DRAFT VERSION MARCH 10, 2017
Preprint typeset using LaTeX style emulateapj v. 12/16/11

PROBABILISTIC INFERENCE OF THE HUBBLE PARAMETER

ALEX MALZ[1], TINA PETERS[2], LLUIS GALBANY, HUMNA AWAN, ANITA BAHMANYAR, AND KARA PONDER

Draft version March 10, 2017

ABSTRACT

The BEAMS framework enables the use of probabilistic supernova classifications to estimate the Hubble parameter quantifying the relationship between distance and redshift over cosmic time. This work extends BEAMS to replace high-confidence spectroscopic redshifts with probabilistic photometric redshifts, enabling inference of the Hubble parameter as a function of two probabilistic variables. By combining posterior probabilities of supernova type and posterior probability distributions over host galaxy redshift, we infer a posterior probability distribution over the redshift-dependent Hubble parameter. This work also produces a code that can be used for other regression problems in astrophysics that involve catalogs of two probabilistic variables.

1. INTRODUCTION

Kunz *et al.* (2007); Kelly *et al.* (2008); Hlozek *et al.* (2012)

2. METHODS

This covers a complete sample, i.e. a catalog of all N supernovae ns in the universe, ever. We'll add in the selection function later.

The goal here is to constrain the posterior distribution of the hyperparameters of interest given the data.
We should give a physical description of what these hyperparameters are: in this case of the Hubble Diagram we are interested in constraining the cosmology so omega matter, omega gamma, and H naut.
We first expand this in terms of Bayes' Rule.

$$p(\vec{\theta}, \underline{\phi} | \{\underline{\ell}_n\}_N, \{\vec{m}_n\}_N) \propto p(\vec{\theta}, \underline{\phi})\ p(\{\underline{\ell}_n\}_N, \{\vec{m}_n\}_N | \vec{\theta}, \underline{\phi}) \tag{1}$$

In plain English we are saying that the likelihood of a certain cosmology and the hyperparameters given some set of observed supernova lightcurves and fluxes is proportional to the prior on the cosomological parameters multiplied by the likelihood of the lightcurves and fluxes given the cosmology and hyperparameters.
Next, we invoke the independence of the supernova parameters and observations; the n^{th} system's parameters and data are assumed to be independent of the $(n+1)^{th}$ system's parameters and data.
In essence this is saying that each supernova observed is independent of every other supernova observed. Is this a safe assumption? What are the caveats of making this assumption?

$$p(\{\underline{\ell}_n\}_N, \{\vec{m}_n\}_N | \vec{\theta}, \underline{\phi}) = \prod_n^N p(\underline{\ell}_n, \vec{m}_n | \vec{\theta}, \underline{\phi}) \tag{2}$$

aimalz@nyu.edu
[1] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, 9th floor, New York, NY 10003, USA
[2] Dunlap Institute & Department of Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, ON M5S 3H4 Canada

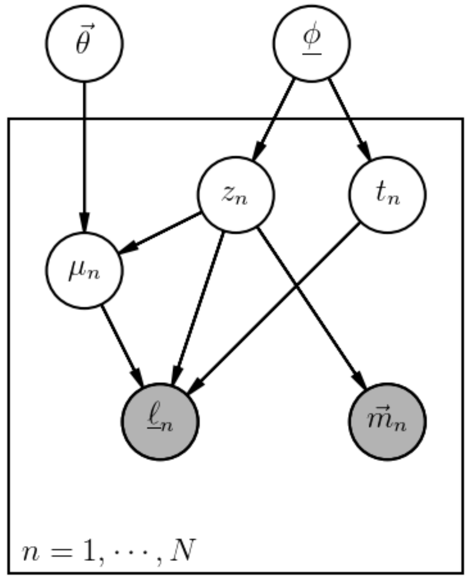

FIG. 1.— This directed acyclic graph corresponds to a probabilistic graphical model for our hierarchical inference of the Hubble parameter. In this graph, all random variables are shown in circles (**What is meant by "random" here?**), with observed variables shown in shaded circles. The box indicates that there are N copies of the relationships between boxed parameters, each independent of all others. The hyperparameters we would like to infer are the cosmological parameters in $\vec{theta}$ and the supernova type-redshift distribution parameters comprising $\underline{\phi}$. Drawn from functions of these hyperparameters are the distance moduli $\{\mu_n\}_N$, redshifts $\{z_n\}_N$, and supernova types $\{t_n\}_N$. Here, we observe host galaxy colors $\{\vec{m}_n\}_N$ and multi-band supernova lightcurves $\{\underline{\ell}_n\}_N$, shown in shaded circles.

