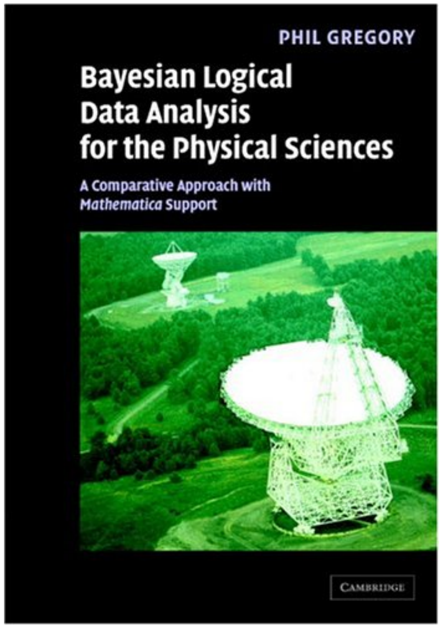
1. Introduction and Theoretical Foundations

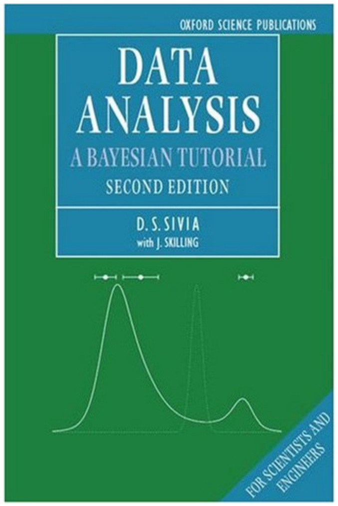
Reasonable thinking?...



PREFACE

The goal of science is to unlock nature's secrets…Our understanding comes through the development of theoretical models capable of explaining the existing observations as well as making testable predictions…Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainities. These topics are commonly referred to as "data analysis".

The most we can hope to do is to make the best inference based on the experimental data and any prior knowledge that we have available.



Reasonable thinking?...

"A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences"

Herodotus, c.500 BC

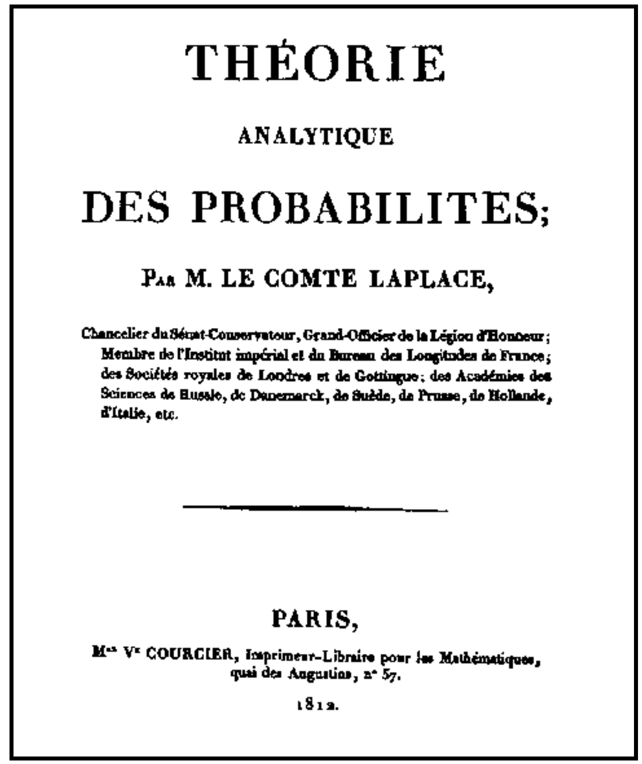
Reasonable thinking?...

"A decision was wise, even though it led to disastrous consequences, if the evidence at hand indicated it was the best one to make; and a decision was foolish, even though it led to the happiest possible consequences, if it was unreasonable to expect those consequences"

Herodotus, c.500 BC

Pierre-Simon Laplace
(1749 – 1827)

"Probability theory is nothing but common sense reduced to calculation"



THÉORIE

ANALYTIQUE

DES PROBABILITES;

Par M. LE COMTE LAPLACE,

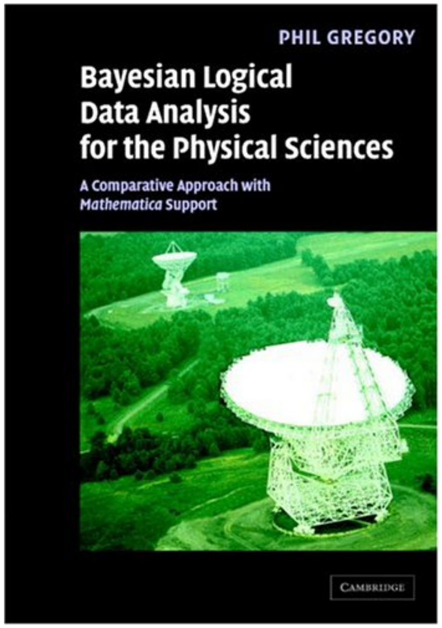
Chancelier du Sénat-Conservateur, Grand-Officier de la Légion d'Honneur; Membre de l'Institut impérial et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, de Hollande, d'Italie, etc.

PARIS,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, nº 57.

1812.

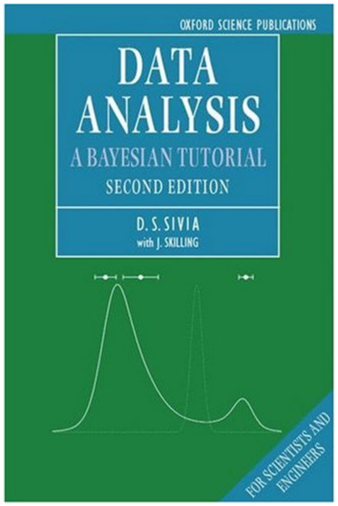
Plausible reasoning?...



PREFACE

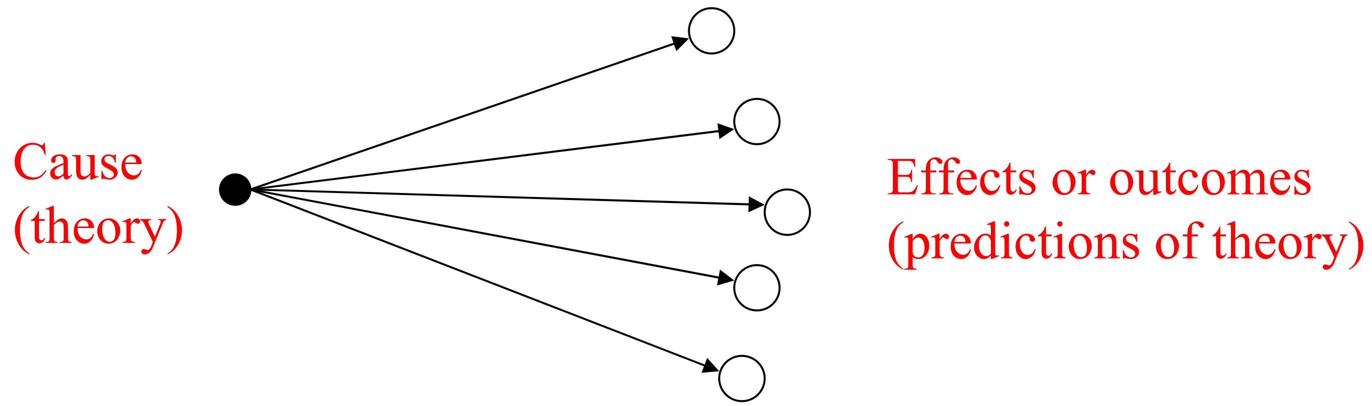
The goal of science is to unlock nature's secrets…Our understanding comes through the development of theoretical models capable of explaining the existing observations as well as making testable predictions…Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainities. These topics are commonly referred to as "data analysis".

The most we can hope to do is to make the best inference based on the experimental data and any prior knowledge that we have available.

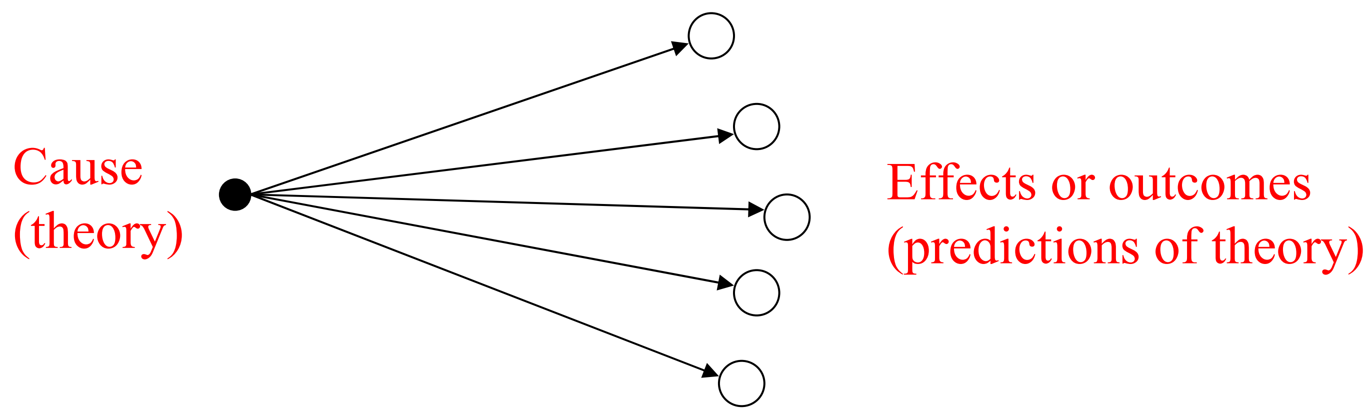


We need to think about the difference between **deductive** and **inductive** logic

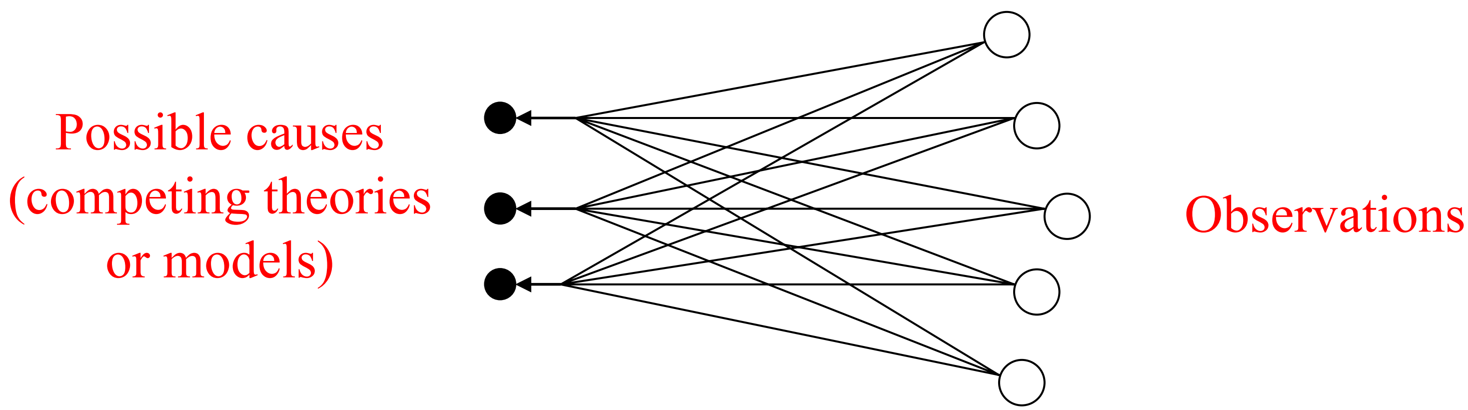
Deductive logic



Deductive logic

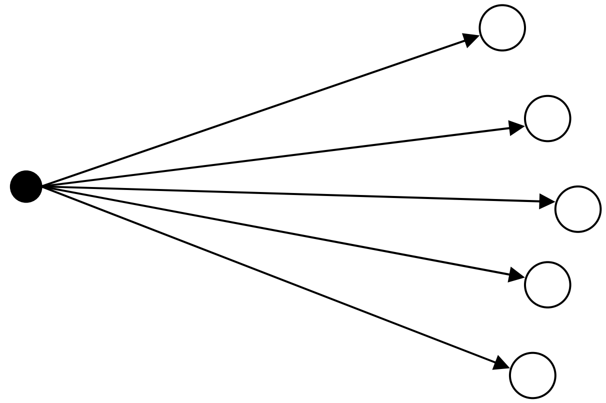


Inductive logic



Deductive logic

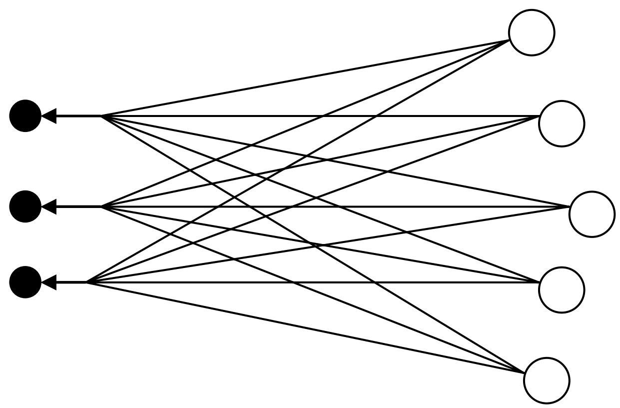
Cause
(theory)



