CS1512
Foundations of Computing Science 2

Lecture 4

Bayes Law; Gaussian Distributions

UNIVERSITY OF ABERDEEN

(Revd. Thomas) Bayes Theorem

$$P(E_1 \text{ and } E_2) = P(E_1) * P(E_2 | E_1)$$

Order of E_1 and E_2 is not important.

(Revd. Thomas) Bayes Theorem

$$P(E_1 \, and \, E_2) = P(E_1) * P(E_2 \mid E_1)$$

Order of E_1 and E_2 is not important.

$$P(E_2 \, and \, E_1) = P(E_2) * P(E_1 \mid E_2) \; .$$

(Revd. Thomas) Bayes Theorem

$$P(E_1 \text{ and } E_2) = P(E_1) * P(E_2 | E_1)$$

Order of E_1 and E_2 is not important.

$$P(E_2 \text{ and } E_1) = P(E_2) * P(E_1 | E_2)$$. So,

$$P(E_2) * P(E_1 | E_2) = P(E_1) * P(E_2 | E_1)$$. So,

$$P(E_1 | E_2) = \frac{P(E_1) * P(E_2 | E_1)}{P(E_2)}$$

(Revd. Thomas) Bayes Theorem

$$P(E_1 \mid E_2) = \frac{P(E_1) * P(E_2 \mid E_1)}{P(E_2)}$$

E.g., once you know P(E1) and P(E2),
knowing *one* of the two conditional probabilities
(that is, knowing either P(E1|E2) or P(E2|E1))
implies knowing *the other* of the two.

(Revd. Thomas) Bayes Theorem

$$P(E_1 | E_2) = \frac{P(E_1) * P(E_2 | E_1)}{P(E_2)}$$

A simple example, *without replacement*:

Two red balls and one black: RRB

P(R)=2/3

P(B)=

P(R|B)=

P(B|R)=

(Revd. Thomas) Bayes Theorem

$$P(E_1 | E_2) = \frac{P(E_1) * P(E_2 | E_1)}{P(E_2)}$$

A simple example:

Two red balls and one black: RRB

P(R)=2/3

P(B)=1/3

P(R|B)=1

P(B|R)=1/2

(Revd. Thomas) Bayes Theorem

$$P(E_1 \mid E_2) = \frac{P(E_1) * P(E_2 \mid E_1)}{P(E_2)}$$

A simple example:

Two red balls and one black: RRB

P(R)=2/3

P(B)=1/3

P(R|B)=1

P(B|R)=1/2. Now calculate P(R|B) using Bayes:

P(R|B)=

(Revd. Thomas) Bayes Theorem

$$P(E_1 \mid E_2) = \frac{P(E_1) * P(E_2 \mid E_1)}{P(E_2)}$$

A simple example:

Two red balls and one black: RRB

P(R)=2/3

P(B)=1/3

P(R|B)=1

P(B|R)=1/2 Now calculate P(R|B) using Bayes:

P(R|B)=(2/3*1/2)/1/3=(2/6)/(1/3)=1 --- correct!

Bayes Theorem

Why bother?

- Suppose E_1 is a disease (say measles)
- Suppose E_2 is a symptom (say spots)
- $P(E_1)$ is the probability of a given patient having measles (before you see them)
- $P(E_2)$ is the probability of a given patient having spots (for whatever reason)
- $P(E_2 \mid E_1)$ is the probability that someone who has measles has spots

Bayes theorem allows us to calculate the probability:

$P(E_1 \mid E_2)$ = P (patient has measles given that the patient has spots)

Discrete random variables

Discrete random variable

- a discrete variable where the values taken are associated with a probability

Discrete probability distribution

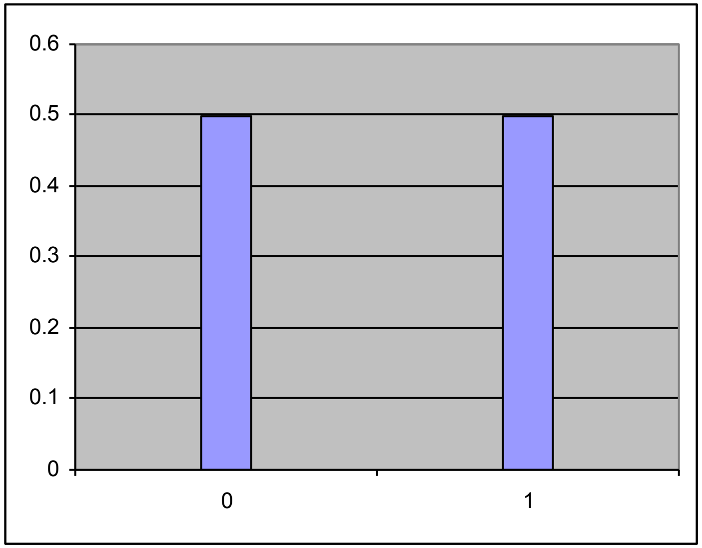
- formed by the possible values and their probabilities
- $P(X$ = <some value>) = ...

Simple example – fair coin

Let variable (*X*) be number of heads in one trial (toss).
Values are 0 and 1

Probability distribution:

Value of *X*	0	1
$P(X = i)$	0.5	0.5



Example – total of two dice

Let variable (X) be the **total** when two fair dice are tossed.

Its values are 2, 3, 4, ... 11, 12

Are these "combination" values equally probable?

Example – total of two dice

Let variable (X) be the **total** when two fair dice are tossed.

Its values are 2, 3, 4, ... 11, 12

Not all these "combination" values are equally probable!

Probability distribution:

$P(X = 2) = 1/36$

$P(X = 3) = 2/36$ (because 3 = 1+2 = 2+1)

$P(X = 4) = 3/36$ (because 4 = 1+3 = 3+1 = 2+2)

...

