EXPLORATORY MEDIATION ANALYSIS WITH MANY POTENTIAL MEDIATORS

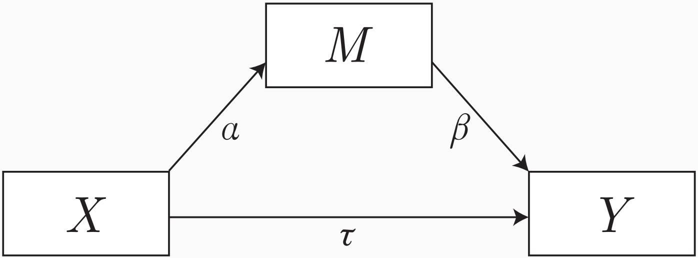
Erik-Jan van Kesteren & Daniel Oberski

March 15, 2018

Utrecht University, Department of Methodology & Statistics

MEDIATION

Q: When is *M* a mediator?



MacKinnon et al. (2002):

1. Causal steps: α & β
2. Difference in coefficients: $\tau - \tau|M$
3. Product of coefficients: $\alpha \times \beta$

VanderWeele (2015, *p.* 46): *"Also take into account $X \cdot M$ interaction!"*

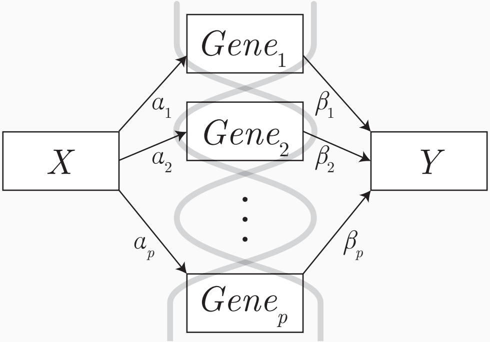
Theory-based decision functions using data from X, M, Y:

$$\mathcal{D}\colon \{\boldsymbol{x}, \boldsymbol{m}, \boldsymbol{y}\} \mapsto \{0, 1\}$$

(0 = not mediator, 1 = mediator)

MANY MEDIATORS

Q: When is $Gene_i$ a mediator?



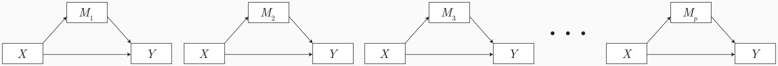
Preacher and Hayes (2008):

1. Fit the full SEM so your parameter estimates take all mediators into account
2. Select mediators using the estimated parameters

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y}) \text{ conditional on } M_{-i}$$

FILTER

p single mediator models



`for (i in 1:`p`)` $\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y})$

The "filter" method (Guyon and Elisseeff, 2003)

Good

- Simple
- Quick
- Flexible

Bad

- Assumes uncorrelated mediators: won't work if mediation only visible conditionally

Jacobucci et al. (2016): We can now penalise SEM parameters

$$F_{\text{regsem}} = F_{\text{ML}} + \lambda P(\cdot)$$

Serang et al. (2017): We can use this to select mediators! Put a lasso penalty on α and β

The “XMed” method

Good

- ”Full” SEM
- Does not assume uncorrelated mediators
- Regularisation is hip

Bad

- What are we actually optimising for?
- Find M for which α OR β but we want α AND β.

Our contribution:

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y}) \text{ conditional on } M_{-i}$$

COORDINATE DESCENT

Insight from regularisation literature (Hastie et al., 2015):

conditional parameter == parameter estimated on residual

Idea:

```
sel ← rep(0, p)

while (!convergence) {
  for (i in 1:p) {
    x_res ← x - M[, sel] %*% beta_x_sel
    y_res ← y - M[, sel] %*% beta_y_sel
    sel[i] ← decisionFunction(x_res, M[, i], y_res)
  }
}
```

COORDINATE-WISE MEDIATION FILTER

for each mediator	Coordinate-wise
perform the decision function	Mediation
throw it out if 0	Filter

conditional on the other selected mediators

repeat until convergence

Good

- Uses theoretically relevant $\mathcal{D}$
- Does not assume uncorrelated mediators

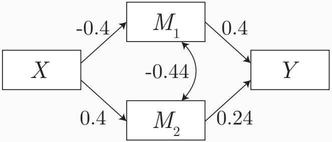
Bad

- Nonconvergence

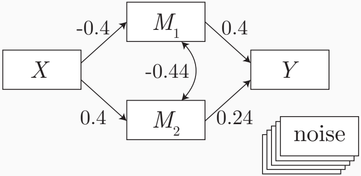
SIMULATION

SIMULATION

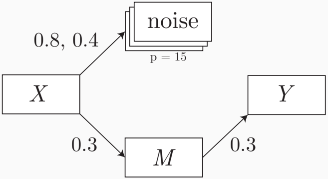
Conditional-only



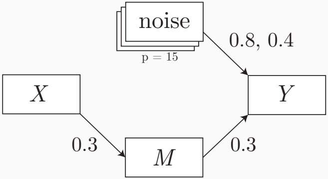
High-dimensional



Noise (α paths)



Noise (β paths)



IMPLEMENTATION

```
> devtools::install_github("vankesteren/cmfilter")
> library(cmfilter)
> res <- cmf(x, M, y, verbose = TRUE)

# CMF Algorithm
#
# -----------
#
# 1 0 0 1 1 1 1 0 1 0 0 0 0 1 1 0 0
# 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
# 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
# 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
#
# Algorithm converged
#
# -----------
```

CONCLUSION

- Novel method for selecting among many mediators
- Flexible choice of $\mathcal{D}$
- Conditional on M_{-i}
- Stable in boundary cases
- Traditional power/type-I tradeoff

FUTURE WORK

- Group lasso
- Bayesian regularisation prior on $\alpha \times \beta$

Guyon, I. and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)*, 3(3):1157–1182.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, Boca Raton.

Jacobucci, R., Grimm, K. J., and McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Structural Equation Modeling*, 23(4):555–566.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, 7(1):83–104.

Preacher, K. J. and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3):879–891.

Serang, S., Jacobucci, R., Brimhall, K. C., and Grimm, K. J. (2017). Exploratory Mediation Analysis via Regularization. *Structural Equation Modeling*, 24(5):733–744.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, 1st edition.

QUESTIONS

Suppression

Method	Power (M_1)	Power (M_2)
Full SEM	1.00	0.99
Filter	0.99	0.13
XMed	1.00	0.99
CMF	1.00	0.91

Noise (α paths)

Method	Power	FPR	PPV
Full SEM	0.20	0.11	0.11
Filter	0.27	0.09	0.17
XMed	0.67	0.34	0.12
CMF	0.17	0.06	0.17

Noise (β paths)

Method	Power	FPR	PPV
Full SEM	0.08	0.01	0.32
Filter	0.44	0.02	0.58
XMed	0.49	0.12	0.22
CMF	0.41	0.02	0.58

High-dimensional data

Method	Power (M_1)	Power (M_2)	FPR	PPV
Full SEM	NA	NA	NA	NA
Filter	0.91	0.07	2.4e-3	0.30
XMed	NA	NA	NA	NA
CMF	0.82	0.06	1.8e-3	0.32