_0 + X_1\beta_1 \in \mathcal{C}(X_0, X_1) : (X_0\beta_0 + X_1\beta_1)'\Pi_0 = 0, \beta_0 \in \mathbb{R}^{p_0}, \beta_1 \in \mathbb{R}^{p_1}\}$

$= \{X_0\beta_0 + X_1\beta_1 \in \mathcal{C}(X_0, X_1) : X_0\beta_0 + \Pi_0 X_1\beta_1 = 0, \forall \beta_0, \forall \beta_1\}$

$\because X_1\beta_1 = (I - \Pi_0) X_0\beta_0$

$= \{(I - \Pi) X_1\beta_1 : \beta_1 \in \mathbb{R}^{p_1}\}$

$= \mathcal{C}(X_{1|0})$

$\therefore \mathcal{C}(X_0) \oplus \mathcal{C}(X_0)^\perp = \mathcal{C}(X_0, X_1) = \mathcal{C}(X_0, X_{1|0})$ & $\mathcal{C}(X_{1|0}) = \mathcal{C}(X_0)^\perp$

$\therefore \mathcal{C}(X_0, X_{1|0}) = \mathcal{C}(X_0) \oplus \mathcal{C}(X_{1|0})$

$\therefore \Pi_{0,1} = \Pi_0 + \Pi_{1|0}$ & $\Pi_0 \Pi_{1|0} = 0$ $\square$