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

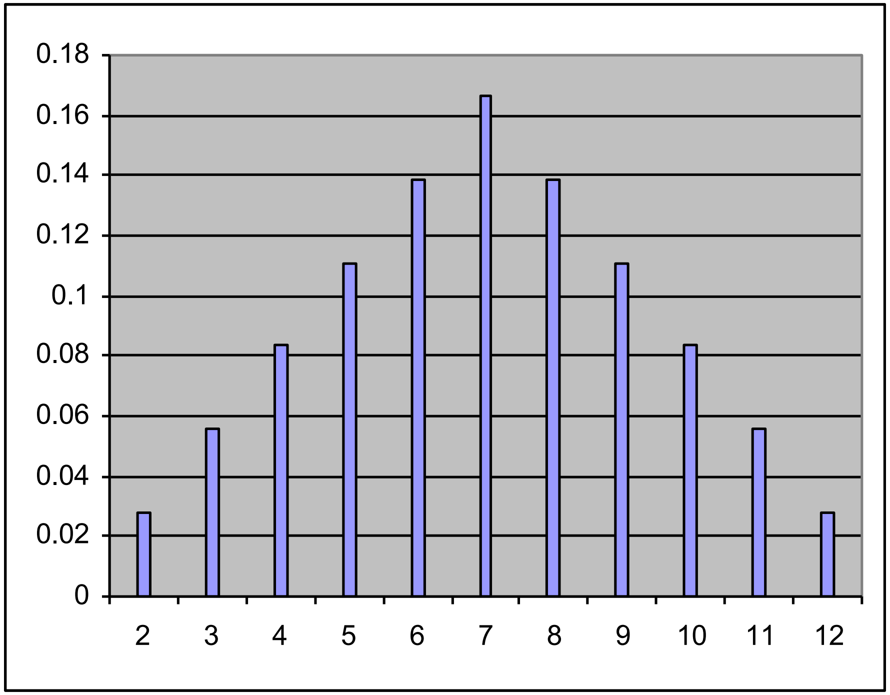
Example – total of two dice

Let variable (X) be the **total** when two fair dice are tossed.
Its values are 2, 3, 4, ... 11, 12

Probability distribution:

$P(X = 2) = 1/36$
$P(X = 3) = 2/36$
...

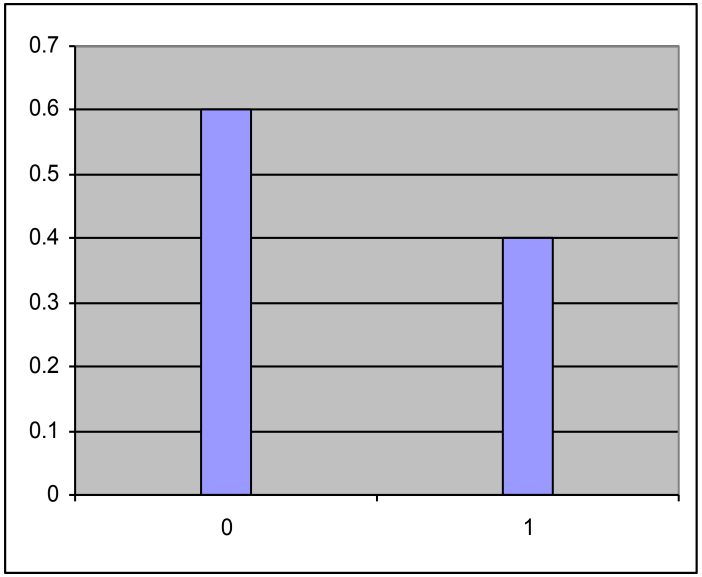
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12



Bernouilli probability distribution

Bernoulli trial:

- one in which there are only two possible and exclusive outcomes;
- call the outcomes 'success' and 'failure';
- consider the variable X which is the number of **successes** in one trial



Probability distribution:

Value of X	0	1
$P(X = i)$	$(1-p)$	p

(fair coin ➔ p=1/2 ;   graph → p=4/10)

Binomial distribution

Carry out n Bernoulli trials. X is the number of successes in n trials.

Possible values are: 0, 1, 2, ... i, ... n-1, n

What is $P(X=r)$? (= the probability of scoring exactly r successes in n trials)

Binomial distribution

Carry out n Bernoulli trials. X is the number of successes in n trials.

Possible values are: 0, 1, 2, ... i, ... n-1, n

What is $P(X=r)$? (= the probability of scoring exactly r successes in n trials)

Let **s**=success and **f**=failure. Then a trial is a tuple like e.g. <**s,s,f,s,f,f,f**> (n=7)

Generally, a trial is an n-tuple consisting of **s** and **f**.

Consider the 7-tuple <**f,s,f,s,f,f,f**> . Its probability is

P(f)*P(s)*P(f)*P(s)*P(f)*P(f)*P(f). That's $P(s)^2 * P(f)^5$.

Probability of a given tuple with r **s**-es and $n-r$ **f**-s in it = $P(s)^r * P(f)^{n-r}$

Binomial distribution

Probability of *a given tuple* with r **s**-es and n-r **f**-s in it = $P(s)^r * P(f)^{n-r}$

We're interested in *P(X=r), that's* the probability of *all tuples together* that have r **s**-es and n-r **f**-s in them (i.e. the sum of their probabilities).

How many ways are there of picking exactly r elements out of a set of n?

Combinatorics says there are

$$\frac{n!}{r!\,(n-r)!} \qquad \textbf{(Def: } n! = n * (n-1) * (n-2) * \ldots * 2 * 1)$$

Try a few values. *E.g., if n=9 and r=1 then 9! / 1!(9-1)! = 9! / 8! = 9.*

So, we need to multiply $P(s)^r * P(f)^{n-r}$ with this factor.

UNIVERSITY OF ABERDEEN

Binomial distribution

Combining what you've seen in the previous slides

Carry out n Bernoulli trials; in each trial the probability of success is $P(s)$ and the probablity of failure is $P(f)$

The distribution of X (i.e., the number of successes) is:

$$P(X=r) = {}^{n}C_{r} * P(s)^{r} * P(f)^{n-r}$$

$${}^{n}C_{r} = \frac{n!}{r!\,(n-r)!}$$

UNIVERSITY OF ABERDEEN

Binomial distribution
(slightly different formulation)

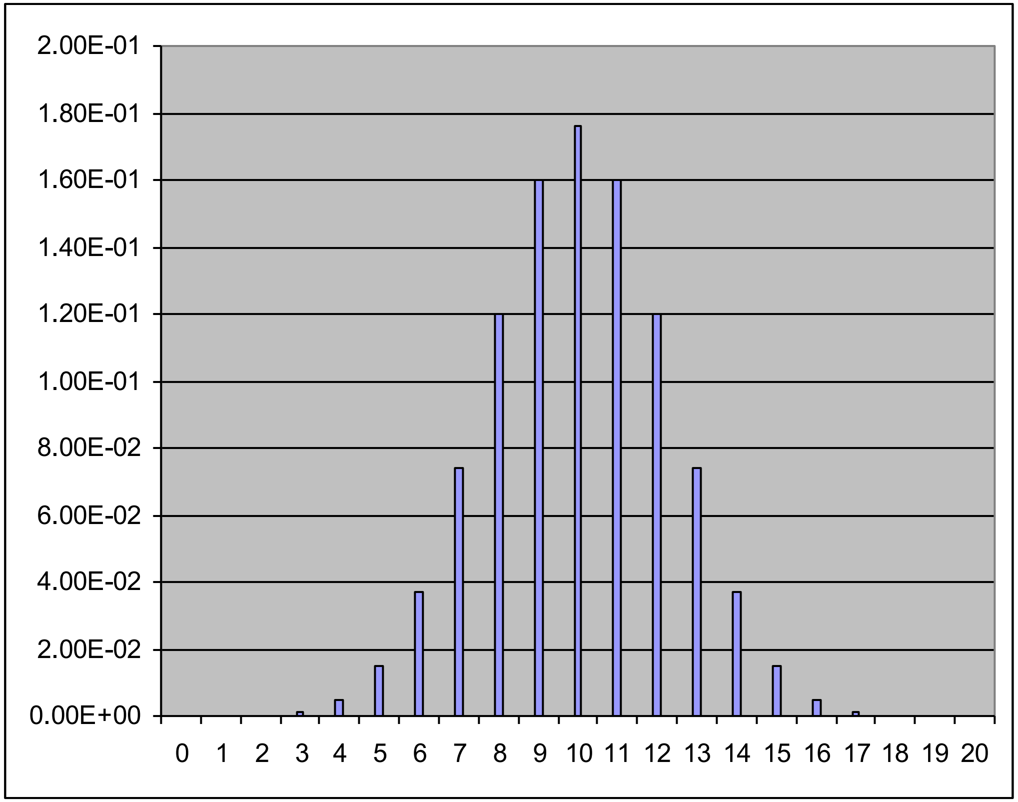
$$P(X=r) = {}^{n}C_{r} * p^{r} * (1-p)^{n-r}$$

