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

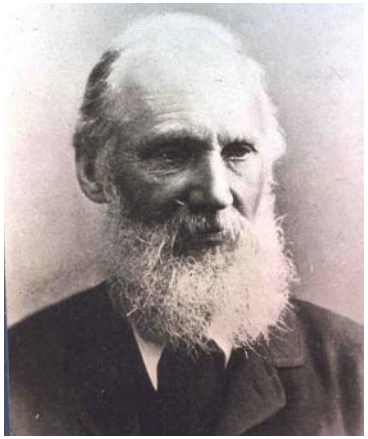
Martin Hendry,
Dept of Physics and Astronomy
University of Glasgow, UK

UNIVERSITY of GLASGOW

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

Who am I?...

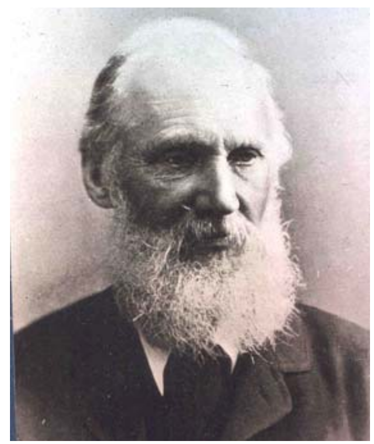
UNIVERSITY
of
GLASGOW



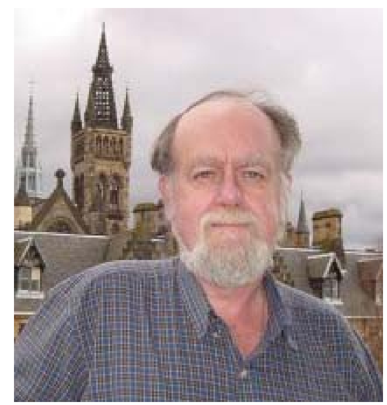
William Thompson
(Lord Kelvin)
1824 - 1907

UNIVERSITY
of
GLASGOW

Who am I?...



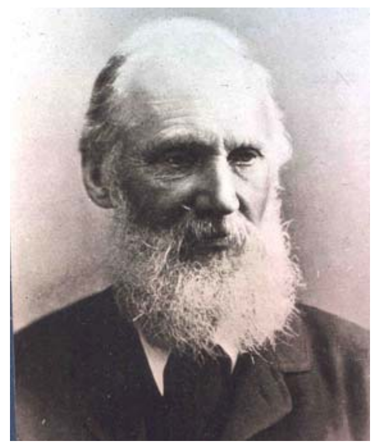
William Thompson
(Lord Kelvin)
1824 - 1907



John Brown
Astronomer Royal
for Scotland

Who am I?...

UNIVERSITY
of
GLASGOW



William Thompson
(Lord Kelvin)
1824 - 1907

"There is nothing new to be discovered in physics now. All that remains is more and more precise measurement"

UNIVERSITY
of
GLASGOW

ASTRONOMICAL SOCIETY OF THE PACIFIC
CONFERENCE SERIES

ASTRONOMICAL SOCIETY OF THE PACIFIC

Volume 126

FROM QUANTUM FLUCTUATIONS TO COSMOLOGICAL STRUCTURES

Proceedings of the First Moroccan School of Astrophysics
Casablanca, Morocco, 1-10 December 1996

Edited by
David Valls-Gabaud, Martin A. Hendry,
Paolo Molaro, and Khalil Chamcham

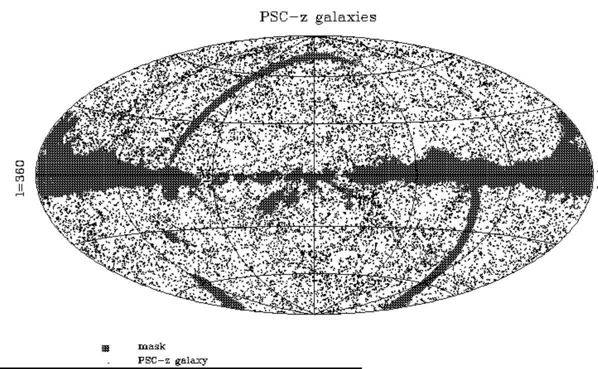
UNIVERSITY of GLASGOW

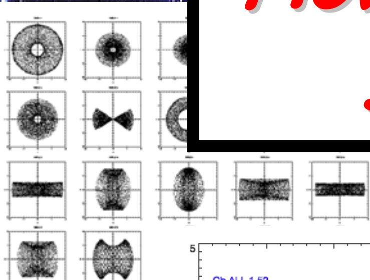
IAU

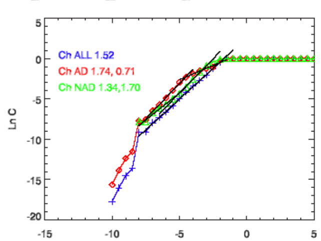
Principal Research Interests:

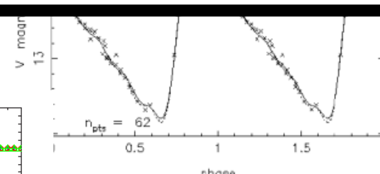
Cosmology – galaxy distances
- cosmic flows
- large-scale structure

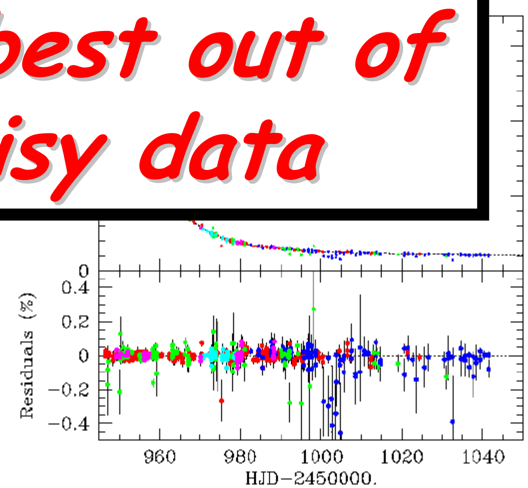
Gravitational microlensing











How to get the best out of sparse and noisy data

UNIVERSITY of GLASGOW

ISYA. Ifrane, 2nd – 23rd July 2004

Lecture Plan

- Why statistical astronomy?
- Introduction to probability theory
- Estimating parameters and testing models
- Observational selection effects
- Robust statistical methods
- Data compression methods
- Monte-Carlo sampling

