

Homework 1

Aiden Kenny

STAT GR5204: Statistical Inference

Columbia University

November 10, 2020

Question 1

When rolling two dice, there are six possible ways for their total to sum up to seven: (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1), so the probability of the sum being seven is $6/36 = 1/6$. If X is the number of trials where the total of both rolls is seven, then we can think of $X \sim \text{Bin}(120, 1/6)$, and so $\mathbb{E}X = 20$ and $\text{Var}X = 50/3$. Using the Central Limit Theorem, we then have

$$\Pr(|X - 20| \leq k) = \Pr\left(\left|\frac{X - 20}{\sqrt{50/3}}\right| \leq k\sqrt{\frac{3}{50}}\right) = 2\Phi\left(k\sqrt{\frac{3}{50}}\right) - 1 \stackrel{\text{set}}{=} 0.95 \implies \Phi\left(k\sqrt{\frac{3}{50}}\right) = 0.975.$$

Using a table of values for $\Phi(z)$, we can see that $k\sqrt{3/50} = 1.96$, and so $k = 1.96\sqrt{50/3} \approx 8$.

Question 2

Let $X \sim \text{Pois}(10)$, and so $\mathbb{E}X = \text{Var}X = 10$. Using the CLT without any continuity correction, we have $(X - 10)/\sqrt{10} \approx N(0, 1)$, and so

$$\Pr(8 \leq X \leq 12) = \Pr\left(\frac{8 - 10}{\sqrt{10}} \leq Z \leq \frac{12 - 10}{\sqrt{10}}\right) = \Pr(|Z| \leq \sqrt{2/5}) \approx 2\Phi(\sqrt{2/5}) - 1 = 0.4714.$$

If we do use continuity correction, then we have

$$\begin{aligned}\Pr(8 \leq X \leq 12) &\approx \Pr(7.5 \leq X \leq 12.5) \\ &= \Pr\left(\frac{7.5 - 10}{\sqrt{10}} \leq Z \leq \frac{12.5 - 10}{\sqrt{10}}\right) = \Pr(|Z| \leq 2.5/\sqrt{10}) \approx 2\Phi(2.5/\sqrt{10}) - 1 = 0.5704.\end{aligned}$$

Question 3

We are assuming that when a program is run, an execution error will occur with probability $\theta \in [0, 1]$. If X is whether or not an execution error occurs, we have $X \sim \text{Ber}(\theta)$, and $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ for $x = \{0, 1\}$. We also believe that $\theta \sim \text{Unif}(0, 1)$, and so $\xi(\theta) = 1$ for $0 \leq \theta \leq 1$.

- (a) After 25 runs of the program we have 10 erros, so $f(\mathbf{x}|\theta) = \theta^{10}(1-\theta)^{15}$. The marginal distribution of \mathbf{X} is given by

$$g_{\mathbf{X}}(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta) \cdot \xi(\theta) \, d\theta = \int_0^1 \theta^{10}(1-\theta)^{15} \cdot 1 \, d\theta = \int_0^1 \theta^{11-1}(1-\theta)^{16-1} \, d\theta = B(11, 16),$$

and so the posterior pdf of θ is

$$\xi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) \cdot \xi(\theta)}{g_{\mathbf{X}}(\mathbf{x})} = \frac{\theta^{10}(1-\theta)^{15} \cdot 1}{B(11, 16)} = \frac{\theta^{11-1}(1-\theta)^{16-1}}{B(11, 16)}.$$

That is, $\theta|\mathbf{x} \sim \text{Beta}(11, 16)$.

- (b) If we are using squared error loss, then our Bayes' estimate is $\delta^*(\mathbf{x}) = \mathbb{E}(\theta|\mathbf{x}) = 11/27$.

Question 4

We believe that $\theta \sim \text{Beta}(3, 4)$, where $\theta \in [0, 1]$ is the proportion of bad apples in the lot. Choosing apples from the lot is essentially sampling from a Bernoulli distribution with parameter θ , and we know that Beta distributions are closed under sampling from a Bernoulli distribution. After choosing 10 apples, we find that three of them are bad, so our posterior distribution becomes $\theta | \mathbf{x} \sim \text{Beta}(3 + 3, 4 + 7) = \text{Beta}(6, 11)$. If we use squared error loss, our Bayes' estimate is then $\delta^*(\mathbf{x}) = \mathbb{E}(\theta | \mathbf{x}) = 6/17$.

