# STA532 - hw6 sample solution

Sheng Jiang (sj156@duke.edu)

1. $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu + \mu - \bar{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}[(Y_i-\mu)^2 + (\mu-\bar{Y})^2 + 2(Y_i-\mu)(\mu-\bar{Y})]$

$= S_0^2 - (\mu-\bar{Y})^2$

$S_n^2 = \frac{n}{n-1} S_1^2$

As $\{Y_i\}$ are iid with finite fourth moment, by wLLN,

$S_0^2 \xrightarrow{P} E[(Y_1-\mu)^2] = \sigma^2$, and $\bar{Y} \xrightarrow{P} E[Y_1] = \mu$.

By Slutsky thm and cts mapping thm, $S_n^2 \xrightarrow{P} \sigma^2$, which is consistency.

Under the same assumptions, by CLT, $\sqrt{n}(S_0^2 - \sigma^2) \xrightarrow{d} N(0, V_0)$, where variance

$V_0 := V((Y_1-\mu)^2) = E[(Y_1-\mu)^4] - [E((Y_1-\mu)^2)]^2 < \infty$, a known constant depending on $P$.

$\sqrt{n}(S_n^2 - \sigma^2) = \sqrt{n}(S_n^2 - \frac{n}{n-1}\sigma^2 + \frac{1}{n-1}\sigma^2) = \sqrt{n}(\frac{n}{n-1}(S_1^2 - \sigma^2) + \frac{1}{n-1}\sigma^2)$

$= \sqrt{n}(\frac{n}{n-1}(S_0^2 - \sigma^2) - \frac{n}{n-1}(\mu-\bar{Y})^2 + \frac{1}{n-1}\sigma^2)$

By Slutsky thm, $\sqrt{n}\frac{n}{n-1}(S_0^2-\sigma^2) \xrightarrow{d} N(0, V_0)$

For the second term, $\sqrt{n}\frac{n}{n-1}(\mu-\bar{Y})^2 = \frac{1}{\sqrt{n}}\frac{n}{n-1}(\sqrt{n}(\bar{Y}-\mu))^2$.

Then by cts mapping and Slutsky thm, $\frac{1}{\sqrt{n}}\frac{n}{n-1}(\sqrt{n}(\bar{Y}-\mu))^2 \xrightarrow{P} 0$.

For the third term, $\sqrt{n}\cdot\frac{1}{n-1}\sigma^2 \to 0$ is real number sequence convergence.

Combining the above, by Slutsky thm, $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, V_0)$, which is Asym. Normality.

2. (a) $E(\hat{\theta}) = E[\bar{W}] = E(\frac{1}{n}\sum_{i=1}^{n} Y_i/x_i) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i)/x_i = \theta$

$V(\hat{\theta}) = V[\bar{W}] = V(\frac{1}{n}\sum_{i=1}^{n} Y_i/x_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i)/x_i^2 + \frac{1}{n^2}\sum_{i\neq j} \mathrm{cov}(\frac{Y_i}{x_i}, \frac{Y_j}{x_j})$

$= \frac{1}{n^2}\sum_{i=1}^{n} x_i^{-2}\sigma^2$

(b) $L(\theta, \sigma^2) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(Y_i - \theta x_i)^2/\sigma^2} \propto \frac{1}{\sigma^n} e^{-\frac{1}{2}\sigma^{-2}[(\Sigma X_i^2)(\theta - (\Sigma X_i^2)^{-1}\Sigma X_i Y_i)^2 + \Sigma Y_i^2 - (\Sigma X_i^2)^{-1}(\Sigma X_i Y_i)^2]}$

$\max_{\theta,\sigma^2} \log L(\theta, \sigma^2) = \max_{\theta,\sigma^2}\left\{-\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sigma^{-2}[(\Sigma X_i^2)(\theta - (\Sigma X_i^2)^{-1}\Sigma X_i Y_i)^2 + \Sigma Y_i^2 - (\Sigma X_i^2)^{-1}(\Sigma X_i Y_i)^2]\right\}$

regardless $\hat{\sigma}^2_{MLE}$, MLE $\hat{\theta}_{MLE} = (\Sigma X_i^2)^{-1}\Sigma X_i Y_i$

$E[\hat{\theta}_{MLE}] = (\Sigma X_i^2)^{-1} \sum_{i=1}^n X_i E(Y_i) = \theta$

$V(\hat{\theta}_{MLE}) = (\Sigma X_i^2)^{-2} V(\Sigma X_i Y_i) = (\Sigma X_i^2)^{-2} \left[ \sum_{i=1}^n X_i^2 \sigma^2 + \sum_{i \neq j} cov(X_i Y_i, X_j Y_j) \right]$

