Random Variables and their distributions

We would like to say that a random variable X is a "numerical outcome of a complicated and/or random experiment." This is not sufficient. For example, suppose you sample 1,500 people at random and find that their average age is 25. Is $X = 25$ a "random variable"? Surely there is nothing random about the number 25!

What is "random" here is the procedure that led to the number 25. This procedure, for a second sample, is likely to lead to a different number. Procedures are functions, and thence

Definition 1. A *random variable* is a function X from Ω to some set D which is usually [for us] a subset of the real line $\mathbf{R}$, or d-dimensional space $\mathbf{R}^d$.

In order to understand this, let us construct a random variable that models the number of dots in a roll of a fair six-sided die.

Define the sample space,

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

We assume that all outcome are equally likely [fair die].

Define $X(\omega) = \omega$ for all $\omega \in \Omega$, and note that for all $k = 1, \ldots, 6$,

$$P(\{\omega \in \Omega : X(\omega) = k\}) = P(\{k\}) = \frac{1}{6}. \tag{7}$$

This probability is zero for other values of k. Usually, we write $\{X \in A\}$ in place of the set $\{\omega \in \Omega : X(\omega) \in A\}$. In this notation, we have

$$P\{X = k\} = \begin{cases} \frac{1}{6} & \text{if } k = 1, \ldots, 6, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

This is a math model for the result of a coin toss.

The distribution of a random variable

Suppose X is a random variable, defined on some probability space Ω. By the *distribution* of X we mean the collection of probabilities $P\{X \in A\}$, as A ranges over all sets that could possibly contain X in them.

If X takes values in a finite, or countably-infinite set, then we say that X is a *discrete random variable*. Its distribution is called a *discrete distribution*. The function

$$f(x) = P\{X = x\}$$

is then called the *mass function* of X. Be warned, however, that your textbook does not use this terminology.

The following simple result tells us that if we know the mass function of X, then we know the entire distribution of X as well.

Proposition 1. *If X is a discrete random variable with mass function f, then $P\{X \in A\} = \sum_{z \in A} P\{X = z\}$ for all at-most countable sets A.*

It might be useful to point out that a set A is "at most countable" if A is either a finite set, or its elements can be labeled as $1, 2, \ldots$. [A theorem of G. Cantor states that the real line is *not* at most countable.]

Proof. The event $\{X \in A\}$ can be written as a disjoint union $\cup_{z \in A}\{X = z\}$. Now apply the additivity rule of probabilities. □

An example. Suppose X is a random variable whose distribution is given by

$$P\{X = 0\} = \frac{1}{2}, \quad P\{X = 1\} = \frac{1}{4}, \quad P\{X = -1\} = \frac{1}{4}.$$

What is the distribution of the random variable $Y := X^2$? The possible values of Y are 0 and 1; and

$$P\{Y = 0\} = P\{X = 0\} = \frac{1}{2}, \text{ whereas } P\{Y = 1\} = P\{X = 1\} + P\{X = -1\} = \frac{1}{2}.$$

The binomial distribution

Suppose we perform n independent trials; each trial leads to a "success" or a "failure"; and the probability of success per trial is the same number $p \in (0\,,1)$.

Let X denote the total number of successes in this experiment. This is a discrete random variable with possible values $0, \ldots, n$. We say then that X is a binomial random variable ["$X = \text{Bin}(n\,,p)$"].

Math modelling questions:

- Construct an Ω.
- Construct X on this Ω.

We saw in Lecture 7 that

$$P\{X = x\} = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

An example. Consider the following sampling question: *Ten percent of a certain population smoke. If we take a random sample [without replacement] of* 5 *people from this population, what are the chances that at least 2 people smoke in the sample?*

Let X denote the number of smokers in the sample. Then $X = \text{Bin}(n\,, p)$ ["success" = "smoker"]. Therefore,

$$\begin{aligned} P\{X \geq 2\} &= 1 - P\{X \leq 1\} \\ &= 1 - P\left(\{X = 0\} \cup \{X = 1\}\right) \\ &= 1 - \left[p(0) + p(1)\right] \\ &= 1 - \left[\binom{5}{0} 0.1^0 (1-0.1)^{5-0} + \binom{5}{1} 0.1^1 0.9^{5-1}\right] \\ &= 1 - 0.9^5 - \left(5 \times 0.1 \times 0.9^4\right). \end{aligned}$$

Alternatively, we can write

$$P\{X \geq 2\} = P\left(\{X = 2\} \cup \cdots \{X = n\}\right) = \sum_{j=2}^{5} P\{X = j\},$$

and then plug in $P\{X = j\} = \binom{5}{j} 0.1^j 0.9^{5-j}$ for $j = 0, \ldots, 5$.

