STA 532 Homework4

Jae Hyun Lee, jl914

09 February, 2020

HW4 for STA-532

1.

The usual arithmetic mean $\bar{y}$ of a sample $\{y_1, \cdots, y_n\}$ is highly sensitive to outliers, so sometimes we analyze $\{x_1, \cdots, x_n\} = \{ln\ y_1, \cdots, ln\ y_n\}$ or compute the sample mean on this scale.

(a)

Show that $e^{\bar{x}} \leq \bar{y}$.

$$
\begin{aligned}
&E(x_i) = \bar{x} = \sum_{i=1}^{n} x_i/n, \quad E(y_i) = \bar{y} = \sum_{i=1}^{n} y_i/n \quad which\ are\ sample\ mean \\
&and \quad f(y) = logy \quad which\ is\ concave\ function \\
&then \quad E(f(y_i)) = E(x_i) = \bar{x} \\
&By\ Jensen's\ inequality \quad E(f(y_i)) \leq f(E(y_i)) \\
&\rightarrow \bar{x} \leq log(\bar{y}) \rightarrow e^{\bar{x}} \leq \bar{y}
\end{aligned}
$$

(b)

Compute magnitude of three different types of means given by $m(y_1, \cdots, y_n) = f^{-1}(\sum f(y_i)/n)$ for f(y) = 1/y, g(y) = ln(y), h(y) = y

With same logic above,

$$
\begin{aligned}
&f(y) = 1/y \quad convex\ function \\
&By\ Jensen's\ inequality \\
&f(\sum y_i/n) \leq \sum f(y_i)/n \rightarrow \bar{y} \leq f^{-1}(\sum f(y_i)/n) = m_1(y_1, \cdots, y_n)
\end{aligned}
$$

$$
\begin{aligned}
&g(y) = ln(y) \quad concave\ function \\
&By\ Jensen's\ inequality \\
&\sum g(y_i)/n \leq g(\sum y_i/n) \rightarrow m_2(y_1, \cdots, y_n) = g^{-1}(\sum f(y_i)/n) \leq \bar{y}
\end{aligned}
$$

$$
\begin{aligned}
&h(y) = y \\
&\rightarrow m_3(y_1, \cdots, y_n) = \sum h(y_i)/n = \sum y_i/n = \bar{y}
\end{aligned}
$$

That is,

$$
m_2(y_1, \cdots, y_n) \leq m_3(y_1, \cdots, y_n) \leq m_1(y_1, \cdots, y_n)
$$

(c)

For each type of mean in (b), compute the sensitivity to outliers by approximating $m(y_1+\delta,\cdots,y_n)$ with $m(y_1,\cdots,y_n)+\delta\times\frac{d}{dy_1}m(y_1,\cdots,y_n)$. Interpret your result.

$$\begin{aligned}
m_1(y_1,\cdots,y_n) &= (\frac{1}{n}\sum_{i=1}^{n} y_i^{-1})^{-1}\\
\frac{d}{dy_1}m_1(y_1\cdots y_n) &= (\frac{1}{n}\sum_{i=1}^{n} y_i^{-1})^{-2}\times\frac{1}{ny_1^2}\\
m_1(y_1+\delta,\cdots,y_n) &\approx \frac{\delta}{n}\frac{m_1(y_1,\cdots y_n)^2}{y_1^2}
\end{aligned}$$

$$\begin{aligned}
&m_2(y_1,\cdots,y_n) = exp\{\sum ln(y_i)/n\}\\
&\frac{d}{dy_1}m_2(y_1\cdots y_n) = exp\{\sum ln(y_i)/n\}\times\frac{d}{dy_1}(\sum ln(y_i)/n) = exp\{\sum ln(y_i)/n\}\times\frac{1}{ny_1}\\
&m_2(y_1+\delta,\cdots,y_n)\approx\frac{\delta}{n}\frac{m_2(y_1,\cdots y_n)}{y_1}
\end{aligned}$$

$$\begin{aligned}
m_3(y_1,\cdots,y_n) &= \sum y_i/n\\
\frac{d}{dy_1}m_3(y_1,\cdots,y_n) &= 1/n\\
m_3(y_1,\cdots,y_n) &\approx \frac{\delta}{n}
\end{aligned}$$

