chapter 8 :《Likelihood theory》

<< March 18, 2015 >> STAT 846.

Chapter8: << Likelihood theory >>

$\Rightarrow$ $L(\theta; x) = f(x \mid \theta)$; L

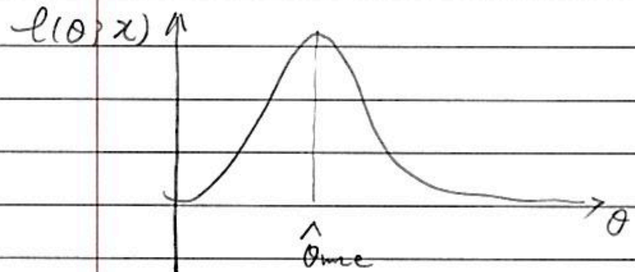
(1) score function :

log-likelihood :

$\ell(\theta, x) = \log L(\theta, x) = \log f(x \mid \theta)$

score function : $S(\theta; x) = \nabla_\theta \ell(\theta; x) = \frac{\partial}{\partial \theta} \log L(\theta, x)$

Maximum likelihood estimator (MLE)



$\hat{\theta}_{mle}(x) = \underset{\theta}{\arg\max}\, \ell(\theta; x)$

$= \arg\max L(\theta, x)$

$\Rightarrow$ An approach to finding $\hat{\theta}_{mle}$

to solve $\nabla_\theta \ell(\theta, x) = 0$

$\Rightarrow$ example 1)

1) $X_1, \dots X_n \mid \theta \sim unif(0, \theta)$

$L(\theta; x) = \prod_{i=1}^{n} f(x_i \mid \theta_i)$

$= \frac{1}{\theta^n} I(X_{(n)} < \theta)$ where $X_{(n)} = \max\{X_1, \dots X_n\}$

$\hat{\theta}_{mle}(x) = X_{(n)}$

$E\,(X_{(n)})$

$= \int_0^{\theta} n \cdot x \left(\frac{x}{\theta}\right)^{n-1} dx$

$= n \cdot \frac{1}{\theta^{n-1}} \cdot \int_0^{\theta} x^n\, dx$

$= \frac{n}{\theta^{n-1}} \cdot \left.\frac{x^{n+1}}{n+1}\right|_0^{\theta} = \frac{n\theta}{n+1} < \theta$

$P((X_{(n)}) \le x \mid \theta)$

$= \left(\frac{x}{\theta}\right)^n$

$\cancel{f_{X_{(n)}}(x) = n \cdot \left(\frac{x}{\theta}\right)^{n-1} \text{ for } x<0}$

$\Rightarrow$ example 2:

$x_1, \cdots x_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$

Likelihood: $L(\theta, x) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}\right)$

log-likelihood. $\ell(\theta, x) = -\frac{n}{2} \cdot \log 2\pi\sigma^2 - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$

Score-function:

$$\frac{\partial L}{\partial \mu} = -\frac{2\,\Sigma(x_i-\mu)}{2\sigma^2} = \frac{\Sigma(x_i-\mu)}{\sigma^2} = 0$$

$$\Rightarrow \mu = \Sigma x_i / n = \bar{x} \quad \Rightarrow \hat{\mu}_{mle} = \mu$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{\partial(2l)}{\partial \sigma^2} = -\Sigma(x_i-\mu)^2 \cdot (\sigma^2)^{-2}(-1) - \frac{n}{\sigma^2} = 0$$

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

$$\sigma^2 = \Sigma(x_i-\bar{x})^2/n \Rightarrow \hat{\sigma}^2_{mle} = \frac{\Sigma(x_i-\bar{x})^2}{n}$$

Note: $E(\bar{x}) = \mu$

$Var(\bar{X}) = \frac{\sigma^2}{n}$

? $E(\hat{\sigma}^2_{mle}) = \frac{1}{n} \left[\frac{n-1}{n} \cdot \sigma^2 \right]$

$S^2 = \Sigma (x_i - \bar{x})^2 / n-1$

$(n-1) S^2 / \sigma^2 \sim \chi^2_{n-1} = \text{Gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right)$

$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1$

$E(S^2) = \sigma^2$; $Var\left(\frac{n-1}{\sigma^2} S^2\right) = 2(n-1)$; $V(S^2) = \left(\frac{\sigma^2}{n-1}\right)^2 \cdot 2(n-1)$

$E(\hat{\sigma}^2_{mle}) = E\left(\frac{n-1}{n} \cdot \frac{\Sigma(x_i - \bar{x})^2}{n-1}\right) = E\left(\frac{n-1}{n} \cdot S^2\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$

$\Rightarrow MSE(\hat{\sigma}^2_{mle})$

$= Var(\hat{\sigma}^2_{mle}) + Bias(\hat{\sigma}^2_{mle})^2$

$= \left(\frac{n-1}{n}\right)^2 \cdot \frac{2(\sigma^2)^2}{n-1} + \left(\frac{n-1}{n} \cdot \sigma^2 - \sigma^2\right)^2$

$= \left[\frac{n-1}{n^2} \times 2 + \frac{1}{n^2}\right](\sigma^2)^2$ $\quad 2n-1 / n^2$

$= \left(\frac{n-1}{n^2} \times 2 + \frac{1}{n^2}\right)(\sigma^2)^2 = \frac{3n-2}{n^2} \cdot (\sigma^2)^2$

