universität wien

MASTERARBEIT / MASTER THESIS

Titel der Masterarbeit / Title of the Master Thesis

"Variance Effective Population Size

in the context of different mating systems"

verfasst von / submitted by

Alina Kadluba, BSc BEd MEd

angestrebter akademischer Grad/ in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Jitka Polechová, Privatdoz. PhD

Contents

List of Figures

Abstract (Deutsch)

Die effektive Populationsgröße ist ein in der Populationssgenetik und in der Biologie weit verbreitetes Konzept. Sie beschreibt die Größe einer idealisierten Population mit denselben genetischen Eigenschaften wie die reale Population. Verschiedene Maße des genetischen Drifts führen zu unterschiedlichen effektiven Populationsgrößen, wobei das gewählte Maß vom Kontext abhängt.

Das Ziel dieser Arbeit ist es, den Einfluss verschiedener Paarungsschemata in Tierpopulationen mit überlappenden Generationen auf die so genannte Varianz Effektive Populationsgröße zu bestimmen. Erstens wird in jedem Paarungsschema angenommen, dass die weibliche Fruchtbarkeit binomialverteilt ist. Im Gegensatz dazu ist die Varianz der männlichen Fruchtbarkeit abhängig vom Paarungssystem und muss daher berechnet werden. So muss beispielsweise in einer Harempopulation die Haremsgröße bei der Bestimmung der Varianz der männlichen Fruchtbarkeit berücksichtigt werden. Bei der Dominanzpolygynie ist der Anteil der dominanten Männchen zu berücksichtigen. Bei der Lotterie-Polygynie spielt die durchschnittliche Anzahl der Partner pro brütendem Weibchen eine wichtige Rolle.

In der vorliegenden Arbeit werden die Modelle verglichen und es wird analysiert, welche Parameter den größten Einfluss auf die effektive Populationsgröße haben. Darüber hinaus wird aufgezeigt, wie die Parameter gewählt werden können, um für verschiedene Paarungssysteme gleichwertige effektive Populationsgrößen zu erhalten.

Die Ergebnisse deuten darauf hin, dass die effektive Populationsgröße bei Dominanz- und Haremspolygynien ähnliche Werte für $r_h = 1 - r_d$ aufweisen, wobei r_h das Geschlechterverhältnis für die Haremspolygynie und r_d das Geschlechterverhältnis für die Dominanzpolygynie sind. Der Hauptunterschied besteht darin, dass der Anteil der Männchen bei der Harempolygynie relativ klein sein sollte, um eine große effektive Populationsgröße zu erhalten. Im Gegensatz dazu ist bei der Dominanzpolygamie ein großer Anteil von Männchen erforderlich, um eine große effektive Populationsgröße zu erreichen. Ein weiteres Ergebnis zeigt, dass der Anteil der brütenden Weibchen einen starken Einfluss auf die effektive Populationsgröße hat. Außerdem hängt bei der Dominanzpolygynie die effektive Populationsgröße stark vom Anteil der dominanten Männchen ab.

Abstract (English)

The effective population size is a widely used concept in conservation genetics and biology. It describes the size of an idealized population with the same genetic properties as the real population. Different measures of genetic drift lead to different effective population sizes where the chosen measure depends on contexts. Thus, there are many ways how an effective population size model can be estimated and computed.

In the research project presented here, the variance effective population size was used to observe a special effective population size model in the context of mating systems. An extension of a variance effective population size model, published in 1972 by Hill, additionally considers overlapping generations. This model can be applied to different mating systems, as the harem polygyny, the lottery polygyny and the dominance polygyny.

My aim in this thesis is to determine the influence of different mating schemes in animal populations with overlapping generations on the so called variance effective population size. Firstly, in every mating scheme the female fecundity is assumed to be binomial distributed. In contrast, the variance of male fecundity depends on the mating system and therefore has to be determined. For example in the harem polygyny the harem size must be taken into account for determining the variance of male fecundity. As other instance, in the dominance polygyny setting the proportion of dominant males has to be considered. In lottery polygyny the average number of mates per breeding female plays an important role in computing the variance of male fecundity. The different variances in fecundity result in different effective population size models.

I compare the models to analyse which parameters have the greatest impact on the effective population size. Furthermore, it is shown how the parameters can be chosen to yield equivalent effective population numbers for different mating systems. Lastly, I show examples of how the effective population size can be computed for special animal species.

The results suggest that in dominance and harem polygyny the effective population size have similar values for $r_h = 1 - r_d$, where r_h and r_d are the sex ratios for the harem and the dominance polygyny, respectively. The main difference is that the sex ratio in the harem polygyny should be relatively small to obtain a large effective population size. In contrast, in a dominance polygyny a high sex ratio is needed to obtain a large effective population size. Another result indicates that the proportion of breeding females strongly influences the effective population size. Furthermore, in dominance polygyny, the effective population size strongly depends on the proportion of dominant males.

Keywords: *effective population size, mating system, sex ratio, survival probability, dominance, genetic drift, variance in fecundity*

1 Introduction

The population size of an animal or a plant population is indicated by its census size N. If this population size is finite, a loss of genetic diversity often occurs due to different reasons as inbreeding, unequal sex ratio, migration, selection etc. All these processes influence the so-called effective population size.

The effective population size is shortened by N_e and describes the size of an idealized population with the same genetic properties as the natural population. The effective population size of an actual population is the number of individuals in a theoretically ideal population with the same amount of genetic drift as the actual population. Since the census size is often the only piece of demographic information one has about a population, the effective population size, N_e, and the ratio $\frac{N_e}{N}$, respectively, are values that cannot be computed and used easily. Unequal sex ratio in a population, fluctuation in population size, and variance in family sizes are, for example, parameters that affect effective population size.

Since N_e is a measure of genetic diversity in a population, it strongly influences conservation genetics and biology. Therefore, many models and formulas have been established for computing effective population sizes. It is important to know that there is not "one" effective population size for a given population, but it depends on the effective population size one wants to estimate and the approach used to compute the model.

In 1972, Hill [13] establishes, among others, what is an extension of Wright's [21] variance effective population size model. This model, unlike many of its predecessors, additionally considers overlapping generations. Nunney [15] extends Hill's formula to apply the model to various mating systems, such as the harem polygyny, the lottery polygyny and the dominance polygyny.

Using different variances of male fecundity in different mating systems shows that the mating behaviour of wild animals likely affects $\frac{N_e}{N}$. Polygamous animals are expected to have lower effective population sizes than monogamous populations, due to unequal sex-ratio and high variance of male gametic contributions. If only few males pass on their genes, genetic diversity shrinks. In this case, the variance of male fecundity is considered. However, the variance of male fecundity depends on the assumed mating system. These different variances result in different effective population size models. The variance of female fecundity is assumed to be binomially distributed in every mating system. Parameters, such as the number of females in one harem, the proportion of males and females, the proportion of breeding females, the survival probabilities of an individual and the average number of offspring per female influence the models. Depending on how the parameters are chosen, the effective population sizes have different magnitudes. This Master's thesis aims to determine the differences and similarities of before-mentioned effective population size models.

Firstly, the general concept of the effective population size shall be introduced. Later, different formulas of the Ne's will be derived to provide clearer understanding of the variance effective population size. Afterwards, the special effective population size for overlapping generations shall be examined in the context of different mating systems. The mating systems in question are the harem polygyny, the lottery polygyny and the dominance polygyny which will be explained later. Each mating system leads to a different outcome regarding the effective population size. Therefore, the following research questions shall be answered in this Master's thesis

Research questions:

RQ1. How do the effective population sizes change in different mating systems?

RQ2. Which parameters lead to differences between the mating systems (harem vs. lottery, harem vs. dominance and dominance vs. lottery)?

RQ3. Which parameters have a strong impact on effective population size models and which parameters do not significantly influence the models?

In the following chapter, the most important general biological terms will be presented.

2 General Biological Terms

In this chapter general biological terms shall be explained. Some of them are of universal interest and some are important to understand the later mathematical explanations.

Diploid and Haploid
Body cells of most animals and plants have two chromosome sets and are therefore called diploid. The chromosomes contain of DNA-strands that carry the genes, i.e. the genetic information. Since one cell contains a double chromosome set from the maternal and the paternal inheritance, they are called diploid ($2N$). Germ cells only have one chromosome set and are therefore called haploid ($1N$). Besides diploid and haploid organisms also polyploids are common among plants and some groups of fish and amphibians. Unless otherwise stated, I will focus on diploids.

Alleles
As already explained the DNA-strands in a chromosome set consists of genes. The genes on a chromosome can vary within a species. This different gene possibilities are called alleles. If one gene-locus, for example, codes for the hair color, there are alleles for brown hair, for blond hair, for black hair etc. The parents of an individual can have different hair colors; the offspring gets the allele of one of the parents.

Inbreeding coefficient

The inbreeding coefficient is often defined as the probability of identity by descent of a pair of randomly sampled alleles. This means when related individuals mate, the same alleles can be inherited and therefore, there is an identity by descent. Two different allelic copies of a given nucleotide site drawn from a population are identical by descent if they trace their ancestry back to a single ancestral copy. At the concept "identical by descent" one starts with a given population at generation 0. Then, t generations later, a new generation emerged from this initial population through inheritance, sampling etc. Two alleles from generation t are identical by descent if they are progeny of the same allele in the initial generation. Measured with respect to an initial generation, in which all the copies of a site are arbitrarily decreed not to be identical by descent. The inbreeding coefficient at time t, f_t, can be computed by the following formula

$$1 - f_t = (1 - f_0)(1 - \frac{1}{2N})^t,$$

where N is the number of individuals in the population. The reproductive success depends on the genetic similarity of the mating individuals. If the genetic similarity is too big (i.e. siblings mating), the inbreeding coefficient is relative high.

3 Basic deviations from the idealised population

3.1 The Wright-Fisher Model

In this chapter the basic Wright-Fisher model shall be introduced. I am referring to a paper by P. Sjödin and other authors [20] as to a paper by Crow and Denniston [3].

Sewall Wright and Ronald A. Fisher were geneticists and biologists at the beginning of the 20^{th} century. Together with Haldane they were the founders of the population genetics. Wright [21] and Fisher [9] [8] established a model to compute an effective population size. This model is known as the "null model" or the "idealised model". The Wright-Fisher model underlies three properties that are used in defining this effective population size. The three properties that are often used in defining N_e are the probability of identity-by-descent of two alleles chosen at random, the variance in offspring allele frequency and the leading non-unit eigenvalue of the allele frequency transition matrix [20]. Therefore the effective population sizes "inbreeding effective population size", "variance effective population size" and "eigenvalue effective population size" can be computed. These N_e's can be similar in some special situations, but they can also be very different or it could be that they do not exist at all.

At this point it is pertinent to give a short motivation why it is so important to use the effective population size. Let us assume a population consists of 5 males and 20 females that live in a wood. After separating the population (i.e. by a motorway) only one male and 15 females live together. In the next generation mating and breeding processes lead to the situation that 5 males and 15 females are in the population again. It seems like nothing has changed from the initial situation to the current situation. However, the big difference is the fact that in generation 2 there are only half-siblings or full siblings since they all have the same father. This leads to a loss of genetic diversity.

Generation	**Males**	**Females**	**Situation**
0	5	20	habitat: wood
1	1	15	because of motorway
2	5	20	only (half)siblings

This example shows that the census size, i.e. the actual population size is not always the same as the effective population size.

Wright and Fisher defined N_e as "the size of [...] an ideal population that has the same rate of increase in homozygosity or gene frequency drift as the actual population under consideration." [3]. The model, established by Wright and Fisher, shall be derived in the following. Wright and Fisher assumed an idealised population in their models.

Idealised in this context means

- Panmixia (random mating)
- Monoecious (i.e. sexes are indistinguishable)
- Constant size (no fluctuations)
- Binomial-distributed "family sizes“.

First of all, one needs to assume that a population with no migration is observed. We assume that there are N individuals in a population and therefore $2N$ gametes. Every individual and therefore every gamete can have progeny. This means that every individual produces a different number of progeny, i.e. the family size can be seen as a random variable. Since constant population size was assumed, $2N$ progeny are produced. The i^{th} individual has k_i progeny and the index i runs from 1 to N. Because of the assumption of an idealised population, the mean $\bar{k_i}$ is 2. Since the variance is defined as the sum of the squared deviations from the mean divided by the total number N, the variance can be defined as

$$\sigma_{k_i}^2 = \frac{1}{N}\sum_{i=1}^{N}(k_i - 2)^2.$$

This is the variance under the gametes that constituted from the individuals of the previous generation. The proportion of cases in which two random gametes come from the same parent is

$$\frac{1}{2N(2N-1)}\sum_{i=1}^{N} k_i(k_i - 1) = \frac{2 + \sigma_{k_i}^2}{4N - 2}. \text{ [3]}$$

This fraction can now be compared to the probability that a random individual is chosen from the set of individuals. The probability of choosing one random individual of the idealised population is $\frac{1}{N_e}$. Therefore,

$$N_e = \frac{4N - 2}{2 + \sigma_{k_i}^2}.$$

When the number of progeny per parent is constant and therefore the variance is zero ($\sigma_{k_i} = 0$), the effective population size reduces to $N_e = 2N - 1$, i.e. for large N the effective population size is approximately double the census size. If the family size k_i is binomial distributed, it approaches the poisson distribution for large N. For the poisson distribution the expected value μ_{k_i} is the same as the variance, i.e.

$$\sigma_{k_i}^2 = \mu_{k_i} = \bar{k_i} = 2.$$

The effective population size becomes approximately $N_e \approx N$ since the number (-2) in the numerator does not play an important role, if N has reached a sufficient level. This basic

Wright-Fisher model can now be observed in different ways.

There are many different effective population sizes, as for example the inbreeding effective population size, the variance effective population size and the eigenvalue effective population size. As already mentioned, in this master thesis the variance effective population size will play an important role. In the following, an investigation will take place as to what will happen if a deviation from the above stated assumption in the Wright-Fisher model, is detected. The first observations refers to the situation when the 1:1 sex ratio does not hold.

3.2 Deviation from the 1:1 sex ratio

One assumption states that the sexes are indistinguishable. Now different numbers of males and females shall be assumed. If the number of males, N_m, and the number of females, N_f, is not equal, i.e. $N_m \neq N_f$, an unequal sex ratio occurs.

Additionally, we know that the probability that 2 genes in generation t derived from a male is $\frac{1}{4}$ and the probability that 2 genes in generation t derived from same male is $\frac{1}{4N_m}$. From this point of view the effective population size can be defined as

$$\frac{1}{4N_m} + \frac{1}{4N_m} = \frac{1}{N_e}$$

and this leads to

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

Consider an example, where $N_m = 1$ and $N_f = 99$; the effective population size is thus $N_e = \frac{4 \cdot 1 \cdot 99}{100} \approx 4$. It can be noted that the effective population size is much smaller than the census size ($4 << 100$) if the sex ratio is very unequal. However, if $N_f = N_m = \frac{N}{2}$, it follows that $N = N_e$ again.

The next assumption for an idealised population is that the population size is constant. If this does not hold, the next deviation from an idealised population can be observed.

3.3 Deviation from the constant population size

For this assumption it is important to work with the heterozygosity through time. If computing this for a population with constant population size, the following formula can be used

$$H_t = (1 - \frac{1}{2N}) \cdot H_{t-1} = (1 - \frac{1}{2N})^{t-1} \cdot H_0.$$

If the population size changes this leads to

$$\frac{H_t}{H_0} = (1 - \frac{1}{2N_1}) \cdot (1 - \frac{1}{2N_2}) \cdot ... \cdot (1 - \frac{1}{2N_{t-1}}).$$

Now the question should be proposed as to what population of constant size would have the same decrease in heterozygosity over the time period involved. This question leads to the definition of the effective population size

$$(1 - \frac{1}{2N_e}) = \prod_{i=1}^{t-1}(1 - \frac{1}{2N_i}).$$

In the next step the logarithm has to be used to make some transformations

$$t \cdot \log(1 - \frac{1}{2N}) = \log(\prod_{i=1}^{t-1}(1 - \frac{1}{2N_i})).$$

To make these terms easier, the fact that $\log(1 - x) \approx x$, if x is small, is used. The term $\frac{1}{2N_i}$ is mostly quite small, because one is divided by a number that is at least 2. This leads to

$$-\frac{t}{2N_e} = \sum_{i=1}^{t-1}(-\frac{1}{2N_i}).$$

After some transformations it can be seen that the effective population size is t divided by the harmonic mean

$$N_e = \frac{t}{\sum_{i=1}^{t-1}\frac{1}{N_i}} = t \cdot \left(\sum_{i=1}^{t-1}\frac{1}{N_i}\right)^{-1}.$$

The harmonic mean is very sensitive to small numbers as the following example shows.

Assuming the population sizes over four generations are the following: $1000, 50, 1000, 1000$; for a set of n numbers $\{x_1, x_2, ..., x_n\}$ the arithmetic mean is

$$\overline{x}_{\text{ari}} = \frac{x_1 + ...x_n}{n}$$

and therefore, the arithmetic mean of this example is

$$N_a = 762{,}5.$$

The formula for the harmonic mean is

$$\overline{x}_{\text{harm}} = \frac{1}{\frac{1}{x_1+...x_n}}$$

and in this example this leads to

$$N_h \approx 174.$$

This number is closer to the small number 50 than to 1000.

