Parameterising and inferring the effect of a continuous exposure using average derivative effects

Oliver Hines
Joint work with Stijn Vansteelandt and Karla Diaz-Ordaz

EuroCIM 19 May 2021

Plan

1. Introduction
2. Derivative effect results
3. Inference

The problem

- Causal effect estimands should be non-parametrically defined
- Binary exposures: ATE, $E(Y^1 - Y^0)$
- Continuous exposures: the dose-response function[1] $\theta(x) = E(Y^x)$.
- We offer an alternative generalization based on derivatives.

[1] Kennedy et al. (2017); Neugebauer and van der Laan (2007); Robins et al. (2001)

The problem

- Causal effect estimands should be non-parametrically defined
- Binary exposures: ATE, $E(Y^1 - Y^0)$
- Continuous exposures: the dose-response function[1] $\theta(x) = E(Y^x)$.
- We offer an alternative generalization based on derivatives.

[1] Kennedy et al. (2017); Neugebauer and van der Laan (2007); Robins et al. (2001)

The problem

- Causal effect estimands should be non-parametrically defined
- Binary exposures: ATE, $E(Y^1 - Y^0)$
- Continuous exposures: the dose-response function[1] $\theta(x) = E(Y^x)$.
- We offer an alternative generalization based on derivatives.

[1] Kennedy et al. (2017); Neugebauer and van der Laan (2007); Robins et al. (2001)

The problem

- Causal effect estimands should be non-parametrically defined
- Binary exposures: ATE, $E(Y^1 - Y^0)$
- Continuous exposures: the dose-response function[1] $\theta(x) = E(Y^x)$.
- We offer an alternative generalization based on derivatives.

[1] Kennedy et al. (2017); Neugebauer and van der Laan (2007); Robins et al. (2001)

Motivating example I

- How does time between two vaccine doses affect blood antibody concentration at 12 months?
- Outcome, Y: blood antibody measurement 12 months after initial dose.
- Continuous exposure, X: Time in days between vaccination doses.
- Single confounder, Z: Indicator of 'risk' e.g. underlying health condition.
- Define potential outcomes, Y^x with standard causal assumptions.

Motivating example I

- How does time between two vaccine doses affect blood antibody concentration at 12 months?
- Outcome, Y: blood antibody measurement 12 months after initial dose.
- Continuous exposure, X: Time in days between vaccination doses.
- Single confounder, Z: Indicator of ‘risk’ e.g. underlying health condition.
- Define potential outcomes, Y^x with standard causal assumptions.

Motivating example I

- How does time between two vaccine doses affect blood antibody concentration at 12 months?
- Outcome, Y: blood antibody measurement 12 months after initial dose.
- Continuous exposure, X: Time in days between vaccination doses.
- Single confounder, Z: Indicator of ‘risk’ e.g. underlying health condition.
- Define potential outcomes, Y^x with standard causal assumptions.

Motivating example II

- How does BMI age 50 affect probability of cardiac event by age 60?
- Outcome, Y: Event indicator.
- Continuous exposure, X: BMI.
- All sorts of socioeconomic and genetic confounders!
- Define potential outcomes, Y^x with standard causal assumptions.

Motivating example II

- How does BMI age 50 affect probability of cardiac event by age 60?
- Outcome, Y: Event indicator.
- Continuous exposure, X: BMI.
- All sorts of socioeconomic and genetic confounders!
- Define potential outcomes, Y^x with standard causal assumptions.

Motivating example II

- How does BMI age 50 affect probability of cardiac event by age 60?
- Outcome, Y: Event indicator.
- Continuous exposure, X: BMI.
- All sorts of socioeconomic and genetic confounders!
- Define potential outcomes, Y^x with standard causal assumptions.