? $MSE(S^2)$

$= Var(S^2)$ $\qquad \frac{3n-2}{n^2} < \frac{2}{n-1}$

$= \frac{2}{n-1} (\sigma^2)^2$

$MSE(\hat{\sigma}^2_{mle}) \leq MSE(S^2)$

$\Rightarrow$ Some properties about $S(Q, x)$

$$E\{S(Q; x) \mid \theta\} = 0$$

$$S_i(\theta; x) = \frac{\partial l(\theta; x)}{\partial \theta_i}$$

$$E\{S_i(\theta; \theta) \mid \theta\} = 0$$

$$\mathrm{cov}(S_i(\theta; x), S_j(\theta; x))$$

$$= E(S_i(\theta; x), S_j(\theta, x))$$

$$= -E\left(\frac{\partial^2 l(\theta; x)}{\partial \theta_i \partial \theta_j}\right)$$

$$I(\theta) = \mathrm{cov}(S(\theta, x))$$

$$= \begin{pmatrix} \mathrm{cov}(S_1, S_1) & \cdots & \mathrm{cov}(S_1, S_d) \\ \vdots & & \\ \mathrm{cov}(S_d, S_1) & \cdots & \mathrm{cov}(S_d, S_d) \end{pmatrix}$$

$I(\theta)$ is a function of θ alone fisher in function

$$\dot{J}(\theta, x) = \left(-\frac{\partial^2}{\partial \theta_k \partial \theta_l} l(\theta; x)\right)$$

$$I(\theta) = E(\dot{J}(\theta; x))$$

$\dot{J}(\theta, x)$ - observed. Fisher information.

<< Mar 20, 2015 >>   STAT 846.

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

$\hat{\mu}_{mle} = \bar{X}$ $\qquad \hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\hat{\sigma}^2_{mle} = \frac{n-1}{n} \cdot \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{n-1}{n} S^2$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \qquad S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1}$

$E(S^2) = \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2$

$Var(S^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2}{n-1}\sigma^4$

$MSE(S^2) = \frac{2}{n-1}\sigma^4$

$$
\begin{aligned}
MSE(\hat{\sigma}^2_{mle}) &= Var(\hat{\sigma}^2_{mle}) + [Bias(\hat{\sigma}^2_{mle})]^2 \\
&= \left(\frac{n-1}{n}\right)^2 V(S^2) + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\
&= \frac{(n-1)^2}{n^2}\frac{2}{n-1}\sigma^4 + \frac{1}{n^2}\sigma^4 \\
&= \left(\frac{2}{n} - \frac{1}{n^2}\right)\sigma^4 \\
&< \frac{2}{n-1}\sigma^4
\end{aligned}
$$

so, $\hat{\sigma}^2_{mle}$ ~~do~~ dominates S^2

$\Rightarrow$ $S_i(\theta; x) = \frac{\partial}{\partial\theta} l(\theta; x)$

theorem: Given the support of X is free of θ

$E_X(S_i(\theta; x) \mid \theta) = 0$

$COV(S_i(\theta; x)\ (S_j(\theta; x)) = E\left(-\frac{\partial l(\theta, x)}{\partial\theta_i} \frac{\partial}{\partial\theta_j} l(\theta, x)\right)$

$= I(\theta)$

proof:

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta, x) = \frac{\partial^2}{\partial\theta_i\partial\theta_j} \log(f(x;\theta))$$

$$= \frac{\partial}{\partial\theta_j}\left(\frac{\frac{\partial}{\partial\theta_j} f(x|\theta)}{f(x|\theta)} \right)$$

$$= \frac{\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x|\theta)\cdot f(x|\theta) - \frac{\partial}{\partial\theta_i} f(x|\theta)\cdot\frac{\partial}{\partial\theta_j} f(x|\theta)}{f^2(x|\theta)}$$

For the R.H.S

$$= E\left(-\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta, x)\right)$$

$$= -\left[\int \frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x|\theta)\, dx - \int \frac{\partial}{\partial\theta_i} f \cdot \frac{\partial}{\partial\theta_j} f \cdot \frac{1}{f(x|\theta)}\, dx \right.$$

$$= \int \frac{\partial f}{\partial\theta_i}\cdot\frac{\partial f}{\partial\theta_j}\, \frac{1}{f(x|\theta)}\, dx$$

For the L.H.S.

$$= Cov(S_i, S_j)$$

$$= E(S_i, S_j) = \int \frac{\partial}{\partial\theta_i} l(\theta, x)\cdot\frac{\partial}{\partial\theta_j} l(\theta, x)\cdot f(x|\theta)\, dx$$

$$= \int \frac{\partial}{\partial\theta_i} f\cdot\frac{\partial f}{\partial\theta_j}\cdot\frac{1}{f(x|\theta)}\, dx.$$

=> Remark: if $X_1, \cdots X_n | \theta \overset{iid}{\sim} f(X_i|\theta)$

Let X to be $X = (X_1 \cdots X_n)$

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

$$l(\theta; x) = \Sigma \log f(x_i|\theta) = \Sigma\, l(\theta; x_i)$$

$S_k(\theta; x) = \frac{\partial}{\partial \theta_k} \ell(\theta; x)$

$= \frac{\partial}{\partial \theta_k} \Sigma \ell(\theta; x_i) = \Sigma \frac{\partial}{\partial \theta_k} \ell(\theta; x_i)$

$I^{k,k}(\theta)$

$\Rightarrow Var(S_k(\theta; x)) = E\left(\Sigma \frac{\partial^2}{\partial \theta_k^2} \ell(\theta, x_i)\right)$

$= \sum_{i=1}^{n} E\left(-\frac{\partial^2}{\partial \theta_k^2} \ell(\theta; x_i)\right) = \sum_{i=1}^{n} I_1^{k,k}(\theta) = n\, I_1(\theta)$

where, $I_1^{k,k}(\theta) = E\left(-\frac{\partial^2}{\partial \theta_k^2} \ell(\theta, x_i)\right)$

⇒ Cramer - Rao Lower bound (CRLB)

Suppose the support of X, Let θ be scalar.

