Cosmologia Física

Ismael Tereno (FCUL, IA)

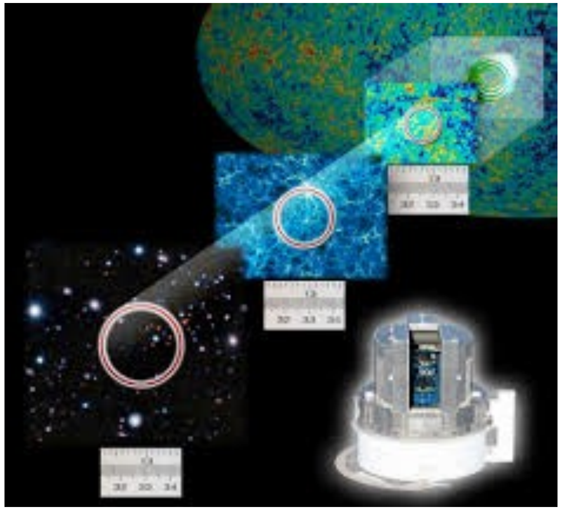
Ciências ULisboa

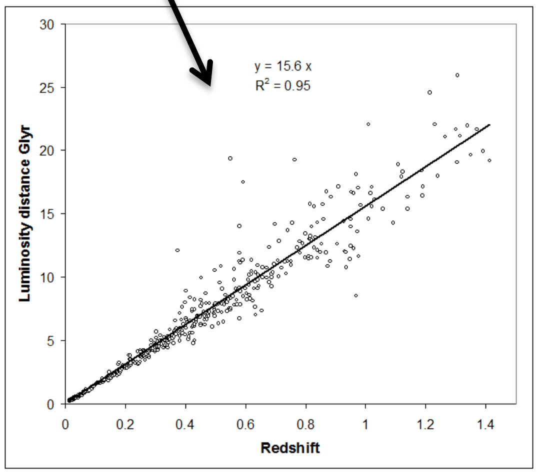
instituto de astrofísica e ciências do espaço

2026

Measuring the Universe

Probes of geometry





e.g. Distance (z)

observables

(fluxes, map of CMB, map of galaxy positions, redshifts, map of galaxy shapes)

estimate physical properties

(distances D(z), power spectra $P_{\delta}(k)$, C_T (l))

Likelihood analysis

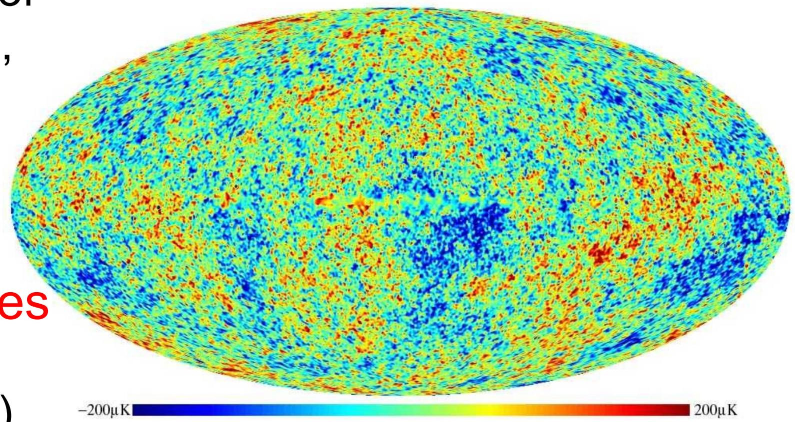
compute physical properties

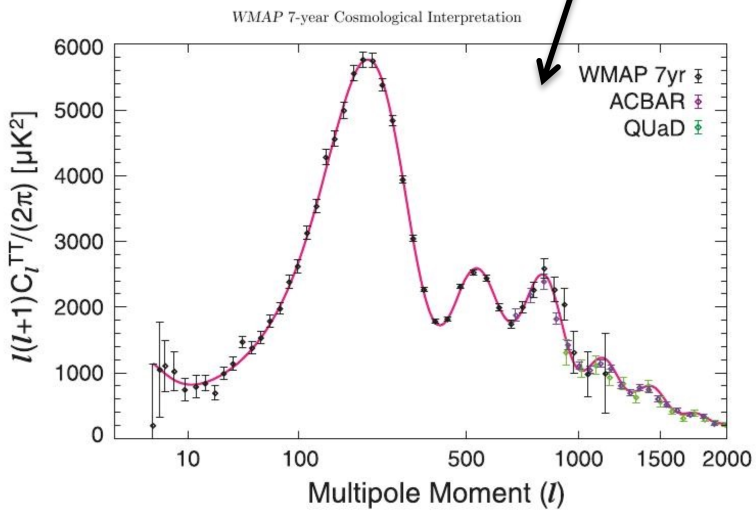
(distances D(z), power spectra $P_{\delta}(k)$, C_T (l))

cosmological models

(equations + Ω_i, w_i, $H_{0,}$ n_s, σ_8)

Probes of structure





e.g. Clustering (r)

Cosmological probes

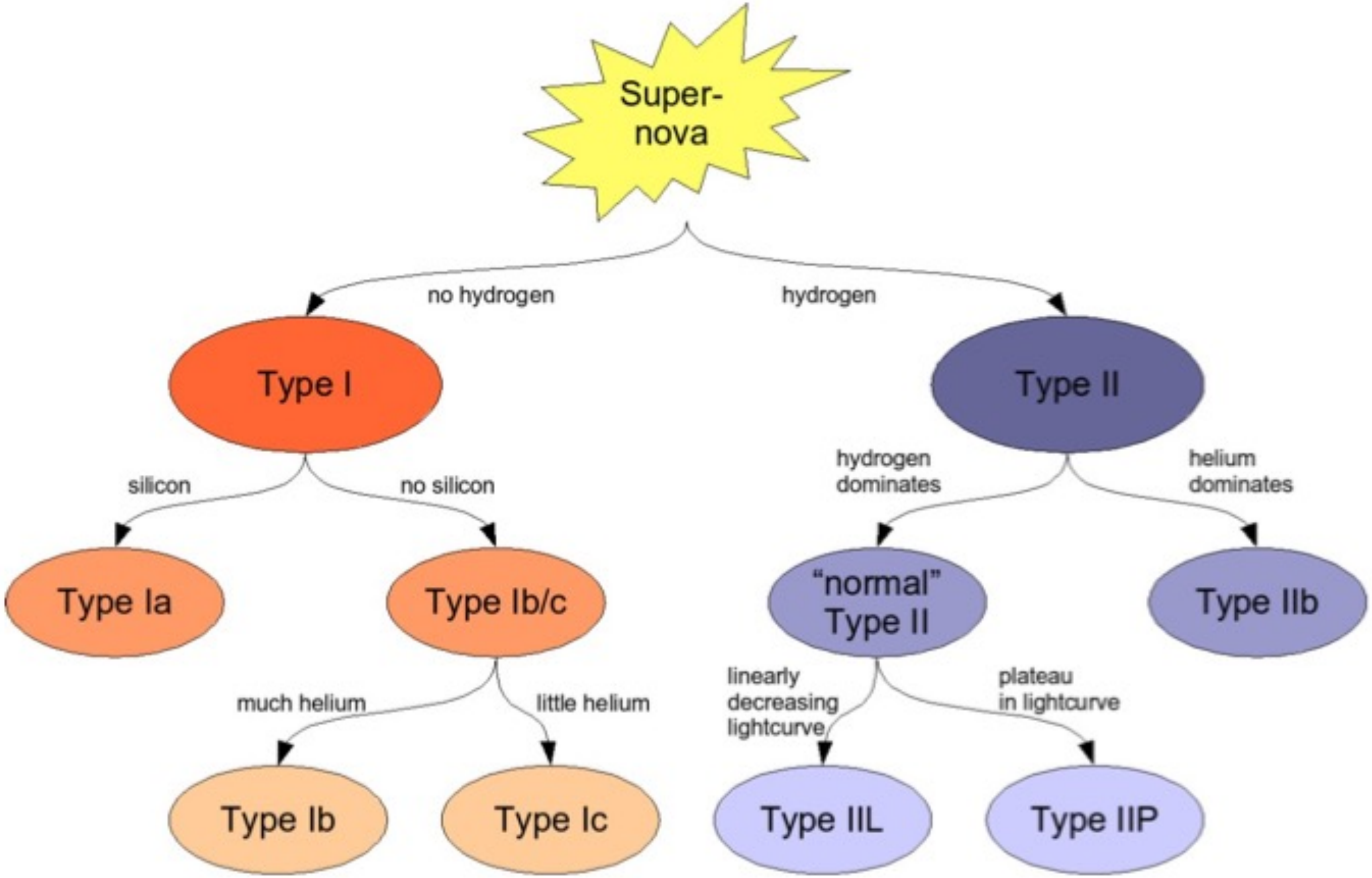
Supernova surveys

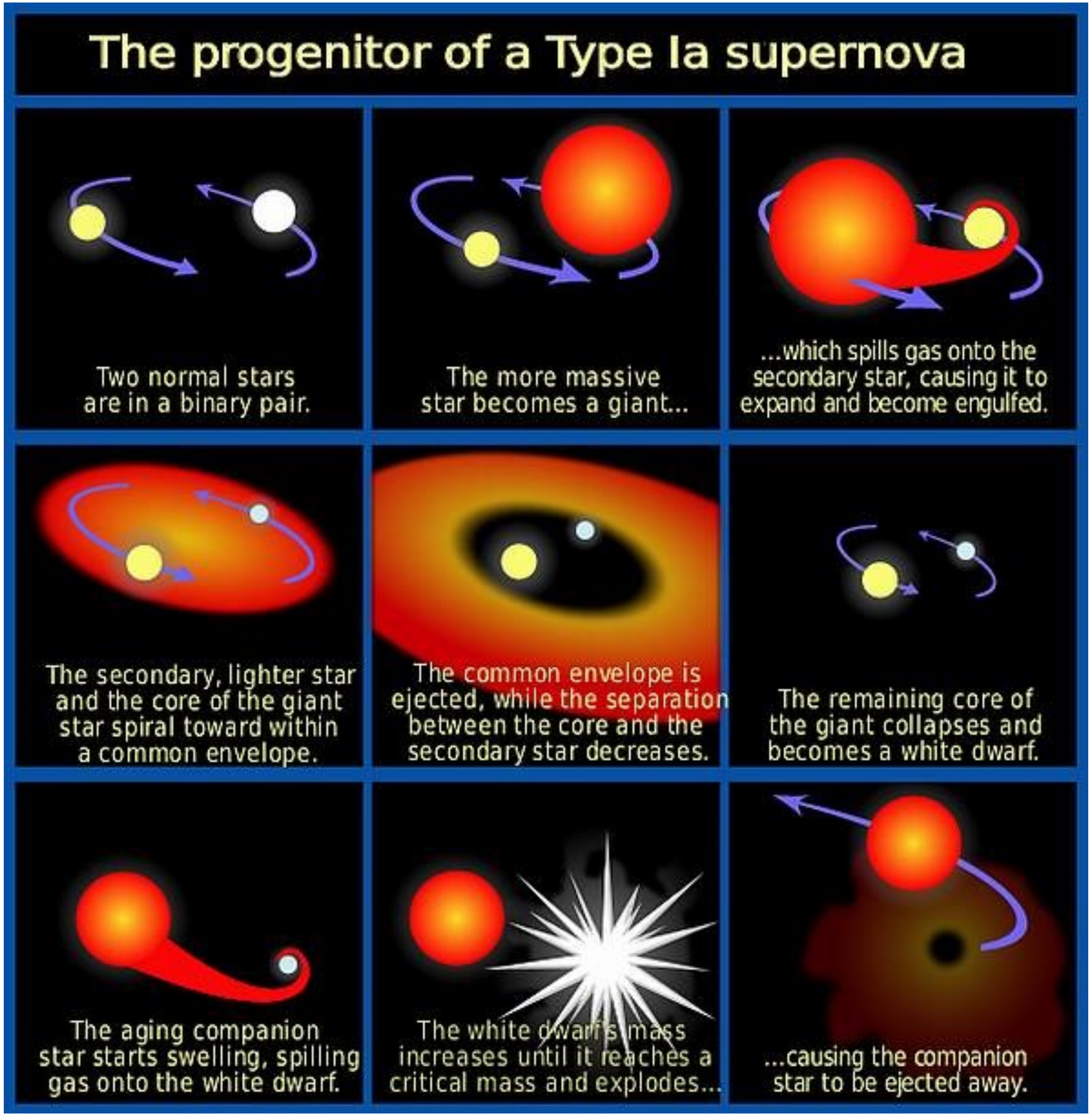
Supernovae of type Ia

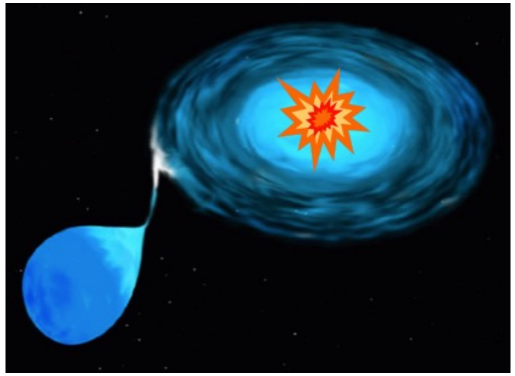
There are different types of supernovae. **Type Ia** is the most luminous one.

Their observations allow us to estimate a cosmological function: the luminosity distance.

This is the most explored probe of geometry (zeroth order)







SNe Ia explosion formed by accretion onto a 1.4 M_{Sun} white dwarf (Chandrasekhar limit : the critical mass for a stable white dwarf)

WD no longer sustained by electron degeneracy pressure.

Becoming unstable, the white dwarf can either collapse further → forming a neutron star or a black hole

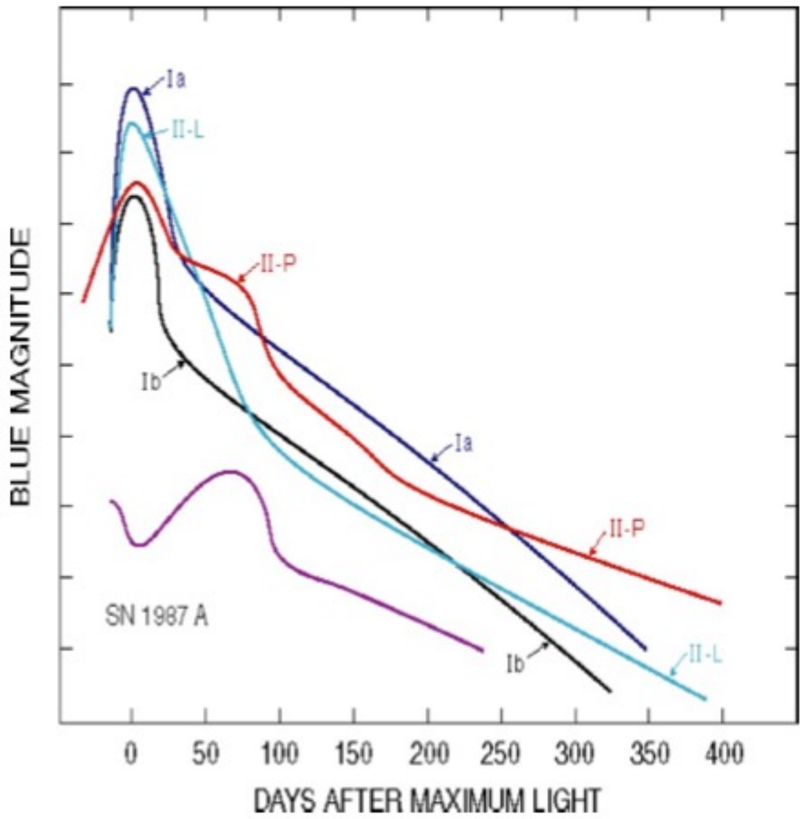
or it can explode → material in the star starts nuclear fusion producing metals, in particular a large amount of ^{56}Ni that decays through ^{56}Co to ^{56}Fe, **producing a very luminous explosion → M ~ -19.**

Other SNe types have different origins and do not require a binary system:

Sne Ib/Ic: collapse of stars with initial masses 8 - 30 M_Sun

Sne II: collapse of stars with initial masses > 30 M_Sun

Each type has a characteristic light curve



The distance modulus

The luminosity distance can be measured through the flux-luminosity relation. This relation is usually written as a difference of magnitudes, known as the **distance modulus.**

Flux is traditionally expressed in magnitudes, m.

Hipparcos separated the magnitude of visible stars in 6 qualitative classes. Since the eye sensitivity is roughly logarithmic → it turns out that apparently equal intervals are in reality equal ratios.

Stars of magnitude 5 have ~ 1% of the flux of stars of magnitude 1. This led to the modern (XIX century) definition of magnitude:

$$\frac{F_1}{F_2} = 100^{(m_2 - m_1)/5} \quad \rightarrow \quad m_1 - m_2 = -2.5 \log_{10}\left(\frac{F_1}{F_2}\right)$$

This is a **relative scale**. To define an absolute scale, the star Vega was chosen as reference → the flux of Vega corresponds to magnitude m=0:

$$m = -2.5 \log_{10} (F / F_{Vega})$$ apparent magnitude

The apparent magnitude depends on the distance, since $F \sim L / D^2$
(Notice that for large distances on an expanding spacetime this distance is the luminosity distance, since it needs to take into account the 'dilution of luminosity').

