Department of Industrial Engineering & Operations Research

IEOR 165 (Spring 2017)

Homework 3

Due: Thursday ~~Mar 16~~ April 6

Question 1. Recall from the previous homework the maximum likelihood estimate derived from iid data $X_1, \ldots, X_n$ with the pdf

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

Find the bias, variance, and mean squared error (MSE) of the estimate $\hat{\theta}_{mle}$.
Hint: You can use the fact that $X_i \sim \text{Erlang}(2, \theta)$ and that $Y \sim \text{Erlang}(2n, \theta)$ if $Y = \sum_{i=1}^n X_i$.

Solution: From the solution set of homework 2 we see that $\hat{\theta}_{mle} = \frac{2n}{\sum_{i=1}^n X_i}$ and we need to find $\text{Bias}(\hat{\theta}_{mle})$ and $\text{Var}(\hat{\theta}_{mle})$. First to compute $\text{Bias}(\hat{\theta}_{mle})$ we need to find $\mathbb{E}[\hat{\theta}_{mle}]$. Using the hint we can find this expectation as follows:

$$\begin{aligned} \mathbb{E}[\hat{\theta}_{mle}] = 2n\mathbb{E}[\frac{1}{Y}] = 2n \int_0^\infty \frac{y^{-1} \cdot y^{2n-1} \cdot \exp(-\theta y) \cdot \theta^{2n}}{(2n-1)!} dy \\ = \frac{2n\theta}{2n-1} \int_0^\infty \frac{y^{2n-2}\theta^{2n-1} \exp(-\theta y)}{(2n-2)!} dy = \frac{2n\theta}{2n-1} \end{aligned} \tag{1}$$

So $\text{Bias}(\hat{\theta}_{mle}) = \mathbb{E}[\hat{\theta}_{mle}] - \theta = \frac{\theta}{2n-1}$. Now let's compute the variance, we note that $\text{Var}(\hat{\theta}_{mle}) = 4n^2 \text{Var}(\frac{1}{Y})$. Since we have already computed the first moment of $\frac{1}{Y}$ we just need to compute the second moment:

$$\begin{aligned} \mathbb{E}[\frac{1}{Y^2}] = \int_0^\infty \frac{y^{2n-3} \exp(-\theta y)\theta^{2n}}{(2n-1)!} dy = \frac{\theta^2}{(2n-1)(2n-2)} \int_0^\infty \frac{y^{2n-3} \exp(-\theta y)\theta^{2n-2}}{(2n-3)!} dy \\ = \frac{\theta^2}{(2n-1)(2n-2)} \end{aligned} \tag{2}$$

Hence:

$$\text{Var}(\hat{\theta}_{mle}) = 4n^2 (\frac{\theta^2}{(2n-1)(2n-2)} - \frac{\theta^2}{(2n-1)^2}) = \frac{4n^2\theta^2}{(2n-1)^2(2n-2)} \tag{3}$$

Finally to compute MSE we use the formula $\text{MSE}(\hat{\theta}_{mle}) = \text{Var}(\hat{\theta}_{mle}) + \text{Bias}(\hat{\theta}_{mle})^2$:

$$\text{MSE}(\hat{\theta}_{mle}) = \frac{4n^2\theta^2}{(2n-1)^2(2n-2)} + \frac{\theta^2}{(2n-1)^2} = \frac{\theta^2(n+1)}{(2n-1)(n-1)} \tag{4}$$

Question 2. In the same setting as problem 1, recall that a shrinkage estimator is an estimator which is a constant multiple of the MLE estimator (i.e. $\hat{\theta}_{shrink} = k\hat{\theta}_{mle}$ for some constant k). Compute an unbiased shrinkage estimator and minimum MSE shrinkage estimator for θ. What is the bias, variance, and MSE of these estimators?