Effects or outcomes
(predictions of theory)

Inductive logic

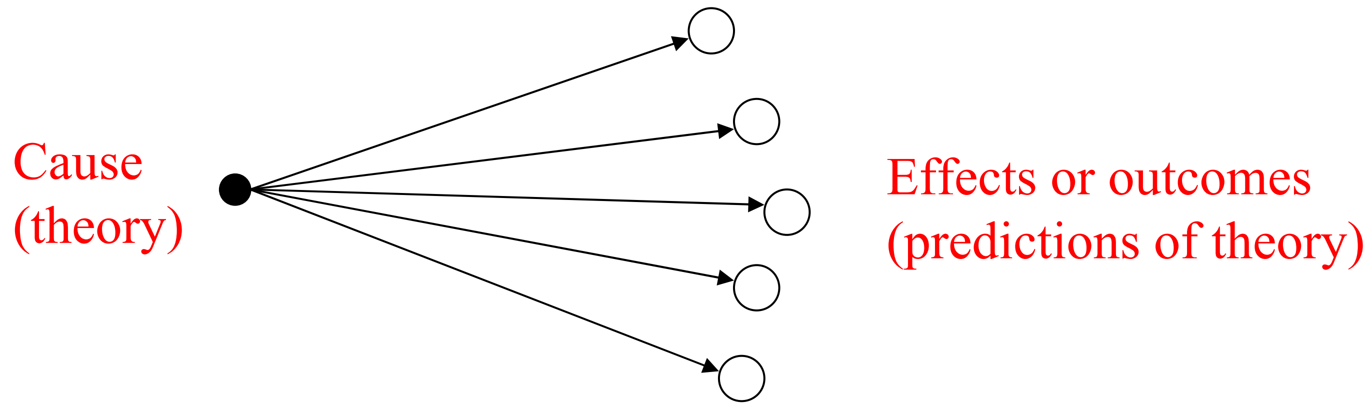
Possible causes
(competing theories
or models)



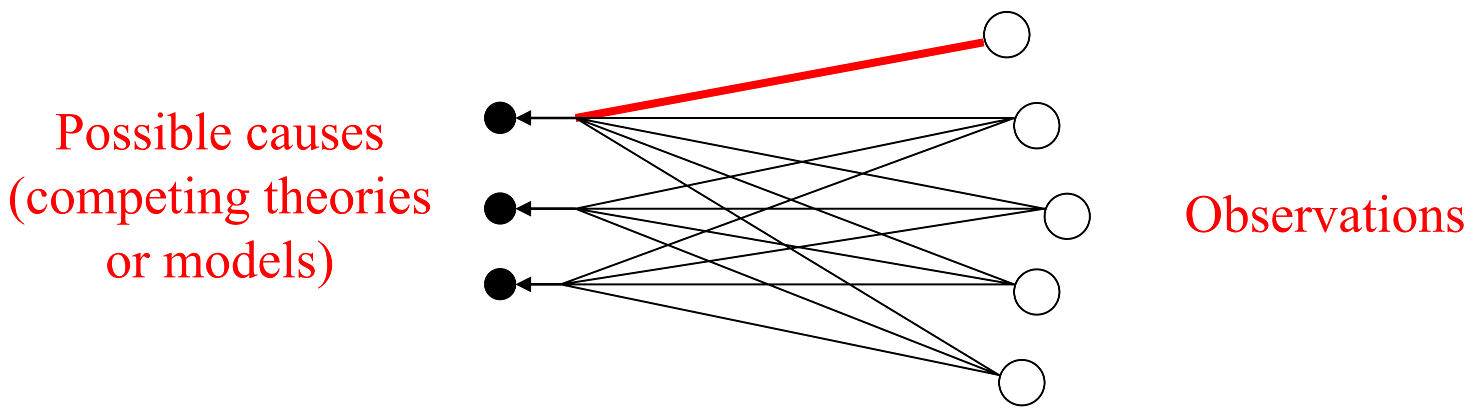
Observations

How do we decide which model is most plausible?

Deductive logic



Inductive logic



How do we decide which model is most **plausible**?

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

Let's suppose that A is true. (Our theory).

- If B is true, then C is true
- If C is false, then B is false

An example of deductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

Let's suppose that A is true. (Our theory).

- If B is true, then C is true
- If C is false, then B is false

C is a logical consequence of A and B

If we set 'true' = 1 and 'false' = 0, we can use the rules of George Boole (1854) to carry out logical operations.
We define

Negation:	$\overline{A}$	'A is false'
Logical product:	AB	'both A and B are true'
Logical sum:	$A+B$	'at least one of A or B is true'

Then

$$A(B+C) = AB + AC$$

$$A + AB = A$$

$$A + \overline{A} = 1$$

$$A + BC = (A+B)(A+C)$$

$$A\overline{A} = 0$$ etc

An example of inductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

What can we say about B if A and C are true?...

(Statement A didn't say that all students who drink Irn Bru have red hair)

An example of inductive logic

Statement A:	All red-haired students drink Irn Bru
Statement B:	Student X has red hair
Statement C:	Student X drinks Irn Bru

What can we say about B if A and C are true?...

(Statement A didn't say that all students who drink Irn Bru have red hair)

We might say, however

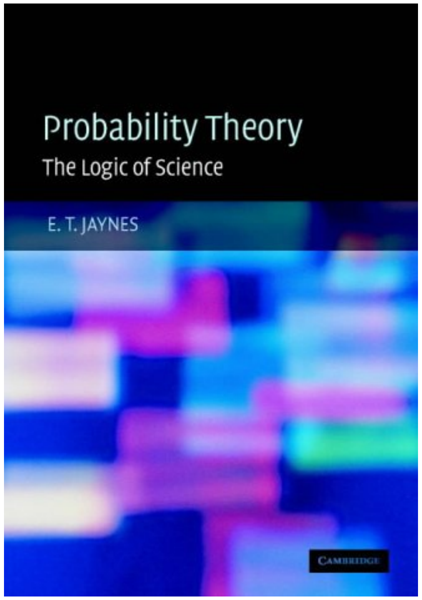
- If C is true, then B is more plausible

In the 1940s and 50s Cox, Polya and Jaynes formalised the mathematics of inductive logic as plausible reasoning

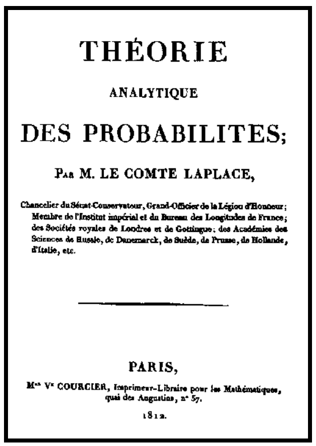
- If we assign degrees of plausibility a real number between 0 and 1, then the rules for combining and operating on inductive logical statements are identical to those for deductive logic $\longrightarrow$ Boolean algebra.

In the 1940s and 50s Cox, Polya and Jaynes formalised the mathematics of inductive logic as plausible reasoning

- If we assign degrees of plausibility a real number between 0 and 1, then the rules for combining and operating on inductive logical statements are identical to those for deductive logic $\longrightarrow$ Boolean algebra.



Ed Jaynes
(1922 - 1998)



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,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

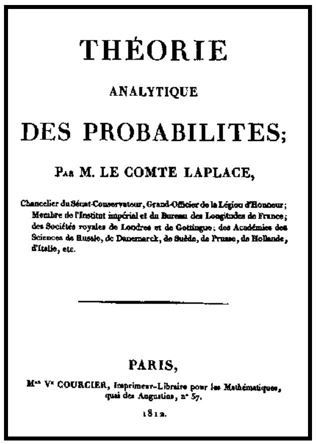
Laplace (1812)

Mathematical framework for probability
as a basis for plausible reasoning:

> Probability measures our degree of belief that something is true

Laplace (1812)

Mathematical framework for probability as a basis for plausible reasoning:



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,

Mme Ve COURCIER, Imprimeur-Libraire pour les Mathématiques, quai des Augustins, n° 57.

1812.

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 $\boxed{\mathrm{Prob}(\, X \mid I \,)}$ "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

Rules for combining probabilities

$$\boxed{p(X \mid I) \quad + \quad p(\bar{X} \mid I) \quad = \quad 1}$$

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

Note: the background information is the *same* in both cases

Rules for combining probabilities

$$\boxed{p(X,Y\,|\,I) \quad = \quad p(X\,|\,Y,I)\times p(Y\,|\,I)}$$

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

Rules for combining probabilities

$$\boxed{p(X,Y \mid I) \quad = \quad p(X \mid Y,I) \times p(Y \mid I)}$$

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

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

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

Also

$$\boxed{p(Y,X \mid I) \quad = \quad p(Y \mid X,I) \times p(X \mid I)}$$

but

$$p(Y,X \mid I) \quad = \quad p(X,Y \mid I)$$

Hence

$$\boxed{p(Y \mid X,I) \quad = \quad \frac{p(X \mid Y,I) \times p(Y \mid I)}{p(X \mid I)}}$$

Bayes' theorem:

$$p(Y \mid X, I) = \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Laplace rediscovered work of
Rev. Thomas Bayes (1763)

Bayesian Inference

Thomas Bayes
(1702 – 1761 AD)

Bayes' theorem:

$$p(Y \mid X, I) \quad = \quad \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

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

Bayes' theorem:

$$p(Y \mid X, I) \quad = \quad \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}$$

Posterior

Likelihood

Prior

$$\underbrace{p(\text{model} \mid \text{data}, I)}_{\text{Posterior}} \quad = \quad \frac{\overbrace{p(\text{data} \mid \text{model}, I)}^{\text{Likelihood}} \times \overbrace{p(\text{model} \mid I)}^{\text{Prior}}}{\underbrace{p(\text{data} \mid I)}_{\text{Evidence}}}$$

Evidence

Bayes' theorem:

$$\boxed{p(Y \mid X, I) \quad = \quad \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}}$$

Posterior

Likelihood

Prior

$$\boxed{p(\text{model} \mid \text{data}, I) \quad = \quad \frac{p(\text{data} \mid \text{model}, I) \times p(\text{model} \mid I)}{p(\text{data} \mid I)}}$$

Evidence

We can calculate these terms

Bayes' theorem:

$$\boxed{p(Y \mid X, I) \quad = \quad \frac{p(X \mid Y, I) \times p(Y \mid I)}{p(X \mid I)}}$$

Posterior

Likelihood

Prior

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

What we know now

Influence of our observations

What we knew before

Bayesian probability theory is simultaneously a very old and a very young field:-

Old : original interpretation of Bernoulli, Bayes, Laplace…

Young: 'state of the art' in data analysis

But BPT was rejected for several centuries.

