

**Discrete random variables****Some revision**

If  $X$  is a discrete random variable then  $X$  may take

finitely many values  $x_1 < x_2 < \cdots < x_n$

or

infinitely many values  $\{x_i : i \in \mathbb{Z}\}$  so long as no two are too close together.

Write  $p_i = \mathbb{P}(X = x_i)$  = probability that  $X = x_i$ . The list of the  $p_i$  is called the *probability mass function*, and

$$\sum_i p_i = 1.$$

The *expectation* of  $X$  is  $\mathbb{E}(X) = \sum_i p_i x_i$ .

If  $g$  is any real function, then  $g(X)$  is a random variable and

$$\mathbb{E}(g(X)) = \sum_i p_i g(x_i).$$

In particular,  $\mathbb{E}(X^2) = \sum_i p_i x_i^2$  and the *variance* of  $X$  is

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

**Bernoulli random variable**

Given  $p$  in  $(0, 1)$ , we say that  $X \sim \text{Bernoulli}(p)$  if  $\mathbb{P}(X = 0) = q$  and  $\mathbb{P}(X = 1) = p$ , where  $q = 1 - p$ .

**Binomial random variable** Given  $p$  in  $(0, 1)$  and a positive integer  $n$ , we say that  $X \sim \text{Bin}(n, p)$  if

$$\mathbb{P}(X = i) = \binom{n}{i} q^{n-i} p^i \quad \text{for } i \in \mathbb{Z} \text{ with } 0 \leq i \leq n,$$

where  $q = 1 - p$ .

**Geometric random variable** Given  $p$  in  $(0, 1)$ , we say that  $X \sim \text{Geom}(p)$  if

$$\mathbb{P}(X = i) = q^{i-1} p \quad \text{for all positive integers } i,$$

where  $q = 1 - p$ .

**Hypergeometric random variable** Suppose that we have  $N$  sheep in a field, of which  $M$  are black and the rest white. We sample  $n$  sheep from the field *without replacement*. Let the random variable  $X$  be the number of black sheep in the sample. Such an  $X$  is called a *hypergeometric* random variable  $\text{Hg}(n, M, N)$ .

Here  $n$ ,  $M$  and  $N$  are positive integers with  $n \leq N$  and  $M \leq N$ .

$$\mathbb{P}(X = i) = \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}} \quad \text{for } 0 \leq i \leq n.$$

If  $n \ll M$  and  $n \ll N - M$  then  $X$  is approximately  $\text{Bin}(n, \frac{M}{N})$ .

**Poisson random variable** Given a positive real number  $\lambda$ , we say that  $X \sim \text{Poisson}(\lambda)$  if

$$\mathbb{P}(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \quad \text{for all non-negative integers } i.$$

## Using tables

The *cumulative distribution function* (cdf)  $F$  of a random variable  $X$  is defined by

$$F(x) = \mathbb{P}(X \leq x) \quad \text{for } x \text{ in } \mathbb{R}.$$

We write  $F_X(x)$  if we need to emphasize  $X$ .

Suppose that all the values taken by  $X$  are

$$x_0 < x_1 < x_2 < \cdots.$$

Then

$$F(x_i) = \mathbb{P}(X \leq x_i) = p_0 + p_1 + \cdots + p_i,$$

so

$$\mathbb{P}(X = x_i) = p_i = F(x_i) - F(x_{i-1}),$$

and

$$\mathbb{P}(X \geq x_i) = 1 - \mathbb{P}(X \leq x_{i-1}) = 1 - F(x_{i-1}).$$

Moreover, if  $x_i < x_j$  then

$$\mathbb{P}(x_i \leq X \leq x_j) = \mathbb{P}(X \leq x_j) - \mathbb{P}(X \leq x_{i-1}) = F(x_j) - F(x_{i-1}).$$

The *New Cambridge Statistical Tables* [1] give the cumulative distribution function for the binomial distribution (Table 1) and the Poisson distribution (Table 2).

**Example** If  $X \sim \text{Bin}(18, 0.3)$  then

$$\begin{aligned} \mathbb{P}(4 \leq X \leq 8) &= \mathbb{P}(X \leq 8) - \mathbb{P}(X \leq 3) \\ &= 0.9404 - 0.1646, \quad \text{from Table 1 of NCST,} \\ &= 0.7758. \end{aligned}$$

### The probability generating function

**Definition** Let  $X$  be a random variable whose values are non-negative integers. The *probability generating function* of  $X$  is defined by

$$G(t) = \sum_i p_i t^i,$$

where  $p_i = \mathbb{P}(X = i)$ .

This is a power series, with dummy variable  $t$ . The sum is over all values  $i$  taken by  $X$ . We write  $G_X(t)$  if we need to emphasize  $X$ .