Solution: To obtain the unbiased shrinkage estimator we note that $\mathbb{E}[\hat{\theta}_{mle}] = \frac{2n}{2n-1}\theta$, hence we can use $\hat{\theta}_{unb} = \frac{2n-1}{2n}\hat{\theta}_{mle}$. Clearly this estimator has zero bias so the MSE is equal to $\text{Var}(\hat{\theta}_{unb}) = (\frac{2n-1}{2n})^2\text{Var}(\hat{\theta}_{mle}) = \frac{\theta^2}{2n-2}$.
Next to find the minimum MSE shrinkage estimator we can solve the problem:

$$\min_k \mathbb{E}[(k\hat{\theta}_{mle} - \theta)^2] \tag{5}$$

By taking the derivative and setting to zero we get that the optimal value for k is given by $k = \frac{\theta\mathbb{E}[\hat{\theta}_{mle}]}{\mathbb{E}[\hat{\theta}^2_{mle}]} = \frac{2n-2}{2n}$. We can compute the bias and variance as follows:

$$\begin{aligned} \text{Bias}(\hat{\theta}_{mse}) &= (\frac{2n-2}{2n})\mathbb{E}[\hat{\theta}_{mle}] - \theta = \frac{-\theta}{2n-1} \\ \text{Var}(\hat{\theta}_{mse}) &= (\frac{2n-2}{2n})^2\text{Var}(\hat{\theta}_{mle}) = \frac{\theta^2(2n-2)}{(2n-1)^2} \end{aligned} \tag{6}$$

So using the formula for MSE we obtain that:

$$\text{MSE}(\hat{\theta}_{mse}) = \frac{(2n-2)\theta^2}{(2n-1)^2} + \frac{\theta^2}{(2n-1)^2} = \frac{\theta^2}{(2n-1)} \tag{7}$$

Question 3. River floods are often measured by their discharges (in units of feet cubed per second). The following table gives the flood discharges of the Blackstone River in Woonsocket, Rhode Island, in each of the years from 1956 to 1965.

Year	Flood Discharge (ft^3/s)
1956	6035
1957	6804
1958	5018
1959	5862
1960	8064
1961	8095
1962	4326
1963	4703
1964	3849
1965	5576

Take the yearly flood discharge as a random variable X. Use the above data to estimate the cdf $F_X(5000)$ using empirical distribution function.

Solution. From the definition,

$$\hat{F}_X(5000) = \frac{1}{10}\sum_{i=1}^{10} \mathbf{1}(X_i \leq 5000) = 0.3.$$

Question 4. In the context of Question 3, estimate the density at 5000, i.e. $f_X(5000)$, using histogram method with bin edges $\{0, 2000, 3000, 4000, 4500, 5000, 5500, 6000, 10000\}$.

Solution. From the definition,

$$\hat{f}_X^{his}(5000) = \frac{C_j}{10(b_j - b_{j-1})}$$

where $b_{j-1} = 5000$ and $b_j = 5500$. Then $C_j = \sum_{i=1}^{10} \mathbf{1}(5000 \leq X_i < 5500) = 1$ and $\hat{f}_X^{his}(5000) = \frac{1}{10 \cdot 500} = 0.0002$.

Question 5. In the context of Question 3, estimate the density at 5000 using kernel density approach. Suppose we choose the uniform kernel and $h = 300$.

Solution. We first compute $K(\frac{u-X_i}{h})$ as follows

$$\begin{aligned}
K((5000-6035)/300) &= 0\\
K((5000-6804)/300) &= 0\\
K((5000-5018)/300) &= 0.5\\
K((5000-5862)/300) &= 0\\
K((5000-8064)/300) &= 0\\
K((5000-8095)/300) &= 0\\
K((5000-4326)/300) &= 0\\
K((5000-4703)/300) &= 0.5\\
K((5000-3849)/300) &= 0\\
K((5000-5576)/300) &= 0
\end{aligned}$$

So the kernel density estimate is

$$\hat{f}_X^{k}(5000) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{u - X_i}{h}) = \frac{1}{10 \cdot 300} = 0.000\overline{33}.$$