Hierarchical Bayesian modeling under covariate shift in supernova cosmology

Presenter: Max Autenrieth
Supervisor: David A. van Dyk
Co-Supervisor: Roberto Trotta, David Stenning

August 3, 2022

Overview

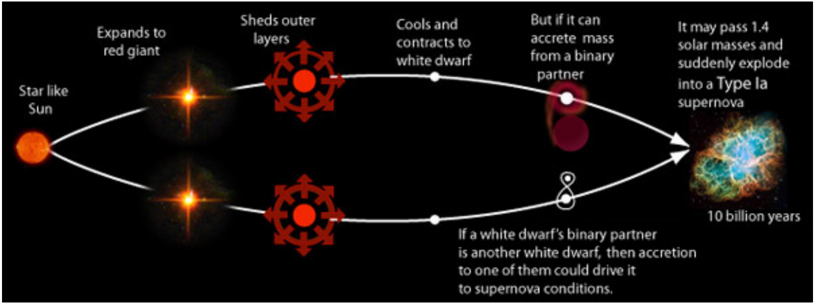


Image Credit: http://hyperphysics.phy-astr.gsu.edu/hbase/astro/snovcn.html

- Supernova type Ia (SNIa) have common "flashpoint" (**standard candles**)
- SNIa allow cosmological parameter estimation (e.g. **dark energy density**).

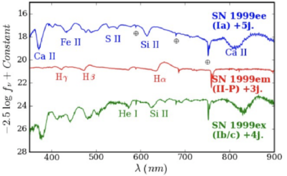
Statistical Challenges:

(i) Reliable classification of SNIa given a non-representative training set.

(ii) Secondary analysis needs to account for contamination arising through (i).

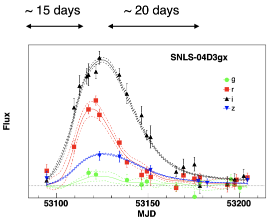
Covariate shift in supernova cosmology

- Confirming Ia is easy with spectra

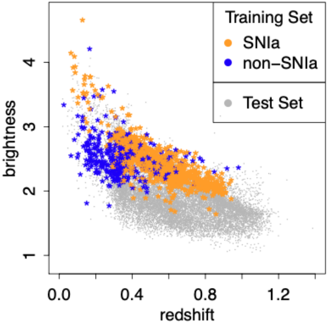


Credit: Julien Guy

- Much harder with just photometry



Guy et al (2007)



Revsbech et al. (2018)

- Probabilistic classification of photometric light curve data
- Confirmed (training) set non-representative

Definitions and Notation:

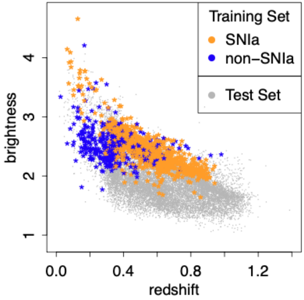
- Feature space $\mathcal{X} \subset \mathbb{R}^n$, and label space $\mathcal{Y}$.
- Different domains: $p(x, y)$ differs in source and target.
- **Labeled source** (training) data $D_S = \{(x_S^{(i)}, y_S^{(i)})\}_{i=1}^{n_s}$.
- **Unlabeled target** data $D_T = \{x_T^{(i)}\}_{i=1}^{n_t}$.

Definition 1.1 (Moreno-Torres et al. (2012))

Covariate shift is defined as $p_S(y|x) = p_T(y|x)$ but $p_S(x) \neq p_T(x)$.

Objective: Accurately predicting target labels y_T, by minimizing target risk

$$\mathcal{R}_T(f) := \mathbb{E}_{(x,y)\sim p_T(x,y)}[\ell(f(x), y)]. \quad (1)$$



Methodology – Stratified Learning (StratLearn):

- We define the propensity score (PS) as:

$$e(x_i) := P(s_i = 1 | x_S, x_T), \text{ with } 0 < e(x_i) < 1. \quad (2)$$

- PS well-established in causal inference (Rosenbaum and Rubin 1983).

Proposition 1 (Learning conditional on the propensity score)

Under covariate shift conditions, conditional on the propensity score:

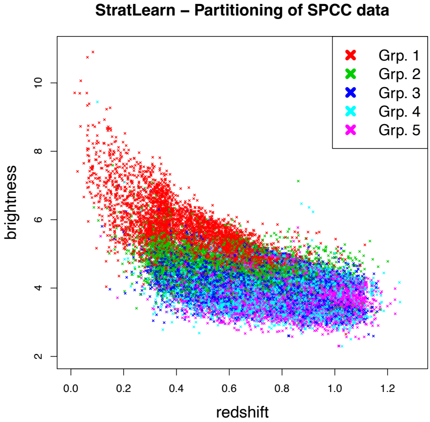
$$p_T(x, y | e(x)) = p_S(x, y | e(x)), \quad (3)$$

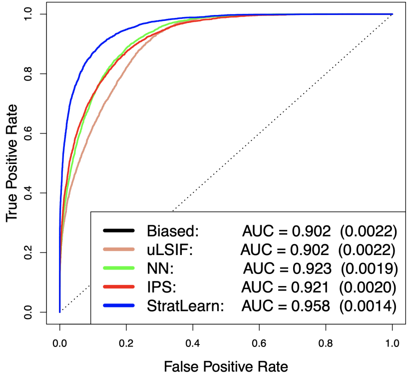
eliminating covariate shift. Thus, for any loss function $\ell = \ell(f(x), y)$*,*

$$E_{(x,y)\sim p_T(x,y|e(x))}[\ell(f(x), y)] = E_{(x,y)\sim p_S(x,y|e(x))}[\ell(f(x), y)]. \quad (4)$$

StratLearn for SN Ia classification

- Stratify source and target data on propensity score.
- Classify separately within strata, via Random Forest.





