Lecture III:
Cosmological data and Likelihood analysis

Dragan Huterer

Physics Department
University of Michigan

In this Lecture:

1. How to test modified gravity proposals for DE
2. Bayesian vs. Frequentist approach to statistics
3. Markov chain Monte Carlo (MCMC)
4. Fisher information matrix

Review of facts about Dark Energy

- DE dominates universe's budget, but not overwhelmingly ($\Omega_{DE} \approx 0.75$)
- DE makes distant SNe dimmer (or more distant) than expected
- DE has strongly negative pressure ($p \approx -\rho$, $w \approx -1$)
- DE domination $\Rightarrow$ suppressed growth of structure (so, had the universe always been DE dominated, there would be no galaxies in the sky!)
- DE is spatially smooth (or nearly so)
- Best measurements of DE obtained with a variety of probes at $z \leq 2$
- Expansion and growth history measurements are crucial; the former described by $w(z)$, and the latter by $g(z)$

What if gravity deviates from GR?

For example:

$$H^2 - F(H) = \frac{8\pi G}{3}\rho, \qquad \text{or} \qquad H^2 = \frac{8\pi G}{3}\left(\rho + \frac{3F(H)}{8\pi G}\right)$$

Modified gravity

Dark energy

Notice: there is no way to distinguish these two possibilities just by measuring H(z)!

How to distinguish between DE and MG

- In standard GR, H(z) determines distances and growth of structure

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi\rho_M\delta = 0$$

- So check if this is true by measuring separately

Distances

(as known as kinematic probes)
(a.k.a. 0^{th} order cosmology)

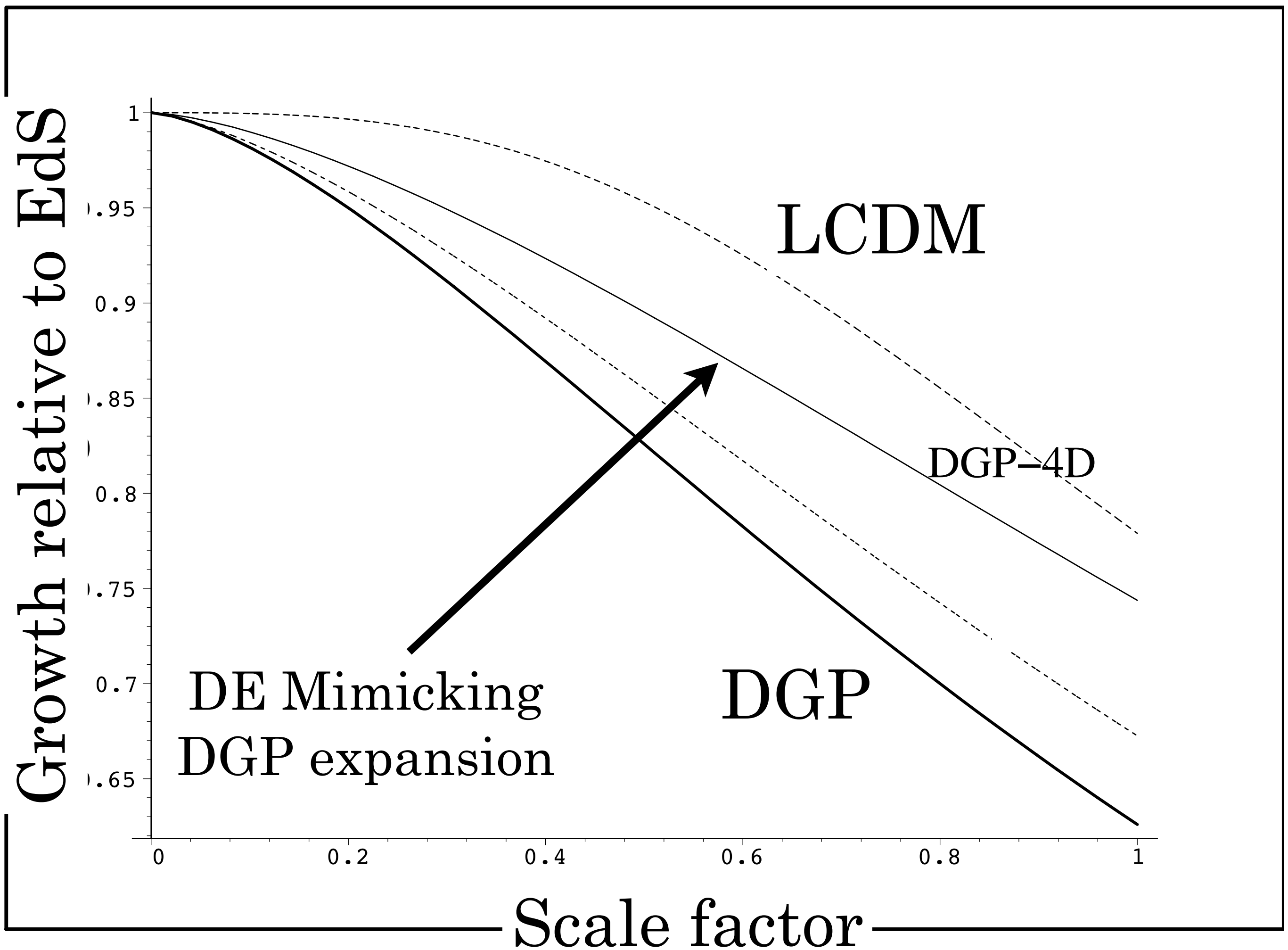
Probed by SN Ia, BAO, CMB, weak lensing, cluster abundance

