abm 4th edition

3.7 The Poisson distribution

This distribution is named after S.D. Poisson (1781–1840), a French mathematician. It is sometimes useful as a limiting form of the binomial, but it is important also in its own right as a distribution arising when events of some sort occur randomly in time, or when small particles are distributed randomly in space.

We shall first consider random events in time. Suppose that a certain type of event occurs repeatedly, with an average rate of λ per unit time but in an entirely random fashion. To make the idea of randomness rather more precise we can postulate that in any very small interval of time of length h (say 1 ms) the probability that an event occurs is approximately proportional to h, say λh. (For example, if h is doubled the very small probability that the interval contains an event is also doubled.) The probability that the interval contains more than one event is supposed to be proportionately smaller and smaller as h gets smaller, and can therefore be ignored. Furthermore, we suppose that what happens in any small interval is independent of what happens in any other small interval which does not overlap the first.

A very good instance of this probability model is that of the emission of radioactive particles from some radioactive material. The rate of emission, λ, will be constant, but the particles will be emitted in a purely random way, each successive small interval of time being on exactly the same footing, rather than in a regular pattern. The model is the analogy, in continuous time, of the random sequence of independent trials discussed in §3.1, and is called the *Poisson process*.

Suppose that we observe repeated stretches of time, of length T time units, from a Poisson process with a rate λ. The number, x, of events occurring in an interval of length T will vary from one interval to another. In fact, it is a random variable, the possible values of which are 0, 1, 2, ... What is the probability of a particular value x?

A natural guess at the value of x would be λT, the rate of occurrence multiplied by the time interval. We shall see later that λT is the mean of the distribution of x, and it will be convenient to denote λT by the single symbol μ.

Let us split any one interval of length T into a large number n of subintervals each of length T/n (Fig. 3.8). Then, if n is sufficiently large, the number of events in the subinterval will almost always be 0, will occasionally be 1 and will hardly ever be more than 1. The situation is therefore almost exactly the same as a sequence of n binomial trials (a trial being the observation of a subinterval), in each of which there is a probability $\lambda(T/n) = \mu/n$ of there being an event and $1 - \mu/n$ of there being no event. The probability that the whole series of n trials provides exactly x events is, in this approximation, given by the binomial distribution:

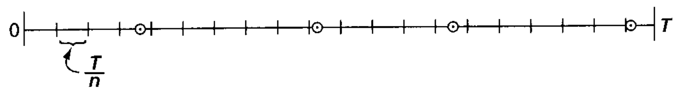

Fig. 3.8 The occurrence of events in a Poisson process, with the time-scale subdivided into small intervals.

$$\frac{n(n-1)\dots(n-x+1)}{x!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}. \tag{3.17}$$

Now, this binomial approximation will get better and better as n increases. What happens to (3.17) as n increases indefinitely? We can replace

$$n(n-1)\dots(n-x+1)$$

by n^x since x will be negligible in comparison with n. Similarly we can replace $(1-\mu/n)^{n-x}$ by $(1-\mu/n)^n$ since $(1-\mu/n)^x$ will approach 1 as n increases. It is a standard mathematical result that, as n increases indefinitely, $(1-\mu/n)^n$ approaches $e^{-\mu}$, where e is the base of natural (or Napierian) logarithms ($e = 2{\cdot}718\dots$).

Finally, then, in the limit as n increases indefinitely, the probability of x events approaches

$$P_x = \frac{n^x}{x!}\left(\frac{\mu}{n}\right)^x e^{-\mu} = \frac{\mu^x e^{-\mu}}{x!}. \tag{3.18}$$

The expression (3.18) defines the Poisson probability distribution. The random variable x takes the values 0, 1, 2,... with the successive probabilities obtained by putting these values of x in (3.18). Thus,

$$\begin{aligned} P_0 &= e^{-\mu} \\ P_1 &= \mu e^{-\mu} \\ P_2 &= \tfrac{1}{2}\mu^2 e^{-\mu}, \text{etc.} \end{aligned}$$

Note that, for $x = 0$, we replace $x!$ in (3.18) by the value 1, as was found to be appropriate for the binomial distribution. To verify that the sum of the probabilities is 1,

$$\begin{aligned} P_0 + P_1 + P_2 + \dots &= e^{-\mu}\left(1 + \mu + \tfrac{1}{2}\mu^2 + \dots\right) \\ &= e^{-\mu} \times e^{\mu} \\ &= 1, \end{aligned}$$

the replacement of the infinite series on the right-hand side by e^{μ} being a standard mathematical result.

