2017

Development and Properties of Kernel-based Methods for the Interpretation and Presentation of Forensic Evidence

Douglas Armstrong
South Dakota State University

Recommended Citation

Armstrong, Douglas, "Development and Properties of Kernel-based Methods for the Interpretation and Presentation of Forensic Evidence" (2017). *Theses and Dissertations*. 2175.
https://openprairie.sdstate.edu/etd/2175

DEVELOPMENT AND PROPERTIES OF KERNEL-BASED METHODS FOR THE INTERPRETATION AND PRESENTATION OF FORENSIC EVIDENCE

BY

DOUGLAS ARMSTRONG

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Computational Science and Statistics

South Dakota State University

2017

DEVELOPMENT AND PROPERTIES OF KERNEL-BASED METHODS FOR THE INTERPRETATION AND PRESENTATION OF FORENSIC EVIDENCE

DOUGLAS ARMSTRONG

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Computational Science and Statistics degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Approved:

Cedric Neumann, Ph.D. Date
Dissertation Advisor

Approved:

Kurt Cogswell, Ph.D. Date
Head, Dept. of Mathematics and Statistics

Approved:

Dean, Graduate School Date

For Darla and your unending love

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor; without his careful guidance, this dissertation would not have been completed. The long days spent doing research, coding simulations and experiments to prove or break theory and models brought the most joy to this process and motivated me when I needed it most.

I would like to thank Dr. Christopher Saunders for the support and encouraging me to pursue my Ph.D.

I would like to Dr. Kurt Cogswell for providing the support and excellent working environment for the duration of my time at SDSU.

I would like to thank my parents, who encouraged my learning and supported any scientific interests I had. The visits to museums, letting me stay up late to watch educational shows, and my father's lectures helped me get to where I am today.

I would like to thank all the graduate students I met, worked with, and became friends with over the years. It would have been lonely without you all.

I would like to thank everyone, both foreign and domestic, who made this research possible. The experience gained through collaboration is invaluable.

This dissertation research was supported in part by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice under awards numbers 2015-R2-CX-0028 and 2014-IJ-CX-K088 . The opinions and conclusions or recommendations expressed in this dissertation are those of the author and do not necessarily represent those of any supporting agency.

Thank you John Miller (George Mason University) for the collaboration and guidance.

CONTENTS

SYMBOLS AND NOTATION

Vectors are given in lowercase bold type with matrices capitalized.

$\boldsymbol{\Sigma}$	covariance matrix
$\boldsymbol{s}$	vector of all pairwise scores between sampled objects
Δ	scoring function between evidentiary objects to reduce dimension
η	natural parameter of the multivariate normal distribution
γ, σ, θ, v, d, R	tunning parameters for kernel functions
κ	a kernel function
λ_v	v^{th} eigenvalue
$\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$	inner product of objects $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$
$\mathbb{N}$	the natural numbers, the postive integers
$\mathbb{R}$	implicit or explicit mapping function
$\mathbf{V}$	set of eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_v\}$
$\mathbf{v}_v$	v^{th} eigenvector
$\mathcal{M}_d$	defense model
$\mathcal{M}_p$	prosecution model
μ or $\boldsymbol{\mu}$	univariate mean or a vector of means
ϕ	implicit or explicit mapping function
Π	probability measure over the space of ψ
ψ	set of parameters characterizing the likelihood function of $\boldsymbol{s}$
ρ, ρ_{ijkl}	correlation coefficient or correlation between the ij and kl objects
σ^2	variance term where subscript notation, such as σ_a^2, refers to the variance of the a effect
$\mathit{\Phi}$	standard normal distribution
$Corr(x, y)$	correlation between two objects, x and y
E	feature space for the evidence
e_a	set of n sets of n_o evidentiary objects sampled from a population of potential sources
e_s	set of evidence objects sampled from a known source
e_{u1}	set of trace evidence objects from unknown source 1
e_{u2}	set of trace evidence objects from unknown source 2
e_u	set of trace evidence objects with unknown source
$exp(x)$	exponential function for some argument x
F	linear feature space

H	Hilbert space
H_d	defense hypothesis
H_p	prosecution hypothesis
ij, kl	object index values for the j^{th} object from the i^{th} source and the l^{th} object from the k^{th}
ln	natural logarithm
N	number of pairwise comparisons
n	number of sources
n_o	number of objects per source
p	object dimension
pr	marginal density function
X	sampling space for evidentiary objects $\boldsymbol{x}$
$\beta_k\left(t\right)$	set of orthonormal functions for$k = 1,\ 2, ...,\ p$ basis elements over the domain t

ABSTRACT

DEVELOPMENT AND PROPERTIES OF KERNEL-BASED METHODS FOR THE INTERPRETATION AND PRESENTATION OF FORENSIC EVIDENCE

DOUGLAS ARMSTRONG

2017

The inference of the source of forensic evidence is related to model selection. Many forms of evidence can only be represented by complex, high-dimensional random vectors and cannot be assigned a likelihood structure. A common approach to circumvent this is to measure the similarity between pairs of objects composing the evidence. Such methods are ad-hoc and unstable approaches to the judicial inference process. While these methods address the dimensionality issue they also engender dependencies between scores when 2 scores have 1 object in common that are not taken into account in these models.

The model developed in this research captures the dependencies between pairwise scores from a hierarchical sample and models them in the kernel space using a linear model. Our model is flexible to accommodate any kernel satisfying basic conditions and as a result is applicable to any type of complex high-dimensional data. An important result of this work is the asymptotic multivariate normality of the scores as the data dimension increases. As a result, we can: 1) model very high-dimensional data when other methods fail; 2) determine the source of multiple samples from a single trace in one calculation. Our model can be used to address high-dimension model selection problems in different situations and we show how to use it to assign Bayes factors to forensic evidence. We will provide examples of real-life problems using data from very small particles and dust analyzed by SEM/EDX, and colors of fibers quantified by microspectrophotometry.

Part I

Introduction

1 Quantification of the weight of forensic evidence

Since the end of the 19th century, forensic scientists have been interested in quantifying the weight of forensic evidence [70]. Several attempts of defining an underlying philosophy for the forensic identification problem were made throughout the 20th century (see [34] for an example of the individualization philosophy, and [35] for a review and discussion of early philosophies). Following the work by early forerunners [70], work done by Parker [50, 51], Finkelstein and Fairley [23], Evett et. al. [20, 21, 22], Kwan [35], and Lindley [37] have laid down the basis for modern forensic evidence interpretation based on Bayesian techniques. Unfortunately, the techniques proposed by these authors, and others during the 1980s-1990s, were seldomly implemented in practice due to the lack of analytical ability to characterize the feature of interest on any given piece of evidence, and lack of computational capacity to perform the required calculations.

Despite the technical difficulties encountered by forensic scientists to develop interpretative techniques, critics of forensic science during the turn of the 21st century emphasized the critical importance of appropriately quantifying the probative value of forensic evidence [14, 40, 57, 59, 60, 61, 62]. In particular, in 2009, a committee from the National Research Council of the National Academy of Sciences spent a considerable amount of effort reviewing and describing several types of forensic evidence, including their interpretation and reporting. As a result, the committee recommended that “quantifiable measures of uncertainty in the conclusions of forensic analyses” be developed ([42], Recommendation 3, pages S-16 and 6-6). Bodies of literature and commentaries spanning more than a century have led forensic scientists and statisticians to consider the issue of the evaluation of the weight of different types of forensic evidence[1], and have resulted in a proliferation of various ad-hoc algorithms for the quantification of the value of forensic evidence.

Following early work performed in the U.K. on the interpretation of blood type comparisons and the measurement of refractive indices from glass [21, 22, 37], many scholars support the use of a special class of statistics, called the Bayes factor (BF), as a reasonable approach for the quantification of the weight of forensic evidence in the context of a case [2]. In its most basic form, a BF is a statistic that compares the marginal likelihood of observing the evidence (usually consisting of trace and control samples, but not necessarily) under two competing propositions: the first proposition (traditionally named the prosecution proposition or H_p) normally states that the trace and the control samples originate from the same source[2], while the alternative proposition (traditionally named the defense proposition or H_d) commonly states that the trace sample does not originate from the same source as the control sample, but from a different source within a relevant population of potential sources (which may or may not be specified) [2]. To calculate a BF, the stochastic manner in which samples are generated by any of the sources considered under H_p and H_d need to be characterized by probability models. The form of these models depends on what is known or assumed under H_p and H_d.

Two main trends for the rigorous calculation of BFs have been proposed in the literature: (1) Plugin-Likelihood ratios (LR) or Neyman-Pearson type LRs (NP-LR) [10, 71, Ch. 16], which typically 'plugs-in' an estimate of the parameters of the prosecution and defense models into assumed likelihood structures, and quantify the

[1] I.J. Good [29] defines the "weight of the evidence" as the base 10 logarithm of the Bayes Factor; however for the purpose of this document, the term "weight of evidence" may be interchanged with Bayes Factor, Likelihood Ratio, or any of their approximations and monotonic transform.

[2] See Kwan [35] for a review of the different definitions of "source" in forensic science. In this research project, the "source"is considered as being a particular object or person.

weight of the evidence as if the estimated parameters are actually known with certainty; and (2) Bayes factors (BF) which includes probability measures used to characterize a scientist's belief about the nature of the likelihood structure under the two competing models [9]. Methods for calculating BFs have been proposed for the quantification of blood type and DNA evidence since the 1980s, and similar methods have been proposed for glass, fingerprint, voice recordings, and other types of forensic evidence. A significant difference between BFs proposed for blood typing and DNA profiling, and the other types of evidence, is the ability to define a likelihood structure and the parameter space of the two competing models in the BF. Indeed, to calculate a BF, the likelihood structure of the models need to be characterized, which, depending on evidence types, can be more or less difficult [30, 63]. While BFs for blood typing and DNA profiling can typically be solved analytically based on population genetics, the likelihood structure of the BFs for the other types of evidence need to be approximated from data [46]. Any of the BFs for the non-biological evidence types is necessarily calculated using an ad-hoc algorithm [30, 35, 46, 63], including most of the recently published BF developments. Unfortunately, likelihood structures become intractable when the evidence forms are characterized using high dimensional variables (such as fingerprint, tool mark, questioned document, spectrogram, chromatograms, and some of the newer quantifications of DNA evidence) [65].

Instead of attempting to quantify the probative value of forensic evidence in its natural feature space, it is possible to consider reducing the dimensionality of the evidence forms. Unfortunately, and contrary to DNA evidence, pattern evidence cannot be characterized by an easily definable and quantifiable set of features. Indeed, while DNA profiles can be described using allele sizes at given loci and their concentration, which are easily measurable, pattern evidence can typically only be represented by high-dimensional and heterogeneous random vectors. This encourages the use of data reduction techniques which, if used recklessly, may lead to gross misrepresentations of the weight of evidence [3].

In Part I, Chapter 2 reviewed score-based Bayes factors. We identified several problems with the use of this methodology to quantify the weight of forensic evidence. These problems are what we corrected with the methodology developed in this research. Chapter 3 introduced the kernel-based methodology we developed in order to correct the issues identified in Chapter 2. We reviewed two developments of the kernel based method that inspired this work. In Chapter 4 we introduced properties of kernels and how they're used in practice. We provided a toy example illustrating how kernel methods are used for dimension reduction.

In Part II, Chapter 5 covered the development of our kernel-based model for a vector of pairwise scores. This is the bulk of this research and includes the final kernel-based model, parameter sampling algorithm for calculating the kernel Bayes factor, and the asymptotic properties of the distribution of the vector of pairwise scores. We tested the asymptotic properties of the score vector with empirical simulations in Chapter 6. In Chapter 7, we reviewed an early application of the kernel-based method to a set of very small particles and then applied the resulting kernel-based Bayes factor to classify blue cotton fibers based on microspectrophotometry.

In Part III, Chapter 8 discussed the kernel-based Bayes factor and its satisfaction of the problems identified in Chapter 2. Chapter 9 concluded the project with contributions to the field, potential for future research, and the flexibility of the model.

2 Score-based Bayes factor (SBF)

A recent technique for summarizing pattern data has emerged, which consists in measuring the level of (dis)similarity between two patterns, and to encompass this level into a similarity score variable [1, 3, 5, 6, 8, 11, 16, 18, 19, 27, 28, 43, 45, 58, 69]. The change in the nature of the forensic evidence resulting from the use of similarity scores in BFs is currently not well understood in a rigorous statistical sense (see [3, 30, 47]). This change of nature and its effect are illustrated hereafter.

Consider the setting where e_u and e_s are some observations respectively made on a single trace and a single control sample. We are interested in quantifying the amount of support that the evidence provides to decide between the two following propositions:

H_p : e_u has been generated by the known donor of e_s

H_d : e_u has not been generated by the donor of e_s but by a person randomly selected from a population of potential sources.

The BF is defined, for a marginal density function pr, as:

$$BF = \frac{pr\left(e_u|H_p\right)}{pr\left(e_u|H_d\right)} \tag{2.1}$$

As mentioned before, when the likelihood structures of this model are entirely defined, the BF can be calculated. However, this is rarely the case with high-dimensional pattern evidence and similarity scores are used in place of the evidence. Let Δ be a scoring function that maps from the feature space of the evidence E to the real line, $\Delta : E \times E \to \mathbb{R}$, for any pair of objects that have the same structure as e_u and e_s. Using this scoring function, it is possible to define a “score-based” BF (SBF) as:

$$SBF = \frac{pr\left(\Delta\left(e_u,\, e_s\right)|H_p\right)}{pr\left(\Delta\left(e_u,\, e_s\right)|H_d\right) \tag{2.2}$$

When comparing the SBF and the BF, we note that the SBF is only concerned with the distribution of scores under the two alternative propositions, while the BF is concerned with the probability distributions of the evidence in its natural feature space. We see that the complexity of the BF depends on the dimension of the feature space, while the complexity of the SBF is controlled by the scoring function. The power of the concept of the SBF lives in the fact that the design of the scoring function can be controlled by the scientist.

However, there are several problems with the SBF:

1. The dimension of the likelihood function of the SBF depends only on the scoring function and not on the number of trace and control samples that may have been observed e.g. it can only handle a single score between a single trace and a single control versus a vector of scores between multiple traces suspected of originating from the same source and control samples from that source;

2. The use of the scoring function to measure the pairwise similarity between a set of objects induces dependencies between the resulting scores. These dependencies are not accounted for in SBF algorithms proposed in literature. This poses two problems:

    (a) SBF techniques assume independent scores and the resulting likelihood structures are wrong. Multiple SBF likelihood structures have been proposed in the literature, all providing different values for the weight of a given score [3, 30, 64];

    (b) If the scores were independent, the resulting weights of evidence could be easily combined. However, since they are not independent they cannot be combined and as such there does not exist a way to combine the evidence;

3. SBFs are extremely dependent on the choice of the control material from the suspected source (the observed e_s), which may be a problem if there exists a large variability between multiple control samples from a given source (e.g., handwriting, tool marks, footwear impressions);

4. None of the SBFs satisfy basic properties shown by the BF. Take for example the coherency principle, given in definition 2.1;

Definition 2.1. Coherency principle:

Given a fixed knowledge base, the evidence can only support one of two mutually exclusive propositions (unless the weight of evidence is 1).

To show that SBFs do not follow the Coherency principle, let $H_A,\ H_B$ be two mutually exclusive propositions for a population of 2 sources $A,\ B$ with a set of observations made on a trace e_u and control objects $e_A,\ e_B$.

$$H_A:\ e_u\ \textit{has been generated by the known donor of}\ e_A$$

$$H_B:\ e_u\ \textit{has been generated by the known donor of}\ e_B$$

The coherency principle imposes that the numerical method can only support one of the propositions, unless the value returned by the method is 1. In particular, the numerical method should not be influenced by which proposition is considered first. Mathematically, this property is equivalent to say that:

$$BF_{A,B} \equiv \frac{pr(e_u|H_A)}{pr(e_u|H_B)} \equiv \frac{1}{BF_{B,A}} \equiv \frac{1}{\frac{pr(e_u|H_B)}{pr(e_u|H_A)}}$$

If the single alternative source's and suspect's roles are interchanged with respect to H_A and H_B, the resulting BF is the inverse of the original BF. Now consider an

SBF approach, first with the suspected source of e_u being A. The BF, with respect to H_A, is calculated using a suspect-anchored SBF (given in Appendix 10.5.1, equation 10)

$$BF_{H_A} = \frac{pr(e_u|H_A)}{pr(e_u|H_B)} \approx \frac{pr\left(\Delta\left(e_A,\, e_u\right),\, A|H_A\right)}{pr\left(\Delta\left(e_A,\, e_u\right),\, A|H_B\right)}$$

Now, consider the suspected source of e_u is B,

$$BF_{H_B} = \frac{pr(e_u|H_B)}{pr(e_u|H_A)} \approx \frac{pr\left(\Delta\left(e_B,\, e_u\right),\, B|H_B\right)}{pr\left(\Delta\left(e_B,\, e_u\right),\, B|H_A\right)} \neq \frac{1}{\frac{pr(\Delta(e_A, e_u), A|H_A)}{pr(\Delta(e_A, e_u), A|H_B)}}$$

this implies that for a fixed pair of propositions and a fixed knowledge base, the observations made on the trace may support both hypotheses at the same time, depending on which sets of evidence are considered for the scoring function of the method. Further explanation and examples of this are given in a draft manuscript found in Appendix 10.5.1. It can be shown that the SBF for a given pair of e_u and e_s genuinely originating from the same source may grossly over- or underestimate the BF for that evidence and that there is no systematic bias[3] that could help predict the lack of convergence (Figures 2.1-2.4, bottom middle and right plots). It can also be shown that the inaccuracy of the SBF approximation of the BF increases as the rarity of the feature of the putative source increases (see Figures 2.2 and 2.4). This inaccuracy was also observed for pairs of e_u and e_s that originated from different sources. In the latter case, it was observed that SBFs favored the prosecution hypothesis when compared with the BF (with the exception of putative sources with rarer characteristics). Additionally, the rate of misleading evidence in favor of the prosecution was higher with SBFs than with BFs. Overall, the result of our experiment showed that some of the proposed SBFs at best converge to the BF in very specific situations, specifically when the within-source variability is much

[3] If there was a way to measure the rarity of the trace, we could predict an under-estimation of the BF when the trace is rare and an over-estimation when the trace was common. However, if we could measure the rarity of the trace in the first place, the SBF would be unnecessary.

smaller than the between-source variability, and at worst that it is consistently biased in favor of the prosecution proposition when the putative source is genuinely not the source of the trace. Please see Appendix 10.5.1 for more details of the development of figures 2.1-2.4.

Figure 2.1: From top left to bottom right: Empirical convergence of Common Source BF, Lindley BF, General SS BF, Evett BF, Suspect anchored SBF, and Non-anchored source SBF for source material with common characteristics $\mu_x = \mu = 1.5182$ and low between to within variance ratio $\frac{\tau^2}{\sigma^2} = 10$

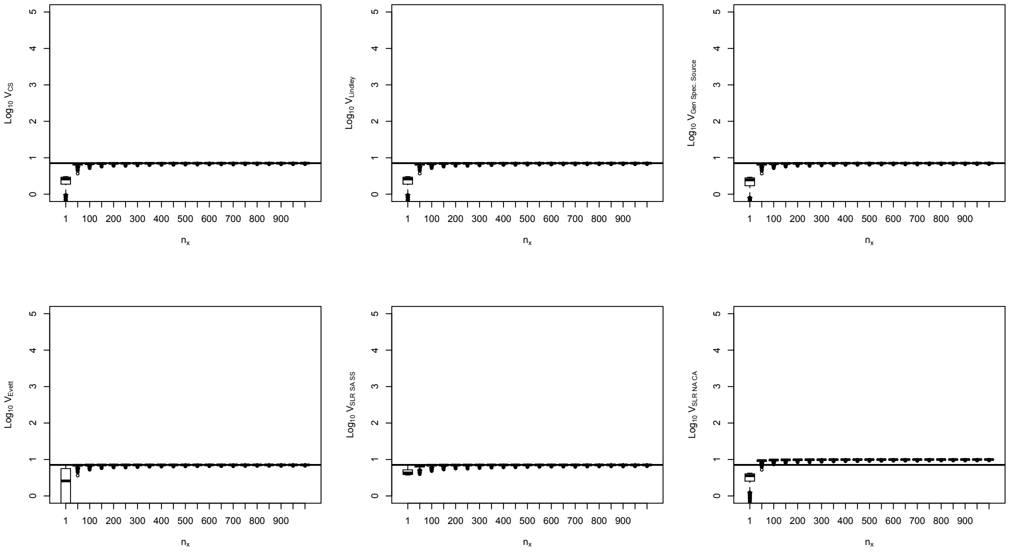


Figure 2.2: From top left to bottom right: Empirical convergence of Common Source BF, Lindley BF, General SS BF, Evett BF, Suspect anchored SBF, and Non-anchored source SBF for source material with rare characteristics $\mu_x = \mu = 1.5302$ and low between to within variance ratio $\frac{\tau^2}{\sigma^2} = 10$

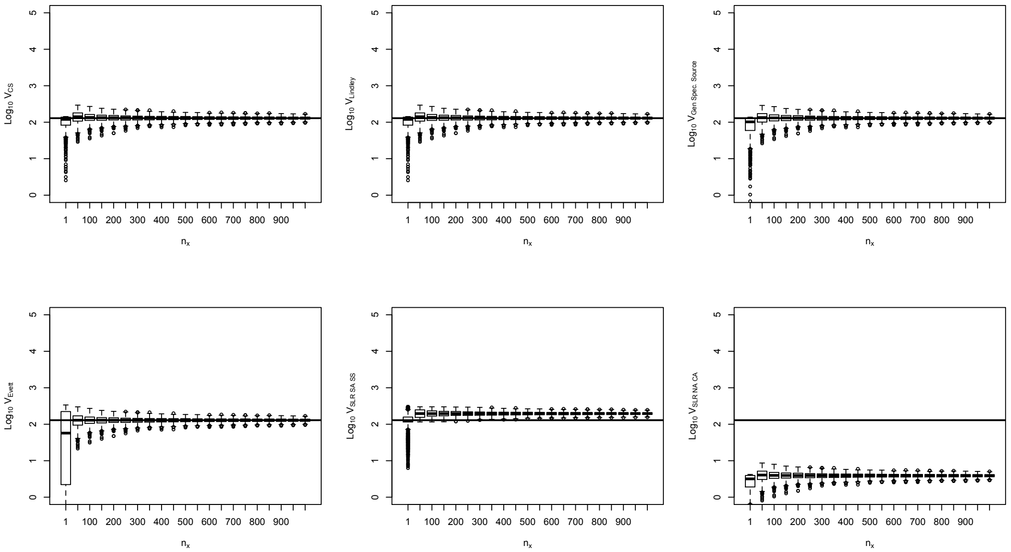


Figure 2.3: From top left to bottom right: Empirical convergence of Common Source BF, Lindley BF, General SS BF, Evett BF, Suspect anchored SBF, and Non-anchored source SBF for source material with common characteristics $\mu_x = \mu = 1.5182$ and high between to within variance ratio $\frac{\tau^2}{\sigma^2} = 10,000$

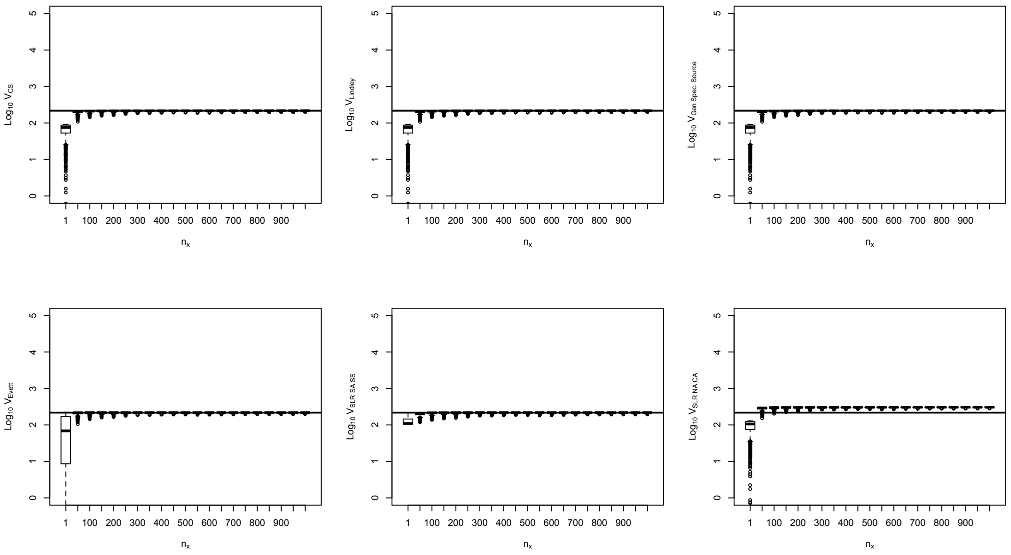
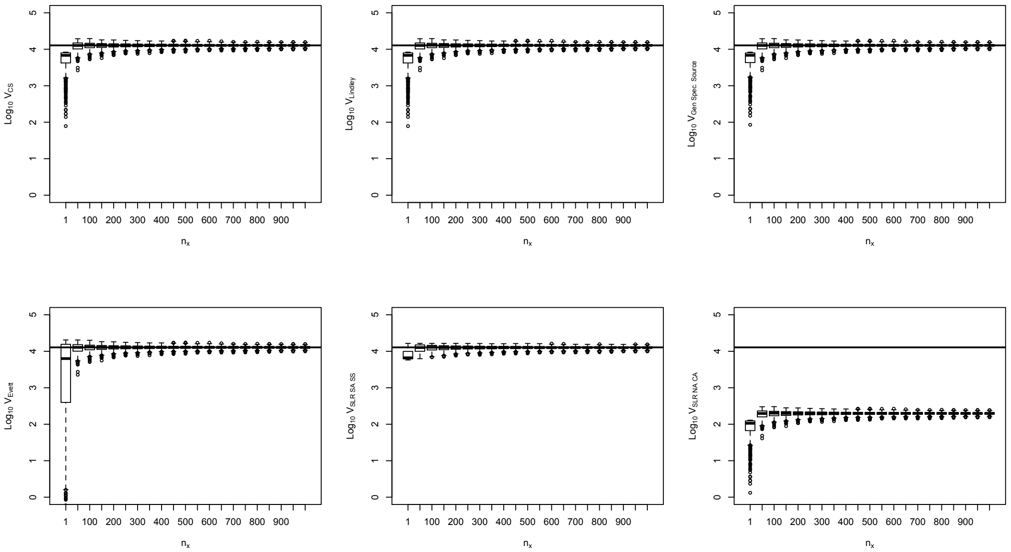


Figure 2.4: From top left to bottom right: Empirical convergence of Common Source BF, Lindley BF, General SS BF, Evett BF, Suspect anchored SBF, and Non-anchored source SBF for source material with common characteristics $\mu_x = \mu = 1.5302$ and high between to within variance ratio $\frac{\tau^2}{\sigma^2} = 10,000$



This lack of convergence likely results from the loss of information when using the scoring function: information is not only lost during the dimension reduction for a given pair of objects, but also from the creation of dependencies between all considered objects (while the objects are assumed i.i.d. in their natural feature space, the scores resulting from all pairwise comparisons are not). The example is an illustration, using a simple situation with an ideal scoring function, of a problem that is potentially magnified when 1) non-ideal scoring functions are used for complex evidence forms (such as AFIS scores for fingerprints, or IBIS scores for firearm/toolmarks) and 2) when samples are used instead of known distributions.

In summary:

1. SBFs do not have the same theoretical properties as BF;
2. SBF values do not necessarily converge with BF values for the same piece of evidence;

3. SBF values should not be misconstrued, intentionally or not, as being equivalent or relative to the weight of forensic evidence;

4. SBFs calculated for pairs of objects that do not originate from the same source have the potential to grossly overestimate the weight of the evidence against an innocent defendant; These points are already problematic when considering a single piece of evidence against a defendant, but they make it impossible to calculate coherently the combined weight of multiple pieces of evidence without taking the risk to be prejudicial to the defendant.

Some researchers have made the argument that SBFs may be “properly calibrated” [41, 53] as to minimize the rates of misleading evidence in favor of the hypothesis of common source. It is not our purpose to claim or show that SBF may not be helpful; however, the lack of convergence of SBFs and BFs limits the field of application of SBF to Bayes classifier [33] for calculating posterior probability of source in computer science, engineering, or biometric contexts.

Given the legal and scientific pressure on the forensic community to develop models to quantify the objective weight of forensic evidence, and the number of “score-based” systems currently developed and presented in the literature [1, 3, 5, 6, 8, 11, 16, 18, 19, 27, 28, 43, 45, 58, 69], this lack of convergence can be considered critical for the fair and balanced interpretation and presentation of forensic evidence by the criminal justice system in the foreseeable future.

3 Aim of the project - development of a kernel-based Bayes factor

Forensic evidence is often characterized by high-dimension random vectors containing different variable types (e.g. categorical for minutiae types, counts of particle, continuous measurements, etc.). Modeling these vectors in the natural feature space would be difficult if not impossible. This situation impedes the use of commonly advocated Bayes factor for quantifying the probative value of the evidence. The aim of the project is to propose an algorithm to quantify the weight of forensic evidence that takes advantage of the data simplification power of similarity measures while accounting for the dependencies between pairwise scores calculated between multiple objects. The problem is as follows.

Let the evidence sets $e_{u1},\ e_{u2}$ be some observations made on multiple objects sampled from two traces of forensic interest, e_a be a set of observations made on i.i.d. objects sampled from a population of potential sources (multiple objects per source), and $\boldsymbol{s}$ be a vector of all $N = \binom{nn_o}{2}$ pairwise scores calculated between n sources and n_o objects per source in $e_{u1},\ e_{u2},\ e_a$. We define the sets $e_{u1},\ e_{u2},\ e_a$ as being simple random samples from a common sample space X and

$$e_{u1} = \{\boldsymbol{x}_{u11},\ \boldsymbol{x}_{u12}, ..., \boldsymbol{x}_{u1n_o}\}$$

$$e_{u2} = \{\boldsymbol{x}_{u21},\ \boldsymbol{x}_{u22}, ..., \boldsymbol{x}_{u2n_o}\}$$

$$e_a = \{e_{a1},\ e_{a2}, ..., e_{an}\}$$

where for $i = 1, ..., n$

$$e_{ai} = \{\boldsymbol{x}_{ai1},\ \boldsymbol{x}_{ai2}, ..., \boldsymbol{x}_{ain_o}\}$$

and, in general $\boldsymbol{x} \in X$.

