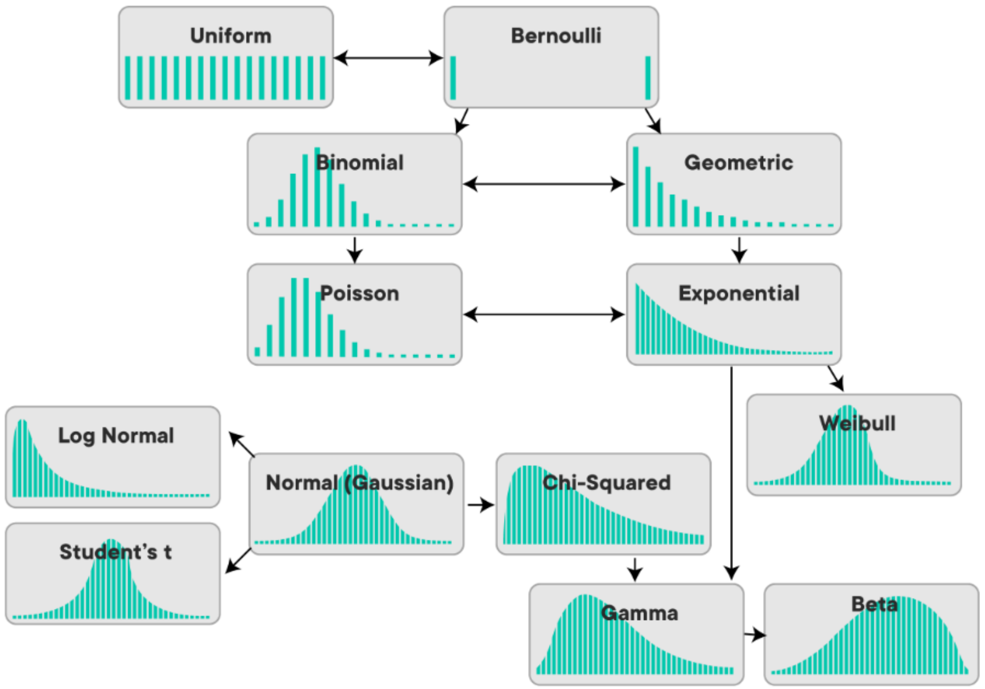
STATISTICAL DISTRIBUTION

~ABHISHEK KUMAR



DISTRIBUTION

- A statistical distribution is a representation of the frequencies of potential events or the percentage of time each event occurs.

TYPES

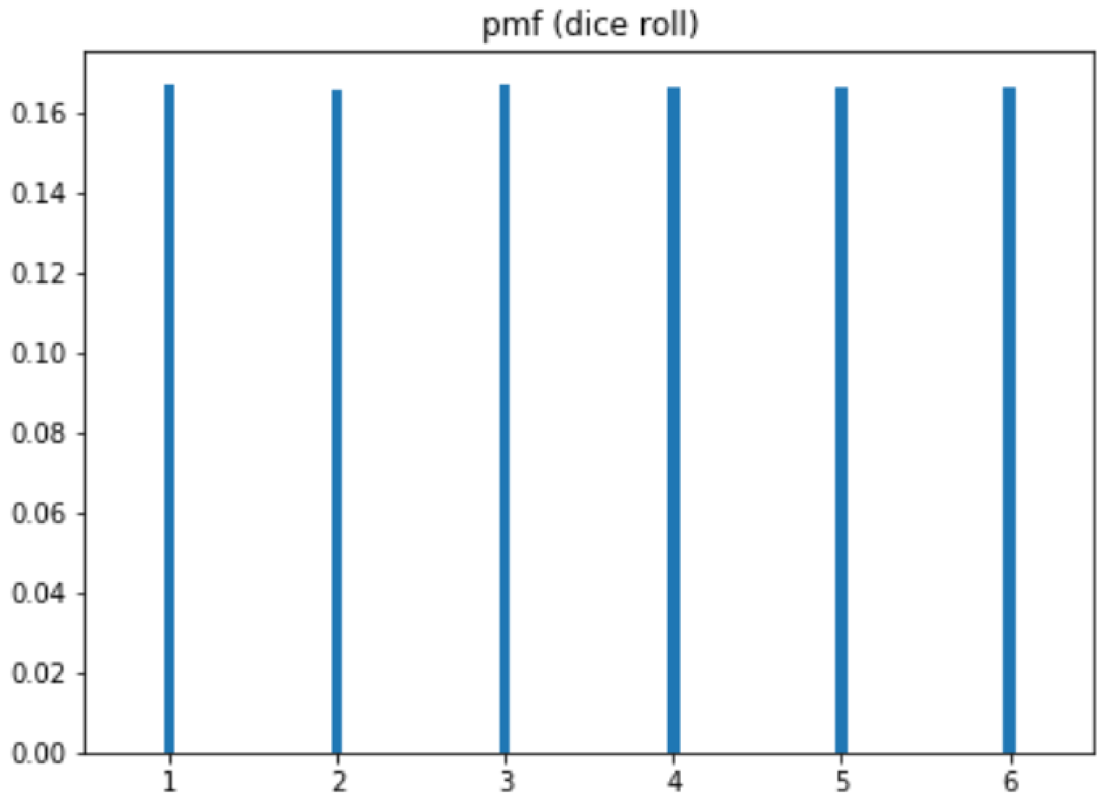
- Discrete distribution
- Continuous distribution

