Exploratory Mediation Analysis with Many Potential Mediators

Erik-Jan van Kesteren
Daniel Oberski

Utrecht University, Netherlands
Department of Methodology & Statistics

Outline

Exploratory Mediation

Current options

Coordinate-wise mediation filter

Implementation

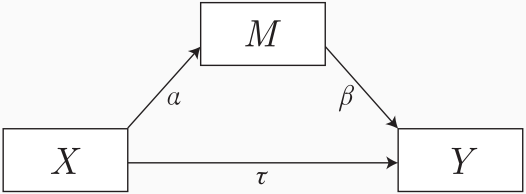
Simulation

Conclusion

Exploratory Mediation

Single mediator model

Q: When is M a mediator?



Single mediator model

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!"*

Single mediator model

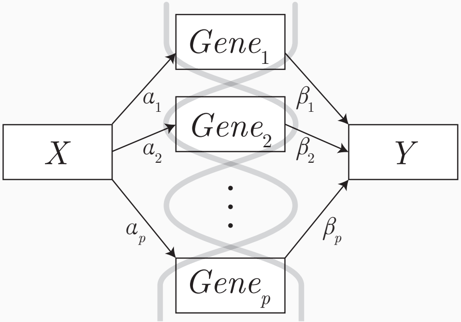
Theory-based decision functions using data from X, M, Y:

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

(0 = not mediator, 1 = mediator)

Many Mediators

Q: When is $Gene_i$ a mediator?



Many Mediators

Preacher and Hayes (2008):

1. Fit the full Structural Equation Model with all M
   $\Rightarrow$ estimates take all mediators into account
2. Perform $\mathcal{D}$ using the estimated parameters

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

Many Mediators

With many mediators $(p > n)$ SEM is unavailable!

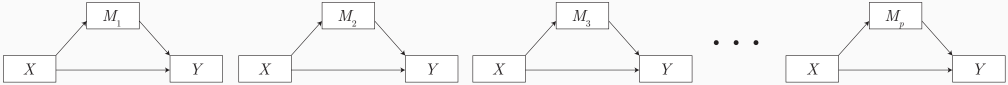
Current options

Three options

- Filter
- XMed
- HIMA

Filter

The filter method: p single mediator models



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

Filter

Good

- Simple
- Quick
- Flexible

Bad

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

XMed

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

XMed

Good

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

Bad

- Find M for which α OR β but we want α AND β.
- Implementation does not handle high-dimensional data.

HIMA

Three-step sequential combination of the above (Zhang et al., 2016):

1. Filter the top $\frac{2n}{\log n}$ M variables based on the β coefficients
2. Estimate remaining β coefficients with sparsity
3. For remaining M variables, perform $\mathcal{D}_{\text{causal steps}}$

HIMA

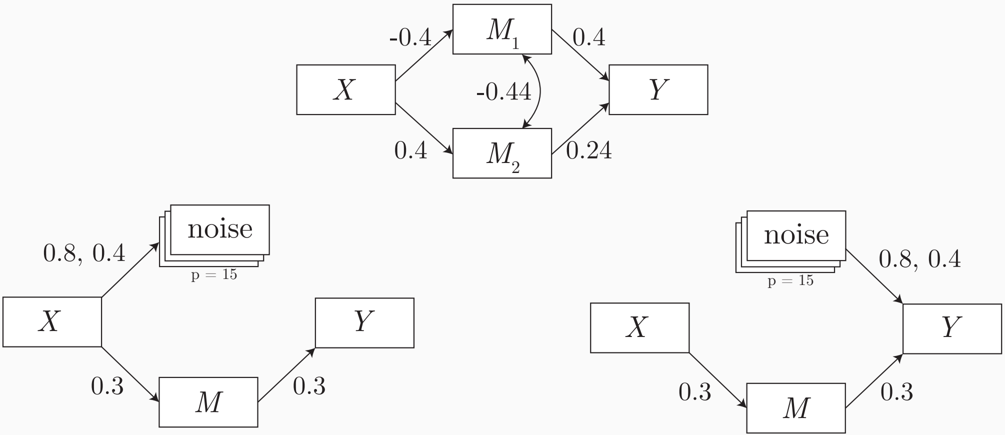
Good

- Very fast implementation
- Promising performance
- Regularisation is hip

Bad

- Very focused on $M \to Y$
- Fixed $\mathcal{D}_{\text{causal steps}}$

Illustrative simulations



Coordinate-wise mediation filter

Coordinate-wise mediation filter

Our contribution:

