The Poisson Distribution

Prerequisites

You should be familiar with exponential functions and should have studied the binomial distribution.

Example (1)
An aircraft has 152 seats. The airline has found, from long experience, that on average 2.5% of people with tickets for a particular flight do not arrive for that flight. If the airline sells 156 seats for a particular flight determine, using the binomial distribution, the probability that the flight is overbooked.

Solution
Let X denote the number of people who do not arrive for the flight.

$X \sim B(156, 0.025)$

There are 4 more people booked for the flight than there are seats. The possibility that the flight is overbooked corresponds to the probability

$P(X < 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$.

These 4 values correspond to the 4 overbooked seats. (If 4 or more people do not arrive for the flight then the flight is not overbooked.) Using the binomial distribution

$$\begin{aligned} P(X < 4) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{156}{0}(0.025)^0(0.975)^{156} + \binom{156}{1}(0.025)^1(0.975)^{155} \\ &\quad + \binom{156}{2}(0.025)^2(0.975)^{154} + \binom{156}{3}(0.025)^3(0.975)^{153} \\ &= 0.019263 + 0.077051 + 0.153115 + 0.201536 \\ &= 0.4510 \text{ (4 d.p.)} \end{aligned}$$

We have solved this problem using the binomial distribution. However, there are certain features of this solution that make the process of solving the problem this way inconvenient and motivate us to find an alternative approach.

(1) Tables of cumulative probabilities for the binomial distribution do not usually list a entries for probabilities as small as $p = 0.025$, nor for samples sizes such as $n = 156$.

For this reason we cannot use tables and have to use the formula $P(X=r)=\binom{n}{r}p^r q^{n-r}$ to evaluate the probabilities. This is tedious.

(2) At the same time, as the example does show, the situation is not uncommon. Airlines wish to know the risks involved when they deliberately overbook their flights. As we shall see below there are other practical situations similar to this one.

The Poisson distribution

The binomial distribution arises from a situation in which there random (independent) trials, in which there are two possible outcomes, one a "success" and the other a "failure". The Poisson distribution also distinguishes between a "success" and a "failure" and the random variable is the number of successes in a given number of trials (or sample). It is used rather than the binomial or some other distribution, when

(1) Events occur randomly in space and time.

(2) Events do not cluster. That is, they do not come in groups or bursts.

(3) "Successes" occur infrequently as a proportion of the total number of events. Hence, the mean frequency of "successes" is very small.

(It is usual to list these three conditions, but conditions (1) and (2) are also conditions of the use of the binomial distribution and to say that events occur randomly is to say that they do not cluster, unless, perhaps, by chance.)

Example (2)

Discuss the conditions under which each of the following might be modelled by the Poisson distribution.

(*a*) Car accidents on a section of motorway.

(*b*) Telephone calls made to a switchboard in a given period of time.

(*c*) Particles emitted by a radioactive source per unit of time.

Solution

(*a*) Car accidents on a section of motorway. It is assumed that the number of accidents per unit of time is a very small proportion of the total units of time for any given time period. From a purely technical point-of-view we regard a "success" as an accident, and a "failure" as a non-accident. However, the Poisson distribution will only be appropriate if the accidents are genuinely only occurring randomly. Thus, for example, a motorway pile up, when "successes" cluster is not appropriately modelled by the Poisson distribution.

(*b*) Telephone calls made to a switchboard in a given period of time. The events may be measured per minute. A “success” is a call and a “failure” is no call. The events are assumed to be infrequent and random. However, the telephone calls to the operators of a World Cup event upon announcing 160,000 extra tickets tend to cluster and the Poisson distribution would not be appropriate.

(*c*) Particles emitted by a radioactive source per unit of time. A Poisson distribution could be appropriate, but not if the radiation comes in bursts – for example, when observing a super nova.

Although these examples demonstrate cases where the Poisson distribution fails, they also show that the Poisson distribution has many applications. As we shall see, it is used because the computations are in fact easier than those for the binomial distribution.

Definition of the Poisson distribution

Let X be a discrete random variable. Then X follows the Poisson distribution if

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \text{ for } x = 0,\ 1,\ 2,\ 3,\ \ldots.$$

where $\lambda \geq 0$ is a parameter. When X follows a Poisson distribution we write

$$X \sim Po(\lambda)$$

This is read, “X follows a Poisson distribution with parameter λ”.

Example (3)

(*a*) Given $X \sim Po(3.9)$ find $P(X < 4)$.

(*b*) The distribution $X \sim Po(3.9)$ is the Poisson approximation to the binomial distribution given in example (1). Compare your answer to part (*a*) with the solution to example (1).

Solution

(*a*)
$$P(X = 0) = e^{-3.9}\frac{3.9^0}{0!} = 0.020242$$
$$P(X = 1) = e^{-3.9}\frac{3.9^1}{1!} = 0.078943$$
$$P(X = 2) = e^{-3.9}\frac{3.9^2}{2!} = 0.153940$$
$$P(X = 3) = e^{-3.9}\frac{3.9^3}{3!} = 0.200127$$

$$P(X<4)=P(X=0)+P(X=1)+P(X=2)+P(X=3)$$
$$=0.020242+0.078943+0.153940+0.200122=0.4532 \ (4 \text{ d.p.})$$

(*b*)

x	0	1	2	3	$P(X<4)$
$X \sim B(156, 0.025)$, $P(X=x)$	0.0193	0.0771	0.1531	0.2015	0.4510
$X \sim Po(3.9)$, $P(X=x)$	0.0202	0.0790	0.1540	0.2001	0.4532

The Poisson approximation gives the same value as the true binomial distribution to 2 decimal places. In most practical examples this would be sufficiently close.

