Poisson Distribution (7 pages; 26/1/19)

(1) This is a discrete distribution, where

$P(X = r) = e^{-\lambda}\frac{\lambda^r}{r!} \quad (r = 0,1,2, \dots)$

This is the probability of r events occurring in a specified interval (usually of time, but sometimes length). [In these notes, a period of time will usually be assumed, but any comments will apply equally to intervals of length.]

We say that $X \sim Po(\lambda)$, and we shall see that $E(X) = \lambda$.

(2) Examples

Number of fires that break out in a certain city over a period of 1 hour

Goals scored in a football match

Number of buses passing a particular point in a given time

Faults in a given length of material

(3) A Poisson model can be thought of as the limiting case of a Binomial model, with $X \sim B(n, p)$, as $n \to \infty$ & $p \to 0$, in such a way that $\lambda = np$ is a constant.

Thus the n trials of the Binomial model become infinitesimal portions of a continuous interval.

(4) It can be shown that $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$

and so $\sum_{r=0}^{\infty} e^{-\lambda}\frac{\lambda^r}{r!} = e^{-\lambda}\sum_{r=0}^{\infty}\frac{\lambda^r}{r!} = e^{-\lambda}.e^{\lambda} = 1$,

as required for a probability distribution.

(5) Conditions required for a Poisson model to be appropriate

(i) Events are ‘rare’ and occur singly.

Thus, in the fire example, there are appreciable gaps between the moments when fires break out. (Even though there may be, say, 3 fires occurring in one hour on average, this counts as 'rare'.)

[However, this is not usually as important as the following two conditions.]

(ii) Events are random and occur independently of each other.

As an example of a (non-Poisson) variable, where the events are random, but not independent: it could be the case that once an event has occurred, the probability of another occurring over a given period is changed. (Two events are independent if the probability of one of them occurring is not affected by the occurrence of the other.)

As an example of a (non-Poisson) variable, where the events are independent but not random, it could be the case that events are bound to occur at certain times.

(iii) Events occur at a uniform rate over the specified period.

There is scope for confusion here in the use of the word 'uniform' (or 'constant'). Note first of all that each Poisson distribution has a specific period associated with it (such that λ is the mean number of occurrences in that period).

Example 1: Goals scored in a football match

Say that the period associated with the distribution is the duration of the match. Suppose that data has been collected, based on a large number of matches, and we are told that, on average, there are 2.3 goals per match. Then, the condition about the uniform rate means that, if we were to consider any part of the match, then the expected number of goals would be proportionately reduced (eg it would be 0.23 for a period of $\frac{1}{10}$ of the match).

One objection to the model could be that the figure of 2.3 was not applicable in a particular match (if one team was particularly strong, say). This may be a valid objection (and it could be described as the rate 'not being constant'), but it isn't anything to do with the condition that 'events occur at a uniform rate over the specified period'.

Example 2: Fires breaking out in a city

Here it could be the case that data has been collected over a period of 24 hours (unwisely), but that the period associated with the distribution is only 3 hours (so that we are interested in the probability of a certain number of fires occurring in a 3 hour period).

An obvious objection to the model is that the rate of fires occurring is going to vary considerably over a 24 hour period, so that the average rate obtained is unlikely to be representative of most 3 hour periods.

Once again, this is a separate issue to the consideration of whether the rate is uniform over the 3 hour period. For most 3 hour periods, we can reasonably assume that the rate will be uniform.

(iv) There is another condition, relating to the variance of the distribution, which will be discussed later.

Conditions (i)-(iii) can often overlap, depending on the situation.

In the case of buses passing a particular point, for example, we could make the following comments:

(a) As buses often arrive together at a stop (ie they bunch up), the events are arguably not always occurring singly, and they are not independent.

(b) The bus timetable means that events don't occur at a uniform rate over the period, and are not random.

Exam mark schemes are not going to be concerned with the finer subtleties of the issues. The points to make will be:

(i) Standard comments; ie

(a) random and independent [don't give 'random' and 'independent' as separate points, if asked for two points]

(b) uniform rate

(c) ‘rare’ and occur singly [this is usually of less importance than (a) and (b)]

(ii) Obvious 'context' points (eg the issue above of the number of goals depending on the standard of the teams in question)

Be wary of making original points, as they won't be on the mark scheme, and may well attract no marks if they cannot be linked to an idea on the mark scheme.

(6) Cumulative probability tables exist (see Appendix), but the 2017 specifications assume that calculators will be used.

If an exam question doesn't specify otherwise, it is usually safe to use a calculator. (However, it is probably best to do simple calculations such as $P(X = 2)$ manually, and just do cumulative probabilities on the calculator.)