Growth

(a.k.a. dynamical probes)
(a.k.a. 1^{st} order cosmology)

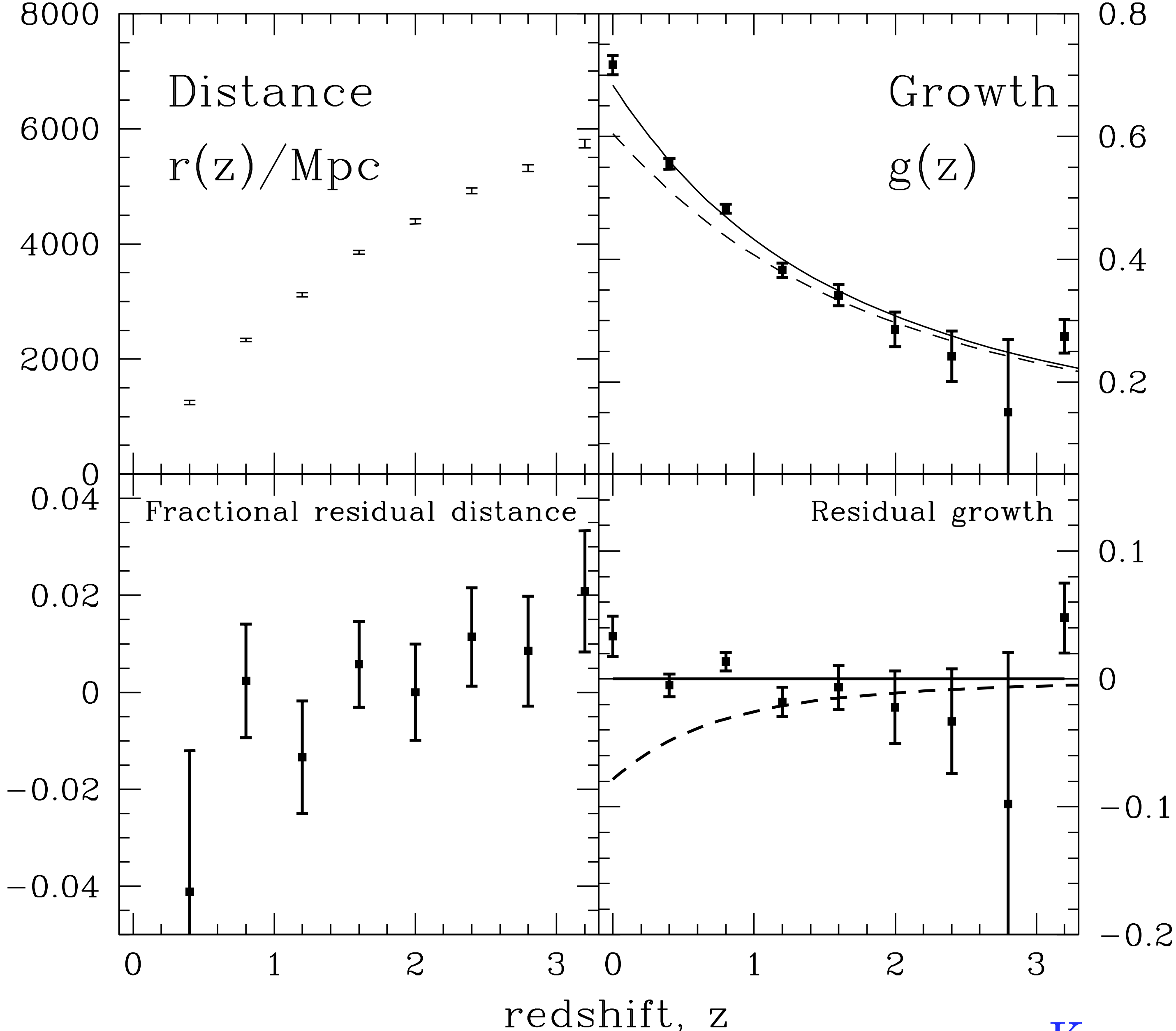
Probed by galaxy clustering, weak lensing, cluster abundance

