Quiz Section 8 – Solutions

Review

1) Markov's Inequality: Let X be a non-negative random variable, and $\alpha > 0$. Then, $\mathbb{P}(X \geqslant \alpha) \leqslant \frac{\mathbb{E}[X]}{\alpha}$.

2) Chebyshev's Inequality: Suppose Y is a random variable with $\mathbb{E}[Y] = \mu$ and Var $(Y) = \sigma^2$. Then, for any $\alpha > 0$, $\mathbb{P}(|Y - \mu| \geqslant \alpha) \leqslant \frac{\sigma^2}{\alpha^2}$.

3) Chernoff Bound (for the Binomial): Suppose $X \sim$ Bin(n, p) and $\mu = np$. Then, for any $0 < \delta < 1$,

- $\mathbb{P}(|X - \mu| \geqslant \delta\mu) \leqslant e^{-\frac{\delta^2\mu}{4}}$

4) 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).

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

Task 1 – Tail bounds

Suppose $X \sim$ Binomial$(6, 0.4)$. We will bound $\mathbb{P}(X \geqslant 4)$ using the tail bounds we've learned, and compare this to the true result.

a) Give an upper bound for this probability using Markov's inequality. Why can we use Markov's inequality?

We know that the expected value of a binomial distribution is np, so: $\mathbb{P}(X \geqslant 4) \leqslant \frac{\mathbb{E}[X]}{4} = \frac{2.4}{4} = 0.6$. We can use it since X is nonnegative.

b) Give an upper bound for this probability using Chebyshev's inequality. You may have to rearrange algebraically and it may result in a weaker bound.

$\mathbb{P}(X \geqslant 4) = \mathbb{P}(X - 2.4 \geqslant 1.6) \leqslant \mathbb{P}(|X - 2.4| \geqslant 1.6)$ we can add those absolute value signs because that only adds more possible values, so it is an upper bound on the probability of $X - 2.4 \geqslant 1.6$. Then, using Chebyshev's inequality we get:
$\mathbb{P}(|X - 2.4| \geqslant 1.6) \leqslant \frac{Var(X)}{1.6^2} = \frac{1.44}{1.6^2} = 0.5625$

c) Give an upper bound for this probability using the Chernoff bound.

$\mathbb{P}(X \geqslant 4) = \mathbb{P}(X \geqslant (1 + \frac{2}{3})2.4) \leqslant e^{-(\frac{2}{3})^2\mathbb{E}[X]/4} = e^{-4\times 2.4/36} \approx 0.77$

d) Give the exact probability.

Since X is a binomial, we know it has a range from 0 to n (or in this case 0 to 6). Thus, the possible values to satisfy $X \geqslant 4$ are 4, 5, or 6. We plug in the PMF for each to get: $\mathbb{P}(X \geqslant 4) = \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) = \binom{6}{4}(0.4)^4(0.6)^2 + \binom{6}{5}(0.4)^5(0.6) + \binom{6}{6}0.4^6 \approx 0.1792$

Task 2 – Exponential Tail Bounds

Let $X \sim$ Exp(λ) and $k > 1/\lambda$. Recall that $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

a) Use Markov's inequality to bound $\mathrm{P}(X \geqslant k)$.

$$\mathbb{P}(X \geqslant k) \leqslant \frac{1}{\lambda k}$$

b) Use Chebyshev's inequality to bound $\mathsf{P}(X \geqslant k)$.

$$\mathbb{P}(X \geqslant k) = \mathbb{P}\left(X - \frac{1}{\lambda} \geqslant k - \frac{1}{\lambda}\right) \leqslant \mathbb{P}\left(\left|X - \frac{1}{\lambda}\right| \geqslant k - \frac{1}{\lambda}\right) \leqslant \frac{1}{\lambda^2(k-1/\lambda)^2} = \frac{1}{(\lambda k - 1)^2}$$

c) What is the exact formula for $\mathsf{P}(X \geqslant k)$?

$$\mathbb{P}(X \geqslant k) = e^{-\lambda k}$$

d) For $\lambda k \geqslant 3$, how do the bounds given in parts (a), (b), and (c) compare?

$$e^{-\lambda k} < \frac{1}{(\lambda k - 1)^2} < \frac{1}{\lambda k}$$

so Markov's inequality gives the worst bound.

Task 3 – 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 4 – 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:

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

Task 5 – Y Me?

Let $y_1, y_2, \ldots y_n$ be i.i.d. samples of a random variable from the family of distributions $Y(\theta)$ with densities

$$f(y;\theta) = \frac{1}{2\theta} \exp\left(-\frac{|y|}{\theta}\right),$$

where $\theta > 0$. Find the MLE for θ in terms of $|y_i|$ and n.

We follow the recipe outlined in class:

$$\begin{aligned}
\mathcal{L}(y_1, \ldots, y_n \mid \theta) &= \prod_{i=1}^{n} \frac{1}{2\theta} \exp(-\frac{|y_i|}{\theta}) \\
\ln \mathcal{L}(y_1, \ldots, y_n \mid \theta) &= \sum_{i=1}^{n} \left[-\ln 2 - \ln \theta - \frac{|y_i|}{\theta}\right] \\
\frac{\mathrm{d}}{\mathrm{d}\theta} \ln \mathcal{L}(y_1, \ldots, y_n \mid \theta) &= \sum_{i=1}^{n} \left[-\frac{1}{\theta} + \frac{|y_i|}{\theta^2}\right] = 0 \\
-\frac{n}{\hat{\theta}} + \frac{\Sigma_{i=1}^{n} |y_i|}{\hat{\theta}^2} &= 0 \\
\hat{\theta} = \frac{\Sigma_{i=1}^{n} |y_i|}{n}
\end{aligned}$$

Task 6 – 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.

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

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

So, for the log-likelihood function we have:

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

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

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

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:

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

So this is a maximum!

[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 $m = 1$.