Discrete Random Variables

So far, we have described probability in terms of events. How are events demarcated? The answer is "by the values of random variables."

One way to think about a random variable is that it is the result of an experiment which has some "random"-ness intrinsic to it. No one would like to admit that their "controlled" experiments have such a random element in them. However, in most experiments there are factors that are beyond the control of the experimenter that contribute to the fact that the outcome is not as deterministic as idealized. A typical example is the standard pendulum experiment in physics which does not give *exactly* the same answer every time.

In general, a random variable can take values anywhere, in integers, in real numbers, in a group and so on. In most cases, in this course we will deal with real-valued random variables. In this lecture, we will primarily deal with the case when the random variable takes "discrete", well-separated values. Usually, these values will be among the integers as those are mathematically nice to deal with.

For example, we have a random variable F associated with a coin flip which takes two values 0 for Tail and 1 for Head; so the event H of Head is the same as $F = 1$.

However, random variables are also most useful when the possible values are infinite. Consider the experiment where we repeatedly flip an unbiased coin until we get a Head. Let W be the random variable that counts the number of flips needed. The event $W = n$ is the same as $H_1^c \wedge \cdots \wedge H_{n-1}^c \wedge H_n$ where H_i represents Head on the i-flip.

As another example, consider the random variable C_k that counts the number of heads obtained in k independent flips of the same coin. If the random variables for the result of each flip in sequence are denoted $F_1, F_2, \ldots, F_k$. Then $C_k = F_1 + \cdots + F_k$.

If we conduct an experiment of asking a randomly chosen student from the class, her/his Hostel of residence, the result R is a random variable as well. The values of R are in the set $\{5, 6, 7, 8\}$ which is discrete.

Probability distribution of a Random Variable

Given a random variable X, an event is described by giving a (restricted) range of values for it; hence there is a probability associated with it.

For example, $P(F = 1) = 1/2$ and $P(F = 0) = 1/2$ for a fair coin. If the coin is not fair, then we have $P(F = 1) = p$ and $P(F = 0) = 1 - p$, where p is the probability of a head so that $(1 - p)$ is the probability of a tail.

As seen above $W = n$ is the event $H_1^c \wedge \cdots \wedge H_{n-1}^c \wedge H_n$ so if $P(H_i) = p$ and the distinct H_i are independent, we see that $P(W = n) = (1-p)^{n-1}p$.

In the previous section we had seen that if 40 students out of 200 are in Hostel 5, then $P(R = 5) = 40/200 = 1/5$.

Also in the previous section we had calculated (for a fair coin) that

$$P(C_k = r) = \frac{\binom{k}{r}}{2^k}$$

We note the identity

$$2^k = (1+1)^k = \sum_{r=0}^{k} \binom{k}{r}$$

This shows us that

$$\sum_{r=0}^{k} P(C_k = r) = 1$$

More generally, when a random variable X takes discrete values (for simplicity we will always assume these are labelled by the integers!), then we probabilities $P(X = i) = p_i$ which satisfy $0 \leq p_i \leq 1$. Now, it is clear that the events $X = i$ for different i are mutually exclusive and, as we run over all i are exhaustive. Thus, we have the condition that $\sum_i p_i = 1$. In fact, if we have an event E that is described by some well-prescribed *subset* of values for X, then $P(E) = \sum_{i \in E} p_i$. The behaviour of the random variable X is largely determined by the collection $(p_i)_i$ which is called its probability distribution. We can "plot" it like the histograms of frequency we plotted earlier.

For example, suppose that S represents the *discrete* collection of sequences of length k of heads and tails. These are all the values of a random variable X that *records* the sequence of heads and tails obtained in an experiment that has performs k independent flips of a coin. Suppose that the probability of obtaining head on that coin is p. For each sequence b, the probability $P(X = b)$ is given by $p^{r(b)}(1-p)^{k-r(b)}$ where $r(b)$ is the number of heads in the sequence b. Now, suppose that as above Y_k is the random variable that counts the number of heads obtained. We see that $Y_k = r$ is the union of the events $X = b$ where b is such that $r(b) = r$. Now, these are mutually exclusive events with $P(X = b) = p^r(1-p)^{k-r}$ *independent* of b! It follows that we can calculate $P(Y_k = r)$ by counting as before

$$P(Y_k = r) = \binom{k}{r} p^r (1-p)^{k-r}$$

We note that

$$\sum_{r=0}^{k} P(Y_k = r) = \sum_{r=0}^{k} \binom{k}{r} p^r (1-p)^{k-r} = 1$$

as expected. The distribution as above is called the Binomial distribution and we say that Y_k is a variable that follows the Bionomial distribution.

Algebra of Random Variables

Whatever algebraic rules exist for combinations of the *values* of random variables also exist for combinations of the random variables themselves. Specifically, if X, Y are real-valued random variables, then we can form cX (for a real number c), $X + Y$, $X \cdot Y$, $\min X, Y$ and so on. If $P(Y = 0) = 0$, then we also have X/Y. There is also a "constant" random variable c which takes the value c with probability 1 and any other value with probability 0.

An important example, is that of the distribution function for the random variable C_k that counts the number of Heads in k flips. As we saw above, $C_k = F_1 + F_2 + \cdots + F_k$, where F_i is the random variable associated with each flip.

This example can be used to understand that F_{X+Y} is *not* $F_X + F_Y$; in fact, the latter is *not* the distribution function of a random variable since it can take values bigger than 1!

Our further study of probability will be based on random variables and the events defined by restrictions on the values of random variables.