example: DGP linear growth



Lue, Scoccimarro & Starkman; Koyama & Maartens; Sawicki, Song & Hu

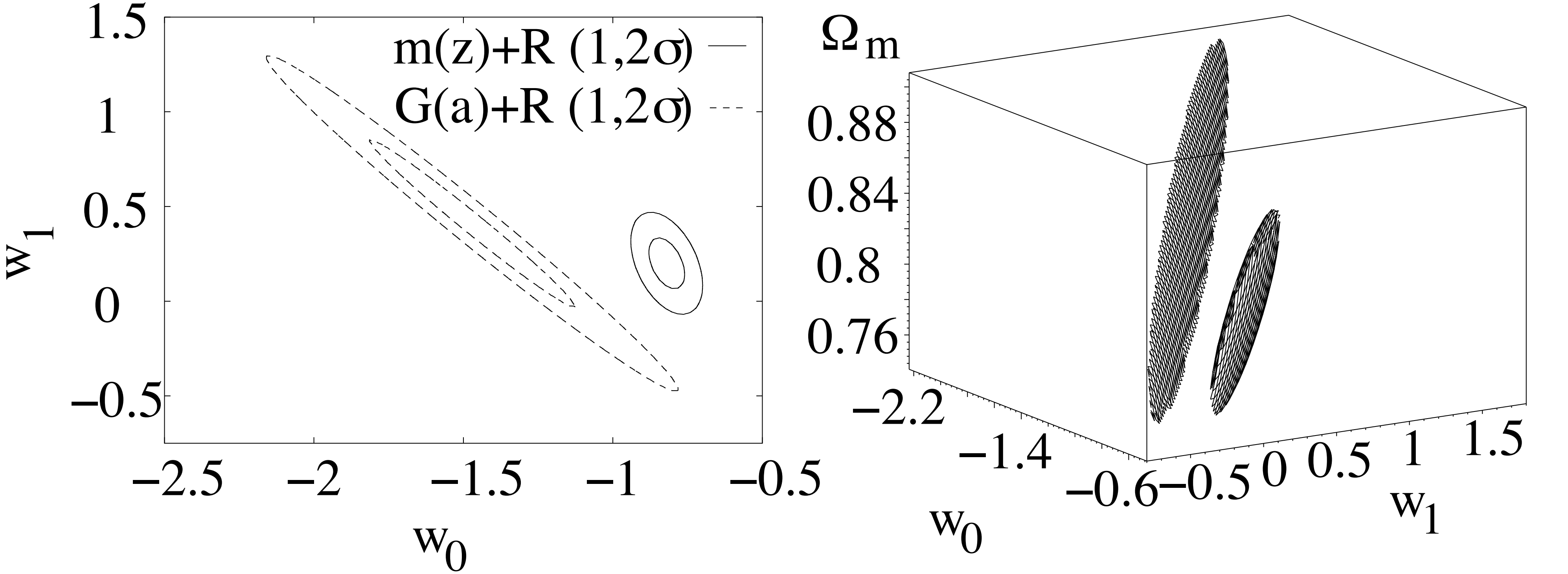
Strategy I: distance (z), growth(z) separately



