A BNP Model for Zero-Inflated Outcomes with Applications in Causal Inference

Arman Oganisian, Jason Roy, Nandita Mitra

Department of Biostatistics, Epidemiology, and Informatics
Division of Biostatistics
University of Pennsylvania

JSM 2018
July 30, 2018

Basic Setting

Consider a cross-sectional study with

- Binary treatment: $A \in \{0, 1\}$
- Continuous outcome: $Y \in \{-\infty, \infty\}$
- Single, continuous confounder: $L \in \{-\infty, \infty\}$
- Goal: estimate Ψ - the average causal effect of A on Y

$$\Psi = E\left[Y^{A=1} - Y^{A=0}\right]$$

If standard causal assumptions (ignorability, consistency, positivity, no interference) are met, can use Standardization (Robins, 2986)

$$E[Y^{A=a}] = \int E[Y|A = a, L; \boldsymbol{\beta}]dF(L; \boldsymbol{\alpha})$$

Bayesian Standardization

Terms of Ψ are computed using the posterior predictive distribution (Keil, 2017),

$$E[\tilde{y}^a|\boldsymbol{Y},\boldsymbol{L}] = \int_{\boldsymbol{\alpha}}\int_{\boldsymbol{\beta}}\int_{\tilde{\boldsymbol{L}}} E[\tilde{y}|A=a,\tilde{\boldsymbol{L}},\boldsymbol{\beta}]\,p(\tilde{\boldsymbol{L}}|\boldsymbol{\alpha})\,p(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{Y},\boldsymbol{L})\,d\tilde{\boldsymbol{L}}\,d\boldsymbol{\beta}\,d\boldsymbol{\alpha} \tag{1}$$

Need to model conditional distribution of Y and distribution of $\boldsymbol{L}$. E.g.,

$$E[Y|A=a,L=l] = \beta_0 + \beta_1 a + \beta_2 L$$

- Imputation model $E[Y|A=a,\boldsymbol{L},\boldsymbol{\beta}]$ needs to be correctly specified.
- Two sets of rigid assumptions: causal assumptions and statistical assumptions.
- Flexible nonparametric methods can at least help us relax the latter.
  - Especially important for modeling cost data.

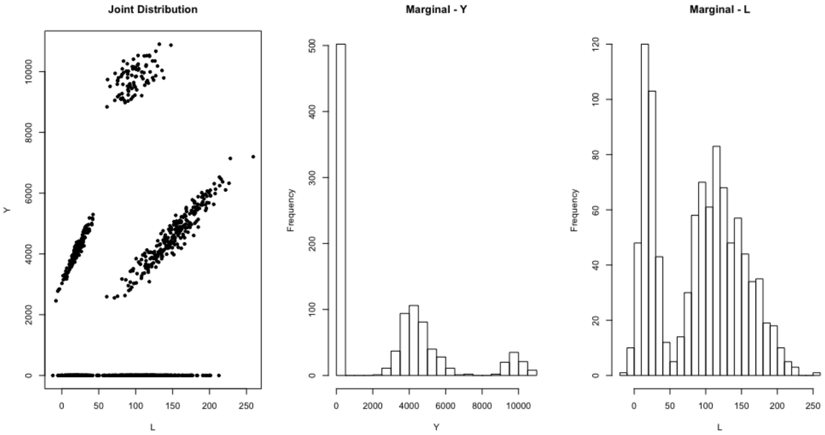
DP Mixture of Zero-Inflated Regression

Building off of previous methods (Hannah, 2011) (Roy, 2018), we propose the following generative model

