Improved Weak Lensing Photometric Redshift Calibration via StratLearn and Hierarchical Modeling

Presenter: Maximilian Autenrieth
Co-Authors: A.H. Wright, R. Trotta, D. A. van Dyk,
D. C. Stenning, B. Joachimi

September 9, 2024

Overview – Cosmic Shear Tomography

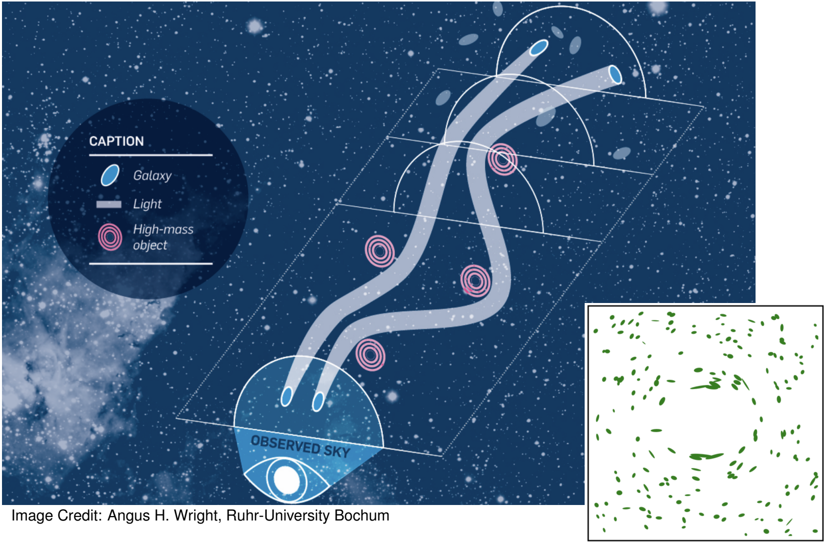
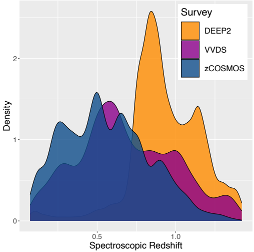


Image Credit: Angus H. Wright, Ruhr-University Bochum

Data and Covariate Shift:

Spectroscopic Data (labelled/source)**:**

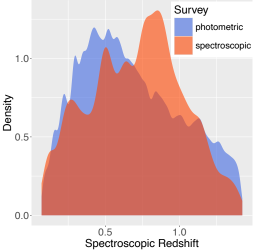
- $\sim$ 21.500 galaxies from three surveys
- x_i vector of 9 photometric magnitudes
- Accurate measurement of redshift z_i
- 100 independent lines-of-sight



Photometric Data (unlabelled/target)**:**

- $\sim$ 12.48 million galaxies, with photometric magnitudes x_i, but *no* redshift z_i.

(Realistic simulations (Wright et al. 2020) mimicking KiDS+VIKING-450.)



Covariate shift:

- We assume $p_S(z|x) = p_T(z|x)$ but $p_S(x) \neq p_T(x)$.
- It generally follows that $p_S(z) \neq p_T(z)$.

The 'photo-z' Estimation Problem:

Conditional density estimation models:

- Fit non-parametric models (hist-NN, ker-NN, Series; Izbicki et al. 2017)
- Comb (combination model):

$$\hat{f}^{\alpha}(z|x) = \sum_{k=1}^{p} \alpha_k \hat{f}_k(z|x), \text{ with constraints (i): } \alpha_i \geq 0, \text{ and (ii): } \sum_{k=1}^{p} \alpha_k = 1, \tag{1}$$

***Generalized* risk** optimization (Izbicki et al. 2017) w.r.t:

$$\hat{R}_S(\hat{f}) = \frac{1}{n_S} \sum_{k=1}^{n_S} \int \hat{f}^2(z|x_S^{(k)})dz - 2\frac{1}{n_S} \sum_{k=1}^{n_S} \hat{f}(z_S^{(k)}|x_S^{(k)}). \tag{2}$$

Issue: Under covariate shift, with $p_S(z,x) \neq p_T(z,x)$, we have $\hat{R}_S(\hat{f}) \neq R_T(\hat{f})$.

Does a New Drug Improve Health Outcomes?

Causal Inference:

- Split subjects: treatment ($A = 1$) and control ($A = 0$) group.
- What if treatment group differs systematically from control group, e.g., in terms of x.

$$p_{\text{treatment}}(x) \overset{?}{=} p_{\text{control}}(x)$$

- Randomization is the gold standard, not always possible.

