Bios 661: 1 – 5; Bios 673: 2 – 6.

1. C&B 7.8

2. (C&B 7.7) Let $X_1, \ldots, X_n$ be a random sample from one of the two probability density functions, namely,

$$f(x|\theta = 0) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$f(x|\theta = 1) = \begin{cases} 1/(2\sqrt{x}), & \text{if } 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

   (a) Find the likelihood function of θ.

   (b) If $n = 10$ and $\sum_{i=1}^n \log(x_i) = -10.7$, find the maximum likelihood estimation (MLE) of θ.

3. C&B 7.10

4. (C&B 7.14) Let $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ be iid random 2-vectors. Assume Y_i and Z_i are independent and follow an exponential distribution with mean λ and μ, respectively, for each i.

   (a) Find MLE for (λ, μ).

   Solution: Since Y is independent of Z, we can find the MLE of λ and μ separately. The likelihood function of $Y = (Y_1, \ldots, Y_n)$ is

$$L(\lambda) = f_Y(y_1, \ldots, y_n|\lambda) = \lambda^{-n} \exp(-\sum_{i=1}^n y_i/\lambda).$$

   The the log-likelihood function is

$$\ell(\lambda) = -n \log \lambda - \sum_{i=1}^n y_i/\lambda.$$

   Taking the first derivative with respect to λ and set the function to 0, one has

$$-n\lambda^{-1} + \lambda^{-2} \sum_{i=1}^n y_i = 0.$$

   That leads to $\hat{\lambda} = \sum_{i=1}^n Y_i/n = \bar{Y}$. Taking the second derivative and replacing λ by $\hat{\lambda} = \bar{y}$, one has

$$n\bar{y}^{-2} - 2\bar{y}^{-3} n\bar{y} = -n\bar{y}^{-2} < 0.$$

One hence can claim that $\hat{\lambda} = \bar{Y}$ is the MLE. Similarly, $\hat{\mu} = \bar{Z}$.

(b) Suppose that we only observe $(X_1, \delta_1), \ldots, (X_n, \delta_n)$, where $X_i = \min(Y_i, Z_i)$ and

$$\delta_i = \begin{cases} 1, & \text{if} \quad X_i = Y_i \\ 0, & \text{if} \quad X_i = Z_i, \end{cases}$$

for $i = 1, \ldots, n$. Find the MLE of (λ, μ).

Solution: We have shown in the previous homework that the joint pdf of (X, δ) is

$$\lambda^{-\delta} \mu^{-(1-\delta)} e^{-(\lambda^{-1} + \mu^{-1})x}.$$

The likelihood function for λ and μ is

$$L(\lambda, \mu) = \lambda^{-\sum_{i=1}^n \delta_i} \mu^{-(n - \sum_{i=1}^n \delta_i)} \exp\left\{ -(\lambda^{-1} + \mu^{-1}) \sum_{i=1}^n x_i \right\},$$

and the log-likelihood function is

$$\ell(\lambda, \mu) = -\sum_{i=1}^n \delta_i \log \lambda - (n - \sum_{i=1}^n \delta_i) \log \mu - (\lambda^{-1} + \mu^{-1}) \sum_{i=1}^n x_i.$$

Therefore, taking the first derivation with respect to λ and setting it to zero, we have

$$-\sum_{i=1}^n \delta_i \lambda^{-1} + \lambda^{-2} \sum_{i=1}^n x_i = 0.$$

Then, $\hat{\lambda} = \sum_{i=1}^n X_i / \sum_{i=1}^n \delta_i$. Similarly, taking the first derivation with respect to μ and setting it to zero, we have

$$-(n - \sum_{i=1}^n \delta_i) \mu^{-1} + \mu^{-2} \sum_{i=1}^n x_i = 0.$$

Then, $\hat{\mu} = \sum_{i=1}^n X_i / (n - \sum_{i=1}^n \delta_i)$. Of course, we need to check the second derivative and the Jacobian matrix so to make sure they are global maximum. However, another way to find the MLE of λ and μ is to write the joint pdf of (X, δ) as

