Matrix Constraints and Multi-Task Learning for Covariate Balance

Eli Ben-Michael and Avi Feller

Berkeley
UNIVERSITY OF CALIFORNIA

Summary

- Recent work in weights which directly balance (high dim) covariates
- Many estimators use multiple sets of weights
- We propose finding all weights in one optimization problem
- Measure covariate balance *jointly* across problems
- Dual view as multi-task learning and hierarchical modeling
- Combine joint and separate constraints $\to$ decomposition in dual

Motivation: Estimating Multiple Means

Many estimators use multiple sets of weights

- **Heterogeneous Treatment Effects**
  - CATT for multiple subgroups
  - Estimate individual-level CATE
- **Obs. Study with Missing Outcomes (IPTW + IPMW)**
- **Multilevel Observational Study**

Instead of fitting weights separately, we can exploit natural structure.

Vector Constraints for Covariate Balance

Goal: Estimate ATT with a weighted estimator Choose weights that balance covariates and are close to uniform

$$\min_{\gamma \in \mathbb{R}^{n_0}} \sum_{T_i=0} f(\gamma_i) + h(\bar{X}_1 - X_0'\gamma)$$

General formulation encompasses several estimators, [1, 2, 3, 4, 5]

- **Dispersion Function**: f penalizes large weights
- **Balance Criterion**: h measures covariate balance

Dual representation as *regularized M-estimator* of p-score

$$\min_{\theta \in \mathbb{R}^d} \sum_{T_i=0}^{n} f^*(X_i'\theta) - \bar{X}_1'\theta + h^*(\theta)$$

Where g^* is the *convex conjugate* of g.

- **Odds Function**: $f^{*\prime}(X'\theta) = \frac{\pi(X)}{1-\pi(X)}$
- **Regularization**: h^* penalizes complex models

Example: Entropy Balancing [1] with L^∞ balance criterion [5].

$$\begin{aligned} \min_{\gamma} \quad & \sum_{T_i=0} \gamma_i \log \gamma_i \\ \text{s.t.} \quad & \|\bar{X}_1 - X_0'\gamma\|_\infty \leq \delta \end{aligned}$$

Dual fits logit p-score with LASSO penalty

$$\min_{\theta} \sum_{T_i=0} \exp(X_i'\theta - 1) - \bar{X}_1'\theta + \delta\|\theta\|_1$$

Matrix Constraints and Multi-Task Learning

Primal: Measure dispersion separately, **balance** jointly

$$\min_{\Gamma} \sum_{k=1}^{m} \sum_{i} (1 - T_i) f(\Gamma_{ik}) + h\left(\bar{X}_1 - X_0'\Gamma\right)$$

Dual: Fit p-score models separately, **regularize** jointly

$$\min_{\Theta} \sum_{k=1}^{m} \sum_{i} \left[(1 - T_i) f^*(\Theta_k' X_i) - \frac{1}{n_{1k}} T_i X_i' \Theta_k\right] + h^*(\Theta)$$

Special Cases

- Estimating subgroup ATT
  Primal: Weighted Frobenius *soft* constraint
  $$\min_{\Gamma} \sum_{k=1}^{m} \sum_{i \in G_k, T_i=0} \Gamma_{ik} \log \Gamma_{ik} + \frac{\lambda}{2} \mathrm{tr}((\bar{X}_1 - X_0'\Gamma)\Omega(\bar{X}_1 - X_0'\Gamma))$$
  Dual: Fully-interacted p-score with **Hierarchical Prior**
  $$\min_{\Theta} \sum_{k=1}^{m} \sum_{i \in G_k, T_i=0} \exp(X_i'\Theta_k) - \mathrm{tr}(\bar{X}_1 \Theta_k') + \frac{\lambda}{2} \mathrm{tr}(\Theta \Omega^{-1} \Theta')$$
  $\Omega \in \mathbb{R}^{m \times m}$ corresponds to *prior covariance*:
  $$\Theta_{j1}, \ldots, \Theta_{jm} \sim \text{Normal}(0, \Omega)$$
- Estimating individual-level CATE
  Primal: Spectral norm *hard* constraint
  $$\begin{aligned} \min_{\Gamma} \quad & \sum_{T_k=1} \sum_{T_i=0} \Gamma_{ik} \log \Gamma_{ik} \\ \text{s.t.} \quad & \sup_{\|u\|_2=1} \|(X_1 - X_0'\Gamma)u\|_2 \leq \delta \end{aligned}$$
  Spectral 2-norm is imbalance in worst-case linear combo of units
  Dual: Individual p-score models have **Low Rank**:
  $$\min_{\Theta} \sum_{T_k=1} \sum_{T_i=0} \exp(X_i'\Theta_k - 1) - X_k'\Theta_k + \delta\|\Theta\|_*$$
  Assume that
  - There are $p \ll n_1$ archetypal models $U \in \mathbb{R}^{d \times p}$
  - Each realization is linear combo with weights $V \in \mathbb{R}p \times n_1$, so $\Theta = UV'$

