Inference methods of high dimensional variance estimator report

Xuelong Wang

2020-01-02

Contents

1 Motivation

2 Subsampling method: Jackknife

2.1 Jackknife Vairance

$S(X_1, \dots, X_n)$ is a statistic of interest, define

$$S_{(i)} = S(X_1, X_{i-1}, X_{i+1} \dots, X_n)$$

as the delete-1 result of S. If we delete each observation, then we will get n $S_{(i)}$. We could use those n subsample to estimate the variance of S on original n dataset as following,

$$\widehat{VAR}\ S(X_1, \dots, X_n) = \frac{n-1}{n} \sum_{i}^{n} (S_{(i)} - S_{(.)})^2$$

, where $S_{(.)} = \frac{\sum_i^n S_{(i)}}{n}$. The variance estimation actually can be considered into a two-step process

1. Estimate the variance of S with n-1 sample:

   $$\widehat{VAR}\ S(X_1, X_{i-1}, X_{i+1} \dots, X_n) := \widetilde{VAR}\ S(X_1, X_{i-1}, X_{i+1} \dots, X_n) = \sum_{i}^{n} (S_{(i)} - S_{(.)})^2,$$

   which could be considered as an modification of the variance estimation corresponding to the dependency of the n delete-1 subsamples. That is originally we need a coefficient $\frac{1}{n-1}$ for sample variance if the samples are indepedent. But the delete-1 subsamples are high dependent to each other, so intuitively the sample variance will underestimate the variance. In order to alleviate the underestimation, it seems that we multiply $n-1$.

$$n-1 \cdot \frac{1}{n-1} \cdot \sum_{i}^{n}(S_{(i)} - S_{(.)})^2 = \sum_{i}^{n}(S_{(i)} - S_{(.)})^2.$$

However, by doing this, the result become overestimated and that will be discussed in the following sections.

2. Modification the variance of $n-1$ samples to n samples by:

$$\widehat{VAR}\ S(X_1, \ldots, X_n) = \frac{n-1}{n}\widetilde{VAR}\ S(X_1, X_{i-1}, X_{i+1} \ldots, X_n).$$

2.2 Bias of Variance estimation

In the Efron 1981's paper, it shows that

$$E\left[\widetilde{VAR}\ S(X_1, X_{i-1}, X_{i+1} \ldots, X_n)\right] \geq VAR\ S(X_1, X_{i-1}, X_{i+1} \ldots, X_n).$$

The details of proof could be found by (Efron and Stein 1981), the idea of the proof is the ANOVA decomposition:

$$\begin{aligned} S\left(X_1, X_2, \cdots, X_n\right) = \mu &+ \sum_i A_i\left(X_i\right) + \sum_{i<i'} B_{i'}\left(X_i, X_{i'}\right) \\ &+ \sum_{i<i'<i''} C_{ii'i''}\left(X_i, X_{i'}, X_{i''}\right) + \cdots + H\left(X_1, X_2, \cdots, X_n\right) \end{aligned}$$

, where $\mu = E(S)$, $A_i\left(x_i\right) = E\left\{S \mid X_i = x_i\right\} - \mu$ and $B_{it}\left(x_i, x_i\right) = E\left(S \mid X_i = x_i, X_i = x_{it}\right\} - E\left\{S \mid X_i = x_i\right) - E\left\{S \mid X_i = x_i\right\} + \mu$. A is the analogy of main effect and B is for the two-term interaction effects. Note that after the ANOVA decompostion, all the terms has mean **0** and correlation **0**. Therefore we have

$$\begin{aligned} S\left(X_1, X_2, \cdots, X_n\right) =&\mu + \frac{1}{n}\sum_i \alpha_i + \frac{1}{n^2}\sum_{i<i'} \beta_{ii} \\ &+ \frac{1}{n^3}\sum_{i<i<i''} \gamma_{ii'i''} + \cdots + \frac{1}{n^n}\eta_{1,2,3,\ldots,n} \end{aligned},$$