DISCRETE DISTRIBUTION

- Rolling a dice

outcome	1	2	3	4	5	6
probability	1/6	1/6	1/6	1/6	1/6	1/6

- Known number of possible outcomes.
- Probability Mass Function (PMF)

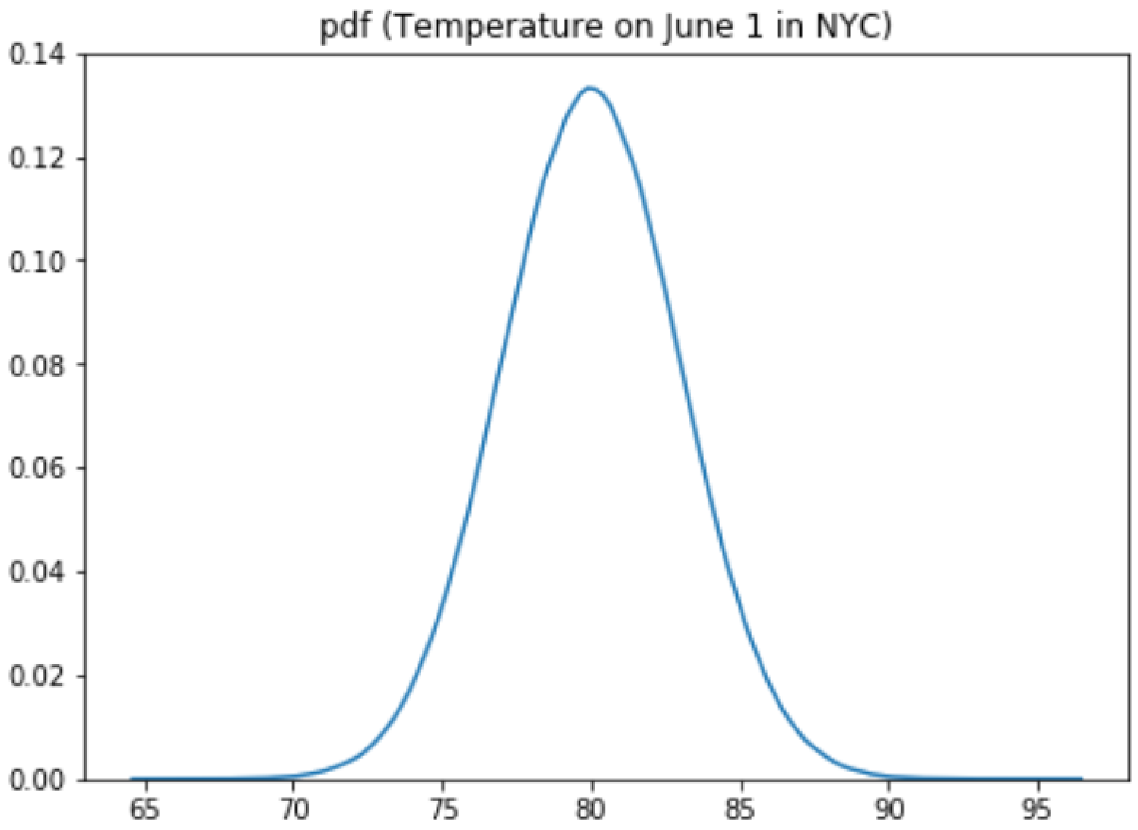


CONTINUOUS DISTRIBUTION

- Temp in New York on Jun 1st

- Probability Density Function (PDF)

this is what we use to find it



HOW TO DESCRIBE IT?

- Expected value or mean
- Variance

EXAMPLES OF DISCRETE DISTRIBUTIONS

- The Bernoulli Distribution: - represents the probability of success for a certain experiment (binary outcome).
- The Poisson Distribution:- represents the probability of n events in a given time period when the overall rate of occurrence is constant.
- The Uniform Distribution:- occurs when all possible outcomes are equally likely.