If you are new to the Poisson distribution and confident with the binomial you might not regard this as much labour saved. However, substituting into the function $e^{-\lambda}\dfrac{\lambda^x}{x!}$ is easier than calculating binomial probabilities. Furthermore, the real labour saving device is the fact that the cumulative probability may be read off directly from tables. What follows is part of one such table, with the cumulative probability we were seeking marked.

	λ						
	3.30	3.40	3.50	3.60	3.70	3.80	3.90
$x =$ 0	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202
1	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992
2	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531
3	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4532
4	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6478
5	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006
6	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995
7	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546
8	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815
9	0.9978	0.9973	0.9963	0.9960	0.9952	0.9942	0.9931

Reading that entry is a good deal easier than calculating

$$P(X<4)=\binom{156}{0}(0.025)^0(0.975)^{156}+\binom{156}{1}(0.025)^1(0.975)^{155}+\binom{156}{2}(0.025)^2(0.975)^{154}+\binom{156}{3}(0.025)^3(0.975)^{153}$$

Mean and variance of a Poisson distribution

The parameter λ in $P(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}$ for $x=0,\ 1,\ 2,\ 3,\ \ldots.$ is both the mean and the variance of a Poisson distribution. For this reason, when a question on the Poisson distribution is introduced it is often by stating its mean. In symbols

If $X \sim Po(\lambda)$ then $\lambda = E(X) = \text{var}(X)$.

Example (4)

The number of insurance claims made to a company per day may be modelled by a Poisson distribution with mean 3.40.

(a) *Without the use of tables*, calculate

(i) $P(X=4)$

(ii) $P(X \le 1)$

(b) *Using tables*, determine $P(3 \le X \le 8)$.

Solution

(a) (i) $P(X=4)=e^{-3.4}\frac{(3.4)^4}{4!}=0.1858$ (4 d.p.)

(ii) $P(X=0)=e^{-3.4}=0.033373$

$P(X=1)=e^{-3.4}\frac{(3.4)^1}{1!}=0.113469$

$P(X \le 1)=P(X=0)+P(X=1)=0.033373+0.113469=0.1468$ (4 d.p.)

(b)

		λ						
	$n=10$	3.30	3.40	3.50	3.60	3.70	3.80	3.90
$x=$	0	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202
	1	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992
	2	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531
	3	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4532
	4	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6478
	5	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006
	6	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995
	7	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546
	8	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815
	9	0.9978	0.9973	0.9963	0.9960	0.9952	0.9942	0.9931

$P(3 \le X \le 8)=P(X \le 8)-P(X \le 2)=0.9917-0.3397=0.6520$ (4 d.p.)

The point of the first part of this last example is that you should be able to calculate a probability in a Poisson distribution even if, in practical applications, you would read values from a table.

Computational formula

There is a *recursion formula* for the calculation of probabilities in a Poisson distribution. When you apply a recursion formula you are repeating a process again and again in order to generate successive values. The formula is

$$P(X=0)=e^{-\lambda}$$

$$P(X=r+1)=\frac{\lambda}{r+1}P(X=r)$$

Example (5)

Given $X \sim Po(1.5)$ use the recursion formula

$$P(X=0)=e^{-\lambda}$$

$$P(X=r+1)=\frac{\lambda}{r+1}P(X=r)$$

to find $P(X=0), P(X=1), \ldots, P(X=5)$. Use your results to find $P(X>4)$

Solution

$$P(X=0)=e^{-1.5}=0.2231$$

$$P(X=1)=\frac{1.5}{1}\times P(X=0)=1.5\times 0.2231=0.3347$$

$$P(X=2)=\frac{1.5}{2}\times P(X=1)=0.2510$$

$$P(X=3)=\frac{1.5}{3}\times P(X=2)=0.1255$$

$$P(X=4)=\frac{1.5}{4}\times P(X=3)=0.0471$$

$$P(X=5)=\frac{1.5}{5}\times P(X=4)=0.0141$$

$$\begin{aligned}P(X>4)&=1-P(x\le 3)\\&=1-P(X=0)+P(X=1)+P(X=2)+P(X=3)\\&=1-0.2231-0.3347-0.2510-0.1255\\&=0.0657\\&=0.066\ (2.S.F.)\end{aligned}$$

Example (6)

The number of hedgehogs that die on a trunk road in Shropshire each day follows a Poisson distribution with mean 1.15. Without using tables, calculate the probability that more than 3 hedgehogs die on a given day on this road. State what needs to be assumed for this situation to be modelled by the Poisson distribution.

Let $X =$ the number of hedgehog deaths per day

Then, $X \sim Po(1.15)$

$$P(X=0) = e^{-1.15} = 0.31664$$

$$P(X=1) = 1.15 \times P(X=0) = 0.36413$$

$$P(X=2) = \frac{1.15}{2} \times P(X=1) = 0.209376$$

$$P(X=3) = \frac{1.15}{3} \times P(X=2) = 0.080261$$

$$\begin{aligned} P(X>3) &= 1 - \{P(X=0) + P(X=1) + P(X=2) + P(X=3)\} \\ &= 1 - \{0.31664 + 0.36413 + 0.20938 + 0.08026\} \\ &= 0.02959 \\ &= 0.030 \ (2.\text{S.F.}) \end{aligned}$$

An assumption for the use of the Poisson distribution is that the events occur randomly in space and time, so that hedgehogs do not cross the road in parties. This may not be valid – for instance, hedgehogs may like to walk in pairs.