Lectures for the 27th IAU ISYA
Ifrane, 2nd – 23rd July 2004

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

Statistical Astronomy

Martin Hendry,
Dept of Physics and Astronomy
University of Glasgow, UK

http://www.astro.gla.ac.uk/users/martin/isya/

UNIVERSITY of GLASGOW

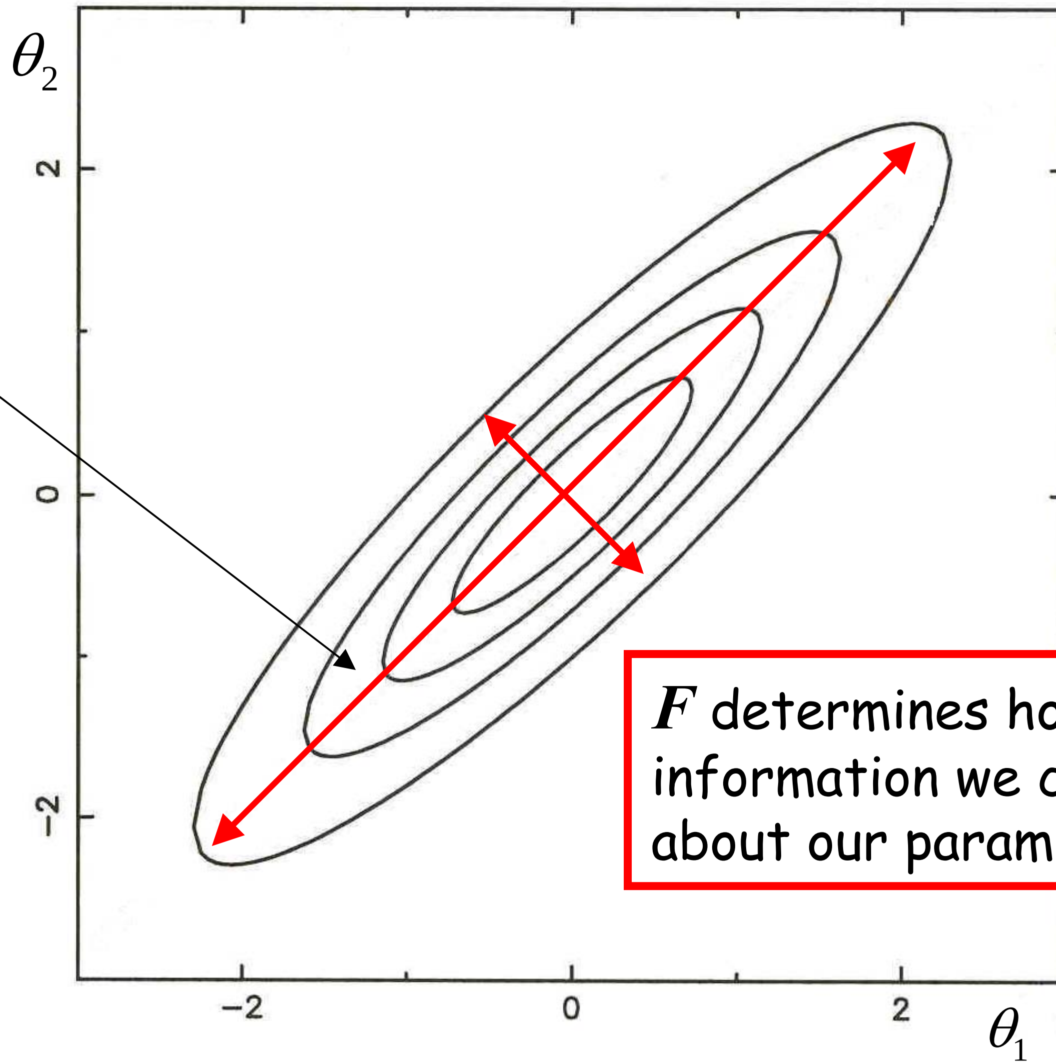
Parameter estimation: 2-D case

Linear combination of θ_1 and θ_2 well constrained by data

Length of axes determined by the **eigenvalues** of the Fisher information matrix

$$F_{ij} = \frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j} = \left[-\sigma_{ij}^2\right]^{-1}$$

$$\boxed{\boldsymbol{F}\,\boldsymbol{\theta} = \lambda\,\boldsymbol{\theta}}$$



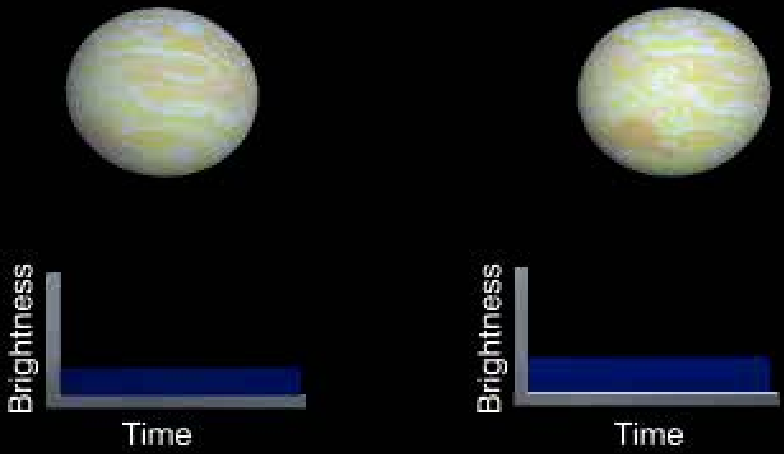
F determines how much information we can learn about our parameters

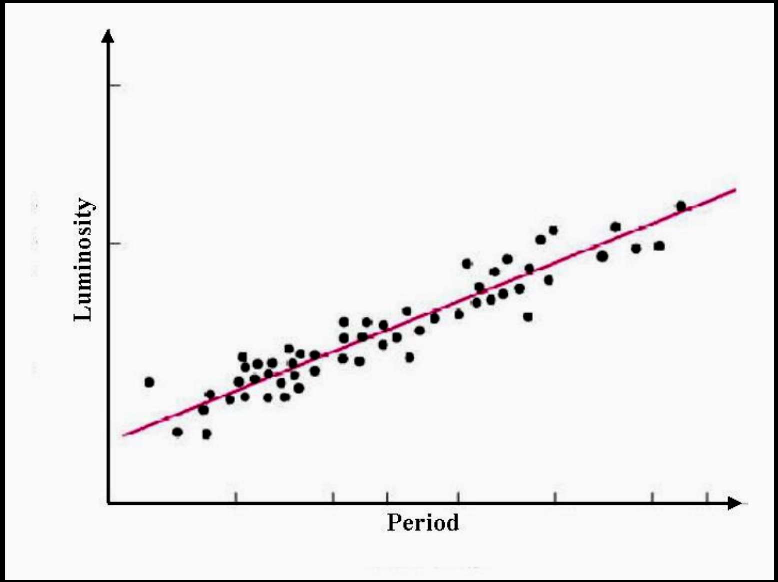
Direction of axes are the *eigenvectors* of $\boldsymbol{F}$

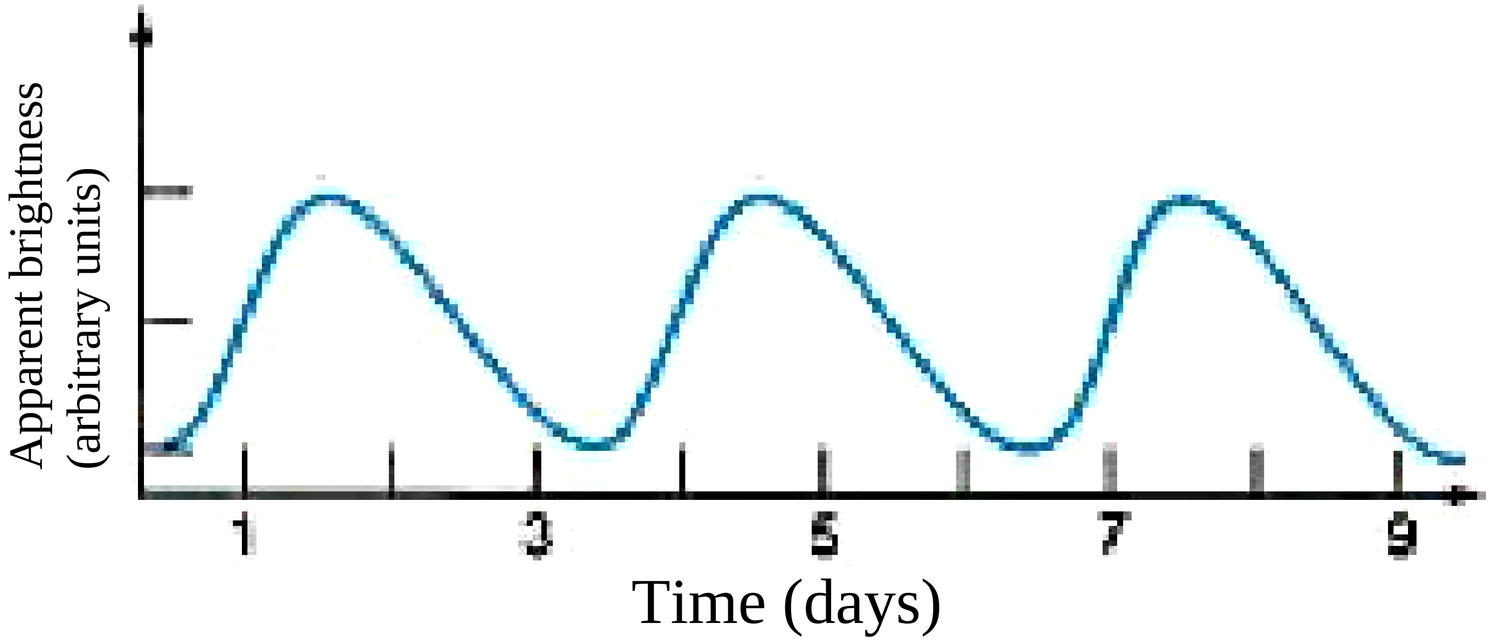
UNIVERSITY of GLASGOW

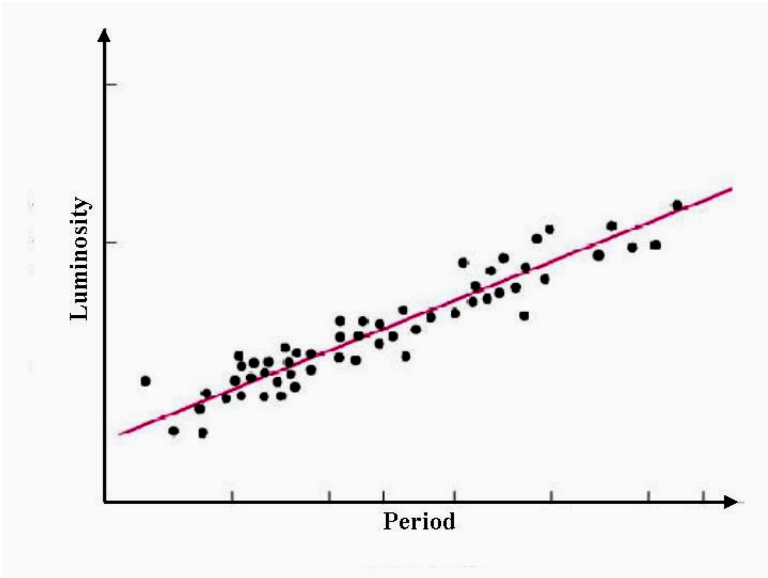
Cepheid Variables: Cosmic Yardsticks

Henrietta Leavitt
1908-1912






Apparent brightness
(arbitrary units)
1
3
5
7
9
Time (days)


Luminosity
Period

Principal Component Analysis templates for Cepheids

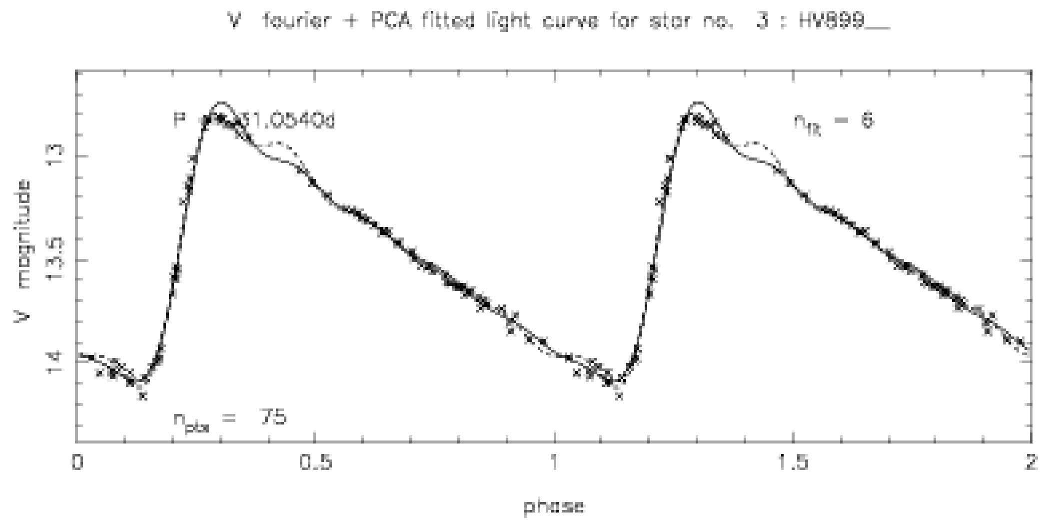
Galactic and LMC V,I data:

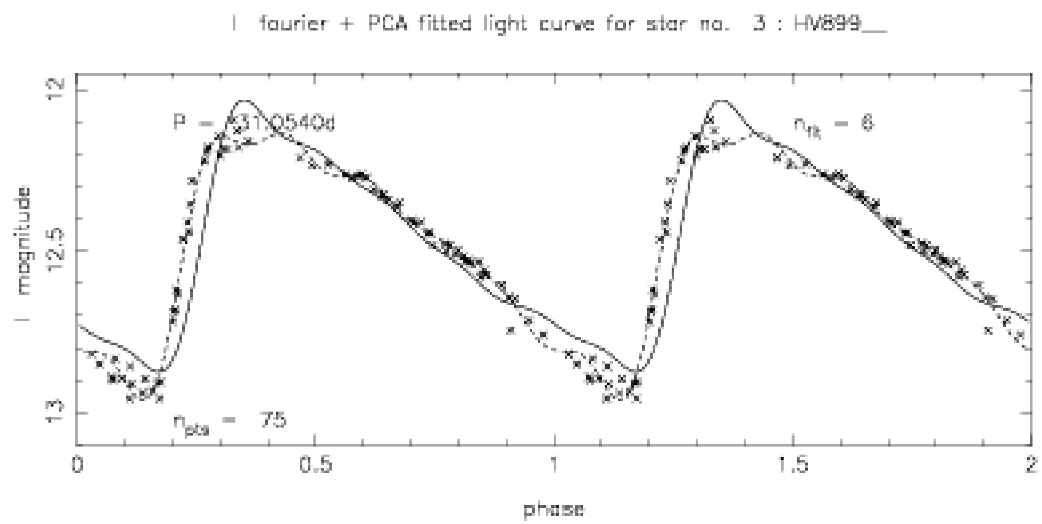
Initially fitted with a 6th order Fourier fit to V and I data $\rightarrow$ 24 parameters

Perform PCA and keep only first two eigenvectors

24-dim problem $\rightarrow$ 2-dim

Can then fit templates to much sparser data



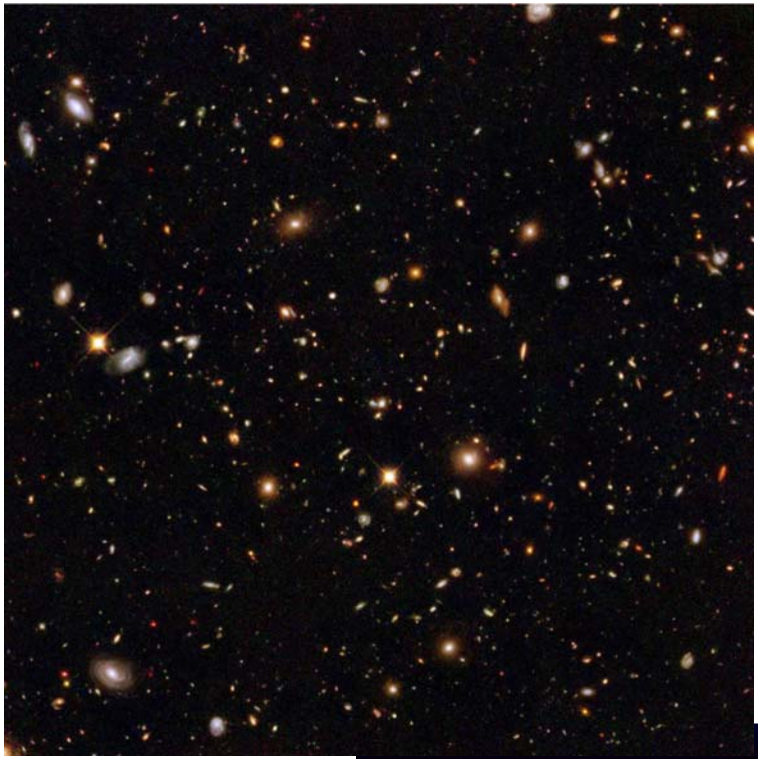


UNIVERSITY of GLASGOW

Dealing with observational selection effects

No matter how good the telescope, there is a limit to the flux that it can reliably detect.

In e.g. galaxy surveys, there is a 'fading out' at large distances



Hubble Ultra Deep Field

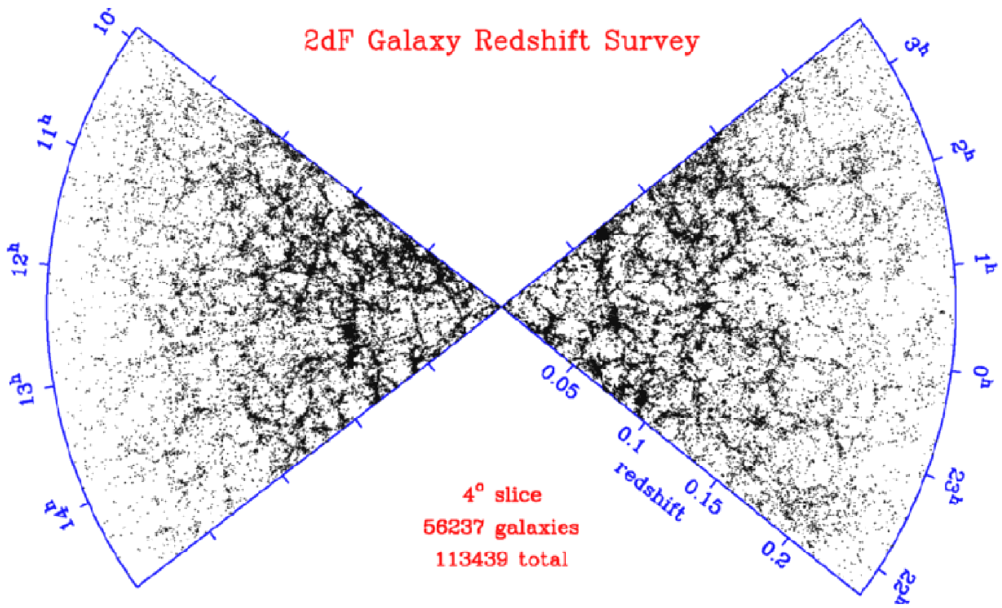
UNIVERSITY of GLASGOW

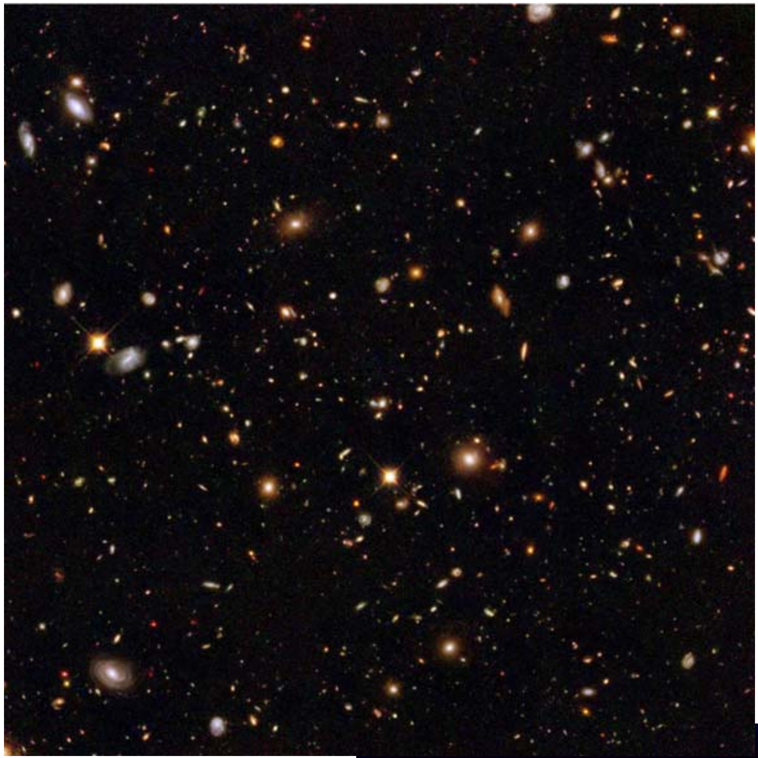
IAU

Dealing with observational selection effects

No matter how good the telescope,
there is a limit to the flux that it
Can reliably detect.

In e.g. galaxy surveys, there is a
'fading out' at large distances





Hubble Ultra Deep Field

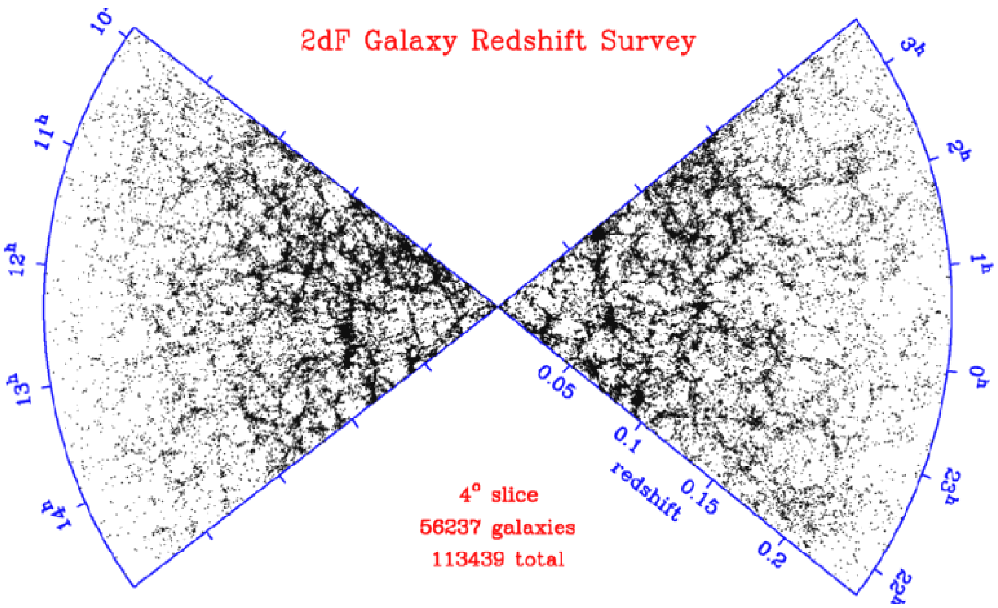
UNIVERSITY of GLASGOW

IAU

Dealing with observational selection effects

No matter how good the telescope, there is a limit to the flux that it Can reliably detect.

