Probability and Probability distributions

Day 2

Do The Math

Chicago, IL
Bangalore, India
www.mu-sigma.com

2016

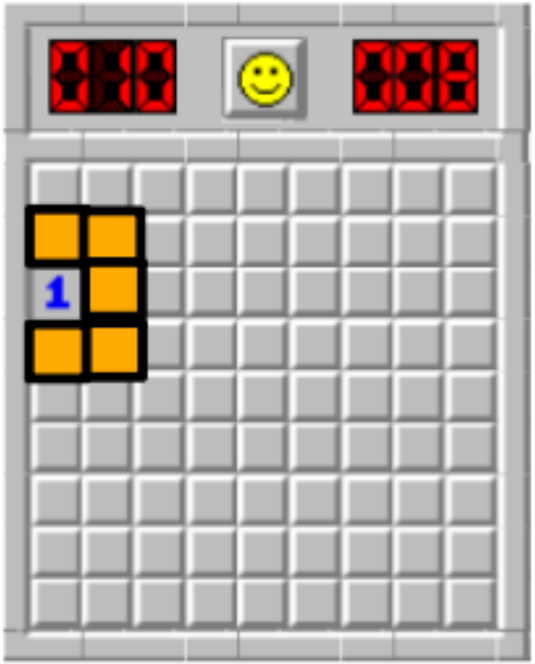
Topics

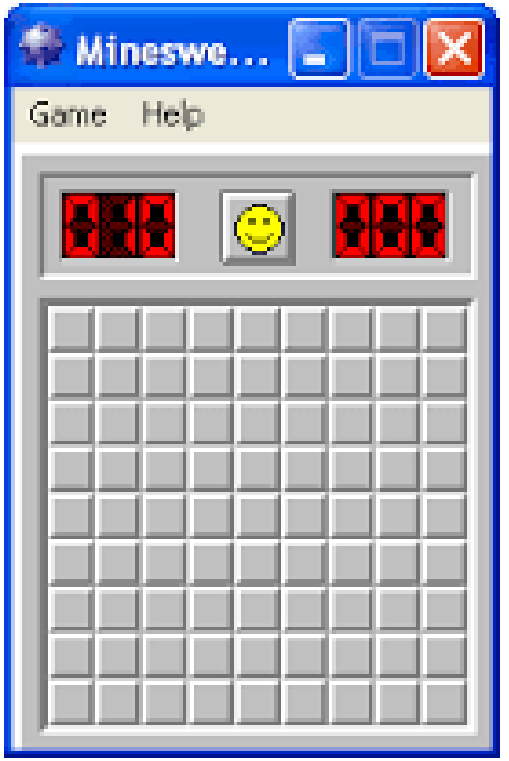
- What is probability?
- An example – Minesweeper
- Rules of probability and Bayes theorem
- An example – Bayes theorem demo
- Frequentist vs. Bayesian probability
- Random variables and probability distributions
- Normal and binomial distributions with examples
- Population and sample
- Sampling distributions and CLT
- An exercise – CLT
- Demo of CLT [Why Normal distribution is so important]

What is probability

Let’s play Minesweeper

- What is the probability of not hitting a mine in the first click?





- Should you choose an orange or non orange cell?

Independence and mutual exclusivity

- What is the probability of hitting a mine and not hitting a mine?

These are mutually exclusive events. They both cannot happen simultaneously.

$$\mathrm{P}(A \cap B) = \ 0$$

- What is the probability of hitting a mine and choosing square A1?

These are independent events. One does not affect the other.

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

Some additional laws of probability

- What is the probability of opening cell A1?
  - For any event A, P(A) ≥ 0.
- What is the probability of opening *atleast one* cell?
  - P(S) = 1.

Conditional probability

- Let’s go back to our minesweeper game

- What is the probability that I will not hit a mine given that I hit square A1 the first time?

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Suppose a large retailer has the following frequencies of people who have bought different combinations of products in the same basket

	Milk	Groceries	Toiletries	**Totals**
Milk	50	33	21	104
Groceries	33	66	45	144
Toiletries	21	45	25	91
Totals	104	144	91	339

*Diagonal represents products were bought alone

Which products are suitable for cross-selling and promotion?

