Global Sensitivity Analysis of Randomized Trials with Non-Monotone Missing Binary Outcomes:

Application to Studies of Substance Use Disorders

Daniel Scharfstein (PI)
Johns Hopkins University
dscharf@jhu.edu

September 17, 2019

Co-Investigators

- Aimee Campbell (Columbia)
- Edward Nunes (Columbia)
- Abigail Matthews (EMMES)
- Aidan McDermott (Johns Hopkins)
- Chenguang Wang (Johns Hopkins)
- Jon Steingrimmson (Brown)

Aims

1. Develop and evaluate a sensitivity analysis methodology for the analysis of randomized clinical trials with repeatedly measured binary outcomes and non-monotone missing data.
2. Develop open-source, user-friendly software.
3. Conduct sensitivity analysis of 29 CTN-sponsored trials with public-use datasets available on the NIDA Data Share website.
4. Link the results to study characteristics in order to identify patterns.
5. Disseminate the methodology and software to researchers interested in substance use disorder clinical trials.

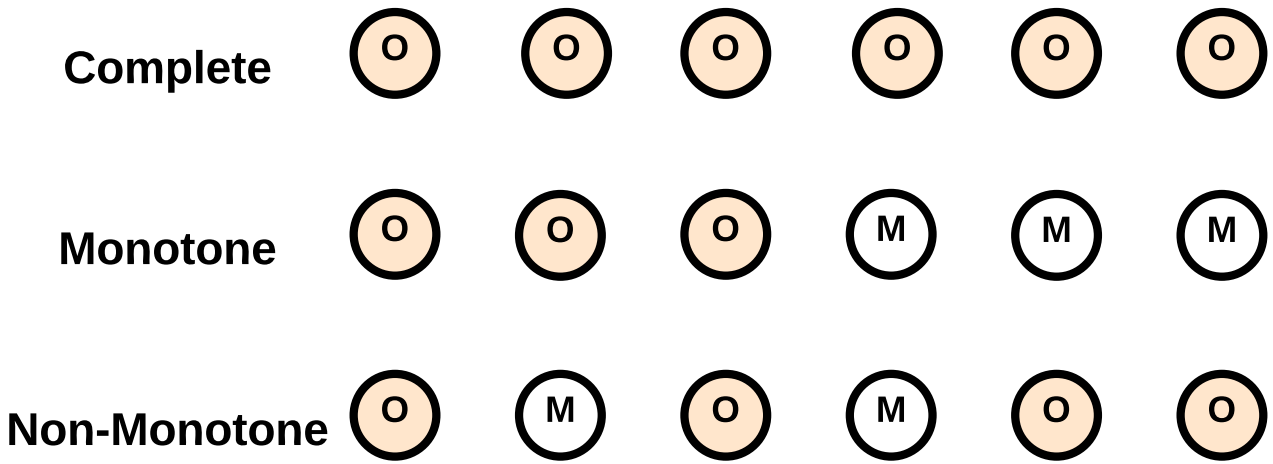
Sensitivity Analysis

- Missing outcome data threaten the validity of randomized clinical trials because inference about treatment effects then necessarily relies on untestable assumptions, which wrongly stated can lead to incorrect conclusions.
- The National Research Council (NRC) in its report entitled "The Prevention Treatment of Missing Data in Clinical Trials" recommended that evaluating the sensitivity of trial results to assumptions about the missing data mechanism should be a mandatory component of reporting.

Sensitivity Analysis

- Chapter 5 of the NRC Report presents an approach whereby one posits a broad class of untestable missing data assumptions that is
  1. indexed by sensitivity analysis parameters,
  2. anchored around a plausible benchmark assumption (sensitivity parameters equal to a reference value), and
  3. sensitivity analysis parameters further from the reference value represent larger deviations from the benchmark assumption.
- The goal of this "global" sensitivity analysis approach is to determine how much deviation from a benchmark assumption is required in order for inferences to change.
- If the deviation is judged to be sufficiently far from the benchmark assumption, then greater credibility is lent to the benchmark analysis; if not, the benchmark analysis can be considered to be fragile.

Missing Data



Non-Monotone Missing Data

- Positing plausible assumptions and specifying flexible models for studies with non-monotone missing data is challenging because of the potentially large number of missingness patterns (as many as $2^K - 1$ patterns, where K is the number of post-baseline assessments).
- Ibrahim and Molenberghs (2009) indicate that "[s]uch data present a considerable modeling challenge for the statistician".
- The NRC report highlighted the need for development and application of "novel, appropriate methods of model specification and sensitivity analyses to handle non-monotone missing data patterns".

Notation

- $Y_k^{(1)}$ is outcome scheduled to be measured at time k
- R_k is the indicator that $Y_k^{(1)}$ is observed.
- Y_k is the observed outcome at time k
  - $Y_k = Y_k^{(1)}$ if $R_k = 1$
  - $Y_k =?$ if $R_k = 0$
- For a time varying quantity Z_k, let
  - $\overleftarrow{Z}_k = (Z_1, \ldots, Z_k)$
  - $\overrightarrow{Z}_k = (Z_{k+1}, \ldots, Z_K)$
- $O_k = (R_k, Y_k)$
- We observe n i.i.d. copies of $\overleftarrow{O}_K$

Learn about functionals of the joint distribution of $\overleftarrow{Y}_K^{(1)}$, e.g.,

- $E[\sum_{k=1}^{K} Y_k^{(1)}]$

Missing at Random (MAR)

- Robins and Gill (1997), Gill and Robins (1997) and Little and Rubin (2014) have argued that MAR is implausible for studies that have non-monotone missing data patterns.