Dose-response curve

- Dose-response function: What would happen if everyone had exposure level x?
- Does this make sense for e.g. BMI? Requires significant **extrapolation**.
- Could be **uninformative** in practice.
- How do we summarize the dose response function?
- Identified by $\theta(x) = E\{m(x, Z)\}$ where $m(x, z) = E(Y|X = x, Z = z)$
- Difficult to estimate[2]

[2] (Kennedy et al., 2017) also keynote speech today

Dose-response curve

- Dose-response function: What would happen if everyone had exposure level x?
- Does this make sense for e.g. BMI? Requires significant **extrapolation**.
- Could be **uninformative** in practice.
- How do we summarize the dose response function?
- Identified by $\theta(x) = E\{m(x, Z)\}$ where $m(x, z) = E(Y|X = x, Z = z)$
- Difficult to estimate[2]

[2] (Kennedy et al., 2017) also keynote speech today

Dose-response curve

- Dose-response function: What would happen if everyone had exposure level x?
- Does this make sense for e.g. BMI? Requires significant **extrapolation**.
- Could be **uninformative** in practice.
- How do we summarize the dose response function?
- Identified by $\theta(x) = E\{m(x, Z)\}$ where $m(x, z) = E(Y|X = x, Z = z)$
- Difficult to estimate[2]

[2] (Kennedy et al., 2017) also keynote speech today

Derivatives effect Proposal

- Derivative effects: Generally speaking should we increase or decrease exposure?
- Captures the notion of changes in exposure around realistic values.
- Delivers a single summary statistic rather than a curve.
- Identified by, e.g. the ADE: $E\{m'(X,Z)\}$
- Derivatives usually generalize finite differences such as, $m(1,z) - m(0,z)$
- We relate to and extend classical derivative effects literature[3].

[3] (Härdle and Stoker, 1989; Powell et al., 1989; Newey and Stoker, 1993)

Derivatives effect Proposal

- Derivative effects: Generally speaking should we increase or decrease exposure?
- Captures the notion of changes in exposure around realistic values.
- Delivers a single summary statistic rather than a curve.
- Identified by, e.g. the ADE: $E\{m'(X,Z)\}$
- Derivatives usually generalize finite differences such as, $m(1,z) - m(0,z)$
- We relate to and extend classical derivative effects literature[3].

[3] (Härdle and Stoker, 1989; Powell et al., 1989; Newey and Stoker, 1993)

Illustrative Example I

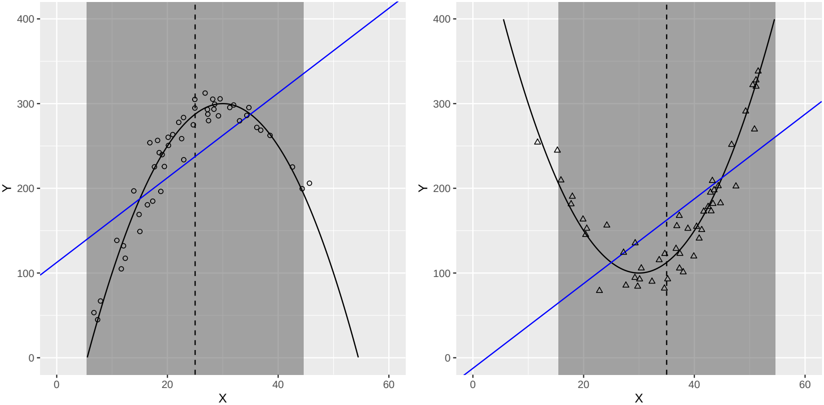

Figure: $m(x, z)$ for $z = 0, 1$ respectively. ADE shown by gradient of line in blue for $X \sim \mathcal{N}(\mu, \sigma)$ with mean and quantiles marked.

Illustrative Example II

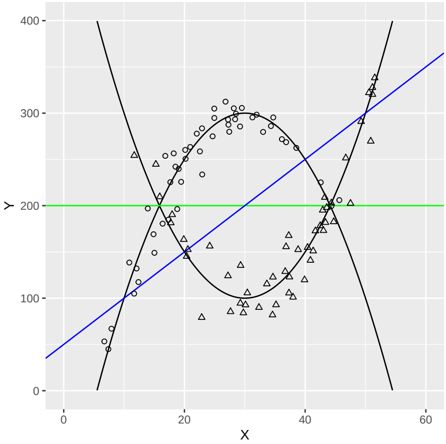

Figure: ADE (blue) and Dose-response curve (green)

Causal interpretation of derivative effects

- Our interpretation relies on the identification result

$$m'(x,z) = \lim_{\epsilon \to 0} \frac{E\{Y^{x+\epsilon} - Y^x | Z = z\}}{\epsilon}$$

- We further consider the weighted generalizations with **exposure** and **subgroup** weights

$$E\{w(Z)w(X|Z)m'(X,Z)\} \leftrightarrow \text{Weighted ATE}$$
$$E\{w(X|Z)m'(X,Z)|Z=z\} \leftrightarrow \text{Heterogeneous TE}$$

- Stochastic intervention[4] interpretation of weights.