Notes:

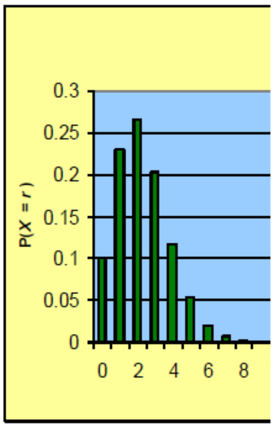
(i) $P(X > x) = 1 - P(X \leq x)$

(ii) $P(X = x) = P(X \leq x) - P(X \leq x - 1)$

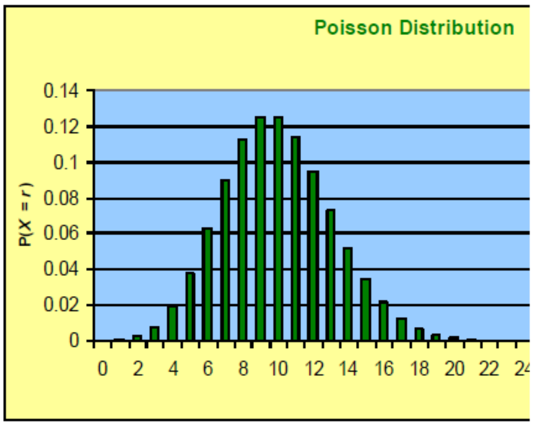
[as an alternative to using the formula $P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$]

(7) Shape of the Poisson distribution

‘$\lambda = 2.3$’



‘$\lambda = 10$’



As λ increases, the Poisson distribution moves from being positively skewed to being approximately symmetrical and Normal (at about $\lambda = 10$), and for this reason the cumulative Poisson tables don't extend beyond $\lambda = 10$, as a Normal approximation can be used instead (see separate note).

(8) Mean of the Poisson distribution

Method 1

$$E(X) = \sum_{r=0}^{\infty} r\, e^{-\lambda}\frac{\lambda^r}{r!} = \sum_{r=1}^{\infty} r\, e^{-\lambda}\frac{\lambda^r}{r!} = \lambda e^{-\lambda}\sum_{r=1}^{\infty} r\frac{\lambda^{r-1}}{r!}$$

$$= \lambda e^{-\lambda}\sum_{r=1}^{\infty}\frac{\lambda^{r-1}}{(r-1)!} = \lambda e^{-\lambda}\sum_{u=0}^{\infty}\frac{\lambda^u}{u!} = \left(\lambda e^{-\lambda}\right)e^{\lambda} = \lambda$$

Method 2

$$\sum_{r=1}^{\infty} r\frac{\lambda^{r-1}}{r!} = \sum_{r=0}^{\infty} r\frac{\lambda^{r-1}}{r!} = \frac{d}{d\lambda}\left(\sum_{r=0}^{\infty}\frac{\lambda^r}{r!}\right) = \frac{d}{d\lambda}\left(e^{\lambda}\right) = e^{\lambda}$$

so that $\text{E(X)} = \lambda e^{-\lambda}\, e^{\lambda} = \lambda$

(9) Variance of the Poisson distribution

$Var(X) = E(X^2) - [E(X)]^2$

$= E[X(X-1)] + E(X) - [E(X)]^2$

$E[X(X-1)] = \sum_{r=0}^{\infty} r(r-1)e^{-\lambda}\frac{\lambda^r}{r!}$

$= \lambda^2 e^{-\lambda} \sum_{r=0}^{\infty} r(r-1)\frac{\lambda^{r-2}}{r!}$

$= \lambda^2 e^{-\lambda}\frac{d^2}{d\lambda^2}\left(\sum_{r=0}^{\infty}\frac{\lambda^r}{r!}\right) = \lambda^2 e^{-\lambda}\frac{d^2}{d\lambda^2}\left(e^{\lambda}\right) = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$

so that $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

(10) Another Poisson condition

Using the fact that $Var(X) = \lambda$, any Poisson data can be studied, to see if $\bar{x}$ and s^2 are suitably close, and (assuming that this is the case) $\bar{x}$ can be used as an estimate for λ.

(11) Sum of Poisson distributions

If $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$, then it can be shown that

$X + Y \sim Po(\lambda + \mu)$, **provided that X and Y are independent**

eg X = number of single-decker buses passing in an hour, and

Y = number of double-decker buses passing in an hour

Note: In theory, the two variables could relate to different time intervals (or even to a time interval and a length), but in that case $X + Y$ would not usually have any practical significance.

Appendix: Cumulative probability tables

POISSON CUMULATIVE DISTRIBUTION FUNCTION

The tabulated value is $P(X \le x)$, where X has a Poisson distribution with parameter λ.

$\lambda =$	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$x = 0$	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.9098	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9856	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.9982	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9998	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	1.0000	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160
6	1.0000	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8	1.0000	1.0000	1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9976	0.9945
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000