PCB sub sampling simulation 2

Xuelong Wang

2019-09-13

Contents

1 Motivation and issues **1**

2 Standardization **1**

3 Simulation result for the main effect **2**
3.1 PCB main effect 2
3.2 PCB total effect 3

4 Simulation study on the small small values **4**
4.1 Normal 4
4.2 Chi 5

1 Motivation and issues

Based on the PCB data simulation results, if we don't do the standardization then there tends to have a larger bias for the estimated total effect than the standardized covariates. So for the following section, we try to figure out what are the exact reasons.

The one feature and issues is that the values of each PCB is very small, which lead to small values in its covariance matrix. However, the correlation coefficients among those PCBs are very large. When I fit the model with those small values of PCBs we have found that the total variance estimation tends to have a bigger bias than I use the stanardized covariates.

I suspected that the small values of the PCBs and their covariance matrix may affect the performance of the main and interaction variance estimation

2 Standardization

- $Var(X^T a) = 8$, rescale a based on the $a^T \Sigma_{emp} a$
- Σ_{emp} is calcuated by all the PCBs from 1999 with $n = 4 * 10^3$ In here the standardization means to normalization:

  $$\tilde{Z} = D(X - \mu_x),$$

  Where D is the scaling matrix, $d_{11} = SD(X_1)$. So after the standardization, the variance explained by $\tilde{Z}$ should be same as X, as following shows

  $$Y = a_0 + X^T a + \epsilon \Rightarrow Y = b_0 + \tilde{Z}^T b + \epsilon$$

  $$Y = b_0 + \tilde{Z}^T b + \epsilon = b_0 + (X^T Db - \mu_x^T Db) + \epsilon$$

In order to have the same model, we only need take linear tranformation of the previous parameters

$$a = Db \text{ and } a_0 = b_0 - \mu_x^T Db.$$

After the reparameterization, we found that the $Var(\tilde{Z}^T b) = Var(X^T a)$, so that it should not affect performance of variance estimation as following shows.

3 Simulation result for the main effect

3.1 PCB main effect

3.1.1 original scale without decorrelation

	var_main_effect	decor	x_dist	n	MSE	est_var	est_mean	NA_main	method
1:	8	FALSE	1999	100	18	15.7	6.5	0	GCTA
2:	8	FALSE	1999	150	22	20.7	6.7	0	GCTA
3:	8	FALSE	1999	231	11	6.6	5.8	0	GCTA
4:	8	FALSE	1999	500	11	4.3	5.4	0	GCTA
5:	8	FALSE	1999	1000	10	2.0	5.1	0	GCTA

3.1.2 original scale decorrelation

	var_main_effect	decor	x_dist	n	MSE	est_var	est_mean	NA_main	method
1:	8	TRUE	1999	100	17.0	13.4	6.1	0	GCTA
2:	8	TRUE	1999	150	17.9	16.7	6.8	0	GCTA
3:	8	TRUE	1999	231	8.7	7.5	6.9	0	GCTA
4:	8	TRUE	1999	500	5.7	5.3	7.4	0	GCTA
5:	8	TRUE	1999	1000	2.9	2.9	7.8	0	GCTA

3.1.3 std covariate without decorrelation

	var_main_effect	decor	x_dist	n	MSE	est_var	est_mean	NA_main	method
1:	8	FALSE	1999	100	67	46.1	13	0	GCTA
2:	8	FALSE	1999	150	76	52.8	13	0	GCTA
3:	8	FALSE	1999	231	33	22.5	11	0	GCTA
4:	8	FALSE	1999	500	22	12.6	11	0	GCTA
5:	8	FALSE	1999	1000	11	5.5	10	0	GCTA

3.1.4 std covariate decorrelation

	var_main_effect	decor	x_dist	n	MSE	est_var	est_mean	NA_main	method
1:	8	TRUE	1999	100	15.3	15.2	7.5	0	GCTA
2:	8	TRUE	1999	150	16.2	16.3	8.1	0	GCTA
3:	8	TRUE	1999	231	7.2	7.1	7.7	0	GCTA
4:	8	TRUE	1999	500	4.3	4.3	7.7	0	GCTA
5:	8	TRUE	1999	1000	2.3	2.3	8.0	0	GCTA

3.2 PCB total effect

After adding the interaction terms, the difference between original and standardized scores are different.

