<Chapter 7> <Slide #4> ①

Example) X_i's $\overset{iid}{\sim} N(\mu, \sigma^2)$. Let $m_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$; jth sample moment.

To find the method of moment estimator (MME), we set

$$\overset{\text{estimator}}{m_1} \overset{set}{=} E[m_1] = \overset{\text{parameter}}{\mu}$$

$$m_2 \overset{set}{=} E[m_2] = \sigma^2 + \mu^2.$$

and solve the above equations to find $\hat{\mu}^{MME}$ and $\hat{\sigma^2}^{MME}$

$$\hat{\mu}^{MME} = m_1, \quad \hat{\sigma^2}^{MME} = m_2 - m_1^2.$$

<Slide #5>.

EX) X_i's iid Gamma (α, β). $\hat{\alpha}^{MME}, \hat{\beta}^{MME} = ?$

Set
$$m_1 = E[m_1] = \alpha\beta \quad — ①$$
$$m_2 = E[m_2] = \alpha\beta^2 + \alpha^2\beta^2 \quad — ②$$

Solve the first one ①: $\beta = \frac{m_1}{\alpha}$. Plug-in into ② provides

$$m_2 = \alpha\cdot\left(\frac{m_1}{\alpha}\right)^2 + \alpha^2\left(\frac{m_1}{\alpha}\right)^2 = \frac{m_1^2}{\alpha} + m_1^2 \Leftrightarrow \hat{\alpha}^{MME} = \frac{m_1^2}{m_2 - m_1^2}$$

$$\therefore \hat{\beta}^{MME} = \frac{m_1}{\hat{\alpha}^{MME}} = \frac{m_2}{m_1} - m_1.$$

EX). X_i's iid Beta (α, β). $\hat{\alpha}^{MME}, \hat{\beta}^{MME} = ?$

Set $m_1 = E(m_1) = \frac{\alpha}{\alpha+\beta}$ —①, $m_2 = E(m_2) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \left(\frac{\alpha}{\alpha+\beta}\right)^2$ —②

Note : $E[m_2] = \left(\frac{\alpha}{\alpha+\beta}\right)\left(1 - \frac{\alpha}{\alpha+\beta}\right)\left(\frac{1}{\alpha+\beta+1}\right) + \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{m_1(1-m_1)}{\alpha+\beta+1} + m_1^2.$

From ① $(\alpha+\beta)m_1 = \alpha$.

From ② $(\alpha+\beta+1)m_2 = m_1 + (\alpha+\beta)m_1^2$.) solving this will give

$$\hat{\alpha}^{MME} = \frac{m_1(m_1 - m_2)}{m_2 - m_1^2}, \quad \hat{\beta}^{MME} = \frac{(m_1 - m_2)(1 - m_1)}{m_2 - m_1^2}.$$

<Slide #14>

X_{ij}'s $\overset{ind}{\sim} N(\mu_i, \sigma^2)$, $i=1,\cdots,s$, $j=1,\cdots,n$.

$$\ln[L(\mu_1,\cdots,\mu_s,\sigma^2)] = \log \prod_{i=1}^{s}\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(X_{ij}-\mu_i)^2}{2\sigma^2}\right\}$$

$$= -\frac{sn}{2}\log 2\pi\sigma^2 - \sum_i\sum_j (X_{ij}-\mu_i)/(2\sigma^2)$$

$$\frac{\partial}{\partial\mu_i}\ln[L(\mu_1,\cdots,\mu_s,\sigma^2)] = 2\sum_{j=1}^{n}(X_{ij}-\mu_i)/(2\sigma^2) \overset{set}{=} 0 \quad i=1,\cdots,s \quad - (a)$$

$$\frac{\partial}{\partial\sigma^2}\ln[L(\mu_1,\cdots\mu_s,\sigma^2)] = -\frac{sn}{2\sigma^2} + \sum_i\sum_j (X_{ij}-\mu_i)^2/(2\sigma^4) \overset{set}{=} 0. \quad - (b)$$

From (a), $\hat{\mu}_i^{MLE} = \sum_{j=1}^{n} X_{ij}/n$, $i=1,\cdots,s$.

From (b), $\hat{\sigma}^{2\,MLE} = \sum_{i=1}^{s}\sum_{j=1}^{n}(X_{ij}-\hat{\mu}_i^{MLE})^2/(sn)$.

