# Project 1: Cosmological Models
## TIF345: Advanced Simulations and Machine Learning


Simon Josefsson and Tomas Lundberg*

*Chalmers University of Technology*


(Dated: December 8, 2021)

### I. Introduction

Measuring distances to heavenly bodies is difficult [1]. One reliable way to do this, however, is with the help of Type Ia supernovae (SNIa). These cosmic explosions are standardized events in the sense that they produce a fairly fixed peak luminosity and can thus be used as standard candles. If we measure the intensity on earth we can infer the (luminosity) distance $d_L$ to the supernova since we know the peak luminosity produced by these types of events. Furthermore, by measuring the spectrum we can in addition obtain data on the cosmological redshift, $z$, as a result of the expansion of the universe. Finally, if we have data of the distance over a wide range of different redshifts, corresponding to different time instances in the age of the universe, we can obtain information on if the expansion of the universe is accelerating or decelerating, as well as test different cosmological models.

In this report we make use of data from the *Supernova Cosmology Project* (SCP) which contains data of the redshift $z$, distance modulus $\mu$ defined as

$$\mu = 5log_{10}(d_L) + 25, \tag{1}$$

and the heteroscedastic errors on $\mu$, $\sigma_\mu$ of $N_d$ =580 different SNIa. In Fig. 1, we display $\mu$ for different values of $z$. We use this data to first infer the value of the Hubble constant $H_0$ and the deceleration parameter $q_0$ in the low $z$ regime ($z < 0.5$). Thereafter, we put two different cosmological models, $\Lambda$CDM and $w$CDM, to test for the full range of $z$-values.

### II. Background

#### A. Cosmology and distances

Following [2] the Hubble parameter as a function of the cosmological redshift $z$ can be written as

$$H(z) = H_0 E(z)^{1/2}, \tag{2}$$

where $E(z)$ depends on our cosmological model. Note that $H(0) = H_0$, i.e. the Hubble parameter today. If we assume a flat universe we can express the luminosity distance as

$$d_L(z) = c(1+z)\int_0^z \frac{dz'}{H(z')}, \tag{3}$$

where $c$ is the speed of light. In the low $z$ regime ($z$<0.5) we can Taylor expand $E(z)$ and obtain

$$d_L \approx \frac{c}{H_0}\left(z + \frac{1}{2}(1-q_0)z^2 + ...\right) \tag{4}$$

where $q_0 > 0$ indicates a decelerating universe expansion and $q_0 < 0$ an accelerating one.

Moreover, The Hubble parameter can be associated with densities of different type of energy; matter, radiation and dark energy. The proportions of which are captured by their corresponding density parameters $\Omega_M$, $\Omega_k$ and $\Omega_\Lambda$ respectively. Of course, we have the constraint that

$$1 = \Omega_M + \Omega_k + \Omega_\Lambda. \tag{5}$$

Again, if we limit ourselves to a flat universe where $\Omega_{k,0} = 0$ the cosmic model $\Lambda$CDM gives

$$E(z) = \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}. \tag{6}$$

Another cosmic model is the $w$CDM

$$E(z) = \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}(1+z)^{3(1+w)}, \tag{7}$$

where we have introduced dark energy equation of state with the additional parameter $w$. Under the assumption that the universe is flat we have that

$$q_0 = \frac{\Omega_{M,0} - 2\Omega_{\Lambda,0}}{2}. \tag{8}$$

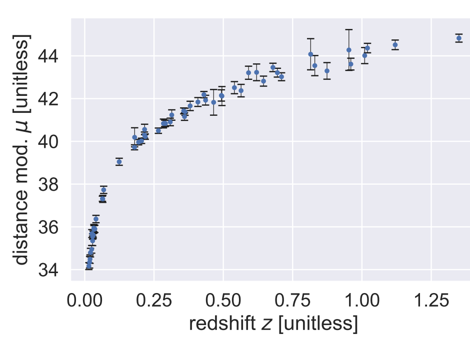


FIG. 1: The distance modulus $\mu$ against the redshift $z$ for every tenth (for a clearer display) point in the dataset with corresponding error bars.


