Jackknife variance estimation corrections

Xuelong Wang

2019-11-08

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,

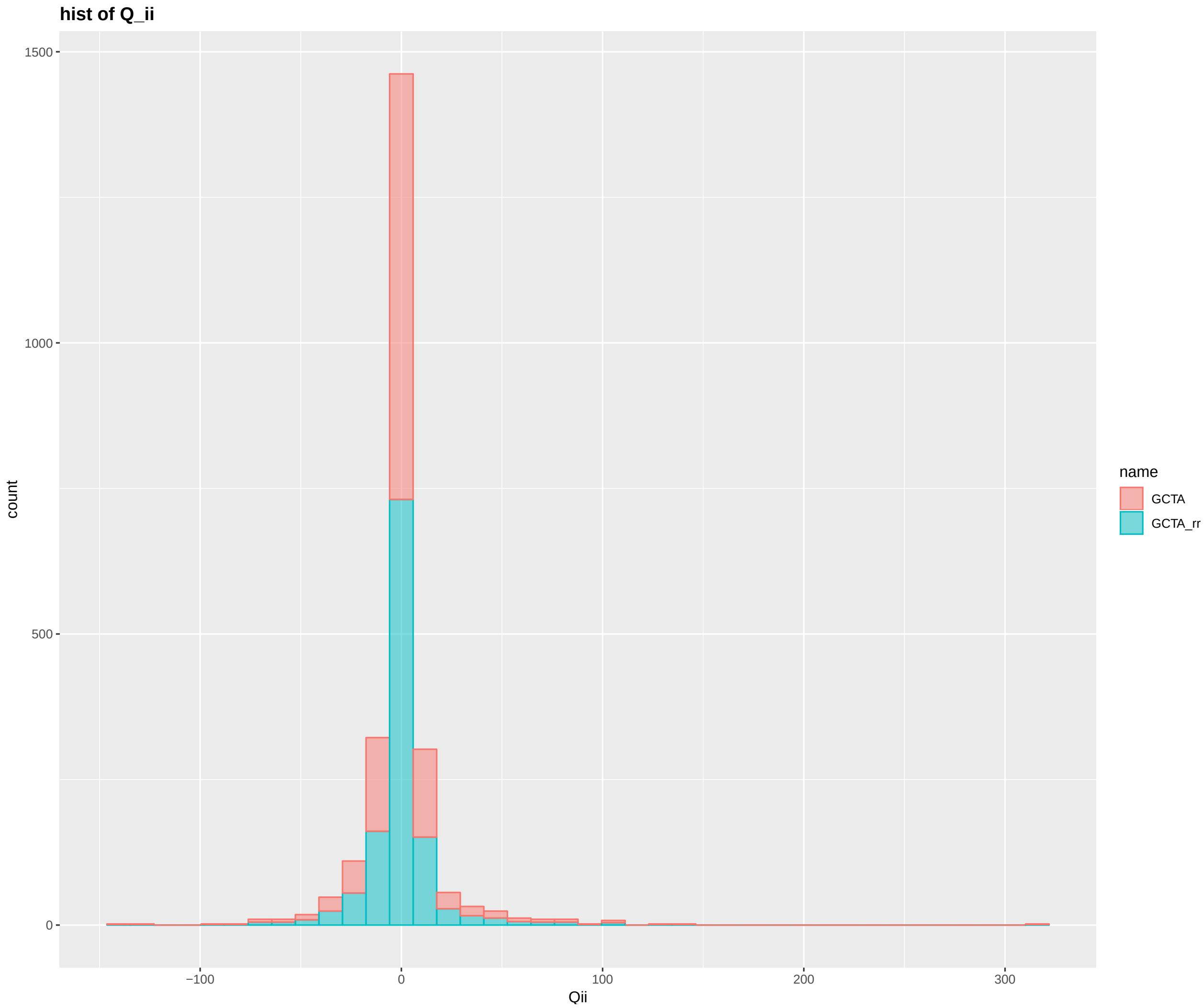
1. when $n < p$ case, those two methods' results are very closed to each other.
2. when $n > p$ case, in terms of effect estimation and jackknife variance estimation those two methods's reuslts are similar to each other. But for the variance corrections are quite different. That is the statistics Q of 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: $Var(X^T\beta) = 8$

