ST210A - Homework 6

Hoang Duong

October 12, 2014

Problem 1. $X_1, ..., X_n \stackrel{iid}{\sim} \mathcal{N}(\theta, \theta)$.

Proof. We have:

$$\begin{aligned}
\mathrm{Var}(\bar{X} - \theta) =& \frac{1}{n}\theta \\
\mathrm{Var}(s^2 - \theta) =& \mathrm{Var}(s^2) = \mathbb{E}s^4 - \mathbb{E}^2\left[s^2\right] \\
\mathbb{E}\left[s^2\right] =& \theta \\
\mathbb{E}\left[s^4\right] =& \mathbb{E}\left[\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\right)\right]^2 \\
=& \frac{1}{(n-1)^2}\mathbb{E}\left[\sum_{i=1}^{n}\left((X_i - \theta)^2 + \left(\theta - \bar{X}\right)^2 + 2\left(X_i - \theta\right)\left(\theta - \bar{X}\right)\right)\right]^2 \\
=& \frac{1}{(n-1)^2}\mathbb{E}\left[\sum_{i=1}^{n}\left(X_i - \theta\right)^2 - n\left(\theta - \bar{X}\right)^2\right]^2 := A
\end{aligned}$$

Now let $Y_i = X_i - \theta$, then $Y_i \stackrel{iid}{\sim} \mathcal{N}(0, \theta)$. Thus $\mathbb{E}Y_i^4 = 3\theta^2$. So:

$$
\begin{aligned}
A =& \frac{1}{(n-1)^2}\mathbb{E}\left[\sum_{i=1}^n Y_i^2 - n\bar{Y}^2\right]^2 \\
=& \frac{1}{(n-1)^2}\mathbb{E}\left[\left(\sum_{i=1}^n Y_i^2\right)^2 - 2n\bar{Y}^2\sum_{i=1}^n Y_i^2 + n^2\bar{Y}^4\right] \\
\mathbb{E}\left(\sum_{i=1}^n Y_i^2\right)^2 =& n\mathbb{E}Y_1^4 + n(n-1)\mathbb{E}Y_1^2\mathbb{E}Y_2^2 \\
=& 3n\theta^2 + n(n-1)\theta^2 = (n^2+2n)\theta^2 \\
-2n\mathbb{E}\left[\bar{Y}^2\sum Y_i^2\right] =& \frac{-2}{n}\mathbb{E}\left[\left(\sum_{i=1}^n Y_i\right)^2\sum_{i=1}^n Y_i^2\right] \\
=& -\frac{2}{n}(n^2+2n)\theta^2 \\
=& -2(n+2)\theta^2 \\
n^2\mathbb{E}\bar{Y}^4 =& \frac{1}{n^2}\mathbb{E}\left[\sum Y_i\right]^4 \\
=& \frac{1}{n^2}3n^2\theta^2 \\
=& 3\theta^2 \\
\Rightarrow A =& \frac{1}{(n-1)^2}\theta^2\left(n^2+2n-2n-4+3\right) \\
=& \frac{n+1}{n-1}\theta^2 \\
\mathrm{Var}\left[s^2\right] =& \mathbb{E}\left[s^4\right] - \mathbb{E}^2\left[s^2\right] \\
=& \frac{n+1}{n-1}\theta^2 - \theta^2 = \frac{2}{n-1}\theta^2
\end{aligned}
$$

So the relative efficiency of S^2 w.r.t. $\bar{X}$ is:

$$
\frac{\mathrm{Var}\bar{X}}{\mathrm{Var}S^2} = \frac{\theta}{n}\frac{n-1}{2\theta^2} = \frac{n-1}{2n\theta}
$$

Roughly speaking, with $\theta > 1/2$, the sample mean is more effective than the sample variance. With $\theta < 1/2$, the sample variance is more effective. □