* simjos@student.chalmers.se, lutomas@student.chalmers.se

TABLE I: Information of the four different MCMC runs (*col* 1) made in this report. *col* 2 indicates the range of $z$-values used. *cols* 3 to 7 shows the sampled parameters, what they are initialized to (*init*) and their priors (*prior*), where applicable. $N$, $U$, $IG$ indicates normal, uniform and inverse gamma distributions. *col* 8-10 are details for the MCMC settings: $N_W$ is the number of walkers, steps taken per walker, and burn-in the discarded samples of each chain.

| run | $z$ | $H_0$ [km/s/Mpc] | | $q_0$ [unitless] | | $\sigma^2$ [unitless] | | $\Omega_{M,0}$ [unitless] | | $w$ [unitless] | | $N_w$ | steps | burn-in |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | init | prior | init | prior | init | prior | init | prior | init | prior | | | |
| 1 | $< 0.5$ | $N(100,10)$ | $U(0,1000)$ | $N(0,1)$ | $U(-10,10)$ | $N(0.1,0.01)$ | $IG(0.01,0.1)$ | - | | - | | 40 | 4000 | 150 |
| 2 | $< 0.05$ | $N(100,10)$ | $U(0,1000)$ | - | | $N(100,10)$ | $IG(0.01,0.1)$ | - | | - | | 40 | 4000 | 100 |
| 3 | full | 70 | | Eq. (8) | | 1 | | $N(0.5,0.1)$ | $U(0,1)$ | - | | 20 | 2000 | 50 |
| 4 | full | 70 | | Eq. (8) | | 1 | | $N(0.5,0.1)$ | $U(0,1)$ | $N(0,1)$ | $U(-10,10)$ | 20 | 2000 | 150 |

### B. Bayes' Theorem

Given a parmeter vector $\theta$ that we want to infer from data $D$ given our prior knowledge $I$ we can express the posterior probability density function (pdf) through Bayes' Theorem

$$\mathrm{p}(\theta \mid D, I) = \frac{\mathrm{p}(D \mid \theta, I)\,\mathrm{p}(\theta \mid I)}{\mathrm{p}(D \mid I)}, \tag{9}$$

where $\mathrm{p}(D \mid \theta, I)$ is the *likelihood*, $\mathrm{p}(\theta \mid I)$ the *prior* and $\mathrm{p}(D \mid I)$ the *evidence*. If we are only interested in the posterior pdf we can note that the evidence is only a normalizing factor and can thus write

$$\mathrm{p}(\theta \mid D, I) \propto \mathrm{p}(D \mid \theta, I)\,\mathrm{p}(\theta \mid I). \tag{10}$$

Furthermore, for simplicity and higher numerical stability we can express the log posterior through

$$\log \mathrm{p}(\theta \mid D, I) \propto \log \mathrm{p}(D \mid \theta, I) + \log \mathrm{p}(\theta \mid I). \tag{11}$$

#### 1. Priors

In this report we will make use of two different priors, a uniform prior and an inverse gamma prior. The uniform prior for a parameter $\theta$ is simply defined as

$$\mathrm{p_U}(\theta \mid I) = \begin{cases} 1 & \text{if } \theta \in [\theta_{min}, \theta_{max}] \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The inverse gamma prior for parameter $\theta$ is

$$\mathrm{p_{IG}}(\theta \mid I) = IG(\theta; a_0, b_0), \tag{13}$$

where $a_0$ and $b_0$ are the shape and scale parameters of the inverse gamma distribution.

