Homework 7

Exercise 1

Let $X_1, X_2, ..., X_n$ be a random sample from the distribution with probability density function

$$f(x) = \frac{2^\alpha \alpha}{x^{\alpha+1}}, \quad x > 2 \quad , \alpha > 0$$

a. (1 pt) Find the Maximum Likelihood Estimator (MLE) for α and show all work.

b. (1 pt) Find the Method of Moments (MoM) estimator for α. You may use the information that this is a Pareto Distribution, with an expected value of $\frac{2\beta}{\beta-1}$. *for an extra challenge, try solving for the expected value and verify that you get the same expression :)*

c. (0.5 pt) Pareto distributions are heavily right skewed distributions, and are commonly used to model variables with a high density on the left end. Suppose we took a sample of 6 from this distribution and got the following values: {2.0, 5.2, 2.7, 2.2, 2.1, 2.7} $n=6$

Using this sample, calculate the Maximum Likelihood (ML) Estimate. $\hat{\alpha}$

hint: In R, you can use vectorized operations in R by defining a vector, and then using that named vector inside an expression.

d. (0.5 pt) Now consider that same sample of 6 observations from above. Using this sample, calculate the MoM estimate.

a). I. $L(\alpha) = \prod_{i=1}^{n} f(x_i) = \frac{\alpha^n \cdot 2^{n\alpha}}{\prod x_i^{\alpha+1}}$

II. $\ln[L(\alpha)] = \ln \alpha^n \cdot 2^{n\alpha} - \ln \prod x_i^{\alpha+1} = n \ln\alpha + n\alpha \ln 2 - \ln \prod x_i^{\alpha+1} = n \cdot \ln\alpha + n\alpha \ln 2 - (\alpha+1) \sum_{i=1}^{n} \ln x_i$

IV. $\frac{\partial \ln[L(\alpha)]}{\partial \alpha} = \frac{n}{\alpha} + n \ln 2 - \sum \ln x_i = 0$

$\therefore \boxed{\hat{\alpha} = \frac{n}{\sum \ln(x_i) - n \ln 2}}$

b). $E[X] = \int_2^\infty x \cdot f(x)\, dx = \int_2^\infty x \cdot \frac{2^\alpha \cdot \alpha}{x^{\alpha+1}} dx = \int_2^\infty \frac{2^\alpha \cdot \alpha}{x^\alpha} dx = 2^\alpha \cdot \alpha \cdot \frac{x^{1-\alpha}}{1-\alpha} \Big|_2^\infty = \begin{cases} \infty & \alpha < 1 \\ \frac{2\alpha}{1-\alpha} & \alpha > 1 \end{cases}$

Let $E[X] = \bar{X}$

$\Downarrow$

$\bar{X} = \frac{2\alpha}{\alpha - 1} \Rightarrow \boxed{\tilde{\alpha} = \frac{\bar{X}}{\bar{X} - 2}}$

c). $\alpha = \frac{6}{\ln(2 \times 5.2 \times 2.7 \times 2.2 \times 2.1 \times 2.7 - 6 \ln 2} = \boxed{3.53}$

d). $\alpha = \frac{\bar{X}}{\bar{X} - 2} = \boxed{3.449}$

Exercise 2 ✓

Let $X_1, X_2, ..., X_n$ be a random sample from a a Gamma(α, θ) distribution. That is,

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, \quad 0 < x < \infty, \quad \alpha > 0, \ \theta > 0$$

Suppose α is known (you can treat it as a constant).

a. (1 pt) Obtain the Maximum likelihood estimator of θ, $\hat{\theta}$ and show your work *(work should include the log likelihood function and derivative expression)*

b. (1 pt) Find $Var(\hat{\theta})$. Your final answer should be in terms of the parameters of your distribution.

