6. Discrete probability distributions. Sums of powers of integers

An important application of sequences is to probability distributions. A probability distribution consists of a set of possible outcomes for an experiment together with a set of associated probabilities. The set of possible outcomes is called the **sample space**. We can always arrange for it to be a set of numbers; for example, if the experiment is the birth of a child and possible outcomes are boy (B) or girl (G), then we can replace sample space {B, G} by {1, 2}, where 1 means boy and 2 means girl. Then an experiment consists of drawing a number at random from the sample space. We call the unknown number a **random variable**, and denote it by X.

A probability distribution whose sample space is a set of integers is said to be **discrete**. We can always arrange for the set of integers to be [0...∞); e.g., in the case of childbirth, we would attach zero probability to $X = 0$ or $X \geq 3$. Then either of two sequences completely specifies the distribution. The first sequence, the **probability density function** or **p.d.f.**, is the nonnegative sequence $\{p_n\}$ defined by

$$p_n \quad = \quad \mathrm{Prob}(X = n), \tag{6.1}$$

where $\mathrm{Prob}(X = n)$ denotes the probability that n is drawn at random from the sample space. For example, in the case of childbirth, if γ (≈ 0.49) is the probability of a girl, then $p_2 = \mathrm{Prob}(X = 2) = \gamma$ and $p_1 = \mathrm{Prob}(X = 1) = 1 - \gamma$. The second sequence, the **cumulative distribution function** or **c.d.f.**, is the nonnegative sequence $\{P_n\}$ defined by

$$P_n \quad = \quad \mathrm{Prob}(X \leq n). \tag{6.2}$$

For example, in the case of childbirth, $P_1 = \mathrm{Prob}(X \leq 1) = 1 - \gamma$ and $P_2 = \mathrm{Prob}(X \leq 2) = 1$. Note that, because $X \geq 0$, we must have

$$P_0 \quad = \quad \mathrm{Prob}(X \leq 0) \quad = \quad \mathrm{Prob}(X = 0) \quad = \quad p_0. \tag{6.3}$$

In fact, for the sake of simplicity, we will assume that X is strictly positive. Then (3) implies

$$P_0 \quad = \quad 0 \quad = \quad p_0. \tag{6.4}$$

Any number in the sample space is independent of any other number (if you draw 1, then you cannot at the same time draw 2, and vice versa). Thus the probability of k or m must always equal that of k plus that of m, and (4) implies

$$\mathrm{Prob}(X \leq n) \quad = \quad \mathrm{Prob}(X = 1 \quad \text{OR} \quad X = 2 \ \text{OR} \cdots \text{OR} \quad X = n-1 \quad \text{OR} \quad X = n)$$
$$\mathrm{Prob}(X \leq n) \quad = \quad \mathrm{Prob}(X = 1) + \mathrm{Prob}(X = 2) + \cdots + \mathrm{Prob}(X = n-1) + \mathrm{Prob}(X = n)$$

for any $n \geq 1$; or, using (1)-(2) and summation notation,

$$P_n \quad = \quad p_1 + p_2 + \cdots + p_{n-1} + p_n \quad = \quad \sum_{k=1}^{n} p_k \tag{6.5}$$

for any $n \geq 1$. Because total probability (i.e., the probability of something happening) is always 1, the sequence $\{P_n\}$ must converge to 1, i.e., we must have

$$P_\infty \quad = \quad \lim_{n\to\infty} P_n \quad = \quad \mathrm{Prob}(X < \infty) \quad = \quad 1, \tag{6.6a}$$

which is usually written as

$$\sum_{k=1}^{\infty} p_k \quad = \quad 1. \tag{6.6b}$$

The easiest way to satisfy (6) is to have $p_k = 0$ and $P_k = 1$ if k is sufficiently large, say, if $k > M$. For example, we can describe the distribution of 489 leaf thicknesses in *D. linearifolia* by setting $M = 15$, X = THICKNESS OF RANDOMLY CHOSEN LEAF and

$$p_k \quad = \quad \mathrm{Prob}(X = k) \quad = \quad \frac{\text{FREQUENCY OF THICKNESS } k/60\text{ mm}}{489}. \tag{6.7}$$

