STATISTICS FOR DATA SCIENCE

Poisson Distribution

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Poisson Distribution

Prof. Uma D

Topics to be covered...

- Poisson Distribution
- Probability Mass Function
- Students t-distribution
- Mean and Variance of Poisson Distribution
- Using the Poisson Distribution to Estimate a Rate
- Computing uncertainty of λ ^

Poisson Distribution

Siméon Denis Poisson
(1781–1840)

First derived Poisson distribution in 1837

Source: https://en.wikipedia.org/wiki/Sim%C3%A9on_Denis_Poisson

Poisson Distribution

A **Poisson distribution** is the probability distribution that results from a **Poisson experiment**.

Attributes of a Poisson Experiment:

1. The experiment results in **outcomes** that can be classified as **successes** or **failures**.
2. The average number of successes(**λ**) that occurs in a region(length, area, volume, period of time) is known.
3. The **probability** that a **success** will occur is **proportional** to the **size of the region**.
4. The probability that a success will occur in an **extremely small region** is virtually **zero**.

Poisson Distribution

- Poisson distribution is used to describe **number of occurrences** of a **(rare)** **event** that occur **randomly** during a specified interval.

- The interval may be time, distance, area, or volume.

- It describes the frequency of “**successes**” in a test where a “success” is a rare event.

- Events with **low frequency** in a large population follow a **Poisson distribution.**

Examples

- The number of deaths by horse kicking in the Prussian army (First application).
- The number of cyclones in a season.
- Arrival of Telephone calls, Customers, Traffic, Web requests.
- Estimating the number of mutations of DNA after exposure to radiation.
- Rare diseases (like Leukemia(cancer of the blood cells), but not AIDS because it is infectious and so not independent).

Examples

- The number of calls coming per minute into a hotels reservation center.
- The number of particles emitted by a radioactive source in a given time.
- The number of births per hour during a given day.
- The number of patients arriving in an emergency room between 11 -12 pm.
- The number of car accidents in a day.

In such situations we are often interested in whether the events occur randomly in time or space.

PES UNIVERSITY ONLINE

Probability Mass Function of a Poisson Distribution

Let X : Number of events that occurs during a given interval of time.

Let λ : The average rate at which the events are occuring in a given interval of time.

$X \sim \text{Poisson}(\lambda)$

$$P(x) = P(X=x) = P(x;\lambda) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

Mean and Variance of a Poisson Distribution

$$X \sim Poisson(\lambda)$$

$$p(x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

$$\mu_X = \lambda \quad ; \quad Var(X) = \sigma_X^2 = \lambda$$

Problem

If X ~ Poisson (3) then compute P(X=2), P(X=10), P(X=0), P(X=-1) and P(X=0.5).

$$p(X=x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

$$P(X=-1) = 0$$

$$P(X=2) = \frac{e^{-3} \cdot 3^{2}}{2!} = 0.2240$$

$$P(X=0.5) = 0$$

$$P(X=10) = \frac{e^{-3} \cdot 3^{10}}{10!} = 0.0008$$

$$P(X=0) = \frac{e^{-3} \cdot 3^{0}}{0!} = 0.0498$$

Probability Mass Function - Example

What is the probability that 8 or more accidents happen?

$$P(x \geq 8) = \mathbf{1 - P(}x\mathbf{ < 8)}$$

$$= \mathbf{1 - P(}x \leq 7)$$

$$= \mathbf{1 - .999 = .001}$$

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

Probability Mass Function

If X ~ Poisson (4) then compute P(X≤2) and P(X>1).

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!} = 0.2381$$

$$P(X > 1) = P(X=2) + P(X=3) + P(X=4) + \cdots$$

$$= 1 - [P(X=0) + P(X=1)]$$

$$= 0.9084$$

Poisson distribution is as an **approximation to the binomial distribution** when **n is large and p is small**.

Example

- A mass contains 10,000 atoms of a radioactive substance. The probability that a given atom will decay in a one- minute time period is 0.0002. Let X represent the number of atoms that decay in one minute. Now each atom can be thought of as a Bernoulli trial, where success occurs if the atom decays.

Poisson Distribution - Example

$X \sim Bin(n, p)$

$X \sim Bin(10000, 0.000$

Poisson Distribution - Example

$X \sim Bin(n, P)$ $\quad P(X=x) = \frac{n!}{x!\,(n-x)!} p^{x} (1-p)^{n-x}$

$X \sim Bin(\underset{n}{10000}, \underset{p}{0.0002})$

$P(X=3) = \frac{10000!}{3!\,9997!} (0.0002)^{3} (0.9998)^{9997}$ $= 0.18046509$

$Y \sim Bin(5000, 0.0004)$

$P(X=3) = \frac{5000!}{3!\,4997!} (0.0004)^{3} (0.4996)^{4997} = 0.\cancel{180}$

$= 0.18048318$

Poisson Distribution - Example

$X \sim Bin(n, p)$ $\qquad X \sim Bin(10000, 0.0002)$

$\mu_X = n * p$ $\qquad Y \sim Bin(5000, 0.0004)$

$\mu_X = 10000 * 0.0002 = 2$

$\mu_Y = 5000 * 0.0004 = 2$ $\qquad X \sim Poisson(2)$

$X \sim Poisson(\lambda)$ $\qquad P(X=3) = \dfrac{e^{-2} \cdot 2^3}{3!} \simeq 0.180$