Thm: $\hat{}$ $W(X)$ is any estimator,

Let $m(\theta) = E(W(X))$, then

$$Var(W(X)) \geq \frac{m'(\theta)}{I(\theta)}$$

where $I(\theta) = E_X\left(-\frac{\partial^2}{\partial \theta^2} \ell(\theta; x)\right)$

⇒ particularly case. Let $m(\theta) = \theta$

For any unbiased estimator of θ, $W(X)$

$Var(W(x)) \geq \frac{1}{I(\theta)}$.

Note that. $\frac{m'(\theta)}{I(\theta)}$ is called CRLB.

⇒ proof: Let $Z(X) = S(\theta; X) = \frac{\partial}{\partial \theta} \ell(\theta; X)$

We know that

$|Cov(X, Y)|^2 \leq Var(X) \cdot Var(Y)$

$Cov(W(X), Z(X)) \leq Var(W(X)) \cdot Var(Z(X))$

$V(Z(x)) = V(S(\theta; x) = E(\frac{\partial}{\partial\theta} \ell(\theta, x)) = I(\theta)$

$\mathrm{Cov}(W(x), S(\theta; x))$

$= E(W(x) \cdot S(\theta; x))$ $\quad$ (b/c $E(S) = 0$)

$\qquad\qquad\qquad\qquad E(W(x)) \cdot E(S(\theta; x)) = 0$

$= \int W(x) \cdot S(\theta; x)\, dx$

$= \int W(x) \frac{\partial f(x;\theta)}{\partial\theta} \cdot \frac{1}{f(x;\theta)} f(x;\theta)\, dx$

$= \int W(x) \frac{\partial f(x;\theta)}{\partial\theta}\, dx$

$= \frac{\partial}{\partial\theta} \left\{ \int W(x) \cdot f(x;\theta)\, dx \right\}$

$= \frac{\partial}{\partial\theta}(m(\theta)) = m'(\theta)$

therefore

$$[m'(\theta)] \le \mathrm{Var}(W(x)) \cdot I(\theta)$$

$\Rightarrow \mathrm{Var}(W(x)) \geq \frac{m'(\theta)}{I(\theta)}$

$\Rightarrow$ example: $\quad x_1 \cdots x_n \overset{iid}{\sim} \exp(\theta)$

where θ is the scaler of x_i

Let $x = (x_1, \cdots x_n)$

$f(x;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}}$

$\ell(\theta; x) = -n \log\theta - \frac{1}{\theta} \Sigma x_i$

$S(\theta; x) = \frac{\Sigma x_i}{\theta^2} - \frac{n}{\theta} = 0 \Leftrightarrow \theta = \bar{x}$

$\hat{\theta}_{mle} = \bar{x}$

$I(\theta) = \mathrm{Var}\left(\frac{\Sigma x_i}{\theta^2} - \frac{n}{\theta} \right) = \frac{1}{\theta^4} n \cdot \mathrm{Var}(x_i) \cdot \theta^2 = \frac{n}{\theta^2}$

$$S'(\theta; x) = \frac{n}{\theta^2} - \frac{2\sum x_i}{\theta^3}$$

$$E(-S'(\theta; X)) = \frac{2}{n^3} \cdot n\theta - \frac{n}{\theta^2} = \frac{n}{\theta^2}$$

mLe $E(\bar{X}) = \theta$

$$Var(\hat{\theta}_{mle}) = Var(\bar{X}) = \frac{1}{n} Var(X_i) = \frac{1}{n}\theta^2$$

Here $Var(\hat{\theta}_{mle}) = \frac{1}{I(\theta)}$

$$CRLB = \frac{1}{I(\theta)}$$

<< Mar 23. 2015 >> STAT 846.

⇒ Asymptotic properties of MLE

$\hat{\theta}_n(x_1, \cdots, x_n) \to \theta$

$f_{\hat{\theta}}(x) \xrightarrow{dis.} f(x)$

☆ $\boxed{\hat{\theta}_n \to N\left(\theta, \frac{1}{I(\theta)}\right)}$ where $I(\theta) = Var(S(\theta(x)))$

$= -E(\underbrace{\nabla \ell(\theta, x)}_{d \times d \text{ matrix}})$

⇒ Review :

1) $X_n \xrightarrow{prob.} X \iff \forall \varepsilon > 0 \; P(|X_n - X| > \varepsilon) \to 0$

2) $X_n \xrightarrow{distri.} X$; $F_{X_n}(x) = P(X_n \le x)$

$F_X(x) = P(X \le x)$

$X_n \xrightarrow{dis.} X \iff F_{X_n}(x) \to F_X(x)$, the continuity point of $F_{X_n}(x)$

3). Central Limit theorem and Large Number Law.

this is weak

LLN: $X_1, \cdots X_n | \theta \overset{iid}{\sim} f(x; \theta)$

$E(X_i) < +\infty$ then $\bar{X} = \sum X_i / n \xrightarrow{P/d} E(X_i)$

this is stronger

4) CLT: Given $X_1, \cdots, X_n | \theta \overset{iid}{\sim} f(x|\theta)$

$E(X_i) = \mu < \infty$

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

Then, $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$

$\Rightarrow$ slutsky's theorem:

if $X_n \xrightarrow{d} X$

$\quad Y_n \xrightarrow{d} a$

a is a non-Random Number without assuming

any relationship btw X_n and Y_n; then

$$g(X_n, Y_n) \xrightarrow{d} g(X, a)$$

In particular:

$$X_n + Y_n \rightarrow X + a$$

$$X_n \cdot Y_n \rightarrow X \cdot a$$

$$X_n / Y_n \rightarrow X / a \quad (a \neq 0)$$

$$X_n + \log(Y_n) \xrightarrow{d} X + \log(a)$$

Example:

