

PS1

Samuel Olson

2025-01-27

Problem 1

Find the method of moment estimators (MMEs) of the unknown parameters based on a random sample X_1, X_2, \dots, X_n of size n from the following distributions:

a)

Negative Binomial $(3, p)$, unknown p :

Noting definitions and the distribution in question:

$$\mu = \frac{3(1-p)}{p}$$

Given a random sample X_1, X_2, \dots, X_n , the sample mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

We then relate the population mean to the sample mean:

$$\frac{3(1-p)}{p} = \bar{X}$$

Solving for p :

$$3(1-p) = \bar{X}p, 3 - 3p = \bar{X}p, 3 = p(\bar{X} + 3), p = \frac{3}{3 + \bar{X}}$$

Thus, the method of moments estimator for p is:

$$\hat{p} = \frac{3}{3 + \bar{X}}$$

b)

Double Exponential (μ, σ) , unknown μ and σ :

Noting definitions and the distribution in question:

$$\mu_1(\mu, \sigma^2) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

and:

$$\mu_2(\mu, \sigma^2) = 2\sigma^2 + \mu^2$$

We equate the population moments to the sample moments:

$$\mu = \bar{X}, \quad 2\sigma^2 = S^2$$

where S^2 is the sample variance (of the random variable):

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Solving for σ :

$$\sigma = \sqrt{\frac{S^2}{2}}$$

Thus, the method of moments estimators are:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma} = \sqrt{\frac{S^2}{2}}$$

Note:

See “Table of Common Distributions” in Casella & Berger (pages 623–623) for the definitions/properties of the above distributions.

Problem 2

7.1, Casella & Berger

Given the pmf table for X :

x	$f(x 1)$	$f(x 2)$	$f(x 3)$
0	$\frac{1}{3}$	$\frac{1}{4}$	0
1	$\frac{1}{3}$	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$
4	$\frac{1}{6}$	0	$\frac{1}{4}$

The MLE $\hat{\theta}$ is found by maximizing $f(x|\theta)$ for each observed x :

0

For $x = 0$:

$$f(0|1) = \frac{1}{3}, \quad f(0|2) = \frac{1}{4}, \quad \text{and} \quad f(0|3) = 0$$

The maximum is $f(0|1) = \frac{1}{3}$, so $\hat{\theta} = 1$.

1

For $x = 1$:

$$f(1|1) = \frac{1}{3}, \quad f(1|2) = \frac{1}{4}, \quad \text{and} \quad f(1|3) = 0$$

The maximum is $f(1|1) = \frac{1}{3}$, so $\hat{\theta} = 1$.

2

For $x = 2$:

$$f(2|1) = 0, \quad f(2|2) = \frac{1}{4}, \quad \text{and} \quad f(2|3) = \frac{1}{4}.$$

There is not a unique maximum, as the maxima are $f(2|2) = f(2|3) = \frac{1}{4}$, so $\hat{\theta} = 2$ or 3.

3

For $x = 3$:

$$f(3|1) = \frac{1}{6}, \quad f(3|2) = \frac{1}{4}, \quad \text{and} \quad f(3|3) = \frac{1}{2}.$$

The maximum is $f(3|3) = \frac{1}{2}$, so $\hat{\theta} = 3$.

4

For $x = 4$:

$$f(4|1) = \frac{1}{6}, \quad f(4|2) = 0, \quad \text{and} \quad f(4|3) = \frac{1}{4}.$$

The maximum is $f(4|3) = \frac{1}{4}$, so $\hat{\theta} = 3$.

Conclusions

Summary of MLE values for $\hat{\theta}$:

x	$\hat{\theta}$
0	1
1	1
2	2 or 3 (Either works)
3	3
4	3

Overall, the MLE $\hat{\theta}$ is:

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \{1,2,3\}} f(x|\theta)$$

And overall MLE is not unique, i.e., it is defined as a function of x (the MLE depends on the value of x), as illustrated above. However, we may describe it as follows, “For $x \in \{0, 1\}$ the MLE $\hat{\theta}$ is 1, otherwise 3, i.e. for $x \in \{2, 3, 4\}$ the MLE $\hat{\theta}$ is 3.

Problem 3

An indicator function $I(A)$ of an event A has the form:

$$I(A) = \begin{cases} 1, & \text{if event } A \text{ holds true,} \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that A_1, \dots, A_n are n separate events. Show that:

$$\prod_{i=1}^n I(A_i) = I(B)$$

where B is the event that $B = \bigcap_{i=1}^n A_i$.

So for this problem, we say the event $B = \bigcap_{i=1}^n A_i$ holds true if and only if all events A_1, A_2, \dots, A_n are true simultaneously.

So we need to prove both directions of the proof to conclude.

