



Tutorials and worked examples for simulation,
curve fitting, statistical analysis, and plotting.
<http://www.simfit.org.uk>

The Poisson distribution is widely used to model the occurrence of events in time or space. It can be presented as a limiting form of the binomial distribution, or more formally by way of the Poisson postulates defined using $P_n(h)$ for the probability of n events occurring in an interval of width h in time (or space) as follows.

1. The number of events in nonoverlapping intervals of time are independent.
2. Probability does not change during the occurrence of the events.
3. Probability of 1 event in a small interval of time is approximately proportional to the size of the interval.
4. Probability of 1 event in a small interval of time is much larger than that for occurrence of multiple events.

Definitions

A random integer variable X that can take all values ≥ 0 is said to obey the Poisson distribution with parameter $\lambda > 0$ if the discrete probability mass function $f(x)$ is

$$f(x) = \frac{\lambda^x}{x!} \exp(-\lambda)$$

with mean and variance both equal to λ .

Example 1: *Counting arbitrary independent events.*

In the case of the binomial distribution with parameters N and p where N is very large and p very small, then the following approximation becomes valid

$$\binom{N}{k} p^k (1-p)^{N-k} \approx \frac{(Np)^k}{k!} \exp(-Np).$$

In other words, the Poisson distribution with one parameter $\lambda = Np$ becomes a good approximation to the binomial distribution with two parameters N and p when $N \rightarrow \infty$ and $p \rightarrow 0$ but Np remains finite.

Example 2: *Counting independent events as a function of time.*

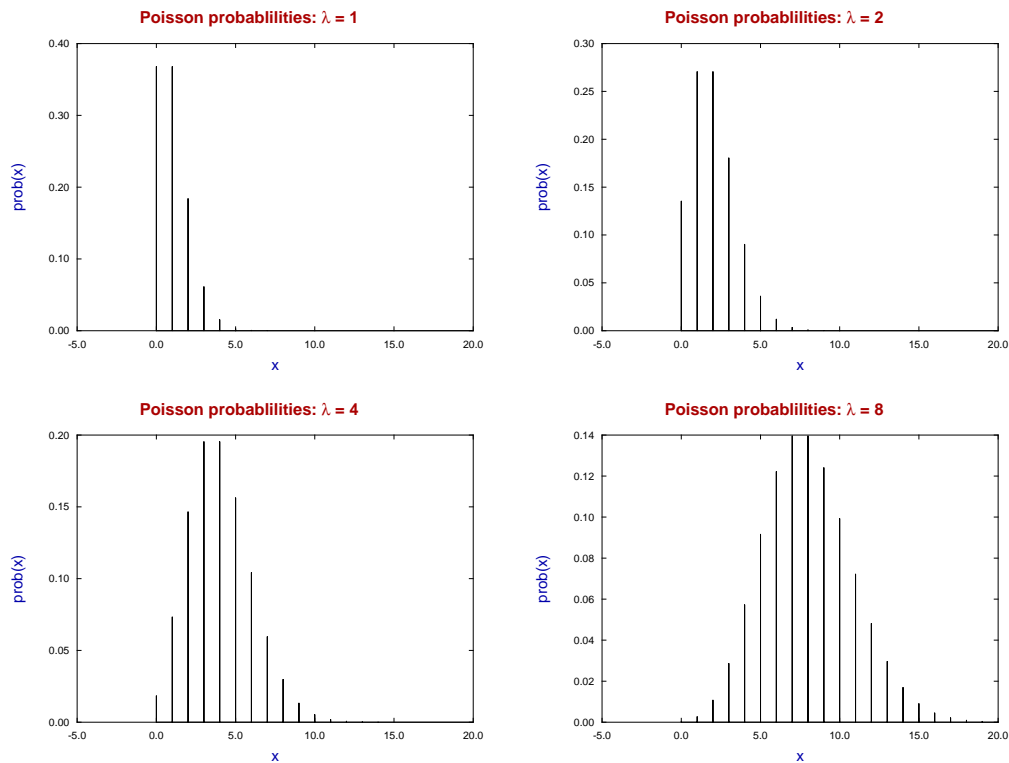
Again, for a process with an average rate of μ events per unit of time, then the probability of k events in time interval t is

$$P_k(t) = \frac{(\mu t)^k}{k!} \exp(-\mu t),$$

which defines a Poisson process with parameter $\lambda = \mu t$.

Plotting Poisson probabilities

The next plots illustrate how the distribution moves to the right as λ increases.



Simfit program binomial

Choose [A/Z] from the main SIMFIT menu and open program **binomial** when the following Poisson options will be available.

```

Input: Poisson x ... calculate pmf(x)
Input: Poisson x ... calculate cdf(x)
Input: Poisson % ... calculate x-critical
Input: Poisson x, estimate lambda and con.lim.
Input: a sample, test if distributed P(lambda)
Calculate: power and sample size
Calculate: change confidence limits (now 95%)
Calculate: using the non-central beta distribution

```

Choosing to analyze test file `poisson.tf1` for consistency with a Poisson distribution using a dispersion test, and also a Fisher exact test first warns that Bonferroni $n = 2$ then outputs these results.

