Poisson Distribution

Travis Beach, Nina Brim, Elliot Weiss

Poisson limit

- Used (before the age of fancy calculators!!!) to approximate hard-to-calculate binomial probabilities when $\boldsymbol{n} \to \boldsymbol{\infty}$, $\boldsymbol{p} \to \mathbf{0}$ and ***np*** remains constant.

For binomial random variable **X** with ***n*** trials having probability of success ***p***:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\text{const.}}} P(X=k) = \frac{e^{-np}(np)^k}{k!}$$

Derivation of Poisson Limit

Starting by rewriting the binomial probability where $\lambda = np$ is a constant :

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

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

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

But since $[1-(\lambda/n)]^n \to e^{-\lambda}$ as $n \to \infty$, we need show only that

$$\frac{n!}{(n-k)!(n-\lambda)^k} \to 1$$

to prove the theorem. However, note that

$$\frac{n!}{(n-k)!(n-\lambda)^k} = \frac{n(n-1)\cdots(n-k+1)}{(n-\lambda)(n-\lambda)\cdots(n-\lambda)}$$

a quantity that, indeed, tends to 1 as $n \to \infty$ (since λ remains constant).

$$\frac{e^{-np}(np)^k}{k!}$$

Example

In 2008, 137.8 million tax returns were filed in the US. Out of these, 1735 people were convicted of tax fraud and went to jail. In a town of 65,000 people, what is the probability at least 3 people went to jail?

$$n = 65000 \qquad p = \frac{1735}{1378000000} = .0000126$$

(wow that's a big ***n*** and little ***p***, it would be very hard to calculate that without a Poisson Limit!!!!!)

$$P(X \geq 3) = 1 - P(X \leq 2)$$

$$= 1 - \sum_{k=0}^{2} \frac{e^{-np}(np)^k}{k!} \qquad \doteq 1 - \sum_{k=0}^{2} e^{-0.819} \frac{(0.819)^k}{k!} = 0.050$$

Poisson Distribution

- In 1898, over 50 years after Poisson's Limit, Ladislaus von Bortkiewicz transformed the limit into a distribution
- The model,

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

is useful in many situations, even when:

-no explicit binomial random variable is present

-values for *n* and *p* are not available

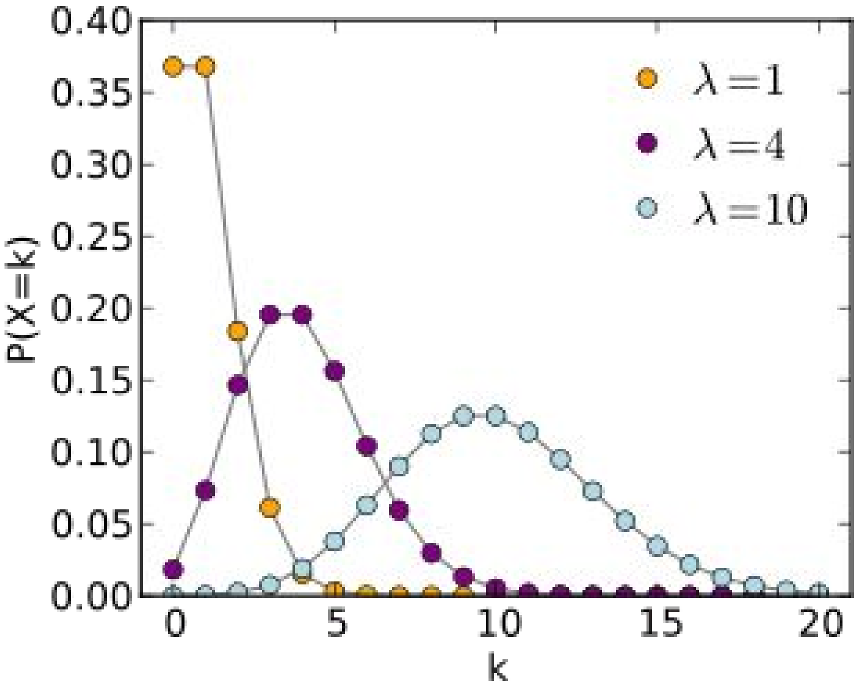
Poisson Distribution

- Used for discrete random variables
- Describes the distribution of the number of times an event occurs in a sample space
- Right-skewed but becomes more symmetric as λ approaches infinity

PDF:

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

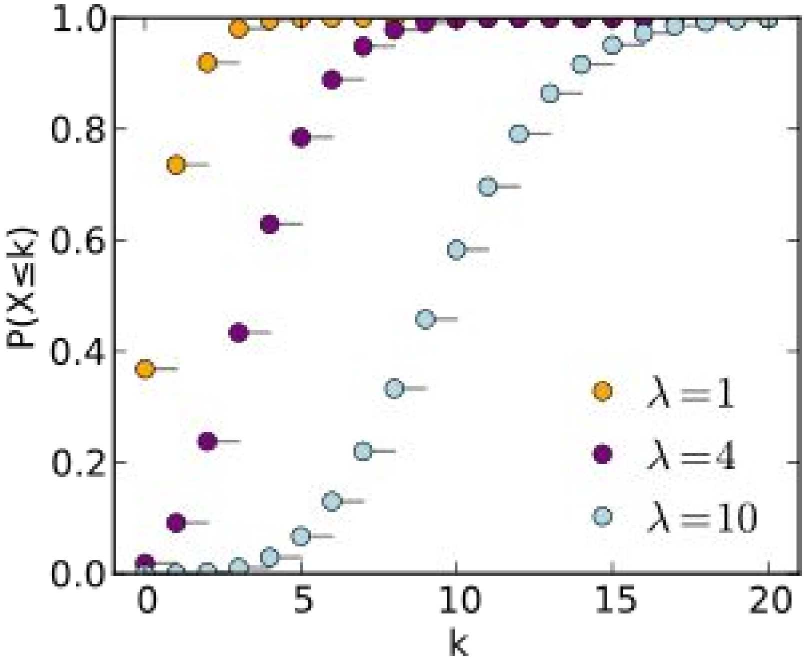
Where k= 0, 1, 2, ... and λ is a positive constant



CDF

Since we are modeling a discrete random variable, the CDF is a sum of the PDFs:

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



MGF

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

$$M_x(t) = \sum_{x=0}^{\infty} e^{tx} p_x(x) \qquad M_x(t) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^k}{k!} \qquad = \sum_{x=0}^{\infty} \frac{(\lambda e^t)e^{-\lambda}}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{a^x e^{-\lambda}}{x!}$$ where $a = \lambda e^t$

$$= \frac{e^{-a}}{e^{-a}} \sum_{x=0}^{\infty} \frac{a^x e^{-\lambda}}{x!} \qquad = \frac{e^{-\lambda}}{e^{-a}} \sum_{x=0}^{\infty} \frac{a^x e^{-a}}{x!}$$

This is the pdf for Poisson dist. where lambda = a! So the summation equals 1!

$$= \frac{e^{-\lambda}}{e^{-a}} \qquad = e^{a-\lambda} \quad = e^{\lambda e^t - \lambda}$$

$$= e^{\lambda(e^t - 1)}$$

Mean and Variance

Mean:

$$M_x(t) = e^{\lambda(e^t-1)}$$

$$E(X) = M_X^{(1)}(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t$$

$$E(X) = M_X^{(1)}(0) = e^{\lambda(e^0-1)} \cdot \lambda e^0$$

$$= e^0 \cdot \lambda$$

$$= \boxed{\lambda}$$

Variance:

$$M_X^{(2)}(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t + \lambda e^t e^{\lambda(e^t-1)} \cdot \lambda e^t$$

$$M_X^{(2)}(0) = e^{\lambda(e^0-1)} \cdot \lambda e^0 + \lambda e^0 e^{\lambda(e^0-1)} \cdot \lambda e^0$$

$$E(X^2) = \lambda + \lambda^2$$

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

$$= \lambda + \lambda^2 - (\lambda)^2$$

$$= \boxed{\lambda}$$

Calculating Poisson Probabilities

Three useful formulas for calculating probabilities:

1. $p_x(k) = e^{-np}\frac{(np)^k}{k!}$ where X is a *binomial* random variable

2. $p_x(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ Where **λ** is E(X)

3. $p_x(k) = e^{-\lambda T}\frac{(\lambda T)^k}{k!}$ Where **T** is a constant such that **λT =** E(X)

All 3 equations can be written as:

$$p_x(k) = e^{-E(X)}\frac{[E(X)]^k}{k!}$$

Applications

When is it an appropriate model? When these assumptions are true:

- K is the number of times an event occurs in an interval and K can take values 0, 1, 2, …
- Events are independent
- The rate at which events occur is constant.
- Two events cannot occur at the exact same instant.
- The probability of an event in an interval is proportional to the length of the interval.

