

Prediction and Estimation of Population Size

—from census to effective population size—

Chang Longxiao

EES program, LMU
Longxiao.Chang@campus.lmu.de

May 12, 2024

Contents

A brief introduction to population size

Census Population Size monitoring

Traditional capture-mark-recapture (CMR)

DNA-based one-sample estimator

Effective Population Size monitoring

Different estimators

Heterozygosity excess

Linkage disequilibrium

Relatedness and relationship

Temporal changes in allele frequency

Multiple sources

Discussion

Effective population size in conservation practices

A shared framework in Evolution and Ecology

A brief introduction to population size

Census Population Size monitoring

- Traditional capture-mark-recapture (CMR)

- DNA-based one-sample estimator

Effective Population Size monitoring

- Different estimators

- Heterozygosity excess

- Linkage disequilibrium

- Relatedness and relationship

- Temporal changes in allele frequency

- Multiple sources

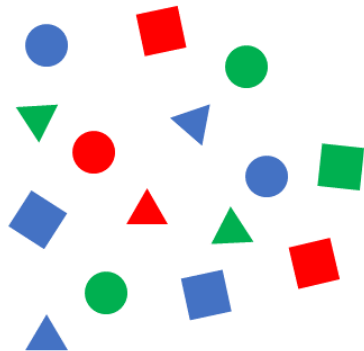
Discussion

- Effective population size in conservation practices

- A shared framework in Evolution and Ecology

Count: the initial impulse of MATH

1,2,3,4,5...
一二三四五 ...



Population Size

The first mathematical abstract when we do research on the visible nature is the population size.

How many individuals are there in the group?

We count the number of individuals as:

N

Important concept in nearly all part of evolution and ecology research

The biological meaning

ECOLOGY

In ecology research, especially in community ecology, the counting of abundance leads to the concept of biodiversity

$$D_{\alpha} = - \sum p_i \cdot \ln(p_i)$$
$$(p_i = \frac{N_i}{\sum N_i})$$

then, based on which, researchers discuss the dynamic and stability of the community.

$$\frac{dN_i}{dt} = r_i N_i + \sum \alpha_{i,j} N_i N_j$$

EVOLUTION

Since population is viewed as the basic unit of reproduction, population size becomes a core parameter in population genetics, especially in the finite model(for example, Wright-Fisher model)

In gene drift model, we find the fixation probability of a single mutation in diploid population is:

$$p = \frac{1}{2N}$$

and then the heterozygosity could be calculated as:

$$\frac{H_t}{H_0} = \left(1 - \frac{1}{2N}\right)^t$$

Two problem of the population size

- ▶ How to count the number?
- ▶ Do all of the nature population follows the same model?

Census and Effective population

Census population N_c

- ▶ The true population size (sometimes only count adults)

Effective population N_e

- ▶ ~~The effective population size (?)~~
- ▶ The effective population size is defined in reference to the Wright-Fisher idealised population
(a hypothetical population with very simplifying characteristics where genetic drift is the only factor in operation, and the dynamics of allelic and genotypic frequencies across generations merely depend on the population census (N) size.)
The effective size of a real population is then defined as the size of an idealised population.
- ▶ $N_e = \frac{4N}{2+s_k^2}$ in which N is the population size in last generation, s_k is the variance of the contributions from parents to progeny

population size estimator

Though population size is one of the most important and core parameter in population genetic as well as conservation biology, it is hard to estimate it[1].

John Shepherd

Counting fish is like counting trees, except they are invisible and they keep moving.

Population sizes are difficult to estimate, especially for secretive and elusive species, ranging from fish to other aquatic organisms and forest-dwelling mammals.

