STA 532 Homework8

Jae Hyun Lee, jl914

27 March, 2020

HW8 for STA-532

1.

(a)

$$
\begin{aligned}
(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n) \sim^{iid} P(x, y) &= P(x) f(y \mid x, \alpha, \beta) \\
f(y_i \mid x_i, \alpha, \beta) &= p_i^{y_i} (1 - p_i)^{1 - y_i} \rightarrow l(y_i, x_i, \alpha, \beta) \\
&= l_{x_i} + log(1 - p_i) + y_i log(\frac{p_i}{1 - p_i}) \\
\rightarrow f(Y \mid X, \alpha, \beta) &= \sum_{i=1}^{n} l_{x_i} + \sum_{i=1}^{n} log f(y_i \mid x_i, \alpha, \beta) \\
&= \sum_{i=1}^{n} l_{x_i} + \sum_{i=1}^{n} log(1 - p_i) + \sum_{i=1}^{n} y_i log(\frac{p_i}{1 - p_i})
\end{aligned}
$$

Before investigate derivative of p_i respect to α, β

$$
\begin{aligned}
\frac{d}{d\alpha} p_i &= \frac{e^{\alpha + \beta x_i}(1 + e^{\alpha + \beta x_i}) - e^{2(\alpha + \beta x_i)}}{(1 + e^{\alpha + \beta x_i})^2} = p_i(1 - p_i) \\
\frac{d}{d\beta} p_i &= \frac{x_i e^{\alpha + \beta x_i}(1 + e^{\alpha + \beta x_i}) - x_i e^{2(\alpha + \beta x_i)}}{(1 + e^{\alpha + \beta x_i})^2} = x_i p_i(1 - p_i)
\end{aligned}
$$

The score function is as follow:

$$
\begin{aligned}
s(x_i, y_i, \alpha, \beta) = \frac{d}{d\theta} l(x_i, y_i, \alpha, \beta) &= \begin{bmatrix} \frac{d}{d\alpha}(x_i, y_i, \alpha, \beta) \\ \frac{d}{d\beta}(x_i, y_i, \alpha, \beta) \end{bmatrix} \\
&= \begin{bmatrix} \frac{\frac{d}{d\alpha} p_i}{(1 - p_i)} + y_i \frac{d}{d\alpha}(\alpha + \beta x_i) \\ \frac{\frac{d}{d\beta} p_i}{(1 - p_i)} + y_i \frac{d}{d\beta}(\alpha + \beta x_i) \end{bmatrix} \\
&= \begin{bmatrix} -p_i + y_i \\ -x_i p_i + x_i y_i \end{bmatrix}
\end{aligned}
$$

The equation that determines MLE is

$$
s(X, Y, \alpha, \beta) = \begin{bmatrix} -\sum p_i + \sum y_i \\ -\sum x_i p_i + \sum x_i y_i \end{bmatrix}
$$

(b)

$$
\begin{aligned}
\nabla^2 l(X,Y,\alpha,\beta) &= \frac{d}{d\theta}s(X,Y,\alpha,\beta) = \begin{bmatrix} -\sum p_i(1-p_i) & -\sum x_i p_i(1-p_i) \\ -\sum x_i p_i(1-p_i) & -\sum x_i^2 p_i(1-p_i) \end{bmatrix} \\
I(\theta) = -E(\nabla^2 l(X,Y,\alpha,\beta)) &= \begin{bmatrix} \sum p_i(1-p_i) & \sum E(x_i) p_i(1-p_i) \\ \sum E(x_i) p_i(1-p_i) & \sum E(x_i^2) p_i(1-p_i) \end{bmatrix} \\
\rightarrow \hat{\theta}_{mle} &= \begin{bmatrix} \hat{\alpha}_{mle} \\ \hat{\beta}_{mle} \end{bmatrix} \sim N(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, I(\theta)^{-1}) \quad \textit{asymptonically}
\end{aligned}
$$

and we can find that at information matrix, x_i are represented with its moments. Thus we can conclude that variance of $\hat{\theta}_{mle}$ depends on first and second moment of x_i's distribution.

