2. A Basic Statistical Toolbox

"**Statistics** is a mathematical science pertaining to the collection, analysis, interpretation, and presentation of data."

Wikipedia definition

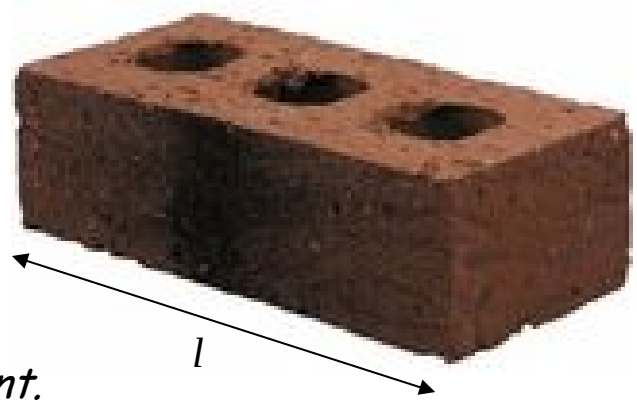
Mathematical statistics: concerned with theoretical foundations (probability theory)

Applied statistics: concerned with modelling of data, and the errors, or uncertainties, in our observations, to make inferences about the physical system we are observing.

Important distinction between random (statistical) error and systematic error.

Statistical error:

Uncertainty in the measurement of a physical quantity that is essentially unpredictable – just as likely to yield a measurement that is too *large* as one that is too *small*.



Example: *Measuring length of the brick in Astronomy 2 asteroid collision experiment.*

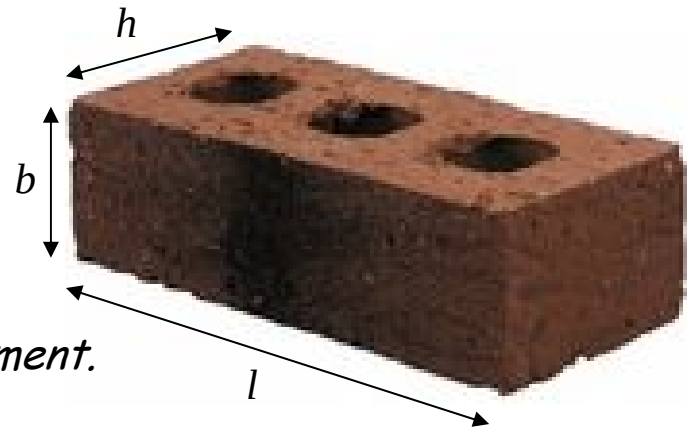
Some students measure brick as slightly longer, others as slightly shorter.

'Common sense' principle: if we repeat our measurements many times and average the results then average length = 'true' length

(We will discuss shortly the basis in probability theory for this common sense result)

Systematic error:

Uncertainty in the measurement of a physical quantity that is always systematically too large or too small. (Measurement is biased)



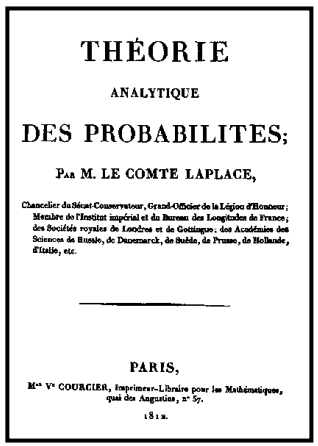
Example: *Measuring **volume** of the brick in Astronomy 2 asteroid collision experiment.*

Suppose we measure the length, breadth and height of the brick, and calculate the volume as $V = lbh$.

This will *always* yield a volume that is systematically too large, because it ignores the fact that the brick has holes.

No matter how often (how accurately) we measure l, b and h we will continue to measure a volume that is too large.

Note that here the systematic error enters not when we make our measurements, but when we analyse them. Systematic flaws in our data analysis methods, rather than in our data themselves, are just as serious (although often easier to fix!)



THÉORIE

ANALYTIQUE

DES PROBABILITES;

PAR M. LE COMTE LAPLACE,

Chancelier du Sénat Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

PARIS,

M^me V^e COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

Laplace (1812)

Mathematical framework for probability as a basis for plausible reasoning:

Probability measures our degree of belief that something is true

$\text{Prob}(X) = 1 \quad \Rightarrow$ we are *certain* that X is true

$\text{Prob}(X) = 0 \quad \Rightarrow$ we are *certain* that X is false

Our degree of belief always depends on the available background information:-

We write $\text{Prob}(X \mid I)$

"Probability that X is true, given I"

Background information

Vertical line denotes conditional probability:

our state of knowledge about X is *conditioned* by background info, I

(We will drop explicit reference to the background information in subsequent equations)

Rules for combining probabilities

$$p(X) + p(\overline{X}) = 1 \qquad (2.1)$$

$\overline{X}$ denotes the proposition that X is false

$$p(X,Y) = p(X \mid Y) \times p(Y) \qquad (2.2)$$

(X,Y) denotes the proposition that X **and** Y are true

$p(X \mid Y)$ = Prob(X is true, given Y is true)

$p(Y)$ = Prob(Y is true, irrespective of X)

In astronomy we generally measure continuous variables which can take on infinitely many values

(e.g. distance, mass, temperature, luminosity, colour etc).

