Quiz Section 9 – Solutions

Review

1) Maximum Likelihood Estimator (MLE): We denote the MLE of θ as $\hat{\theta}_{\text{MLE}}$ or simply $\hat{\theta}$, the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \mathcal{L}(x_1, \ldots, x_n \mid \theta) = \arg\max_{\theta} \ln \mathcal{L}(x_1, \ldots, x_n \mid \theta)$$

2) An estimator $\hat{\theta}$ for a parameter θ of a probability distribution is **unbiased** iff $\mathbb{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta$

Task 1 – Mystery Dish!

A fancy new restaurant has opened up which features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability θ, dish C with probability 2θ, and dish D with probability $0.5 - 3\theta$. Each diner is served a dish independently. Let x_A be the number of people who received dish A, x_B the number of people who received dish B, etc, where $x_A + x_B + x_C + x_D = n$. Find the MLE for θ, $\hat{\theta}$.

The data tells us, for each diner in the restaurant, what their dish was. We begin by computing the likelihood of seeing the given data given our parameter θ. Because each diner is assigned a dish independently, the likelihood is equal to the product over diners of the chance they got the particular dish they got, which gives us:

$$\mathcal{L}(x \mid \theta) = 0.5^{x_A} \theta^{x_B} (2\theta)^{x_C} (0.5 - 3\theta)^{x_D}$$

From there, we just use the MLE process to get the log-likelihood, take the first derivative, set it equal to 0, and solve for $\hat{\theta}$.

$$\ln \mathcal{L}(x \mid \theta) = x_A \ln(0.5) + x_B \ln(\theta) + x_C \ln(2\theta) + x_D \ln(0.5 - 3\theta)$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(x \mid \theta) = \frac{x_B}{\theta} + \frac{x_C}{\theta} - \frac{3x_D}{0.5 - 3\theta}$$

$$\frac{x_B}{\hat{\theta}} + \frac{x_C}{\hat{\theta}} - \frac{3x_D}{0.5 - 3\hat{\theta}} = 0$$

Solving yields $\hat{\theta} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$.

Task 2 – A Red Poisson

Suppose that $x_1, \ldots, x_n$ are i.i.d. samples from a Poisson(θ) random variable, where θ is unknown. In other words, they follow the distributions $\mathbb{P}(k; \theta) = \theta^k e^{-\theta}/k!$, where $k \in \mathbb{N}$ and $\theta > 0$ is a positive real number.

Find the MLE of θ.

We follow the recipe given in class:

$$\begin{aligned}
\mathcal{L}(x_1, \ldots, x_n \mid \theta) &= \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!} \\
\ln \mathcal{L}(x_1, \ldots, x_n \mid \theta) &= \sum_{i=1}^{n} [-\theta - \ln(x_i!) + x_i \ln(\theta)] \\
\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(x_1, \ldots, x_n \mid \theta) &= \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\theta}\right] \\
-n + \frac{\sum_{i=1}^{n} x_i}{\hat{\theta}} &= 0 \\
\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}
\end{aligned}$$

Task 3 – A biased estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2\right)$$

where $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

Let $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\mathbb{E}\left[\hat{\theta}_2\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i^2 - 2X_i\overline{X} + \overline{X}^2)\right]$$

which by linearity of expectation (and distributing the sum) is

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\frac{2}{n} \overline{X} \sum_{i=1}^{n} X_i\right] + \mathbb{E}\left[\overline{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - 2\mathbb{E}\left[\overline{X}^2\right] + \mathbb{E}\left[\overline{X}^2\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\overline{X}^2\right] . \qquad (**)
\end{aligned}$$

