Accounting for Selection Effects in Supernova Cosmology with Simulation-Based Inference and Hierarchical Bayesian Modelling

Benjamin M. Boyd [1] **Matthew Grayling** [1] **Stephen Thorp** [2] **Kaisey S. Mandel** [1]

Abstract

Type Ia supernovae (SNe Ia) are thermonuclear exploding stars that can be used to put constrains on the nature of our universe. One challenge with population analyses of SNe Ia is Malmquist bias, where we preferentially observe the brighter SNe due to limitations of our telescopes. If untreated, this bias can propagate through to our posteriors on cosmological parameters. In this paper we develop a novel technique of using a normalising flow to learn the non-analytical likelihood of observing a SN Ia for a given survey from simulations, that is independent of any cosmological model. The learnt likelihood is then used in a hierarchical Bayesian model with Hamiltonian Monte Carlo sampling to put constraints on different sets of cosmological parameters conditioned on our observed data. We verify this technique on toy model simulations finding excellent agreement with analytically-derived posteriors to within 1σ.

1. Introduction

Type Ia supernovae (SNe Ia) are violent explosions of white dwarf stars when they interact with their binary partner. All SNe Ia share a similar peak brightness and can be used as "standardisable candles". The difference between the modelled absolute brightness and apparent brightness of a standard candle is used in astronomy to estimate distance. The redshifts of SNe, caused by the universe's expansion, can be plotted against the modelled distances in a Hubble diagram to put constraints on cosmology.

In standardisation, it has been observed that SNe Ia are brighter if they are bluer, and also if they evolve more slowly around peak brightness. These effects are conventionally incorporated into a linear model (Tripp, 1998) that describes the relationship between SNe Ia absolute brightnesses M_s (in magnitudes where lower is brighter) and observable parameters:

$$M_s = m_s - \mu_s(\hat{z}_s | \boldsymbol{C}) = M_0 + \alpha \times x_s + \beta \times c_s + \epsilon \quad (1)$$

where μ_s represents the distance modulus and m_s represents the peak apparent B band magnitude of SN s. The other observable SNe parameters include the peak $B-V$ apparent colour c_s (where lower is bluer and higher is redder) and stretch parameter x_s, encoding how long the light curve takes to decay (where higher is longer). ϵ is the additional scatter coming from residual variance and peculiar velocities. The three parameters $\boldsymbol{d_s} = (m_s, c_s, x_s)^T$ can be defined and fitted using the SALT light curve model (Guy et al., 2007; 2010), that is widely adopted in SNe Ia analysis. The effect that the stretch and colour has on the absolute magnitude is determined by global hyperparameters α and β. These hyperparameters, the average absolute magnitude M_0 and cosmology $\boldsymbol{C}$ can be inferred by jointly analysing the overall SN population and their redshifts $\hat{z}_s$.

When conducting any type of cosmological inference using SNe Ia, it is important to consider selection effects, otherwise they can propagate to biases in cosmological parameter inference. A common type of bias is known as Malmquist bias (Malmquist, 1922), where surveys are more likely to detect bright objects. If this is true for SNe Ia surveys, then it would also create a bias in detecting bluer and longer lasting SNe as they are brighter.

The leading method to handle this bias is BEAMS with Bias Corrections (BBC, Kessler & Scolnic 2017) that uses state-of-the-art survey simulations from *SNANA* (Kessler et al., 2009). One-dimensional BBC involves splitting simulated SNe Ia into redshift bins. Each simulated SN Ia is fit by a model to estimate its distance modulus. For the SNe within each bin, the differences between fitted distance moduli and true distance moduli are calculated and averaged to make $\Delta\mu_{\text{bias}}(\hat{z}_s)$. To apply bias corrections on real SNe Ia, the learnt $\Delta\mu_{\text{bias}}(\hat{z}_s)$ is added to every distance modulus in the same bin. After this, the corrected Hubble diagram may be plotted and cosmology inferred. BBC may also be done in multiple dimensions, for example SNe Ia can also be be split into bins according to colour $\hat{c}_s$ and stretch $\hat{x}_s$.

[1]University of Cambridge, United Kingdom [2]Stockholm University, Sweden. Correspondence to: Benjamin M. Boyd <bmb41@cam.ac.uk>.

Accepted at the AI for Science Workshop at the 41st International Conference on Machine Learning (ICML). Copyright 2024 by the author(s).

The process works the same where the bias in distance for each bin is averaged and future real SNe Ia are corrected accordingly. Importantly, in this method the true distance moduli used in this method rely on assumed cosmology, which is carried through computing bias corrections and used to infer real cosmology which we would like to avoid in future analysis. Furthermore, this method performs analysis in separate stages where the data is modified using BBC before the cosmological model is fit.

Another approach to SN Ia cosmological analysis involves modelling the observable parameters in Equation (1) within a hierarchical Bayesian model (HBM) (March et al., 2011; Shariff et al., 2016). Malmquist bias can be accounted for by assuming the selection function $P(I_s = 1 \big| \hat{\boldsymbol{d}}_{\boldsymbol{s}})$ is analytical, such as a step function or Gaussian cumulative distribution function. We indicate a SN has been detected with $I_s = 1$. The selection function can be incorporated in a HBM framework (Rubin et al., 2015; March et al., 2018) where the data likelihood $P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} | \, \boldsymbol{\Theta})$ can be analytically modified to down-weight the probability of detecting a dim SN and up-weight the probability of detecting a bright one, such that:

$$P\big(\boldsymbol{\Theta} \big| \, \hat{\boldsymbol{D}}\big) \propto \left[\prod_{s=1}^{N_{\text{SN}}} P\big(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \big| \, I_s = 1, \hat{z}_s, \boldsymbol{\Theta}\big) \right] P\big(\boldsymbol{\Theta}\big) \tag{2}$$

$$P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} | \, I_s = 1, \hat{z}_s, \boldsymbol{\Theta}) = \frac{P(I_s = 1 \big| \, \hat{\boldsymbol{d}}_{\boldsymbol{s}}) P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} | \, \hat{z}_s, \boldsymbol{\Theta})}{\int_{-\infty}^{\infty} P(I_s = 1 \big| \, \hat{\boldsymbol{d}}_{\boldsymbol{s}}) P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} | \, \hat{z}_s, \boldsymbol{\Theta}) \, \mathrm{d}\hat{\boldsymbol{d}}_{\boldsymbol{s}}} \tag{3}$$

where $\hat{\boldsymbol{D}} = \{\hat{\boldsymbol{d}}_{\boldsymbol{s}}, \hat{z}_s, I_s = 1\}_{s=1}^{N_{\text{SN}}}$ and $\boldsymbol{\Theta} = (\boldsymbol{C}, M_0, \alpha, \beta, ...)$. These methodologies have their benefits since they are each one integrated analysis. While the assumption of a simple analytic ansatz for the selection function may be appropriate to first order, the true survey selection function induced by Malmquist bias is not known to be analytic.

Simulation-based inference (SBI) provides a way forward for the field by which the state-of-the-art simulations that capture the complex selection effects can be utilised and applied in cosmological analysis. SBI can be used to learn probability distributions and then apply them to real data to perform inference. This can be done using neural density estimation (Rezende & Mohamed, 2015; Papamakarios et al., 2021) or neural ratio estimation (Hermans et al., 2020). Previous examples of using SBI to address SN Ia selection effects in cosmological analysis have modelled the overall posterior conditioned on the entire SN population at once $P\big(\boldsymbol{\Theta} \big| \, \hat{\boldsymbol{D}} \, \big)$ (Alsing et al., 2019; Karchev et al., 2023a;b; 2024). Since the current combined SN datasets are of the order of thousands, the current methods rely on data compression to learn the complex relationship between the full SN sample and cosmology. Importantly these methods need to be trained on every cosmological parameter value for a given cosmological model, which is expensive in terms of simulation. The non-hierarchical approaches are also sensitive to the size of the sample where the model is required to be retrained if a single SN is added or removed from the sample.

In this paper we propose a new method that combines SBI with hierarchical Bayesian modelling for SN Ia cosmological analysis. This allows a non-analytical three-dimensional likelihood $P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} | \, I_s = 1, \hat{z}_s, \boldsymbol{\Theta})$ to be trained to learn the survey selection effect independent of cosmology. The same likelihood may then be re-used in HBMs to allow for constraints on different cosmological models. The HBM framework is flexible to allow for varying sample-sizes without the need to retrain the likelihood.

2. Method

In this section we outline the method used for our hierarchical simulation-based inference model used in our SN Ia cosmological analysis. In Section 2.1 we define the HBM, while in Section 2.2 we describe the normalising flow used in likelihood estimation and finally in Section 2.3 we outline the analytical toy simulations used to verify the method.