Lectures 1 & 2

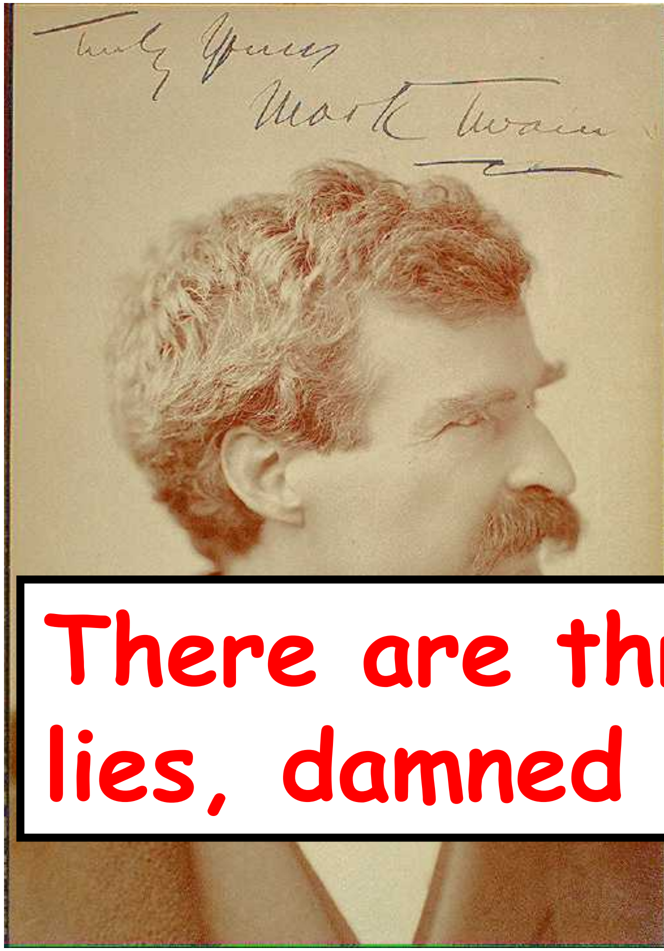
Lectures 3 & 4

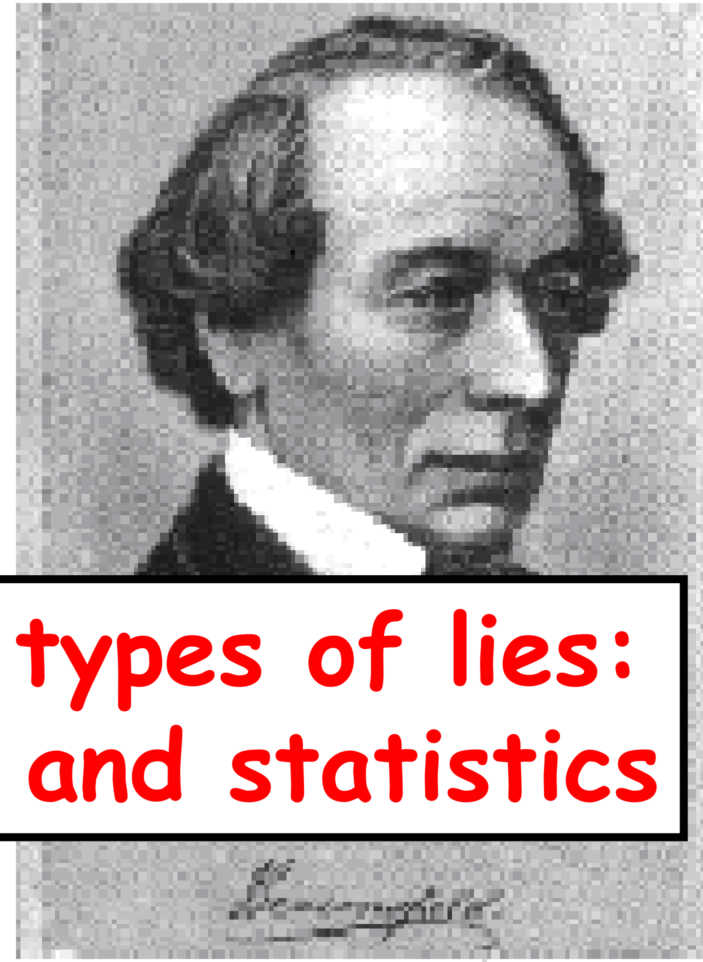
UNIVERSITY *of* GLASGOW

Why statistical astronomy?...

UNIVERSITY of GLASGOW

Why statistical astronomy?...





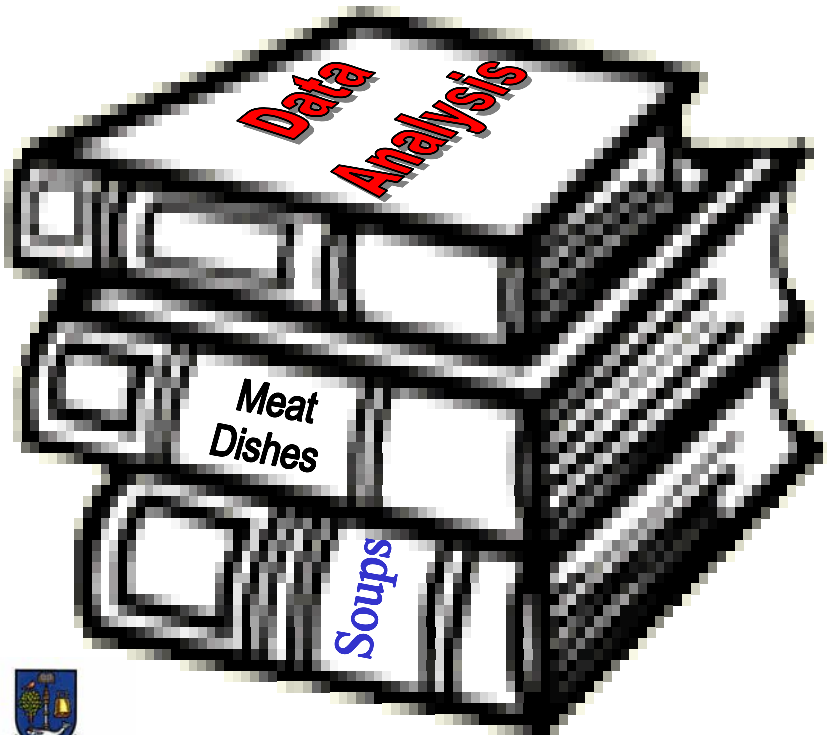
There are three types of lies: lies, damned lies and statistics

Mark Twain

Benjamin Disraeli

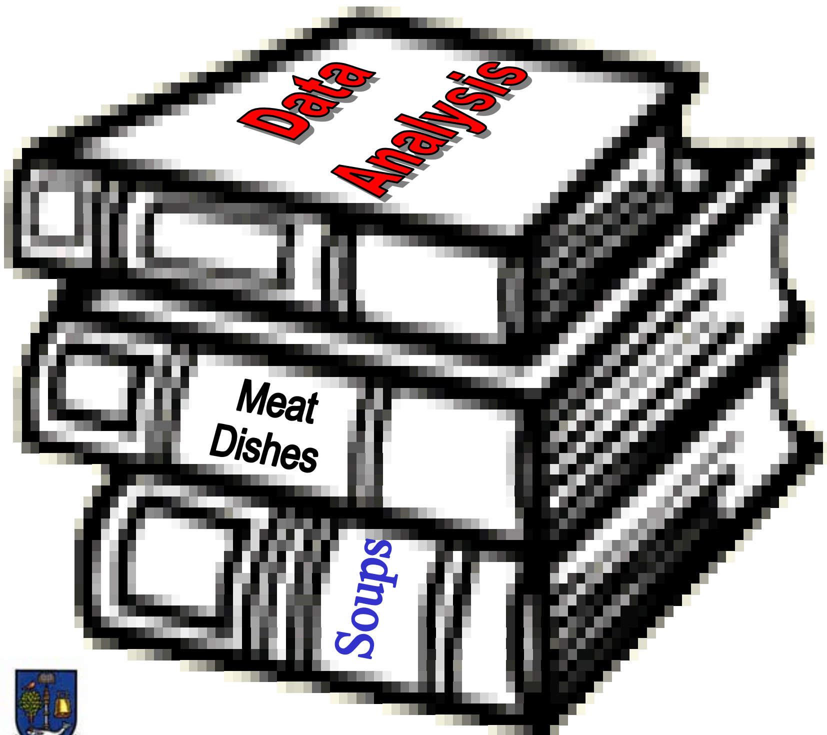
Why statistical astronomy?...

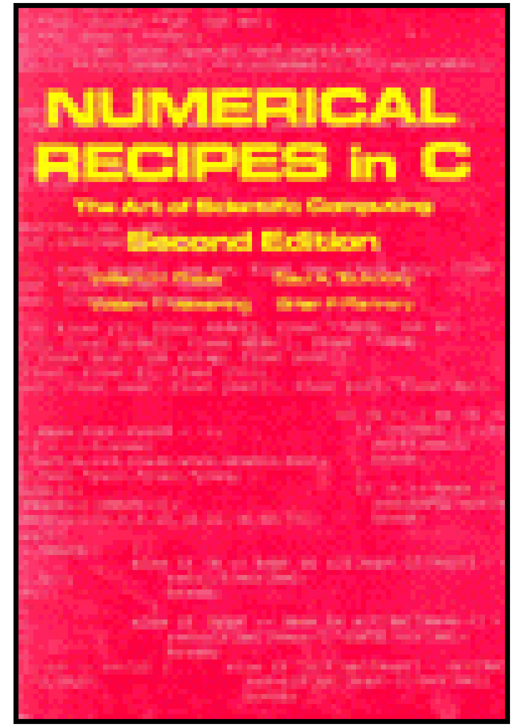
Data analysis methods are often regarded as simple recipes…



Why statistical astronomy?...

Data analysis methods are often regarded as simple recipes...





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

UNIVERSITY of GLASGOW

Why statistical astronomy?...

Data analysis methods are often regarded as simple recipes...

...but in astronomy, sometimes the recipes don't work!!!

- Low number counts
- Distant sources
- Correlated 'residuals'
- Incorrect assumptions

SYSTEMATIC ERRORS

How fast is the Universe expanding?

