A Bayesian Causal Inference Approach in Observational Studies with Missingness in Covariates and Outcomes

Huaiyu Zang

Heart Institute Research Core
Cincinnati Children's Hospital Medical Center

Joint work with Drs. Hang Kim (UC), Bin Huang (CCHMC) and Rhonda Szczesniak (CCHMC)

Joint Statistical Meetings, August, 2022

Motivating Problem

- Motivated by clinical effectiveness studies of different therapeutic approaches using data of Juvenile idiopathic arthritis (JIA) and Cystic Fibrosis (CF).
  - Goal: evaluate the comparative effectiveness of treatments for chronic disease management.
  - Data characteristics:
    - Mixed continuous and categorical covariates.
    - Missing values in covariates and outcomes
    - Complex data distribution.
    - In the CF study, there are two outcomes of interest with substantial missing values in both.

Introduction

Proposed a Bayesian nonparametric (BNP) causal model:

- Extending hierarchically coupled mixture model of Murray and Reiter (2016) to causal inference.
- Simultaneously imputing missing values, accounting for multiple outcomes, and estimating causal effects.
  - Propagating the uncertainty of missing data to the final causal inference.
  - Handling missing values exist not only in the covariates but also in outcome variables.
  - Capturing complex data distribution
  - Allowing for the mixed-type covariates.

Previous Causal Models with Missing Data

- A primary focus on causal inference considering incomplete data is in the context of propensity score (PS) analysis (Lu and Ashmead, 2018)
  - Generalized propensity score (D'Agostino Jr and Rubin, 2000)
  - Imputing missing data first, followed by applying PS method to imputed data (Leyrat et al 2019).
  - Suffering if the propensity model is mis-specified.
- Kapelner and Bleich (2016) suggested a modified version of the Bayesian additive regression trees (BART) method to handle the missing covariates
  - Incorporating missingness into BART trees.
  - R package `bartMachine`.
  - Performance of `bartMachine` with missing covariates not formally evaluated in causal inference setting.

Previous Causal Models with Missing Data

- Mayer et al. (2020) proposed a doubly robust method to average treatment effect estimation with missing covariates.
    - Their causal estimate may be more biased than that of a single, misspecified model under misspecification of both outcome and propensity score models (Kang and Schafer, 2007)
- Roy et al 2018 adopting an enriched Dirichlet process approach to causal inference with missing covariates.
    - Flexible and data-adaptive
    - Not clear if their approach can address the missing covariates problem in causal inference when the data have a continuous outcome.
    - Limited where missing values exist only in covariates

Basic Notation and Assumption

- Causal notations
  - For a patient record i, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ip_y})^\top$ denotes a vector of p_y outcomes
  - A_i denotes a binary treatment indicator ($A_i = 1$: treated; $A_i = 0$: control).
  - $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip_x})^\top$ denotes a vector of p_x baseline covariates.
  - $\boldsymbol{y}_i^1 = (y_{i1}^1, \ldots, y_{ip_y}^1)^\top$ denotes potential outcomes of treated.
  - $\boldsymbol{y}_i^0 = (y_{i1}^0, \ldots, y_{ip_y}^0)^\top$ denotes potential outcomes of control.
- Causal estimand: average treatment effect (ATE) on p-th outcome.
  - Characterized as $E(\boldsymbol{y}_{ip}^1 - \boldsymbol{y}_{ip}^0)$.
- Causal inference assumption
  - Consistency: $\boldsymbol{y}_i^a = \boldsymbol{y}_i$ with $A_i = a$ for all i.
  - Positivity: $p(A_i = a|\boldsymbol{x}_i) > 0$ if $p(\boldsymbol{x}_i) > 0$.
  - Unconfoundedness: $(\boldsymbol{y}_i^1, \boldsymbol{y}_i^0) \perp A_i|\boldsymbol{x}_i$
- Missing at random (MAR) given covariates assumption: the missingness can be fully accounted for by covariates $\boldsymbol{x}$.

