8장. 모수의 추정 (Parameter Estimation)

Some definitions.

Def. (*Parameter* : 모수)

Unknown target constant that we are interested in.

Def. (*Parameter space* : 모수공간)

The set of all space that the parameter can take.

Def. (*Point Estimator* : 점추정량)

A point estimator is any function $\hat{\theta}(X_1, X_2, \cdots, X_n)$ of a sample.

That is, any statistic is a point estimator.

c.f. An estimate (추정치) is the realized value of an estimator.

Ex. $X_1, X_2, \cdots, X_n \overset{iid}{\sim} f(x \ : \theta)$ and $\overline{X}$ is an estimator of θ.

==> $x_1, x_2, \cdots, x_n$ (observed sample value) & $\overline{x}$ is an estimate of θ.

8.2 최대가능도추정법 (Maximum Likelihood Estimator : MLE)

Def. (*Likelihood Function* : 가능도함수)

Let $f(\boldsymbol{x}|\theta)$ be the joint pmf or pdf of $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$.

(i.e, $f(\boldsymbol{x}|\theta) = f(x_1|\theta) f(x_2|\theta) \cdots f(x_n|\theta)$)

Then, given $\boldsymbol{X} = \boldsymbol{x}$(observed), the function of θ defined by

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$$

is called the likelihood function.

Note1. For $\boldsymbol{X}$ is discrete, $L(\theta|\boldsymbol{x}) = P_\theta(\boldsymbol{X} = \boldsymbol{x})$.

If $P_{\theta_1}(\boldsymbol{X} = \boldsymbol{x}) = L(\theta_1|\boldsymbol{x}) > L(\theta_2|\boldsymbol{x}) = P_{\theta_2}(\boldsymbol{X} = \boldsymbol{x})$ then it suggests that

this observation $\boldsymbol{x}$ is more likely to have occurred if $\theta = \theta_1$

than if $\theta=\theta_2$.

Note2. In $L(\theta|\boldsymbol{x})$, we consider $\boldsymbol{x}$ to be the observed sample point and θ to be varying over all possible parameter values.

Def. (*Maximum Likelihood Estimator* : MLE, 최대가능도추정량)

Maximum likelihood estimator (MLE) of θ is the value of $\hat{\theta}(\boldsymbol{x})$ which maximizes $L(\theta|\boldsymbol{x})$.

(i.e., $\hat{\theta}^{MLE}(\boldsymbol{x})=\underset{\theta}{\operatorname{argmax}}\, L(\theta|\boldsymbol{x})$)

When $\boldsymbol{X}$ is discrete, $L(\theta|\boldsymbol{x}) = P_\theta(\boldsymbol{X}=\boldsymbol{x})$.

Given $\boldsymbol{X}=\boldsymbol{x}$, $\hat{\theta}(\boldsymbol{x})$ gives maximum value of $L(\theta|\boldsymbol{x})$.

Note. From now on, we will denote $\hat{\theta}=\hat{\theta}(\boldsymbol{x})$, $L(\theta)=L(\theta|\boldsymbol{x})$.

Ex 8.7 (Read text p.179 and Explain)

Ref. When we are finding MLE, normally it is easier to maximize $\log L(\theta)$ instead of maximizing $L(\theta)$.

$$\begin{aligned}\hat{\theta}&=\operatorname{argmax}_\theta \log L(\theta)\\&=\operatorname{argmax}_\theta l(\theta)\\&=\operatorname{argmax}_\theta \sum_{i=1}^{n}\log f(x_i|\theta).\end{aligned}$$

Example of Log-Likelihood : See Fig 8.1 (concave & has maximum)

Ex 8.8 (MLE for Bernoulli)

$$X_1, X_2, \cdots X_n \overset{iid}{\sim} Bernoulli(p), \text{ MLE of } p?$$

Sol. (i) Using likelihood

$$L(p) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum x_i}(1-p)^{n-\sum x_i}, \ 0 \le p \le 1$$

$$\frac{dL(p)}{dp} = 0 \ ;$$

$$(\sum x_i)p^{\sum x_i - 1}(1-p)^{n-\sum x_i} - (n - \sum x_i)p^{\sum x_i}(1-p)^{n-\sum x_i - 1} = 0$$

$$\Rightarrow \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \quad \Rightarrow \quad \sum x_i - np = 0$$

$$\therefore \ p = \frac{\sum x_i}{n} = \bar{x}, \quad i.e. \ \hat{p}^{MLE} = \overline{X}$$

(ii) Using log-likelihood

See text p. 180 & Fill out by yourself

Ex 8.9

$$X_1, X_2, \cdots X_n \overset{iid}{\sim} Poisson(\lambda), \text{ MLE of } \lambda?$$

Sol.

Likelihood : $L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}, \ 0 < \lambda$

log-likelihood :

$$l(\lambda) = \ln L(\lambda) = \sum_{i=1}^{n} \log f(x_i|\lambda) = \sum_{i=1}^{n} (x_i \log\lambda - \lambda - \log(x_i!))$$

$$= \log\lambda \sum_{i=}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log(x_i!)$$

$$\frac{dl(\lambda)}{d\lambda} = 0 \Rightarrow \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$$

$$\therefore \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad \therefore \hat{\lambda}^{MLE} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Question. In this example, is the likelihood function concave?

Answer : Yes, $\dfrac{d^2 l(\lambda)}{d\lambda^2} = -\dfrac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$ holds.

Not all cases but most of cases, the likelihood function is concave, which means finding a root of equation is enough. But, if we have more than one solution, we should concern it more carefully.

