STA532 - hw8 sample solution

Sheng Jiang (sj156@duke.edu)

1. (a) $p_i \equiv \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} = 1 - \frac{1}{1+e^{\alpha+\beta x_i}} \Rightarrow e^{\alpha+\beta x_i} = \frac{1}{1-p_i} - 1 = \frac{p_i}{1-p_i}$

By chain rule, $\frac{\partial p_i}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left(1 - \frac{1}{1+e^{\alpha+\beta x_i}}\right) = \frac{1}{(1+e^{\alpha+\beta x_i})^2} \cdot e^{\alpha+\beta x_i} = (1-p_i)p_i$

$\frac{\partial p_i}{\partial \beta} = \frac{\partial}{\partial \beta}\left(1 - \frac{1}{1+e^{\alpha+\beta x_i}}\right) = \frac{1}{(1+e^{\alpha+\beta x_i})^2} e^{\alpha+\beta x_i} \cdot x_i = (1-p_i)p_i x_i$

Likelihood of $\{(X_i, Y_i)\}_{i=1}^n$ given $\{X_i = x_i\}_{i=1}^n$ is $\prod_{i=1}^n p_i^{Y_i}(1-p_i)^{1-Y_i}$

log likelihood of $\{(X_i, Y_i)\}_{i=1}^n$ given $\{X_i = x_i\}_{i=1}^n$ is

$$l_n(\alpha,\beta) \equiv \sum_{i=1}^n \left[Y_i \log p_i + (1-Y_i)\log(1-p_i)\right] = \sum_{i=1}^n \left[Y_i \log\frac{p_i}{1-p_i} + \log(1-p_i)\right]$$

Score fcn $S_n(\alpha,\beta) = \begin{bmatrix} \frac{\partial}{\partial\alpha} l_n(\alpha,\beta) \\ \frac{\partial}{\partial\beta} l_n(\alpha,\beta) \end{bmatrix}$

where $\frac{\partial}{\partial\alpha} l_n(\alpha,\beta) = \frac{\partial}{\partial\alpha}\left\{\sum_{i=1}^n \left[Y_i \log\frac{p_i}{1-p_i} + \log(1-p_i)\right]\right\}$

$= \sum_{i=1}^n \left(Y_i + \frac{-1}{1-p_i}\frac{\partial p_i}{\partial\alpha}\right) = \sum_{i=1}^n (Y_i - p_i)$

$\frac{\partial}{\partial\beta} l_n(\alpha,\beta) = \frac{\partial}{\partial\beta}\left\{\sum_{i=1}^n \left[Y_i \log\frac{p_i}{1-p_i} + \log(1-p_i)\right]\right\}$

$= \sum_{i=1}^n \left(Y_i x_i + \frac{-1}{1-p_i}\frac{\partial p_i}{\partial\beta}\right) = \sum_{i=1}^n \left[x_i(Y_i - p_i)\right]$

To find the MLE of (α,β), set $S(\alpha,\beta) = 0$.

That is, $(\hat\alpha_{MLE}, \hat\beta_{MLE})$ solves $\begin{cases} \sum_{i=1}^n p_i = \sum_{i=1}^n Y_i \\ \sum_{i=1}^n p_i x_i = \sum_{i=1}^n x_i Y_i \end{cases}$

(b) $\frac{\partial^2}{\partial\alpha^2} l_n(\alpha,\beta) = \frac{\partial}{\partial\alpha}\left(\frac{\partial}{\partial\alpha} l_n(\alpha,\beta)\right) = \frac{\partial}{\partial\alpha}\left[\sum_{i=1}^n (Y_i - p_i)\right] = -\sum_{i=1}^n \frac{\partial}{\partial\alpha} p_i = -\sum_{i=1}^n (1-p_i)p_i$

$E\frac{\partial^2}{\partial\alpha^2} l_n(\alpha,\beta) = -n\,E[(1-p_1)p_1]$, by iid assumption.

