December 2016

Estimating the Selection Gradient of a Function-valued Trait

Tyler John Baur
University of Wisconsin-Milwaukee

ESTIMATING THE SELECTION GRADIENT OF A FUNCTION-VALUED TRAIT

by

Tyler J. Baur

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in

Mathematics

at

The University of Wisconsin-Milwaukee

December 2016

ABSTRACT

ESTIMATING THE SELECTION GRADIENT OF A FUNCTION-VALUED TRAIT

by

Tyler J. Baur

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Jay H. Beder

Kirkpatrick and Heckman initiated the study of function-valued traits in 1989. How to estimate the selection gradient of a function-valued trait is a major question asked by evolutionary biologists. In this dissertation, we give an explicit expansion of the selection gradient and construct estimators based on two different samples: one consisting of independent organisms (the independent case), and the other consisting of independent families of equally related organisms (the dependent case).

In the independent case we first construct and prove the joint consistency of sieve estimators of the mean and covariance functions of a Gaussian process, based on previous developments by Beder. From this we prove the consistency of the estimator of the selection gradient. This is supported by simulations. Using this estimator of the selection gradient, the estimated between-generation change in the mean phenotype is shown in simulations to be consistent.

In the dependent case we are able to estimate both the phenotypic and the genetic covariance functions. Simulations indicate consistency of these estimators, but appear not to support the consistency of the estimator of the selection gradient, nor of the estimator of the between-generation change in the mean phenotype. A probable source of this problem is identified.

To Kylee, Everett and Baby Baur

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

δ_t	Delta Function at t i.e. the Point Evaluation Functional at t
δ_{ij}	Kronecker Delta
$\ell_c^2(B)$	The subset of square summable sequences indexed by B: $\{\{a_k\}_k \in \ell^2(B) \colon \inf_k a_k > -1, k \in B\}$
$\langle \cdot, \cdot \rangle_K$	Inner Product in the RKHS $\mathcal{H}(K)$
$(\cdot, \cdot)_H$	Inner product in the Hilbert space H
$\mathfrak{P}$	Model
$\mathbb{N}$	Natural Numbers, $\mathbb{N} = 1, 2, 3, \ldots$
$\mathbb{R}$	Real Numbers
$\mathcal{H}(K)$	Reproducing Kernel Hilbert Space (RKHS) with Kernel K
P	Probability Measure
$\bar{z}$	Mean of the process $\{Z_t \colon t \in T\}$
H_{P}	Hilbert Space Associated to the Process $\{Z_t \colon t \in T\}$ under the measure P
P	Phenotypic Covariance function
CONS	Complete Orthonormal Sequence

ONS	Orthonormal Sequence

ACKNOWLEDGEMENTS

First of all, I cannot give enough thanks to my advisor Professor Jay Beder for his continued guidance and encouragement during my research. My appreciation for his willingness to spend time and share ideas to help me is beyond words. Without his help, this dissertation would have never gotten off of the ground. I would also like to thank the rest of my dissertation committee: Professor Jugal Ghorai, Professor Chao Zhu, Professor Kevin McLeod and Professor Daniel Gervini.

I would also like to thank the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee for providing support and an excellent place to study and develop as a Mathematician. The faculty has been very supportive of me; Dr. Eric Key deserves a special thank you for always being interested in what I was doing during my stay in the department.

Last but not least I would like to thank my parents and parents-in-law for always believing in me. My deepest thanks go to my wonderful wife, Kylee, and my ever-joyful son, Everett, for making me smile when I needed it most.

Chapter 1

Introduction

One of the most important topics in evolutionary biology is the way in which the physical traits, also called *phenotypes*, of an organism change between generations. The phenotype is often written as the sum of a genetic effect and an uncorrelated environmental effect. Together with selection, the genetic effect, called the *genotype*, determines the way that the phenotypes evolve.

If a trait is able to be entirely described by a vector in some $\mathbb{R}^n$, it is called *finite-dimensional*. The evolutionary change in the mean, the change in the mean trait between generations, is described by a well-known equation (3.3), the *Breeder's Equation*, in which the effects of selection are quantified by a vector called the *selection gradient*. However, many traits are better described by functions and are thus called *function-valued traits.* Gomulkiewicz and Beder [12] have been able to extend the selection gradient to function-valued traits. Furthermore, a version of the Breeder's Equation has also been extended [7] for use with such traits.

Our goal is to estimate the selection gradient for function-valued traits. To do so, we refine the definition of the selection gradient given in [12] and show that it lies in a certain weak completion, $\mathcal{M}$, of a subset of a set of square-integrable functions. Next, we prove several properties of the selection gradient and show that it has a unique expansion in $\mathcal{M}$. To

estimate the selection gradient, we must first estimate the covariance function of a Gaussian process. Using the Gaussian Dichotomy Theorem (see [20]), Beder [5, 6] gives explicit *sieve estimators* for the mean function of a Gaussian process with known covariance and also for the covariance function of a zero mean process. In most applications, zero mean cannot be assumed. We are able to use the Gaussian Dichotomy Theorem (GDT) to construct a joint estimator for the mean and covariance of a Gaussian process and prove several asymptotic properties.

One issue with using the GDT to estimate the covariance of a Gaussian process is that it requires independent observations. In many of the experiments performed by biologists, the pedigree of the individuals is purposefully selected and thus some individuals are related. When a organisms are equally related, such as siblings, we are able to transform the observations to independent ones and estimate the covariance functions of these new observations. This approach is able to estimate the original covariance function, but in many experiments, there are several unrelated families. With this in mind, we extend the estimation of P for a sample consisting of one family to one involving independent families, each consisting of equally-related organisms. From these estimates, we are able to create an estimator of the phenotypic covariance function and also of the genetic covariance function.

This dissertation is organized as follows. In Chapter 2 we review the concept of a reproducing kernel Hilbert space (RKHS) and the connection to a stochastic process. We also state the Gaussian Dichotomy Theorem in two parts and show how it may be used to estimate the covariance function of a stochastic process. Chapter 3 reviews biological traits, both finite-dimensional and function-valued, and introduces the (functional) selection gradient, β. In Chapter 4, we develop a joint estimator of the mean and covariance functions of a stochastic process and prove the consistency of the estimators. We also construct an estimator for the selection gradient and prove some consistency results, ending with a conjecture. Chapter 5 deals with estimation of a covariance function, genetic covariance function and the selection gradient based on a sample of independent families consisting of equally related organisms.

In Chapter 6, we conduct a simulation study for both the independent and dependent cases.

Chapter 2

Background on Sieve Estimation

2.1 Reproducing Kernel Hilbert Spaces

The concept of a reproducing kernel Hilbert space is paramount in estimating the selection gradient. In this section, we let $\mathcal{H}$ be a Hilbert space of real-valued functions on an arbitrary set T with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. We begin with a definition.

Definition 2.1.1. Let T be a set and K be a real-valued function on $T \times T$. K is called a *reproducing kernel* for the Hilbert space $\mathcal{H}$ if it satisfies the following two properties:

(i)

$$K_t = K(\cdot, t) \in \mathcal{H} \text{ for all } t \in T$$

(ii)

$$\langle f, K_t \rangle_{\mathcal{H}} = f(t) \text{ for all } t \in T \text{ and all } f \in \mathcal{H}. \tag{2.1}$$

The Hilbert space $\mathcal{H}$ is called the *reproducing kernel Hilbert space* (RKHS) with kernel K and is denoted by $\mathcal{H}(K, T)$ or $\mathcal{H}(K)$ when T is understood.

The property (2.1) is called the *reproducing property*. Note that if $\mathcal{H}$ is separable, then for a complete orthonormal sequence (CONS) $\{g_k\}_{k=1}^{\infty}$, it follows that for all $s, t \in T$,

$K_t(s) = \sum_{k=1}^{\infty} \langle g_k, K_t \rangle_{\mathcal{H}} g_k(s)$. By the reproducing property, we have $\langle g_k, K_t \rangle_{\mathcal{H}} = g_k(t)$. Therefore,

$$K(s,t) = \sum_{k=1}^{\infty} g_k(s) g_k(t) \text{ for all } s, t \in T. \tag{2.2}$$

2.1.1 Constructing Orthonormal Sets in a RKHS

A useful way to construct an orthonormal sequence (ONS), $\{g_1, g_2, \dots\}$, in $\mathcal{H}(K)$ is by choosing a sequence $t_1, t_2, \dots \in T$ such that $K_{t_1}, K_{t_2}, \dots$ are linearly independent and applying the Gram-Schmidt process. This gives us expressions

$$g_k = \sum_{i=1}^{k} q_{ki} K_{t_i}, \ k = 1, 2, 3, \dots \tag{2.3}$$

where the coefficients q_{ki} are computed from values of K. For example,

$$g_1 = \frac{1}{||K_{t_1}||} K_{t_1} = \frac{1}{\sqrt{K(t_1, t_1)}} K_{t_1}.$$

We will use this throughout the dissertation.

2.1.2 Stochastic Processes and Reproducing Kernel Hilbert Spaces

Let T be a set and K be a real-valued function on $T \times T$. K is called a *covariance kernel* if it is symmetric and nonnegative-definite, where symmetric means that for all $t, s \in T$,

$$K(t,s) = K(s,t)$$

and nonnegative-definite means that for all $n \in \mathbb{N}$, all $t_1, \dots, t_n \in T$ and all $a_1, \dots a_n$, we have

$$\sum_{i,j=1}^{n} a_i a_j K(t_j, t_i) \geq 0.$$

Definition 2.1.2. A real-valued *stochastic process* defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ is a family $\{Z_t \colon t \in T\}$ of real-valued random variables on Ω. The process is of p^{th} order if $\mathbb{E}_{\mathrm{P}}\left[|Z_t|^p\right] < \infty$ for all $t \in T$, where expectation is

$$\mathbb{E}_{\mathrm{P}}\left[\cdot\right] = \int_{\Omega} \cdot \quad d\mathrm{P}.$$

The process $\{Z_t \colon t \in T\}$ is called a *Gaussian process* if any finite linear combination of elements Z_t, $t \in T$ is a Gaussian random variable.

If a process is of second order, it has mean function $\mu(t) = \mathbb{E}_{\mathrm{P}}\left[Z_t\right]$ and covariance function $K(s,t) = \mathtt{Cov}_{\mathrm{P}}\left(Z_s, Z_t\right) = \mathbb{E}_{\mathrm{P}}\left[(Z_t - \mu(t))(Z_s - \mu(s))\right]$. Note that a second order process is contained in $L^2(\Omega, \mathcal{A}, \mathrm{P})$ and we may consider the closure of $\{Z_t \colon t \in T\}$, $H = H_{\mathrm{P}} = \overline{\{Z_t | t \in T\}}_{\mathrm{P}}$, in $L^2(\Omega, \mathcal{A}, \mathrm{P})$. Then H is the smallest Hilbert space contained in $L^2(\Omega, \mathcal{A}, \mathrm{P})$ that contains $\{Z_t \colon t \in T\}$ and is called the *Hilbert space spanned by the process* $\{Z_t | t \in T\}$. That is, H is the space of linear combinations of the random variables Z_t, $t \in T$ and their L^2-limits. In the case that $\{Z_t \colon t \in T\}$ is Gaussian, the space H will be called the *Gaussian space associated to* $\{Z_t \colon t \in T\}$.

Remark 2.1.1. From this definition, it is clear that we may write H as the $L^2(\Omega, \mathcal{A}, \mathrm{P})$ completion of the space V defined by

$$V = \left\{ X \in L^2(\Omega, \mathcal{A}, \mathrm{P}) : X = \sum_{i=1}^{n} c_i Z_{t_i} \text{ for some } n \in \mathbb{N}, c_1, \ldots, c_n \in \mathbb{R} \text{ and } t_1, \ldots, t_n \in T \right\}. \tag{2.4}$$

Now if K is the covariance function of a second-order process $\{Z_t \colon t \in T\}$, then K is symmetric as

$$\mathtt{Cov}_{\mathrm{P}}\left(Z_s, Z_t\right) = \mathtt{Cov}_{\mathrm{P}}\left(Z_t, Z_s\right), \quad \forall s, t \in T.$$

Furthermore, if $n \in \mathbb{N}$, $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and $t_1, t_2, \ldots, t_n \in T$, then

$$\sum_{i,j=1}^{n} a_j a_i K(t_i, t_j) = \mathtt{Var}_{\mathrm{P}} \left(\sum_{i=1}^{n} a_i Z_{t_i} \right) \geq 0.$$

Therefore, K is also non-negative definite and is thus a covariance kernel. We also note that if K is the covariance function of the process $\{Z_t \colon t \in T\}$, then $\{Z_t \colon t \in T\}$ is of second order. To connect a Gaussian process and a RKHS, we use the following proposition:

Proposition 2.1.1 ([20, Proposition 3.2])**.** *Let T be a set and $\{Z_t \colon t \in T\}$ be a zero mean Gaussian process with covariance function K and let H be the Hilbert space spanned by $\{Z_t \colon t \in T\}$. There is a unique reproducing kernel Hilbert space with kernel K, $\mathcal{H}(K)$, consisting of functions on T. Moreover, the Hilbert spaces H and $\mathcal{H}(K)$ are isomorphic via the isometry $\Lambda : H \to \mathcal{H}(K)$ defined by*

$$\Lambda(Y)[t] = \mathbb{E}_P \left[Y Z_t \right]. \tag{2.5}$$

The mapping Λ in (2.5) has been called the *Loève map* [6, 12]. If $(\cdot, \cdot)_{\mathrm{P}}$ is the inner product in H, then it follows that the inner product in $\mathcal{H}(K)$ can be written as

$$\langle f, g \rangle_K = \left(\Lambda^{-1} f, \Lambda^{-1} g \right)_{\mathrm{P}}, \text{ for all } f, g \in \mathcal{H}(K) \tag{2.6}$$

Furthermore, note that we are able to write

$$\Lambda(Z_s)[\cdot] = \mathbb{E}_{\mathrm{P}} \left[Z_s Z_\cdot \right] = K(s, \cdot) = K_s(\cdot). \tag{2.7}$$

We will use (2.7) throughout the dissertation to construct an orthonormal set in H: if an ONS $\{g_k\}$ in $\mathcal{H}(K)$ is given by (2.3), i.e. ,

$$g_k = \sum_{i=1}^{k} q_{ki} K_{t_i},$$

for all k, then the sequence $\{U_k\}$ defined by

$$U_k = \sum_{i=1}^{k} q_{ki} Z_{t_i}$$

is an ONS in H.

Remark 2.1.2. The inner products on H_{P} and $\mathcal{H}(K)$ shall be denoted by $(\cdot,\cdot)_{\mathrm{P}}$ and $\langle\cdot,\cdot\rangle_K$, respectively. When it is clear, the subscripts may be dropped.

2.2 The Gaussian Dichotomy Theorem

The Gaussian Dichotomy Theorem (GDT) is an essential tool used in the estimation of both the mean and covariance of a Gaussian process. The GDT says roughly that two Gaussian measures are either equivalent (mutually absolutely continuous) or mutually singular. We are particularly interested in the treatment of the GDT given by Neveu [20]. The theorem is stated in two parts, one where the two measures have the same covariance and the other when the measures both have mean zero and different covariances. Since the equivalence of measures is an equivalence relation on sets of measures, it is enough to consider both of these cases to arrive at the equivalence or singularity of any two Gaussian measures.

Before stating the GDT, we shall first introduce the concept of a σ-algebra generated by a Gaussian space. To any Gaussian space H in $L^2(\Omega, \mathcal{A}, \mathrm{P})$ is associated the sub-σ-algebra $\mathcal{B}(H)$ of $\mathcal{A}$ generated by the random variables that belong to H. This space $\mathcal{B}(H)$ is very closely connected to the process $\{Z_t \colon t \in T\}$ as shown in the following lemma [20, Lemma 2.3]:

Lemma 2.2.1. *If $\{Z_t \colon t \in T\}$ is the Gaussian process that generates the Gaussian space $H \subset L^2(\Omega, \mathcal{A}, \mathrm{P})$, then the σ-algebra $\mathcal{B}(H)$ coincides with the σ-algebra $\overline{\mathcal{B}\,(Z_t, t \in T)}$ generated by the random variables Z_t, $t \in T$ and the* P*-negligible sets.*

We now are able to discuss the first part of the GDT given in [20, Proposition 8.1].

2.2.1 The Gaussian Dichotomy Theorem, Part 1

Theorem 2.2.1. *Let $\{Z_t : t \in T\}$ be a zero mean Gaussian process defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. Let K be the covariance function of this process and let H be the Gaussian space associated to it. We assume that $\mathcal{A} = \mathcal{B}(H)$.*

Any probability measure Q on $(\Omega, \mathcal{A})$ such that $\{Z_t : t \in T\}$ is a Gaussian process with covariance function K is either singular to the probability measure P or equivalent to P. For the measures Q and P to be equivalent, it is necessary and sufficient that there exist a Y in H such that

$$\mathbb{E}_{\mathrm{Q}}[Z_t] = \mathbb{E}_{\mathrm{P}}[YZ_t] \text{ for all } t \in T. \tag{2.8}$$

Equivalently, it is necessary and sufficient that the mean $\mu_{\mathrm{Q}}(\cdot) = \mathbb{E}_{\mathrm{Q}}[Z_\cdot]$ of the process $\{Z_t : t \in T\}$ belong to the reproducing kernel Hilbert space $\mathcal{H}(K)$. Furthermore, when these conditions are fulfilled, the Radon-Nikodym derivative of Q with respect to P is

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \exp\left\{Y - \frac{1}{2}\mathbb{E}_{\mathrm{P}}\left[Y^2\right]\right\} \tag{2.9}$$

on $(\Omega, \mathcal{A}, \mathrm{P})$.

Conversely, for any $Y \in H$, the probability measure Q on $(\Omega, \mathcal{A})$ defined by

$$d\mathrm{Q} = \exp\left\{Y - \frac{1}{2}\mathbb{E}_{\mathrm{P}}\left[Y^2\right]\right\} d\mathrm{P}$$

makes the process $\{Z_t : t \in T\}$ Gaussian; the mean of this process is $\mathbb{E}_{\mathrm{Q}}[Z_t] = \mathbb{E}_{\mathrm{P}}[YZ_t]$ and the process has covariance K.

This theorem is useful since it states exactly what the Radon-Nikodym derivative (2.9) is and tells us how to find Y. This derivative is the relative density, and allows us to use maximum likelihood estimation. The theorem also tells us that the parameter set for μ_{Q} is the RKHS $\mathcal{H}(K)$.

Before stating the second half of the GDT, we must first introduce the necessary material

found in [20, pages 109-138].

2.2.2 Tensor Products of Hilbert Spaces

Let E_1 and E_2 be two vector spaces. We begin by taking the closure, $E_1 \circ E_2$, of the vector space of linear combinations of elements of $E_1 \times E_2$. If $f_1 \circ f_2$ denotes an arbitrary element of $E_1 \times E_2$, then the elements of $E_1 \circ E_2$ can be written in the form $\sum_{k=1}^{n} c_k f_1^k \circ f_2^k$, where $c_k \in \mathbb{R}$, $f_1^k \in E_1$ and $f_2^k \in E_2$. Now, denote by N the sub-vector space of $E_1 \circ E_2$ that is generated by the elements that are of one of the following forms:

$$\left(\sum_{k=1}^{n} c_k f_1^k\right) \circ f_2 - \sum_{k=1}^{n} c_k f_1^k \circ f_2,$$

$$f_1 \circ \left(\sum_{k=1}^{n} c_k f_2^k\right) - \sum_{k=1}^{n} c_k f_1 \circ f_2^k.$$

The *tensor product of* E_1 *and* E_2, $E_1 \otimes E_2$, is defined as the quotient $E_1 \circ E_2 / N$. We denote by $f_1 \otimes f_2$ the equivalence class of $f_1 \circ f_2$ in $E_1 \otimes E_2$; an arbitrary element of $E_1 \otimes E_2$ can be written as $\sum_{k=1}^{n} c_k f_1^k \otimes f_2^k$.

When E_1 and E_2 are vector spaces of functions, the following proposition [20, Proposition 6.2] will be useful in understanding the general elements of $E_1 \otimes E_2$.

Proposition 2.2.1. *If E_1 and E_2 are two vector spaces of real-valued functions defined on the sets T_1 and T_2, respectively, then the tensor product $E_1 \otimes E_2$ is isomorphic to the vector space of functions on $T_1 \times T_2$ generated by the functions $f_1 \otimes f_2$ defined by $f_1 \otimes f_2(t_1, t_2) = f_1(t_1) f_2(t_2)$, where $f_1 \in E_1$, $f_2 \in E_2$, $t_1 \in T_1$ and $t_2 \in T_2$.*

To define the tensor product of two Hilbert spaces, we need to relate the tensor product and the inner products of two inner product spaces.

Lemma 2.2.2 (Lemma 6.3,[20])**.** *If E_1 and E_2 are two separable inner product spaces with inner products $(\cdot,\cdot)_1$ and $(\cdot,\cdot)_2$, respectively, then $E_1 \otimes E_2$ is a separable inner product space*

with inner product defined by

$$(f_1 \otimes f_2, f_1' \otimes f_2') = (f_1, f_1')_1 (f_2, f_2')_2 . \tag{2.10}$$

Using this lemma, we may then define the *Hilbert space tensor product of two Hilbert spaces* H_1 *and* H_2 as the completion of the space $H_1 \otimes H_2$ with respect to the inner product (2.10). To denote this Hilbert space, we use a slight abuse of notation: when the context is clear, $H_1 \otimes H_2$ will denote the Hilbert space tensor product of H_1 and H_2. We can also write a complete orthonormal sequence (CONS) of $H_1 \otimes H_2$ given two CONS of H_1 and H_2 [20, Lemma 6.5]. If $\{f_i, i \in I\}$ and $\{g_j, j \in J\}$ are CONS of H_1 and H_2, then $\{f_i \otimes g_j, (i,j) \in I \times J\}$ is a CONS of $H_1 \otimes H_2$.

As developed here, the tensor product of two functions $f \otimes g$ is more of an algebraic construct. If E_1 and E_2 are function spaces, we will also use the notation $f_1 \otimes f_2$ for the function on $T_1 \times T_2$ defined by

$$f_1 \otimes f_2(t_1, t_2) = f_1(t_1) f_2(t_2), \tag{2.11}$$

where $f_1 \in E_1$, $f_2 \in E_2$, $t_1 \in T_1$ and $t_2 \in T_2$. This is justified by the following [20, Proposition 6.2]:

Proposition 2.2.2. *If* E_1 *and* E_2 *are two vector spaces of real-valued functions defined on the sets* T_1 *and* T_2*, respectively, then the tensor product* $E_1 \otimes E_2$ *is isomorphic to the vector space generated by functions on* $T_1 \otimes T_2$ *of the form (2.11).*

Using this proposition and a slight abuse of notation, when the context is clear, $f \otimes g$ will denote the image of the tensor product $f \otimes g$ and is a function on $T_1 \times T_2$.

We may extend the tensor product to any finite number of Hilbert spaces $H_1 \otimes \cdots \otimes H_n$, with inner product

$$(f_1 \otimes \cdots \otimes f_n, f_1' \otimes \cdots \otimes f_n') = (f_1, f_1')_{H_1} \ldots (f_n, f_n')_{H_n} .$$

We will be mostly concerned with the case where $H_i = H$ for all i. In this case, $H^{n\otimes}$ will denote the n-fold tensor product of H with itself. The elements $f \otimes \cdots \otimes f$ of $H^{n\otimes}$ are written as $f^{n\otimes}$ and the inner product of $H^{n\otimes}$ is given by

$$\left(f^{n\otimes}, g^{n\otimes}\right)_{n\otimes} = \left((f,g)_H\right)^n.$$

Let $\mathfrak{S}_n$ be the nth symmetric group. For any $\sigma \in \mathfrak{S}_n$, there exists [20, Lemma 6.9] a unique invertible operator , U_σ, on $H^{n\otimes}$ such that

$$U_\sigma[f_1 \otimes \cdots \otimes f_n] = f_{\sigma_1} \otimes \cdots \otimes f_{\sigma_n}. \tag{2.12}$$

Using this equation, we may define a special subspace of $H^{n\otimes}$.

Definition 2.2.1. Let H be a Hilbert space and $n \in \mathbb{N}$. The n^{th} *symmetric tensor product of* H is the subspace of $H^{n\otimes}$ defined by

$$H^{n\odot} = \{f | f \in H^{n\otimes}, U_\sigma f = f \text{ for any } \sigma \in \mathfrak{S}_n\}.$$

If $f_1, \ldots, f_n$ are in H, we define

$$f_1 \odot \cdots \odot f_n = \frac{1}{\sqrt{n!}} \sum_\sigma f_{\sigma_1} \otimes \cdots \otimes f_{\sigma_n}.$$

From this definition, we see that $f_1 \odot \cdots \odot f_n \in H^{n\odot}$ when $f_i \in H$ for $i = 1, \ldots, n$. The inner product of elements of $H^{n\odot}$ of the form $f_1 \odot \cdots \odot f_n$ is given by

$$(f_1 \odot \cdots \odot f_n, f_1' \odot \cdots \odot f_n')_{n\odot} = \sum_\sigma \left(f_1, f_{\sigma_1}'\right)_H \ldots \left(f_n, f_{\sigma_n}'\right)_H. \tag{2.13}$$

In particular, when $n = 2$, we have

$$(f_1 \odot f_2, f_1' \odot f_2')_{2\odot} = (f_1, f_1')_H (f_2, f_2')_H + (f_1, f_2')_H (f_2, f_1')_H. \tag{2.14}$$

If H_1 and H_2 are Hilbert spaces, let H_1' denote the dual space of H_1, and if $\tilde{U} : H_1 \to H_2'$ is a linear operator, let $\tilde{U}_h$ denote $\tilde{U}(h)$ for all $h \in H_1$. In order to state the GDT, we need the following proposition [20, Proposition 6.16]:

Proposition 2.2.3. *Let H_1 and H_2 be two Hilbert spaces. To any element U of $H_1 \otimes H_2$ corresponds a unique continuous linear operator $\tilde{U}$ of H_1 into H_2' such that*

$$\tilde{U}_{h_1}(h_2) = (h_1 \otimes h_2, U)_{H_1 \otimes H_2}, \quad \forall h_1 \in H_1, h_2 \in H_2,$$

where $\tilde{U}_{h_1}$ is the functional in H_2' that maps $h_2 \in H_2$ to $(h_1 \otimes h_2, U)_{H_1 \otimes H_2}$. This operator satisfies

$$\sum_I ||\tilde{U}_{h_i}||^2_{H_2'} = ||U||^2_{H_1 \otimes H_2} < \infty, \tag{2.15}$$

for any complete orthonormal sequence $\{h_i, i \in I\}$ of H_1.

When the spaces H_1 and H_2 are equal, say $H_1 = H_2 = H$, the tensor $U \in H^{2\otimes}$ is symmetric, that is, belongs to $H^{2\odot}$ if and only if the operator $\tilde{U}$ of H into H' is self-adjoint.

Because of (2.15), the operators $\tilde{U}$ of H_1 into H_2' associated to an element U of $H_1 \otimes H_2$ are Hilbert-Schmidt operators. The Hilbert-Schmidt norm of the operator is the common value of the expressions $\left[\sum_I ||\tilde{U}_{h_i}||^2_{H_2'}\right]^{1/2}$, where $\{h_i, i \in I\}$ is a CONS of H_1. As Hilbert spaces are self-dual, it should be understood that for $g \in H_2$, the equation $\tilde{U}_h = g$ means that g is the corresponding operator in H_2'.

Of further interest is how to write an arbitrary element of the space $H_1 \otimes H_2$ or an element of $H_1^{2\odot}$. We turn to [20, Proposition 6.18]

Proposition 2.2.4. *Let H_1 and H_2 be two Hilbert spaces. Any element $U \in H_1 \otimes H_2$ can be written in the form*

$$U = \sum_I a_i f_i \otimes g_i,$$

where $\{f_i, i \in I\}$ and $\{g_i, i \in I\}$ are complete orthonormal sequences of H_1 and H_2, respectively, and $\{a_i, i \in I\}$ is a sequence of scalars such that $\sum_I |a_i|^2 < \infty$. The Hilbert-Schmidt

operator $\tilde{U}$ *of* H_1 *into* H_2' *associated to* U *satisfies* $\tilde{U}_{f_i} = a_i g_i$ *and* $\tilde{U}'_{g_i} = a_i f_i$ *for* $i \in I$.

When $H = H_1 = H_2$ *and* U *is symmetric, we may write*

$$U = \frac{1}{2} \sum_I a_i f_i^{2\odot}, \tag{2.16}$$

for a complete orthonormal sequence $\{f_i, i \in I\}$ *of* H. *In addition,* $\tilde{U}$ *satisfies*

$$\tilde{U}_{f_i} = a_i f_i, \quad \forall i \in I. \tag{2.17}$$

This proposition says that we may write any $U \in H^{2\odot}$ in the form (2.16) and that the coefficients of this expansion are the eigenvalues of the associated linear operator $\tilde{U}$.

To fully understand the consequences of the GDT, we end this subsection with two results that allow us to write a symmetric tensor $U^{2\odot}$ in terms of the random variable U. We begin with [20, Proposition 7.3]:

Proposition 2.2.5. *Let* H *be a Gaussian space in* $L^2(\Omega, \mathcal{A}, \mathrm{P})$ *and let* $\mathcal{B}(H)$ *be the sub-*σ*-algebra of* $\mathcal{A}$ *generated by* H.

There exists a unique isomorphism ϕ *of the Hilbert space direct sum* $\oplus_{n\geq 0} H^{n\odot}$ *onto* $L^2(\Omega, \mathcal{B}(H), \mathrm{P})$ *such that for each* $X \in H$,

$$\phi\left[\exp \odot(X)\right] = \exp\left[X - \frac{1}{2}\mathbb{E}_{\mathrm{P}}\left[X^2\right]\right], \tag{2.18}$$

where $\exp \odot(X) = \sum_{n\geq 0} \frac{1}{n!} X^{n\odot}$ *in* $\oplus_{n\geq 0} H^{n\odot}$.

The next proposition [20, Proposition 7.5] will allow us write $\phi(X^{2\odot})$ in terms of X:

Proposition 2.2.6. *If* H *is a Gaussian space in* $L^2(\Omega, \mathcal{A}, \mathrm{P})$ *and* $X \in H$, *then the image under* ϕ *of* $X^{2\odot}$ *is given by*

$$\phi(X^{2\odot}) = X^2 - \mathbb{E}_{\mathrm{P}}\left[X^2\right]. \tag{2.19}$$

In particular, if U is a standard normal random variable under P, we have

$$\phi(U^{2\odot}) = U^2 - 1. \tag{2.20}$$

By abuse of notation, we shall write $U^{2\odot} = U^2 - 1$.

We are now able to conclude the Gaussian Dichotomy Theorem, where we consider two measures that give the process $\{Z_t \colon t \in T\}$ zero mean and different covariances.

2.2.3 The Gaussian Dichotomy Theorem, Part 2

The second part of the GDT is given in [20, Proposition 8.6]

Theorem 2.2.2. *Let $\{Z_t \colon t \in T\}$ be a real-valued zero mean Gaussian process on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and such that $\mathcal{A} = \mathcal{B}\,(\{Z_t \colon t \in T\})$; denote by K_{P} and H, respectively, the covariance and the Gaussian space associated to this process. Let Q be a second probability measure on $(\Omega, \mathcal{A})$ for which $\{Z_t \colon t \in T\}$ is a zero mean Gaussian process; denote by K_{Q} the covariance of $\{Z_t \colon t \in T\}$ with respect to Q.*

Then the probability measures P and Q are either singular or they are equivalent. For these measures to be equivalent, it is necessary and sufficient that there exist an element U in $H^{2\odot}$ such that

$$K_{\mathrm{Q}}(s,t) - K_{\mathrm{P}}(s,t) = (Z_s \odot Z_t, U)_{2\odot}\,, \qquad s,t \in T, \tag{2.21}$$

and furthermore, that the Hilbert-Schmidt operator $\tilde{U}$ associated to H has eigenvalues strictly greater than -1. A condition equivalent to (2.21) is

$$K_{\mathrm{Q}}(s,t) - K_{\mathrm{P}}(s,t) = \sum_k a_k g_k(s) g_k(t) \quad \forall s,t \in T, \tag{2.22}$$

where $\{g_k\}$ is an ONS in $\mathcal{H}(K)$.

When the probability measures P and Q are equivalent, the Gaussian space H_{Q} generated by the process $\{Z_t \colon t \in T\}$ in $L^2(\Omega, \mathcal{A}, \mathrm{Q})$ coincides with the vector space H equipped with

the inner product

$$(Y_1, Y_1)_{H_{\mathrm{Q}}} = (Y_1, Y_2)_H + (Y_1 \odot Y_2, U)_{2^\odot}, \qquad Y_1, Y_2 \in H. \tag{2.23}$$

Furthermore, the function spaces $\mathcal{H}(K_{\mathrm{P}})$ *and* $\mathcal{H}(K_{\mathrm{Q}})$ *are composed of the same functions.*[1]

When the probability measures P *and* Q *are equivalent, the Radon-Nikodym derivative of* Q *with respect to* P *equals*

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \frac{\exp(X)}{\mathbb{E}_{\mathrm{P}}[\exp(X)]}, \tag{2.24}$$

where X *denotes the element of* $H^{2^\odot}$ *defined by*

$$X = \frac{1}{2}\sum_k \lambda_k U_k^{2^\odot} \text{ if } U = \frac{1}{2}\sum_k a_k U_k^{2^\odot} \text{ and } (1-\lambda_k)(1+a_k) = 1, \tag{2.25}$$

and $\{U_k\}_k$ *is an orthonormal sequence (ONS) in* H *such that* $\Lambda_{\mathrm{P}} U_k = g_k$*, where* $\{g_k\}$ *is given in (2.22).*

Proposition 2.2.4 and this theorem tell us that the coefficients $\{a_k\}$ are strictly greater than -1, as $\{a_k\}$ is the set of eigenvalues of the operator $\tilde{U}$. Furthermore, the sequence $\{a_k\}$ satisfies $\sum_k a_k^2 < \infty$.

Proof of (2.22). Let $\{U_k\}$ be the ONS in H given after (2.25) and let $g_k = \Lambda_{\mathrm{P}}(U_k)$ for all k. It follows from equation (2.6), that $\{g_k\}$ is an ONS in $\mathcal{H}(K_{\mathrm{P}})$. Using equation (2.14) the

[1]The inner product in $\mathcal{H}(K_{\mathrm{Q}})$ is given terms of those in $\mathcal{H}(K_{\mathrm{P}})$ and $\mathcal{H}(K_{\mathrm{P}})^{2^\odot}$ by the formula

$$\langle h_1, h_2\rangle_{\mathcal{H}(K_{\mathrm{Q}})} = \langle h_1, h_2\rangle_{\mathcal{H}(K_{\mathrm{P}})} + \langle h_1 \odot h_2, K_{\mathrm{Q}} - K_{\mathrm{P}}\rangle_{\mathcal{H}(K_{\mathrm{P}})^{2^\odot}}.$$

right hand side of (2.21) becomes

$$\begin{aligned}
(Z_s \odot Z_t, U)_{2\odot} &= \left(Z_s \odot Z_t, \frac{1}{2}\sum_k a_k U_k^{2\odot} \right)_{2\odot} \\
&= \frac{1}{2}\sum_k a_k \left(Z_s \odot Z_t, U_k^{2\odot} \right)_{2\odot} \\
&= \frac{1}{2}\sum_k a_k \left((Z_s, U_k)_H (Z_t, U_k)_H + (Z_s, U_k)_H (Z_t, U_k)_H \right) \\
&= \sum_k a_k (Z_s, U_k)_H (Z_t, U_k)_H \\
&= \sum_k a_k \Lambda_{\mathrm{P}}(U_k)[s] \Lambda_{\mathrm{P}}(U_k)[t] \\
&= \sum_k a_k g_k(s) g_k(t).
\end{aligned}$$

Therefore, we may write

$$K_{\mathrm{Q}}(s,t) - K_{\mathrm{P}}(s,t) = \sum_k a_k g_k(s) g_k(t) \quad \forall s,t \in T.$$

□

This theorem gives a form for the Radon-Nikodym derivative of Q with respect to P, namely

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \frac{\exp(X)}{\mathbb{E}_{\mathrm{P}}[\exp(X)]},$$

where $X = \frac{1}{2}\sum_k \lambda_k U_k^{2\odot}$ for some ONS $\{U_k\}$ in H. Similar to the first part of the GDT, this derivative is the relative density and allows us to use maximum likelihood estimation. The denominator of this derivative is given [20, Proposition 8.5] by

$$\mathbb{E}_{\mathrm{P}}[\exp(X)] = \exp\left\{ -\frac{1}{2}\sum_k \lambda_k \right\} \prod_k (1-\lambda_k)^{-1/2}.$$

Therefore,

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \exp\left\{\frac{1}{2}\sum_k \left(\lambda_k U_k^{2\odot} + \ln(1-\lambda_k) + \lambda_k\right)\right\}.$$

To deal with the $U_k^{2\odot}$ term, we use the discussion after Proposition 2.2.6 to obtain $U_k^{2\odot} = U_k^2 - 1$. This implies that

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \exp\left\{\frac{1}{2}\sum_k \left(\lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\} \tag{2.26}$$

Finally, since we know that the $\{U_k\}$ are i.i.d. standard normal random variables under P (they are orthonormal in H with respect to P) and $\mathbb{E}_{\mathrm{Q}}[Z_t] = 0$ for all $t \in T$, we can gain information about the distribution under Q of the $\{U_k\}$.

Proposition 2.2.7. *Under the hypotheses of Theorem 2.2.2, we have*

$$\mathbb{E}_{\mathrm{Q}}[U_k] = 0 \tag{2.27}$$

and

$$\mathtt{Cov}_{\mathrm{Q}}(U_i, U_j) = \mathbb{E}_{\mathrm{Q}}[U_i U_j] = \delta_{ij}(1+\lambda_j), \tag{2.28}$$

where δ_{ij} is the Kronecker δ.

Proof. Let V be as in (2.4), the linear span of $\{Z_t : t \in T\}$. Let X be an element of H_{Q}, so that X is either an element of V or an $L^2(\Omega, \mathcal{A}, \mathrm{Q})$ limit of a sequence in V. It is clear that if $X \in V$ then $\mathbb{E}_{\mathrm{Q}}[X] = 0$. If $X \in H_{\mathrm{Q}} \backslash V$, then X there exists a sequence $\{X_n, n \in \mathbb{N}\}$ in V such that $\mathbb{E}_{\mathrm{Q}}[(X_n - X)^2] \to 0$ as $n \to \infty$. By the Cauchy-Schwarz inequality,

$$\left(\mathbb{E}_{\mathrm{Q}}[X_n - X]\right)^2 \leq \mathbb{E}_{\mathrm{Q}}\left[(X_n - X)^2\right],$$

and thus $\mathbb{E}_{\mathrm{Q}}[X_n] \to \mathbb{E}_{\mathrm{Q}}[X]$. Now $\mathbb{E}_{\mathrm{Q}}[X_n] = 0$ for all n, which implies that $\mathbb{E}_{\mathrm{Q}}[X] = 0$. Since $U_j \in H_{\mathrm{Q}}$, it follows that U_j has zero mean for all j. This proves (2.27).

Similarly by (2.23), we have

$$\begin{aligned}
(U_i, U_j)_{\mathrm{Q}} &= (U_j, U_j)_{\mathrm{P}} + (U_i \odot U_j, U)_{2\odot} \\
&= \delta_{ij} + \frac{1}{2}\sum_k \lambda_k \left(U_i \odot U_j, U_k^{2\odot}\right)_{2\odot} \\
&= \delta_{ij} + \frac{1}{2}\sum_k \lambda_k ((U_i, U_k)_{\mathrm{P}} (U_j, U_k)_{\mathrm{P}} + (U_i, U_k)_{\mathrm{P}} (U_j, U_k)_{\mathrm{P}} \\
&= \delta_{ij} + \frac{1}{2}\sum_k \lambda_k 2((U_i, U_k)_{\mathrm{P}} (U_j, U_k)_{\mathrm{P}} \\
&= \delta_{ij} + \lambda_j \delta_{ij} \\
&= \delta_{ij}(1 + \lambda_j),
\end{aligned}$$

which proves (2.28). □

This proposition shows that under Q, U_i and U_j are uncorrelated (and therefore independent) if $i \neq j$ and U_j has variance $1 + \lambda_j$.

2.3 Estimation of the Covariance Function

In this section, we will define and discuss the need for a sieve estimator. We shall also review the properties of such an estimator of a covariance function of a Gaussian process given in [6].

Let $\{Z_t \colon t \in T\}$ be a stochastic process on a measurable space $(\Omega, \mathcal{A})$ and let

$$\ell_c^2(B) = \{\{a_\alpha\} \in \ell^2(B) \colon \inf(a_\alpha) > -1\}, \tag{2.29}$$

where B is usually a finite set or $\mathbb{N}$. Recall that $f \otimes g$ will denote a function on $T \times T$ defined by $f \otimes g(s,t) = f(s)g(t)$. In light of the Gaussian Dichotomy Theorem, we will consider the largest set $\mathfrak{P}$ of probability measures on $(\Omega, \mathcal{A})$ such that

(A1) the process is Gaussian under every $\mathrm{Q} \in \mathfrak{P}$,

(A2) the mean function of the process is identically zero under every $\mathrm{Q} \in \mathfrak{P}$,

(A3) all of the measures in $\mathfrak{P}$ are equivalent (mutually absolutely continuous), and

(A4) the true probability measure belongs to $\mathfrak{P}$.

If a measure is arbitrarily chosen from the model $\mathfrak{P}$, say P, then $\{Z_t \colon t \in T\}$ is a zero mean Gaussian process defined on $(\Omega, \mathcal{A})$ with covariance K_{P} and has the associated Gaussian space H_{P}. Let the RKHS with kernel K_{P} be denoted by $\mathcal{H}_{\mathrm{P}} = \mathcal{H}(K_{\mathrm{P}}, T)$ and the Loève map between H_{P} and $\mathcal{H}_{\mathrm{P}}$ by Λ_{P}.

