ConGen Report
An Attempt to Improve CKMR Estimator by Multi-Siblings Set

Longxiao Chang, Yakup Ferit Ayranci

July 14, 2024

Abstract

The effective population size (N_e) is a model to regulate the real population to fit the Wright-Fisher model and conduct further prediction. It can be used to guiding conservation efforts. CKMR (the close-kin mark-recapture method) provide an efficient way to estimate N_e by sibling pairs in samples. However, it generally disregarded the cases of unequal distribution of offspring, in which there are more multi-siblings sets (more than two individuals are the offspring of same parent). By using multi-siblings sets, we generated a new CKMR N_e estimator based on the linear relation between siblings sets number and power of N_e. The new estimator shows robustness in both the unequal and equal populations.

Contents

1 Introduction 3

2 Models and Methods 4
2.1 Population simulation . . . 4
2.2 Estimation of real effective population size . . . 4
2.3 Estimator by real siblings sets . . . 5
2.4 Improved CKMR by observed siblings sets . . . 5

3 Results 6
3.1 Estimator by different observed siblings set . . . 6
3.2 Weighted linear regression of observed siblings set . . . 6

4 Discussion 7
4.1 Real siblings set provides a bad estimator . . . 7
4.2 Siblings set under unbalanced population . . . 7
4.3 Estimate effective population size from the linear combination and exponentiation . . . 7

Appendix A Real number of sibling pairs and Siblings Sets 9

Appendix B Real and observed siblings set 10

Appendix C Documentation of Code 11

1 Introduction

N_e of a population refers to the number of individuals that actively participate in producing the next generation. The contribution of genetic material to the next generation is strongly linked with the genetic diversity of the population and both the genetic diversity and the N_e are widely used to estimate the population's viability (Scheiner, 2024).

It is crucial to make informed decisions when planning conservation programs because of the limited resources available, and having a concrete estimation of a population's viability is quite helpful in that regard. If said programs can estimate the Ne effectively, and relate it to the population viability, attention can be focused on the topics deserving it most.

There are various ways to estimate the Ne for a given population, and they all have positives and negatives which should be weighed to decide on the selection of one. Among the methods, including comparing the amount of excess heterozygosity in a population (Pudovkin et al., 1996), checking the amount of linkage disequilibrium between a neutral and an unlinked selected locus (Hill et al., 1981), or measuring the amount of genetic relatedness or relationship between individuals, we chose to focus on the latter and improve its estimation power on populations with a particular structure.

CKMR (the close-kin mark-recapture method) estimates the effective population size by the proportion of sibling pairs in the sample (Bravington et al., 2016b; Waples and Feutry, 2022). In a haploid Wright-Fisher population (population size N), the probability that two individuals are siblings is given by:

$$p_{sib} = \frac{1}{N} \tag{1}$$

As a result, if we take a sample (with size n) from the population, and find Y_2 pairs of siblings, then the estimation of sibling proportion is:

$$\hat{p}_{sib} = \frac{Y_2}{R(n)} \tag{2}$$

where $R(n) = \frac{n(n-1)}{2}$ is the number of potential pairs in the sample. Thus, we could estimate population size by the sibling pairs in the sample. Furthermore, because the frequency of siblings in the population shows effective reproduction in the last generation and genetic biodiversity in the sibling's generation, the same method could be used to estimate effective population size (N_e) in non-WF populations (equation 3).

$$\hat{N}_e = \frac{R_(n)}{Y_2} \tag{3}$$

This is an efficient and stable method estimator for recent effective population size, especially for huge populations like marine fishes (Bravington et al., 2016a; Hillary et al., 2018; Ruzzante et al., 2019). The rapid development of sequencing technology has improved the accuracy of the CKMR method by increasing the resolution of sibling identification.

However, theoretically, the CKMR method is based on an implicit assumption that all the sibling pairs in the sample are independent. Considering a $n \times n$ matrix ($\mathbf{C}$) as comparison matrix of the sample, if individuals i and j are sibling, then $\mathbf{C}_{i,j} = 1$, otherwise it is 0 (Figure A1). The CKRM method assumes that the values of each element in matrix $\mathbf{C}$ are independent of each other. However, the sibling relationships between individuals are not independent. The probability bias caused by a slight correlation between pairs decreases as the effective population size increases, but also increases as the number of multi-sibling sets (three or more individuals are siblings of each other) increases. In the computer lab of the ConGen course, we attempt to improve the CKMR method by introducing multi-sibling sets.