R_1	R_2	R_3	$Y_1^{(1)}$	$Y_2^{(1)}$	$Y_3^{(1)}$	$P[\overleftarrow{R}_3 = \overleftarrow{r}_3 \mid \overleftarrow{Y}_3^{(1)} = \overleftarrow{y}_3^{(1)}]$
1	1	0	0	0	0	1.0
1	0	1	1	0	0	1.0
1	1	1	0	1	0	1.0
1	1	0	0	0	1	1.0
1	0	1	1	1	0	1.0
1	1	1	1	0	1	1.0
0	1	1	0	1	1	1.0
0	1	1	1	1	1	1.0

- $P[\overleftarrow{R}_3 = (1,1,0) \mid \overleftarrow{Y}_3^{(1)}]$ does not depend on $Y_3^{(1)}$
- $P[\overleftarrow{R}_3 = (1,0,1) \mid \overleftarrow{Y}_3^{(1)}]$ does not depend on $Y_2^{(1)}$
- $P[\overleftarrow{R}_3 = (0,1,1) \mid \overleftarrow{Y}_3^{(1)}]$ does not depend on $Y_1^{(1)}$

Missing at Random (MAR)

Simulation of the MAR process:

- Observe $Y_3^{(1)}$
  - If $Y_3^{(1)} = 1$, observe $Y_2^{(1)}$
    - If $Y_2^{(1)} = 1$, do not observe $Y_1^{(1)}$
    - If $Y_2^{(1)} = 0$ observe $Y_1^{(1)}$. If $Y_1^{(1)} = 0$, hide $Y_3^{(1)}$
  - If $Y_3^{(1)} = 0$, observe $Y_1^{(1)}$
    - If $Y_1^{(1)} = 1$, do not observe $Y_2^{(1)}$
    - If $Y_1^{(1)} = 0$ observe $Y_2^{(1)}$. If $Y_2^{(1)} = 0$, hide $Y_3^{(1)}$

- Need information on potentially hidden value of $Y_3^{(1)}$ to simulate data.
- "MAR is not want it seems" (Gill and Robins, 1997)

Complete-Case Missing Value

- For each possible pattern with missing observations, the conditional distribution of the missing outcomes given the observed outcomes is equal to corresponding distribution for the pattern with no missing observations.
- Tchetgen-Tchetgen, Wang and Sun (2017) developed a global sensitivity analysis procedure anchored at CCMV.

$$P[\overleftarrow{Y}_K^{(1)} = \overleftarrow{y}_K^{(1)} | \overleftarrow{R}_K = \overleftarrow{r}_K, \overleftarrow{Y}_K = \overleftarrow{y}_K]$$
$$= P[\overleftarrow{Y}_K^{(1)} = \overleftarrow{y}_K^{(1)} | \overleftarrow{R}_K = \overleftarrow{1}_K, \overleftarrow{Y}_K = \overleftarrow{y}_K]$$

For example,

$$P[Y_1^{(1)} = y_1^{(1)}, Y_2^{(1)} = y_2^{(1)} | R_1 = R_2 = 0, R_3 = 1, Y_3 = y_3]$$
$$= P[Y_1^{(1)} = y_1^{(1)}, Y_2^{(1)} = y_2^{(1)} | R_1 = R_2 = R_3 = 1, Y_3 = y_3]$$

Vansteelandt, Rotnitzky and Robins (2007)

- For individuals who have the same observed data prior to a scheduled visit, the distribution of the outcome for those missing the visit is the same as the distribution of the outcome for those who attend the visit.
- They developed a global sensitivity analysis procedure anchored at this assumption.
- Linero and Daniels (2018) built a Bayesian synthesis procedure.

$$P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{O}_{k-1}, R_k = 0] = P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{O}_{k-1}, R_k = 1]$$

Zhou, Little, Kalbfleisch et al. (2010)

- For individuals who share the same outcomes (observed or not) and same missingness pattern prior to a scheduled visit, the distribution of the outcome for those missing the visit is the same as the distribution of the outcome for those who attend the visit
- No global sensitivity analysis procedure was developed.

$$P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, \overleftarrow{R}_{k-1}, R_k = 0]$$
$$= P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, \overleftarrow{R}_{k-1}, R_k = 1]$$

Sadinle and Reiter (2017a); Shpitser (2016)

- For individuals who have the same outcomes (observed or not) and same missingness pattern prior to and after a scheduled visit, the distribution of the outcome for those missing the visit is the same as the distribution of the outcome for those who attend the visit.
- Cannot be represented as a directed acyclic graph.
- Sadinle and Reiter (2017a) developed a Bayesian global sensitivity analysis procedure anchored at this assumption.

$$
\begin{aligned}
&P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, \overleftarrow{R}_{k-1}, R_k = 0, \overrightarrow{Y}_k^{(1)}, \overrightarrow{R}_k] \\
&\quad = P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, \overleftarrow{R}_{k-1}, R_k = 1, \overrightarrow{Y}_k^{(1)}, \overrightarrow{R}_k]
\end{aligned}
$$

Robins (1997); Sadinle and Reiter (2017b)

- For individuals who share the same outcomes (observed or not) prior to a scheduled visit and the same observed data after the visit, the distribution of the outcome for those missing the visit is the same as the distribution of the outcome for those attending the visit.
- No global sensitivity analysis procedure was developed.

$$P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, R_k = 0, \overrightarrow{O}_k] \quad (1)$$

$$= P[Y_k^{(1)} = y_k^{(1)} | \overleftarrow{Y}_{k-1}^{(1)}, R_k = 1, \overrightarrow{O}_k] \quad (2)$$

In Other Words ...

- Imagine the stratum of people who share the same outcomes prior to visit k (*observed or not*) and who share the same *observed* outcomes after visit k.
  - Sub-stratum A: people who show up at visit k
  - Sub-stratum B: people who do not show up at visit k
- Probability of outcome at visit k is the same for those in sub-stratum A and those in sub-stratum B.

Sensitivity Analysis

$$P(Y_k^{(1)} = 1 | \underbrace{\overleftarrow{Y}_{k-1}^{(1)}, R_k = 0, \overrightarrow{O}_k}_{\text{Stratum B}})$$
$$\propto P(Y_k^{(1)} = 1 | \underbrace{\overleftarrow{Y}_{k-1}^{(1)}, R_k = 1, \overrightarrow{O}_k}_{\text{Stratum A}}) \exp(\alpha_k)$$

where α_k is the sensitivity analysis parameter.

- When $\alpha_k > 0$ (< 0), it is assumed that stratum B individuals are more (less) likely to have $Y_k^{(1)} = 1$ than stratum A individuals.
- As $\alpha_k \to \infty$ ($-\infty$), it is assumed that all individuals in stratum B have $Y_k^{(1)} = 1$ ($Y_k^{(1)} = 0$).
- Notice that when $\alpha_k = 0$ for all k, the benchmark assumption is obtained.