$\lambda = n * p = 2$ $\qquad \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \simeq \dfrac{e^{-\lambda} \cdot \lambda^x}{x!}$

Students t-Distribution

William Sealy Gosset (pen name Student)
(1876 – 1937)
English statistician
Famous for Student's t-distribution

Students t-Distribution

- The first biological application of the Poisson distribution was given by 'Student' (1907) in his paper on the **error of counting yeast cells in a haemocytometer**(instrument for counting the no. of cells in a cell suspension.), although he was unaware of the work of Poisson and von Bortkiewicz and derived the distribution afresh.

Students t-Distribution

The t-distribution plays a role in a number of widely used statistical analyses, including Student's t-test for assessing the statistical significance of

- the difference between two sample means,
- the construction of confidence intervals for the difference between two population means, and
- in linear regression analysis.

Students t-Distribution

A normal distribution describes a full population,
t-distributions describe samples drawn from a full population;

Accordingly, the t-distribution for each sample size is different, and the larger the sample, the more the distribution resembles a normal distribution.

Examples

If electricity power failures occur according to a Poisson distribution with an average of 3 failures every twenty weeks, calculate the probability that there will not be more than one failure during a particular week.

Solution:

Number of failures every twenty weeks = 3

X : No. of failures per week

$X \sim Poisson(\lambda t)$

$X \sim Poisson(3/20)$

$X \sim Poisson(0.15)$

$P(X \le 1) = P(X=0) + P(X=1)$

$= \frac{e^{-0.15} \cdot 0.15^0}{0!} + \frac{e^{-0.15} \cdot 0.15^1}{1!}$

$= 0.98981$

Examples

A life insurance salesman sells on the average 3 life insurance policies per week.

Use Poisson's law to calculate the probability that in a given week he will sell.

1) Some policies $P(X > 0) = 1 - P(X \leq 0) = 1 - P(X = 0)$

2) 2 or more policies but less than 5 policies. $P(2 \leq X < 5) = P(X=2) + P(X=3) + P(X=4)$

3) Assuming that there are 5 working days per week, what is the probability that in

   a given day he will sell one policy? $X \sim \text{Poisson}(3/5)$

$$P(X=1) = \frac{e^{-0.6} \cdot (0.6)^1}{1!}$$

Using the Poisson Distribution to Estimate a Rate

Let **λ** denote the **mean number of events** that occur in **one unit** of **time or space**.

Let **X** denote the **number of events** that are observed to occur in **t units** of **time or space**.

Then,

X ~ Poisson(λt)

$$\hat{\lambda} = \frac{X}{t}$$

where λ is estimated with λ^ = X / t

Example

A microbiologist wants to estimate the concentration of a certain type of bacterium in a wastewater sample.

She puts a 0.5 mL sample of the waste-water on a microscope slide and counts 39 bacteria.

Estimate the concentration of bacteria per mL, in this waste-water.

$X \sim \text{Poisson}(\lambda t)$

$X \sim \text{Poisson}(0.5\lambda)$

$\hat{\lambda} = \frac{X}{t} = \frac{39}{0.5} = 78$

$X \sim \text{Poisson}(78)$

Computing bias of λ^

Bias – is intentional or unintentional favoring of one outcome over the other in the population.

In statistics, **Bias** of an estimator is the **difference** between estimator's **expected value** and **true value** of parameter being estimated.

$$\text{Bias} = \mu_{\hat{\lambda}} - \lambda$$

Computing uncertainty of λ^

Uncertainty – is the **standard deviation of sample proportion**.

$$\sigma_{\hat{\lambda}} = \frac{\sigma_x}{t} = \frac{\sqrt{\lambda t}}{t} = \sqrt{\frac{\lambda}{t}}$$

$$X \sim \text{Poisson}(\lambda t)$$

As λ is unknown when computing uncertainty , we approximate it with λ ^.

Example

A 5 mL sample of a suspension is withdrawn, and 47 particles are counted. Estimate the mean number of particles per mL, and find the uncertainty in the estimate.

$$\hat{\lambda} = \frac{X}{t} = \frac{47}{5} = 9.4$$

$$\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda}{t}} = \sqrt{\frac{9.4}{5}} = 1.4$$

Do It Yourself !!!

1. The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident during a one-week period.

2. A suspension contains particles at an unknown concentration of λ per mL. The suspension is thoroughly agitated, and then 8mL are withdrawn and 22 particles are counted. Estimate λ.

Problem

Do It Yourself !!!

A certain mass of a radioactive substance emits alpha particles at a mean rate of λ particles per second. A physicist counts 1594 emissions in 100 seconds. Estimate λ, and find the uncertainty in the estimate.

THANK YOU

Prof.Uma D

Department of Computer Science and Engineering