Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Ingri Benedicte Rønning

A new latent-variable statistical method for estimating the cost of phenotypic plasticity

Master's thesis in Applied Physics and Mathematics
Supervisor: Jarle Tufto
July 2020

NTNU
Norwegian University of Science and Technology

Abstract

A new latent-variable statistical method for detecting the cost of plasticity in a population has been tested using simulations. Individual fitness was modeled by a Poisson distribution with expectation according to the selective mechanisms in the relevant population. The simulated populations were under a phenotypic stabilizing selection and a cost of plasticity. The individual breeding values and plasticity values were latent variables in the model. Thus, the model inferred them by applying the result of the infinitesimal model to the pedigree chart of the current population. The objective was to find the conditions under which the model attains its highest statistical power for detecting the cost of plasticity in the population. This was done by varying the family structure of the populations, the model for the environmental contribution to the phenotype, and the level of plasticity variance. Results were obtained using maximum likelihood estimation on the marginal likelihood function in R software with the open R-package **TMB**, and using the asymptotic normality of the ML estimators in a power analysis.

The results revealed that, depending on the exact assumptions on the effect of the cost of plasticity, there is an optimal family structure which maximizes the asymptotic power of each model. Two different models for the environmental contribution to the phenotype were applied to the simulated populations. The environmental model which split each family of the population into two distinct groups was the superior choice in the models which assumed that the cost of plasticity had both a linear and quadratic effect. The other environmental model drew environmental contributions independently from a standard normal distribution. In the statistical selection models which considered either an exclusive linear or an exclusive quadratic effect of the cost of plasticity, the environmental models performed equally well. The results also confirmed that, in a population where the plasticity variance is at a higher level, the potential cost of plasticity is more detectable.

Sammendrag

En ny latent variabel-metode for å oppdage kostnaden av plastisitet i en population har blitt testet ved hjelp av simuleringer. Individuell fitness ble modellert av en Poisson-fordeling med forventningsverdi i henhold til de selektive mekanismene i den gjeldende populasjonen. De simulerte populasjonene var under fenotypisk stabiliserende seleksjon og en kostnad av plastisitet. De individuelle avlsverdiene og plastisitetsverdiene var latente variabler i modellen. Modellen måtte derfor inferere dem ved å anvende den infinitesimale modellen på stamtavlen over familiestrukturene til den gjeldende populasjonen. Målet var å finne forholdene der modellen oppnår høyest statistisk styrke for å oppdage kostnaden av plastisitet i populasjonen. Dette ble gjort ved å variere familiestrukturen i populasjonene, modellen for miljøpåvirkningen til fenotypen, og varians av plastisitet. Resultater ble til ved å bruke 'Maximum likelihood estimation' på den marginale rimelighetsunkfjonen med `R` sotftware og den åpne `R`-pakken **TMB**, og ved å benytte den asymptotiske normalfordelingen til ML-estimatorene i en analyse av statistisk teststyrke.

Resultatene avslørte at, avhenig av antagelsene på effekten av kostanden av plastisitet, finnes det en optimal familiestruktur som maksimerer den asymptotiske styrken til hver modell. To ulike modeller for miljøpåvirkningen på fenotypen ble anvendt på de simulerte populasjonene. Miljømodellen som deler hver familie i populasjonen inn i to distinkte grupper var det overlegne valget i modellene som antok at kostnaden av plastisitet både hadde en lineær og kvadratisk effekt. Den andre miljømodellen samplet miljøbidragene uavhengig fra en standard normalfordeling. I de statistiske seleksjonsmodellene som antok enten en utelukkende lineær eller utelukkende kvadratisk effekt av kostnaden av plastisitet, fungerte begge miljømodellene like bra. Resultatene bekreftet også at kostnaden av plastisitet er lettere å oppdage i en populasjon med høyere varians i plastisitet.

Preface

This master thesis concludes my five years as a student enrolled in the *Physics and Mathematics* program with specialization in *Industrial mathematics* at the Norwegian University of Science and Technology (NTNU) in Trondheim.

The thesis is a contribution to the field of quantitative genetics, a field which was first introduced to me during the writing of my specialization project (Rønning 2020). This thesis continues my exploration in this interesting and relevant branch. I have learned many new things and gotten new insights while working on these projects, which has been both challenging and rewarding.

I want to thank my supervisor, Jarle Tufto, Professor at the Department of Mathematical Sciences at NTNU, for introducing me to a new field and coming up with an interesting topic for this thesis, and, for helping me underway. I also want to thank my classmates for all the great memories they have given me over the past five years, and for the support I received during the writing of my thesis.

Ingri Benedicte Rønning
Trondheim, July 2020

Contents

Chapter 1

Introduction

In 1859, Charles Darwin introduced the theory of natural selection, which states that some individuals survive longer and produce more offspring than others due to differences in phenotype. The measurement of the survival and reproduction skills of an individual is known as its biological fitness, or *Darwinian* fitness. A selection is fluctuating if there is variation in the optimal trait values of the population over a relatively brief period of time. The fluctuations mainly occur as a result of environmental changes, and, consequently, populations living in heterogeneous environments are subject to fluctuating selection. Species living in heterogeneous environments are thereby dependent on mechanisms which allow them to adapt to the fluctuating conditions they are experiencing. Changes in gene frequencies by natural selection is an adaptation in itself, however, it is extremely time-demanding. Today, many species are experiencing environmental change at a critical rate which may exceed the capacity of genetic change of the population, and they must rely on other adaptation methods. Another relevant evolutionary response to fluctuating selection is phenotypic plasticity (Chevin *et al.* 2010; King and Hadfield 2018), which means the ability of a single individual to express different phenotypes in different environments.

Phenotypic plasticity provides an individual with the capacity to change its phenotype as a direct response to the fluctuations it is subject to, such that the plastic individual can adapt to the changes of its heterogeneous surroundings. Adaptive plasticity can be detrimental for individual survival, and, can thus also be significant to the survival of entire species. At first glance, it may appear that plasticity is exclusively advantageous. One could thereby expect that evolution by natural selection caused an increase of phenotypic plasticity in all individuals, such that they could exhibit the optimal phenotype at all times. However, empirical studies suggest constraints on the evolution towards higher plasticity. One plausible constraint is that the exhibition of plasticity poses a cost to fitness. A cost of plasticity must be defined as a direct reduction to individual fitness as a consequence of exhibiting plasticity beyond the direct fitness effect of the phenotype (Dewitt *et al.* 1998).

Unfortunately, no matter the source of the cost of plasticity, it is not directly

measurable, and there have therefore been several attempts at estimating the cost of plasticity by several scientists (Dewitt *et al.* 1998; Scheiner and Berrigan 1998; Donohue *et al.* 2000; Agrawal *et al.* 2002; Relyea 2002). The results indicate that costs of plasticity are not common. This is however not necessarily true, as the applied methods might be infested with hidden bias. These attempts have mainly considered models for biological fitness in a population and included a plasticity covariate which is assumed to have a decreasing effect. However, the plasticity covariates of the models are not actual observations, they are best linear unbiased predictors (BLUPs) which are provided from an external linear mixed model. Best linear unbiased prediction is a method for obtaining point estimates of a random effect in a mixed effect model, and it has been frequently used for estimation of individual breeding values and plasticity values. The statistical model for fitness makes no compensation for the uncertainty of the estimators, which may lead to inaccuracies in the model. An article by Jarrod D. Hadfield (Hadfield *et al.* 2009) highlights the problem of using BLUPs from a linear mixed model as explanatory variables in a separate model. It shows that BLUP often provides anticonservative and biased estimators for parameters of selection. Thus, there is a demand for a selection model which takes the inaccuracy of the unobserveable variables into account.

This thesis presents a novel approach for estimating the cost of plasticity, where individual breeding values and plasticity levels are latent variables in a joint statistical model. The degree of plasticity in individuals is included as an explanatory variable in both the submodel for the observed phenotype and the submodel for the observed selection. Plasticity is not a directly observeable trait, so, instead, the model infers the plasticity values by making assumptions on their distributions. The distribution of plasticity in an individual is determined by its genetic background, which in turn is determined by the family relations of the respective individual. Hence, the model relies on observations of related individuals. It is in particular the family structure of populations and its impact on the performance of the statistical model which is the centre of this thesis.

The model is made with the intention of being applied to a set of observations of an actual population. In this thesis we thus try to uncover the circumstances under which the model works best. So, at a later time, if someone decides to perform an experiment for estimating the cost of plasticity, they can, hopefully, use the results of this thesis as a guidance for the design of their experiment. The results may also be used for assessing the value of a data set of observations of a population which are already made. The thesis evaluates the performance of the model on simulated populations using statistical power analysis. The main objective is to answer the following question:

> *Given a set of N individuals of the same species which are separated into m distinct families, in which each family consists of n full siblings, what are the concurrent optimal values of m and n for uncovering the cost of plasticity in the population?*

Other conditions which are expected to affect the power of the statistical model are also examined. These conditions are: the environmental contributions to individual development, the variance of plasticity in the population, and the strength of the cost of plasticity in terms of selection.

The thesis starts by presenting the relevant theoretical results of quantitative genetics which are directly applied in the statistical model. It then proceeds to introduce necessary computational and statistical theory for assessing the model. Chapter 3 provides a detailed description of the aspects of the model, and explains the procedure of assessing the model using power analysis. In addition, all the different conditions of the populations and selections which will be explored are introduced in this chapter. The results of the power analysis are presented in chapter 4. Finally, chapters 5 and 6 respectively contain the discussion and conclusion of the thesis.

Chapter 2

Background theory

The purpose of this chapter is to present the applied theory which justifies the method of the thesis. This includes theory which is specific to the field of quantitative genetics, and statistical theory which is relevant to analysis of statistical power.

Section 2.1 presents the infinitesimal model, which accounts for the distributional assumptions on the individuals of the simulated populations. Section 2.2 then goes on to explain the mechanisms of selection in detail, and the impacts of a stabilizing phenotypic selection on a population. The cost of plasticity and the possible sources for it are accounted for in 2.3. Section 2.4 describes the process of performing maximum likelihood on a latent-variable model using computational software tools. Finally, 2.5 derives the asymptotic distribution of maximum likelihood estimators, which will be applied in the power analysis.

2.1 The infinitesimal model

The infinitesimal model, introduced for the first time by Ronald Fisher (Fisher 1918), is a statistical model for the inheritance of quantitative traits. The individual breeding value with respect to a trait is the expected trait value of its offspring, measured as the deviation from the population mean (Etterson and Kliman 2016). Under the conditions of the infinitesimal model, the breeding value of an individual is approximately normally distributed around the mean breeding value of its parents with a variance which remains constant in the population despite selection (Barton *et al.* 2017).

Assume a large population of N individuals that is outcrossing, which entails that only individuals with no common ancestors can mate. The population is under linkage equilibrium and Hardy-Weinberg equilibrium, such that the allele frequencies at two or more loci are independent and the population-wide genetic variation remains constant through generations. Assume that we are considering m traits on each individual of the population. For $k = 1, \ldots, N$, let the vector of breeding values of individual k be denoted by $\mathbf{x}_k$, $\mathbf{x}_k \in \mathbf{R}^m$. Let n denote the number of loci of each individual of the population, and, for all $i = 1, \ldots, n$, denote the additive contribution of locus i in individual k to its breeding value by $\mathbf{x}_k^{(i)}$. Then, $\mathbf{x}_k$ is defined as

$$\mathbf{x}_k = \mathbf{x}_k^{(1)} + \mathbf{x}_k^{(2)} + \ldots + \mathbf{x}_k^{(n)}.$$

Denote the average parental breeding value of individual k by $\bar{\mathbf{x}}_k$. Now, let n tend towards infinity while, for each $i = 1, \ldots, n$, $\mathbf{x}_k^{(i)}$ tends towards the zero vector. In this limit, the breeding value has the following conditional probability distribution;

$$\mathbf{x}_k \mid \bar{\mathbf{x}}_k \sim N\left(\bar{\mathbf{x}}_k, \frac{1}{2}V\right), \qquad \text{for } k = 1, \ldots, N,$$

where V is the genetic variance-covariance matrix. The genetic variance-covariance matrix is the sum of all the variance-covariance matrices of the additive effects of each loci (Barton *et al.* 2017). A proof which reaffirms the applicability of the approximations of the infinitesimal model is given in the same article by Barton. It relies on some restrictive model conditions, and, applying the central limit theorem to the breeding values for achieving normality.

The infinitesimal model is only applicable under certain conditions. The number of loci must be sufficiently large, and each contribution must be sufficiently small for the infinity-infinitesimal approximation to hold. It is necessary that the population is large and outcrossing, because inbreeding decreases genetic variance, and the population must be sufficiently large to remain outcrossing and avoid inbreeding. Several common genetic phenomena are not accounted for by the model applied here, such as mutation and dominance. The effect of these must thus be sufficiently small.

An illustration of the distributions of breeding values for $m = 1$ trait in a population under the infinitesimal model is given in figure 2.1. Each breeding

value is normally distributed with the same variance, while the mean values differ between individuals.

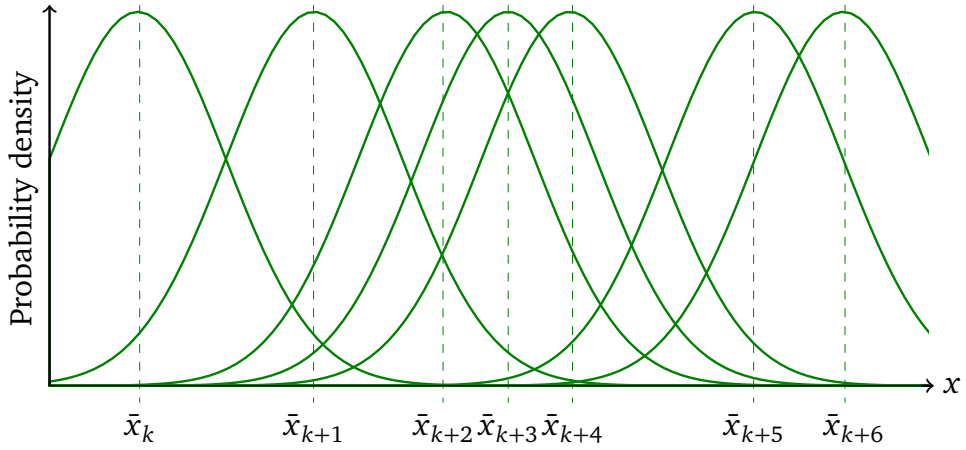


Figure 2.1: Probability density functions of the distributions of the breeding values of individuals in a population under the infinitesimal model.

2.2 Selection and fitness

Selection is the differential survival and reproduction skills of individuals caused by trait differences. In a population under the influence of selection, selected individuals survive longer and produce more offspring than other individuals. The population will then experience an accompanied change in gene frequencies. The biological fitness of an individual is given by the genetic contribution of the individual to the next generation of the population in proportion to the contributions of other individuals in the same generation of the population. It is thus a measure of an individual's reproductive success.

