MA181 INTRODUCTION TO STATISTICAL MODELLING
RANDOM VARIABLES

Definition A *random variable* X is a real-valued function defined on a sample space.

Examples 1. The number on a die,

2. Height of a 5 year old boy,

3. $x = \begin{cases} 0, & \text{black hair}, \\ 1, & \text{brown hair}, \\ 12, & \text{red/fair hair}. \end{cases}$

If the range of X is discrete, i.e. it consists of a finite or countably infinite number of points, then X is called a *discrete* random variable, otherwise it is said to be *continuous*.

Notation We let the upper case letter (X) denote the definition of a random variable and a lower case letter (x) denote a value taken by it. Thus it makes sense to ask if $X = x$ or if $Y = 3$.

Probability function

Definition Let $p(x) = P(X = x)$. Then $p(x)$ is called the *probability function* (pf) of X.

Notes 1. $0 \leq p(x) \leq 1$ for all x,

2. $\sum_{\text{all } x} p(x) = 1$.

Example $p(x) = \begin{cases} 0.3, & x = 1 \\ 0.4, & x = 3, \\ 0.1, & x = 6, \\ 0.2, & x = 10. \end{cases}$

Distribution function

Definition Let $F(x) = P(X \leq x)$. Then $F(x)$ is called the *(cumulative) distribution function* (cdf) of X.

Notes 1. $F)(x)$ is defined over the whole line, i.e. for $-\infty < x < \infty$.

2. $F(x)$ is a monotonic increasing function of x such that $F(-\infty) = 0$ and $F(\infty) = 1$.

3. $F(x)$ is continuous from the right but not necessarily from the left. So, for $\varepsilon > 0$, $F(x) = \lim_{\varepsilon \to 0} F(x + \varepsilon)$.

Example Consider the example with the probability function defined above. Then

$$F(x) = \begin{cases} 0, & -\infty < x < 1, \\ 0.3, & 1 \le x < 3, \\ 0.7, & 3 \le x < 6, \\ 0.8, & 6 \le x < 10, \\ 1, & 10 \le x < \infty. \end{cases}$$

Note that, if X takes integer values, $p(x) = F(x) - F(x-1)$.

Transformations of random variables Occasionally we may know the distribution of a random variable X but require the distribution of a function Y of X. If the function is one-one, then the problem is easily solved.

Example Let X be the number of heads showing when four coins are tossed and let $Y = 3X + 5$ be my winnings. Then $P(Y = y) = P[X = (y-5)/3]$, so that

$$p_X(x) = \begin{cases} \frac{1}{16}, & x = 0, \\ \frac{1}{4}, & x = 1, \\ \frac{3}{8}, & x = 2, \\ \frac{1}{4}, & x = 3, \\ \frac{1}{16}, & x = 4. \end{cases} \quad \text{and} \quad P_Y(y) = \begin{cases} \frac{1}{16}, & x = 5, \\ \frac{1}{4}, & x = 8, \\ \frac{3}{8}, & x = 11, \\ \frac{1}{4}, & x = 14, \\ \frac{1}{16}, & x = 17. \end{cases}$$

Many-one functions need a little more care. Suppose $Y = X^2$. Then, for example, $P(Y = 16) = P(X^2 = 16) = P(X = -4) + P(X = 4)$ since there are two mutually exclusive ways for Y to equal 16. Similarly, $P(Y \le 25) = P(-5 \le X \le 5)$.