Directed Acyclic Graph (DAG)

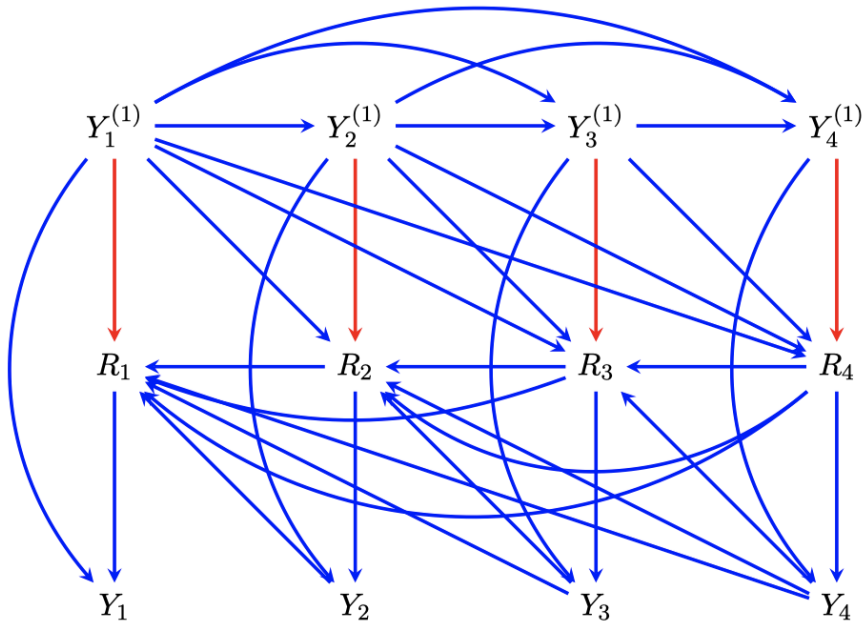


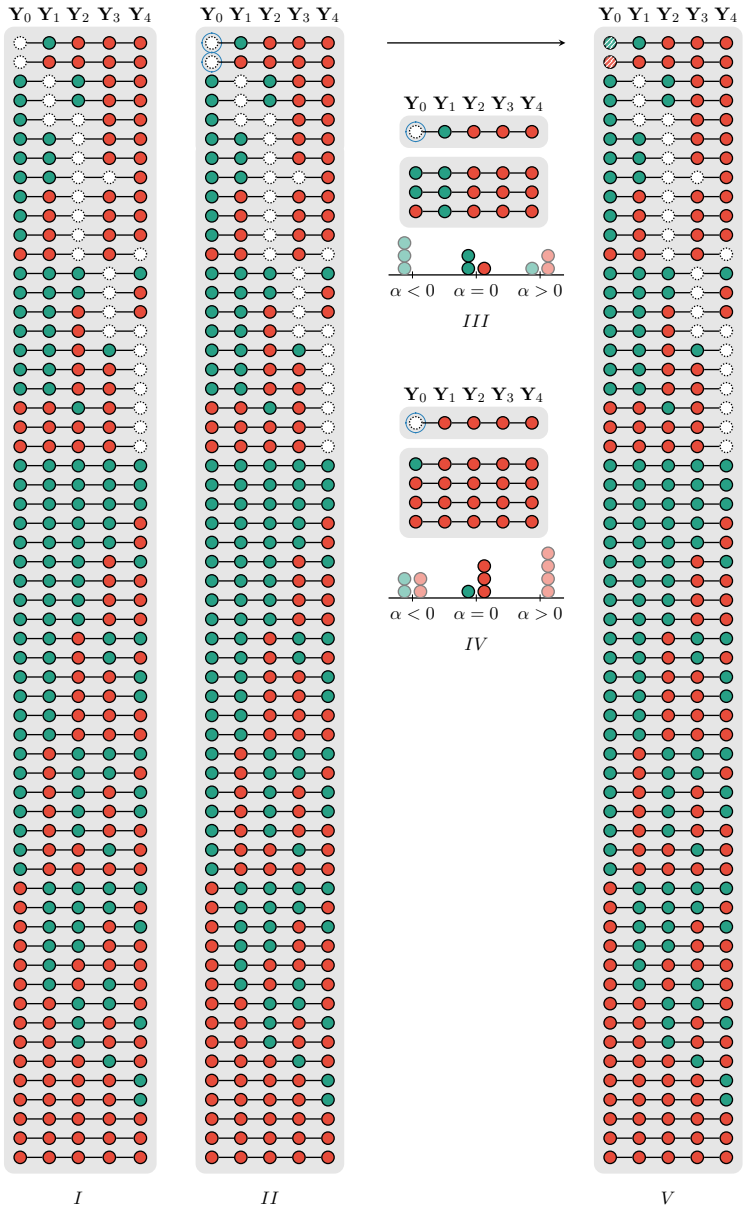
Figure 1: Directed acyclic graph representation (DAG) of (1) with $K = 4$. In this DAG, there are arrows into Y_k from $Y_k^{(1)}$ and R_k since Y_k is a deterministic function of these latter variables. The red arrow from $Y_k^{(1)}$ into R_k represents the dependence implied by the sensitivity analysis parameter α_k. The red arrows will be absent when $\alpha_k = 0$ for all k.