hint: First, check your continuous distributions table to find the variance of X when X follows a gamma distribution

c. (1 pt) Calculate the $MSE(\hat{\theta})$

a).
$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \left[\frac{1}{\Gamma(\alpha)}\right]^n \cdot \theta^{-n\alpha} \cdot \prod_{i=1}^{n} (x_i^{\alpha-1}) \cdot e^{-\frac{x_i}{\theta}}$$

$$\downarrow$$

$$\ln[L(\theta)] = n \ln\left[\frac{1}{\Gamma(\alpha)}\right] - n\alpha \cdot \ln\theta + \sum \ln\left[(x_i)^{\alpha-1} \cdot e^{-\frac{x_i}{\theta}}\right]$$

$$\downarrow \quad = n\ln\left[\frac{1}{\Gamma(\alpha)}\right] - n\alpha \cdot \ln\theta + (\alpha-1) \cdot \sum \ln(x_i) - \frac{1}{\theta} \cdot \sum x_i$$

$$\frac{\partial \ln[L(\theta)]}{\partial\theta} = 0 - \frac{n\alpha}{\theta} + 0 + \frac{1}{\theta^2} \sum x_i \overset{set}{=} 0$$

$$\downarrow$$

$$\boxed{\hat{\theta} = \frac{\bar{X}}{\alpha}}$$

b).
$$Var[\hat{\theta}] = Var\left[\frac{\bar{X}}{\alpha}\right] = \frac{1}{\alpha^2} \cdot \overset{iid}{Var[\bar{x}]} = \frac{1}{\alpha^2} \cdot \frac{\sigma_X^2}{n} = \frac{1}{\alpha^2} \cdot \frac{\alpha \cdot \theta^2}{n} = \boxed{\frac{\theta^2}{n \cdot \alpha}}$$

伽马分布 Gamma Distribution Γ

概念：第n[th]次事件发生需要多少时间

λ 是单位时间内事件发生次数

连续型随机变量

$X \sim Gamma(\alpha, \theta)$

$\Gamma(t) = \int_0^{\infty} y^{t-1} e^{-y}\, dy, \quad 0 < t$

$\Gamma(1) = 1$

$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \cdot x^{\alpha-1} \cdot e^{-\frac{x}{\theta}}, \quad x \geq 0$

$f(x) = \frac{1}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot \lambda^{\alpha} \cdot e^{-\lambda x}$

$E[X] = \alpha\theta$

$Var[X] = \alpha\theta^2$

c).
$$\left.\begin{aligned} MSE[\hat{\theta}] &= Var[\hat{\theta}] + Bias^2[\hat{\theta}] \\ Bias[\hat{\theta}] &= E[\hat{\theta}] - \theta \end{aligned}\right\} \Rightarrow MSE[\hat{\theta}] = Var[\hat{\theta}] + \left(E[\hat{\theta}] - \theta\right)^2$$

$$E[\hat{\theta}] = \overset{iid}{E\left[\frac{\bar{X}}{\alpha}\right]} = \frac{1}{\alpha} \cdot E[X] = \frac{1}{\alpha} \cdot \alpha \cdot \theta = \theta \Rightarrow Bias[\hat{\theta}] = 0$$

$$\Rightarrow \boxed{MSE[\hat{\theta}] = \frac{\theta^2}{n \cdot \alpha}}$$

Exercise 3 ✓

(1.5 pt) Consider the following statistic as a possible estimator of σ^2.