Interpretation : When y_1 increased as much as δ. If $y_1 > m_1(y_1,\cdots,y_n), \rightarrow \frac{m_1(y_1\cdots y_n)}{y_1} < 1$ which means that $m_1(y_1,\cdots,y_n)$ is less sensitive to the change of y_1. On the other hand, if $y_1 < m_1(y_1,\cdots,y_n)$, it indicates that $m_1(y_1,\cdots,y_n)$ is more sensitive to the change of y_1. That is $m_1(y_1,\cdots,y_n)$ changes more sensitively when y_1 becomes closer to its mean, but $m_1(y_1,\cdots,y_n)$ changes less sensitively when y_1 becomes far from its mean. Thus we can conclude that $m_1(y_1,\cdots,y_n)$ is less sensitive to outlier. It is same for $m_2(y_1,\cdots,y_n)$. But $m_3(y_1,\cdots,y_n)$ changes by constant rate $\frac{\delta}{n}$.

For comparison sensitivity between m_1, m_2, we do not know exact relationship between $\frac{m_1^2}{y_1^2}, \frac{m_2}{y_1}$ in the case that $\frac{m_2}{y_1}\leq\frac{m_1}{y_1}<1$. But we can assume that effect of y_1's changes on m_1 might be much less than m_2 because it squared its effect.

2.

Let $Y_1\cdots Y_n$ be real-valued random variables with $E(Y_i)=\mu, Var[Y_i]=\sigma^2$ for all i = 1,. . . ,n. Compute the variance $\bar{Y}=\sum Y_i/n$ when

(a)

$Cor[Y_i,Y_j]=0\ for\ all\ i,j:$

$$\begin{aligned}
&Var(\sum Y_i) = \sum Var(Y_i)+2\sum\sum Cov(Y_i,Y_j)\\
&When\ Cor(Y_i,Y_j) = Cov(Y_i,Y_J)=0\\
&Var(\sum Y_i) = \sum Var(Y_i) = n\sigma^2\\
&\rightarrow Var(\bar{Y}) = \frac{1}{n^2}\times n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

(b)

$Cor(Y_i, Y_j) = \rho$ for all i,j

$$\begin{aligned}
&Var(\sum Y_i) = \sum Var(Y_i) + 2\sum\sum Cov(Y_i, Y_j)\\
&When\ Cor(Y_i, Y_j) = \rho \rightarrow Cov(Y_i, Y_J) = \rho\sigma^2\\
&Var(\sum Y_i) = \sum Var(Y_i) + 2\sum\sum Cov(Y_i, Y_j) = n\sigma^2 + 2n(n-1)\rho\sigma^2\\
&\rightarrow Var(\bar{Y}) = \frac{1}{n^2} \times Var(\sum y_i) = \frac{\sigma^2}{n} + 2(n-1)\rho\sigma^2/n
\end{aligned}$$

(c)

$$\begin{aligned}
&Var(\sum Y_i) = \sum Var(Y_i) + 2\sum\sum Cov(Y_i, Y_j)\\
&When\ Cor(Y_i, Y_j) = \rho \rightarrow Cov(Y_i, Y_J) = \rho\sigma^2 for\ subdiagonal\\
&Var(\sum Y_i) = \sum Var(Y_i) + 2\sum\sum Cov(Y_i, Y_j) = n\sigma^2 + n(n-1)\rho\sigma^2\\
&\rightarrow Var(\bar{Y}) = \frac{1}{n^2} \times Var(\sum y_i) = \frac{\sigma^2}{n} + 2(n-1)\rho\sigma^2/n^2
\end{aligned}$$

And,

$$\begin{aligned}
as \quad n \rightarrow \infty \quad &(a)Var(\bar{Y}) = 0\\
&(b)Var(\bar{Y}) = 2\rho\sigma^2\\
&(c)Var(\bar{Y}) = 0
\end{aligned}$$

By, Chebyshev's inequality

$$Pr(|\bar{Y} - \mu| > \epsilon) \leq Var(\bar{Y})/\epsilon^2$$

Thus (a) and (c) for any ϵ, $Pr(|\bar{Y} - \mu| > \epsilon)$ as $n \rightarrow \infty$.

That is as n increases, probability that distance between $\bar{Y}, \mu$ is larger than any ϵ which indicate that $\bar{Y}$ is consistent estimate of μ.

