2. More Theoretical Foundations: Marginalisation

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

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

2. 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.

$$p(x_1, y \mid I) = p(x_1 \mid y, I)\, p(y \mid I)$$

$$\vdots$$

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

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

$$= 1$$

Then

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

$$= 1$$

Marginal probability

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

This extends to the *continuum limit* :

x can take infinitely many values

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

This extends to the *continuum limit* :

x can take infinitely many values

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

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

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

with obvious extension to continuum limit for y

This extends to the *continuum limit* :

x can take infinitely many values

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

Also

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

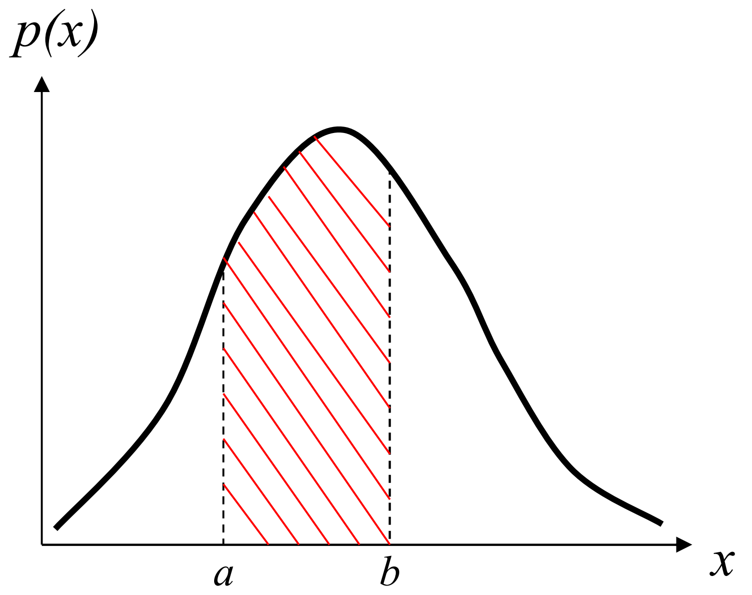
Normalisation condition

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.

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

'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 objects counted by a survey

r = number of detections

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 objects counted by a survey

r = number of detections

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

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 objects counted by a survey

r = number of detections

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

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

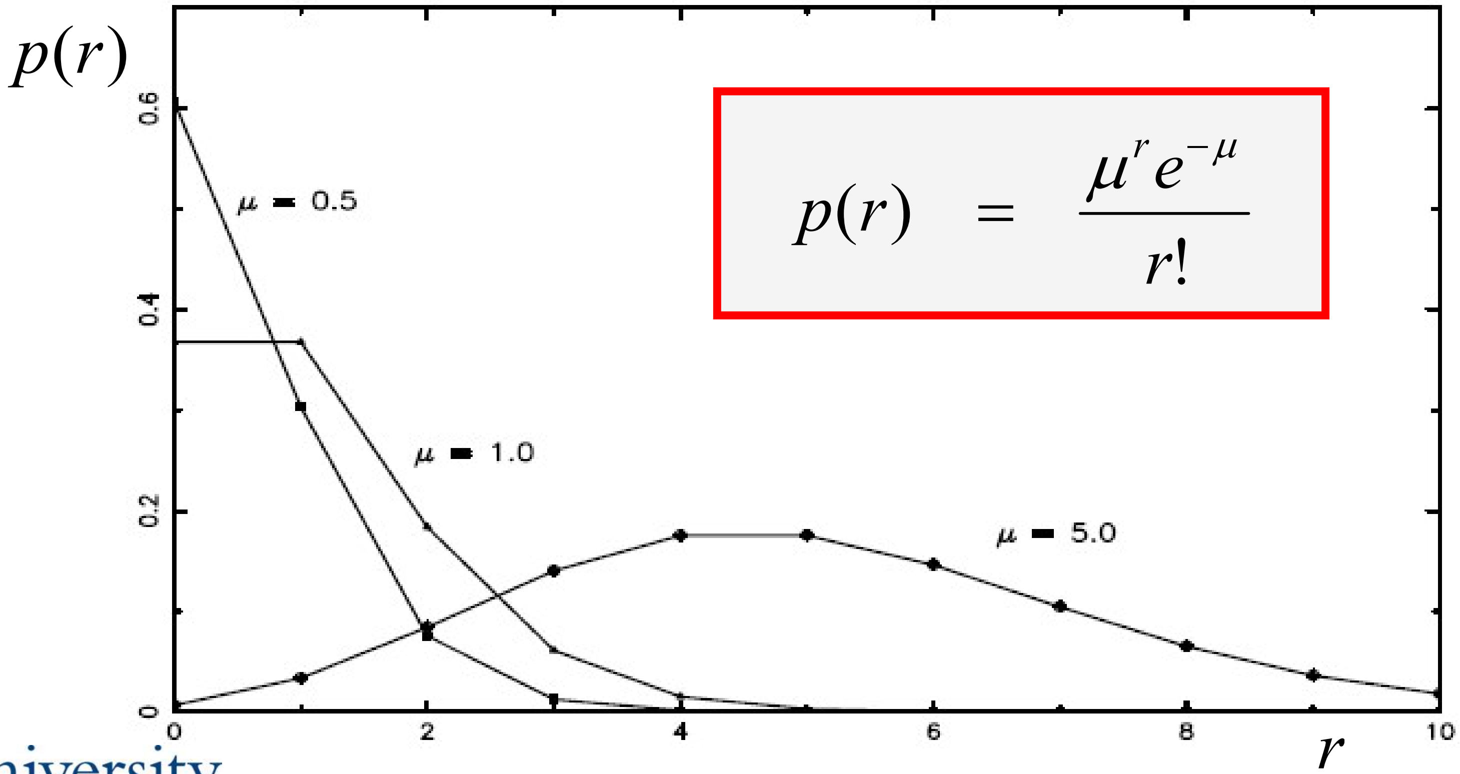
Can show that $\sum_{r=0}^{\infty} p(r) = 1$

