Alex Tsun
CSE 312: Foundations of Computing II
Section #7
August 6th, 2020

Section #7 Solutions

Shreya Jayaraman, Luxi Wang, Alex Tsun

1. The Pareto distribution, was discovered by Vilfredo Patero 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). The PDF is given by:

$$f_X(x; m, \alpha) = \frac{\alpha m^\alpha}{x^{\alpha+1}}$$

where $x \geq m$ and real $\alpha, m > 0$. m describes the minimum value that X takes on (scale) and α is the shape. So the range of X is $\Omega_X = [m, \infty)$. Assume that m is given and that $x_1, x_2, \ldots, x_n$ are i.i.d. samples from the Pareto distribution. Find the MLE estimation of α.

Solution

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

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

So, for the log-likelihood function we have:

$$\begin{aligned} l(\alpha) &= \sum_{i=1}^{n} (\ln(\frac{\alpha m^\alpha}{x_i^{\alpha+1}})) \\ &= \sum_{i=1}^{n} (\ln(\alpha m^\alpha) - \ln(x_i^{\alpha+1})) \\ &= \sum_{i=1}^{n} (\ln(\alpha) + \alpha \ln(m) - (\alpha+1)\ln(x_i)) \\ &= n\ln(\alpha) + n\alpha\ln(m) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i) \end{aligned}$$

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

$$\frac{\partial l(\alpha)}{\partial \alpha} = \frac{n}{\alpha} + n\ln(m) - \sum_{i=1}^{n}\ln(x_i)$$

And then by setting to zero we get:

$$\frac{n}{\hat{\alpha}} + n\ln(m) - \sum_{i=1}^{n} \ln(x_i) = 0$$

$$\begin{aligned}
\frac{n}{\hat{\alpha}} &= \sum_{i=1}^{n} \ln(x_i) - n\ln(m) \\
\hat{\alpha} &= \frac{n}{\sum_{i=1}^{n} \ln(x_i) - n\ln(m)} \\
&= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i) - \ln(m)} \\
&= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \ln(x_i) - \ln(m)} \\
&= \frac{1}{\overline{\ln(x)} - \ln(m)}
\end{aligned}$$

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

$$\frac{\partial^2 l(\alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0$$

So this is a maximum!

2. 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?

Solution

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.

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

Then, we use this to determine the log likelihood.

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

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

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

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

$$\frac{55}{\hat{\theta}_1} - \frac{100}{\hat{\theta}_2 - \hat{\theta}_1} = 0$$

$$55(\hat{\theta}_2 - \hat{\theta}_1) - 100\hat{\theta}_1 = 0$$

$$55\hat{\theta}_2 = 155\hat{\theta}_1$$

$$\hat{\theta}_1 = \frac{11\hat{\theta}_2}{31}$$

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

$$\frac{\partial}{\partial \theta_2} \ln L(x_1, x_2, ..., x_n | \theta_1, \theta_2) = \frac{100}{\theta_2 - \theta_1} - \frac{50}{\frac{1}{2} - \theta_2}$$

Setting this equal to 0, we solve for $\hat{\theta}_2$:

$$\frac{100}{\hat{\theta}_2 - \hat{\theta}_1} - \frac{50}{\frac{1}{2} - \hat{\theta}_2} = 0$$

$$100(\frac{1}{2} - \hat{\theta}_2) - 50(\hat{\theta}_2 - \hat{\theta}_1) = 0$$

$$50 - 150\hat{\theta}_2 + 50\hat{\theta}_1 = 0$$

$$\hat{\theta}_2 = \frac{50\hat{\theta}_1 + 50}{150} = \frac{\hat{\theta}_1 + 1}{3}$$

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

$$\hat{\theta}_2 = \frac{\theta_1 + 1}{3}$$

Plugging in the equation for θ_1, we find

$$\hat{\theta}_2 = \frac{\frac{11\hat{\theta}_2}{31} + 1}{3}$$

$$3\hat{\theta}_2 = \frac{11\hat{\theta}_2}{31} + 1$$

$$93\hat{\theta}_2 = 11\hat{\theta}_2 + 31$$

$$\hat{\theta}_2 = \frac{31}{82}$$

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

$$\hat{\theta}_1 = \frac{11\frac{31}{82}}{31} = \frac{11}{82}$$

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!