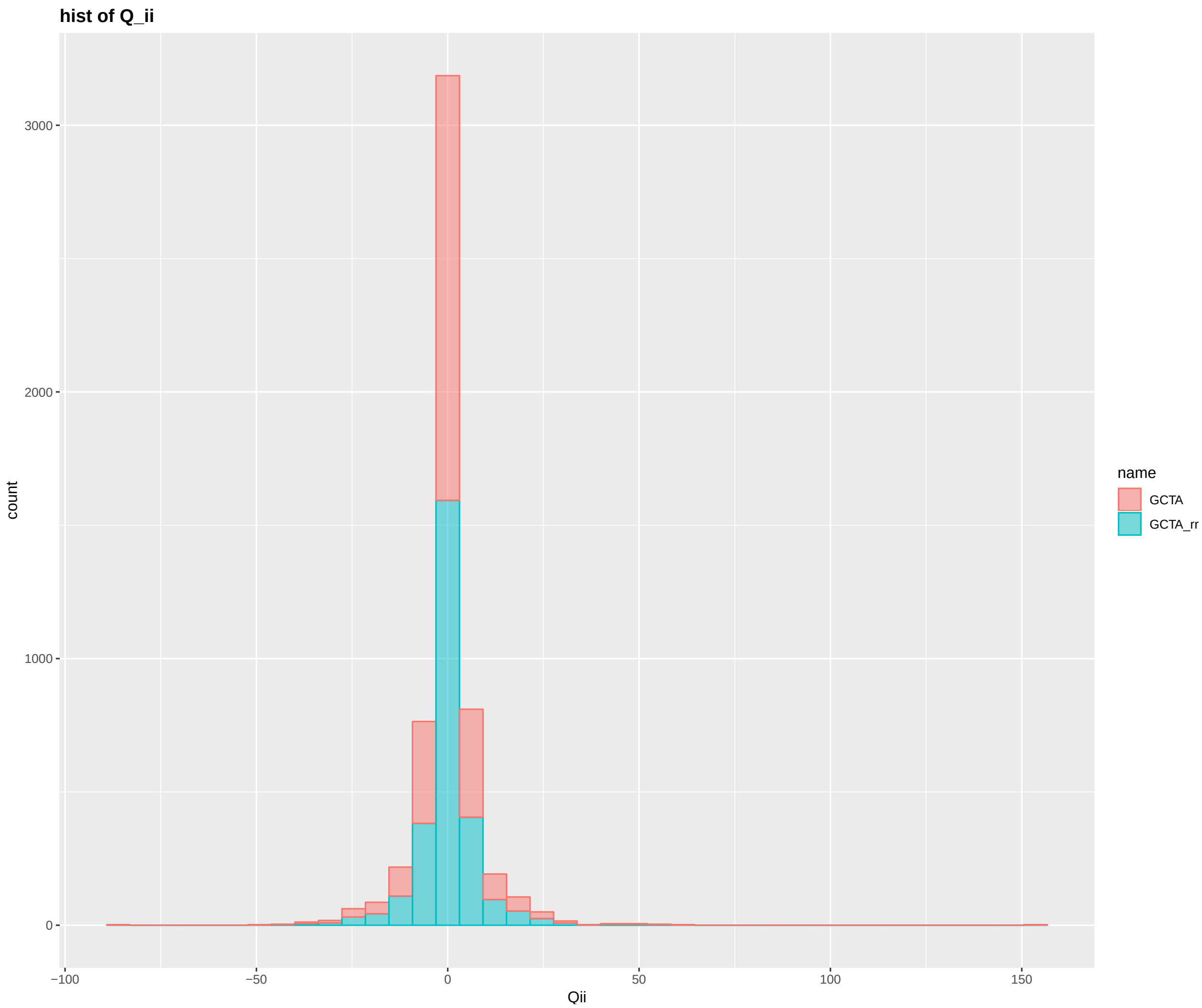
2.0.2 $\mathbf{n = 50}$

```
        main_effect_GCTA v_jack v_jack_corr
GCTA                6.29    124       -92.6
GCTA_rr             6.29    124       -92.6
```



