STA 532 Homework7

Jae Hyun Lee, jl914

19 March, 2020

HW7 for STA-532

1.

(a)

$$\begin{aligned}
E(Y_i) = E(e^{X_i}) &= \int e^{x_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} dx_i \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i^2 - 2x_i\mu - 2\sigma^2 x_i + \mu^2)}{2\sigma^2}} dx_i \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-(\mu+\sigma^2))^2}{2\sigma^2}} dx_i \times e^{-\frac{\mu^2-(\mu+\sigma^2)^2}{2\sigma^2}} \\
&= e^{-\frac{-\sigma^4-2\sigma^2\mu}{2\sigma^2}} = e^{\sigma^2/2+\mu} = \phi \\
E(Y_i^2) = E(e^{2x_i}) &= e^{-\frac{\mu^2-(\mu+2\sigma^2)^2}{2\sigma^2}} \\
&= e^{-\frac{-4\sigma^4-4\sigma^2\mu}{2\sigma^2}} = e^{2\sigma^2+2\mu} \\
V(Y_i) = E(Y_i^2) - E(Y_i)^2 &= e^{2\sigma^2+2\mu} - e^{\sigma^2+2\mu} = e^{\sigma^2+2\mu}(e^{\sigma^2}-1)
\end{aligned}$$

$$\begin{aligned}
E(\bar{Y}) &= \frac{\sum E(Y_i)}{n} = e^{\sigma^2/2+\mu} \\
V(\bar{Y}) &= \frac{\sum V(Y_i)}{n^2} (because\ of\ iid) = \frac{e^{\sigma^2+2\mu}(e^{\sigma^2}-1)}{n}
\end{aligned}$$

(b)

$$\begin{aligned}
&P(Y_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma Y_i} exp\{-\frac{(logY_i - \mu)^2}{2\sigma^2}\} \quad by\ change\ of\ variable. \\
&\phi = e^{\sigma^2/2+\mu} \to log\phi = \sigma^2/2 + \mu \to \mu = -\sigma^2/2 + log\phi \quad where\ 0 < \phi < \infty \\
&P(Y_i \mid \phi, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma Y_i} exp\{-\frac{(logY_i - log\phi)^2}{2\sigma^2}\} e^{-1/4} \\
&l(y, \phi, \sigma^2) = c - \frac{n}{2} log\sigma^2 - \frac{1}{2\sigma^2} \sum (log(Y_i/\phi))^2 \\
&\frac{d}{d\phi} l(y, \phi, \sigma^2) = \frac{1}{\phi\sigma^2} \sum log(Y_i/\phi) = 0 \to log\phi = \frac{1}{n} \sum logY_i \to \hat{\phi}_{mle} = (\prod y_i)^{1/n}
\end{aligned}$$

$$\nabla l(y,\phi,\sigma^2) = \begin{pmatrix} \frac{1}{\phi\sigma^2}\sum log(Y_i/\phi) \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum(log(Y_i/\phi))^2 \end{pmatrix} \rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum(log(Y_i/\phi))^2$$
$$\nabla^2 l(y,\phi,\sigma^2) = \begin{pmatrix} -\frac{1}{\sigma^2\phi^2}\sum log(Y_i)+\frac{n}{\phi^2\sigma^2}log(\phi)-\frac{n}{\phi^2\sigma^2} & -\frac{1}{\sigma^4\phi}\sum log(Y_i/\phi) \\ -\frac{1}{\sigma^4\phi}\sum log(Y_i/\phi) & \frac{1}{2\sigma^4}-\frac{1}{4\sigma^6}\sum(log(Y_i/\phi)^2) \end{pmatrix}$$
$$I_n(\phi,\sigma^2) = -E(\nabla^2 l(y,\phi,\sigma^2)) = \begin{pmatrix} \frac{n}{\phi^2\sigma^2} & 0 \\ 0 & V(\sigma^2) \end{pmatrix} \rightarrow V(\hat{\phi}) = \phi^2\sigma^2/n$$
$$V(\bar{Y}) = e^{\sigma^2+2\mu}(e^{\sigma^2}-1)/n, V(\hat{\phi}) = e^{\sigma^2+2\mu}\sigma^2/n$$
$$\rightarrow compare\ e^{\sigma^2}-1\ and\ \sigma^2$$
$$\rightarrow if\ 0<\sigma^2<1\ V(\bar{Y})\ is\ smaller,\ o.w\ V(\hat{\phi})\ is\ smaller$$

2.

