Causal Effects of Multidimensional Exposures

Razieh Nabi, PhD

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

razieh.nabi@emory.edu

Joint work with Todd McNutt and Ilya Shpitser (JHU)

August 9, 2022

Outline

1. Single and multidimensional exposures/treatments
2. Sufficient dimension reduction
3. Causal sufficient dimension reduction
4. Simulations and data application

Single dimensional exposures

- Observed data $O = \{C, A, Y\} \sim P \in \mathcal{M}$
- $Y(a)$: potential outcome Y had A been assigned to a

Single dimensional exposures

- Observed data $O = \{C, A, Y\} \sim P \in \mathcal{M}$
- $Y(a)$: potential outcome Y had A been assigned to a
- **If A is binary:**

$$\text{ACE} = \mathbb{E}[Y(A=1)] - \mathbb{E}[Y(A=0)]$$

- Identifiability under standard assumptions:
  - Consistency: $Y = A \times Y(1) + (1 - A) \times Y(0)$,
  - (Conditional) Ignorability: $Y(a) \perp\!\!\!\perp A \mid C$,
  - Positivity: $p(A = a \mid C) > 0$.

$$\text{ACE} = \mathbb{E}\Big[\mathbb{E}[Y \mid A=1, C] - \mathbb{E}[Y \mid A=0, C]\Big]$$

- **If A is continuous:**
  - look at the entire dose-response relationship

Multidimensional exposures

- **Radiation oncology**
  - Head and neck cancer patients
  - Target volumes in close proximity to sensitive salivary glands
  - Radiation morbidities such as xerostomia affect quality of life
  - Such morbidities can lead to severe reduction in food intake
  - Minor variations in dose may improve secondary outcomes

Multidimensional exposures

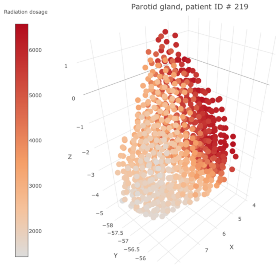
- **Radiation oncology**
  - Head and neck cancer patients
  - Target volumes in close proximity to sensitive salivary glands
  - Radiation morbidities such as xerostomia affect quality of life
  - Such morbidities can lead to severe reduction in food intake
  - Minor variations in dose may improve secondary outcomes
- Prospective studies for evaluating risk factors for weight loss (Cacicedo et al., 2014)
- **Cause-effect relation between radiation and weight loss**

Multidimensional exposures

- **Radiation oncology**
  - Head and neck cancer patients
  - Target volumes in close proximity to sensitive salivary glands
  - Radiation morbidities such as xerostomia affect quality of life
  - Such morbidities can lead to severe reduction in food intake
  - Minor variations in dose may improve secondary outcomes
- Prospective studies for evaluating risk factors for weight loss (Cacicedo et al., 2014)
- **Cause-effect relation between radiation and weight loss**
- Natural language processing: effects of high dimensional text data (Gentzkow et al., 2019), (Feder et al., 2022)
- Neuroimaging: neuronal network activity to cognitive processing and behavior (Ramsey et al., 2010), (Mather et al., 2013)

Radiation therapy as exposure

- Raw records of radiation:
  3D voxel maps of radiation doses on different glands



- Summarize the exact dose localization information
  - A multidimensional vector of radiation dosages
  - E.g., radiation dose on k^{th} percentile of the gland's volume, $k = 1, \ldots, 100$
- Even such summaries complicate establishing clinically relevant causal relationships

Dimension reduction of exposure

- **Assumption:** There exists a lower dimensional representation of A that preserves the effect of A on Y
- Let $g : A \in \mathbb{R}^p \to g(A) \in \mathbb{R}^d$, $d < p$ s.t., $\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a))]$
- Assume $g(.;\beta)$, e.g., $g(A;\beta) = \beta^T A$ where $\beta \in \mathbb{R}^{p\times d}$

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a;\beta))].$$

Dimension reduction of exposure

- **Assumption:** There exists a lower dimensional representation of A that preserves the effect of A on Y
- Let $g : A \in \mathbb{R}^p \to g(A) \in \mathbb{R}^d$, $d < p$ s.t., $\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a))]$
- Assume $g(.; \beta)$, e.g., $g(A; \beta) = \beta^T A$ where $\beta \in \mathbb{R}^{p \times d}$

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a; \beta))].$$

- Given n i.i.d. samples, how to estimate β?
  - ML methods (PCA and variants) fail because they ignore treatment-outcome relationships
  - Preserving associational relations between A and Y is insufficient
    - $\mathbb{E}[Y(a)] \neq \mathbb{E}[Y \mid A = a]$ due to the confounding bias