Some important pdfs: Discrete case

1) Poisson pdf

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

number of objects 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 (e.g. 'Heads' and 'Tails')

r = number of 'successes'

θ = probability of 'success' for single observation

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

Some important pdfs: Discrete case

2) Binomial pdf

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

r = number of 'successes'

θ = probability of 'success' for single observation

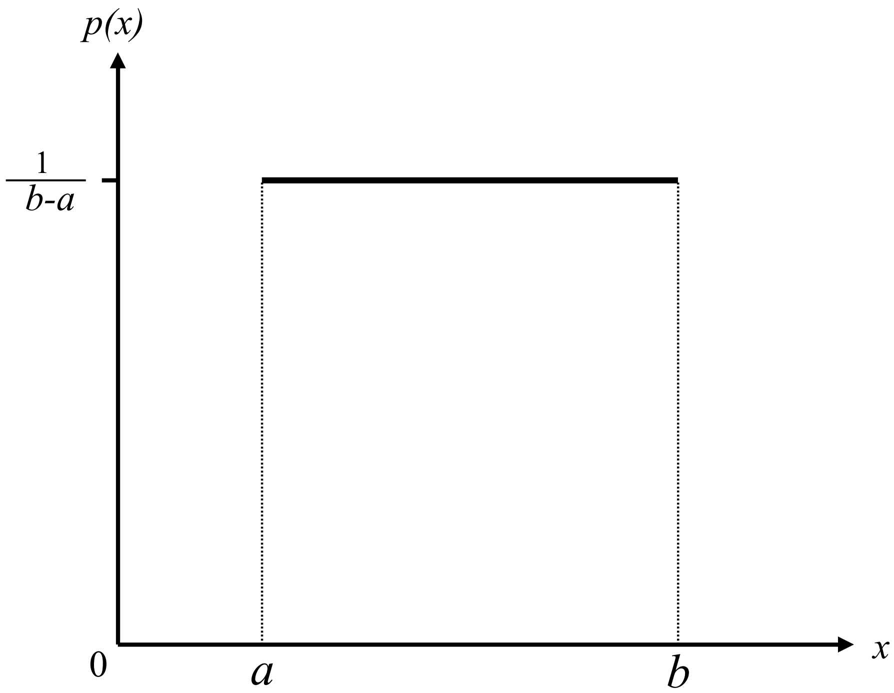
$$p_N(r) = \frac{N!}{r!(N-r)!}\theta^r(1-\theta)^{N-r}$$

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

Some important pdfs: Continuous case

1) Uniform pdf

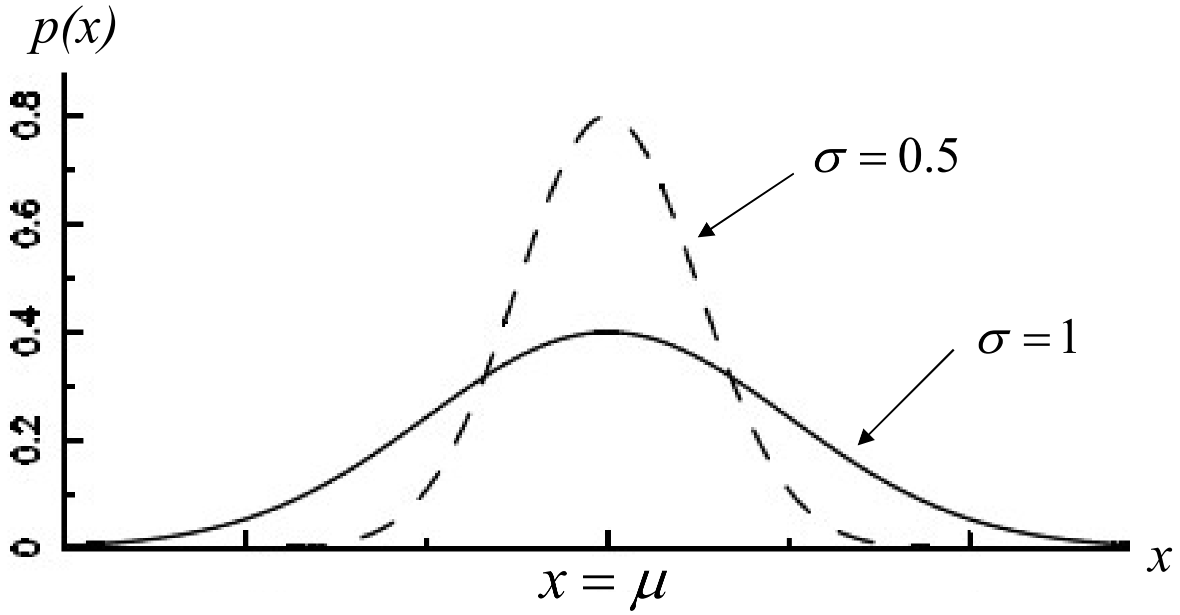
$$p(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$