3.4 Deviation from the binomial distribution

Deriving the effective population size by observing the distribution of the family sizes, an observation $\sigma^2_{k_i}$, that will be the variance of the contributions from parents to progeny is necessary.

family size k_i = life-time reproductive success of individual i In an idealised population the mean and the variance would be

- mean(k_i) = 2 (one individual produces two gametes to replace itself)
- variance(k_i) = 2 (since for the poisson distribution variance = mean holds)

In the framework of family size distribution, there are still different situations in populations that lead to different effective population sizes. If there is an equal number of males and females, N_e can be written as follows

$$N_e = \frac{4N}{2 + \sigma^2_{k_i}}.$$

If the numbers of males and females are not equal, the following formula can be implemented, as was already used in the first case (unequal males and females)

$$N_e = \frac{\frac{16 \cdot N_m \cdot N_f}{N_m + N_f}}{N_m + N_f}.$$

If additionally self-fertilization is permitted with a proportion of β it leads to

$$N_e = \frac{4N - 2}{2(1 - \alpha) + \sigma^2_{k_i}(1 + \alpha)},$$

where $\alpha = \frac{\beta}{2-\beta}$. It is obvious that the higher the variance, the smaller the effective population size. This makes also sense in a biological way, since an unequal distribution leads to the fact that the genes of some individuals are not represented in the gene pool and the genes of some few individuals are highly represented. The gene diversity therefore shrinks. If a binomial distribution takes place with

$$\sigma^2_{k_i} = 2N \cdot \frac{1}{N} \cdot (1 - \frac{1}{N})$$

it follows that

$$N_e = N$$

again.

Now we have observed what happens if there are single deviations from the idealised population. If we return to the assumptions for an idealised population

- Panmixia (random mating)
- Monoecious
- Constant size (no fluctuations)
- Binomial-distributed "family sizes“.

one sees that we already considered the deviations for the last three assumptions. The deviation from the first assumption was not in consideration yet. It is not an easy task, to take another mating system as the panmixia (random mating) into account, because there are different mating systems that can appear in a population. Therefore, models of polygamy (as for example harem polygyny, dominance polygny and lottery polygyny) shall be observed in the next chapters. A special variance effective population size model will be applied to different mating systems which shows how the $N_e's$ change if not random mating but another mating system is considered.

Before starting with the general derivation of the variance effective population size, it is interesting to provide a table of actual effective population numbers such that one sees, how different the result of effective population size models can be.

species	$\frac{N_e}{N}$**-ratio**	**variables**	**population size used**
pacific oyster	10^{-6}	SR, VFS, FPS, o	A
rusty lizard	0.21	VFS, o	A
tiger	0.41	SR, VRS, n	B
snail	0.75	VFS, n	B
horse	0.8	SR, n	B
rainbow trout	0.9	SR, VFS, FPS, o	A

SR ... sex ratio

VFS ... variation in family sizes

FPS ... fluctuations in population over generations

o ... overlapping generations

n ... non-overlapping generations

A ... adults

B ... breeding adults

It can be seen that the N_e-N-ratio reaches from very small numbers (i.e. 10^-6) to numbers that are approximately 1 (i.e. $N_e \approx N$).

This overview about the different effective population sizes that results from deviating from the idealised population has shown basic ways how to work with N_e's. In the next chapter the attention shall lie on general models of the variance effective population size.

4 Variance Effective Population Size

The effective population size models that will be examined in the following chapters are derivations of the variance effective population size. Therefore, in the next chapter some important information about the variance N_e shall be explained.
Different assumptions lead to different effective population size models. Therefore, in the following chapter the variance effective population size will be examined under different expectations. It is just an overview about the different types of models that can be used to discover the variance effective population size. The basic assumptions in this chapter are

1. discrete generations
2. random mating
3. diploidy.

In the following chapters the assumption *discrete generations* will be extended to the assumption *overlapping generations*. This leads to very interesting results in the context of mating systems.

In general, the variance effective population size can be found by asking for the size of an ideal population with the same random gene frequency drift. The gene frequency drift effects the variance between groups in an increasing way [3], where the groups are the different allele groups that coexists in a population. The *between-group* variance of the allele frequency, p, in an idealised population at time t is

$$\sigma_t^2 = (1 - \frac{1}{2N})\sigma_{t-1}^2 + \frac{p(1-p)}{2N}.$$

The variance effective population size is defined by treating the variance of the change in allele frequency

$$\mathrm{Var}(\Delta p) = E((\Delta p)^2) - E(\Delta p)^2 = E((\Delta p)^2)$$

as equivalent to its value in an idealised population, where Δp just states the change in allele frequency. From earlier studies [2] we know that the expected value of the change in allele frequency in an idealised population is

$$\frac{p(1-p)}{2N_e},$$

and this leads to

$$N_e = \frac{p(1-p)}{2E((\Delta p)^2)}.$$

It is assumed that the number of parents and the number of gametes contributed by each parent are known but their genotypes are not. It is the aim to find the variance of

allele-frequency drift if one samples only from the parent population with a certain numbers of gametes but the genotypes are distributed randomly among the parents [3]. If males and females are separated, the notation $N_{m,t}$ and $N_{f,t}$ for the number of male respectively female individuals at time t is used. Let k_i be the number of progenies that individual i has, where i runs through the adults. The expected value of progenies per individual is $\bar{k_i}$ and we will indicate the variance as $\sigma^2_{k_i}$. In the following subsection the variance effective population size will be examined under the special case of an isogamous monoecious diploid population.

4.1 Isogamous Monoecious Diploids

If a population is isogamous it means, that the fusing gametes do not differ in size shape or function. This means, that one cannot distinguish between male and female gametes. Examples for isogamous populations are algae, fungi and protozoans. Further we assume monocious instead of diocious populations. While monoecious means that there are male and female sex organs in the same individual, dioecious means that one individual can only have either the male or the female sex organs. An example for a monoecious animal is the sea star. This means that we observe a population with individuals that have both sex organs, i.e. they can mate with themselves, and that cannot be distinguished in males and females. Additionally, a random amount of self fertilization is assumed.

For the derivation of the variance effective population size one picks one allele of interest which will be designated by A_1. The other possible alleles will be designated by A_x and the proportion of the allele combination A_1A_1 is P_{11}, the proportion of the allele combination A_1A_x is $2P_{1x}$. With this notation one can calculate the allele frequency p of allele A_1

$$p = P_{11} + P_{1x}.$$

In the next step, we define three random variables l, a and b, where l is binomial distributed and a and b are characteristic functions for determining if A_1 is homozygotes or heterozygotes [3]. Therefore, let

$$l_i \sim \text{Bin}(\frac{1}{2}, k_i)$$

$$a_i = \begin{cases} 1 & \text{if i is a } A_1A_1 \text{ individual} \\ 0 & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} 1 & \text{if i is a } A_1A_x \text{ individual} \\ 0 & \text{otherwise} \end{cases}$$

By denoting $n_{11} = N_{t-1} \cdot P_{11}$ as the number of A_1A_1 individuals and $n_{1x} = 2N_{t-1} \cdot P_{1x}$ as the number of A_1A_x individuals, this leads to

$$n_{11} = \sum_{i=1}^{N_{t-1}} a_i$$

and

$$n_{1x} = \sum_{i=1}^{N_{t-1}} b_i$$

and further $E(l_i) = \frac{k_i}{2}$, $E(l_i - \frac{k_i}{2})^2 = \frac{k_i^2}{2}$.

Now it is the aim to compute the expected values for the random variables a_i, b_i and a_ib_j. Assuming that $N_{t-1} = N$ it follows that there are n_{11} of A_1A_1 individuals, n_{1x} of the A_1A_x individuals and $N - (n_{11} + n_{1x})$ of the remaining combinations. This means that

$$E(a_i) = \frac{n_{11}}{N}, \;\; E(b_i) = \frac{n_{1x}}{N} \text{ and } E(a_ib_i) = 0.$$

Since the random variables are independent of each other, $E(a_ib_j)$ can also easily be computed if $i \neq j$

$$E(a_ia_j) = \frac{n_{11}(n_{11} - 1)}{N(N-1)}$$

$$E(b_ib_j) = \frac{n_{1x}(n_{1x} - 1)}{N(N-1)}$$

and

$$E(a_ib_j) = \frac{n_{11}n_{1x}}{N(N-1)}.$$

Additionally, some calculations for k_i can be done. Since k_i is the number of progenies that individual i has, one can rewrite the formulas for the expected value μ_{k_i} and the variance $\sigma^2_{k_i}$

$$N \cdot \mu_{k_i} = \sum_i k_i$$

$$N \cdot \sigma^2_{k_i} = \sum_i (k_i - \mu_{k_i})^2$$

and

$$N(N-1) \cdot \mathrm{Cov}(k_i, k_j) = \sum_{i \neq j} (k_i - \mu_{k_i})(k_j - \mu_{k_j}),$$

where $\mathrm{Cov}(k_i, k_j)$ is the covariance between the progeny numbers of different parents. The number of A_1 gametes contributed to the next generation is $\sum_i a_ik_i + \sum_j b_jl_j$ [3]. Therefore,

the new allele frequency is $\frac{\sum_i a_i k_i + \sum_j b_j l_j}{N\mu_{k_i}}$ and with this the change in allele frequency can be computed easily

$$\Delta p = \frac{\sum_i a_i k_i + \sum_j b_j l_j}{N\mu_{k_i}} - p.$$

With some rewritings this leads to

$$\Delta p = \frac{\sum_i a_i k_i + \sum_j b_j l_j}{N\mu_{k_i}} - \frac{2\sum_i a_i + \sum_j b_j}{2N}.$$

Now $N\mu_{k_i}\Delta p$ and $(N\mu_{k_i})^2 E(\Delta p)^2$ can be computed. For the exact calculations I refer to [3]. The results can be summarized as follows

$$N\mu_{k_i}\Delta p = \sum_i a_i(k_i - \mu_{k_i}) + \underbrace{\frac{1}{2}\sum_i b_i(k_i - \mu_{k_i}) + \sum_i b_i(l_i - \frac{k_i}{2})}_{=\sum_i b_i(l_i - \frac{k_i}{2} + \frac{k_i}{2} - \mu_{k_i}) = \sum_i b_i(l_i - \mu_{k_i})}$$

and

$$(N\mu_{k_i})^2 E(\Delta p)^2 = E(\sum_i a_i(k_i - \mu_{k_i}))^2 + \frac{1}{4}E(\sum_i b_i(k_i - \mu_{k_i}))^2 + E(\sum_i b_i(l_i - \frac{k_i}{2}))^2 +$$

$$+E(\sum_i a_i(k_i - \mu_{k_i}) \cdot \sum_i b_i(k_i - \mu_{k_i})) + 2E(\sum_i a_i(k_i - \mu_{k_i}) \cdot \sum_i b_i(l_i - \frac{k_i}{2})) + E(\sum_i b_i(k_i - \mu_{k_i}) \cdot \sum_i b_i(l_i - \frac{k_i}{2})).$$

Since b_i and l_i are two independent random variables, the expected values can be separated. Additionally, $E(l_i - \frac{k_i}{2}) = 0$ and therefore, the last two terms vanish and by squaring the other terms, we get

$$\underbrace{E(\sum_i a_i^2(k_i - \mu_{k_i})^2}_{=N\sigma^2_{k_i}E(a_i^2)} + \underbrace{\sum_{i\neq k} a_i a_j (k_i - \mu_{k_i})(k_j - \mu_{k_j})}_{=N(N-1)\operatorname{Cov}(k_i,k_j)E(a_i a_j)} + \underbrace{\frac{1}{4}E(\sum_i b_i^2(k_i - \mu_{k_i})^2 + \sum_{i\neq j} b_i b_j (k_i - \mu_{k_i})(k_j - \mu_{k_j})}_{\frac{1}{4}N\sigma^2_{k_i}E(b_i^2) + \frac{1}{4}N(N-1)\operatorname{Cov}(k_i,k_j)E(b_i,b_j)} +$$

$$+ \underbrace{E(\sum_i b_i^2(l_i - \frac{k_i}{2})^2}_{\sum_i E(l_i - \frac{k_i}{2})^2 E(b_i^2)} + \underbrace{\sum_{i\neq j} b_i b_j (l_i - \frac{k_i}{2})(l_j - \frac{k_j}{2})}_{\sum_{i\neq j} E((l_i - \frac{k_i}{2})(l_j - \frac{k_j}{2}))E(b_i b_j)} + \underbrace{E(\sum_i a_i b_i (k_i - \mu_{k_i})^2}_{N\sigma_{k_i}E(a_i b_i)} + \underbrace{\sum_{i\neq j} a_i b_j (k_i - \mu_{k_i})(k_j - \mu_{k_j}))}_{N(N-1)\operatorname{Cov}(k_i,k_j)E(a_i b_j)}.$$

The therein contained term

$$\sum_{i\neq j} E((l_i - \frac{k_i}{2})(l_j - \frac{k_j}{2})) \cdot E(b_i b_j)$$

is zero, since meioses in different individuals are independent, and the term

$$N\sigma_{k_i}^2 E(a_i b_i)$$

is zero too because the expected value $E(a_i b_i) = 0$ [3]. However, these terms vanish and the remaining terms can be rewritten to $(N\mu_{k_i})^2 E(\Delta p)^2 =$

$$= n_{11}\sigma_{k_i}^2 + n_{11}(n_{11}-1)\operatorname{Cov}(k_i,k_j) + \frac{n_{1x}\sigma_{k_i}^2}{4} + \frac{n_{1x}(n_{1x}-1)\operatorname{Cov}(k_i,k_j)}{4} + \frac{n_{1x}\mu_{k_i}}{4} + n_{11}n_{1x}\operatorname{Cov}(k_i,k_j).$$

Now we want to compute the covariance $\operatorname{Cov}(k_i,k_j)$ with a trick, such that the covariance can be inserted in the above formula

$$N\sigma_{k_i}^2 + N(N-1)\operatorname{Cov}(k_i,k_j) = \underbrace{\sum_i (k_i-\mu_{k_i})^2}_{=0,\ \text{since}\ \sum_i (k_i-\mu_{k_i})=0} + \underbrace{\sum_{i\neq j}(k_i-\mu_{k_i})(k_j-\mu_{k_j})}_{=0,\ \text{since}\ \sum_i (k_i-\mu_{k_i})=0} = 0.$$

This formula can now be rewritten

$$\operatorname{Cov}(k_i,k_j) = \frac{-\sigma_{k_i}^2}{N-1} = \frac{-\sigma_{k_i}^2}{N_{t-1}-1}.$$

We insert this covariance and get

$$(N_{t-1}\mu_{k_i})^2 \operatorname{Var}(\Delta p) = \frac{1}{N_{t-1}-1}\sigma_{k_i}^2[N_{t-1}(n_{11}+\frac{n_{1x}}{4}) - (n_{11}+\frac{n_{1x}}{2})^2] + \frac{n_{1x}\mu_{k_i}}{4}.$$

Now the two new parameters α and s_k^2 shall be introduced. While α is a measure of departure from Hardy-Weinberg ratio in the parent generation, the variance $s_{k_i}^2$ is a correction of $\sigma_{k_i}^2$. [3] We have

$$n_{11} = N_{t-1}[p^2 + p(1-p)\alpha]$$

$$n_{1x} = N_{t-1}[2p(1-p)(1-\alpha)]$$

and

$$s_{k_i}^2 = \sigma_{k_i}^2[\frac{N_{t-1}}{N_{t-1}-1}].$$

With these new parameters and dividing trough $(N_{t-1}\mu_{k_i})^2$, this leads to

$$\operatorname{Var}(\Delta p) = p(1-p)\cdot\frac{1}{2N_{t-1}\mu_{k_i}^2}[(1-\alpha)\mu_{k_i} + (1+\alpha)s_{k_i}^2].$$

From the definition of variance N_e,

$$\operatorname{Var}(\Delta p) = \frac{p(1-p)}{2N_e},$$

the new computed variance can be equaled to this variance. This leads to

$$\frac{1}{N_e} = \frac{1}{2N_t}[1 - \alpha + \frac{(1+\alpha)s_{k_i}^2}{\mu_{k_i}}].$$

Setting the departure from Hardy-Weinberg ratio and the corrected variance equal to zero, $\alpha = 0$ and $s_{k_i}^2 = 0$, leads to an effective population size of $2N_e = N_t$, which shows that in a population with no fluctuating population size, with a binomial progeny distribution and with the parents in Hardy-Weinberg proportions, half the variance in gene-frequency drift is due to variable progeny numbers, and half is due to segregation from heterozygotes [3]. But if in the population random mating is assumed and alleles are distributed randomly among the parent generation, then $\alpha = -\frac{1}{2N_{t-1}-1}$ and this leads to

$$N_e = \frac{4N_t - \mu_{k_i}}{2(1 + \frac{\sigma_{k_i}^2}{\mu_{k_i}})}.$$

Again a binomial distribution would lead to $\frac{\sigma_{k_i}^2}{\mu_{k_i}} = 1 - \frac{1}{N_{t-1}}$ and $N_e = N$ would be the result.