2.1. Hierarchical Bayesian Model

In this section we define the HBM (Mandel et al. 2017 Appendix D) used in our cosmological analysis. The model takes the same form as Equation (2) where $\boldsymbol{\Theta} = (\boldsymbol{C}, M_0, \alpha, \beta, \sigma_{\text{res}}^2, c_0, \alpha_c, \sigma_c^2, x_0, \sigma_x^2)$. The population hyperparameters are described as follows:

- $\boldsymbol{C}$: vector containing cosmological parameters to be fit for a given cosmological model. For example, in a flat wCDM model we might fit w and Ω_{m0} or for a ΛCDM model we might fit for Ω_{m0} and $\Omega_{\Lambda 0}$.
- M_0: the absolute magnitude constant shown in Equation (1) is the expected absolute magnitude for a SN with light curve shape $x_s = 0$ and color $c_s = 0$.
- α: the slope of the trend of absolute magnitude vs. light curve shape from Equation (1).
- β: the slope of the trend of absolute magnitude vs. colour from Equation (1).
- σ_{res}^2: the remaining residual variance around the mean trend of absolute magnitude vs. light curve shape and color.
- c_0: the expected color for a SN Ia with light curve shape x_s = 0. If $\alpha_c = 0$ then c_0 is the population mean color.

- α_c: the slope of the trend of color vs. light curve shape.
- σ_c^2: the variance around the mean trend of color vs. light curve shape.
- x_0: the mean of the x_s light curve shape population distribution.
- σ_x^2: the variance of the x_s light curve shape population distribution.

The latent peak apparent magnitude m_s is modelled using Equation (1). The additional scatter is is drawn from $\epsilon \sim N(0, \sigma_{\text{res}}^2 + \sigma_{\mu|\,z,s}^2)$ where $\sigma_{\mu|\,z,s}^2$ is additional variance from the peculiar velocities. The latent peak $B-V$ colour c_s is modelled using:

$$c_s = c_0 + \alpha_c x_s + \epsilon_c \tag{4}$$

The additional colour scatter is drawn from $\epsilon_c \sim N(0, \sigma_c^2)$. The latent stretch x_s is modelled using:

$$x_s = x_0 + \epsilon_x \tag{5}$$

the additional stretch scatter is drawn from $\epsilon_x \sim N(0, \sigma_x^2)$.

The likelihood function of the observed data $\hat{\boldsymbol{d}}_{\boldsymbol{s}}$ given the latent parameters $\boldsymbol{d_s} = (m_s, c_s, x_s)^T$ is modelled as a multivariate Gaussian such that:

$$P(\hat{\boldsymbol{d}}_{\boldsymbol{s}} |\, \boldsymbol{d_s}) = N(\hat{\boldsymbol{d}}_{\boldsymbol{s}} |\, \boldsymbol{d_s}, \boldsymbol{W_s}) \tag{6}$$

Where $\boldsymbol{W_s}$ is the error covariance matrix from the SALT light curve fitter (Guy et al., 2007; 2010). In our method we implicitly marginalise over latent variables $\boldsymbol{d_s}$ by only modelling relationships between observed parameters $\hat{\boldsymbol{d}}_{\boldsymbol{s}}$ and hyperparameters. For the toy model in Section 2.3 we can perform the marginalisation analytically.

The graphical model in Figure 1 illustrates the relationship between the above hyperparameters, latent and observed variables in our hierarchical model. We define the hyperpriors in Table 1. For sampling the hierarchical model we use Hamiltonian Monte Carlo (Duane et al., 1987), a technique that uses gradient information (something that neural network architectures also use) for faster convergence on the posterior. We use the No-U-Turn Sampler (NUTS) in *numpyro* (Bingham et al., 2019; Phan et al., 2019), a probabilistic programming package for automatic differentiation and just-in-time (JIT) compilation built with *JAX* (Bradbury et al., 2018).

For the observed data likelihood we write:

$$P\big(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \big|\, I_s = 1, \hat{z}_s, \boldsymbol{\Theta}\big) = g\big(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \big| m_0^s = M_0 + \mu_s(\hat{z}_s|\boldsymbol{C}), \boldsymbol{W_s}, \boldsymbol{\Theta}_{\text{SN}}\big) \tag{7}$$

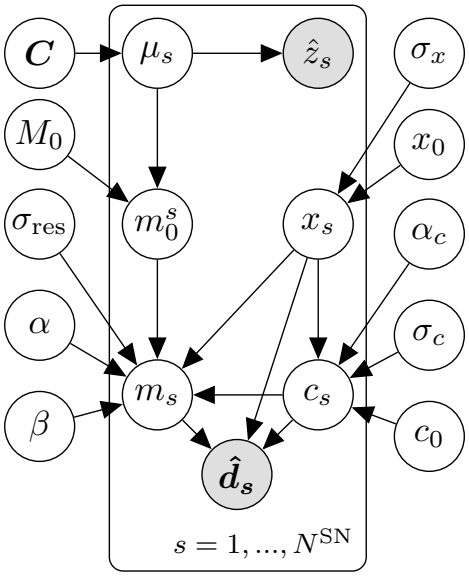

Figure 1. Graphical model illustrating the dependencies between hyperparameters, latent parameters and the observed data (shaded in grey) in the hierarchical Bayesian model.

where $\boldsymbol{\Theta}_{\text{SN}} = (\alpha, \beta, \sigma_{\text{res}}^2, c_0, \alpha_c, \sigma_c^2, x_0, \sigma_x^2)$ contains the SN hyperparameters and m_s^0 is the expected apparent magnitude for a SN with zero colour and stretch such that $m_s^0 = \text{E}[m_s|c_s = 0, x_s = 0, \hat{z}_s, M_0, \boldsymbol{C}]$. It is useful to define the auxiliary density $g\big(\hat{\boldsymbol{d}}|m_0, \boldsymbol{W}, \boldsymbol{\Theta}_{\text{SN}}\big)$, to take advantage of the fact that the observed data likelihood depends on $(\hat{z}_s, M_0, \boldsymbol{C})$ only through the combination $m_0^s = M_0 + \mu(\hat{z}_s|\boldsymbol{C})$. This same auxiliary density can be reused to define the likelihood for constraining different cosmological models.

The auxiliary density can be learnt with simulation-based inference as we describe in Section 2.2 or for some special cases derive it analytically as we do in Section 2.3.

2.2. Training the Normalising Flow to Approximate the Likelihood

To learn our three-dimensional auxiliary density $g\big(\hat{\boldsymbol{d}}|m_0, \boldsymbol{W}, \boldsymbol{\Theta}_{\text{SN}}\big)$ we use a normalising flow (Tabak & Vanden-Eijnden, 2010; Tabak & Turner, 2013). We select a normalising flow instead of a mixture density network (Bishop, 1994) since we require the likelihood is flexible enough to capture the non-analytical selection effects from realistic survey simulations. The normalising flow q_ϕ, used to approximate g, is defined such that:

$$q_\phi\big(\hat{\boldsymbol{d}}|m_0, \boldsymbol{W}, \boldsymbol{\Theta}_{\text{SN}}\big) = N(\boldsymbol{u_0} |\, \mathbf{0,I}) \prod_{l=1}^{L} \left| \frac{\partial f_l^{\boldsymbol{\phi}}}{\partial \boldsymbol{u_{l-1}}} \right|^{-1} \tag{8}$$

where $\boldsymbol{u_0} \sim N(\mathbf{0,I})$, $\boldsymbol{u_l} = f_l^{\boldsymbol{\phi}}(\boldsymbol{u_{l-1}}, \boldsymbol{x})$ and $\hat{\boldsymbol{d}} = f_L^{\boldsymbol{\phi}}$. The weights and biases $\boldsymbol{\phi}$ for the bijective transforms f_l

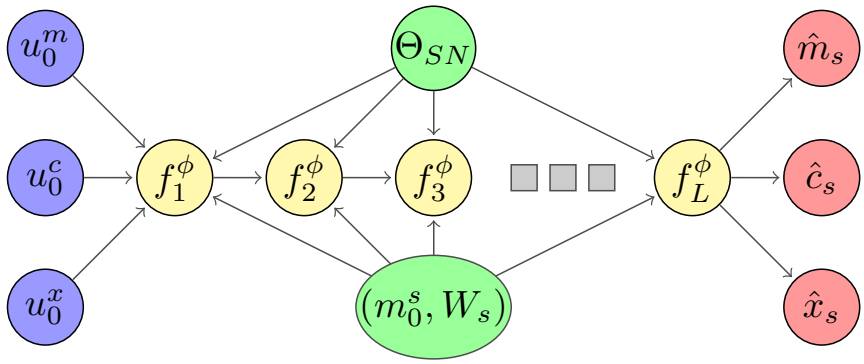

Figure 2. Illustration of how parameters interact with the normalising flow that models $q_\phi\left(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \middle| m_0^s, \boldsymbol{W_s}, \boldsymbol{\Theta}_{\text{SN}}\right)$ to approximate the auxiliary density g for observed data likelihood. The three u_0 parameters are all drawn from unit normal distributions and passed through L transforms f_l with learnable parameters ϕ. The transforms are conditioned on the expected apparent magnitude m_0^s for a given supernova (at zero stretch and colour), measurement covariances $\boldsymbol{W_s}$ and supernova hyperparameters $\boldsymbol{\Theta}_{\text{SN}}$. The outputs after f_L are the three SALT parameters representing observed apparent magnitude $\hat{m}_s$, colour $\hat{c}_s$ and stretch $\hat{x}_s$

Table 1. Hyperparameters and respective hyperpriors used in our hierarchical Bayesian model. Detailed also are the ranges used in the likelihood training. $U(B_L, B_U)$ is a uniform distribution between B_L and B_U, while $HN(0, 1^2)$ represents a positive half-normal distribution centred at 0 with a variance of 1.