#### 2. Likelihood

As we mentioned in Section I, the data of the distance modulus comes with a corresponding error $\sigma_\mu$. We want the data points that have the least uncertainty to have a higher weight in our likelihood. We will therefore apply weights that are inversely proportional to $\sigma_\mu^2$ and normalized such that $\sum_{i=1}^{N_d} w_i = N_d$, where $N_d$ is the number of data points. In other words, the errors are distributed as $\varepsilon_i \sim N(0, \sigma^2/w_i)$. With this established, we define our likelihood as

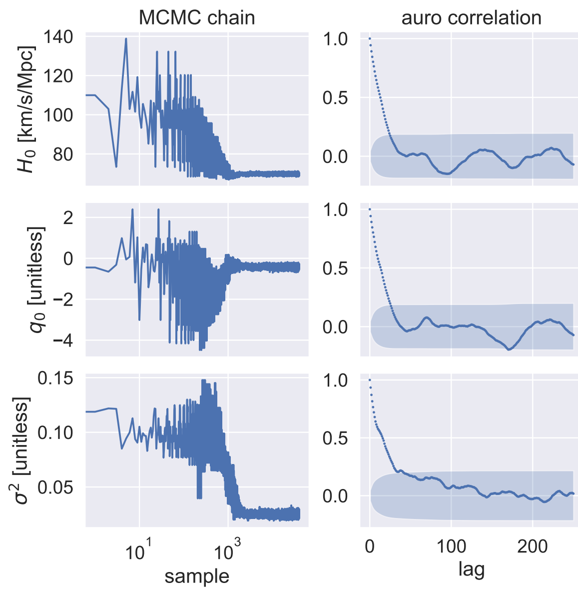


FIG. 2: The MCMC chain (*left*) and auto correlation (*right*) of a walker of run 1 (see Table I) for $H_0$ (*top*), $q_0$ (*middle*) and $\sigma^2$ (*bottom*). The shaded region of the auto correlation indicates 95% confidence interval.

$$\mathrm{p}(D \mid \theta, I) = \prod_{i=1}^{N_d} \frac{1}{\sqrt{2\pi\sigma^2/w_i}} \exp\left\{-\frac{w_i}{2\sigma^2}(\mu_i - \hat{\mu}(\theta; z_i)^2\right\}, \tag{14}$$

where $\mu_i$ is the $i$:th distance modulus in our data and $\hat{\mu}(\theta; z_i)$ our model value at $z_i$ given parameters $\theta$ and transformed through Eq. (1). The corresponding log likelihood becomes

$$\ln \mathrm{p}(D|\theta) = -\frac{1}{2}\sum_{i=0}^{N_d}\left[\frac{w_i}{\sigma^2}(\mu_i - \hat{\mu}(\theta; z_i))^2 - \ln(2\pi\sigma^2/w_i)\right]. \tag{15}$$

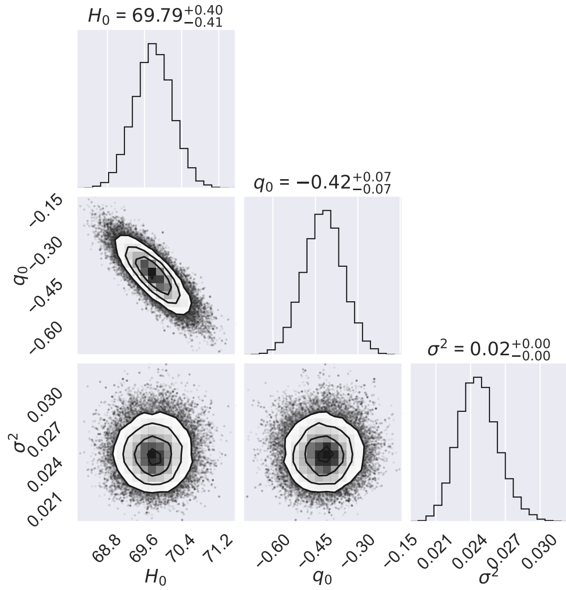


FIG. 3: Corner plot of run 1 (see Table I) with the joint posterior distribution for parameters $H_0$ [km/s/Mpc], $q_0$ [unitless] and $\sigma^2$ [unitless], for $z < 0.5$ with median± standard deviation values indicated above the marginalized posterior distributions.