We know that for any random variable Y, since $\mathrm{Var}(Y) = \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2$ it holds that

$$\mathbb{E}\left[Y^2\right] = \mathrm{Var}(Y) + (\mathbb{E}[Y])^2.$$

Also, we have $\mathbb{E}[X_i] = \mu,\ \mathrm{Var}(X_i) = \sigma^2\ \forall i$ and $\mathbb{E}\left[\overline{X}\right] = \mu,\ \mathrm{Var}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. Combining these facts, we get

$$\mathbb{E}\left[X_i^2\right] = \sigma^2 + \mu^2 \quad \forall i \qquad \text{and} \qquad \mathbb{E}\left[\overline{X}^2\right] = \frac{\sigma^2}{n} + \mu^2.$$

Substituting these equations into (**) we get

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[\overline{X}^2\right] = \sigma^2+\mu^2-\left(\frac{\sigma^2}{n}+\mu^2\right)$$
$$= \left(1-\frac{1}{n}\right)\sigma^2.$$

Thus $\hat{\theta}_2$ is not unbiased.

Task 4 – Weather Forecast

A weather forecaster predicts sun with probability θ_1, clouds with probability $\theta_2 - \theta_1$, rain with probability $\frac{1}{2}$ and snow with probability $\frac{1}{2} - \theta_2$. This year, there have been 55 sunny days, 100 cloudy days, 160 rainy days and 50 snowy days. What is the maximum likelihood estimator for θ_1 and θ_2?

We want to find the likelihood of the data samples given the parameter θ. To do this, we take the following product over all the data points.

$$\mathcal{L}(x_1,...,x_{365} \mid \theta_1,\theta_2) = \theta_1^{55}(\theta_2-\theta_1)^{100}\left(\frac{1}{2}\right)^{160}\left(\frac{1}{2}-\theta_2\right)^{50}$$

Then, we use this to determine the log likelihood.

$$\ln\mathcal{L}(x_1,...,x_{365} \mid \theta_1,\theta_2) = \ln\theta_1^{55}(\theta_2-\theta_1)^{100}\left(\frac{1}{2}\right)^{160}\left(\frac{1}{2}-\theta_2\right)^{50}$$
$$= \ln\theta_1^{55} + \ln(\theta_2-\theta_1)^{100} + \ln\left(\frac{1}{2}\right)^{160} + \ln\left(\frac{1}{2}-\theta_2\right)^{50}$$
$$= 55\ln\theta_1 + 100\ln(\theta_2-\theta_1) + 160\ln\left(\frac{1}{2}\right) + 50\ln\left(\frac{1}{2}-\theta_2\right)$$

Then, we take the derivative of the log likelihood with respect to θ_1.

$$\frac{\partial}{\partial\theta_1}\ln\mathcal{L}(x_1,...,x_{365} \mid \theta_1,\theta_2) = \frac{55}{\theta_1} - \frac{100}{\theta_2-\theta_1}$$

Setting this equal to 0, we solve for $\hat{\theta}_1$:

$$\frac{55}{\hat{\theta}_1} - \frac{100}{\hat{\theta}_2-\hat{\theta}_1} = 0$$
$$55(\hat{\theta}_2-\hat{\theta}_1) - 100\,\hat{\theta}_1 = 0$$
$$55\,\hat{\theta}_2 = 155\,\hat{\theta}_1$$
$$\hat{\theta}_1 = \frac{11}{31}\,\hat{\theta}_2$$

Then, we take the derivative of the log likelihood with respect to θ_2.

$$\frac{\partial}{\partial\theta_2}\ln\mathcal{L}(x_1,...,x_{365} \mid \theta_1,\theta_2) = \frac{100}{\theta_2-\theta_1} - \frac{50}{\frac{1}{2}-\theta_2}$$

Setting this equal to 0, we solve for $\hat{\theta_2}$:

$$\frac{100}{\hat{\theta_2} - \hat{\theta_1}} - \frac{50}{\frac{1}{2} - \hat{\theta_2}} = 0$$

$$100\left(\frac{1}{2} - \hat{\theta_2}\right) - 50\ (\hat{\theta_2} - \hat{\theta_1}) = 0$$

$$50 - 150\ \hat{\theta_2} + 50\ \hat{\theta_1} = 0$$

$$\hat{\theta_2} = \frac{\hat{\theta_1} + 1}{3}$$

We can now solve the simultaneous equations we have for θ_1 and θ_2 to obtain the maximum likelihood estimators for each parameter.

