

# 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  $\theta$  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,  $\theta \sim \text{Beta}(11, 16)$ .

- (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  $\theta$ , 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  $\theta \sim \text{Beta}(3 + 3, 4 + 7) = \text{Beta}(6, 11)$ . If we use squared error loss, our Bayes' estimate is then  $\delta^*(\mathbf{x}) = \mathbb{E}(\theta | \mathbf{x}) = 6/17$ .

### Question 5

Let  $X_1, \dots, X_n$  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  $\theta$  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  $\theta$ , we were able to restrict  $\theta$  to a new parameter space  $\hat{\Theta} = [x_{(n)}/2, x_{(1)}]$ . We can see that our likelihood function is monotone decreasing, and so it will be maximized by the smallest possible value of  $\theta$ . Therefore, the MLE of  $\theta$  is  $\hat{\theta} = x_{(n)}/2$ .

### 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  $\theta$ , setting to 0, and solving for  $\theta$  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  $\psi$  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  $\psi | \mathbf{x} \sim \text{Gamma}(\alpha + 3, \beta + \varphi(\mathbf{x}))$ .