3. Let $X_1, \ldots, X_n$ be a random sample from the distribution with PDF $f_X(x \mid \theta) = (\theta^2+\theta)x^{\theta-1}(1-x)$ for $0 < x < 1$ and $\theta > 0$. What is the MOM estimator for θ?

   Solution

   First, we need to determine the first moment of X, E[X]:

$$\begin{aligned} E[X] &= \int_0^1 x(\theta^2+\theta)x^{\theta-1}(1-x)\ dx \\ &= \int_0^1 (\theta^2+\theta)x^{\theta}(1-x)dx \\ &= (\theta^2+\theta)\int_0^1 x^{\theta} - x^{\theta+1}dx \\ &= (\theta^2+\theta)\left[\frac{x^{\theta+1}}{\theta+1} - \frac{x^{\theta+2}}{\theta+2}\right]_0^1 \\ &= \frac{\theta(\theta+1)}{(\theta+1)(\theta+2)} \\ &= \frac{\theta}{\theta+2} \end{aligned}$$

   We then set the first true moment to the first sample moment as follows:

$$\frac{\theta}{\theta+2} = \bar{x}$$

   Solving for θ, we get

$$\theta = (\theta+2)\bar{x}$$

$$\theta - \theta\bar{x} = 2\bar{x}$$

$$\theta = \frac{2\bar{x}}{1-\bar{x}}$$

   (Notice, though, that the original PDF looks a lot like the beta distribution PDF.
   In fact, $X \sim Beta(\alpha,\beta)$ with $\alpha = \theta$ and $\beta = 2$, for which we know $E[X] = \frac{\alpha}{\alpha+\beta} = \frac{\theta}{\theta+2}$.)

4. Let $x_1, x_2, ..., x_n$ be a random sample from a distribution with PDF $f(x; \theta, k) = \theta \frac{k^\theta}{x^{\theta+1}}$ where $x \geq k$ and k, θ are positive real numbers. The expectation is given by

$$E[X] = \frac{k\theta}{\theta - 1}$$

Determine

(a) the maximum likelihood estimator of k and θ

(b) the method of moments estimator of θ

(c) suppose the observed samples were $\mathbf{x} = (0.18, 0.94, 0.54, 0.11, 0.62, 0.45)$. What are the MLE and MOM estimators of k and θ to 3 decimal places? (Use the same value of k from the MLE estimator in (a) for both of your calculations.)

Solution

a. We want to find the likelihood of the data samples given the parameter θ. To do this, we take the following product, from i equal to 1 to n (all the data points). We can just plug in the PDF here to get this:

$$\begin{aligned} L(x_1, ..., x_n \mid k, \theta) &= \prod_{i=1}^{n} \left(\frac{\theta k^\theta}{x_i^{\theta+1}} \right) \\ &= \theta^n \cdot k^{n\theta} \cdot \prod_{i=1}^{n} \frac{1}{x_i^{\theta+1}} \end{aligned}$$

Then, we find the log likelihood:

$$\begin{aligned} \ln L(x_1, ..., x_n \mid k, \theta) &= \ln \left(\theta^n \cdot k^{n\theta} \cdot \prod_{i=1}^{n} \frac{1}{x_i^{\theta+1}} \right) \\ &= \ln \theta^n + \ln k^{n\theta} + \sum_{i=1}^{n} \ln \frac{1}{x_i^{\theta+1}} \\ &= n \ln \theta + n\theta \ln k + \sum_{i=1}^{n} \ln 1 - \ln x_i^{\theta+1} \\ &= n \ln \theta + n\theta \ln k - (\theta + 1) \sum_{i=1}^{n} \ln x_i \end{aligned}$$

We want to maximize the log likelihood. First, let's only consider the term k, for which we know that to maximize the log likelihood, the term $n \ln k$ must be as large as possible (since $n, k > 0$). So k must have the maximum value possible. From the question statement, $k \leq x$, so the largest value k can have is the smallest x_i

$$\hat{k} = \min_i x_i$$

Now, to consider θ, we take the derivative of the log likelihood with respect to θ:

$$\frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_n \mid k, \theta) = \frac{n}{\theta} + n \ln k - \sum_{i=1}^{n} \ln x_i$$

Letting this equal 0, we solve for θ:

$$\frac{n}{\hat{\theta}} + n \ln k - \sum_{i=1}^{n} \ln x_i = 0$$

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln x_i - n \ln k}$$

We wouldn't ask you to consider the second derivative test in this multivariate case, but here, the second derivative test is actually a bit easier. We already know that $\hat{k}$ is a maximum based on the earlier explanation. So, we just need to take the second derivative in respect to θ and we have:

$$\frac{\partial^2}{\partial\theta^2} \ln L(x_1, \ldots, x_n \mid k, \theta) = -\frac{n}{\theta^2} < 0$$