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

Coordinate-wise mediation filter

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

conditional parameter == parameter estimated on residual

Coordinate-wise mediation filter

```r
sel ← rep(0, p)

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

Coordinate-wise mediation filter

for each mediator	**C**oordinate-wise
perform the decision function	**M**ediation
throw it out if 0	**F**ilter

conditional on the other selected mediators

repeat until convergence

Coordinate-wise mediation filter

Good

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

Bad

- Nonconvergence
  $\Rightarrow$ weak learner

Nonconvergence

Aggregating the weak learner:

- Multiple random starts (parallel processing)
  $\Rightarrow$ empirical selection probability
- Randomly order variables within iterations
- Consider only $\sqrt{p}$ variables at each step
- Early stopping
- Convergence after > 1 unchanged iteration

Implementation

Implementation

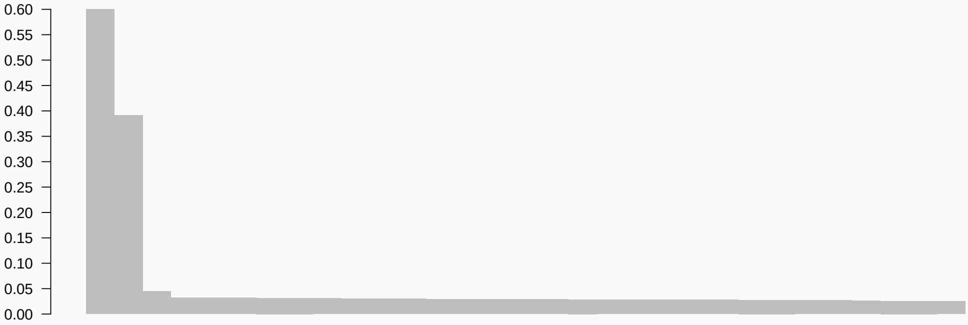
```
> library(cmfilter)

> # Perform the cmf algorithm
> result ← cmf(dataset, nStarts = 10000)

  |+++++++++++++++++++++++++                                  | 51% ~52s
```

Implementation

```
> screeplot(result)
```



Implementation

```
> result <- setCutoff(result, 0.2)
> result

CMF Algorithm Results

----------------------

call:
cmf(x = d, nStarts = 10000, cutoff = 0.2)

Algorithm converged.
variables selected: 2
number of starts: 10000
cutoff probability: 0.2

----------------------
```

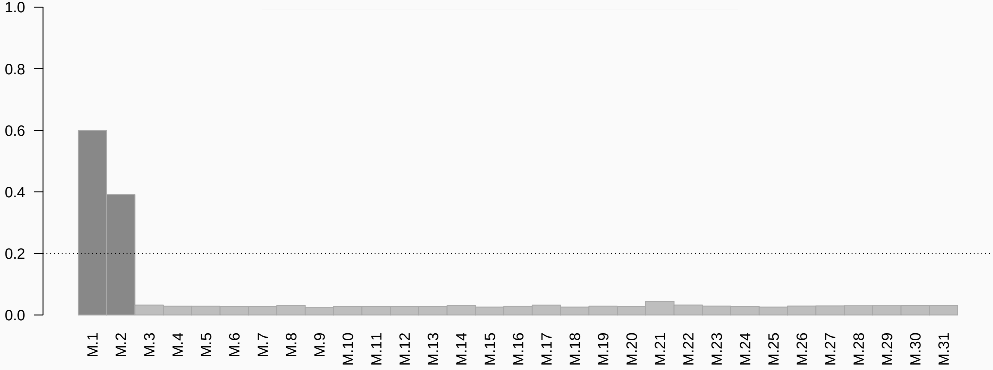
```
----------------------

Top 10:
     SelectionRate Selected
M.1         0.6001     TRUE
M.2         0.3911     TRUE
M.21        0.0446    FALSE
M.22        0.0324    FALSE
M.3         0.0323    FALSE
M.17        0.0321    FALSE
M.31        0.0317    FALSE
M.30        0.0316    FALSE
M.8         0.0311    FALSE
M.14        0.0304    FALSE