Galaxies are clustered

Clustering caused by gravity

Clustering distorts the Hubble expansion

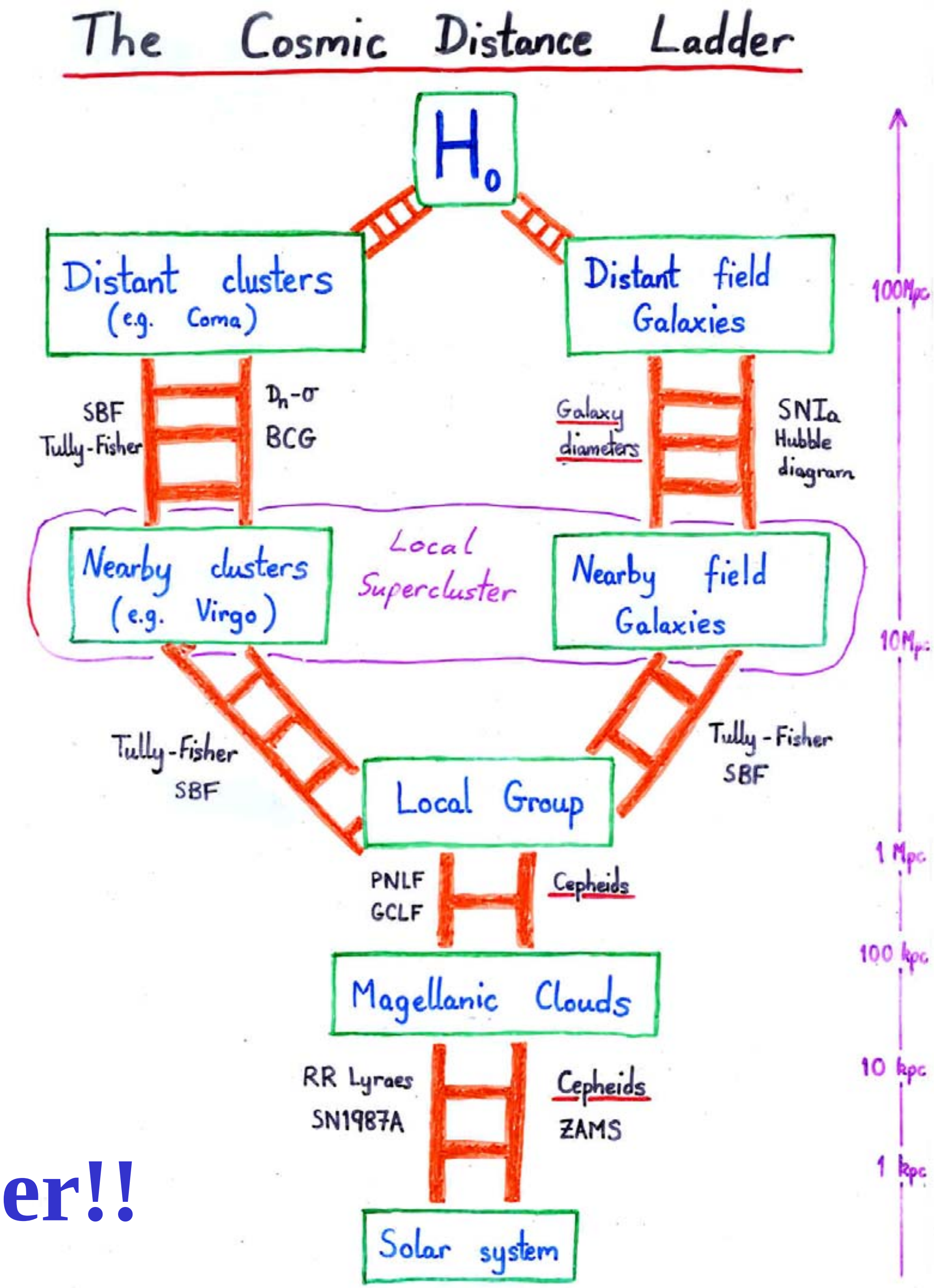
Problem:

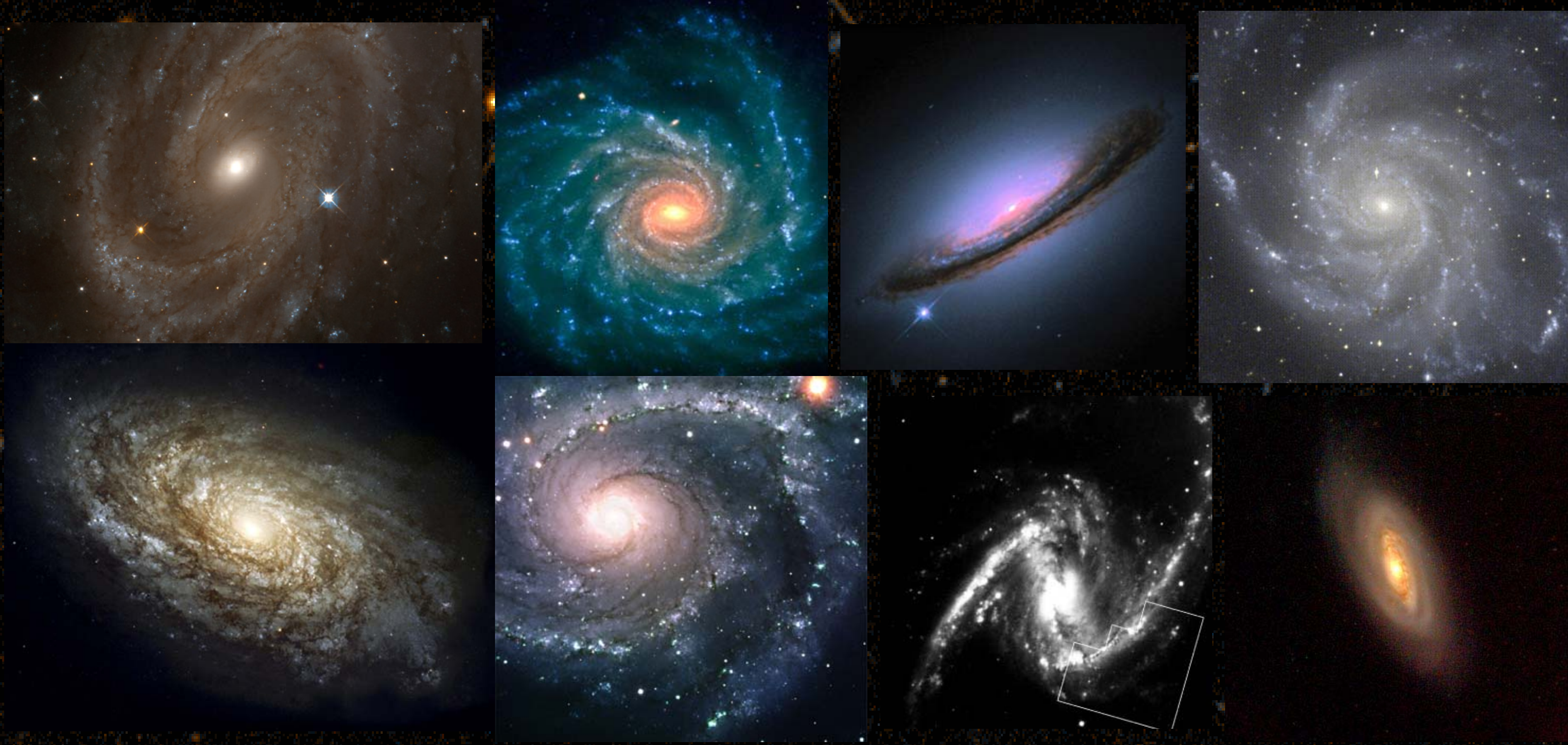
Need to determine H_0 from distant galaxies, where distortions are less important

….**but….**

Our most reliable distance methods can only be used nearby

Need Distance Ladder!!





Key Project of the Hubble Space Telescope

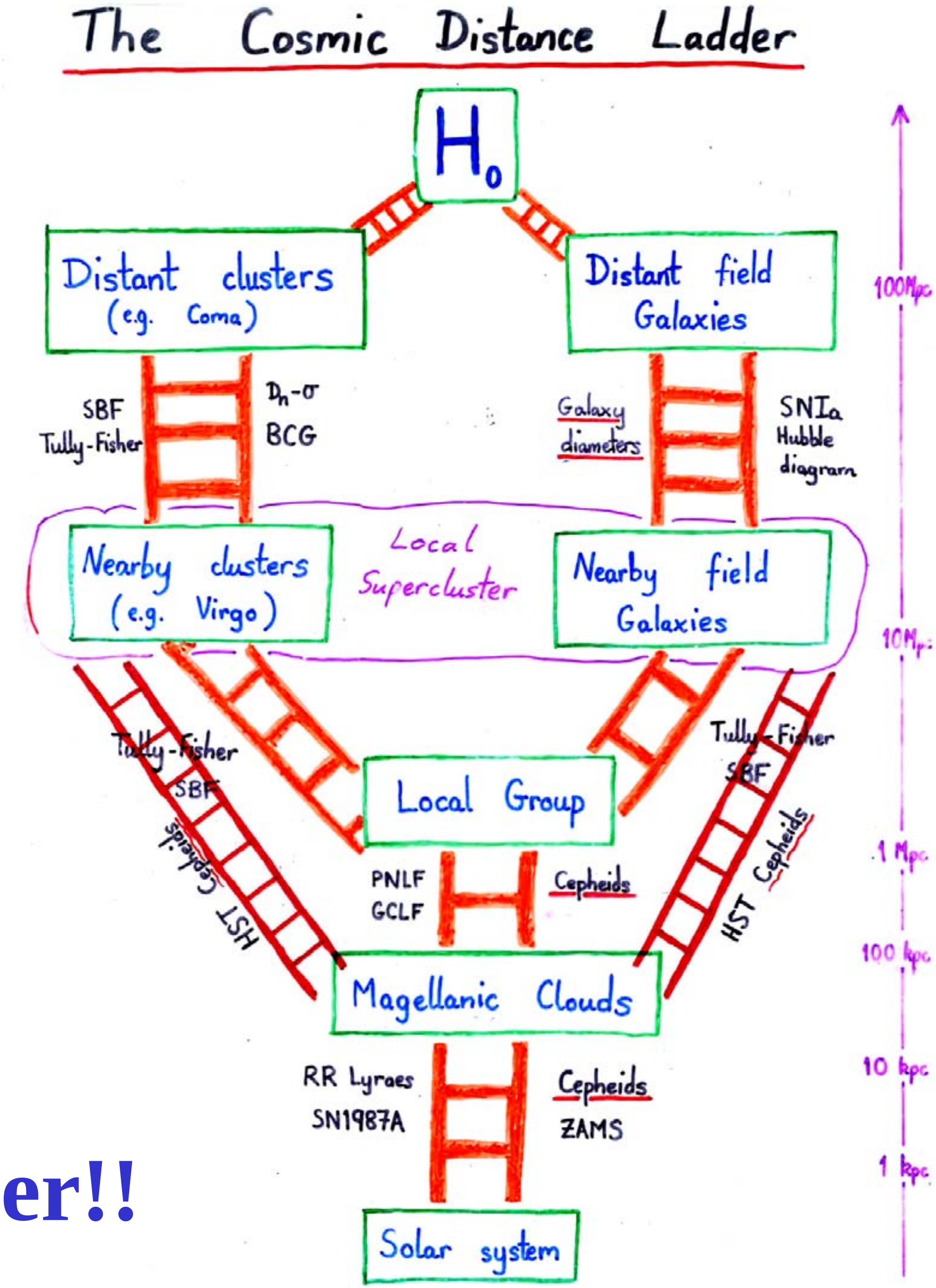
Problem:

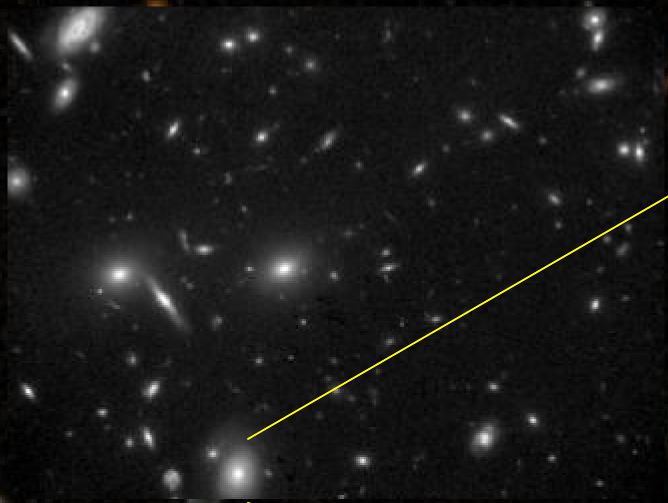
Need to determine H_0 from distant galaxies, where distortions are less important

….**but….**

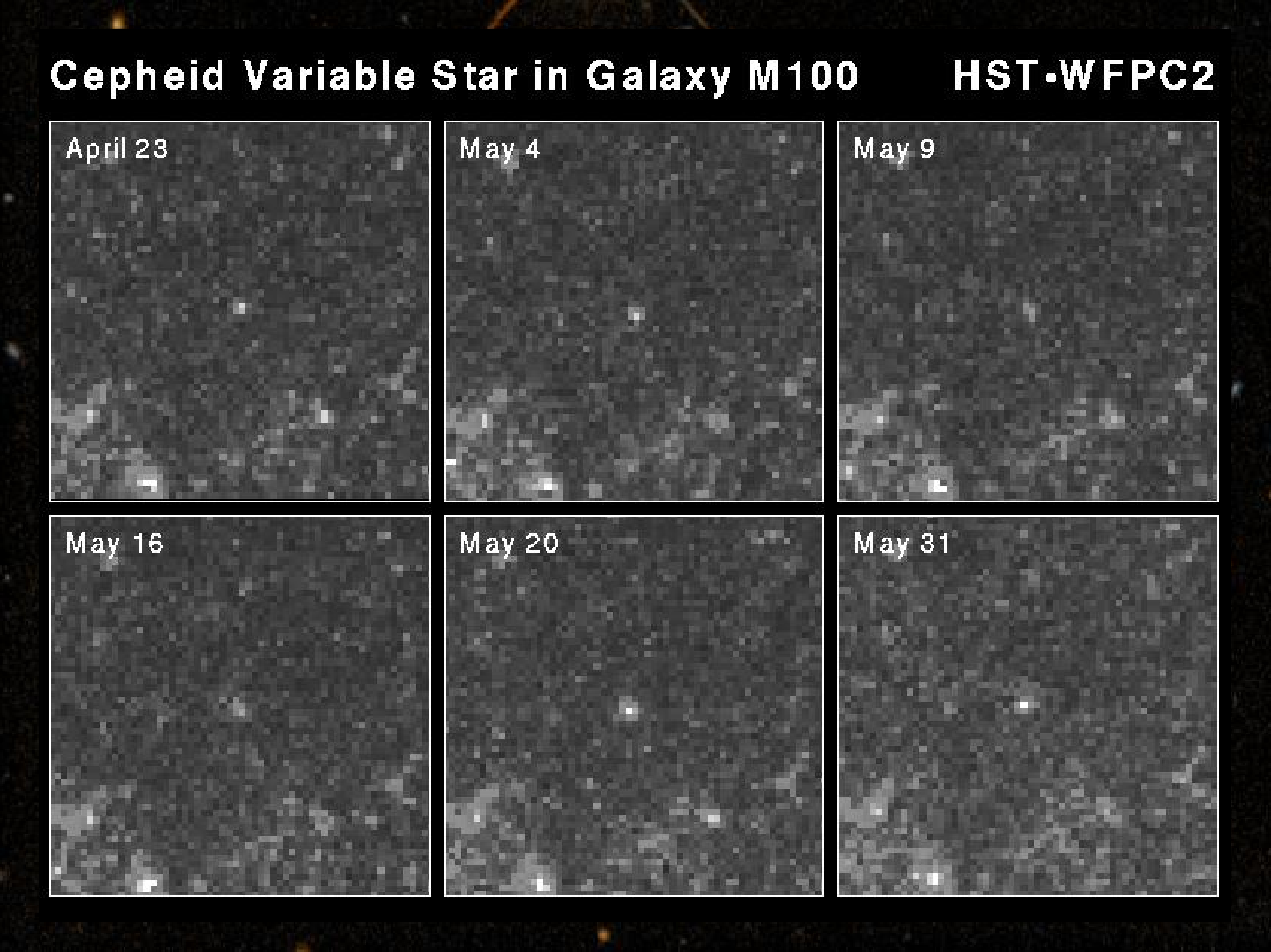
Our most reliable distance methods can only be used nearby

Need Distance Ladder!!





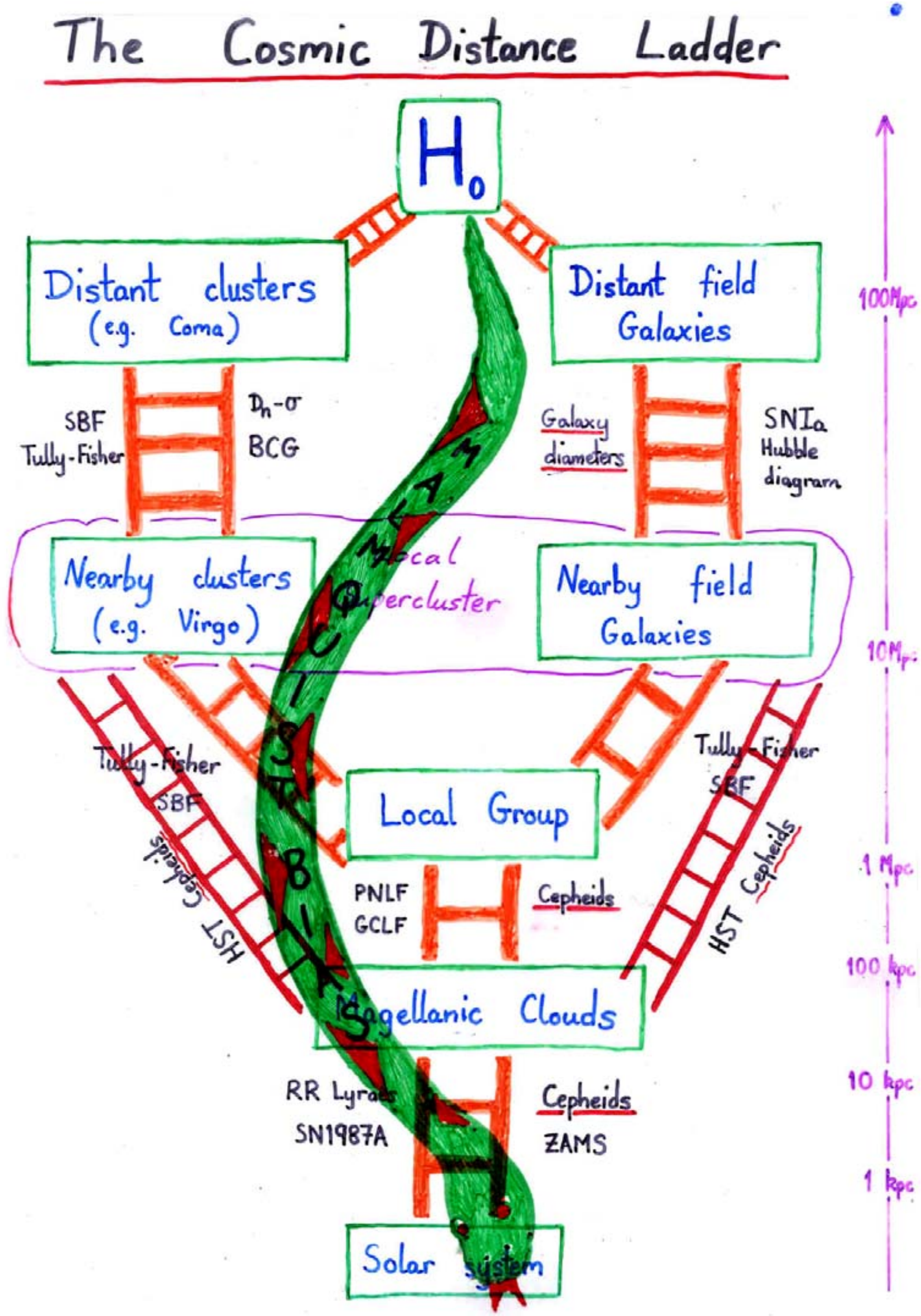
Virgo Cluster galaxy
M100, 60 million light years distant…..


