Jackknife variance estimation corrections

Xuelong Wang

2019-12-06

Contents

1 Jackknife variance correction

If we assume the S is a smooth functions of emperical CDF, especially a quadratic functions, then it can be shown the leading terms of $E(\tilde{Var}(S(X_1, \ldots, S_{n-1}))) \geq Var(S(X_1, \ldots, S_{n-1}))$ is a quadratic term in expecation. Therefore we could try to estimate the quadratic term and correct the bias for the jackknife variance estimation.

Define $Q_{ii'} \equiv nS - (n-1)(S_i + S_{i'}) + (n-2)S_{(ii')}$, then the correction will be

$$\hat{Var}^{corr}(S(X_1, \ldots, X_n)) = \hat{Var}(S(X_1, \ldots, X_n)) - \frac{1}{n(n-1)} \sum_{i<i'} (Q_{ii'} - \bar{Q})^2$$

where $\bar{Q} = \sum_{i<i'} (Q_{ii'})/(n(n-1)/2)$

2 Simulation study compare two GCTA and GCTA_rr

GCTA_rr is the `mixed.solve` function from `rrBLUP` r package.
Based on the following simulation results,

1. when $n < p$ case, those two methods' results are very closed to each other.
2. when $n > p$ case, in terms of effect estimation and jackknife variance estimation those two methods's reuslts are similar to each other. But for the variance corrections are quite different. That is the statistics Q of our method has a very large variance which leads to negative correction result.

2.0.1 setup

- Independent
- Normal
- $p = 100$
- $n = \{50, 75, 100, 150, 200\}$
- with interaction terms
- main effect: $Var(X^T\beta) = \{0, 8, 100\}$

2.0.2 Simulation result

2.0.3 $Var(X^T\beta) = \{0\}$

	n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack_1	GCTA_v_jack_2	GCTA_v_corr
1:	50	3.40	1.83	1.32	0	0.59	9.55	9.62	-9.28
2:	75	1.19	0.98	0.56	0	0.46	2.73	2.68	-5.64
3:	100	1.08	0.84	0.57	0	0.44	1.35	1.35	-0.77
4:	150	0.28	0.19	0.32	0	-1.09	0.86	0.67	-64.08
5:	200	0.21	0.12	0.32	0	-1.60	0.78	0.69	-46.14

	n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack_1	GCTA_rr_v_jack_2	GCTA_rr_v_corr
1:	50	3.40	1.83	1.32	0	0.60	9.55	9.47	-3.560
2:	75	1.19	0.98	0.56	0	0.46	2.73	2.68	-5.643
3:	100	1.08	0.84	0.57	0	0.44	1.35	1.35	-0.770
4:	150	0.28	0.19	0.33	0	-0.17	0.62	0.61	-1.204
5:	200	0.21	0.12	0.33	0	0.28	0.61	0.61	-0.041

2.0.4 $Var(X^T\beta) = \{100\}$

	n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack_1	GCTA_v_jack_2	GCTA_v_corr
1:	50	9247	1784	87	0	66	8795	8793	-3687
2:	75	10077	1863	92	0	103	5170	5109	-3122
3:	100	11839	2142	100	0	84	2072	2081	194
4:	150	10953	443	103	0	31	1280	1148	-80475
5:	200	9778	245	98	0	30	725	673	-32124

	n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack_1	GCTA_rr_v_jack_2	GCTA_rr_v_corr
1:	50	9247	1784	87	0	66	8795	8787	-3492
2:	75	10077	1863	92	0	103	5170	5109	-3124
3:	100	11839	2142	100	0	84	2072		
4:	150	11194	414	104	0	103	969		
5:	200	9854	238	98	0	98	616		

3:	2081	194
4:	970	158
5:	616	220

2.0.5 $Var(X^T\beta) = \{8\}$

	n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack_1	GCTA_v_jack_2	GCTA_v_corr
1:	50	90	25.8	8.0	0	8.5	74.1	73.8	-190.67
2:	75	70	13.1	7.5	0	7.5	32.1	31.9	-25.67
3:	100	68	6.3	7.8	0	7.5	13.7	13.8	-0.97
4:	150	70	4.0	8.1	0	8.4	9.2	8.1	-502.59
5:	200	65	2.5	7.9	0	7.6	4.6	4.3	-214.51

	n	MSE	est_var	est_mean	NA_main
1:	50	24.0	24.0	8.0	0
2:	75	13.8	13.8	7.9	0
3:	100	8.6	8.6	8.1	0
4:	150	3.7	3.7	8.0	0
5:	200	2.7	2.7	8.0	0