Sample size	40
Sample total	44
Sample sum of squares	80
Sample mean	1.1
Lower 95% confidence limit	0.7993
Upper 95% confidence limit	1.477
Sample variance	0.8103
Dispersion (D)	28.73
$P(\chi^2 \geq D)$	0.88632
Degrees of freedom	39
Fisher exact Probability	0.91999

Note that the Bonferroni $n = 2$ declaration is a warning not to use both test statistics uncritically. Actually SIMFIT often lists the results of several tests at the same time, but this is only for convenience, and users should always take note if a Bonferroni correction is required.

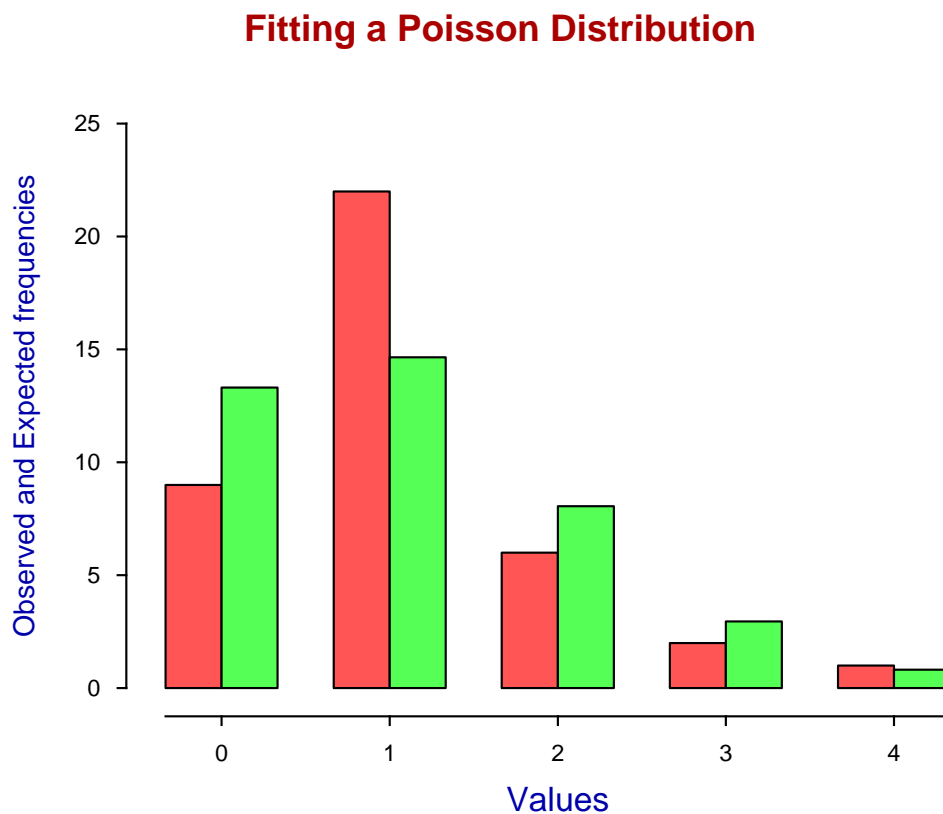
It is frequently required to confirm that it is sensible to use the Poisson distribution, with all the associated assumptions that are involved, as a model when analyzing a given data set. The dispersion test examines if there is any evidence that the dispersion D

$$D = \sum_{i=1}^n (x_i - \bar{x})^2 / \bar{x}$$

is significantly greater than 1 (indicating over-dispersion, i.e. clumping or clustering) or significantly less 1 (indicating under-dispersion, i.e. too evenly scattered), while the Fisher exact test, which can only be done with small samples, estimates the probability of the sample based on all partitions consistent with the sample size, mean, and total. In this case there seems no evidence to reject the null hypothesis

H_0 : the sample is consistent with a Poisson distribution.

The following plot compares the observed and expected frequencies in order to visualize the goodness of fit.



Note the use of a Poisson distribution to assess the significance of k , a small number of counts for one outcome, out of total number number $n > k$, by the rule of thumb of taking a 95% confidence range for the population parameter K as $k - 2\sqrt{k} \leq K \leq k + 2\sqrt{k}$.

Simfit program chisqd

Radioactive decay is an exponential process but, during a sufficiently small time interval where the decay rate can be regarded as approximately constant, particle emission follows a Poisson distribution. In a famous experiment Rutherford counted k , the number of particles emitted in 2608 intervals of 7.5 seconds to obtain the following results, where the expected values were calculated using $\lambda = 10094/2608 = 3.87$.

k	Observed	Expected
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
≥ 10	16	17.075

The SIMFIT program **chisqd** was used to analyze the observed and expected frequencies to obtain these results.

Number of partitions (bins)	11	
Number of degrees of freedom	9	
Chi-square test statistic C	12.88	
$P(\chi^2 \geq C)$	0.1679	Consider accepting H_0
Upper tail 5% critical point	16.92	
Upper tail 1% critical point	21.67	

and the following bar chart.