Cepheid Variable Star in Galaxy M100
HST·WFPC2
April 23
May 4
May 9
May 16
May 20
May 31

Must check that distant galaxy data are free from **Selection Biases**

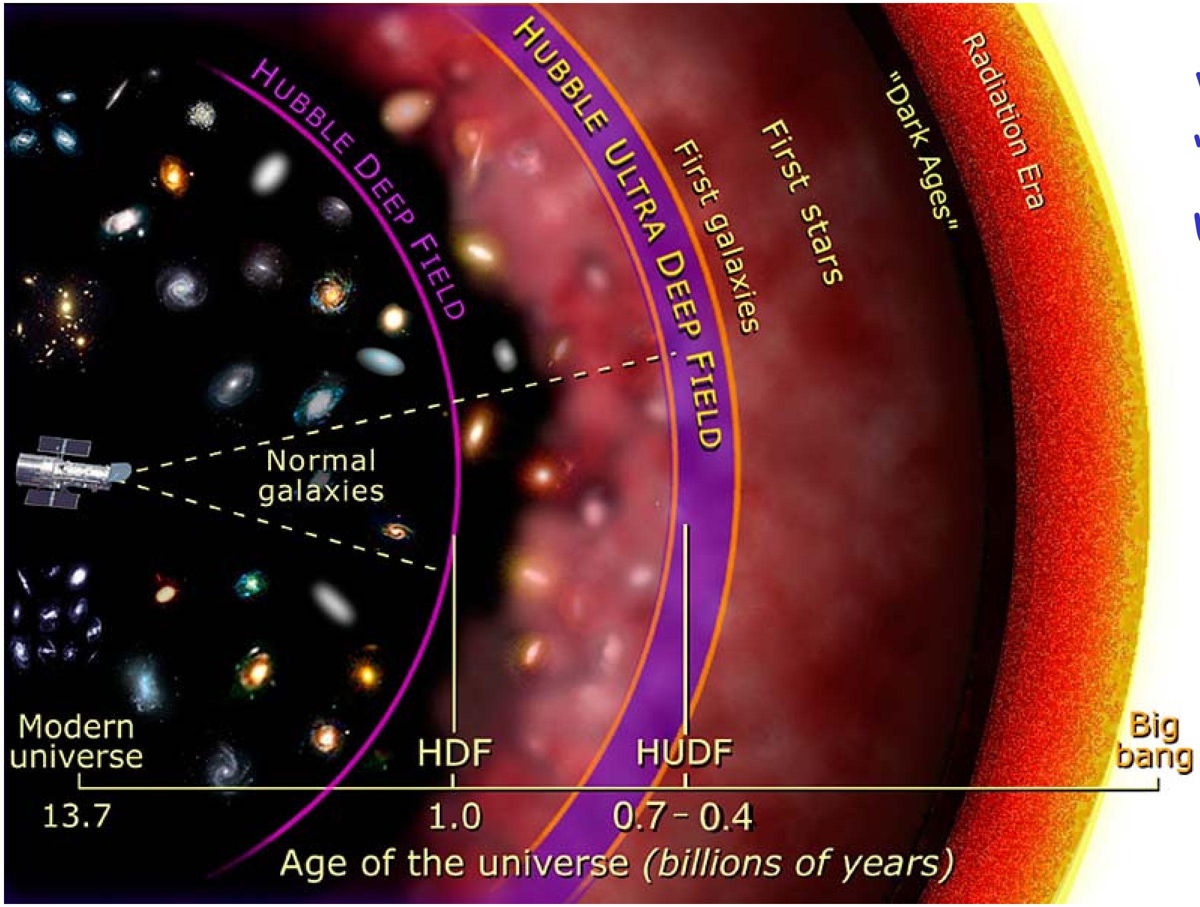
e.g. intrinsically brighter or bigger?…

Malmquist Bias



Why statistical astronomy?…

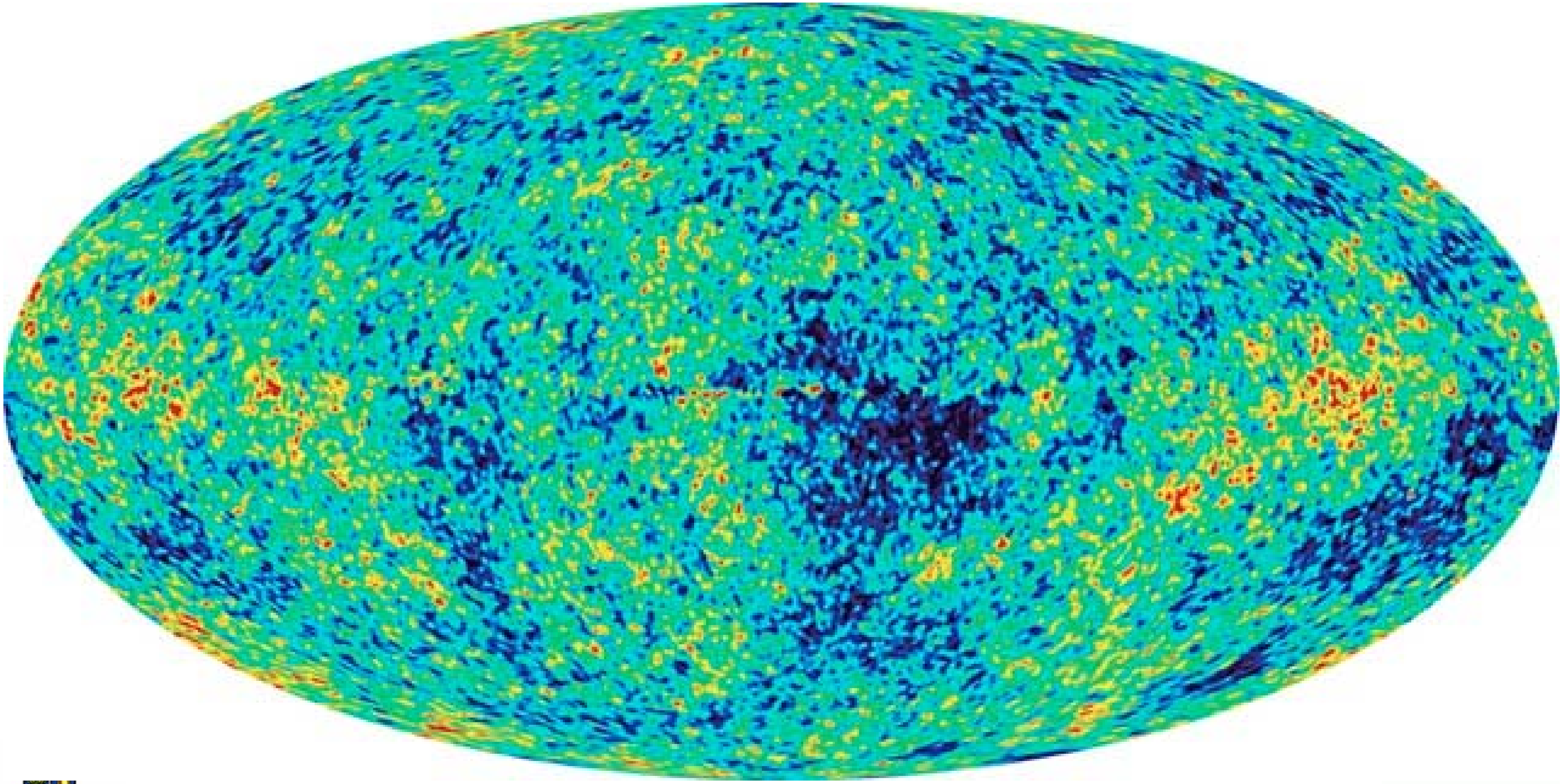
Astrophysics is *Remote sensing*



We can't ask:
"What happens if we change *this*?"