The n-fold product measure $\mathrm{Q} \times \cdots \times \mathrm{Q}$ will be denoted by $\mathrm{Q}^{n\otimes}$ and $\mathcal{A}^{n\otimes}$ will denote the σ-algebra generated by sets of the form $B_1 \times \cdots \times B_n$, where $B_i \in \Omega$ for $i = 1, \ldots, n$. We will also let $\mathfrak{P}^{(n)} = \{\mathrm{Q}^{n\otimes}, \mathrm{Q} \in \mathfrak{P}\}$.

According to the Gaussian Dichotomy Theorem (GDT), for each $\mathrm{Q} \in \mathfrak{P}$,

(i) There are a countable orthonormal sequence $\{g_k\}$ in $\mathcal{H}_{\mathrm{P}}$ and a sequence $\mathbf{a} = \{a_k\} \in \ell_c^2$, both depending on Q, such that

$$K_{\mathrm{Q}} = K_{\mathrm{P}} + \sum_k a_k g_k \otimes g_k. \tag{2.30}$$

This means that we may write the set of possible covariances as specified by the model, $\mathcal{C}$, as

$$\mathcal{C} = \left\{ K_{\mathrm{Q}} = K_{\mathrm{P}} + \sum_k a_k g_k \otimes g_k, \mathbf{a} \in \ell_c^2(\mathbb{N}), \{g_k\} \text{ countable and orthonormal in } \mathcal{H}_{\mathrm{P}} \right\} \tag{2.31}$$

(ii) There are a countable orthonormal sequence $\{U_k\}$ in H_{P} and a sequence $\boldsymbol{\lambda} = \{\lambda_k\}$ with $-\boldsymbol{\lambda} \in \ell_c^2$ such that

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \exp\left\{ \frac{1}{2} \sum_k \left(\lambda_k U_k^2 + \ln(1 - \lambda_k) \right) \right\}, \tag{2.32}$$

where $(1+a_k)(1-\lambda_k)=1$ and $g_k=\Lambda_{\mathrm{P}}(U_k)$ for all k.

(iii) Furthermore,

$$\frac{d\mathrm{Q}^{n\otimes}}{d\mathrm{P}^{n\otimes}}=\exp\left\{\frac{n}{2}\sum_k\left(\lambda_k S_k^2+\ln(1-\lambda_k)\right)\right\}, \tag{2.33}$$

where $S_k^2=\frac{1}{n}\sum_{i=1}^n U_{ki}^2$.

(iv) Under Q, the sequence

$$\{U_k/\sqrt{1+a_k}\}_k \text{ is i.i.d. } N(0,1). \tag{2.34}$$

Given n realizations of this process, the variable U_{ki} is the value of the ith realization of U_k for all k. We wish to maximize (2.33) by fixing $\omega\in\Omega^n$ and allowing $\{S_k^2, k\in B\}$ and $-\boldsymbol{\lambda}\in\ell_c^2(B)$ to vary. We begin by fixing $\{g_k\}$ in $\mathcal{H}(P_0)$ and the corresponding set $\{U_{k,i}\}_{k,i}$. The following lemma is due to [6, Theorem 3.1]:

Lemma 2.3.1. *Fix $\{U_{k,i}\}_{k,i}$ for $k\in B$ and $i=1,\ldots,n$. If B is an infinite set, the likelihood (2.33) is unbounded almost surely $\mathfrak{P}^{(n)}$. On the other hand, when B is a finite set, then this likelihood is maximized at*

$$\hat{\lambda}_k=\begin{cases}1-S_k^{-2} & \text{if } k\in B\\ 0 & \text{otherwise}\end{cases}, \tag{2.35}$$

with corresponding estimate

$$\hat{a}_k=\begin{cases}S_k^2-1 & \text{if } k\in B\\ 0 & \text{otherwise}\end{cases}. \tag{2.36}$$

2.3.1 Sieve Estimation

Since (2.33) is unbounded over an infinite set and can be maximized over a finite set, we shall consider the method of sieves. Let $\mathcal{D}=\{\mathrm{P}_\theta,\theta\in\Theta\}$ be a dominated family of probability measures, that is, the densities of the measures exist with respect to some measure.

Definition 2.3.1. A *sieve* in Θ is a collection $\{\mathcal{S}_d\}$ of subsets of Θ indexed by a parameter d such that

(i) $d' > d \Rightarrow \mathcal{S}_{d'} \supset \mathcal{S}_d$,

(ii) $\cup \mathcal{S}_d$ is dense in Θ, and

(iii) the likelihood can be maximized at $\hat{\theta}_d$ over each $\mathcal{S}_d$ for some sample size n.

This estimator $\hat{\theta}_d$ over each $\mathcal{S}_d$ is called a *sieve estimator of* θ and d is called the *sieve parameter/size*.

The parameter space $\mathcal{C}$ in (2.31) involves not only letting the coefficients $\mathbf{a}$ vary, but also the CONS $\{g_k\}_k$ in $\mathcal{H}(K_{\mathrm{P}})$. This space is very large and it is difficult to find a sieve in $\mathcal{C}$. This leads us to consider a subset, $\mathcal{C}_0$ of $\mathcal{C}$ defined by

$$\mathcal{C}_0 = \left\{ K_{\mathrm{Q}} \in \mathcal{C} \colon K_{\mathrm{Q}} = K_{\mathrm{P}} + \sum_k a_k g_k \otimes g_k, \mathbf{a} \in \ell^2_c \right\}, \tag{2.37}$$

where $\{g_k\}_k$ is a fixed CONS in $\mathcal{H}(K_{\mathrm{P}})$.

To make finding a sieve in $\mathcal{C}_0$ easier, we make use of the one-to-one correspondence between $\mathcal{C}_0$ and ℓ^2_c. To see this correspondence, suppose that

$$\sum_k a_k g_k \otimes g_k = \sum_k \tilde{a}_k g_k \otimes g_k.$$

Then for each $s \in T$,

$$\sum_k a_k g_k(s) g_k = \sum_k \tilde{a}_k g_k(s) g_k.$$

Applying the linear functional $\langle \cdot, g_m \rangle_{\mathcal{H}(K_{\mathrm{P}})}$ to both sides of this equation, we obtain

$$a_m g_m(s) = \tilde{a}_m g_m(s).$$

Now,we let s vary in T. One more application of the linear functional $\langle \cdot, g_m \rangle_{\mathcal{H}(K_{\mathrm{P}})}$ gives us

$$a_m = \tilde{a}_m.$$

Therefore, the mapping $\mathbf{a} \to K_{\mathrm{Q}} - K_{\mathrm{P}}$ is one-to-one.

Define the sets $\mathcal{S}_d$ by

$$\mathcal{S}_d = \{\mathbf{a} \in \ell_c^2 \colon a_k = 0 \text{ for } k > d\} \quad d \in \mathbb{N}.$$

Then $\{\mathcal{S}_d, d \in \mathbb{N}\}$ is a sieve in ℓ_c^2. We want a sieve in $\mathcal{C}_0$, but we may still consider $\{\mathcal{S}_d, d \in \mathbb{N}\}$ as a sieve in $\mathcal{C}_0$. The sieve estimator for $\mathbf{a}$ in ℓ_c^2 is given by $\hat{\mathbf{a}}_{n,d} = \{\hat{a}_k\}_k$, where

$$\hat{a}_k = \begin{cases} S_k^2 - 1 & \text{if } k \leq d \\ 0 & \text{otherwise} \end{cases}, \tag{2.38}$$

where S_k^2 is given right after (2.33) and the sieve estimator in $\mathcal{C}_0$ is

$$\hat{K}_{\mathrm{Q}} = K_{\mathrm{P}} + \sum_{k=1}^{d} \hat{a}_k g_k \otimes g_k. \tag{2.39}$$

Because of this new sieve estimator, we shall add an assumption to the model:

(A5) The true covariance belongs to $\mathcal{C}_0$.

This means that instead of considering the large $\mathfrak{P}$, we are assuming that the true measure belongs to a subset $\mathfrak{P}_0$, corresponding to $\mathcal{C}_0$. In addition, we may parameterize $\mathcal{C}_0$ by ℓ_c^2, so that the measure that endows $\{Z_t \colon t \in T\}$ with the covariance function

$$K_{\mathrm{P}} + \sum_{k=1}^{\infty} a_k g_k \otimes g_k \tag{2.40}$$

can be denoted by $\mathrm{P}_{\mathbf{a}}$ instead of Q.

For $\mathrm{Q} \in \mathfrak{P}_0$ and each $(s,t) \in T \times T$,the estimator $\hat{K}_{\mathrm{Q}}(s,t)$ has been shown to be

asymptotically unbiased as $d \to \infty$ and weakly and mean-square consistent for $K_{\mathrm{Q}}(s,t)$ provided that $d = O(n)$ [6, Corollary 4.1]. Furthermore, the following theorem and corollary follow from [6, Theorem 5.1, Corollary 5.1]:

Theorem 2.3.1. *Let $d \to \infty$ and $d/n^{\sigma} \to \beta < \infty$ for some $\sigma \in (0,1)$. Then for any $\epsilon > 16\beta/e^2$, we have $\mathrm{P}_{\mathbf{a}}\left(\sum_{k=1}^{d}(\hat{a}_k - a_k)^2 < \epsilon \quad i.o.\right) = 1$.*

Corollary 2.3.1. *If $d \to \infty$ and $d/n^{\sigma} \to 0$ for some $\sigma \in (0,1)$, then $||\hat{\mathbf{a}}_{n,d} - \mathbf{a}||_{\ell^2} \to 0$ a.s. $\mathrm{P}_{\mathbf{a}}$.*

Therefore, we say that $\hat{K}_{\mathrm{P}}$ converges to K_{P} $\mathrm{P}_{\mathbf{a}}$-almost surely in the sense that $\hat{\mathbf{a}}_{n,d}$ converges to $\mathbf{a}$ $\mathrm{P}_{\mathbf{a}}$-almost surely.

Chapter 3

Phenotypic Traits and the Selection Gradient

The evolutionary change of a biological trait from one generation to the next is a topic of interest of evolutionary biologists. When the trait of interest can be described as a vector, the result is given by a well-known equation (3.3), the so-called Breeder's Equation [13]. If the trait is better described by a function, a version, (3.6), of the Breeder's equation holds as well [14].

We assume that evolution occurs in two steps. Evolution first occurs by selection, determined by the fitness (i.e., survivorship). The fitness is determined by the observed trait called the *phenotype* and describes an organism's ability to survive and procreate. Once selection has occurred, inheritance is the second step and is determined by the mating patterns and genetics of the survivors.

For the majority of quantitative genetics, biologists are interested in the genetic makeup, or *genotype*, of an organism. Unfortunately, the genotype of an organism, in general, is not observable, but the phenotype is, by definition. The phenotype of an organism is its physical characteristics, which are what give an individual an advantage/disadvantage when it comes to its ability to survive and reproduce. Therefore, selection is the force that acts upon the

phenotype, rather than on the genotype. Since the genotype represents the genetics and the phenotype represents observable characteristics, we may decompose the phenotype (which is a random variable) into the sum of two uncorrelated random variables (assuming that there is no genotype-environment interaction),

$$z = g + e,$$

where g is called the additive genetic effect and e is called the environmental effect [14]. In a sample, the i^{th} individual has phenotype $z_i = g_i + e_i$ and it is often assumed that organisms are raised in similar but independent environments, so the e_i's are independent and identically distributed.

3.1 Finite Dimensional Traits

The simplest traits are those that can be described by a finite number of measurements. Such traits are called finite-dimensional. For example, "stature" which is given by height and weight is a two-dimensional trait. We may consider a finite-dimensional trait as a random (column) vector $\mathbf{z}$ in $\mathbb{R}^k$ (realizations of $\mathbf{z}$ are the phenotypes). We assume that $\mathbf{z}$ is a Gaussian random variable with mean $\overline{\mathbf{z}} = \mathbb{E}[\mathbf{z}] \in \mathbb{R}^k$ and covariance $\mathbf{P} \in \mathbb{R}^{k\times k}$. Note that this notation for the expected value is usually taken to be the sample mean, but biologists prefer to use it for expectation. Since this dissertation has applications in evolutionary biology, we shall be faithful to this notation. That is, when $\bar{\mathbf{z}}$ appears, it is <u>not</u> a random quantity. This notation only pertains to the trait $\mathbf{z}$; for any other random variables, say $W_1, \ldots, W_n$, $\overline{W} = \frac{1}{n}\sum_{i=1}^n W_i$ denotes the usual sample mean. Additionally, note that the only assumption made on $\bar{\mathbf{z}}$ is that it belongs to $\mathbb{R}^k$.

We assume that $\mathbf{z}_i = \mathbf{g}_i + \mathbf{e}_i \in \mathbb{R}^k$ as in the previous section, where $\mathbf{g}_i$ and $\mathbf{e}_i$ are the additive-genetic trait and the environmental effect of the ith observation, respectively. We also assume that observations are raised in independent, but similar environments, so that

$\{\mathbf{e}_i\}$ is an independent and identically distributed set of random vectors.

Let $\mathbf{G} = \texttt{Cov}\,(\mathbf{g}_i) \in \mathbb{R}^{k\times k}$ and $\mathbf{E} = \texttt{Cov}\,(\mathbf{e}_i) \in \mathbb{R}^{k\times k}$ and set $\mathbf{Z} = [\mathbf{z}_1^T \dots \mathbf{z}_n^T]^T \in \mathbb{R}^{nk}$. Then

$$\begin{aligned}\texttt{Cov}\left(\left[\mathbf{g}_i\right]_{i=1}^{n}\right) &= \left[\texttt{Cov}\,(\mathbf{g}_i, \mathbf{g}_j)\right]_{i,j=1}^{n} \\ &= \left[a_{ij}\mathbf{G}\right]_{i,j=1}^{n} \\ &= \mathbf{A}\otimes\mathbf{G},\end{aligned} \tag{3.1}$$

where $\mathbf{A} = [a_{ij}]_{i,j=1}^{n}$ is the *additive-genetic relationship matrix* that is discussed in the next subsection and $B \otimes C$ denotes the Kronecker product of B and C. Similarly, we may write

$$\texttt{Cov}\left(\left[\mathbf{e}_i\right]_{i=1}^{n}\right) = \mathbf{I}_n \otimes \mathbf{E}, \tag{3.2}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

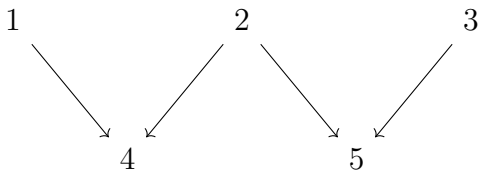
3.1.1 The Relationship Matrix

The additive-genetic relationship matrix $\mathbf{A}$ is a matrix that indicates the different relations in a sample. Standard arguments [13] show that

$$\texttt{Cov}\,(\text{sibling,sibling}) = \texttt{Cov}\,(\text{parent, offspring}) = \frac{1}{2}\mathbf{G},$$

$$\texttt{Cov}\,(\text{half-sibling,half-sibling}) = \frac{1}{4}\mathbf{G}.$$

That is, the covariance between the traits of two siblings is equal to one-half of the genetic covariance. As an example [19, page 757], consider the pedigree



where organisms 1, 2, and 3 are unrelated and 4 and 5 are their offspring. Then the relationship matrix is given by

$$A = \begin{bmatrix} 1 & 0 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 1 & 1/4 \\ 0 & 1/2 & 1/2 & 1/4 & 1 \end{bmatrix}.$$

3.1.2 The Selection Gradient

The probability density function of $\mathbf{z}$ shall be denoted by $p_{\overline{\mathbf{z}}}(\mathbf{z})$ and the expected value with respect to $p_{\overline{z}}(\mathbf{z})$ by $\mathbb{E}_{\overline{\mathbf{z}}}$. The fitness of an individual with phenotype $\mathbf{z}$ is denoted by $W(\mathbf{z})$ or W if the context is clear.

The normal distribution with mean $\overline{\mathbf{z}}$ and covariance matrix $\mathbf{P}$ is assumed to be the *pre-selection* distribution. Additionally, we assume that

$$W > 0 \text{ and}$$

$$\mathtt{Var}_{\overline{\mathbf{z}}}(W) < \infty, \text{ for all } \overline{\mathbf{z}} \in \mathbb{R}^n.$$

These assumptions allow us to define the *post-selection* distribution of $\mathbf{z}$. The post- selection distribution of $\mathbf{z}$ is the distribution with density

$$p^*_{\bar{Z}}(\mathbf{z}) = \frac{W(\mathbf{z}) p_{\bar{Z}}(\mathbf{z})}{\mathbb{E}_{\overline{\mathbf{z}}}[W]}.$$

It should be noted that if $p_{\bar{Z}}(\mathbf{z})$ is normal density function, $p^*_{\bar{Z}}(\mathbf{z})$ may not be a normal density. In addition, it may be the case that W depends on $\overline{\mathbf{z}}$ or other parameters of the pre-selection distribution of $\mathbf{z}$. If this isn't the case, W is said to be *frequency independent.*

Next, we denote other means as follows:

$$\begin{aligned} \overline{\mathbf{z}}^* &= \text{the mean of the trait } \mathbf{z} \text{ after selection, i.e. with respect to } p^*_{\overline{Z}}(\mathbf{z}), \text{ and} \\ \overline{\mathbf{z}}' &= \text{the mean among newborns of the following generation.} \end{aligned}$$

Let $\mathbf{s} = \overline{\mathbf{z}}^* - \overline{\mathbf{z}}$ be the *selection differential*, the within-generation change in the mean phenotype due to selection. The between-generation change in the mean is called the *evolutionary response to selection* and is denoted by

$$\Delta\mathbf{z} = \overline{\mathbf{z}}' - \overline{\mathbf{z}}.$$

If $\mathbf{z}$ follows a $N(\overline{\mathbf{z}}, \mathbf{P})$ distribution, then we have the *Breeder's Equation* [15] :

$$\Delta\mathbf{z} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}, \tag{3.3}$$

where $\mathbf{G}$ is the *additive genetic covariance matrix*. The product $\mathbf{P}^{-1}\mathbf{s}$ has been called [15] the *selection gradient* of the trait at $\overline{\mathbf{z}}$ and is denoted by $\boldsymbol{\beta}$.

Except in artificially managed populations, $\overline{\mathbf{z}}^*$ (and therefore $\mathbf{s}$) is difficult, if not impossible, to estimate. Fortunately, the Robertson-Price Identity [13] states that

$$\mathbf{s} = \mathtt{Cov}\left(w, \mathbf{z}\right), \tag{3.4}$$

where w is the *relative fitness* and is defined by

$$w = \frac{W}{\mathbb{E}_{\overline{\mathbf{z}}}\left[W\right]}.$$

Both W and $\mathbf{z}$ are observable, so $\mathbf{s}$ may be estimated using this identity. Now, from the equation $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathtt{Cov}\left(w, \mathbf{z}\right)$, we see that $\boldsymbol{\beta}$ consists of the partial regression coefficients when regressing w on $\mathbf{z}$, since $\mathbf{P} = \mathtt{Cov}\left(\mathbf{z}\right)$. Furthermore, the ith component of $\boldsymbol{\beta}$ indicates the

force of directional selection acting directly on the ith component of the trait [16], hence the term selection gradient.

3.2 Function Valued Traits

In contrast to the finite-dimensional traits of the previous section, a *function-valued trait* cannot be described by a finite number of components. Such a trait is better represented by a function $Z(t)$ with parameter t. For example, $Z(t)$ could be the size of an organism at age t, or the shape of a wing, where $Z(t)$ is the length of a wing at a polar angle of t. When endowed with a probability distribution, such a trait becomes a stochastic process $\{Z_t \colon t \in T\}$, where T is a set. Similar to the finite-dimensional case, a realization (sample path) of $\{Z_t \colon t \in T\}$ is called the *phenotype* of an organism. The pre-selection mean of $\{Z_t \colon t \in T\}$ will be denoted by $\bar{Z}_t$ and the pre-selection distribution by $\mathrm{P}_{\bar{z}}$. We assume that under $\mathrm{P}_{\bar{z}}$, $\{Z_t \colon t \in T\}$ is a Gaussian process with mean $\bar{z}$ and *phenotypic covariance function* P. It should be noted that the roman P will denote a probability measure, while the italic P will denote a phenotypic covariance function. Both notations are common notations and will be used in this dissertation. As with the finite-dimensional case, we assume that for all $t \in T$, we may write $Z_t = G_t + E_t$, where $\{g_t, t \in T\}$ is the additive-genetic process and $\{e_t, t \in T\}$ is the environmental process. Furthermore, we assume that there is no genotype-environmental interaction. Then $\{g_t, t \in T\}$ and $\{e_t, t \in T\}$ are uncorrelated. It follows that for all $s, t \in T$,

$$P(s,t) = G(s,t) + E(s,t), \tag{3.5}$$

where $G(s,t)$ is the *additive-genetic covariance function* and $E(s,t)$ is the *environmental covariance function*. We also define

$$\begin{aligned}
\overline{z}^*(t) &= \text{ the mean of the trait } z \text{ after selection,} \\
s(t) &= \overline{z}^*(t) - \overline{z}(t) = \text{selection differential,} \\
\bar{z}^\dagger(t) &= \text{ the mean of the trait of newborns in the next generation, and} \\
\Delta\overline{z}(t) &= \bar{z}^\dagger(t) - \overline{z}(t) = \text{ the evolutionary response due to selection.}
\end{aligned}$$

Kirkpatrick and Heckman [14] state that a version of the Breeder's Equation (3.3) holds for function-valued traits. We have

$$\Delta\bar{z} = \mathcal{G}\mathcal{P}^{-1}s, \tag{3.6}$$

where $\mathcal{P}$ and $\mathcal{G}$ are the integral operators with kernels $P(s,t)$ and $G(s,t)$, respectively. Here G is the additive-genetic covariance function given in (3.5). Kirkpatrick and Heckman also suggest that the selection gradient should be formally defined as

$$\beta = \mathcal{P}^{-1}s. \tag{3.7}$$

This equation defines β if s is in the range of $\mathcal{P}$, but not if $s \notin \text{range}(\mathcal{P})$.

The main goal of this section is to determine what sort of mathematical object the selection gradient is and where it resides. A few of the results in this section are due to [12], as noted, while the rest expand upon the results of [12].

3.2.1 Restrictions on T and the Covariance Function

We would like to define an integral operator with kernel P, but to do so, we must first introduce a few conditions. Most of the results in this subsection are due to [9].

We begin with a few assumptions on T and P:

(B1) $(T, \mathcal{T}, \mu)$ is a σ-finite measure space,

(B2) P is a measurable covariance kernel on $T \times T$ that satisfies

$$\int_T P(t,t)d\mu(t) < \infty, \tag{3.8}$$

(B3) in the RKHS $\mathcal{H} = \mathcal{H}(P)$ the only μ-negligible function is the zero function on T.

If the condition (3.8) is satisfied, we say that P has *finite trace*. Note that the integral in (3.8) is nonnegative since $P(t,t)$ is a nonnegative function. The conditions **(B1)-(B3)** are met, for example, when P is continuous, T is a compact interval $[a,b]$ in $\mathbb{R}$, and μ is Lebesgue measure [12]. The following lemma is due to [9, Theorem 26]:

Lemma 3.2.1. *Under the conditions **(B1)-(B3)**, every $f \in \mathcal{H}$ is $\mathcal{T}$-measurable and square integrable on $(T, \mathcal{T}, \mu)$ and the map j that maps $f \in \mathcal{H}$ to its μ-equivalence class $[f] \in L^2(T, \mathcal{T}, \mu)$ is one-to-one and Hilbert-Schmidt.*

In light of this lemma, we may consider $\mathcal{H}$ as a vector subspace of $L^2(T) = L^2(T, \mathcal{T}, \mu)$. Let

$$M = \overline{\text{range}(j)}, \tag{3.9}$$

where $\overline{E}$ is the closure of E in the ambient space (here, the closure is in $L^2(T)$). It follows that M is a Hilbert space with the $L^2(T)$ inner product restricted to M.

Define the integral operator $\mathcal{P}$ by

$$\mathcal{P}f(s) = \int_T P(s,t)f(t)d\mu(t), \tag{3.10}$$

for a square-integrable function f. If $(\cdot,\cdot)_{L^2(T)}$ denotes the inner product in $L^2(T)$, then (3.10) may be written as

$$\mathcal{P}f(s) = (f, P_s)_{L^2(T)}.$$

Furthermore, if j^* is the adjoint of j, then [9, Lemma 29]

$$j^*([f]) = \mathcal{P}f. \tag{3.11}$$

This shows that we may consider $\mathcal{P}$ as a map from $M \subset L^2(T)$ to $\mathcal{H}(P)$, so that range$(\mathcal{P})$ = range(j^*) and also that $\mathcal{P}f \in \mathcal{H}(P)$. An immediate consequence of (3.11) is that

$$\text{range}(\mathcal{P}) \subset \mathcal{H}(P), \tag{3.12}$$

since the adjoint maps M to $\mathcal{H}(P)$. The result (3.12) can be strengthened slightly. Since j is one-to-one (Lemma 3.2.1), it follows that $\overline{\text{range}(\mathcal{P})} = 0^{\perp} = \mathcal{H}(P)$. Hence,

$$\text{range}(\mathcal{P}) \text{ is } \mathcal{H}\text{-dense in } \mathcal{H}. \tag{3.13}$$

The following lemma is essentially due to [9] and connects the inner products of the spaces $\mathcal{H}(P)$ and $L^2(T)$. The proof must be slightly modified to suit our hypotheses.

Lemma 3.2.2. *Let $\mathcal{P}$ be defined as in (3.10). Then for any square-integrable f and any $\eta \in \mathcal{H}(P)$, $\mathcal{P}f \in \mathcal{H}(P)$ and*

$$\langle \mathcal{P}f, \eta \rangle_P = (f, \eta)_{L^2(T)}. \tag{3.14}$$

Proof. It has already been shown that $\mathcal{P}f \in \mathcal{H}(P)$ for any square integrable function f. Let $\eta \in \mathcal{H}(P)$. Then η is square-integrable and also

$$
\begin{aligned}
||\eta||^2_{L(T)} &= \int_T |\eta(t)|^2 d\mu(t) \\
&= \int_T |\left\langle \eta, P_t \right\rangle_P|^2 d\mu(t) \\
&\leq \int_T ||\eta||^2_P ||P_t||^2_P d\mu(t) \\
&= ||\eta||^2_P \int_T \left\langle P_t, P_t \right\rangle_P d\mu(t) \\
&= ||\eta||^2_P \int_T P(t,t) d\mu(t).
\end{aligned}
$$

Therefore, there exists an $N = \left[\int_T P(t,t)d\mu(t)\right]^{1/2} \geq 0$ such that

$$
||\eta||_{L^2(T)} \leq N||\eta||^2_P. \tag{3.15}
$$

It follows from this and the Cauchy-Schwarz inequality that

$$
|(f,\eta)_{L^2(T)}| \leq ||f||_{L^2(T)}||\eta||_{L^2(T)} \leq N||f||_{L^2(T)}||\eta||_P.
$$

Therefore, $(f,\cdot)$ is a bounded linear functional on $\mathcal{H}(P)$. By the Riesz Representation Theorem, there exists a $g \in \mathcal{H}(P)$ such that

$$
(f,\eta)_{L^2(T)} = \left\langle g, \eta \right\rangle_P.
$$

Setting $\eta = P_t$ and using the reproducing property of $\mathcal{H}(P)$, we have

$$
\mathcal{P}f(t) = (f, P_t)_{L^2(T)} = \left\langle g, P_t \right\rangle_P = g(t). \tag{3.16}
$$

Thus $g = \mathcal{P}f$, so (3.14) follows from (3.16). □

From now on, we will write f to denote both a function and its class $[f]$ in $L^2(T)$. It will

be clear from context which is meant. From now on, when we write $\mathcal{P}$, it is meant to be the operator $\mathcal{P} : M \to \mathcal{H}$. Using (3.14), we are able to prove the following lemma:

Lemma 3.2.3. *$\mathcal{P}$ maps M one-to-one into $\mathcal{H}(P)$.*

Proof. Let $f \in M$ be such that $\mathcal{P}f = 0$. By (3.14), it follows that

$$0 = \langle \mathcal{P}f, \eta \rangle_P = (f, \eta)_{L^2(T)},$$

for all $\eta \in \mathcal{H}(P)$ and, since the map j is one-to-one, for all $\eta \in M$. Therefore, $f \in M^\perp$. By assumption f is also an element of M, so that $f = 0$. Then nullspace$(\mathcal{P}) = 0$ and $\mathcal{P}$ is one-to-one. □

Remark 3.2.1. In [12] it is stated that $\mathcal{P}$ as a map from $L^2(T)$ to $\mathcal{H}(P)$ is one-to-one. However, this doesn't hold if $M^\perp$ is nontrivial.

3.2.2 Weak Limits

In order to properly define the selection gradient, we must begin by defining weak topologies and weak limits, see e.g. [23].

Let H be a Hilbert space with inner product $(\cdot, \cdot)$. A sequence $\{f_k\}$ in H is *weakly convergent* if

$$(f_k, \eta) \text{ converges for all } \eta \in H. \tag{3.17}$$

We define the linear functional $\langle f, \cdot \rangle$ by $\langle f, \eta \rangle = \lim_k (f_k, \eta)$ for $\eta \in H$ and call f the *weak limit* of $\{f_k\}$. Note that as of yet, it is unclear whether the weak limits of H are elements of H. We shall let $\bar{H}$ denote the *weak completion of H*, that is H with its weak limits. We show that the weak limits are actually elements of H.

Lemma 3.2.4. *If $\overline{H}$ is the weak closure of the Hilbert space H, then $\overline{H} = H$.*

Remark 3.2.2. In a RKHS, weak convergence implies pointwise convergence. This follows immediately from this lemma and the reproducing property by setting $\eta = P_t$. Furthermore, it is easy to show that the converse holds as well.

Proof of Lemma 3.2.4. Let $h \in \overline{H}$. Then there is a sequence $\{h_n\}$ in H such that the bounded linear functionals $(h_n, \eta)_H$ converge to (h, η) for all $\eta \in H$. In particular, there exists a $y \in \mathbb{R}$ such that $|(h_n, \eta)_H - y|$ converges to 0. On the other hand, by the Banach-Steinhaus Theorem, see e.g. [10, Theorem III.14.6], there is a bounded linear functional L such that

$$|(h_n, \eta)_H - L\eta| \to 0$$

for all $\eta \in \mathcal{H}(P)$. Therefore, we see that $L\eta = y = \langle h, \eta \rangle$ and thus $L = \langle h, \cdot \rangle$ is a bounded linear functional on $\mathcal{H}(P)$ and by the Riesz Representation Theorem, h may be identified with an element $g \in H$. Hence $\overline{H} = H$. □

More generally, let $\mathcal{F}$ be a set of linear functionals on H and consider the weakest topology $\sigma(H, \mathcal{F})$ that makes all of the linear functionals of $\mathcal{F}$ continuous. A sequence $\{f_k\}$ in H is convergent in this new topology if

$$(f_k, h) \text{ converges for every } h \in \mathcal{F}.$$

Similar to the weak completion, this defines a linear functional $(f, \cdot) = \lim_k (f_k, \cdot)$ and we call f the $\sigma(H, \mathcal{F})$-limit of $\{f_k\}$. It may be possible for the $\sigma(H, \mathcal{F})$-limit of a sequence to not exist in H. This is because the linear functionals $(f, \cdot)$ are continuous in $\sigma(H, \mathcal{F})$, but not necessarily in the natural topology of H. Thus, f may not be able to be identified with an element of H. A specific example of such a limit is given below in Remark 3.2.3. We will combine the set of all $\sigma(H, \mathcal{F})$-limits and H to form the completion of H in the $\sigma(H, \mathcal{F})$-topology.

For example, let H be $L^2(T)$ and suppose that restrictions **(B1)-(B3)** of Subsection

3.2.1 hold. Consider the set of linear functionals $\mathcal{F}$ on $L^2(T)$ defined by

$$\mathcal{F} = \{(\cdot, \eta)_{L^2(t)} : \eta \in \mathcal{H}, \}$$

and let $\mathcal{L}$ be the $\sigma(L^2(T), \mathcal{F})$-completion of $L^2(T)$. If F is a functional in $\mathcal{F}$, then $F(\cdot) = (\cdot, \eta)_{L^2(T)}$ for $\eta \in \mathcal{H}$. F is extended to $\mathcal{L}$ by setting $(f, \eta)_{L^2(T)} = \lim_k (f_k, \eta)_{L^2(T)}$ with $f_k \to f$ in $\mathcal{L}$. This is no longer an inner product unless f is in $L^2(T)$. Let M be as in (3.9). The completion of M in the $\sigma(M, \mathcal{F})$-topology will be denoted by $\mathcal{M}$. It follows that $\mathcal{M} \subset \mathcal{L}$. We now extend $\mathcal{P}$ to an operator on $\mathcal{M}$.

Theorem 3.2.1. *The integral operator $\mathcal{P}$ given by (3.10) has a unique extension to a linear operator $\bar{\mathcal{P}}$, mapping $\mathcal{M}$ bijectively on $\mathcal{H}(P)$. Moreover, for all $f \in \mathcal{M}$ and $\eta \in \mathcal{H}(P)$, we have*

$$\left\langle \bar{\mathcal{P}} f, \eta \right\rangle_P = (f, \eta)_{L^2(T)}. \tag{3.18}$$

Remark 3.2.3. The right hand side of (3.18) is the value of the functional $(f, \cdot)_{L^2(T)}$ at η. We also note that (3.18) extends the validity of equation (3.14) and that this theorem extends Lemma 3.2.2. Moreover, this shows that linear functionals contained in $\mathcal{M}$ are continuous on $\mathcal{H}(P)$.

This theorem is similar to [12, Proposition 4.2]. In the proposition, it is claimed that $\bar{\mathcal{P}}$ maps $\mathcal{L}$ one-to-one into $\overline{\mathcal{H}(P)}$, but if $\mathcal{P} : L^2(T) \to \mathcal{H}(P)$ is not one-to-one (Remark 3.2.1), $\bar{\mathcal{P}} : \mathcal{L} \to \overline{\mathcal{H}(P)}$ fails to be one-to-one.

Using this theorem, we can show that it is possible for $\mathcal{M}$ to contain generalized functions; in particular, the δ functions. Here, the phrase δ *function at* t means the point evaluation functional at t, i.e. δ_t satisfies

$$(\delta_t, \eta)_{L^2(T)} = \eta(t)$$

for all $t \in T$ and $\eta \in \mathcal{H}(P)$. Since $P_t \in \mathcal{H}(P)$, there is a unique $f_t \in \mathcal{M}$ such that $\bar{\mathcal{P}} f_t = P_t$.

For all $\eta \in \mathcal{H}(P)$, we have

$$\left\langle \bar{\mathcal{P}} f_t, \eta \right\rangle_P = \langle P_t, \eta \rangle_P = \eta(t).$$

On the other hand, by (3.18),

$$\left\langle \bar{\mathcal{P}} f_t, \eta \right\rangle_P = (f_t, \eta)_{L^2(T)}.$$

Therefore, $(f_t, \eta)_{L^2(T)} = \eta(t)$, which shows that f_t is a δ function. Since these functionals are not elements of $L^2(T)$, they are not elements of M.

Proof of Theorem 3.2.1. We begin by extending $\mathcal{P}$ to an operator $\bar{\mathcal{P}}$ from $\mathcal{M}$ onto $\overline{\mathcal{H}(P)}$ by letting $f \in \mathcal{M}$ and $\{f_n\}_n$ be such that $f_n \to f$ in $\mathcal{M}$ with $f_n \in M$ for all n. If $r_n = \mathcal{P} f_n$, then by (3.14), for all $\eta \in \mathcal{H}(P)$

$$\langle r_n, \eta \rangle_P = \langle \mathcal{P} f_n, \eta \rangle_P = (f_n, \eta)_{L^2(T)} \to (f, \eta)_{L^2(T)}.$$

Therefore, $\{r_n\}$ is weakly convergent in $\mathcal{H}(P)$. Define $\bar{\mathcal{P}} f$ to be the weak limit of $\{r_n\}$ in $\overline{\mathcal{H}(P)}$. Clearly, we have

$$\left\langle \bar{\mathcal{P}} f, \eta \right\rangle_P = (f, \eta)_{L^2(T)}.$$

Since $\bar{\mathcal{P}} f$ is an element of $\overline{\mathcal{H}(P)}$ it follows that

$$\mathcal{H}(P) \subseteq \text{range}(\bar{\mathcal{P}}) \subseteq \overline{\mathcal{H}(P)}. \tag{3.19}$$

By Lemma 3.2.4, we have $\overline{\mathcal{H}(P)} = \mathcal{H}(P)$ and hence $\mathcal{H}(P) = \text{range}(\bar{\mathcal{P}})$, so that $\bar{\mathcal{P}}$ maps $\mathcal{M}$ onto $\mathcal{H}(P)$. Furthermore, since weak limits are unique [11, Theorem 5.8], and $\mathcal{P}$ maps M one-to-one into $\mathcal{H}(P)$ (Lemma 3.2.3), it follows that $\bar{\mathcal{P}}$ maps $\mathcal{M}$ onto $\mathcal{H}(P)$ in a one-to-one fashion. Thus, $\bar{\mathcal{P}} : \mathcal{M} \to \mathcal{H}(P)$ is a bijection. □

This theorem states that if $\{g_k\}$ is a basis, that is, a spanning and linearly independent

set in $\mathcal{H}(P)$, we have a basis $\{\gamma_k\}$ in $\mathcal{M}$ where γ_k is such that

$$\bar{\mathcal{P}}\gamma_k = g_k \text{ for all } k. \tag{3.20}$$

This may be rephrased as follows:

Let γ_k be such that $\bar{\mathcal{P}}\gamma_k = g_k$. For any $\eta \in \mathcal{H}$, we may write

$$\eta = \sum_{k=1}^{\infty} \theta_k g_k.$$

Since $\bar{\mathcal{P}}$ is bijective, it follows that

$$\eta = \bar{\mathcal{P}}\left(\sum_{k=1}^{\infty} \theta_k \gamma_k\right),$$

and hence for any $\eta \in \mathcal{H}$ there exists a unique $f = \sum_{k=1}^{\infty} \theta_k \gamma_k \in \mathcal{M}$ such that $\bar{\mathcal{P}} f = \eta$.

3.2.3 Biological Assumptions

Analogous to the finite-dimensional fitness, the fitness of $\{Z_t \colon t \in T\}$, W, is a positive function of the trait. Also, W may depend on $\bar{z}$ or other parameters of $\mathrm{P}_{\bar{z}}$. When it does not, it is said to be *frequency independent*. We assume that W is frequency independent and that

$$\mathtt{Var}_{\mathrm{P}_{\bar{z}}}(W) < \infty \text{ for all } \bar{z} \in \mathcal{H}(P).$$

Define the post-selection distribution $\mathrm{P}^*_{\bar{z}}$ of $\{Z_t \colon t \in T\}$ by

$$d\mathrm{P}^*_{\bar{z}} = \frac{W}{\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]} d\mathrm{P}_{\bar{z}}.$$

Since W is positive, $\mathrm{P}^*_{\bar{z}}$ is a probability measure and is also absolutely continuous with respect to $\mathrm{P}_{\bar{z}}$. Again, this post-selection distribution may not be a Gaussian probability measure

even if $\mathrm{P}_{\bar{z}}$ is. As in the finite-dimensional case, we define

$$\begin{aligned}
\bar{z}^* &= \text{the post-selection mean of } \{Z_t \colon t \in T\},\\
s = \bar{z}^* - \bar{z} &= \text{the selection differential (the within-generation change in mean)},\\
\bar{z}^\dagger &= \text{the mean of } \{Z_t \colon t \in T\} \text{ among newborns in the following generation, and}\\
\Delta\bar{z} = \bar{z}^\dagger - \bar{z} &= \text{the evolutionary change in the mean}\\
&\qquad \text{(the between-generation change in mean)}.
\end{aligned}$$

The last quantity $\Delta\bar{z}$ has also been called the *evolutionary response to selection*. We shall make the assumption that $\{Z_t \colon t \in T\}$ is a Gaussian process with *phenotypic covariance* $P(s,t)$ and mean $\bar{z} \in \mathcal{H}(P)$, where P and T satisfy **(B1)-(B3)** in Subsection (3.2.1). Further, we assume that W is a positive random variable $\mathrm{P}_{\bar{z}}$-almost surely, $W \in L^2(\Omega, \mathcal{A}, \mathrm{P}_{\bar{z}})$ for each $\bar{z} \in \mathcal{H}(P)$ and $\mathcal{A} = \sigma\left(\{Z_t \colon t \in T\}\right)$.

Once again, save in artificially managed populations, s may be difficult to estimate as it is unknown if selection has occurred. Fortunately, there is a function-valued version of the Robertson-Price Identity given in (3.4), see e.g. [13]:

$$s(t) = \mathtt{Cov}_{\mathrm{P}_{\bar{z}}}\left(Z_t, w\right), \tag{3.21}$$

where $w = W/\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]$ is the *relative fitness function*. In order to use this equation, we must know in what space s resides. The following proposition [12, Proposition 4.1] tells us where:

Proposition 3.2.1. *Let $\bar{z} \in \mathcal{H}(P)$. Then $\bar{z}^*, s \in \mathcal{H}(P)$.*

This proposition and (3.13) show that s is "almost" in range($\mathcal{P}$), as it is in the closure of range($\mathcal{P}$).

3.2.4 The Selection Gradient

We are now able to define the selection gradient, β, if $s \notin$ range($\mathcal{P}$).