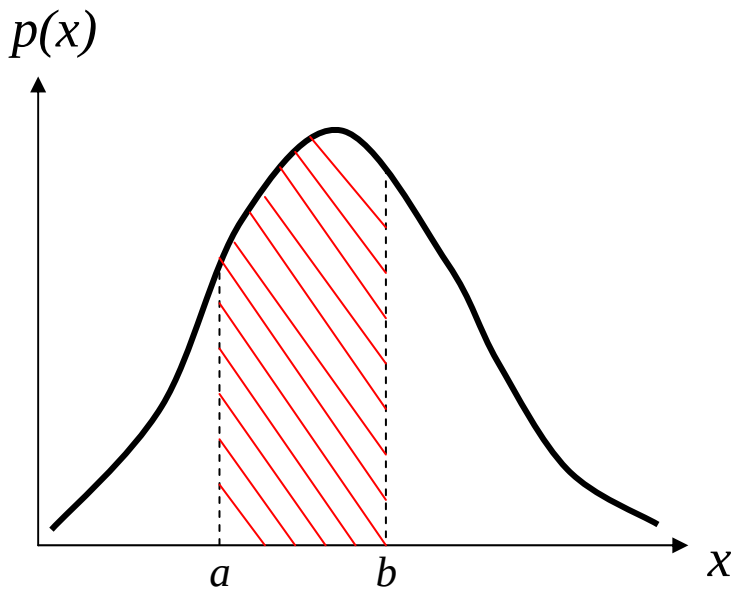
In this case $p(X)$ is no longer a probability, but a *probability density function* $p(x)$

Probabilities are never negative, so $p(x) \geq 0$ for all x

We compute probabilities by measuring the area under the pdf curve, i.e.

$$\text{Prob}(a \leq x \leq b) = \int_a^b p(x)dx \qquad (2.3)$$

'Normalisation' $$\int_{-\infty}^{\infty} p(x)dx = 1 \qquad (2.4)$$



We can also define joint pdfs of two (or more) variables, e.g.

Marginal pdf

$$p(y) = \int_{-\infty}^{\infty} p(x, y)\, dx \qquad (2.5)$$

$$p(x) = \int_{-\infty}^{\infty} p(x, y)\, dy$$

Conditional pdf

$$p(x \mid y) = \frac{p(x, y)}{p(y)} \qquad (2.6)$$

$$p(y \mid x) = \frac{p(x, y)}{p(x)}$$

Statistical Independence

If the conditional pdf of y given x does not depend on x, this means that x and y are statistically independent, since the observed value of y is unaffected by the observed value of x.

Equivalently, x and y are independent if and only if the joint pdf of x and y can be written as the product of their marginal pdfs, i.e.

$$p(x, y) = p(x)\, p(y) \qquad (2.7)$$

Some important pdfs: Discrete case

1) Poisson pdf

e.g. number of photons / second counted by a CCD,

number of galaxies / degree^2 counted by a galaxy survey

(For proof, see non-examinable handout)

r = number of detections

$$p(r) = \frac{\mu^r e^{-\mu}}{r!} \qquad (2.8)$$

Poisson pdf assumes detections are independent, and there is a constant *rate* μ

Can show that $$\sum_{r=0}^{\infty} p(r) = 1 \qquad (2.9)$$

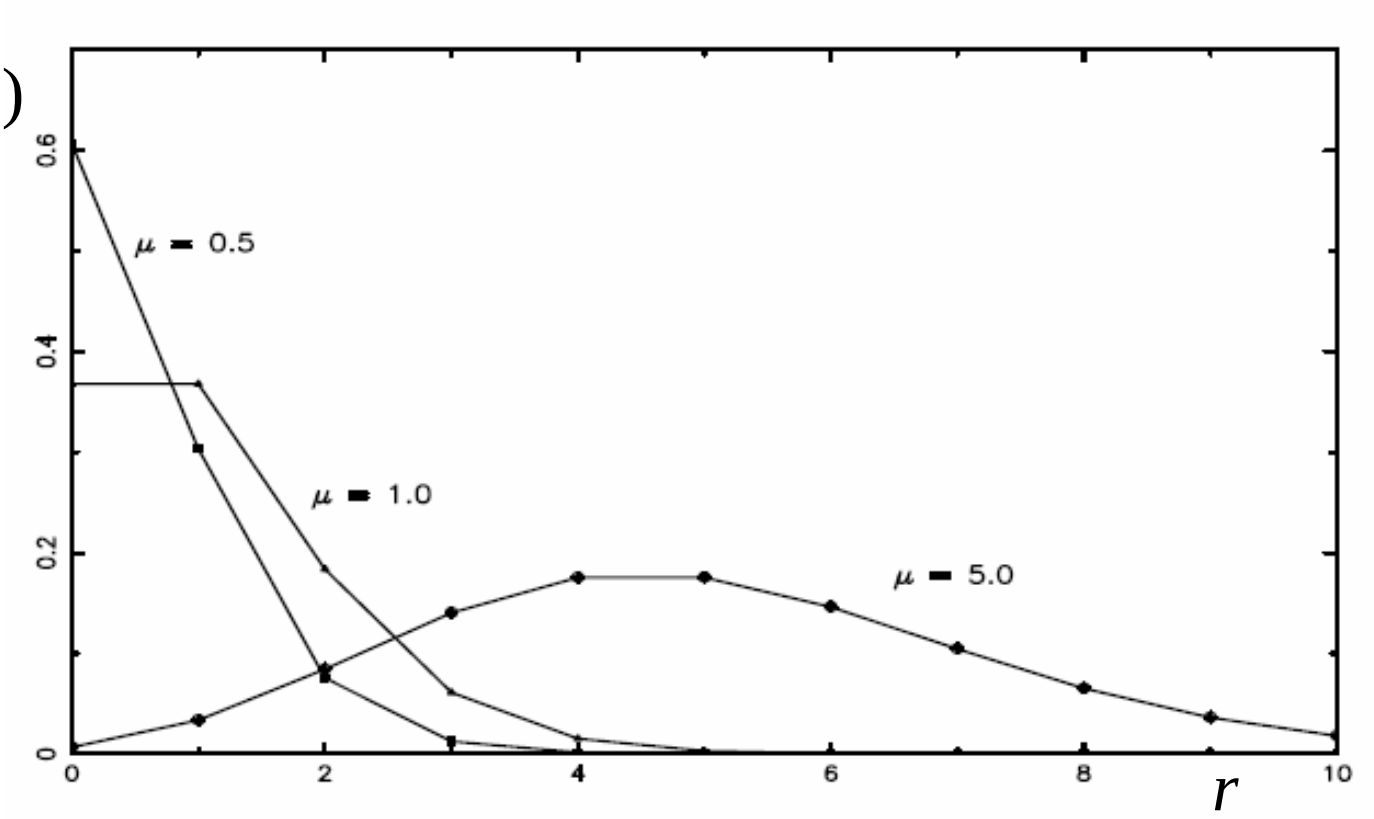
(See examples sheet 1)

