M339J/M389J *Probability Models*
University of Texas at Austin
HW #11
Instructor: Milica Čudina

Note: You **must** show all your work. Numerical answers without a proper explanation or a clearly written down path to the solution will be assigned zero points.

Problem 11.1. (5 points) You want to fit to the observed values

$$4, 5, 7$$

a two-parameter Pareto distribution with parameters $\alpha = 4$ and θ unknown using maximum likelihood estimation. Write down **clearly** an **explicit** expression for the loglikelihood function (of course, as a function of θ).

Solution: Note that we are dealing with complete, individual data. We get the likelihood and loglikelihood functions:

$$L(\theta) = \prod_{j=1}^{n} f_{X_j}(x_j|\theta),$$
$$l(\theta) = \sum_{j=1}^{n} \ln[f_{X_j}(x_j|\theta)],$$

where $n = 3$ is the number of observations, and $x_1 = 4, x_2 = 5, x_3 = 7$ (as given in the problem) while $f_{X_j}(\cdot\,|\,\theta)$ denotes the density function of the Pareto distribution with parameters $\alpha = 4$ and θ (unknown) for $j = 1, 2, 3$. So, using our tables, we get

$$\begin{aligned} l(\theta) &= \sum_{j=1}^{3} \ln[\frac{\alpha\theta^{\alpha}}{(x_j+\theta)^{\alpha+1}}] \\ &= \sum_{j=1}^{3} [\ln(\alpha) + \alpha\ln(\theta) - (\alpha+1)\ln(x_j+\theta)] \\ &= 3(\ln(4) + 4\ln(\theta)) - (\alpha+1)[\ln(4+\theta) + \ln(5+\theta) + \ln(7+\theta)]\,. \end{aligned}$$

Problem 11.2. (10 points) Consider the following individual observed values:

$$5, 8, 10$$

of a random variable Y such that $Y = X^{-1}$ with $X \sim Gamma(\alpha = 2, \theta)$.

Calculate $\hat{\theta}_{MLE}$, the Maximum Likelihood Estimate of θ based on the above observed values.

Solution: The above observations of Y give us the corresponding observations of X:

$$1/5, 1/8, 1/10.$$

In our usual notation, the likelihood function is

$$L(\theta) = \frac{(1/(5\theta))^2 e^{-1/(5\theta)}}{(1/5)\Gamma(2)} \cdot \frac{(1/(8\theta))^2 e^{-1/(8\theta)}}{(1/8)\Gamma(2)} \cdot \frac{(1/(10\theta))^2 e^{-1/(10\theta)}}{(1/10)\Gamma(2)}$$
$$\propto \theta^{-6} e^{-(1/(5\theta)+1/(8\theta)+1/(10\theta))},$$

where we decided to use proportionality so that we do not have to write the part of the likelihood function which does not depend on θ.

Taking the natural logarithm of the final expression above, we get that the loglikelihood can be written as

$$l(\theta) = C + (-6)\ln(\theta) - (1/(5\theta) + 1/(8\theta) + 1/(10\theta))$$

with C being a constant which may depend on the observed values or the given value $\alpha = 2$, but does not depend on θ. Then,

$$l'(\theta) = -\frac{6}{\theta} + \frac{1}{5\theta^2} + \frac{1}{8\theta^2} + \frac{1}{10\theta^2}.$$

Setting the above equal to zero and solving for θ, we get

$$\hat{\theta} = \frac{1/5 + 1/8 + 1/10}{6} = \frac{17}{240}.$$

Problem 11.3. (10 points) A sample of n independent observations

$$x_1, x_2, \ldots, x_n$$

came from a distribution with the probability density function $f_x(x) = 2\theta e^{-\theta x^2}, x > 0$. Determine the maximum likelihood estimator of θ.

Solution: (c)
The density function is given by

$$f(x) = 2\theta e^{-\theta x^2} \qquad x > 0.$$

Since we are dealing with individual observations, we can use the first formula in section 15.2.2 in the textbook to write the likelihood and the loglikelihood functions as

$$L(\theta) = \prod_{j=1}^{n} (2\theta e^{-\theta x_j^2})$$
$$= 2^n \theta^n e^{-\theta \sum_{j=1}^{n} x_j^2},$$
$$l(\theta) = n\ln(2) + n\ln(\theta) - \theta \sum_{j=1}^{n} x_j^2.$$

Then, the derivative with respect to θ of the loglikelihood function is

$$l'(\theta) = \frac{n}{\theta} - \sum_{j=1}^{n} x_j^2.$$

Setting the above equal to zero and solving for θ we get that the MLE for θ equals

$$\hat{\theta} = \frac{n}{\sum_{j=1}^{n} x_j^2}.$$

Problem 11.4. (10 points) Consider a random variable Y such that $Y = e^X$ with $X \sim Gamma(\alpha = 2, \theta)$. Your colleague was playing with the collected data and the only things you still know about the observations from Y are:

(i) There was a total of 20 observations;
(ii) The product of all observations was $5,000$.

