How to obtain valid tests and confidence intervals for treatment effects after confounder selection?

Stijn Vansteelandt
Ghent University, Belgium
London School of Hygiene and Tropical Medicine, U.K.
joint work with Oliver Dukes, Vahe Avagyan

Outline

1. Introduction
2. A dissection of the problem
3. Proposal
4. Numerical results
5. Discussion

Introduction (1)

- I will consider the problem of estimating the effect of some exposure A on an outcome Y based on data from an observational study.

Introduction (1)

- I will consider the problem of estimating the effect of some exposure A on an outcome Y based on data from an observational study.
- Exposed and unexposed subjects in such studies usually differ in many observed (pre-exposure) characteristics L.
- This can make it difficult to make contrasts of the mean outcome between exposed and unexposed subjects with the same characteristics.

Introduction (1)

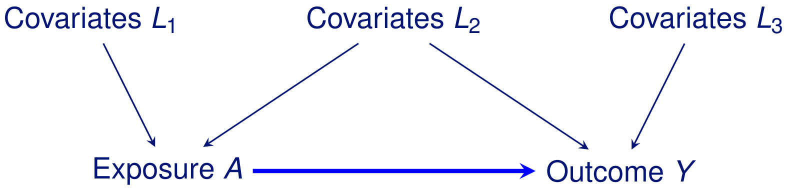
- I will consider the problem of estimating the effect of some exposure A on an outcome Y based on data from an observational study.
- Exposed and unexposed subjects in such studies usually differ in many observed (pre-exposure) characteristics L.
- This can make it difficult to make contrasts of the mean outcome between exposed and unexposed subjects with the same characteristics.
- The curse of dimensionality thus forces us to adopt some form of modelling.
- E.g. a linear model

$$E(Y|A,L) = \psi A + \beta' L$$

Introduction (2)

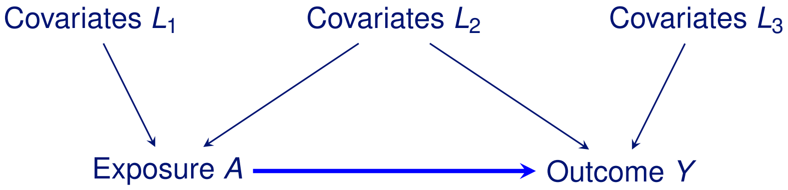
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

- Adjusting for all available characteristics L can be detrimental, or even impossible.
- It can inflate bias and variance.



Introduction (2)

- Adjusting for all available characteristics L can be detrimental, or even impossible.
- It can inflate bias and variance.



- There may be more covariates than observations.
- This is not uncommon, considering the possible need for interactions or other higher-order terms...

Introduction (3)

- Stepwise variable selection strategies and penalisation methods (e.g. the lasso) are therefore routinely employed.

Introduction (3)

- Stepwise variable selection strategies and penalisation methods (e.g. the lasso) are therefore routinely employed.
- One common strategy is to adjust for L iff it is significantly associated with outcome, conditional on exposure, at e.g. the 5% level.
- A related common strategy is the lasso, without penalisation of the exposure effect.

Introduction (3)

- Stepwise variable selection strategies and penalisation methods (e.g. the lasso) are therefore routinely employed.
- One common strategy is to adjust for L iff it is significantly associated with outcome, conditional on exposure, at e.g. the 5% level.
- A related common strategy is the lasso, without penalisation of the exposure effect.
- *How well does this work?*

Outcome-based selection

- Suppose that the exposure has no effect.
- Suppose that L has a moderate effect on outcome, but a strong effect on exposure.
- Then when fitting model

$$E(Y|A, L) = \psi A + \beta L$$

one will typically have little power to detect that $\beta \neq 0$.