Some important pdfs: Discrete case

1) Poisson pdf

e.g. number of photons / second counted by a CCD,

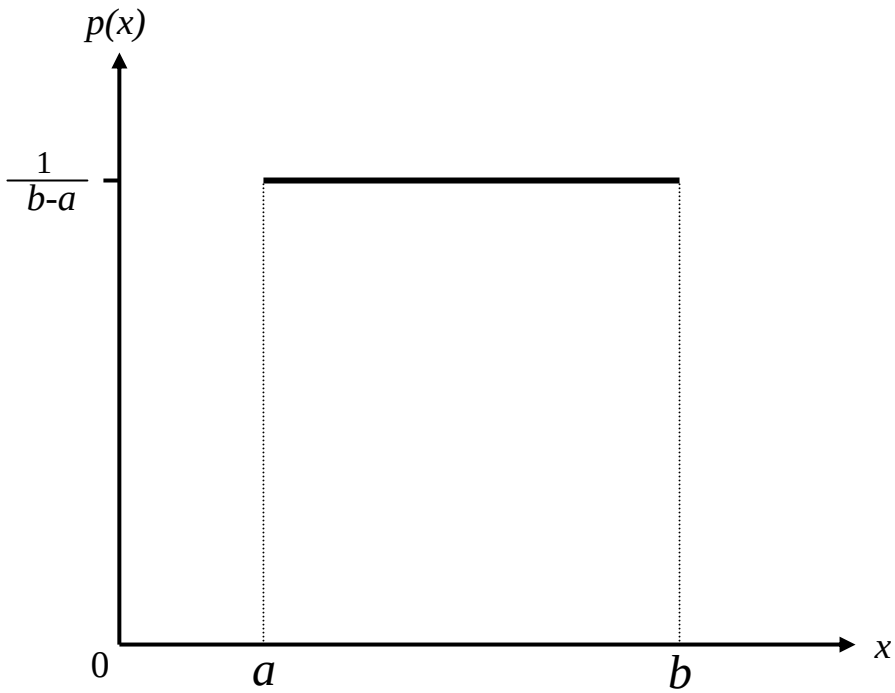
number of galaxies / degree^2 counted by a galaxy survey



Some important pdfs: Continuous case

1) Uniform pdf

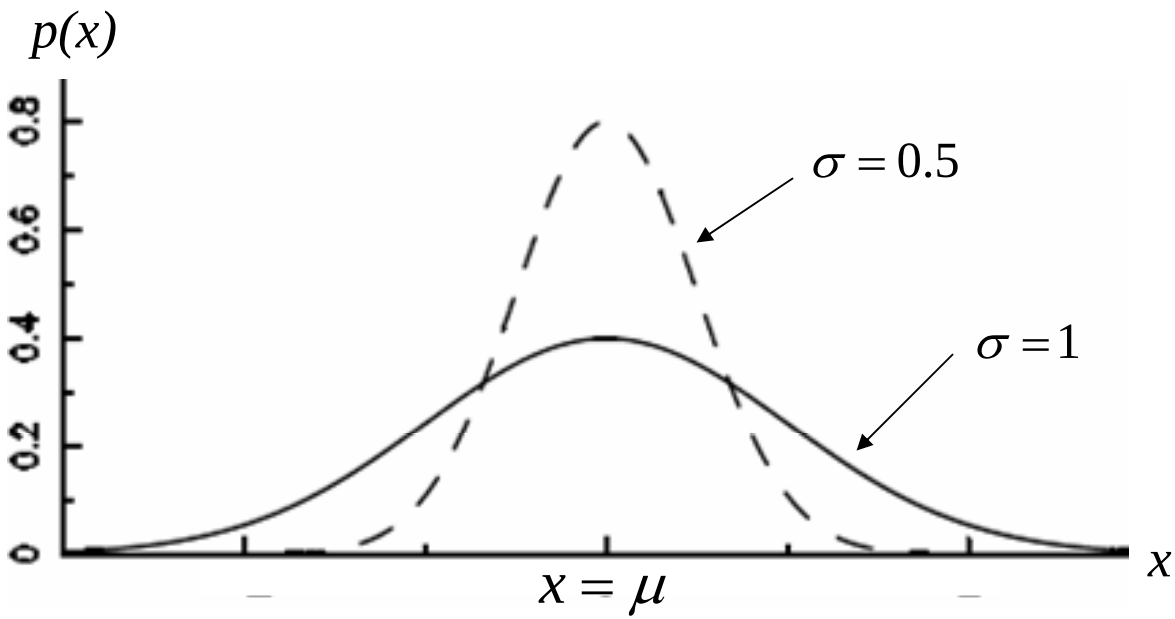
$$p(x) = \begin{cases} \frac{1}{b-a} & a < X < b \\ 0 & \text{otherwise} \end{cases} \qquad (2.10)$$