Measure r(z), g(z), see if they agree

Knox, Song & Tyson 2005

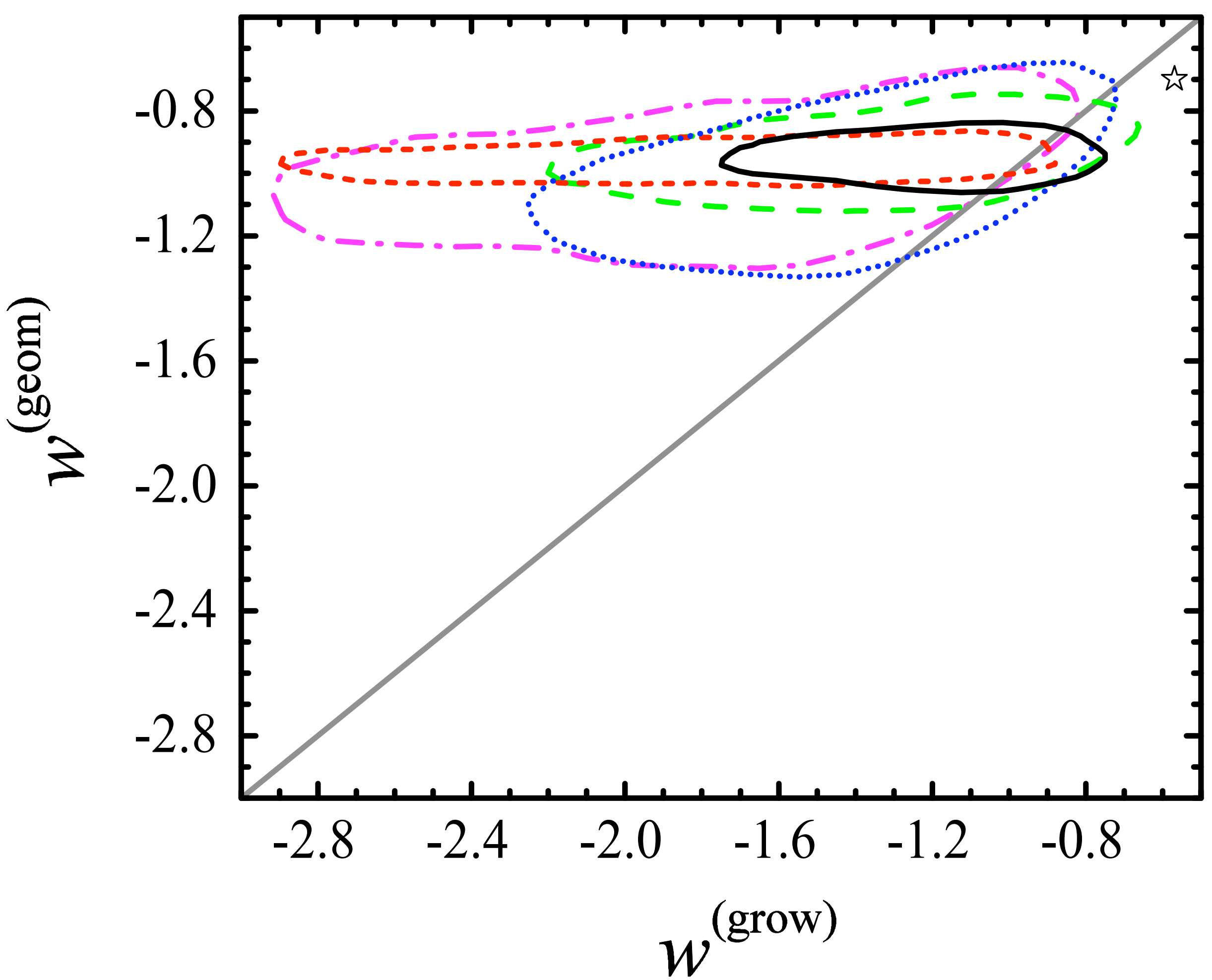
Strategy II: Measure (Ω_m, w_0, w_a) separately for growth and distance



Measure w_0 and w_1=w_a for growth and distance, see if they agree

Ishak, Upadhye & Spergel 2005, others...