$\frac{\partial^2}{\partial\alpha\partial\beta} l_n(\alpha,\beta) = \frac{\partial}{\partial\alpha}\left(\frac{\partial}{\partial\beta} l_n(\alpha,\beta)\right) = \frac{\partial}{\partial\alpha}\left\{\sum_{i=1}^n [x_i(Y_i - p_i)]\right\} = -\sum_{i=1}^n x_i \frac{\partial p_i}{\partial\alpha} = -\sum_{i=1}^n x_i(1-p_i)p_i$

$E\frac{\partial^2}{\partial\alpha\partial\beta} l_n(\alpha,\beta) = -n\,E[X_1(1-p_1)p_1]$ also by iid assumption.

$$\frac{\partial^2}{\partial\beta^2} l_n(\alpha,\beta) = \frac{\partial}{\partial\beta}\left(\frac{\partial}{\partial\beta} l_n(\alpha,\beta)\right) = \frac{\partial}{\partial\beta}\left\{\sum_{i=1}^{n}[x_i(Y_i - p_i)]\right\} = -\sum_{i=1}^{n} x_i \frac{\partial p_i}{\partial\beta} = -\sum_{i=1}^{n} x_i^2(1-p_i)p_i$$

$$E\frac{\partial^2}{\partial\beta^2} l_n(\alpha,\beta) = -n\, E[X_1^2(1-p_1)p_1], \text{ by iid assumption.}$$

Fisher information matrix

$$I_n(\alpha,\beta) = -\begin{bmatrix} E\frac{\partial^2}{\partial\alpha^2} l_n(\alpha,\beta) & E\frac{\partial^2}{\partial\alpha\partial\beta} l_n(\alpha,\beta) \\ E\frac{\partial^2}{\partial\alpha\partial\beta} l_n(\alpha,\beta) & E\frac{\partial^2}{\partial\beta^2} l_n(\alpha,\beta) \end{bmatrix}$$

$$= n\begin{bmatrix} E_X[p_1(1-p_1)] & E_X[X_1 p_1(1-p_1)] \\ E_X[X_1 p_1(1-p_1)] & E_X[X_1^2 p_1(1-p_1)] \end{bmatrix} \equiv n\, I_1(\alpha,\beta)$$

By the theory of MLE, $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, I_1(\theta)^{-1})$, where $\theta = \begin{bmatrix}\alpha\\\beta\end{bmatrix}$.

Note $p_1(1-p_1) = V(Y_1|X_1)$,

$$I_1(\alpha,\beta)^{-1} = \frac{1}{E_X[X_1^2 V(Y_1|X_1)]\,E[V(Y_1|X_1)] - \{E_X[X_1 V(Y_1|X_1)]\}^2}\begin{bmatrix} E_X[X_1^2 V(Y_1|X_1)] & -E_X[X_1 V(Y_1|X_1)] \\ -E_X[X_1 V(Y_1|X_1)] & E_X[V(Y_1|X_1)] \end{bmatrix}.$$

Larger variability of X_i gives smaller asym. variance.

(c). observed Fisher information matrix

$$\hat{I}_1(\hat\alpha,\hat\beta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(1-\hat p_i)\hat p_i & \frac{1}{n}\sum_{i=1}^{n} x_i(1-\hat p_i)\hat p_i \\ \frac{1}{n}\sum_{i=1}^{n} x_i(1-\hat p_i)\hat p_i & \frac{1}{n}\sum_{i=1}^{n} x_i^2(1-\hat p_i)\hat p_i \end{bmatrix}$$

where $\hat p_i = \frac{e^{\hat\alpha+\hat\beta x_i}}{1+e^{\hat\alpha+\hat\beta x_i}}$, $(\hat\alpha,\hat\beta)$ is MLE. $\hat I_n(\hat\alpha,\hat\beta) = n\hat I_1(\hat\alpha,\hat\beta)$.

Recall $(\hat\alpha,\hat\beta)$ solves $\begin{cases} \sum_{i=1}^{n}(Y_i - p_i) = 0 \\ \sum_{i=1}^{n} x_i(Y_i - p_i) = 0 \end{cases}$

Suppose (α_m, β_m) is m^{th} iteration of the Newton-Raphson algorithm,

$$\begin{bmatrix}\alpha_{m+1}\\\beta_{m+1}\end{bmatrix} = \begin{bmatrix}\alpha_m\\\beta_m\end{bmatrix} + \hat I_n(\alpha_m,\beta_m)^{-1}\begin{bmatrix}\sum_{i=1}^{n}(Y_i - p_{i,m}) \\ \sum_{i=1}^{n} x_i(Y_i - p_{i,m})\end{bmatrix}$$

where $p_{i,m} = \frac{e^{\alpha_m+\beta_m x_i}}{1+e^{\alpha_m+\beta_m x_i}}$.

(d) see attached R script.