2 Models and Methods

Notation

N_c	Census population size
N_e	Effective population size
N	Population size (in WF population, when $N_e = N_c$)
n	Sample size
$\binom{i}{j}$	Combinatorial number, choose j from i
S_i	number of real i-siblings set (i individuals that are siblings and all of the offspring from the same parent)
Y_i	number of observed i-siblings set (i individuals that are siblings but may not have all of the offspring from the same parent)
$\mathbb{E}$	Expectation value
$\mathbf{C}$	Comparison matrix of a sample the element $\mathbf{C}_{i,j}$ is 1 when the ith and jth individual in the sample are sibling

2.1 Population simulation

To test the CKMR method and other estimators, we simulated some simple populations in R. To simplify the condition, the population is viewed as a haploid population, with a constant census population size. The parents are numbered without duplication. The offspring have the same serial number as their parents. Then the sibling relationship is identified by the serial number. Three types of populations have been simulated with different Population sizes as well as different sample sizes.

Firstly, we generated haploid Wright-Fisher populations where each offspring randomly and independently chooses an individual in the last generation as its parent. Secondly, populations with a dominant individual were simulated. In these populations, the dominant individual is the parent for the k proportion of the offspring. Other individuals still have the same probability of being the parent of an offspring. Furthermore, the third type of population has q proportion of the individuals that are the parents of k proportion of the offspring.

n offspring were then sampled from the generated population. A comparison matrix is then calculated by whether offspring share the same parent. Sibling numbers of individuals are given by the column sum.

2.2 Estimation of real effective population size

For the WF population, the effective population size is the same as the census population size (because of the definition of N_e). For the non-random population, effective population size could also be estimated by the strength of genetic drift in the population. The fixation time of alleles is the direct result of genetic drift. In a haploid population with constant size, the relation between fixation time and effective population size is:

$$\mathbb{E}(T_{fix}) = \frac{-2N_e(1-p)ln(1-p)}{p} \tag{4}$$

in which $\bar{T}_{fix}$ is the mean fixation time, p is the initial frequency of allele. Generally, we use $T_{fix} \simeq 2N_e$ to estimate the real value of N_e.

The same is true in the population simulation section. We first numbered the parents without duplication and used the rule talked about before to generate offspring. The offspring inherit the same serial number as their parents. The simulation was repeated on the offspring until all the individuals had the same serial number. We counted the number of generations needed as fixation time.

2.3 Estimator by real siblings sets

The sibling pairs are not totally independent, because some of the sibling pairs (actually, most of them), are from multi-sibling sets. For example, a 3-sibling set has 3 sibling pairs, but the three pairs are related. As a result, we first attempted to find a new estimator based on the number of multi-sibling sets.

The sibling numbers of each individual are given by the column sum of the comparison matrix. The individual that has a sibling number as i must be in an i-sibling set. Thus we could statistic the number of multi-sibling sets.

Considering in a WF population, the proportion of i-sibling set equals the probability that a parent has i offspring. This probability follows a binomial distribution.

$$\mathbb{E}(S_i) = N \cdot \binom{N}{i}(\frac{1}{N})^i(1-\frac{1}{N})^{N-i} \tag{5}$$

in which, S_i is the number of real i-sibling set.

Notice that:

$$\begin{aligned}
\mathbb{E}(S_0)/N &= (1-\frac{1}{N})^N \\
\mathbb{E}(S_1)/N &= (1-\frac{1}{N})^{N-1} \\
\mathbb{E}(S_2)/N &= \frac{N(N-1)}{2}(\frac{1}{N})^2(1-\frac{1}{N})^{N-2} \\
&= \frac{1}{2}(1-\frac{1}{N})^{N-1}
\end{aligned}$$

The expected proportion of the real i-sibling set is a function of N. Especially, $0, 1, 2$-sibling sets provide an easy way to estimate $(1 - \frac{1}{N})$.

2.4 Improved CKMR by observed siblings sets

The CKMR method itself is accurate with large population size and low selfing. Considering that, we attempt to improve it with the same idea. The observed multi-sibling set is used to improve the estimator.

Firstly, the number of the 2-sibling set, the sibling pair, gives the estimation of N_e. Similarly, the number of multi-sibling sets (Y) also gives the estimation of the power of N_e. The probability that i individuals are siblings is the $i-1$ power of $\frac{1}{N}$.

$$\hat{N}_e^{i-1} = \frac{\binom{n}{i}}{Y_i} \tag{6}$$

As a result, Y_i provides more estimation of N_e, which could be used to improve the single estimation.

