

Testing the mutual consistency of different supernovae surveys

N.V. Karpenka¹, F. Feroz² and M.P. Hobson²

¹The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden
email: nkarp@fysik.su.se

²Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge CB3 0HE, UK

Abstract. It is now common practice to constrain cosmological parameters using supernovae (SNe) catalogues constructed from several different surveys. Before performing such a joint analysis, however, one should check that parameter constraints derived from the individual SNe surveys that make up the catalogue are mutually consistent. We describe a statistically-robust mutual consistency test, which we calibrate using simulations, and apply it to each pairwise combination of the surveys making up, respectively, the UNION2 catalogue and the very recent JLA compilation by Betoule *et al.* We find no inconsistencies in the latter case, but conclusive evidence for inconsistency between some survey pairs in the UNION2 catalogue.

Keywords. methods: data analysis — methods: statistical — supernovae: general

1. Introduction

The original discovery of the accelerated expansion of the universe (Riess *et al.* 1998; Perlmutter *et al.* 1999), prompted many other teams to start searches for high redshift SNe. This resulted in SNe compilations containing data from different telescopes. Although such analyses result in tighter constraints on parameters, one needs to be careful when using data from different sources, as the presence of unaccounted systematics in any of these data-sets can lead to misleading results. Therefore, it is extremely important to establish whether different data-sets are consistent with each other before performing a joint analysis using them.

A common method to check for consistency between different data-sets, in general, is to perform a χ^2 analysis and compare the best-fit values χ_{\min}^2 . It is, however, impossible to perform such a test in the standard framework of SNe analysis. Moreover tests based on χ_{\min}^2 depend only on the best-fit parameters, and are insensitive to the likelihood over the rest of the parameter space. Thus one must seek an alternative test for testing consistency between SNIa data-sets. We use a method based on the Bayesian consistency test described in Marshall *et al.* (2006).

2. Analysis methodology

Estimation of cosmological parameters from flux measurements is a two-step procedure. First, a SN light-curve fitting algorithm is applied in order to ‘standardise’ these observations. Second, cosmological parameters are estimated using the output from the light-curve fitting method. In our analysis, this is followed by the application of a consistency test to the constraints derived both on the cosmological parameters and those associated with the SNe population.

Light-curve fitting. They are many different algorithms to perform light-curve fitting, in this paper we employ the SALT-II SN light-curve fitter. When fitting SN light-curves

Parameter	Symbol	Prior	Simulated value
Matter density	$\Omega_{m,0}$	$\mathcal{U}(0, 1)$	0.3
Absolute magnitude	M_0	$\mathcal{U}(-20.3, -18.3)$	-19.3
Stretch multiplier	α	$\mathcal{U}(0, 1)$	0.12
Colour multiplier	β	$\mathcal{U}(1, 5)$	2.6

Table 1. Priors assumed on the parameters, where $U(a, b)$ indicates the uniform distribution in the range $[a, b]$. The final column gives the value assumed in generating the simulated data.

with SALT-II, the outputs are the best-fit values \hat{m}_B^* of the apparent rest frame B -band magnitudes of the SNe at maximum luminosity, the light-curve shape parameter \hat{x}_1 , the colour \hat{c} in the B -band at maximum luminosity, and the covariance matrix of the uncertainties in the estimated light-curve parameters \hat{C} . Combining these quantities with the estimated redshift \hat{z} of the SN, our basic input data for each SN are thus

$$D_i \equiv \{\hat{z}_i, \hat{m}_{B,i}^*, \hat{x}_{1,i}, \hat{c}_i\} \tag{2.1}$$

for $(i = 1, \dots, N_{\text{SN}})$. The vector $(\hat{m}_{B,i}^*, \hat{x}_{1,i}, \hat{c}_i)$ for each SN is then assumed to be distributed as a multivariate Gaussian about their respective true values, with covariance matrix \hat{C}_i .

Parameter estimation. The output from the light-curve fitting constitutes the data used to estimate parameters. In this paper we employ the χ^2 method, since it is the most common approach.

In this approach, one defines the χ^2 misfit function to be

$$\chi^2(\mathcal{C}, \alpha, \beta, M_0, \sigma_{\text{int}}) = \sum_{i=1}^N \frac{[\mu_i^{\text{obs}}(\alpha, \beta, M_0) - \mu(\hat{z}_i, \mathcal{C})]^2}{(\sigma_{\mu,i}^z)^2 + \sigma_{\text{int}}^2 + \sigma_{\text{fit},i}^2(\alpha, \beta)}, \tag{2.2}$$

where $\mu(\hat{z}_i, \mathcal{C})$ is the predicted distance modulus. In this expression, the ‘observed’ distance modulus for the i th SN is

$$\mu_i^{\text{obs}} = \hat{m}_{B,i}^* - M_0 + \alpha \hat{x}_{1,i} - \beta \hat{c}_i, \tag{2.3}$$

where M_0 is the (unknown) B -band absolute magnitude of the SNe, and α, β are (unknown) nuisance parameters controlling the stretch and colour corrections. The three dispersion components are: (i) the error $\sigma_{\mu,i}^z$ in the distance modulus; (ii) the intrinsic dispersion σ_{int} , which describes the global variation in the SNIa absolute magnitudes; and (iii) the fitting error $\sigma_{\text{fit},i}(\alpha, \beta)$.

