Comparison of methods with fitting by coefficient and outcome

2020-02-02

In our previous results, we tried to calculate the Kullback Leibler Divergence with normal distributions estimated by coefficient of the LME or the outcome of the LME. The results do not have very big differences. Therefore, we tried to increase the residuals ϵ and decrease the random effect covariance value of the concanvity part. We would like to check whether they have differences in the new setting.

Simulation setting

With 1000 repetitions. The parameters are used the same parameters estimated in EMBARC dataset.

- n = 100
- $\epsilon_{drg} \sim N(0, \sigma^2), \epsilon_{pbo} \sim N(0, \sigma^2)$, where $\sigma = 10$
- $\alpha_0 = [\cos(\theta), \sin(\theta)] = [\frac{\sqrt(2)}{2}, \frac{\sqrt(2)}{2}]$, where $\theta = \frac{\pi}{4}$
- $\beta_{drg} = [18.63, -2.29, 0.17]$, $\beta_{pbo} = [18.59, -1.86, 0.14]$
- $\Gamma_{drg} = [0, 1, 0]$, $\Gamma_{pbo} = [0, \cos(\eta), \sin(\eta)], \eta = 0, 60, 120, 180$
- $D_{drg} =$

```
##             (Intercept)         tt    I(tt^2)
## (Intercept)   8.8686336  2.7662417 -0.3588474
## tt            2.7662417  1.0157249 -0.1040613
## I(tt^2)      -0.3588474 -0.1040613  0.0010000
```

- $D_{pbo} =$

```
##             (Intercept)         tt     I(tt^2)
## (Intercept)  9.50686040  1.0123563 -0.09265618
## tt           1.01235632  3.8560459 -0.39529489
## I(tt^2)     -0.09265618 -0.3952949  0.00100000
```

where we changed $D_{drg}[3, 3] = D_{pbo}[3, 3] = 0.001$

Result

- eta: the angles between Γ_1 and $\Gamma 2$;
- true_theta: the true angle for generation of α_0;
- est_theta: the mean of the estimated θ;
- sd_theta: the standard deviation of the estimated θ;
- ratio_theta: the fraction of est_theta and sd_theta;
- cossim: cosine similarity of the estimated α and α_0
- true_purity: the purity calculated by using the true $\alpha, \beta_1, \beta_2, \Gamma_1, \Gamma_2, D1, D2, \mu_x, \Sigma_x$. * est_purity: the mean of the esimated purity;
- sd_purity: the standard devation of the estimated purity.

Estimate the coefficient as $z = (\beta + \Gamma(\alpha' x) + b) \sim MVN(\beta + \Gamma(\alpha' x), D)$ and then calculate the purity:

Table 1: coefficient

eta	true_theta	est_theta	sd_theta	ratio_theta	cossim	true_purity	est_purity	sd_purity
0	0.785	0.794	0.233	3.404	0.947	0.542	3357633622	21480994330

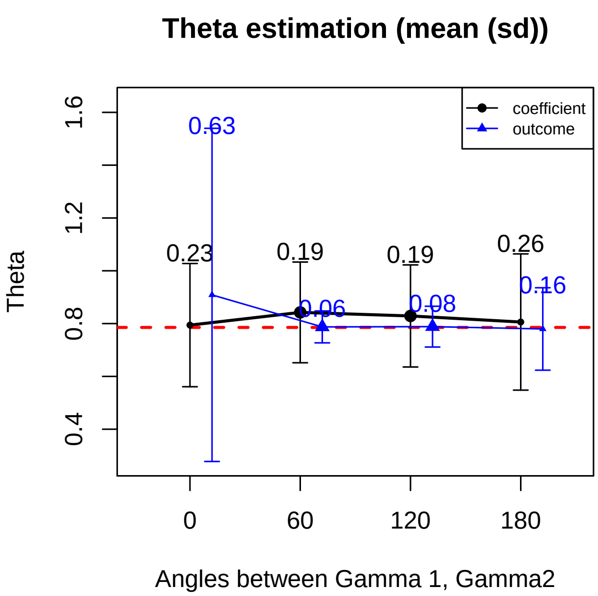
eta	true_theta	est_theta	sd_theta	ratio_theta	cossim	true_purity	est_purity	sd_purity
60	0.785	0.842	0.191	4.417	0.968	477.908	456131539	2347682445
120	0.785	0.829	0.193	4.288	0.969	369.217	208225516	934111953
180	0.785	0.806	0.258	3.124	0.951	38.184	37768620777	240242725431