Theorem

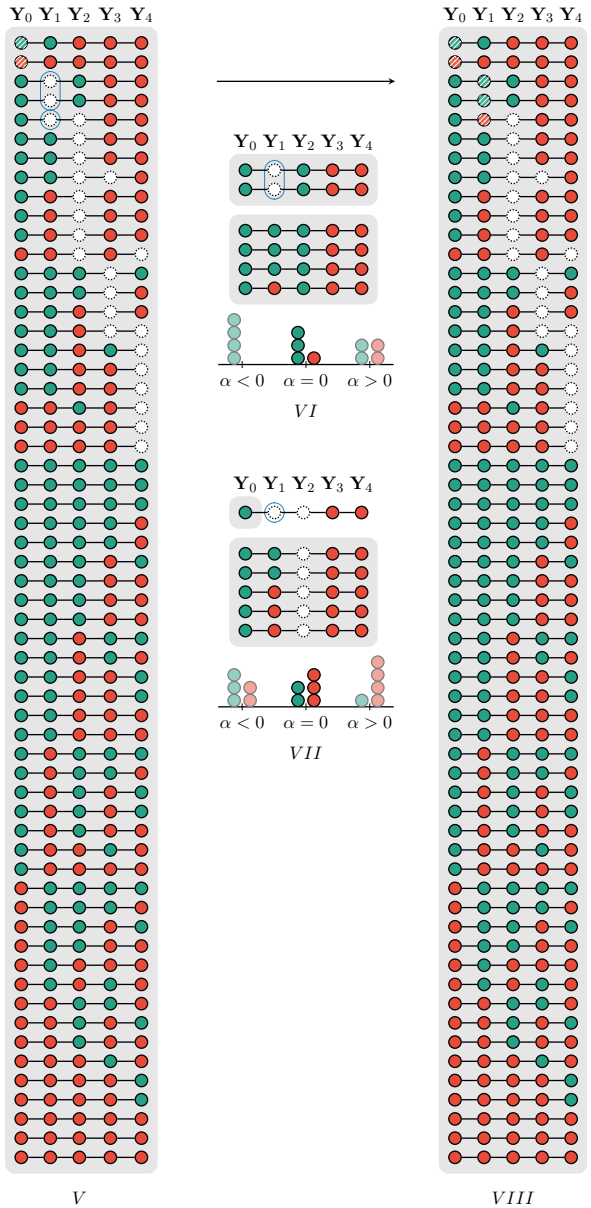
The distribution of $\overleftarrow{Y}_K^{(1)}$ is identified for specified $\overleftarrow{\alpha}_K$.

- Identification means that there is mathematical mapping from the distribution of the observed data $\overleftarrow{O}_K$ to the distribution of $\overleftarrow{Y}_K^{(1)}$.
- The proof follows by induction.

