STAT 323 Assignment 3

Instructor: Claudia Marie Mahler
Name: Yifeng Pan
UCID: 30063828

Summer 2019

1 Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a Poisson distribution with mean λ. Consider $\hat{\lambda}_1 = \frac{X_1+X_2}{2}$ and $\hat{\lambda}_2 = \overline{X}$. Find $RE(\hat{\lambda}_1, \hat{\lambda}_2)$ for $n = 25$ and interpret the meaning of the RE in the context of this question.

B$(\hat{\lambda}_1)$ = E$(\hat{\lambda}_1) - \lambda$ = E$(\frac{X_1+X_2}{2}) - \lambda = \frac{\mathsf{E}(X_1)+\mathsf{E}(X_2)}{2} - \lambda = 2\lambda/2 - \lambda = 0$
B$(\hat{\lambda}_2)$ = E$(\hat{\lambda}_2) - \lambda$ = E$(\overline{X}) - \lambda = \lambda - \lambda = 0$
So they are both unbised.
Var$(\hat{\lambda}_1)$ = Var$(\frac{X_1+X_2}{2})$ = Var$(X_1 + X_2)/4 = \frac{\mathsf{Var}X_1+\mathsf{Var}X_2+\mathsf{Cov}(X_1,X_2)}{4} = \frac{\lambda+\lambda+0}{4} = \frac{\lambda}{2}$
Var$(\hat{\lambda}_2)$ = Var$(\overline{X})$ = Var$(\sum_{i=1}^n X_i)/n^2 = n\lambda/n^2 = \lambda/n$
RE$(\hat{\lambda}_1, \hat{\lambda}_2)$ = Var$(\hat{\lambda}_1)$/Var$(\hat{\lambda}_1) = \frac{n\lambda}{2\lambda} = n/2$
For $n = 25$, RE$(\hat{\lambda}_1, \hat{\lambda}_2) = 25/2$. This means $\hat{\lambda}_2$ is a much better estimator then $\hat{\lambda}_1$, by a factor of $25/2$ times.

2 The Rayleigh density function is given by

$$f(y) = \begin{cases} (\frac{2y}{\theta})e^{\frac{-y^2}{\theta}} & \textbf{for } y > 0 \\ 0 & \textbf{elsewhere} \end{cases}$$

The quantity Y^2 has an exponential distribution with mean θ. If $Y_1, Y_2, \ldots, Y_n$ denotes a random sample from a Rayleigh distribution, show that $W_n = \frac{1}{n}\sum_{i=1}^n Y_i^2$ is a consistent estimator for θ.

Proof that $\lim_{n\to\infty} P(|W_n - \theta| \le \epsilon) = 1, \forall \epsilon \in \mathbb{R}^+$:
Let $Z = Y^2$ be the exponential distribution with mean θ.

$$\begin{aligned}
P(|W_n - \theta| \le \epsilon) &= P(-\epsilon \le W_n - \theta \le \epsilon) = P(\theta - \epsilon \le W_n \le \theta + \epsilon) \\
&= P(\theta - \epsilon \le \frac{1}{n}\sum_{i=1}^n Y_i^2 \le \theta + \epsilon) = P(\theta - \epsilon \le \overline{z} \le \theta + \epsilon) \\
&= P(\frac{\theta - \epsilon - \mu_{\overline{z}}}{\sigma_{\overline{z}}} \le \frac{\overline{z} - \mu_{\overline{z}}}{\sigma_{\overline{z}}} \le \frac{\theta + \epsilon - \mu_{\overline{z}}}{\sigma_{\overline{z}}}) \text{ Using CLT, because } n \to \infty. \\
&= P(\frac{\theta - \epsilon - \theta}{\theta/\sqrt{n}} \le \frac{\overline{z} - \mu_{\overline{z}}}{\sigma_{\overline{z}}} \le \frac{\theta + \epsilon - \theta}{\theta/\sqrt{n}}) \\
&= P(\frac{-\epsilon\sqrt{n}}{\theta} \le \frac{\overline{z} - \mu_{\overline{z}}}{\sigma_{\overline{z}}} \le \frac{\epsilon\sqrt{n}}{\theta}) \\
&= pnorm(\frac{\epsilon\sqrt{n}}{\theta}) - pnorm(\frac{-\epsilon\sqrt{n}}{\theta})
\end{aligned}$$

It's easy to see that $\lim_{n\to\infty}(pnorm(\frac{\epsilon\sqrt{n}}{\theta}) - pnorm(\frac{-\epsilon\sqrt{n}}{\theta})) = 1$.
As ϵ and θ are fixed finite positive values, and $\sqrt{n}$ is tends to ∞.
Therefore W_n is a consistent estimator of θ.