We are interested in quantifying the amount of support that the evidence, represented by the random vector $\boldsymbol{s}$, provides to decide between the two following models of simple random samples (SRS) of the evidence $e_{u1},\ e_{u2}$:

$\mathcal{M}_p$: $e_{u1},\ e_{u2}$ are SRS from a common randomly selected source in the population

$\mathcal{M}_d$: $e_{u1},\ e_{u2}$ are SRS from 2 different random sources in the population of sources

We wish to evaluate the following KBF:

$$KBF = \frac{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_p\right) d\Pi\left(\psi\right)}{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_d\right) d\Pi\left(\psi\right)}$$

where ψ is a set of parameters characterizing the likelihood functions of $\boldsymbol{s}$ under both models and Π is a probability measure over the parameter space of ψ. We note that the likelihood functions of $\boldsymbol{s}$ are high-dimensional, however they account for the dependencies between multiple pairwise scores sharing one common object and can account for multiple objects sampled from a given trace. Note that the KBF of interest in this project is a common source one [48].

In practice, the project consists in proposing a likelihood structure for $\boldsymbol{s}$ and the algorithm necessary to estimate the marginal distributions of $\boldsymbol{s}$ under both models.

3.1 Previous work to the kernel-based Bayes factor (KBF)

In order to take advantage of the data simplification power of similarity measures, while accounting for the dependencies between the objects in the population of potential sources created by the scoring function, Lock & Morris [39] and Saunders et al. [24] have considered “kernel-based” methods, which are common in pattern classification [31]. It is possible to consider the scoring function as a kernel. A kernel function is a special function used to compare two objects defined by vectors $\boldsymbol{x}_i,\ \boldsymbol{x}_j$ and project that comparison onto $\mathbb{R}$. Note that the notation for

$\boldsymbol{x}_i,\ \boldsymbol{x}_j$ is used only for this section since the developments reviewed here considered objects sampled from a single source instead of the multi-source situation we considered.

A kernel uses an explicit or implicit mapping function, ϕ, that projects objects from a complex non-linear input space, X, into a linear feature space, F, before comparing the objects. This projection allows for the use of linear techniques (i.e. regression, PCA, LDA) on otherwise non-linear data [25, 67]. A kernel κ is defined for any two vectors $\boldsymbol{x}_i,\ \boldsymbol{x}_j \in X$:

Definition 3.1. Kernel

A kernel function, κ, may be defined such that for all $\boldsymbol{x}_i,\ \boldsymbol{x}_j \in X,\ (i,\, j \in \mathbb{N}):$

$$\kappa :\ X \times X \rightarrow \mathbb{R}\ s.t.\ \kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \kappa\left(\boldsymbol{x}_j, \boldsymbol{x}_i\right)$$

or with respect to machine learning for an explicit mapping function
ϕ:$\boldsymbol{x}$ $\in$X$\rightarrow$ $\phi\left(\boldsymbol{x}\right)$ $\in$F:

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left\langle \phi\left(\mathrm{x}_i\right),\, \phi\left(\mathrm{x}_j\right)\right\rangle$$

The kernel function will be revisited in more detail in Chapter 2. For our purposes, a kernel function is used interchangeably with a scoring function, and the resulting projection from the kernel may be referred to as a score.

The main difference between SBFs and the method developed in this research program is that the KBFs simultaneously considers the pairwise comparisons between all objects in the input space. This enables the KBF method to:

1. Account for the structure of the relationships between the scores created by the kernel;

2. Take into account situations where the evidence consists of multiple trace

and/or control objects with or without a balanced number of samples[4];

3. Take into account situations where the variability between different control objects from a given source is not null [25, 19];

4. Allow for a solution when only one sample per alternative object is available in a similar fashion to SBFs.

The choice of kernel function, κ, among the many classes of kernels available, [25, 56, 67] can be user-specified to target a specific problem; hence it can be customized to any particular complex evidence form.

At least 3 different methods based on kernels have been investigated in forensic science. The method proposed by Neumann et al. [46] to quantify the weight of fingerprint evidence can be considered as an initial step towards a kernel based method. While they do not compute all cross-comparisons between the fingerprints considered in their study, they used a scoring function to project them on a new space which origin is the considered latent print.

Lock and Morris in 2013 [39] developed and used a two-stage approach in order to analyze forensic toolmark data for the significance of the angle left behind by tools, and infer the source (i.e. tools) of trace toolmarks. A hypothesis test was used to compare sets of scores from trace-control comparisons and control-control comparisons. The samples to be compared were digitized marks from multiple tools with 4 resamples per tool to get an estimate of within-tool variance. The authors used Chumbley's algorithm [13] as a kernel function to compare pairs of digitized tool marks,$\left(\boldsymbol{y}_i,\, \boldsymbol{y}_j\right)$ and obtain correlation scores y_{ij}, which were then calibrated to follow an approximate normal distribution using scores.

The resulting sets of data were organized into the sets $\mathbf{y}_{0j},\ \mathbf{y}_{ij}$ where $\mathbf{y}_{0j}$ is a vector of calibrated scores between the trace and j^{th} control and $\mathbf{y}_{ij}$ is the vector of

[4] As long as the parameters of the model are estimated using a balanced training sample.

scores from all control-control comparisons $ij \neq 0$. In order to account for dependency between $\mathbf{y}_{0j},\ \mathbf{y}_{ij}$, Lock and Morris defined the $N* \times N*$, where $N* = n + \begin{pmatrix} n \\ 2 \end{pmatrix}$, correlation matrix $\mathbf{R}$ where the entries are defined as:

$$Corr\left(y_{ij},\, y_{kl}\right) = \begin{cases} 0, & if\ i \neq k,\, i \neq l,\, j \neq k,\, j \neq l \\ \rho, & if\ i = k\ or\ i = l\ or\ j = k\ or\ j = l,\ and\ \ (i,\, j) \neq (k,\, l) \\ 1, & if\ i = k\ and\ j = l \end{cases}$$

and $\rho \in [0,\, 0.5)$ to ensure nonnegative variances in their calculations. The complete model for a dataset with $n = 4$ controls and a single trace under the hypothesis that the trace and control originate from the same source, is calculated as:

$$\begin{aligned} \mathbf{y} &= \left(y_{01},\, y_{02},\, y_{03},\, y_{04},\, y_{12},\, y_{13},\, y_{14},\, y_{23},\, y_{24},\, y_{34}\right)' \\ \boldsymbol{\mu} &= \left(\mu_0,\, \mu_0,\, \mu_0,\, \mu_0,\, \mu_1,\, \mu_1,\, \mu_1,\, \mu_1,\, \mu_1,\, \mu_1\right)' \end{aligned}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho & \rho & 0 & 0 & 0 \\ \rho & 1 & \rho & \rho & \rho & 0 & 0 & \rho & \rho & 0 \\ \rho & \rho & 1 & \rho & 0 & \rho & 0 & \rho & 0 & \rho \\ \rho & \rho & \rho & 1 & 0 & 0 & \rho & 0 & \rho & \rho \\ \rho & \rho & 0 & 0 & 1 & \rho & \rho & \rho & \rho & 0 \\ \rho & 0 & \rho & 0 & \rho & 1 & \rho & \rho & 0 & \rho \\ \rho & 0 & 0 & \rho & \rho & \rho & 1 & 0 & \rho & \rho \\ 0 & \rho & \rho & 0 & \rho & \rho & 0 & 1 & \rho & \rho \\ 0 & \rho & 0 & \rho & \rho & 0 & \rho & \rho & 1 & \rho \\ 0 & 0 & \rho & \rho & 0 & \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

Because the scores were calibrated to be approximately normal, the authors assumed $\mathbf{y} \sim MVN\left(\boldsymbol{\mu},\, \sigma^2\mathbf{R}\right)$. A formal hypothesis test for whether or not a pair of tool marks was made by the same tool was set up:

$$\begin{aligned} H_0 : \mu_0 &= \mu_1 \\ H_A : \mu_0 &< \mu_1 \end{aligned}$$

A likelihood ratio test was used to compare the null and alternative models by calculating the likelihood ratio statistic, λ, (equation 3.1, $l\left(\cdot\right)$ is the normal likelihood function) and then calculating its p-value by using the asymptotic distribution of $-2ln\left(\lambda\right)$, which is chi-bar [12] under H_0.

$$\lambda = \frac{l\left(\hat{\mu},\ \hat{\sigma}^2,\ \hat{\rho}\right)}{l\left(\hat{\mu}_0,\ \hat{\mu}_1,\ \hat{\sigma}^2,\ \hat{\rho}\right)} \tag{3.1}$$

Gantz and Saunders [24] used a linear random effects model to describe the dependency between similarity scores betwen pairs of objects. They were interested in estimating the random match probability of objects using all pairwise scores. This was later used to predict if a set of scores, $\mathbf{s}_m$, resulting from comparing a trace with multiple control objects from a single source is from the same distribution as the set of scores between control objects, $\mathbf{s}_n$, resulting from comparing the control samples to each other [4]. They defined a linear model for the score s_{ij}, given in equation 3.2:

$$s_{ij} = \theta + a_i + a_j + \varepsilon_{ij} \tag{3.2}$$

where θ is the grand mean, $a_i,\ a_j$ are i.i.d. random variables assumed to be distributed $N\left(0,\, \sigma_a^2\right)$, and ε_{ij} is assumed distributed $N\left(0,\, \sigma_e^2\right)$. The authors were interested in deriving the multivariate normal distribution of $\mathbf{s}_n$, the vector of scores resulting from all control-control comparisons. The details of this development may be found in [4].

The main takeaway from the Gantz and Saunders work is the joint model for the vector of scores $\mathbf{s} = \theta\mathbf{1}_N + \mathbf{Pa} + \mathbf{e}$, which, under normality assumptions, gives us $\boldsymbol{s} \sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ with covariance matrix $\boldsymbol{\Sigma}$ given in equation 3.3

$$\boldsymbol{\Sigma} = \sigma_a^2\mathbf{PP}^t + \sigma_\varepsilon^2\mathbf{I}_N. \tag{3.3}$$

Gantz and Saunders were able to estimate the parameters θ, σ_a^2, σ_e^2 in closed form for any kernel considered using this model. In order to predict whether a new set of scores, $\mathbf{s}_m$, comes from the same distribution that gave rise to $\mathbf{s}_n$, the conditional distribution for $\mathbf{s}_m|\mathbf{s}_n$ is obtained using [33].

Under the null hypothesis, both groups of authors were ultimately concerned with the density value associated with $f\left(y_{n+1},\, y_n|H_0\right)$; both Lock and Morris, and Gantz and Saunders, made the same assumption of the normality of the distribution of scores; however, the two models differ in their construction of the covariance matrix[5].

When comparing the covariance structures between the two models, assuming that $f_{LM}\left(y_{n+1},\, y_n|H\right) = f_{GS}\left(y_{n+1},\, y_n|H\right)$ and using the σ^2 estimate from the Lock and Morris model, we get:

$$\begin{aligned}
\sigma^2\mathbf{R} &= \sigma^2\left(\rho\mathbf{P}\mathbf{P}^t - 2\rho\mathbf{I}\right) \\
&= \rho\sigma^2\mathbf{P}\mathbf{P}^t - 2\rho\sigma^2\mathbf{I} \\
&= \sigma_a^2\mathbf{P}\mathbf{P}^t + o_e^2\mathbf{I}.
\end{aligned}$$

This would imply that $\sigma_a^2 = \rho\sigma^2$ and $o_e^2 = -2\rho\sigma^2$, which is not possible. The modeling assumption and constraints in the Lock and Morris approach are not as attractive when compared with those of the Gantz and Saunders approach. It appears that the covariance matrix in Lock and Morris is erroneously designed. Furthermore, parameters in the Gantz and Saunders approach have closed form solutions while Lock and Morris requires an optimization step to ensure $\mu_0 < \mu_1$. Finally, the proof that the assumption of normality made by Gantz and Saunders holds for data of sufficiently high dimension for most kernels used, as shown in section 5.7.

[5] I attempt to align the notation used by all authors. y_{n+1} denotes the y_{0j} sets, y_n denotes the y_{ij} sets from control-control comparisons.

Both developments by Lock and Morris, and Gantz and Saunders consider the scores in a multivariate space and the dependencies that exist between them. However, both models only consider a sample of control objects from a single source and do not account for a hierarchical sampling scheme with multiple samples from multiple sources. It is the purpose of this project to develop a kernel model, in an analogous manner to the Gantz and Saunders model, that captures the dependency structure existing between scores from hierarchically-sampled objects in order to quantify the weight of forensic evidence.

4 Kernel theory and classes

Before beginning the developmet of a kernel model, we introduce kernel methods in more detail, how they work, and the types of problems they are typically used for. Kernel methods are typically leveraged for their computational efficiency and data reduction power while providing some measure of similarity between objects. According to Shawe-Taylor and Cristiani [67], the selection of a kernel in task comes down to two properties:

1. Does the selected kernel capture the measure of similarity appropriate to the task at hand?
2. Is its evaluation significantly less computationally demanding than an explicit evaluation of the corresponding feature mapping, ϕ?

If both of these conditions are met, then that kernel is appropriate to use.

4.1 Properties of kernels

Kernel-based methods, for pattern recognition and the purposes of this research, are prevalent in the machine learning community as a means to classify objects belonging to a (often) non-linear, high dimensional space. As such, much of the terminology doesn't always line up with classic statistics terminology. The definitions found below are what we will use for the duration of this paper. There are some terms that arise in both the machine-learning and the statistics fields but hold different meanings, and these will be pointed out as they are covered.

The workhorse of the entire method is of course the kernel itself. Recall the kernel definition 3.1 for vectors $\boldsymbol{x}_i$, $\boldsymbol{x}_j \in X$.

Definition. Kernel

A kernel function, κ, may be defined such that for all $\boldsymbol{x}_i,\ \boldsymbol{x}_j \in X,\ (i,\ j \in \mathbb{N}):$

$$\kappa :\ X \times X \to \mathbb{R}\ s.t.\ \kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \kappa\left(\boldsymbol{x}_j, \boldsymbol{x}_i\right)$$

or with respect to machine learning [73] for an explicit mapping function $\phi :\ \boldsymbol{x} \in X \to \phi\left(\boldsymbol{x}\right) \in F$:

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left\langle \phi\left(\mathbf{x}_i\right),\ \phi\left(\mathbf{x}_j\right)\right\rangle$$

Often times, calculating the full inner product $\left\langle \phi\left(\mathbf{x}_i\right),\ \phi\left(\mathbf{x}_j\right)\right\rangle$ is computationally intensive as each mapping for $\phi\left(\mathbf{x}_i\right),\ \phi\left(\mathbf{x}_j\right)$ must be calculated. Instead, $\left\langle \phi\left(\mathbf{x}_i\right),\ \phi\left(\mathbf{x}_j\right)\right\rangle$ is replaced by $\kappa\left(\mathbf{x}_i,\ \boldsymbol{x}_j\right)$, bypassing the need to calculate the mapping ϕ. This replacement is referred to as the *kernel trick* in the machine learning community and can be found in use in almost all modern kernel-based techniques [31]. To show how the kernel trick is used, a toy problem is explored in two examples: Example 4.1 demonstrates how a simple explicit mapping, ϕ, works and Example 4.3 exploits the kernel trick to circumpass the explicit mapping seen in Example 4.1. This is the favorite toy problem to develop intuition of the power of kernels.

Example 4.1. Using the explicit mapping function $\phi :\ \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$

Let $\mathbf{x} \in \mathbb{R}^2$ with class labels $l \in \{1,\ 2\}$. We draw 400 independent samples from a multivariate normal distribution with $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Class labels are assigned according to the rule,

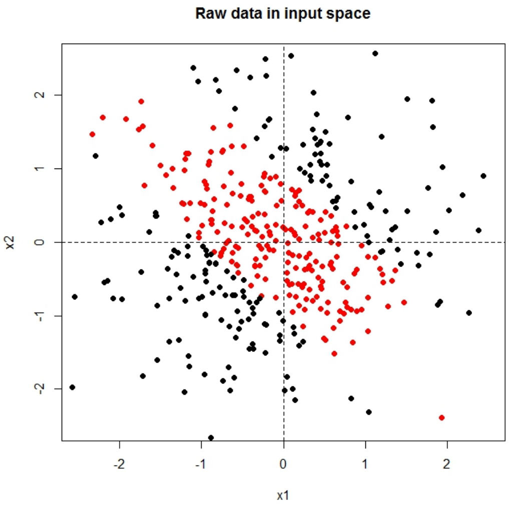
$$l = \begin{cases} 1 \quad if \quad x_1^2 + x_2^2 + 2x_1x_2 \leq 1 \\ 2 \quad if \qquad otherwise \end{cases}$$

By design, the data for this example has no linear separator in its input space, as can be seen in Figure 4.1. Often in practice, the exact form of the separator is unknown but for this toy example, a keen eye will notice a possible elliptical separator. The first five samples of the data is found in Table 4.1.

Table 4.1: Sample of data for Example 4.1

$$\mathbf{X} = \left[\begin{array}{ccc} x_1 & x_2 & l \\ -1.393 & -0.761 & 1 \\ -1.220 & 0.525 & 2 \\ -1.212 & -0.541 & 2 \\ -0.225 & 2.492 & 1 \\ 0.440 & 1.350 & 1 \\ \vdots & \vdots & \vdots \end{array}\right]$$

Figure 4.1: Scatterplot of data for Example 4.1 where red is class 1 and black is class 2



If one did not know the scheme of the class separation, it would be reasonable to assume a quadratic separator because of the elliptical boundary between the classes. Hence, the inner product $\langle \phi(\mathbf{x}_i), \phi(\boldsymbol{x}_j) \rangle$ needs to have a quadratic form, implying that ϕ could take the following form:

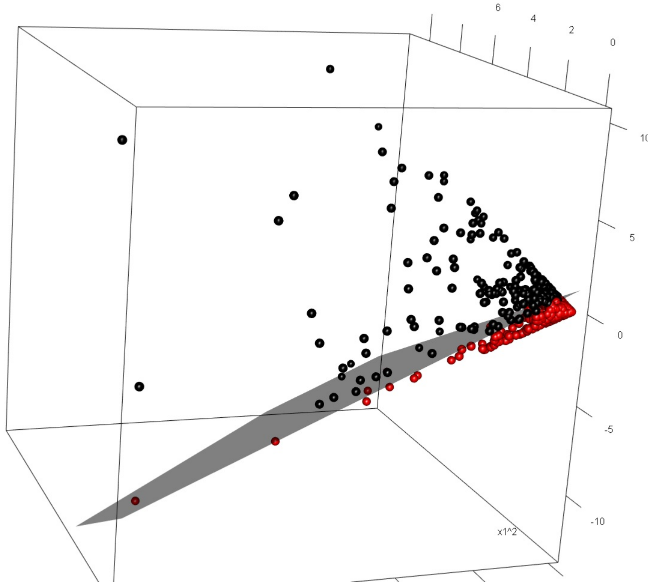
$$\begin{aligned} \phi : \mathbf{x} \in \mathbb{R}^2 &\mapsto \phi(\mathbf{x}) \in F = \mathbb{R}^3, \text{ where} \\ \phi(\boldsymbol{x}) &= \left(x_1^2,\, x_2^2,\, \sqrt{2}x_1x_2\right) \end{aligned} \tag{4.1}$$

and we have

$$\begin{aligned} \kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \langle \phi(\mathbf{x}_i),\, \phi(\boldsymbol{x}_j)\rangle &= \left\langle \left(x_{i1}^2,\, x_{i2}^2,\, \sqrt{2}x_{i1}x_{i2}\right),\, \left(x_{j1}^2,\, x_{j2}^2,\, \sqrt{2}x_{j1}x_{j2}\right)\right\rangle \\ &= x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\ &= \left(\boldsymbol{x}_i^t\boldsymbol{x}_j\right)^2 \end{aligned} \tag{4.2}$$

Applying the mapping function ϕ from equation 4.1 to the example data results in an explicit mapping into $\mathbb{R}^3$ and a linearly separable data cloud (in our toy example). By design of our example, we know the plane $1 - x_1^2 - x_2^2 - 2x_1x_2 = 0$ separates the classes in $\mathbb{R}^3$. Otherwise, without this prior knowledge a numerical method may be used to fit a best-separating plane. The separating plane can be seen in Figure 4.2. For this toy example, the mapping function has increased the number of dimensions by 1.

Figure 4.2: Data mapped into feature space, F, with linear separator



While this is a toy example of small dimension to demonstrate the basic use of an explicit mapping function in kernels, it is important to take note of the computational expense required for its use namely the storage of all the coordinates involved. If the data were of higher dimension, or an explicit mapping of higher order used, the computational resources would increase while efficiency decreases. The aforementioned kernel trick bypasses the explicit mapping step, saving on computational resources and effort in most cases [6]. Indeed, we note in equation 4.2 that the mapping doesn't have to be explicit, that the kernel is symmetrical, and that it returns a scalar.

Instead of working with data in F, when using the kernel trick, we only use the resulting inner products between all objects in the input space. In the case of our toy example, $(\boldsymbol{x}_i^t \boldsymbol{x}_j)^2$ is the resulting scalar between two objects. The resulting scores from all pairwise comparisons in the input space can be organized into a kernel[7] matrix, $\mathbf{K}$ whose ij^{th} entry $\mathbf{K}_{ij} = \kappa\left(\mathbf{x}_i,\, \mathbf{x}_j\right)$.

[6] While this is not the case in this toy example, as data grows more complex the explicit mapping ϕ is often intractable.

[7] In literature, this may be reffered to as a Gram matrix but to keep language consistent we will call it a kernel matrix as it organizes the resulting scores from kernels.

Definition 4.2. Gram (Kernel) matrix

$$\mathbf{K} = \begin{bmatrix} \kappa\left(\mathbf{x}_1,\, \mathbf{x}_1\right) & \kappa\left(\mathbf{x}_1,\, \mathbf{x}_2\right) & \cdots & \kappa\left(\mathbf{x}_1,\, \mathbf{x}_n\right) \\ \kappa\left(\mathbf{x}_2,\, \mathbf{x}_1\right) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \kappa\left(\mathbf{x}_n,\, \mathbf{x}_1\right) & \cdots & \cdots & \kappa\left(\mathbf{x}_n,\, \mathbf{x}_n\right) \end{bmatrix}$$

The kernel matrix resulting from a kernel need not be positive semi-definite. However, the kernel matrix resulting from a covariance function, a subclass of kernels, is positive semi-definite. The kernel matrix contains all information pertaining to pairwise distances between objects of the data set. There is a loss of information with the kernel matrix, namely the orientation of the original data and its axes. This is due to the property of the stationary kernel and that the resulting kernel matrix is rotationally invariant with respect to the original data. In other words, if the original data is arbitrarily rotated in its coordinates, **K** will not change.

Example 4.3. Using the kernel trick

Continuing from Example 1, **K**, resulting from the kernel trick, coupled with a Kernel Principal Component Analysis (KPCA) gives us the same separation as the explicit mapping did, with the added benefit of the KPCA suggesting the number of dimensions of the implicit mapping function required to describe the data in a linear space. First, we calculate **K** using the squared inner product as before. A sample of the top-left 5x5 corner of **K** is given in Table 4.2.

Table 4.2: 5x5 sample of **K**

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$
$\mathbf{x}_1$	6.35	2.51	2.69	2.95	8.82
$\mathbf{x}_2$	2.51	39.16	10.66	1.25	9.27
$\mathbf{x}_3$	2.69	10.66	4.07	1.29	5.29
$\mathbf{x}_4$	2.95	1.25	1.29	1.37	4.13
$\mathbf{x}_5$	8.82	9.27	5.29	4.13	14.17

The resulting 400x400 symmetric matrix of inner products can be reduced further in dimension by using KPCA (see [33, pgs 609-613] for further detail). The eigen decomposition results in 3 non-zero eigenvalues and 397 zero-valued eigenvalues, as seen in Figure 4.3, suggesting projection into a lower 3-dimensional space. Using this information, the scores for the objects in 3 dimensions are calculated and plotted. Aside from a new coordinate system due to the KPCA, the structure of the data cloud in Figure 4.4 is nearly identical to the one observed in Figure 4.2. Calculating a linear separator on the KPCA coordinates ($x_1^\star$, $x_2^\star$, $x_3^\star$) using logistic regression, we get an approximate solution of $36x_1^\star - 72.40x_2^\star + 16.17x_3^\star + 1331.23 = 0$ for a separating hyperplane.

Figure 4.3: Eigenvalues for **K** in toy Example. Note there are three non-zero eigenvalues and 397 zero-valued eigenvalues

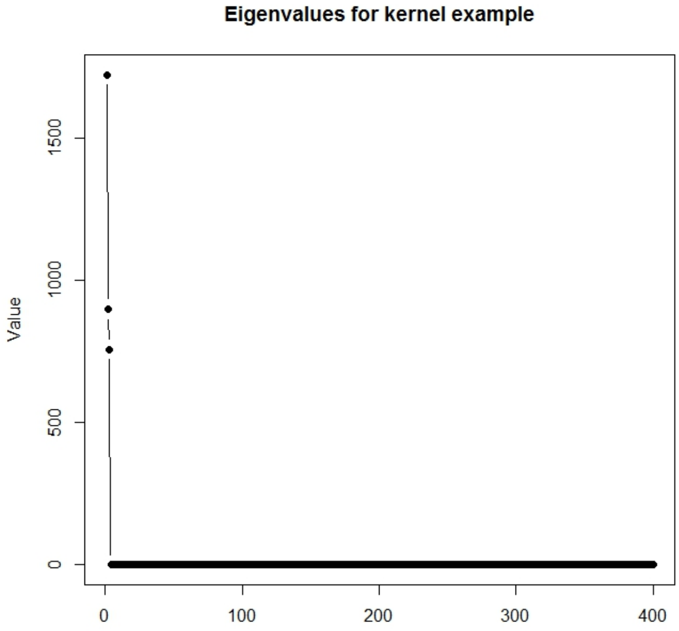
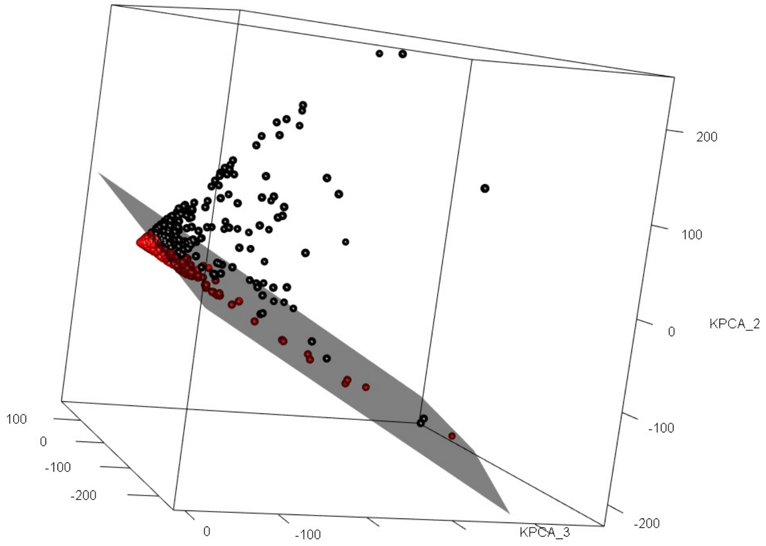


Figure 4.4: Projection using scores from KPCA in 3 dimensions



Definition 4.4. Positive semi-definite kernel

A kernel is PSD if it produces a positive semi-definite kernel matrix [56]

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \kappa\left(\mathbf{x}_i,\, \mathbf{x}_j\right) = \mathbf{a}'\mathbf{K}\mathbf{a} \geq 0$$

where $\mathbf{a}$ is any nonzero vector of scalars and $\forall$ finite n; $\mathbf{x}_1, ..., \mathbf{x}_n \in X$; $a_1, ..., a_n \in \mathbb{R}$. [67]

Definition 4.5. Hilbert Space

A Hilbert space, H, is a separable and complete inner product space. Completeness states that every Cauchy sequence $\{h_n\} \in H$ converges to an element $h \in H$. Separable states there is a countable set of elements, $h_1, ..., h_i, ...$ of H such that for all $h \in H$ and $\varepsilon > 0$, $\|h_i - h\| < \varepsilon$.

The properties of completeness and separability allow for a coordinate system to be defined in the space, H. This coordinate system will allow for a likelihood structure to be defined on the space, an important step in constructing the KBF.

4.1.1 Closure properties of kernels:

An important set of properties used to build new kernels or extend their embeddings is known as the closure properties [67].

Let κ_1 and κ_2 be kernels over $X \times X$, $a \in \mathbb{R}^+$, $f(\cdot)$a real-valued function on X, $\phi :\ X \to F$ with κ_3 a kernel over $F \times F$, and $\mathbf{B}$ a symmetric positive semi-definite $N \times N$ matrix. Then the following are kernels:

(i) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \kappa_1(\boldsymbol{x}_i,\, \boldsymbol{x}_j) + \kappa_2(\boldsymbol{x}_i,\, \boldsymbol{x}_j)$

(ii) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = a\kappa_1(\boldsymbol{x}_i,\, \boldsymbol{x}_j)$

(iii) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \kappa_1(\boldsymbol{x}_i,\, \boldsymbol{x}_j)\,\kappa_2(\boldsymbol{x}_i,\, \boldsymbol{x}_j)$

(iv) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = f(\boldsymbol{x}_i)\, f(\boldsymbol{x}_j)$

(v) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \kappa_3(\phi(\boldsymbol{x}_i)\,,\, \phi(\boldsymbol{x}_j))$

(vi) $\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \boldsymbol{x}_i^t \mathbf{B} \boldsymbol{x}_j$

Closure property (i) states that the sum of two kernels is in itself a kernel. If κ_1 and κ_2 are positive definite, then the resulting kernel will also be positive definite.