[4] Díaz and van der Laan (2012); Kennedy (2019)

Causal interpretation of derivative effects

- Our interpretation relies on the identification result

$$m'(x,z) = \lim_{\epsilon \to 0} \frac{E\{Y^{x+\epsilon} - Y^{x} | Z = z\}}{\epsilon}$$

- We further consider the weighted generalizations with **exposure** and **subgroup** weights

$$E\{w(Z)w(X|Z)m'(X,Z)\} \leftrightarrow \text{Weighted ATE}$$
$$E\{w(X|Z)m'(X,Z)|Z=z\} \leftrightarrow \text{Heterogeneous TE}$$

- Stochastic intervention[4] interpretation of weights.

[4] Díaz and van der Laan (2012); Kennedy (2019)

Causal interpretation of derivative effects

- Our interpretation relies on the identification result

$$m'(x,z) = \lim_{\epsilon \to 0} \frac{E\{Y^{x+\epsilon} - Y^{x} | Z = z\}}{\epsilon}$$

- We further consider the weighted generalizations with **exposure** and **subgroup** weights

$$E\{w(Z)w(X|Z)m'(X,Z)\} \leftrightarrow \text{Weighted ATE}$$
$$E\{w(X|Z)m'(X,Z)|Z = z\} \leftrightarrow \text{Heterogeneous TE}$$

- Stochastic intervention[4] interpretation of weights.

[4] Díaz and van der Laan (2012); Kennedy (2019)

Plan

Motivation

- We can rewrite the ATE in an IPW-like form

$$E\left\{\frac{X-\pi(Z)}{\pi(Z)\{1-\pi(Z)\}}Y\right\}$$

- Likewise the ADE can be rewritten[5]

$$E\left\{\frac{-d\log(f(X|Z))}{dx}Y\right\}$$

- We call these red terms **conditional contrast functions** as they share similar properties.

[5] Under regularity conditions (Powell et al., 1989)

Motivation

- We can rewrite the ATE in an IPW-like form

$$E\left\{\frac{X-\pi(Z)}{\pi(Z)\{1-\pi(Z)\}}Y\right\}$$

- Likewise the ADE can be rewritten[5]

$$E\left\{\frac{-d\log(f(X|Z))}{dx}Y\right\}$$

- We call these red terms **conditional contrast functions** as they share similar properties.

[5] Under regularity conditions (Powell et al., 1989)

Motivation

- We can rewrite the ATE in an IPW-like form

$$E\left\{\frac{X-\pi(Z)}{\pi(Z)\{1-\pi(Z)\}}Y\right\}$$

- Likewise the ADE can be rewritten[5]

$$E\left\{\frac{-d\log(f(X|Z))}{dx}Y\right\}$$

- We call these red terms **conditional contrast functions** as they share similar properties.

[5] Under regularity conditions (Powell et al., 1989)

Contrast effects

- We propose new effect estimands which also have this property e.g.

$$\Lambda = E\left\{\frac{\text{Cov}(X, Y|Z)}{\text{Var}(X|Z)}\right\} = E\left\{\frac{X - \pi(Z)}{\text{Var}(X|Z)}Y\right\}$$

- This reduces to the ATE for binary X and to the ADE under differentiability and normality.
- But, this estimand is **better-defined** and **eaiser to estimate** than the ADE (and the dose response curve).

Contrast effects

- We propose new effect estimands which also have this property e.g.

$$\Lambda = E\left\{\frac{\mathsf{Cov}(X, Y|Z)}{\mathsf{Var}(X|Z)}\right\} = E\left\{\frac{X - \pi(Z)}{\mathsf{Var}(X|Z)}Y\right\}$$

- This reduces to the ATE for binary X and to the ADE under differentiability and normality.
- But, this estimand is **better-defined** and **eaiser to estimate** than the ADE (and the dose response curve).