Outcome-based selection

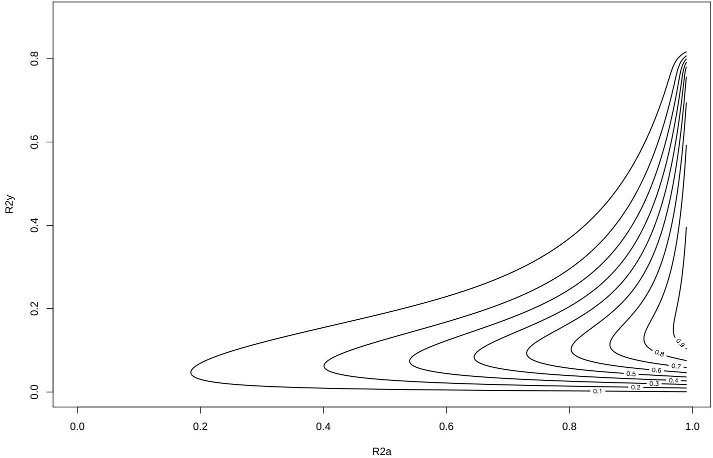
- Suppose that the exposure has no effect.
- Suppose that L has a moderate effect on outcome, but a strong effect on exposure.
- Then when fitting model

$$E(Y|A,L) = \psi A + \beta L$$

  one will typically have little power to detect that $\beta \neq 0$.
- Upon removing L from the model, one is likely to find 'strong evidence' of an exposure effect.
- This can result in highly inflated Type I error rates

Type I error rate inflation ($n = 100$)

$R2y$: R^2 of Y-L association; $R2a$: R^2 of A-L association



Convergence with increasing sample size

- This problem persists at all sample sizes.
- No matter how large the sample size,
  one can always choose correlations between Y-L and A-L,
  at which outcome-based selection inflates Type I error rates.

Convergence with increasing sample size

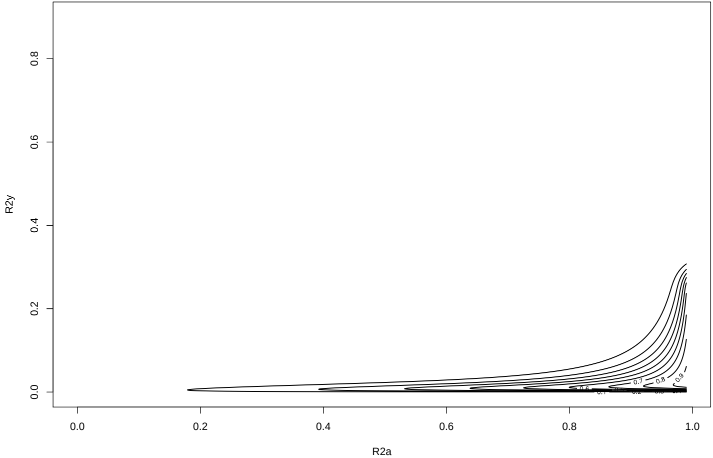
- This problem persists at all sample sizes.
- No matter how large the sample size,
  one can always choose correlations between Y-L and A-L,
  at which outcome-based selection inflates Type I error rates.
- We therefore say that convergence of the test statistic
  to a normal limit (centered around the truth) is non-uniform.

Convergence with increasing sample size

- This problem persists at all sample sizes.
- No matter how large the sample size,
  one can always choose correlations between Y-L and A-L,
  at which outcome-based selection inflates Type I error rates.
- We therefore say that convergence of the test statistic
  to a normal limit (centered around the truth) is non-uniform.
- Lack of uniform convergence is a concern.
- It implies that we can never guarantee
  that the procedure will do well in finite samples.

Outcome-based selection ($n = 1000$)

$R2y$: R^2 of Y-L association; $R2a$: R^2 of A-L association



Propensity-score-based selection (1)

- One key reason why this procedure is problematic, is that it prioritises the exposure: it prioritises the elimination of covariates over the elimination of the exposure.

  (Robins and Greenland, 1986)

- This problem can be overcome using propensity scores.