4.2 Separated Sexes

While in the first case it was sufficient to use the notation k_i for the progenies of individual i and N or N_{t-1} for the number of individuals in the parent generation, we now need to distinguish between k_{smi} and k_{sfi} (the number of sons and daughters of the i^{th} parent of sex s) and also between $N_{m,t-1}$ and $N_{f,t-1}$. For simplicity, we assume that $N_{s,t} = N_s$, where $s \in \{m, f\}$.

The idea is similar as before. One tries to find $\mathrm{Var}(\Delta p_{sm})$ and $\mathrm{Var}(\Delta p_{sf})$ to equal this variance with the idealised one and receive the effective population size. With similar calculations as before, one computes

$$\mathrm{Var}(\Delta p_{sf}) = p_s(1-p_s) \cdot \frac{(1-\alpha_s)\mu_{sf} + (1+\alpha_s)s_{sf}^2}{2N_s\mu_{sf}^2},$$

where μ_{sf} is the mean number of daughters of a parent of sex s. In general, the first subscript always indicates the parent's sex and the second one indicates the progeny's sex. Further, one can define $\mathrm{Cov}(k_{smi}, k_{sfi})$, which is the covariance between the number of sons and daughters of parent i with sex s. It is now possible to compute this covariances

$$N_s \,\mathrm{Cov}(k_{smi}, k_{sfi}) = \sum_i (k_{smi} - \mu_{sm})(k_{sfi} - \mu_{sf})$$

and

$$N_s(N_s-1)\operatorname{Cov}(k_{smi},k_{sfj})=\sum_{ij}(k_{smi}-\mu_{sm})(k_{sfj}-\mu_{sf}).$$

With similar steps as before, one can compute $N_s^2\mu_{sm}\mu_{sf}E(\Delta p_{sm}\Delta p_{sf})$, the variance $\operatorname{Var}(\Delta p_s)$ and finally $N_{e,s}$

$$\frac{1}{N_{e,s}}=\frac{1}{4N_s}\cdot[(1-\sigma_s)(\frac{1}{\mu_{sm}}+\frac{1}{\mu_{sf}})+(1+\alpha_s)\cdot(\frac{s_{sm}^2}{\mu_{sm}^2}+\frac{2\operatorname{Cov}(k_{sm},k_{sf})}{\mu_{sm}\mu_{sf}}+\frac{s_{sf}^2}{\mu_{sf}^2})].$$

For the detailed calculation I refer to [3]. As before, we assume that the parents have random mating. However, when the numbers of male and female parents are fixed, then $\alpha_s=-\frac{1}{2N_s-1}$ and this leads to

$$\frac{1}{N_{e,s}}=\frac{1}{2(2N_s-1)}\cdot[\frac{1}{\mu_{sm}}+\frac{1}{\mu_{sf}}+\frac{s_{sm}^2}{\mu_{sm}^2}+\frac{2\operatorname{Cov}(k_{sm},k_{sf})}{\mu_{sm}\mu_{sf}}+\frac{s_{sf}^2}{\mu_{sf}^2}].$$

One can see that some special assumptions lead to some easier models again.

For this section we only considered discrete populations. Some mathematicians already found some models for the variance effective population size for overlapping generations. Hill published two papers, one in 1972 [13] and one in 1979 [12], where overlapping generations are considered. Among other papers they will be the basis for the models of variance effective population sizes in the context of mating systems.

5 Variance Effective Population Size and different mating systems

Until now the variance effective population size has been discussed in a very general way. In this chapter the effective population size shall be observed in the context of mating systems. If one wants to examine a special population in a mathematical way, one needs to make assumptions such that it is possible drawing up formulas. In the following chapter the formulas shall be derived by considering the required assumptions.

Before one can work with the concept of mating systems and the variance effective population size one should get some general information of the biological sexual conflict in nature.

5.1 Conflict between sexes and mating systems

The following chapter will refer to the book by Davies, Krebs and Stuart [5]. Most of the biological information is from chapters 6 and 7.

According to Davies, Krebs and Stuart, it has long been believed that mating is a process in which females and males are in balanced symbiosis. Both sexes have the goal of passing on their genes and this is what happens when individuals get offspring. However, it is now known that the mating process is often subject to conflict between the sexes. For reproduction gametes are first formed, which eventually fuse to form zygotes. In most multicellular animals the female gametes can be clearly distinguished from the male ones. The female gametes are immobile and are called eggs; the male gametes are mobile, small and are called sperm. However, there are also species in which gametes are not sexually different. These species are called isogamous. The conflict between the sexes just mentioned underlies these differences in the gametes in anisogamous species. In anisogamous animals the female invests a lot of time and energy in supplying the egg with nutrients. In contrast, the male gametes - sperm - are produced quickly. They are able to move quickly in large numbers in order to find the female's egg. As a result the male has more opportunities to inseminate a female and produce offspring than the female, which needs weeks to months to produce an egg or to carry the offspring to term [5]. It is therefore an advantage for males to mate with different females as often and as quick as possible in order to father many offspring and thereby pass on their own genes. On the other hand, females have to think carefully about which males they want to breed with, as they invest more energy and time in this one brood. Of course, this is a simplification. There are also cases of (exclusive) male care or population where both sexes take care of the progenies. This juxtaposition of sperms and eggs shall just show in which way the conflict between the sexes occurs in nature.

The parents' entire investment in their offspring is called "parental benefit". While fe-

males invest their "parental benefit" in the form of breeding and upbringing, the males use their "parental benefit" to courtship and copulate in an appealing way for females. For this reason, it also makes sense that greater sexual selection leads to a more unequal division of gender roles. If the gestation period of the females of a certain mammal species is particularly long, as for example with the seal elephants, this leads to a very pronounced polygyny: the harem polygyny.

For example, after an 11-month pregnancy, female seal elephants come to breeding sites to give birth to a single pup. During this time one male can fertilize many more females. For this reason, the elephant seals live in harems. The maximum number of offspring a male elephant seal can produce during his life time is around 200, while the maximum number of a female elephant seal is around 15. An interesting adaption to sperm competition is the method of the damselfy *Calopteryx Maculata*, a dragonfly species, where the males use the most bizarre methods to pass on their own genes. Since the female stores the male's sperm in a kind of bag called the abdomen, the males have developed their own scoop-like structures at the end of their penis with which they can scrape out the sperm of other males during the mating act in order to place their own sperm there.

In the case of the *Xylocoris Maculipennis* bug, a bug species, the male pierces the female's body wall during copulation and injects it with his sperm. The sperm swims around in the female until they hit an egg. In addition, the males often inject other males with their sperm and when the male victim mates with a female, the sperm of the other male is also transmitted. [5]

These examples show that it is the aim of every individual to pass on as many genes as possible. There are different species of animals with different methods and mating systems. Each sex tries to get the best out of him or her. In some cases, the male invests more (for example, when the female scorpion fly receives a nourishing gift from the male before mating) and in other cases, the female invests a lot without any success in passing on her genes (when for example the male lion kills other males' cubs). The mating conflict between the sexes has many facets and can be viewed on different levels. At the end, this conflict is mostly influenced by ecological conditions, such as distribution of resources etc. The ecological, species-specific and situation-related influences naturally also affect the mating system. Often the breeding behavior and the mating system are related. It is not always clear which situation causes the other. For example, one can ask whether in monogamous starlings the male helps with brood care because he lives in a monogamous relationship, or whether the birds live in a monogamous relationship because the male participates in brood care. In order to examine the main factors that influence brood care, the three vertebrate classes, birds, fish and mammals, shall be observed.

5.1.1. Birds

In birds reproductive success is strongly related to the amount of food the young get from their parents. Therefore, it makes sense for pairs of birds to work together instead of leaving the brood care to just one sex. In the case of the kittiwake couples who have bred together have a higher reproductive success than newly wed couples. For this reason, in this species it is a fitness advantage to stay faithful to their partner. There are also bird species - such as the weaver bird - that live sometimes in monogamous populations and sometimes in polygamous populations. For example, if there is enough food source - fruit and grain eaters - one parent can provide the offspring with sufficient food. Then it is usually the male who leaves the breeding site [5]. There are two reasons why the males usually leave the female alone with the brood care and not the other way round. Firstly, after internal insemination, the female still has the offspring in her body and cannot escape like the male. Secondly, due to the differences in gametes, it makes more sense for the male to inseminate several females and leave them alone to raise the brood. While birds are often biparental (or sometimes monogamous), mammals are often polygamous.

5.1.2. Mammals

Since the offspring grows inside the females for a longer period, they, in contrast to the males, contribute disproportionally more to the development of the young. At best, the males can protect the female and provide it with food. Since it is advantageous for the male to fertilize as many females as possible and at the same time protect them, it is convenient if all of the females are at one place. Therefore, pack formations often occur. After birth, only the female produces the nourishing milk for the young. For these reasons, polygenic mating systems often occur in mammals and the females mainly take care of the brood. There are also mammal species where both parents are responsible for the brood care.

5.1.3. Fish

In a study by Breder and Rosen [5], several species of fish were examined and the following result was found: while with internal insemination of the eggs only two males, 10 females and 0 times both sexes take care of brood care, with external insemination 28 males and only 6 the females take care of the brood [5]. If the eggs are inseminated externally, the females can swim away while the males are still inseminating the eggs. So the male has to take care of the brood, since the female already ran off. Further reasons for paternal brood care are that the males may have taken a lot of time to impress the females with beautifully built nests and have therefore already invested more in the brood. Additionally, most of the time no other male fish is present at the time of insemination, so the males can be sure of their paternity. This study by Breder and Rosen was conducted with some fish

species. This is not a general result for all fish.

In this chapter an overview over the mating behaviour of the three different vertebrate species were given. Now the effective population size models, that are used to analyse N_e in the context of mating systems, will be derived in the next chapter.

5.2 Derivation of the formulas

Models of the effective population size in the context of mating systems has been already examined by different researchers, as for example the biologist Leonard Nunney. In 1993 Nunney published a paper [16] that shows how to compute the effective population size with respect to different mating systems.

In this chapter firstly these formulas shall be derived. Afterwards, the formulas will be discussed and different results by varying the parameters will be shown. The outcome and interpretations will also take place in this chapter. Before introducing the formulas, some proofs respectively derivations shall help to understand where the formulas come from. One way how an effective population size model can be derived is to observe the variance of the productive success in a lifetime of an individual, σ_k^2. In other words, the family sizes of the individuals is in the focus of interest. This effective population size can be written as

$$N_e = \frac{N}{2 + \sigma_k^2}.$$

William G. Hill [13], James F. Crow and Motoo Kimura [4] extended Wright's formulas by additionally assuming overlapping generations. The general formula for the variance effective population size by considering overlapping generations was derived by Hill in 1972. His statements are summarized in *Theorem 1.* The theorem describes a more general result for the variance effective number. The variance of family size shall be specified in this model. The assumption are constant population size and the existence of an age distribution. These and some other small assumptions lead to the fact that the final result is just an extension of the basic model. Further, the author considers an annual effective population size alternative to the current effective population size N_e.

> "For comparison of alternative methods of maintaining controls we shall find it useful not only to consider the effective population size N_e, which is the size of an idealised population leading to the same variance of gene frequency change per generation, but also an annual effective size N_a, which is the size of an idealised population with a generation interval of one year leading to the same variance of gene frequency change per year." [13]

One can easily see that using L for the generation interval and assuming a constant increment in variance, then $N_a = L \cdot N_e$.

For the derivation of the model, we first consider by a general stochastic process. After doing this we put the model in a genetic context and invoking variable family size in a haploid model. Finally we consider diploid models for the last result.

Theorem 1. *(N_e for overlapping generation-populations) Let L be the generation interval, i.e. the average age of parents, N_f and N_m the number of females respectively males, σ^2_{ij} the variance of the pathway for gametes if a parent of sex i has a progeny of sex j, with $i, j \in \{m, f\}$ and $\mathrm{Cov}(mm, mf)$ and $\mathrm{Cov}(fm, ff)$ are covariances between the ages of a male/female parent, when he/she gets his/her male/female progenies.*
Then the effective population size is

$$\frac{1}{N_e} = \frac{1}{16N_mL}[2 + \sigma^2_{mm} + 2\frac{N_m}{N_f}\,\mathrm{Cov}(mm, mf) + (\frac{N_m}{N_f})^2\sigma^2_{mf}]+$$

$$\frac{1}{16N_fL}[2 + \sigma^2_{ff} + 2\frac{N_f}{N_m}\,\mathrm{Cov}(fm, ff) + (\frac{N_f}{N_m})^2\sigma^2_{fm}].$$

One could also rewrite the formula as a sum $\frac{1}{N_e} = \sum_i \frac{1}{16N_iL}\,[...]$, where $i \in \{f, m\}$.

Proof:

It is easier to understand the derivation of the formula if we first consider a haploid model, then a monoecious model and at last the dioecious diploid model. Therefore, the proof shall start with a haploid model.

5.2.1. Haploid model:

We assume that there are N individuals born in one time period that are summarized to one cohort. The term *cohort* refers to the time structured group of the individuals. We assume that the cohort of N newborns had frequencies x_1, ..., x_N of some neutral gene at some time in the past, where x_i can only have the values 0 or 1, $i \in \{1, ..., N\}$ since this is the haploid model. In the diploid case x_i could be 0, $\frac{1}{2}$ or 1. The frequencies x_i are random variables which are independent and identically distributed (iid). For this model one can conceive a box with two alleles. For every newborn one allele is pulled out of the box. Therefore, the random variable underlies the Bernoulli-distribution. The expected value of the random variable and the mean frequency of this gene in one cohort is

$$q = \sum_{i=1}^{N} \frac{x_i}{N}.$$

N is the number of the newborns. Since $E(x_i) = q$ and the random variable is Bernoulli-distributed, it follows that $\mathrm{Var}(x_i) = q(1 - q)$.
The number of progenies from individual i in his or her whole life-time is k_i, which can be separated in k_{ij} progenies at age j. This means

$$k_i = \sum_{j=1}^{s} k_{ij}$$

where s is the age after which all reproduction ceases.

For any newborn gene or - in the haploid case - for any newborn individual, the probability that it stays in the population is $\frac{1}{LN}$. [7] This is the contribution by the cohort - which is $\frac{1}{L}$ - divided by the number of genes respectively individuals in one cohort N. Now we can compute the mean contribution by the whole cohort which we define as Q. The mean contribution by the cohort is the mean gene frequency times the probability that the gene remains

$$Q = \frac{1}{LN} \sum_{i=1}^{N} k_i x_i.$$

The mean contribution Q will change because of differential reproduction among the individuals and create genetic drift. [12] As already mentioned, the mean of the random variable x_i is q and since the population size remains constant, the mean family size is $\mu_{k_i} = 1$. Thus, the expected value of Q is

$$E(Q) = \frac{1}{LN} \cdot N \cdot E(k_i) \cdot E(x_i) = \frac{N \cdot 1 \cdot q}{LN} = \frac{q}{L}.$$

Now it is easy to compute the deviation, i.e. the genetic drift, of Q from the expected value

$$Q - \frac{q}{L} = \frac{1}{LN} \sum_{i=1}^{N} (k_i x_i - q) = \frac{1}{LN} \sum_{i=1}^{N} (k_i - \frac{q}{x_i}) x_i.$$

This is the random variable that describes the deviation from the mean by sampling from the binomial distribution. Thus, we multiply the probability that a gene remains with the average number of genes in the population (since k_i is the number of newborns that carry the gene with frequency x_i). Another quantity we need is the expected family size μ_{k_i}. The variance of the family size can be defined as

$$\sigma^2_{k_i} = \sum_{i=1} \frac{(k_i - \mu_{k_i})^2}{N},$$

but in the Bernoulli-distributed case the variance is $\mathrm{Var}(x_i) = q(1-q)$, as already described above. Therefore, the variance of $Q - \frac{q}{L}$ can be rewritten

$$\mathrm{Var}(Q - \frac{q}{L}) = q(1-q) \frac{\sigma^2_{k_i}}{NL^2}.$$

To now calculate the effective population size, one needs to define the effective population size through the same increment in drift variance.