Ex 8.9-2 (MLE for Exponential)

$$X_1, X_2, \cdots X_n \overset{iid}{\sim} f(x:\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \ 0 < x < \infty, \ 0 < \theta < \infty, \text{ MLE of } \theta ?$$

Likelihood : $L(\theta) = (\frac{1}{\theta}e^{-\frac{x_1}{\theta}}) \cdots (\frac{1}{\theta}e^{-\frac{x_n}{\theta}}) = \frac{1}{\theta^n}e^{-\Sigma\frac{x_i}{\theta}}$, $0 < \theta < \infty$

log-likelihood : $l(\theta) = \ln L(\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$, $0 < \theta < \infty$

$$\frac{dl(\theta)}{d\theta} = 0 \Rightarrow -\frac{n}{\theta} + \frac{\sum x_i}{\theta^2} = 0$$

$$\therefore \theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad \therefore \hat{\theta}^{MLE} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Ex 8.10 (MLE for Double Exponential)

Fill out by yourself

Ex 8.11-0 (Normal with known variance)

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2), \quad \sigma^2 : \text{known, MLE of } \mu?$$

Sol. (i) Using likelihood

$$L(\mu) = \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma}\right)\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\} \qquad \text{-----} \ (*)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$

Because e^{-y} is decreasing, maximizing $L(\mu)$ requires minimizing $\sum_{i=1}^{n}(x_i-\mu)^2$.

==> $\hat{\mu} = \bar{x}$ minimizes $\sum_{i=1}^{n}(x_i-\mu)^2$.

So, $\hat{\mu}^{MLE} = \bar{X}$.

(i) Using log-likelihood

Recall (*) in the above,

$$l(\mu) = \ln L(\mu) = \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

$$\frac{dl(\mu)}{d\mu} = 2\sum(x_i-\mu)/2\sigma^2 = 0 \quad ==> \hat{\mu}^{MLE} = \bar{X}.$$

$$\frac{d^2 l(\mu)}{d\mu^2} = -\frac{n}{\sigma^2}\Big|_{\mu=\bar{X}} = -\frac{n}{\sigma^2} < 0.$$

Therefore, $\hat{\mu}^{MLE}$ is the global maximizer.

Thm 8.1 (*Invariance property of MLE*)

If $\hat{\theta}$ is the MLE of θ, then for any function $h(\theta)$, the MLE of $h(\theta)$ is $h(\hat{\theta})$.

Note. The invariance property of MLE also holds in the multi-parameter case.
That is, if the MLE of $(\theta_1,\theta_2,\cdots,\theta_k)$ is $(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)$,
then MLE of $h(\theta_1,\theta_2,\cdots,\theta_k)$ is $h(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)$.

e.g.1 $X_1,X_2,\cdots,X_n$: iid from $N(\mu,\sigma^2)$ with known σ^2.
$\overline{X}$: MLE of μ.
Then, MLE of μ^2 is $\overline{X}^2$ and MLE of μ^3-1 is $\overline{X}^3-1$.

e.g.2 $X_1,X_2,\cdots,X_n$: iid Bernoulli(p)

==> MLE of p ==> $\hat{p}=\frac{1}{n}\sum_{i=1}^{n}X_i$

What is the MLE of $\sqrt{p(1-p)}=\sqrt{Var(X_1)}$?
==> By invariance property of MLE
MLE of $\sqrt{p(1-p)}$ is $\sqrt{\hat{p}(1-\hat{p})}=\sqrt{\overline{X}(1-\overline{X})}$.

Ex 8.11 (Normal with unknown variance)

$$X_1,\ X_2,\cdots,X_n \overset{iid}{\sim} N(\mu,\ \sigma^2),\qquad \Omega=\{(\mu,\sigma^2)\ :\ -\infty<\mu<\infty\ ,\ 0<\sigma^2<\infty\}$$

MLE of $\mu,\ \sigma^2$?

Sol.

$$L(\mu\ ,\ \sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}\ ,\quad (\mu\ ,\ \sigma^2)\in\Omega$$

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

$$\frac{\partial l(\mu,\sigma^2)}{\partial \mu} = 0 \Rightarrow \frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = 0 \Rightarrow \frac{\sum(x_i-\mu)}{\sigma^2} = 0 \qquad \therefore\ \mu = \frac{1}{n}\sum x_i = \bar{x}$$

$$\frac{\partial l(\mu,\sigma^2)}{\partial \sigma^2} = 0 \Rightarrow \frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} = 0 \Rightarrow -\frac{n}{2\sigma^2} + \frac{1}{2{\sigma^2}^2}\sum(x_i-\mu)^2 = 0$$

$$\therefore\ \sigma^2 = \frac{1}{n}\sum(x_i-\mu)^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$$

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

참고1. We need to check whether the following matrix is negative definite.

$$\begin{pmatrix} \frac{\partial^2 l(\mu,\sigma^2)}{\partial \mu^2} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 l(\mu,\sigma^2)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 l(\mu,\sigma^2)}{\partial^2 \sigma^2} \end{pmatrix}$$

참고2. What is the definition of negative definite matrix?

Fill out by yourself

Ex 8.12 (In case of likelihood function is not differentiable)

$$X_1,\, X_2, \cdots, X_n \overset{iid}{\sim} f(x|\theta) = \exp(-(x-\theta)),\ x \ge \theta$$

MLE of θ?

Sol.

$$f(x_1, x_2, \ldots, x_n|\theta)$$

$$= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} I(x_1 \ge \theta) I(x_2 \ge \theta) \cdots I(x_n \ge \theta)$$

$$= e^{-(x_1-\theta)} e^{-(x_2-\theta)} \cdots e^{-(x_n-\theta)} I(x_{(1)} \ge \theta)$$

Then, the likelihood function is in Figure 8.2.

Therefore, $\hat{\theta}^{MLE} = X_{(1)}$.

8.1 적률추정법 (Method of Moments Estimator : MME)

$X_1, X_2, \cdots, X_n$: iid from $f(x|\theta_1, \theta_2, \cdots, \theta_k)$

Calculate

$$EX_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k)$$

$$EX_1^2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k)$$

$$\vdots$$

$$EX_1^k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k)$$

Set sample moments to population moments.

$$m_1 = \frac{\sum_{i=1}^{n} X_i}{n} = \mu_1(\theta_1, \theta_2, \cdots, \theta_k)$$

$$m_2 = \frac{\sum_{i=1}^{n} X_i^2}{n} = \mu_2(\theta_1, \theta_2, \cdots, \theta_k)$$

$$\vdots$$

$$m_k = \frac{\sum_{i=1}^{n} X_i^k}{n} = \mu_k(\theta_1, \theta_2, \cdots, \theta_k)$$

==> that is a function of $\theta_1, \theta_2, \cdots, \theta_k$.

