Probability Distribution

Frequency distribution

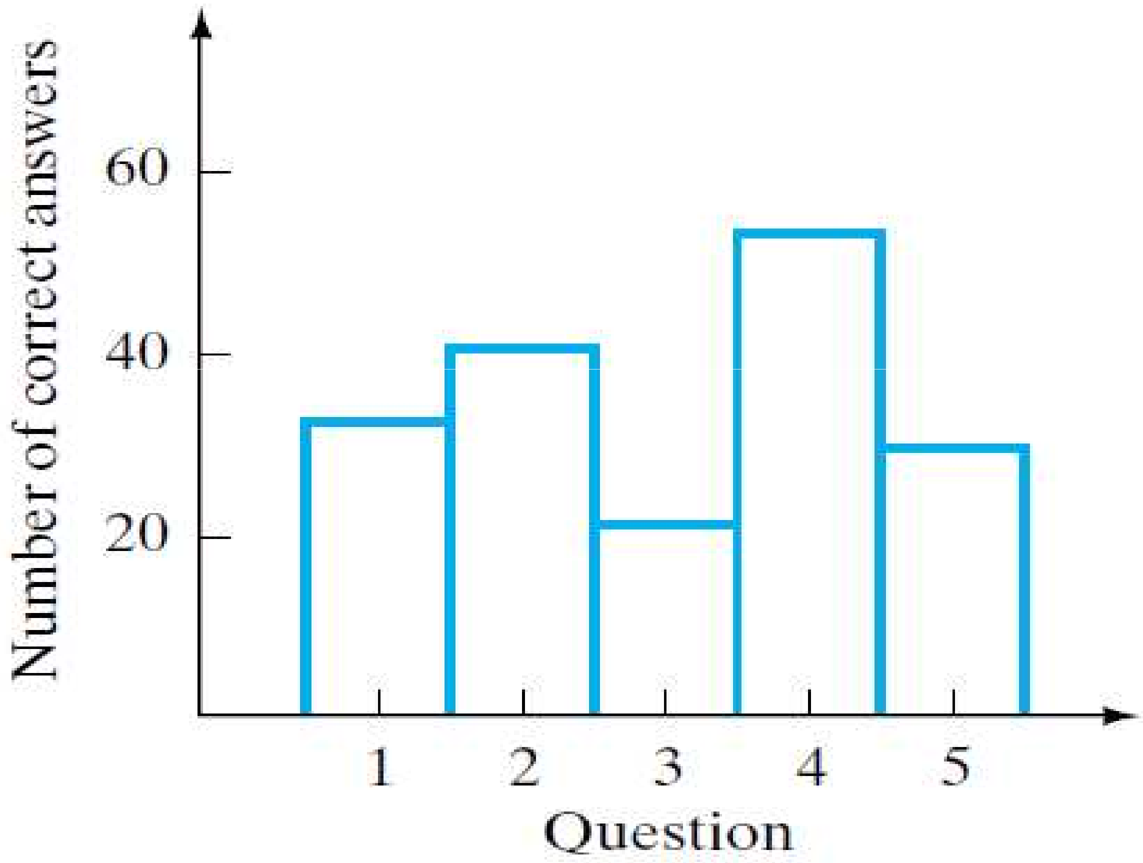
Quantitative data

Score	Number of Students Making Score
0–20	0
21–40	18
41–60	36
61–80	83
81–100	110
101–120	121
121–140	73
141–160	16
Total	457

Qualitative data

Major	Number of Students
Science	429
Arts	132
Languages	41
Social sciences	631
Engineering	344
Total	1577

Histogram – no. of correct answers for 5 questions in an exam

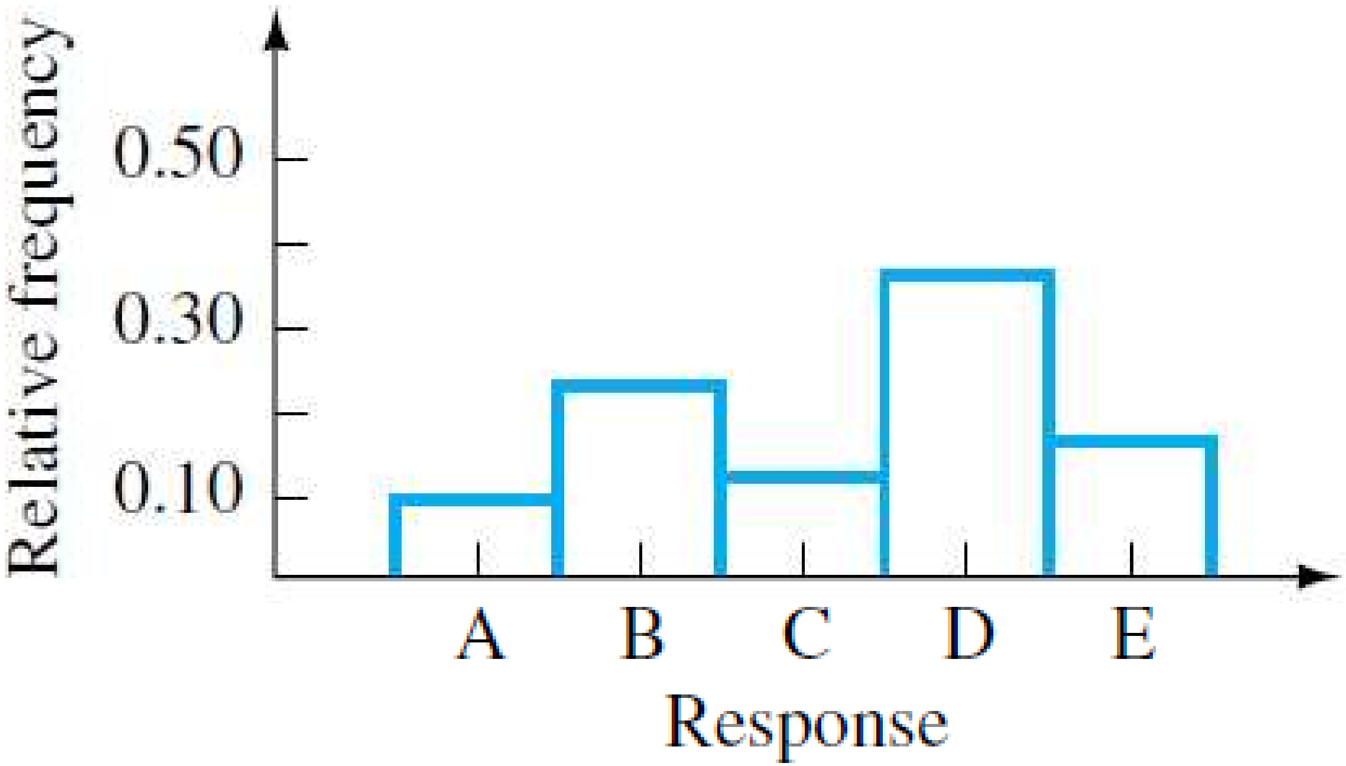


Relative frequency

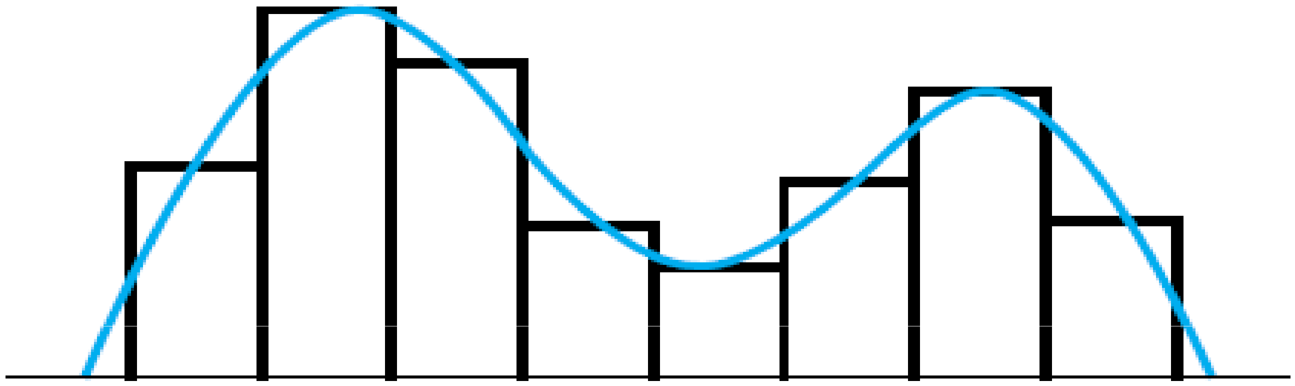
- Relative frequency counts the fractional part of the data that belong to a category

An exam has five possible grades:
A, B, C, D, and E

Response	Frequency	Relative frequency
A	6	6/60=0.1
B	14	14/60=0.23
C	8	8/60=0.13
D	22	22/60=0.37
E	10	10/60=0.17
	Total = 60	Total = 1

