Week 20 Combinatorics and Probabilities continued Lecture Note

Notebook:	Computational Mathematics
Created:	2020-04-21 2:48 PM
Updated:	2020-07-31 4:06 PM
Author:	SUKHJIT MANN
URL:	https://www.coursera.org/learn/uol-cm1015-computational-mathematics/home/week/13

Cornell Notes	**Topic:** **Combinatorics and Probabilities continued**	Course: BSc Computer Science Class: Computational Mathematics[Lecture] Date: July 31, 2020

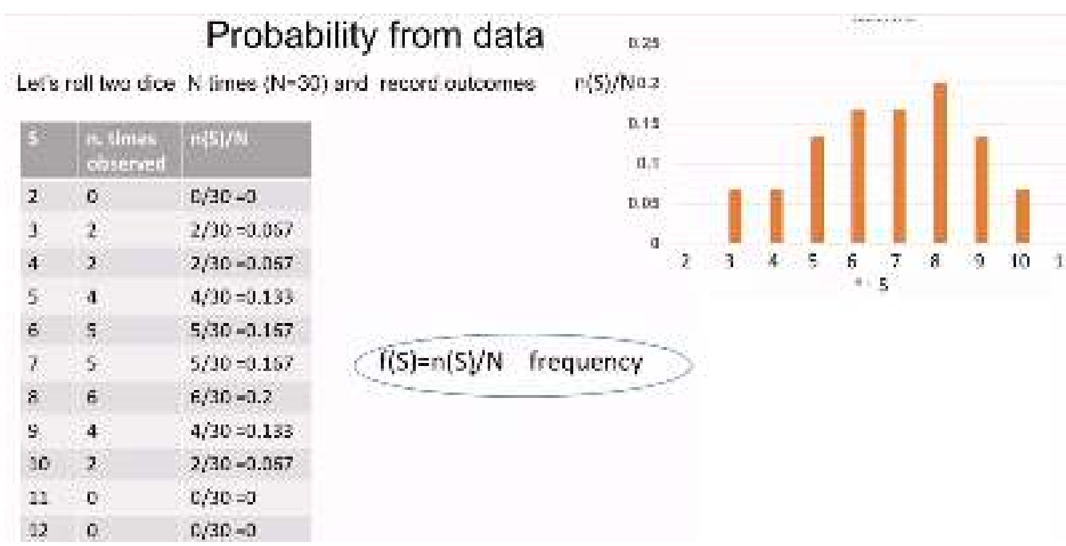
Essential Question:

What is the probability of an event and how can we use the principles of counting to evaluate such probabilities?

Questions/Cues:

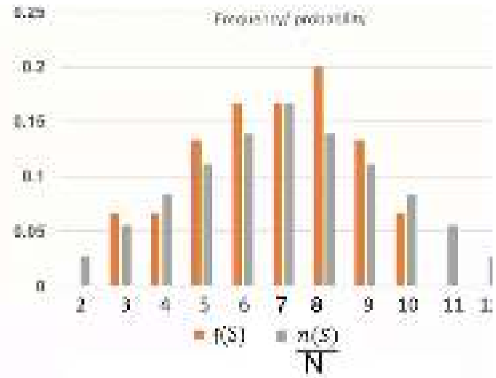
- How can we calculate the probability of an event from data and what is frequency?
- What is the experimental definition of probability?
- What is the mean, median and mode?
- What is variance and standard deviation?
- What is normal distribution?
- What is Chebyshev's theorem?

Notes



Probability data vs theory

S	n. times observed	n(S)/N	P(S)
2	0	0/30	1/36
3	2	2/30	2/36
4	3	2/30	3/36
5	4	4/30	4/36
6	5	5/30	5/36
7	5	5/30	6/36
8	6	6/30	5/36
9	4	4/30	4/36
10	2	2/30	3/36
11	0	0/30	2/36
12	0	0/30	1/36



Experimental def of probability:

$P(S) = \lim_{N\to\infty} f(S) = \lim_{N\to\infty} \frac{n(S)}{N}$

Statistical analysis of data

Let's say we roll two dice N times and we record outcomes

$S_1\ S_2\ S_3\ \ldots\ S_N$

Mean

$m = (S_1 + S_2 + S_3 \ldots S_N)/N = \frac{1}{N}\sum_{i=1}^{N} S_i$

→if only M different results $m = \sum_{i=1}^{M} S_i\, n(S_i)/N$

Median

S_m that separates your ordered data set in two halves

$S_1\ S_2\ S_3 \ldots S_m \ldots S_{N-2} S_{N-1} S_N$

If N odd median=$S_{(N+1)/2}$
If N even median=$(S_{N/2}+S_{N/2+1})/2$

Mode or most probable value
Outcome of maximal probability, that happens most frequently

Mean,median and mode

Let's roll 2 dice N times (N=30) and record outcome

S_i	n. times observed	n(S)/N
2	0	0/30
3	2	2/30
4	2	2/30
5	4	4/30
6	5	5/30
7	5	5/30
8	6	6/30
9	4	4/30
10	2	2/30
11	0	0/30
12	0	0/30

$$m=\frac{1}{N}\sum_{i=1}^{N} S_i = \frac{2x0+3x2+4x2+5x4+6x5+7x5+8x6+9x4+10x2+11x0+12x0}{30}=$$