Then

$$P_k \quad = \quad \mathrm{Prob}(X \le k) \quad = \quad \frac{\text{FREQUENCY OF THICKNESS } \le k/60\text{ mm}}{489}. \tag{6.8}$$

See Table 1 and Figure 1. (It is possible, however, to have $p_k > 0$ for all positive k; see Exercise 6.)

k	p_k	P_k	k	p_k	P_k
0		0	8	55/163	84/163
1	0	0	9	30/163	114/163
2	0	0	10	118/489	460/489
3	0	0	11	17/489	159/163
4	3/163	3/163	12	3/163	162/163
5	5/489	14/489	13	0	162/163
6	28/489	14/163	14	1/489	487/489
7	15/163	29/163	15	2/489	1

Table 6.1 Leaf thickness distribution for *Dicerandra linearifolia*

Now, if (5) holds for all $n \ge 1$, then

$$P_{n-1} \quad = \quad \sum_{k=1}^{n-1} p_k \tag{6.9}$$

must hold for all $n - 1 \ge 1$ or $n \ge 2$. The only difference between the right-hand sides of (5) and (9), however, is that (5) contains the term p_n, whereas (9) does not. Hence

$$P_n \quad - \quad P_{n-1} \quad = \quad \sum_{k=1}^{n} p_k \quad - \quad \sum_{k=1}^{n-1} p_k \quad = \quad p_n \tag{6.10}$$

for any $n \ge 2$. But (10) also holds for $n = 1$, because $P_1 - P_0 = P_1 - 0 = p_1$, by (1)-(4). So (10) holds for $n \ge 1$. The upshot is that we can always obtain the p.d.f. from the c.d.f. by using

$$p_n \quad = \quad P_n \quad - \quad P_{n-1}, \quad n \ge 1 \tag{6.11a}$$

and we can always obtain the c.d.f. from the p.d.f by using

$$\begin{aligned} P_0 \quad &= \quad 0 \\ P_n \quad &= \quad \sum_{k=1}^{n} p_k, \quad n \ge 1. \end{aligned} \tag{6.11b}$$

For example, in Table 1 we have $P_9 = 114/163$ and $P_{10} = 460/489$, implying $p_{10} = P_{10} - P_9 = 118/489$, by (31a); and we have $p_k = 0$ for $k \le 3$, $p_4 = 3/163$ and $p_5 = 5/489$, so that $P_5 = p_1 + p_2 + p_3 + p_4 + p_5 = 14/489$, by (31b). Similarly, if X is a clutch size drawn at random from Hussell's (1972) Lapland Longspur data, then Table 5.2 and (11) imply

$$\begin{aligned} &p_1 = 0 \quad p_2 = \tfrac{1}{54} \quad p_3 = \tfrac{2}{27} \quad p_4 = \tfrac{2}{9} \\ &p_5 = \tfrac{8}{27} \quad p_6 = \tfrac{17}{54} \quad p_7 = \tfrac{2}{27} \quad p_n = 0, n \ge 8 \end{aligned} \tag{6.12a}$$

and

$$\begin{aligned} &P_0 = 0 \quad P_1 = 0 \quad P_2 = \tfrac{1}{54} \quad P_3 = \tfrac{5}{54} \\ &P_4 = \tfrac{17}{54} \quad P_5 = \tfrac{11}{18} \quad P_6 = \tfrac{25}{27} \quad P_n = 1, n \ge 7. \end{aligned} \tag{6.12b}$$

More generally, any sequence $\{p_n\}$ on $[1 \ldots \infty)$ is potentially the p.d.f. of some distribution if it satisfies only two conditions, namely,

$$\begin{aligned} p_n &\geq 0, \quad 1 \leq n < \infty \\ \sum_{n=1}^{\infty} p_n &= 1. \end{aligned} \tag{6.13a}$$

Correspondingly, any sequence $\{P_n\}$ on $[0 \ldots \infty)$ is potentially the c.d.f. of a distribution if it satisfies only three conditions, namely,

$$\begin{aligned} P_0 &= 0 \\ P_n &\geq P_{n-1}, \quad 1 \leq n < \infty \\ P_\infty &= \lim_{n\to\infty} P_n = 1 \end{aligned} \tag{6.13b}$$

These two sets of conditions are equivalent, by (11).