On the other hand, in (b), $Pr(|\bar{Y} - \mu| > \epsilon) \leq 2\rho\sigma^2 \rightarrow 1 - Pr(|\bar{Y} - \mu| < \epsilon) \geq 1 - 2\rho\sigma^2$ which indicates that regardless of how large n is, there is probabilty that distance between $\bar{Y}, \mu$ is larger than ϵ. It means that $\bar{Y}$ is not consistent estimate of μ.

3.

Suppose E(Y) = μ, and Var(Y) = σ^2. Consider the estimator $\hat{\mu} = (1-w)\mu_0 + wY$, where $mu_0 \neq 0$ and $w \in (0, 1)$ are numbers.

(a)

Find the expectation, variance, bias, and MSE of $\hat{\mu}$ as function of μ

$$\begin{aligned}
&E(\hat{\mu}) = E[(1-w)\mu_0 + wY] = (1-w)\mu_0 + wE(Y) = (1-w)\mu_0 + w\mu\\
&Var(\hat{\mu}) = Var[(1-w)\mu_0 + wY] = Var(wY) = w^2\sigma^2\\
&B(\hat{\mu}, \mu) = E[(\mu - E(\hat{\mu}))] = \mu - (1-w)\mu_0 - w\mu = (1-w)(\mu - \mu_0)\\
&MSE(\hat{\mu}, \mu) = B(\hat{\mu}, \mu)^2 + Var(\hat{\mu}) = w^2\sigma^2 + (1-w)^2(\mu - \mu_0)^2
\end{aligned}$$

(b)

FOr what values of μ does $\hat{\mu}$ have lower MSE than Y? Interpret your results.

$$E(Y) = \mu \rightarrow B(Y,\mu) = E(\mu - E(Y)) = 0$$
$$MSE(Y,\mu) = Var(Y) = \sigma^2$$

$$\begin{aligned}
&MSE(Y,\mu) > MSE(\hat{\mu},\mu) \\
&\rightarrow \sigma^2 > w^2\sigma^2 + (1-w)^2(\mu-\mu_0)^2 \\
&\rightarrow (1-w^2)\sigma^2 > (1-w)^2(\mu-\mu_0)^2 \\
&\rightarrow \frac{1+w}{1-w}\sigma^2 > (\mu-\mu_0)^2 \\
&\rightarrow |\mu-\mu_0| < \pm\sqrt{\frac{1+w}{1-w}\sigma^2}
\end{aligned}$$

If we consider μ_0 as prior estimator of μ and w as weight that applied in prior estimator and new observation when make weighted average. If we have good prior information about μ which is closer than $\pm\sqrt{\frac{1+w}{1-w}\sigma^2}$, it gives us better estimator for μ than using only data for estimation.

4.

Let $\hat{\theta}$ be an estimator for some unknown quantity θ. Derive a Chebyshev-like bound on $Pr(|\hat{\theta}-\theta| > \epsilon)$ in terms of the MSE of $\hat{\theta}$.

$$\begin{aligned}
&MSE(\hat{\theta},\theta) = E[(\theta-\hat{\theta})^2] \\
&Pr(|\hat{\theta}-\theta| > \epsilon) \le E[(\hat{\theta}-\theta)^2]/\epsilon^2 \quad (by\ Chebyshev\ inequality) \\
&= MSE(\hat{\theta},\theta)/\epsilon^2
\end{aligned}$$

5.

(a)

$$\begin{aligned}
E(\hat{\mu}_n) &= (1-w_n)\mu_0 + w_n\mu \\
E(\hat{\mu}_n^2) &= (1-w_n)^2\mu_0^2 + 2w_n(1-w_n)\mu_0\mu + w_n^2\mu^2
\end{aligned}$$

$$\begin{aligned}
E[(\hat{\mu}_n-\mu)^2] &= E(\hat{\mu}^2) - 2\mu E(\hat{\mu}) + \mu^2 \\
&= (1-w_n)^2\mu_0^2 + 2w_n(1-w_n)\mu_0\mu + w_n^2\mu^2 - 2\mu(1-w_n)\mu_0 - 2w_n\mu^2 + \mu^2 \\
&= \mu^2(w_n^2 - 2w_n + 1) - 2\mu((1-w_n)\mu_0 - w_n(1-w_n)\mu_0) + (1-w_n)^2\mu_0^2 \\
&= \mu^2(1-w_n)^2 - 2\mu\mu_0(1-w_n)^2 + (1-w_n)^2\mu_0^2 \\
&= (1-w_n)^2(\mu - 2\mu\mu_0 + \mu_0^2) \\
&= (1-w_n)^2(\mu-\mu_0)^2
\end{aligned}$$

