10 Introduction to Random Variables

Suppose we have a sample space S. Sometimes we are interested not in the exact outcome but only some consequence of it (for example I toss 3 coins but I only care about how many heads occur not the exact outcome). In this sort of situation random variables are useful.

Definition A *random variable* is a function from S to $\mathbb{R}$.

Notation We usually use capital letters for random variables e.g. X, Y, Z, even though they are functions.

Remark An (informal) way of thinking is to regard a random variable as a question about the outcome which has an answer which is a real number (e.g. how many heads occur in a coin tossing experiment) or a measurement made on the outcome.

Given a random variable X on S, statements like "$X = 3$" or "$X \leq 3$" are events. Specifically "$X = 3$" is the event $\{s \in S : X(s) = 3\}$ i.e. the set of all outcomes in S for which which X takes the value 3. Similarly "$X \leq 3$" is the event $\{s \in S : X(s) \leq 3\}$ i.e. that is the set of all outcomes in S for which X takes a value at most 3.

Definition For a random variable X, the *range of* X is the set of all values taken by X. We denote it by $\mathit{Range}(X)$. Thus

$$\mathit{Range}(X) = \{x \in \mathbb{R} \ : \ X(s) = x \text{ for some } s \in S\}.$$

Definition Let X be a random variable X. The *probability mass function* (or pmf) of X is the function from $\mathit{Range}(X)$ to $\mathbb{R}$ given by

$$x \mapsto \mathbb{P}(X = r) \text{ for each } r \in \mathit{Range}(X).$$

Let X be a random variable X. The *cumulative distribution function* (or cdf) of X is the function from $\mathbb{R}$ to $\mathbb{R}$ given by

$$x \mapsto \mathbb{P}(X \leq r) \text{ for each } r \in \mathbb{R}.$$

Note that we are taking the domain of the pmf of X to be $\mathit{Range}(X)$ and the domain of the cdf of X to be the whole of $\mathbb{R}$.

Example A biased coin which has probability p of coming up heads is tossed three times. Let X be the number of heads observed. So for example we have

$X(hht) = 2$ (remember that X is a function from S to $\mathbb{R}$ so each element of S is mapped by X to a real number, in this case to 2). The event "$X = 2$" is the set $\{hht, hth, thh\}$ so

$$\mathbb{P}(X = 2) = \mathbb{P}(\{hht, hth, thh\}) = 3p^2(1-p).$$

We have $Range(X) = \{0, 1, 2, 3\}$. We can work $\mathbb{P}(X = r)$ for the other values of $r \in \{0, 1, 2, 3\}$ and obtain the following probability mass function:

r :	0	1	2	3
$P(X = r)$:	$(1-p)^3$	$3p(1-p)^2$	$3p^2(1-p)$	p^3

Definition Let X be a random variable defined on a sample space S. We say X is *discrete* if there exists a real number $c > 0$ such that $|r_1 - r_2| \geq c$ for all $r_1, r_2 \in Range(X)$.

Informally X is discrete if the values X takes are separated by "gaps". It follows that if $Range(X)$ is finite, or $Range(X) = \mathbb{N}$ (for example when X equals the number of tosses of a coin until the first head is seen), then X is discrete. If $Range(X)$ is equal to a subinterval of $\mathbb{R}$ such as $\{x \in \mathbb{R} \,:\, x \geq 0\}$ (for example the time I wait until the DLR arrives) then X is not discrete.

Remark An important property of a discrete random variable X is that its range is either finite or countably infinite. This means that if we know the pmf for X then we can use Kolmogorov's third axiom to calculate the probability of any event by adding together the probabilities of the outcomes which belong to it. In particular, for any $r_0 \in \mathbb{R}$, we have

$$\mathbb{P}(X \leq r_0) = \sum_{r \in Range(X),\, r \leq r_0} \mathbb{P}(X = r)$$

so we can calculate the cdf of X from its pmf. We saw an example in lectures which showed that this need not be true when X is not discrete.

11 Discrete Random variables

In this section we define the expectation and variance of a discrete random variable and deduce some basic properties.