Some important pdfs: Continuous case

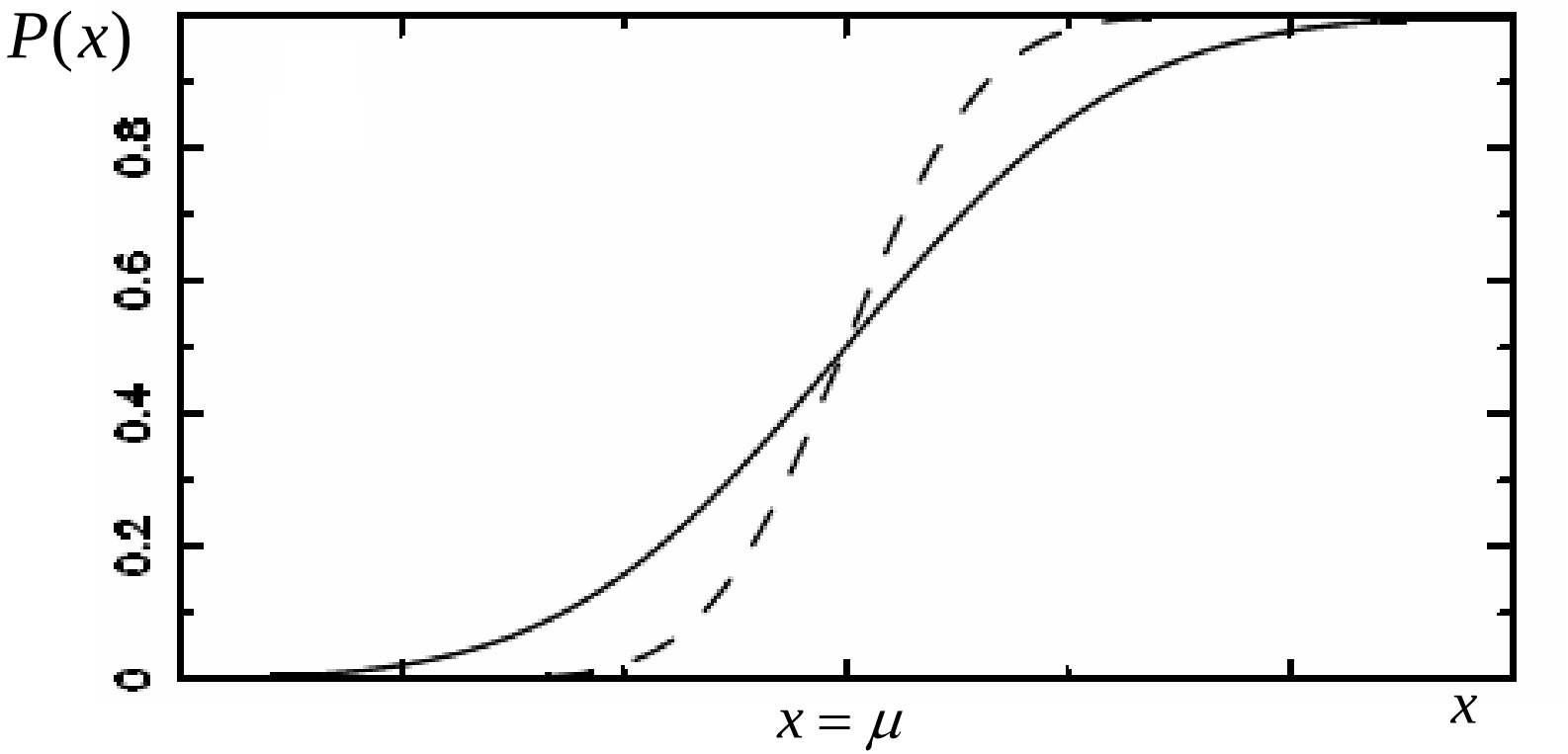
1) Central, or normal pdf
(also known as *Gaussian*)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad (2.11)$$



