3.2 Distributions and probability mass functions

Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

> **Definition 3.2.1 (Discrete random variable).**
>
> A random variable X is said to be *discrete* if there is a finite list of values $a_1, a_2, \dots, a_n$ or an infinite list of values $a_1, a_2, \dots$ such that $P(X = a_j \text{ for some } j) = 1$. If X is a discrete r.v., then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called the *support* of X.

Most commonly in applications, the support of a discrete r.v. is a set of integers. In contrast, a *continuous* r.v. can take on any real value in an interval (possibly even the entire real line); such r.v.s are defined more precisely in Unit 4. It is also possible to have an r.v. that is a *hybrid* of discrete and continuous, such as by flipping a coin and then generating a discrete r.v. if the coin lands Heads and generating a continuous r.v. if the coin lands Tails. For example, imagine that a customer in a store flips a coin to decide whether to make a purchase. If the coin lands Heads, the customer doesn't buy anything; if Tails, the customer spends some random positive real amount of money. But the starting point for understanding such r.v.s is to understand discrete and continuous r.v.s.

Given a random variable, we would like to be able to describe its behavior using the language of probability. For example, we might want to answer questions about the probability that the r.v. will fall into a given range: if L is the lifetime earnings of a randomly chosen U.S. college graduate, what is the probability that L exceeds a million dollars? If M is the number of major earthquakes in California in the next five years, what is the probability that M equals 0? The *distribution* of a random variable provides the answers to these questions; it specifies the probabilities of all events associated with the r.v., such as the probability of it equaling 3 and the probability of it being at least 110. We will see that there are several equivalent ways to express the distribution of an r.v. For a discrete r.v., the most natural way to do so is with a *probability mass function*, which we now define.

> **Definition 3.2.2 (Probability mass function).**

The *probability mass function* (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P(X = x)$. Note that this is positive if x is in the support of X, and 0 otherwise. Here $X = x$ denotes an *event*, consisting of all outcomes s to which X assigns the number x.

Let's look at a few examples of PMFs.

Example 3.2.3 (Coin tosses continued).

In this example we'll find the PMFs of all the random variables in Example 3.1.3, the example with two fair coin tosses. Here are the r.v.s we defined, along with their PMFs:

- X, the number of Heads. Since X equals 0 if TT occurs, 1 if HT or TH occurs, and 2 if HH occurs, the PMF of X is the function p_X given by

$$\begin{aligned} p_X(0) &= P(X=0) = 1/4, \\ p_X(1) &= P(X=1) = 1/2, \\ p_X(2) &= P(X=2) = 1/4, \end{aligned}$$

  and $p_X(x) = 0$ for all other values of x.

- $Y = 2 - X$, the number of Tails. Reasoning as above or using the fact that

$$P(Y = y) = P(2 - X = y) = P(X = 2 - y) = p_X(2 - y),$$

  the PMF of Y is

$$\begin{aligned} p_Y(0) &= P(Y=0) = 1/4, \\ p_Y(1) &= P(Y=1) = 1/2, \\ p_Y(2) &= P(Y=2) = 1/4, \end{aligned}$$

  and $p_Y(y) = 0$ for all other values of y.

  Note that X and Y have the same PMF (that is, p_X and p_Y are the same function) even though X and Y are not the same r.v. (that is, X and Y are two different functions from $\{HH, HT, TH, TT\}$ to the real line).

- I, the indicator of the first toss landing Heads. Since I equals 0 if TH or TT occurs and 1 if HH or HT occurs, the PMF of I is

$$\begin{aligned} p_I(0) &= P(I=0) = 1/2, \\ p_I(1) &= P(I=1) = 1/2, \end{aligned}$$

  and $p_I(i) = 0$ for all other values of i.