<Slide #19>

$X|\theta \sim N(\theta,\sigma^2)$, σ^2 known. $\theta \sim N(\mu,\tau^2)$, μ,τ^2 known. Find $[\theta|X]$.

$$[\theta|X] \overset{\text{Bayes' theorem}}{\propto} [\theta][X|\theta]$$

$$\propto \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}\exp\left\{-\frac{(X-\theta)^2}{2\sigma^2}\right\}$$

$$\propto \exp\left\{-\theta^2\left(\frac{1}{2\tau^2}+\frac{1}{2\sigma^2}\right)+\theta\left(\frac{2\mu}{2\tau^2}+\frac{2x}{2\sigma^2}\right)\right\}$$

$$\propto \exp\left\{-\left(\frac{1}{2\tau^2}+\frac{1}{2\sigma^2}\right)\left[\theta-\frac{\frac{\mu}{2\tau^2}+\frac{x}{2\sigma^2}}{\frac{1}{2\tau^2}+\frac{1}{2\sigma^2}}\right]^2\right\}$$

$$= \exp\left\{\left[\theta-\frac{\sigma^2\mu+\tau^2x}{\sigma^2+\tau^2}\right]^2\Big/\left(\frac{1}{2\tau^2}+\frac{1}{2\sigma^2}\right)^{-1}\right\}$$

$$= \exp\left\{\left(\theta-\frac{\sigma^2\mu+\tau^2x}{\sigma^2+\tau^2}\right)^2\Big/2\left(\frac{\tau^2\sigma^2}{\sigma^2+\tau^2}\right)\right\}.$$

$\therefore [\theta|X] \sim N\left(\frac{\sigma^2\mu+\tau^2x}{\sigma^2+\tau^2}, \frac{\tau^2\sigma^2}{\sigma^2+\tau^2}\right)$. Note, the posterior mean is a weighted average of the prior mean and the data. $\left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)\cdot\mu + \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\cdot x$.

Ex). X_i's $\overset{iid}{\sim} f(x|\theta) = \theta e^{-\theta x}$, $\pi(\theta) = \beta e^{-\beta\theta}$, $\theta > 0$. <Slide#20>

$$\theta | \underset{\sim}{X} \propto \pi(\theta) f(\underset{\sim}{X}|\theta)$$

$$\propto e^{-\beta\theta} \cdot \theta^n e^{-\theta \Sigma X_i}$$

$$\propto \theta^n e^{-\theta(\Sigma X_i + \beta)}$$

$$= \theta^{(n+1)-1} e^{-\theta / [\Sigma X_i + \beta]^{-1}}$$

$$\sim \text{Gamma}\left(n+1, \frac{1}{\Sigma X_i + \beta}\right).$$

Note : $E[\theta] = \frac{1}{\beta}$. $E[\theta | \underset{\sim}{X}] = \frac{n+1}{\Sigma X_i + \beta} = \frac{1 + 1/n}{\bar{X} + \beta/n}$,

Posterior mean is mainly dependent on the data $\bar{x}$ unless β is an extremly large value.

<slide #32>. Proof of Lemma 7.3.11.

Note that $E\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right] = 0$ if differentiation and integration are exchangeable.

Then $\frac{\partial}{\partial\theta} E\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right] = E\left[\frac{\partial}{\partial\theta}\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right]\right]$

$$= \int \frac{\partial}{\partial\theta}\left\{\frac{\partial \log f(x|\theta)}{\partial\theta} \cdot f(x|\theta)\right\} dx$$

$$= \int \frac{\partial^2 \log f(x|\theta)}{\partial\theta^2} f(x|\theta)\, dx + \int \frac{\partial \log f(x|\theta)}{\partial\theta} \cdot \frac{\partial f(x|\theta)/\partial\theta}{f(x|\theta)} \cdot f(x|\theta)\, dx$$

$$= E\left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2}\right] + E\left[\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2\right] = 0.$$

$$\therefore E\left[\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2}\right].$$

<Additional Exercizes for UMVUE>

Ex) X_i's $\overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 known. Find UMVUE of μ^2.

Distribution of $\bar{X}$ belongs to the exponential family.

Thus $\bar{X}$ is a complete statistic. [and sufficient by factorization Thm]

Consider $Var(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2 = \sigma^2/n$.

$$\Leftrightarrow \quad E[\bar{X}^2] = \mu^2 + \sigma^2/n.$$

Thus $\bar{X}^2 - \frac{\sigma^2}{n}$ is an UE of μ^2. ($\frac{\sigma^2}{n}$ is known here)

By Lehmann & Scheffé Theorem, $\bar{X}^2 - \frac{\sigma^2}{n}$ is a UMVUE of μ^2.

