

Nearly unbiased estimator of N_e/N based on kinship relationships

Tetsuya Akita^{a*}

^aNational Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa, 236-8648, Japan.

Preprint, December 11, 2019

Abstract

This study develops a nearly unbiased estimator of the ratio of the contemporary effective mother size to the census size in a population, as a proxy of the ratio of contemporary effective size to census size (N_e/N). The proposed estimator is based on both known mother-offspring (MO) and maternal-sibling (MS) relationships observed within the same cohort, in which sampled individuals in the cohort probably share MO relationships with sampled mothers. The rationale is that the frequency of MO and MS pairs contains information regarding the contemporary effective mother size and the (mature) census size, respectively. Therefore, the estimator can be obtained only from genetic data. Moreover, We also evaluate the performance of the estimator by running an individual-based model. The results of this study provide the following: i) parameter range for satisfying the unbiasedness, and ii) guidance for sample sizes to ensure the required accuracy and precision, especially when the order of the ratio is available. Furthermore, the results demonstrate the usefulness of a sibship assignment method for genetic monitoring, providing insights for interpreting environmental and/or anthropological factors fluctuating N_e/N , especially in the context of conservation biology and wildlife management.

Running head: Estimator of N_e/N

*Corresponding author: National Research Institute of Fisheries Science, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa, 236-8648, Japan. E-mail: aktiatetsuya1981@affrc.go.jp

1 INTRODUCTION

The Estimation of the ratio of the contemporary effective population size to the census size (N_e/N) has attracted much research attention for providing information about a current population, especially in the context of conservation biology and wildlife management (Frankham, Bradshaw, & Brook, 2014; Palstra & Fraser, 2012). Small N_e/N demonstrates large variance in reproductive success (Akita, 2019; Wang, Santiago, & Caballero, 2016; Waples, 2016), resulting from the variance of reproductive potential (e.g., the big old fat fecund female fish hypothesis; Hixon, Johnson, & Sogard, 2014) or from the situation in which only some families successfully reproduce (referred to as the “Sweepstakes reproductive success” hypothesis, Hedgecock & Pudovkin, 2011). Moreover, if N_e/N is invariant across years, then N_e may behave like an index of N , and vice versa (Luikart, Ryman, Tallmon, Schwartz, & Allendorf, 2010). However, if N_e/N fluctuates across years, the trends can clarify the interpretation of environmental and/or anthropological factors, causing the variance of reproductive potential, family-correlated survivorship, or fluctuating population dynamics.

The estimation of N_e/N has been performed by utilizing the estimated values of contemporary effective population size (N_e) and census size (N), unless complete pedigree data and/or full census data are available. Additionally, there are numerous methods for estimating N_e from genetic markers (Wang et al., 2016, and the references contained therein). There are also numerous methods for estimating N , such as a mark-recapture method or population dynamics modeling with survey data (e.g., Kéry & Schaub, 2011; Methot & Wetzel, 2013; Quinn & Deriso, 1999; Seber, 1982). It is known that there are large variations in both estimators; thus, their combination (i.e., the estimator of N_e/N) also shows large variation (Marandel et al., 2018; Palstra & Fraser, 2012). There is currently a little theoretical foundation for the estimator of N_e/N , indicating no guidance for a

sample size to ensure the required accuracy and precision.

Close-kin mark-recapture (CKMR) is a recently developed method for estimating N that utilizes the information about kinship in a sample. This was possible owing to the recent advances in genetic methods for kinship determination (Bravington, Grewe, & Davies, 2016; Bravington, Skaug, & Anderson, 2016; Hillary et al., 2018; Skaug, 2017) although similar methods have been proposed in the beginning of the 21st century (Nielsen, Mattila, Clapham, & Palsbøll, 2001; Pearse, Eckerman, Janzen, & Avise, 2001; Skaug, 2001). Besides, the rationale is that the presence of a kinship pair in the sample is analogous to the recapture of a marked individual in mark-recapture. Therefore, kinship pairs in the sample are less likely to be observed in larger populations; thus, the number of kinship pairs may reflect N . Since the original CKMR is designed to estimate adult abundance (i.e., N), the monitoring data for CKMR also produce the estimator of N_e by detecting half-sibling (HS) pairs within the same cohort (Akita, 2019). Since kinship determination is more accurate, this kinship-oriented estimation of N_e was presented in the context of the sibship assignment method (Wang, 2009) and is expected to provide a much more accurate estimator.

In this study, we propose a new method for estimating the ratio of contemporary effective mother size to the census size ($N_{e,m}/N_m$) in a population, as a proxy of N_e/N . Assuming that kinships are genetically detected without any error, this method is based on the numbers of maternal-sibling (MS) and mother-offspring (MO) pairs in a sample. Thus, sampling is completed at a single breeding time; sampling offspring within the same cohort and mothers probably shares MO relationship with sampled offspring. Our model provides a nearly unbiased estimator of $N_{e,m}/N_m$ that explicitly incorporates two types of overdispersed reproduction (i.e., parental and nonparental variations; Akita, 2019), making it possible to target a species that shows iteroparity (i.e., multiple reproductive cycles during the lifetime) and/or sweepstakes reproductive success. First, we explain the modeling assumption and sampling scheme. Then, we analytically determine (nearly) the un-

biased estimators of $N_{e,m}$, $1/N_m$, and $N_{e,m}/N_m$, which are based on the numbers of MS and/or MO pairs. Finally, by running an individual-based model, we investigate the performance of the estimator and provide a guide for a sample size. Moreover, it is noteworthy that our modeling framework can be applied to diverse animal species. However, the description of the model focuses on fish species, which are presently the best candidate target of our proposed method.

2 THEORY

The main symbols used in this paper are summarized in Table 1.

2.1 Hypothetical population

We suppose that there is a hypothetical population comprising N_m mothers and there is also no population subdivision or spatial structure. In this study, a mature female is called a mother even if she does not produce offspring. For mathematical tractability, we assume that only one spawning ground exists in which the mothers remain for the entire spawning season. Following Akita (2019), our modeling framework employs two types of overdispersed reproduction: parental and nonparental variations. Thus, the former indicates a variation caused by the mother's covariates, such as age, weight, and residence time on the spawning ground, while the latter indicates a variation caused by non-random stochastic events during a series of reproductive episodes, which are independent of the mother's covariates, such as family-correlated survivorship or the mating behavior effects (e.g., competition for males/females and correlation between mating opportunities of the mother and the number of her offspring). **Figure 1** illustrates a schematic diagram of the effects of parental and nonparental variations exemplified by age-dependent reproduction and family-correlated survival on kinship relationships in a population. Detailed definitions of parental

and nonparental variations are stated in Akita (2019).