Closure property (ii) states that multiplying kernel by a scalar results in a scaled kernel. This is sometimes used for normalizing a kernel space.

Closure property (iii) states that the multiplication of two kernels, κ_1 and κ_2 , results in a kernel as well. This property is utilized in the Matérn kernel.

These three properties were utilized in application of the kernel model, as discussed in Section 7.

4.2 Classes of kernels

The flexibility of kernels comes from a combination of the closure properties and the vast number of classes of kernels available for use [31, 52, 56, 67]. Kernels may be designed to measure specific components of data and then combined to create new kernels. They can be used for continuous or discrete data and even the combination of both. There exists hundreds of kernels and as such, the major classes of kernels are given below. Additional examples of kernels can be found in [31, 56, 67].

One major distinguishing factor between classes of kernels is the treatment of the input vectors $\boldsymbol{x}_i$, $\boldsymbol{x}_j$. The two treatments are either nonstationary or stationary. Nonstationary kernels are popular in classification of objects where a natural "zero" or origin doesn't make sense. These kernels do not rely on distance between objects and the resulting diagonal elements of $\mathbf{K}$ will be non-negative. Because of this, the results of an analysis using a nonstationary kernel is unique to the data it was built on[8]. Some examples of this include binary input data or greyscale images, normalized to $[-1, 1]$[56]. Conversely, stationary kernels rely upon a difference between input vectors and are used when the data has some form of a natural zero or origin. As a result, the diagonal elements of $\mathbf{K}$ resulting from the use of a stationary kernel will all equal the same value, often 0 or 1. In our experience, forensic data has a stationary form and is exploited in SBF methods already.

4.2.1 Stationary kernels

Stationary kernels are functions of $\boldsymbol{d} = \boldsymbol{x}_i - \boldsymbol{x}_j$, the difference between vectors $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ in the input space. By design, the output of stationary kernels has an intuitive interpretation: the closer a score is to the origin (e.g. 0 or 1), the more similar those objects are. There are two types of stationary kernels; anisotopic, which measures magnitude and direction, and isotropic, which measures magnitude only.

Definition 4.6. Anisotropic stationary kernel

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \kappa(\mathbf{x}_i - \mathbf{x}_j)$$

[8] Unique in the sense of the scale of the input data.

A class of kernels which emphasizes magnitude and direction of a lag vector between two objects in the input space.

Definition 4.7. Isotropic stationary kernel

$$\kappa(\boldsymbol{x}_i,\, \boldsymbol{x}_j) = \kappa(\|\mathbf{x}_i - \mathbf{x}_j\|)$$

A kernel which uses a normed vector, $\|\mathbf{d}\| = \|\mathbf{x}_i - \mathbf{x}_j\|$, resulting in a measure of magnitude only.

Below are the major families of stationary kernels used in practice.

Definition 4.8. The γ-exponential kernel family
The γ-exponential family of kernels defines any kernel with the for

$$\kappa\left(\boldsymbol{x}_i,\, \boldsymbol{x}_j\right) = exp\left(-\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma}\right)^{\gamma}\right),\; 0 < \gamma \leq 2,\; \sigma > 0. \tag{4.3}$$

Within the γ-exponential family is the popular Gaussian kernel where $\gamma = 2$ and $\sigma > 0$.

Definition 4.9. The Gaussian kernel for $\gamma = 2$ and $\sigma > 0$.

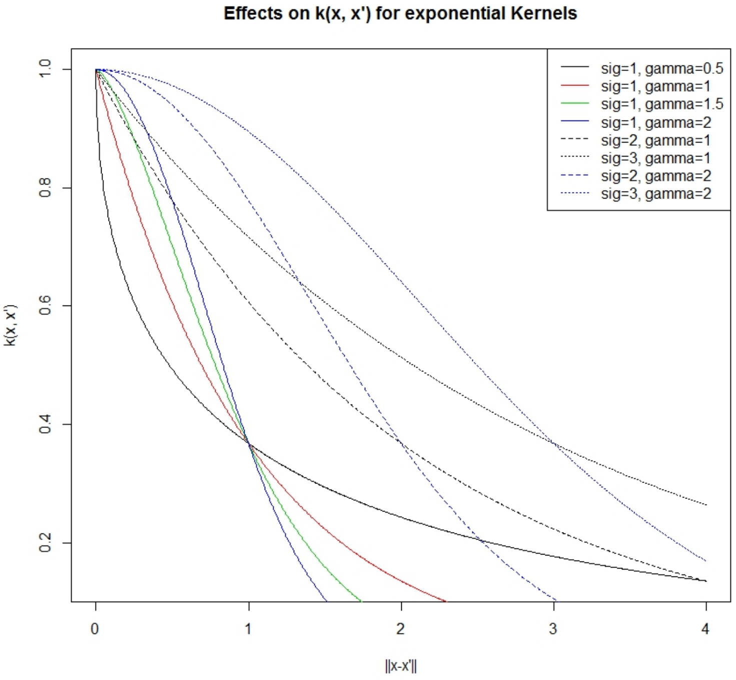
$$\kappa\left(\boldsymbol{x}_i,\, \boldsymbol{x}_j\right) = exp\left(-\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma}\right)^{2}\right) \tag{4.4}$$

In the exponential family, σ and γ controls the flexibility[9] of the kernel. Small σ will amplify the weight of large distances. As the value of σ grows large, the kernel will be forced to 1 as $\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2} \to 0$. This effect is seen in Figure 4.5 as a varying degree of paramter values are used for the γ-Exponential family of kernels. The x-axis is values of $\|\mathbf{x}_i - \mathbf{x}_j\|$, describing a distance between 2 objects. As the distance increases, the strength of the relationship between input vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ decreases. How it decreases depends on the parameterization of the kernel.

[9] Flexibility of kernels refers to the number of parameters available for tuning. The more tuning parameters a kernel has, the more flexible the kernel

Numerical optimization techniques are often used to assign the values of tuning parameters.

Figure 4.5: The strength of relationship of the exponential kernel for multiple parameter selections as a function of $\|\mathbf{x}_i - \mathbf{x}_j\|$



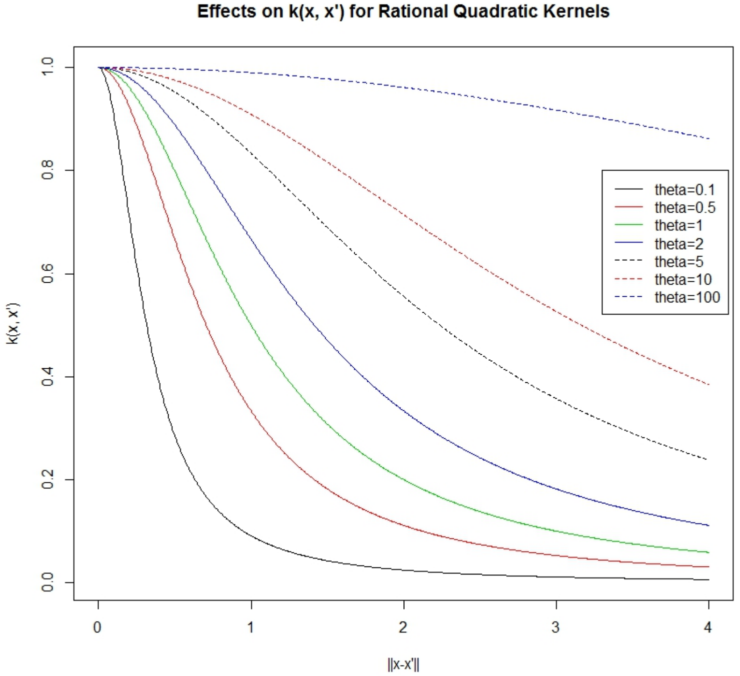
Definition 4.10. The rational quadratic kernel

for $\theta > 0$,

$$\kappa\left(\mathbf{x}_i,\, \mathbf{x}_j\right) = 1 - \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\|\mathbf{x}_i - \mathbf{x}_j\|^2 + \theta}$$

The rational quadratic kernel is considered one of the least flexible models due to having one parameter to optimize and being bounded in (0, 1). The effect of θ can be seen in Figure 4.6 below. Small values of θ cause the strength of the relationship to drop off very quickly to 0; and as θ increases, the strength of the relationship also increases towards 1. As θ gets large the kernel will converge to 1.

Figure 4.6: The strength of the relationship of the rational quadratic kernel for multiple parameter values



Definition 4.11. The Matérn kernel

for $v,\ \sigma > 0,\ \boldsymbol{d} = \|\mathbf{x}_i - \mathbf{x}_j\|$ and K_v is a modified Bessel function.

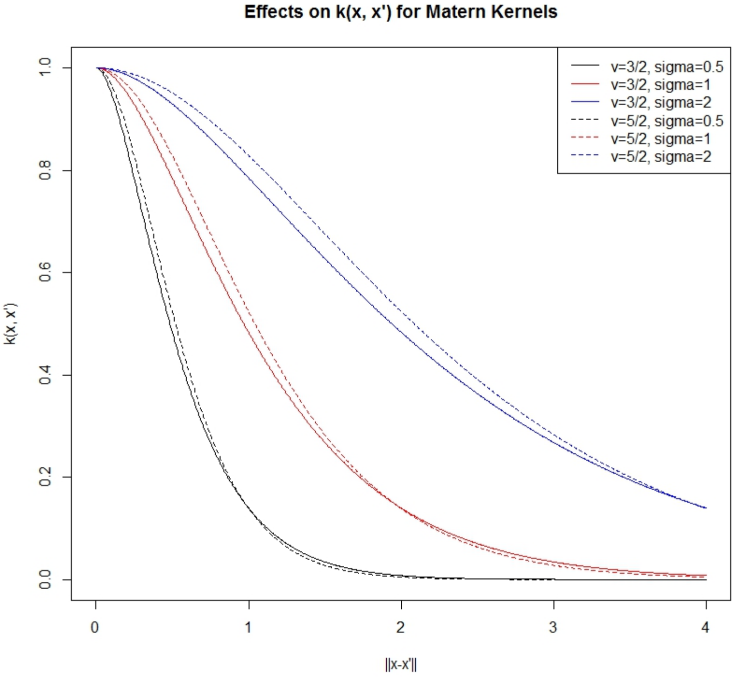
$$\kappa_v\left(\boldsymbol{d}\right) = \frac{2^{1-v}}{\Gamma(v)}\left(\frac{\sqrt{2v}\boldsymbol{d}}{\sigma}\right)^v K_v\left(\frac{\sqrt{2v}\boldsymbol{d}}{\sigma}\right)$$

The Matérn kernel class has become increasingly popular in machine learning due to the flexibility of the class. It is the resulting product of an exponential and polynomial kernel of order d. The simplest cases of the Matérn kernel are when v is a half integer value, defined as $v = d + \frac{1}{2}$, for $d \in \mathbb{N}$ which is the polynomial order. The general form of the Matérn kernel for a half integer is given by [56],

$$\kappa_{v=d+\frac{1}{2}}\left(\boldsymbol{d}\right) = exp\left(-\frac{\sqrt{2v}\boldsymbol{d}}{\sigma}\right)\frac{\Gamma\left(d+1\right)}{\Gamma\left(2d+1\right)}\sum_{i=0}^{d}\frac{(d+i)!}{i!\left(n-i\right)!}\left(\frac{\sqrt{8v}\boldsymbol{d}}{\sigma}\right)^{d-i}$$

However, as noted in [56], for the case of $v = \frac{1}{2}$, the Matérn class becomes a rough exponential function with $\gamma = 1$ and, conversely, as $v \to \infty$ the Matérn kernel becomes an overly smooth squared exponential or Gaussian kernel. Additionally, for $v \geq \frac{7}{2}$, unless prior knowledge of the underlying process is known, it is difficult to distinguish between $\frac{7}{2}$ and ∞ in practice. In light of this, the two most popular choices for v are $\frac{3}{2}$ and $\frac{5}{2}$ [56]. In Figure 4.7 the effect of different Matérn parameterizations is seen. The effect of v for the example is subtle when contrasted with the effect σ has on the strength of the relationship. While v shifts the tail weights for smaller values of σ, it is σ that increases the strength of the relationship over larger distances.

Figure 4.7: The strength of the relationship of the Matérn kernel for multiple parameter values



4.2.2 Nonstationary kernels

The most basic non-stationary kernel is the polynomial kernel. Nonstationary kernels do not make use of a difference between inputs as with stationary vectors, and their use is usually with data types that don't have a natural "origin" such as images.

Definition 4.12. Polynomial kernels:

$$\kappa\left(\mathbf{x}_i,\mathbf{x}_j\right)=\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle^d=\left(\sum_{k=1}^{p}\left[x_i\right]_k\left[x_j\right]_k\right)^d \tag{4.5}$$

where $d\in\mathbb{N},\ \mathbf{x}_i,\mathbf{x}_j\in X,\ \boldsymbol{x}_i=\{x_{i1},...,\ x_{ip}\}$ and p is the object dimension.

Another representation of the polynomial kernel is:

$$\kappa\left(\mathbf{x}_i,\mathbf{x}_j\right)=\left(\left\langle \mathbf{x}_i,\mathbf{x}_j\right\rangle+R\right)^d \tag{4.6}$$

where R is an additional tuning parameter. This representation allows the scientist to control the relative weightings of the degree monomials. Larger values for R decreases the weight of higher order polynomials, and equation 4.6 may be rewritten as

$$\begin{aligned} \kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) &= \sum_{s=0}^{p} \left(\begin{array}{c} p \\ s \end{array}\right) R^{p-s} \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle^s \\ &= \sum_{s=0}^{p} a_s \hat{\kappa}\left(\mathbf{x}_i, \mathbf{x}_j\right) \end{aligned} \tag{4.7}$$

where

$$a_s = \left(\begin{array}{c} p \\ s \end{array}\right) R^{p-s}$$

$$\hat{\kappa}\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle^s$$

Another popular type of kernel is the Analysis of Variance (ANOVA) kernel, which is a specific type of convolution kernel. The convolution kernel is explained in [31] as a kernel defined on structured objects, and analogous to ANOVA in statistics, it is a decomposition of kernels on some object $x \in X$ where x is composed of $x_p \in X_p$ for $p \in P$. What this means is that the object x is a combination of its decompositions, x_p. For instance, if we have the string $x = XYZ$, we can choose $p = 2$ and decompose x into the parts $x_1 = X$ and $x_2 = YZ$ and apply kernels accordingly. The set of allowed decompositions is noted as the relation $R\left(x_1, ..., x_p,\ x\right)$ or the parts $x_1, ..., x_p$ that constitute the composite object x. The ANOVA kernel is defined as [31]:

Definition 4.13. ANOVA kernels

Consider $X = S^N$ for some set S, and kernels $\kappa^{(i)}$ on $\mathcal{S} \times \mathcal{S}$, for $i = 1, ..., N$. For $P = 1, ..., \ N$, the ANOVA kernel of order P is

$$\kappa \left(\mathbf{x}_k, \mathbf{x}_l \right) \quad = \quad \sum_{1 \leq i_1 < ... < i_p \leq N} \prod_{p=1}^{P} \kappa^{(i_p)} \left(\boldsymbol{x}_{k,i_p}, \, \boldsymbol{x}_{l,i_p} \right) \tag{4.8}$$

ANOVA kernels are typically used with moderate values of P . A simple example for an ANOVA kernel with two inputs $\mathbf{x}_i, \ \mathbf{x}_j$ may look like:

$$\begin{aligned} \kappa \left(\mathbf{x}_i, \mathbf{x}_j \right) &= \mathbf{x}_i^t \mathbf{x}_i + \mathbf{x}_j^t \mathbf{x}_j + \mathbf{x}_i^t \mathbf{x}_j \\ &= \kappa \left(\mathbf{x}_i, \mathbf{x}_i \right) + \kappa \left(\mathbf{x}_j, \mathbf{x}_j \right) + \kappa \left(\mathbf{x}_i, \mathbf{x}_j \right) \end{aligned}$$

which looks very similar to two “source” effects and an interaction component.

4.2.3 String kernels

A third popular type of kernel are string kernels. These kernels are designed to count occurances typically in character strings such as handwriting or DNA.

Definition 4.14. Intersection kernel

This kernel is used to measure the intersecting elements of subsets, A_1, A_2, in domain D with a corresponding measure, μ.

$$\kappa \left(A_1, \, A_2 \right) = \mu \left(A_1 \bigcap A_2 \right)$$

This kernel is used often in problems with discrete data. A specific family of the intersection kernel is the string kernel.

Definition 4.15. String kernels [56]

If we have a set of strings of interest A and two strings of letters $\boldsymbol{x}_i,\ \boldsymbol{x}_j$ in some document. $\boldsymbol{s} \in A$ is a substring of $\boldsymbol{x}$ of length $|\boldsymbol{x}|$ if we can write $\boldsymbol{x} = \boldsymbol{usv}$ for some strings $\boldsymbol{u},\ \boldsymbol{s},\ \boldsymbol{v}$ (possibly of length 0). Let $\phi_s(\boldsymbol{x})$ denote the number of times that a substring $\boldsymbol{s}$ appears in string $\boldsymbol{x}$. The kernel between two strings $\boldsymbol{x}_i,\ \boldsymbol{x}_j$ is defined as:

$$\kappa(\boldsymbol{x}_i,\ \boldsymbol{x}_j) = \sum_{\boldsymbol{s} \in A} w_s \phi_s(\boldsymbol{x}_i)\, \phi_s(\boldsymbol{x}_j)$$

where w_s is a non-negative weight for substring $\boldsymbol{s}$. If we were interested in giving shorter substrings more weight, we could set $w_s = \lambda^{|\boldsymbol{s}|},\ 0 < \lambda < 1$.

Within the family of string kernels, there are 3 popular special cases:

Case 1. The bag-of-characters kernel. Given when $w_s = 0$ for $|\boldsymbol{s}| > 1$. In this case, A is a set of unique letters. This kernel measures the number of times each character in $\mathcal{A}$ appears in $\boldsymbol{x}$.

Case 2. The bag-of-words kernel. Popular in text analysis, one is concerned with the frequency of word occurence. In this case A is a set of unique words, and each $\boldsymbol{s}$ represents a word. The entire text (papers, books, etc.) is searched and the number of times that each word appears is measured.

Case 3. k-spectrum kernel. Any substrings of length k is considered. For example taking the 2-spectrum kernel for the strings $\boldsymbol{x}_i = ABCDE$, $\boldsymbol{x}_j = CDEFG$ would result in $\kappa(\boldsymbol{x}_i,\ \boldsymbol{x}_j) = 2$ since CD and DE are found in both strings.

4.3 Applications of kernel methods to non-forensic problems

The use of kernel methods is found throughout numerous scientific fields. The flexibility of the models, and relative ease of construction of kernels has allowed scientists to take advantage of the power of kernels in pattern recognition, classification (image, speech, etc.)

Industrial processes can take advantage of kernel methods in order to speed up production, control for quality, or discover new formulations to improve existing products. In [38], the authors developed a kernel learning method to predict: the quality of crystallization processes on crystal size distribution; the quality of polymerization processes based on molecular weight distributions; the quality of powders based on particle size distributions; the quality of papers based on pulp fiber length distribution. Past research into these quality controls uses lump-sum or sufficient statistics, but the authors instead represented the data as a distribution and the shape of the distribution was used in the kernel method. Traditional methods were unable to score differences in shapes, but using a Gaussian kernel allowed the authors to compare the distribution shapes. This kernel-learning method improved production efficiency by predicting batch quality faster than traditional means, and the authors were able to do this online (without interupting the workflow) during the process.

A recent kernel based method in drug discovery accelerates the discovery of highly active peptides [26]. The method works to help sort through the vast combinations of chemicals used to create drug compunds. Doing so helps reduce cost, complexity and time to discovery of new drugs. The data used in this study were strings of amino acids from peptides and their associated validated bioactivity such as binding affinity and IC_{50}. Ultimately, the authors want to predict which strings of amino acids have highly active peptides for new drugs. To do this, a kernel called the Generic String Kernel was created using the closure properties of kernels. The authors combined a k-spectrum kernel to measure k-mers and a Gaussian kernel to penalize the distance that the k-mers were from each other in the strings. The authors chose to use a single response due to unforseen complications and chose to use affinity. They observed some promising results and are working to add multiple objective measures to ensure amino acid strings meet more stringent criteria before physical processing.

Improving a manufactured product is not the sole use of kernel methods in industry. The highly mechanized manufacturing processes, regardless of product, could benefit from the use of kernel methods designed to reduce downtime, decrease repair time, and offer a safer environment. In [17] an industrial process involving many controllers and sensors is monitored for changes that the automated controllers are unable to handle. These unpermitted deviations of at least one characteristic property or variable in the industrial system is defined as a fault. The authors hope to detect early faults using the large amount of (often noisy) feedback from controllers and sensors.

The data was first preprocessed to remove noise and outliers and then reduced in dimensionality for better facilitation of early fault detection. The authors selected the Gaussian kernel to reduce their data. Then the authors one of two kernel techniques: Kernel Fisher Discriminant Analysis (KFDA) and Kernel Principal Component Analysis (KPCA) to further reduce the dimension and remove redundancy. On the backend of both of these techniques, an Artificial Neural Network was used to classify the input into normal and not normal process behavior. It was found that the KFDA method was able to separate the input data the best and improve fault detection tasks.

In [52], the authors introduce the PQ kernel to obtain a better pairwise similarity between visual word histograms for object classification in computer vision. The authors hope to improve image recognition using this method. The technique begins with an image and a regular grid imposed onto the image. In each part of the grid, a scale-invariant feature transformation (SIFT) is computed, followed by a vector quantization process to assign a visual word to that part of the grid. A bag-of-words kernel is applied to the quantized image, and the frequency of each word is recorded in a histogram. The histogram is now representative of the image. The PQ kernel then measures the number of concordant pairs among two histograms (denoted as P) and the number of discordant pairs (denoted as Q). The resulting kernel is defined as $\kappa\left(X,\, Y\right) = 2\left(P - Q\right)$ for images $X,\, Y.$ This kernel was found to outperform other several other popular kernels used for image classification on 2 sets of benchmark images.

A kernel method was used by researchers for determining the spread of the avian flu through poultry farms located in the Netherlands[7]. Researchers are interested with the status of farms located throughout the Netherlands. The status of a farm at anytime falls into one of four levels: uninfected, infected but not yet infectious, infected and infectious, removed (culled). As farms became infected, researchers attempted to gather key demographic characteristics (number of barns, number of animals, type of animals, age of the animals) and data of epidemiological interest (number of dead animals per day, number of sick animals per day, food and water intake per day) to supplement the status of the farm whether or not certain demographics helped predict the spread of the avian flu. The researchers were unable to collect this data for all farms, and as such some of their data is incomplete. All of this data was tracked and updated daily, so researchers could track the spread of the virus and hopefully protect farms with higher infection hazards.

To assess the infection risk, the researchers estimated $p\left(r_{ij}\right)$, the probability that an uninfected farm j will be infected by an infected farm i. Using the Euclidean distance defined as $r_{ij} = |r_i - r_j|$ and an infectious period defined as $T_i \sim Gamma(c,\, T/c)$, the researchers estimated $p\left(r_{ij}\right) = 1 - exp\left(-h\left(r_{ij}\right)T_i\right)$ where $h\left(r_{ij}\right)$ is the transmission kernel. The transmission kernel is a 3-parameter logistic expression, given by

$$h(r;\;h_0,\, r_0,\, \alpha) = \frac{h_0}{1 + \left(r/r_0\right)^{\alpha}}$$

where the parameters $h_0,\ r_0,\ \alpha$ are estimated from the data via maximum likelihood. They could also be estimated via numerical optimization techniques, possibly reducing computational effort. In actuality, the researchers are using two kernels, one within the other as the entire $p\left(r_{ij}\right)$ can be represented as a γ-exponential kernel using a modified distance metric. The second kernel is used to incorporate infection period information not included in the transmission kernel.

The results of modelling the avian flu in this manner allowed authorities to implement local control measures such as vaccination or culling of local farms to delay the spread of the virus. Of the two control measures, researchers found merit in both control measures and the cost of the action is the ultimate prize.

As seen, the modern use of kernels is often concerned with predicting outcomes or monitoring processes. The main benefits of kernels are dimension reduction and linearization of data so that techniques such as regression, PCA, LDA, etc. may be used. In our case, we want to leverage the dimension reduction and linearization benefits of kernels in a Hilbert space; in addition, we want to develop a likelihood structure to perform model selection tasks in a Hilbert space which appears to be a novel contribution to the field.

Part II

The kernel model

5 Development of the kernel model

As mentioned in Section 3, the project consists of proposing a likelihood structure for $\boldsymbol{s}$, a vector of all pairwise comparisons between sets of i.i.d. objects $\boldsymbol{e}_{u1},\ \boldsymbol{e}_{u2},\ \boldsymbol{e}_{a}$, and the algorithm necessary to estimate the marginal distributions of $\boldsymbol{s}$ under the two following models:

$\mathcal{M}_p : \boldsymbol{e}_{u1},\ \boldsymbol{e}_{u2}$ originates from a common randomly selected source in the population

$\mathcal{M}_d : \boldsymbol{e}_{u1},\ \boldsymbol{e}_{u2}$ originates from 2 different random sources in the population of sources

The development of the kernel model is outlined here before details of each step are given. A representation of the general process that we studied to derive a score model which embodies the set of assumptions regarding the distribution G of $\boldsymbol{s}$ is given below:

$$\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl} \xrightarrow{\kappa} s_{ijkl} \rightarrow \boldsymbol{s} \sim G \tag{5.1}$$

In the Representation 5.1 we start with objects $\boldsymbol{x}_{ij},\ \boldsymbol{x}_{kl}$ where $i,\ k$ represent the i^{th} and k^{th} sources in the population and $j,\ l$ represent the j^{th} and l^{th} samples of their respective sources. We reduce their dimension to a pairwise score s_{ijkl} using an isotropic stationary kernel κ. The vector of scores resulting from all pairwise comparisons, $\boldsymbol{s}$, has some distribution G that we wish to model. The following general approach was used to define G:

1. First stage:

    (a) We defined a hierarchical sampling model for $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{kl}$, and a basic isotropic stationary kernel $\kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right) = \left(\boldsymbol{x}_{ij} - \boldsymbol{x}_{kl}\right)^2 = s_{ijkl}$;

   (b) We calculated the expectations and covariances of pairs of scores s_{ijkl} and $s_{i'j'k'l'}$ for different situations (e.g. $i = i'$, $ij = i'j'$, etc.);

   (c) We wrote the covariance matrix, $\boldsymbol{\Sigma}$, of the "sampling-driven model" of $\boldsymbol{s}$ in terms of design matrices and the sampling model variances, σ_a^2, σ_r^2, defined in (a) above.

2. Second stage:

   (a) We studied the eigenstructure of $\boldsymbol{\Sigma}$;

   (b) We defined a piecewise linear model for the score s_{ijkl}, which we called the "score model", that would result in the same covariance matrix $\boldsymbol{\Sigma}$;

   (c) We verified analytically and by simulation that the different mean and covariance components of the score model and sampling model corresponded exactly;

   (d) We verified by simulation that the different mean and covariance components for the score models hold for non-trivial sampling models and different stationary kernels.

3. Third stage:

   (a) We derived closed-form solutions for the estimation of parameters for the score model;

   (b) We implemented the Method of Composition to obtain posterior distributions for the parameters of the score model.

5.1 Sampling-driven model

5.1.1 Hierarchical sampling model and definition of a sampling-driven model

We began studying the score model by defining a very simple univariate sampling model with finite moments (up to the 4th moment) that allows us to derive the scores, and their means and covariances directly[10]. Let x_{ij} be generated by a hierarchical model, given in definition 5.1.

Definition 5.1. Sampling model

Let the j^{th} sample from the i^{th} source be defined as a simple random effects model:

$$x_{ij} = \mu + a_i + r_{ij} \tag{5.2}$$

with overall mean μ, source effect $a_i \sim N\left(0,\, \sigma_a^2\right)$ for $i = 1, ..., n$ sources, and within-source effect, $r_{ij} \sim N\left(0,\, \sigma_r^2\right)$ for $j = 1, ..., n_o$ samples per source. Let $N = \binom{nn_o}{2}$ be the number of pairwise comparisons.

In definition 5.1 we consider that each object is univariate. This choice is made in order to explore analytically the covariance structure of the score distribution, which dimension is defined by the number of pairwise comparisons, N, and not by the dimension of the original objects. We will discuss the impact of the dimension of the original object on the score model later. We begin by studying the sampling-driven model using the squared Euclidean distance as our kernel.

For the j^{th} sample from the i^{th} source and the l^{th} sample from the k^{th} source, let

$$\begin{aligned} s_{ijkl} = k\left(x_{ij},\, x_{kl}\right) &= \left(x_{ij} - x_{kl}\right)^2 \\ &= x_{ij}^2 - 2x_{ij}x_{kl} + x_{kl}^2 \end{aligned}$$

[10] We discuss later the extension of the sampling model to situations that cannot be solved directly.

where

$$\begin{aligned}
x_{ij}^2 &= \mu^2 + a_i^2 + r_{ij}^2 + 2\mu a_i + 2\mu r_{ij} + 2a_i r_{ij} \\
x_{kl}^2 &= \mu^2 + a_k^2 + r_{kl}^2 + 2\mu a_k + 2\mu r_{kl} + 2a_k r_{kl} \\
2x_{ij}x_{kl} &= 2\left(\mu\mu + \mu a_k + \mu r_{kl} + \mu a_i + a_i a_k + a_i r_{kl} + \mu r_{ij} + a_k r_{ij} + r_{ij} r_{kl}\right).
\end{aligned}$$

Upon further simplification, we obtain the following sampling-driven model[11] 5.3:

$$\begin{aligned}
s_{ijkl} &= a_i^2 + a_k^2 + r_{ij}^2 + r_{kl}^2 + 2a_i r_{ij} + 2a_k r_{kl} - 2a_i a_k - 2a_i r_{kl} - 2a_k r_{ij} - 2r_{ij} r_{kl} \\
&= \left(a_i - a_k\right)^2 + \left(r_{ij} - r_{kl}\right)^2 + 2\left(a_i - a_k\right)\left(r_{ij} - r_{kl}\right) \\
&= \left(\left(a_i - a_k\right) + \left(r_{ij} - r_{kl}\right)\right)^2 \qquad (5.3)
\end{aligned}$$

5.1.2 Covariance of the sampling-driven model

Before we can began calculating covariances, we first determined the number of unique situations for $ijkl$ vs. $i'j'k'l'$. These situations provide the basic structure of dependencies in the covariance matrix, $\mathbf{\Sigma}$, of $\boldsymbol{s}$. The situations were defined as a function of the source indices $(i,\ k,\ i',\ k')$ and object indices $(j,\ l,\ j',\ l')$ within these sources. For example, if two scores have a source index in common, $i = i'$ or $i = k'$or $k = i'$ or $k = k'$, then they share an effect a_i and will covary. Similar situations will arise when two scores share more than one source or directly share objects. We counted 13 unique situations listed in Table 5.1. These situations occur with a minimum of $n = 4$ and $n_o = 4$. No new situations occur when n or n_o are increased.