Propensity-score-based selection (2)

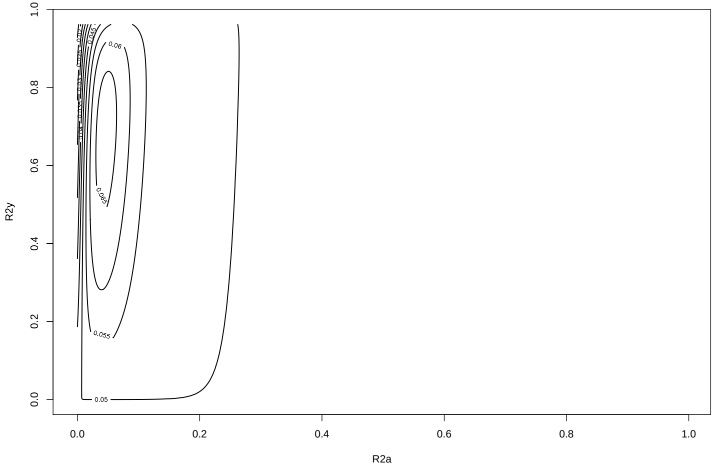
- Consider stepwise selection in a propensity score model, then regressing outcome on exposure and propensity score.
- By always adjusting for the propensity score, this strategy does not prioritise the exposure.

Propensity-score-based selection (2)

- Consider stepwise selection in a propensity score model, then regressing outcome on exposure and propensity score.
- By always adjusting for the propensity score, this strategy does not prioritise the exposure.
- With linear models for Y and A, and a single covariate L, this strategy is tantamount to adjusting for L iff it is significantly associated with exposure, at e.g. the 5% level.

Type I error rate inflation ($n = 100$)

$R2y$: R^2 of Y-L association; $R2a$: R^2 of A-L association



Propensity-score-based selection (2)

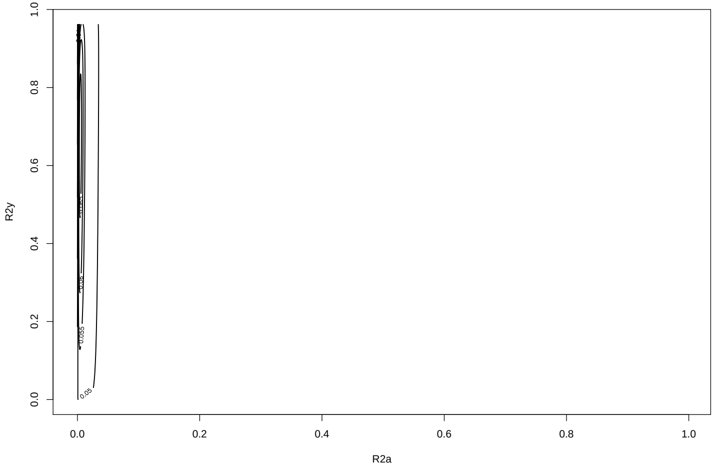
- By not prioritising the exposure,
  the problem of Type I error inflation is much less severe.

Propensity-score-based selection (2)

- By not prioritising the exposure,
  the problem of Type I error inflation is much less severe.
- In fact, ignoring the variable selection process
  often results in conservative inferences.
- This is line with the property
  that ignoring estimation of the propensity score
  typically results in conservative inferences.
- Also this persists at all sample sizes.

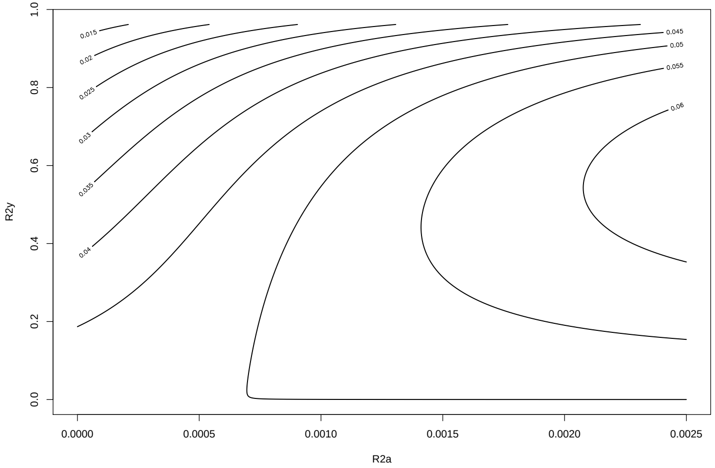
Propensity-score-based selection ($n = 1000$)

$R2y$: R^2 of Y-L association; $R2a$: R^2 of A-L association



Propensity-score-based selection ($n = 1000$)

$R2y$: R^2 of Y-L association; $R2a$: R^2 of A-L association



Problem solved?