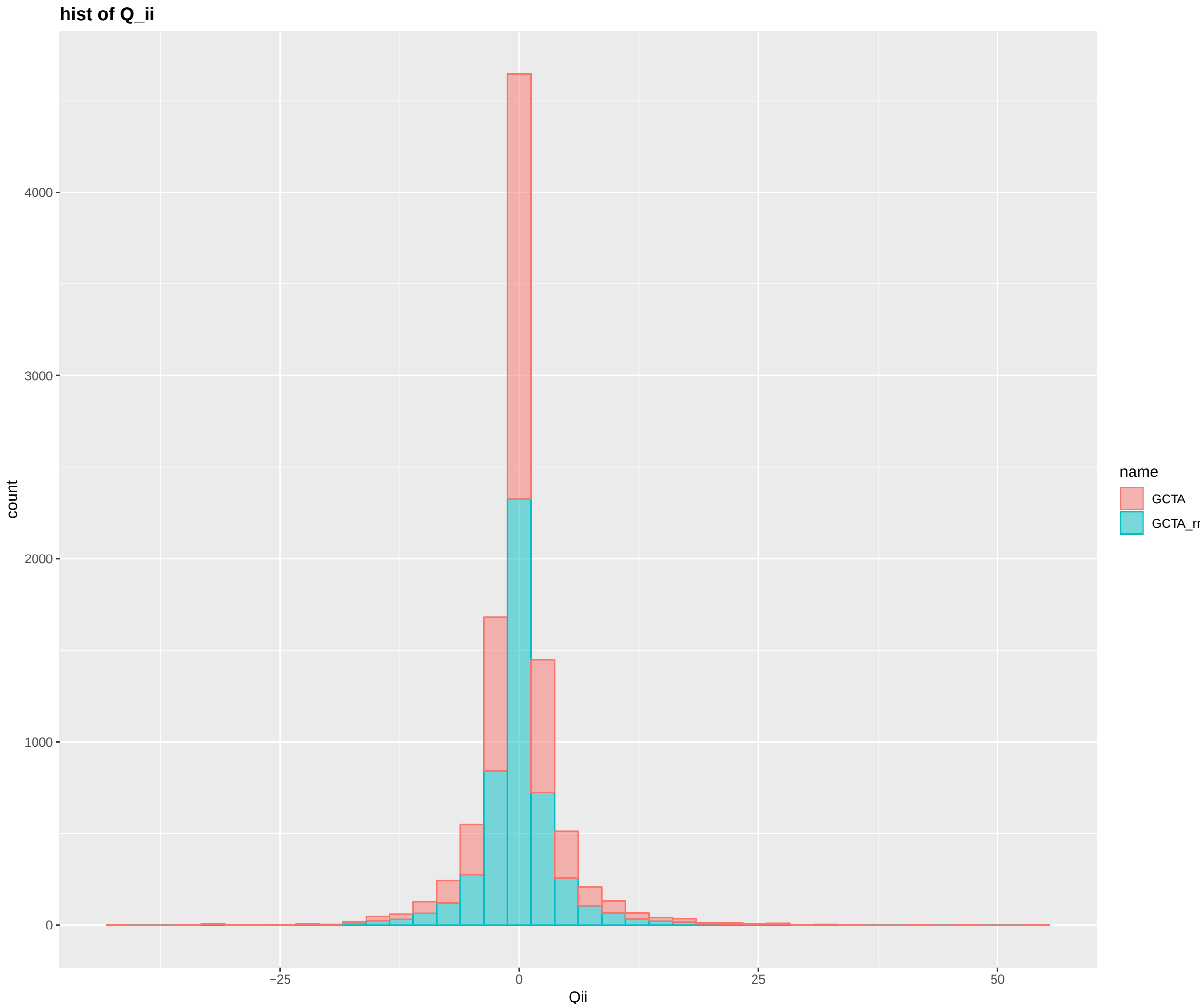
2.0.3 $n = 75$

```
       main_effect_GCTA v_jack v_jack_corr
GCTA                3.7     28        -9.4
GCTA_rr             3.7     28        -9.4
```



