ep 5. $X \sim N(\mu, \sigma^2)$

$\hat{\mu}_{MLE} = \bar{X}$

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

$\frac{\hat{\mu}_{MLE} - \mu}{S} \xrightarrow{P} t_{n-1}$ where $S = \sqrt{\frac{1}{n-1}(X_i - \bar{X})^2}$

So Confidence interval for μ.

$\hat{\mu} \pm \sqrt{n}\, t_{n-1}(\frac{\alpha}{2}) S$ is a $(1-\alpha)100\%$ CI for μ

$P\left[\frac{n\hat{\sigma}^2_{MLE}}{\sigma^2} \in [0, \chi^2_{n-1}(\alpha)]\right] = \alpha$.

For σ^2,

$\frac{n\hat{\sigma}^2_{MLE}}{\sigma^2} \xrightarrow{P} \chi^2_{n-1}$

Let $\chi^2_{n-1}(\alpha)$ be the α-th. upper quatile of a χ^2_{n-1} distribution. Then a $100\alpha\%$. CI for σ^2 is

$$\left[\frac{n\sigma^2_{mLE}}{\chi_{n-1}(\alpha)}, \infty\right]$$

2. The use of theorem B (from last lecture) to construct CI's for θ.

Then $\sqrt{nI(\theta)}\,(\hat{\theta}_{MLE} - \theta) \longrightarrow N(0,1)$

Collollary: A $(1-\alpha)100\%$ CI for θ is

$$\hat{\theta}_{MLE} \pm z(1-\tfrac{\alpha}{2}) / \sqrt{nI(\hat{\theta}_{MLE})}$$

$\longrightarrow$ normal cut of. fat. $(1-\frac{\alpha}{2})$ quatile.

Ex1:

Poisson case.

$X_1 \ldots X_n$'s are i.i.d. Poisson(λ)

$$P[X_i = x] = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$L(\theta) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod x_i!} e^{-n\lambda}.$$

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

$$l'(\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n.$$

$$\Rightarrow \hat{\lambda}_{MLE} = \bar{x}$$

$$l''(\lambda) = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} \Rightarrow -E[l''(\lambda)] = \left[\frac{E\sum_{i=1}^{n} x_i}{\lambda^2}\right]$$

$$= \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

$$\boxed{-E[l''(\lambda)]} = nI(\lambda) \quad \text{[Lemma from last lecture]}.$$

Therefore a 95% CI for λ is $\hat{\lambda}_{MLE} \pm 1.96 / \sqrt{n/\hat{\lambda}_{MLE}}$

$$= \hat{\lambda}_{MLE} \pm 1.96 \sqrt{\hat{\lambda}_{MLE}} / \sqrt{n}.$$

1 obs.

$\log f(x|\theta) = l(\theta)$

n obs.

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

Ex 2:

Hardy - Weinberg Equilibrium.

AA Aa aa

$(1-\theta)^2$ $2\theta(1-\theta)$ θ^2.

n subjects X_1 = total number of AA's

X_2 = total number of Aa's

X_3 = total number of aa's.

$$L(\theta) = \frac{n!}{X_1!\,X_2!\,X_3!}(1-\theta)^{2X_1}\left[2\theta(1-\theta)\right]^{X_2}(\theta^2)^{X_3}.$$

$$l(\theta) = \log\left[\frac{n!}{n_1!\,n_2!\,n_3!}\right] + 2X_1\log(1-\theta) + X_2\log(2\theta(1-\theta))$$
$$+\ 2X_3\log\theta.$$

$$l'(\theta) = \frac{-2X_1}{1-\theta} + \frac{X_2}{\theta} - \frac{X_2}{1-\theta} + \frac{2X_3}{\theta}.$$

$$\hat{\theta}_{MLE} = \frac{2X_3+X_2}{n}.$$

$$l''(\theta) = \frac{-2X_1}{(1-\theta)^2} - \frac{X_2}{\theta^2} - \frac{X_2}{(1-\theta)^2} - \frac{2X_3}{\theta^2}.$$

$$-E^{\theta}[l''(\theta)] = \frac{2E[X_1]}{(1-\theta)^2} + \frac{E[X_2]}{\theta^2} + \frac{E[X_2]}{(1-\theta)^2} + \frac{2E[X_3]}{\theta^2}.$$