Ex) X_i's $\overset{iid}{\sim} U(-\theta, \theta)$. Find UMVUE of θ, $\theta > 0$

Ex 7.37

$$\prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{(2\theta)^n} \prod_{i=1}^{n} I(-\theta < x_i < \theta) = \frac{1}{(2\theta)^n} I(0 \le \max_{1\le i\le n} |x_i| < \theta)$$

Thus, by factorization theorem $\max_{1\le i\le n} |X_i|$ is a sufficient statistic —ⓐ

Let $Y = \max_{1\le i\le n} |X_i|$. Then $Y \sim U[0, \theta)$, $f_Y(y) = n y^{n-1}/\theta^n$, $0 \le y < \theta$.

Suppose $g(y)$ is a function s.t.

$$E[g(Y)] = \int_0^\theta \frac{n y^{n-1}}{\theta^n} g(y)\, dy = 0, \quad \forall \theta$$

$$\Rightarrow \frac{\partial}{\partial\theta} E[g(Y)] = \frac{\partial}{\partial\theta} \int_0^\theta \frac{n y^{n-1}}{\theta^n} g(y)\,dy = 0 \Rightarrow n\theta^{n-1} g(\theta) = 0 \quad \forall\theta.$$

So $g(\theta) = 0$ for all θ. Therefore Y is a complete sufficient statistic —ⓑ

From ⓐ, ⓑ, Y is a C.S.S.

$$E[Y] = \int_0^\theta y \cdot \frac{n y^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta \quad \therefore \frac{n+1}{n} Y \text{ is U.E. of } \theta.$$

Then by Lehmann-Scheffé Theorem, $\frac{n+1}{n} Y$ is a UMVUE of θ.

EX1). X_i's $\overset{iid}{\sim} N(\mu, \sigma^2)$.

$$\frac{\partial}{\partial \sigma^2} \log f(\underset{\sim}{x} \mid \sigma^2) = \frac{\partial}{\partial \sigma^2} \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(X_i-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2}\log 2\pi\sigma^2 - \frac{\Sigma(X_i-\mu)^2}{2\sigma^2}\right]$$

$$= -\frac{n}{2\sigma^2} + \frac{\Sigma(X_i-\mu)^2}{2\sigma^4}$$

$$= \underbrace{\frac{n}{2\sigma^4}}_{a(\sigma^2)}\left\{\underbrace{\frac{\Sigma(X_i-\mu)^2}{n}}_{W} - \underbrace{\sigma^2}_{\tau(\sigma^2)}\right\}.$$

Thus, if μ is known, the variance of $\frac{\Sigma(X_i-\mu)^2}{n}$ attains the CRLB.
(If μ is unknown, then $\Sigma(X_i-\mu)^2/n$ does not attain the CRLB).

EX2). X_i's $\overset{iid}{\sim}$ Exp(λ).

$$\frac{\partial}{\partial \lambda} \log f(\underset{\sim}{X} \mid \lambda) = \frac{\partial}{\partial \lambda} \log \prod_{i=1}^{n} \frac{1}{\lambda} e^{-x_i/\lambda} = \frac{\partial}{\partial \lambda}\left(-n\log\lambda - \Sigma X_i/\lambda\right)$$

$$= -\frac{n}{\lambda} + \frac{\Sigma X_i}{\lambda^2} = \underbrace{\frac{n}{\lambda^2}}_{a(\lambda)}\left(\underbrace{\frac{\Sigma X_i}{n}}_{W} - \underbrace{\lambda}_{\tau(\lambda)}\right).$$

Thus $W = \bar{X}$ attains the CRLB.

<Slide #35>.

Rao-Blackwell Theorem says, we need to condition on a sufficient stat.
If not....??

EX). Let $X_1, X_2 \overset{iid}{\sim} N(\mu, 1)$. $\bar{X} = (X_1+X_2)/2$. Note X_1 is not sufficient.

Let $\phi(X_1) = E[\bar{X} \mid X_1]$. Then, $E\phi(X_1) = E[E[\bar{X} \mid X_1]] = E[\bar{X}] = \mu$.

Here, $Var(\phi(X_1)) = Var[E(\bar{X} \mid X_1)] = Var[E[\frac{1}{2}X_1 + \frac{1}{2}X_2 \mid X_1]] = Var(\frac{1}{2}X_1 + \frac{\mu}{2}) = \frac{1}{4}$

which is smaller than $Var(\bar{X}) = \frac{1}{2}$.

But $\phi(X_1) = \frac{1}{2}X_1 + \frac{\mu}{2}$ is <u>not</u> a statistic. So, we need to consider only estimators based on a sufficient statistic.