In e.g. galaxy surveys, there is a 'fading out' at large distances



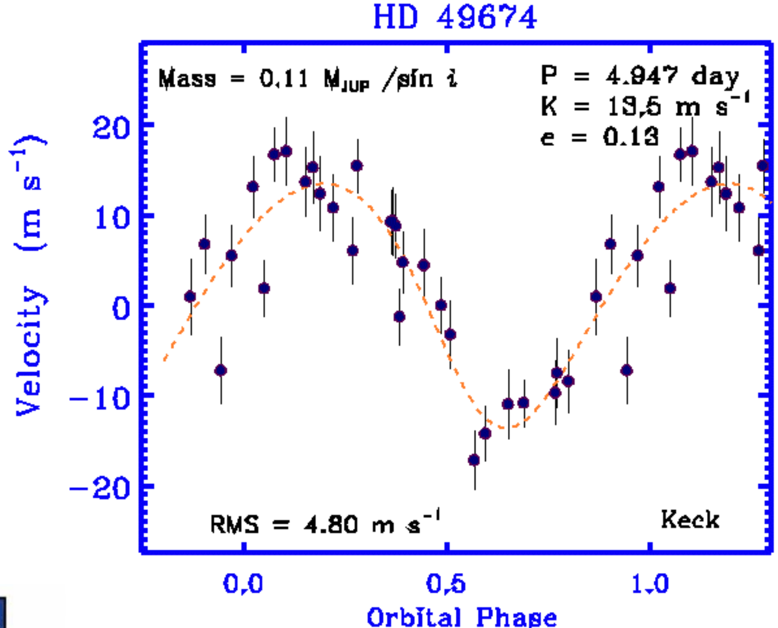
Properties of sampled objects (e.g. luminosity, colour) change with increasing distance

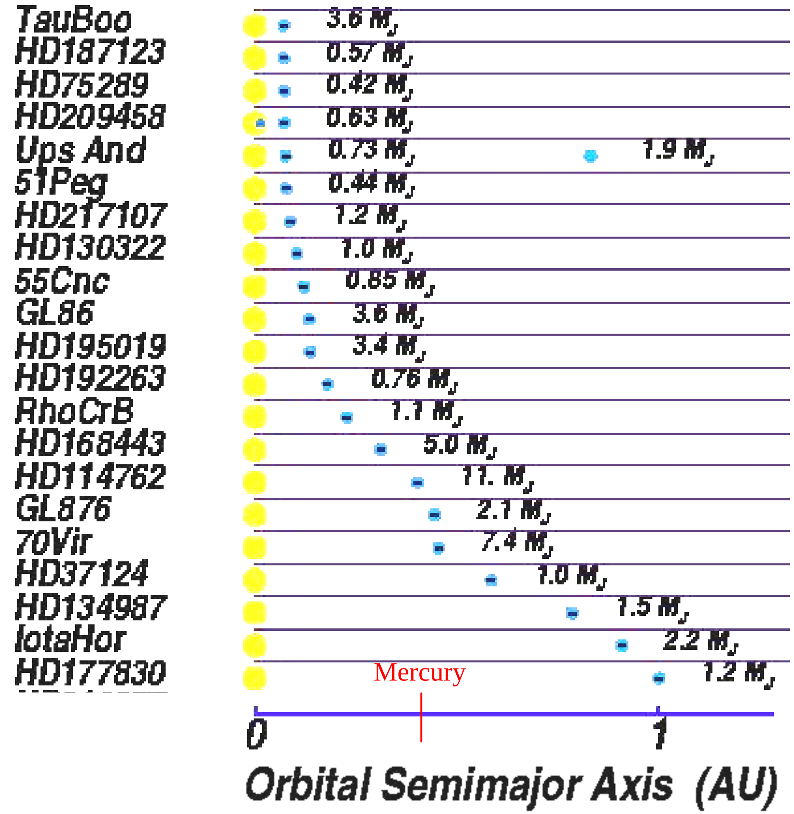
Malmquist bias

Many other examples of observational selection effects in astronomy:

e.g. masses and semimajor axes of extra-solar planets

$$v_S = \left(\frac{2\pi G}{T}\right)^{1/3} m_S^{-2/3} m_P$$





UNIVERSITY of GLASGOW

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$\boxed{p_{\mathrm{obs}}(\mathrm{data}\,|\,\mathrm{model},I) \quad = \quad p(\mathrm{data}\,|\,\mathrm{model},I)\times S(\mathrm{data},I)}$$

The 'actual' likelihood

The 'ideal' likelihood

The selection function

UNIVERSITY *of* GLASGOW

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$\boxed{p_{\text{obs}}\left(\text{data}\,|\,\text{model}, I\right) \;\propto\; p\left(\text{data}\,|\,\text{model}, I\right)\times S\left(\text{data}, I\right)}$$

The 'actual' likelihood The 'ideal' likelihood The selection function

Selection function measures the probability that an object with particular data characteristics* would be observable

(* e.g. apparent magnitude, colour, surface brightness, angular size)

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$\boxed{p_{\mathrm{obs}}(\mathrm{data} \mid \mathrm{model}, I) \;\propto\; p(\mathrm{data} \mid \mathrm{model}, I) \times S(\mathrm{data}, I)}$$

The 'actual' likelihood   The 'ideal' likelihood   The selection function

$$\boxed{p(\mathrm{data} \mid \mathrm{model}, I) \;\propto\; \frac{p_{\mathrm{obs}}(\mathrm{data} \mid \mathrm{model}, I)}{S(\mathrm{data}, I)}}$$

Problems: *need to know* $S(\mathrm{data}, I)$ *accurately*

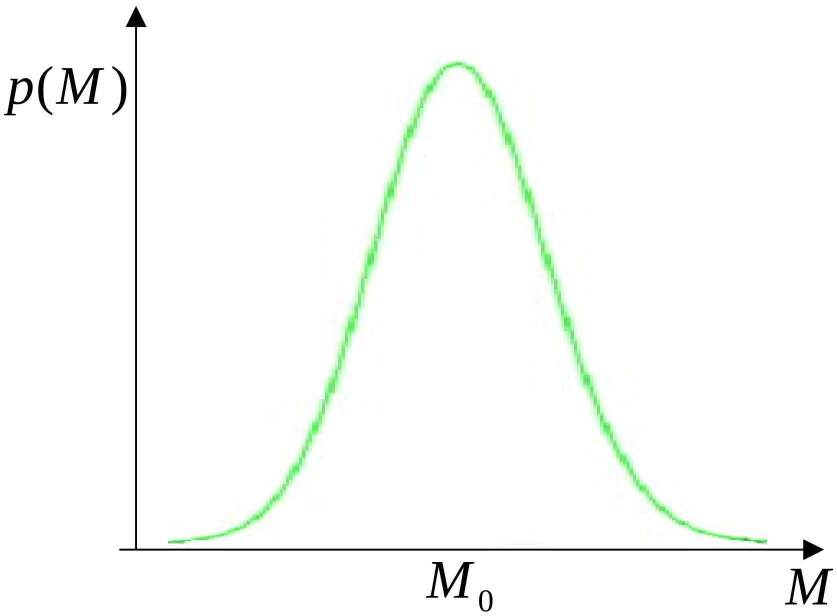
$S(\mathrm{data}, I)$ *may depend on* different *data than the likelihood functon*

UNIVERSITY *of* GLASGOW

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$

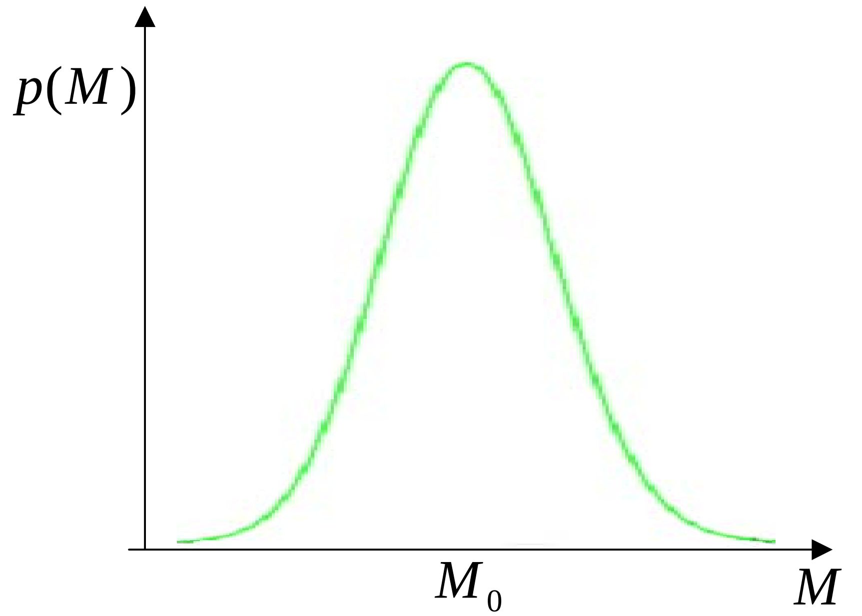


UNIVERSITY of GLASGOW

Dealing with observational selection effects

Example: Galaxy **luminosity function**

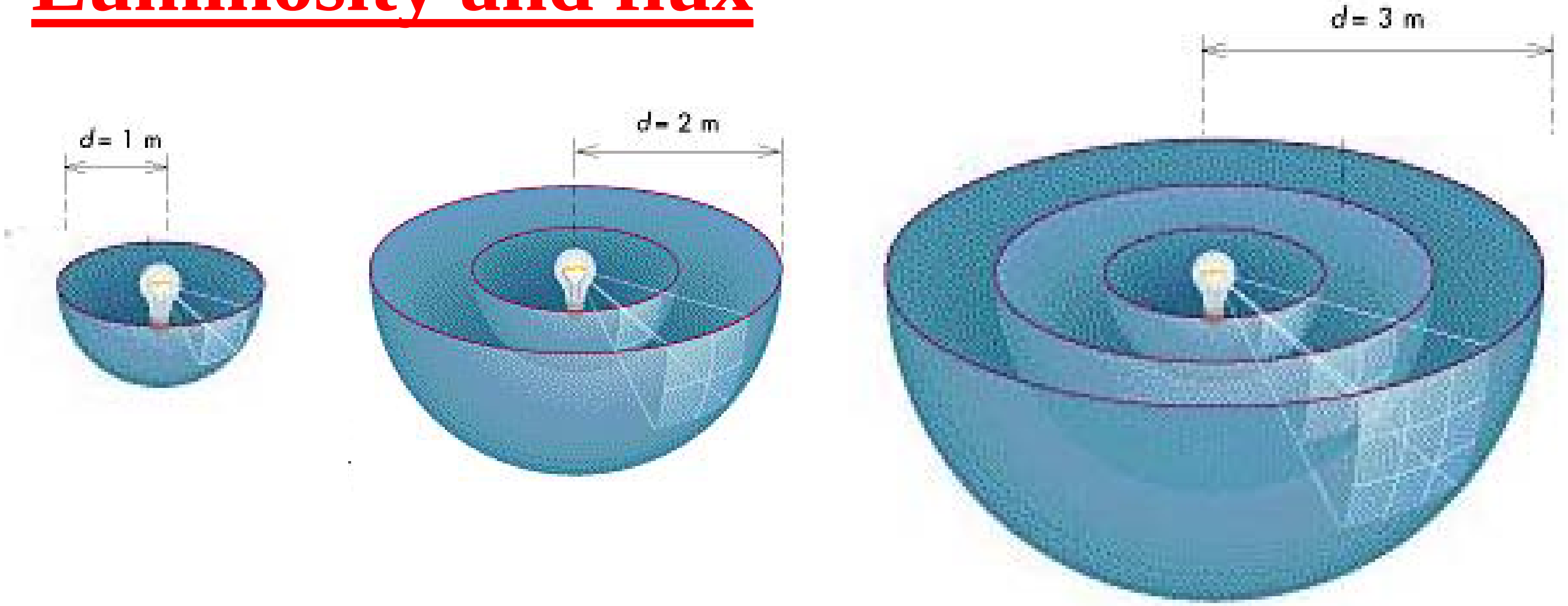
$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from
The apparent magnitude and distance (modulus)

UNIVERSITY of GLASGOW

Luminosity and flux



Apparent brightness, or **flux**, falls off with the square of the distance, because the surface area of a sphere increases with the square of its radius

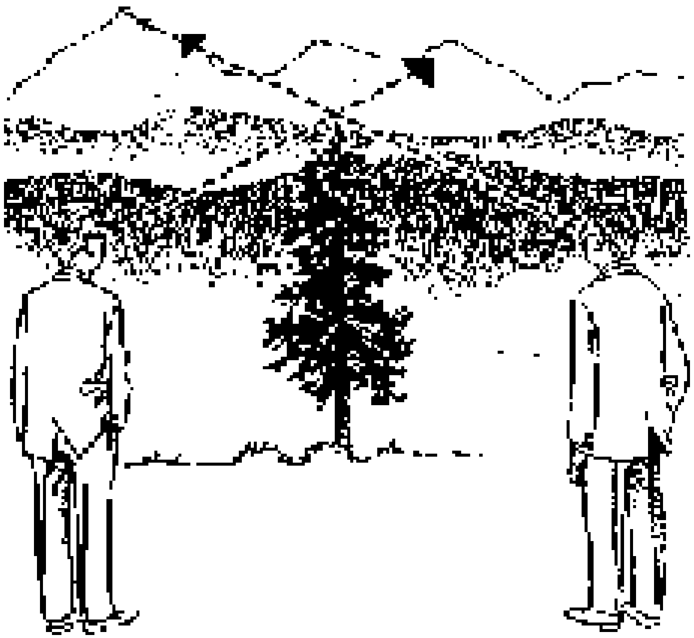
Distance, (metres)

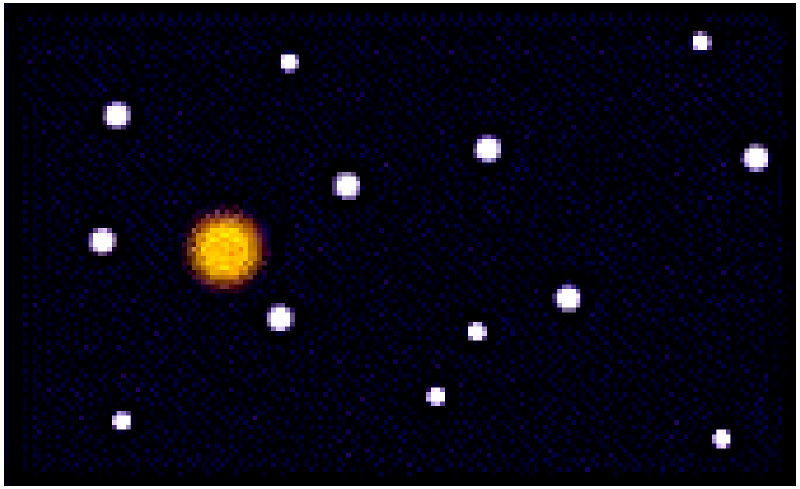
$$\boxed{L = 4\pi D^2 F}$$

Luminosity, (watts)

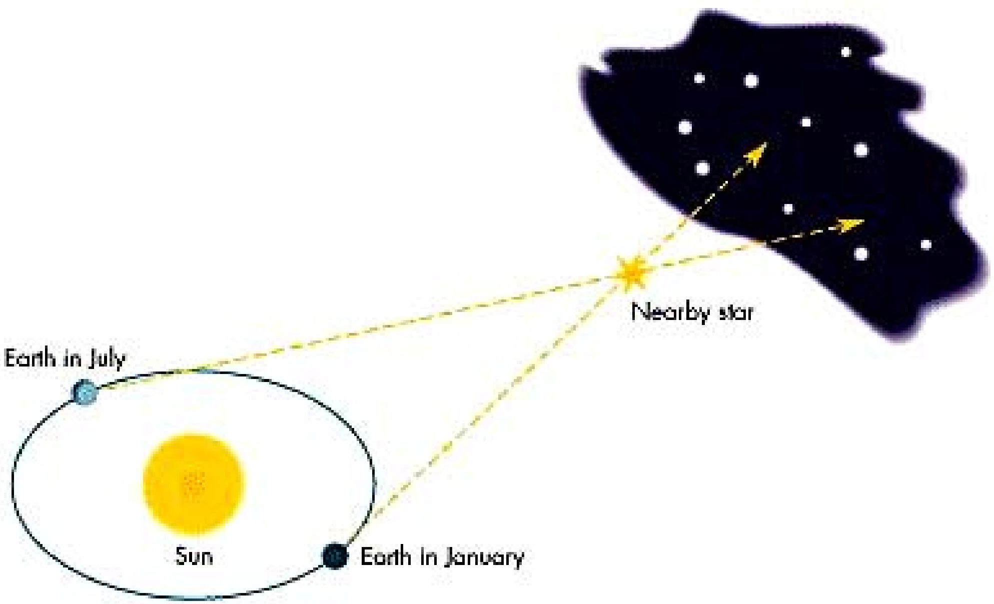
Flux, (watts / square metre)

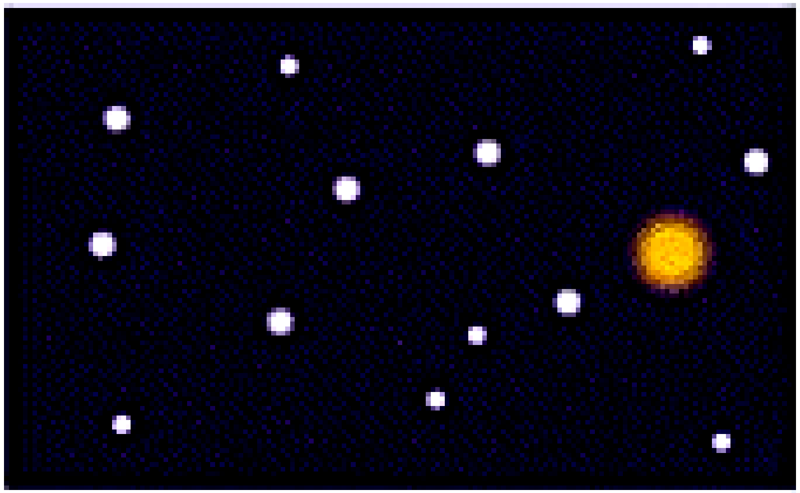
Measuring Astronomical Distances: Parallax





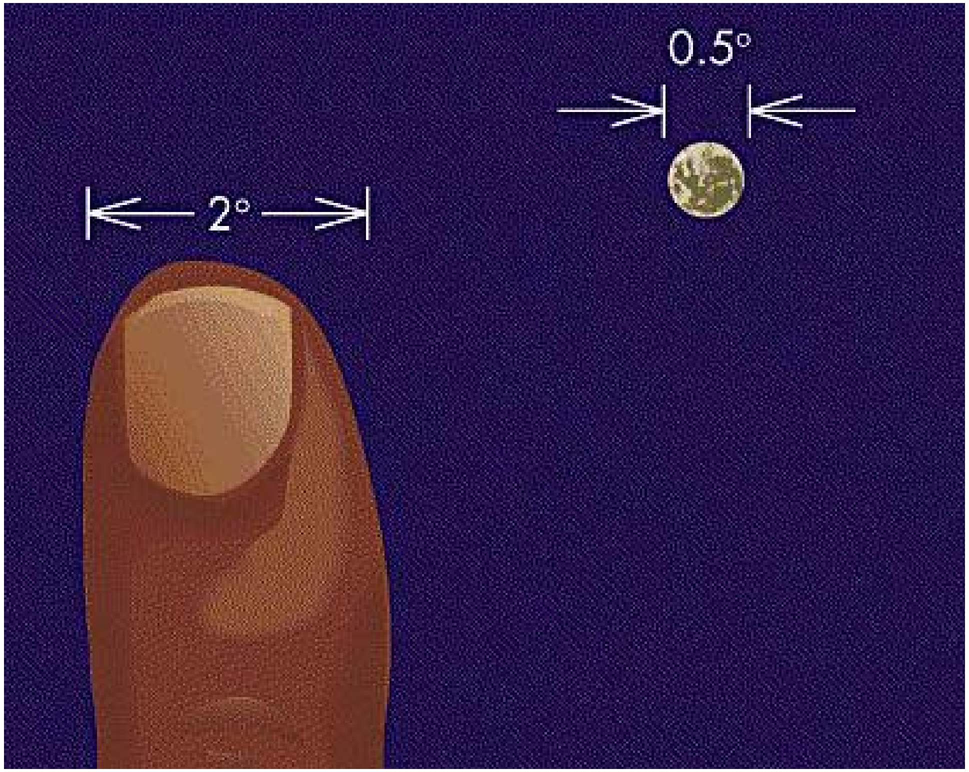
View from the Earth in January

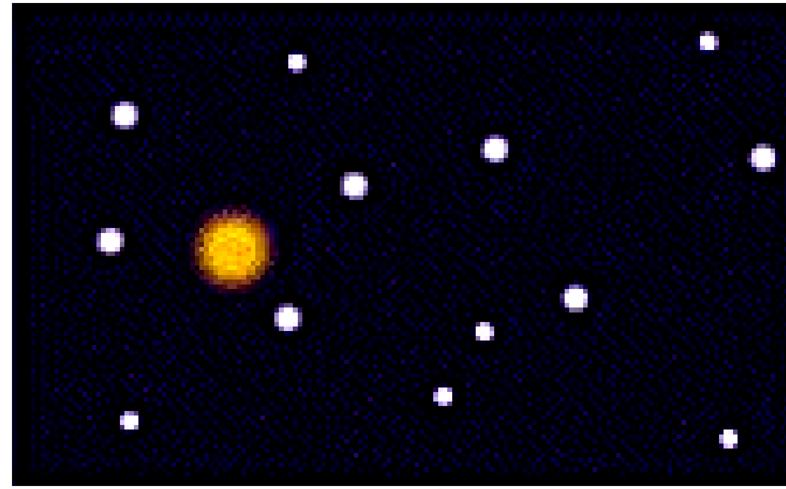




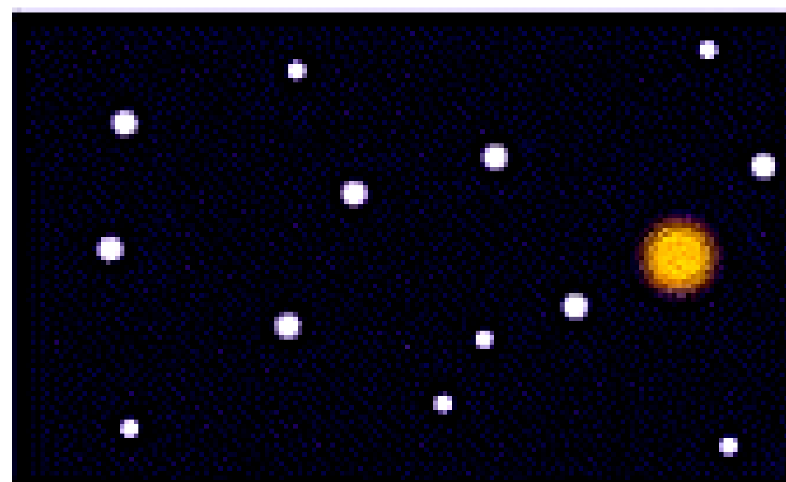
View from the Earth in July

Measuring Astronomical Distances: Parallax





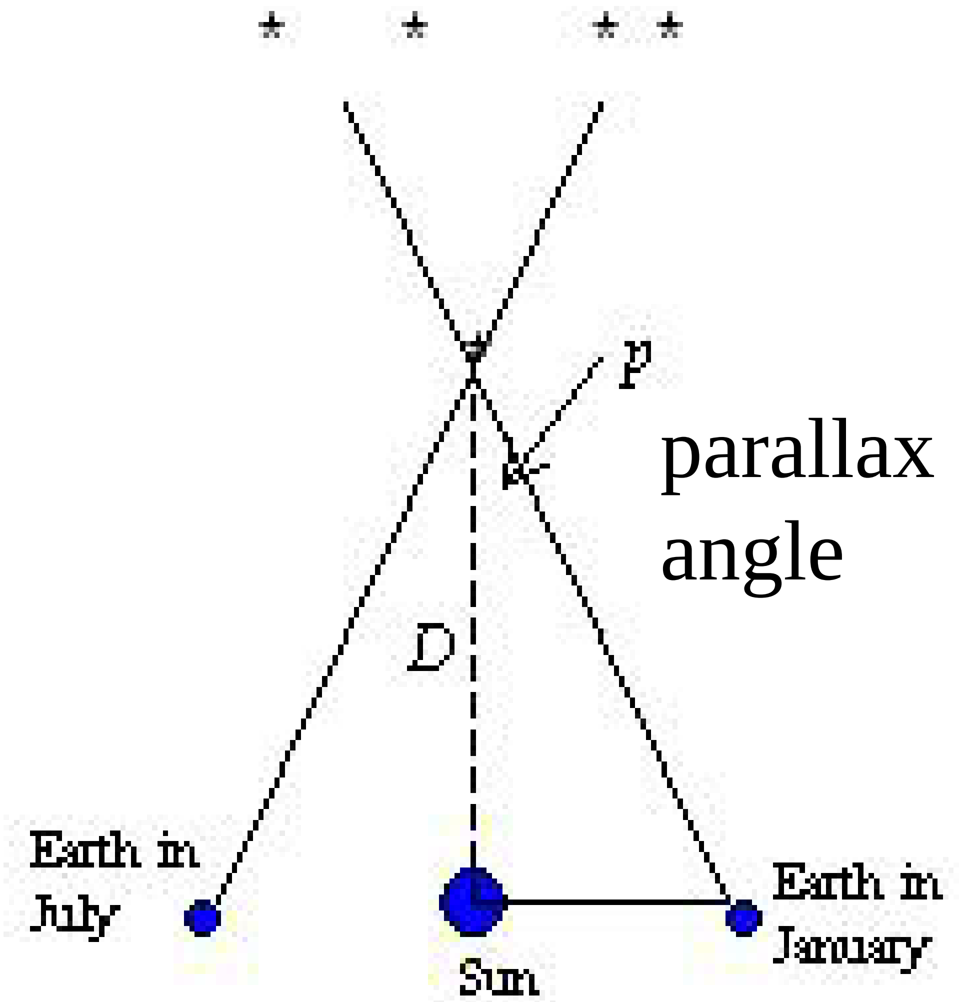
View from the Earth in January



View from the Earth in July

Even the nearest star shows a parallax shift of only 1/2000th the width of the full Moon