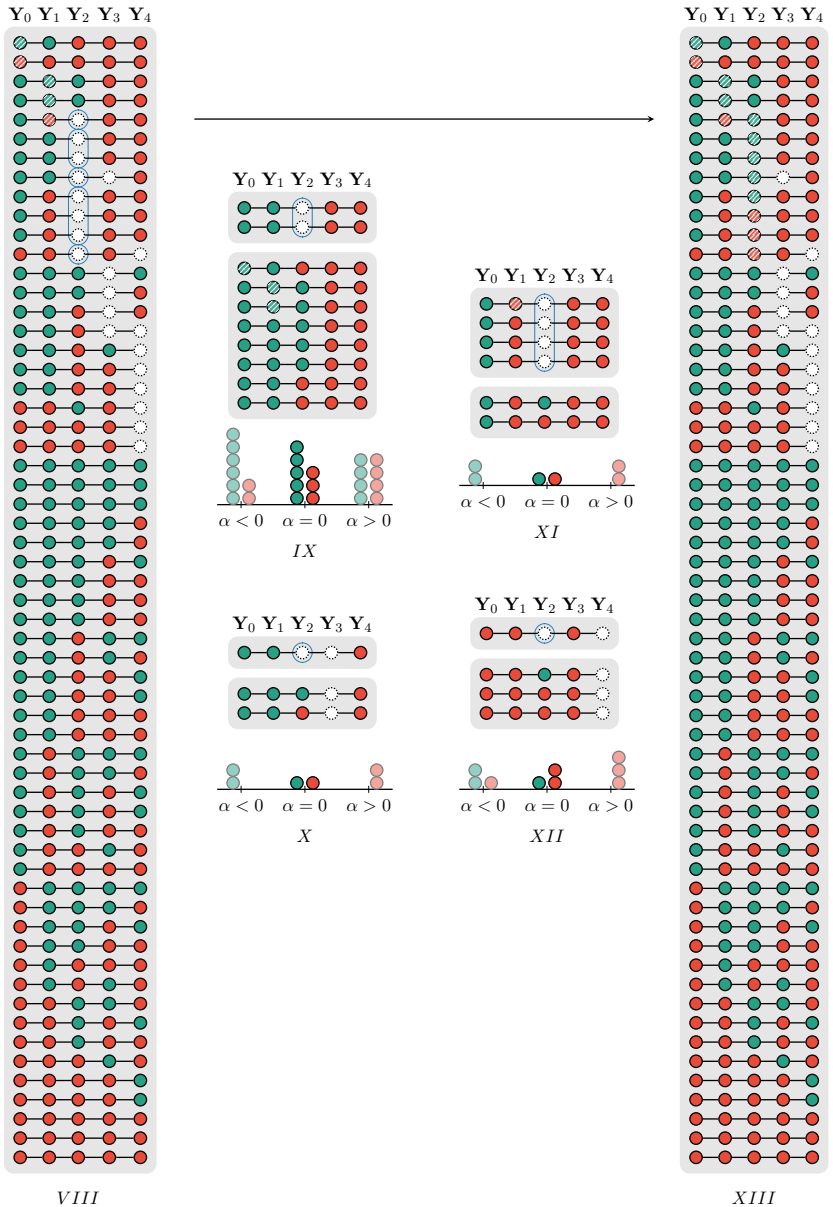
Identification



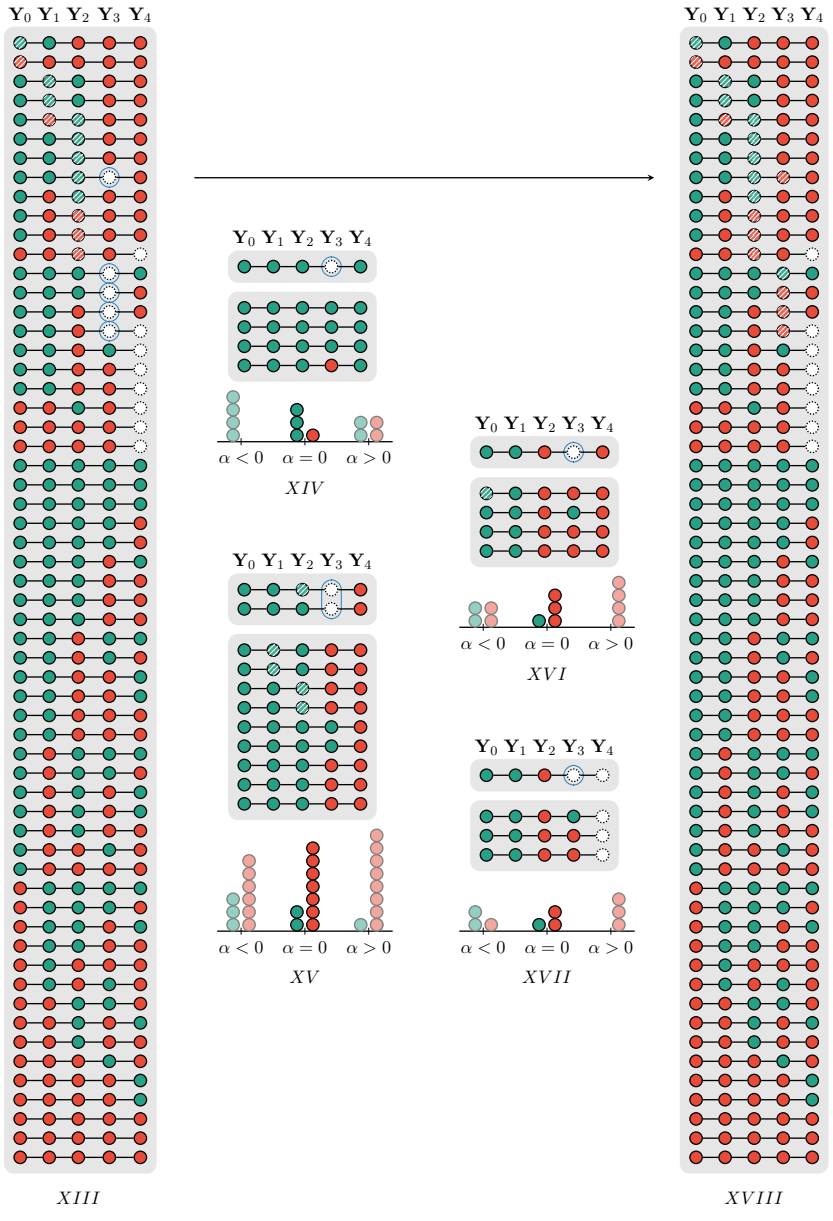
Identification



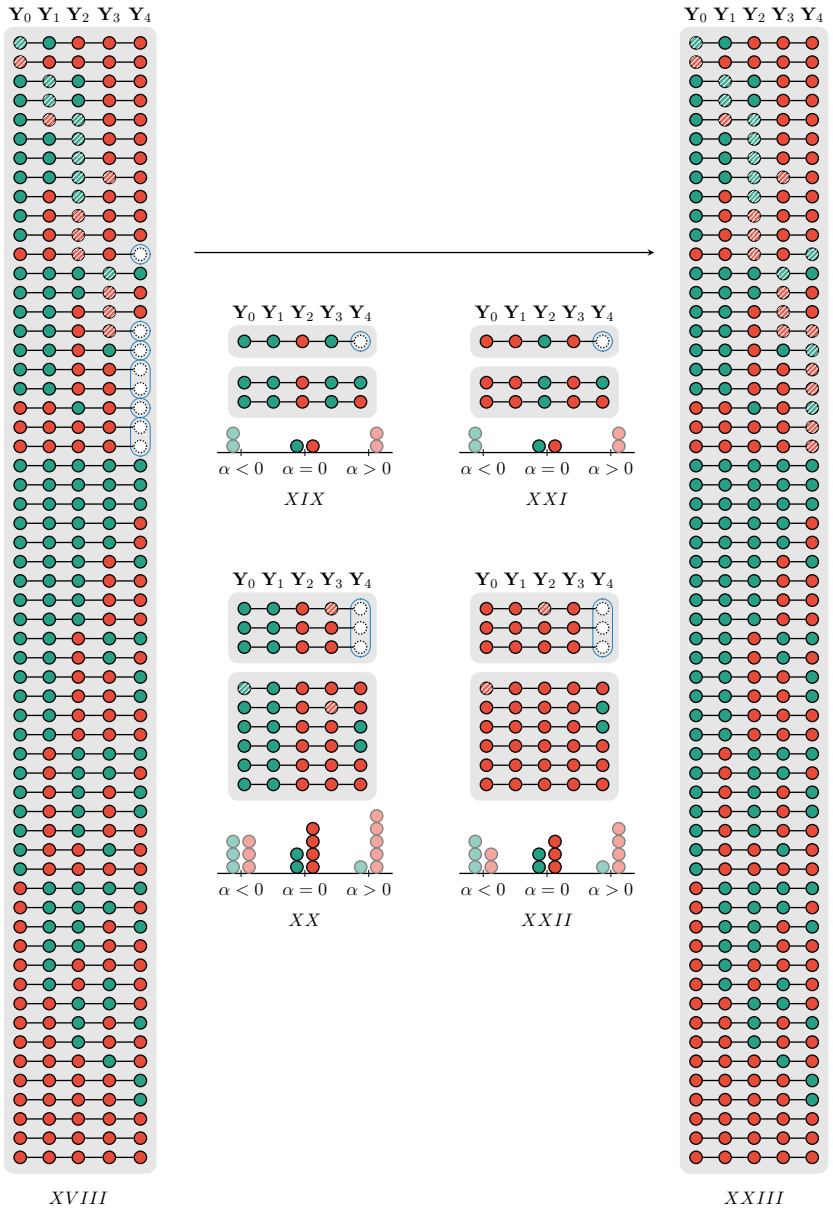
Identification



Identification



Identification



Modeling the Distribution of $\overleftarrow{O}_K$

$$P[\overleftarrow{O}_K = \overleftarrow{o}_K] = P[O_1 = o_1] \prod_{k=2}^{K} P[O_k = o_k | \overleftarrow{O}_{k-1} = \overleftarrow{o}_{k-1}]$$

- Model $P[O_k = o_k | \overleftarrow{O}_{k-1} = \overleftarrow{o}_{k-1}]$ using random forests.
- We show that the resulting estimator of the joint distribution of the observed data is consistent and asymptotically normal.
- Thus, the plug-in estimator of functional of interest will be consistent and asymptotically normal.

