Quantifying cosmological tensions

Interpreting the DES evidence ratio

Will Handley
wh260@cam.ac.uk

Astrophysics Group
Cavendish Laboratory
University of Cambridge

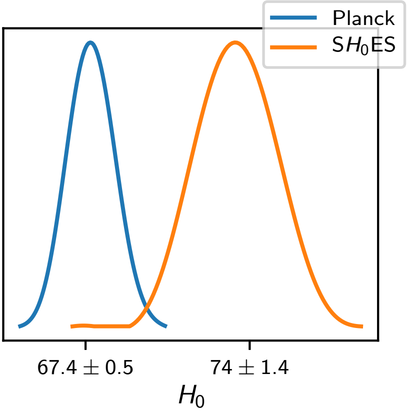
24$^{\text{th}}$ July 2019

Handley & Lemos arXiv:1902.04029, arXiv:1903.06682

github.com/williamjameshandley/anesthetic

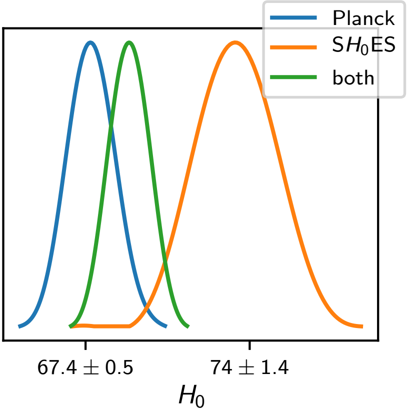
The Hubble H_0 tension

- CMB cosmologists (Planck) infer $H_0 = 67 \pm 0.5$ km s^{-1}Mpc^{-1}
- Supernovae data (SH_0ES) measure $H_0 = 74 \pm 1.4$
- $> 4\sigma$ discrepancy could be due to:
  - Systematic error
  - Problem with standard model of cosmology (ΛCDM)
- Inconsistent datasets shouldn't be combined



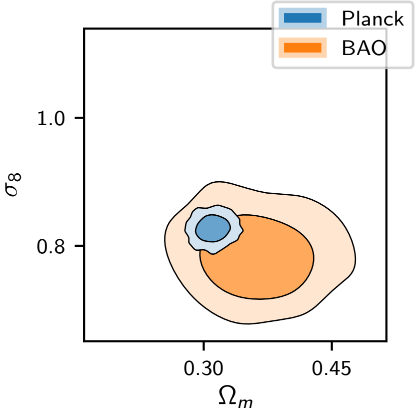
The Hubble H_0 tension

- CMB cosmologists (Planck) infer $H_0 = 67 \pm 0.5$ km s^{-1}Mpc^{-1}
- Supernovae data (SH_0ES) measure $H_0 = 74 \pm 1.4$
- $> 4\sigma$ discrepancy could be due to:
  - Systematic error
  - Problem with standard model of cosmology (ΛCDM)
- Inconsistent datasets shouldn't be combined



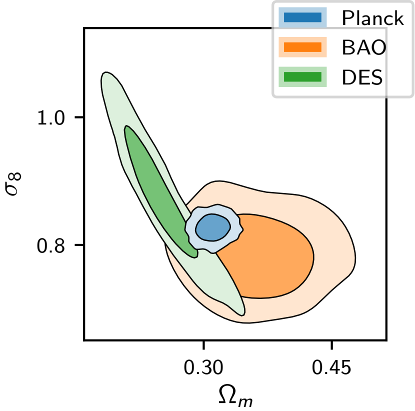
σ_8-Ω_m tension

- Matter density Ω_m and RMS matter fluctuations σ_8 are constrained by
- BAO and Planck look consistent



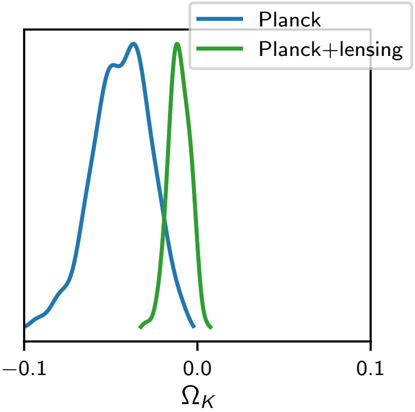
σ_8-Ω_m tension

- Matter density Ω_m and RMS matter fluctuations σ_8 are constrained by
- BAO and Planck look consistent
- DES is less clear
- How do you define a tension in terms of "sigma" for this case?