For one direction, consider the definition of the indicator function:

$$I(B) = I\left(\bigcap_{i=1}^n A_i\right) = \begin{cases} 1, & \text{if all } A_i \text{ hold true, i.e., } A_1 \cap A_2 \cap \dots \cap A_n, \\ 0, & \text{otherwise.} \end{cases}$$

For the product $\prod_{i=1}^n I(A_i)$, we then have:

$$\prod_{i=1}^n I(A_i) = I(A_1) * I(A_2) * \dots * I(A_n)$$

Where $I(A_i)$ is 1 if A_i is true, and 0 otherwise, for all i .

It follows then that the product $\prod_{i=1}^n I(A_i)$ will equal 1 iff $I(A_i) = 1$, for all i , i.e., all events A_i are true.

If any A_i is false, then $I(A_i) = 0$ for an i , which then makes the entire product 0.

Therefore, the product $\prod_{i=1}^n I(A_i)$ is therefore 1 iff all events A_1, A_2, \dots, A_n are true, which is equivalent to the definition of $I(B)$.

If any event A_i is false, the product is 0, again being equivalent to the definition of $I(B)$.

We then conclude that:

$$\prod_{i=1}^n I(A_i) = I(B), \quad \text{where } B = \bigcap_{i=1}^n A_i.$$

Problem 4

Maximum-Likelihood & Indicator Functions

Given a random sample X_1, \dots, X_n from a pdf/pmf $f(x|\theta)$, $\theta \in \Theta \subset \mathbb{R}$, we know that the likelihood function will generically be

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta), \quad \theta \in \Theta,$$

but there's one subtle point to again highlight about how to exactly write the likelihood expression depending on the support of $f(x|\theta) > 0$.

- Recall the support or range of $f(x|\theta)$ is a set

$$S_\theta = \{x \in \mathbb{R} : f(x|\theta) > 0\},$$

which could possibly depend on $\theta \in \Theta$. For example, an exponential distribution has a pdf

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

with a parameter $\theta > 0$, and in this case the support $S_\theta = (0, \infty)$ doesn't depend on $\theta \in \Theta = (0, \infty)$.

On the other hand, the pdf (1):

(1)

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x \leq \theta, \\ 0, & \text{otherwise,} \end{cases}$$

with parameter $\theta > 0$, does have a support $S_\theta = (0, \theta]$ depending on $\theta \in \Theta = (0, \infty)$.

- It's always true that $f(x|\theta) = f(x|\theta)I(x \in S_\theta)$ for all $x \in \mathbb{R}$ and so always true that (2):

(2)

$$L(\theta) = \prod_{i=1}^n [f(x_i|\theta)I(x_i \in S_\theta)] = \left(\prod_{i=1}^n f(x_i|\theta) \right) I(x_1, \dots, x_n \text{ are all in } S_\theta).$$

Questions

a)

If X_1, \dots, X_n are a random sample from an exponential pdf $f(x|\theta)$, $\theta > 0$ (and so X_1, \dots, X_n are positive values), show that the likelihood function (2) can be written as

$$L(\theta) = \frac{1}{\theta^n} e^{-\sum_{i=1}^n x_i/\theta},$$

and that the MLE of θ is \bar{X}_n .

(Message here: The support of an exponential doesn't depend on θ , so we don't have to worry about indicating the support.)

The exponential pdf is:

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0, \theta > 0.$$

The likelihood function for the sample X_1, \dots, X_n would then be given by:

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \left(\frac{1}{\theta} e^{-x_i/\theta} \right) = \frac{1}{\theta^n} e^{-\sum_{i=1}^n x_i/\theta}$$

To find the MLE of θ , we maximize the log-likelihood function (log instead of just likelihood for ease of analysis):

$$L(\theta) = \log L(\theta) = -n \log \theta - \frac{\sum_{i=1}^n x_i}{\theta}$$

To find the maximum, we differentiate $L(\theta)$ with respect to θ and solve for setting this relation equal to 0:

$$\frac{\partial L(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} - \frac{n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} = 0$$

Rearranging and solving for θ , we have:

$$\sum_{i=1}^n x_i = n\theta, \quad \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}_n$$

And we conclude the MLE of θ is:

$$\hat{\theta} = \bar{X}_n$$

Additional Check To ensure this is a maximum, another validation is to check that:

$$\left. \frac{d^2 \log L(\theta)}{d\theta^2} \right|_{\hat{\theta}} < 0$$

To that end, we have the second derivative of:

$$\frac{\partial^2 L(\theta)}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2 \sum_{i=1}^n x_i}{\theta^3}$$

With note of $\hat{\theta} = \bar{X}_n = \frac{\sum_{i=1}^n x_i}{n}$, we have:

$$\left. \frac{\partial^2 L(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}} = \frac{n}{\hat{\theta}^2} - \frac{2 \sum_{i=1}^n x_i}{\hat{\theta}^3} = \frac{n}{\hat{\theta}^2} - \frac{2n\hat{\theta}}{\hat{\theta}^3} = \frac{n}{\hat{\theta}^2} - \frac{2n}{\hat{\theta}^2} = -\frac{n}{\hat{\theta}^2}$$

Given $n > 0$ and $\hat{\theta}^2 > 0$, the second derivative is always negative, such that we have confirmed that $\hat{\theta} = \bar{X}_n$ is a maximum.

b)

If X_1, \dots, X_n are a random sample from the pdf

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2} & 0 < x \leq \theta, \\ 0 & \text{otherwise} \end{cases}$$

(and so $X_1, \dots, X_n > 0$ are less than or equal to θ), show that the likelihood function (2) can be written as

$$L(\theta) = \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} I\left(\max_{1 \leq i \leq n} x_i \leq \theta\right)$$

and that the MLE of θ is $\max_{1 \leq i \leq n} X_i$.