Question 5

Let $\mathbf{X} = (X_1, \dots, X_n)^T$ be a random sample from $X \sim \text{Unif}(\theta, 2\theta)$, where $\theta > 0$. The likelihood function is then given by $f(\mathbf{x} | \theta) = 1/\theta^n$ when $\theta \leq x_i \leq 2\theta$ for $i \in \{1, \dots, n\}$. We can re-frame the boundaries of the likelihood function using order statistics. Since we need every observation $x_i \in [\theta, 2\theta]$, it follows that $\theta \leq x_{(1)} \leq \dots \leq x_{(n)} \leq 2\theta$, where $x_{(j)}$ is the j th order statistics; namely, we have $\theta \leq x_{(1)}$ and $x_{(n)} \leq 2\theta$. From the second inequality, we have $x_{(n)}/2 \leq \theta$, and so the possible values of θ are $x_{(n)}/2 \leq \theta \leq x_{(1)}$. In other words, even though we had the original parameter space $\Theta = (0, \infty)$, because the bounds of the density functions depended on θ , we were able to restrict θ to a new parameter space $\tilde{\Theta} = [x_{(n)}/2, x_{(1)}]$. We can see that our likelihood function is monotone decreasing, and so it will be maximized by the smallest possible value of θ . Therefore, the MLE of θ is $\hat{\theta}(\mathbf{X}) = X_{(n)}/2$.

Question 6

Suppose that $\mathbf{X} = (X_1, X_2, X_3)^T$ are each exponentially distributed with $\mathbb{E}X_i = i\theta$, where $\theta > 0$. This implies that $X_i \sim \text{Exp}(1/i\theta)$, and so $f(x_i | \theta) = e^{-x_i/i\theta}/i\theta$.

(a) The likelihood function is given by

$$f(\mathbf{x} | \theta) = \prod_{i=1}^3 f(x_i | \theta) = \prod_{i=1}^3 \frac{e^{-x_i/i\theta}}{i\theta} = \frac{1}{6\theta^3} \exp\left(-\frac{1}{\theta} \sum_{i=1}^3 \frac{x_i}{i}\right),$$

and the corresponding log-likelihood function is given by $\ell(\mathbf{x} | \theta) = -3 \log(6\theta) - \frac{1}{\theta} \sum_{i=1}^3 x_i/i$. Differentiating $\ell(\mathbf{x} | \theta)$ with respect to θ , setting to 0, and solving for θ gives us the MLE:

$$\frac{\partial \ell}{\partial \theta} = -\frac{3}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^3 \frac{x_i}{i} \stackrel{\text{set}}{=} 0 \implies \hat{\theta}(\mathbf{X}) = \frac{1}{3} \sum_{i=1}^3 \frac{X_i}{i}$$

(b) Let $\psi = 1/\theta$, and we believe that $\psi \sim \text{Gamma}(\alpha, \beta)$, i.e. $\xi(\psi) = \frac{\beta^\alpha}{\Gamma(\alpha)} \psi^\alpha e^{-\beta\psi}$ for $\psi > 0$. For notational ease, let $\varphi(\mathbf{x}) = \sum_{i=1}^3 x_i/i$; the likelihood function of ψ is then given by $f(\mathbf{x} | \psi) = \psi^3 e^{-\varphi(\mathbf{x})\psi}/6$. We have

$$\xi(\psi | \mathbf{x}) \propto f(\mathbf{x} | \psi) \cdot \xi(\psi) = \frac{1}{6} \psi^3 e^{-\varphi(\mathbf{x})\psi} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \psi^\alpha e^{-\beta\psi} \propto \psi^{\alpha+3-1} e^{-(\beta+\varphi(\mathbf{x}))\psi}.$$

This is very similar to a Gamma distribution with parameters $\tilde{\alpha} = \alpha + 3$ and $\tilde{\beta} = \beta + \varphi(\mathbf{x})$, and adding in the normalizing constants would make it so. Therefore, we conclude that $\psi | \mathbf{x} \sim \text{Gamma}(\alpha + 3, \beta + \varphi(\mathbf{x}))$.

Question 7

We assume that our parameter θ has a prior density $\xi(\theta) = \theta e^{-\theta}$ for $\theta > 0$. Let $X \sim \text{Unif}(0, \theta)$, and so $f(x | \theta) = 1/\theta$ for $0 \leq x \leq \theta$. We note that while the original parameter space was $\Theta = (0, \infty)$, sampling from a uniform distribution whose bound depended on θ resulted in a new parameter space $\tilde{\Theta} = [x, \infty)$. The marginal distribution of X is given by

$$g_X(x) = \int_{\tilde{\Theta}} f(x | \theta) \cdot \xi(\theta) d\theta = \int_x^\infty \frac{1}{\theta} \cdot \theta e^{-\theta} d\theta = \int_x^\infty e^{-\theta} d\theta = \left[-e^{-\theta}\right]_x^\infty = e^{-x},$$

and so our posterior density for θ is given by

$$\xi(\theta | x) = \frac{f(x | \theta) \cdot \xi(\theta)}{g_X(x)} = \frac{1}{\theta} \cdot \theta e^{-\theta} \cdot \frac{1}{e^{-x}} = e^{-(\theta-x)}.$$

This density corresponds to a “shifted” exponential distribution with $\lambda = 1$; instead of starting at zero, we are starting at x .