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

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

proof: By LLN $\hat{\mu} = \bar{X} \xrightarrow{P} E(X_i) = \mu$

$$\hat{\sigma}^2 = \frac{1}{n}\left(\Sigma X_i^2 - n\bar{X}^2\right) = \frac{\Sigma X_i^2}{n} - \bar{X}^2$$

By LLN (Law Learge Number)

$$\Sigma X_i^2 / n \xrightarrow{P} E(X_i^2) = \sigma^2 + \mu^2$$

$$\bar{X} \xrightarrow{P} \mu$$

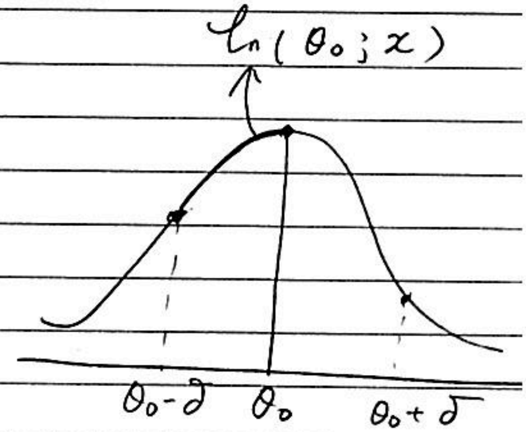
By slutsky's theorem.

$$\hat{\sigma}^2 \rightarrow \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

$\Rightarrow$ Thm: Given $x_1 \cdots x_n \mid \theta \overset{iid}{\sim} f(x \mid \theta)$

$\hat{\theta}_n$ is the MLE biased on $x_1, \cdots, x_n$

$\hat{\theta}_n \xrightarrow{P} \theta$



proof: Let $X = (x_1, \cdots x_n)$

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

For any fixed θ

$l_n(\theta; x)$ is a R.V

indeed is a sample mean of $Z_i = l(\theta; x_i)$

$= \log[f(x_i \mid \theta)]$

we will show that

$[l_n(\theta_0 - \delta) - l_n(\theta_0))]/n \to \mu_1 < 0$

$[l_n(\theta_0 + \delta) - l_n(\theta_0)]/n \to \mu_2 < 0$

proof: $\forall\ \theta \neq \theta_0$

$[l(\theta; x_n) - l_n(\theta_n; x_n)]/n$

$= [\Sigma[l(\theta; x_i) - l(\theta_0; x_i)]]/n$

$\xrightarrow{LLN} E[(l(\theta; x_i) - l(\theta_0; x_i)) \mid \theta]$

$E\left(\log \frac{f(x_i \mid \theta)}{f(x_i \mid \theta_0)} \,\middle|\, \theta_0\right) \le \log E\left(\frac{f(x_i \mid \theta)}{f(x_i \mid \theta_0)} \,\middle|\, \theta\right)$

$= \log\left\{\int \frac{f(x \mid \theta)}{f(x \mid \theta_0)} \cdot f(x \mid \theta_0)\, dx\right) = \log\left(\int f(x \mid \theta)\, dx\right)$

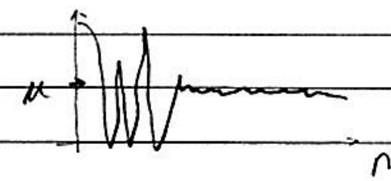
$= \log 1 = 0$

<< Mar. 25 2015 >> STAT 846

=> Review:

1) LLN (Law large Number)

$X_1 \cdots x_n \mid \theta \overset{iid}{\sim} f(x\mid\theta)$



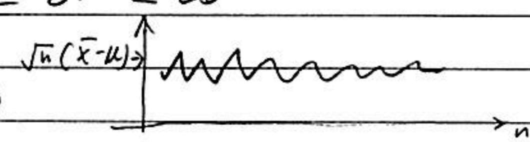
$E(x_i) = \mu < +\infty \qquad \bar{X}_n \xrightarrow{d/p} \mu$

2) CLT:

$X_1, \cdots x_n \mid \theta \overset{iid}{\sim} f(x\mid\theta)$

$E(x_i) = \mu < \infty \qquad V(x_i) = \sigma^2 < \infty$

$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$



3). Slutsky's theorem:

Given $X_n \xrightarrow{d} X \qquad Y_n \xrightarrow{d} C$

then: $X_n + Y_n \xrightarrow{d} X + C$

$X_n \cdot Y_n \xrightarrow{d} X \cdot C$

$X_n / Y_n \xrightarrow{d} X / C$

=> example: $X_1 \cdots x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \Sigma(x_i - \bar{x})^2 / n \end{cases}$$

$\hat{\mu} \xrightarrow{d/p} \mu$ by (LLN)

$\hat{\sigma}^2 = \Sigma(x_i - \mu - (\bar{x} - \mu))^2 / n$

$= \Sigma(x_i - \mu)^2 / n - (\bar{x} - \mu)^2$

$\qquad\quad \downarrow \qquad\qquad\qquad\quad \downarrow$

$E((x_i - \mu)^2) = \sigma^2 \qquad (\mu - \mu)^2$

$\sqrt{n}\,(\bar{X}-\mu) \xrightarrow{d} N(0, \sigma^2)$ by CLT.

