Lecture 8

1. Random Variables

We would like to say that a random variable X is a "numerical outcome of a complicated 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? The procedure that led to 25. This procedure, for a second sample, is likely to lead to a different number. Procedures are functions, and thence

Definition 8.1. 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 $k = 1, \ldots, 6$,

$$\mathrm{P}\left(\{\omega \in \Omega : X(\omega) = k\}\right) = \mathrm{P}(\{k\}) = \frac{1}{6}. \tag{5}$$

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

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

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

2. General notation

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

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

$$p(x) = \mathrm{P}\{X = x\}$$

is then called the *mass function* of X. Note that $p(x) = 0$ for all but a countable number of values of x. The values x for which $p(x) > 0$ are called the *possible values* of X.

Some important properties of mass functions:

- $0 \le p(x) \le 1$ for all x. [Easy]
- $\sum_x p(x) = 1$. Proof: $\sum_x p(x) = \sum_x \mathrm{P}\{X = x\}$, and this is equal to $\mathrm{P}(\cup_x \{X = x\}) = \mathrm{P}(\Omega)$, since the union is a countable disjoint union.

3. 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 ["$X = \mathrm{Bin}(n\,,p)$"].

Math modelling questions:

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

Let us find the mass function of X. We seek to find $p(x)$, where $x = 0, \ldots, n$. For all other values of x, $p(x) = 0$.

Now suppose x is an integer between zero and n. Note that $p(x) = \mathrm{P}\{X = x\}$ is the probability of getting exactly x successes and $n-x$ failures. Let S_i denote the event that the ith trial leads to a success. Then,

$$p(x) = \mathrm{P}\left(S_1 \cap \cdots \cap S_x \cap S_{x+1}^c \cap \cdots S_n^c\right) + \cdots$$

where we are summing over all possible ways of distributing x successes and $n - x$ failures in n spots. By independence, each of these probabilities is

$p^x(1-p)^{n-x}$. The number of probabilities summed is the number of ways we can distributed x successes and $n-x$ failures into n slots. That is, $\binom{n}{x}$. Therefore,

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

Note that $\sum_x p(x) = 1$ by the binomial theorem. So we have not missed anything.

3.1. 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 $X = \mathrm{Bin}(n\,,p)$ ["success" = "smoker"]. Therefore,

$$\begin{aligned}
\mathrm{P}\{X \geq 2\} &= 1 - \mathrm{P}\{X \leq 1\} \\
&= 1 - \mathrm{P}\left(\{X = 0\} \cup \{X = 1\}\right) \\
&= 1 - [p(0) + p(1)] \\
&= 1 - \left[\binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1}\right] \\
&= 1 - \left[1 - p + np(1-p)^{n-1}\right] \\
&= p - np(1-p)^{n-1}.
\end{aligned}$$

Alternatively, we can write

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

and then plug in $p(j) = \binom{n}{j} p^j (1-p)^{n-j}$.

4. The geometric distribution

A *p-coin* is a coin that tosses heads with probability p and tails with probability $1-p$. Suppose we toss a p-coin until the first time heads appears. Let X denote the number of tosses made. Then X is a so-called geometric random variable ["$X = \mathrm{Geom}(p)$"].

Evidently, if n is an integer greater than or equal to one, then $\mathrm{P}\{X = n\} = (1-p)^{n-1}p$. Therefore, the mass function of X is given by

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{if } x = 1, 2, \ldots\,, \\ 0 & \text{otherwise.} \end{cases}$$