Remember CS1015, combinations & permutations

Example of binomial distribution

Probability of r heads in n trials (n = 20)

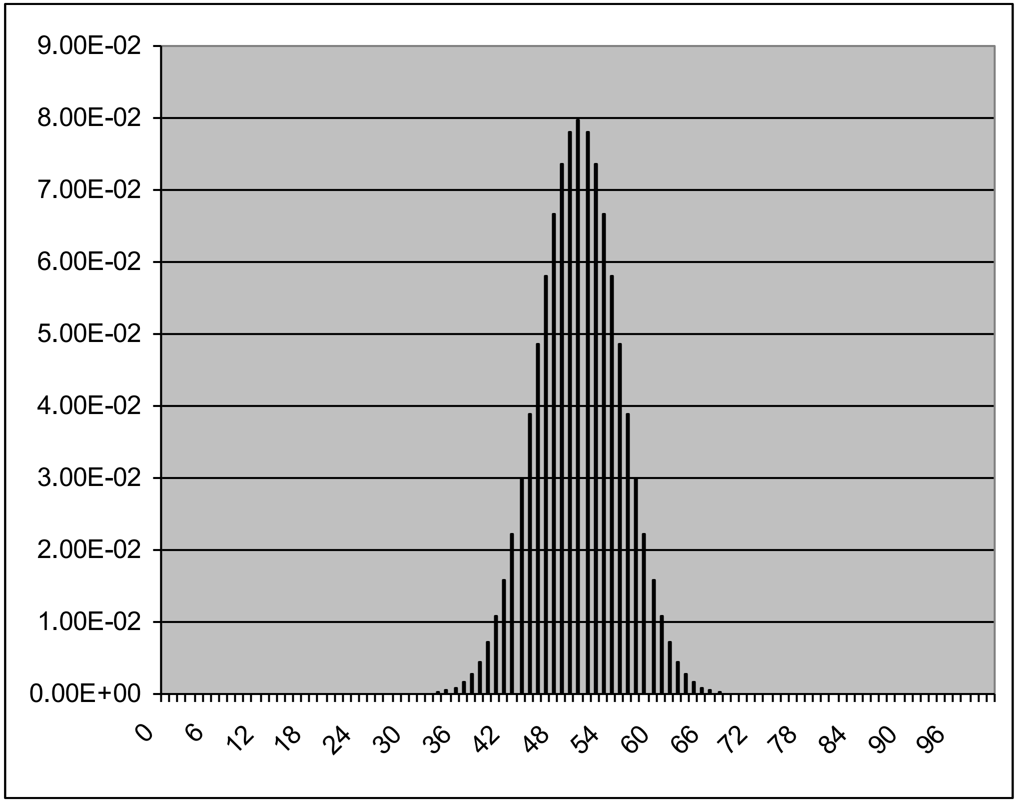
$P(X = r)$



Example of binomial distribution

Probability of r heads in n trials (n = 100)

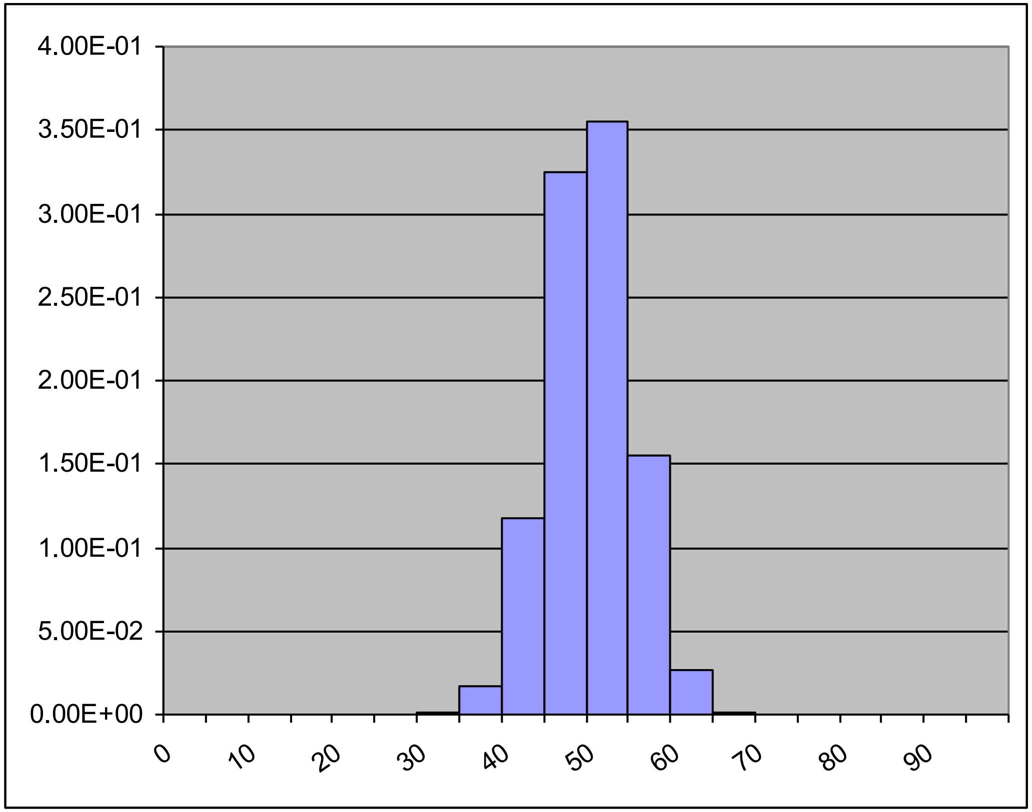
$P(X = r)$



Probability of being in a given range

Probability of r heads in n trials (n = 100)