HYPERPARAMETER	HYPERPRIOR	TRAINING RANGE
w_0 (FLAT wCDM)	$U(-2, 2)$	-
Ω_{m0} (FLAT wCDM)	$U(0, 1)$	-
w_0 (ΛCDM)	$U(0, 2)$	-
$\Omega_{\Lambda 0}$ (ΛCDM)	$U(0, 2)$	-
M_0	$U(-\infty, \infty)$	-
α	$U(-\infty, \infty)$	$U(-0.55, 0.25)$
β	$U(-\infty, \infty)$	$U(0, 4)$
σ_{res}	$HN(0, 1^2)$	$U(0, 0.2)$
c_0	$U(-\infty, \infty)$	$U(-0.4, 0.4)$
α_c	$U(-\infty, \infty)$	$U(-0.03, 0.03)$
σ_c	$HN(0, 1^2)$	$U(0, 0.1)$
x_0	$U(-\infty, \infty)$	$U(-1, 1)$
σ_x	$HN(0, 1^2)$	$U(0, 2)$

that need to be optimised by maximising the log likelihood via stochastic gradient descent. It is simple to evaluate normalising flows with the chain rule as their architectures are designed to have lower triangular Jacobian matrices, meaning the determinant is the product of diagonal terms. Figure 2 illustrates how the parameters are given to the normalising flow.

For this work we use the Masked Autoregressive Flow (MAF) architecture (Papamakarios et al., 2017) that involves stacking Masked Autoencoder for Distribution Estimation (MADE) blocks (Germain et al., 2015). MADE blocks were thought to be advantageous for this problem since they use one dimension to predict the next such that $P(\boldsymbol{d}|\theta) = \prod_n P(d_n|\, d_{1:n-1}, \theta)$, which is the case in our problem where $P(\hat{m}_s|\, \hat{c}_s, \hat{x}_s, \boldsymbol{\Theta}_{\text{SN}})$ and $P(\hat{c}_s|\, \hat{x}_s, \boldsymbol{\Theta}_{\text{SN}})$.

To train the normalising flow with simulations the $\boldsymbol{\Theta}_{\text{SN}}$ hyperparameters are sampled from a wide non-informative prior. We also sample m_0^s to ensure the normalising flow is not influenced by any cosmology $\boldsymbol{C}$. Furthermore it is necessary to put an emphasis on sampling the dimmest m_0^s values closer to the detection limit where Malmquist bias has the greatest effect. After these parameters are drawn the SNe may be forward simulated according to Equation (1) and the further equations in Section 2.1. The survey simulation then determines which SNe are then detected. For state-of-the-art survey simulations such as those in *SNANA* (Kessler et al., 2009), this includes multiple complex factors such as telescope pointing plans, systematics and the weather. For the purpose of our tests the selection is determined by an analytical function of the observed data $\hat{\boldsymbol{d}}_{\boldsymbol{s}}$ defined in Equation (9). It is only the data and parameters of the "observed" simulated SN ($I_s = 1$) that are given to the normalising flow for training the auxiliary density $q_\phi\left(\hat{\boldsymbol{d}} \middle| m_0, \boldsymbol{W}, \boldsymbol{\Theta}_{\text{SN}}\right)$. Further details on how we train the normalising flow for our toy problem, including architectures and runtime, are detailed at the end of Section 2.3.

2.3. Analytical Toy Simulations

In this we section define a toy Malmquist bias simulation that has an analytical likelihood to validate the method outlined above.

We begin by sampling 10,000 simulated SN from the redshift distribution $P(\hat{z}_s) \propto (\hat{z}_s + 1)^{3/2}$. We convert these into distance $\mu_s(\hat{z}_s|\, \boldsymbol{C})$ using *JAX-cosmo* (Campagne et al., 2023) conditioned on our cosmology $\boldsymbol{C}$. We define our true cosmology as a flat ΛCDM cosmology ($h = 0.7324$, $\Omega_{m0} = 0.28$, $\Omega_{\Lambda 0} = 0.72$, $w_0 = -1$ and $w_a = 0$). We later infer posteriors on a flat wCDM universe where $\boldsymbol{C_1} = (w_0, \Omega_{m0})$ and a potentially un-flat ΛCDM universe where $\boldsymbol{C_2} = (\Omega_{m0}, \Omega_{\Lambda 0})$. Additional scatter is added to μ_s according to $N(0, \sigma^2_{\mu|\, z,s})$ where $\sigma^2_{\mu|\, z,s}$ is variance from

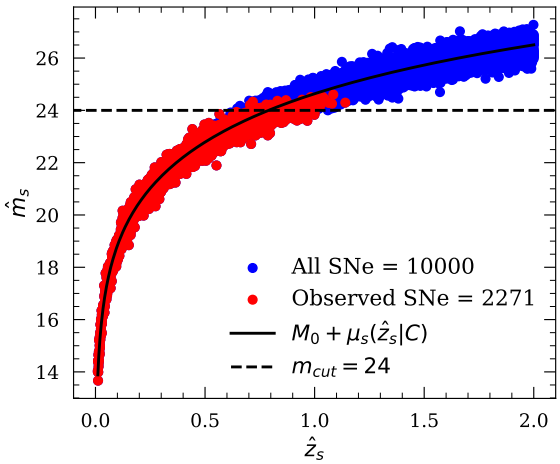

Figure 3. Toy simulations showing the effects of using the Gaussian cumulative distribution selection function to model Malmquist bias on the Hubble diagram. Plotted on the horizontal axis are the simulated spectroscopic redshifts and on the vertical axis are simulated apparent magnitudes of the SNe. In blue are all the SNe Ia simulated according to a flat ΛCDM universe ($h = 0.7324, \Omega_{m_0} = 0.28, \Omega_{\Lambda_0} = 0.72, w_0 = -1$). Overplotted in red are the SNe that would have been detected according to a Gaussian cumulative distribution selection function centred at $m_{\text{cut}} = 24$ with standard deviation $\sigma_{\text{cut}} = 0.25$. The black solid line shows the effect of the true underlying cosmology that needs to be recovered.

peculiar velocities of host galaxies. This is approximated as $\sigma^2_{\mu|\,z,s} \approx [5/(\hat{z}_s \ln 10)]^2[\hat{\sigma}^2_{z,s} + \sigma^2_{pec}/c^2]$ where c is the speed of light in kms^{-1}, $\hat{\sigma}_{z,s}$ is the observed redshift measurement uncertainty and σ_{pec} is the peculiar velocity dispersion that we set to 200 kms^{-1}. Our simulations are designed to replicate high-precision spectroscopic redshifts so we set $\hat{\sigma}_{z,s} = 0$.

After we have the distance modulus μ_s for each SN, we forward simulate the observed parameters $\hat{\boldsymbol{d}}_{\boldsymbol{s}}$ using Equation (1) and the equations detailed at the beginning of Section 2.1. The true values of each hyperparameter used in the simulation are shown in Table 2. We assume a diagonal measurement covariance matrix $\boldsymbol{W}_s$. The measurement variances for apparent magnitudes are sampled according to $\ln W_s^{mm} \sim N(0.2(m_s - 56), 1.2^2)$, where m_s is the apparent magnitude before the observed variance is applied. For apparent colour measurement variances we sample from $\ln W_s^{cc} \sim N(-7.0, 0.6^2)$ and for apparent stretch measurement variances we sample from $\ln W_s^{xx} \sim N(-3, 1^2)$.