Solving the above equations w.r.t. $\theta_1, \theta_2, \cdots, \theta_k$ yields the MME.

Examples :

1. $X_1, X_2, \cdots, X_n$: random sample from Bernoulli(p)

$E(X_1) = p$ ==> $\frac{\sum_{i=1}^{n} X_i}{n} = \overline{X} = \hat{p}$

==> $\hat{p}^{MME} = \overline{X}$.

2. $X_1, X_2, \cdots, X_n$: random sample from a Gamma(α, β)

$$EX_1 = \alpha\beta, \quad EX_1^2 = Var(X_1) + (EX_1)^2 = \alpha\beta^2 + (\alpha\beta)^2 = (\alpha + \alpha^2)\beta^2$$

==> Solve

$$\alpha\beta = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad (\alpha + \alpha^2)\beta^2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

$$\text{==>} \quad \left(\frac{\overline{X}}{\beta} + \frac{\overline{X}^2}{\beta^2}\right)\beta^2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

$$\text{==>} \quad \overline{X}\beta + \overline{X}^2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

$$\text{==>} \quad \hat{\beta}^{MME} = \left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2\right)/\overline{X} \quad \text{and}$$

$$\hat{\alpha}^{MME} = \frac{\overline{X}}{\hat{\beta}} = \frac{\overline{X}^2}{\frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2}$$

Note : $\hat{\beta}^{MME} = \left(\frac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2\right)/\ \overline{X} = S^2 / \overline{X},$

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

Sample variance with n in denominator instead of $n-1$.

Ex 8.1 (Binomial)

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} Bin(m, \theta), \ \hat{\theta}^{MME} \ ?$$

Sol.

Ex 8.2 (Poisson)

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} Poisson(\theta),\ \hat{\theta}^{MME}\ ?$$

Sol.

Ex 8.3 (Gamma)

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} \mathrm{Gamma}(2, \theta),\ \hat{\theta}^{MME}\ ?$$

Sol.

Ex 8.4

$$X_1, X_2, \cdots, X_n \overset{iid}{\sim} f(x|\theta) = \exp(-(x-\theta)),\ x \geq \theta,\ \hat{\theta}^{MME}\ ?$$

Sol.

Ex 8.5 (Normal)

$X_1, \cdots, X_n$: iid from $N(\theta, \sigma^2)$

$$E(X_1) = \theta$$

$$E(X_1^2) = Var(X_1) + (EX_1)^2$$

$$= \sigma^2 + \theta^2$$

$$\frac{1}{n}\sum X_i = \hat{\theta}$$

$$\frac{1}{n}\sum X_i^2 = \widehat{\sigma^2} + \hat{\theta}^2 \ = \widehat{\sigma^2} + \overline{X}^2$$

$$\therefore \quad \hat{\theta}^{MME} = \overline{X},$$

$$\widehat{\sigma^2}^{MME} = \frac{1}{n}\sum X_i^2 - \overline{X}^2 \ = \frac{1}{n}\sum (X_i - \overline{X})^2$$

Ex 8.6 (Normal)

See handout page 10.

Note. Method of moment estimator is not so popular since it does not have good statistical properties.

8.3 추정량의 성질 (Properties of Estimators)

Question. When we have several estimators for single parameter θ, which estimator is preferred?

Answer. To answer this question, we need to study criteria for the comparison and some statistical properties of estimators.

Def 8.1 (*Unbiased estimator & Consistent estimator*)

(i) (Unbiase estimator: 불편추정량, 비편향추정량, U.E.)

$\hat{\theta}$: an estimator for θ

If $E(\hat{\theta})=\theta$, then $\hat{\theta}$ is an unbiased estimator for θ.

(ii) (Consistent estimator: 일치추정량, C.E.)

$\hat{\theta}$: an estimator for θ

If $\hat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$, then $\hat{\theta}$ is an consistent estimator for θ.

Note. $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$ (편향, 편차). Then, unbiased estimator has bias 0.

Ex 8.13 $X_1, X_2, \cdots, X_n$: r.s from the population with mean μ, variance σ^2.

(a) X_1 : U.E. for θ but not C.E. for θ

Sol.

(b) $\overline{X}$: U.E. for θ and C.E. for θ

Sol.

(c) $\frac{n}{n+1}\overline{X}$: not U.E. for θ but C.E. for θ

Sol.

Question. How to check whether an estimator has consistency property?

Answer1. Use Markov inequality.

Def. (*Markov inequality* : 마코프 부등식)

For a random variable X,

$$P(|X| \geq \epsilon) \leq \frac{E(X^2)}{\epsilon^2}.$$

Def. (*Mean squared error* : 평균제곱오차, MSE)

$\hat{\theta}$: an estimator for θ

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 \\ &= E(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2 \\ &= E(\hat{\theta}-E(\hat{\theta}))^2+2E(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)+(E(\hat{\theta})-\theta)^2 \\ &= E(\hat{\theta}-E(\hat{\theta}))^2+(E(\hat{\theta})-\theta)^2 \\ &= Var(\hat{\theta})+(Bias(\hat{\theta}))^2 \end{aligned}$$

Answer2. Therefore, $\hat{\theta} \xrightarrow{p} \theta$ if $MSE(\hat{\theta}) \to 0$. i.e., $Var(\hat{\theta}) \to 0$ and $Bias(\hat{\theta}) \to 0$.

Ex 8.14 $X_1, X_2, \cdots, X_n$: r.s from $Poisson(\theta)$. Show that $\hat{\theta} = \overline{X}$ is a C.E. for θ.

Sol.

Ex 8.15 $X_1, X_2, \cdots, X_n$: r.s from $N(\mu, \sigma^2)$.

(a) Show that $\hat{\mu}^{MLE}$ is a C.E. for μ.

Sol.

(b) Show that $\widehat{\sigma}^{2\,MLE}$ is a C.E. for σ^2.