Proposition 11.1. *Let X be a discrete random variable on a sample space S. Then*

$$\sum_{r \in Range(X)} \mathbb{P}(X = r) \;=\; 1$$

Proof The fact that the range of X is either finite or countably infinite means that we can use Kolmogorov's second and third axioms to deduce that

$$\sum_{r \in Range(X)} \mathbb{P}(X = r) \;=\; P(S) \;=\; 1$$

•

This proposition is a useful to check we have calculated the probability mass function of X correctly.

Definition The *expectation* of a discrete random variable X is

$$\mathrm{E}(X) = \sum_{r \in Range(X)} r\, \mathbb{P}(X = r)$$

Thus the expectation of X is a 'weighted average' of the values taken by X, where each value $r \in Range(X)$ is weighted by the probability it occurs.

Sometimes it is useful to consider functions of random variables. If X is a random variable on S and $g : \mathbb{R} \to \mathbb{R}$ then $Y = g(X)$ is a new random variable on X which maps S into $\mathbb{R}$ by $Y(s) = g(X(s))$ for all outcomes $s \in S$. For example $Y = 2X - 7$ or $Z = X^2$ are both new random variables on S.

Proposition 11.2 (Properties of expectation)**.** *Let X be a discrete random variable.*

(a) Suppose $b, c \in \mathbb{R}$. Then $\mathrm{E}(bX + c) = b\mathrm{E}(X) + c$.

(b) If $m, M \in \mathbb{R}$ and $m \leq X(s) \leq M$ for all $s \in S$ then

$$m \leq \mathrm{E}(X) \leq M.$$

(c) If there exists $k \in \mathbb{R}$ with $\mathbb{P}(X = k - t) = \mathbb{P}(X = k + t)$ for all $t \in \mathbb{R}$ then $\mathrm{E}(X) = k$.

Proof (a) We have

$$\begin{aligned}
\mathrm{E}(bX+c) &= \sum_{r\in Range(X)} (br+c)\mathbb{P}(X=r) \\
&= \sum_{r\in Range(X)} br\mathbb{P}(X=r) + \sum_{r\in Range(X)} c\mathbb{P}(X=r) \\
&= b\sum_{r\in Range(X)} r\mathbb{P}(X=r) + c\sum_{r\in Range(X)} \mathbb{P}(X=r) \\
&= b\mathrm{E}(X)+c
\end{aligned}$$

where the fourth equality uses the definition of $\mathrm{E}(X)$ for the first term and Proposition 11.1 for the second term.

(b) See Exercises 7, Q3.

(c) We have

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{r\in Range(X)} r\mathbb{P}(X=r) \\
&= k\mathbb{P}(X=k) + \sum_{\substack{t>0\\ k-t\in Range(X)}} [(k-t)\mathbb{P}(X=k-t)+(k+t)\mathbb{P}(X=k+t)] \\
&= k\mathbb{P}(X=k) + \sum_{\substack{t>0\\ k-t\in Range(X)}} [k\mathbb{P}(X=k-t)+k\mathbb{P}(X=k+t)] \\
&= k\sum_{r\in Range(X)} \mathbb{P}(X=r) \\
&= k
\end{aligned}$$

where:

- the second and fourth equalities follow by putting $r=k-t$ when $r<k$ and $r=k+t$ when $r>k$;
- the third equality holds because $\mathbb{P}(X=k-t)=\mathbb{P}(X=k+t)$ so the terms involving t cancel;
- the fifth equality uses Proposition 11.1.

•

Definition Let X be a discrete random variable with $\mathrm{E}(X) = \mu$. Then the *variance* of X is

$$\mathrm{Var}(X) = \sum_{r \in Range(X)} [r - \mu]^2 \, \mathbb{P}(X = r).$$

Thus $\mathrm{Var}(X)$ is the expected value of $[X - \mu]^2$ i.e. the square of the distance from X to $E(X)$. It measures how the probability distribution of X is concentrated about its expectation. A small variance means X is sharply concentrated and a large variance means X is spread out.