EXAMPLES OF CONTINUOUS DISTRIBUTIONS

gaussian exists in
nature

- The Normal or Gaussian distribution.

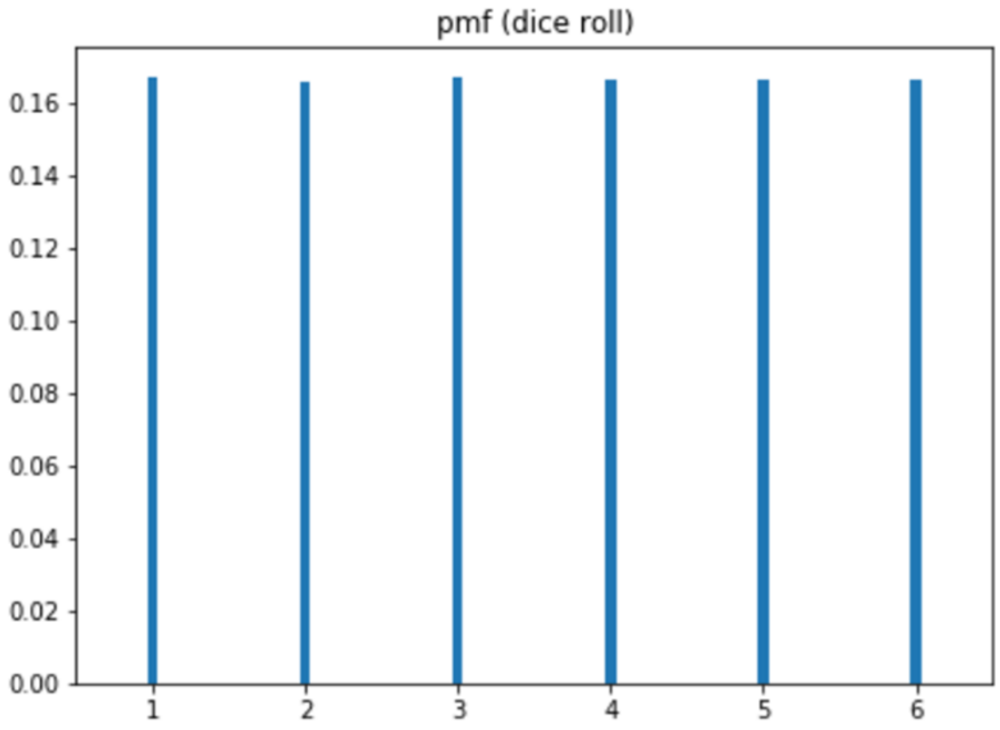
PROBABILITY MASS FUNCTION (PMF)

- Frequency function or **probability distribution**
- Associate probabilities with discrete random variables

$f(x) = P(X = x)$

$R_x = \{x_1, x_2, x_3, \ldots\}$

where $x_1, x_2, x_3, \ldots$ are the possible values of x



PROBABILITY MASS FUNCTION (PMF)

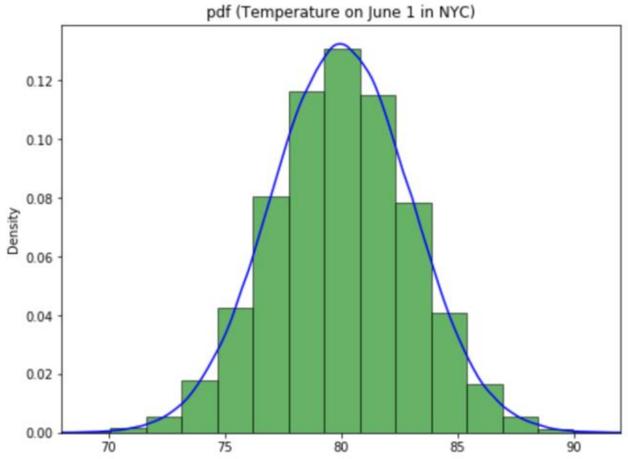
- To convert any random variable's frequency into a probability, we need to perform the following steps:
  - Get the frequency of every possible value in the dataset
  - Divide the frequency of each value by the total number of values (length of dataset)
  - Get the probability for each value

PROBABILITY MASS FUNCTION (PMF)