Sol. Recall $\widehat{\sigma}^{2\,MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Note that $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$.

Therefore,

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

$$Var\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = 2(n-1).$$

The bias and variance of $\widehat{\sigma}^{2\,MLE}$ are :

$$Bias(\widehat{\sigma^2}) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$$

$$Var(\widehat{\sigma^2}) = Var\left(\frac{\sigma^2}{n}\cdot\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}\right)$$

$$= \frac{\sigma^4}{n^2}Var\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2}\right) = \frac{2(n-1)}{n^2}(\sigma^2)^2$$

Confidence Intervals for Means

$X_1, X_2, X_3, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$일 때 μ 의 $100(1-\alpha)\%$ C.I ?

(i) σ^2 Known :

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ 임을 이용}$$

$$P\left\{-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} = 1-\alpha$$

$$\therefore \overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

(ii) σ^2 Unknown :

① Large sample (no need to assume normality)

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \doteq N(0,1) \text{ 임을 이용}$$

$$\Rightarrow \overline{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$$

② Small sample

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \text{ 임을 이용}$$

$$\Rightarrow \overline{X} \pm t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$$

Confidence Intervals for difference of two means

$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_X, \sigma_X^2)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} N(\mu_Y, \sigma_Y^2)$ *are independent*

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y \ , \ \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

(i) σ_X^2 , σ_Y^2 : Known

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

(ii) σ_X^2 , σ_Y^2 : Unknown

① Large Sample

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \overset{\circ}{\underset{\circ}{\sim}} N(0, 1)$$ 임을 이용

② Small Sample

$\sigma_X^2 = \sigma_Y^2 = \sigma^2$ 임을 가정

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1)$$

$$U = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(n+m-2)$$

Then, Z and U are independent .

$$\therefore\ T = \frac{Z}{\sqrt{U/(n+m-2)}} = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\dfrac{(n-1)S_X^2 - (m-1)S_Y^2}{n+m-2}\ (\dfrac{1}{n}+\dfrac{1}{m})}}$$

$$= \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{S_P\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim t\,(n+m-2)$$ 임을 이용

$$\Rightarrow\ \overline{X}-\overline{Y} \pm t_{\alpha/2}(n+m-2)S_P\sqrt{\frac{1}{n}+\frac{1}{m}}$$

- $\sigma_X^2 \neq \sigma_Y^2$ 일때, Welch's $t-dist^n$ 이용

참고) Paired comparison

$(X_1, Y_1), \ldots, (X_n, Y_n)$

$D_i \equiv X_i - Y_i,\quad i = 1,2,\ldots,n \quad , \quad \mu_D = \mu_X - \mu_Y$

$T = \dfrac{\overline{D} - \mu_D}{S_D/\sqrt{n}} \sim t\,(n-1)$ 임을 이용

Confidence Intervals for variances

$X_1, X_2, X_3, \ldots, X_n \sim iid\ N(\mu, \sigma^2)$

$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ and $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2\,(n-1)$

- C.I. for σ^2

: $P\left(a \le \dfrac{(n-1)S^2}{\sigma^2} \le b\right) = 1-\alpha$

(ex) take $a = \chi^2_{1-\alpha/2}(n-1)$, $b = \chi^2_{\alpha/2}(n-1)$

$$1-\alpha = P\left(\frac{(n-1)S^2}{b} \le \sigma^2 \le \frac{(n-1)S^2}{a} \right)$$

$\therefore 100(1-\alpha)\%$ $C.I.$ for σ^2

: $[\ (n-1)S^2/b\ ,\ (n-1)S^2/a\ ]$, $a = \chi^2_{1-\alpha/2}(n-1)$, $b = \chi^2_{\alpha/2}(n-1)$

For σ , $\left[\sqrt{\frac{(n-1)S^2}{b}} \ ,\ \sqrt{\frac{(n-1)S^2}{a}} \right]$

- C.I. for variance ratio

$Y_1, \ldots, Y_m \sim^{iid} N(\mu_Y, \sigma_Y^2)$ and $X_1, \ldots, X_n \sim^{iid} N(\mu_X, \sigma_X^2)$ *are independent*

$$\frac{\frac{S_Y^2}{\sigma_Y^2}}{\frac{S_X^2}{\sigma_X^2}} = \frac{\left[\frac{(m-1)S_Y^2}{\sigma_Y^2}\right] / (m-1)}{\left[\frac{(n-1)S_X^2}{\sigma_X^2}\right] / (n-1)} \sim F\,(m-1\,,\,n-1)$$

$$\therefore\ 1-\alpha = P\left(c \le \frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} \le d\right)$$

$$= P\left(c\,\frac{S_X^2}{S_Y^2} \le \frac{\sigma_X^2}{\sigma_Y^2} \le d\,\frac{S_X^2}{S_Y^2}\right)$$

ex : take $c = F_{1-\alpha/2}(m-1,n-1) = \dfrac{1}{F_{\alpha/2}(n-1,m-1)}$

$d = F_{\alpha/2}(m-1,n-1)$

Confidence Intervals for Proportions

$Y \sim B(n, p)$

- C.I. for p

$$\frac{Y-np}{\sqrt{np(1-p)}} = \frac{Y/n-p}{\sqrt{p(1-p)/n}} \overset{\cdot}{\sim} N(0,1) \quad as \quad n \to \infty$$

$$\therefore \; P\left(-z_{\alpha/2} \le \frac{Y/n-p}{\sqrt{p(1-p)/n}} \le z_{\alpha/2}\right) \approx 1-\alpha$$

$\therefore$ 100 $(1-\alpha)\%$ *approx. C.I.* for p

$$: \left[\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \;,\; \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \quad , \quad \hat{p} = \frac{Y}{n}$$

-C.I. for difference of proportion

$Y_1 \sim B(n_1, p_1)$ and $Y_2 \sim B(n_2, p_2)$ *are independent*

$$\frac{Y_1/n_1 - Y_2/n_2 - (p_1-p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \;\overset{\circ}{\underset{\circ}{\sim}}\; N(0,1)$$

$\therefore$ 100$(1-\alpha)\%$ *approx. C.I.* for (p_1-p_2)

$$: \frac{Y_1}{n1} - \frac{Y_2}{n_2} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

8.4 추정량의 효율과 크래머-라오 부등식
(Efficiency of Estimators and Cramér-Rao Inequality)

Question. Suppose we have two different estimators, say $\hat{\theta}$ and $\tilde{\theta}$. Which one do we use ?

