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

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

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 $(X - 10)/\sqrt{10} \approx \text{N}(0,1)$, and so

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

If we do use continuity correction, then we have

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

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

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

and so the posterior pdf of θ is

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

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

(b) If we are using squared error loss, then our Bayes' estimate would be $\hat{\theta} = \mathbb{E}(\theta \,|\, \mathbf{x}) = 11/(11 + 16) = 11/27$.