Measuring Astronomical Distances: Parallax

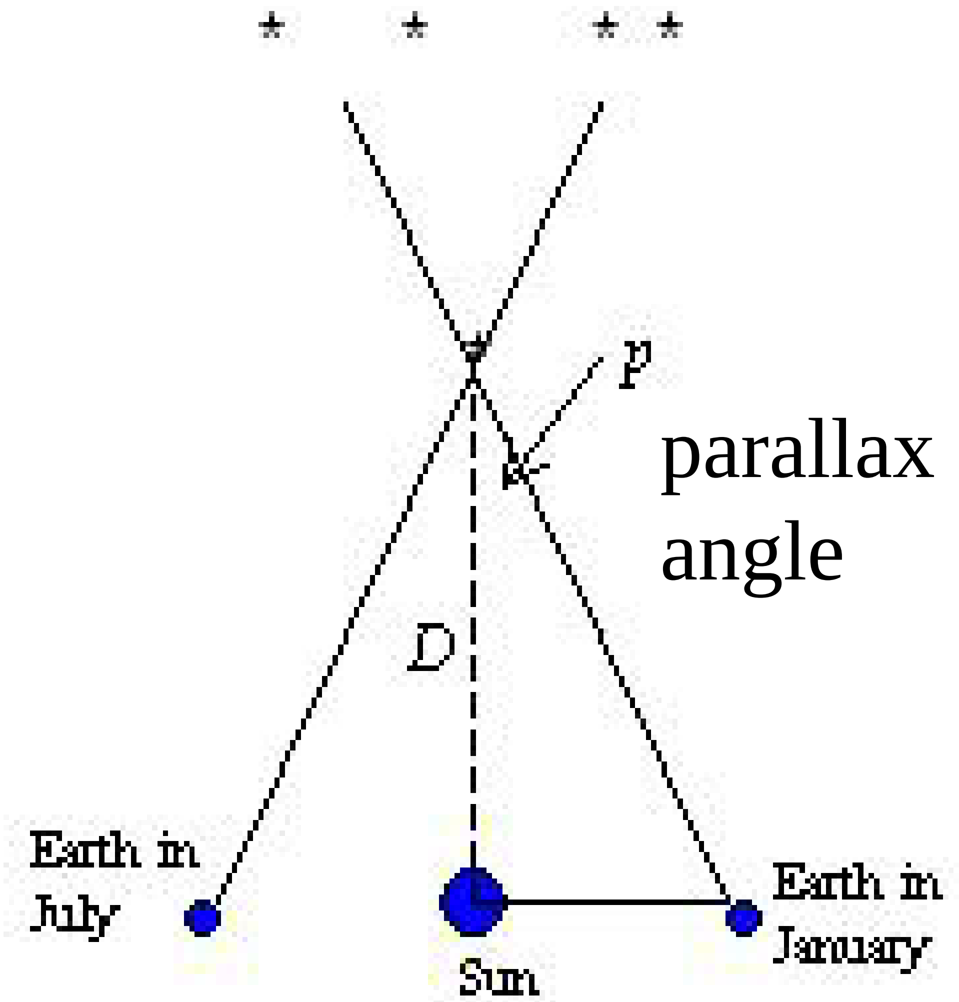


$$D = \frac{1}{\tan p} \cong \frac{1}{p} \ \text{A.U.}$$

$$D = \frac{206265}{p''} \ \text{A.U.}$$

A star at a distance of 1 parsec shows a parallax angle of one arc second

Measuring Astronomical Distances: Parallax



parallax angle

$$D = \frac{1}{\tan p} \cong \frac{1}{p} \ \text{A.U.}$$

$$D = \frac{206265}{p''} \ \text{A.U.}$$

$$1\,\text{pc} = 206265\ \text{A.U.}$$

$$= 3.086 \times 10^{16}\ \text{m}$$

$$= 3.262\ \text{light years}$$

Apparent and Absolute Magnitude

Expressing flux in terms of distance and luminosity:-

$$m_1 - m_2 = -2.5\log_{10}\frac{4\pi D_2^2 L_1}{4\pi D_1^2 L_2}$$

$$= 5\log_{10} D_1 - 5\log_{10} D_2$$
$$+ 2.5\log_{10} L_2 - 2.5\log_{10} L_1$$

Suppose L_1 and L_2 are equal:-

$$m_1 = m_2 + 5\log_{10} D_1 - 5\log_{10} D_2$$

Apparent and Absolute Magnitude

Absolute magnitude = apparent magnitude which a star *would* have if it were at a distance of **ten parsecs**

$$\boxed{m = M + 5\log_{10} D - 5}$$

$$m - M = \mu \quad = \text{distance modulus}$$

Apparent and Absolute Magnitude

In cosmology we often measure distances in **Megaparsecs**

$$\boxed{1\,\text{Mpc} = 1\,\text{million parsecs} = 10^6\ \text{pc}}$$

Apparent and Absolute Magnitude

In cosmology we often measure distances in **Megaparsecs**

$$1\,\text{Mpc} = 1\,\text{million parsecs} = 10^6\,\text{pc}$$

in Mpc

$$m = M + 5\log_{10}(D \times 10^6) - 5$$

$$= M + 5\log_{10} D + 5\log_{10} 10^6 - 5$$

Apparent and Absolute Magnitude

In cosmology we often measure distances in **Megaparsecs**

$$\boxed{1\,\text{Mpc} = 1\,\text{million parsecs} = 10^6\ \text{pc}}$$

in Mpc

$$m = M + 5\log_{10}(D \times 10^6) - 5$$

$$= M + 5\log_{10} D + 5\log_{10} 10^6 - 5$$

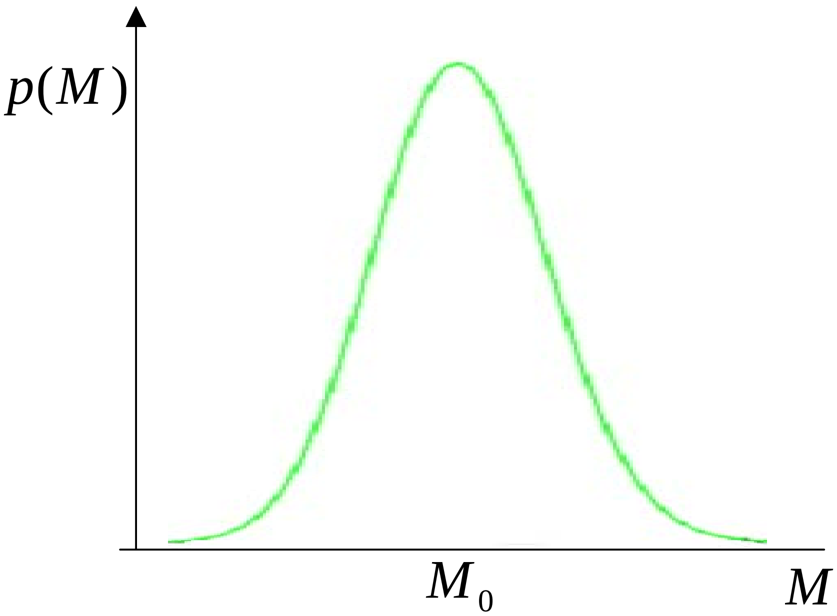
in Mpc

$$\boxed{m = M + 5\log_{10} D + 25}$$

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from
The apparent magnitude and distance (modulus)

Simplest form of observational selection:

sharp apparent magnitude limit

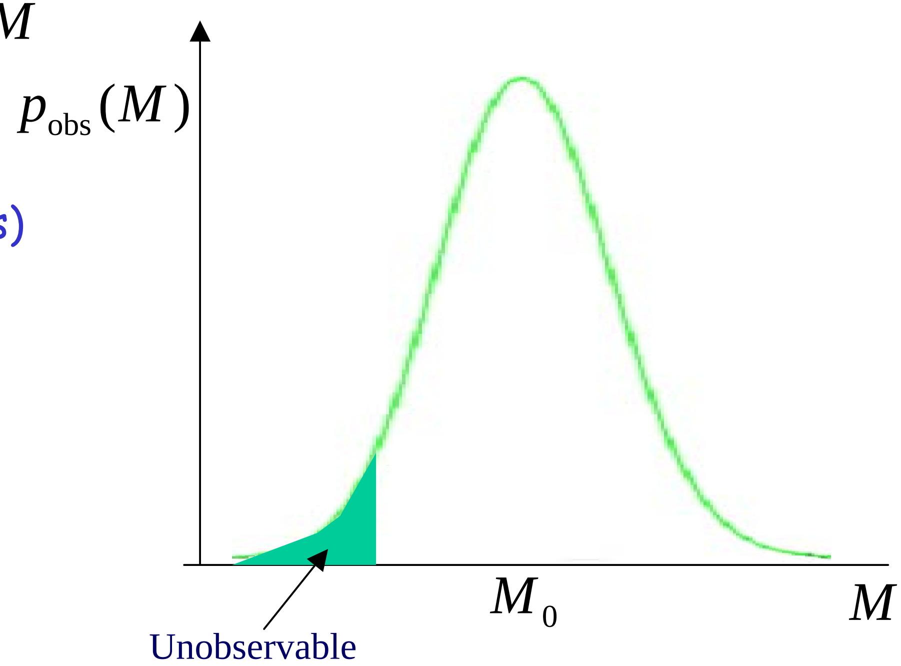
$$S(m) = \begin{cases} 1 & \text{if} \quad m < m_{\mathrm{LIM}} \\ 0 & \text{otherwise} \end{cases}$$

UNIVERSITY of GLASGOW

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from
The apparent magnitude and distance (modulus)

Simplest form of observational selection:

sharp apparent magnitude limit

$$\boxed{S(m) = \begin{cases} 1 & \text{if} \quad m < m_{\text{LIM}} \\ 0 & \text{otherwise} \end{cases}}$$

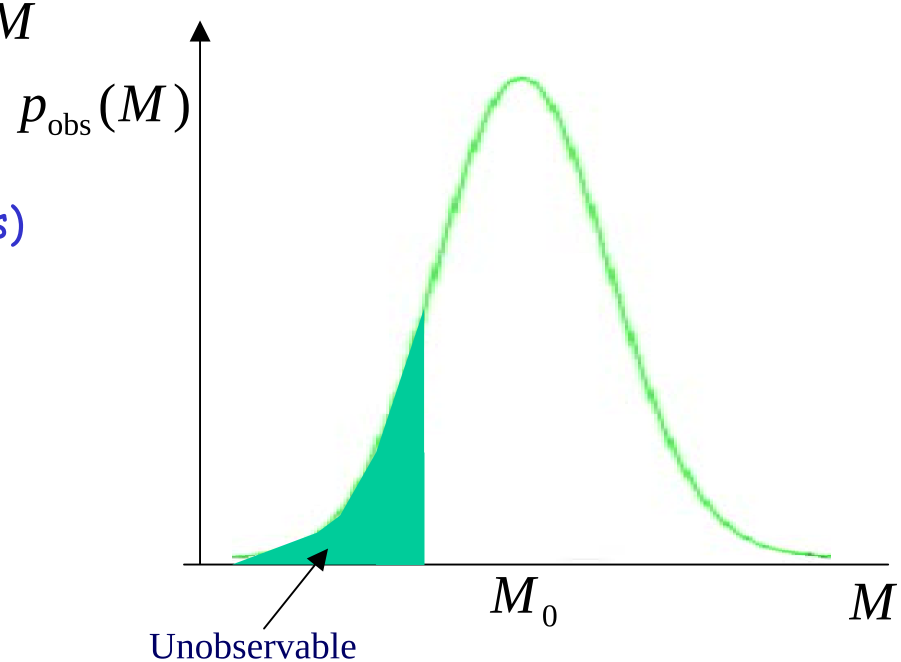
$$M_{\text{LIM}} = m_{\text{LIM}} - \mu$$

Distance modulus

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from
The apparent magnitude and distance (modulus)

Simplest form of observational selection:

sharp apparent magnitude limit

$$S(m) = \begin{cases} 1 & \text{if} \quad m < m_{\text{LIM}} \\ 0 & \text{otherwise} \end{cases}$$

$$M_{\text{LIM}} = m_{\text{LIM}} - \mu$$

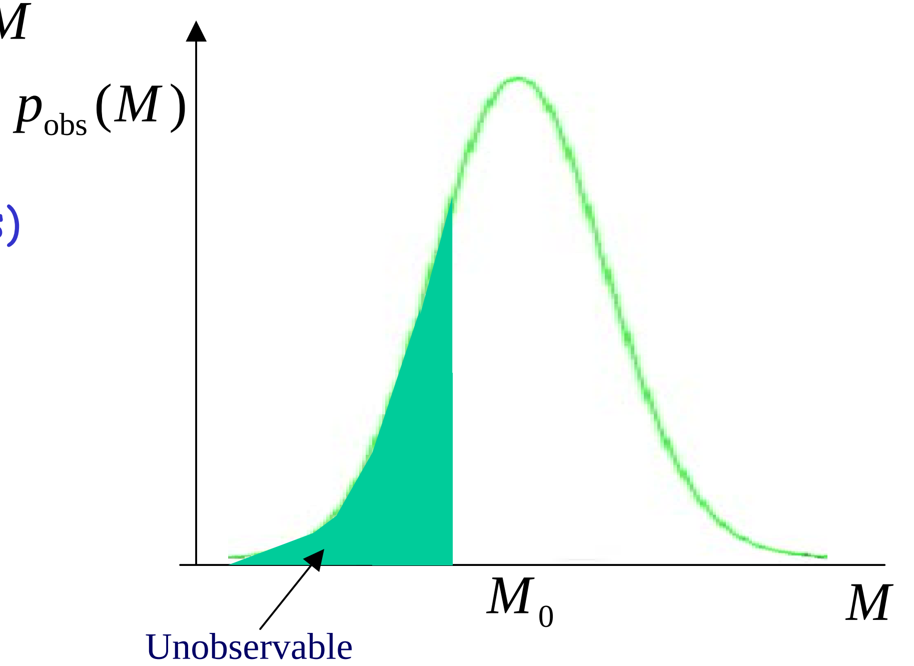
Distance modulus

UNIVERSITY of GLASGOW

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from
The apparent magnitude and distance (modulus)

Simplest form of observational selection:

sharp apparent magnitude limit

$$S(m) = \begin{cases} 1 & \text{if} \quad m < m_{\text{LIM}} \\ 0 & \text{otherwise} \end{cases}$$

$$M_{\text{LIM}} = m_{\text{LIM}} - \mu$$

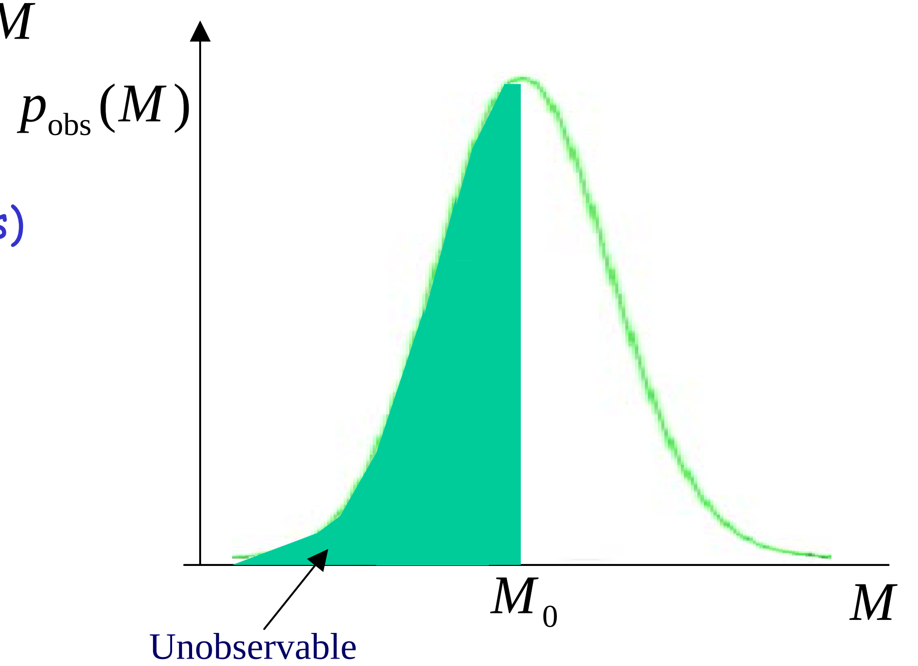
Distance modulus

UNIVERSITY of GLASGOW

Dealing with observational selection effects

Example: Galaxy **luminosity function**

$p(M)dM$ = fraction of galaxies with absolute magnitude between M and $M + dM$



But we don't observe M. We infer it from The apparent magnitude and distance (modulus)

Simplest form of observational selection:

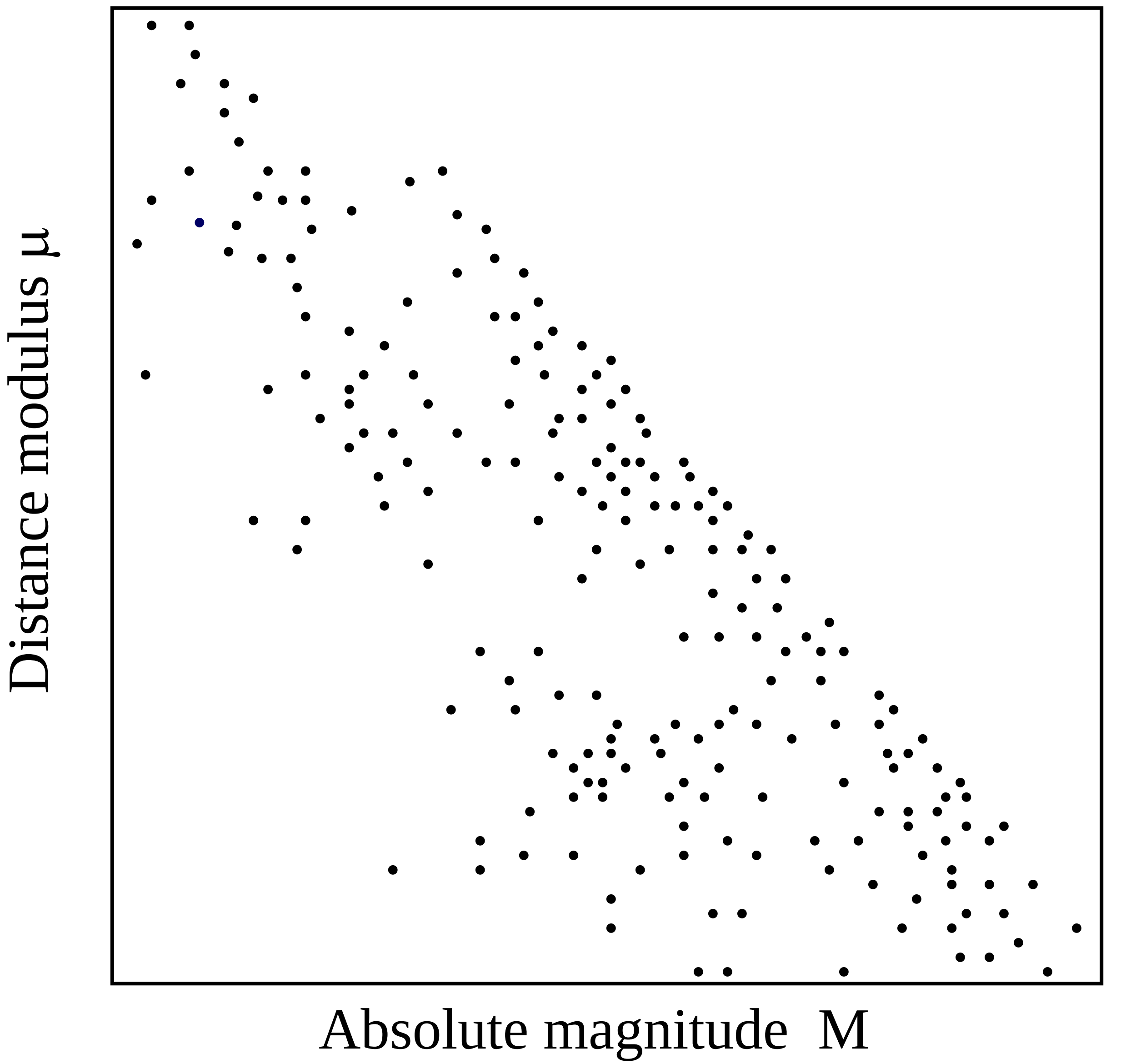
sharp apparent magnitude limit

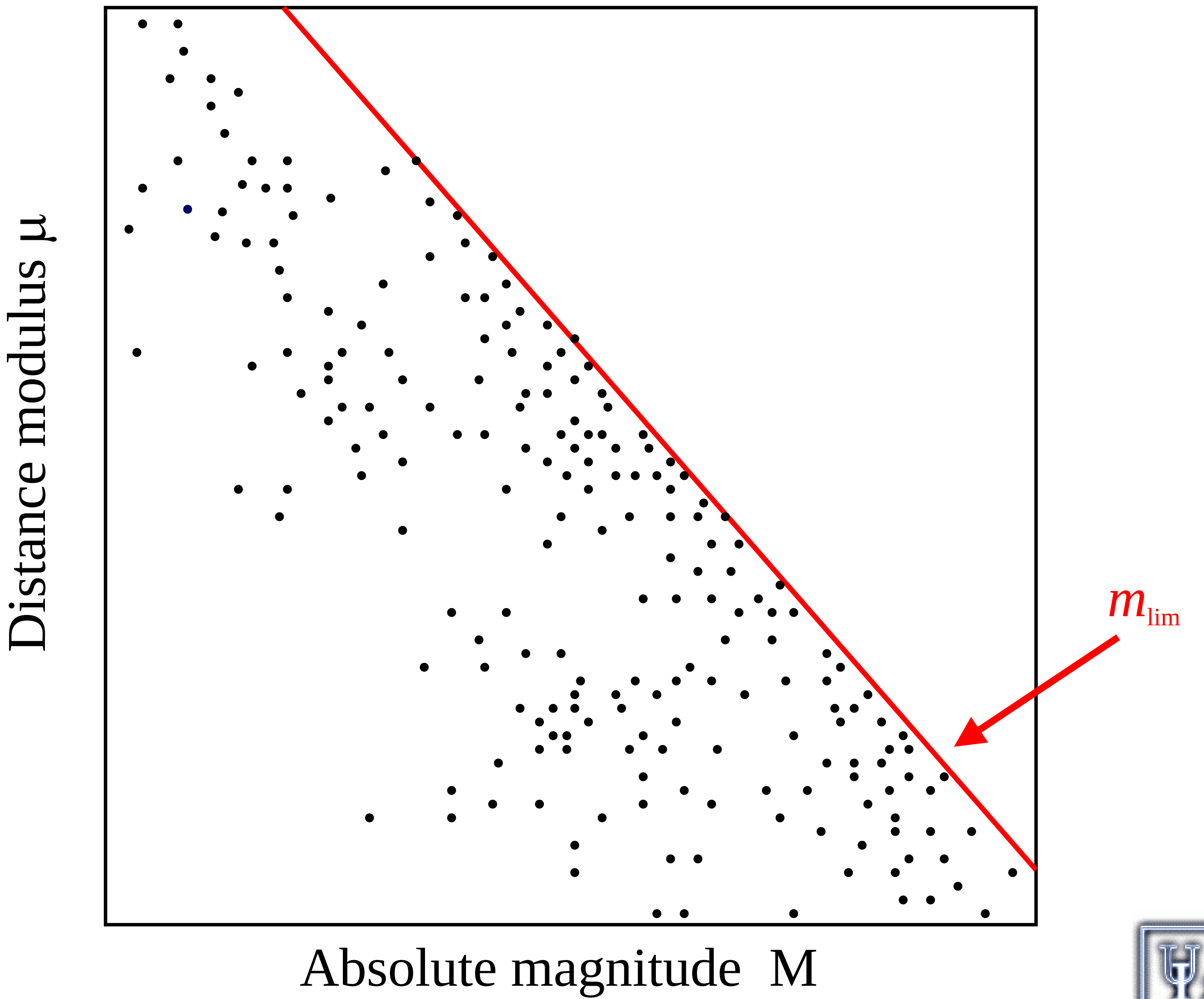
$$S(m) = \begin{cases} 1 & \text{if} \quad m < m_{\mathrm{LIM}} \\ 0 & \text{otherwise} \end{cases}$$

$$M_{\mathrm{LIM}} = m_{\mathrm{LIM}} - \mu$$

Distance modulus

UNIVERSITY of GLASGOW


Distance modulus μ
Absolute magnitude M



Distance modulus μ
m_{lim}
Absolute magnitude M

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$p_{\text{obs}}(\text{data}\,|\,\text{model},I) \;\propto\; p(\text{data}\,|\,\text{model},I)\times S(\text{data},I)$$