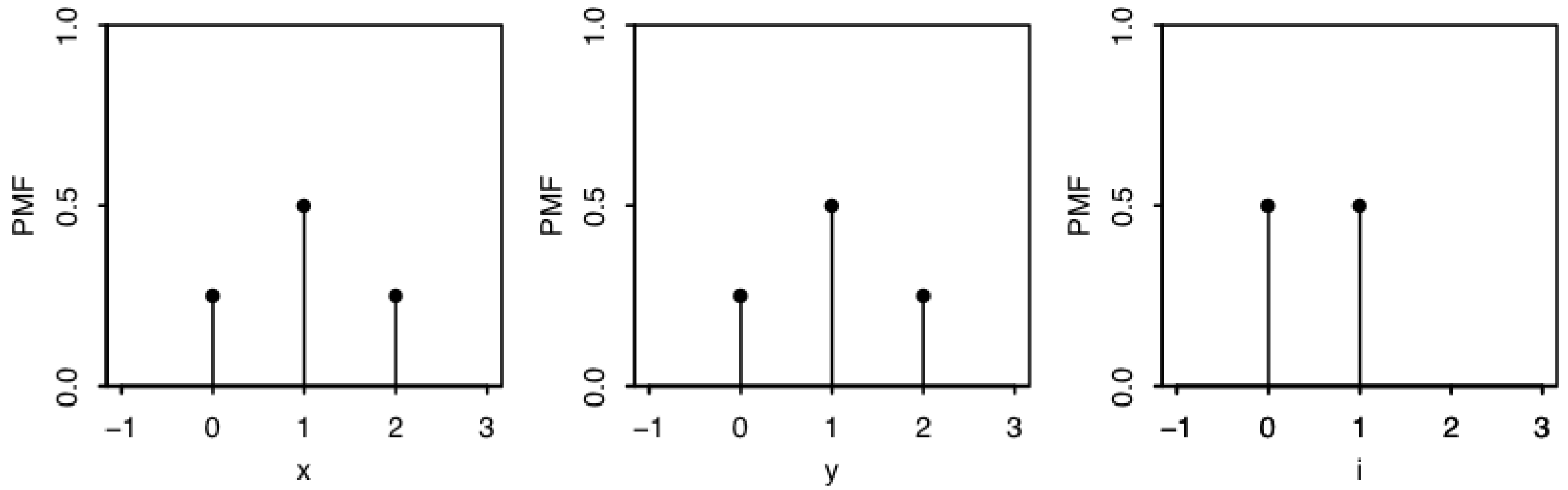

Figure 3.2.4: Left to right: PMFs of X, Y, and I, with X the number of Heads in two fair coin tosses, Y the number of Tails, and I the indicator of Heads on the first toss.

View Larger Image
Image Description

The PMFs of X, Y, and I are plotted in Figure 3.2.4. Vertical bars are drawn to make it easier to compare the heights of different points.

We will now state the properties of a valid PMF.

> **Theorem 3.2.5 (Valid PMFs).**
>
> Let X be a discrete r.v. with support $x_1, x_2, \ldots$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite).
>
> The PMF p_X of X must satisfy the following two criteria:
>
> - Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise;
> - Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

Proof

The first criterion is true since probability is nonnegative. The second is true since X must take on *some* value, and the events $\{X = x_j\}$ are disjoint, so

$$\sum_{j=1}^{\infty} P(X = x_j) = P\left(\bigcup_{j=1}^{\infty}\{X = x_j\}\right) = P(X = x_1 \text{ or } X = x_2 \text{ or } \ldots) = 1.$$

Conversely, if distinct values $x_1, x_2, \ldots$ are specified and we have a function satisfying the two criteria above, then this function *is* the PMF of some r.v.

The PMF is one way of expressing the distribution of a discrete r.v. This is because once we know the PMF of X, we can calculate the probability that X will fall into a given subset of the real numbers by summing over the appropriate values of x. Given a discrete r.v. X and a set B of real numbers, if we know the PMF of X we can find $P(X \in B)$, the probability that X is in B, by summing up the heights of the vertical bars at points in B in the plot of the PMF of X. *Knowing the PMF of a discrete r.v. determines its distribution.*

Example 3.2.6 Poisson Distribution.

An r.v. X has the *Poisson distribution* with parameter λ, where $\lambda > 0$, if the PMF of X is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

We write this as $X \sim \textrm{Pois}(\lambda)$. The Poisson is one of the most widely used distributions in all of statistics, and is a very common choice of model (or building block for more complicated models) for data that *counts* the number of occurences of some kind. The Poisson is discussed in much more detail in Unit 5. The Poisson also arises through the *Poisson process*, a model that is used in a wide variety of problems in which events occur at random points in *time*. Poisson processes are introduced in Unit 4 .