BNP Causal Model: Notation

For unit i,

- $\boldsymbol{u}_i$ (of size $1 \times p_u$): categorical covariates of $\boldsymbol{x}_i$; transformed to $\boldsymbol{u}_i^*$ of size p_u^*.
- $\boldsymbol{v}_i$ (of size $1 \times p_v$): p_y outcome variables $\boldsymbol{y}_i$, and $(p_v - p_y)$ continuous covariates of $\boldsymbol{x}_i$
- $\boldsymbol{D}_i$ (of size $p_v \times p_y$): treatment indicator matrix.
    - For a treated unit, $\boldsymbol{D}_i$ is defined by stacking the identity matrix of size $p_y \times p_y$ on the zero matrix of size $(p_v - p_y) \times p_y$.
    - For a control unit, $\boldsymbol{D}_i$ is the zero matrix of size $(p_v \times p_y)$.
- $\boldsymbol{B}_r = (\boldsymbol{\beta}_{1r} \cdots \boldsymbol{\beta}_{p_v r})^\top$ of size $p_v \times p_u^*$ and $\boldsymbol{\Sigma}_r$ of size $p_v \times p_v$.
- $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p_y})^\top$ of size p_y.
- $r_i \in \{1, \ldots, R\}$ label the mixture component for $\boldsymbol{v}_i$.
- $s_i \in \{1, \ldots, S\}$ label the mixture component for $\boldsymbol{u}_i$.

BNP Causal Model: Data Model

The data model assumes that:

$$
\begin{aligned}
&[\boldsymbol{v}_i \mid \boldsymbol{u}_i, \boldsymbol{D}_i, \boldsymbol{\delta}, r_i, \{\boldsymbol{B}_r\}, \{\boldsymbol{\Sigma}_r\}] \sim \mathrm{N}(\ \boldsymbol{B}_{r_i}\boldsymbol{u}_i^* + \boldsymbol{D}_i\boldsymbol{\delta}\ ,\ \boldsymbol{\Sigma}_{r_i}\ ), \quad r_i = 1, \ldots, R, \\
&[u_{il} \mid s_i, \{\boldsymbol{\psi}_{ls}\}] \sim \mathrm{Categ}(\boldsymbol{\psi}_{ls_i}), \quad l = 1, \ldots, p_u,\ s_i = 1, \ldots, S,
\end{aligned}
\tag{1}
$$

- The response surface of the dth outcome of unit i can be expressed as $E(y_{id}|\boldsymbol{x}_i, A_i, \boldsymbol{B}_{r_i}, \boldsymbol{\Sigma}_{r_i}, \delta_d) = f_d(\boldsymbol{x}_i, \boldsymbol{B}_{r_i}, \boldsymbol{\Sigma}_{r_i}) + A_i\delta_d$ where the flexible functional form is derived from the mixture regression in (1) to capture nonlinearities in the response surface.
- $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p_y})^\top$ have the desired ATE interpretation following three standard causal assumptions (Imbens and Rubin, 2015).

Cincinnati Children's changing the outcome together

BNP Causal Model: Dirichlet process priors for mixture components

We write the likelihood and prior for mixture components as follows:

1. Hierarchical prior for component indexes

$$[k_i|\boldsymbol{\pi}] \sim \text{Categorical}(\boldsymbol{\pi}) \tag{2}$$

$$[r_i, s_i \mid k_i, \{\boldsymbol{\eta}_k\}, \{\boldsymbol{\lambda}_k\}] = [r_i \mid k_i, \{\boldsymbol{\eta}_k\}][s_i \mid k_i, \{\boldsymbol{\lambda}_k\}] \quad \text{where} \tag{3}$$

$$[r_i \mid k_i, \{\boldsymbol{\eta}_k\}] \sim \text{Categorical}(\boldsymbol{\eta}_{k_i}), \qquad [s_i \mid k_i, \{\boldsymbol{\lambda}_k\}] \sim \text{Categorical}(\boldsymbol{\lambda}_{k_i})$$