Assume that there is selection acting on a phenotype in a given population. This implies that there is at least one optimal trait value, and that individuals with trait value equal to this optimum have enhanced fitness. The selection is fluctuating if this optimal trait value is variant in the population, which means that different individuals may experience different optimal values due to environmental fluctuations (Haldane and Jayakar 1963). In practice, all natural selection is fluctuating, but they fluctuate in different proportions. The changes of the optimum with the environment is called the environmental sensitivity of selection, and this feature of the selection determines how fluctuating it is. Low environmental sensitivity of selection on a population indicates that there is little spatial variation in the population habitat, and consequently the different individuals experience selection towards trait values of little diversification. A non-fluctuating selection is called a static selection, and means that the optimal trait value is constant in time and space.

Stabilizing selection is the process in which non-extreme trait values are selected above extreme values. Non-extreme means that they are close to the average trait value of the population, and thus, individuals which are closer to average are selected above other individuals. Stabilizing selection on a trait will eventually decrease the trait variance in the population, as illustrated in figure 2.2. It can be modeled by letting fitness be a Gaussian function centered around the non-extreme optimum (Bull 1987). Let the fitness and trait value of an individual be denoted by w and z, respectively. Gaussian stabilizing selection on the trait z towards its optimum θ is given by

$$w(z) = e^{-\frac{1}{2}\left(\frac{z-\theta}{\omega}\right)^2}, \tag{2.1}$$

where ω is the width of the fitness function on the trait. The strength of a Gaussian stabilizing selection is determined by the width of the fitness function and the variance of the selected-upon trait. The narrower the fitness function and the larger the trait variance, the stronger the selection.

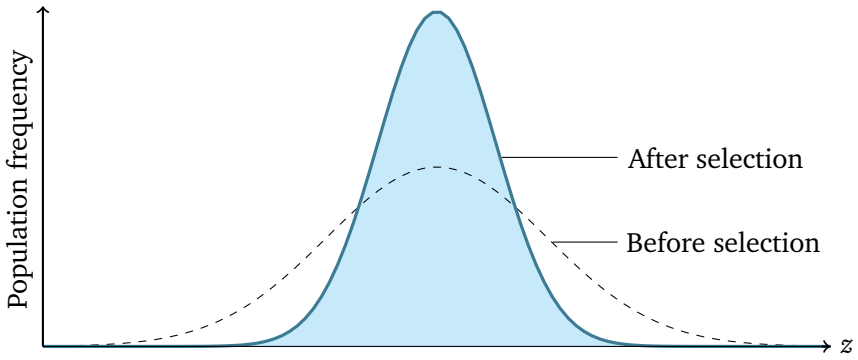


Figure 2.2: The distribution of trait values of a given quantitative trait z in a population before and after a stabilizing selection on z.

2.3 The cost of plasticity

Phenotypic plasticity quantifies the interaction between the phenotype and the surrounding environment, and it is a genetically determined component. Let individual phenotypic plasticity be denoted by b until further. A cost of plasticity is a direct reduction in biological fitness caused by the act of exhibiting non-optimal plasticity. It is thus a direct relation between plasticity and fitness that acts independently of the phenotypic selection which persists in the population (DeWitt and Scheiner 2003). This implies that there is an independent selection on the plasticity trait. This can be modeled by an additional fitness function which depends solely on plasticity, such that the fitness function of an individual which is under a phenotypic selection and a cost of plasticity is equal to the product of the two fitness functions of the independent selections. This has, for example, been performed by King and Hadfield (King and Hadfield 2018), and was explored in the final chapter of my specialization project (Rønning 2020).

Some concepts of selection were introduced in the previous section, and they apply here as well. An independent selection in b implies that there must be at least one optimum which maximizes the fitness function on plasticity. Optima in this sense means the plasticity values which are least costly. The value of the optima and the formulation of the fitness function of the selection on b depend on the mechanism of the cost. An article by Thomas J. DeWitt lists five plausible reasons for a cost — maintenance costs, production costs, information acquisition costs, developmental instability and genetic costs (Dewitt *et al.* 1998). The definition of each cost is taken from the same article. The effects of the different types of costs were discussed in my specialization project (Rønning 2020).

A maintenance cost of plasticity is the energy consumption and consequent fitness reduction which comes from constantly perceiving the environment and regulating the plastic phenotype accordingly. All degrees of plasticity exhibition demand maintenance costs, so it follows that the optimum would be positioned at 0 if a maintenance cost is the only active cost.

Production costs are the costs from producing the phenotype, and it is not specific to plastic phenotypes. Production costs constitute a cost to plasticity only when the cost of production is greater for a plastic phenotype than for a non-plastic phenotype under the exact same conditions. The opposite — that plasticity to a certain extent is the least costly alternative — is also possible. Thus, with respect to production costs, there might be several, non-zero least costly degrees of plasticity.

Information acquisition costs include costs which are forced upon the individual from acquiring information about its surroundings. These include, among others, energetic costs, risks of being eaten by predators and the absence of fitness-enhancing activities, for example mating. The presence of information acquisition costs pleads an optimal value of 0 to the fitness function on plasticity.

Developmental instability means high phenotypic variance, which may pose a cost and lower fitness. This occurs if plastic development (the evolution of b)

is more variable than fixed development, such as the evolution of the breeding value. These costs are however related to genetic variance, and they don't imply that there is an independent fitness function in b, or a clear optimum of such a function.

Genetic phenomena can also lead to a cost of plasticity. If, for example. the is population is under linkage disequilibrium, and a plasticity allele is linked to some alleles which lower fitness, the exhibition of plasticity directly lowers fitness. Epistatic effects between different loci, for example suppression of each other gene effects, may also reduce fitness. Genetic costs may arise in a number of ways, and can imply several optimal values of plasticity.

2.4 Maximum Likelihood Evaluation of a latent variable model with TMB

A latent variable model is a model that makes statistical assumptions on a set of variables, a subset of which are latent variables, which means they can not be measured directly. It is assumed that the latent variables determine the dependency of all the observations, and that the observeable variables of the model have nothing in common after removing the effect of the latent variables. Some examples of latent variable models are linear mixed models (LMMs), generalized linear mixed models (GLMMs) and state space models (SSMs). LMMs and GLMMs are 'Random effects'-models, in which the latent variables serve as random effects, which model the heterogeneity between different groups of the data. In SSMs, the latent variables represent a hidden state, and the objective is to compute the optimal estimate of the hidden state given the observed data.

Assume that we have a latent variable model. Because of the latent variables, some crucial data points are missing from the respective likelihood function. Hence, the likelihood function must be marginalized on the observeable variables. This means integrating the likelihood function with respect to the latent variables, such that the marginal likelihood becomes a function of the observeable variables and the model parameters. Then, we can apply maximum likelihood estimation (MLE) to the model and obtain estimators for the model parameters. Let the data (observations of the observeable variables) of the model be denoted by $\mathbf{y}$, let $\mathbf{u}$ denote the latent variables and $\boldsymbol{\theta}$ be the model parameters. The marginal likelihood of the model, denoted by $L(\boldsymbol{\theta})$, is then given by the marginal distribution of the joint probability distribution of the model,

$$L(\boldsymbol{\theta}) = f(\mathbf{y}) = \int f(\mathbf{y} \mid \mathbf{u}) f(\mathbf{u}) \, d\mathbf{u}, \tag{2.2}$$

where $f(\mathbf{y})$ is the marginal probability density function (pdf) of the data, $f(\mathbf{y} \mid \mathbf{u})$ is the conditional pdf of the data, and $f(\mathbf{u})$ is the pdf of the latent variables. This equation is often without an analytic solution, and the solution is thus usually approximated by some integration approximation method. For example, Markov Chain Monte Carlo methods can be applied to obtain a numerical approximation, or we can use other numerical integration methods. Or, if the integrand of (2.2) is written as an exponential, the Laplace approximation can be applied.

TMB (Template Model Builder) is an open `R` package that uses the Laplace approximation to obtain the marginal likelihood of latent variable models. The approximation is obtained using automatic differentiation, which is a method for computing the derivatives of a computer-implemented function. The use of the Laplace approximation and automatic differentiation will be further explained in the next two subsections. Applying MLE to a latent variable model using **TMB**

requires a R-file and a C++-file, and is done as follows. The joint negative log-likelihood of a model is defined as a C++ template function in a C++ file. The associated R-file sends the required data and initial parameter values to the C++-file, which, by means of the Laplace approximation and automatic differentiation, evaluates the negative marginal log-likelihood, the score vector and the Hessian matrix of the model, and returns it to the R-file. All further work, such as minimization of the negative log-likelihood and further analysis, is performed by the R-file. An overview chart of the process is given in figure 2.3.

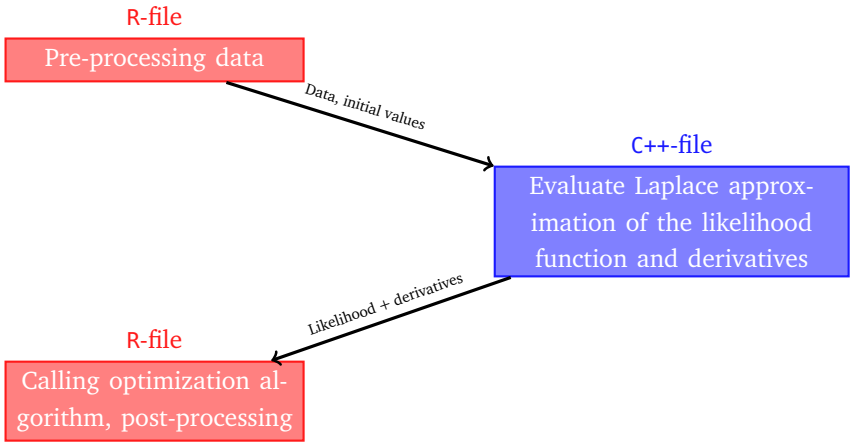


Figure 2.3: The process of performing MLE on a latent variable model using **TMB**.

2.4.1 The Laplace approximation of the marginal likelihood

For a latent variable model with data $\mathbf{y}$, latent variables $\mathbf{u} \in \mathbb{R}^n$ and parameters $\boldsymbol{\theta} \in \mathbb{R}^m$, where $n, m \in \mathbb{N}$, let $f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})$ be the negative joint log-likelihood of the model. The marginal likelihood function of the model is given by

$$L(\boldsymbol{\theta}) = \int_{\mathbb{R}^n} \exp(-f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}))\, d\mathbf{u}. \tag{2.3}$$

The Laplace approximation, introduced by the mathematician Pierre-Simon Laplace (Laplace 1774), is a method for estimating the solution to an integral of an exponential. The method is used by **TMB** to obtain the approximated likelihood function of the model.

The following derivation is taken from two articles (Fog 2008; Kristensen *et al.* 2016). Let $\hat{\mathbf{u}}(\boldsymbol{\theta})$ be the minimizer of $f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})$ with respect to $\mathbf{u}$, such that the score vector evaluated at $\hat{\mathbf{u}}$ is the zero vector,

$$\hat{\mathbf{u}}(\boldsymbol{\theta}) = \arg\min_{\mathbf{u}} f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}).$$

Denote by $H(\boldsymbol{\theta})$ the Hessian of $f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta})$ with respect to $\mathbf{u}$ and evaluated at $\hat{\mathbf{u}}(\boldsymbol{\theta})$,

$$H(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \mathbf{u} \partial \mathbf{u}^T} f(\mathbf{y}, \mathbf{u}; \boldsymbol{\theta}) \Big|_{\mathbf{u}=\hat{\mathbf{u}}(\boldsymbol{\theta})}.$$

Applying the Taylor expansion around $\hat{\mathbf{u}}$ to f yields an approximation given by

$$f(\mathbf{y},\mathbf{u};\boldsymbol{\theta}) \approx f(\mathbf{y},\hat{\mathbf{u}};\boldsymbol{\theta}) + \frac{1}{2}H(\boldsymbol{\theta})(\mathbf{u}-\hat{\mathbf{u}})^2. \tag{2.4}$$

The Laplace approximation to the marginal likelihood function in (2.3) is then obtained by replacing the function f with the Taylor approximation of f, as given in (2.4). Then, a Gaussian integral arises, which is solved analytically using substitution in two variables, and exploiting that the integrand is proportional to the pdf of the Gaussian distribution. The approximation is

$$\begin{aligned} L^*(\boldsymbol{\theta}) &= \int_{\mathbb{R}^n} \exp\left(f(\mathbf{y},\hat{\mathbf{u}};\boldsymbol{\theta}) + \frac{1}{2}H(\boldsymbol{\theta})(\mathbf{u}-\hat{\mathbf{u}})^2 \right) d\mathbf{u}, \\ &= \exp(f(\mathbf{y},\hat{\mathbf{u}};\boldsymbol{\theta})) \int_{\mathbb{R}^n} \exp\left(\frac{1}{2}H(\boldsymbol{\theta})(\mathbf{u}-\hat{\mathbf{u}})^2 \right) d\mathbf{u}, \\ &= \sqrt{2\pi}^n \det(H(\boldsymbol{\theta}))^{-\frac{1}{2}} \exp(-f(\mathbf{y},\hat{\mathbf{u}};\boldsymbol{\theta})). \end{aligned}$$

Now we have obtained a formulation for an approximation of the marginal likelihood of the model. The ML estimator for the parameter vector, denoted by $\hat{\boldsymbol{\theta}}$, is finally found by minimizing the negative log of the Laplace approximation to the likelihood function,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left(-\log L^*(\boldsymbol{\theta}) \right) = \arg\min_{\boldsymbol{\theta}} \left(-n\log\sqrt{2\pi} + \frac{1}{2}\log\det(H(\boldsymbol{\theta})) + f(\mathbf{y},\hat{\mathbf{u}};\boldsymbol{\theta}) \right).$$

Observe that, obtaining the final solution poses a nested optimization problem; the minimizer of f with respect to the latent variables, $\hat{\mathbf{u}}$, is required for minimizing the negative log of $L^*(\boldsymbol{\theta})$. So, **TMB** must solve the inner optimization problem of minimizing f before solving the outer optimization problem and obtaining the ML estimator of the parameters. Solving this nested optimization problem requires evaluation of many partial derivatives, and this is performed by **TMB** using automatic differentiation.

2.4.2 Optimization using automatic differentiation

Automatic differentiation (AD) is a technique that enables the evaluation of derivatives of any objective function defined by an algorithm such that no user-defined derivative code is required. **TMB** uses AD to obtain the Hessian matrix of the likelihood and to solve the nested optimization problem associated with MLE of the Laplace approximation.

When a computer program uses the AD technique for differentiating a function, it starts by decomposing the computational sequence of the user-defined objective function into elements. Every objective function defined by an algorithm is a sequence of elementary operations, such as multiplication or addition, and, elementary functions, such as the exponential function or trigonometric functions.