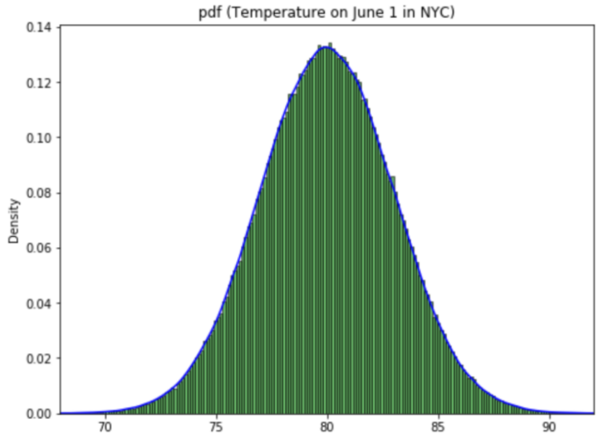
$$E(X) = \mu = \sum_i p(x_i)x_i$$

$$E((X-\mu)^2) = \sigma^2 = \sum_i p(x_i)(x_i - \mu)^2$$

PROBABILITY DENSITY FUNCTION (PDF)



- Continuous variables can take on any real value.
- Helps identify the regions in the distribution where observations are more likely to occur



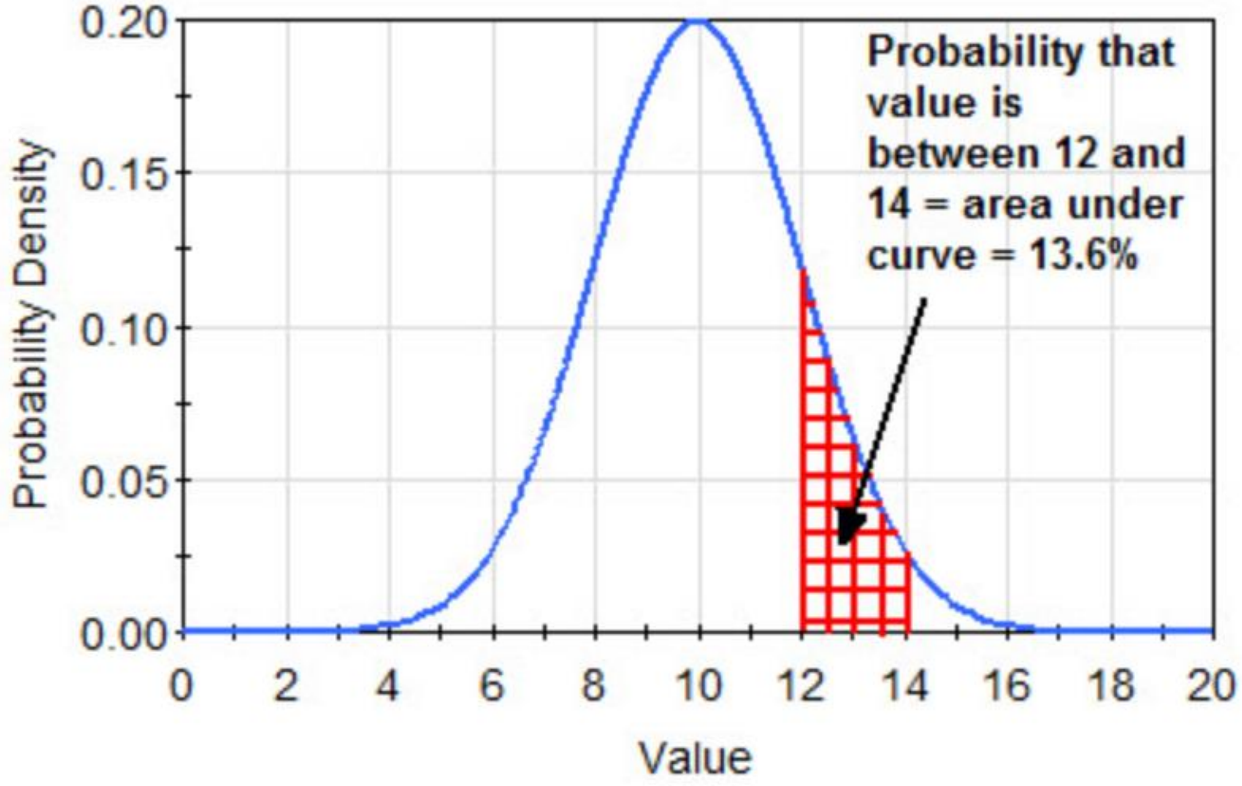
PROBABILITY DENSITY FUNCTION (PDF)

$$E(X) = \mu = \int_{-\infty}^{+\infty} p(x) x dx$$

$$E((X-\mu)^2) = \sigma^2 = \int_{-\infty}^{+\infty} p(x)(x-\mu)^2 dx$$

To obtain exact number, you would get a 1-dimensional line down which isn't really an "area".
For this reason, $P(X\text{=n})\text{=}0$

HOW TO INTERPRET IT?