Problem 2. Exponential Distribution

Proof. 1. Consider the CDF of $X_{(2)}$,

$$
\begin{aligned}
\mathbb{P}\left[X_{(2)} \le x/n^p\right] =& 1 - \mathbb{P}\left[X_2 > x/n^p\right] \\
\mathbb{P}\left[X_{(2)} > x/n^p\right] =& \mathbb{P}\left[X_{(1)} > x/n^p\right] + \mathbb{P}\left[X_{(1)} \le x/n^p < X_{(2)}\right] \\
=& \mathbb{P}\left[X_i > x/n^p, \forall i\right] + n\mathbb{P}\left[X_1 \le x/n^p\right]\mathbb{P}\left[X_i > x/n^p, \forall i \ge 2\right] \\
=& \exp\left(-nx/n^p\right) + n\left(1-\exp\left(-x/n^p\right)\right)\exp\left\{-(n-1)x/n^p\right\} \\
=& \frac{1+n\exp\left(x/n^p\right)-n}{\exp\left(nx/n^p\right)} := \frac{A_n}{B_n}
\end{aligned}
$$

We note that $\mathbb{P}\left[n^pX_{(2)} \le x\right]$ converges to a value between 0 and 1 iff $\mathbb{P}\left[n^pX_{(2)} > x\right]$ converges to a value between 0 and 1.

Now if $p > 1$, using the fact that $\lim_{x\to 0} \frac{e^x - 1}{x} = 1$, we have:

$$\begin{aligned}
\lim A_n =& 1 + \lim n(\exp(x/n^p) - 1) \\
=& 1 + \lim_{n\to\infty} \frac{\exp(x/n^p) - 1}{\frac{x}{n^p}} \frac{x}{n^{p-1}} \\
=& 1 + 1 \times 0 = 1 \\
\lim B_n =& 1 \\
\Rightarrow \lim \frac{A_n}{B_n} =& 1
\end{aligned}$$

If $0 < p < 1$, then:

$$\begin{aligned}
\lim A_n =& 1 + \lim_{n\to\infty} \frac{\exp(x/n^p) - 1}{\frac{x}{n^p}} \frac{x}{n^{p-1}} = \infty \\
\lim B_n =& \infty \\
\lim \frac{A_n'}{B_n'} =& \lim \frac{\exp(x/n^p) - nxpn^{-p-1}\exp(x/n^p) - 1}{(1-p)n^{-p} x \exp(n^{1-p}x))} \\
\lim \frac{\partial A_n}{\partial n} =& \lim \left(\exp(x/n^p) - npxn^{-p-1}\exp(x/n^p) - 1\right) \\
=& \lim \left(\exp(x/n^p) - px\frac{\exp(xn^{-p})}{n^p} - 1\right) \\
=& 1 - 0 - 1 = 0 \\
\lim \frac{\partial B_n}{\partial n} =& \lim \frac{(1-p)x\exp\left(n^{1-p}x\right)}{n^p} = \infty. \text{ (exp} \to \infty \text{ faster than polynomial)} \\
\Rightarrow \lim \frac{A_n}{B_n} =& 0 \text{ (L'Hospital's Rule)}
\end{aligned}$$

Finally for $p = 1$, then:

$$\begin{aligned}
\lim \frac{A_n}{B_n} =& \lim \frac{1 + n\exp(x/n) - n}{\exp(x)} \\
=& \lim \exp(-x)\left(1 + \frac{\exp(x/n) - 1}{\frac{1}{n}}\right) \\
=& \exp(-x)(1 + x)
\end{aligned}$$

So $p = 1$. With this choice of p, $\mathbb{P}\left[X_{(2)}n \le x\right] = 1 - \frac{1+x}{\exp(x)}$

2. We have:

$$\begin{aligned}
\mathbb{P}\left[X_{(n)} - \log n \le x\right] =& \mathbb{P}\left[X_{(n)} \le x + \log n\right] \\
=& \mathbb{P}\left[X_i \le x + \log n, \forall i\right] \\
=& \mathbb{P}^n\left[X_1 \le x + \log n\right] \\
=& \left(1 - \exp(-x - \log n)\right)^n \\
=& \left(1 - \exp(-x)/n\right)^n
\end{aligned}$$