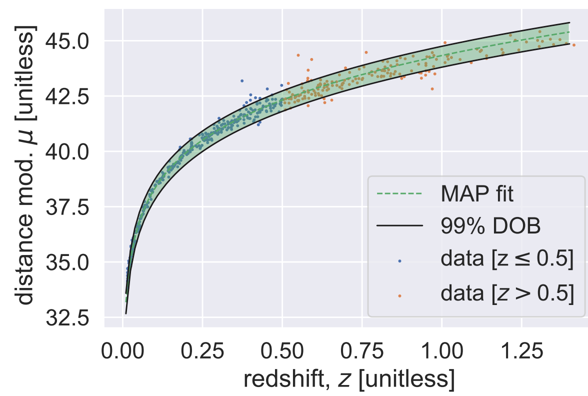


FIG. 4: Our model with the fitted MAP parameters and a 99% degree of belief (DOB) for the distance modulus versus redshift. Including the data for our z-regime and the entire z-regime.

### C. Model comparison

To compare models we take advantage of two different types of information criteria (IC), the Bayesian IC (BIC) and the Akaike IC (AIC) [2]. These are only valid in the large data limit and we will assume that our $N_d = 580$ will be sufficiently close to this. The BIC and AIC are defined as

$$AIC = 2\ln[\mathrm{p}(D|\theta_*)] - 2N_p, \tag{16}$$

and

$$BIC = 2\ln[\mathrm{p}(D|\theta_*)] - N_p \ln(N_d), \tag{17}$$

where $N_p$ is the number of parameters. This gives a score for each of the models given their maximum likelihood, $p(D|\theta_*)$, i.e. where $\theta_*$ is the max a posteriori model values. The higher the score the better.

## III. Methods

### A. Low $z$ regime

To gain a first insight of the acceleration/deceleration of the expansion of the universe we numerically calculate the joint distribution of $H_0$, $q_0$ together with the error scale parameter $\sigma^2$ in the low $z$ regime. To this end, we make use of $d_L$ as described by Eq. (4) up to and including the second order. For $H_0$ and $q_0$ we adopted a uniform prior, (12) with $H_{0,min} = 0$, $H_{0,max} = 1000$ (km/s/Mpc) and $q_{0,min} = -10$ and $q_{0,max} = 10$ respectively. Moreover, the prior for $\sigma^2$ was assumed to be inverse gamma, Eq. (13), with parameters $a_0 = 0.01$ and $b_0 = 0.1$. We obtained the joint posterior distribution for $H_0$, $q_0$ and $\sigma^2$ using Markov Chain Monte Carlo (MCMC) sampling. In Table I, we summarize the relevant information of the MCMC simulation.

Finally, to check the extraction of $H_0$, we limited ourselves to $z < 0.05$ such that we could drop the second order term in Eq. (4) and thereby also $q_0$. Using the same settings for the MCMC sampling, see Table I, yielded the posterior for $H_0$ and $\sigma^2$ for $z < 0.05$.

### B. Different cosmologies

Now, assuming $H_0 = 70$ km/s/Mpc and $\sigma^2 = 1$ (with $w_i = 1/\sigma^2_{i,\mu}$) for the whole range of $z$ values, i.e. the full dataset, we compared two different cosmological models, namely, ΛCDM and $w$CDM. These models have expressions for $d_L$ through the integral in Eq. (3) combined with their corresponding expressions for $E(z)$ given by Eq. (6) for ΛCDM and Eq. (7) for $w$CDM. Note that in the assumed flat universe, $\Omega_{M,0} = 1 - \Omega_{\Lambda,0}$. Using the `quad` module of `scipy` we could numerically integrate Eq. (3) and, similarily as before, obtain the posteriors $\mathrm{p}_{\Lambda CDM}(\Omega_{M,0} \mid D, I)$ and $\mathrm{p}_{wCDM}(\Omega_{M,0}, w \mid D, I)$ for the corresponding models using MCMC sampling. We use uniform priors for $\Omega_{M,0}$ and $w$ where $\Omega_{M,0}, \Omega_{\Lambda,0} \in [0,1]$ and $w \in [-10, 10]$. See Table I for details on the MCMC simulations.