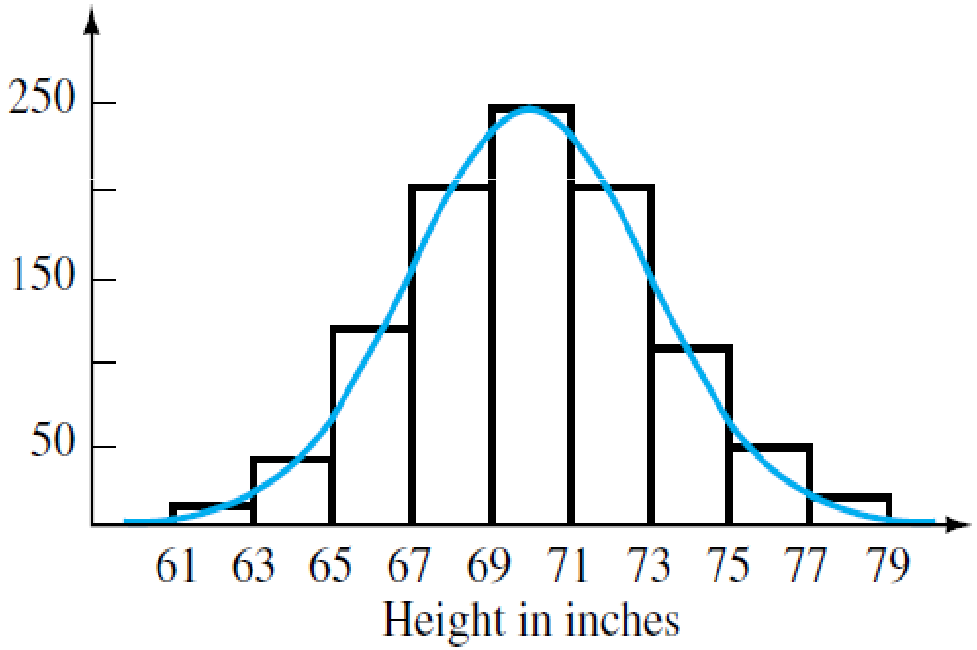

Relative frequency
0.50
0.30
0.10
A
B
C
D
E
Response

When dealing with continuous data…



- A smooth curve conveys the impression of continuous data better than a histogram.
- Sketch a smooth curve based on a histogram <u>by drawing it through the midpoints at the top of the bars</u>

Continuous data – example

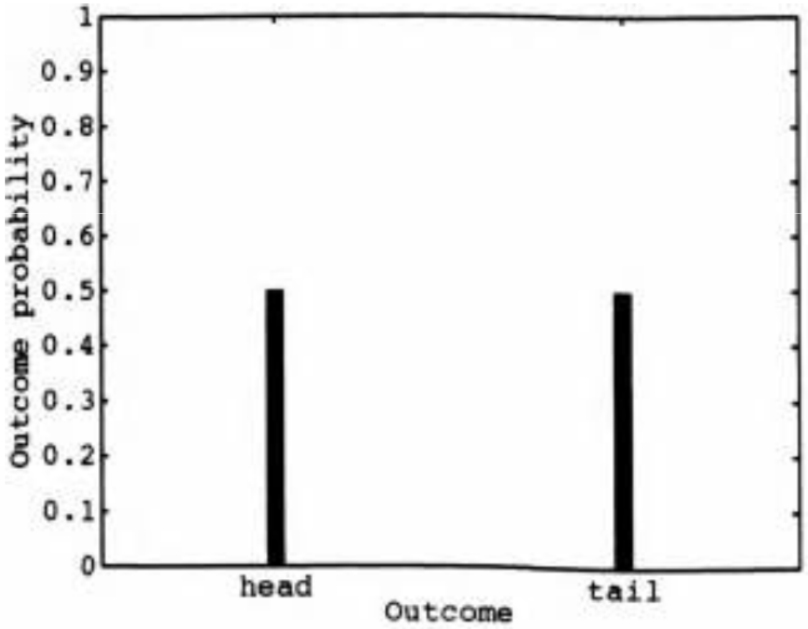


Height (Inches)	Frequency
61–62.9	10
63–64.9	51
65–66.9	115
67–68.9	200
69–70.9	240
71–72.9	195
73–74.9	104
75–76.9	42
77–78.9	15

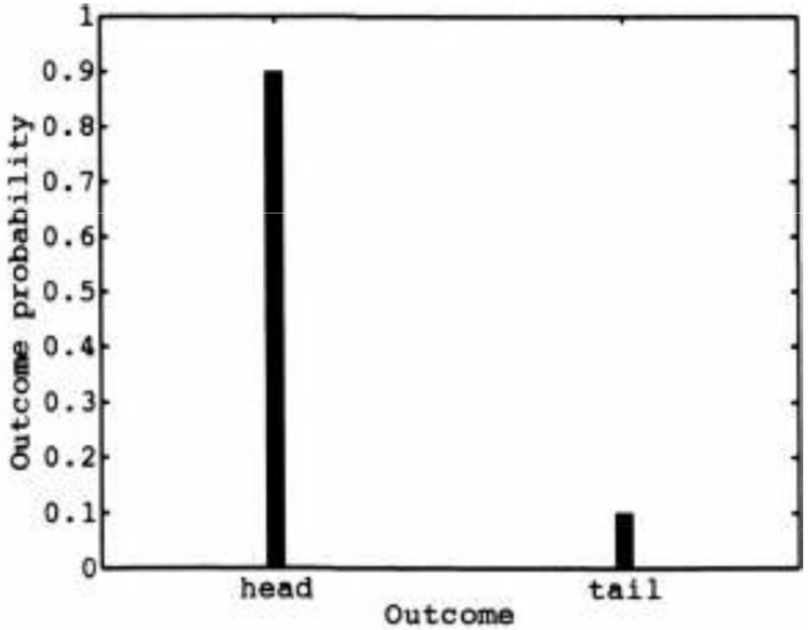
Probability distributions of two coins

p(head)=0.5

p(head)=0.9



Unbiased coin



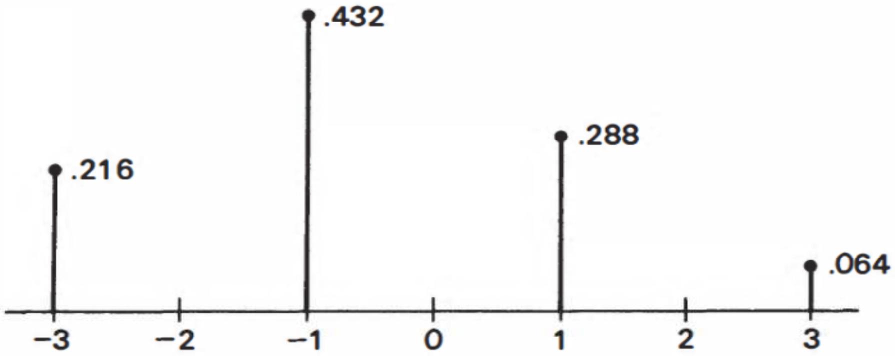
Biased coin

Example - RV

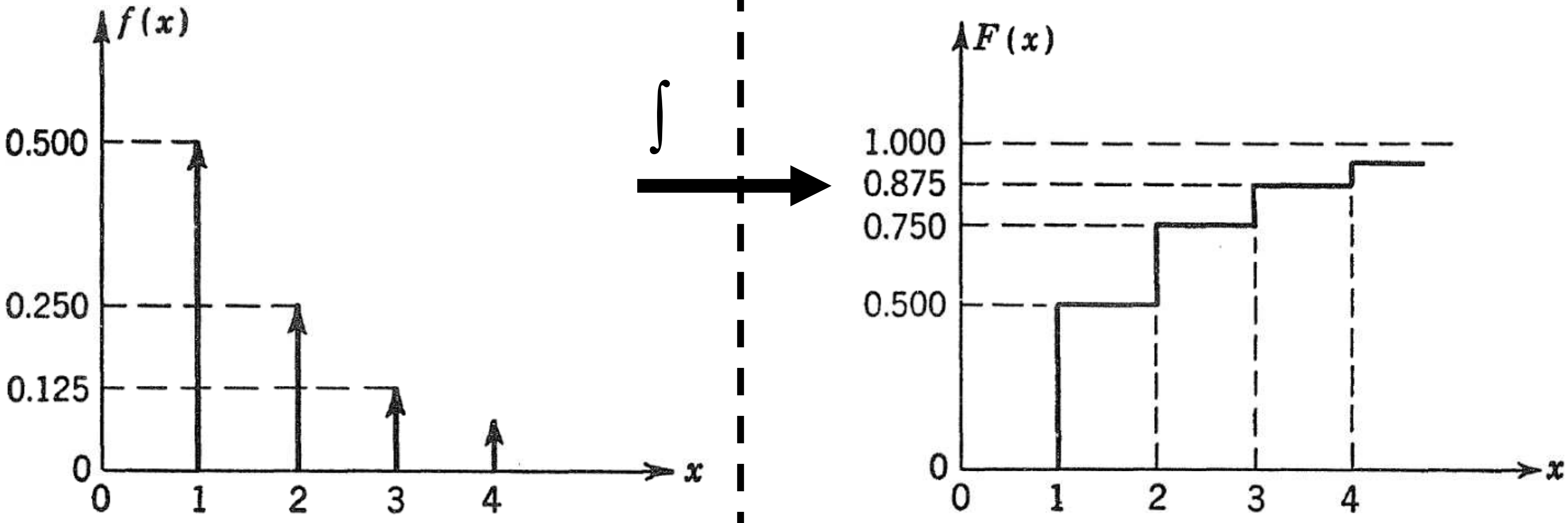
A coin is flipped 3 times
3 heads earn 3 points
2 heads earn 1 point
3 tails earn -3 points
2 tails earn -1 point

