STAT111: Mathematical Statistics II **Spring 2024**

Homework 5

David Yang

1. **Suppose $X_1, \dots, X_n$ are i.i.d. Normal with mean μ and variance σ^2.**

   a) **Show the MLE for σ is the square root of the MLE for σ^2 (in general, if $\hat{\theta}$ is the MLE of θ, then $\hat{\theta}$ is the MLE of θ, then $\hat{\varphi} = g(\hat{\theta})$ is the MLE for $\varphi = g(\theta)$).**

   Solution. If $X_1, \dots, X_n$ are i.i.d. Normal with mean μ and variance σ^2,

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}. \end{aligned}$$

The log-likelihood is

$$\begin{aligned} l(\mu, \sigma^2) &= n \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\sum(x_i-\mu)^2}{2\sigma^2} \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{\sum(x_i-\mu)^2}{2\sigma^2}. \end{aligned}$$

Setting the partials to 0 to solve for the MLEs, we get that

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= 2\frac{\sum(x_i-\mu)}{2\sigma^2} \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi - \frac{1}{2} \cdot \left(\sum(x_i-\mu)^2\right) \cdot \left(-\frac{1}{\sigma^4}\right) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 \end{aligned}$$

The first equation is equivalent to $\sum(x_i - \mu) = 0$, so solving for μ gives

$$\hat{\mu} = \frac{\sum x_i}{n} = \overline{x}.$$

The second equation is equivalent to $\frac{n}{2\sigma^2} = \frac{1}{2\sigma^4}\sum(x_i-\mu)^2$, and solving for σ^2 gives

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{n}.$$

Plugging in the MLE for μ, we get the MLE for $\hat{\sigma}^2$:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum(x_i-\mu)^2}{n} \\ &= \frac{1}{n}\left(\sum(x_i-\overline{x})^2\right). \end{aligned}$$

Similarly, if we find the MLE for σ, we get that

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{2}\frac{1}{2\pi\sigma^2} \cdot 4\pi\sigma - \frac{\sum(x_i - \mu)^2}{4\sigma^3}.$$

Equating this to 0, plugging in $\hat{\mu} = \overline{x}$, and solving for σ, we find that

$$\sigma^2 = \frac{1}{n}\left(x_i - \overline{x}\right)^2,$$

so $\hat{\sigma} = \sqrt{\frac{1}{n}\left(x_i - \overline{x}\right)^2} = \sqrt{\hat{\sigma}^2}$. Thus, the MLE for σ is the square root of the MLE for σ^2, as desired. ■

b) **Show that $\hat{\sigma}^2 = \sum \frac{(X_i - \overline{X})^2}{n}$ has smaller mean square error than the unbiased estimate $s^2 = \sum \frac{(X_i - \overline{X})^2}{n-1}$.**

Solution. Note that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$, so

$$\mathbb{E}\left[\frac{(n-1)s^2}{\sigma^2}\right] = n - 1 \text{ and } \mathrm{Var}\left[\frac{(n-1)s^2}{\sigma^2}\right] = 2(n-1).$$

It follows that

$$\mathbb{E}\left[s^2\right] = \sigma^2 \text{ and } \mathrm{Var}\left[s^2\right] = \frac{2\sigma^4}{n-1}.$$

Furthermore, we know that $\hat{\sigma}^2 = \frac{n-1}{n}s^2$. Consequently,

$$\mathbb{E}\left[\hat{\sigma}^2\right] = \frac{n-1}{n}\mathbb{E}\left[s^2\right] = \frac{n-1}{n}\sigma^2$$

and

$$\mathrm{Var}\left[\hat{\sigma}^2\right] = \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\left[s^2\right] = \frac{2\sigma^4(n-1)}{n^2}.$$

Since s^2 is an unbiased estimate, its mean square error is equal to its variance:

$$MSE(s^2) = \mathrm{Var}\left[s^2\right] = \frac{2\sigma^4}{n-1}.$$

On the other hand,

$$\begin{aligned} MSE(\hat{\sigma}^2) &= \mathrm{Var}\left[\hat{\sigma}^2\right] + (\mathbb{E}\left[\hat{\sigma}^2\right] - \sigma^2)^2 \\ &= \left(\frac{2\sigma^4(n-1)}{n^2}\right) + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 \\ &= \frac{2\sigma^4(n-1)}{n^2} + \frac{\sigma^4}{n^2} \\ &= \frac{(2n-1)\sigma^4}{n^2}. \end{aligned}$$

Note that by direct comparison, since $\frac{2n-1}{n^2} < \frac{2}{n-1}$, it follows that $MSE(\hat{\sigma}^2) < MSE(s^2)$, as desired. ■

c) **Find an expression for the expected value of the sample standard deviation $s = \sqrt{s^2}$. Use this to construct an unbiased estimate for σ. You could check your answer by generating sample variances from the Gamma distribution implied by a given n and σ.**

Solution. We know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. Suppose that $X \sim \chi^2_{n-1}$. Then

$$\mathbb{E}\left[\sqrt{\frac{(n-1)s^2}{\sigma^2}}\right] = \mathbb{E}\left[\sqrt{x}\right].$$