Secondly, observed multi-sibling sets are a set of linear combinations of real multi-sibling sets. Also, this linear combination is a set of basis in the vector space (S). On one hand, this fact means different Y_i gives different estimations. And on the other, Y_i can be calculated by S_i, the latter is directly obtained from the comparison matrix.

$$\mathbb{E}(Y_i) = \sum_{j=i}^{n} \binom{j}{i} \cdot \mathbb{E}(S_j)$$

To combine the different estimations of N_e in (6), we first take the logarithm of the equation.

$$(i-1) \cdot ln(\hat{N}_e) = ln\binom{n}{i} - ln(Y_i) \ldots\ldots (i = 2, 3, 4 \ldots) \tag{7}$$

let $f(x) = ln\binom{n}{(x+1)} - ln(Y_{x+1})$, fit $f(x)$ to x, the slope is the estimation of $ln(N_e)$. Because equation 7 shows a straight line through the origin, we used the linear regression with intercept eauqls to 0 as our new estimator. And considering the effection of sample size to regression (different distribution of Y_i), we used Y_i as the weight in the linear regression model. We also tested linear regression without limitation on intercept or unweighted.

3 Results

3.1 Estimator by different observed siblings set

We compared the CKMR estimator with our new estimator based on linear regression of different multi-siblings set (Figure 1). In the WF population and multi-dominant population (k = 0.2, q = 0.1), the new estimator nearly have the same performance with the CKMR estimator (Figure 1a, 1b), which is better than $E3$ (the estimator based on 3-siblings set). But in the single-dominant population (k = 0.1), the new estimator is more similar to $E3$, which is better than CKMR estimator in this case (Figure 1c).

This result indicate that along with the increasing of selfing proportion, estimator based on 3-silings set get better than estimator based on sibling pair. However, since our new estimator provide a combination of different siblings sets, it could always give a better estimation.

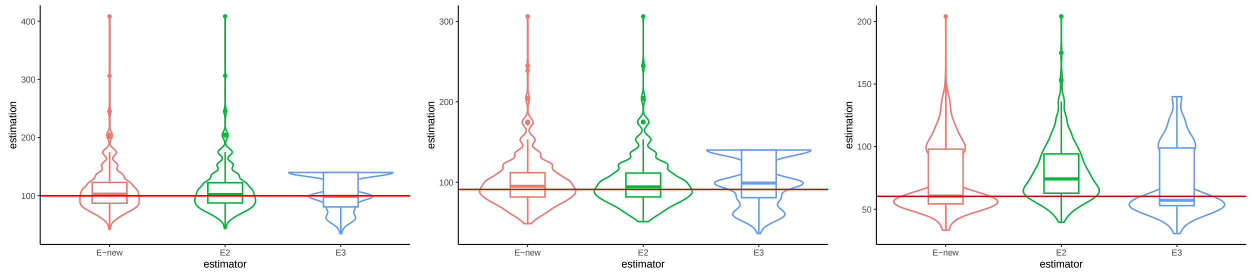


Figure 1: Estimator in different population. $E-new$ is the new estimator based on the linear regression of different observed siblings sets. $E2$ is the normal CKMR estimator, based on sibling pairs. $E3$ is the estimator based on 3-siblings sets. (a) is in the WF population, (b) is in the population that 10% individuals have 20% offsprings, (c) is in the populatio that only 1 parent have 10% offsprings. Red line in the figure is N_e that calculated by fixation time.

3.2 Weighted linear regression of observed siblings set

Unweighted linear regression was then be examined. Because the selfing proportion in single-dominant population changes slowly and gently, we tested estimators in this type of populations ($q = 0.1$, $k = 0.2, 0.5, 0.8$).