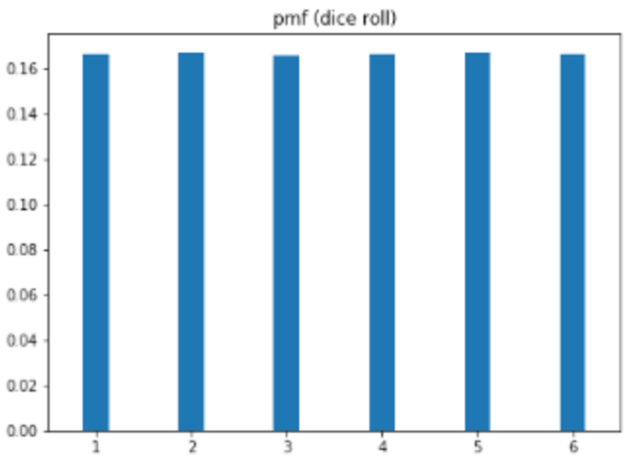


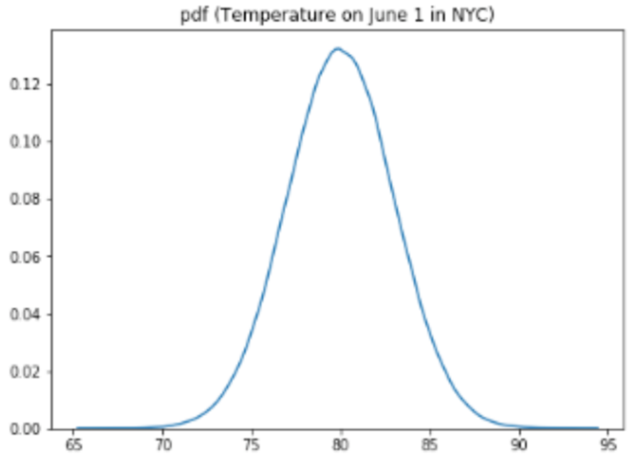
CUMULATIVE DISTRIBUTION FUNCTION

n=we cannot have an absolute value here

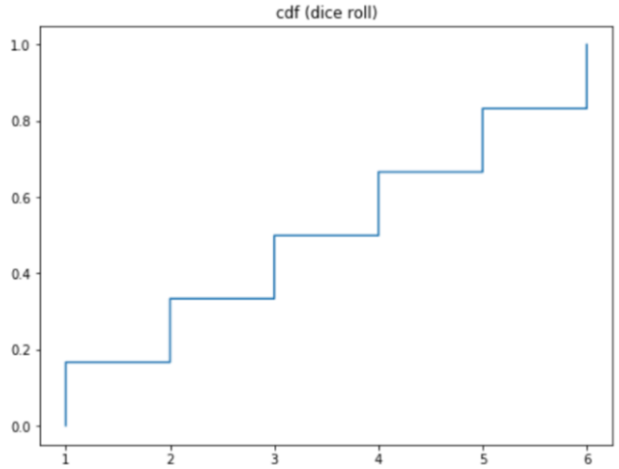
- **Percentile probability function**
- For continuous random variables, obtaining probabilities for observing a specific outcome is not possible
- Have to be careful with interpretation in PDF