Strategy II.5: Measure w separately; example from real data



Nice work, but current constraints are weak

Wang, Hui, May & Haiman, 2007

Strategy III: “Minimalist Modified Gravity”

$$g(a) \equiv \frac{\delta}{a} = \exp\left[\int_0^a d\ln a[\Omega_M(a)^\gamma - 1]\right]$$

Excellent fit to standard DE cosmology with

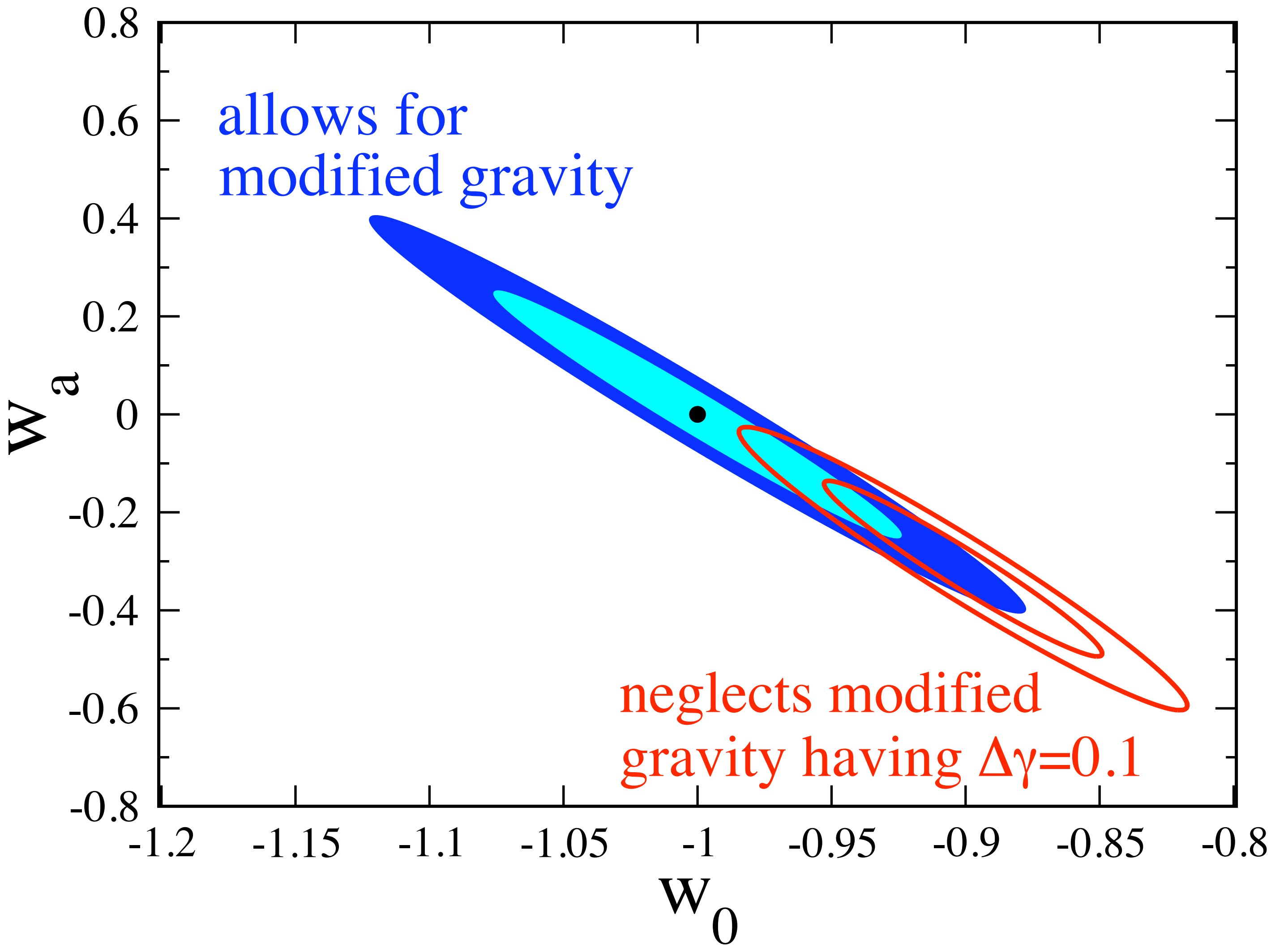
$$\gamma = 0.55 + 0.05[1 + w(z = 1)]$$

Linder 2005

- Gamma is a new parameter - the growth index - and we should measure it!
- E.g. fits DGP with value different from GR by $\Delta\gamma$=0.13
- Strategy: measure γ, see if it differs from ~0.55 or not

Huterer & Linder, 2007

Price of ignorance of MG



Huterer & Linder 2007

What about fluctuations in DE?