Let k_i denote the number of surviving offspring of mother i ($i = 1, 2, \dots, N_m$) during sampling. It is noteworthy that k_i is assumed to be observed at the sampling, as explained in the next subsection. Following Akita (2019) and giving the expected number of the surviving offspring per mother during the sample timing, λ_i (> 0), k_i is assumed to follow a negative binomial distribution through a conventional parametrization:

$$\Pr[k_i|\lambda_i] = \frac{\Gamma[k_i + \phi]}{k_i! \Gamma[\phi]} \left(\frac{\lambda_i}{\phi + \lambda_i} \right)^{k_i} \left(\frac{\phi}{\phi + \lambda_i} \right)^{\phi}, \quad (1)$$

where ϕ (> 0) denotes the overdispersion parameter that describes the degree of the nonparental variation. Presently, ϕ is assumed to be constant across mothers, whereas the expected number of the surviving offspring (λ_i) varies across mothers. The mean and variance of this distribution are denoted by λ_i and $\lambda_i + \lambda_i^2/\phi$, respectively. In the limit of infinite ϕ , this distribution becomes a Poisson distribution, which is given by $\Pr[k_i|\lambda_i] = \lambda_i^{k_i} e^{-\lambda_i} / (k_i!)$. We assumed that λ_i is independent and identically distributed with the density function $f(\lambda)$, which produces a parental variation. The shape of the density function is often complex, but it may be described by information, for example, the mother's weight composition in the population. The specific form of $f(\lambda)$ is given in **Supporting Information** and is used for running an individual-based model.

2.2 Sampling

To obtain the estimator of $N_{e,m}/N_m$, we utilize the number of MS and MO pairs observed in a sample. For the mathematical tractability, only one reproductive season is targeted for sampling. Thus, whether the sampling method does not affect our modeling framework whether it is invasive or noninvasive. In the population, n_M mothers are randomly sampled immediately after the end of

the reproductive season. Additionally, in the population, n_O offspring are also randomly sampled but allowed to die before sampling. The numbers of MS and MO pairs are determined by pairwise comparison of all the sample individuals ($n_O C_2$ and $n_M n_O$ comparisons, respectively). As depicted in **Fig 1**, six offspring and three mothers are sampled and two MS and three MO pairs are observed. In our modeling framework, if an MS pair also shares a paternal-sibling (PS) relationship, we count it as an MS pair and ignore the existing full-sibling (FS) relationship.

2.3 Effective mother size ($N_{e,m}$) and the ratio to census size ($N_{e,m}/N_m$)

Akita (2019) derived the approximate probability showing that two offspring share an MS relationship with an arbitrary mother (denoted by π_{MS}) as follows:

$$\pi_{MS} \approx \frac{c}{N_m + c - 1}, \quad (2)$$

where

$$c = (1 + \phi^{-1}) \frac{\mathbb{E}[\lambda^2]}{\mathbb{E}[\lambda]^2}.$$

Without both parental and nonparental variations (i.e., λ is constant among mothers and $\phi \rightarrow \infty$), π_{MS} converges to $1/N$, corresponding to the Poisson variance for all mothers in a population. Moreover, the effect of the two factors causing a deviation from the Poisson variance can be combined as parameter c (≥ 1). In the sequel, we refer to “overdispersion” as the distribution of the offspring number that results from this combined effect. By applying the combined effect, the

variance of the offspring number can be given by

$$\begin{aligned}\mathbb{V}[k] &= \mathbb{E}[\mathbb{V}[k|\lambda]] + \mathbb{V}[\mathbb{E}[k|\lambda]] \\ &= \mathbb{E}[\lambda + \lambda^2/\phi] + \mathbb{V}[\lambda] \\ &= \mathbb{E}[\lambda] + \mathbb{E}[\lambda]^2(c-1),\end{aligned}\tag{3}$$

suggesting that the variance substantially increases with c .

Akita (2019) defined the contemporary effective mother size as follows:

$$N_{e,m} = \frac{1}{\pi_{MS}}.\tag{4}$$

Besides, this definition is related to the inbreeding effective population size (Nordborg & Krone, 2002; Wang, 2009). When sampling from a single cohort in a population with overlapping generations, the effective mother size in our definition corresponds to the effective breeding mother size that produces a single cohort. We obtain the ratio of the effective mother size to census size using Eqs. 2 and 4 (Akita, 2019), and it is given by

$$\begin{aligned}\frac{N_{e,m}}{N_m} &= \frac{1}{\pi_{MS}} \frac{1}{N_m} \\ &\approx \frac{1}{c},\end{aligned}\tag{5}$$

where $N_m \gg 1$ is assumed for approximation.

2.4 Estimator of $N_{e,m}/N_m$

This subsection proposes the estimator of $N_{e,m}/N_m$ as follows:

$$\widehat{\left(\frac{N_{e,m}}{N_m}\right)} = \widehat{N_{e,m}} \widehat{\left(\frac{1}{N_m}\right)}. \quad (6)$$

A “hat” denotes the estimator of a variable in this study. The requisite condition that satisfies Eq. 6 is independent of $\widehat{N_{e,m}}$ and $\widehat{1/N_m}$. This property will be shown later in this subsection. Akita (2019) derived the nearly unbiased estimator of $N_{e,m}$, which is given by

$$\begin{aligned} \widehat{N_{e,m}} &= \widehat{\left(\frac{1}{\pi_{MS}}\right)} \\ &= \frac{\binom{n_O}{2} + 1}{H_{MS}^{obs} + 1}, \end{aligned} \quad (7)$$

where H_{MS}^{obs} denotes the observed number of MS pairs in a sample. This estimator works well unless n_O is very small. Akita (2018) obtained a probability in which a randomly sampled mother and her offspring shares an MO relationship, which is given by

$$\pi_{MO} = \frac{1}{N_m}. \quad (8)$$

This indicates that π_{MO} is independent of the distribution of the offspring number. By definition of π_{MO} , we can set its estimator by $H_{MO}^{obs}/(n_M n_O)$, where H_{MO}^{obs} denotes the observed number of MO pairs in a sample. Thus, the estimator of $1/N_m$ can be determined as follows:

$$\widehat{\left(\frac{1}{N_m}\right)} = \frac{H_{MO}^{obs}}{n_M n_O}. \quad (9)$$

Finally, substituting $\widehat{N}_{e,m}$ (Eq. 7) and $1/\widehat{N}_m$ (Eq. 9) into Eq. 6, we obtain the estimator of $N_{e,m}/N_m$ as follows:

$$\left(\frac{\widehat{N}_{e,m}}{\widehat{N}_m}\right) = \frac{\binom{n_O}{2} + 1}{H_{MS}^{obs} + 1} \frac{H_{MO}^{obs}}{n_M n_O}. \quad (10)$$

Let both n_M and n_O be given. We numerically confirmed that there is no correlation between H_{MO}^{obs} and H_{MS}^{obs} (results are not shown). To intuitively explain this independency, we consider three mothers ($i = 1, 2, 3$) and their offspring, and assume that $(k_1, k_2, k_3) = (3, 1, 1)$ and $(n_M, n_O) = (1, 3)$. Moreover, when the three offspring born to the first mother are sampled (i.e., $H_{MS}^{obs} = 3$), the expected number of MO relationship is one ($= 1/3 \times 3 + 1/3 \times 0 + 1/3 \times 0$). Meanwhile, when an offspring is sampled from each mother's offspring (i.e., $H_{MS}^{obs} = 0$), the expected number of MO relationship is also one ($= 1/3 \times 1 + 1/3 \times 1 + 1/3 \times 1$). Therefore, we conclude that both $\widehat{N}_{e,m}$ and $1/\widehat{N}_m$ are independent, and $\widehat{N}_{e,m}/\widehat{N}_m$ is expected to work well (see details in the **RESULTS** section).