Dimension reduction of exposure

- **Assumption:** There exists a lower dimensional representation of A that preserves the effect of A on Y
- Let $g : A \in \mathbb{R}^p \to g(A) \in \mathbb{R}^d$, $d < p$ s.t., $\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a))]$
- Assume $g(.;\beta)$, e.g., $g(A;\beta) = \beta^T A$ where $\beta \in \mathbb{R}^{p \times d}$

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(g(a;\beta))].$$

- Given n i.i.d. samples, how to estimate β?
  - ML methods (PCA and variants) fail because they ignore treatment-outcome relationships
  - Preserving associational relations between A and Y is insufficient
    - $\mathbb{E}[Y(a)] \neq \mathbb{E}[Y \mid A = a]$ due to the confounding bias
- N, McNutt, and Shpitser, *Semiparametric causal sufficient dimension reduction of multidimensional treatments*, UAI 2022.
  - Functional form of g is known; assume $g(A;\beta) = \beta^T A, \beta \in \mathbb{R}^{p \times d}$

Sufficient dimension reduction (SDR)

- Notation: $Y \in \mathbb{R}, X \in \mathbb{R}^p, \beta \in \mathbb{R}^{p \times d}, d < p$.
- Central mean space model assumption:
  - Mean of Y relates to covariates X only through $\beta^T X$:

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \beta^T X]$$

or equivalently:

$$Y = m(\beta^T X) + \epsilon, \ \mathbb{E}(\epsilon \mid X) = 0,$$

where $m(\beta^T X)$ unspecified smooth function

- Central mean space $S_{\mathbb{E}(Y|X)}$: span of the columns in β
- Objective of SDR: estimate $S_{\mathbb{E}(Y|X)}$

Three broad approaches to SDR

1. Inverse regression based methods
   - Ordinary Least Squares (OLS), Principal Hessian Direction (PHD)
   - Rely on linearity condition and/or constant variance condition

2. Nonparametric estimation
   - Minimum Average Variance Estimation (MAVE)
   - Rely on continuity of each covariate
     (could be relaxed with more computational complexity)

3. **Semiparametric estimation**
   - Yanyuan Ma and Liping Zhu
     (A Semiparametric Approach to Dimension Reduction, JASA 2012)

A semiparametric approach to SDR

- $S_{\mathbb{E}(Y|X)}$ is a semiparametric model:

$$f_{X,Y}(X, Y; \beta, \eta) = \eta_1(X) \times \eta_2(Y, X; \beta),$$

  where $\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \beta^T X]$.

- β: parameter of interest, η_1, η_2: nuisance

A semiparametric approach to SDR

- $S_{\mathbb{E}(Y|X)}$ is a semiparametric model:

$$f_{X,Y}(X,Y;\beta,\eta) = \eta_1(X) \times \eta_2(Y,X;\beta),$$

  where $\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \beta^T X]$.

- β: parameter of interest, η_1, η_2: nuisance

- Regular Asymptotically Linear (RAL) estimator:

$$n^{1/2}(\widehat{\beta} - \beta) = n^{-1/2} \sum_{i=1}^{n} \phi(O_i) + o_p(1),$$

  where $\phi(O)$ is the influence function

- Estimate β by solving: $\sum_{i=1}^{n} \phi(O_i) = 0$
- $\widehat{\beta}$ is asymptotically normal and $\sqrt{n}\widehat{\beta} = \mathbb{E}[\phi(O)\phi^T(O)]$

A semiparametric approach to SDR

- $S_{\mathbb{E}(Y|X)}$ is a semiparametric model:

$$f_{X,Y}(X, Y; \beta, \eta) = \eta_1(X) \times \eta_2(Y, X; \beta),$$

  where $\mathbb{E}[Y \mid X] = \mathbb{E}[Y \mid \beta^T X]$.