Why statistical astronomy?...

We observe only one Universe

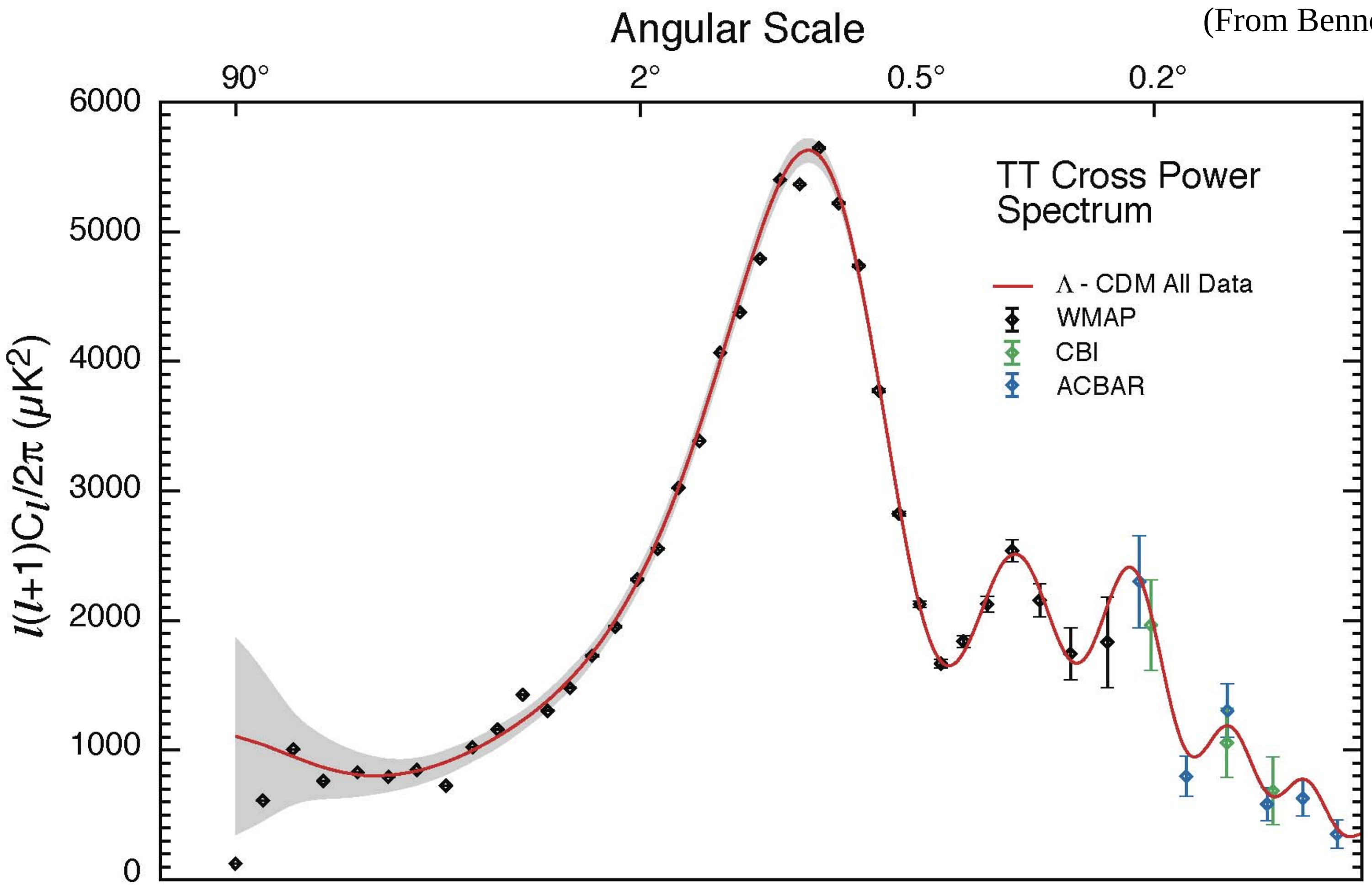


UNIVERSITY of GLASGOW

Why statistical astronomy?...

We observe only one Universe

(From Bennett et al 2003)



The Astrophysicist's Shopping List

We want tools for:

- dealing with very faint sources
- handling very large data sets
- correcting for selection effects
- diagnosing systematic errors
- avoiding unnecessary assumptions

- **estimating parameters and testing models**

Why statistical astronomy?...

Key question:

How do we infer properties of the Universe from incomplete and imprecise astronomical data?

Answer:

Use probability theory!

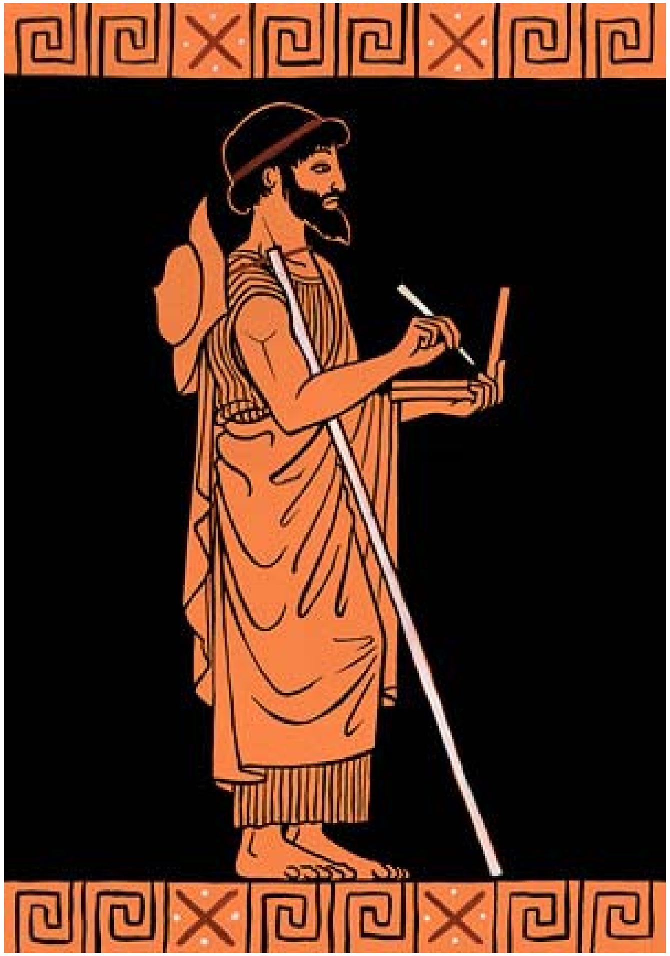
Why statistical astronomy?...

Answer:

Use probability theory!

Our goal:

To make the best inference, based on our observed data and any prior knowledge, reserving the right to revise our position if new information comes to light.



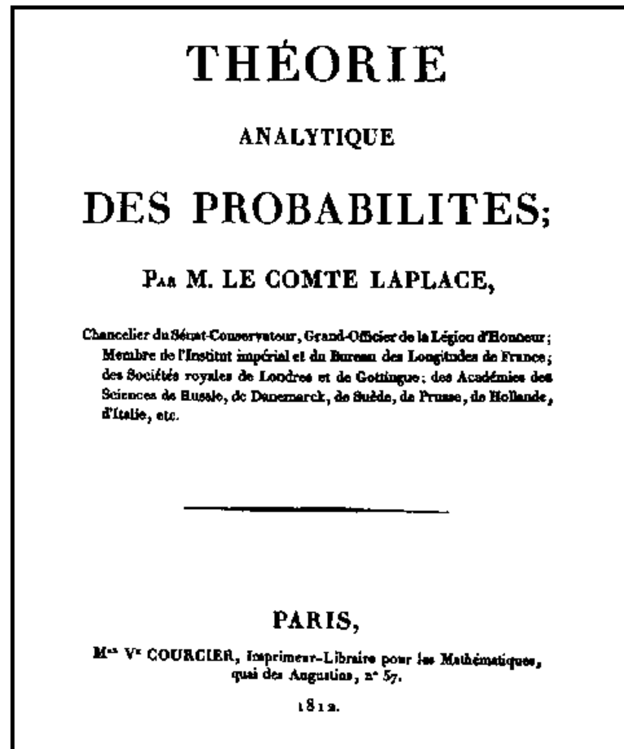
Herodotus, c.500 BC

"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"

UNIVERSITY of GLASGOW

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, n° 57.

1812.

UNIVERSITY of GLASGOW

William of Ockham
(1288 – 1348 AD)

Ockham's Razor

"Frustra fit per plura, quod fieri potest per pauciora."

"It is vain to do with more what can be done with less."

Everything else being equal, we favour models which are ***simple***.

