Chapter 8

Poisson population distribution $X \sim P(\lambda)$

8.1 Definition of a Poisson distribution, $X \sim P(\lambda)$

If the random variable X has a Poisson population distribution, i.e., $X \sim P(\lambda)$ then its probability function is given by

$$\boxed{p(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}} \qquad \text{for } x = 0, 1, 2, 3, \ldots\ldots, \quad \lambda > 0$$

A Poisson distribution is often a good model for many naturally occurring discrete random variables which are counts, e.g.

X = number of events occurring in a unit of time (or unit of space)

X = number of air collisions per month
X = number of cases of a rare disease in a city per month (or a country per day)
X = number of cars passing a point on a road per minute
X = number of telephone calls to a telesales company per minute
X = number of alpha particles emitted by a (slowly decaying) radioactive source per second
X = number of plants of a particular species per square metre of land

NOTE the range of possible values of the variable X is $x = 0, 1, 2, 3, \ldots\ldots,$
i.e. X is a discrete variable which takes non-negative integer values.

Examples of Poisson distributions

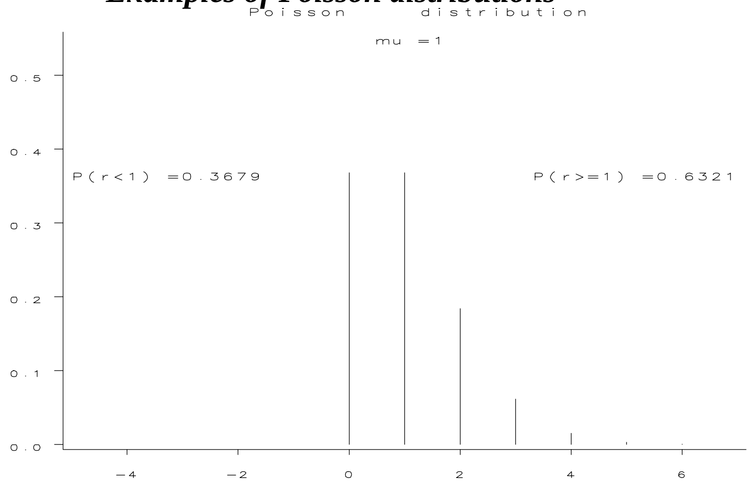


Figure 1 Poisson distribution with μ =1

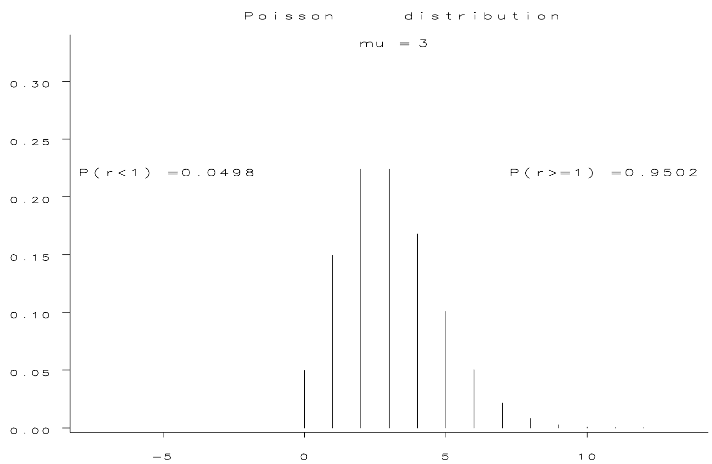


Figure 2: Poisson distribution with μ =3

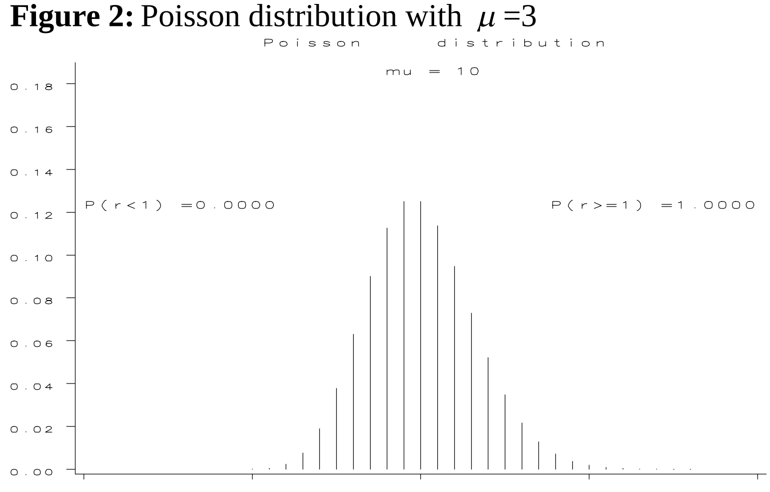


Figure 3 Poisson distribution with μ =10

8.2 Theoretical conditions leading to a Poisson random variable $X \sim P(\lambda)$

Theorem 1
It can be proved that if X counts the number of events occurring in a unit of time (or space), where