[11] As above, we chose to use the squared-Euclidean distance as it allows us to derive the parameters of the model analytically.

Table 5.1: Score situations for sampling-driven model

# sources involved	# objects in common	$ijkl$	$i'j'k'l'$
1	2	1112	1112
	1	1112	1113
	0	1112	1314
2	2	1121	1121
	1	1121	1221
	0	1121	1222
2 ($i = k$, $i' \neq k'$)	1	1112	1121
	0	1112	1321
2 ($i = k$, $i' = k'$, $i \neq i'$)	0	1112	2122
3 ($i = i'$)	1	1121	1131
	0	1121	1231
3 ($i = k$)	0	1112	2131
4	0	1121	3141

The sampling-driven model is used to calculate the covariances found in $\mathbf{\Sigma}$ for our toy example. Additionally, two expectations are needed for the covariance calculations, given below.

$$\begin{aligned}
E\left(s_{ijkl}|i=k\right) &= E\left[\left(a_i - a_i\right)^2 + \left(r_{ij} - r_{il}\right)^2 + 2\left(a_i - a_i\right)\left(r_{ij} - r_{il}\right)\right] \\
&= E\left(r_{ij}^2 + r_{il}^2 - 2r_{ij}r_{il}\right) \\
&= E\left(r_{ij}^2\right) + E\left(r_{il}^2\right) - E\left(2r_{ij}r_{il}\right) \\
&= 2\sigma_r^2
\end{aligned}$$

$$\begin{aligned}
E\left(s_{ijkl}|i \neq k\right) &= E\left[\left(a_i - a_k\right)^2 + \left(r_{ij} - r_{kl}\right)^2 + 2\left(a_i - a_k\right)\left(r_{ij} - r_{kl}\right)\right] \\
&= E\left[\left(a_i - a_k\right)^2 + \left(r_{ij} - r_{kl}\right)^2\right] \\
&= E\left(a_i^2 + a_k^2 - 2a_ia_k + r_{ij}^2 + r_{il}^2 - 2r_{ij}r_{il}\right) \\
&= 2\sigma_a^2 + 2\sigma_r^2
\end{aligned}$$

All covariances have been calculated in collaboration with John Miller (George Mason University). The calculations are provided in Appendix 10.1.1 and their final results are reported in Table 5.2. Note that the covariances $Cov\left(s_{1112},\, s_{1314}\right)$ and $Cov\left(s_{1112},\, s_{1321}\right)$ are both equivalent to 0 in this due to conditional independence and that the source effect is removed as an effect of the toy example.

Table 5.2: Covariance components of sampling-driven model

#	ijkl	i'jk'l	Covariance
1	1112	1112	$8\sigma_r^4$
2	1112	1113	$2\sigma_r^4$
3	1112	1314	0
4	1121	1121	$8\sigma_a^4 + 16\sigma_a^2\sigma_r^2 + 8\sigma_r^4$
5	1121	1221	$8\sigma_a^4 + 8\sigma_a^2\sigma_r^2 + 2\sigma_r^4$
6	1121	1222	$8\sigma_a^4$
7	1112	1121	$2\sigma_r^4$
8	1112	1321	0
9	1121	1131	$2\sigma_a^4 + 4\sigma_a^2\sigma_r^2 + 2\sigma_r^4$
10	1121	1231	$2\sigma_a^4$
11	1112	2122	0
12	1112	2131	0
13	1121	3141	0

A method to calculate $\mathbf{\Sigma}$ of $\boldsymbol{s}$ in terms of design matrices based on the sampling-driven model and its variances was explored next. From Table 5.2, we note that $\mathbf{\Sigma}$ can be decomposed as a function of 3 components, two solitary covariance parameters σ_a^4 and σ_r^4 and the cross product term, $\sigma_a^2\sigma_r^2$, and three suitably chosen design matrices:

$$\mathbf{\Sigma} = \mathbf{A}\sigma_a^4 + \mathbf{C}\sigma_a^2\sigma_r^2 + \mathbf{R}\sigma_r^4 \tag{5.4}$$

where $\mathbf{A}$, $\mathbf{R}$, and $\mathbf{C}$ are design matrices. Note the decomposition 5.4 has a quadratic form and we used this to reduce the number of design matrices required by one. To do this, we made use of a Schur/Hadamard[12] product [32] for matrices:

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathbf{A}\sigma_a^4 + \mathbf{C}\sigma_a^2\sigma_r^2 + \mathbf{R}\sigma_r^4 \\
&= \left[\mathbf{A}^{\frac{1}{2}}\sigma_a^2 + \mathbf{R}^{\frac{1}{2}}\sigma_r^2\right] \circ \left[\mathbf{A}^{\frac{1}{2}}\sigma_a^2 + \mathbf{R}^{\frac{1}{2}}\sigma_r^2\right] \\
&= \sigma_a^4\left[\mathbf{A}^{\frac{1}{2}} \circ \mathbf{A}^{\frac{1}{2}}\right] + 2\sigma_a^2\sigma_r^2\left[\mathbf{A}^{\frac{1}{2}} \circ \mathbf{R}^{\frac{1}{2}}\right] + \sigma_r^4\left[\mathbf{R}^{\frac{1}{2}} \circ \mathbf{R}^{\frac{1}{2}}\right]
\end{aligned}$$

where $\mathbf{A}^{\frac{1}{2}}$ is the square root of the elements of $\mathbf{A}$ and $\mathbf{R}^{\frac{1}{2}}$ is the square root of the elements of $\mathbf{R}$. Using this form, the design matrix $\mathbf{C} = 2\left[\mathbf{A}^{\frac{1}{2}} \circ \mathbf{R}^{\frac{1}{2}}\right]$ and we need only to construct design matrices $\mathbf{A}$ and $\mathbf{R}$. We detail the construction of the three design matrices $\mathbf{A}$, $\mathbf{R}$, $\mathbf{C}$ below. The design matrix for $\mathbf{R}$ is given first since its construction is the simplest, using a similar construction to that of the Gantz-Saunders model [24, 4]. Construction of $\mathbf{A}$ uses a modification of the construction used for $\mathbf{R}$, which is shown second. Finally, since $\mathbf{C}$ uses both $\mathbf{A}$ and $\mathbf{R}$, it was constructed last.

Design matrix R The design matrix $\mathbf{R}$ accounts for the contribution of the within-source variance to the covariance of scores sharing at least one object. We collected the contributing covariance terms from Table 5.2 involving σ_r^4 and listed them in Table 5.3.

Table 5.3: Contributing components of σ_r^4

ijkl	i'j'k'l'	Covariance term
1112	1112	$8\sigma_r^4$
1112	1113	$2\sigma_r^4$
1121	1121	$8\sigma_r^4$
1121	1221	$2\sigma_r^4$
1112	1121	$2\sigma_r^4$
1121	1131	$2\sigma_r^4$

[12] The Schur or Hadamard product is an elementwise matrix multiplication of two matrices, $\mathbf{A}$ and $\mathbf{R}$ with equivalent dimensions. The operation is denoted in one of three ways: $\mathbf{A} \circ \mathbf{R}$, $[\mathbf{A} \circ \mathbf{R}]_{ij}$, or $[\mathbf{A}]_{ij}\,[\mathbf{R}]_{ij}$ where ij denotes the row and column index of the elements being multiplied.

There are two terms with a scalar multiplier of 8: $Cov\,(s_{1112},\, s_{1112})$ and $Cov\,(s_{1121},\, s_{1121})$ which are simply the variance of s_{ijkl} for any $ijkl$ and are the diagonal terms in the covariance matrix $\mathbf{\Sigma}$. There are four covariance terms with only one object in common within $\mathbf{\Sigma}$ and have a scalar multiplier of 2. All other covariance terms do not have aσ_r^4 component. This leads to the logical rule to build the design matrix $\mathbf{R}$:

$$\mathbf{R}\left[s_{ijkl},\, s_{i'j'k'l'}\right] = \begin{cases} 8 & if\ ijkl = i'j'k'l' \\ 2 & if\ \ (ij = i'j' \parallel k'l') \oplus (kl = i'j' \parallel k'l') \\ 0 & otherwise. \end{cases}$$

An example of the design matrix $\mathbf{R}$ built with $n = 3$ and $n_o = 3$ per source is given in the the Appendix 10.3.2.

Upon inspection of the structure of $\mathbf{R}$, it can be seen that it is very similar to that of the structure of $\mathbf{PP}^t$ in [24, 4] (in fact, $\mathbf{PP}^t$ may be constructed for our purposes if we consider all objects to come from 1 source with nn_o samples.) If we remove 4 from the diagonal of $\mathbf{R}$ and divide the remaining elements by 2, then in fact we do get $\mathbf{PP}^t$ with size $N = \binom{nn_o}{2}$. This results in the design matrix for $\mathbf{R}$ being:

$$\mathbf{R} = 2\mathbf{PP}^t + 4\mathbf{I}_N.$$

Design matrix A The design matrix $\mathbf{A}$ accounts for the contribution of the between-source variance to the covariance of scores sharing at least one source. We collected the contributing covariance terms from Table 5.2 involving σ_a^4 and listed them in Table 5.4.

Table 5.4: Theoretical covariance components of between source model

ijkl	i’j’k’l’	Covariance term
1121	1121	$8\sigma_a^4$
1121	1221	$8\sigma_a^4$
1121	1222	$8\sigma_a^4$
1121	1131	$2\sigma_a^4$
1121	1231	$2\sigma_a^4$

There are three terms with a scalar multiplier of 8: $Cov\left(s_{1121},\, s_{1121}\right),\ Cov\left(s_{1121},\, s_{1221}\right)$ and $Cov\left(s_{1121},\, s_{1222}\right)$. These are covariance terms for between-source scores calculated between objects from the same two independent sources such that $i = i'$, $k = k'$, $i \neq k$. There are two covariance terms for scores with only one source in common and both have a scalar multiplier of 2. All other covariance scalar multipliers, with respect to σ_a^4, are equal to zero. This leads to the logical rules that build the design matrix $\mathbf{A}$:

$$\mathbf{A}\left[s_{ijkl},\, s_{i'j'k'l'}\right] = \begin{cases} 8 & if\ i = i',\, k = k',\, i \neq k \\ 2 & if\ \ [(i = i' \parallel k') \oplus (k = i' \parallel k')]\ \&\ (i \neq k,\, i' \neq k') \\ 0 & otherwise. \end{cases}$$

An example of the design matrix $\mathbf{A}$ built with 3 sources and 3 samples per source is given in 10.1.

The major difference in the structure of $\mathbf{A}$ and $\mathbf{R}$ is most obvious on the diagonal, where there are numerous zero terms for the variance of within-source scores in $\mathbf{A}$. The structure of $\mathbf{A}$ can be recreated with the use of a matrix similar to $\mathbf{P}$ [24, 4] but instead of within-source comparisons, we use source-level comparisons. This design is captured by the $\mathbf{Q}$ matrix, which in general will have the indexing $ijkl$ given in N rows and n columns, each labeled by the source id. The rows in $\mathbf{Q}$ will follow one of two patterns; (1) if $i = k$, the row will contain all zeroes (same source); if $i \neq k$, the row will contain all zeroes except for in the i^{th} and k^{th} columns, which will be one. Note that the samples given a source, indexed by j and

l, do not have an effect since this is for the source level. An example of the $\mathbf{Q}$ matrix for the example of $n = 3$ and $n_o = 3$ is given in Table 5.5.

Table 5.5: Example of the $\mathbf{Q}$ matrix for $n = 3,\ n_o = 3$

	1	2	3
1112	0	0	0
1113	0	0	0
1121	1	1	0
1122	1	1	0
1123	1	1	0
1131	1	0	1
1132	1	0	1
1133	1	0	1
1213	0	0	0
1221	1	1	0
1222	1	1	0
1223	1	1	0
1231	1	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Then, taking the outer product $\mathbf{QQ}^t$, the matrix given in Table 10.3 in Appendix 10.3.3 is produced. The matrix $\mathbf{QQ}^t$ has a similar structure as the one seen in $\mathbf{A}$ but with different scalar multipliers than the 8 and 2 observed in Table 5.4. By using a Schur product with $\mathbf{QQ}^t$ and multiplying by 2, we recover the design matrix $\mathbf{A}$,

$$\mathbf{A} = 2\left[\mathbf{QQ}^t\right] \circ \left[\mathbf{QQ}^t\right].$$

Design matrix C As calculated earlier, we have that $\mathbf{C} = 2\left[\mathbf{A}^{\frac{1}{2}} \circ \mathbf{R}^{\frac{1}{2}}\right]$. To confirm this, we observe the covariance terms that include $\sigma_a^2\sigma_r^2$ in Table 5.2. These specific components are given in Table 5.6

Table 5.6: Theoretical covariance components of $\mathbf{C}$

ijkl	i'j'k'l'	Covariance term
1121	1121	$16\sigma_a^2\sigma_r^2$
1121	1221	$8\sigma_a^2\sigma_r^2$
1121	1131	$4\sigma_a^2\sigma_r^2$

and are the only terms where both σ_a^4 and σ_r^4 show up together. The logical rules to build $\mathbf{C}$ are:

$$\mathbf{C}\left[s_{ijkl},\, s_{i'j'k'l'}\right] = \begin{cases} 16 & if\ ijkl = i'j'k'l' \ \&\ i \neq k \\ 8 & if\ i,\, k = i',\, k' \ \&\ i \neq k \ \&\ \ (ij = i'j' \text{ or } k'l') \oplus (kl = i'j' \text{ or } k'l') \\ 4 & if\ i \neq k \ \&\ i' \neq k' \ \&\ \ (i = (i' \text{ or } k') \oplus k = (i' \text{ or } k')) \\ 0 & otherwise. \end{cases}$$

For our example of $n = 3$ and $n_o = 3$ results in the design matrix $\mathbf{C}$, given in Appendix 10.3.4 Table 10.4.

The decomposition of $\boldsymbol{\Sigma}$ in terms of matrices $\mathbf{A},\ \mathbf{R},\ \mathbf{C}$ is of no direct interest for the end-product of the estimation of the score model below. However, it is a critical tool to study the exact eigenstructure of $\boldsymbol{\Sigma}$ for our toy example, and simulate it for different n, n_o and σ_a^2, σ_r^2.

5.2 Studying the eigenstructure to help build the score model

Once we were able to exactly produce the covariance matrix $\boldsymbol{\Sigma}$ for our simple scenario, we attempted to interpret its eigenstructure. The eigenstructure of $\boldsymbol{\Sigma}$ may be useful when solving for the REML estimates of the final score model in a similar fashion as in the Gantz/Saunders approach [4, 24]. Additionally, understanding this structure may give us clues as to what form the score model will take and the number of terms it will require. We studied the effect of increasing the number of sources n while letting $n_o = 4$. The number of repeating eigenvalues are counted, in descending order, and tabulated in Table 5.7.

Accounting for the multiplicity of the eigenvalues was important because this offered the first view into the structure of the linear model. Eigenvalues of a covariance matrix each account for some amount of the total variance in the model and as such repeated eigenvalues may imply the existence of an effect in a linear model and the number of repetitions informs on the dimension of the subspace where that effect lives (i.e. degrees of freedom).

Table 5.7: Counts of repeating eigenvalues for $(n, 4)$

Rank	(4, 4)	(5, 4)	(6, 4)	(7, 4)	Multiplicity
1	1	1	1	1	1
2	3	4	5	6	$n-1$
3	2	5	9	14	$\binom{n}{2}-n$
4	12	15	18	21	$n\left(n_o-1\right)$
5	24	45	72	105	$n\left(n-2\right)\left(n_o-1\right)$
6	4	5	6	7	n
7	12	15	18	21	$n\left(n_o-1\right)$
8	62	100	147	203	$\binom{n(n_o-1)}{2}-n$
Total	120	190	276	378	$\binom{nn_o}{2}$

The formulas to calculate the number of repeating eigenvalues shown in the last column of Table 5.7 were confirmed for values of $n_o \neq 4$. Multiplicity of the eigenvalues given in Table 5.7, resemble that of a linear random effects model, as given in Equation 5.5. For example, the first subspace defined by the first eigenvector has dimension 1 which could correspond to a grand mean; the second subspace defined by the next $n-1$ eigenvectors could correspond to a main effect due to different sources; and so on. Our interpretation of the various subspaces is reported in Table 5.8.

Table 5.8: Possible score model terms

Rank	Multiplicity	Score model term	Interpretation
1	1	θ_b	Between score mean
2	$n-1$	$b_i,\ b_k$	Source effect
3	$\binom{n}{2}-n$	d_{ik}	Source interaction effect
4	$n\left(n_o-1\right)$	$t_{i:ij},\ t_{i:kl},\ t_{k:ij},\ t_{k:kl}$	Interaction between source and sample
5	$n\left(n-2\right)\left(n_o-1\right)$	e^b_{ijkl}	Between source error
6	n	θ_w	Within source mean
7	$n\left(n_o-1\right)$	$w_{ij},\ w_{kl}$	Within source effect
8	$\binom{n(n_o-1)}{2}-n$	e^w_{ijkl}	Within source error
Total	$\binom{nn_o}{2}$		

Based on the preceeding analysis, we hypothesized that the linear model for the score will have 2 fixed effects (means $\theta_b,\ \theta_w$) and 6 random effects, $b,\ d,\ t,\ w,\ e_w,\ e_b$ as shown in Table 5.8. The resulting model is a piecewise linear model for the score s_{ijkl}, pieced by between source scores $(i \neq k)$ and within source scores $(i = k)$:

$$s_{ijkl} = \begin{cases} \theta_b + b_i + b_k + d_{ik} + t_{i:ij} + t_{i:kl} + t_{k:ij} + t_{k:kl} + w_{ij} + w_{kl} + e^b_{ijkl} & if\ i \neq k \\ \theta_w + w_{ij} + w_{kl} + e^w_{ijkl} & if\ i = k \end{cases} \tag{5.5}$$

where $b_i \sim N\left(0,\, \sigma_b^2\right)$, $d_{ik} \sim N\left(0,\, \sigma_d^2\right)$, $w_{ij} \sim N\left(0,\, \sigma_w^2\right)$, $t_{i:ij} \sim N\left(0,\, \sigma_t^2\right)$, $e^b_{ijkl} \sim N\left(0,\, \sigma^2_{e^b}\right)$, and $e^w_{ijkl} \sim N\left(0,\, \sigma^2_{e^w}\right)$. The distribution for the vector of scores $\boldsymbol{s}$ converges asymptotically to $MVN\left(\boldsymbol{\theta},\, \boldsymbol{\Sigma}\right)$ as the dimension, p, of the original object increases. This is shown later in Theorem 5.4.

It is trivial to calculate the covariance of s_{ijkl} from the model described in equation 5.5. These covariances are reported in the middle column of Table 5.9 where they can be compared with the covariance previously obtained from the sampling-driven model. Their exact calculations are given in the Appendix 10.1.1.

Table 5.9 shows that a system of 6 equations can be used to solve for the values of the six variance parameters of the score model in equation 5.5, and the remaing two can be used to check the results.

To check that there is a direct correspondance between the sampling-driven model and the score model proposed in equation 5.5, the system of equations was used to solve for the variance parameters of the score model for different values of σ_a and σ_r $(\sigma_a > 0,\ \sigma_r > 0)$. An example of $\sigma_a = 5,\ \sigma_r = 2$ is given in Table 5.9 where the resulting solutions for the score model variances is:

$$\begin{aligned}
\sigma_b^2 &= 1250 \\
\sigma_d^2 &= 2500 \\
\sigma_w^2 &= 32 \\
\sigma_t^2 &= 400 \\
\sigma_{e^b}^2 &= 64 \\
\sigma_{e^w}^2 &= 64.
\end{aligned}$$

Table 5.9: Theoretical covariance components of between source model for $\sigma_a = 5,\ \sigma_r = 2$

ijkl	i'j'k'l'	Sampling-driven model	Score model	Example values
1112	1112	$8\sigma_r^4$	$2\sigma_w^2 + \sigma_{e^w}^2$	128
1112	1113	$2\sigma_r^4$	σ_w^2	32
1112	1314	0	0	0
1121	1121	$8\sigma_a^4 + 16\sigma_a^2\sigma_r^2 + 8\sigma_r^4$	$2\sigma_b^2 + \sigma_d^2 + 4\sigma_t^2 + 2\sigma_w^2 + \sigma_{e^b}^2$	6728
1121	1221	$8\sigma_a^4 + 8\sigma_a^2\sigma_r^2 + 2\sigma_r^4$	$2\sigma_b^2 + \sigma_d^2 + 2\sigma_t^2 + \sigma_w^2$	5832
1121	1222	$8\sigma_a^4$	$2\sigma_b^2 + \sigma_d^2$	5000
1112	1121	$2\sigma_r^4$	σ_w^2	32
1112	1321	0	0	0
1121	1131	$2\sigma_a^4 + 4\sigma_a^2\sigma_r^2 + 2\sigma_r^4$	$\sigma_b^2 + \sigma_t^2 + \sigma_w^2$	1682
1121	1231	$2\sigma_a^4$	σ_b^2	1250
1112	2122	0	0	0
1112	2131	0	0	0
1121	3141	0	0	0

5.3 Empirical simulation studies to confirm covariance structure of $\boldsymbol{\Sigma}$

While the score model was developed through the use of the dependency structure using univariate objects from a hierarchical normal sample, in practice it will be applied to data that is inherently non-normal and high-dimensional. We tested that the dependency structure for $\boldsymbol{s}$ was the same for samples from multivariate normal and multivariate non-normal distributions. Furthermore, we tested the structure of the covariance matrix for other types of stationary kernels. To confirm that the structure was the same, we used an empirical simulation to create large samples of $\boldsymbol{s}$ from which we calculate the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}$. We then decompose $\hat{\boldsymbol{\Sigma}}$ into its eigenstructure and observe the eigenvalues. If the multiplicity of the eigenvalues for $\hat{\boldsymbol{\Sigma}}$ matches that of $\boldsymbol{\Sigma}$, then we are satisfied that the covariance structures are the same. We also visibly compare the covariance structures by use of heatmap images. As discussed earlier in 5.1.2 we only need to check for the situation with $n = 4$, $n_o = 4$. We give the general algorithm for the empirical simulation in Algorithm 1

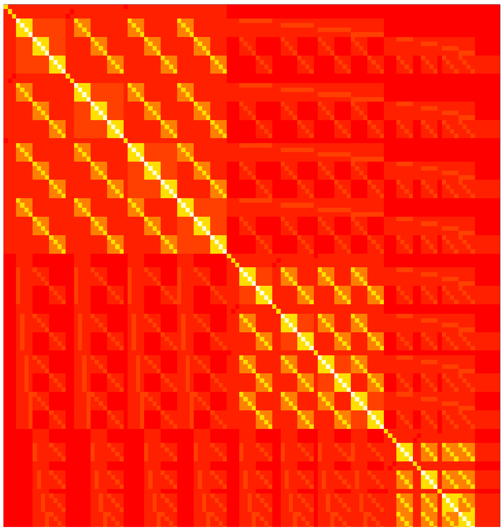
Algorithm 1 General object sampling for covariance structure check simulation

```
Let $\boldsymbol{x}_{ij} = \boldsymbol{a}_i + \boldsymbol{r}_{ij}$ where $\boldsymbol{a}_i$ is a source effect with some distribution $F$ and $\boldsymbol{r}_{ij}$ is a within-source effect with distribution $G$. Select a stationary kernel $\kappa(\boldsymbol{x}_{ij}, \boldsymbol{x}_{kl})$. Set $n.sims = 1,000,000$.

 1. Sample $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{kl}$ for $i, k = 1, 2, 3, 4$ and $j, l = 1, 2, 3, 4$
     (a) Compute the score vector $\boldsymbol{s}$ of all pairwise comparisons between the objects, where $s_{ijkl} = \kappa(\boldsymbol{x}_{ij}, \boldsymbol{x}_{kl})$
     (b) Store $\mathbf{s}$ from (a)
 2. Repeat step (1) $n.sims$ times and store in a matrix $\mathbf{S}$
 3. Calculate the empirical covariance matrix $\hat{\boldsymbol{\Sigma}}$ for $\boldsymbol{s}$ from $\mathbf{S}$
 4. Check that the eigenstructure of $\hat{\boldsymbol{\Sigma}}$ matches that of $\boldsymbol{\Sigma}$
```

The heatmap for $\boldsymbol{\Sigma}$ is given in Figure 5.1. In the heatmap, lighter colors represent larger covariance terms, mainly on the diagonal. The covariance matrix has a distinct structure that is easily recognizable in visual comparison.

Figure 5.1: Heatmap for covariance matrix $\boldsymbol{\Sigma}$ where lighter colors are larger covariance terms



5.3.1 Multivariate normal simulations

Checking the covariance structure obtained from a hierarchical multivariate normal sample was completed for varying sizes of dimension p. For the MVN distribution being sampled from, we used two different types of covariance structures; (1) diagonal matrices for independent dimensions and (2) banded matrices representing dependencies between dimensions. For the diagonal matrices, we define a between source variance σ_a^2 and within source variance σ_w^2 to obtain that

$$\begin{aligned} \boldsymbol{a}_i \sim & MVN\left(\boldsymbol{0}_p, \, \sigma_a^2 \boldsymbol{I}_p\right) \\ \boldsymbol{r}_{ij} \sim & MVN\left(\boldsymbol{a}_i, \, \sigma_r^2 \boldsymbol{I}_p\right). \end{aligned}$$

For each simulation iteration, $\boldsymbol{a}_i$ were sampled $n = 4$ times and $\boldsymbol{r}_{ij}$were sampled $n_o = 4$ times per $\boldsymbol{a}_i$.

A simple band matrix with a bandwidth of two was constructed for the second simulation test. The diagonal was set to be σ_a^2 and the elements in the band are set to be $\frac{1}{2}\sigma_a^2$. To obtain the within-source covariance matrix, $\mathbf{\Sigma}_r^{Band}$, the source-level matrix$\mathbf{\Sigma}_a^{Band}$ was scaled by a constant c such that $0 < c < 1$. An example for the covariance matrices $\mathbf{\Sigma}_a^{Band}$ and $\mathbf{\Sigma}_r^{Band}$ with $\sigma_a^2 = 10$ and $p = 5$ is given below.

$$\begin{aligned}\mathbf{\Sigma}_a^{Band} &= \begin{bmatrix} 10 & 5 & 5 & 0 & 0 \\ 5 & 10 & 5 & 5 & 0 \\ 5 & 5 & 10 & 5 & 5 \\ 0 & 5 & 5 & 10 & 5 \\ 0 & 0 & 5 & 5 & 10 \end{bmatrix} \\ \mathbf{\Sigma}_r^{Band} &= c\mathbf{\Sigma}_a^{Band}\end{aligned}$$

The resulting sampling distributions for generating $\boldsymbol{x}_{ij}$, $\boldsymbol{x}_{kl}$ are:

$$\begin{aligned}\boldsymbol{a}_i &\sim MVN\left(\mathbf{0}_p, \, \mathbf{\Sigma}_a^{Band}\right) \\ \boldsymbol{r}_{ij} &\sim MVN\left(\boldsymbol{a}_i, \, \mathbf{\Sigma}_r^{Band}\right).\end{aligned}$$

We tested using three kernels: squared Euclidean, exponential, and rational quadratic. They all gave similar results, reported in Table 5.10. These results were the same regardless of the dimension p used for either of the MVN distributions. Note that, for each different dimension of p the tuning parameters for the exponential and rational quadratic needed to be adjusted to scale the resulting score.

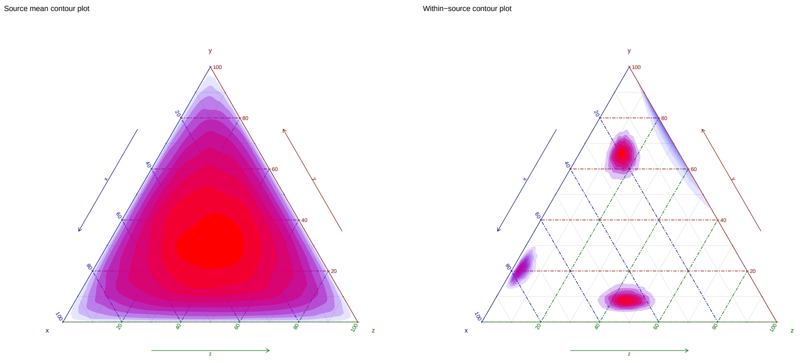
Table 5.10: Eigenvalue multiplicity for Multivariate-score simulation covariance. The counts for expected and observed eigenvalues are given in each column for respective kernels

Rank	Expected	Sq-Euclid.	Exponential	Rat. Quad
1	1	1	1	1
2	3	3	3	3
3	2	2	2	2
4	12	12	12	12
5	24	24	24	24
6	4	4	4	4
7	12	12	12	12
8	62	62	62	62

5.3.2 Non-normal simulations

To check the covariance structure obtained from a non-normal hierarchical sample, we used non-normal distributions to produce the objects $\boldsymbol{x}_{ij}$, $\boldsymbol{x}_{kl}$. In our first simulation, we used a Dirichlet distribution to produce samples. This distribution can be easily scaled up for any dimension p. The source distribution for $\boldsymbol{a}_i$ was chosen to have a weight vector with all entries equal to 0.5. To calculate within-source samples, a constant c was used to concentrate the within-source samples around $\boldsymbol{a}_i$ so that $\boldsymbol{x}_{ij} \sim Dirichlet\,(c\boldsymbol{a}_i)$. Examples for a three-dimensional Dirichlet distribution are given in Figure 5.2.