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

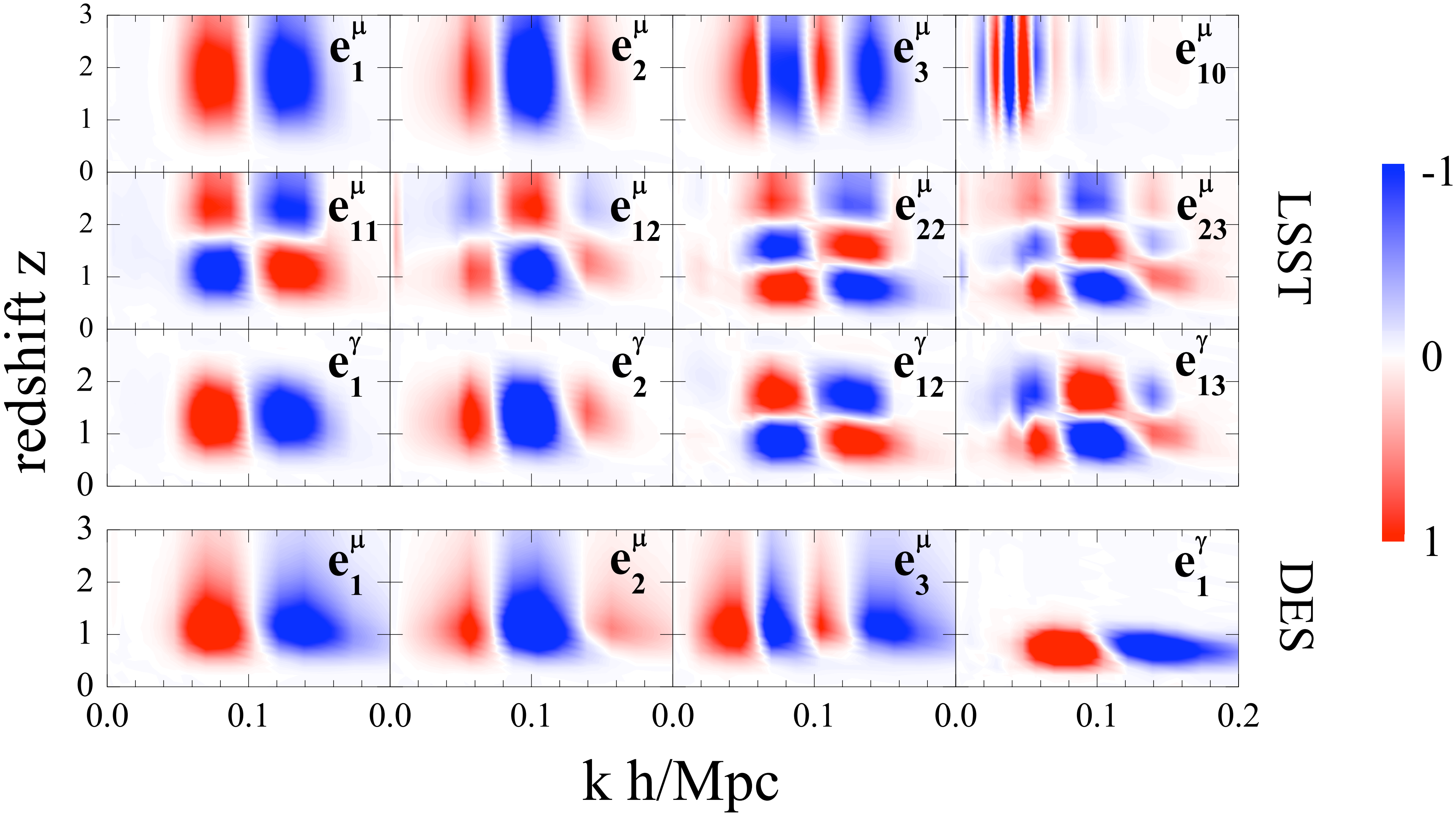
- Lensing probes φ+ψ
- velocity (and dynamical measure) probe ψ

Measuring ψ and φ separately
also tests models of modified gravity

e.g. Jain & Zhang 2007

Principal components of MG

$$\frac{\Phi}{\Psi} = \gamma(k, a), \quad k^2\Psi = -\mu(k, a)4\pi G a^2 \rho \Delta$$



Zhao, Pogosian, Silverstri, Zylberberg, arXiv:0905.1326

Bayesian statistics

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.

Bayes' theorem

(D=data, M=model)

Posterior probability: model given data

Likelihood (data given model)

Prior probability (of models)

$$P(M|D) = \frac{P(D|M)\,P(M)}{P(D)}$$

Probability of data (usually constant)

Bayesian vs. Frequentist: Example 1

Say we have measurements of H_0=(72±8) kms/Mpc.
What would the two statisticians say?