Some important pdfs: Continuous case

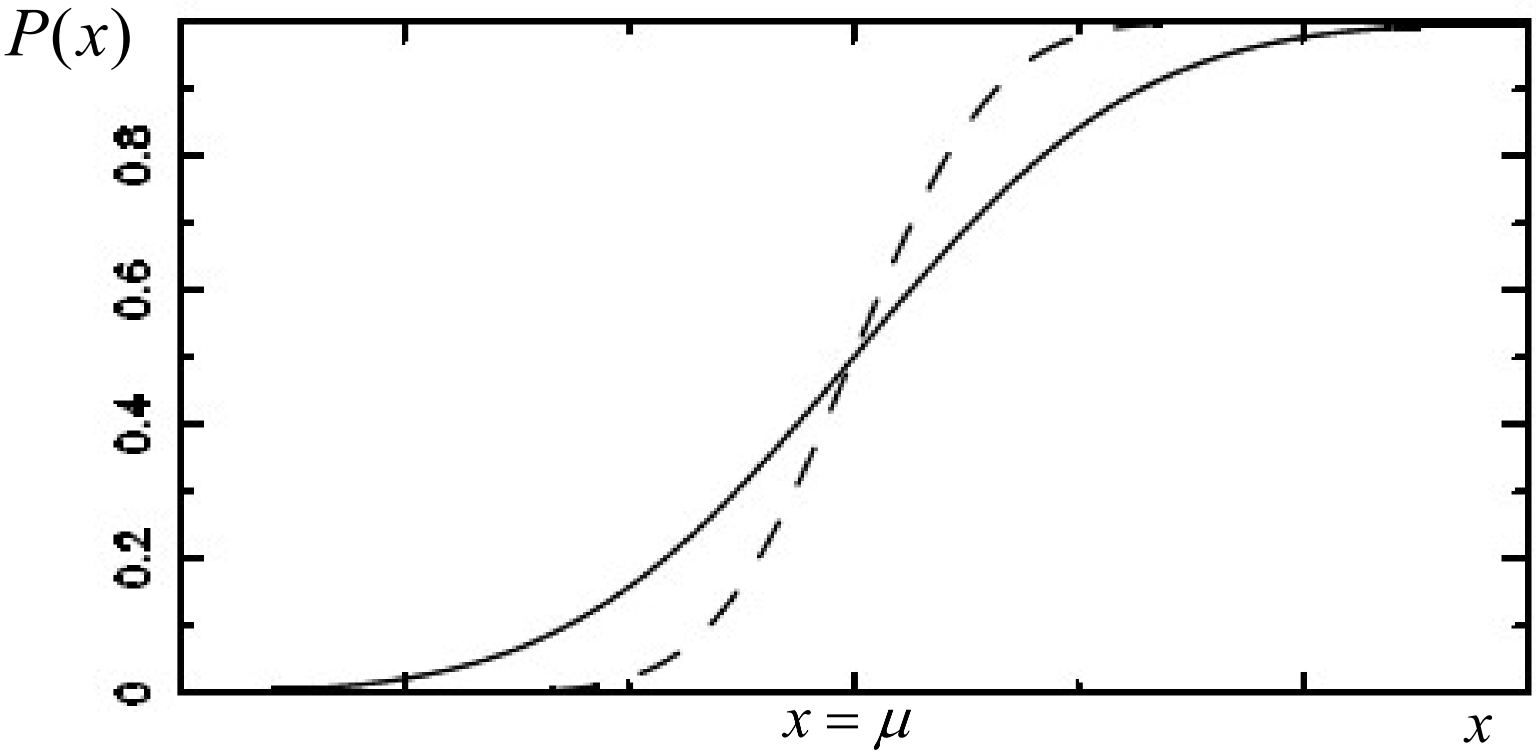
2) Central, or normal pdf
(also known as *Gaussian*)

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



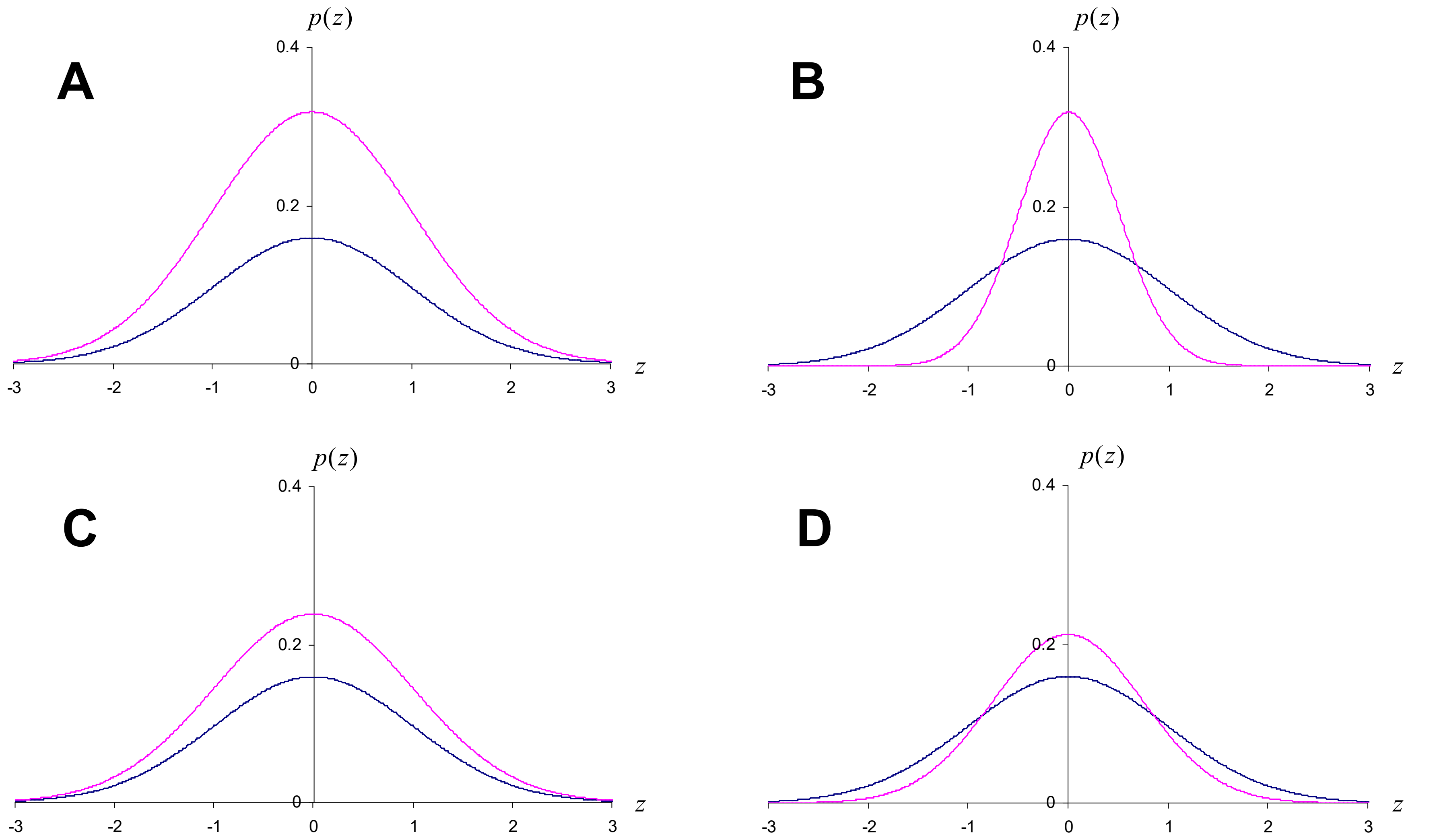
Cumulative distribution function (CDF)

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



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 $\sigma = 0.5$?