(a)

$$P(Y_i) = \frac{1}{\Gamma(a)}b^a Y_i^{a-1}e^{bY_i}$$
$$Let\ \theta = (a,b)$$
$$l(\theta,y) = -nlog\Gamma(a)+nalog(b)+(a-1)\sum logY_i - b\sum Y_i$$
$$\nabla l(\theta,y) = \begin{pmatrix} -n\frac{d}{da}log\Gamma(a)+nlogb+\sum logY_i \\ \frac{na}{b}-\sum Y_i \end{pmatrix}$$
$$\rightarrow \hat{b}_{mle} = \frac{a}{\bar{Y}}, \hat{a}_{mle} = solution\ of\ first\ equation.$$

(b)

$$\nabla^2 l(\theta,y) = \begin{pmatrix} -n\frac{d^2}{da^2}log\Gamma(a) & n/b \\ n/b & -\frac{na}{b^2} \end{pmatrix}$$
$$\rightarrow I_n(\theta) = n\begin{pmatrix} \frac{d^2}{da^2}log\Gamma(a) & -1/b \\ -1/b & \frac{a}{b^2} \end{pmatrix}$$
$$by\ using\ result\ of\ asymptotic\ distribution\ of\ MLE$$
$$\begin{pmatrix} \hat{a}_{mle} \\ \hat{b}_{mle} \end{pmatrix} \sim N\left(\begin{pmatrix} a \\ b \end{pmatrix}, I_n^{-1}(\theta)\right)$$
$$where\ I_n^{-1}(\theta) = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

(c)

$$
\begin{aligned}
\hat{\mu}_{mle} &= \frac{\hat{a}_{mle} \sim N(a, \sigma_a^2)}{\hat{b}_{mle} \sim N(b, \sigma_b^2)} \\
&= \frac{a + \hat{a} \sim N(0, \sigma_a^2)}{b + \hat{b} \sim N(0, \sigma_b^2)} \\
&= \frac{a}{b} \times \frac{1 + \hat{a}/a}{1 + \hat{b}/b} \\
\rightarrow log(\hat{\mu}_{mle}) &= log(a/b) + log(1 + \hat{a}/a) + log(1 + \hat{b}/b) \\
since\ log(1+\delta) &= \delta - \delta^2/2 + \delta^3/3 \cdots \\
log(\hat{\mu}_{mle}) &= log(a/b) + \hat{a}/a - \hat{b}/b \quad asymptotically \\
&\sim N(log(a/b), \frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2}) \\
\rightarrow \hat{\mu}_{mle} &\sim logN(log(a/b), \frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2})
\end{aligned}
$$

From result of previous question, we can find that

$$
\begin{aligned}
V(\hat{\mu}_{mle}) &= exp\{\frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2} + 2log\mu\}(exp\{\frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2}\} - 1) \\
V(\bar{Y}) &= \frac{a}{nb^2}
\end{aligned}
$$

Since we cannot have closed form of Fisher information matrix, we cannot directly compare the variance of two estimators. But I expect var($\hat{\mu}_{mle}$) to be smaller because it based on assumption about parametric space.

3.

(a)