The bias of $\widehat{N}_{e,m}/\widehat{N}_m$ is defined by b , which is approximately given by (see **APPENDIX** for the derivation)

$$\begin{aligned} b &= \mathbb{E} \left[\left(\frac{\widehat{N}_{e,m}}{\widehat{N}_m} \right) \right] - \left(\frac{N_{e,m}}{N_m} \right) \\ &\approx - \left(\frac{N_{e,m}}{N_m} \right) \left(1 - \frac{1}{N_{e,m}} \right)^{\binom{n_O}{2} + 1}. \end{aligned} \quad (11)$$

It is noteworthy that $\widehat{N}_{e,m}/\widehat{N}_m$ is downwardly biased, especially when n_O is very small. However, this bias may be ignored for a wide range of parameters (see details in the **RESULTS** section). Theoretically, b asymptotically converges to zero as n_O increases, making $\widehat{N}_{e,m}/\widehat{N}_m$ a nearly unbiased estimator. Moreover, as demonstrated in the **RESULTS** section, it is observed that an extremely skewed reproduction breaks down the unbiasedness (e.g., in the case that $c = 20$ and 100 in the

results).

2.5 Individual-based model

We developed an individual-based model that tracks kinship relationships to evaluate the estimator's performance (e.g., $\widehat{N_{e,m}/N_m}$). The population structure was assumed to be identical to that in the development of the estimators. In addition, the population comprised mothers and their offspring, and it was assumed to follow a Poisson or negative binomial reproduction. The expected number of the surviving offspring of a mother followed the density distribution $f(\lambda)$ (see **Supporting Information** for details). We calculated overdispersion parameter (c) from ϕ and $f(\lambda)$, as well as confirmed numerically that the value of c gives the same statistics of the estimators even if the combination of ϕ and $f(\lambda)$ differs (results are not shown). Therefore, each offspring retained the mother's ID, making it possible to trace an MS and MO relationship.

Let a parameter set (n_O , n_M , N_m , ϕ , and parameters that determine $f(\lambda)$) be given. We simulated a population history in which N_m mothers generated offspring; this process was repeated 100 times. The sampling process for each history was repeated 10,000 times, acquiring 1,000,000 data points that were utilized to construct the distribution of the estimators for each parameter set. Furthermore, true value of $N_{e,m}$ was calculated from N_m and c (Eqs. 2 and 4).

3 RESULT

We numerically evaluated the performance of $\widehat{N_{e,m}/N_m}$ for the case in which the number of mothers, N_m , and the combined effect of deviation from the Poisson, c , were unknown. Moreover, we changed the parameter values for N_m (10^3 and 10^4) and c (1, 10, 20, and 100). In addition, based on a given parameter set (N_m and c), we mainly addressed the number of samples (n_M and

n_O) required to obtain adequate accuracy and precision. In this study, we evaluated the performance of $\widehat{N_{e,m}/N_m}$ for specific ranges of the sample sizes (50-200 when $N_m = 10^3$, and 200-1000 when $N_m = 10^4$). Meanwhile, other estimators (i.e., $\widehat{N_{e,m}}$ and $\widehat{1/N_m}$) are also evaluated and provided in **Supporting Information**.

First, we evaluated the accuracy of estimators based on their relative bias calculated by applying the individual-based model, which is defined as follows: “(averaged estimator – true value)/true value.” For a given combination of N_m and c , the value of the relative error of $\widehat{N_{e,m}/N_m}$ is represented on a heatmap as a function of n_M and n_O , as depicted in **Fig. 2**. For most of the investigated parameter sets, we observed that their relative error is less than 10%. As expected, the relative error is not affected by n_M since $\widehat{1/N_m}$ is exactly an unbiased estimator of $1/N_m$ (see Eq. A2 in **APPENDIX** and also **Fig. S2** in **Supporting Information**). Meanwhile, $\widehat{N_{e,m}}$ is downwardly biased when n_O is relatively small to true $N_{e,m}$ (e.g., see $c = 1$ in **Fig. 2** and also **Fig. S1** in **Supporting Information**), as presented in Akita (2019); thus, $\widehat{N_{e,m}/N_m}$ is downwardly biased. Contrary to the theoretical prediction for the direction of the bias (Eq. 11), relatively strong overdispersion results in an upwardly bias for $\widehat{N_{e,m}/N_m}$ when c is relatively large (e.g., $c = 20$ and 100 in **Fig. 2a**). This inconsistency may be caused by the breakdown of the approximation for deriving $\widehat{N_{e,m}}$ (Eq. S14 in Akita, 2019). Thus, as described in Eq. 3, extremely large c results in a large variance of offspring number, generating a situation in which the behavior of random variable H_{MS} far deviates from the binomial distribution.

Next, we evaluated the precision of estimators based on their coefficient of variation. As demonstrated in **Fig. 3**, the value of the coefficient of variation of $\widehat{N_{e,m}/N_m}$ is also represented on a heatmap as a function of n_M and n_O . For the investigated parameter sets, the degree of the coefficient of variation strongly depends on the sample sizes. As shown in **Figs. S3 and S4** in **Supporting Information**, the dependency results from the combined effects of variation of

both $\widehat{1/N_m}$ and $\widehat{N_{e,m}}$. As c increases, it is noteworthy that the parameter space of sample sizes demonstrating large variation of $\widehat{1/N_m}$ (e.g., $CV > 30\%$) expands; however, when c is small (e.g., $c = 1$), relatively small n_O results in large variation of $\widehat{N_{e,m}}$ because of a relatively large $N_{e,m}$.