Propensity Scores:

- Rosenbaum and Rubin (1983) define propensity scores:

$$e(x) = P(A = 1|x).$$

- Demonstrate that $e(x)$ is a balancing score:

$$p_{\text{treatment}}(x|e(x)) = p_{\text{control}}(x|e(x)).$$

StratLearn – Principled Learning under Covariate Shift:

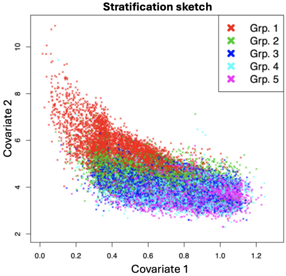
- Estimate the **propensity score**

$$\hat{e}(x_i) = P(\text{target set}|\text{covariates}). \tag{3}$$

- Under covariate shift conditions (Autenrieth et al. 2024a):

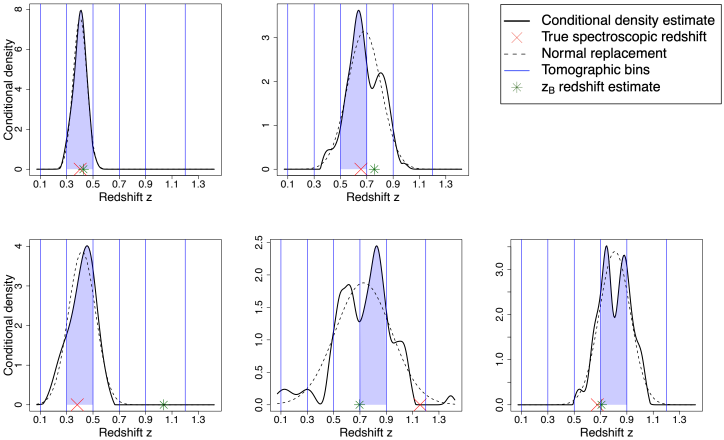
$$p_T(x, z|\hat{e}(x)) \approx p_S(x, z|\hat{e}(x)). \tag{4}$$

- **Stratify** source/target data sets on $\hat{e}(x)$.
- Within strata j, $p_{Tj}(x, z) \approx p_{Sj}(x, z)$, thus $\hat{R}_S(\hat{f}) \approx R_T(\hat{f})$.
- Fit conditional density models within **covariate balanced** strata.



Reduce covariate shift and thus expected classification/prediction error.

Photo-*z* Estimates and Tomographic Binning:

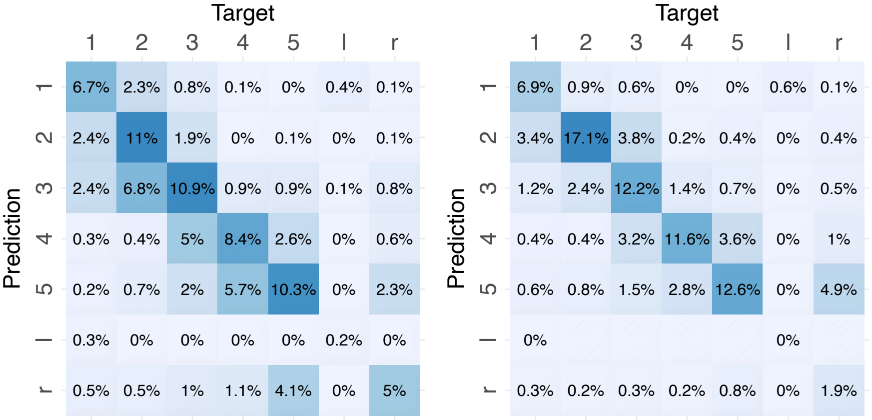


- Tomographic bin assignment:

$$b(i) = \operatorname{argmax}_m \int_{B(m)} \hat{f}(z_i|x_i)dz, \quad m = 1, \ldots, 5, l, r. \tag{5}$$

Improved Bin Assignment Accuracy

- Confusion matrices for z_B (left) and *StratLearn* (right):