ω	$X(\omega)$	$P\{\omega\}$
HHH	3	p^3
HHT	1	$p^2(1-p)$
HTH	1	$p^2(1-p)$
THH	1	$p^2(1-p)$
HTT	-1	$p(1-p)^2$
THT	-1	$p(1-p)^2$
TTH	-1	$p(1-p)^2$
TTT	-3	$(1-p)^3$

If p = 0.4,



Probability distribution function – another look



Example

- 15 cards with 5 colored red, 3 colored black, and 7 colored green
- The cards are shuffled, and 2 cards selected
  - No points if both cards different colors
  - Five points if both cards are green
  - 10 points are if both cards are red
  - 5 points are if both cards are black

Random variable

Outcome	Value of X (points)
Red and black	0
Red and green	0
Red and red	10
Black and black	15
Black and green	0
Green and green	5

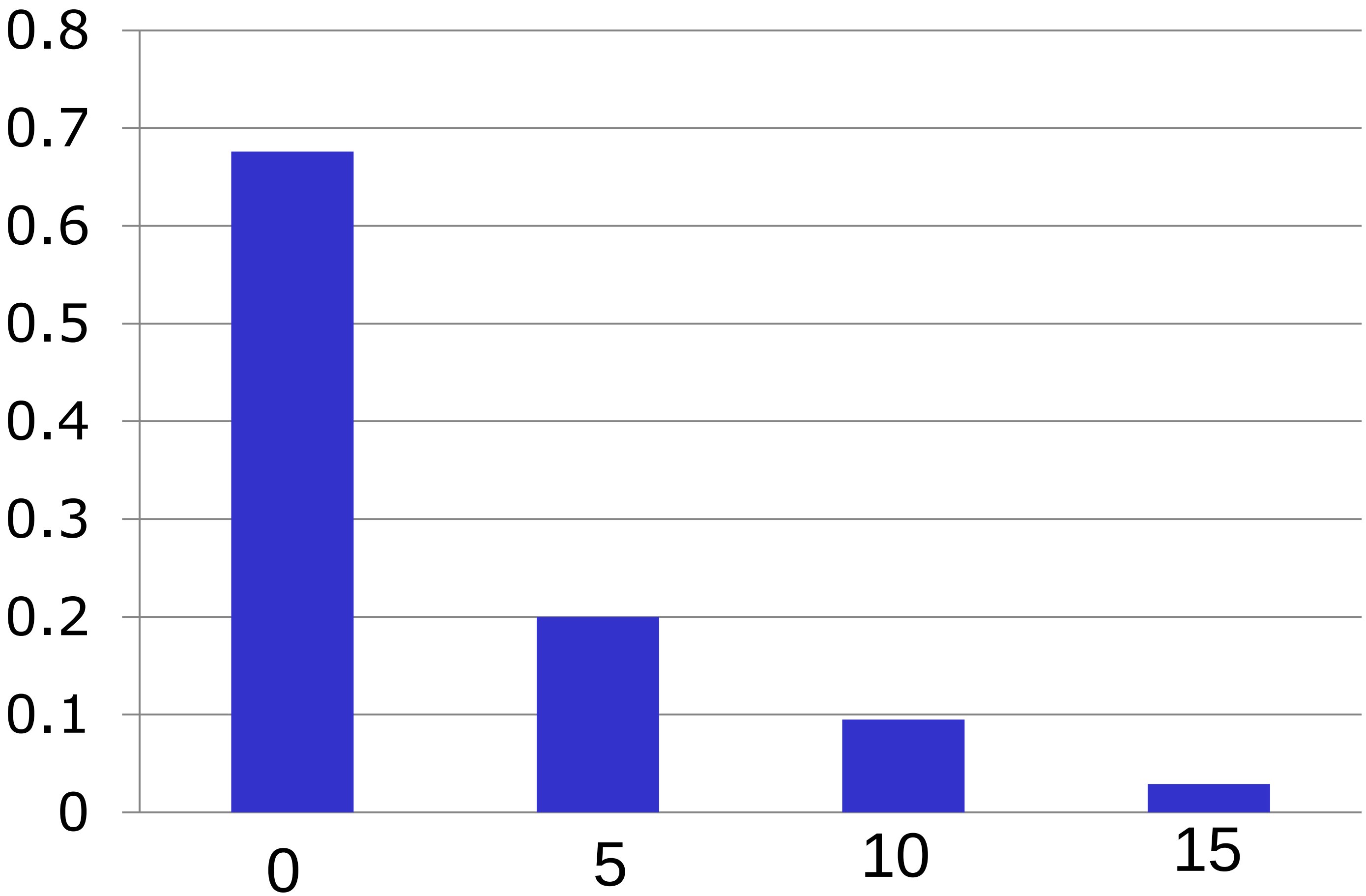
The probability of drawing two greens is $\dfrac{C(7,2)}{C(15,2)} = 0.200$

of two reds $\dfrac{C(5,2)}{C(15,2)} = 0.095$

of two blacks $\dfrac{C(3,2)}{C(15,2)} = 0.029$

X	$P(X)$
0	0.676
5	0.200
10	0.095
15	0.029

1- (0.2+0.095+0.029)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15

One more example

- Two people are selected from a group of five men and four women
- A random variable X is the number of women selected
- What is the range of RV?
- X={0, 1, 2}
- Range is 0 to 2

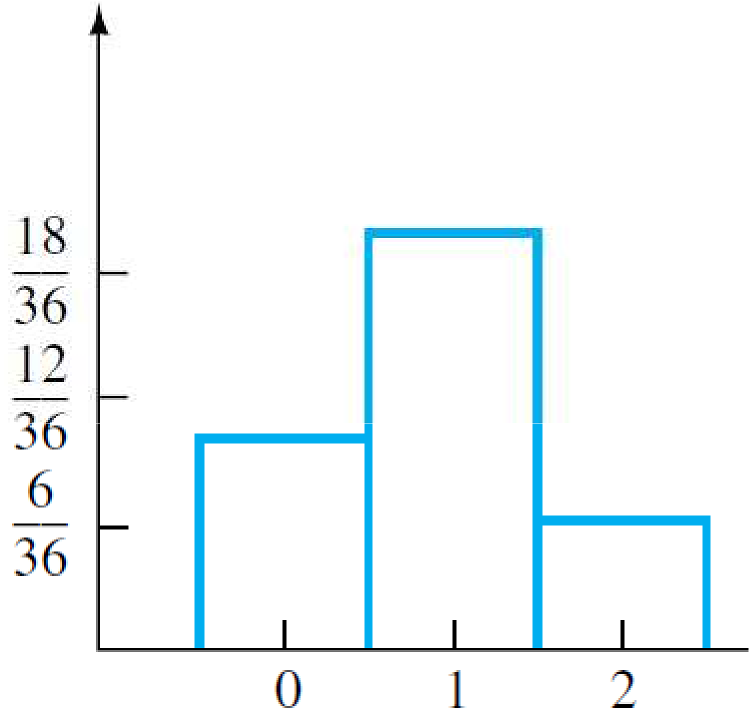
Outcome	X
Three men	0
One man and one woman	1
Two women	2

Outcome	X	Possible ways
Two men	0	C(5,2) = 5!/(3!x2!) = 10
One man and one woman	1	C(5,1).C(4,1)=5!/(4!x1!) x 4!/(3!x1!)=20
Two women	2	C(4,2)=4!/(2!2!)=6

Two people from nine: C(9,2) = 9!/(7!x2!) = 36

Probability for {0, 1, 2}

X	$P(X)$
0	$\frac{10}{36}$
1	$\frac{20}{36}$
2	$\frac{6}{36}$



Average

- N numbers are given
- Add all the numbers and divide by N

Expectation

- Kind of average for random variable
- Dealing with probabilities

Difference between average and expectation - mathematically

Average of N numbers:

- Weighting factor is (1/N)

Expectation of random variable with N outcomes

- Weighting factor is probability

Average

Average = $(x_1+x_2+\ldots.x_n)/n$

$=(1/n).x_1 + (1/n).x_2 + \ldots(1/n).x_n$

- All the data are weighed equally with (1/n)
- (1/n) can also be called as weighing function

Expectation

- Let X represent the outcome of a roll of an unbiased six-sided die
- Possible values for X are 1, 2, 3, 4, 5, & 6
- Each having the probability of occurrence of 1/6