$$
\begin{aligned}
P(Y \mid \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y-\mu)^2}{2\sigma^2}} \\
logP(Y \mid \mu, \sigma^2) &= C - \frac{1}{2} log\sigma^2 - \frac{(Y-\mu)^2}{2\sigma^2} \\
and\ E(Y) &= a/b\ E(Y^2) = \frac{a^2 + a}{b^2} \\
E[logP(Y \mid \mu, \sigma^2)] &= C - \frac{1}{2} log\sigma^2 - \frac{1}{2\sigma^2}[E(Y^2) - 2E(Y)\mu + \mu^2] \\
&= C - \frac{1}{2} log\sigma^2 - \frac{1}{2\sigma^2} \times \frac{a^2 + a}{b^2} + \frac{a\mu}{\sigma^2 b} - \frac{\mu^2}{2\sigma^2} \\
\nabla E[logP(Y \mid \mu, \sigma^2)] &= \begin{pmatrix} \frac{1}{\sigma^2}(\frac{a}{b} - \mu) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(\frac{a^2+a}{b^2} - \frac{2a}{b}\mu + \mu^2) \end{pmatrix} \\
&\rightarrow \hat{\mu} = a/b, \\
\hat{\sigma}^2 &= \frac{a^2 + a}{b^2} - \frac{2a}{b}\hat{\mu} + \hat{\mu}^2 = \frac{a^2 + a}{b^2} - \frac{2a^2}{b^2} + a^2/b^2 = \frac{a}{b^2}
\end{aligned}
$$

(b)

$$
\nabla E[logP(Y \mid \mu, \sigma^2)] = E[\nabla logP(Y \mid \mu, \sigma^2)] = E(S(\mu, \sigma^2))
$$

In class, we have shown that expectation of score function is zero at true parameter. It indicates that true parameter $\mu_0 = a/b, \sigma_0^2 = a/b^2$ as we have shown above. Also if we assume well seperated maximum and uniformly convergence, MLE converges to true parameter in probability. Thus $\hat{\mu}_{mle} \to^p a/b, \hat{\sigma}^2_{mle} \to^p a/b^2$.

(c)

$$\nabla^2 E[logP(Y \mid \mu, \sigma^2)] = \begin{pmatrix} -1/\sigma^2 & \frac{1}{\sigma^4}(-\frac{a}{b}+\mu) \\ \frac{1}{\sigma^4}(-\frac{a}{b}+\mu) & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(\frac{a^2+a}{b^2} - \frac{2a}{b}\mu + \mu^2) \end{pmatrix}$$
$$\to I(\mu, \sigma^2) = -\nabla^2 E[logP(Y \mid \mu, \sigma^2) \to I_n(\mu, \sigma^2) = -n\nabla^2 E[logP(Y \mid \mu, \sigma^2)]$$

By asymptotic normality of MLE and plug-in approach,

$$\begin{pmatrix} \hat{\mu}_{mle} \\ \hat{\sigma^2}_{mle} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_0 \\ \sigma_0^2 \end{pmatrix}, I_n(\hat{\mu}, \hat{\sigma}^2)^{-1}\right) \quad where\ \hat{\mu} = a/b, \hat{\sigma}^2 = a/b^2$$
$$\to \sim N\left(\begin{pmatrix} a/b \\ a/b^2 \end{pmatrix}, \begin{pmatrix} \frac{a}{nb^2} & 0 \\ 0 & \frac{a^2}{2nb^4} \end{pmatrix}\right)$$
$$\to \quad V(\hat{\mu}_{mle}) = \frac{a}{nb^2}$$

We can find that $V(\hat{\mu}_{mle}) = V(\bar{Y})$ which expected to be larger than previous $V(\hat{\mu}_{mle})$. Thus $V(\hat{\mu}_{mle})$ based on normal model will be larger than $V(\hat{\mu}_{mle})$ based on gamma as in previous question.

(d)

According tto findings at Q2 and Q3, this model misspecification shows that estimator from misspecified model has same variance of estimator based on nonparametric setup. In this case, misspecification effect seems not much severe. But if we misspecify model in other case, it might produce worse estimator than nonparametric inference. Therefore, in this example, I could find the importance of correct specification of model in parametric inference.

4.

(a)

$$f(\tilde{x} \mid \theta) = \prod f(x_i \mid \theta)$$
$$\to l(\theta, \tilde{x}) = \sum logf(x_i \mid \theta)$$
$$f(y_i \mid \theta) = f(g^{-1}(y_i \mid \theta))\frac{d}{dy_i}g^{-1}(y_i)$$
$$\to l(\theta, \tilde{y}) = \sum logP(y_i \mid \theta) = \sum log[f(g^{-1}(y_i \mid \theta))\frac{d}{dy_i}g^{-1}(y_i)]$$