2. Each mixture component probabilities are assigned a truncated stick-breaking process

$$\pi_k = \tilde{\pi}_k \prod_{k'=1}^{k-1}(1-\tilde{\pi}_{k'}), \quad \tilde{\pi}_k \overset{iid}{\sim} \text{Beta}(1, \alpha^{(k)}) \text{ for } k = 1, \ldots, K-1, \quad \tilde{\pi}_K \equiv 1$$

$$\eta_{kr} = \tilde{\eta}_{kr} \prod_{r'=1}^{r-1}(1-\tilde{\eta}_{kr'}), \quad \tilde{\eta}_{kr} \overset{iid}{\sim} \text{Beta}(1, \alpha^{(r)}) \text{ for } r = 1, \ldots, R-1, \quad \tilde{\eta}_{kR} \equiv 1$$

$$\lambda_{ks} = \tilde{\lambda}_{ks} \prod_{s'=1}^{s-1}(1-\tilde{\lambda}_{ks'}), \quad \tilde{\lambda}_{ks} \overset{iid}{\sim} \text{Beta}(1, \alpha^{(s)}) \text{ for } s = 1, \ldots, S-1, \quad \tilde{\lambda}_{kS} \equiv 1$$

BNP Causal Model: Graphical Representation

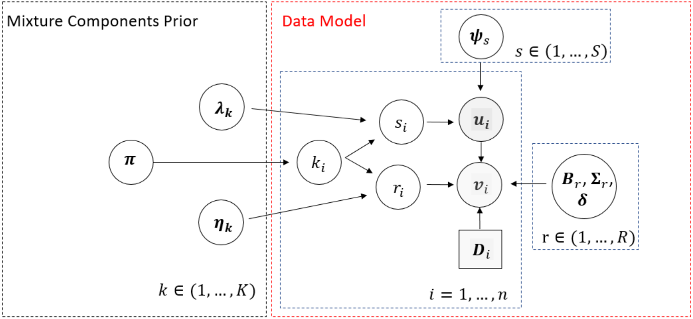

Figure: Graphical representation of Bayesian nonparametric causal analysis model. The joint distribution of the categorical covariates $\boldsymbol{u}_i$ is modeled by Dirichlet process (DP) mixture multinomial distributions, then the conditional distribution of continuous variables $\boldsymbol{v}_i$, including the outcome variables and continuous covariates, is modeled by the DP mixture regression given $\boldsymbol{u}_i$ and the treatment indicator matrix $\boldsymbol{D}_i$.

Method Compared

- Existing causal methods:
  1. Bayesian additive regression tree (BART)
  2. Linear model (LM) with all covariates
  3. Inverse probability of treatment weighting (IPTW)
  4. Targeted maximum likelihood estimation (TMLE)
- Three approaches compared (across 1000 simulated datasets):
  1. `bartMachine` with missing covariates
  2. Existing causal models applied to the completed data (MICE + existing causal model).
     - Imputing missing data via multiple imputation by chained equations (MICE).
     - Pooling the treatment effect estimates using Rubin's rule.
  3. BNP causal model (BNPc)
     - Upper bound of occupied components $K = 20$, $R = 30$, and $S = 50$.
     - The ATE estimate is obtained by averaging over 10,000 MCMC iterations from the Gibbs sampler, after discarding the first 2,000 iterations as burn-ins
     - The MCMC draws were enough to accurately capture the posterior.

Simulation Setting 1: Outcome With a Bimodal Distribution

- Data setup (mimic the simulation setting in Roy et al. 2018)
    - Four continuous covariates with multivariate normal of mean 0, variance 1, and correlation 0.3.
    - Two categorical covariates that are related to the continuous covariates.
    - Potential outcomes come from a bi-modal distribution.
    - Introducing missing data in the covariates (The percentage of complete cases is around 9.1%).
- Reporting the point estimate (Est) of the ATE, standard error (SE), root mean square error (RMSE), median absolute error (MAE), 95% confidence interval coverage, and averaged standard error estimate (SEE) as evaluation metrics.

Simulation Setting 1: Causal Inference Results

Method	Est	SE	RMSE	MAE	Coverage	SEE
Ref	1.487	0.120	0.121	0.083	0.974	0.144
BNPc	1.542	0.132	0.139	0.095	0.950	0.141
bartMachine	1.684	0.351	0.396	0.265	0.912	0.356
MICE+ BART	1.370	0.324	0.349	0.240	0.959	0.352
MICE+ LM	1.514	0.440	0.440	0.304	0.949	0.438
MICE+ IPTW	1.473	0.445	0.446	0.313	0.986	0.574
MICE+ TMLE	1.479	0.452	0.452	0.318	0.949	0.449

Table: ATE results over 1000 repetitions from Simulation 1 with the true ATE of 1.5 ($n = 500$). BNPc denotes the proposed Bayesian nonparameteric causal inference method; bartMachine denotes the bartMachine model with missing covariates; MICE+ denotes the existing causal inference methods (BART, LM, IPTW, and TMLE) applied to the imputed data generated from the MICE; Ref denotes the proposed BNP method applied to the simulation data set *before introducing missingness*.