where $\begin{array}{l} \alpha_i \equiv \alpha\left(X_i\right) \equiv nA\left(X_i\right), \quad \beta_{ii} = \beta\left(X_i, X_i\right) \equiv n^2 B\left(X_i, X_{i'}\right) \\ \gamma_{ui^*} = \gamma\left(X_i, X_{i'}, X_{i^*}\right) = n^3 C\left(X_i, X_i, X_{i^*}\right), \cdots \end{array}$. Then since all of them are uncorrelated, we could take the variance on both side and have

$$\operatorname{Var} S\left(X_1, X_2, \cdots, X_n\right) = \frac{\sigma_a^2}{n} + \left(\begin{array}{c} n-1 \\ 1 \end{array}\right)\frac{\sigma_\beta^2}{2n^3} + \left(\begin{array}{c} n-1 \\ 2 \end{array}\right)\frac{\sigma_\gamma^2}{3n^5} + \cdots + \frac{\sigma_n^2}{n^{2n}}.$$

It can also shown that

$$\begin{aligned} E\left(\widetilde{\mathrm{VAR}} S\left(X_1, X_2, \cdots, X_{n-1}\right)\right) =&\frac{\sigma_\alpha^2}{n-1} \\ &+ \left(\begin{array}{c} n-2 \\ 1 \end{array}\right)\frac{\sigma_\beta^2}{(n-1)^2} + \left(\begin{array}{c} n-2 \\ 2 \end{array}\right)\frac{\sigma_r^2}{(n-1)^3} + \cdots \end{aligned}$$

, so we have

$$\begin{aligned} &E\left(\widetilde{\mathrm{VAR}} S\left(X_1, X_2, \cdots, X_{n-1}\right)\right\} - \operatorname{Var} S\left(X_1, X_2, \ldots, X_{n-1}\right) \\ &= \frac{1}{2}\left(\begin{array}{c} n-2 \\ 1 \end{array}\right)\frac{\sigma_N^2}{(n-1)^3} + \frac{2}{3}\left(\begin{array}{c} n-2 \\ 2 \end{array}\right)\frac{\sigma_r^2}{(n-1)^s} + \cdots \end{aligned}$$

. Note that bias of the variance comes from the variance of high order interactions. If S is a **linear** functional the emprical cumulative density function, the the bias is 0. However, if it is not, then there will be a non-zero bias. Although Efron suggested a bias correction method, but it is not very practical which I will mention in the next section.

For certain types of S, the bias of the variance will be reduced by increasing of n.

$$E\hat{Var} = Var^{(n)} + \{\frac{n-1}{n}Var^{(n-1)} - Var^{(n)})\} + O(1/n^3),$$

2.2.1 Functionals of emprical distribution function

2.3 Bias correction

2.3.1 Using delete-1-2 method

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

1. One potential issue of this method is that it cannot guarantee the corrected variance is positive. In other words, some times the bias correction is overestimating the bias so that ending a negative variance. This issue is not unexpected, because the correction is based only on the quadratic form.
2. Another issue is the computational time. To calculate the variance correction, one needs to do $\binom{n}{2}$ times iteration, which will be time comsuming for large n.

2.3.2 Delete-d method

The delete-d jackknife method is porposed In (Shao, Wu, and others 1989), The delete-d jackknife varinace estimator is

$$\mathcal{V}_{J(d)} = \frac{n-d}{d} \cdot \frac{1}{N}\sum_{S}(\hat{\theta}_S - \hat{\theta}_{S.})$$

, where $N = \binom{n}{d}$ and S is subset of $x_1, \ldots, x_n$ with size $n-d$. Note that delete-1 jackknife will be a special case of delete-d case variance estimation:

$$\mathcal{V}_{J(1)} = \frac{n-1}{1} \cdot \frac{1}{N}\sum_{S}(\hat{\theta}_S - \hat{\theta}_{S.})$$

where $N = \binom{n}{1} = n$. But how could we explain the 2-steps estimation in Eforn's 1989 paper?

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 $m = n^{1.5}$.