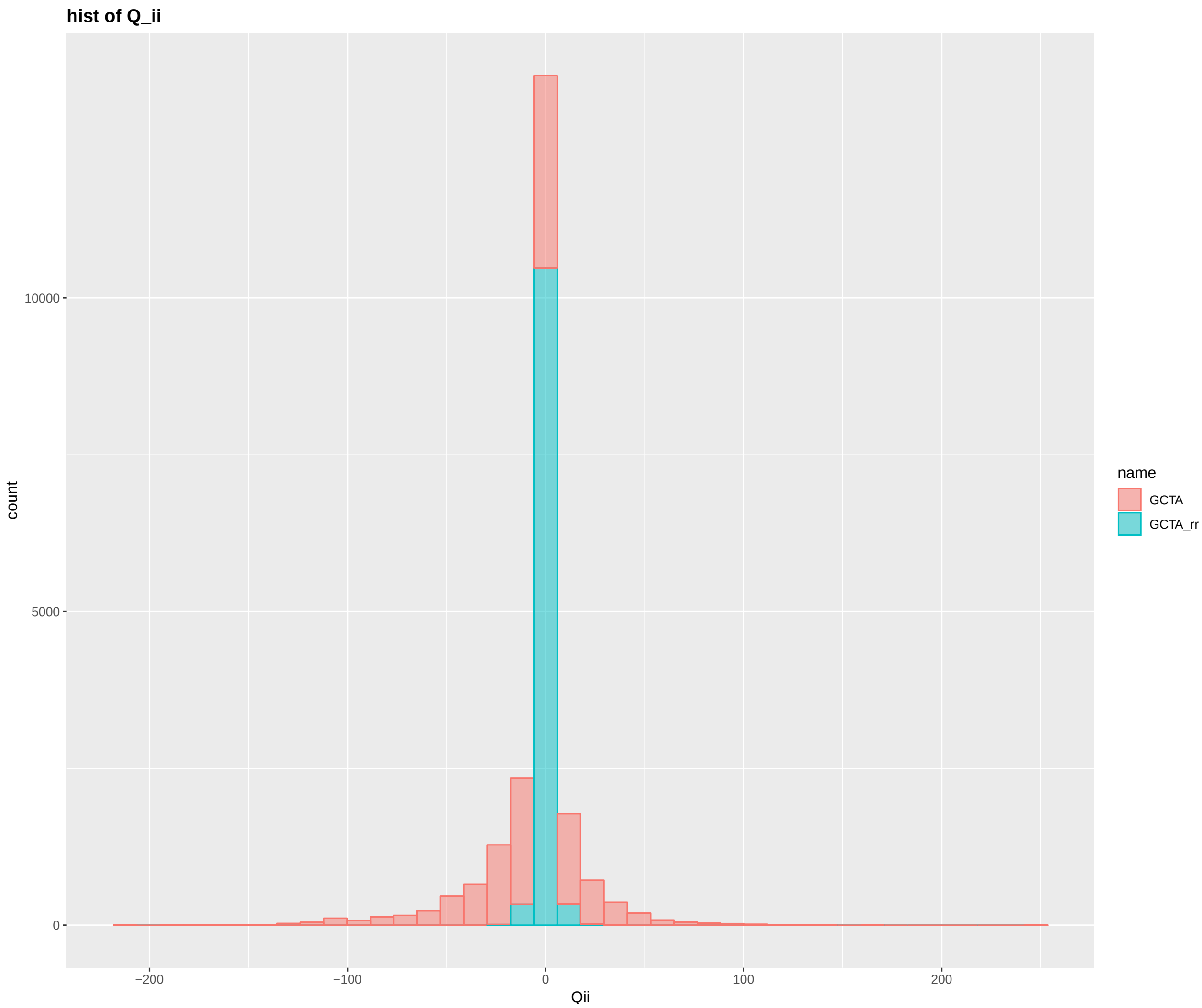
2.0.4 n = 100

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	8.41	9.68	-1.08
GCTA_rr	8.41	9.68	-1.08