Need to integrate out over distance modulus ('nuisance parameter'), since the selection function depends on both M and μ

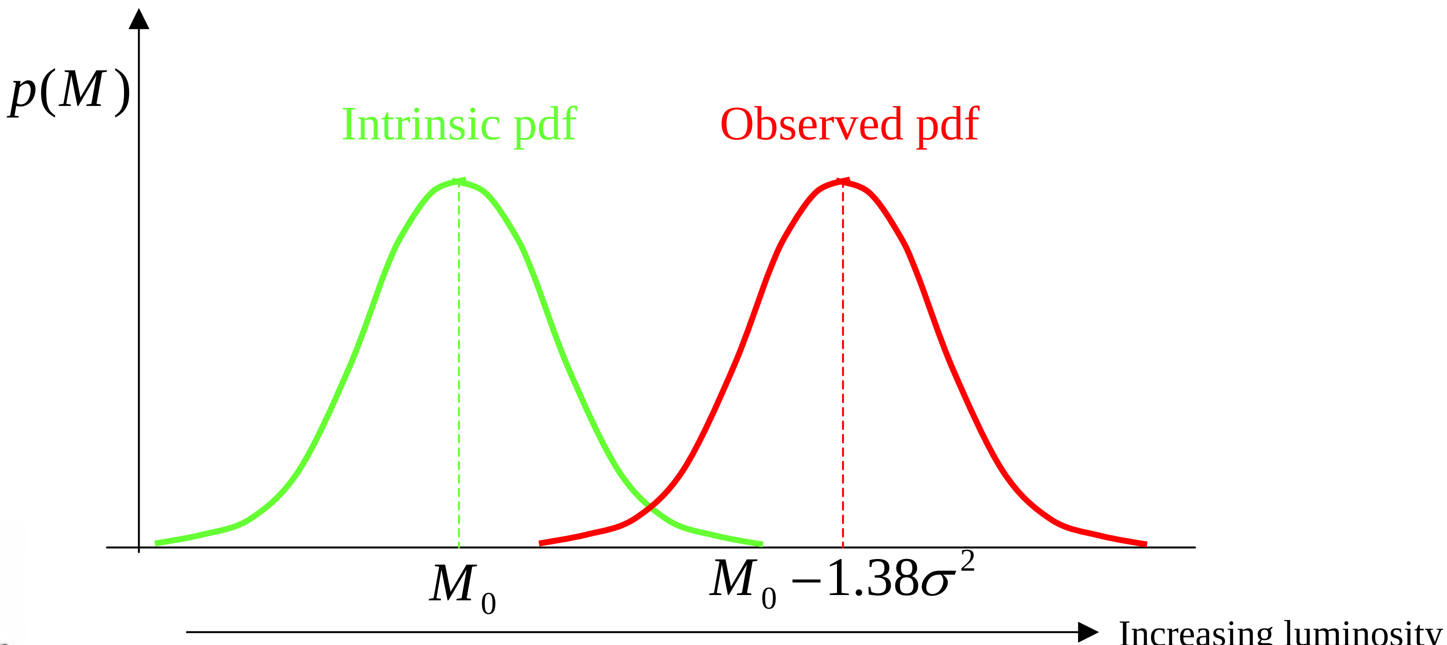
$$p_{\text{obs}}(M) \;\propto\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\tfrac{1}{2}\left[\frac{M-M_0}{\sigma}\right]^2 \right) p(\mu) S(M,\mu)\, d\mu$$

UNIVERSITY of GLASGOW

To make any further progress we need to adopt a model for $p(\mu)$

Assuming that galaxies are uniformly distributed in space, we can show that

$$p_{\rm obs}(M) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\tfrac{1}{2}\left[\frac{M-\left(M_0-1.38\sigma^2\right)}{\sigma}\right]^2\right)$$



UNIVERSITY of GLASGOW

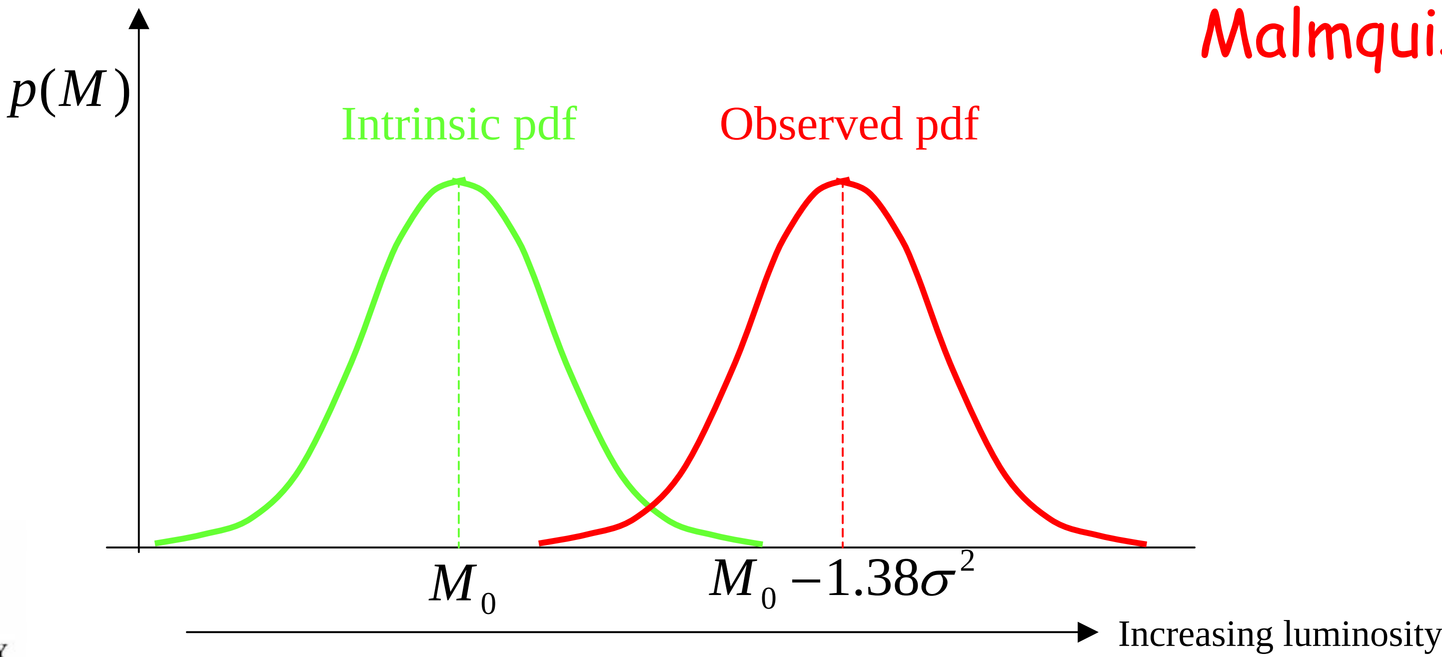
To make any further progress we need to adopt a model for $p(\mu)$

Observed galaxies are intrinsically more luminous

Assuming that galaxies are uniformly distributed in space, we can show that

$$p_{\rm obs}(M) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\tfrac{1}{2} \left[\frac{M - \left(M_0 - 1.38\sigma^2\right)}{\sigma} \right]^2 \right)$$

Malmquist Bias



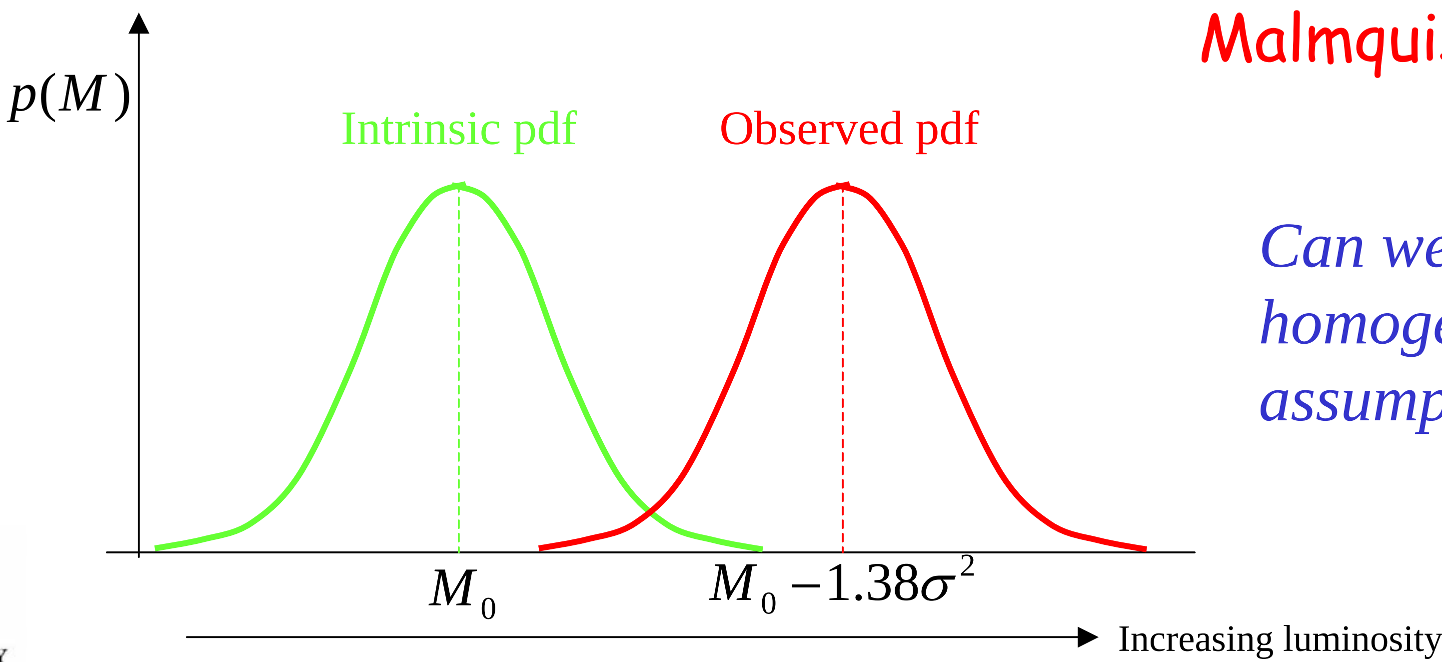
To make any further progress we need to adopt a model for $p(\mu)$

Assuming that galaxies are uniformly distributed in space, we can show that

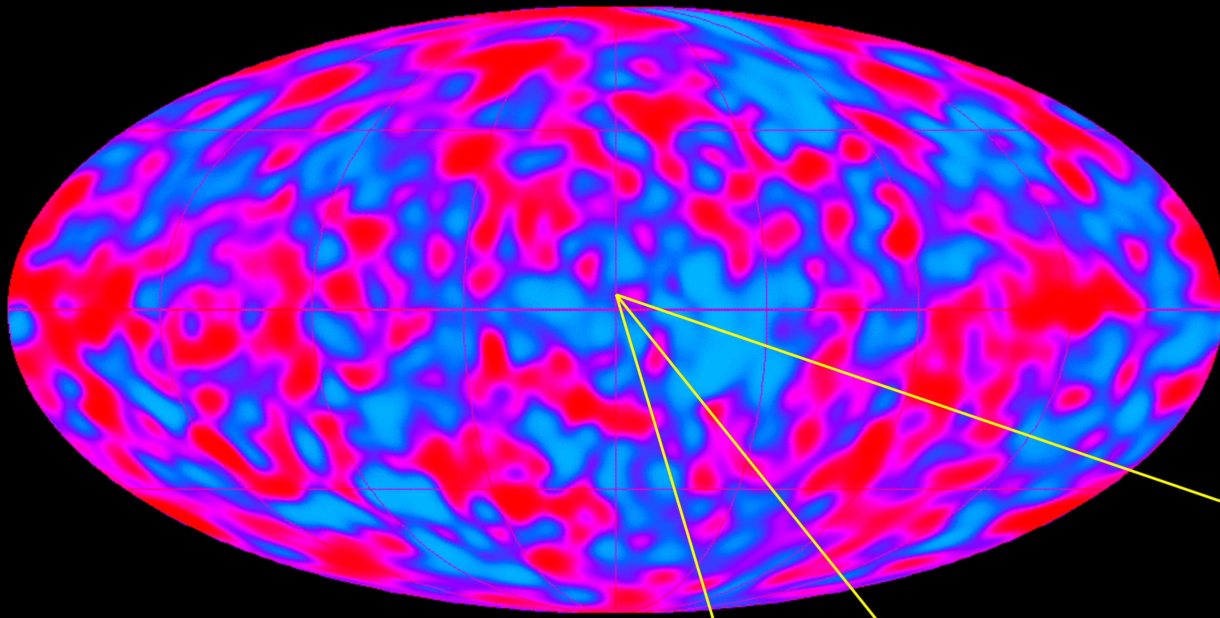
Observed galaxies are intrinsically more luminous

$$p_{\mathrm{obs}}(M) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\tfrac{1}{2} \left[\frac{M - \left(M_0 - 1.38\sigma^2\right)}{\sigma} \right]^2 \right)$$

Malmquist Bias

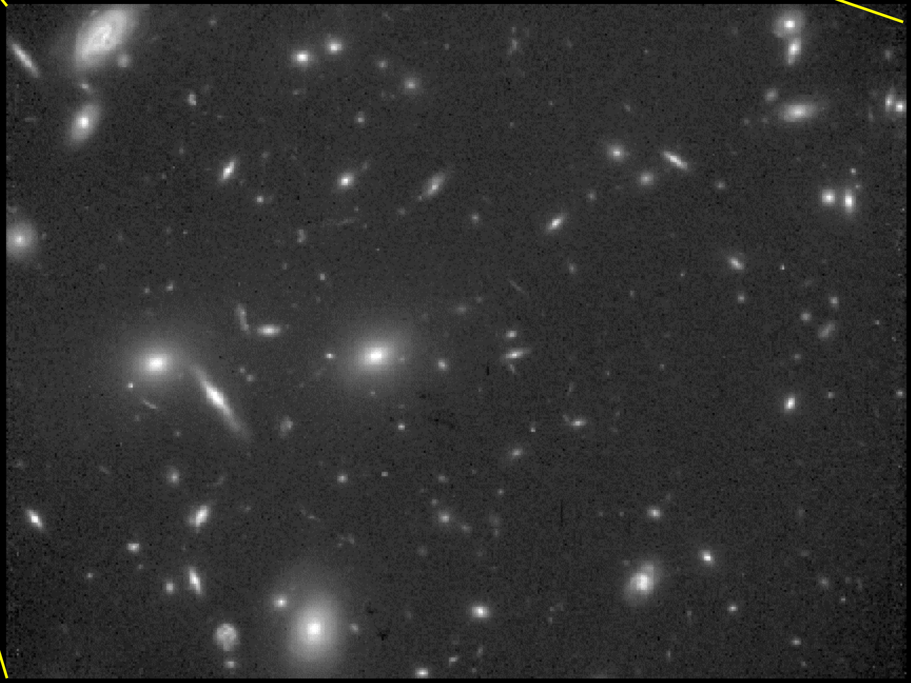


Can we avoid homogeneity assumption?



CMBR ‘ripples’ are the seeds of today’s galaxies

Galaxy formation is highly sensitive to the pattern, or power spectrum, of CMBR temperature ripples and the dark matter in the Universe



Velocity – Density Reconstructions

We can compare observed peculiar velocities with the reconstructed density and velocity field from all-sky redshift surveys, via linear theory relations:-

$$\mathbf{v}_{\mathrm{pec}}(\mathbf{r}) = \frac{\Omega_m^{0.6}}{4\pi} \int d^3\mathbf{r}' \frac{\delta(\mathbf{r}')(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3}$$

$$\nabla \cdot \mathbf{v}_{\mathrm{pec}} = -\Omega_m^{0.6}\, \delta$$

- **density-density comparisons**
- **velocity-velocity comparisons**

UNIVERSITY *of* GLASGOW

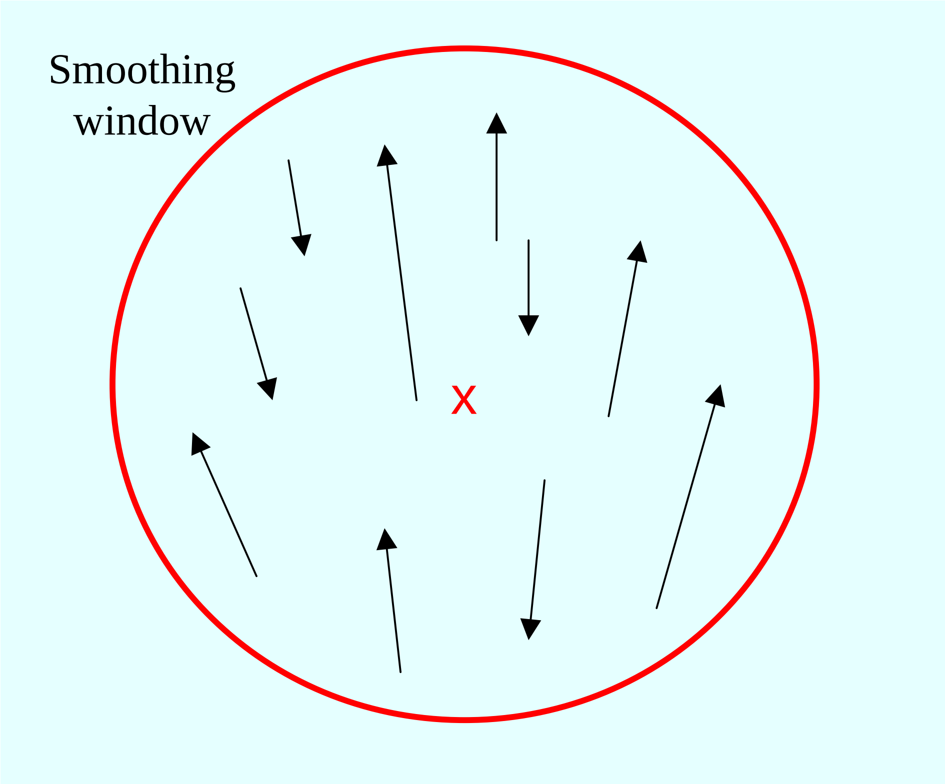
Density – density comparisons

Archetype is POTENT (Bertschinger & Dekel 1988; Dekel et al 1999)

$$\mathbf{v}_{\mathbf{pec}} = -\nabla\Phi_V$$

$$\Phi_V(\mathbf{r}) = -\int_0^{\mathbf{r}} u(r', \theta, \phi)\, dr'$$

Need *only* radial components, but everywhere! Interpolate $u(\mathbf{r})$ on a regular grid



Density – density comparisons

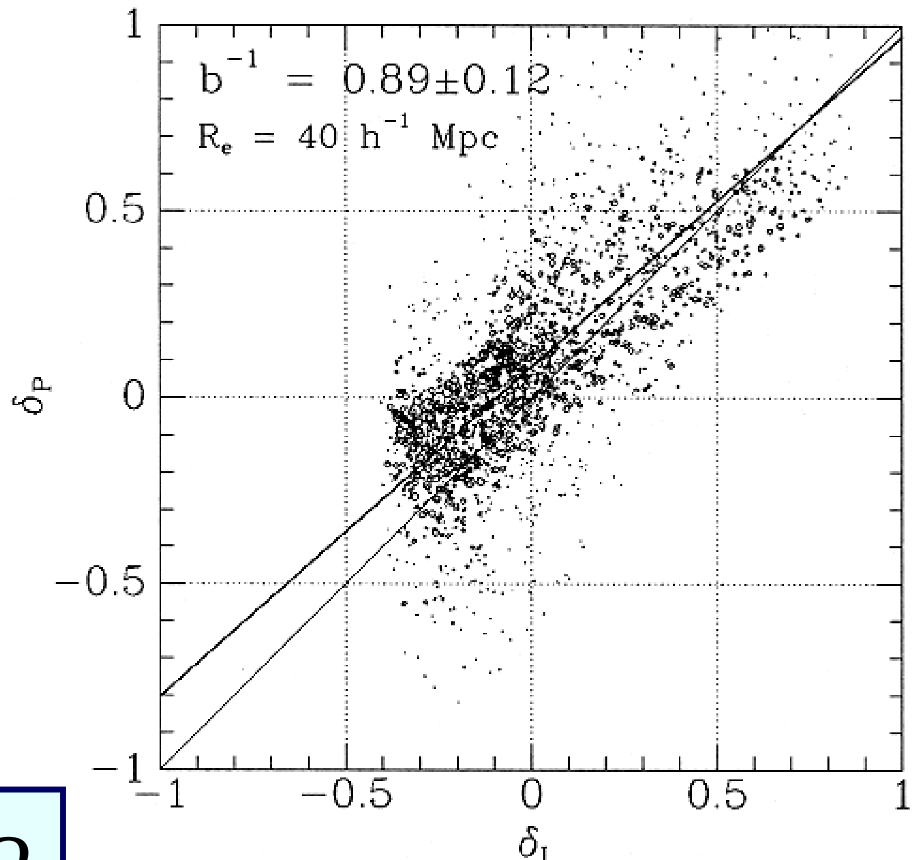
Archetype is POTENT (Bertschinger & Dekel 1988; Dekel et al 1999)

Compare $\mathbf{v}_{\mathbf{pec}}$ with e.g. IRAS δ-field. Assume *linear biasing:* $\delta_{\text{gal}} = b\,\delta$

$\nabla \cdot \mathbf{v}_{\mathbf{pec}}$ versus δ

has slope $\boxed{\beta = \Omega_m^{0.6} / b}$

$$\boxed{\beta_{\text{I}} = 0.89 \pm 0.12}$$



Sigad et al. (1998)

UNIVERSITY *of* GLASGOW

Density – density comparisons

POTENT is vulnerable to a number of statistical biases:-

- Calibration bias
- Inhomogeneous Malmquist bias
- Tensor window bias
- Sampling gradient bias

See e.g. Strauss & Willick (1995), Hendry & Simmons (1995), Hendry (2001)

Density – density comparisons

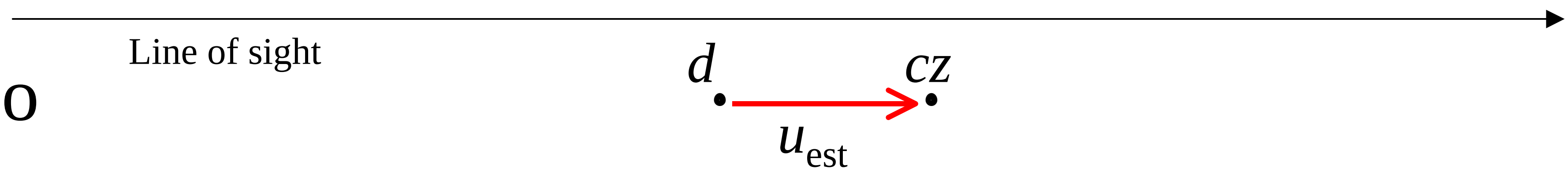
POTENT is vulnerable to a number of statistical biases:-

- Calibration bias
- Inhomogeneous Malmquist bias
- Tensor window bias
- Sampling gradient bias

See e.g. Strauss & Willick (1995), Hendry & Simmons (1995), Hendry (2001)