This leads to the definition of a distance-independent magnitude: the absolute magnitude, M: *the apparent magnitude an object would have if placed at a distance of 10 parsec.*

SNIa have M ~ -19 → they are very bright.

We can compare their magnitudes with the apparent magnitudes of the full moon (m = -12) or of the Sun (m = -26.7) → knowing the distance to the Sun, we find that $L_{SN} = 3 \times 10^9 L_{Sun}$

This implies that a SN Ia would appear as bright as the Sun if it was placed at a distance of D_L =31.6 pc ~ 100 lyr (this is the distance to some of the well-known night-sky stars in the Milky Way).

Luckily for us SN Ia are rare events - only 5 records in our galaxy in the past 1000 years → 1006, 1054, 1181, 1572, 1604

If the SN occurs in a nearby galaxy from the local group (D_L ~15 Mpc) → its apparent magnitude is m = 12 (this was the case of SN1987, a SN II that appeared in the LMC in 1987).

If the SN occurs at a distant galaxy (z = 1) → D_L ~ 6.7 Gpc (concordance model) → 450 x D_L (local galaxy) → its apparent magnitude is m = 25

In both cases, the apparent magnitude of the SN Ia is similar to the apparent magnitude of the whole galaxy

The difference between apparent and absolute magnitudes is a L/F ratio and **it is thus a direct measure of luminosity distance**.
This difference is known as the distance modulus:

$\mu = \mathbf{m - M} = -2.5 \log_{10} (F/F_{Vega}) + 2.5 \log_{10} (F_{10}/F_{Vega})$

$= -2.5 \log_{10} (F/F_{10})$

$= -2.5 \log_{10} [(L / D_L^2) / (L/10^2)]$

$= -5 \log_{10} (10/D_L)$

$= \mathbf{5\log_{10} (D_L) +25}$ (for D_L in Mpc, i.e.,we used 10 pc = 10^{-5} Mpc)

The goal of SN Ia cosmological surveys is to measure μ from the data, and then fit the cosmological predictions of μ (i.e. D_L) to the measured μ.

For this, we need **to define an estimator of μ from the observed quantities. An optimal estimator is one that gets an accurate (i.e, without bias) and precise (i.e. with high signal-to-noise ratio) measurement of μ.**

SNe surveys

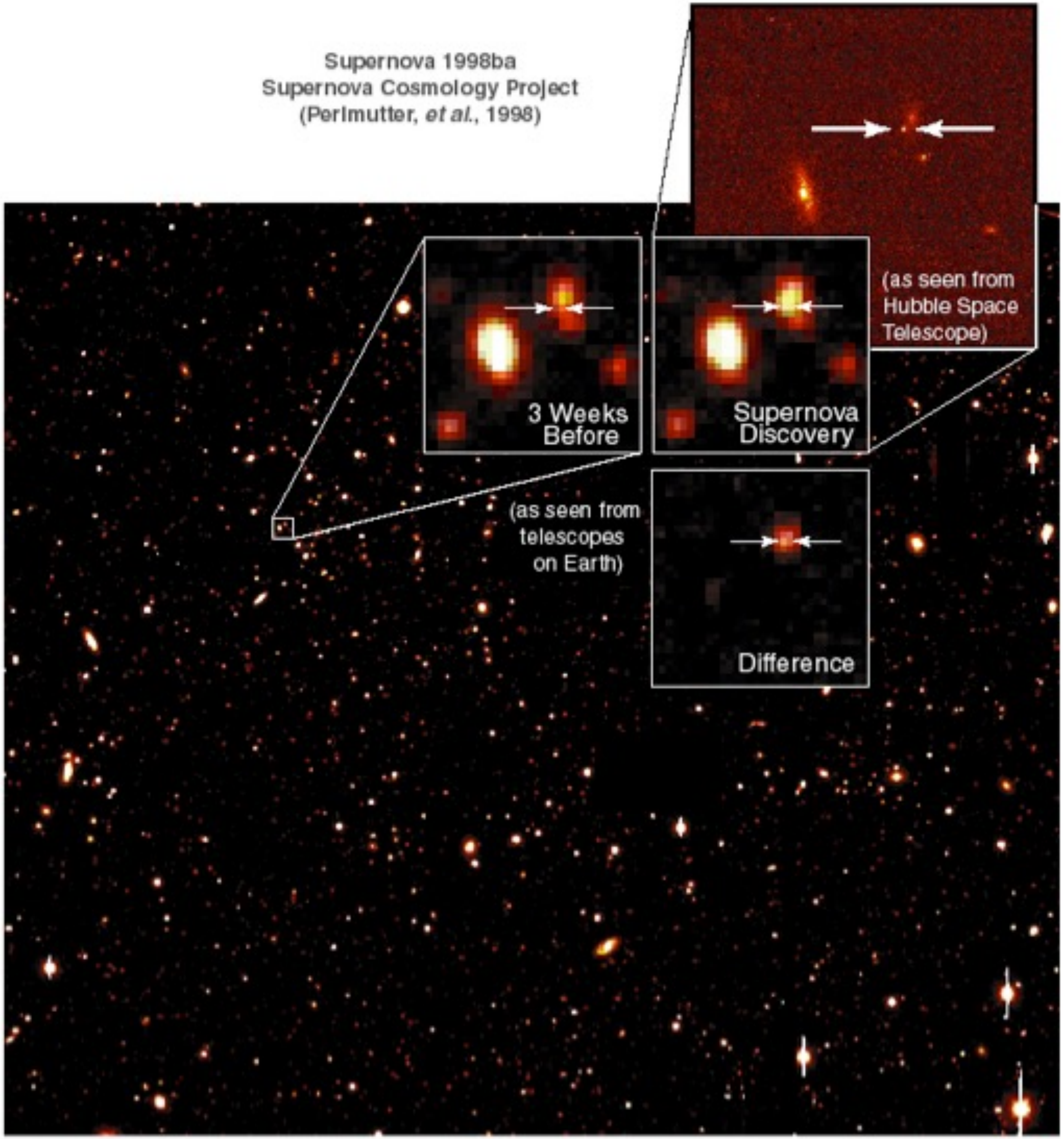
Building a survey to find SN Ia candidates

-- image the same part of the sky repeatedly looking for them
(monitoring required every few days)

-- subtract current image from image earlier to time to look for time variable event
(this can be a challenge with ground-based telescopes since smoothing from turbulent sky conditions may change from night to night)

In nearby galaxies

In distant galaxies

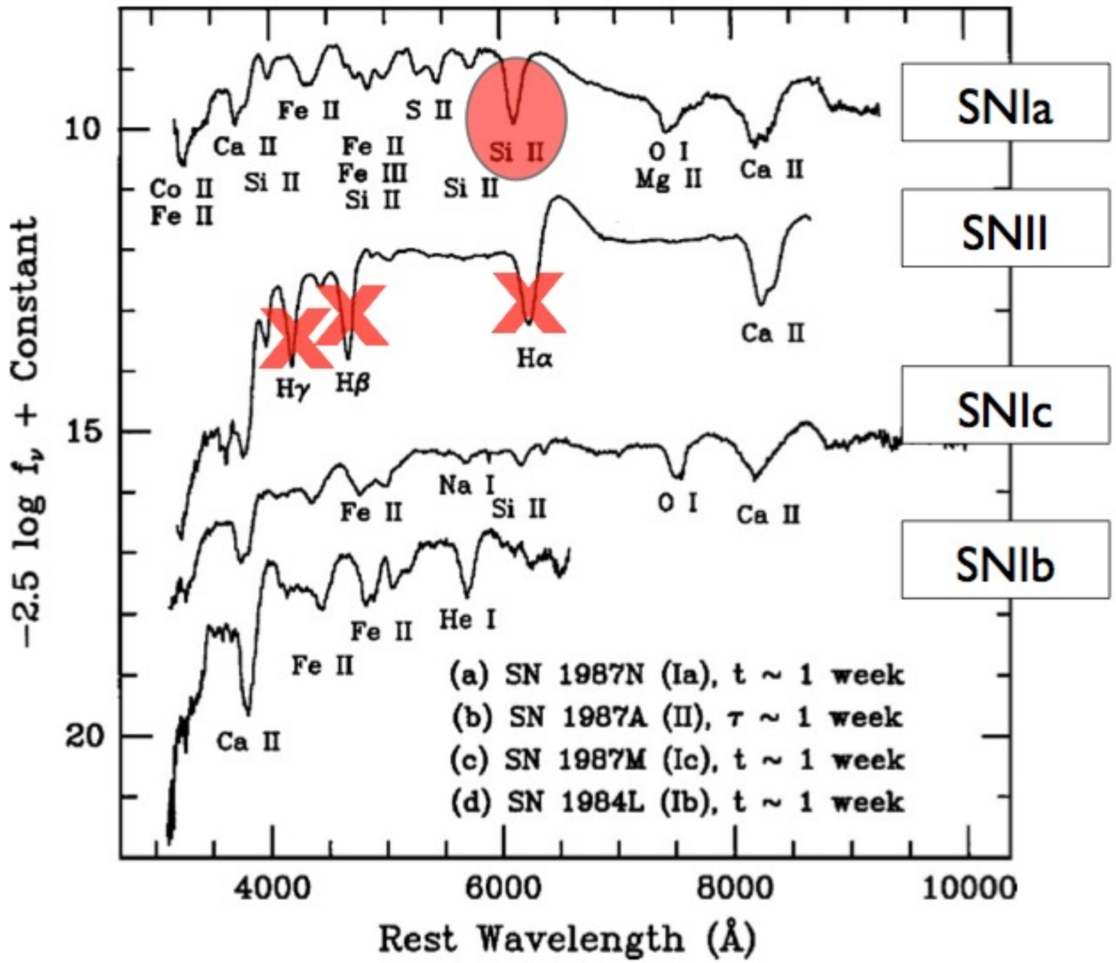


Identifying the SNe of type Ia and measuring their redshifts

- Determine the SN type from its **spectrum**

(this is challenging for distant SNe since they have faint magnitudes ~24)

To distinguish SNe Ia from SNe Ib or Ic, need to check for the presence of Silicon lines (at ~6500 Angstroms rest-frame) → need spectroscopy in near-IR



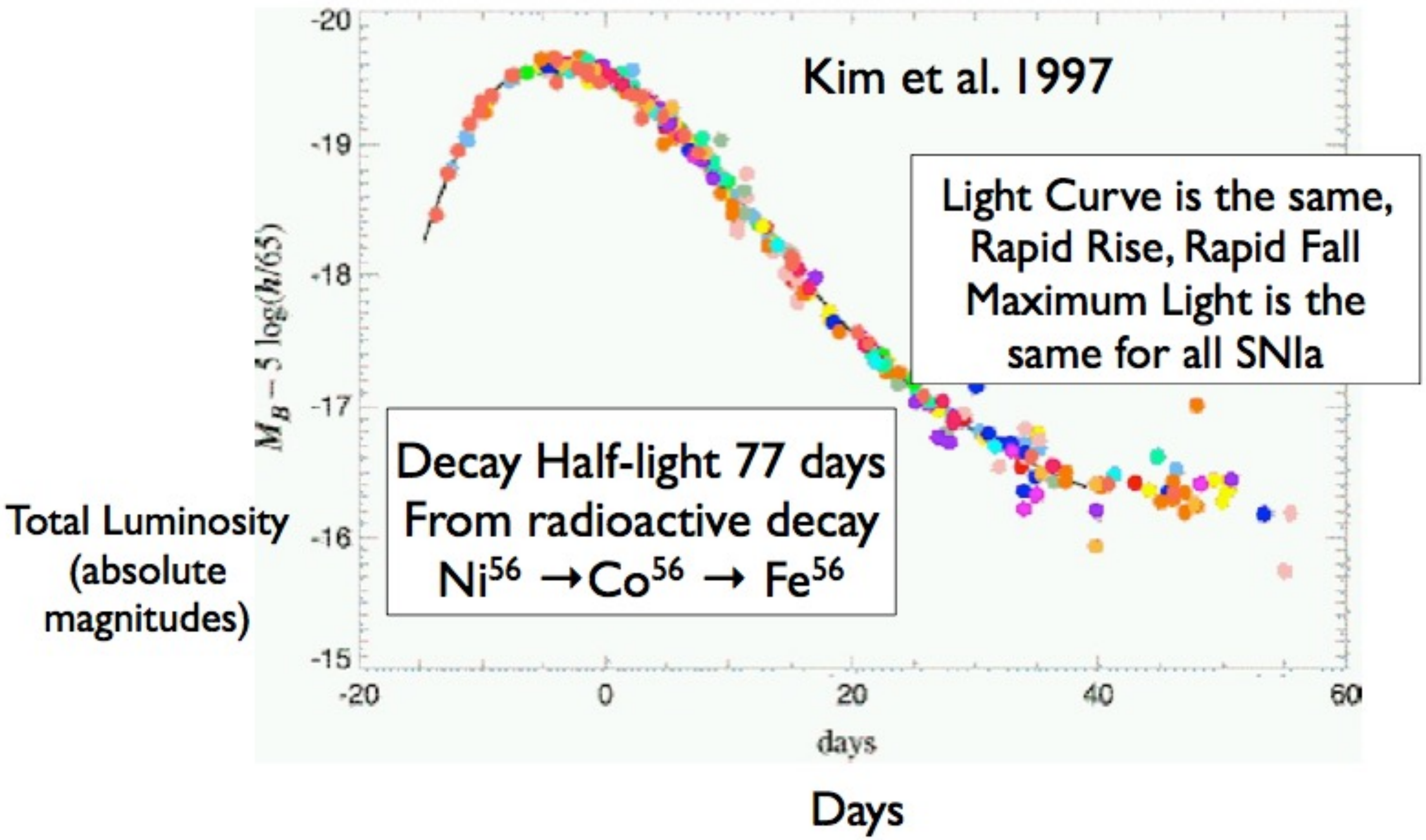
- Use the spectrum also to **measure the redshift**

- Check if the **colors** of the source are consistent with being a SN Ia

- Check if the **light-curve** is consistent with being a SN Ia

The distance modulus estimator

Follow the SN event to get its peak amplitude and also get as much points in the light-curve as possible.



Since all SN Ia are formed in the same way, they are in first approximation assumed to be standard candles, i.e., all SN Ia would have the same light-curve: same absolute magnitude at the peak and same duration in time.

In this assumption, measuring the flux at the peak gives directly the distance, since the absolute magnitude can be determined in advance :

The universal absolute magnitude is computed in advance from observations of SNe in galaxies at known distances (from a **distance ladder** method → it is a **model-independent calibration** procedure).

It relies on:

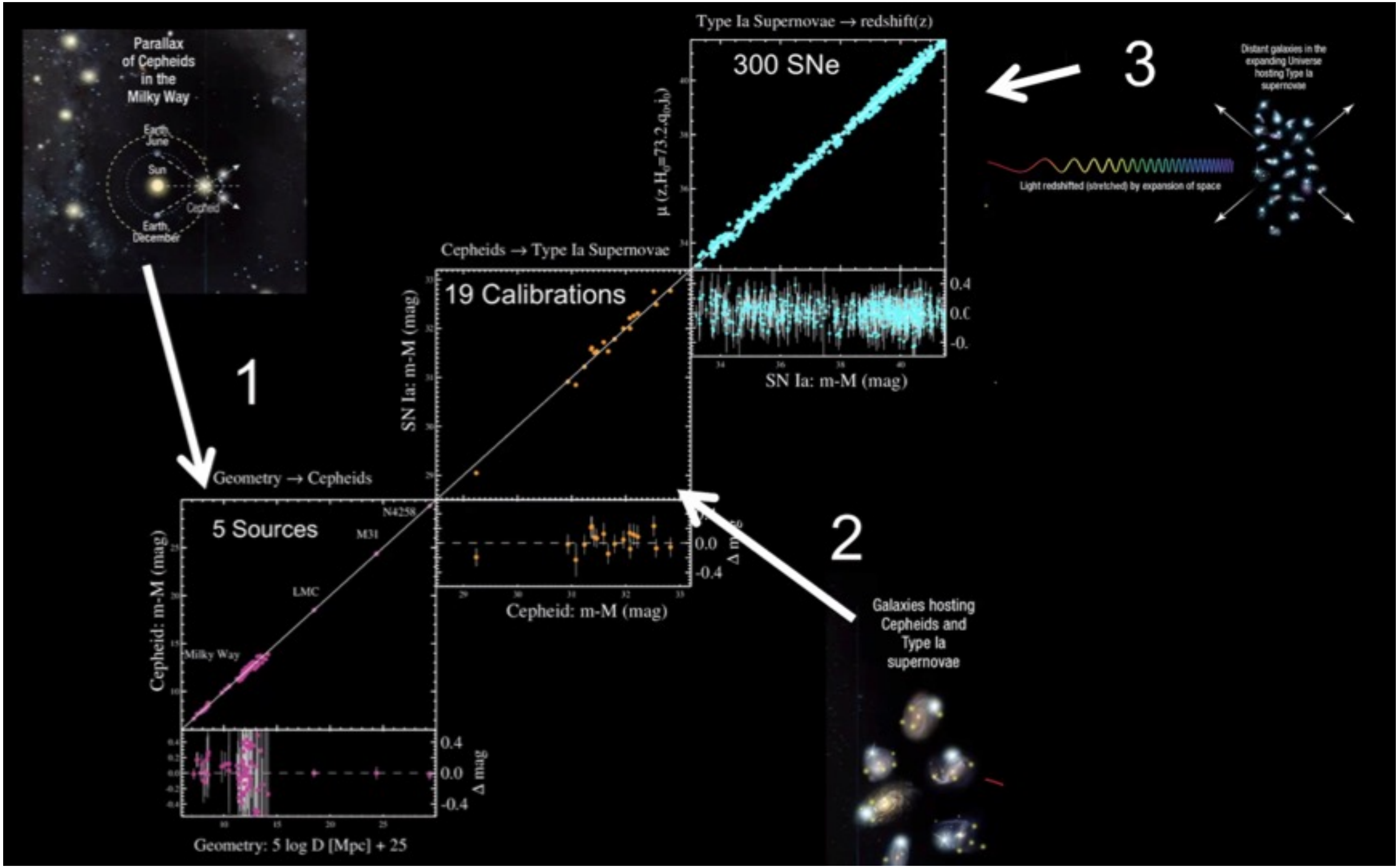
Anchor galaxies: key nearby systems with precise geometric determination of distances: Milky Way (Gaia parallaxes), Large Magellanic Cloud [LMC] (eclipsing binaries), SMC, NGC 4258 (water maser – a microwave laser -).