Following the simulation of SN observables $\hat{\boldsymbol{d}}_{\boldsymbol{s}}$ we pass them through an analytical selection function to model Malmquist bias (Rubin et al., 2015) such that:

$$P(I_s = 1|\,\hat{\boldsymbol{d}}_{\boldsymbol{s}}) = \Phi\left(\frac{m_{\text{cut}} - (\hat{m}_s + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma_{\text{cut}}}\right) \tag{9}$$

where $\Phi(\cdot)$ is a unit normal cumulative distribution function. m_{cut} represents the limiting magnitude of the survey and σ^2_{cut} represents the dispersion around the limiting magnitude. a_{cut} is included to allow the possibility that longer lasting SNe are more likely to be detected and b_{cut} allows the colour of the SN to influence the selection probability if the survey was not done in a B band filter. For the purpose of our simulations we set $m_{\text{cut}} = 24$, $\sigma_{\text{cut}} = 0.25$, $a_{\text{cut}} = -0.1$ and $b_{\text{cut}} = -1$. After the selection function is defined we perform one Bernoulli trial for each SN with their respective selection probability to determine if they are selected. In our subsequent analysis we then only use the 2,271 SN that have been selected after the Bernoulli trial ($I_s = 1$). Figure 3 shows how the Malmquist bias selection function affects the SN that are observed.

The auxiliary density for the observed SNe in the simulations defined above can be derived analytically. This can done similarly to Equation (3) such that:

$$\begin{aligned} g(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|m_0^s, \boldsymbol{W}_{\boldsymbol{s}}, \boldsymbol{\Theta}_{\text{SN}}) &= \\ &\frac{P(I_s = 1|\,\hat{\boldsymbol{d}}_{\boldsymbol{s}})P(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|\,m_0^s, \boldsymbol{\Theta}_{\text{SN}})}{\int_{-\infty}^{\infty} P(I_s = 1|\,\hat{\boldsymbol{d}}_{\boldsymbol{s}})P(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|\,m_0^s, \boldsymbol{\Theta}_{\text{SN}})\,\mathrm{d}\hat{\boldsymbol{d}}_{\boldsymbol{s}}} \\ &= \frac{P(I_s = 1|\,\hat{\boldsymbol{d}}_{\boldsymbol{s}})N(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|\,\boldsymbol{\Psi}_{\boldsymbol{d}}, \boldsymbol{\Sigma}_{\boldsymbol{d}})}{\int_{-\infty}^{\infty} P(I_s = 1|\,\hat{\boldsymbol{d}}_{\boldsymbol{s}})N(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|\,\boldsymbol{\Psi}_{\boldsymbol{d}}, \boldsymbol{\Sigma}_{\boldsymbol{d}})\,\mathrm{d}\hat{\boldsymbol{d}}_{\boldsymbol{s}}} \end{aligned} \tag{10}$$

$N(\hat{\boldsymbol{d}}_{\boldsymbol{s}}|\,\boldsymbol{\Psi}_{\boldsymbol{d}}, \boldsymbol{\Sigma}_{\boldsymbol{d}})$ is a multivariate Gaussian with:

$$\boldsymbol{\Psi}_{\boldsymbol{d}} = \begin{pmatrix} \mathrm{E}[\hat{m}_s] \\ \mathrm{E}[\hat{c}_s] \\ \mathrm{E}[\hat{x}_s] \end{pmatrix} = \begin{pmatrix} m_0^s + \alpha x_0 + \beta(\alpha_c x + c_0) \\ \alpha_c x + c_0 \\ x_0 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{d}} = \begin{pmatrix} \mathrm{Var}[\hat{m}_s] & \mathrm{Cov}[\hat{m}_s, \hat{c}_s] & \mathrm{Cov}[\hat{m}_s, \hat{x}_s] \\ \mathrm{Cov}[\hat{m}_s, \hat{c}_s] & \mathrm{Var}[\hat{c}_s] & \mathrm{Cov}[\hat{c}_s, \hat{x}_s] \\ \mathrm{Cov}[\hat{m}_s, \hat{x}_s] & \mathrm{Cov}[\hat{c}_s, \hat{x}_s] & \mathrm{Var}[\hat{x}_s] \end{pmatrix}$$

$$\begin{aligned} \mathrm{Var}[\hat{m}_s] &= \sigma^2_{\text{res}} + \beta^2\sigma_c^2 + (\alpha + \beta\alpha_c)^2\sigma_x^2 + W_s^{mm} \\ \mathrm{Var}[\hat{c}_s] &= \sigma_c^2 + \alpha_c^2\sigma_x^2 + W_s^{cc} \\ \mathrm{Var}[\hat{x}_s] &= \sigma_x^2 + W_s^{xx} \\ \mathrm{Cov}[\hat{m}_s, \hat{c}_s] &= \alpha\alpha_c\sigma_x^2 + \beta(\sigma_c^2 + \alpha_c^2\sigma_x^2) + W_s^{mc} \\ \mathrm{Cov}[\hat{m}_s, \hat{x}_s] &= \alpha\sigma_x^2 + \beta\alpha_c\sigma_x^2 + W_s^{mx} \\ \mathrm{Cov}[\hat{c}_s, \hat{x}_s] &= \alpha_c\sigma_x^2 + W_s^{cx} \end{aligned}$$

The integral in the denominator of Equation (10) can be solved using the conditional properties of multivariate Gaussian distributions to split the likelihood into the product of

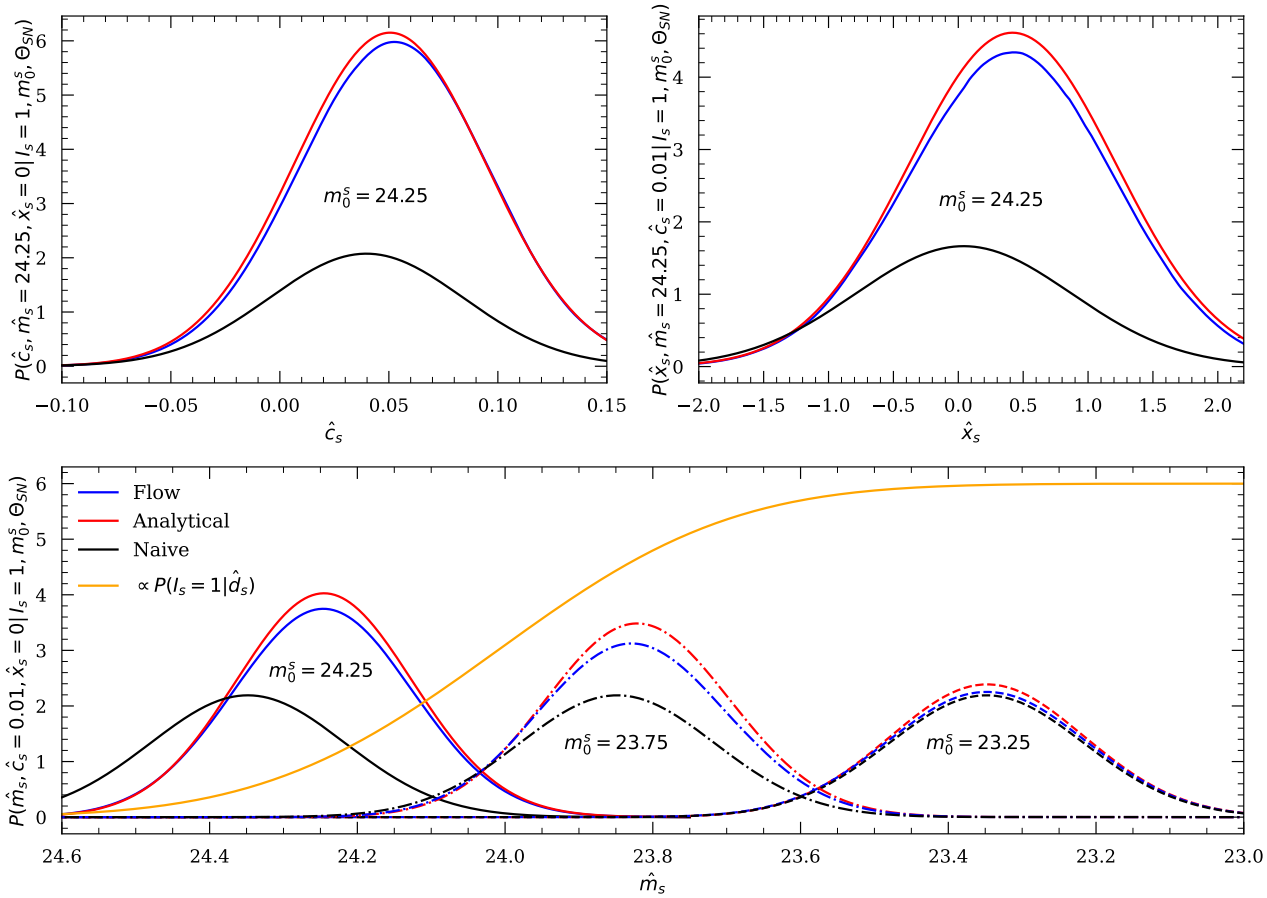

Figure 4. This plot compares the learnt normalising flow likelihood (blue) with the analytical data likelihood (red). We also plot the naive likelihood when it is just assumed to be Gaussian without considering selection effects in black. The bottom panel illustrates the likelihood as a function of apparent brightnesses for three different SNe. The selection function (orange) has a greater effect on dimmer SNe to the left of the plot. The top two panels illustrate the likelihood of the dimmest SN as function of observed colour and stretch. The other parameters used to generate these plots were set arbitrarily as $\boldsymbol{\Theta}_{\text{SN}} = (c_0 = 0.1, x_0 = -0.5, \sigma_{\text{res}} = 0.1, \sigma_c = 0.05, \sigma_x = 1, \alpha = -0.1, \beta = 2.7, \alpha_c = 0)$, $W_s^{mm} = 0.042$, $W_s^{cc} = 0.031$ and $W_s^{xx} = 0.25$.

