Lecture I: Discovery of the Accelerating Universe

Recommended reading:

- "Measurements of the cosmological parameters Ω_M and Λ from 42 high-redshift supernovae", S. Perlmutter et al. (The Supernova Cosmology Project), Astrophysical Journal **517**, 565 (1999) — *a classic.*
- "Supernovae, Dark Energy, and the Accelerating Universe", S. Perlmutter, Physics Today, April 2003 — *very nice popular account.*
- "Measuring Cosmology with Supernovae", Saul Perlmutter and Brian P. Schmidt, Supernovae & Gamma Ray Bursts, K. Weiler, Ed., Springer, Lecture Notes in Physics, astro-ph/0303428 — *intermediate level overview.*
- "Improved Dark Energy Constraints from $\sim$ 100 New CfA Supernova Type Ia Light Curves", M. Hicken et al., arXiv:0901:4804 — *latest and greatest SN Ia data set.*
- "Dark Energy and the Accelerating Universe", J. Frieman, M. Turner and D. Huterer, Ann. Rev. Astron. Astrophys. **46**, 385 (2008) (`http://huterer8.physics.lsa.umich.edu/~huterer/Papers/ARAA_DE.pdf`) — *a review of DE for a general-practice physicist or astronomer.*

Introduction. Type Ia supernovae are interesting objects. They have been studied extensively by the famous American-Swiss astronomer Fritz Zwicky (there is a notable paper by Baade and Zwicky from 1934); Zwicky gave them their name. They have been known to have nearly uniform luminosity; this feature is easily understood from the currently favored explanation for the physics of these events: these are white dwarf stars accreting matter from a companion, going over the Chandrasekhar limit, and undergoing explosion.

Explosions of type Ia supernovae are extremely luminous events that can be seen across the observable universe. At their peak, SNe Ia can be as luminous as the entire galaxy in which they reside.

Standard candles. It is very difficult to measure *distances* in astronomy. You can get redshift of an object from its spectrum, but how do you get the distance? There are many empirical — and uncertain — ways to do so (surface brightness fluctuations, period-luminosity relation of Cepheids, proper motions, etc). Typically, astronomers construct an unwieldy "distance ladder" to measure distance to a distant galaxy: they use some of these relations (say, proper motions) to calibrate distances to more nearby objects, then go from those objects to more distant ones using other relations that work better in that distance regime. This procedure is clearly not robust.

A "standard candle" is a hypothetical object that has a fixed luminosity (that is, fixed intrinsic power that it radiates). Having a standard candle would be useful since then you could infer distances from objects just by using the inverse square law, $f = L/(4\pi d^2)$, by measuring the flux f and knowing the luminosity L from the standard candle property. In fact, you don't even to know the luminosity of the standard candle to be able to infer *relative* distances to objects – all you need to know that thus luminosity is the same for all objects.

The fact that SNe Ia can potentially be used as a standard candle has been realized long ago (at least as far back as the 1970, as far as I know). Briefly, SNe Ia are extremely bright explosions that are presumed to be cases when a white dwarf accretes matter from a (binary) companion, and exceeds the Chandrasekhar mass limit and explodes. These SN type is defined by the absence of Hydrogen in their spectra, but the unmistakable presence of a Silicon line at 6150 Angstroms.

Finding SNe. However, a real problem was scheduling telescopes to detect and "follow-up" SNe that are discovered. Basically, if you point a telescope at a galaxy and wait for the SN to go off, you will wait an average of 500 years. There has been a program in the 1980s to do that (Norgaard-Nielsen, 1989) and it discovered only one SN, and after the peak!

A real breakthrough came in the 1990s when two teams of SN researchers, Supernova Cosmology Project (SCP; lead by Saul Perlmutter) and High-z Supernova Team (Highz; lead, at the time, by Brian Schmidt) made careful use of world's most powerful telescope working in concert to discover and follow up SNe, essentially guaranteeing to funding agencies that they will find batches of SNe in each run.

Some of the early results came out in 1997, which paradoxically indicated that the universe is matter-dominated and consistent with being flat. But those early measurements were based on only 7 high-z SNe and had large errors. The definitive results came out in 1998 (High-z team) and 1999 (SCP, though they had results earlier) and indicated that the universe is dominated by a component with negative pressure.

Broader is brighter. Another major breakthrough came in 1993 by Mark Phillips (astronomer working in Chile). He noticed that the *intrinsic* SN brightness (or, SN luminosity) is correlated with the decay time of SN light curve. Phillips considered the following quantity: Δm_{15}, the decay of the light curve 15 days after the maximum. He found that Δm_{15} is strongly correlated with the intrinsic brightness of SNe. In other words, he found the "Phillips relation" which roughly goes as

Broader is brighter.

(see left panel of Fig. 1). Phillips found that the intrinsic dispersion of SNe, which is ~ 0.5 magnitudes (depending on which band you look at, etc), can be brought down to ~ 0.2 magnitudes once you *correct* each SN luminosity by this relation, using of course its measured Δm_{15}. In astronomers' language, the Phillips relation is

$$M_{\rm max} = a + b\,(\Delta m_{15}) \tag{1}$$

where $m_{\rm max}$ is the absolute magnitude of SN, a and b are some constants, and the dispersion in this relation around the mean is small as mentioned above.

The Phillips relation was the second key breakthrough that enabled SNe Ia to achieve precision needed to probe dark energy.

Observable and inferred quantities with SNe Ia. The astronomers use apparent magnitudes to measure apparent (measured) brightness of an object. The difference between the apparent and absolute magnitude is the *distance modulus*. In particular,

$$DM \equiv m - M = 5\log_{10}\left(\frac{d_L}{10\,\rm pc}\right) \tag{2}$$

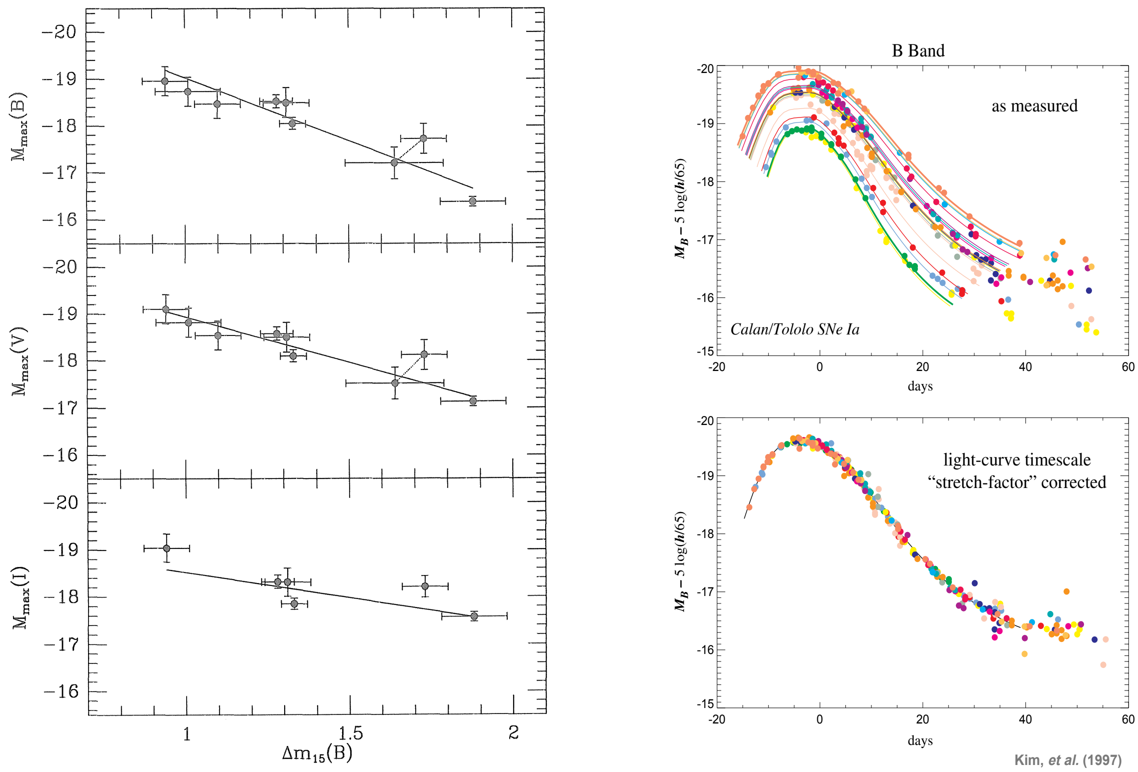

Figure 1: Left panel: Phillips relation, from his 1993 paper. The (apparent) magnitude of type Ia supernovae is correlated with Δm_{15}, the decay of the light curve 15 days after the maximum. Right panel: light curves of a sample of SNe Ia before correction for the Phillips relation (top), and after (bottom).

where d_L is the luminosity distance

$$d_L = (1+z)\frac{H_0^{-1}}{\sqrt{|\Omega_K|}}\text{sinn}\left[\sqrt{|\Omega_K|}\int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + \Omega_{\rm DE}(1+z')^{3(1+w)} + \Omega_R(1+z')^4}}\right] \tag{3}$$

where sinn(x) is equal to $\sin(x)$ (closed universe; $\Omega_K > 0$) or $\sinh(x)$ (open universe; $\Omega_K < 0$) or just x for a flat universe with $\Omega_K = 0$.

This equation can be re-written as

$$m = M + 5\log_{10}(H_0 d_L) - 5\log_{10}(H_0 \times 10\,\text{pc}) \tag{4}$$

or

$$\boxed{m \equiv 5\log_{10}(H_0 d_L) + \mathcal{M}} \tag{5}$$

where the "script-M" factor is defined as

$$\boxed{\mathcal{M} \equiv M - 5\log_{10}\left(\frac{H_0}{\text{Mpc}^{-1}}\right) + 25.} \tag{6}$$

Note that $\mathcal{M}$ is a dummy parameter that captures *two* uncertain quantities: the absolute magnitude (i.e. intrinsic luminosity) of a supernova, M, and the Hubble constant H_0. We typically do

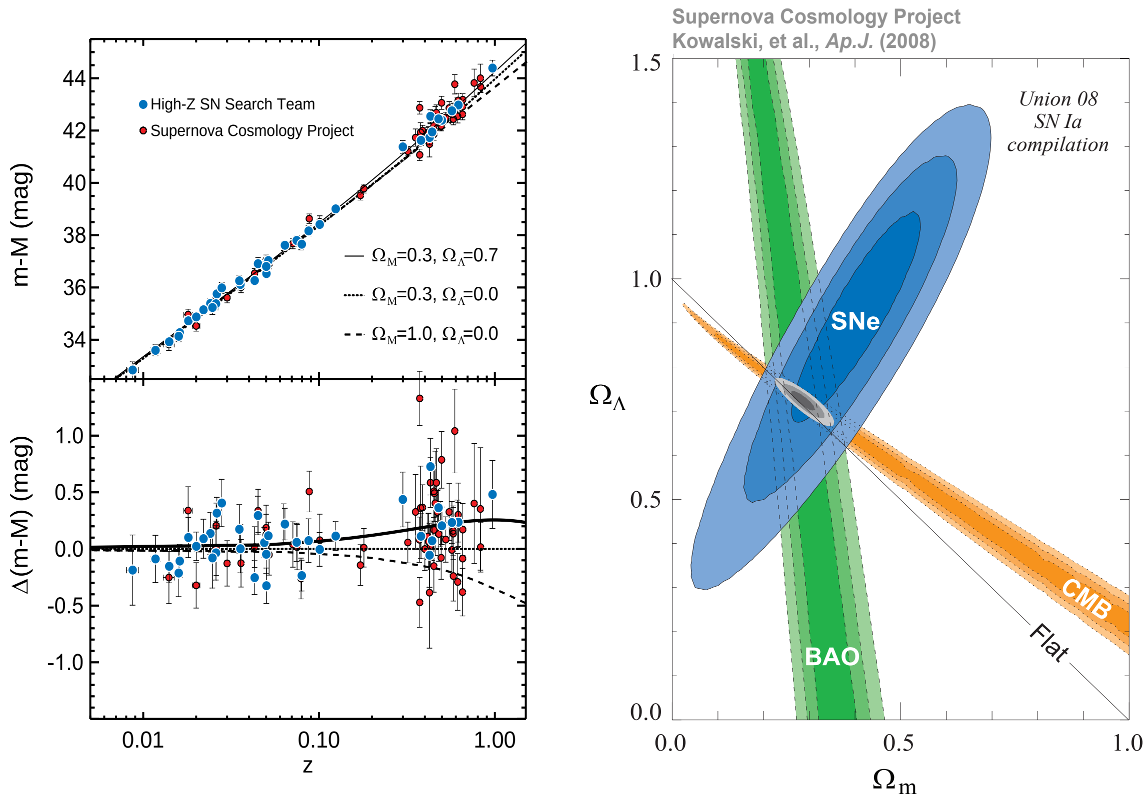

Figure 2: Left panel: combined Hubble diagram data from both SN teams, circa 2003. Right panel: constraints on Ω_M and Ω_Λ from all SN data, circa 2008.

not know $\mathcal{M}$, and we need to marginalize (i.e. integrate) over this parameter in the cosmological analysis.