$= (\Sigma X_i^2)^{-1} \sigma^2$

(c) For $\hat{\theta}$, its MSE is $\frac{1}{n^2} \sum_{i=1}^n X_i^{-2} \sigma^2 = \frac{1}{n} \frac{1}{n} \sum_{i=1}^n X_i^{-2} \sigma^2$

For $\hat{\theta}_{MLE}$, its MSE is $(\Sigma X_i^2)^{-1} \sigma^2 = \frac{1}{n} (\frac{1}{n} \sum_{i=1}^n X_i^2)^{-1} \sigma^2$

By Jensen's inequality, $\frac{1}{n} \Sigma (X_i^2)^{-1} \geqslant (\frac{1}{n} \Sigma X_i^2)^{-1}$, where "=" holds when $x_i$'s are identical.

MLE is weakly better in all cases and strictly better unless $x_i$'s are identical.

(d) $V(\hat{\theta}_{MLE}) \cdot n \xrightarrow{P} (E X_i^2)^{-1} \sigma^2 = (\mu_x^2 + \sigma_x^2)^{-1} \sigma^2$

To minimize the estimation error, better to choose large $|\mu_x|$ and $\sigma_x^2$.

3. (a) We follow the convention $0 \cdot \log 0 = 0$.

i. Binomial : $Y \sim Bin(n, p)$

$L(p) = \binom{n}{Y} p^Y (1-p)^{n-Y} = e^{Y \cdot \log(p/(1-p)) + n \log(1-p)}$

$c(Y) = \binom{n}{Y}$, $\theta = \log \frac{p}{1-p}$, $t(Y) = Y$, $A(\theta) = n \log(1-p) = n \log \frac{e^\theta}{1+e^\theta}$

ii. Poisson : $Y \sim Pois(\mu)$

$L(\mu) = e^{-\mu} \cdot \frac{\mu^Y}{Y!} = (Y!)^{-1} e^{Y \cdot \log(\mu) - \mu}$

$c(Y) = (Y!)^{-1}$, $\theta = \log \mu$, $t(Y) = Y$, $A(\theta) = \mu = e^\theta$

iii Normal : $Y \sim N(\mu, \sigma^2)$

$L(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(Y-\mu)^2/\sigma^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} Y^2 \cdot \sigma^{-2} + Y\mu\sigma^{-2} - \frac{1}{2}\mu^2\sigma^{-2} - \frac{1}{2}\log\sigma^2}$

$c(Y) = \frac{1}{\sqrt{2\pi}}$, $\theta = [\sigma^{-2} \ \ \mu\sigma^{-2}]^T$, $t(Y) = [-\frac{1}{2}Y^2 \ \ Y]^T$, $A(\theta) = \frac{1}{2}\mu^2\sigma^{-2} + \frac{1}{2}\log\sigma^2$

iv. Beta : $Y \sim Beta(a, b)$

$L(a, b) = \frac{1}{B(a,b)} Y^{a-1} (1-Y)^{b-1} = Y^{-1} (1-Y)^{-1} e^{a \log Y + b \log(1-Y) - \log B(a,b)}$

$c(Y) = (Y(1-Y))^{-1}$, $\theta = [a \ \ b]^T$, $t(Y) = [\log Y \ \ \log(1-Y)]^T$, $A(\theta) = \log B(a,b)$

(b) n-sample log-likelihood $\ell_n(\theta) = \log \prod_{i=1}^{n} L(\theta) = \theta^T\left(\sum_{i=1}^{n} t(Y_i)\right) - nA(\theta) + \sum_{i=1}^{n} \log c(Y_i)$

$= n\left(\theta^T \overline{t_n(Y)} - A(\theta)\right) + \sum_{i=1}^{n} \log c(Y_i)$, where $\overline{t_n(Y)} = \frac{1}{n}\sum_{i=1}^{n} t(Y_i)$

(c). Set $\frac{\partial}{\partial\theta} \ell_n(\theta) = 0$, then it's equivalent to $\frac{\partial}{\partial\theta} A(\theta) = \overline{t_n(Y)}$.

MLE is the solution to the equation "$\frac{\partial}{\partial\theta} A(\theta) = \overline{t_n(Y)}$".

(d) $\int\left[\frac{\partial}{\partial\theta} p(y|\theta)\right] dy = \frac{\partial}{\partial\theta}\left[\int p(y|\theta)\, dy\right] = 0.$

then, by chain rule, $\int p(y|\theta)\left(t(y) - \frac{\partial}{\partial\theta} A(\theta)\right) dy = 0$, that is, $\frac{\partial}{\partial\theta} A(\theta) = \mathbb{E}\, t(Y)$

$\overline{t_n(Y)}$ is the sample analogue of $\mathbb{E}t(Y)$.

