Chapter 2

Discrete Random Variables

2.1 Definition and Distributions

We are often interested in a function of the outcome of a random experiment, rather than in the outcome itself.

Examples

(i) Suppose two fair distinguishable dice are tossed: we might be interested in the sum of scores. The sample space is

$$\mathcal{S} = \{(d_1, d_2); \quad d_1, d_2 = 1, ..., 6\}$$

and the sum $x = d_1 + d_2, \qquad x = 2, ..., 12.$

(ii) Suppose we are interested in whether the lifetime of a piece of equipment exceeds 100 hours. The sample space is

$$\mathcal{S} = \{t; \quad 0 \leq t < \infty\}$$

and we could define

$$x = \begin{cases} 1, & \text{if } t > 100 \\ 0, & \text{if } t \leq 100 \end{cases}.$$

Given a probability space $(\mathcal{S}, \mathcal{F}, \mathrm{P})$, a *discrete random variable* X is defined to be a mapping of $\mathcal{S}$ into the set R of real numbers (i.e. with every $E \in \mathcal{S}$ there is associated a real number $X(E)$), such that

(i) $\mathcal{S}$ is mapped into a *countable* set of real numbers, D (the *image* of $\mathcal{S}$ under X);

(ii) if A_x denotes the subset of outcomes in $\mathcal{S}$ which are mapped into a real number x (i.e. $A_x = \{E \in \mathcal{S} : X(E) = x\}$), then

$$A_x \in \mathcal{F} \qquad \text{for all } x \in D$$

(or, indeed, for all $x \in \mathcal{R}$, since $A_x = \emptyset$ if $x \notin D$, and always $\emptyset \in \mathcal{F}$.

Then the probability that X takes the value x, denoted by $\mathrm{P}(X = x)$, is given by

$$\boxed{\mathrm{P}(X = x) = \mathrm{P}(A_x). \qquad (2.1)}$$

It is common to speak of the event '$X = x$', though we should always remember that we are really referring to the event A_x in our event space $\mathcal{F}$.

Let $D = \{x_1, x_2, ...\}$ (where $x_1, x_2, ...$ are assumed to be in increasing order).

The *probability distribution* of X is given by

$$\boxed{\mathrm{P}(X = x_i) = p_X(x_i), \qquad i = 1, 2, ... \qquad (2.2)}$$

The function $p_X(\cdot)$ is called the *probability (mass) function* of X: it maps $\mathcal{R}$ into $[0, 1]$.

(The suffix 'X' may be dropped when there is no ambiguity as to the random variable concerned). We observe that $p_X(\cdot)$ has the following properties (readily derived from those of P):

(i) $p_X(x) = 0$ if $x \notin D$ (since then $A_x = \emptyset$); (2.3)

(ii) $p_X(x_i) = \mathrm{P}(A_{x_i}) \geq 0, \qquad i = 1, 2, ...;$ (2.4)

(iii) $\sum\limits_{x \in D} p_X(x) = \mathrm{P}(\bigcup\limits_{x \in D} A_x) = \mathrm{P}(\mathcal{S}) = 1.$ (2.5)

The *cumulative (probability) distribution function* of X is given by

$$\begin{aligned} F_X(y) &= \mathrm{P}(E \in \mathcal{S} : X(E) \leq y) \\ &= \mathrm{P}(\bigcup_{x \in D; x \leq y} A_x) \\ &= \sum_{x \in D; x \leq y} \mathrm{P}(A_x) \end{aligned}$$

i.e.

$$\boxed{F_X(y) = \sum_{x_i \leq y} p_X(x_i), \qquad -\infty < y < \infty \qquad (2.6)}$$

and we often simply write

$$\boxed{F_X(y) = \mathrm{P}(X \leq y), \qquad (2.7)}$$

where again $X \leq y$ is 'shorthand' for $\{E \in \mathcal{S} : X(E) \leq y\}$).

$F_X(\cdot)$ is a step (or staircase) function which is continuous from the right but not from the left: the size of the 'step' at x_i is $p_X(x_i)$. Also it is easily shown that

$$\boxed{\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a) \qquad (a < b). \qquad (2.8)}$$

Transformations Given a function $g(X)$ of X, it is easily shown that $Y = g(X)$ is also a discrete random variable. Let x_{i_1}, x_{i_2}, ...be the values of X having the property

$$g(x_{i_j}) = y_i \text{ for } j = 1, 2, ...$$

Then

$$\mathrm{P}(Y = y_i) = \mathrm{P}(X = x_{i_1}) + \mathrm{P}(X = x_{i_2}) + ... \tag{2.9}$$

(Note: In future, we shall generally adopt the abbreviation 'r.v.' for random variable).