A general theory

- We consider a general theory for estimands of the form

$$E\left[w(Z)l(X|Z)Y\right]$$

- Contrast function: $E\{l(X|Z)|z\} = 0$ and $E\{l(X|Z)X|z\} = 1$.
- Why would we do this? Reduce to weighted ATE when X is binary and to a weighted ADE when X is continuous.
- However, usually **better-defined** and **eaiser to estimate** than the ADE.

A general theory

- We consider a general theory for estimands of the form

$$E\left[w(Z)l(X|Z)Y\right]$$

- Contrast function: $E\{l(X|Z)|z\} = 0$ and $E\{l(X|Z)X|z\} = 1$.
- Why would we do this? Reduce to weighted ATE when X is binary and to a weighted ADE when X is continuous.
- However, usually **better-defined** and **eaiser to estimate** than the ADE.

Plan

1. Introduction
2. Derivative effect results
3. Inference

Marginal Effects

- Consider estimators constructed from the Efficient Influence Curve
- Unbiased asymptotically normal estimators, usually invoking sample splitting.
- E.g. Λ may be estimated based on models for $m(z), \pi(z), \text{Var}(X|z)$ and

$$\lambda(z) = \frac{\text{Cov}(X, Y|z)}{\text{Var}(X|z)}$$

- This approach generalizes the AIPW which is usually based on $m(x, z), \pi(z)$

Marginal Effects

- Consider estimators constructed from the Efficient Influence Curve
- Unbiased asymptotically normal estimators, usually invoking sample splitting.
- E.g. Λ may be estimated based on models for $m(z), \pi(z), \text{Var}(X|z)$ and

$$\lambda(z) = \frac{\text{Cov}(X, Y|z)}{\text{Var}(X|z)}$$

- This approach generalizes the AIPW which is usually based on $m(x, z), \pi(z)$

Conditional Effects

- The R-learner[6] of heterogeneous treatment effect is based on
$$\arg\min_{\tau(.)} E\left(\left[Y - m(Z) - \tau(Z)\{X - \pi(Z)\}\right]^2\right)$$
- Generally speaking this is $\lambda(z)$ - so we can just use the R-learner for continuous exposures!
- For certain subgroup weightings, however, we can bypass learning of $\lambda(z)$ altogether

[6] Robinson (1988); Nie and Wager (2017)

Conditional Effects

- The R-learner[6] of heterogeneous treatment effect is based on

$$\arg\min_{\tau(.)} E\left(\left[Y - m(Z) - \tau(Z)\{X - \pi(Z)\}\right]^2\right)$$

- Generally speaking this is $\lambda(z)$ - so we can just use the R-learner for continuous exposures!
- For certain subgroup weightings, however, we can bypass learning of $\lambda(z)$ altogether

[6] Robinson (1988); Nie and Wager (2017)

Conditional Effects

- The R-learner[6] of heterogeneous treatment effect is based on
$$\arg\min_{\tau(.)} E\left(\left[Y - m(Z) - \tau(Z)\{X - \pi(Z)\}\right]^2\right)$$
- Generally speaking this is $\lambda(z)$ - so we can just use the R-learner for continuous exposures!
- For certain subgroup weightings, however, we can bypass learning of $\lambda(z)$ altogether

[6] Robinson (1988); Nie and Wager (2017)

Marginal Effects II

- A much simpler estimand[7] is

$$\frac{E\{\text{Cov}(X,Y|Z)\}}{E\{\text{Var}(X|Z)\}} = E\left\{\frac{\text{Var}(X|Z)}{E\{\text{Var}(X|Z)\}}\frac{X-\pi(Z)}{\text{Var}(X|Z)}Y\right\}$$

- This is also a weighted ADE! It also solves the least squares problem:

$$\arg\min_{\beta} E\left(\left[Y - m(Z) - \beta\{X-\pi(Z)\}\right]^2\right)$$

- And has the one-step estimator (partialling out estimator[8])

$$\frac{\sum_{i=1}^{n}\{x_i-\hat{\pi}(z_i)\}\{y_i-\hat{m}(z_i)\}}{\sum_{i=1}^{n}\{x_i-\hat{\pi}(z_i)\}^2}$$