(c)

for observed information $\hat{I}(\theta)$

$$
\hat{I}(\theta) = -\frac{1}{n}\nabla^2 l(X,Y,\alpha,\beta) = \begin{bmatrix} \frac{1}{n}\sum p_i(1-p_i) & \frac{1}{n}\sum x_i p_i(1-p_i) \\ \frac{1}{n}\sum x_i p_i(1-p_i) & \frac{1}{n}\sum x_i^2 p_i(1-p_i) \end{bmatrix}
$$

Newton Raphson method uses first order approximation. We want to find $\tilde{\theta}$ that makes $l(\tilde{\theta}) = 0$ Let θ_0 be initial guess and $\tilde{\theta} = \theta_0 + h$ then $l(\tilde{\theta}) = 0 \approx l(\theta_0) + hl'(\theta_0) \rightarrow h = -\frac{l(\theta_0)}{l'(\theta_0)}$. Thus our next guess θ_1 is $\theta_1 = \theta_0 - \frac{l(\theta_0)}{l'(\theta_0)} \rightarrow \theta_n = \theta_{n-1} - \frac{l(\theta_{n-1})}{l'(\theta_{n-1})}$. In this case, $\theta = (\alpha, \beta)$, $l'(\theta)^{-1} \rightarrow \nabla l(\theta)^{-1} = \begin{bmatrix} \frac{d}{d\alpha} l(\theta) \\ \frac{d}{d\beta} l(\theta) \end{bmatrix}^{-1}$.

In the same way, if we replace log likelihood function with observed score function, we can find $\hat{\theta}_{mle}$ that makes $s(\hat{\theta}_{mle}) = 0$ and $\theta_n = \theta_{n-1} - s(\theta) \times s'(\theta)^{-1} = \theta_{n-1} - \nabla l(\theta) \times \nabla^2 l(\theta)^{-1}$

2.

(a)

$$
\begin{aligned}
&E(Y) = \int_0^1 \theta y^\theta dy = \frac{\theta}{\theta+1} \\
&\mu = \frac{\theta}{\theta+1} \rightarrow \mu = 1 - \frac{1}{\theta+1} \rightarrow \theta + 1 = \frac{1}{1-\mu} \rightarrow \theta = \frac{\mu}{1-\mu} \\
&\rightarrow P(y \mid \mu) = \frac{\mu}{1-\mu} y^{\frac{\mu}{1-\mu}-1}
\end{aligned}
$$

(b)

$$
\begin{aligned}
V(Y) = E(Y^2) - E(Y)^2 &= \frac{\theta}{\theta+2} - (\frac{\theta}{\theta+1})^2 \\
&= \frac{\frac{\mu}{1-\mu}}{\frac{2-\mu}{1-\mu}} - \mu^2 \\
&= \frac{\mu}{2-\mu} - \mu^2 \\
&= \frac{\mu(1-\mu)^2}{2-\mu} \\
\rightarrow V(\bar{Y}) = \frac{\mu(1-\mu)^2}{n(2-\mu)}
\end{aligned}
$$

(c)

$$
\begin{aligned}
L(Y,\mu) &= (\frac{\mu}{1-\mu})^n (\prod y_i)^{\frac{\mu}{1-\mu}-1} \\
l(Y,\mu) &= nlog(\frac{\mu}{1-\mu}) + (\frac{\mu}{1-\mu}-1)\sum log(y_i) \\
\frac{d}{d\mu} l(Y,\mu) &= n(\frac{1-\mu}{\mu})\frac{1}{(1-\mu)^2} + \frac{1}{(1-\mu)^2}\sum log(y_i) \\
\frac{d^2}{d\mu^2} l(Y,\mu) &= -n[1/\mu^2 - 1/(1-\mu)^2] + \frac{2}{(1-\mu)^3}\sum log(y_i) \\
nI(\mu) &= -E(\frac{d^2}{d\mu^2} l(Y,\mu)) \\
&= \frac{n(-2\mu+1)}{\mu^2(1-\mu)^2} - \frac{2}{(1-\mu)^3} E(\sum log(y_i)) \\
&= \frac{n(-2\mu+1)}{\mu^2(1-\mu)^2} - \frac{2}{(1-\mu)^3} \sum E(log(y_i))
\end{aligned}
$$

