The Poisson Distribution

Math-570 Project

Mathematical Theory

- Bernoulli (binomial) distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- N and p combinations under the constraint that np = 1

n	p	$X=0$	$X=1$	$X=2$	$X=3$	$X=4$	$X=5$	$X=6$
4	0.25	0.316	0.422	0.211	0.047	0.004		
5	0.20	0.328	0.410	0.205	0.051	0.006	0.000	
10	0.10	0.349	0.387	0.194	0.058	0.011	0.001	0.000
20	0.05	0.359	0.377	0.189	0.060	0.013	0.002	0.000
100	0.01	0.366	0.370	0.185	0.061	0.014	0.003	0.001
1000	0.001	0.368	0.368	0.184	0.061	0.015	0.003	0.001
10000	0.0001	0.368	0.368	0.184	0.061	0.015	0.003	0.001

From the binomial distribution :

- $\lim\limits_{n\to\infty} P(X = k) = \lim\limits_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-\mathrm{k}}$

$$= \left(\frac{\lambda^k}{k!}\right) \lim_{n\to\infty} \underbrace{\frac{n!}{(n-k)!} \left(\frac{1}{n^k}\right)}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to \mathrm{e}^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-\mathrm{k}}}_{\to 1}$$

$$= \left(\frac{\lambda^k}{k!}\right) (1)\, e^{-\lambda}\, (1)$$

Formula

The Poisson distribution

$$P(X = k) = \frac{\lambda^k \, e^{-\lambda}}{k!}$$

Cumulative distribution

$$P(X \leq k) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$$

Computations

The Poisson distribution model can be used if the following set of canonical assumptions are valid:

- The number of occurrences for a certain event in an interval k can take values 0, 1, 2,..
- Events occur independently.
- The average rate at which events occur is independent of any occurrences. This is usually assumed to be constant.
- Two events cannot occur at exactly the same instant.

Example 1

- If the random variable X follows a Poisson distribution with mean 3.4 as an example, you can find the probability for k = 6 occurrences as:

$$P(X=6) = \frac{e^{-\lambda}\lambda^{6}}{6!}$$

$$= \frac{e^{-3.4}(3.4)^{6}}{6!} \quad (\text{mean}, \lambda = 3.4)$$

$$= 0.071\,604\,409 = 0.072 \quad (\text{to 3 d.p.}).$$

Example 2

- The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with a mean of 0.5. In a particular week, the probability of having less than two accidents can be looked at as: $k < 2$ is not an exact one number, the probability can be computed as:

$$
\begin{aligned}
P(A < 2) &= P(A = 0) + P(A = 1) \\
&= e^{-0.5} + \frac{e^{-0.5} \times 0.5}{1!} \\
&= \frac{3}{2} e^{-0.5} \\
&\approx 0.9098.
\end{aligned}
$$

- To get the probability for more than 2 accidents: $P(X > 2) = 1 - P(X \leq 2)$

$$1 - \left[P(A=0) + P(A=1) + P(A=2)\right]$$

$$= 1 - \left[e^{-0.5} + e^{-0.5}0.5 + \frac{e^{-0.5}(0.5)^2}{2!}\right]$$

$$= 1 - e^{-0.5}(1 + 0.5 + 0.125)$$

$$= 1 - 1.625\, e^{-0.5}$$

$$\approx 0.0144\,.$$

- If for example, we want the probability in a three week period there will be no accidents.

$$P\,(0 \text{ in } 3 \text{ weeks}) = \left(e^{-0.5}\right)^3 \approx 0.223\,.$$

Algorithm

1

1. Evaluating a Poisson distribution $P(k,\lambda)$

2

2. Drawing random numbers according to a given Poisson distribution.

Implementation:

- the standard definition of P(k, λ) has the two terms λ^k and $k!$ that can **overflow** on computers very easily.
- They can also give a large rounding error compared to $e^{-\lambda}$ that will produce an erroneous result.