For simplicity we used `scipy` to obtain the maximum log likelihood values for the computation of the AIC and the BIC, as explained in Section II C, that were used to compare the two models.

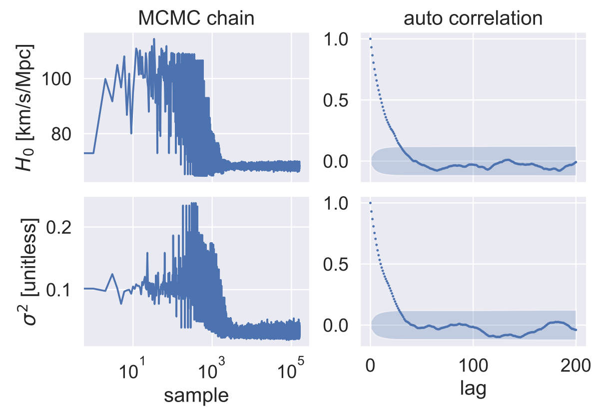


FIG. 5: The MCMC chain (*left*) and auto correlation (*right*) of a walker of run 2 (see Table I) for $H_0$ (*top*) and $\sigma^2$ (*bottom*). The shaded region of the auto correlation indicates 95% confidence interval.

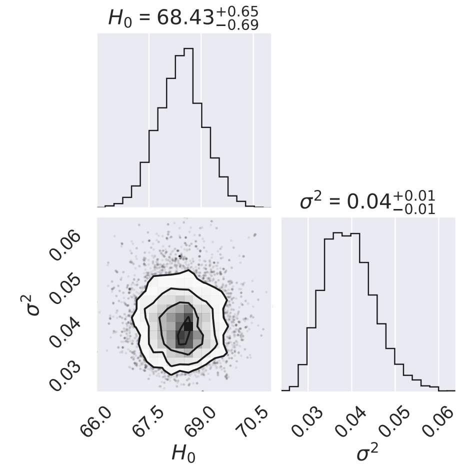


FIG. 6: Corner plot of run 2 (see Table I) with the joint posterior distribution for parameters $H_0$ [km/s/Mpc] and $\sigma^2$ [unitless], for $z < 0.05$ with median$\pm$ standard deviation values indicated above the marginalized posterior distributions.

## IV. Results and discussion

### A. Low $z$ regime

First, in Fig. 2, we see the full MCMC chains (*left*) for run 1 from Table I together with the auto correlation (*right*) of a single chain for all three parameters $H_0$, $q_0$ and $\sigma^2$. The chains should motivate a burn-in of 150 and the short decoherence length indicates a fair sampling.

In Fig. 3 we see a corner plot of the joint posterior pdf of $H_0$, $q_0$ and $\sigma^2$ in the low $z$ regime. We see that the *max a posteriori* (MAP) values are $H_0 = 69.79$ km/s/Mpc, $q_0 = -0.42$ and $\sigma^2 = 0.02$. Also, note the rather tight margin of error for $H_0$ and $q_0$ as indicated by the standard deviations. First, the negative value of $q_0$ indicates that the expansion of the universe is indeed accelerating, as is confirmed by the literature [3]. Second, the value of $H_0$=70 km/s/Mpc achieved in the SCP is included in the standard deviation interval which indicates confidence in our simulation. In our inference of $H_0$ and $q_0$ we only make use of $z < 0.5$ in an approximate model. To better improve the inference of $H_0$ and $q_0$ we could try and make use of the full dataset with for instance $w$CDM or $\Lambda$CDM. Another way would be to put more work in specifying a prior that better represents are knowledge, not just a uniform prior. However, given that we have so much data, this probably would not improve the inference by too much.