We use the limit equality that: $\lim\left(1 + \frac{a}{n}\right)^n = \exp(a)$, then:

$$\lim \mathbb{P}\left[X_{(n)} - \log n \le x\right] = \lim\left(1 + \frac{-\exp(-x)}{n}\right)^n = \exp\left(-\exp(-x)\right)$$

So the limiting CDF for $X_{(n)} - \log n$ is $\exp\left(-\exp(-x)\right)$. $\square$

Lemma 1. *Leibniz's Rule*

Let $\phi(\alpha) = \int_{u_1(\alpha)}^{u_2(\alpha)} f(x,\alpha)dx, \alpha \in [a,b]$ *then:*

$$\frac{d\phi}{d\alpha} = \int_{u_1}^{u_2} \frac{\partial f}{\partial \alpha} dx + f(u_2,\alpha)\frac{du_2}{d\alpha} - f(u_1,\alpha)\frac{du_1}{d\alpha}$$

for $\alpha \in [a,b]$.

Lemma 2. *Let* X *be a random variable with CDF* F. *Let* $k_1, k_2 \in \mathbb{R}^+$. *Define:*

$$L(X,a) = \begin{cases} k_1 \left|X-a\right|, & a \le X \\ k_2 \left|X-a\right|, & a > X \end{cases}$$

Then a *that minimizes* $\mathbb{E}L(X,a)$ *is the* $F^{-1}(\frac{k_2}{k_1+k_2})$.

Proof. We have:

$$\begin{aligned}
\mathbb{E}L(X,a) &= \int_{-\infty}^{a} k_1(a-x)dF + \int_{a}^{\infty} k_2(x-a)dF \\
\frac{d\mathbb{E}L(X,a)}{da} &= \int_{-\infty}^{a} k_1 dF + k_1(a-a) + \int_{a}^{\infty} -k_2 dF - k_2(a-a) \\
&= \int_{-\infty}^{a} k_1 dF - \int_{a}^{\infty} k_2 dF \\
\frac{d\mathbb{E}L(X,a)}{da} &= 0 \\
\Leftrightarrow \int_{-\infty}^{a} k_1 dF &= \int_{a}^{\infty} k_2 dF \\
\Leftrightarrow k_1 F(a) &= k_2(1-F(a)) \\
\Leftrightarrow F(a) &= \frac{k_2}{k_1+k_2} \\
\Leftrightarrow a &= F^{-1}\left(\frac{k_2}{k_1+k_2}\right)
\end{aligned}$$

To verify that we have the minimum at the point where first derivative is equal to zero, we need to check the second derivative:

$$\frac{d^2\mathbb{E}L(X,a)}{da^2} = k_1 + k_2 > 0.$$

So $\mathbb{E}L(X,a)$ is minimized when $a = F^{-1}\left(\frac{k_2}{k_1+k_2}\right)$. □

Problem 3. Loss Function that is minimized at quantile

Proof. Applying Lemma 2 we just proved, and Theorem 7.1 about Bayes estimator in Keener's book, we have the Bayes estimator is the $\frac{k_2}{k_1+k_2}$ quantile of the posterior distribution. □

Problem 4. Normal Distribution

Proof. For i.i.d random variable, the aggregate log likelihood function is just sum of individual likelihood function. Thus:

$$-l(\theta) = -\log \mathcal{L}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{r}\left(\log \sigma + \frac{1}{2\sigma^2}\left(X_{ij} - \mu_i\right)^2\right) + C$$

Looking at μ_i, the negative log-likelihood function is minimized when $\mu_i = \bar{X}_i$. So now we need to find the σ that minimize:

$$-l(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{r}\left(\log \sigma + \frac{1}{2\sigma^2}\left(X_{ij} - \bar{X}_i\right)^2\right)$$