> "To compute the effective population size, consider an idealized population of N haploid individuals with no differential viability or fertility and discrete

generations, where the variance in family size equals 1. [...] The *annual effective population size*, N_a, the effective size of a population having discrete generations of length one time period [...] and the same increment in drift variance is thus

$$N_a = \frac{q(1-q)}{\mathrm{Var}(Q - \frac{q}{L})} = \frac{NL^2}{\sigma^2_{k_i}}.\text{" [12]}$$

Since we want to calculate the effective population size N_e and not only the annual effective population size, one needs to divide through the generation length L. This leads to

$$N_e = \frac{N_a}{L} = \frac{NL}{\sigma^2_{k_i}}.$$

5.2.2. Monoecious diploids:
If we now assume a diploid population, we need to double the number of genes. Therefore, we let $N \to 2N$. The probability of the survival of a gene is now $\frac{1}{2NL}$. We see that the haploid case is very similar to the current diploid case. To compute Q one needs to consider one big difference between these two cases. We need to introduce a random variable δ_{ij} that shows the difference in frequency between the jth sampled gene and its parental value x_i. If the parent is homozygote, $\delta_{ij} = 0$ (no difference between the parent and the progeny) and if the parent is heterozygote it is $\pm\frac{1}{2}$. Therefore, this value δ_{ij} needs to be introduced in the formula of Q

$$Q = \frac{1}{2LN}[\sum_{i=1}^{N} k_i x_i + \sum_{j=1}^{k_i} \delta_{ij}].$$

Since Q is the mean contribution of the whole cohort, i.e. of one age-group, one needs to add the deviation from the homozygote case to the already known Q. If there is no difference in frequency between the sampled gene and the parental value, then $\delta_{ij} = 0$ and Q is the same as in 5.2.1. If there is a difference, the mean contribution changes as much as the progeny-gene deviates from the parent-gene. Therefore, the sum $\sum \delta_{ij}$ needs to be added. Noting that in an idealized population the drift variance per generation is $\mathrm{Var}(x_i) = \frac{q(1-q)}{2}$, we can again derive the effective population size but we additionally need the variance of δ_{ij} now. Since $E(\delta_{ij}) = \frac{1}{4}\cdot 0 + \frac{1}{4}\cdot 0 + \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot(-\frac{1}{2}) = 0$ and $\mathrm{Var}(\delta_{ij}) = E(\delta^2_{ij}) - E(\delta_{ij})^2$, we just need to derive $E(\delta^2_{ij})$. However, $E(\delta^2_{ij})$ = frequency of heterozygotes $\cdot$ variance given heterozygotes and we see that

$$\mathrm{Var}(\delta_{ij}) = 2q(1-q)\cdot\frac{1}{4}.$$

If we insert this variances in the formula above, we get

$$\mathrm{Var}(Q - \frac{q}{L}) = \frac{q(1-q)(\sigma^2_k + 2)}{8NL^2}$$

and therefore,

$$N_e = \frac{N_a}{L} = \frac{4NL}{\sigma^2_{k_i} + 2}.$$

However, we now derived the Wright-Fisher model by considering the generation length L. If $L = 1$, we exactly get the effective population size by Wright that does not consider overlapping generations.

5.2.3. Two sexes:
The last step considers two different sexes. We will now derive this last formula. The following overview is fundamental for further understanding.

Parent-Progeny	**mean age**	**family size**	**heterozy. sampling**	**variance**
male-male	L_{mm}	k_{mmi}	δ_{mmij}	σ^2_{mmi}
male-female	L_{mf}	k_{mfi}	δ_{mfij}	σ^2_{mfi}
female-male	L_{fm}	k_{fmi}	δ_{fmij}	σ^2_{fmi}
female-female	L_{ff}	k_{ffi}	δ_{ffij}	σ^2_{ffi}

The covariance Cov(mm, mf) comes from the first two variances and the covariance Cov(fm, ff) comes from the last two variances. However, the correlation consists between the ages of a male/female parent, when he/she gets his/her male/female progenies. It is possible but unlikely that a parent gets his male progenies at the age of 1 and then the female progenies at the age of 100. Thus, the covariance measures the correlation between the age of producing male and female progenies. Thus, the covariances Cov(mm, mf) and Cov(fm, ff) come from variation in the age of a parent when his/her offspring of either sex is born.

$$Q = \frac{1}{4N_m L}[\sum_{i=1}^{N_m} k_{mmi}x_{mi} + \sum_{j=1}^{k_{mmi}} \delta_{mmij} + \sum_{i=1}^{N_f} k_{fmi}x_{fi} + \sum_{j=1}^{k_{fmi}} \delta_{fmij}]+$$

$$\frac{1}{4N_f L}[\sum_{i=1}^{N_m} k_{mfi}x_{mi} + \sum_{j=1}^{k_{mfi}} \delta_{mfij} + \sum_{i=1}^{N_f} k_{ffi}x_{fi} + \sum_{j=1}^{k_{ffi}} \delta_{ffij}].$$

The same definition of the effective population size as before is used to derive N_e. Some short transformations lead to the effective population size that we wanted to derive

$$\frac{1}{N_e} = \frac{1}{16N_m L}[2 + \sigma^2_{mm} + 2\frac{N_m}{N_f}Cov(mm, mf) + (\frac{N_m}{N_f})^2\sigma^2_{mf}]+$$

$$\frac{1}{16N_f L}[2 + \sigma^2_{ff} + 2\frac{N_f}{N_m}Cov(fm, ff) + (\frac{N_f}{N_m})^2\sigma^2_{fm}]. \quad (1)$$

One can easily see that with a 1:1 sex ratio and variances of 2 and covariances of 1 the effective population size is NL again.

This general formula derived by Hill can now be used for our aims. Nunney worked with this model by considering some assumptions such as the possibility to calculate the variances and covariances. Therefore, another theorem can be formulated. The following derivation refers to Nunney's paper from 1991. [15]

The models are created for populations with overlapping generations and the mating systems are considered in combination with age-independent fecundity. Thus, the formula considers a population with overlapping generations, a constant size of N adults and a sex-ratio of r males to $(1-r)$ females. Further, we assume a stable age distribution. The newborns are already able to produce own juveniles in the next breeding season, but only a limited number of juveniles are recruited to the population (since the population size is constant). Another assumption is that the sex ratio of juveniles is $1:1$ and the adult fecundity and mating success does not depend on the age. The fecundity is a random variable that acts independent each season. This means that the reproductive success of an individual in one season has no influence on how well it will be the next. [15] With this information the theorem can now be stated.

Theorem 2. *Lets assume a population with the following properties:*

- *adult numbers are constant*
- *fecundity is age-independent*
- *survival is age-independent*
- *sex ratio of juveniles is 1:1*
- *maturity is achieved at age 1*

Then under these assumptions the effective population size is

$$N_e = \frac{4r(1-r)NL}{L_m(1-r) + L_f r - \frac{2r}{\alpha_f} + s_m^2(1-r) + s_f^2 r + L_m \cdot s_{L_m}^2(1-r) + L_f \cdot s_{L_f}^2 r}, \tag{2}$$

where r is the male sex ratio, α_f is the female average fecundity per breeding season, L_i is the mean generation length of sex i, s_i^2 is the standardized variance of the fecundity of sex i and $s_{L_i}^2$ is the standardized variance of the mean generation length of sex i.

Proof:

The model shall be derived for a population with overlapping generations. But the census size of the population stays the same, i.e. $N(t) = N(t+1)$ if $N(t)$ is the population size at time t. The proportion of the males is r and the proportion of the females is $(1-r)$, i.e. with the old notation this means

$$N \cdot r = N_f \text{ and } N \cdot (1-r) = N_f.$$

We need to work with some parameters, if we want to consider the assumption above. First of all, the generation time (in breeding seasons) of the males is L_m and the females' one is L_f and one can also calculate the mean generation time $L = \frac{L_m+L_f}{2}$. Therefore, two variances can be defined σ^2_{Lm} and σ^2_{Lf}.

Another situation that needs to be considered is the average fecundity of the sexes. Therefore, we choose α_m for the mean male fecundity and α_f for the mean female fecundity. This two parameters come together in the following relation

$$\alpha_f \cdot (1-r) = \alpha_m \cdot r.$$

Also the fecundity can vary and therefore we define two parameters for the variance in fecundity σ^2_f and σ^2_m. If the variances are divided through the mean squared, the standardized variances are received, which are defined by s^2_{Lm}, s^2_{Lf} respectively s^2_m, s^2_f, where s stands for "standardized". It is also important to know that in this model the fecundity is independent from one season to the next. This means that a high reproductive success of an individual has no impact to the success in the next season. Now there is a need to derive the variances and the covariances such that they can be inserted in (1). After doing so, we will receive the formulas we are going to work with.

Since the population size and also the sex relation r stay the same from one breeding season to the next, the number of recruited offspring (of one sex) has to be the same as the number of adults (of the same sex) dying. The number of adults dying is $\frac{N_i}{L_i}$, where $i \in \{m, f\}$ since after L_i seasons an adult of sex i dies and therefore we need to divide through this generation time if we want to know the individuals dying in one season. If α_i ($i \in \{m, f\}$) is the proportion of breeding individuals, $\frac{\alpha_i N_i}{2}$ is the offspring of each sex produced per season. We will denote this by $B = \frac{\alpha_i N_i}{2}$. We already denoted k_{ij} as the number of offspring of sex j from a parent of sex i. If we want to include the age t of the parent at the breeding time, we need to add t so the number of offspring is k_{ijt}. The quantity k_{ijt} can be viewed as a random variable with mean

$$E(k_{ijt}) = \frac{\alpha_i}{2}$$

and the variance that is assumed to just depend on i is

$$\sigma^2_{k_{ijt}} = \frac{\sigma_i^2}{4}.$$

Since not all of the offspring is recruited we need to distinguish between x_{ijt} (the offspring that is recruited to the adult population) and k_{ijt} (all of the offspring of sex j from a parent of sex i at time t). All in all, there is $\frac{N_i}{L_i}$ offspring of sex i recruited in the adult population since that is the number of dying adults. Therefore, the probability that one of the k_{ijt} offspring is to be chosen, is $\frac{k_{ijt}}{B}$. [15] The expected value of x_{ijt} under the knowledge of k_{ijt} is the product of these two terms

$$E(x_{ijt}|k_{ijt}) = \frac{N_j}{L_j} \cdot \frac{k_{ijt}}{B}.$$

Without knowing that the chosen individual is one of the k_{ijt} the expected value is

$$E(x_{ijt}) = \frac{N_j}{L_j \cdot N_i}.$$

With the formula for hypergeometrical sampling one can compute $E(x^2_{ijt}|k_{ijt})$ by using the fact that $E(X^2) = E(X)^2 + \mathrm{Var}(X)$ for any random variable X. The general formula for the hypergeometrical expected value and variance are

$$E(X) = n \cdot \frac{M}{N}$$

and

$$\mathrm{Var}(X) = n \cdot \frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1},$$

where N is the whole number of elements, $M < N$ is the number of elements with a certain property and n is the sample. By rewriting the parameters to $n = \frac{N_j}{L_j}$, $M = k_{ijt}$ and $N = B$, this leads to

$$E(x^2_{ijt}|k_{ijt}) = \underbrace{(\frac{N_j k_{ijt}}{L_j B})^2}_{E(x_{ijt}|k_{ijt})^2} + \underbrace{\frac{N_j k_{ijt}}{L_j B} \cdot (1 - \frac{k_{ijt}}{B}) \cdot \frac{B - \frac{N_j}{L_j}}{B-1}}_{\mathrm{Var}(x_{ijt}|k_{ijt})}.$$

One can simplify this to

$$E(x^2_{ijt}|k_{ijt}) = [\frac{N_j k_{ijt}}{L_j B} + \frac{(B - \frac{N_j}{L_j})(1 - \frac{k_{ijt}}{B})}{B-1}]\frac{N_j k_{ijt}}{L_j B}.$$

To make this formula easier, we assume that the total number of offspring is very high $B >> 1$, such that terms of the form $\frac{1}{B}$ can be ignored and we average over all k_{ijt}. One

can now rewrite the expected value to

$$E(x_{ijt}^2) = \frac{\frac{N_j^2(\alpha_j L_j - 2)}{\alpha_j} + (1 + s_i^2)(N_j - L_j)N_j}{(N_i L_j)^2}$$

by using the approximation to the hypergeometrical distribution through the binomial distribution. [15]
With the given parameters one can also compute the number of offspring of one individual that are recruited to the adults if one knows how many seasons this individual lives/breeds. Let us assume that an individual of sex i that lives for t_i seasons which is a random variable again: The random variable has a distribution of mean L_i and variance s_{Li}^2. Then X_{ij} is the number of the described offspring and this number of offspring in dependence of the individual live time is

$$X_{ij}|t_i = \sum_{t=1}^{t_i} x_{ijt}.$$

Since the expected value and the variance of x_{ijt} are known, one can easily compute these two values for X_{ij}

$$E(X_{ij}|t_i) = t_i \cdot \frac{N_j}{L_j N_i}$$

and

$$\mathrm{Var}(X_{ij}|t_i) = t_i \cdot \mathrm{Var}(x_{ijt}).$$

Using, again, the fact that the variance of a random variable X can be computed by

$$\mathrm{Var}(X) = E(X^2) - E(X)^2$$

leads to

$$E(X_{ij}^2|t_i) = \frac{t_i[\frac{N_j^2(\alpha_j L_j - 2)}{\alpha_j} + (1 + s_i^2)(N_j - L_j)N_j - N_j^2 + t_i N_j^2]}{(L_j N_i)^2}.$$

If we now average over t_i, we get

$$E(X_{ij}) = \frac{L_i N_j}{L_j N_i}$$

and

$$\mathrm{Var}(X_{ij}) = \frac{[\frac{N_j(\alpha_j L_j - 2)}{\alpha_j} - L_j + s_i^2(N_j - L_j) + s_{Li}^2 L_i N_j]L_i N_j}{(L_j N_i)^2}.$$

For detailed calculation I refer to [15]. The covariance for parents of sex i that live for t_i seasons is

$$\mathrm{Cov}(X_{if}, X_{im}|s_i) = \frac{t_i s_i^2 N_f N_m}{L_f L_m N_i^2}$$

and by averaging again over t_i we get

$$\mathrm{Cov}(X_{if}, X_{im}) = \frac{(L_i s_i^2 + \sigma_{Li}^2)^2 N_f N_m}{L_f L_m N_i^2}.$$

The formulas of the covariances are derived in detail in the original paper. [15] This can now be inserted in the model that was derived earlier (1) and this leads to

$$N_e = \frac{4r(1-r)NL}{L_m(1-r) + L_f r - \frac{2r}{\alpha_f} + s_m^2(1-r) + s_f^2 r + L_m s_{Lm}^2(1-r) + L_f s_{Lf}^2 r}.$$

5.3 Application to specific mating systems

Now we know where the models of the effective population size come from. In the following, we will observe the formula in the context of different mating systems. In the information box, one can find an overview of the variables that will be used in the following formulas.

h ... number of females in one harem
r ... proportion of males in the population
α_m ... proportion of males that are dominant
α_f ... proportion of females that breed
n ... average number of mates per female
v_f ... probability that females survive to the next breeding season
v_m ... probability that males survive to the next breeding season
s_m^2 ... standardized variance of male fecundity
s_f^2 ... standardized variance of female fecundity

The formula we received from the derivations by Hill and Nunney is

$$N_e = \frac{4r(1-r)NL}{L_m(1-r) + L_f r - \frac{2r}{\alpha_f} + s_m^2(1-r) + s_f^2 r + L_m s_{Lm}^2(1-r) + L_f s_{Lf}^2 r}. \tag{3}$$

This formula can be rewritten using the new parameters.

We assume that there is only a proportion of α_f females that breed. This leads to a binomial standardized variance of

$$\frac{1-\alpha_f}{\alpha_f}.$$

If the breeding females are approximately Poisson distributed, the variance would be $\frac{1}{\alpha_f}$. Therefore, by setting $s_f^2 = \frac{1-\alpha_f}{\alpha_f}$, the term $\frac{2r}{\alpha_f}$ cancels out and can be ignored. The term s_m^2 is the standardized variance of the male fecundity. Further, we assume that the survival rate is constant, the juveniles mature at age 1 and the mean life span L_m and L_m are the same as the generation length L and there are $L_i = \frac{1}{1-v_i}$ breeding seasons, with $i \in \{m, f\}$. The standardized variance of the life span is $s_{Li}^2 = v_i$. [16] By inserting this in the model (3), this leads to

$$N_e = \frac{4r(1-r)N}{\underbrace{\frac{L_m}{L}}_{1} \cdot [(1-r) + \underbrace{s_{Lm}^2}_{v_m}(1-r) + \underbrace{\frac{1}{L_m}}_{(1-v_m)} s_m^2(1-r)] + \underbrace{\frac{L_f}{L}}_{1} \cdot [r + \underbrace{s_{Lf}^2}_{v_f} r + \underbrace{\frac{1}{L_f}}_{(1-v_f)} \cdot r \cdot \underbrace{s_f^2}_{\frac{1-\alpha_f}{\alpha_f}}]}$$

and therefore,

$$N_e = \frac{4r(1-r)N}{1 + v_m(1-r) + v_f r + (1-r)(1-v_m)s_m^2 + r(1-v_f)\frac{1-\alpha_f}{\alpha_f}}. \tag{4}$$

To make it easier to work with a big formula like this, we will split the model in two different cases. In the first case the differences between males and females refer to differences in the survival probability and in the second case the differences refer to recruitment differences. We will later discuss what this difference means in a biological way and how these formulas can be applied to real populations. Sometimes one of the two formulas will fit better than the other one and sometimes even none of the formulas can be applied.