$X_1, \cdots, X_n \overset{iid}{\sim}$ Binomial(k, θ). Find UMVUE of ① θ, ② $k\theta(1-\theta)^{k-1}$.

①. $f(x|\theta) = \binom{k}{x} \theta^x (1-\theta)^{k-x}$, $x = 0, \cdots, k$

$$= \underbrace{\binom{k}{x}}_{h(x)} \cdot \underbrace{(1-\theta)^k}_{c(\theta)} \cdot \exp\{\underbrace{x}_{t_1(x)=x} \underbrace{\log\frac{\theta}{1-\theta}}_{w(\theta)=\log\frac{\theta}{1-\theta}}\}$$

Thus X_i's are from an exponential family. Then $\sum_{i=1}^n X_i$ is CSS.

[This is a result of Theorem 6.2.10 and Theorem 6.2.25].

$E[\sum X_i/(nk)] = \theta$ as $\sum X_i \sim$ Binomial(nk, θ).

Thus $\frac{\sum X_i}{nk}$ is a UMVUE of θ by Lehmann-Scheffé Theorem.

② Notice that $P(X_1 = 1) = \binom{k}{1}\theta^1(1-\theta)^{k-1} = k\theta(1-\theta)^{k-1}$.

Thus, consider an estimator based on this. $h(X_1) = \begin{cases} 1 & \text{if } X_1 = 1 \\ 0 & \text{o/w} \end{cases}$.

Then $E[h(X_1)] = \sum_{x=0}^{k} h(x)\binom{k}{x}\theta^x(1-\theta)^{k-x} = k\theta(1-\theta)^{k-1}$; unbiased.

Since $T = \sum X_i$ is CSS, $\phi(T) = E[\underset{\text{unbiased}}{h(X_1)} \mid \underset{\text{CSS}}{\sum X_i}]$ is a UMVUE of $k\theta(1-\theta)^{k-1}$.

[See Note.2 in slide #36]

Then when we observe $\sum_{i=1}^n X_i = t$, $\phi(t) = E[h(X_1)|\sum X_i = t] = P(X_1 = 1|\sum X_i = t)$

$$= \frac{P(X_1 = 1 \ \& \ \sum_{i=2}^n X_i = t-1)}{P(\sum X_i = t)} = \frac{k\theta(1-\theta)^{k-1} \cdot \binom{k(n-1)}{t-1}\theta^{t-1}(1-\theta)^{k(n-1)-(t-1)}}{\binom{kn}{t}\theta^t(1-\theta)^{kn-t}}$$

$$= \frac{k\binom{k(n-1)}{t-1}}{\binom{kn}{t}}. \Rightarrow \text{Therefore } \phi(T) = k\frac{\binom{k(n-1)}{\sum X_i - 1}}{\binom{kn}{\sum X_i}}.$$

X_i's $\overset{iid}{\sim}$ Geometric (p), $i=1,\dots,n$. Find the UMVUE of p.

$$f_X(x|p) = (1-p)^{x-1} p \quad , x = 1, 2, \cdots$$

$$= p \cdot \exp\{(x-1)\log(1-p)\} \rightarrow \text{exponential family.}$$

$T = \sum_{i=1}^{n} X_i$ is CSS. $\quad T \sim NB(n, p)$.

$$f_T(t|p) = \binom{n+t-1}{n-1} p^t (1-p)^n.$$

we need to find an unbiased estimator of p. To get some hint, let's find $\hat{p}^{MLE}$.

$$\Rightarrow \log f_T(t|p) \propto t\log p + n\log(1-p).$$

$$\Rightarrow \frac{\partial}{\partial p}\log f_T(t|p) = \frac{t}{p} + \frac{n}{1-p}(-1) \overset{set}{=} 0 \Rightarrow \hat{p}^{MLE} = \frac{T}{n+T}.$$

$$E[\hat{p}^{MLE}] = \sum_{t=0}^{\infty} \left(\frac{t}{n+t}\right)\binom{n+t-1}{n-1} p^t (1-p)^n$$

$$= \sum_{t=0}^{\infty} \frac{t}{n+t} \frac{(n+t-1)!}{(n-1)!\,t!} p^t (1-p)^n.$$

$\Rightarrow$ The final line provide a clue that there will be a cancellation when $\frac{t}{n+t-1}$ is used.

$$\Rightarrow E\left[\frac{T}{n+T-1}\right] = \sum_{t=1}^{\infty} \frac{t}{n+t-1} \cdot \frac{(n+t-1)!}{(n-1)!\,t!} p^t (1-p)^n$$

$\ast$ $t=0$ has no effect on the sum.

$$= \sum_{t=1}^{\infty} \frac{(n+t-2)!}{(n-1)!\,(t-1)!} p^{t-1}(1-p)^n \cdot p$$

$$= p\sum_{u=0}^{\infty} \underbrace{\frac{(n+u-1)!}{(n-1)!\,u!} p^u (1-p)^n}_{\text{pmf of } NB(n-1,\, p)}. \quad \text{by letting } t-1 = u$$

$= p$. $\therefore$ unbiased

By Lehmann - Scheffé Theorem, $\dfrac{\sum X_i}{n + \sum X_i - 1}$ is an UMVUE of p.