Corollary 3.2.1. *For all $s \in \mathcal{H}(P)$, there is a unique $\beta \in \mathcal{M}$ such that*

$$\bar{\mathcal{P}}\beta = s, \tag{3.22}$$

and for all $\eta \in \mathcal{H}(P)$, we have

$$\langle s, \eta \rangle_P = (\beta, \eta)_{L^2(T)}. \tag{3.23}$$

If $s \in$ range$(\mathcal{P})$, then $\beta \in L^2(T)$ (in particular, $\beta \in M$).

Proof. Since $s \in \mathcal{H}(P)$, (3.22) and (3.23) follow from Theorem 3.2.1. The final statement follows from Lemma 3.2.3. □

It is important to note that given two covariance kernels, P, P_0, corresponding to the equivalent measures, P, P_0, the reproducing kernel Hilbert spaces $\mathcal{H}(P), \mathcal{H}(P_0)$ consist of exactly the same functions, as stated in Theorem 2.2.2. Since $s \in \mathcal{H}(P)$, we then have $s \in \mathcal{H}(P_0)$, so that s has an expansion in $\mathcal{H}(P_0)$. By Assumption **(A5)** on page 23, we also have that $P = P_0 + \sum_k a_k g_k \otimes g_k$ for some $\{a_k\} \in \ell_c^2$ and a fixed ONS $\{g_k\}$ in $\mathcal{H}(P_0)$. Let $\bar{\mathcal{P}}_0$ be the extension of the integral operator $\mathcal{P}_0$ with kernel P_0. Theorem 3.2.1 shows that the extended operator $\bar{\mathcal{P}}_0$ maps $\mathcal{M}$ bijectively to $\mathcal{H}(P_0)$.

Proposition 3.2.2. *Let P_0 and P be equivalent measures corresponding to the covariance kernels P_0 and P, respectively, with corresponding operators $\mathcal{P}_0$ and $\mathcal{P}$. Let g_k be an orthonormal sequence in $\mathcal{H}(P_0)$ and let γ_k be such that $\bar{\mathcal{P}}_0 \gamma_k = g_k$ for all k. Then*

$$(\gamma_j, g_k)_{L^2(T)} = \delta_{jk}. \tag{3.24}$$

Furthermore, let $s = \sum_k c_k g_k$. Then

$$\bar{\mathcal{P}}\beta = s,$$

if and only if

$$(a_k + 1)b_k = c_k,$$

where $\beta = \sum_k b_k \gamma_k$.

Proof. The first result is obtained from the equality

$$(\gamma_j, g_k)_{L^2(T)} = \left\langle \bar{\mathcal{P}}_0 \gamma_j, g_k \right\rangle_{P_0} = \langle g_j, g_k \rangle_{P_0} = \delta_{jk}.$$

Next, note that if $P = P_0 + \sum_k a_k g_k \otimes g_k$, we have

$$\mathcal{P} f = \mathcal{P}_0 f + \sum_k a_k (f, g_k)_{L^2(T)} g_k,$$

for $f \in M$. Let $f_n \to f$ in $\mathcal{M}$, then for each $t \in T$,

$$\mathcal{P} f_n(t) = (f_n, P_t)_{L^2(T)} = \langle \mathcal{P} f_n, P_t \rangle_P \to \left\langle \bar{\mathcal{P}} f, P_t \right\rangle_P = \bar{\mathcal{P}} f(t). \tag{3.25}$$

On the other hand, we have

$$\begin{aligned} \mathcal{P}_0 f_n(t) + \sum_k a_k (f_n, g_k)_{L^2(T)} g_k &= (f_n, P_{0t})_{L^2(T)} + \sum_k a_k (f_n, g_k)_{L^2(T)} g_k(t) \\ &= \langle \mathcal{P}_0 f_n, P_{0t} \rangle_{P_0} + \sum_k a_k (f_n, g_k)_{L^2(T)} g_k(t). \end{aligned} \tag{3.26}$$

Similar to (3.25), the first term in (3.26) converges to $\bar{\mathcal{P}}_0 f(t)$. Since $f_n \to f$ in $\mathcal{M}$, it follows that the second term in (3.26) converges to

$$\sum_k a_k (f, g_k)_{L^2(T)} g_k(t).$$

It follows that for each $t \in T$,

$$\bar{\mathcal{P}} f(t) = \bar{\mathcal{P}}_0 f(t) + \sum_k a_k (f, g_k)_{L^2(T)} g_k(t).$$

Therefore, we may write

$$\bar{\mathcal{P}} f = \bar{\mathcal{P}}_0 f + \sum_k a_k (f, g_k)_{L^2(T)} g_k, \tag{3.27}$$

for all $f \in \mathcal{M}$. Thus,

$$\begin{aligned}
\bar{\mathcal{P}}\beta &= \bar{\mathcal{P}}_0 \beta + \sum_k a_k (\beta, g_k)_{L^2(T)} g_k \\
&= \bar{\mathcal{P}}_0 \left(\sum_k b_k \gamma_k \right) + \sum_k a_k \left(\sum_k b_k \gamma_k, g_k \right)_{L^2(T)} g_k \\
&= \sum_k b_k g_k + \sum_k a_k b_k g_k \\
&= \sum_k (1 + a_k) b_k g_k \\
&= \sum_k c_k g_k.
\end{aligned}$$

Hence, $\bar{\mathcal{P}}\beta = s$ if and only if $(1 + a_k)b_k = c_k$. □

Setting $b_k = c_k/(1 + a_k)$, it is easy to see that $\{b_k\}$ is square-summable since $\{a_k + 1\}$ is bounded away from zero. By this proposition and the discussion following 3.2.1, we may write

$$\beta = \sum_{k=1}^{\infty} b_k \gamma_k,$$

where $b_k = c_k/(1 + a_k)$ for all k.

Chapter 4

Estimation of β from an Independent Sample

In this chapter, we deal with estimating the selection gradient given an independent sample. Let $(\Omega, \mathcal{A})$ be a measurable space and let P and P_0 denote the probability measures that endow a process $\{Z_t \colon t \in T\}$ on $(\Omega, \mathcal{A})$ with a Gaussian distribution with mean zero and covariances P and P_0, respectively. Further, we shall assume that $\mathcal{H}(P_0)$ is a separable Hilbert space. For example, $\mathcal{H}(P_0)$ is separable when P_0 is bounded and $T \subset \mathbb{R}$ is separable [8, page 35].

Recall from (3.22) that the selection gradient β is the solution to the equation

$$\bar{\mathcal{P}}\beta = s,$$

where $\bar{\mathcal{P}}$ is an extension of the integral operator with kernel P and therefore $\bar{\mathcal{P}}$ depends on P.

Let $\{g_k\}_k$ be a CONS of $\mathcal{H}(P_0)$, Λ_{P_0} be the Loève map between H_{P_0} and $\mathcal{H}(\mathrm{P}_0)$ and $\{U_k\}_k$ be such that $U_k = \Lambda_{\mathrm{P}_0}^{-1} g_k$. Recall that under the Loève map, $\Lambda_{\mathrm{P}_0}(Z_t) = P_{0t}$, where Z_t

is the observed trait. Since $P_0 = \sum_k g_k \otimes g_k$, it follows that we may write

$$Z(t) = \sum_k U_k g_k(t) \text{ in } \mathcal{H}(P_0).$$

By the Robertson-Price identity (3.21), we see that

$$s(t) = \mathtt{Cov}_{\mathrm{P}}\left(Z(t), w\right) = \sum_k \mathtt{Cov}_{\mathrm{P}}\left(U_k, w\right) g_k(t),$$

where $w = W/\mathbb{E}_{\mathrm{P}}\left[W\right]$ is the relative fitness function. Therefore,

$$c_k = \mathtt{Cov}_{\mathrm{P}}\left(U_k, w\right), \tag{4.1}$$

and given a sample of phenotypes $Z_1(t), \ldots, Z_n(t)$, Proposition 3.2.2 implies that we may estimate b_k by

$$\hat{b}_{kn} = \frac{\hat{c}_{kn}}{1 + \hat{a}_{kn}}, \tag{4.2}$$

where $\hat{c}_{kn}$ and $\hat{a}_{kn}$ are estimators of c_k and a_k, respectively. Furthermore, we estimate β by

$$\hat{\beta}_n = \sum_{k=1}^{\infty} \hat{b}_{kn} \gamma_k,$$

where γ_k is such that $\bar{\mathcal{P}}_0 \gamma_k = g_k$ for all k.

Therefore, to estimate the selection gradient, we need only estimate the sequences $\{c_k\}_k$ and $\{a_k\}_k$. If the true probability measure P endows $\{Z_t \colon t \in T\}$ with mean zero and covariance P, then we are able to use the estimators $\hat{a}_{kn}$ given by (2.38) and the natural estimator $\hat{c}_{kn}$ of c_k given by

$$\hat{c}_{kn} = \widehat{\mathtt{Cov}}_{\mathrm{P}}(w, U_k) = \frac{1}{n} \sum_{j=1}^{n} (\hat{w}_j - \bar{w})(U_{kj} - \bar{U}_k), \tag{4.3}$$

where

$$\hat{w_j} = \frac{W_j}{\frac{1}{n}\sum_{k=1}^{n} W_k},$$

and $\bar{w}$ is the sample average of $\hat{w_j}$. Note that we may write

$$\hat{c}_{kn} = \frac{1}{n}\sum_{j=1}^{n} \hat{w_j}(U_{kj} - \bar{U}_k).$$

On the other hand, if the true distribution of $\{Z_t \colon t \in T\}$ does not have mean zero, the estimator $\hat{c}_{kn}$ is still an appropriate estimator of c_k, but the estimator $\hat{a}_{kn}$ is no longer appropriate. Thus, we must estimate the sequence $\{a_k\}_k$ if the mean $\bar{z}$ is nonzero.

4.1 Joint Estimation of the Mean and Covariance Functions

Let $\{Z_t \colon t \in T\}$ be a function-valued trait on $(\Omega, \mathcal{A})$ with mean $\bar{z}$ and covariance function P. Recall from (2.29) that $\ell_c^2(B) = \{\{a_k\} \colon \inf_k a_k > -1, k \in B\}$ and if $B = \mathbb{N}$ it is written ℓ_c^2. In most applications, we cannot assume that $\bar{z} = 0$. Thus, we must estimate both $\bar{z}$ and P simultaneously. We shall use the results of Theorems 2.2.1 and 2.2.2 to do so. These theorems lead us to initially consider the largest collection $\mathfrak{P}'$ of probability measures on $(\Omega, \mathcal{A})$ such that:

(C1) the process is Gaussian under every $\mathrm{P} \in \mathfrak{P}'$,

(C2) the measures in $\mathfrak{P}'$ are mutually absolutely continuous (equivalent), and

(C3) the mean function $\bar{z}$ belongs to $\mathcal{H}(P)$ for every $P \in \mathfrak{P}'$.

These assumptions imply that $\mathfrak{P}'$ is composed of the family of measures $\mathfrak{P}$ (Section 2.3) and the family of measures with same covariance functions of $\mathfrak{P}$, but with mean $\bar{z}$. Furthermore, if P_0 is the covariance of the process corresponding to the measure P_0, the covariance function

P under $\mathrm{P} \in \mathfrak{P}'$ may be written in the form

$$P = P_0 + \sum_k a_k g_k \otimes g_k, \tag{4.4}$$

where $\{g_k\}$ is a CONS of $\mathcal{H}(P_0)$ and $\{a_k\} \in \ell^2_c$.

Consider the following measures belonging to $\mathfrak{P}'$ such that:

- $\mathrm{P}_{\bar{z}}$ endows $\{Z_t \colon t \in T\}$ with mean $\bar{z}$ and covariance P,
- P endows $\{Z_t \colon t \in T\}$ with mean 0 and covariance P, and
- P_0 endows $\{Z_t \colon t \in T\}$ with mean 0 and covariance P_0.

If $\{g_k\}_k$ is an CONS of $\mathcal{H}(P_0)$ and $\Lambda_{\mathrm{P}_0} : H_{\mathrm{P}_0} \to \mathcal{H}(P_0)$ is the Loève map, and if $U_k = \Lambda^{-1}_{\mathrm{P}_0} g_k$, then $\{U_k\}_k$ is a CONS of H_{P_0}. Let H_{P} be the Gaussian space generated by $\{Z_t\}$ under P.

We need to calculate

$$\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}_0}.$$

We use the following theorem to calculate this likelihood:

Theorem 4.1.1. *Let $\mathrm{P}_{\bar{z}}, \mathrm{P}_0 \in \mathfrak{P}'$ be such that $\mathrm{P}_{\bar{z}}$ endows the process $\{Z_t \colon t \in T\}$ with mean $\bar{z}$ and covariance P and P_0 endows $\{Z_t \colon t \in T\}$ mean zero and covariance P_0. Then*

$$\frac{dP_{\bar{z}}}{d\mathrm{P}_0} = \exp\left\{\frac{1}{2}\sum_k \left(2\theta_k U_k - \theta_k^2 \frac{1}{(1-\lambda_k)} + \lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\},$$

where $\{\theta_k\} \in \ell^2$, $\{-\lambda_k\} \in \ell^2_c$ and $\{U_k\}$ is orthonormal in $\mathcal{H}(P_0)$.

Proof. To obtain the result, we use the formula

$$\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}_0} = \frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}} \frac{d\mathrm{P}}{d\mathrm{P}_0}.$$

From (2.22) and the definition of $\mathfrak{P}'$ we have

$$P = P_0 + \sum_k a_k g_k \otimes g_k,$$

where $\{g_k\}$ is an ONS in $\mathcal{H}(P_0)$ and $\{a_k\} \in \ell^2_c$. Furthermore, there exists a sequence $\boldsymbol{\lambda} = \{\lambda_k\}_k$ with $-\boldsymbol{\lambda} \in \ell^2_c$ such that

$$\frac{d\mathrm{P}}{d\mathrm{P}_0} = \exp\left\{\frac{1}{2}\sum_k \lambda_k U_k^2 + \ln(1-\lambda_k)\right\}$$

and $(1+a_k)(1-\lambda_k) = 1$ for all k, where $\Lambda_{\mathrm{P}_0} U_k = g_k$ for all k, as in (2.26).

By a separate application of the Gaussian Dichotomy Theorem (Theorem 2.2.1), we have

$$\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}} = \exp\left\{Y - \frac{1}{2}||Y||^2_P\right\}, \tag{4.5}$$

where $Y \in H_{\mathrm{P}}$ is unique, $||\cdot||_{\mathrm{P}}$ is the norm in H_{P} and that $\bar{z} = \Lambda_{\mathrm{P}}(Y)$.

By Theorem 2.2.2, H_{P} and H_{P_0} are equal as sets and Y as given in (4.5) has the expansion $Y = \sum_k \theta_k U_k$ in H_{P_0}. It follows that

$$\begin{aligned}\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}} &= \exp\left\{Y - \frac{1}{2}||Y||^2_P\right\} \\ &= \exp\left\{\sum_k \theta_k U_k - \frac{1}{2}(Y,Y)_P\right\}.\end{aligned}$$

By the same theorem the inner products of H_{P} and H_{P_0} are related by

$$(Y_1, Y_2)_P = (Y_1, Y_2)_{P_0} + (Y_1 \odot Y_2, U)_{2^\odot},$$

where $U = \frac{1}{2}\sum a_k U_k^{2^\odot} \in H^{2^\odot}_{\mathrm{P}_0}$. To compute the inner product on $H^{2^\odot}_{\mathrm{P}_0}$ we make use of (2.14),

which states that the inner product on $H^{2\odot}$ is given by

$$(f_1 \odot f_2, f_1' \odot f_2') = (f_1, f_1')_H (f_2, f_2')_H + (f_1, f_2')_H (f_2, f_1')_H .$$

Now,

$$\begin{aligned}
(Y,Y)_P &= \sum_{i,j} \theta_i \theta_j (U_i, U_j)_P \\
&= \sum_{i,j} \theta_i \theta_j \left[(U_i, U_j)_{P_0} + (U_i \odot U_j, U) \right] \\
&= \sum_i \theta_i^2 + \sum_{i,j} \theta_i \theta_j \left(U_i \odot U_j, U \right) \\
&= \sum_i \theta_i^2 + \sum_{i,j} \theta_i \theta_j \sum_k \frac{1}{2} a_k \left(U_i \odot U_j, U_k^{2\odot} \right) \\
&= \sum_i \theta_i^2 + \sum_{i,j} \theta_i \theta_j \sum_k \frac{1}{2} a_k [(U_i, U_k)_{P_0} (U_j, U_k)_{P_0} + (U_i, U_k)_{P_0} (U_j, U_k)_{P_0}] \\
&= \sum_i \theta_i^2 + \sum_{i,j} \theta_i \theta_j \sum_k a_k (U_i, U_k)_{P_0} (U_j, U_k)_{P_0} \\
&= \sum_i \left(\theta_i^2 + \theta_i^2 a_i \right) \\
&= \sum_i \theta_i^2 (1 + a_i).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}} &= \exp \left\{ \sum_k \left(\theta_k U_k - \frac{1}{2} \theta_k^2 (1 + a_k) \right) \right\} \\
&= \exp \left\{ \sum_k \left(\theta_k U_k - \frac{1}{2} \theta_k^2 \frac{1}{(1 - \lambda_k)} \right) \right\},
\end{aligned}$$

and hence

$$\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}_0} = \exp\left\{\frac{1}{2}\sum_k \left(2\theta_k U_k - \theta_k^2 \frac{1}{(1-\lambda_k)} + \lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\}. \tag{4.6}$$

□

To develop estimators of $\bar{z}$ and P that have nice properties, we first change the parameterization of the likelihood (4.6). To do so, we wish to determine the expansion of $\bar{z}$ in $\mathcal{H}(P_0)$, but from (2.8), we know that $\bar{z} = \Lambda_P(Y)$, where $Y = \sum_k \theta_k U_k$ in H_{P_0}. Then $\bar{z} = \sum_k \theta_k \Lambda_P(U_k)$. Now, we can write $\Lambda_P(U_k) = (U_k, Z.)_P$. By (2.14),

$$\begin{aligned}
(U_k, Z.)_P &= (U_k, Z.)_{P_0} + (U_k \odot Z., U)_{2\odot} \\
&= g_k + \frac{1}{2}\sum_j a_j[(U_k, U_j)_{P_0}(U_k, Z.)_{P_0} + (U_k, U_j)_{P_0}(U_k, Z.)_{P_0}] \\
&= g_k + \sum_j a_j (U_k, U_j)_{P_0}(U_k, Z.)_{P_0} \\
&= g_k + \sum_j a_j \delta_{jk} \Lambda_{\mathrm{P}_0}(U_k) \\
&= g_k + a_k g_k \\
&= (1 + a_k) g_k
\end{aligned}$$

Therefore, it follows that in $\mathcal{H}(P_0)$ we can write

$$\bar{z} = \sum_k \theta_k (1 + a_k) g_k.$$

We can then set

$$\mu_k = \theta_k (1 + a_k), \tag{4.7}$$

so that $\bar{z} = \sum_k \mu_k g_k$. Since $\bar{z} \in \mathcal{H}(P_0)$, it follows that $\boldsymbol{\mu} = \{\mu_k\}_k \in \ell^2$. In light of this, the

parameter space for $\mathfrak{P}'$ is given by

$$\mathcal{C}' = \Big\{(\bar{z}, P)\colon P = P_0 + \sum_k a_k g_k \otimes g_k, \mathbf{a} \in \ell^2_c, \bar{z} = \sum_k \mu_k g_k, \\ \boldsymbol{\mu} \in \ell^2, \{g_k\} \text{ countable and orthonormal in } \mathcal{H}(P_0).\Big\} \tag{4.8}$$

To determine the distribution of U_k for each k under $\mathrm{P}_{\bar{z}}$, we need to use the following:

Proposition 4.1.1 ([5, Corollary 2.2])**.** *For any $X, Y \in H_{\mathrm{P}}$,*

$$\mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(X, Y) = \mathtt{Cov}_{\mathrm{P}}(X, Y).$$

Lemma 4.1.1. *Under $\mathrm{P}_{\bar{z}}$, $\{U_k\}$ is an independent set of Gaussian random variables and for each k, $U_k \sim N(\mu_k, 1 + a_k)$*

Proof. Since the measures P and $\mathrm{P}_{\bar{z}}$ are equivalent,

$$\mathbb{E}_{\mathrm{P}_{\bar{z}}}[U_k] = (U_k, Y)_{\mathrm{P}}, \tag{4.9}$$

for a unique $Y = \sum_j \theta_j U_j$ in H_{P} given by (4.5). Further, by (2.34), $\{U_k\}$ is a sequence of independent normal random variables such that for each k, $U_k \sim N(0, 1 + a_k)$. Therefore, $(U_k, U_j)_{\mathrm{P}} = (1 + a_k)\delta_{jk}$ where δ_{jk} is the Kronecker δ. The expectation (4.9) then becomes

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\bar{z}}}[U_k] &= (U_k, Y)_{\mathrm{P}} \\
&= \sum_j \theta_j (U_k, U_j)_{\mathrm{P}} \\
&= \sum_j \theta_j (1 + a_k)\delta_{jk} \\
&= \theta_k (1 + a_k) \\
&= \mu_k.
\end{aligned}$$

To show that $\mathtt{Var}_{\mathrm{P}_{\bar{z}}}(U_k) = 1 + a_k$, use (2.34) and Proposition 4.1.1 to obtain

$$\mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(U_k, U_j) = \mathtt{Cov}_{\mathrm{P}}(U_k, U_j) = (1 + a_k)\delta_{jk}. \tag{4.10}$$

This proves the lemma. □

Continuing, we may rewrite

$$\begin{aligned}
\frac{d\mathrm{P}_{\bar{z}}}{d\mathrm{P}_0} &= \exp\left\{\frac{1}{2}\sum_k \left(2\theta_k U_k - \theta_k^2 \frac{1}{(1-\lambda_k)} + \lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\} \\
&= \exp\left\{\frac{1}{2}\sum_k \left(2\frac{\mu_k}{1+a_k} U_k - \left(\frac{\mu_k}{1+a_k}\right)^2 \frac{1}{(1-\lambda_k)} + \lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\} \\
&= \exp\left\{\frac{1}{2}\sum_k \left(2\mu_k(1-\lambda_k)U_k - \mu_k^2(1-\lambda_k) + \lambda_k U_k^2 + \ln(1-\lambda_k)\right)\right\}.
\end{aligned}$$

From this, we obtain the likelihood on $(\Omega^n, \mathcal{A}^{n\otimes})$,

$$L = \frac{d\mathrm{P}_{\bar{z}}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}} = \exp\left\{\frac{n}{2}\sum_k \left(2(1-\lambda_k)\mu_k \bar{U}_k - \mu_k^2(1-\lambda_k) + \lambda_k S_k^2 + \ln(1-\lambda_k)\right)\right\} \tag{4.11}$$

where $S_k^2 = \sum_{i=1}^n U_{ki}^2$. Note that L is the product of terms of the form,

$$\exp\left\{\frac{n}{2}\left(2(1-\lambda_k)\mu_k \bar{U}_k - \mu_k^2(1-\lambda_k) + \lambda_k S_k^2 + \ln(1-\lambda_k)\right)\right\}. \tag{4.12}$$

Thus, to maximize L it suffices to maximize (4.12) for all k. Let $\mathfrak{P}'^{(n)} = \{\mathrm{Q}^{n\otimes} \colon \mathrm{Q} \in \mathfrak{P}'\}$.

Theorem 4.1.2. *Let U_{ki} $i = 1, \ldots, n$, $k \in B$, be fixed. If B is finite, the likelihood* (4.11) *is maximized over $\ell^2(B) \times \ell_c^2(B)$ almost surely $(\mathfrak{P}'^{(n)})$ at*

$$\hat{\mu}_{kn} = \begin{cases} \bar{U}_k & \text{if } k \in B, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{\lambda}_{kn} = \begin{cases} 1 - \left[\frac{1}{n}\sum_{i=1}^{n}(U_{mi} - \bar{U}_m)^2\right]^{-1} & \text{if } k \in B, \\ 0 & \text{otherwise,} \end{cases}$$

with corresponding estimator

$$\hat{a}_{kn} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(U_{mi} - \bar{U}_m)^2 - 1 & \text{if } k \in B, \\ 0 & \text{otherwise} \end{cases}$$

If B is infinite, (4.11) *is unbounded almost surely $\mathfrak{P}'^{(n)}$.*

Proof. First, suppose that B is infinite. By construction, $\frac{d\mathrm{P}_{\bar{z}}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}} = \frac{d\mathrm{P}_{\bar{z}}^{n\otimes}}{d\mathrm{P}^{n\otimes}}\frac{d\mathrm{P}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}}$. By [6, Theorem 3.1], $\frac{d\mathrm{P}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}}$ is unbounded almost surely $\mathrm{P}_0^{n\otimes}$ for $\mathrm{P}_0 \in \mathfrak{P} \subset \mathfrak{P}'$. On the other hand, [5, Theorem 3.1] shows that $\frac{d\mathrm{P}_{\bar{z}}^{n\otimes}}{d\mathrm{P}^{n\otimes}}$ is unbounded almost surely $\mathrm{P}^{n\otimes}$. Since $\mathfrak{P}'$ contains equivalent measures, $\frac{d\mathrm{P}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}}$ is unbounded $\mathrm{P}_0^{n\otimes}$-almost surely. it follows that $\frac{d\mathrm{P}_{\bar{z}}^{n\otimes}}{d\mathrm{P}_0^{n\otimes}}$ is unbounded almost surely $\mathrm{P}_0^{n\otimes}$ and hence almost surely $\mathfrak{P}'^{(n)}$.

If B is finite, it suffices to maximize the natural logarithm of (4.12),

$$\ell_k = \frac{n}{2}\left(2(1-\lambda_k)\mu_k\bar{U}_k - \mu_k^2(1-\lambda_k) + \lambda_k S_k^2 + \ln(1-\lambda_k)\right), \tag{4.13}$$

for all k. Taking derivatives, it follows that

$$\frac{\partial \ell_k}{\partial \mu_k} = n(1-\lambda_m)\bar{U}_k - n\mu_k(1-\lambda_k).$$

Since $\sup_k \lambda_k < 1$, $1 - \lambda_k > 0$ for all k and therefore, $\frac{\partial \ell_k}{\partial \mu_k} = 0$ if and only if

$$\mu_k = \bar{U}_k. \tag{4.14}$$

Noting that $\frac{\partial^2 \ell_k}{\partial \mu_k^2} = -n(1-\lambda_k)$, it follows that for each λ_k, ℓ_k is maximized at $\mu_k = \bar{U}_k$.

On the other hand,

$$\begin{aligned}\frac{\partial \ell_k}{\partial \lambda_k} &= \frac{n}{2}(-2\mu_k \bar{U}_k + \mu_k^2 + S_k^2 - \frac{1}{1-\lambda_k}) \\ &= \frac{n}{2}\left(-2\mu_m \bar{U}_k + \mu_k^2 + S_k^2 - 1 - a_k\right) \qquad ((1+a_k)(1-\lambda_k) = 1) \\ &= \frac{n}{2}\left(-2\mu_m \bar{U}_m + \mu_m^2 + S_m^2 - 1 - a_k\right).\end{aligned}$$

Setting this equal to zero, we find that

$$a_k = S_k^2 - 2\mu_k \bar{U}_k + \mu_k^2 - 1.$$

Since ℓ_k is maximized when $\mu_k = \bar{U}_k$, this becomes

$$a_m = \frac{1}{n}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2 - 1, \tag{4.15}$$

Since $1 + a_k = 1/(1+\lambda_k)$ and $\frac{1}{n}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2 > 0$ $\mathrm{P}_0^{n\otimes}$ - almost surely,

$$\lambda_k = 1 - \left(\frac{1}{n}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2\right)^{-1}.$$

Then ℓ_k is maximized $\mathrm{P}_0^{n\otimes}$ - almost surely at $\mu_k = \bar{U}_k$ since

$$\frac{\partial^2 \ell_k}{\partial \lambda_k^2} = -\frac{n}{2}\frac{1}{(1-\lambda_k)^2}.$$

Therefore, (4.14) and (4.15) $\mathrm{P}_0^{n\otimes}$ - almost surely maximize ℓ_k, so that (4.11) is maximized $\mathrm{P}_0^{n\otimes}$ (and thus $\mathfrak{P}^{\prime(n)}$) - almost surely at

$$\hat{\mu}_{kn} = \begin{cases} \bar{U}_k & \text{if } k \in B, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{a}_{kn} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(U_{mi} - \bar{U}_m)^2 - 1 & \text{if } k \in B, \\ 0 & \text{otherwise} \end{cases}$$

□

Note that in this theorem, fixing U_{ki} corresponds to fixing the CONS $\{g_k\}_k$ in $\mathcal{H}(P_0)$.

4.1.1 The Model $\mathfrak{P}_0'$

Since the original parameter space $\mathcal{C}'$ (4.8) is very large, we currently do not have a sieve in $\mathcal{C}'$. However, by Theorem 4.1.2, we can create a sieve in a subset of $\mathcal{C}'$ by fixing the CONS $\{g_k\}_k$ in $\mathcal{H}(P_0)$ and considering the subset $\mathcal{C}_0'$ of $\mathcal{C}'$ defined by

$$\mathcal{C}_0' = \Big\{(\bar{z}, P)\colon P = P_0 + \sum_k a_k g_k \otimes g_k, \mathbf{a} \in \ell_c^2, \bar{z} = \sum_k \mu_k g_k, \boldsymbol{\mu} \in \ell^2.\Big\} \tag{4.16}$$

This parameterizes a subset $\mathfrak{P}_0'$ of $\mathfrak{P}'$. As there is a one-to-one correspondence between $\mathcal{C}$ given in (2.37) and ℓ_c^2 and a one-to-one correspondence between $\bar{z}$ and $\boldsymbol{\mu}$, it follows that there is a one-to-one correspondence between $\mathcal{C}_0'$ and $\ell^2 \times \ell_c^2$. Consider the set $\mathcal{S}_d = \{(\boldsymbol{\mu}, \mathbf{a}) \in \ell^2 \times \ell_c^2\colon \mu_k = a_k = 0 \text{ for } k > d\}$. It is easy to see that $\{\mathcal{S}_d\colon d \in \mathbb{N}\}$ is a sieve in $\ell^2 \times \ell_c^2$ and may be considered as a sieve in $\mathcal{C}_0'$. Therefore, using Theorem 4.1.2 we consider the sieve estimator of (μ_k, a_k) given by

$$\hat{\mu}_{kn} = \begin{cases} \bar{U}_k & \text{if } k \le d, \\ 0 & \text{otherwise} \end{cases} \qquad \textit{and} \tag{4.17}$$

$$\hat{a}_{kn} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(U_{mi} - \bar{U}_m)^2 - 1 & \text{if } k \le d, \\ 0 & \text{otherwise} \end{cases} \tag{4.18}$$

in $\ell^2 \times \ell_c^2$. Note that by Lemma 4.1.1, $\mathbb{E}_{\mathrm{P}_{\bar{z}}}[\hat{\mu}_{kn}] = \mu_k$ and

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\bar{z}}}[\hat{a}_{kn}] &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\frac{1}{n}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2\right] - 1 \\
&= \frac{1}{n}\mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\sum_{i=1}^{n} U_{ki}^2\right] - \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\bar{U}_k^2\right] - 1 \\
&= \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[U_{ki}^2\right] - \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\bar{U}_k^2\right] \\
&= \mathtt{Var}_{\mathrm{P}_{\bar{z}}}(U_{ki}) - \mathtt{Var}_{\mathrm{P}_{\bar{z}}}\left(\bar{U}_k\right) \\
&= (1 + a_k) - \frac{1}{n}(1 + a_k) - 1 \\
&= \frac{n-1}{n}(1 + a_k) - 1 \\
&= \frac{n-1}{n}a_k - \frac{1}{n} \neq a_k.
\end{aligned}$$

Therefore, $\hat{a}_{kn}$ is a biased estimator of a_k and on average will underestimate a_k. To fix this, consider the estimator

$$\tilde{a}_{kn} = \begin{cases} \frac{1}{n-1}\sum_{i=1}^{n}(U_{mi} - \bar{U}_m)^2 - 1 & \text{if } k \leq d, \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

It follows that

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\bar{z}}}[\tilde{a}_{kn}] &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2\right] - 1 \\
&= \frac{n}{n-1}\mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[\frac{1}{n}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2\right] - 1 \\
&= \frac{n}{n-1}\frac{n-1}{n}(1 + a_k) - 1 \\
&= a_k,
\end{aligned}$$

which shows that $\tilde{a}_{kn}$ is an unbiased estimator of a_k. This is the estimator that we shall use for a_k. To the estimators $\hat{\boldsymbol{\mu}}_{nd} = \{\hat{\mu}_{1n}, \ldots, \hat{\mu}_{dn}, 0, 0, \ldots\}$ and $\tilde{\mathbf{a}}_{nd} = \{\hat{a}_{1n}, \ldots, \hat{a}_{dn}, 0, 0, \ldots\}$

corresponds the estimator of $(\bar{z}, P)$ given by

$$\hat{\bar{z}} = \sum_{k=1}^{d} \hat{\mu}_{kn} g_k \qquad \textit{and} \tag{4.20}$$

$$\hat{P} = P_0 + \sum_{k=1}^{d} \tilde{a}_{kn} g_k \otimes g_k. \tag{4.21}$$

in $\mathcal{C}_0'$. We shall discuss the properties of $(\hat{\bar{z}}, \hat{P})$ and $(\hat{\boldsymbol{\mu}}_{nd}, \tilde{\mathbf{a}}_{nd})$ under the additional assumption:

(C4) the true mean and covariance belong to $\mathcal{C}_0'$.

We may parameterize $\mathcal{C}_0'$ by $\ell^2 \times \ell_c^2$, so that the measure that endows $\{Z_t \colon t \in T\}$ with mean $\sum_k \mu_k g_k$ and covariance $P = P_0 + \sum_k a_k g_k \otimes g_k$ may be denoted by $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$. Thus the assumption **(C4)** is that the true measure belongs to $\mathfrak{P}_0' \subset \mathfrak{P}'$.

4.1.2 Asymptotics

To show the pointwise consistency of $(\hat{\bar{z}}, \hat{P})$, we need the Toeplitz Lemma (see e.g. [17]):

Lemma 4.1.2. *Let $x_n' = n^{-1} \sum_{k=1}^{d} x_n$ for sequences $\{x_n\}_n$ and $\{d\} = \{d_n\}$. If $x_n \to 0$ and $d = O(n)$ as $n \to \infty$, then $x_n' \to 0$. If $x_n \to x$ and $d/n \to \alpha$ as $n \to \infty$, then $x_n' \to \alpha x$.*

Also, for estimators $(\hat{\theta}_1, \hat{\theta}_2) \in \mathbb{R}^2$, weak and mean-square consistency are taken to be in the norm of $\mathbb{R}^2$. That is, $(\hat{\theta}_1, \hat{\theta}_2)$ is a P-weakly consistent estimator for (θ_1, θ_2) if for all $\epsilon > 0$, $\mathrm{P}\left(||(\hat{\theta}_1, \hat{\theta}_2) - (\theta_1, \theta_2)||_{\mathbb{R}^2} > \epsilon \right) \to 0$ as $n \to \infty$ and is mean-square consistent if $\mathbb{E}_{\mathrm{P}}\left[||(\hat{\theta}_1, \hat{\theta}_2) - (\theta_1, \theta_2)||_{\mathbb{R}^2}^2 \right] \to 0$ as $n \to \infty$.

Proposition 4.1.2. *For each $d \in \mathbb{N}$, $(\boldsymbol{\mu}, \mathbf{a}) \in \ell^2 \times \ell_c^2$, $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} \in \mathfrak{P}_0'$ and for each $(s,t) \in T \times T$, $(\hat{\bar{z}}, \hat{P})$ is asymptotically unbiased as $d \to \infty$ and is weakly and mean-square consistent for $(\bar{z}, P)$ if $d = O(n)$ as $d \to \infty$.*

Proof. To show the unbiasedness, note that

$$\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\hat{\bar{z}}(t)\right] = \sum_{k=1}^{d} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\hat{\mu}_{kn}\right] g_k(t) \text{ and} \tag{4.22}$$

$$\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\hat{P}(s,t)\right] = P_0(s,t) + \sum_{k=1}^{d} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\tilde{a}_{kn}\right] g_k(s) g_k(t). \tag{4.23}$$

By Lemma 4.1.1, (4.22) and (4.23) become

$$\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\hat{\bar{z}}(t)\right] = \sum_{k=1}^{d} \mu_k g_k(t) \text{ and} \tag{4.24}$$

$$\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\hat{P}(s,t)\right] = P_0(s,t) + \sum_{k=1}^{d} a_k g_k(s) g_k(t). \tag{4.25}$$

Therefore as $d \to \infty$, (4.24) and (4.25) converge to

$$\bar{z}(t) = \sum_{k=1}^{\infty} \mu_k g_k(t) \text{ and} \tag{4.26}$$

$$P(s,t) = P_0(s,t) + \sum_{k=1}^{\infty} a_k g_k(s) g_k(t). \tag{4.27}$$

Thus, $(\hat{\bar{z}}, \hat{P})$ is an asymptotically unbiased estimator for $(\bar{z}, P)$.

For mean-square consistency, it is sufficient to show that $\hat{\bar{z}}(t)$ and $\hat{P}(s,t)$ are mean-square consistent estimators for $\bar{z}(t)$ and $P(s,t)$, respectively. In turn, since mean-square error = variance + bias^2 it suffices to show that the variance of $\hat{\bar{z}}(t)$ and $\hat{P}(s,t)$ converge to 0. Since $\{U_{ki}\}_{k,i}$ is a sequence of independent normal random variables, it follows that $\{\hat{\mu}_{kn}\}_k$ and $\{\hat{a}_{kn}\}_k$ are sequences of independent normal random variables. Therefore, by Lemma 4.1.1,

$$\begin{aligned} \mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\hat{\bar{z}}(t)\right) &= \sum_{k=1}^{d} \mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\hat{\mu}_{kn}\right)\left(g_k(t)\right)^2 \\ &= \frac{1}{n}\sum_{k=1}^{d}(1+a_k)(g_k(t))^2 \end{aligned} \tag{4.28}$$

Note that since $\{a_k\}_k \in \ell_c^2$, it follows that $a_k \to 0$ as $k \to \infty$. Also, note that $\sum_{k=1}^{\infty}(g_k(t))^2$ converges (and equals $P_0(t,t)$) and hence $(g_k(t))^2 \to 0$ as $k \to \infty$. Thus, by the Toeplitz Lemma (Lemma 4.1.2), the expression (4.28) and hence $\mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\hat{\bar{z}}(t)\right)$ converges to 0 as $n \to \infty$. Along with $\hat{\bar{z}}$ being asymptotically unbiased, this shows that $\hat{\bar{z}}(t)$ is mean-square consistent.

By Lemma 4.1.1, $Z_{kn} = \frac{1}{1+a_k}\sum_{i=1}^{n}(U_{ki} - \bar{U}_k)^2$ are independent χ^2_{n-1} random variables. Thus,

$$\begin{aligned}
\mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\hat{P}(s,t)\right) &= \sum_{k=1}^{d}\mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\tilde{a}_{kn}\right)\left(g_k(s)g_k(t)\right)^2 \\
&= \sum_{k=1}^{d}\frac{(1+a_k)^2}{(n-1)^2}\mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(Z_{kn}\right)\left(g_k(s)g_k(t)\right)^2 \\
&= \sum_{k=1}^{d}\frac{(1+a_k)^2}{(n-1)^2}2(n-1)(g_k(s)g_k(t))^2 \\
&= \frac{1}{n-1}\sum_{k=1}^{d}(1+a_k)^2(g_k(s)g_k(t))^2. \qquad (4.29)
\end{aligned}$$

Now, we have $(1+a_k^2) \to 1$ and $(g_k(s)g_k(t))^2 \to 0$ as $k \to \infty$. Therefore, by Lemma 4.1.2, $\frac{1}{n-1}\sum_{k=1}^{d}(1+a_k)^2(g_k(s)g_k(t))^2 \to 0$ as $n \to \infty$. It follows that $\mathtt{Var}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left(\hat{P}(s,t)\right)$ converges to 0 as $n \to \infty$. This and $\hat{P}(s,t)$ being asymptotically unbiased proves the mean-square consistency of $(\hat{\bar{z}}(t), \hat{P}(s,t))$.

Weak consistency follows from the Chebyshev Inequality:

$$\begin{aligned}
&\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}\Big(||(\hat{\bar{z}}(t),\hat{P}(s,t)) - (\bar{z}(t),P(s,t))||_{\mathbb{R}^2} > \epsilon\Big) \\
&\qquad \le \frac{1}{\epsilon^2}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[||(\hat{\bar{z}}(t),\hat{P}(s,t)) - (\bar{z}(t),P(s,t))||^2_{\mathbb{R}^2}\right] \\
&\qquad = \frac{1}{\epsilon^2}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left(\hat{\bar{z}}(t) - \bar{z}(t)\right)^2\right] + \frac{1}{\epsilon^2}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left(\hat{P}(s,t) - P(s,t)\right)^2\right]. \qquad (4.30)
\end{aligned}$$

Since $\hat{\bar{z}}(t)$ and $\hat{P}(s,t)$ are mean-square consistent, it follows that (4.30) converges to 0. Hence for each $s,t \in T$, $(\hat{\bar{z}}(t), \hat{P}(s,t))$ is weakly consistent for $(\bar{z}(t), P(s,t))$. □

This proposition can be seen as a "local" result in the sense that the consistency holds for each $(\boldsymbol{\mu}, \mathbf{a})$ and each $s, t \in T$. We wish to find a "global" consistency result. That is, a result that holds for all $(\boldsymbol{\mu}, \mathbf{a}) \in \ell^2 \times \ell_c^2$. Strong consistency is taken to be in the norm of $\ell^2 \times \ell^2$, so that $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ is strongly consistent for $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ if

$$\mathrm{P}\left(||(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)||_{\ell^2 \times \ell^2} \to 0\right) = 1.$$

As the norm in $\ell^2 \times \ell^2$ is given by $||(\boldsymbol{\mu}, \mathbf{a})||_{\ell^2 \times \ell^2}^2 = ||\boldsymbol{\mu}||_{\ell^2}^2 + ||\mathbf{a}||_{\ell^2}^2$, it shall suffice to show that $\hat{\boldsymbol{\mu}}_{nd}$ and $\tilde{\mathbf{a}}_{nd}$ are $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ - strongly consistent estimators of $\boldsymbol{\mu}$ and $\mathbf{a}$, respectively.