This assumption is necessary for us to easily combine the contributions to likelihoods of the individual supernova. Here we are saying the the likelihood of the set of lightcurves and fluxes is simply the product of each of the individual likelihoods.

Next, we use marginalization of the latent variables, because we do not wish to estimate their values directly.

$$p(\underline{\ell}_n, \vec{m}_n|\vec{\theta}, \underline{\phi}) = \iiint p(\underline{\ell}_n, \vec{m}_n|\mu_n, z_n, t_n)\ p(\mu_n, z_n, t_n|\vec{\theta}, \underline{\phi})\ d\mu_n\ dz_n\ dt_n \tag{3}$$

Now we expand the expression to include the latent variables, the intermediate values that are not directly measured but calculated from the measured data. In this case are latent variables are distance modulus, redshift, and type the values that we are producing in the simulated data.

We note that the equation calls for likelihoods $\{p(\underline{\ell}_n, \vec{m}_n|\mu_n, z_n, t_n)\}_N$ (**the likelihood of a certain measured lightcurve and flux given some distance modulus, redshift, and type.**), but what we have are interim posteriors $\{p(\mu_n, z_n, t_n|\underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*)\}_N$ (**likelihood of some distance modulus, redshift, and type given the lightcurve, flux, and the priors that are used in the fitting of the lightcurve to a suite of templates. This would be a good time to go into a discuss about how lightcurve fitters work.**) for some interim priors $\vec{\theta}^*, \underline{\phi}^*$. The interim priors represent the beliefs about the hyperparameters $\vec{\theta}$ and $\underline{\phi}$ that go into our calculation of the posteriors from the data; whether explicitly chosen (as in methods relying on a template library) or implicitly derived (as in methods relying on a training data set), the interim priors are always present in the calculation of posterior distributions from data. We will thus have to transform the math to be in terms of quantities we actually have. We do this by multiplying the likelihood by an inspired factor of unity, in terms of the interim posteriors.

$$p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n) = p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n) \; \frac{p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*)}{p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*)} \tag{4}$$

We will expand the denominator of that factor of unity according to Bayes' Rule.

$$p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) = \frac{p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*) \; p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n, \vec{\theta}^*, \underline{\phi}^*)}{p(\underline{\ell}_n, \vec{m}_n | \vec{\theta}^*, \underline{\phi}^*)} \tag{5}$$

By the independence of different hierarchical levels in the probabilistic graphical model, we may split up the most daunting term in the above expression.

$$p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n, \vec{\theta}^*, \underline{\phi}^*) = p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n) \; p(\underline{\ell}_n, \vec{m}_n | \vec{\theta}^*, \underline{\phi}^*) \tag{6}$$

Noting the presence of $p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n)$ and $p(\underline{\ell}_n, \vec{m}_n | \vec{\theta}^*, \underline{\phi}^*)$ in both the numerator and denominator for $p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n)$, we cancel the like terms to express the individual likelihoods in terms of known quantities.

$$p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n) = \frac{p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*)}{p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*)} \tag{7}$$

We are now ready to plug the individual likelihoods into the marginalization.

$$p(\underline{\ell}_n, \vec{m}_n | \vec{\theta}, \underline{\phi}) = \iiint p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) \; \frac{p(\mu_n, z_n, t_n | \vec{\theta}, \underline{\phi})}{p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*)} \; d\mu_n \; dz_n \; dt_n \tag{8}$$

Now we can plug the marginalization back into the product.

$$p(\{\underline{\ell}_n, \vec{m}_n\}_N | \vec{\theta}, \underline{\phi}) = \prod_n^N \; \iiint p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) \; \frac{p(\mu_n, z_n, t_n | \vec{\theta}, \underline{\phi})}{p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*)} \; d\mu_n \; dz_n \; dt_n \tag{9}$$

And finally, we can plug the product back into Bayes' Rule.

$$p(\vec{\theta}, \underline{\phi} | \{\underline{\ell}_n, \vec{m}_n\}_N) \propto p(\vec{\theta}, \underline{\phi}) \; \prod_n^N \; \iiint p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) \; \frac{p(\mu_n, z_n, t_n | \vec{\theta}, \underline{\phi})}{p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*)} \; d\mu_n \; dz_n \; dt_n \tag{10}$$

This is the posterior we will attempt to sample!