Answer. Select the one that has smallest mean squared error; that is, if, for example,

$$E(\hat{\theta}-\theta)^2 < E(\tilde{\theta}-\theta)^2,$$

then, $\hat{\theta}$ is preferred.

If $\hat{\theta}$ and $\tilde{\theta}$ are unbiased estimators, then we would select the one with the smallest variance.

Question. Are there any possible tools to determine the variance of unbiased estimators?

Answer. Yes, there is an interesting result by Cramer-Rao.
(Carmer-Rao Inequality : 크래머-라오 부등식)

Thm 8.2

Let $X \sim f(x|\theta)$. Then, under some assumptions, the followings are hold.

$$E\left(\frac{d}{d\theta}\log f(X|\theta)\right) = 0 \quad \text{..............................} \quad (1)$$

$$E\left(\frac{d}{d\theta}\log f(X|\theta)\right)^2 = -E\left(\frac{d^2}{d\theta^2}\log f(X|\theta)\right) \quad \text{......} \quad (2)$$

Proof.

(1)

(2) fill out by yourself

Thm 8.4 (*Carmer-Rao Inequality*)

$X_1, X_2, \cdots, X_n$: random sample from a distribution with pdf $f(x|\theta)$.
If $T = T(X_1, X_2, \cdots, X_n)$ is an U.E. of θ, then

$$Var(T) \geq \frac{1}{n\int_{-\infty}^{\infty}\left[\frac{d\log f(x|\theta)}{d\theta}\right]^2 f(x|\theta)d\theta}$$

$$= -\frac{1}{n\int_{-\infty}^{\infty}\left[\frac{d^2\log f(x|\theta)}{d\theta^2}\right] f(x|\theta)d\theta}$$

Note. The term in the denominator is called Fisher information number.

$$i.e.,\quad I(\theta) = E\left\{\left[\frac{d\log f(x|\theta)}{d\theta}\right]^2\right\} = \int_{-\infty}^{\infty}\left[\frac{d\log f(x|\theta)}{d\theta}\right]^2 f(x|\theta)d\theta$$

$$= -E\left(\frac{d^2}{d\theta^2}\log f(X|\theta)\right)$$

Therefore, Cramer-Rao Inequality becomes :

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

Ex 8.18 $X_1, X_2, \cdots, X_n$: random sample from the Bernoulli(θ).
Find the Fisher information number.
Then, what is the lower bound of variance of an U.E. ?

Sol.

$f(x) =$

Since $\log f(x|\theta) = \ldots$

Ex 8.20 $X_1, X_2, \cdots, X_n$: random sample from the Bernoulli(θ).

Then, show that $\hat{\theta} = \overline{X}$ has the smallest variance among unbiased estimators.

Example. $X_1, X_2, \cdots, X_n$: random sample from the $Poisson(\theta)$.

Find the Fisher information number.

Then, what is the lower bound of variance of an U.E. ?

Sol. *Fill out by yourself*

Since $\log f(x|\theta) = \ldots$

Ex 8.19 $X_1, X_2, \cdots, X_n$: random sample from the $Poisson(\theta)$.

Then, show that $\hat{\theta} = \overline{X}$ has the smallest variance among unbiased estimators.

Sol. *Fill out by yourself*

Example. $X_1, X_2, \cdots, X_n$: random sample from the exponential distribution with pdf

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}},\ 0 < x < \infty,\ 0 < \theta < \infty.$$

Then, what is the lower bound of variance of an U.E. ?

Sol.

Since $\log f(x|\theta) = \ldots$

8.6 최대가능도추정량의 점근적 성질 (Asymptotic Properties of Maximum Likekihood Estimators)

Def (*Asymptotic property : 점근적 성질*)

Property of an estimator when the sample size gets larger.

: 표본의 크기가 커질 때 추정량이 가지는 성질

Assumptions

(i) Log-likelihood function is differentiable

(ii) The order of differentiation and integration of log-likelihood can be changed.

Thm 8.7 (*Consistency of MLE*) Under some conditions,

$X_1, X_2, \cdots, X_n$: random sample from $f(x|\theta)$.

Let $\hat{\theta}_n$ is the MLE of θ. Then, $\hat{\theta}_n$ is a consistent estimator of θ.

(*i.e.,* $\hat{\theta}_n^{MLE} \xrightarrow{p} \theta$)

Thm 8.10 (*Asymptotic Normality of MLE* : MLE의 점근정규성)

Under some conditions, let $\hat{\theta}_n$ is the MLE of θ. Then,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right).$$

Implication of Thm 8.10

최대가능도추정량 $\hat{\theta}_n^{MLE}$는 일치추정량이고 정리 8.10에 주어진 점근분포로부터 최대가능도추정량의 점근분산은 크래머-라오 하한과 같다. 즉, 점근적인 관점에서 볼 때 최대가능도추정량은 비편향추정량이고 분산은 크래머-라오 하한과 같으므로 최소의 분산을 가지는 최적의 추정량임.

Ex 8.26 $X_1, X_2, \cdots, X_n$: random sample from the $Poisson(\theta)$.

Find the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, where $\hat{\theta}_n = \overline{X}$.

Sol.

Since $\hat{\theta}_n = \overline{X}$ is the MLE and $I(\theta) = 1/\theta$, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \theta)$.

Go to Chapter 7 at this stage...

8.5 최소분산비편향추정량

In English, minimum variance unbiased estimator,
best unbiased estimator,
uniformly minimum variance unbiased estimator
(UMVUE)

Note1. Comparison of two different unbiased estimators :

MSE criterion ⇔ Variance criterion.