$$
\begin{aligned}
y_i \mid z_i, \boldsymbol{x}_i &\sim \begin{cases} \delta_0(y_i), & z_i = 1 \\ N(\ \boldsymbol{x}_i^T \boldsymbol{\beta}_i\ ,\ \phi_i), & z_i = 0 \end{cases} \\
z_i | \boldsymbol{x}_i &\sim Ber(\ expit(\ \boldsymbol{x}_i^T \boldsymbol{\gamma}_i\ )\ ) \\
x_{i,j} \sim g_{j,i} &= \begin{cases} N(\lambda_{j,i}, \tau_{j,i}), & x \text{ is continuous} \\ Ber(p_{j,i}), & x \text{ is binary} \end{cases}, \quad j \in \{1, 2, \dots, d\} \\
\left(\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \boldsymbol{\lambda}_i, \boldsymbol{\tau}_i, \boldsymbol{p}_i\right) | G &\sim G \\
G &\sim DP(\alpha G_0)
\end{aligned}
\tag{2}
$$

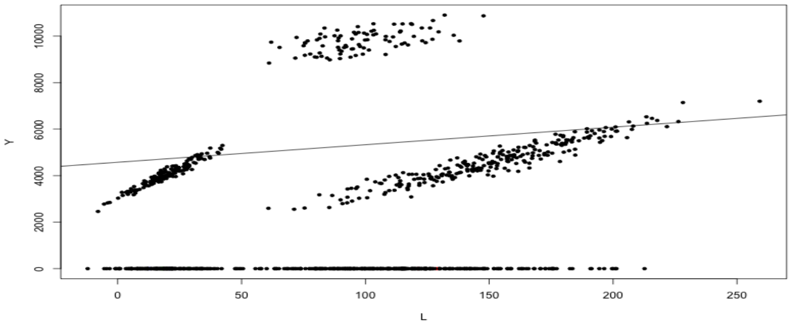
- Nonparametric in the sense that there are infinitely many potential parameters.
- But DP prior induces clustering. Infinitely many possible clusters.

Some Simulated Data

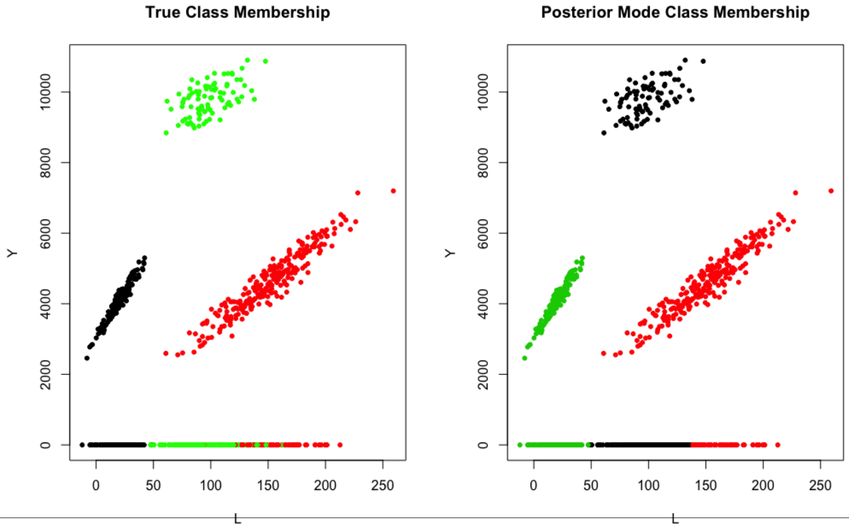


Estimation Using MCMC Methods

We use a Metropolis-within-Gibbs approach extended from (Neal, 2000) and similar to (Roy, 2018).