3. MOCK DATA

In order to simulate a mock dataset of three-dimensional posterior distributions $\{p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*)\}_N$ over supernova type, redshift, and distance modulus, we will have to employ a forward model.

Since we are aiming to construct posteriors, we know from Bayes' Rule that they will take the form of Eq. 11.

$$p(\mu_n, z_n, t_n | \underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) \propto p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n, \vec{\theta}^*, \underline{\phi}^*) \; p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*) \tag{11}$$

We will discuss the two terms separately, starting with the second. According to Fig. 1, we have Eq. 12.

$$p(\mu_n, z_n, t_n | \vec{\theta}^*, \underline{\phi}^*) = p(\mu_n | z_n, \vec{\theta}^*) \; p(z_n, t_n | \underline{\phi}^*) \tag{12}$$

The first term of Eq. 11 may easily be reduced to $p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n)$, as there is no direct dependence of the data on the hyperparameters is nontrivial. Fig. 1 tells us how to break it up, resulting in Eq. 13.

$$p(\underline{\ell}_n, \vec{m}_n | \mu_n, z_n, t_n) = p(\underline{\ell}_n | \mu_n, z_n, t_n) \; p(\vec{m}_n | z_n) \tag{13}$$

We will set true values of the hyperparameters *(**The generating cosmology of our simulation and what we are trying to recover: e.g. Hnaut, omega matter, omega lambda.**)$\vec{\theta}_0$ and $\underline{\Phi}_0$ (**True type (Ia, Ibc, II) and true redshift of the supernova.**)that we would like to recover by performing hierarchical inference on our mock data. We will first draw pairs of true parameters (T_n^0, z_n^0) from $\underline{\Phi}_0$. Then we will calculate the true μ_n^0 according to [insert distance modulus equation here (**astropy.cosmology**)] from z_n^0 and $\vec{\theta}_0$.

What does it mean physically to sample from this posterior? We are marginalizing over mu, z, and type, but that is what is generated by our simulation. We are recovering cosmology given lightcurve and fluxes, but we never model lightcurves and fluxes.

Next, we will construct likelihoods for each supernova. Still not sure I'm thinking about this part correctly…

There are many ways to formulate the various terms in Eqs. 13 and 12, but we will restrict ourselves to the simplest version for now, outlined below in Sec. 3.1.

3.1. *Model Specifics*

We establish the following domains for the latent variables $\{\mu_n\}$, $\{z_n\}$, and $\{t_n\}$. There are T possible values τ for t, defining a discrete space of types. z and μ are defined in binned spaces with Z bins ζ of widths $\vec{\Delta}_z$ and D bins ν of widths $\vec{\Delta}_\mu$ respectively.

$\underline{\phi}$ may then be represented as a $T \times Z$ two-dimensional array satisfying $\sum_\tau^T \underline{\phi} \cdot \vec{\Delta}_z = 1$, i.e. $p(z_\zeta, t_\tau|\underline{\phi}) \equiv \phi_{\zeta\tau}$. For simplicity, we shall state that $\vec{\theta} \equiv H_0$, assuming all other cosmological parameters are perfectly known and forbidding redshift evolution. Rather than writing the form of the distance modulus as a function of redshift and cosmological parameters, we will instead use the shorthand $\mu = f_{\vec{\theta}}(z)$ here. Thus $p(\mu_\nu|z_\zeta, \vec{\theta}) \equiv \delta_{f_{\vec{\theta}}(z_\zeta)}(\mu_\nu)$ is a delta function centered at $\mu_n = f_{\vec{\theta}}(z_n)$. Under the binned parametrization established above, this makes the product $p(\mu_n, z_n, t_n|\vec{\theta}, \underline{\phi})$ of Eq. 12 a $T \times Z \times D$ three-dimensional array $\underline{S}$ that is sparse in the μ dimension, with $S_{\tau\zeta\nu} = \phi_{\zeta\tau}\delta_{f_{\vec{\theta}}(z_\zeta)}(\mu_\nu)$.