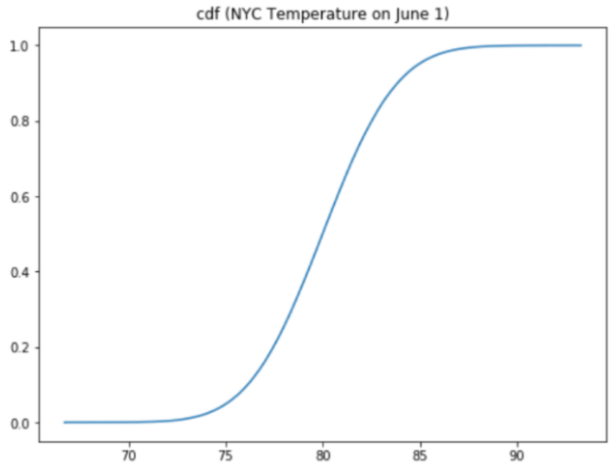
CUMULATIVE DISTRIBUTION FUNCTION





Step functions for discrete random variables





Smooth curves for continuous random variables

CUMULATIVE DISTRIBUTION FUNCTION

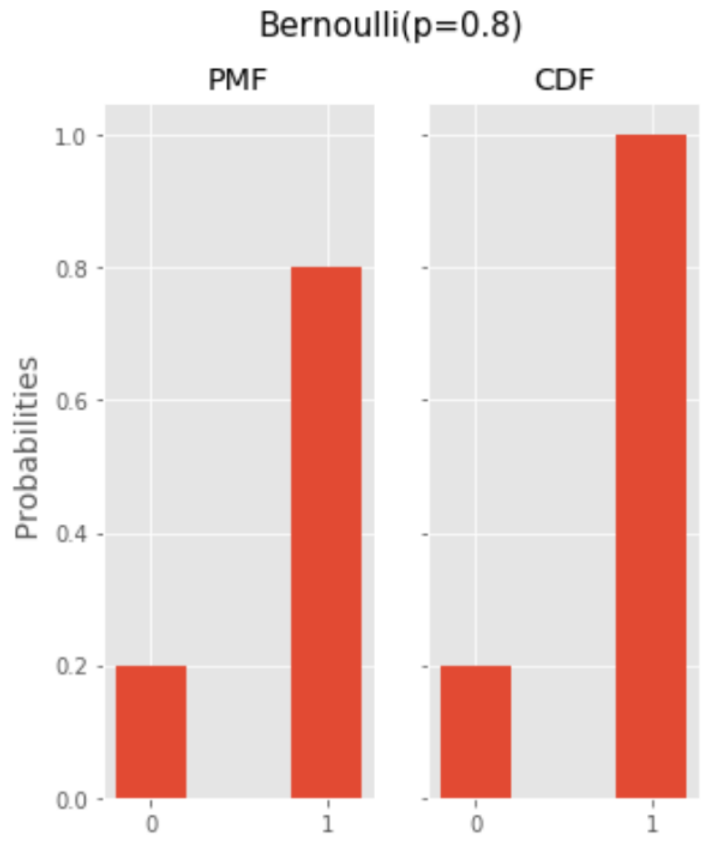
- What is the probability that you throw a value $\leq$ 4 when throwing a dice?
- What is the probability that the temperature in NYC is $\leq$ 79?

BERNOULLI OR BINARY DISTRIBUTION

- A simple experiment in which there is a binary outcome:

0-1, success-failure, heads-tails, etc.

$$E(X) = p \text{ and } \sigma^2 = p * (1 - p).$$



n *independent* Bernoulli trials

BINOMIAL DISTRIBUTION

- If we repeat this process multiple times
- n independent Bernoulli trials

Eg:

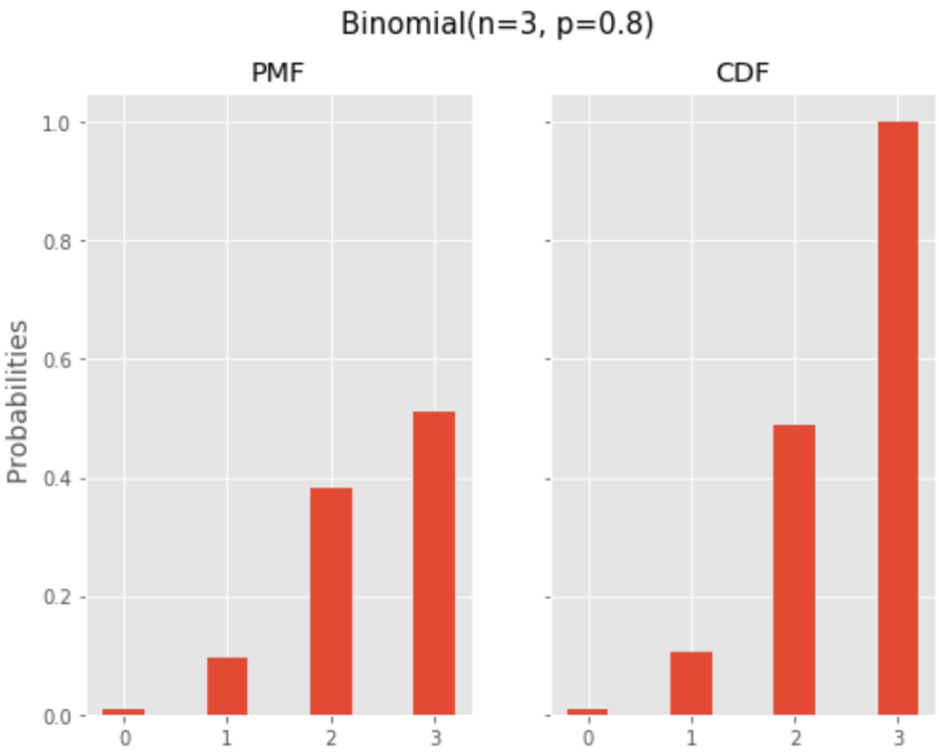
$P(Y=0)$ (or the soccer player doesn't score a single time)?