2.2 Expectation

The *expected value* or *expectation* of a discrete r.v. X, denoted by $\mathrm{E}(X)$ or μ, is defined as

$$\mathrm{E}(X) = \sum_{x \in D} x\mathrm{P}(X = x) = \sum_i x_i \mathrm{P}(X = x_i) = \sum_i x_i p_X(x_i) \tag{2.10}$$

provided this series is absolutely convergent (i.e. $\sum_i |x_i p_X(x_i)|$ is finite, so that when D is an infinite set, the series takes the same value irrespective of the order in which we add up the terms). $\mathrm{E}(X)$ is also referred to as the *population mean* or the *mean of the distribution* or the *mean* of X. It is analogous to the idea of 'centre of gravity' in mechanics.

The E operator has the following properties:

(i) If a and b are constants, $\mathrm{E}(a + bX) = a + b\mathrm{E}(X)$. (2.11)

(ii) If $L \leq X \leq U$, then $L \leq \mathrm{E}(X) \leq U$. (2.12)

(iii) $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$. (2.13)

(iv) If $Y = g(X)$, it is readily proved that

$$\boxed{\mathrm{E}(Y) = \mathrm{E}[g(X)] = \sum_{x \in D} g(x)\mathrm{P}(X = x) = \sum_i g(x_i)\mathrm{P}(X = x_i), \qquad (2.14)}$$

provided again that the series is absolutely convergent. Note that this is a *result* of the original definition of $\mathrm{E}(X)$ and not a *definition* in its own right: Ross terms this the 'law of the unconscious statistician' (since many people think it is a definition). The result is very useful, since it enables us to find $\mathrm{E}[g(X)]$ without first finding the probability distribution of $Y = g(X)$ in order to determine $\mathrm{E}(Y) = \sum_j y_j p_Y(y_j)$.

(v) $\mathrm{E}\{g(X) + h(X)\} = \mathrm{E}\{g(X)\} + \mathrm{E}\{h(X)\}$. (2.15)

The *variance* of the r.v. X is defined as

$$\boxed{\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[\{X - \mathrm{E}(X)\}^2] = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2, \qquad (2.16)}$$

and its positive square root, σ, is called the *standard deviation* of X (or the *population standard deviation* or *standard deviation of* X). It has the following properties:

(i) $\mathrm{Var}(X) \geq 0$. (2.17)

(ii) $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$. (2.18)

(iii) If $\mathrm{Var}[g(X)] = 0$, then $g(X) =$ constant. (2.19)

The rth *moment about the origin* is defined as $\mathrm{E}(X^r)$; the rth *moment about the mean* is defined as $\mathrm{E}[(X - \mu)^r]$, $\quad r = 1, 2, ...$ Usually we are interested in at most the first four moments.

2.3 Important discrete distributions

Here we review the common discrete distributions, nearly all of which have already been studied at Level 1.

Consider a sequence of repeated independent trials where there are only two possible outcomes for each trial, 'success' and 'failure', and the probabilities of success and failure, p and $q = 1 - p$, remain the same throughout the trials (Bernoulli sequence). Then the following discrete distributions are defined on such an experiment:

Binomial distribution

Let X be the number of successes in n trials, where n is fixed. Then

$$\mathrm{P}(X = x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, ..., n \tag{2.20a}$$

and

$$\mathrm{E}(X) = np, \qquad \mathrm{Var}(X) = npq. \tag{2.20b}$$

Geometric distribution

Let X be the number of trials required to obtain the first success. Then

$$\mathrm{P}(X = x) = pq^{x-1}, \qquad x = 1, 2, ... \tag{2.21a}$$

and

$$\mathrm{E}(X) = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{q}{p^2}. \tag{2.21b}$$

If on the other hand we define X to be the number of failures obtained before the first success is obtained, X has the *modified geometric distribution*

$$\mathrm{P}(X = x) = pq^x, \qquad x = 0, 1, ... \tag{2.22a}$$

and

$$\mathrm{E}(X) = \frac{1}{p} - 1 = \frac{q}{p}, \tag{2.22b}$$

the variance being unchanged.

Negative binomial distribution

Let X be the number of trials required to obtain r successes, where r is fixed. Then

$$\mathrm{P}(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \qquad x = r, r+1, ... \tag{2.23a}$$

and

$$\mathrm{E}(X) = \frac{r}{p}, \quad \mathrm{Var}(X) = \frac{rq}{p^2}. \tag{2.23b}$$

Example 2.1 *The Banach Match Problem*

A pipe-smoking mathematician carries two match-boxes, one in his left-hand pocket and the other in his right-hand pocket: initially each box contains N matches. Each time the mathematician requires a match he is equally likely to take it from either box. We ask: at the moment when he first *discovers* one of the boxes to be empty, what is the probability that there are exactly k matches in the other box, $k = 0, 1, ..., N$?