i) the rate at which the events occur is constant over time (or space)
ii) the occurrence of events are independent of each other

then X will have a Poisson distribution. [This is called a Poisson process]

Theorem 2
It can also be proved that a Poisson $P(\lambda)$ distribution is a limiting distribution of a Binomial $B(n, p)$ distribution where $p = \frac{\lambda}{n}$ as $n \to \infty$, i.e.

$$P(\lambda) = \lim_{n\to\infty} B(n, \frac{\lambda}{n})$$

Hence the Poisson distribution provides a good approximation to a Binomial distribution when n is large and p is small.

Consequently if X counts the number the number of rare events out of a large number of trials, then X will have approximately a Poisson distribution.

This relates to theorem 1 since a time (or space) interval can be subdivided into a large number of smaller time intervals in each of which the probability of an event occurring is very small, i.e. n is large and p is small.

8.3 Population summary measures for a Poisson random variable $X \sim P(\lambda)$

mean $\mu_X = \lambda$

variance $\sigma_X^2 = \lambda$

standard deviation $\sigma_X = \sqrt{\lambda}$

Derivation of the population mean for a Poisson random variable $X \sim P(\lambda)$

$$\mu = \sum_{x=0}^{\infty} x p(X = x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$
$$= \lambda e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right] = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

8.4 Example

8.4.1 Data

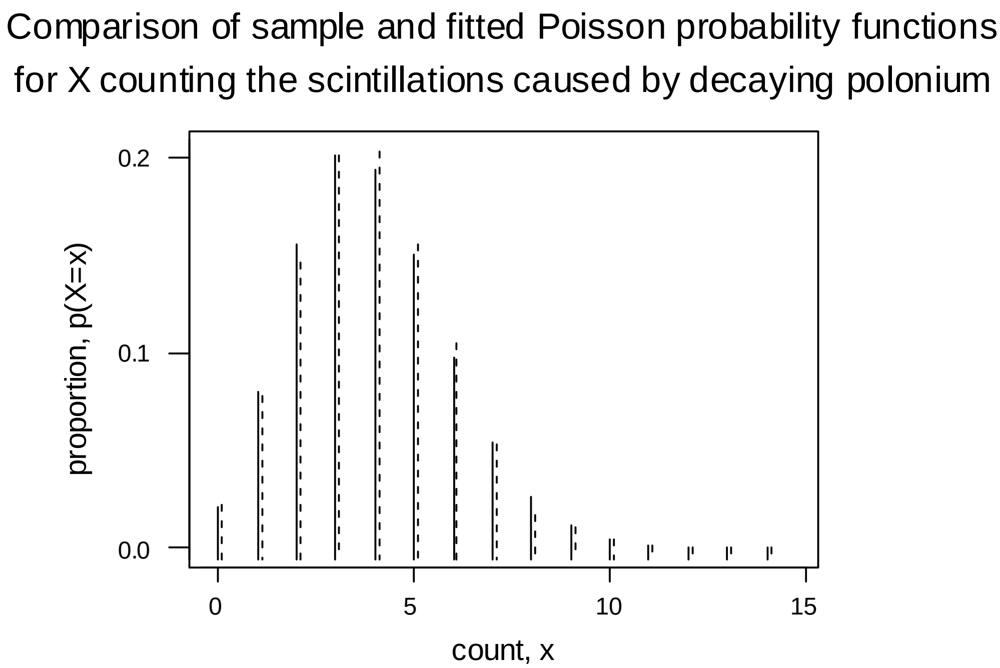
The count *X* of the number of scintillations caused by the radioactive decay of polonium was recorded in a sequence of 2608 intervals giving the following frequency counts:

Count, *x*	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Freq, *f*	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1

[From: Rutherford, E and Geiger, M (1910) *Philosophical Magazine, Series 6*, **20**, 698-704

and Hand et al. 'Small data sets' p223, COUNTING.DAT, data set 279]

8.4.2 Comparison of the sample and fitted Poisson probability functions for *X* counting the number of scintillations in decaying polonium



Conclusion

The Poisson model $P(3.87)$ with mean 3.87 provides a very good model for the variable X, from the comparison of the sample and Poisson $P(3.87)$ model probability functions.

Comparison of the sample and fitted Poisson probability functions for *X* counting the number of scintillations in decaying polonium using MINITAB

1) Type the above counts and frequencies into column 1 and 2 respectively of MINITAB.

2) Now calculate the sample mean from the frequency data,

i.e. $\bar{x} = \frac{1}{n}\sum_{j=1}^{J} f_j x_j$ where $n = \sum_{j=1}^{J} f_j$ and store the result in K4

> Calc > Calculator | Store in **'C3'** | Expression **'C2*C1'** | OK

> Calc > Column Statistics | ⊙ Sum | Input variable **'C2'** | Store in **'K2'** | OK

> Calc > Column Statistics | ⦿ Sum | Input variable **'C3'** | Store in **'K3'** | OK

> Calc > Calculator | Store in **'K4'** | Expression **'K3/K2'** | OK

> Manip > Display Data | **'K4'** | OK

3) Now calculate the sample and fitted Poisson probability functions (in C4 and C5)

> Calc > Calculator | Store in **'C4'** | Expression **'C2/K2'** | OK

> Calc > Random Distributions > Poisson

| ⦿ Probability | Mean **'K4'** | Input Variable **'C1'** | Output Variable **'C5'** | OK

4) Now plot the sample and fitted Poisson probability functions

[The *x* values in C1 are shifted by 0.1 to give C6 to separate the sample and fitted pf's]

> Calc > Calculator | Store in **'C6'** | Expression **'C1+0.1'** | OK

> Graph > Plot

| Y **'C4'** | X **'C1'**| Y **'C5'**| X **'C6'** | Display > Project

| Frame > Axis | Label 1 **'x'** | Label 2 **'probability p(X=x)'** | OK

| Frame > Multiple Graphs | ⦿ Overlay | OK

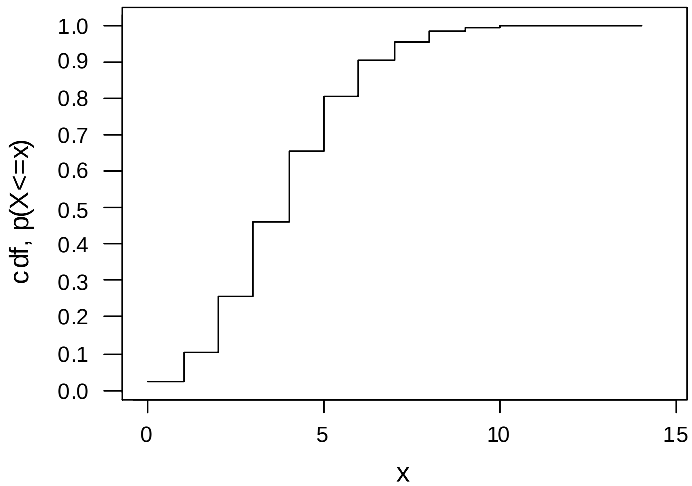
Now calculate and plot the fitted cumulative distribution function (cdf)

| ⦿ Cumulative Prob | Mean **'3.84'** | Input Variable **'C1'** | Output Variable **'C7'** | OK