The situation is now clear: astronomers measure m, which could be peak magnitude of a SN Ia. Then they measure the redshift of a SN. With a bunch of SNe, they can marginalize over the parameter $\mathcal{M}$ and be left with, effectively, measurements of luminosity distance vs. redshift. A plot of either $m(z)$ or $d_L(z)$ is called Hubble diagram.

The discovery of dark energy. Two aforementioned teams, the Highz team and the SCP, published their findings in 1998 and 1999 respectively (the SCP team was a year late since they were notoriously slow with getting papers out; in fact the discoveries had been made around the same time). The results agreed and indicated that more distant SNe are dimmer than would be expected in a matter-only universe. In other words, the universe is speeding up its expansion. This was a watershed event in modern cosmology, and these two papers are among the most cited physics papers ever.

These results have been greatly strengthened since, with many hundreds of SNe Ia currently indicating same results, but with smaller errors, as in the original 1998-9 papers. Meanwhile, other cosmological probes have come in with results confirming the SN results (see the right panel of Fig. 2).

Systematic errors. There are many systematic errors that can creep up in SN observations. At their most pernicious, these systematics may be a cause of the apparent acceleration of the universe (that is, dimming of distant SNe). More generally, they stand in the way of making SNe Ia a more precise tool of cosmology. Here we list a few prominent sources of error, and ways in which they are controlled:

- Extinction: there is dust between SN and dust (remember, some of them are thousands of Mpc away!); is it possible that they appear dimmer simply because of extinction and not dark energy? Well, there are ways to control (and correct for) extinction, basically by looking at SNe in different colors. But also, if extinction were to be responsible for the appearance of dimming, then you'd expect more distant SNe to dim more, without limit. In fact, a "turnover" in the SN Hubble diagram has been observed - basically a signature of universe being matter dominated at high z. This turnover cannot easily be explained by extinction.

- Evolution: is it possible that SNe evolve, so that you are seeing a different population at higher redshift that simply is intrinsically dimmer (violating the assumption of a standard candle)? Well, SNe Ia do not own a "cosmic clock" by which they say "oops, 5 billion years from Big Bang, time to get brighter". Rather, they respond to their local environment, in addition to being ruled by physics of accretion/explosion. So, by observing various signature, in particular in SN spectra, researchers can identify local environmental conditions, and even go so far to compare only like-to-like SNe (resulting, potentially, in several Hubble diagrams, one for each subspecies). First such divisions have been made recently, and indicate that DE results are insensitive to what subspecies of SNe is used to obtain them.

- Typing: is it possible that non-Ia supernovae have crept in the samples used for dark energy analysis? Well this question is easy to answer: type Ia supernova possess a characteristic Silicon 6150 angstrom line (in rest frame), and this line, and a few others, are telltale signs of SNe Ia. This question, however, becomes more relevant for SN surveys which cannot afford to take spectra of all SNe (for example the LSST for most of their SNe, or the Dark Energy Survey for 75% of their SNe). Then one must exercise a lot of care in studying the light curves and trying to establish whether or not a given SN is SN Ia.

- K-corrections: As SNe Ia are observed at larger and larger redshifts, their light is shifted to longer wavelengths. Since astronomical observations are normally made in fixed band passes on Earth, corrections need to be applied to account for the differences caused by the spectrum shifting within these band passes. These corrections take the form of integrating the spectrum of an SN over the relevant band passes, shifting the SN spectrum to the correct redshift, and re-integrating.

- Gravitational lensing: distant SNe are gravitationally lensed by matter along the line of sight, making them magnified or demagnified. This is bad, since of course we use the apparent luminosity of each SN at maximum light to determine how far away it is. Lensing is most effective over large distances; the effect goes roughly as z^2 is non-negligible only for high-z SNe; $z \gtrsim 1.2$. The *mean* magnification is zero (owing to a theorem that the total light in the universe is conserved), but the distribution is skewed, meaning that most SNe get demagnified but occasional ones get strongly magnified. The way to protect against biases due to gravitational lensing is to seek "safety in numbers": simply put, if you collect enough SNe at any given redshift (in practice, ~ 50 SNe per $\Delta z = 0.1$), the effects of gravitational lensing will average down to near zero. See Fig. 3.

Type II SNe. Other types of SNe may potentially also be useful for probing dark energy. Type II SNe (also called the core-collapse SNe) are caused by a different mechanism. Among

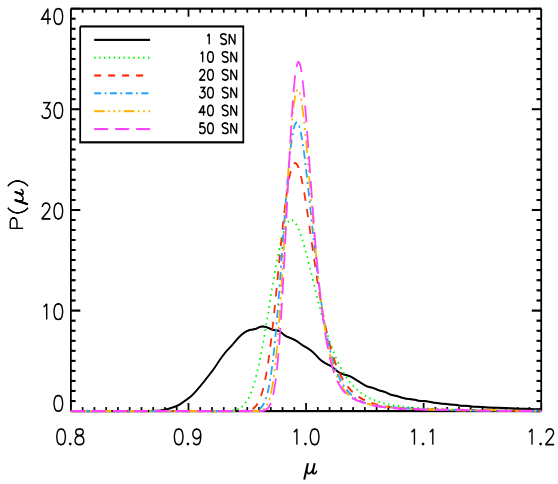

Figure 3: Magnification distribution for lensing of a supernova at $z = 1.5$ in the usual ΛCDM cosmology (black curve). Other curves show how the distribution both narrows and becomes more gaussian as you average over more SNe. Adopted from Holz & Linder (2004).

other things, they do have hydrogen lines in their spectra (unlike SNe Ia) and are more numerous than SNe Ia. Recent work has produced Hubble diagrams using only SNe II.

Other probes of dark energy. In addition to type Ia supernovae, there are several other important probes of dark energy. These probes operate using very different physics, and have very different systematic errors. Therefore, when we have several probes independently indicating the presence dark energy – or measuring one of its properties – we can be assured that the effect is real. At this time, not every one of these probes independently indicates DE, but there are at least three independent lines of evidence for DE.

The principal probes, in addition to SNe Ia, are: baryon acoustic oscillations, weak gravitational lensing, and galaxy cluster abundance. I will now discuss each of those in turn.

Additionally, there are secondary probes of dark energy — ones that might be useful for DE, but are currently not as well developed as the primary probes. I will discuss these at the end.

Baryon acoustic oscillations (BAO). Power spectrum of density perturbations in dark matter, $P(k)$, is mainly sensitive to the density in matter (relative to critical), Ω_M. If we assume a flat universe (from inflationary "prior", and indicated by the CMB observations), then $\Omega_{\rm DE} = 1 - \Omega_M$ and you get the dark energy density, though not much more (for DE) from the broad-band $P(k)$.

However, the *oscillations* in the power spectrum provide much more DE information. Basically, these oscillations correspond to a single scale - the sound horizon - at the redshift at which you make the observations. Therefore, the BAO measure the angular diameter distance, $d_A(z)$, at some redshift z (for example, $z \sim 0.2$ for 2dF survey, and $z \sim 0.35$ for the SDSS). So this is not too different from the quantity that type Ia SNe measure, which is the luminosity distance $d_L(z)$.

Key to successful application of baryon acoustic oscillations are redshift measurements of galaxies in the sample. You need the galaxy redshift in order to know at where radially to "put it", and thus to reconstruct the baryonic oscillations without bias. Another systematic that needs to be understood is the bias of galaxies in your sample (whose clustering you measure) and the underlying dark matter (whose clustering you can predict); it the bias has features in scale, then the systematic errors creep in.

Future surveys that plan to mine this method typically propose measuring redshifts of millions of galaxies, and the goal is to go deep ($z \sim 1$, and beyond) and have wide angular coverage as well.

Coding (that is, computer programming) the power spectrum is very useful. We will consider not just $P(k)$ defined in the inflation lecture, but also the dimensionless form

$$\boxed{\Delta^2(k) \equiv \frac{k^3 P(k)}{2\pi^2}} \tag{7}$$

which one can show to be the contribution to variance of density perturbations per log wavenumber. Without proof, we present the final formula for the power spectrum of dark matter density perturbations in standard FRW cosmology:

$$\boxed{\Delta^2(k) = A\,\frac{4}{25}\,\frac{1}{\Omega_M^2}\left(\frac{k}{k_{\rm piv}}\right)^{n-1}\left(\frac{k}{H_0}\right)^4 [aG(a)]^2\, T^2(k)\, T_{\rm nl}(k)} \tag{8}$$

notice that $\Delta^2 \propto k^{n+3}$, and thus $P(k) \propto k^n$, was predicted by Harrison, Zeldovich and Peebles in the late 1960s (well before inflation *predicted* that $n \simeq 1$!). In this equation:

- A is the normalization of the power spectrum (for the concordance cosmology, $A \simeq 2.4 \times 10^{-9}$)

- $k_{\rm piv}$ is the "pivot" around which we compute the spectral index; for WMAP $k_{\rm piv} = 0.002\,{\rm Mpc}^{-1}$ is used (beware — occasionally $k = 0.05\,h\,{\rm Mpc}^{-1}$ is used too, which is actually closer to the true pivot and anyway changes which amplitude A is appropriate)

- $aG(a)$ is the linear growth of perturbations. Note that in the EdS model $G(a) = 1$ identically and at all times, and in ΛCDM model $G(a)$ at recent times drops, down to the value of ≈ 0.75 at $a = 1$. Note that $aG(a)$ is related to the also much-used growth function $D(a)$, defined as
$$\delta(a) = D(a)\delta(a = 1) \tag{9}$$
via
$$\boxed{D(a) \equiv \frac{aG(a)}{G(1)},} \tag{10}$$
so that $D(1) = 1$ as Eq. (9) requires.

- $T(k)$ is the transfer function: you can use fits (e.g. Hu & Eisenstein, 1997) or else the exact output out of computer codes that solve the coupled Einstein-Boltzmann equations CAMB (`http://cosmologist.info`) or CMBFAST (`http://cmbfast.org`)

- $T_{\rm nl}$ is prescription for a *nonlinear* power spectrum. The nonlinearities are important, for example, today at scales $k \gtrsim 0.2\,h\,{\rm Mpc}^{-1}$. The nonlinearities add power in a complicated way, and are calibrated from numerical simulations, and given to theorists as fitting formulae[1]. The most popular recent fitting formulae are to be found in R. Smith et al. (2003) paper.

[1]That is, fits to simulation results are given in terms of a formula for $T_{\rm nl}(k)$ that contains few universal functions, such as the mass of amplitude fluctuations.