- StratLearn balances covariates (and outcome) within strata.
- Performance close to unbiased “Gold Standard” (AUC: 0.965).

Bayesian Modelling for Supernova Type Ia Cosmology

- **Scientifically justified** model **for SNIa** available.
- Account for measurement errors via **Bayesian hierarchical model**.

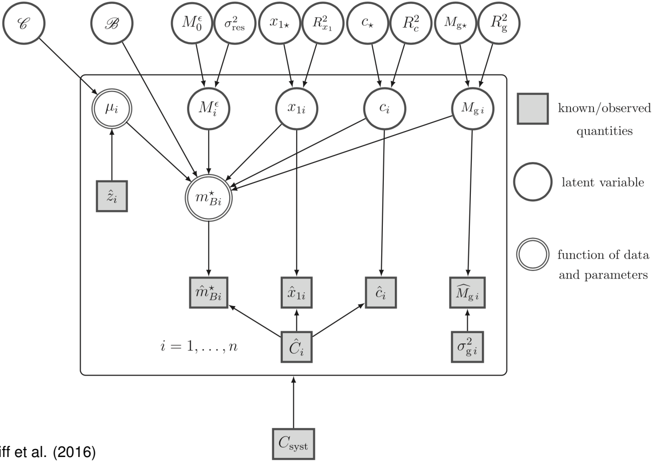
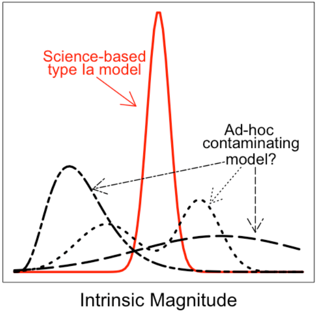


Image Credit: Shariff et al. (2016)

Supernova Cosmology with Contaminated Data

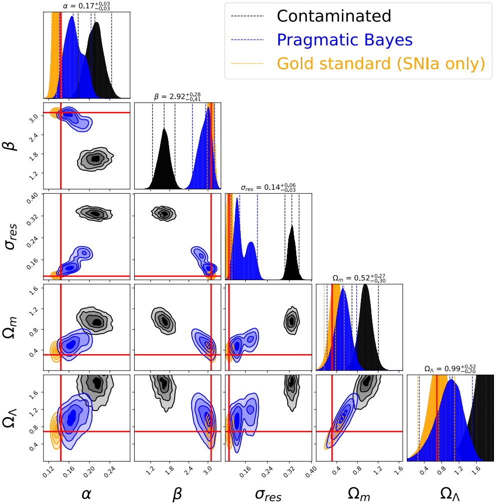
Secondary analysis: needs to account for classification uncertainties!



- Fully Bayes: with target distribution $p(\theta \mid y, g(D))\, p(y \mid f(D), g(D))$
- Requires model specification for contaminants (unknown).
- Parts of data used twice for type probabilities.
- We avoid this using **Pragmatic Bayes!**

- **Pragmatic** target distribution: $p(\theta \mid y = \text{SNIa}, g(D))\, p(y = \text{SNIa} \mid f(D))$
- **Assumption:** Contaminants non-informative for parameters of interest θ.

SN Cosmology with Contamination: A pragmatic Bayesian Approach



(Results for 500 simulated SNe (SNcosmo) with 5% contamination.)

Iteratively:

(i) Resample y, with $p(y = \text{SNIa}|f(D))$

(ii) Fit SNIa model to obtain posterior sample (e.g. via Multinest).

Ongoing and Future Work

Cosmological parameter estimation (secondary analysis):

- Comparison to fully Bayes (with/without model misspecification).
- Incorporation of (photometric) redshift uncertainties.
- Consideration of selection effects (photometry not representative).

StratLearn:

- Scientific applications (e.g. redshift calibration for weak lensing).
- Balance diagnostics via predicted marginal outcome distributions.

References I

Autenrieth, M., Van Dyk, D. A., Trotta, R., and Stenning, D. C. (2021). Stratified learning: a general-purpose statistical method for improved learning under covariate shift. *arXiv preprint arXiv:2106.11211*.

Bickel, S., Brückner, M., and Scheffer, T. (2009). Discriminative learning under covariate shift. *Journal of Machine Learning Research*, 10(Sep):2137–2155.

Kunz, M., Bassett, B. A., and Hlozek, R. A. (2007). Bayesian estimation applied to multiple species. *Physical Review D*, 75(10):103508.

