Friday, May 10, 2019 9:38 PM

Learning contents: basic statistics, binomial distribution, Poisson distribution, law of large number, normal distribution

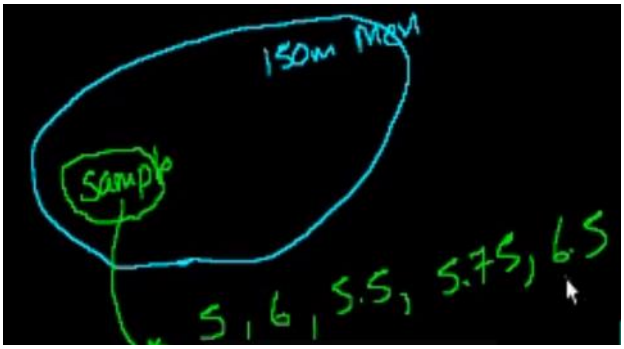
Learning materials: https://www.youtube.com/playlist?list=PL1328115D3D8A2566

Sample and Population:

Population: The average height of all man in America (150 million)

Sample: the average height of some men in America

Randomly selected from the population



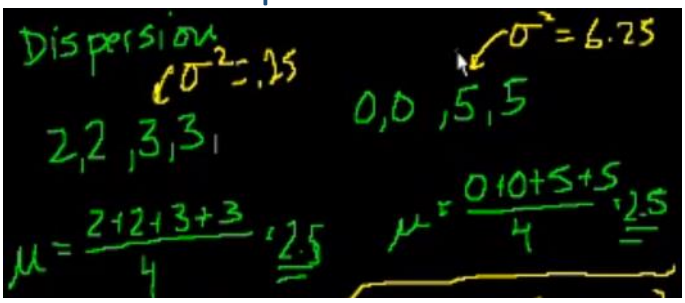
$$\mu = \text{population mean} = \sum_{i=1}^{N} x_i / N$$

$$\bar{x} = \text{sample mean} = \sum_{i=1}^{n} x_i / n$$

Central tendency of the data

We try to pick up some numbers that are most representative of all numbers

Measure of dispersion



Variance of the population

$$\sigma^2 = \sum_{i=1}^{N} (X_i - u)^2 / N$$

Variance of the sample:

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

The sample variance will always under estimate the actual variance of the population :

- because your sample average is always within your data sample, that is when you calculate the sample variance using this formula, your sample is always relatively clustering around your average. So the sample variance will always underestimate the population variance.

So the better estimate is

$$s^2 = s_{n-1}^2 = \frac{\sum (X_i - \bar{x})^2}{n-1}$$

Standard deviation

Standard deviation

$$\sigma = \sqrt{\sigma^2}$$

$$S = \sqrt{S^2}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - u)^2}{N}$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - 2u\sum_{i=1}^{N} x_i + u^2 N\right)$$

$$= \frac{\sum_{i=1}^{N} x_i^2}{N} - 2u^2 + u^2$$

$$= \frac{\sum x_i^2}{N} - u^2$$

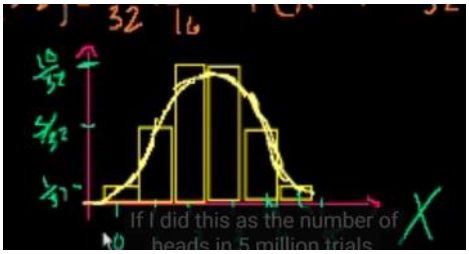
Random variable:

Discrete and continuous

Binomial distribution:

$$A(n,m) = \frac{n!}{(n-m)!}$$

$$C(n,m) = \frac{n!}{m!(n-m)!}$$



Sometimes people make that assumption that the distribution is a binomial distribution or normal distribution, but actually it is not.
So be careful to make assumptions

$$P(x=n) = \frac{5!}{n!(5-n)!} = \binom{5}{n}$$

Factorial in python:

```
>>> import math
>>> math.factorial(3)
6
```

Permutation and combination in python:

```
from itertools import permutations
perm = permutations([1,2,3])
print(len(list(perm)))
```

```
6
```

```
from itertools import combinations
com = combinations([1,2,3],2)
print(len(list(com)))
```

```
3
```

```
from itertools import permutations
perm = permutations([1,2,3],2)
print(list(perm))
```

```
[(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)]
```

Expected value of binormal distribution:

Sometimes the expected value is the population mean

Given the weights/frequency of all possible value, then calculate the average of the population

We can not add up an infinite number of data points and Divide by infinite number

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

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} \cdot p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-k)!} \cdot p^{k-1} (1-p)^{n-k}$$

$$a = k - 1$$

$b = n-1$

$n-k = a-b$

$$= np \underbrace{\sum_{a=0}^{b} \frac{b!}{a!(b-a)!} \cdot p^a (1-p)^{b-a}}_{=1}$$

$$= np$$

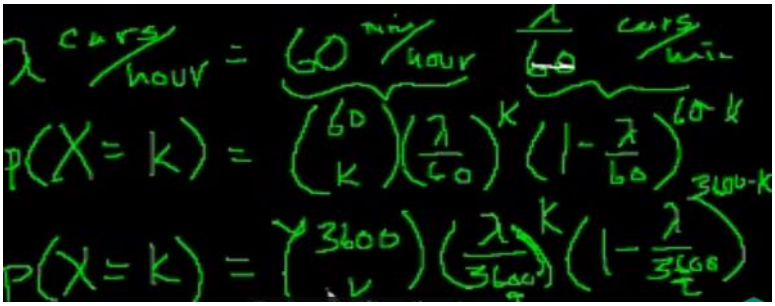
Poisson process:

X = number of cars pass in an hour
Two arbitrary assumption:

- Any hour is no different from any other hour in this street
- Any hour is independent.

just try to use binomial distribution

E(X) = np
Lambda :the cars passing in one hour
n : the cars passing in one minute/second
P : the probability of car passing



As the number of intervals gets bigger…

$$\lim_{x\to\infty}\left(1+\frac{a}{x}\right)^x = e^a$$

$$\frac{1}{n} = \frac{a}{x} \qquad x = na$$

$$\lim_{x\to\infty}\left(1+\frac{1}{n}\right)^{na}$$

$$= e^a$$

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

$$= \lim_{n\to\infty}\frac{n\cdot(n-1)(n-2)\cdots(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}$$

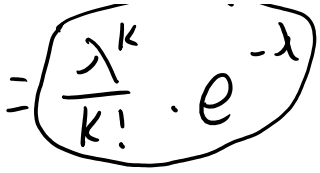
$$= \lim_{n\to\infty}\frac{n^k+\cdots}{n^k}\cdot\left(\frac{\lambda^k}{k!}\right)\cdot\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

this is when we multiple it out (n^k is the highest dimension)

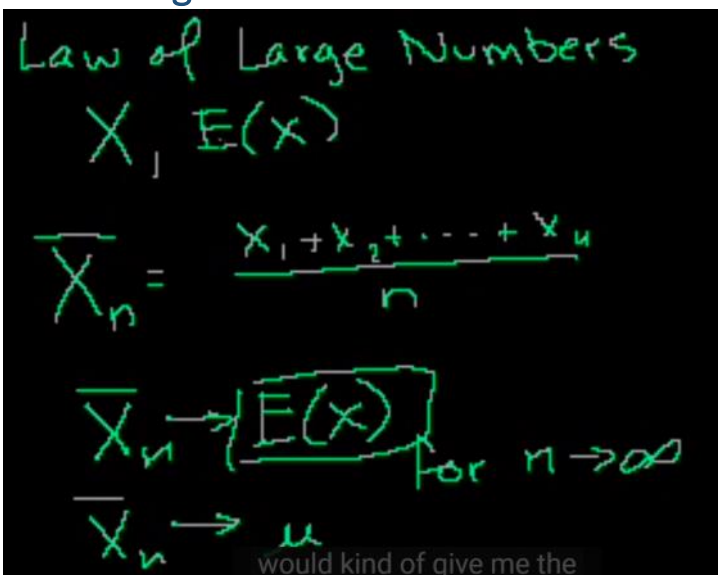
in this case, $a = -\lambda$

$$= 1\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1$$

$$= \frac{\lambda^k}{k!}e^{-\lambda}$$



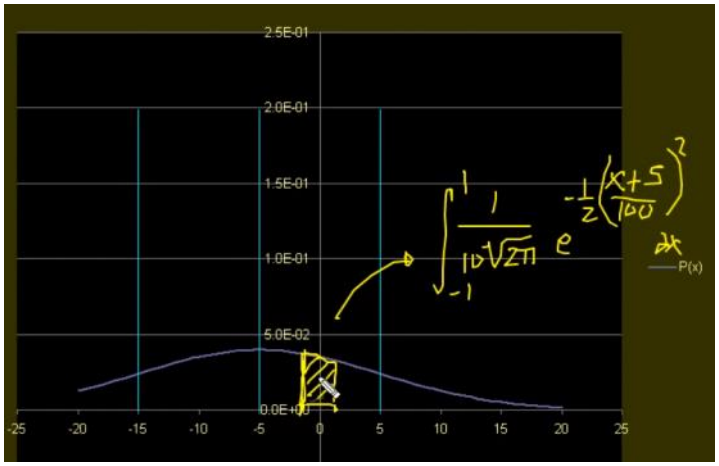
Law of large number:



When n is reaching infinity, the sample mean is equal to E(x), the population mean

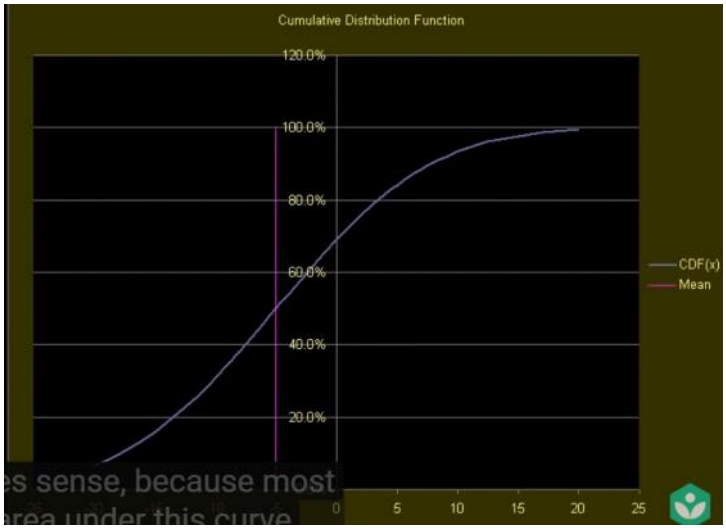
Normal distribution/Gaussian distribution :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Cumulative distribution function(AUC

of normal
distribution)



$$CDF(x) = \int_{-1}^{1} p(x) \cdot dx$$