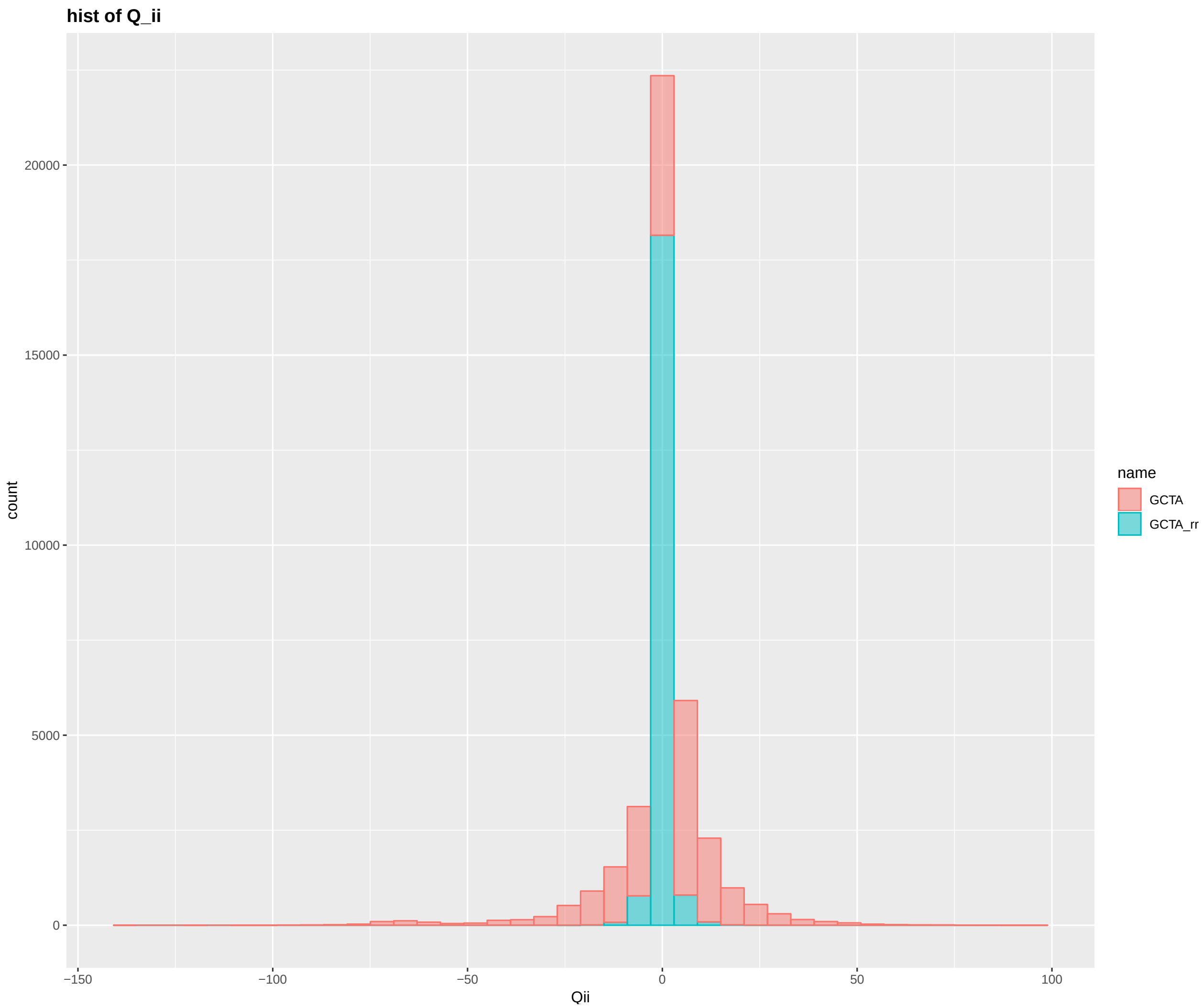
2.0.5 n = 150

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	8.47	8.73	-478.33
GCTA_rr	8.50	6.35	1.57



2.0.6 n = 200

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	8.56	3.96	-137.52
GCTA_rr	8.58	3.64	1.61

