Sparse covariance estimation

Xuelong Wang

2019-09-20

Contents

1 Motivation

After decorrelation with the information of the historical data, the correlation between the covariate is reduced. However, there are still some correlation coeifficients are large. That may suggest that after using the historical data, the decorrelated data still is not uncorrelated. But the correlation structure becomes a sparse and symmetric. Therefore, we could apply another decorrelation to further reudce the non-zero correlation, so that we may have a better performance on the following variance estimation procedure.

2 Simulation

2.1 Simulation procedure

2.1.1 Standardization will not change total variance

1. Standardize the X $\tilde{Z}_m = (X - \mu)A_1$
2. Generate the interaction based on the $\tilde{Z}_{int} = \tilde{Z}_m * \tilde{Z}_m$ without the square terms and set $\tilde{Z}_t = (\tilde{Z}_m, \tilde{Z}_{int})$
3. Generate the Y based on the $\tilde{Z}_t$
4. Estimate the $Var(\tilde{Z}_t\beta_t)$ by $Z = \tilde{Z}_tA_2$, where A_2 is for decorrelation

Note that the $Var(\tilde{Z}_t\beta_t) = Var(Z\gamma)$, $A_2 = \hat{\Sigma}_h^{-1/2}$ or $A_2 = \hat{\Sigma}_h^{-1/2}\hat{\Sigma}_s^{-1/2}$