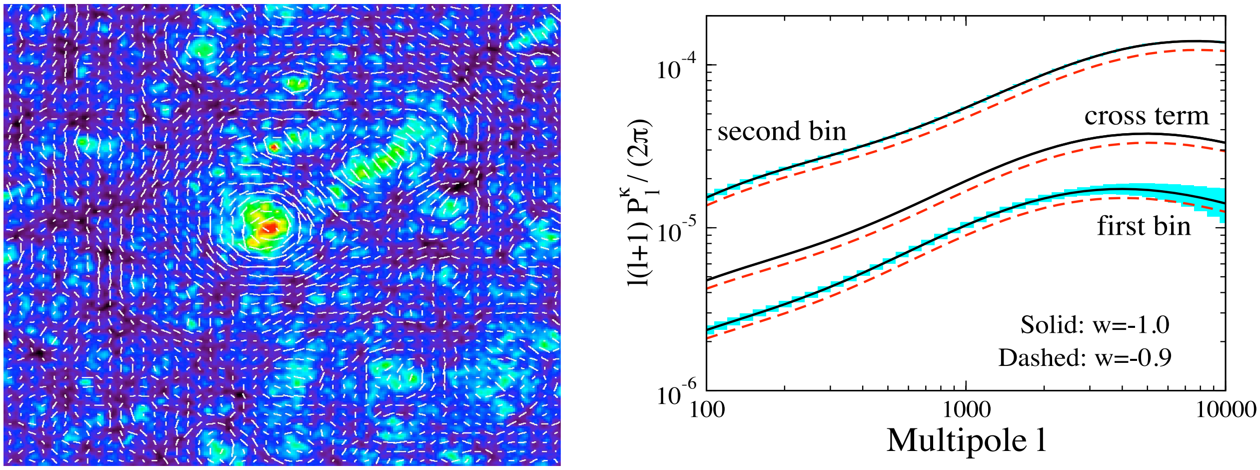

Figure 4: *Left panel:* Cosmic shear field (white ticks) superimposed on the projected mass distribution from a cosmological N-body simulation: overdense regions are bright, underdense regions are dark. Note how the shear field is correlated with the foreground mass distribution. Figure courtesy of T. Hamana. *Right panel:* Cosmic shear angular power spectrum and statistical errors expected for LSST for $w = -1$ and -0.9. For illustration, results are shown for source galaxies in two broad redshift bins, $z_s = 0-1$ (first bin) and $z_s = 1-3$ (second bin); the cross-power spectrum between the two bins (cross term) is shown without the statistical errors.

Finally, it is useful to mention how to calculate the linear growth function in a cosmology with an arbitrary dark energy equation of state $w(z)$. Basically you can re-write the usual equation for the growth of fluctuations

$$\ddot{\delta}_k + 2H\dot{\delta}_k - 4\pi G\rho_M\delta_k = 0 \tag{11}$$

(where δ_k is the Fourier amplitude of the density perturbation — this equation equally holds in real space because it is linear in δ) to get

$$\boxed{2\frac{d^2g}{d\ln a^2} + [5 - 3w(a)\Omega_{\rm DE}(a)]\,\frac{dg}{d\ln a} + 3\,[1 - w(a)]\,\Omega_{\rm DE}(a)g = 0} \tag{12}$$

where $g(a) \equiv D(a)/a$ is the growth rate relative to that in an EdS universe.

Equations (7) and (12) enable you to calculate the power spectrum of density perturbations at any redshift and any scale, and for any cosmological model. This is useful!

Weak gravitational lensing. The gravitational bending of light by structures in the Universe distorts or shears the images of distant galaxies; see the left panel of Fig. 4. This distortion allows the distribution of dark matter and its evolution with time to be measured, thereby probing the influence of dark energy on the growth of structure (for detailed review, see e.g. Bartelmann & Schneider 2001).

We work in the Newtonian Gauge, where the perturbed Friedmann-Robertson-Walker metric reads

$$ds^2 = -\,(1 + 2\Phi)\,dt^2 + a^2(t)\,(1 - 2\Phi)\,\left[d\chi^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{13}$$

where we have set $c = 1$, χ is the radial distance, Φ is the gravitational potential, and $k = 1$, 0, -1 for closed, flat and open geometry respectively. We also use the comoving distance r which is defined as

$$r(\chi) = \begin{cases} (-K)^{-1/2}\sinh[(-K)^{1/2}\chi], & \text{if } \Omega_{\rm TOT} < 1; \\ \chi, & \text{if } \Omega_{\rm TOT} = 1; \\ K^{-1/2}\sin(K^{1/2}\chi), & \text{if } \Omega_{\rm TOT} > 1. \end{cases} \tag{14}$$

where K is the curvature, $\Omega_{\rm TOT}$ is the total energy density relative to critical, and $K = (\Omega_{\rm TOT} - 1)H_0^2$. (Note that this exactly agrees with Eq. (3)).

Gravitational lensing produces distortions of images of background galaxies. These distortions can be described as mapping between the source plane (S) and image plane (I)

$$\delta x_i^S = A_{ij}\delta x_j^I \tag{15}$$

where $\delta\mathbf{x}$ are the displacement vectors in the two planes and A is the distortion matrix

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{16}$$

The deformation is described by the convergence κ and complex shear (γ_1, γ_2); the total shear is defined as $|\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}$. We are interested in the weak lensing limit, where κ, $|\gamma| \ll 1$. Magnification is also given in terms of κ and $\gamma_{1,2}$ as

$$\mu = \frac{1}{|1-\kappa|^2 - |\gamma|^2} \approx 1 + 2\kappa + O(\kappa^2, \gamma^2) \tag{17}$$

where the second approximate relation holds again in the weak lensing limit.

But how do you theoretically predict convergence and shear, given some source galaxies and the foreground distribution on the sky? The convergence in any particular direction on the sky $\hat{\mathbf{n}}$ is given by the integral along the line-of-sight

$$\kappa(\hat{\mathbf{n}}, \chi) = \int_0^{\chi} W(\chi')\,\delta(\chi')\,d\chi' \tag{18}$$

where δ is the relative perturbation in matter energy density and

$$W(\chi) = \frac{3}{2}\,\Omega_M\,H_0^2\,g(\chi)\,(1+z) \tag{19}$$

is the weight function that "assigns" lensing efficiency to foreground galaxies. Furthermore

$$g(\chi) = r(\chi)\int_{\chi}^{\infty} d\chi' n(\chi')\frac{r(\chi'-\chi)}{r(\chi')} \longrightarrow \frac{r(\chi)r(\chi_s-\chi)}{r(\chi_s)} \tag{20}$$

where $n(\chi)$ is the distribution of source galaxies in redshift (normalized so that $\int dz\, n(z) = 1$) and the expression after the arrow holds only if all sources are at a single redshift z_s. The function g peaks about halfway between the observer and the source (that is, at $\chi \sim \chi_s/2$. Therefore, *the most efficient lenses lie about half-way between us and the source galaxies.*

The statistical signal due to gravitational lensing by large-scale structure is termed "cosmic shear." The cosmic shear field at a point in the sky is estimated by locally averaging the shapes of large numbers of distant galaxies. The primary statistical measure of the cosmic shear is the shear angular power spectrum measured as a function of source-galaxy redshift z_s. (Additional information is obtained by measuring the correlations between shears at different redshifts or with foreground lensing galaxies.)

You can typically measure galaxy shear (from the galaxy shapes), but you can theoretically predict the convergence. Fortunately, in the weak lensing limit, convergence and shear are equal[2].

The convergence can be transformed into multipole space (e.g. Bartelmann & Schneider 2001)

$$\kappa_{lm} = \int d\hat{\mathbf{n}}\, \kappa(\hat{\mathbf{n}}, \chi)\, Y_{lm}^{*}(\hat{\mathbf{n}}), \tag{21}$$

and the power spectrum is defined as the two-point correlation function (of convergence, in this case)

$$\langle \kappa_{\ell m} \kappa_{\ell' m'} \rangle = \delta_{\ell\ell'}\, \delta_{mm'}\, P_\ell^\kappa. \tag{22}$$

The angular power spectrum is

$$\boxed{P_\ell^\gamma(z_s) \simeq P_\ell^\kappa(z_s) = \int_0^{z_s} \frac{dz}{H(z) d_A^2(z)} W(z)^2 P\left(k = \frac{\ell}{d_A(z)}; z\right)}, \tag{23}$$

where ℓ denotes the angular multipole, the weight function $W(z)$ is the efficiency for lensing a population of source galaxies and is determined by the distance distributions of the source and lens galaxies, and $P(k, z)$ is the usual power spectrum of density perturbations. Notice *integral* along the line of sight — essentially, weak lensing

The dark-energy sensitivity of the shear angular power spectrum comes from two factors:

- *geometry*—the Hubble parameter, the angular-diameter distance, and the weight function= $W(z)$; and
- *growth of structure*—through the redshift evolution of the power spectrum of density perturbations (really, function $G(a)$ from Eq. (8)).

It is also possible to use the *three*-point correlation function of cosmic shear is also sensitive to dark energy (power spectrum is the *two*-point correlation function.

The *statistical* uncertainty in measuring the shear power spectrum on large scales is

$$\Delta P_\ell^\gamma = \sqrt{\frac{2}{(2\ell + 1) f_{\rm sky}}} \left[P_\ell^\gamma + \frac{\sigma^2(\gamma_i)}{n_{\rm eff}} \right] \quad , \tag{24}$$

where $f_{\rm sky}$ is the fraction of sky area covered by the survey (that is, $f_{\rm sky} = 0.5$ for half-sky, etc), $\sigma^2(\gamma_i)$ is the variance in a single component of the (two-component) shear (this number is ~ 0.2 for typical measurements), and $n_{\rm eff}$ is the effective number density per steradian of galaxies with well-measured shapes.

The first term in brackets dominates on *large scales*, and comes from *cosmic variance* of the mass distribution. The second term dominates on *small scales*, and represents the shot-noise from both the variance in galaxy ellipticities ("shape noise") combined with a finite number of galaxies (hence inverse proportionality to $n_{\rm eff}$).

The right panel of Fig. 4 shows the dependence on the dark energy of the shear power spectrum and an indication of the statistical errors expected for a survey such as LSST, assuming a survey area of 15,000 sq. deg. and effective source galaxy density of $n_{\rm eff} = 30$ galaxies per sq. arcmin. Current survey cover more modest hundreds of square degrees, with a comparable or slightly lower galaxy density. Note that the proportionality of errors to $f_{\rm sky}^{-1/2}$ means that large sky coverage is at a premium.

[2]That is, $\kappa \simeq \sqrt{\gamma_2^2 + \gamma_2^2}$. Corrections to this equality are of order κ^2 and γ^2, and are therefore small since $|\kappa|, |\gamma| \lesssim 0.01$.

Clusters of galaxies. Galaxy clusters are the largest virialized objects in the Universe. Therefore, not only can they be observed, but also their number density can be *predicted* quite reliably, both analytically, and from numerical simulations. Comparing these predictions to large-area cluster surveys that extend to high redshift ($z \gtrsim 1$) can provide precise constraints on the cosmic expansion history.

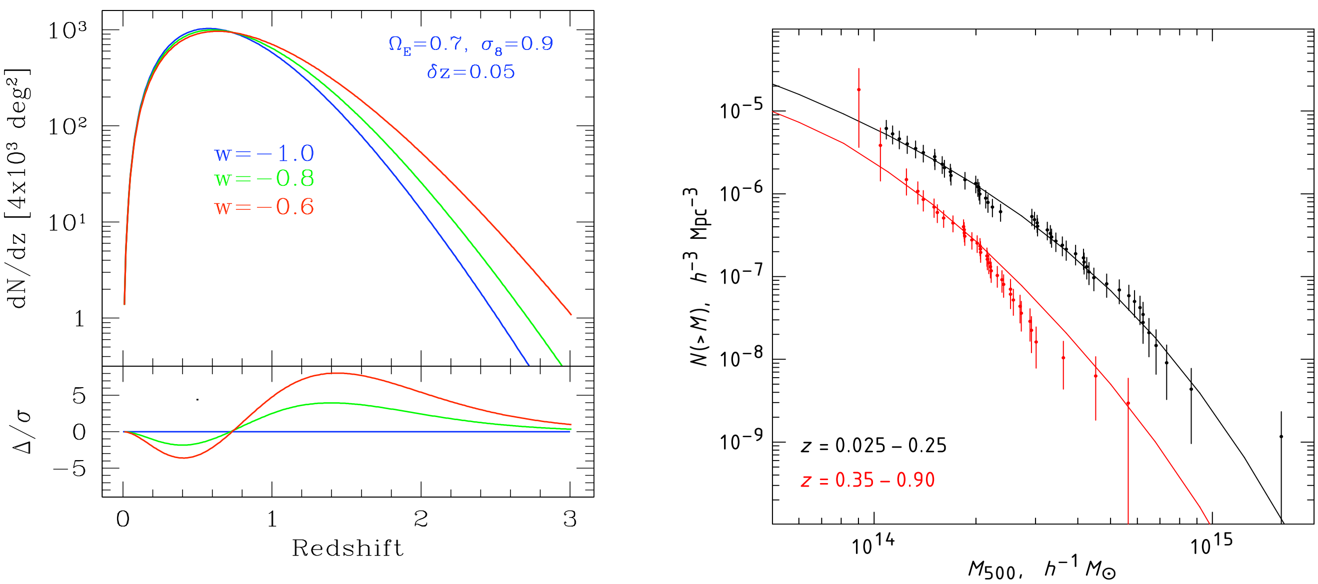

Figure 5: *Left panel:* Predicted cluster counts for a survey covering 4,000 sq. deg. that is sensitive to halos more massive than $2 \times 10^{14} M_{\odot}$, for 3 flat cosmological models with fixed $\Omega_{\rm M} = 0.3$ and $\sigma_8 = 0.9$. Lower panel shows fractional differences between the models in terms of the estimated Poisson errors. From Mohr (2005). *Right panel:* Measured mass function – $n(z, M_{\rm min}(z))$ – in our notation – from the 400 square degree survey of ROSAT clusters followed up by Chandra. Adopted from Vikhlinin et al. (2009).