three univariate Gaussian distributions. Appendix A shows the three dimensions can be marginalised over one after another using the identity:

$$\int_{-\infty}^{\infty} \Phi((\mu_1 - x)/\sigma_1) N(x | \mu_2, \sigma_2^2)\, \mathrm{d}x = \Phi\left(\frac{\mu_1 - \mu_2}{\sigma_1} \times \left(1 + \frac{\sigma_2^2}{\sigma_1^2}\right)^{-\frac{1}{2}}\right) \quad (11)$$

We then proceed to training a normalising flow to learn from simulations the same likelihood we found analytically using the steps described in Section 2.2. The training set is constructed by sampling from wide range of $\boldsymbol{\Theta}_{\text{SN}}$ hyperparameter values as defined in Table 1. We sample 80% of m_0^s from a negative half-normal distribution $HN(m_{\text{cut}} + 3\sigma_{\text{cut}}, 1^2)$ to ensure there are enough examples of dim objects at the detection limit where the selection effect has the greatest influence. The remaining 20% of m_0^s examples are drawn from $U(11, m_{\text{cut}})$ to ensure the normalising flow has seen the full range of magnitudes.

It helps in training to normalise the dimensions by first centering the data $\hat{\boldsymbol{d}}_{\boldsymbol{s}} \to \hat{\boldsymbol{d}}_{\boldsymbol{s}} - \boldsymbol{d_0}$ where $\boldsymbol{d_0} = (m_0^s, c_0, x_0)^T$. After this we do minimum/maximum normalisation across both data and parameters. In our example we stack $L = 4$ MADE blocks, reversing the conditional order of the three dimensions in between each block. The architecture of the blocks includes two hidden layers of size 128 and ReLu activation functions. In total the normalising flow contains 143,384 learnable parameters. We use batch sizes of 2000 and train on 250,000 random batches generated on-the-fly taking approximately 4 hours on a GPU. We implement the normalising flow using *JAX* (Bradbury et al., 2018) to allow for compatibility with the HBM. Sampling the HBM that uses the normalising flow takes less than 20 CPU minutes with Hamiltonian Monte Carlo to produce 500 warm-up samples followed by 500 posterior samples for 4 chains.

After both the analytical likelihood is derived and the normalising flow likelihood is trained, we alternately use them in the hierarchical model defined in Section 2.1 to put con-

Table 2. In this table we present the mean and the standard deviation of the posteriors from our hierarchical models when applied to the toy simulations. We compare the true parameter values with the inferred normalising flow and analytical model posterior estimates.

TEST		FLAT wCDM		ΛCDM	
HYPERPARAMETER	TRUTH	FLOW	ANALYTICAL	FLOW	ANALYTICAL
w_0	-1	-0.95±0.11	-0.96±0.13	FIXED	FIXED
Ω_{m0}	0.28	0.28±0.04	0.28±0.05	0.28±0.04	0.28±0.04
$\Omega_{\Lambda 0}$	0.72	FIXED	FIXED	0.67±0.07	0.67±0.08
M_0	-19.392	-19.386±0.011	-19.382±0.013	-19.386±0.012	-19.381±0.011
α	-0.154	-0.155±0.003	-0.157±0.003	-0.155±0.003	-0.157±0.003
β	2.252	2.3±0.05	2.27±0.05	2.3±0.05	2.28±0.05
σ_{res}	0.1	0.101±0.003	0.101±0.003	0.101±0.003	0.1±0.003
c_0	-0.061	-0.063±0.002	-0.06±0.002	-0.063±0.002	-0.06±0.002
α_c	-0.008	-0.007±0.001	-0.008±0.001	-0.007±0.001	-0.008±0.001
σ_c	0.065	0.065±0.001	0.065±0.001	0.065±0.001	0.065±0.001
x_0	-0.432	-0.44±0.026	-0.443±0.027	-0.441±0.026	-0.444±0.026
σ_x	1.124	1.152±0.02	1.122±0.018	1.155±0.02	1.121±0.018

straints on different cosmological parameters and compare their results.

3. Results and Discussion

In this section we present the results of our method on our toy simulations and discuss future work.

Before analysing the hierarchical model posteriors we can take an intermediate step of validation by comparing the analytically derived likelihood and the normalising flow likelihood. Figure 4 compares the two for a given set of parameters. It is clear that the selection effect only comes into play for the dimmer SNe with magnitudes greater than m_{cut}. Brighter SN have likelihoods that are closer to a typical multivariate Gaussian. The normalising flow was able to account for selection effects in all three dimensions with a probability distribution functions close to the analytical solution. Despite having different centerings and normalisations, the likelihoods for the toy example look approximately Gaussian in shape. However, if σ_{cut} decreases or more realistic simulations are used the shape of the likelihood becomes more skewed, hence why it is still important to use a flexible normalising flow.

After the normalising flow was validated on the analytical likelihood solution, both were then used in hierarchical models to put constraints on cosmological parameters within two types of cosmological models as well as the other SN parameters. Table 2 displays the means and standard deviations of the resulting posteriors as well as the true simulated values. Figure 5 shows the inferred posteriors for the key hyperparameters. The posteriors utilising the normalising flow were able to recover the true simulated cosmology and supernova hyperparameters to within 2σ and agreed with the analytical posteriors within 1σ. The normalising flow likelihood was also flexible enough to be reused for both cosmological model fits reproducing almost identical posteriors on the SN hyperparameters. It is clear from these results that this technique combining SBI and hierarchical Bayesian modelling is effective in reproducing constraints on different cosmological parameters by correctly accounting for Malmquist bias.

Now that the method has been validated on simulations with analytical likelihoods, the next stage is to apply it to more realistic survey simulations with non-analytical likelihoods. The natural comparison is to train the normalising flow with the simulations from *SNANA* (Kessler et al., 2009) and then compare inferred posteriors with the popular BBC method that uses binning. This work is already in progress and will provide useful insights to the effectiveness of BBCs uncertainty estimation and extent that it is independent of the fiducial simulated cosmology.

Other avenues for this work include hierarchical analysis on SNe from multiple different surveys at once. This is would be done by splitting the likelihood for N_{Surveys} such that:

$$P\big(\boldsymbol{\Theta}\big|\,\hat{\boldsymbol{D}}\big) \propto \left[\prod_{j=1}^{N_{\text{Surveys}}}\prod_{s=1}^{N_{\text{SN}}^{j}} P_j\big(\hat{\boldsymbol{d}}_{\boldsymbol{s}}\big|\,I_s=1,\hat{z}_s,\boldsymbol{\Theta}\big)\right]P\big(\boldsymbol{\Theta}\big) \quad (12)$$

where $P_j\big(\hat{\boldsymbol{d}}_{\boldsymbol{s}}\big|\,I_s=1,\hat{z}_s,\boldsymbol{\Theta}\big)$ is the observed data likelihood for survey j. Each survey has its own selection effect so this would require training N_{Survey} normalising flows to model the individual likelihoods. SN Ia datasets that could be used for this type of analysis include the Pantheon+ (Scolnic et al., 2022) and Union (Rubin et al., 2023) compilations.