DP-Induced Clustering



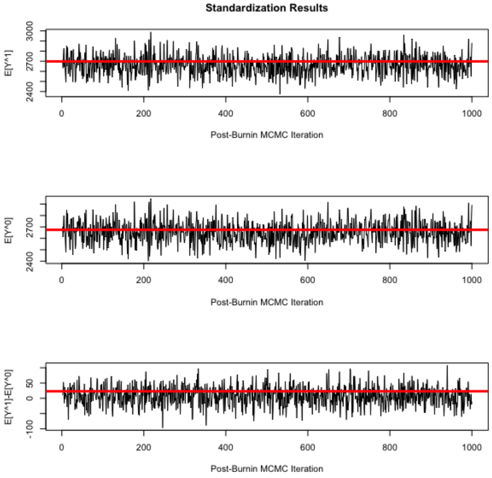
Standardization - Drawing from posterior predictive under different interventions

$$p(\tilde{y}^a|\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}) = \sum_{k=1}^{\infty} \int_{\boldsymbol{\theta}_{x,k}} \int_{\boldsymbol{\beta}_k} \int_{\phi_k} \int_{\boldsymbol{\gamma}} \int_{\tilde{\boldsymbol{x}}} \sum_{l\in\{0,1\}} p(\tilde{y}|\boldsymbol{\beta}_k, \phi_k, c=k, \tilde{z}=l, \tilde{\boldsymbol{x}}^a) \cdot p(\tilde{z}=l|c=k, \tilde{\boldsymbol{x}}^a, \boldsymbol{\gamma_k})$$
$$\cdot\, p(\tilde{\boldsymbol{x}}^a|\boldsymbol{\theta_{x,k}}, \boldsymbol{c=k}) \cdot p(\boldsymbol{\beta}_k, \phi_k, \boldsymbol{\gamma_k}, \boldsymbol{\theta}_{x,k}, c=k|\boldsymbol{y},\boldsymbol{x},\boldsymbol{z}=l)\; d\tilde{\mathbf{x}}d\boldsymbol{\gamma_k}d\phi_k d\boldsymbol{\beta}_k d\boldsymbol{\theta}_{x,k} \tag{3}$$

$$\Psi = E\left[p(\tilde{y}^1|\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\right] - E\left[p(\tilde{y}^0|\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})\right]$$

- Developed Monte Carlo procedure for evaluation of this integral.
- Can compute other causal contrasts, e.g. $E[Y^1]/E[Y^0]$, easily.
- Can compute conditional causal effects using appropriately conditional posterior predictive.
- Interval estimates constructed in the usual ways.

Standardization - MCMC Chains



References

- Roy Jason, Lum Kirsten J., Zeldow Bret, Dworkin Jordan D., Re Vincent Lo, and Daniels Michael J. Bayesian nonparametric generative models for causal inference with missing at random covariates. Biometrics, 0(0).
- Alexander P Keil, Eric J Daza, Stephanie M Engel, Jessie P Buckley, and Jessie K Edwards. A bayesian approach to the g-formula. Statistical Methods in Medical Research, 0(0):0962280217694665, 0. PMID: 29298607.
- Lauren A Hannah, David M Blei, and Warren B Powell. Dirichlet process mixtures of generalized linear models. Journal of Machine Learning Research, 12(Jun):1923?1953, 2011.
- Radford M. Neal. Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics, 9(2):249?265, 2000.

Appendix 1: Causal Assumptions

- Ignorability: $Y_i^{A_i=a} \perp A_i = a | L_i$. Conditional on observed covariates, potential cost is independent of treatment assignment. In randomized control trials, this conditional independence holds by virtue of randomization.
- Consistency: cost Y_i observed under the actual treatment $A_i = a$ is equal to $Y_i^{A_i=a}$. Specifically, $Y_i^{A_i=a} = Y_i | A_i = a$.
- No interference: one patients treatment assignment does not impact another's potential outcome - $Y_i^{A_i=a} \perp A_j, \ \ \forall i \neq j$. A common example of interference is a setting in which Y represents someone's infection status and A represents someone's vaccination status against the disease in question. It is reasonable to consider that one patient's vaccination status may impact another's infection status.
- Positivity: no patient has a deterministic treatment. That is, the probability is strictly between 0 and 1 $0 < P(A_i = 1 | L_i) < 1$. If this assumption did not hold, then one of subject i's potential outcomes would be undefined.