Cumulative distribution function (CDF)

$$P(a) = \int_{-\infty}^{a} p(x)dx = \text{Prob}(\, x < a \,) \qquad (2.12)$$



Note: $P(-\infty) = 0$ and $P(\infty) = 1$

Measures and moments of a pdf

The nth moment of a pdf is defined as:-

$$\left\langle x^n \right\rangle = \sum_{x=0}^{\infty} x^n \, p(x)$$

Discrete case

(2.13)

$$\left\langle x^n \right\rangle = \int_{-\infty}^{\infty} x^n \, p(x) dx$$

Continuous case

Measures and moments of a pdf

The 1st moment is called the **mean** or **expectation value**:-

$$E(x) = \langle x \rangle = \sum_{x=0}^{\infty} x\, p(x)$$

Discrete case

(2.14)

$$E(x) = \langle x \rangle = \int_{-\infty}^{\infty} x\, p(x) dx$$

Continuous case

Measures and moments of a pdf

The 2nd moment is called the **mean square**:-

$$\left\langle x^2 \right\rangle = \sum_{x=0}^{\infty} x^2 \, p(x) \qquad \text{Discrete case}$$

$$\left\langle x^2 \right\rangle = \int_{-\infty}^{\infty} x^2 \, p(x) dx \qquad \text{Continuous case}$$

(2.15)

Measures and moments of a pdf

The **variance** is defined as:-

$$\mathrm{var}[x] = \sum_{x=0}^{\infty} \left(x - \langle x \rangle\right)^2 p(x) \qquad \text{Discrete case}$$

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x)\,dx \qquad \text{Continuous case} \qquad (2.16)$$

and is often written as σ^2

$\sigma = \sqrt{\sigma^2}$ is called the **standard deviation**

In general

$$\mathrm{var}[x] = \langle x^2 \rangle - \langle x \rangle^2 \qquad (2.17)$$

Measures and moments of a pdf

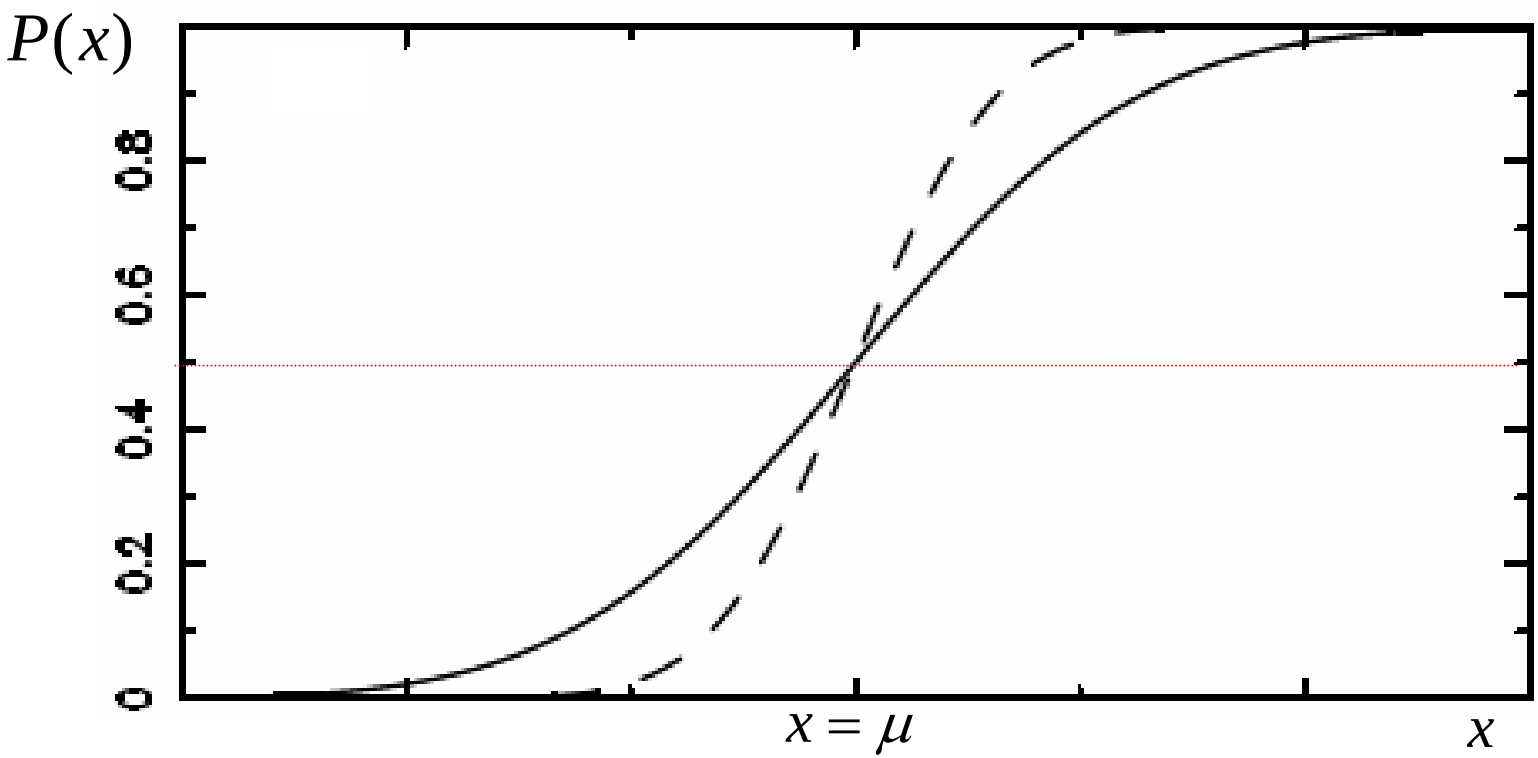
pdf	mean	variance
Poisson $p(r) = \frac{\mu^r e^{-\mu}}{r!}$	μ	μ
Uniform $p(X) = \frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Normal $p(X) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$	μ	σ^2