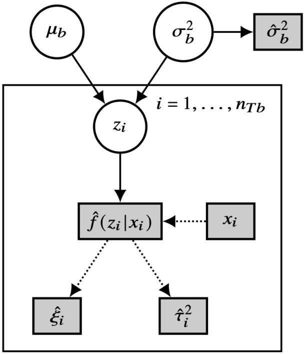


	1	2	3	4	5	l	r
1	6.7%	2.3%	0.8%	0.1%	0%	0.4%	0.1%
2	2.4%	11%	1.9%	0%	0.1%	0%	0.1%
3	2.4%	6.8%	10.9%	0.9%	0.9%	0.1%	0.8%
4	0.3%	0.4%	5%	8.4%	2.6%	0%	0.6%
5	0.2%	0.7%	2%	5.7%	10.3%	0%	2.3%
l	0.3%	0%	0%	0%	0%	0.2%	0%
r	0.5%	0.5%	1%	1.1%	4.1%	0%	5%

	1	2	3	4	5	l	r
1	6.9%	0.9%	0.6%	0%	0%	0.6%	0.1%
2	3.4%	17.1%	3.8%	0.2%	0.4%	0%	0.4%
3	1.2%	2.4%	12.2%	1.4%	0.7%	0%	0.5%
4	0.4%	0.4%	3.2%	11.6%	3.6%	0%	1%
5	0.6%	0.8%	1.5%	2.8%	12.6%	0%	4.9%
l	0%					0%	
r	0.3%	0.2%	0.3%	0.2%	0.8%	0%	1.9%

Overall binning accuracy improved from 0.526 (z_B) to 0.622 (*StratLearn*).

Bayesian Hierarchical Modelling of Conditional Densities



- **Population level** distribution within bin b:

$$z_i|\mu_b, \sigma_b \overset{\text{indep.}}{\sim} N(\mu_b, \sigma_b^2), \tag{6}$$

with redshift population variance σ_b^2.

- **Object level:**

$$z_i|x_i \overset{\text{indep.}}{\sim} N(\hat{\zeta}_i, \hat{\tau}_i^2), \tag{7}$$

Gaussian replacement of conditional density estimates.

- With $\hat{X}_{\mathbf{n_{Tb}}} := \{\hat{\zeta}_i, \hat{\tau}_i\}_{i=1}^{n_{Tb}}$, obtain the (joint) marginal posterior

$$p(\mu_b, \sigma_b|\hat{X}_{\mathbf{n_{Tb}}}) \propto p(\mu_b, \sigma_b) \prod_i N(\hat{\zeta}_i|\mu_b, \hat{\tau}_i^2 + \sigma_b^2). \tag{8}$$

StratLearn-Bayes:

- Conditional posterior distribution of μ_b given σ_b:

$$\mu_b|\sigma_b, \hat{X}_{\mathbf{n_{Tb}}} \sim N(\tilde{\mu}_b, V_{\mu_b}), \tag{9}$$

with

$$\tilde{\mu}_b = \frac{\sum_i \frac{1}{\hat{\tau}_i^2+\sigma_b^2}\hat{\zeta}_i}{\sum_i \frac{1}{\hat{\tau}_i^2+\sigma_b^2}} \quad \text{and } V_{\mu_b}^{-1} = \sum_i \frac{1}{\hat{\tau}_i^2 + \sigma_b^2}, \tag{10}$$

with $V_{\mu_b}^{-1}$ being the total precision.

- Empirical Bayesian approach by setting σ_b to fixed estimate from data (via stacked estimator), i.e., by choosing $p(\sigma_b) = \delta(\sigma_b - \hat{\sigma}_b)$.
- Given σ_b, analytically derive $\tilde{\mu}_b$ as MAP estimate of μ_b.

Stacked Variance Estimator:

- Obtain estimate $\hat{p}_b^{\text{stack}}(z)$ of $p_b(z)$ by averaging (stacking) the conditional densities within bin, i.e.,

$$\hat{p}_b^{\text{stack}}(z) = \frac{1}{n_{Tb}} \sum_j \hat{f}(z_j|x_j). \tag{11}$$

- Estimate for redshift population variance σ_b^2 then obtained as variance of $\hat{p}_b^{\text{stack}}(z)$, i.e.,

$$\hat{\sigma}_b^2 = \frac{1}{\sum_k \hat{p}_{b,m(k)}^{\text{stack}}(z)} \sum_k \hat{p}_{b,m(k)}^{\text{stack}}(z)(m(k) - \hat{\mu}_b^{\text{stack}})^2. \tag{12}$$

Results – Evaluation Metrics:

- Mean discrepancy (bias):

$$\mathbb{E}[\hat{\mu}_b - \mu_b^{\text{true}}] \simeq \frac{1}{L}\sum_{l=1}^{L}(\hat{\mu}_{b,l} - \mu_{b,l}^{\text{true}}), \quad (= \widehat{\text{bias}}_b) \tag{13}$$

  where $L = 100$ is the number of LoS.