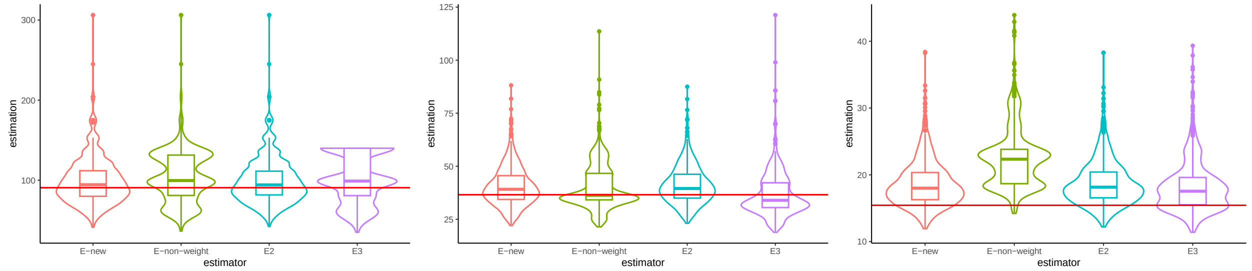


Figure 2: Compare weighted and non-weighted estimator. $E-new$ is the new estimator based on the linear regression of different observed siblings sets. As well as $E-non-weighted$ is the same estimator but with non-weighted linear regression. The three conditions are in the population that 10% individuals have (a) 20%, (b) 50%, (c) 80% offsprings. Red line in the figure is N_e that calculated by fixation time.

Again, same to Figure 1, $E2$ performes well in a low selfing condition, while $E3$ works better in a high selfing condition. And our new estimator always works similar to the better one (Figure 2). In an intermediate selfing condition ($k = 0.5$, Figure 2b), estimator based on unweighted linear regression ($E - non - weighted$) works best. However, it is even worse in other conditions ($k = 0.2, 0.8$, Figure 2a, 2c).

We also tested the same new estimator but without limiting the intercept as 0. In all condition it performes worse, because the high variance of the number of multi-siblings sets affect it a lot.

4 Discussion

4.1 Real siblings set provides a bad estimator

Number of real siblings set (equation 5) is always a function of N_e, it seems could provide more estimation than simply use CKMR method. However, it is a bad estimator because of its very low resolution. Take the situation of real sibling pair as an example, we could get the estimation of $(1 - \frac{1}{N})^{N-1}$ from the pair number. But this term cahnges slowly when N is big. If we compare the condition $N = 100$ and $N = 1000$, the difference of $(1 - \frac{1}{N})^{N-1}$ lower than 0.01, which is very easy to be covered by random variance. This estimator is only effective when $N < 20$.

4.2 Siblings set under unbalanced population

It can be noticed that the CKMR method works very well in the WF population and the population with low selfing proportion. This is because of the relatively more sibling pairs in WF population. But when the population don not follows the condition that individuals have same probability to have offspring, which means that some individuals have more offsprings then others, CKMR estimator works worse. Also, the estimator by 3-siblings set get better in the unbalance condition.

This shift of estimation effect is based on the change of multi-siblings sets proportion. With the population deviating WF population, multi-siblings sets get relatively more, based on which, made a better estimator. As a result, as the combination of all the multi-siblings sets, our new estimator has robustness of estimation. Especially, because the lower number of certain multi siblings set always has a higher variance, the unweighted estimator could only have a good estimation in certain conditions but always worse than our new estimator.

4.3 Estimate effective population size from the linear combination and exponentiation

Our new estimator is not a incalid improvement of CKMR. The real siblings sets provide more information than only the observed pair number, though both of them could be easily draw from the comparison matrix. As we discussed in the Methods and Models section, the number of observed siblings sets (Y_i) provide a set of basis of the space construct by the number of real siblings sets (S_i). The whole process could be written as the fomula that the power of N_e is a linear combination of S_i.

$$\mathbf{M} \times \mathbf{S} = [N_e, N_e^2, \dots N_e^n]^T$$

where $\mathbf{M}$ is a matrix of constant (also, a combination of combination number), $mathbfS$ is the vector consist of Y_i. We assume this form could provide more Rigorous statistical inference of N_e. Our new estimator just a simple attempt of it.

References

Bravington, M. V., Grewe, P. M., and Davies, C. R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications*, 7(1):13162. https://doi.org/10.1038/ncomms13162.

Bravington, M. V., Skaug, H. J., and Anderson, E. C. (2016b). Close-Kin Mark-Recapture. *Statistical Science*, 31(2):259–274. https://doi.org/10.1214/16-STS552.

Hill, W., Amsden, A. A., and Cook, J. L. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetical Research*, 38(3):209–216. https://doi.org/10.1017/S0016672300020553.

Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P., Bradford, R., Feutry, P., Gunasekera, R., Peddemors, V., Werry, J., Francis, M. P., Duffy, C. a. J., and Bruce, B. D. (2018). Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. *Scientific Reports*, 8(1):2661. https://doi.org/10.1038/s41598-018-20593-w.

Pudovkin, A., Zaykin, D., and Hedgecock, D. (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*, 144(1):383–387. https://doi.org/10.1093/genetics/144.1.383.

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., Mills Flemming, J., and Skaug, H. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods in Ecology and Evolution*, 10(9):1445–1453. https://doi.org/10.1111/2041-210X.13243.

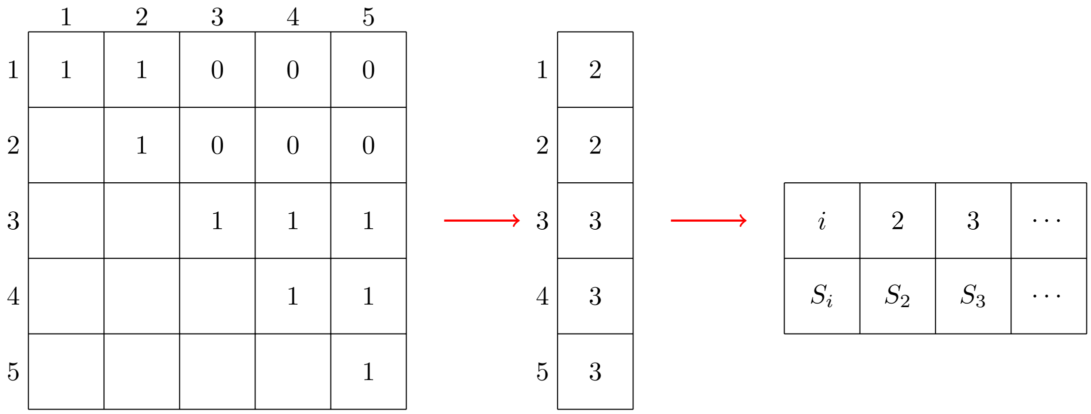
Scheiner, S. (2024). Encyclopedia of biodiversity. *Elsevier, Academic Press*.

Waples, R. S. and Feutry, P. (2022). Close-kin methods to estimate census size and effective population size. *Fish and Fisheries*, 23(2):273–293. https://doi.org/10.1111/faf.12615.

Author Contributions
Longxiao Chang:
The basic idea and mathematical derivation, Coding and programming, writing of the report.
Yakup Ferit Ayranci:
Writing of the report, troubleshooting and proofreading, formatting of the bibliography.

Appendix A Real number of sibling pairs and Siblings Sets

The real number of siblings sets could easily calculated from the comparison matrix. As a result, our new estimator do not need extra investigation. In CKMR, the sum of comparison matrix is calculated. But in the new estimator, we calculate column sum and than count the same numbers of column sum (Figure A1).



$$
\begin{array}{c|ccccc|}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 1 & 1 & 0 & 0 & 0 \\
2 & & 1 & 0 & 0 & 0 \\
3 & & & 1 & 1 & 1 \\
4 & & & & 1 & 1 \\
5 & & & & & 1 \\
\hline
\end{array}
\longrightarrow
\begin{array}{c|c|}
\hline
1 & 2 \\
2 & 2 \\
3 & 3 \\
4 & 3 \\
5 & 3 \\
\hline
\end{array}
\longrightarrow
\begin{array}{|c|c|c|c|}
\hline
i & 2 & 3 & \cdots \\
\hline
S_i & S_2 & S_3 & \cdots \\
\hline
\end{array}
$$

Figure A1: From comparison matrix to siblings sets

The CKRM method assumes that the values of each element in matrix $\mathbf{C}$ are independent of each other. But let us considering three individuals (a, b, c), the values of $\mathbf{C}_{a,b}$, $\mathbf{C}_{a,c}$ and $\mathbf{C}_{b,c}$ are related. Only in the case that $\mathbf{C}_{a,b}, \mathbf{C}_{a,c} = 0$, the probability that b, c are sibling ($Pr(\mathbf{C}_{b,c}) = 1$) is $\frac{1}{N_e}$. The probability bias caused by a slight correlation between pairs decreases as the effective population size increases, but also increases as the number of multi-sibling sets (three or more individuals are siblings of each other) increases.