We can exploit this equivalence to obtain expressions for sums of powers of positive integers, which we need in Lecture 10. We will obtain an expression for the sum of squares, leaving analogous results for cubes and other powers to the exercises; see Exercises 2-4. Accordingly, consider the sequence defined on $[0 \ldots \infty)$ by

$$P_n = \begin{cases} \dfrac{n(n+1)(2n+1)}{M(M+1)(2M+1)} & \text{if } 0 \leq n \leq M \\ 1 & \text{if } M+1 \leq n < \infty. \end{cases} \tag{6.14}$$

You can see by inspection that $P_0 = 0$, $P_\infty = 1$ and $P_n \geq P_{n-1}$ (in fact $P_n > P_{n-1}$ for $1 \leq n \leq M$). Thus $\{P_n\}$ defines a probability distribution, implying in particular that (11a) and (13a) must hold. Note that $P_M = 1$. Thus $P_n = 1$ for $n \geq M$, implying $P_{n-1} = 1$ for $n \geq M + 1$, so that (11a) implies $p_n = P_n - P_{n-1} = 1 - 1 = 0$ for $n \geq M + 1$. Hence (13a) reduces to

$$\sum_{n=1}^{M} p_n = 1, \tag{6.15}$$

and it follows immediately from (11a) that

$$\sum_{n=1}^{M} \{P_n - P_{n-1}\} = 1. \tag{6.16}$$

For $n \leq M$, however, (14) implies

$$\begin{aligned} P_n - P_{n-1} &= \frac{n(n+1)(2n+1)}{M(M+1)(2M+1)} - \frac{(n-1)(n-1+1)(2\{n-1\}+1)}{M(M+1)(2M+1)} \\ &= \frac{n(n+1)(2n+1)}{M(M+1)(2M+1)} - \frac{(n-1)n(2n-1)}{M(M+1)(2M+1)} \\ &= \frac{n}{M(M+1)(2M+1)}\left\{(n+1)(2n+1) - (n-1)(2n-1)\right\} \\ &= \frac{n}{M(M+1)(2M+1)}\left\{2n^2+3n+1 - (2n^2-3n+1)\right\} \\ &= \frac{n \cdot 6n}{M(M+1)(2M+1)}. \end{aligned} \tag{6.17}$$

Substituting into (15), we find that

$$\sum_{n=1}^{M} \frac{6n^2}{M(M+1)(2M+1)} = 1, \tag{6.18}$$

implying

$$\frac{6}{M(M+1)(2M+1)} \sum_{n=1}^{M} n^2 = 1 \tag{6.19}$$

because anything that does not depend on n can be brought outside the summation sign. So the sum of the squares of the first M positive integers is

$$\sum_{n=1}^{M} n^2 = \frac{1}{6}M(M+1)(2M+1). \tag{6.20}$$

For example, $1^2+2^2+3^2 = 3(3+1)(2\cdot 3+1)/6 = 3\cdot 4\cdot 7/6 = 14$, $1^2+2^2+3^2+4^2 = 4(4+1)(2\cdot 4+1)/6 = 4\cdot 5\cdot 9/6 = 30$, and so on. We will need (20) and similar results in Lectures 10-11.

References

Gross, Alan J. & Virginia A. Clark (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. Wiley, New York.

Hussell, D.J.T. (1972) Factors affecting clutch size in Arctic passerines. Ecological Monographs **42**, 317-364.

Jean, Roger V. (1994). Phyllotaxis: A Systemic Study in Plant Morphogenesis. Cambridge University Press, U.K.

MacDonald, E.J. (1963). The epidemiology of melanoma. *Annals of the New York Academy of Sciences* **100**, 4-15.

