
3.1.3 Probability Mass Function (PMF)

If X is a discrete random variable then its range R_X is a countable set, so, we can list the elements in R_X . In other words, we can write

$$R_X = \{x_1, x_2, x_3, \dots\}.$$

Note that here x_1, x_2, x_3, \dots are possible values of the random variable X . While random variables are usually denoted by capital letters, to represent the numbers in the range we usually use lowercase letters such as x, x_1, y, z , etc. For a discrete random variable X , we are interested in knowing the probabilities of $X = x_k$. Note that here, the event $A = \{X = x_k\}$ is defined as the set of outcomes s in the sample space S for which the corresponding value of X is equal to x_k . In particular,

$$A = \{s \in S | X(s) = x_k\}.$$

The probabilities of events $\{X = x_k\}$ are formally shown by the probability mass function (pmf) of X .

Definition 3.1

Let x be a discrete random variable with range $R_X = \{x_1, x_2, x_3, \dots\}$ (finite or countably infinite). The function

$$P_X(x_k) = P(X = x_k), \text{ for } k = 1, 2, 3, \dots,$$

is called the *probability mass function (PMF)* of X .

Thus, the PMF is a probability measure that gives us probabilities of the possible values for a random variable. While the above notation is the standard notation for the PMF of X , it might look confusing at first. The subscript x here indicates that this is the PMF of the random variable X . Thus, for example, $P_X(1)$ shows the probability that $X = 1$. To better understand all of the above concepts, let's look at some examples.

Example 3.3

I toss a fair coin twice, and let X be defined as the number of heads I observe. Find the range of X , R_X , as well as its probability mass function P_X .

- Solution
 - Here, our sample space is given by

$$S = \{HH, HT, TH, TT\}.$$

The number of heads will be 0, 1 or 2. Thus

$$R_X = \{0, 1, 2\}.$$

Since this is a finite (and thus a countable) set, the random variable X is a discrete random variable. Next, we need to find PMF of X . The PMF is defined as

$$P_X(k) = P(X = k) \text{ for } k = 0, 1, 2.$$

We have

$$P_X(0) = P(X = 0) = P(TT) = \frac{1}{4},$$

$$\begin{aligned} P_X(1) &= P(X = 1) = P(\{HT, TH\}) = \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2}, \end{aligned}$$

$$P_X(2) = P(X = 2) = P(HH) = \frac{1}{4}.$$

Although the PMF is usually defined for values in the range, it is sometimes convenient to extend the PMF of X to all real numbers. If $x \notin R_X$, we can simply write $P_X(x) = P(X = x) = 0$. Thus, in general we can write

$$P_X(x) = \begin{cases} P(X = x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$

To better visualize the PMF, we can plot it. Figure 3.1 shows the PMF of the above random variable X . As we see, the random variable can take three possible values 0, 1 and 2. The figure also clearly indicates that the event $X = 1$ is twice as likely as the other two possible values. The Figure can be interpreted in the following way: If we repeat the random experiment (tossing a coin twice) a large number of times, then about half of the times we observe $X = 1$, about a quarter of times we observe $X = 0$, and about a quarter of times we observe $X = 2$.