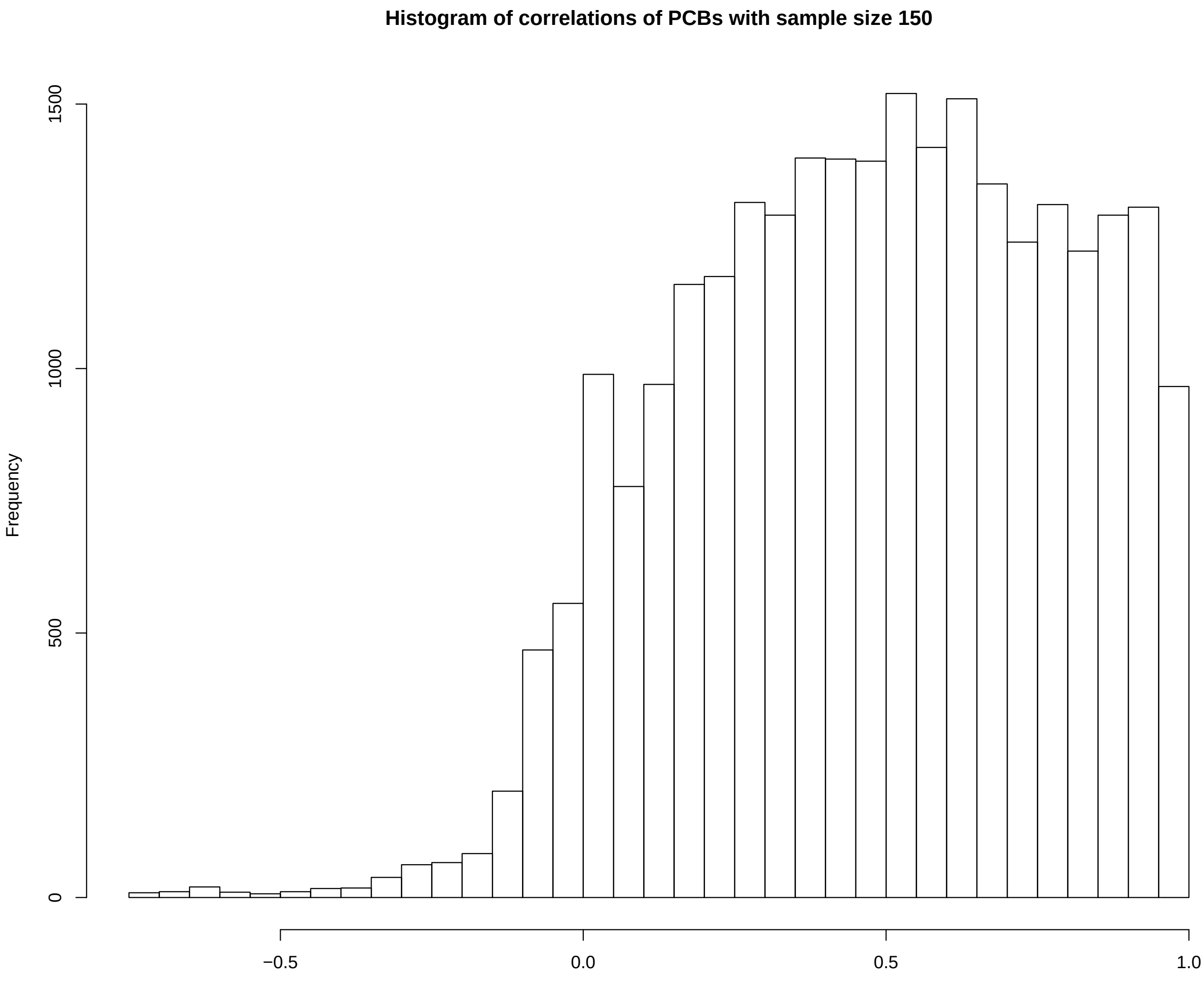
2.2 Decorrelation steps

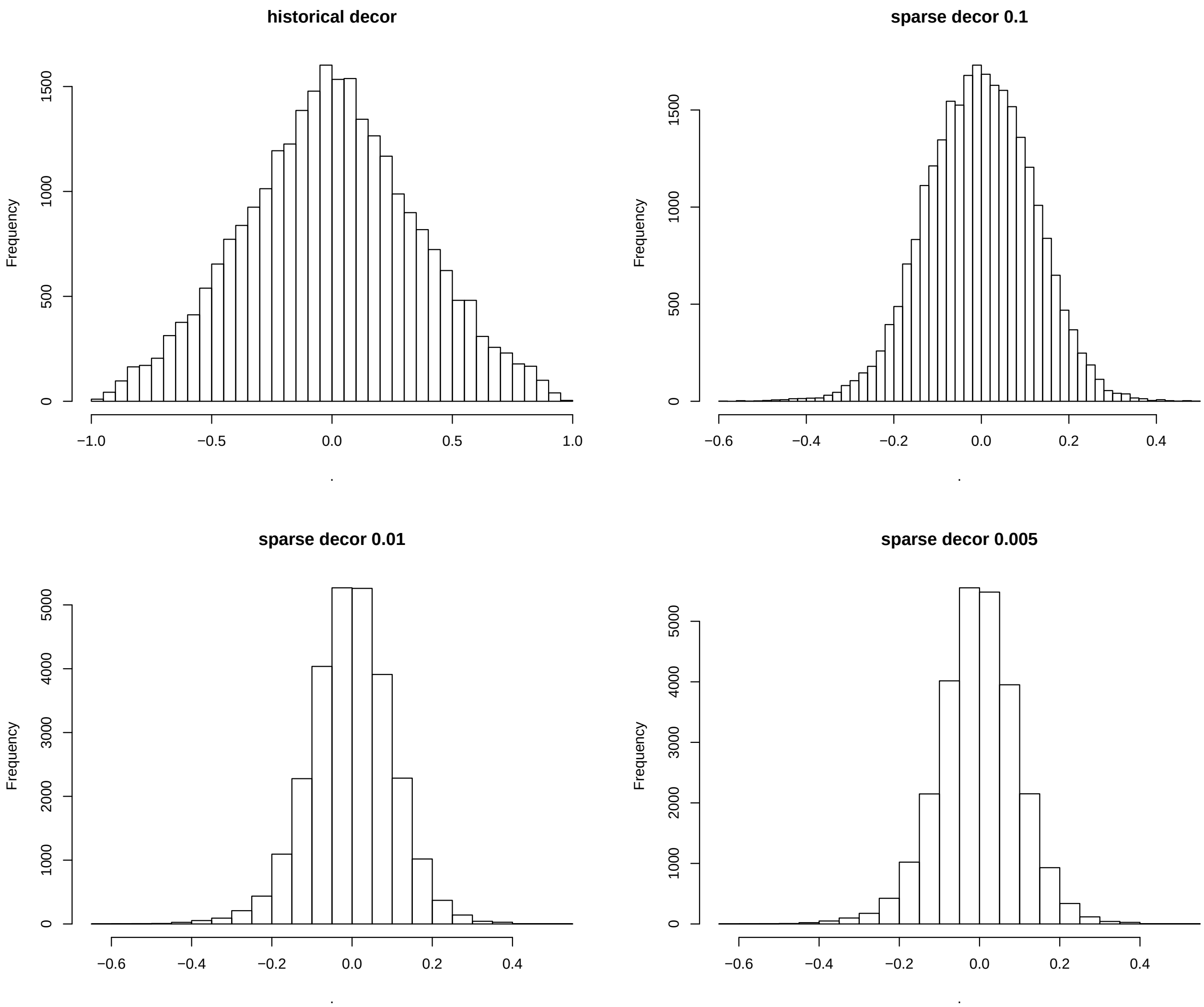
2.3 two steps decorrelation

1. Decorrelation by covariance matrix estimated by historical data
2. If after the step 1, the correlation is still large, then we may need a second decorrelation by sparse precision method

2.4 Dpglasso: The Graphical Lasso: New Insights and Alternatives

1. An Alternatives for Glasso
2. Glasso works on W the covariance matrix, but its alg cannot make sure the precision matrix Θ is positive definite.
3. Dpglasso can provide both W and Θ be positive definite




historical decor
Frequency
0
500
1000
1500
−1.0
−0.5
0.0
0.5
1.0
sparse decor 0.1
Frequency
0
500
1000
1500
−0.6
−0.4
−0.2
0.0
0.2
0.4
sparse decor 0.01
Frequency
0
1000
2000
3000
4000
5000
−0.6
−0.4
−0.2
0.0
0.2
0.4
sparse decor 0.005
Frequency
0
1000
2000
3000
4000
5000
−0.6
−0.4
−0.2
0.0
0.2
0.4