Theorem 4.1.3. *Let $\beta \geq 0$ and $\sigma \in (0, 1)$. For all $(\boldsymbol{\mu}, \mathbf{a}) \in \ell^2 \times \ell_c^2$ and for every sequence $\{d_n\}_{n=1}^{\infty}$ of integers such that $d_n \to \infty$ and $d_n/n^{\sigma} \to \beta$ as $n \to \infty$, one has $||\hat{\boldsymbol{\mu}}_{nd} - \boldsymbol{\mu}||_{\ell^2}^2 \to 0$ $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost surely.*

Proof. Note that if $d_n/n^{\sigma} \to \beta$, then $d_n/n = (1/n^{1-\sigma})(d_n/n^{\sigma}) \to 0$. Also

$$||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||_{\ell^2}^2 = \sum_{k=1}^{d} (\hat{\mu}_{kn} - \mu_k)^2 + \sum_{k>d} \mu_k^2 = X_{n,d} + \sum_{k>d} \mu_k^2,$$

for $X_{n,d} = \sum_{k=1}^{d} (\hat{\mu}_{kn} - \mu_k)^2$. Now, $\sum_{k>d} \mu_k^2 \to 0$ as long as $d \to \infty$. For each $n \in \mathbb{N}$, it one has

$$\begin{aligned} nX_{n,d} &= n \sum_{k=1}^{d} (\hat{\mu}_{kn} - \mu_k)^2 \\ &= \sum_{k=1}^{d} (1 + a_k) \frac{n(\hat{\mu}_{kn} - \mu_k)^2}{1 + a_k}. \end{aligned} \tag{4.31}$$

By Lemma 4.1.1, $\hat{\mu}_{kn} \sim N(\mu_k, 1 + a_k)$ are i.i.d. for $k = 1, \ldots, d$ under $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$. Therefore,

$$Z_{n,k} = \frac{n(\hat{\mu}_{kn} - \mu_k)^2}{1 + a_k}$$

is a set of i.i.d. χ_1^2 random variables and

$$\begin{aligned} X_{n,d} &= \frac{d}{n}\left(\frac{1}{d}\sum_{k=1}^{d}(1+a_k)Z_{n,k}\right) \\ &= \frac{d}{n}\left(\frac{1}{d}\sum_{k=1}^{d}(1+a_k)(Z_{n,k}-1)\right)+\frac{1}{n}\sum_{k=1}^{d}(1+a_k). \end{aligned} \tag{4.32}$$

To deal with the second term in (4.32), note that $1+a_k \to 1$ as $k \to \infty$ and thus by the second statement in Lemma 4.1.2 with $\alpha = 0$,

$$\frac{1}{n}\sum_{k=1}^{d}(1+a_k) \to 0 \text{ as } n \to \infty. \tag{4.33}$$

Let $Y_{n,d} = \frac{1}{d}\sum_{k=1}^{d}(1+a_k)(Z_{n,k}-1)$. The Markov inequality implies that for $\epsilon > 0$,

$$\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}\left((|Y_{n,d}| > \epsilon\right) \leq \frac{1}{\epsilon^4}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[Y_{n,d}^4\right]. \tag{4.34}$$

As $\{Z_{n,k}\}_{k=1}^{d}$ are independent, it follows that

$$\begin{aligned} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[Y_{n,d}^4\right] &= \frac{1}{d^4}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[\sum_{k=1}^{d}(1+a_k)(Z_{n,k}-1)\right]^4\right] \\ &= \frac{1}{d^4}\sum_{m_1+\cdots+m_d=4}\binom{4}{m_1,m_2,\ldots,m_d}(1+a_1)^{m_1}\ldots(1+a_d)^{m_d} \\ &\qquad \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,1}-1)^{m_1}\right]\ldots\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,d}-1)^{m_d}\right]. \end{aligned} \tag{4.35}$$

Since $Z_{n,k}-1$ is mean zero and $\sum m_i = 4$, the only terms in (4.35) that are nonzero are those that have an even degree, that is,

$$\begin{aligned} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[Y_{n,d}^4\right] = \frac{1}{d^4}\Bigg(&\sum_{k=1}^{d}(1+a_k)^4\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,k}-1)^4\right] \\ &+\sum_{k=1}^{d}\sum_{i\neq k}\frac{4!}{2!2!}(1+a_k)^2(1+a_i)^2\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,i}-1)^2\right]\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,k}-1)^2\right]\Bigg). \end{aligned} \tag{4.36}$$

Now, $\{a_k\}_k \in \ell_c^2$ and thus $\sup_k(1+a_k) < c_1$ for some $c_1 > 0$ and $\{Z_{n,k}\}_{k=1}^d$ are identically distributed, so that (4.36) is bounded by

$$\begin{aligned}\frac{1}{d^4}\Big(dc_1^4\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,1}-1)^4\right] &+6d(d-1)c_1^4\left[\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,1}-1)^2\right]\right]^2\Big)\\ &\leq \frac{1}{d^4}(dc_2 + d(d-1)c_2)\\ &= \frac{c_2}{d^2}, \end{aligned} \tag{4.37}$$

where $c_2 = \max\left\{c_1^4\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,1}-1)^4\right], 6c_1^4\left[\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[(Z_{n,1}-1)^2\right]\right]^2\right\}$ is independent of n. If $d_n/n^\sigma \to \beta$ as $n\to\infty$, then by limit comparison, (4.37) is summable. Therefore,

$$\sum_{n=1}^{\infty}\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}\left((|Y_{n,d}|>\epsilon\right) \leq \sum_{n=1}^{d}\frac{c_2}{d^2} < \infty.$$

Thus, if $d_n/n^\sigma \to \beta$ the Borel-Cantelli Lemma implies that $Y_{n,d}\to 0$ and hence $X_{n,d}\to 0$ $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost surely. □

Next, we consider the strong consistency of the estimator $\tilde{\mathbf{a}}_{n,d}$. To show strong consistency, we follow the approach of [6]. The following lemma, taken from [6, Lemma 5.3], finds a bound for the even central moments of a χ^2 random variable.

Lemma 4.1.3. *Let Z be a χ_q^2-distributed random variable with $q\geq 2$ and let $s\in\mathbb{N}$. Then*

$$\mathbb{E}\left[(Z-n)^{2s}\right] \leq [4q]^s(2s)!\frac{q}{q-1}$$

Note that

$$||\tilde{\mathbf{a}}-\mathbf{a}||_{\ell^2}^2 = \sum_{k=1}^{d}(\tilde{a}_{kn}-a_k)^2 + \sum_{k>d}a_k^2 = V_{n,d} + \sum_{k>d}a_k^2,$$

where the second term vanishes as $d\to\infty$. We shall let $d=d_n$ be such that $d_n\to\infty$ as

$n \to \infty$ and then only consider $n \to \infty$. We have

$$\begin{aligned} \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}\left(V_{n,d} > \epsilon\right) &= \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}\left(\left[(n-1)^2 V_{n,d}\right]^r > \left[(n-1)^2\epsilon\right]^r\right) \\ &\leq \frac{\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[(n-1)^2 V_{n,d}\right]^r\right]}{\left[(n-1)^2\epsilon\right]^r}. \end{aligned}$$

Write

$$(n-1)^2 V_{n,d} = \sum_{k=1}^{d} (1+a_k)^2 \left[\Upsilon_{n,k} - (n-1)\right]^2,$$

where $\Upsilon_{n,k} = \frac{(n-1)(\tilde{a}_{kn}+1)}{(1+a_k)}$. By Lemma 4.1.1 and the definition of $\tilde{a}_{kn}$ (4.19), the variables $\Upsilon_{n,k}$ are i.i.d. χ^2_{n-1} random variables for $k = 1, 2, \ldots, d$. Since the variables $\Upsilon_{n,k} - (n-1)$ are independent, the Multinomial Theorem implies

$$\begin{aligned} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[(n-1)^2 V_{n,d}\right]^r\right] = \sum_{r_1+\cdots+r_d=r} \Bigg(&\binom{r}{r_1,\ldots,r_d} \\ &\prod_{k=1}^{d} (1+a_i)^{2r_k} \prod_{k=1}^{d} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[\Upsilon_{n,k} - (n-1)\right]^{2r_i}\right] \Bigg) \end{aligned} \tag{4.38}$$

Note that $\binom{2r}{2r_1,\ldots,2r_d} = (2r)!/\left(\prod_{k=1}^{d}(2r_k)!\right) \geq 1$, so that $(2r)! \geq \prod_{k=1}^{d}(2r_k)!$. By Lemma 4.1.3, it follows that

$$\begin{aligned} \prod_{k=1}^{d} \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[\Upsilon_{n,k} - (n-1)\right]^{2r_i}\right] &< \left[4(n-1)\right]^r \left(\frac{n-1}{n-2}\right)^d \prod_{i=1}^{d} (2r_i)! \\ &\leq \left[4(n-1)\right]^r \left(\frac{n-1}{n-2}\right)^d (2r)! \end{aligned} \tag{4.39}$$

This inequality and (4.38) imply that

$$\begin{aligned}\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}\left[\left[(n-1)^2 V_{n,d}\right]^r\right] &< [4(n-1)]^r \left(\frac{n-1}{n-2}\right)^d (2r)! \\ &\qquad \sum_{r_1+\cdots+r_d=r} \binom{r}{r_1,\ldots,r_d} \prod_{k=1}^{d} (1+a_i)^{2r_k} \\ &= [4(n-1)]^r \left(\frac{n-1}{n-2}\right)^d (2r)! \left(\sum_{k=1}^{d}(1+a_i)^2\right)^r \\ &= \left[4(n-1)\sum_{k=1}^{d}(1+a_i)^2\right]^r \left(\frac{n-1}{n-2}\right)^d (2r)! \qquad (4.40)\end{aligned}$$

This proves a lemma similar to [6, Lemma 5.2]:

Lemma 4.1.4. *For each $\epsilon > 0$ and $r \in \mathbb{N}$, one has $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}(V_{n,d} > \epsilon) < \alpha_{ndr}(\epsilon)$, where*

$$\alpha_{ndr}(\epsilon) = \left[\frac{4}{(n-1)\epsilon}\sum_{k=1}^{d}(1+a_i)^2\right]^r \left(\frac{n-1}{n-2}\right)^d (2r)! \qquad (4.41)$$

We shall choose the variable $r = r_n$ in such a way that makes $\{\alpha_{ndr}\}$ summable over n. As in Theorem 4.1.3, we shall also choose $d = d_n$ such that $d_n/n \to \beta$ as $n \to \infty$. We say that $f_n \sim g_n$ if $f_n/g_n \to 1$ as $n \to \infty$. Parallel to [6, Lemma 5.4], we have the following:

Lemma 4.1.5. *Let $\{r_n\}_n$ and $\{d_n\}_n$ be such that for $\beta \geq 0$ and for some positive scalars σ, τ such that $\sigma + 2\tau = 1$, one has*

(i) $d/n^\sigma \to \beta$ and

(ii) $(r/n^\tau)^r n^{1/4}$ and $r \leq n$ for n large.

Then $\{\alpha_{ndr}(\epsilon)\}$ given in (4.41) is a summable sequence for $\epsilon > 16\beta/e^2$.

Proof. As in Theorem 4.1.3 $1 + a_k \to 1$ as $k \to \infty$. Fix ϵ and set $\alpha_n = \alpha_{ndr}(\epsilon)$. Since $d/n^\sigma \to \beta$, d tends towards infinity at a slower rate than n. Therefore, as $n \to \infty$,

$$\left(\frac{n-1}{n-2}\right)^d \to 1.$$

This and Stirling's Formula, $(2r)! \sim 2\sqrt{\pi}(4r^2/e^2)^r r^{1/2}$, imply that

$$\frac{\alpha_n}{2\sqrt{\pi}} \sim \left[\frac{16}{e^2\epsilon}\frac{1}{(n-1)^\sigma}\sum_{k=1}^{d}(1+a_i)^2\right]^r \left(\frac{r}{(n-1)^\tau}\right)^{2r} r^{1/2},$$

but then it follows that

$$\left[\frac{16}{e^2\epsilon}\frac{1}{(n-1)^\sigma}\sum_{k=1}^{d}(1+a_k)^2\right]^r \left(\frac{r}{(n-1)^\tau}\right)^{2r} r^{1/2} \sim \left[\frac{16}{e^2\epsilon}\frac{1}{n^\sigma}\sum_{k=1}^{d}(1+a_k)^2\right]^r \left(\frac{r}{n^\tau}\right)^{2r} r^{1/2}.$$

By Lemma 4.1.2 and $d/n^\sigma \to \beta$ as $n \to \infty$, it follows that $1/n^\sigma \sum_{k=1}^{d}(1+a_k) \to \beta$ as $n \to \infty$. Then if $\epsilon > 16\beta/e^2$, for n large, one has

$$\frac{16}{e^2\epsilon}\frac{1}{n^\sigma}\sum_{k=1}^{d}(1+a_k) < 1.$$

Note that by limit comparison with $(r/n^\tau)^r n^{1/4}$, the sequence $\left(\frac{r}{n^\tau}\right)^{2r} r^{1/2}$ is summable. Therefore, $\alpha_n/2\sqrt{\pi}$ is summable by limit comparison with $\left(\frac{r}{n^\tau}\right)^{2r} r^{1/2}$. □

Remark 4.1.1. As an example of the sequence $(r/n^\tau)^r n^{1/4}$, let $r = \ln(n)$. Then $(\ln(n)/n^\tau)^{\ln(n)} n^{1/4}$ is summable [6, Example A.1].

From Lemmas 4.1.4 and 4.1.5 and the Borel-Cantelli Lemma, the following theorem is immediate:

Theorem 4.1.4. *If $\{d_n\}_n$ is such that $d_n \to \infty$ and $d_n/n^\sigma \to \beta \geq 0$, for $\sigma \in (0,1)$, then for $\epsilon > 16\beta/e^2$, $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}(V_{n,d} > \epsilon \ i.o.) = 0$.*

By setting $\beta = 0$, we obtain the desired consistency result for $\tilde{\mathbf{a}}_{nd}$:

Corollary 4.1.1. *If $\{d_n\}_n$ is such that $d_n \to \infty$ and $d_n/n^\sigma \to \beta \geq 0$, for $\sigma \in (0,1)$, then $||\tilde{\mathbf{a}}_{nd} - \mathbf{a}||_{\ell^2} \to 0$ $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ - almost surely as $n \to \infty$.*

We end this section with the desired strong consistency result, which follows from Theorem 4.1.3 and Corollary 4.1.1:

Theorem 4.1.5. *Let $\sigma \in (0,1)$. For all $(\boldsymbol{\mu}, \mathbf{a}) \in \ell^2 \times \ell_c^2$ and for every sequence $\{d_n\}_{n=1}^{\infty}$ of integers such that $d_n \to \infty$ and $d_n/n^\sigma \to 0$ as $n \to \infty$, one has $||(\hat{\boldsymbol{\mu}}_{nd}, \tilde{\mathbf{a}}_{nd}) - (\boldsymbol{\mu}, \mathbf{a})||^2_{\ell^2 \times \ell_c^2} \to 0$ $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost surely.*

Remark 4.1.2. Theorem 3.1 in [1] makes a similar assertion to Theorem 4.1.5, although with a strong assumption on the sample paths of $\{Z_t \colon t \in T\}$. However, the theorem in [1] states that for $\beta \geq 0$ such that $d_n/n \to \beta$ as $n \to \infty$, one has $||(\hat{\boldsymbol{\mu}}_{nd}, \tilde{\mathbf{a}}_{nd}) - (\boldsymbol{\mu}, \mathbf{a})||_{\ell^2 \times \ell^2} \to 3\beta$ $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ almost surely. This result holds only for $\beta > 0$ and is not sufficient to show strong consistency of the estimator. A different argument would be necessary for $\beta = 0$, possibly with a different growth rate for d_n. Corollary 4.1.1 avoids this problem.

4.2 Estimation of β

As before, let P and P_0 be the measures that give $\{Z_t \colon t \in T\}$ mean zero and covariance P and P_0, respectively, and let γ_k be such that $\bar{\mathcal{P}}_0 \gamma_k = g_k$ for all k. In addition, let $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ be the true measure that gives $\{Z_t \colon t \in T\}$ mean $\bar{z}$ and covariance P. As stated at the beginning of the chapter, we shall estimate β by

$$\hat{\beta}_n = \sum_{k=1}^{\infty} \hat{b}_{kn} \gamma_k,$$

where $\hat{b}_{kn}$ is of the form (4.2) given by

$$\hat{b}_{kn} = \frac{\hat{c}_{kn}}{1 + \hat{a}_{kn}}.$$

The estimator $\hat{c}_{kn}$ is given in (4.3) and the estimator $\hat{a}_{kn}$ is given by either (2.38) or (4.15), depending on the assumption $\bar{z} = 0$.

4.2.1 Consistency of $\hat{\beta}_n$

Since $\beta \in \mathcal{M}$ and $\mathcal{M}$ is the weak completion of M as in Subsection 3.2.4, to demonstrate convergence of $\hat{\beta}_n$ we must show that it converges in the topology of $\mathcal{M}$. That is, we must show that

$$(\hat{\beta}_n, \eta) \to (\beta, \eta), \text{(almost surely, in probability, etc.)}$$

for all $\eta \in \mathcal{H}(P)$.

Let $\eta \in \mathcal{H}(P)$ be nonzero. Then $\eta \in \mathcal{H}(P_0)$ and has the expansion $\eta = \sum_k \theta_k g_k$ for some square-summable sequence $\{\theta_k\}$, where $\{g_k\}$ is the orthonormal sequence in $\mathcal{H}(P_0)$ fixed by the model $\mathfrak{P}_0'$. Then

$$\begin{aligned}
|(\hat{\beta}_n, \eta) - (\beta, \eta)| &= \left|(\hat{\beta}_n - \beta, \eta)\right| \\
&= \left|\left(\sum_{k=1}^{\infty}(\hat{b}_{kn} - b_k)\gamma_k, \sum_{j=1}^{\infty}\theta_j g_j\right)\right| \\
&= \left|\sum_{k=1}^{\infty}(\hat{b}_{kn} - b_k)\left(\gamma_k, \sum_{j=1}^{\infty}\theta_j g_j\right)\right| \\
&= \left|\sum_{k=1}^{\infty}(\hat{b}_{kn} - b_k)\sum_{j=1}^{\infty}\theta_j(\gamma_k, g_j)\right| \qquad (4.42) \\
&= \left|\sum_{k=1}^{\infty}(\hat{b}_{kn} - b_k)\theta_k\right| \qquad (4.43)
\end{aligned}$$

Note that (4.42) follows from the fact that the functional $(\gamma_k, \cdot)$ is continuous on $\mathcal{H}(P_0)$ (see Remark 3.2.3).

If $\eta = g_k$ for some $k \in \mathbb{N}$, (4.43) becomes

$$|(\hat{\beta}_n, g_k) - (\beta, g_k)| = |\hat{b}_{kn} - b_k|,$$

so that the convergence of $\hat{\beta}_n$ is equivalent to componentwise convergence of the coefficients $\{\hat{b}_{kn}\}$. Note that it is assumed that P and $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ are equivalent measures, so that if an

estimator is P - consistent, then it is also $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ - consistent. In particular, this applies to the estimator $\hat{a}_{kn}$ given in (2.38), which is P - consistent and thus $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ - consistent.

Proposition 4.2.1. *For each $k \in \mathbb{N}$, $\hat{b}_{kn}$ is a $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$ - strongly consistent estimator for b_k.*

Proof. Begin by using a standard argument to write $\hat{c_k}$ in a different form by adding and subtracting $\mathbb{E}_{\mathrm{P}}[U_k]$.

$$\begin{aligned}
\hat{c}_{kn} &= \frac{1}{n}\sum_{j=1}^{n}\left(\hat{w}_j(U_{kj} - \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] + \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] - \bar{U}_k)\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}(\hat{w}_j U_{kj}) - \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] + \left(\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] - \bar{U}_k\right) \\
&= \frac{1}{\frac{1}{n}\sum_{j=1}^{n} W_j}\frac{1}{n}\sum_{j=1}^{n} W_j U_{kj} - \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] + \left(\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] - \bar{U}_k\right) \qquad (4.44)
\end{aligned}$$

For each $j = 1, \ldots, n$, U_{kj} is determined by the j^{th} individual and W_j is observed from the j^{th} individual. Hence, for each $k = 1, \ldots, d$, $\{W_j U_{kj}\}$ is an independent set of random variables.

Clearly, $W_j U_{kj}$ has finite expectation, since U_{kj} is a normal random variable and W_j was assumed to have finite second moment. Therefore, the strong law of large numbers applies. Thus,

$$\frac{1}{n}\sum_{j=1}^{n} W_j U_{kj} \to \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[WU_k], \quad \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} - a.s \text{ as } n \to \infty.$$

Next, by the strong law of large numbers and the continuous mapping theorem,

$$\frac{1}{\frac{1}{n}\sum_{j=1}^{n} W_j} \to \frac{1}{\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[W]}, \quad \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} - a.s \text{ as } n \to \infty.$$

The third term in (4.44) converges to 0 almost surely by the strong law of large numbers.

Hence,

$$\begin{aligned}\hat{c}_{kn} &\to \frac{\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[WU_k]}{\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[W]} - \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] \quad \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} - a.s \text{ as } n \to \infty. \\ &= \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[wU_k] - \mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[w]\,\mathbb{E}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}[U_k] \\ &= \mathtt{Cov}_{\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}}(w, U_k) \\ &= c_k\end{aligned}$$

Thus, we have the strong consistency of $\hat{c}_{kn}$.

Returning to the sieve estimator, an application of the continuous mapping theorem implies that

$$\frac{1}{1+\hat{a}_{kn}} \to \frac{1}{1+a_k} \quad \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} - a.s \text{ as } n \to \infty.$$

Hence,

$$\hat{b}_k = \frac{\hat{c}_{kn}}{1+\hat{a}_{kn}} \to \frac{c_k}{1+a_k} = b_k \quad \mathrm{P}_{\boldsymbol{\mu},\mathbf{a}} - a.s \text{ as } n \to \infty.$$

□

This proves the strong consistency of the estimator $\hat{b}_{kn}$. Let V denote the linear span of the CONS $\{g_k\}$, that is, the set of finite linear combinations of $\{g_k\}$.

Corollary 4.2.1. *$(\hat{\beta}_n, \eta)_{L^2(T)}$ is a $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-strongly consistent estimator of (β, η) for $\eta \in V$.*

Proof. If $\eta \in V$, that is, if $\eta = \sum_{k=1}^{m} \theta_k g_{j_k}$ for a finite subsequence $\{j_k\}$ of $\mathbb{N}$ then

$$\begin{aligned}|(\hat{\beta}_n, \eta) - (\beta, \eta)| &= \left| \sum_{k=1}^{\infty} (\hat{b}_{kn} - b_k) \sum_{i=1}^{m} \theta_{j_i} (\gamma_k, g_{j_i}) \right| \\ &= \left| \sum_{k=1}^{m} \theta_{j_k} (\hat{b}_{j_k n} - b_{j_k}) \right| \\ &\leq \sum_{k=1}^{m} |\theta_{j_k}| \left| \hat{b}_{j_k n} - b_{j_k} \right|. \end{aligned} \tag{4.45}$$

This inequality and Proposition 4.2.1 show that $|(\hat{\beta}_n, \eta)_{L^2(T)} - (\beta, \eta)_{L^2(T)}|$ is $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost

surely convergent to zero. □

Since the finite linear combinations of $\{g_k\}$ are a dense set in $\mathcal{H}(P_0)$, it follows that $(\hat{\beta}_n, \eta)_{L^2(T)}$ converges $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost surely to (β, η) for η in a dense set in $\mathcal{H}(P_0)$. The general case $\eta \in \mathcal{H}(P)$ is at present unproven:

Conjecture 4.2.1. *The estimator $\hat{\beta}_n$ is a $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-strongly consistent estimator for β.*

Remark 4.2.1. It is easy to see that if $\sum_{k=1}^{\infty}(\hat{b}_{kn} - b_k)^2$ converges to 0 almost surely, so does (4.43). This almost sure convergence is also unproven and the conjecture is left in its most general form.

If it can be shown that $\{\hat{b}_{kn}\}_k$ converges uniformly $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-almost surely to $\{b_k\}_k$, then this conjecture is easily proven. The problem lies in definition of b_k, that is, that the estimators $\hat{b}_{kn}$ are not independent and that W is relatively unknown.

4.3 Evolutionary Response to Selection

Equation (3.6) is undefined if the selection gradient β is not in the range of $\mathcal{P}$. We end the chapter with an extension of (3.6) for arbitrary selection gradients. Let P be the phenotypic covariance, G be the additive-genetic covariance and E be the environmental covariance functions of the process $\{Z_t : t \in T\}$. The following lemma due to [2, Corollary IV.2] allows us to compare the spaces $\mathcal{H}(P)$ and $\mathcal{H}(G)$:

Lemma 4.3.1. *Let K_0 and K_1 be two covariance kernels. To have $\mathcal{H}(K_0) \subset \mathcal{H}(K_1)$ (as sets) it is necessary and sufficient that there exist a constant $B > 0$ such that $BK_1 - K_0$ is a covariance kernel.*

Noting that $E = P - G$, it follows from this lemma that $\mathcal{H}(G) \subset \mathcal{H}(P)$ as sets. Suppose that G has finite trace. Let $\mathcal{L}_G$ be the $\sigma(L^2(T), \mathcal{F}_G)$-completion, where $\mathcal{F}_G = \{\langle \cdot, \eta \rangle_G : \eta \in \mathcal{H}(G)\}$. Then we have [7, Proposition A.3]

Proposition 4.3.1. *If P, G, and $\mathcal{L}_G$ are as above and let $\mathcal{L}$ be as in Subsection 3.2.2. If $\bar{\mathcal{G}}$ is the $\mathcal{L}_G$ extension of $\mathcal{G}$, then $\mathcal{L} \subset \mathcal{L}_G$ and for $\beta \in \mathcal{L}$,*

$$\bar{\mathcal{G}}\beta(t) = (\beta, G_t)_{L^2(T)}\,, \tag{4.46}$$

where $G_t(\cdot) = G(t, \cdot)$.

Since $\mathcal{M}$ (Subsection 3.2.4) is a subset of $\mathcal{L}$, (4.46) holds for all $\beta \in \mathcal{M}$. Therefore, we extend the Breeder's Equation (3.6) to $\mathcal{M}$ and obtain

$$\Delta\bar{z} = \bar{\mathcal{G}}\beta. \tag{4.47}$$

This gives us a formula for estimating the evolutionary response to selection if we have an estimate of G and β. The only issue with estimating $\Delta\bar{z}$ using independent organisms, is that there is no clear way to estimate G. In this case, one must either use a previous estimator or assume a given form for G; the latter is used in Chapter 6. Recall that V denotes the linear span of the CONS $\{g_k\}$. The following proposition follows immediately from Corollary 4.2.1, (4.46) and (4.47).

Proposition 4.3.2. *For each $t \in T$ such that $G_t \in V$, the estimator*

$$\widehat{\Delta\bar{z}}(t) = \bar{\mathcal{G}}\hat{\beta}_n(t) \tag{4.48}$$

is a $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-strongly consistent estimator for $\Delta\bar{z}(t)$.

It is currently unclear what conditions guarantee that G_t is an element of the linear span of $\{g_k\}$. On the other hand, since $\mathcal{H}(G) \subset \mathcal{H}(P)$, it follows that $G_t \in \mathcal{H}(P)$. We are then led to the following conjecture:

Conjecture 4.3.1. *For each $t \in T$, the estimator $\widehat{\Delta\bar{z}}(t)$ given in (4.48) is a $\mathrm{P}_{\boldsymbol{\mu},\mathbf{a}}$-strongly consistent estimator for $\Delta\bar{z}(t)$.*

If Conjecture 4.2.1 holds, then this conjecture automatically holds due to (4.46) and (4.47).

Chapter 5

Estimation of β from a Dependent Sample

As in the previous chapter, to estimate the selection gradient β, we must first be able to estimate the covariance function P of a Gaussian process $\{Z(t)\colon t \in T\}$. Since the observations in this chapter will no longer be independent, we must find a new estimator of P. Recall that in the finite-dimensional case, the covariance matrix of the sample was dependent on the relationship matrix $\mathbf{A}$. Let $Z_1(t), Z_2(t), \ldots, Z_n(t)$ be a sample of traits of organisms. Similar to the finite-dimensional case, we assume that for each i, $Z_i(t) = g_i(t) + e_i(t)$, where $g_i(t)$ is the additive-genetic process and $e_i(t)$ is the environmental process of the ith observation, see e.g. [13]. We assume that each organism is raised in independent but similar environments, so that $\{e_i(t)\}_{i=1}^n$ are independent and identically distributed processes. Let $\mathbf{Z}(t) = \begin{bmatrix} Z_1(t) \ldots Z_n(t) \end{bmatrix}^T$, $\mathbf{g}(t) = \begin{bmatrix} g_1(t) \ldots g_n(t) \end{bmatrix}^T$ and $\mathbf{e}(t) = \begin{bmatrix} e_1(t) \ldots e_n(t) \end{bmatrix}^T$, so that $\mathbf{Z}(t) = \mathbf{g}(t) + \mathbf{e}(t)$. It follows that

$$\mathtt{Cov}\left(\mathbf{g}(s), \mathbf{g}(t)\right) = \mathbf{A}G(s,t), \tag{5.1}$$

where $\mathbf{A}$ is the relationship matrix discussed in subsection 3.1.1 and $G(s,t)$ is the additive-genetic covariance function. If the organisms are equally related, it then follows that $A =$

$(1-a)\mathbf{I}_n + a\mathbf{J}_n$, where a is the relationship coefficient of the organisms, $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n^T$ is the $n \times n$ matrix of 1's. The matrix $\mathbf{A}$ is equal to 1 on the diagonal and a on the off-diagonal. Furthermore,

$$\texttt{Cov}\left(\mathbf{e}(t), \mathbf{e}(s)\right) = \mathbf{I}_n E(s,t), \tag{5.2}$$

where $E(s,t)$ is the environmental covariance function. Assuming no genetic-environmental interaction, we have

$$\texttt{Cov}\left(\mathbf{Z}(s), \mathbf{Z}(t)\right) = \mathbf{A}G(s,t) + \mathbf{I}_n E(s,t) \tag{5.3}$$

Then (5.3) becomes

$$\texttt{Cov}\left(\mathbf{Z}(t), \mathbf{Z}(s)\right) = (G(s,t) + E(s,t))\mathbf{I}_n + aG(s,t)(\mathbf{J}_n - \mathbf{I}_n). \tag{5.4}$$

Thus the covariance matrix of $\mathbf{Z}(t)$ is equal to $G(s,t) + E(s,t)$ on the diagonal and $aG(s,t)$ on the off-diagonal. Thus, we shall let

$$P(s,t) := G(s,t) + E(s,t), \tag{5.5}$$

as usual and $\Psi(s,t)$ be defined by

$$\Psi(s,t) := aG(s,t). \tag{5.6}$$

Since G is a covariance function and a is a nonnegative constant, we see that

$$\Psi \text{ is a covariance function.} \tag{5.7}$$

5.1 Estimation of P

Let $Z_1(t), Z_2(t), \ldots, Z_n(t)$ be a sample of traits of equally related organisms. Let $\mathbf{Z}(t) = \begin{bmatrix} Z_1(t) & Z_2(t) & \ldots & Z_n(t) \end{bmatrix}^T$ be defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $\mathcal{A} = \sigma(Z_i(t), \forall i, t \in T)$. The measure $\mathbb{P}$ gives $\mathbf{Z}(t)$ zero mean and covariance

$$\mathbf{P}(s,t) := \mathtt{Cov}_{\mathbb{P}}\left(\mathbf{Z}(t), \mathbf{Z}(s)\right) = (P(s,t) - \Psi(s,t))\mathbf{I}_n + \Psi(s,t)\mathbf{J}_n. \tag{5.8}$$

This matrix has the covariance function $P(s,t)$ on the diagonal and the *cross-covariance function*, $\Psi(s,t)$ on the off diagonal. For each $s, t \in T$, the matrix $\mathbf{P}(s,t)$ has only two eigenvalues: $P(s,t) + (n-1)\Psi(s,t)$ corresponding to the eigenvector $\mathbf{1}_n$ and $P(s,t) - \Psi(s,t)$ of multiplicity $n-1$ with eigenvectors in the orthocomplement of $\mathbf{1}_n$ (see e.g. [4]).

We can diagonalize the covariance matrix (5.8) by

$$\mathbf{D}(s,t) = V^{-1}\mathbf{P}(s,t)V,$$

where $D(s,t)$ is such that $D_{11}(s,t) = P(s,t) + (n-1)\Psi(s,t)$, $D_{ii}(s,t) = P(s,t) - \Psi(s,t)$ for $i = 2, \ldots, n$ and $D_{ij}(s,t) = 0$ for $i \neq j$ and V, is given by $V = [\mathbf{v}_1 \ldots \mathbf{v}_n]$, with columns given by the eigenvectors of $\mathbf{P}(s,t)$. For example, we may use the orthonormal eigenvectors:

$$\begin{aligned}
\mathbf{v}_1 &= \begin{bmatrix} \frac{1}{\sqrt{n}}, & \frac{1}{\sqrt{n}}, & \ldots & , \frac{1}{\sqrt{n}} \end{bmatrix}^T \\
\mathbf{v}_2 &= \begin{bmatrix} \frac{1}{\sqrt{2}}, & \frac{-1}{\sqrt{2}}, & 0, & \ldots, & 0 \end{bmatrix}^T \\
&\vdots \\
\mathbf{v}_i &= \begin{bmatrix} \frac{1}{\sqrt{i(i-1)}}, & \ldots, & \frac{1}{\sqrt{i(i-1)}}, & \frac{-(i-1)}{\sqrt{i(i-1)}}, & 0, & \ldots, & 0 \end{bmatrix}^T \\
&\vdots \\
\mathbf{v}_n &= \begin{bmatrix} \frac{1}{\sqrt{n(n-1)}}, & \ldots, & \frac{1}{\sqrt{n(n-1)}}, & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{bmatrix}^T.
\end{aligned}$$

We may transform $\mathbf{Z}(t)$ to a new random variable by

$$\mathbf{Y}(t) = V^{-1}\mathbf{Z}(t). \tag{5.9}$$

Then $\mathbf{Y}(t)$ satisfies $\mathtt{Cov}_{\mathbb{P}}(\mathbf{Y}(s), \mathbf{Y}(t)) = \mathbf{D}(\mathbf{s}, \mathbf{t})$. It is important to note that $P - \Psi$ is the covariance function of the process $\{Y_i(t) \colon t \in T\}$ for $i = 2, \ldots, n$. Therefore,

$$P - \Psi \text{ is a covariance kernel.} \tag{5.10}$$

Now, let $\mathbb{P}_0$ be the measure on $(\Omega, \mathcal{A})$ that gives $\mathbf{Z}(t)$ zero mean and covariance

$$\mathbf{P}_0(s,t) = (P_0(s,t) - \Psi_0(s,t))\mathbf{I}_n + \Psi_0(s,t)\mathbf{J}_n.$$

Similar to (5.10), we see that

$$P_0 - \Psi_0 \text{ is a covariance kernel.} \tag{5.11}$$

Furthermore, an argument similar to that showing (5.5) and (5.6) shows that

$$P_0(s,t) = G_0(s,t) + E_0(s,t), \text{ and} \tag{5.12}$$

$$\Psi_0(s,t) = aG_0(s,t). \tag{5.13}$$

Therefore,

$$\Psi_0 \text{ is a covariance function.} \tag{5.14}$$

Now, if $\mathbb{P} \sim \mathbb{P}_0$ on $\mathcal{A}$ and $\mathcal{A}_i = \sigma(Y_i(t), t \in T)$, then $\mathcal{A}_i \subset \mathcal{A}$ is a sub-sigma-algebra and $\mathbb{P} \sim \mathbb{P}_0$ on $\mathcal{A}_i$. Let $\mathbb{P}|_{\mathcal{A}_i}$ and $\mathbb{P}_0|_{\mathcal{A}_i}$ be the restrictions of $\mathbb{P}$ and $\mathbb{P}_0$, respectively, to $(\Omega, \mathcal{A}_i)$. It follows that for $i = 2, \ldots, n$, $\mathbb{P}|_{\mathcal{A}_i}$ and $\mathbb{P}_0|_{\mathcal{A}_i}$ give Y_i zero mean and covariance $P - \Psi$ and $P_0 - \Psi_0$, respectively. This allows us to use the equivalent measures $\mathbb{P}|_{\mathcal{A}_i}$ and $\mathbb{P}_0|_{\mathcal{A}_i}$ on each

$\mathcal{A}_i$ to estimate the covariance $P - \Psi$. Similarly, we use the equivalent measures $\mathbb{P}|_{\mathcal{A}_1}$ and $\mathbb{P}_0|_{\mathcal{A}_1}$ to find an estimator of $P + (n-1)\Psi$.

Note that if $K_i(s,t) = \mathtt{Cov}_{\mathbb{P}}(Y_i(s), Y_i(t))$ and $K_{0i}(s,t) = \mathtt{Cov}_{\mathbb{P}_0}(Y_i(s), Y_i(t))$, we can write $K = K_0 + \sum_k \lambda_k h_k \otimes h_k$ for $\{h_k\}$ orthonormal functions in $\mathcal{H}(K_{0i})$. On the other hand K is $P + (n-1)\Psi$ or $P - \Psi$ and K_{0i} is equal to $P_0 + (n-1)\Psi_0$ or $P_0 - \Psi_0$, depending on i. This leads us to consider the spaces $\mathcal{H}(P_0 + (n-1)\Psi_0)$ and $\mathcal{H}(P_0 - \Psi_0)$.

As Y_1 has covariance $P + (n-1)\Psi$, we consider the RKHS $\mathcal{H}(P_0 + (n-1)\Psi_0)$. Let $\Lambda_1 : H_1 \to \mathcal{H}(P_0 + (n-1)\Psi_0)$ be the Loève map, where $H_1 = \overline{\langle Y_1(t), t \in T \rangle}$ (closure with respect to $\mathbb{P}_0|_{\mathcal{A}_1}$). The GDT (Theorem 2.2.2) gives a countable sequence of orthonormal functions, $\{f_k^+\}_{k=1}^{\infty}$ in $\mathcal{H}(P_0 + (n-1)\Psi_0)$ and a square-summable sequence $\{a_k^{+,n}\}$ with $\inf_k a_k^{+,n} > -1$ such that

$$P + (n-1)\Psi = P_0 + (n-1)\Psi_0 + \sum_k a_k^{+,n} f_k^{+,n} \otimes f_k^{+,n}.$$

We also have a sequence $\{\lambda_{k,n}^+\}_k$ with $\sup_k \lambda_{k,n}^+ < 1$ such that

$$\frac{d\mathbb{P}|_{\mathcal{A}_1}}{d\mathbb{P}_0|_{\mathcal{A}_1}} = \exp\left\{ 2^{-1} \sum_k \lambda_{k,n}^+ (U_k^+)^2 + \ln(1 - \lambda_{k,n}^+) \right\},$$

where $(1 + a_k^{+,n})(1 - \lambda_{k,n}^+) = 1$ and $U_{k,1}^{+,n} = \Lambda_1^{-1} f_k^{+,n}$ is a countable orthonormal sequence in H_1.

Similarly, since $Y_i(t)$, $i = 2, \ldots, n$, have covariance $P(s,t) - \Psi(s,t)$, we now consider the RKHS $\mathcal{H}(P_0 - \Psi_0)$. Let $\Lambda_i : H_i \to \mathcal{H}(P_0 - \Psi_0)$ be the Loève map, where $H_i = \overline{\langle Y_i(t), t \in T \rangle}$. From the GDT, we obtain a countable sequence of orthonormal functions, $\{f_k^-\}_k$, in $\mathcal{H}(P_0 - \Psi_0)$ and a square-summable sequence $\{a_k^-\}_k$ with $\inf_k a_k^- > -1$ such that

$$P - \Psi = P_0 - \Psi_0 + \sum_k a_k^- f_k^- \otimes f_k^-.$$

We also obtain a square-summable sequence $\{\lambda_k^-\}$ with $\sup_k \lambda_k^- < 1$ such that

$$\frac{d\mathbb{P}|_{\mathcal{A}_i}}{d\mathbb{P}_0|_{\mathcal{A}_i}} = \exp\left\{2^{-1}\sum_k \lambda_k^-(U_{k,i}^-)^2 + \ln(1-\lambda_k^-)\right\},$$

where $(1+a_k^-)(1-\lambda_k^-) = 1$ and $U_{k,i}^- = \Lambda_i^{-1} f_k^-$ is a countable orthonormal set in H_i.