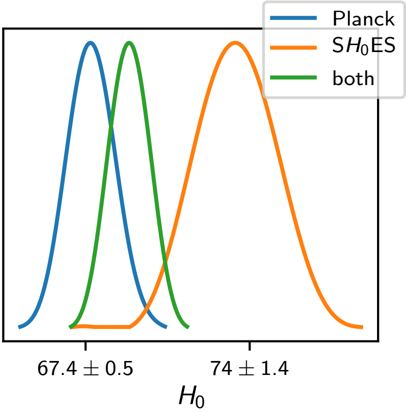
Ω_K tension

- Models with spatial curvature Ω_K.
- Best-kept secret of Planck: only $1/10,000$ MCMC samples $\Omega_K > 0$.
- How consistent do Planck and CMB lensing look?
- Await likelihood release (beginning of next month)



Ω_K tension

- Models with spatial curvature Ω_K.
- Best-kept secret of Planck: only $1/10,000$ MCMC samples $\Omega_K > 0$.
- How consistent do Planck and CMB lensing look?
- Await likelihood release (beginning of next month)



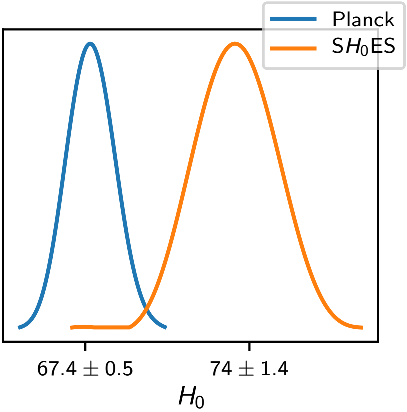
Quantifying tension

Gaussians

For 1D Gaussian distributions, tension is pretty easy to define:

$$X = \frac{\mu_A - \mu_B}{\sqrt{\sigma_A^2 + \sigma_B^2}},$$

where μ and σ are the respective parameter means and standard deviations.

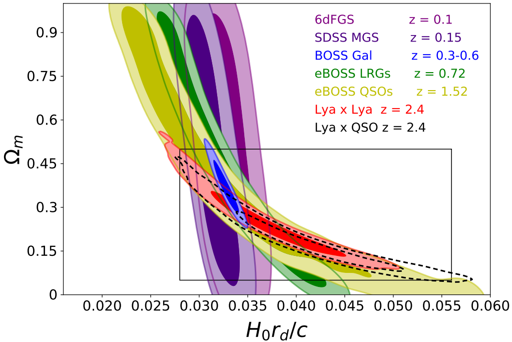


The multivariate d-dimensional equivalent to this tension would be:

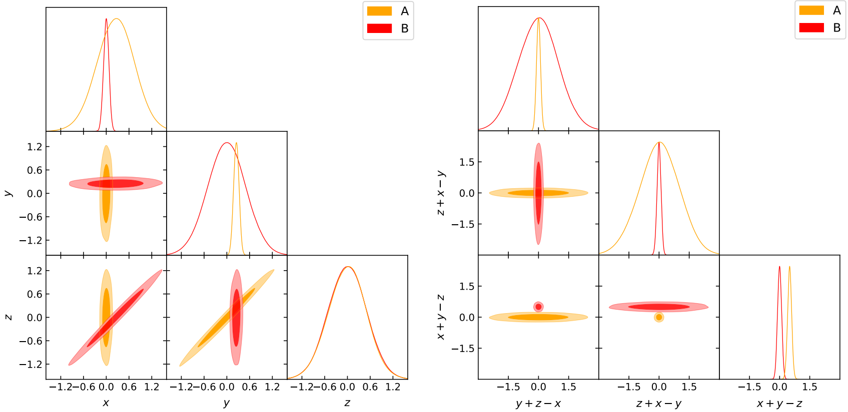
$$X_d^2 = (\mu_A - \mu_B)^T(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B),$$

where Σ is in general a covariance matrix.

Things become less clear when distributions become "banana like" (arXiv:1906.11628), or worse, multimodal.



Many attempts to generalise the Gaussian case result in a parameterisation-dependent quantity.



- In high dimensions, things can look good when projected into 2D.
- We need a systematic way of seeking out tension, without relying on inspired choices of parameters to reveal them

The DES evidence ratio R

- The Dark Energy Survey (arXiv:1708.01530) quantifies tension between two datasets A and B using the Bayes ratio:

$$R = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A \mathcal{Z}_B}$$

  where $\mathcal{Z}$ is the Bayesian evidence.
- Many attractive properties:
  - Symmetry
  - Parameterisation independence
  - Dimensional consistency
  - Use of well-defined Bayesian quantities
- What does it mean?

Bayesian evidence $\mathcal{Z}$

- Bayes theorem for parameter estimation:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \quad \longrightarrow \quad \text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

- Normalising constant $\equiv$ Bayesian evidence $\equiv$ $P(D)$ is hard to compute:

$$P(D) = \int P(D|\theta)P(\theta)d\theta = \langle \text{Likelihood} \rangle_{\text{Prior}}$$

- Traditionally used to compare models using the same data
- For DES, it is used to compare different data with the same model.
- Computed using nested sampling (`MultiNest`, `PolyChord`, `dynesty`), simulated annealing (`emcee`), or from MCMC using `MCEvidence`.