Let $x_i = -log(y_i)$ then

$$
\begin{aligned}
P(x_i \mid \theta) &= P(y_i \mid \theta)\frac{dy}{dx} \\
&= \theta e^{-(\theta-1)x_i} \times |\frac{d}{dx} e^{-x_i}| \\
&= \theta e^{-\theta x_i} \rightarrow E(x_i) = \frac{1}{\theta} = \frac{1-\mu}{\mu} \\
-\sum E(log(y_i)) &= \frac{n(1-\mu)}{\mu} \\
\rightarrow nI(\mu) &= \frac{n(-2\mu+1)}{\mu^2(1-\mu)^2} + \frac{2n}{\mu(1-\mu)^2} = \frac{n}{\mu^2(1-\mu)^2}
\end{aligned}
$$

Comparing variance of CR lower bound and variance of $\bar{Y}$

$$
\begin{aligned}
&V(\bar{Y}) = \frac{\mu(1-\mu)^2}{n(2-\mu)}, (nI(\mu))^{-1} = \frac{\mu^2(1-\mu)^2}{n} \\
&\rightarrow nI(\mu)V(\bar{Y}) = \frac{1}{\mu(2-\mu)} \geq 1 \rightarrow -\mu^2 + 2\mu \leq 1 \rightarrow -(\mu-1)^2 \leq 0 \\
&Thus \quad V(\bar{Y}) \geq \frac{1}{nI(\mu)}
\end{aligned}
$$

(d)

$\hat{\mu}_{mle}$ is point that $\frac{d}{d\mu} l(y,\hat{\mu}) = 0$

$$
\begin{aligned}
\frac{d}{d\mu} l(y,\mu) &= \frac{n}{\mu(1-\mu)} + \frac{\sum log(y_i)}{(1-\mu)^2} \\
&= \frac{n(1-\mu)\sum log(y_i)}{\mu(1-\mu)^2} = 0 \\
&\rightarrow n + (1-\hat{\mu}_{mle})\sum log(y_i) = 0 \\
&\rightarrow \hat{\mu}_{mle} = \frac{n + \sum log(y_i)}{\sum log(y_i)}
\end{aligned}
$$

In class, we have shown that asymptonicaaly $\hat{\mu}_{mle} \sim N(\mu, \frac{1}{nI(\mu)})$ and we have shown that $V(\bar{Y}) \geq \frac{1}{nI(\mu)}$. Thus we can conclude that $V(\hat{\mu}_{mle}) \leq V(\bar{Y})$.

3.

(a)

As we have shown for unbiased estimator of $\hat{\theta}$, consider correlation of $t = t(y_1, \cdots, y_n)$ and $l'(\theta_0) = \sum s(y_i, \theta_0)$. Then

$$\begin{aligned}
& -1 \leq cor(t, l'(\theta_0)) = \frac{cov(t, l'(\theta_0))}{\sqrt{V(t)V(l'(\theta_0))}} \leq 1 \\
& \rightarrow \frac{cov(t, l'(\theta_0))^2}{V(t)V(l'(\theta_0))} \leq 1 \quad and \quad V(l'(\theta_0)) = nI(\theta_0) \\
& \rightarrow \frac{cov(t, l'(\theta_0))^2}{n} \leq V(t)
\end{aligned}$$

and

$$\begin{aligned}
cov(t, l'(\theta_0)) &= E(tl'(\theta_0)) - E(t)E(l'(\theta_0)) \\
&= E(t, l'(\theta_0)) \\
&= \int t \frac{d}{d\theta} log(\prod f(y_i \mid \theta)) \prod f(y_i \mid \theta) dy_1 \cdots dy_n \\
&= \int t \frac{\frac{d}{d\theta} \prod f(y_i \mid \theta)}{\prod f(y_i \mid \theta)} \times \prod f(y_i \mid \theta) dy_1 \cdots dy_n \\
&= \int t \frac{d}{d\theta} \prod f(y_i \mid \theta) dy_1 \cdots dy_n \\
&= \frac{d}{d\theta} E(t) \\
\rightarrow V(t) &\geq \frac{(\frac{d}{d\theta} E(t))^2}{nI(\theta_0)}
\end{aligned}$$

(b)

Now

$$\begin{aligned}
t &= \frac{n\bar{y}}{n + 1/\tau^2}, \theta = \mu \\
E(t) &= \frac{n}{n + 1/\tau^2} \mu \rightarrow (\frac{d}{d\mu} E(t))^2 = (\frac{n}{n + 1/\tau^2})^2 \\
\rightarrow V(\hat{\mu}_b) &\geq \frac{(\frac{n}{n+1/\tau^2})^2}{nI(\theta_0)}
\end{aligned}$$

