Crossing Statistics: Dealing with Unknown Errors in the Dispersion of Data

Shafieloo, Arman
Institute for the Early Universe, Ewha Womans University
Seoul, 120-750, South Korea
E-mail: arman@ewha.ac.kr

Clifton, Timothy
Astrophysics Department, University of Oxford
Denys Wilkinson Building, Oxford, OX1 3RH, UK
E-mail: tclifton@astro.ox.ac.uk

Ferreira, Pedro
Astrophysics Department, University of Oxford
Denys Wilkinson Building, Oxford, OX1 3RH, UK
E-mail: p.ferreira1@physics.ox.ac.uk

We propose a new statistic that has been designed to be used in situations where the intrinsic dispersion of a data set is not well known: 'The Crossing Statistic'. This statistic is in general less sensitive than χ^2 to the intrinsic dispersion of the data, and hence allows us to make progress in distinguishing between different models using goodness of fit to the data even when the errors involved are poorly understood. We show that this statistic can easily distinguish between different models in cases where the χ^2 statistic fails. We also show that the last mode of Crossing Statistic is identical to χ^2, so that one can consider it as a generalization of χ^2.

SN Ia act to some degree like standardized candles, and are widely used in cosmology to probe the expansion history of the universe, and hence to investigate the properties of dark energy. Indeed it is from observations of SN Ia that the first direct evidence for an accelerating universe was found [2], and although this result has far reaching physical consequences, a complete understanding of the physics of SN Ia is still lacking. This lack of understanding is manifest in the largely unaccounted for intrinsic dispersion of SN Ia, which affect almost any subsequent statistical analysis that one attempts to perform [3]. Given that the intrinsic dispersion of SN Ia, $\sigma_{(int)}$, typically constitute a large fraction of the total error on a data point, σ_i, this is a serious problem.

One procedure that is often used to find the *a priori* unknown intrinsic dispersion is to look for the value of $\sigma_{(int)}$ that gives a reduced χ^2 of 1 for a particular model, and then to use this value to determine the likelihood of the data given that model. Such an approach does indeed allow one to distinguish between different models using the likelihood function, but at the expense of losing much of the original concept of 'goodness of fit' (which is the essence of a χ^2 analysis). Rather than directly answering the question of which model actually fits the data best, we are then left with answering the question of which model can be made to give an ideal fit to the data by adding the smallest possible error bars. This gives us no direct information about which model best fits the data, as the error bars have been adjusted by hand so that they all fit perfectly. Furthermore, by treating error bars in this way it becomes very difficult to detect any features that may be present in the data.

If we want to determine the goodness of fit of different models to the data, we must therefore take a different approach. Standard statistics, such as χ^2, however, are only reliable when the assumed parameterization of the model is correct, and when the errors on the data are properly estimated. Given that the true nature of dark energy is still not known, and that we have no reliable theoretical derivation of $\sigma_{(int)}$, the application of χ^2 statistics to the SN Ia data is not at all straightforward. These problems persist even when using non-parametric or model independent approaches [4].

To address these difficulties we propose a new statistic, which we call *the Crossing Statistic* [1]. This statistic is significantly less sensitive than χ^2 to intrinsic dispersion of the data, and can therefore be used more easily to check the consistency between a given model and a data set with largely unknown errors. The Crossing Statistic does not compare two models directly, but rather determines the probability of getting the observed data given a particular theoretical model. It works with the data directly, and makes use of the shape and trends in a model's predictions when comparing it with the data.

In the following we will discuss the concept of goodness of fit and show how the χ^2 statistic is sensitive to the size of unknown errors, as well as how it fails to distinguish between different cosmological models when errors are not prescribed in a definite manner. We will then introduce the Crossing Statistic and show how it can be used to distinguish between different cosmological models when the standard χ^2 analysis fails to do so. For simplicity, we will restrict ourselves to four theoretical models: (i) a best fit flat ΛCDM model, (ii) a smooth Lemaître-Tolman-Bondi void model with simultaneous big bang [5], (iii) a flat ΛCDM model with $\Omega_{0m} = 0.22$, and (iv) an open, empty 'Milne' universe. We will use the Constitution supernovae data set [6] consisting of 147 supernovae at low redshifts and 250 supernovae at high redshifts, from the SuperNovae Legacy Survey, the ESSENCE survey, the HST data set, as well as some older data sets, and fitted for using the SALT light-curve fitter. We adjust the size of the error-bars in this data set by considering additional intrinsic errors (added quadratically). By comparing with χ^2 we then show that the Crossing Statistic is relatively insensitive to the unknown intrinsic error, as well as being more reliable in distinguishing between different cosmological models. In a companion paper, we will test a number of other dark energy models using this statistic.