Bayesian evidence $\mathcal{Z}$: Prior dependency

- Bayesian evidences are prior dependent:

$$\mathcal{Z} = \int P(D|\theta)P(\theta)d\theta \approx \langle \text{Likelihood} \rangle_{\text{Posterior}} \times \frac{\text{Posterior volume}}{\text{Prior volume}}$$

- They balance "goodness of fit" via likelihood with "complexity" through Occam penalty.
- Models that include too many fine-tuned parameters are disfavoured, unless they provide a much better fit.
- Corollary: Unconstrained parameters are not penalised.
- Widen prior $\Rightarrow$ reduce evidence (providing prior does not cut into posterior).
- Bayesians vs Frequentists $\leftrightarrow$ Feature vs Bug.

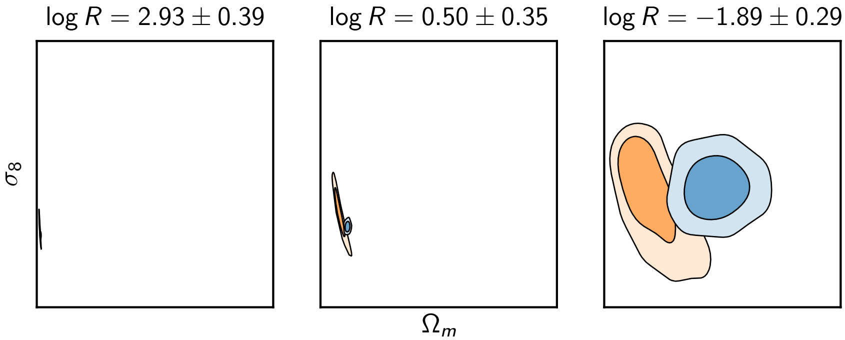
The meaning of the DES evidence ratio R

- The Dark Energy Survey collaboration (arXiv:1708.01530) quantify tension between two datasets A and B using the Bayes ratio:

$$R = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A\mathcal{Z}_B} = \frac{P(A\cap B)}{P(A)P(B)} = \frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

- R gives the relative change in our confidence in data A in light of having seen B (and vice-versa).
- $R > 1$ implies we have more confidence in A having received B.
- Like evidences, it is prior-dependent
- Increasing prior widths $\Rightarrow$ increasing confidence.

The DES evidence ratio R: Prior dependency



- What does it mean if increasing prior widths $\Rightarrow$ increasing confidence?
- Wide priors mean *a-priori* the parameters could land anywhere.
- We should be proportionally more reassured when they land close to one another if the priors are wide

How do we deal with the prior dependency in R?

Option 1 Take the Bayesian route, accept the prior dependency, and spend time trying to justify why a given set of priors are "physical".

Option 2 Try to find a principled way of removing this prior dependency

- One of the critical observations is that one can only hide tension by widening priors. Narrowing them will only ever show tension if it is present.
- If we could define "Narrowest reasonable priors" and find that $R < 1$, then this would indicate tension.

R: a Gaussian example

- Given two Gaussians with parameter means μ_A, μ_B and parameter covariances Σ_A, Σ_B and a prior with volume V_π:

$$\log R = -\frac{1}{2}(\mu_A - \mu_B)(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B) \\ + \log V_\pi - \log \sqrt{|2\pi(\Sigma_A + \Sigma_B)|}$$

- Like evidence, R composed of "Goodness of fit", and "Occam factor".
- Ideally want would remove this Occam factor (ratio of prior to posterior volume).

KL divergence $\mathcal{D}$, Information $\mathcal{I}$, suspiciousness S

- The KL divergence quantifies the compression from prior to posterior:

$$\mathcal{D} = \int P(\theta|D) \log \frac{P(\theta|D)}{P(\theta)} d\theta = \left\langle \log \frac{\text{Posterior}}{\text{Prior}} \right\rangle_{\text{Posterior}}$$

- It bears many similarities to an Occam factor, for a Gaussian:

$$\mathcal{D} = \log V_\pi - \log \sqrt{|2\pi\Sigma|} - \frac{1}{2} d$$

- Can define equivalent of R for KL divergence, the information ratio $\mathcal{I}$

$$\log R = \mathcal{Z}_{AB} - \mathcal{Z}_A - \mathcal{Z}_B$$
$$\log \mathcal{I} = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB}$$