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:

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

Discrete case

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

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

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

Measures and moments of a pdf

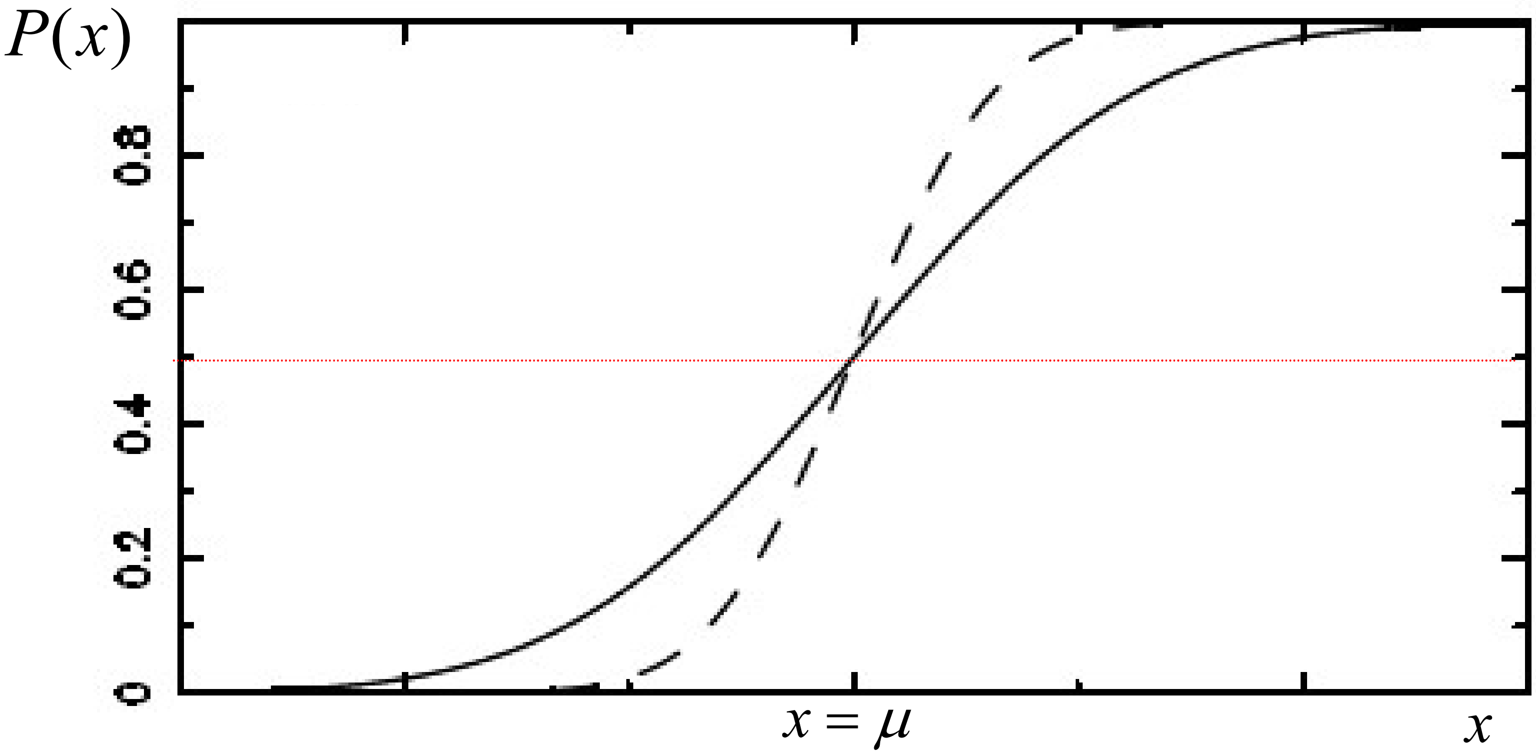
pdf	mean	variance
Poisson $p(r) = \frac{\mu^r e^{-\mu}}{r!}$	μ	μ
Binomial $p_N(r) = \frac{N!}{r!(N-r)!}\theta^r(1-\theta)^{N-r}$	$N\theta$	$N\theta(1-\theta)$
Uniform $p(x) = \frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
Normal $p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$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