- The conclusion so far is that propensity-score based selection is much less vulnerable to Type I error inflation than outcome-based selection.
- *Problem solved?*

Problem solved?

- The conclusion so far is that propensity-score based selection is much less vulnerable to Type I error inflation than outcome-based selection.
- *Problem solved?*
- Its typical conservatism implies a lack of power.
- *What if there are many covariates?*
- *What if the models are non-linear?*

Problem solved?

- The conclusion so far is that propensity-score based selection is much less vulnerable to Type I error inflation than outcome-based selection.
- *Problem solved?*
- Its typical conservatism implies a lack of power.
- *What if there are many covariates?*
- *What if the models are non-linear?*
- In view of this, the aim of this talk will be to develop uniformly valid tests that incorporate selection.
- The propensity score will continue to play a crucial role...

Post-selection inference

- This problem of post-selection inference has been quite thoroughly studied for some selection strategies.

  (e.g. Leeb and Pötscher, 2005; Berk et al., 2013; Taylor et al., 2014; ...)
- Most proposed solutions infer the distribution of the estimator or test statistic after selection.

  (e.g. Claeskens and Hjört, 2006)
- This has the disadvantage that the results
  - are often complex,
  - not immediately accessible for routine data analysis,
  - and sometimes dependent on the choice of procedure.

Aim

- Inspired by others,

  (Chernozhukov et al., 2017; Farrell, 2015)

  I will instead propose specific tests for treatment effect in combination with a specific selection strategy.
- Their combination is such that the test statistic converges uniformly to a normal distribution centred at the truth.

Outline

1. Introduction
2. A dissection of the problem
3. Proposal
4. Numerical results
5. Discussion

Hypothesis-test-based selection

- Reconsider model $E(Y|A, L) = \psi A + \beta L$ (where A and L are mean centred).
- Perform a score test of $\psi = 0$ based on the test statistic

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i(Y_i - \hat{\beta} L_i)$$

where $\hat{\beta}$ is the OLS estimator if we have selected L and 0 otherwise.

Hypothesis-test-based selection

- Reconsider model $E(Y|A,L) = \psi A + \beta L$
  (where A and L are mean centred).
- Perform a score test of $\psi = 0$ based on the test statistic

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i (Y_i - \hat{\beta} L_i)$$

  where $\hat{\beta}$ is the OLS estimator if we have selected L
  and 0 otherwise.
- *What is the distribution of the test statistic?*
- Consider outcome-based selection...

What is the distribution of the test statistic? (1)

By a Taylor expansion,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \hat{\beta}' L_i\}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \beta' L_i\} + \sqrt{n}(\hat{\beta} - \beta)\left\{\frac{1}{n}\sum_{i=1}^{n} A_i L_i\right\}$$

$$+ \text{Remainder}$$

What is the distribution of the test statistic? (1)

By a Taylor expansion,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \hat{\beta}' L_i\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \beta' L_i\} + \sqrt{n}(\hat{\beta} - \beta)\left\{\frac{1}{n}\sum_{i=1}^{n} A_i L_i\right\} \\
&\quad + \text{Remainder}
\end{aligned}$$

- When β is of the order $1/\sqrt{n}$,
  we will often erroneously set $\hat{\beta}$ to zero.

What is the distribution of the test statistic? (1)

By a Taylor expansion,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \hat{\beta}' L_i\}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \beta' L_i\} + \sqrt{n}(\hat{\beta}-\beta)\left\{\frac{1}{n}\sum_{i=1}^{n} A_i L_i\right\}$$

$$+ \text{Remainder}$$

- When β is of the order $1/\sqrt{n}$, we will often erroneously set $\hat{\beta}$ to zero.
- This results in bias, which affects the score test.