$\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2)$

$= \sqrt{n}\,\left(\Sigma (X_i-\mu)^2/n - (\bar{X}-\mu)^2 - \sigma^2 \right)$

$= \underbrace{\sqrt{n}\left(\frac{\Sigma\, \cancel{E}[(X_i-\mu)^2]}{n} - \sigma^2 \right)}_{\text{viewing } (X_i-\mu)^2 \text{ as } X_i \text{ in CLT}} - \sqrt{n}\,(\bar{X}-\mu)^2$

$= N(0, V[(X_i-\mu)^2]) - \frac{1}{\sqrt{n}}\,[\sqrt{n}\,(\bar{X}-\mu)]^2$

$\because [\sqrt{n}\,(\bar{X}-\mu)]^2 \xrightarrow{d} [N(0, \sigma^2)]^2$

$\frac{1}{\sqrt{n}} \xrightarrow{p/d} 0$

$\Rightarrow \sqrt{n}\,(\bar{X}-\mu)^2 \to 0$

By slusky theorem:

$$\boxed{\sqrt{n}\,(\hat{\sigma}^2 - \sigma^2) \to N(0, 2\sigma^4)}$$

$Y_i = X_i - \mu \sim N(0, \sigma^2)$; $Z_i \sim N(0, 1)$

$V(Y_i^2) = V(\sigma^2, Z_i^2)$

$= \sigma^4 \left(E(Z_i^4) - (E(Z_i^2))^2 \right) = \sigma^4 (3 - 1)$

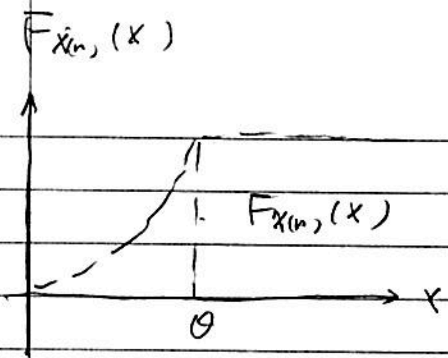
In formally we say

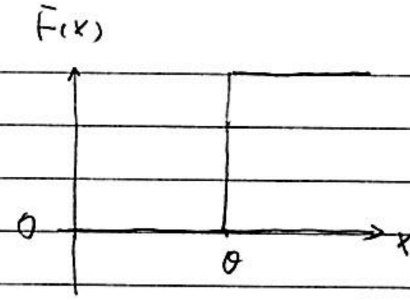
$\hat{\sigma}^2 \overset{\text{approximation}}{\sim} N(\sigma^2, \frac{2\sigma^4}{n})$

$\Rightarrow$ example: $X_1 \cdots X_n | \theta \sim unif(0, \theta)$

$\hat{\theta} = X_{(n)} = \max(X_1, \cdots, X_n)$

$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(\text{all } X_i \le x) = \left(\frac{x}{n}\right)^n$





~~For $0 < x \le 1$~~ ~~$F(x)$ is the CDF of~~

$F_{X_{(n)}}(x) \to 0$

~~For $x \geq 1$~~

$F_{X_{(n)}}(x) \to 1$

X	θ
prob.	1

$\hat{\theta} \xrightarrow{d} \theta$

$$Y_n = \sqrt{n}(\hat{\theta} - \theta)$$

$$F_{Y_n}(y) = P(\sqrt{n}(\hat{\theta} - \theta) \le y) = P\left(\hat{\theta} \le \frac{y}{\sqrt{n}} + \theta\right)$$

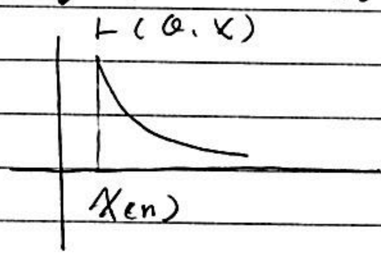
$$= \left(\frac{y}{\sqrt{n}} + \theta\right)^n \quad \text{for } y/\sqrt{n} + \theta < \theta$$

$\to \infty$

$\Rightarrow$ why $\sqrt{n}(\hat{\theta} - \theta)$ doesn't converge to $N(\ ;\ )$

1) the suppose of X_i depend on θ

2) $\hat{\theta}$ is not obtained at $\ell'(\theta) = 0$



$\Rightarrow$ Asymptotic Normality of MLE conditions:

1) the support of $f(x|\theta)$ doesn't depend on θ

2) $\ell'(\hat{\theta}_n) = 0$

~~3) ℓ is twice cont. differentiable at each θ~~

4) $|\ell'''(\theta; x)| \le g(x)$ for each θ, x.