In the first case where only the different survival probability and not the sex ratios play a role, i.e. $r = \frac{1}{2}$, the formula becomes

$$N_e = \frac{N}{1 + \frac{1}{2}[v_m + v_f + (1-v_m)s_m^2 + (1-v_f)\frac{1-\alpha_f}{\alpha_f}]}. \tag{5}$$

In the second case the differences of females and males refer to recruitment differences. "This means that the number of adults of each sex is regulated separately rather than survival determines the adult sex ratio." [15] It depends how many males respectively females are recruited to the adult population and this causes different sex ratios. The different survival probabilities do not play a role anymore and $v = \frac{v_m + v_f}{2}$. Therefore, the formula is in this case

$$N_e = \frac{4r(1-r)N}{1 + v + [(1-r)s_m^2 + r \cdot \frac{1-\alpha_f}{\alpha_f}](1-v)}. \tag{6}$$

The first of the two formulas can be interpreted in a biological way. The denominator can be explained by looking at the probability that males and females are surviving or males and females are not surviving. In order to contribute to N_e they have to mate and breed in this breeding season. For this case, one has to multiply the probabilities with the standardized variance in male/female fecundity, because then it matters, how the matings distribute in the current breeding season. If the adults certainly survive to the next breeding season, it follows $v_m = v_f = 1$ and one can see, that $N_e = \frac{N}{2}$. If $v_m = v_f = 0$, the adults never survive to the next breeding season and it follows

$$N_e = \frac{N}{1 + \frac{1}{2}(s_m^2 + \frac{1-\alpha_f}{\alpha_f})}.$$

The bigger the variance, the smaller the effective population size, since a high variance means that some of the males mate very often and others do not. So the genetic diversity will shrink. But on the other hand the fraction $\frac{1-\alpha_f}{\alpha_f}$ has to be small to receive a big

N_e. The fraction is high if we have a high numerator and a low denominator, i.e. the fraction is near one if there are a lot of non-breeding females. This leads to the biological interpretation that the less females breed (for example if more females are infertile), the smaller is the effective population size. If we additionally assume that $s_m^2 = 0$ (no variance in male fecundity) and $\alpha_f = 1$ (every female breeds), it follows that $N_e = N$ because we have a completely balanced mating system.

In the second formula, where males and females are distinguished with respect to differences in the recruitment, the formula is

$$N_e = \frac{4r(1-r)N}{1 + v + [(1-r)s_m^2 + r \cdot \frac{1-\alpha_f}{\alpha_f}](1-v)}$$

as already explained. This formula can be interpreted in a similar way as before.

The special case of $v = 1$, every adult survives the next breeding season, leads to $N_e = 2r(1-r)N$ and if there is additionally a 1:1 sex ratio it follows $N_e = \frac{N}{2}$. If no adults survive to the next breeding generation, $v = 0$ and we have a 1:1 sex-ratio, it leads to $N_e = N$, because the mating system is "balanced" again. Before applying the formulas (4), (5) and (6) to the different mating systems, one should have a closer look at the difference between the survival and the recruitment based case. In the survival based case, the male sex-ratio was set to be $r = \frac{1}{2}$. This means that the number of males and females is the same in this model. The difference only refers to the survival probability. If the two probabilities are very different, i.e. $v_m << v_f$ or the other way round, in every new generation there are new male individuals but the females stay generally the same. Of course, this has an effect on the genetic pool because the same genes of the female can be inherited very often. This could be important in populations where a special disease can only be inherited by one sex. If one applies this model to an actual population, one should be careful since this model only makes sense if the survival probabilities are actually very different and if the sex-ratio is approximately one half. If one sex is more represented in this model, this could lead to the wrong result since the sex ratio is not taken into account in this model.

In the following chapters the models will be used to analyse the effective population size under different mating systems. Since we have three different models to observe (the whole model (4) and the recruitment based respectively the sex based model (5) and (6)), not all of the models can be discussed in detail. Therefore, the reduced models shall just be discussed to introduce the different mating systems. The whole model (4) will be represented with 3D plots and contour plots such that one can see the differences between models applied to various mating systems.

5.3.1 Harem Polygyny

In this section it is the aim to observe the effective population size if the population has the mating system of a harem population. In a harem population a non-random distribution of reproductive success occurs since males compete with each other and the strongest has the possibility to reproduce. In other words, a sexual selection occurs.

The simplest model would occur if only one male wins against every other competitor and creates one offspring with every female in his harem. This means that in the next generation there are only N_f progenies and the population size would shrink. That is a very simplified example of a harem population which normally does not occur like this in nature. In real populations the male may survive but often there happens a turnover in the position of the lead male. It is important to find out which standardized variance in male fecundity, s_m^2, is the right one for the harem population situation. Since the standardized variance is $\frac{\text{variance}}{\text{mean}^2}$ one can compute s_m^2 by dividing $rN\alpha_m(1-\alpha_m)$ through the mean^2, which is $rN\alpha_m^2$. It follows

$$s_m^2 = \frac{1-\alpha_m}{\alpha_m}.$$

The concept of a harem population tells us that

$$\alpha_f \cdot (1-r) = h\alpha_m r$$

must hold. It is simpler to assume that half of the individuals can breed, i.e. $\alpha_f = 0.5$. Inserting $\alpha_m = \frac{\alpha_f(1-r)}{hr}$ in the standardized variance s_m^2 this leads to

$$s_m^2 = \begin{cases} \frac{hr}{1-r} - 1 & r \geq \frac{1}{h+1} \\ 0 & r < \frac{1}{h+1}. \end{cases} \tag{7}$$

The first case means that there are more males in proportion to the females, i.e. to the harem size. The second case means that there is an excess of females and the males don't have to worry about owning a harem. The standardized variance is zero. It is more realistic to assume that the case $r \geq \frac{1}{h+1}$ appears, because more often the males have to compete for owning a harem. Inserting this standardized variance in the formulas (5) and (6), leads to

$$N_e = \frac{2N}{2(v_m + v_f) + (1-v_f)(h+1)}$$

in the survival based situation and to

$$N_e = \frac{4r(1-r)N}{2v + r(1-v)(h+1)}$$

in the recruitment based situation. The recruitment based situation shows that for $v = 0$ it follows

$$N_e = \frac{4(1-r)N}{h+1}.$$

Since $(1-r)N$ is the number of females in a population, the fraction $\frac{(1-r)N}{h+1}$ is approximately the number of harems in a population. This leads to the fact, that the effective population size is about four times the number of the harems in a population. For other mating systems one can find different standardized variances and therefore different N_e's.

If the trivial case occurs where $v_f = 0$, i.e. the (female) generations do not overlap, the survival based effective population size behaves almost linear with respect to v_m. The N_e to N ratio is quite small and does not show any sufficiently big changes if v_m increases. Even if one changes the harem size from $h = 10$ to $h = 50$ and finally to $h = 100$, there is no big difference in the way how the graph behaves; it is just noticeable that the ratio shrinks with increasing h, but this is not surprising since the genetic diversity shrinks if there are big harems that are in the property of only one male.

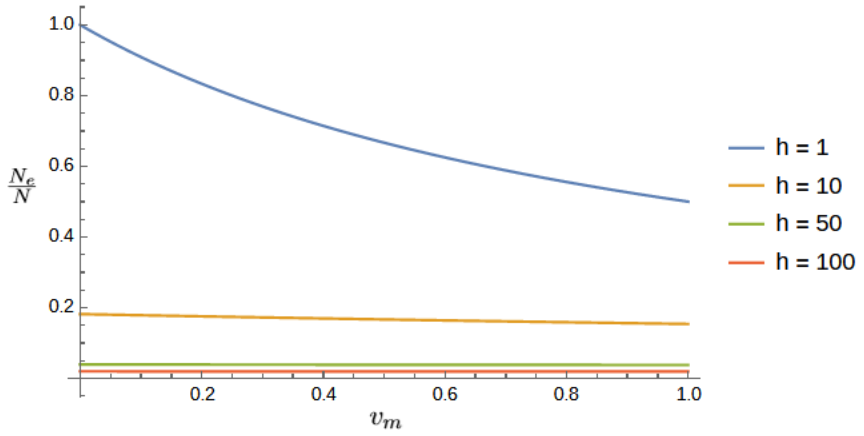

Figure 1: **Survival based N_e in harem polygyny for different harem sizes.** The survival based effective population size does not strongly depend on the male survival probability if the harem size is bigger that 1. If the harem size is 1, i.e. monogamous population, the $\frac{N_e}{N}$-ratio shrinks from 1 to 0.5 with increasing male survival probability. Parameter: $v_f = 0$

The only case, where the survival based effective population size changes a lot with respect to v_m is the case where $h = 1$, i.e. where a monogamous population occurs. The graph behaves exponential and the N_e to N ratio is 1 if $v_m = 0$ (completely non-overlapping generation). This is interesting, since this shows that there is a big difference between an actual harem population ($h > 1$) and a monogamous ($h = 1$) population.

Similar analyses can be done for the recruitment based effective population size. As

in the survival based model the parameters shall have the values $v = 0$. This leads to

$$N_e = \frac{4r(1-r)N}{2v + r(1-v)(h+1)} \underbrace{=}_{v=0} \frac{4r(1-r)N}{r \cdot (h+1)} = \frac{4(1-r)N}{h+1}.$$

We will analyse the model with respect to the sex ratio r. At first sight it could be confusing that the effective population size is not zero if $r = 0$. This does not make any sense, since there are no males that can mate with females if $r = 0$. The reason for this mistake lies in equation (7). There we said that we have to distinguish between $r \geq \frac{1}{h+1}$ and $r < \frac{1}{h+1}$. If one observes the model for the case where $h = 10$, one needs to use the formula

$$N_e = \frac{4r(1-r)N}{1 + v + [(1-r)(\frac{10r}{1-r} - 1) + r \cdot \frac{1-\alpha_f}{\alpha_f}](1-v)} \underbrace{=}_{v=0,\alpha_f=0.6} \frac{4(1-r)N}{11}$$

for $r \geq \frac{1}{11}$ and the formula

$$N_e = \frac{4r(1-r)N}{1 + v + [r \cdot \frac{1-\alpha_f}{\alpha_f}](1-v)} \underbrace{=}_{v=0,\alpha_f=0.5} \frac{4r(1-r)}{1+r}$$

for $r < \frac{1}{11}$. One can easily see that in this case

$$N_e = \frac{1}{1+r} \cdot 4 \underbrace{r(1-r)}_{=0, \text{ if } r=0} N = 0.$$

This is why both cases ($r < \frac{1}{h+1}$ and $r \geq \frac{1}{h+1}$) have to be in mind if one wants to calculate with this formula.

Now it can be interesting to analyze the models by keeping one harem size $h = 10$ fixed, but varying the female survival probability. Raising the probability of survival of females from $v_f = 0.1$ to $v_f = 0.9$ does not lead to big differences in the survival based model.

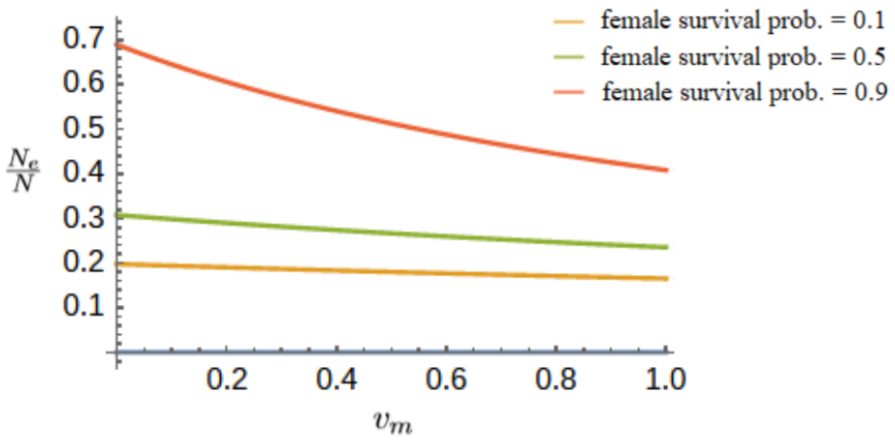

Figure 2: **Survival based N_e in harem polygyny for different female survival probabilities.** The higher the female survival probability is, the more the $\frac{N_e}{N}$-ratio depends on the male survival probability. Parameter: $h = 10$

An increasing male survival probability leads to a smaller effective population size while an increment in female survival probability leads to a bigger effective population size.

In the recruitment based situation the highest point of the graph moves to the right, if the female survival change increases. This means that for reaching a high N_e it is better to have more females than males if the probability for a female to survive the next breeding season is low. Since in a harem population few males are responsible for many females, it is better if the sex ratio tends to a small male proportion r, and a big female proportion, $1-r$.

In the next chapter another mating system will be observed. The mating system is called lottery polygyny.

5.3.2 Lottery Polygyny

The lottery polygyny is a mating system in which receptive females randomly meet males and then the either mate or not. The probability of mating is fixed. [16] There is an obvious difference between the harem polygyny and the lottery polygyny. Therefore, one also needs another variance of male fecundity. All males attempt to mate and the breeding females mate an average of n times; therefore, a Possion approximation to binomial sampling can be assumed. [16] We now do not observe one male that mates with a group of females and the females are not allowed to mate with anybody else, but two main steps have to be considered

- probability for a male individual to be chosen for the mating process: $\frac{1}{rN}$
- successful matings: $\underbrace{n}_{\text{average number of mates per female}} \cdot \underbrace{\alpha_f}_{\text{fertile females}} \cdot \underbrace{N(1-r)}_{\text{number of females}}$.

So the standardized variance is

$$s_m^2 = \frac{r}{n\alpha_f(1-r)}$$

and this can be inserted in the formula (5) and (6).
This leads to

$$N_e = \frac{N}{1 + \frac{v_m+v_f}{2} + (n^{-1} + 1 - \alpha_f)\frac{1-v_f}{2\alpha_f}} \tag{8}$$

or

$$N_e = \frac{4r(1-r)N}{1 + v + (n^{-1} + 1 - \alpha_f)\frac{r}{\alpha_f}(1-v)}. \tag{9}$$

In the first step, we will consider that there are non-overlapping generations.

5.3.2.1. Non-overlapping generations

If the generations do not overlap, the individuals die after one breeding season. This means that $v_m = v_f = v = 0$. The formula (9) therefore transforms to

$$N_e = \frac{N}{1 + (n^{-1} + 1 - \alpha_f)\frac{1}{2\alpha_f}}.$$

In the survival based differences situation the effective population size only depends on the breeding success of the females and the number of mates per female that occurs. Let us assume there is no limit and $n \to \infty$ and every female can breed, i.e. $\alpha_f = 1$. This means that $n^{-1} \to 0$ and $N_e = N$. In the special case where $n = 1$ and $\alpha_f = 1$, i.e. every female breeds once, the effective population size is

$$N_e = \frac{N}{1 + (2 - \alpha_f)\frac{1}{2\alpha_f}} \underbrace{=}_{\alpha_f=1} \frac{N}{1 + \frac{1}{2}} = \frac{2}{3} \cdot N.$$

This means that N_e is by a third smaller that the census size N.

The same can be done for the recruitment based situation. In this situation it plays a very important role if the sex ratio is 1:1 or if the sex ratio is very unequal. In the non-overlapping situation $v = 0$ again and this leads to

$$N_e = \frac{4r(1-r)N}{1 + (n^{-1} + 1 - \alpha)\frac{r}{\alpha_f}}.$$

The very obvious case, where one sex does not exist at all, i.e. $r = 0$ or $1 - r = 0$, is not worth mentioning because then the individuals cannot mate and $N_e = 0$ (and also $N = 0$ after one breeding season because no new progenies are born and the adults die because of the assumption $v = 0$). If the sex-ratio is 1:1, i.e. $r = 1 - r = \frac{1}{2}$ and all females mate, this leads to

$$N_e = \frac{N}{1 + \frac{1}{2n}}.$$

The effective population size therefore again depends on the number of mates per female. If $n = 1$ we have the same situation as before

$$N_e = \frac{2}{3} \cdot N.$$

One can see, that the two situations

1. sex differences because of survival based differences
2. sex differences because of recruitment based differences

are equal under several assumptions. But in extreme cases the effective population sizes can be very different. This shall be shown in the following where overlapping generations shall be observed.

5.3.2.2. Overlapping generations

In overlapping generations v_m, v_f and v have to be unequal zero. The simplest cases are again the situations where $\alpha_f = 1$, i.e.

$$N_e = \frac{N}{1 + \frac{v_m + v_f}{2} + \frac{1}{n} \cdot \frac{1 - v_f}{2}}$$

and

$$N_e = \frac{4r(1-r)N}{1 + v + n^{-1} \cdot r(1-v)}.$$

Since the first equation is the survival based equation, it is interesting how N_e changes with a change of v_m and v_f. And since the second equation is the recruitment based equation, it should be in the focus how r changes the effective population size. For this reason, we first let $v_f = 1, \alpha_f = 1$ and $n = 1$ and therefore

$$N_e = \frac{N}{1 + \frac{v_m + v_f}{2} + \frac{1 - v_f}{2}} = \frac{N}{1 + \frac{v_m + 1}{2}}$$

and

$$N_e = \frac{4r(1-r)N}{1 + v + r(1-v)}.$$

It is an easement that in the special case of $\alpha_f = n = 1$ the survival rate of the females does not play any role.

Now the question can be asked, which changes will occur if n does not equal 1. For this situation, we will consider different situations with various n. Lets look at the survival based effective population size, when $v_f = 0$, $\alpha_f = 0.5$ and n has different values from $n = 1$ to $n = 1000$.