Random Forests

- The random forest algorithm is built on top of the classification and regression tree (CART) algorithm, which creates a risk prediction model by recursively partitioning the covariate space $\overleftarrow{O}_{k-1}$ using binary splits.
- With ternary outcomes (O_k), the decision to split is made by minimizing a measure of impurity (e.g., Gini impurity).
- For fully grown trees, splitting is continued until each terminal node has at most D observations, for a pre-determined integer D.
- To improve prediction accuracy, an ensemble algorithm, referred to as bagging, averages fully grown CART trees built using different bootstrap samples.
- To de-correlate the trees in the ensemble, the random forest algorithm modifies bagging by only considering a random subset of the covariates at each splitting decision.

Case Study: CTN -0044

- Two-arm randomized trial designed to evaluate a new approach to reducing substance use among patients entering outpatient addiction treatment.
- Treatment-as-usual (TAU) vs. treatment-as-usual plus a computerized therapeutic education system and contingent incentives (TAU+).
- TAU: individual and group counseling.
- TAU+: substituted 2 hours of usual care per week with computer-interactive modules covering skills for achieving and maintaining abstinence and prize-based motivational incentives contingent on abstinence and treatment adherence.
- Urine samples scheduled to be collected twice weekly.
- Outcome: number of negative urine samples during first 6 weeks.

Case Study: CTN -0044

- Among the 252 individuals randomized to TAU
  - 42 (16.7%) had a complete record of urine samples
  - 11 (4.4%) had no urine samples
  - 28 (11.1%) had at least one urine sample and a monotone missing data pattern
  - 171 (67.9%) individuals had an intermittent missing data pattern
- Among the 255 individuals randomized to TAU+
  - 81 (31.8%) had a complete record of urine samples,
  - 3 (1.2%) had no urine samples
  - 18 (7.1%) had at least one urine sample and a monotone missing data pattern
  - 153 (60.0%) individuals had an intermittent missing data pattern

Case Study: CTN -0044

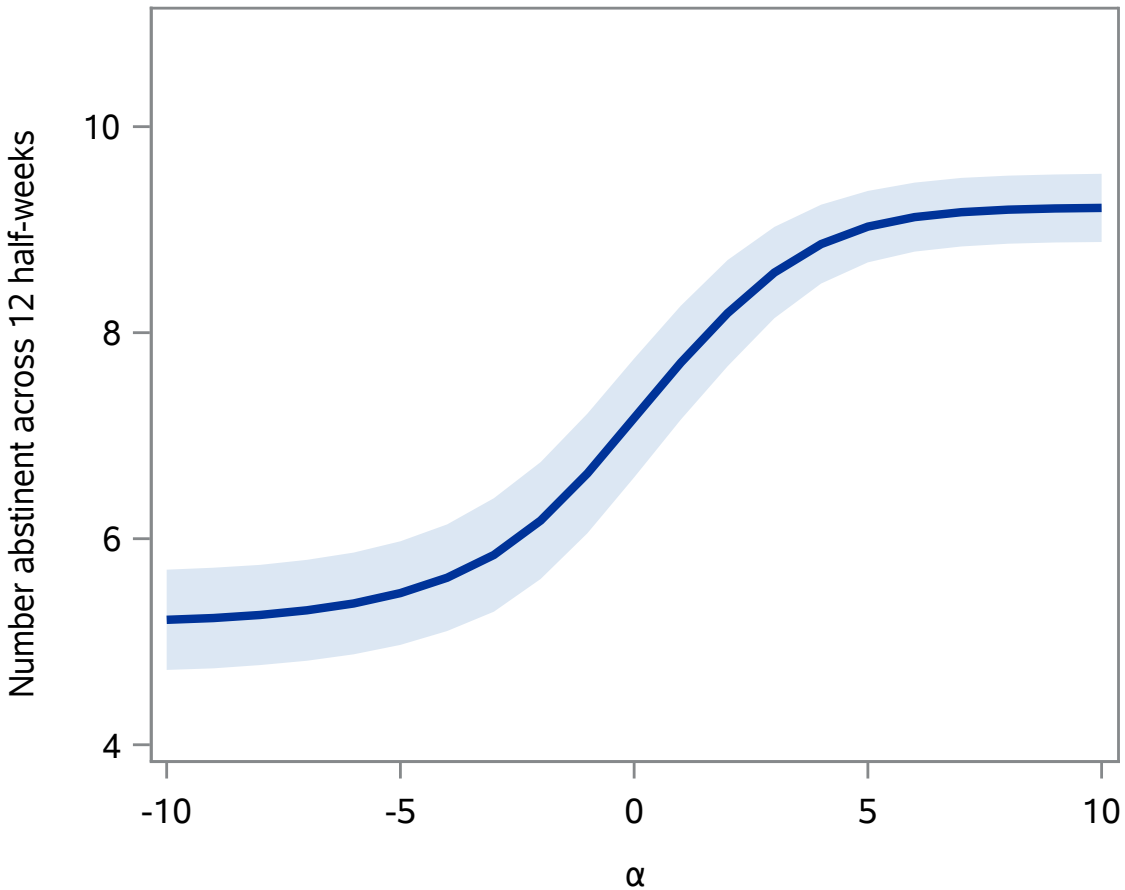
We first used the random forest algorithm to estimate the distribution of the observed data. We used 1000 trees.

- To evaluate the model fit, we compared empirical and model-based estimates of the joint distribution of the observed data at all 66 pairs of time points.
- For each pair, the joint distribution is represented by the cell probabilities of a three by three table.
- For each table, we computed the maximum of the absolute differences between the empirical and model-based estimates of the cell probabilities.
- The largest of these maximums over the 66 tables was 1.82%.
- In contrast, the largest of the maximums based on a first-order Markov model was 12.98%.