Simulation Setting 2: Outcome and Covariates With a Mixture Distribution

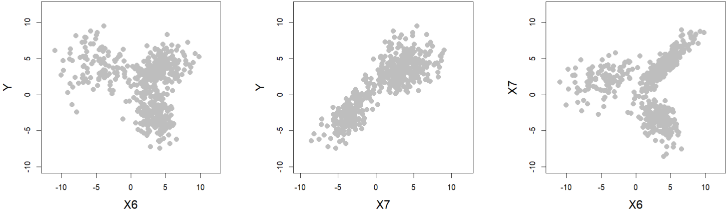

Figure: Bivariate plots of continuous variables in one simulation data of Simulation 2.

Simulation Setting 2: Causal Inference Results

Method	Est	SE	RMSE	MAE	Coverage	SEE
Ref	1.489	0.107	0.108	0.076	0.968	0.117
BNPc	1.496	0.126	0.126	0.085	0.957	0.130
bartMachine	1.438	0.165	0.177	0.115	0.912	0.155
MICE+ BART	1.261	0.146	0.280	0.236	0.748	0.172
MICE+ LM	1.308	0.176	0.260	0.188	0.828	0.186
MICE+ IPTW	1.148	0.355	0.500	0.369	0.998	0.713
MICE+ TMLE	1.331	0.212	0.271	0.185	0.947	0.264

Table: ATE results over 1000 repetitions from Simulation 2 with the true ATE of 1.5 ($n = 500$). The average percentage of complete cases over the repeated simulations is around 22.2%.

Simulation Setting 2: Imputation Performance Comparison

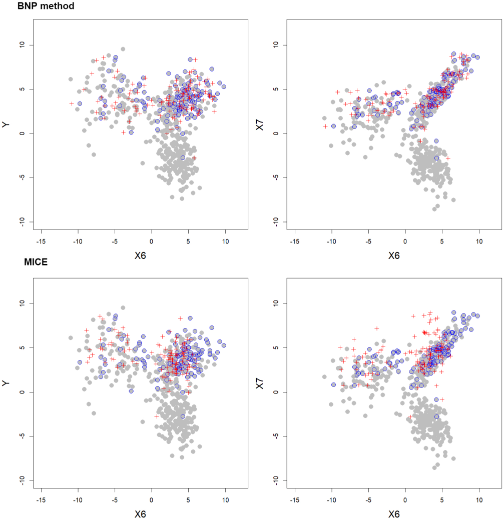

Figure: The solid dots represent original data points; the empty dots represent missing data points; and the crosses represent imputed data points drawn from each method.

Simulation Setting 3: Missingness in Both Outcomes and Covariates

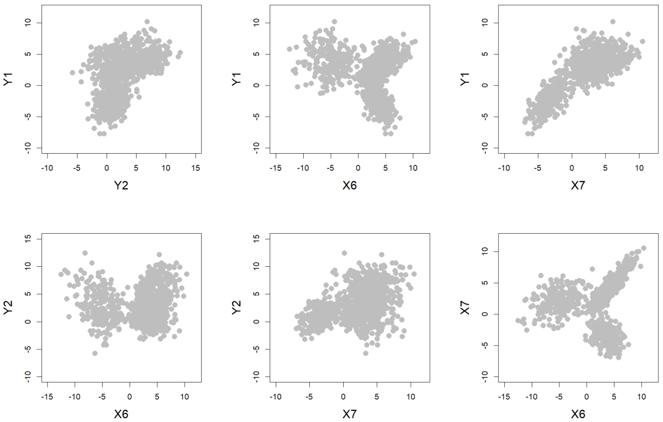

Figure: Bivariate plots of continuous variables in one simulation dataset from Simulation Study 3.