Before conducting credible large scale cosmological analyses with this method it is important to think about how survey systematics can be modelled. Survey systematics include correlations between SNe due to connecting factors,

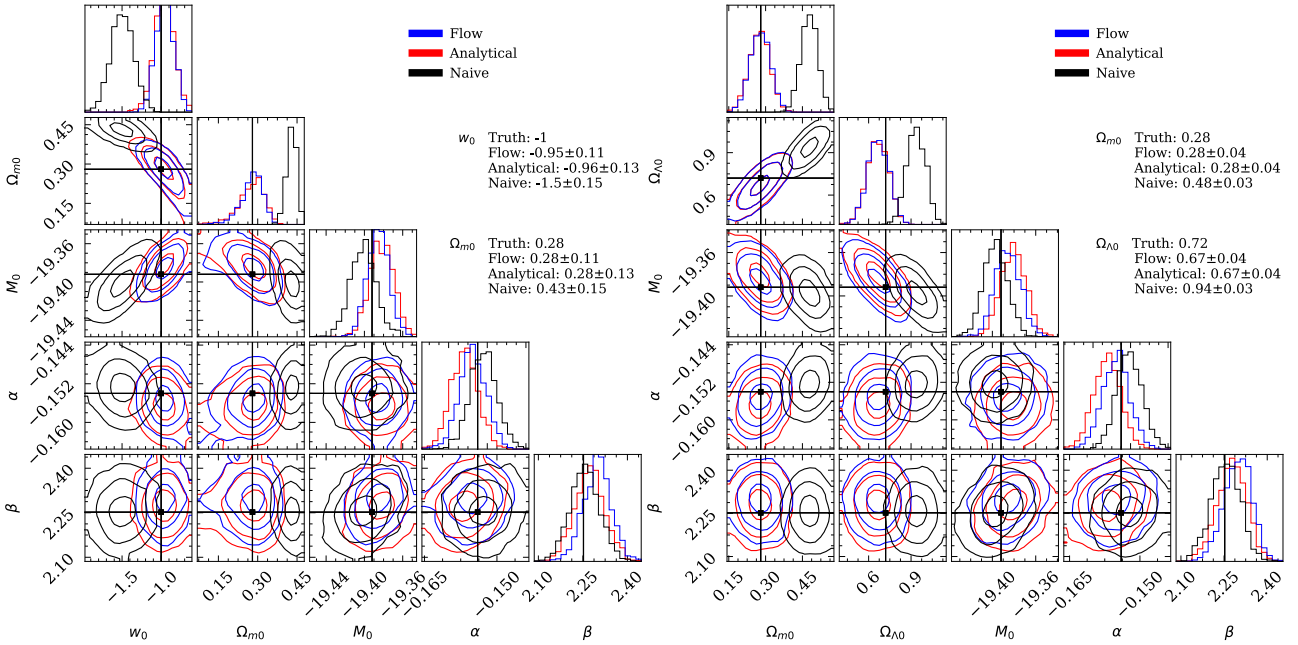

Figure 5. Corner plots showing inferred posteriors for our toy simulations. The left contours infer constrains on a flat wCDM cosmology whilst the right contours infer constraints on a ΛCDM cosmology. The other three parameters inferred are from Equation (1) and describe the relationship between SN characteristics and absolute brightness. The blue contours are from the HBM that uses the normalising flow likelihood whilst the red contours use the analytical likelihood. We also include black contours that use a naive Gaussian likelihood that does not take selection effects into account. The true hyperparameters used in the simulations are indicated with black straight lines.

such as them being observed on the same evening or belonging to the same host galaxy. Traditionally, this information is captured by a $N_{\text{SN}} \times N_{\text{SN}}$ Gaussian covariance matrix, however, since we assume SNe are conditionally independent given SN hyperparameters, we currently do not include this. Explicitly conditioning on such latent common factors within our likelihoods could potentially solve this problem, but this requires further work.

Another use for this methodology would be to provide constraints on the astrophysics of SNe Ia. *Simple-BayeSN* (Mandel et al., 2017) uses a similar hierarchical model framework but separates variations in intrinsic SN properties and extrinsic dust effects. This allows the model to make inferences on SN Ia dust environments including the "mass-step" (Kelly et al., 2010; Sullivan et al., 2010; Lampeitl et al., 2010; Childress et al., 2013) phenomenon where SNe Ia in larger host galaxies appear brighter. In principle our methodology is robust and flexible enough to be extended to these problems to allow for further analysis whilst still correcting for Malmquist bias.

4. Conclusion

We have developed a simple methodology for Type Ia supernova cosmology that accounts for Malmquist bias using simulation-based inference and hierarchical Bayesian modelling. This method is unique compared to other simulation-based inference applications in the field, because it can infer different sets of cosmological parameters and is robust to varying SN sample sizes, without needing to be retrained. The likelihood to be learnt is only over three dimensions, does not require data compression beyond the three SN light curve summary parameters and importantly is trained independent of cosmology. These factors allow the likelihood to be easier to learn in a domain where simulations may be expensive.

We validate our method on a toy simulation that has an analytical likelihood solution. Results showed that the method was able to recover true simulated cosmology and most supernova hyperparameters and agreed with the analytical posteriors to within 1σ. The method provided almost identical posteriors on supernova hyperparameters when inferring parameters from two different cosmological models.

This technique can be extended to provide useful insights on the astrophysics of SNe as it is flexible enough to incorporate intrinsic and extrinsic host galaxy environment effects. Furthermore, we are training the model on state-of-the-art survey simulations to learn non-analytical likelihoods caused by more complex selection effects. These learnt likelihoods may be used in future hierarchical cosmological analyses combining multiple surveys to further constrain the nature of our universe.

Acknowledgements

BMB is supported by the Cambridge Centre for Doctoral Training in Data-Intensive Science funded by the UK Science and Technology Facilities Council (STFC) and a G-Research early career researchers grant used for equipment. MG and KSM are supported by the European Union's Horizon 2020 research and innovation programme under European Research Council (ERC) Grant Agreement No. 101002652 and Marie Skłodowska Curie Grant Agreement No. 873089. ST was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101018897 CosmicExplorer).

References

Alsing, J., Charnock, T., Feeney, S., and Wandelt, B. Fast likelihood-free cosmology with neural density estimators and active learning. *Mon. Not. R. Astron. Soc.*, 488 (3):4440–4458, September 2019. doi: 10.1093/mnras/stz1960.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P. A., Horsfall, P., and Goodman, N. D. Pyro: Deep universal probabilistic programming. *J. Machine Learning Res.*, 20:28:1–28:6, 2019. URL http://jmlr.org/papers/v20/18-403.html.

Bishop, C. M. Mixture density networks. Technical Report NCRG/94/004, Aston University, 1994. URL https://publications.aston.ac.uk/id/eprint/373/1/NCRG_94_004.pdf.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Campagne, J.-E., Lanusse, F., Zuntz, J., Boucaud, A., Casas, S., Karamanis, M., Kirkby, D., Lanzieri, D., Peel, A., and Li, Y. JAX-COSMO: An End-to-End Differentiable and GPU Accelerated Cosmology Library. *Open J. Astrophys.*, 6:15, April 2023. doi: 10.21105/astro.2302.05163.

Childress, M., Aldering, G., Antilogus, P., Aragon, C., Bailey, S., Baltay, C., Bongard, S., Buton, C., Canto, A., Cellier-Holzem, F., Chotard, N., Copin, Y., Fakhouri, H. K., Gangler, E., Guy, J., Hsiao, E. Y., Kerschhaggl, M., Kim, A. G., Kowalski, M., Loken, S., Nugent, P., Paech, K., Pain, R., Pecontal, E., Pereira, R., Perlmutter, S., Rabinowitz, D., Rigault, M., Runge, K., Scalzo, R., Smadja, G., Tao, C., Thomas, R. C., Weaver, B. A., and Wu, C. Host Galaxy Properties and Hubble Residuals of Type Ia Supernovae from the Nearby Supernova Factory. *Astrophys. J.*, 770(2):108, June 2013. doi: 10.1088/0004-637X/770/2/108.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. Hybrid monte carlo. *Phys. Lett. B*, 195(2):216–222, 1987. ISSN 0370-2693. doi: 10.1016/0370-2693(87)91197-X.

Germain, M., Gregor, K., Murray, I., and Larochelle, H. Made: Masked autoencoder for distribution estimation. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 881–889, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/germain15.html.

Guy, J., Astier, P., Baumont, S., Hardin, D., Pain, R., Regnault, N., Basa, S., Carlberg, R. G., Conley, A., Fabbro, S., Fouchez, D., Hook, I. M., Howell, D. A., Perrett, K., Pritchet, C. J., Rich, J., Sullivan, M., Antilogus, P., Aubourg, E., Bazin, G., Bronder, J., Filiol, M., Palanque-Delabrouille, N., Ripoche, P., and Ruhlmann-Kleider, V. SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *Astron. Astrophys.*, 466(1):11–21, April 2007. doi: 10.1051/0004-6361:20066930.