2.3.2.1 An example of delete-d and delete-1: median

$S_n = F_n^{-1}(1/2)$ The simulation setup is following

n	MSE	est_var	est_mean	NA_main	median_main_jack	median_v_jack	relative_ratio	relative_ratio_var	d
50	0.03	0.03	-0.01	0	-0.01	0.07	1.17	1.05	1
75	0.02	0.02	0.00	0	0.05	0.03	0.41	0.09	1
100	0.02	0.02	0.00	0	0.00	0.03	0.79	0.20	1
150	0.01	0.01	0.00	0	0.00	0.02	0.76	0.13	1
200	0.01	0.01	0.00	0	0.00	0.02	1.01	0.18	1
50	0.03	0.03	-0.01	0	0.42	0.03	0.11	0.01	25
75	0.02	0.02	0.00	0	1.15	0.02	0.02	0.01	38
100	0.02	0.02	0.00	0	0.46	0.02	0.04	0.00	50
150	0.01	0.01	0.00	0	-0.48	0.01	0.04	0.00	75
200	0.01	0.01	0.00	0	2.34	0.01	0.03	0.00	100

2.4 functional of distribution functions

2.5 Jackknife variance estimation on high dimension signal estimation

3 Jackknife variance estimation on high dimension signal estimation

Different methods have their own ### Jackknife variance estimation's bias and sample size n

3.0.0.1 setup

- Independent
- Normal
- $p = \{100, 1000\}$
- $n = \{50, 75, 100, 150, 200, 500, 750, 1000, 1500\}$
- $d = \{0.5, 0.75\} \times n$ or $d = 25$
- $n_{repeat} = n^{1.5}$ for delete d jackknife and $n_{repeat} = n$ for delete 1 jackknife
- main effect: $Var(X^T\beta) = 8$

3.0.0.2 result

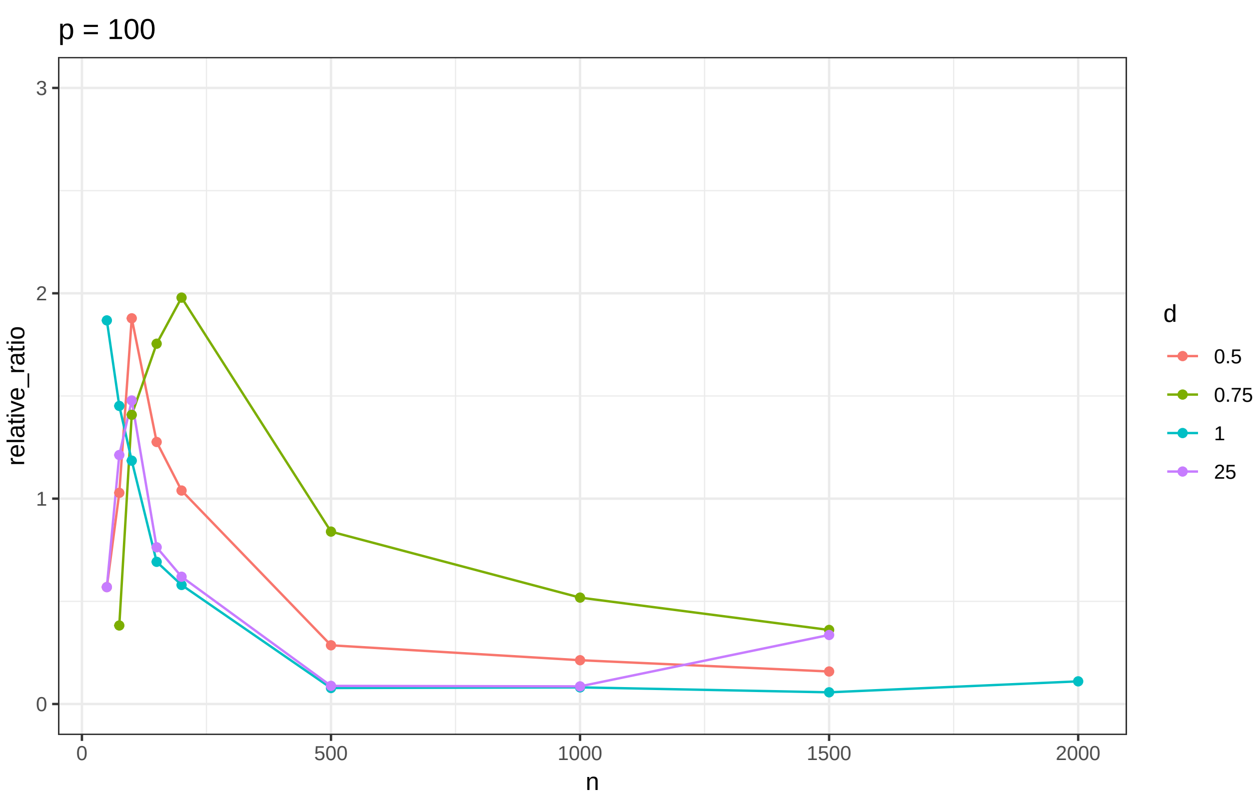
Based on the previous simulation results, we find there is a bias among all the jackknife variance estimation. Based on the Efron's result, the overestimation is because the statistics S is not a smooth function of the distribution function, so that the correct coefficient actually inflate the variance estimation.