An extreme example to dependent probability of sibling relation is the probability that all the individuals are siblings. Considering that we sample the whole population, then $n = N$, and the probability that all individuals are siblings is actually equation A1

$$P_{all-sib} = \frac{1}{N^N} \tag{A1}$$

But if we calculate it by sibling pairs, the probability that all elements in the comparison matrix are 1 is given by equation A2

$$P_{all-sib-mat} = \frac{1}{N^{R(N)}} \tag{A2}$$

Obviously, they are different.

Appendix B Real and observed siblings set

We define i individuals that are siblings of each other as a i-siblings set. In figure A2, red individuals are a 2-siblings set (also, sibling pair), as well as blue individuals are a 3-siblings set.

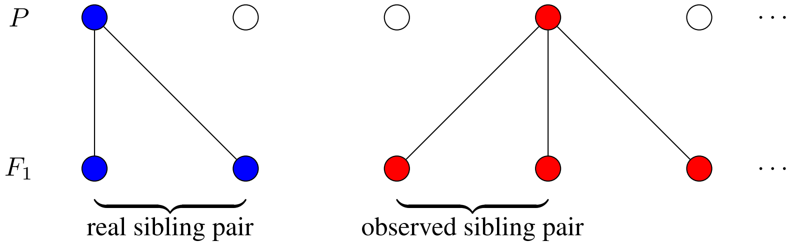


Figure A2: Real and observed sibling pair

In the sample or the population, any i individuals that are siblings of each other are an observed i-siblings set. But, only the i-siblings set that contains all the offspring from the same parent in the sample or population is a real i-siblings set. For example, in figure A2, red pair is a real sibling pair, but the marked blue pair is a observed sibling pair.

Also, to count observed i-siblings set, it is not needed to compare every i individuals. The number of observed siblings sets could be calculated by number of real siblings sets. For example, there are 6 sibling pairs or 4 3-siblings sets (Figure A3).

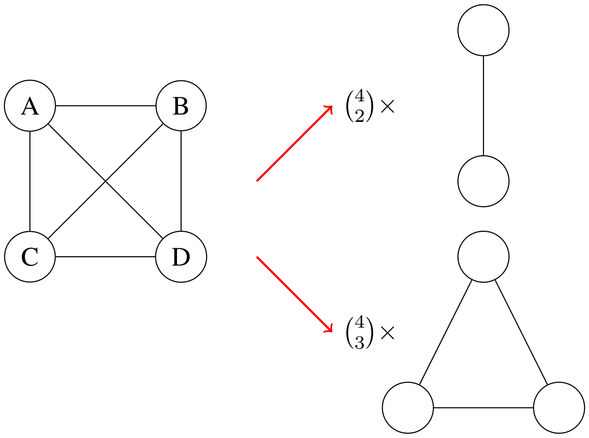


Figure A3: Calculate observed siblings set from real siblings set

Appendix C   Documentation of Code

Main document of R code

```r
#---Main document to do all the analysis---

#set workspace
setwd("E:/LMU-EES/ConGen/multi_siblings")

#import functions
source("Toolbox_pop_simu.R")
source("Toolbox_Ne.R")
source("Toolbox_estimator.R")

library(ggplot2)

#parameter
t<-100
pop_size<-100
sample_size<-50

#An example of WF-population
#population simulation
WF_result<-SimuMain(t,pop_size,sample_size,WF_func)
WF_sibling_stat<-RealSibling(WF_result)

#Ne estimation by CKMR, new estimator
WF_Ne_est<-NeEstimator(WF_sibling_stat,sample_size)

#calculate the "true value" of Ne by fixation time
WF_Ne<-Ne_TrueValue(pop_size, WF_func)

#organize the estimations and draw figures
WF_graph_data<-data.frame(c(WF_Ne_est$estimator,WF_Ne_est$estimator3,WF_Ne_est$new_est),c(
names(WF_graph_data)<-c("estimation","estimator")

ggplot(data = WF_graph_data,
       aes(x = estimator, y = estimation, color = estimator))+
  geom_violin(width = 0.9, size = 0.75)+
  geom_boxplot(width = 0.3, size = 0.75)+
  geom_hline(yintercept = 100, color = "red", size = 0.75)+
  theme_classic()+
  theme(legend.position = "none")
```