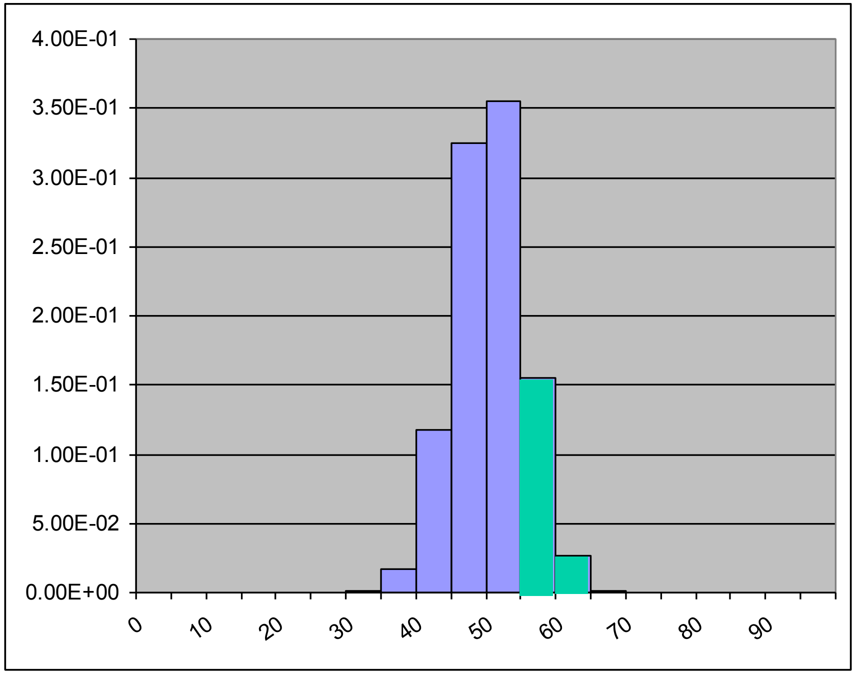
$P(0 \geq X < 5)$
$P(5 \geq X < 10)$
…



Probability of being in a given range

***P*(55 ≥ *X* < 65) = *P*(55 ≥ *X* < 60) + *P*(60 ≥ *X* < 65)** (mutually exclusive events)

Add the heights of the columns



Continuous random variable

Consider a large population of individuals
e.g. all males in the UK over 16

Consider a continuous attribute
e.g. Height: X

Select an individual at random so that any individual is as likely to be selected as any other, observe his/her Height

X is said to be a continuous random variable

Continuous random variable

Suppose the heights in your sample range
from 31cm (a baby) to 207cm
Continuous → can't list every height separately → need bins.

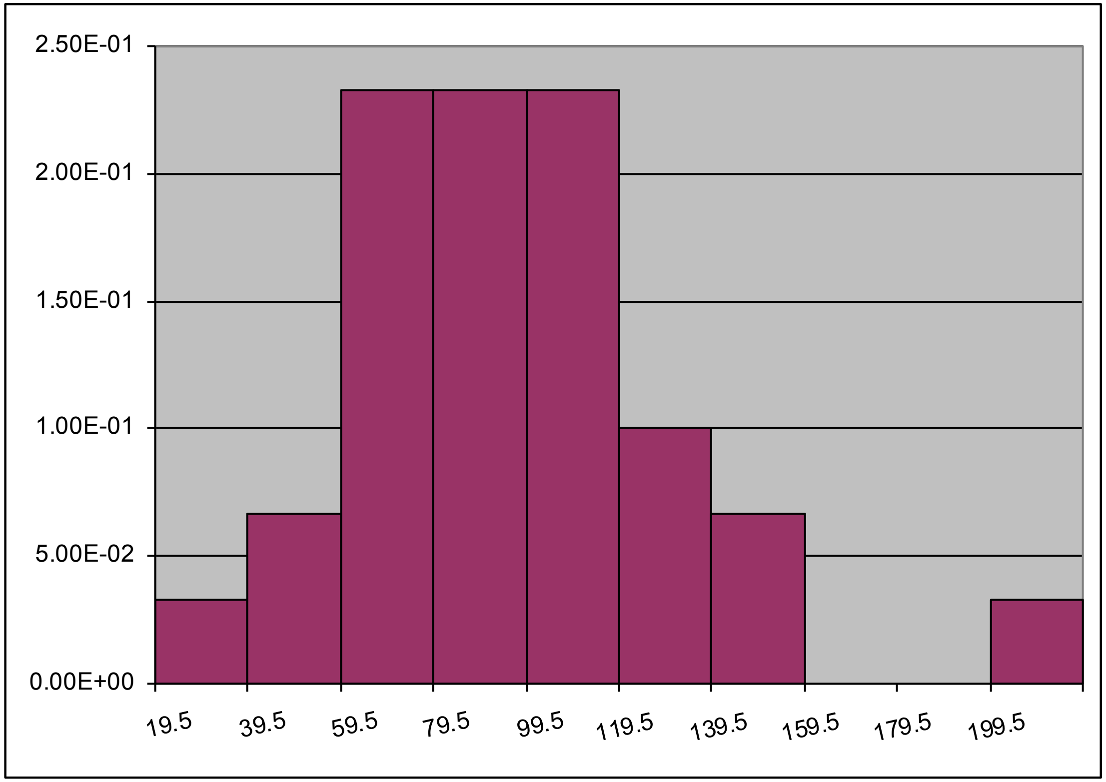
If you are satisfied with a rough division into a few
dozen categories, you might decide to use bins as follows:

- 19.5 - 39.5cm: relative frequency = 3/100
- 39.5 - 59.5cm: relative frequency = 7/100
- 59.5 - 79.5cm: relative frequency = 23/100
- *etc*.