And,

$$\begin{aligned}
&Pr(|\hat{\mu}_n-\mu| > \epsilon) \le E[(\hat{\mu}-\mu)^2]/\epsilon^2 \quad (by\ Chebyshev\ inequality) \\
&For\ \lim_{n\to\infty} Pr(|\hat{\mu}-\mu| > \epsilon) \rightarrow 0, \quad no\ matter\ \mu's\ value \\
&it\ should\ (1-w_n)^2 \rightarrow 0 \quad n \rightarrow \infty \\
&\rightarrow |1-w_n| \rightarrow 0, |w_n| \rightarrow 1
\end{aligned}$$

(b)

$P(\mu) \sim N(\mu_0, \tau^2), P(\bar{Y_n} \mid \mu) \sim N(\mu, \frac{\sigma^2}{n})$

i.

$$
\begin{aligned}
P(\mu, \bar{Y_n}) &= P(\bar{Y_n}) \mid \mu)P(\mu) \\
&= \sqrt{\frac{n}{2\pi\sigma^2}} exp\{-\frac{n}{2}(\bar{Y_n} - \mu)^2\} \times \frac{1}{\sqrt{2\pi\tau^2}} exp\{-\frac{1}{2\tau^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2)\} \\
&\propto \frac{1}{2\pi}\sqrt{\frac{n}{\sigma^2\tau^2}} exp\{-\frac{1}{2}(\mu^2(\frac{n}{\sigma^2} + \frac{1}{\tau^2}) - 2\mu(\frac{n\bar{Y}_n}{\sigma^2} + \frac{\mu_0}{\tau^2}) + (\frac{n\bar{Y}_n^2}{\sigma^2} + \frac{\mu_0}{\tau^2}))\} \\
&\propto \frac{1}{2\pi}\sqrt{\frac{n}{\sigma^2\tau^2}} \underbrace{exp\{-\frac{1}{2}(\frac{n}{\sigma^2} + \frac{1}{\tau^2})(\mu - (\frac{n}{\sigma^2} + \frac{1}{\tau^2})^{-1})(\frac{n\bar{Y}_n}{\sigma^2} + \frac{\mu_0}{\tau^2}))^2\}}_{kernel\ of\ \mu|\bar{Y}_n \sim N(\mu_n, \tau_n^2)} \times \underbrace{exp\{-\frac{1}{2}(\frac{n}{\sigma^2}\bar{Y}_n^2 + \frac{\mu_0^2}{\tau^2} - (\frac{n}{\sigma^2} + \frac{1}{\tau^2})^{-1}(\frac{nY_n}{\sigma^2} + \frac{\mu_0}{\tau^2}}_{kernel\ of\ \bar{Y}_n}
\end{aligned}
$$

$$
where \quad \tau_n^2 = (\frac{n}{\sigma^2} + \frac{1}{\tau^2})^{-1}, \quad \mu_n = \tau_n^2(\frac{n\bar{Y}}{\sigma^2} + \frac{\mu_0}{\tau^2})
$$

ii.

$$
E(\mu \mid \bar{Y}_n) = (\frac{n}{\sigma^2} + \frac{1}{\tau^2})^{-1}(\frac{n\bar{Y}}{\sigma^2} + \frac{\mu_0}{\tau^2}) = w_n\bar{Y}_n + (1 - w_n)\mu_0 \quad where\ w_n = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}
$$

We have check that $Pr(|\hat{\mu}_n - \mu| > \epsilon) \to 0 \quad as\ n \to \infty \quad if \quad \lim_{n\to\infty} |w_n| \to 1$. In addition, we can confirm that $\lim_{n\to\infty} w_n \to 1$. Thus we can conclude that posterior mean of $\mu \mid \hat{Y}_n$ is consistent estimator for μ