$$\left(\frac{\mu}{\mu + \lambda} \right)^{\delta} \left(\frac{\lambda}{\mu + \lambda} \right)^{1-\delta} (\lambda^{-1} + \mu^{-1}) e^{-(\lambda^{-1} + \mu^{-1})x},$$

then re-parameterize $\theta_1 = \mu/(\mu + \lambda)$ and $\theta_2 = (\lambda^{-1} + \mu^{-1})^{-1} = \lambda\mu/(\lambda + \mu)$. Since the joint pdf can be written in two functions and the range of the variables does not depend on one and another, we can claim X and δ are independent and maximize the likelihood separately. That is, the likelihood function for θ_1 and θ_2 are

$$\begin{aligned} L(\theta_1, \theta_2) &= \prod_{i=1}^{n} \theta_1^{\delta_i}(1-\theta_1)^{1-\delta_i} \prod_{i=1}^{n} \frac{1}{\theta_2} \exp\left(-\frac{x_i}{\theta_2}\right) \\ &= L_1(\theta_1|\boldsymbol{\delta}) L_2(\theta_2|\boldsymbol{x}). \end{aligned}$$

We can maximize $L_1(\theta_1|\boldsymbol{\delta})$ and $L_2(\theta_2|\boldsymbol{x})$ separately and get $\hat{\theta}_1 = n^{-1}\sum_{i=1}^{n}\delta_i = \bar{\delta}$, and $\hat{\theta}_2 = n^{-1}\sum_{i=1}^{n} X_i = \bar{X}$. We can also rewrite $\lambda = \theta_2/\theta_1$ and $\mu = \theta_2/(1-\theta_1)$. By invariance of MLE, one has $\hat{\lambda} = \bar{X}/\bar{\delta}$ and $\mu = \bar{X}/(1-\bar{\delta})$.

5. Suppose that n observations, $X_1, \ldots, X_n$, are taken from $N(\mu, 1)$ with an unknown μ. If one can only records the value when the observation is positive, find the MLE for μ.

   (a) One possible approach is that one can still observe a complete data on $Y_1, \ldots, Y_n$, where $Y_i = I(X_i > 0)$, even though one did not observe a complete data. What is the distribution of Y?

   Solution: Since Y_i is binary, it follows a binomial distribution with mean

   $$\theta = P(X_i > 0) = P(X_i - \mu > -\mu) = 1 - \Phi(-\mu).$$

   (b) Given the distribution you identified in (a), find the MLE of the parameter in the distribution.

   Solution: The MLE of θ is $\bar{Y}$.

   (c) We have learned that the MLE has a nice invariance property. Comment on how one can use the property to find the MLE of μ.

   Solution: By invariance property, the MLE of μ is $\hat{\mu} = -\Phi^{-1}(1-\bar{Y})$.

6. For a certain African village, available data strongly suggests that the expected number of new cases of AIDS developing in any particular year is directly proportional to the expected number of new AIDS cases that developed during the immediately preceding year. An important statistical goal is to estimate the value of this unknown proportionality constant θ ($\theta > 1$). To accomplish this goal, the following statistical model is to be used: For $j = 0, 1, \ldots, n$ consecutive years of data, let Y_j be the random variable denoting the number of new AIDS cases developing in year j.

Further, suppose that the $(n+1)$ random variables $Y_0, Y_1, \ldots, Y_n$ are such that the conditional distribution of Y_{j+1}, given $Y_k = y_k$ for $k = 0, 1, \ldots, j$, depends only on y_j and is Poisson with $E(Y_{j+1}|Y_j = y_j) = \theta y_j$, $j = 0, 1, \ldots, (n-1)$. Further, assume that the distribution of the random variable Y_0 is Poisson with $E(Y_0) = \theta$, where $\theta > 1$.

(a) Using all $(n+1)$ random variables $Y_0, Y_1, \ldots, Y_n$, develop an explicit expression for the MLE $\hat{\theta}$ of the unknown proportionality constant θ.