What is the distribution of the test statistic? (1)

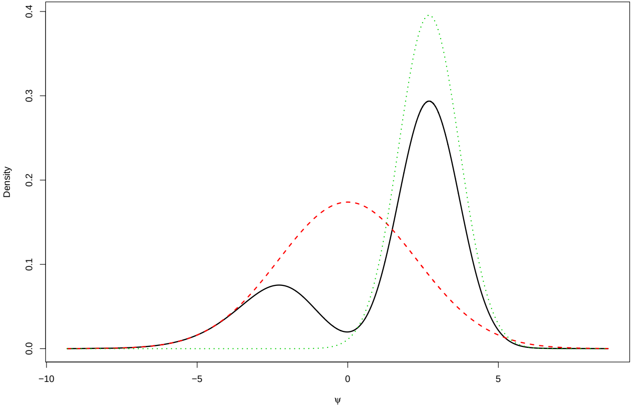
By a Taylor expansion,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \hat{\beta}' L_i\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \beta' L_i\} + \sqrt{n}(\hat{\beta}-\beta)\left\{\frac{1}{n}\sum_{i=1}^{n} A_i L_i\right\} \\
&\quad + \text{Remainder}
\end{aligned}
$$

- When β is of the order $1/\sqrt{n}$, we will often erroneously set $\hat{\beta}$ to zero.
- This results in bias, which affects the score test.
- $\sqrt{n}(\hat{\beta}-\beta)$ then moreover has a complex distribution.

Inference after variable selection (2)

This may cause bias, excess variability, and may invalidate inference.



What is the distribution of the test statistic? (2)

- Convergence of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A_i\{Y_i - \hat{\beta}' L_i\}$$

to a mean zero normal distribution is therefore non-uniform.

- We will remedy this
using bias-reduced double-robust estimators.

(Vermeulen and Vansteelandt, 2015)

Outline

1 Introduction

2 A dissection of the problem

3 Proposal

4 Numerical results

5 Discussion

Double-robust estimation

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Consider the test statistic

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i-\pi(L_i;\gamma)\}\{Y_i-m(L_i;\beta)\}$$

where we use
  - a parametric propensity score model $\mathcal{A}$:

$$E(A|L)=\pi(L;\gamma)$$

    e.g. $\text{expit}(\gamma' L)$ for binary A.
  - a parametric outcome model $\mathcal{B}$:

$$E(Y|L)=m(L;\beta)$$

    e.g. $\beta' L$ for continuous Y.

Double-robust estimation

- Consider the test statistic

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i - \pi(L_i;\gamma)\}\{Y_i - m(L_i;\beta)\}$$

where we use

  - a parametric propensity score model $\mathcal{A}$:

$$E(A|L) = \pi(L;\gamma)$$

    e.g. $\text{expit}(\gamma' L)$ for binary A.

  - a parametric outcome model $\mathcal{B}$:

$$E(Y|L) = m(L;\beta)$$

    e.g. $\beta' L$ for continuous Y.

- This test statistic has mean zero under the null when either model $\mathcal{A}$ or model $\mathcal{B}$ is correct.
- We therefore call it double-robust.

(Robins and Rotnitzky, 2001; see Rotnitzky and Vansteelandt, 2014, for a review)

What is the distribution of the test statistic now?

What is the distribution of the test statistic now?

- In practice, we need estimators of γ and β.
- Then

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\hat{\gamma},\hat{\beta}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\gamma,\beta) + \sqrt{n}(\hat{\gamma}-\gamma)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta)\right\} \\
&\quad + \sqrt{n}(\hat{\beta}-\beta)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta)\right\} + \text{Remainder}
\end{aligned}
$$

What is the distribution of the test statistic now?

- In practice, we need estimators of γ and β.
- Then

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\hat{\gamma},\hat{\beta}) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\gamma,\beta) + \sqrt{n}(\hat{\gamma}-\gamma)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta)\right\} \\
&\quad + \sqrt{n}(\hat{\beta}-\beta)\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta)\right\} + \text{Remainder}
\end{aligned}$$

- If we could set those gradients to zero, then local changes in these estimators would not affect the double-robust test.

