More on Fisher information : Asymptotically normal of MLE

① $I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]$, & fact : $E\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right] = 0$

② $\hat{\theta}_{MLE} \xrightarrow{D} N(\theta_0, \frac{1}{nI(\theta_0)})$

Brief recall :

θ_0 — the true value of θ

$\hat{\theta}_{MLE}$ — the MLE of θ

Assume $X_1, \cdots, X_n$ i.i.d random samples

Assume $\ell(\theta) = \log L(\theta) = \log \prod_{i=1}^{n} f(X_i|\theta)$

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

score function $\frac{\partial}{\partial\theta}\ell(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(X_i|\theta) = \sum_{i=1}^{n}\underbrace{\frac{\partial}{\partial\theta}\log f(X_i|\theta)}_{\text{Score function}} = 0$

Score function : $\frac{\partial}{\partial\theta}\log f(x|\theta)$

$E\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) = 0$, $Var\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) = E\left[(\quad)^2\right] - \left(E[\quad]\right)^2 = I(\theta) - 0 = I(\theta)$

Estimating function : $\frac{\partial}{\partial\theta}\log f(X_i|\theta) = \frac{\partial\ell(\theta)}{\partial\theta}$

$E\left[\frac{\partial\ell(\theta)}{\partial\theta}\right] = 0$, $Var\left(\frac{\partial}{\partial\theta}\ell(\theta)\right) = nI(\theta)$

Thm Under conditions on f,
(Rice: 360.5.5, Hogg)
or (Hogg, Theorem 6.3.1) $\sqrt{nI(\theta_0)}\,(\hat{\theta} - \theta_0) \xrightarrow{D} N(0,1)$ as $n \to \infty$

or the large sample distribution of $\hat{\theta}_{MLE}$ is approximately normal with mean θ_0 and variance $1/(nI(\theta_0))$

proof Expanding the function $\ell(\theta)$ into a Taylor series of order 2 about θ_0 and evaluating it at $\hat{\theta}_{MLE}$, we get

$$\ell(\hat{\theta}) = \ell(\theta_0) + (\hat{\theta} - \theta_0)\frac{\ell'(\theta_0)}{1!} + (\hat{\theta} - \theta_0)^2\frac{\ell''(\theta^*)}{2!}$$

where θ^* is between $\hat{\theta}$ and θ_0

Notice that $\ell'(\hat{\theta}) = 0$. then

$$(\hat{\theta} - \theta_0)\left[\ell'(\theta_0) + (\hat{\theta} - \theta_0)\frac{\ell''(\theta^*)}{2}\right] = -\ell'(\theta_0)$$

$$\sqrt{n}\,(\hat{\theta} - \theta_0) = -\frac{\ell'(\theta_0)/\sqrt{n}}{\underbrace{\ell''(\theta_0)/n}_{-I(\theta_0)} + \underbrace{(\hat{\theta} - \theta_0)\frac{\ell''(\theta^*)}{2n}}_{\xrightarrow{P} 0}}$$

$\xrightarrow{P} 0$, $\ell''(\theta^*)$ bound
$\hat{\theta} \xrightarrow{P} \theta_0$, consistency

$-I(\theta_0)$ by W.L.L.N

Slutsky's Lemma:
If $X_n \to c$ in probability and $Y_n \to Y$ in distribution,
then $X_n Y_n \to cY$ in distribution

$$\approx \frac{-\ell'(\theta_0)/\sqrt{n}}{\ell''(\theta_0)/n}$$

$$\sqrt{n}\,(\hat{\theta} - \theta_0) \approx \frac{\ell'(\theta_0)/\sqrt{n}}{I(\theta_0)} \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)$$

since

$$Var\left(\frac{\ell'(\theta_0)/\sqrt{n}}{I(\theta_0)}\right) = \frac{1}{I(\theta_0)^2} \cdot I(\theta_0) = \frac{1}{I(\theta_0)}$$

$$E\left(\frac{\ell'(\theta_0)/\sqrt{n}}{I(\theta_0)}\right) = 0$$

W.L.L.N
i.i.d. random sample $Y_1, \cdots, Y_n$, $E(Y_i) = \mu$
$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} \mu$ as $n \to \infty$

consider

$$\ell'(\theta_0)/n = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{\partial}{\partial\theta}\log f(X_i|\theta_0)}_{Y_i} \xrightarrow{P} \frac{1}{n}\left(-I(\theta_0)\right) = -I(\theta_0)$$

C.L.T
i.i.d. random sample $Y_1, \cdots, Y_n$, $E(Y_i) = \mu$, $Var(Y_i) = \sigma^2$
$(\bar{Y} - \mu) \xrightarrow{D} N(0, \frac{\sigma^2}{n})$

consider

$$\ell'(\theta_0)/\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{\frac{\partial}{\partial\theta}\log f(X_i|\theta_0)}_{Y_i} \xrightarrow{D} N(0, I(\theta_0))$$

$E\left(\frac{\ell'(\theta_0)}{\sqrt{n}}\right) = 0$, $Var\left(\ell'(\theta_0)/\sqrt{n}\right) = \frac{1}{n}Var(\ell'(\theta_0)) = \frac{1}{n}\,nI(\theta_0) = I(\theta_0)$

The large sample distribution of $\hat{\theta}_{MLE}$ is approximately normal with mean θ_0 and asymptotic variance of $\hat{\theta}_{MLE}$ $\frac{1}{nI(\theta_0)}$

asymptotically unbiased $\hat{\theta}_{MLE} \sim N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$

$$\sqrt{n}\,(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)$$

$$\sqrt{nI(\theta_0)}\,(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{D} N(0, 1)$$

△ $100(1-\alpha)\%$ confidence interval of θ based on MLE.

