EE5137 2020/21 (Sem 2): Quiz 1 (Total 40 points)

Name:______________________________

Matriculation Number:_______________

Score:______________________________

You have 1.5 hours for this quiz. There are THREE (3) printed pages. You're allowed 1 sheet of handwritten notes. Please provide *careful explanations* for all your solutions.

1. [Conditional Expectation] Mark is running a bicycle shop. Every day, he is unable to run his business due to the rain which occurs with with probability $p \in [0, 1]$. Mark works everyday except rainy days, which he takes as holidays. Let X be the number of consecutive days Mark has to work between rainy days. Let Y be the number of customers who go to Mark's bicycle shop in this period of X days. Conditioned on the event $\{X = x\}$, the distribution of Y is

$$\Pr(Y = y \mid X = x) = \frac{e^{-\lambda x}(\lambda x)^y}{y!}, \quad y = 0, 1, 2, \ldots$$

This is known as the Poisson distribution with rate λx where $\lambda > 0$. You can use without proof the facts that Poisson distribution with rate μ has expectation μ and variance also μ.

(a) (2 points) Write down the distribution (probability mass function) of X. Be sure to specify the set of values that X takes on.

(b) (4 points) Find $\mathbb{E}[X]$ and $\mathrm{Var}(X)$. You may use without proof the facts that

$$\sum_{k=1}^{\infty} k a^k = \frac{a}{(1-a)^2}, \qquad \sum_{k=1}^{\infty} k^2 a^k = \frac{a(1+a)}{(1-a)^3}, \qquad |a| < 1.$$

(c) (3 points) By writing $\mathrm{Var}(Y)$ as $\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$, show that

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X])$$

where $\mathrm{Var}(Y|X)$ is the random variable that takes on the value $\mathrm{Var}(Y|X = x)$ with probability $\Pr(X = x)$.

(d) (3 points) Using the law of iterated expectations find $\mathbb{E}[Y]$.

(e) (3 points) Using Part (c), or otherwise, find $\mathrm{Var}(Y)$.

2. [Convergence in Distribution] Suppose that $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables with cumulative distribution function

$$F_{X_i}(x) = \Pr(X_i \leq x) = \begin{cases} \frac{x}{x+1} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

Let $M_n = \max\{X_1, X_2, \ldots, X_n\}$.

(a) (2 points) Find the cumulative distribution function of M_n.

(b) (2 points) Find the cumulative distribution function of M_n/n.

(c) (6 points) The exponential distribution of rate θ has probability density function

$$f(w) = \begin{cases} \theta e^{-\theta w} & w \geq 0 \\ 0 & w < 0 \end{cases}.$$

Show that the sequence of random variables $\{M_n/n\}_{n=1}^{\infty}$ converges in distribution to a random variable Y whose inverse is an exponential random variable. Identity the rate of the exponential random variable $1/Y$.

You may use without proof the fact that for any $a > 0$,

$$\lim_{n\to\infty} \left(1 - \frac{a}{n}\right)^n = e^{-a}.$$

3. [Balls and Bins] Suppose we have n balls and n bins. We throw each ball uniformly at random into one of the n bins. The throws of the n balls into the n bins are done independently. In this problem, we would like to prove the following fact.

Theorem 1. *Let $n \geq 200$. With probability at least $1 - 1/n$, the bin that contains the most balls contains no more than*

$$\alpha = \frac{10 \ln n}{\ln(\ln n)}$$

balls.

(a) (2 points) Let $X_1, \ldots, X_n$ be independent indicator random variables where $X_i = 1$ if the i^{th} ball lands in bin 1 and $X_i = 0$ otherwise. Find the expectation of $B_1 = \sum_{i=1}^{n} X_i$, the total number of balls in bin 1.

(b) (6 points) Use the Chernoff bound to find a and b (in terms of α and n) such that

$$\Pr(B_1 \geq \alpha) \leq \exp\big(-nD(a\|b)\big) \quad \text{where} \quad D(a\|b) = a \ln \frac{a}{b} + (1-a) \ln \frac{1-a}{1-b}.$$

(c) (3 points) The following facts can be used without proof.

$$D(a\|b) \geq a\left(\ln \frac{a}{b} - 1\right), \quad \alpha \ln \alpha \geq 8 \ln n, \quad \exp(\exp(10/6)) \approx 199.2.$$

Use above facts to show that

$$\Pr(B_1 \geq \alpha) \leq \frac{1}{n^2}$$

(d) (4 points) Use part (c), symmetry and the union bound to prove Theorem 1.

We remark that α is tight up to constant factors. More precisely, the bin that contains the most balls contains $\frac{\ln n}{\ln(\ln n)} \cdot (1 + o(1))$ balls with probability $= 1 - o(1)$. The balls and bins model is super important for analyzing hashing schemes in security and cryptography.