Cosmological models

Erik Andersson, Sebastian Holmin

November 2020

In this project, we will use utilize the Supernova Cosmology Project (SCP) data, to analyze and compare cosmological models. The SCP 2.1 dataset contains detailed measurement of the redshift, z, and distance moduli μ, of several supernovea, which we will utilize to perform Bayesian parameter estimations and model comparisons.

1 Parameter estimation

In a flat universe, and in the small redshift z regime (about $z \leq 0.5$), one can approximately relate the distance to a star with the redshift of the incoming light through

$$d_L(z) = c(1+z)\int_0^z \frac{dz'}{H(z')} \approx \frac{c}{H_0}\left(z + \frac{1}{2}(1-q_0)z^2 + \dots\right). \tag{1}$$

We will use Bayesian parameter estimation to extract posterior probability distributions for the parameters H_0 and q_0 through the SCP 2.1 dataset, which contains redshift, distance modulus, and distance modulus error. The distance modulus is related to the distance (in Mpc) by

$$\mu = 5\log_{10}(d_L) + 25.$$

We will use MCMC sampling through the package `emcee` to sample the posterior

$$p(H_0, q_0 \mid \mathcal{D}) = \int \frac{p\left(\mathcal{D} \mid H_0, q_0, \sigma^2\right) p\left(H_0, q_0, \sigma^2\right)}{p(\mathcal{D})} d\sigma^2 \tag{2}$$

where the marginalization is automatically taken care of by the `emcee` sampler function and the evidence $p(\mathcal{D})$ is disregarded since it is only a constant normalizing factor and the sampler normalizes automatically. We assume that the likelihood for the data is normally and independently distributed with weights proportional to the inverse measurement errors squared, i.e.

$$\begin{aligned} p\left(\mathcal{D} \mid H_0, q_0, \sigma^2\right) &= \mathcal{N}\left(\mathcal{D} \mid \mu(\mathcal{Z}, H_0, q_0), \sigma^2\mathbf{W}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{N_d/2}|\mathbf{W}|^{-1/2}} \exp\left\{-\frac{1}{2\sigma^2}(d_i - \mu(z_i, H_0, q_0))^2 w_i\right\}, \end{aligned}$$

where $\mathcal{D} = \{d_i\}_{i=0}^{N_d}$ is the data for the distance modulus, $\mathcal{Z} = \{z_i\}_{i=0}^{N_d}$ are the corresponding redshifts and $\boldsymbol{W}$ is diagonal with entries $\{w_i \propto 1/\sigma_{\mathrm{e},i}^2\}_{i=0}^{N_d}$. For the prior we use $p(H_0, q_0, \sigma^2) = \mathcal{U}(H_0)\mathcal{U}(q_0)\mathcal{IG}(\sigma^2|\alpha, \beta)$, where we again do not care about the overall normalization. We initially

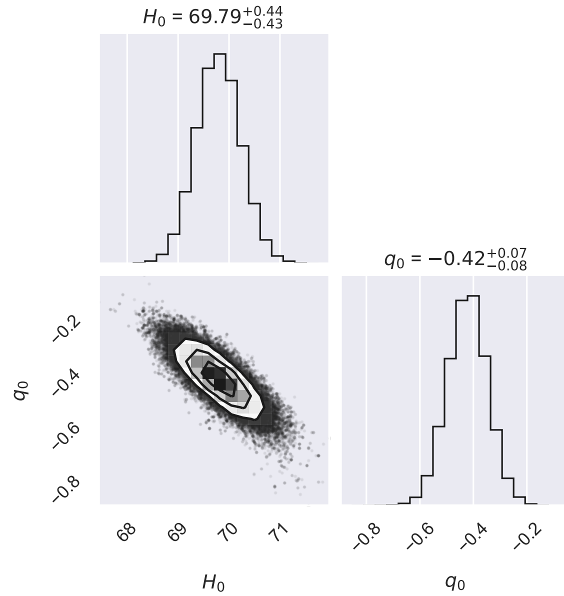

Figure 1: The joint distribution (bottom left) and the corresponding marginal distributions (top and right) for the Hubble constant H_0 [km/(s Mpc)] and the deceleration parameter q_0 [a.u.] using 40 walkers with 4000 steps each. The error estimation roughly corresponds to a central 1σ interval, i.e. a 68% credibility interval where the limits are such that the probability is split equally above and below the mean.

chose values for α and β such that σ^2 is expected to lie close to the square of the experimental uncertainty by setting the mode of $\mathcal{IG}$ to 0.2 and the mean at 1 so the pdf would be wide and lenient. We also verified that this choice didn't have any significant impact on the resulting distributions for H_0 and q_0 by simply running a few samplings using slightly shifted values for α and β.