$\exists$ a $g(x)$ with $E(g(x)) < +\infty$

then we have

$$\sqrt{n}\,(\hat{\theta}_{mle} - \theta) \xrightarrow{d} (0 \quad \frac{1}{I_1(\theta)})$$

where, $I_1(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2} \ell(\theta, x_i)\right) = V\left(\frac{\partial}{\partial\theta}\ell(\theta, x_i)\right)$

That is $\hat{\theta}_{mle} \overset{approx}{\sim} N(\theta, \frac{1}{nI_1(\theta)})$

$\Rightarrow$ proof:

suppose θ_0 is the true parameter, $\hat{\theta}_n$ is the MLE

$x = (x_1, \cdots x_n)$ (Taylor extention)

$$\ell'(\theta, x) = \ell'(\hat{\theta}_n; x) + (\theta - \hat{\theta}_n)\,\ell''(\hat{\theta}_n^*)$$

where $\hat{\theta}_n^*(x)$ lies between θ_0 and $\hat{\theta}_n$

$$\ell'(\theta_0, x) = \ell'(\hat{\theta}_n; x) + (\theta_0 - \hat{\theta}_n)\,\ell''(\hat{\theta}_n^*).$$

$\ell'(\hat{\theta}_n) = 0$

$$\ell'(\theta_0, x) = (\theta_0 - \hat{\theta}_n)\,\ell''(\hat{\theta}_n^*) \Rightarrow \theta_0 - \hat{\theta}_n = \frac{\ell'(\theta_0, x)}{\ell''(\hat{\theta}_n^*)}$$

$$\sqrt{nI_1(\theta_0)}\,(\hat{\theta}_n - \theta_0) = -\frac{\ell'(\theta_0; x)}{\ell''(\hat{\theta}_n^*)} \cdot \sqrt{nI_1(\theta_0)}$$

$$= \left(\frac{\ell'(\theta_0; x)}{\sqrt{nI_1(\theta_0)}}\right) \cdot \left(\frac{\ell''(\theta_0)}{\ell''(\hat{\theta}_n^*)}\right)\left(-\frac{nI_1(\theta_0)}{\ell''(\theta_0)}\right)$$

We will show

$$C_1: \quad \frac{\ell'(\theta_0, x)}{\sqrt{nI_1(\theta)}} \to N(0, 1)$$

$$\sqrt{\frac{n}{\sigma^2}}\left(\bar{x}-\mu\right) \to N(0,1)$$

C_2 : $\dfrac{-\ell''(\theta_0)}{-\ell''(\hat{\theta}_n^*)} \to 1$

C_3 : $\dfrac{n I_1(\theta_0)}{\ell''(\theta_0)} \to 1$

$\Rightarrow$ proof $C_1 \to N(0,1)$

$$\sqrt{n I_1^{-1}(\theta_0)}\left(\frac{\Sigma \ell'(\theta_0; x_i)}{n}\right) \xrightarrow{CLT} N(0,1)$$

because $E(\ell'(\theta_0; x_i)) = 0$; $Var(\ell'(\theta_0; x_i)) = I_1(\theta_0)$

$\Rightarrow$ proof C_3:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(\ell''(\theta_0; x_i))}{I_1(\theta)} \to \frac{E(-\ell''(\theta_0; x_i))}{I_1(\theta_0)}$$

$$\to 1$$

<< Mar 27. 2015 >> STAT 846

=> Asymptotic Normality under some regularity condition

$$\sqrt{n}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta))$$

$$\sqrt{n}\,(I(\theta_0))^{\frac{1}{2}}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} Nor\left[\underset{p\times p}{0}, \underset{p\times p}{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}}\right]$$

=> A generalization

$$\sqrt{n}\,(\hat{I}(\hat{\theta}_{mle}))^{\frac{1}{2}}\,(\hat{\theta}_{mle} - \theta_0) \rightarrow N(\underset{p\times p}{0}, I_p)$$

where, $\hat{I}(\hat{\theta}_{mle}) = -\sum_{i=1}^{n} \underbrace{\nabla_\theta^2\, \ell(\hat{\theta}_{mle}; x_i)}_{p\times p} / n \xrightarrow{p/d} I(\theta_0)$ by LLN

note $I(\theta) = E(-\nabla_\theta^2\, \ell(\theta; x))$

=> proof:

assume $p = 1$.

$$\sqrt{n\cdot \hat{I}(\hat{\theta}_{mle})}\,(\hat{\theta}_{mle} - \theta_0) = \sqrt{n I(\theta_0)}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N$$

$$\times \sqrt{\frac{\hat{I}(\hat{\theta}_{mle})}{I(\theta_0)}} \xrightarrow{p/d} 1$$

$\xrightarrow{d} N(0, 1)$ By Slutsky's theorem

=> example:

$X_1, \cdots x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

We have shown directly

$$\sqrt{n}\,(\hat{\sigma}^2_{mle} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$$

Now we apply our theorem

$$L(\theta; x_i) = (\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ell(\theta; x_i) = -\frac{1}{2}\log\sigma^2 - \frac{(x_i-\mu)^2}{2\sigma^2}$$

Score function : $\frac{\partial}{\partial \sigma^2} \ell(\theta; x_i) = -\frac{1}{2} \cdot \frac{1}{\sigma^2} + \frac{(x_i-\mu)^2}{2(\sigma^2)^2}$

$$\frac{\partial}{\partial \mu} \ell(\theta; x_i) = \frac{2(x-\mu)}{2\sigma^2}$$

Hessian function : $\frac{\partial^2}{\partial\mu\partial\sigma^2}\ell = -\frac{(x_i-\mu)}{(\sigma^2)^2}$

$$\frac{\partial}{\partial(\sigma^2)^2}\ell = \frac{1}{2}\frac{1}{(\sigma^2)^2} - \frac{(x_i-\mu)^2}{(\sigma^2)^3}$$

$$\frac{\partial^2\ell}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$I(\theta) = -\begin{bmatrix} E\frac{\partial^2\ell}{\partial\mu^2} & E\frac{\partial^2\ell}{\partial\mu\partial\sigma^2} \\ E\frac{\partial^2\ell}{\partial\mu\partial\sigma^2} & E\frac{\partial^2\ell}{\partial(\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{bmatrix}$$