First let us consider the χ^2 statistic. For a given data set $(\mu_i^e, i = 1 \cdots N)$ we have that χ^2 is given by

$$\chi^2 = \sum_i^N \frac{(\mu_i^t - \mu_i^e)^2}{\sigma_i^2}, \tag{1}$$

where μ_i^t is the model prediction that we are comparing the data set to, and σ_i are the corresponding variances (σ has units of magnitudes throughout). If the data-points are uncorrelated and have a Gaussian distribution around the distribution mean, then we have a χ^2 distribution with $N - N_P$ degrees of freedom (where N_P is the number of parameters in the theoretical model).

Now let us now calculate the χ^2 goodness of fit for two of our cosmological models: a flat best fit ΛCDM model, and a Milne universe. Let us also assume an additional intrinsic error, $\sigma_{(int)}$, on top of the error prescribed in the Constitution data set, $\sigma_{i(data)}$, so that the total error is $\sigma_i^2 = \sigma_{i(data)}^2 + \sigma_{(int)}^2$. This will allow us to check how sensitive our analysis is to coherent changes in error-bars. In Fig. 1 we plot the χ^2 goodness of fit for our two theoretical models as a function of $\sigma_{(int)}$. It can be seen that these two models cannot be easily distinguished from each other using χ^2 alone, unless the additional intrinsic error is known. We also note that the χ^2 goodness of fit for the standard flat ΛCDM model given the Constitution data without any additional intrinsic errors is less than 0.6% ($\chi^2 = 465.5$ for 397 data points).

If the real Universe differs from the assumed theoretical model, one would hope that it would be possible to develop a statistical test that would be able to pick up on this. To these ends we consider the 'crossings' between the predictions of a given model, and the real Universe from which data is derived. Figure 2 shows a schematic picture of what we mean by one crossing (left panel), and two crossings (right panel). In what follows we will use the existence of this type of crossing to develop a new statistic that can be used to determine the goodness of fit between an assumed model and the

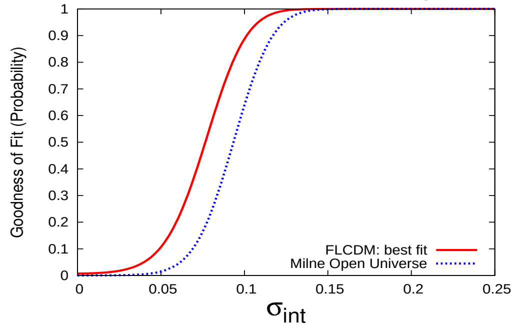

Figure 1: The χ^2 goodness of fit of the Constitution supernovae data [6] to a flat ΛCDM model (red line) and a Milne universe (blue line), assuming additional intrinsic errors added quadratically to the errors specified in the data set. The χ^2 goodness of fit for these two models can be seen to be comparable for different values of additional intrinsic error, making them difficult to distinguish without any knowledge of what $\sigma_{(int)}$ should be.

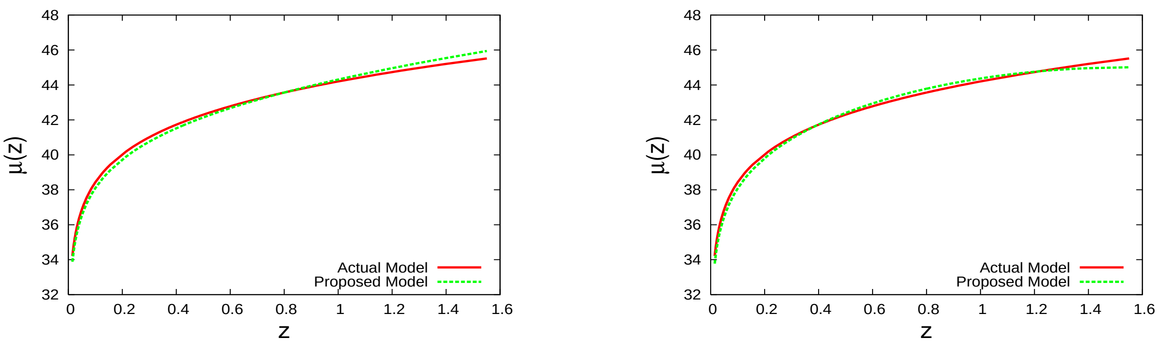

Figure 2: An idealized schematic plot of one crossing (left panel) and two crossings (right panel) between a proposed theoretical model and the actual model of the Universe when comparing magnitudes as a function of redshift, $\mu(z)$. In reality the actual Universe is observed in the form of data with error bars, of course.

real Universe.