[7] Vansteelandt and Dukes (2020); Robins et al. (2008); Newey and Robins (2018)

[8] Robinson (1988)

Marginal Effects II

- A much simpler estimand[7] is

$$\frac{E\{\mathrm{Cov}(X,Y|Z)\}}{E\{\mathrm{Var}(X|Z)\}} = E\left\{\frac{\mathrm{Var}(X|Z)}{E\{\mathrm{Var}(X|Z)\}}\frac{X-\pi(Z)}{\mathrm{Var}(X|Z)}Y\right\}$$

- This is also a weighted ADE! It also solves the least squares problem:

$$\arg\min_{\beta} E\left([Y - m(Z) - \beta\{X - \pi(Z)\}]^2\right)$$

- And has the one-step estimator (partialling out estimator[8])

$$\frac{\sum_{i=1}^{n}\{x_i - \hat{\pi}(z_i)\}\{y_i - \hat{m}(z_i)\}}{\sum_{i=1}^{n}\{x_i - \hat{\pi}(z_i)\}^2}$$

[7] Vansteelandt and Dukes (2020); Robins et al. (2008); Newey and Robins (2018)

[8] Robinson (1988)

Marginal Effects II

- A much simpler estimand[7] is

$$\frac{E\{\mathrm{Cov}(X, Y|Z)\}}{E\{\mathrm{Var}(X|Z)\}} = E\left\{\frac{\mathrm{Var}(X|Z)}{E\{\mathrm{Var}(X|Z)\}}\frac{X-\pi(Z)}{\mathrm{Var}(X|Z)}Y\right\}$$

- This is also a weighted ADE! It also solves the least squares problem:

$$\arg\min_{\beta} E\left([Y - m(Z) - \beta\{X - \pi(Z)\}]^2\right)$$

- And has the one-step estimator (partialling out estimator[8])

$$\frac{\sum_{i=1}^{n}\{x_i - \hat{\pi}(z_i)\}\{y_i - \hat{m}(z_i)\}}{\sum_{i=1}^{n}\{x_i - \hat{\pi}(z_i)\}^2}$$

[7] Vansteelandt and Dukes (2020); Robins et al. (2008); Newey and Robins (2018)

[8] Robinson (1988)

A summary

- Establish a link between causal estimands and classical derivative effects.
- Unify derivative effects and treatment effects through contrast functions.
- Establish a duality between contrast funtions and weighted ADEs.
- Propose new estimands based on their contrast functions which are easier to estimate and generally well defined.
- Derive estimators based on influence curve methods
- In our paper[9]: motivate derivative effects with reference to projections of $m(x, z)$ on to a sort of Taylor series.

[9] Arxiv version will be shared on twitter: @hines8

References

Díaz, I. and van der Laan, M. (2012). Population Intervention Causal Effects Based on Stochastic Interventions. *Biometrics*, 68(2):541–549.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995.

Kennedy, E. H. (2019). Nonparametric Causal Effects Based on Incremental Propensity Score Interventions. *Journal of the American Statistical Association*, 114(526):645–656.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 79(4):1229–1245.

Neugebauer, R. and van der Laan, M. (2007). Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434.

Newey, W. K. and Robins, J. M. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv*, pages 1–43.

Newey, W. K. and Stoker, T. M. (1993). Efficiency of Weighted Average Derivative Estimators and Index Models. *Econometrica*, 61(5):1199.

Nie, X. and Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. *arXiv*, (2019):1–50.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric Estimation of Index Coefficients. *Econometrica*, 57(6):1403.

Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. *Probability and Statistics: Essays in Honor of David A. Freedman*, 2:335–421.

Robins, R. D., M, G., and James. (2001). Causal Inference for Complex Longitudinal Data : The Continuous Case. *Annals of Statistics*, 29(6):1785–1811.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4):931.

Vansteelandt, S. and Dukes, O. (2020). Assumption-lean inference for generalised linear model parameters.