Figure 5.2: Density plots for source level and within-source level 3D Dirichlet distributions. On the left, the density for the distribution of $\boldsymbol{a}_i$ with highest density towards the middle. On the right, within-source densities for 4 sources.



We tested using three kernels: squared Euclidean, exponential, and rational quadratic. They all gave similar results, reported in Table 5.11. These results were the same regardless of the dimension p used for either of the MVN distributions. Note that, for each different dimension of p the tuning parameters for the exponential and rational quadratic needed to be adjusted to scale the resulting score.

Table 5.11: Eigenvalue multiplicity for Dirichlet-score simulation covariance. The counts for expected and observed eigenvalues are given in each column for respective kernels

Rank	Expected	Sq-Euclid.	Exponential	Rat. Quad
1	1	1	1	1
2	3	3	3	3
3	2	2	2	2
4	12	12	12	12
5	24	24	24	24
6	4	4	4	4
7	12	12	12	12
8	62	62	62	62

5.4 Multivariate score model form

Having a model for s_{ijkl}, we turned to studying its multivariate form. We first decomposed $\boldsymbol{\Sigma}$ in a manner similar to that used in the sample-driven model given in equation 5.4. It was found that the decomposition has the form,

$$\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^t\sigma_b^2 + \mathbf{D}\mathbf{D}^t\sigma_d^2 + \mathbf{T}\mathbf{T}^t\sigma_t^2 + \mathbf{W}\mathbf{W}^t\sigma_w^2 + \mathbf{I}_N\mathbf{1}_{(i\neq k)}\sigma_{e^b}^2 + \mathbf{I}_N\mathbf{1}_{(i=k)}\sigma_{e^w}^2$$

where $\mathbf{I}_N$ is an identity matrix of size $N = \binom{nn_o}{2}$, $\mathbf{B},\ \mathbf{D},\ \mathbf{T},\ \mathbf{W}$ are design matrices, and

$$\mathbf{1}_{(i\neq k)} = \begin{cases} 1 & i \neq k \\ 0 & otherwise \end{cases}$$

$$\mathbf{1}_{(i=k)} = \begin{cases} 1 & i = k \\ \\ 0 & otherwise \end{cases}$$

These design matrices were built as follows in the following Sections 5.4.1-5.4.4

5.4.1 Score-design matrix B

The design matrix **B** captures source effects b_i and b_k in the score model and $\mathbf{BB}^t$ gives the structure for the σ_b^2 components of $\mathbf{\Sigma}$. The structure for $\mathbf{BB}^t$ has the following rule for the r^{th} row and c^{th} column taken from observing where σ_b^2 shows up in the covariance components Table 5.9:

$$\mathbf{BB}^t_{r,c} = \begin{cases} 2 & \text{for } 1121 \ vs.\ 1121 \\ 2 & \text{for } 1121 \ vs.\ 1221 \\ 2 & \text{for } 1121 \ vs.\ 1222 \\ 1 & \text{for } 1121 \ vs.\ 1131 \\ 1 & \text{for } 1121 \ vs.\ 1231 \\ 0 & \text{otherwise} \end{cases}$$

which is simplified to 3 rules:

$$\mathbf{BB}^t_{r,c} = \begin{cases} 2 & \text{if } i \neq k \ \& \ i' \neq k' \ \& \ \#\{i,\, k,\, i',\, k'\}_{\neq} = 2 \\ 1 & \text{if } i \neq k \ \& \ i' \neq k' \ \& \ \#\{i,\, k,\, i',\, k'\}_{\neq} = 3 \\ 0 & \text{otherwise} \end{cases}$$

where $\#\{i,\, k,\, i',\, k'\}_{\neq}$ denotes the cardinality of the set of unique source indices between s_{ijkl} and $s_{i'j'k'l'}$. Using this, we were able to solve for the individual design matrix **B**, which was done by observing the number of sources combined together in the multiplication of the design matrices. It turns out that **B** has the same form as the **Q** matrix used in the sampling model covariance structure. It is denoted as **B** here to maintain consistency with the score model components. The design matrix is given in Table 5.12 for $n = 4$ and $n_o = 3$. In each row, when $i \neq k$, a 1 is entered in the i^{th} and k^{th} columns. If $i = k$, then that row contains all zeroes.

Table 5.12: Design matrix $\mathbf{B}$ for $n = 4$, $n_o = 3$

	1	2	3	4
1112	0	0	0	0
1113	0	0	0	0
1121	1	1	0	0
1122	1	1	0	0
1123	1	1	0	0
1131	1	0	1	0
1132	1	0	1	0
1133	1	0	1	0
1141	1	0	0	1
1142	1	0	0	1
1143	1	0	0	1
1213	0	0	0	0
1221	1	1	0	0
1222	1	1	0	0
1223	1	1	0	0
1231	1	0	1	0
1232	1	0	1	0
1233	1	0	1	0
1241	1	0	0	1
1242	1	0	0	1
1243	1	0	0	1
1321	1	1	0	0
1322	1	1	0	0
1323	1	1	0	0
1331	1	0	1	0
1332	1	0	1	0

The full form of $\mathbf{BB}^t$ is given in Appendix 10.3.5.

5.4.2 Score-design matrix D

The design matrix $\mathbf{D}$ describes the source interaction effect d_{ik} in the score model and $\mathbf{DD}^t$ gives the structure for the σ_d^2 components of $\mathbf{\Sigma}$. The structure for $\mathbf{DD}^t$ has the following rule for the r^{th} row and c^{th} column taken from observing where σ_b^2 shows up in the covariance components Table 5.9:

$$\mathbf{DD}^t_{r,c} = \begin{cases} 1 & \text{for } 1121 \textit{ vs. } 1121 \\ 1 & \text{for } 1121 \textit{ vs. } 1221 \\ 1 & \text{for } 1121 \textit{ vs. } 1222 \\ 0 & \text{otherwise} \end{cases}$$

which is simplified to 2 rules:

$$\mathbf{DD}^t_{r,c} = \begin{cases} 1 & \text{if } i \neq k \;\&\; i' \neq k' \;\&\; \#\{i,\, k,\, i',\, k'\}_{\neq} = 2 \\ 0 & \text{otherwise} \end{cases}$$

Using this, we were able to solve for the individual design matrix $\mathbf{D}$, which was done by observing the number of sources combined together in the multiplication of the design matrices. The design matrix $\mathbf{D}$ is a $N \times \binom{n}{2}$ matrix whose rows are the pairwise comparisons s_{ijkl} and the columns are the possible pairwise combinations of the source indices i and k for a sample with n sources. For example, if we had $n = 4$ sources, the columns would be: $12,\ 13,\ 14,\ 23,\ 24,\ 34$. In $\mathbf{D}$ a 1 is placed everywhere a source index combination ik of $ijkl$ shows up and the remaining elements are equal to zero. If $i = k$, then the row will consist of all zeroes since all scores represent within-source comparisons. An example of $\mathbf{D}$ with $n = 4$ and $n_o = 3$ is given in Table 5.13.

Table 5.13: Design matrix **D** for $n = 4$, $n_o = 3$

D	12	13	14	23	24	34
1112	0	0	0	0	0	0
1113	0	0	0	0	0	0
1121	1	0	0	0	0	0
1122	1	0	0	0	0	0
1123	1	0	0	0	0	0
1131	0	1	0	0	0	0
1132	0	1	0	0	0	0
1133	0	1	0	0	0	0
1141	0	0	1	0	0	0
1142	0	0	1	0	0	0
1143	0	0	1	0	0	0
1213	0	0	0	0	0	0
1221	1	0	0	0	0	0
1222	1	0	0	0	0	0
1223	1	0	0	0	0	0
1231	0	1	0	0	0	0
1232	0	1	0	0	0	0
1233	0	1	0	0	0	0
1241	0	0	1	0	0	0
1242	0	0	1	0	0	0
1243	0	0	1	0	0	0
1321	1	0	0	0	0	0
1322	1	0	0	0	0	0
1323	1	0	0	0	0	0
1331	0	1	0	0	0	0
1332	0	1	0	0	0	0

An example of the full $\mathbf{DD}^t$ for $n = 4$ and $n_o = 3$ is given in Appendix 10.3.6.

5.4.3 Score-design matrix T

The design matrix **T** describes the interaction between source and samples $t_{i:ij}$, $t_{i:kl}$, $t_{k:ij}$, and $t_{k:kl}$ in the score model and $\mathbf{TT}^t$ gives the structure for the σ_t^2 components of $\boldsymbol{\Sigma}$.

$$\mathbf{TT}^t_{r,c} = \begin{cases} 4 & \text{for } 1121 \ vs. \ 1121 \\ 2 & \text{for } 1121 \ vs. \ 1221 \\ 1 & \text{for } 1121 \ vs. \ 1131 \\ 0 & \text{otherwise} \end{cases}$$

which is given by 4 rules:

$$\mathbf{TT}^t_{r,c} = \begin{cases} 4 & \text{if } ijkl = i'j'k'l' \ \& \ i \neq k \\ 2 & \text{if } i \neq k \ \& \ i' \neq k' \ \& \ \# \{i, \, k, \, i', \, k'\}_{\neq} = 2 \\ 1 & \text{if } i \neq k \ \& \ i' \neq k' \ \& \ \# \{i, \, k, \, i', \, k'\}_{\neq} = 3 \\ 0 & \text{otherwise} \end{cases}$$

For a single score s_{ijkl}, $\mathbf{T}$ is a $N \times (n^2 n_o)$ matrix where the rows are the pairwise comparisons s_{ijkl} and the columns are the combinations $i : ij, \ i : kl, \ k : ij, \ k : kl$ for all sources and samples. In $\mathbf{T}$, a 1 is placed everywhere an index combination $i : ij, \ i : kl, \ k : ij, \ k : kl$ of $ijkl$ shows up. For example, if we have the combination $ijkl = 1122$, the columns for combinations $1 : 11, \ 1 : 22, \ 2 : 11, \ 2 : 22$ would have a 1. If $i = k$, then the row will consist of all zeroes since all scores represent within-source comparisons. We give a short example of the design matrix $\mathbf{T}$ for $n = 3$, $n_o = 2$ in Table 5.14.

Table 5.14: Design matrix $\mathbf{T}$

	1:11	1:12	1:21	1:22	1:31	1:32	2:11	2:12	2:21	2:22	2:31	2:32	3:11	3:12	3:21	3:22	3:31	3:32
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
1122	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
1131	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
1132	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
1221	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
1222	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0
1231	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0
1232	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0
2132	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1
2231	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0
2232	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

A full example of $\mathbf{TT}^t$ for $n = 3$, $n_o = 2$ can be found in Appendix 10.3.7.

5.4.4 Score-design matrix W

The design matrix **W** describes the within source effect w_{ij} and w_{kl} in the score model and $\mathbf{W}\mathbf{W}^t$ gives the structure for the σ_w^2 components of $\mathbf{\Sigma}$. It was found that the design matrix **W** is a larger version of **P** using nn_o samples instead of just n_o samples. For the design matrix **W**, the number of columns is equal to nn_o and contains the indices ij for all objects. The rows are the pairwise comparisons s_{ijkl}. Each row contains all zeroes except for a one in the ij^{th} and kl^{th} columns. An example for $n = 4$ and $n_o = 3$ is given in Table 5.15.

Table 5.15: Design matrix **W** for $n = 4$, $n_o = 3$

	11	12	13	21	22	23	31	32	33	41	42	43
1112	1	1	0	0	0	0	0	0	0	0	0	0
1113	1	0	1	0	0	0	0	0	0	0	0	0
1121	1	0	0	1	0	0	0	0	0	0	0	0
1122	1	0	0	0	1	0	0	0	0	0	0	0
1123	1	0	0	0	0	1	0	0	0	0	0	0
1131	1	0	0	0	0	0	1	0	0	0	0	0
1132	1	0	0	0	0	0	0	1	0	0	0	0
1133	1	0	0	0	0	0	0	0	1	0	0	0
1141	1	0	0	0	0	0	0	0	0	1	0	0
1142	1	0	0	0	0	0	0	0	0	0	1	0
1143	1	0	0	0	0	0	0	0	0	0	0	1
1213	0	1	1	0	0	0	0	0	0	0	0	0
1221	0	1	0	1	0	0	0	0	0	0	0	0
1222	0	1	0	0	1	0	0	0	0	0	0	0
1223	0	1	0	0	0	1	0	0	0	0	0	0
1231	0	1	0	0	0	0	1	0	0	0	0	0
1232	0	1	0	0	0	0	0	1	0	0	0	0
1233	0	1	0	0	0	0	0	0	1	0	0	0
1241	0	1	0	0	0	0	0	0	0	1	0	0
1242	0	1	0	0	0	0	0	0	0	0	1	0
1243	0	1	0	0	0	0	0	0	0	0	0	1
1321	0	0	1	1	0	0	0	0	0	0	0	0
1322	0	0	1	0	1	0	0	0	0	0	0	0

5.5 Calculating estimates for model parameters

Up to this point, we have defined a covariance structure that holds for stationary kernels under certain conditions (see above) and proposed a linear model for the score s_{ijkl}. Our next step is to estimate the parameters for the model. A technique to estimate the variance parameters for the within-source model $(i = k)$ was described by [24, 4]. We turn our focus to the between-source model $(i \neq k)$ and its parameters. We start by subsetting the scores so that we have only $N_b = n\binom{n_o}{2}$ between-source scores and assign them to the score vector $\boldsymbol{s}_b$, which will have a covariance matrix $\mathbf{\Sigma}_b$,

$$\mathbf{\Sigma}_b = \mathbf{B}^*\mathbf{B}^{*t}\sigma_b^2 + \mathbf{D}^*\mathbf{D}^{*t}\sigma_d^2 + \mathbf{T}^*\mathbf{T}^{*t}\sigma_t^2 + \mathbf{W}^*\mathbf{W}^{*t}\sigma_w^2 + \mathbf{I}_{N_b}\sigma_{e^b}^2 \tag{5.6}$$

where $\mathbf{B}^*,\ \mathbf{D}^*,\ \mathbf{T}^*,\ \mathbf{W}^*,\ \mathbf{I}_{N_b}$ are the design matrices for between-source scores only. These design matrices correspond directly to the matrices $\mathbf{B},\ \mathbf{D},\ \mathbf{T},\ \mathbf{W},\ \mathbf{I}_N$ where the rows for the within scores were removed. We began by working with the likelihood for the multivariate normal distribution as given in equation 5.7. We expressed the likelihood of the multivariate normal in terms of the eigenvalues and eigenvectors of the covariance matrix $\mathbf{\Sigma}_b$ in order to reduce computational complexity. The likelihood function for the set of parameters $\psi = \left\{\sigma_b^2\ \sigma_d^2,\ \sigma_t^2,\ \sigma_w^2,\ \sigma_{e^b}^2,\ \theta_b\right\}$, using the eigenvalue representations for $|\mathbf{\Sigma}_b|$ and $\mathbf{\Sigma}_b^{-1}$ [33], is:

$$\begin{aligned}
-2l\left(\psi|\mathbf{s}_b\right) =& ln\left(|\mathbf{\Sigma}_b|\right) + \left(\mathbf{s}_b - \theta_b\mathbf{1}_{N_b}\right)^t \mathbf{\Sigma}_b^{-1}\left(\mathbf{s}_b - \theta_b\mathbf{1}_{N_b}\right) + N_b ln\left(2\pi\right) \\
=& ln\left(\prod_{v=1}^{N_b}\lambda_v\right) + \left(\mathbf{s}_b - \theta_b\mathbf{1}_{N_b}\right)^t \sum_{v=1}^{N_b}\lambda_v^{-1}\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b - \theta_b\mathbf{1}_{N_b}\right) + N_b ln\left(2\pi\right) \qquad (5.7) \\
=& \sum_{v=1}^{N_b}\left[ln\left(\lambda_v\right) + \lambda^{-1}\left(\left(\boldsymbol{s}_b - \theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\right)^2 + ln\left(2\pi\right)\right]. \qquad (5.8)
\end{aligned}$$

The number of eigenvalues and their respective multiplicity is given in Table 5.17. This is used to break equation 5.7 into smaller components. Additionally, in Table 5.16, the end points $a_2, ..., a_6$ for the sets of eigenvalues are given. These are the right side indices for the sets of eigenvalues $\lambda_2, ..., \lambda_6$ and their respective eigenvectors $\mathbf{v}_v$. For example, λ_2 is repeated from $[2, a_2]$. These help to organize our work and with programming later.

Table 5.16: Cut points for sets of eigenvalues

endpoint	Value
a_1	1
a_2	n
a_3	nn_o
a_4	$n\left(\frac{2n_o+n-3}{2}\right)$
a_5	$n\left(\frac{(2n_o-1)(n-1)}{2}\right)$
a_6	N_b

Table 5.17: Counts of eigenvalues for $\mathbf{\Sigma}_b$

Rank	# roots
λ_1	1
λ_2	$n-1$
λ_3	$n\left(n_o-1\right)$
λ_4	$\binom{n}{2}-n$
λ_5	$n\left(n-2\right)\left(n_o-1\right)$
λ_6	$N_b-\sum_{k=1}^{5} N_{\lambda_k}$

We considered the first two summands (the third is a constant) separately. The first summand is given in equation 5.9:

$$\begin{aligned} ln\left(\prod_{v=1}^{N_b}\lambda_v\right) =& ln\left(\lambda_1\right)+\left(n-1\right)ln\left(\lambda_2\right)+\left(n\left(n_o-1\right)\right)ln\left(\lambda_3\right) \\ &+\left(\binom{n}{2}-n\right)ln\left(\lambda_4\right)+\left(n\left(n-2\right)\left(n_o-1\right)\right)ln\left(\lambda_5\right)+\left(N_b-\sum_{k=1}^{5}N_{\lambda_k}\right)ln\left(\lambda_6\right) \end{aligned} \tag{5.9}$$

and the second summand is given in equation 5.10:

$$
\begin{aligned}
\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\sum_{v=1}^{N_b}\lambda^{-1}\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)=&\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_1\mathbf{v}_1^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_1}\\
&+\sum_{v=2}^{a_2}\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_2}\\
&+\sum_{v=a_2+1}^{a_3}\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_3}\\
&+\sum_{v=a_3+1}^{a_4}\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_4}\\
&+\sum_{v=a_4+1}^{a_5}\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_5}\\
&+\sum_{v=a_5+1}^{a_6}\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_6}
\end{aligned}
\tag{5.10}
$$

Equation 5.10 is simplified further by the knowledge that $\mathbf{v}_1=\frac{\mathbf{1}_{N_b}}{\sqrt{N_b}}$ which implies for $v\neq 1$, $\mathbf{1}_{N_b}^t\mathbf{v}_v=0$:

$$
\begin{aligned}
\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\sum_{v=1}^{N_b}\lambda^{-1}\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)=&\frac{\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\mathbf{v}_1\mathbf{v}_1^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)}{\lambda_1}\\
&+\sum_{v=2}^{a_2}\frac{\left(\mathbf{s}_b^t\mathbf{v}_v\right)^2}{\lambda_2}\\
&+\sum_{v=a_2+1}^{a_3}\frac{\left(\mathbf{s}_b^t\mathbf{v}_v\right)^2}{\lambda_3}\\
&+\sum_{v=a_3+1}^{a_4}\frac{\left(\mathbf{s}_b^t\mathbf{v}_v\right)^2}{\lambda_4}\\
&+\sum_{v=a_4+1}^{a_5}\frac{\left(\mathbf{s}_b^t\mathbf{v}_v\right)^2}{\lambda_5}\\
&+\sum_{v=a_5+1}^{a_6}\frac{\left(\mathbf{s}_b^t\mathbf{v}_v\right)^2}{\lambda_6}.
\end{aligned}
\tag{5.11}
$$

The numerator of the first summand of equation 5.11 can be simplified further:

$$\begin{aligned}
&\left(\mathbf{s}_b - \theta_b \mathbf{1}_{N_b}\right)^t \mathbf{v}_1 \mathbf{v}_1^t \left(\mathbf{s}_b - \theta_b \mathbf{1}_{N_b}\right)\\
&= \left(\mathbf{v}_1^t \left(\mathbf{s}_b - \theta_b \mathbf{1}_{N_b}\right)\right)^t \left(\mathbf{v}_1^t \left(\mathbf{s}_b - \theta_b \mathbf{1}_{N_b}\right)\right)\\
&= \left(\mathbf{v}_1^t \left(\mathbf{s}_b - \theta_b \mathbf{1}_{N_b}\right)\right)^2\\
&= \left(\mathbf{s}_b^t \mathbf{v}_1 - \theta_b \mathbf{1}_{N_b}^t \mathbf{v}_1\right)^2\\
&= \left(\frac{\mathbf{s}_b^t \mathbf{1}_{N_b}}{\sqrt{N_b}} - \frac{\theta_b \mathbf{1}_{N_b}^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)^2\\
&= \left(\frac{\mathbf{s}_b^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)^2 - \left(\frac{\mathbf{s}_b^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)\left(\frac{\theta_b \mathbf{1}_{N_b}^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)\\
&\quad - \left(\frac{\theta_b \mathbf{1}_{N_b}^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)\left(\frac{\mathbf{s}_b^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right) + \left(\frac{\theta_b \mathbf{1}_{N_b}^t \mathbf{1}_{N_b}}{\sqrt{N_b}}\right)^2\\
&= \left(\frac{\sum_{i=1}^{N_b} s_{bi}}{\sqrt{N_b}}\right)^2 - \left(\frac{\sum_{i=1}^{N_b} s_{bi}}{\sqrt{N_b}}\right)\left(\frac{N_b \theta_b}{\sqrt{N_b}}\right)\\
&\quad - \left(\frac{N_b \theta_b}{\sqrt{N_b}}\right)\left(\frac{\sum_{i=1}^{N_b} s_{bi}}{\sqrt{N_b}}\right) + \left(\frac{N_b \theta_b}{\sqrt{N_b}}\right)^2\\
&= \frac{\left(\sum_{i=1}^{N_b} s_{bi}\right)^2}{N_b} - 2\theta_b \sum_{i=1}^{N_b} s_{bi} + N_b \theta_b^2\\
&= N_b \left(\frac{\left(\sum_{i=1}^{N_b} s_{bi}\right)^2}{N_b^2} - \frac{2\theta_b \sum_{i=1}^{N_b} s_{bi}}{N_b} + \theta_b^2\right)\\
&= N_b \left(\bar{s_b}^2 - 2\theta_b \bar{s_b} + \theta_b^2\right)\\
&= N_b \left(\bar{s_b} - \theta_b\right)^2 \qquad (5.12)
\end{aligned}$$

Upon further inspection of equation 5.11, we note that 1) we can decompose the likelihood into sums of squares terms $\sum \mathbf{s}_b^t \mathbf{v}_v \mathbf{v}_v^t \mathbf{s}_b = \sum \left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$; 2) we need to find the form of the eigenvectors when $v \neq 1$; 3) we need assign values to $\lambda_1, ..., \lambda_6$. To accomplish this, we used the multiplicity of the eigenvalues and that the sets of projections $\mathbf{v}_v^t \mathbf{s}_b$ for a particular eigenvalue λ_v will have variance equal to λ_v. For $v \neq 1$:

$$\begin{aligned}\mathbf{s}_b \sim MVN\left(\theta_b \mathbf{1}_{N_b},\ \mathbf{\Sigma}_b\right) &\implies \\ \mathbf{v}_v^t \mathbf{s}_b \sim N\left(\theta_b \mathbf{v}_v^t \mathbf{1}_{N_b},\ \mathbf{v}_v^t \mathbf{\Sigma}_b \mathbf{v}_v\right) &\iff \\ \mathbf{v}_v^t \mathbf{s}_b \sim N\left(0,\ \lambda_v\right) & \end{aligned} \tag{5.13}$$

where λ_v is the eigenvalue for the v^{th} eigenvector. We can also expand $\mathbf{\Sigma}_b$ by using the results from equation 5.6.

$$\begin{aligned}\mathbf{v}_v^t \mathbf{s}_b \sim & N\left(0,\ \lambda_v\right) \\ &= N\left(0,\ \mathbf{v}_v^t \left(\mathbf{B}^*\mathbf{B}^{*t}\sigma_b^2 + \mathbf{D}^*\mathbf{D}^{*t}\sigma_d^2 + \mathbf{T}^*\mathbf{T}^{*t}\sigma_t^2 + \mathbf{W}^*\mathbf{W}^{*t}\sigma_w^2 + \mathbf{I}_{N_b}\sigma_{e^b}^2\right)\mathbf{v}_v\right) \\ &= N\left(0,\ \mathbf{v}_v^t\mathbf{B}^*\mathbf{B}^{*t}\mathbf{v}_v\sigma_b^2 + \mathbf{v}_v^t\mathbf{D}^*\mathbf{D}^{*t}\mathbf{v}_v\sigma_d^2 + \mathbf{v}_v^t\mathbf{T}^*\mathbf{T}^{*t}\mathbf{v}_v\sigma_t^2 + \mathbf{v}_v^t\mathbf{W}^*\mathbf{W}^{*t}\mathbf{v}_v\sigma_w^2 + \mathbf{v}_v^t\mathbf{I}_{N_b}\mathbf{v}_v\sigma_{e^b}^2\right)\end{aligned} \tag{5.14}$$

Therefore, if $\boldsymbol{v}_v$ is an eigenvector of $\mathbf{B}^*\mathbf{B}^{*t},\ \mathbf{D}^*\mathbf{D}^{*t},\ \mathbf{T}^*\mathbf{T}^{*t},\ \mathbf{W}^*\mathbf{W}^{*t},\ \mathbf{I}_{N_b}$, then λ_v is simply the sum of the corresponding eigenvalues of the components of $\mathbf{\Sigma}_b$, since for two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same eigenvector, $\mathbf{v}$, and respective eigenvalues $\lambda_A,\ \lambda_B$ then,

$$\left(\mathbf{A}+\mathbf{B}\right)\mathbf{v} = \mathbf{A}\mathbf{v} + \mathbf{B}\mathbf{v} = \lambda_A\mathbf{v} + \lambda_B\mathbf{v} = \left(\lambda_A + \lambda_B\right)\mathbf{v} \tag{5.15}$$

We also see that λ_v is a function of the design matrices multipled by the parameters of interest. Hence, if we know the eigenvalues for $\mathbf{\Sigma}_b$ and its components for different values of v, we may be able to provide estimates for the variance components of the model.

5.5.1 Eigenstructure of Σ_b

Let $\mathbf{A} = \sum \mathbf{A}_i$ be a matrix (linear operator) that we wish to diagonalize and $\mathbf{V} = \{\mathbf{v}_1, ..., \mathbf{v}_v\}$ a set of eigenvectors of the vector space P, then:

Theorem 5.2. *Simultaneous Diagonalization [15]*

If $\mathbf{A}_1, ..., \mathbf{A}_r$ are linear operators on P and each $\mathbf{A}_i$ is diagonalizable, they are simultaneously diagonalizable if and only if they commute.

The conditions to use simultaneous diagonalization are thus:

1. Each matrix $\mathbf{A}_1, ..., \mathbf{A}_r$ can be diagonalized, as per definition 5.3;

2. All the matrices $\mathbf{A}_1, ..., \mathbf{A}_r$ pairwise commute.

Definition 5.3. Diagonalization

The linear operator $\mathbf{A} : P \to P$ is diagonalizable when it admits a diagonal matrix representation with respect to some basis of P. In other words, there is a basis $\mathbf{V} = \{\mathbf{v}_1, ..., \mathbf{v}_v\}$ of P such that the matrix $[\mathbf{A}]_\mathbf{V}$ is diagonal.

We can show that all the matrices $\mathbf{B}^*\mathbf{B}^{*t},\ \mathbf{D}^*\mathbf{D}^{*t},\ \mathbf{T}^*\mathbf{T}^{*t},\ \mathbf{W}^*\mathbf{W}^{*t},\ \mathbf{I}_{N_b}$, pairwise commute and are individually diagonalizable with respect to some $\mathbf{V}$, thus there exists a of the vector space P such that $[\mathbf{\Sigma}_b]_\mathbf{V} = \mathbf{V}^{-1}\mathbf{\Sigma_b}\mathbf{V}$ and

$$[\mathbf{\Sigma}_b]_\mathbf{V} = \left[\mathbf{B}^*\mathbf{B}^{*t}\right]_\mathbf{V} \sigma_b^2 + \left[\mathbf{D}^*\mathbf{D}^{*t}\right]_\mathbf{V} \sigma_d^2 + \left[\mathbf{T}^*\mathbf{T}^{*t}\right]_\mathbf{V} \sigma_t^2 + \left[\mathbf{W}^*\mathbf{W}^{*t}\right]_\mathbf{V} \sigma_w^2 + [\mathbf{I}_{N_b}]_\mathbf{V} \sigma_{e^b}^2. \tag{5.16}$$

It turned out that $\mathbf{\Sigma}_b$ shares the same eigenvectors as the design covariance matrix $\mathbf{\Gamma}$, defined as:

$$\mathbf{\Gamma} = \mathbf{B}^*\mathbf{B}^{*t} + \mathbf{D}^*\mathbf{D}^{*t} + \mathbf{T}^*\mathbf{T}^{*t} + \mathbf{W}^*\mathbf{W}^{*t} + \mathbf{I}_{N_b}. \tag{5.17}$$

which enables us to find a set of eigenvectors $\mathbf{V}$ from $\mathbf{\Gamma}$ irrespectively of the values of the parameters of interest. It also allows us to obtain the corresponding eigenvalues of the components of $\mathbf{\Sigma}_b$.

The eigenvalues of the diagonalized design matrices using $\mathbf{V}$ are given in Table 5.18. These eigenvalues were determined empirically by generating $\mathbf{\Gamma}$ for different values of n and n_o, numerically obtaining its eigenvectors $\mathbf{V}$, and looking for patterns in the eigenvalues of the diagonalized components of $\mathbf{\Gamma}$.