Guy, J., Sullivan, M., Conley, A., Regnault, N., Astier, P., Balland, C., Basa, S., Carlberg, R. G., Fouchez, D., Hardin, D., Hook, I. M., Howell, D. A., Pain, R., Palanque-Delabrouille, N., Perrett, K. M., Pritchet, C. J., Rich, J., Ruhlmann-Kleider, V., Balam, D., Baumont, S., Ellis, R. S., Fabbro, S., Fakhouri, H. K., Fourmanoit, N., González-Gaitán, S., Graham, M. L., Hsiao, E., Kronborg, T., Lidman, C., Mourao, A. M., Perlmutter, S., Ripoche, P., Suzuki, N., and Walker, E. S. The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *Astron. Astrophys.*, 523:A7, November 2010. doi: 10.1051/0004-6361/201014468.

Hermans, J., Begy, V., and Louppe, G. Likelihood-free MCMC with amortized approximate ratio estimators. In Daumé III, H. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4239–4248. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/hermans20a.html.

Karchev, K., Trotta, R., and Weniger, C. SICRET: Supernova Ia Cosmology with truncated marginal neural Ratio EsTimation. *Mon. Not. R. Astron. Soc.*, 520(1): 1056–1072, March 2023a. doi: 10.1093/mnras/stac3785.

Karchev, K., Trotta, R., and Weniger, C. SimSIMS: Simulation-based Supernova Ia Model Selection with

thousands of latent variables. *arXiv e-prints*, art. arXiv:2311.15650, November 2023b. doi: 10.48550/arXiv.2311.15650.

Karchev, K., Grayling, M., Boyd, B. M., Trotta, R., Mandel, K. S., and Weniger, C. SIDE-real: Supernova Ia Dust Extinction with truncated marginal neural ratio estimation applied to real data. *Mon. Not. R. Astron. Soc.*, 530(4): 3881–3896, June 2024. doi: 10.1093/mnras/stae995.

Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., and Kirshner, R. P. Hubble Residuals of Nearby Type Ia Supernovae are Correlated with Host Galaxy Masses. *Astrophys. J.*, 715(2):743–756, June 2010. doi: 10.1088/0004-637X/715/2/743.

Kessler, R. and Scolnic, D. Correcting Type Ia Supernova Distances for Selection Biases and Contamination in Photometrically Identified Samples. *Astrophys. J.*, 836(1):56, February 2017. doi: 10.3847/1538-4357/836/1/56.

Kessler, R., Bernstein, J. P., Cinabro, D., Dilday, B., Frieman, J. A., Jha, S., Kuhlmann, S., Miknaitis, G., Sako, M., Taylor, M., and Vanderplas, J. SNANA: A Public Software Package for Supernova Analysis. *Publ. Astron. Soc. Pacific*, 121(883):1028, September 2009. doi: 10.1086/605984.

Lampeitl, H., Smith, M., Nichol, R. C., Bassett, B., Cinabro, D., Dilday, B., Foley, R. J., Frieman, J. A., Garnavich, P. M., Goobar, A., Im, M., Jha, S. W., Marriner, J., Miquel, R., Nordin, J., Östman, L., Riess, A. G., Sako, M., Schneider, D. P., Sollerman, J., and Stritzinger, M. The Effect of Host Galaxies on Type Ia Supernovae in the SDSS-II Supernova Survey. *Astrophys. J.*, 722(1):566–576, October 2010. doi: 10.1088/0004-637X/722/1/566.

Malmquist, K. G. On some relations in stellar statistics. *Medd. Lunds Astron. Obser. Serie I*, 100:1–52, March 1922.

Mandel, K. S., Scolnic, D. M., Shariff, H., Foley, R. J., and Kirshner, R. P. The Type Ia Supernova Color-Magnitude Relation and Host Galaxy Dust: A Simple Hierarchical Bayesian Model. *Astrophys. J.*, 842(2):93, June 2017. doi: 10.3847/1538-4357/aa6038.

March, M. C., Trotta, R., Berkes, P., Starkman, G. D., and Vaudrevange, P. M. Improved constraints on cosmological parameters from Type Ia supernova data. *Mon. Not. R. Astron. Soc.*, 418(4):2308–2329, December 2011. doi: 10.1111/j.1365-2966.2011.19584.x.

March, M. C., Wolf, R. C., Sako, m., D'Andrea, C., and Brout, D. A Bayesian approach to truncated data sets: An application to Malmquist bias in Supernova Cosmology. *arXiv e-prints*, art. arXiv:1804.02474, April 2018. doi: 10.48550/arXiv.1804.02474.

Papamakarios, G., Pavlakou, T., and Murray, I. Masked autoregressive flow for density estimation. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Machine Learning Res.*, 22(57):1–64, 2021. URL http://jmlr.org/papers/v22/19-1028.html.

Phan, D., Pradhan, N., and Jankowiak, M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *arXiv e-prints*, art. arXiv:1912.11554, December 2019. doi: 10.48550/arXiv.1912.11554.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/rezende15.html.

Rubin, D., Aldering, G., Barbary, K., Boone, K., Chappell, G., Currie, M., Deustua, S., Fagrelius, P., Fruchter, A., Hayden, B., Lidman, C., Nordin, J., Perlmutter, S., Saunders, C., Sofiatti, C., and Supernova Cosmology Project, T. UNITY: Confronting Supernova Cosmology's Statistical and Systematic Uncertainties in a Unified Bayesian Framework. *Astrophys. J.*, 813(2):137, November 2015. doi: 10.1088/0004-637X/813/2/137.

Rubin, D., Aldering, G., Betoule, M., Fruchter, A., Huang, X., Kim, A. G., Lidman, C., Linder, E., Perlmutter, S., Ruiz-Lapuente, P., and Suzuki, N. Union Through UNITY: Cosmology with 2,000 SNe Using a Unified Bayesian Framework. *arXiv e-prints*, art. arXiv:2311.12098, November 2023. doi: 10.48550/arXiv.2311.12098.

Scolnic, D., Brout, D., Carr, A., Riess, A. G., Davis, T. M., Dwomoh, A., Jones, D. O., Ali, N., Charvu, P., Chen, R., Peterson, E. R., Popovic, B., Rose, B. M., Wood, C. M., Brown, P. J., Chambers, K., Coulter, D. A., Dettman, K. G., Dimitriadis, G., Filippenko, A. V., Foley, R. J., Jha, S. W., Kilpatrick, C. D., Kirshner, R. P., Pan, Y.-C., Rest, A., Rojas-Bravo, C., Siebert, M. R., Stahl, B. E., and Zheng, W. The Pantheon+ Analysis: The Full Data Set and Light-curve Release. *Astrophys. J.*, 938(2):113, October 2022. doi: 10.3847/1538-4357/ac8b7a.

Shariff, H., Jiao, X., Trotta, R., and van Dyk, D. A. BAHAMAS: New Analysis of Type Ia Supernovae Reveals Inconsistencies with Standard Cosmology. *Astrophys. J.*, 827(1):1, August 2016. doi: 10.3847/0004-637X/827/1/1.

Sullivan, M., Conley, A., Howell, D. A., Neill, J. D., Astier, P., Balland, C., Basa, S., Carlberg, R. G., Fouchez, D., Guy, J., Hardin, D., Hook, I. M., Pain, R., Palanque-Delabrouille, N., Perrett, K. M., Pritchet, C. J., Regnault, N., Rich, J., Ruhlmann-Kleider, V., Baumont, S., Hsiao, E., Kronborg, T., Lidman, C., Perlmutter, S., and Walker, E. S. The dependence of Type Ia Supernovae luminosities on their host galaxies. *Mon. Not. R. Astron. Soc.*, 406(2): 782–802, August 2010. doi: 10.1111/j.1365-2966.2010.16731.x.

Tabak, E. G. and Turner, C. V. A family of nonparametric density estimation algorithms. *Comm. Pure Applied Math.*, 66(2):145–164, 2013.

Tabak, E. G. and Vanden-Eijnden, E. Density estimation by dual ascent of the log-likelihood. *Comm. Math. Sci.*, 8(1): 217–233, 2010.

Tripp, R. A two-parameter luminosity correction for Type IA supernovae. *Astron. Astrophys.*, 331:815–820, March 1998.