By constantly applying the chain rule and differentiating with respect to a single elementary operation or a single elementary function at a time, the function is finally decomposed into partial derivatives (Kristensen *et al.* 2016). After obtaining the first derivative of an objective function, the program can proceed to calculate the second derivative of the objective function by the same means.

A clear advantage of AD is that the user doesn't have to define the derivatives of the objective function analytically. This is however not a unique quality, as it also holds for symbolic differentiation and numerical differentiation. Symbolical differentiation also applies the chain rule to decompose an expression, however, it manipulates mathematical expressions to obtain a formulation of the derivative of those expressions. AD has the advantage that it manipulates a computer algorithm to obtain numerical values. So, as opposed to AD, symbolic differentiation can lead to complicated expressions and thereby produce inefficient calculations. Numerical differentiation means evaluating the objective function at many places to measure the slope of the function without using the formulation of the objective function, merely sampled values of it. This method is very sensitive to the amount of data needed. A function of many dimensions may require a lot of computation and numerical differentiation may thus be quite inefficient. Thus, AD is preferable to both symbolic differentiation and to numerical differentiation (Neidinger 2010).

2.5 Asymptotic distribution of Maximum Likelihood Estimators

The purpose of this section is to give a brief review on the derivation of the asymptotic distribution of maximum likelihood estimators.

The derivation is taken from a book by Simon Wood (Wood 2015). Assume that a statistical model of data from a continuous probability distribution with parameter vector $\boldsymbol{\theta}$ is given. The log-likelihood function is denoted by l, and the accompanied maximum likelihood estimator for $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$. Let the true parameter value be denoted by $\boldsymbol{\theta}_t$. First, we present some necessary results regarding the log-likelihood function. Under certain restrictions, the following holds,

$$E\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}\right) = 0. \tag{2.5}$$

The proof of (2.5) rests upon the definition of the expected value, and the Leibniz integral rule for interchanging the differentiation and integration operators, so the pdf of the data and its first derivative must be continuous (Protter and Morrey 1985). By writing out the expression for the variance-covariance matrix of the score, applying (2.5), and, applying the chain rule to the definition of the expected value, we obtain

$$\mathrm{Cov}\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}_t}\right) = E\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}_t}\frac{\partial l}{\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}_t}\right) = -E\left(\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}_t}\right) = \boldsymbol{\mathcal{I}},$$

where $\boldsymbol{\mathcal{I}}$ denotes the Fisher information matrix. It is called the Fisher information because it is a way of measuring the amount of information the data carries about $\boldsymbol{\theta}$, and it is named after the aforementioned statistician, Ronald Fisher, who emphasized the role of this matrix in MLE. Now, apply the Taylor approximation to the score vector around $\boldsymbol{\theta}_t$. This gives

$$\frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \simeq \frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} + \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right). \tag{2.6}$$

By the definition of $\hat{\boldsymbol{\theta}}$, the left hand side must be the zero vector. Denote the sample size (number of observations in the data set) by n. Asymptotic distribution means the distribution of $\hat{\boldsymbol{\theta}}$ in the large sample limit, which is when $n \to \infty$. ML estimators are asymptotically unbiased, so, in this limit, $\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right)$ tends towards $\mathbf{0}$. Assume that $\boldsymbol{\mathcal{I}}$ increases without limit in a way such that $\boldsymbol{\mathcal{I}}/n$ is constant in the large sample limit. Then, as $n \to \infty$,

$$\frac{1}{n}\frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} \to \frac{1}{n}\boldsymbol{\mathcal{I}}, \qquad \text{and,} \qquad \frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} \sim \boldsymbol{\mathcal{I}},$$

which means that the score vector evaluated at the true parameter value is a random vector with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\mathcal{I}}$ in the large sample

limit. Finally, if we apply all the asymtptoic results for $n \to \infty$ to (2.6), we get that

$$\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_t\right) \overset{\text{asymp.}}{\sim} \mathcal{I}^{-1} \left.\frac{\partial l}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}.$$

So, $\hat{\boldsymbol{\theta}}$ has some large-sample-limit distribution with mean $\boldsymbol{\theta}_t$ and variance $\mathcal{I}^{-1}$.

When the observations are independent, the log-likelihood is a sum of independent contributions from each observation. It follows that $\partial l/\partial \boldsymbol{\theta}$ is a sum of independent identically distributed random variables. If the sample size is sufficiently large, the central limit theorem, which states that the sum of n i.i.d. random variables tends towards a normal distribution as n goes to infinity, can be applied to the sum. Then, the asymptotic distribution of the ML estimator is obtained,

$$\hat{\boldsymbol{\theta}} \overset{\text{asymp.}}{\sim} N\left(\boldsymbol{\theta}_t, \mathcal{I}^{-1}\right). \tag{2.7}$$

This distribution can also be achieved when the likelihood is not a sum of independent contributions. As long as $\mathcal{I}$ increases without limit as the sample size increases, such that $\mathcal{I}/n$ remains constant in the $n \to \infty$ limit, (2.7) often holds anyway.

Chapter 3

Methods

This chapter covers all of the methods which are applied in the thesis, and describes the implementation and computation of these methods.

The first section describes all the assumptions of the composite statistical model in detail — including the assumptions on the populations and the selective forces acting on it. Section 3.2 goes on to explain how the data sets onto which the model will be applied are simulated. Then, section 3.2 elucidates how the power analysis is performed on the model, particularly how the asymptotic power of a model of certain assumptions on the cost of plasticity is obtained. Section 3.4 presents the different conditions which will be applied to the model, and explains our expectations. The exact details around these conditions are given in 3.5. Section3.6 justifies the choice of parameter values in the model, and finally, 3.7 describes the computational process for obtaining results.

3.1 Statistical model

The statistical model attempts to explain the relationship between the biological fitnesses, phenotypes and levels of plasticity of individuals in a given population under selection. The observed biological fitness of an individual is the number of offspring produced, and this is the response variable of the statistical model. The observeable variables of the model are the number of offspring and the phenotypes of individuals, while individual breeding values and plasticity values are latent variables. The populations to which the statistical model is applied are assumed to have non-overlapping generations, and for each population, the observations are collected from one single generation.

This statistical model is a latent variable model which differs from (generalized) linear mixed models and state space models. As in mixed models, there are hierarchical structures to all the relevant populations; observations are sampled from individuals within different families. In this model, however, there are latent variables at two levels. Individuals are divided into groups according to relation, and their relation determines the distribution of their latent variables. In addition, parameters of the distribution of the latent variables of individuals are in turn latent variables. The upper-level latent variables can not be considered random effects, so, the hierarchical structure of this particular model is extended from those of LMMs and GLMMs. This hierarchy is also what contradicts an SSM — the hidden states of an SSM are assumed to have observeable distributional parameters, while the distributional parameters are latent variables in this model.

The next subsection describes the hierarchical structure of the distribution of the latent variables in detail. Then, the chosen model for the phenotype is presented in 3.1.2, the selection assumptions are presented in 3.1.3, and, lastly, the likelihood function of the model is given in 3.1.4.

3.1.1 Distribution of the latent variables

The latent variables of a model are those variables that can't be measured directly. In a statistical selection model such as this one, these are the breeding values and plasticity values of individuals. Together, the breeding value and plasticity value of an individual constitute the genotype of the individual, which is the genetic contribution to its phenotype. The relation between the latent variables, the phenotype and the phenotypic selection will be presented in the next two subsections. For now, it is sufficient to know that they are genetically determined components which are decisive on the phenotype, and that they are latent variables in the statistical model. The distributional assumptions on the latent variables are obtained from knowing the family structure of the population and applying the infinitesimal model, as presented in section 2.1. As it is a genetically determined component, the plasticity value is assumed to be subject to the results of the infinitesimal model.

Each considered population has a family structure as follows. Let N be the

population size. The N individuals are split into m distinct families, where each family consists of n full siblings. Two arbitrary individuals of this given generation of the population are thus either completely unrelated or have the same parents. Recall that, under the assumption of the infinitesimal model, the breeding value of an individual is normally distributed around the midparental breeding value, and that the variance is constant in the population. Thus, in this statistical model, the genotypes of the n full siblings of any arbitrary family are i.i.d.. Let i, $i = 1, \dots, m$ be the family index and j, $j = 1, \dots, n$ denote which sibling. Denote the breeding value of sibling j of family i by a_{ij}, and let b_{ij} denote the plasticity level in the same individual, for all $i = 1, \dots, m$, $j = 1, \dots, n$. The vector $[a_{ij}, b_{ij}]^T$ constitutes the genotype of individual ij. The genotype of any sibling of a given family i has conditional distribution

$$\begin{bmatrix} a_{ij} \\ b_{ij} \end{bmatrix} \Bigg| \begin{bmatrix} \bar{a}_i \\ \bar{b}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{a}_i \\ \bar{b}_i \end{bmatrix}, \frac{1}{2} \begin{bmatrix} G_{aa} & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix} \right), \quad \text{for } j = 1, \dots, n, \tag{3.1}$$

where $\bar{a}_i$ and $\bar{b}_i$ are the midparental breeding value and plasticity value, respectively, of the parents of family i, G_{aa} is the population-wide variance of the breeding value, G_{bb} is the population-wide variance of plasticity and G_{ab} is their population-wide covariance.

The parental generations are assumed to be panmictic, which means that reproduction occured without selection and all individuals had equal expected fitness. With consecutive generations of random mating, the population as a whole rapidly converges towards a Gaussian distribution (Barton *et al.* 2017). Thus, the genotypes of the immediate parental generation are normally distributed in the population around some overlying population mean value, denoted by $[\bar{\bar{a}}, \bar{\bar{b}}]^T$,

$$\begin{bmatrix} \bar{a}_i \\ \bar{b}_i \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\bar{a}} \\ \bar{\bar{b}} \end{bmatrix}, \frac{1}{2} \begin{bmatrix} G_{aa} & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix} \right), \quad \text{for } i = 1, \dots, m. \tag{3.2}$$

The genotypes of the parental generation are just as immeasurable as those of the current generation. Hence, the family-specific mean values of the breeding value and plasticity are also latent variables of the model, while the overlying means, $\bar{\bar{a}}$ and $\bar{\bar{b}}$, are parameters. An illustration of the hierarchical distributional assumptions of the latent variables is given in figure 3.1.

3.1.2 The phenotype

Individual phenotypes are dependent on the respective genotypes and on environmental impact. Let z_{ij} denote the phenotype of sibling j of family i of the population, and let its genotype be given by a_{ij} and b_{ij}, for all $i = 1, \dots, m$, $j = 1, \dots, n$. The environmental contribution to the phenotype of individual ij at development is denoted by ε_{ij}. In all models, the unconditional expected value of the environmental contribution is 0. The environmental component is drawn from a distinct model, and is simply a given value to the statistical model. Thus, the model makes

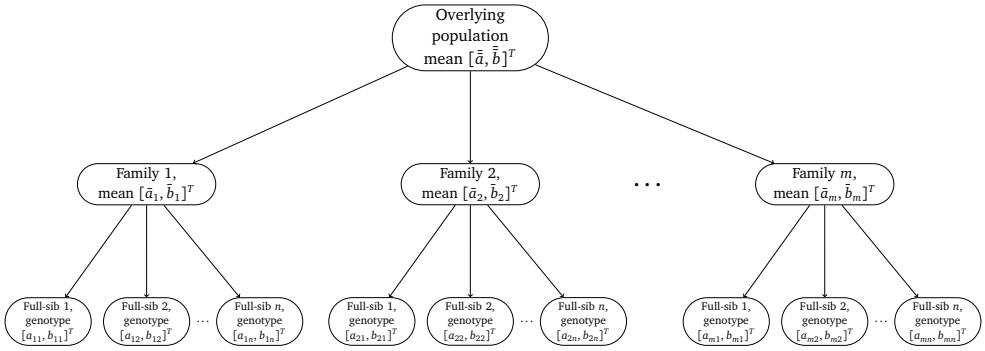


Figure 3.1: Hierarchical structure of the genotypic distributions of the theoretical populations, as given in (3.1) and (3.2). The genotypes of the child nodes are drawn from distributions centered around the values of the respective parental nodes. The lower level represents the observed generation under selection, the middle level represents the parental generation, and the upper node represents all the previous generations.

no attempts at estimating the parameters and formulations of the models of the environmental contributions. Recall that plasticity is the interaction of the phenotype with the environment, while the breeding value constitutes an elevation. The model assumes a linear relationship with the environment. Hence, the phenotype is modeled as

$$z_{ij} = a_{ij} + b_{ij}\varepsilon_{ij} + e_{ij}, \tag{3.3}$$

where $e_{ij} \sim N(0, \sigma_e^2)$. The last term can represent environmental variation which is not captured by ε_{ij} or observational errors, and imposes a lower limit to the variance of the phenotype. Note that the phenotype is a linear combination of normally distributed variables, and has a normal distribution itself. For all individuals $i = 1, \ldots, m$, $j = 1, \ldots, n$, the conditional phenotypic distribution is

$$z_{ij} \mid a_{ij}, b_{ij}, \varepsilon_{ij} \sim N\left(a_{ij} + b_{ij}\varepsilon_{ij},\ \sigma_e^2\right).$$

3.1.3 Selection

The selection which persists in the notional population consists of a stabilizing phenotypic selection and a cost of plasticity. A stabilizing selection on the phenotype is chosen because it is thought to be the most common type of selection in most populations (Charlesworth *et al.* 1982). Selection is assumed to occur simultaneously for all individuals and to be static.

The number of offspring produced by any given individual of the relevant generation is assumed to have a Poisson distribution with an expectation that depends directly on the phenotype and on plasticity. The Poisson distribution is a discrete probability distribution which models the probability of events occurring in an interval, so, given that all individuals of the same generation produce offspring in the same temporal interval, and, that there is selection present such that the distribution parameter (expected number of offspring) is inhomogeneous

in the population, this distribution fits well. The Poisson distribution assumption is inspired by the mating patterns of idealized Wright-Fisher (WF) populations as described by Felsenstein (Felsenstein 2015). Here, the number of offspring of each individual is binomially distributed, and is thus approximately Poisson distributed in the large population limit. An idealised WF population is under the random mating assumption, so the number of offspring of individuals are i.i.d.. In the model presented here, however, the producing generation is under selection, so the expected number of offspring of individuals varies according to their trait- and plasticity values. Additionally, this model poses no constraint on the number of offspring, while idealized WF populations have constant size.

For all $i = 1, \dots, m$, $j = 1, \dots, n$, let y_{ij} denote the number of offspring produced by sibling j of family i, and denote its phenotype and plasticity value by z_{ij} and b_{ij}, respectively. Denote the expected number of offspring of individual ij by w_{ij}. It follows from the selection assumptions that w_{ij} is a function of z_{ij} and b_{ij}. For all $i = 1, \dots, m$, $j = 1, \dots, n$, the number of offspring of individual ij has the following conditional Poisson distribution,

$$y_{ij} \mid z_{ij}, b_{ij} \sim \text{Poisson}\big(w_{ij}(z_{ij}, b_{ij})\big).$$

The parameter of interest is the expected number of offspring of each individual. We will apply the canonical link function, which is the log-link function, such that selection is given by a log-linear Poisson model. This entails that $\log w_{ij}$ is a linear function in the covariates of the model, and imposes a lower bound $0 \leq w_{ij}$. The exact covariates of the model will be presented in the next three paragraphs.