- β: parameter of interest, η_1, η_2: nuisance

- Regular Asymptotically Linear (RAL) estimator:

$$n^{1/2}(\widehat{\beta} - \beta) = n^{-1/2} \sum_{i=1}^{n} \phi(O_i) + o_p(1),$$

  where $\phi(O)$ is the influence function

- Estimate β by solving: $\sum_{i=1}^{n} \phi(O_i) = 0$
- $\widehat{\beta}$ is asymptotically normal and $\sqrt{n}\widehat{\beta} = \mathbb{E}[\phi(O)\phi^T(O)]$

- Find influence function $\rightarrow$ find RAL estimator

The class of influence functions

- The orthogonal complement of the nuisance tangent space: (Ma and Zhu, 2012)

$$\Lambda_\eta^\perp = \Big\{\big(Y - \mathbb{E}[Y \mid \beta^T X]\big) \times \big(\alpha(X) - \mathbb{E}[\alpha \mid \beta^T X]\big)\Big\}$$

  (i.e., the class of all influence functions)

- Given $\alpha(X)$, estimate β by solving the sample version of:

$$\mathbb{E}\Big[\big(Y - \mathbb{E}[Y \mid \beta^T X]\big) \times \big(\alpha(X) - \mathbb{E}[\alpha \mid \beta^T X]\big)\Big] = 0.$$

- The above estimator has a **double robustness property**:
  - If misspecify either $\mathbb{E}[Y \mid \beta^T X]$ or $\mathbb{E}[\alpha \mid \beta^T X]$, the resulting estimator remains consistent

Marginal structural models

- If exposure A is randomized:

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(\beta^T a)] \longrightarrow \mathbb{E}[Y \mid a] = \mathbb{E}[Y \mid \beta^T a]$$

i.e., estimate β using the semiparametric approach to SDR

Marginal structural models

- If exposure A is randomized:

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(\beta^T a)] \;\longrightarrow\; \mathbb{E}[Y \mid a] = \mathbb{E}[Y \mid \beta^T a]$$

  i.e., estimate β using the semiparametric approach to SDR

- If A is not randomized, but conditional ignorability still holds:
  - Use **marginal structural models** (MSMs) to estimate β (Robins, 1999)

Marginal structural models

- If exposure A is randomized:

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(\beta^T a)] \longrightarrow \mathbb{E}[Y \mid a] = \mathbb{E}[Y \mid \beta^T a]$$

i.e., estimate β using the semiparametric approach to SDR

- If A is not randomized, but conditional ignorability still holds:
  - Use **marginal structural models** (MSMs) to estimate β (Robins, 1999)

- Assume MSM $\mathbb{E}[Y(a)] = f(a; \beta)$, an IPW estimator of β:

$$\mathbb{P}_n\Big[\frac{P^*(a)}{W_a(C; \widehat{\eta}_a)} \times \big(Y - f(a; \beta)\big)\Big] = 0,$$

where $W_a(C; \widehat{\eta}_a) := P(A = a \mid C)$, and $\mathbb{P}_n[.] = \sum_{i=1}^{n}[.]$

Marginal structural models

- If exposure A is randomized:

$$\mathbb{E}[Y(a)] = \mathbb{E}[Y(\beta^T a)] \longrightarrow \mathbb{E}[Y \mid a] = \mathbb{E}[Y \mid \beta^T a]$$

  i.e., estimate β using the semiparametric approach to SDR

- If A is not randomized, but conditional ignorability still holds:
  - Use **marginal structural models** (MSMs) to estimate β (Robins, 1999)

- Assume MSM $\mathbb{E}[Y(a)] = f(a;\beta)$, an IPW estimator of β:

$$\mathbb{P}_n\Big[\frac{P^*(a)}{W_a(C;\widehat{\eta}_a)} \times \big(Y - f(a;\beta)\big)\Big] = 0,$$

  where $W_a(C;\widehat{\eta}_a) \coloneqq P(A = a \mid C)$, and $\mathbb{P}_n[.] = \sum_{i=1}^n [.]$

- MSM for a binary treatment: $\mathbb{E}[Y(a)] = \beta_0 + \beta_a \times a$
  - ACE = β_a
  - Can interpret an MSM as a “causal regression"

Causal sufficient dimension reduction (Causal SDR)

- MSM and causal regressions:

$$\mathbb{E}_q[Y \mid a] = f(a; \beta),$$

where $q(C, A, Y) = P(C) \times P^*(A) \times P(Y \mid A, C)$

- MSM in our setup:

$$\mathbb{E}_q[Y \mid a] = \mathbb{E}_q[Y \mid \beta^T a]$$

Causal sufficient dimension reduction (Causal SDR)

- MSM and causal regressions:

$$\mathbb{E}_q[Y \mid a] = f(a; \beta),$$

where $q(C, A, Y) = P(C) \times P^*(A) \times P(Y \mid A, C)$

- MSM in our setup:

$$\mathbb{E}_q[Y \mid a] = \mathbb{E}_q[Y \mid \beta^T a]$$

- An IPW estimator for β:

$$\mathbb{P}_n\Big[\frac{P^*(a)}{W_a(C; \widehat{\eta}_a)} \times \big(Y - \mathbb{E}[Y \mid \beta^T A]\big)\Big] = 0,$$

where $W_a(C; \widehat{\eta}_a) \coloneqq P(A = a \mid C)$.