The absolute number of clusters in a survey of size $\Omega_{\rm survey}$ centered at redshift z and in the shell of thickness Δz is given by

$$N(z, \Delta z) = \Omega_{\rm survey} \int_{z-\Delta z/2}^{z+\Delta z/2} n(z, M_{\rm min}(z)) \, \frac{dV(z)}{d\Omega \, dz} \, dz. \tag{25}$$

where $M_{\rm min}$ is the minimal mass of clusters in the survey (it's usually of order $10^{14} M_{\odot}$). Note that knowledge of the minimal mass is extremely important, since the "mass function" $n(z, M_{\rm min}(z))$ is *exponentially* decreasing with M, so that most of the contribution comes from a small range of masses just above $M_{\rm min}$. The mass function is key to theoretical predictions, and it is usually obtained from a combination of analytic and numerical results; famous mass functions used in cosmology are the Press-Schechter mass function that dates back to 1970s, or the more recent Sheth-Tormen mass function (these are basically just different fitting function). The volume element can easily be related to comoving distance $r(z)$ and the expansion rate $H(z)$ via

$$\frac{dV(z)}{d\Omega \, dz} = \frac{r^2(z)}{H(z)} \tag{26}$$

and it is basically known exactly for a fixed cosmological model, with no theoretical uncertainty (unlike the mass function which is usually known to a few percent at best, at a given M and z).

The sensitivity of cluster counts to dark energy arises – as in the case of weak lensing – from two factors:

- *geometry*, the term $dV(z)/(d\Omega\, dz)$ in Eq. (25) is the comoving volume element
- *growth of structure*, $n(z, M_{\min}(z))$ depends on the evolution of density perturbations, cf. Eq. (11).

This last point is worth emphasizing further. Fitting functions for the cluster mass function (e.g. Press-Schechter) relate it to the primordial spectrum of density perturbations. The mass function's near-exponential dependence upon the power spectrum is at the root of the power of clusters to probe dark energy. In particular, the mass function explicitly depends on the *amplitude of mass fluctuations* smoothed on some scale R

$$\boxed{\sigma^2(R) = \int_0^\infty \Delta^2(k) \left(\frac{3j_1(kR)}{kR}\right)^2 d\ln k} \tag{27}$$

where R is traditionally taken to be $\sim 8\,\mathrm{h}^{-1}\mathrm{Mpc}$, the characteristic size of galaxy clusters. Here of course $\Delta^2(k)$ is our dear power spectrum from Eq. (7).

Fig. 5 shows the sensitivity to the dark energy equation-of-state parameter of the expected cluster counts for the South Pole Telescope and the Dark Energy Survey. At modest redshift, $z < 0.6$, the differences are dominated by the volume element; at higher redshift, the counts are most sensitive to the growth rate of perturbations.

Summary of dark energy constraints. Figure 6, adopted from Vikhlinin et al. (2009), summarizes constraints in the $\Omega_{\rm DE} - w$ plane (assuming a flat universe) from clusters, CMB, BAO and SNe Ia.

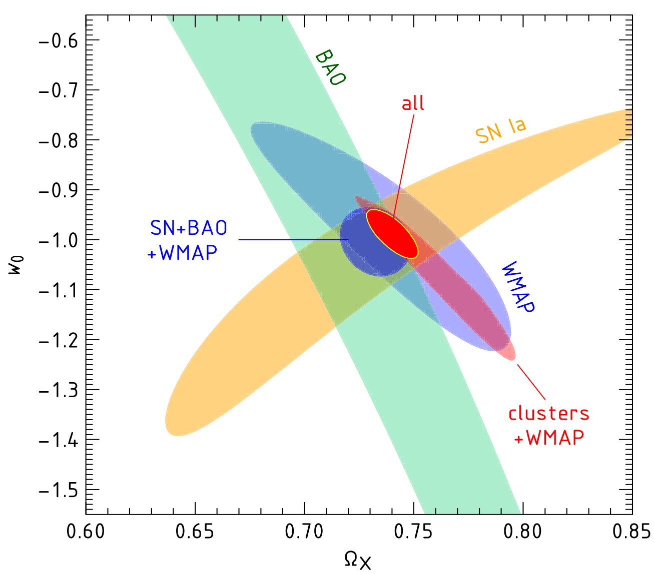

Figure 6: Summary of constraints in the $\Omega_{\rm DE} - w$ plane (assuming a flat universe) from clusters, CMB, BAO and SNe Ia. Note that the dark energy density to critical today, $\Omega_{\rm DE}$, is labeled as Ω_X in this graph. Adopted from Vikhlinin et al. (2009).

The four principal probes of DE: systematics summary. In Table 1 we list the principal strengths and weaknesses of the four principal probes of DE. Control of systematic

Table 1: Comparison of dark energy probes, adopted from Frieman, Turner and Huterer [2008]. CDM refers to Cold Dark Matter paradigm, FoM is the Figure-of-Merit for dark energy surveys defined in the Dark Energy Task Force (DETF) report, while SZ refers to Sunyaev-Zeldovich effect.

Method	Strengths	Weaknesses	Systematics
WL	growth+geometry, Large FoM	CDM assumptions	Shear systematics, Photo-z
SN	pure geometry, mature	complex physics	evolution, dust extinction
BAO	pure geometry, low systematics	coarse-grained information	bias, non-linearity, redshift distortions
CL	growth+geometry, X-ray+SZ+optical	CDM assumptions	mass-observable, selection function

errors — observational, instrumental and theoretical — is crucial for these probes to realize their intrinsic power in constraining dark energy.

Role of the CMB. While the CMB provides precise cosmological constraints, by itself it has little power to probe dark energy. The reason is simple: the CMB provides a single snapshot of the Universe at a time when dark energy contributed but a tiny part of the total energy density (a part in 10^9 if dark energy is the vacuum energy, or when $w = -1$). Nonetheless, the CMB plays a critical supporting role by determining other cosmological parameters, such as the spatial curvature and matter density, to high precision, thereby considerably strengthening the power of the methods discussed above. Essentially, what you get from the CMB is a *single* measurement of the angular diameter distance to recombination, $d_L(z \approx 1000)$ — therefore you get a *single* very accurate measurement of the parameters Ω_M, $\Omega_{\rm DE}$ (if you do not assume a flat universe), and w (or $w(z)$ if you don't assume that the equation of state is constant). Therefore, while from the CMB alone there is degeneracy between the DE parameters, CMB is very useful to *break* degeneracy from other cosmological probes (see Frieman, Huterer, Linder & Turner 2005 for details).

Data from the Planck CMB mission, launched earlier in 2009, will complement those from dark energy surveys. If the Hubble parameter can be directly measured to better than a few percent, in combination with Planck it would also provide powerful dark energy constraints.

Secondary probes. There are a number of secondary probes of dark energy. Here we review a few of them. You are welcome to learn more about them at your leisure if you are interested.

- The Integrated Sachs-Wolfe (ISW) effect provided a confirmation of cosmic acceleration (e.g. Scranton et al. 2003 from the SDSS). ISW impacts the large-angle structure of the CMB anisotropy, but low-ℓ multipoles are subject to large cosmic variance, limiting their power. Nevertheless, ISW is of interest because it may be able to show the imprint of large-scale dark-energy perturbations (Hu & Scranton 2004).

- Gravitational radiation from inspiraling binary neutron stars or black holes can serve as "standard sirens" to measure absolute distances (Holz & Hughes 2005). If their redshifts can be determined, then they could be used to probe dark energy through the Hubble diagram (Dalal et al. 2006).

- Long-duration gamma-ray bursts have been proposed as standardizable candles (e.g. Schaefer 2003), but their utility as cosmological distance indicators that could be competitive with or complementary to SNe Ia has yet to be established. The angular size-redshift relation for double radio galaxies has also been used to derive cosmological constraints that are consistent with dark energy.

- The optical depth for strong gravitational lensing (multiple imaging) of QSOs or radio sources has been proposed and used to provide independent evidence for dark energy, though these measurements depend on modeling the density profiles of lens galaxies.

- The Sandage-Loeb effect (Sandage 1962, Loeb 1998), the redshift change of an object measured using extremely high-resolution spectroscopy over a period of 10 years or more, will some day be useful in constraining the expansion history at higher redshift, $2 \lesssim z \lesssim 5$ (Corasaniti, Huterer & Melchiorri 2005).

- Polarization measurements from distant galaxy clusters in principle provide a sensitive probe of the growth function and hence dark energy (Cooray, Huterer & Baumann 2004).

- The relative ages of galaxies at different redshifts, if they can be determined reliably, provide a measurement of dz/dt and, from

$$t(z) = \int_0^{t(z)} dt' = \int_z^{\infty} \frac{dz'}{(1+z')H(z')}, \tag{28}$$

  measure the expansion history directly (Jimenez & Loeb 1998). Measurements of the abundance of lensed arcs in galaxy clusters, if calibrated accurately, provide a probe of dark energy.

Surveys overview. Many ambitious surveys are planned for the next decade or so. These include ground-based and space-based surveys, and include the four principal cosmological probes (SNe Ia, BAO, weak lensing and clusters) and other. These surveys, their principal specifications and main probes with which they are probing dark energy, are given in Table 2.

Table 2: Dark energy projects proposed or under construction. Stage refers to the DETF time-scale classification. Adopted from Frieman, Turner & Huterer (2008).

Survey	Description	Probes	Stage
Ground-based:			
ACT	SZE, 6-m	CL	II
APEX	SZE, 12-m	CL	II
SPT	SZE, 10-m	CL	II
VST	Optical imaging, 2.6-m	BAO,CL,WL	II
Pan-STARRS 1(4)	Optical imaging, 1.8-m($\times$4)	All	II(III)
DES	Optical imaging, 4-m	All	III
Hyper Suprime-Cam	Optical imaging, 8-m	WL,CL,BAO	III
ALPACA	Optical imaging, 8-m	SN, BAO, CL	III
LSST	Optical imaging, 6.8-m	All	IV
AAT WiggleZ	Spectroscopy, 4-m	BAO	II
HETDEX	Spectroscopy, 9.2-m	BAO	III
PAU	Multi-filter imaging, 2-3-m	BAO	III
SDSS BOSS	Spectroscopy, 2.5-m	BAO	III
WFMOS	Spectroscopy, 8-m	BAO	III
HSHS	21-cm radio telescope	BAO	III
SKA	km^2 radio telescope	BAO, WL	IV
Space-based:			
JDEM Candidates			
ADEPT	Spectroscopy	BAO, SN	IV
DESTINY	Grism spectrophotometry	SN	IV
SNAP	Optical+NIR+spectro	All	IV
Proposed ESA Missions			
DUNE	Optical imaging	WL, BAO, CL	
SPACE	Spectroscopy	BAO	
eROSITA	X-ray	CL	
CMB Space Probe			
Planck	SZE	CL	
Beyond Einstein Probe			
Constellation-X	X-ray	CL	IV

Lecture II: Descriptions of Dark Energy

Dark Energy: review of important facts. To review the last part of my lecture which was not in these notes, here are the five important things you should know about dark energy:

1. Dark energy has negative pressure. It can be described with its energy density relative to critical today $\Omega_{\rm DE}$, and equation of state $w \equiv p_{\rm DE}/\rho_{\rm DE}$; the cosmological constant (or vacuum energy) has $w = -1$ precisely and at all times. More general explanations for dark energy may have constant or time dependent equation of state. Assuming constant w, current constraints roughly give $w \approx -1 \pm 0.2$. Measuring the equation of state (and its time dependence) may help understand the nature of dark energy, and is a key goal of modern cosmology.