3.2.1 original scale without decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8               2                   2.4              15
```

```
       n MSE est_var est_mean NA_total     method
 1:  100 166     149     10.5        1 EigenPrism
 2:  150 247     246     13.0        0 EigenPrism
 3:  231 150     138     11.1        0 EigenPrism
 4:  500 NaN      NA      NaN      100 EigenPrism
 5: 1000 NaN      NA      NaN      100 EigenPrism
 6:  100  95      68      9.6        0       GCTA
 7:  150 222     209     10.9        0       GCTA
 8:  231  95      70      9.7        0       GCTA
 9:  500 117     105     11.2        0       GCTA
10: 1000  65      60     12.4        0       GCTA
```

3.2.2 original scale decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8               2                   2.4              15
```

```
       n MSE est_var est_mean NA_total     method
 1:  100 105      92     11.2        0 EigenPrism
 2:  150 121     107     10.9        0 EigenPrism
 3:  231  64      31      9.0        0 EigenPrism
 4:  500 NaN      NA      NaN      100 EigenPrism
 5: 1000 NaN      NA      NaN      100 EigenPrism
 6:  100 141     130     11.4        0       GCTA
 7:  150 138     127     11.3        0       GCTA
 8:  231  64      37      9.6        0       GCTA
 9:  500  53      30     10.0        0       GCTA
10: 1000  32      16     10.7        0       GCTA
```

3.2.3 std without decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8               2                  0.62              11
```

```
       n MSE est_var est_mean NA_total     method
 1:  100 201     143       19        4 EigenPrism
 2:  150 500     348       24        0 EigenPrism
 3:  231 324     182       23        0 EigenPrism
 4:  500 NaN      NA      NaN      100 EigenPrism
 5: 1000 NaN      NA      NaN      100 EigenPrism
 6:  100 150     135       15        0       GCTA
 7:  150 710     614       21        0       GCTA
 8:  231 281     193       21        0       GCTA
 9:  500 665     380       28        0       GCTA
```

```
10: 1000 576     211       30        0       GCTA
```

3.2.4 original scale decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.62               11
```

```
       n MSE est_var est_mean NA_total     method
 1:  100  49    49.6     11.0        0 EigenPrism
 2:  150  38    38.5     10.7        0 EigenPrism
 3:  231  19    16.2      9.5        0 EigenPrism
 4:  500 NaN      NA      NaN      100 EigenPrism
 5: 1000 NaN      NA      NaN      100 EigenPrism
 6:  100  53    52.4     10.2        0       GCTA
 7:  150  42    41.6     10.3        0       GCTA
 8:  231  19    16.7      9.5        0       GCTA
 9:  500  12     9.9      9.8        0       GCTA
10: 1000   7     5.3      9.9        1       GCTA
```

4 Simulation study on the small small values

4.1 Normal

4.1.1 original scale without decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.046               10
```

```
       n  MSE est_var est_mean NA_total     method
 1:  100 21.9    20.1     11.5        0 EigenPrism
 2:  150 16.3    16.1      9.5        0 EigenPrism
 3:  231 13.0    11.2      8.7        0 EigenPrism
 4:  500  NaN      NA      NaN      100 EigenPrism
 5: 1000  NaN      NA      NaN      100 EigenPrism
 6:  100 25.5    15.8     13.2        0       GCTA
 7:  150 10.9     8.8     11.6        0       GCTA
 8:  231  6.0     5.3     11.0        0       GCTA
 9:  500  3.5     2.3      9.0        0       GCTA
10: 1000  4.8     1.0      8.1        0       GCTA
```

4.1.2 original scale with decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.046               10
```

```
       n  MSE est_var est_mean NA_total     method
 1:  100 20.8   20.98     10.1        0 EigenPrism
 2:  150 13.3   13.45     10.2        0 EigenPrism
 3:  231  9.4    8.30     11.2        0 EigenPrism
 4:  500  NaN      NA      NaN      100 EigenPrism
 5: 1000  NaN      NA      NaN      100 EigenPrism
 6:  100 26.2   25.03      8.9        0       GCTA
```

```
 7:  150 17.6   16.81      9.1        0       GCTA
 8:  231  8.7    6.47      8.6        0       GCTA
 9:  500  2.4    2.33     10.4        0       GCTA
10: 1000  2.3    0.97     11.2        0       GCTA
```

4.2 Chi

4.2.1 original scale without decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                   1.9               14
```

```
       n MSE est_var est_mean NA_total     method
 1:  100 139      88       21        1 EigenPrism
 2:  150 157      75       23        0 EigenPrism
 3:  231 194      67       25        0 EigenPrism
 4:  500 NaN      NA      NaN      100 EigenPrism
 5: 1000 NaN      NA      NaN      100 EigenPrism
 6:  100 116     102       18        0       GCTA
 7:  150 131      88       20        0       GCTA
 8:  231 146      76       22        0       GCTA
 9:  500 172      39       25        0       GCTA
10: 1000 151      23       25        0       GCTA
```

4.2.2 original scale with decorrelation

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                   1.9               14
```

```
       n  MSE est_var est_mean NA_total     method
 1:  100 22.4    21.2       13        0 EigenPrism
 2:  150 17.3    16.6       13        0 EigenPrism
 3:  231 12.5    11.4       13        0 EigenPrism
 4:  500  NaN      NA      NaN      100 EigenPrism
 5: 1000  NaN      NA      NaN      100 EigenPrism
 6:  100 16.5    15.5       13        0       GCTA
 7:  150 14.6    14.6       13        0       GCTA
 8:  231  7.1     6.8       13        0       GCTA
 9:  500  4.0     2.9       13        0       GCTA
10: 1000  2.9     1.3       12        1       GCTA
```