The resulting posterior, using only data points with $z \leq 0.5$, is shown in a corner plot in fig. 1. The error estimation is a central 1σ Bayesian credibility interval, i.e. a 68% credibility interval where the limits are such that the probability is split equally above and below the mean. The MCMC sampling was done with 40 walkers of 4000 steps each and a burn in period of 100 steps. The result is very robust and looks practically identical even with a much smaller sample size (20 walkers of 2000 steps). A value $q_0 < 0$ implies an accelerating universe and the evidence for this is very convincing seeing as the entire distribution for q_0, sampled ~160000 times, is located below zero.

We can verify our calculation of H_0 by looking at the very low limit of z and dropping the second order term in eq. (1), which gives us

$$H_0 = cz \cdot 10^{(25-\mu)/5} \text{ (for } z \approx 0). \tag{3}$$

Plotting this for decreasing upper limits on the z data gives us the plot in fig. 2. This becomes a better approximation the lower the limit, but also more unstable from the lack of data. Nevertheless,

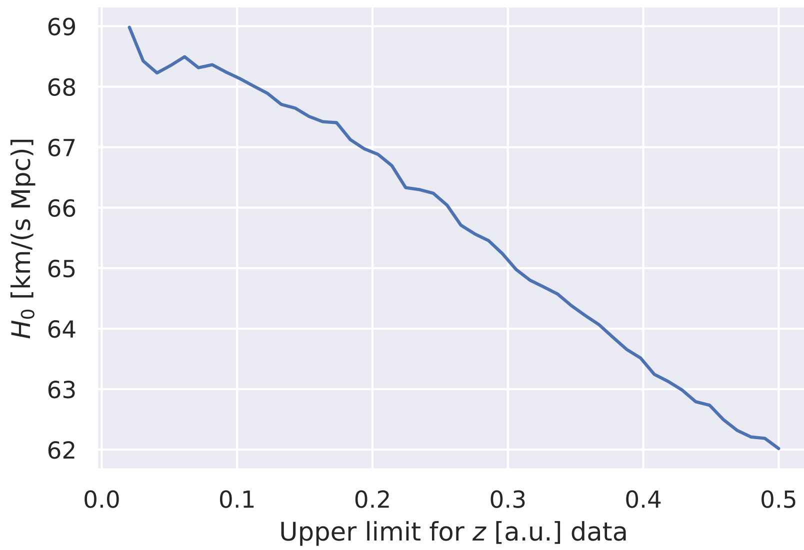

Figure 2: H_0 value for the very small z approximation eq. (3), given for a range of upper limits on the z data.

we can see that the curve roughly trends towards the value of $H_0 = 69.8 \pm 0.4$ km/(s Mpc) given in fig. 1.

We check the model using the sampled posterior distributions with a posterior predictive plot shown in fig. 3. This was done by calculating the model prediction for the sampled H_0 and q_0 values and then plotting a central 99% credibility interval. The model seems to work relatively well for $z < 0.5$, which is the data it was fitted against, but less so in the higher $z > 0.5$ region. We calculated that about 56% of the data point error intervals overlap with the mean prediction in the $z > 0.5$ region, compared to 72% for $z < 0.5$ (as a reference, this value should be 68% for the true curve and an infinite amount of data points).

Clearly we would need a model more suited for higher z value, e.g. more terms in eq. (1), along with more data at high z values to improve the inference of the H_0 and q_0 parameters.

2 Comparing ΛCDM and wCDM

To investigate the two models, ΛCDM and wCDM, we use the full expression in eq. (1), i.e.

$$d_L(z) = \frac{c(1+z)}{H_0} \int_0^z \frac{dz'}{\sqrt{E\left(z'\right)}}.$$

where $H(z) = H_0\sqrt{E(z)}$, and we will use $H_0 = 70$ km/(s Mpc) for which the SCP dataset is calibrated. In a flat universe the Friedman equation can be written as

$$1 = \Omega_M + \Omega_\Lambda \tag{4}$$

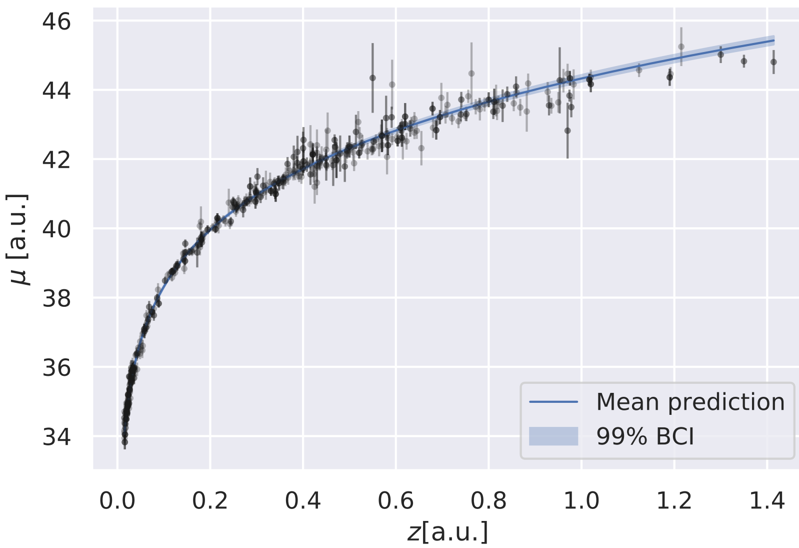

Figure 3: Posterior predictive for the physical model, eq. (1), using parameters from the MCMC sampling of eq. (2).