As previously mentioned, the statistical model presented here is intended to be relevant for an authentic biological experiment on the cost of plasticity. This notional experiment is the study of a population under phenotypic stabilizing selection and an alleged cost of plasticity in some controlled habitat. The selection on the phenotype is modeled as a static selection, so it is important that the habitat of the experiment is customized in such a way that the environmental sensitivty of selection is sufficiently low, such that the optimal phenotype is virtually equal for all individuals of the population. The stabilizing phenotypic selection is modeled by a Gaussian function, as presented in section 2.2. Thus, for each individual ij, $i = 1, \dots, m$, $j = 1, \dots, n$, the expected fitness must obey

$$w_{ij} \propto e^{-\frac{1}{2}\left(\frac{z_{ij}-\theta}{\omega}\right)^2},$$

for some population-wide optimum θ and width ω of the fitness function on the phenotype. It follows that $\log w_{ij}$ is linear to both z_{ij} and z_{ij}^2, and thus both z_{ij} and z_{ij}^2 are included as model covariates (provided that θ is non-zero, if $\theta = 0$, then z_{ij} can be omitted). For $\theta \neq 0$, this assumption also imposes an intercept component to the model.

We have established that both the linear and quadratic effect of the phenotype are covariates in the statistical model. The remaining covariates depend on how

the costliness of plasticity unfolds. The cost of plasticity is assumed to arise from one of the causes presented in section 2.3, and it is assumed to be static in the population. It is reasonable to assume that a cost of plasticity which persists in a population is static. Individuals of the same population are virtually equally exposed for developmental instability and a cost of plasticity as a result of some genetic phenomena. They may also have similar maintenance costs, production costs and information acquisition costs.

Now, let's consider the fitness effect of the cost of plasticity on each individual. For example, plasticity could have a strict linear relationship to the logarithmic fitness, which implies that an additional linear plasticity covariate to the model is sufficient. Alternatively, there could be a Gaussian stabilizing selection on plasticity towards 0 due to maintenance costs, as modeled by King and Hadfield (King and Hadfield 2018). Then, the expected fitness w_{ij} of sibling j of family i with phenotype z_{ij} and plasticity b_{ij} is

$$w_{ij} \propto e^{-\frac{1}{2}\left(\frac{z_{ij}-\theta}{\omega}\right)^2 - \frac{1}{2}\left(\frac{b_{ij}}{\omega_b}\right)^2},$$

for all $i = 1, \ldots, m$, $j = 1, \ldots, n$, where ω_b is the population-wide width of the fitness function of plasticity. In this case, only the quadratic effect is significant, such that, in the model for fitness of individual ij, b_{ij}^2 is a covariate of the model while b_{ij} is not. A third option is that there is a Gaussian stabilizing selection on plasticity towards a non-zero least costly degree of plasticity, which would imply both a linear and quadratic dependence. The statistical model accommodates for all aforementioned possibilities of a cost of plasticity (and some others). Both linear and quadratic effects of the cost plasticity are relevant, as well as the intercept. The phenotypic stabilizing selection and the cost of plasticity are assumed to act independently of each other. In other words, an interaction covariate is not included in the model. Phenotypic selection is assumed to be an 'exterior' contributor, as it depends on the environmental conditions, while the cost of plasticity is a self-inflicting mechanism which only depends on an individual's plasticity level. By reparameterizing the previous equations and assumptions, the following model for expected fitness is obtained; for all $i = 1, \ldots, m$, $j = 1, \ldots, n$, sibling j of family i has expected fitness given by

$$\log w_{ij} = \beta_0 + \beta_1 z_{ij} + \beta_2 z_{ij}^2 + \beta_3 b_{ij} + \beta_4 b_{ij}^2, \tag{3.4}$$

where β_0, β_1, β_2, β_3 and β_4 are the unknown model parameters that determine selection in the population.

Note that, as plasticity is detrimental on the phenotype, a change in the plasticity component of individual ij, b_{ij}, inflicts a change to the phenotype z_{ij} for all $i = 1, \ldots, m, j = 1, \ldots, n$ (unless individual environmental contribution at development is valued at zero, but this is not the general case). Hence, alterations of the degree of plasticity will influence fitness whether plasticity is costly or not, by changing the phenotype. However, if either of the two latter coefficients of (3.4), β_3 and β_4, are nonzero, this would imply a direct dependence between the degree

of plasticity and the biological fitness of each individual. This direct relationship is the cost of plasticity. If β_1 and/or β_2 are nonzero, on the other hand, this means that plasticity has an indirect impact on fitness through dependence of the phenotype. This significance does however not imply that there is a cost of plasticity in the population.

3.1.4 The likelihood function

At this point, the relevant assumptions on the selection and the distribution of the relevant populations have been presented. There are a total of 11 unknown parameters. These are the β-values, β_0, β_1, β_2, β_3 and β_4, that determine selection, σ_e^2, which determines the conditional phenotypic variance, and the parameters of the genotypic distribution, $\bar{\bar{a}}$, $\bar{\bar{b}}$, G_{aa}, G_{ab} and G_{bb}. The latent variables are the individual genotypes, a_{ij} and b_{ij}, and the genotypic family means $\bar{a}_i$ and $\bar{b}_i$, for all $j = 1,\dots,n$, $i = 1,\dots,m$. The input data of the model consists of the number of offspring produced by each individual, the phenotype, the given environmental contribution to the phenotype and, lastly, the pedigree chart of the family structure of the population, which determines the distribution of the lower-level latent variables. Let

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^T, \qquad \mathbf{G} = \begin{bmatrix} G_{aa} & G_{ab} \\ G_{ab} & G_{bb} \end{bmatrix}.$$

The marginal log-likelihood function of the model is given by

$$l(\boldsymbol{\beta}, \sigma_e^2, \bar{\bar{a}}, \bar{\bar{b}}, \mathbf{G}) = \sum_{i=1}^{m} \iint \Bigg(\log f(\bar{a}_i, \bar{b}_i) + \sum_{j=1}^{n} \iint \log f(y_{ij}, z_{ij} \mid a_{ij}, b_{ij}) + \log f(a_{ij}, b_{ij} \mid \bar{a}_i, \bar{b}_i)\, da_{ij} db_{ij} \Bigg) d\bar{a}_i d\bar{b}_i,$$

where $f(\bar{a}_i, \bar{b}_i)$ is the pdf of the distribution of the genotypic mean of family i, $f(y_{ij}, z_{ij} \mid a_{ij}, b_{ij})$ is the conditional joint pdf of the number of offspring and the phenotype of sibling j of family i, and $f(a_{ij}, b_{ij} \mid \bar{a}_i, \bar{b}_i)$ is the distribution of the genotype of sibling j of family i, for $i = 1,\dots,m$, $j = 1,\dots,n$.

By inserting the formulations of the pdf's into the expression, we will find that the marginal log-likelihood function has a quite awkward formulation. Obtaining it also requires a lot of integration — integration must be done with respect to two parameters for each population member, and, with respect to two parameters for each family. The objective function can not be computed explicitly, and the ML estimators of the model parameters are thus obtained by using the approximation methods of the **TMB** package as described in section 2.4.

3.2 Data simulation

This section explains how each data set is obtained. Each data set is thought to imitate observations on an actual population. A data set of a single population is realized as follows. First, all the necessary parameter values are set. This includes the population size N, the number of families, m, and the number of siblings, n, as well as $\bar{\bar{a}}$, $\bar{\bar{b}}$, G_{aa}, G_{bb}, G_{ab}, σ_e^2, β_0, β_1, β_2, β_3, and β_4. Then, the m family means are sampled from the distribution of (3.2) and n distinct genotypes, corresponding to the n full siblings, are sampled from the distribution of (3.1) for each of the m families. Now the genotypes of all individuals are obtained. The model of the environmental contribution to the phenotype is decided, and the necessary parameters for this model are set. The contributions are drawn from the respective model, and the e_{ij}-values are drawn from a $N(0, \sigma_e^2)$ distribution. The phenotype is computed, along with expected fitness, and the number of offspring produced by each individual is sampled from a Poisson distribution. Realizations of all the observeable variables in the population are now obtained, and the data set is thus complete.

In a real-life case study, the same exact data set could be obtained in the following way. The initial population consists of $2m$ genetically uncorrelated individuals which are the product of several consecutive generations of random mating. By selective breeding, monogamous mating occurs, such that m *couples* of mating partners are formed, and each of the m couples of parents produce exactly n offspring (without breaking the random mating assumption, so the monogamous parents must be paired at random). Thus, a generation of $N = m \times n$ offspring is produced. These individuals are placed in controlled and measurable environments. The phenotype of each individual of the second generation is measured. Selection occurs, and each of the N individuals produce a certain number of offspring each. The mating patterns of this selection are irrelevant, as the observed fitness is solely determined by total quantity of offspring. It is essential that the different environmental contributions and the family affiliation of each individual is known at all times.

Clearly, the simulated populations have a simplified structure that does not easily appear on its own. In the generation which constitutes the dataset, two individuals picked at random are either full siblings or come from different families, in which case they are unrelated and genetically uncorrelated. Realistically, two individuals of the same population can be genetically correlated in an immense number of ways, and there is always some correlation. (This amount can however be of a negligible size.) Fortunately, the results and knowledge obtained in this thesis are easily generalized to populations of a more intricate family structure. The family structure mainly affects the distribution of the latent variables of individuals of the population, and an intricate family structure can lead to a hierarchy with more levels and more branching than the one presented in figure

3.1. Thus, the likelihood function may get an even more awkward formulation for those populations. Nevertheless, the model is applicable to those populations as well. Additionally, recall that the the model is thought to be applied to case studies, in which artificial selection can be applied to obtain less challenging pedigree charts than those that may persist in nature.

3.3 Model assessment and test power

This section explains the power analysis which will be used for assessing the statistical model on different simulated populations. When the statistical model is applied to the observations, estimators for each of the unknown model parameters are obtained using MLE. The model seeks to establish if there is a cost of plasticity present in the observed population or not, which is done by performing a significance test on the associated model parameters. The significance tests are given by

$$H_0 : \beta_3 = \beta_4 = 0 \qquad \text{vs.} \qquad H_1 : \beta_3 \neq 0 \cup \beta_4 \neq 0, \tag{3.5}$$

and each test has a chosen significance level α. The statements of the alternative hypothesis are united with a union sign as opposed to an intersection, because, finding that only one the statements holds is sufficient for concluding that plasticity is costly. One-dimensional tests on either β_3 and β_4 are also relevant in this thesis. The first subsection of this section derives the asymptotic power of a one-dimensional significance test, while the second derives the asymptotic power of a two-dimensional test.

The power of a hypothesis test is defined as the probability of correctly rejecting H_0, which, with respect to this statisticl model, means the probability of detecting a cost of plasticity when such a cost is present. The test power for the parameters of this model cannot be found explicitly, so, we exploit what we know about ML estimators and hypothesis testing to obtain approximations and assess the models thereafter. We will use the asymptotic normality of ML estimators, as presented in section 2.5. The pdf of the observations of the model are continuous, and observations from different individuals are independent (though not identically dsitributed), so the asymtptoic distribution applies to the parameters of this statistical model.

An alternative approach to obtain the model power is to perform simulation-based analysis (Bolker 2008). This method involves simulating many populations for which H_1 is true and perform tests on all the data sets. Each test is viewed as a Bernoulli trial, where success means the rejection of H_0. The number of rejections out of n trials is then binomially distributed with success probability p equal to the model power and n as parameters. MLE can be performed on the binomial distribution to obtain estimates for the model power. This method has not been chosen because it requires the execution of a large number of tests and is thus much more computer intensive. Also, the outcome of each test is based on the comparison of α to the p-value of the test. The p-value is taken from the computations performed by **TMB**, and it is obtained based on the asymptotic normality of the test parameters and on the Fisher information matrix. Thus, we are reliant on the asymptotic normality of the test parameters no matter which method is chosen. One benefit of a simulation-based power analysis as opposed to the method which we will apply, is that we can, to some extent, control the accuracy of the asymptotic power by averaging over several simulations.

3.3.1 Asymptotic power of the reduced model

The reduced model refers to the case where one of the two 'cost of plasticity'-parameters is fixed at zero, such that the significance test is applied to only one parameter. For example, we could have a reduced model where $\beta_4 = 0$ is fixed, and β_3 is unknown. The model then assumes that the logarithm of expected fitness of each individual i, j, $i = 1, \ldots, m$, $j = 1, \ldots, n$, follows

$$\log w_{ij} = \beta_0 + \beta_1 z_{ij} + \beta_2 z_{ij}^2 + \beta_3 b_{ij},$$

and thus performs estimation on a subset of the parameters of the full model. The hypothesis test on the cost of plasticity is, in this case, a one-dimensional, two-sided test on β_3 with null value $\beta_3 = 0$. As the sample size is increasing, the ML-estimator for β_3, denoted by $\hat{\beta}_3$, converges towards a normal distribution. Thus, given that H_0 is correct,

$$\hat{\beta}_3 \overset{\text{asymp.}}{\sim} N\left(0, \mathrm{SE}(\hat{\beta}_3)^2\right) \mid H_0,$$

where $\mathrm{SE}(\hat{\beta}_3)$ is the standard error of $\hat{\beta}_3$, as estimated by the computations of **TMB**. The null hypothesis is rejected if the estimator deviates too much from this distribution, which means that H_0 is rejected if either

$$\hat{\beta}_3 \leq -z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3), \quad \text{or} \quad \hat{\beta}_3 \geq z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3),$$

where $z_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution.

Let the true value of β_3 be known, and let it be non-zero such that H_0 is in fact false. The estimator then has a shifted asymptotic normal distribution with non-zero mean β_3 and the same standard error under H_1. Denote the cumulative distribution function of the standard normal distribution by Φ. The asymptotic test power is then the probability of rejecting H_0, when the test estimator follows the distribution actual distribution as given by H_1,

$$\begin{aligned}\text{Power} &\overset{\text{asymp.}}{=} P\left(\hat{\beta}_3 \leq -z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3) \mid \beta_3\right) + P\left(\hat{\beta}_3 \geq z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3)) \mid \beta_3\right), \\ &= \Phi\left(\frac{-z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3) - \beta_3}{\mathrm{SE}(\hat{\beta}_3)}\right) + \left(1 - \Phi\left(\frac{z_{\alpha/2}\ \mathrm{SE}(\hat{\beta}_3) - \beta_3}{\mathrm{SE}(\hat{\beta}_3)}\right)\right).\end{aligned}$$

Figure 3.2 shows the area that constitutes the asymptotic power of a significance test on β_3. The asymptotic test power of a one-dimensional significance test on β_4 can be derived in an equivalent way, using the asymptotic normal distribution and the standard error of β_4.