Solution Let A denote the event that the mathematician first discovers that the right-hand box is empty and there are k matches in the left-hand box at that moment. Think of choosing the right-hand pocket as a 'success' ($p = \frac{1}{2}$). Then A will occur if and only if exactly $N - k$ 'failures' precede the $(N + 1)$th 'success', i.e. the $(N + 1)$th 'success' occurs at trial number $(2N - k + 1)$. Using (2.23a) with $p = \frac{1}{2}, r = N + 1, x = 2N - k + 1$, we deduce that

$$\mathrm{P}(A) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1}.$$

Since there is an equal probability that it is the left-hand box that is first discovered to be empty and there are k matches in the right-hand box at that moment, the required probability is

$$2\mathrm{P}(A) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}. \qquad \square$$

Other common discrete distributions include:

Poisson distribution

A Poisson r.v. has

$$\mathrm{P}(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad x = 0, 1, ...; \quad \lambda > 0, \tag{2.24a}$$

and

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda. \tag{2.24b}$$

The Poisson distribution with $\lambda = np$ provides a useful approximation to the binomial distribution when n is large and p is small such that $np \leq 5$.

Hypergeometric distribution

Suppose a population of N items consisting of N_1 type 1 items and $N_2 = N - N_1$ type 2 items is randomly sampled n times *without replacement*. Let X be the number of type 1 items in the sample. Then

$$\mathrm{P}(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \tag{2.25a}$$

where $0 \leq x \leq N_1, 0 \leq n-x \leq N_2, 0 \leq n \leq N$ i.e. $x = \max(0, n-N_2), ..., \min(n, N_1)$. In practice, however, n is usually smaller than N_1, N_2, so that $x = 0, ..., n$. Also

$$\mathrm{E}(X) = \frac{nN_1}{N}, \qquad \mathrm{Var}(X) = \frac{nN_1N_2(N-n)}{N^2(N-1)}. \tag{2.25b}$$

When n is small compared with N_1, N_2, the hypergeometric distribution can (as expected) be approximated by the binomial distribution with parameters n and $p = \dfrac{N_1}{N}, q = \dfrac{N_2}{N}$.

Example 2.2 *Estimating the Size of an Animal Population*

The number of animals inhabiting a certain region (N, say) is unknown. To estimate N, an ecologist catches m animals, marks them and then releases them. After giving the marked animals time to disperse through the region, a second catch of n animals is carried out. Stating any assumptions made, give the probability distribution of X, the number of marked animals in the second catch. If in fact $X = k$, find the corresponding maximum likelihood estimate of N, i.e. the value of N which maximises $\mathrm{P}(X = k)$.

Solution We assume that

(i) the number of animals in the region remains unchanged between the times of the first and second catches;

(ii) each time an animal is caught, it is equally likely to be any one of the remaining uncaught animals.

Then X is a hypergeometric r.v. with

$$\mathrm{P}(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, \quad i = 0, 1, ..., n.$$

It is convenient to write $\mathrm{P}(X = i) \equiv p_i(N)$. To find the (integer) value of N which maximises $p_k(N)$, we observe that

$$\frac{p_k(N)}{p_k(N-1)} = \frac{(N-m)(N-n)}{N(N-m-n+k)}$$

which is ≥ 1 iff $N \leq \dfrac{mn}{k}$. So the maximum likelihood estimate of N is $\left\lfloor \dfrac{mn}{k} \right\rfloor$.

(The same result is obtained if we suppose that the proportion of marked animals in the second catch is approximately the same as the proportion in the region as a whole.) ♢

Uniform distribution

Here X takes any one of a finite set of values with equal probability, e.g.

$$\mathrm{P}(X = x) = \frac{1}{N} \qquad \text{for } x = 1, 2, ..., N \tag{2.26a}$$

with

$$\mathrm{E}(X) = \frac{1}{2}(N+1), \quad \mathrm{Var}(X) = \frac{1}{12}(N^2 - 1). \tag{2.26b}$$

Zeta (or Zipf) distribution

In this case

$$\mathrm{P}(X = x) = \frac{C}{x^{\alpha+1}}, \quad x = 1, 2, ..., \alpha > 0 \tag{2.27a}$$

with

$$C = \left[\sum_{x=1}^{\infty} \left(\frac{1}{x}\right)^{\alpha+1}\right]^{-1} = \zeta(\alpha+1), \tag{2.27b}$$

where $\zeta(s)$ is the Riemann zeta function.