$P(Y=1)$ (or the soccer player scores exactly once)?

$P(Y=2)$ (or the soccer player scores exactly twice)?

$P(Y=3)$ (or the soccer player scores exactly three times)?

BINOMIAL DISTRIBUTION



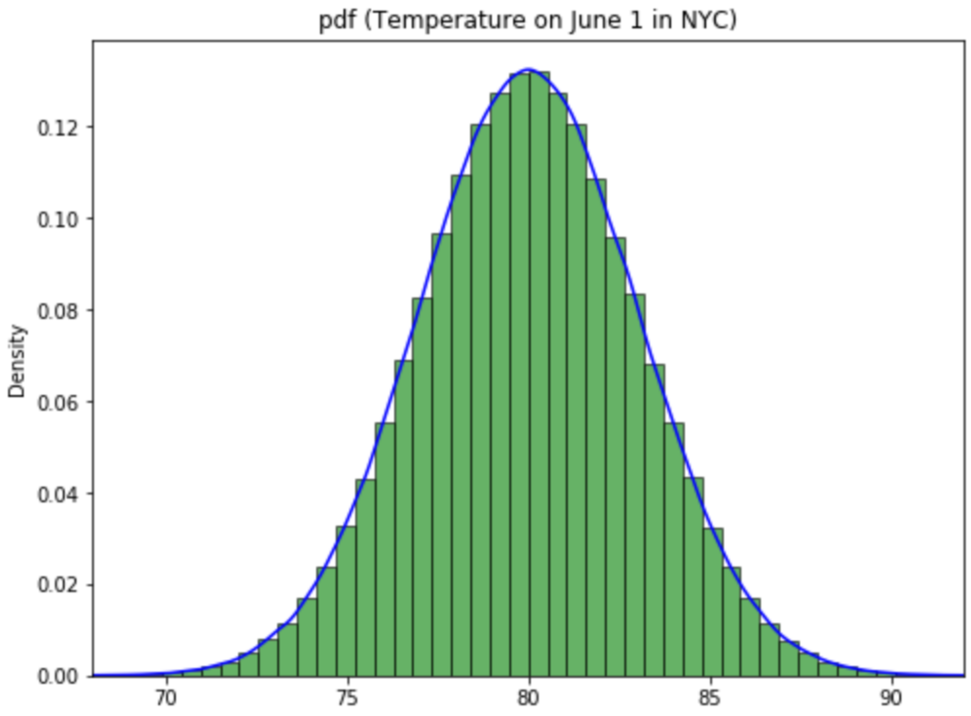
NORMAL DISTRIBUTION

- Most important and most widely used

here mean, median, and mode are all similar.
we can calculate this distribution by variance which is dev ^2

- "Gaussian curve" after the German mathematician

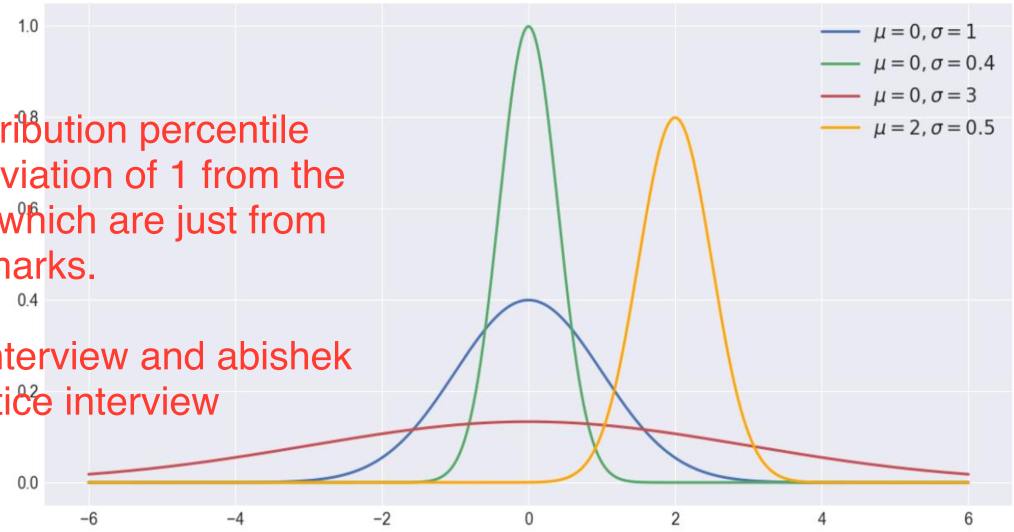
Karl Friedrich Gauss.