Note: for a normal pdf all the higher moments are zero

Measures and moments of a pdf

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

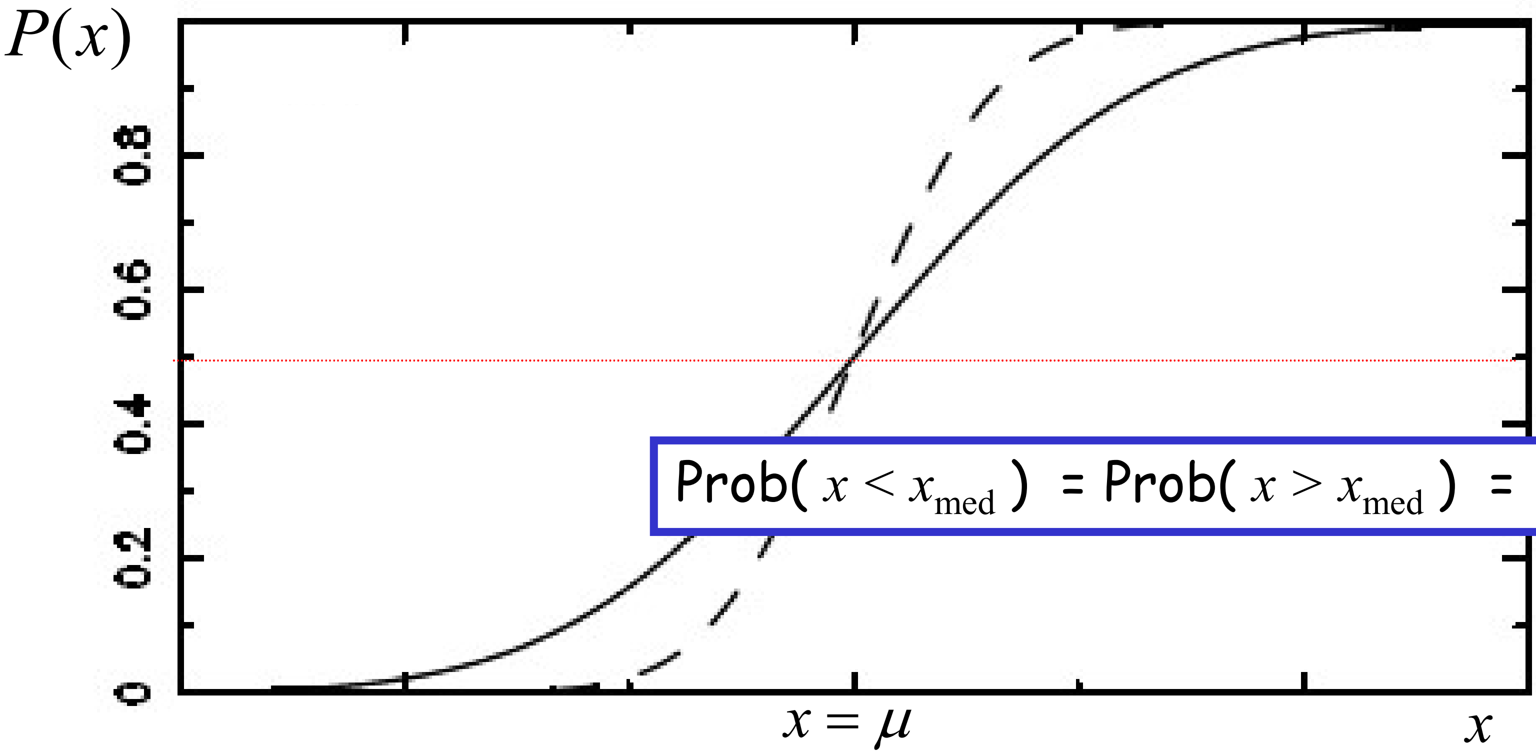
$$P(x_{\mathrm{med}}) = \int_{-\infty}^{x_{\mathrm{med}}} p(x')dx' = 0.5$$



Measures and moments of a pdf

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

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



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

Question 2: The following table lists the absolute magnitude (related to the log of luminosity) of the brightest galaxy in 15 distant galaxy clusters. The galaxies are arranged from brightest to faintest.

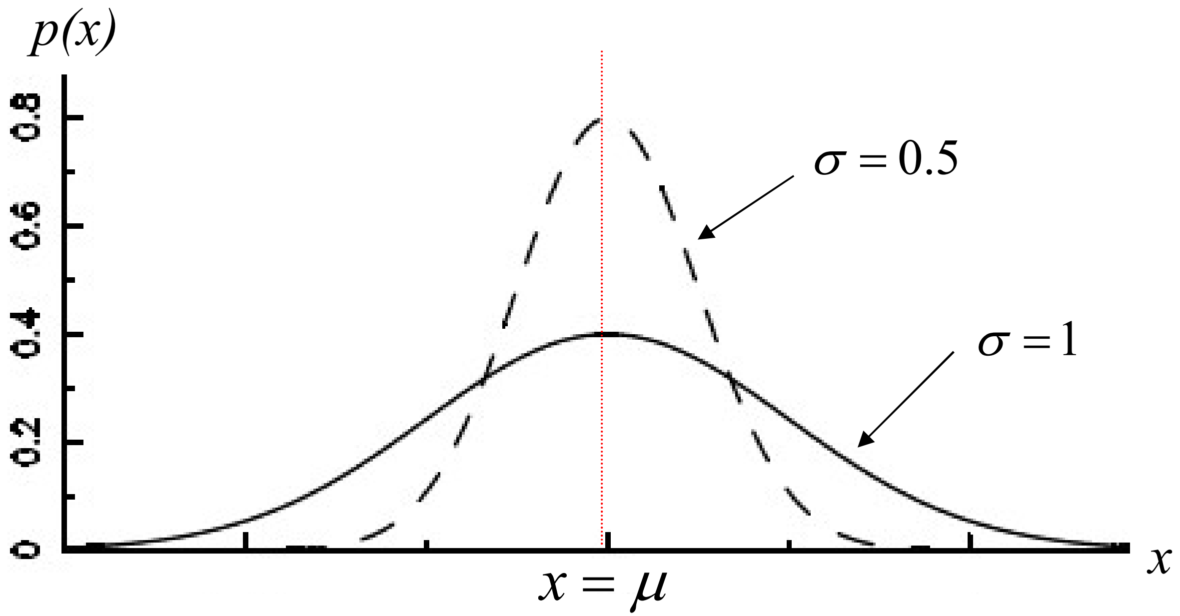
Cluster #	M_V	Cluster #	M_V	Cluster #	M_V
1	-23.36	6	-22.37	11	-21.14
2	-23.23	7	-21.89	12	-21.13
3	-22.50	8	-21.33	13	-21.11
4	-22.48	9	-21.23	14	-21.08
5	-22.46	10	-21.17	15	-21.04