By LLN, when sample size n is sufficiently large, $\mathbb{E}t(Y) \approx \overline{t_n(Y)}$.

If $\frac{\partial}{\partial\theta} A(\theta)$ is cts, solving "$\frac{\partial}{\partial\theta} A(\theta) = \overline{t_n(Y)}$" gives similar solution(s) to solving "$\frac{\partial}{\partial\theta} A(\theta) = \mathbb{E}\, t(Y)$", in probability.

4. (a). $L(\theta) = \prod_{i=1}^{n} \theta^{Y_i}(1-\theta)^{1-Y_i} = \theta^{\Sigma Y_i}(1-\theta)^{n-\Sigma Y_i}$

Set $\frac{d}{d\theta} \log L(\theta) = 0$, i.e., $\frac{d}{d\theta}\left(\Sigma Y_i \log\theta + (n-\Sigma Y_i)\log(1-\theta)\right) = 0$,

the solution is $\hat{\theta}_{MLE} = \frac{1}{n}\Sigma Y_i$.

By checking the second order condition, we know the solution is unique.

(b). By invariance of MLE, $\hat{\varphi}_{MLE}$ satisfies "$\frac{e^{\hat{\varphi}_{MLE}}}{1+e^{\hat{\varphi}_{MLE}}} = \hat{\theta}_{MLE}$".

then $\hat{\varphi}_{MLE} = \log\left(\frac{\frac{1}{n}\Sigma Y_i}{1-\frac{1}{n}\Sigma Y_i}\right) = \log\left(\frac{\Sigma Y_i}{n-\Sigma Y_i}\right)$

5. Suppose $\exists\, \theta_0 \in \Theta$, s.th. $f_{\theta_0}(y) \geq f_\theta(y)$ for all $\theta \in \Theta$, and $f_{\theta_0}(y) > f_\theta(y)$ for some $\theta \in \Theta$.

The MLE based on $\mathcal{P}_1$ is denoted as $\hat{\theta} = \arg\max_\theta f_\theta(y)$. Note the solution can be not unique.

With the mapping $h: \Theta \to \Psi$, $f_\theta(y) = f_{h(\psi)}(y) = g_\psi(y)$.

Therefore, $\hat{\psi} \equiv \arg\max_\psi g_\psi(y) = \arg\max_\psi f_{h(\psi)}(y)$ with $h(\hat{\psi}) = \hat{\theta}$.

We prove the claim by contradiction. Now suppose $h(\hat{\psi}) \neq \hat{\theta}$.

Pick $\psi_0 \in \Psi$ s.th. $h(\psi_0) \in \hat{\theta}$ and $h(\psi_0) \notin h(\hat{\psi})$

Since $\hat{\psi}$ solves $\max_\psi g_\psi(y) = \max_\psi f_{h(\psi)}(y)$, and $\psi_0$ is not among the solutions, then $g_{\hat{\varphi}}(y) = f_{h(\hat{\varphi})}(y) > g_{\varphi_0}(y) = f_{h(\varphi_0)}(y)$.

On the other hand, $h(\varphi_0) \in \hat{\theta}$ solves " $\max_{\theta} f_\theta(y)$ " and $h(\hat{\varphi})$ is not among the solutions,

then $f_{h(\varphi_0)}(y) > f_{h(\hat{\varphi})}(y)$, which yields a contradiction.

The other case where $\varphi_0 \in \Psi$ s.th. $h(\varphi_0) \notin \hat{\theta}$ and $h(\varphi_0) \in h(\hat{\varphi})$ follows similarly.

Therefore, the claim $h(\hat{\varphi}) = \hat{\theta}$ holds.

As $h$ is 1-1, $\hat{\varphi} = h^{-1}(\hat{\theta})$.

If there does not exist $\theta_0 \in \Theta$ ~~to~~ maximize $f_\theta(y)$,

then we replace the MLE as $\hat{\theta} = \arg\sup_{\theta \in \Theta} f_\theta(y)$, that is, the solution is on the boundary of $\Theta$, and $\exists \{\theta_n\} \subseteq \Theta$ s.th. $\lim_{n\to\infty} \theta_n = \hat{\theta}$.

Then define $\varphi_n \subseteq \Psi$ s.th. $h(\varphi_n) = \theta_n$.

$\lim_{n\to\infty} \varphi_n$ is the MLE for $g_\varphi(y)$

Uniqueness of $\lim_{n\to\infty} \varphi_n$ can be proved by contradiction.