(See examples sheet 1)

Measures and moments of a pdf

The **Median** divides the CDF into two equal halves

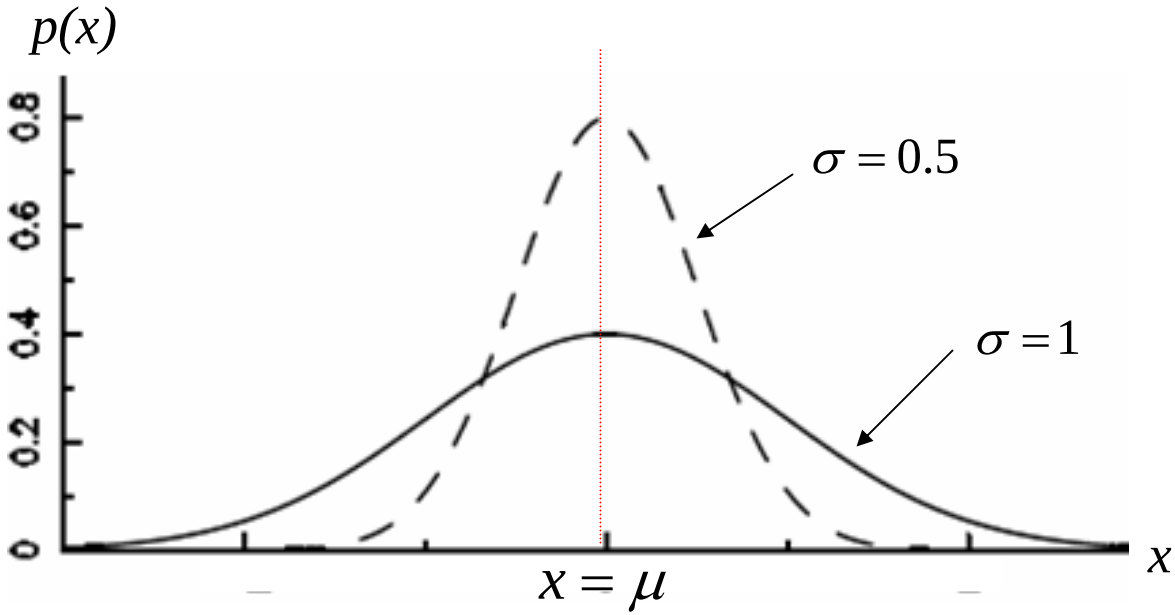
$$P(x_{\text{med}}) = \int_{-\infty}^{x_{\text{med}}} p(x')dx' = 0.5 \qquad (2.18)$$



Prob($x < x_{\text{med}}$) = Prob($x > x_{\text{med}}$) = 0.5

Measures and moments of a pdf

The **Mode** is the value of x for which the pdf is a ***maximum***



For a normal pdf, mean = median = mode = μ

Consider again Laplace's definition of probability:

Probability measures our degree of belief that something is true

This approach to probability theory has recently become very popular again, and is the basis of **Bayesian Probability Theory**

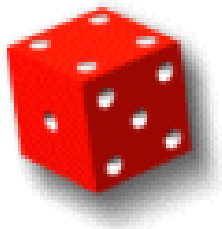
But BPT was rejected for several centuries.

Probability ≡ degree of belief was seen as too subjective

Frequentist approach

Probability = 'long run relative frequency' of an event

in principle can be measured objectively

e.g. rolling a die.  What is $p(1)$?

If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

Can imagine rolling die M times.

Number rolled is a random variable – different outcome each time.

We define $p(1) = \lim_{M \to \infty} \frac{n(1)}{M}$ If $p(1) = \frac{1}{6}$ die is 'fair'

But objectivity is an illusion:

$$p(1) = \lim_{M \to \infty} \frac{n(1)}{M}$$ assumes each outcome equally likely