By Asymptotic Normality, we conclude :

$$\left[\begin{pmatrix} \hat{\mu}_{mle} \\ \hat{\sigma}^2_{mle} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}\right]\sqrt{n} \xrightarrow{d} N\left(\underset{2\times 2}{0}, \underset{2\times 2}{\begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}}\right)$$

=> $\boxed{\text{Three Asymptotic properties of MLE}}$

1) Consistency : $\hat{\theta}_{mle} \xrightarrow{p/d} \theta_0$

2) Normality : $\sqrt{n}\,(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$

3) efficiency : $n\,Var(\hat{\theta}_{mle}) \to I^{-1}(\theta_0)$

=> likelihood Ratio test:

Suppose that: $\theta = (\theta_1, \cdots \theta_p)$

Null hypothesis H_0: $\theta_1 = \theta_1^0, \cdots, \theta_m = \theta_m^0$

alternative H_1: no restriction

=> example1 (ANOVA)

$H_0: \mu_1 = \mu_2 = \cdots = \mu_p$ OR $H_0: \theta_i = \mu_i - \mu_1 = 0$

$i = 2, \cdots p.$

=> example 2 : $X_1, \cdots, x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

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

H_0: $\mu = \mu_0$ σ^2 can be anything

H_1: $\mu \neq \mu_0$ σ^2 ——————.

point null hypothesis

=> Defination:

Likelihood Ratio Test

$$\Lambda = \max_{\theta \in H} f(x;\theta) \Big/ \max_{\theta \in H_0} f(x;\theta)$$

$$= f(x;\hat{\theta}_{mle1}) \Big/ f(x;\hat{\theta}_{mle0})$$

⇒ Wilk's theorem:

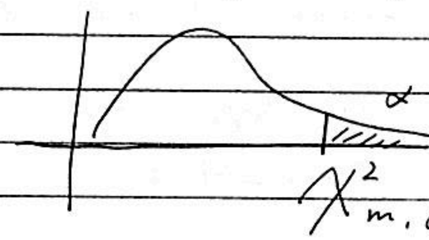
$$-2\log(\Lambda)\,\Big|\,\theta\in H_0 \xrightarrow{d} \chi^2_m$$

m = reduction of dimensions of θ from H_1 to H_0

$= \mathrm{Dim}(H_1) - \mathrm{Dim}(H_0)$

$\mathrm{Dim}(H_0) = p-m.$ $\quad \mathrm{Dim}(H_1) = p$

Rejection Region of size α, $\quad -2\log(\Lambda) > \chi^2_{m,\alpha}.$



⇒ example:

$x_1, \cdots, x_n \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$

$H_0: \mu = \mu_0 \quad$ vs $\quad H_1: \mu \neq \mu_0$

Given $\theta \in H_0$, $\quad f(x;\theta) = \dfrac{1}{(\sqrt{2\pi})^n} \cdot \dfrac{1}{(\sigma^2)^{\frac{n}{2}}} \exp -\dfrac{\Sigma(x_i-\mu)^2}{2\sigma^2}$

$\mu = \mu_0$

$\hat{\sigma}_0^2 = \dfrac{\Sigma(x_i-\mu_0)^2}{n}.$

$$\Rightarrow f(x|\hat{\theta}_0) = \frac{1}{(\sqrt{2\pi})^n}\frac{1}{(\hat{\sigma}_0^2)^{\frac{n}{2}}}\, e^{-\frac{\Sigma(x_i-\mu_0)^2}{2\hat{\sigma}_0^2}} = e^{-\frac{n}{2}}$$

Given $\theta \in H_1 \quad \hat{\mu}_1 = \bar{x}$

$\hat{\sigma}_1^2 = \Sigma(x_i - \bar{x})^2 / n$

$$\Rightarrow f(x|\hat{\theta}_1) = \frac{1}{(\sqrt{2\pi})^n}\frac{1}{(\hat{\sigma}_1^2)^{\frac{n}{2}}}\, e^{-\frac{n}{2}}$$

$$\text{So} \quad \Lambda = f(x \mid \hat{\theta}_1) \big/ f(x \mid \hat{\theta}_0) = \frac{(\hat{\sigma}_0^2)^{\frac{n}{2}}}{(\hat{\sigma}_1^2)^{\frac{n}{2}}}$$

$$= \left[\frac{\Sigma (x_i - \mu_0)^2}{\Sigma (x_i - \bar{x})^2} \right]^{\frac{n}{2}} = \left[\frac{\Sigma (x_i - \bar{x} + \bar{x} - \mu_0)^2}{\Sigma (x_i - \bar{x})^2} \right]^{n/2}$$

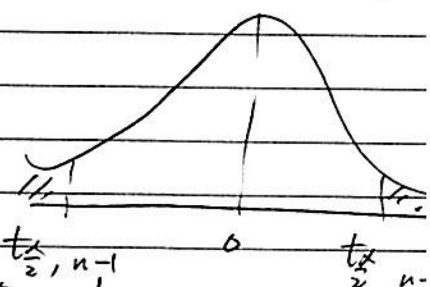
$$= \left[1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2} \right]^{n/2}$$

look at Λ directly

$$\Lambda > \lambda \iff \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2 / n} > K/n$$

$$\boxed{\frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2 / n} \text{ is T-test} \quad \left(\because \ t = \frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \right)}$$

$$\iff \left| t = \frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \right| > t^*$$



To determine t^* we use.

$$\frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \sim t_{n-1} \implies t^* = t_{\frac{\alpha}{2}, n-1}$$

<< Mar 30, 2015 >> STAT 846

Schdule:

① No class 3rd Apr.

② Review on 6th Apr.

③ Test 2. on 8th Apr.