4 DISCUSSION

We theoretically developed a nearly unbiased estimator of the ratio of contemporary effective mother size to the census size ($N_{e,m}/N_m$) in a population (Eq. 10). The proposed estimator is based on known MO relationship and MS relationships observed within the same cohort, in which sampled individuals in the cohort probably share MO relationships with sampled mothers (**Fig 1**). Moreover, the performance of the estimator (accuracy and precision) was quantitatively evaluated by running an individual-based model (**Figs. 2 and 3**). Meanwhile, the bias is analytically provided (Eq. 11). Our modeling framework utilizes two types of reproductive variations (Akita, 2019): variance of the average offspring number per mother (parental variation, denoted by $f(\lambda)$), and variance of the offspring number across mothers with the same reproductive potential (nonparental variation, denoted by ϕ). Additionally, these two effects result in a skewed distribution of offspring number and are summarized into one parameter (c) in the framework. Thus, our estimator can be calculated from sample sizes of mother and offspring (n_M and n_O , respectively) and the observed numbers of MS and MO pairs (H_{MS}^{obs} and H_{MO}^{obs} , respectively), and it does not require other parameters. The rationale for this is the following: i) the frequency of MS and MO pairs contains information about $N_{e,m}$ and N_m , respectively; ii) the estimators of $N_{e,m}$ and $1/N_m$ are independently determined based on a pedigree structure in the population and sample sizes, generating the estimator of $N_{e,m}/N_m$ by multiplying both estimators ($\widehat{N_{e,m}}$ and $\widehat{1/N_m}$). In this study, although $N_{e,m}/N_m$ is considered as a proxy of N_e/N , our theoretical results can easily

be extended to the estimator of the ratio of contemporary effective father size to the census size if fathers are also sampled. The comparison of both ratios could clarify the underlying processes that differentiate between the sexes in the context of reproductive ecology.

The novelty of this study is that $\widehat{N_{e,m}/N_m}$ can be obtained only from the genetic data, and there are numerous advantages in using the proposed estimator instead of separately estimating N_e (via genetical method) and N (via non-genetical method). First, sampling and analyzing designs have become substantially simplified. Moreover, requirements for the proposed estimator are sampling of mothers and (potentially) their offspring in an appropriate time, and the extraction of their DNA that satisfies an adequate number of markers for kinship detection. In addition, both MO and MS pairs can be detected by a applying unified framework of genetic analyzes (there are many algorithms to detect kinship pairs from single nucleotide polymorphisms (SNPs) or short tandem repeats (STRs)), although an MS pair involves many more DNA markers (e.g., several thousands of SNPs are required for detection) than an MO pair (e.g., several hundreds of SNPs are required for detection). Second, our theoretical results guide sample sizes (n_M and n_O) to ensure the required accuracy and precision, especially if the order of the number of effective mothers is approximately known. This is due to the simple formulation of the estimator determined only by the observed values (Eq. 10). Third, the proposed estimator directly reflects the amounts of $N_{e,m}$ and N_m at the same timing (i.e., immediately after the end of the reproductive season), leading to a clear interpretation of the results, especially for genetic monitoring. For example, when the strong cohort is added to the spawning population in the beginning of the year, the estimator of N_e without reflecting this addition may results in an inappropriate estimation of N_e/N (details of the temporal scale relevant to estimated N_e for each method were discussed in Wang et al., 2016).

Our modeling framework is presented by combining the context of the sibship assignment method (for estimating $N_{e,m}$) and the CKMR method (for estimating $1/N_m$), which defines a

kinship-oriented estimation of effective/census population size. Therefore, improvements to these methods directly contribute to the estimation of N_e/N . Furthermore, the original theory of the sibship assignment method requires HS and FS pairs but does not require a distinction between the MS and PS pairs. This is a significant advantage due to the difficulty of the distinction from genetic data. However, the limitation of using MS or PS pair enables us to employ a nearly unbiased estimator of N_e for particular sex (Akita, 2019), which greatly improves the accuracy of the estimation of the $N_{e,m}$ in this study and thus that of $N_{e,m}/N_m$. It is noteworthy that the estimator of $1/N$ is given by

$$\widehat{\left(\frac{1}{N}\right)} = \frac{H_{PO}^{obs}}{2n_P n_O}, \quad (12)$$

where n_P and H_{PO}^{obs} denotes the sample size of the parent and the observed number of parent–offspring (PO) pairs in a sample, respectively (Bravington, Skaug, & Anderson, 2016). The development of the unbiased estimator of N_e without a distinction between MS and PS pairs that could provide an unbiased estimator of N_e/N coupled with Eq. 12, is a study for the future. Furthermore, using cross-cohort HS pairs, the CKMR method also provides the estimator of N (Bravington, Skaug, & Anderson, 2016) that does not require the sampling of the parent, which probably provides the estimator of N_e/N only from unmatured samples. This perspective of the study will also be conducted in the future.

Acknowledgments

The author thanks R. Nakamichi for fruitful discussions. This work was supported by JSPS KAKENHI Grant Number 19K06862.

References

- Akita, T. (2018). Statistical test for detecting overdispersion in offspring number based on kinship information. *Population Ecology*, 60(4), 297–308. Retrieved from <https://doi.org/10.1007/s10144-018-0629-2> doi: 10.1007/s10144-018-0629-2
- Akita, T. (2019). Nearly unbiased estimator of contemporary effective mother size using within-cohort maternal sibling pairs incorporating parental and nonparental reproductive variations. *Heredity*. Retrieved from <http://dx.doi.org/10.1038/s41437-019-0271-6> doi: 10.1038/s41437-019-0271-6
- Bravington, M. V., Grewe, P. M., & Davies, C. R. (2016). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nat Commun*, 7, 13162. Retrieved from <http://dx.doi.org/10.1038/ncomms13162> doi: 10.1038/ncomms13162
- Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Stat Sci*, 31(2), 259–274.
- Frankham, R., Bradshaw, C. J., & Brook, B. W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biological Conservation*, 170, 56 - 63. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0006320713004576> doi: <https://doi.org/10.1016/j.biocon.2013.12.036>
- Hedgcock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *B Mar Sci*, 87(4), 971–1002.
- Hillary, R. M., Bravington, M. V., Patterson, T. A., Grewe, P., Bradford, R., Feutry, P., ... et al. (2018). Genetic relatedness reveals total population size of white sharks in eastern australia and new zealand. *Scientific Reports*, 8(1), 2661. Retrieved from

- <http://dx.doi.org/10.1038/s41598-018-20593-w> doi: 10.1038/s41598-018-20593-w
- Hixon, M. A., Johnson, D. W., & Sogard, S. M. (2014). BOFFFFs: on the importance of conserving old-growth age structure in fishery populations. *ICES Journal of Marine Science*, 71(8), 2171–2185. Retrieved from <http://dx.doi.org/10.1093/icesjms/fst200> doi: 10.1093/icesjms/fst200
- Kéry, M., & Schaub, M. (2011). *Bayesian population analysis using winbugs: a hierarchical perspective*. Academic Press.
- Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv Genet*, 11(2), 355–373.
- Marandel, F., Lorance, P., Berthelé, O., Trenkel, V. M., Waples, R. S., & Lamy, J.-B. (2018). Estimating effective population size of large marine populations, is it feasible? *Fish and Fisheries*, 20(1), 189–198. Retrieved from <http://dx.doi.org/10.1111/faf.12338> doi: 10.1111/faf.12338
- Methot, R. D., & Wetzel, C. R. (2013). Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research*, 142, 86–99.
- Nielsen, R., Mattila, D. K., Clapham, P. J., & Palsbøll, P. J. (2001). Statistical approaches to paternity analysis in natural populations and applications to the north atlantic humpback whale. *Genetics*, 157(4), 1673–1682.
- Nordborg, M., & Krone, S. M. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern developments in theoretical population genetics: The legacy of gustave malécot* (pp. 194–232). Oxford University Press Oxford.
- Palstra, F. P., & Fraser, D. J. (2012). Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution*, 2(9), 2357–2365. Retrieved from