3 Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a Pareto distribution. $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$ has the cumulative distribution function given by:

$$F_{(1)}(x) = \begin{cases} 1 - (\frac{\beta}{x})^{an} & \textbf{for } x > \beta \\ 0 & \textbf{for } x \leq \beta \end{cases}$$

Show that $X_{(1)}$ is a consistent estimator of β (hint: use the method described in Exercise 9.26).

Proof that $\lim_{n\to\infty} P(\left|X_{(1)} - \beta\right| \leq \epsilon) = 1, \forall \epsilon \in \mathbb{R}^+$:

$$\begin{aligned} P(\left|X_{(1)} - \beta\right| \leq \epsilon) = P(-\epsilon \leq X_{(1)} - \beta \leq \epsilon) &= P(\beta - \epsilon \leq X_{(1)} \leq \beta + \epsilon) \\ &= F(\beta + \epsilon) - F(\beta - \epsilon) \\ &= (1 - (\frac{\beta}{\beta + \epsilon})^{an}) - F(\beta - \epsilon) \\ &= (1 - (\frac{\beta}{\beta + \epsilon})^{an}) - 0 \end{aligned}$$

$$\begin{aligned} \lim_{n\to\infty} P(\left|X_{(1)} - \beta\right| \leq \epsilon) &= \lim_{n\to\infty} (1 - (\frac{\beta}{\beta + \epsilon})^{an}) \\ &= 1 - \lim_{n\to\infty} (\frac{\beta}{\beta + \epsilon})^{an} \\ &= 1 - 0 \text{ assuming } a \text{ is positive, and we know } \epsilon \text{ is positive.} \\ &= 1 \end{aligned}$$

Therefore $X_{(1)}$ is a consistent estimator of β.

4 Consider the density function

$$f(x) = \begin{cases} e^{\beta - x} & \textbf{for } x > \beta \\ 0 & \textbf{elsewhere} \end{cases}$$

Determine the method of moments estimator for β.

First theoretical moment: $\mathsf{E}(X) = \int_\beta^\infty x e^{\beta - x} dx = \beta + 1$
First sample moment: $\overline{x}$.
$\boxed{\hat{\beta}_{mm} = \overline{x} - 1}$.

5 A certain type of electronic component has a lifetime Y (in hours) with probability density function given by

$$f(y \mid \theta) = \begin{cases} (\frac{1}{\theta^2}) y e^{\frac{-y}{\theta}} & \textbf{for } y > 0 \\ 0 & \textbf{elsewhere} \end{cases}$$

That is, Y has a gamma distribution with parameters $\alpha = 2$ and θ. Let $\hat{\theta}$ denote the MLE of θ. Find the MLE of θ based on three such componets (tested independently) with lifetimes of 120, 130, and 128 hours of lifetime.