Simulation Setting 3: Causal Inference Results

Estimand	Method	Est	SE	RMSE	MAE	Coverage	SEE
ATE on Y_1	BNPc	1.474	0.094	0.098	0.066	0.951	0.095
	MICE+ BART	1.138	0.161	0.396	0.361	0.485	0.176
	MICE+ LM	1.123	0.182	0.418	0.371	0.466	0.176
	MICE+ IPTW	0.991	0.481	0.700	0.528	0.909	0.570
	MICE+ TMLE	1.263	0.209	0.315	0.253	0.852	0.228
ATE on Y_2	BNPc	0.472	0.090	0.094	0.063	0.942	0.093
	MICE+ BART	-0.124	0.271	0.680	0.625	0.542	0.318
	MICE+ LM	-0.189	0.298	0.751	0.705	0.477	0.324
	MICE+ IPTW	-0.054	0.490	0.740	0.572	0.873	0.621
	MICE+ TMLE	0.046	0.357	0.577	0.460	0.836	0.404

Table: ATE estimation over 1000 repetitions for two outcome variables in Simulation 3 ($n = 1,000$). The true ATE on Y_1 and Y_2 are 1.5 and 0.5 respectively. The average percentage of complete cases over the repeated simulations is around 19%.

Cincinnati Children's changing the outcome together

JIA Data

- Treatment: “early aggressive approach" ($n = 135$) vs. “conservative approach" ($n = 330$).
- Outcome: clinical Juvenile Arthritis Disease Activity Score (cJADAS) at 6 months.
  - Disease severity score calculated based on three core clinical measures.
- Five continuous covariates and eight categorical covariates were considered as confounding factors based on our clinical knowledge.
- Suffering from missing data in the outcome as well as some covariates
  - Missing rates for of cJADAS at 6 months, uveitis ever-positive indicator and the quality of life total index are 0.53, 0.50, and 0.48, respectively.

Results for JIA Data

Method	Est	SE	95% CI	p-value	$P(\text{ATE} < 0)$
BNPc	-1.26	0.60	(-2.42, -0.06)	0.018	0.980
MICE+BART	-1.76	0.94	(-3.72, 0.21)	0.077	0.944
MICE+LM	-1.47	0.87	(-3.28, 0.35)	0.107	
MICE+IPTW	-1.65	0.96	(-3.63, 0.33)	0.098	
MICE+TMLE	-0.38	0.53	(-1.41, 0.65)	0.470	

Table: Results for JIA Data. 95% CI indicates the credible interval for BNPc and the confidence intervals for other methods. p-value for our proposed model is calculated based on Bayesian asymptotic theory using posterior mean and variance. The $P(\text{ATE} < 0)$ denotes the posterior probability of ATE less than 0, suggesting prescribing early aggressive plan being effective in decreasing the disease activity. $P(\text{ATE} < 0)$ for MICE + BART is calculated based on one imputed data generated from the MICE.

Results for JIA Data

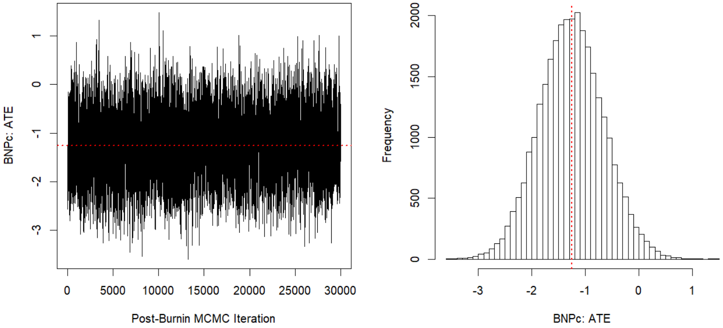

Figure: Trace plot and histogram of the ATE estimate drawn from the proposed BNP method with the JIA data. The dotted lines represent estimated causal effects.

CF Data

- Goal: evaluate the effects of two treatment strategies according to the timing of inhaled tobramycin (Tobi) delivery.
- Treatment: early Tobi ($n = 289$) vs. late Tobi ($n = 266$).
- Outcomes: FEV_1 measured at 6 months and 12 months after the first pseudomonas aeruginosa infection.
- Time window: set a certain time window defined with the time of interest plus or minus a margin, then consider the measurements of the visits within the time window as those at the time of interest.
  - Trade-off between the size of the time window and the missing rate in outcome variables: as the time window increased, we had less accurate measurements, but we had reduced the number of missing values.
  - Examining four different time window sizes: 1 week, 2 weeks, 1 month, and 1.5 months.