	n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack_1	GCTA_rr_v_jack_2	GCTA_rr_v_corr
1:	50	90	25.8	8.0	0	8.5	74.1	73.6	-177.35
2:	75	70	13.1	7.5	0	7.5	32.1	31.8	-16.78
3:	100	68	6.3	7.8	0	7.5	13.7	13.8	-0.97
4:	150	70	4.1	8.1	0	8.1	6.9	6.9	1.49
5:	200	65	2.5	7.9	0	7.9	3.9	3.9	1.38

	n	MSE	est_var	est_mean	NA_main
1:	50	23.8	23.9	8.0	0
2:	75	13.7	13.7	7.9	0
3:	100	8.6	8.6	8.1	0
4:	150	3.8	3.8	8.0	0
5:	200	2.7	2.7	8.1	0

2.0.6 correlation test $

	n	MSE	est_var	est_mean	NA_main	cor_main_jack	cor_v_jack_1
1:	50	0.0131	0.0130	0.49	0	0.49	0.0127
2:	75	0.0083	0.0083	0.50	0	0.50	0.0079
3:	100	0.0057	0.0057	0.50	0	0.50	0.0059
4:	150	0.0038	0.0038	0.50	0	0.50	0.0039
5:	200	0.0030	0.0030	0.50	0	0.50	0.0029

```
   cor_v_jack_2 cor_v_corr
1:       0.0128     0.0120
2:       0.0079     0.0076
3:       0.0059     0.0057
4:       0.0039     0.0038
5:       0.0029     0.0029
```

2.1 compare the performance of delete 1 and delete d in variance estimation

The delete-d jackknife varinace estimator is

$$\sqsubseteq_{J(d)} = \frac{n-d}{d} \cdot \frac{1}{S} \sum_{S} (\hat{\theta}_s - \hat{\theta}_{s.})$$

, where $S = \binom{n}{d}$. Note that S could a very large value, so in the following simulation, only $S = 1000$ is used. In Jun Shao's another paper, he proposed an approximation of the deletel-d variance estimation. That is just select m from $S = \binom{n}{d}$ sub-samples and in that paper it recommended $m = n^{1.5}$.

2.1.1 setup

- Independent
- Normal
- $p = \{100, 1000\}$
- $n = \{50, 75, 100, 150, 200, 500, 750, 1000, 1500, 2000\}$
- $d = 0.5 \times n$
- $n_{repeat} = 1000$ for delete d jackknife
- main effect: $Var(X^T\beta) = 8$

2.1.2 GCTA with p = 100

n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack	GCTA_v_jack_var	d	n_sub
50	25.6	25.8	8.0	0	8.5	74.1	8383.8	1.0	NA
75	13.2	13.1	7.5	0	7.5	32.1	685.1	1.0	NA
100	6.2	6.3	7.8	0	7.5	13.7	102.2	1.0	NA
150	4.0	4.0	8.1	0	8.4	9.2	16.4	1.0	NA
200	2.5	2.5	7.9	0	7.6	4.6	2.1	1.0	NA
50	25.6	25.8	8.0	0	45.5	41.2	365.2	0.5	NA
75	13.2	13.1	7.5	0	-177.5	27.1	99.7	0.5	NA
100	6.2	6.3	7.8	0	-237.3	18.5	38.1	0.5	NA
150	4.0	4.0	8.1	0	-13.8	9.4	7.5	0.5	NA
200	2.5	2.5	7.9	0	17.3	5.0	1.4	0.5	NA
50	25.6	25.8	8.0	0	35.1	41.1	366.6	0.5	354
75	13.2	13.1	7.5	0	-107.6	27.0	100.1	0.5	650
100	6.2	6.3	7.8	0	-237.3	18.5	38.1	0.5	1000
150	4.0	4.0	8.1	0	-20.2	9.3	7.0	0.5	1837
200	2.5	2.5	7.9	0	53.4	5.1	1.3	0.5	2828

2.1.3 GCTA with p = 1000

n	MSE	est_var	est_mean	NA_main	GCTA_main_jack	GCTA_v_jack	GCTA_v_jack_var	d
500	2.88	2.91	8.0	0	7.8	4.65	1.08	1.0
750	1.29	1.30	8.0	0	8.0	2.26	0.15	1.0
1000	0.77	0.78	8.0	0	8.0	1.28	0.04	1.0
1500	0.47	0.48	7.9	0	6.7	0.80	0.01	1.0
500	2.88	2.91	8.0	0	-79.1	6.56	1.17	0.5
750	1.29	1.30	8.0	0	-5.9	3.04	0.13	0.5
1000	0.77	0.78	8.0	0	40.8	1.71	0.05	0.5
1500	0.41	0.41	8.0	0	9.9	0.80	0.01	0.5
2000	0.31	0.31	8.0	0	25.6	0.48	0.00	0.5