in terms of the density parameters

$$\Omega_M \equiv \frac{\rho_M}{\rho_c}, \quad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c}.$$

The models ΛCDM and wCDM can then be summarized by

$$\begin{aligned} E_{\Lambda\mathrm{CDM}} &= \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0} \\ E_{w\mathrm{CDM}} &= \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}(1+z)^{3(1+w)} \end{aligned}$$

where we can use eq. (4) to eliminate $\Omega_{\Lambda,0}$, thus the models have one and two free parameters respectively. Given these models, we simply utilize the `minimize` (single paramater) and `differential_evolution` (multiple paramaters) of `scipy.optimize` to optimize the log likelihood

$$\log\left[p(\mathcal{D}|\Omega_M, w)\right] \propto -\frac{1}{2}\sum \frac{(d_i - \mu(z_i, \Omega_M, w))^2}{\sigma_i^2}.$$

where setting $w = -1$ constant gives the likelihood for the ΛCDM model.

Using the optimized parameter values we then extract the maximum log-likelihood for the different models and calculate the Bayesian[3] and Akaike[1] information criteria

$$\begin{aligned} \mathrm{BIC} &\equiv 2\log\left[p\left(\mathcal{D} \mid \boldsymbol{\theta}_\star\right)\right] - N_p \log\left(N_d\right) \\ \mathrm{AIC} &\equiv 2\log\left[p\left(\mathcal{D} \mid \boldsymbol{\theta}_\star\right)\right] - 2N_p. \end{aligned}$$

In this context of comparing models, as we cannot determine the overall scale of the variance we will not take it into account and just set it to 1. This may impact the balance between the max likelihood term and the penalty term in the information criteria, which may in turn affect the comparison of the models as they have a different number of parameters.

The result is shown in table 1, we can clearly see that both information criteria favor the simpler ΛCDM, since the score is higher for both, which is to be expected as they perform almost identically despite wCDM having an extra parameter. In fact, using the given data, the wCDM model nearly reverts to ΛCDM as it does analytically when $w_{w\text{CDM}} = -1$. Also, note that the difference in score comes almost entirely from the parameter penalization term, i.e. the maximum likelihood is the same for both models but the more complex one is penalized for having an extra parameter. This means that neglecting the overall scale of the variance has no effect in this case. These values are close to the values $\Omega_M = 0.2855$ and $w_{w\text{CDM}} = -1.073$ found by [2], which also takes into account other data such as CMB (Cosmic Microwave Background).

Table 1: Information criteria and parameters for the two competing models. Comparisons can only be made horizontally, between models, and higher scores imply a better model.

	ΛCDM	wCDM
AIC	-18.78	-20.78
BIC	-23.14	-29.51
Ω_M	0.278	0.280
$w_{w\text{CDM}}$	–	-1.004

The posterior probability for Ω_M for the ΛCDM model is shown in fig. 4. This distribution was sampled using 20 walkers of 2000 steps each and a burn-in period of 100 steps. The prior used for $\Omega_{M,0}$ was simply a uniform one of height 1 in the interval $[0, 1]$ (values outside this range are unphysical). For the variance, the measurement errors squared was used.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[2] G. Hinshaw, D. Larson, E. Komatsu, D. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, L. Page, K. M. Smith, J. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. Meyer, G. Tucker, Edward J. Wollack, and E. L. Wright. Nine-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological parameter results. *Astrophysical Journal Supplement Series*, 208:19, 2013.

[3] Gideon Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2):461–464, 03 1978.

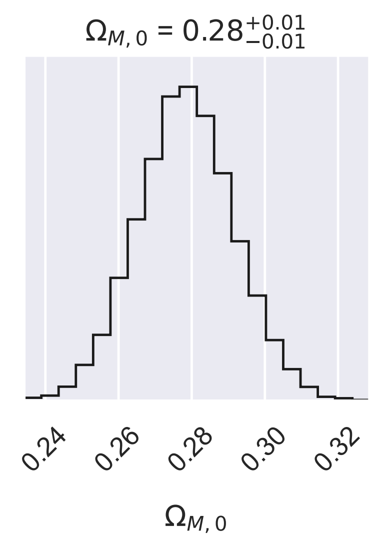

Figure 4: Posterior probability distribution for the matter density parameter $\Omega_{M,0}$ [a.u.] in the ΛCDM model sampled using 20 walkers of 2000 steps. The errors define a central 68% credibility interval as before.