We also assume that dark energy is in the form of a cosmological constant ($w = -1$) and that the universe is spatially-flat; thus we vary only the parameters $\{\Omega_{m,0}, M_0, \alpha, \beta, \sigma_{\text{int}}\}$.

In keeping with standard practice, we define the ‘likelihood’ to be simply

$$\mathcal{L}(\Omega_{m,0}, M_0, \alpha, \beta, \sigma_{\text{int}}) = \exp[-\frac{1}{2}\chi^2(\Omega_{m,0}, M_0, \alpha, \beta, \sigma_{\text{int}})] \tag{2.4}$$

which is clearly not properly normalized. More importantly, this ‘likelihood’ is *not* (proportional to) the probability $\Pr(\mathbf{D}|\Omega_{m,0}, M_0, \alpha, \beta, \sigma_{\text{int}})$.

Adopting the initial value $\sigma_{\text{int}} = 0.1$, we use the nested sampling algorithm MULTINEST (Feroz and Hobson 2008; Feroz *et al.* 2009) to sample from the (unnormalised) ‘posterior’ $\mathcal{P}(\Theta, \sigma_{\text{int}}) = \mathcal{L}(\Theta, \sigma_{\text{int}})\pi(\Theta)$ assuming the separable, uniform priors $\pi(\Theta)$ listed in Table 1. The value of σ_{int} is estimated by adjusting it to obtain $\chi_{\text{min}}^2/N_{\text{dof}} \sim 1$.

Test for mutual consistency between data-sets. The standard approach to analysing multiple data-sets jointly is simply to assume that they are mutually consistent. We represent this (null) hypothesis by H_0 . It may be the case, however, that the data-sets are inconsistent with one another, resulting in each one favouring a different region of the

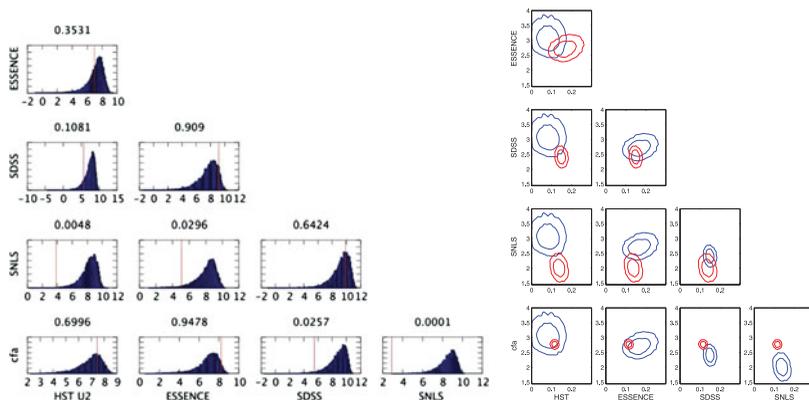


Figure 1. Left panel: Results of the consistency test applied to surveys pairs in the UNION2 compilation. The blue histograms show the distribution of R values obtained from 10^4 consistent simulations of each pair, and the red vertical line indicates the R value obtained from the real data. The corresponding one-sided p -value is given above each panel. Right panel: Two-dimensional marginalised constraints on the parameters (α, β) obtained from the individual constituent surveys contained in the UNION2 catalogue. The red (blue) contours denote the 68 and 95 per cent confidence regions for the survey in that row (column).

model parameter space. We represent this (alternative) hypothesis by H_1 . In this case, a joint analysis would lead to completely misleading results (see, for example, Appendix A in Feroz *et al.* 2008 for a demonstration).

In order to determine which one of these hypotheses is favoured by the data, one can perform Bayesian model selection between H_0 and H_1 . Assuming that hypothesis H_0 and H_1 are equally likely *a priori*, this can be achieved by calculating the ratio

$$\mathcal{R} = \frac{\Pr(\mathbf{D}|H_0)}{\Pr(\mathbf{D}|H_1)} = \frac{\Pr(\mathbf{D}|H_0)}{\prod_i \Pr(D_i|H_1)}. \quad (2.5)$$

where the probabilities $\Pr(\mathbf{D}|H)$, called Bayesian evidences or marginal likelihoods.