Before proceeding to further consideration of the properties of the Poisson distribution we may note that a similar derivation may be applied to the situation in which particles are randomly distributed in space. If the space is one-dimensional (for instance the length of a cotton thread along which flaws may occur with constant probability at all points), the analogy is immediate. With two-dimensional space (for instance a microscopic slide over which bacteria are distributed at random with perfect mixing technique) the total area of size A may be divided into a large number n of subdivisions each of area A/n; the argument then carries through with A replacing T. Similarly, with three-dimensional space (bacteria well mixed in a fluid suspension), the total volume V is divided into n small volumes of size V/n. In all these situations the model envisages particles distributed at random with density λ per unit length (area or volume). The number of particles found in a length (area or volume) of size l (A or V) will follow the Poisson distribution (3.18) where the parameter $\mu = \lambda l$ (λA or λV).

The shapes of the distribution for $\mu = 1$, 4 and 15 are shown in Fig. 3.9. Note that for $\mu = 1$ the distribution is very skew, for $\mu = 4$ the skewness is much less and for $\mu = 15$ it is almost absent.

The distribution (3.18) is determined entirely by the one parameter, μ. It follows that all the features of the distribution in which one might be interested are functions only of μ. In particular the mean and variance must be functions of μ. The mean is

$$\begin{aligned} E(x) &= \sum_{x=0}^{\infty} x P_x \\ &= \mu, \end{aligned}$$

this result following after a little algebraic manipulation.

By similar manipulation we find

$$E(x^2) = \mu^2 + \mu$$

and

$$\begin{aligned} \operatorname{var}(x) &= E(x^2) - \mu^2 \\ &= \mu \end{aligned} \tag{3.19}$$

Thus, the variance of x, like the mean, is equal to μ. The standard deviation is therefore $\sqrt{\mu}$.

Much use is made of the Poisson distribution in bacteriology. To estimate the density of live organisms in a suspension the bacteriologist may dilute the suspension by a factor of, say, 10^{-5}, take samples of, say, 1 cm^3 in a pipette and drop the contents of the pipette on to a plate containing a nutrient medium on which the bacteria grow. After some time each organism dropped on to the plate will have formed a colony and these colonies can be counted. If the original suspension was well mixed, the volumes sampled are accurately determined and

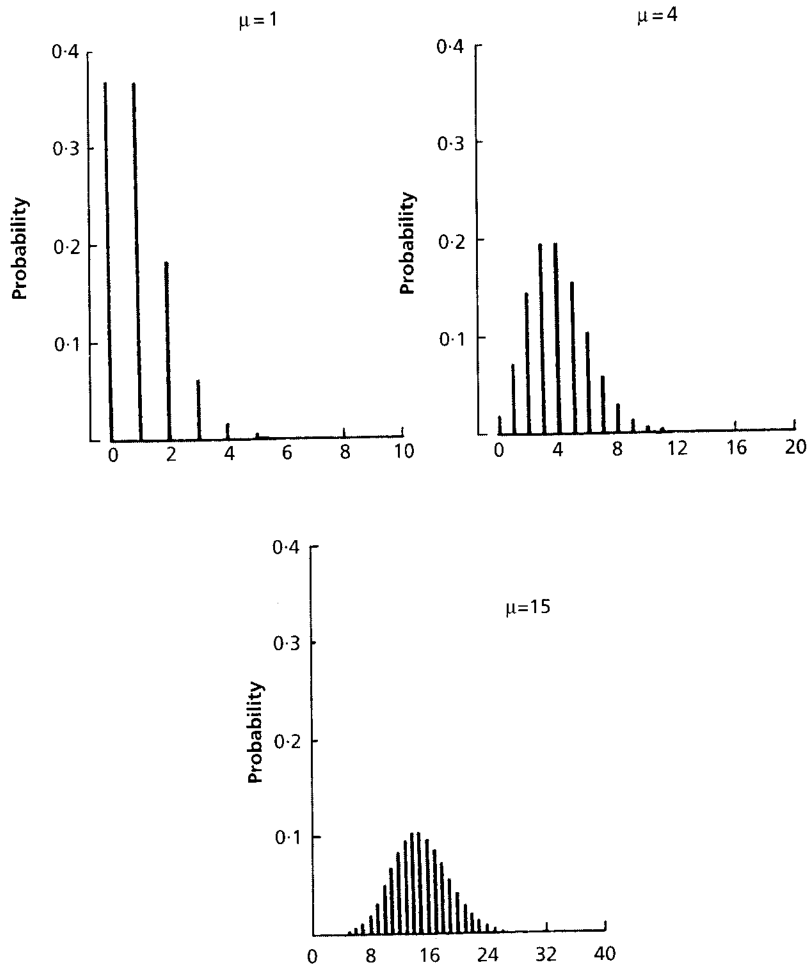

Fig. 3.9 Poisson distribution for various values of μ. The horizontal scale in each diagram shows values of x.

the medium is uniformly adequate to sustain growth, the number of colonies in a large series of plates could be expected to follow a Poisson distribution. The mean colony count per plate, $\bar{x}$, is an estimate of the mean number of bacteria per $10^{-5}\,\mathrm{cm}^3$ of the original suspension, and a knowledge of the theoretical properties of the Poisson distribution permits one to measure the precision of this estimate (see §5.2).