Bayesian probability theory is simultaneously a very old and a very young field:-

Old : original interpretation of Bernoulli, Bayes, Laplace…

Young: 'state of the art' in data analysis

But BPT was rejected for several centuries.

Probability $\equiv$ degree of belief was seen as too subjective

Frequentist approach

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

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

Probability = 'long run relative frequency' of an event

in principle, it was thought, 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.

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

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

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

These probabilities are fixed (but unknown) numbers.

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

Probability = 'long run relative frequency' of an event

in principle, it was thought, 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.

But objectivity is an illusion:

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

(i.e. equally probable)

Probability = 'long run relative frequency' of an event

in principle, it was thought, 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.

But objectivity is an illusion:

Also assumes infinite series of *identical* trials;

why can't probabilities change?

Probability = 'long run relative frequency' of an event

in principle, it was thought, can be measured objectively

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

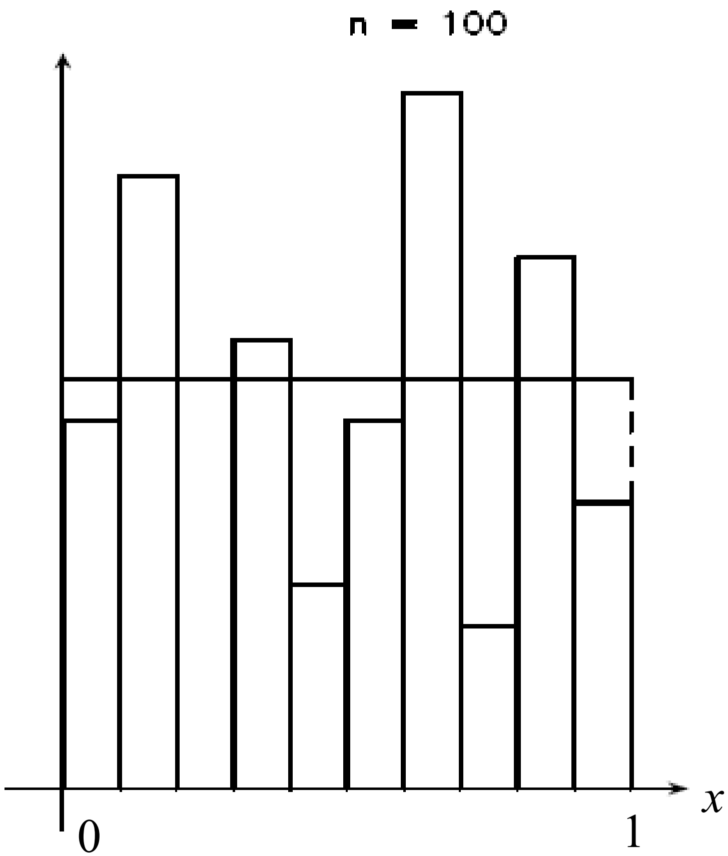
If die is 'fair' we expect $p(1) = p(2) = \ldots = p(6) = \frac{1}{6}$

These probabilities are fixed (but unknown) numbers.

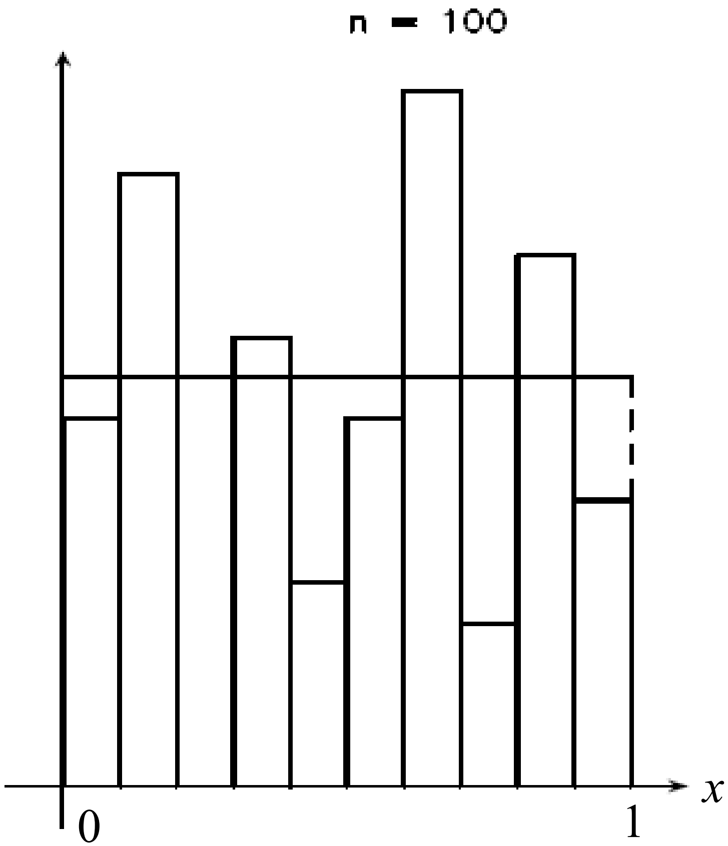
But objectivity is an illusion:

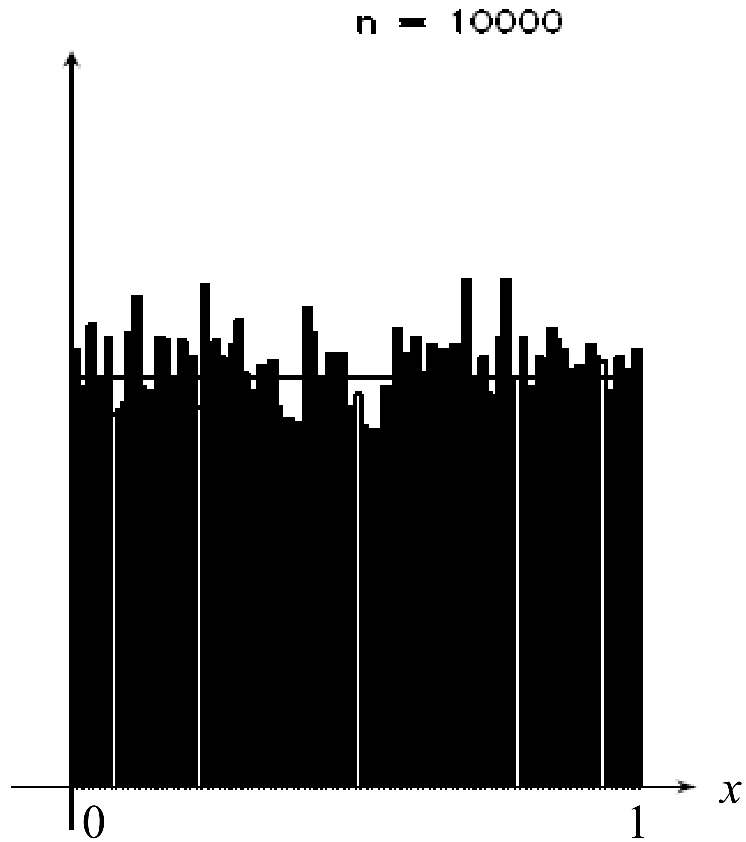
What can we say about the fairness of the die after (say) 5 rolls, or 10, or 100 ?

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



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



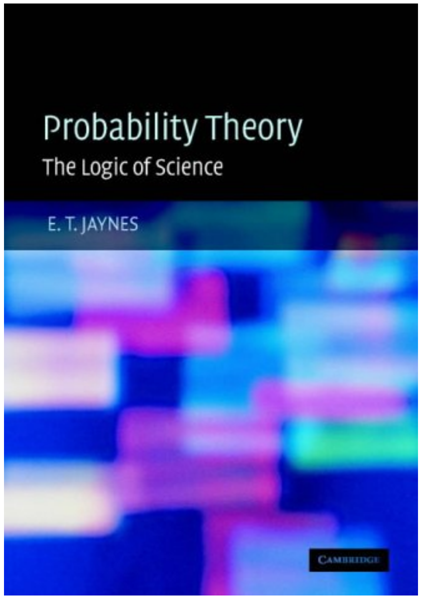


Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!!!

Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!!!

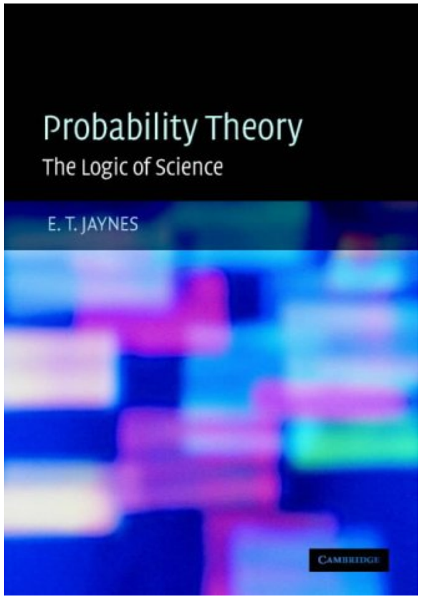


Ed Jaynes
(1922 – 1998)

Probability *is* subjective; it depends on the available information

Bayesian versus Frequentist statistics: Who is right?

Frequentists are correct to worry about subjectiveness of assigning probabilities – Bayesians worry about this too!!!



Ed Jaynes
(1922 – 1998)

Probability *is* subjective;
it depends on the available information

Subjective ≠ arbitrary

Given the *same* background information, two observers should assign the *same* probabilities

More Theoretical Foundations: Marginalisation

Suppose there are a set of M propositions $\{x_k : k = 1, \ldots, M\}$

Then $$\sum_{k=1}^{M} p(x_k \mid I) = 1$$

More Theoretical Foundations: Marginalisation

Suppose there are a set of M propositions $\{x_k : k = 1, \ldots, M\}$

Then $$\sum_{k=1}^{M} p(x_k \mid I) \;=\; 1$$

Suppose we introduce some additional proposition Y

Use Bayes' theorem. $$p(x_1, y \mid I) = p(x_1 \mid y, I)\, p(y \mid I)$$

$$\vdots$$

$$p(x_M, y \mid I) = p(x_M \mid y, I)\, p(y \mid I)$$

More Theoretical Foundations: Marginalisation

Then $$\sum_{k=1}^{M} p(x_k, y \mid I) = \underbrace{\left[\sum_{k=1}^{M} p(x_k \mid y, I) \right]}_{=\ 1} p(y \mid I)$$

More Theoretical Foundations: Marginalisation

Then

$$\sum_{k=1}^{M} p(x_k, y \mid I) = \underbrace{\left[\sum_{k=1}^{M} p(x_k \mid y, I) \right]}_{=\ 1} p(y \mid I)$$

Marginal probability

$$\boxed{p(y \mid I) = \sum_{k=1}^{M} p(x_k, y \mid I)}$$

More Theoretical Foundations: Marginalisation

This extends to the *continuum limit* :

x can take infinitely many values

$$\boxed{p(y \mid I) = \int_{-\infty}^{\infty} p(x, y \mid I)\, dx}$$

More Theoretical Foundations: Marginalisation

This extends to the *continuum limit* :

x can take infinitely many values

$$\boxed{p(y \mid I) = \int_{-\infty}^{\infty} p(x, y \mid I)\, dx}$$

$p(x, y \mid I)$ is no longer a probability, but a *probability density*

$$\boxed{\mathrm{Prob}(a \leq x \leq b \text{ and } y \text{ is true} \mid I) = \int_a^b p(x, y \mid I) dx}$$

with obvious extension to continuum limit for y

More Theoretical Foundations: Marginalisation

This extends to the *continuum limit* :

x can take infinitely many values

$$\boxed{p(y \mid I) = \int_{-\infty}^{\infty} p(x, y \mid I)\, dx}$$

Also
$$\int_{-\infty}^{\infty} p(x \mid y, I) dx \quad = \quad 1$$

Normalisation condition

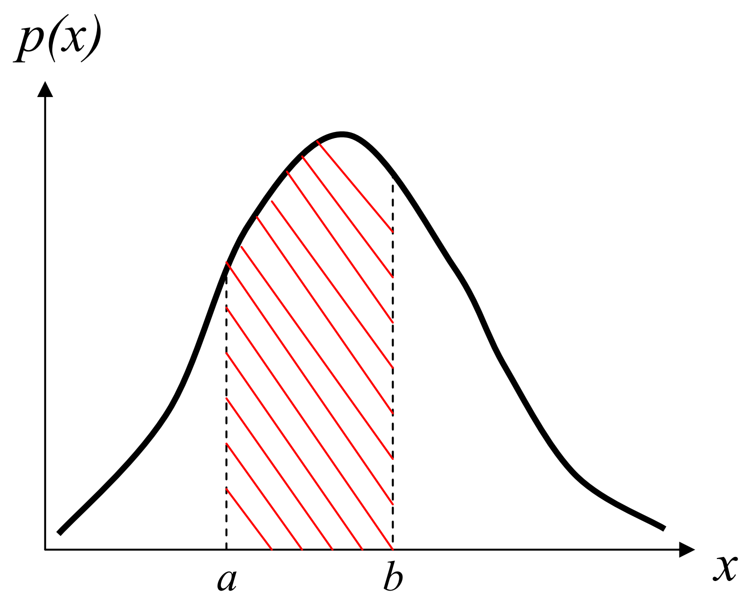
More Theoretical Foundations: Marginalisation

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.