3.3.2 Asymptotic power of the full model

Now consider the bivariate test on both parameters, as presented in (3.5). For this subsection, let $\boldsymbol{\beta} = [\beta_3, \beta_4]^T$, and let $\hat{\boldsymbol{\beta}}$ denote the associated ML estimator,

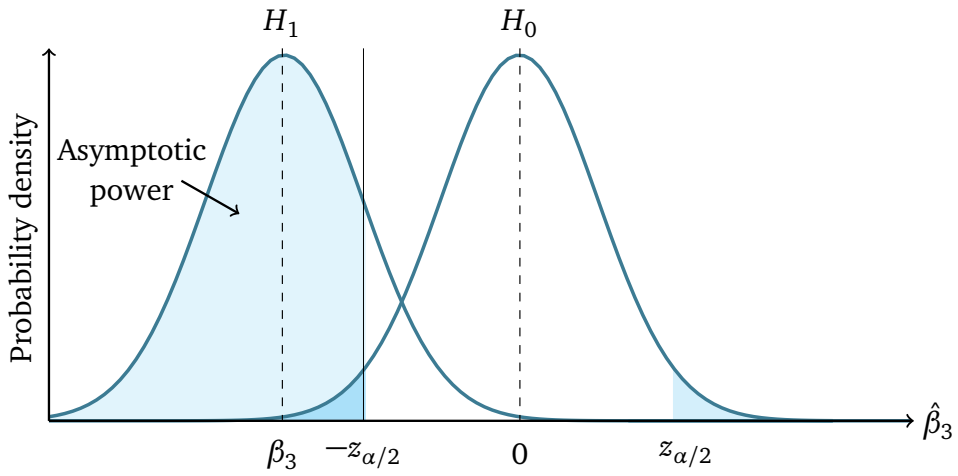


Figure 3.2: One-dimensional significance test on β_3 with $\beta_3 \leq 0$.

$\hat{\boldsymbol{\beta}} = [\hat{\beta}_3, \hat{\beta}_4]^T$. The submatrix of the Hessian matrix of the model which contains the variances and covariance of $\hat{\beta}_3$ and $\hat{\beta}_4$ is denoted by $\boldsymbol{\mathcal{I}}$,

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix} \mathrm{SE}(\hat{\beta}_3) & \widehat{\mathrm{Cov}(\hat{\beta}_3, \hat{\beta}_4)} \\ \widehat{\mathrm{Cov}(\hat{\beta}_3, \hat{\beta}_4)} & \mathrm{SE}(\hat{\beta}_4) \end{bmatrix}.$$

The estimator $\hat{\boldsymbol{\beta}}$ then has an asymptotic bivariate normal distribution under H_0 with mean equal to the zero vector and a variance of $\boldsymbol{\mathcal{I}}^{-1}$. The Wald statistic, denoted by W, is defined as the quadratic form of $\hat{\boldsymbol{\beta}}$ under H_0,

$$W = \hat{\boldsymbol{\beta}}^T \boldsymbol{\mathcal{I}} \hat{\boldsymbol{\beta}}, \tag{3.6}$$

such that W has a chi-squared distribution with 2 degrees of freedom under the asymptotic distribution of the ML-estimator under H_0. Thus, the null hypothesis is rejected if

$$W \geq \chi^2_{1-\alpha,2},$$

where $\chi^2_{1-\alpha,2}$ is the $(1-\alpha)$-quantile of the chi-squared distribution with 2 degrees of freedom.

Again, let the true parameter values β_3 and β_4 be known, and let at least one of them be non-zero such that H_1 is true. Then, under H_1 and given the true parameter values, the ML estimator has a shifted asymptotic normal distribution with mean $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}} \overset{\text{asymp.}}{\sim} N\left(\boldsymbol{\beta}, \boldsymbol{\mathcal{I}}^{-1}\right) \mid H_1. \tag{3.7}$$

It follows that the distribution of the Wald statistic is also shifted under H_1. In fact, W has non-central chi-squared distribution under H_1 (Hélie 2007). A sum of k squared independent standard normal variables is chi-squared distributed with

k degrees of freedom. A sum of k squared independent normal variables with unit variances and means μ_i, $i = 1, \ldots, k$, where at least one μ_i is non-zero, has a non-central chi-squared distribution with non-centrality parameter

$$\lambda = \sum_{i=1}^{k} \mu_i^2.$$

Thus, W has a non-central chi-squared distribution under H_1 with some non-centrality parameter, λ. We transform the normally distributed random variable of (3.7) to obtain a vector variable of uncorrelated, unit-variance normally distributed variables. This gives us

$$\mathcal{I}^{1/2}\hat{\boldsymbol{\beta}} \overset{\text{asymp.}}{\sim} N\left(\mathcal{I}^{1/2}\boldsymbol{\beta}, I_2\right) \mid H_1,$$

where I_2 denotes the 2×2 identity matrix. Finally, under H_1, and given the true values of β_3 and β_4, the Wald statistic has a non-central chi-squared distribution with non-centrality parameter λ, where

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mathcal{I}^{1/2}\boldsymbol{\beta}, \quad \text{and } \lambda = \mu_1^2 + \mu_2^2.$$

The asymptotic power of a two-dimensional test is then

$$\begin{aligned} \text{Power} &\overset{\text{asymp.}}{=} P\left(W > \chi^2_{1-\alpha,2} \mid W \sim \chi^2_2(\lambda)\right) \\ &= 1 - P\left(W \leq \chi^2_{1-\alpha,2} \mid W \sim \chi^2_2(\lambda)\right). \end{aligned}$$

An illustration which shows the area that constitutes the asymptotic power of a two-dimensional test using the Wald statistic is given in figure 3.3. The two dashed lines represent the expected value of W under H_0, which is 2, and, the expected value of W under H_1, which is $(2 + \lambda)$, respectively. The colored area is the asymptotic power.

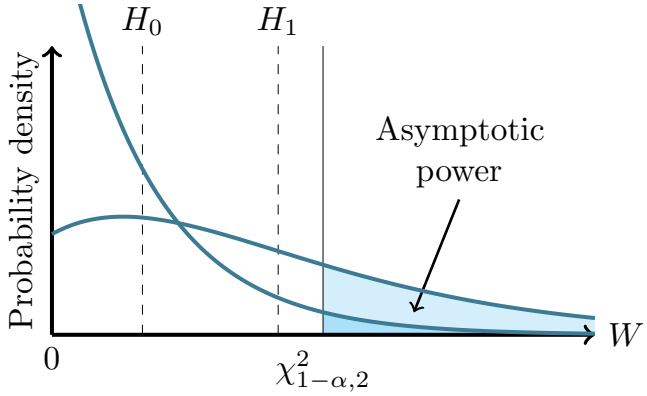


Figure 3.3: Two-dimensional significance test on β_3 and β_4 with at least one significant parameter.

3.4 Experimental design

This thesis attempts to discover the optimal way to design a field experiment for which the objective is to estimate the cost of plasticity using the statistical model presented in 3.1. This is done by varying some of the factors which are possible to control when performing such an experiment, and observe how different conditions affect the asymptotic test power of the associated models. An expedient distribution of individuals into families of full siblings is of biggest interest. Different models for the environmental contribution is also worth considering, as it is a manipulable factor. Several studies have shown that sufficient genetic variation is a prerequisite for evolution of all traits. Although this factor may be hard to control or manipulate in a case study, different levels of genetic variation and their influence on the model will be explored. Finally, recall that the expected fitness of an individual as of (3.4) states that both a direct linear effect and a direct quadratic effect of individual plasticity is possible. However, costs may arise in different ways and persist in different manners in different populations. The values of the 'cost of plasticity'-parameters will be varied to test how the model performs under different types of costs.

The relationship between the population arrangement of relatives and the power of a hypothesis test on the cost of plasticity is elucidated by simulating several dissimilar populations of N individuals for which the number of families, m, and thus the number of full siblings, n, is varied while N and all the other parameters remain unchanged, and computing the asymptotic test power for each population setup. The optimal values of m and n for a set of parameters are those which produce the highest test power. The family structure of a population is expected to be decisive for the test power, because it determines the heterogeneity between individuals in the data set. It is expected that, for each value of N, there are some concurrent m, n-values which maximize the test power of the statistical model on the associated populations. The data sets represent actual populations, so m and n need to take suitable values such that a similar pedigree chart can be recreated.

When performing a case study it may also be possible to place individuals into different surroundings, which directly affects the environmental influence on the phenotype. It is worth exploring whether this factor is decisive on the test power of the statistical model. Two different modelings of the environmental contributions to the phenotype will be tested in this experiment. Both options produce individual contributions with an unconditional expected value of 0. The first option is to draw contributions independently from zero-mean a normal distribution, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, for $i = 1, \ldots, m$, $j = 1, \ldots, n$. This can model a design in which all individuals of the population move freely in some restricted area with microenvironmental fluctuations. The alternative approach which will be considered in the thesis is to let $\varepsilon_{ij} \in \{\varepsilon_1, \varepsilon_2\}$, for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$, where $\varepsilon_1 \neq \varepsilon_2$. This could model the case for which the population is divided 50 : 50 into two

separate areas of complete distinction. It would then be necessary to divide the siblings of each family by 50 : 50 in the two distinct locations, such that environmental correlation is not mistaken for genetic correlation by the model. The latter model is expected to produce the highest power. This is because the environment more distinctly influences the phenotypes of different individuals, such that the model can distinguish between the environmental impact to the phenotype and the genetic contributions more easily. The two different models are applied to the same populations, and the associated test powers are observed and compared.

Sufficient variation of the explanatory variables of a statistical model is a necessary presupposition for the model to detect the significance of the explanatory variables. Thus, populations will be simulated from different values on G_{bb} in order to observe the importance of variance in plasticity for detecting a cost of plasticity. Higher values of G_{bb} are assumed to increase the test power.

Several alternatives of the cost of plasticity will be tested. This entails varying the values of the model parameters β_3 and/or β_4. It is expected that the model power increases when the absolute value of the 'cost of plasticity'-parameter(s) increases.

Each statistical model for fitness is obtained by performing MLE on the data of a simulated population. Because each population data set is made from sampling from distributions with a significant variance parameter, two populations can be quite dissimilar even though they are obtained from the same parameter- and m, n-values. Thus, fitness models which are derived from the data of two different populations which are sampled from the exact same parameter values and m, n-values may be slightly different from each other. The extent is unknown until experiments are completed, but, hopefully, fitness models derived from populations of the same fundamental values are closely related, and easily distinguishable from models derived from populations of other values.

3.5 Implementation

All results are obtained from simulating a population under a cost of plasticity, such that the alternative hypothesis, H_1, is true, and calculating the asymptotic power. All tests have a significance level of $\alpha = 0.05$, and all tests are two-sided with a null-value of zero (or the zero vector for the multidimensional case). The results will be presented as plots showing the asymptotic test power of models from different simulations, with asymptotic power either as a function of n, or as a function of β_3 or β_4. Details on the different simulations will be explained in the following subsections.

3.5.1 Selection assumptions

Simulations are executed for three different assumptions on the effect of the cost of plasticity. Firstly, the assumption of an exclusive linear cost effect to the logarithm of expected fitness, such that the expected fitness of sibling j of family i of a population is

$$\log w_{ij} = \beta_0 + \beta_1 z_{ij} + \beta_2 z_{ij}^2 + \beta_3 b_{ij},$$

where z_{ij} and b_{ij} is the phenotypic trait- and plasticity value, respectively, for all $i = 1, \dots, m$, $j = 1, \dots, n$. For these experiments, β_4 is fixed at zero and the model makes no attempt at estimating β_4. The second assumption is that the cost has an exclusive quadratic effect, such that sibling j of family i with trait- and plasticity value z_{ij} and b_{ij}, respectively, has an expected fitness which follows

$$\log w_{ij} = \beta_0 + \beta_1 z_{ij} + \beta_2 z_{ij}^2 + \beta_4 b_{ij}^2,$$

for all $i = 1, \dots, m$, $j = 1, \dots, n$. For these experiments, β_3 is fixed at zero and the model makes no attempt at estimating β_3. Finally, the last assumptions is the full model, in which both β_3 and β_4 are assumed to be significant and are inferred by the model. For this case, individual expected fitness obeys (3.4).

As mentioned in the previous section, we expect that the asymptotic power of the significance test will increase as the deviation from the null value of the given test parameters increase. This will be tested for the first two assumptions on the cost of plasticity, by plotting the asymptotic test power as a function of β_3/β_4 on a given population. Recall that we compute the asymptotic power by computing the probability of rejecting the null hypothesis when the true value of the test parameter is given. When the test parameter is zero, such that H_0 is actually true, rejecting the null is equivalent to committing a type I-error. The probability for committing a type I-error is equal to the significance level of the test, α. Thus, we expect that the asymptotic power is α when the test parameter is 0, and that the asymptotic power increases as the absolute value of the test parameter increases.

3.5.2 Population structure

The values of m and n must be chosen such that both m and n are positive integers, the product $m \cdot n$ is exactly N, where N denotes the population size, and

n is an even number. (The latter requirement is added because a subset of the experiments entail the division of n siblings into two equally sized groups, so $n/2$ should be an integer.) At the same time, we also want to apply the model to as many populations of different structures as possible. The population sizes are set to be $N = 10000$ for all simulated populations. This number produces a sufficient sample size, while keeping the computational time at an acceptable level. The applied test values are given in the following table , where each column represents a population setup.

n-value	2	4	8	10	16	20	40	50	80	100	200	250	500	1000
m-value	5000	2500	1250	1000	625	500	250	200	125	100	50	40	20	10

3.5.3 Environmental models

Recall that the parameters of the environmental contribution is not studied by the model. The individual contributions are drawn from their respective models and are simply treates as given values. However, they are still expected to have an impact on the model results. Simulations are implemented for the two different environmental models introduced in the previous section, which will be denoted by ‘env 1’ and ‘env 2’. More specifically,

- Model ‘env 1’ is the case where all environmental contributions are drawn independently from a standard normal distribution, and

- Model ‘env 2’ models the even division of the population and full siblings into two distinct environments. For each family of a simulated population, half of the full siblings have an environmental contribution valued at 1 and the other half have a contribution valued at -1.

3.5.4 Variance of plasticity

Populations will be simulated from several distributions of different levels of plasticity variance. Three different levels will be applied to population of all the different selection assumptions. The associated asymptotic powers will be plotted into the same figures and compared.