2. Dark energy suppresses the growth of density perturbation: whenever dark energy dominates, structures do not grow.

3. Dark energy comes to dominate the density of the universe only recently: at $z \gtrsim 2$, dark energy is negligible. This has to be so not only from the observations, but also from the fact that, *if* dark energy dominated throughout the history of the universe, we would have had no galaxies and clusters today.

4. Dark energy is spatially smooth. It affects both the geometry (that is, distances in the universe) and the growth of structure (that is, clustering and abundance of galaxies and clusters of galaxies).

5. Dark energy bring two puzzling problems with it. The first one, the *coincidence problem*, is the fact that the dark energy to dark matter ratio today is $O(1)$, and not huge or near zero; this is not necessarily a problem depending on who you talk to. The second problem is one of the most famous puzzles in physics, and is called the *cosmological constant problem*: the fact that the dark energy density, $\rho_{\rm DE} \sim {m_{\rm Pl}}^2 H_0^2$, is about **120 orders of magnitude** smaller than the natural Planck scale, $\rho_P \sim {m_{\rm Pl}}^4$.

Parametrizations of dark energy: Introduction. The absence of a consensus model for cosmic acceleration presents a challenge in trying to connect theory with observations. For dark energy, the equation-of-state parameter w provides a useful phenomenological description Because it is the ratio of pressure to energy density, it is also closely connected to the underlying physics. However, w is not fundamentally a function of redshift, and if cosmic acceleration is due to new gravitational physics, the motivation for a description in terms of w disappears. On the practical side, determining a free function is more difficult than measuring parameters. We now review a variety of formalisms that have been used to describe and constrain dark energy.

First, let us recall some basics. From continuity equation

$$\dot\rho + 3H(p+\rho) = 0, \quad \text{or} \tag{29}$$

$$\frac{d\ln\rho}{d\ln a} = -3(1+w), \tag{30}$$

we can calculate the dark energy density

$$\rho_{\rm DE}(a) = \rho_{\rm DE,0}\exp\left(-3\int_1^a (1+w(a'))d\ln a'\right) \tag{31}$$

$$\rho_{\rm DE}(z) = (1-\Omega_M)\rho_{\rm crit,0}\exp\left(3\int_0^z (1+w(z'))d\ln(1+z')\right) \tag{32}$$

where in the second line we assumed a flat universe in which $\Omega_{\rm DE} = 1-\Omega_M$. Let us also write possibly the most useful form

$$\boxed{\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = \exp\left(3\int_0^z (1+w(z'))d\ln(1+z')\right).} \tag{33}$$

Parametrizations. The simplest parameterization of dark energy is

$$w = \text{const.} \tag{34}$$

This form fully describes vacuum energy ($w = -1$) or topological defects ($w = -N/3$ with N an integer dimension of the defect – 0 for monopoles, 1 for strings, 2 for domain walls). Together with $\Omega_{\rm DE}$ and $\Omega_{\rm M}$, w provides a 3-parameter description of the dark-energy sector (2 parameters if flatness is assumed). However, it does not describe scalar field or modified gravity models which generically have a *time-varying* w.

A number of two-parameter descriptions of w have been explored in the literature, e.g.,

$$w(z) = w_0 + w'z \tag{35}$$

$$w(z) = w_0 + b\ln(1+z). \tag{36}$$

For low redshift they are all essentially equivalent, but for large z, some lead to unrealistic behavior, e.g., $w \ll -1$ or $\gg 1$. The energy density with these parametrizations can easily be worked out starting with the continuity equation $\dot\rho + 3H(p+\rho) = 0$; for example, for the first parametrization (with w') we have

$$\rho(z) = \rho_0(1+z)^{3(1+w_0-w')}e^{3w'z}. \tag{37}$$

The parametrization

$$\boxed{w(a) = w_0 + w_a(1-a) = w_0 + w_a z/(1+z)} \tag{38}$$

avoids this problem and leads to the most commonly used description of dark energy, namely $(\Omega_{\rm DE}, \Omega_{\rm M}, w_0, w_a)$. The energy density is then

$$\rho(a) = \rho_0 a^{-3(1+w_0+w_a)}e^{-3(1-a)w_a} \tag{39}$$

More general expressions have been proposed, for example, Padé approximants or the transition between two asymptotic values w_0 (at $z \to 0$) and w_f (at $z \to \infty$):

$$w(z) = w_0 + \frac{(w_f - w_0)}{1+\exp[(z-z_t)/\Delta]}. \tag{40}$$

However one problem with introducing more parameters are a) that additional parameters are very hard to measure, and b) that the parametrizations are *ad hoc*, and not well motivated from either theory or measurements' point of view.

Pivot w and pivot redshift. The two-parameter descriptions of $w(z)$ that are linear in the parameters entail the existence of a "pivot" redshift z_p at which the measurements of the two parameters are uncorrelated and the error in $w_p \equiv w(z_p)$ reaches a minimum; see the left panel of Fig. 7.

Here is how you compute the pivot redshift (assuming, for example, the (w_0, w_a) parametrization). Note that we are looking for the equation of state of the form

$$w(a) = \boxed{w_p + (a_p - a)w_a} \tag{41}$$

$$= w_p + \frac{z_p}{1+z_p} w_a \tag{42}$$

Then you proceed as follows:

- Compute the covariance matrix for all parameters (in the Fisher matrix formalism, this is just the inverse of the Fisher matrix). For two parameters p_i and p_j, the covariance is defined as
$$\mathrm{Cov}(p_i, p_j) \equiv \langle (p_i - \bar{p_i})(p_j - \bar{p_j}) \rangle \tag{43}$$
where this is just the likelihood-weighted average over all values of (p_i, p_j) and the overbar indicates the *mean* value of that parameter from the data.

- Marginalize (that is, integrate the likelihood) over all other parameters, leaving only the 2×2 matrix on w_0 and w_a

- Assuming $w_p = w_0 - (a_p - 1)w_a$, we have $0 = \mathrm{Cov}(w_p, w_a) = \mathrm{Cov}(w_0, w_a) - (a_p - 1)\mathrm{Cov}(w_a, w_a)$, implying that $a_p - 1 = \mathrm{Cov}(w_0, w_a)/\mathrm{Cov}(w_a, w_a)$. Therefore
$$a_p = 1 + \mathrm{Cov}(w_0, w_a)/\mathrm{Cov}(w_a, w_a) \tag{44}$$

- Finally, you can compute the error in w_p as (using in second line Eq. 44)
$$\mathrm{Cov}(w_p, w_p) = \mathrm{Cov}(w_0, w_0) - 2(a_p - 1)\mathrm{Cov}(w_0, w_a) + (a_p - 1)^2\mathrm{Cov}(w_a, w_a) \tag{45}$$
$$= \mathrm{Cov}(w_0, w_0) - \frac{\mathrm{Cov}(w_0, w_a)^2}{\mathrm{Cov}(w_a, w_a)} \tag{46}$$

The redshift of this sweet spot varies with the cosmological probe and survey specifications; for example, for current SN Ia surveys $z_p \approx 0.25$, while for weak lensing and baryon acoustic oscillations $z_p \sim 0.5$-0.7 because of the way in which dark energy enters the physical quantities featured in these probes.

Note that forecast constraints for a particular experiment on w_p are numerically equivalent to constraints one would derive on constant w. This is because almost all information is concentrated in w_p (which was optimally chosen precisely using that criterion).

Direct reconstruction. Another approach is to directly invert the redshift-distance relation $r(z)$ measured from SN data to obtain the redshift dependence of $w(z)$ in terms of the coordinate distance $r(z)$ that is measured, for example, by SNe Ia[3]. Note that, since $r(z) = \int dz/H(z)$, we have

$$H(z) = \frac{1}{(dr/dz)} \tag{47}$$

[3]SNe Ia, of course, measure the luminosity distance $d_L(z)$, but in a flat universe the two are only offset by a factor of $1 + z$.

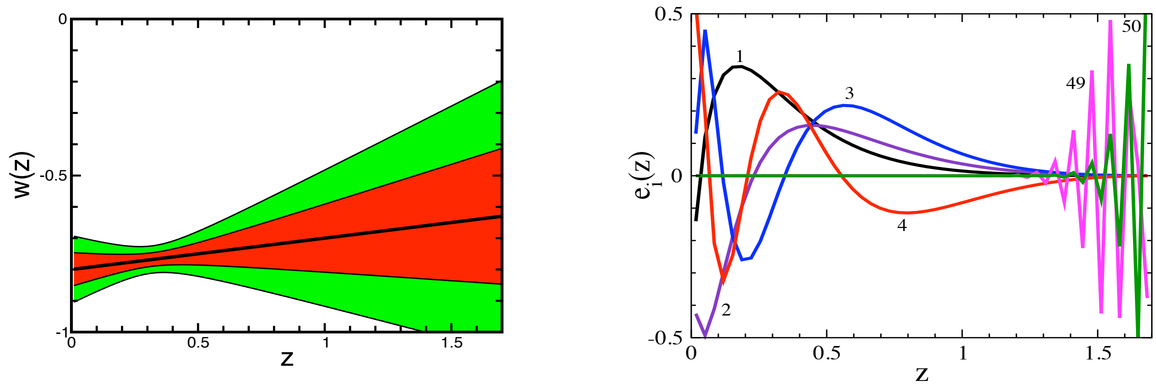

Figure 7: *Left panel:* Example of forecast constraints on $w(z)$, assuming $w(z) = w_0 + w'z$. The "pivot" redshift, $z_p \simeq 0.3$, is where $w(z)$ is best determined. *Right panel:* The four best-determined (labelled $1-4$) and two worst-determined (labelled $49, 50$) principal components of $w(z)$ for a future SN Ia survey such as SNAP, with several thousand SNe in the redshift range $z = 0$ to $z = 1.7$.

Furthermore, it is almost as easy to reconstruct the dark energy density; since

$$H^2 = H_0^2 \left[\Omega_M (1+z)^3 + (1-\Omega_M)\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)}\right] \tag{48}$$

then

$$\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = \frac{1}{1-\Omega_M}\left[\frac{H^2}{H_0^2} - \Omega_M(1+z)^3\right] \tag{49}$$

or

$$\boxed{\frac{\rho_{\rm DE}(z)}{\rho_{\rm DE}(0)} = \frac{1}{1-\Omega_M}\left[\frac{1}{(d(H_0 r)/dz)^2} - \Omega_M(1+z)^3\right]} \tag{50}$$

As an exercise, you can show that the equation of state can also be reconstructed as

$$\boxed{1 + w(z) = \frac{1+z}{3}\,\frac{3H_0^2\Omega_{\rm M}(1+z)^2 + 2(d^2r/dz^2)/(dr/dz)^3}{H_0^2\Omega_{\rm M}(1+z)^3 - (dr/dz)^{-2}}\ .} \tag{51}$$

Note that H_0 only enters in the combination $\Omega_M H_0^2$ which is measured by the CMB extremely well. Therefore, the main sources of error should be in the derivatives of the comoving distance — more about that in a second.