Bias-reduced double-robust estimation

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Bias-reduced double-robust estimators achieve this by estimating γ by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta)=0$$

and β by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta)=0.$$

(Vermeulen and Vansteelandt, 2015)

Bias-reduced double-robust estimation

- Bias-reduced double-robust estimators achieve this by estimating γ by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta)=0$$

and β by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta)=0.$$

(Vermeulen and Vansteelandt, 2015)

- *Is this a valid proposal?*

Bias-reduced double-robust estimation

- Bias-reduced double-robust estimators achieve this by estimating γ by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta)=0$$

  and β by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta)=0.$$

  (Vermeulen and Vansteelandt, 2015)
- *Is this a valid proposal?*
- Suppose model $\mathcal{A}$ is correct with true value γ^*.
- Then $U_i(\gamma^*,\beta)$ has mean zero for all β, so that

$$E\left\{\frac{\partial}{\partial\beta}U_i(\gamma^*,\beta)\right\}=0$$

Incorporating variable selection

- So far, I have not considered variable selection.

Incorporating variable selection

- So far, I have not considered variable selection.
- We will incorporate it by penalising the estimating equations with a bridge penalty:

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}U_i(\gamma,\beta) + \lambda_\beta\delta|\beta|^{\delta-1}\circ\operatorname{sign}(\beta)$$

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}U_i(\gamma,\beta) + \lambda_\gamma\delta|\gamma|^{\delta-1}\circ\operatorname{sign}(\gamma)$$

where $\lambda_\gamma > 0$ and $\lambda_\beta > 0$ are penalty parameters, and $\delta \to 1+$.

(Avagyan and Vansteelandt, 2017; Dukes, Avagyan and Vansteelandt, 2018)

Incorporating variable selection

- So far, I have not considered variable selection.
- We will incorporate it by penalising the estimating equations with a bridge penalty:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \beta} U_i(\gamma, \beta) + \lambda_\beta \delta |\beta|^{\delta-1} \circ \operatorname{sign}(\beta)$$

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \gamma} U_i(\gamma, \beta) + \lambda_\gamma \delta |\gamma|^{\delta-1} \circ \operatorname{sign}(\gamma)$$

where $\lambda_\gamma > 0$ and $\lambda_\beta > 0$ are penalty parameters, and $\delta \to 1+$.

(Avagyan and Vansteelandt, 2017; Dukes, Avagyan and Vansteelandt, 2018)

- Standard choices of penalty (of the order $\sqrt{\log(p)/n}$) make these gradients sufficiently close to zero.

Example - Y continuous, A binary

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

- Consider models $\pi(L;\gamma) = \text{expit}(\gamma' L)$ and $m(L;\beta) = \beta' L$.
- Then we estimate γ and β as the solutions to

$$0 = \frac{1}{n}\sum_{i=1}^{n}\{A_i - \text{expit}(\gamma' L_i)\}L_i + \lambda_\gamma \delta |\gamma|^{\delta-1} \circ \text{sign}(\gamma)$$

$$0 = \frac{1}{n}\sum_{i=1}^{n} w_i(\gamma)\{Y_i - \beta' L_i\}L_i + \lambda_\beta \delta |\beta|^{\delta-1} \circ \text{sign}(\beta)$$

where $w_i(\gamma) = \text{expit}(\gamma' L_i)\{1 - \text{expit}(\gamma' L_i)\}$.

Example - Y continuous, A binary

- In practice, we let $\delta \to 1+$ and solve the following problems:

$$\min_{\gamma} \mathcal{F}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} \log\{1+\exp(\gamma' L_i)\} - A_i(\gamma' L_i) + \lambda_\gamma ||\gamma||_1$$

$$\min_{\beta} \mathcal{F}(\beta) = \frac{1}{2n}\sum_{i=1}^{n} [\hat{w}_i\{Y_i - \beta' L_i\}^2] + \lambda_\beta ||\beta||_1$$

- Components of $\hat{\eta}$ may be shrunk to zero, in view of which we recommend refitting the selected model.
- The test statistic is then

$$T_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\{A_i - \text{expit}(\hat{\gamma}' L_i)\}\{Y_i - \hat{\beta}' L_i\}}{\sqrt{\frac{1}{n}\{\frac{1}{n-1}\sum_{i=1}^{n}[\{A_i - \text{expit}(\hat{\gamma}' L_i)\}\{Y_i - \hat{\beta}' L_i\}]^2\}}}$$

Asymptotic properties - both models correct

- Let s_γ and s_β be the sparsity indices of models $\mathcal{A}$ and $\mathcal{B}$.
- Suppose that (in addition to mild regularity conditions), the following sparsity assumptions hold:
  (i) $s_\gamma \log(p) = o(n)$
  (ii) $s_\beta \log(p) = o(n)$
  (iii) $s_\gamma s_\beta \log^2(p) = o(n)$.