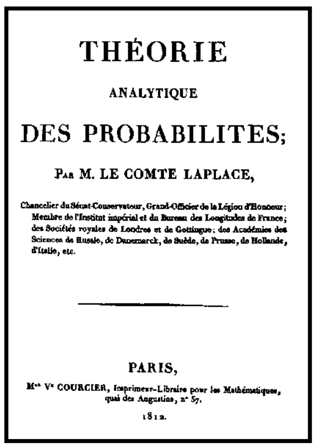
UNIVERSITY of GLASGOW

A brief history of probability theory

Early 1700s

"How can we apply the rules of games of chance to inference in everyday life?"

Johann Bernoulli
(1667 – 1748 AD)



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

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

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

UNIVERSITY of GLASGOW

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

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

"Probability that X is true, given I"

Background information

Vertical line denotes conditional probability:

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

UNIVERSITY of GLASGOW

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

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

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)

UNIVERSITY of GLASGOW

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

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

UNIVERSITY of GLASGOW

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)

Thomas Bayes
(1702 – 1761 AD)

UNIVERSITY of GLASGOW

Bayes' theorem:

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

Posterior Likelihood Prior

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

Evidence

We can calculate these terms

UNIVERSITY of GLASGOW

Bayes' theorem:

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

Posterior

Likelihood

Prior

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

What we know now

Influence of our observations

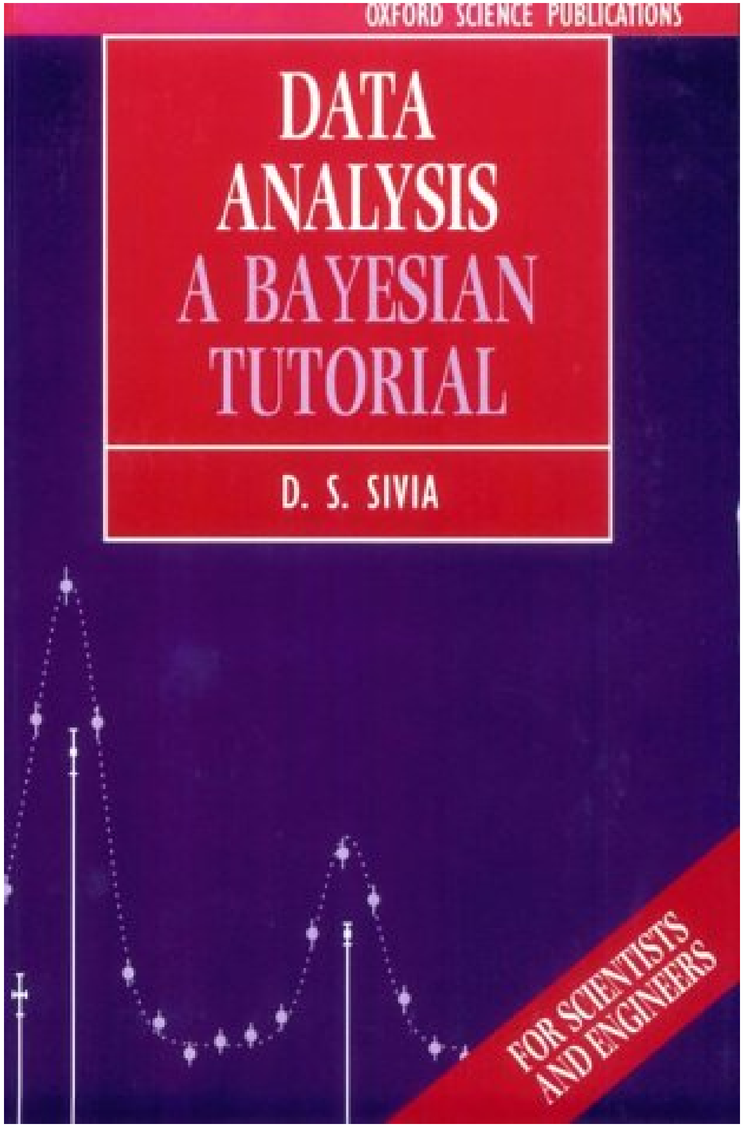
What we knew before

UNIVERSITY of GLASGOW

This equation is the key to ***Bayesian Inference*** – the methodology upon which (most) astronomical data analysis is now founded.

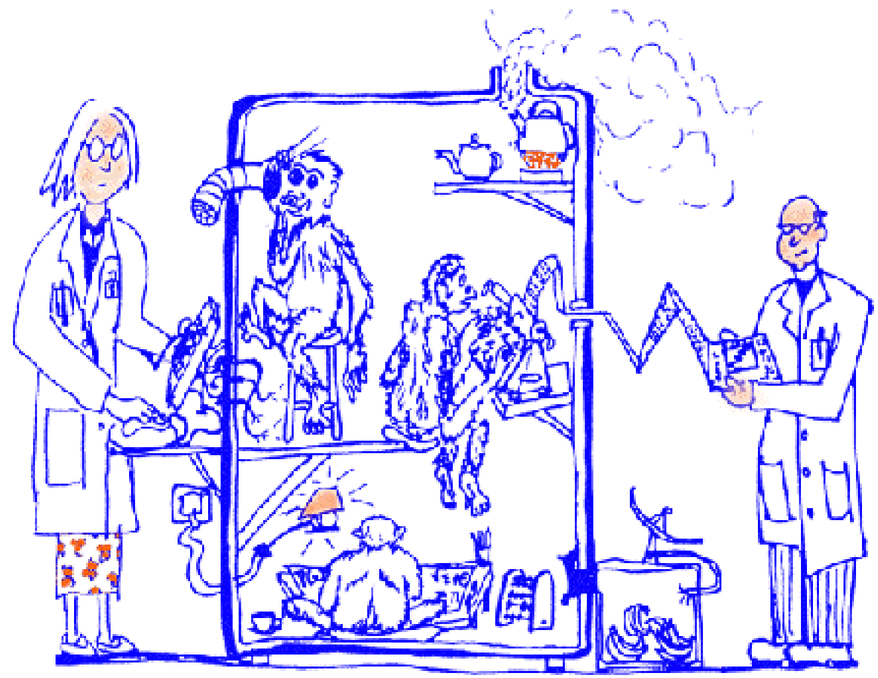
Clear introduction in **Sivia**

(~ $45 from amazon)



UNIVERSITY of GLASGOW

This equation is the key to ***Bayesian Inference*** – the methodology upon which (most) astronomical data analysis is now founded.



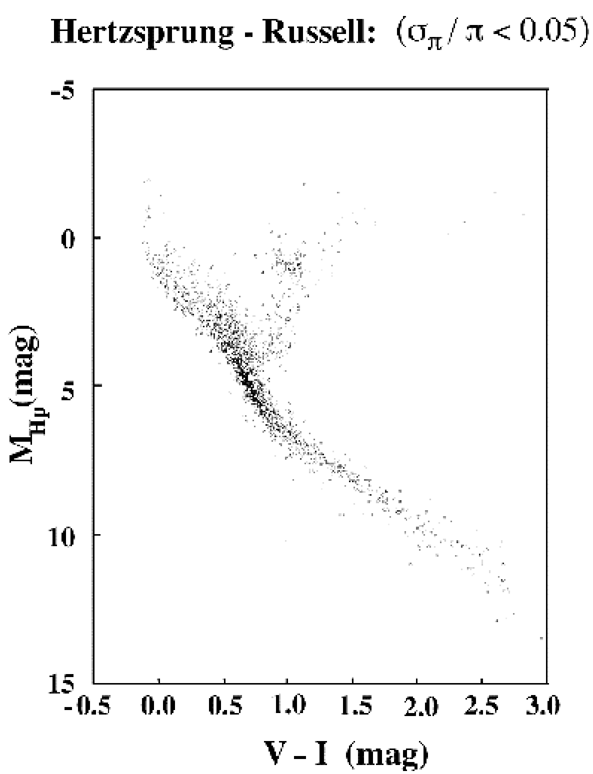
See also **free book** by Praesenjit Saha (QMW, London).