The functions used to simulate population

```r
#---Toolbox of population simulation---
#simple haploid constant population size condition


WF_func<-function(pop_seq){
  #reproduction: randomly pick parent
```

```
  m<-length(pop_seq)
  return(sample(pop_seq,m,replace = T))
}

SinExtra<-function(pop_seq){
  #reproduction: k of the population are offspring of same parent
  #pick super_parent
  m<-length(pop_seq)
  s_parent<-sample(pop_seq,1)
  s_offs<-round(k*m)
  parent<-c(rep(s_parent,s_offs),sample(pop_seq[-which(pop_seq==s_parent)],(m-s_offs),repla
  return(parent)
}

MultiExtra<-function(pop_seq){
  m<-length(pop_seq)
  s_parent<-sample(pop_seq,round(q*m))
  s_offs<-round(k*m)
  parent<-c(sample(s_parent,s_offs,replace = T),sample(pop_seq[!pop_seq %in% s_parent],(m-
  return(parent)
}

SimuMain<-function(t,pop_size,sample_size,func){

  simu_times<- c(1:t)
  n<-c(1:pop_size)
  compare_matrix<-matrix(0,sample_size,sample_size)

  #record simulation result
  simu_record<-numeric(sample_size)

  for (s in simu_times) {

    parent<-func(n)

    #capture in CKMR, sample individuals
    capture_ind<-sample(n,sample_size)

    #calculate the comparison matrix to check sampled pairs
    for (i in 1:sample_size) {
      for (j in 1:sample_size) {
        compare_matrix[i,j]<-(parent[capture_ind[i]]==parent[capture_ind[j]])
      }
    }

    #result record
    simu_record<-rbind(simu_record,colSums(compare_matrix))
  }

  return(simu_record)
}
```

The functions used to calculate fixation time

```
#---Toolbox of Ne calculation---

Ne_TrueValue<-function(pop_size, func, t=1000){
  simu_times<- c(1:t)
  n<-c(1:pop_size)
  fixation_num<-numeric(t)
  fixation_time<-numeric(t)
  k=0.1

  for (s in simu_times) {
    genotype<-n
    ft=1
    while(!all(genotype == genotype[1])){
      ft<-ft+1
      genotype<-genotype[func(n)]
    }
    fixation_num[s]<-genotype[1]
    fixation_time[s]<-ft
  }
  fixation_frame<-data.frame(fixation_time,fixation_num)
  Mean_T<-mean(fixation_time)
  SD_T<-sd(fixation_time)
  Ne_est<-Mean_T/2
  return(list(fixation_frame,Mean_T,SD_T,Ne_est))
}
```

The functions used to calculate the estimator

```
#---Toolbox of siblings set and the estimators---

f <- function(array) {
  #function used for RealSibling-apply
  return(as.data.frame(table(array)))
}

RealSibling<-function(simu_record){
  #statistic real siblings set
  simu_record<-simu_record[-1,]
  stat_simu<-apply(simu_record, 1, f)
  return(stat_simu)
}

NeEstimator<-function(stat_simu, sample_size, intercept0 = TRUE, weighted = TRUE){
  #Estimators
  #stat_simu is list from RealSibling function
  #sample_size is the size of sample
  #intercept0 = T means the intercept of linear regression is 0
  #weighted control whether the regression is weighted
  obs_pairs<-list()
  estimator<-numeric()
```

```
  estimator3<-numeric()
  new_est<-numeric()
  t<-length(stat_simu)
  for (i in 1:t) {
    freq<-stat_simu[[i]]$Freq
    array<-as.numeric(stat_simu[[i]]$array)
    k<-max(array)
    f_obs<-function(j){
      num_set<-sum(freq/array*choose(array,j))
      return(num_set)
    }
    obs_pairs[[i]]<-sapply(2:k,f_obs)
    Yi<-choose(sample_size,2:k)/obs_pairs[[i]]
    if(weighted){
      if(intercept0){
        a<-lm(log(Yi)~0+c(1:(k-1)),weights = 1/Yi )
      }
      else{
        a<-lm(log(Yi)~c(1:(k-1)),weights = 1/Yi )
      }
    }
    else{
      if(intercept0){
        a<-lm(log(Yi)~0+c(1:(k-1)))
      }
      else{
        a<-lm(log(Yi)~c(1:(k-1)))
      }
    }

    new_est[i]<-exp(a$coefficients)
    estimator[i]<-Yi[1]
    if(is.na(Yi[2])){
      estimator3[i]<-NA
    }
    else{
      estimator3[i]<-sqrt(Yi[2])
    }
  }
  return(data.frame(new_est,estimator,estimator3))
}
```