Solution: The likelihood function is

$$\begin{aligned} L(\theta|\boldsymbol{y}) &= f(y_0)f(y_1|y_0)f(y_2|y_0, y_1)\ldots f(y_n|y_0, y_1, \ldots, y_{n-1}) \\ &= f_{Y_0}(y_0)\prod_{i=1}^{n} f_{Y_i|Y_{i-1}}(y_i) \\ &= \frac{\theta^{y_0}e^{-\theta}}{y_0!}\prod_{i=1}^{n}\frac{(\theta y_{i-1})^{y_i}e^{-\theta y_{i-1}}}{y_i!}. \end{aligned}$$

The log-likelihood function becomes

$$\ell(\theta) \propto y_0 \log\theta - \theta + \log\theta\sum_{i=1}^{n} y_i - \theta\sum_{i=1}^{n} y_{i-1}.$$

By taking the first derivative,

$$\frac{\partial}{\partial\theta}\ell(\theta) = y_0\theta^{-1} - 1 + \theta^{-1}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_{i-1} = 0$$

Then, the MLE for θ equals

$$\hat{\theta} = \frac{\sum_{i=0}^{n} y_i}{1 + \sum_{i=0}^{n-1} y_i}.$$

The second derivative equals $-\theta^{-2}\sum_{i=0}^{n} y_i$, which is negative for all θ. The $\hat{\theta}$ is the global maximizer.

(b) Find $E(Y_1)$, $E(Y_2)$, and a general expression of $E(Y_j)$, $j = 0, 1, \ldots, n$.

Solution: Using double expectation, $E(Y_1) = EE(Y_1|Y_0) = E(\theta Y_0) = \theta^2$. Similarly, $E(Y_2) = EE(Y_2|Y_1) = E(\theta Y_1) = \theta^3$. It is not hard to find the general expression as $E(Y_j) = \theta^{j+1}$, $j = 0, 1, \ldots, n$.

(c) Given that the inverse of the variance of $\hat{\theta}$, $V(\hat{\theta})^{-1}$, equals $-E(\partial^2\ell(\theta)/\partial\theta^2)$, where $\ell(\theta)$ is the log-likelihood function derived in (a), and

$$\frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}} \to_d N(0,1),$$

where $\hat{V}(\hat{\theta})$ is the estimator of $V(\hat{\theta})$ with θ replaced by $\hat{\theta}$ (we have seen this kind of approach before), derive the 95% CI of θ when n is large.

Solution: We have

$$V(\hat{\theta})^{-1} = -E(\partial^2\ell(\theta)/\partial\theta^2) = \theta^{-2}\sum_{i=0}^{n} E(Y_i) = \theta^{-2}\sum_{i=0}^{n}\theta^{i+1} = \theta^{-1}\frac{1-\theta^{n+1}}{1-\theta}.$$

When n is large, one may use normality to approximate the distribution. Hence,

$$0.95 = P\left(-1.96 < \frac{\hat{\theta}-\theta}{\sqrt{\hat{V}(\hat{\theta})}} < 1.96\right),$$

where

$$\hat{V}(\hat{\theta}) = \frac{\hat{\theta}(1-\hat{\theta})}{1-\hat{\theta}^{n+1}}.$$

The 95% CI for θ is $(\hat{\theta} - 1.96\sqrt{\hat{V}(\hat{\theta})}, \hat{\theta} + 1.96\sqrt{\hat{V}(\hat{\theta})})$.

7. (Bios 673/740 in class, C&B 7.37) Let $X_1, \ldots, X_{n+1}$ be iid Bernoulli(p), and define the function $h(p)$ by

$$h(p) = P\left(\sum_{i=1}^{n} X_i > X_{n+1}|p\right),$$

which is the probability that the first n observations exceed the $(n+1)$st.

(a) Show that

$$T(X_1, \ldots, X_{n+1}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i > X_{n+1} \\ 0 & \text{otherwise} \end{cases}$$

is an unbiased estimator of $h(p)$.

(b) Find the best unbiased estimator of $h(p)$.