Similarly, for total counts of live and dead organisms, repeated samples of constant volume may be examined under the microscope and the organisms

Example 3.7

As an example, Table 3.3 shows a distribution observed during a count of the root nodule bacterium (*Rhizobium trifolii*) in a Petroff–Hausser counting chamber. The 'expected' frequencies are obtained by calculating the mean number of organisms per square, $\bar{x}$, from the frequency distribution (giving $\bar{x} = 2{\cdot}50$) and calculating the probabilities P_x of the Poisson distribution with μ replaced by $\bar{x}$. The expected frequencies are then given by $400\,P_x$. The observed and expected frequencies agree quite well. This organism normally produces gum and therefore clumps readily. Under these circumstances one would not expect a Poisson distribution, but the data in Table 3.3 were collected to show the effectiveness of a method of overcoming the clumping.

In the derivation of the Poisson distribution use was made of the fact that the binomial distribution with a large n and small π is an approximation to the Poisson with mean $\mu = n\pi$.

Conversely, when the correct distribution is a binomial with large n and small π, one can approximate this by a Poisson with mean $n\pi$. For example, the number of deaths from a certain disease, in a large population of n individuals subject to a probability of death π, is really binomially distributed but may be taken as approximately a Poisson variable with mean $\mu = n\pi$. Note that the standard deviation on the binomial assumption is $\sqrt{[n\pi(1-\pi)]}$, whereas the Poisson standard deviation is $\sqrt{(n\pi)}$. When π is very small these two expressions are almost equal. Table 3.4 shows the probabilities for the Poisson distribution with $\mu = 5$, and those for various binomial distributions with $n\pi = 5$. The similarity between the binomial and the Poisson improves with increases in n (and corresponding decreases in π).

Table 3.3 Distribution of counts of root nodule bacterium (*Rhizobium trifolii*) in a Petroff–Hausser counting chamber (data from Wilson and Kullman, 1931).

Number of bacteria per square	Number of squares	
	Observed	Expected
0	34	32·8
1	68	82·1
2	112	102·6
3	94	85·5
4	55	53·4
5	21	26·7
6	12	11·1
7–	4	5·7
	400	399·9

Table 3.4 Binomial and Poisson distributions with $\mu = 5$.

	π	0·5	0·10	0·05	
r	n	10	50	100	Poisson
0		0·0010	0·0052	0·0059	0·0067
1		0·0098	0·0286	0·0312	0·0337
2		0·0439	0·0779	0·0812	0·0842
3		0·1172	0·1386	0·1396	0·1404
4		0·2051	0·1809	0·1781	0·1755
5		0·2461	0·1849	0·1800	0·1755
6		0·2051	0·1541	0·1500	0·1462
7		0·1172	0·1076	0·1060	0·1044
8		0·0439	0·0643	0·0649	0·0653
9		0·0098	0·0333	0·0349	0·0363
10		0·0010	0·0152	0·0167	0·0181
>10		0	0·0094	0·0115	0·0137
		1·0000	1·0000	1·0000	1·0000

Probabilities for the Poisson distribution may be obtained from many statistical packages.

3.8 The normal (or Gaussian) distribution

The binomial and Poisson distributions both relate to a discrete random variable. The most important continuous probability distribution is the *Gaussian* (C.F. Gauss, 1777–1855, German mathematician) or, as it is frequently called, the *normal* distribution. Figures 3.10 and 3.11 show two frequency distributions, of height and of blood pressure, which are similar in shape. They are both approximately symmetrical about the middle and exhibit a shape rather like a bell, with a pronounced peak in the middle and a gradual falling off of the frequency in the two tails. The observed frequencies have been approximated by a smooth curve, which is in each case the probability density of a normal distribution.

Frequency distributions resembling the normal probability distribution in shape are often observed, but this form should not be taken as the norm, as the name 'normal' might lead one to suppose. Many observed distributions are undeniably far from 'normal' in shape and yet cannot be said to be abnormal in the ordinary sense of the word. The importance of the normal distribution lies not so much in any claim to represent a wide range of observed frequency distributions but in the central place it occupies in sampling theory, as we shall see in Chapters 4 and 5. For the purposes of the present discussion we shall regard the normal distribution as one of a number of theoretical forms for a ... variable, and proceed to describe some of its properties