The corresponding product measure on $(\Omega, \mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_n)$ is

$$\frac{d\mathbb{P}|_{\mathcal{A}_2}}{d\mathbb{P}_0|_{\mathcal{A}_2}} \cdots \frac{d\mathbb{P}|_{\mathcal{A}_n}}{d\mathbb{P}_0|_{\mathcal{A}_n}} = \exp\left\{2^{-1}\sum_k \lambda_k^- \sum_{i=2}^n (U_{k,i}^-)^2 + (n-1)\ln(1-\lambda_k^-)\right\}.$$

Thus, the likelihood function for the covariance is given by

$$\frac{d\mathbb{P}}{d\mathbb{P}_0} = \exp\left\{2^{-1}\sum_k \lambda_{k,n}^+(U_{k,1}^+)^2 + \ln(1-\lambda_{k,n}^+) + \lambda_k^- \sum_{i=2}^n (U_{k,i}^-)^2 + (n-1)\ln(1-\lambda_k^-)\right\} \quad (5.15)$$

This is the likelihood that we wish to maximize. We have a situation similar to that of Lemma 2.3.1:

Lemma 5.1.1. *Let $\{U_k^+\}_k$ and $\{U_{k,i}^-\}_k$, $i = 2, \ldots, n$, $k \in B$, be fixed. If B is infinite the likelihood (5.15) is unbounded over $\ell_c^2(B) \times \ell_c^2(B)$ $\mathbb{P}$-almost surely. If B is finite, the maximum occurs at*

$$\lambda_{k,n}^+ = \begin{cases} 1 - \left(U_k^+\right)^{-2} & k \in B \\ 0 & \textit{otherwise} \end{cases} \quad (5.16)$$

and

$$\lambda_k^- = \begin{cases} 1 - \left[\frac{1}{n-1}\sum_{i=2}^n (U_{k,i}^-)^2\right]^{-1} & k \in B \\ 0 & \textit{otherwise} \end{cases}, \quad (5.17)$$

with corresponding estimates

$$\hat{a}_k^{+,n} = \begin{cases} \left(U_k^+\right)^2 - 1 & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

and

$$\hat{a}_k^- = \begin{cases} \frac{1}{n-1}\sum_{i=2}^n (U_{k,i}^-)^2 - 1 & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

Proof. If B is infinite, note that (5.15) is the product of two likelihoods that are each unbounded almost surely with respect to $\mathbb{P}|_{\mathcal{A}_1}$ and $\mathbb{P}|_{\mathcal{A}_2} \times \cdots \times \mathbb{P}|_{\mathcal{A}_n}$. This and Lemma 2.3.1 imply that (5.15) is unbounded $\mathbb{P}$-almost surely.

Let $\ell = \ln \frac{d\mathbb{P}}{d\mathbb{P}_0} = \sum_k (\ell_k^+ + \ell_k^-)$. It suffices to show that (5.16) and (5.17) maximize the corresponding summands in ℓ. If B is finite, then

$$\frac{\partial \ell_k^+}{\partial \lambda_{k,n}^+} = 2^{-1}\left((U_{k,1}^+)^2 - \frac{1}{1-\lambda_{k,n}^+}\right).$$

Setting this equal to zero yields

$$\lambda_{k,n}^+ = 1 - \left(U_k^+\right)^{-2}.$$

Now,

$$\frac{\partial^2 \ell_k^+}{\partial (\lambda_{k,n}^+)^2} = -\frac{1}{(1-\lambda_{k,n}^+)^2} < 0,$$

which shows that ℓ_k^+ is maximized at $1 - \left(U_k^+\right)^{-2}$ $\mathbb{P}|_{\mathcal{A}_1}$ - almost surely.

Similarly, taking the derivative of ℓ_k^- with respect to λ_k^- yields

$$\frac{\partial \ell_k^-}{\partial \lambda_k^-} = 2^{-1}\left(\sum_{i=2}^n (U_{k,i}^-)^2 - (n-1)\frac{1}{1-\lambda_k^-}\right).$$

Therefore, $\frac{\partial \ell_k^-}{\partial \lambda_k^-} = 0$ when

$$\lambda_k^- = 1 - (n-1)\left[\sum_{i=2}^n (U_{k,i}^-)^2\right]^{-1}.$$

Then

$$\frac{\partial^2 \ell_k^-}{\partial(\lambda_n^-)^2} = -(n-1)\frac{1}{(1-\lambda_k^-)^2} < 0,$$

and thus ℓ_k^- is maximized $\mathbb{P}|_{\mathcal{A}_2} \times \cdots \times \mathbb{P}|_{\mathcal{A}_n}$ - almost surely. Therefore, ℓ is maximized $\mathbb{P}$ - almost surely at (5.16) and (5.17). □

Although it is valid to construct an estimator of $\mathbf{P}$ using only one family, we only have one observation Y_1 that can be used to estimate the covariance function $P + (n-1)\Psi$. Such an estimator is inherently independent of the sample size n and nothing can be said about the asymptotic properties of this estimator and hence of an estimator of $\mathbf{P}$. In order to have multiple observations of the process Y_1, we consider a sample consisting of unrelated families.

Suppose that we have a sample of traits of m unrelated families, with the j^{th} family consisting of n_j equally related organisms. Let $\mathbf{Z} = [\mathbf{Z}_1(t)^T, \ldots, \mathbf{Z}_m(t)^T]^T$ be defined on the probability space $(\Omega, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$, where $\mathbf{Z}_j(t) = [Z_{j,1}(t), Z_{j,2}(t), \ldots, Z_{j,n_j}(t)]^T$, $j = 1, \ldots, m$ and where $\tilde{\mathcal{A}} = \sigma(Z_{j,1}(t), Z_{j,2}(t), \ldots, Z_{j,n_j}(t), \forall j, t \in T)$. For each j, $\tilde{\mathbb{P}}$ gives $\mathbf{Z}_j(t)$ the covariance of the form in (5.8). It follows that the covariance function of $\mathbf{Z}$ under $\tilde{\mathbb{P}}$ is a block diagonal matrix with entries of the form (5.8). The vectors

$$\{\mathbf{Z}_j(t)\}_{j=1}^m = \left\{ \begin{bmatrix} Z_{j,1}(t) & Z_{j,2}(t) & \ldots & Z_{j,n_j}(t) \end{bmatrix}^T \right\}_{i=1}^m$$

are independent and may be transformed by using the transformation (5.9) to a set of random variables $\{\mathbf{Y}_j(t)\}_{j=1}^m$. Now, each of these $\mathbf{Y}_j(t)$ can be considered itself as a random sample of unrelated individuals and thus, we may now consider the new restrictions of $\tilde{\mathbb{P}}$ to $\tilde{\mathcal{A}}_j = \sigma(Y_{j,1}(t), Y_{j,2}, \ldots, Y_{j,n_j}, \forall t \in T)$. Note that $\tilde{\mathcal{A}}_i$ is the same σ-algebra as $\mathcal{A}$ above. The restricted probability measure, $\tilde{\mathbb{P}}\Big|_{\tilde{\mathcal{A}}_j}$ on $(\Omega, \tilde{\mathcal{A}}_j)$ is the measure $\mathbb{P}$ in the computation of (5.15). Let $\Lambda_{j,i}$ be the Loève map corresponding to $Y_{j,i}$ for $j = 1, \ldots, m$ and $i = 1, \ldots, n_j$.

Thus, the likelihood function of the covariance is given by

$$\frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{P}}_0} = \exp\Bigg\{ 2^{-1} \sum_k \Bigg(\sum_{j=1}^{m} \left[\lambda_{k,n_j}^{+} (U_{k,j}^{+})^2 + \ln(1 - \lambda_{k,n_j}^{+}) \right] + \lambda_k^{-} \sum_{j=1}^{m} \left[\sum_{i=2}^{n_j} (U_{k,j,i}^{-})^2 + (n_j - 1)\ln(1 - \lambda_k^{-}) \right] \Bigg) \Bigg\}, \tag{5.18}$$

where $U_{k,j,i}^{-}$ is the realization of the ith observation in the jth family of $U_{k,j}^{-}$ and satisfies $\Lambda_{j,i} U_{k,j,i}^{-} = f_k^{-}$, for $j = 1, \ldots, m$ and $i = 2, \ldots, n_j$, where $\{f_k^{-}\}_k$ is a countable ONS in $\mathcal{H}(P_0 - \Psi_0)$. Also, $\{U_{k,j}^{+}\}_k$ is the realization of U_k^{+} corresponding to the jth family and satisfies $\Lambda_{j,1}(U_{k,j}^{+}) = f_k^{+,j}$ for $\{f_k^{+,j}\}_k$ a countable ONS of $\mathcal{H}(P_0 + (n_j - 1)\Psi_0)$. Let

$$N = \sum_j n_j, \tag{5.19}$$

and let $\ell_c^2(B)^{(m+1)\otimes} = \ell_c^2(B) \times \cdots \times \ell_c^2(B)$ denote the $m+1$-fold product of $\ell_c^2(B)$. We maximize the likelihood given by (5.18):

Theorem 5.1.1. *Let $\{U_{k,j}^{+}\}_{k,j}$ and $\{U_{k,j,i}^{-}\}_{k,j,i}$, $i = 2, \ldots, n$, $j = 1, \ldots, m$ and $k \in B$, be fixed. If B is infinite, the likelihood (5.18) is $\tilde{\mathbb{P}}$-almost surely unbounded over $\ell_c^2(B)^{(m+1)\otimes}$. If B is finite, (5.18) is $\tilde{\mathbb{P}}$-almost surely maximized over $\ell_c^2(B)^{(m+1)\otimes}$ at*

$$\lambda_{k,n_j}^{+} = \begin{cases} 1 - \left(U_{k,j}^{+}\right)^{-2} & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

and

$$\lambda_k^{-} = \begin{cases} 1 - (N - m)\left[\sum_{j=1}^{m} \sum_{i=2}^{n_j} (U_{k,j,i}^{-})^2\right]^{-1} & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

with corresponding estimates

$$\hat{a}^+_{k,n_j} = \begin{cases} \left(U^+_{k,j}\right)^2 - 1 & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

and

$$\hat{a}^-_k = \begin{cases} \frac{1}{N-m}\left[\sum_{j=1}^m \sum_{i=2}^{n_j} (U^-_{k,j,i})^2\right] - 1 & k \in B \\ 0 & \textit{otherwise} \end{cases}$$

Proof. The likelihood (5.18) is the product of m likelihoods of the form (5.15) that are each almost surely unbounded with respect to $\tilde{\mathbb{P}}\Big|_{\tilde{A}_j}$. Therefore, if follows that (5.18) is unbounded $\tilde{\mathbb{P}}$-almost surely.

The proof of the maximality of the above estimates is very similar to that of Lemma 5.1.1 and is omitted. □

In light of this theorem, we fix the orthonormal sequences $\{f^-_k\}_k$ and $\{f^{+,j}_k\}_k$, $j = 1, \dots m$ and consider the sets $\mathcal{S}_d$ defined by

$$\mathcal{S}_d = \Big\{ \big(\{a^+_{k,n_1}\}_k, \dots, \{a^+_{k,n_m}\}_k, \{a^-_k\}_k\big) \in \ell^2_c(\mathbb{N})^{(m+1)\otimes} : \\ a^-_k = a^+_{k,n_1} = \cdots = a^+_{k,n_m} = 0 \text{ for } k > d \Big\}.$$

It follows that the collection $\{\mathcal{S}_d \colon d \in \mathbb{N}\}$ is a sieve in $\ell^2_c(\mathbb{N})^{(m+1)\otimes}$. We use the sieve estimators

$$\hat{a}^+_{k,n_j} = \begin{cases} \left(U^+_{k,j}\right)^2 - 1 & k \le d \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{a}^-_k = \begin{cases} \frac{1}{N-m}\left[\sum_{j=1}^m \sum_{i=2}^{n_j} (U^-_{k,j,i})^2\right] - 1 & k \le d \\ 0 & \text{otherwise} \end{cases}$$

Presently, there is not much that can be said about the asymptotic properties of the estimator

$\hat{a}^+_{k,n_j}$, as it is computed using only one random variable. On the other hand, we have the following lemma:

Lemma 5.1.2. *Let $m = m_N$ be such that $N - m_N \to \infty$ and $m_N \to \infty$ as $N \to \infty$. The estimator $\{\hat{a}^-_k\}$ is a $\tilde{\mathbb{P}}$ - strongly consistent estimator for $\{a^-_k\}$ in ℓ^2_c.*

Proof. The lemma is immediate, since $\hat{a}^-_k$ is the same form as $\hat{a}_k$ in Lemma 2.3.1. □

We now find an estimator of P using the estimators of $P - \Psi$ and $P + (n_j - 1)\Psi$. Now, for each j, we can write $n_j P = (P + (n_j - 1)\Psi) + (n_j - 1)(P - \Psi)$, so that

$$NP = \sum_j (P + (n_j - 1)\Psi) + \sum_j (n_j - 1)(P - \Psi)$$

and therefore,

$$P = \sum_j \frac{1}{N}(P + (n_j - 1)\Psi) + \frac{N - m}{N}(P - \Psi). \tag{5.20}$$

Combining the expansions for $P + (n_j - 1)\Psi$ and $P - \Psi$, we have

$$P - P_0 = \sum_k \sum_j \frac{a^+_{k,n_j}}{N} f^{+,j}_k \otimes f^{+,j}_k + \sum_k \frac{N - m}{N}(a^-_k) f^-_k \otimes f^-_k.$$

As in the independent case, we have the following expansion of P in $\mathcal{H}(P_0)$:

$$P = P_0 + \sum_k a_k g_k \otimes g_k,$$

where $\{g_k\}$ is a fixed CONS of $\mathcal{H}(P_0)$. Therefore, we have

$$\sum_k a_k g_k \otimes g_k = \sum_k \sum_j \frac{a^+_{k,n_j}}{N} f^{+,j}_k \otimes f^{+,j}_k + \sum_k \frac{N - m}{N}(a^-_k) f^-_k \otimes f^-_k. \tag{5.21}$$

With our assumptions, we note that $P_0 - (P_0 - \Psi_0) = \Psi_0$ and for each $n \in \mathbb{N}$, $nP_0 - (P + (n-1)\Psi_0) = (n-1)(P_0 - \Psi_0)$ are both covariances by (5.14) and (5.11), respectively. Therefore, by Lemma 4.3.1 we have proven the following:

Lemma 5.1.3. *As sets of functions,* $\mathcal{H}(P_0 - \Psi_0) \subset \mathcal{H}(P_0)$ *and* $\mathcal{H}(P_0 + (n-1)\Psi_0) \subset \mathcal{H}(P_0)$ *for all* $n \in \mathbb{N}$.

Therefore, since $\{f_k^{+,j}\}_{k,j=1}^{\infty,m}$ are functions in $\mathcal{H}(P_0 + (n_j - 1)\Psi_0)$ for $j = 1, \ldots, m$, it follows that $f_k^{+,j} \in \mathcal{H}(P_0)$ for all $k \in \mathbb{N}$ and $j = 1, \ldots, m$. Similarly, $f_k^- \in \mathcal{H}(P_0)$ for all $k \in \mathbb{N}$.

For each $s \in T$, apply the linear functional $\langle \cdot, g_\ell \rangle_{P_0}$ to both sides of the equation

$$\sum_k a_k g_k(s) g_k = \sum_k \sum_j \frac{a_{k,n_j}^+}{N} f_k^{+,j}(s) f_k^{+,j} + \sum_k \frac{N-m}{N} (a_k^-) f_k^-(s) f_k^-,$$

to obtain the equation

$$a_\ell g_\ell = \sum_k \sum_j \left[\frac{a_{k,n_j}^+}{N} \left\langle f_k^{+,j}, g_\ell \right\rangle_{P_0} f_k^{+,j} \right] + \frac{N-m}{N} (a_k^-) \left\langle f_k^-, g\ell \right\rangle_{P_0} f_k^-.$$

Another application of $\langle \cdot, g_\ell \rangle_{P_0}$ yields

$$a_\ell = \sum_k \left(\sum_{j=1}^m \left[\frac{a_{k,n_j}^+}{N} \left\langle f_k^{+,j}, g_\ell \right\rangle_{P_0}^2 \right] + \frac{N-m}{N} (a_k^-) \left\langle f_k^-, g\ell \right\rangle_{P_0}^2 \right). \tag{5.22}$$

We can use this to estimate a_ℓ by using the sieve estimators for $a_k^{+,j}$ and a_k^-, so that

$$\hat{a}_\ell = \begin{cases} \sum_k \left(\sum_{j=1}^m \left[\frac{\hat{a}_{k,n_j}^+}{N} \left\langle f_k^{+,j}, g_\ell \right\rangle_{P_0}^2 \right] + \frac{N-m}{N} (\hat{a}_k^-) \left\langle f_k^-, g_\ell \right\rangle_{P_0}^2 \right) & \text{if } k \le d \\ 0 & \text{otherwise} \end{cases} \tag{5.23}$$

Estimation when $n_j = n$

When the families are the same size, i.e. $n_j = n$ for all j, (5.18) becomes

$$\frac{d\tilde{\mathbb{P}}}{d\tilde{\mathbb{P}}_0} = \exp\Bigg\{ 2^{-1} \sum_k \Bigg(\lambda_{k,n}^+ \sum_j (U_{k,j}^{+,n})^2 + m \ln(1 - \lambda_{k,n}^+) + \\ \lambda_k^- \sum_{j=1}^m \sum_{i=2}^n (U_{k,j,i}^-)^2 + m(n-1) \ln(1 - \lambda_k^-) \Bigg) \Bigg\}, \tag{5.24}$$

The sieve estimators of $\lambda_{k,n}^+$ and λ_k^- are then

$$\hat{\lambda}_{k,n}^+ = \begin{cases} 1 - \left\{ \frac{1}{m} \sum_j \left(U_{k,j}^{+,n} \right)^2 \right\}^{-1} & k \leq d \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{\lambda}_k^- = \begin{cases} 1 - \left\{ \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=2}^n (U_{k,j,i}^-)^2 \right\}^{-1} & k \leq d \\ 0 & \text{otherwise} \end{cases},$$

corresponding to the estimators

$$\hat{a}_{k,n}^+ = \begin{cases} \frac{1}{m} \sum_j \left(U_{k,j}^{+,n} \right)^2 - 1 & k \leq d \\ 0 & \text{otherwise} \end{cases}$$

and

$$\hat{a}_k^- = \begin{cases} \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=2}^n (U_{k,j,i}^-)^2 - 1 & k \leq d \\ 0 & \text{otherwise} \end{cases}$$

From this, (5.22) becomes

$$a_\ell = \sum_k \left[\frac{a_k^{+,n}}{n} \left\langle f_k^{+,n}, g_\ell \right\rangle_{P_0}^2 + \frac{(n-1)}{n} (a_k^-) \left\langle f_k, g_\ell \right\rangle_{P_0}^2 \right]. \tag{5.25}$$

The asymptotic properties of $\{\hat{a}_k\}_k$ in either case are currently unknown. The next chapter will discuss the convergence of $||\{\hat{a}_k\} - \{a_k\}||_{\ell^2}$ based on simulations.

5.1.1 Estimation of the Genetic Covariance Function

In the case of independent families of equally-related organisms, we are also able to construct an estimate for the additive-genetic covariance function G. Similarly to (5.20), for each j we may write

$$n_j \Psi = P + (n_j - 1)\Psi - (P - \Psi). \tag{5.26}$$

Summing over j and solving for Ψ, we have

$$\Psi = \frac{1}{N}\sum_{j=1}^{m}(P + (n_j - 1)\Psi) - \frac{m}{N}(P - \Psi).$$

Therefore, we may estimate Ψ by

$$\hat{\Psi} = \frac{1}{N}\sum_{j=1}^{m}\widehat{P + (n_j - 1)\Psi} - \frac{m}{N}\widehat{P - \Psi}. \tag{5.27}$$

On the other hand, by (5.6) $\Psi = aG$. Since $a \neq 0$ for related individuals, it follows that $G = \Psi/a$ and thus G may be estimated by

$$\hat{G} = \frac{1}{a}\hat{\Psi} = \frac{1}{aN}\sum_{j=1}^{m}\widehat{P + (n_j - 1)\Psi} - \frac{m}{aN}\widehat{P - \Psi} \tag{5.28}$$

As with the estimation of P, we do not have any theory about any consistency of $\hat{G}$. The next chapter will discuss convergence based on simulations of equally related individuals.

5.2 Estimation of β

In estimating β using an independent sample, we used the estimator $\hat{c}_{kn}$ given in (4.3). In the dependent case, we lack an estimator for c_k as the distribution of the fitness function W and hence of $\hat{w}_i$ is unknown. Lacking such information, we shall use the same estimator of c_k as in the independent case. That is,

$$\hat{c}_{kn} = \frac{1}{n}\sum_{j=1}^{n}\hat{w}_j(U_{kj} - \bar{U}_k).$$

Let $\{g_k\}$ be a CONS of $\mathcal{H}_{P_0}$. As in Section 4.2, we estimate β by

$$\hat{\beta}_n = \sum_{k=1}^{\infty}\hat{b}_{kn}\gamma_k,$$

where γ_k is such that $\bar{\mathcal{P}}_0(\gamma_k) = g_k$ for all k and $\hat{b}_{kn} = \hat{c}_{kn}/(1 + \hat{a}_{kn})$. Consistency of this estimator will be studied based on simulations in the next chapter.

Chapter 6

Simulations

We now turn our focus from theory to application. We begin with the computational aspects of the simulations, in particular, computing the CONS $\{g_k\}_k$ and the values of the coefficients $c_k = \texttt{Cov}\,(w, U_k)$ in the expansion of the selection differential s in $\mathcal{H}(P_0)$.

6.1 Computation

6.1.1 Orthonormal Functions and Generalized Functions

Both the mean function $\bar{z}$ and the covariance function P can be written in terms of a complete orthonormal sequence $\{g_k\}_k$ in the space $\mathcal{H}(P_0)$, see (4.26) and (4.27). However, the sequence $\{g_k\}_k$ has been relatively arbitrary. We wish to find a CONS that is easy to calculate and allows us to find the sequence $\{\gamma_k\}_k$ such that $\bar{\mathcal{P}}_0\gamma_k = g_k$ for all k. By [21, Theorem 5B] the set $\{P_{0t}, t \in T\}$ is a spanning set in $\mathcal{H}(P_0)$. Define a pseudo-metric on T by

$$d_{P_0}(s,t) = ||P_{0t} - P_{0s}||_{P_0} = P_0(s,s) - 2P_0(s,t) + P_0(t,t). \tag{6.1}$$

A sufficient condition for d_{P_0} to be a metric is that $\{P_{0t}, t \in T\}$ be linearly independent [18, Lemma 4.2]. Furthermore [18, Lemma 4.3], $\mathcal{H}(P_0)$ is separable if and only if (T, d_{P_0}) is separable. Thus, if $\{P_{0t}, t \in T\}$ is linearly independent and $\mathcal{H}(P_0)$ is separable, there exists

a countable dense set $T_0 = \{t_i\}_i$ such that $\{P_{0t}, t \in T_0\}$ is dense in $\mathcal{H}(P_0)$. In general, the set $\{P_{0t}, t \in T\}$ is not linearly independent. However, suppose that P_0 is of the form

$$P_0(s,t) = u(s \wedge t)v(s \vee t) \quad s,t \in [a,b], \tag{6.2}$$

where u, v are nonzero on (a,b), $\frac{u}{v}$ is positive and strictly increasing on (a,b), $s \wedge t = \min(s,t)$ and $s \vee t = \max(s,t)$. Such a covariance function is said to be a *product type covariance function*. We have the following lemma, given in [3, Lemma 5.1.5]:

Lemma 6.1.1. *Let $\{Z_t \colon t \in [a,b]\}$ be a stochastic process with product type covariance function K. Then for all $r \in \mathbb{N}$ and $a < t_1 < \cdots < t_r < b$ the set of functions $\{K_{t_i}\}_{i=1}^r$ are linearly independent.*

For practical applications, we are only able to create a finite number of functions in the CONS $\{g_k\}_k$. To compute the functions $\{g_k\}_{k=1}^r$, we take r to be large and choose $T_r = \{t_i\}_{i=1}^r \subset T_0$. Next, we perform Gram-Schmidt orthonormalization on $\{P_{0t}, t \in T_r\}$. Usually, to perform Gram-Schmidt orthonormalization, one needs to use an algorithm and compute the functions sequentially. Fortunately, in a RKHS, we are able to compute all of the functions $\{g_k\}_{k=1}^r$ at once. Let $\mathbf{P} = [P_0(t_i,t_j)]_{i,j=1}^{m,m}$ be the matrix of P_0 restricted to $T_r \times T_r$, which is invertible as the rows are composed of linearly independent functions restricted to T_r. Since $\mathbf{P}$ is a symmetric positive-definite matrix, it has a Cholesky decomposition $\mathbf{P} = \mathbf{N}\mathbf{N}^T$, where $\mathbf{N}$ is an invertible lower triangular matrix. Now, Gram-Schmidt orthonormalization of the functions $\{P_{0t}, t \in T_r\}$ is equivalent [24, Definition 5.10] to computing $g_k = \sum_{i=1}^r q_{ki} P_{0t_i}$, where

$$q_{ki} = \left(\mathbf{N}^{-1}\right)_{ki}. \tag{6.3}$$

Alternatively, we may write

$$\begin{bmatrix} g_1(\cdot) \\ \vdots \\ g_r(\cdot) \end{bmatrix} = \mathbf{N}^{-1} \begin{bmatrix} P_{0t_1}(\cdot) \\ \vdots \\ P_{0t_r}(\cdot) \end{bmatrix}. \tag{6.4}$$

In addition to calculating the sequence $\{g_k\}_k$, we must also calculate the sequence $\{U_k\}_k$. This calculation is almost immediate if we use (2.7), which states that $\Lambda_{\mathrm{P}_0}(Z_s) = P_0(s,\cdot)$. Then

$$\begin{aligned}
\begin{bmatrix} U_1(\cdot) \\ \vdots \\ U_r(\cdot) \end{bmatrix} &= \Lambda_{\mathrm{P}_0}^{-1} \left(\mathbf{N}^{-1} \begin{bmatrix} P_{0t_1}(\cdot) \\ \vdots \\ P_{0t_r}(\cdot) \end{bmatrix} \right) \\
&= \mathbf{N}^{-1} \begin{bmatrix} \Lambda_{\mathrm{P}_0}^{-1} P_{0t_1}(\cdot) \\ \vdots \\ \Lambda_{\mathrm{P}_0}^{-1} P_{0t_r}(\cdot) \end{bmatrix} \\
&= \mathbf{N}^{-1} \begin{bmatrix} Z_{t_1} \\ \vdots \\ Z_{t_r} \end{bmatrix}. \qquad (6.5)
\end{aligned}$$

In order to calculate the selection gradient, β, we must find the sequence $\{\gamma_k\}_k$ such that $\bar{\mathcal{P}}_0 \gamma_k = g_k$ for all k. Let $\{g_k\}_{k=1}^r$ be as in (6.4) and let δ_t be the δ function at t. To find γ_k, note that $g_k = \sum_{i=1}^r q_{ki} P_{0t_i}$. The discussion following Lemma 3.2.2 shows that $\bar{\mathcal{P}}_0 \delta_t = P_{0t}$, which implies that

$$\begin{aligned}
g_k &= \sum_{i=1}^k q_{ki} P_{0t_i} \\
&= \sum_{i=1}^k q_{ki} \bar{\mathcal{P}}_0 \delta_{t_i} \\
&= \bar{\mathcal{P}}_0 \left(\sum_{i=1}^k q_{ki} \delta_{t_i} \right).
\end{aligned}$$

As $\bar{\mathcal{P}}_0$ is a bijection between $\mathcal{M}$ and $\mathcal{H}(P_0)$, it follows that for each $k = 1, \ldots, m$,

$$\gamma_k = \sum_{i=1}^k q_{ki} \delta_{t_i}.$$

Alternatively, we may symbolically write

$$\begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_r \end{bmatrix} = \mathbf{N}^{-1} \begin{bmatrix} \delta_{t_1} \\ \vdots \\ \delta_{t_r} \end{bmatrix}$$

When estimating the covariance function of a set of unrelated families consisting of equally related individuals (Section 5.1), we must compute the inner products $\left\langle f_k^{+,j}, g_\ell \right\rangle_{P_0}$ and $\left\langle f_k^{-}, g_\ell \right\rangle_{P_0}$. We compute g_ℓ by (6.4), which states that

$$g_\ell = \sum_{i=1}^{r} q_{\ell i} P_{0t_i},$$

where $q_{\ell i} = \left(\mathbf{N}^{-1}\right)_{\ell i}$. Therefore, for any $\eta \in \mathcal{H}(P_0)$,

$$\langle \eta, g_\ell \rangle_{P_0} = \sum_{i=1}^{r} q_{\ell i} \langle \eta, P_{0t_i} \rangle_{P_0} = \sum_{i=1}^{r} q_{\ell i} \eta(t_i), \tag{6.6}$$

where the last equality follows from the reproducing property of $\mathcal{H}(P_0)$. Then we may write

$$\begin{bmatrix} \left\langle f_k^{+,j}, g_1 \right\rangle_{P_0} \\ \vdots \\ \left\langle f_k^{+,j}, g_r \right\rangle_{P_0} \end{bmatrix} = \mathbf{N}^{-1} \mathbf{f}_k^{+,j},$$

where

$$\mathbf{f}_k^{+,j} = \begin{bmatrix} f_k^{+,j}(t_1) \\ \vdots \\ f_k^{+,j}(t_r) \end{bmatrix}.$$

It follows that if $\mathbf{F}^{+,j} = \begin{bmatrix} \mathbf{f}_1^{+,j} & \dots & \mathbf{f}_r^{+,j} \end{bmatrix}$, then

$$\begin{bmatrix} \left\langle f_1^{+,j}, g_1 \right\rangle_{P_0} & \dots & \left\langle f_r^{+,j}, g_1 \right\rangle_{P_0} \\ \vdots & \ddots & \vdots \\ \left\langle f_1^{+,j}, g_r \right\rangle_{P_0} & \dots & \left\langle f_r^{+,j}, g_r \right\rangle_{P_0} \end{bmatrix} = \mathbf{N}^{-1}\mathbf{F}^{+,j}.$$

Similarly, we may write

$$\begin{bmatrix} \left\langle f_1^{-}, g_1 \right\rangle_{P_0} & \dots & \left\langle f_r^{-}, g_1 \right\rangle_{P_0} \\ \vdots & \ddots & \vdots \\ \left\langle f_1^{-}, g_r \right\rangle_{P_0} & \dots & \left\langle f_r^{-}, g_r \right\rangle_{P_0} \end{bmatrix} = \mathbf{N}^{-1}\mathbf{F}^{-},$$

where

$$\mathbf{F}^{-} = \begin{bmatrix} f_1^{-}(t_1) & \dots & f_r^{-}(t_1) \\ \vdots & \ddots & \vdots \\ f_1^{-}(t_r) & \dots & f_r^{-}(t_r) \end{bmatrix}$$

6.1.2 Computing Parameter Values

To be able to run any meaningful simulations, we must be able to construct a covariance function of the form (4.4), a mean function $\bar{z}$ in $\mathcal{H}(P_0)$, the selection differential, s, and the selection gradient β. To create these, we need to know the true values of $\{a_k\}$, $\{b_k\}$ and $\{c_k\}$. By Lemma 3.2.2, once c_k and a_k are known, then b_k is easily found. The only restriction on the sequence $\{a_k\}$ is that it belongs to ℓ_c^2, which means that if we wish to create the true covariance function, we need only specify $\{a_k\}$. Similarly, to construct the true mean function, we need only specify the sequence $\{\mu_k\}$ in ℓ^2. Unfortunately, the sequence $\{c_k\}$ is not an arbitrary sequence. By (4.1), we see that $c_k = \mathtt{Cov}_{\mathrm{P}}(w, U_k)$ for all k. We already know how to write U_k in terms of $\{Z_{t_i}\}_{i=1}^r$. As for the relative fitness, we must know something about the fitness function W. As in Subsection 3.2.3, we assume that W is frequency independent. Further, we assume that $W = \nu(X)$ for some $X \in H_{\mathrm{P}}$. We shall consider

several different forms of ν and give specific biological applications of such functions. As in [7] we assume that X is of one of the following two forms:

$$X = Z(t^*), \tag{6.7}$$

for some $t^* \in T$, or

$$X = \int_T Z(t) f(t) d\mu(t), \tag{6.8}$$

where f is a square integrable function on $(T, \mathfrak{T}, \mu)$. If reproduction occurs at a single stage in life, then the ability to reproduce depends only on the trait at time of reproduction and the most appropriate form of X is (6.7). Such is the case in annual plants and many species of Pacific salmon. If the ability to reproduce at any age depends on the trait, then X of the form (6.8) is appropriate. Since X is defined in terms of $Z(t)$, it is not independent of $Z(t)$. Let μ_X and σ_X^2 denote the mean and variance of X and let $\rho_{i,X}$ denote the correlation of Z_{t_i} and X under the measure $\mathrm{P}_{\bar{z}}$. By definition, the mean and covariance of the random variable Z_{t_i} are $\bar{z}(t_i)$ and $\sigma_i^2 = P(t_i, t_i)$ and thus if X is of the form (6.7), we have $\mu_X = \bar{z}(t^*)$ and $\sigma_X^2 = P(t^*, t^*)$. To calculate μ_X, σ_X^2 and $\mathtt{Cov}\,(Z_{t_i}, X)$ when X is of the form (6.8), we make use of [9, Section 47]:

Proposition 6.1.1. *Let X be as in (6.8) and suppose that $\{Z_t \colon t \in T\}$ is of second order with covariance P and mean zero under the measure* P*. Then*

$$(X, U)_{H_{\mathrm{P}}} = \int_T (Z_t, U)_{H_{\mathrm{P}}} f(t) d\mu(t),$$

for every $U \in H$. Furthermore,

$$\mathtt{Var}_{\mathrm{P}}\,(X) = \iint_{T \times T} P(t, s) f(t) f(s) d\mu(t) d\mu(s).$$

Recall that $\mathrm{P}_{\bar{z}}$ is the measure that endows $\{Z_t \colon t \in T\}$ with covariance P and mean $\bar{z}$ and that $\mathrm{P}_{\bar{z}}$ and P are equivalent measures. Then (2.8) tells us how to calculate μ_X and Proposition 4.1.1 tells us how to calculate σ_X^2. This is stated in the following:

Corollary 6.1.1. *Let X be as in Proposition 6.1.1. Then X is Gaussian under $\mathrm{P}_{\bar{z}}$ and*

$$\mu_X = \mathbb{E}_{\mathrm{P}_{\bar{z}}}[X] = \int_T \bar{z}(t) f(t) d\mu(t),$$

and

$$\sigma_X^2 = \mathtt{Var}_{\mathrm{P}_{\bar{z}}}(X) = \iint_{T\times T} P(t,s) f(s) f(t) d\mu(t) d\mu(s).$$

To calculate $\mathtt{Cov}(Z_{t_i}, X)$, we use Proposition 4.1.1 and (6.1.1) to see that

$$\begin{aligned} \mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(Z_{t_i}, X) &= \mathtt{Cov}_{\mathrm{P}}(Z_{t_i}, X) \\ &= \int_T (Z_{t_i}, Z_t)_{H_{\mathrm{P}}} f(t) d\mu(t) \\ &= \int_T P(t_i, t) f(t) d\mu(t) \end{aligned} \tag{6.9}$$

The conditional distribution of $Z_{t_i}|X$ is Gaussian with mean

$$\bar{z}(t_i) + \frac{\sigma_i}{\sigma_X} \rho_{i,X} (X - \mu_X),$$

and variance

$$(1 - \rho_{i,X}^2)\sigma_i^2.$$

To calculate c_k, we use U_k given in (6.5), so that

$$\begin{aligned} c_k &= \mathtt{Cov}_{\mathrm{P}_{\bar{z}}}\left(\frac{W}{\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]}, \sum_{i=1}^{r} q_{ki} Z_{t_i}\right) \\ &= \frac{1}{\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]} \sum_{i=1}^{r} q_{ki} \mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(W, Z_{t_i}) \\ &= \frac{1}{\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]} \sum_{i=1}^{r} q_{ki} \left(\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W Z_{t_i}] - \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]\, \bar{z}(t_i)\right), \end{aligned} \tag{6.10}$$

where q_{ki} is the k, i^{th} element of $\mathbf{N}^{-1}$ given in (6.3). It follows that for each i, we have

$$\begin{aligned} \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W Z_{t_i}] &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}[\mathbb{E}[W Z_{t_i}|X]] \\ &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W \mathbb{E}[Z_{t_i}|X]] \\ &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[W\left(\bar{z}(t_i) + \frac{\sigma_i}{\sigma_X}\rho_{i,X}(X - \mu_X)\right)\right] \\ &= \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]\, \bar{z}(t_i) + \frac{\sigma_i}{\sigma_X}\rho_{i,X}\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] - \mu_X \frac{\sigma_i}{\sigma_X}\rho_{i,X}\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]. \end{aligned} \tag{6.11}$$

Substituting (6.11) into (6.10),

$$c_k = \frac{1}{\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]} \sum_{i=1}^{r} q_{ki} \left(\frac{\sigma_i}{\sigma_X}\rho_{i,X}\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] - \mu_X \frac{\sigma_i}{\sigma_X}\rho_{i,X}\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]\right). \tag{6.12}$$

Therefore, we must calculate $\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW]$ and $\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]$. We now consider 3 different types of selection.

Directional Selection: $W = e^X$

Directional selection refers to the type of selection in which the mean trait of the parental subpopulation that contributes to the mean trait of the offspring population and the mean trait of the parental population differ [22]. That is, the individuals that successfully reproduce have a different mean trait than the population as a whole. This can be thought of as saying that those with a favorable extreme phenotype reproduce while those with an average

phenotype do not. As in [7], we use a fitness function of the form

$$W = e^X$$

to describe directional selection. This fitness function states that lifetime reproductive success is the logarithm of fitness. Then

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] &= \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x e^x \exp\left\{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2\right\} dx \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{-\frac{1}{2\sigma_X^2}(x^2 - 2\mu_X x + \mu_X^2 + -2\sigma_X^2 x)\right\} dx \\
&= \exp\{-\frac{\mu_X^2}{2\sigma_X^2}\}\frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{-\frac{1}{2\sigma_X^2}(x^2 - 2(\mu_X + \sigma_X^2)x)\right\} dx \\
&= \exp\{-\frac{\mu_X^2}{2\sigma_X^2}\}\frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{-\frac{1}{2\sigma_X^2}\left[(x-(\mu_X+\sigma_X^2))^2 - (\mu_X+\sigma_X^2)^2\right]\right\} dx \\
&= \exp\left\{-\frac{\mu_X^2 - (\mu_X+\sigma_X^2)^2}{2\sigma_X^2}\right\}\frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{-\frac{1}{2\sigma_X^2}\left[(x-(\mu_X+\sigma_X^2))^2\right]\right\} dx \\
&= \exp\left\{-\left(\mu_X + \frac{\sigma_X^2}{2}\right)\right\}(\mu_X + \sigma_X^2). \qquad (6.13)
\end{aligned}$$

Note that e^X is log-normal, so that $\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W] = \mathbb{E}_{\mathrm{P}_{\bar{z}}}\left[e^X\right] = \exp\left\{\mu_X + \frac{\sigma_X^2}{2}\right\}$. It follows that (6.13) becomes

$$\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] = \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W](\mu_X + \sigma_X^2). \qquad (6.14)$$

Substituting (6.14) into (6.12),

$$\begin{aligned}
c_k &= \sum_{i=1}^{r} q_{ki}\left(\frac{\sigma_i}{\sigma_X}\rho_{i,X}(\mu_X + \sigma_X^2) - \mu_X\frac{\sigma_i}{\sigma_X}\rho_{i,X}\right) \\
&= \sum_{i=1}^{r} q_{ki}\sigma_X\sigma_i\rho_{i,X} \\
&= \sum_{i=1}^{r} q_{ki}\mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(Z_{t_i}, X) \\
&= \sum_{i=1}^{r} q_{ki}\int_T P(t_i,t)f(t)d\mu(t), \qquad (6.15)
\end{aligned}$$

where the last equality follows from (6.9). It follows that

$$\begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix} = \mathbf{N}^{-1} \begin{bmatrix} \int_T P(t_1,t) f(t) d\mu(t) \\ \vdots \\ \int_T P(t_r,t) f(t) d\mu(t) \end{bmatrix}$$

Stabilizing Selection: $W = e^{-X^2/2v^2}$**.**

Stabilizing selection is a type of selection where the variability of the trait decreases and stabilizes to the population mean trait. One way to model this type of selection is to let

$$W = e^{-X^2/2v^2},$$

where $v \in \mathbb{R}$ is constant and $X \in H_{\mathrm{P}_{\tilde{z}}}$. For a fitness function of this form, we have

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\tilde{z}}}[XW] &= \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{ -\frac{1}{2\sigma_X^2}(x - mx)^2 - \frac{1}{2v^2}x^2 \right\} dx \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{ -\frac{x^2 - 2\mu_X x + \mu_X^2}{2\sigma_X^2} - \frac{1}{2v^2}x^2 \right\} dx \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{ -\frac{1}{2}\left[\frac{\sigma_X^2 + v^2}{\sigma_X^2 v^2}x^2 - 2\frac{\mu_X}{\sigma_X^2}x + \frac{\mu_X^2}{\sigma_X^2} \right] \right\} dx \\
&= \exp\left\{ -\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2} \right\} \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{\mathbb{R}} x \exp\left\{ -\frac{1}{2\sigma_{X,v}^2}\left[x^2 - \frac{2\mu_X\sigma_{X,v}^2}{\sigma_X^2}x \right] \right\} dx, \qquad (6.16)
\end{aligned}$$

where $\sigma_{X,v}^2 = \frac{\sigma_X^2 v^2}{\sigma_X^2 + v^2}$. Focusing on the exponential in (6.16), we have

$$\begin{aligned}
-\frac{1}{2\sigma_{X,v}^2}\left[x^2 - \frac{2\mu_X\sigma_{X,v}^2}{\sigma_X^2}x \right] &= -\frac{1}{2\sigma_{X,v}^2}\left[x^2 - \frac{2\mu_X\sigma_{X,v}^2}{\sigma_X^2}x + \frac{\mu_X^2\left(\sigma_{X,v}^2\right)^2}{\left(\sigma_X^2\right)^2} \right] + \frac{\mu_X^2}{2\sigma_{X,v}^2(\sigma_X^2)^2}\left(\sigma_{X,v}^2\right)^2 \\
&= -\frac{1}{2\sigma_{X,v}^2}\left[x - \frac{\mu_X\sigma_{X,v}^2}{\sigma_X^2} \right]^2 + \frac{\mu_X^2\sigma_{X,v}^2}{2(\sigma_X^2)^2} \\
&= -\frac{1}{2\sigma_{X,v}^2}\left[x - \frac{\mu_X v^2}{\sigma_X^2 + v^2} \right]^2 + \frac{\mu_X^2 v^2}{2(\sigma_X^2 + v^2)\sigma_X^2}. \qquad (6.17)
\end{aligned}$$

Substituting (6.17) into the integral in (6.16), we have

$$\begin{aligned}
&\int_{\mathbb{R}} x \exp\left\{ -\frac{1}{2\sigma_{X,v}^2}\left[x^2 - \frac{2\mu_X\sigma_{X,v}^2}{\sigma_X^2}x\right]\right\}dx \\
&\qquad = \exp\left\{\frac{\mu_X^2 v^2}{2(\sigma_X^2+v^2)\sigma_X^2}\right\}\int_{\mathbb{R}} x\exp\left\{-\frac{1}{2\sigma_{X,v}^2}\left[x-\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^2\right\}dx \\
&\qquad = \exp\left\{\frac{\mu_X^2 v^2}{2(\sigma_X^2+v^2)\sigma_X^2}\right\}\sqrt{2\pi\sigma_{X,v}^2}\frac{\mu_X v^2}{\sigma_X^2+v^2}. \qquad (6.18)
\end{aligned}$$

Substituting (6.18) into (6.16) yields

$$\begin{aligned}
&\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2}\right\}\frac{1}{\sqrt{2\pi\sigma_X^2}}\int_{\mathbb{R}} x\exp\left\{-\frac{1}{2\sigma_{X,v}^2}\left[x^2-\frac{2\mu_X\sigma_{X,v}^2}{\sigma_X^2}x\right]\right\}dx \\
&\qquad = \exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2}\right\}\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left\{\frac{\mu_X^2 v^2}{2(\sigma_X^2+v^2)\sigma_X^2}\right\}\sqrt{2\pi\sigma_{X,v}^2}\frac{\mu_X v^2}{\sigma_X^2+v^2} \\
&\qquad = \exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2}\right\}\left[\frac{\sigma_{X,v}^2}{\sigma_X^2}\right]^{1/2}\exp\left\{\frac{\mu_X^2 v^2}{2(\sigma_X^2+v^2)\sigma_X^2}\right\}\sqrt{2\pi\sigma_{X,v}^2}\frac{\mu_X v^2}{\sigma_X^2+v^2} \\
&\qquad = \left[\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^{3/2}\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2}+\frac{\mu_X^2 v^2}{2(\sigma_X^2+v^2)\sigma_X^2}\right\} \\
&\qquad = \left[\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^{3/2}\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2}\left[1-\frac{v^2}{(sx+v^2}\right]\right\} \\
&\qquad = \left[\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^{3/2}\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2+v^2}\right\}
\end{aligned} \qquad (6.19)$$

Therefore,

$$\mathbb{E}_{\mathrm{P}_{\tilde{z}}}[XW] = \left[\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^{3/2}\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2+v^2}\right\} \qquad (6.20)$$

A very similar calculation shows that

$$\mathbb{E}_{\mathrm{P}_{\tilde{z}}}[W] = \left[\frac{\mu_X v^2}{\sigma_X^2+v^2}\right]^{1/2}\exp\left\{-\frac{1}{2}\frac{\mu_X^2}{\sigma_X^2+v^2}\right\}.$$

Thus

$$\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] = \left[\frac{\mu_X v^2}{\sigma_X^2 + v^2}\right] \mathbb{E}_{\mathrm{P}_{\bar{z}}}[W]. \tag{6.21}$$

From this and (6.12), we obtain

$$\begin{aligned}
c_k &= \sum_{i=1}^{r} q_{ki} \left(\frac{\sigma_i}{\sigma_X} \rho_{i,X} \left[\frac{\mu_X v^2}{\sigma_X^2 + v^2}\right] - \mu_X \frac{\sigma_i}{\sigma_X} \rho_{i,X} \right) \\
&= \sum_{i=1}^{r} q_{ki} \mu_X \frac{\sigma_i}{\sigma_X} \rho_{i,X} \left(\frac{v^2}{\sigma_X^2 + v^2} - 1 \right) \\
&= \sum_{i=1}^{r} q_{ki} \frac{-\mu_X}{\sigma_X^2 + v^2} \sigma_i \sigma_X \rho_{i,X} \\
&= \sum_{i=1}^{r} q_{ki} \frac{-\mu_X}{\sigma_X^2 + v^2} \mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(Z_{t_i}, X) \\
&= \sum_{i=1}^{r} q_{ki} \frac{-\mu_X}{\sigma_X^2 + v^2} \int_T P(t_i, t) f(t) d\mu(t).
\end{aligned} \tag{6.22}$$

Writing this in matrix form, we have

$$\begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix} = \frac{-\mu_X}{\sigma_X^2 + v^2} \mathbf{N}^{-1} \begin{bmatrix} \int_T P(t_1, t) f(t) d\mu(t) \\ \vdots \\ \int_T P(t_r, t) f(t) d\mu(t) \end{bmatrix}.$$

Note that if $\mu_X = 0$, then $c_k = 0$ and hence $b_k = 0$ for all k. In other words, there is no force of selection acting on the trait.