The sample median of these data is

A -23.36 **B** -21.04 **C** -21.33 **D** -21

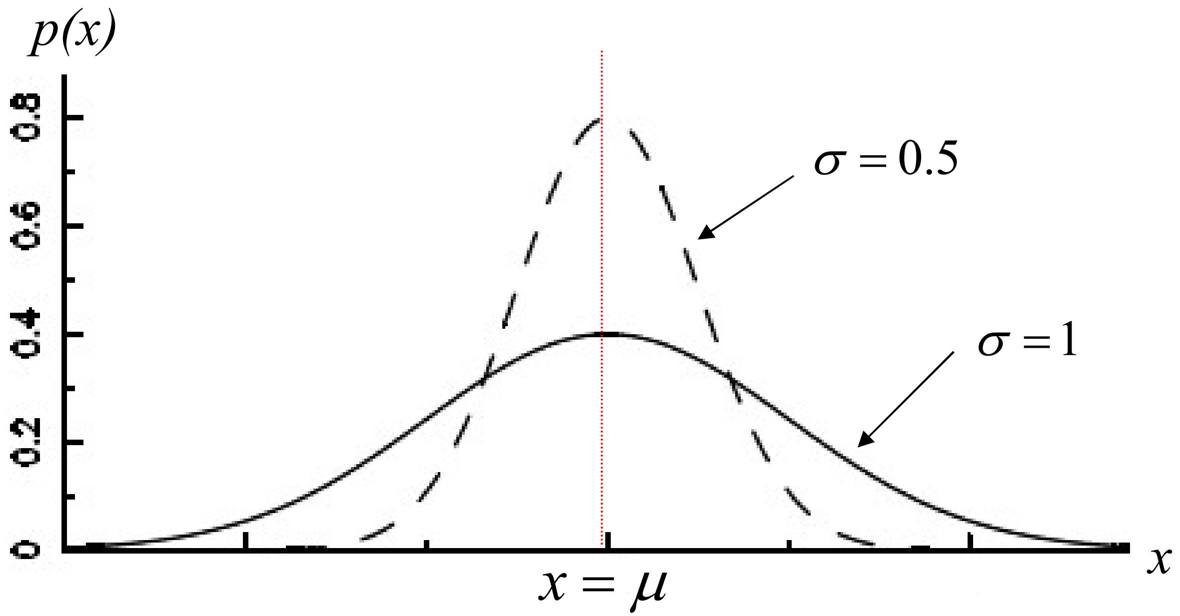
Measures and moments of a pdf

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

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

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

See also the SUPAIDA course

Question 3: 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 $\sigma_{\ln x} \sim x \, \sigma_x$

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,

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

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. $\text{Prob}(a < x_1 < b) \quad = \quad \int_a^b p_1(x_1) dx_1$
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.

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

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

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

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

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

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

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

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.

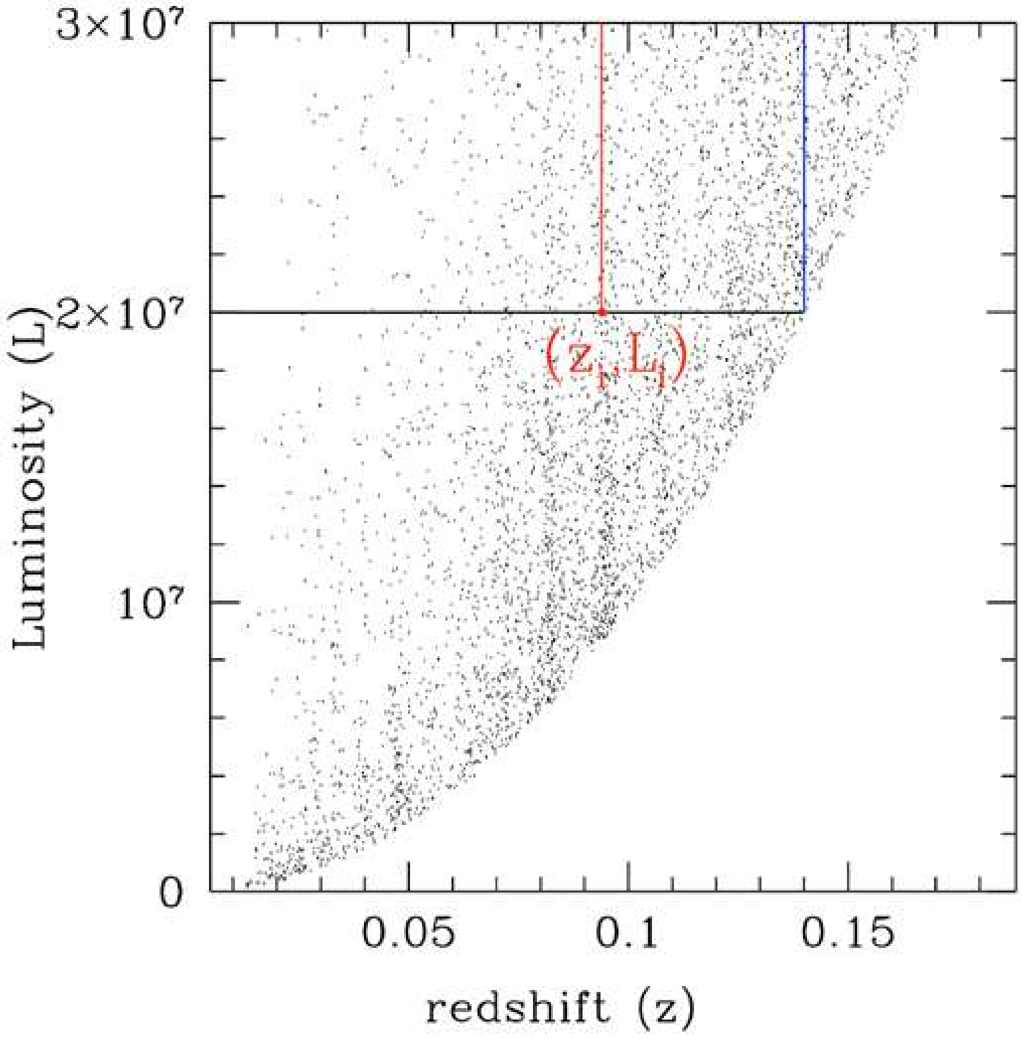
$$p(x_1, x_2) = p_1(x_1)\, p_2(x_2)$$

Example application

Is the **luminosity** distribution of galaxies independent of **redshift**, or do galaxies evolve?

