Bayesian Hierarchical Models for parameter inference with missing data: Supernova cosmology case study.

Marisa Cristina March, University of Pennsylvania
New Perspectives, Fermilab 18\textsuperscript{th}-19\textsuperscript{th} June 2018
Mission: To understand the nature of dark energy

- Using the Dark Energy Camera to search for Supernovae Ia
- Using Bayes Theory to do the statistical analysis in order to understand the nature of dark energy.
- Specific challenge addressed in this talk:
  - How to deal with missing data (magnitude limited survey) in a Bayesian way, in order to:
    - Use Supernovae Ia to do Bayesian Model Selection
    - Understand and reduce systematics
Physics concept: Using standard candles to measure dark energy

If you have objects of a standard brightness, you can work out how far away they are based on how bright they appear to be.

Define the ‘observed’ distance modulus, to be the difference between the apparent (observed) and absolute magnitudes (brightness) of your standard object:

\[ \mu_{\text{observed}} = m_B - M_0 \]

The theoretical distance modulus depends on the redshift and the cosmological parameters:

\[ \mu_{\text{theory}} = f\{z, \Omega_m, \Omega_k, \Omega_\Lambda, w(z)\} \]

Recipe:
1. Measure the apparent magnitude.
2. Measure the redshift.
3. Work out what values the cosmological parameters must be to get:

\[ \mu_{\text{theory}} = \mu_{\text{observed}} \]
Evidence for Cosmic Acceleration

Distance modulus

Redshift

Using Supernovae Type Ia as Standard Candles

Use the stretch and color of the SNe light curves to apply small corrections (i.e. to standardize) their brightness.

\[
\mu_{\text{observed}} = m_B - M_0 + \alpha x_1 - \beta c
\]

SNe Ia thermonuclear explosions come from white dwarf binary mass transfer.

\[
m_B = M_{0} + \alpha x_1 - \beta c
\]
Data: Supernova Light Curves

DES13C3hwb
Host specz = 0.60647

Plot credit: Chris D'Andrea
Prelim DES Results! Flat $w$CDM

$w = -1.002 \pm 0.057$

$\sigma_w = 0.041(STAT), 0.040(SYS)$

The beginning of an era dominated by systematic uncertainties

$\Omega_M = 0.314 \pm 0.017$

Slide & Plot Credit: Thanks to Dillon Brout!
Beyond the preliminary results:

• Systema Ācs?

• Model Selection on?

Use Bayes Theory!
Supernova Bayesian Hierarchical Model

Allows use of Supernova data for Bayesian Model Selection.

- **Which model best explains dark energy?** LCDM, Modified Gravity? Scalar Field? Chameleon Field?
- Uses latent or hidden variables and priors to model observational data.

Marisa Cris) na March
Truncated data sets and Malmquist bias in SN cosmology

- Problem is that supernova data sets are **incomplete** in magnitude space. Limit of magnitude is set by instrument and environmental conditions.
- One solution is to **discard data** below a magnitude threshold. Disadvantage is **loss of information**.
- Another solution is to **simulate** surveys and “correct” mB data points to recover correct cosmology. Disadvantage is that this cannot be used for Bayesian model selection.
- **Alternative way:** Bayesian Hierarchical Model.
Analytic solution for Malmquist bias (missing data) in Supernova Bayesian Hierarchical Model

\[ p(\theta, \alpha, \beta | x_{1, \text{obs}}, c_{\text{obs}}, m_{B, \text{obs}}, z_{\text{obs}}, m_{B, \text{thresh}} | I, M) \]

\[ \propto \int_{N_{\text{obs}}}^{\inf} dN \int dR_x dx_\star \frac{1}{N} \left( \frac{N}{N_{\text{obs}}} \right)^{N_{\text{obs}}} \prod_{i} \left| 2\pi \Sigma_{v, i} \right|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\tilde{w}_i - q_i)^T \Sigma_{v, i} (\tilde{w}_i - q_i) \right) \]

\[ \times \prod_{i} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{m_{B, \text{thresh}}}^{2^{\text{max}}} \int_{x_{1, i}}^{\tilde{x}_{1, i}} \int_{\tilde{c}_i}^{c_{\text{mis}}} \int_{\tilde{m}_{B,i}}^{m_{\text{mis}}} \int_{\tilde{z}_{i}}^{z_{\text{mis}}} \left| 2\pi \Sigma_{v, i} \right|^{-\frac{1}{2}} \]

\[ \times \exp \left( -\frac{1}{2} (\tilde{w}_i - q_i)^T \Sigma_{v, i} (\tilde{w}_i - q_i) \right) \]

\[ \times p(R_x, x_\star | I, M) p(\theta, \alpha, \beta | I, M) \]

\[ x_\star = [x_{1, \text{star}}, c_\star] \in \mathbb{R}^2 \]

\[ \tilde{w}_i = \begin{bmatrix} \tilde{m}_{B,i} \\ \tilde{x}_{1,i} \\ \tilde{c}_i \end{bmatrix} \in \mathbb{R}^3 \]

\[ \Sigma_{c,i} = \begin{bmatrix} \sigma_{m_1}^2 & \sigma_{m_1,x_1} & \sigma_{m_1,c} \\ \sigma_{m_1,x_1} & \sigma_{x_1}^2 & \sigma_{x_1,c} \\ \sigma_{m_1,c} & \sigma_{x_1,c} & \sigma_{c_1}^2 \end{bmatrix} \in \mathbb{R}^{3 \times 3} \]

\[ R_x = \begin{bmatrix} R_{x_1} & 0 \\ 0 & R_c \end{bmatrix} \in \mathbb{R}^{2 \times 2} \]

Posterior probability of parameters in a truncated data set

arXiv:1804.02474
Parameter estimation with simulated truncated supernovae data sets.

- Constant, $M_0$
- Slope (SALT-II) Parameters $\alpha$, $\beta$
- Intrinsic dispersion
- Standard deviations and means of latent $x_1$ and $c$
  independent variables
- Matter density
- Curvature density
- Dark Energy density
- $H_0$
Summary:

- If you want to do Bayesian **Model selection**, you need to have the correct Bayesian **Posterior**.
- How do you account for **missing data** in a Bayesian way?
- See arXive: **1804.02474**

Done:

- Analy(c) solution for missing data. Tested on basic simulations.

Next Steps:

- Include refined selection function, test on SNANA DES like simulations.
- Account for uncertainty in **typing**.