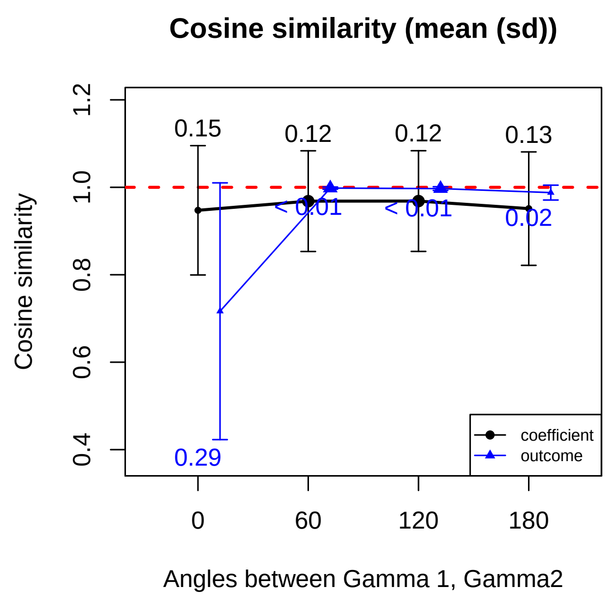
Estimate the outcomes as $Y = X(\beta + \Gamma(\alpha' x)) + Zb \sim MVN(X(\beta + \Gamma(\alpha' x)), ZDZ') + \sigma^2 I$ and then calculate the purity

Table 2: outcome

eta	true_theta	est_theta	sd_theta	ratio_theta	cossim	true_purity	est_purity	sd_purity
0	0.785	0.909	0.631	1.440	0.717	6.124	5.319	0.127
60	0.785	0.787	0.060	13.044	0.998	72.372	16.853	1.827
120	0.785	0.788	0.077	10.205	0.997	57.143	12.371	1.196
180	0.785	0.780	0.156	4.994	0.988	12.719	7.890	0.632

To make the results more intuitive, we draw the following plots about estimation of theta and estimation of cosine similarity. The plot of purity is not shown here since the sd is too large.





Comparison with previous setting

$\sigma = 1$ and $D_{drg}[3,3] = 0.15$, $D_{pbo}[3,3] = 0.04$

Estimate the coefficient as $z = (\beta + \Gamma(\alpha' x) + b) \sim MVN(\beta + \Gamma(\alpha' x), D)$ and then calculate the purity

eta	true_theta	est_theta	sd_theta	ratio_theta	cossim	true_purity	est_purity	sd_purity
0	0.785	0.800	0.253	3.163	0.965	1.195	1.505	0.533
60	0.785	0.785	0.032	24.630	0.999	241.835	261.166	51.682
120	0.785	0.785	0.034	23.338	0.999	187.044	201.892	41.082
180	0.785	0.787	0.072	10.935	0.997	20.170	21.870	3.551

Estimate the outcomes as $Y = X(\beta + \Gamma(\alpha' x)) + Zb \sim MVN(X(\beta + \Gamma(\alpha' x)), ZDZ') + \sigma^2 I$ and then calculate the purity

eta	true_theta	est_theta	sd_theta	ratio_theta	cossim	true_purity	est_purity	sd_purity
0	0.785	0.786	0.225	3.500	0.976	11.375	6.145	0.366
60	0.785	0.786	0.036	21.994	0.999	395.923	164.405	20.099
120	0.785	0.786	0.038	20.818	0.999	336.868	131.551	15.510
180	0.785	0.781	0.076	10.332	0.997	33.371	21.010	2.618