(Message here: The support in this case depends on θ , so we should think about indicator functions in writing the likelihood.)

The given pdf is:

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2} & 0 < x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

The likelihood function for a random sample X_1, \dots, X_n is:

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{2x_i}{\theta^2} I(x_i \leq \theta)$$

Simplifying gives us:

$$L(\theta) = \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} I(x_1 \leq \theta, x_2 \leq \theta, \dots, x_n \leq \theta)$$

Note: The indicator function $I(x_1 \leq \theta, \dots, x_n \leq \theta)$ is equivalent to $I(\max_{1 \leq i \leq n} x_i \leq \theta)$ because θ must be greater than or equal to all observed values for the likelihood to be nonzero.

We may then write the above likelihood function as:

$$L(\theta) = \frac{2^n \prod_{i=1}^n x_i}{\theta^{2n}} I\left(\max_{1 \leq i \leq n} x_i \leq \theta\right)$$

As the likelihood function includes the indicator $I(\max_{1 \leq i \leq n} x_i \leq \theta)$, then this relation must have θ satisfy $\theta \geq \max_{1 \leq i \leq n} x_i$ for $L(\theta) > 0$. For $\theta \geq \max_{1 \leq i \leq n} x_i$, the likelihood decreases as θ increases because the denominator θ^{2n} grows. To maximize the likelihood, set θ to the smallest value that satisfies the condition $\theta \geq \max_{1 \leq i \leq n} x_i$, which is done when $\max_{1 \leq i \leq n} x_i = \theta$.

We then conclude that the MLE as specified is:

$$\hat{\theta} = \max_{1 \leq i \leq n} x_i$$

Additional Check To ensure this is a maximum, another validation is to check that:

$$\left. \frac{d^2 \log L(\theta)}{d\theta^2} \right|_{\hat{\theta}} < 0$$

However, because of the inclusion of the Indicator function, we are not able to differentiate and solve using the typical “Calculus Method”.

For that reason, it is important to note that analytically, we are maximizing the function in question globally, such that despite not being able to validate our answer via differentiation, we do in fact have the MLE as defined, $\hat{\theta}$, maximizing the likelihood function specified.

Problem 5

Problem 7.6(b)-(c), Casella & Berger (Skip part (a).)

Let X_1, \dots, X_n be a random sample from the pdf

$$f(x|\theta) = \theta x^{-2}, \quad 0 < \theta \leq x < \infty.$$

(b)

Find the MLE of θ .

The likelihood function for the random sample X_1, \dots, X_n is:

$$L(\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \theta x_i^{-2} = L(\theta) = \theta^n \prod_{i=1}^n x_i^{-2}$$

The support of the distribution depends on θ , so we should use an indicator function within the likelihood function to ensure that $\theta \leq x_{(1)}$, where we define the first order statistic in the usual manner, i.e. $x_{(1)} = \min(X_1, \dots, X_n)$.

Using this approach we rewrite as:

$$L(\theta) = \theta^n \prod_{i=1}^n x_i^{-2} I_{[\theta, \infty)}(x_{(1)})$$

The term θ^n is increasing in θ , so to maximize $L(\theta)$, we want the largest θ . However, the indicator function $I_{[\theta, \infty)}(x_{(1)})$ ensures $L(\theta) = 0$ for $\theta > x_{(1)}$. Thus, the maximum likelihood occurs at the largest possible value of θ satisfying $\theta \leq x_{(1)}$, which is $x_{(1)}$.

Taken together, we then know the MLE of θ is:

$$\hat{\theta} = x_{(1)}$$

(c)

Find the method of moments estimator of θ .

To find the method of moments estimator (MME) of θ , we use the given pdf:

$$f(x|\theta) = \theta x^{-2}, \quad 0 < \theta \leq x < \infty$$

The first moment (mean) of X is:

$$E[X] = \int_{\theta}^{\infty} x f(x|\theta) dx = \int_{\theta}^{\infty} x (\theta x^{-2}) dx = \int_{\theta}^{\infty} \theta x^{-1} dx = \theta \int_{\theta}^{\infty} x^{-1} dx = \theta [\ln x]_{x=\theta}^{\infty}$$

Evaluating this gives us

$$E[X] = \theta(\ln(\infty) - \ln(\theta))$$

Since $\ln(\infty) \rightarrow \infty$, the expected value $E[X]$ is infinite. This indicates that the first moment does not exist.

Because the first moment does not exist, the method of moments estimator cannot be defined. Thus, the MME for θ does not exist.