Other possible dual regularization: Group LASSO

Future Work

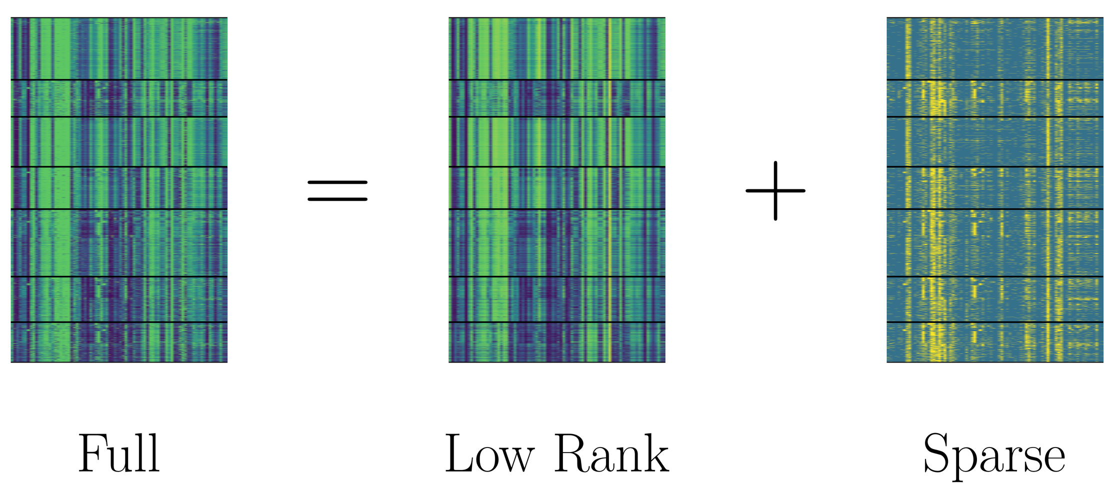
- Weights are fully design-based; combine with outcome information
  - AIPW, balancing Prognostic score, double selection
  - Ensemble methods with low-rank constraints
- For HTE, include penalty for large weights on far-away donors

Combining Matrix and Vector Constraints

Example: Spectral 2-norm and an L^∞ constraint

$$\begin{aligned} \min_{\Gamma} \quad & \sum_{T_k=1} \sum_{T_i=0} \Gamma_{ik} \log \Gamma_{ik} \\ \text{s.t.} \quad & \max_{jk} |X_{1jk} - X_{0\cdot j}'\Gamma_{\cdot k}| \leq \delta_1 \\ & \sup_{\|u\|_2=1} \|(X_1 - X_0'\Gamma)u\|_2 \leq \delta_2 \end{aligned}$$

Results in robust PCA decomposition, used in image processing [6]