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

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



Some important pdfs: Discrete case

1) Poisson pdf

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

number of galaxies / degree^2 counted by a survey

r = number of detections

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

BBC News - The first glimpse of dark matter? - Mozilla Firefox

http://news.bbc.co.uk/1/hi/8420089.stm

BBC | Low graphics | Help | Search | Explore the BBC

NEWS | LIVE BBC NEWS CHANNEL

- News Front Page
- World
- UK
- England
- Northern Ireland
- Scotland
- Wales
- Business
- Politics
- Health
- Education
- Science & Environment
- Technology
- Entertainment
- Also in the news
- Video and Audio
- Have Your Say
- Magazine
- In Pictures
- Country Profiles
- Special Reports

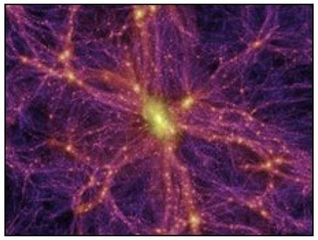
Page last updated at 11:52 GMT, Friday, 18 December 2009

E-mail this to a friend | Printable version

The first glimpse of dark matter?

By Victoria Gill
Science reporter, BBC News

US scientists have reported the detection of signals that could indicate the presence of dark matter.



Dark matter may make up most of the "cosmic web" of the Universe

A team announced on Thursday detecting two events with characteristics "consistent with" what physicists believe make up the elusive matter.

The main announcement came from the Department of Energy's Fermi National Accelerator Laboratory near Chicago.

The scientists were keen to stress that they could not confirm that what they had seen was definitely dark matter.

"While this result is consistent with dark matter, it is also consistent with backgrounds," said Fermilab's director, Pier Oddone.

SEE ALSO
- Signals could be from dark matter
  01 Apr 09 | Science & Environment
- Cosmic crash unmasks dark matter
  30 Aug 08 | Science & Environment
- Giant black holes just got bigger
  09 Jun 09 | Science & Environment
- Hubble makes 3D dark matter map
  07 Jan 07 | Science & Environment

RELATED INTERNET LINKS
- Fermilab
- CDMS
- Soudan Underground Laboratory

The BBC is not responsible for the content of external internet sites

TOP SCIENCE & ENVIRONMENT STORIES
- Devil cancer source 'identified'
- DNA analysed from early European

Done

Some important pdfs: Discrete case

1) Poisson pdf

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

number of galaxies / degree^2 counted by a survey

r = number of detections

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

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

Some important pdfs: Discrete case

1) Poisson pdf

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

number of galaxies / degree^2 counted by a survey

r = number of detections

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

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

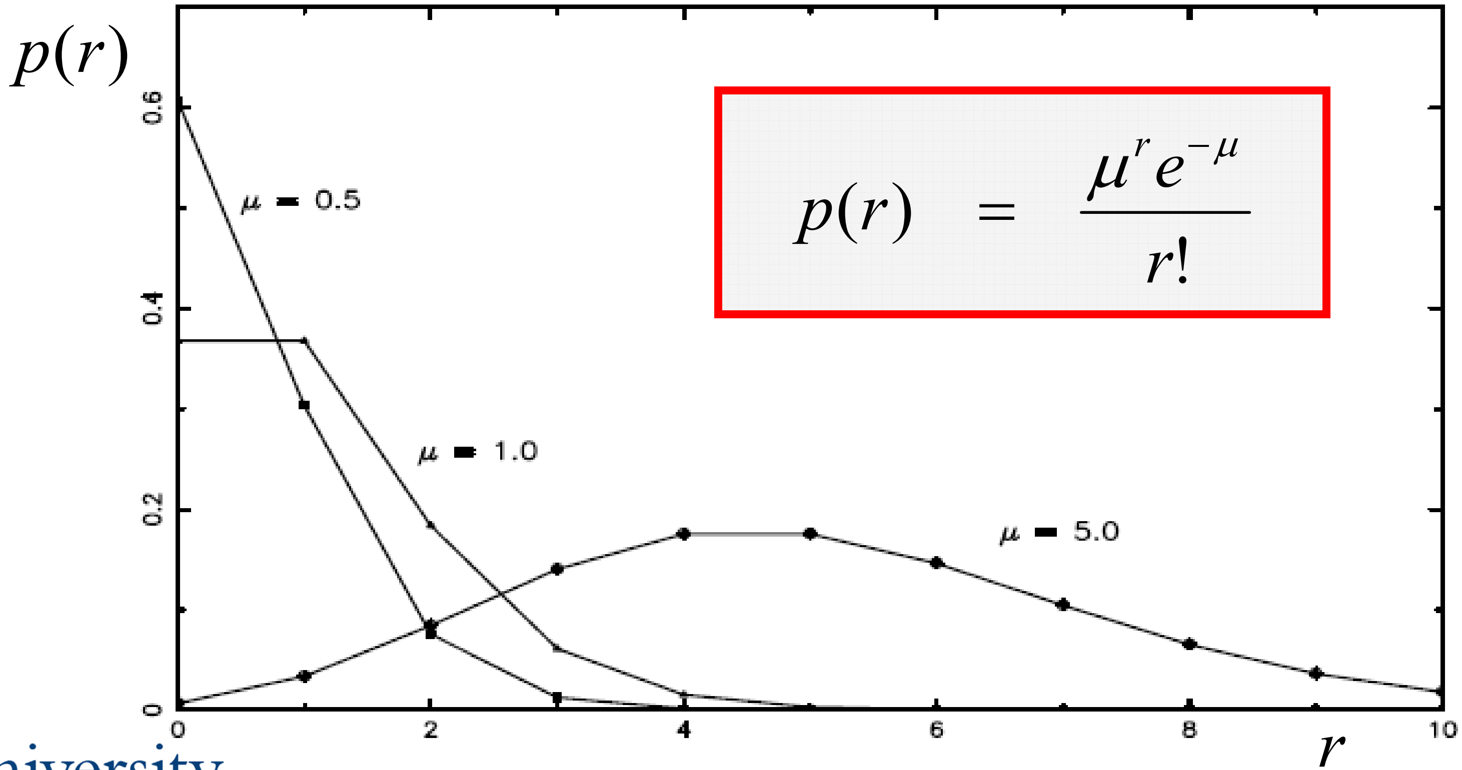
Can show that $\sum_{r=0}^{\infty} p(r) = 1$

Some important pdfs: Discrete case

1) Poisson pdf

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

number of galaxies / degree^2 counted by a survey



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

Some important pdfs: Discrete case

2) Binomial pdf

number of 'successes' from N observations, for two mutually exclusive outcomes ('Heads' and 'Tails')

r = number of 'successes'

θ = probability of 'success' for single observation

$$\boxed{p_N(r) = \frac{N!}{r!(N-r)!}\theta^r(1-\theta)^{N-r}}$$

Some important pdfs: Discrete case

2) Binomial pdf

number of 'successes' from N observations, for two mutually exclusive outcomes ('Heads' and 'Tails')

r = number of 'successes'

θ = probability of 'success' for single observation

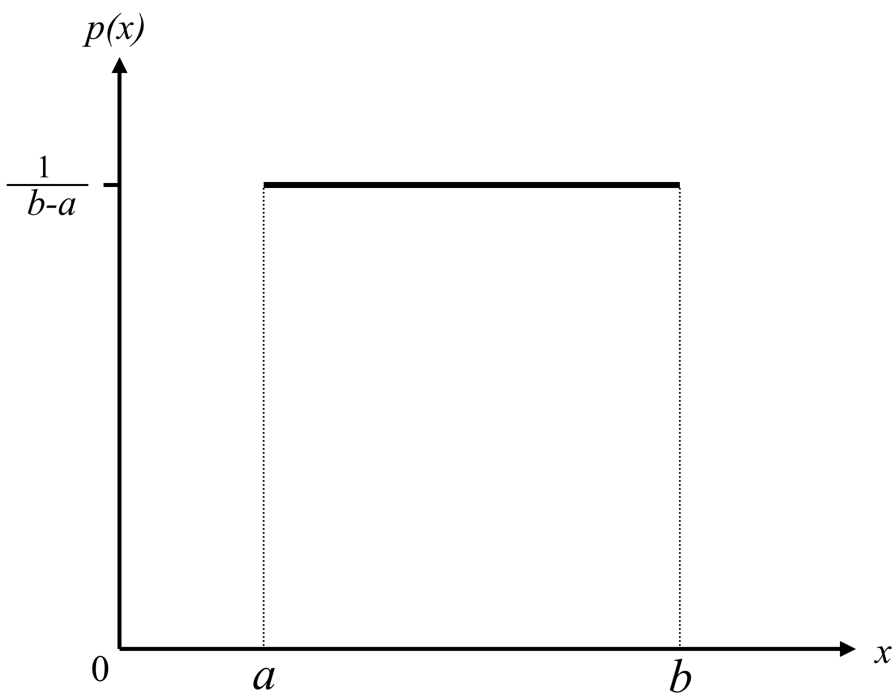
$$\boxed{p_N(r) = \frac{N!}{r!(N-r)!}\theta^r(1-\theta)^{N-r}}$$

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

Some important pdfs: Continuous case

1) Uniform pdf

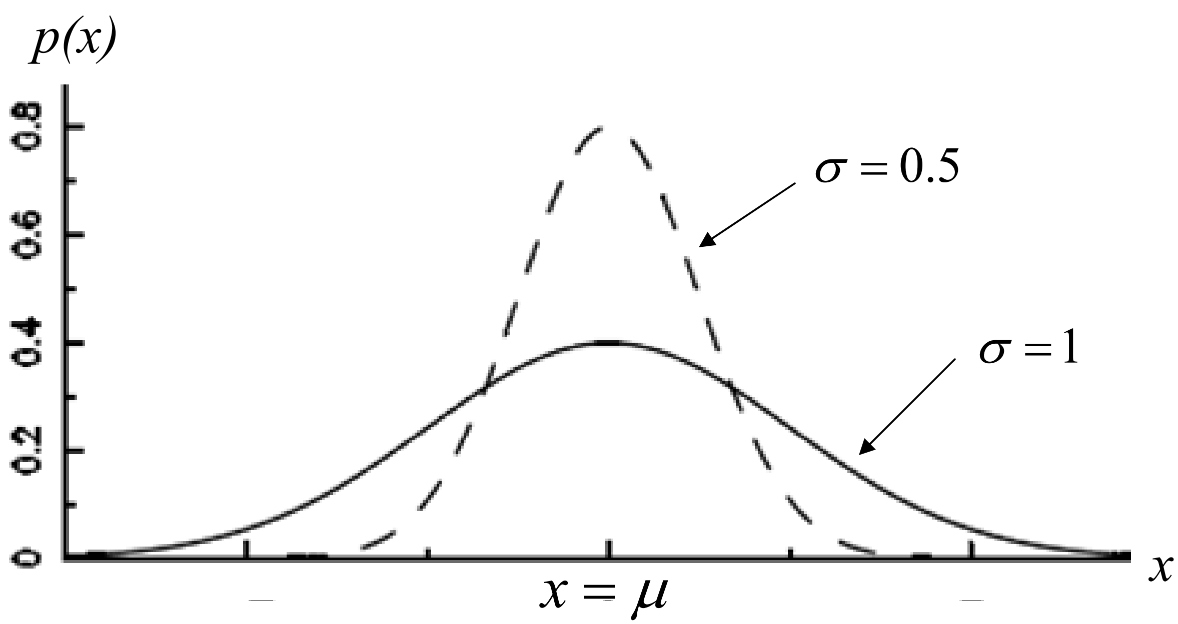
$$p(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$