A brief introduction to population size

Census Population Size monitoring

Traditional capture-mark-recapture (CMR)

DNA-based one-sample estimator

Effective Population Size monitoring

Different estimators

Heterozygosity excess

Linkage disequilibrium

Relatedness and relationship

Temporal changes in allele frequency

Multiple sources

Discussion

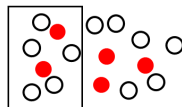
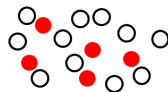
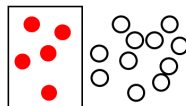
Effective population size in conservation practices

A shared framework in Evolution and Ecology

Capture-Mark-Recapture (CMR)

- ▶ mark individuals in an initial capture session
- ▶ quantify the proportion of marked individuals in subsequent recapture sessions
- ▶ estimate N_c from the ratio of marked to unmarked individuals in recapture sessions

(assuming that all individuals (marked and unmarked) are randomly mixed and sampled and thus all are equally catchable during the recapture sessions)



Noninvasive or remote genetic techniques allow marking of many animals by using DNA collected from hair, feathers, urine, menstrual blood, snail slime tracks, or other tissue samples [2].

one-sample estimator

Main Principal

If the same multilocus genotype, i.e., individual, is observed ('captured') two or more times in the same single sampling session, all captures beyond the first are considered 'recaptures'. The ability to estimate N_c from a single sampling session is extremely helpful for species that are costly or time consuming to sample.

An accumulation curve is then used to estimate the number of unique genotypes, which is viewed as individuals.

also, currently, maximum likelihood and Bayesian methods can make use of the information provided by multiple occurrences of individual genotypes within a session

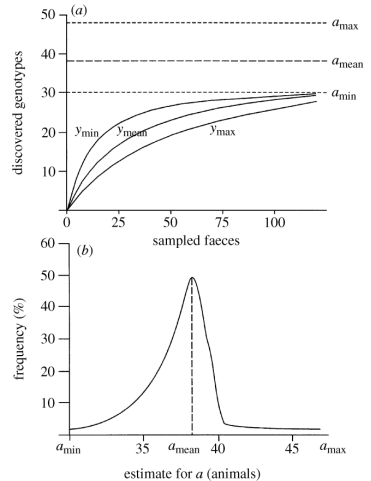
one-sample estimator: example

Population size inference based on the rarefaction curve used to estimate the number of coyotes from feces sampled in an area near Los Angeles [3].

average number of unique genotypes (y) discovered as a function of the number of samples (x)

$$y = \frac{ax}{b + x}$$

where a is the population size asymptote, and b a constant which is the rate of decline in the value of the slope



A brief introduction to population size

Census Population Size monitoring

Traditional capture-mark-recapture (CMR)

DNA-based one-sample estimator

Effective Population Size monitoring

Different estimators

Heterozygosity excess

Linkage disequilibrium

Relatedness and relationship

Temporal changes in allele frequency

Multiple sources

Discussion

Effective population size in conservation practices

A shared framework in Evolution and Ecology

Conditions of effective population size

Effective population size is a model to regulate the real population to fit the Wright-Fisher model and conduct further prediction.

As a result, with different condition, the prediction, or the calculation of N_e should be different.[4]

Conditions of effective population size

- Variation in offspring number

$$N_e = \frac{4N}{2 + s_k^2}$$

- Division into two sexes

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

- partially selfed population

$$N_e = \frac{4N}{2(1 - \alpha) + s_k^2(1 + \alpha)}$$

in which $\alpha = \beta(2 - \beta)$, β is the proportion of self mating.

- Age- and stage-structure ...
- ...

- Changes in population size (temporal)

$$\begin{aligned} \frac{H_t}{H_0} &= \left(1 - \frac{1}{2N_e}\right)^t \\ &= \left(1 - \frac{1}{2N_0}\right) \left(1 - \frac{1}{2N_1}\right) \left(1 - \frac{1}{2N_2}\right) \cdots \left(1 - \frac{1}{2N_{t-1}}\right) \end{aligned}$$

since we have:

$$\lim_{x \rightarrow 0} \left(1 - \frac{1}{x}\right)^x = e \text{ and } \lim_{x \rightarrow 0} e^x = 1 + x$$

then,

$$N_e = \frac{t}{\sum_{i=0}^{t-1} \frac{1}{N_i}}$$

classical or gene-based method

- ▶ Given the concepts of effective size, different approaches can be used to predict the effective size of a population from its demographic parameters, such as census size and variance of reproductive success.
- ▶ In parallel, different methodologies can also be developed to estimate the realized effective size of a population from its genetic properties revealed by genetic markers, such as temporal changes in allele frequency and linkage disequilibrium.