Relative frequencies may be plotted as follows:

Relative Frequency Histogram

relative frequency



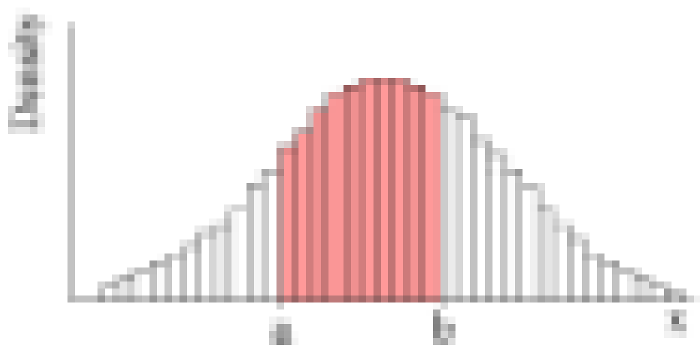
height of the bar gives the relative frequency

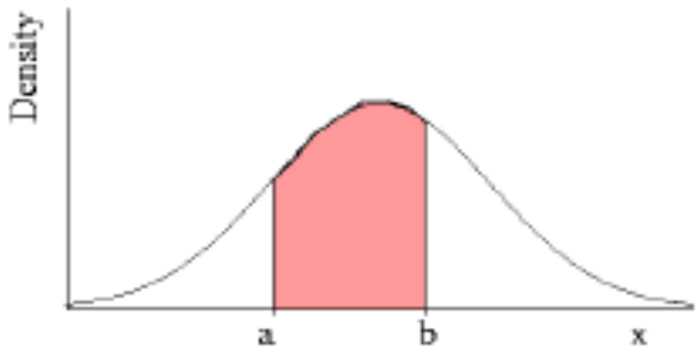
UNIVERSITY OF ABERDEEN

Probability density function

The probability distribution of X is the function ***P*** such that, for all ***a*** and ***b*** :

$P(\{x : a \geq x > b\})$ = the area under the curve between a and b

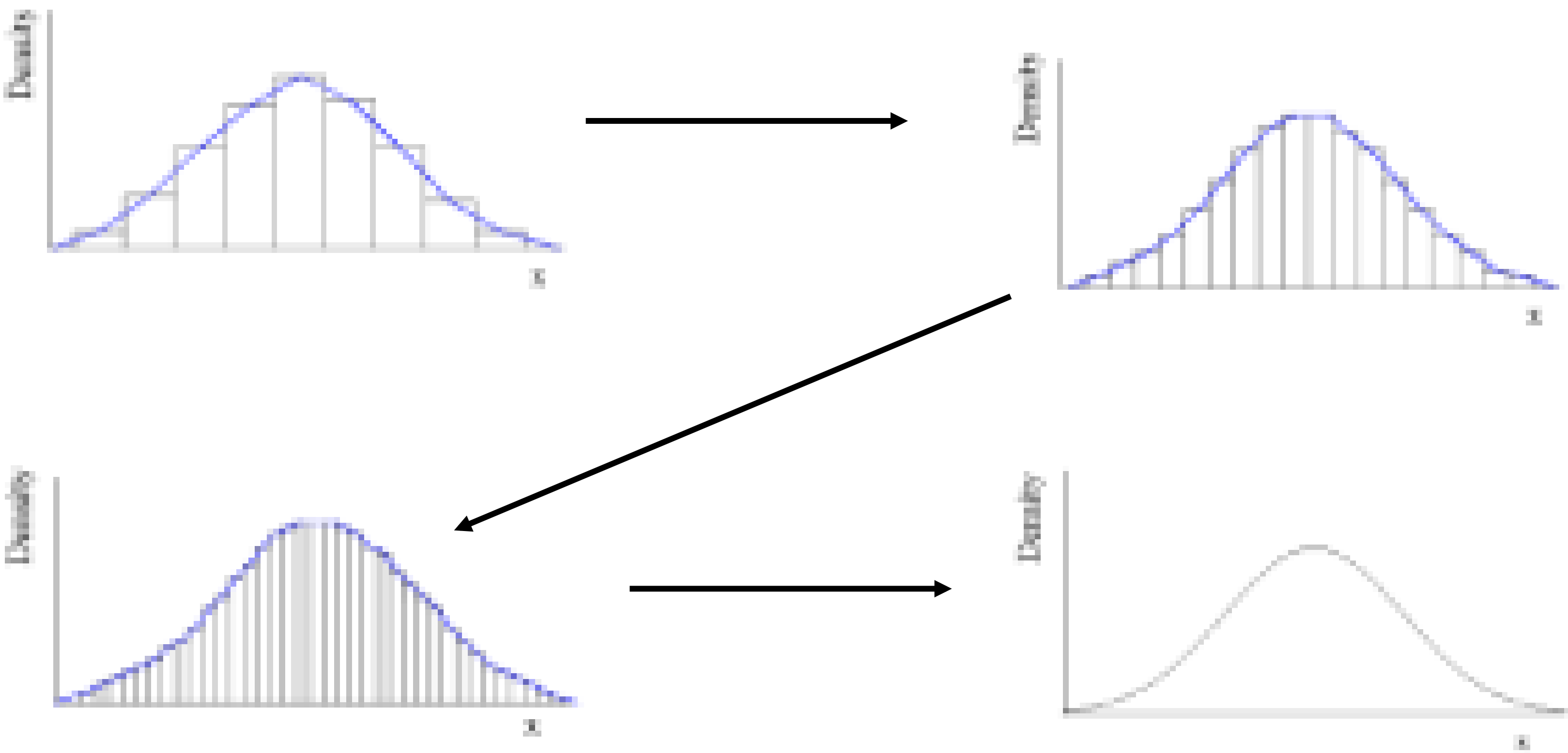




P (total area under curve) = 1.0

Increase the number of samples

... and decrease the width of the bin ...



The ‘normal’ or Gaussian distribution

Very common distribution:

- variable measured for large number of nominally identical objects;
- variation assumed to be caused by a large number of factors;
- each factor exerts a small random positive or negative influence;
- e.g. height: age, diet, genes

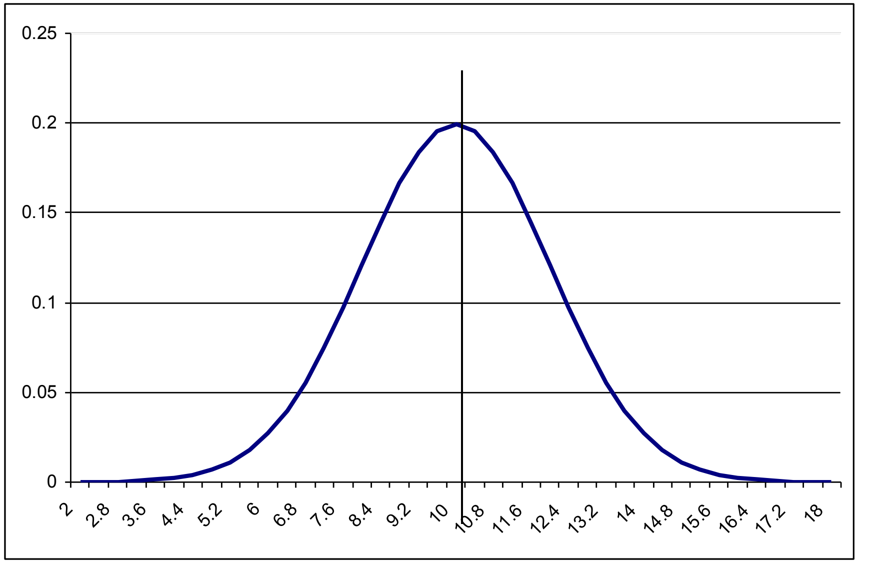
-- Symmetric about mean μ

-- Monotonically increasing where < μ

-- Monotonically descreasing where > μ

-- (Unimodal: only one peak)