Theorem

When model $\mathcal{A}$ and $\mathcal{B}$ are correct, the test statistic T_n converges uniformly to a standard normal distribution.

A closer look at the conditions...

- Conditions
  (i) $s_\gamma \log(p) = o(n)$
  (ii) $s_\beta \log(p) = o(n)$
  are quite standard to guarantee consistency of the lasso-based estimators.

A closer look at the conditions...

- Conditions
  (i) $s_\gamma \log(p) = o(n)$
  (ii) $s_\beta \log(p) = o(n)$
  are quite standard to guarantee consistency
  of the lasso-based estimators.
- Condition
  (iii) $s_\gamma s_\beta \log^2(p) = o(n)$
  determines the rate of convergence of the estimators.
- It suggests that if one model is sparse,
  the other can be more dense.

A closer look at the conditions...

- Conditions
  (i) $s_\gamma \log(p) = o(n)$
  (ii) $s_\beta \log(p) = o(n)$
  are quite standard to guarantee consistency
  of the lasso-based estimators.
- Condition
  (iii) $s_\gamma s_\beta \log^2(p) = o(n)$
  determines the rate of convergence of the estimators.
- It suggests that if one model is sparse,
  the other can be more dense.
- When evaluating medical treatments, this is arguably satisfied as clinicians may use a limited number of variables to decide on treatment, whereas outcome may be affected by many more.

Relationship to existing literature

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Compared with other recent proposals from high-dimensional inference in GLMs:

(van de Geer et al., 2014; Belloni et al., 2016)

- We have weakened the assumptions on sparsity by making use of double robustness.

(see also Farrell, 2015, for the ATE)

- Other approaches usually require ultra-sparsity, e.g. $s_\gamma \sqrt{\log(p)} = o(\sqrt{n})$ instead of $s_\gamma \log(p) = o(n)$.

Relationship to existing literature

Compared with other recent proposals from high-dimensional inference in GLMs:

(van de Geer et al., 2014; Belloni et al., 2016)

- We have weakened the assumptions on sparsity by making use of double robustness.

(see also Farrell, 2015, for the ATE)

- Other approaches usually require ultra-sparsity, e.g. $s_\gamma \sqrt{\log(p)} = o(\sqrt{n})$ instead of $s_\gamma \log(p) = o(n)$.
- Unlike others, we do not require sample-splitting to obtain weaker rates.

(Chernozhukov et al., 2017)

Asymptotic properties - model misspecification

Suppose that (in addition to the previous conditions), the following sparsity assumptions hold:

(iv) Either (a) $s_\gamma\sqrt{\log(p)} = o(\sqrt{n})$ (if model $\mathcal{A}$ is correct)
or (b) $s_\beta\sqrt{\log(p)} = o(\sqrt{n})$ (if model $\mathcal{B}$ is correct).

Theorem

When model $\mathcal{A}$ or $\mathcal{B}$ is correct, the test statistic T_n converges uniformly to a standard normal distribution.

Note the tradeoff between modelling and sparsity conditions.

Relationship to existing literature

- Other proposals from high-dimensional inference in GLMs assume $\mathcal{A}$ and $\mathcal{B}$ to be linear,
  and $\mathcal{B}$ to be correctly specified and ultra-sparse.

  (van de Geer et al., 2014; Belloni et al., 2016; Shah and Bühlmann, 2017)
- By using specific bias-reduction strategies, our tests
  - allow arbitrary conditional mean models for $\mathcal{A}$ and $\mathcal{B}$,
  - remain valid when $\mathcal{A}$ or $\mathcal{B}$ is misspecified,
  - use weaker sparsity assumptions.

Relationship to existing literature

- Other proposals from high-dimensional inference in GLMs assume $\mathcal{A}$ and $\mathcal{B}$ to be linear,
  and $\mathcal{B}$ to be correctly specified and ultra-sparse.