Expectation

$$E(X)=(1/6)*1+(1/6)*2+(1/6)*3+(1/6)*4+(1/6)*5+(1/6)*6$$

$$E(X) = 3.5$$

Average

- Suppose that in a sequence of ten rolls of the die, if the outcomes are 5, 2, 6, 2, 2, 1, 2, 3, 6, 1,
- Average (arithmetic mean) of the results is given by

  (5+2+6+2+2+1+2+3+6+1)/10=3.0

Average and expectation

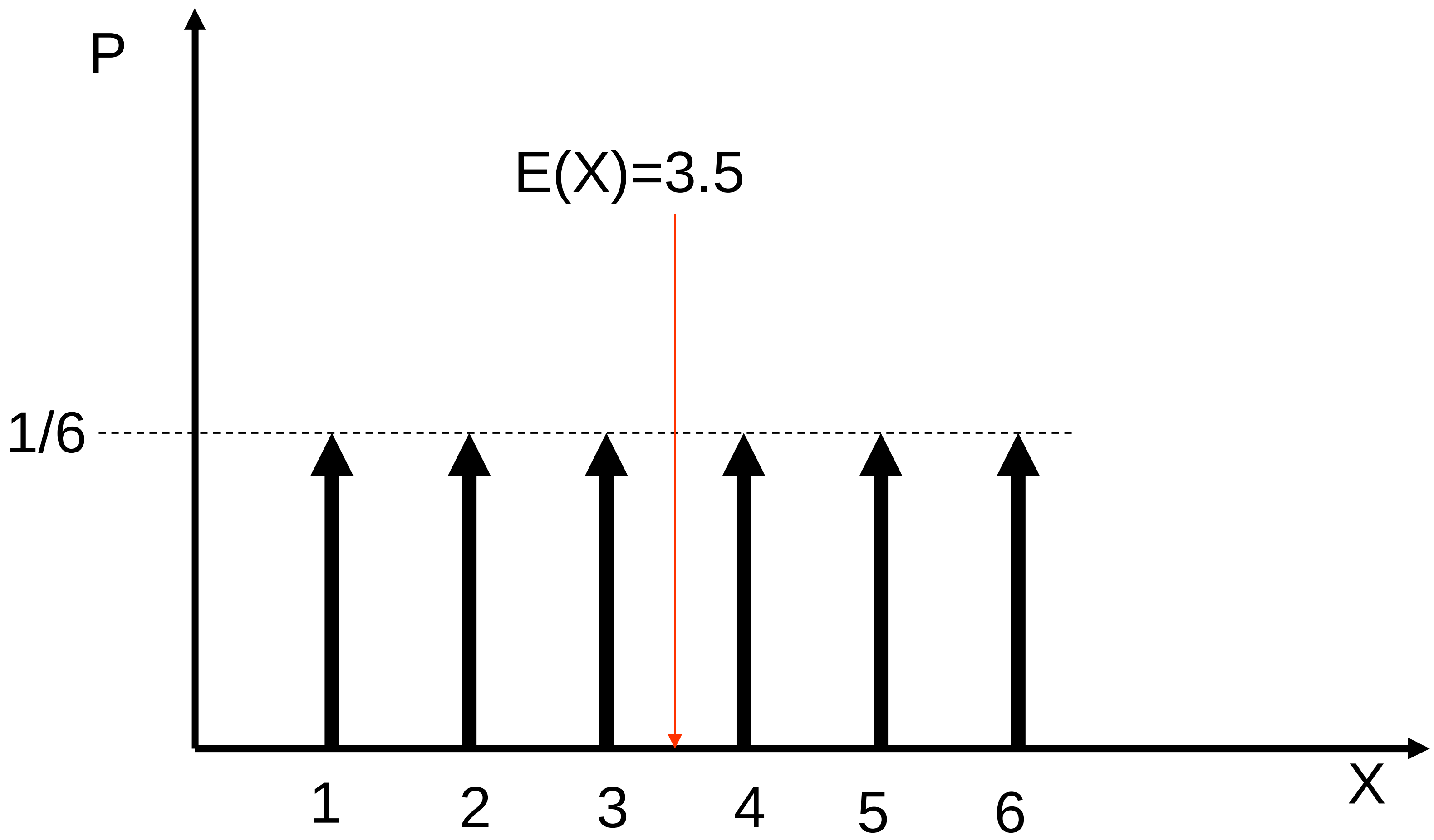
- Expectation: Outcome is not sure i.e. in the form of probability
- Average: Deterministic data values

<u>Previous E.g.</u>

- Average value is 3.0, with the distance of 0.5 from the expectation value of 3.5
- Roll the die N times, N is very large
- Then the average will converge to the expected value,

$$\text{Average }(X) \rightarrow \text{Expected}(X)$$

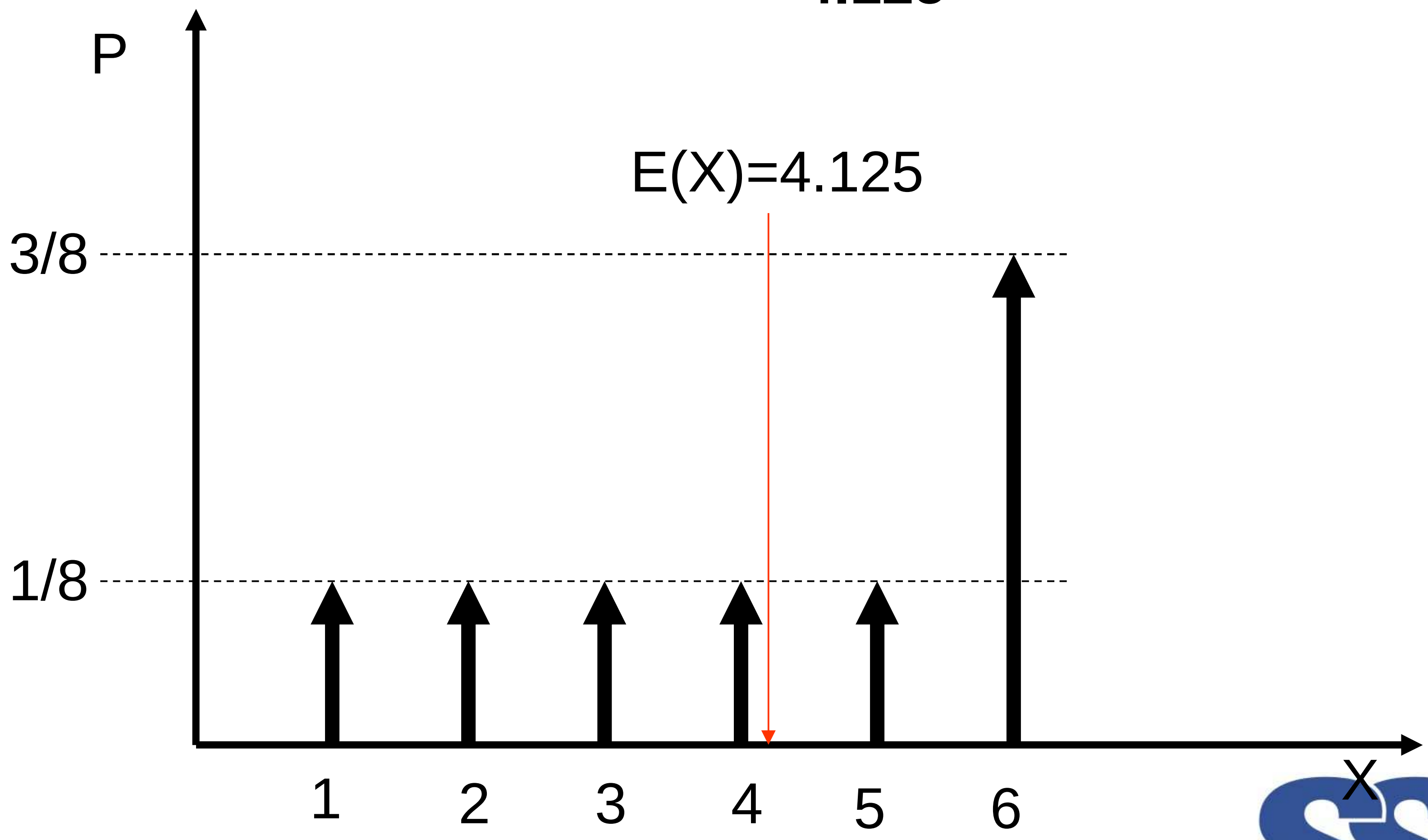
Expectation – rolling fair dice



Expectation – rolling biased dice

<u>Expectation</u>

E(X)=(1/8)*1+(1/8)*2+(1/8)*3+(1/8)*4+(1/8)*5+(3/8)*6 = 4.125



Expectation

- Consider a random variable
- **<u>Probability function</u>**

$[x_1, x_2, \ldots.., x_n]$

$[p_1, p_2, \ldots.., p_n]$

- Average of X

$$\overline{X} = \sum_{k=1}^{n} p_k x_k$$

Expectation - example

x_i	2	5	9	24
p_i	0.4	0.2	0.3	0.1

$$
\begin{aligned}
E(X) &= 0.4 \times 2 + 0.2 \times 5 + 0.3 \times 9 + 0.1 \times 24 \\
&= 0.8 + 1.0 + 2.7 + 2.4 \\
&= 6.9
\end{aligned}
$$