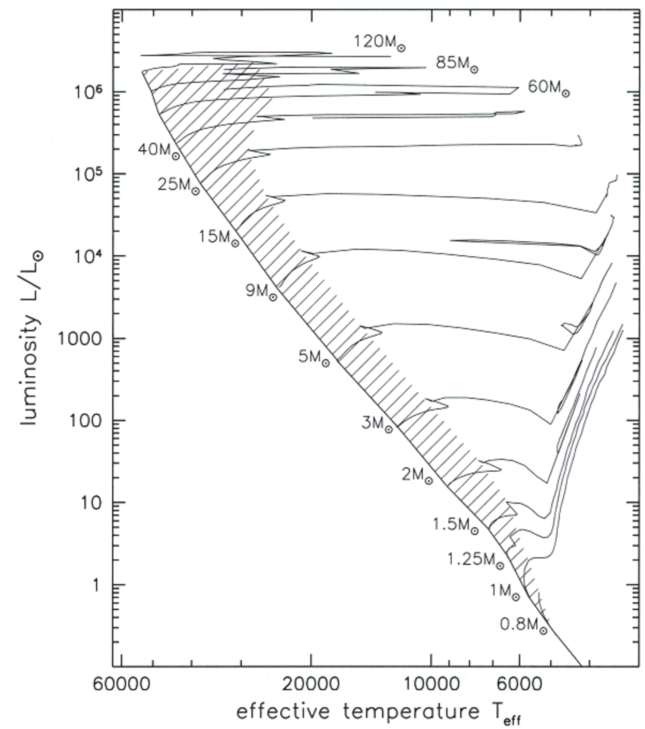
Can be downloaded from

http://ankh-morpork.maths.qmw.ac.uk/%7Esaha/book

Or see ISYA website / MAH ISYA webpage

Example: Stellar Evolution





Model = theoretical evolutionary tracks

+ $(L, T_{\text{eff}}) \Leftrightarrow (M, V - I)$

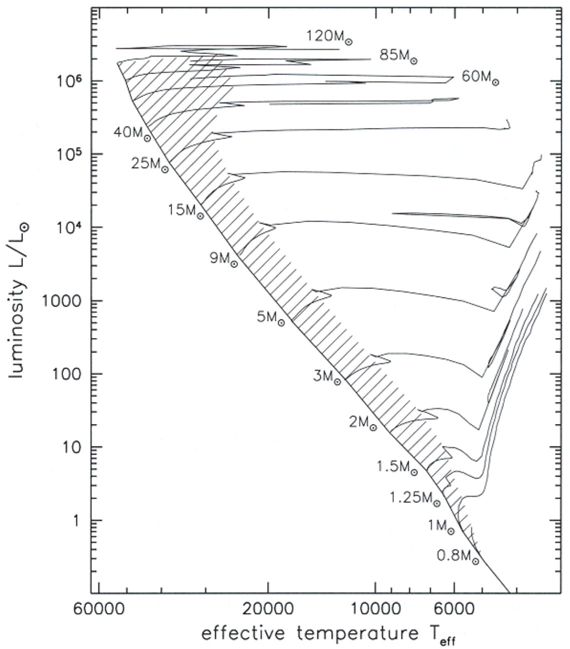
+ $(m, (V - I)_{\text{obs}}, \pi) \Leftrightarrow (M, V - I)$

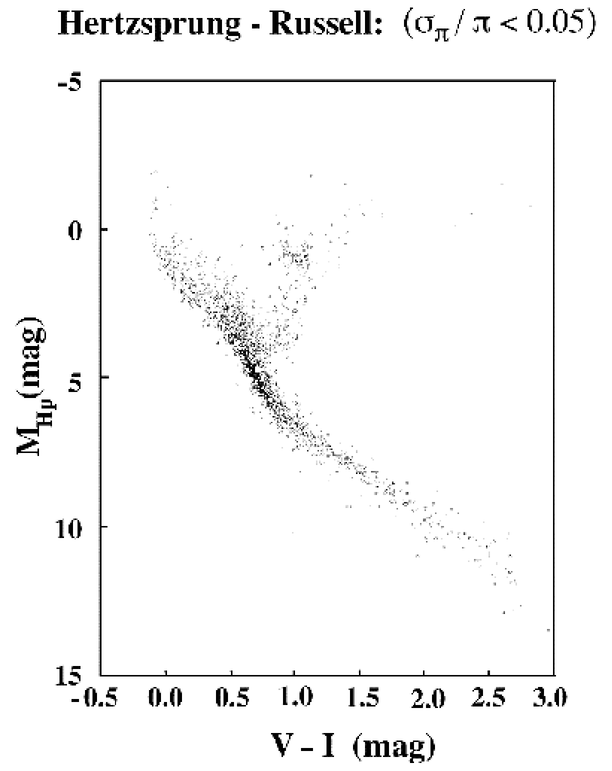
Data = 'observed' HIPPARCOS colour-magnitude diagram

UNIVERSITY of GLASGOW

Example: Stellar Evolution

Likelihood: $p(\mathrm{data} \mid \mathrm{model}, I)$





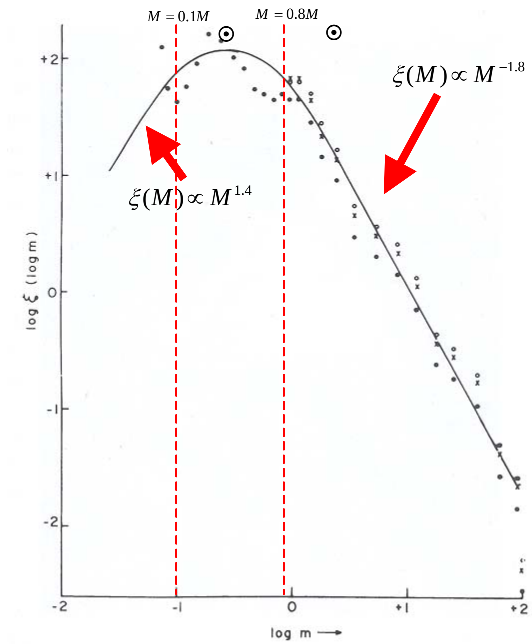
Example: Stellar Evolution

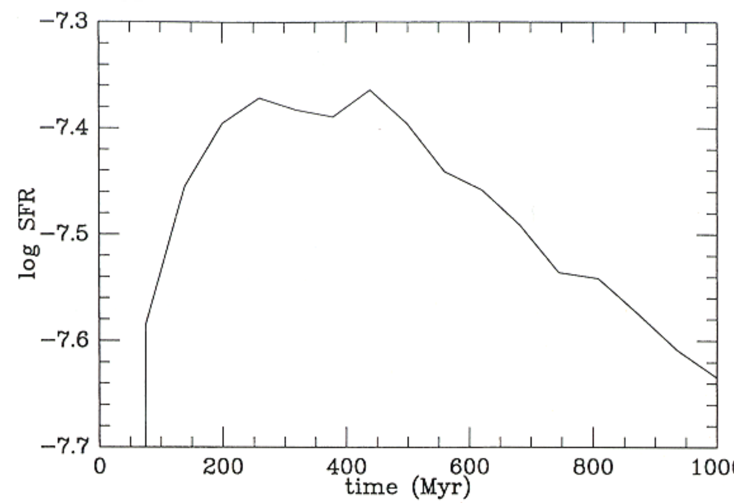
Prior: $p(\text{model} \mid I)$

Expresses our prior belief / assumptions about the model *before* our observations

e.g. stellar IMF, SFR,
ages, metallicities,
extinction law,
parallax uncertainties

Also background assumptions –
e.g. separable IMF, SFR
instantaneous recycling





Marginalisation

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

e.g. X_k = team k wins Euro 2004

Then $$\sum_{k=1}^{16} p(X_k \mid I) = 1$$

I = state of knowledge at 12/06/04

Marginalisation

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

e.g. X_k = team k wins Euro 2004

Then $\displaystyle\sum_{k=1}^{16} p(X_k \mid I) = 1$

I = state of knowledge at 24/06/04

Marginalisation

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

e.g. X_k = team k wins Euro 2004

Then $$\sum_{k=1}^{16} p(X_k \mid I) \;=\; 1$$

I = state of knowledge at 30/06/04

UNIVERSITY of GLASGOW

Marginalisation

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

e.g. X_k = team k wins Euro 2004

Then $\displaystyle\sum_{k=1}^{16} p(X_k \mid I) = 1$

I = state of knowledge at 04/07/04

Marginalisation

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

e.g. X_k = team k wins Euro 2004

Then $\displaystyle\sum_{k=1}^{16} p(X_k \mid I) \; = \; 1$

I = state of knowledge at 05/07/04

UNIVERSITY of GLASGOW

Marginalisation

Let Y = winning team wears red

Then $p(Y \mid I) = 0$

I = state of knowledge at 04/07/04

Marginalisation

Let Y = winning team wears red

What about $p(Y \mid I)$ at 12/06/04 ?

Marginalisation

Let Y = winning team wears red

What about $p(Y \mid I)$ at 12/06/04 ?

Use Bayes' theorem.

$$p(X_1, Y \mid I) = p(X_1 \mid Y, I)\, p(Y \mid I)$$

$$\vdots$$

$$p(X_{16}, Y \mid I) = p(X_{16} \mid Y, I)\, p(Y \mid I)$$

UNIVERSITY of GLASGOW

Marginalisation

Let Y = winning team wears red

What about $p(Y \mid I)$ at 12/06/04 ?

Use Bayes' theorem.

$$\sum_{k=1}^{16} p(X_k, Y \mid I) = \underbrace{\left[\sum_{k=1}^{16} p(X_k \mid Y, I) \right]}_{=\,1} p(Y \mid I)$$

Marginalisation

Let Y = winning team wears red

What about $p(Y|I)$ at 12/06/04 ?

Use Bayes' theorem.

$$\sum_{k=1}^{16} p(X_k, Y | I) = \underbrace{\left[\sum_{k=1}^{16} p(X_k | Y, I) \right]}_{=\,1} p(Y | I)$$

Marginal probability

$$\boxed{p(Y|I) = \sum_{k=1}^{16} p(X_k, Y | I)}$$