Stat 113 (Introduction to Mathematical Statistics)

Lesson 2.1: Discrete Random Variables and Their Probability Distributions

Learning Outcomes

At the end of the lesson, students must be able to

1. Explain what is a random variable,
2. Differentiate discrete and continuous random variables, and
3. Construct the probability mass function and cumulative distribution function of a discrete random variable.

Introduction

In practice it is convenient to represent random outcomes numerically. If the outcome is numerical and one-dimensional we call the outcome a random variable. If the outcome is multi-dimensional we call it a random vector. Random variables are one of the most important and core concepts in probability and statistics. It provides a way to quantify the outcomes of random experiments.

As an example consider a coin flip which has the possible outcome H or T. We can write the outcome numerically by setting the result as $X = 1$ if the coin is heads and $X = 0$ if the coin is tails. The object X is random since its value depends on the outcome of the coin flip.

Definition:

A *random variable* is a function which assigns a numerical value to every outcome of a random experiment.

Example 2.1:

Consider tossing 2 coins. The possible outcomes are

$$\Omega = \{HH, HT, TH, TT\}$$

Let us count the number of heads and label it as Y. Then

$$Y = \begin{cases} 0, & \text{if } \omega = TT \\ 1, & \text{if } \omega = TH \text{ or } \omega = TH \\ 2, & \text{if } \omega = HH \end{cases}$$

Remarks:

1. In the introductory example, the values of X are random since being equal to 0 or 1 depends on the random outcomes in tossing a coin. Hence, X is a random variable. The same is true in Example 2.1
2. By convention we use uppercase letters to denote random variables, say X, Y, or Z, and the corresponding lowercase letters for the realized value.

Example 2.2:

Consider rolling a pair of dice. Let Z be the sum of the number of dots on each die. What are the possible values of Z?

Types of random variables

A random variable which assumes finite or countably infinite number of possible values is called a *discrete* random variable.

Random variables X, Y, and Z are examples of discrete random variables. Other examples are

1. X = number of vehicles owned by a random sample of politicians
2. Y = number of houses affected by a typhoon
3. Z = number of mobile phones owned by a random sample of students

In the above examples, it is clear that a discrete random variable assumes values corresponding to the natural numbers or counting numbers, otherwise called the whole numbers.

Meanwhile, a random variable which assumes values in an interval of values is referred to as *continuous* random variable. Examples are

1. X = height of a random tree in a forest reserve
2. Y = salary of a randomly selected VSU employee
3. Z = length of time a random student finishes an exam

Probability distribution of a discrete random variable

Every random variable has an associated *probability distribution*, or simply *distribution.* The distribution of the random variable, say X, refers to the assignment of probabilities to all events defined in terms of this random variable.

Definition:

Suppose that X is a discrete random variable. The function $p_X(x) = P(X = x)$ is called the *probability mass function* (pmf) for X.

Remarks:

1. The pmf of X has two parts:

   a. the domain of X, and
   b. the probability assignment $P(X = x)$ for every $x \in \mathbb{R}$

2. For any discrete probability distribution, the following must be true:

   a. $p_X(x) = P(X = x) \geq 0$, and
   b. $\sum p_X(x) = 1, \forall\, x \in \mathbb{R}$

3. The pmf of a discrete random variable can be presented as a table, a graph, or a formula.

4. Given the pmf of a discrete random variable we can derive its cumulative distribution function (CDF), and vice versa.

Definition:

The cumulative distribution function (CDF) of a random variable, say X, is given by $F_X(x) = P(X \leq x)$.

Example 2.3:

In Example 2.1 we defined Y as the number if heads recorded in tossing two coins and the possible values of Y are $0, 1,$ or 2. Then the pmf of Y is given by

Y	0	1	2
$P(Y = y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

while the CDF of Y is given by

$$F_Y(y) = \begin{cases} 0, \text{ if } y < 0 \\ \frac{1}{4}, \text{ if } 0 \le y < 1 \\ \frac{3}{4}, \text{ if } 1 \le y < 2 \\ 1, \text{ if } y \ge 2 \end{cases}$$

QUESTIONS:

1. How did I come up with this pmf?
2. How did I come up with the CDF?

Example 2.4:

Consider the random experiment of tossing 3 coins. Define X as the number of heads recorded in this experiment.

a. Derive the pmf of X.
b. Derive the CDF of X.