One more example

- A tray of fruits contains nine good and three rotten.
- Two fruits selected at random
- What is the expected number of rotten?
- Let the random variable X be the number of defective
- X can have the value 0, 1, or 2

Solution

$$P(0) = \text{probability of no defective (both good)}$$

$$= \frac{C(9,2)}{C(12,2)} = \frac{36}{66} = \frac{12}{22}$$

$$P(1) = \text{probability of one good and one defective}$$

$$= \frac{C(9,1)C(3,1)}{C(12,2)} = \frac{27}{66} = \frac{9}{22}$$

$$P(2) = \text{probability of two defective}$$

$$= \frac{C(3,2)}{C(12,2)} = \frac{3}{66} = \frac{1}{22}$$

$$E(X) = \frac{12}{22}(0) + \frac{9}{22}(1) + \frac{1}{22}(2) = \frac{11}{22} = \frac{1}{2}$$

Interpretation:

If the experiment is repeated many times

- You expect to get no defectives a little less than half the time
- Either one or two the rest of the time

Standard deviation of a RV

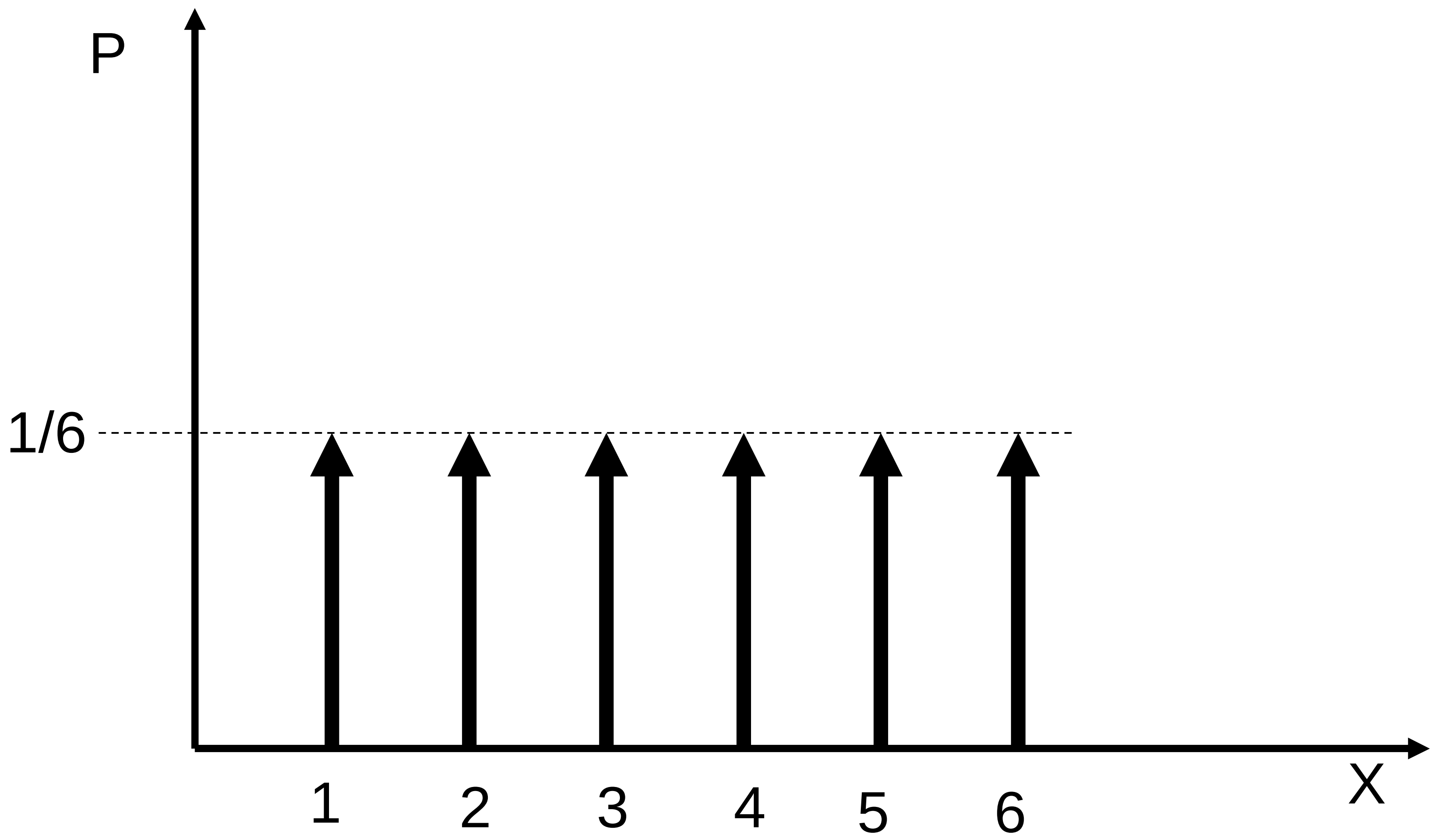
x_i	4	7	10	8
p_i	0.2	0.2	0.5	0.1

x_i	p_i	$p_i x_i$	$x_i - \mu$	$(x_i - \mu)^2$	$p_i(x_i - \mu)^2$
4	0.2	0.8	−4	16	3.2
7	0.2	1.4	−1	1	0.2
10	0.5	5.0	2	4	2.0
8	0.1	0.8	0	0	0
		$\mu = 8.0$			$\sigma^2(X) = 5.4$

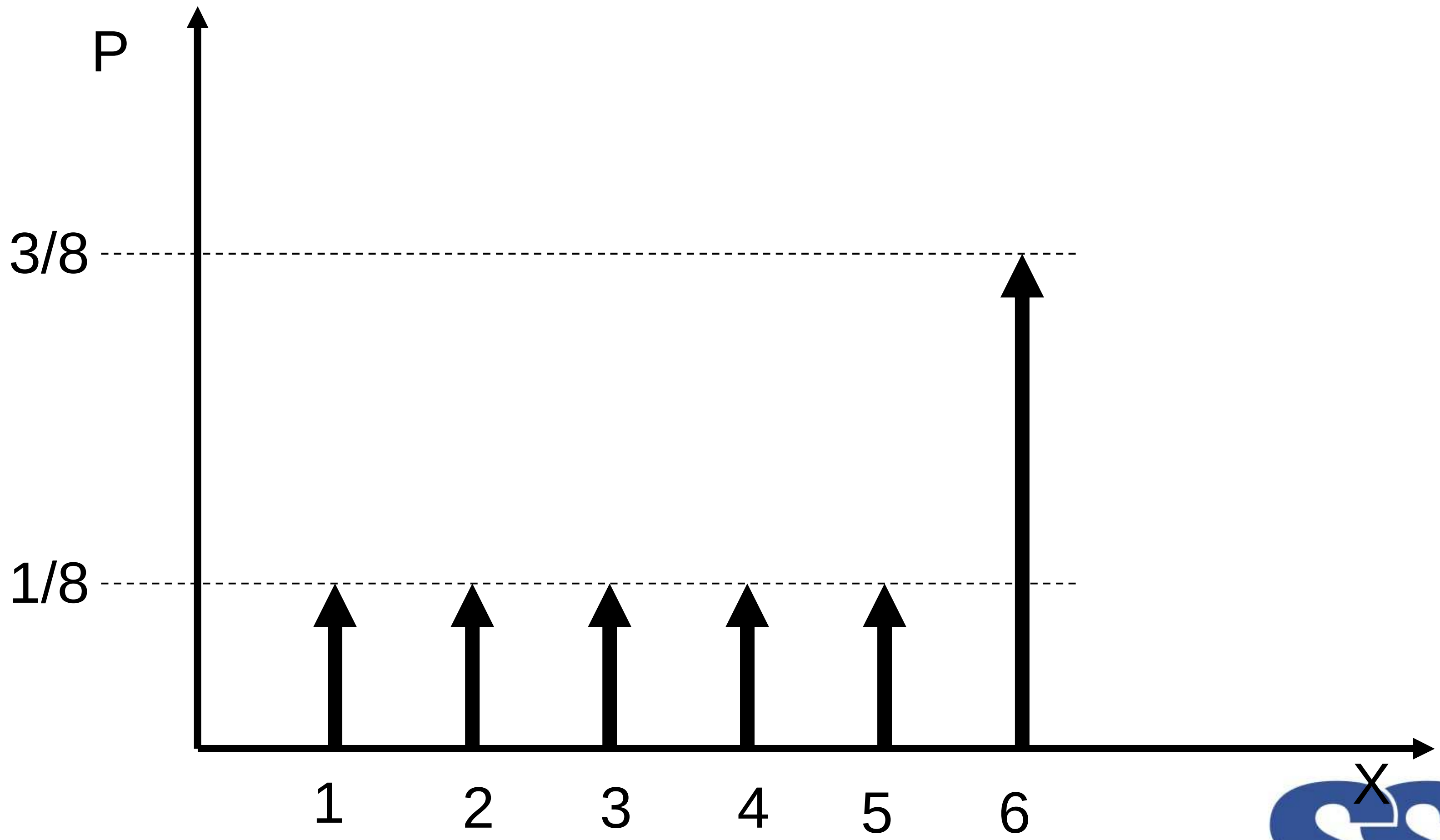
The variance of $\sigma^2(X) = 5.4$. So the standard deviation is