It follows from LOTUS and integral by recognition that

$$\begin{aligned}
\mathbb{E}\left[\sqrt{x}\right] &= \int_0^\infty \sqrt{x} f_X(x)\, dx \\
&= \int_0^\infty \sqrt{x} \frac{x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)}\, dx \\
&= \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}\, dx \\
&= \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}
\end{aligned}$$

where the final step follows by recognizing the integrand as the pdf for a χ^2 variable. It follows that

$$\mathbb{E}\left[\sqrt{\frac{(n-1)s^2}{\sigma^2}}\right] = \mathbb{E}\left[\frac{s}{\sigma}\sqrt{n-1}\right] = \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Using linearity of expectation, we find that

$$\mathbb{E}[s] = \frac{\sigma}{\sqrt{n-1}} \frac{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Thus, an unbiased estimate for σ is simply

$$\boxed{\hat{\sigma} = \frac{\sqrt{n-1}}{\sigma\sqrt{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}}.$$

■

2. **Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson(θ).**

a) **Let $Y = \sum X_i$. If $n = 3$, find $P(X_1 = 2, X_2 = 3, X_3 = 0 \mid Y = 5, \theta = 2)$. How does this change if $\theta = 3$?**

Solution. Since the sum of Poissons is a Poisson, we know that $Y \sim \text{Poisson}(3\theta)$. For $\theta = 2$, we calculate

$$\begin{aligned}
P(X_1 = 2, X_2 = 3, X_3 = 0 \mid Y = 5) &= \frac{P(X_1 = 2, X_2 = 3, X_3 = 0 \cap Y = 5)}{P(Y = 5)} \\
&= \frac{P(X_1 = 2)P(X_2 = 3)P(X_3 = 0)}{P(Y = 5)}
\end{aligned}$$

where the second line follows from the fact that each X_i are i.i.d.

Plugging in the respective probabilities, we get

$$\begin{aligned} P(X_1 = 2, X_2 = 3, X_3 = 0 \mid Y = 5) &= \frac{P(X_1 = 2)P(X_2 = 3)P(X_3 = 0)}{P(Y = 5)} \\ &= \frac{\frac{e^{-2}2^2}{2!}\frac{e^{-2}2^3}{3!}\frac{e^{-2}2^0}{0!}}{\frac{e^{-6}6^5}{5!}} \\ &= \boxed{\frac{10}{243}}. \end{aligned}$$

For $\theta = 3$, we use the same logic to get that

$$\begin{aligned} P(X_1 = 2, X_2 = 3, X_3 = 0 \mid Y = 5) &= \frac{P(X_1 = 2)P(X_2 = 3)P(X_3 = 0)}{P(Y = 5)} \\ &= \frac{\frac{e^{-3}3^2}{2!}\frac{e^{-3}3^3}{3!}\frac{e^{-3}3^0}{0!}}{\frac{e^{-9}9^5}{5!}} \\ &= \boxed{\frac{10}{243}} \end{aligned}$$

and so the probability does not change depending on the value of θ. ■

b) **Show that $\overline{X}$ is the MLE for θ, and that the reciprocal Fisher information gives the exact variance of $\hat{\theta}$. Verify that Poisson(θ) is a 1-parameter exponential family, so you know you can use the second derivative formula for the Fisher information.**

Solution. Note that

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}$$

so the log-likelihood is

$$l(\theta) = -n\theta + \sum x_i \log\theta - \sum \log(x_i!).$$

Setting $\frac{\partial l}{\partial \theta}$ to 0, we get that

$$\frac{\partial l}{\partial \theta} = -n + \frac{\sum x_i}{\theta} = 0$$

so $\boxed{\hat{\theta} = \frac{\sum x_i}{n} = \overline{x}}$ as desired.

We will now verify that Poisson(θ) is a 1-parameter exponential family, so you know you can use the second derivative formula for the Fisher information. Note that if $X \sim$ Poisson(θ), then

$$\begin{aligned} f_X(x) &= \frac{\theta^x e^{-\theta}}{x!} \\ &= \exp\left(\log\left(\frac{\theta^x e^{-\theta}}{x!}\right)\right) \\ &= \exp((x\log\theta) - \log(x!) - \theta) \end{aligned}$$

which is precisely the form of the exponential family; $T(x) = x$, $\eta(\theta) = \log(\theta)$, $A(\theta) = \theta$, and $B(x) = \log(x!)$.

Thus, Poisson(θ) is a 1-parameter exponential family.