$$= \frac{2n(1-\theta)^2}{(1-\theta)^2} + \frac{2\theta(1-\theta)n}{\theta^2} + \frac{2\theta(1-\theta)n}{(1-\theta)^2} + \frac{2n\theta^2}{\theta^2}.$$

$$2n + \frac{2(1-\theta)}{\theta}n + \frac{2\theta}{(1-\theta)}n + 2n = 2n\left[2 + \frac{2\theta^2-2\theta+1}{\theta(1-\theta)}\right].$$

$$= 2n/\theta(1-\theta)$$

Hence $nI(\theta) = 2n/\theta(1-\theta)$

A 95% CI for θ is $\hat{\theta}_{MLE} \pm 1.96/\sqrt{2n/\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}$.

For instance If AA Aa aa
50 25 25.

$$\hat{\theta}_{MLE} = \frac{2X_3 + X_2}{2n} = \frac{3}{8}$$

95% CI for θ in this case is $\frac{3}{8} \pm \left(1.96 / \sqrt{100}\right) \cdot \sqrt{\frac{\frac{3}{8}(1-\frac{3}{8})}{2}}$.

8.6. $$P[A|B] = \frac{P(B|A)\,P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{\sum P(B|A_i)\cdot P(A_i)}$$

8.7. Efficiency and Cramer-Rao Lower Bound.

Suppose $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of θ.
Then the relate efficiency of $\hat{\theta}_1$ versus $\hat{\theta}_2$ is defined as

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

$\Rightarrow$ eff$(\hat{\theta}_1, \hat{\theta}_2)$ is ratio of sample sizes for $\hat{\theta}_1$ and $\hat{\theta}_2$ to achieve the same accuracy.

$$\sqrt{n}\,(\hat{\theta}_1 - \theta) \longrightarrow N\left(0, \frac{n_1 Var(\hat{\theta}_1)}{}\right)$$

$$\sqrt{n}\,(\hat{\theta}_2 - \theta) \longrightarrow N\left(0, \frac{n_2 Var(\hat{\theta}_2)}{}\right)$$

$$\boxed{\frac{Var(\hat{\theta}_1)}{n_1} \approx \frac{Var(\hat{\theta}_2)}{n_2}}$$

$$Var(\hat{\theta}_2) \approx \text{eff}(\hat{\theta}_1, \hat{\theta}_2) \cdot Var(\hat{\theta}_1)$$

$$N\left(0, \frac{n_2 Var(\hat{\theta}_1) \cdot \text{eff}(\hat{\theta}_1, \hat{\theta}_2)}{}\right)$$

Now if $n_1 = n_2$,

$$MSE(\hat{\theta}_1) = Var(\hat{\theta}_1)$$

$$MSE[\hat{\theta}_2] = Var(\hat{\theta}_2) = eff(\hat{\theta}_1, \hat{\theta}_2)\, Var(\hat{\theta}_1)$$

$$MSE[\hat{\theta}_1] / MSE[\hat{\theta}_2] = 1 / eff(\hat{\theta}_1, \hat{\theta}_2)$$

Now suppose the sample size is increasing by a factor of $eff(\hat{\theta}_1, \hat{\theta}_2)$. Then $MSE[\hat{\theta}_2]$ in the new sample size will decrease by a factor of $eff(\hat{\theta}_1, \hat{\theta}_2)$

Hence, the $MSE(\hat{\theta}_2)$ in the new sample size $= MSE(\hat{\theta}_1)$ in the old sample size $\square$.

Ex: Estimating a population Mean.

$X_1 \dots X_n$ are i.i.d $E[X_i] = \mu$.

$Var(X_i) = \sigma^2 < \infty$.

$$\hat{\mu}_{(a_1 \cdots a_n)} = a_1X_1 + a_2X_2 + \cdots + a_nX_n.$$

where $a_1 + \cdots + a_n = 1$

(1) all $\mu_{(a_1 \cdots a_n)}$ are unbiased.