The greater the number of trials (n), the more closely does the binomial Distribution approximate a normal Distribution

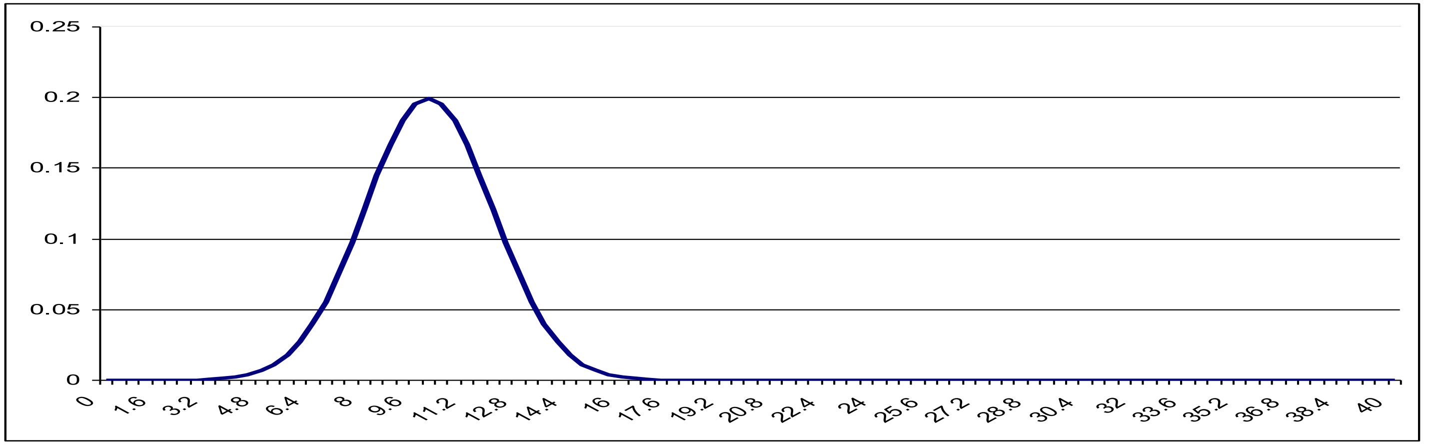


UNIVERSITY OF ABERDEEN

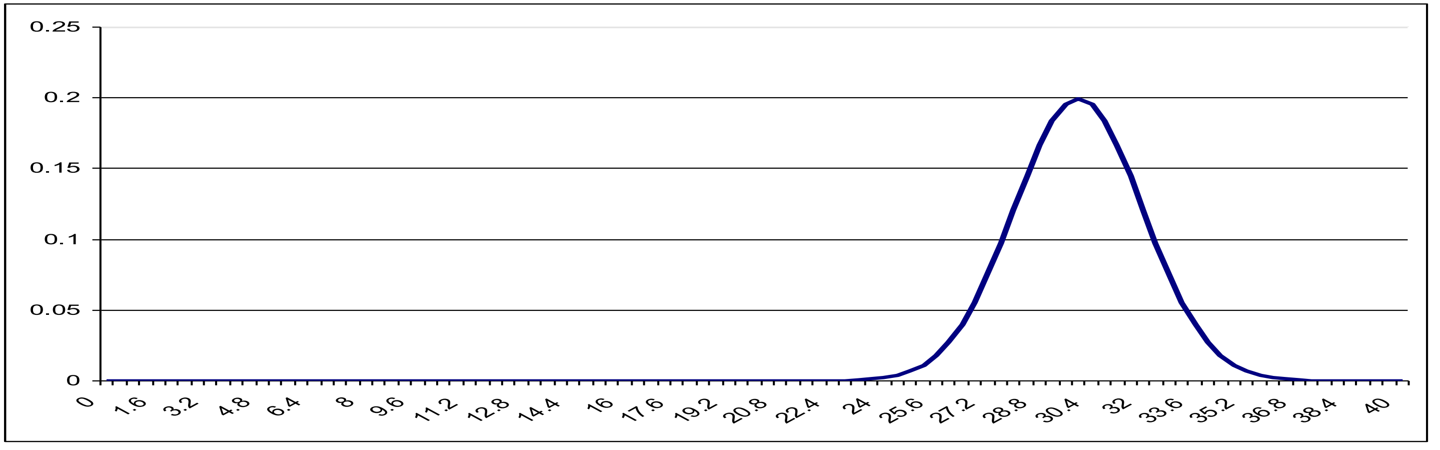
Mean

Mean determines the centre of the curve:

Mean = 10



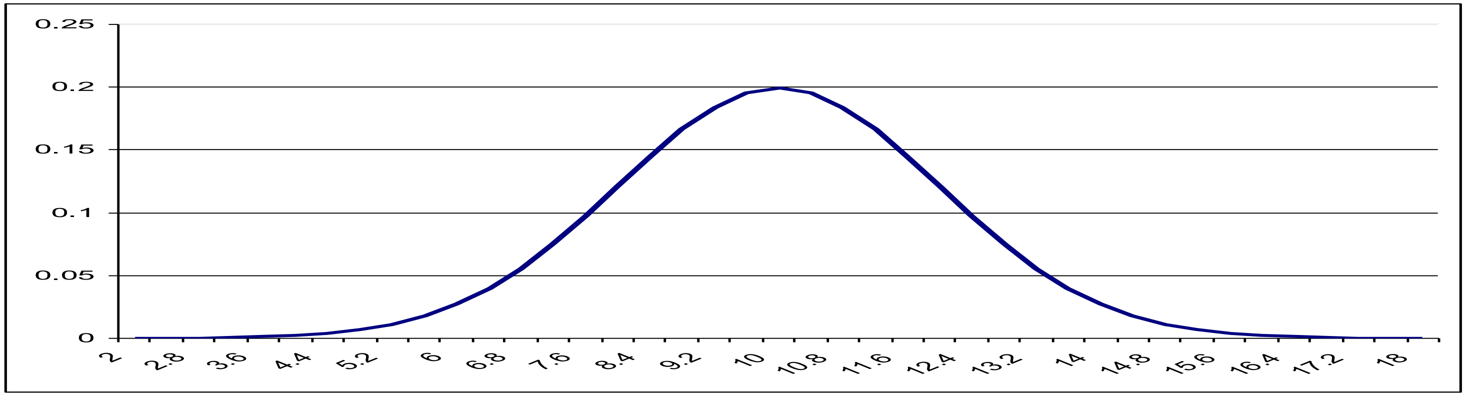
Mean = 30



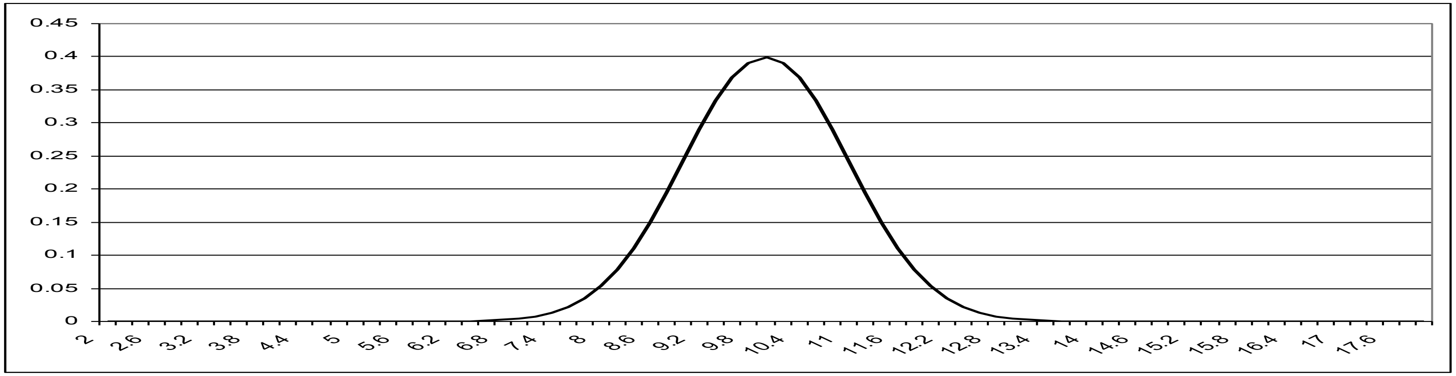
Standard deviation

Standard deviation determines the 'width' of the curve:

Std. Devn. = 2



Std. Devn. = 1



Normal distribution

Given a perfectly normal distribution, mean and standard deviation together determine the function uniquely.

Certain things can be proven about *any* normal distribution

This is why statisticians often assume that their sample space is approximately normally distributed .... even though this is usually very difficult to be sure about!

Example (without proof): In normally distributed data, the probability of lying <u>within 2 standard deviations of the mean,</u> on either side, is approximately <u>0.95</u>

Probability of sample lying within mean ± 2 standard deviations

Example:
Mean (μ)= 10.0
Std. devn (σ)= 2.0

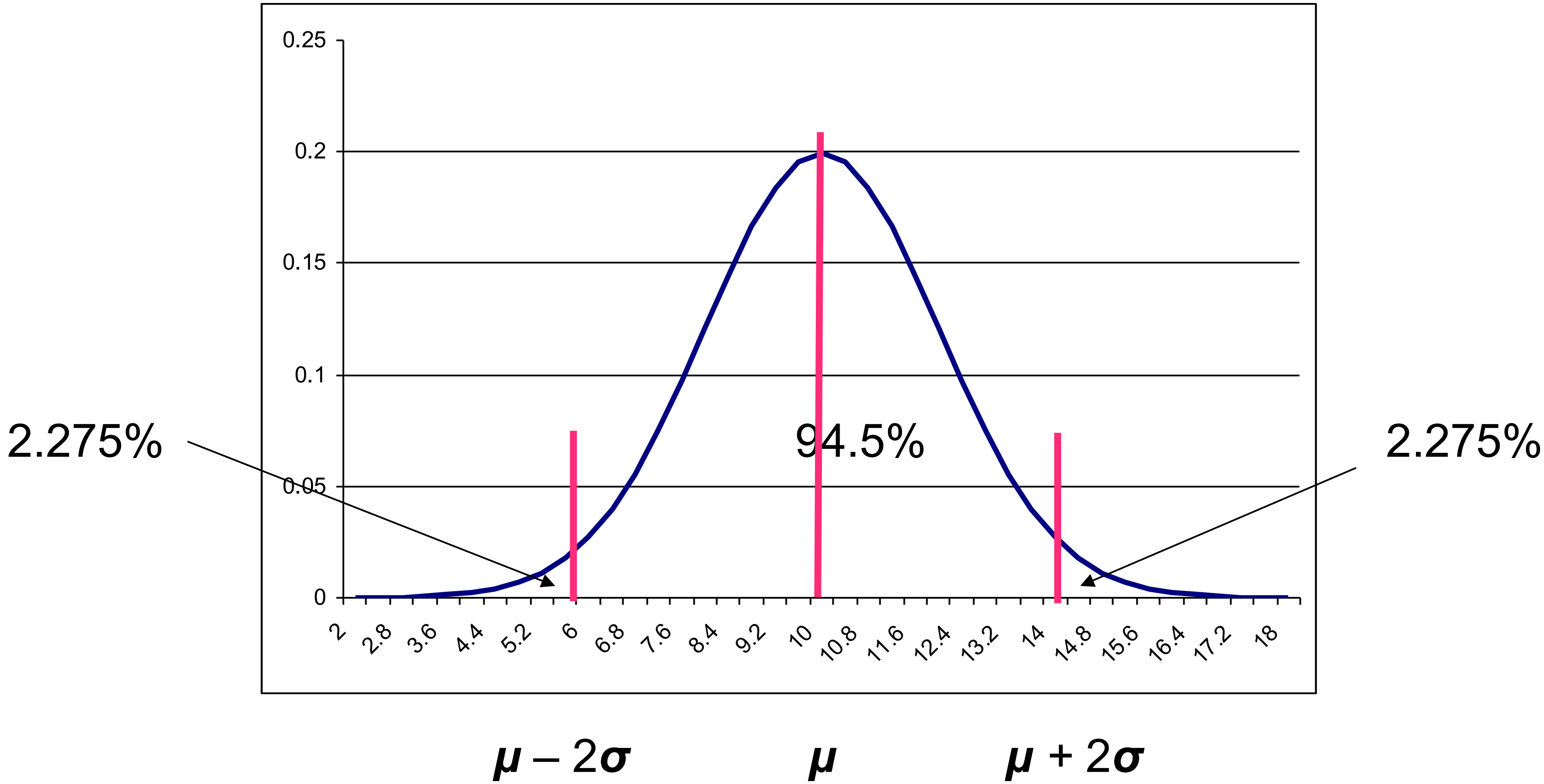
$P(X < \mu - \sigma) = 15.866\%$

$P(X < \mu - 2\sigma) = 2.275\%$

$P(\mu - 2\sigma < X < \mu + 2\sigma) =$
$= (100 - 2 * 2.275)\%$
$= 94.5\%$

x	Prob Dist	Cum Prob Dist
2.0	0.00013383	00.003%
2.4	0.000291947	00.007%
2.8	0.000611902	00.016%
3.2	0.001232219	00.034%
3.6	0.002384088	00.069%
4.0	0.004431848	00.135%
4.4	0.007915452	00.256%
4.8	0.013582969	00.466%
5.2	0.02239453	00.820%
5.6	0.035474593	01.390%
6.0	0.053990967	02.275%
6.4	0.078950158	03.593%
6.8	0.110920835	05.480%
7.2	0.149727466	08.076%
7.6	0.194186055	11.507%
8.0	0.241970725	15.866%
8.4	0.289691553	21.186%
8.8	0.333224603	27.425%
9.2	0.36827014	34.458%
9.6	0.391042694	42.074%
10.0	0.39894228	50.000%