Anchor galaxies need to have: the possibility to measure their distances; and contain a **distance calibrator** (e.g. a Cepheid)

Host galaxies: galaxies that have simultaneously a SN and a calibrator

There are only 42 such SNe known (found with the SHoES project, Riess et al 2022), increased from a previous sample of 19 → these calibrate the luminosity of all SNe Ia observed.

The distance ladder accuracy is a key aspect to solve the **Hubble tension**.



Parallax of Cepheids in the Milky Way
Earth, June
Sun
Cepheid
Earth, December
1
Geometry → Cepheids
5 Sources
N4258
M31
LMC
Milky Way
Cepheid: m-M (mag)
Geometry: 5 log D [Mpc] + 25
Δ mag
Cepheids → Type Ia Supernovae
19 Calibrations
SN Ia: m-M (mag)
Cepheid: m-M (mag)
2
Galaxies hosting Cepheids and Type Ia supernovae
Type Ia Supernovae → redshift(z)
300 SNe
SN Ia: m-M (mag)
3
Distant galaxies in the expanding Universe hosting Type Ia supernovae
Light redshifted (stretched) by expansion of space

For the **host galaxies**, measuring the SN peak flux and knowing the distance → obtain the universal absolute magnitude (that would remain a fixed quantity for all subsequent observations of SN in galaxies with unknown distances, if M is universal).

→ with this information, the distance modulus of a given galaxy can be accurately computed from its definition: $\mathbf{\mu_i = m_i - M}$

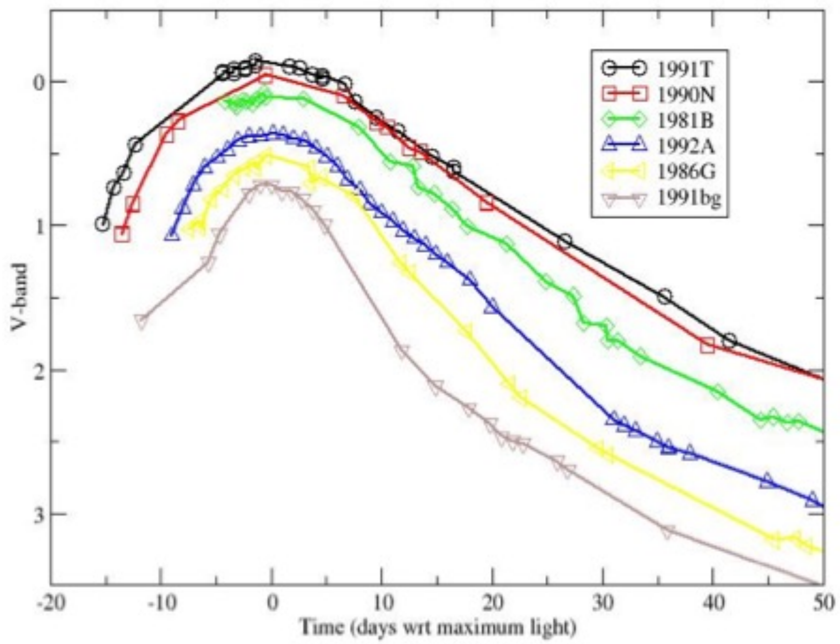
where the **magnitude m** is the measurement made from the flux of the galaxy obtained in this survey (with a certain uncertainty) and the **magnitude M** is a constant value known a priori from a distance ladder calibration previously made (also with an uncertainty).

So this expression $\hat{\mu}_i = m_i - M$ seems a good choice to be the estimator of the distance modulus of a galaxy from the data.

Building an unbiased estimator

In general, there are additional astrophysical effects that also contribute to the cosmological estimator. For example, there may be other effects contributing to the magnitude such that m-M is not just the cosmological contribution.

In SN surveys, an important effect to consider is the fact that **in reality the SNe are not standard candles,** i.e., the approximation of universal luminosity is not very accurate**.**



Each SN has its own light curve.

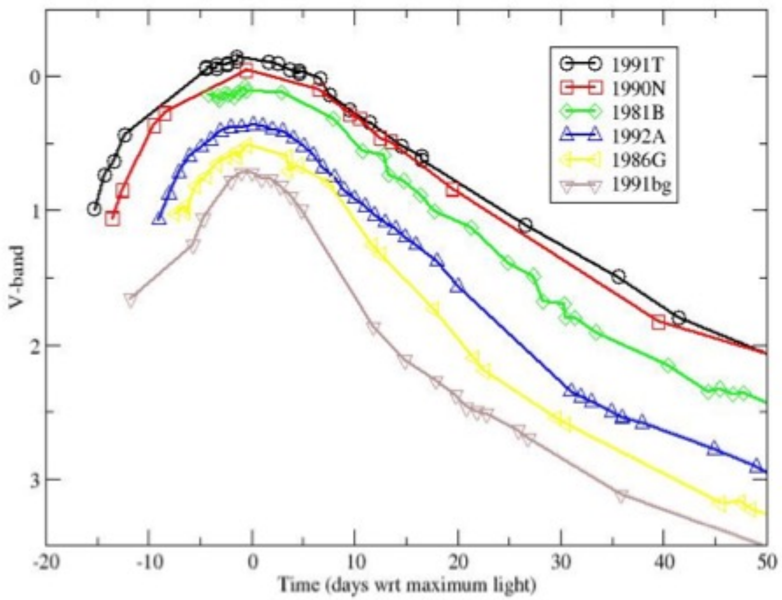
They are not universal after all → the value of M obtained for the calibrator SN is not valid for all → each SN may have its own value of M.

In this way, if we apply the estimator **$\hat{\mu}_i = m_i - M$,** we will get a wrong value for the distance modulus of the galaxy - a **biased** value, i.e., there will be a **systematic error** on all estimated distance modulus.

A more accurate (**unbiased**) estimator would be:

$$\hat{\mu}_i = m_i - M_i$$

Does this mean that we need to find the absolute magnitude of each SN? This is not possible. We only have very few SNe where M can be measured (the calibrators).



Fortunately, the SNe are not completely different. There is a correlation in the behaviour of their light-curves:

There seems to be a **shape- luminosity** relation in the light-curves → **the luminosities are indeed different but the peak amplitude depends on the decay time** → the brighter ones are systematically slower.

The existence of this (empirical) relation means that

the SNe are standardizable → their M values can be related with the standard “universal” M value.

So, if we apply a stretch factor to a light-curve, its peak will go up and reach the “universal light-curve” → the one that corresponds to the (**standard**) universal luminosity.

The question now is: if we would stretch the observed light-curve of a SN by a certain factor, this would correspond to change the luminosity by how much?

We do not know this! But the important point is that the impact of the stretching of the light-curve on the luminosity is the same for all SNe.

So we can **model** this effect with an arbitrary function of an arbitrary amplitude, and apply the model consistently to all SNe.

The **response of magnitude to stretch** is usually taken to be **linear**, i.e.,

$$\Delta M_i = M - M_i = \alpha (s_i - 1)$$

So, a galaxy that requires a stretch s for its light-curve to become identical to the standard one, has an absolute magnitude that differs ΔM_i from the universal value.
(s = 1 means no stretch, it is a SN already standard)

This means that the estimator of the distance modulus of a galaxy is not:

$\hat{\mu}_i = m_i - M$ → this gives a biased result;

it is also not:

$\hat{\mu}_i = m_i - M_i$ → this gives an unbiased result, but it is impossible to measure, so it cannot be an estimator;

but it is:

$\hat{\mu}_i = m_i - M + \alpha (s_i - 1)$ → this gives an unbiased result.

Notice that this method introduces one unknown parameter in the analysis:

the **stretch response parameter α**

Its value is unknown. How can we find out its value?

option 1:

If the luminosities (i.e., the absolute magnitudes) of some of the stretched SNe were known, that information could be used to calibrate the relation (i.e., to find the value of α).

However, this is not the case.

option 2:

An alternative would be to predict the absolute magnitudes of SNe from **astrophysical theory**.

However this cannot be done with enough precision, and it depends on many assumptions and astrophysical modelling (which would introduce additional astrophysical parameters, and would just move the problem to another place).

option 3:

The usual approach is to leave α as an additional free parameter of the model (introduced to model an extra effect), to be treated in the same way as the cosmological parameters.

This type of parameter is known as a nuisance parameter.

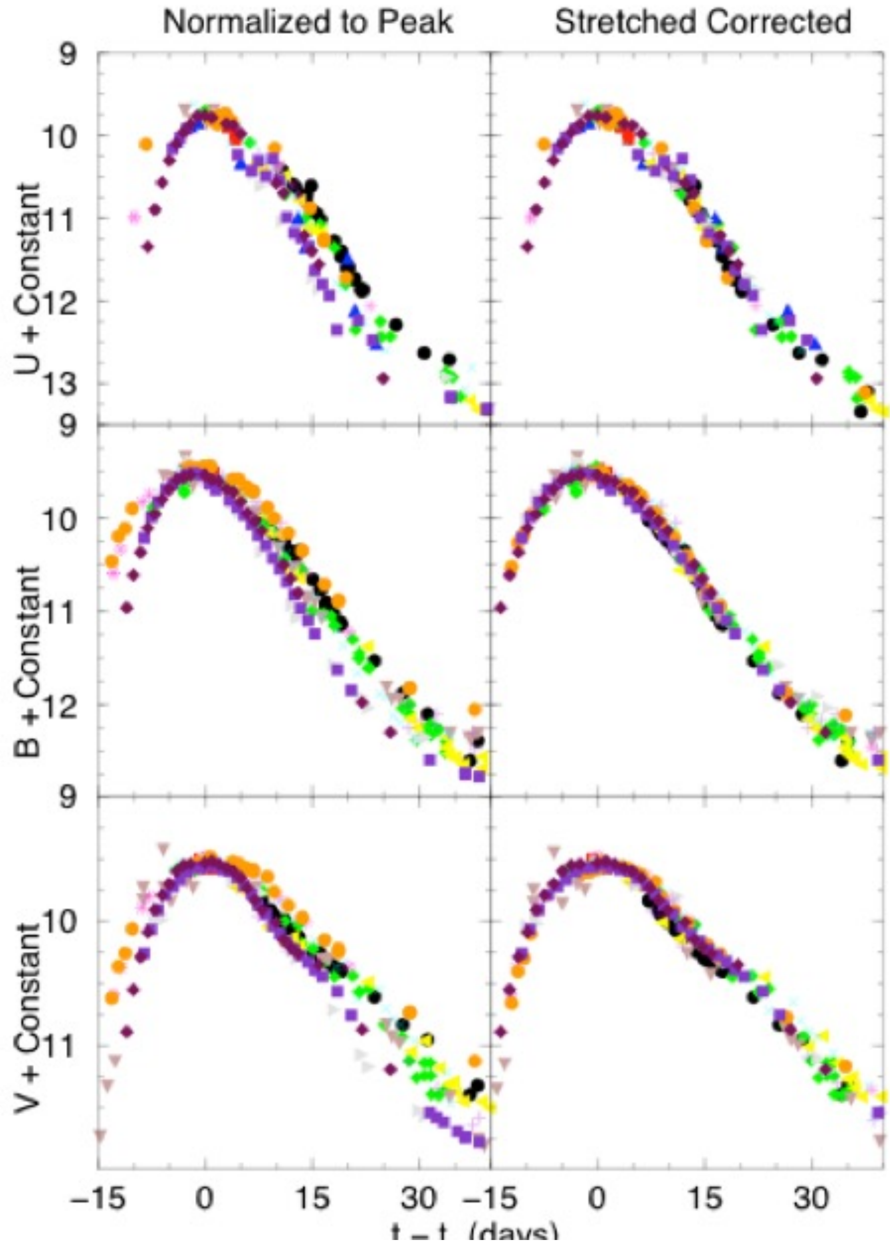
Despite the name, nuisance parameters are very important → without them the cosmological analysis would be biased → they model (and correct) a systematic effect.

In addition to the shape-luminosity relation, the light-curves also show a **color-luminosity relation,**

i.e., two SNe of different colours, if stretched by the same amount will not reach the same peak amplitude.

The bluer ones (the ones with higher amplitudes on the blue filters compared to redder filters) have larger luminosities.

So, for each measured SN, after stretching the light curve by a factor s, the amplitude needs to be further increased for it to match the standard light curve.



Moreover, by comparing the fluxes of one SN on different filters, sometimes it is found that the flux ratios are too different from the standard case to be explained just by this intrinsic color variation → differences also arise because of dust extinction in the galaxy host, which does not affect equally all bands.

So the amplitude of the light-curve is increased by a color factor, c, (that accounts both for color and for dust) in order to match the standard amplitude.

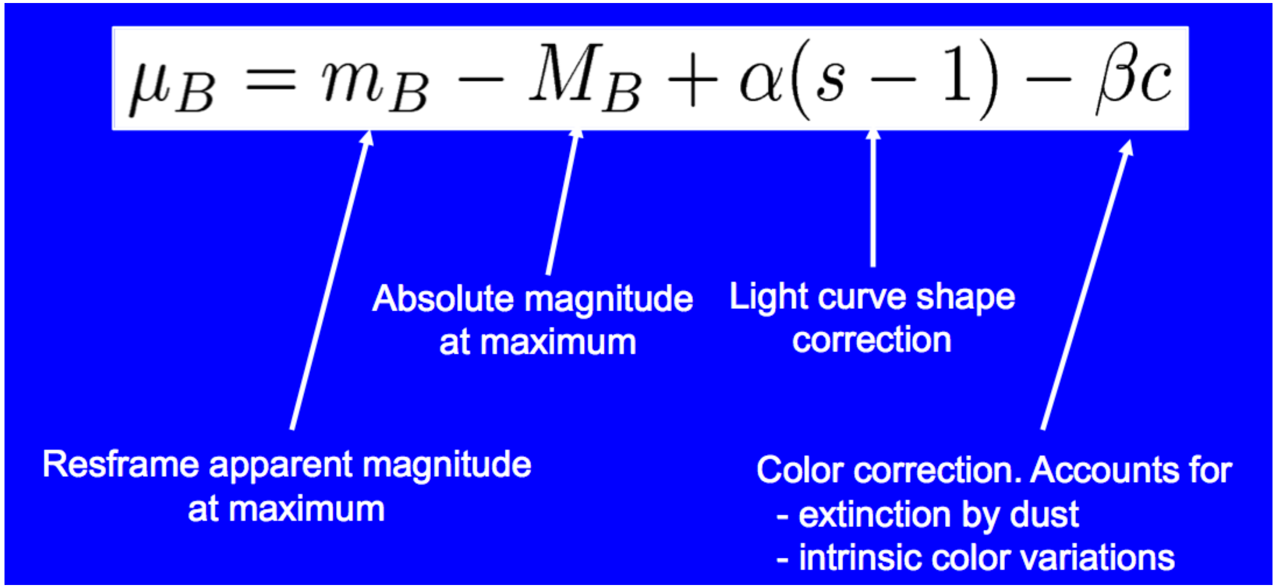
Once again, we need to find out what is the impact of this shift in the absolute magnitude.

This introduces another nuisance parameter : **the response factor β, that models a linear response:**

$$\Delta M_i = \beta c_i$$

The **estimator** of the distance modulus then becomes:

$$\hat{\mu}_i = m_i - M + \alpha\,(s_i - 1) - \beta\, c_i$$



The estimator involves 3 measured quantities for each SN:

m_i : observed magnitude of each SN → i.e. the measured flux

s_i, c_i : stretch and color factors of each SN → measured from the light curves

Notice that the absolute magnitude of a galaxy with no stretch (s=1) and no color correction (c=0) is the reference value M, while the absolute magnitude of a galaxy with a corrected light curve is $M - \alpha (s_i - 1) + \beta c_i$

The estimator also involves 3 (global) model parameters:

M : reference absolute magnitude, known from the calibrators with an uncertainty → it may treated as a free parameter with a prior

α : response of magnitude to stretch → a free parameter

β: response of magnitude to color and dust → a free parameter

The nuisance parameters are not necessarily universal across the whole sample of SN Ia. They may be different for SNe at different redshifts - evolution - or for SNe in different host galaxies - environment -

In this case, the analysis needs to be done with the SN separated in various sub-samples, with different parameter values in each.

