EEE-485

Problem Set 1

Kerem Ayöz
21501569

Q1

a.) $\hat{\lambda}_{MLE} = \underset{\lambda}{\arg\max}\, I(\lambda) \Rightarrow L(\lambda) = p(D|\lambda) = \prod_{i=1}^{n} p(y_i|x_i, \lambda)$

$$I(\lambda) = \sum_{i=1}^{n} \ln(p(y_i|x_i,\lambda)) = \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right) = \sum_{i=1}^{n} -\lambda + x_i\ln(\lambda) - \ln(x_i!)$$

$$I'(\lambda) = \sum_{i=1}^{n} -1 + \frac{x_i}{\lambda} = 0 \Rightarrow -n + \frac{\bar{x}n}{\lambda} = 0 \Rightarrow \boxed{\hat{\lambda}_{MLE} = \bar{x}}$$

b.) $y_i = \begin{cases} 0 & \text{if } x_i = 0 \\ 1 & \text{if } x_i > 0 \end{cases}$, $P(x_i > 0) = 1 - P(x_i = 0) = 1 - \frac{e^{-\lambda}\lambda^0}{1} = 1 - e^{-\lambda}$

$\Rightarrow y_i \sim Ber(p = 1 - e^{-\lambda})$, $f(y_i) = p^{y_i}(1-p)^{1-y_i}$

$L(p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$, $I(p) = \sum_{i=1}^{n} y_i \ln p + (1-y_i)\ln(1-p)$

$$I'(p) = \sum_{i=1}^{n} \frac{y_i}{p} + \frac{1-y_i}{p-1} = \frac{\bar{y}n}{p} + \frac{n - \bar{y}n}{p-1} = 0 \Rightarrow \hat{p} = \bar{y}$$

Then $\bar{y} = 1 - e^{-\hat{\lambda}} \Rightarrow \boxed{\hat{\lambda}_{MLE} = -\ln(1-\bar{y})}$

c.) $L(\lambda) = \prod_{i=1}^{n} \frac{e^{e^{-\lambda}}(e^{-\lambda})^{x_i}}{x_i!}$, $I(\lambda) = \sum_{i=1}^{n} e^{-\lambda} - \lambda x_i - \ln x_i!$

$$I'(\lambda) = \sum_{i=1}^{n} -e^{-\lambda} - x_i = -ne^{-\lambda} - \bar{x}n = 0 \Rightarrow \boxed{\hat{\lambda}_{MLE} = -\ln(-\bar{x})}$$

Q2

a.) To calculate the likelihood, define $\underline{k} = (k_1, k_2, \dots k_6)$ where k_i represents the # of outcomes where die comes i.

$$L(\theta) = \theta_1^{k_1} \cdot \theta_2^{k_2} \cdots \theta_6^{k_6} = \theta_1^{k_1} (1-\theta_1)^{n-k_1}$$

$$l(\theta) = k_i \ln\theta_i + (n-k_i)\ln(1-\theta_i)$$

$$\frac{\partial l(\theta)}{\partial \theta_i} = \frac{k_i}{\theta_i} + \frac{n-k_i}{\theta_i - 1} = 0 \Rightarrow \hat{\theta}_i = \frac{k_i}{n}$$

Then $\hat{\theta} = \left(\frac{k_1}{n}, \frac{k_2}{n}, \dots, \frac{k_6}{n}\right) = \frac{1}{n}(k_1, k_2, \dots k_6) = \boxed{\frac{\underline{k}}{n}}$

b.) posterior $\propto$ likelihood $\times$ prior , $\underline{\alpha} = (\alpha_1, \alpha_2, \dots \alpha_6)$

$$f(\theta) \propto L(\theta) \times p(\theta)$$

$$f(\theta) \cong \theta_1^{k_1} \cdot \theta_2^{k_2} \cdots \theta_6^{k_6} \cdot \frac{\theta_1^{\alpha_1 - 1} \cdots \theta_6^{\alpha_6 - 1}}{A(\theta_1 \dots \theta_6)}$$

$$f(\theta) \cong \theta_1^{k_1+\alpha_1-1} \cdot \theta_2^{k_2+\alpha_2-1} \cdots \theta_6^{k_6+\alpha_6-1} \quad (A \text{ is independent})$$

$$\ln f(\theta) \cong (k_1+\alpha_1-1)\ln\theta_1 + \dots + (k_6+\alpha_6-1)\ln\theta_6$$