Heterozygosity excess

Compared with an infinitely large population at Hardy-Weinberg equilibrium, a finite population is expected to show a deficit of homozygotes and **an excess of heterozygotes** at a neutral locus because of binomial sampling error.

$$E(H_{obs}) = 2pq + pq/2N_e$$

The heterozygosity excess(D) is expected to be

$$D = \frac{H_{exp}}{H_{exp} - H_{obs}} = \frac{1}{2N_e}$$

Measuring the heterozygosity excess, D, at a number of marker loci in a population yields an estimate of the parental population effective size [5]:

$$\hat{N}_e = \frac{1}{2\hat{D}} + \frac{1}{2(\hat{D} + 1)}$$

Methods based on Linkage disequilibrium

Random genetic drift also leads to associations between alleles of different loci. For neutral loci unlinked with selected loci in an isolated population under random mating, LD would come exclusively from genetic drift and can be used to estimate N_e [6].

linkage disequilibrium

Considering 2 loci A and B with 2 alleles in both loci (as frequency p, q), the linkage disequilibrium could be represent as:

$$D = p(AB) - pq$$

also, the correlation of gene frequencies is:

$$r = \frac{D}{\sqrt{p(1-p)q(1-q)}}$$

and the combination fraction between A and B is c .

Methods based on Linkage disequilibrium

Though the expectation of r becomes 0 when the population becomes a stable stage, the variance of r could maintain a constant value if the population is still segregated.

$$Var(r) = \frac{(1 - c)^2 + c^2}{2N_e c(2 - c)} + \frac{1}{n}$$

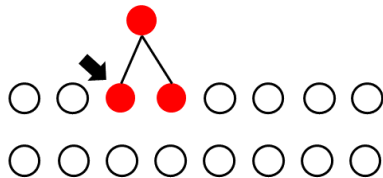
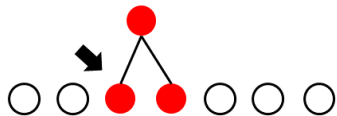
where n is the number of samples.

Using the genotypes at a number of loci of n sampled individuals, we can estimate $V(r)$, which can then be inserted in the equation to obtain an estimate of N_e if the recombination fraction c between loci is known

Relatedness and relationship

The pattern of genetic relatedness or relationship between individuals in a population has a direct functional relationship with the inbreeding effective size of the population[7] .

Two individuals taken at random from a population with a smaller N_e will have a higher probability of sharing the same father, mother or both. More generally, the mean and variance in pairwise relatedness within a generation are expected to increase with decreasing N_e .



Method based on Relatedness and relationship

A more robust and powerful method is to estimate the frequencies of half-sib (sharing a single parent, Q_{HS}) and full-sib (sharing both parents, Q_{FS}) dyads in a sample taken at random from a single cohort of a population[8]

$$\frac{1}{N_e} = \frac{1+3\alpha}{4}(Q_{HS} + 2Q_{FS}) - \frac{\alpha}{2}\left(\frac{1}{N_f} + \frac{1}{N_m}\right)$$

which is also used as close-kin mark-recapture (CKMR) methods[9]

Temporal changes in allele frequency

In the absence of the action of all of the systematic forces in a population, any observed allele frequency change must come solely from genetic drift and can thus be used to infer the rate of drift or the N_e of the population [10].

Moment estimators calculate a standardised variance in the temporal changes of allele frequency, F , from marker genotypes in two temporally spaced samples.

(F is essentially similar to Wright's F_{st})

One of the F estimator could be:

$$\hat{F} = \frac{1}{k} \sum_{i=1}^k \frac{(x_i - y_i)^2}{(x_i + y_i)/2 - x_i y_i}$$

for a locus with k alleles, where x_i and y_i are the observed frequencies of allele i in the first and second samples, respectively.