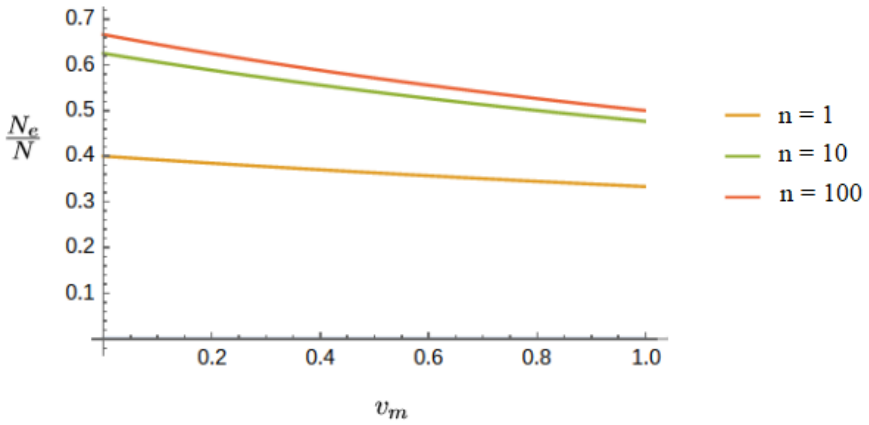

Figure 3: **Survival based N_e in lottery polygyny with $a_f = 0.5$.** While the effective population size shrinks with respect to the male survival probability, v_m, it increases with respect to the average number of progeny per breeding female, n. Parameters: $v_f = 0$, $\alpha_f = 0.5$

It makes sense that the effective population size is higher if there are more progeny per female since the genetic pool is larger in this case. But if just half of the females breed, the other half has no progeny at all. Therefore, it is interesting to compare this result with the case where every female breeds, i.e. $\alpha_f = 1$. In this case the behaviour of the graphs is similar as in the previous case, but the $\frac{N_e}{N}$-ratios start in general higher with respect to v_m. If every female ($\alpha_f = 1$) has one progeny, this leads to a similar effective population size as if half of the females ($\alpha_f = 0.5$) have in average 100 progeny. This indicates that a small proportion of breeding females results in high loss of genetic diversity and the population needs many progeny per breeding female to catch up with this loss.

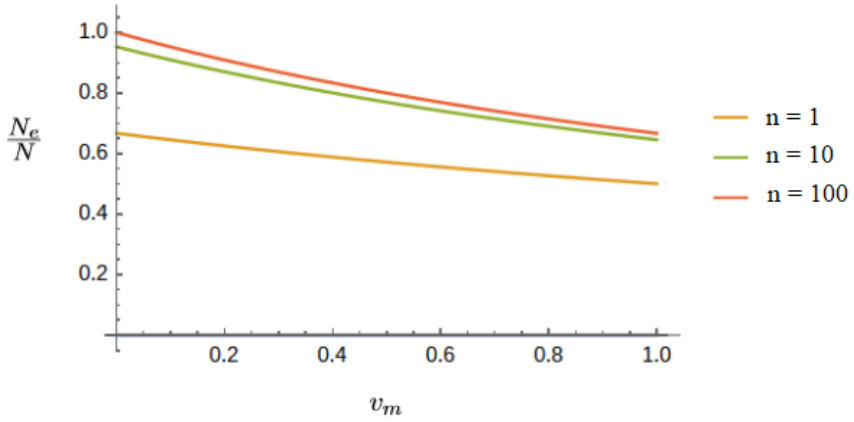

Figure 4: **Survival based N_e in lottery polygyny with $a_f = 1$.** If every female has one progeny (yellow line, figure 4), this leads to a similar effective population size as if half of the females have in average 100 progeny (red line, figure 3). Parameters: $v_f = 0$, $\alpha_f = 1$

In general, the survival based effective population size does not strongly depend on the male survival rate if there is a small number of average mates per breeding female, since the graphs behave quite constant. In the next section a third polygyny shall be discussed.

5.3.3 Dominance Polygyny

The dominance polygyny is similar to the lottery polygyny. The males mate again with different females and the females have a limited number of mates. The main difference to the other cases is the fact that the males compete for a status symbol. This can be dominance or a physical resource as strength or a save food/breeding place. This status symbol is a prerequisite of competing for female mating, i.e. a male has no chance to mate without 'dominance'. Therefore, next to the parameter α_f we need the parameter α_m, which is the proportion of dominant males. Since the dominance prerequisite is the only difference to the lottery polygyny, $\alpha_m = 1$ means that we have the lottery situation again. To compute the probability that a dominant male mates, we need to divide the number of mating males through the number of mates in general. This leads to

$$\frac{r\alpha_m}{n\alpha_f(1-r)}.$$

For the standardized variance in mating success we need to add the chance of not being dominant $1 - \alpha_m$ and divide both of the term through the proportion of mating males α_m. The standardized variance in mating success is therefore similar to the one in the lottery polygyny

$$\frac{r}{n\alpha_f(1-r)}$$

but a bit different

$$s_m^2 = \frac{r}{n\alpha_f(1-r)} + \frac{1-\alpha_m}{\alpha_m}.$$

Now we insert s_m^2 in the equations (5) and (6) again and receive

$$N_e = \frac{2N}{2(v_m + v_f) + (1-v_f)(1+n^{-1}) + \frac{1-v_m}{\alpha_m}}$$

for the survival based case and

$$N_e = \frac{4r(1-r)N}{2v + (1-v)[r(1+n^{-1}) + \frac{1-r}{\alpha_m}]}$$

for the recruitment based case. For the dominance polygyny we set $\alpha_f = 1$, because otherwise we would have a proliferation of parameters. In the dominance polygyny the males are more important, because the development of the effective population size depends on the dominance of the males. First of all, we see that the the effective population sizes in the special cases $v_m = v_f = 0$ and $n = \alpha_f = \alpha_m = 1$ are N again.

It is interesting to observe the survival based effective population size again with respect to v_m. If $v_f = 1$ then n has no influence on the effective population size anymore. Therefore, we choose $v_f = 1$, such that the formula only depends on v_m and α_m. We will see, that the effective population size behaves very different depending on how α_m was chosen. If every male is dominant, $\alpha_m = 1$, the effective population size shrinks with an increasing v_m. If the proportion of dominant males is low, $\alpha_m = 0.1$, the survival based effective population size increases with an increasing v_m.

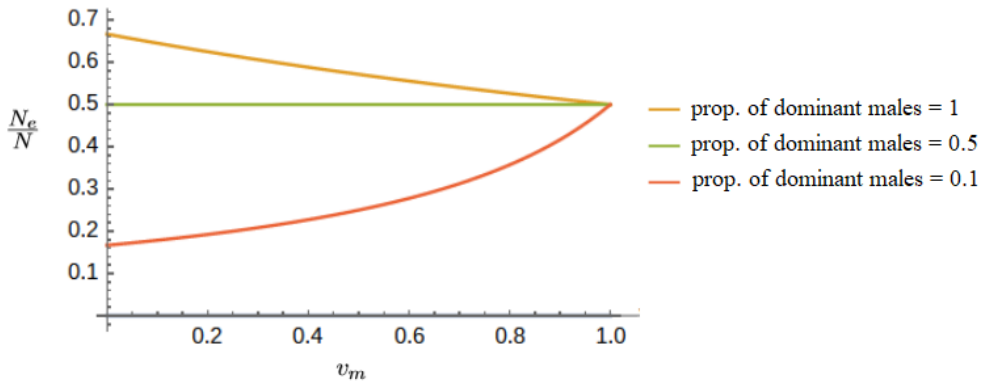

Figure 5: **Survival based N_e in dominance polygyny.** For a high/low proportion of dominant males the effective population decreases/increases with respect to the male survival probability. If the proportion of dominant males is 0.5, the effective population size does not depend on the male survival rate at all. Parameter: $v_f = 1$

This means that a small amount of dominant males leads in general to a smaller effective population size but the size increases with increasing male survival probability if there is less than a half of dominant males. This is interesting as dominance probability $> \frac{1}{2}$ relates it to harem polygyny. If all males are dominant, the effective population size is higher in general, but it decreases with increasing survival probability. For $v_m = 1$ the proportion of male dominance has no influence on the effective population size. In every case (no matter if α_m is high or low) the $\frac{N_e}{N}$-ratio is 0.5.

In the recruitment based case the effective population size behaves similar with respect to r if one chooses $v = 1$. First the $\frac{N_e}{N}$-ratio increases with an increasing r and then it shrinks again (i.e. a parable) occurs. But the proportion of dominant males only changes the slope of the increment and shrinkage and the position of the highest effective population size.

However, one sees that the impact that α_m has on the recruitment based $\frac{N_e}{N}$-ratio is not that important. The difference between the highest point of the N_e to N ratio if $\alpha_m = 0.5$ or $\alpha_m = 1$ is only rounded 0.06.

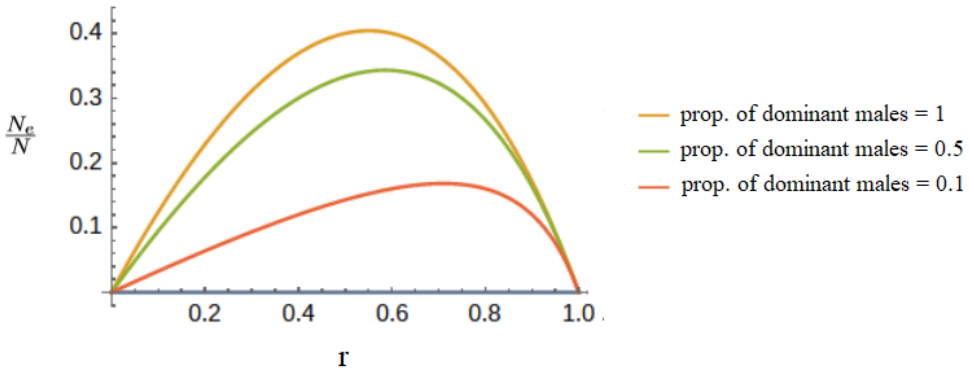

Figure 6: **Recruitment based N_e in dominance polygyny.** The smaller the proportion of dominant males, the higher the sex ratio is needed to reach the maximum effective population size. If all males are dominant, the sex ratio has to be $r = 0.5$ for the highest possible effective population size. Parameter: $v = 1$

If one compares this result with the survival based ratio, one sees that this changes more. While for $\alpha_m = 0.1$ the function is monotonically increasing, the functions decreases for $\alpha_m = 1$ with respect to v_m. The reason for this differences is the fact that in the survival based case it depends on how long the individuals survive. The females survive until the next breeding season for sure $v_f = 1$. If the males have very little mating success, i.e. only 10% of the males are dominant and can mate, it is better if the males survive until the

next breeding season such that the mates per male individual spread through different generations. Therefore, the graph increases with the survival chance of the males. If there are many dominant males in the population, different males mate with different females, and the variance of the matings spreads anyway. Then it is better of the N_e if the males die after one breeding season, such that in the next generation there are different males to mate with. The higher the probability that the male individuals survive, i.e. the bigger v_m gets, the more N_e decreases.

After discussing the three mating systems with the effective population size models separately, it is interesting to compare two models. However, we will compare the harem polygyny with the dominance polygyny. The only difference between these two models is the parameter α_m that is 1 if we observe a lottery polygyny. To make the two mating systems comparable, one could find a relationship between h, the harem size, and α_m, the proportion of the dominant males. It makes sense to set $h = \frac{1}{\alpha_m}$ since the bigger the proportion of dominant males, the smaller the harem size gets. If the harem size is small, more males have the chance to ‚own' a harem, i.e. to mate. The other way round, if the proportion α_m is small, the harem size is big and therefore there are only few males that are dominant respectively that ‚own' a harem and can mate. Lets compare the $\frac{N_e}{N}$-ratios for the recruitment based situation. This leads to the two formulas

$$N_e = \frac{4r(1-r)N}{2v + r(1-v)(h+1)}$$

and

$$N_e = \frac{4r(1-r)N}{2v + (1-v)[r(1+n^{-1}) + (1-r)h]},$$

where in the last equation $\frac{1}{\alpha_m} = h$.
Since in the harem polygyny n was set to be 1 the two situations are only comparable, if $n = 1$ in the dominance polygygy as well. We want to compare the two mating systems by varying the parameter v from 0 to 1. Further, we choose $h = 10$ and $n = 1$ for the remaining parameters.

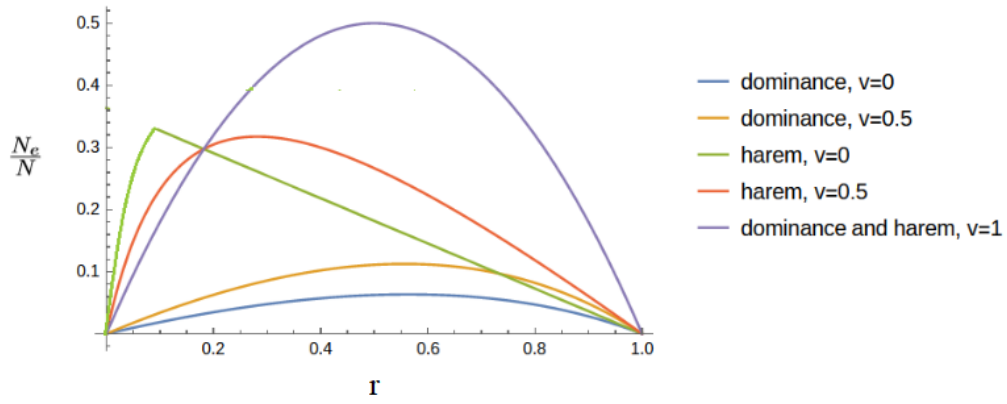

Figure 7: **Recruitment based N_e in harem and dominance polygyny.** If every individual survives to the next generation ($v = 1$), the effective population sizes are the same in a dominance and a harem polygyny. In this case the effective population size just depends on the sex ratio r. Parameters: $h = 10$, $n = 1$

The harem effective population size and the dominance effective population size behave mirrored with respect to r. In the case where $v = 1$ the two effective population sizes are equal. This means that the effective population sizes for these two mating systems are the more equal the higher the survival probability is.

Another case that could be observed is the situation where $v = 1$ (i.e. every individual survives until the next breeding season), $h = 10$ and $n = 1$ occurs. If every individual survives and the male dominance has the 'same proportion' since we set $\alpha_m = \frac{1}{10}$ as we do indirectly in the harem case, the two models are identical. This means under the given assumptions the only difference between harem polygyny and dominance polygyny is the survival rate that affects the harem polygyny more strongly. It is also noticeable that the highest possible $\frac{N_e}{N}$-ratio is 0.5 reached at $r = 0.5$. This means that for a 1:1 sex ratio the effective population size is half of the actual population if $v = 1$. It is interesting to observe that the harem size respectively the proportion of dominant male does not play any role any more, since

$$N_e = \frac{4r(1-r)N}{2v + r(1-v)(h+1)} \underbrace{=}_{v=1} 2r(1-r)N$$

and

$$N_e = \frac{4r(1-r)N}{2v + (1-v)[r(1+n-1) + (1-r)h]} \underbrace{=}_{v=1} 2r(1-r)N.$$

If we have a totally overlapping population the harem size can increase to ∞, but it still has no influence to the effective population size. This is quite interesting, since this means that

there is also no difference between a monogamous population (i.e. $h = 1$) and a polygamous population.

After explaining the three types of mating systems in the context of the survival based and the recruitment based effective population size, it is now important to look at the whole model (4). Since there are more parameters included in one model, we will use 3D-plots and contour plots to find an appropriate representation of the model.

5.4 Results

Since we already analyzed the effective population sizes in the sense of the two separated models (survival based and recruitment based N_e), it is now interesting to observe the whole model without any differentiation (4). We will state the model again such that one remembers what the whole formula looked like

$$N_e = \frac{4r(1-r)N}{1 + v_m(1-r) + v_f r + (1-r)(1-v_m)s_m^2 + r(1-v_f)\frac{1-\alpha_f}{\alpha_f}}.$$

Plotting the model with respect to r and α_f and manipulating the graph, shows that the most interesting result is achieved if $v_m = 0$, $v_f = 0.5$ and s_m^2 depends on the mating system. Choosing the parameters in this way and inserting the different variances s_m^2 for the harem and the lottery polygyny, leads to the result that the model in the context of the harem polygyny is quite different than the model in the context of the lottery polygyny.