Some important pdfs: Continuous case

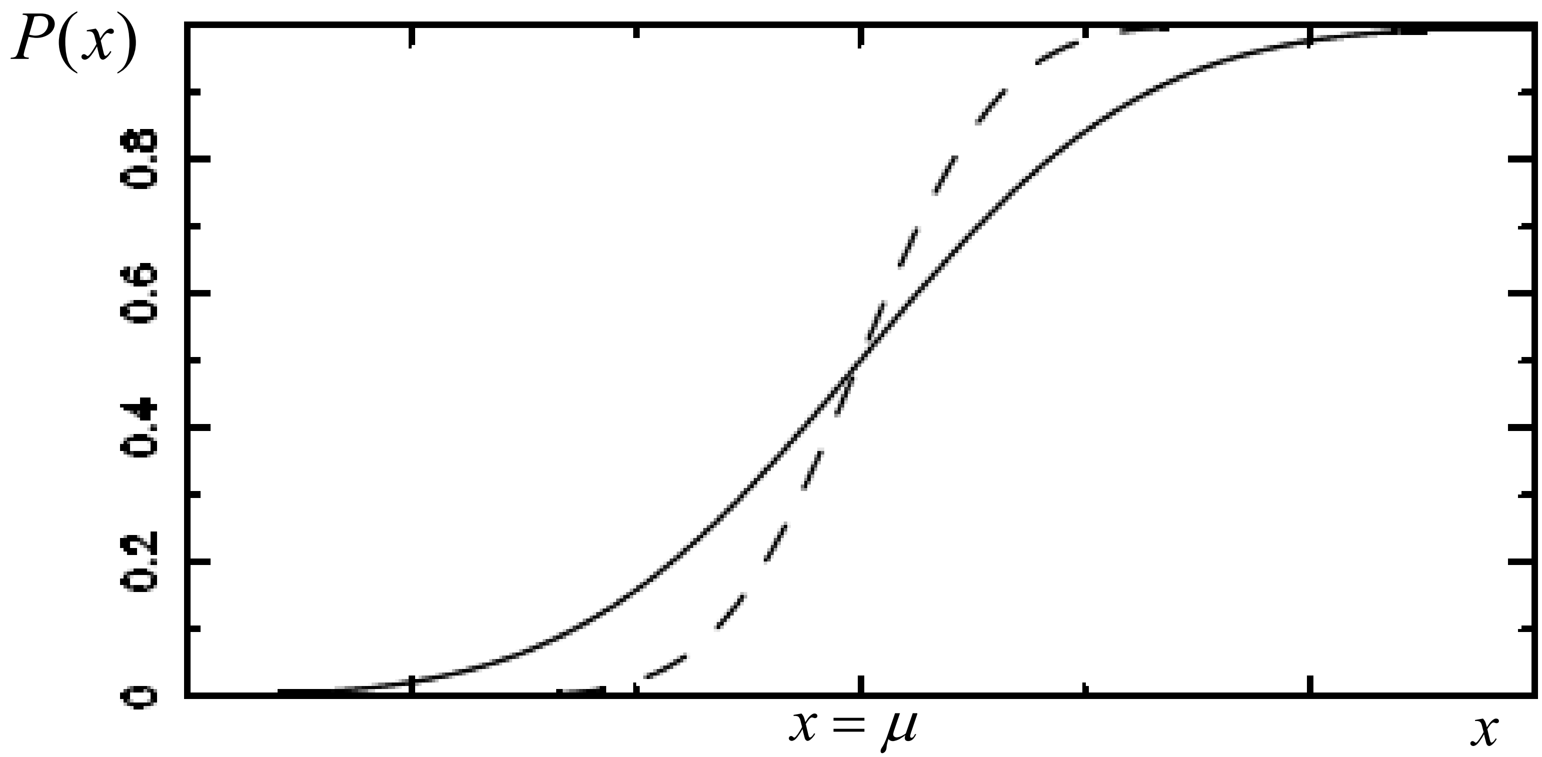
2) Central, or normal pdf
(also known as *Gaussian*)

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



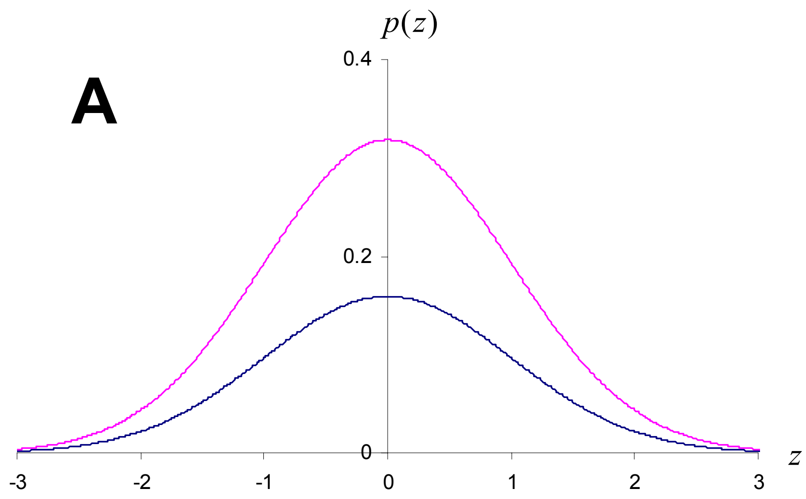
Cumulative distribution function (CDF)

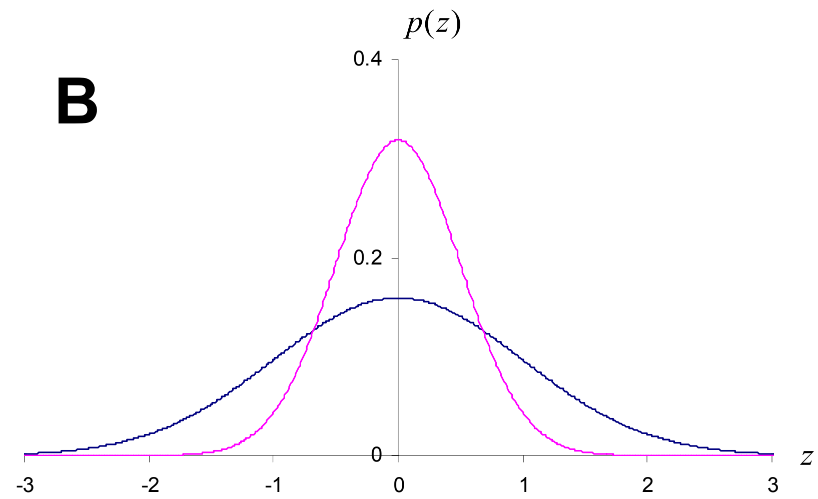
$$P(a) \;=\; \int_{-\infty}^{a} p(x)dx \;=\; \text{Prob}(\, x < a \,)$$

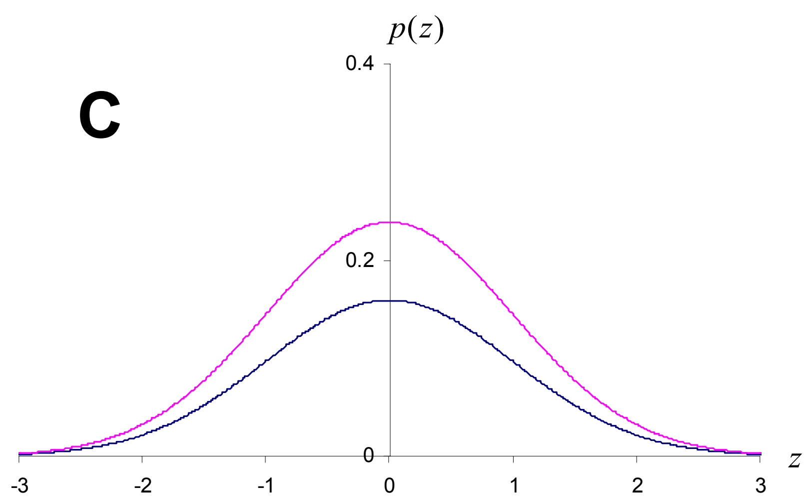


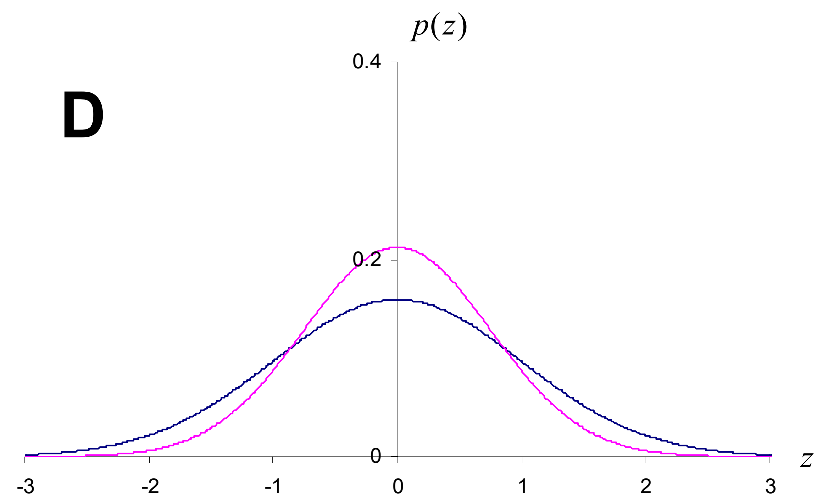
Question 1: In the figures below, the blue curves show a normal distribution with mean zero and $\sigma = 1$.

Which of the pink curves shows a normal distribution with mean zero and $\sigma = 0.5$?