Bayes' Theorem

- Let's go back to our minesweeper game

- What is the probability of having selected square A1 given that you did not hit a mine?

$$P(A_1 \mid B) = \frac{P(A_1)\,P(B \mid A_1)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)}$$

Example – Part 1

- A large online retailer [say Amazon] just rolled out a digital ad campaign on their website. For the people who click on the ad, Amazon wants to customize the post-click engagement
- Amazon has High and Low Value customers (20% vs. 80%)
- Amazon has to choose between investment in 2 engagements strategies [They can pick only one]
  - Customized [Call] - This is for High value customers
  - Generic [Email] - This is for low value customers
- Amazon has to decide which strategy to invest in based on whether a customer who clicked is more likely to be a High value vs. Low value customer
- Click rates are:
  - High value customers – 40%
  - Low value customers – 20%

Bayes theorem example – Part 2

Type of customer	Population proportion	Clicked	Didn't click
High value	20%	40%	60%
Low value	80%	20%	80%

$$P(\text{High} \mid \text{Clicked}) = \frac{P(\text{Clicked} \mid \text{High}) * P(\text{High})}{P(\text{Clicked} \mid \text{High}) * P(\text{High}) + P(\text{Clicked} \mid \text{Low}) * P(\text{Low})}$$

$$= \frac{0.08}{0.24} = 0.33$$

Similarly, P (Low | Clicked) = 0.66

Theoretical probability

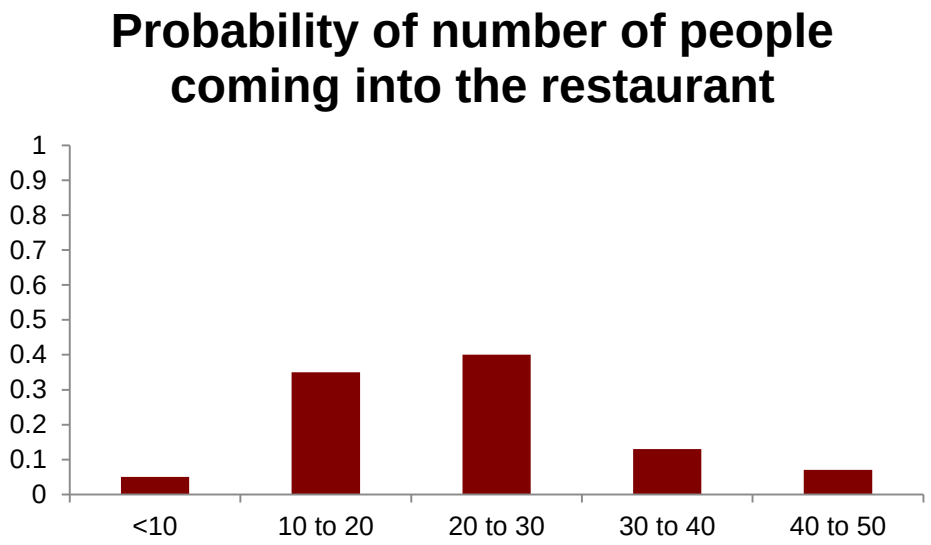
- **Theoretical Probability** of an event is the number of ways that the event can occur, divided by the total number of outcomes. It is finding the **probability** of events that come from a sample space of known equally likely outcomes.

Frequentist probability

- Experimental probability of an event is the ratio of the number of times the event occurs to the total number of trials

- This is called the frequentist approach to probability

Random variables and Probability distributions

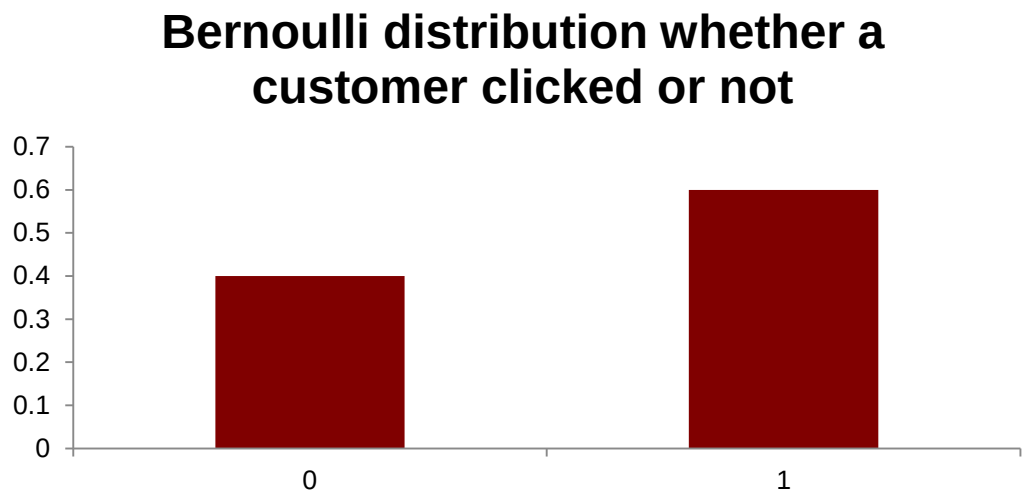


Number of people	Random variable - X
<10	1
10 to 20	2
20 to 30	3
30 to 40	4
40 to 50	5