Moreno-Torres, J. G., Raeder, T., Alaiz-RodríGuez, R., Chawla, N. V., and Herrera, F. (2012). A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530.

Revsbech, E. A., Trotta, R., and van Dyk, D. A. (2018). Staccato: a novel solution to supernova photometric classification with biased training sets. *Monthly Notices of the Royal Astronomical Society*, 473(3):3969–3986.

References II

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Shariff, H., Jiao, X., Trotta, R., and van Dyk, D. A. (2016). Bahamas: new analysis of type ia supernovae reveals inconsistencies with standard cosmology. *The Astrophysical Journal*, 827(1):1.

Shimodaira, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244.

Zadrozny, B. (2004). Learning and evaluating classifiers under sample selection bias. In *Proceedings of the twenty-first international conference on Machine learning*, page 114. ACM.

Thank you very much for your time!

Future work – Balance diagnostics via predicted outcome

Remark 1 (Outcome balance:)

In covariate shift framework

- *Potential outcomes are identical ($Y_0 \equiv Y_1$), no "treatment effect"*
- *Only source data is observed ($Y_1 \equiv Y$)*
- *Given $e(x)$, with $0 < e(x) < 1$, and covariate shift conditions, source data assignment is 'strongly ignorable'*
- *Then, conditional on PS, source and target outcome are the same in expectation [invoking Rosenbaum and Rubin (1983), Theorem 4].*

- Target labels y_T in practice not observed, only source labels y_S
- Target and source label predictions ($\hat{y}_T$ and $\hat{y}_S$) given.
- Can we use predicted outcomes to test for remaining confounding?

Future work – Balance diagnostics via predicted outcome

Table: **StratLearn** strata composition (including all 102 covariates in PS).

Stratum	Set	Number of SNe	Number of SNIa	Prop. of SNIa
1	Source	958	518	**0.54**
	Target	3306	1790	**0.54**
2	Source	120	28	**0.23**
	Target	4144	927	**0.22**
3	Source	13	4	0.31
	Target	4250	540	0.13
4	Source	7	4	0.57
	Target	4257	610	0.14
5	Source	4	4	1
	Target	4259	662	0.16

- Outcome proportions balanced within stratum 1 and 2.

Table: Composition with two covariates (redshift and brightness) in PS.

Stratum	Set	Number of SNe	Number of SNIa	Prop. of SNIa
1	Source	947	652	**0.69**
	Target	2519	1242	**0.49**
2	Source	245	181	**0.74**
	Target	3221	1147	**0.36**
3	Source	17	12	0.71
	Target	3449	754	0.22
4	Source	6	6	1
	Target	3460	342	0.10
5	Source	2	0	0
	Target	3464	107	0.03

- Imbalance due to remaining confounding.

Previous methods – Importance weighting:

Under **covariate shift** conditions:

Proposition 2 (Shimodaira (2000), Bickel et al. (2009))

If the support of $p_T(x)$ is contained in $p_S(x)$, then

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_T}\left[\ell(f(x),y)\right] = \mathbb{E}_{(x,y)\sim\mathcal{D}_S}\left[\frac{p_T(x)}{p_S(x)}\ell(f(x),y)\right]. \quad (5)$$

Proposition 3 (Bias Correction (Zadrozny 2004))

Let (x, y, s) be examples drawn from a distribution $\mathcal{D}$, with feature-label-selection space $\mathcal{X} \times \mathcal{Y} \times \mathcal{S}$. Then,

$$\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\ell(f(x),y)\right] = \mathbb{E}_{(x,y)\sim\hat{\mathcal{D}}}\left[\ell(f(x),y)|s=1\right], \quad (6)$$

$$\text{with } \hat{\mathcal{D}}(x,y,s) := \frac{P(s=1)}{P(s=1|x)}\mathcal{D}(x,y,s). \quad (7)$$

The target risk $\hat{R}_T(\hat{f})$ is computed as

$$\hat{R}_T(\hat{f}) = \frac{1}{n_T}\sum_{k=1}^{n_T}\int \hat{f}^2(z|x_T^{(k)})dz - 2\frac{1}{n_T}\sum_{k=1}^{n_T}\hat{f}(z_T^{(k)}|x_T^{(k)}), \quad (8)$$

where z_T is the true target redshift, used for evaluation purposes only.

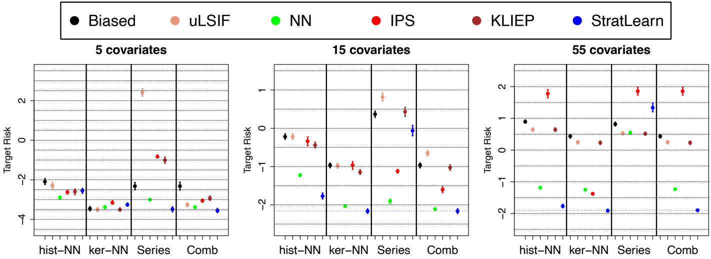


Figure: Target risk ($\hat{R}_T$) of photometric redshift estimation.