④ Final on 14th Apr.

⑤ Tuesday 9-12. AGRI 2D

=> Wilk Theorem:

$x_1, \cdots x_n \mid \theta \overset{iid}{\sim} f(x;\theta)$; $\theta = (\theta_1, \cdots, \theta_p)$

H_0 : $\theta_1 = \theta_1^0, \cdots, \theta_m = \theta_m^0$

H_1 : no restriction for θ

$$\Lambda = \max_{\theta \in H_1} f(x;\theta) \Big/ \max_{\theta \in H_0} f(x;\theta)$$

$$2\log\Lambda \mid \theta \in H_0 \xrightarrow{d} \chi^2_m$$

m = reduction of dimension of θ from H_1 to H_0

=> example. (ANOVA). $x_{ij} \mid \mu_i\, \sigma^2 \overset{iid}{\sim} N(\mu_i, \sigma^2)$, for $i = 1, 2, \cdots K$

H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$

H_1: No restriction.

$m = \dim(H_1) - \dim(H_0)$

$= (K+1) - (1+1) = K-1$

Let $\theta_1 = \mu_1$

$\theta_2 = \mu_2 - \mu_1$

$\vdots$

$\theta_k = \mu_k - \mu_1$

$H_0: \theta_2 = 0, \quad \theta_3 = 0, \quad \cdots \quad \theta_k = 0$

H_1: no - restriction

$m = k - 1$

=> example:

$$x_1, \cdots x_n \mid \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

$H_0: \quad \mu = \mu_0$

$H_1: \quad \mu \neq \mu_0$

$$\Lambda = \left[\frac{\sum_{i=1}^{n} (x_i - \mu_0)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]^{\frac{n}{2}}$$

$$\Lambda = \left[1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2} \right]^{\frac{n}{2}} \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

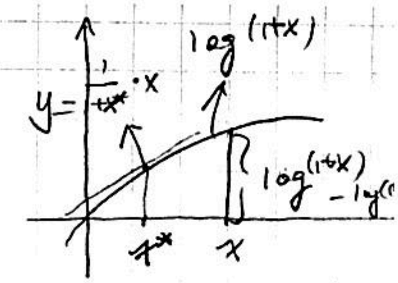
Rejection Region give size α.

$$\{ \Lambda > \lambda_\alpha \} \Leftrightarrow \left| \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}} \right| > t_{\frac{\alpha}{2}, n-1}$$

$t_{\frac{\alpha}{2}, n-1}$ is determined by α and

the fact: $\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}} \Big| \mu = \mu_0 \sim t_{n-1}$

An illustration of Wilk's Theorem



- $2\log(\Lambda) = n\log\left[1 + \frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}\right]$

$\log(1+x) = \log(1+0) + (x^* - 0)\left.\frac{\partial \log(1+x)}{\partial x}\right|_{x=x^*} = 0 + x \cdot \frac{1}{1+x^*}$

$= x \cdot \frac{1}{1+x^*} \qquad (0 < x^* < x)$

$$2\log\Lambda = n \cdot \log\left[1 + \left(\frac{\bar{x}-\mu_0}{\hat{\sigma}}\right)^2\right]$$

$$= n \cdot \frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2} \cdot \frac{1}{1+T^*} \quad \text{for some}$$

$$0 < T^* < \frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2} \quad \text{(which is } x \text{ is mean value theorem)}$$

- $2\log\Lambda = \frac{[\sqrt{n}(\bar{x}-\mu_0)]^2}{\sigma^2} \cdot \frac{\sigma^2}{\hat{\sigma}^2} \cdot \frac{1}{1+T^*}$

$$\underbrace{\frac{[\sqrt{n}(\bar{x}-\mu_0)]^2}{\sigma^2}}_{\downarrow \text{CLT} \;\; [N(0,1)]^2} \cdot \underbrace{\frac{\sigma^2}{\hat{\sigma}^2}}_{\downarrow \;\; 1} \cdot \underbrace{\frac{1}{1+T^*}}_{\downarrow \;\; 1}$$

$$\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2} = \underbrace{\frac{(\bar{x}-\mu_0)^2}{\sigma^2}}_{\downarrow \text{LNN} \;\; 0} \cdot \underbrace{\frac{\sigma^2}{\hat{\sigma}^2}}_{\downarrow \;\leftarrow \text{(converge)} \;\; 1}$$

$$\text{B/C} \quad 0 < T^* < \underbrace{\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}}_{\downarrow \;\; 0}, \Rightarrow T^* \to 0$$

- By Slutsky's Theorem $2\log\Lambda \xrightarrow{d} Z^2$ where $Z \sim N(0,1)$ that is, $2\log\Lambda \xrightarrow{d} \chi^2_1$

$\Rightarrow$ proof : Wilk's Theorem.

Let $d = m = 1$ for simplicity

θ is a scalar.

$H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$

$\Lambda_n = 2\log (f(x; \hat{\theta}_n) - f(x; \theta_0))$

$\hat{\theta}_n$ is M.L.E of θ under H_1

$\Lambda_n = 2 [\ell(\hat{\theta}_n) - \ell(\theta_0)]$ where x is omitted.

$= 2 (\hat{\theta}_n - \theta_0) \ell'(\theta_n) + (\hat{\theta}_n - \theta_0)^2 \cdot f''(\theta_n^*)$ $\quad \theta_0 < \theta_n^* < \theta_n$

$= n I_1(\theta_0) (\hat{\theta}_n - \theta_0)^2 \dfrac{\ell''(\theta_0)}{-n I_1(\theta_0)} \cdot \ell$