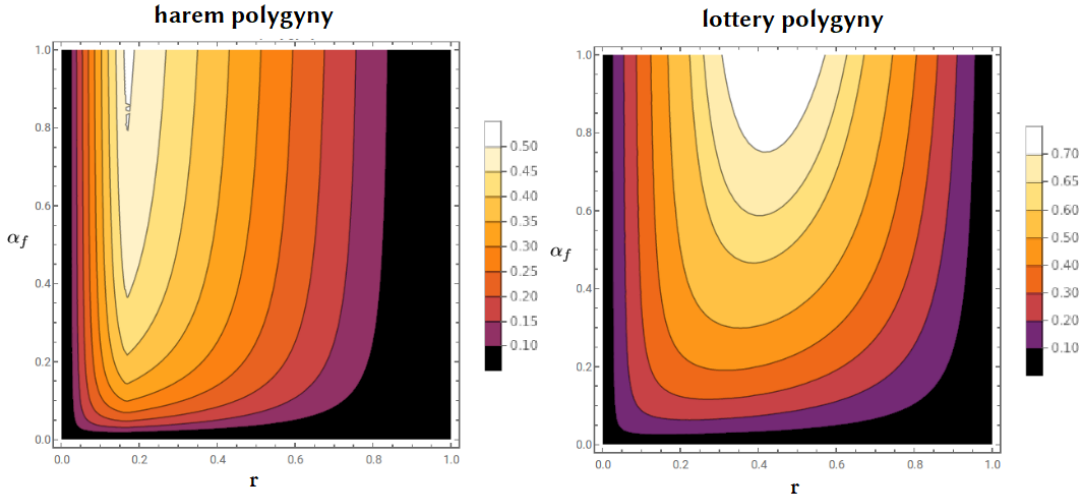

Figure 8: **Harem and lottery polygyny.** For reaching a high effective population size, in a lottery polygyny their individuals need a more balanced sex ratio than in a harem polygyny. Parameters: $h = n = 5$, $v_m = 0$, $v_f = 0.5$

One can see that the contour plot is quite different for two different mating systems, even if the parameters were chosen equally. If the goal is a high effective population size, the lottery polygyny needs a more balanced sex ratio than the harem polygyny. This makes sense since the harem polygyny benefits from a high amount of females and a small amount of males. The lottery polygyny reaches the highest effective population size when many females breed, $\alpha_f \approx 1$, and the sex ratio is approximately $1:1$. The harem polygyny

benefits from a low ratio of males, $r << 1$ and a high proportion of breeding females. One has to be careful since the contour plots have different legend distributions, i.e. in the harem polygyny the $\frac{N_e}{N}$-ratio is in average lower than in the lottery polygyny.

Now the same can be done for the harem polygyny and the dominance polygyny. This time the contour plot shall depend on r and on $v = v_m = v_f$, i.e. the survival probabilities are equal. The remaining parameters are $\alpha_f = 0.9$ and s_m^2 where s_m^2 again depend on the mating system.

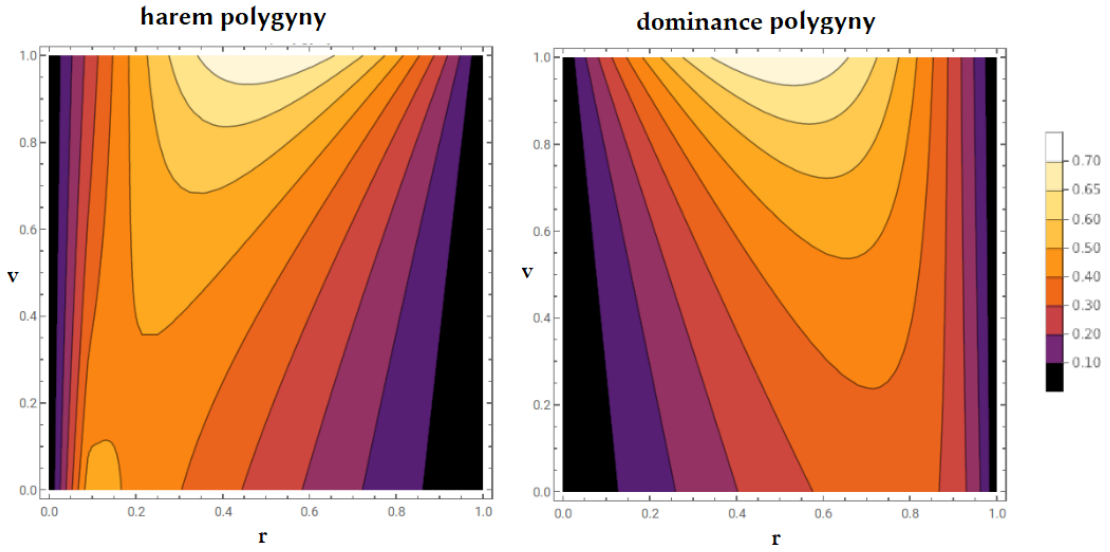

Figure 9: **Harem and dominance polygyny.** The harem and the lottery polygyny behave similar but vice versa with respect to the sex ratio. While the survival rate has less influence on the $\frac{N_e}{N}$-ratio, the effective population size, depends strongly on the sex ratio. Parameters: $h = 10 = \frac{1}{\alpha_m}$, $n = 100$, $\alpha_f = 0.9$

One sees that the effective population sizes of the two mating systems behave similar but vice versa with respect to v and r. This is interesting since it means that in the harem polygyny a small r (less males than females) causes in general a higher $\frac{N_e}{N}$-ratio and for the dominance polygyny a higher r is better for a higher effective population size. This can be explained by the fact that a harem polygyny needs more females that males. Depending on how high the proportion of dominant males is, a higher r is better for the dominance polygyny since the population needs *many* males if there are just a few dominant males.

Similar to the harem and the lottery polygyny, the lottery and the dominance polygyny can be compared. Since the lottery polygyny is a special case of the dominance

polygyny with $\alpha_m = 1$, the two mating systems shall be compared by setting $\alpha_m = 0.1$ in the dominance polygyny. However, in the lottery case every male is dominant and in the dominance case only a proportion of the males is dominant.

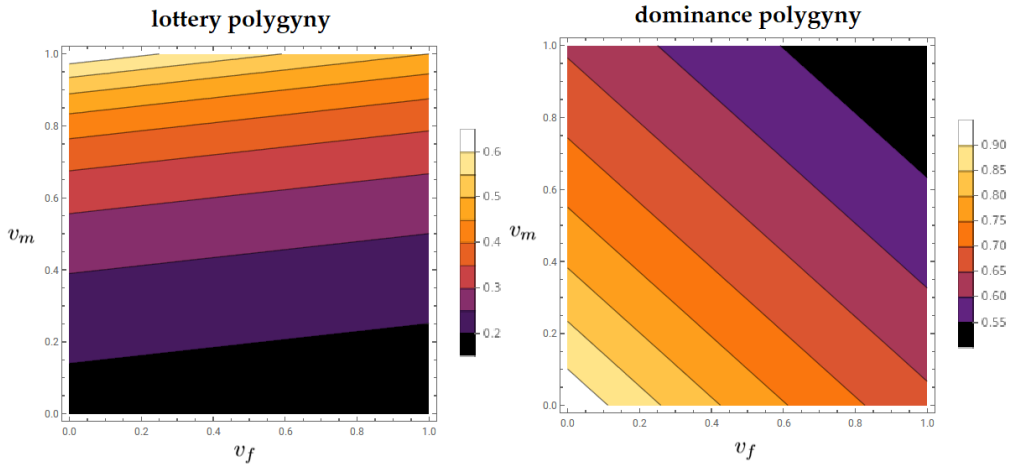

Figure 10: **Lottery and dominance polygyny.** In a lottery polygyny the effective population size is nearly independent of the female survival probability. In a dominance polygyny the female survival probability indeed influences the $\frac{N_e}{N}$-ratio. Parameters: $r = 0.5$, $n = 100$, $\alpha_f = 0.9$, $\alpha_m = 0.1$

The results indicate that, in a lottery polygyny, the effective population size is nearly independent of the female's survival probability, v_f. It only changes with respect to the survival probability of the males, v_m. On the other hand, v_f indeed influences the $\frac{N_e}{N}$-ratio in the dominance polygyny model. This means that the female survival probability gets important as soon as there is only a small proportion of mating males.

One can also compare lottery polygyny representation of figure 9 with that of figure 10. In the first one, the survival probability was equal for males and females, $v_m = v_f = v$ and in the second one, we assumed $1 : 1$ sex ratio. Since differences in sex can occur due to survival differences or due to recruitment differences, these two plots give a good overview about these two difference types. One can see that in the first case the effective population size was seriously reduced by a high or low sex ratio, i.e. differences in recruitment influences the $\frac{N_e}{N}$-ratio strongly. The second plot shows that the female survival probability does not influence the effective population size, i.e. the sex differences resulting from survival differences have no high affect on N_e. These results confirm Nunneys hypothesis that the potential for a sex-ratio bias to change N_e is much reduced given a survival bias. [16]

Another interesting result that is worth to be examined, is the effective population size for

the dominance polygyny with respect to α_m and v_m. Manipulating the contour plot showed that effective population size in the context of a dominance polygyny depends strongly on the proportion of dominant males. Therefore, the model is plotted with respect to the male dominance proportion, α_m and the male survival probability, v_m. The remaining parameters are $v_f = 0.9$, $r = 0.5$ and $\alpha_f = 1$. Surprisingly, the contour plot does not only strongly depend on the proportion of dominant males but also on the number of progenies per female. The contour plot nearly does not change if one varies between $n = 10$ and $n = 50$ or $n = 1000$. But there is a quite big difference between $n = 1$ and $n = 10$. The following two contour plots show both the strong dependence on α_m and the impact of the number of progenies per female.

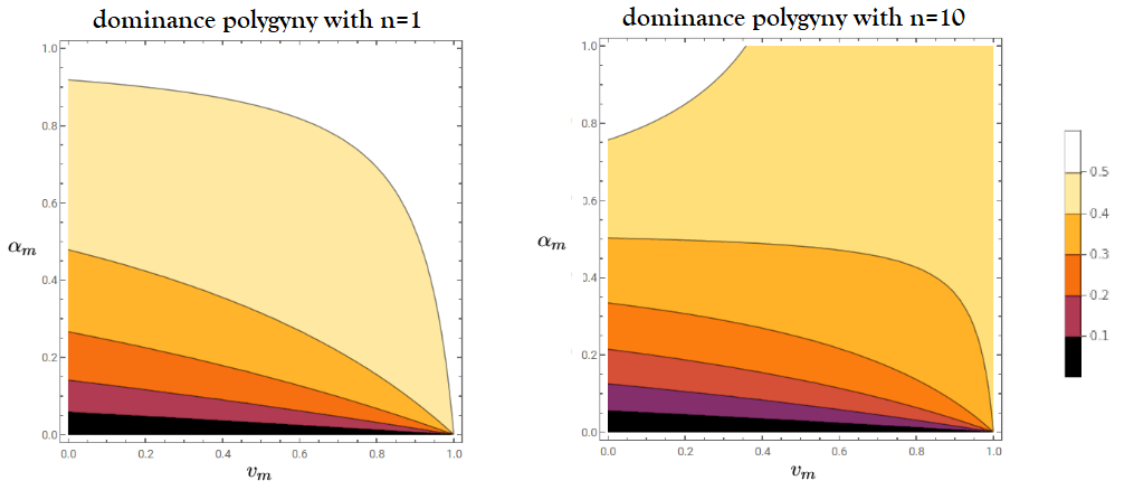

Figure 11: **Dominance polygyny**. One progeny per female results a high effective population size if nearly all males are dominant (independent of the male survival probability). If a breeding female has in average 10 progeny, the male survival probability must be low to reach a high effective population size. Parameters: $v_f = 0.9$, $r = 0.5$, $\alpha_f = 1$

The contour plots show the dominance effective population size with respect to v_m and α_m. It is interesting how the $\frac{N_e}{N}$-ratio changes if the proportion of dominant males varies. In the case where $n = 1$, one can see that a high proportion of dominant males lead to a high effective population size in any case, no matter if v_m is high or low and a high probability for males to survive also leads to a high effective population size, no matter if the proportion of dominant males is high or low. In the case where $n = 10$ a similar situation occurs for a high α_m. Nevertheless, in this case, an increasing male survival probability leads to a shrinking $\frac{N_e}{N}$-ratio if α_m is high.

In summary it can be said that examining the effective population size models in detail

has shown some very interesting results. Answering the research questions shall summarize these results.

RQ1. *How do the effective population sizes change if different mating systems are considered and when are they equal?*
The difference between the effective population sizes in the context of the three mating systems can be significant. While a smaller r leads to a higher N_e in the harem polygyny, the lottery polygyny needs an equal 1:1 sex ratio for a high effective population size. The effective population size models of the harem and the dominance polygyny behave similarly but vice versa with respect to the sex ratio: By fixing the survival probability, the effective population size is approximately the same for both the harem and lottery polygyny if the sex ratios have the relation $r_h = 1 - r_d$. However, r_h denotes the sex ratio in the harem polygyny and r_d in the dominance polygyny. In general, one reaches the maximum of the effective population size with increasing sex ratio before decreasing. Nevertheless, in a harem polygyny one needs a relatively low sex ratio, r_h, and in a dominance polygyny, one needs a relatively high sex ratio, $r_d = 1-r_h$, to reach a high effective population size.

Since the three models only differ in the standardized variance in male fecundity, the effective population sizes are the same if

$$h = \frac{1}{n}$$

and if

$$\alpha_f = \alpha_m = 1.$$

If every male and female individual can produce offspring and the harem size is reciprocal to the number of progeny per female, it does not matter which model is used.

RQ2. *Which parameters lead to the strongest differences between mating systems (lottery vs. harem, dominance vs. harem and lottery vs. dominance)?*
The difference between the effective population sizes in the context of lottery polygyny and in the context of harem polygyny is due to the sex ratio. As already interpreted, this comes from the fact that in a harem polygyny an unequal sex ratio ($r >> \frac{1}{2}$) leads to an excess of males and to only a small amout of harem owners. The difference between the dominance and the harem polygyny is remarkable since the formulas are similar but vice versa with respect to v and r. The lottery and the dominance polygyny differ in that way that the lottery polygyny is a special case of dominance polygyny with $\alpha_m = 1$. This is leads to the result that the female survival probability only effects the effective population size if α_m is small. Otherwise, i.e. in the lottery case, where $\alpha_m = 1$, the female survival probability has nearly no influence on N_e.

RQ3. *Which parameters have a strong impact on the effective population size of the models assesed and which parameters do not influence the models significantly?*
Observe the effective population sizes in the context of the three mating systems shows that the proportion of dominant males has a significant impact on the dominance $\frac{N_e}{N}$-ratio.

Also the number of progeny per female n and the male survival probability have a strong influence on the dominance effective population size if the other parameters are chosen in a certain way. In the lottery and harem polygyny the N_e model is strongly influenced by the sex ratio r. However, the male survival probability has, under the assumption of a high harem size, h, and a high average number of progeny per female, nearly no impact on the effective population size.

In general, the comparisons of the N_e's in the context of different mating systems showed that the effective population size models are equal under the assumption that $h = \frac{1}{n}$ and $\alpha_f = \alpha_m = 1$ but they can also be very different: In a harem polygyny, a low harem size and a relatively low sex ratio lead to a high effective population size. In the lottery polygyny, a high proportion of breeding females, a 1:1 sex ratio and a high male survival probability are needed to get a high effective population size. In a dominance polygyny, a high proportion of dominant males (if $1 - r > r$) and a low female survival probability are necessary for a high effective population size. However, comparing the case where recruitment leads to sex differences with the case where survival leads to sex differences shows that the chance for a sex-ratio bias to change N_e is much reduced given a survival bias. However, recruitment has a higher influence on the effective population size that survival probabilities.

5.5 Outlook for further research

As it was done for the variance effective population size models, it would be interesting to examine models for the inbreeding effective population size under the prospective of different mating systems. Unfortunately, it does not work to just take the models and formulas from the variance effective population size and use them for the inbreeding effective population size. Since one uses the change in heterozygosity for computing the inbreeding effective population size rather than considering the variance in allele frequency, one cannot do an inbreeding correction to the variance N_e. Hill confirms this in one of his papers:

> "The calculations have been made solely for the variance effective number. A similar argument can be adopted using the contribution of one cohort to the subsequent increase in inbreeding to show that inbreeding effective numbers are the same, as they must be in random mating populations of constant size." [12]

However, it would be interesting to find a possibility how the different mating systems can be included in an inbreeding effective population size. This would be a meaningful project for future research.

6 Examples: Different mating systems and their effective population size

Until now, the theoretical models of the effective population size were introduced. Now it is the aim to find some actual examples where the mating system of the populations play an important role. It is the aim to address the question of what effects ecology has on the mating systems (and partly vice versa). Since the males maximize their reproductive success if they fertilize many females, it is often useful for males to use scarce resources to attract or to control females. Therefore,

- spatial distribution and
- temporal distribution

play an important role in the development of an appropriate mating system. [5] When the females are spatially evenly distributed, it is difficult for one male to control many females because there are not several females at one place. Also the temporal distribution has different effects. If all females are ready to conceptualize semen at the same time, one male is not able to fertilize all of them at the same time. However, if the females become receptive one after the other, the male can fertilize the females one after each other. For this reason, these two factors naturally have a major impact on the mating systems.

6.1. Polygyny by defense of resources: yellow-bellied marmots

Male yellow-bellied marmots build their own burrows, which are particularly suitable for mating and for caring young animals. The sizes of these habitats ranging from very small areas (0.01 ha) occupied by only one female to large areas (7.1 ha) where more females can live at the same time. [18] The better the males build these caves, the more females are attracted by its constructs. For the females, however, it would be better if the males focused only on them specifically, i.e. monogamy would be better for the females. This is a good example of the mating conflict between males and females described earlier. However, the quality difference between a bad but monogamous cave and a good cave in which the female would have to share the male with a rival must be large enough such that the female voluntarily chooses the non-monogamous cave. [5] The unequal goals of the two sexes can be made clear again with the following graphic.