To answer this question is not straightforward. The *observed* distributions of luminosity and redshift are in any case not independent – due to what are known as **selection effects**.

At higher redshift only the more luminous galaxies are bright enough to be detected, so the observed luminosity distribution *appears* to evolve – but is this evolution real?...



Question 4: 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 $p(x,y) \propto \exp\left[-\frac{1}{2}(x+y)\right]; \qquad 0 \le x < y,\ 0 \le y < \infty$

The bivariate normal distribution

Let x and y be RVs with the following joint pdf

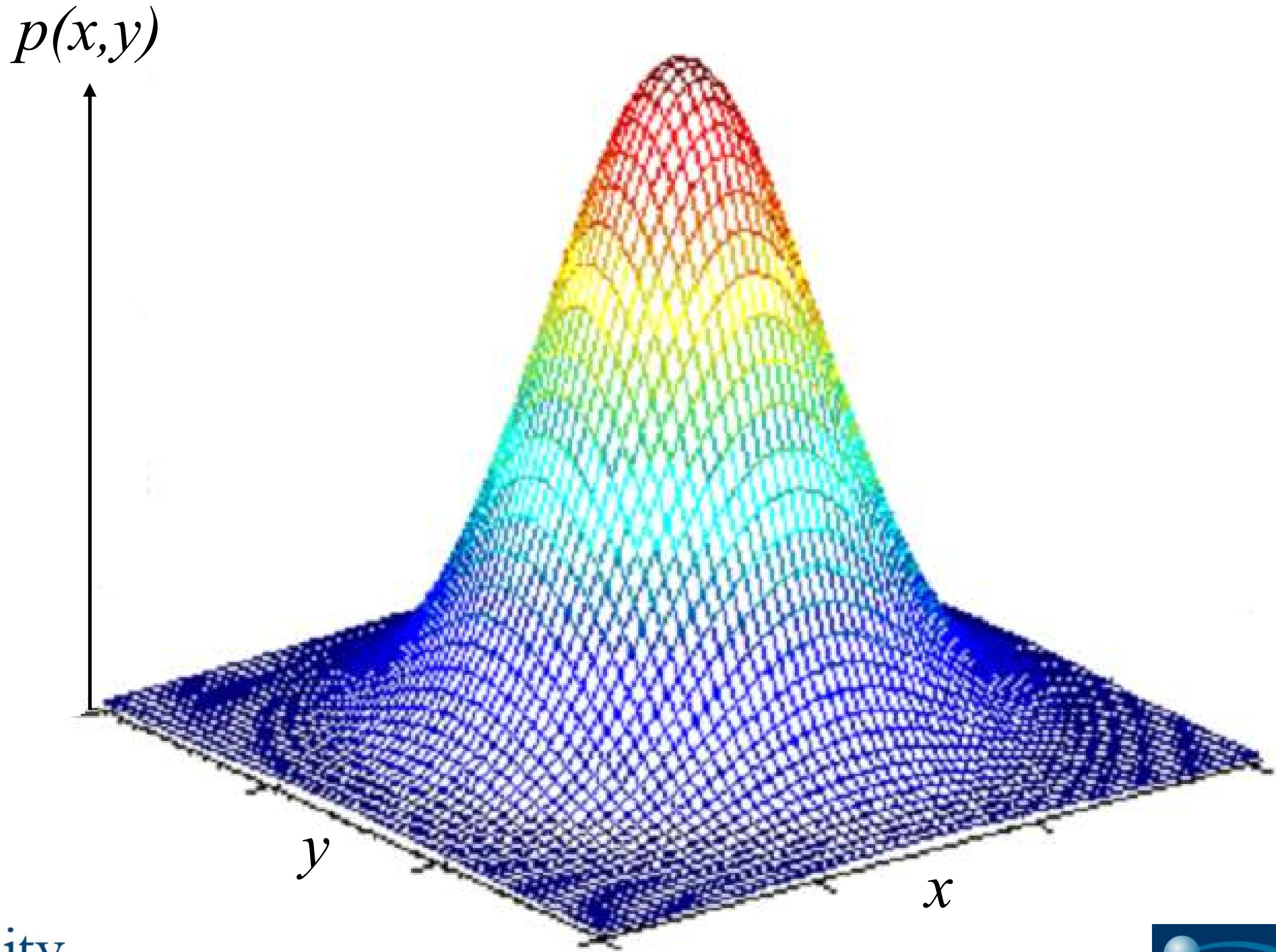
$$p(x,y) \quad = \quad \frac{1}{2\pi\sigma_{\rm x}\sigma_{\rm y}\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}Q(x,y)\right]$$

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

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

Then $p(x,y)$ is known as the **bivariate normal pdf** and is specified by the 5 parameters $\mu_{\rm x}$, $\mu_{\rm y}$, $\sigma_{\rm x}$, $\sigma_{\rm y}$ 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. $E(x) = \mu_{\rm x}$
2. $E(y) = \mu_{\rm y}$
3. $\mathrm{var}(x) = \sigma_{\rm x}^2$
4. $\mathrm{var}(y) = \sigma_{\rm y}^2$

The bivariate normal distribution

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

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

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

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

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 $\mathrm{cov}(x, y)$.