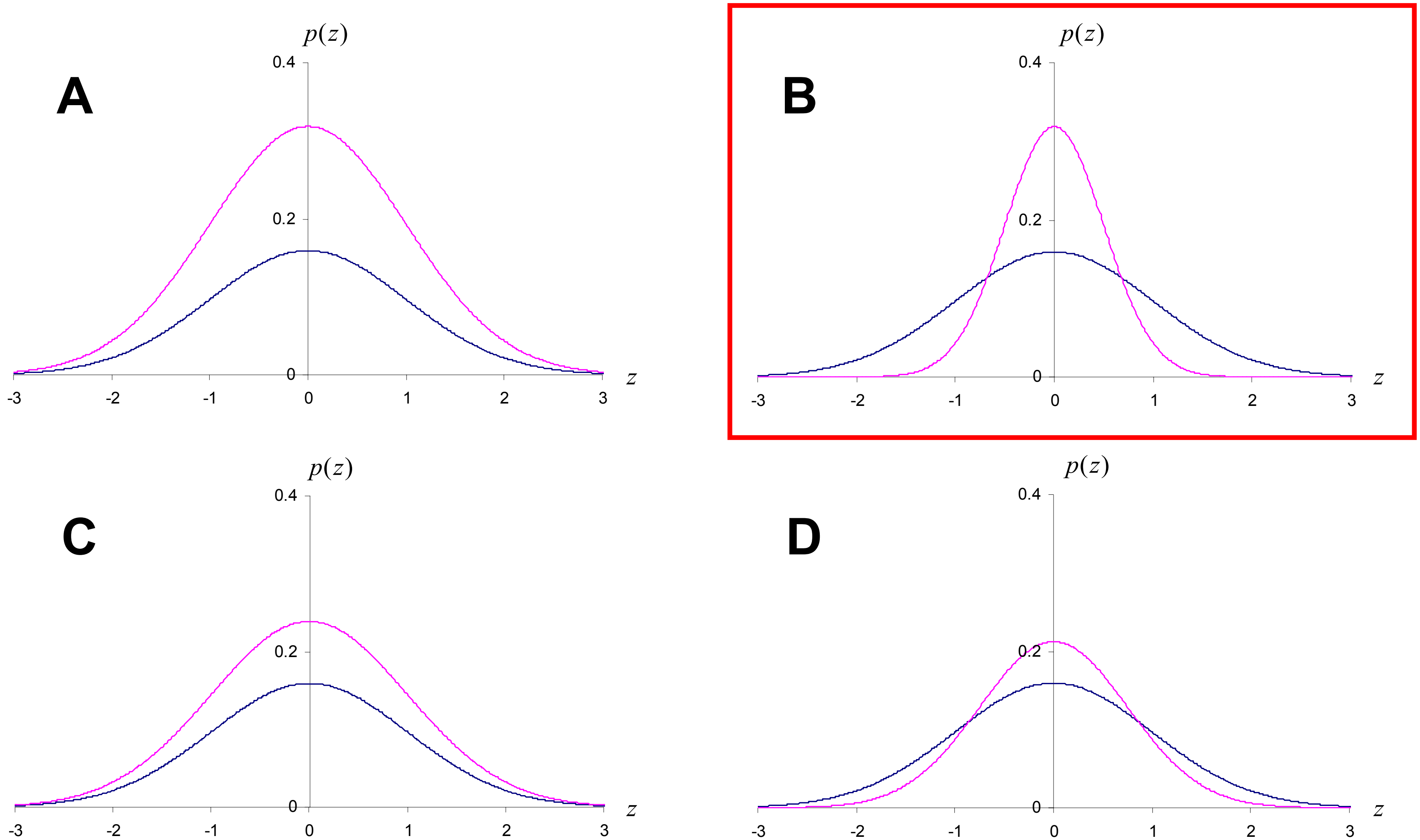






Question 1: In the figures below, the blue curves show a normal distribution with mean zero and $\sigma = 1$.

Which of the pink curves shows a normal distribution with mean zero and $\sigma = 0.5$?



Measures and moments of a pdf

The nth moment of a pdf is defined as:-

$$\langle x^n \rangle = \sum_{x=0}^{\infty} x^n p(x \mid I)$$

Discrete case

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

Continuous case

Measures and moments of a pdf

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

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

Discrete case

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

Continuous case

Measures and moments of a pdf

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

$$\langle x^2 \rangle = \sum_{x=0}^{\infty} x^2 \, p(x \mid I)$$

Discrete case

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 \, p(x \mid I) dx$$

Continuous case

Measures and moments of a pdf

The **variance** is defined as:

$$\mathrm{var}[x] = \sum_{x=0}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I)$$

Discrete case

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I) dx$$

Continuous case

and is often written as σ^2

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

Measures and moments of a pdf

The **variance** is defined as:

$$\mathrm{var}[x] = \sum_{x=0}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I)$$

Discrete case

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} \left(x - \langle x \rangle\right)^2 p(x \mid I) dx$$

Continuous case

In general

$$\boxed{\mathrm{var}[x] = \langle x^2 \rangle - \langle x \rangle^2}$$

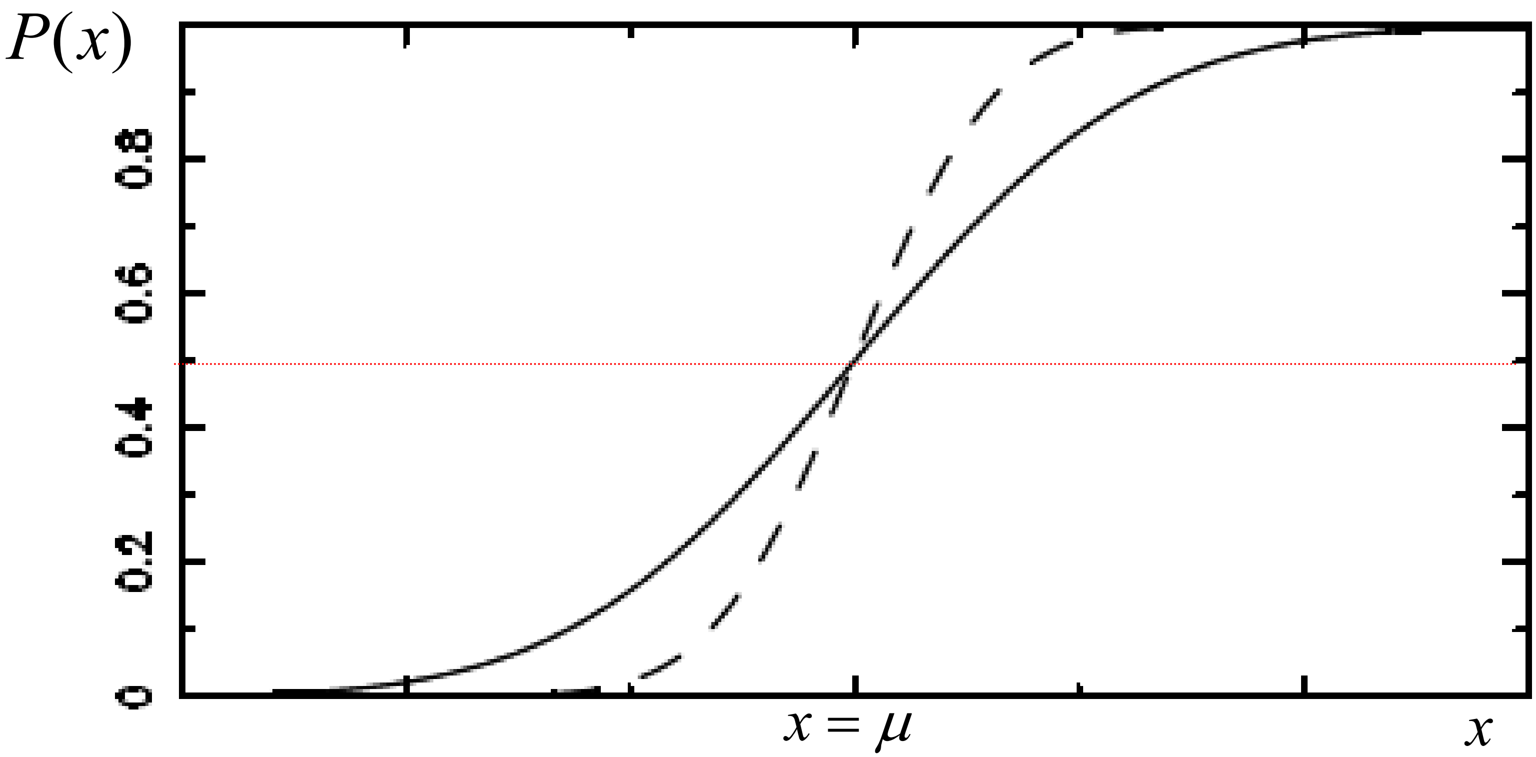
Measures and moments of a pdf

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

Measures and moments of a pdf

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

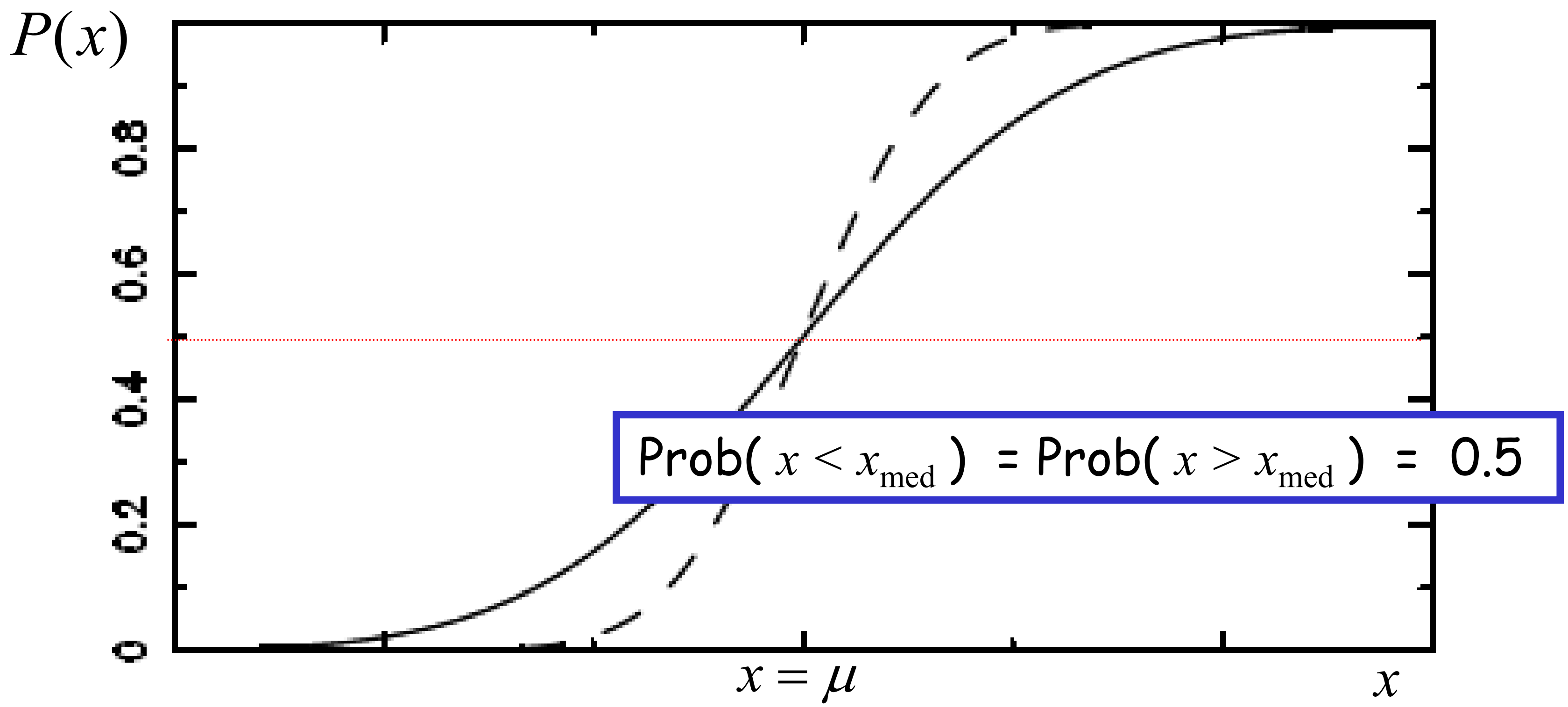
$$P(x_{\rm med}) \quad = \quad \int_{-\infty}^{x_{\rm med}} p(x')dx' \quad = \quad 0.5$$