Real-life Scenarios

Just how important is the Poisson distribution? Here are just a few real-life scenarios that the Poisson distribution can model (or predict based on previous data):

- The number of patients admitted to an emergency room between 8-9pm
- The number of times a river floods in a year
- The number of meteors larger than 1m in diameter that hits Earth each year
- The number of people diagnosed with Leukemia in Minneapolis over the course of one year

Example problem

Entomologists estimate that an average person consumes almost a pound of bug parts each year. For peanut butter, the legal limit of contamination is 30 insect fragments per 100 grams.

You buy a snack that contains 20 grams of peanut butter. What are the chances that your snack will include at least five crunchy critters?

Let X denote the number of bug parts in 20 grams of peanut butter. Assume the contamination level is the legal limit (30 insect fragments per 100 grams or 0.30 fragment / gram).

Example problem solution

E(X) = λT = (.3 fragment / g) x (20 g) = 6.0

E(X) = 6.0

Since E(x) = λT and $p_x(k) = e^{-\lambda T}\frac{(\lambda T)^k}{k!}$

We find that

$$P(X \geq 5) = 1 - P(X \leq 4) \quad = 1 - \sum_{k=0}^{4} \frac{e^{-6.0}(6.0)^k}{k!}$$

$$= 1 - 0.29$$

$$= 0.71$$

Exponential interval

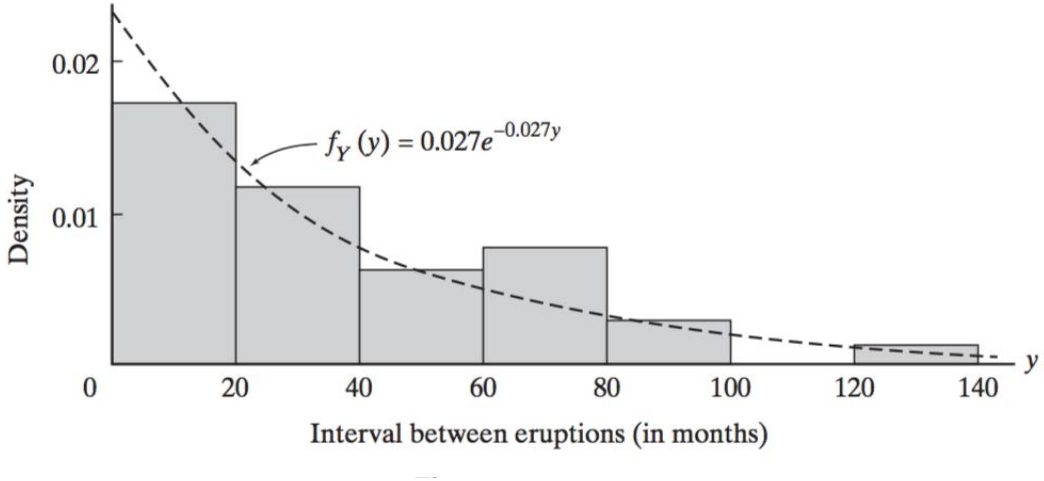
Theorem: Suppose a series of events satisfying the Poisson model are occurring at the rate of λ per unit time. Let the random variable Y denote the interval between consecutive events. Then Y has the exponential distribution:

$$f_Y(y) = \lambda e^{-\lambda y}$$

Note that although the Poisson distribution is only for discrete random variables, these intervals are modeled by a continuous random variable.

Application

Over "short" geological periods, a volcano's eruptions are believed to be Poisson events. (Independent at a constant rate). This means that the interval *between* eruptions can be modeled by $f_Y(y) = \lambda e^{-\lambda y}$.



Mauna Loa, a Hawaiian volcano, erupted 37 times from 1832-1950. The time between each eruption is displayed in the graph below. We can see that the data is consistent with the poisson distribution where lambda is calculated from the eruption rate of 0.027 eruptions / month during that period.

Example Problem

As shown previously, Mauna Loa erupted 37 times over a 118 year period (1832-1950). If an eruption occurs on January 1st, 1951, what is the probability that another eruption will occur in 1951?

$$\lambda = \frac{37}{118} = 0.31 eruptions/year$$

$$P(0 \leq Y \leq 1) = \int_0^1 0.31e^{-0.31y} dy$$

$$f_y(y) = 0.31e^{-0.31y}$$

$$= -e^{-0.31(1)} + e^{-.31(0)}$$

= 0.27

Questions ????????????????????

????

???????? :)

:)

???????????

?????????