# Problem Set 2

Due: Week 6

1. Which of the following best describes a random variable?
   (a) A deterministic outcome in an experiment
   (b) A function that assigns a numerical value to each outcome in a sample space
   (c) The probability of an event occurring
   (d) A measure of central tendency in a data set
2. Which of the following is best suited to be modeled as a random variable?
   (a) The distance between the campus to the city center
   (b) The probability of winning a lottery
   (c) The lifespan of a cat
   (d) None of the above
3. Which of the following statements is true about the probability distribution of a discrete random variable?
   (a) The sum of all possible probabilities must equal 1
   (b) The possible values cannot be negative
   (c) The distribution must be symmetrical
   (d) Both (a) and (b)
4. For a continuous random variable $X$, the probability that $X$ takes on a specific value is:
   (a) Given by the probability density function (PDF)
   (b) Given by the area under the PDF over an interval
   (c) Always equal to 1
   (d) Always equal to 0
5. What is the primary characteristic of a probability mass function (PMF)?
   (a) It can only be used for continuous random variables
   (b) It integrates to 1 over the entire range of the variable
   (c) It assigns probabilities to each possible value of a discrete random variable
   (d) It represents the cumulative probability up to a certain value

6. Suppose $X$ is a random variable with the CDF:

| X | 1 | 3 | 5 | 7 |
|---|---|---|---|---|
| CDF | 0.5 | 0.75 | 0.9 | 1 |

What is $P(X = 3)$?

(a) 0.15

(b) 0.25

(c) 0.5

(d) 0.75

7. Which of the following scenarios can be modeled by a Bernoulli distribution?

(a) The number of customers arriving at a store in an hour.

(b) The number of heads when flipping a coin three times.

(c) Whether a student passes or fails an exam.

(d) The number of defective items in a batch of 10.

8. Which of the following is a key characteristic of a Binomial distribution?

(a) It models the number of successes in a fixed number of independent Bernoulli trials.

(b) It models the time until the first success in a sequence of independent trials.

(c) It models the number of successes in an infinite number of trials.

(d) It models the number of events occurring in a fixed interval of time.

9. There are 100 balls in a box, numbered 1,2,...,100 (no number appears more than once). Five balls are drawn, one at a time, *with replacement.* Answer the following questions.

(a) What is the distribution of the number of balls with a value of at least 80?

(b) What is the distribution of the value of the $j$-th draw $(1 \leq j \leq 5)$?

(c) What is the probability that the number 100 is drawn at least once?

Now consider random sampling *without replacement.* Answer the same set of questions.

(d) What is the distribution of the number of balls with a value of at least 80?

(e) What is the distribution of the value of the $j$-th draw $(1 \leq j \leq 5)$?

(f) What is the probability that the number 100 is drawn at least once?

10.

(a) Show that $p(k) = \left(\frac{1}{2}\right)^{k+1}$ for $k = 0, 1, 2, \ldots$ is a valid PMF for a discrete random variable.

(b) Find the CDF of a random variable with the PMF from (a).

# Solutions

## Question 1-8

BCAD CBCA

## Question 9

(a) Let $I_j$ be an indicator that the $j$-th draw has a value of at least 80. Then $I_j \sim Bern(21/100)$. Let $X$ be the random variable in question. $X = \sum_{j=1}^{5} I_j$. Therefore, $X \sim Bin(5, 0.21)$.

(b) Let $X_j$ be the value of the $j$-th draw. Since all values are equally likely, $X_j \sim DUnif(1, 2, ..., 100)$.

(c) $P(100 \text{ is drawn at least once}) = 1 - P(100 \text{ is never drawn})$. By the classical approach, this is $1 - (99/100)^5 \approx 0.049$.

(d) By definition, this is a Hypergeometric distribution $HGeom(21, 79, 5)$.

(e) Let $Y_j$ be the value of the $j$-th draw. The distribution is still discrete uniform. $Y_j \sim DUnif(1, 2, ..., 100)$. Here $Y_1, ..., Y_5$ are not independent. But the $j$-th draw is equally likely to be any of the numbers.

If you are not convinced, think about a two-ball case. $P(Y_1 = a) = \sum_{b \neq a} P(Y_1 = a | Y_2 = b) P(Y_2 = b) = \sum_{b \neq a} P(Y_1 = a, Y_2 = b) = 99 \times \frac{1}{P_{100}^2} = \frac{1}{100}$.

(f) $1 - \frac{C_{99}^5}{C_{100}^5} = 1 - \frac{95}{100} = 0.05$. But realizing in this case $Y_1 = 100, ..., Y_5 = 100$ are disjoint, we have $P(Y_j = 100 \text{ for some } j) = \sum_{j=1}^{5} P(Y_j = 100) = 0.05$.

## Question 10

(a) For a PMF to be valid, it has to be (i) nonnegative and (ii) all values sum up to 1. $p(k) \geq 0$ satisfies the first requirement trivially. It suffices to show the second requirement:

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k+1} = \frac{1}{1 - 1/2} - 1 = 1.$$

(b) Let $F$ be the CDF. Then

$$F(n) = \sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \left(\frac{1}{2}\right)^{k+1} = \frac{\frac{1}{2}(1 - 1/2^{n+1})}{\frac{1}{2}} = 1 - \frac{1}{2^{n+1}}.$$