$$f(k;\lambda) = \exp[k \ln \lambda - \lambda - \ln \Gamma(k+1)]$$

- Mathematically equivalent to the standard but **numerically stable**
- **Built-in functions:** R, MATLAB, SciPy, Excel

```
# Documentation:
#https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.poisson.html

# The Poisson Distribution Function takes two arguments: k and λ as:
# scipy.stats.poisson.pmf(k, λ)

answer = scipy.stats.poisson.pmf(6,3.4)

# Print Result
print(answer)
```

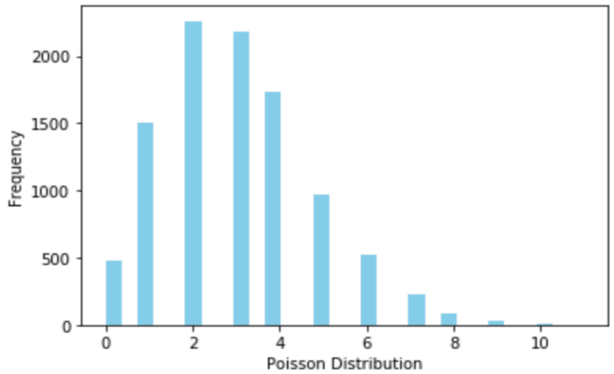
```
0.07160440945982202
```

```
def Poisson_Distribution(lamda, k):

    # Probability at most k
    atmost = scipy.stats.poisson.pmf(k, lamda)
    morethan = 1 - atmost

    print('Probability for k <= λ: ',atmost)
    print('Probability for k > λ: ',morethan)

Poisson_Distribution(2.0, 0.5)
```

```
Probability for k <= λ:  0.0
Probability for k > λ:  1.0
```

```python
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import poisson

data_poisson = poisson.rvs(mu=3, size=10000)

# poisson.rvs() method takes arguments:
# 1. μ: as a shape parameter and is nothing but the λ equasion.
# 2. size: decides the number of random variates in the distribution.

# To maintain reproducibility, include a random_state argument assigned to a number.


# Documentation:
# https://seaborn.pydata.org/generated/seaborn.distplot.html

# Plotting Function
ax = sns.distplot(data_poisson,
                  bins=30,
                  kde=False,
                  kde_kws={"lw": 2},
                  color='skyblue',
                  hist_kws={"linewidth": 15,'alpha':1})

sns.set_color_codes()
ax.set(xlabel='Poisson Distribution', ylabel='Frequency')
```

```
[Text(0, 0.5, 'Frequency'), Text(0.5, 0, 'Poisson Distribution')]
```



Applications

- One famous application of the Poisson distribution was During World War II. A number of sectors in London were being hit by bombs. In total, 537 bombs were recorded, with 576 sectors hit at least once.
- On average, each sector was hit $\lambda = 537/576 = 0.9323$ times.
- Shown in the table below, the number of hits for each sector and the expected hits with the probabilities:

Number of hits (k)	0	1	2	3	4	≥ 5
$\mathbb{P}(k) = \frac{\mu^k}{k!}e^{-\mu}$	0.39365	0.36700	0.17108	0.05316	0.01239	0.00272
Expected number ($573 \times \mathbb{P}(k)$)	226.74	211.39	98.54	30.62	7.14	1.57
Actual number	229	211	93	35	7	1

Other Applications

- The assumption is that the events are independent and λ is constant, generally the Poisson Distribution is being used in:
-
  - The number of mutations on a strand of DNA in a time unit.
  - The number of network crashes per day.
  - The number of filed bankruptcies in one month.
  - The number of arrivals at a car wash in one hour.
  - The number of file server virus infection at a data center during a day.
  - The number of Airbus aircraft engine shutdowns per 100,000 flight hours.
  - The number of birth, deaths, marriages, divorces, suicides, and homicides over a period of time.
  - The number of visitors to a Web site per minute

Thank you