So we have a maximum!

b. Setting the first true moment to the first sample moment, we solve for θ:

$$\frac{k\theta}{\theta - 1} = \bar{x}$$

$$k\theta = (\theta - 1)\bar{x}$$

$$\theta(\bar{x} - k) = \bar{x}$$

$$\hat{\theta} = \frac{\bar{x}}{\bar{x} - k}$$

c. Recall that $\mathbf{x} = (0.18, 0.94, 0.54, 0.11, 0.62, 0.45)$. For the MLE, we have $\hat{k} = \min(\mathbf{x}) = 0.11$ and:

$$\begin{aligned}
\hat{\theta} &= \frac{n}{\sum_{i=1}^{n} \ln x_i - n \ln k} \\
&= \frac{6}{(\ln(0.18) + \ln(0.94) + \ln(0.54) + \ln(0.11) + \ln(0.62) + \ln(0.45)) - 6(\ln(0.11))} \\
&= 0.814
\end{aligned}$$

Then, for MOM we have:

$$
\begin{aligned}
\hat{\theta} &= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} x_i - k} \\
&= \frac{\frac{1}{6}(0.18 + 0.94 + 0.54 + 0.11 + 0.62 + 0.45)}{\frac{1}{6}(0.18 + 0.94 + 0.54 + 0.11 + 0.62 + 0.45) - 0.11} \\
&= 1.303
\end{aligned}
$$

So the final answers are:
MLE: $\hat{\theta} = 0.814$
MOM: $\hat{\theta} = 1.303$

5. Suppose $x_1, \ldots, x_{2n}$ are iid realizations from the Laplace density (double exponential density)

$$f_X(x \mid \theta) = \frac{1}{2} e^{-|x-\theta|}$$

Find the MLE for θ. For this problem, you need not verify that the MLE is indeed a maximizer. You may find the **sign** function useful:

$$\text{sgn}(x) = \begin{cases} +1, & x > 0 \\ -1, & x < 0 \end{cases}$$

(in our case undefined at 0)

Solution

$$\begin{aligned} L(x_1, \ldots, x_{2n} \mid \theta) &= \prod_{i=1}^{2n} \frac{1}{2} e^{-|x_i-\theta|} \\ \ln L(x_1, \ldots, x_{2n} \mid \theta) &= \sum_{i=1}^{2n} \left[-\ln 2 - |x_i - \theta|\right] \\ \frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_{2n} \mid \theta) &= \sum_{i=1}^{2n} \text{sgn}(x_i - \theta) = 0 \\ \hat{\theta} &= \text{any value in } \left[x'_n, x'_{n+1}\right] \end{aligned}$$

where x'_i is the i^{th} order statistic: the i^{th} smallest observation.

Intuitively (ignoring the edge cases) this is because if $\theta \in [x'_n, x'_{n+1}]$, for $i \in \{1, \ldots, n\}$, $\text{sgn}(x'_i - \theta) = -1$ and for $i \in \{n+1, \ldots, 2n\}$, $\text{sgn}(x'_i - \theta) = 1$. So the sum of these will be zero.

If you want to argue that this is a global maximizer, note that the log likelihood is the sum of concave functions (negative absolute value), so every critical point is a global maximizer.

However, if you want to argue this more rigorously considering edge cases, we need to show that it is a maximizer, but the second derivative test is inconclusive because the second derivative is 0 except at $x_1, x_2, \ldots, x_{2n}$, where it is undefined. We inspect the log likelihood $\ln L(x_1, \ldots, x_{2n} \mid \theta)$ directly, ignoring the constant $-\ln 2$ terms:

$$S = -\sum_{i=1}^{2n} |x_i - \theta|$$

If $\theta \in [x'_n, x'_{n+1}]$, $S = -\sum_{i=1}^{n} (x'_{n+i} - x'_{n+1-i})$. When θ crosses an endpoint of $[x'_n, x'_{n+1}]$, the term $|x'_{n+1} - x'_n|$ in this sum is replaced by something greater, so S decreases. Therefore, the log likelihood is maximized when $\theta \in [x'_n, x'_{n+1}]$. This is also why we need to include the end points in our interval.