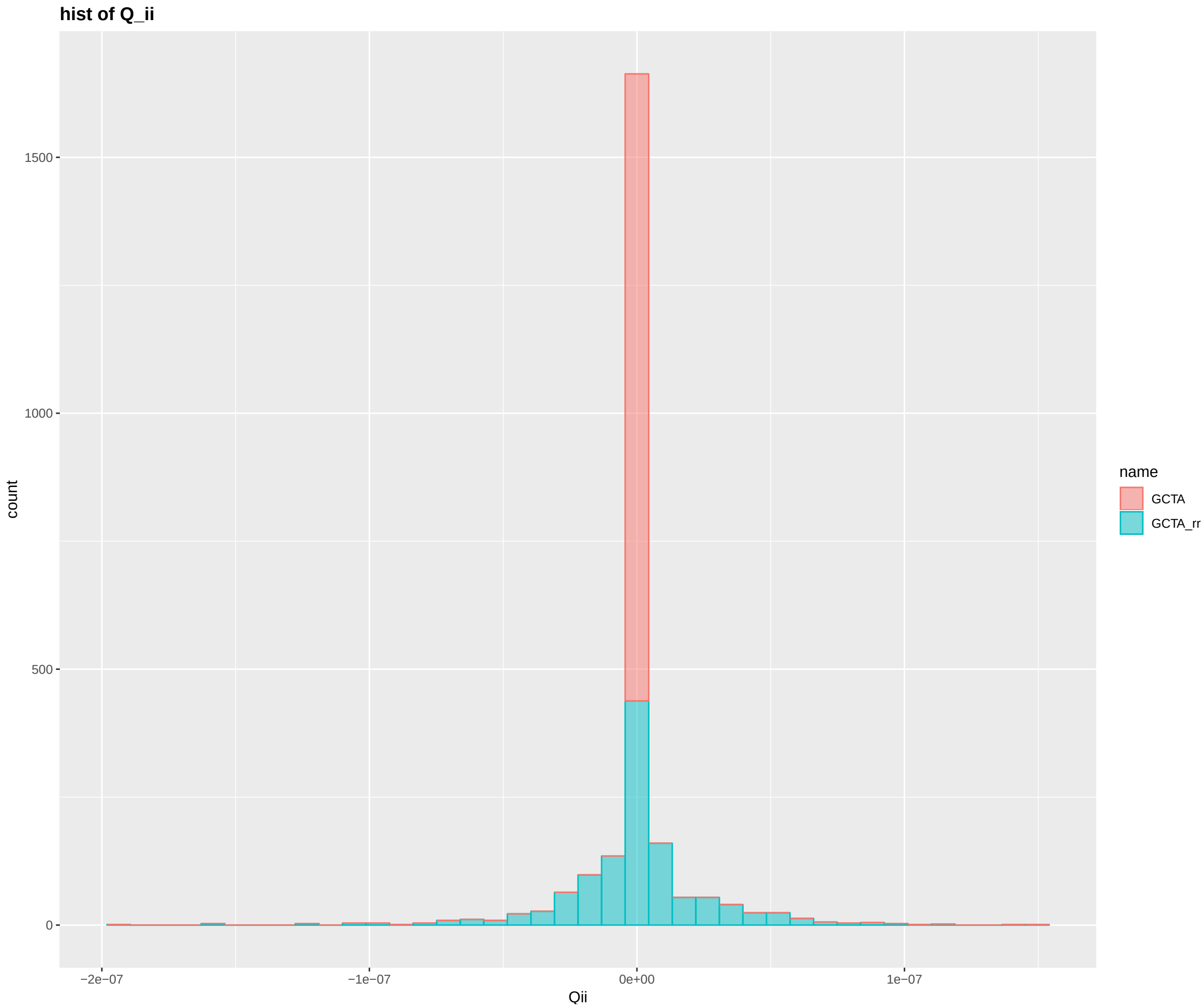


2.0.7 setup

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

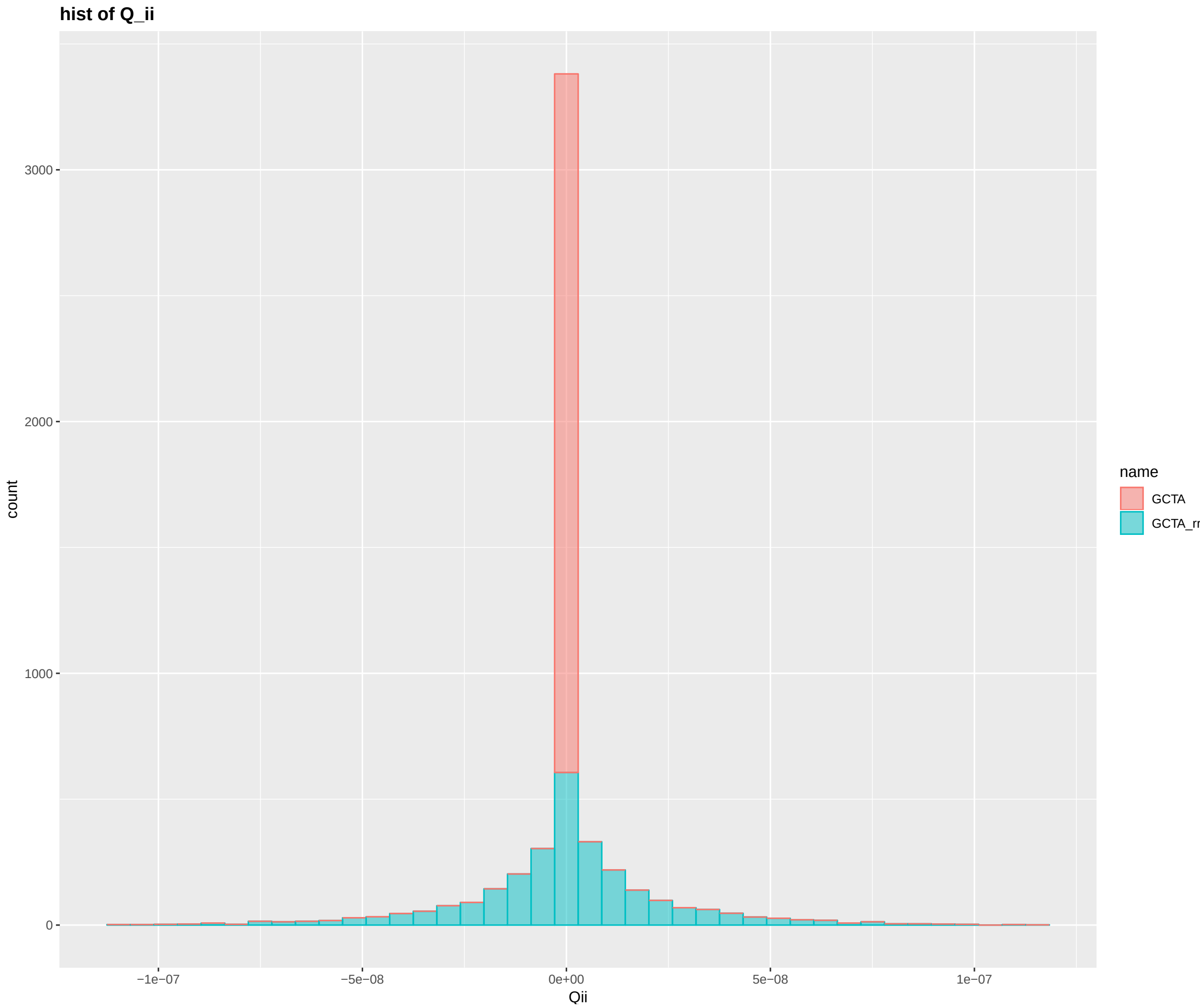
2.0.8 n = 50

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	0.0e+00	0.00e+00	0.0e+00