Truncation Selection: $W = \mathbb{1}_A(X)$

Truncation selection [22] occurs when individuals exceeding a particular value are chosen to reproduce. One way to model this type of selection is to let

$$W = \mathbb{1}_{[\alpha,\infty)}(X) = \begin{cases} 1 & \text{if } X > \alpha \\ 0 & \text{otherwise.} \end{cases}$$

In this case, we have

$$\begin{aligned}
\mathbb{E}_{\mathrm{P}_{\bar{z}}}[XW] &= \int_{\alpha}^{\infty} x\left[2\pi\sigma_X^2\right]^{-1/2}\exp\left\{\frac{-1}{2\sigma_X^2}(x-\mu_X)^2\right\}dx \\
&= [2\pi]^{-1/2}\int_{\frac{\alpha-\mu_X}{\sigma_X}}^{\infty}(\mu_X+\sigma_X u)\exp\left\{\frac{-1}{2}u^2\right\}du \\
&= \mu_X[2\pi]^{-1/2}\int_{\frac{\alpha-\mu_X}{\sigma_X}}^{\infty}\exp\left\{\frac{-1}{2}u^2\right\}du + \sigma_X[2\pi]^{-1/2}\int_{\frac{\alpha-\mu_X}{\sigma_X}}^{\infty}u\exp\left\{\frac{-1}{2}u^2\right\}du \\
&= \mu_X\mathrm{P}_{\bar{z}}[X>\alpha] + \frac{\sigma_X}{\sqrt{2\pi}}\left[-\exp\left\{\frac{-u^2}{2}\right\}\right]_{\frac{\alpha-\mu_X}{\sigma_X}}^{\infty} \\
&= \mu_X\mathrm{P}_{\bar{z}}[X>\alpha] + \frac{\sigma_X}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\alpha-\mu_X}{\sigma_X}\right)^2\right\}. \qquad (6.23)
\end{aligned}$$

It is easy to see that $\mathbb{E}_{\mathrm{P}_{\bar{z}}}[W] = \mathrm{P}_{\bar{z}}[X>\alpha] = 1-\Phi\left(\frac{\alpha-\mu_X}{\sigma_X}\right)$, where Φ is the standard normal cumulative distribution function. It follows that (6.12) equals

$$\begin{aligned}
c_k &= \sum_{i=1}^{r} q_{ki}\frac{\sigma_i}{\sigma_X}\rho_{i,X}\frac{\sigma_X}{\sqrt{2\pi}}\frac{\exp\left\{-\frac{1}{2}\left(\frac{\alpha-\mu_X}{\sigma_X}\right)^2\right\}}{1-\Phi\left(\frac{\alpha-\mu_X}{\sigma_X}\right)} \\
&= \sum_{i=1}^{r} q_{ki}\frac{\mathtt{Cov}_{\mathrm{P}_{\bar{z}}}(Z_{t_i},X)}{\sigma_X\sqrt{2\pi}}\frac{\exp\left\{-\frac{1}{2}\left(\frac{\alpha-\mu_X}{\sigma_X}\right)^2\right\}}{1-\Phi\left(\frac{\alpha-\mu_X}{\sigma_X}\right)} \\
&= \sum_{i=1}^{r} q_{ki}\frac{\int_T P(t_i,t)f(t)d\mu(t)}{\sigma_X\sqrt{2\pi}}\frac{\exp\left\{-\frac{1}{2}\left(\frac{\alpha-\mu_X}{\sigma_X}\right)^2\right\}}{1-\Phi\left(\frac{\alpha-\mu_X}{\sigma_X}\right)}. \qquad (6.24)
\end{aligned}$$

In matrix form, we have

$$\begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix} = \frac{1}{\sqrt{2\pi\sigma_X^2}}\frac{\exp\left\{-\frac{1}{2}\left(\frac{\alpha-\mu_X}{\sigma_X}\right)^2\right\}}{1-\Phi\left(\frac{\alpha-\mu_X}{\sigma_X}\right)}\mathbf{N}^{-1}\begin{bmatrix} \int_T P(t_1,t)f(t)d\mu(t) \\ \vdots \\ \int_T P(t_r,t)f(t)d\mu(t) \end{bmatrix}.$$

6.2 Independent Sample

For simulations there are many parameters that are allowed to vary: the measure space $(T, \mathcal{T}, \mu)$, the candidate covariance function P_0, the coefficients $\mathbf{a}$ in the expansion of P, the coefficients $\boldsymbol{\mu}$ in the expansion of $\bar{z}$, the sieve parameter d, the form of X, the function f in the definition of X, the type of selection, the selection parameter and the finite grid $T_r = \{t_1, \ldots, t_r\}$. To make things simpler, we shall fix a few of these parameters. In particular, we shall let $T = [0, 1]$, $\mathcal{T}$ be the Borel σ-algebra on $[0, 1]$, μ be Lebesgue measure and we will let T_r consist of 100 equally spaced points between 0 and 1, not including 0. For the candidate covariance function P_0, we consider two different functions: the Wiener covariance, $P_0(s, t) = \min(s, t)$ and the Ornstein-Uhlenbeck covariance, $P_0(s, t) = \exp\{-|s - t|\}$. To ensure that the measures P_0 and $\mathrm{P}_{\bar{z}}$ are equivalent, we construct

$$P = P_0 + \sum_{k=1}^{100} a_k g_k \otimes g_k \text{ and}$$

$$\bar{z} = \sum_{k=1}^{100} \mu_k g_k,$$

where $\{g_k\}_{k=1}^{100}$ are computed by Gram-Schmidt orthonormalization of the set $\{P_{0t_i, t \in T_r'}\}$ and $\mathbf{a}$ and $\boldsymbol{\mu}$ are either $\{1/k\}_k$ or $\{1/k^2\}_k$. We then restrict $\bar{z}$ and P to T_r' to create a vector in $\mathbb{R}^r$ and a matrix in $\mathbb{R}^{r \times r}$, respectively. We use this vector and matrix to generate a sample of 10000 multivariate normal random variables, say $\{z_i\}_{i=1}^{10000}$. Each of these generated random variables lies on a sample path of $\{Z_t \colon t \in T\}$.

Further, we will only consider the case when X is an integral of the form (6.8). The function f in the definition of X will be assumed to be one of three forms, $f(s) = 1$, $f(s) = s$ or $f(s) = \exp\{s\}$. Letting $f(s) = 1$ makes biological sense, as X is then the average value of $\{Z_t \colon t \in T\}$ on $[0, 1]$. To construct X, we need to integrate over the entire sample path $\{Z_t \colon t \in T\}$ but we only have a finite number of points on this sample path. Therefore, we estimate X by using the trapezoidal method on the grid T_r. After this, we take

the generated random variables and perform the estimation of the desired parameters on an increasing sequence of nested subsets of $\{z_i\}_{i=1}^{10000}$ and explore the asymptotic properties of these random variables. To more accurately depict a sample with unknown parameters, the sieve parameter d is chosen to be $d(n) = \sqrt{n}$. Then $\sqrt{n} \leq 100$ for all $n = 1, \ldots, 10000$, so that we are not estimating every a_k. In other words, the only time that we use all 100 orthonormal functions $\{g_k\}$ to estimate P is when $n = 10000$.

Additionally, we look at the estimates of the evolutionary response to selection given by the Breeder's Equation (4.47). As we do not have an estimator of the genetic covariance, we shall take $G = P_0$. As a measure of closeness, we shall compute the $L^2([0,1])$ norm of the difference $\hat{\bar{z}}_n^\dagger - \bar{z}^\dagger$, where $\bar{z}^\dagger$ is the mean of $\{Z_t \colon t \in T\}$ among newborns in the following generation and $\hat{\bar{z}}_n^\dagger$ is computed via (4.47) using $\hat{\beta}_n$.

To demonstrate consistency of $\hat{\beta}_n$, we must show that (4.43) converges to 0 for all $\eta \in \mathcal{H}(P_0)$. We wish to find a way to study the consistency of $\hat{\beta}_n$ independent of η. This follows from the Cauchy-Schwarz inequality and (4.43). We have

$$\begin{aligned} |(\hat{\beta}_n, \eta) - (\beta, \eta)| &= \left| \sum_{k=1}^{\infty} (\hat{b}_{kn} - b_k)\theta_k \right| \\ &\leq \left[\sum_{k=1}^{\infty} (\hat{b}_{kn} - b_k)^2 \right]^{1/2} \left[\sum_{k=1}^{\infty} \theta_k^2 \right]^{1/2}. \end{aligned} \tag{6.25}$$

Therefore, we shall study the behavior of the square-error of $\hat{\mathbf{b}}_n$,

$$SE(\hat{\mathbf{b}}_n) = \sum_{k=1}^{\infty} (\hat{b}_{kn} - b_k)^2,$$

as n increases. We perform these iterations 100 times each and estimate the mean square error (MSE) of $\hat{\bar{z}}_n^\dagger$, denoted by $MSE(\hat{\bar{z}}_n^\dagger)$, and the mean square error of the sequence $\hat{\mathbf{b}}_n = \{\hat{b}_{kn}\}$, denoted by $MSE(\hat{\mathbf{b}}_n)$. We similarly define $SE(\hat{\mathbf{a}}_n)$ and $MSE(\hat{\mathbf{a}}_n)$, to be used in the next section. For each table below, we find the column corresponding to the chosen values of

a_k and μ_k. Next, we find the rows under the chosen function f. The intersection of rows and columns is the MSE of the indicated estimator when $n = 1000, 5000, 10000$. For example, in Table 6.1, for $a_k = 1/k^2$, $\mu_k = 1/k$ and $f(s) = e^s$, then $MSE(\hat{\mathbf{b}}_n)$ is 0.5744 for $n = 1000$, 8.9466×10^{-2} for $n = 5000$ and 2.5866×10^{-2} for $n = 10000$.

The rest of this section is dedicated to the results of such simulations. Within each subsection, we consider one type of selection: directional, stabilizing or truncation. For each type of selection, we begin with the results of $MSE(\hat{\mathbf{b}}_n)$ followed by those of $MSE(\hat{\bar{z}}_n^{\dagger})$ with P_0 being the Wiener covariance function. In the case of stabilizing and truncation selection, we also have a "selection parameter" for which we select two values. Considering the same tpe of selection, we then shift to the case when P_0 is the Ornstein-Uhlenbeck covariance and repeat the same format. There is a discussion involving various comparisons between the different parameter choices following the each of the results of a simulation.

The simulations conducted seem to indicate convergence of both $MSE(\hat{\mathbf{b}}_n)$ and $MSE(\hat{\bar{z}}_n^{\dagger})$ to 0 for all of the different combinations of model parameters. While, some values of the MSE appear to be converging faster than others, these results do seem to support the conclusions about the selection gradient given in Chapter 4. Even though we do not have any theoretical results about the evolutionary response to selection, these results also indicate that the estimates $\hat{\bar{z}}_n^{\dagger}$ are good estimators of $\bar{z}^{\dagger}$ and are converging to $\bar{z}^{\dagger}$.

6.2.1 Directional Selection

We shall first consider a fitness function of the form

$$W(X) = \exp(X),$$

and $P_0(s,t) = \min(s,t)$ and we let $\mathbf{a}$, $\boldsymbol{\mu}$ and f vary.

Wiener Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	0.1225	0.1224	0.1219	0.1217
5000	1.4809×10^{-2}	1.4803×10^{-2}	1.4694×10^{-2}	1.4724×10^{-2}
10000	4.2758×10^{-3}	4.1648×10^{-3}	4.0922×10^{-3}	4.1014×10^{-3}
$f(s) = s$				
$n = 1000$	6.5048×10^{-2}	6.4903×10^{-2}	6.4883×10^{-2}	6.5282×10^{-2}
5000	9.1233×10^{-3}	9.1081×10^{-3}	9.1097×10^{-3}	9.0574×10^{-3}
10000	1.4707×10^{-3}	1.4393×10^{-3}	1.4689×10^{-3}	1.4629×10^{-3}
$f(s) = e^s$				
$n = 1000$	0.5720	0.5741	0.5744	0.5746
5000	9.1537×10^{-2}	9.0228×10^{-2}	8.9466×10^{-2}	8.7613×10^{-2}
10000	2.7612×10^{-2}	2.6674×10^{-2}	2.5866×10^{-2}	2.5828×10^{-2}

Table 6.1: $MSE(\hat{\mathbf{b}}_n)$ for Directional Selection with $P_0(s,t) = \min(s,t)$

Based on the results in Table 6.1, it appears that convergence is independent of our choice of f, $\mathbf{a}$ and $\boldsymbol{\mu}$; convergence occurs at approximately the same rate. Also, the values of $MSE(\hat{\mathbf{b}}_n)$ is roughly equal for all choices of $\mathbf{a}$ and $\boldsymbol{\mu}$. On the other hand, the values of the MSE are dependent on the choice of the function f. The MSE when $f(s) = 1$ is approximately twice as large as when $f(s) = s$ and the MSE when $f(s) = e^s$ is around 4.5 times as large as when $f(s) = 1$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	2.3122×10^{-2}	1.9547×10^{-2}	2.3448×10^{-2}	2.1356×10^{-2}
5000	4.4497×10^{-4}	2.8617×10^{-4}	5.6439×10^{-5}	3.3820×10^{-4}
10000	5.6973×10^{-5}	6.1121×10^{-5}	1.1575×10^{-5}	4.6812×10^{-5}
$f(s) = s$				
$n = 1000$	1.4218×10^{-2}	1.1484×10^{-2}	1.4898×10^{-2}	1.3223×10^{-2}
5000	4.1053×10^{-3}	2.9830×10^{-4}	4.3009×10^{-4}	4.5719×10^{-4}
10000	2.0427×10^{-5}	3.4002×10^{-5}	3.1159×10^{-5}	6.5659×10^{-5}
$f(s) = e^s$				
$n = 1000$	0.1047	9.7056×10^{-2}	0.1082	0.1053
5000	2.7160×10^{-3}	2.3479×10^{-3}	3.3986×10^{-3}	3.7008×10^{-3}
10000	3.7224×10^{-4}	2.6767×10^{-4}	3.0223×10^{-4}	5.1505×10^{-4}

Table 6.2: $MSE(\hat{\bar{z}}_n^{\dagger})$ for Directional Selection with $P_0(s,t) = \min(s,t)$

From Table 6.2, we see that as with $MSE(\hat{\mathbf{b}}_n)$, the convergence of $MSE(\hat{\bar{z}}_n^{\dagger})$ appears to be independent of the choice of $\mathbf{a}$ and $\boldsymbol{\mu}$. We see that when $f(s) = 1$, $MSE(\hat{\bar{z}}_n^{\dagger})$ is anywhere from 0.4 to 2.8 times as large as when $f(s) = s$. If $f(s) = e^s$, we see that $MSE(\hat{\bar{z}}_n^{\dagger})$ is the larger than the other choices of f by around 400-800%.

Ornstein-Uhlenbeck Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	0.1918	0.1924	0.1944	0.1966
5000	4.3958×10^{-2}	4.3160×10^{-2}	4.2745×10^{-2}	4.3139×10^{-2}
10000	2.0971×10^{-2}	2.0801×10^{-2}	2.0612×10^{-2}	2.0794×10^{-2}
$f(s) = s$				
$n = 1000$	8.0974×10^{-2}	8.1053×10^{-2}	8.1225×10^{-2}	8.1230×10^{-2}
5000	1.5437×10^{-2}	1.5534×10^{-2}	1.5576×10^{-2}	1.5400×10^{-2}
10000	3.0610×10^{-3}	3.0099×10^{-3}	2.9980×10^{-3}	3.0024×10^{-3}
$f(s) = e^s$				
$n = 1000$	1.1432	1.1266	1.0735	1.1247
5000	0.3686	0.3586	0.3378	0.4499
10000	0.2087	0.1971	0.1844	0.2398

Table 6.3: $MSE(\hat{\mathbf{b}}_n)$ for Directional Selection with $P_0(s,t) = \exp\{-|s-t|\}$

If $P_0(s,t) = \exp(-|s-t|)$, Table 6.3 shows that many of the same things that were said about the results of Table 6.1 are valid. The main difference is that the values of $MSE(\hat{\mathbf{b}}_n)$ are 25-100% larger in Table 6.3. Further, the rate of convergence appears to be slightly lower when $P_0(s,t) = \exp(-|s-t|)$ versus when $P_0(s,t) = \min(s,t)$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.8975×10^{-2}	3.5582×10^{-2}	4.1519×10^{-2}	3.6310×10^{-2}
5000	1.6546×10^{-3}	1.9009×10^{-3}	2.2580×10^{-3}	1.5684×10^{-3}
10000	3.8197×10^{-4}	6.1750×10^{-4}	5.0943×10^{-4}	3.6211×10^{-4}
$f(s) = s$				
$n = 1000$	1.9274×10^{-2}	1.6576×10^{-2}	1.8842×10^{-2}	1.8101×10^{-2}
5000	8.4371×10^{-4}	5.9262×10^{-4}	1.0137×10^{-3}	8.6174×10^{-4}
10000	5.1704×10^{-5}	5.4553×10^{-5}	2.7251×10^{-4}	8.8599×10^{-5}
$f(s) = e^s$				
$n = 1000$	0.1673	0.1537	0.1602	0.1598
5000	1.1575×10^{-2}	1.1102×10^{-2}	1.0958×10^{-2}	1.5888×10^{-2}
10000	4.5098×10^{-3}	3.6610×10^{-3}	3.1488×10^{-3}	5.6310×10^{-2}

Table 6.4: $MSE(\hat{\bar{z}}_n^{\dagger})$ for Directional Selection with $P_0(s,t) = \exp\{-|s-t|\}$

Table 6.4 shows that the results in Table 6.2 hold as well. As with $MSE(\hat{\mathbf{b}}_n)$, the values of $MSE(\hat{\bar{z}}_n^{\dagger})$ in Table 6.4 are approximately 25-50% larger than those in Table 6.2.

Under directional selection, it appears that the convergence of $MSE(\hat{\mathbf{b}}_n)$ and $MSE(\hat{\bar{z}}_n^{\dagger})$ is independent of both $\mathbf{a}$ and $\boldsymbol{mu}$ and that the choice of covariance function only effects the magnitude of the MSEs. The most noticeable difference in these tables is that function f has the largest effect on the magnitude of the MSEs, but still does not affect the rate of convergence.

6.2.2 Stabilizing Selection

We now consider the a fitness function of the form

$$W(X) = \exp\left(-\frac{X^2}{2v^2}\right),$$

where v is either equal to 1 or 5 and we let the other parameters vary.

Wiener Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	2.6926×10^{-4}	6.1251×10^{-6}	2.6972×10^{-4}	1.2892×10^{-2}
5000	3.2361×10^{-5}	2.0086×10^{-6}	3.2041×10^{-5}	1.7864×10^{-3}
10000	8.6775×10^{-6}	2.8452×10^{-7}	8.4210×10^{-6}	6.4975×10^{-4}
$f(s) = s$				
$n = 1000$	5.5046×10^{-5}	6.1992×10^{-6}	5.5143×10^{-5}	6.1390×10^{-6}
5000	7.8021×10^{-6}	1.0070×10^{-6}	7.7751×10^{-6}	9.9563×10^{-7}
10000	1.3543×10^{-6}	2.8375×10^{-7}	1.2984×10^{-6}	2.7012×10^{-7}
$f(s) = e^s$				
$n = 1000$	4.0015×10^{-3}	4.8765×10^{-4}	4.0359×10^{-3}	4.8754×10^{-4}
5000	5.4249×10^{-4}	7.5724×10^{-5}	5.4328×10^{-4}	7.7099×10^{-5}
10000	1.0429×10^{-4}	2.2908×10^{-5}	1.0199×10^{-4}	2.3178×10^{-5}

Table 6.5: $MSE(\hat{\mathbf{b}}_n)$ for Stabilizing Selection with $P_0(s,t) = \min(s,t)$ and $v = 5$

Table 6.5 shows that the convergence of $MSE(\hat{\mathbf{b}}_n)$ is occurring at about the same rate for all choices of f, $\mathbf{a}$, and $\boldsymbol{\mu}$. With this model, it appears that the choice of $\mathbf{a}$ and $\boldsymbol{\mu}$ has a clear impact on the magnitude of $MSE(\hat{\mathbf{b}}_n)$. If $\mu_k = 1/k$, the MSEs are approximately equal when $a_k = 1/k$ and $a_k = 1/k^2$. If $\mu_k = 1/k^2$ and $a_k = 1/k$, the MSE is around 1000-4500% smaller than the MSE when $\mu_k = 1/k$. When $a_k = 1/k^2$ and $\mu_k = 1/k^2$, $MSE(\hat{\mathbf{b}}_n)$ for $f(s) = 1$ is approximately 4700% larger than when $\mu_k = 1/k$. On the other hand, when $f(s) = s$ and $f(s) = \exp(s)$, it is approximately 1000% smaller than when $\mu_k = 1/k$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	1.9912×10^{-3}	7.5292×10^{-5}	2.2322×10^{-3}	1.0794×10^{-3}
5000	1.5326×10^{-4}	4.1739×10^{-5}	2.9297×10^{-4}	1.3768×10^{-4}
10000	3.6563×10^{-5}	3.8729×10^{-5}	1.5651×10^{-4}	2.0290×10^{-4}
$f(s) = s$				
$n = 1000$	2.0113×10^{-3}	7.0968×10^{-5}	1.9998×10^{-3}	3.5721×10^{-5}
5000	1.1220×10^{-4}	2.8817×10^{-5}	1.8415×10^{-4}	6.9237×10^{-6}
10000	9.5988×10^{-6}	2.6467×10^{-5}	5.7790×10^{-5}	6.1700×10^{-6}
$f(s) = e^s$				
$n = 1000$	2.9697×10^{-3}	7.7609×10^{-5}	2.6531×10^{-3}	2.3635×10^{-4}
5000	2.2225×10^{-4}	9.2786×10^{-6}	1.6662×10^{-4}	4.3401×10^{-5}
10000	7.9374×10^{-5}	1.4852×10^{-5}	3.5408×10^{-5}	3.3476×10^{-5}

Table 6.6: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Stabilizing Selection with $P_0(s,t) = \min(s,t)$ and $v = 5$

When viewing the results in Table 6.6, we see that there are a couple of times when $MSE(\hat{\tilde{z}}_n^{\dagger})$ actually increases when the sample size is increased. This could indicate that for certain combinations of f, $\mathbf{a}$ and $\boldsymbol{\mu}$, $MSE(\hat{\tilde{z}}_n^{\dagger})$ does not converge to zero. Another item of note is that the rates of convergence depend on the choice of $\mathbf{a}$ and $\boldsymbol{\mu}$. When $a_k = 1/k$ and $\mu_k = 1/k^2$, $MSE(\hat{\tilde{z}}_n^{\dagger})$ is decreasing, but at a much slower rate than for other choices of $\mathbf{a}$ and $\boldsymbol{\mu}$. The magnitude differences are approximately the same as those in Table 6.5.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	9.6532×10^{-2}	1.2951×10^{-2}	9.9619×10^{-2}	1.2913×10^{-2}
5000	1.1324×10^{-2}	1.8054×10^{-3}	1.1123×10^{-2}	1.7371×10^{-3}
10000	2.9723×10^{-3}	6.6124×10^{-4}	2.9124×10^{-3}	6.2034×10^{-4}
$f(s) = s$				
$n = 1000$	2.6370×10^{-2}	2.9333×10^{-3}	2.6791×10^{-2}	2.9691×10^{-3}
5000	3.7018×10^{-3}	4.8208×10^{-4}	3.7504×10^{-3}	4.8430×10^{-4}
10000	5.8838×10^{-4}	1.2707×10^{-4}	5.9548×10^{-4}	1.3083×10^{-4}
$f(s) = e^s$				
$n = 1000$	0.5094	6.1399×10^{-2}	0.5422	6.4885×10^{-2}
5000	7.2036×10^{-2}	1.0065×10^{-2}	7.6066×10^{-2}	1.0425×10^{-2}
10000	1.6478×10^{-2}	3.1741×10^{-3}	1.6804×10^{-2}	3.2860×10^{-3}

Table 6.7: $MSE(\hat{\mathbf{b}}_n)$ for Stabilizing Selection with $P_0(s,t) = \min(s,t)$ and $v = 1$

When we let $v = 1$, Table 6.7 shows that $MSE(\hat{\mathbf{b}}_n)$ is converging to 0 at approximately the same rate as the values in Table 6.5. The only possible exception to this is that when $f(s) = \exp(s)$, the values of $MSE(\hat{\mathbf{b}}_n)$ maybe converging slightly slower than the other two choices of f. We also note that when $\mu_k = 1/k^2$, the values of the MSE are about 4-8 times smaller than when $\mu_k = 1/k$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	1.6858×10^{-2}	2.6747×10^{-3}	1.7513×10^{-2}	1.9950×10^{-3}
5000	3.1766×10^{-4}	1.6736×10^{-4}	4.3057×10^{-4}	2.8872×10^{-5}
10000	2.6647×10^{-5}	7.4162×10^{-5}	2.1553×10^{-5}	1.5941×10^{-5}
$f(s) = s$				
$n = 1000$	7.3363×10^{-3}	3.0316×10^{-4}	6.2345×10^{-3}	7.6329×10^{-4}
5000	3.8368×10^{-4}	7.3470×10^{-5}	1.7910×10^{-4}	6.6090×10^{-5}
10000	8.4281×10^{-5}	8.0649×10^{-5}	1.1306×10^{-5}	3.4159×10^{-5}
$f(s) = e^s$				
$n = 1000$	9.4061×10^{-2}	1.0939×10^{-2}	9.2111×10^{-2}	1.1517×10^{-2}
5000	2.3025×10^{-3}	3.5984×10^{-4}	1.9072×10^{-3}	4.3636×10^{-4}
10000	1.0142×10^{-4}	4.8831×10^{-5}	1.1982×10^{-4}	6.5969×10^{-5}

Table 6.8: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Stabilizing Selection with $P_0(s,t) = \min(s,t)$ and $v = 1$

Similarly to when $v = 5$, there is one case ($\mu_k = 1/k^2, a_k = 1/k, f(s) = s$) when $MSE(\hat{\tilde{z}}_n^{\dagger})$ increases with increases in sample size. Other than that case, convergence appears to be occurring at the same rate as the values of $MSE(\hat{\tilde{z}}_n^{\dagger})$ in Table 6.6. We also note that different choices of μ_k lead to differing magnitudes of $MSE(\hat{\tilde{z}}_n^{\dagger})$. For example, when $\mu_k = 1/k^2$, $MSE(\hat{\tilde{z}}_n^{\dagger})$ is approximately 2-9 times smaller than when $\mu_k = 1/k$.

When comparing Tables 6.5 and 6.7, the most noticeable difference is in the magnitude of $MSE(\hat{\mathbf{b}}_n)$. On average the magnitude of $MSE(\hat{\mathbf{b}}_n)$ is much (100-10000 times) smaller when $v = 5$ than when $v = 1$. The rates of convergence in these tables is approximately equal throughout. Similar observations can be made when comparing Tables 6.6 and 6.8.

Ornstein-Uhlenbeck Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	7.7989×10^{-4}	2.4904×10^{-4}	7.9518×10^{-4}	2.5272×10^{-4}
5000	1.4689×10^{-4}	5.2332×10^{-5}	1.4793×10^{-4}	5.3271×10^{-5}
10000	5.8885×10^{-5}	2.3580×10^{-5}	5.9510×10^{-5}	2.3830×10^{-5}
$f(s) = s$				
$n = 1000$	1.0312×10^{-4}	2.4528×10^{-5}	1.0416×10^{-4}	2.4625×10^{-5}
5000	1.9876×10^{-5}	5.0389×10^{-6}	1.9850×10^{-5}	5.0373×10^{-6}
10000	3.8721×10^{-6}	1.311×10^{-6}	3.9243×10^{-6}	1.3297×10^{-6}
$f(s) = e^s$				
$n = 1000$	7.9782×10^{-3}	2.2203×10^{-3}	2.6301×10^{-3}	2.2603×10^{-3}
5000	1.5157×10^{-3}	4.6230×10^{-4}	1.5395×10^{-3}	4.6775×10^{-4}
10000	3.9151×10^{-4}	1.4936×10^{-4}	3.9267×10^{-4}	1.5184×10^{-4}

Table 6.9: $MSE(\hat{\mathbf{b}}_n)$ for Stabilizing Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $v = 1$

The rates of convergence in Table 6.9 are all approximately equal. The magnitude of $MSE(\hat{\mathbf{b}}_n)$ when $\mu_k = 1/k^2$ is slightly smaller than when $\mu_k = 1/k$. Also, $MSE(\hat{\mathbf{b}}_n)$ is increasing over $f(s) = s$, $f(s) = 1$ and $f(s) = \exp(s)$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.2329×10^{-3}	1.2702×10^{-3}	2.6301×10^{-3}	3.1982×10^{-4}
5000	2.9824×10^{-4}	9.7567×10^{-4}	3.7877×10^{-4}	2.7316×10^{-4}
10000	1.0294×10^{-4}	9.4816×10^{-4}	1.9719×10^{-4}	2.7551×10^{-4}
$f(s) = s$				
$n = 1000$	2.6958×10^{-3}	9.5261×10^{-5}	2.5726×10^{-3}	1.0516×10^{-4}
5000	1.7376×10^{-4}	9.6448×10^{-5}	4.2034×10^{-4}	4.7385×10^{-5}
10000	1.6838×10^{-5}	9.4340×10^{-5}	2.3512×10^{-4}	4.4916×10^{-5}
$f(s) = e^s$				
$n = 1000$	3.8553×10^{-3}	5.2854×10^{-4}	3.4065×10^{-3}	3.4670×10^{-4}
5000	3.2185×10^{-4}	4.9947×10^{-5}	1.9632×10^{-4}	3.6792×10^{-5}
10000	7.1483×10^{-5}	1.6230×10^{-5}	2.1954×10^{-5}	2.4019×10^{-5}

Table 6.10: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Stabilizing Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $v = 5$

From Table 6.10, we see that $MSE(\hat{\tilde{z}}_n^{\dagger})$ converges more slowly when $\mu_k = 1/k^2$ and actually increases slightly when $a_k = 1/k^2$ and $f(s) = 1$ and also when $a_k = 1/k$ and $f(s) = s$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	0.1837	5.8169×10^{-2}	0.1844	5.7935×10^{-2}
5000	5.6188×10^{-2}	1.8093×10^{-2}	5.1477×10^{-2}	1.7730×10^{-2}
10000	3.8606×10^{-2}	1.2643×10^{-2}	3.6561×10^{-2}	1.2400×10^{-2}
$f(s) = s$				
$n = 1000$	4.0263×10^{-2}	9.4875×10^{-3}	4.0880×10^{-2}	9.6945×10^{-3}
5000	7.6377×10^{-3}	1.9418×10^{-3}	7.7958×10^{-3}	1.9684×10^{-3}
10000	1.3820×10^{-3}	4.8332×10^{-4}	1.4165×10^{-3}	4.7585×10^{-4}
$f(s) = e^s$				
$n = 1000$	0.3783	0.1051	0.3935	0.1102
5000	8.3034×10^{-2}	2.3831×10^{-2}	8.5062×10^{-2}	2.4411×10^{-2}
10000	2.8134×10^{-2}	9.3972×10^{-3}	2.9506×10^{-2}	9.3865×10^{-3}

Table 6.11: $MSE(\hat{\mathbf{b}}_n)$ for Stabilizing Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $v = 1$

As with the case $P_0(s,t) = \min(s,t)$, Table 6.11 shows that the values of $MSE(\hat{\mathbf{b}}_n)$ increase substantially when changing from $v = 5$ to $v = 1$. Other than this observation, the comments from Table 6.9 hold here as well.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	7.8630×10^{-2}	2.1735×10^{-2}	7.5700×10^{-2}	2.3137×10^{-2}
5000	2.9203×10^{-2}	7.5384×10^{-3}	2.5564×10^{-2}	8.4615×10^{-3}
10000	2.4974×10^{-2}	6.4458×10^{-3}	2.1686×10^{-2}	7.3903×10^{-3}
$f(s) = s$				
$n = 1000$	1.2039×10^{-2}	2.0245×10^{-3}	8.5953×10^{-3}	1.7694×10^{-3}
5000	9.6891×10^{-4}	1.4790×10^{-4}	3.5711×10^{-4}	5.1445×10^{-4}
10000	2.8360×10^{-4}	3.0930×10^{-5}	1.2745×10^{-4}	4.6307×10^{-4}
$f(s) = e^s$				
$n = 1000$	6.3787×10^{-2}	2.0377×10^{-2}	7.7600×10^{-2}	2.0696×10^{-2}
5000	2.4847×10^{-3}	1.1559×10^{-3}	3.5251×10^{-3}	1.2399×10^{-3}
10000	3.2815×10^{-4}	2.1095×10^{-4}	3.6606×10^{-4}	3.0738×10^{-4}

Table 6.12: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Stabilizing Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $v = 1$

The results in Table 6.12 have the same interpretation as those in Table 6.10, apart from the values of $MSE(\hat{\tilde{z}}_n^{\dagger})$ being roughly 20 times larger.

When comparing Tables 6.5-6.12, the main thing that stands out is that for $P_0(s,t) = \exp(-|s-t|)$, the respective values of MSE are larger than those for $P_0(s,t) = \min(s,t)$. Another item of importance is the convergence rates, which do not appear to follow any pattern. It is difficult to determine if the convergence can be made faster by choosing a specific parameter value over another.

6.2.3 Truncation Selection

We now consider a fitness function of the form $\mathbb{1}_{[\alpha,\infty)}(X)$.