We wish to simulate interim posteriors, which means we must construct the likelihoods of Eq. 13 and make them compatible with the arrays comprising Eq. 12 that were established above. However, to do this, we will first have to set true values for the latent variables. To do that, we first choose true values of the hyperparameters $\vec{\theta}'$ and $\underline{\phi}'$ that we would like to recover by performing hierarchical inference on our mock data. Once $\vec{\theta}'$ and $\underline{\phi}'$ are set, the array $S'_{\tau\zeta\nu} = \phi'_{\zeta\tau}\delta[\mu_\nu, f_{\vec{\theta}'}(z_\zeta)]$ is also set. With the elements of $\underline{\phi}'$ as weights, we sample N pairs of indices (τ'_n, ζ'_n) specifying the true values of (t'_n, z'_n). Since there is no uncertainty in $f_{\vec{\theta}}(z'_n)$, we also know μ'_n and thus ν'_n.

Now we are ready for a generative model of the lightcurves and galaxy photometry! We will again try to consider the simplest possible model.

For each n, $p(\underline{\ell}_n|\mu_n, z_n, t_n)$ will be a $T \times Z \times D$ three-dimensional array $\underline{L}^n$. We assume that a lightcurve classifier/fitter has been specified, and that its confusion matrix $\underline{C}$ is known. The elements $C_{\tau'\hat{\tau}} = p(t'_n|\hat{t}_n)$ of the confusion matrix represent the probability that a randomly chosen supernova of true type t' is classified as type $\hat{t}$. For all T^2 combinations of τ' and $\hat{\tau}$, there is a function $\mathcal{F}_{\tau'\hat{\tau}}(z, \mu)$ that maps pairs (z', μ') onto pairs $(\hat{z}, \hat{\mu})$, which corresponds to the output of the lightcurve fitting function when using a template of type $\hat{\tau}$ to fit a lightcurve that is truly of τ', in the absence of observational errors. We assume that the lightcurve fitter produces an accurate multivariate Gaussian likelihood estimate for each fit n with covariance $\underline{\Sigma}$ We draw maximum likelihood values for $(\hat{z}^\ell, \hat{\mu}^\ell)$ from the multivariate Gaussian distribution $\mathcal{N}_{(\hat{z}_n^\ell, \hat{\mu}_n^\ell), \underline{\Sigma}_n}$. Thus, the lightcurve likelihood may be written as $L^n_{\tau\zeta\nu} = C_{\tau'_n\tau}\mathcal{N}_{(\hat{z}^\ell, \hat{\mu}^\ell), \underline{\Sigma}_n}(z_\zeta, \mu_\nu)$.

Next, we tackle $p(\vec{m}_n|z_n)$, which will be a length Z array $\vec{M}^n$, which is the likelihood that forms the basis for what is commonly reported as the photo-z PDF $p(z_n)$. We will assume a very simple model for photo-z likelihoods, that of a Gaussian distribution. We draw a maximum likelihood photo-z $\hat{z}_n^m$ from the distribution $\mathcal{N}_{z'_n, \sigma_n^2}$. Then $M^n_\zeta = \mathcal{N}_{\hat{z}_n^m, \sigma_n^2}(z_\zeta)$.

Finally, in Eq 14, we put these pieces together to express the form of the individual interim posteriors of the form of Eq. 11.

$$p_n(\mu_\nu, z_\zeta, t_\tau|\underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) = KC_{\tau'_n\tau}\mathcal{N}_{(\hat{z}^\ell, \hat{\mu}^\ell), \underline{\Sigma}_n}(z_\zeta, \mu_\nu)\mathcal{N}_{\hat{z}_n^m, \sigma_n^2}(z_\zeta)\phi_{\zeta\tau}\delta_{f_{\vec{\theta}}(z_\zeta)}(\mu_\nu) \tag{14}$$

The constant of proportionality K here will be set such that $\sum_\nu^D \sum_\zeta^Z \sum_\tau^T p_n(\mu_\nu, z_\zeta, t_\tau|\underline{\ell}_n, \vec{m}_n, \vec{\theta}^*, \underline{\phi}^*) = 1$.

REFERENCES

M. Kunz, B. A. Bassett, and R. A. Hlozek, Physical Review D **75**, 103508 (2007).

B. C. Kelly, X. Fan, and M. Vestergaard, The Astrophysical Journal **682**, 874 (2008).

R. Hlozek, M. Kunz, B. Bassett, M. Smith, J. Newling, M. Varughese, R. Kessler, J. P. Bernstein, H. Campbell, B. Dilday, B. Falck, J. Frieman, S. Kuhlmann, Hubert Lampeitl, J. Marriner, R. C. Nichol, A. G. Riess, M. Sako, and D. P. Schneider, The Astrophysical Journal **752**, 79 (2012)