3.6 Appropriate parameter values

This section justifies the choice of parameter values of the simulated populations. The parameters will be presented as estimators by the model, but, as this experiment considers simulated populations rather than real ones, the parameter values must also be set prior to simulation. To imitate the scenario of a real case study to the biggest possible extent, and to enable a decent data analysis, some constraints should be applied to the range of the different parameter values. There needs to be some compliance between the trait values which can be observed in an actual population and the simulated trait values, such that the obtained knowledge of this thesis is transmissible to a real experiment. The model parameters include those of the genotype distribution, σ_e^2, and theβ-values. The different considerations regarding the different parameters will be presented in the next three subsections.

3.6.1 Genetically determined parameters

The parameters $\bar{\bar{a}}$, $\bar{\bar{b}}$, G_{aa}, G_{bb}, G_{ab} are those that constitute the genotype distribution. The overlying mean breeding value, $\bar{\bar{a}}$, represents the mean phenotype across all environments, and $\bar{\bar{b}}$ represents the impact to the trait of the environment. The exact values of $\bar{\bar{a}}$ and $\bar{\bar{b}}$ are not that interesting in a simulation study — they can easily be scaled to represent any trait which is given by a model equal to (3.3). It is first and foremost the variance components of the distribution which are worth considering. The variance parameters, G_{aa} and G_{bb}, must be sufficiently large such that genetic adaptation is possible. Insufficient genetic variation may constrain the evolution of traits in the population. It is assumed that the breeding value and the plasticity value in any given individual are uncorrelated due to genetic canalization in the mean environment (Lande 2009). It follows that the population-wide genotype covariance, G_{ab}, is minimized. All the populations considered in this thesis will be simulated from distributions of $G_{ab} = 0$.

3.6.2 Heritability and additive variance

The heritability of a phenotypic trait, denoted by h^2, is the fraction of the trait variance that is due to additive, genetic variance in the population. It has been shown that the selection response is proportional to h^2 (Lande and Arnold 1983). Thus, if h^2 is close to zero, then there is approximately no genetic variability in the population, and selection becomes negligible. If h^2 becomes close to 1 this means that the environment contributes nothing to the phenotypic variance, and modeling selection in this way is unconvenient. It is desirable that the heritability of the phenotypic trait under selection is approximately 0.50, as $h^2 = 0.50$ implies that the inherited genes and the environment with other possible factors contribute equally to the phenotypic variance. The approach will be to insure that the expected fraction for each individual is 0.50. Recall that a commonality of all the applied

environmental models is that the contribution experienced by each individual is zero-meaned; for individual i, j, $E(\varepsilon_{ij}) = 0$, for all $i = 1, \ldots, m$, $j = 1, \ldots, n$. Thus,

$$h^2 = \frac{\text{additive variance}}{\text{phenotypic variance}} = \frac{G_{aa}}{G_{aa} + \sigma_e^2} = 0.50,$$

which gives $\sigma_e^2 = G_{aa}$. This restriction will be applied to all the simulations.

3.6.3 Selection

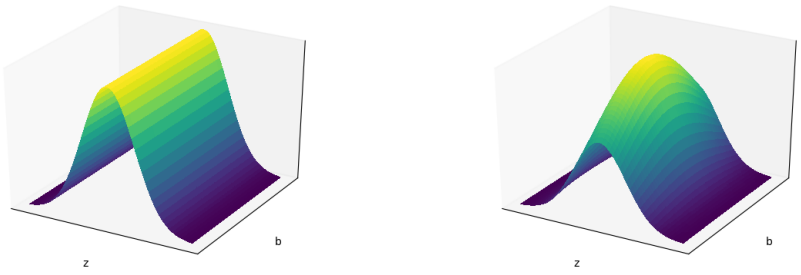
The presence of the Gaussian phenotypic stabilizing selection is realized through parameters β_1 and β_2, with analytical values given by

$$\beta_1 = \frac{1}{\omega^2}\theta, \qquad \beta_2 = -\frac{1}{2\omega^2},$$

as of the definition in (2.1). (This selection may of course also affect the intercept, but, as the β_0-value is also influenced by other mechanisms which are unrelated to the phenotypic selection, the phenotypic selection forces no constraints onto β_0.) The selection is thought to be stabilizing, which entails that θ should be non-extreme relative to the trait values of the population. Thus, θ should be valued such that it is close-to-centered in the trait distribution of the population. The ω^2-parameter affects the strength of the selection, and it is inverse proportional to both β_1 and β_2. Thus, high values of ω^2 lowers the significance of the trait values to fitness. This is equivalent to widening the fitness function.

A cost of plasticity is present if either β_3 or β_4 are non-zero. These should be valued such that they reflect that non-optimal plasticity is costly. A cost can have several effects. For example, there can be either a linear or quadratic effect, or both, and, the optimal levels of plasticity can take different values. Various options of a cost of plasticity will be imposed on the populations, and the two latter β-parameters will take values accordingly.

An example on how the cost of plasticity may affect individual fitness is given in figure 3.4. It shows a surface plot of the fitness function with a stabilizing phenotypic selection in two cases. In the leftmost figure, there is no cost of plasticity present, and, in the rightmost figure, there is a stabilizing selection on plasticity. Observe that the surface of the fitness function is curved in the b-dimension when a cost of plasticity is introduced.



Plasticity is not costly Plasticity is costly

Figure 3.4: 3-dimensional surface plots of fitness as a function of the phenotype and plasticity. In the plots, z denotes the phenotype and b denotes the plasticity, respectively, of a given individual.

3.7 Computation

The process of computing the results are briefly described in this section. Results are produced using R, and C++ via the R-package **TMB**. Fixation of parameter values and data simulation is performed in R. The input of the model, which is the data representation of a population, is simulated from sampling data values based on the respective distributional assumptions in R. The ML estimators of each of the unknown parameters of the statistical model is obtained by the means described in section 2.4 — the data set is sent to a C++-file which returns the negative marginal log-likelihood of the model, which is later minimized in an R-file. The asymptotic test power is computed straightforwardly in R. All plots given in the figures of chapter 4 are obtained using the open R-package `ggplot2`.

Chapter 4

Results

This chapter contains all the results which were produced. Each result is presented as a plotted figure which show the asymptotic power of models. They are given as drawn lines between data points, where each point represents the asymptotic power of the model applied to a population simulated under the given parameters and model assumptions. Asymptotic power is plotted either as a function of the logarithm of n, or as a function of either β_3 or β_4. A logarithmic scale for n was chosen to narrow some of the intervals between different n-values on the figures. The results showing asymptotic power as a function of n are easier to interpret in a log-plot.

The results are shown for statistical models of three different assumptions on the cost of plasticity, as introduced in subsection 3.5.1. First, the model which assumes that the cost of plasticity has an exclusive linear effect, such that β_3 is non-zero and estimated by the model while $\beta_4 = 0$ is fixed. The results on the statistical models of this assumption are given in section 4.1. The test results for the model which assumes that the quadratic effect is significant while the linear effect is insignificant, such that $\beta_3 = 0$ is fixed in each model and β_4 is inferred, are then given in 4.2. Finally, the asymptotic powers of the models which assume that the cost of plasticity has both a linear and quadratic effect, are shown in section 4.3.

Under the first two selection assumptions, the significance test for the cost of plasticity is a one-dimensional test. Plots of the asymptotic test power as a function of the respective test parameter are made for the two one-dimensional significance tests. All of the three model assumptions are explored in different circumstances. They are explored for several m, n-values, under two environmental models, 'env 1' and 'env 2', respectively, and, three levels of plasticity variance, G_{bb}.

Some model parameter values remain constant in all the simulations. These are: $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = -0.20$, $\sigma_e^2 = 1$, $\bar{\bar{a}} = 0$, $\bar{\bar{b}} = 0.5$, and, $G_{aa} = 1$ and $G_{ab} = 0$. The populations of every figure, except for the figures which illstrutate the effect of different levels of plasticity variance, that is, 4.3, 4.7 and 4.11, are simulated under a plasticity variance value of $G_{bb} = 0.1$.

4.1 Linear effect

The results for the model assumptions of this section are given in figures 4.1, 4.2, 4.3 and 4.4. In the first three figures, the cost of plasticity is given by $\beta_3 = -0.20$. The quadratic effect is non-existent for this case, so $\beta_4 = 0$ is fixed in all the figures.

Figure 4.1 reveals that the family size is quite decisive on the test power of the models with an exclusive linear cost of plasticity. It is plain to see that a high number of full siblings per family is the preferable population structure. This claim is also supported by figures 4.2 and 4.3. According to figure 4.2, the environmental models have a virtually equal impact on the power of these models. There are some minor fluctuations to observe, but no environmental model stands out as the superior one. Figure 4.3 shows that a higher variance of plasticity gives models of higher power. The graph of figure 4.4 shows how the asymptotic power evolves as the strength of the cost increases for a population of $n = 100$ in 'env 1'. Just as expected, it starts at $\alpha = 0.05$ for $\beta_3 = 0$, and then increases towards 1 as β_3 increases.

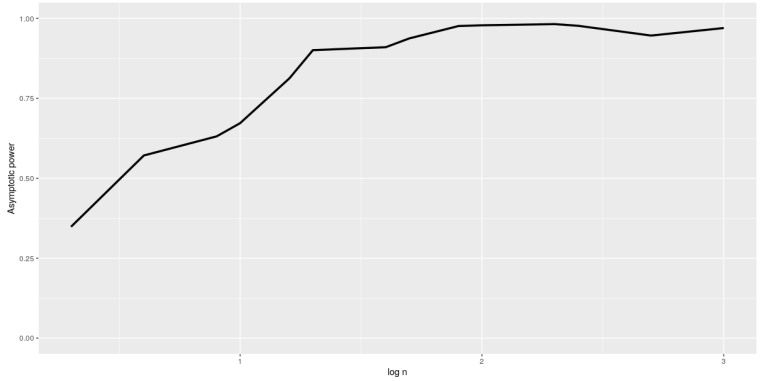


Figure 4.1: Asymptotic power as a function of $\log_{10}(n)$ for a linear cost of $\beta_3 = -0.20$ and $\beta_4 = 0$ fixed for populations in 'env 1'.

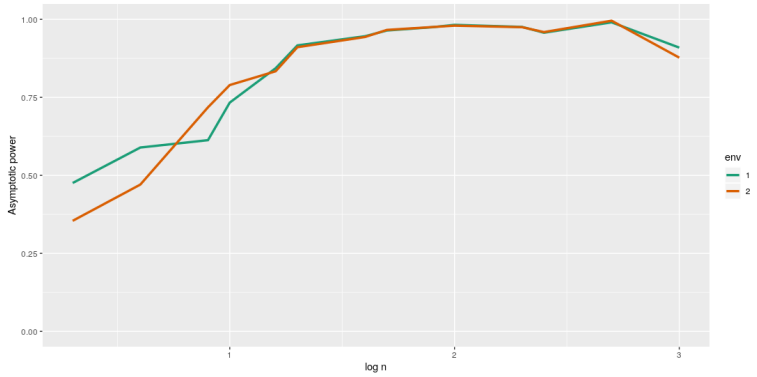


Figure 4.2: Asymptotic power as a function of $\log_{10}(n)$ for a linear cost of $\beta_3 = -0.20$ and $\beta_4 = 0$ fixed. Populations are simulated in different environmental models as indicated to the right.

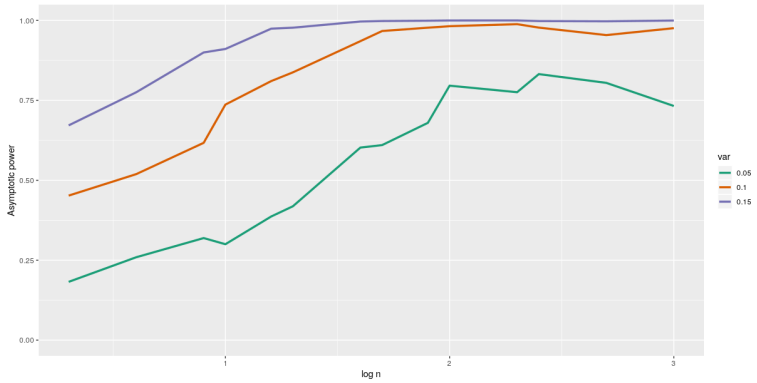


Figure 4.3: Asymptotic power as a function of $\log_{10}(n)$ for a linear cost of $\beta_3 = -0.20$ and $\beta_4 = 0$ fixed for populations in 'env 1'. Populations are simulated from different values of G_{bb} as indicated to the right.

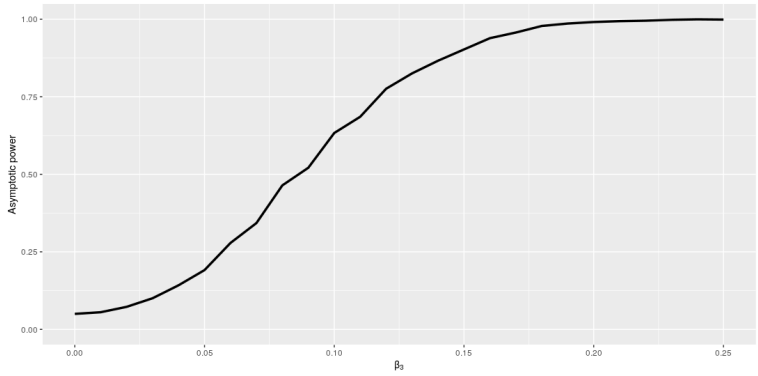


Figure 4.4: Asymptotic power as a function of β_3 for a population of $n = 100$ and $\beta_4 = 0$ fixed in 'env 1'.

4.2 Quadratic effect

The results for the model assumptions of this section are given in figures 4.5, 4.6, 4.7, and 4.8. The cost of plasticity is given by $\beta_4 = -0.15$ in the first three figures, and $\beta_3 = 0$ is fixed in this section.

We can detect a modest pattern of the effect of the values of m and n on the model power for models of an exclusive quadratic in figure 4.5. The same pattern is also identifiable in both figures 4.6 and 4.7. There is less variation in asymptotic power to observe in the figures of this section than those of the previous section, and it appears that the choices of m and n are less decisive on the power of models for which $\beta_3 = 0$ and β_4 is significant. Figure 4.6 reveal that the two environmental models have a similar effect on the models of these section as well. Figure 4.7 shows that an increase in plasticity variance poses an increase in asymptotic model power. The development of the asymptotic test power as the strength of selection increases is shown in 4.8 for a population with $n = 100$ in 'env 1'. Again, asymptotic power is at $\alpha = 0.05$ for $\beta_4 = 0$ and the increases towards 1, just as expected.