GCTA_rr	7.1e-07	2.44e-14	2.4e-14



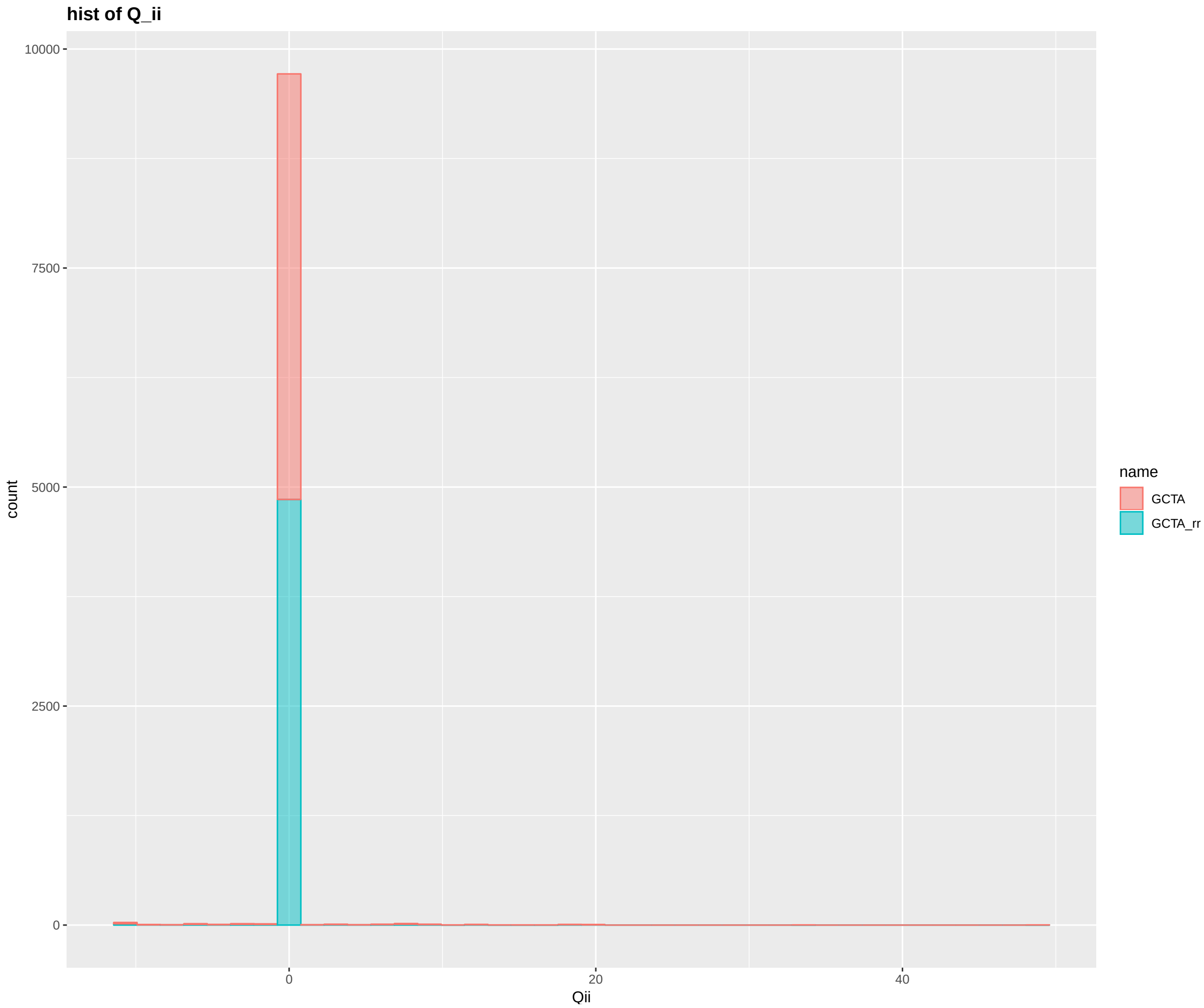
2.0.9 n = 75

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	0.0e+00	0.00e+00	0.00e+00
GCTA_rr	9.5e-07	1.72e-14	1.69e-14