Results for CF Data

Time window	Estimand	Est	SE	95% CI	$P(\text{ATE} > 0)$
1 week	ATE at 6 month	6.04	2.37	(1.36, 10.72)	0.994
	ATE at 12 month	2.71	1.98	(-1.24, 6.45)	0.908
2 weeks	ATE at 6 month	2.90	1.79	(-0.61, 6.39)	0.946
	ATE at 12 month	-0.68	1.66	(-4.00, 2.58)	0.341
1 month	ATE at 6 month	1.51	1.44	(-1.36, 4.28)	0.852
	ATE at 12 month	1.77	1.34	(-0.85, 4.40)	0.906
1.5 months	ATE at 6 month	0.66	1.24	(-1.76, 3.09)	0.705
	ATE at 12 month	0.95	1.21	(-1.42, 3.37)	0.787

Table: Results of CF study for ATE estimates of lung function at two time points using the proposed BNP method with varying time windows. 95% credible interval are presented for the proposed method. Time windows are used to define the measurements at a particular time point.

Strengths/Importance

- Flexible and data adaptive, minimizing model mis-specification issues.
- Considering the missing values that exist not only in the covariates but also in the outcome variables.
- Allowing for the mixed-type covariates and multiple outcomes.
- Presenting a good summary (e.g., posterior probability) for communicating the results and providing comprehensive distributional information.

Thank you!

Contact Information

Huaiyu Zang

Huaiyu.Zang@cchmc.org

Heart Institute Research Core

Cincinnati Children's Hospital Medical Center

Key References

- Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1), 217-240.
- Imbens, Guido W., and Donald B. Rubin. (2015) Causal inference in statistics, social, and biomedical sciences. *Cambridge University Press*.
- Si, Y., and Reiter, J. P. (2011). A comparison of posterior simulation and inference by combining rules for multiple imputation. *Journal of Statistical Theory and Practice*, 5(2), 335-347.
- Murray, J. S., and Reiter, J. P. (2016). Multiple imputation of missing categorical and continuous values via Bayesian mixture models with local dependence. *Journal of the American Statistical Association*, 111(516), 1466-1479.
- Lu, B., and Ashmead, R. (2018). Propensity score matching analysis for causal effects with MNAR covariates. *Statistica Sinica*, 28(4), 2005-2025.
- Kang, J. D., and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science*, 22(4), 523-539.

Key References

- D'Agostino Jr, R. B., and Rubin, D. B. (2000). Estimating and using propensity scores with partially missing data. *Journal of the American Statistical Association*, 95(451), 749-759.
- Roy, J., Lum, K. J., Zeldow, B., Dworkin, J. D., Re III, V. L., and Daniels, M. J. (2018). Bayesian nonparametric generative models for causal inference with missing at random covariates. *Biometrics*, 74(4), 1193-1202.
- Leyrat, C., Seaman, S. R., White, I. R., Douglas, I., Smeeth, L., Kim, J. and Williamson, E. J. (2019). Propensity score analysis with partially observed covariates: How should multiple imputation be used?. *Statistical methods in medical research*, 28(1), 3-19.
- Kapelner A, Bleich J (2016). "bartMachine: Machine Learning with Bayesian Additive Regression Trees." *Journal of Statistical Software*, 70(4), 1–40.
- Mayer, I., Sverdrup, E., Gauss, T., Moyer, J. D., Wager, S., and Josse, J. (2020). Doubly robust treatment effect estimation with missing attributes. *The Annals of Applied Statistics*, 14(3), 1409-1431.
- Ishwaran, H., and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453), 161-173.

Cincinnati Children's™
changing the outcome together

Appendix: Causal Inference Notation/Estimands ($p_y = 1$)

- $\boldsymbol{x}_i$: a vector of p baseline (pre-treatment) covariates.
- A_i: binary treatment indicator.
  - $A_i = 1$ for a unit assigned to the treatment and $A_i = 0$ for a unit assigned to the control.
- y_i^0 and y_i^1: potential outcomes if a unit would be assigned to the treatment and to the control, respectively (Neyman 1923, Rubin 1974).
- Causal effects are functions of y_i^1 and y_i^0.
  - For example: $\text{ATE} = E(y_i^1 - y_i^0)$ (our causal estimand of interest).
- Fundamental Problem of Causal Inference: we can only observe one potential outcome for each unit.