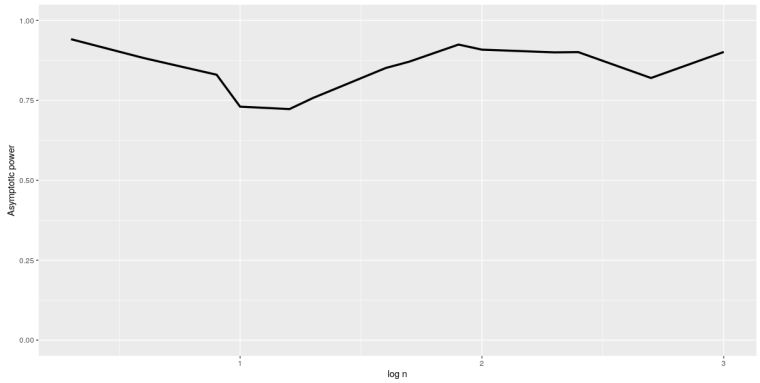


Figure 4.5: Asymptotic power as a function of $\log_{10}(n)$ for a quadratic cost of $\beta_4 = -0.15$ and $\beta_3 = 0$ fixed for populations in 'env 1'.

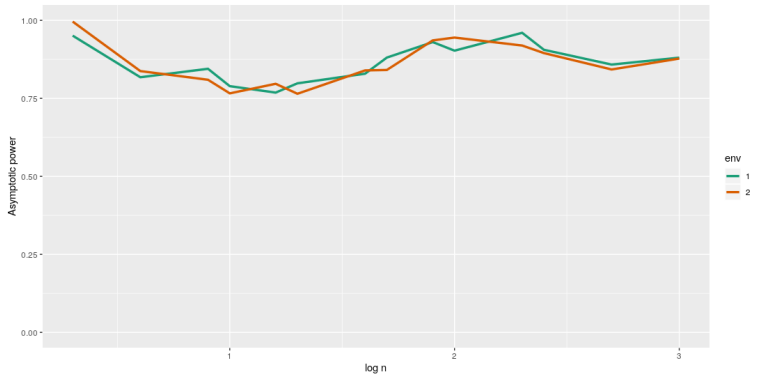


Figure 4.6: Asymptotic power as a function of $\log_{10}(n)$ for a quadratic cost of $\beta_4 = -0.15$ and $\beta_3 = 0$ fixed. Populations are simulated in different environmental models as indicated to the right.

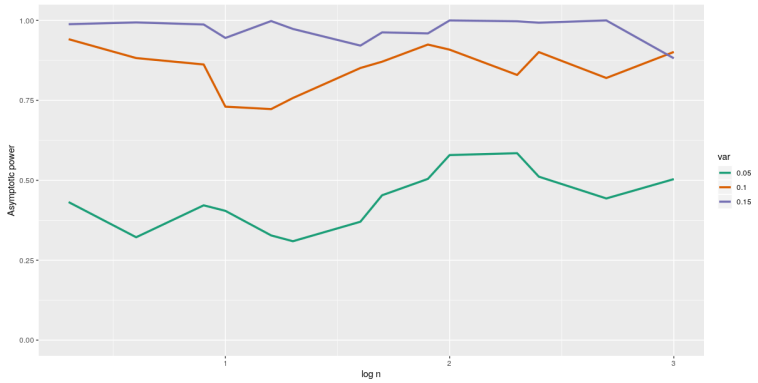


Figure 4.7: Asymptotic power as a function of $\log_{10}(n)$ for a linear cost of $\beta_4 = -0.15$ and $\beta_3 = 0$ fixed for populations in 'env 1'. Populations are simulated from different values of G_{bb} as indicated to the right.

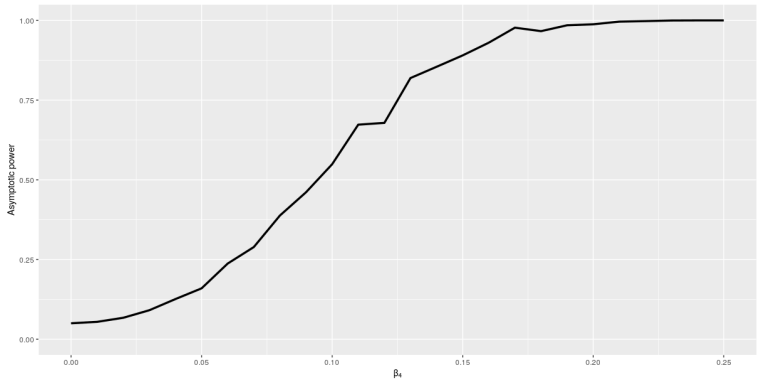


Figure 4.8: Asymptotic power as a function of β_4 for a population of $n = 100$ and $\beta_3 = 0$ fixed in 'env 1'.

4.3 The full model

This section considers all the results for the models which assume that both the linear and quadratic effect of the cost of plasticity is significant. The cost is given by $\beta_3 = 0.01$ and $\beta_4 = -0.05$. The results are shown in figures 4.9, 4.10 and 4.11. Note that the results for this section only consider a subset of the m, n-values which were considered by the models of the two previous sections. This is because the optimization algorithm generally did not converge for small values of n ($n \leq 20$) when analysing models where both β_3 and β_4 were significant. Hence, the results of this section are presented for populations of $n \in [40, 1000]$.

According to figure 4.9, a large number of full siblings is the most beneficial choice. The asymptotic model power is mostly increasing with n. The two other figures, 4.10 and 4.11, support the assertion — the model power reaches its maximum in the populations of the highest number of full siblings per family in both environmental models and all the considered levels of plasticity variance. In contrast to the results of the one-dimensional tests, the choice of an environmental model for the population seems to be quite crucial to the performance of the statistical model which assumes both a linear and quadratic effect. For the higher values of n, the 'env 2' model is the superior choice by far, and constitutes a remarkable improvement to the model power. Similar to what we observed in figures 4.3 and 4.7, figure 4.11 reveals that the model power increases with the G_{bb}-value for the full model as well.

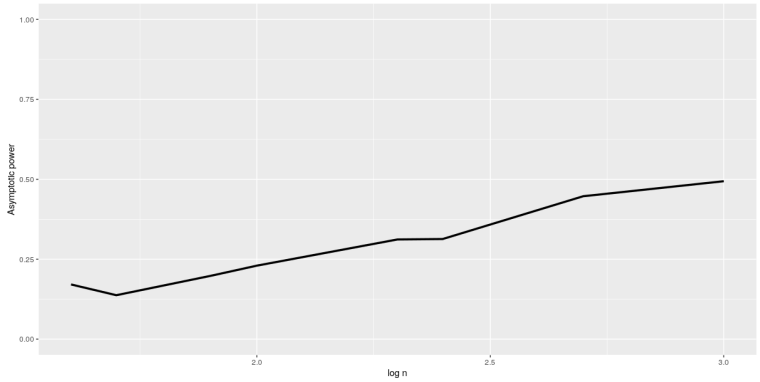


Figure 4.9: Asymptotic power as a function of $\log_{10}(n)$ for a cost of $\beta_3 = 0.01$ and $\beta_4 = -0.05$ for populations in 'env 1'.

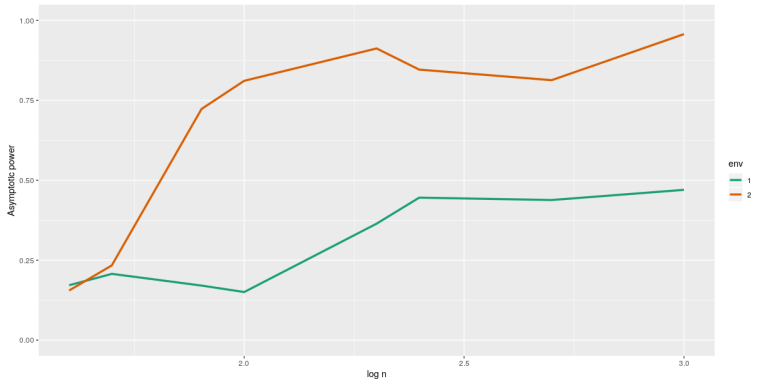


Figure 4.10: Asymptotic power as a function of $\log_{10}(n)$ for a cost of $\beta_3 = 0.01$ and $\beta_4 = -0.05$. Populations are simulated in different environmental models as indicated to the right.

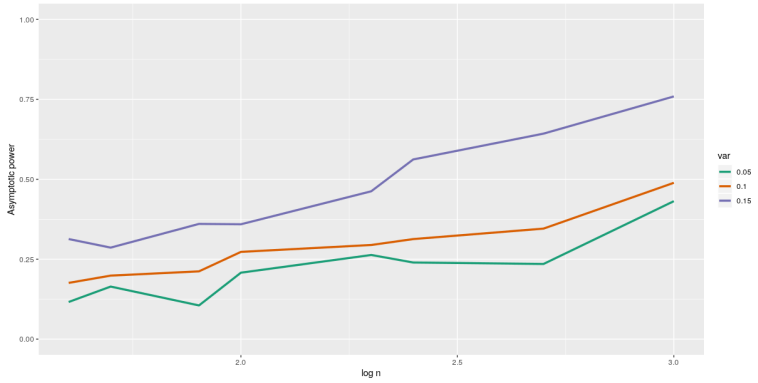


Figure 4.11: Asymptotic power as a function of $\log_{10}(n)$ for a cost of $\beta_3 = 0.01$ and $\beta_4 = -0.05$ for populations in 'env 1'. Populations are simulated from different values of G_{bb} as indicated to the right.

Chapter 5

Discussion

Just as we suspected, the results reveal that the various conditions which were tested on the statistical model do have a conclusive effect on the asymptotic model power. As mentioned before, the most gripping result of this thesis is the relation between model power and the number of full siblings per family of the populations. Few predictions were made on this relation in advance, other than that the choice of the m- and n-values of a population would be decisive on the power of the applied model. It is safe to say that this expectation was satisfied. However, it remains to establish how the m, n-values affect the model power, and why they affect models of different assumptions on the cost of plasticity differently. This relation will be further discussed in the next section.

It was expected that the 'env 2'-model for the environmental contribution would be the better choice in all the models, however, this was only true for the models which considered both a linear and quadratic effect of the cost. This is a surprising result, and the occurrence of this distinction is uncertain. Section 5.2 will examine the impact of the two environmental models deeper.

The results confirmed our expectation about the variance of plasticity, and, they clearly state that sufficient variation in the plasticity trait is a necessity for detecting a cost of plasticity in a population. This is a general result which holds for all the model assumptions which were explored. Some dissimilarities in the exact outlet of this result can be observed by comparing the figures — this will be further discussed in section 5.3.

It is evident that, in the tests concerning models which assume either a linear or a quadratic effect of the cost of plasticity, where the significance tests are one-dimensional, the model power increases as the relevant test parameter increases in absolute value. This is just as expected, and will not be further discussed.

Some problems were encountered underway. The biggest obstacle was making the optimization algorithm converge for the fitting of the full model onto the data sets where both β_3 and β_4 were significant. Often, it would return a Hessian matrix which was not positive definite. The algorithm would never converge if the model assumed a significant β_3 and β_4, and had a small n-value. Thus, those

values of n had to be neglected from the power analysis for those models. Different parameter values and initial values were attempted, but the same problem occured for all of them. However, there might exist some parameter values, or seed for the optimization algorithm which would fix this problem.

The figures of chapter 4 are not necessarily as smooth as one would want them to be. The lines on all the plots are slightly stuttering. This is partly due to the amount of data points per plot — each of the plots showing asymptotic power as a function of $\log n$ has data points for only 14 different n-values (8 in the results of the final section), because of the restrictions we forced upon the choices for different n-values in the population. Another factor is that the data points are based on simulations. So, if there is a perfectly smooth function which describes the relation between the model power and m and n and the rest of model parameters, we have only been able to sample from it. Even though two distinct populations are simulated from the exact same parameter values, there is a sizeable chance that the two populations can result in two statistical models with a considerable difference in statistical power. We might have removed this problem by performing simulation-based power analysis on all the models, because, then we could control the uncertainty of the model power by increasing the number of tests. Nevertheless, simulations from the same base parameters always returned similar patterns. By comparing the first three figures associated to each model in the respective sections of chapter 4 to one another, we can observe the same pattern, that is, the way the asymptotic power moves with $\log n$, in all three of them. So, the variation of results due to the simulation inaccuracy is not problematic. Additionally, figures 4.4 and 4.8 are slightly stuttering as well, yet they reveal a satisfying and correct pattern.

Some of the model parameters might be slightly inconsistent. For example, the model uses family sizes in the entire range from 2 to 1000. The possibility of having 1000 full siblings per family might be somewhat inconsistent with the parameters on selection in the simulated population — the results are made for model parameters such that sibling j of family i produces y_{ij} siblings with $y_{ij} \in [0, 15]$, for all $i = 1, \ldots, m$ $j = 1, \ldots, n$. Hopefully, the model can be reparameterized in an uncomplicated matter such that the results are applicable for populations of species which produce even more offspring. This was nonetheless never tested for.

Overall, the new latent-variable statistical method has performed well. Under the optimal conditions, as revealed by the results of this thesis, one can achieve a model of an adequate power level. A two-dimensional test might be preferable in the sense that neither of the parameters need to be discarded beforehand — if we suspect a cost of plasticity in a log-linear Poisson model, we can apply the model without considering whether the cost is linear or quadratic. However, if the type of effect is virtually known beforehand, and it leads to a one-dimensional test, it is worth conditioning on, as the one-dimensional tests are easier to perform and gives higher power in the case that one of the parameters is in fact 0.

5.1 Family distribution

The influence of the different choices of m- and n-values on the asymptotic powers of the models is clearly dependent on the selection assumptions of the respective models. For the models of a one-dimensional cost with a significant β_3-parameter, it is evident that the model power increases with $\log n$. In the one-dimensional case with a significant β_4-parameter, however, the model power is alternating weakly in the upper layer of its range. The full model which assumes that both the linear and quadratic effect of plasticity is significant to log fitness achieves its maximum power for the larger values of n. For the latter case, it appears that asymptotic power increases with $\log n$, though with some minor alternations underway.

So, if we were to design an experiment for estimating the cost of plasticity in a population, the optimal choices of the number of families and the sizes of each family would depend somewhat on our assumptions and knowledge on the cost of plasticity in the population. (We must also consider the restrictions of the procreation habits of the relevant population — not all species have the capacity to produce any quantity of offspring in the entire range between 2 and 1000.) One commonality of all the results, is that the power is relatively high for large values of n. Hence, across all the selection assumptions on the cost of plasticity, for a population of $N = 10000$ individuals, $n = 1000$ full siblings per family (if possible) is an optimal choice. If we assume that the cost of plasticity is linear, or, both linear and quadratic, we would try to maximize n within the framework of the population structure, as model power increases with n. If the cost is assumed to be exclusive quadratic, however, we might be more considerate. For example, in a population of $N = 10000$ individuals where the size of each family ranges within $[10, 100]$, we would prefer $n = 100$, which is the largest possible value of n. However, if the range is $[2, 10]$, we would prefer 2, which is the smallest possible value of n, so, the approach is less straight-forward in this case.