$=203/30=6.8$

Median=7

$$\underbrace{3,3,4,4,5,5,5,5,6,6,6,6,6,7}_{14},7,7,\underbrace{7,7,8,8,8,8,8,8,9,9,9,9}_{14}$$

Mode=8

Variance and standard deviation

Variance

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(S_i-m)^2 \quad \text{also} \quad \sum_{i=1}^{M}(S_i-m)^2\, n(S_i)/N$$

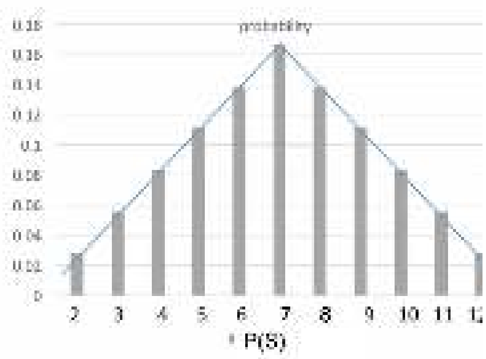
on data $\longrightarrow$ $\frac{1}{N-1}\sum_{i=1}^{N}(S_i-m)^2$

standard deviation

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i-m)^2} \quad \text{on data} \longrightarrow \sqrt{\frac{1}{(N-1)}\sum_{i=1}^{N}(S_i-m)^2}$$

Probability distribution

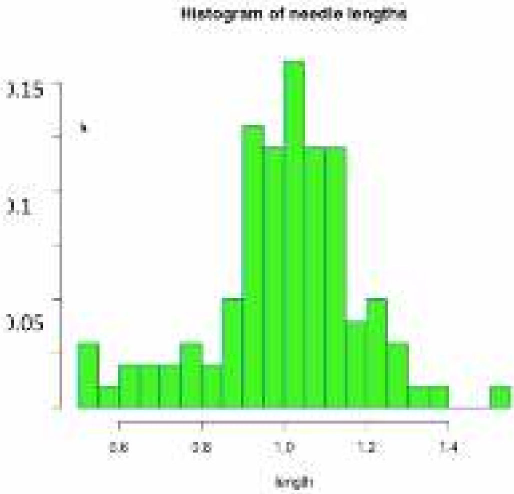
S	P(S)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



$P(S)=(S-1)/36$ for $2<=S<=7$
$P(S)=(13-S)/36$ for $7<S<=12$

Verify $m=\sum_{i=1}^{11} S_i\, P(S_i)=7$; $\sigma^2=\sum_{i=1}^{11}(S_i-m)^2\, P(S_i)=5.8$; mode=7

Continuous variables: Normal Distribution



100 Needles lengths

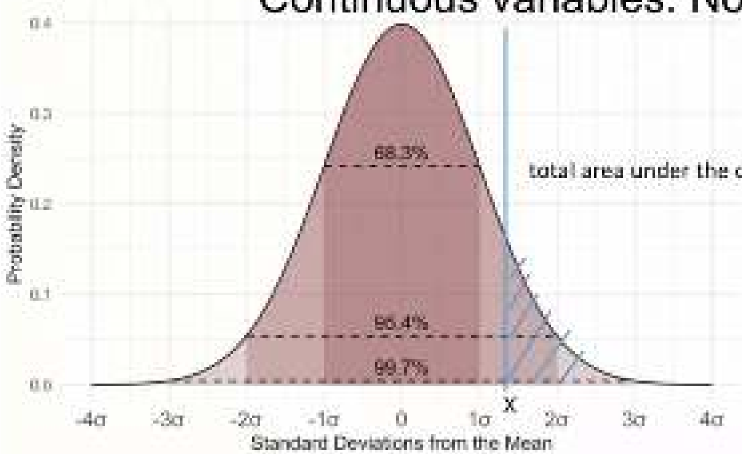
N.	length
1	1.2 cm
2	0.95 cm
3	1.1 cm
4	1.3 cm
5	0.9 cm
6	1.15 cm
7	1.0 cm
...	....
100	0.92 cm

From data: $m=1 \quad \sigma^2=0.2$

dx = width of histograms='bin'

Frequency=(n. of observed values in bin)/100(total measurements)

Continuous variables: Normal Distribution



total area under the curve=1 (100%)

$$p(x)=\frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

- p(X)dx = probability that length falls in interval dx around value X
- p(x>X) = Area marked by blue lines

Chebyshev's theorem

$$p(|x-m|\geq k\sigma) \leq 1/k^2 \quad \text{with } k>0$$

for a wide class of probability distributions k>1

⇒ you can take the mean value m as an estimate of the outcome and the variance σ as an estimate of the uncertainty in the outcome

Summary

In this week, we learned about how to calculate probability from data and what frequency is, what the experimental definition of probability is, what mean, median and mode are, what variance and standard deviation are, what normal distribution is, and what Chebyshev's theorem is.