Probability distribution is a function which captures the probabilities of occurrence of all the events of an experiment

Random variables are mathematical constructs which help map events to a real number

Discrete distributions - Bernoulli



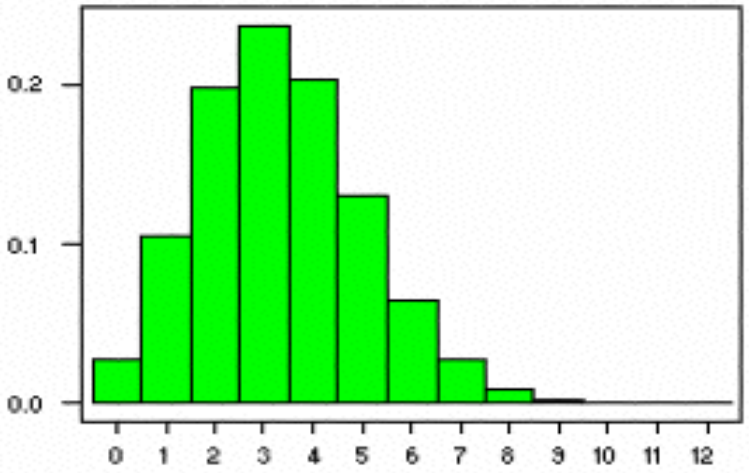
- Bernoulli distribution has two possible outcomes, success and failure
- The distribution is defined by the probability of success (p)
- Probability distribution for the outcome of a toss of a coin

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0. \end{cases}$$

Binomial distribution

D:\Training\MSU
)\binomialExample

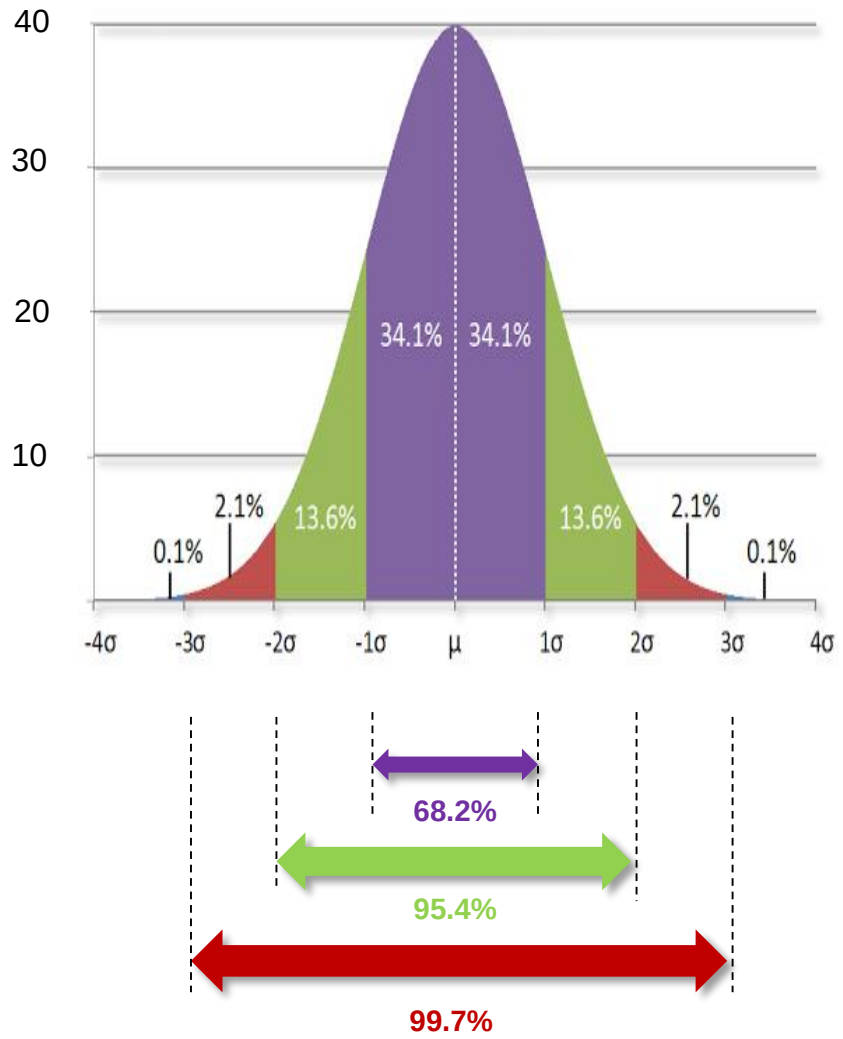
Bin(20, 1/6)