The following reuslt is trying to see the relation between the bias and the sample size n

Note: 1. For delete-1 jackknife, the variance estimation becomes better when the sample size is increasing 1. However, for delete-d, it does not show the similar pattern, the relative ratio becomew worse when n is large, which is what we expected. One factor could the number of covariates, that is when p is large then it will be hard to make the jackknife work well??

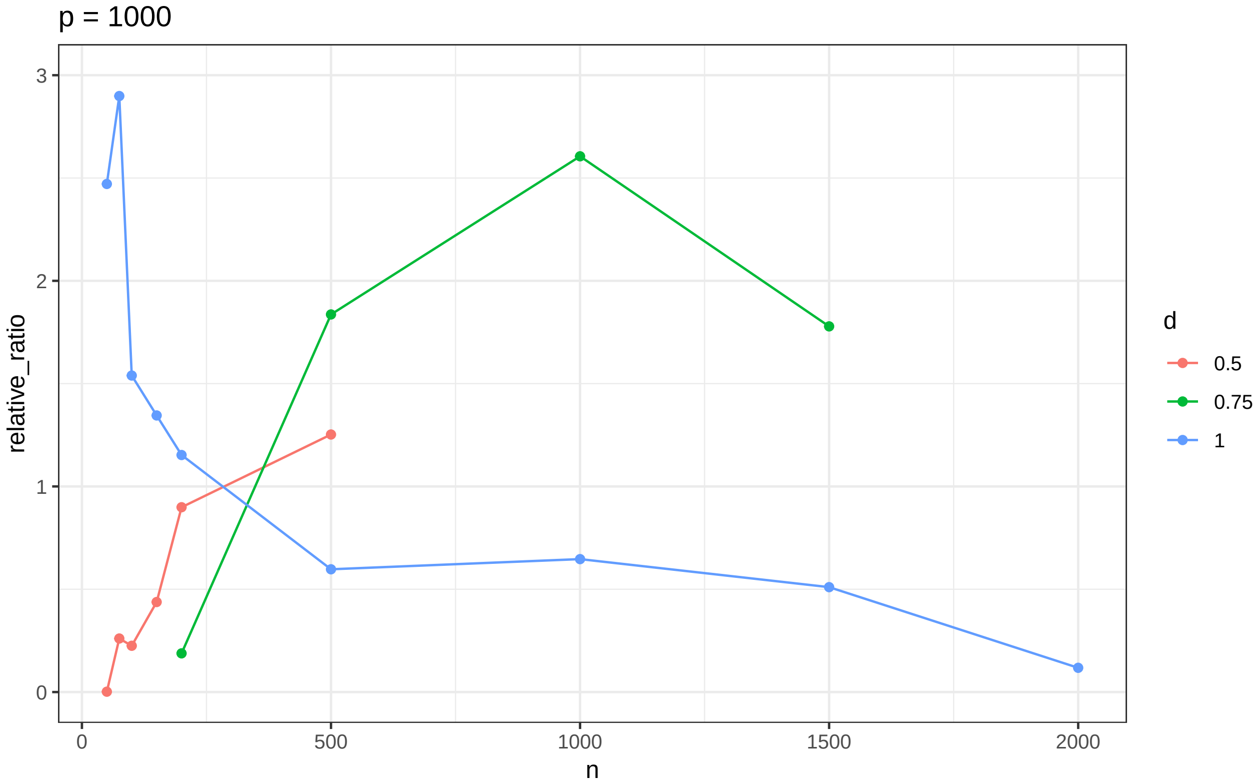
3.0.0.3 GCTA with p = 100

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	relative_ratio	relative_ratio_var	N	d
50	25.58	25.84	8.0	0	35.2	40.54	363.40	0.57	14.07	100	0.50
75	13.18	13.10	7.5	0	-120.5	26.57	100.83	1.03	7.70	100	0.50
100	6.25	6.29	7.8	0	-257.6	18.09	38.63	1.88	6.14	100	0.50
150	4.07	4.09	8.1	0	-17.0	9.31	7.11	1.28	1.74	100	0.50
200	2.48	2.49	7.9	0	76.2	5.08	1.31	1.04	0.53	100	0.50
500	0.83	0.83	8.1	0	4.9	1.06	0.03	0.29	0.04	100	0.50
1000	0.33	0.32	8.1	0	95.6	0.39	0.00	0.21	0.01	100	0.50
1500	0.21	0.20	8.1	0	85.9	0.23	0.00	0.16	0.00	99	0.50
50	25.58	25.84	8.0	0	53.8	21.44	75.63	-0.17	2.93	100	0.75
75	13.18	13.10	7.5	0	-154.6	18.11	39.61	0.38	3.02	100	0.75
100	6.25	6.29	7.8	0	-403.7	15.14	20.26	1.41	3.22	100	0.75