- Standard deviation (SD) of the mean differences across the 100 LoS:

$$\text{SD}(\hat{\mu}_b - \mu_b^{\text{true}}) = \sqrt{\frac{\sum_{l=1}^{L}(\mu_{b,l}^{\text{diff}} - \widehat{\text{bias}}_b)^2}{L-1}}. \tag{14}$$

Main Results:

- **Main objective** – population mean discrepancy **(bias)**:

	Binning	Galaxies [M]	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	|Average|
StratLearn-Bayes (A)	*SL*	12.02	0.0123	-0.0076	0.0053	-0.0010	0.0001	0.0053
StratLearn-Bayes (B)	*SL*	12.02	0.0095	-0.0092	0.0047	-0.0013	0.0012	0.0052
SOM	*SL* (gold)	11.48	-0.0084	0.0022	0.0156	0.0117	0.0148	0.0105
StratLearn-Bayes (A)	z_B	10.90	0.0259	0.0127	0.0084	0.0003	-0.0231	0.0141
StratLearn-Bayes (B)	z_B	10.90	0.0228	0.0117	0.0071	-0.0002	-0.0236	0.0131
SOM	z_B (gold)	10.17	-0.0005	0.0036	0.0135	0.0147	-0.0102	0.0085

- *StratLearn*-Bayes leads to around 40% bias reduction compared to state-of-the-art, SOM with z_B binning.
- z_B binning leads to slightly lower variance.
- *StratLearn*-Bayes approach leads to an increase of galaxies available for science of ∼18%.

Population Distribution Estimates:

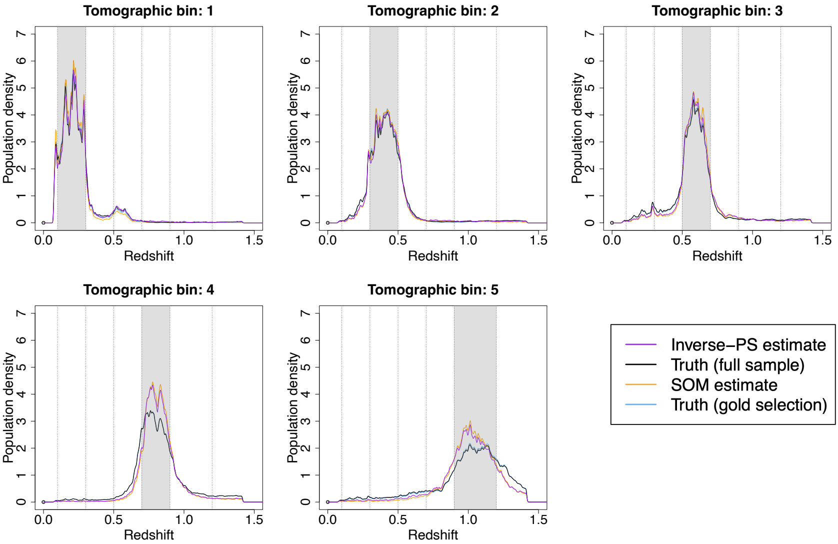
- **Direct redshift calibration:** Obtain an estimate of the binned joint target distribution $p_{Tb}(z, x)$ via

$$p_{Tb}(z, x) = \omega_b(x) p_{Sb}(z, x), \tag{15}$$

  with $p_{Sb}(z, x)$ being the binned joined source distribution of bin b.
- Weight estimation via (inverse) propensity scores:

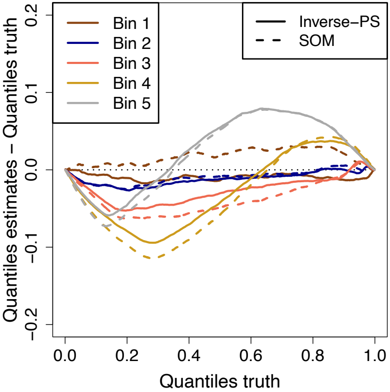
$$\omega_b(x) = \frac{p_{Tb}(x)}{p_{Sb}(x)} = \frac{p(s_b = 1)}{p(s_b = 0)} \frac{p(s_b = 0|x)}{p(s_b = 1|x)} \propto \left(\frac{1}{p(s_b = 1|x)} - 1 \right). \tag{16}$$

Results – Redshift Population Distribution Estimates:



True photometric population distribution shapes reasonably well recovered.