Upon inspection, we note that several eigenvalues are 0. These correspond to eigenvectors that live in the nullspace for the particular component of interest. However, because of the identity matrix, there is no nullspace of $\mathbf{\Sigma}_b$ once the diagonalized components have been summed.

The diagonalized components of $\mathbf{\Sigma}_b$ in Table 5.18 are linearly combined to obtain the eigenvalues of $\mathbf{\Sigma}_b$. The representations of the diagonl matrices that are added together to obtain the diagonal matrix $[\mathbf{\Sigma}_b]_{\mathbf{V}}$ is given in Equation 5.18. These eigenvalues are given in Table 5.19.

Table 5.18: Corresponding eigenvalues of the diagonalized components of $\mathbf{\Sigma}_b$ using common basis $\mathbf{V}$

$\mathbf{V}^{-1}\mathbf{B}^{*}\mathbf{B}^{*t}\mathbf{V}\sigma_b^2$		$\mathbf{V}^{-1}\mathbf{D}^{*}\mathbf{D}^{*t}\mathbf{V}\sigma_d^2$		$\mathbf{V}^{-1}\mathbf{T}^{*}\mathbf{T}^{*t}\mathbf{V}\sigma_t^2$		$\mathbf{V}^{-1}\mathbf{W}^{*}\mathbf{W}^{*t}\mathbf{V}\sigma_w^2$		$\mathbf{V}^{-1}\mathbf{I}_{N_b}\mathbf{V}\sigma_{e^b}^2$	
λ	# roots	λ	# roots	λ	# roots	λ	# roots	λ	# roots
$2\,(n-1)\,n_o^2\sigma_b^2$	1	$n_o^2\sigma_d^2$	n	$2nn_o\sigma_t^2$	1	$2\,(n-1)\,n_o\sigma_w^2$	1	$\sigma_{e^b}^2$	N_b
$(n-2)\,n_o^2\sigma_b^2$	$n-1$	0	$n\,(n_o-1)$	$nn_o\sigma_t^2$	nn_o-1	$(n-2)\,n_o\sigma_w^2$	$n-1$		
0	N_b-n	$n_o^2\sigma_d^2$	$\binom{n}{2}-n$	$2n_o\sigma_t^2$	$\binom{n-1}{2}-1$	$(n-1)\,n_o\sigma_w^2$	$n\,(n_o-1)$		
		0	$N_b-n(n_o-1)-\binom{n}{2}$	$n_o\sigma_t^2$	$n\,(n-2)\,(n_o-1)$	0	N_b-nn_o		
				0	$N_b-(n-1)\,(nn_o-1)$				

$$
\begin{bmatrix}
2\left(n-1\right)n_o^2\sigma_b^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \left(n-2\right)n_o^2\sigma_b^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \ddots & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
+\begin{bmatrix}
n_o^2\sigma_d^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & n_o^2\sigma_d^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
+\begin{bmatrix}
2nn_o\sigma_t^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & nn_o\sigma_t^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2n_o\sigma_t^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & n_o\sigma_t^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
+\begin{bmatrix}
2\left(n-1\right)n_o\sigma_w^2 & 0 & 0 & 0 & 0 & 0 \\
0 & \left(n-2\right)n_o\sigma_w^2 & 0 & 0 & 0 & 0 \\
0 & 0 & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & \left(n-1\right)n_o\sigma_w^2 & 0 & 0 \\
0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
+\begin{bmatrix}
\sigma_{e^b}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \ddots & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma_{e^b}^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \ddots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma_{e^b}^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e^b}^2
\end{bmatrix}
\tag{5.18}
$$

Table 5.19: Eigenvalues of $\mathbf{\Sigma}_b$

Rank	# roots	λ	Starting Index	Ending Index
λ_1	1	$2\left(n-1\right)n_o^2\sigma_b^2+n_o^2\sigma_d^2+2nn_o\sigma_t^2+2\left(n-1\right)n_o\sigma_w^2+\sigma_{e^b}^2$	1	1
λ_2	$n-1$	$\left(n-2\right)n_o^2\sigma_b^2+n_o^2\sigma_d^2+nn_o\sigma_t^2+\left(n-2\right)n_o\sigma_w^2+\sigma_{e^b}^2$	2	n
λ_3	$n\left(n_o-1\right)$	$nn_o\sigma_t^2+\left(n-1\right)n_o\sigma_w^2+\sigma_{e^b}^2$	$n+1$	nn_o
λ_4	$\binom{n}{2}-n$	$n_o^2\sigma_d^2+2n_o\sigma_t^2+\sigma_{e^b}^2$	nn_o+1	$n\left(\frac{2n_o+n-3}{2}\right)$
λ_5	$n\left(n-2\right)\left(n_o-1\right)$	$n_o\sigma_t^2+\sigma_{e^b}^2$	$n\left(\frac{2n_o+n-3}{2}\right)+1$	$n\left(\frac{(2n_o-1)(n-1)}{2}\right)$
λ_6	$N_b-\sum_{k=1}^{5}N_{\lambda_k}$	$\sigma_{e^b}^2$	$n\left(\frac{(2n_o-1)(n-1)}{2}\right)+1$	N_b

The REML estimates for the variance parameters $\hat{\sigma}_b^2$ $\hat{\sigma}_d^2$, $\hat{\sigma}_t^2$, $\hat{\sigma}_w^2$, $\hat{\sigma}_{e^b}^2$, $\hat{\sigma}_{e^w}^2$ and mean parameters for $\hat{\theta}_b$, $\hat{\theta}_w$ can be calculated as described in the next sections.

5.5.2 Mean parameters estimation

The within-source score mean $\hat{\theta}_w$ for all s_{ijkl} such that $i=k$ is estimated by

$$\hat{\theta}_w=\frac{\sum^N s_{ijkl}I_{(i=k)}}{N_w}$$

where $N=\binom{nn_0}{2}$ the total number of scores, $N_w=n\binom{n_o}{2}$ the total number of within-source scores, and $I_{(i=k)}$ is an indicator function for when a within-source score is used. The between-source score mean $\hat{\theta}_b$ for all s_{ijkl} such that $i\neq k$ is estimated by

$$\hat{\theta}_b=\frac{\sum^N s_{ijkl}I_{(i\neq k)}}{N_b}$$

where $N_b=N-N_w$. The results are then combined to create the mean vector $\hat{\boldsymbol{\theta}}$ for the full score model:

$$\hat{\boldsymbol{\theta}}=\hat{\theta}_w\mathbf{1}_{\boldsymbol{N}(i=k)}+\hat{\theta}_b\mathbf{1}_{\boldsymbol{N}(i\neq k)} \tag{5.19}$$

5.5.3 Variance parameters estimation

To solve for the REML estimates for the variance parameters, we use the same strategy described in section 5.5.1. At first, the design covariance matrix $\mathbf{\Gamma}$ is created for the appropriate number of sources n and objects per source n_o[13]. A set

[13] Note that the model requires a balanced sample in order to estimate the parameters.

of eigenvectors $\mathbf{V}$ is obtained for $\mathbf{\Gamma}$. This set of eigenvectors can be divided into the six sets given in Table 5.19 based on the multiplicity of the eigenvalues $\lambda_1, ..., \lambda_6$. Each set of eigenvectors can be used to obtain the projections $\mathbf{v}_v^t \mathbf{s}_b$ and $\sum_v \left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$ where v is an index of the eigenvectors belonging to a particular λ_r-eigenspace $(r = 1, ..., 6)$.

Projections using eigenvectors for the same eigenvalue will belong to the same subspace and as a result have the same variance. We use this to calculate the sums of squares $\sum \left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$ for each of these sets in order to estimate their respective variances. The resulting sums of squares will be $SS_2,\ SS_3,\ SS_4,\ SS_5,\ SS_6$. An example of SS_2 and SS_3 is given here:

$$SS_2 = \sum_{v=2}^{n} \left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$$

$$SS_3 = \sum_{v=n+1}^{nn_o} \left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$$

and the remaining sums of squares are calculated in a similar fashion for their respective set. We can now complete the second summand of the likelihood function from equation 5.11

$$\begin{aligned}
&\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)^t\sum_{v=1}^{N_b}\lambda^{-1}\mathbf{v}_v\mathbf{v}_v^t\left(\mathbf{s}_b-\theta_b\mathbf{1}_{N_b}\right)\\
=&\frac{N_b\left(\bar{s_b}-\theta_b\right)^2}{2\left(n-1\right)n_o^2\sigma_b^2+n_o^2\sigma_d^2+2nn_o\sigma_t^2+2\left(n-1\right)n_o\sigma_w^2+\sigma_{e^b}^2}\\
+&\frac{SS_2}{\left(n-2\right)n_o^2\sigma_b^2+n_o^2\sigma_d^2+nn_o\sigma_t^2+\left(n-2\right)n_o\sigma_w^2+\sigma_{e^b}^2}\\
+&\frac{SS_3}{nn_o\sigma_t^2+\left(n-1\right)n_o\sigma_w^2+\sigma_{e^b}^2}\\
+&\frac{SS_4}{n_o^2\sigma_d^2+2n_o\sigma_t^2+\sigma_{e^b}^2}\\
+&\frac{SS_5}{n_o\sigma_t^2+\sigma_{e^b}^2}\\
+&\frac{SS_6}{\sigma_{e^b}^2}
\end{aligned}\tag{5.20}$$

We calculated the ANOVA Table for the likelihood function in 5.20 to obtain the expected mean squares from the different sources. The Table is given below in 5.20. We give examples of the calculation of $E\left(MS_6\right)$ and $E\left(MS_5\right)$, used in the creation of the ANOVA Table. Here, $N_6=N_b-\sum_{k=1}^{5}N_{\lambda_k}$ and $N_5=n\left(n-2\right)\left(n_o-1\right)$. Note that because we have independent normal projections $\mathbf{v}_v^t\mathbf{s}_b$, we can calculate these for each λ-eigenspace.

$$\begin{aligned}
\frac{SS_6}{\sigma_{e^b}^2}&\sim\chi^2_{df=N_6}\\
E\left(\frac{SS_6}{\sigma_{e^b}^2}\right)&=N_6\\
E\left(SS_6\right)&=\sigma_{e^b}^2N_6\\
\frac{E\left(SS_6\right)}{N_6}&=\sigma_{e^b}^2\\
E\left(MS_6\right)&=\sigma_{e^b}^2
\end{aligned}$$

$$\begin{aligned} \frac{SS_5}{n_o\sigma_t^2 + \sigma_{e^b}^2} &\sim \chi^2_{df=N_5} \\ E\left(\frac{SS_5}{n_o\sigma_t^2 + \sigma_{e^b}^2}\right) &= N_5 \\ E\left(SS_5\right) &= \left(n_o\sigma_t^2 + \sigma_{e^b}^2\right) N_5 \\ \frac{E\left(SS_5\right)}{N_5} &= n_o\sigma_t^2 + \sigma_{e^b}^2 \\ E\left(MS_5\right) &= n_o\sigma_t^2 + \sigma_{e^b}^2 \end{aligned}$$

The ANOVA Table for the effects of the between-source score model is given below in Table 5.20 for the effects $b,\ w,\ d,\ t,\ e^b$. Note that the ordering of the effects is based off of the ordering of the sums of squares in equation 5.20.

Table 5.20: ANOVA Table for estimates

Source	SS	df	$E\left(MS\right)$	
b	SS_2	$n-1$	$(n-2)\,n_o^2\sigma_b^2 + n_o^2\sigma_d^2 + nn_o\sigma_t^2 + (n-2)\,n_o\sigma_w^2 + \sigma_{e^b}^2$	η_2
w	SS_3	$n\left(n_o-1\right)$	$nn_o\sigma_t^2 + (n-1)\,n_o\sigma_w^2 + \sigma_{e^b}^2$	η_3
d	SS_4	$\binom{n}{2} - n$	$n_o^2\sigma_d^2 + 2n_o\sigma_t^2 + \sigma_{e^b}^2$	η_4
t	SS_5	$n\left(n-2\right)\left(n_o-1\right)$	$n_o\sigma_t^2 + \sigma_{e^b}^2$	η_5
e^b	SS_6	$N_b - \sum_{k=1}^{5} N_{\lambda_k}$	$\sigma_{e^b}^2$	η_6

Calculating the REMLs can be done by solving the following system of equations:

$$\begin{bmatrix} (n-2)\,n_o^2 & n_o^2 & (n-2)\,n_o & nn_o & 1 \\ 0 & 0 & (n-1)\,n_o & nn_o & 1 \\ 0 & n_o^2 & 0 & n_o^2 & 1 \\ 0 & 0 & 0 & n_0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_b^2 \\ \hat{\sigma}_w^2 \\ \hat{\sigma}_d^2 \\ \hat{\sigma}_t^2 \\ \hat{\sigma}_{e^b}^2 \end{bmatrix} = \begin{bmatrix} MS_2 \\ MS_3 \\ MS_4 \\ MS_5 \\ MS_6 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\sigma}_b^2 \\ \hat{\sigma}_w^2 \\ \hat{\sigma}_d^2 \\ \hat{\sigma}_t^2 \\ \hat{\sigma}_{e^b}^2 \end{bmatrix} = \begin{bmatrix} (n-2)\,n_o^2 & n_o^2 & (n-2)\,n_o & nn_o & 1 \\ 0 & 0 & (n-1)\,n_o & nn_o & 1 \\ 0 & n_o^2 & 0 & n_o^2 & 1 \\ 0 & 0 & 0 & n_0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} MS_2 \\ MS_3 \\ MS_4 \\ MS_5 \\ MS_6 \end{bmatrix} \tag{5.21}$$

However, this can only be accomplished when the 5 × 5 matrix of weights in equation 5.21 is invertable[14]. If we cannot invert that matrix, it is possible to use the method of moments estimators below in a stepwise manner, while ensuring that all estimates remain non-negative.

$$\widetilde{\sigma}_{e^b}^2 = \frac{SS_6}{N_b - \sum_{v=1}^{5} N_{\lambda_v}} \tag{5.22}$$

$$\widetilde{\sigma}_t^2 = \frac{\left(\frac{SS_5}{n(n-2)(n_o-1)} - \hat{\sigma}_{e^b}^2\right)}{n_o} \tag{5.23}$$

$$\widetilde{\sigma}_d^2 = \frac{\left(\frac{SS_4}{\binom{n}{2}-n} - 2n_o\hat{\sigma}_t^2 - \hat{\sigma}_{e^b}^2\right)}{n_o^2} \tag{5.24}$$

$$\widetilde{\sigma}_w^2 = \frac{\left(\frac{SS_3}{n(n_o-1)} - nn_o\hat{\sigma}_t^2 - \hat{\sigma}_{e^b}^2\right)}{n_o\,(n-1)} \tag{5.25}$$

$$\widetilde{\sigma}_b^2 = \frac{\left(\frac{SS_2}{n-1} - n_o^2\hat{\sigma}_d^2 - nn_o\hat{\sigma}_t^2 - (n-2)\,n_o\hat{\sigma}_w^2 - \hat{\sigma}_{e^b}^2\right)}{n_o^2\,(n-2)} \tag{5.26}$$

$$\widetilde{\sigma}_{e^w}^2 = \widetilde{\sigma}_{e^b}^2 \tag{5.27}$$

where $\sum_{v=1}^{5} N_{\lambda_v}$ is the sum of the number of projections not in the sixth set. While the last estimate for $\sigma_{e^w}^2$ could be obtained from the between source model, we can see from Table 5.9 that it is equal to $\sigma_{e^b}^2$.

[14] Some empirical testing showed this is not invertable for $n < 3$ and $n_0 < 2$, which are not situations we would expect this model to be used for.

5.6 Kernel Bayes factor

The REML equation given in 5.21 can be used to calculate the KBF shown in equation 5.31. Given the evidence sets $e_{u1},\ e_{u2},\ e_a$, the sets of scores in the kernel model are defined as follows:

For a given kernel κ, let

1. $\boldsymbol{s}_m = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\boldsymbol{x}_{ij} \in e_{u1},\, \boldsymbol{x}_{kl} \in e_{u2}$ (this is trace vs. trace)
2. $\boldsymbol{s}_n = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\boldsymbol{x}_{ij} \in \{e_{u1},\, e_{u2}\},\, \boldsymbol{x}_{kl} \in e_a$ (this is trace vs. controls)
3. $\boldsymbol{s}_c = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl} \in e_a$ (this is control vs. control)
4. $\boldsymbol{s} = \begin{pmatrix} \boldsymbol{s}_m \\ \boldsymbol{s}_n \\ \boldsymbol{s}_c \end{pmatrix}$

Under the common source inference, the models for the prosecution, $\mathcal{M}_p$, and defense, $\mathcal{M}_d$, for the vector of scores $\boldsymbol{s}$ are:

$$\boldsymbol{s}|\mathcal{M}_p \sim MVN\left(\begin{pmatrix} \boldsymbol{\theta}_w \\ \boldsymbol{\theta}_b \\ \boldsymbol{\theta}_{bw} \end{pmatrix},\ \begin{bmatrix} \boldsymbol{\Sigma}_w & \boldsymbol{\Sigma}_b & \boldsymbol{0} \\ \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_b \\ \boldsymbol{0} & \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bw} \end{bmatrix}\right) \tag{5.28}$$

$$\boldsymbol{s}|\mathcal{M}_d \sim MVN\left(\begin{pmatrix} \boldsymbol{\theta}_{bw} \\ \boldsymbol{\theta}_b \\ \boldsymbol{\theta}_{bw} \end{pmatrix},\ \begin{bmatrix} \boldsymbol{\Sigma}_{bw} & \boldsymbol{\Sigma}_b & \boldsymbol{0} \\ \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_b \\ \boldsymbol{0} & \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bw} \end{bmatrix}\right) \tag{5.29}$$

Model selection between $\mathcal{M}_p$ and $\mathcal{M}_d$ is completed using the kernel Bayes factor for the common source [49], calculated below for the parameters
$\psi = \left\{\theta_w,\, \theta_b,\, \sigma^2_b,\, \sigma^2_d,\, \sigma^2_t,\, \sigma^2_w,\, \sigma^2_{e^b},\, \sigma^2_{e^w}\right\}$:

$$
\begin{align}
KBF &= \frac{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_p\right) d\pi\left(\psi\right)}{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_d\right) d\pi\left(\psi\right)} \nonumber \\
&= \frac{\int f\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi,\, \mathcal{M}_p\right) f\left(\boldsymbol{s}_c|\psi\right) d\pi\left(\psi\right)}{\int f\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi,\, \mathcal{M}_d\right) f\left(\boldsymbol{s}_c|\psi\right) d\pi\left(\psi\right)} \nonumber \\
&= \frac{\int f_p\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right) d\pi\left(\psi|\boldsymbol{s}_c\right)}{\int f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right) d\pi\left(\psi|\boldsymbol{s}_c\right)} \tag{5.30} \\
&= \int \frac{f_p\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)}{m_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c\right)} \times \frac{f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)}{f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)} d\pi\left(\psi|\boldsymbol{s}_c\right) \nonumber \\
&= \int \frac{f_p\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)}{f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)} \times \frac{f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)}{m_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c\right)} d\pi\left(\psi|\boldsymbol{s}_c\right) \nonumber \\
&= \int \frac{f_p\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)}{f_d\left(\tbinom{\boldsymbol{s}_m}{\boldsymbol{s}_n}|\boldsymbol{s}_c,\, \psi\right)} d\pi\left(\psi|\boldsymbol{s}_c,\, \binom{\boldsymbol{s}_m}{\boldsymbol{s}_n},\, \mathcal{M}_d\right) \nonumber \\
&= \int \lambda\left(\binom{\boldsymbol{s}_m}{\boldsymbol{s}_n};\, \boldsymbol{s}_c,\, \psi\right) d\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right) \tag{5.31}
\end{align}
$$

We note that to carry out the calculations outlined from equation 5.21 to equation 5.27, the sample size n_o within-source needs to be equal for all sources. We also note that the Bayes factor in equation 5.31 requires to be estimated by Markov Chain Monte Carlo (MCMC) methods using a posterior sample from $\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right)$. Equation 5.31 is convenient as it marginalizes the likelihoods of $\boldsymbol{s}$ with respect to one distribution; therefore it can be evaluated using a single MCMC integration, which minimizes the MCMC error when compared to the two integrations required by 5.30. However, $\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right)$ is updated using samples from e_{u1}, e_{u2} and e_a and can only be assigned if the number of within-source samples is the same in e_{u1}, e_{u2}, and each source in e_a. This might prove to be unrealistic in practice where the numbers of samples in the trace evidence sets e_{u1} and e_{u2} are defined by what is gathered at the crime scene. Fortunately, it is realistic to consider that sources in e_a

can be studied through an equivalent number of samples. In cases where the number of samples for each source in e_a is balanced and the number of samples in e_{u1} and e_{u2} is not, we have to use equation 5.30 to assign the Bayes factor at the cost of a greater MCMC error (which can be reduced by increasing the number of posterior samples).

In general, to calculate the KBF 5.31, we need to sample from the posterior distribution $\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right)$ or from $\pi\left(\psi|\boldsymbol{s}_c\right)$. Their updated parameters are described below.

When using normal conjugate priors for the mean parameters, $N\left(\mu_w,\, \tau_w^2\right)$ and $N\left(\mu_b,\, \tau_b^2\right)$, we obtain the following results for the mean parameter posterior distributions:

$$\begin{aligned}
\pi\left(\theta_w|\boldsymbol{s}_w\right) &= N\left(\left(\frac{\frac{\mu_w}{\tau_w^2}+\frac{\sum \boldsymbol{s}_w}{\sigma^2_{\boldsymbol{s}_w}}}{\frac{1}{\tau_w^2}+\frac{n\binom{n_o}{2}}{\sigma^2_{\boldsymbol{s}_w}}}\right),\, \left(\frac{1}{\tau_w^2}+\frac{n\binom{n_o}{2}}{\sigma^2_{\boldsymbol{s}_w}}\right)^{-1}\right)\\
\pi\left(\theta_b|\boldsymbol{s}_b\right) &= N\left(\left(\frac{\frac{\mu_b}{\tau_b^2}+\frac{\sum \boldsymbol{s}_b}{\sigma^2_{\boldsymbol{s}_b}}}{\frac{1}{\tau_b^2}+\frac{\binom{nn_o}{2}-n\binom{n_o}{2}}{\sigma^2_{\boldsymbol{s}_b}}}\right),\, \left(\frac{1}{\tau_b^2}+\frac{\binom{nn_o}{2}-n\binom{n_o}{2}}{\sigma^2_{\boldsymbol{s}_b}}\right)^{-1}\right)
\end{aligned}$$

We can consider each mean sums of squares in the ANOVA Table 5.20 as the natural parameters $\eta_2, ..., \eta_6$ for the multivariate normal distribution of $\boldsymbol{s}$. These parameters are independent from one another and can be mapped to the parameters σ_b^2, σ_d^2, σ_t^2, σ_w^2, $\sigma_{e^b}^2$ as follows:

$$\begin{bmatrix} \sigma_b^2 \\ \sigma_w^2 \\ \sigma_d^2 \\ \sigma_t^2 \\ \sigma_{e^b}^2 \end{bmatrix} = \begin{bmatrix} (n-2)\,n_o^2 & n_o^2 & (n-2)\,n_o & nn_o & 1 \\ 0 & 0 & (n-1)\,n_o & nn_o & 1 \\ 0 & n_o^2 & 0 & n_o^2 & 1 \\ 0 & 0 & 0 & n_0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{bmatrix} \tag{5.32}$$

To draw samples from the posterior distributions of the natural parameters

$\eta_2, ..., \eta_6$, we construct the posterior distribution for a given η as:

$$\pi\left(\eta|\mathbf{v}_v^t \boldsymbol{s}_b\right) \propto f\left(\mathbf{v}_v^t \boldsymbol{s}_b|\eta\right) \pi\left(\eta\right)$$

We assumed that the likelihood for $\mathbf{v}_v^t \boldsymbol{s}_b$, $v \neq 1$, is normal with mean 0[15] and variance η. We choose the inverse-gamma as the conjugate prior $\pi\left(\eta|\alpha,\, \beta\right)$, with hyperparameters α and β. This results in an inverse-gamma posterior $\pi\left(\eta|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha,\, \beta\right)$ with parameters $\alpha + \frac{N}{2}$, $\beta + \frac{SS}{2}$ where N is the number of observations used to update $\pi\left(\eta|\alpha,\, \beta\right)$. N is given by the number of degrees of freedom in Table 5.20. SS is the sums of squares $\sum\left(\mathbf{v}_v^t \mathbf{s}_b\right)^2$. Using this, we can independently sample the natural parameters $\eta_2, ..., \eta_6$, from the following posterior distributions:

$$\begin{aligned}
\eta_2|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha_2,\, \beta_2 &\sim IG\left(\alpha_2 + \frac{N_2}{2},\, \beta_2 + \frac{SS_2}{2}\right)\\
\eta_3|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha_3,\, \beta_3 &\sim IG\left(\alpha_3 + \frac{N_3}{2},\, \beta_3 + \frac{SS_3}{2}\right)\\
\eta_4|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha_4,\, \beta_4 &\sim IG\left(\alpha_4 + \frac{N_4}{2},\, \beta_4 + \frac{SS_4}{2}\right)\\
\eta_5|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha_5,\, \beta_5 &\sim IG\left(\alpha_5 + \frac{N_5}{2},\, \beta_5 + \frac{SS_5}{2}\right)\\
\eta_6|\mathbf{v}_v^t \boldsymbol{s}_b,\, \alpha_6,\, \beta_6 &\sim IG\left(\alpha_6 + \frac{N_6}{2},\, \beta_6 + \frac{SS_6}{2}\right)
\end{aligned}$$

The samples for the natural parameters can be mapped to the parameters $\sigma_b^2,\, \sigma_d^2,\, \sigma_t^2,\, \sigma_w^2,\, \sigma_{e^b}^2$ using equation 5.32. Repeating this sampling will produce a sample from either of the posterior distributions$\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right)$ or $\pi\left(\psi|\boldsymbol{s}_c\right)$, which can then be used to perform MCMC integrations of the KBF in equation 5.30 or 5.31.

We note that if the design matrix, $\boldsymbol{Z}$, is not invertable, a more complicated Gibbs sampler[36] may be required to sample from the posteriors $\pi\left(\psi|\boldsymbol{s},\, \mathcal{M}_d\right)$ or

[15] Due to orthogonality to the mean.

$\pi(\psi|\boldsymbol{s}_c)$, using the estimates given in Equations 5.22-5.27.

5.7 Asymptotic assumption of normality

The assumption of normality of the scores for this model may be satisfied in one of two ways: (1) the choice of kernel or; (2) increasing the intrinsic dimensionality, p, of the original objects. It can formally be shown that as the dimension, p, of the object increases, the score vector converges to multivariate normal for a fixed sample size. Several simulation experiments were also performed to illustrate this effect by sampling functional objects with increasing number of basis functions before calculating their pairwise scores. To observe the normality convergence, three objects were compared at a time. We began by formally proving that as the dimension, p, of the object increases, the score vector converges to multivariate-normal for a fixed sample size.

Theorem 5.4. *Normality of kernel scores*

Let $\mathbf{x}_i \sim MVN\left(\mathbf{0},\, \frac{1}{2}\mathbf{I}_p\right)$ *for* p *dimensions,* $s_{ij} = \Phi^{-1}\left(F_p\left(\|\mathbf{x}_i - \mathbf{x}_j\|^2,\ p\right)\right)$, F_p *denotes the CDF of a* χ^2_p *distribution,* $\bar{y_{ij}} = \frac{\sum_{k=1}^{p}\left(x_{ik}-x_{jk}\right)^2}{p}$, *and* $i,\, j = 1,\, 2,\, 3$. *Then for* $\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \boldsymbol{x}_3$,

$$\mathbf{s} = \begin{pmatrix} \Phi^{-1}\left(F_p\left(\|\mathbf{x}_1 - \mathbf{x}_2\|^2,\ p\right)\right) \\ \Phi^{-1}\left(F_p\left(\|\mathbf{x}_1 - \mathbf{x}_3\|^2,\ p\right)\right) \\ \Phi^{-1}\left(F_p\left(\|\mathbf{x}_2 - \mathbf{x}_3\|^2,\ p\right)\right) \end{pmatrix} \rightsquigarrow MVN\left(\mathbf{0},\, \boldsymbol{\Sigma}\right)$$

where for $\gamma^2 < 1$,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \gamma & \gamma \\ \gamma & 1 & \gamma \\ \gamma & \gamma & 1 \end{pmatrix}$$

Proof. We first note that

$$\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 = \sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2 = p\bar{y}_{ij}$$

and that

$$\bar{y}_{ij} \sim \chi_1^2, \ E\left(\bar{y}_{ij}\right) = 1$$

Then note for $\boldsymbol{x}_i$, $\boldsymbol{x}_j$:

$$\begin{aligned}
&\Phi^{-1}\left(F_p\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2, \ p\right)\right) \\
&\qquad = \Phi^{-1}\left(Pr\left(\left\|\mathbf{x}_i^* - \mathbf{x}_j^*\right\|^2 \leq \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 \mid \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right)\right) \\
&\qquad = \Phi^{-1}\left(Pr\left(p\bar{y}_{ij}^* \leq p\bar{y}_{ij} | \bar{y}_{ij}\right)\right) \\
&\qquad = \Phi^{-1}\left(Pr\left(\frac{\sqrt{p}\left(\bar{y}_{ij}^* - 1\right)}{\sqrt{2}} \leq \frac{\sqrt{p}\left(\bar{y}_{ij} - 1\right)}{\sqrt{2}} | \bar{y}_{ij}\right)\right) \\
&\qquad = \Phi^{-1}\left(G_p\left(\frac{\sqrt{p}\left(\bar{y}_{ij} - 1\right)}{\sqrt{2}}\right)\right) && (5.33) \\
&\qquad = \Phi^{-1}\left(Z_p\right) && (5.34)
\end{aligned}$$

where $Z_p = G_p\left(\frac{\sqrt{p}(\bar{y}_{ij}-1)}{\sqrt{2}}\right)$. Note that $G_p\left(\frac{\sqrt{p}(\bar{y}_{ij}-1)}{\sqrt{2}}\right) \rightsquigarrow N\left(0, 1\right)$ as $p \to \infty$. These results are used in the remaining parts of the proof.