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

When X is normally distributed, this is the MLE for X. However, it's a biased statistic. Calculate the bias of this statistic as an estimator for σ^2, and show your work.

$$\text{Bias}[\hat{\sigma}^2] = E[\hat{\sigma}^2] - \sigma^2$$

$$= E\left[\frac{1}{n}\sum(x_i - \bar{x})^2\right] - \sigma^2 = \frac{1}{n}E\left[\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2)\right] - \sigma^2$$

$$= \frac{1}{n}E\left[\sum x_i^2 - 2\sum x_i\bar{x} + \sum\bar{x}^2\right] - \sigma^2$$

$$= \frac{1}{n}E\left[\sum x_i^2 - 2\bar{x}\cdot\sum x_i + n\bar{x}^2\right] - \sigma^2$$

$$= \frac{1}{n}E\left[\sum x_i^2 - 2\bar{x}\cdot n\cdot\bar{x} + n\bar{x}^2\right] - \sigma^2 = \frac{1}{n}E\left[\sum x_i^2 - n\bar{x}^2\right] - \sigma^2$$

$$= \frac{1}{n}E\left[\sum x_i^2 - \sum\bar{x}^2\right] - \sigma^2$$

$$= \frac{1}{n}E\left[\sum(x_i^2 - \bar{x}^2)\right] - \sigma^2$$

$$= \frac{1}{n}\left[E[\sum x_i^2] - E[\sum\bar{x}^2]\right] - \sigma^2$$

$$= \frac{1}{n}\left[n\cdot E[x_i^2] - E[n\bar{x}^2]\right] - \sigma^2$$

$$= E[x_i^2] - E[\bar{x}^2] - \sigma^2$$

$$= E[x^2] - E[\bar{x}^2] - \sigma_x^2$$

$$\because \sigma_x^2 = E[x^2] - (E[x])^2$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \qquad \mu_{\bar{x}} = \mu_x$$

$$\therefore \text{Original} = \cancel{\sigma_x^2} + \cancel{\mu_x^2} - \left[\frac{\sigma_x^2}{n} + \cancel{\mu_x^2}\right] - \cancel{\sigma_x^2}$$

$$\boxed{= -\frac{\sigma_x^2}{n} \neq 0}$$

Exercise 4 ✓

Let $X_1, X_2, ..., X_n \overset{i.i.d}{\sim} f(x)$, where

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

a. (1 pt) Find an expression for the maximum likelihood estimator of $\theta, \hat{\theta}$, and show all your work. MLE

b. (0.5 pt) Suppose we obtain the following sample of size $n = 5$ from this distribution: $\{4, 2.5, 6, 5, 3\}$.

Use these data points to calculate a value of the maximum likelihood **estimate** of θ.

a). $L(\theta) = \prod_{i=1}^{n} f(x_i) = \theta^{-n} \prod x_i \cdot e^{-\frac{1}{2\theta} \cdot \sum x_i^2}$

$\downarrow$

$\ln[L(\theta)] = -n \cdot \ln\theta + \sum \ln x_i - \frac{1}{2\theta} \cdot \sum x_i^2$

$\downarrow$

$\frac{\partial \ln[L(\theta)]}{\partial\theta} = -\frac{n}{\theta} + 0 + \frac{1}{2\theta^2} \cdot \sum x_i^2 \overset{\text{set}}{=} 0$

$\downarrow$

$\boxed{\hat{\theta} = \frac{1}{2n} \cdot \sum x_i^2}$

b). $\theta = \frac{1}{10} \cdot [16 + 6.25 + 36 + 25 + 9] = \boxed{9.225}$

Exercise 5 (Use R)

(1 pt) Using the 5 data points from the previous exercise, use R to plot the likelihood function as a function of θ. Show your code and plot. $L(\theta)$

- x-axis: θ from 0 to 12 in intervals of 0.5
- y-axis: likelihood, Either the Likelihood function $L(\theta)$ or the log likelihood function $ln[L(\theta)]$)]

Hint: Make a sequence for 'theta' in R. Then make a function or expression in terms of theta and run that. Be sure both vectors are assigned to a name, and plot these two named vectors.

```r
rm(list=ls()) #remove all data

x=c(4,2.6,6,5,3)
n=length(x)

L=function(theta){
  prod(x)*theta^(-n)*exp(-1/2/theta*sum(x^2))
}

theta=seq(from=0,to=12,by=0.5)

windows()

plot(theta,L(theta))
```