Results – Redshift Population Distribution Estimates:



- Evaluation via probability-probability plots (pp-plots) with respective underlying true distributions.
- Based on new binning, Inverse-PS closer to underlying truth than SOM (without need of quality cuts).

Conclusion:

Summary:

- Two-stage analysis:
  1. Development of principled methodology for covariate shift correction to obtain accurate object level photo-*z* distributions.
  2. Employ suitable Bayesian tools to achieve desired frequentist properties of population means/distributions (particularly in terms of bias) validated by comprehensive/realistic MC studies.
- Bias reduction of $\sim$ 40%, while maintaining $\sim$ 18% more data for scientific analysis.

Ongoing and Future work:

- Applying method for present and upcoming cosmic shear analysis.
- Account for quality cuts in spectroscopic sample (consider potential violations of covariate shift assumption).
- Improvement of *StratLearn* via more general matching approach.

References I

Autenrieth, M., Van Dyk, D. A., Trotta, R., and Stenning, D. C. (2024a). Stratified learning: A general-purpose statistical method for improved learning under covariate shift. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 17(1):e11643.

Autenrieth, M., Wright, A. H., Trotta, R., van Dyk, D. A., Stenning, D. C., and Joachimi, B. (2024b). Improved weak lensing photometric redshift calibration via stratlearn and hierarchical modeling. *arXiv preprint arXiv:2401.04687*.

Izbicki, R., Lee, A. B., Freeman, P. E., et al. (2017). Photo-z estimation: An example of nonparametric conditional density estimation under selection bias. *The Annals of Applied Statistics*, 11(2):698–724.

Kanamori, T., Hido, S., and Sugiyama, M. (2009). A least-squares approach to direct importance estimation. *Journal of Machine Learning Research*, 10(Jul):1391–1445.

References II

Kremer, J., Gieseke, F., Pedersen, K. S., and Igel, C. (2015). Nearest neighbor density ratio estimation for large-scale applications in astronomy. *Astronomy and Computing*, 12:67–72.

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.

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.

Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P. V., and Kawanabe, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in neural information processing systems*, pages 1433–1440.

References III

Wright, A. H., Hildebrandt, H., Van den Busch, J. L., and Heymans, C. (2020). Photometric redshift calibration with self-organising maps. *Astronomy & Astrophysics*, 637:A100.

Thank you very much for your time!

Previous methods – Weighting:

Under **covariate shift** conditions, if the support of $p_T(x)$ is contained in $p_S(x)$, following Shimodaira (2000),

$$\mathbb{E}_{(x,y)\sim\mathcal{D}_T}\left[\ell(f(x),y)\right] = \mathbb{E}_{(x,y)\sim\mathcal{D}_S}\left[\frac{p_T(x)}{p_S(x)}\ell(f(x),y)\right]. \tag{17}$$

Various weighting methods:

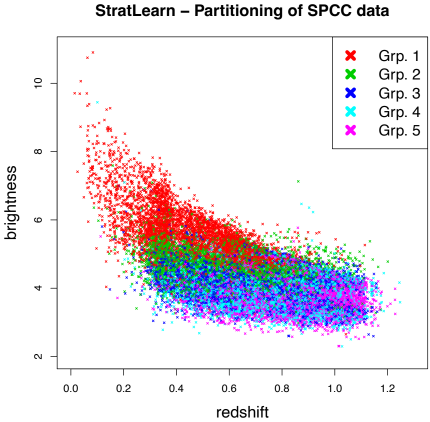
- KLIEP (Sugiyama et al. 2008);
- uLSIF (Kanamori et al. 2009);
- NN – Nearest-Neighbor (Kremer et al. 2015);
- IPS – probabilistic classification (Kanamori et al. 2009);

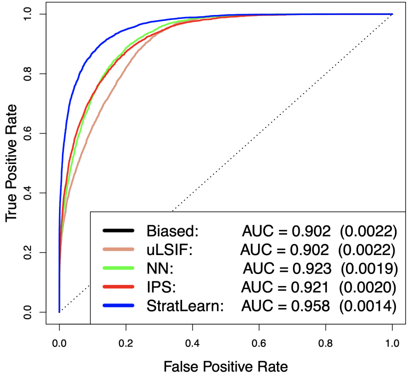
Issue of weighting:

- Estimation of weights is difficult (especially in high dimensions).
- Large (noisy) weights cause **high variance** and unreliable target predictions.

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).