Using the MCMC samples of $H_0$, $q_0$ and $\sigma^2$ for our model and errors we plot a posterior prediction for the full $z$ interval which can be seen in Fig. 4. That is, we plot $\mu(z) = M(H_{0,i}, q_{0,i}; z) + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$ for every $i$:th sample of the MCMC chain. We display the 99% degree of belief (DOB). The predictive model optimized in the low $z$-regime predicts the whole data set fairly well as shown in Fig. 4. However, we show a 99% DOB but there are more data points outside our DOB than would be expected, especially for higher $z$. We could improve this in (at least) two ways. We assumed $\varepsilon_i \sim N(0, \sigma^2)$ but, as hinted by Fig. 1, the errors are larger for higher values of $z$, we could try and model this fact to obtain better predictions with $\varepsilon(z)$. Furthermore, we have assumed the model discrepancy $\delta(z_i)$ to be 0, i.e. that our models are perfect. This clearly is not 100% true, and thus we could choose to model this discrepancy as well. One way to implement the model discrepancy would be to assume that $\delta_i(z)$ is normally distributed and use Gaussian processes to describe random functions, which lets us avoid having a specific function for the model discrepancy. We could even improve the model discrepancy if we have prior knowledge of the problem, e.g $\delta(0) = 0$. To summarize, we could improve the predictions and obtain a more realistic DOB with

$$\mu(z) = M(\theta; z) + \varepsilon(z) + \delta(z) \tag{18}$$

where $M(\theta; z)$ is our model given our parameters $\theta$ ($H_0$ and $q_0$ in the case of Fig. 4).

For run 2 in Table I, with $z < 0.05$, we present the MCMC chains together with the auto correlation of $H_0$ and $q_0$ where we used a burn-in of 100 in Fig. 5. We show a corner plot in Fig. 6 where we can see that we obtain a similar distribution of $H_0$ as when we included the second order of $z$. This tells us that the Taylor expansion in Eq. (4) is a sufficiently good approximation.

### B. Different cosmologies

In Fig. 7 and Fig. 8 we show the MCMC chains and the auto correlations for the corresponding parameters of run

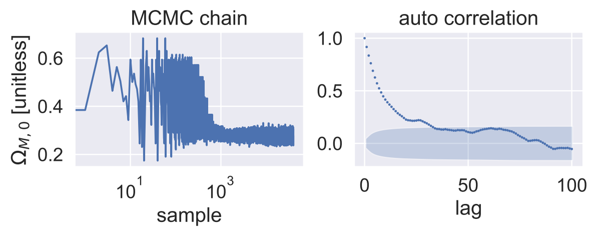


FIG. 7: The MCMC chain (*left*) and auto correlation (*right*) of a walker of run 3 (see Table I) for $\Omega_{M,0}$. The shaded region of the auto correlation indicates 95% confidence interval.

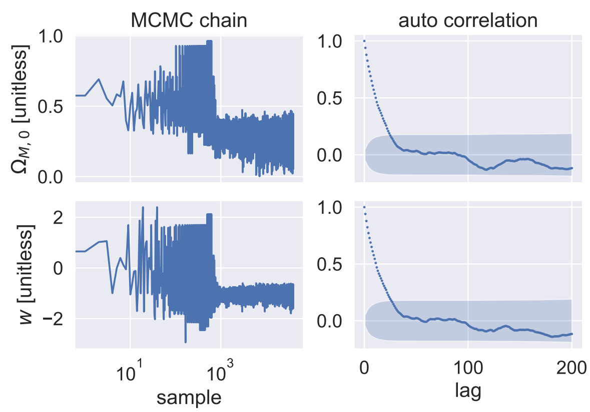


FIG. 8: The MCMC chain (*left*) and auto correlation (*right*) of a walker of run 4 (see Table I) for $\Omega_{M,0}$ (*top*) and $w$ (*bottom*). The shaded region of the auto correlation indicates 95% confidence interval.

3 and run 4 in Table I. These figures should serve as a motivation of the burn-ins of 50 and 150 respectively as well as accurate samplings.