Note2. The estimator with smaller variance is preferred among unbiased estimators. (즉, 비편향추정량 중에서는 분산이 작은 것이 선호됨)

Def. (*Uniformly Minimum Variance Unbiased Estimator*: UMVUE, 균일최소분산비편향추정량)

$\tau(\theta)$: Target parameter

W^* is the uniformly minimum variance unbiased estimator of $\tau(\theta)$

if $E_\theta(W^*) = \tau(\theta)$ for all θ and,

for any other estimator W with $E_\theta(W) = \tau(\theta)$,

$Var_\theta(W^*) \leqq Var_\theta(W)$ for all θ.

W^* is also called best unbiased estimator.

즉, 비편향추정량 중에서 분산을 최소로 하는 것.

Question. How to find UMVUE ?

Answer. Step1. Find a unbiased estimator.
Step2. Calculate the variance of the estimator in step1.
Step3. If the result in step2 is equal to the Cramer-Rao lower bound, it is the UMVUE.

Question. Is it easy to find UMVUE by the above approach?

Answer. Generally NOT.

Question. Is there any tool for finding UMVUE? If so, what is it?

Answer. We don't need to consider all possible unbiased estimators.
We can restrict the class of U.E.s by the following theorem.

Thm 8.5 (Rao-Blackwell theorem, 라오-블랙웰 정리)

W : any unbiased estimator of θ, T : a sufficient statistic for θ.
Define $\phi(T) = E(W|T)$.
Then, $E_\theta\phi(T) = \theta$ and $Var_\theta\phi(T) \leqq Var_\theta W$ for all θ,
that is, $\phi(T)$ is uniformly better than W for all θ .

Proof)

Use the fact $E(X) = E[E(X|Y)]$ and

$$Var(X) = Var[E(X|Y)] + E[Var(X|Y)].$$

That is, $\theta = E_\theta W = E_\theta[E(W|T)] = E_\theta\,\phi(T)$ and

$$\begin{aligned} Var_\theta W &= Var_\theta[E(W|T)] + E_\theta[Var(W|T)] \\ &= Var_\theta\phi(T) + E_\theta[Var(W|T)] \\ &\geqq Var_\theta\phi(T) \end{aligned}$$

Is $\phi(T)$ an estimator ? Yes, since it does not depend on θ.

Implication Rao-Blackwell Thm

Conditioning any U.E on a S.S will result in a uniform improvement.
Therefore, we need consider only statistics that are function of a S.S.

(즉, 좋은 추정량을 선택할 때는 충분통계량의 함수인 것들만 생각하면 됨)

Ex 8.21 $X_1, X_2, \cdots, X_n$: random sample from the $Poisson(\theta)$.

(a) Calculate the variance of $S(X) = X_1$.

Sol.

Since $X_1 \sim Poisson(\theta)$, $Var(S(X)) = \theta$.

(b) Derive the estimator $S^*(X) = E(S(X)|T(X))$, where $T(X) = \sum_{i=1}^{n} X_i$.

Sol.

Since $\sum_{i=1}^{n} X_i \sim Poisson(n\theta)$, the conditional distn of X_1 given $T(X) = t$ is as follows :

$$P(X_1 = x_1 | \sum_{i=1}^{n} X_i = t) = \frac{P(X_1 = x_1, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{P(X_1 = x_1, \sum_{i=2}^{n} X_i = t - x_1)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{\frac{\theta^{x_1} e^{-\theta}}{x_1 !} \frac{((n-1)\theta)^{t-x_1} e^{-(n-1)\theta}}{(t-x_1)!}}{\frac{(n\theta)^t e^{-n\theta}}{t!}}$$

$$= \frac{t!}{x_1!(t-x_1)!} \left(\frac{1}{n}\right)^{x_1} \left(1 - \frac{1}{n}\right)^{t-x_1}$$

Therefore, $X_1 \left| \sum_{i=1}^{n} X_i = t \right. \sim Bin\left(t, \frac{1}{n}\right)$.

Hence,

$$S^*(X) = E(S(X)|T(X)) = E\left(X_1 \left| \sum_{i=1}^{n} X_i \right.\right) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

(c) Show that $S^*(X)$ in (b) is unbiased estimator and calculate the variance.

Sol.

Since $S^*(X) = E(S(X)|T(X)) = \overline{X}$ and $E(\overline{X}) = \theta$,

$S^*(X)$ is an U.E. for θ.

And $Var(\overline{X}) = \dfrac{\theta}{n}$.

Note that $Var(S^*(X)) = \dfrac{\theta}{n} < \theta = Var(S(X))$.

Thm 8.6 (Lehmann-Scheffe theorem, 레만-쉐페 정리)

$S(X)$: unbiased estimator of θ, $T(X)$: C.S.S. for θ.

$\Rightarrow S^*(X) = E(S(X)|T(X))$: UMVUE for θ.

Proof.

$T(X)$: S.S., 라오-블랙웰 정리에 의하여

$S^*(X)$: θ의 U.E. and $Var(S^*(X)) \le Var(S(X))$.

$U(X)$: θ의 임의의 비편향추정량이라 하자.

그러면, $U^*(X) = E(U(X)|T(X))$는 라오-블랙웰 정리에

의하여 θ의 비편향추정량이고, $Var(U^*(X)) \le Var(U(X))$.

$S^*(X), U^*(X)$: θ의 비편향추정량이므로 모든 θ에 대하여

$E(S^*(X) - U^*(X)) = 0$ 성립.

$S^*(X), U^*(X)$:완비통계량 $T(X)$의 함수 $\Rightarrow$ $S^*(X) = U^*(X)$.

$\therefore$, $Var(S^*(X)) = Var(U^*(X)) \le Var(U(X))$

즉, $S^*(X)$는 최소분산비편향추정량이다.

Summary : How to find the UMVUE?

(i) 비편향추정량(UE) $S^*(X)$의 분산이 크래머-라오 하한과 같을 때 $S^*(X)$는 θ의 UMVUE.

(ii) $T(X)$: 완비충분통계량이고 $h(T(X))$는 θ의 비편향추정량일 때 $h(T(X))$는 θ 의 최소분산비편향추정량이다.