↳ Exact Method (very exceptional, need some luck and clarness)

E.g. Normal distribution , $X \sim N(\mu, \sigma^2)$

An i.i.d. random sample $\{X_1, X_2, \cdots, X_n\}$, then

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

① To construct C.I. for μ :

The key is "what is the sampling distribution of $\hat{\mu}_{MLE}$"?

$\hat{\mu}_{MLE} = \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ unknown

exactly follows parameter

The fact from Chap 1 (Thm 1.7.35) Student T :

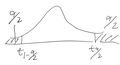
① $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

② $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

③ $\frac{Z}{\sqrt{V/r}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}} \sim t(n-1)$

i.e. $\frac{(\bar{X} - \mu)\sqrt{n}}{S} \sim t(n-1)$ d.f. $= n-1$

So $P\left(t_{1-\frac{\alpha}{2}} < \frac{(\bar{X} - \mu)\sqrt{n}}{S} < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$



$$\frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}} < \bar{X} - \mu < \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}$$

$$\bar{X} - \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}} < \mu < \bar{X} - \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}$$

$$\bar{X} - \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}$$

t is symmetric distribution

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

$$\bar{X} - \frac{\hat{\sigma}_{MLE}}{\sqrt{n-1}}t_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{\hat{\sigma}_{MLE}}{\sqrt{n-1}}t_{\frac{\alpha}{2}}$$

⇐ symmetric about $\bar{X}$

② To construct C.I. for σ^2 based on S^2 (or $\hat{\sigma}^2$)

The key is $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. d.f. $= n-1$

parameter

$$P\left(\chi^2_{1-\frac{\alpha}{2}}(n-1) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

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

i.e. $\frac{n\hat{\sigma}^2_{MLE}}{\chi^2_{\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{n\hat{\sigma}^2_{MLE}}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}$ ⇐ notice that it is generally not symmetric about $\hat{\sigma}^2_{MLE}$.

△ Approximate Method (based on "asymptotically normal" distribution of $\hat{\theta}_{MLE}$)

$$\hat{\theta}_{MLE} \overset{approx}{\underset{as\ n\to\infty}{\sim}} N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$$

Since θ_0 is unknown, we will use $I(\hat{\theta}_{MLE})$ instead of $I(\theta_0)$

$$-z_{\frac{\alpha}{2}} < \frac{\hat{\theta}_{MLE} - \theta_0}{\sqrt{1/nI(\hat{\theta}_{MLE})}} < z_{\frac{\alpha}{2}}$$

$$\hat{\theta}_{MLE} - \sqrt{\frac{1}{nI(\hat{\theta}_{MLE})}}\,z_{\frac{\alpha}{2}} < \theta_0 < \hat{\theta}_{MLE} + \sqrt{\frac{1}{nI(\hat{\theta}_{MLE})}}\,z_{\frac{\alpha}{2}}$$

Which is an approximate $100(1-\alpha)\%$ C.I. for θ_0.

E.g. (Bernoulli(p)) $X_i \sim$ Bernoulli(p)

Recall: $\hat{p}_{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Approx. $100(1-\alpha)\%$ C.I. for p :

$$\hat{p}_{MLE} \pm \sqrt{\frac{1}{nI(\hat{p}_{MLE})}} \cdot z_{\frac{\alpha}{2}}$$

$$I(p) = E\left[\left(\frac{\partial}{\partial p}\log f(X|p)\right)^2\right] = -E\left[\frac{\partial^2}{\partial p^2}\log f(X|p)\right]$$

$$= -E\left[-\frac{X}{p^2} - \frac{(1-X)}{(1-p)^2}\right]$$

$$= \frac{1}{p(1-p)}$$

$$\log f(X|p) = \log\left[p^X(1-p)^{1-X}\right]$$

$$= X\log p + (1-X)\log(1-p)$$

$$\frac{\partial}{\partial p}\log f(X|p) = \frac{X}{p} + \frac{X-1}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\log f(X|p) = -\frac{X}{p^2} - \frac{1-X}{(1-p)^2}$$

$$I\left(\frac{\partial^2}{\partial p^2}\cdots\right) = -\left[-\frac{1}{p^2}\cdot p + \frac{1}{(1-p)^2}(1-p)\right] = -\left[-\frac{1}{p} + \frac{1}{1-p}\right]$$

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

$$\left(\hat{p}_{MLE} \pm \sqrt{\frac{1}{nI(\hat{p}_{MLE})}} \cdot z_{\frac{\alpha}{2}}\right)$$

i.e. $\left(\hat{p} \pm \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,z_{\frac{\alpha}{2}}\right)$

$\left(\bar{X} \pm \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\,z_{\frac{\alpha}{2}}\right)$