To build our Crossing Statistic in the case of SN Ia data, we must first pick a theoretical or phenomenological model of dark energy (e.g. ΛCDM) and a data set of SN Ia distance moduli $\mu_i(z_i)$ (e.g. the Constitution data set [6]). As in [7], we use the χ^2 statistic to find the best fit form of the assumed model, and from this we then construct the *error normalized difference* of the data from the best fit distance modulus $\bar{\mu}(z)$:

$$q_i(z_i) = \frac{\mu_i(z_i) - \bar{\mu}(z_i)}{\sigma_i(z_i)}. \tag{2}$$

Let us now consider the *one-point Crossing Statistic*, which tests for a model and a data set that cross each other at only one point. We must first try to find this crossing point, which we label by n_1^{CI} and z_1^{CI}. To achieve this we define

$$T(n_1) = Q_1(n_1)^2 + Q_2(n_1)^2, \tag{3}$$

where $Q_1(n_1)$ and $Q_2(n_1)$ are given by

$$\begin{aligned} Q_1(n_1) &= \sum_{i=1}^{n_1} q_i(z_i) \\ Q_2(n_1) &= \sum_{i=n_1+1}^{N} q_i(z_i), \end{aligned} \tag{4}$$

and where N is the total number of data points. If n_1 is allowed to take any value from 1 to N (when the data is sorted by red shift) then we can maximize $T(n_1)$ by varying with respect to n_1^{CI}. We then