> Graph > Plot

| Y **'C7'** | X **'C1'** | Display > Connect | > Edit Attributes | ⦿ Step > left | OK

| OK



C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
x	**f**	**x.f**	**sample pf**	**pf P(3.87)**	**x+0.1**	**cdf P(3.87)**
0	57	0	0.021856	0.020826	0.1	0.020826
1	203	203	0.077837	0.080629	1.1	0.101455
2	383	766	0.146856	0.156080	2.1	0.257535
3	525	1575	0.201304	0.201424	3.1	0.458959
4	532	2128	0.203988	0.194955	4.1	0.653914
5	408	2040	0.156442	0.150956	5.1	0.804870
6	273	1638	0.104678	0.097406	6.1	0.902276
7	139	973	0.053298	0.053873	7.1	0.956149
8	45	360	0.017255	0.026071	8.1	0.982220
9	27	243	0.010353	0.011215	9.1	0.993436
10	10	100	0.003834	0.004342	10.1	0.997778
11	4	44	0.001534	0.001528	11.1	0.999306
12	0	0	0.000000	0.000493	12.1	0.999799
13	1	13	0.000383	0.000147	13.1	0.999946
14	1	14	0.000383	0.000041	14.1	0.999986

8.4.3 Applying the model

The model could now be used to obtain estimated answers to questions relating to random variable X, assuming

$$X \sim P(3.87)$$

Use MINITAB to output the values of the pf and cdf for the $P(3.87)$ model

Probability Function

> Calc > Random Distributions > Poisson

| ⦿ Probability | Mean **'3.87'** | Input Variable **'C1'** | OK

```
Poisson with mu = 3.87000

        x      P( X = x)
     0.00        0.0209
     1.00        0.0807
     2.00        0.1562
     3.00        0.2015
     4.00        0.1949
     5.00        0.1509
     6.00        0.0973
     7.00        0.0538
     8.00        0.0260
     9.00        0.0112
    10.00        0.0043
    11.00        0.0015
    12.00        0.0005
    13.00        0.0001
    14.00        0.0000
```

Cumulative Distribution Function

> Calc > Random Distributions > Poisson

| ⦿ Cumulative Prob | Mean **'3.87'** | Input Variable **'C1'** | OK

```
Poisson with mu = 3.87000

        x     P( X <= x)
     0.00        0.0209
     1.00        0.1016
     2.00        0.2578
     3.00        0.4593
     4.00        0.6542
     5.00        0.8051
     6.00        0.9024
     7.00        0.9562
     8.00        0.9823
     9.00        0.9935
    10.00        0.9978
    11.00        0.9993
    12.00        0.9998
    13.00        0.9999
    14.00        1.0000
```

Applying the model

The pf and pdf of $X \sim P(3.87)$ output from MINITAB can be used to answer the following:

Q1 Estimate the probability that the number scintillations X is less than or equal to 5 $p(X \leq 5) = F_X(5) = 0.8051$

Q2 Estimate the probability that the number scintillations X is greater than 8
$p(X \leq 8) = F_X(8) = 0.9823$
$p(X > 8) = 1 - p(X \leq 8) = 1 - F_X(8)$ = 1 - 0.9823 = 0.0177

Q3 Estimate the probability that the number scintillations X is from 6 and 8 inclusively $p(6 \leq X \leq 8) = p(X \leq 8) - p(X \leq 5) = F_X(8) - F_X(5)$ = 0.9823 - 0.8051 = 0.1772

Q4 Estimate the lower quartile Q_1 of X
$F_X(Q_1) = p(X \leq Q_1) = 0.25$, and hence $Q_1 = 2$

8.4.4 MINITAB commands for calculating pdf, cdf and inverse cdf values

Calculating cdf $F_X(x) = p(X \leq x) = p$ using MINITAB

Q1 To find $F_X(5) = p(X \leq 5)$ where $X \sim P(3.87)$
> Calc > Probability Distributions > Poisson
| ⦿ Cumulative Probability | Mean **'3.87'** | Input constant **'5'** | OK

Calculating inverse cdf $x = F_X^{-1}(p)$ using MINITAB

Q2 To find the lower quartile Q_1 of X when $X \sim P(3.87)$
$F_X(Q_1) = p(X \leq Q_1) = 0.25$ so that $Q_1 = F_X^{-1}(0.25)$
> Calc > Probability Distributions > Poisson
| ⦿ Inverse Cumulative | Mean **'3.87'** | Input constant **'0.25'** | OK

Calculating pf $p(X = x)$ using MINITAB

Q3 To find $p(X = 5)$ when $X \sim P(3.87)$
> Calc > Probability Distributions > Poisson
| ⦿ Probability | Mean **'3.87'** | Input constant **'5'** | OK