Appendix: Bayesian Causal Model

- Rubin (1978) first introduced the ideas of Bayesian analysis for causal modelling.
  - Considering the missing potential outcomes as unobserved random variables.
  - $Pr(y^1, y^0|\boldsymbol{x})$
- Bayesian additive regression tree (BART) used in causal inference; Hill (2011)
  - Modeling the response surface: $(y|A = a, \boldsymbol{x} = x) = f(A = a, \boldsymbol{x} = x) + \epsilon = \sum_{j=1}^{m} g(A = a, \boldsymbol{x} = x; T_j, M_j) + \epsilon.$
  - Allowing for main effects for each covariate, as well as their interactions.
  - Producing accurate estimates of ATE.
- Many Bayesian causal methods could not handle the missing data directly.
  - In practice, (1) causal inference with complete-case only, i.e., disregarding the missing values OR (2) Imputing missing data first, followed by applying causal inference methods to imputed data.

Appendix: Bayesian Nonparametrics Approaches

- Bayesian Nonparametrics (BNP) model:
  - Flexible and adaptive to the different data's characteristics.
  - Lessening the model mis-specification problems.
  - Among all BNP models, the Dirichlet process (DP) is commonly chosen as the prior.

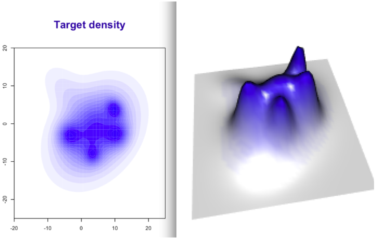

Figure: A **mixture** of the normal dist'ns is useful to capture complex form.

Appendix: Stick Breaking

Sethuraman (1994) showed that the DP can be characterized through the stick-breaking representation.

$$P = \sum_{k=1}^{\infty} \pi_k \delta_{\Theta_k} \tag{4}$$

$$\pi_k = \nu_k \prod_{g=1}^{k-1} (1 - \nu_g) \tag{5}$$

$$\nu_k \sim \mathrm{Beta}(1, \alpha), \quad \Theta_k \sim P_0 \tag{6}$$

- π_k is constructed through sequential breaks of a stick of length one.
- Encouraging the weight π_k to decrease stochastically; most of the probability in π is allocated to the first few components.
- Ishwaran and James (2001) introduced the blocked Gibbs sampler and suggested using the truncation of the stick breaking prior distributions.

Appendix: Multiple imputation

- Multiple imputation (MI)
  - Creating $m > 1$ completed datasets.
  - Combining them using simple rules to get pooled estimates.
    - Estimate $\hat{\theta}^{(l)}$ and its variance $u^{(l)}$ for each completed data for the point estimator θ and its variance estimates u.
    - $\bar{\theta}_m = \frac{1}{m} \sum_{l=1}^{m} \hat{\theta}^{(l)}$, $\bar{u}_m = \frac{1}{m} \sum_{l=1}^{m} u^{(l)}$, $b_m = \frac{1}{m-1} \sum_{l=1}^{m} \left(\hat{\theta}^{(l)} - \bar{\theta}_m\right)^2$
    - Rubin (1987)

$$\theta - \bar{\theta}_m \sim t_\nu(0, T_m), \qquad \text{where} \quad T_m = \bar{u}_m^2 + \left(1 + \tfrac{1}{m}\right) b_m$$

    - Si and Reiter (2011) justified to use the posterior variance as the variance estimate $u^{(l)}$ used in Rubin's combining rule formula for Bayesian inference after MI.
- Benefits of using MI:
  - Accounting for the imputation uncertainty.
- Performing MI can be based on different imputation strategies (eg., sequential modeling or joint modeling).

Cincinnati Children's
changing the outcome together

Appendix: BNP Causal Model Properties

- Capturing irregular and complicated relationships within or between categorical and continuous variables.
  - The associations among continuous variables $\boldsymbol{v}_i$ can be captured via r_i.
  - The associations among categorical variables $\boldsymbol{u}_i$ can be captured via s_i.
  - The dependence between $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$ can be captured in two ways:
    - Captured via component-specific regression functions and covariance matrices
    - Captured via the hierarchical structure of mixture components.
- Representing a wide variety of shapes for the surface of outcome variable on baseline covariates.