(a) Using squared error loss, our Bayes’ estimate is the mean of the posterior distribution:

$$\delta^*(x) = \mathbb{E}(\theta | x) = \int_x^\infty \theta \cdot e^{-(\theta-x)} d\theta = \left[-\theta e^{-(\theta-x)} \right]_x^\infty + \int_x^\infty e^{-(\theta-x)} d\theta = x + 1.$$

(b) Using absolute error loss, our Bayes’ estimate is the median of the posterior distribution. The posterior cdf is easily seen to be $\Xi(\theta | x) = 1 - e^{-(\theta-x)}$. To get the median, we need $\Xi(\delta^*(x)) = 1 - e^{-(\delta^*(x)-x)} = 1/2$, and solving for $\delta^*(x)$ gives us $\delta^*(x) = x + \log 2$.

Question 8

Because $0 \leq \beta \leq 1$, we have $1/3 \leq \theta \leq 2/3$. If we are sampling from $X \sim \text{Ber}(\theta)$, then the likelihood and log-likelihood functions are given by $f(\mathbf{x} | \theta) = \theta^{n\bar{x}}(1 - \theta)^{n(1-\bar{x})}$ and $\ell(\mathbf{x} | \theta) = n\bar{x} \log \theta + n(1 - \bar{x}) \log(1 - \theta)$. Differentiating $\ell(\mathbf{x} | \theta)$, setting equal to zero, and solving for θ gives the MLE as

$$\frac{\partial \ell(\mathbf{x} | \theta)}{\partial \theta} = \frac{n\bar{x}}{\theta} - \frac{n(1 - \bar{x})}{1 - \theta} = n \left(\frac{\bar{x} - \theta}{\theta(1 - \theta)} \right) \stackrel{\text{set}}{=} 0 \implies \hat{\theta} = \bar{x}.$$

We must be cautious; because each $x_i \in \{0, 1\}$, we can have $\bar{x} \in [0, 1]$, and so the maximum of $\ell(\mathbf{x} | \theta)$ can occur at $\hat{\theta} \in [0, 1]$. However, because of the constraints placed on θ by β , this maximum can potentially fall outside the range of possible values. To remedy this, we will consider two cases:

1. $\bar{x} < 1/3$: for all values $\theta \in [1/3, 2/3]$, we have $\partial \ell / \partial \theta < 0$, so $\ell(\mathbf{x} | \theta)$ is a decreasing function. Then the maximum value of $\ell(\mathbf{x} | \theta)$ is obtained when $\theta = 1/3$.
2. $\bar{x} > 2/3$: for all values $\theta \in [1/3, 2/3]$, we have $\partial \ell / \partial \theta > 0$, so $\ell(\mathbf{x} | \theta)$ is an increasing function. Then the maximum value of $\ell(\mathbf{x} | \theta)$ is obtained when $\theta = 2/3$.

Therefore, the MLEs for both θ and β are given by

$$\hat{\theta} = \begin{cases} \bar{X} & \text{if } 1/3 \leq \bar{X} \leq 2/3, \\ 1/3 & \text{if } \bar{X} < 1/3, \\ 2/3 & \text{if } \bar{X} > 2/3. \end{cases} \quad \text{and} \quad \hat{\beta} = 3\hat{\theta} - 1 = \begin{cases} 3\bar{X} - 1 & \text{if } 1/3 \leq \bar{X} \leq 2/3, \\ 0 & \text{if } \bar{X} < 1/3, \\ 1 & \text{if } \bar{X} > 2/3. \end{cases}$$

This is because $\beta = 3\theta - 1$, which means the MLE of β is given by $\hat{\beta} = 3\hat{\theta} - 1$.

Question 9

We are sampling from a “shifted” exponential distribution, i.e. its density is given by $f(x | \beta, \theta) = \beta e^{-\beta(x-\theta)}$ for $x \geq \theta$. The likelihood function is then given by $f(\mathbf{x} | \beta, \theta) = \beta^n e^{-n\beta(\bar{x}-\theta)}$ when $x_i \geq \theta$ for $i \in \{1, \dots, n\}$. If every observation $x_i \geq \theta$, then it is also true that the lowest value for each observation is as well, so $x_{(1)} \geq \theta$. We can incorporate this condition into the likelihood function using an indicator function:

$$f(\mathbf{x} | \beta, \theta) = \beta^n e^{-n\beta(\bar{x}-\theta)} \cdot \mathbb{I}_{[\theta, \infty)}(x_{(1)}).$$

By the Factorization Theorem, \bar{X} and $X_{(1)}$ are a pair of jointly sufficient statistics for β and θ .