<http://dx.doi.org/10.1002/ece3.329> doi: 10.1002/ece3.329

- Pearse, D., Eckerman, C., Janzen, F., & Avise, J. (2001). A genetic analogue of ‘mark–recapture’ methods for estimating population size: an approach based on molecular parentage assessments. *Molecular ecology*, *10*(11), 2711–2718.
- Quinn, T. J., & Deriso, R. B. (1999). *Quantitative fish dynamics*. oxford university Press.
- Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters* (Vol. 8). Blackburn press Caldwell, New Jersey.
- Skaug, H. J. (2001). Allele-sharing methods for estimation of population size. *Biometrics*, *57*(3), 750–756.
- Skaug, H. J. (2017). The parent–offspring probability when sampling age-structured populations. *Theor popul biol*, *118*, 20–26.
- Wang, J. (2009). A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology*, *18*(10), 2148–2164.
- Wang, J., Santiago, E., & Caballero, A. (2016). Prediction and estimation of effective population size. *Heredity (Edinb)*, *117*(4), 193–206. doi: 10.1038/hdy.2016.43
- Waples, R. S. (2016). Tiny estimates of the N_e/N ratio in marine fishes: Are they real? *J Fish Biol*, *89*(6), 2479–2504. Retrieved from <http://dx.doi.org/10.1111/jfb.13143> doi: 10.1111/jfb.13143

Table 1: The list of mathematical symbols employed in the main text

n_M	Sample number of mother
n_O	Sample number of offspring
N_m	Number of mothers in the population when sampled offspring are born
$N_{e,m}$	Effective number of mothers in the population
ϕ	Overdispersion parameter under negative binomial reproduction
λ_i	Expected number of surviving offspring of mother i at sampling
$f(\lambda)$	Frequency of λ for all mothers.
c	Combined effect of deviation from the Poisson (= $(1 + \phi^{-1})\mathbb{E}[\lambda^2]/\mathbb{E}[\lambda]^2$)
k_i	Number of surviving offspring born to mother i
H_{MO}	Number of mother–offspring pairs observed in samples
H_{MS}	Number of maternal-sibling pairs observed in samples
π_{MO}	Probability that a randomly selected pair (mother and offspring) shares a mother-offspring relationship
π_{MS}	Probability that a randomly selected pair (two offspring) shares a maternal-sibling relationship
b	Bias of $\widehat{N_{e,m}}/N_m$

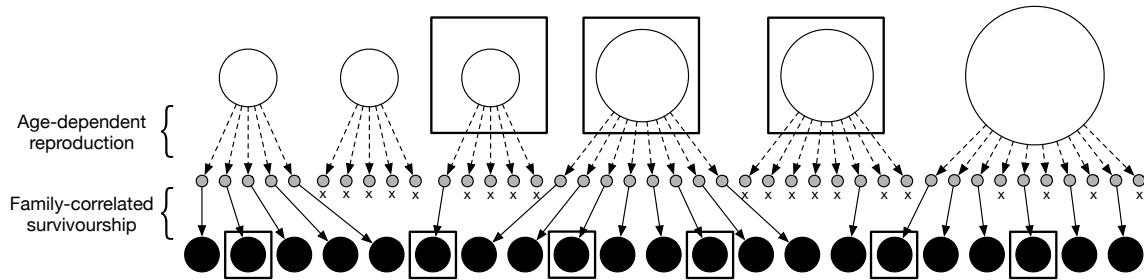


FIGURE 1 Example of relationships between mothers and their offspring number. The open, gray, and black circles represent mothers, their eggs, and their offspring, respectively. The area of an open circle indicates the degree of reproductive potential of each mother (i.e., λ_i). The dotted and thin arrows denote mother-egg and egg-offspring relationships, respectively. The symbol x denotes a failure to survive at sampling. Sampled individuals are denoted with a bold line.

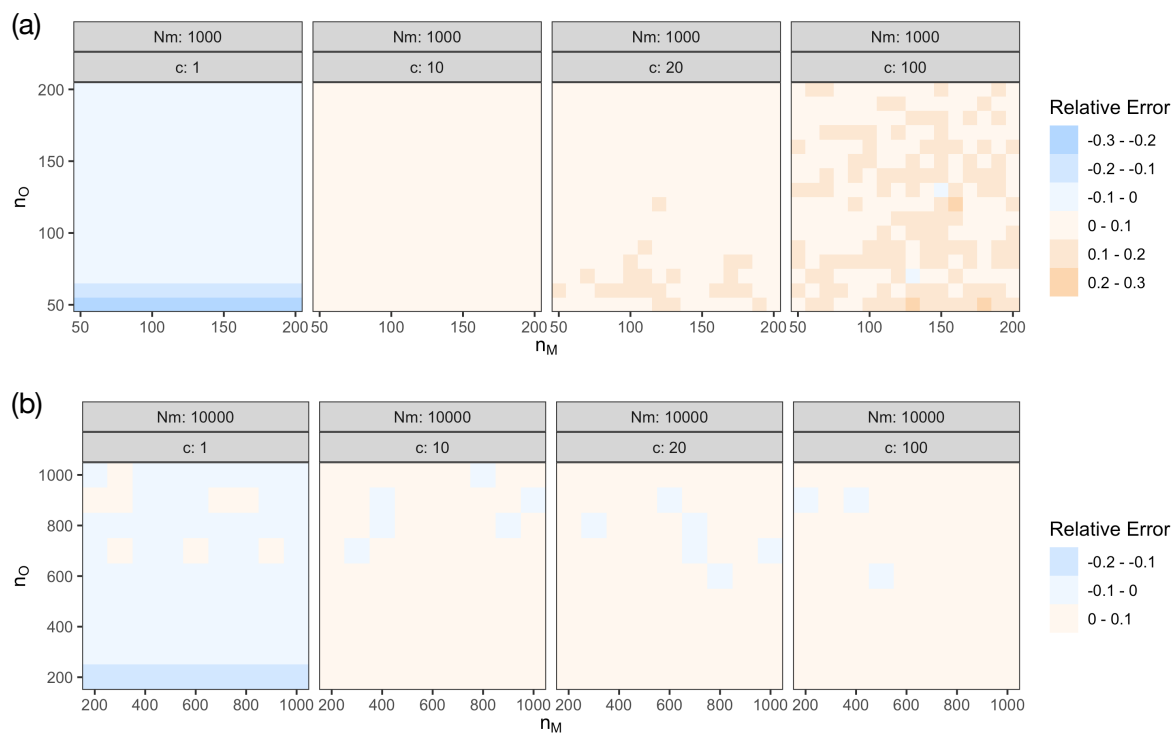


FIGURE 2 Heatmap showing the relative error of $\widehat{N_{e,m}}/N_m$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.