K-correction

The observational procedures in general introduce additional biases. In high-z measurements, **the use of an observing filter bias the flux-luminosity relation.**

Remember that the flux-luminosity relation, relates observed flux (in the observer's frame o) with the corresponding luminosity. But that luminosity is not the equal to the intrinsic luminosity (in the source rest-frame e), due to the expansion of the Universe (causing energy dilution and time dilation).

$$E_o = \frac{E_e}{1+z} \qquad \frac{\Delta t_o}{a(t_o)} = \frac{\Delta t_e}{a(t_e)}$$

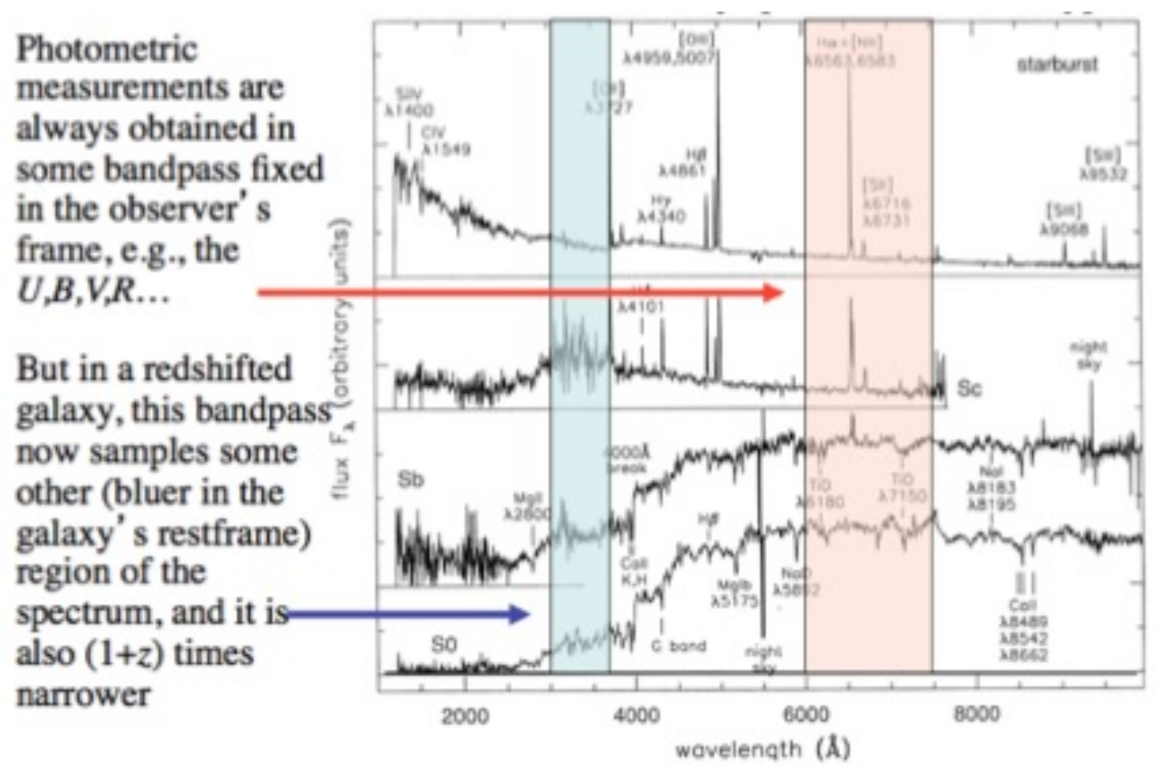
This led to the additional factor of $(1+z)^2$ that was absorbed in the distance, leading to the definition of luminosity distance.

$$L_o = \frac{L}{(1+z)^2}$$

However, this reasoning is only valid when we consider the total luminosity emitted by the source (and the total measured flux), the so-called bolometric quantities.

In practice, the fluxes are measured within a filter → **only a part of the energy is detected** by the observer, i.e., an integral over the field spectral domain of a **flux density** per frequency bin ($F(\nu_0)\, d\nu_0$).

This flux corresponds to an emission of a different frequency in the galaxy's rest-frame ($L(\nu_e)\, d\nu_e$) due to the expansion redshift.



The flux on a detected frequency corresponds to an emission from a higher frequency

$$\nu_e = [1 + z]\, \nu_o$$

The filter broadens at the observer's frame → the "unit" $d\nu_0$ shrinks

$$d\nu_e = (1+z)\, d\nu_0$$

The detected flux is an integration of the flux density and thus of the spectral energy distribution (SED) on a portion of the spectrum (i.e. within a filter, that provides a weighting function: the filter throughput),

This implies that the luminosity distance needs to be redefined, i.e., the $(1+z)^2$ factor valid for the bolometric case needs to be removed and replaced by a frequency-dependent function.

In practice, what is done is to keep the same luminosity distance and apply instead a correction to the flux-luminosity relation, i.e. to **unbias the luminosity distance**. This effectively introduces an extra term in the definition of the distance modulus, named the **K-correction**.

The K-correction is thus the difference between the observed magnitude for a source at redshift z *(for a specific filter and a specific source SED)* **and the magnitude that would be observed if there was no expansion.**

How can we compute the K-correction for a given SED and filter?

Let us consider the flux measured in a filter centered in v_0, per unit frequency dv_0 (i.e., the **flux density**)

It was emitted in the rest-frame of the source as frequencies centered on a redshift v_e , per unit frequency dv_e.

The standard bolometric flux-luminosity relation (already including the $(1+z)^2$ factor in the definition of luminosity distance) is:

$$F(\nu_0)\, d\nu_0 = \frac{L(\nu_e)\, d\nu_e}{4\pi D_L^2}$$

We now want to write the rest-frame luminosity in terms of the observed-frame luminosity. **For this, we need to introduce the two effects: frequency shift and filter broadening:**

$$F(\nu_0)\, d\nu_0 = \frac{L\left(\nu_0 (1+z)\right) (1+z)\, d\nu_0}{4\pi D_L^2}$$

The flux observed is the integral of this **flux density**, within the filter (or band).

$$\int_{band} F(\nu_o)\, d\nu_o$$

Using the flux-luminosity relation this is:

$$\int_{band} F(\nu_o)\, d\nu_o = \frac{\int L(\nu_o(1+z))\,(1+z)\, d\nu_o}{4\pi D_L^2}$$

If the luminosity density (the SED) is constant within the filter (e.g. in the case of a narrow filter), then the numerator just contains a $(1+z)^2$ factor that cancels out the one implicit in D_L^2 and we recover the standard relations

$$F_o = \frac{L_o}{4\pi D_C^2} \qquad F_o = \frac{L_e}{4\pi D_L^2}$$

In the more relevant case of a broad filter, we can multiply and divide the expression for the observed flux by the integral of L_o and write:

$$\int_{band} F(\nu_0)\,d\nu_0 = \frac{\int L(\nu_0(1+z))\,(1+z)\,d\nu_0}{4\pi D_L^2}$$

$$= \frac{\int^{band} L(\nu_0)\,d\nu_0}{4\pi D_L^2} \; \frac{\int^{band} L(\nu_0(1+z))\,(1+z)\,d\nu_0}{\int_{band} L(\nu_0)\,d\nu_0}$$

↓ the previous result — ↘ the correction

The result is then: the correction due to the use of a filter is a factor (1+z), times the ratio f_ν between the integrated luminosities on the emission and observed bands, where

$$f_\nu = \frac{\int_{\nu_1}^{\nu_2} L[\nu(1+z)]d\nu}{\int_{\nu_1}^{\nu_2} L(\nu)d\nu}$$

This means that when we only measure part of the flux (using a filter) **the correct flux-luminosity relation is no longer**

$F = L / (4\pi D_L^2)$ (the bolometric relation)

but

$$F = \frac{L}{4\pi D_L^2} f_\nu (1+z)$$

We can also write the flux-magnitude relation in terms of magnitude difference:

We start with the definition: $$m - M = -2.5 \log_{10}\left(\frac{F}{F_*}\right) + 2.5 \log_{10}\left(\frac{F_{10}}{F_*}\right)$$

where now $$F = \frac{L}{4\pi D_L^2} f_\nu (1+z)$$

For D = 10pc there is no correction (very low redshift) → z=0 and $f_v = 1$

$$\Rightarrow \quad m - M = -2.5 \log_{10}\left(\frac{k f_\nu(1+z)}{4\pi D_L^2}\frac{(10^{-5}\,\mathrm{Mpc})^2 4\pi}{k}\right) = 2.5 \log_{10}\left(\frac{D_L^2 (10^5)^2}{(\sqrt{f_\nu})^2 (1+z)}\right)$$

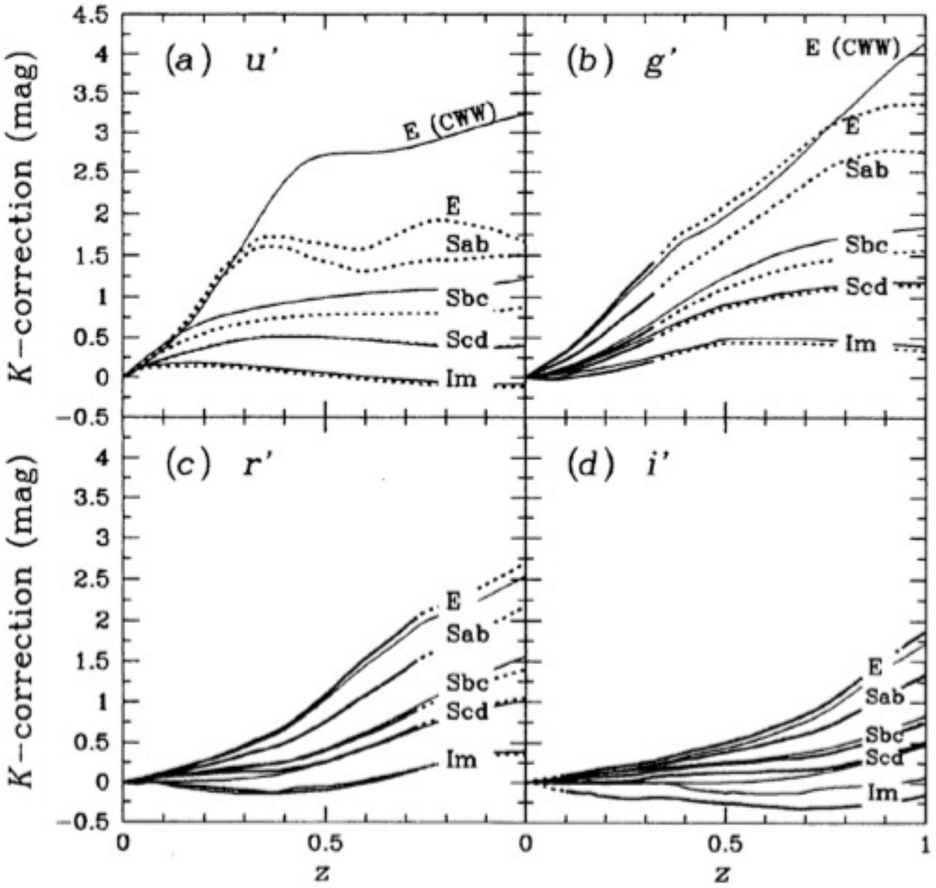
$$\Rightarrow \quad m - M = 5 \log_{10} D_L + 25 + 5 \log_{10}\left(\frac{1}{\sqrt{f}\sqrt{1+z}}\right)$$

$$\Rightarrow \quad m - M = 5 \log_{10}(D_L) + 25 - 2.5 \log_{10}(f_\nu (1+z))$$

$$\boxed{m - M = 5\log_{10}(D_L) + 25 - 2.5\log_{10}(f_\nu) - 2.5\log_{10}(1+z)}$$

The correction in the distance modulus is thus a constant shift in magnitude:

$$K_\nu(z) = 2.5 \log_{10}(f_\nu (1+z))$$



K-correction (in magnitude) for various filters and types of objects (spectra) as function of redshift.

The correction can be large, but it does not introduce modeling of an effect with a new free parameter.

It can be computed and directly applied. It depends on the observing filter, the redshift of the source, and the shape of the spectrum (SED).

We conclude that a **better estimator of the distance modulus is:**

$$\hat{\mu} = m - M + \alpha (s - 1) - \beta c - K_v(z)$$

Other systematics

Besides stretch, color and K-correction, there are many other possible sources of bias (also known as systematic effects) that impact the SN measurements:

- Peculiar velocities for low-z SN → the measured redshift are not only due to expansion → need to be corrected

- Contamination by core collapse SN for high-z SN → SN of other types mixed in the analysis

- Evolution of color-luminosity relation with redshift → the modeling of this bias should be done with $\beta(z)\,c$, and so β is no longer a single parameter

- Evolution of SNe with z → light-curves should not be matched to a single universal template

- Gravitational magnification → lensing effect changing the flux of the SNe

- Malmquist bias → a SN sample is biased towards the brightest objects, needs to correct the distance modulus by adding a magnitude of 1.38 σ^2_M

Assume we have a **flux-limited sample** of SNe and they have a distribution of absolute magnitudes $M_0 \pm \sigma_M$.

Then for high-z SNe the ones at the tail of the distribution may be outside of the limit → we observe a **biased sample, not representative of the full distribution → we lose systematically the faint objects, the sample is incomplete but not in a random way** → introducing a bias

We will think that those high-z SNe are brighter than they really are → we need to correct by **adding a magnitude value that increases with the width of the distribution** (the correction is computed from first principles, by integrating the SNe 'magnitude function').

(Note that if we lose objects in a random way - for example because of not observing the full sky - then there is no bias)

- Dependence on mass of the host galaxy → SNe appear systematically brighter when they are in massive galaxies by ~0.1 mag

Two possible simplest ways to proceed:

1) Add a further linear host term, H, to the analysis:

$$m_B = m_B - M_B + a(s-1) - bc + gH$$

– *Requires very precise measure of H, and robust errors*

2) Use two M_B – one for high-mass galaxies and one for low-mass

$$m_B = m_B - M_B^1 + a(s-1) - bc \quad \text{when } H < H_{\text{split}}$$

$$m_B = m_B - M_B^2 + a(s-1) - bc \quad \text{when } H > H_{\text{split}}$$

There are about 200 systematic effects identified in SNe analyses!

The dominant source of bias is the calibration of the universal absolute magnitude M (which depends on the distance ladder determinations and on the light-curves template-fitting).

Description	Ω_m	w	Rel. Area [a]
Stat only	$0.19^{+0.08}_{-0.10}$	$-0.90^{+0.16}_{-0.20}$	1
All systematics	0.18 ± 0.10	$-0.91^{+0.17}_{-0.24}$	1.85
Calibration	$0.191^{+0.095}_{-0.104}$	$-0.92^{+0.17}_{-0.23}$	1.79
SN model	$0.195^{+0.086}_{-0.101}$	$-0.90^{+0.16}_{-0.20}$	1.02
Peculiar velocities	$0.197^{+0.084}_{-0.100}$	$-0.91^{+0.16}_{-0.20}$	1.03
Malmquist bias	$0.198^{+0.084}_{-0.100}$	$-0.91^{+0.16}_{-0.20}$	1.07
non-Ia contamination	$0.19^{+0.08}_{-0.10}$	$-0.90^{+0.16}_{-0.20}$	1
MW extinction correction	$0.196^{+0.084}_{-0.100}$	$-0.90^{+0.16}_{-0.20}$	1.05
SN evolution	$0.185^{+0.088}_{-0.099}$	$-0.88^{+0.15}_{-0.20}$	1.02
Host relation	$0.198^{+0.085}_{-0.102}$	$-0.91^{+0.16}_{-0.21}$	1.08

We finally found an estimator that should be better than the naïve one $\hat{\mu}_i = m_i - M$

This estimator is:

$$\hat{\mu}_i = m_i - M + \alpha (s_i - 1) - \beta c_i + \gamma H_i - K_v(z) + \text{other biases}$$

It takes into account the fact that

measured m = true m + $K_v(z)$

and

universal M = true M + $\alpha (s - 1) - \beta c + \gamma H$ + other biases

This estimator should give the true value of the distance modulus, i.e., if we average the measurements of N SNe Ia (at the same redshift), we should get the true μ :

(averaging over large N) → $\langle\hat{\mu}\rangle$ = true μ

an estimator with this property is called an **unbiased estimator.**

On the contrary, for the original estimator ^μ^= m - M,

(averaging over large N) → <^μ^> ≠ true μ

In that case the estimator is called a **biased estimator.**

Notice that the reason for averaging over many observations, is because the estimator has an uncertainty (error bars).

An **optimal** estimator should be **unbiased** (i.e., an estimator that provides **accurate** measurements)

and at the same time it should be measured with a high **signal-to-noise ratio** (i.e. an estimator that provides **precise** measurements).

The estimator uncertainty

The values of the unbiased estimator averaged over the sample give the estimate for the quantity of interest (in our case $\mu(z)$).

But this is not enough to fully describe the measurement. **We also need to quantify the uncertainty of the estimator.**

This can be addressed in two different ways:

Measuring the variance of the sample

Consider a set of N_z SN Ia at the same redshift

All of these SNe should have the same value of μ_z.

However, the measured μ_{zi} for each SN will not be the same because the measurement process (including emission and correction factors) is a **random process**.

So, if we measure N_z SNe at the same redshift, they will constitute **N_z independent measurements of the same quantity**, where ^μ_z^ is a **random variable** and each of the measured values μ_{zi}, is a realization of the probability distribution of the random variable ^μ_z^.

If the distribution is Gaussian (which is always the case if N_z is large due to the **Central Limit Theorem**), the distribution is described by only two parameters (the lower order moments of a distribution). These are the mean, α, (i.e. the true value of μ), and the variance, σ^2

The larger the variance of a distribution, the most likely to observe a value μ_{zi} far from the true value α.