$$\sigma(X) = \sqrt{5.4} = 2.32$$

Uniform distribution



Non-uniform distribution



Random variable

Discrete

Probability mass function

Discrete

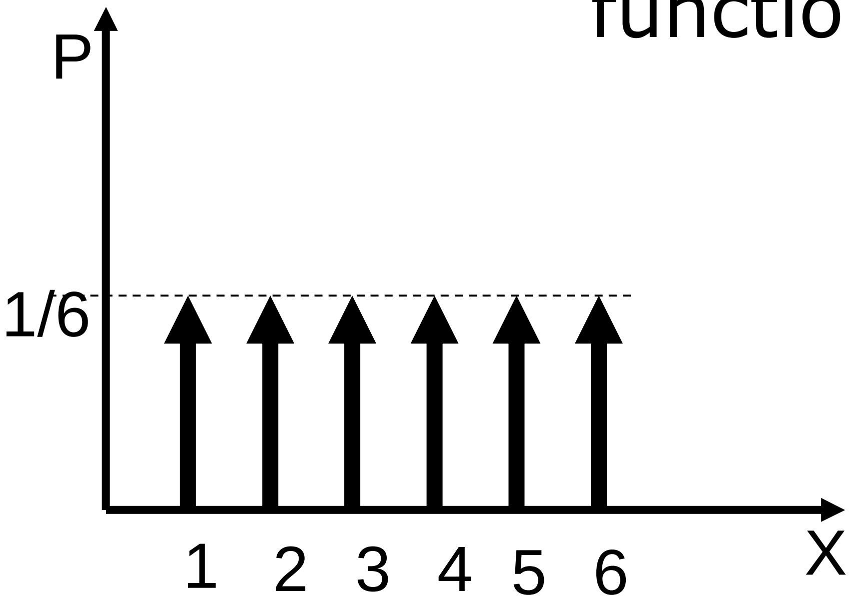
Continuous

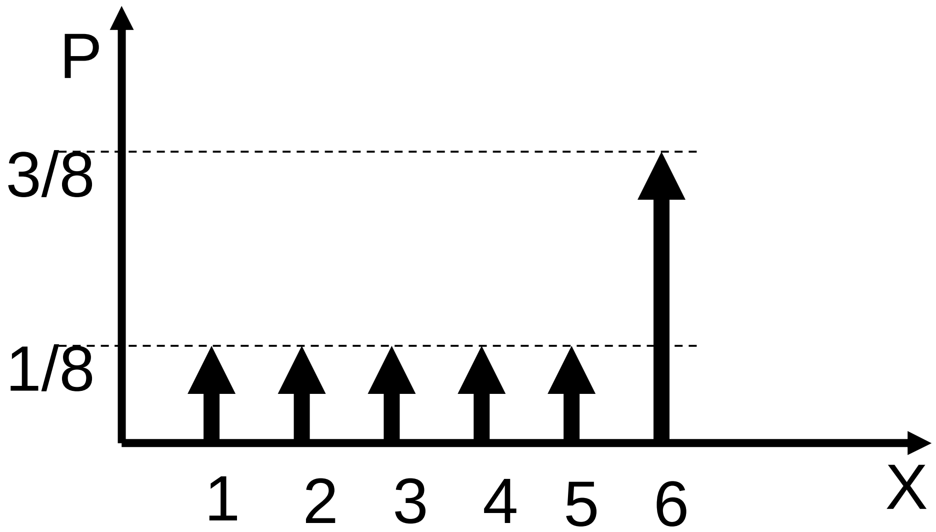
Probability density function

Continuous

Probability distribution

Discrete random variable: Probability distribution – Probability Mass function - PMF

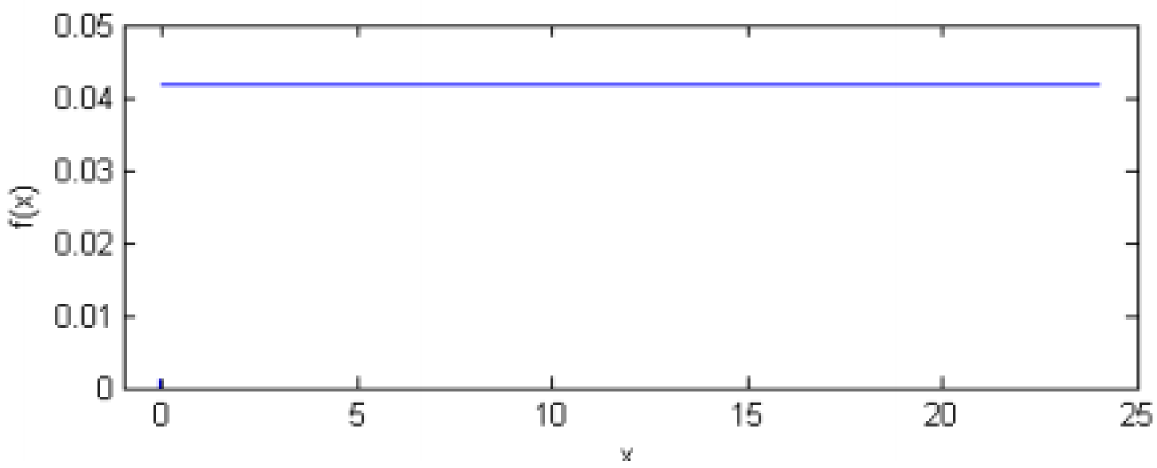




$$\Sigma\, p(i) = 1$$

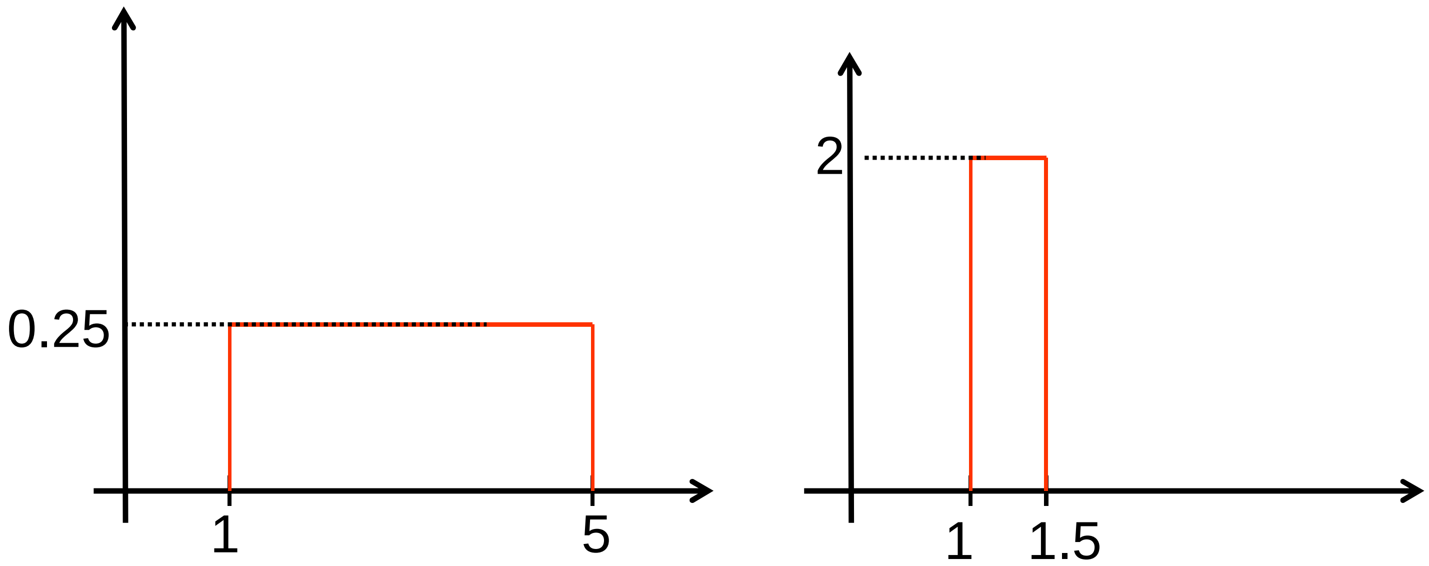
Continuous random variable: Probability distribution - Probability Density function - PDF

- In PDF – what should not be asked?
- What is the probability @ some value of x?
- Ask – what is the probability for x=x1 to x=x2 i.e. interval?