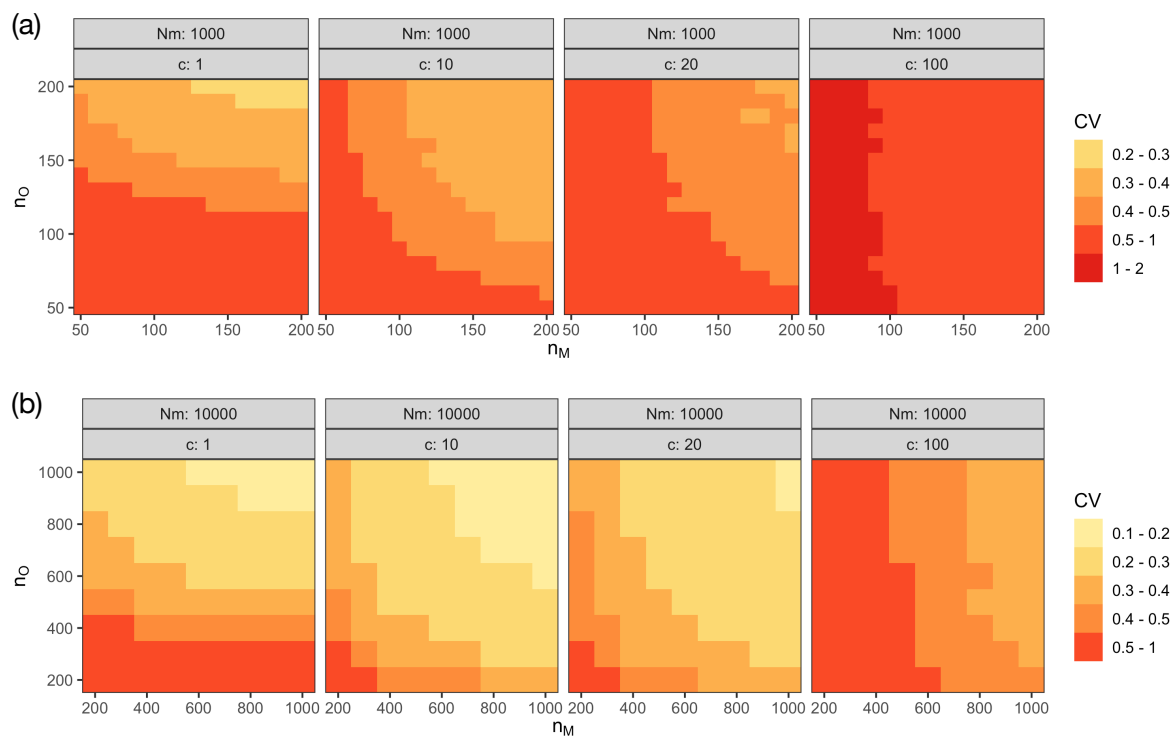


FIGURE 3 Heatmap showing the coefficient of variation of $\widehat{N_{e,m}}/N_m$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.

APPENDIX

Derivation of the bias of $\widehat{N_{e,m}/N_m}$

For calculation of the bias of $\widehat{N_{e,m}/N_m}$, we require an expectation of the estimator given by

$$\mathbb{E} \left[\left(\frac{\widehat{N_{e,m}}}{\widehat{N_m}} \right) \right] = \text{COV} \left[\left(\frac{1}{\widehat{N_m}} \right), \widehat{N_{e,m}} \right] + \mathbb{E} \left[\left(\frac{1}{\widehat{N_m}} \right) \right] \mathbb{E} \left[\widehat{N_{e,m}} \right]. \quad (\text{A1})$$

As stated in the main text, both $\widehat{N_{e,m}}$ and $1/\widehat{N_m}$ are independent. Thus, the first term in the right-hand side of Eq. A1 can be ignored. The expectation of $1/\widehat{N_m}$ is given by

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{\widehat{N_m}} \right) \right] &= \frac{\mathbb{E} [H_{MO}]}{n_M n_O} \\ &= \frac{1}{N_m}. \end{aligned} \quad (\text{A2})$$

From the first to the second line of Eq. A2, we applied the relationship $\pi_{MO} = \mathbb{E}[\pi_{MO}] = \mathbb{E}[H_{MO}]/(n_M n_O)$ and Eq. 8. Equation A2 indicates that $1/\widehat{N_m}$ is the unbiased estimator. The expectation of $\widehat{N_{e,m}}$ is given by

$$\mathbb{E}[\widehat{N_{e,m}}] = N_{e,m} - N_{e,m} \left(1 - \frac{1}{N_{e,m}} \right)^{\binom{n_O}{2} + 1}, \quad (\text{A3})$$

which is illustrated in Appendix D of Akita (2019). Together with these relationships, we can obtain the bias of $\widehat{N_{e,m}/N_m}$ described in Eq. 11.

Supporting Information

Probability density function and its moment of λ

As stated in the main text, our modeling framework does not require the specific form of $f(\lambda)$; it only requires the ratio of the second moment to the squared first moment ($\mathbb{E}[\lambda^2]/\mathbb{E}[\lambda]^2$) instead. However, the specific form is required for evaluating the theoretical results (i.e., calculating the moment or running the individual-based model). Herein, we model an age-structured fish population that serves as a representative example, demonstrating both parental and nonparental variations. The following contents are almost the same as those of Akita (2019) except for the parameter values that produce the setting $c = 20$ and 100.

Suppose that the mean fecundity of a mother depends on her age. Let λ_a denote the mean fecundity, which is a function of age (denoted by a). The moment can be defined as $\mathbb{E}[\lambda^m] = \sum_{a=0}^{a_{\max}} \lambda_a^m h_{\text{mat}}(a)$, where $h_{\text{mat}}(a)$ is the frequency of mature mothers at a given age, and a_{\max} denotes the maximum age. Thus, we can numerically obtain the moment by applying λ_a and $h_{\text{mat}}(a)$.

For marine species with a type-III survivorship curve, it is generally assumed that individual fecundity is proportional to weight. By utilizing the von Bertalanffy growth equation for body weight, λ_a is explicitly defined as a function of age as follows:

$$\lambda_a \propto (1 - \exp[-\kappa(a - a_0)])^\beta, \quad (\text{S1})$$

where κ , a_0 , and β are conventionally used parameters in the von Bertalanffy equation, and they denote the growth rate, the adjuster of the equation for the initial size of the animal, and the allometric growth parameter, respectively. To obtain a specific value of λ , a coefficient value of 10 multiplied by the right-hand side of Eq. S1 was used when running the individual-based model.