The σ of the distribution (i.e., the square root of the variance σ^2, also called the dispersion or the root-mean-square rms) is the error bar of the single observation μ_{zi}

This is the average error of each μ_{zi} measurement, but we are interested in the error of the $\hat{\mu}_z$ estimate.

$\hat{\mu}_z$ is estimated from the individual measurements μ_{zi}. It is known that the maximum likelihood estimator of the mean of a Gaussian distribution is **the average of the realizations**:

$$\hat{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\mu_i$$

↓ all at the same z

So, the uncertainty we are looking for is the error on the estimation of the mean.

The **variance of an estimator of a mean** is a well known result, and is given by:

$$\sigma_{\alpha}^{2} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\mu_i - \hat{\alpha})^{2}$$

Basically this is the variance of the random variable μ_{zi} divided by N_z

$$\sigma_{\alpha}^{2} = \frac{\sigma^{2}}{N}$$

This makes sense, since the estimated value of the mean will be closer to the true mean if we have a better sampling of the distribution (i.e., larger N_z).

The square root of this variance is the **uncertainty** associated with the estimator ^μ_z^ → **the error bar**.

This error is usually called the **statistical error** or also the **noise**.

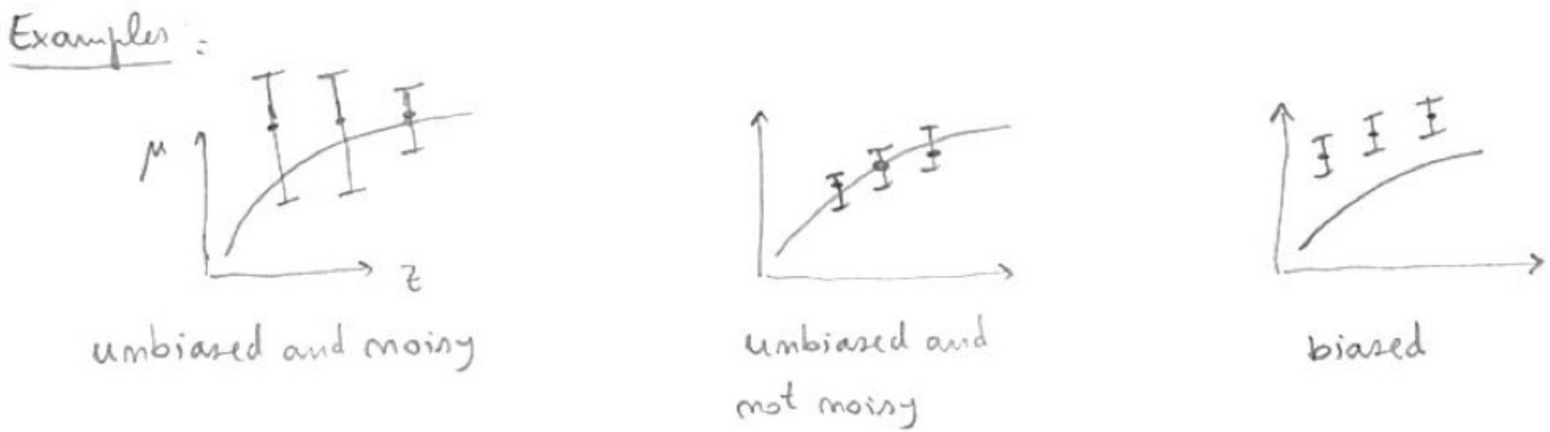
The uncertainty introduced by the bias corrections, is in general not included in the variance of the random variable and needs to be taken into account separately as an extra contribution to the total error: the **systematic error**.

We just saw that **the uncertainty of ^μ_z^ decreases with the square root of the number of SNe observed at that z.**

If the variance of the random variable is large, then many measurements are needed (a large N_z) in order to obtain a small error on the mean, i.e, for the estimated value ^μ_z^ to be close to the true value μ_z.

In this case, we say that the estimator is **noisy** → it has a low **signal-to-noise ratio (S/N)**.

Note that a biased estimator is not necessarily noisy. On the contrary it can have a low statistical error if its measurements have little dispersion (but around a wrong value, since it is biased). The fact that the value is wrong means there is a large systematic error (it is biased).



→ it is possible to have a high precision measurement with low accuracy.

Consider now a set of N SNe Ia at various redshifts

For each one of the redshift bins (let us assume there are Nb bins), there is a different random variable μ_{zb}.

We measure N_{zb} SN for each of the bins zb, obtaining the various measurements μ_{zbi}.

We have then a vector of random variables, that is described by a multi-dimensional Gaussian (of dimension N_b).

The mean of a multi-dimensional Gaussian is a vector, that we want to estimate: ^(μ_{z1}, μ_{z2}, ... μ_{znb})^

The variance of a multi-dimensional Gaussian is a matrix, that consists of the variances of each random variable (which are the diagonal terms of the matrix), and the correlations between the various random variables (which are the off-diagonal terms of the matrix, also called the covariances) → this defines the **covariance matrix**.

The covariance matrix of a mean vector is computed like in the case of a single random variable, but considering all correlations:

$$\sigma^2_{z_i z_j} = \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \left(\mu_k(z_i) - \langle \mu(z_i) \rangle_k \right) \left(\mu_l(z_j) - \langle \mu(z_j) \rangle_l \right)$$

This defines a Nb x Nb covariance matrix:

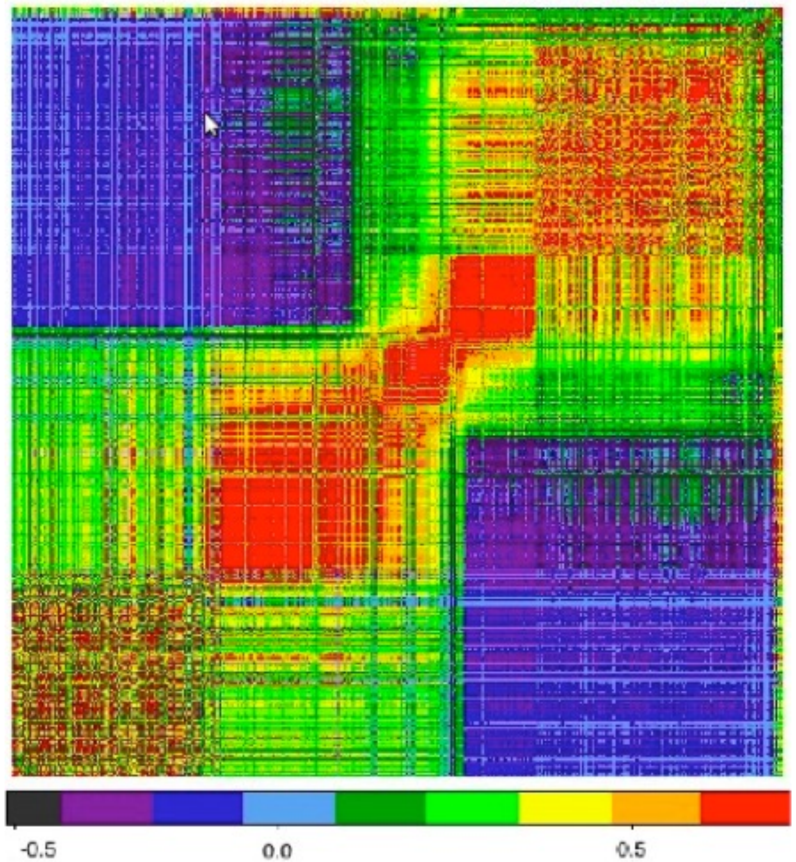
$$\sigma^2_{\mu_{zz'}} = \begin{bmatrix} \sigma^2_{z_1 z_1} & \sigma^2_{z_1 z_2} & \sigma^2_{z_1 z_3} & \cdots \\ \shortparallel & \sigma^2_{z_2 z_2} & \sigma^2_{z_3 z_3} & \\ & & & \end{bmatrix}$$

The error bars for the measured function ^μ(z)^ are defined as the square root of the diagonal of the covariance matrix.

However the error bars do not quantify all the uncertainty of the measurement, for that we need the full covariance matrix.

In principle, the measurements of SNe at different redshifts should be uncorrelated → off-diagonal of the covariance matrix is zero and the diagonal contains all the information on the uncertainty of the estimator.

However some of the bias corrections introduce a correlation between redshift bins (for example the evolution effects). For that reason, the covariance matrix is in general not diagonal.



Measuring the uncertainty with mock data

We just saw that we can compute the covariance matrix by computing the dispersion of the measurements.

However, this method does not give reliable results when the sample is small and is not a good representation of a complete sample → the results will suffer from sample variance.

In those cases it is better to measure the uncertainty of the estimator from simulations:

Simulate a random distribution of SN at various redshifts, with a distribution of fluxes and luminosities. From here we can build different realizations of SN samples, including noise, and measure the corresponding ^μ(z)^ for each realization. The samples may be generated with bootstrapping methods. This will result in a set of **mock data**, that contains all the noise properties of the true data.

The uncertainty is then the covariance matrix of the mock ^μ(z)^ data.

Computing the variance of the estimator

Instead of measuring the variance directly on the sample (or on mock data), the variance of the estimator can be analytically computed from its definition:

$$\sigma^2_{\mu_{zz'}} = \left\langle \left[(m_{z,i} - M + bias) - \langle m_{z,i} - M + bias \rangle_{N_z} \right] \left[(m_{z',j} - M + bias) - \langle m_{z',j} - M + bias \rangle_{N_{z'}} \right] \right\rangle_{ij}$$

This is the most consistent way to find the variance, since it is the formal definition and allows to consider not only the statistical error of the measurement (as in the first method) but also **the uncertainty of all terms contributing to the estimator**, such as:

- the uncertainty of the bias correction factors (the **systematic errors**)

- the intrinsic uncertainty of the true value. This is an important contribution for cosmological structure formation probes (not for the SN method), since the parameter values of the Universe are considered to be realizations of an unknown true value. So even the mean of a distribution has an intrinsic error (beyond the standard statistical error that decreases with the size of the sample). This contribution is called the **cosmic variance**.

The various contributions to the estimator uncertainty form the **error budget**.

Taylor expanding the formula of the variance of the estimator, we can write a linear expression for the variance, showing the explicit contributions of the error budget. This is the well-known **error propagation** formula. Assuming that all the terms are independent effects, the uncertainty of the estimator ^μ_z^ (for a given z) may be written as: (only a few terms are shown)

$$\sigma_\mu^2 = \left[\left(\frac{\partial \mu}{\partial m}\right)^2 \sigma_m^2 + \left(\frac{\partial \mu}{\partial M}\right)^2 \sigma_M^2 + \sigma^2_{\text{template fitting } s,\, c} + \sigma^2_{\text{spectroscopic identification}} \right]$$

(σ_m^2: uncertainty in the measured flux; σ_M^2: uncertainty in the intrinsic luminosity)

Note that this method allows us to compute the error of the estimator from the various statistical and systematic error contributions, but does not tell us how to compute those.

Each of the error contributions for the budget need to be computed according to the physical process associated to that contribution.

For example, let us **consider the statistical error associated with the measurement of μ_z**, which propagates from the statistical error of the measurements of flux (magnitude), denoted in the error propagation formula by σ_m.

To compute it we need to realize that the measured signal is determined by the number of SN photons detected per pixel: it is a Poisson process.

The noise in a Poisson process **is the square-root of the number of detections.**
So, if signal is N_photons_per_pixel

→ noise is sqrt (N_photons_per_pixel)

→ signal-to-noise ratio (S/N) is also sqrt (N_photons_per_pixel)

→ the dimensionless relative error is noise/signal, i.e., the inverse of the S/N, i.e.,

$$\sigma_m = 1/\text{sqrt (N\_photons\_per\_pixel)}$$

or, in percentage, the relative error is

$$\sigma_m = 100/\text{sqrt (N\_photons\_per\_pixel)} \quad (\%)$$

We see that, for example, a S/N of 5 (also called a 5-sigma detection) means a relative error of 20%

We also see that the error decreases with the number of detected photons.
So in astronomical observations, **we can decrease the error by increasing the exposure time.**

When applying for telescope time, it is very important to compute in advance what is the needed exposure time. This is determined by the S/N that we want to achieve, but also depends on the specific filter and telescope used.

"Exposure time calculators" (ETC) are codes that compute this for different observational configurations.

An ETC example

Assume we want to prepare the observation of one SN at redshift z=0.8 using the William Herschel 4m telescope at the Observatorio del Roque de los Muchachos (Canary islands).

The observation will be made in the B band and we want to obtain a S/N = 10 at the tail of the light curve

What is the exposure time that we need?

- The signal is determined by the number of SN photons detected per pixel.

Pointing to the host galaxy, we will detect N photons per pixel, approximately half of which come from the SN → signal is N/2

The noise depends on all the photons detected → noise is sqrt(N)

→ S/N = sqrt(N) / 2

We want S/N = 10 → **N = 400 photons per pixel**

- Now, consider the luminosity at the peak:

M_{SN} ~ -19 → $L_{SN} = 4 \times 10^9$ L_Sun = 1.5×10^{36} W= 1.5×10^{43} erg/s

We want a S/N of 10 at the end of the light-curve (~1 month after the peak) in order to have a good detection of the full light-curve. At the end of the light-curve, the luminosity is around 2 magnitudes fainter than at the peak → a factor of 6 in luminosity → L_tail = 2.5×10^{42} erg/s

- Consider now the distance to the SN

$z = 0.8 \rightarrow D_L = 5$ Gpc (assuming the concordance model)

So the expected flux from the tail of the light-curve is:

$F = L / (4 \pi D_L^2) = 8.7 \times 10^{-16}$ erg/s/cm^2

This is the flux we will get in a square centimeter of the telescope

- Consider now the size of the telescope (diameter of 4 meters)

In 1 second, the telescope receives an energy of

$8.7 \times 10^{-16} \times (400)^2 = 1.4 \times 10^{-10}$ erg from the SN

However, the telescope optics and the CCD do not have 100% efficiency.
Assuming the combined efficiency is only 30%

→ the energy detected from the SN in 1 second is 4.2×10^{-11} erg

- Consider now that we observe using the B filter,

The mean wavelength of the filter is 442 nm → frequency of 6.78×10^{14} Hz

This means that the mean energy of a photon detected in this filter is
E = h v (where h is Planck's constant) = 4.5×10^{-12} erg

→ the telescope detects
$4.2 \times 10^{-11} / 4.5 \times 10^{-10}$ = **9.3 SN photons per second**

- Consider now that the size of the SN in the image is 4 pixels

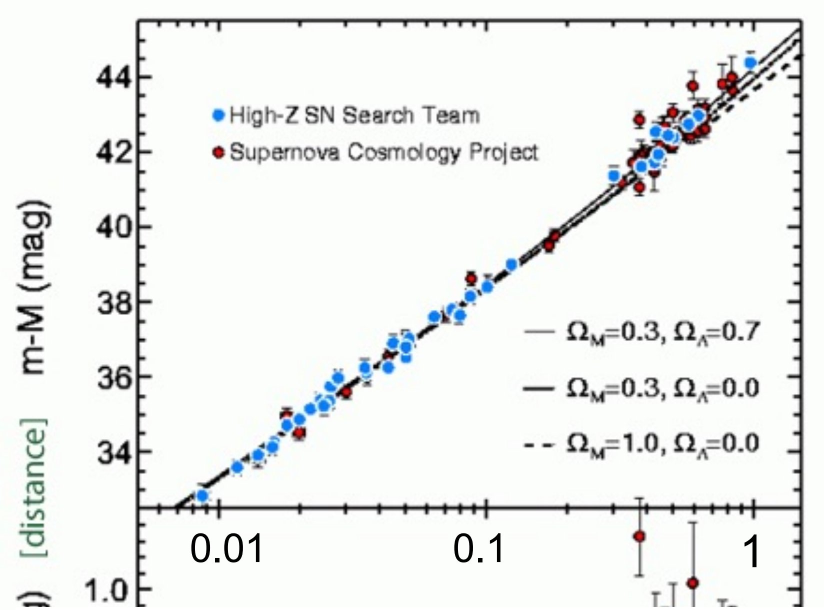
→ in 1 second, we get 9.3 / 4 = 2.3 SN photons per pixel

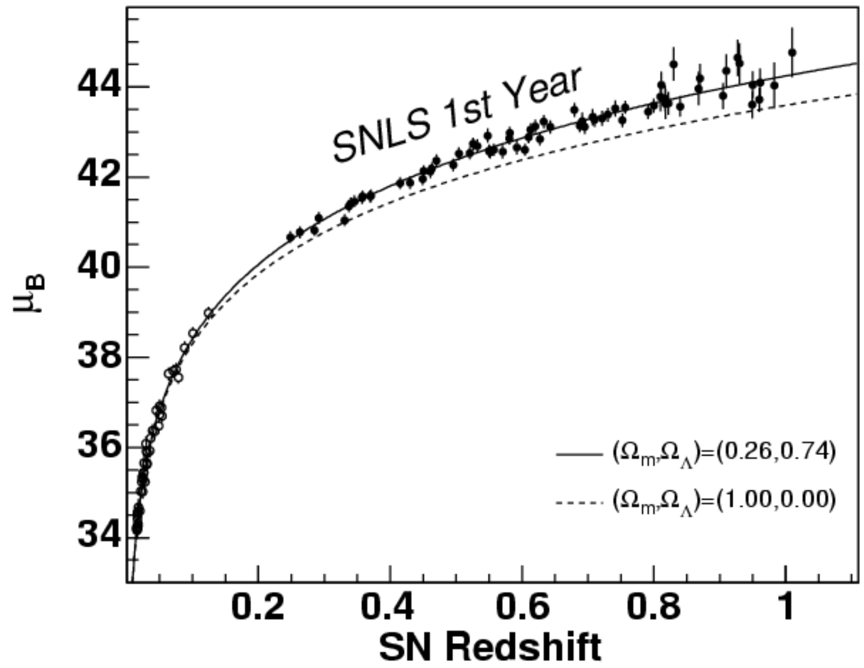
Conclusion: knowing that we need 400 SN photons per second and per pixel, we need an exposure time of 400/2.3 = 170 s ~3 minutes

