Cell-type specific mediation project update

Andrea Lane

2/22/2020

Outline

1. Project idea reminder
2. EM max.iter and tol parameters
3. TOAST results
4. EM results

Project idea reminder

$$Y \sim N(\theta_0 + \theta_1 E + \sum_k \theta_2^k m_k, \ \gamma^2)$$

$$M \sim N(\sum_k m_k \pi_k, \ \sigma^2)$$

$$m_k \sim N(\beta_0^k + \beta_1^k E, \ \tau_k^2)$$

- Use TOAST to get β initial values and reduce number of sites to go through EM
- Run EM algorithm one site at a time to get estimates of Θ
- Calculate observed indirect effect $\hat{\beta}_1 \hat{\theta}_2$
- Use percentile bootstrap to assess significance of the indirect effect for each site and cell type

EM initial values

I wanted to evaluate the efficiency and accuracy of the EM algorithm to set the tolerance and maximum iteration parameters. Up to this point, I had set max.iter=500 and tol=0.001.

Table 1: Final iteration Theta2 estimates (true values in parentheses)

	Cell 1 (0)	Cell 2 (0)	Cell 3 (0)	Cell 4 (0.4)
Initial values: true values				
500; .01	-0.1799810	0.0351150	0.0358092	0.4047087
500; .001	-0.1799810	0.0351150	0.0358092	0.4047087
2000; .0001	0.0053486	0.0317210	-0.0786960	0.3854218
Initial values: 0				
500; .01	-0.1979723	0.0688971	-0.2234888	0.1747376
500; .001	-0.0321125	0.0561223	-0.3536296	0.1920167
2000; .0001	0.0041869	0.0366775	-0.3592493	0.1854096

Comments:

- odd that convergence is faster with 0s as initial values than true values as initial values?
- estimates not as good with 0s as initial values (to be expected)
- have been using max.iter=500, tol=.001

EM results

To assess the EM/boostrap procedure, I ran it with only the mediating sites using the true values as initial values.

Settings/Notes:

- N=500
- To account for multiple testing, the (.025/Ncell*Nsite,1-(.025/Ncell*Nsite)) quantile is used for bootstrap samples.
- for EM, max.iter=500 and tol=0.001
- EM initial values:
    - beta: from TOAST
    - theta: true values
    - tau, sigma: sampled from inverse gamma distribution with mean 0.001 (based on HIRE, which states that variance initial values are sampled from inverse gamma distribution with small mean but does not specify the mean)
    - gamma: true value 0.0004
- Scripts used: "Lane_Mediation_sim12821_cluster.R" and "mediation_sim_functions_1120.R" (functions script contains EM julia code)

Table 2: 1 site - 200 bootstrap

Theta = 0.2				Theta = 0.4			
Cell 1	Cell 2	Cell 3	Cell 4	Cell 1	Cell 2	Cell 3	Cell 4
317	78	167	194	**874**	94	162	94
315	**678**	115	68	158	**998**	79	37
80	24	**942**	65	23	13	**998**	15
57	14	42	**984**	19	7	13	**999**

Table 3: 2 sites - 200 bootstrap

Theta = 0.2				Theta = 0.4			
Cell 1	Cell 2	Cell 3	Cell 4	Cell 1	Cell 2	Cell 3	Cell 4
358	112	209	213	**895**	71	163	65
317	2	2	0	**790**	3	3	0
320	**710**	165	77	352	**995**	159	79
4	**277**	1	2	7	**711**	2	2
160	52	**896**	51	40	16	**990**	79
83	37	**886**	14	45	31	**997**	12
51	12	21	**831**	19	8	52	**987**
155	47	36	**999**	101	33	37	**1000**

Table 4: Average across mediating and nonmediating sites/cell types 200 bootstrap

	1 site	2 sites	1 site	2 sites
	Theta = 0.2		Theta = 0.4	
Power	0.730	0.659	0.967	0.921
Type 1 error	0.102	0.076	0.060	0.057

Table 5: 1 site - 1000 bootstrap

Theta = 0.2				Theta = 0.4			
Cell 1	Cell 2	Cell 3	Cell 4	Cell 1	Cell 2	Cell 3	Cell 4
307	65	133	166	**855**	73	132	82
273	**659**	104	47	113	**998**	51	28
61	15	**934**	50	20	6	**999**	12
35	14	36	**981**	9	4	11	**999**

Table 6: Average across mediating and nonmediating sites/cell types 1000 bootstrap

	Theta = 0.2	Theta = 0.4
Power	0.720	0.963
Type 1 error	0.083	0.045

For reference, I also present the power and type 1 error from the last set of results where I used 0 for the initial theta values.

Table 7: Average across mediating and nonmediating sites/cell types 0s as initial values (theta=0.4)

	1 site	2 sites	3 sites
Power	0.691	0.696	0.692
Type 1 error	0.308	0.160	0.155