150	4.07	4.09	8.1	0	-342.2	11.27	8.25	1.75	2.02	100	0.75
200	2.48	2.49	7.9	0	-243.4	7.41	2.18	1.98	0.88	100	0.75
500	0.83	0.83	8.1	0	-98.3	1.52	0.05	0.84	0.06	100	0.75
1000	0.33	0.32	8.1	0	246.4	0.49	0.00	0.52	0.01	100	0.75
1500	0.21	0.20	8.1	0	128.3	0.28	0.00	0.36	0.00	100	0.75
50	25.58	25.84	8.0	0	8.5	74.10	8378.63	1.87	324.31	100	1.00
75	13.18	13.10	7.5	0	7.5	32.11	685.27	1.45	52.32	100	1.00
100	6.25	6.29	7.8	0	7.5	13.74	102.20	1.18	16.26	100	1.00
150	4.07	4.09	8.1	0	8.1	6.92	8.46	0.69	2.07	100	1.00
200	2.48	2.49	7.9	0	7.9	3.93	1.32	0.58	0.53	100	1.00
500	0.83	0.83	8.1	0	8.1	0.89	0.02	0.08	0.03	100	1.00
1000	0.33	0.32	8.1	0	8.1	0.35	0.00	0.08	0.00	100	1.00
1500	0.21	0.20	8.1	0	8.1	0.22	0.00	0.06	0.00	99	1.00
2000	0.14	0.14	8.0	0	8.1	0.15	0.00	0.11	0.00	100	1.00
50	25.58	25.84	8.0	0	35.2	40.54	363.40	0.57	14.07	100	25.00
75	13.18	13.10	7.5	0	-60.2	28.98	185.76	1.21	14.18	100	25.00
100	6.25	6.29	7.8	0	-98.7	15.58	52.81	1.48	8.40	100	25.00
150	4.07	4.09	8.1	0	3.1	7.21	7.25	0.76	1.77	100	25.00
200	2.48	2.49	7.9	0	26.6	4.03	1.29	0.62	0.52	100	25.00
500	0.83	0.83	8.1	0	6.6	0.90	0.03	0.09	0.03	100	25.00
1000	0.33	0.32	8.1	0	8.7	0.35	0.00	0.09	0.00	100	25.00
1500	0.15	0.16	8.0	0	11.4	0.21	0.00	0.34	0.00	20	25.00