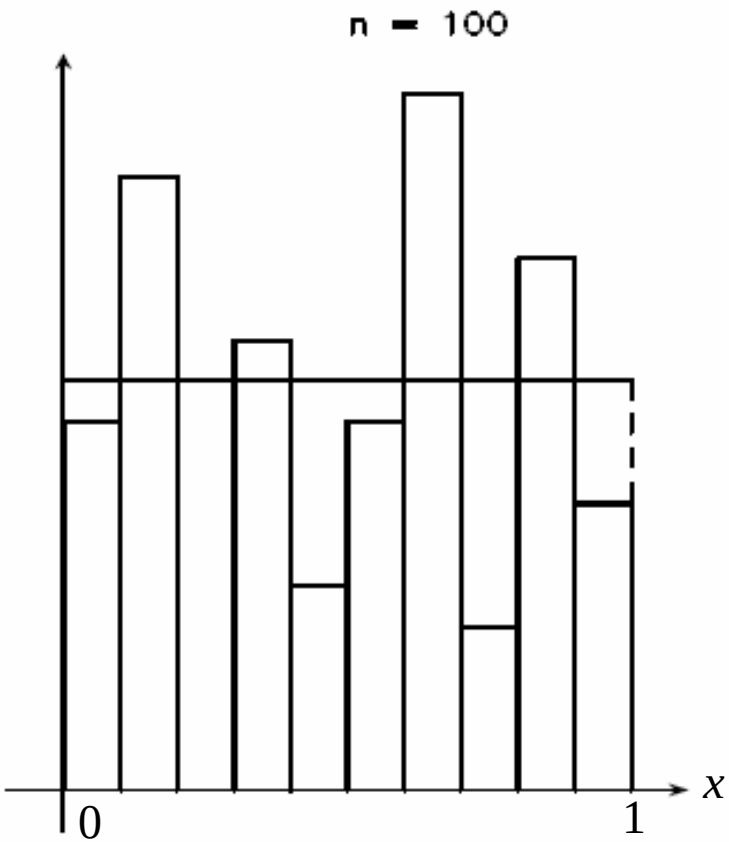
(i.e. equally probable)

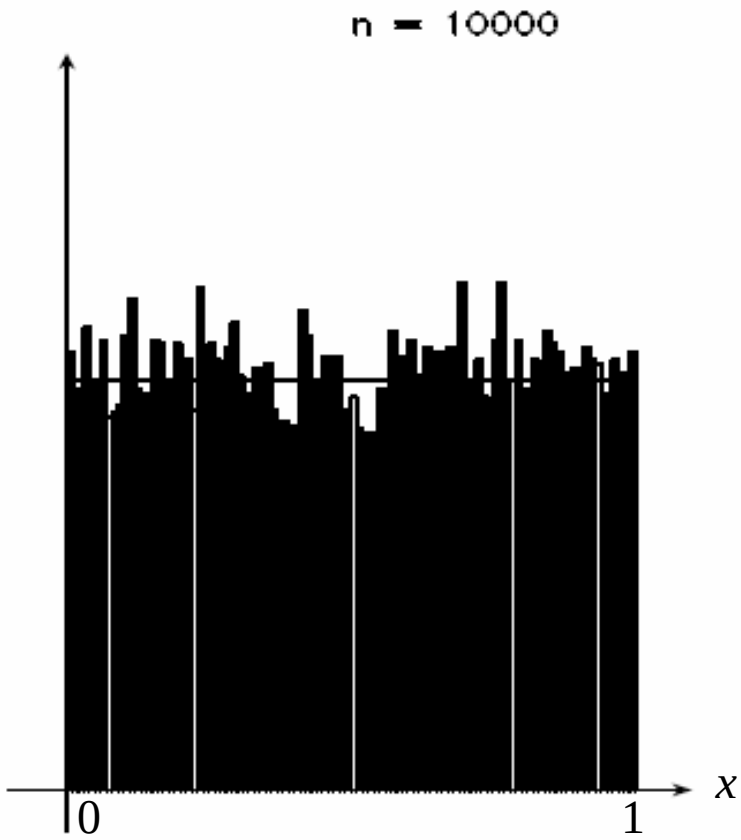
Also assumes infinite series of *identical* trials;

What can we say about the fairness of the die after (say) 5 rolls, or 10, or 100 ?

In the frequentist approach, a lot of mathematical machinery is defined to let us address this type of question:

- **Model underlying population by probability density function**
- **Observed data is a random sample of size M, drawn from pdf**
- **Compute sampling distribution, derived from pdf** (depends on M)
- **Define an *estimator* – function of sample data used to estimate properties of pdf**
- **Carry out Hypothesis test to decide if estimator is 'acceptable' for the given sample size, i.e. to test Goodness of fit of our data.**





Example:

Sample of random numbers from EXCEL random number generator.

These numbers are assumed drawn from U[0,1].

Is the histogram of sampled values sufficiently close to the model pdf?

Examples of Estimators: The Sample Mean

$\{x_1, \ldots, x_M\}$ = random sample from pdf $p(x)$ with mean μ and variance σ^2

μ and σ^2 are fixed (but unknown) parameters

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i = \textbf{sample mean} \qquad (2.19)$$

Can show that $E(\hat{\mu}) = \mu$ ***unbiased estimator*** (2.20)

(No systematic error)