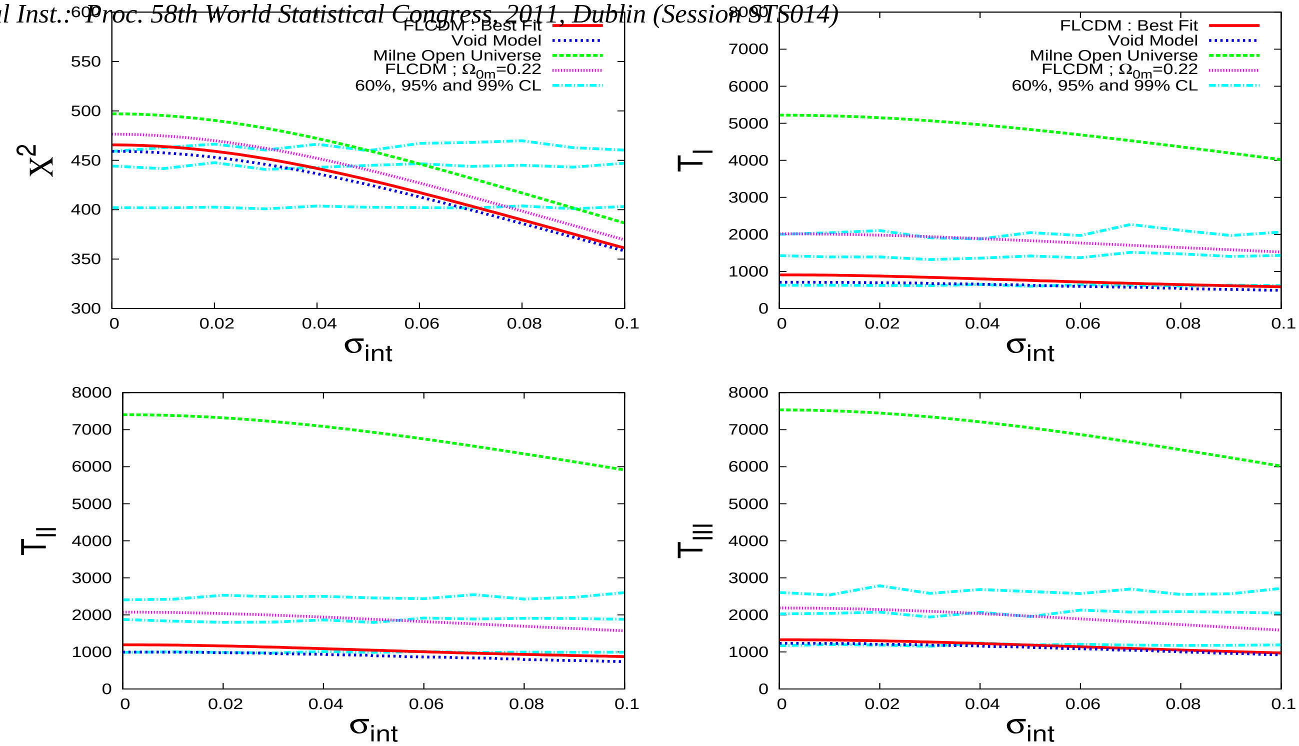

Figure 3: The χ^2, T_I, T_{II} and T_{III} statistics for a best fit flat ΛCDM model (red lines), a void model (blue dashed lines), the Milne universe (green dashed lines) and a flat ΛCDM model with $\Omega_m = 0.22$ (pink dotted lines). The analyses are performed using the Constitution supernovae data [6], and by assuming various different additional intrinsic errors. The confidence limits derived from 1000 Monte Carlo realizations of the data error around the zero mean value. It can be seen the χ^2 statistic fails to distinguish between these models with any degree of significance, and that by assuming additional intrinsic errors all models can be made to show a good consistency with the data. The T_I crossing statistic, on the other hand, rules out the Milne universe and also the flat ΛCDM model with $\Omega_m = 0.22$ to a high degree of confidence, even when the amount of additional intrinsic error is large.

write the maximum value of $T(n_1)$ as T_I. Finally, we can then use Monte Carlo simulations to find how often we should expect to obtain a T_I larger or equal to the value derived from the observed data. In particular, for each Monte Carlo dataset we can find the best fit form of $\mu(z)$ and then follow the steps above to find the corresponding T_I. In doing this we find the fraction of Monte Carlo data sets leading to $T_I^{M.C} \geq T_I^{data}$, which we will use as an estimate of the probability that the data set should be realized from the particular best fit cosmological model we have been considering.

This approach can be extended to models with more than one crossing point by the *two-point Crossing Statistic*. In this case we assume that the model and the data cross each other at two points and, as above, we try to find the two crossing points and their red shifts, which we now label n_1^{CII}, z_1^{CII} and n_2^{CII}, z_2^{CII}. This is achieved by defining

(5) $$T(n_1, n_2) = Q_1(n_1, n_2)^2 + Q_2(n_1, n_2)^2 + Q_3(n_1, n_2)^2,$$

where the $Q_i(n_1, n_2)$ are now given by

$$Q_1(n_1, n_2) = \sum_{i=1}^{n_1} q_i(z_i)$$

$$Q_2(n_1, n_2) = \sum_{i=n_1+1}^{n_2} q_i(z_i)$$

(6) $$Q_3(n_1, n_2) = \sum_{i=n_2+1}^{N} q_i(z_i).$$

We can then maximize $T(n_1, n_2)$ by varying with respect to n_1 and n_2, to get T_{II}. Comparing T_{II} with the results from Monte Carlo realizations then allows us to determine how often we should expect a two-point crossing statistic that is greater than or equal to the T_{II} obtained from real data. The *three-point Crossing Statistic*, and higher statistics, can be defined in a similar manner. This can continue up to the *N-point Crossing Statistic* which is, in fact, identical to χ^2. We also note that the zero-point Crossing Statistic, $T_0 = (\sum_i^N q_i)^2$, is very similar to the Median Statistic developed by Gott *et al.* [8]. The Crossing Statistic can therefore be thought of as generalizing both the χ^2 and Median Statistics, which it approaches in different limits.

Let us now apply our Crossing Statistics to a suite of different models. We calculate χ^2, T_I, T_{II} and T_{III} for (i) the best fit flat ΛCDM model (with $\Omega_{0m} = 0.288$ when $\sigma_{(int)} = 0$), (ii) a best fit asymptotically flat void model (with $\Omega_{0m} = 0.28$ at the centre, and with FWHM at $z = 0.66$ when $\sigma_{(int)} = 0$), (iii) a flat ΛCDM model with $\Omega_{0m} = 0.22$, and (iv) the Milne open universe. We use the Constitution data set [6], and vary the additional intrinsic error, $\sigma_{(int)}$, between 0 and 0.25 magnitudes. In Fig. 3 we compare these statistics with the confidence limits that result from 1000 Monte Carlo realizations of the data error around a zero mean value for each $\sigma_{(int)}$. One could use other models for these confidence limits, but we expect them to be relatively insensitive to the particular model that is chosen.

It can be seen from Fig. 3 that the χ^2 statistic (upper left panel) cannot easily be used to distinguish between the different models with a high degree of confidence, especially if we do not know $\sigma_{(int)}$. Indeed, if we add $\sigma_{(int)} = 0.1$ magnitudes to the data then all four models become a good fit, at the 60% confidence level. Alternatively, with $\sigma^2_{(int)} = 0$ all four models are outside of the 99% confidence level. This illustrates the ineffectiveness of χ^2 as a statistic for determining the goodness of fit when the errors on the data are not well known.