In adapting this Bayesian test to apply it to the standard χ^2 analysis of SNIa data, one replaces $\Pr(\mathbf{D}|\Theta, H)$ by the ‘likelihood’ $\mathcal{L}(\Theta, \sigma_{\text{int}})$ given in (2.4) and $\Pr(\Theta|H)$ by the prior $\pi(\Theta)$ listed in Table 1. In so doing, however, one cannot interpret the terms in (2.5) directly as probabilities, and thus the value of the \mathcal{R} cannot be compared with the normal Jeffrey’s scale. Nonetheless, one still expects \mathcal{R} values to be higher for consistent data-sets and lower for inconsistent ones (March *et al.* 2011), so \mathcal{R} may be used as a test statistic that is ‘calibrated’ with simulations to perform a standard one-sided frequentist hypothesis test.

Thus, we construct the distribution of \mathcal{R} under the null hypothesis H_0 by evaluating it for the simulations in which the individual surveys are mutually consistent. The \mathcal{R} value obtained by analysing the real data can then be used to calculate the p -value as follows:

$$p = \frac{N(\mathcal{R}_s < \mathcal{R}_r)}{N_{\text{tot}}}, \quad (2.6)$$

where \mathcal{R}_s and \mathcal{R}_r are the \mathcal{R} values obtained by analysing simulated and real data-sets respectively, $N(\mathcal{R}_s < \mathcal{R}_r)$ is the number of simulations with \mathcal{R} values less than that obtained by analysing the real data and N_{tot} is the total number of simulations.

3. Data-sets

Real data. In this paper we consider SNe from the UNION2 catalogue (Amanullah *et al.* 2010) and the ‘joint light-curve analysis’ (JLA) compilation given in Betoule *et al.* (2014). These data-sets consist of SNe observed by different telescopes. Since our aim in

this paper is to check for consistency between different SNe surveys/data-sets, we divide these SNe according to the telescope with which they have been observed.

SNANA simulations. In order to calibrate our test statistic \mathcal{R} and calculate the resultant p -values we simulate SNe using the publicly available SNANA package[†]. Our null hypothesis H_0 is that different data-sets are consistent with each other. Thus, we simulate all data-sets with exactly the same parameter values, which are listed in Table 1, and an intrinsic dispersion $\sigma_{\text{int}} = 0.106$.

4. Results and conclusions

Constraints on $\Omega_{\text{m},0}$ and M_0 . In analysing compilations of SNe data, it is usual to plot the combined constraints on cosmological parameters obtained from a full joint analysis. Here we instead calculating the constraints imposed by each constituent survey in the UNION2 and JLA compilations. The key observation for our purposes is that the constraints from the constituent surveys are in good agreement, as indicated by the large degree of overlap of the confidence contours for each pairwise comparison for both the UNION2 catalogue and the JLA.

Constraints on α and β . We now consider the constraints on the stretch and colour corrections multipliers α and β , which are not often presented in the analysis of SNe data. The resulting two-dimensional marginalised constraints on the parameters (α, β) are shown in right panel Figs 1 for UNION2. One sees that for several pairings the confidence contours for the two surveys are significantly displaced from one another, and do not overlap at all in some cases. For the JLA compilation for each pairing the confidence contours for the two surveys overlap well, indicating mutual consistency, in sharp contrast to the UNION2 results.

Mutual consistency test between survey pairs. The results of the consistency test for the UNION2 compilation are shown in left panel of Fig. 1, together with the corresponding p -values. One sees that the p -values obtained agree very closely with the degree of overlap between the confidence contours plotted in Fig. 1. In particular, we find very strong evidence, at more than the 99 per cent significance level, for inconsistency between the survey pairs CfA/SNLS and SNLS/HST.

The same analysis for JLA shows that none of the survey pairs shows evidence for inconsistency, even at the 90 per cent significance level, which is in agreement with the good degree of overlap between the confidence contours.

The consistency between the survey pairs in the JLA compilation is in sharp contrast to inconsistencies present in the UNION2 compilation. This difference must result from the different SNe selection criteria and calibration techniques used in the construction of the two compilations. The level of inconsistency exhibited by the UNION2 compilation suggests that one must exercise caution when interpreting cosmological constraints derived from it with the usual joint analysis.

References

- Amanullah R., *et al.* 2010, *ApJ*, 716, 712
 Betoule M., *et al.* 2014, *ArXiv e-prints*
 Feroz F., *et al.* 2008, *Journal of High Energy Physics*, 10, 64
 Feroz F., Hobson M. P. 2008, *MNRAS*, 384, 449
 Feroz F., Hobson M. P., Bridges M. 2009, *MNRAS*, 398, 1601
 March M. C., Trotta R., Amendola L., Huterer D. 2011, *MNRAS*, 415, 143
 Marshall P., Rajguru N., Slosar A. 2006, *Phys.Rev.D*, 73, 067302
 Perlmutter S., *et al.* 1999, *ApJ*, 517, 565
 Riess A. G., *et al.* 1998, *AJ*, 116, 1009

[†] The SNANA package: <http://sdssdp62.fnal.gov/sdssn/SNANA-PUBLIC/>