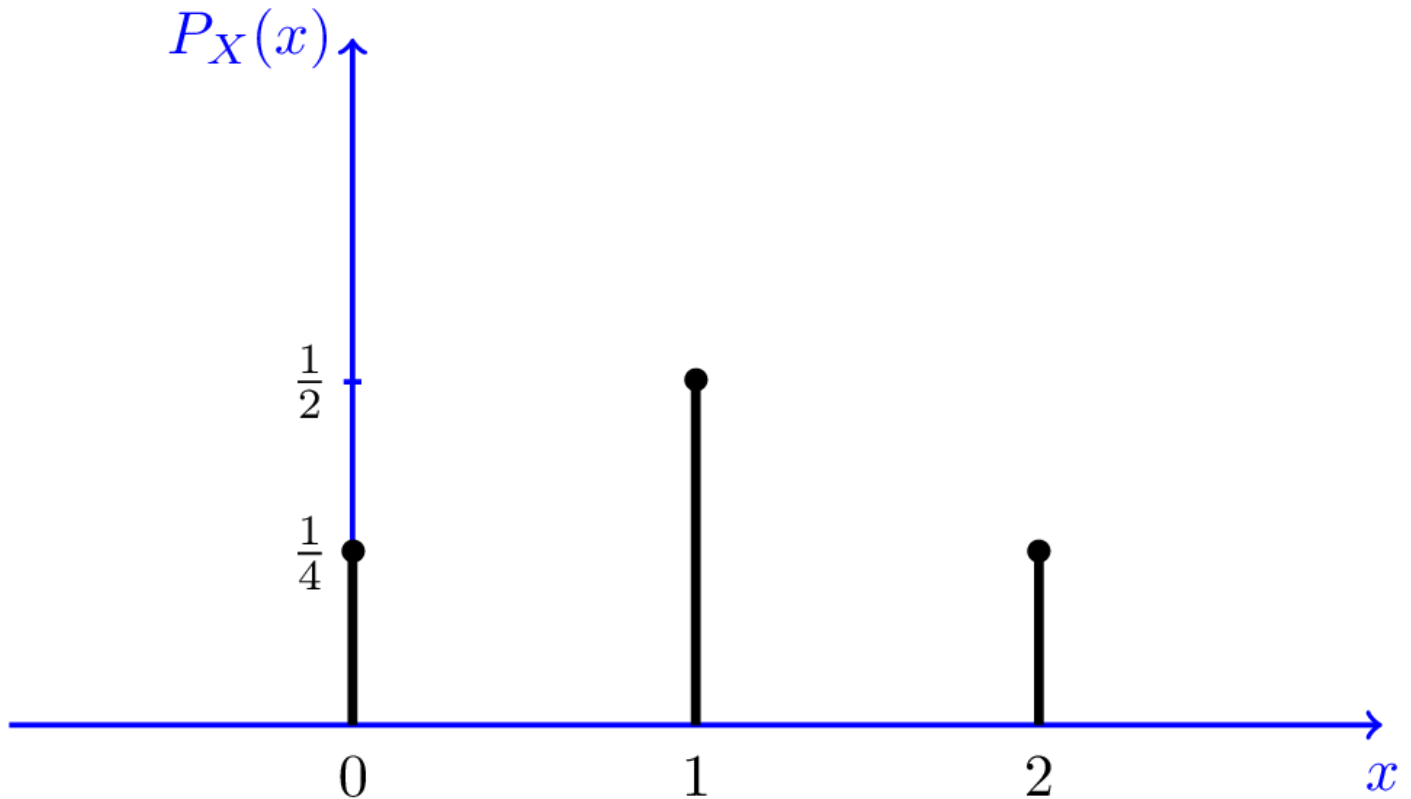


Fig.3.1 - PMF for random Variable x in Example 3.3.

For discrete random variables, the PMF is also called the probability distribution. Thus, when asked to find the probability distribution of a discrete random variable X , we can do this by finding its PMF. The phrase *distribution function* is usually reserved exclusively for the cumulative distribution function CDF (as defined later in the book). The word *distribution*, on the other hand, in this book is used in a broader sense and could refer to PMF, probability density function (PDF), or CDF.

Example 3.4

I have an unfair coin for which $P(H) = p$, where $0 < p < 1$. I toss the coin repeatedly until I observe a heads for the first time. Let Y be the total number of coin tosses. Find the distribution of Y .

- Solution
 - First, we note that the random variable Y can potentially take any positive integer, so we have $R_Y = \mathbb{N} = \{1, 2, 3, \dots\}$. To find the distribution of Y , we need to find $P_Y(k) = P(Y = k)$ for $k = 1, 2, 3, \dots$. We have

$$P_Y(1) = P(Y = 1) = P(H) = p,$$

$$P_Y(2) = P(Y = 2) = P(TH) = (1 - p)p,$$

$$P_Y(3) = P(Y = 3) = P(TTH) = (1 - p)^2 p,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot$$

$$P_Y(k) = P(Y = k) = P(TT \dots TH) \\ = (1 - p)^{k-1} p.$$

Thus, we can write the PMF of Y in the following way

$$P_Y(y) = \begin{cases} (1 - p)^{y-1} p & \text{for } y = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

Consider a discrete random variable X with $\text{Range}(X) = R_X$. Note that by definition the PMF is a probability measure, so it satisfies all properties of a probability measure. In particular, we have

- $0 \leq P_X(x) \leq 1$ for all x , and
- $\sum_{x \in R_X} P_X(x) = 1$.

Also note that for any set $A \subset R_X$, we can find the probability that $X \in A$ using the PMF

$$P(X \in A) = \sum_{x \in A} P_X(x).$$

Properties of PMF:

- $0 \leq P_X(x) \leq 1$ for all x ;
- $\sum_{x \in R_X} P_X(x) = 1$;
- for any set $A \subset R_X$, $P(X \in A) = \sum_{x \in A} P_X(x)$.

Example 3.5

For the random variable Y in Example 3.4,

1. Check that $\sum_{y \in R_Y} P_Y(y) = 1$.
2. If $p = \frac{1}{2}$, find $P(2 \leq Y < 5)$.

- Solution
 - In Example 3.4, we obtained

$$P_Y(k) = P(Y = k) = (1 - p)^{k-1}p, \text{ for } k = 1, 2, 3, \dots$$

Thus,

1. to check that $\sum_{y \in R_Y} P_Y(y) = 1$, we have

$$\begin{aligned} \sum_{y \in R_Y} P_Y(y) &= \sum_{k=1}^{\infty} (1 - p)^{k-1}p \\ &= p \sum_{j=0}^{\infty} (1 - p)^j \\ &= p \frac{1}{1 - (1 - p)} && \text{Geometric sum} \\ &= 1; \end{aligned}$$

2. if $p = \frac{1}{2}$, to find $P(2 \leq Y < 5)$, we can write

$$\begin{aligned} P(2 \leq Y < 5) &= \sum_{k=2}^4 P_Y(k) \\ &= \sum_{k=2}^4 (1 - p)^{k-1}p \\ &= \frac{1}{2} \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} \right) \\ &= \frac{7}{16}. \end{aligned}$$

← [previous](#)
[next](#) →

The print version of the book is available through Amazon [here](#).