  (van de Geer et al., 2014; Belloni et al., 2016; Shah and Bühlmann, 2017)
- By using specific bias-reduction strategies, our tests
  - allow arbitrary conditional mean models for $\mathcal{A}$ and $\mathcal{B}$,
  - remain valid when $\mathcal{A}$ or $\mathcal{B}$ is misspecified,
  - use weaker sparsity assumptions.
- Weaker sparsity assumptions do not suffice for Wald tests.

Outline

1. Introduction
2. A dissection of the problem
3. Proposal
4. **Numerical results**
5. Discussion

Simulation study

- $n = 200$
- linear models with $Z_1, ..., Z_p$ for $p = 140$ mutually independent, standard normal variates.
- 19 confounders, generally strongly associated with exposure, and more weakly with outcome.
- No pure exposure predictors.
- 40 pure outcome predictors.
- Covariates explain 80% of the variability in exposure and outcome.
- 1000 simulation experiments.
- Penalty parameters chosen via cross-validation (1 SE).

Simulation results: $n = 200, p = 100$

Correct outcome model

Method	Type I error
Standard naïve	0.212
`hdm` DS	0.470
`hdm` OI	0.451
Proposal	0.063
Proposal (Unweighted)	0.063

Simulation results: $n = 200, p = 200$

Correct outcome model

Method	Type I error
Standard naïve	0.399
`hdm` DS	0.454
`hdm` OI	0.435
Proposal	0.074
Proposal (Unweighted)	0.087

Simulation results: $n = 200, p = 100$

Misspecified outcome model

Method	Type I error
Standard naïve	0.156
`hdm` DS	0.194
`hdm` OI	0.191
Proposal	0.072
Proposal (Unweighted)	0.059

Simulation results: $n = 200, p = 200$

Misspecified outcome model

Method	Type I error
Standard naïve	0.266
`hdm` DS	0.233
`hdm` OI	0.233
Proposal	0.060
Proposal (Unweighted)	0.067

Outline

1. Introduction
2. A dissection of the problem
3. Proposal
4. Numerical results
5. Discussion

Discussion

- Routine outcome-based variable selection strategies are problematic.
- Propensity-score-based selection has much greater validity, but is not guaranteed to result in tests with the nominal size.

Discussion

- Routine outcome-based variable selection strategies are problematic.
- Propensity-score-based selection has much greater validity, but is not guaranteed to result in tests with the nominal size.
- Double-robust tests enable uniformly valid inference in high-dimensional settings with correct model specification.

  (Chernozhukov et al., 2017; Farrell, 2015)

Discussion

- Routine outcome-based variable selection strategies are problematic.
- Propensity-score-based selection has much greater validity, but is not guaranteed to result in tests with the nominal size.
- Double-robust tests enable uniformly valid inference in high-dimensional settings with correct model specification.

  (Chernozhukov et al., 2017; Farrell, 2015)
- For testing the null, we have shown that weaker conditions are attainable without the need for sample-splitting.

Discussion

- Routine outcome-based variable selection strategies are problematic.
- Propensity-score-based selection has much greater validity, but is not guaranteed to result in tests with the nominal size.
- Double-robust tests enable uniformly valid inference in high-dimensional settings with correct model specification.

  (Chernozhukov et al., 2017; Farrell, 2015)
- For testing the null, we have shown that weaker conditions are attainable without the need for sample-splitting.
- We have extended this to allow for model misspecification.
- This required the use of special ‘bias-reduced’ fitting strategies.

  (Vermeulen and Vansteelandt, 2015)

This talk was based on...

Avagyan, V. and Vansteelandt, S. (2017). Honest data-adaptive inference for the average treat- ment effect under model misspecification using penalised bias-reduced double-robust estimation. arXiv:1708.03787

Dukes, O., Avagyan, V. and Vansteelandt, S. (2018). High-Dimensional Doubly Robust Inference for Regression Parameters. Technical Report.

Vermeulen, K. and Vansteelandt, S. (2015). Bias-Reduced Doubly Robust Estimation. Journal of the American Statistical Association, 110(511):1024-1036.

Bias in function of the nuisance parameters