Assuming that dark energy is due to a single rolling scalar field, the scalar potential can also be reconstructed,

$$V[\phi(z)] \quad = \quad \frac{1}{8\pi G}\left[\frac{3}{(dr/dz)^2} + (1+z)\frac{d^2r/dz^2}{(dr/dz)^3}\right] - \frac{3\Omega_{\rm M}H_0^2(1+z)^3}{16\pi G} \tag{52}$$

Direct reconstruction is the only approach that is truly model-independent. However, it comes at a price – taking derivatives of noisy data. In practice, one must fit the distance data with a smooth function — e.g., a polynomial, Padé approximant, or spline with tension; for example,

$$H_0 r(z) = \frac{z + a_2 z^2}{1 + b_1 z + b_2 z^2} \tag{53}$$

(which is a Padé approximant fit, made so that $H_0 r \to z$ when $z \ll 1$).

The fitting process introduces systematic biases – it is simply very hard to take second derivative of noisy data (if on a given day you want to be frustrated a little extra, try it!). While a variety of methods have been pursued it appears that direct reconstruction is too challenging and not robust even with SN Ia data of excellent quality. Although the expression for $\rho_{\rm DE}(z)$ involves only first derivatives of $r(z)$, it contains little information about the nature of dark energy. For a review of dark energy reconstruction and related issues, see Sahni & Starobinsky 2006.

Principal components. The most general, and the only really non-parametric method, to reconstruct properties of DE is that of *principal components* of dark energy. The cosmological function that we are trying to determine — $w(z)$, $\rho_{\rm DE}(z)$, or $H(z)$ — can be expanded in terms of principal components, a set of functions that are uncorrelated and orthogonal by construction. In this approach, *the data determine* which aspects of a cosmological function — which linear combinations of $w(z_i)$, in this case — are measured best.

For example, suppose we parametrize $w(z)$ in terms of piecewise constant values w_i ($i = 1, \ldots, N$), each defined over a small redshift range (z_i, $z_i + \Delta z$). Then you get the principal components as follows

- Create the covariance matrix of all parameters, including the w_i
- Marginalize over all non-w parameters, and be left with the $N \times N$ covariance matrix in the w_i — call it C; also compute its inverse, C^{-1}
- Diagonalize the inverse covariance: $C^{-1} = W^T \Lambda W$, where Λ is diagonal and W is orthogonal
- Rows of W^T are the eigenvectors e_i — the "principal components" — while the elements of Λ give the accuracies in how well each PC is measured. The eigenvectors are of course orthonormal so that $\int e_i(z) e_j(z) dz = \delta_{ij}$.

In other words

$$\boxed{1 + w(z) = \sum_{i=1}^{N} \alpha_i \, e_i(z) \ ,} \tag{54}$$

with

$$\sigma(\alpha_i) = \lambda_i^{-1/2} \tag{55}$$

(note that we wisely added unity to w in its expansion, so that the fiducial expansion parameters are zero around ΛCDM).

In the limit of small Δz this recovers the shape of an arbitrary dark energy history (in practice, $N \gtrsim 20$ is sufficient), but the estimates of the w_i from a given dark energy probe will be very noisy for large N. Principal Component Analysis extracts from those noisy estimates the best-measured features of $w(z)$.

The coefficients α_i, which can be computed via the orthonormality condition

$$\alpha_i = \int (1 + w(z)) e_i(z) dz \tag{56}$$

Examples of these components are shown for a future SN survey in the right panel of Fig. 7.

There are multiple advantages of using the PCs of dark energy (note you can also calculate PCs of $\rho_{\rm DE}(z)$, $H(z)$, etc):

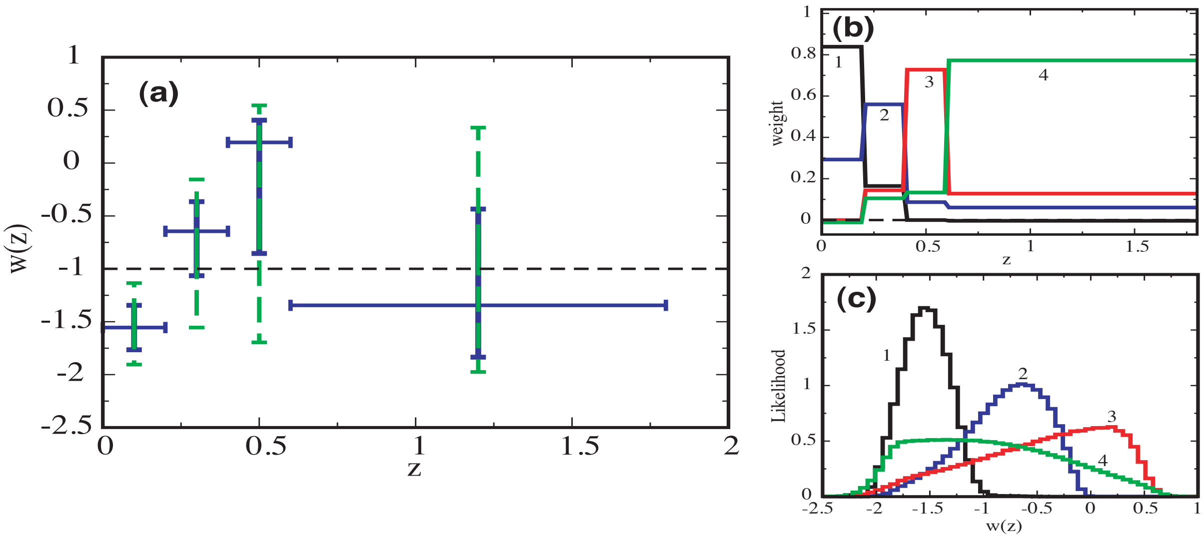

Figure 8: Uncorrelated band-power estimates of the equation of state $w(z)$ of dark energy are shown in panel (a), based on SN data available circa 2004. Vertical error bars show the 1 and 2-σ error bars (the full likelihoods are shown in panel (c)), while the horizontal error bars represent the approximate range over which each measurement applies. The full window functions in redshift space for each of these measurements are shown in panel (b); they have small leakage outside of the original redshift divisions. The window functions and the likelihoods are labeled in order of increasing redshift of the band powers in panel (a). Adopted from Huterer & Cooray (2004).

- The method is as close as it gets to "model independent"
- Data tells you what you measure, and how well — no arbitrary parametrizations
- One can use this approach to design a survey that is most sensitive to the dark energy equation-of-state parameter in some specific redshift interval...
- ... or to study how many independent parameters are measured well by a combination of cosmological probes (i.e. how many PCs have $\sigma(\alpha_i)$, or $\sigma(\alpha_i)/\alpha_i$, less than some desired threshold value)

There are a variety of extensions of this method. A popular "relative" of the PCs is the approach where one parametrizes the equation of state in piecewise constant values, but then uses the diagonalization similar to that above to find linear, *nearly localized and 100% uncorrelated* linear combinations of these bins. A plot of these uncorrelated band-powers is easy to visualize in terms of checking of $w(z)$ varies with redshift. See Fig. 8 and Huterer & Cooray (2004) for details.

Figures of Merit. We finally discuss the so-called figures of merit (FoMs) for dark energy experiments. A FoM is a number, or collection of numbers, that serves as simple and quantifiable metrics by which to evaluate the accuracy of constraints on dark energy parameters from current and proposed experiments. For example, marginalized accuracy in the (constant) equation of state, w, could serve as a figure of merit – since a large FoM is "good", you would probably want something like FoM $= 1/\sigma_w$, or $1/\sigma_w^m$ where m is some positive power.

The most commonly discussed figure of merit is that proposed by the Dark Energy Task Force (Albrecht et al 2007, though this proposal goes back to Huterer & Turner 2001 paper),

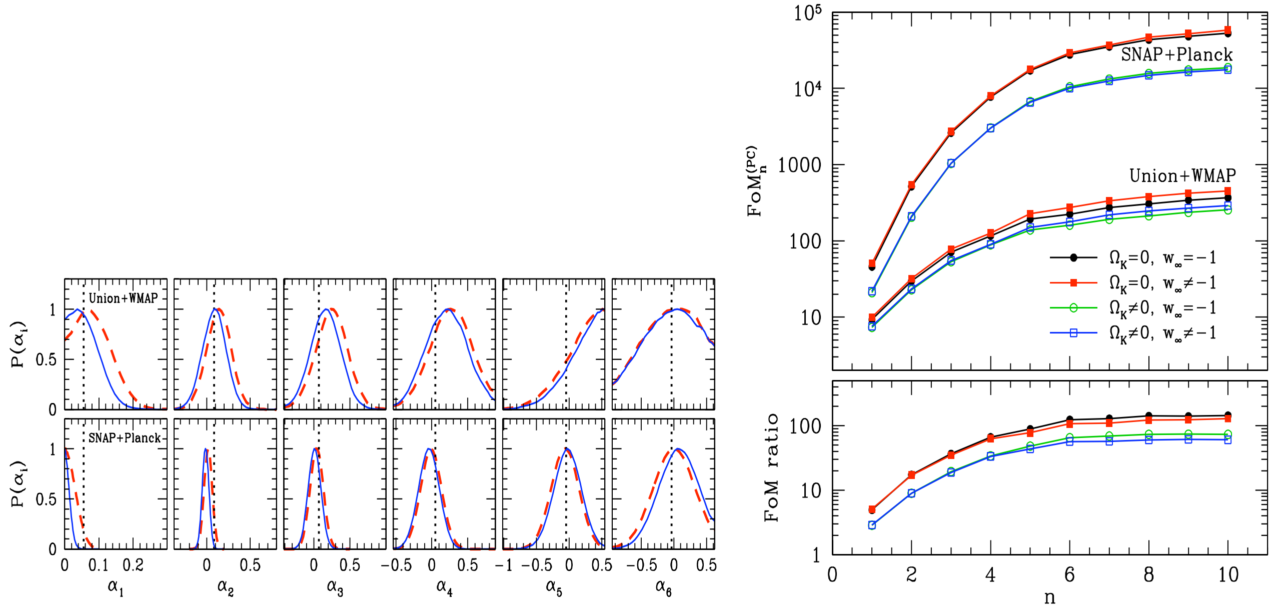

Figure 9: Figure of merit with PCs. **Left plot**: Marginalized 1D posterior distributions for the first 6 PCs of flat (solid blue curves) and nonflat (dashed red curves) quintessence models. Top row: current Union+WMAP data; bottom row: forecasts for SNAP+Planck assuming a realization of the data with $\alpha_i = 0$. **Right plot:** PC-based FoM comparisons. Top panel: PC figures of merit $\text{FoM}_n^{(\text{PC})}$ with forecasted uncertainties for SNAP+Planck and with measured uncertainties for Union+WMAP. Bottom panel: Ratios of $\text{FoM}_n^{(\text{PC})}$ forecasts to current values. In both panels, point types indicate different quintessence model classes: flat (solid points) or non-flat (open points), either with (squares) or without (circles) early dark energy. Adopted from Mortonson, Huterer & Hu (2010).

which is essentially inverse area in the $w_0 - w_a$ plane. For uncorrelated w_0 and w_a this would be $\propto 1/(\sigma_{w_0} \times \sigma_{w_a}$; because the two are typically correlated, the FoM can be defined as

$$\text{FoM}^{(w_0 - w_a)} \equiv (\det \mathbf{C})^{-1/2} \approx \frac{6.17\pi}{A_{95}}, \tag{57}$$

Note the constant of proportionality is really not that important, since typically you compare the FoM from different surveys, and the constant disappears when you take the ratio.

The standard "DETF FoM" defined in Eq. (57) keeps some information about the dynamics of DE (that is, the time variation of $w(z)$). But we can do better, and several more general FoMs have been proposed. My favorite one is, surprisingly, from a paper that I coauthored; Mortonson, Huterer & Hu (2010) proposed taking the FoM to be inversely proportional to the volume of the n-dimensional *ellipsoid* in the space of principal component parameters

$$\text{FoM}_n^{(\text{PC})} \equiv \left(\frac{\det \mathbf{C}_n}{\det \mathbf{C}_n^{(\text{prior})}} \right)^{-1/2}. \tag{58}$$

where the prior covariance matrix is proportional to the (squares of) the product of prior ranges in the principal components. As with the multiplicative constant in Eq. (57), the prior matrix is unimportant since the $\det \mathbf{C}_n^{(\text{prior})}$ term cancels out once you take the ratio of FoMs of two surveys you are comparing. See Fig. 9 for an illustration.