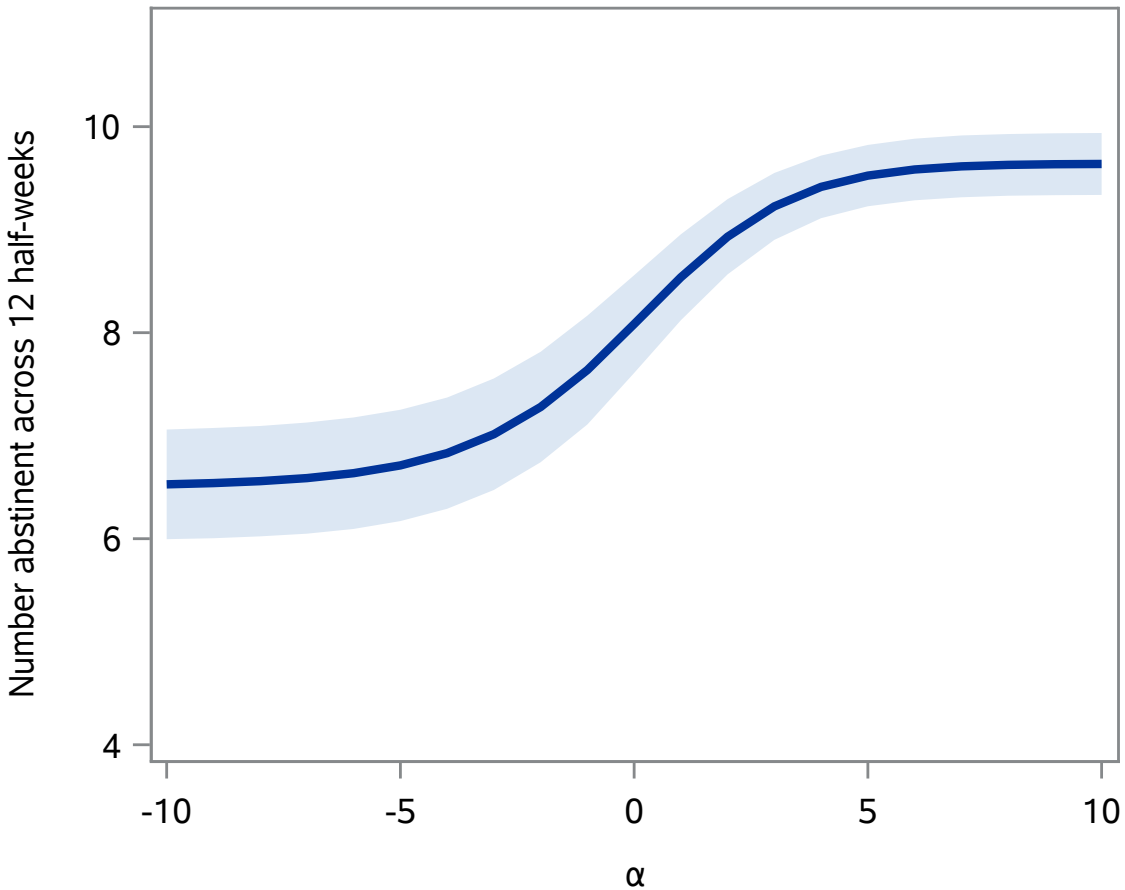
Case Study: CTN -0044

Assumption	TAU	TAU+	Difference
MCAR	7.86 (7.25, 8.47)	8.83 (8.28, 9.38)	0.97 (0.17, 1.76).
Missing=Non-Abstinent	5.14 (4.60, 5.69)	6.48 (5.90, 7.06)	1.34 (0.58, 2.10)
Missing=Abstinent	9.27 (8.87, 9.67)	9.64 (9.24, 10.04)	0.37 (-0.18, 0.92)
Benchmark	7.17 (6.60, 7.75)	8.08 (7.61, 8.56)	0.91 (0.06, 1.76)