Probability of sample lying within mean ± 2 standard deviations



UNIVERSITY of ABERDEEN

Non-normal distributions

Some non-normal distributions are so close to 'normal' that it doesn't harm much to pretend that they are normal

But some distributions are very far from 'normal'

As a stark reminder, suppose we're considering the variable X= "error when a person's age is stated". Error $=_{def}$ real age – age last birthday

Range(X)= 0 to 12 months

Which of these values is most probable?

Non-normal distributions

Some non-normal distributions are so close to ‘normal’ that it doesn’t harm much to pretend that they normal

But some distributions are very far from ‘normal’

As a stark reminder, suppose we’re considering the variable X= “error when a person’s age is state”. Error $=_{def}$ real age – age last birthday

Range(X)= 0 to 12 months

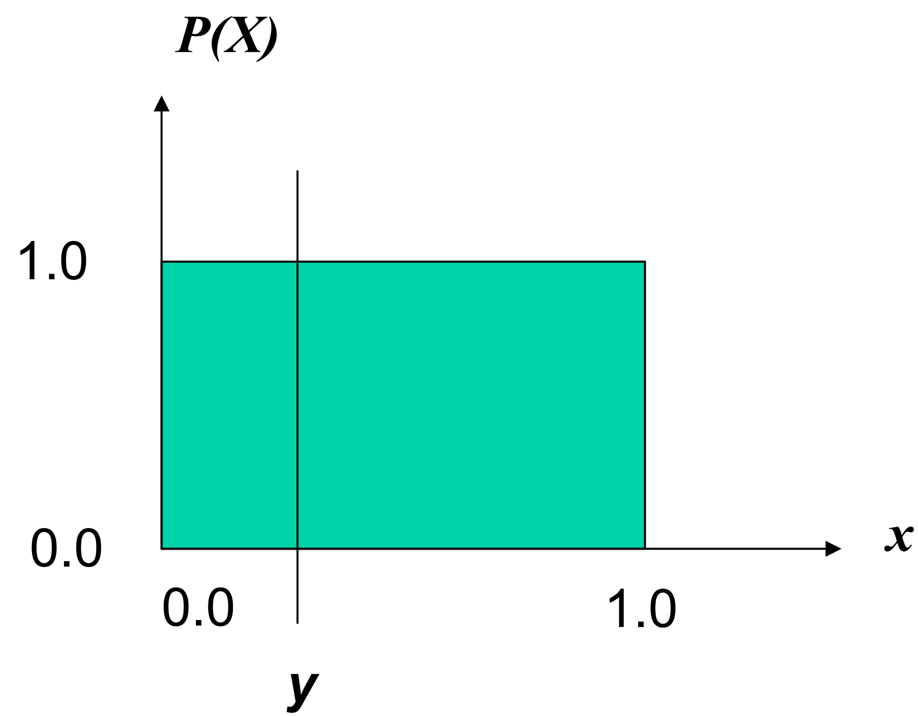
Which of these values is most probable? None: always P(x)=1/12

How would you plot this?

Uniform probability distribution

Also called rectangular distribution

$$P(X < y) = 1.0 * y$$
$$= y$$

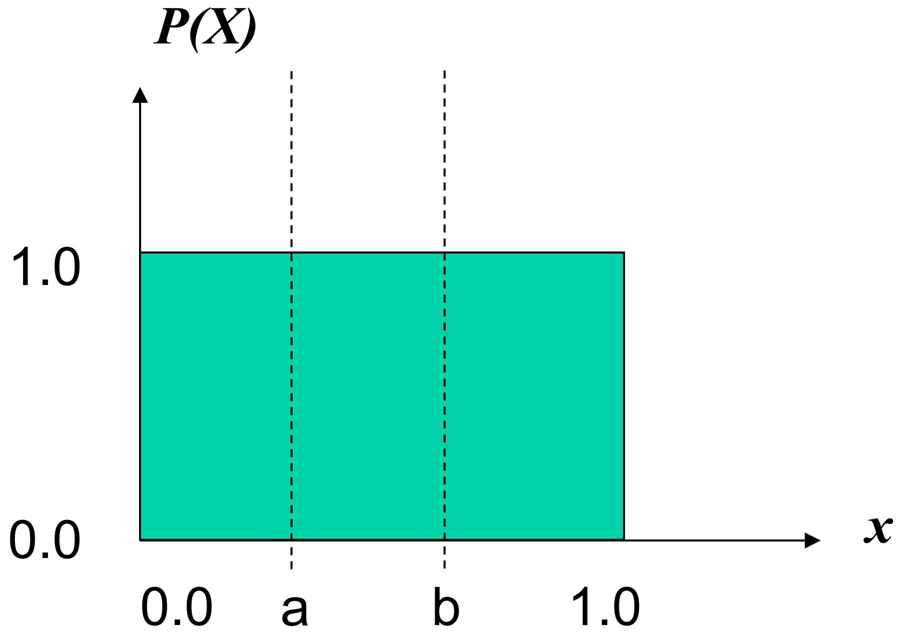


Uniform probability distribution

P(X < a) = a

P(X < b) = b

P(a ≤ X < b) = b - a



Towards Inferential Statistics

If this was a course on inferential statistics, we would now discuss the role of normal distributions in the statistical testing of hypotheses.

The rough idea (in one representative type of situation):
Your hypothesis: Brits are taller than Frenchmen

- Take two samples: heights of Brits, heights of Frenchmen.
- Assume both populations (as a whole) are normally distributed.
- Compute the mean values, μ_b and μ_f , for these two samples.
- Suppose you find $\mu_b > \mu_f$ This is promising …. but it could be an accident!
- Now compute the probability p of measuring these means if, in fact, $\mu_b = \mu_f$ in the two populations as a whole (i.e., your finding was an accident).
- If $p < 0.05$ then conclude that Brits are taller than Frenchmen.

Summing up this lecture

- <u>Discrete</u> random variable
- Benouilli trial; series of Bernouilli trials
- Binomial distribution
- Probability of *r* successes out of *n* trials
- <u>Continuous</u> random variable
- Normal (= Gaussian) distribution