Supplementary Material: Accounting for Selection Effects in Supernova Cosmology with Simulation-Based Inference and Hierarchical Bayesian Modelling

A. Analytical Solution to Toy Model

To analytically write the likelihood for the toy model outlined in Section 2.3, we need to solve the integral in the denominator of Equation (10). To begin we use the conditional properties of the multivariate Gaussian to write:

$$N\left(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \middle| \boldsymbol{\Psi}_{\boldsymbol{d}}, \boldsymbol{\Sigma}_{\boldsymbol{d}}\right) = N\left(\hat{m}_s \middle| \mathrm{E}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s], \mathrm{Var}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s]\right) N\left(\hat{c}_s \middle| \mathrm{E}[\hat{c}_s|\, \hat{x}_s], \mathrm{Var}[\hat{c}_s|\, \hat{x}_s]\right) N\left(\hat{x}_s \middle| \mathrm{E}[\hat{x}_s], \mathrm{Var}[\hat{x}_s]\right) \tag{13}$$

where

$$\begin{aligned}
\mathrm{E}[\hat{c}_s|\, \hat{x}_s] &= \mathrm{E}[\hat{c}_s] + \frac{\mathrm{Cov}[\hat{c}_s, \hat{x}_s]}{\mathrm{Var}[\hat{x}_s]}\left(\hat{x}_s - \mathrm{E}[\hat{x}_s]\right) \\
\mathrm{Var}[\hat{c}_s|\, \hat{x}_s] &= \mathrm{Var}[\hat{c}_s] - \frac{\mathrm{Cov}[\hat{c}, \hat{x}]^2}{\mathrm{Var}[\hat{x}_s]} \\
\mathrm{E}[\hat{m}_s|\, \hat{x}_s] &= \mathrm{E}[\hat{m}_s] + \frac{\mathrm{Cov}[\hat{m}_s, \hat{x}_s]}{\mathrm{Var}[\hat{x}_s]}\left(\hat{x}_s - \mathrm{E}[\hat{x}_s]\right) \\
\mathrm{Var}[\hat{m}_s|\, \hat{x}_s] &= \mathrm{Var}[\hat{m}_s] - \frac{\mathrm{Cov}[\hat{m}, \hat{x}]^2}{\mathrm{Var}[\hat{x}_s]} \\
\mathrm{Cov}[\hat{m}_s, \hat{c}_s|\, \hat{x}_s] &= \mathrm{Cov}[\hat{m}_s, \hat{c}_s] - \frac{\mathrm{Cov}[\hat{m}_s, \hat{x}_s]\mathrm{Cov}[\hat{c}_s, \hat{x}_s]}{\mathrm{Var}[\hat{x}]} \\
\mathrm{E}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s] &= \mathrm{E}[\hat{m}_s|\, \hat{x}_s] + \frac{\mathrm{Cov}[\hat{m}_s, \hat{c}_s|\, \hat{x}_s]}{\mathrm{Var}[\hat{c}_s|\, \hat{x}_s]}\left(\hat{c}_s - \mathrm{E}[\hat{c}_s|\, \hat{x}_s]\right) \\
\mathrm{Var}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s] &= \mathrm{Var}[\hat{m}_s|\, \hat{x}_s] - \frac{\mathrm{Cov}[\hat{m}_s, \hat{c}_s|\, \hat{x}_s]^2}{\mathrm{Var}[\hat{c}_s|\, \hat{x}_s]}
\end{aligned}$$

Marginal expectations and covariances are defined in Section 2.3.

We can then solve the integral using Equation (11) by integrating each dimension one at a time with the selection function:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\hat{m}_s + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma_{\text{cut}}}\right) N\left(\hat{\boldsymbol{d}}_{\boldsymbol{s}} \middle| \boldsymbol{\Psi}_{\boldsymbol{d}}, \boldsymbol{\Sigma}_{\boldsymbol{d}}\right) \mathrm{d}\hat{d}_s \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\Phi\left(\frac{m_{\text{cut}} - (\hat{m}_s + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma_{\text{cut}}}\right) \times \right. \\
&\quad \left. N\left(\hat{m}_s \middle| \mathrm{E}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s], \mathrm{Var}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s]\right) N\left(\hat{c}_s \middle| \mathrm{E}[\hat{c}_s|\, \hat{x}_s], \mathrm{Var}[\hat{c}_s|\, \hat{x}_s]\right) N\left(\hat{x}_s \middle| \mathrm{E}[\hat{x}_s], \mathrm{Var}[\hat{x}_s]\right)\right] \mathrm{d}\hat{m}_s\, \mathrm{d}\hat{c}_s\, \mathrm{d}\hat{x}_s
\end{aligned} \tag{14}$$

The first dimension $\hat{m}_s$ is integrated to get:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\hat{m}_s + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma_{\text{cut}}}\right) N\left(\hat{m}_s \middle| \mathrm{E}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s], \mathrm{Var}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s]\right) \mathrm{d}\hat{m}_s \\
&\qquad = \Phi\left(\frac{m_{\text{cut}} - (\mathrm{E}[\hat{m}_s|\, \hat{c}_s, \hat{x}_s] + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma'_{\text{cut}}}\right)
\end{aligned} \tag{15}$$

where

$$\sigma'_{cut} = \sigma_{cut}\sqrt{1 + \frac{\text{Var}[\hat{m}_s|\,\hat{c}_s, \hat{x}_s]}{\sigma^2_{cut}}}$$

The second dimension $\hat{c}_s$ is then integrated out to get:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s|\,\hat{c}_s, \hat{x}_s] + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma'_{\text{cut}}}\right) N\left(\hat{c}_s \middle|\, \text{E}[\hat{c}_s|\,\hat{x}_s], \text{Var}[\hat{c}_s|\,\hat{x}_s]\right) \mathrm{d}\hat{c}_s \\
&\quad = \int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s|\,\hat{x}_s] + \gamma_c(\hat{c}_s - \text{E}[\hat{c}_s|\,\hat{x}_s]) + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\hat{c}_s)}{\sigma'_{\text{cut}}}\right) N\left(\hat{c}_s \middle|\, \text{E}[\hat{c}_s|\,\hat{x}_s], \text{Var}[\hat{c}_s|\,\hat{x}_s]\right) \mathrm{d}\hat{c}_s \\
&\quad = \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s|\,\hat{x}_s] - \gamma_c\text{E}[\hat{c}_s|\,\hat{x}_s] + a_{\text{cut}}\hat{x}_s + (b_{\text{cut}} + \gamma_c)\text{E}[\hat{c}_s|\,\hat{x}_s])}{\sigma''_{\text{cut}}}\right) \\
&\quad = \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s|\,\hat{x}_s] + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\text{E}[\hat{c}_s|\,\hat{x}_s])}{\sigma''_{\text{cut}}}\right) \qquad (16)
\end{aligned}$$

where

$$\begin{aligned}
\sigma''_{cut} &= \sigma'_{cut}\sqrt{1 + \frac{(b_{\text{cut}} + \gamma_c)^2\text{Var}[\hat{c}_s|\,\hat{x}_s]}{{\sigma'_{cut}}^2}} \\
\gamma_c &= \frac{\text{Cov}[\hat{m}_s, \hat{c}_s|\,\hat{x}_s]}{\text{Var}[\hat{c}_s|\,\hat{x}_s]}
\end{aligned}$$

Then the third dimension $\hat{x}_s$ is integrated to get the final result:

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s|\,\hat{x}_s] + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}\text{E}[\hat{c}_s|\,\hat{x}_s])}{\sigma''_{\text{cut}}}\right) N\left(\hat{x}_s \middle|\, \text{E}[\hat{x}_s], \text{Var}[\hat{x}_s]\right) \mathrm{d}\hat{x}_s \\
&= \int_{-\infty}^{\infty} \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s] + \gamma_x(\hat{x}_s - \text{E}[\hat{x}_s]) + a_{\text{cut}}\hat{x}_s + b_{\text{cut}}(\text{E}[\hat{c}_s] + \delta_x(\hat{x}_s - \text{E}[\hat{x}_s]))}{\sigma''_{\text{cut}}}\right) N\left(\hat{x}_s \middle|\, \text{E}[\hat{x}_s], \text{Var}[\hat{x}_s]\right) \mathrm{d}\hat{x}_s \\
&\quad = \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s] - \gamma_x\text{E}[\hat{x}] + (a_{\text{cut}} + \gamma_x + \delta_x b_{\text{cut}})\text{E}[\hat{x}_s] + b_{\text{cut}}(\text{E}[\hat{c}_s] - \delta_x E[\hat{x}])}{\sigma'''_{\text{cut}}}\right) \\
&\quad = \Phi\left(\frac{m_{\text{cut}} - (\text{E}[\hat{m}_s] + a_{\text{cut}}\text{E}[\hat{x}_s] + b_{\text{cut}}\text{E}[\hat{c}_s])}{\sigma'''_{\text{cut}}}\right) \qquad (17)
\end{aligned}$$

where

$$\begin{aligned}
\sigma'''_{cut} &= \sigma''_{cut}\sqrt{1 + \frac{(a_{\text{cut}} + \gamma_x + \delta_x b_{\text{cut}})^2\text{Var}[\hat{x}_s]}{{\sigma''_{cut}}^2}} \\
\gamma_x &= \frac{\text{Cov}[\hat{m}_s, \hat{x}_s]}{\text{Var}[\hat{x}_s]} \\
\delta_x &= \frac{\text{Cov}[\hat{c}_s, \hat{x}_s]}{\text{Var}[\hat{x}_s]}
\end{aligned}$$