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} (\frac{1}{\theta^2}) y_i e^{\frac{-y_i}{\theta}} \\
\ln(L(\theta)) &= \ln(\prod_{i=1}^{n} (\frac{1}{\theta^2}) y_i e^{\frac{-y_i}{\theta}}) = \sum_{i=1}^{n} \ln((\frac{1}{\theta^2}) y_i e^{\frac{-y_i}{\theta}}) \\
&= \sum_{i=1}^{n} \ln(\frac{1}{\theta^2}) + \ln(y_i e^{\frac{-y_i}{\theta}}) = n \ln(\frac{1}{\theta^2}) + \sum_{i=1}^{n} \ln(y_i e^{\frac{-y_i}{\theta}}) \\
&= n \ln(\frac{1}{\theta^2}) + \sum_{i=1}^{n} \ln(y_i) + \frac{-y_i}{\theta} \ln(e) = n \ln(\frac{1}{\theta^2}) + \sum_{i=1}^{n} \ln(y_i) + \frac{-y_i}{\theta} \\
&= n \ln(\frac{1}{\theta^2}) + (\sum_{i=1}^{n} \ln(y_i) + \sum_{i=1}^{n} \frac{-y_i}{\theta}) \\
&= n \ln(\frac{1}{\theta^2}) + (\sum_{i=1}^{n} \ln(y_i) + \frac{-1}{\theta} \sum_{i=1}^{n} (y_i)) = n \ln(\frac{1}{\theta^2}) + \frac{-n\overline{y}}{\theta} + \sum_{i=1}^{n} \ln(y_i) \\
\frac{\partial}{\partial \theta}(\ln(L(\theta))) &= \frac{\partial}{\partial \theta}(n \ln(\frac{1}{\theta^2}) + \frac{-n\overline{y}}{\theta} + \sum_{i=1}^{n} \ln(y_i)) \\
&= \frac{\partial}{\partial \theta}(n \ln(\frac{1}{\theta^2})) + \frac{\partial}{\partial \theta}(\frac{-n\overline{y}}{\theta}) + \frac{\partial}{\partial \theta}(\sum_{i=1}^{n} \ln(y_i)) = \frac{-2n}{\theta} + \frac{n\overline{y}}{\theta^2} + 0
\end{aligned}$$

Solve for θ in $\frac{-2n}{\theta} + \frac{n\overline{y}}{\theta^2} = 0$:

$$\begin{aligned}
\frac{2n}{\theta} &= \frac{n\overline{y}}{\theta^2} \\
\theta^2 2n &= n\overline{y}\theta \\
\theta &= \overline{y}/2
\end{aligned}$$

Second derivative test:

$$\begin{aligned}
\frac{\partial^2}{\partial \theta^2}(\ln(L(\overline{y}/2))) &= \frac{\partial}{\partial \theta}(\frac{-2n}{\theta} + \frac{n\overline{y}}{\theta^2})\Big|^{\overline{y}/2} \\
&= \frac{2n}{\theta^2} - \frac{2n\overline{y}}{\theta^3}\Big|^{\overline{y}/2} = \frac{2^3 n}{\overline{y}^2} - \frac{2^4 n\overline{y}}{\overline{y}^3} \\
&= -\frac{2^3 n}{\overline{y}^2} < 0
\end{aligned}$$

therefore $\hat{\theta}_{ML}$ is a maximum.

$\hat{\theta}_{ML} = \overline{y}/2$. $\overline{y} =$ mean$(120, 130, 128) = 126$. Therefore $\boxed{\hat{\theta}_{ML} = 126/2 = 63}$.

6 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} (\theta+1)y^{\theta} & \textbf{for } 0 < y < 1, \theta > -1 \\ 0 & \textbf{elsewhere} \end{cases}$$

Find the MLE of θ.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n}(\theta+1)y_i^{\theta} \\ \ln(L(\theta)) &= \ln(\prod_{i=1}^{n}(\theta+1)y_i^{\theta}) = \sum_{i=1}^{n}\ln((\theta+1)y_i^{\theta}) = \sum_{i=1}^{n}\ln(\theta+1) + \sum_{i=1}^{n}\ln(y_i^{\theta}) \\ &= n\ln(\theta+1) + \theta\sum_{i=1}^{n}\ln(y_i) \\ \frac{\partial}{\partial\theta}(\ln(L(\theta))) &= \frac{\partial}{\partial\theta}(n\ln(\theta+1) + \theta\sum_{i=1}^{n}\ln(y_i)) = \frac{n}{\theta+1} + \sum_{i=1}^{n}\ln(y_i) \end{aligned}$$

Solve for θ in $\frac{n}{\theta+1} + \sum_{i=1}^{n}\ln(y_i) = 0$:

$$\begin{aligned} -\frac{n}{\theta+1} &= \sum_{i=1}^{n}\ln(y_i) \\ -\frac{n}{\sum_{i=1}^{n}\ln(y_i)} - 1 &= \theta \end{aligned}$$