Winn, Alice A. (1996). The contributions of programmed developmental change and phenotypic plasticity to within-individual variation in leaf traits in *Diceranda linearifolia.* Unpublished manuscript.

Exercises 6

6.1 Table 5.2 shows clutch sizes observed among four species of arctic passerine. For each species, produce the analogues of Table 1 and Figure 1.

6.2 Use the c.d.f. defined by

$$P_n = \begin{cases} \dfrac{n(n+1)}{M(M+1)} & \text{if } 0 \le n \le M \\ 1 & \text{if } M+1 \le n < \infty \end{cases}$$

and the method of this lecture to establish that

$$\sum_{n=1}^{M} n = \frac{1}{2}M(M+1).$$

6.3 Use the c.d.f. defined by

$$P_n = \begin{cases} \dfrac{n^2(n+1)^2}{M^2(M+1)^2} & \text{if } 0 \le n \le M \\ 1 & \text{if } M+1 \le n < \infty \end{cases}$$

and the method of this lecture to establish that

$$\sum_{n=1}^{M} n^3 = \frac{1}{4}M^2(M+1)^2.$$

6.4 Use the c.d.f. defined by

$$P_n = \begin{cases} \dfrac{n(n+1)(2n+1)(3n^2+3n-1)}{M(M+1)(2M+1)(3M^2+3M-1)} & \text{if } 0 \le n \le M \\ 1 & \text{if } M+1 \le n < \infty \end{cases}$$

and the method of this lecture to establish that

$$\sum_{n=1}^{M} n^4 = \frac{1}{30}M(M+1)(2M+1)(3M^2+3M-1).$$

6.5 A discrete probability density function is defined by

$$p_n = \begin{cases} bn & \text{if } n = 1, 2, \ldots, M \\ 0 & \text{if } n \ge M+1 \end{cases}$$

where b is a constant. What must be the value of b?

6.6 A discrete probability density function is defined by

$$p_n = \frac{6}{\pi^2 n^2}, \quad n \ge 1.$$

(i) Sketch the graph of the c.d.f. $\{P_n\}$ on subdomain [0 ... 10].

(ii) What must be the sum of the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \ldots \text{ ?}$$

Answers and Hints for Selected Exercises

6.3 For $n = M$ we have $P_n = P_M = \frac{M^2(M+1)^2}{M^2(M+1)^2} = 1$. For $n \geq M + 1$ we have $P_n = 1$. So for $n \geq M$ we have $P_n = 1$, implying $P_{n-1} = 1$ for $n \geq M + 1$. So for $n \geq M + 1$ we have $P_n - P_{n-1} = 1 - 1 = 0$. That is, $p_n = 0$ for $n > M$, by (3.31a). So, by (3.35),

$$\begin{aligned}
1 \quad &= \quad \sum_{n=1}^{M}\{P_n - P_{n-1}\} \quad = \quad \sum_{n=1}^{M}\left\{\frac{n^2(n+1)^2}{M^2(M+1)^2} - \frac{(n-1)^2(\{n-1\}+1)^2}{M^2(M+1)^2}\right\} \\
&= \quad \sum_{n=1}^{M}\left\{\frac{n^2(n+1)^2}{M^2(M+1)^2} - \frac{(n-1)^2 n^2}{M^2(M+1)^2}\right\} \\
&= \quad \frac{1}{M^2(M+1)^2}\sum_{n=1}^{M} n^2\left\{(n+1)^2 - (n-1)^2\right\} \\
&= \quad \frac{1}{M^2(M+1)^2}\sum_{n=1}^{M} n^2\left\{n^2 + 2n + 1 - (n^2 - 2n + 1)\right\} \\
&= \quad \frac{1}{M^2(M+1)^2}\sum_{n=1}^{M} 4n^3 \quad = \quad \frac{4}{M^2(M+1)^2}\sum_{n=1}^{M} n^3,
\end{aligned}$$

which implies the result.

6.5 From Exercise 2, $b \quad = \quad \frac{2}{M(M+1)}$.

6.6 (ii) $\frac{\pi^2}{6}$