Causal SDR ctd.

- A more efficient approach to MSM is to use influence functions
  - Derive RAL estimators based on deriving the $\Lambda_\eta^\perp$
- $\Lambda_\eta^\perp$ that satisfies $\mathbb{E}_q[Y \mid a] = \mathbb{E}_q[Y \mid \beta^T a]$:

$$\Lambda_\eta^\perp = \Big\{ \frac{U(O;\beta)}{W_a(C)} - \phi(A,C) + \mathbb{E}[\phi(A,C) \mid C] \Big\},$$

where $U(O;\beta) = \big(Y - \mathbb{E}[Y \mid \beta^T A]\big) \times \big(\alpha(A) - \mathbb{E}[\alpha \mid \beta^T A]\big)$

Causal SDR ctd.

- A more efficient approach to MSM is to use influence functions
  - Derive RAL estimators based on deriving the $\Lambda_\eta^\perp$
- $\Lambda_\eta^\perp$ that satisfies $\mathbb{E}_q[Y \mid a] = \mathbb{E}_q[Y \mid \beta^T a]$:

$$\Lambda_\eta^\perp = \Big\{ \frac{U(O;\beta)}{W_a(C)} - \phi(A,C) + \mathbb{E}[\phi(A,C) \mid C] \Big\},$$

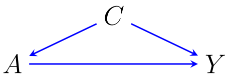
where $U(O;\beta) = \big(Y - \mathbb{E}[Y \mid \beta^T A]\big) \times \big(\alpha(A) - \mathbb{E}[\alpha \mid \beta^T A]\big)$

- For a fixed $\alpha(A)$, the most efficient estimator in this class:

$$\phi^{\text{opt}}(A,C) = \mathbb{E}\Big[\frac{U(O;\beta)}{W_a(C)} \; \Big| \; A, C \Big].$$

- **Robustness property:** the estimator for β is consistent if one of $P(A|C)$, $\mathbb{E}[U(O;\beta)|A,C]$, and one of $\mathbb{E}[Y \mid \beta^T A]$, $\mathbb{E}[\alpha(A) \mid \beta^T A]$ is correctly specified.

Simulations



- C: four confounders
  - generated from a standard multivariate normal distribution
- T: treatment dimension $p = 6, 12$
  - **Case 1.** linearity and constant covariance conditions are violated
  - **Case 2.** these assumptions are satisfied
- Y: generated from a normal distribution
  - causal structure $d = 2$
- Estimate β such that $\mathbb{E}[Y(a)] = \mathbb{E}[Y(\beta^T a)]$

Simulations ctd.

- Comparing different estimation strategies for β

1. **Reg SDR:** ignoring the confounding issue

$$\mathbb{P}_n\Big[\big(Y - \mathbb{E}[Y \mid \beta^T A]\big) \times \big(\alpha(A) - \mathbb{E}[\alpha \mid \beta^T A]\big)\Big] = 0$$

2. **IPW:** (ignore robustness; not efficient)

$$\mathbb{P}_n\Big[\frac{P^*(a)}{W_a(C;\widehat{\eta}_a)} \times \big(Y - \mathbb{E}[Y \mid \beta^T A]\big)\Big] = 0,$$

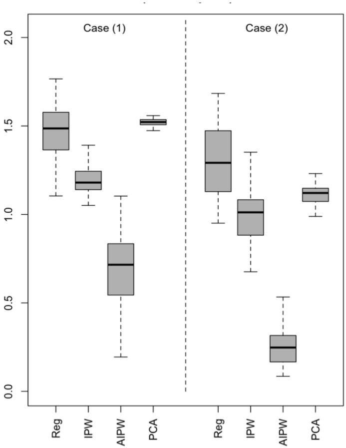
3. **AIPW:** influence-function based estimator (RAL)

$$\mathbb{P}_n\Big[\frac{U(\beta)}{W_a(C)} - \phi(A,C) + \mathbb{E}[\phi(A,C) \mid C]\Big] = 0$$