$$\hat{\theta_2} = \frac{\hat{\theta_1} + 1}{3}$$

Plugging in the equation for θ_1, we find

$$\hat{\theta_2} = \frac{\frac{11}{31}\ \hat{\theta_2} + 1}{3}$$

$$3\ \hat{\theta_2} = \frac{11}{31}\ \hat{\theta_2} + 1$$

$$93\ \hat{\theta_2} = 11\ \hat{\theta_2} + 31$$

$$\hat{\theta_2} = \frac{31}{82}$$

Plugging in the value for θ_2 into the equation for θ_1,

$$\hat{\theta_1} = \frac{11}{31} \cdot \frac{31}{82} = \frac{11}{82}$$

To confirm that this is in fact a maximum, we could do a second derivative test. We won't ask you do this for this multivariate case, but it would still be good to check!

Task 5 – Pareto

The Pareto distribution was discovered by Vilfredo Pareto and is used in a wide array of fields but particularly social sciences and economics. It is a density function with a slowly decaying tail, for example it can describe the wealth distribution (a small group at the top holds most of the wealth). We consider its special form given by the family of Pareto distributions Pareto$(1, \alpha)$ with densities[1]

$$f(x; \alpha) = \frac{\alpha}{x^{\alpha+1}}$$

where $x \geqslant 1$ and the real number $\alpha \geqslant 0$ is the parameter. Moreover, $f(x; \alpha) = 0$ for $x < 1$. You are given i.i.d. samples $x_1, x_2, \ldots, x_n$ from the Pareto distribution with parameter α. Find the MLE estimation of α.

[1] The more general Pareto distribution depends on an additional real positive parameter m and follows the density $f(x; \alpha, m) = \frac{\alpha \cdot m^\alpha}{x^{\alpha+1}}$ for $x \geqslant m$, and is 0 for $x < m$. Here, we consider the special case with $m = 1$.

We first need to solve for the likelihood function for which we have:

$$\mathcal{L}(x_1, \dots, x_n \mid \alpha) = \prod_{i=1}^{n} \frac{\alpha}{x_i^{\alpha+1}}$$

So, for the log-likelihood function we have:

$$\begin{aligned}\ln \mathcal{L}(x_1, \dots, x_n \mid \alpha) &= \sum_{i=1}^{n} \ln\left(\frac{\alpha}{x_i^{\alpha+1}}\right) \\ &= \sum_{i=1}^{n} (\ln(\alpha) - \ln(x_i^{\alpha+1})) \\ &= \sum_{i=1}^{n} (\ln(\alpha) - (\alpha+1)\ln(x_i)) \\ &= n\ln(\alpha) - (\alpha+1)\sum_{i=1}^{n} \ln(x_i)\end{aligned}$$

So, for the derivative with respect to α we have:

$$\frac{\mathrm{d}\ln \mathcal{L}(x_1, \dots, x_n \mid \alpha)}{\mathrm{d}\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \ln(x_i)$$

And then by setting to zero we get:

$$\begin{aligned}\frac{n}{\hat{\alpha}} - \sum_{i=1}^{n} \ln(x_i) = 0 \\ \frac{n}{\hat{\alpha}} &= \sum_{i=1}^{n} \ln(x_i) \\ \hat{\alpha} &= \frac{n}{\sum_{i=1}^{n} \ln(x_i)}\,.\end{aligned}$$

Now, let's (optionally) do a second derivative test to prove this is in fact a maximum. We have:

$$\frac{\mathrm{d}^2\ln \mathcal{L}(x_1, \dots, x_n \mid \alpha)}{\mathrm{d}\alpha^2} = -\frac{n}{\alpha^2} < 0$$

So this is a maximum!