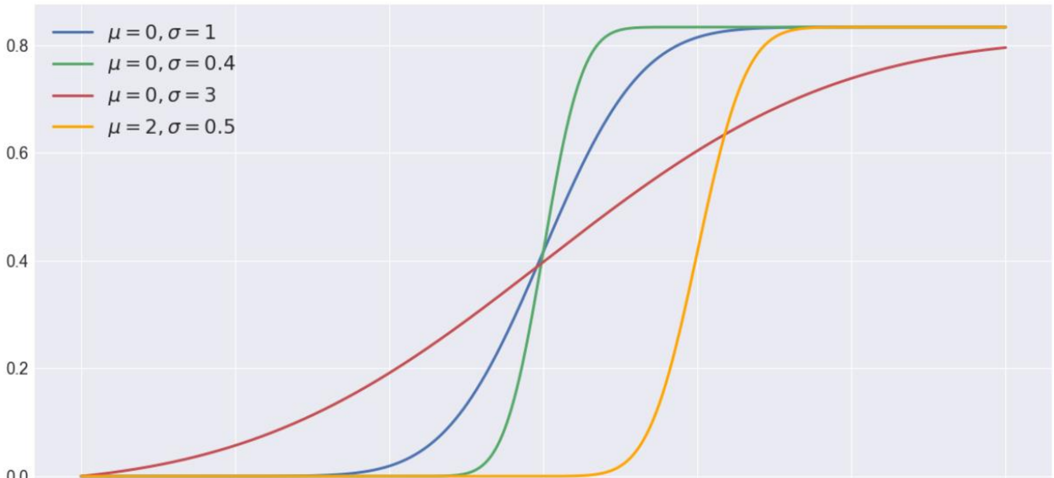


NORMAL DISTRIBUTION

68, 95, 99.7 are normal distribution percentile values, based on standard deviation of 1 from the center vs 50/75th percentile which are just from those quartile marks.

remeber this, this matters in interview and abishek will as us this in proactice interview





NORMAL DISTRIBUTION

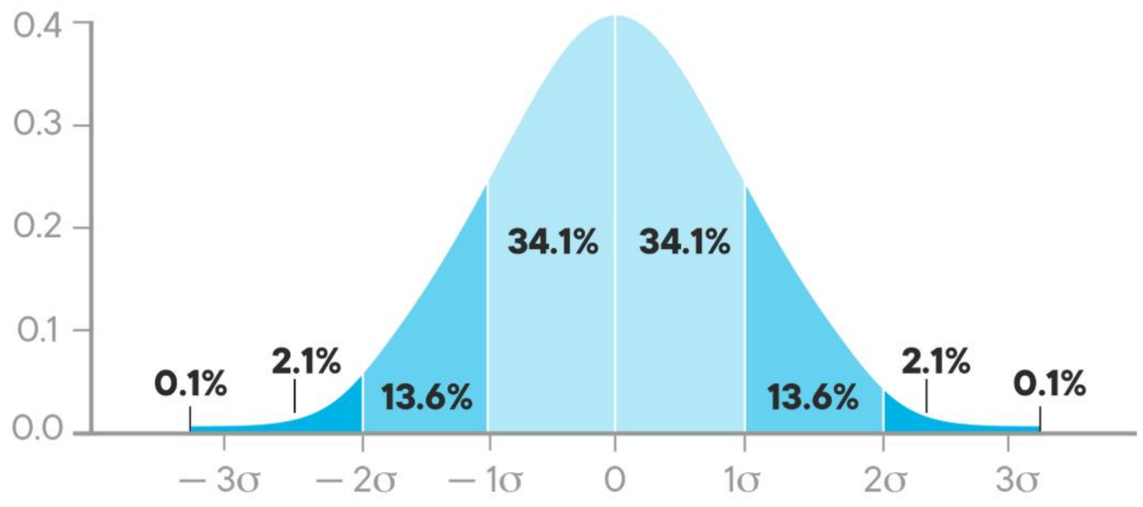
- Central Limit Theorem:

When you add a large number of independent random variables, irrespective of the original distribution of these variables, their sum tends towards a normal distribution.

STANDARD NORMAL DISTRIBUTION

in standard normal dist the mean is always 0 and the deviation is always 1

- mean of 0 and a standard deviation of 1.



look up Z score and Z test and theorems mentioned here, make flashcards for formulas on deviation, variance, etc.

DISCUSSION

THANK YOU!!