(iii) $T(X)$: 완비충분통계량이고 $S(X)$는 θ의 비편향추정량일 때,

$$S^*(X) = E(S(X)|T(X))$$

는 θ 의 최소분산비편향추정량이다.

Ex 8.22 $X_1, X_2, \cdots, X_n$: random sample from the $Bin(m, \theta)$.
Find the UMVUE of θ.

Sol.

$T(X) = \sum_{i=1}^{n} X_i$: CSS for θ and $E(T(X)) = nm\theta$.

Hence, $\frac{1}{m}\frac{T(X)}{n} = \frac{1}{m}\overline{X}$ is an U.E. & function of CSS.

Therefore, by *Summary* (ii), $\frac{1}{m}\overline{X}$ is the UMVUE of θ.

Ex 8.23 $X_1, X_2, \cdots, X_n$: random sample from the $Poisson(\theta)$.

(a) Find the UMVUE of θ.

Sol.

$T(X) = \sum_{i=1}^{n} X_i$: CSS for θ and $E(T(X)) = n\theta$.

Hence, $\frac{T(X)}{n} = \overline{X}$ is an U.E. & function of CSS.

Therefore, by *Summary* (ii), $\overline{X}$ is the UMVUE of θ.

(b) Find the UMVUE of θ^2.

Sol.

$T(X) = \sum_{i=1}^{n} X_i$: CSS for θ and $T(X) \sim Poisson(n\theta)$.

Hence, $E(T(X)) = n\theta$ and $E(T(X)^2) = n^2\theta^2 + n\theta$.

$(\because\ Var(T(X)) = E(T(X)^2) - (ET(X))^2)$

That is, $E\left(\frac{T(X)^2 - T(X)}{n^2}\right) = \theta^2$.

$\frac{T(X)^2 - T(X)}{n^2} = \frac{1}{n^2}\left((\sum_{i=1}^{n} X_i)^2 - \sum_{i=1}^{n} X_i\right)$: U.E. & function of CSS.

$\therefore$, by *Summary* (ii), $\frac{1}{n^2}\left((\sum_{i=1}^{n} X_i)^2 - \sum_{i=1}^{n} X_i\right)$ is the UMVUE of θ^2.

Ex 8.24 $X_1, X_2, \cdots, X_n$: random sample from the $U(0, \theta)$.
Find the UMVUE of θ.

Sol.

$T(X) = X_{(n)}$: CSS for θ and

$$E_\theta(X_{(n)}) = \int_0^\theta tn\frac{t^{n-1}}{\theta^n}dt = \frac{n}{n+1}\theta.$$

That is, $\frac{n+1}{n}X_{(n)}$ is an U.E of θ and function of CSS.

Therefore, $\frac{n+1}{n}X_{(n)}$ is the UMVUE of θ.

Ex 8.25 $X_1, X_2, \cdots, X_n$: random sample from the $N(\mu, \sigma^2)$.

(a) Find the UMVUE of μ.

Sol. *Fill out by yourself.*

(b) Find the UMVUE of σ^2.

Sol. *Fill out by yourself.*

Example of applying *Summary* (iii) (Lehmann-Scheffe Theorem) :

$X_1, X_2, \cdots, X_n$: random sample from the *Bernoulli* (p).

Find the UMVUE of p.

Sol.

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

exponential family ==> $T(X) = \sum_{i=1}^{n} X_i$ is a C.S.S.

Let $S(X) = X_1$. Then, $S(X)$ is an U.E

and $E(S(X)|T(X))$ is the UMVUE of p.

Now, calculate $E(X_1|\sum_{i=1}^{n} X_i)$.

First, we need the conditional distribution $f(X_1|\sum_{i=1}^{n} X_i)$.

$$P(X_1 = x_1|\sum_{i=1}^{n} X_i = t) = \frac{P(X_1 = x_1, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{P(X_1 = x_1, \sum_{i=2}^{n} X_i = t - x_1)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{p^{x_1}(1-p)^{1-x_1}\binom{n-1}{t-x_1}p^{t-x_1}(1-p)^{(n-1)-(t-x_1)}}{\binom{n}{t}p^t(1-p)^{n-t}}$$

$$= \frac{\binom{n-1}{t-x_1}}{\binom{n}{t}} = \begin{cases} \dfrac{\binom{n-1}{t-1}}{\binom{n}{t}} = \dfrac{t}{n}, & x_1 = 1 \\ \dfrac{\binom{n-1}{t}}{\binom{n}{t}} = \dfrac{n-t}{n}, & x_1 = 0 \end{cases}$$

That is,

$$P(X_1 = 1 | \sum_{i=1}^{n} X_i = t) = \frac{t}{n}$$

$$P(X_1 = 0 | \sum_{i=1}^{n} X_i = t) = 1 - \frac{t}{n}.$$

Therefore,

$$E(X_1 | \sum_{i=1}^{n} X_i = t) = 1 \times P(X_1 = 1 | \sum_{i=1}^{n} X_i = t) + 0 \times P(X_1 = 0 | \sum_{i=1}^{n} X_i = t)$$

$$= \frac{t}{n}.$$

$\therefore$, $E(X_1 | \sum_{i=1}^{n} X_i = t) = \dfrac{\sum_{i=1}^{n} X_i}{n}$: UMVUE of p.

참고 : (Different solution)

$T(X) = \sum_{i=1}^{n} X_i$ is a C.S.S.

$E\left(\sum_{i=1}^{n} X_i\right) = np.$ So, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ is an U.E and function of CSS.

Therefore, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ is the UMVUE of p.

Extra Homework

1. $X_1, X_2, \cdots, X_n$이 random sample from the $Bernoulli\,(p)$.