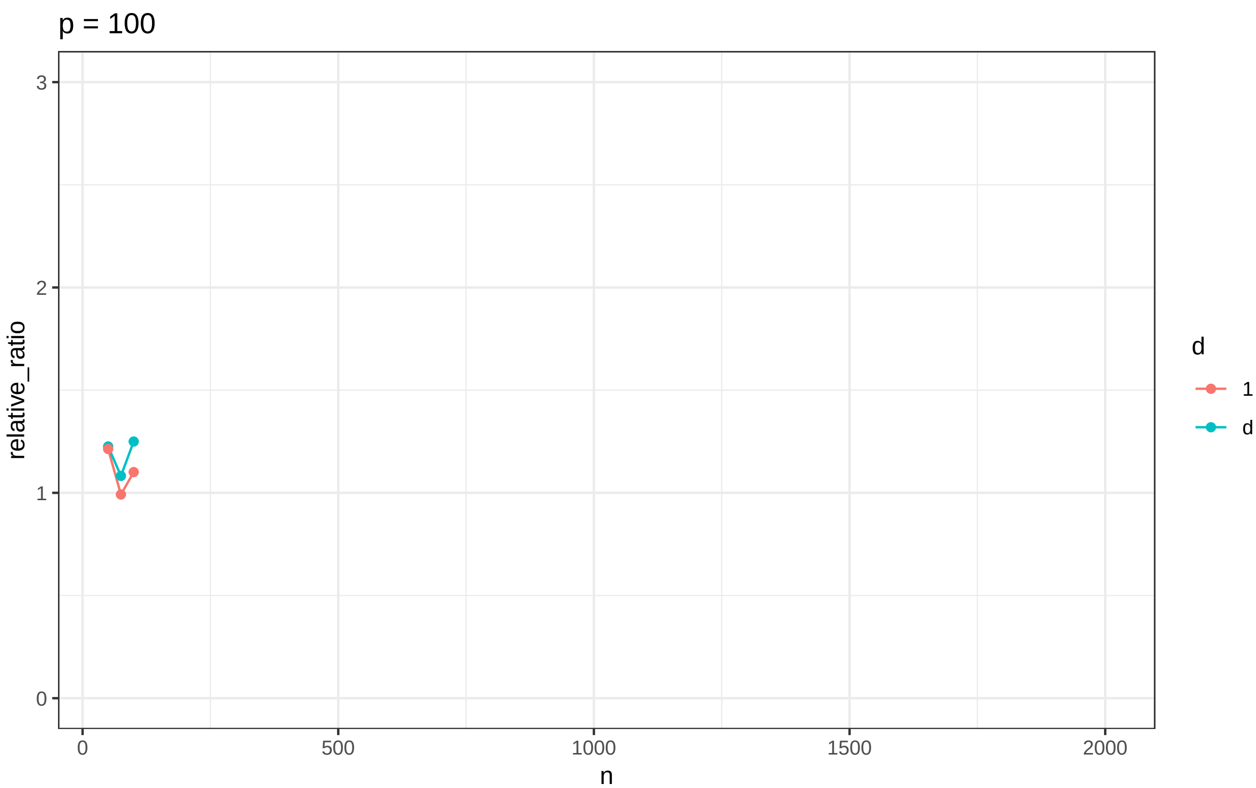
3.0.0.4 GCTA with p = 1000

n	MSE	est_var	est_mean	NA_main	GCTA_rr_main_jack	GCTA_rr_v_jack	GCTA_rr_v_jack_var	relative_ratio	relative_ratio_var	N	d
50	55.85	55.27	9.1	0	136.6	55.37	518.78	0.00	9.39	100	0.50
75	34.53	34.17	7.2	0	-289.0	43.07	340.53	0.26	9.96	100	0.50
100	32.50	32.08	7.1	0	-268.2	39.29	243.29	0.23	7.58	100	0.50
150	21.94	21.72	7.3	0	-267.1	31.23	67.58	0.44	3.11	100	0.50
200	13.23	13.25	7.7	0	-340.4	25.15	39.00	0.90	2.94	100	0.50
500	2.88	2.91	8.0	0	-1102.5	6.56	1.05	1.25	0.36	100	0.50
50	55.85	55.27	9.1	0	263.1	25.39	109.31	-0.54	1.98	100	0.75
75	34.53	34.17	7.2	0	-345.6	21.20	63.47	-0.38	1.86	100	0.75
100	32.50	32.08	7.1	0	-464.3	19.65	36.61	-0.39	1.14	100	0.75
150	21.94	21.72	7.3	0	-579.1	17.45	14.49	-0.20	0.67	100	0.75
200	13.23	13.25	7.7	0	-637.8	15.74	9.71	0.19	0.73	100	0.75
500	2.88	2.91	8.0	0	-1993.2	8.26	1.24	1.84	0.43	100	0.75
1000	0.77	0.78	8.0	0	1213.7	2.80	0.09	2.61	0.11	100	0.75
1500	0.46	0.46	8.0	0	178.8	1.27	0.01	1.78	0.02	81	0.75
50	55.85	55.27	9.1	0	8.2	191.85	61399.17	2.47	1110.88	100	1.00
75	34.53	34.17	7.2	0	6.8	133.23	13844.25	2.90	405.13	100	1.00
100	32.50	32.08	7.1	0	6.6	81.45	4350.37	1.54	135.63	100	1.00
150	21.94	21.72	7.3	0	7.5	50.94	759.61	1.35	34.97	100	1.00
200	13.23	13.25	7.7	0	7.5	28.52	121.00	1.15	9.14	100	1.00
500	2.88	2.91	8.0	0	7.8	4.65	1.08	0.60	0.37	100	1.00
1000	0.77	0.78	7.9	0	8.0	1.28	0.04	0.65	0.05	99	1.00
1500	0.41	0.41	8.0	0	8.0	0.62	0.00	0.51	0.01	100	1.00
2000	0.33	0.34	8.0	0	8.1	0.38	0.00	0.12	0.00	62	1.00