The posterior distributions $p_{\Lambda CDM}(\Omega_{M,0} \mid D,I)$ and $p_{wCDM}(\Omega_{M,0}, w \mid D,I)$ are shown in Fig. 9 and Fig. 10 respectively. Both models results in the same MAP values for $\Omega_{M,0} = 0.28$ (and $\Omega_{\Lambda,0}$) even though the error of $\Lambda$CDM is smaller. Our simple analysis yields results that are close to those found in the literature, [3] reports $\Omega_{M,0}$=0.308 $\pm$ 0.012, that is, a dark energy dominated universe. The discrepancy might be due to different data and or different models. Moreover the MAP value for $w$ in Fig. 10 is also in agreement with the literature: $w = -1.01 \pm 0.04$. This is very close the value $w = -1$ where $w$CDM reduce to $\Lambda$CDM. Using the MCMC samplings for $\Omega_{M,0}$ we calculated $q_0$ with Eq. (8) for comparison with run 1 and obtained $q_0^{\Lambda CDM} = -0.58 \pm 0.008$ and $q_0^{wCDM} = -0.58 \pm 0.04$, that is, slightly smaller value of $q_0$, indicating a faster acceleration of the expansion of the universe.

TABLE II: The parameter values (*col* 2-3) obtained through maximum likelihood (*col* 4) for the two tested cosmological models $\Lambda$CDM and $w$CDM together with the BIC and AIC (*col* 5-6).

| Model | $\Omega_{M,0}$ | $w$ | ML | BIC | AIC |
|---|---|---|---|---|---|
| $\Lambda$CDM | 0.278 | - | 118.740 | 231.1 | 235.5 |
| $w$CDM | 0.280 | -1.004 | 118.741 | 224.8 | 233.5 |

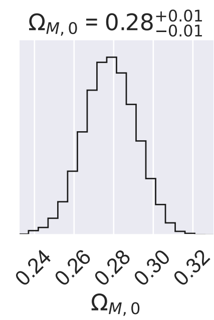


FIG. 9: Corner plot of run 3 (see Table I) with the posterior distribution for parameters $\Omega_{M,0}$ [unitless], with median$\pm$ standard deviation values indicated above the posterior distribution.

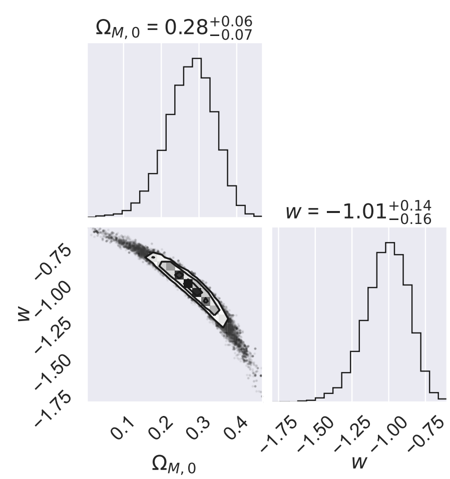


FIG. 10: Corner plot of run 4 (see Table I) with the joint posterior distribution for parameters $\Omega_{M,0}$ [unitless] and $w$ [unitless], with median$\pm$ standard deviation values indicated above the marginalized posterior distribution.

Finally, in Table II, we present the parameter values of the two models together with the maximum likelihood (ML) and the BIC and AIC values. We note that the two models have almost identical maximum likelihoods why the difference in the BIC and AIC scores, see Section II C, comes down to the fact that $w$CDM has an additional parameter for which a penalty is payed. Therefore, we can express a slightly higher preference for $\Lambda$CDM. Also, rembember that for $w = -1$, $w$CDM reduces to $\lambda$CDM and so the more complex $w$CDM does not seem to make use of its additional parameter, supporting the conclusion of the IC scores showed in Table II. However, remember that we assumed the error scale to be known as $\sigma^2 = 1$ which effects the value of the maximum likelihood. To obtain more accurate results we could try and sample the error scale as well.

[1] M. Thabet, SSRN Electronic Journal (2013), 10.2139/ssrn.2355164.
[2] A. Ekström, "Tif345/fym345 project 1: Cosmological models," (2020).
[3] J. Rademacker, Physical Review D **98**, 030001 (2018).