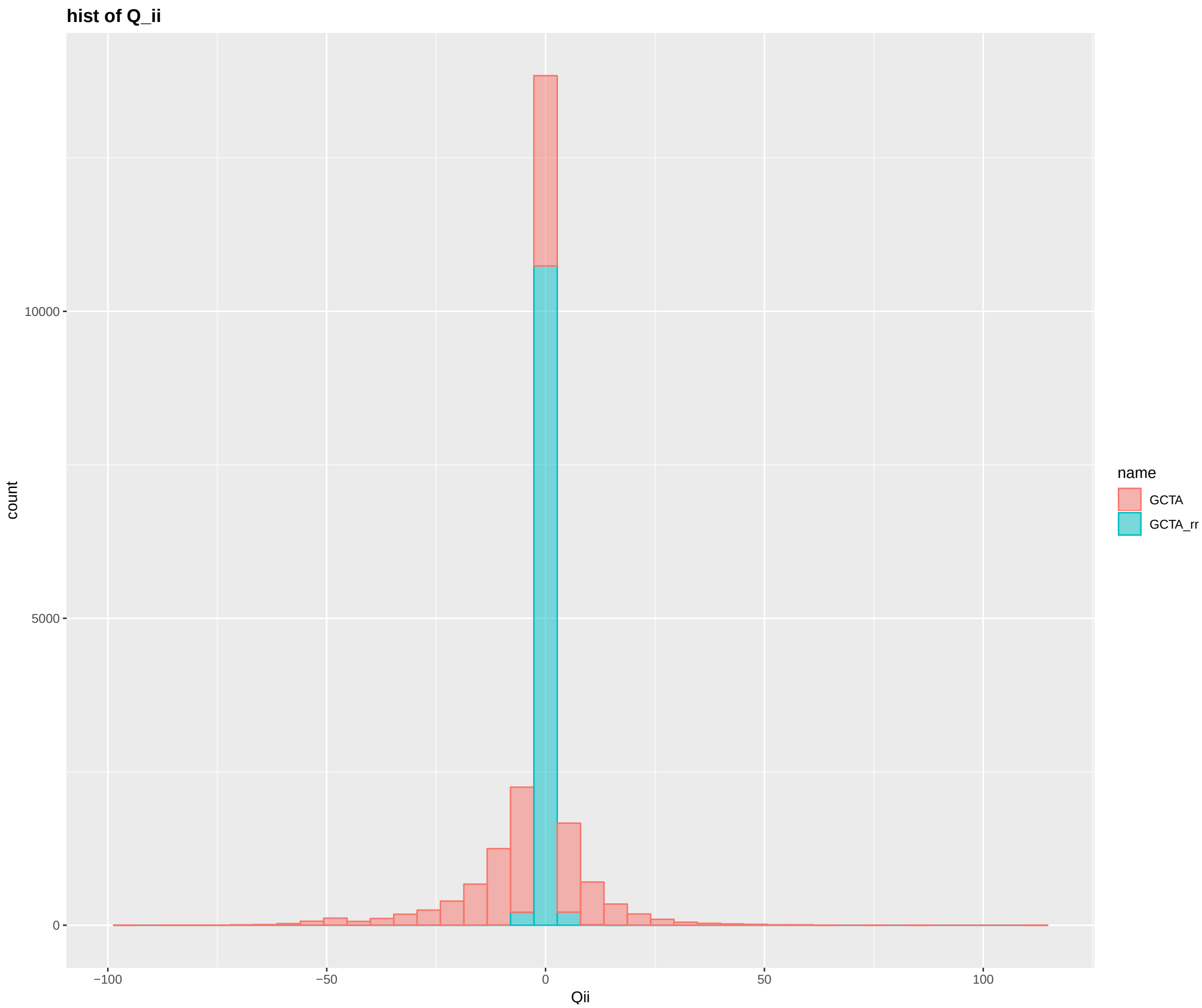
2.0.10 $n = 100$

```
         main_effect_GCTA v_jack v_jack_corr
GCTA              0.0e+00 0.0104       -1.11
GCTA_rr           7.7e-07 0.0104       -1.11
```



2.0.11 $n = 150$

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	0.595	1.305	-100.040
GCTA_rr	0.609	0.807	0.191



2.0.12 $n = 200$

	main_effect_GCTA	v_jack	v_jack_corr
GCTA	0.106	0.696	-57.22
GCTA_rr	0.118	0.482	-0.95