The frequency of mature mothers at a given age can be given as the following:

$$h_{\text{mat}}(a) \propto h(a)Q(a), \quad (\text{S2})$$

satisfying $\sum_{a=0}^{a_{\text{max}}} h_{\text{mat}}(a) = 1$, where $h(a)$ and $Q(a)$ denote the frequency and maturity at a given age, respectively. Although $f(a)$ is affected by historical population dynamics and age-dependent survival, for simplicity, the mortality rate is assumed to be constant (i.e., age independent):

$$h(a) \propto \begin{cases} S^a & \text{if } a < a_{\text{max}} \\ 0 & \text{if } a = a_{\text{max}} \end{cases}, \quad (\text{S3})$$

where S denotes a survival probability. The maturity at age ($Q(a)$) is assumed to be a knife-edge function, which is given by

$$Q(a) = \begin{cases} 1 & \text{if } a \geq a_{\text{mat}} \\ 0 & \text{otherwise} \end{cases}, \quad (\text{S4})$$

where a_{mat} denotes the mature age.

To calculate $\mathbb{E}[\lambda^2]/\mathbb{E}[\lambda]^2$, the required parameter set is $(a_{\text{max}}, \kappa, a_0, \beta, S, a_{\text{mat}})$. In this study, for the purpose of representation, we fixed the values of several parameters as follows: $a_{\text{max}} = 20$, $\kappa = 0.3$, $a_0 = 0$, $S = 0.5$ and $a_{\text{mat}} = 0$. In addition, we selected parameter value c ($= (1 + \phi^{-1})\mathbb{E}[\lambda^2]/\mathbb{E}[\lambda]^2$) to be 1, 10, 20, and 100 for comparison with the results in the main text that are derived from the parameter set $(\phi, \beta) = (1000, 0.0009)$, $(0.1302, 0.9)$, $(0.06111, 0.9)$, and $(0.01165, 0.9)$, respectively.

Finally, we provide specific forms of $f(\lambda)$; thus, when λ_a and $h_{\text{mat}}(a)$ are obtained, $f(\lambda)$ is

given by

$$f(\lambda) = \begin{cases} h_{\text{mat}}(a) & \text{if } \lambda = \lambda_a \\ 0 & \text{otherwise} \end{cases} . \quad (\text{S5})$$

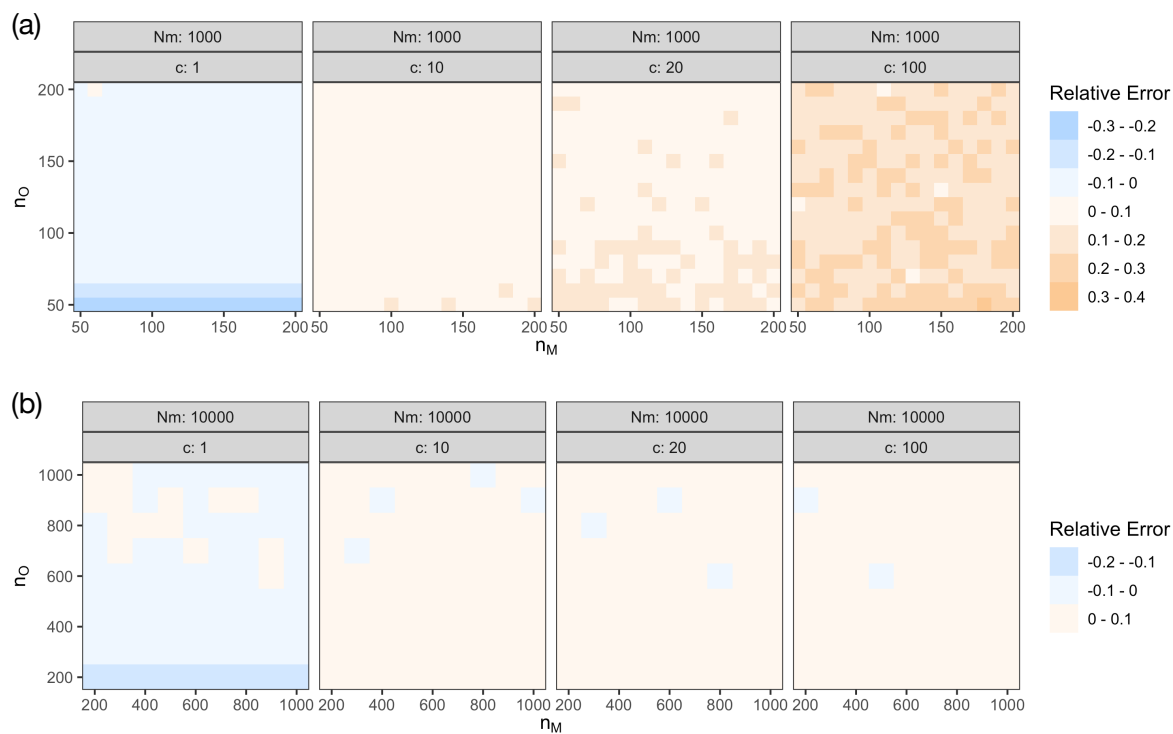


FIGURE S1 Heatmap showing the relative error of $\widehat{N}_{e,m}$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.

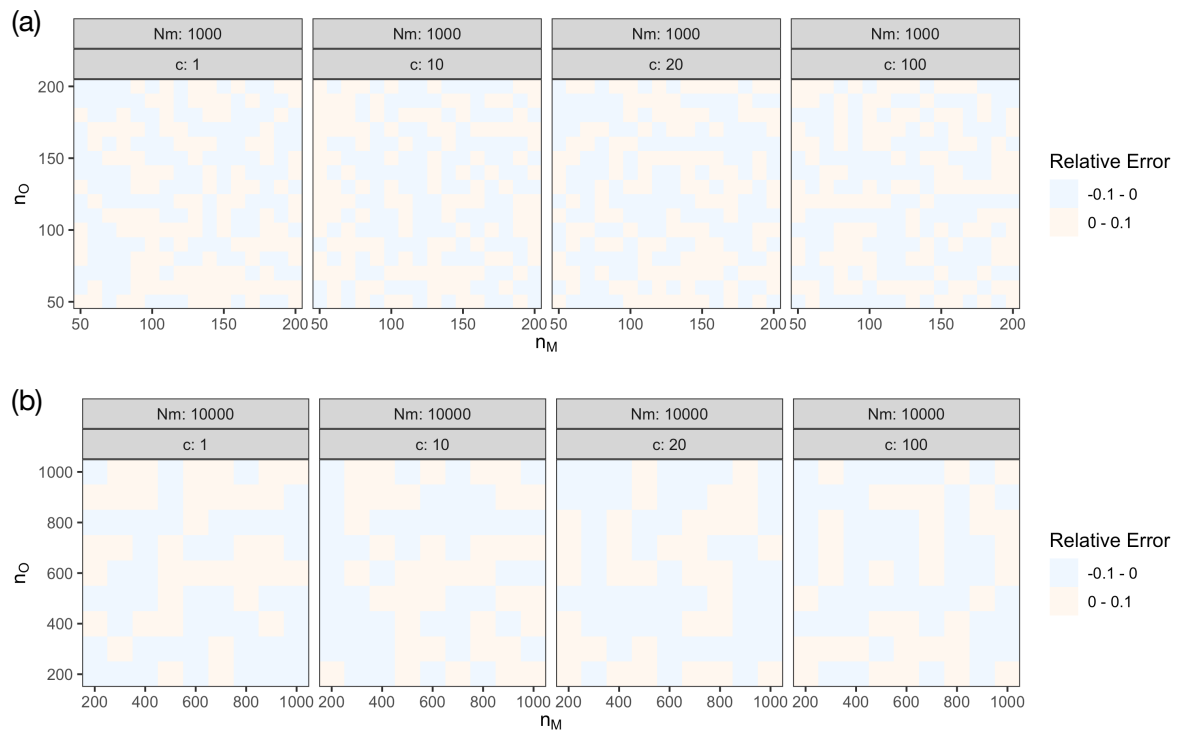


FIGURE S2 Heatmap showing the relative error of $1/\widehat{N}_m$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.