To find the Fisher information for $X_1, \ldots, X_n$, we will use the second derivative formula. We found that

$$\frac{\partial l}{\partial \theta} = -n + \frac{\sum x_i}{\theta}$$

so

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{\sum x_i}{\theta^2}.$$

It follows that

$$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2 l}{\partial \theta^2}\right] = -\mathbb{E}\left[-\frac{\sum x_i}{\theta^2}\right] = \frac{\mathbb{E}\left[\sum x_i\right]}{\theta^2} = \frac{n}{\theta}.$$

However, we also know that

$$\text{Var}\left[\hat{\theta}\right] = \text{Var}\left[\overline{X}\right] = \frac{n\text{Var}\left[X_i\right]}{n^2} = \frac{\theta}{n}.$$

Thus, we conclude that $\frac{1}{I_n(\theta)} = \text{Var}\left[\hat{\theta}\right]$ so the reciprocal Fisher information gives the exact variance of $\hat{\theta}$, as desired. ■

c) **If $n = 10$ and $\theta = 10$, check the Normal approximation to the distribution of $\overline{X}$. Use the continuity correction to approximate $P(\overline{X} = 10)$ and compare this to the exact probability based on the associated Poisson distribution.**

Solution. Note that $\mathbb{E}\left[\overline{X}\right] = \mathbb{E}\left[X_i\right] = 10$. On the other hand, $\text{Var}\left[\overline{X}\right] = \frac{1}{n}\text{Var}\left[X_i\right] = \frac{10}{10} = 1$. It follows that the normal approximation for $\overline{X}$ is $N(10, 1)$.

Using the continuinity correction, we can approximate $P(\overline{X} = 10)$ as $P(9.95 < \overline{X} < 10.05)$, so

$$\begin{aligned} P(\overline{X} = 10) &= P(9.95 < \overline{X} < 10.05) \\ &= pnorm(10.05, 10, 1) - pnorm(10.05, 10, 1) \approx 0.0399. \end{aligned}$$

On the other hand, note that $P(\overline{X} = 10) = P(Y = 100)$, where $Y \sim \text{Poisson}(10 \cdot 10)$, so

$$P(Y = 100) = \frac{100^{100}e^{-100}}{100!} \approx 0.04.$$

We see that the exact probability based on the associated Poisson distribution and the Normal approximation are very close. ■

d) **Show that $\theta \sim \text{Gamma}\left(\alpha, \frac{\alpha}{u}\right)$ defines a conjugate family of prior distributions for θ, and that the posterior mean for $\theta \mid \overline{x}$ is a weighted average of $\overline{x}$ and the prior mean μ, with more weight on $\overline{x}$ as n increases.**

Solution. Note that $\overline{X}$ is a sufficient statistic for $X_1, \ldots, X_n$ which are i.i.d Poisson. Consequently, conditioning on the sufficient statistic $\overline{X}$ is equivalent to conditioning on $X_1, \ldots, X_n$. We know that

$$f_{\theta|X_1,\ldots,X_n} = f_\theta L(X_1, \ldots, X_n \mid \theta)$$

where

$$f_\theta = \frac{\left(\frac{\alpha}{\mu}\right)^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{\frac{-\alpha}{\mu}\theta}$$

and

$$L(X_1, \ldots, X_n \mid \theta) = \prod \frac{\theta^{x_i}}{x_i!} e^{-\theta} = e^{-n\theta} \frac{\theta^{\sum x_i}}{\prod x_i!}.$$

It follows that

$$\begin{aligned} f_{\theta|X_1,\ldots,X_n} &= f_\theta L(X_1, \ldots, X_n \mid \theta) \\ &= \left(\frac{\left(\frac{\alpha}{\mu}\right)^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{\frac{-\alpha}{\mu}\theta} \right) \left(e^{-n\theta} \frac{\theta^{\sum x_i}}{\prod x_i!} \right) \\ &\propto \theta^{\alpha-1+\sum x_i} e^{-\theta\left(\frac{\alpha}{\mu}+n\right)} \end{aligned}$$

which is the kernel of a Gamma $\left(\alpha + \sum x_i, \frac{\alpha}{\mu} + n\right)$ random variable.

Thus, we know that $f_{\theta|X_1,\ldots,X_n} = f_{\theta|\overline{X}} \sim$ Gamma $\left(\alpha + \sum x_i, \frac{\alpha}{\mu} + n\right)$. Since the prior and posterior distributions are in the same family (Gamma) of distributions, we know that θ defines a conjugate family of prior distributions for θ, as desired.

Finally, note that

$$\mathbb{E}\left[\theta \mid \overline{X}\right] = \frac{\alpha + \sum x_i}{\frac{\alpha}{\mu} + n}.$$

Manipulating this expression, we find that

$$\begin{aligned} \mathbb{E}\left[\theta \mid \overline{X}\right] &= \frac{\alpha + \sum x_i}{\frac{\alpha}{\mu} + n} \\ &= \frac{\alpha + n\overline{x}}{\frac{\alpha}{\mu} + n} \\ &= \frac{\alpha}{\frac{\alpha}{\mu} + n} + \overline{x}\left(\frac{n}{\frac{\alpha}{\mu} + n}\right) \\ &= \mu\left(\frac{\alpha}{\alpha + n\mu}\right) + \overline{x}\left(\frac{n\mu}{\alpha + n\mu}\right), \end{aligned}$$

so the posterior mean for $\theta \mid \overline{x}$ is a weighted average of $\overline{x}$ and the prior mean μ. Finally, we see that if n increases, the weight on the second term (corresponding to $\overline{x}$) increases, as desired. ■