I assume that frequentist's estimate is MLE. For model $y_1 \cdots y_n \sim N(\mu, 1)$, $\hat{\mu}_{mle} = \bar{y}$ and $V(\hat{\mu}_{mle}) = \frac{1}{nI(\mu)}$ but since $1/\tau^2 > 0 \rightarrow \frac{n}{n+1/\tau^2} \leq 1 \rightarrow \frac{(\frac{n}{n+1/\tau^2})^2}{nI(\mu)} \leq \frac{1}{nI(\mu)}$. Thus we can conclude that $V(\hat{\mu}_b) \leq V(\hat{\mu}_{mle})$.

4.

(a)

$$
\begin{aligned}
&Y_i \cdots Y_n \sim^{iid} N(\mu, \sigma^2) \rightarrow E(\bar{Y}), V(\bar{Y}) = \sigma^2/n \\
&By\ properties\ of\ normal\ distribution \quad \bar{Y} \sim N(\mu, \sigma^2/n) \\
&Then\ \bar{Y} - \mu \sim N(0, \sigma^2/n) \rightarrow \sqrt{n}(\bar{Y} - \mu)/\sigma \sim N(0,1) : Standard\ normal
\end{aligned}
$$

(b)

$$
\begin{aligned}
(n-1)S^2/\sigma^2 &= \sum\left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2 \\
&= \frac{1}{\sigma^2}\sum(Y_i - \mu - (\bar{Y} - \mu))^2 \\
&= \frac{1}{\sigma^2}\sum[(Y_i - \mu)^2 - 2(Y_i - \mu)(\bar{Y} - \mu) + (\bar{Y} - \mu)^2] \\
&= \sum\left(\frac{Y_i - \mu}{\sigma}\right)^2 - \frac{n(\bar{Y} - \mu)^2}{\sigma^2}
\end{aligned}
$$

We know that $\frac{Y_i - \mu}{\sigma} = Z_i \sim N(0,1) \rightarrow Z_i^2 \sim \chi_1^2$ and $\frac{\sqrt{n}(\bar{Y} - \mu)}{\sigma} = \bar{Z} \sim N(0,1) \rightarrow \bar{Z}^2 \sim \chi_1^2$

In conclusion: $(n-1)S^2/\sigma^2 = \sum Z_i^2 - \bar{Z}^2 \sim \chi_{n-1}^2$

(c)

$$
\sqrt{n}(\bar{Y} - \mu)/S = \frac{\sqrt{n}(\bar{Y} - \mu)/\sigma}{\sqrt{\S^2/\sigma^2}} = \frac{\sqrt{n}(\bar{Y} - \mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} \sim t_{n-1}
$$

by confirmed facts and definition provided in Question.

(d)

We have found that $\sqrt{n}(\bar{Y} - \mu)/S \sim t_{n-1}$ Thus for $T \sim t_{n-1}$

$$
\begin{aligned}
&Pr(-t_{n-1,\alpha/2} \le T \le t_{n-1,\alpha/2}) = 1 - \alpha \\
&\rightarrow Pr(-t_{n-1,\alpha/2} \le \sqrt{n}(\bar{Y} - \mu)/S \le t_{n-1,\alpha/2}) = 1 - \alpha \\
&\rightarrow Pr(-t_{n-1,\alpha/2} \le \sqrt{n}(\bar{Y} - \mu_0)/S \le t_{n-1,\alpha/2} \mid H) = 1 - \alpha \\
&\rightarrow Pr(-t_{n-1,\alpha/2} \times S/\sqrt{n} \le (\bar{Y} - \mu_0) \le t_{n-1,\alpha/2} \times S/\sqrt{n} \mid H) = 1 - \alpha \\
&\rightarrow Pr(\bar{Y} - t_{n-1,\alpha/2} \times S/\sqrt{n} \le \mu_0 \le \bar{Y} + t_{n-1,\alpha/2} \times S/\sqrt{n} \mid H) = 1 - \alpha
\end{aligned}
$$

In conclusion, test statistics $= \bar{Y}$ and acceptance region $= (\bar{Y} - t_{n-1,\alpha/2} \times S/\sqrt{n}, \bar{Y} + t_{n-1,\alpha/2} \times S/\sqrt{n})$