2.1.4 GCTA_rr_rr with p = 100

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d	n_sub
50	25.6	25.8	8.0	0	8.5	74.1	8378.6	1.0	NA
75	13.2	13.1	7.5	0	7.5	32.1	685.3	1.0	NA
100	6.2	6.3	7.8	0	7.5	13.7	102.2	1.0	NA
150	4.1	4.1	8.1	0	8.1	6.9	8.5	1.0	NA
200	2.5	2.5	7.9	0	7.9	3.9	1.3	1.0	NA
50	25.6	25.8	8.0	0	52.5	40.6	363.1	0.5	NA
75	13.2	13.1	7.5	0	-198.0	26.6	100.2	0.5	NA
100	6.2	6.3	7.8	0	-257.6	18.1	38.6	0.5	NA
150	4.1	4.1	8.1	0	-11.9	9.3	7.5	0.5	NA
200	2.5	2.5	7.9	0	25.4	5.0	1.4	0.5	NA
50	25.6	25.8	8.0	0	35.2	40.5	363.4	0.5	354
75	13.2	13.1	7.5	0	-120.5	26.6	100.8	0.5	650
100	6.2	6.3	7.8	0	-257.6	18.1	38.6	0.5	1000
150	4.1	4.1	8.1	0	-17.0	9.3	7.1	0.5	1837
200	2.5	2.5	7.9	0	76.2	5.1	1.3	0.5	2828

2.1.5 GCTA_rr with p = 1000

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d
500	2.88	2.91	8.0	0	7.8	4.65	1.08	1.0

(continued)

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	d
750	1.29	1.30	8.0	0	8.0	2.26	0.15	1.0
1000	0.77	0.78	8.0	0	8.0	1.28	0.04	1.0
1500	0.48	0.48	7.9	0	8.0	0.62	0.00	1.0
500	2.88	2.91	8.0	0	-79.1	6.56	1.17	0.5
750	1.29	1.30	8.0	0	-5.9	3.04	0.13	0.5
1000	0.77	0.78	8.0	0	40.8	1.71	0.05	0.5
1500	0.41	0.41	8.0	0	11.8	0.80	0.01	0.5
2000	0.31	0.31	8.0	0	24.4	0.48	0.00	0.5

2.1.6 cor with n = 200

n	MSE	est_var	est_mean	NA_main	cor_main_jack	cor_v_jack	d
50	0.01252	0.01265	0.50001	0	0.50432	0.01229	1.0
75	0.00774	0.00782	0.50050	0	0.50323	0.00815	1.0
100	0.00607	0.00613	0.50148	0	0.50334	0.00582	1.0
150	0.00383	0.00385	0.49584	0	0.49709	0.00391	1.0
200	0.00281	0.00284	0.49930	0	0.50027	0.00288	1.0
50	0.01252	0.01265	0.50001	0	5.32154	0.01282	0.5
75	0.00774	0.00782	0.50050	0	3.26213	0.00844	0.5
100	0.00607	0.00613	0.50148	0	2.50378	0.00595	0.5
150	0.00383	0.00385	0.49584	0	1.59064	0.00396	0.5
200	0.00281	0.00284	0.49930	0	1.46439	0.00293	0.5

2.1.7 median with n = 200

n	MSE	est_var	est_mean	NA_main	median_main_jack	median_v_jack	d
50	0.03138	0.03135	-0.00775	0	-0.00775	0.06818	1.0
75	0.02211	0.02212	-0.00212	0	0.05228	0.03113	1.0
100	0.01523	0.01523	-0.00378	0	-0.00378	0.02720	1.0
150	0.01072	0.01072	-0.00279	0	-0.00279	0.01885	1.0
200	0.00804	0.00804	-0.00051	0	-0.00051	0.01614	1.0
50	0.03138	0.03135	-0.00775	0	5.04459	0.03477	0.5
75	0.02211	0.02212	-0.00212	0	8.44376	0.02248	0.5
100	0.01523	0.01523	-0.00378	0	2.68868	0.01587	0.5
150	0.01072	0.01072	-0.00279	0	-1.47581	0.01110	0.5
200	0.00804	0.00804	-0.00051	0	3.96797	0.00827	0.5