4. **PCA**: find $d = 2$ principal components of A and ignore Y

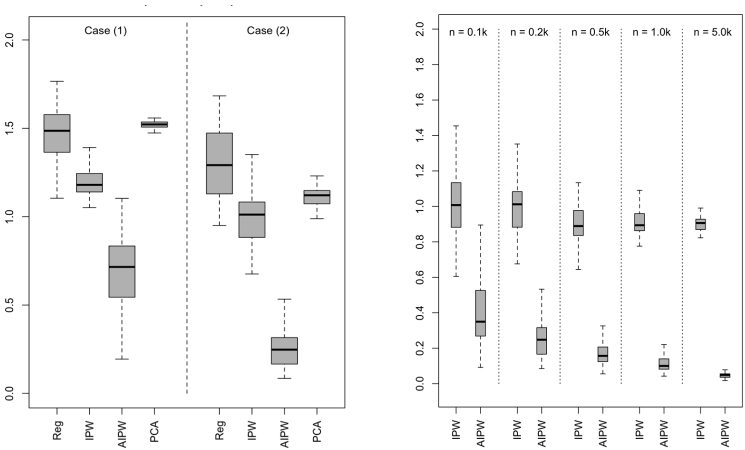
Simulations ctd.

- Comparing different estimation strategies for β ($p = 6, n = 200$)
  - Frobenius norm between true β and estimated β, (50 replications)



Simulations ctd.

- Comparing different estimation strategies for β ($p = 6, n = 200$)
  - Frobenius norm between true β and estimated β, (50 replications)



Data application

- A cohort of patients treated with radiation therapy for head and neck cancer
- Cohort consists of 613 patients who received radiation therapy at the Hopkins hospital prior to 2016
- **Exposure A:** a vector of radiation doses on the parotid glands
  - Summary measures from cumulative dose-volume histograms
  - In particular, 5 equally spaced percentages of volume
- **Outcome Y:** weight loss
  - difference between weight measured within 100 to 160 days after the completion of treatment and the weight measured during consultation before the start of treatment.
- **Confounders C:** age, sex, race, and baseline clinical factors: feeding tubes, chemotherapy before radiation

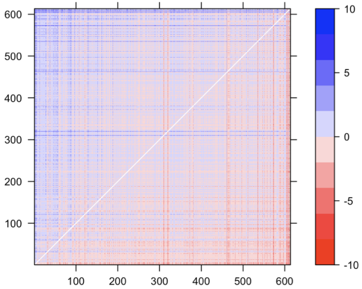
`https://github.com/raziehna/multidimensional-treatments`

Data application ctd.

- Structural dimension $d = 1$, linear mapping, AIPW estimator
  - mean dose to the parotid glands strongly correlates with risk factors of weight loss (Deasy et al., 2010)

- $n \times n$ heatmap
- Radiation doses sorted in increasing values along x, y
- $(k, i)^{\text{th}}$ coordinate:

$$\mathbb{E}[Y(g(a_k; \beta)) - Y(g(a_i; \beta))]$$

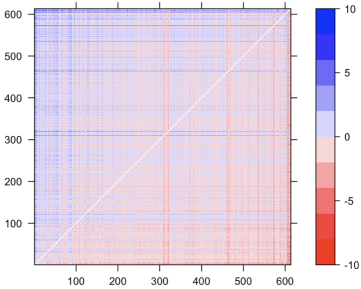


Data application ctd.

- Structural dimension $d = 1$, linear mapping, AIPW estimator
  - mean dose to the parotid glands strongly correlates with risk factors of weight loss (Deasy et al., 2010)

- $n \times n$ heatmap
- Radiation doses sorted in increasing values along x, y
- $(k, i)^{\text{th}}$ coordinate:

$$\mathbb{E}[Y(g(a_k; \beta)) - Y(g(a_i; \beta))]$$



- If $k > i$, then a red dot at (k, i) coordinate implies that an increase in radiation doses leads to an increase in weight loss
- As amount of radiation increases, severity of weight loss increases
- **Radiation therapy is a (potential) cause of weight loss in cancer patients**

Future directions

- Choosing the structural dimension d (besides heuristic ways)
- Extensions to nonparametric mappings: $\beta^T X \to u(X)$
- Incorporating sparsity techniques within the semiparametric framework to deal with higher dimensional treatments
  - Hypothesis: current methods hold for $p = o(n^{1/2})$
- Causal inference in longitudinal studies where multiple time points render a collection of treatments a multidimensional object
- Generalization to high-dimensional confounders, mediators, outcomes

Semiparametric Causal Sufficient Dimension Reduction of Multidimensional Treatments

Razieh Nabi[1] **Todd McNutt**[2] **Ilya Shpitser**[3]

[1]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA
[2]School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[3]Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA

Razieh Nabi, PhD (she/her/hers)
Rollins Assistant Professor
Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University

@raziehnabi
razieh.nabi@emory.edu
https://raziehnabi.com