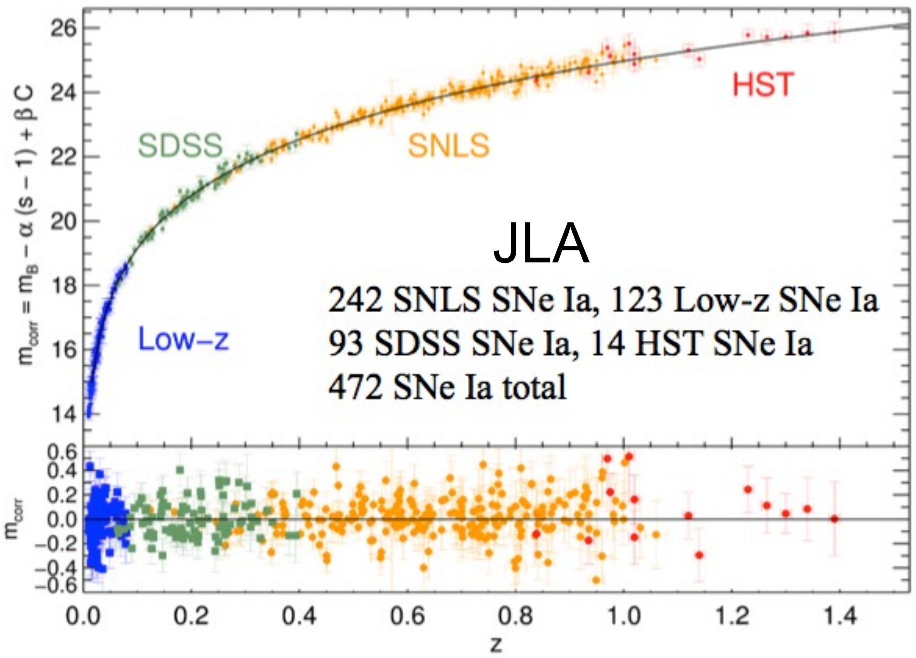
An exposure of 3 minutes in a 4m telescope, with these characteristics and in the B band, forms the image of a z=0.8 SN with a quality of S/N =10 at the tail of the light-curve

The data vector

We have finally obtained the data vector, i.e., the distance modulus vector and its covariance matrix/error bars







1998: Original sample	N ~ 40
2006: SNLS DR1	N ~ 100
2014: JLA compilation	N ~ 500
2018: Pantheon compilation	N ~ 1000
2025: Union3 compilation	N ~ 2000

What do we do with the **data vector**?

We compare it with the theoretical computation, **“the theory vector”:**

$$\mu(z) = 5 \log_{10} [D_L (z; H0, \Omega, w)] + 25$$

and estimate the model free parameters in a **statistical inference analysis**, involving the computation of **likelihoods**.

Hopefully the data vector is measured with an **optimal estimator** **(accurate and precise)**

The estimator is unbiased and no extra uncertainty is introduced

(accurate and precise estimation of the cosmological parameters)

As we saw, in order to find an unbiased estimator of the distance modulus we needed to make a careful consideration of all sources of **systematics**.

We saw that in general **the systematic effects can be treated in different ways:**

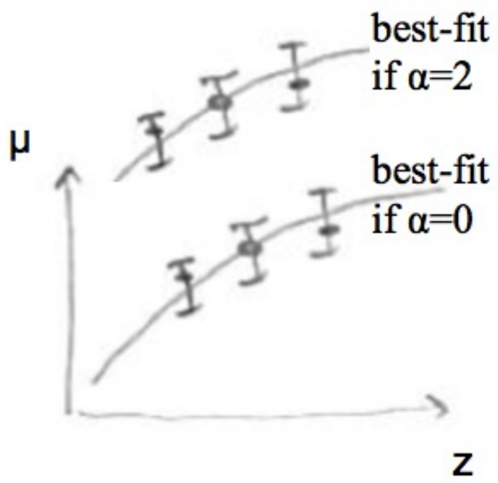
a) **directly computed** (with perfect knowledge) and their values subtracted to the measurements

(e.g. the K-correction)

b) **modeled with a function** (perfectly known) that introduces additional **nuisance parameters** with unknown values

(e.g. the nuisance parameters)

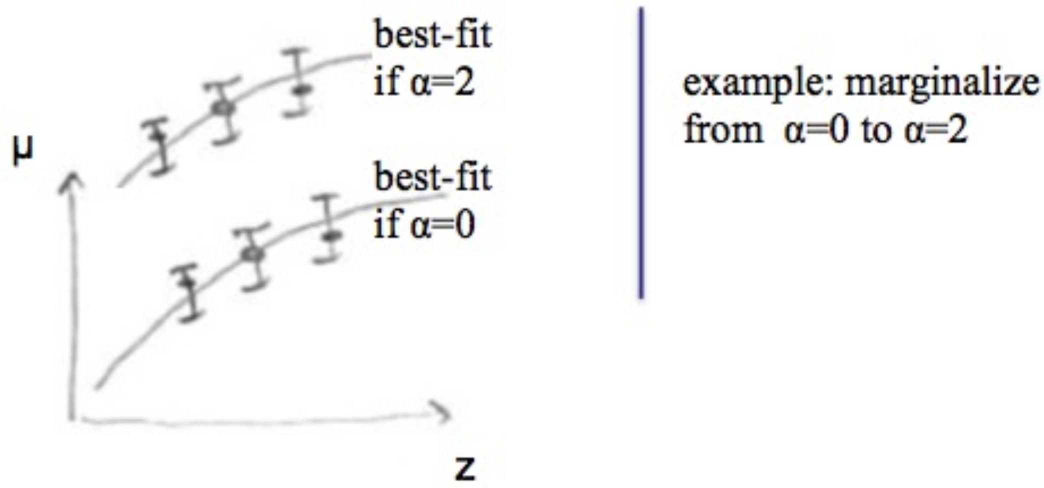
The nuisance parameters can be **estimated** in the statistical inference analysis together with the cosmological parameters.



Depending on the value of α, the data points will be higher or lower in the μ(z) plot. → the best-fit cosmological model will depend on the value of α → **there is a degeneracy between α and the cosmological parameters.**

This makes it difficult to estimate simultaneously the nuisance and cosmological parameters with great precision.

- **marginalized**: usually there is no interested in finding the values of the nuisance parameters. We only want **to estimate the cosmological parameters, but need to consider all the possible range of values of the nuisance parameters.**



This implies that the estimate obtained for the cosmological parameters will have a larger uncertainty than if there were no nuisance parameters → **unbiasing the estimator results in decreasing the precision of the result** → but it increases its accuracy and the result is more reliable.

(accurate but not precise estimation of the cosmological parameters)

The presence of nuisance parameters worsens the constraints on the cosmological parameters.

In terms of Figure-of-Merit (FOM) - area of the confidence contours in the cosmological parameters space, where small area means strong constraints - this means that without the bias correction, the FoM will increase (is better) → but it is a wrong result, too optimistic.

c) **known unknowns**: **systematic error**

In reality, the computation of the corrections is not perfect, there is an uncertainty.

Also the astrophysical modeling may have uncertainties.

It may also happen that we are able to identify some systematics but are not able to compute them directly or to model them.

In these cases we are using a more or less biased estimator. To be able to make a meaningful analysis, in the presence of these **known unknowns**, we need to include additional uncertainties in the error bars: **systematic errors** σ_{sys} → the bias is replaced by an increase in the error bars.

In this case we can still get the right values inside our estimated interval, but it was just because we increased the uncertainty.

Notice that contrary to the statistical error (noise), systematic errors do not average out to zero with large N.

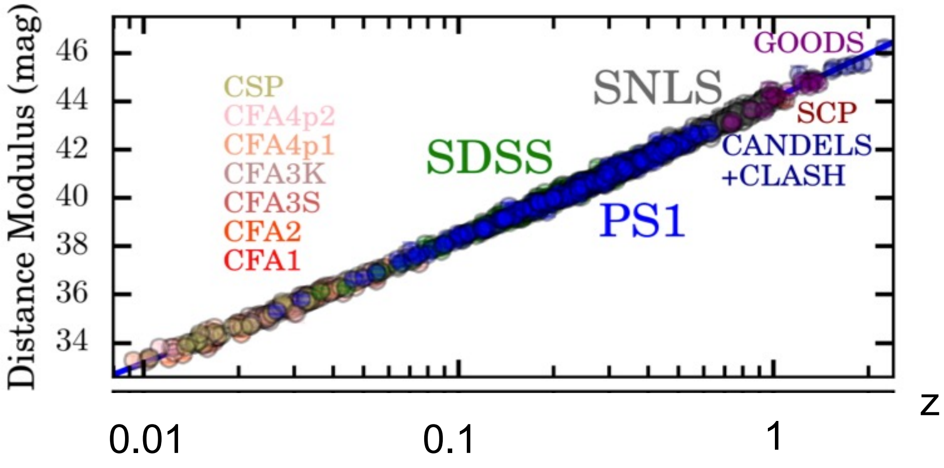
Notice also that if we underestimate the systematic uncertainty that should be allocated, we may end up with the wrong result

d) **unknown unknowns**: do nothing

In the case we are not aware that there are still additional effects and just use the biased estimator (without realising) without doing any correction or modelling, and do not even include systematic uncertainties,

then we can find precise results (if the estimator has a high signal-to-noise ratio) but they will be inaccurate.

The Pantheon compilation



Sample	Number	Mean z
CSP	26	0.024
CFA3	78	0.031
CFA4	41	0.030
CFA1	9	0.024
CFA2	18	0.021
SDSS	335	0.202
PS1	279	0.292
SNLS	236	0.640
SCP	3	1.092
GOODS	15	1.120
CANDELS	6	1.732
CLASH	2	1.555
Tot	1048	

SNLS: SN Legacy Survey
SDSS: Sloan Digital Sky Survey
PS1: Pan-STARRS 1

Precision:
The large number of data points (N increased by a factor of 25) → leads to a factor of 5 improvement in precision.

Accuracy:
Systematic errors from bias corrections decreased because of better corrections, but did not decrease with N → a systematics floor may eventually be reached in future large surveys.

"Precision cosmology is hard, accurate cosmology is even harder"

- Michael Turner -

Cosmological constraints: the cosmographic approach

The cosmological information of the measured distance modulus is contained in the luminosity distance:

$$\mu(z) = 5 \log_{10} (D_L (z; H0, \Omega, w)) + 25$$

with $D_L = (1+z) D_C$ (for a flat Universe)

Let us investigate what is the **cosmological information** that the distance-modulus contains:

First, the comoving distance from the observer at t_0 to the source at t is computed from the metric as

$$D_c = \int_t^{t_0} \frac{1}{a} dt = \int_t^{t_0} (1+z) dt = \int \frac{1}{H(a)} \frac{1}{a^2} da$$

which depends on H_0 and the density parameters (that define the cosmological model).

Alternatively, in order to access the cosmological information in a more fundamental **model-independent** way, let us consider first the **cosmographic approach,** where a(t) is expanded as:

$$a(t) = a_0 \left[1 + H_0 (t - t_0) - \frac{1}{2} q_0 H_0^2 (t-t_0)^2 + \frac{1}{3!} j_0 H_0^3 (t-t_0)^3 + \frac{1}{4!} s_0 H_0^4 (t-t_0)^4 + \mathcal{O}(t-t_0)^5 \right]$$

From here, we can also write an expansion for z(t), since a-1 = - z/(z+1):

$$a - 1 = -\frac{z}{z+1} \Rightarrow \quad z = \frac{1}{a} - 1 \Rightarrow \quad z = (z+1) \left[H_0 \overset{(-1)}{(t_0 - t)} \overset{-1}{+} \frac{1}{2} q_0 H_0^2 (t_0 - t)^2 + \dots \right]$$

$$\Rightarrow \quad z = \underbrace{H_0 (t_0 - t)} + \frac{1}{2} q_0 H_0^2 (t_0 - t)^2 + \underbrace{z H_0 (t_0 - t)} + \frac{1}{2} \underbrace{z q_0 H_0^2 (t_0 - t)^2} + \dots$$

the lowest-order term is order $\mathcal{O}(t)$

order $\mathcal{O}(t^2) + \dots$

order $\mathcal{O}(t^3) + \dots$

Keeping only order $\mathcal{O}(t^2)$ we may use $z \sim H_0 (t_0 - t)$ to insert here

$$\Rightarrow \quad \boxed{z = H_0 (t_0 - t) + \left(\frac{1}{2} q_0 H_0^2 + H_0^2 \right) (t_0 - t)^2 + \mathcal{O}(t^3)}$$

Inserting the z(t) expansion in the comoving distance, we find, to second order:

$$D_c = \int_t^{t_0} [1 + H_0 (t_0 - t)]\, dt$$

(to second order we just need to consider order t in the integrand, because the integral will be order t^2)

$$\boxed{D_c = (t_0 - t) + \frac{H_0}{2} (t_0 - t)^2}$$

At this point it would be useful to invert the expansion z(t), to be able to find an expression for $D_c(z)$ instead of $D_c(t)$

Writing now for t :

$$\Rightarrow \quad t_0 - t = \frac{z}{H_0} - \left(\frac{q_0}{2} H_0^2 + H_0^2\right) \frac{1}{H_0} (t_0 - t)^2 + o(t^3)$$

↓

again the same trick, using $(t_0 - t) \sim \frac{z}{H_0}$

$$\Rightarrow \quad t_0 - t = \frac{z}{H_0} - \left(\frac{q_0 H_0^2}{2} + H_0^2\right)\frac{z^2}{H_0^3} + \ldots$$

$$\text{or} \quad \boxed{t_0 - t = \frac{1}{H_0} z - \frac{1}{H_0}\left(1 + \frac{q_0}{2}\right) z^2 + \sigma(z^3)}$$

Now, Since $D_c = (t_0 - t) + \frac{H_0}{2}(t_0 - t)^2 + \sigma(t^3)$,

we get

$$D_c^{(z)} = \frac{1}{H_0} z - \left(1 + \frac{q_0}{2}\right)\frac{z^2}{H_0} + \frac{H_0}{2}\frac{z^2}{H_0^2} + \sigma(z^3)$$

$$\text{or} \quad D_c = \frac{1}{H_0} z - \left(1 + \frac{q_0}{2} - \frac{1}{2}\right)\frac{z^2}{H_0} + \sigma(z^3)$$

$$\Rightarrow \quad \boxed{D_c^{(z)} = \frac{1}{H_0}\left[z - \frac{1}{2}(1 + q_0)\, z^2\right]}$$

We arrive then at the expression for the luminosity distance that we were looking for:

$$D_L = \left[\frac{z}{H_0} - \frac{z^2}{H_0} \left(\frac{1}{2} + \frac{1}{2} q_0 \right) \right] (1+z) + \sigma(z^3)$$

$$= \frac{z}{H_0} - \frac{z^2}{H_0} \left(\frac{1}{2} + \frac{1}{2} q_0 \right) + \frac{z^2}{H_0} + \sigma(z^3)$$

$$\text{or} \quad \left[D_L = \frac{1}{H_0} \left(z + \frac{1}{2} (1 - q_0) z^2 \right) \right] \quad (\text{flat})$$

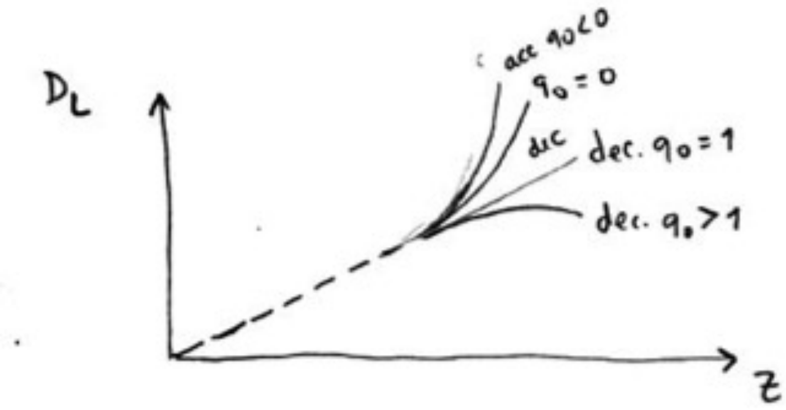
We see that, up to second-order, the luminosity distance:

- at low-z measures the (constant) **velocity of the Universe** (with (c)z/H_0 being the Doppler velocity)

- at high-z measures the (constant) **acceleration of the Universe** (q_0)

So, in order to detect an acceleration of the Universe we need to observe SNe at **high redshifts**.