Note that  $G(1) = \sum_i p_i = 1$ .

Here is some notation that we need in the next theorem. First,

$$\frac{d^n G(t)}{dt^n}$$

means the result of differentiating  $G(t)$  with respect to  $t$ ,  $n$  times. Then

$$\left. \frac{d^n G(t)}{dt^n} \right|_{t=1}$$

means the result of substituting  $t = 1$  into that.

**Theorem 3** Let  $G(t)$  be the probability generating function of a random variable  $X$ . If  $n$  is a positive integer, then

$$\left. \frac{d^n G(t)}{dt^n} \right|_{t=1} = \mathbb{E}[X(X-1)\cdots(X-n+1)].$$

**Proof**  $G(t) = \sum p_i t^i$ , so

$$\frac{d^n G(t)}{dt^n} = \sum p_i i(i-1)\cdots(i-n+1)t^{i-n}.$$

Substituting  $t = 1$  gives

$$\left. \frac{d^n G(t)}{dt^n} \right|_{t=1} = \sum p_i i(i-1)\cdots(i-n+1) = \mathbb{E}[X(X-1)\cdots(X-n+1)]. \quad \blacksquare$$

**Corollary** (a)  $\mathbb{E}(X) = \left. \frac{dG(t)}{dt} \right|_{t=1}$ .

$$(b) \text{ Var}(X) = \left. \frac{d^2 G(t)}{dt^2} \right|_{t=1} + \mathbb{E}(X) - [\mathbb{E}(X)]^2.$$

Let  $\mu = \mathbb{E}(X)$ . For positive integers  $n$ , the quantities  $\mathbb{E}(X^n)$  are called the *moments* of  $X$ , while the quantities  $\mathbb{E}(X - \mu)^n$  are called the *central moments*. The quantities  $\mathbb{E}[X(X - 1) \cdots (X - n + 1)]$  in Theorem 3 are called *factorial moments* of  $X$ .

**Example** Let  $X \sim \text{Bin}(n, p)$ . Put  $q = 1 - p$ . Then

$$G(t) = \sum_{i=0}^n \binom{n}{i} q^{n-i} p^i t^i = (q + pt)^n$$

by the Binomial Theorem. Hence

$$\frac{dG(t)}{dt} = np(q + pt)^{n-1},$$

and so

$$\mathbb{E}(X) = \left. \frac{dG(t)}{dt} \right|_{t=1} = np(q + p)^{n-1} = np1^{n-1} = np.$$

Continuing, we have

$$\frac{d^2G(t)}{dt^2} = n(n-1)p^2(q + pt)^{n-2},$$

so

$$\left. \frac{d^2G(t)}{dt^2} \right|_{t=1} = n(n-1)p^2,$$

and therefore

$$\begin{aligned} \text{Var}(X) &= n(n-1)p^2 + np - (np)^2 \\ &= np[(n-1)p + 1 - np] \\ &= np(np - p + 1 - np) \\ &= npq. \end{aligned}$$

**Example** Let  $X \sim \text{Geom}(p)$  and  $q = 1 - p$ . Then

$$\begin{aligned} G(t) &= \sum_{i=1}^{\infty} q^{i-1} p t^i \\ &= pt \sum_{i=1}^{\infty} q^{i-1} t^{i-1} \\ &= pt \sum_{i=0}^{\infty} (qt)^i \\ &= \frac{pt}{1 - qt}. \end{aligned}$$

Then

$$\frac{dG(t)}{dt} = \frac{(1-qt)p - pt(-q)}{(1-qt)^2} = \frac{p}{(1-qt)^2},$$

so

$$\mathbb{E}(X) = \left. \frac{dG(t)}{dt} \right|_{t=1} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Furthermore

$$\frac{d^2G(t)}{dt^2} = \frac{-2p(-q)}{(1-qt)^3} = \frac{2pq}{(1-qt)^3},$$

and so

$$\left. \frac{d^2G(t)}{dt^2} \right|_{t=1} = \frac{2pq}{(1-q)^3} = \frac{2pq}{p^3} = \frac{2q}{p^2}.$$

Then

$$\text{Var}(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2}(2q + p - 1) = \frac{q}{p^2}.$$

**Example** Let  $X \sim \text{Poisson}(\lambda)$ . Then

$$\begin{aligned} G(t) &= \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} t^i \\ &= e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda t)^i}{i!} \\ &= e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}. \end{aligned}$$

Hence

$$\frac{dG(t)}{dt} = \lambda e^{\lambda(t-1)}$$

and

$$\frac{d^2G(t)}{dt^2} = \lambda^2 e^{\lambda(t-1)}.$$

Substituting  $t = 1$  gives  $\mathbb{E}(X) = \lambda$  and  $\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$ .

