STATISTICS FOR DATA SCIENCE

Poisson Distribution

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering

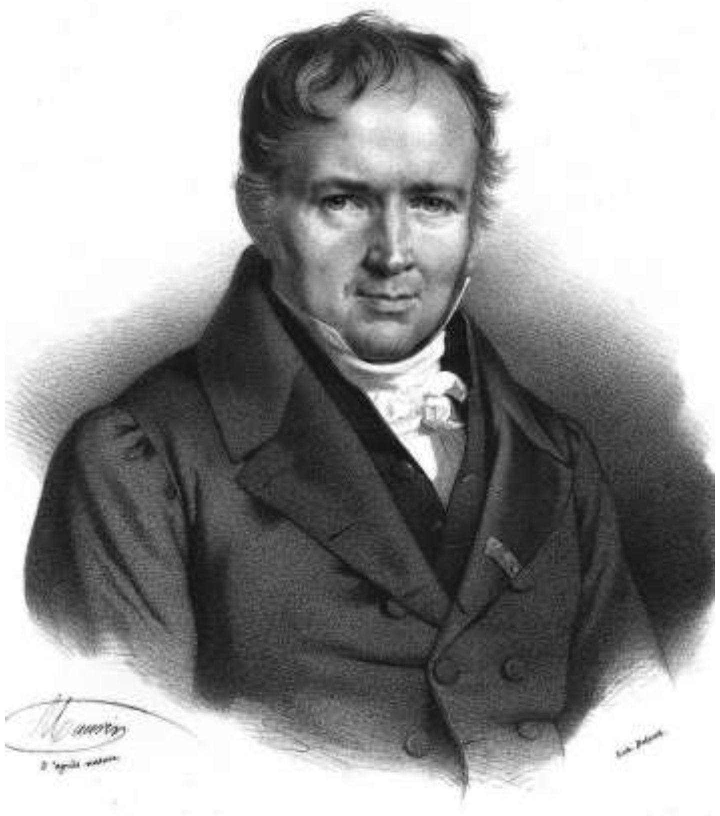
STATISTICS FOR DATA SCIENCE

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 ^



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**.

Source: www.stattrek.com

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.

Probability Mass Function of a Poisson Distribution

Mean and Variance of a Poisson Distribution

Problem

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

Probability Mass Function - Example

What is the probability that 8 or more accidents happen?

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

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

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

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

If $X \sim$ Poisson (4) then compute $P(X \leq 2)$ and $P(X > 1)$.

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

Poisson Distribution - Example

Poisson Distribution - Example

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:

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

2) 2 or more policies but less than 5 policies.

3) Assuming that there are 5 working days per week, what is the probability that in a given day he will sell one policy?

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)

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.

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.

Computing uncertainty of λ^

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

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.

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