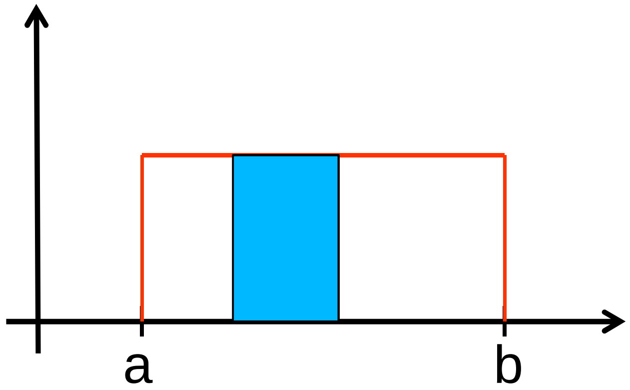


Some more points on PDF

- PMF values cannot be greater than 1
- But PDF values can be greater than 1
- Area under PDF = 1 (only condition)
- Consider following uniform distributions



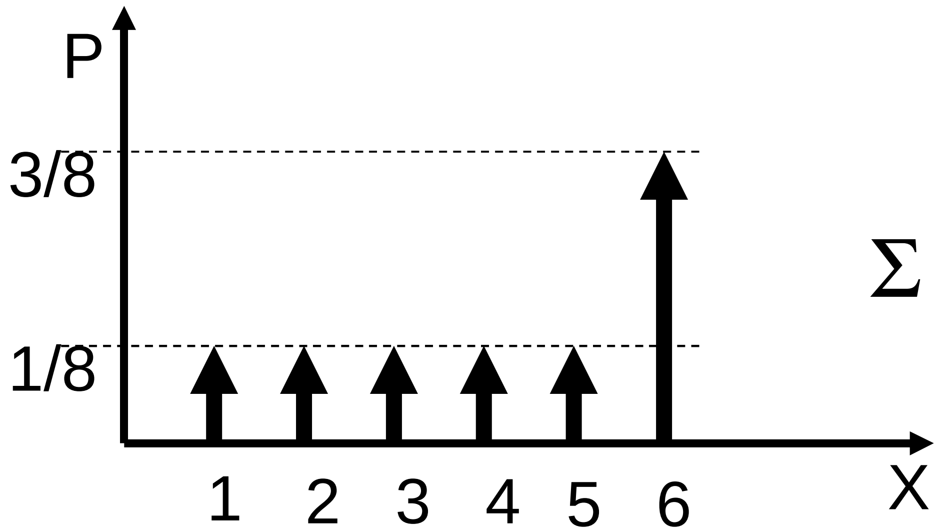
PDF - Area gives probability



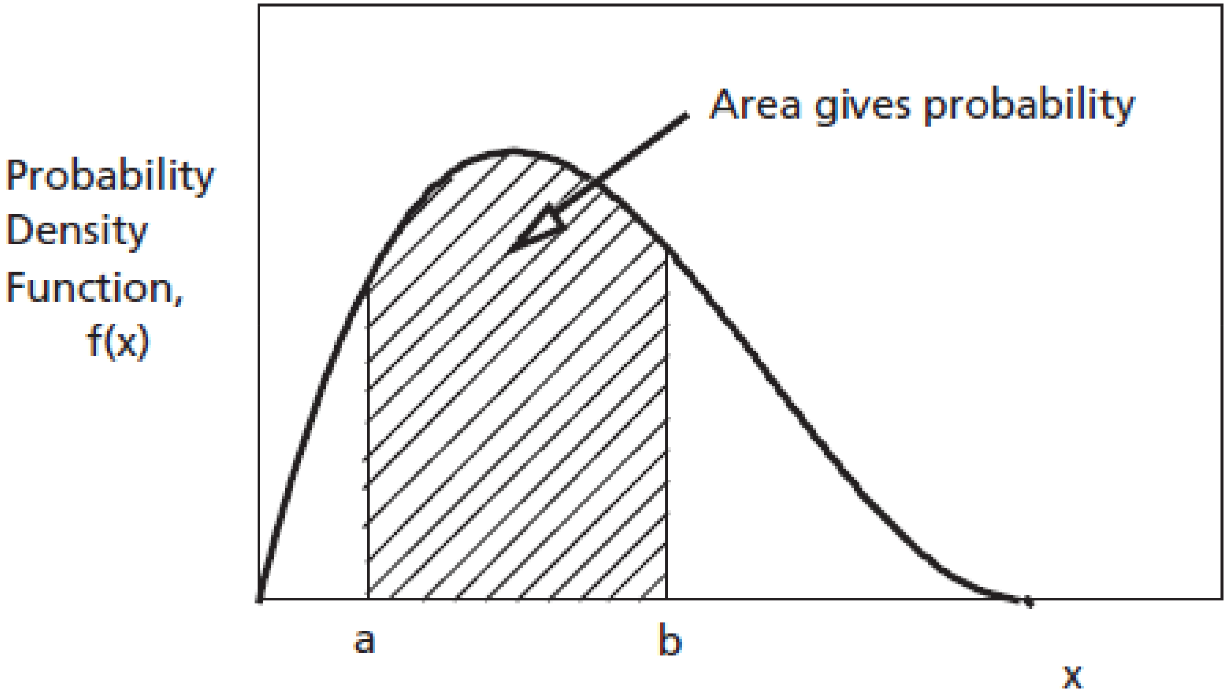
$$p(x1<x<x2)= \int_{x1}^{x2} p(x)\ dx$$

$$\int_a^b p(x)\ dx = 1$$

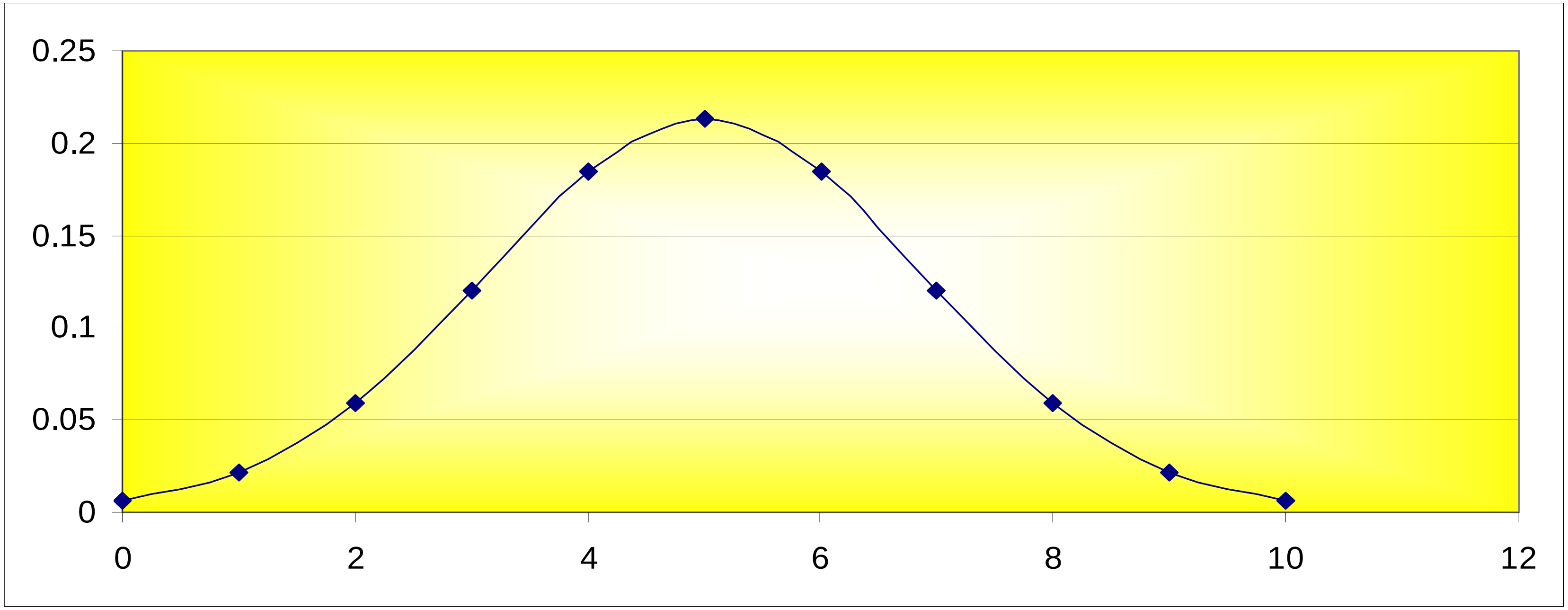
PMF – probability values can be directly read