Lecture III: Cosmological data and likelihood analysis

Recommended reading: There are some excellent resources on topics covered here.

- "Numerical Recipes - the Art of Scientific Computing", Press, Teukolsky, Vetterling & Flannery — *A famous book that doesn't disappoint on topics of interest here — see its chapter 15.*
- "Bayes in the sky: Bayesian inference and model selection in cosmology", R. Trotta, arXiv:0803.4089 — *A fairly complete scripture of various Bayesian techniques from one of the Apostles.*
- "A practical guide to Basic Statistical Techniques for Data Analysis in Cosmology", L. Verde, arXiv:0712.3028 — *A good, broad but rushed overview of various numerical/statistical topics in cosmology, with a good list of references.*
- "Karhunen-Loeve Eigenvalue Problems in Cosmology: How Should We Tackle Large Data Sets?", M. Tegmark, A. Taylor and A. Heavens, ApJ, 480, 22 (1997) — *one of the papers that introduced the Fisher matrix to cosmology; explained well and major bonus materials on data compression in cosmology if you are interested.*
- "Unified approach to the classical statistical analysis of small signals", G.J. Feldman and R.D. Cousins, PRD, 57, 3873 (1998) — *If you would like to be a frequentist, read this very clear and important paper, which also gives applications to neutrino oscillation data.*

Bayesian vs. frequentist. There are two principal approaches to statistics, and their competition is as famous as that between the Montagues and Capulets, or the Lakers and the Celtics. These are the Bayesian and frequentist approaches.

Frequentist interpretation of probability defines an event's probability as the limit of its relative frequency in a large number of trials. So I observe the event unfold many times and, in the limit when that number goes to infinity, the relative frequency of its outcome becomes its probability.

Bayesian probability interprets the concept of probability as 'a measure of a state of knowledge, and not as a frequency. One of the crucial features of the Bayesian view is that a probability can be assigned to a hypothesis, which is not possible under the frequentist view, where a hypothesis can only be rejected or not rejected.

More formally, the Bayesian probability calculus makes use of Bayes' formula - a theorem that is valid in all common interpretations of probability - in the following way:

$$\boxed{P(M|D) = \frac{P(D|M)\,P(M)}{P(D)}} \quad \text{(Bayes' theorem)} \tag{59}$$

where M represents a model (or a hypothesis) and D is data. Here

- $P(M)$ is a prior probability of M the probability that M is correct before the data D was seen

- $P(D|M)$ is the conditional probability of seeing the data D given that the hypothesis H is true. $P(D|M)$ is called the *likelihood*
- $P(D)$ is the *a priori* probability of witnessing the data D under all possible hypotheses. It is a normalizing constant that only depends on the data, and which in most cases does not need to be computed explicitly. It is also called the marginal probability, and given by $P(D) = \int P(D|M)P(M)dM$
- $P(M|D)$ is the *posterior probability*: the probability that the hypothesis is true, given the data and the previous state of belief about the hypothesis.

The key thing to note is that we are most often interested in the probability of a model given data, $P(M|D)$, while what we can most often calculate from the data is the likelihood of the data given the model, $P(D|M)$. Bayes' theorem lets you go form from the latter to the former

$$P(M|D) \propto P(D|M)\, P(M) \quad \text{(Bayesian evidence)} \tag{60}$$

Note that the two are equal if the prior in the model space is flat. However, if the prior is not flat, the two will in general differ.

Bayesian approach has many advantages, and has been near-universally accepted in cosmology since the data boom in the 1990s.

- The Bayesian approach is well founded theoretically, and is consistent
- It allows easy incorporation of different data sets. For example, you can have one data set impose an effective prior on the model space M, and then this prior probability is updated with a new data set using the Bayes' theorem.
- In frequentist statistics, a hypothesis can only be rejected or not rejected. In Bayesian statistics, a probability can be assigned to a hypothesis (provided you know or can calculate the marginal probability of the data, $P(D)$).

Bayesian-frequentist example. Say for example that we have a measurement of the Hubble constant of (72 ± 8)km/s/Mpc. What would the Bayesian and the frequentist say?

- Bayesian: the posterior distribution for H_0 has 68% if its integral between 64 and 80km/s/Mpc. The posterior can be used as a prior on a new application of Bayes' theorem.
- Frequentist: Performing the same procedure will cover the real value of H_0 within the limits 68% of the time. But how do I repeat the same procedure (generate a new H_0 out of the underlying model) if I only have one Universe?

Let us give another example. Say I would like to measure Ω_M and Ω_Λ from SN data (let us ignore $\mathcal{M}$ for the moment and just assume these two parameters). What would the two statisticians do?

- Bayesian: Take some prior (say, uniform prior in both Ω_M and Ω_Λ). Then, for each model $M = (\Omega_M, \Omega_\Lambda)$, compute the likelihood of the data, $P(D|M)$ using, for example, the chi-square statistic. Obtain the posterior probability on the two parameters using Bayes' theorem; $P(M|D) \propto P(D|M)\, P(M)$.

- Frequentist: Calibrate your statistic by assuming a model within the range you are exploring (say, $\Omega_M = \Omega_\Lambda = 0.5$) and running many realizations of SN data with that underlying model. Each realization i of the data (points, and errors) will given you χ_i^2. Histogramming χ_i^2 will calibrate the likelihood. Now calculate the χ^2 statistic for the *real* data, assuming the same model, and compare to the histogram — this will give you a (relative likelihood) for that model. Repeat for each model $M = (\Omega_M, \Omega_\Lambda)$.

The latter approach is also called the Feldman-Cousins approach, referring to an excellent paper that I encourage you to read. It is computationally very demanding, since it requires a suite of realizations of data for *each* model M. To make it less demanding, you can hope for the best and assume the histogram of the statistic to be the same for each model, and only do it for one model.

So what prior do I use? In general, the results will depend on the prior. For example, you can consider using a flat prior on some parameter p (equal probability per dp), or a prior flat in the log of p (so equal probability per $d\ln p$). However, *when* the data is very informative, it will completely dominate over the prior and the prior itself will be irrelevant. In other words, when you see people arguing about which prior is "more physical" given that they lead to different final parameter constraints, you conclude that the data they are using is quite weak and probably cannot lead to robust cosmological parameter constraints.

Maximum likelihood Often times, due to the central limit theorem, the likelihood of measuring data given the model can be represented in terms of the Gaussian likelihood

$$\mathcal{L} = \frac{1}{(2\pi)^{n/2}|\det C|^{1/2}} \exp\left[-\frac{1}{2}(d-\bar{d})_i^T C_{ij}^{-1}(d-\bar{d})_j\right] \tag{61}$$

Note that the term in the exponent is just minus one half times chi-squared

$$\chi^2 \equiv (d-\bar{d})_i^T C_{ij}^{-1}(d-\bar{d})_j \tag{62}$$

(where here and below we implicitly sum over the repeated indices). Note that maximum a Gaussian likelihood, and maximizing chi-square statistic, are one and the same in the limit when the covariance matrix C doesn't depend on the cosmological parameters.

Markov chain Monte Carlo. Say you want to constrain N cosmological parameters; let us take $N = 10$ typical for cosmology. Say, for simplicity, that you want to allow each parameter to take M discrete values; let us take $M = 10$ which is the bare barest minimum you would want to do. Then the total number of models to explore (and calculate observables for) is $M^N = 10^{10}$, which is huge — this might be doable for a simpler data set, but if you consider running CAMB (which actually only takes seconds per model), this is about 100 years. And if you want to allow the still-modest $M = 20$ values per parameter, then likelihood calculations would take 100,000 years, which means that an early Neanderthal starting the chains would make it just in time for his paper to be published this year.

Markov chain Monte Carlo (MCMC) methods are an incredibly powerful tool to overcome these problems[4]. MCMC are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a large number of steps is then used as a sample from the desired distribution. The quality of the sample improves as a function of the number of steps.

[4] A *Markov process* (or a Markov chain) is a process where the future states only depend on the present state, but not on the past states.

Instead of going exponentially with the number of parameters, the MCMC calculation goes approximately linearly with N.

Usually it is not hard to construct a Markov Chain with the desired properties. The more difficult problem is to determine how many steps are needed to converge to the stationary distribution within an acceptable error. A good chain will have rapid mixing - the stationary distribution is reached quickly starting from an arbitrary position.

MCMC: the science. We will only consider the Metropolis-Hastings algorithm here, which is the most simple variant of MCMC.

The Metropolis-Hastings algorithm draws samples from the probability distribution $P(x)$. How does it do that? The algorithm generates a Markov chain where each state x^{t+1} depends only on the previous state x^t. The algorithm uses a proposal density $Q(x'|x^t)$ which depends on the current state x^t to generate a new proposed sample x'. This proposal is accepted as the next value (so, $x^{t+1} = x'$) if α, drawn from a uniform distribution $U[0, 1]$, satisfies

$$\alpha < \frac{P(x')Q(x^t|x')}{P(x^t)Q(x'|x^t)} \tag{63}$$

If the proposal is not accepted, then the current value of x is retained, so that $x^{t+1} = x^t$.

So the step-by-step instructions for the Metropolis-Hastings algorithm are as follows: given you are at some parameter value x^t, you

- calculate the value $a_1 a_2$, where $a_1 = P(x')/P(x)$ is the likelihood ratio between the proposed sample x' and the previous sample x^t, and $a_2 = Q(x^t|x')/Q(x'|x^t)$ is the ratio of the proposal density in two directions (from x^t to x' and vice versa). Note that $a_2 = 1$ if the proposal density is symmetric, which we often assume
- if $a \geq 1$, then move to the proposed point; $x^{t+1} = x'$; repeat
- if $a < 1$, then draw a random number $\alpha \in U[0, 1]$. If $a > \alpha$ then move to the proposed point; $x^{t+1} = x'$. If $a < \alpha$, do NOT move to the proposed point; $x^{t+1} = x^t$;
- repeat

MCMC: the art. The Metropolis-Hastings algorithm (above) is about 10 lines of computer code. So what's all the fuss about? Well, to make a successful MCMC, you need to take care of a number of things.

1. You need to assure that the *burn-in* stage is not included in the final results. This typically means running the MCMC for a number of steps (say, 10,000), discarding those results, and then doing a "production run" (with, say, a million steps)
2. You need to ensure that the MCMC is *efficient* — ideally, it will move from x^t to the proposal value x' about 1/3 of the time. Imagine if you had two highly degenerate parameters — say, Ω_M and h in CMB measurements where only the combination $\Omega_M h^2$ is well determined. Say you use an otherwise reasonable proposal distribution which is a multivariate Gaussian with standard deviation equal to the *guessed* error in each parameter, and without correlation between parameters. Steps in Ω_M or h separately will lead to rejection of the proposed steps vast majority of the time!. However, if you are clever and reparameterize the problem so that you have parameter $\Omega_M h^2$ (and, say, Ω_M separately), then the acceptance will be much better, and the asymptotic distribution will be reached sooner. Equivalently,

making the proposal function be an "off-diagonal" Gaussian with eigendirections along the axes of the (pre-computed, rough) estimate for the target distribution $P(x)$ will help — for example, having used a short chain to discover that Ω_M and h are highly correlated, you could step along the gaussian whose axes are short and long axes of the $\Omega_M h^2$ contour in 2D.

3. Finally, you need to make sure *mixing* of your chain. To do so, you can *thin* the chain, writing out every 100th (for example) value, so that you decrease the (otherwise very high) correlation between the steps. Likewise, you should run several (say, four) chains, and test convergence using one of the criteria (say, the Gelman-Rubin criterion) that typically compare variance within a chain with variance between different chains.

MCMC: enjoying the fruits of labor. MCMC is really a fantastic tool, enabling exploration of the multi-dimensional likelihoods that cannot be even contemplated using a naive multi-dimensional gridding of the parameter space.