1. Bayesian:

- The posterior distribution for H_0 has 68% of its integral between 64 and 80 km/s/Mpc.
- The posterior can be used as a prior on a new application of Bayes' theorem.

Bayesian vs. Frequentist: Example 1

Say we have measurements of H_0=(72±8) kms/Mpc.
What would the two statisticians say?

2. 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?

Bayesian vs. Frequentist: Example 2

Say I would like to measure Ω_M and Ω_Λ from SN data.
What would the two statisticians do?

1. 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 χ^2 statistics
- Obtain the posterior probability on the two parameters using Bayes' theorem:

$$P(M|D) \propto P(D|M)\, P(M)$$

Bayesian vs. Frequentist: Example 2

Say I would like to measure Ω_M and Ω_Λ from SN data.
What would the two statisticians do?

2. Frequentist:

Feldman & Cousins, PRD, 1997

- Calibrate your statistic: for each model within the range you are exploring, generate many realizations of data with that underlying model. Each realization of the data (points, and errors) gives you a χ^2.
- Histogram χ^2 to 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 $M=(\Omega_M, \Omega_\Lambda)$

Statistics: philosophy

- When data are informative, Bayesian and frequentist approach will give very similar results
- But when data are 'weak', the two will generally differ
- No 'right answer' as to which one is better
- Given that we have 1 universe and cannot get arbitrary amount of data, **Bayesian approach seems more appropriate**
- In particular, Bayesian enables answering questions about model selection (e.g. is a dark energy model with w(z) a better fit to the data than w=const) A
- Also Bayesian enables easily adding new information (new data)

e.g. Trotta, arXiv:0804:4089

Markov chain Monte Carlo (MCMC)

- Say we'd like to constraint cosmological parameters using some CMB or LSS data
- We have ~10 parameters; say we consider 20 values in each parameter to get smooth contours
- $\rightarrow 20^{10}$ ($\sim 10^{13}$) parameter combinations
- CAMB and WMAP likelihood take seconds to run per model $\rightarrow$ a total of 100 million years CPU time
- A better strategy of the likelihood exploration is needed!

Markov chain Monte Carlo (MCMC)

- MCMC: A method invented at Los Alamos lab in the 1950s by physicists
- Instead of mapping out the likelihood, try **sampling from** the likelihood
- Metropolis algorithm:
- given the parameter set at some step t, x^t, draw the next step x^{t+1} from some given proposal density $Q(x^{t+1} \mid x^t)$
- Now draw a random number $\alpha = U[0, 1]$
- If $\alpha < P(x^{t+1}) / P(x^t)$, $x^t \rightarrow x^{t+1}$ *
- If $\alpha < P(x^{t+1}) / P(x^t)$, $x^t \rightarrow x^t$ (and repeat)

*Note: if $P(x^{t+1}) > P(x^t)$, you **always** move to the proposed parameter value

Fisher Information Matrix

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

(minus) Hessian
of likelihood

Cramér-Rao inequality:
best errors you can achieve in cosmological parameters are

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

Fisher matrix can be rewritten as
(Tegmark Taylor & Heavens 1997)

covariance
of data

data

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

Fisher Matrix: examples

SN Ia: observable is magnitude m(z)