$$\Sigma\ p(i) = 1$$


Area gives probability
Probability
Density
Function,
f(x)
a
b
x

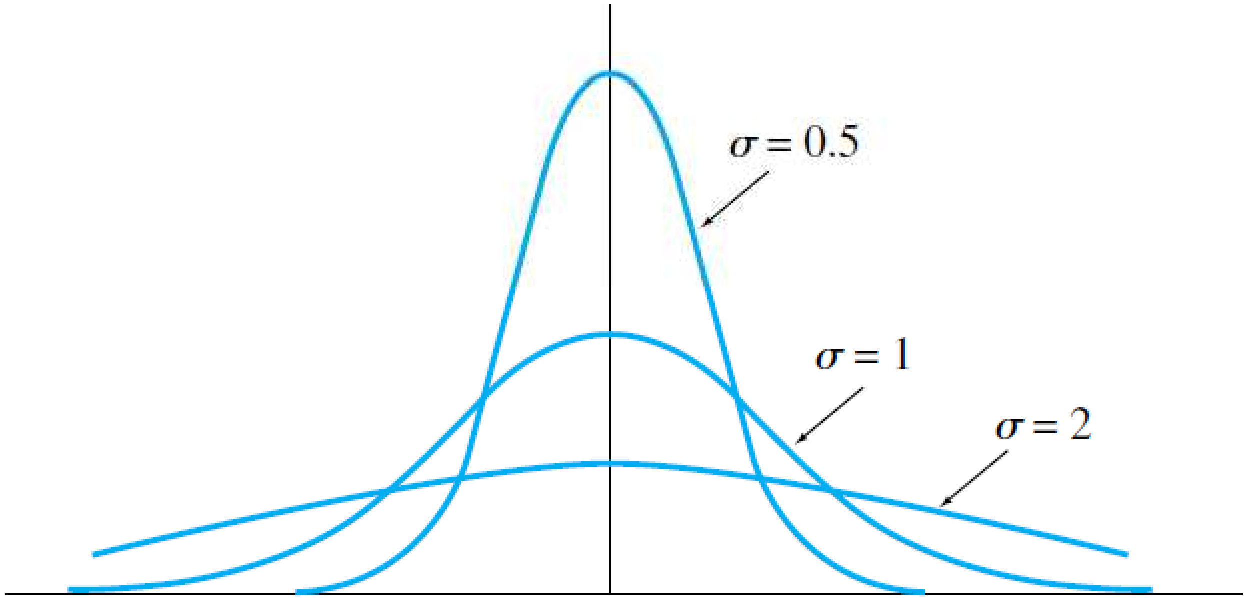
Gaussian distribution – mean and standard deviation



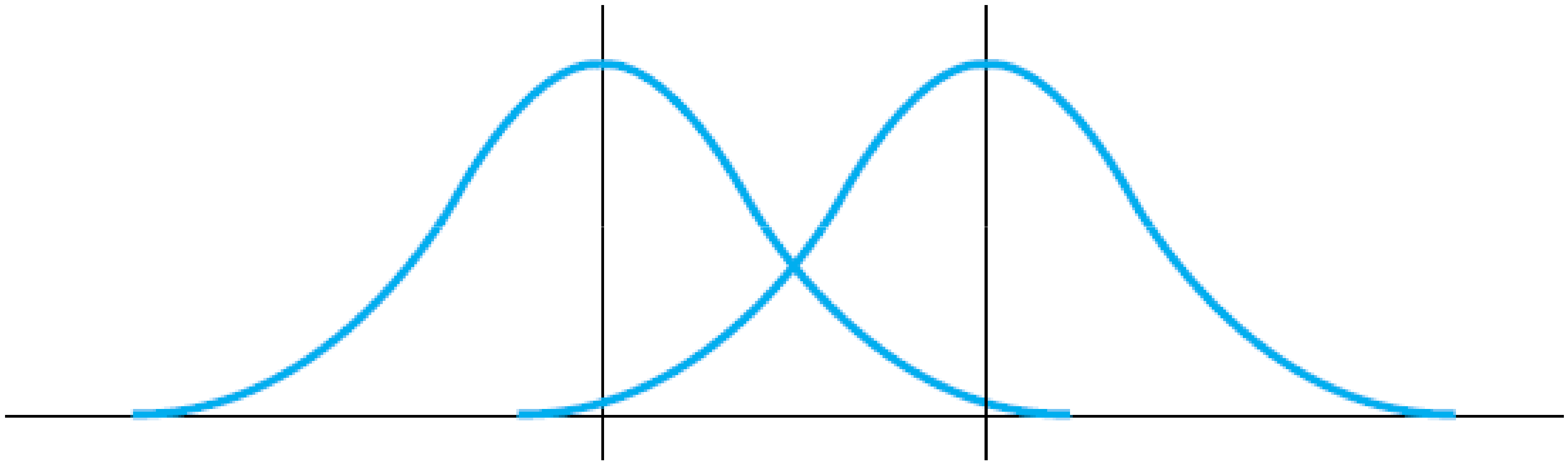
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Defined from
-∞ to +∞

Same mean – different standard deviations



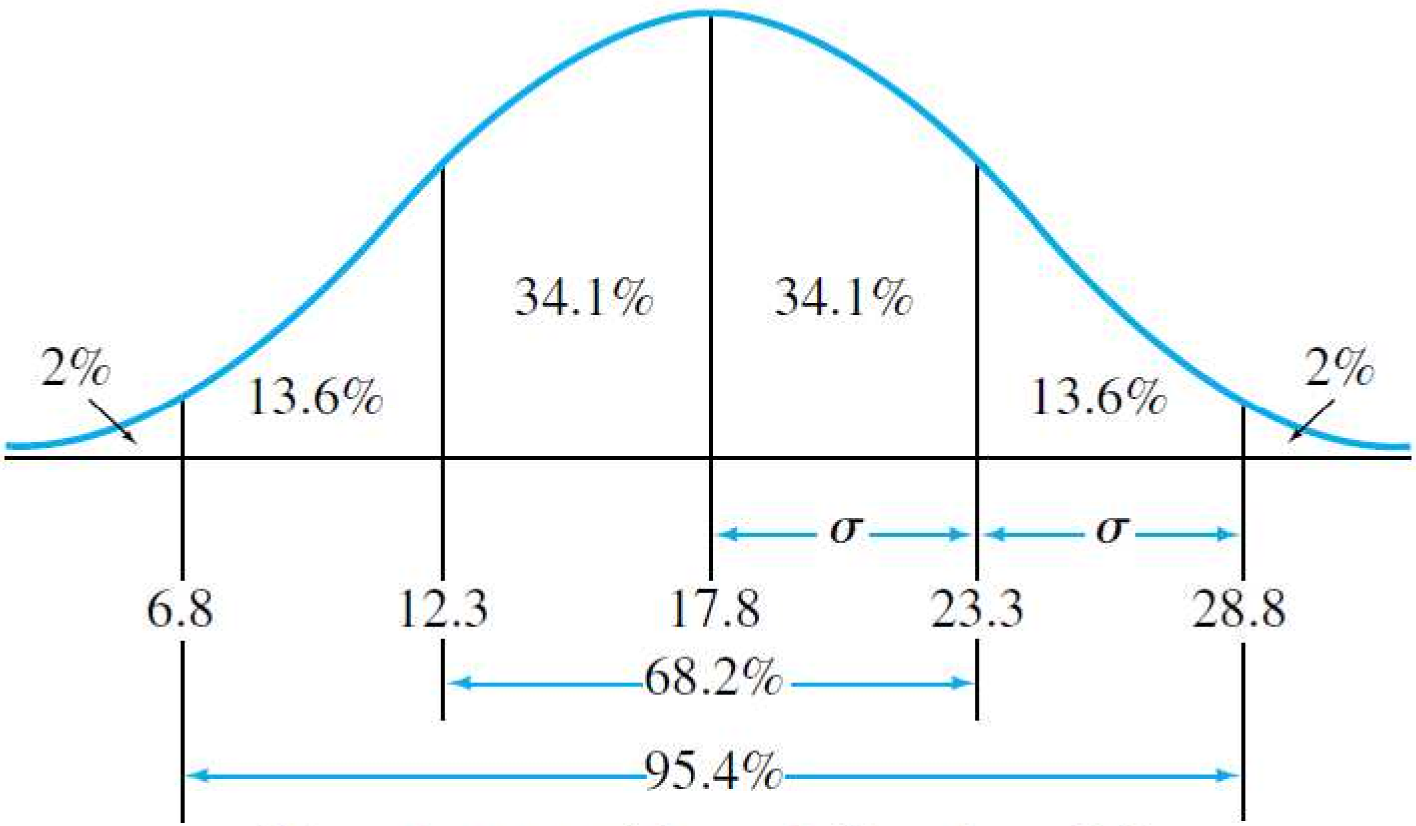
Same standard deviation – different means



Standard normal deviation

- Make, $\sigma = 1$
- Area under the curve becomes 1
- can be used as probability measure

Area under the normal curve



Normal curve with $\mu = 17.8$ and $\sigma = 5.5$

$$\overline{X} = \sum_{k=1}^{n} p_k x_k$$

More generic form,

$$\overline{\psi(X)} = \sum_{k=1}^{n} p_k . \psi(x_k)$$

Moments

$$E(X) = \bar{X} = \text{first-order moment of } X = \sum_{k=1}^{n} p_k x_k$$

$$E(X^2) = \overline{X^2} = \text{second-order moment of } X = \sum_{k=1}^{n} p_k {x_k}^2$$

$$E(X^3) = \overline{X^3} = \text{third-order moment of } X = \sum_{k=1}^{n} p_k {x_k}^3$$