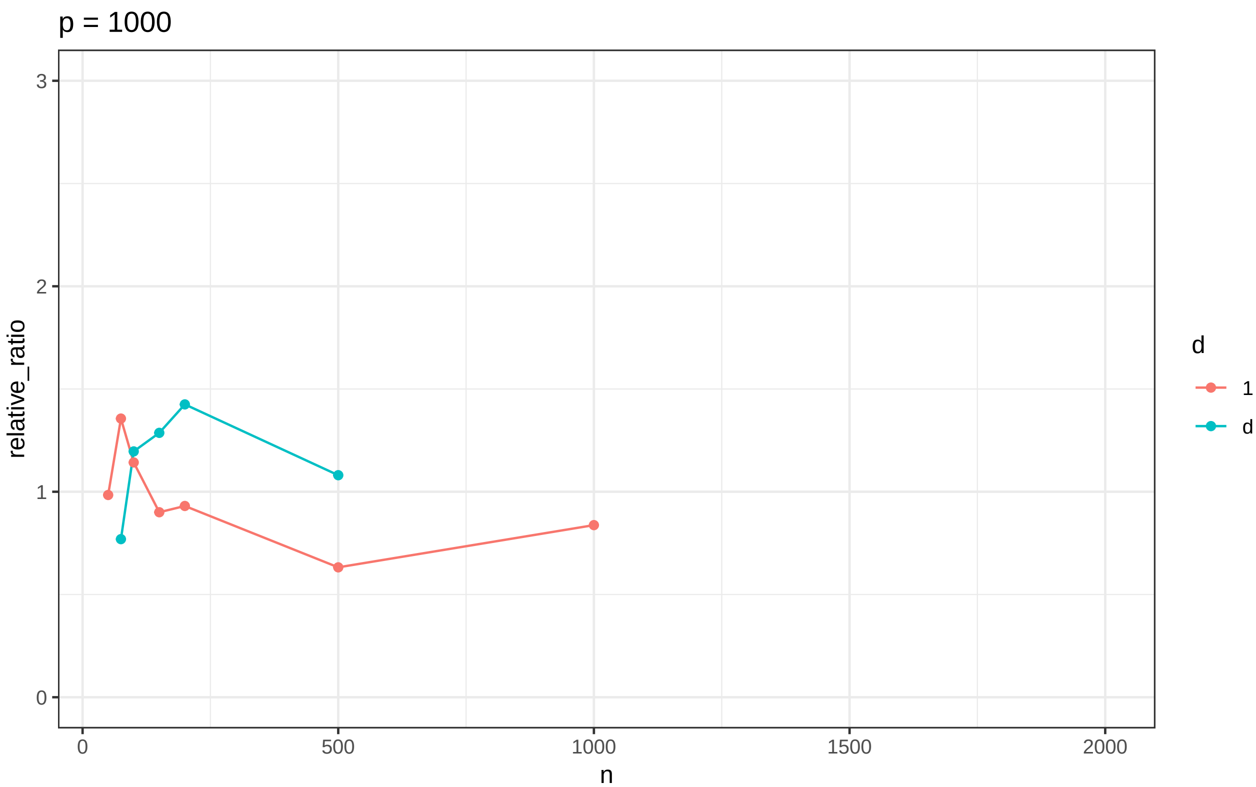
3.0.0.5 Eg with p = 100

n	MSE	est_var	est_mean	NA_main	EigenPrism_main_jack	EigenPrism_v_jack	EigenPrism_v_jack_var	relative_ratio	relative_ratio_var	N	d
50	21.6	21.7	8.4	0	-366.9	48.4	463.04	1.23	21.3	100	0.5
75	12.2	12.3	7.9	0	-463.3	25.6	77.41	1.08	6.3	100	0.5
100	7.1	7.2	8.0	0	-710.9	16.2	33.41	1.25	4.6	100	0.5
150	NaN	NA	NaN	100	NaN	8.4	4.89	NA	NA	100	0.5
200	NaN	NA	NaN	100	NaN	5.0	0.95	NA	NA	100	0.5
500	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	0.5
1000	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	0.5
1500	NaN	NA	NaN	99	NaN	NaN	NA	NA	NA	99	0.5
50	21.6	21.7	8.4	0	7.6	48.1	454.15	1.21	20.9	100	1.0
75	12.2	12.3	7.9	0	7.3	24.5	95.00	0.99	7.7	100	1.0
100	7.1	7.2	8.0	0	7.1	15.1	40.84	1.10	5.7	100	1.0
150	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0
200	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0
500	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0
1000	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0
1500	NaN	NA	NaN	99	NaN	NaN	NA	NA	NA	99	1.0
2000	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0



3.0.0.6 Eg with p = 1000

n	MSE	est_var	est_mean	NA_main	EigenPrism_main_jack	EigenPrism_v_jack	EigenPrism_v_jack_var	relative_ratio	relative_ratio_var	N	d