Wiener Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.4860×10^{-3}	1.8542×10^{-2}	1.4761×10^{-3}	6.1725×10^{-2}
5000	2.2651×10^{-3}	4.7443×10^{-3}	3.9920×10^{-4}	8.6463×10^{-3}
10000	2.0396×10^{-3}	2.5037×10^{-3}	2.4128×10^{-4}	3.0788×10^{-3}
$f(s) = s$				
$n = 1000$	3.5951×10^{-3}	4.1495×10^{-2}	2.2952×10^{-3}	9.7812×10^{-2}
5000	9.3790×10^{-4}	8.8718×10^{-3}	5.4385×10^{-4}	1.5052×10^{-2}
10000	4.2381×10^{-4}	2.5032×10^{-4}	2.5972×10^{-4}	3.5369×10^{-3}
$f(s) = e^s$				
$n = 1000$	2.1951×10^{-3}	2.8597×10^{-2}	1.8830×10^{-3}	8.0129×10^{-2}
5000	5.0380×10^{-4}	6.3479×10^{-3}	4.6295×10^{-4}	1.2122×10^{-2}
10000	2.6579×10^{-4}	2.3093×10^{-3}	2.4135×10^{-4}	3.2573×10^{-3}

Table 6.13: $MSE(\hat{\mathbf{b}}_n)$ for Truncation Selection with $P_0(s,t) = \min(s,t)$ and $\alpha = 0$

In contrast to stabilizing selection, Table 6.13 shows that the values of $MSE(\hat{\mathbf{b}}_n)$ are, on average, larger for $\mu_k = 1/k^2$ than for $\mu_k = 1/k$. The convergence rates are roughly all equal and are not as large as for other types of selection.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	2.7379×10^{-3}	2.6656×10^{-3}	2.5133×10^{-3}	1.1134×10^{-2}
5000	6.0192×10^{-4}	1.6278×10^{-4}	1.6677×10^{-4}	3.2086×10^{-4}
10000	5.2763×10^{-4}	1.1122×10^{-4}	3.7204×10^{-5}	5.6239×10^{-5}
$f(s) = s$				
$n = 1000$	3.3987×10^{-3}	7.5794×10^{-3}	2.4680×10^{-3}	2.0304×10^{-2}
5000	2.6435×10^{-4}	4.4267×10^{-4}	1.4677×10^{-4}	9.4315×10^{-4}
10000	2.1124×10^{-5}	1.2995×10^{-4}	2.7558×10^{-5}	1.9441×10^{-4}
$f(s) = e^s$				
$n = 1000$	2.3876×10^{-3}	5.2262×10^{-3}	2.5250×10^{-3}	1.3109×10^{-2}
5000	1.2304×10^{-4}	2.2798×10^{-4}	1.4007×10^{-4}	2.7198×10^{-2}
10000	1.9533×10^{-5}	3.7217×10^{-5}	1.6749×10^{-5}	4.0268×10^{-5}

Table 6.14: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Truncation Selection with $P_0(s,t) = \min(s,t)$ and $\alpha = 0$

For all choices of $\mathbf{a}$, $\boldsymbol{\mu}$ and f, Table 6.14 shows that the convergence rates are all approximately equal. Similar to Table 6.13, the values of $MSE(\hat{\tilde{z}}_n^{\dagger})$ tend to be slightly larger when $\mu_k = 1/k^2$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	0.1324	0.9071	0.1339	0.9131
5000	1.7481×10^{-2}	0.1407	1.7584×10^{-2}	5.2871×10^{-2}
10000	6.1061×10^{-3}	5.9537×10^{-2}	5.8713×10^{-3}	5.2871×10^{-2}
$f(s) = s$				
$n = 1000$	0.7966	3.6887	0.8528	4.1904
5000	0.1239	0.8708	0.1339	1.0103
10000	3.0098×10^{-2}	0.4331	3.2539×10^{-2}	0.5035
$f(s) = e^s$				
$n = 1000$	2.8240×10^{-2}	0.3543	2.7573×10^{-2}	0.3790
5000	4.6683×10^{-3}	5.2631×10^{-2}	4.5403×10^{-3}	5.5131×10^{-2}
10000	1.5331×10^{-3}	1.3149×10^{-2}	1.4881×10^{-3}	1.3984×10^{-2}

Table 6.15: $MSE(\hat{\mathbf{b}}_n)$ for Truncation Selection with $P_0(s,t) = \min(s,t)$ and $\alpha = 1$

Table 6.15 shows some of the most extreme results yet. First, we note that the rates of convergence appear to follow no distinct pattern. For each function f, the rates of convergence are approximately the same, whereas for different functions, the rates appear to follow no concrete pattern. We also note that much like in Table 6.13, the values of $MSE(\hat{\mathbf{b}}_n)$ are larger when $\mu_k = 1/k^2$ and, in this case, are relatively large (3.6887 and 4.1904) when compared to the case when the truncation parameter, α, is 0. Furthermore, we note that $f(s) = \exp(s)$, has the smallest values of $MSE(\hat{\mathbf{b}}_n)$, followed by $f(s) = 1$ and $f(s) = s$ with the largest values. This is almost the exact opposite of what we saw with other types of selection.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.1501×10^{-2}	0.1798	2.6295×10^{-2}	0.1627
5000	1.5859×10^{-3}	1.2178×10^{-2}	6.9960×10^{-4}	5.4292×10^{-3}
10000	5.7249×10^{-4}	5.7510×10^{-3}	5.7818×10^{-5}	1.2978×10^{-3}
$f(s) = s$				
$n = 1000$	0.1387	0.4673	0.1554	0.5223
5000	3.0725×10^{-3}	1.0051×10^{-2}	3.8634×10^{-3}	1.3709×10^{-2}
10000	1.0087×10^{-4}	5.7800×10^{-4}	1.1378×10^{-4}	8.4938×10^{-4}
$f(s) = e^s$				
$n = 1000$	7.2263×10^{-3}	5.8776×10^{-2}	5.9711×10^{-3}	6.5795×10^{-2}
5000	3.1895×10^{-4}	1.0443×10^{-3}	2.4989×10^{-4}	1.4834×10^{-3}
10000	4.5053×10^{-5}	1.0832×10^{-4}	1.0516×10^{-4}	6.5617×10^{-5}

Table 6.16: $MSE(\hat{\bar{z}}_n^{\dagger})$ for Truncation Selection with $P_0(s,t) = \min(s,t)$ and $\alpha = 1$

When switching from $\alpha = 0$ to $\alpha = 1$, we see from Table 6.16 that $MSE(\hat{\bar{z}}_n^{\dagger})$ follows the same kind of patterns as in Table 6.14. The only real difference is the size of the MSE. In Table 6.14, the values of $MSE(\hat{\bar{z}}_n^{\dagger})$ are all roughly the same, while in Table 6.16 the values vary by factors of 10 based on the selected parameters.

Ornstein-Uhlenbeck Covariance

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.1254×10^{-2}	0.1845	3.1146×10^{-2}	0.1834
5000	6.8956×10^{-3}	5.7709×10^{-2}	6.9980×10^{-3}	5.6017×10^{-2}
10000	3.4218×10^{-3}	4.0431×10^{-2}	3.2796×10^{-3}	3.8443×10^{-2}
$f(s) = s$				
$n = 1000$	0.2609	0.8609	0.2678	0.8981
5000	5.3077×10^{-2}	0.1915	5.4082×10^{-2}	0.1988
10000	1.3070×10^{-2}	5.9624×10^{-2}	1.3076×10^{-2}	6.3613×10^{-2}
$f(s) = e^s$				
$n = 1000$	1.1955×10^{-2}	8.6821×10^{-2}	1.5315×10^{-2}	8.8292×10^{-2}
5000	3.5330×10^{-3}	1.8100×10^{-2}	3.5472×10^{-3}	1.8562×10^{-2}
10000	1.4152×10^{-3}	6.0537×10^{-3}	1.4381×10^{-3}	6.2172×10^{-3}

Table 6.17: $MSE(\hat{\mathbf{b}}_n)$ for Truncation Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $\alpha = 0$

As was the case with $P_0(s,t) = \min(s,t)$, we see that from Table 6.17, $\mu_k = 1/k^2$ produces larger values of $MSE(\hat{\mathbf{b}}_n)$. We also notice that the MSE is the smallest with $f(s) = \exp(s)$ and largest with $f(s) = s$. The convergence rates appear to all be roughly the same; the only real difference between different selections of parameters is the magnitude of $MSE(\hat{\mathbf{b}}_n)$.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	9.8474×10^{-3}	8.2243×10^{-2}	7.8141×10^{-3}	7.2711×10^{-2}
5000	7.2361×10^{-4}	3.1401×10^{-2}	4.0759×10^{-4}	2.6010×10^{-2}
10000	1.7541×10^{-4}	2.7255×10^{-2}	9.6947×10^{-5}	2.2434×10^{-2}
$f(s) = s$				
$n = 1000$	5.3927×10^{-2}	0.1526	5.7281×10^{-2}	0.1570
5000	2.4308×10^{-3}	6.2806×10^{-3}	2.3480×10^{-3}	5.5702×10^{-3}
10000	3.4662×10^{-4}	3.4864×10^{-4}	1.3628×10^{-4}	5.0073×10^{-4}
$f(s) = e^s$				
$n = 1000$	5.9323×10^{-3}	1.4623×10^{-2}	6.5083×10^{-3}	1.8964×10^{-2}
5000	3.7535×10^{-4}	6.7774×10^{-4}	6.2080×10^{-4}	1.0356×10^{-3}
10000	9.8372×10^{-5}	2.0751×10^{-4}	2.1836×10^{-4}	2.1915×10^{-4}

Table 6.18: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Truncation Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $\alpha = 0$

The analysis of Table 6.18 shows that $\mu_k = 1/k^2$ produces larger values of $MSE(\hat{\tilde{z}}_n^{\dagger})$. It appears that the convergence for $\mu_k = 1/k^2$ and $f(s) = 1$ is much slower than the convergence for other parameters.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	3.0999×10^{-2}	0.1851	3.1855×10^{-2}	0.1865
5000	6.7268×10^{-3}	5.7231×10^{-2}	7.1184×10^{-3}	5.7585×10^{-2}
10000	3.2156×10^{-3}	4.0283×10^{-2}	3.3876×10^{-3}	3.9180×10^{-2}
$f(s) = s$				
$n = 1000$	0.2627	0.8465	0.2679	0.8941
5000	5.2670×10^{-2}	0.1886	5.4464×10^{-2}	0.1946
10000	1.2689×10^{-2}	6.0060×10^{-2}	1.3537×10^{-2}	6.2579×10^{-2}
$f(s) = e^s$				
$n = 1000$	1.5039×10^{-2}	8.7618×10^{-2}	1.5340×10^{-2}	8.8507×10^{-2}
5000	3.5627×10^{-3}	1.8121×10^{-2}	3.6606×10^{-3}	1.8478×10^{-2}
10000	1.4612×10^{-3}	6.0057×10^{-3}	1.5189×10^{-3}	6.0924×10^{-3}

Table 6.19: $MSE(\hat{\mathbf{b}}_n)$ for Truncation Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $\alpha = 1$

We see immediately that Table 6.19 looks very similar to Table 6.17. This may be due to a similar number of observations have an X value that exceed 1 and 0, respectively. The interpretation of Table 6.19 is identical to that of Table 6.17.

	$a_k = 1/k$		$a_k = 1/k^2$	
	$\mu_k = 1/k$	$\mu_k = 1/k^2$	$\mu_k = 1/k$	$\mu_k = 1/k^2$
$f(s) = 1$				
$n = 1000$	1.0993×10^{-2}	6.7500×10^{-2}	9.2418×10^{-3}	6.1490×10^{-2}
5000	1.2081×10^{-3}	2.2719×10^{-2}	5.7729×10^{-4}	1.8682×10^{-2}
10000	6.0880×10^{-4}	1.9591×10^{-2}	1.3386×10^{-4}	1.5071×10^{-2}
$f(s) = s$				
$n = 1000$	5.5430×10^{-2}	0.1594	5.5920×10^{-2}	0.1670
5000	2.2897×10^{-3}	6.6654×10^{-3}	2.1104×10^{-3}	7.8236×10^{-3}
10000	1.4397×10^{-4}	2.9305×10^{-4}	1.4376×10^{-4}	6.8808×10^{-4}
$f(s) = e^s$				
$n = 1000$	5.8113×10^{-3}	1.5266×10^{-2}	4.9146×10^{-3}	1.7911×10^{-2}
5000	3.4023×10^{-4}	4.7935×10^{-4}	5.6392×10^{-4}	8.1923×10^{-4}
10000	3.0427×10^{-5}	1.1883×10^{-4}	3.4190×10^{-4}	1.1914×10^{-4}

Table 6.20: $MSE(\hat{\tilde{z}}_n^{\dagger})$ for Truncation Selection with $P_0(s,t) = \exp\{-|s-t|\}$ and $\alpha = 1$

Again, Table 6.20 is very similar to 6.18 and thus the interpretation is identical.

One thing to note is the wildly different behavior of $MSE(\hat{\mathbf{b}}_n)$ and $MSE(\hat{\tilde{z}}_n^{\dagger})$ when comparing the cases $a = 0$ and $a = 1$ for the Wiener and Ornstein-Uhlenbeck covariance functions. For the Ornstein-Uhlenbeck covariance, the values of the MSE do not change much between $a = 0$ and $a = 1$, while for the Wiener covariance, the values of the MSE increase by as much as 8900%. This may have to do with the magnitude of the covariance functions; the Wiener covariance, $P_0(s,t) = \min(s,t)$, is contained in the interval [0,1] for all $s, t \in [0,1]$, while the Ornstein-Uhlenbeck covariance, $P_0(s,t) = \exp(-|s-t|)$, is contained in the interval $[1, e]$ for all $s, t \in [0,1]$. This implies that a process with the Ornstein-Uhlenbeck covariance function has a larger variability and thus the number of processes that have a corresponding X that exceeds 1 is similar to the number of processes that have an X exceeding 0.

6.2.4 Comparing Selection

We have already compared $MSE(\hat{\mathbf{b}}_n)$ and $MSE(\hat{\tilde{z}}_n^{\dagger})$ for differing parameters under the same type of selection. We now compare the values of the MSEs under different types of selection. It appears that the rates of convergence between the directional selection and truncation selection are approximately equal. For most of the parameter choices of stabilizing selection, the rate of convergence appears to be the same as well. However, there are certain cases, such as $a_k = 1/k$, $\mu_k = 1/k^2$, $P_0(s,t) = \exp(-|s-t|)$ and $f(s) = 1$ (Table 6.11), where $MSE(\hat{\mathbf{b}}_n)$ is decreasing at a much slower rate than the corresponding values for other types of selection. It is because of this that it is hard to find a particular rate at which the MSEs are converging to 0. As for the values of the MSEs, there appears to be no way to determine if one set of parameters will produce larger values than another set of parameters. In the previous tables, for one type of selection, the values follow a particular pattern and for a different type of selection, it follows the opposite pattern, as in Tables 6.7 and 6.1.

6.3 Dependent Sample

As in the independent case, we have many parameters that are allowed to vary. As before, we shall let $T = [0,1]$, $\mathfrak{T}$ be the Borel sets on $[0,1]$ and μ be Lebesgue measure. Further, we let T_r be 50 equally spaced points between 0 and 1, not including 0. Also, we construct P_0 by summing G_0 and E_0, where G_0 and E_0 are either the Wiener covariance or the Ornstein-Uhlenbeck covariance. This guarantees that Ψ will be a covariance function. We also only consider the case when X is given as an integral (6.8) and $f(s) = 1$ or $f(s) = \exp(s)$. Due to the extensive computation time, we shall use $a_k^- = 1/k$ and $a_k^+ = 1/k$ for all k. Next, we construct P by first computing the sequence $\{a_k\}$ by using (5.25) and computing P by

$$P = P_0 + \sum_{k=1}^{50} a_k g_k \otimes g_k.$$

The cross-covariance Ψ is constructed by using (5.26) and the expansions

$$P + (n-1)\Psi = P_0 + (n-1)\Psi_0 + \sum_k a_k^{+,n} f_k^{+,n} \otimes f_k^{+,n}, \text{ and}$$

$$P - \Psi = P_0 - \Psi_0 + \sum_k a_k^{-} f_k^{-} \otimes f_k^{-},$$

for large n. For simplicity, we let $\{a_k^{+,n}\}_k$ be equal for all n. Once this is complete, we restrict P and Ψ to $T_r \times T_r$ and obtain the matrices $\mathbf{P}_r$ and $\boldsymbol{\Psi}_r$. Next, we construct the matrix

$$\boldsymbol{\Sigma}_{r,n} = \mathbf{I}_n \otimes \mathbf{P}_r + (\mathbf{J}_n - \mathbf{I}_n) \otimes \boldsymbol{\Psi}_r,$$

where $\otimes$ is the Kronecker product and $n \in \mathbb{N}$. The Cholesky decomposition of $\boldsymbol{\Sigma}_{r,n} = \mathbf{C}_{r,n}\mathbf{C}_{r,n}^T$ is computed, where $\mathbf{C}_{r,n}$ is a lower triangular matrix. We then generate an $rn \times 1$ vector of i.i.d. standard normal variables, $\mathbf{X}$. To obtain a vector with the variables we desire, we compute $\mathbf{Y} = \mathbf{C}_{r,n}\mathbf{X}$. We split this vector into n vectors in $\mathbb{R}^r$ by letting $z_i = [Y_{(i-1)r+1}, \ldots, Y_{ir}]$. Then the set of vectors $\{z_i\}_{i=1}^n$ is the desired set of random vectors with covariance $\mathbf{P}_r$ and cross-covariance $\boldsymbol{\Psi}_r$. We consider this set $\{z_i\}_{i=1}^n$ as one family of organisms. This process is repeated m times to create a sample of m independent families consisting of equally related organisms. This process is very resource intensive; creating the matrix $\boldsymbol{\Sigma}_{50,500}$ (6.25×10^8 elements) requires approximately 7 gigabytes of memory. Therefore, we keep the family size at a level that makes biological sense. The family size will be kept at a maximum of around 50 individuals. As for the estimation, we start with the largest number of families that we will consider, 200, and we generate 200 Poisson random variables with mean 50, say $\{n_j\}_{j=1}^{200}$. Similar to the independent case, we take an increasing sequence of nested subsets of this sample. We perform the estimation of $\mathbf{a}$, $\mathbf{b}$, G and $\bar{z}'$ on each of the sets in the sequence. As for the sieve parameter, d, we take d to be the minimum of the family sizes over all families. Due to the extensive computational time for one iteration (12-18 hours) and the number of different models, we repeat for 30 iterations. The computation

time indicates that it should be attempted to improve the computational time, although the majority of the computation was spent computing the $L^2(T)$ and $L^2(T \times T)$ norms. That is, in practical applications, the estimation should take far less time.

We shall explore the consistency results of the estimated mean square error of these estimators as m (and thus n) becomes large. Since we are increasing m, we shall write the estimates of $\mathbf{a}$, $\mathbf{b}$, G and $\bar{z}'$ as $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{G}_m$ and $\hat{\bar{z}}'_m$, respectively. In the subsections below, we consider one type of selection and choose $G_0(s,t) = \min(s,t)$ and $E_0(s,t) = \exp(-|s-t|)$, compare the various mean square errors as m increases. Once this discussion is complete, we choose $G_0(s,t) = \exp(-|s-t|)$ and $E_0(s,t) = \min(s,t)$ and discuss the results. We also plot the graphs of $\hat{G}_m$, G, $\hat{\bar{z}}'_m$ and $\bar{z}^\dagger$ for a few different models to visualize what the corresponding square errors are indicating.

The tables and figures below show that $MSE(\hat{\mathbf{a}}_m)$ and $MSE(\hat{G}_m)$ are both decreasing as we increase the number of families. This shows that $\hat{\mathbf{a}}_m$ and $\hat{G}_m$ are good estimators for $\mathbf{a}$ and G, respectively, at least for the various parameter values chosen. On the other hand, the MSEs of the proposed estimators $\hat{\mathbf{b}}_m$ and $\hat{\bar{z}}'$ are typically increasing as m increases. In some cases $MSE(\hat{\mathbf{b}}_m)$ and $MSE(\hat{\bar{z}}'_m)$ are of the order 10^3, which indicates that the estimated values are far from the corresponding true values. We shall address this issue in Chapter 7.

6.3.1 Directional Selection

The resulting mean square errors for directional selection are:

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\tilde{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s)=1$				
$m=50$	2.1840×10^{-1}	3.0162×10^{-1}	1.1174×10^{-2}	8.5957×10^{-4}
100	8.3607×10^{-2}	3.9035×10^{-1}	5.2018×10^{-2}	1.9597×10^{-4}
200	2.2531×10^{-2}	5.0202×10^{-1}	1.2941×10^{-1}	4.0381×10^{-5}
$f(s)=e^s$				
$m=50$	2.1827×10^{-1}	1.5866	1.8170×10^{-2}	7.7907×10^{-4}
100	9.1430×10^{-2}	1.3458	1.6342×10^{-1}	2.4655×10^{-4}
200	3.2142×10^{-2}	1.9084	4.2922×10^{-1}	4.1639×10^{-5}

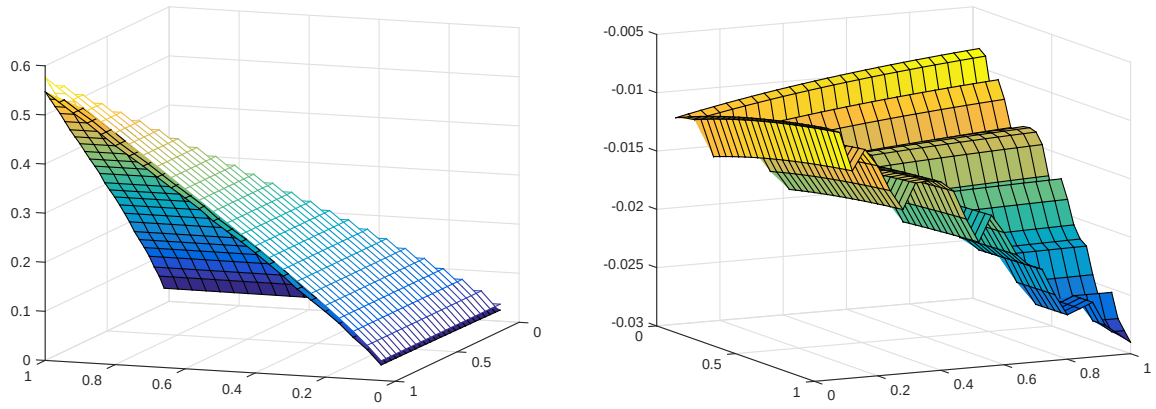
Table 6.21: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\tilde{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t)=\min(s,t)$ and $E(s,t)=\exp(-|s-t|)$ under directional selection

From Table 6.21, we see that $MSE(\hat{\mathbf{a}}_m)$ is decreasing to zero as n increases. Similarly, $MSE(\hat{G}_m)$ is also decreasing as m increases. As should be expected, it does not seem to matter the form of f, as the rate of decrease appears to be the same if we consider $f(s)=1$ or $f(s)=e^s$. On the other hand, $MSE(\hat{\mathbf{b}}_m)$ is actually increasing with m. This is most likely due to the naive estimator of c_k that was used. As $\hat{\tilde{z}}^{\dagger}_n$ depends on the selection gradient β, it should come at no surprise that $MSE(\hat{\tilde{z}}'_m)$ is also increasing as m increases. We see similar properties from Table 6.22.

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\tilde{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s)=1$				
$m=50$	2.7316×10^{-1}	4.0020×10^{-1}	2.1411	1.0640×10^{-2}
100	1.0742×10^{-1}	2.0046×10^{-1}	3.2478	1.0051×10^{-2}
200	3.8298×10^{-2}	4.1832×10^{-2}	3.7769	8.8998×10^{-4}
$f(s)=e^s$				
$m=50$	2.7316×10^{-1}	4.0020×10^{-1}	8.899×10^{-1}	3.7769×10^{-3}
100	1.0741×10^{-1}	2.0046×10^{-1}	1.0640	3.2478×10^{-3}
200	3.8298×10^{-2}	4.1823×10^{-2}	1.0504	2.1411×10^{-3}

Table 6.22: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\tilde{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t)=\exp(-|s-t|)$ and $E(s,t)=\min(s,t)$ under directional selection

For visualization of the square errors, we choose $G_0(s,t) = \min(s,t)$ and $f(s) = 1$. Figures 6.1a and 6.1b show that $\hat{G}_{200}$ is a fairly decent estimator of G, although it tended to slightly underestimate the true G.



(a) (b)

Figure 6.1: Directional Selection: (a) Graphs of the average of $\hat{G}_{200}$ and G. The shaded region represents the average $\hat{G}_m$, while the wireframe represents the true genetic covariance G. (b) Difference of the average $\hat{G}_{200}$ and G.

In Figure 6.2, we see that $\hat{\bar{z}}'_{200}$ is a poor estimator of $\bar{z}^\dagger$. The two functions are close near 0, but the farther from 0, the worse the estimation.

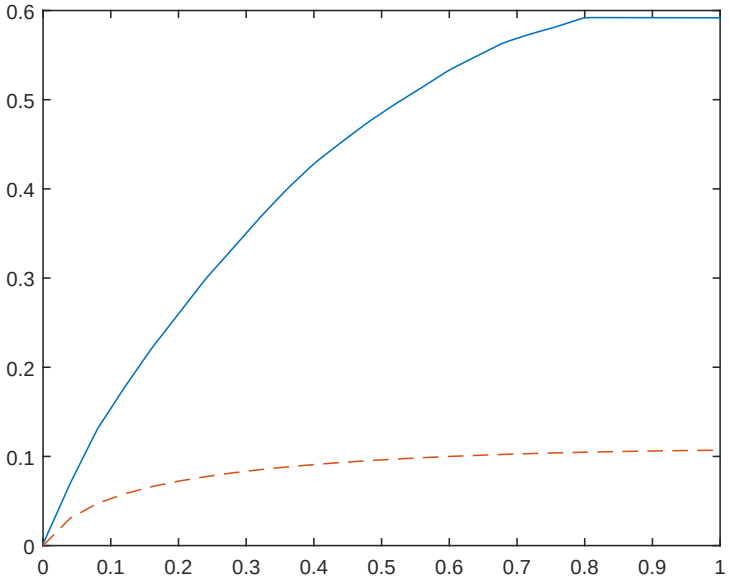


Figure 6.2: Graphs of the average of $\hat{\bar{z}}'_{200}$ and $\bar{z}^{\dagger}$. The dashed line represents $\bar{z}^{\dagger}$ and the solid line represents the average of the estimators $\hat{\bar{z}}'_{200}$.

6.3.2 Stabilizing Selection

For stabilizing selection, we consider the case when $v = 5$.

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\bar{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s) = 1$				
$m = 50$	2.1122×10^{-1}	8.3417×10^{-6}	1.1929×10^{-7}	7.9710×10^{-4}
100	8.4723×10^{-2}	1.0024×10^{-5}	4.7407×10^{-7}	2.5954×10^{-4}
200	2.1919×10^{-2}	5.2370×10^{-6}	4.2195×10^{-7}	4.2041×10^{-5}
$f(s) = e^s$				
$m = 50$	1.5362×10^{-1}	6.5057×10^{-5}	1.8429×10^{-6}	8.0354×10^{-4}
100	7.7518×10^{-2}	4.7304×10^{-5}	2.0582×10^{-6}	1.9246×10^{-4}
200	2.3017×10^{-2}	4.1483×10^{-5}	3.1784×10^{-6}	1.7316×10^{-4}

Table 6.23: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\bar{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t) = \min(s,t)$ and $E(s,t) = \exp(-|s-t|)$ under stabilizing selection with $v = 5$

We see from Table 6.23 that in either case, $MSE(\hat{\mathbf{a}}_m)$ and $MSE(\hat{G}_m)$ are both decreasing

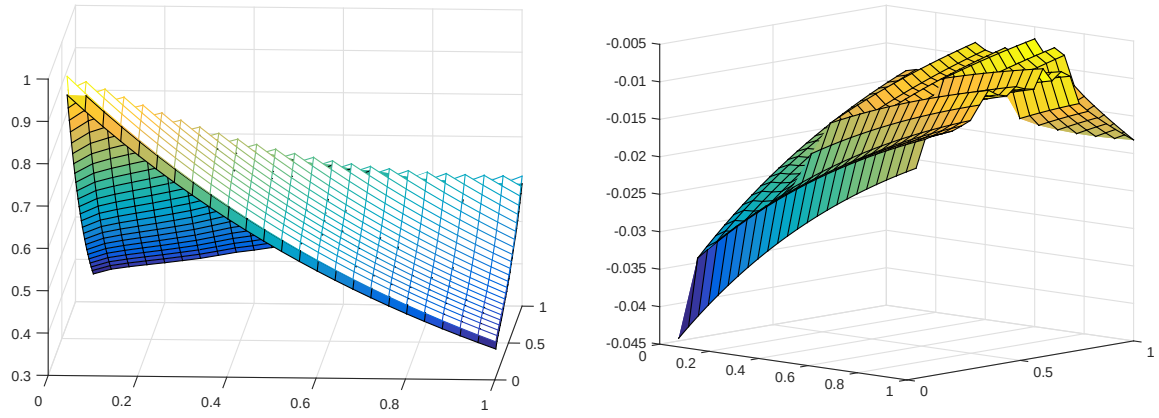
at about the same rate as the number of families increases. We also note that when $f(s) = 1$, $MSE(\hat{\mathbf{b}}_m)$ increases from $m = 50$ to $m = 100$ and decreases from $m = 100$ to $m = 200$ with $MSE(\hat{\mathbf{b}}_{200})$ being the smallest value. On the other hand, $MSE(\hat{\bar{z}}'_m)$ increases and then decreases, but $MSE(\hat{\bar{z}}'_{50})$ is the smallest value. When $f(s) = e^s$, we see that $MSE(\hat{\mathbf{b}}_m)$ is decreasing while $MSE(\hat{\bar{z}}'_m)$ is increasing.

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\bar{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s) = 1$				
$m = 50$	2.7316×10^{-1}	2.5001×10^{-5}	4.2267×10^{-5}	1.3256×10^{-2}
100	1.0742×10^{-1}	1.2872×10^{-5}	1.1028×10^{-5}	1.0695×10^{-3}
200	3.8298×10^{-2}	1.4508×10^{-5}	6.6728×10^{-6}	9.3578×10^{-4}
$f(s) = e^s$				
$m = 50$	2.7316×10^{-1}	1.5559×10^{-4}	1.8419×10^{-4}	1.3256×10^{-2}
100	1.0743×10^{-1}	8.1948×10^{-5}	6.5425×10^{-5}	1.0690×10^{-3}
200	3.8297×10^{-2}	9.4144×10^{-5}	4.1791×10^{-5}	9.3599×10^{-4}

Table 6.24: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\bar{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t) = \exp(-|s-t|)$ and $E(s,t) = \min(s,t)$ under stabilizing selection with $v = 5$

Table 6.24 shares much of the same interpretation as Table 6.23 for $MSE(\hat{\mathbf{a}}_m)$ and $MSE(\hat{G}_m)$, but is very different for $MSE(\hat{\mathbf{b}}_m)$ and $MSE(\hat{\bar{z}}'_m)$. For both choices of f, we see that $MSE(\hat{\mathbf{b}}_m)$ is decreasing from $m = 50$ to $m = 100$ and increasing slightly from $m = 100$ to $m = 200$. If we consider $MSE(\hat{\bar{z}}'_m)$, we see that the values are decreasing as m increases. This doesn't appear consistent with Table 6.23. To visualize the square errors, we let $f(s) = e^s$ and $G(s,t) = \exp(-|s-t|)$.

Figure 6.3a shows the difference of the average $\hat{G}_{200}$ and G. From Figure 6.3b we see that $\hat{G}_{200}$ tends to underestimate G slightly, with the largest difference when $s, t = 0$ and the smallest around $s, t = 1$.



(a) (b)

Figure 6.3: (a) Graphs of the average of $\hat{G}_{200}$ and G. The shaded region represents the average $\hat{G}_m$, while the wireframe represents the true genetic covariance G. (b) Difference of the average $\hat{G}_{200}$ and G.

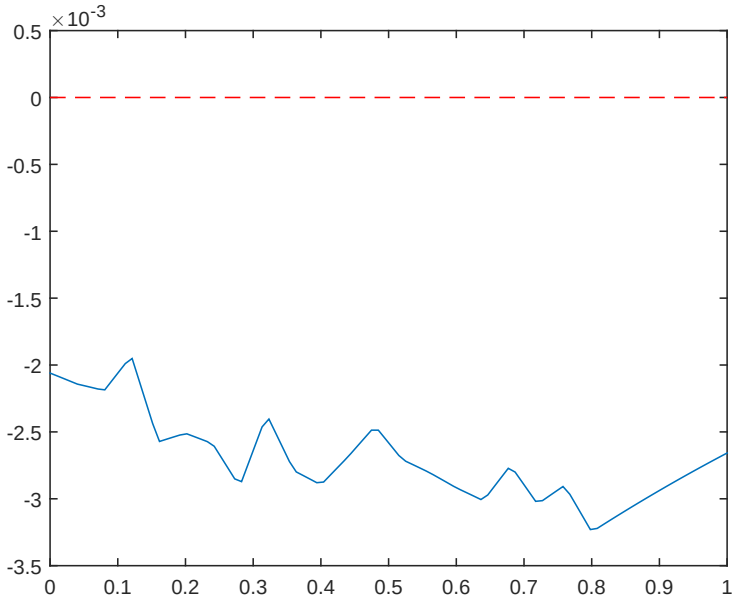


Figure 6.4: Graphs of the average of $\hat{\bar{z}}'_{200}$ and $\bar{z}^{\dagger}$. The dashed line represents $\bar{z}^{\dagger}$ and the solid line represents the average of the estimators $\hat{\bar{z}}'_{200}$.

Figure 6.4 shows that on average $\hat{\bar{z}}'_{200}$ tends to underestimate $\bar{z}^{\dagger}$, although by a maximum of about 3.3×10^{-3}.

6.3.3 Truncation Selection

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\tilde{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s)=1$				
$m=50$	1.9113×10^{-1}	2.3966×10^{3}	1.6245×10^{2}	2.9452×10^{-4}
100	7.9772×10^{-2}	2.3941×10^{3}	1.6112×10^{2}	1.9307×10^{-4}
200	2.4072×10^{-2}	2.3907×10^{3}	1.5885×10^{2}	9.7957×10^{-5}
$f(s)=e^s$				
$m=50$	1.9112×10^{-1}	9.0831×10^{2}	5.3118×10^{1}	2.9452×10^{-4}
100	7.9771×10^{-2}	9.0733×10^{2}	5.2425×10^{1}	1.9307×10^{-4}
200	2.4070×10^{-2}	9.0575×10^{2}	5.1460×10^{1}	9.7959×10^{-5}

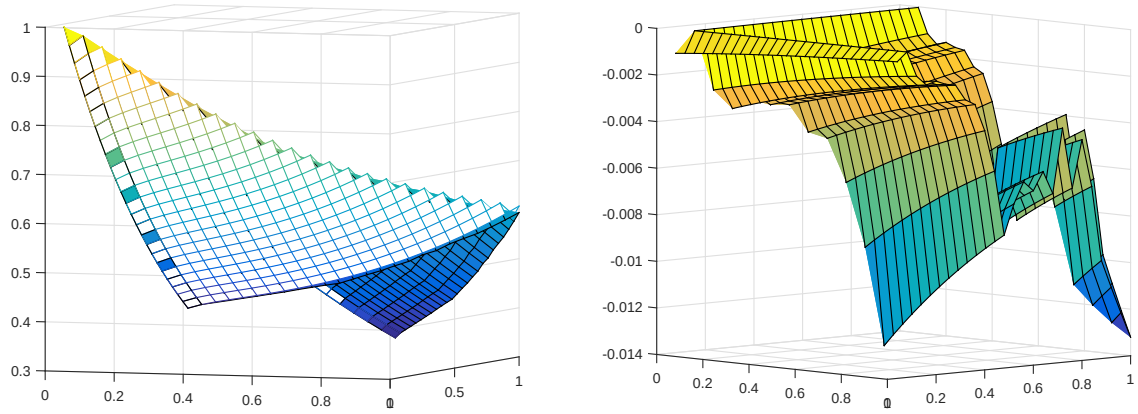
Table 6.25: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\tilde{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t)=\min(s,t)$ and $E(s,t)=\exp(-|s-t|)$ under Truncation Selection with $\alpha=0$

Table 6.25 shows that the mean square errors for $\hat{\tilde{z}}'_m$ and $\hat{\mathbf{b}}_m$ are very large. This indicates that the estimation of β proposed in this dissertation should not be used for truncation selection with a relatively "weak" selection parameter and possibly for truncation selection in general. We also see that $MSE(\hat{\mathbf{a}}_m)$ and $MSE(\hat{G}_m)$ are both decreasing with very similar values for both choices of f.

	$MSE(\hat{\mathbf{a}}_m)$	$MSE(\hat{\mathbf{b}}_m)$	$MSE(\hat{\tilde{z}}'_m)$	$MSE(\hat{G}_m)$
$f(s)=1$				
$m=50$	1.7446×10^{-1}	9.0756×10^{2}	1.7566×10^{3}	1.1908×10^{-3}
100	8.9326×10^{-1}	9.0589×10^{2}	1.7546×10^{3}	3.3510×10^{-3}
200	3.9930×10^{-2}	9.06056×10^{2}	1.7540×10^{3}	8.7931×10^{-5}
$f(s)=e^s$				
$m=50$	2.7316×10^{-1}	2.4000×10^{3}	4.6195×10^{3}	1.3256×10^{-2}
100	1.0742×10^{-1}	2.3660×10^{3}	4.6156×10^{3}	1.0695×10^{-3}
200	3.8298×10^{-2}	2.3640×10^{3}	4.6078×10^{3}	9.3578×10^{-4}

Table 6.26: Mean Square Error of $\hat{\mathbf{a}}_m$, $\hat{\mathbf{b}}_m$, $\hat{\tilde{z}}'_m$ and $\hat{G}_m$ with $G_0(s,t)=\exp(-|s-t|)$ and $E(s,t)=\min(s,t)$ under Truncation Selection with $\alpha=0$

Table 6.26 shares the same interpretation as that of Table 6.25. To visualize the square errors, we let $G_0 = \exp(-|s-t|)$ and $f(s) = 1$. Figures 6.5a and 6.5b show that the estimation of G is fairly good, although appears to underestimate G slightly. Overall, $\hat{G}_{200}$ is a good estimator of G.



(a) (b)

Figure 6.5: (a) Graphs of the average of $\hat{G}_{200}$ and G. The shaded region represents the average $\hat{G}_m$, while the wireframe represents the true genetic covariance G. (b) Difference of the average $\hat{G}_{200}$ and G.

Figure 6.6 shows the reason that the MSE of $\hat{\bar{z}}'_m$ is quite large. The estimate $\hat{\bar{z}}'_{200}$ stays close to 0, with some deviations. On the other hand, $\bar{z}$ is sometimes around 55 times the corresponding estimate. This indicates that another estimator of $\mathbf{c}$ should be found.

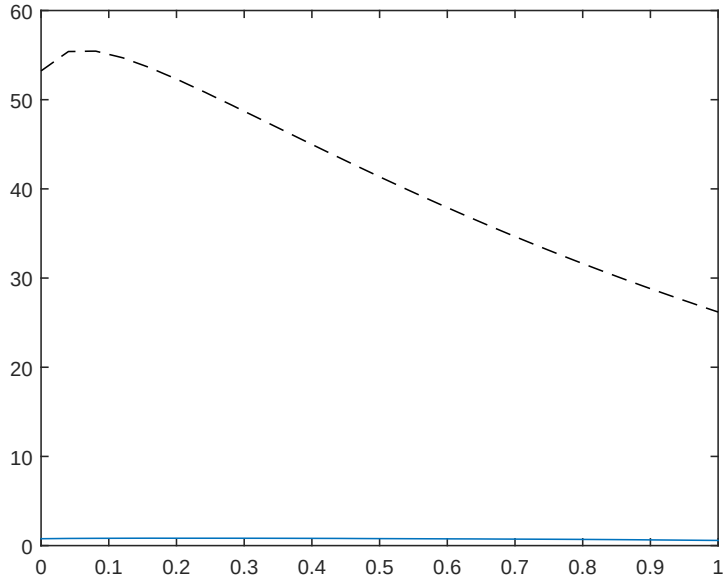


Figure 6.6: Graphs of the average of $\hat{\bar{z}}'_{200}$ and $\bar{z}^{\dagger}$. The dashed line represents $\bar{z}^{\dagger}$ and the solid line represents the average of the estimators $\hat{\bar{z}}'_{200}$.

6.3.4 Comparing Selection

We see that overall, $\hat{\mathbf{a}}_m$ and $\hat{G}_m$ are good estimators for $\mathbf{a}$ and G, respectively. The mean square errors for both estimators are decreasing as the number of families increases. It appears that the estimators for $\bar{z}^{\dagger}$ and $\mathbf{b}$ are not optimal. They are acceptable in certain cases, such as the stabilizing selection used in Subsection 6.3.2, but should be avoided with truncation selection. These results indicate that a different estimator of $\mathbf{c}$ should be found.

Chapter 7

Conclusion and Discussion

We now recall the main results of this dissertation and discuss some of the open problems that we found. We first looked to refine the definition of the selection gradient β. Assuming that the covariance function P of a Gaussian process is of finite trace, $(T, \mathcal{T}, \mu)$ is a σ-finite measure space and that the only μ-negligible function in the reproducing kernel Hilbert space $\mathcal{H}(P)$ is the zero function, we found that the integral operator $\mathcal{P}$ maps M, the image of the inclusion map $j : \mathcal{H}(P) \to L^2(T)$ one-to-one into a dense subset of $\mathcal{H}(P)$. We then extended $\mathcal{P}$ to $\bar{\mathcal{P}} : \mathcal{M} \to \mathcal{H}(P)$, where $\mathcal{M}$ is the $\sigma(M, \mathcal{F})$-completion of the space M. It is then shown that the map $\bar{\mathcal{P}}$ is a bijection between $\mathcal{M}$ and $\mathcal{H}(P)$ and that β lies in $\mathcal{M}$. Furthermore, β is the unique solution to the equation $\bar{\mathcal{P}}\beta = s$, where s is the selection differential. Given a complete orthonormal sequence in $\mathcal{H}(P)$, we find an expansion of β in the space $\mathcal{M}$ and find that its coefficients depend on $\mathtt{Cov}\,(w, Z)$ and $\mathbf{a}$.

Since we only had an estimator of the covariance function P when the mean function $\bar{z}$ is identically zero, we found a joint estimator $(\hat{\bar{z}}, \hat{P})$ of the mean and covariance functions of a Gaussian process by using sieve estimation and the Gaussian Dichotomy Theorem. It is then shown that the joint estimator is asymptotically unbiased and weakly and mean-square consistent for $(\bar{z}, P)$ as long as the sieve parameter, d is $O(n)$. Furthermore, the coefficients in the expansions of $\hat{\bar{z}}$ and $\hat{P}$, $(\hat{\boldsymbol{\mu}}_n, \tilde{\mathbf{a}}_n)$ are strongly consistent for the coefficients $(\boldsymbol{\mu}, \mathbf{a})$ in

$\ell^2 \times \ell_c^2$ as long as $d = o(n^\sigma)$ for some $\sigma \in (0,1)$. We then construct an estimator of β and show that this estimator is strongly consistent in the topology of $\mathcal{M}$. For η in the linear span of the orthonormal functions $\{g_k\}$, we have shown that $\left(\hat{\beta}_n, \eta\right)_{L^2(T)}$ is strongly consistent for $(\beta, \eta)_{L^2(T)}$. We have not shown the general case for $\eta \in \mathcal{H}(P)$. This leads to the first question:

Question 1: Is Conjecture 4.2.1 true?

Simulations point towards yes, but it is unclear on the growth rate on d or the restrictions on the fitness functions required to prove this.

Our next goal was to find an estimator of the selection gradient when the organisms involved in the sample are related. This first led to finding an estimator of the phenotypic covariance function P. We considered a family of n equally related organisms and transformed the corresponding phenotypes to a sample of independent processes, albeit with covariances $P + (n-1)\Psi$ and $P - \Psi$. The likelihood function $\dfrac{d\mathbb{P}}{d\mathbb{P}_0}$ was maximized over $\ell_c^2(B) \times \ell_c^2(B)$, for a finite set B. The issue with this was that there is only one observation used to estimate $a_k^{+,n}$ and we could say nothing about the asymptotic properties of such an estimator. Thus, we needed to consider m independent families of equally-related organisms. Using this sample scheme and the estimation for one family, we were able to find estimators for the phenotypic covariance and the additive-genetic covariance functions. We then come upon more questions:

Question 2: Are the estimators for **a** given in (5.25) consistent in any sense?

Question 3: Is the estimator $\hat{G}$ of the genetic covariance function consistent in any sense?

The simulations in Section 6.3 seem to indicate that the answer to both of these questions is yes, there is some sort of consistency occurring.