$$\ln f(\theta) \cong (k_i+\alpha_i-1)\ln\theta_i + \left(n-k_i-5+\sum_{j\neq i}\alpha_j\right)\ln(1-\theta_i)$$

$$\frac{\partial \ln f(\theta)}{\partial \theta_i} \cong \frac{k_i+\alpha_i-1}{\theta_i} + \frac{n-k_i-5+\sum_{j\neq i}\alpha_j}{\theta_i - 1} = 0$$

$$-\cancel{\theta_i k_i} - \theta_i\alpha_i + \theta_i + k_i + \alpha_i - 1 = \theta_i n - \cancel{\theta_i k_i} - 5\theta_i + \theta_i \sum_{j\neq i}\alpha_j$$

$$\theta_i\left(\sum_{j=1}^{6}\alpha_j + n - 6\right) = \alpha_i + k_i - 1$$

$$\Rightarrow \hat{\theta}_{i_{MAP}} = \frac{\alpha_i + k_i - 1}{\left(\sum_{j=1}^{6}\alpha_j\right) + n - 6}$$

$$\boxed{\hat{\theta}_{MLE} = \frac{1}{S_\alpha + n - 6}(\underline{\alpha} + \underline{k} - \underline{1}) \text{ where } \underline{\alpha}, \underline{k}, \underline{1} \text{ are all vectors, where } S_\alpha = \sum_{i=1}^{6}\alpha_i}$$

Q3

a.)

Sample / μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n=100	96	89	88	95	92	95	94	95	98
n=1000	94	95	98	96	95	91	95	94	95
n=10000	97	96	95	92	98	95	98	91	97

where $X \sim N(\mu, 1)$ $\hat{\mu} = \begin{cases} 1.0754 & \text{when } n=100 \\ 1.0458 & \text{when } n=1000 \\ 0.9918 & \text{when } n=10000 \end{cases}$

Theory suggests that %95 of the estimations should be in interval, by looking to the results above, we could say that experiment is consistent with theory.

b.) Yes, because of the law of large numbers, the estimated mean converges to the true mean which is 1.

c.) $X \sim N(\mu, 0.1) \Rightarrow \hat{\mu} = \begin{cases} 1.0172, & n=100 \\ 1.0250, & n=1000 \\ 1.0159, & n=10000 \end{cases}$

$X \sim N(\mu, 1) \Rightarrow \hat{\mu} = \begin{cases} 1.0754, & n=100 \\ 1.0458, & n=1000 \\ 0.9918, & n=10000 \end{cases}$

$X \sim N(\mu, 10) \Rightarrow \hat{\mu} = \begin{cases} 1.1933, & n=100 \\ 0.9909, & n=1000 \\ 1.0044, & n=10000 \end{cases}$

Q4

a.) $y_i = a x_i + \varepsilon_i \qquad RSS(a) = \sum_{i=1}^{n} (y_i - a x_i)^2$

$$\hat{a} = \underset{a}{\arg\min}\, RSS(a) \implies RSS'(a) = -2 \sum_{i=1}^{n} x_i (y_i - a x_i) = 0$$

$$\implies \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} a x_i \implies \boxed{\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}$$

b.) $y_i = a^{true} x_i + \varepsilon_i$ where a^{true} and x_i are fixed. Since $\varepsilon_i \sim N(0, \sigma^2)$ y_i is also normally distributed. $y_i \sim N(a^{true} x_i, \sigma^2)$.

- $$E[\hat{a}] = E\left[\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\right] = E\left[\frac{\sum_{i=1}^{n} x_i (a^{true} x_i + \varepsilon_i)}{\sum_{i=1}^{n} x_i^2}\right] = E\left[\frac{\sum_{i=1}^{n} a^{true} x_i^2 + \sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}\right]$$

$$E[\hat{a}] = a^{true} + E\left[\frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}\right] = \boxed{a^{true}}$$

- $$Var(\hat{a}) = Var\left(\frac{\sum_{i=1}^{n} a^{true} x_i^2 + x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}\right) = Var\left(a^{true} + \frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}\right) = \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2} Var\left(\sum_{i=1}^{n} x_i \varepsilon_i\right)$$

$$Var(\hat{a}) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \boxed{\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}}$$

Then $\boxed{\hat{a} \sim N\left(a^{true}, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right)}$