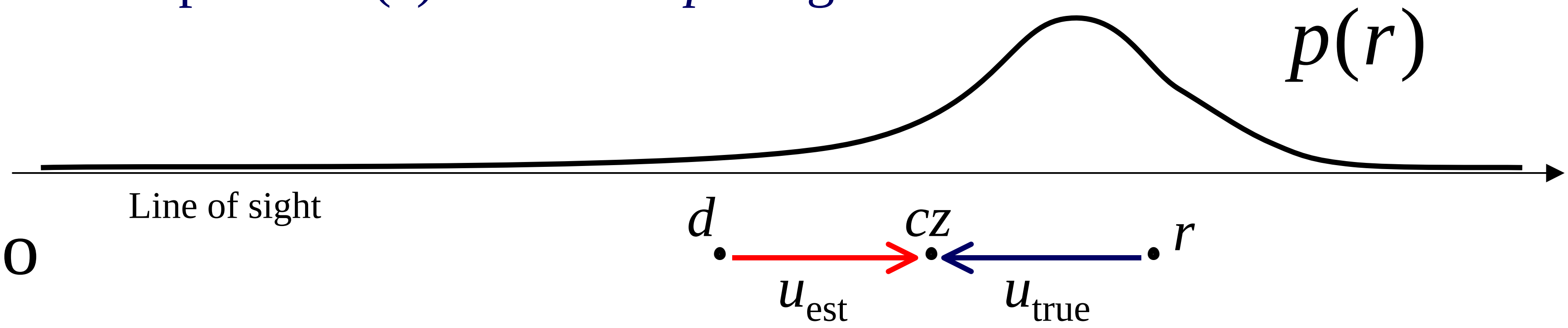
Inhomogeneous Malmquist bias

Interpolate $u(\mathbf{r})$ on a *real space* grid



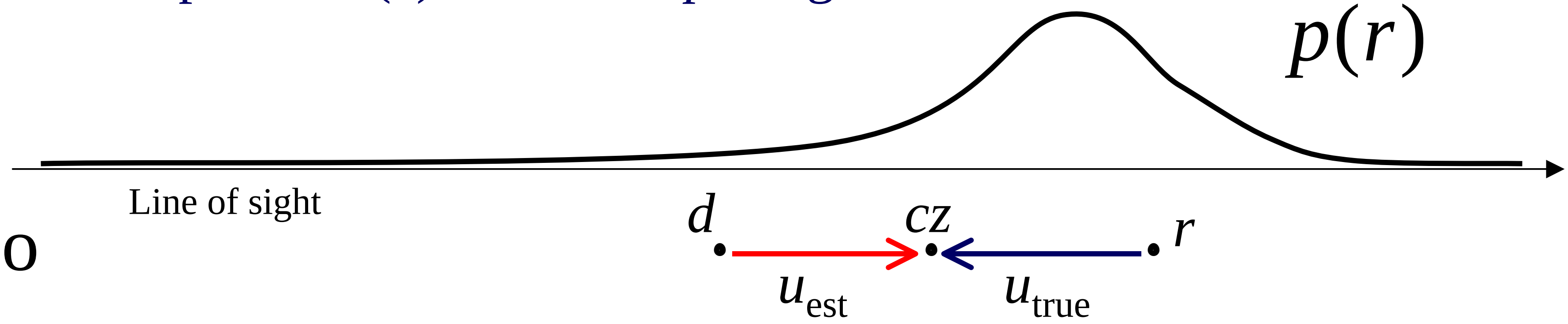
Inhomogeneous Malmquist bias

Interpolate $u(\mathbf{r})$ on a *real space* grid



Inhomogeneous Malmquist bias

Interpolate $u(\mathbf{r})$ on a *real space* grid



In general $\boxed{E(r \mid d) \neq d}$

Bias correction depends on $p(r)$

UNIVERSITY of GLASGOW

Velocity – velocity comparisons

Archetype is VELMOD (Willick & Strauss 1997, Willick et al 1998)

Maximise likelihood of observing Tully-Fisher data, given a velocity field and TF model

'Forward' VELMOD

$$L = \prod p(m_i \mid \eta_i, cz_i; \Theta)$$

'Inverse' VELMOD

$$L = \prod p(\eta_i \mid m_i, cz_i; \Theta)$$

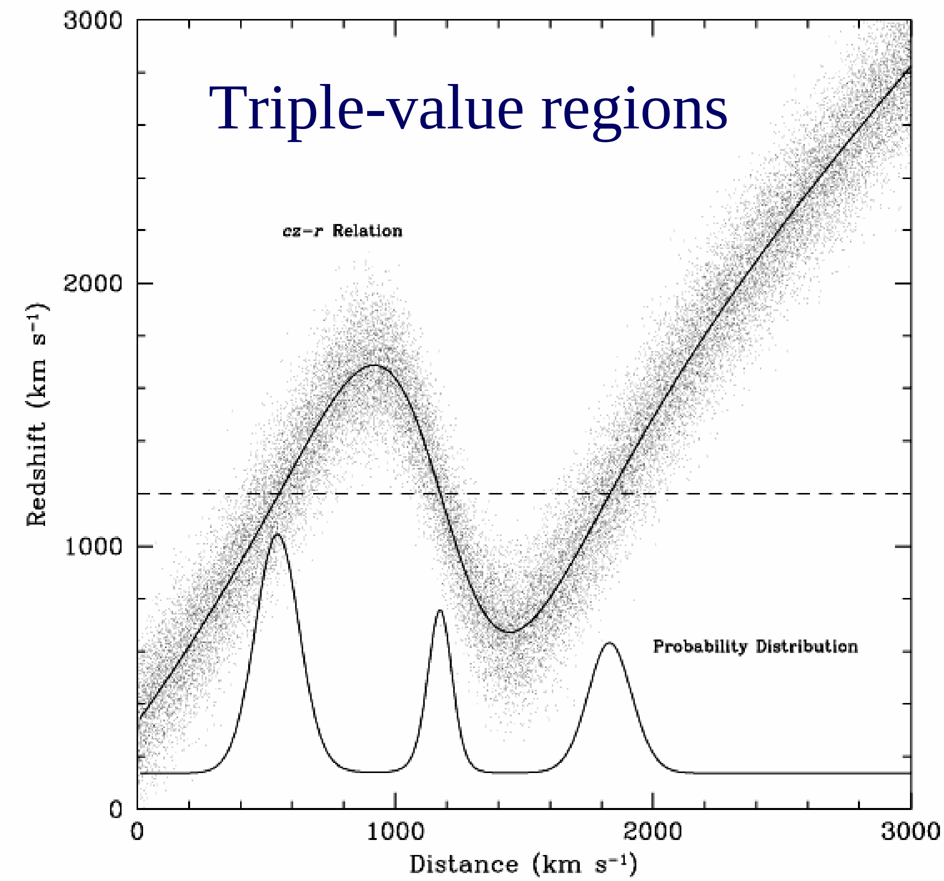
Θ = parameters of TF relation and velocity model

VELMOD also requires a parametric model for

$S(m, \eta, r)$, LF, $p(cz \mid r)$

UNIVERSITY of GLASGOW

Velocity – velocity comparisons



Strauss & Willick (1995)

VELMOD also requires a parametric model for

$$S(m,\eta,r), \quad \text{LF}, \quad p(cz\,|\,r)$$

UNIVERSITY *of* GLASGOW

Robust Method

Assumption: luminosity function is **Universal**

$$\text{Prob} \quad \propto \quad \psi(m,r,l,b)\,\rho(r,l,b)\,p(M)$$

Selection effects

Spatial distribution

Luminosity function

We want to test our model for the selection effects

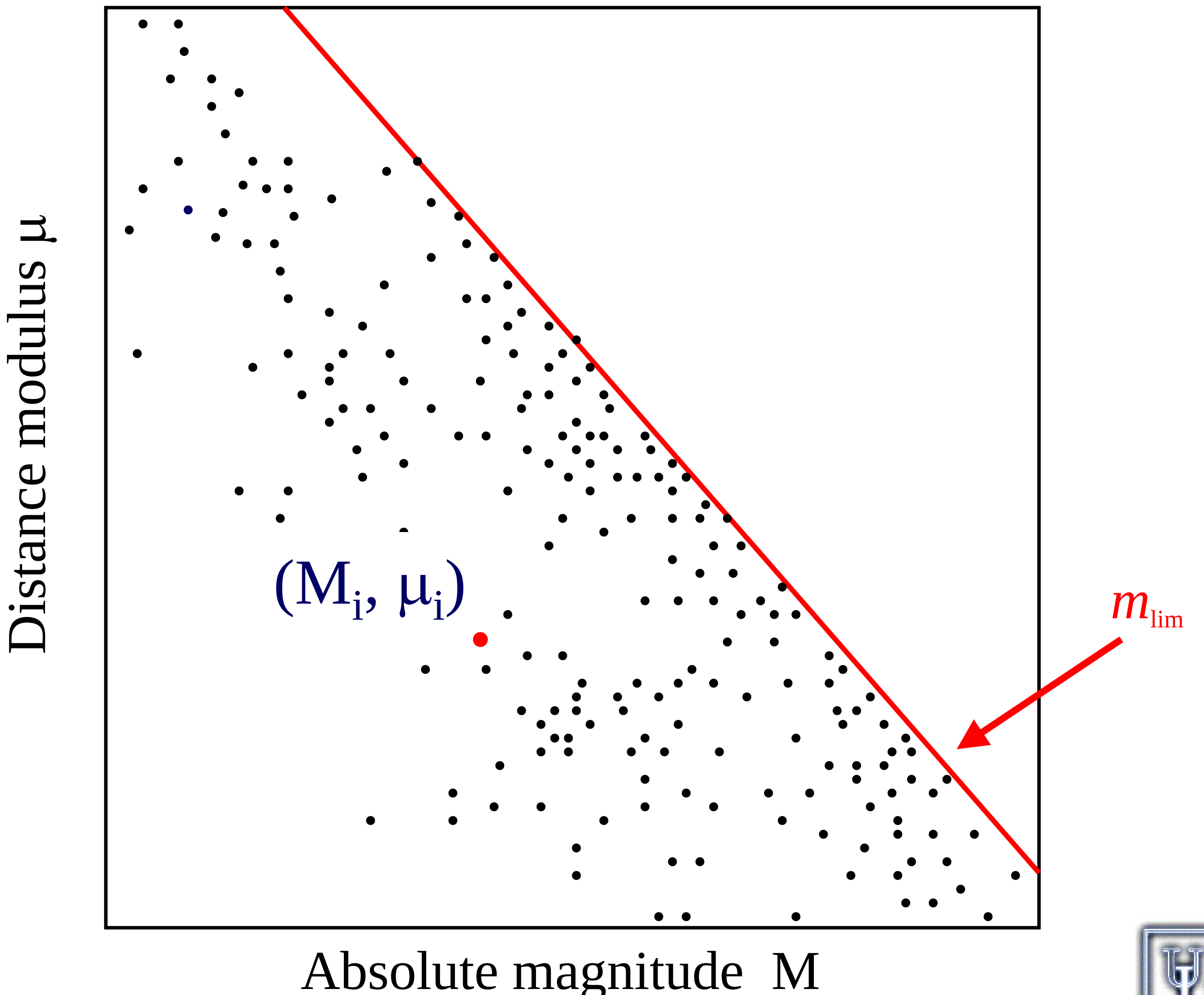
$$\psi(m,z,l,b) \equiv \theta(m_{\mathrm{lim}} - m) \times \phi(z,l,b)$$