Because $E[\hat{\mu}_{(a_1 \cdots a_n)}] = a_1\mu + \cdots a_n\mu.$

$= \mu$

~~$Var(\hat{\mu}_{(a_1 \cdots a_n)})$~~

$$Var(\hat{\mu}_{(a_1 \cdots a_n)}) = a_1^2\sigma^2 + a_2^2\sigma^2 + \cdots a_n^2\sigma^2$$
$$= (a_1^2 + \cdots + a_n^2)\sigma^2.$$

(2) Suppose I have two sequences $a_1^*, a_2^*, \dots, a_n^*$ and $a_1^\circ, \dots, a_n^\circ$

Then the relative efficiency: $eff(\hat{\mu}_{(a_1^* \cdots a_n^*)}, \hat{\mu}_{(a_1^\circ \cdots a_n^\circ)})$

$$= \frac{(a_1^*)^2 + (a_2^*)^2 \cdots + (a_n^*)^2}{(a_1^\circ)^2 + \cdots + (a_n^\circ)^2}.$$

Note $\frac{a_1 + \cdots + a_n}{n} \le \sqrt{\frac{a_1^2 + \cdots + a_n^2}{n}}$ (*) ~~(exercise)~~

For any real numbers

$$\sqrt{\frac{a_1^2 + \cdots + a_n^2}{n}} \geq \frac{1}{n} \text{ if } a_1 + \cdots + a_n = 1.$$

$\Rightarrow Var(\quad) \geq \frac{1}{n}\sigma^2$.

and the equality holds in (*) iff
$a_1 = a_2 = \cdots = a_n$.

Hence the sample mean is the most efficient in this class of estimator.

— Cramer- Rao Inequality

Theorem: If $X_1 \cdots X_n$'s are i.i.d. with density $f(x|\theta)$. Suppose $\hat{\theta} = T(X_1 \cdots X_n)$ is an unbiased estimator of θ. Then under some smoothness assumption on $f(x|\theta)$ we have

$$Var(\hat{\theta}) \geq \frac{1}{nI(\theta)} \leftarrow \text{Cramer-Rao Lower Bound.}$$

(1). Note

$$Var(\hat{\theta}_{MLE}) \approx \frac{1}{nI(\theta)}$$

Hence $\hat{\theta}_{MLE}$ is ~~ass~~ asumptotically efficients.

Proof: $Z = \frac{\partial}{\partial\theta}\left[\sum_{i=1}^{n} \log f(X_i|\theta)\right]$

Now we see that $|corr(Z, T)| \le 1$

$\Rightarrow Cov^2(Z, T) \le Var(T)\, Var(Z)$

$Var(Z) = nI(\theta)$

$$Var(T) \geq \frac{Cov^2(Z,T)}{nI(\theta)}$$

$$var(\hat{\theta}) \geq \frac{Cov^2(Z,T)}{nI(\theta)}.$$

prove $Cov(T,Z) = 1$.

Note that $E[Z] = 0$ (last lecture)

Hence, $Cov(T,Z) = E[TZ]$.

$$= \int T(x_1 \cdots x_n)\, Z(x_1 \cdots x_n)\, f(x_1 \cdots x_n|\theta)\, dx_1 \cdots dx_n.$$

$$= \int T(x_1 \cdots x_n) \frac{\partial}{\partial\theta}\left[\sum \log f(x_i|\theta)\, f(x_2|\theta) \cdots f(x_n|\theta) \, dx_1 \cdots dx_n\right].$$

Note : $\frac{\partial}{\partial\theta}\left[\sum_{i=1}^{n} \log f(x_i|\theta)\right] \prod_{i=1}^{n} f(x_i|\theta)$ ①

$$= \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i|\theta) \quad ②$$

Proof : $\frac{\partial}{\partial\theta}\left[\sum_{i=1}^{n} \log f(x_i|\theta)\right] = \sum_{i=1}^{n} \frac{f'(x_i|\theta)}{f(x_i|\theta)}$

$$\Rightarrow \left[\frac{\partial}{\partial\theta} \sum_{i=1}^{n} \log f(x_i|\theta)\right] \prod_{i=1}^{n} f(x_i|\theta)$$

$$\Rightarrow \sum_{i=1}^{n} \frac{f'(x_i|\theta)}{f(x_i|\theta)} \prod_{i=1}^{n} f(x_i|\theta) \Rightarrow ② \text{ holds.}$$