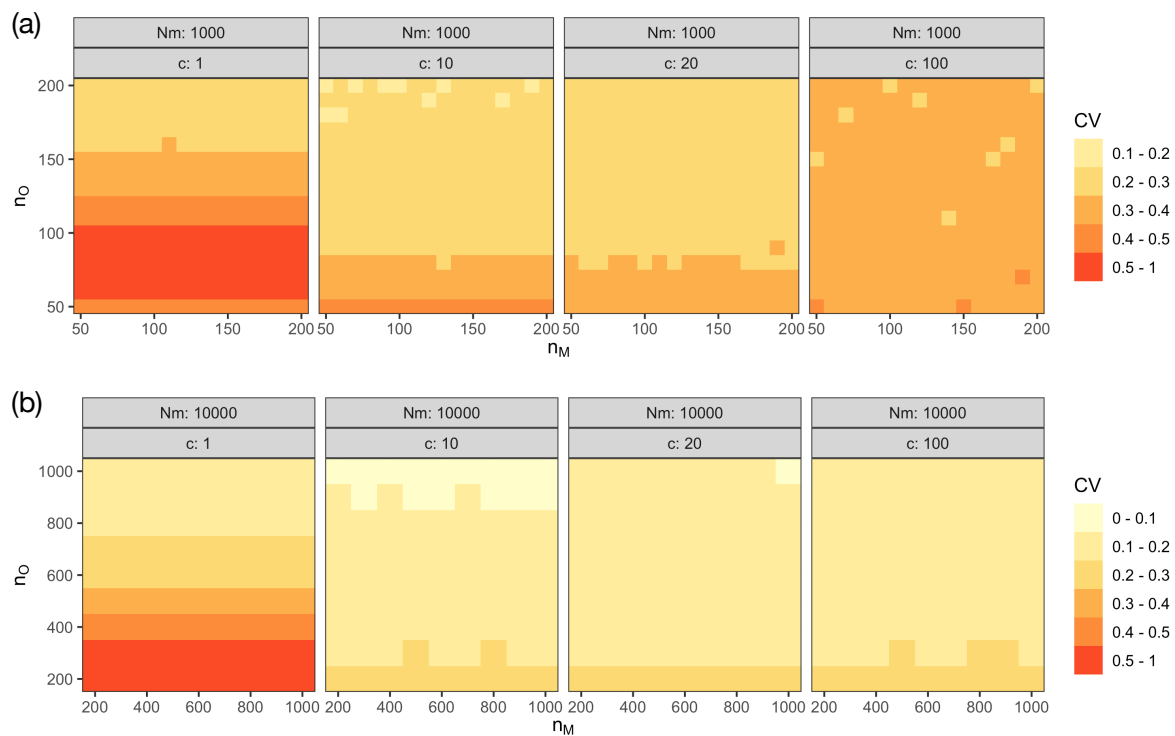


FIGURE S3 Heatmap showing the coefficient of variation of $\widehat{N}_{e,m}$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.

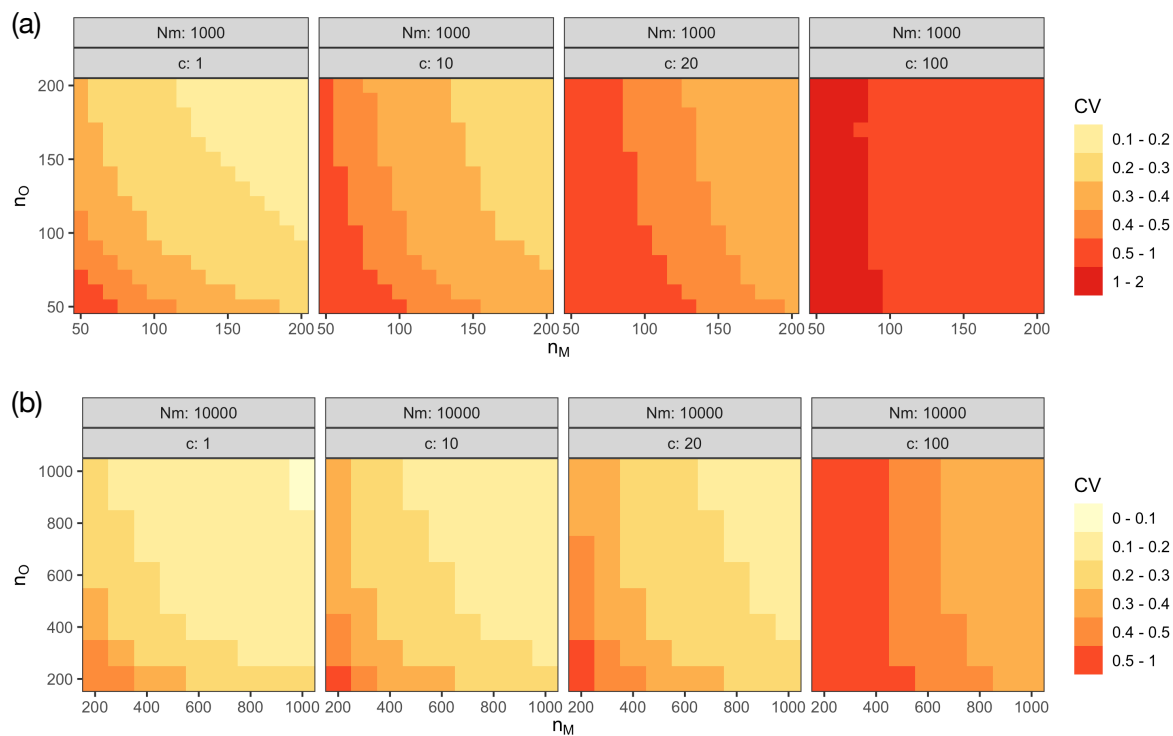


FIGURE S4 Heatmap showing the coefficient of variation of $1/\widehat{N}_m$ as a function of both n_M and n_O : (a) $N_m = 1,000$, (b) $N_m = 10,000$.