- Subtracting the two removes prior dependency, giving suspiciousness:

$$\log S = \log R - \log \mathcal{I}$$

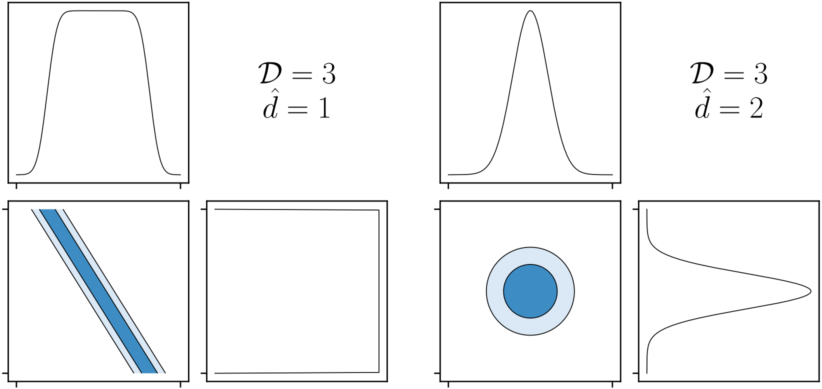
Suspiciousness S

- For a Gaussian:

$$\log S = \frac{d}{2} - \frac{1}{2}(\mu_A - \mu_B)(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B).$$

- We thus find that our original idea for tension $X_d^2 = d - 2\log S$.
- However S is composed of evidences $\mathcal{Z}$ and KL divergences $\mathcal{D}$, which are Gaussian-independent concepts.
- The only thing remaining to determine is d, the "number of parameters".

Dimensionality d



- Intuition should tell us that the d we need is the effective number of parameters (i.e. should not include unconstrained ones).
- Like the evidence, or the KL divergence, this "Model dimensionality" should be a sought-after inference quantity.

Dimensionality $\tilde{d}$

- KL divergence is the mean of the Shannon information I:

$$\mathcal{D} = \int P(\theta|D) \log \frac{P(\theta|D)}{P(\theta)} d\theta = \left\langle \log \frac{\text{Posterior}}{\text{Prior}} \right\rangle_{\text{Posterior}}$$

$$I = \log \frac{\text{Posterior}}{\text{Prior}}$$

- Model dimensionality proportional to variance of Shannon information:

$$\frac{\tilde{d}}{2} = \text{var}\left(\frac{\text{Posterior}}{\text{Prior}} \right)_{\text{Posterior}}$$

- Examples from real data:

$$\tilde{d}_{\text{Planck}} = 15.8 \pm 0.3 \qquad (21)$$
$$\tilde{d}_{\text{DES}} = 14.0 \pm 0.3 \qquad (26)$$
$$\tilde{d}_{\text{BAO}} = 2.95 \pm 0.07 \qquad (6)$$
$$\tilde{d}_{\text{S}H_0\text{ES}} = 0.93 \pm 0.03 \qquad (6)$$

Headline results

- Can calibrate X_d^2 as on the same scale as χ_d^2 to give a p-value-like quantity, termed "Tension probability" p

$$\begin{aligned} \text{Planck+BAO}: &\qquad p = 42 \pm 4\% \\ \text{Planck+DES}: &\qquad p = 3.2 \pm 1.0\% \\ \text{Planck+S}H_0\text{ES}: &\qquad p = 0.25 \pm 0.17\% \end{aligned}$$

- Under this metric, SH_0ES is unambiguously inconsistent, although not quite as brutal as $> 4\sigma$. BAO is consistent, and *DES* is inconsistent, but only just. This is pleasingly similar to ones intuition.

Extensions

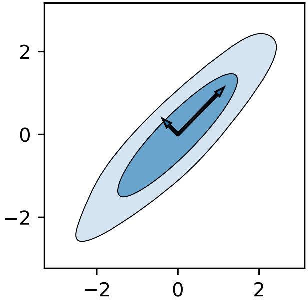
- In light of these results, there are two natural questions to ask:

1. $\tilde{d}_{\mathrm{BAO}} = 2.95 \pm 0.07$ out of a possible 6. Which ~ 3 are these?
2. Is there a direction in parameter space which is "most in tension"

- These are questions which people would usually answer with a PCA-type approach.

The problem with Principle Component Analysis

- Compute eigenvectors and eigenvalues of covariance matrix.
- These aim to describe "directions" in parameter space
- This procedure is not covariant:



"Principal Component Analysis" is a dimensionally invalid method that gives people a delusion that they are doing something useful with their data. If you change the units that one of the variables is measured in, it will change all the "principal components"! It's for that reason that I made no mention of PCA in my book. I am not a slavish conformist, regurgitating whatever other people think should be taught. I think before I teach. David J C MacKay.

Conclusions

- The DES ratio R is a principled thing to work with, but its prior dependency must be acknowledged
- Using KL divergences and model dimensionalities, R may be calibrated into something akin to the tension we desire.
- The "inference triple" of $\mathcal{Z}, \mathcal{D}, \tilde{d}$ should be considered in all model comparison analyses.
- All three can be computed from nested sampling runs using the `anesthetic` package.
- Be careful when applying principle component analysis!