Second derivative test:

$$\begin{aligned} \frac{\partial^2}{\partial\theta^2}(\ln(L(-\frac{n}{\sum_{i=1}^{n}\ln(y_i)} - 1))) = &\ \frac{\partial}{\partial\theta}(\frac{n}{\theta+1} + \sum_{i=1}^{n}\ln(y_i))\Big|^{-\frac{n}{\sum_{i=1}^{n}\ln(y_i)}-1} \\ = &\ \frac{-n}{(\theta+1)^2}\Big|^{-\frac{n}{\sum_{i=1}^{n}\ln(y_i)}-1} \\ = &\ \frac{-n}{(-\sum_{i=1}^{n}\ln(y_i) - 1 + 1)^2} < 0 \end{aligned}$$

therefore $\hat{\theta}_{ML}$ is a maximum.

Therefore: $\boxed{\hat{\theta}_{ML} = -(\frac{n}{\sum_{i=1}^{n}\ln(y_i)} + 1)}$. Note: $\ln(y)$ is always negative.

7 The hourly wages in a particular industry are normally distributed with a mean of $\$13.20$ and a standard deviation of $\$2.50$. A company in this industry employs 40 workers, paying them an average of $\$12.20$ per hour. Can this company be accused of paying substandard wages? Use an $\alpha = 0.01$ level hypothesis test to test this claim.

Yes they can, accusations don't require proof.
Is there statistical evidence to back up the accusation?:
$n = 40, df = 39, \mu = 13.2, \sigma = 2.5, \overline{x} = 12.2, \alpha = 0.01$
$z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} = \frac{12.2-13.2}{2.5/\sqrt{40}} = -\sqrt{40}/2.5$
$H_0 : \mu \geq 13.3, H_a : \mu < 13.3.$
Therefore p-value $= pnorm(-sqrt(42)/2.5) \approx 0.48\% < \alpha = 1\%.$
Therefore the null hypothesis is rejected. And there is evidence to suggest that the company is paying wages below the industry mean.

8 A study by Children's Hospital in Boston indicates that about 67% of American adults and about 15% of American children and adolescents are overweight. Thirteen children in a random sample of 100 were found to be overweight. Based on this sample, is there sufficient evidence to indicate that the percentage reported by Children's Hospital is too high? Carry out an $\alpha = 0.05$ level hypothesis test.

$H_0 : p \geq 0.15, H_a : p < 0.15, p = 0.15, \overline{x} = 13/100, n = 100, \alpha = 0.05$
p-value $= pbinom(13, size = 100, prob = 0.15) \approx 34.74\% > \alpha = 5\%$
Not sufficient evidence for alternative hypothesis.

9 A single observation is randomly selected from an exponentially distributed population. The value of the observation is used to test the hypothesis that the mean of the population is 2 against the alternative hypothesis that the mean is 5. The null hypothesis is not rejected if and only if the observed value is less than 3. Find the probabilities of committing Type I and Type II error and interpret these probabilities.

Type 1: $P(R\ H_0|H_0) = \alpha.$
$\alpha = P(\overline{x} \geq 3|\mu = 2) = 1 - pexp(3, 1/2) \approx 22.31\%.$
Type 2: $P(R\ H_a|H_a) = \beta.$
$\beta = P(\overline{x} < 3|\mu = 5) = pexp(3, 1/5) \approx 45.12\%.$

10 An Ipsos-Reid poll in 2005 revealed that the mean amount of time internet-using Canadians spend online for personal reasons (that is, not work-related) was 12.7 (hours) per week. A recent random sample of $n = 85$ internet-using Canadians was taken. Each person was asked the following question: "In a typical week, approximately how many hours are you using the internet for personal reasons?" The data can be found in the file InternetUse.R on D2L. Does this sample suggest that, on average, Canadians spend more time online now compared to 2005? Conduct a hypothesis test using $\alpha = 0.05$.

$n = 85, \overline{x} \approx 17.75, s \approx 16.32, \mu = 12.7, \alpha = 0.05$
$H_0 : \mu \leq 12.7$
$H_a : \mu > 12.7$
$1 - \alpha$ confidence interval is $(\overline{x} - qt(\alpha, n-1) * s/\sqrt{n}, \infty) \approx (14.80916, \infty).$
Because $(14.80916, \infty)$ does not contain 12.7, the null hyposis is rejected.