Step function

Angular and radial Selection function

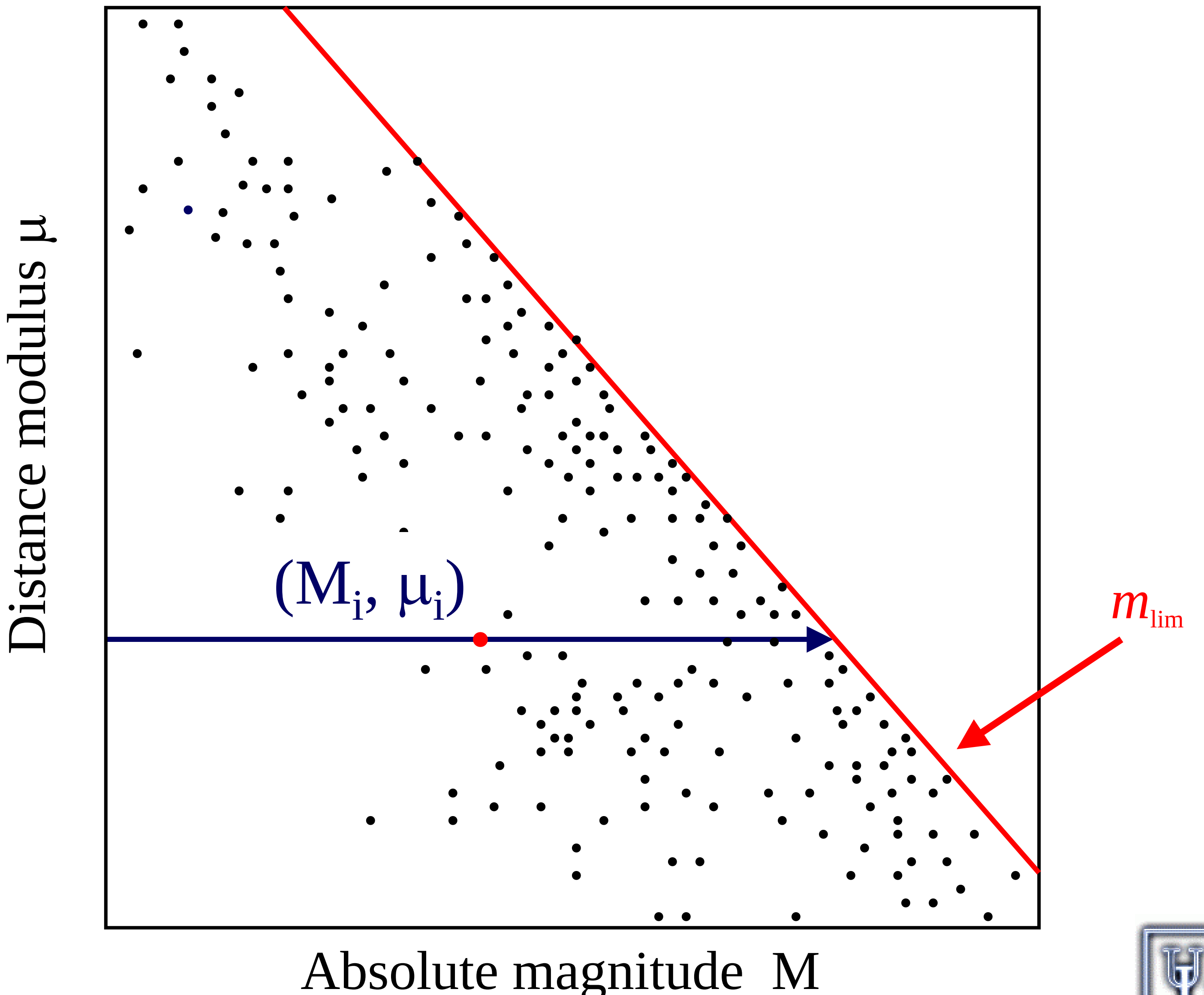
UNIVERSITY of GLASGOW

Robust Method: Completeness

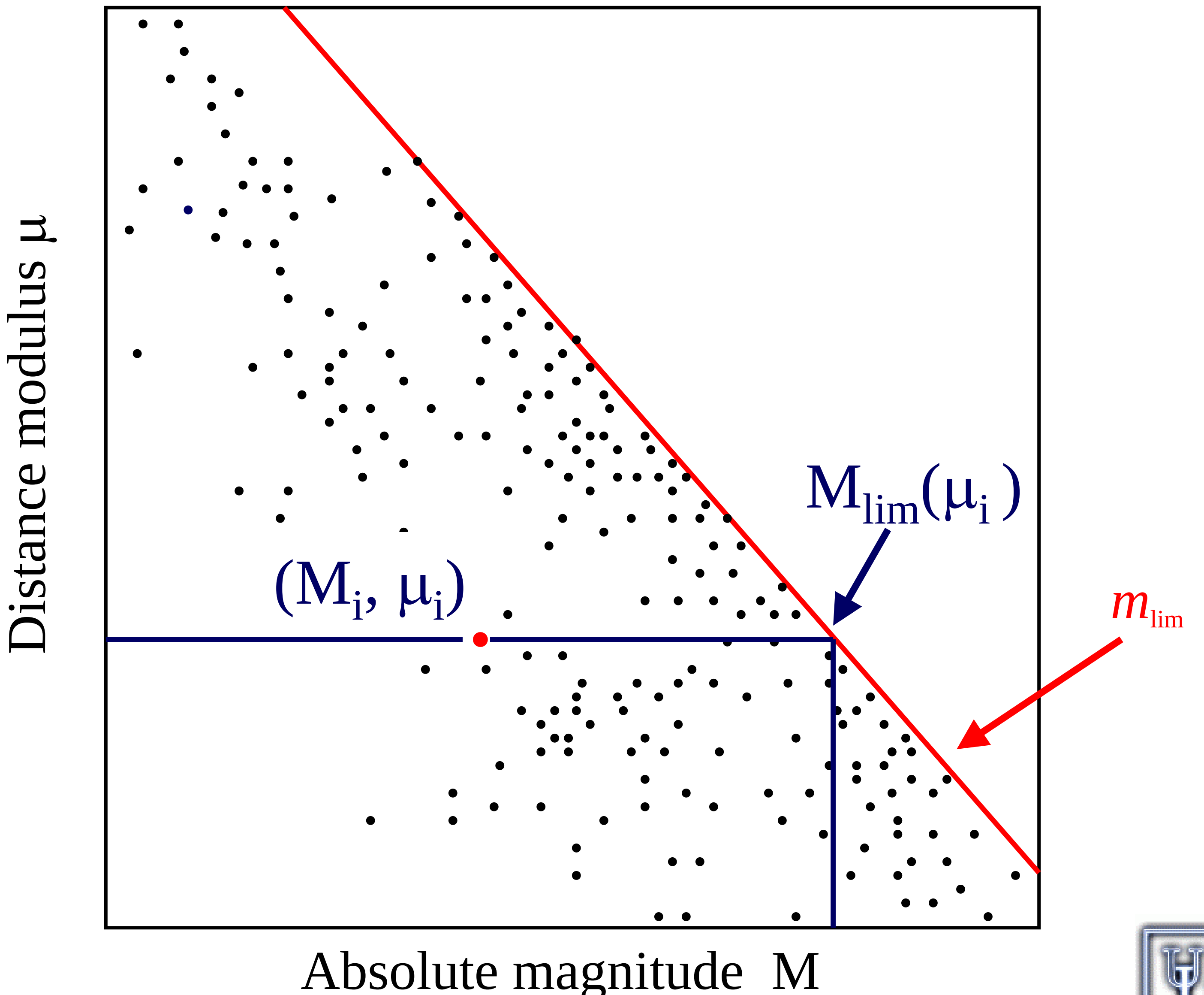


UNIVERSITY of GLASGOW

Robust Method: Completeness

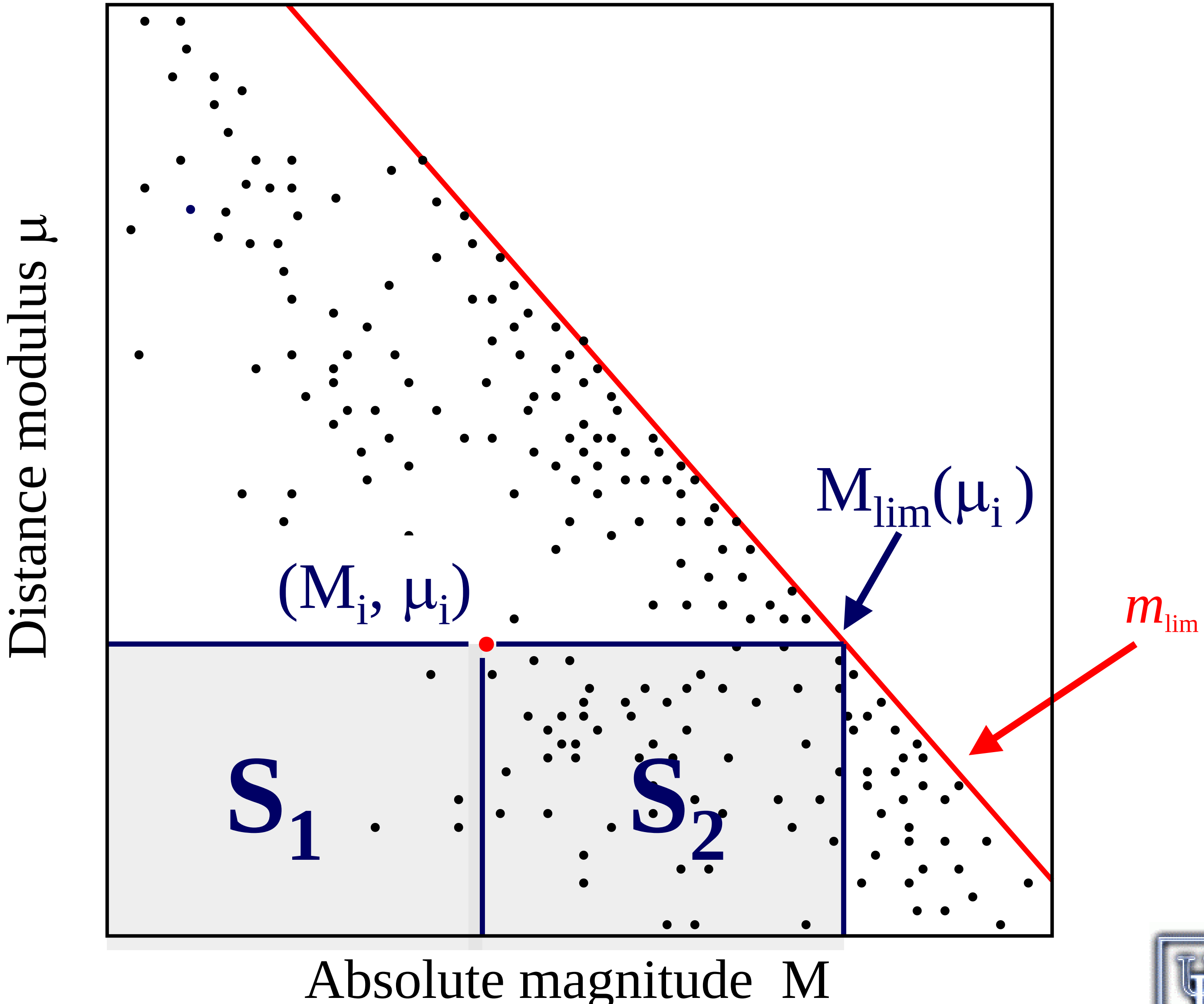


Robust Method: Completeness



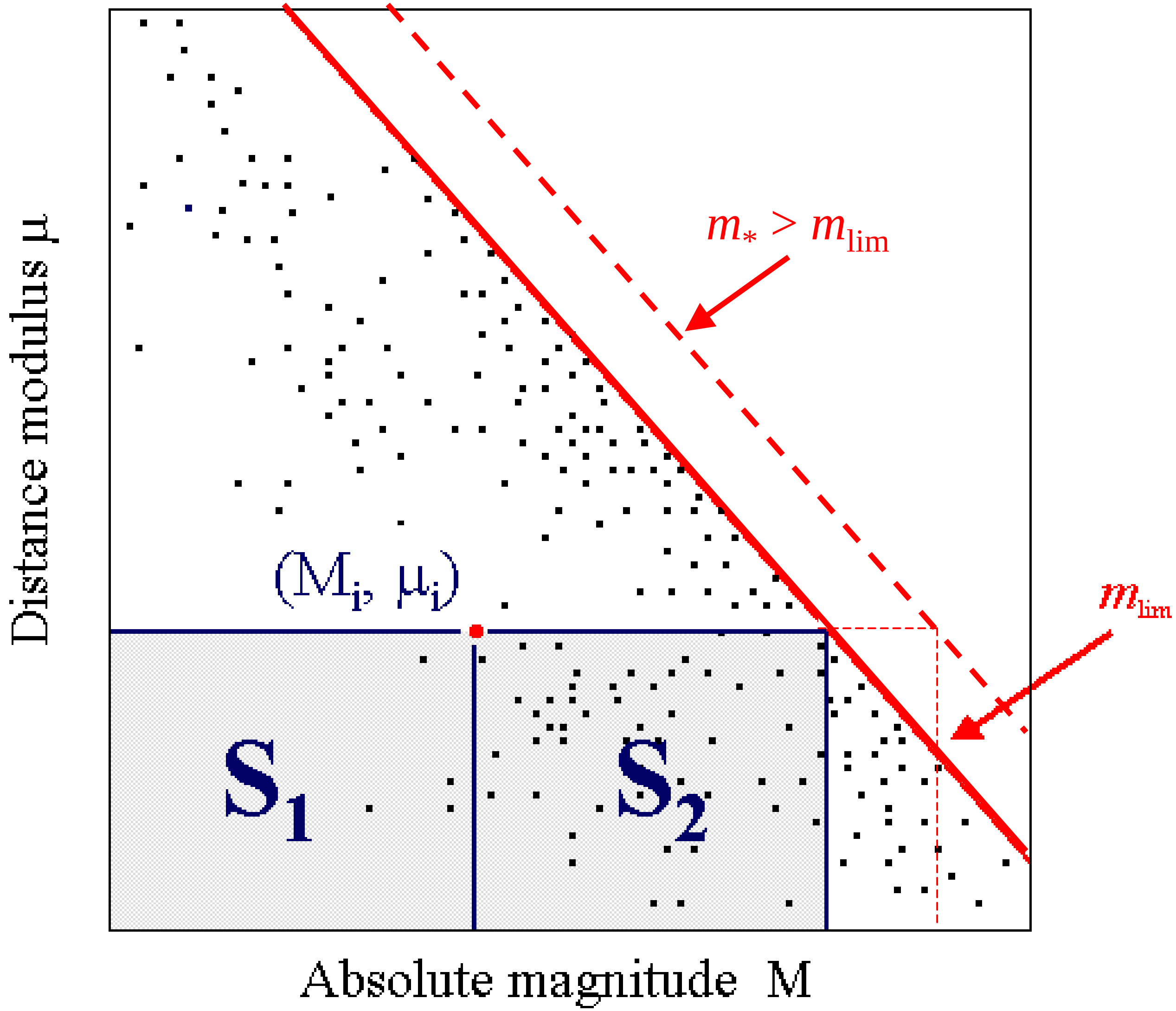
UNIVERSITY of GLASGOW

Robust Method: Completeness



UNIVERSITY of GLASGOW

Robust Method: Completeness



UNIVERSITY of GLASGOW

Robust Method: Completeness

Define:- $$\zeta = \frac{F(M)}{F(M_{\lim})}$$

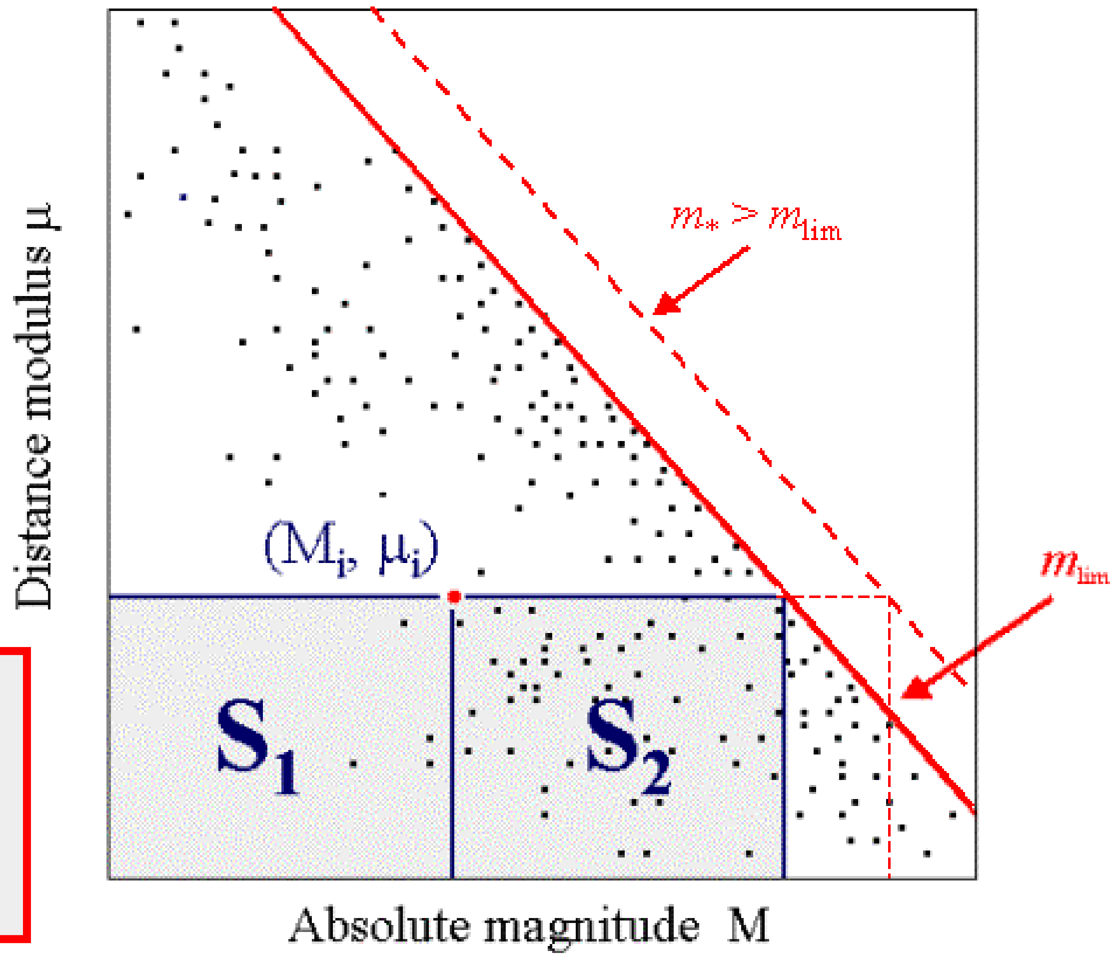
where

$$F(M) = \int_{-\infty}^{M} f(x)dx$$

Can show:-

P1: $\zeta \in \mathrm{U}[0,1]$

P2: ζ, μ uncorrelated



Robust Method: Completeness

Also:-

$$\boxed{\hat{\zeta}_i = \frac{r_i}{n_i + 1}}$$

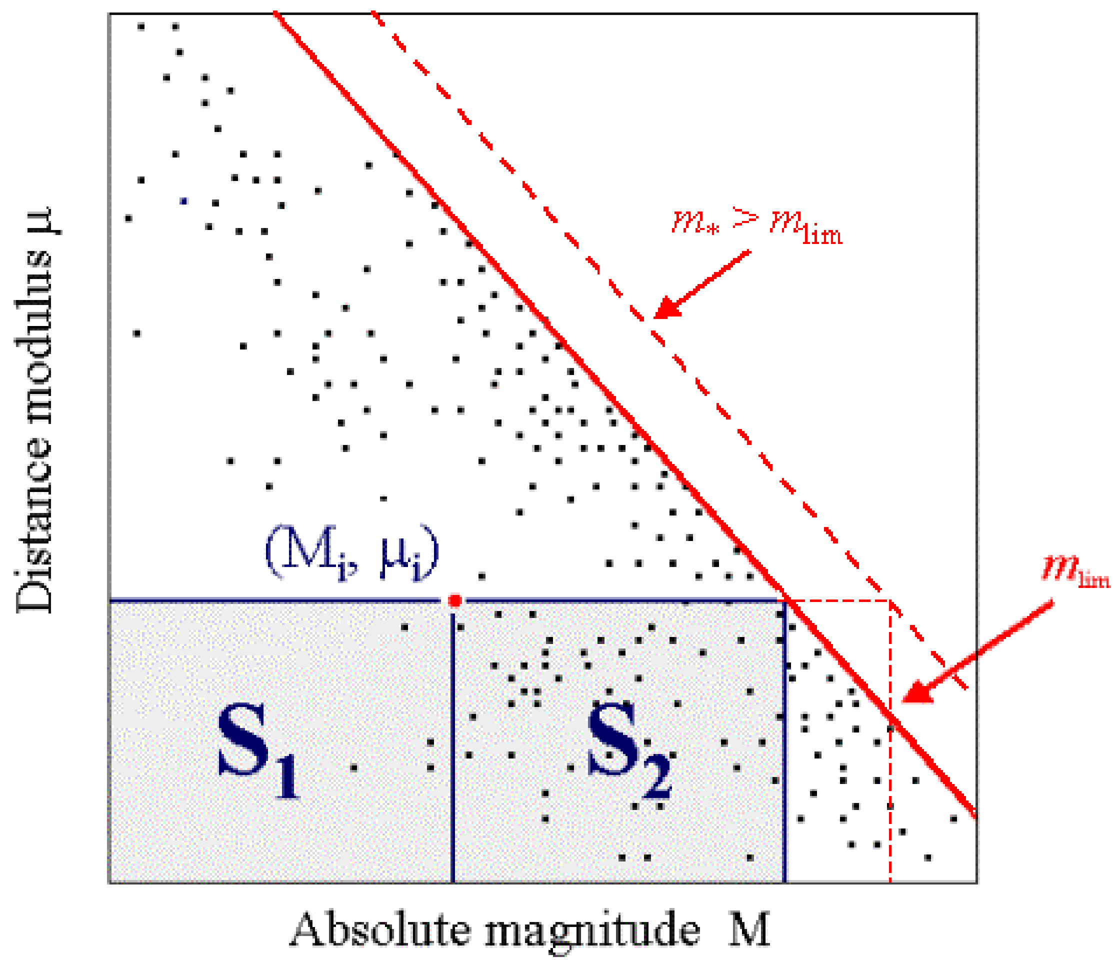
$$r_i = n(S_1)$$

$$n_i = n(S_1 \cup S_2)$$

$$E_i = \frac{1}{2} \qquad V_i = \frac{1}{12}\frac{n_i - 1}{n_i + 1}$$

but *only* for

$$\boxed{m_* \le m_{\lim}}$$



Robust Method: Completeness

Also:-

$$\boxed{\hat{\zeta}_i = \frac{r_i}{n_i + 1}}$$

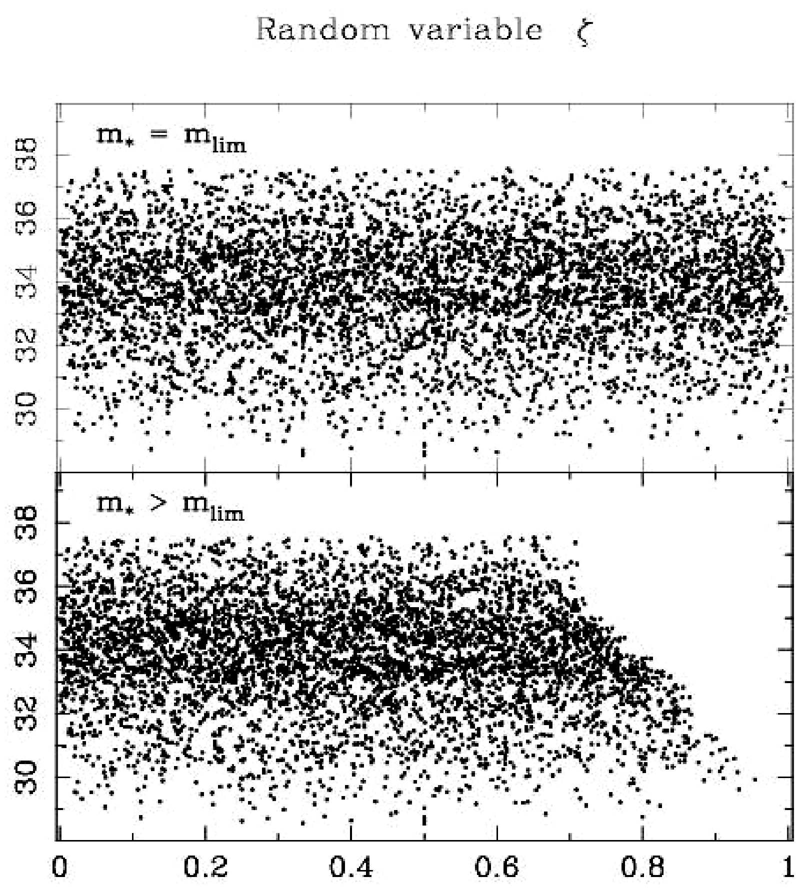
$$r_i = n(S_1)$$

$$n_i = n(S_1 \cup S_2)$$

$$E_i = \frac{1}{2} \qquad V_i = \frac{1}{12}\frac{n_i - 1}{n_i + 1}$$

but *only* for

$$\boxed{m_* \leq m_{\text{lim}}}$$



Robust Method: Completeness

Also:-

$$\boxed{\hat{\zeta}_i = \frac{r_i}{n_i + 1}}$$

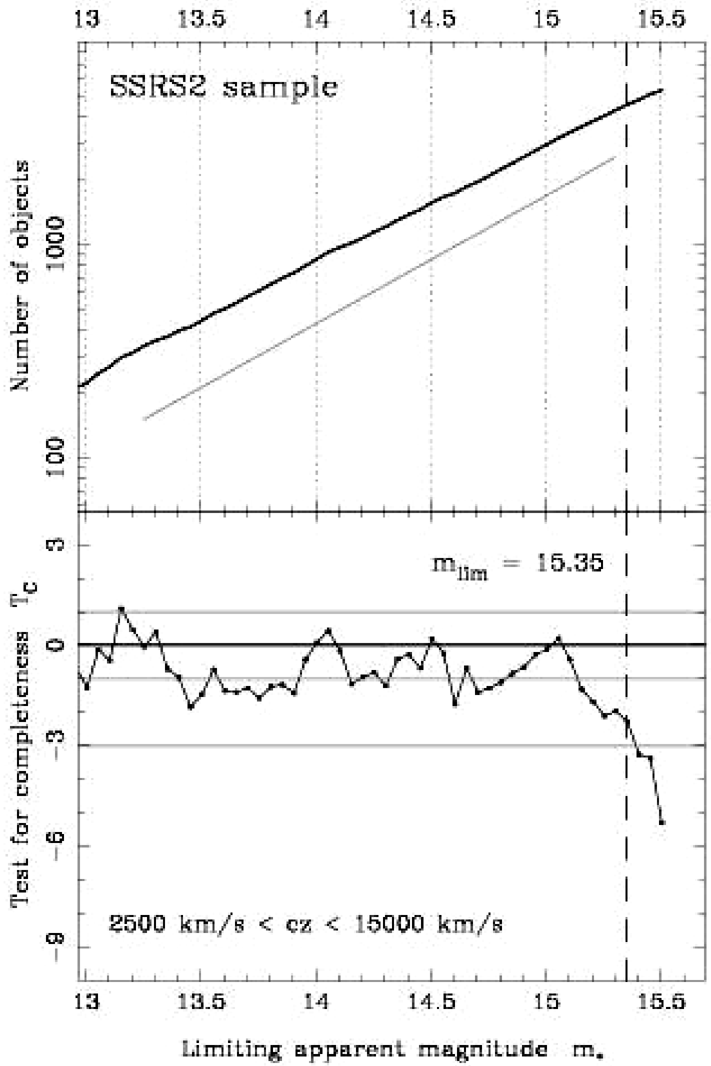
$$r_i = n(S_1)$$

$$n_i = n(S_1 \cup S_2)$$

$$E_i = \frac{1}{2} \qquad V_i = \frac{1}{12}\frac{n_i - 1}{n_i + 1}$$

but *only* for

$$\boxed{m_* \le m_{\lim}}$$



UNIVERSITY of GLASGOW

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$p_{\mathrm{obs}}(\mathrm{data}\,|\,\mathrm{model}, I) \;=\; p(\mathrm{data}\,|\,\mathrm{model}, I) \times S(\mathrm{data}, I)$$

The 'actual' likelihood

The 'ideal' likelihood

The selection function

More generally, the selection function can be much more complicated

- 'Zone of avoidance'
- Surface brightness
- Galaxy diameters
- Colour
- Redshift

UNIVERSITY of GLASGOW

IAU

Dealing with observational selection effects

Easy *in principle* to correct for selection effects

$$p_{\text{obs}}(\text{data}\,|\,\text{model}, I) \;=\; p(\text{data}\,|\,\text{model}, I)\times S(\text{data}, I)$$

The 'actual' likelihood

The 'ideal' likelihood

The selection function

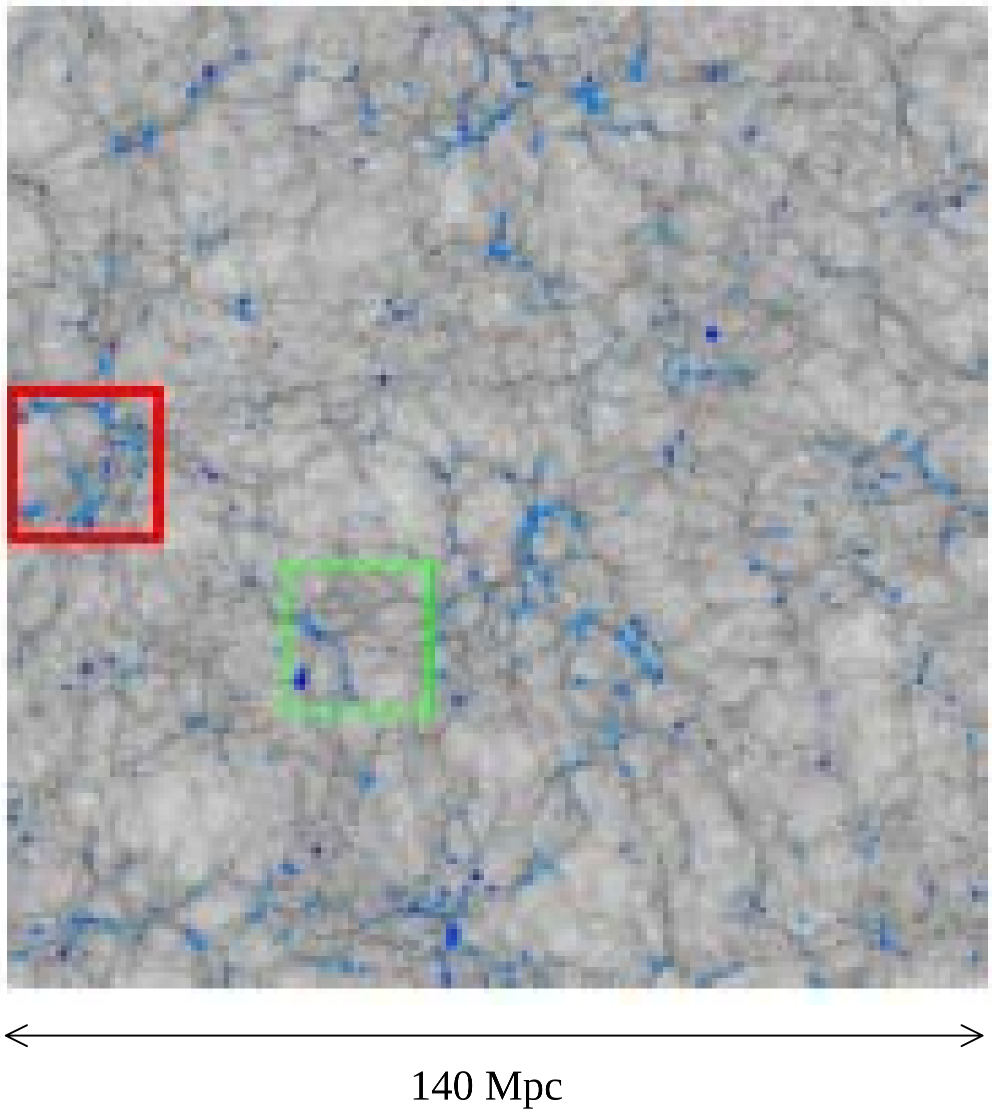
More generally, the selection function can be much more complicated

- 'Zone of avoidance'
- Surface brightness
- Galaxy diameters
- Colour
- Redshift

Too difficult to model analytically, but we can use **Monte Carlo simulation** to generate 'Mock' datasets

UNIVERSITY of GLASGOW

IYAU

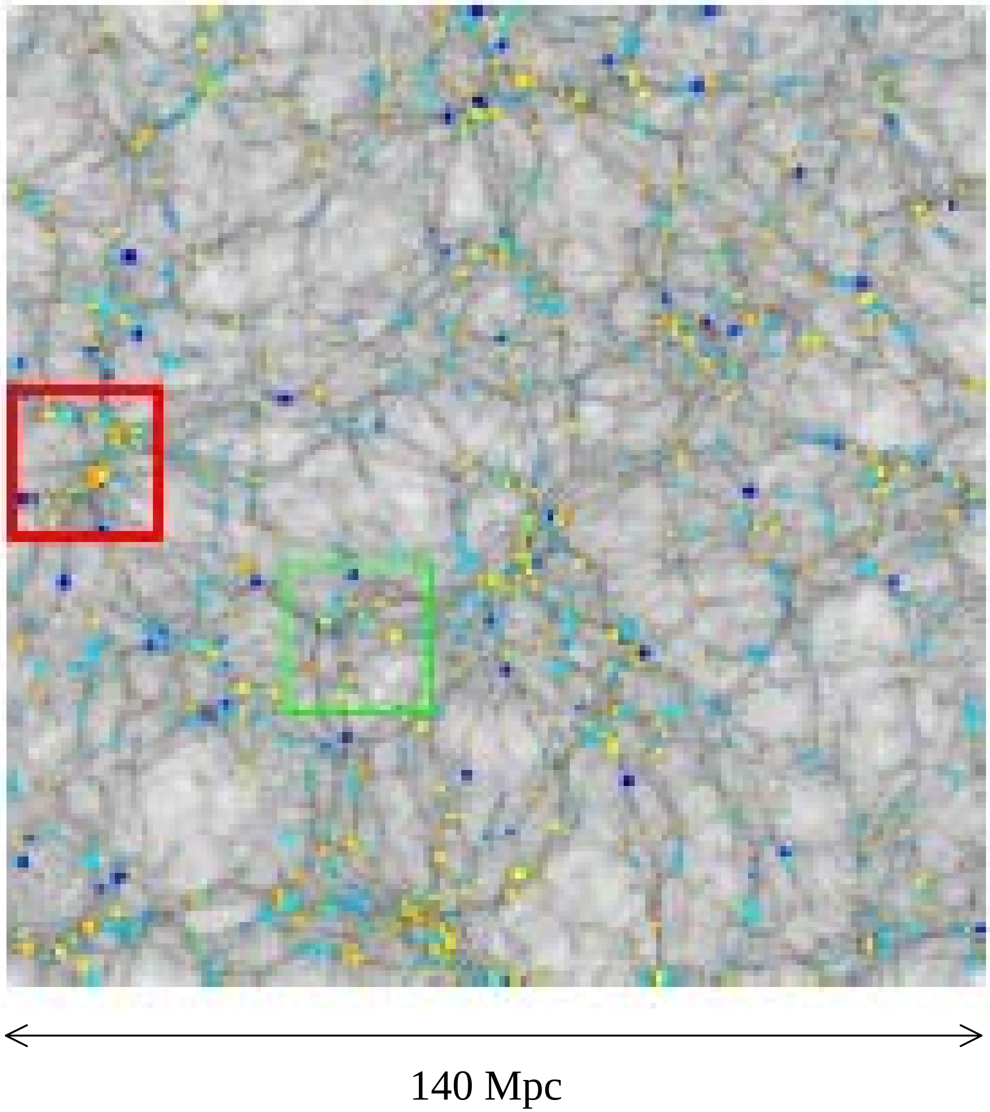


11 Gyr ago

Hierarchical clustering:

Galaxies form out of the mergers of fragments: CDM halos at high redshift.

Clusters form where filaments and sheets of matter intersect



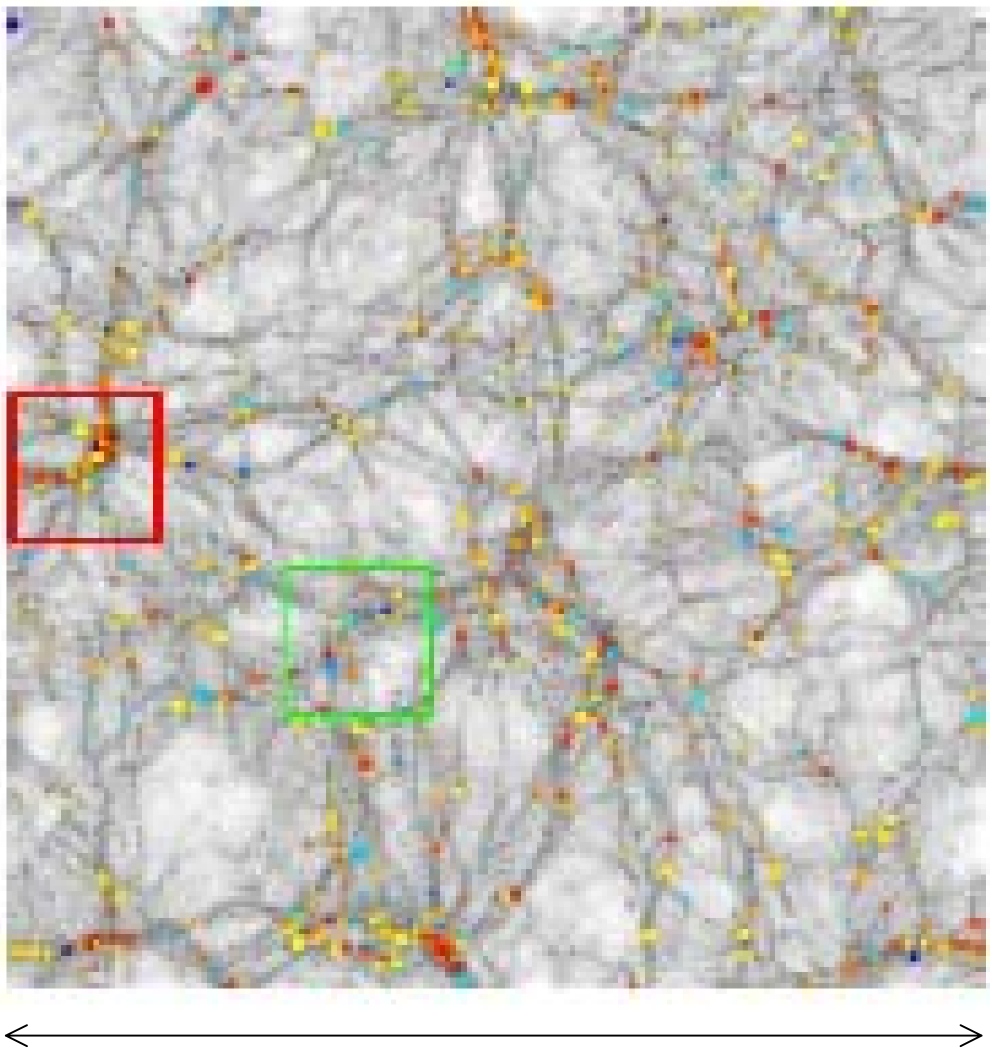
Hierarchical clustering:

Galaxies form out of the mergers of fragments: CDM halos at high redshift.