(b)

$$\hat{\theta}_{mle}\ from\ A = argmax_\theta \sum logf(x_i \mid \theta)$$
$$\rightarrow s(\theta, \tilde{x}) = \frac{d}{d\theta} \sum logf(x_i \mid \theta)$$
$$\hat{\theta}_{mle}\ from\ B = argmax_\theta \sum log[f(g^{-1}(y_i \mid \theta))\frac{d}{dy_i}g^{-1}(y_i)]$$
$$\rightarrow s(\theta, \tilde{y}) = \frac{d}{d\theta} \sum log[f(g^{-1}(y_i \mid \theta))\frac{d}{dy_i}g^{-1}(y_i)]$$
$$\rightarrow s(\theta, \tilde{y}) = \frac{d}{d\theta} \sum log[f(g^{-1}(y_i \mid \theta)) \quad because \frac{d}{d\theta}log[\frac{d}{dy_i}g^{-1}(y_i)] = 0$$
$$\rightarrow s(\theta, \tilde{y}) = \frac{d}{d\theta} \sum log[f(g^{-1}(y_i \mid \theta)) = \frac{d}{d\theta} \sum logf(x_i \mid \theta) = s(\theta, \tilde{x})$$
$Thus\ \hat{\theta}_{mle}\ that\ makes\ score\ function\ zero\ is\ identical$
$$I_n(\theta)\ from\ A = -E[\frac{d^2}{d\theta^2} \sum logf(x_i \mid \theta)] = -E[\frac{d}{d\theta}s(\theta, \tilde{x})]$$
$$I_n(\theta)\ from\ B = -E[\frac{d^2}{d\theta^2} \sum log[f(g^{-1}(y_i \mid \theta))\frac{d}{dy_i}g^{-1}(y_i)]] = -E[\frac{d}{d\theta}s(\theta, \tilde{y})]$$
$since\ score\ functions\ are\ identical\ above\ two\ Fisher\ information\ function\ will\ be\ identical\ too.$
$\rightarrow V(\hat{\theta}_{mle})\ is\ same$

5.

(a)

$$\begin{aligned}
&\bar{Y}\ is\ estimator\ of\ \theta/2 \rightarrow \hat{\theta} = 2\bar{Y} \\
&P(y_i) = \frac{1}{\theta}I(0 < y_i < \theta) \\
&\rightarrow E(y_i) = \theta/2, E(y_i^2) = \theta^2/3 \\
&\rightarrow V(y_i) = \theta^2/12, V(\bar{Y}) = \theta^2/12n \rightarrow V(\hat{\theta}) = \theta^2/3n
\end{aligned}$$

(b)

$$\begin{aligned}
L(\theta, \tilde{y}) &= 1/\theta^n \prod I(0 < y_i < \theta) \\
&= 1/\theta^n I(0 < y_{(n)} < \theta) \quad where\ y_{(n)}\ is\ the\ largest\ sample \\
and\ \hat{\theta}_{mle} &= y_{(n)}
\end{aligned}$$

(c)

$$\begin{aligned}
&F_{y_{(n)}} = P(y_{(n)} \leq y) = \prod P(y_i \leq y) = F(y)^n \\
&f_{y_{(n)}} = \frac{d}{dy}F(y)^n = nF(y)^{n-1}f(y) = n\frac{y^{n-1}}{\theta^n} \\
&P(\hat{\theta}_{mle}) = n\frac{\hat{\theta}_{mle}^{n-1}}{\theta^n}
\end{aligned}$$

(d)

$$
\begin{aligned}
E(y_{(n)}) &= \int_0^\theta ny\frac{y^{n-1}}{\theta^n}dy = \frac{n}{n+1}\theta \\
E(y_{(n)}^2) &= \int_0^\theta ny^2\frac{y^{n-1}}{\theta^n}dy = \frac{n}{n+2}\theta^2 \\
V(y_{(n)}) &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta^2}{(n+2)(n+1)^2} \\
MSE = bias^2 + V(y_{(n)}) &= \frac{1}{(n+1)^2}\theta^2 + \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{(n+1)(n+2)}
\end{aligned}
$$

$\hat{\theta}$ from (a) is unbiased and $V(\hat{\theta}) = \theta^2/3n$. Thus compared to (a), MSE decrease much faster for $\hat{\theta}_{mle}$