Joint distributions

Suppose X and Y are both random variables defined on the same sample space Ω [for later applications, this means that X and Y are defined simultaneously in the same problem]. Then their *joint distribution* is the collection of all probabilities of the form $P\{X \in A\,, Y \in B\}$, as A and B range over all possible sets, respectively in which X and Y could possibly land. An argument similar to the one that implied Proposition 1 shows that

$$P\{X \in A\,, Y \in B\} = \sum_{x \in A, y \in B} P\{X = x\,, Y = y\}.$$

Therefore, in order to know the joint distribution of X and Y it suffices to know the *joint mass function*

$$f(x\,,y) = P\{X = x\,, Y = y\} \qquad \text{for all } x, y.$$

[Warning: Your textbook writes $f(x\,,y)$ as $P(x\,,y)$, and does not refer to it as a joint mass function.]

If we know the joint mass function, then we can compute the individual [also known as *marginal*] mass functions of X and Y as follows:

$$P\{X = x\} = \sum_b P\{X = x\,, Y = b\}$$

$$P\{Y = y\} = \sum_c P\{X = c\,, Y = y\}.$$

The term "marginal" comes from examples of the following type:

An example (Two draws at random, Pitman, p. 144). We make two draws at random, without replacement, from a box that contains tickets numbered 1, 2, and 3. Let X denote the value of the first draw and Y the value of the second draw. The following tabulates the function $f(x\,,y) = P\{X = x\,, Y = y\}$ for all possible values of x and y:

		possible value for X			dist of Y
		1	2	3	(row sums)
possible values for Y	3	1/6	1/6	0	1/3
	2	1/6	0	1/6	1/3
	1	0	1/6	1/6	1/3
	dist of X (column sums)	1/3	1/3	1/3	1 (total sum)

Thus, for example,

$$P\{X = Y\} = f(1\,,1) + f(2\,,2) + f(3\,,3) = 0 + 0 + 0 = 0,$$

$$P\{X > Y\} = f(2\,,1) + f(3\,,1) + f(3\,,2) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

and

$$P\{X + Y = 3\} = f(1\,,2) + f(2\,,1) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

In fact, the mass function for $Z := X + Y$ is given below:

$$P\{Z = 3\} = \frac{1}{3}, \;\; P\{Z = 5\} = \frac{2}{3}.$$

When we know the joint mass function of X and Y, we are also able to make conditional computations. For instance,

$$P(Y = y \mid X = x) = \frac{P\{X = x\,, Y = y\}}{P\{X = x\}} = \frac{f(x\,,y)}{\sum_b P\{X = x\,, Y = b\}} = \frac{f(x\,,y)}{\sum_b f(x\,,b)}.$$

Note that as a function of y, the preceding gives a distribution of probabilities. But as a function of x it has no particular structure.

More on the previous example. In the previous example,

$$P(Y = 1 \mid X = 2) = \frac{f(2\,,1)}{f(2\,,1) + f(2\,,2) + f(2\,,3)} = \frac{1/6}{1/3} = \frac{1}{2}.$$

And

$$P(Y = 3 \mid X = 2) = \frac{1}{2} \qquad \text{also.}$$

In other words:

(1) The [marginal] distribution of Y is 1/3 probability on each of the possible values 1, 2, and 3;

(2) However, if we are told that $X = 2$, then the [conditional] distribution of Y is 1/2 probability on the values 1 and 3 each.

(3) What if we are told that $X = 3$? How about the conditional distribution of X, given that $Y = 1$?

Independence

If X and Y are jointly distributed random variables, then [by the conditional probabilities formula]

$$\begin{aligned} P\{X = x\,, Y = y\} &= P\{X = x\} \cdot P(Y = y \mid X = x) \\ &= P\{Y = y\} \cdot P(X = x \mid Y = y). \end{aligned} \tag{9}$$

Definition 2. X and Y are said to be *independent* if

$$P\{X = x\,, Y = y\} = P\{X = x\} \cdot P\{Y = y\} \qquad \text{for } \underline{\text{all}} \text{ possible } x, y.$$

Thanks to (9):

(1) X and Y are independent if

$$P(X = x \mid Y = y) = P\{X = x\} \qquad \text{for all possible values of } x, y.$$

(2) Equivalently, X and Y are independent if

$$P(Y = y \mid X = x) = P\{Y = y\} \qquad \text{for all possible values of } x, y.$$

By Rule 3 of probabilities,

$$P\{X \in A\,, Y \in B\} = \sum_{x \in A} \sum_{y \in B} \mathrm{P}\{X = x\,, Y = y\}.$$

Therefore, if X and Y are independent, then $P\{X \in A\,, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}$. And the converse is obvious if you consider sets A and B that have one element in them each.

Several random variables

If $X_1, \ldots, X_n$ are random variables, then we can consider their joint mass function

$$f(x_1, \ldots, x_n) := P\{X_1 = x_1, \ldots, X_n = x_n\}.$$

And $X_1, \ldots, X_n$ are *independent* if

$$f(x_1, \ldots, x_n) = P\{X_1 = x_1\} \times \cdots \times P\{X_n = x_n\} \qquad \text{for all } x_1, \ldots, x_n.$$