SNe at **low redshifts** measure H_0 , i.e., the slope of the D_L (z) straight line



To determine **the value of the acceleration**, both high and low redshift measurements are needed, to break the **degeneracy** between H_0 (the Hubble law slope, important at lower z) and the acceleration (important at higher z)

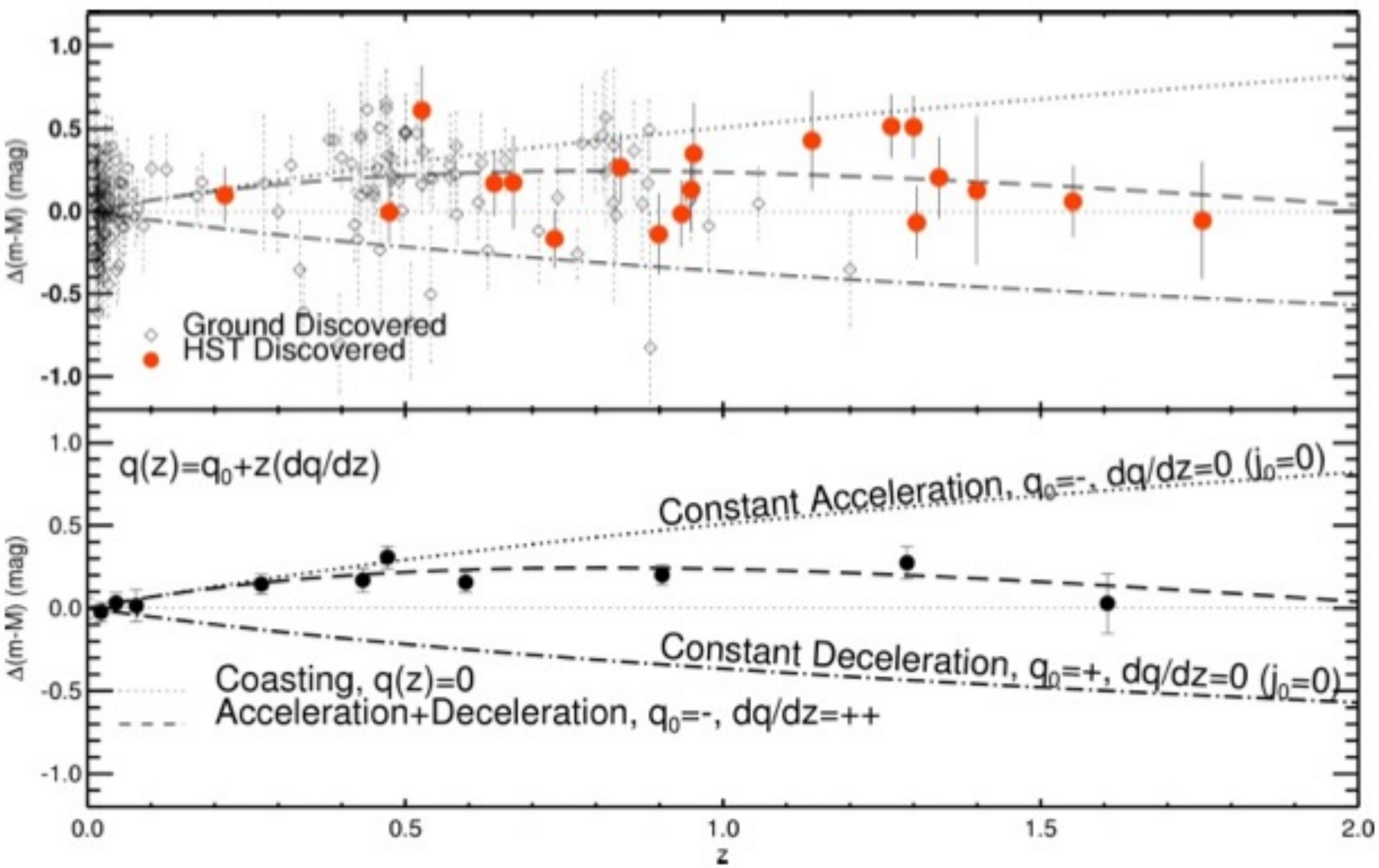
Note that the evidence for acceleration is based only on the shape of the function.

Even if the absolute values determined for the distances were not precise (i.e., if H_0 was estimated with a large uncertain) we could still find evidence for acceleration using only high-z SNe.

This result is only an approximation. **If we consider higher-orders**, we find

$$D_c(z) = \frac{z}{H_0}\left[1 - \left(1+\frac{q_0}{2}\right)z + \left(1+q_0+\frac{q_0^2}{2}-\frac{j_0}{6}\right)z^2 - \left(1+\frac{3}{2}q_0(1+q_0)+\frac{5}{8}q_0^3-\frac{1}{2}j_0-\frac{5}{12}q_0 j_0-\frac{s_0}{24}\right)z^3 + O(z^4)\right]$$

The introduction of higher orders shows that the acceleration is not necessarily constant, i.e., the quadratic term depends also on j_0, i.e., there is a non-zero dq/dz.



The data seem to prefer a model with varying q_0, such that

$q_0 > 0$ at high-z

and

$q_0 < 0$ at low-z.

This plot shows the dynamic behaviour of the Universe (independently of the values of the density parameters) → clear **evidence for a model with acceleration for $z < 1$ and deceleration for $z > 1$** → proof of **late-time acceleration of the universe**

Now, the cosmographic analysis was very useful to get an insight of the dynamic behaviour of the Universe, but in order to estimate cosmological parameters this analysis is not needed.

What we need is just to compute the observable cosmological function (i.e. the luminosity distance) from vectors of cosmological parameter values and compare the various theoretical D_L obtained with the observed one through the computation of likelihoods in the parameter space.

Cosmological constraints: parameter estimation

The dependence of the luminosity distance on the cosmological parameters can be most easily seen by writing the distance as an integral over redshift:

In general, $D_L(z) = (1+z)\, D_M$,

considering the case of flat Universe, we have

$$D_L(z) = (1+z)\, D_C(z) = (1+z) \int_{t}^{t_0} \frac{1}{a}\, dt = (1+z) \int_0^z \frac{c\,dz}{H(z)}$$

and the Hubble function is found in terms of the parameters of the cosmological fluid (densities of the various sources) through Friedmann's equation:

$$H^2(a) = H_0^2 \left(\Omega_r (1+z)^4 + \Omega_{\rm m} (1+z)^3 + \Omega_{\rm K} (1+z)^2 + \Omega_\Lambda \right)$$

(here including a cosmological constant)

From this, the luminosity distance can be computed for any values of the vector of cosmological parameters, and for the redshifts of the various SN.

Then all terms of the distance-modulus estimator are added (introducing a large number of nuisance parameters)

and finally the theoretical distance-modulus $\mu(z)$ is found.

Its likelihood can then be computed by comparing with the distance-modulus data.

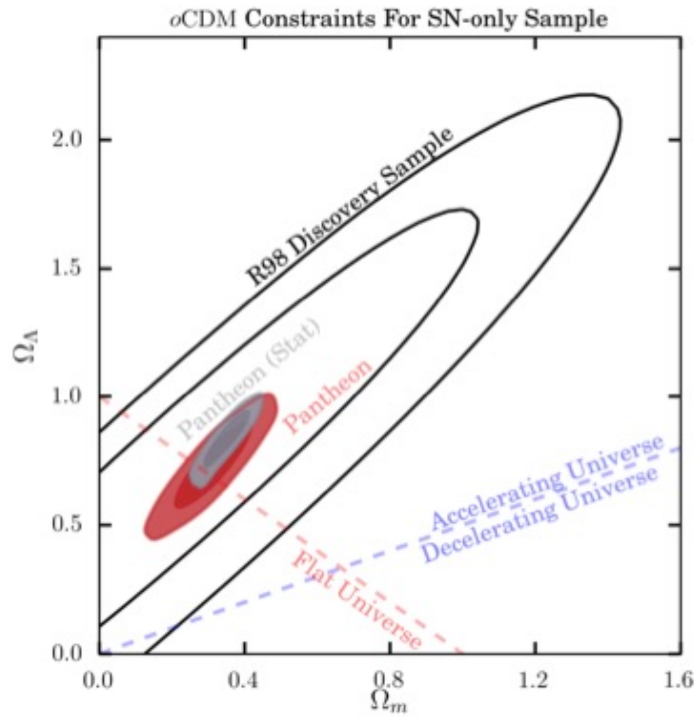
In a sampling method, the procedure is then repeated for millions of points in the parameter space. The cosmological parameters estimates are finally found by marginalizing over the nuisance parameters.

It is important to realize that the results depend on the cosmology assumed (it is a working hypothesis).

Constraints in the ΛCDM scenario

flat ΛCDM → 2 independent background cosmological parameters: H_0 , Ω_m, since Ω_r and Ω_K are fixed and $\Omega_\Lambda = 1 - \Omega_m$. If curvature is not fixed *a priori* then there are 3 free parameters: H_0 , Ω_m, Ω_Λ, and this is historically called **oCDM** (open CDM, even though the fit is free to have any curvature - flat, open or closed)

Constraints in the (Ω_m, Ω_Λ) plane: after marginalizing over the nuisance parameters and H_0, the constraints on the two density parameters are given as confidence contours in the 2D parameter space.



Some notes:

- The large contours are from the **first SN results** of 1998. They show quite large 1σ and 2σ probability contours, since the data has large error bars.

- The smaller contours (also showing 1σ and 2σ contours) are for the **recent Pantheon results.**

- Notice also the impact of considering or not **the contribution of the systematic effects** for the data error bars:

Grey contours - analysis done using Pantheon data with error bars including only the errors (statistical uncertainties)

Red contours - analysis done using Pantheon data with error bars including statistical + systematic uncertainties)

So, with larger error bars, a larger region of the parameter space has a "good likelihood" and is included inside the confidence contours (red larger than grey).

- If H_0 was known (**fixed** in the analysis instead of marginalized), the (Ω_m , Ω_Λ) contours would be smaller (tighter constraints)

- Models such that $\Omega_\Lambda = 1 - \Omega_m$ (i.e., $\Omega_K = 0$), lie on the straight line marked "**flat**"

- There is also a line dividing **accelerating** and **decelerating** models.

Notice the contours are all aligned on a preferred direction. Why is this?

To answer this question, let us remember that to first approximation, we are measuring the acceleration of the Universe, i.e., as we saw, D_L depends directly on q_0

$$q_0 = - \frac{\ddot{a}}{a}\Big|_{t_0} \frac{1}{H_0^2}$$

From Raychadhuri's equation, we can write the acceleration parameter in terms of the source parameters:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p)$$

$$= -\frac{8\pi G}{3H_0^2} \frac{H_0^2}{2} \rho(1+3W(a)) \Leftrightarrow \frac{\ddot{a}}{a} = -\frac{H_0^2}{2}\left[\Omega_m a^{-3} + \Omega_{DE}(1+3w(a))\right]$$

(for a general dark energy fluid)

For the case of a cosmological constant:

$$\Rightarrow \frac{\ddot{a}}{a}\Big|_{t_0} = -\frac{H_0^2}{2}\left[\Omega_m + \Omega_{DE}(1+3w)\right]$$

$$\text{or, for } \Omega_{DE} = \Omega_\Lambda \Rightarrow \frac{\ddot{a}}{a}\Big|_{t_0} = -\frac{H_0^2}{2}(\Omega_m - 2\Omega_\Lambda)$$

and so $\boxed{q_0 = \frac{1}{2}\Omega_m - \Omega_\Lambda}$

(note that q_0 is independent of H_0, which is consistent, since they are directly different orders of the Taylor expansion → acceleration is independent of velocity)

So, models with the same acceleration (same q_0) all lie in a line

$$y = ax + b$$

where $y = \Omega_\Lambda$, $x = \Omega_m$, $a = 1/2$, $b = -q_0$ (which is >0 for an accelerated model)

This line defines the direction of the contour (with some width due to the uncertainty on the measured acceleration)

This shows that Ω_m and Ω_Λ are correlated in the acceleration they produce. The two parameters define a straight line along which all models have exactly the same acceleration and will have exactly the same likelihood values → a degeneracy direction

Consider a model 1. Now, if a model 2 has a higher Ω_Λ with respect to model 1, then by also increasing its Ω_m value the acceleration produced by model 2 will be the same as for model 1 → they are **correlated (positively correlated)** → contours from bottom-left to top-right.

If to keep the acceleration constant when one of the parameters increase, the other would need to decrease, then they would be → **anti-correlated (negatively correlated)** → contours from top-left to bottom-right

Two different ways of increasing luminosity distance:

1) Increase Ω_Λ
2) Decrease Ω_m

This causes the degeneracy between Ω_Λ and Ω_m

It is then impossible to distinguish those 2 models (or any model along the degeneracy direction) with SN measurements (or any other D_L based method).

In general, **cosmological probes are very good in constraining degeneracy directions** (i.e. combinations of cosmological parameters) but not so good in constraining individual parameters.

In our case, SN measurements are good in constraining the orthogonal direction to the degeneracy direction i.e., the deviation from the acceleration line (or the width of the contours).

Note that a parameter defined along the width of the contours would be highly constrained - this parameter corresponds to the last principal components in a PCA analysis of the parameter space covariance matrix.

Notice that ($0.5\ \Omega_m - \Omega_\Lambda$ = constant) is a perfect degeneracy. Why then do the contours close and do not extend infinitely along the degeneracy direction?

This is because the linear dependence of D_L on q_0 is only a good approximation at second-order of the a (t) expansion. **In reality, there are other terms and degeneracy is not perfect → the contours close and show a preference for Ω_m < 1 (and Ω_Λ > 0)**

Could we get better estimates for individual parameters?

The Pantheon results are:

Analysis	Model	Ω_m	Ω_Λ
SN-stat	ΛCDM	0.284 ± 0.012	0.716 ± 0.01
SN-stat	oCDM	0.348 ± 0.040	0.827 ± 0.06
SN	ΛCDM	0.298 ± 0.022	0.702 ± 0.02
SN	oCDM	0.319 ± 0.070	0.733 ± 0.11

Notice the constraints are looser (worse) if:

- systematics are included in the error budget

- curvature is left free (one more free parameter to add to the general degeneracies)

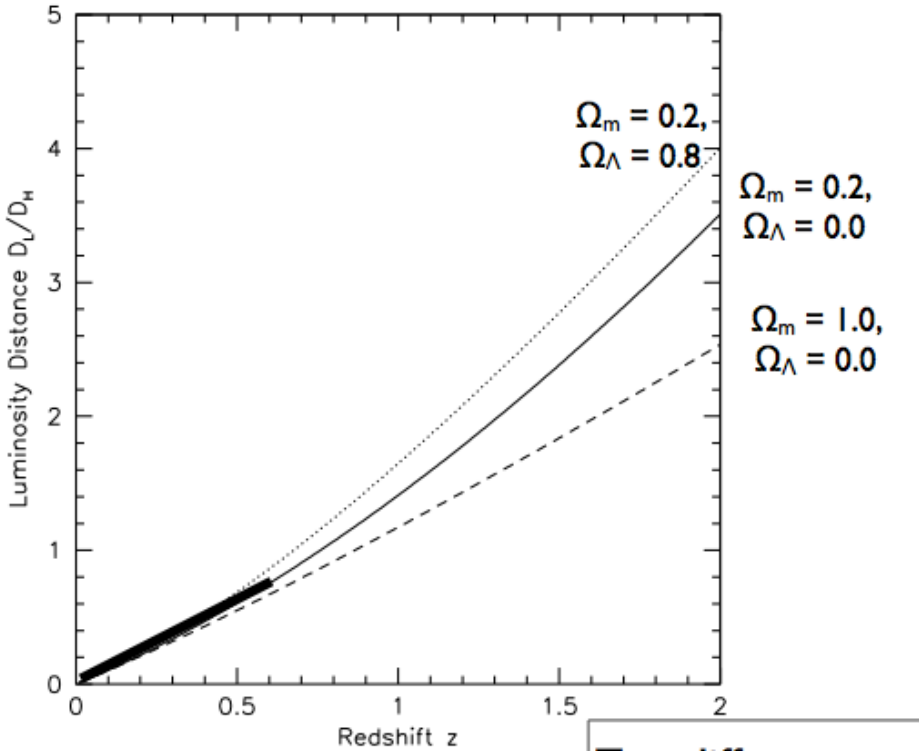
Notice also that in the (flat) ΛCDM case, the result for Ω_Λ is just $1-\Omega_m$

We can improve the constraints by combining various cosmological probes in order to break the degeneracy.

For example, consider an observable that would depend directly on the curvature of the Universe. In the (Ω_m, Ω_Λ) plane we see that lines of constant curvature are more or less orthogonal (i.e. complementary) to lines of constant acceleration.

The joint likelihood analysis of those two datasets would produce contours in the intersection of the two directions → i.e. potentially small round contours → constraining simultaneously the two parameters Ω_m and $\Omega_{\Lambda.}$

Do the SN data prove the existence of dark energy?



The evidence from the data is for acceleration (based on the shape of the D_L (z) function).