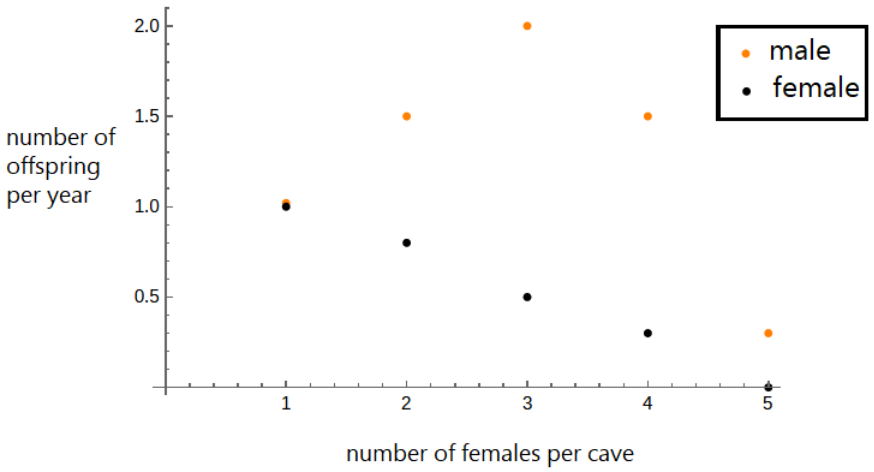

Figure 12: **Yellow-bellied marmots: Mating success of males and females.** The number of offspring per year in dependence of the number of females per cave. There is a difference between the male success (black) and the female success (orange).

In figure 12 one can see that the male succeeds if he has 2-3 females in his cave. For the female it is better if there are no other females in the territory. This example demonstrates the narrow degree between a polygamous and a monogamous population and how the defense of resources can lead a successful mating. In the next chapter another type of polygyny will be discussed.

6.2 Polygyny by defence of females: elephant seals and horses

6.2.1. Elephant seals

Hareming is particularly common in mammals. The already described pressure of the males to mate with many females and the pressure of the females to protect and raise the few offspring they get - compared to the males - leads to useful harem formations. With those formations many females can be protected at the same time, so the probability of being caught in an attack is relatively small. The males can mate with many females and pass on their genes to many offspring.

An example of animals in the context of this mating system is the elephant seal. The mating system of southern elephant seals is strongly polygynous. One male holds the control of a group of females and this male is the only one that has access to the females. Depending on parameters like the size of the harems, the density of males and the level of control of

the harem holder, the effective population size varies.

Every breeding time the females meet at one of the rare breeding places where they give birth to their offspring. If the females are already collected at one place, one male is able to control them all. There are many males that want to be this "lucky male", but only the strongest and the most dominant male can be the owner of the whole harem. Most harems tend to be larger than fifty females, and every season there are at least two harems with more than one hundred females. [14] The females vary in time of their oestrous which allows the males to mate with multiple females one after each other. In populations where very big harems occur, one male can be the father of up to 95% of all offspring. [14] Therefore, the harem holder has to be a good strategist such that he does not loose the control of his harem. Since this job is very exhausting, harem holders are often only 1-2 years in this "holder situation".

6.2.1.1. Isolated couple

If females are not part of a harem but are pregnant respectively give birth to a cub, they normally have difficulties participating a harem again. One reason for that is the fact that the progenies are not mobile enough to join the harem. Another reason are the secondary males that try to keep the lonely females away from a harem, maybe to establish a own harem. [10]

6.2.1.2. Breeding season

After the 11-month pregnancy, the females give birth to only one cub. [14] Afterwards the breeding time starts. While the females are breeding, the male harem holder protects the females and the offspring. During mating, the male sometimes accidentally crushed the newborn cubs. This is of course a big loss for the mother of the cub, since she used a lot of time and energy to get this progeny. But for the male the cost is on average lower, since the eras of the harem holders are short enough such that this offspring is possibly not his.

In 2005, Oliveira published a study about the effective population size in a fur seal population. [6] The researchers observed an elephant seal population in Peru between 1951 and 1979. The census size of the elephant seals, was between 40 and 20.255 individuals. Few years later (from 1992 and 1996) the population was observed again and one found out that the census size grew up to around 25.000 individuals. However, from the beginning 40 individuals the population grew enormously. Estimations showed that for this population the effective population size was

$$N_e \approx 2153.$$

Obviously, the census size of this population is much higher than the effective population size and the $\frac{N_e}{N}$-ratio is

$$\frac{N_e}{N} \approx 0.0861.$$

This big difference between the census size and the effective population size can be explained by the fact that a big number of individuals have only a small number of individuals as ancestors. Another example for a population that lives in harem formations is the horse.

6.2.2. Wild Horses

Wild horses often live in harem populations. The harem size differs from the one of elephant seals. In a study by Berger [1] some wild horses were observed over six years. At the beginning 49 individuals were observed and at the end of the study the number of individuals almost tripled. There was no immigration and emigration noted during the six years. The horses lived in harems and from 54 adult males only 16 were stallions in possession of harems of one or more females. The harems consisted of one male and a different number of females. Some corrections by Nunney (concerning the breeding adults) lead to a female mature time of 3 years and a male mature time of 5 years. The mean of the average age of the parents is 7.86 respectively 7.23 years. The average harem size was $h = 3.6$ and the sex ratio was $r = 0.48$, i.e. a bit more females than males. The proportion of breeding females was $\alpha_f = 1$. This means that every female was able to breed. This calculations led to an $\frac{N_e}{N}$-ratio of 0.82. [17] However, the effective population size was quite high. The relatively small harem sizes and the fact that there was no $1:1$ sex ratio could cause this high effective population size. This example indicates that a harem population does not have automatically a small effective population size. The above elephant seals example exhibited a smaller effective population size. However, the effective population size always strongly depends on the different parameters of the formulas. In the next chapter another type of polygamy will be discussed.

6.3. Polygyny by male dominance: greater sage-grouse and bison

6.3.1. Greater Sage-Grouse

Two examples of polygyny have now been presented. Now there is another possibility of polygyny in which the male does not defend resources or females, but simply has to have a dominant feature in order to conquer a female. The dominant male transmits some dominance signals to attract females. This can be a particularly noticeable appearance but it can also be the most beautiful song (birds) or the best dance (some types of chicken). An example of a species that exhibits this male dominance is the greater sage-grouse. During the courtship season, males and females gather in the courtship areas. The males then perform dance-like fights to defend certain territories. [5] From a resource point of view,

however, these areas are meaningless. They have neither particularly good protection nor other advantages. In fact, the dance just seems to be about defending a fictitious territory. Interestingly, those males who have fought for a territory in the middle of the square - although this middle is neither better nor worse than the edge - have an increased number of mating successes. [5] The females are more likely to chose a male who has his territory in the middle of the square. In the following figure it can be seen that the males achieve up to 50% (!) of all copulations if they were successful in fighting for the middle territories. The difference between place 1 and place 2 is very large which is why the curve can only be roughly described as exponential.

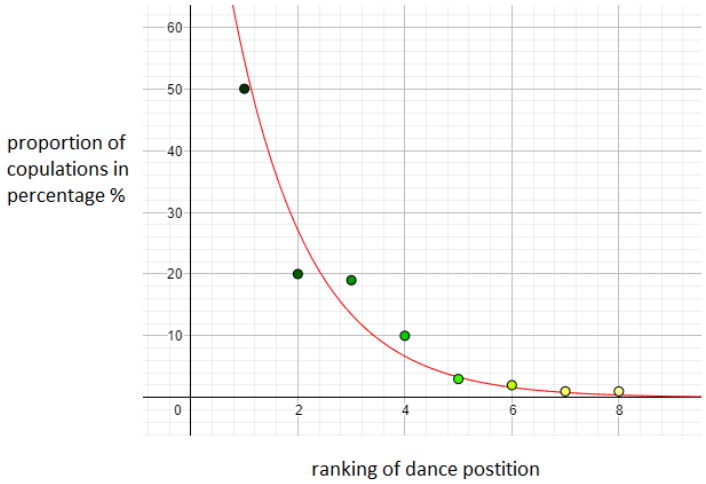

Figure 13: **Greater sage-grouse: mating success depending on the dance position.** The females are more likely to chose a male who has his territory in the middle of the square. The mating success nearly decreases exponentially with respect to the dance position.

The exponential function has the form $f(x) = 110 \cdot e^{-0,7 \cdot x}$, where the function f describes the copulation-proportion and the argument x describes the rank of the dance. This means that, in average, the proportion of copulation is approximately halved per rank. Therefore, it is the aim of every male to win the competition or at least not be at one of the last ranks. It is interesting that there are no actual advantages in the highly competitive male territories. This means that the females actually only choose the "middle" males because of their dominance and not because of good breeding or feeding grounds. Now another animal population shall be discussed in the next chapter before expounding an example for a polyandry population.

6.3.2. Bison

Another example for polygyny by male dominance can be founded in bison. Bison are animals that live in herds. Researchers observed some bison herds in the Wichita Mountains Yellowstone Area to study their social structure, reproductive behaviour and to compute the effective population size. As already mentioned, male bison are more or less successful depending on their dominance status. While the great sand-grouse define their dominance status via the rank of the dance position, the male bison is dominant if he is heavier or older than other bison. [11] Again important parameters like the mature time was included in the calculations to get the effective population size. Researchers determined that the majority of cows had their first calf at age 3 years; but 13 percent had their first calf at age 2 years. [19] To compute the effective population size the parameters were chosen in the following way: For the male and female mature time 3 years were chosen since the majority of the animals had their first calf at this age. The mean of the average age was also the same for males and females. They chose $T_m = T_f = 7$ for both sexes. The sex ratio was $r = 0.5$, i.e. same amount of males and females and the proportion of breeding females respectively dominant males was $\alpha_f = 1$ and $\alpha_m = 0.6$. This means that all of the females was able to breed but only less that two third of the males had dominant features. As already mentioned the dominance characteristic in bison is age and weight. This means that approximately 60% of the males were significantly heavier or older than other bison such that they were dominant and the others were not. These values lead to an $\frac{N_e}{N}$-ratio of 0.72. [17] The effective population size has been computed by observing a special bison population and determining the relevant parameters. In the next chapter an example for a polyandry population will be presented.

6.4. Polyandry in birds: sandpiper

Because of the already described differences in male and female gametes, polygyny occurs more often than polyandry. Therefore, it is a special case when polyandry exists in bird species. One example of birds where polyandry can be observed is the sandpiper. The sandpiper is located in Nord America. The animals have a size of 12 to 60 cm and the females have the ability to produce eggs very rapidly. [5] This is the reason why it works that the female birds mate with different males. There are also female sandpipers that live with only one male (monogamous). Since the females can produce eggs very quickly it makes sense for them to mate with different males such that they can act as breeder and educator. The female therefore is the owner of different nests with different males. After mating with one the of males and laying the eggs, the male breeds and feeds the progenies. Sometimes the female is responsible for one nest but often each nest is provided by one of the males. The female is also responsible for the defense of the nests and her territory. She lays eggs for up to four different males at one time.

6.5. Polygamy and Monogamy in one species: weaver birds

Another example for interesting mating behaviour in one species is the weaver bird. This species does not only live in monogamy but also in polygamous constellations. The way how they mate and breed depend on the place where they breed. Weaver birds are seasonal breeders and breed during June to August and they are home in Africa and Asia. [5]
A special feature about this bird is the nest type they build.

Figure 14: **Weaver bird nest**

The mating system of the weaver birds depends on their ecological situation. If the weaver birds live in an environment in which the animals increasingly feed on fruits due to the flora and fauna, it is advantageous for the birds to join together in polygamous "packs". The fruits usually grow in large quantities at one place. For this reason it is advantageous if the birds defend the feeding grounds together by forming a larger group. The fact that many individuals (all members of the group) eat from this one place is not a problem, as there is enough fruit at one place such that all group members get enough. However, if the weaver bird's territory is a place where many grains grow, they usually live in monogamous constellations. Since grains can be found scattered over a larger area, it would not make sense to defend the feeding places, since there is not one large feeding place that is worth to be defended. Therefore, it is an advantage if the monogamous couple is on their own, looking for food when needed without having to consider other group members. Finally, this leads to different results: While in the polygamous situation the males are larger than the females because they have to assert themselves against other males in order to convince the females of their "pack", the males in the monogamous situation look similar to the females.

7 Summary

Mating system, sex ratio, survival probability, and fecundity strongly influence effective population size. However, the exact extent of these relations were unknown until the second half of the last century. In 1972, Hill established an effective population size model for overlapping generations. Accordingly, Nunney [16] applied it to various mating systems such that the effective population size could be calculated more precisely and directly. First of all, with reference to that, this thesis discussed this model with respect to its components in order to investigate the influence of single parameters and special cases.

Secondly, sex differences resulting from recruitment and from differences in survival lead to different behaviors of the effective population size. Moreover, it was shown that the three different models are the same if it holds that the harem size is the reciprocal value of the number of progenies per female, all females breed and all males are dominant.

Lastly, the lottery polygyny and the harem polygyny differ strongly with respect to sex ratio. Furthermore, an unequal sex ratio affects a harem polygyny less than a lottery polygyny, since by definition a harem consists of only one male and many females. Additionally, in dominance and harem polygyny the effective population size have similar values for $r_h = 1 - r_d$, where r_h and r_d are the sex ratios for the harem and the dominance polygyny, respectively. In contrast, due to the strong dependence on the female survival rate in the dominance polygyny and its negliglible influence in the lottery polygyny the effective population size difference is quite distinct. On the one hand, the number of progeny per female, n, and the male survival probability, v_m, have a strong influence on the dominance effective population size. On the other hand, in lottery and harem polygyny the N_e-model is strongly influenced by the sex ratio r. The survival probabilities have, under certain assumptions, nearly no impact on the effective population size.

In conclusion, investigating the influences of different parameters leads to a better understanding of internal structures of a population. As the examples in the last chapter show, animal species seem to "adapt" the mating systems that best fit their morphological, ecological and behavioral traits.

References

[1] Joel Berger. "Reproductive fates of dispersers in a harem-dwelling ungulate: the wild horse". In: *Mammalian dispersal patterns* (1987), pp. 41–54.

[2] James F. Crow et al. "Breeding structure of populations. II. Effective population number". In: *Statistics and mathematics in biology* 543 (1954), p. 556.

[3] James F. Crow and Carter Denniston. "Inbreeding and variance effective population numbers". In: *Evolution* 42.3 (1988), pp. 482–495.

[4] James F. Crow and Motoo Kimura. "The effective number of a population with overlapping generations: a correction and further discussion." In: *American Journal of Human Genetics* 24.1 (1972), p. 1.

[5] Nicholas B. Davies, John R. Krebs, and Stuart A. West. *An introduction to behavioural ecology*. John Wiley & Sons, 2012.

[6] Larissa Rosa De Oliveira et al. "Effective population size in a bottlenecked fur seal population". In: *Biological Conservation* 131.4 (2006), pp. 505–509.

[7] Ted H. Emigh. "The dynamics of finite haploid populations with overlapping generations. I. Moments, fixation probabilities and stationary distributions". In: *Genetics* 92.1 (1979), pp. 323–337.

[8] Ronald A. Fisher. "XVII.—The distribution of gene ratios for rare mutations". In: *Proceedings of the Royal Society of Edinburgh* 50 (1931), pp. 204–219.

[9] Ronald A. Fisher. "XXI.—On the dominance ratio". In: *Proceedings of the royal society of Edinburgh* 42 (1923), pp. 321–341.

[10] F. Galimberti, L. Boitani, and I. Marzetti. "The frequency and costs of harassment in southern elephant seals". In: *Ethology Ecology & Evolution* 12.4 (2000), pp. 345–365.

[11] C. Cormack Gates et al. *American bison: status survey and conservation guidelines 2010*. IUCN, 2010.

[12] William G. Hill. "A note on effective population size with overlapping generations". In: *Genetics* 92.1 (1979), pp. 317–322.

[13] William G. Hill. "Effective size of populations with overlapping generations". In: *Theoretical population biology* 3.3 (1972), pp. 278–289.

[14] Rus Hoelzel et al. "Alpha-male paternity in elephant seals". In: *Behavioral Ecology and Sociobiology* 5 (1999), pp. 298–306.

[15] Leonard Nunney. "The influence of age structure and fecundity on effective population size". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 246.1315 (1991), pp. 71–76.

[16] Leonard Nunney. "The influence of mating system and overlapping generations on effective population size". In: *Evolution* 47.5 (1993), pp. 1329–1341.

[17] Leonard Nunney and Diane R. Elam. “Estimating the effective population size of conserved populations”. In: *Conservation biology* 8.1 (1994), pp. 175–184.

[18] Orlando A. Schwartz, Kenneth B Armitage, and Dirk Van Vuren. “A 32-year demography of yellow-bellied marmots (Marmota flaviventris)”. In: *Journal of Zoology* 246.3 (1998), pp. 337–346.

[19] Alisa M. Shull and Alan R. Tipton. “Effective population size of bison on the Wichita Mountains Wildlife Refuge”. In: *Conservation Biology* 1.1 (1987), pp. 35–41.

[20] P. Sjödin et al. “On the meaning and existence of an effective population size”. In: *Genetics* 169.2 (2005), pp. 1061–1070.

[21] Sewall Wright. “Evolution in Mendelian populations”. In: *Genetics* 16.2 (1931), p. 97.