Find $\hat{\theta}_{MLE}$, i.e., the Maximum Likelihood Estimate of θ based on the observed values.

Solution: Any observation y_k of Y gives us the corresponding observation $x_k = \ln(y_k)$ of X. In our usual notation, $n = 20$ and the likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{k=1}^{n} \frac{(x_k/\theta)^{\alpha} e^{-x_k/\theta}}{x_k \Gamma(\alpha)} \\ &\propto \prod_{k=1}^{n} \theta^{-\alpha} e^{-x_k/\theta} \\ &= \theta^{-n\alpha} e^{-\frac{1}{\theta}\sum_{k=1}^{n} x_k} \end{aligned}$$

where we decided to use proportionality so that we do not have to write the part of the likelihood function which does not depend on θ.

Taking the natural logarithm of the final expression above, we get that the loglikelihood can be written as

$$l(\theta) = -n\alpha \ln(\theta) - \frac{1}{\theta}\sum_{k=1}^{n} x_k.$$

Then,

$$l'(\theta) = -\frac{n\alpha}{\theta} + \frac{1}{\theta^2}\sum_{k=1}^{n} x_k.$$

Setting the above equal to zero and solving for θ, we get

$$\hat{\theta} = \bar{x}/\alpha$$

where $\bar{x}$ denotes the sample mean.

However,

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} \ln(y_k) = \ln(\prod_{k=1}^{n} y_k).$$

So, using (i) and (ii) from the problem statement, we get

$$\hat{\theta} = \frac{\frac{\ln(5000)}{20}}{2} \approx 0.2129.$$

Problem 11.5. (5 points) Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with an unknown mean θ. Consider these as individual, unmodified data. What is the expression for the maximum likelihood estimator of θ denoted by $\hat{\theta}_{MLE}$?

Solution: The probability density function of the exponential distribution is $f(x) = \frac{1}{\theta}e^{-x/\theta}, x > 0$. So, the likelihood function with the fiven data can be written as

$$L(\theta) = \prod_{i=1}^{n}\left(\frac{1}{\theta}e^{-X_i/\theta}\right) = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n} X_1}.$$

Thus, the log-likelihood function is

$$l(\theta) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} X_i.$$

We differentiate the log-likelihood function and equate it to zero in order to find the maximum likelihood estimator.

$$l'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} X_i = 0.$$

The MLE estimator is, hence,

$$\hat{\theta}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Problem 11.6. (10 points) *Source: Sample STAM problem #179.* The time to an accident follows an exponential distribution. A random sample of size two has a mean time of 4. Let Y denote the average of a new sample of size two from the same distribution. Calculate the maximum likelihood estimate of $\mathbb{P}[Y > 8]$.

<u>Hint:</u> Remember that the sum of independent, identically distributed exponential random variables has a gamma distribution. You can convince yourselves of this fact using moment generating functions.

Solution: As we have seen in a previous problem, and as we should remember for the future, the MLE for the parameter of the exponential distribution is exactly the sample mean. So, $\hat{\theta}_{MLE} = 4$.

Using the hint, we conclude that $2Y \sim Gamma(\alpha = 2, \theta = 4)$. Then, we consult our STAM tables and see that

$$\mathbb{P}[Y > 8] = \mathbb{P}[2Y > 16] = \int_{16}^{\infty} \frac{(x/4)^2 e^{-x/4}}{x}\, dx = \frac{1}{16}\int_{16}^{\infty} xe^{-x/4}\, dx.$$

At this point, we can focus in the integral in the expression above. Let W denote an exponential random variable with mean 4. Then,

$$\begin{aligned}\int_{16}^{\infty} x(\frac{1}{4}e^{-x/4})\, dx &= \mathbb{E}[W\mathbb{I}_{[W>16]}] = 4 - \mathbb{E}[W\mathbb{I}_{[W\leq 16]}] = 4 - \mathbb{E}[W\mathbb{I}_{[W\leq 16]}] - 16\mathbb{P}[W > 16] + 16\mathbb{P}[W > 16] \\ &= 4 - \mathbb{E}[W \wedge 16] + 16S_W(16).\end{aligned}$$

Using the STAM tables again, we get

$$\int_{16}^{\infty} x(\frac{1}{4}e^{-x/4})\, dx = 4 - 4(1 - e^{-16/4}) + 16e^{-16/4} = 20e^{-4}.$$

So,

$$\mathbb{P}[Y > 8] = \frac{1}{4}\int_{16}^{\infty} x(\frac{1}{4}e^{-x/4})\,dx = \frac{1}{4}(20e^{-4}) = 5e^{-4} \approx 0.09158.$$