**Note:** If we know  $G_X(t)$  then we know all the coefficients  $p_i$  so we know the distribution of  $X$ . In particular, if  $G_X(t) = G_Y(t)$  then  $X$  and  $Y$  have the same distribution.

If  $X$  and  $Y$  are discrete random variables then  $X$  and  $Y$  are *independent* if

$$\mathbb{P}(X = i \text{ and } Y = j) = \mathbb{P}(X = i) \mathbb{P}(Y = j)$$

for all values  $i$  of  $X$  and  $j$  of  $Y$ .

**Theorem 4** Let  $X$  and  $Y$  be two random variables whose values are non-negative integers. Let  $G_X(t)$ ,  $G_Y(t)$  and  $G_{X+Y}(t)$  be the probability generating functions of  $X$ ,  $Y$  and  $X + Y$  respectively. If  $X$  and  $Y$  are independent of each other then  $G_{X+Y}(t) = G_X(t)G_Y(t)$ .

**Proof** If  $X + Y = i$  then there is some integer  $j$  with  $0 \leq j \leq i$  such that  $X = i - j$  and  $Y = j$ . If  $X$  and  $Y$  are independent then

$$\begin{aligned} \mathbb{P}(X + Y = i) &= \sum_{j=0}^i \mathbb{P}(X = i - j \text{ and } Y = j) \\ &= \sum_{j=0}^i \mathbb{P}(X = i - j) \mathbb{P}(Y = j). \end{aligned}$$

Hence

$$G_{X+Y}(t) = \sum_{i=0}^{\infty} \left[ \sum_{j=0}^i \mathbb{P}(X = i - j) \mathbb{P}(Y = j) \right] t^i.$$

On the other hand,

$$\begin{aligned} G_X(t)G_Y(t) &= \left( \sum_{k=0}^{\infty} \mathbb{P}(X = k) t^k \right) \left( \sum_{j=0}^{\infty} \mathbb{P}(Y = j) t^j \right) \\ &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \mathbb{P}(X = k) \mathbb{P}(Y = j) t^{k+j}. \end{aligned}$$

To get the coefficient of  $t^i$  in  $G_X(t)G_Y(t)$  we need all pairs  $(k, j)$  with  $k + j = i$ , so we need to take  $k = i - j$  and the coefficient is

$$\sum_{j=0}^i \mathbb{P}(X = i - j) \mathbb{P}(Y = j),$$

which is exactly the same as the coefficient of  $t^i$  in  $G_{X+Y}(t)$ .

This is true for all  $i$ , and so  $G_{X+Y}(t) = G_X(t)G_Y(t)$ . ■

**Theorem 5** If  $X$  and  $Y$  are independent random variables and  $X \sim \text{Bin}(n, p)$  and  $Y \sim \text{Bin}(m, p)$  then  $X + Y \sim \text{Bin}(n + m, p)$ .

**Proof** The probability generating functions of  $X$  and  $Y$  are  $G_X(t) = (q + pt)^n$  and  $G_Y(t) = (q + pt)^m$ , where  $q = 1 - p$ .

By Theorem 4,

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (q + pt)^n(q + pt)^m = (q + pt)^{n+m}.$$

Hence

$$\begin{aligned} \mathbb{P}(X + Y = i) &= \text{coefficient of } t^i \text{ in } (q + pt)^{n+m} \\ &= \binom{n+m}{i} q^{n+m-i} p^i \quad \text{for } 0 \leq i \leq n+m, \end{aligned}$$

and so  $X + Y \sim \text{Bin}(n + m, p)$ . ■

**Theorem 6** If  $X$  and  $Y$  are independent random variables and  $X \sim \text{Poisson}(\lambda)$  and  $Y \sim \text{Poisson}(\mu)$  then  $X + Y \sim \text{Poisson}(\lambda + \mu)$ .

**Proof** The probability generating functions of  $X$  and  $Y$  are  $G_X(t) = e^{\lambda(t-1)}$  and  $G_Y(t) = e^{\mu(t-1)}$ .

By Theorem 4,

$$G_{X+Y}(t) = G_X(t)G_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}.$$

Therefore

$$\begin{aligned} \mathbb{P}(X + Y = i) &= \text{coefficient of } t^i \text{ in } e^{(\lambda+\mu)(t-1)} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^i}{i!} \quad \text{for non-negative integers } i, \end{aligned}$$

and so  $X + Y \sim \text{Poisson}(\lambda + \mu)$ . ■

---

[1] D. V. Lindley and W. F. Scott, *New Cambridge Statistical Tables*, Second edition, Cambridge University Press.