A simple regression problem

$$Model \;:\; Y_i = \alpha + \beta x_i + \epsilon_i \;,\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Y_i = \alpha_0 + \beta(x_i - \bar{x}) + \epsilon_i \;,\; \alpha_0 = \alpha + \beta\bar{x} \;,\; \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Likelihood \;:\; L(\alpha_0, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{ -\frac{[y_i - \alpha_0 - \beta(x_i - \bar{x})]^2}{2\sigma^2} \right\}$$

$$= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{ -\frac{\sum [y_i - \alpha_0 - \beta(x_i - \bar{x})]^2}{2\sigma^2} \right\}$$

$$MLE \Leftrightarrow maximize \; L(\alpha_0, \beta, \sigma^2) \Leftrightarrow minimize \quad -\ln L(\alpha_0, \beta, \sigma^2)$$

$$-\ln L(\alpha_0, \beta, \sigma^2) = \frac{n}{2} ln(2\pi\sigma^2) + \frac{\sum [y_i - \alpha_0 - \beta(x_i - \bar{x})]^2}{2\sigma^2}$$

$MLE \; of \; \alpha_0, \beta \; minimize$

$$H(\alpha_0, \beta) = \sum_{i=1}^{n} \left[y_i - \alpha_0 - \beta(x_i - \bar{x}) \right]^2 \;:\; Least \;\; square \;\; estimates$$

$$\frac{\partial H(\alpha_0, \beta)}{\partial \alpha_0} = 2\sum \left[y_i - \alpha_0 - \beta(x_i - \bar{x}) \right](-1) = 0 \qquad (1)$$

$$\frac{\partial H(\alpha_0, \beta)}{\partial \beta} = 2\sum \left[y_i - \alpha_0 - \beta(x_i - \bar{x}) \right]\left[-(x_i - \bar{x}) \right] = 0 \qquad (2)$$

$$From \; (1), \; \sum y_i - n\alpha_0 = 0 \;,\; i.e \;\; \hat{\alpha}_0 = \frac{\sum y_i}{n} = \bar{Y}$$

$$From \; (2), \; \sum (y_i - \bar{y})(x_i - \bar{x}) - \beta \sum (x_i - \bar{x})^2 = 0$$

$$\therefore \quad \hat{\beta} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$MLE \ \ of \ \ \sigma^2$:

$$\frac{\partial\left[-\ln L(\alpha_0,\beta,\sigma^2)\right]}{\partial\sigma^2}=\frac{n}{2\sigma^2}-\frac{\left[y_i-\alpha_0-\beta(x_i-\overline{x})\right]^2}{2(\sigma^2)^2}=0$$

$$\therefore\ \hat{\sigma}^2=\frac{1}{n}\sum\left[y_i-\hat{\alpha}_0-\hat{\beta}(x_i-\overline{x})\right]^2=\frac{1}{n}\sum\left[y_i-\hat{y}_i\right]^2$$

$*Properties \ \ of \ \ \widehat{\alpha_0}, \hat{\beta}$

$$E(\widehat{\alpha_0})=E\left(\frac{1}{n}\sum Y_i\right)=\frac{1}{n}\sum E(Y_i)=\frac{1}{n}\sum_{i=1}^{n}[\alpha_0+\beta(x_i-\overline{x})]=\alpha_0$$

$$Var(\widehat{\alpha_0})=\sum_{i=1}^{n}\frac{1}{n^2}Var(Y_i)=\frac{\sigma^2}{n}$$

$$\therefore\ \widehat{\alpha_0}\ \sim\ N\left(\alpha_0,\frac{\sigma^2}{n}\right)\quad(\because\ \hat{\alpha}\ is\ a\ \text{linear comb. of } \mathrm{Y_i})$$

$$\therefore\ \frac{\widehat{\alpha_0}-\alpha}{\sigma/\sqrt{n}}\ \sim\ N(0,1)$$

since $\dfrac{\sum\left[y_i-\widehat{\alpha_0}-\hat{\beta}(x_i-\overline{x})\right]^2}{\sigma^2}\sim\chi^2_{(n-2)}$,

$$\frac{\sqrt{n}(\widehat{\alpha_0}-\alpha)/\sigma}{\sqrt{\dfrac{n\widehat{\sigma^2}}{\sigma^2(n-2)}}}=\frac{\widehat{\alpha_0}-\alpha}{\sqrt{\dfrac{\widehat{\sigma^2}}{n-2}}}\ \sim\ t(n-2)$$

$$\therefore\quad 100(1-\alpha)\%\ \ C.I.\ \ of\ \ \alpha_0\ :\ \widehat{\alpha_0}\pm t_{\alpha/2}(n-2)$$

$$E(\hat{\beta})=\frac{\sum(x_i-\overline{x})E(Y_i)}{\sum(x_i-\overline{x})^2}=\frac{\sum(x_i-\overline{x})\left[\alpha+\beta(x_i-\overline{x})\right]}{\sum(x_i-\overline{x})^2}$$

$$=\frac{\alpha\sum(x_i-\overline{x})+\beta\sum(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}=\beta$$

$$Var(\hat{\beta}) = \sum_{i=1}^{n}\left[\frac{x_i - \overline{x}}{\sum_{j=1}^{n}(x_j - \overline{x})^2}\right]^2 Var(Y_i)$$

$$= \frac{\sum(x_i - \overline{x})^2}{\left[\sum(x_i - \overline{x})^2\right]^2}\,\sigma^2 = \frac{\sigma^2}{\sum(x_i - \overline{x})^2}$$

$$\therefore\ \hat{\beta} \sim N\left(\beta\,,\, \frac{\sigma^2}{\sum(x_i - \overline{x})^2}\right) \quad (\because\ \hat{\beta}\ is\ a\ lin.\ comb.\ of\ Y_i)$$

$$\therefore\ \frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum(x_i - \overline{x})^2}}} \sim N(0\,,1)$$

$$\therefore\ \frac{(\hat{\beta} - \beta)/\sqrt{\sigma^2}}{\sqrt{\frac{n\hat{\sigma^2}}{\sigma^2/(n-2)}}} = \frac{\hat{\beta} - \beta}{\sqrt{\frac{n\hat{\sigma^2}}{(n-2)\sum(x_i - \overline{x})^2}}} \sim t(n-2)$$

$\therefore\ 100(1-\alpha)\%$ $C.I.$ of β

$$:\ \hat{\beta} \pm t_{\alpha/2}(n-2)\sqrt{\frac{n\hat{\sigma}^2}{(n-2)\sum(x_i - \overline{x})^2}}$$