Since the organisms in such a sample are dependent, their corresponding fitness functions are dependent random variables. As we do not have much information on the distribution of W, it is difficult, if not impossible to find an estimator for $c_k = \texttt{Cov}\,(w, U_k)$. Lacking such

information, we use the naive estimator, the same estimator of c_k used for an independent sample. The simulations show that the mean square errors for $\hat{\mathbf{b}}_m$ and $\hat{\bar{z}}'_m$ are not converging to zero and in some cases are quite large. This indicates that we need to find a different estimator for c_k.

Question 4: How can we estimate $c_k = \texttt{Cov}\,(w, U_k)$ for a sample of dependent organims?

This question itself leads to other questions. For example, what restrictions on W are required to obtain consistency of $\hat{\mathbf{c}}_m$?

In summary, our results provide a joint estimator of the mean and covariance functions of a Gaussian process that is strongly consistent. We also construct an estimator the selection gradient of a function-valued trait that is "almost" strongly consistent, in the sense that it is strongly consistent on a dense set. Furthermore, we provide an explicit estimator for both the phenotypic covariance and genotypic covariance functions that can be used in samples consisting of equally-related organisms.

References

[1] A. Antoniadis and J.H. Beder. Joint estimation of the mean and the covariance of a Banach-valued gaussian vector. *Statistics*, 20(1):77–93, 1989.

[2] N. Aronszjan. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[3] J.Y. Bae. *Estimation of the Additive Genetic Covariance Function*. PhD thesis, University of Wisconsin - Milwaukee, 2009.

[4] A. Basilevsky. *Statistical Factor Analysis and Related Methods: Theory and Applications*. John Wiley and Sons, Inc., 1994.

[5] J. H. Beder. A sieve estimator for the mean of a Gaussian process. *Annals of Statistics*, 15:59–78, 1987.

[6] J. H. Beder. A sieve estimator for the covariance of a Gaussian process. *Annals of Statistics*, 16:648–660, 1988.

[7] J. H. Beder and R. Gomulkiewicz. Computing the selection gradient and evolutionary response of an infinite-dimensional trait. *Journal of Mathematical Biology*, 36:299–319, 1998.

[8] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, 2011.

[9] P. Cartier. Une étude des covariances mesurables. In L. Nachbin, editor, *Mathematical Analysis and Applications*, pages 267–316. Academic Press, 1981.

[10] J.B. Conway. *A Course in Functional Analysis*. Springer-Verlag, 2nd edition, 1990.

[11] G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, 2nd edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[12] R. Gomulkiewicz and J. H. Beder. The selection gradient of an infinite-dimensional trait. *SIAM Journal of Applied Math*, 56:509–523, 1996.

[13] N. Heckman. Functional data analysis in evolutionary biology. *Recent Advances and Trends in Nonparametric Statistics*, pages 49–60, 2003.

[14] M. Kirkpatrick and N. Heckman. A quantitative genetic model for growth, shape, reation norms, and other infinite-dimensional characters. *Journal of Mathematical Biology*, 27:429–450, 1989.

[15] R. Lande. Quantitative genetics analysis of multivariate evolution, applied to brain:body allometry. *Evolution*, 33:402–416, 1979.

[16] R. Lande and S. J. Arnold. The measurement of selection on correlated characters. *Evolution*, 37:1210–1226, 1983.

[17] M. Loève. *Probability Theory I*. Springer-Verlag, New York, 4th edition, 1977.

[18] M. Lukić and J. H. Beder. Stochastic processes with sample paths in reproducing kernel hilbert spaces. *Transactions of the American Mathematical Society*, 353(10):3945–3969, 2001.

[19] M. Lynch and B. Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Incorporated, 1998.

[20] J. Neveu. *Processus Aléatoires Gaussiens*. Publications du Séminaire de Mathématiques Supérieures. Les Presses de l'Université Montréal, 1968.

[21] E. Parzen, Stanford University. Applied Mathematics, Statistics Laboratory, and United States. Office of Naval Research. *Statistical Inference on Time Series by Hilbert Space Methods, I.* Number v. 1 in Applied Mathematics and Statistics Laboratory Stanford, Calif..: Technical report. Applied Mathematics and Statistics Laboraotry, Stanford University, 1959.

[22] D. Roff. *Evolutionary Quantitative Genetics*. Evolutionary biology : Genetics. Springer US, 1997.

[23] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 1973.

[24] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Kernel Methods for Pattern Analysis. Cambridge University Press, 2004.

Appendix

MATLAB Code

Independent Simulations

The following MATLAB program was used for the simulations using independent data.

```
function [ sq_err,N_vec,estimate,zstar_hat,Z_all,varargout ]...
                    = sim_ind(n,N,iter,t,t_range,P0,G,a,alpha,...
                              d,f,Comp_P,selection,select_param,...
                              X_type,varargin)
%
%%%%%%Input%%%%%%%
% N = Number of times to increase sample size (scalar)
% n = amount sample size increase (scalar)
% iter = number of iterations (scalar)
% f = anonymous function handle or vector
%                         (anonymous function or Mx1)
% t = grid of time points (M x 1)
% t_range =  vector containing the (finite) range
%                                  of time points (2x1 or 1x2)
% P_0 = Candidate Covariance function (anonymous function)
```

```
% G = Genetic Covariance Function (anonymous function)
% a = vector of coefficients in the expansion of P (M x 1)
% alpha = vector of coefficients in the expansion
%                                          of zbar (M x 1)
% d = sieve parameter (anonymous function)
% selection = Type of selection ('directional',
%                              'truncated' or 'stabilizing')
% select_param = Selection parameter (scalar)
% X_type = type of X ('integral' or 'sum')
%%% varargin/varargout %%%
% params = struct containing the computed parameters
%   params.P0_matrix = matrix with entries P0(ti,tj) (M x M)
%   params.P_matrix = matrix with entries P(ti,tj) (M x M)
%   params.b = vector containing parameters {bk} (M x 1)
%   params.c = vector containing parameters {ck} (M x 1)
%   params.g = Gram-Schmidt Orthonormalized functions P_ti
%                          (M x 1 cell of symbolic functions)
%   params.zbar = mean of newborns in current generation
%                                       (symbolic function)
%   params.zbar_vec = vector of zbar evaluated at t (M x1)
%   params.zstar = mean of newborns in next generation
%                                        (symbolic function)
%%%%%%%%Output%%%%%%%%
% sq_error = struct containing:
%   sq_error.a = Matrix containing square errors of a -
%       columns contain independent iterations (N x iter)
%   sq_error.b = Matrix containing square errors of b -
```

```
%       columns contain independent iterations (N x iter)
%   sq_error.c = Matrix containing square errors of c -
%       columns contain independent iterations (N x iter)
%   sq_error.alpha = Matrix containing square errors of alpha -
%       columns contain independent iterations (N x iter)
%
%
% N_vec = vector containing sample sizes in each iteration
%                                                   (N x 1)
% estimate = struct containing:
%   estimate.a = Matrix containing final estimate of a
%                                                (M x 1 x iter)
%   estimate.b = Matrix containing final estimate of b
%                                                (M x 1 x iter)
%   estimate.c = Matrix containing final estimate of c
%                                                (M x 1 x iter)
%   estimate.alpha = Matrix containing final estimate of alpha
%                                                (M x 1 x iter)
%
%zstar_hat = cell containing z*_hat(N x iter;
%                                          anonymous functions)
%Z_all = matrix containing all data used (M x n*N x iter)

selection=lower(selection);
X_type=lower(X_type);
%%%%%%%%%%%%%%%%%%%% Input check %%%%%%%%%%%%%%%%%%%%
if (size(t,1)>1 && size(t,2)>1)
```

```
    disp(' t must be a vector')
    return
else
    if size(t,2)>1
    t=t';
    end
end

if (size(a,1)>1 && size(a,2)>1)
    disp('a must be a vector')
else
    if size(a,2)>1
        a=a';
    end
end
if (size(alpha,1)>1 && size(alpha,2)>1)
    disp('alpha must be a vector')
else
    if size(alpha,2)>1
        alpha=alpha';
    end
end

switch X_type
    case 'sum'
    if isa(f,'function_handle')
        disp('f cannot be a function')
```

```
        return
    end
    if (size(f,1)>1&& size(f,2)>1)
        disp('f must be a vector if X is a sum')
        return
    else
        if size(f,2)>1
            f=f';
        end
    end
    case 'integral'

    if (strcmp(X_type,'integral')&& ~isa(f,'function_handle'))
        disp('f must be a function handle if X is an integral')
        return
    end


end
%%%%%%%%%%%%%%%%%%%% End Input Check %%%%%%%%%%%%%%%%%%%%




%Create cell array of sections of the genetic covariance
%                                          (symbolic functions)
G_t=cell(length(t),1);
```

```
syms s r
 for i=1:length(t)
     G_t{i}=subs(G(s,r),r,t(i));
 end


%%%%% Output Initialization %%%%%%
sq_err=struct('a',zeros(N,iter),'b',zeros(N,iter),'c',...
                        zeros(N,iter),'alpha',zeros(N,iter));
N_vec=zeros(N,1);
zstar_hat=cell(N,1);
Comp_P=lower(Comp_P);
switch Comp_P
    case 'compute'

        [ P0_matrix,~, P_matrix , g] = compute_P(t,P0,a );
        [T,S]=meshgrid(t,t);
        G_mat=makeSymmetric(G(T,S),'G_mat');
        [~,p]=chol(P_matrix-G_mat);
        if (p~=0)
            error('P-G is not pd,so E=P-G is not a covariance')
        end
        [zbar,zbar_vec]=compute_zbar(alpha,g,t);
        [ c,b ] = compute_b( t,  t_range ,P0,g, f,a,zbar,...
                    P0_matrix,selection,select_param,alpha);
                        chol_P0=chol(P0_matrix,'lower');
        inv_chol_P0=chol_P0\eye(length(t));
```

```
gamma_G=cell(length(t),1);
 for j=1:length(t)
gamma_G{j}=compute_lincomb(G_t,inv_chol_P0(j,:));
 end
        [~,zstar]=compute_zstar(gamma_G,inv_chol_P0,b,zbar);
        params=struct('P0_matrix',P0_matrix,'P_matrix',...
             P_matrix, 'b',b,'c',c,'g',{g},'zbar',zbar,...
                      'zbar_vec',zbar_vec, 'zstar',zstar);

        varargout{1}=params;

    case 'given'
     params=varargin{1};
     P0_matrix=params.P0_matrix;
     P_matrix=params.P_matrix;
     b=params.b;
     c=params.c;
     g=params.g;
     zbar=params.zbar;
     zbar_vec=params.zbar_vec;
     zstar=params.zstar;
    otherwise
      error('Comp_P must either be ''given'' or ''compute'' ')

end
```

```
a_hat=zeros(length(t),1,iter);
b_hat=a_hat;
c_hat=a_hat;
alpha_hat=a_hat;
%%%% End Output Initialization %%%%

%%%%%%%%% Start Iterations %%%%%%%%%
Z_all=zeros(length(t),n*N,iter);
for j=1:iter
    Z_sim=mvnrnd(zbar_vec,P_matrix,n*N)';%Z_i is
    %                                   in the ith column
    Z_all(:,:,j)=Z_sim;
%%%%%%%%%%%%%%%%%%%%%%Main Simulation %%%%%%%%%%%%%%%%%%%%%
    chol_P0=chol(P0_matrix,'lower');
    for i=1:N
        n_sim=n*i;
        dn=floor(d(n_sim));
        if dn>=length(t);
            dn=length(t);
        end
        Z=Z_sim(:,1:n_sim); %Z_i is in the ith column
        [ W ] =compute_W( Z,t,f,selection,select_param ,...
                                              n_sim,X_type);
        [a_hat(:,:,j), b_hat(:,:,j),c_hat(:,:,j),...
            alpha_hat(:,:,j)] = estimate_b( Z,W,chol_P0...
                                          ,n_sim,alpha,dn);
```

```
        sq_err.b(i,j)=sum((b_hat(:,:,j)-b).^2);
        sq_err.c(i,j)=sum((c_hat(:,:,j)-c).^2);
        sq_err.a(i,j)=sum((a_hat(:,:,j)-a).^2);
        sq_err.alpha(i,j)=sum((alpha_hat(:,:,j)-alpha).^2);

        N_vec(i)=n_sim;
        zbar_hat=compute_lincomb(g,alpha_hat);
        [~,zstar_anon]=compute_zstar(gamma_G,inv_chol_P0,b_hat,
          ↪ zbar_hat);
        zstar_hat{i,j}=zstar_anon;


    end
%%%%%%%%%End Main Simulation%%%%%%%%%
end
%%%%%%%%%End Iterations%%%%%%%%%
estimate=struct('a',a_hat,'b',b_hat,'c',c_hat,...
                                        'alpha',alpha_hat);



end

%%%%%%%%%%% AUXILLARY FUNCTIONS %%%%%%%%%%%

function [ W ] =compute_W( Z,t,f,selection,select_param...
```

```
                                        ,n,X_type )


switch X_type
    case 'integral'
        X=zeros(n,1);
        f_vec=f(t);
        for i=1:n
            X(i)=trapz(t, Z(:,i).*f_vec); %approximates
            %   the integral int z_t f(t)dt by trapezoidal rule
        end
    case 'sum'
        X=Z'*f;
end

switch selection
    case 'directional'
        W=exp(X); %Directional selection
    case 'stabilizing'
        W=exp(-(X.^2)./(2*select_param.^2));
    case 'truncated'
        W=(X>select_param);
end


end

function [a_hat, b_hat,c_hat,alpha_hat ] = ...
```

```
                                estimate_b( Z,W,chol_P0,n,alpha,dn)
%Estimates the vectors a,b,c and alpha

 U=chol_P0\Z; %U_k is in the kth row
 U_bar=mean(U,2);

if any(alpha~=0)
    U_Ubar=zeros(size(U)); %Uk-Ubar
    for i=1:n
        U_Ubar(:,i)=U(:,i)-U_bar;
    end
    U_sq=sum(U_Ubar.^2,2)./n;
else
    U_sq=sum(U.^2,2)./n;
end

a_hat=[(U_sq(1:dn)-1);zeros(numel(U_sq)-dn,1)];
W_U=U*W./n;
w_U=W_U./mean(W);
c_hat=w_U-U_bar;
c_hat=[c_hat(1:dn);zeros(numel(U_sq)-dn,1)];
b_hat=[c_hat(1:dn)./U_sq(1:dn);zeros(numel(U_sq)-dn,1)];
alpha_hat=[U_bar(1:dn);zeros(numel(U_sq)-dn,1)];

end

function M=makeSymmetric(M,name)
```

```
%remove rounding error from P_matrix to make P_matrix symmetric
 issym=@(x) all(all(x==x.'));
 i=20;
 while ~issym(M)
     M=round(M,i);
     i=i-1;
 end

disp([name, ' has been rounded to ',num2str(i+1), ' digits'])

end

function [ P0_matrix,P, P_matrix , g] = compute_P( t, P0,a )
% computes a covariance function of the form
% P(s,t)=P0(s,t)+sum(a_i g_i(s)g_i(t)), where {g_i} are
% orthonormal in the RKHS corresponding to P0
n_t=length(t);
[T,S]=meshgrid(t,t);
P0_t=cell(n_t,1);
g=cell(n_t,1);
P0_matrix=P0(S,T);
chol_P0=chol(P0_matrix,'lower');
inv_chol_P0=chol_P0\eye(n_t);


%create a cell array of sections of P0
syms s r
```

```
 for i=1:n_t
     P0_t{i}=subs(P0(s,r),r,t(i));
 end

 %orthonormalize P0 to calculate g
 for j=1:n_t
g{j}=compute_lincomb(P0_t,inv_chol_P0(j,:));
 end

 %calculate P from a and g
 P=0;
for i=1:n_t
    P=P+a(i)*g{i}*subs(g{i},r);
end
P=P0(s,r)+P;

% create a matrix P(ti,tj) to generate random sample
P_anon=matlabFunction(P);
P_matrix=P_anon(T,S);

P_matrix=makeSymmetric(P_matrix,'P_matrix');

%check positive definiteness of P_matrix
[~,p]=chol(P_matrix);
if (p~=0)
    error('P_matrix is not positive definite')
```

```
else disp('P_matrix is positive definite')
    return

end

end


function f=compute_lincomb(X,v)
%computes linear combinations of the form sum(a_i*g_i) where
%g is a cell array of symbolic functions
f=0;
for i=1:length(v)
    f=f+v(i)*X{i};
end


end
function [ c,b ] = compute_b( t,t_range,P0,g, f,a,zbar,...
                        P0_matrix,selection,select_param,alpha)
% calculates the parameters of the selection differential (c)
% and theselection gradient (b)
syms s r
n_t=length(t);
% compute the integral of g_i*f
int_gf=zeros(n_t,1);
digits(25)
```

```
for j=1:n_t
    int_gf(j)=vpa(int(g{j}*f(s),t_range(1),t_range(2) ) );

end
%%%% Compute {g_i(tj)} %%%%
g_t=zeros(n_t,n_t);
for i=1:n_t

    for j=1:n_t
        g_t(i,j)=vpa(subs(g{i},s,t(j)));
    end

end

int_P0f=zeros(n_t,1);
for j=1:n_t
    int_P0f(j)=vpa(int(subs(P0(s,r),r,t(j))*f(s),...
                        t_range(1),t_range(2)));
end

int_Pf=int_P0f+g_t'*(a.*int_gf);
chol_P0=chol(P0_matrix,'lower');
if any(alpha~=0)
muX=vpa(int(zbar.*f(s),t_range(1),t_range(2)));
else
    muX=0;
```

```
end

varX=vpa(int(int(P0(s,r).*f(s).*f(r),r,t_range(1),...
    t_range(2)),s,t_range(1),t_range(2)))+(a.*int_gf)'*int_gf;
switch selection
    case 'directional'

         c=chol_P0\int_Pf;

    case 'stabilizing'

            c=(-muX/(select_param.^2+varX))*(chol_P0\int_Pf);
    case 'truncated'

            c=exp(-.5.*((-muX+select_param).^2)./varX)./...
               (sqrt(2.*pi.*varX).*normcdf(select_param,...
                muX,sqrt(varX),'upper')).*(chol_P0\int_Pf);

end



b=c./(1+a);
display(' b has been computed')
end
function [zbar,zbar_vec]=compute_zbar(alpha,g,t)
if any(alpha~=0)
    disp('zbar~=0')
```

```
 zbar=compute_lincomb(g,alpha);
 zbar_anon=matlabFunction(zbar);
 zbar_vec=zbar_anon(t);
else
    zbar=0;
    zbar_vec=zeros(size(t));
    disp('zbar=0')
end;

end


function [zstar,zstar_anon]=compute_zstar(gamma_G,inv_chol_P0,b
   ↪ ,zbar)
%compute/estimate zbar*
%G_t = cell array of symbolic functions of sections of the
%                   genetic covariance function
%b = coefficients of selection gradient
%zbar = mean process zbar(t) - symbolic function

G_beta=compute_lincomb(gamma_G,b);
zstar=G_beta+zbar;
zstar_anon=matlabFunction(zstar);
end
```

Dependent Simulations

The following MATLAB program was used for the simulations using equally-related organisms.

```
function [ Psi,Psi_hat,estimates,param,sq_err,P_added,...
        P_computed,num_fam_vec,Psi_int,zprime,zprime_hat,...
         zprime_int,zprime_vec,zprime_hat_vec,varargout]= ...
          related_anon( n,N,family,iter,reln,t,t_range,...
                    G0,E0, a_minus,a_plus,d,f,selection,...
               select_param ,X_type,smpl_type,generate,...
                                                 varargin)


% Example Commands:
%
%[ Psi,Psi_hat,estimates,param,sq_err,P_added,...
%       P_computed,num_fam_vec,Psi_int,zprime,zprime_hat...
%        ,zprime_int]= related( 50,4,200,0.5,t,t_range,...
%                G0,E0,a_minus,a_plus,d,f,selection,1,...
%                                    'integral',smpl_type)
%
%%%%%%Input%%%%%%%
%
% N = Number of times to increase sample size (scalar)
% n = mean family size (scalar)
% family = total number of families (scalar)
% iter = number of different simulations (scalar)
% t = grid of time points (M x 1)
% t_range =  vector containing the (finite) range of
```

```
%                                              time points (2x1)
% f = anonymous function handle or vector
%                                    (anonymous function or Mx1)
% reln = relationship coefficient (scalar)
% G0 = candidate genetic covariance function
%                                           (anonymous function)
% E0 = candidate environmental covariance function
%                                           (anonymous function)
% a_minus = vector of coefficients in the expansion of
%                                                 P-Psi (M x 1)
% a_plus = vector of coefficients in the expansion of
%                                             P+(n-1)Psi (M x 1)
% d = sieve parameter (anonymous function)
% selection = Type of selection ('directional','truncated' or
%                                                'stabilizing')
% select_param = Selection parameter (scalar)
% X_type = type of X ('integral' or 'sum')
% smpl = type of sample ('rnd' or 'given')
% generate = string ('generate' or 'given') signifies
%      whether or not the data is to be generated or is given.
% varargin{1} = vector containing the number of
%                        observations in each family (family x
   ↪ 1)
% varargin{2} = Generated data (family x iter)
%
%%%%%%%%Output%%%%%%%%
% Psi = true cross-covariance = aG(s,t) (anonymous function)
```

```
% Psi_hat = cell array containing estimates of Psi
%                                (anonymous functions, N x 1)
% estimates = struct containing:
%  estimates.a = Matrix containing estimates of a
%                                          (M x 1 x N x iter)
%  estimates.b = Matrix containing estimates of b
%                                          (M x 1 x N x iter)
%  estimates.c = Matrix containing estimates of c
%                                          (M x 1 x N x iter)
%  estimates.a_minus = Matrix containing estimates of a_minus
%                                          (M x 1 x N x iter)
%  estimates.a_plus = Matrix containing estimate of a_plus
%                                          (M x 1 x N x iter)
% param = struct containing the computed parameters
%  param.a = vector containing parameters {ak} (M x 1)
%  param.b = vector containing parameters {bk} (M x 1)
%  param.c = vector containing parameters {ck} (M x 1)
% sq_err = struct containing:
% sq_err.a = Matrix containing square errors of a
%                                          (N x 1 x iter)
% sq_err.b = Matrix containing square errors of b
%                                          (N x 1 x iter)
% sq_err.c = Matrix containing square errors of c
%                                          (N x 1 x iter)
% P_added = covariance function P found by adding
%             (P+(n-1)Psi+(n-1)(P-Psi))/n (anonymous function)
% P_computed = covariance function P found by using param.a
```

```
%                and ONS {gk} (anonymous function)
% num_fam_vec = vector of number of families between increase
%                            in sample size (N x iter)
% Psi_int = Vector containing the L^2(TxT) norm of Psi_hat-Psi
%                                                  (N x iter)
% zprime = mean of newborns in next generation
%                                          (anonymous function)
% zprime_hat = cell containing z*_hat(N x iter;
%                                          anonymous functions)
% zprime_int = matrix containg the L^2(T) norm of zprime_hat-
   ↪ zprime
%                                                    (N,iter)
% zprime_vec = vector containing values of zprime at t,
%       where t=linspace(t_range(1),t_range(2),100) (100 x 1)
% zprime_hat_vec = matrix containing values of zprime_hat at t
%  iterations are in the second dim       (N x iter x 100)
% varargout{1} = vector of number of observations in
%                                       each family (family x 1)
% varargout{2} = cell array containing the family data
%                                              (family x iter)

%%%%%Input check%%%%%%
selection=lower(selection);
X_type=lower(X_type);

if (size(t,1)>1 && size(t,2)>1)
    disp(' t must be a vector')
```

```
    return
else
    if size(t,2)>1
        t=t';
    end
end

if (size(a_plus,1)>1 && size(a_plus,2)>1)
    disp('a_plus must be a vector')
else
    if size(a_plus,2)>1
        a_plus=a_plus';
    end
end
if (size(a_minus,1)>1 && size(a_minus,2)>1)
    disp('a_minus must be a vector')
else
    if size(a_minus,2)>1
        a_minus=a_minus';
    end
end


switch X_type
    case 'sum'
        if isa(f,'function_handle')
            disp('f cannot be a function')
```

```
            return
        end
        if (size(f,1)>1&& size(f,2)>1)
            disp('f must be a vector if X is a sum')
            return
        else
            if size(f,2)>1
                f=f';
            end
        end
    case 'integral'

        if (strcmp(X_type,'integral')&& ~isa(f,...
                                    'function_handle'))
         disp('f must be a function handle if X is an integral'
            ↪ )
            return
        end


end
%%%%%%%%%%%%%%%%%%End Input Check%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%%%%%% Output Generation %%%%%%
estimates=struct('a',zeros(length(t),1,N,iter),'b',...
    zeros(length(t),1,N,iter),'c',zeros(length(t),1,...
```

```
        N,iter),'a_plus',zeros(length(t),1,N,family,...
           iter),'a_minus',zeros (length(t),1,N,iter));

param=struct('a',zeros(length(t),1),'b',zeros(...
                    length(t),1),'c',zeros(length(t),1));

num_fam_vec=zeros(N,iter);
sq_err=struct('a',zeros(N,1,iter),'b',...
                    zeros(N,1,iter),'c',zeros(N,1,iter));
Psi_hat=cell(N,iter);
Psi_int=zeros(N,iter);
zprime_int=zeros(N,iter);
Z=cell(family,iter);
G_t=cell(length(t),1);
zprime_hat=cell(N,1,iter);
zprime_hat_vec=zeros(N,iter,100);
tt=linspace(t_range(1),t_range(2),100);
%%%%%% End Output Generation %%%%%%

%%%%%% Preliminary Computation %%%%%%
disp('preliminary computation')
P0=@(s,t) G0(s,t)+E0(s,t);
Psi0=@(s,t)reln.*G0(s,t);
P0_minus=@(s,t)(1-reln).*G0(s,t)+E0(s,t);
[T,S]=meshgrid(t,t);
P0_matrix=P0(T,S);
chol_P0=chol(P0_matrix,'lower');
```

```
inv_chol_P0=chol_P0\eye(length(t));
g=ortho(t,P0);
f_minus=ortho(t,P0_minus);
[P_minus, ~ ] = compute_P( t, P0_minus,a_minus,...
                                            'P-Psi',f_minus );
P0_minus_mat=P0_minus(T,S);
chol_minus=chol(P0_minus_mat,'lower');
f_minus_temp=cell2mat(cellfun(@(x) x(t),f_minus,...
                                    'UniformOutput',false));
f_minus_mat=reshape(f_minus_temp,length(t),length(t));
%   functions are in columns

% round the inner product
Ndecimals = 12;
H = 10.^Ndecimals ;
rkhs_ip_minus = round(H*(chol_P0\f_minus_mat))/H;



%Precursory simulations indicated that P0_plus was
% converging to a single function as n-> infinity
P0_plus=@(s,t)P0(s,t)+(n.*iter.*family.*N-1).*Psi0(s,t);
f_plus=ortho(t,P0_plus);
[P_plus, ~ ] = compute_P( t, P0_plus,a_plus, ['P+(',...
                        num2str(n.*iter.*family.*N-1),')Psi'],
                          ↦ f_plus);
f_plus_temp=cell2mat(cellfun(@(x) x(t),f_plus,...
                                    'UniformOutput',false));
```

```
f_plus_mat=reshape(f_plus_temp,length(t),length(t));
%   functions are in columns

%round the inner product
Ndecimals = 12;
H = 10.^Ndecimals ;
rkhs_ip_plus = round(H*(chol_P0\f_plus_mat))/H;
%                                  <fk,gl> is the l,k entry

param.a=((rkhs_ip_plus.^2*a_plus)./(n.*iter.*family.*N)+...
     ((n.*iter.*family.*N)-1)*(rkhs_ip_minus.^2*a_minus)...
                                        ./(n.*iter.*family.*N));
[param.c,param.b]=compute_b(t,t_range,P0,g,f,param.a ,...
                         P0_matrix, selection,select_param);

[P_computed,~] = compute_P( t, P0,param.a,'P',g );

disp('computing G')
tic;
G=@(s,t)1./(n.*iter.*family.*N).*(P_plus(s,t)-...
                                             P_minus(s,t))./reln;


 for i=1:length(t)
     G_t{i}=@(s)G(t(i),s);
 end
toc;
```

```
Psi=@(s,t)((1./(n.*iter.*family.*N)).*(P_plus(s,t)-...
                                                P_minus(s,t)));

P_added=@(s,t)((1./(n.*iter.*family.*N)).*...
      (((n.*iter.*family.*N)-1).*P_minus(s,t)+P_plus(s,t)));
if strcmp(selection,'stabilizing')
    zprime=@(s) 0;
    zprime_vec=zeros(1,length(tt));
else
[zprime,zprime_vec]=compute_zprime(G_t,inv_chol_P0,param.b,tt);
end


switch generate
    case 'generate'
    switch smpl_type
    case 'rnd'

        n_smpl= poissrnd(n,family,1);
        while any(n_smpl==0)  % ensures that n_smpl is nonzero
            n_smpl=poissrnd(n,family,1);
        end

        n_smpl=floor(n_smpl./N); %increase within family
        %%%%%% Generate Data %%%%%%
tic;
for k=1:iter
```

```
disp('Generating Data')

for jj=1:family

    Z{jj,k}=gen_related(n_smpl(jj)*N,P_computed,Psi,t);



end
end
varargout{2}=Z;
toc;
%%%%%% End Generate Data %%%%%%

    case 'given'
        n_smpl=(n).*ones(1,family);
        n_smpl=floor(n_smpl./N);



    end
    case 'given'

        Z=varargin{2};
        n_smpl=varargin{1};
end
%%%%%% End Preliminary Computation %%%%%%
```

```
%%%%%% Main Simulation %%%%%%
disp('Running Main Simulation')
[T,S]=meshgrid(t,t);
for k=1:iter
    disp(['Beginning Iteration No. ',num2str(k)])
    tic;
for j=1:N
    num_fams=floor(family/N*j);
    dn=floor(d(min(n_smpl(1:num_fams)*j)));
    if dn>=length(t);
        dn=length(t);
    end



    num_fam_vec(j,k)=num_fams;
    Z_all=cell(num_fams,1);
    P_plus_hat=cell(num_fams,1);
    sum_P_plus_hat=@(s,t)0;
    num_obs=sum(n_smpl(1:num_fams)*j);
    a_minus_temp=zeros(length(t),1,num_fams);
    for i=1:num_fams
        n_fam=n_smpl(i)*j;
        Z_temp=Z{i,k};
        Z_all{i}=Z_temp(1:n_fam,:);
```

```
P0_plus=@(s,t)P0(s,t)+(n_fam-1).*Psi0(s,t);
f_plus=ortho(t,P0_plus);
P0_plus_mat=P0_plus(S,T);
chol_plus=chol(P0_plus_mat,'lower');
f_plus_temp=cell2mat(cellfun(@(x) x(t),f_plus,...
                                'UniformOutput',false));
f_plus_mat=reshape(f_plus_temp,length(t),length(t));
                        %functions are in columns

 % round inner product
Ndecimals = 12;
H = 10.^Ndecimals ;
rkhs_ip_plus = round(H*(chol_P0\f_plus_mat))/H;
           %inner product <fk,gl> is the l,k entry

Transform=transform(n_fam);
Y=Transform'*Z_all{i};
Y1=Y(1,:);
Yend=Y(2:end,:);
a_minus_temp(:,:,i)=estimate_a( Yend',chol_minus,dn);


aplus_temp=estimate_a(Y1',chol_plus,dn)-[ones(dn,1);...
                                  zeros(length(t)-dn,1)];
aplus_temp((dn+1):end)=zeros(length(t)-dn,1);
estimates.a_plus(:,:,j,i,k)=aplus_temp;
[P_plus_hat{j,k},~]=compute_P(t,P0_plus,aplus_temp,...
```

```
                                        'P_plus_hat',f_plus);
    estimates.a(:,:,j,k)=estimates.a(:,:,j,k)+...
        (rkhs_ip_plus.^2*aplus_temp);
    P_plus_temp=P_plus_hat{j,k};
    sum_P_plus_hat=@(s,t)sum_P_plus_hat(s,t)+...
                                    P_plus_temp(s,t);
    pause(.005)%allow user to stop execution (ctrl + c)
end
Z_all_mat=vertcat(Z_all{:});

estimates.a_minus(:,:,j,k)=sum(a_minus_temp,3)./(num_obs-
   ↪ num_fams)-...
                    [ones(dn,1);zeros(length(t)-dn,1)];

estimates.a_minus((dn+1):end,:,j,k)=zeros(length(t)-dn,1);

W=compute_W( (Z_all_mat)',t,f,selection,select_param,...
    num_obs,X_type );

[P_minus_hat,~]=compute_P(t,P0_minus,...
        estimates.a_minus(:,:,j,k),'P_minus_hat',f_minus);

estimates.a(:,:,j,k)=estimates.a(:,:,j,k)./num_obs+...
         (rkhs_ip_minus.^2)*(estimates.a_minus(:,:,j,k))...
                        .* (num_obs-num_fams)./(num_obs);

estimates.a((dn+1):end,:,j,k)=zeros(length(t)-dn,1);
```

```
estimates.c(:,:,j,k)=estimate_c(Z_all_mat',W,chol_P0,...
                                                num_obs,dn);

estimates.b(:,:,j,k)=estimate_b(estimates.a(:,:,j,k),...
                                    estimates.c(:,:,j,k));
%
sq_err.a(j,k)=(estimates.a(:,:,j,k)-param.a)'*...
                (estimates.a(:,:,j,k)-param.a);

sq_err.b(j,k)=(estimates.b(:,:,j,k)-param.b)'*...
                (estimates.b(:,:,j,k)-param.b);

sq_err.c(j,k)=(estimates.c(:,:,j,k)-param.c)'*...
                (estimates.c(:,:,j,k)-param.c);


Psi_hat{j,k}=@(s,t)(sum_P_plus_hat(s,t))./num_obs-...
    P_minus_hat(s,t).*num_fams./num_obs;
G_hat=@(s,t)Psi_hat{j,k}(s,t)./reln;

G_hat_t=cell(length(t),1);
for i=1:length(t)

    G_hat_t{i}=@(s)G_hat(t(i),s);

end
```

```
    [ zprime_hat{j,k},zprime_hat_vec(j,k,:)]=compute_zprime(...
                G_hat_t,inv_chol_P0,estimates.b(:,:,j,k),tt);

    disp('integrating functions')
    hat_Psi_sq=@(s,t) ((Psi_hat{j,k}(s,t)-Psi(s,t)).^2);


     Psi_int(j,k)=integral2(hat_Psi_sq,t_range(1),...
         t_range(2),t_range(1),t_range(2),'method',...
         'tiled','AbsTol',1e-5,'RelTol',1e-3);

    zprime_sq=@(x)((zprime_temp(x)-zprime(x)).^2);
     zprime_int(j,k)=integral(zprime_sq,t_range(1),...
            t_range(2),'AbsTol',1e-5,'RelTol',1e-3);

                    runtime=toc;
    disp(['integration complete: number=',num2str(j)])
    %%%%%% End Main Simulation %%%%%%

end
disp(['Iteration No. ',num2str(k),' Completed. ','Took ',...
    num2str(runtime), 'seconds'])
end

end
```

```
%%%%%%%% Auxillary Functions %%%%%%%%%


function X=gen_related(n,P,Psi,t)
%n=sample size
%P=(Anonymous function) Covariance function
%Psi=(Anonymous function) Cross-covariance function
%t = grid of time points
%X=Random correlated data, where X_i is in the ith row
 [T,S]=meshgrid(t,t);
Cov=kron(eye(n),P(T,S)-Psi(S,T))+kron(ones(n,n),Psi(T,S));
Z=mvnrnd(zeros(length(t)*n,1),Cov);
X=(reshape(Z,length(t),n))';

end

function g=ortho(t,P0)
% Orthonormalize the sections P_0t
P0_t=cell(length(t),1);
g=cell(length(t),1);
[T,S]=meshgrid(t,t);
P0_matrix=P0(S,T);
chol_P0=chol(P0_matrix,'lower');
inv_chol_P0=chol_P0\eye(length(t));
%create a cell array of sections of P0

for i=1:length(t)
```

```
    P0_t{i}=@(s)P0(t(i),s);
end

%orthonormalize P0 to calculate g
for j=1:length(t)
    g{j}=compute_lincomb(P0_t,inv_chol_P0(j,:));
end



end

function [P,P_matrix]=compute_P(t,P0,a,cov,g)
%Compute P=P0+sum ak gk*gk
[T,S]=meshgrid(t,t);
P=@(s,t)0;
for i=1:length(t)
    f=g{i};
    if a(i)==0
        continue
    else
    P=@(s,t)P(s,t)+a(i).*f(s).*f(t);
    end
end
P=@(s,t)P0(s,t)+P(s,t);

% create a matrix P(ti,tj) to generate random sample
P_matrix=P(T,S);
```

```
P_matrix=makeSymmetric(P_matrix,cov);

%check positive definiteness of P_matrix
[~,p]=chol(P_matrix);
if (p~=0)
    error([cov ' is not positive definite'])
end
end

function g=compute_lincomb(X,v)
%computes linear combinations of the form sum(v_i*X_i)
%where X is a cell array of symbolic functions
g=@(s)0;
for i=1:length(v)
    if v(i)==0
        continue
    else
    g=@(s)(g(s)+v(i).*X{i}(s));
    end
end



end

function M=makeSymmetric(M,~)
%remove rounding error from P_matrix to make P_matrix symmetric
```

```
issym=@(x) all(all(x==x.'));
i=20;
while ~issym(M)
    Ndecimals = i ;
    H = 10.^Ndecimals;
    M = round(H*M)/H;

   % M=round(M,i);
    i=i-1;
end
end

function a_hat=estimate_a(Z,chol_P0,dn)
%estimates ahat+1
%Z_i must be in columns
U=chol_P0\Z; %U_k is in the kth row
U_sq=sum(U.^2,2);
a_hat=[(U_sq(1:dn)); zeros(numel(U_sq)-dn,1)];
end

function c_hat=estimate_c(Z,W,chol_P0,n,dn)
U=chol_P0\Z; %U_k is in the kth row
U_bar=mean(U,2);
W_U=U*W./n;
w_U=W_U./mean(W);
c_hat=w_U-U_bar;
c_hat=[c_hat(1:dn); zeros(numel(U_bar)-dn,1)];
```

```
end

function [ b_hat ] = estimate_b(a_hat,c_hat)
b_hat=c_hat./(a_hat+1);

end

function T= transform(n)
T = gallery('orthog',n,4)';
return;
end

function [ W ] =compute_W( Z,t,f,selection,select_param...
                                                    ,n,X_type )
% Compute the true fitness values

switch X_type
    case 'integral'
        X=zeros(n,1);
        f_vec=f(t);
        for i=1:n
            X(i)=trapz(t, Z(:,i).*f_vec);
            %approximates the integral int z_t f(t)dt
            %                       by trapezoidal rule
        end
    case 'sum'
```

```
        X=Z'*f;

end

switch selection
    case 'directional'
        W=exp(X); %Directional selection
    case 'stabilizing'
        W=exp(-(X.^2)./(2*select_param.^2));
    case 'truncated'
        W=(X>select_param);
end

end

function [ c,b ] = compute_b( t,t_range,P0,g, f,a ,...
    P0_matrix,selection,select_param)
% calculates the true coefficients of the selection
% differential (c) and the selection gradient (b)


n_t=length(t);
%compute the integral of g_i*f
int_gf=zeros(n_t,1);
digits(25)
for i=1:n_t
    gi=g{i};
```

```
    int_gf(i)=integral(@(x)f(x).*gi(x),t_range(1),t_range(2)) ;

end
%%%% Compute {g_i(tj)} %%%%
g_t=zeros(n_t,n_t);
for i=1:n_t
    f=g{i};
    for j=1:n_t
        g_t(i,j)=f(t(j));
    end

end

int_P0f=zeros(n_t,1);
for j=1:n_t
    int_P0f(j)=integral(@(s)P0(t(j),s).*f(s),t_range(1),...
                                                t_range(2));
end

int_Pf=int_P0f+g_t'*(a.*int_gf);
chol_P0=chol(P0_matrix,'lower');

muX=0;


varX=integral2(@(s,t)P0(s,t).*f(s).*f(t),t_range(1),...
    t_range(2), t_range(1),t_range(2),'method','tiled'...
```

```
                                ,'AbsTol',1e-6,'RelTol',1e-4);
switch selection
    case 'directional'

        c=chol_P0\int_Pf;

    case 'stabilizing'

        c=zeros(length(t),1);
    case 'truncated'

        c=exp(-.5.*((select_param).^2)./varX)./...
            (sqrt(2.*pi.*varX).*normcdf(select_param,muX,...
                    sqrt(varX),'upper')).*(chol_P0\int_Pf);

end



b=c./(1+a);
display(' b has been computed')
end
function [zprime,zprime_vec]=compute_zprime(G_t,inv_chol_P0,b,t
   ↪ )
%compute/estimate zbardagger
%G_t = cell array of anonymous functions of sections of the
   ↪ genetic covariance function
%b = coefficients of selection gradient
```

```
%zbar = mean process zbar(t) - anonymous function
n_t=size(G_t,1);
gamma_G=cell(n_t,1);
gamma_G_vec=zeros(n_t,length(t));
if size(t,1)>1
    t=t';
end
 for j=1:n_t
gamma_G{j}=compute_lincomb(G_t,inv_chol_P0(j,:));
gamma_G_vec(j,:)=G_t{j}(t);
 end
zprime_vec=b'*inv_chol_P0*gamma_G_vec;
zprime=compute_lincomb(gamma_G,b);



end
```

Curriculum Vitae

Name

Tyler J. Baur

Education

Ph.D., Mathematics, University of Wisconsin-Milwaukee, December 2016

M.S., Mathematics, University of Wisconsin-Milwaukee, August 2011

B.S., Mathematics, University of Wisconsin-La Crosse, May 2009

Dissertation Title: Estimating the Selection Gradient of a Function-Valued Trait

Affiliations

Member, American Mathematical Society

Member, American Statistical Association