Appendix: Evaluation Metrics

- Average of the estimates (Est): $\frac{1}{n}\sum_{r=1}^{n}\hat{\delta}_r$, where $\hat{\delta}_r$ is a method-specific point estimate of δ in replication r and n is the number of replications.
- Standard Error (SE) of the mean estimates: $\sqrt{\frac{1}{(n-1)}\sum_{r=1}^{n}\{\hat{\delta}_r - (\frac{1}{n}\sum_{r=1}^{n}\hat{\delta}_r)\}^2}$.
- Root mean square error (RMSE): $\sqrt{\frac{1}{n}\sum_{r=1}^{n}(\hat{\delta}_r - \delta)^2}$.
- Median absolute error (MAE): median($\mid \hat{\delta}_r - \delta \mid$).
- 95% nominal coverage probability (Coverage): $\frac{1}{n}\sum\limits_{r=1}^{n} I\{L(\hat{\delta}_r) < \delta < U(\hat{\delta}_r)\}$, where $L(\hat{\delta}_r)$ and $U(\hat{\delta}_r)$ are lower and upper endpoints of the 95% confidence interval estimate, respectively.
- Average standard error estimate (SEE): $\frac{1}{n}\sum_{r=1}^{n}\widehat{\text{se}}(\hat{\delta}_r)$, where $\widehat{\text{se}}(\hat{\delta}_r)$ is the associated standard error estimator of $\hat{\delta}_r$ in replication r.

Appendix: Simulation Setting 1

Table: Average treatment effects (ATE) estimation in Simulation Study 1 (n=500) with the true ATE of 1.5, comparing existing methods. A method name followed by CC denotes the causal method applied to the the complete-case only dataset.

Method	Est	SE	RMSE	MAE	Coverage	SEE
bartMachine	1.684	0.351	0.396	0.265	0.912	0.356
BART CC	1.387	0.677	0.686	0.436	0.935	0.638
MICE + BART	1.370	0.324	0.349	0.240	0.959	0.352
LM CC	1.509	0.969	0.968	0.627	0.950	0.916
MICE + LM	1.514	0.440	0.440	0.304	0.949	0.438
IPTW CC	1.522	1.021	1.020	0.601	0.963	0.971
MICE + IPTW	1.473	0.445	0.446	0.313	0.986	0.574
TMLE CC	1.531	1.048	1.048	0.632	0.870	0.732
MICE + TMLE	1.479	0.452	0.452	0.318	0.949	0.449

Appendix: Simulation Setting 2

Table: Average treatment effects (ATE) estimation in Simulation Study 2 (n=500) with the true ATE of 1.5, comparing existing methods. A method name followed by CC denotes the causal method applied to the the complete-case only dataset.

Method	Est	SE	RMSE	MAE	Coverage	SEE
bartMachine	1.438	0.165	0.177	0.115	0.912	0.155
BART CC	1.277	0.320	0.390	0.263	0.901	0.334
MICE+ BART	1.261	0.146	0.280	0.236	0.748	0.172
LM CC	1.426	0.386	0.393	0.251	0.942	0.391
MICE+ LM	1.308	0.176	0.260	0.188	0.828	0.186
IPTW CC	1.377	0.885	0.893	0.516	0.952	0.865
MICE+ IPTW	1.148	0.355	0.500	0.369	0.998	0.713
TMLE CC	1.529	0.595	0.595	0.393	0.665	0.290
MICE+ TMLE	1.331	0.212	0.271	0.185	0.947	0.264

Appendix: Future Direction

- Investigating heterogeneous causal effects across subgroups
  - Allowing the causal parameters to vary by the mixture components, i.e., $\boldsymbol{\delta}_r$
  - Challenges: mixture components of the BNP model do not necessarily represent subclusters and show label switching phenomenon that needs additional post-processing.
- Accounting for the time-varying confounding setting
- Extension to a dynamic treatment regime
- Generalizing our data model to account for mixed types of outcomes, i.e., continuous and categorical outcomes