$\sqrt{n}\left(\begin{bmatrix}\hat\alpha\\\hat\beta\end{bmatrix} - \begin{bmatrix}\alpha\\\beta\end{bmatrix}\right) \overset{d}{\approx} N(0, \hat I_1(\hat\alpha,\hat\beta)^{-1})$, then $\begin{bmatrix}\hat\alpha\\\hat\beta\end{bmatrix} - \begin{bmatrix}\alpha\\\beta\end{bmatrix} \overset{d}{\approx} N(0, \hat I_n(\hat\alpha,\hat\beta)^{-1})$

2(a). $E_\theta Y = \int_0^1 y\, p(y|\theta)\, dy = \int_0^1 \theta y^\theta\, dy = \frac{\theta}{\theta+1}$

$\mu = \frac{\theta}{\theta+1} \Rightarrow \theta = \frac{\mu}{1-\mu}$, $\mu \in (0,1)$

$p(y|\mu) = \frac{\mu}{1-\mu} y^{\frac{\mu}{1-\mu}-1}$ for $y \in (0,1)$

(b) $E_\theta Y^2 = \int_0^1 y^2 p(y|\theta)\, dy = \int_0^1 \theta y^{\theta+1}\, dy = \frac{\theta}{\theta+2}$

Then $E Y^2 = \frac{\mu}{\mu+2(1-\mu)} = \frac{\mu}{2-\mu}$ in terms of μ.

$V(Y) = EY^2 - (EY)^2 = \frac{\mu}{2-\mu} - \mu^2$

$V(\bar{Y}) = \frac{1}{n} V(Y) = \frac{1}{n}\left(\frac{\mu}{2-\mu} - \mu^2\right)$

(c) log likelihood of $\{Y_i\}_{i=1}^n$: $l_n(\mu) = \sum_{i=1}^n \left(\left(\frac{\mu}{1-\mu}-1\right)\log Y_i + \log\frac{\mu}{1-\mu}\right)$

score fcn : $S_n(\mu) = \frac{d}{d\mu} l_n(\mu) = \sum_{i=1}^n \left(\frac{1}{(1-\mu)^2}\log Y_i + \frac{1}{\mu} + \frac{1}{1-\mu}\right)$

Fisher information : $I_n(\mu) = -E\left[\frac{d}{d\mu} S_n(\mu)\right] = -E\left[\sum_{i=1}^n \left(\frac{2}{(1-\mu)^3}\log Y_i - \frac{1}{\mu^2} + \frac{1}{(1-\mu)^2}\right)\right]$

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

where $E_\theta \log Y = \int_0^1 \log y \cdot \theta y^{\theta-1}\, dy$ $\quad x = \log y$

$= \int_{-\infty}^0 x \cdot \theta e^{(\theta-1)x} e^x\, dx$ $\quad z = -\theta x$

$= -\frac{1}{\theta}\int_0^\infty z e^{-z}\, dz = -\theta^{-1}$

Cramer-Rao lower bound is $I_n(\theta)^{-1} = \frac{1}{n}\mu^2(1-\mu)^2$

$V(\bar{Y}) > I_n(\mu)^{-1}$ for $\mu \in (0,1)$

$\frac{\mu}{2-\mu} - \mu^2 - \mu^2(1-\mu)^2 = \frac{\mu}{2-\mu}\{1-(2-\mu)\mu[1+(1-\mu)^2]\}$

$= \frac{\mu}{2-\mu}(1-\mu)^4 > 0$

(d). First find MLE of θ, then invoke invariance principle of MLE to find MLE of μ.

log likelihood in terms of θ is $l_n(\theta) = \sum_{i=1}^n ((\theta-1)\log Y_i + \log\theta)$

F.O.C. : $\sum_{i=1}^n (\log Y_i + \theta^{-1}) = 0 \Rightarrow \hat{\theta} = \left(\frac{-1}{n}\sum_{i=1}^n \log Y_i\right)^{-1}$

S.O.C. : $-\theta^{-2} < 0$

Therefore, $\hat{\mu}_{MLE} = \frac{\hat{\theta}}{\hat{\theta}+1}$

Simulation is in the R script.

I use inverse CDF method to simulate $\{Y_i\}$.

CDF calculation : $F_Y(t) = \int_0^t \theta y^{\theta-1} dy = t^\theta$ for $t \in (0,1)$.

If we use Cramer-Rao lower bound, we need to compute $E[\hat{\mu}_{MLE}]$.

Alternatively, we know asymptotically, $\sqrt{n}(\hat{\mu}_{MLE} - \mu) \xrightarrow{d} N(0, I_1^{-1}(\mu))$. We can compare the asymptotic variance.