The results for the one-point Crossing Statistic are shown in the upper right panel in Fig. 3. In terms of this statistic it can be seen that the best fit flat ΛCDM model and the best fit void model are now very much consistent with the data, even with no additional intrinsic error. At the same time, it is also clear that the Milne universe and the flat ΛCDM model with $\Omega_{0m} = 0.22$ lie well outside the 99% and 95% confidence levels, even when $\sigma_{(int)}$ is large. Based on our new statistic, these last two models are therefore ruled out with high confidence. In the lower panels in Fig. 3 we see the results for the two-point and three-point Crossing Statistics, respectively. The Milne Universe remains outside the 99% confidence level in each of these, for the range of $\sigma_{(int)}$ considered, while the flat ΛCDM model with $\Omega_{0m} = 0.22$ now lies mostly within the 60-99% confidence region.

This difference in probability of the different Crossing Statistics for the ΛCDM model with $\Omega_{0m} = 0.22$ is due to this model having only one 'crossing' with the data. Adding extra hypothetical crossings then has little affect on T_i, as the extra crossing points all cluster around the same z. A model that fits the data better, with many crossings, however, should be expected to have T_i statistics that increase with i. On this basis, one can then argue that for a model to be considered consistent with the data it must show consistency across all crossing modes. The point here is that if there is a significant crossing of the data and the model, then it should show up in the Crossing Statistics as a failure of T_i to decrease sufficiently with decreasing i. A flat ΛCDM model with $\Omega_{0m} = 0.22$ is therefore considered non-viable because of the discrepancy in T_I, even though T_{II} and T_{III} show some degree of consistency.

In summary, we have presented a new statistic that can be used to distinguish between different cosmological models using their goodness of fit with supernovae data. We have shown that the different Crossing Statistics are sensitive to the shapes and trends of the data and the assumed theoretical model, and are in general less sensitive to the unknown intrinsic dispersion of the data than χ^2. This is exemplified by the fact that the consistency between a model and a data set does not change much, even when we assume large additional intrinsic errors. The Crossing Statistic appears to us to be a promising method of confronting cosmological models with supernovae observations, and can be straightforwardly generalized to other datasets where similar problems occur.

AS acknowledges the support of the EU FP6 Marie Curie Research and Training Network "UniverseNet" (MRTN-CT-2006-035863) and Korea World Class University grant no. R32-10130.

References

[1] A. Shafieloo, T. Clifton and P. G. Ferreira, JCAP **08**, 017 (2011).

[2] A. G. Riess *et al.*, Astron. J. **116**, 1009 (1998); S. J. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999).

[3] A. G. Kim, E. V. Linder, R. Miquel and N. Mostek, Mon. Not. Roy. Astron. Soc. **347**, 909 (2004); E. V. Linder, Phys. Rev. D **79**, 023509 (2009); R. P. Kirshner, arXiv:0910.0257.

[4] R. A. Daly and S. G. Djorgovski, Astrophys. J. **597**, 9 (2003); Y. Wang and P. Mukherjee, Astrophys. J. **606**, 654 (2004); V. Sahni, A. Shafieloo and A. A. Starobinsky, Phys. Rev. D **78**, 103502 (2008); C. Zunckel and C. Clarkson, Phys. Rev. Lett **101**, 181301 (2008); A. Shafieloo, V. Sahni and A. A. Starobinsky, Phys. Rev. D **80 R**, 101301 (2009); A. Shafieloo, U. Alam, V. Sahni and A. A. Starobinsky, Mon. Not. Roy. Astron. Soc. **366**, 1081 (2006); A. Shafieloo, Mon. Not. Roy. Astron. Soc. **380**, 1573 (2007); Y. Wang and M. Tegmark, Phys. Rev. D **71**, 103513 (2005); A. Shafieloo and C. Clarkson, Phys. Rev. D **81**, 083537 (2010); S. Nesseris and A. Shafieloo, Mon. Not. Roy. Astron. Soc. **408**, 1879 (2010).

[5] H. Alnes, M. Amarzguioui and Ø. Grøn, Phys. Rev. D **73**, 083519 (2006); H. Alnes and M. Amarzguioui, Phys. Rev. D **75**, 023506 (2007); T. Clifton, P. G. Ferreira and K. Land, Phys. Rev. Lett. **101** 131302 (2008).

[6] M. Hicken *et al.*, Astrophys.J. **700**, 1097 (2009); M. Hicken *et al.*, Astrophys.J. **700**, 331 (2009).

[7] L. Perivolaropoulos and A. Shafieloo, Phys. Rev. D **79**, 123502 (2009).

[8] J. R. Gott III *et al.* , Astrophys. J. **549**, 1 (2001).

99