2.5 PCB

2.5.1 None

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.62               11
   structure decor x_dist
1:        un FALSE   1999
```

```
       n MSE est_var est_mean NA_total     method
 1:  100 206     112       21        3 EigenPrism
 2:  100 153     108       18        0       GCTA
 3:  150 304     112       25        0 EigenPrism
 4:  150 278     149       23        0       GCTA
 5:  231 276      83       25        0 EigenPrism
 6:  231 217      86       23        0       GCTA
 7:  500 NaN      NA      NaN      100 EigenPrism
 8:  500 472     166       29        0       GCTA
 9: 1000 NaN      NA      NaN      100 EigenPrism
10: 1000 495     113       31        0       GCTA
```

2.5.2 None with hist

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.62               11
   structure decor x_dist
1:        un  TRUE   1999
```

```
       n MSE est_var est_mean NA_total     method
 1:  100  31      32       11        3 EigenPrism
 2:  100  22      22       11        0       GCTA
 3:  150  26      26       11        0 EigenPrism
 4:  150  21      21       11        0       GCTA
 5:  231  19      20       11        0 EigenPrism
 6:  231  14      14       11        0       GCTA
 7:  500 NaN      NA      NaN      100 EigenPrism
 8:  500  31      31       12        0       GCTA
 9: 1000 NaN      NA      NaN      100 EigenPrism
10: 1000  44      39       13        1       GCTA
```

2.5.3 None with hist + sparse

```
   var_main_effect var_inter_effect cov_main_inter_effect var_total_effect
1:               8                2                  0.62               11
   structure decor x_dist
1:        un  TRUE   1999
```

```
       n  MSE est_var est_mean NA_total     method
 1:  100 14.6   14.68     11.4        0 EigenPrism
 2:  100 14.0   14.11     11.2        0       GCTA
 3:  150  8.6    7.91     10.4        0 EigenPrism
 4:  150  6.9    6.78     10.8        0       GCTA
 5:  231 11.6    7.42      9.2        0 EigenPrism
 6:  231  5.2    4.37     10.3        0       GCTA
```

	n	MSE	est_var	est_mean	NA_total	method
7:	500	NaN	NA	NaN	100	EigenPrism
8:	500	3.1	1.53	10.0	0	GCTA
9:	1000	NaN	NA	NaN	100	EigenPrism
10:	1000	2.4	0.99	10.1	1	GCTA

2.6 Chi

2.6.1 None

	var_main_effect	var_inter_effect	cov_main_inter_effect	var_total_effect	structure	decor	x_dist
1:	8	2	1.8	14	un	FALSE	chi

	n	MSE	est_var	est_mean	NA_total	method
1:	100	108	67	20	3	EigenPrism
2:	100	77	68	17	0	GCTA
3:	200	213	90	25	7	EigenPrism
4:	200	189	109	23	0	GCTA
5:	500	NaN	NA	NaN	100	EigenPrism
6:	500	131	31	24	0	GCTA
7:	1000	NaN	NA	NaN	100	EigenPrism
8:	1000	134	20	24	0	GCTA

2.6.2 hist

	var_main_effect	var_inter_effect	cov_main_inter_effect	var_total_effect	structure	decor	x_dist
1:	8	2	1.8	14	un	TRUE	chi

	n	MSE	est_var	est_mean	NA_total	method
1:	100	23.4	21.5	12	0	EigenPrism
2:	100	18.5	16.5	12	0	GCTA
3:	200	15.5	12.9	12	0	EigenPrism
4:	200	9.3	8.2	13	0	GCTA
5:	500	NaN	NA	NaN	100	EigenPrism
6:	500	4.1	2.8	13	0	GCTA
7:	1000	NaN	NA	NaN	100	EigenPrism
8:	1000	3.0	1.3	12	1	GCTA

2.6.3 hist + sparse

	var_main_effect	var_inter_effect	cov_main_inter_effect	var_total_effect	structure	decor	x_dist
1:	8	2	1.8	14	un	TRUE	chi

	n	MSE	est_var	est_mean	NA_total	method
1:	100	27.7	27.6	14	0	EigenPrism
2:	100	26.9	27.1	14	0	GCTA
3:	200	12.4	10.7	12	0	EigenPrism
4:	200	8.9	8.8	13	0	GCTA

```
5:  500  NaN      NA      NaN      100 EigenPrism
6:  500  4.5     2.5       12        0       GCTA
7: 1000  NaN      NA      NaN      100 EigenPrism
8: 1000  4.1     1.8       12        0       GCTA
```