Simulation Study

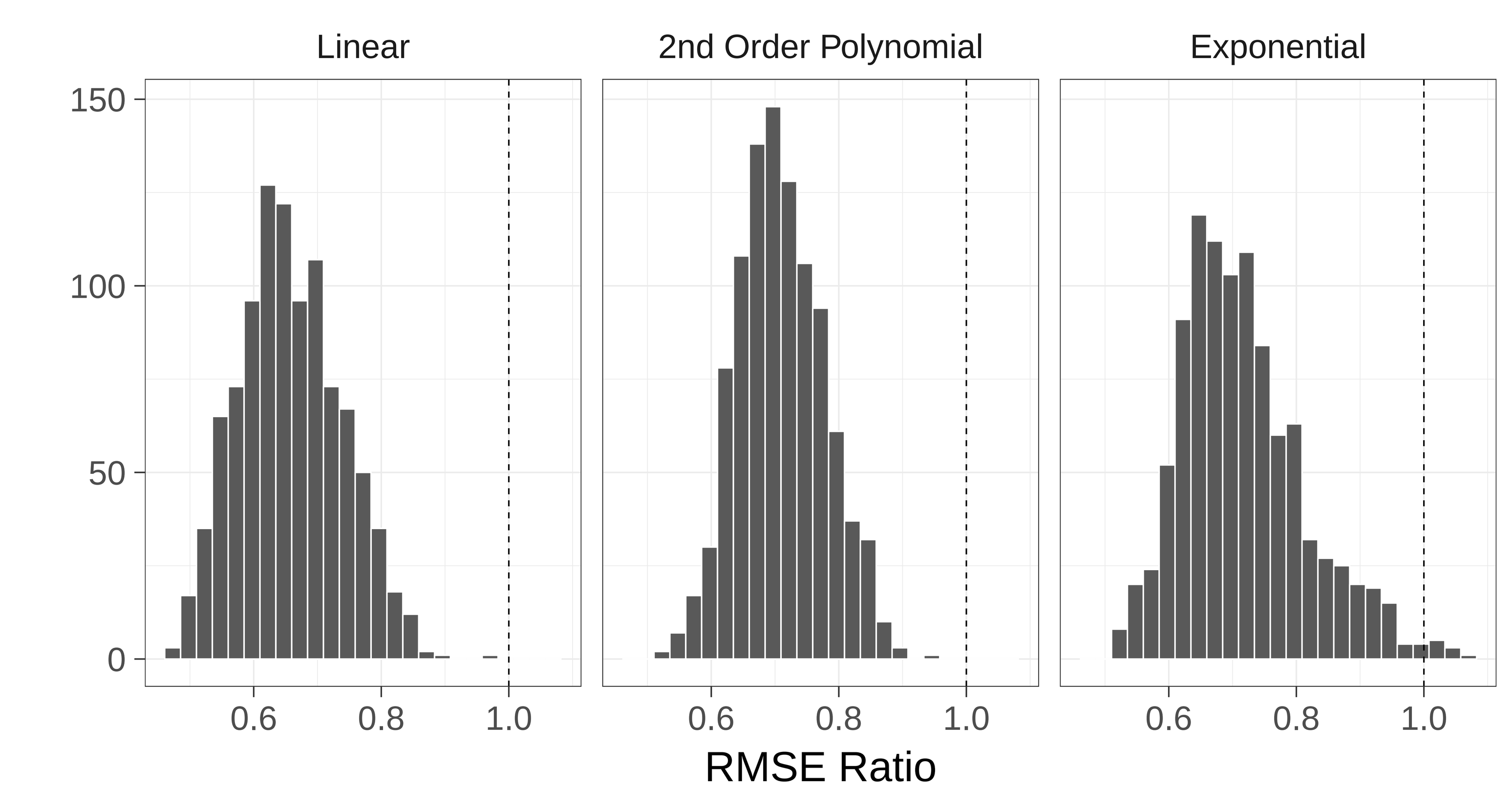


Figure: RMSE ratio of Nuclear vs L1 penalty for simulations specifications in [7].

References

[1] J. Hainmueller, "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, vol. 20, pp. 25–46, 2011.

[2] J. R. Zubizarreta, "Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data," *Journal of the American Statistical Association*, vol. 110, no. 511, pp. 910–922, 2015.

[3] Q. Zhao and D. Percival, "Entropy Balancing is Doubly Robust," *Journal of Causal Inference*, 2016.

[4] Q. Zhao, "Covariate Balancing Propensity Score by Tailored Loss Functions," 2017.

[5] Y. Wang and J. R. Zubizarreta, "Minimal Approximately Balancing Weights: Asymptotic Properties and Practical Considerations." 2018.

[6] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *Journal of the ACM*, vol. 58, no. 3, pp. 11:1—-11:37, 2011.

[7] J. L. Hill, "Bayesian nonparametric modeling for causal inference," *Journal of Computational and Graphical Statistics*, vol. 20, no. 1, pp. 217–240, 2011.