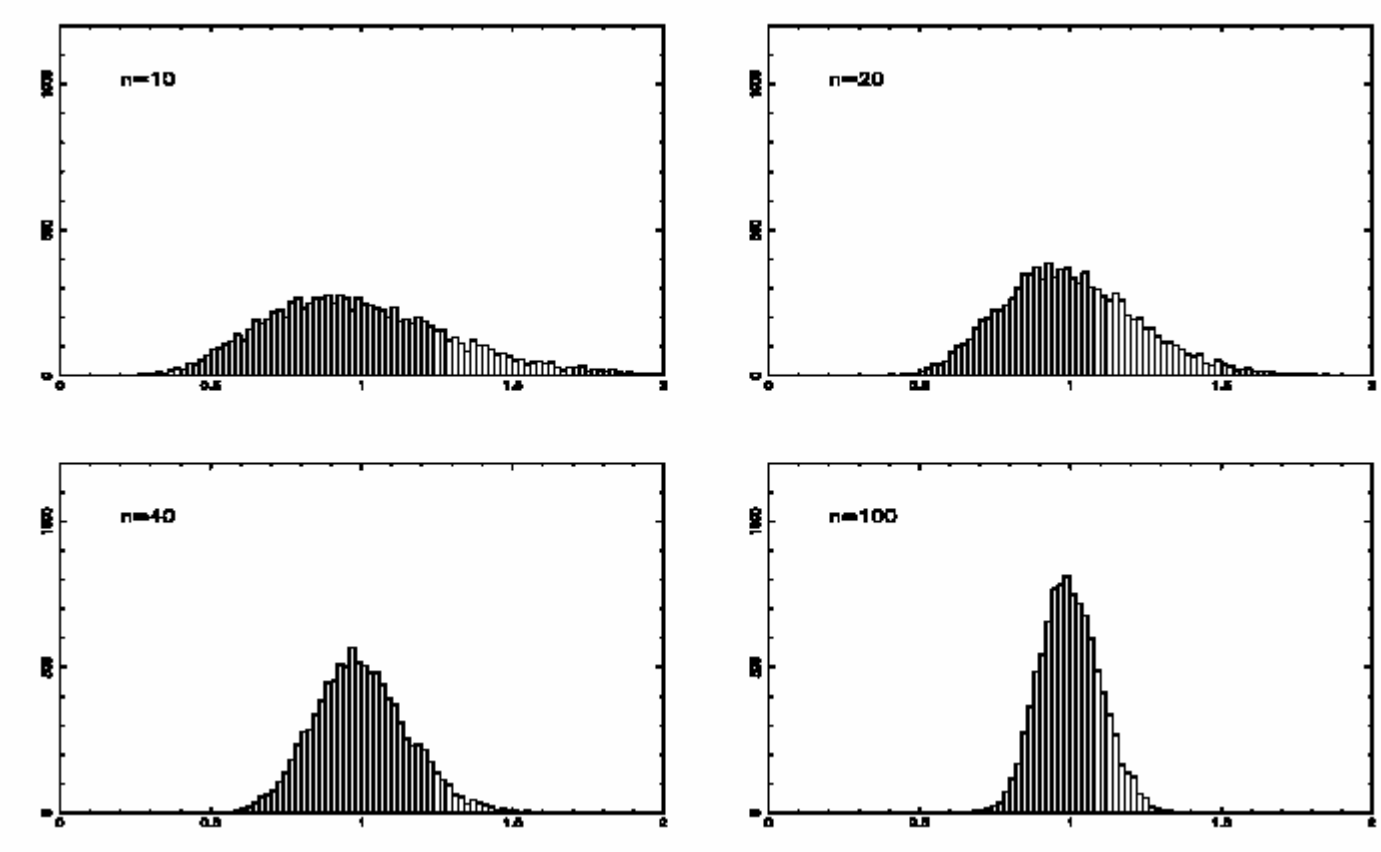
and $\mathrm{var}[\hat{\mu}] = \frac{\sigma^2}{M}$ (2.21)

as sample size increases, sample mean increasingly concentrated near to true mean

The Central Limit Theorem

For *any* pdf with finite variance σ^2, as $M \to \infty$

$\hat{\mu}$ follows a normal pdf with mean μ and variance σ^2 / M



The Central Limit Theorem

Explains importance of normal pdf in statistics.

But still based on asymptotic behaviour of an infinite ensemble of samples that we didn't actually observe!

A Bayesian approach to probability avoids some of these issues:

- No philosophical distinction between observed data (random variables) and pdf model parameters.
- Can make inferences about our model parameters using

  Bayes formula

Recall eqn. (2.6) for conditional pdf of y given x

$$p(y|x) = \frac{p(y|x) \times p(y)}{p(x)}$$

Thomas Bayes
(1702 – 1761 AD)

This is known as Bayes' formula.
We can re-write it in the following form:

Posterior

Likelihood

Prior

$$p(\text{model}|\text{data}, I) = \frac{p(\text{data}|\text{model}, I) \times p(\text{model}|I)}{p(\text{data}|I)}$$

(2.22)

Evidence

Recall eqn. (2.6) for conditional pdf of y given x

$$p(y \mid x) = \frac{p(y \mid x) \times p(y)}{p(x)}$$

Thomas Bayes
(1702 – 1761 AD)

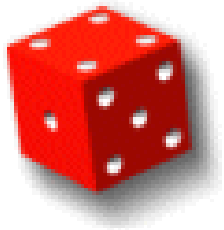
This is known as Bayes' formula.

Or, equivalently:

Posterior — Likelihood — Prior

$$p(\text{model} \mid \text{data}, I) \propto p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)$$

What we know now — Influence of our observations — What we knew before



Consider again our example of rolling a die, and trying to decide if it is fair.

In the Bayesian approach, we can test our model, in the light of our data (i.e. rolling the die) and see how our degree of belief in its 'fairness' evolves, for any sample size, considering only the data that we *did* actually observe

Posterior — Likelihood — Prior

$$p(\text{model}\,|\,\text{data}, I) \quad \propto \quad p(\text{data}\,|\,\text{model}, I) \times p(\text{model}\,|\,I)$$

What we know now — Influence of our observations — What we knew before

Deeper discussion contrasting the Bayesian vs Frequentist approach will be considered in MSci course.

For some practical illustrations of the Bayesian approach, see the example sheets.