Method based on Temporal changes in allele frequency

For multiple loci, F is obtained by averaging single-locus estimates. The expectation of F is a function of N_e and sample sizes (S_0 and S_1) to account for genetic drift and sampling effects.

$$\hat{N}_e = \frac{t - 2}{2 \left(\hat{F} - \left(\frac{1}{S_1} + \frac{1}{S_0} \right) \right)}$$

Solving the expectation equation of for N_e yields the temporal estimate of the (harmonic) mean N_e during the sampling period

Multiple sources of drift/inbreeding information

Each piece of information reflects a facet of the stochastic process (genetic drift or inbreeding) and combining multiple pieces of information may potentially allow for a better delineation of the process and thus yield a more accurate estimate

An approximate Bayesian computation (ABC) is used to estimate N_e

A brief introduction to population size

Census Population Size monitoring

- Traditional capture-mark-recapture (CMR)

- DNA-based one-sample estimator

Effective Population Size monitoring

- Different estimators

- Heterozygosity excess

- Linkage disequilibrium

- Relatedness and relationship

- Temporal changes in allele frequency

- Multiple sources

Discussion

- Effective population size in conservation practices

- A shared framework in Evolution and Ecology

Effective population size in conservation practices

The concept of effective size is key to conservation genetic practices, as it summarises the past history of the population regarding inbreeding and genetic drift and provides the prospects for the sustainability of the population if the current effective size is maintained in the future.[8]

Minimising the loss of genetic variation is one of the main objectives of captive breeding programmes. This is achieved through minimising genetic drift and, therefore, maximising N_e .

$$N_e = \frac{4N}{2 + s_k^2}$$

A classical strategy to follow is the equalisation of family sizes. By choosing one couple from each pair of parents, the variance of parental contributions is null ($s_k^2 = 0$), which makes $N_e \simeq 2N$, the twice size as randomly mating group.

A conservation strategy is supportive breeding, where a population is typically divided into a captive and a wild group and the offspring of the captive group are released into the wild habitat to mix with the offspring from the wild group.

Because the captive group (permanent or transitional) is bred to produce a lot of offspring that are released into the wild group at each generation, the variance in family size is greatly elevated artificially ($k_{captive} \gg k_{wild}$) and thus the N_e of the entire population is reduced.

$$N_e = \frac{4N}{2 + s_k^2}$$

Ironically, the more successful the supportive programme is in augmenting the wild population, the greater the reduction in N_e and the greater the loss of genetic diversity in the total population (supportive+wild). This paradox is overcome only when successful supportive breeding in augmenting the wild population is carried out over a long period of time such that the excessive drift and inbreeding in the initial generations of supportive breeding is compensated for by weaker drift and inbreeding in later generations because of the increase in census size

Similar mathematical form: A shared framework

population size: N_c and N_e

CMR and CKMR

Neutral theory

References

- [1] G. Luikart, N. Ryman, D. A. Tallmon, M. K. Schwartz, F. W. Allendorf, Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches 11 (2) 355–373.
- [2] P. Taberlet, L. P. Waits, G. Luikart, Noninvasive genetic sampling: look before you leap 14 (8) 323–327.
- [3] M. H. Kohn, E. C. York, D. A. Kamradt, G. Haught, R. M. Sauvajot, R. K. Wayne, Estimating population size by genotyping faeces 266 (1420) 657–663, publisher: Royal Society.
- [4] B. Charlesworth, Effective population size and patterns of molecular evolution and variation 10 (3) 195–205, publisher: Nature Publishing Group.
- [5] A. I. Pudovkin, D. V. Zaykin, D. Hedgecock, On the potential for estimating the effective number of breeders from heterozygote-excess in progeny 144 (1) 383–387.
- [6] W. G. Hill, Estimation of effective population size from data on linkage disequilibrium 38 (3) 209–216.
- [7] M. Nei, F. Tajima, GENETIC DRIFT AND ESTIMATION OF EFFECTIVE POPULATION SIZE 98 (3) 625–640.
- [8] J. Wang, E. Santiago, A. Caballero, Prediction and estimation of effective population size 117 (4) 193–206.
- [9] R. S. Waples, P. Feutry, Close-kin methods to estimate census size and effective population size 23 (2) 273–293, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/faf.12615>.
- [10] R. S. Waples, A generalized approach for estimating effective population size from temporal changes in allele frequency. 121 (2) 379–391.

Prediction and Estimation of Population Size

—from census to effective population size—

Chang Longxiao

EES program, LMU
Longxiao.Chang@campus.lmu.de

May 12, 2024