Looking at the derivative:

$$\begin{aligned}
\frac{\partial - l(\theta)}{\partial \sigma} =& \frac{nr}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 \\
\frac{\partial - l(\theta)}{\partial \sigma} =& 0 \\
\Leftrightarrow \sigma^2 =& \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 := \sigma^2_{MLE}
\end{aligned}$$

To make sure that σ^2_{MLE} is the minimizer, we check the second derivative, which is:

$$\begin{aligned}
\frac{\partial^2 - l(\theta)}{\partial \sigma^2} =& -\frac{nr}{\sigma^2} + \frac{3}{\sigma^4}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 \geq 0 \\
\Leftrightarrow \sigma^2 \leq& \frac{3}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 \ \ (*)
\end{aligned}$$

The last statement, (*), is not always true. However since $\mathbb{E}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 = n(r-1)\sigma^2$, we can safely say that it is true with high probability. Assuming this then the MLE estimator for σ^2 is $\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2$. Now:

$$\mathbb{E}\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\left(X_{ij} - \bar{X}_i\right)^2 = \frac{n(r-1)}{nr}\sigma^2 = \frac{r-1}{r}\sigma^2$$

So $\lim_{n\to\infty}\sigma^2_{MLE} = \frac{r-1}{r}\sigma^2 \neq \sigma^2$. Thus the MLE estimator for σ^2 is not consistent. □

Problem 5. Mixture Distribution

Proof. (a) Again we observe that log likelihood function of i.i.d random variables is the sum of individual log likelihood function. Thus:

$$\begin{aligned}
l(\theta) =& \sum_{i=1}^{n}\log\left(\theta f_1(x_i) + (1-\theta) f_2(x_i)\right) \\
\Rightarrow \frac{\partial l(\theta)}{\partial \theta} =& \sum_{i=1}^{n}\frac{f_1(x_i) - f_2(x_i)}{\theta f_1(x_i) + (1-\theta) f_2(x_i)}
\end{aligned}$$

Now we note that for any linear function $g(\theta)$ of θ. The second derivative:

$$
\begin{aligned}
\frac{\partial \log g(\theta)}{\partial \theta} &= \frac{g'(\theta)}{g(\theta)} \\
\Rightarrow \frac{\partial^2 \log g(\theta)}{\partial \theta^2} &= \frac{g''(\theta)g(\theta) - (g'(\theta))^2}{g^2(\theta)} = -\frac{(g'(\theta))^2}{g^2(\theta)} \leq 0
\end{aligned}
$$

In our case the sum of n such second derivative is also non-positive. It is zero iff $f_1(x_i) = f_2(x_i), \forall i$. We will make the weak assumption that this does not happen (otherwise the the solution for MLE might not be unique). With this assumption, the first derivative of our log likelihood is a decreasing function on $(0, 1)$. Thus it has a unique root iff $\lim_{\theta \to 0^+} \frac{\partial l(\theta)}{\partial \theta} > 0$ and $\lim_{\theta \to 1^-} \frac{\partial l(\theta)}{\partial \theta} < 0$. The first condition is equivalent to:

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{f_1(x_i) - f_2(x_i)}{f_2(x_i)} &> 0 \\
\Leftrightarrow \sum_{i=1}^{n} \frac{f_1(x_i)}{f_2(x_i)} &> n
\end{aligned}
$$

While the second condition is equivalent to:

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{f_1(x_i) - f_2(x_i)}{f_1(x_i)} &< 0 \\
\Leftrightarrow \sum_{i=1}^{n} \frac{f_2(x_i)}{f_1(x_i)} &> n
\end{aligned}
$$

Since we've already shown that the second derivative is negative, the solution to the first derivative equal to zero is the maximizer of the function.

(b) When the score equation has no solution, there are two cases.

First case, the first derivative of the log likelihood is always bigger than zero, then the log likelihood is maximized as $\theta \to 1$. The MLE for θ is 1.

Second case, the first derivative of the log likelihood is always smaller than zero, then the log likelihood is maximized as $\theta \to 0$. The MLE for θ is 0. $\square$