3 (a). Define $E_\theta[t(\vec{Y})] = g(\theta)$, and $S_n(\theta) = \frac{d}{d\theta} \log P_\theta(\vec{Y})$

Since P_θ is differentiable, $g(\theta)$ is differentiable.

By Cauchy-Schwartz inequality,

$$\text{cov}(t(\vec{Y}), S_n(\theta)) \le \sqrt{V(t(\vec{Y}))} \sqrt{V(S_n(\theta))} \quad \cdots\cdots\cdots (*)$$

Note $\text{cov}(t(\vec{Y}), S_n(\theta)) = E(t(\vec{Y}) S_n(\theta)) - E[t(\vec{Y})]E[S_n(\theta)]$

$$E(t(\vec{Y}) S_n(\theta)) = \int t(\vec{Y}) S_n(\theta) P_\theta(\vec{Y}) d\vec{Y}$$

$$= \int t(\vec{Y}) \left(\frac{d}{d\theta} P_\theta(\vec{Y})\right) d\vec{Y} = \frac{d}{d\theta} \int t(\vec{Y}) P_\theta(\vec{Y}) d\vec{Y}$$

$$= g(\theta)'$$

$$E[S_n(\theta)] = \int S_n(\theta) P_\theta(\vec{Y}) d\vec{Y} = \int \frac{d}{d\theta} P_\theta(\vec{Y}) d\vec{Y} = \frac{d}{d\theta} \int P_\theta(\vec{Y}) d\vec{Y} = 0$$

$$V(S_n(\theta)) = I_n(\theta)$$

Therefore, inequality $(*)$ implies $V(t(\vec{Y})) \geq \frac{(g'(\theta))^2}{I_n(\theta)}$

(b) $E\hat{\mu} = E\left(\frac{n}{n+\tau^{-2}} \bar{Y}\right) = \frac{n}{n+\tau^{-2}} E\bar{Y} = \frac{n}{n+\tau^{-2}} \mu$

Following (a)'s notation, $g(\mu) = \frac{n}{n+\tau^{-2}} \mu$, $g(\mu)' = \frac{n}{n+\tau^{-2}}$

log-likelihood $l_n(\mu) = \sum_{i=1}^n -\frac{1}{2}(Y_i - \mu)^2 + \text{const}$.

score fcn : $S_n(\mu) = \sum_{i=1}^n (Y_i - \mu)$

Fisher information : $I_n(\mu) = n$

Lower bound on the variance is $\frac{n}{(n+\tau^{-2})^2}$

For comparison, $V(\hat{\mu}) = V\left(\frac{n}{n+\tau^{-2}} \bar{Y}\right) = \left(\frac{n}{n+\tau^{-2}}\right)^2 V(\bar{Y}) = \frac{n}{(n+\tau^{-2})^2}$,

which equals the lower bound.

4. (a) $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$, then in vector notation, $\vec{Y}_n \sim N(\mu 1_n, \sigma^2 I_n)$.

For any linear transformation A, $A\vec{Y}_n \sim N(\mu A 1_n, \sigma^2 AA^T)$.

Then, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n} 1_n^T \vec{Y}_n \sim N(\mu, \frac{1}{n}\sigma^2)$, therefore $\sqrt{n}(\bar{Y}-\mu)/\sigma \sim N(0,1)$.

(b) $(n-1)S^2/\sigma^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma} - \frac{\bar{Y}-\mu}{\sigma}\right)^2 \sim \chi^2_{n-1}$ by fact (a).

(c) $\sqrt{n}(\bar{Y}-\mu)/S = \sqrt{n}\frac{(\bar{Y}-\mu)}{\sigma} \Big/ \sqrt{S^2/\sigma^2}$

from part (a), $\sqrt{n}(\bar{Y}-\mu)/\sigma \sim N(0,1)$

from part (b), $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

by fact (a), $\sqrt{n}(\bar{Y}-\mu)/\sigma$ and S^2/σ^2 are independent.

by fact (b), $\sqrt{n}(\bar{Y}-\mu)/S \sim t_{n-1}$

(d) Test statistic is $T_n = \frac{\sqrt{n}(\bar{Y}-\mu_0)}{S}$.

Level α test is to accept "$\mu=\mu_0$" if $|T_n| \le Q_{t,n-1}(1-\frac{\alpha}{2})$

where $Q_{t,n-1}(1-\frac{\alpha}{2})$ is the $(1-\frac{\alpha}{2})^{th}$ quantile of t-distribution with d.f. $n-1$.