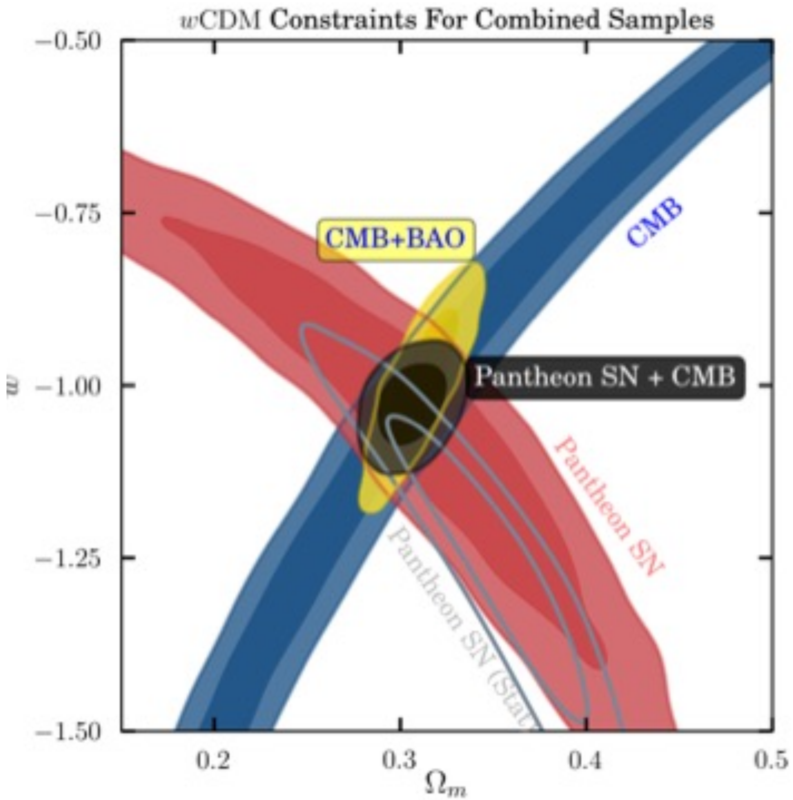
The "evidence for dark energy" is a model-dependent conclusion (i.e. based on the assumption of an underlying cosmology) and therefore less robust than the evidence for acceleration.

Constraints in the wCDM scenario

where dark energy has a constant equation of state, but not necessarily equal to -1 (which would be ΛCDM).

wCDM → there are 4 independent background cosmological parameters: H_0 , Ω_m, Ω_{DE}, w (or alternatively H_0 , Ω_m, Ω_K, w), or only 3: H_0 , Ω_m, w, if flatness is also assumed (Ω_K = 0 and Ω_{DE} = 1 - Ω_m)

Constraints on the (Ω_m, w) plane
(after marginalizing over the other parameters)



<u>Some notes:</u>

- The SN-Pantheon contours (red) are in a very **different direction** than the contours in the (Ω_m, Ω_Λ) plane that we saw previously.

This is because that (as before) they are determined by the acceleration parameter q_0 , which now (from Raychaudhuri's eq.) is,

$$q_0 = \frac{1}{2}\Omega_m + \Omega_{DE}\left(\frac{1+3w}{2}\right)$$

i.e., Ω_m and w add, instead of subtracting (contrary to the relation between Ω_m and Ω_Λ), and so they are anti-correlated. (Note that this is just an effect of w being negative)

- The **contour is no longer an ellipse** (it is curved).

This is because the line of constant luminosity distance (which in our $O(z^2)$ approximation is the line of constant acceleration) is no longer a straight line in the (Ω_m, w) plane.

When we move along a straight line in this plane, a change on Ω_m induces a change on Ω_{DE} → the dependence of q_0 on the parameters is no longer linear.

Indeed, if we replace $\Omega_\Lambda = 1 - \Omega_m$ in the expression for q_0, we get

$$q_0 = \frac{1}{2} + \frac{3}{2} w \left(1 + \Omega_m\right)$$

i.e., $y = a\,(1-x)^{-1}$, where $y = w$, $x = \Omega_m$, $a = 2/3\,(q_0 - 1/2)$ → corresponding to the red curved contour

We recover the result that only in the case the cosmological function (in this case, distance modulus, luminosity distance, acceleration) **depends linearly on the parameters, is the posterior distribution in the parameters space a Gaussian (leading to elliptical contours).**

- Contours from CMB-Planck measurements are orthogonal to the SN ones (they do not measure the luminosity distance or acceleration but different observables, like the angular size of the sound horizon at recombination) → they are complementary probes, and the joint contours are much reduced.

- Baryonic Acoustic Oscillations (BAO) measurements are similar to the SN ones (they measure the angular diameter distance) and also complementary to CMB.

Sample	w
CMB+BAO	-0.991 ± 0.074
CMB+H0	-1.188 ± 0.062
CMB+BAO+H0	-1.119 ± 0.068
SN+CMB	-1.026 ± 0.041
SN+CMB+BAO	-1.014 ± 0.040
SN+CMB+H0	-1.056 ± 0.038
SN+CMB+BAO+H0	-1.047 ± 0.038

SNe Ia distances combined with CMB and/or BAO remain the best probe to constraint the DE equation of state :

- a 5% measure of a constant DE EoS, w, is achievable
- currently little sensitivity to w(z)

Including systematics and combined with BAO and CMB : w (cte) = -1.018 ± 0.057 (~6%) compatible with a cosmological constant

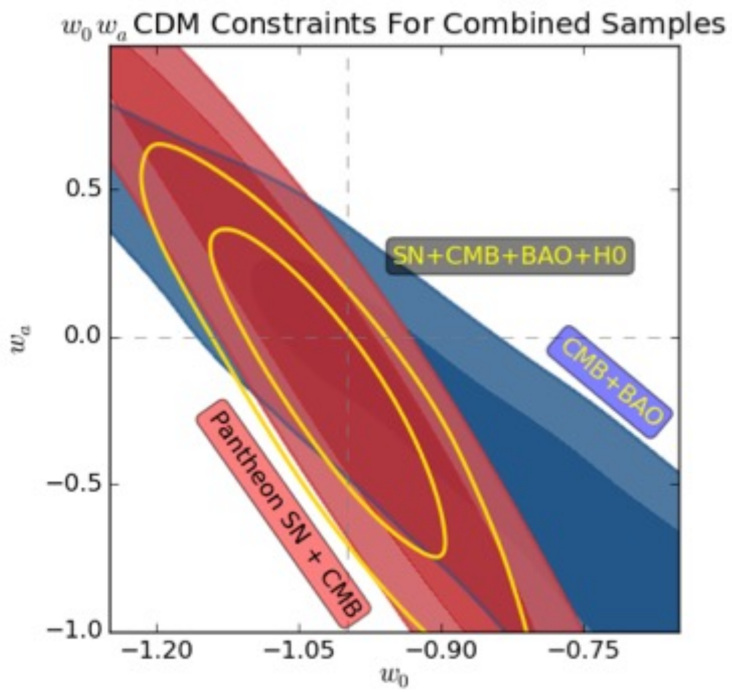
Constraints in the w(z)CDM scenario

where dark energy has an evolving equation of state

w(z)CDM → there are now 5 independent cosmological parameters: H_0 , Ω_m, Ω_{DE}, w_0, w_a (or only 4 if flatness is assumed)

The evolution of the dark energy equation of state is parameterized as $w(z) = w_a$ which is a first-order Taylor expansion in the scale factor: $w_0 + w_a (1 - a)$

Constraints in the (w_0, w_a) plane
(after marginalizaing over the other parameters)



Some notes:

- The effect of w_a on the geometric observables is very weak → probes of structure are more useful, since the evolution of dark energy affects structure formation
- Due to the weak constraints, the figure only shows combined contours: SN+CMB, BAO+CMB, SN+CMB+BAO+H0_prior

- ΛCDM is a point in this plane (w_0 = -1, w_a =0), and is inside all the contours

Sample	w_0	w_a	Ω_m	H_0	FoM
CMB+BAO	-0.616 ± 0.262	-1.108 ± 0.771	0.343 ± 0.025	64.614 ± 2.447	14.5
CMB+H0	-1.024 ± 0.347	-0.789 ± 1.338	0.265 ± 0.015	73.397 ± 1.961	9.1
CMB+BAO+H0	-0.619 ± 0.270	-1.098 ± 0.781	0.343 ± 0.026	64.666 ± 2.526	14.5
SN+CMB	-1.009 ± 0.159	-0.129 ± 0.755	0.308 ± 0.018	68.188 ± 1.768	31.4
SN+CMB+BAO	-0.993 ± 0.087	-0.126 ± 0.384	0.308 ± 0.008	68.076 ± 0.858	65.0
SN+CMB+H0	-0.905 ± 0.101	-0.742 ± 0.465	0.287 ± 0.011	70.393 ± 1.079	54.2
SN+CMB+BAO+H0	-1.007 ± 0.089	-0.222 ± 0.407	0.300 ± 0.008	69.057 ± 0.796	63.2

The **dark energy figure-of-merit** (FOM) is defined as the inverse of the area of the 1-sigma contour - or more precisely, it is the area of an ellipse that fits the countour.

The larger the FoM → the smaller the contour → the stronger the constraint.

The most powerful combination in the table is SN+CMB+BAO.

Model selection (Goodness-of-fit)

The estimation of the cosmological parameter contours (also called credible intervals - mean values and uncertainties -) is not the last step of the cosmological data analysis process.

Using the SN - Pantheon example, let us look at its results:

Nuisance parameters

(notice the uncertainties are much larger if only low-z SN are used)

Survey	α	β	γ
Pantheon	0.154 ± 0.006	3.02 ± 0.06	0.053 ± 0.009
Low-z	0.154 ± 0.011	2.99 ± 0.15	0.076 ± 0.030

Cosmological parameters

constraints are worse if the full (stat+sys) errors are used (more realistic)

Analysis	Model	w	Ω_m	Ω_Λ
SN-stat	ΛCDM		0.284 ± 0.012	0.716 ± 0.01
SN-stat	oCDM		0.348 ± 0.040	0.827 ± 0.06
SN-stat	wCDM	-1.251 ± 0.144	0.350 ± 0.035	
SN	ΛCDM		0.298 ± 0.022	0.702 ± 0.02
SN	oCDM		0.319 ± 0.070	0.733 ± 0.11
SN	wCDM	-1.090 ± 0.220	0.316 ± 0.072	

From the table, it is clear that the results depend on the scenario assumed:

- ΛCDM (Ω_m) - with few free parameters, the constraints are tighter
- oCDM (Ω_m Ω_Λ) - not only parameter uncertainties are larger, but the central values can change a lot (central values for ΛCDM are not even contained in the oCDM 1σ confidence intervals)
- wCDM (Ω_m w Ω_K) - constraints closer to the oCDM ones

So, what is the final result? What is our finding, is it Ω_m 0.30 or 0.32?

This is a question of **goodness-of-fit**. Among the various best-fits which one is the best?

We turn to Bayesian inference to answer this question by performing **model comparison** tests.

There are different ways to evaluate the goodness-of-fit. The classic way is to look at the **chi-square**, while the most rigorous way is to use the **evidence**.

Chi-square

Criteria based on the chi-square values are standard in determining the best model in all branches of physics.

The most usual quantity is the reduced chi-square of the best-fit, i.e., the chi-square normalised by the number of degrees-of-freedom,

$N_{dof} = N_d - N_p$ (where N_d is the number of datapoints - for example the number of redshift bins in the SN data - and N_p is the number of parameters in the model)

In this criterium, **the best model** (i.e., the favoured one) **is the one where the best-fit has the lowest reduced chi-square,**

$$\chi^2_{red} = \chi^2 / N_{dof}$$

Evidence

is the integral of the likelihood on the parameters space of a given cosmological model → it indicates the ‘average likelihood of a model“.

It may happen that a certain set of parameter values are a very good fit to the data (high likelihood values in that region of the parameter space), but overall this model can have a worse evidence than another one (for example because of having a larger number of parameters, or a large region of small likelihood values).

The evidence is thus a global way to characterize the goodness-of-fit of a model, beyond the simple assessment of finding which model has the “best best-fit”.

The evidence is a good number to show the balance between **best-fit vs. model complexity.**

In this approach, the best model is the one with the highest Bayes factor, computed from the evidences of the 2 models under comparison:

B = (Evidence_1 * Prior_1) / (Evidence_2 * Prior_2)

The **Jeffrey's scale** classifies the degree of preference for a model over another, based on the values of $\ln B$:

<1 → inconclusive
1 - 2.5 → substantial evidence for one of the models
2.5 - 5 → strong evidence
> 5 → decisive evidence

The evidence is difficult to compute in practice with high precision, since it is **a multi-dimensional integral** of a possibly complex posterior distribution function.

Moreover, by **sampling the posterior** with a grid or a Markov chain Monte Carlo (MCMC) method, we only know a rough sample of it, which may be good enough to find the parameter constraints, but not precise enough to compute the total integral.

By design, **MCMC algorithms** only sample with high resolution the region near the maximum of likelihood. The tails of the distribution are usually badly sampled because they are not needed for parameter inference.

So the sample obtained with MCMC is not complete enough to compute the evidence. We need other Monte Carlo sampling methods to solve the multi-dimensional integral.

A popular algorithm for this is the **Nested Sampling:**

1. Sample N points randomly from within the prior, and evaluate their likelihoods. Initially we will have the full prior range available, i.e., $(0, X_0 = 1)$.

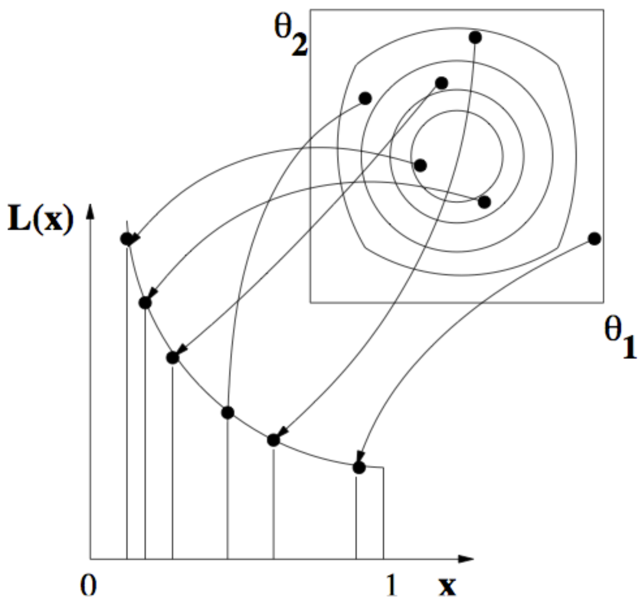
2. Select the point with the lowest likelihood (L_j). The prior volume corresponding to this point, X_j, can be estimated probabilistically. The average volume decrease is given as $X_j/X_{j-1} = t$, where t is the expectation value of the largest of N random numbers from uniform (0, 1), which is $N/(N+1)$.

3. Increment the evidence by $E_j = L_j(X_{j-1} - X_{j+1})/2$.

4. Discard the lowest likelihood point and replace it with a new point, which is uniformly distributed within the remaining prior volume $(0, X_j)$. The new point must satisfy the hard constraint on likelihood of $L > L_j$.

5. Repeat steps 2–4, until the evidence has been estimated to some desired accuracy.

This means: find iso-regions of likelihood. If they are 'nested' the integrand is monotonic → the integral reduces to 1-dimension.



For each layer → $$E_j = \frac{L_j}{2}(X_{j-1} - X_{j+1}).$$

The total evidence is → $$E = \sum_{j=1}^{m} E_j.$$

Information criteria

Besides the evidence, there are alternative approximate methods, much simpler to compute, that can also be used for model selection and quantify the balance of best-fit vs. model complexity. Some popular **information criteria** are:

Akaike information criterion: AIC = -2 ln L_bestfit + 2 n_p = χ^2_bestfit + 2n_p
(this formula is the result of a minimisation of entropy criterium)

Bayesian information criterion: BIC = -2 ln L_bestfit + n_p ln(n_d)
(based on an approximation of the evidence)
BIC penalizes more the complexity than AIC does.

Deviance information criterion: DIC = 2 χ^2_mean - χ^2_bestfit
(it is like an effective χ^2 , sensitive to the difference between the best-fit and the full distribution).

For all information criteria, **the best model is the one with the lowest value**.

Model selection results: an example

In this example, SN data was used to test two very different scenarios: ΛCDM and UDM (model where DM and DE are one single fluid.

This model has one density parameter less, but two new additional parameters - so one parameter more than ΛCDM in total).

Two different UDM models were tested and (like ΛCDM) both are able to produce D_L (z) functions that allow for good fits to the SN data for certain values of their parameters.

Various model selection criteria were computed:

The question is, is there enough evidence to select UDM over ΛCDM?

	UDM	ΛCDM	UDM_{ph}
χ^2_{min}	552.59	552.77	552.75
χ^2_{red}	0.9478	0.9449	0.9481
$\ln B_{U\Lambda}$	−0.0196	0	0.6850
BIC	584.485	571.902	584.644
DIC	553.250	552.770	552.814

- The first UDM model is the one with the smallest best-fit χ^2 , i.e., it contains a vector of parameter values that produced the closest fit to the data.

However, since this model has more cosmological parameters than ΛCDM it is penalized, and the lowest reduced chi-square turns out to be the one of ΛCDM.
The complexity of the model (having more free parameters) is always penalized in these criteria. This is because increasing the number of parameters naturally helps in finding a closer fit (in a potentially artificial way).

- UDM_ph is the model with largest evidence. Indeed, the Bayes factor of the second UDM model with respect to ΛCDM is positive, although smaller than one → the analysis shows a very slight inconclusive preference for this model UDM_ph

- BIC shows a reasonable preference for ΛCDM.

- DIC shows a slight preference for ΛCDM.

The analysis does not show a conclusive preference for any of the models
(but given the close results, it shows that it is useful to compute all the criteria).