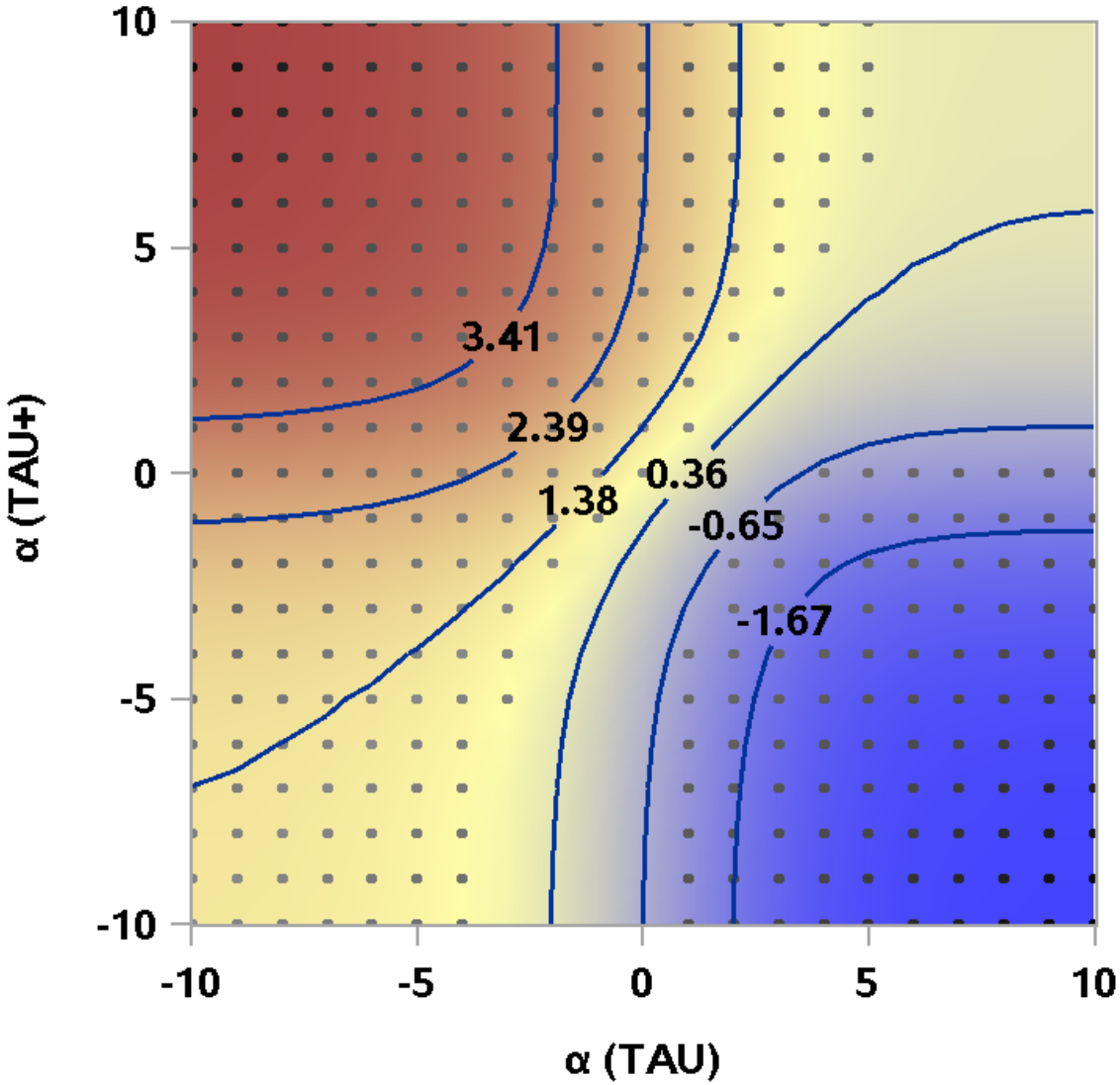
Case Study: CTN -0044



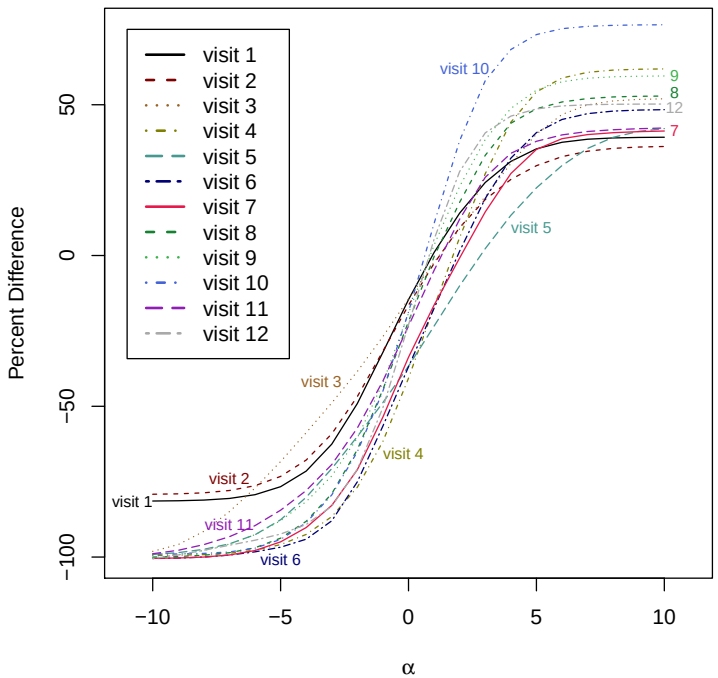
Case Study: CTN -0044

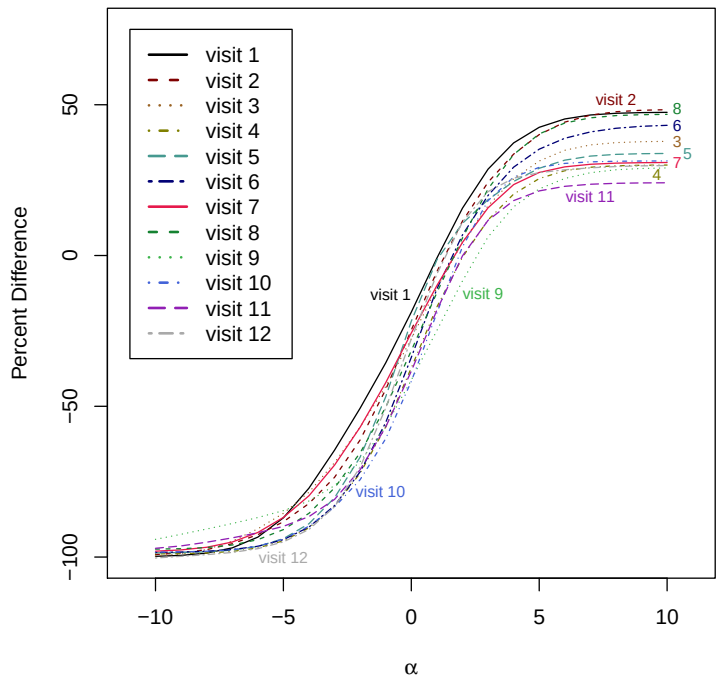


Case Study: CTN -0044



Case Study: CTN -0044





Simulation

	TAU					TAU+				
α	Truth	Mean	Std. Dev.	$\sqrt{\text{MSE}}$	Coverage	Truth	Mean	Std. Dev.	$\sqrt{\text{MSE}}$	Coverage
-5	5.47	5.51	0.282	0.286	0.954	6.71	6.78	0.306	0.313	0.942
-4	5.62	5.67	0.285	0.289	0.952	6.82	6.90	0.307	0.316	0.936
-3	5.84	5.90	0.289	0.295	0.954	7.00	7.08	0.307	0.316	0.936
-2	6.18	6.24	0.297	0.302	0.952	7.27	7.34	0.306	0.314	0.940
-1	6.63	6.68	0.306	0.310	0.952	7.64	7.69	0.301	0.305	0.936
0	7.18	7.22	0.309	0.311	0.954	8.10	8.10	0.292	0.292	0.938
1	7.72	7.76	0.298	0.301	0.948	8.54	8.52	0.278	0.279	0.934
2	8.20	8.23	0.279	0.281	0.944	8.92	8.89	0.259	0.261	0.938
3	8.59	8.60	0.261	0.261	0.946	9.21	9.16	0.240	0.244	0.938
4	8.86	8.84	0.248	0.248	0.948	9.40	9.35	0.225	0.230	0.942
5	9.03	9.00	0.238	0.240	0.942	9.50	9.47	0.215	0.218	0.946

Discussion

- Software has been developed
- Key computational limitation
  - Requires storage and operation on a 3^K vector of probabilities.
  - Computationally infeasible when $K > 15$.
  - Reduce dimension by introducing Markovian-type conditional independence restrictions.
- Extension to continuous outcomes.
  - Asymptotic theory for the random forest estimator is substantially more complex .
  - Different strategy will be needed, e.g., using an influence function-based approach.