It is not easy to recognize exactly how the values of m and n affect the power of the significance tests of the 'cost of plasticity'-parameters. We start by considering how the choices of m and n affect the inference of the model. A prerequisite for achieving a high power is precise inference. This applies to all the parameters and latent variables of the model, even though only one or two of them are actually being tested. There is an interaction between all of the model parameters, and, if the model either underestimates or overestimates the effect of one of them, it has most likely misunderstood the effect of several other parameters. For the statistical model of this thesis, the plasticity component is a latent variable in two submodels of the selection on each individual — it provides an interaction between the individual phenotypes and the environment, and thus has an indirect effect on fitness through the phenotypic selection, and, it is the main component of the cost of plasticity in each individual. Thus, the parameters of the two distinct selections are interacting in some sense, and it is safe to say that there is a relation between all the model parameters.

The family-determined mean genotype values, which, for family i, $i = 1, \ldots, m$, of the population, are denoted by $\bar{a}_i$ and $\bar{b}_i$, are easier to infer the more they appear explicitly in the likelihood function. Each pair of mean values appear exactly $n+1$ times in the likelihood function of the model (n times as the mean of the genotype distribution of each sibling of the respective family, and, one additional time as a latent variable of the overlying genotype distribution of the population). Thus, when n is large, better estimators are obtained for $\bar{a}_i$ and $\bar{b}_i$, for all $i = 1, \ldots, m$. However, the family-determined mean genotypes are not just parameters of distributions of indviduals — they are also latent variables in an overlying distribution of the population. For all $i = 1, \ldots, m$, $[\bar{a}_i, \bar{b}_i]^T$ are random variables drawn from the overlying genotype distribution centered around the model parameters $[\bar{\bar{a}}, \bar{\bar{b}}]^T$. We might think that, just as for the family-dependent mean values, $\bar{\bar{a}}$ and $\bar{\bar{b}}$ are more easily inferred upon when they appear more frequently in the likelihood function. They appear in the likelihood exactly m times. The parameters do nonetheless not appear as the mean values of known observations, but, as the mean values of latent variables, which are also inferred upon by the model. The quality of inference of $\bar{\bar{a}}$ and $\bar{\bar{b}}$ is largely determined by the quality of inference of each of the $[\bar{a}_i, \bar{b}_i]^T$-values, for $i = 1, \ldots, m$.

It thus makes sense that high values of n is the preferable choice, because it enhances the quality of inference in each of the m different mean values. This was observed in all the results. The results of sections 4.1 and 4.3 are consistent in the sense that the power is virtually monotonically increasing with n. The same is not observed in the results of 4.2. These result shows that asymptotic power is alternating with n, yet it is hard to know why.

5.2 Environmental models

Two distinct environmental models, 'env 1' and 'env 2', as defined in subsection 3.5.3, were applied to the same populations. We expected that making a population subject to 'env 2', as opposed to 'env 1', would significantly increase the test power of the statistical model. This expectation was, for some reason, only satisfied in the models which considered both a linear and quadratic effect of the cost of plasticity.

The (lack of) impact of the choice of environmental models in the models of the restricted cost of plasticity is a surprising result. The rationale was that 'env 2' would make it easier for the model to identify the environmental contribution to the traits. There is no apparent reason why this should remain true in the cases where both β_3 and β_4 are significant and remain false in the cases where only one of them is significant while the other is not. However, as they perform equally well in the one-dimensional tests, no model is superior to the other in this case. Thus, in a case study of either a linear or a quadratic cost, we may choose the most convenient option. If the assumption of the study is that both parameters are significant, however, we should strive to make the population subject to environmental model 'env 2' (at least if we are able to have a sufficiently high n-value such that the models perform differently) to increase the power.

5.3 Variance of plasticity

The most compelling result was the impact of variation of plasticity on the model power. A commonality of all the models is that an increase in variance of plasticity induces an increase in the asymptotic model power.

The exact relation between the change in the G_{bb}-parameter and the following change in the model power is however still mostly unknown. The increase in asymptotic power from changing G_{bb} from 0.05 to 0.1 is dissimilar to the increase in asymptotic power from changing G_{bb} from 0.1 to 0.15, so it is not a linear relationship. For the models which assumed either a linear or a quadratic effect of cost, the increase in model power from changing G_{bb} from 0.05 to 0.1 was bigger than increase from changing G_{bb} from 0.1 to 0.15, while we observed the opposite in the model which assumed that both effects of the cost were significant. Thus, it might be that the relation between asymptotic test power and plasticity depends on the model assumptions on the effects of the cost of plasticity as well.

Chapter 6

Conclusion

The latent-variable approach for detecting the cost of plasticity in a population has been successful. As the breeding values and plasticity values of individuals are treated as latent variables, all aspects of the uncertainty and variation of the parameters regarding the population and the associated selection are included in the model. Thus, there is no hidden bias to spoil the findings of the statistical model on the cost of plasticity. Additionally, we can attain models of a relatively high power when using this approach. A latent-variable statistical model is undoubtedly a suitable model for estimating the cost of plasticity in a given population.

The results uncovered how different distributions of related individuals, different environmental models, and different levels of plasticity variance affected the asymptotic power of a latent-variable statistical model for detecting the cost of plasticity in a population. They revealed that the applied conditions have a significant effect on the asymptotic power, and, that, if we were to perform an experiment on a population with the objective of estimating the cost of plasticity, we should take the state of these conditions into account. The statistical model might have been tested for a wider range of parameter values. Also, it would have been valuable to try performing the power analysis using a simulation study to validate our findings even further.

If we are to perform a case-study on a population to estimate the cost of plasticity, we should apply the results in the following way. If the assumption is that the cost of plasticity is log-linear to fitness, we should maximize the number of relatives within each family as much as possible. If, on the other hand, we assume that the cost of plasticity exclusively consists of a quadratic effect, we should evaluate the situation further, and use figure 4.5, which shows the asymptotic power of populations as a function of $\log n$ for a significant β_4-parameter, as guidance for obtaining the best value of n given the value of N. For the first two cases, the two environmental models had a more or less equal effect on the model power, so, it would be unnecessary to manipulate the environmental contribution in one way or the other. Lastly, if we assume that the cost of plasticity can consist of both a

linear and a quadratic effect, we should maximize the number of relatives within each family as well as we can, and, use a model for the environmental model which resembles 'env 2' in the best possible way. We should also take into account that sufficient genetic variation is a necessity for detecting the cost of plasticity in all the models of the aforementioned assumptions. If variance of plasticity in the population is too low, it might spoil the chances for detecting the cost of plasticity.

Bibliography

[1] I. B. Rønning, *Evolutionary responses to fluctuating selection*, NTNU Trondheim, 2020.

[2] L.-M. Chevin, R. Lande and G. M. Mace, 'Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory', *PLoS Biology*, vol. 8, no. 4, 2010.

[3] J. G. King and J. D. Hadfield, 'The evolution of phenotypic plasticity when environments fluctuate in time and space', *Evolution Letters*, vol. 3, no. 1, pp. 15–27, 2018.

[4] T. J. Dewitt, A. Sih and D. S. Wilson, 'Costs and limits of phenotypic plasticity', *Trends in ecology and evolution*, vol. 13, no. 2, pp. 77–81, 1998.

[5] S. M. Scheiner and D. Berrigan, 'The genetics of phenotypic plasticity. viii. the cost of plasticity in daphnia pulex', *Evolution*, vol. 52, no. 2, pp. 368–378, 1998.

[6] K. Donohue, E. H. Pyle, D. Messiqua, M. S. Heschel and J. Schmitt, 'Adaptive divergence in plasticity in natural populations of impatiens capensis and its consequences for performance in novel habitats', *Evolution*, vol. 55, no. 4, pp. 692–702, 2000.

[7] A. A. Agrawal, J. K. Conner, M. T. J. Johnson and R. Wallsgrove, 'Ecological genetics of an induced plant defense against herbivores: Additive genetic variance and costs of phenotypic plasticity', *Evolution*, vol. 56, no. 11, pp. 2206–2213, 2002.

[8] R. A. Relyea, 'Local population differences in phenotypic plasticity: Predator-induced changes in wood frog tadpoles', *Ecological monographs*, vol. 72, no. 1, pp. 77–93, 2002.

[9] J. D. Hadfield, A. J. Wilson, D. Garant, B. C. Sheldon and L. E. Kruuk, 'The misuse of blup in ecology and evolution', *The American naturalist*, vol. 175, no. 1, 2009.

[10] R. Fisher, 'The correlation between relatives on the supposition of mendelian inheritance', *Transactions of the Royal Society of Edinburgh*, vol. 52, no. 399–433, pp. 77–81, 1918.

[11] J. Etterson and R. M. Kliman, *Encyclopedia of Evolutionary Biology*. Elsevier Inc., 2016, pp. 277–286.

[12] N. Barton, A. Etheridge and A. Véber, 'The infinitesimal model: Definition, derivation, and implications', *Theoretical population biology*, vol. 118, pp. 50–73, 2017.

[13] J. B. S. Haldane and S. D. Jayakar, 'Polymorphism due to selection of varying direction', *Journal of Genetics*, vol. 58, no. 2, pp. 237–242, 1963.

[14] J. J. Bull, 'Evolution of phenotypic variance', *Evolution*, vol. 41, no. 2, pp. 303–315, 1987.

[15] T. J. DeWitt and S. M. Scheiner, *Phenotypic Plasticity: Functional and Conseptual Approaches*. Oxford University Press, 2003.

[16] P.-S. Laplace, 'Memoir on the probability of causes of events', *Statistical Science*, vol. 1, no. 3, pp. 366–367, 1774.

[17] A. Fog, 'Calculation methods for wallenius' noncentral hypergeometric distribution', *Communications in Statistics - Simulation and Computation*, vol. 37, no. 2, pp. 258–273, 2008.

[18] K. Kristensen, A. Nielsen, C. W. Berg, H. Skaug and B. M. Bell, 'Tmb: Automatic differentiation and laplace approximation', *Journal of Statistical Software*, vol. 70, no. 5, 2016.

[19] R. D. Neidinger, 'Introduction to automatic differentiation and matlab object-oriented programming', *SIAM Review*, vol. 52, no. 3, pp. 545–563, 2010.

[20] S. N. Wood, *Core Statistics*. Cambridge University Press, 2015.

[21] M. H. Protter and C. B. J. Morrey, *Intermediate Calculus (Second ed.)* Springer, 1985, pp. 421–426.

[22] B. Charlesworth, R. Lande and M. Slatkin, 'A neo-darwinian commentary on macroevolution', *Evolution; International Journal of Organic Evolution*, vol. 36, no. 3, pp. 474–498, 1982.

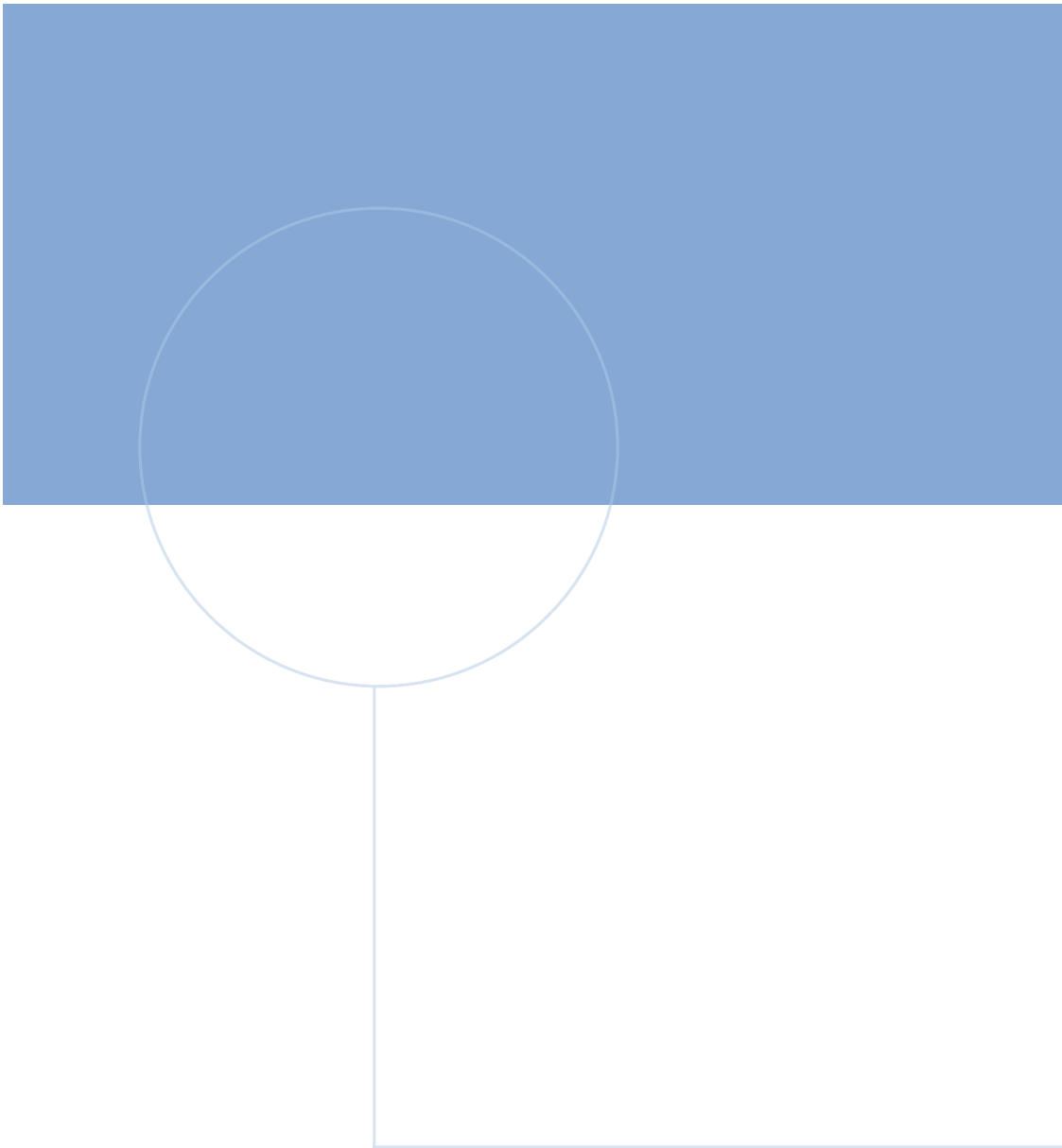
[23] J. Felsenstein, *Theoretical Evolutionary Genetics*. Department of genome sciences and department of biology, university of Washington, 2015.

[24] B. Bolker, *Ecological models and Data in R*. Princeton University Press, Princeton Oxford, 2008.

[25] S. Hélie, 'Understanding statistical power using noncentral probability distributions: Chi-squared, g-squared and anova', *Tutorials in quantitative methods for Psychology*, vol. 3, no. 2, pp. 63–69, 2007.

[26] R. Lande, 'Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation', *Journal of Evolutionary Biology*, vol. 22, no. 7, pp. 1435–1446, 2009.

[27] R. Lande and S. J. Arnold, 'The measurement of selection on correlated characters', *Evolution*, vol. 37, no. 6, pp. 1210–1226, 1983.



NTNU

Norwegian University of
Science and Technology