<Slide #39>

X_i's $\overset{iid}{\sim}$ Bernoulli(p). $i=1,\cdots,n$. Find the UMVUE of p^2.

The distribution of X_i's belong to an exponential family.

Thus $T=\Sigma X_i$ is a CSS, and $T \sim$ Binomial(n, p).

We know $E[T] = np$, $E[T^2] = np(1-p) + n^2p^2$.

From the above, we construct $E[T^2 - T] = n(n-1)p^2$.

$$\Rightarrow E\left[\frac{T(T-1)}{n(n-1)}\right] = p^2. \quad \therefore \frac{\sum_{i=1}^{n} X_i \left(\sum_{i=1}^{n} X_i - 1\right)}{n(n-1)} \text{ is an UMVUE of } p^2.$$

<Slide #50>

X_i's $\overset{iid}{\sim}$ Bernoulli(θ), $\theta \sim$ Uniform$(0,1)$

$$[\theta \mid \underset{\sim}{X}] \propto [\theta][\underset{\sim}{X} \mid \theta]$$

$$\propto \theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i}$$

$$\sim \text{Beta}(\Sigma X_i + 1, n - \Sigma X_i + 1)$$

The posterior mean minimizes the squared error loss.

$$\Rightarrow \hat{\theta}^{Bayes} = \frac{\Sigma X_i + 1}{n+2}$$

The posterior median minimizes the absolute error loss

$$\Rightarrow \hat{\theta}^{Bayes} = m \text{ s.t. } \int_0^m \frac{\Gamma(n+2)}{\Gamma(\Sigma X_i + 1)\Gamma(n - \Sigma X_i + 1)} \theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i} \, d\theta = \frac{1}{2}$$

[Solution should be found numerically]

< slide #46 >

X_i's $\overset{iid}{\sim} N(\mu, \sigma^2)$, $i=1,\cdots,n$. $\delta_b(X) = bS^2$ is an estimator of σ^2.

The risk under the squared error loss is

$R((\mu,\sigma^2), \delta_b(X)) = E[(\delta_b(X)-\sigma^2)^2] = E[(bS^2-\sigma^2)^2]$

$= Var(bS^2) + bias(bS^2)$

$= b^2 Var(S^2) + E[bS^2 - \sigma^2]^2$, $\quad \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \Rightarrow Var(S^2) = \dfrac{2\sigma^4}{(n-1)}$

$= \sigma^4 \left[\dfrac{2b^2}{(n-1)} + (b-1)^2 \right]$

$\Rightarrow$ This will be minimized when $\dfrac{2b^2}{(n-1)} + (b-1)^2$ is minimum.

$\Rightarrow \dfrac{1}{n-1}\left(2b^2 + (n-1)b^2 - 2b(n-1) + (n-1)\right) = \dfrac{1}{n-1}\left((n+1)b^2 - 2b(n-1) + (n-1)\right)$

$= \dfrac{n+1}{n-1}\left(b - \dfrac{n-1}{n+1}\right)^2 + const.$ $\quad \therefore b = \dfrac{n-1}{n+1}$ will minimize the risk among the estimators in the form of bS^2.

$\therefore \delta_b(X) = \dfrac{n-1}{n+1} S^2$ minimizes the risk.

Now, we consider the Stein's loss.

$$R(\sigma^2, \delta_b(X)) = E\left[\frac{bS^2}{\sigma^2} - 1 - \log\frac{bS^2}{\sigma^2}\right]$$

$$= bE\left[\frac{S^2}{\sigma^2}\right] - 1 - E\left[\log\frac{bS^2}{\sigma^2}\right]$$

$$= b - 1 - \log b + \underbrace{E\left[\log\frac{S^2}{\sigma^2}\right]}_{\text{not a function of } b}$$

$\Rightarrow$ will be minimized when $b=1$ by differentiating w.r.t. b and setting to zero.

$\therefore \delta_b(X) = S^2$ minimizes the risk under the Stein's loss.