MAST20004 Probability

Assignment 2

- Assignment boxes are located on the ground floor in the Peter Hall Building (near Wilson computer lab) and your tutor's name and box number for submission are on the signs above the assignment boxes.
- Your solutions to the assignment should be left in the MAST20004 assignment box set up for your tutorial group.
- **Don't forget** to staple your solutions and to print your name, student ID, the subject name and code, and your tutor's name on the first page.
- The submission deadline is **3 pm on Friday 12 April**.
- There are 5 questions, of which 2 randomly chosen questions will be marked. Note you are expected to submit answers to all questions, otherwise a mark penalty will apply. Working and reasoning **must** be given to obtain full credit. Give clear and concise explanations. Clarity, neatness, and style count.

---

1. Let X be a random digit sampled from $\{0, 1, \ldots, 9\}$ with equal probability.

   (a) Let Y_1 be the remainder obtained by dividing X^2 by 10, for example, 7^2 divided by 10 has remainder 9. Calculate the pmf of Y_1.

   (b) Let Y_2 be the remainder obtained by dividing Y_1^2 by 10. Derive the pmf of Y_2.

   (c) Let Y_3 be the remainder obtained by dividing Y_2^2 by 10. Calculate the pmf of Y_3.

2. The following is the cumulative distribution function of a random variable X:

$$F_X(x) = \begin{cases} 0, & \text{if } x < -1, \\ 0.1, & \text{if } -1 \le x < 1, \\ \frac{x}{5}, & \text{if } 1 \le x < 3, \\ 0.9, & \text{if } 3 \le x < 6, \\ 1, & \text{if } x \ge 6. \end{cases}$$

   (a) Plot the cdf of X.

   (b) Is X a discrete random variable or continuous random variable or neither? Justify your answer.

   (c) What are the possible values of X?

   (d) Using the cumulative distribution function calculate (i) $\mathbb{P}(X \le 2)$; (ii) $\mathbb{P}(1 < X \le 2)$; (iii) $\mathbb{P}(1 \le X \le 2)$; (iv) $\mathbb{P}(X > 1.5)$.

[Please Turn Over!]

3. (a) Modify the proof of the formula for computing higher moments via tail probabilities to show that if $\mathbb{P}(X \geq 0) = 1$, X has pdf f and $\mathbb{E}(X^2) < \infty$, then

$$\int_0^\infty x[1 - F_X(x)]dx < \infty.$$

(b) If X is a discrete random variable with pmf $p_X(i) = cp^i$, $i = 3, 4, 5, \ldots$, where $0 < p < 1$.

(i) Determine the value of c in terms of p.

(ii) Compute the cdf of X.

(iii) Use the formula for computing moments via tail probabilities to calculate $\mathbb{E}[X]$.

4. Assume that X is a continuous random variable with cumulative distribution function

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{3x}{2}, & \text{if } 0 \leq x < \frac{1}{2}, \\ cx + \frac{1}{2}, & \text{if } \frac{1}{2} \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

(a) Determine the value c and find the pdf of X.

(b) What are the possible values of X?

(c) Compute the mean and variance of X.

5. A factory has two production lines A_1 and A_2 with A_1 having double the production capacity of that in A_2. It is known that the products from line A_1 normally have defective rate 10% while 16% of those from line A_2 are usually defective. As a quality control, an inspector randomly selects n items for inspection. (In MAST20005, you will see that the distribution of the number of defective items in the sample is of critical importance in determining whether this batch of products has higher defective rate than anticipated.) We now consider two methods for selecting the random sample: 1) mix up all the products and then select n from them and the number of defectives is denoted by X_n; 2) with probability in proportion to the number of items produced, we select a production line, then select n items from the chosen line and we let Y_n be the number of defectives in the sample.

(a) For Method 1, identify the distribution of X_n and find its mean and variance.

(b) For Method 2, the distribution of Y_n is called a *mixed distribution*. Instead of working on it analytically, we do simulations for a rough idea of the distribution. Using the Matlab file Assignment2Ex5c_2019.m, available on the LMS, adding suitable commands where indicated in the Matlab file, estimate the pmf, mean and variance of Y_1. Compare your findings with those of X_1 in (a) and comment on your findings.

[*Please Turn Over!*]

(c) Now use the Matlab file to find the pmf of Y_{20}, and estimate the mean and variance of Y_{20}. Report your estimates of the mean and variance of Y_{20} in the assignment.
Please print out your code for Part (c) and include it with your assignment. A mark penalty will apply to assignments without code

(d) For $n = 20$, with your findings in (a) and (c), which method is more consistent?