Not only that, but computing constraints on any quantities of interest, once you have run your chains, is trivial. *Post-processing the MCMC output is easy.* What you need to do is write out chains, together with the "weight" (number of times the chain is "stuck" at that value if the proposed move was rejected) for each step. Then, for any parameter set of choice — a single parameter (e.g. Ω_M), joint contour of two parameters (e.g. (Ω_M, w)), a function of a few parameters (e.g. $w(a) = w_0 + w_a(1-a)$), whatever — you just look at their weights, rank order them, and add them until you get 68% or 95% or whatever fraction of the total weight. Remember, weight is proportional to posterior likelihood owing to ergodic property of MCMC.

Moreover, let us say that, after this hard work, that you decide you would like to combine your constraints with some other. That's easy — you just use constraints from your chain as a prior, and combine with new constraints to get a new posterior.

Fisher information matrix. Fisher matrix presents an excellent tool to forecast errors from a given experiment. Even though we have argued that the MCMC itself is "easy" and "fast" compared to brute force methods for exploring the likelihood, in comparison the Fisher matrix is still *much* easier and faster tool to forecast the likelihood distribution, given some expected experimental data.

Let us assume that we have cosmological measurements, and that the associated likelihood in the data can be represented by the likelihood $\mathcal{L}$. The Fisher matrix is formally defined as the curvature of the likelihood — that is, matrix of second derivatives of the log likelihood around the peak[5]

$$\boxed{F_{ij} = \left\langle -\frac{\partial^2 \ln \mathcal{L}}{\partial p_i \partial p_j} \right\rangle} \tag{64}$$

where $\{p_i\}$ is the set of cosmological parameters.

When doing the Fisher matrix, we always[6] assume that the data are distributed according to a multivariate Gaussian — that is, that the covariance matrix of the data C has all information. In particular, we assume Eq. (61)

$$\mathcal{L} = \frac{1}{(2\pi)^{n/2} |\det C|^{1/2}} \exp\left[-\frac{1}{2}(d-\bar{d})_i^T C_{ij}^{-1} (d-\bar{d})_j\right] \tag{65}$$

[5] You will notice that Fisher matrix is the negative of the Hessian of the log likelihood.

[6] As far as I know! To impress me greatly, work out the Fisher matrix for e.g. the case when the likelihood is assumed to be something different — e.g. an approximation to the top-hat (i.e. flat in an interval). Might be hard.

where d_i are the data (and $\bar{d}_i$ is the mean for each i) and C_{ij} is the covariance. Then you could show (exercise!) that the Fisher matrix evaluates to

$$\boxed{F_{ij} = \frac{1}{2}\mathrm{Tr}[C^{-1}C_{,i}C^{-1}C_{,j}] + \bar{d}_{,i}^T C^{-1}\bar{d}_{,j}} \tag{66}$$

where $_{,i}$ is partial derivative with respect to p_i.

Fisher matrix as an estimate of parameter errors. Most of the time, Fisher matrix users rely on the Cramer-Rao inequality, which says that an error in a cosmological parameter p_i will be greater or equal to the corresponding Fisher matrix element

$$\boxed{\sigma(p_i) \geq \begin{cases} \sqrt{(F^{-1})_{ii}} & \text{(marginalized)} \\ 1/\sqrt{F_{ii}} & \text{(unmarginalized)} \end{cases}} \tag{67}$$

where "marginalized" is the uncertainty marginalized over all other $N-1$ parameters, while the "unmarginalized" case is when you ignore the other parameters, assuming them effectively fixed and known. Note that the marginalized case has inverse of F which lets the parameters "talk to each other" about degeneracies. Most often in cosmology we are interested in the marginalized errors; the unmarginalized ones are often much larger and not practically achievable.

So while the Cramer-Rao just tells us about the best possible error (so using the best possible estimator etc), we often just assume that it gives *the* error from data of the given quality.

Examples. Let us give some examples of the expressions for probe-specific Fisher matrices.

For type Ia supernova observations, the covariance matrix of SNe doesn't depend on cosmological parameters, and in fact it's often taken to be constant (remember, $C_{ij} \to \sigma_m^2\delta_{ij}$ with $\sigma_m \sim 0.15\,\mathrm{mag}$). So then

$$F_{ij}^{\mathrm{SNe}} = \sum_{n=1}^{N_{\mathrm{SNe}}} \frac{1}{\sigma_m^2}\frac{\partial m(z_n)}{\partial p_i}\frac{\partial m(z_n)}{\partial p_j} \tag{68}$$

where $m(z) = m(z, \Omega_M, \Omega_\Lambda, \mathcal{M}...)$ is the theoretically expected apparent magnitude. Notice that, if you had a full off-diagonal covariance matrix C_{ij} that is still independent of the cosmological parameters, the equation above would generalize trivially. For the cluster counts the simple-minded equation is similar; let N_k be number of clusters in the kth bin and $O(z)$ be an observable (say, SZ flux), then

$$F_{ij}^{\mathrm{clus}} = \sum_{k=1}^{Q} \frac{N_k}{\sigma_O(z_k)^2}\frac{\partial O(z_k)}{\partial \theta_i}\frac{\partial O(z_k)}{\partial \theta_j} \tag{69}$$

Now consider a case of measurements of the CMB (or weak lensing) power spectrum where the mean (temperature or shear) is zero and doesn't depend on cosmology, but the covariance carries all cosmological information. Then you can show that

$$F_{ij}^{\mathrm{WL}} = \sum_{\ell} \frac{\partial \mathbf{C}}{\partial p_i}\,\mathbf{Cov}^{-1}\,\frac{\partial \mathbf{C}}{\partial p_j}, \tag{70}$$

where $\mathbf{Cov}^{-1}$ is the inverse of the covariance matrix between the observed power spectra whose elements are given by

$$\mathrm{Cov}\left[C_{ij}^{\kappa}(\ell), C_{kl}^{\kappa}(\ell)\right] = \frac{\delta_{\ell\ell'}}{(2\ell+1)\, f_{\mathrm{sky}}\,\Delta\ell}\,\left[C_{ik}^{\kappa}(\ell)C_{jl}^{\kappa}(\ell) + C_{il}^{\kappa}(\ell)C_{jk}^{\kappa}(\ell)\right]. \tag{71}$$

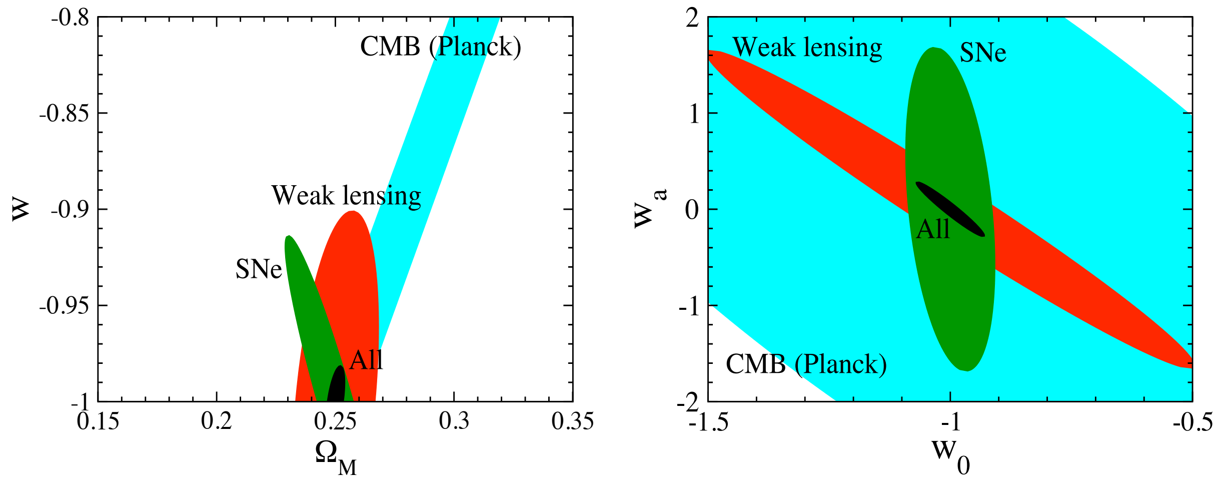

Figure 10: Illustration of forecast constraints on dark energy parameters, taken from the Frieman, Turner & Huterer review of DE. All contours have been computed using the Fisher matrix. Note that the contours are by definition ellipses, and one (Planck) is nearly completely degenerate — meaning, long.

where $C^{\kappa}_{kl}(\ell)$ is the covariance of convergence κ between galaxies in the kth and lth redshift bin, on scales corresponding to a multipole bin centered at ℓ with width $\Delta\ell$, in a survey covering $f_{\rm sky}$ fraction of the sky. You can tell by eyeballing this covariance-of-covariance four-point correlation function that it was computed using Wick's theorem, that is, assuming Gaussianity of C.

Marginalization over parameters. If you have, say, N, cosmological parameters, how do you marginalize over $N-M$ of them to be left with a desired joint constraints on M parameters? This is easy:

- Calculate the full $N \times N$ matrix F
- Invert it to get F^{-1}
- Take the desired $M \times M$ subset of F^{-1}, and call it G^{-1} note that this matrix is M dimensional
- Invert G^{-1} to get G

and voilà — the matrix G is the projected Fisher matrix onto the M-dimensional space. That was easy!

Fisher ellipses. How do you plot the Fisher matrix contour? To plot a 2D ellipse, you first want to project down to that space, and be left with a marginalized 2x2 Fisher matrix, call it G. The equation for the 2D ellipse is

$$G_{11}p_1^2 + G_{12}p_1p_2 + G_{22}p_2^2 = \frac{1}{f} \tag{72}$$

where $f = 0.434$ for a 68% CL ellipse, and $f = 0.167$ for a 95% ellipse (these numbers can easily be calculated using Gaussian statistics).

The Fisher ellipse are also show how much information is carried by the data on the parameters — the smaller the ellipse, the more information. More generally, the volume of an n-dimensional ellipsoid is

$$\boxed{\text{Volume} \propto (\det F)^{-1/2}}. \tag{73}$$

You can find this useful if you are estimating relative amounts of information in surveys, etc.

Fisher bias. Another great application of the Fisher matrix is to calculate the bias in parameters p_i given biases in the observables. This can be derived easily again assuming the same Gaussian distribution in the data; the result (for weak lensing) is

$$\delta p_i = F_{ij}^{-1} \sum_{\ell} \left[C_{\alpha}^{\kappa}(\ell) - \bar{C}_{\alpha}^{\kappa}(\ell)\right] \, \text{Cov}^{-1} \left[\bar{C}_{\alpha}^{\kappa}(\ell), \bar{C}_{\beta}^{\kappa}(\ell)\right] \, \frac{\partial \bar{C}_{\beta}^{\kappa}(\ell)}{\partial p_j}, \tag{74}$$

where $[\bar{C}_{\alpha}^{\kappa}(\ell), \bar{C}_{\beta}^{\kappa}(\ell)]$ is the bias in the "observable" shear covariance due to any reason.

Perhaps a simpler example would be that for SNeIa, where the bias in the parameters takes the form

$$\delta p_i = F_{ij}^{-1} \sum_{n} \frac{1}{\sigma_m^2} \left[m(z_n) - \bar{m}(z_n)\right] \frac{\partial \bar{m}(z_n)}{\partial p_j} \tag{75}$$

where $[m(z_n) - \bar{m}(z_n)]$ is the bias in the observed apparent magnitudes.

The bias formula is extremely useful if you would like to see what effect on cosmological parameter errors an arbitrary systematic effect makes. So, given some biases in the observable quantities, for example $[C_{\alpha}^{\kappa}(\ell) - \bar{C}_{\alpha}^{\kappa}(\ell)]$ in Eq. (74), you can find biases in the cosmological parameters δp_i. Then you can compare those biases with the statistical errors in the cosmological parameters $\sigma(p_i)$ and impose requirements on the control of your systematic effect so that $|\delta p_i|/\sigma(p_i)$ is no larger than some threshold, say 0.3 (corresponding to $< 30\%$ bias in the parameters). See the paper by Huterer & Takada (2006) for application to how well theoretical prediction for the power spectrum $P(k)$ needs to be know in order not to "mess up" cosmological parameter determinations.