Measures and moments of a pdf

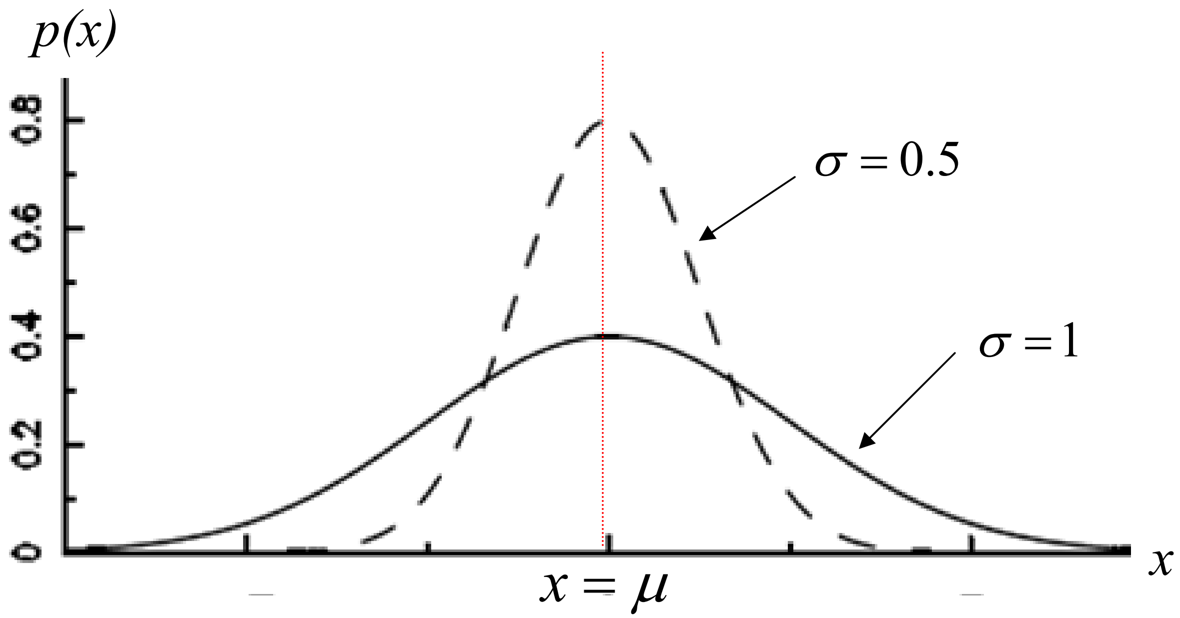
The **Median** divides the CDF into two equal halves

$$P(x_{\rm med}) \quad = \quad \int_{-\infty}^{x_{\rm med}} p(x')dx' \quad = \quad 0.5$$



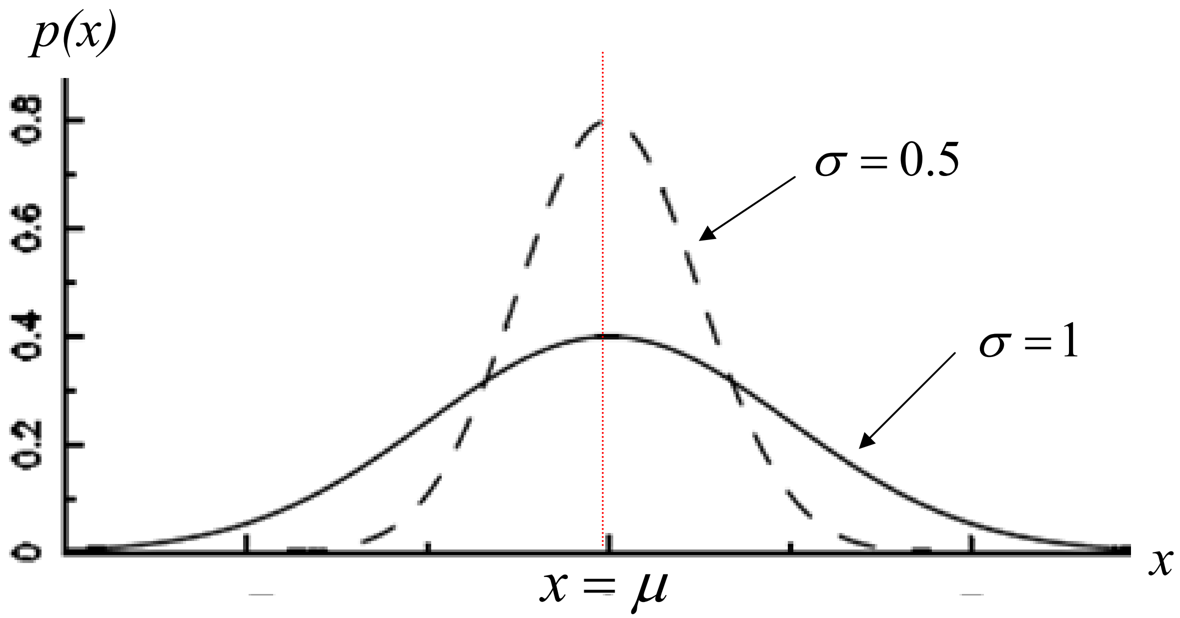
Measures and moments of a pdf

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



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

Variance of a Function of a RV

The variance, $\mathrm{var}[f(x)]$, of an arbitrary function of x can be approximated to second order by the following expression

$$\boxed{\mathrm{var}[f(x)] \quad = \quad \mathrm{var}(x) \left(\frac{\partial f}{\partial x}\right)^2_{x=\overline{x}}}$$

This expression is the basis for the 'error propagation' formulae we use in e.g. undergraduate physics labs

See also the SUPAIDA course

Question 2: Which expression correctly approximates the error on the *natural logarithm* of a variable x ?

A $\sigma_{\ln x} \sim \sigma_x^{\,2} / x^2$

B $\sigma_{\ln x} \sim x^2 \, \sigma_x^{\,2}$

C $\sigma_{\ln x} \sim \sigma_x / x$

D $\sigma_{\ln x} \sim x \, \sigma_x$

Question 2: Which expression correctly approximates the error on the *natural logarithm* of a variable x ?

A $\sigma_{\ln x} \sim {\sigma_x}^2 / x^2$

B $\sigma_{\ln x} \sim x^2 \, {\sigma_x}^2$

C $\boxed{\sigma_{\ln x} \sim \sigma_x / x}$

D $\sigma_{\ln x} \sim x \, \sigma_x$

Multivariate Distributions

Thus far we have considered only the pdf of a single (univariate) RV. We now extend to the **multivariate** case of two or more RVs.

Joint pdf

The **joint pdf** of two RVs, x_1 and x_2 is $p(x_1, x_2)$. Then,

$$\boxed{\mathrm{Prob}(a_1 < X_1 < b_1 \ \text{ and } \ a_2 < X_2 < b_2) \quad = \quad \int_{a_1}^{b_1} \int_{a_2}^{b_2} p(x_1, x_2)\, dx_1 dx_2}$$

Extension to more than two RVs is carried out in the obvious way.

Marginal Distributions

The **marginal pdf**, $p_1(x_1)$ of x_1 is defined by

$$p_1(x_1) \quad = \quad \int_{-\infty}^{\infty} p(x_1, x_2)\, dx_2$$

and is a pdf in the usual sense that

1. $p_1(x_1) \geq 0$, for all x_1
2. $\mathrm{Prob}(a < x_1 < b) \quad = \quad \int_a^b p_1(x_1) dx_1$
3. $\int_{-\infty}^{\infty} p_1(x_1) dx_1 \quad = \quad 1$

Marginal Distributions

Similarly, the marginal pdf of x_2 is

$$p_2(x_2) = \int_{-\infty}^{\infty} p(x_1, x_2)\, dx_1$$

In general, given any multivariate pdf, we may find the marginal pdf of any subset of the $x_1, ..., x_n$ by integrating over all other variables.

e.g.

$$\boxed{p_{13}(x_1, x_3) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} p(x_1, ..., x_n)\, dx_2 dx_4 dx_5 ... dx_n}$$

Conditional Distributions

Consider the joint pdf, $p(x_1, x_2)$, of x_1 and x_2. Suppose we observe x_1, but do not observe x_2. We want a function that describes the pdf of x_2, given the observed value of x_1 (usually simply stated as 'given x_1'). This function is known as the **conditional** pdf of x_2, written as $p(x_2|x_1)$, and defined by

$$\boxed{p(x_2|x_1) \quad = \quad \frac{p(x_1, x_2)}{p_1(x_1)}}$$

i.e. the conditional pdf is obtained by dividing the joint pdf of x_1 and x_2 by the marginal pdf of x_1 (provided $p_1(x_1) \neq 0$).

Conditional Distributions

Similarly

$$\boxed{p(x_1|x_2) \quad = \quad \frac{p(x_1, x_2)}{p_2(x_2)}}$$

Extension to more than 2 RVs is again straightforward. For example,

$$\boxed{p(x_1, x_3|x_2, x_4) \quad = \quad \frac{p(x_1, x_2, x_3, x_4)}{p_{24}(x_2, x_4)}}$$

Statistical Independence

If the conditional pdf of x_2 given x_1 does *not* depend on x_1, this means that x_1 and x_2 are statistically independent, since the observed value of x_2 is unaffected by the observed value of x_1.

Equivalently, x_1 and x_2 are independent if and only if the joint pdf of x_1 and x_2 can be written as the product of their marginal pdfs, i.e.

$$\boxed{p(x_1, x_2) = p_1(x_1)\, p_2(x_2)}$$

Question 3: Which of the following joint pdfs describe variables x and y which are statistically independent?

A $p(x,y) \propto \frac{1}{2}(x+y);\qquad 0 \le x, y < \infty$

B $p(x,y) \propto \exp\left[-\frac{1}{2}(x+y)\right];\qquad 0 \le x, y < \infty$

C $p(x,y) \propto \log(x+y);\qquad 0 < x, y < \infty$

D $p(x,y) \propto \exp\left[-\frac{1}{2}(x+y)\right];\qquad 0 \le x < y,\ 0 \le y < \infty$

Question 3: Which of the following joint pdfs describe variables x and y which are statistically independent?

A $p(x,y) \propto \frac{1}{2}\left(x+y\right)$; $\quad 0 \le x, y < \infty$

B $p(x,y) \propto \exp\left[-\frac{1}{2}\left(x+y\right)\right]$; $\quad 0 \le x, y < \infty$

C $p(x,y) \propto \log\left(x+y\right)$; $\quad 0 < x, y < \infty$

D $p(x,y) \propto \exp\left[-\frac{1}{2}\left(x+y\right)\right]$; $\quad 0 \le x < y,\ 0 \le y < \infty$

The bivariate normal distribution

Let x and y be RVs with the following joint pdf

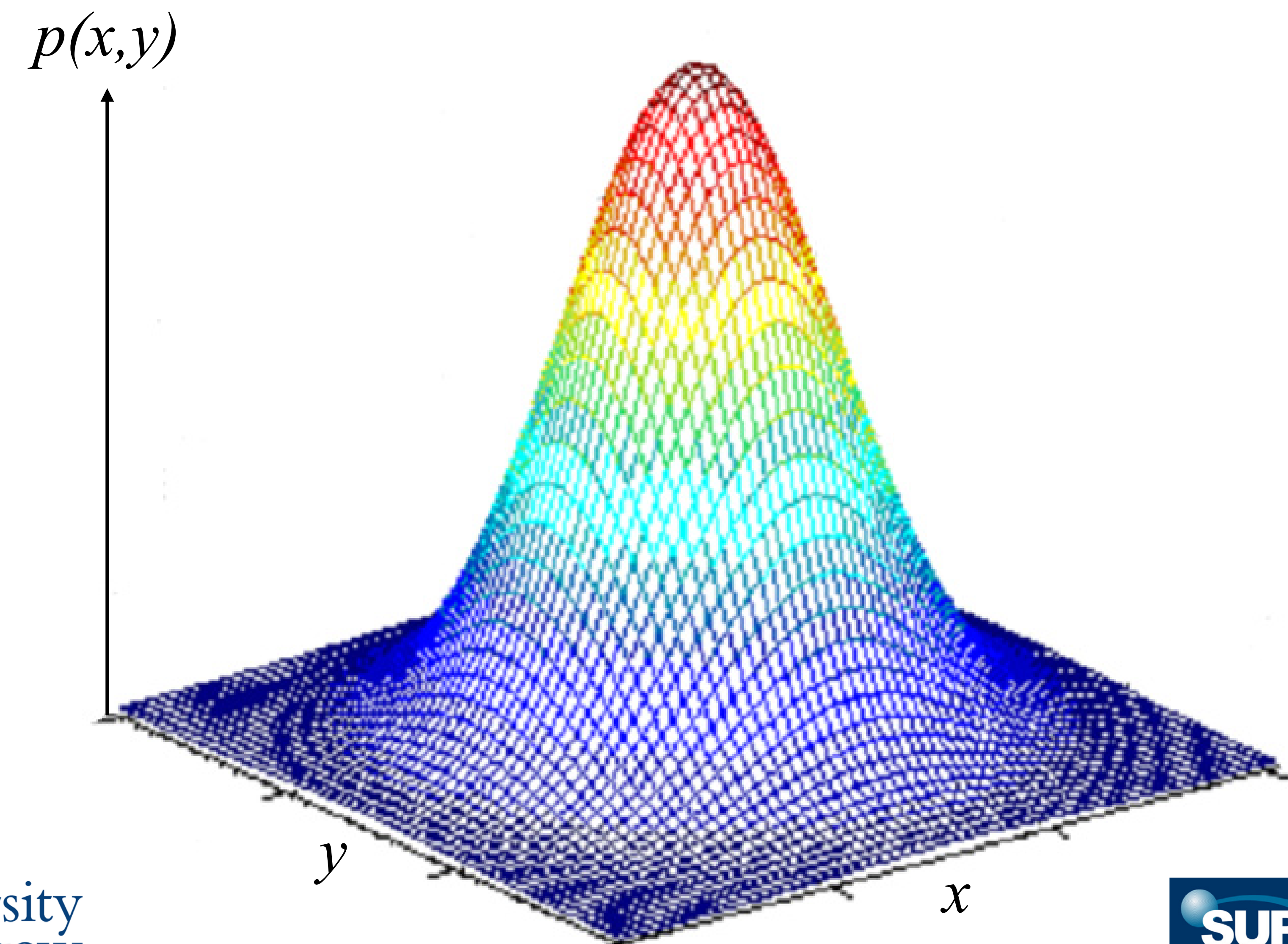
$$p(x,y) \quad = \quad \frac{1}{2\pi\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}Q(x,y)\right]$$