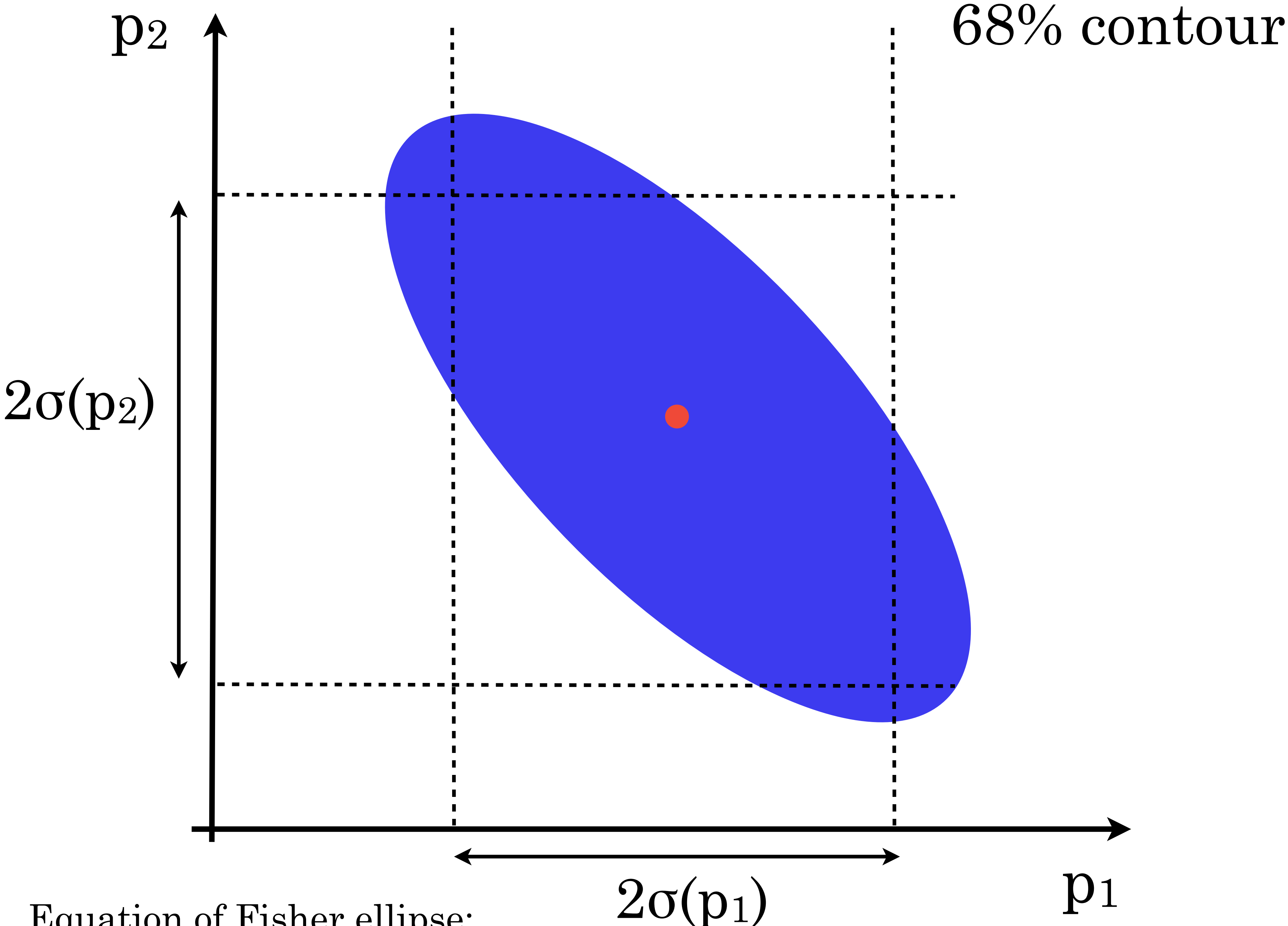
$$F_{ij}^{\mathrm{SNe}} = \sum_{n=1}^{N_{\mathrm{SNe}}} \frac{1}{\sigma_m^2} \frac{\partial m(z_n)}{\partial p_i} \frac{\partial m(z_n)}{\partial p_j}$$

Cluster counts: observable is O(z) (say X-ray or SZ flux etc)

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

Weak lensing power spectrum: observable is tomographic power spectrum $C_{ab}(\ell)$

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



Equation of Fisher ellipse:

$$F_{11}p_1^2 + 2F_{12}p_1p_2 + F_{22}p_2^2 = \chi^2$$

χ^2=2.3 (68%)
χ^2=6.1 (95%)

Fisher Matrix: facts

- Extremely useful tool for forecasting errors (and also Figures of Merit, in defining PCs, in the quadratic estimator method, etc)
- Easy to calculate: - only need one calculation of the observables for the fiducial model, and its derivatives wrt cosmological parameters
- Assumes that the likelihood (in parameters) is Gaussian: good approximation near the peak of likelihood (i.e. when the parameter errors are small)

Marginalizing over parameters with Fisher

Say you have N, cosmological parameters.
How do you marginalize over N-M of them to be left with a desired joint constraints on M parameters?

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

And voilà -the matrix G is the projected Fisher matrix onto the M-dimensional space

Bias in parameters using Fisher matrix

Say you have biases (say, systematic errors) in observables.
How do you calculate the resulting bias in cosmological parameters p_i?
Easily! Can derive formula from first principles.

SN Ia example:

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

Weak lensing example:

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