Let $\boldsymbol{a} \in \mathbb{R}^3$, we want to show that $\boldsymbol{a}^t\boldsymbol{s} \rightsquigarrow N\left(0, \sigma^2\right)$ as $p \to \infty$. We first define $\boldsymbol{s}'$:

$$\boldsymbol{s}' = \sqrt{\frac{p}{2}}\left(\begin{pmatrix} \bar{y}_{12} \\ \bar{y}_{13} \\ \bar{y}_{23} \end{pmatrix} - \mathbf{1}_3\right) \rightsquigarrow MVN\left(\mathbf{0}, \mathbf{\Sigma}\right)$$

and

$$h_p^{\boldsymbol{a}}(\boldsymbol{c}) = \boldsymbol{a}^t \begin{pmatrix} \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_1)) \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_2)) \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_3)) \end{pmatrix}.$$

For every $\boldsymbol{c} \to \boldsymbol{c}_0$, $h_p^{\boldsymbol{a}}(\boldsymbol{c}) \to h^{\boldsymbol{a}}(\boldsymbol{c}_0)$ where $h^{\boldsymbol{a}}(\boldsymbol{c}_0) = \boldsymbol{a}^t\boldsymbol{c}_0$. Consider:

$$\begin{aligned} &\left|h_p^{\boldsymbol{a}}(\boldsymbol{c}) - h^{\boldsymbol{a}}(\boldsymbol{c}_0)\right| \\ &\qquad = \left|h_p^{\boldsymbol{a}}(\boldsymbol{c}) - h^{\boldsymbol{a}}(\boldsymbol{c}) + h^{\boldsymbol{a}}(\boldsymbol{c}) - h^{\boldsymbol{a}}(\boldsymbol{c}_0)\right| \\ &\qquad \leq \left|h_p^{\boldsymbol{a}}(\boldsymbol{c}) - h^{\boldsymbol{a}}(\boldsymbol{c})\right| + \left|h^{\boldsymbol{a}}(\boldsymbol{c}) - h^{\boldsymbol{a}}(\boldsymbol{c}_0)\right| \\ &\qquad = \left|\boldsymbol{a}^t \begin{pmatrix} \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_1)) - \boldsymbol{c}_1 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_2)) - \boldsymbol{c}_2 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_3)) - \boldsymbol{c}_3 \end{pmatrix}\right| + \left|\boldsymbol{a}^t(\boldsymbol{c} - \boldsymbol{c}_0)\right| \end{aligned}$$

We know by the extended continuous mapping Theorem [72, pg. 67] that for $t \to t_0$, $\mathit{\Phi}^{-1}(G_p(\boldsymbol{t})) \to t_0$ as $p \to \infty$. This will give us that as $p \to \infty$,

$$\boldsymbol{a}^t \begin{pmatrix} \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_1)) - \boldsymbol{c}_1 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_2)) - \boldsymbol{c}_2 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_3)) - \boldsymbol{c}_3 \end{pmatrix} \to \boldsymbol{a}^t \begin{pmatrix} \boldsymbol{c}_{01} - \boldsymbol{c}_1 \\ \boldsymbol{c}_{02} - \boldsymbol{c}_2 \\ \boldsymbol{c}_{03} - \boldsymbol{c}_3 \end{pmatrix}$$

then

$$\left|\boldsymbol{a}^t \begin{pmatrix} \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_1)) - \boldsymbol{c}_1 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_2)) - \boldsymbol{c}_2 \\ \mathit{\Phi}^{-1}(G_p(\boldsymbol{c}_3)) - \boldsymbol{c}_3 \end{pmatrix}\right| + \left|\boldsymbol{a}^t(\boldsymbol{c} - \boldsymbol{c}_0)\right| \to 0 \,\text{as}\, p \to \infty.$$

which implies that

$$h_p^{\boldsymbol{a}}(Z_p) \to h^{\boldsymbol{a}}(Z) \sim N\left(0,\, \boldsymbol{a}^t\boldsymbol{\Sigma}\boldsymbol{a}\right)$$

So for $\boldsymbol{a}^t\boldsymbol{s} = h_p^{\boldsymbol{a}}(\boldsymbol{s}')$, we note by the extended continuous mapping theorem, we have that $\boldsymbol{a}^t\boldsymbol{s} = h_p^{\boldsymbol{a}}(\boldsymbol{s}') \rightsquigarrow N(0,\, \boldsymbol{a}^t\boldsymbol{\Sigma}\boldsymbol{a})$, which by Cramer-Wold implies

$$\boldsymbol{s} \sim MVN\left(\mathbf{0},\, \boldsymbol{\Sigma}\right)$$

□

We use the extended mapping theorem, given in [72, pg. 67] in the proof for Theorem 5.4. It is given below:

Theorem 5.5. *Extended continuous mapping. Let $\mathbb{D}_n \subset \mathbb{D}$ and $g_n : \mathbb{D}_n \mapsto \mathbb{E}$ satisfy the following statements: if $x_n \to x$ with $x_n \in \mathbb{D}_n$ for every n and $x \in \mathbb{D}_0$, then $g_n\left(x_n\right) \to g\left(x\right),$ where $\mathbb{D}_0 \subset \mathbb{D}$ and $g : \mathbb{D}_0 \mapsto \mathbb{E}$. Let X_n be maps with values in $\mathbb{D}_n$, let X be Borel measurable and separable, and take values in $\mathbb{D}_0$. Then*

$$(i) X_n \rightsquigarrow X \text{ implies that } g_n\left(X_n\right) \rightsquigarrow g\left(X\right);$$
$$(ii) X_n \xrightarrow{P} X \text{ implies that } g_n\left(X_n\right) \xrightarrow{P} g\left(X\right);$$
$$(iii) X_n \xrightarrow{as} X \text{ implies that } g_n\left(X_n\right) \xrightarrow{as} g\left(X\right).$$

Finally, we note that we only need to use three objects $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ to show the convergence of $\boldsymbol{s}$ to multivariate normality.

Corollary 5.6. *Triplicate scores*

Let $\mathbf{s}$ be a vector of pairwise scores between n objects $\boldsymbol{x}_i \sim MVN\left(\boldsymbol{\mu}_p,\, \boldsymbol{\Sigma}_{p,p}\right)$, $i = 1, 2, ..., n$ and dimension p. Show that for any $n > 1$, $\mathbf{s} \rightsquigarrow MVN\left(\mathbf{0},\, \boldsymbol{\Sigma}\right)$ as $p \to \infty$ where $\boldsymbol{\Sigma}$ is a covariance matrix with the structure:

$$Cov\left(s_{ij},\, s_{ij}\right) = 1$$
$$Cov\left(s_{ij},\, s_{ik}\right) = \gamma$$
$$Cov\left(s_{ij},\, s_{kl}\right) = 0$$

where $i,\ j,\ k,\ l$ are indices for objects taken from a single source and $\gamma^2 < 1$

Proof. This will be a proof by induction. We have shown that for $n = 3$ this holds. Assuming that $n = m$ holds, we show that $n = m + 1$ holds. Let $\mathbf{s}_{m+1}$ be the vector of pairwise scores from $m + 1$ objects

$$\begin{aligned}
\mathbf{s}_{m+1} &= \begin{pmatrix} s_{12} \\ s_{13} \\ \vdots \\ s_{1(m+1)} \\ s_{23} \\ \vdots \\ s_{m(m+1)} \end{pmatrix} \\
&= \sqrt{\frac{p}{2}} \left(\frac{1}{p} \sum_{k=1}^{p} \begin{pmatrix} \left(x_{1k} - x_{2k}\right)^2 \\ \left(x_{1k} - x_{3k}\right)^2 \\ \vdots \\ \left(x_{1k} - x_{(m+1)k}\right)^2 \\ \left(x_{2k} - x_{3k}\right)^2 \\ \vdots \\ \left(x_{mk} - x_{(m+1)k}\right)^2 \end{pmatrix} - \mathbf{1}_{\binom{m+1}{2}} \right) \\
&= \sqrt{\frac{p}{2}} \left(\sum_{k=1}^{p} \begin{pmatrix} \bar{Y}_{12} \\ \bar{Y}_{13} \\ \vdots \\ \bar{Y}_{1(m+1)} \\ \bar{Y}_{23} \\ \vdots \\ \bar{Y}_{m(m+1)} \end{pmatrix} - \mathbf{1}_{\binom{m+1}{2}} \right)
\end{aligned}$$

Let $M = \binom{m+1}{2}$,$\mathbf{a} \in \mathbb{R}^M$, then

$$
\begin{aligned}
&\sqrt{\frac{p}{2}}\begin{pmatrix} a_1 & a_2 & \cdots & a_M \end{pmatrix}\begin{pmatrix} \bar{Y}_{12} - 1 \\ \bar{Y}_{13} - 1 \\ \vdots \\ \bar{Y}_{m(m+1)} - 1 \end{pmatrix} \\
=&\sqrt{\frac{p}{2}}\left[a_1\left(\bar{Y}_{12} - 1\right) + a_2\left(\bar{Y}_{13} - 1\right) + \cdots + a_M\left(\bar{Y}_{m(m+1)} - 1\right)\right] \\
=&\sqrt{\frac{p}{2}}\left[a_1\bar{Y}_{12} - a_1 + a_2\bar{Y}_{13} - a_2 + \cdots + a_M\bar{Y}_{m(m+1)} - a_M\right] \\
=&\sqrt{\frac{p}{2}}\left[a_1\bar{Y}_{12} + a_2\bar{Y}_{13} + \cdots + a_M\bar{Y}_{m(m+1)} - (a_1 + a_2 + \cdots + a_M)\right] \\
=&\sqrt{\frac{p}{2}}\left[\frac{1}{p}\sum_{k=1}^{p} a_1\left(x_{1k} - x_{2k}\right)^2 + \cdots + \frac{1}{p}\sum_{k=1}^{p} a_M\left(x_{mk} - x_{(m+1)k}\right)^2 - \sum_{i=1}^{M} a_i\right] \\
=&\sqrt{\frac{p}{2}}\left[\frac{1}{p}\sum_{k=1}^{p}\left[a_1\left(x_{1k} - x_{2k}\right)^2 + a_2\left(x_{1k} - x_{3k}\right)^2 + \cdots + \sum_{k=1}^{p} a_M\left(x_{mk} - x_{(m+1)k}\right)^2\right] - \sum_{i=1}^{M} a_i\right].
\end{aligned}
\tag{5.35}
$$

Let $J_k = a_1\left(x_{1k} - x_{2k}\right)^2 + a_2\left(x_{1k} - x_{3k}\right)^2 + \cdots + a_M\left(x_{mk} - x_{(m+1)k}\right)^2$, we are interested in $E\left(J_k\right)$ with respect to the k^{th} dimension. We know that $\left(x_{ik} - x_{jk}\right)^2 \sim \chi^2_{df=1}$ and that $E\left(\left(x_{ik} - x_{jk}\right)^2\right) = 1$. This gives us $E\left(J_k\right) = a_1 + a_2 + \cdots + a_M$. Then, by the Central Limit Theorem as $p \to \infty$,

$$
\sqrt{\frac{p}{2}}\left[\frac{1}{p}\sum_{k=1}^{p} J_k - \sum_{i=1}^{M} a_i\right] \rightsquigarrow N\left(0,\, 1\right)
$$

This implies:

$$
\mathbf{s} \rightsquigarrow MVN\left(\mathbf{0},\, \mathbf{\Sigma}\right).
$$

□

The results of corollary 5.6 lets us prove and test convergence of $\boldsymbol{s}$ with triplicate scores instead of larger sets of scores since the convergence to MVN will occur no matter the size n. We can increase computational efficinecy by using $n = 3$.

6 Empirical simulations for convergence

Theorem 5.4 is useful when the scores are calculated with a Euclidean metric between objects known to have a multivariate normal distribution. However, this assumption can almost never be satisfied. We empirically tested the extension of Theorem 5.4 to other types of data and stationary kernels. For these simulations, we created high-dimensional objects using Fourier and B-spline basis transforms and used a squared-Euclidean and cross-correlation kernels to create the score vectors $\boldsymbol{s}$.

6.1 Normality Theorem 5.4 simulation

The first simulation checks the marginal and bivariate normality of the distribution for a score vector created by using 3 independent objects. If all objects come from the same source, we may write the score vector as:

$$\mathbf{s} = \begin{pmatrix} s_{12} \\ s_{13} \\ s_{23} \end{pmatrix}$$

and if they come from independent sources, then

$$\mathbf{s} = \begin{pmatrix} s_{1121} \\ s_{1131} \\ s_{2131} \end{pmatrix}$$

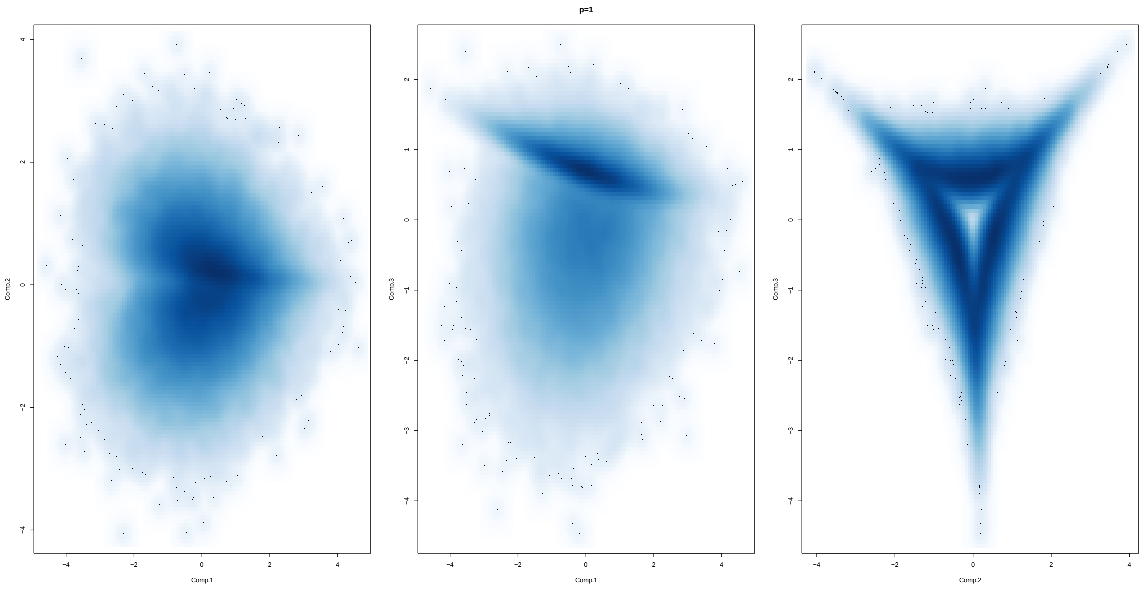
The two main situations we studied were: 1) 3 objects from a single source (given in Theorem 5.4); 2) 3 objects from 3 sources. The general algorithm for this simulation is given in Algorithm 2. We first sampled three objects from a p-MVN distribution and then computed the score vector $\boldsymbol{s}$ between these objects. After generating a large sample of score vectors, we computed the empirical covariance matrix $\hat{\mathbf{\Sigma}}$. We then used an eigen-decomposition of $\hat{\mathbf{\Sigma}}$ to get eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ which provided a rotation of the data cloud. This rotation was used to test for the normality of the marginal projections $\mathbf{v}^t \boldsymbol{s}$.

Algorithm 2 Kernel normality convergence simulation for within source

```
1. Sample $\mathbf{x}_i \sim MVN\left(\mathbf{0},\, \mathbf{I}_p\right)$ for $i = 1,\, 2,\, 3$
    (a) Compute $\mathbf{s} = \begin{pmatrix} s_{12} \\ s_{13} \\ s_{23} \end{pmatrix}$ for $s_{ij} = \Phi^{-1}\left(F_p\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2,\ p\right)\right)$
    (b) Store $\mathbf{s}$ from (a)
2. Repeat step (1) $n.sims = 50,000$ times and store in an $n.sim \times 3$ matrix $\mathbf{S}$
3. Calculate the covariance matrix $\hat{\mathbf{\Sigma}}$ of $\boldsymbol{s}$ from $\mathbf{S}$
4. Compute the eigenvectors $\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3$ of $\hat{\mathbf{\Sigma}}$ and calculate the principal scores
   $\mathbf{s}_1^* = \mathbf{v}_1\mathbf{S}^t,\ \mathbf{s}_2^* = \mathbf{v}_2\mathbf{S}^t,\ \mathbf{s}_3^* = \mathbf{v}_3\mathbf{S}^t$
5. Repeat steps 1-4 for increasing $p$.
```

The simulation was run for $p = 1,\ 2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ 192,\ 256$. The pairwise joint distributions for $\mathbf{s}_1^*$, $\mathbf{s}_2^*$, and $\mathbf{s}_3^*$ are plotted to visually check for normality, as seen in Figure 6.1 for $p = 1$. If normality for the score vector $\boldsymbol{s}$ held, we would expect these plots to be spherical or ellipsoidal with no outstanding structure in the point clouds. As can be seen in Figure 6.1, a clear departure from multivariate-normality can be seen for all pairwise joint distributions, especially that of $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$, which appears to be a hollow triangular distribution.

Figure 6.1: pairwise joint distributionsfor $\mathbf{s}_1^*$, $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$ with the "hollow-triangle" distribution between $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$ for $p = 1$. From left to right are $\mathbf{s}_1^*$ vs. $\mathbf{s}_2^*$, $\mathbf{s}_1^*$ vs. $\mathbf{s}_3^*$, $\mathbf{s}_2^*$ vs. $\mathbf{s}_3^*$.



We tested the normality of the marginal distributions of $\mathbf{s}_1^*$, $\mathbf{s}_2^*$, and $\mathbf{s}_3^*$using a Kolmogorov–Smirnov (K-S) test for normality[16]. The K-S test statistic is defined as:

$$D_n = \underset{x}{sup} \left| F_n(x) - F(x) \right|$$

which is the maximum distance between the empirical CDF $F_n(x)$ and a given CDF $F(x)$. For this, test $F(x)$ is the normal CDF for $\mathbf{s}_1^*$, $\mathbf{s}_2^*$, and $\mathbf{s}_3^*$. The hypotheses for this test are:

$$H_0 \text{ :The scores } \boldsymbol{s}^* \text{ are normally distributed}$$

$$H_1 \text{ :The scores } \boldsymbol{s}^* \text{ are not normally distributed}$$

We rejected H_0 when any of the p-values for the margins is less than the Bonferroni corrected significance level for multiple comparisons, $\alpha_{bonf} = 0.016 \approx \frac{0.05}{3}$.

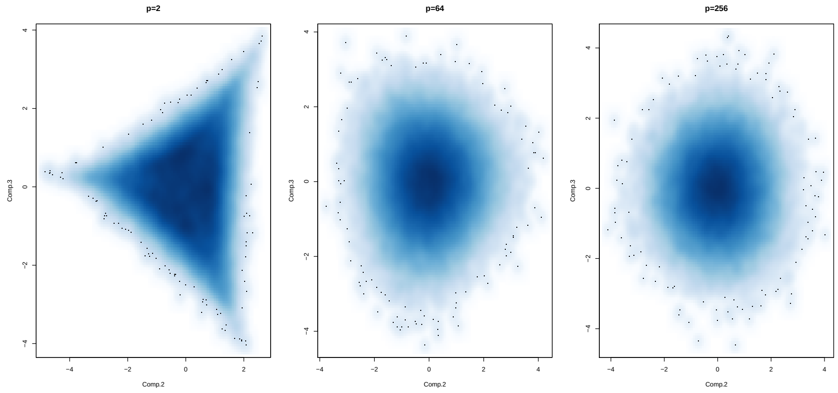
[16] This test was chosen due to its ability to efficiently handle the large number of principal scores we computed in the simulation.

6.1.1 Within-source simulation results

In the results, we focused on the marginal convergence of normality for $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$. It was found that the marginal distribution for $\mathbf{s}_1^*$ converged very quickly and we consistently failed to reject H_0 for $\mathbf{s}_1^*$. However, this certainly was not the case for the remaining margins.

As p increases, the hollow-triangle distribution fills in and converges to a spherical/ellipsoidal bivariate normal distribution. This convergence can be seen in Figure 6.2 for $p = 2,\ 64,\ 256$. For $p = 64$, the K-S test on the marginal distributions for $\mathbf{s}_1^*$ and $\mathbf{s}_3^*$ failed to reject the null hypothesis, however, the K-S test for normality of the margin for $\mathbf{s}_2^*$ rejected the null hypothesis. By the time $p = 256$, we had stopped noticing significant departures from multivariate normality.

Figure 6.2: Effect of increasing $p = 2,\ 64,\ 256$ on the structure of the bivariate distribution between $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$



The plots for the bivariate distributions of $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$ for

$p = 1,\ 2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ 192,\ 256$ are given in Appendix 10.4. The K-S statistics for $\mathbf{s}_2^*$ and $\mathbf{s}_3^*$ for $p = 64,\ 256$ are given in Table 6.1 below. We see that for this example, at $p = 256$, $\mathbf{s} \sim MVN(\boldsymbol{\theta},\, \boldsymbol{\Sigma})$.

Table 6.1: K-S statistics results for marginals

	p	D_{KS}	p-value	MVN decision
$\mathbf{s}_2^*$	64	0.007514	0.007058	Reject
$\mathbf{s}_3^*$	64	0.005663	0.08091	
$\mathbf{s}_2^*$	256	0.005183	0.1364	Fail to reject
$\mathbf{s}_3^*$	256	0.003102	0.722	

6.1.2 Between-source simulation results

Using the same simulation approach used for the within-source samples, we tested the convergence to normality of the vector of pairwise comparisons between three objects sampled from three sources. The score vector considered is:

$$\mathbf{s} = \begin{pmatrix} s_{1121} \\ s_{1131} \\ s_{2131} \end{pmatrix}$$

Convergence to normality for between-source scores occured just as it did for within-source. This can be attributed to the sampling of objects from independent sources instead of independent objects sampled from a single source.

6.2 Extension of simulations to non-normal data

We studied Theorem 5.4 by the use of a simulation similar to the one used for Theorem 5.4. However, unlike in Algorithm 2, we transformed data into a complex high-dimensional space by using different basis transformations. The specific basis transforms we used were the Fourier and B-spline. These were selected to cover two main classes of basis expansion that describe signals: 1) continuous and periodic and 2) continuous and aperiodic [54].

Let $\{\beta_k(t)\}$ be a set of orthonormal functions for $k = 1, 2, ..., p$ basis elements over the domain $t \in [-3, 3]$. Let $\mathbf{\Sigma}_{p,p}^{AR}$ be a banded covariance matrix and $\mathbf{\Gamma}_{p,p}^{AR}$ be banded covariance matrix so that $\mathbf{\Gamma}_{p,p}^{AR} = c\mathbf{\Sigma}_{p,p}^{AR}$ for some scalar $0 < c < 1$.

Algorithm 3 Non-normal score simulation

```
Choose a kernel$\kappa$ and a set of basis functions $\{\beta_k(t)\}$,
  1. Sample the $i^{th}$ source mean $\mathbf{a}_i \sim MVN\left(\boldsymbol{\mu},\ \mathbf{\Sigma}_{p,p}^{AR}\right)$for $i = 1, ..., n$
  2. Sample the coefficients $\mathbf{b}_{ij} \sim MVN\left(\mathbf{a}_i,\ \mathbf{\Gamma}_{p,p}^{AR}\right)$ for $j = 1, ..., n_o$
  3. Calculate the function values $x_{ij}(t) = \sum_{k=1}^{p} b_{ij}\beta_k(t)$ for all $t \in [-3,\ 3]$
  4. Calculate and store $\mathbf{s} = \begin{pmatrix} s_{1112} \\ \vdots \\ s_{n,n_{o-1},\ n,n_o} \end{pmatrix}$
  5. Repeat steps 1-4 $n.sims$ times
```

We constructed $\mathbf{\Sigma}_{p,p}^{AR}$ with a lag[17] equal to two, main diagonal values equal to 10, and super/sub diagonal elements equal to 5. An example for $p = 5$ is given below:

$$\mathbf{\Sigma}_{p,p}^{AR} = \begin{bmatrix} 10 & 5 & 5 & 0 & 0 \\ 5 & 10 & 5 & 5 & 0 \\ 5 & 5 & 10 & 5 & 5 \\ 0 & 5 & 5 & 10 & 5 \\ 0 & 0 & 5 & 5 & 10 \end{bmatrix}$$

We let $c = \frac{1}{10}$ for the construction of $\mathbf{\Gamma}_{p,p}^{AR}$.

The basis functions are used to represent a function f by use of a linear expansion, and generally is defined as

$$f(t) = \sum_{k=1}^{p} b_k\beta_k(t)$$

where $b_k \in \mathbb{R}$ and β_k an element of the functional vector $\boldsymbol{\beta}$ defined by the basis set being used.

[17] The number of super/sub diagonals in the matrix.

6.2.1 Fourier basis simulation

For the Fourier basis simulation, the series of elements defining $\boldsymbol{\beta}$ are

$$sin\left(t\right),\ cos\left(t\right),\ sin\left(2t\right),\ cos\left(2t\right),\ sin\left(3t\right),\ cos\left(3t\right),...,sin\left(pt\right),\ cos\left(pt\right)$$

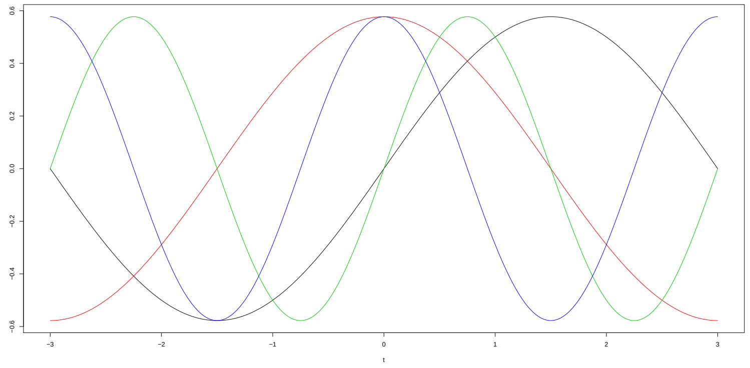
and are used in the linear expansion for f_{ij}

$$f_{ij}\left(t\right) \quad = \quad \sum_{k=1}^{p} b_{ijk}sin\left(kt\right)I\left(k\ odd\right)+b_{ijk}cos\left(kt\right)I\left(k\ even\right)$$

where $p \in \mathbb{N}$ and the effective dimension of each object is p. The coefficients, denoted by b_{ijk}, are randomly distributed as defined in Step 2 of Algorithm 3. Because the Fourier basis requires alternating between $sin\left(pt\right)$ and $cos\left(pt\right)$, the indicator functions $I\left(k\ odd\right)$ and $I\left(k\ even\right)$ control the basis element being used.

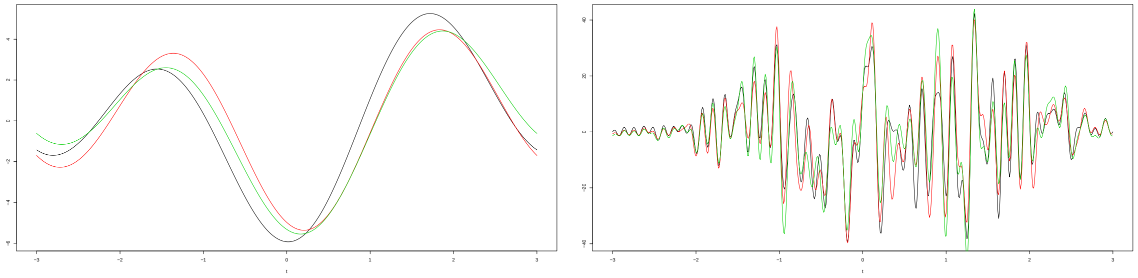
The Fourier basis functions for $p = 2$, before being multiplied by the random coefficients, is given in Figure 6.3

Figure 6.3: The Fourier basis set for 4 basis elements



Using this set of basis values in Step 3 of Algorithm 3, with three samples of coefficients from the same source, resulted in the signals found in Figure 6.4.

Figure 6.4: Fourier-based random function for $p = 4$ (left) and $p = 100$ (right)



Within-source Fourier simulation The simulations showed that convergence to normality occured as the dimension p increased. We did not observe the "hollow triangle" as we did with earliler simulations, but rather the cloud of PCA scores took a 3D cone shape. This is seen in Figure 6.5 in the leftmost pair of plots for $p = 4$. By $p = 512$ $\boldsymbol{s}$, we cannot comment on any departure from multivariate normality.

Figure 6.5: Fourier basis and with Euclidean kernel convergence simulation for within-source scores. In pairs from left to right, $p = 4,\ 64,\ 512$

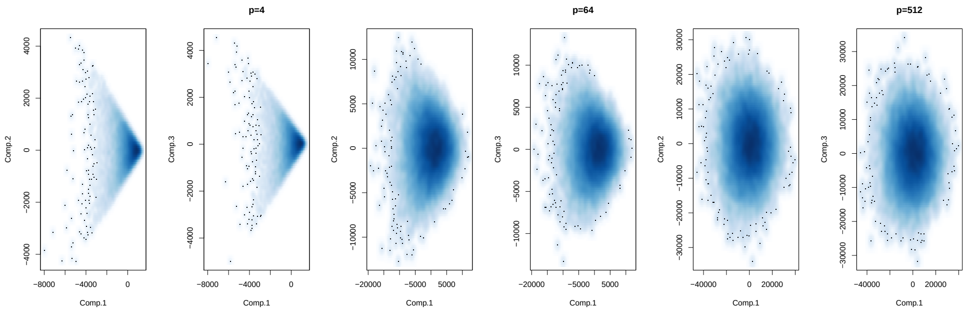
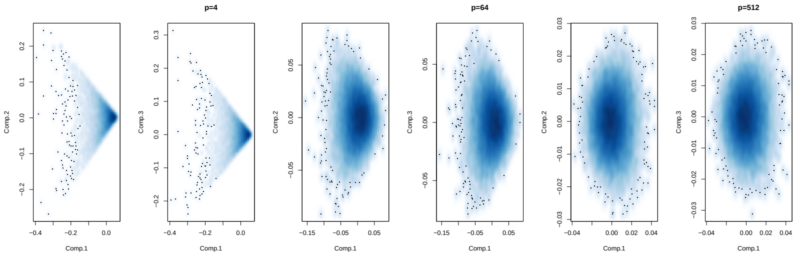


Figure 6.6: Fourier basis and with cross correlation kernel convergence simulation for within-source scores. In pairs from left to right, $p = 4,\ 64,\ 512$



The full set of convergence plots for the Fourier basis with a cross correlation score are given in Appendix 10.4.2 and the full set of convergence plots for the Fourier basis with a Euclidean score are given in Appendix 10.4.3

Between-source Fourier simulation As before, the simulations showed that convergence to normality occured as the dimension p increased between objects from different sources. However, the convergence occured by $p = 128$, much earlier than it did within-source. This convergence can be seen in Figure 6.7 from the leftmost pair of plots for $p = 4$ to the rightmost pair of plots for $p = 128$.