Clusters form where filaments and sheets of matter intersect

8 Gyr ago

UNIVERSITY of GLASGOW



140 Mpc

Present day

Hierarchical clustering:

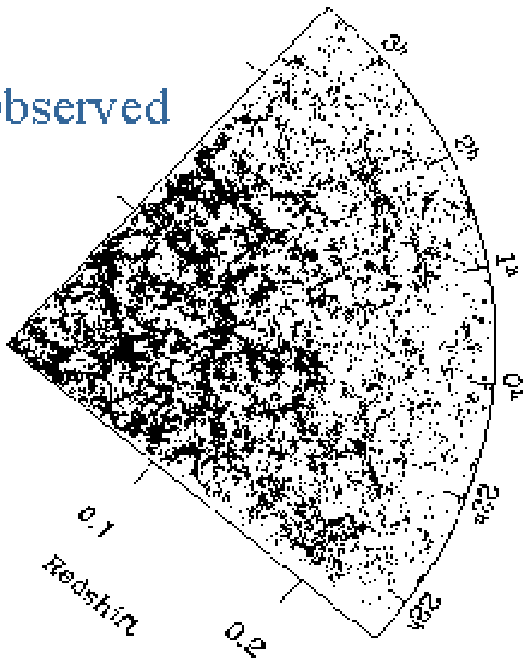
Galaxies form out of the mergers of fragments: CDM halos at high redshift.

Clusters form where filaments and sheets of matter intersect

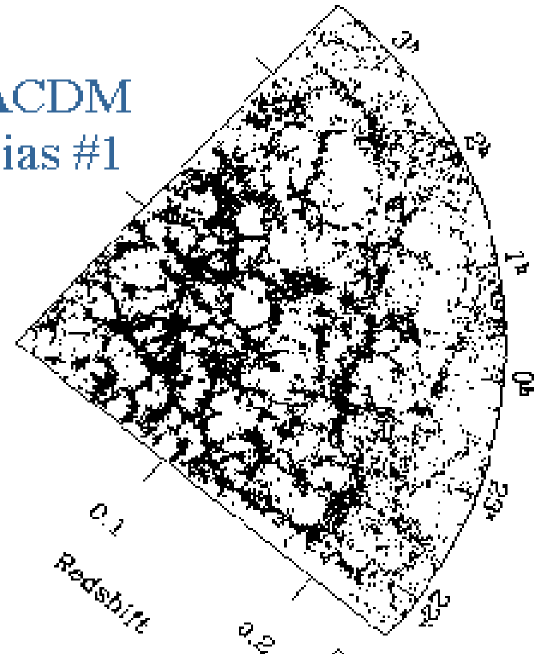
UNIVERSITY of GLASGOW

Models vs observations

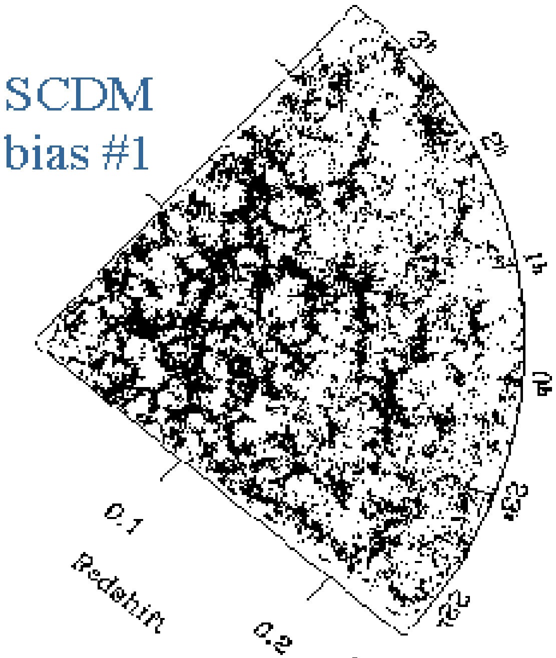
Observed



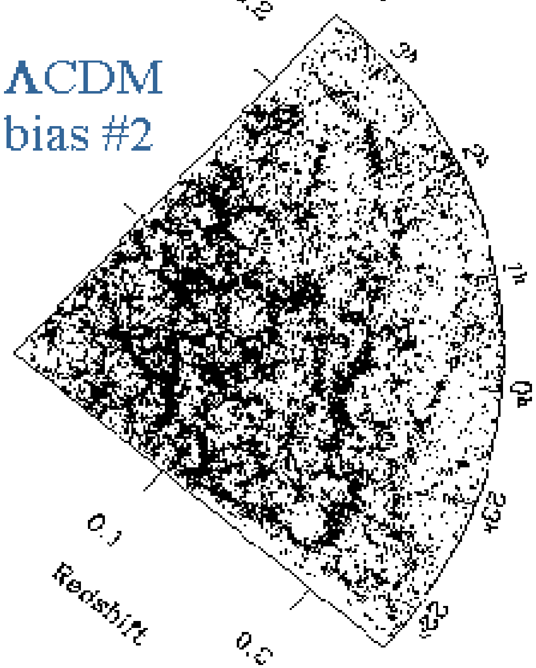
ΛCDM
bias #1



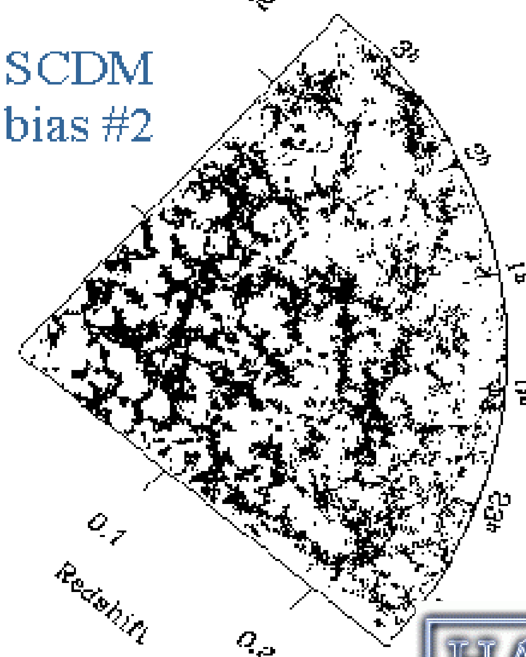
SCDM
bias #1



ΛCDM
bias #2



SCDM
bias #2



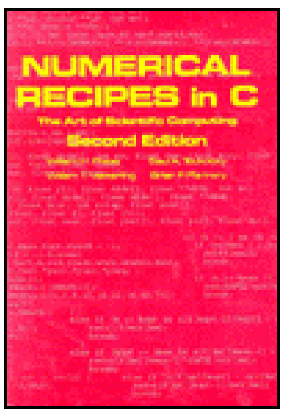
UNIVERSITY
of
GLASGOW

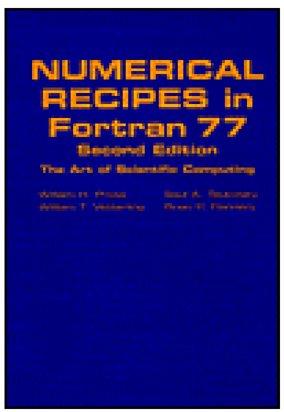
Monte Carlo Sampling

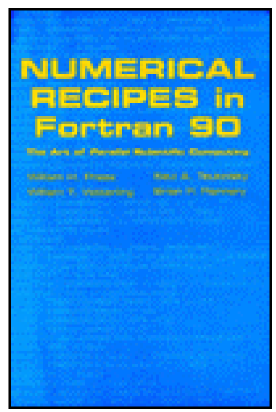
Generating random variables

1. Uniform random number, U[0,1]

See Numerical Recipes!







http://www.numerical-recipes.com/

Portable Random Number Generators

Park and Miller [1] have surveyed a large number of random number generators that have been used over the last 30 years or more. Along with a good theoretical review, they present an anecdotal sampling of a number of inadequate generators that have come into widespread use. The historical record is nothing if not appalling.

There is good evidence, both theoretical and empirical, that the simple multiplicative congruential algorithm

$$I_{j+1} = aI_j \pmod{m} \tag{7.1.2}$$

can be as good as any of the more general linear congruential generators that have $c \neq 0$ (equation 7.1.1) — *if* the multiplier a and modulus m are chosen exquisitely carefully. Park and Miller propose a "Minimal Standard" generator based on the choices

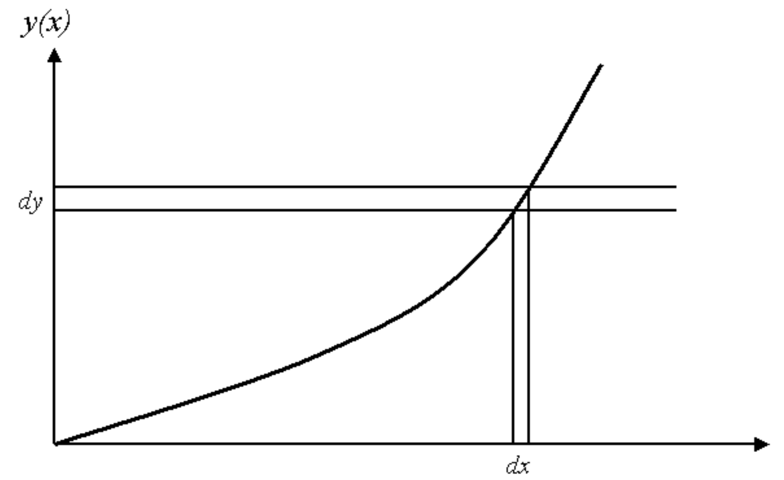
$$a = 7^5 = 16807 \qquad m = 2^{31} - 1 = 2147483647 \tag{7.1.3}$$

First proposed by Lewis, Goodman, and Miller in 1969, this generator has in subsequent years passed all new theoretical tests, and (perhaps more importantly) has accumulated a large amount of successful use. Park and Miller do not claim that the generator is "perfect" (we will see below that it is not), but only that it is a good minimal standard against which other generators should be judged.

It is not possible to implement equations (7.1.2) and (7.1.3) directly in a high-level language, since the product of a and $m - 1$ exceeds the maximum value for a 32-bit integer. Assembly language implementation using a 64-bit product register is straightforward, but not portable from machine to machine. A trick due to Schrage [2,3] for multiplying two 32-bit integers modulo a 32-bit constant, without using any intermediates larger than 32 bits (including a sign bit) is therefore extremely interesting: It allows the Minimal Standard generator to be implemented in essentially any programming language on essentially any machine.

Schrage's algorithm is based on an *approximate factorization* of m,

Monte Carlo Sampling



Generating random variables

2. Transformed Random Variables

Suppose we have $x \sim U[0,1]$

Let $y = y(x)$

Then $\boxed{p(y)dy = p(x)dx}$

Probability of number between y and $y+dy$

Probability of number between x and $x+dx$

$$\boxed{p(y) = \frac{p(x(y))}{|dy/dx|}}$$

Because probability must be positive

Monte Carlo Sampling

Generating random variables

2. Transformed Random Variables

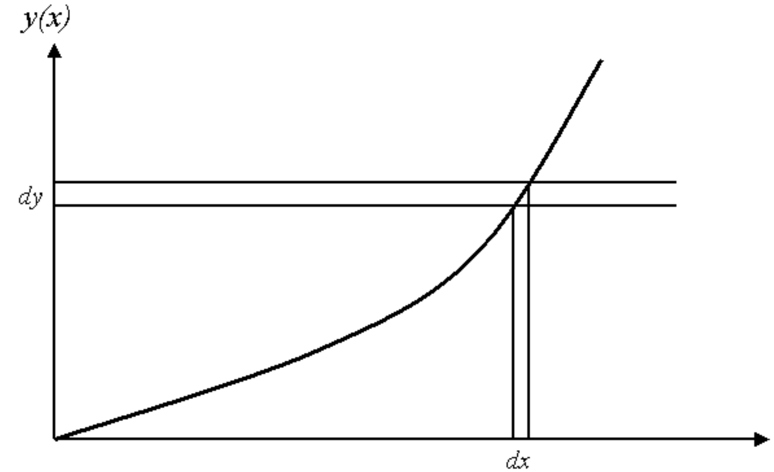
Suppose we have $x \sim U[0,1]$

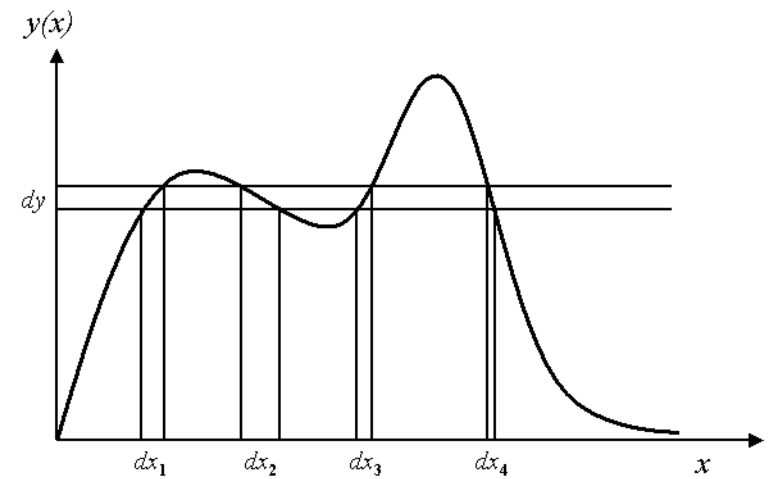
Let $y = y(x)$

Then $\boxed{p(y)dy = p(x)dx}$

Probability of number between y and $y+dy$

Probability of number between x and $x+dx$





UNIVERSITY of GLASGOW

Monte Carlo Sampling

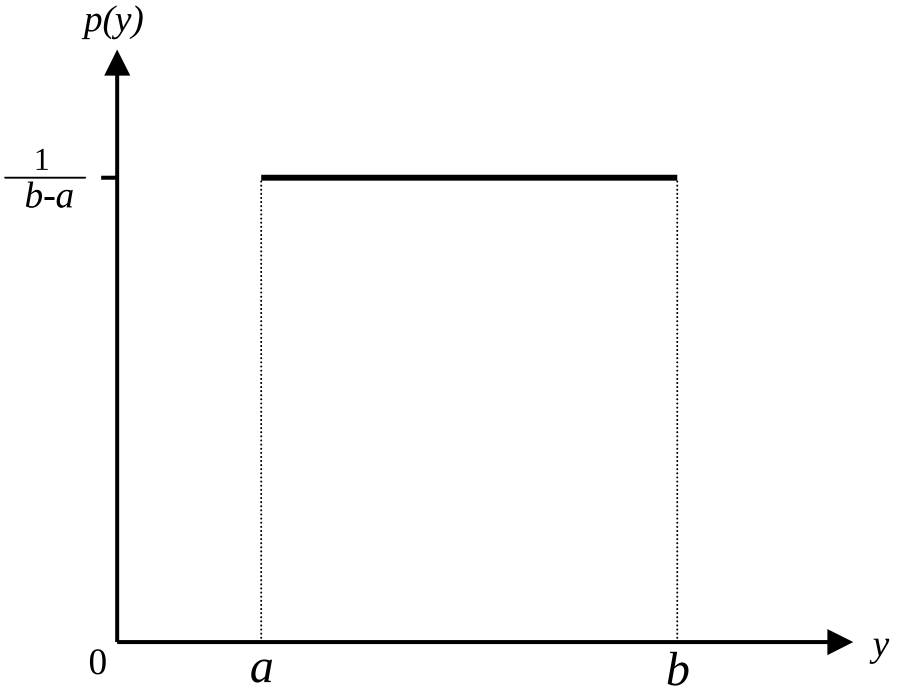
<u>Generating random variables</u>

2. Transformed Random Variables

Suppose we have $x \sim U[0,1]$

Let $y = a + (b-a)x$

Then $y \sim U[a,b]$



Monte Carlo Sampling

Generating random variables

2. Transformed Random Variables

Numerical Recipes uses the transformation method to provide $x \sim N(0,1)$:

Normal distribution with mean zero and standard deviation unity

Define $z = \mu + \sigma x$

$$x \sim N(\mu,\sigma)$$

UNIVERSITY of GLASGOW

Monte Carlo Sampling

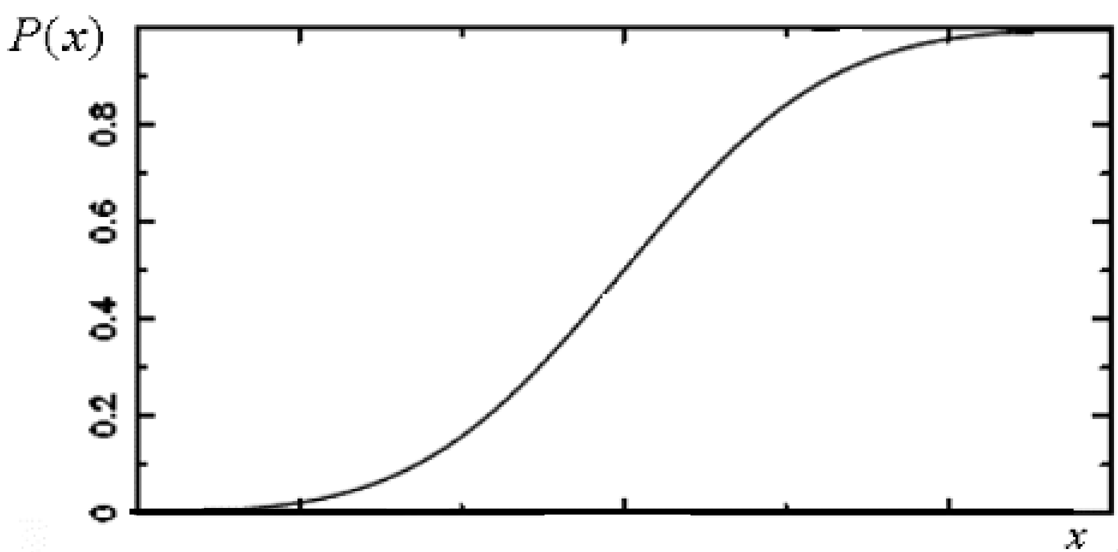
Generating random variables

3. Probability Integral Transform

Cumulative distribution function (CDF)

Suppose we can compute the CDF of some desired random variable

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



Monte Carlo Sampling

Generating random variables

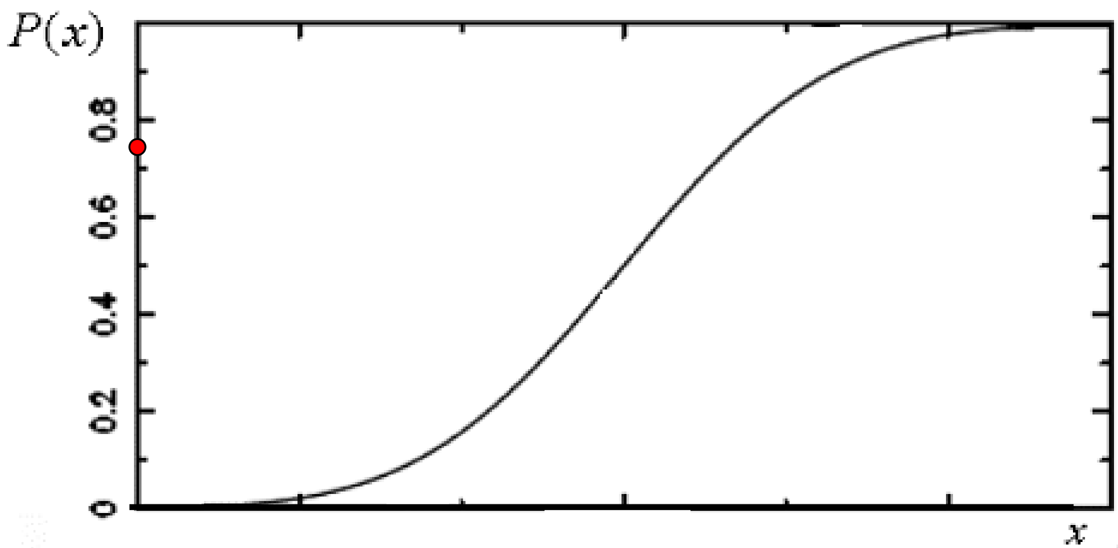
3. Probability Integral Transform

Cumulative distribution function (CDF)

Suppose we can compute the CDF of some desired random variable

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

1) $y \sim U[0,1]$



UNIVERSITY of GLASGOW

Monte Carlo Sampling

Generating random variables

3. Probability Integral Transform

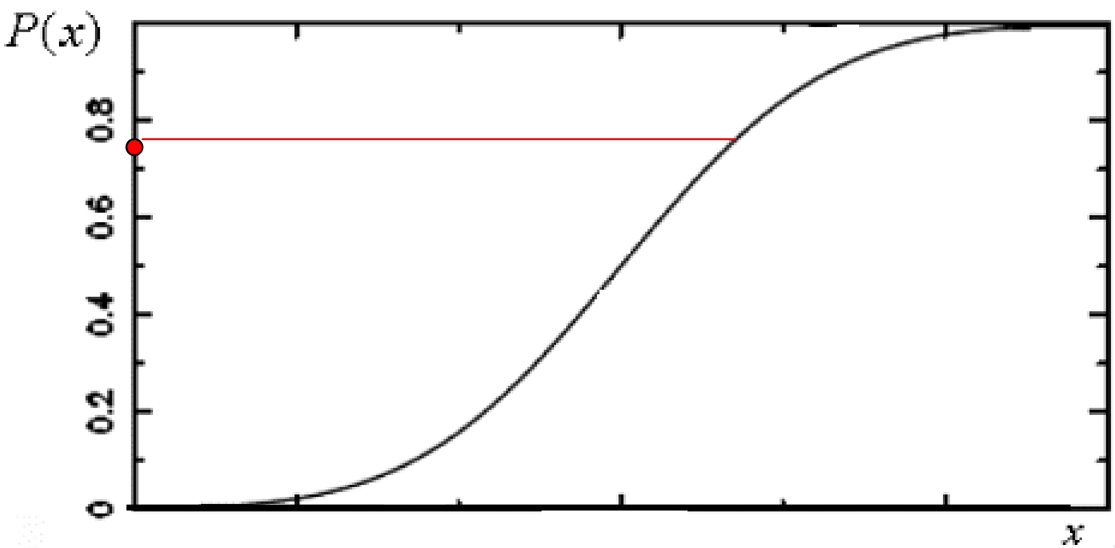
Cumulative distribution function (CDF)

Suppose we can compute the CDF of some desired random variable

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

1) $y \sim U[0,1]$

2) Compute
$x = P^{-1}(y)$



UNIVERSITY of GLASGOW

Monte Carlo Sampling

Generating random variables

3. Probability Integral Transform

Cumulative distribution function (CDF)