$$Cov(T,Z) = \int T(x_1 \cdots x_n) \frac{\partial}{\partial\theta} f(x_1 \cdots x_n|\theta)\, dx_1 \cdots dx_n$$

$$= \frac{\partial}{\partial\theta} \int T(x_1 \cdots x_n)\, f(x_1 \cdots x_n|\theta)\, dx_1 \cdots dx_n.$$

$$= \frac{\partial}{\partial\theta}\left[E\,T(x_1 \cdots x_n)\right].$$

$= \frac{\partial}{\partial \theta}(\theta) = 1 \Rightarrow Var(\hat{\theta}) \geq \frac{1}{nI(\theta)}$

Ex: Efficiency of Poisson mean.

$X_1 \cdots X_n$ are Poisson(λ)

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. have the following variance.

$Var(\bar{X}) = \frac{1}{n} Var(X_i) = \frac{\lambda}{n}$.

According Cramer-Rao inequality for any unbiased estimator $\hat{\lambda}$ of λ.

$$Var(\hat{\lambda}) \geq \frac{1}{nI(\lambda)} = \frac{\lambda}{n}.$$

8.8. Sufficient Statistics

Def: A statistic $T(X_1 \cdots X_n)$ is called a sufficient Statistic for θ if the conditional distribution of $X_1 \cdots X_n$ given $T = t$ does not depend on θ. for any t

Theorem (Factorization Thm):

$T(X_1 \cdots X_n)$ is a ~~sufficient~~ Sufficient Statistic for θ if and only if

$$f(X_1 \cdots X_n | \theta) = g(T(X_1 \cdots X_n), \theta) h(X_1 \cdots X_n) \quad (*)$$

no θ in h function.

Proof: Firstly prove that if $(*)$ holds, then T is sufficient

We only ~~proof~~ prove the case for discrete r.v.'s.

For any t

$$P(T(x_1 \cdots x_n) = t) = \sum_{T(x_1 \cdots x_n) = t} P((X_1 \cdots X_n) = (x_1 \cdots x_n))$$

$$\overset{*}{=} g(t, \theta) \sum_{T(x_1 \cdots x_n)} h(x_1 \cdots x_n)$$

Therefore $P[X_1 = x_1 \cdots X_n = x_n \mid T = t]$

$$= \frac{P[X_1 = x_1 \cdots X_n = x_n \; T = t]}{P[T = t]}$$

$$\overset{*}{=} \frac{\cancel{g(t,\theta)}\, h(x_1 \cdots x_n)}{\cancel{g(t,\theta)} \sum_{T(x_1 \cdots x_n) = t} h(x_1 \cdots x_n)} = \frac{h(x_1 \cdots x_n)}{\sum_{T(x_1 \cdots x_n) = t} h(x_1 \cdots x_t)}$$

doesn't depend on θ.

Second we prove that if T is sufficient, then $\circledast$ holds.

Let $t = T(x_1 \cdots x_n)$

Let ~~$h(x) =$~~ $h(x_1 \cdots x_n) = P[X_1 = x_1 \cdots X_n = x_n \mid T = t]$

Beause of assumption h does not depend on θ.

Let $g(t, \theta) = P[T = t]$.

hence $P[X_1 = x_1 \cdots X_n = x_n \mid \theta]$

$$= P[X_1 = x_1 \cdots X_n = x_n \mid T = t] \cdot P(T = t)$$

$$= h(x_1 \cdots x_n)\, g(t, \theta)$$

Prob... landing heads $= \theta$.

Ex: Toss an unfair coin for n times.

$$Pmf = \theta^{x_1}(1-\theta)^{1-x_1} \cdots\cdots \theta^{x_n}(1-\theta)^{1-x_n}.$$

$$= \theta^{x_1+\cdots x_n}(1-\theta)^{n-(x_1+\cdots x_n)}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{x_1+\cdots+x_n}(1-\theta)^n \Rightarrow x_1+\cdots+x_n \text{ is sufficient for } \theta.$$

About midterm:

coverage:

Chapter 7 (1.2.3.4)

Chapter 8 (1.2, 3, 4.5, 7, 8.8 part. to P309)