- Binomial distribution is arrived at by performing ‘n’ successive Bernoulli trials, probability of success in each trial is given by ‘p’

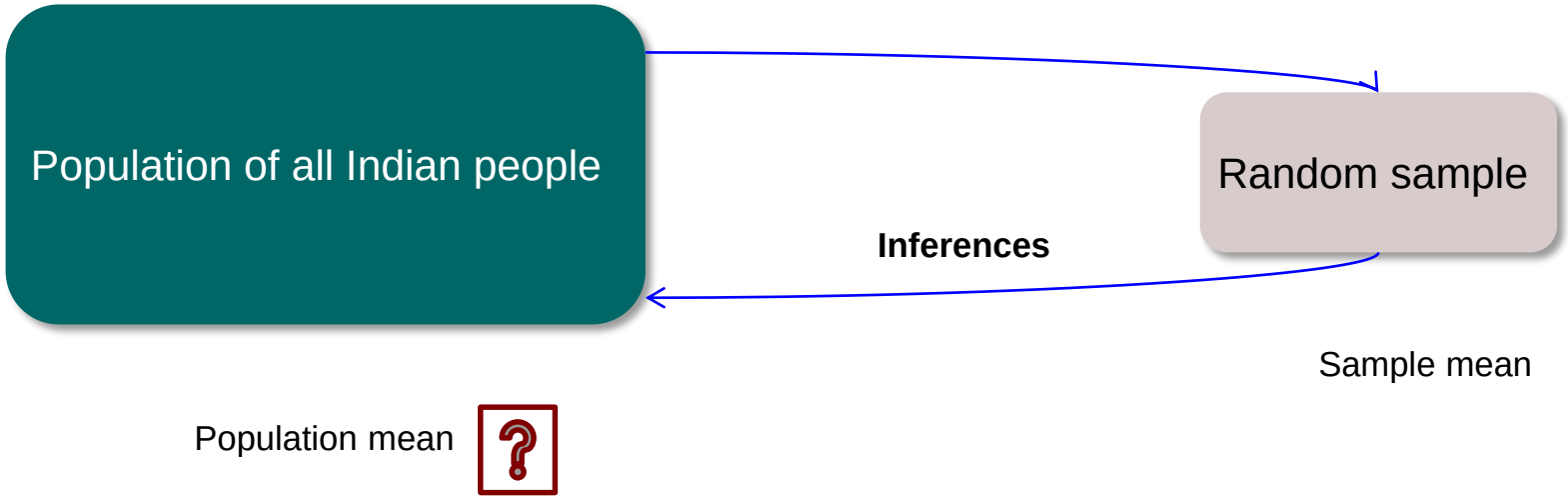
Continuous distributions – Normal distribution

- Normal distribution is a continuous frequency distribution which is symmetric around the mean
- Normal distribution is completely described by **mean μ** and **variance, σ^2**
- μ is the center and σ determines the spread (the larger the σ the more spread the curve is)



Why sampling?

What is the average age of India?



Randomization

- What is randomization?
  - The process of selection from the population such that every possible sample has an equal chance of being selected as the sample
  - This translates to equal chance of every unit of the population having an equal chance of being in the sample
  - It is the process of generating itself that is random, not the sample itself

- What is a ‘representative’ sample?
  - There is no such thing! It is a misnomer
  - Though intuitively appealing, a sample can **never** be truly representative of the population
  - It can be representative on one dimension, but there are so many unknown dimensions on which it is not representative. How do we know which dimensions? How do we decide?

An exercise

Height
168
165
171
173
166

Population mean = 168.6

Mean of sample means= 168.463

{168, 165, 171} **168**	{168, 171, 173} **170.66**	{168, 173, 166} **169**	{165, 171, 173} **169.66**
{168, 165, 173} **168.66**	{168, 171, 166} **168.33**	{171, 173, 166} **170**	{165, 171, 166} **167.33**
	{173, 161, 166} **166.66**	{168, 165, 166} **166.33**	

The mean of the sample means of all samples of a particular size is close to the population mean

Central Limit Theorem Demo

- [Link below] (OR)
- R script

D:\Training\MSU
\Day2-Probability a

Appendix

Bayes theorem in action – An example

D:\Training\MSU
BayesTheoremInAc