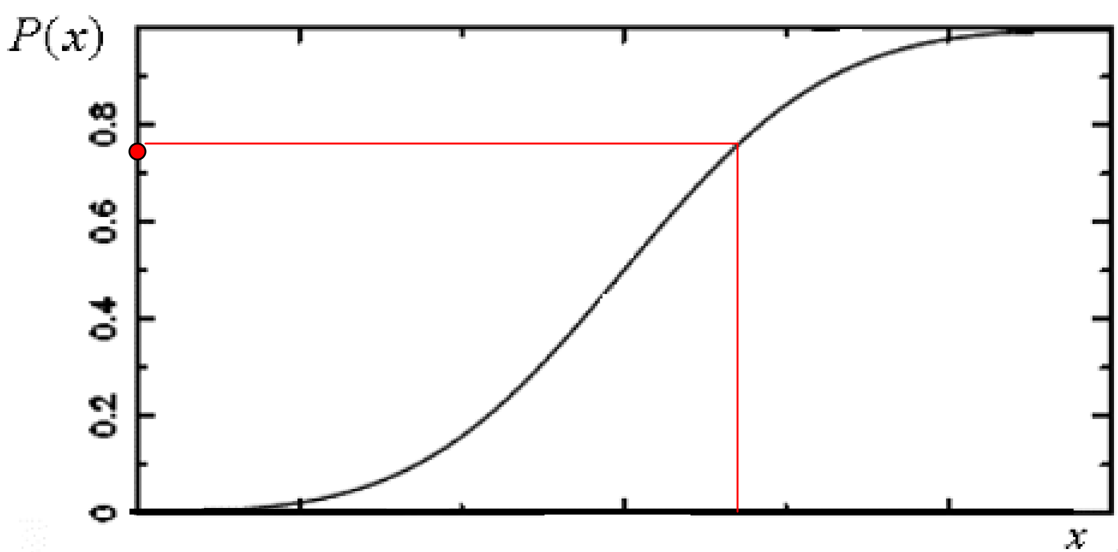
Suppose we can compute the CDF of some desired random variable

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

1) $y \sim U[0,1]$

2) Compute $x = P^{-1}(y)$

3) Then $x \sim p(x)$



UNIVERSITY of GLASGOW

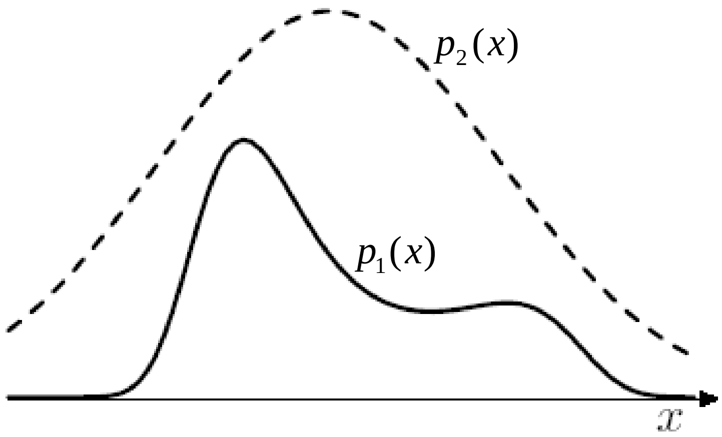
Monte Carlo Sampling

<u>Generating random variables</u>

4. Rejection Sampling

   Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$



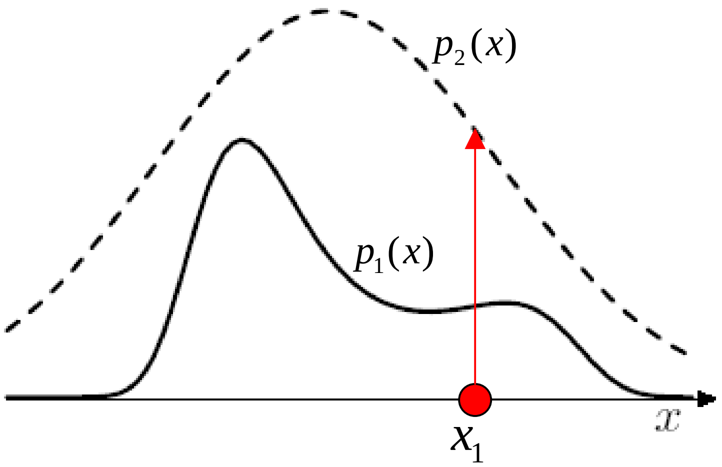
Monte Carlo Sampling

<u>Generating random variables</u>

4. Rejection Sampling

Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$



1) Sample x_1 from $p_2(x)$

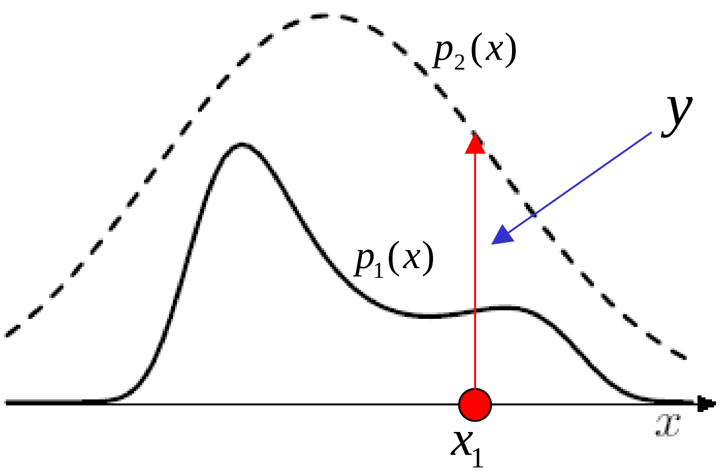
Monte Carlo Sampling

<u>Generating random variables</u>

4. Rejection Sampling

Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$



1) Sample x_1 from $p_2(x)$

2) Sample $y \sim U[0, p_2(x_1)]$

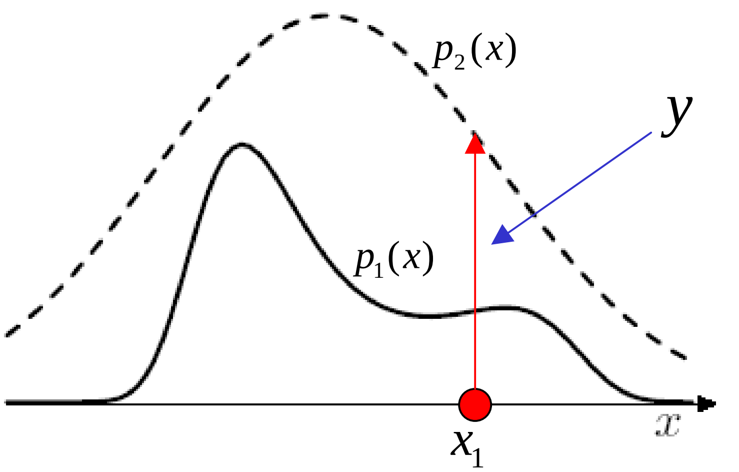
Monte Carlo Sampling

Generating random variables

4. Rejection Sampling

Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$



1) Sample x_1 from $p_2(x)$

2) Sample $y \sim U[0, p_2(x_1)]$

3) If $y < p_1(x)$ **ACCEPT**

otherwise **REJECT**

Monte Carlo Sampling

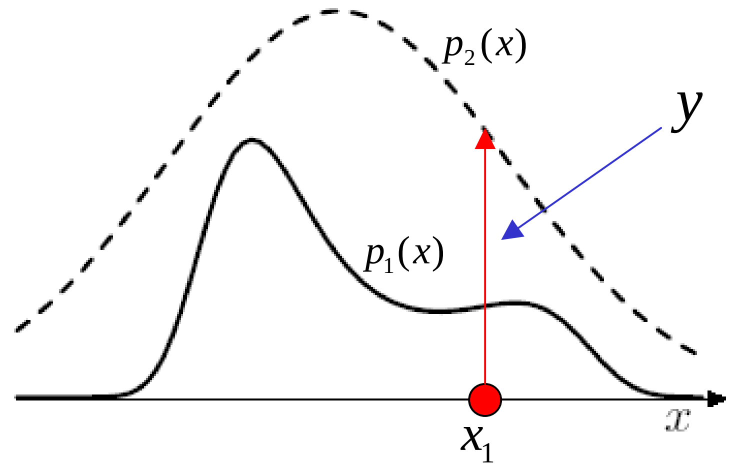
Generating random variables

4. Rejection Sampling

Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$

Set of accepted values $\{x_i\}$ *are a sample from* $p_1(x)$



1) Sample x_1 from $p_2(x)$

2) Sample $y \sim U[0, p_2(x_1)]$

3) If $y < p_1(x)$ **ACCEPT**

otherwise **REJECT**

Monte Carlo Sampling

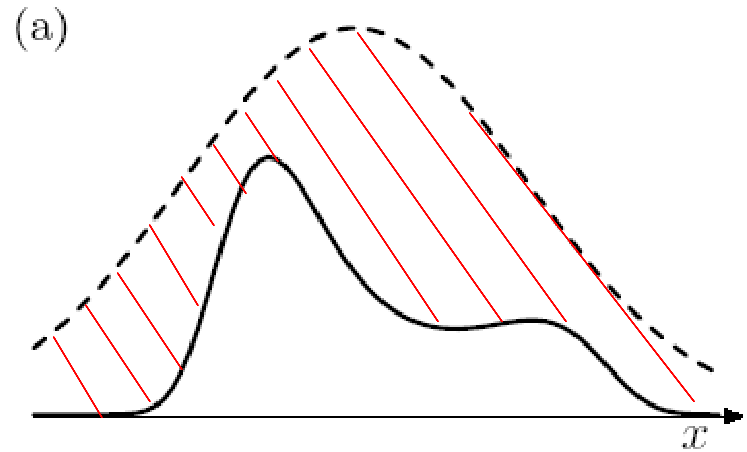
<u>Generating random variables</u>

4. Rejection Sampling

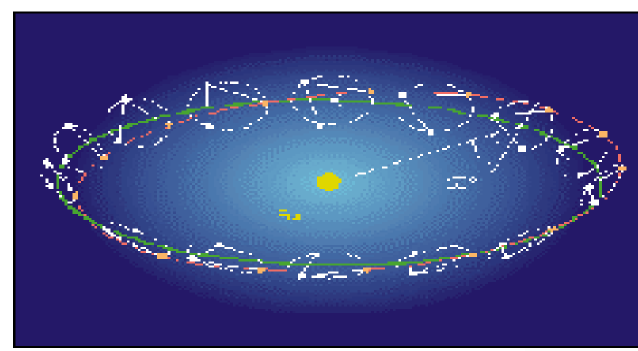
Suppose we want to sample from some pdf $p_1(x)$ and we know that

$$p_1(x) < p_2(x) \quad \forall x$$

Set of accepted values $\{x_i\}$ *are a sample from* $p_1(x)$



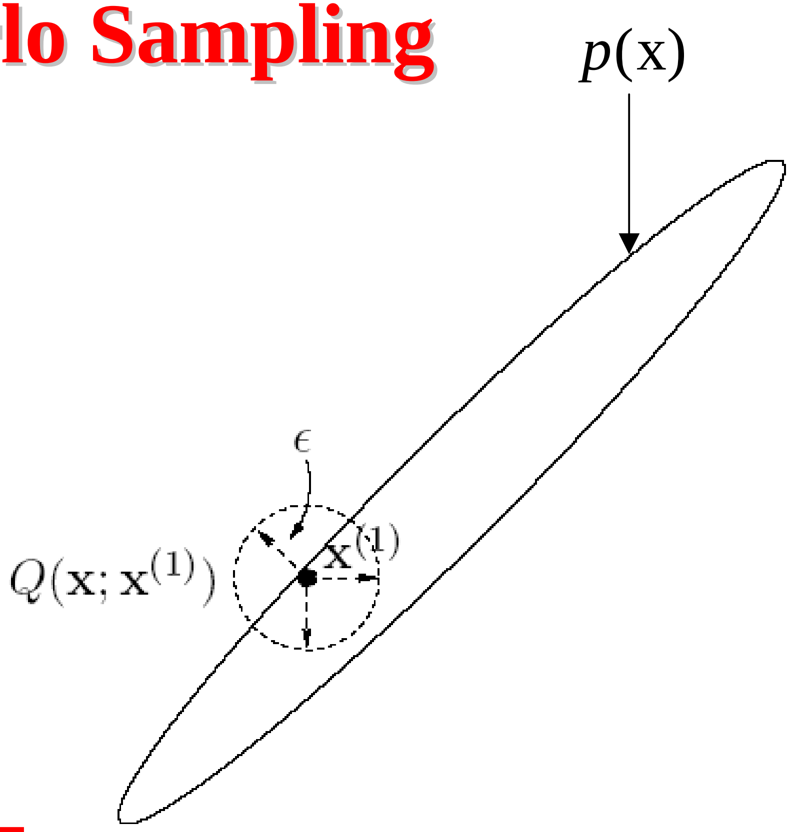
Method can be very slow if the shaded region is too large - particularly in high-N problems



LISA

UNIVERSITY of GLASGOW

Monte Carlo Sampling

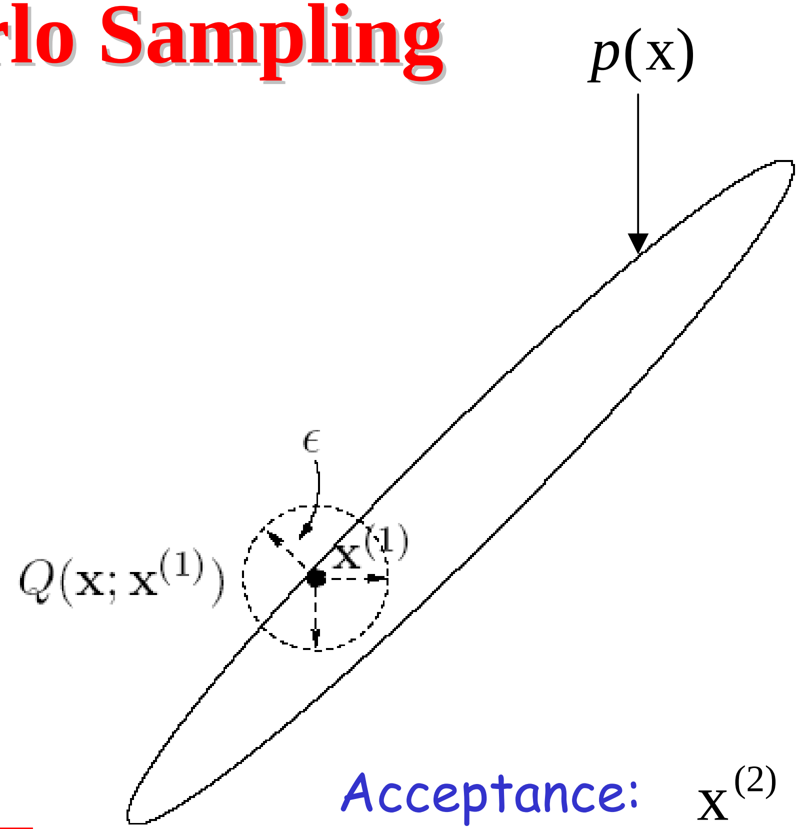


Generating random variables

5. Metropolis-Hastings Algorithm

- Sample initial point $x^{(1)}$
- Sample tentative new state from $Q(x', x^{(1)})$ (e.g. Gaussian)
- Compute $$\boxed{a = \frac{p(x')\,Q(x', x^{(1)})}{p(x^{(1)})\,Q(x^{(1)}, x')}}$$
- If $a > 1$ Accept
  Otherwise Accept with probability a

Monte Carlo Sampling



Generating random variables

5. Metropolis-Hastings Algorithm

- Sample initial point $\mathrm{x}^{(1)}$
- Sample tentative new state from $Q(\mathrm{x}',\mathrm{x}^{(1)})$ (e.g. Gaussian)
- Compute $\boxed{a = \dfrac{p(\mathrm{x}')\,Q(\mathrm{x}',\mathrm{x}^{(1)})}{p(\mathrm{x}^{(1)})\,Q(\mathrm{x}^{(1)},\mathrm{x}')}}$
- If $a > 1$ Accept
  Otherwise Accept with probability a

Acceptance: $\mathrm{x}^{(2)} = \mathrm{x}'$

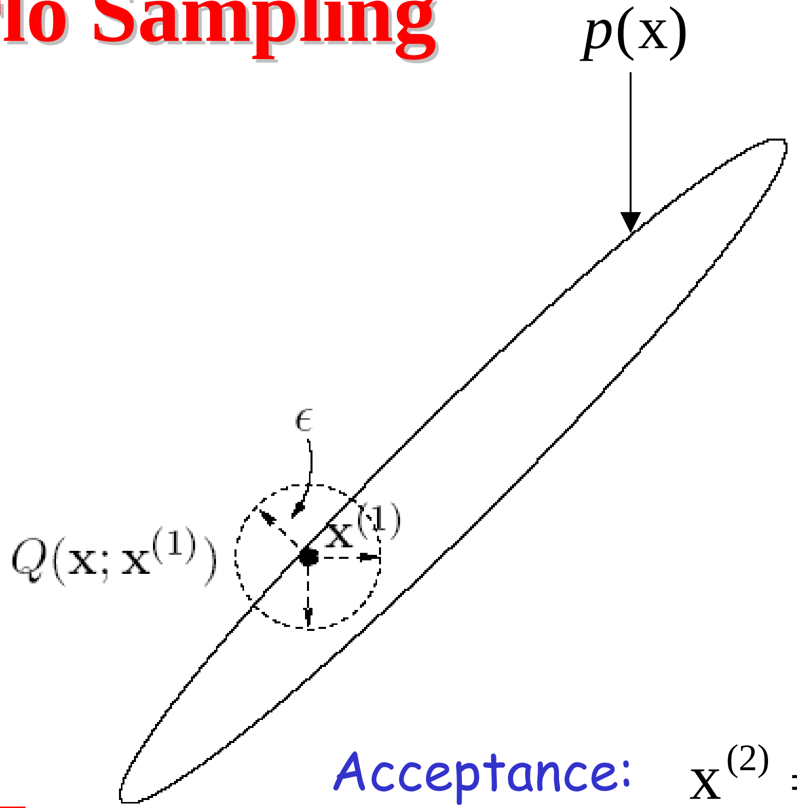
Rejection: $\mathrm{x}^{(2)} = \mathrm{x}^{(1)}$

Monte Carlo Sampling

Generating random variables

5. Metropolis-Hastings Algorithm

- Sample initial point $x^{(1)}$
- Sample tentative new state from $Q(x', x^{(1)})$ (e.g. Gaussian)
- Compute $\boxed{a = \dfrac{p(x')\,Q(x', x^{(1)})}{p(x^{(1)})\,Q(x^{(1)}, x')}}$
- If $a > 1$ Accept
  Otherwise Accept with probability a

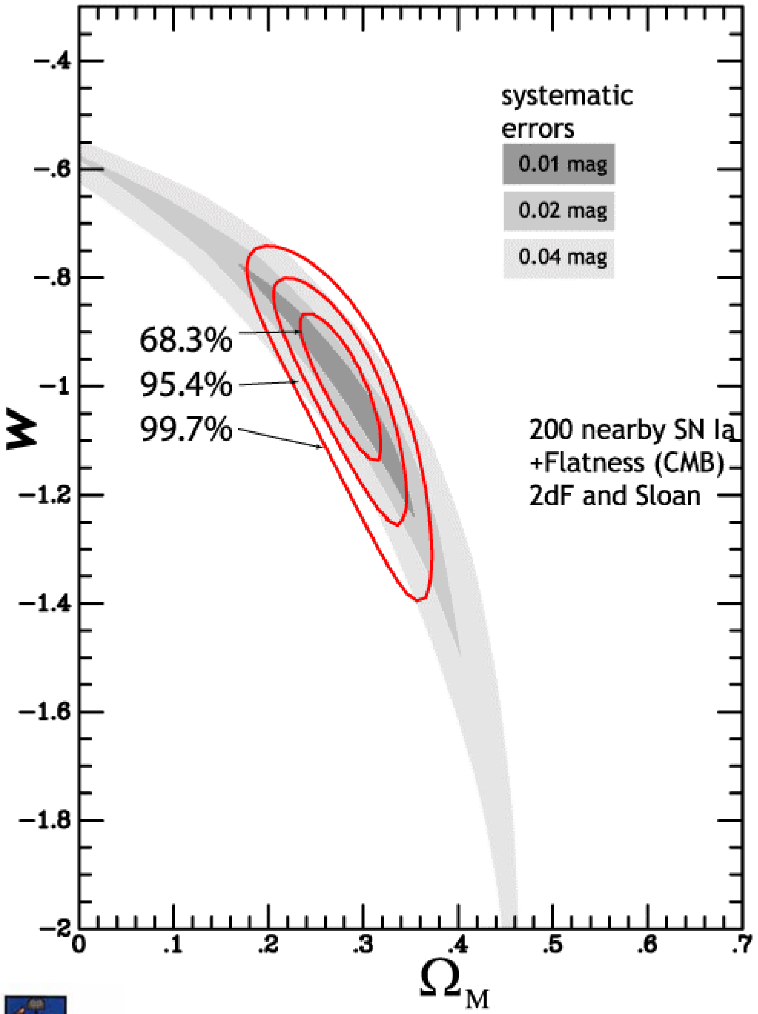


Acceptance: $x^{(2)} = x'$

Rejection: $x^{(2)} = x^{(1)}$

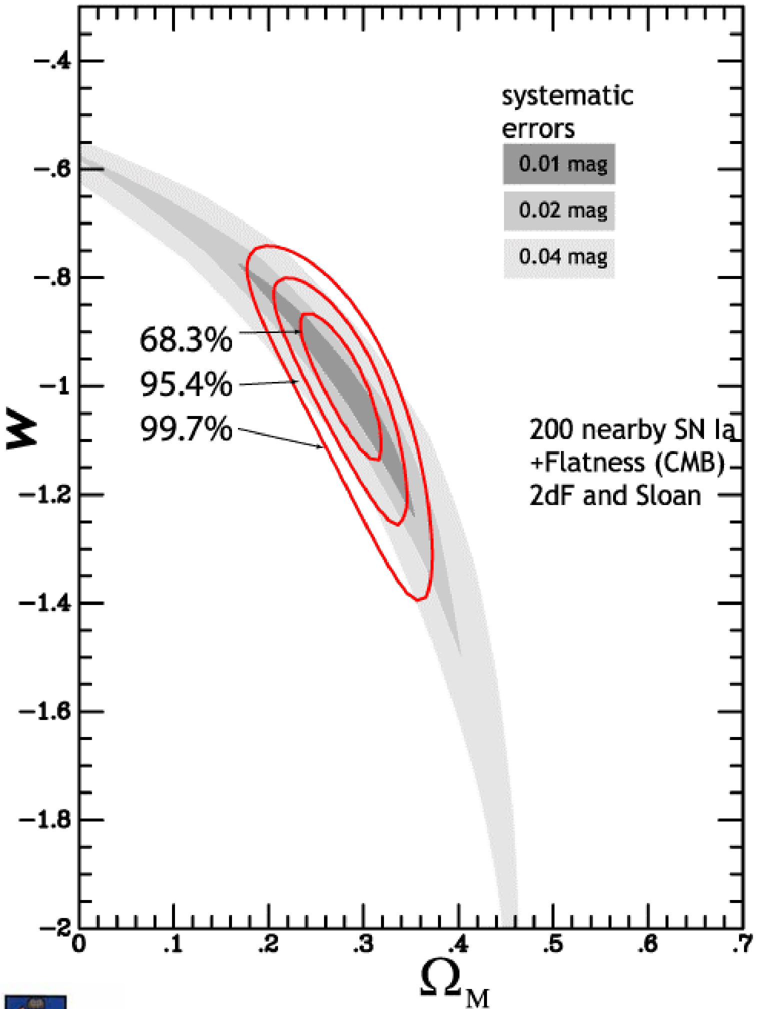
Markov Chain

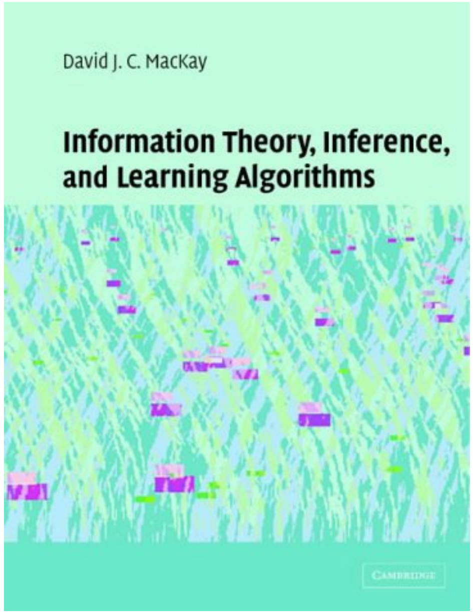
Monte Carlo Sampling

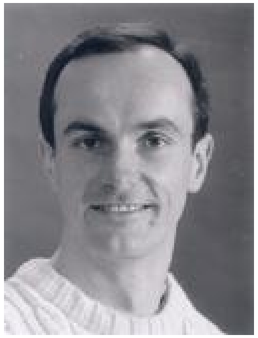


UNIVERSITY *of* GLASGOW

Monte Carlo Sampling







David Mackay

Cavendish Lab
Cambridge

http://www.inference.phy.cam.ac.uk/mackay/

http://www.statslab.cam.ac.uk/~mcmc/pages/links.html

UNIVERSITY of GLASGOW

IAU

Enjoy the ISYA, and keep in touch!!...