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

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

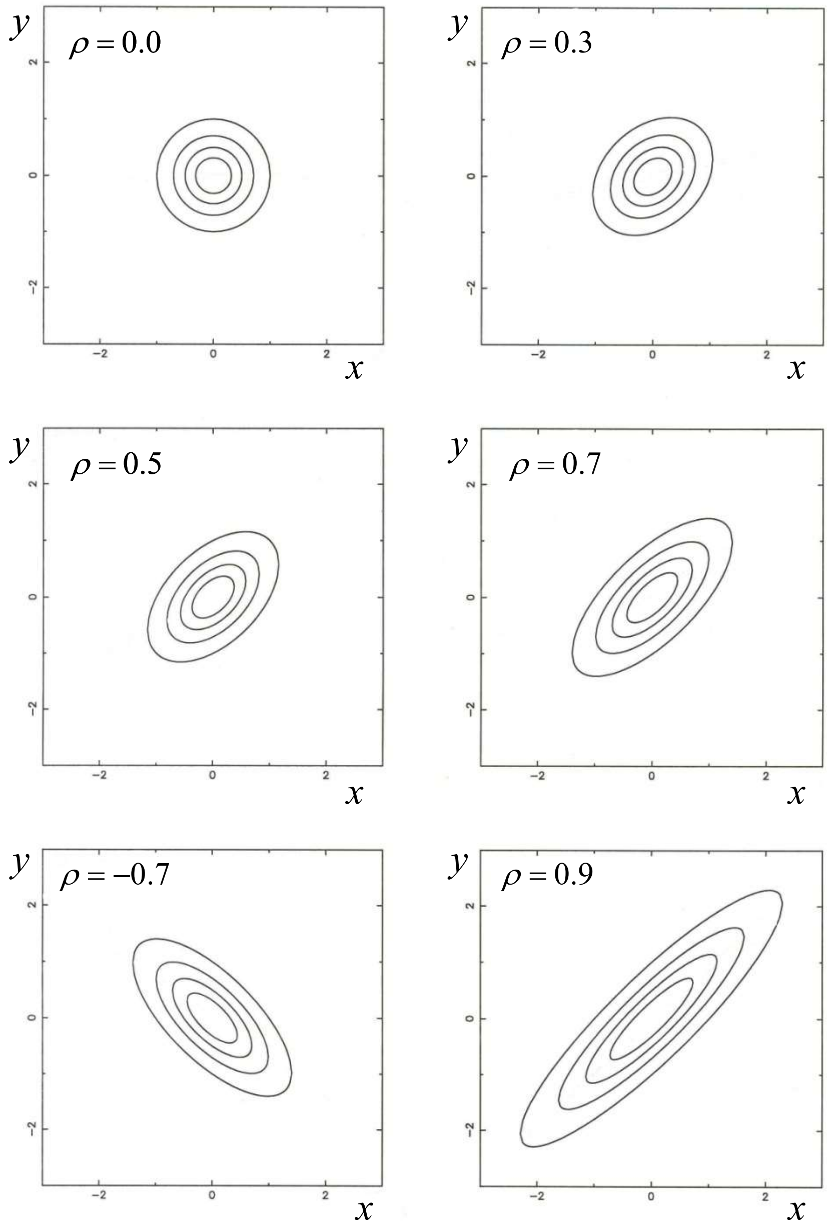
Isoprobability contours for the bivariate normal pdf

$\rho > 0$: 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

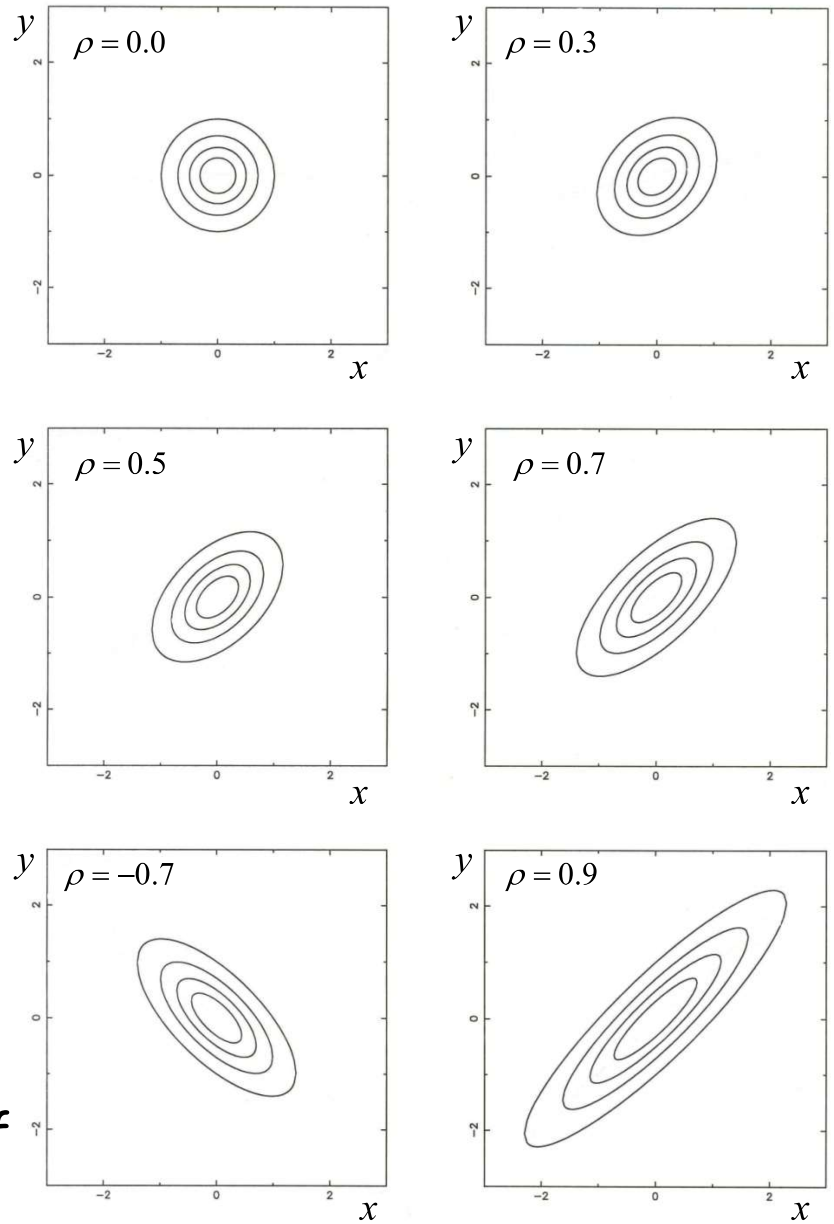
y tends to decrease as x increases

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.

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

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

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

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

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

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