where the quadratic form, $Q(x,y)$ is given by

$$Q(x,y) \quad = \quad \left(\frac{x-\mu_{\mathrm{x}}}{\sigma_{\mathrm{x}}}\right)^2 - 2\rho\left(\frac{x-\mu_{\mathrm{x}}}{\sigma_{\mathrm{x}}}\right)\left(\frac{y-\mu_{\mathrm{y}}}{\sigma_{\mathrm{y}}}\right) + \left(\frac{y-\mu_{\mathrm{y}}}{\sigma_{\mathrm{y}}}\right)^2$$

Then $p(x,y)$ is known as the **bivariate normal pdf** and is specified by the 5 parameters μ_{x}, μ_{y}, σ_{x}, σ_{y} and ρ. This pdf is used often in the physical sciences to model the joint pdf of two random variables.

The bivariate normal distribution



The bivariate normal distribution

The first 4 parameters of the bivariate normal pdf are, in fact, equal to the following expectation values:-

1. $E(x) = \mu_{\mathrm{x}}$
2. $E(y) = \mu_{\mathrm{y}}$
3. $\mathrm{var}(x) = \sigma_{\mathrm{x}}^2$
4. $\mathrm{var}(y) = \sigma_{\mathrm{y}}^2$

The bivariate normal distribution

The parameter ρ is known as the **correlation coefficient** and satisfies

$$\boxed{E[(x-\mu_{\mathrm{x}})(y-\mu_{\mathrm{y}})] \quad = \quad \rho\sigma_{\mathrm{x}}\sigma_{\mathrm{y}}}$$

Note that if $\rho = 0$ then x and y are independent.

$E[(x-\mu_{\mathrm{x}})(y-\mu_{\mathrm{y}})]$ is known as the **covariance** of x and y and is often denoted by $\mathrm{cov}(x, y)$.

The bivariate normal distribution

The parameter ρ is known as the **correlation coefficient** and satisfies

$$\boxed{E[(x-\mu_{\rm x})(y-\mu_{\rm y})] \quad = \quad \rho\sigma_{\rm x}\sigma_{\rm y}}$$

Note that if $\rho = 0$ then x and y are independent.

$E[(x-\mu_{\rm x})(y-\mu_{\rm y})]$ is known as the **covariance** of x and y and is often denoted by $\mathrm{cov}(x,y)$.

In fact, for *any* two variables x and y, we define

$$\boxed{\mathrm{cov}(x,y) \quad = \quad E\left[\left(x-E(x)\right)\left(y-E(y)\right)\right]}$$

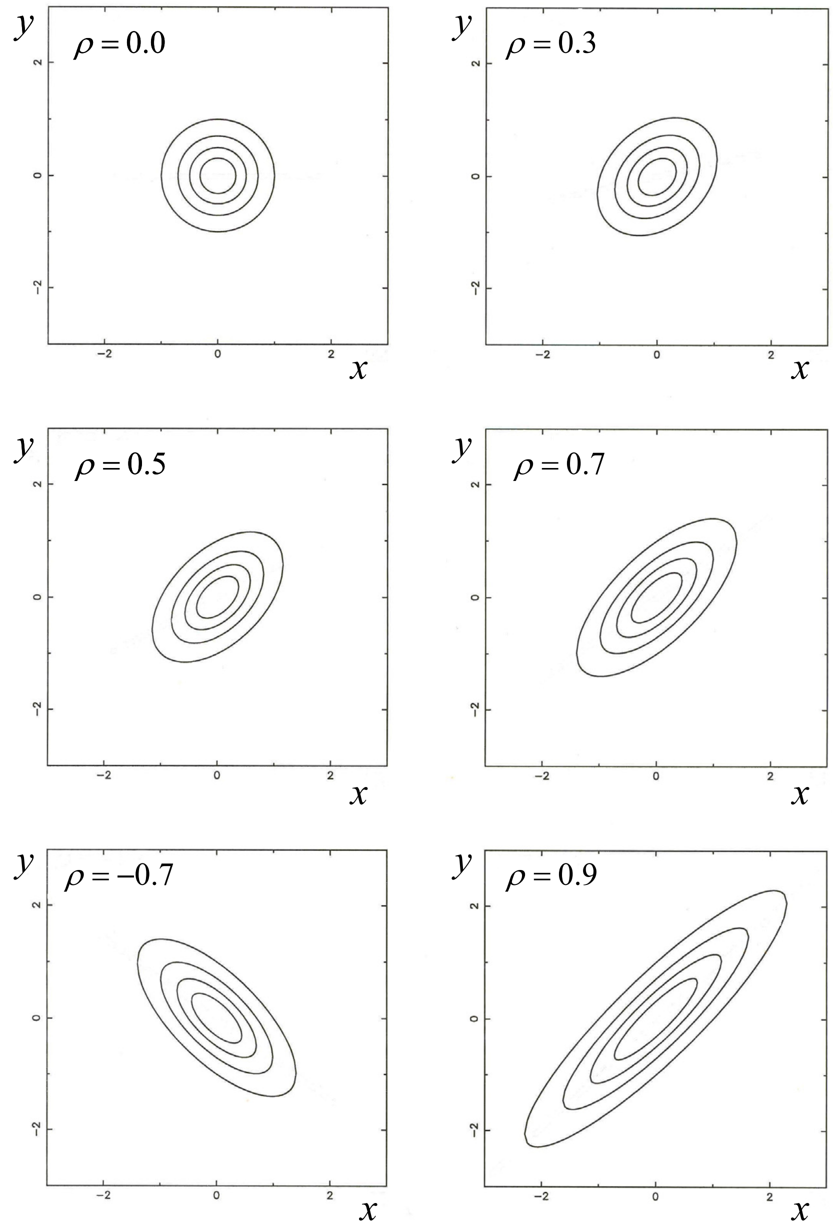
Isoprobability contours for the bivariate normal pdf

$\rho > 0$: positive correlation

y tends to increase as x increases

$\rho < 0$: negative correlation

y tends to decrease as x increases



Isoprobability contours for the bivariate normal pdf

$\rho > 0$:   positive correlation

y tends to increase as x increases

$\rho < 0$:   negative correlation

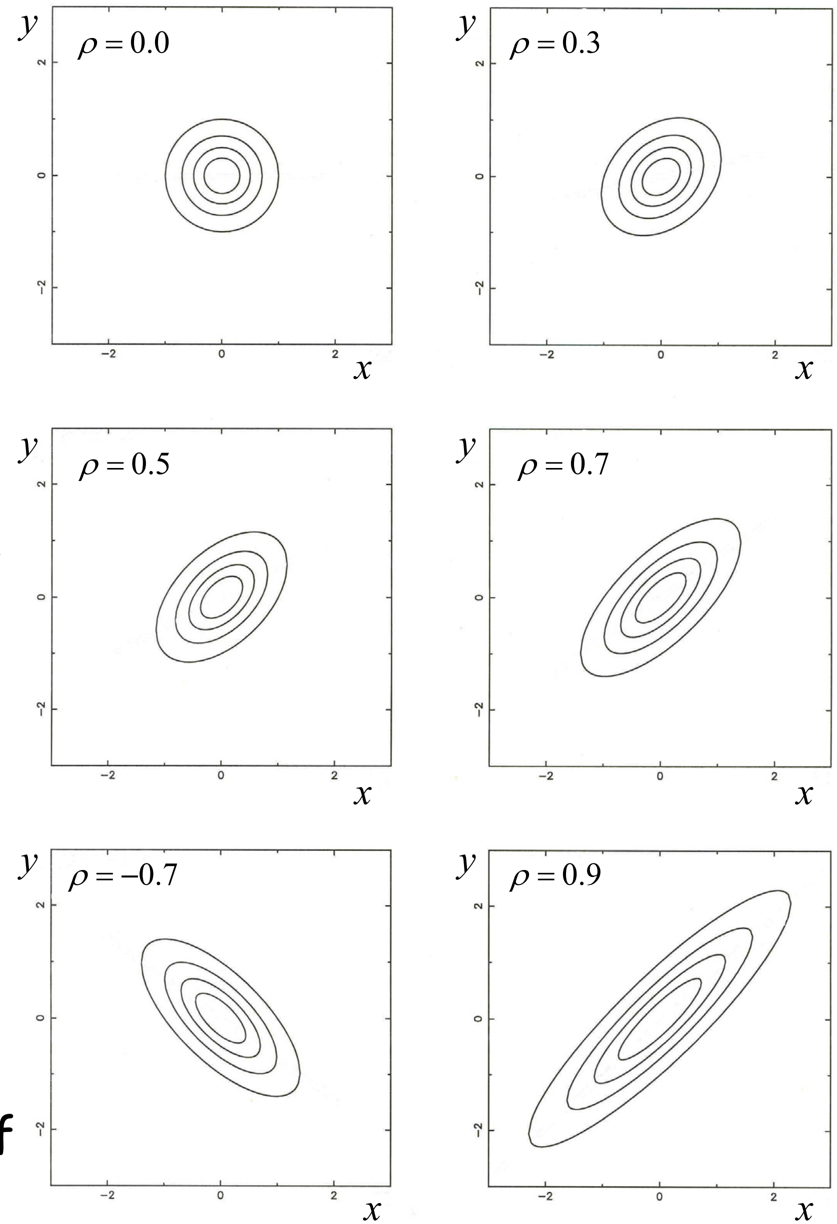
y tends to decrease as x increases

Contours become narrower and steeper as

$$|\rho| \rightarrow 1$$

$\Rightarrow$ stronger (anti) correlation between x and y.

i.e.   *Given* value of x, value of y is tightly constrained.



The bivariate normal distribution

The marginal pdfs of x and y are just the univariate normal pdfs, i.e.

$$p_x(x) = N(\mu_\mathrm{x}, \sigma_\mathrm{x}) \qquad p_y(y) = N(\mu_\mathrm{y}, \sigma_\mathrm{y})$$

The conditional pdf of y given x is also a univariate normal pdf, viz:-

$$\boxed{p(y|x) \quad = \quad N(\mu_\mathrm{y} + \frac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}\rho(x - \mu_\mathrm{x}), \sigma_\mathrm{y}\sqrt{1-\rho^2})}$$

with the corresponding expression for $p(x|y)$.

The bivariate normal distribution

$\mu_y + \frac{\sigma_y}{\sigma_x}\rho(x - \mu_x)$ is often referred to as the **conditional expectation** (value) of y given x, and the equation

$$\boxed{y = \mu_y + \frac{\sigma_y}{\sigma_x}\rho(x - \mu_x)}$$

is called the **regression line** of y on x.

(see also Section 2)