50	139.7	125.05	12.0	0	-784.3	113.0	3039.25	-0.10	24.30	100	0.5
75	57.4	57.44	8.7	0	-1435.4	101.6	2347.86	0.77	40.87	100	0.5
100	39.0	39.36	7.7	0	-1558.8	86.5	853.63	1.20	21.69	100	0.5
150	20.9	21.06	7.9	0	-1415.7	48.2	142.18	1.29	6.75	100	0.5
200	12.4	12.56	7.9	0	-2090.4	30.4	49.02	1.43	3.90	100	0.5
500	3.0	2.99	8.0	0	-3215.4	6.2	0.86	1.08	0.29	100	0.5
50	139.7	125.05	12.0	0	9.3	248.1	18031.74	0.98	144.19	100	1.0
75	57.4	57.44	8.7	0	6.9	135.3	4335.86	1.36	75.48	100	1.0
100	39.0	39.36	7.7	0	6.1	84.3	954.62	1.14	24.25	100	1.0
150	20.9	21.06	7.9	0	7.3	40.0	120.60	0.90	5.73	100	1.0
200	12.4	12.56	7.9	0	7.2	24.2	39.43	0.93	3.14	100	1.0
500	3.0	2.99	8.0	0	7.7	4.9	0.83	0.63	0.28	100	1.0
1000	0.8	0.81	8.0	0	8.0	1.5	0.03	0.84	0.04	99	1.0
1500	NaN	NA	NaN	100	NaN	NaN	NA	NA	NA	100	1.0
2000	NaN	NA	NaN	62	NaN	NaN	NA	NA	NA	62	1.0



4 Subsampling method: bootstrap

4.1 non-parameteric bootstrap

parametric bootstrap

Efron, Bradley, and Charles Stein. 1981. "The Jackknife Estimate of Variance." *The Annals of Statistics*. JSTOR, 586–96.

Shao, Jun, CF Jeff Wu, and others. 1989. "A General Theory for Jackknife Variance Estimation." *The Annals of Statistics* 17 (3). Institute of Mathematical Statistics: 1176–97.