----------------------
```

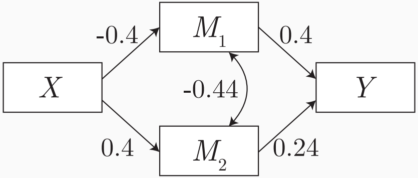
Implementation

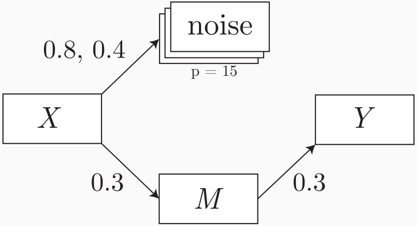
```
> plot(result)
```

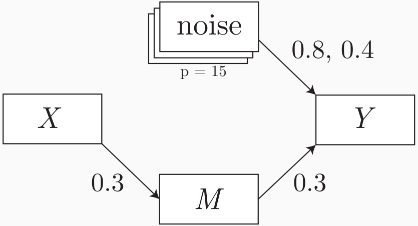


Simulation

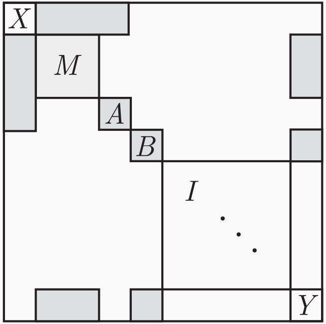
Illustrative simulations







High-Dimensional Simulation



Method	TPR	FPR	PPV
CMF	.55	.005	.52
Filter	.22	.002	.52
HIMA	.06	.009	.03

Conclusion

Conclusion

- New algorithmic method for exploratory mediation analysis
- Flexible choice of $\mathcal{D}$
- Conditional on $M_{\text{-}i}$
- Performs at benchmark-level (including in boundary cases)
- Works for high-dimensional data
- Implemented in R package `cmfilter`

e.vankesteren1@uu.nl
github.com/vankesteren
@ejvankesteren

References

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.

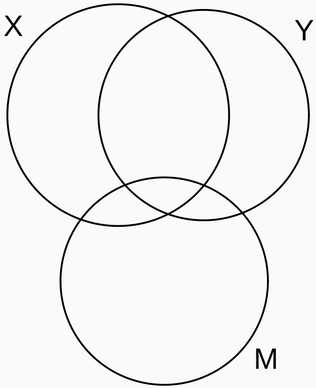
Serang, S., Jacobucci, R., Brimhall, K. C., and Grimm, K. J. (2017). Exploratory Mediation Analysis via Regularization. *Structural Equation Modeling*, 24(5):733–744.

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

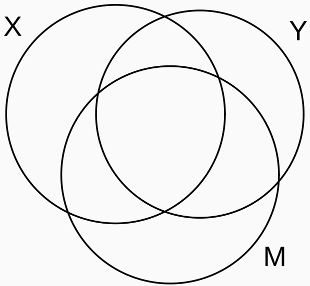
Zhang, H., Zheng, Y., Zhang, Z., Gao, T., Joyce, B., Yoon, G., Zhang, W., Schwartz, J., Just, A., Colicino, E., Vokonas, P., Zhao, L., Lv, J., Baccarelli, A., Hou, L., and Liu, L. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics*, 32(20):3150–3154.

Single mediator model

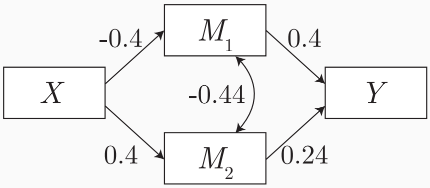
Weak mediation



Strong mediation

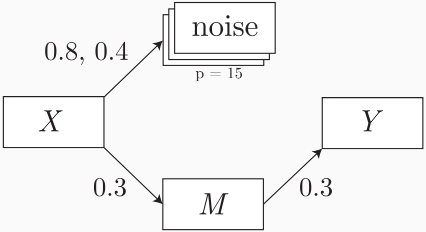


Conditional-only



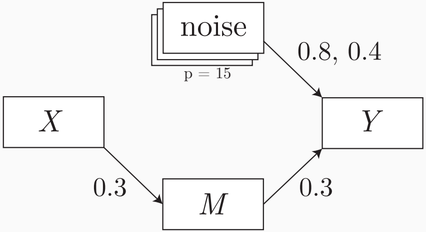
Method	M1	M2
SEM	100	100
Filter	100	.
XMed	100	100
HIMA	100	100
CMF	100	100

Noise in α paths



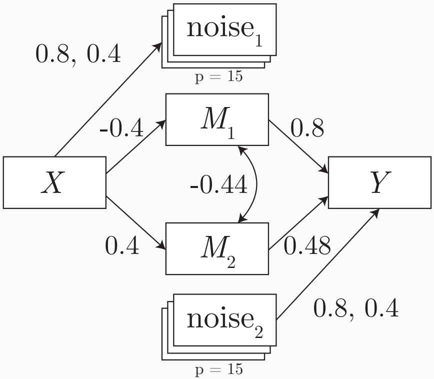
Method	TPR	FPR
SEM	100	.
Filter	100	17
XMed	77	.
HIMA	100	.
CMF	100	.

Noise in β paths



Method	TPR	FPR
SEM	100	.
Filter	100	.
XMed	100	.
HIMA	.	.
CMF	100	.

Everything combined



Method	M1	M2	FPR	PPV
SEM	1	1	.	1
Filter	1	.	0.02	0.27
XMed	1	1	0.1	0.77
HIMA	1	1	.	1
CMF	1	1	.	1