Figure 6.7: Fourier basis and with Euclidean kernel convergence simulation for between-source scores. In pairs from left to right, $p = 4,\ 64,\ 128$

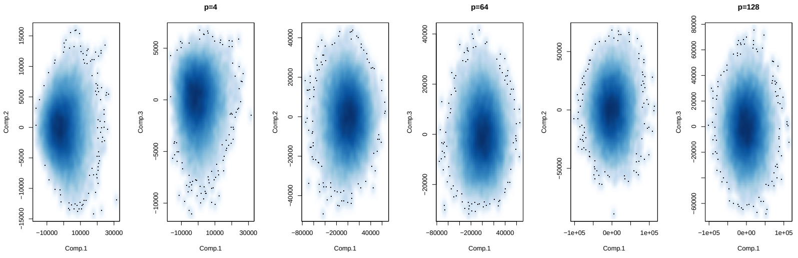
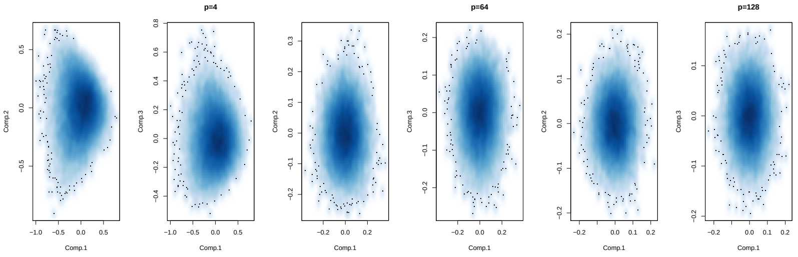


Figure 6.8: Fourier basis and with cross correlation kernel convergence simulation for between-source scores. In pairs from left to right, $p = 4,\ 64,\ 128$



There was no noticeable difference in the convergence of the distribution of scores for both kernels considered.

The full set of convergence plots for the Fourier basis with a cross correlation score are given in Appendix 10.4.2 and the full set of convergence plots for the Fourier basis with a Euclidean score are given in Appendix 10.4.3

6.2.2 B-spline basis simulation

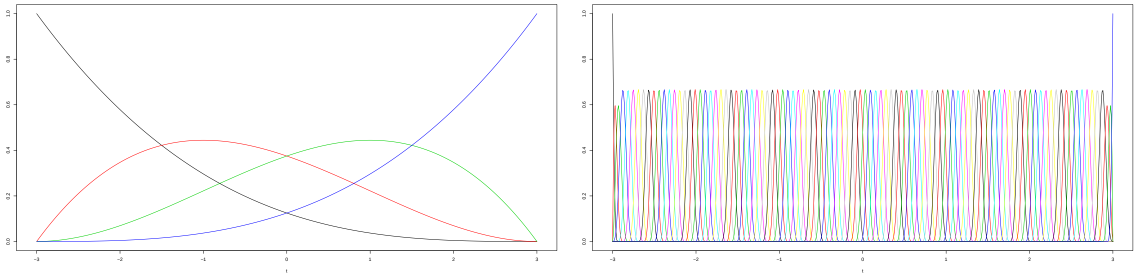
We use a B-spline of order n defined over a domain (in this case $t \in [-3, 3]$). Specifically, we use the B-spline basis used in the fda library [55] for R, described in [54]. The B-spline is defined as

$$f_{ij,n}(t) = \sum_{k=1}^{p} b_{ijk} \beta_{k,n}(t)$$

where the coefficients b_{ijk} are randomly selected as defined in Step 2 of Algorithm 3 and $\beta_{k,n}$ is the k^{th} polynomial of order n. For the example given below, cubic polynomials are used.

The B-spline basis functions for $p = 4$ before being multiplied by the random coefficients are given in Figure 6.9

Figure 6.9: The B-spline basis set for $p = 4$ and $p = 100$ basis elements



Using this set of basis values in Step 3 of Algorithm 3 with three samples of coefficients from the same source will result in the signals found in Figure 6.10. As the dimension p increases, the complexity of the signal increases as seen by the number of spikes or valleys present in the right plot compared to the left.

Figure 6.10: B-spline based random function for $p = 4$ (left) and $p = 100$ (right)

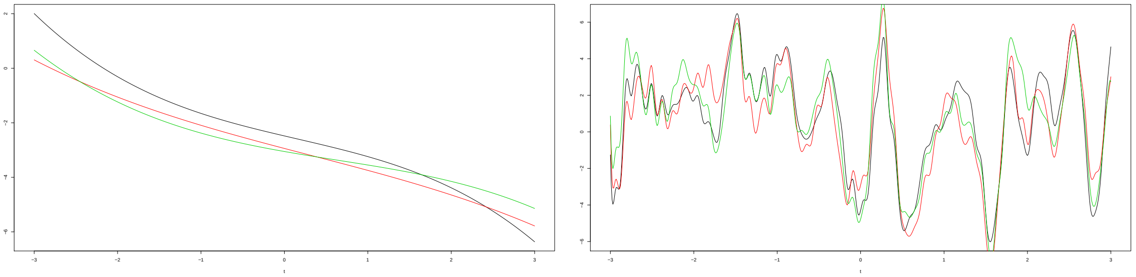


Figure 6.11: B-spline basis and with Euclidean kernel convergence simulation for within-source scores. In pairs from left to right, $p = 4,\ 64,\ 512$

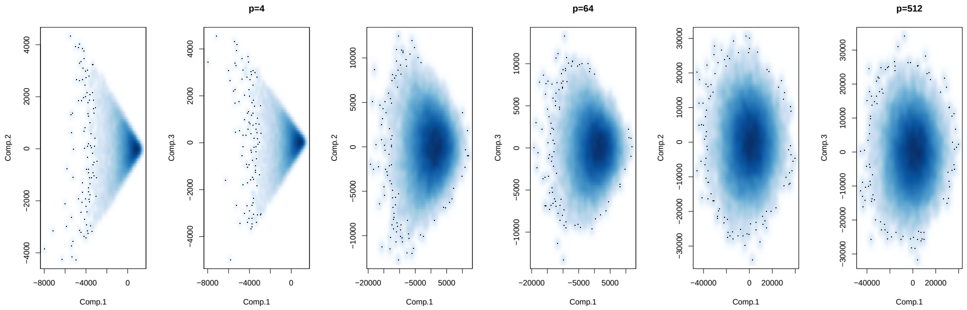
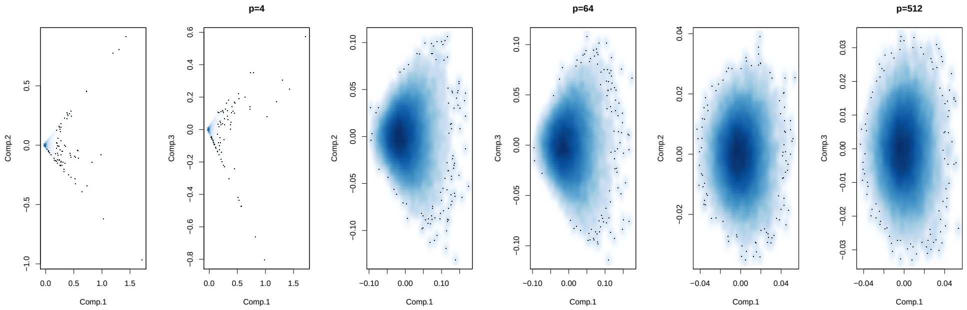


Figure 6.12: B-spline basis and with cross correlation kernel convergence simulation for within-source scores. In pairs from left to right, $p = 4,\ 64,\ 512$



Within-source B-spline simulation The convergence plots for the B-spline basis with a cross correlation score are given in Appendix 10.4.4. The convergence plots for the B-Spline basis with a squared-Euclidean score are given in Appendix 10.4.5. It is noteable that all convergence plots look similar to each other and have similar convergence results.

Between-source B-spline simulation As before, the simulations showed that convergence to normality occured as the dimension p increased between objects from different sources. However, the convergence occured by $p = 128$, much earlier than it did within-source. This convergence can be seen in Figure 6.13 from the leftmost pair of plots for $p = 4$ to the rightmost pair of plots for $p = 128$.

Figure 6.13: B-spline basis and with Euclidean kernel convergence simulation for between-source scores. In pairs from left to right, $p = 4,\ 64,\ 128$

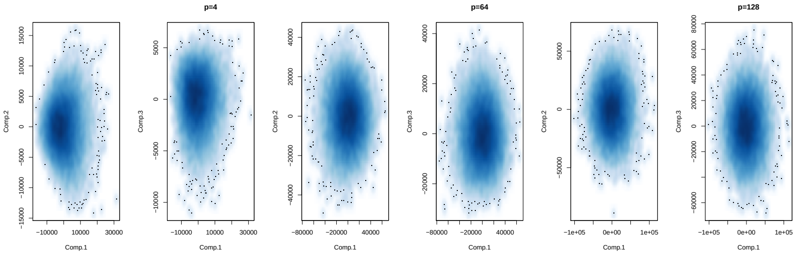
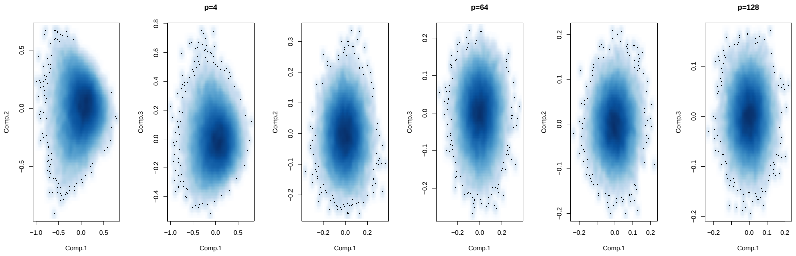


Figure 6.14: B-spline basis and with cross correlation kernel convergence simulation for between-source scores. In pairs from left to right, $p = 4,\ 64,\ 128$



There was no noticeable difference in the convergence of the distribution of scores for both kernels considered.

The full set of convergence plots for the Fourier basis with a cross correlation score are given in Appendix 10.4.2 and the full set of convergence plots for the Fourier basis with a Euclidean score are given in Appendix 10.4.3

7 Application of kernel method to a forensic problem

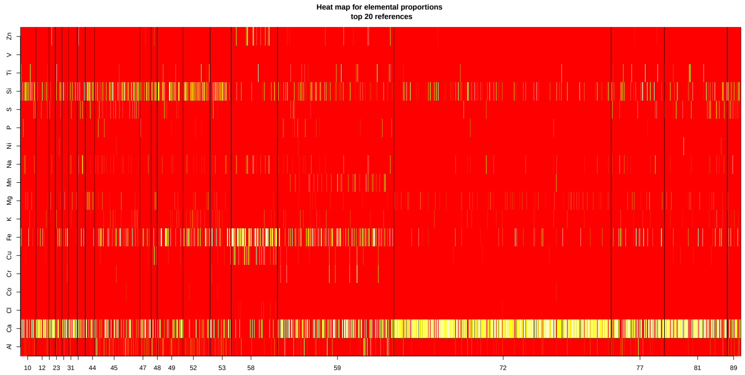
7.1 Very small particle project

The single source Gantz/Saunders model was applied to a dataset collected by Stoney et. al [68] concerning very small particles (VSPs) on carpet fibers. This application was a proof-of-concept for the Gantz/Saunders model as it had not been applied to real-world data. This dataset was selected because of its similarities to handwriting and firearms, and for the difficulty of the problem. The data is compositional, unbalanced, and complex in nature, making it an ideal candidate for kernel based methods. The application of this method was published in Chemometrics and Intelligent Laboratory Systems [4].

This application proved that the kernel-based method was able to infer the source of VSPs quite successfully. On the training dataset we acheived 100% correct classification and on the test dataset we acheived 55% correct classification. This is better than chance alone (5%) and is a slight improvement over Stoney et. al [68], who achieved a 50% correct classification rate with this dataset. What was especially encouraging was that this model required only 3 parameters to be estimated from the scores.

However, VSPs must be handled a bit differently than other types of trace evidence (fingerprints, toolmarks, spectra, etc.) for statistical analysis, at least for kernel methods. The sources of VSPs had widely varying sample sizes, anywhere from 100 to 18,170 particles. This unbalanced and rather complex sampling is seen in Figure 7.1, where black vertical lines separate the samples for sources. Lighter colors indicate a higher proportion of that element being present in a particular particle. In this form, we would be unable to perform a kernel-based approach without the use of a supercomputer. In the training dataset, we had 60,277 VSPs which would result in 1,816,628,226 pairwise comparisons, and the covariance matrix describing the dependencies betewen all these could be as large as 3.3 exabytes in storage. It is currently unrealistic to process this amount of data, and additional data processing is required.

Figure 7.1: Heat map for top 20 sources, selected by most numerous samples of VSPs. Vertical black lines separate VSPs by source (x-axis). Lighter colors represent greater proportions of the corresponding elements (y-axis) for each VSP.



The kernel defined for this project was built after initial work showed that well known classes of kernels, as listed in the previous section, did not meet normality assumptions of the model and did not correctly classify more than 50% of the sources in the validation dataset. Our main challenge was that two VSP sets do not necessarily contain an equivalent number of particles and as mentioned before, we cannot currently compare every pair of particle. Therefore, the first step in defining a kernel to measure the simlarity of unbalanced sets of VSPS required us to map the sets of VSPs into a common space.

Let $\mathbf{X}_i^l$ and $\mathbf{X}_j^l$ be two $p_i \times 18$ and $p_j \times 18$ matrices representing two sets of VSPs from reference source l in $1, ..., 20$, with p_i and p_j particles respectively, and 18 chemical elemental relative concentrations. Let $\mathbf{x}_{ir}^l$ be the set of measurements for the r^{th} ($r = 1$, $2, ...,$ 18) chemical element across all particles on the i^{th} ($i = 1, ...,$ n_l) VSP set of source l. For a pair of VSP sets, our kernel compares the empirical cummulative distribution functions of the relative concentration of each element taken in turn and then aggregates the elementwise scores to obtain s_{ij} for that pair of VSP sets.

We defined κ as the function $\kappa\left(\mathbf{F}_{\mathbf{Y}_i^l}, \mathbf{F}_{\mathbf{Y}_j^l}\right)$ where $\mathbf{F}_{\mathbf{Y}_i^l}, \mathbf{F}_{\mathbf{Y}_j^l}$ are the sets of empirical cummulative distribution functions (ECDF) based off of $\mathbf{Y}_i^l$ and $\mathbf{Y}_j^l$, where

$$\mathbf{Y}_i^l = \mathbf{A}\mathbf{X}_i^l.$$

$\mathbf{A}$ is a matrix of bases that can be suitably chosen to weight the original data or reduce the dimension of the number of measurements taken on each particle. The outside operator of $\kappa\left(\mathbf{F}_{\mathbf{Y}_i^l}, \mathbf{F}_{\mathbf{Y}_j^l}\right)$ is:

$$\kappa\left(\mathbf{F}_{\mathbf{Y}_i^l}, \mathbf{F}_{\mathbf{Y}_j^l}\right) = \sum_{r=1}^{18} log\left(\int\left(F_{\mathbf{y}_{ir}^l}(t) - F_{\mathbf{y}_{jr}^l}(t)\right)^2 dt\right)$$

where $\mathbf{y}_{ir}^l$ is the r^{th} ($r = 1, 2, ..., 18$) column of $\mathbf{Y}_i^l$ corresponding to the transformed measurements of the r^{th} chemical element in matrix $\mathbf{X}_i^l$.

As mentioned at the beginning of section 5.7, the normality assumption for scores may be satisfied in the design of the kernel. For this project, the dimension was too low to assure asymptotic convergence to normality. However, the kernel we designed did satisfy this assumption, especially once the logarithm was used.

Through the use of mapping functions and our kernel method, we were able to simplify what seemed to be a very large and complex problem, into a simpler and easier-to-handle problem. The method was computationally efficient and it corrected issues we had with sample sizes and compositional data.

7.2 Microspectrophotometry of blue cotton fibers

The new model proposed in this dissertation was tested on microspectrophotometry (MSP) data collected by Dr. Patrick Buzzini of Sam

Houston State University. The objectives of this analysis were to quantify and classify blue cotton fibers based off of spectra using the KBF method. We are collaborating with Dr. Buzzini to publish this work.

When evaluating fibers recovered in connection with a crime, one of the first steps taken is to compare the color of the recovered fibers with reference fibers from a known source. Color is one of the characteristics that varies most between different textile fibers and has the potential to provide highly probative value to the weight of evidence. Colors in textile fibers come from a mixture of dyes used by a manufacturer and MSP provides an objective description of this physical characteristic.

Currently, the evaluation and comparison of MSP spectra is done by trained forensic technicians. As in any human-based activity, the conclusion reached by forensic technicians when considering MSP spectra is susceptible to vary uncontrollably within an individual and between different individuals (see [44] for an example related to fingerprint comparison.) The criteria that examiners use in the evaluation of MSP spectra are: general spectra shape; locations of maxima, minima, and points of flexion; intersection of curves at multiple locations (used for exclusion). To form their conclusions, examiners may observe multiple spectra from multiple fibers from the recovered and reference fibers. This visual comparison process makes the unrealistic assumption that examiners are capable to subjectively quantify sources of variability from within and between sources through a mere visual examination of the spectra.

The use of the KBF method allows us to simultaneously compare multiple sets of fibers from multiple source and objectively quantify the within and between-source variations observed between the spectra. We can design a kernel to duplicate the criteria used by examiners, reduce the time needed to evaluate spectra, and increase objectivity of the results.

7.2.1 MSP data

Microspectrophotometry is used to measure the intensity of reflected, transmitted, or absorbed light at different wavelengths. MSP typically measures intensity across the visible light spectrum (350nm-700nm) and sometimes will include measurements in the near-infared and ultraviolet. For this project, $n_o = 3$ fibers were sampled from $n = 20$ blue cotton t-shirts for a total of $nn_o = 60$ fibers. The individual fibers were remeasured three times each, and their results averaged together[18]. We display the raw, unscaled spectra in Figure 7.2. Several of the spectra in this dataset are distinctly different than others and should be easy to separate from the other sources. In contrast, there are several spectra that are very similar to each other[19] and may be assigned to a wrong source.

Before application of the model, rescaling was completed so that we could use a kernel that measured the relative difference in intensity between the spectra at selected wavelength. Without rescaling, this may result in an erroneous analysis due to possible difference in dye concentration between multiple fibers of a given source. For rescaling, we rescaled each spectra respective of their intensity at wavelength 700nm, which had the least variance observed. The resulting rescaled data can be seen in Figure 7.3.

[18] This averaging is necessary as our model does not support a third level in the hierarchy.

[19] This is due to the sampling method.

Figure 7.2: Unscaled blue cotton MSP spectra. Like colors represent samples from same sources.

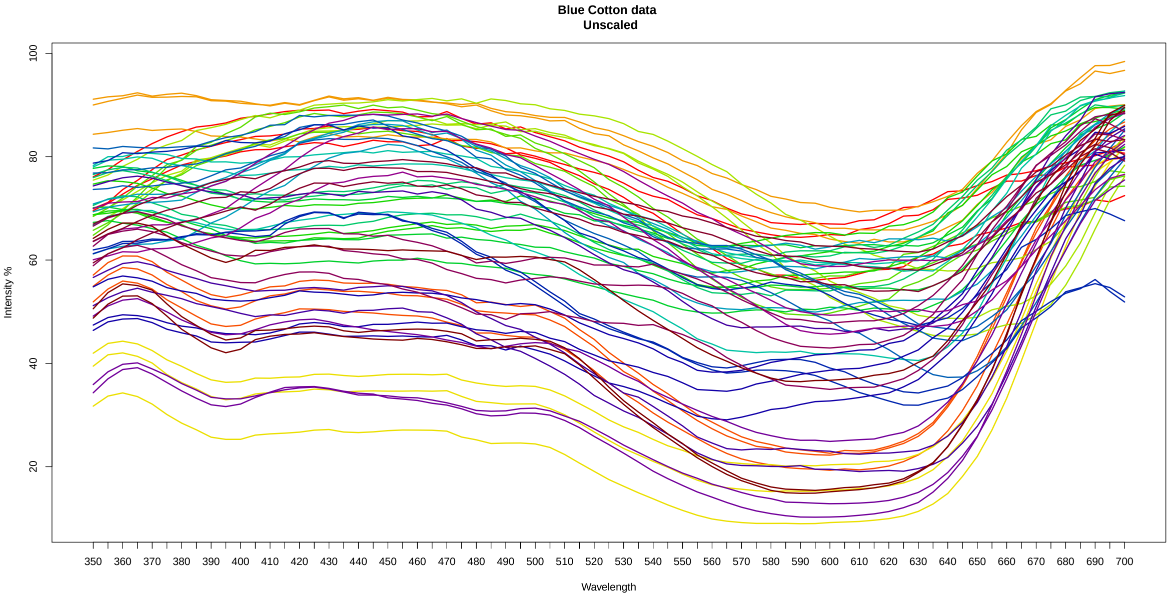
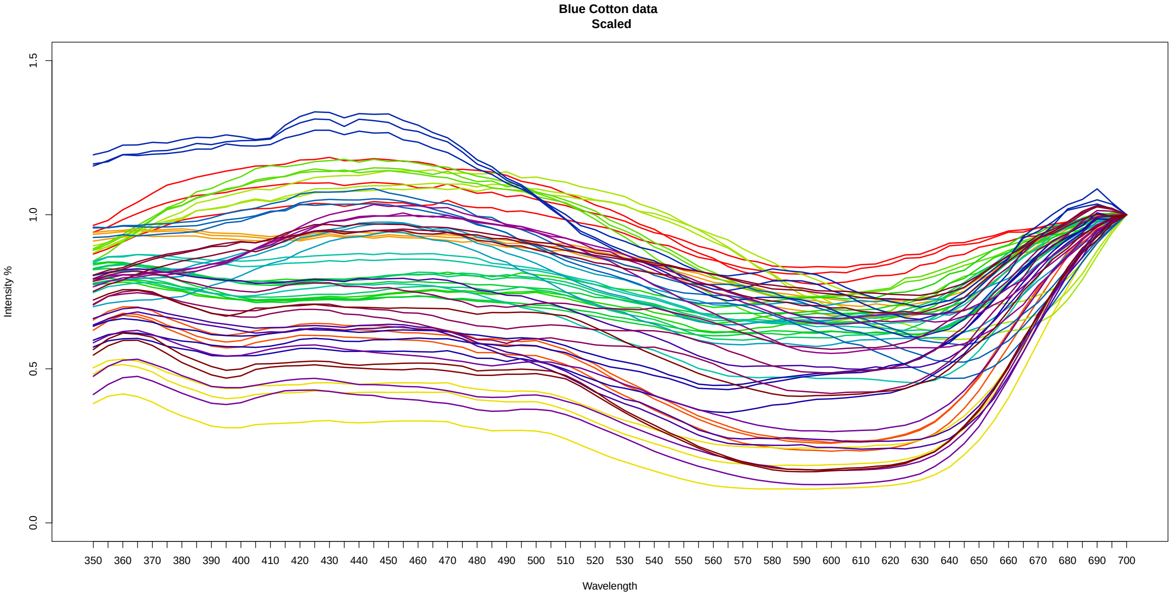


Figure 7.3: Rescaled blue cotton MSP spectras. Like colors represent samples from same sources. Note because of rescaling all spectra equal 1 at 700nm.



7.2.2 Methodology

We use the KBF method as described earlier for quantification purposes with the addition of using the method for objective source classification. We implemented a simple kernel to capture two pieces of information in the MSP data: 1) the shape of the spectra measurements and 2) the closeness of the spectra based on intensity. The similarity in shape of two spectra $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{kl}$ is measured using the absolute value of the cross correlation $\rho_{ijkl} = |\rho(\boldsymbol{x}_{ij}, \boldsymbol{x}_{kl})| \in [0, 1]$. The absolute value is used to ensure the kernel will satisfy definition 3.1. To capture the closeness of spectra, a squared Euclidean distance $d_{ijkl} = \|\boldsymbol{x}_{ij} - \boldsymbol{x}_{kl}\|^2$ is used. These are combined into the kernel via the closure properties:

$$\kappa(\boldsymbol{x}_{ij}, \boldsymbol{x}_{kl}) = \frac{1}{\rho_{ijkl}} d_{ijkl}$$

where the inverse of ρ_{ijkl} weights the distance. The kernel will return small values when spectra are close in intensity and have similar shapes. Moderate scores may be returned if spectra have simlar shapes but express different intensity levels or if spectra are similar in intensity but have distinct shapes. Large scores will be returned for spectra that have different shapes and are well separated by intensity. Theoretically, it is possible to have an undefined kernel if $\rho_{ijkl} = 0$; in practice it can be corrected by adding a small amount of noise to measurements equal to 0.

We have already discussed the KBF methodology in the previous chapters, therefore the quantification of the weight of fiber evidence is trivial. In this section, we are concerned with the task of classification. Note that our model does not enable us to classify a new sample into a specific class as in traditional Bayes classifiers. The model described in the previous chapters only allows us to determine if two sets of objects originate from the same, unkown source. For a series of sources, our model allows us to assign the posterior probability that two sets of objects are from the same source without having to study all possible sources as in a traditional Bayes classifier. Our model is limited in the sense that it will not enable us to assign a new set of objects to a specific source, however its power lives in its ability to consider an open set of sources. In practice, if we have a finite series of sources and we know the origin of one of the two sets of objects (say e_{u1}), we will assign the sets of objects of unknown origin (say e_{u2}) to the source with highest posterior probability $P\left(M_p|\boldsymbol{s}\right)$. We express $P\left(M_p|\boldsymbol{s}\right)$ as:

$$P\left(M_p|\boldsymbol{s}\right) = \frac{P\left(\boldsymbol{s}|M_p\right)P\left(M_p\right)}{P\left(\boldsymbol{s}|M_p\right)P\left(M_p\right) + P\left(\boldsymbol{s}|M_d\right)P\left(M_d\right)}$$

For ease of notation, let $A = P\left(\boldsymbol{s}|M_p\right)P\left(M_p\right)$ and $B = P\left(\boldsymbol{s}|M_d\right)P\left(M_d\right)$:

$$\begin{aligned} P\left(M_p|\boldsymbol{s}\right) =& \frac{A}{A+B} \\ =& \frac{1}{1+\frac{B}{A}} \end{aligned}$$

Our decision rule will assign a source of e_{u2} based off the maximum posterior probability $P\left(M_p|\boldsymbol{s}\right)$ calculated for different e_{u1}. We are interested in:

$$\begin{aligned}
\underset{e_{u1}}{argmax}\left(P\left(M_p|\boldsymbol{s}\right)\right) =& \underset{e_{u1}}{argmax}\left(\frac{1}{1+\frac{B}{A}}\right)\\
=& ln\left(\underset{e_{u1}}{argmax}\left(\frac{1}{1+\frac{B}{A}}\right)\right)\\
=& \underset{e_{u1}}{argmax}\left(ln\left(\frac{1}{1+\frac{B}{A}}\right)\right)\\
=& \underset{e_{u1}}{argmax}\left(ln\left(1\right)-ln\left(1+\frac{B}{A}\right)\right)\\
=& \underset{e_{u1}}{argmax}\left(-ln\left(1+\frac{B}{A}\right)\right)\\
=& \underset{e_{u1}}{argmin}\left(\frac{B}{A}\right)\\
=& \underset{e_{u1}}{argmax}\left(\frac{A}{B}\right)\\
=& \underset{e_{u1}}{argmax}\left(ln\left(\frac{A}{B}\right)\right)\\
=& \underset{e_{u1}}{argmax}\left(ln\left(A\right)-ln\left(B\right)\right)
\end{aligned}$$

Substituting A and B, and dropping the argmax for ease of notation:

$$\begin{aligned}
& ln\left(P\left(\boldsymbol{s}|M_p\right)P\left(M_p\right)\right)-ln\left(P\left(\boldsymbol{s}|M_d\right)P\left(M_d\right)\right)\\
=& ln\left(P\left(\boldsymbol{s}|M_p\right)\right)-ln\left(P\left(\boldsymbol{s}|M_d\right)\right)+ln\left(\frac{P\left(M_p\right)}{P\left(M_d\right)}\right)\\
=& ln\left(P\left(\boldsymbol{s}_m\boldsymbol{s}_n\boldsymbol{s}_cM_p\right)\right)-ln\left(P\left(\boldsymbol{s}_m\boldsymbol{s}_n\boldsymbol{s}_c|M_d\right)\right)+C\\
=& ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n\boldsymbol{s}_c,M_p\right)P\left(\boldsymbol{s}_n\boldsymbol{s}_c|M_p\right)\right)-ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n\boldsymbol{s}_c,M_d\right)P\left(\boldsymbol{s}_n\boldsymbol{s}_c|M_d\right)\right)+C\\
=& ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n,M_p\right)P\left(\boldsymbol{s}_n\boldsymbol{s}_c|M_p\right)\right)-ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n,M_d\right)P\left(\boldsymbol{s}_n\boldsymbol{s}_c|M_d\right)\right)+C \qquad (7.1)\\
=& ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n,M_p\right)\right)-ln\left(P\left(\boldsymbol{s}_m|\boldsymbol{s}_n,M_d\right)\right)+C \qquad (7.2)
\end{aligned}$$

The result in equation 7.1 is a result of the conditional independence between $\boldsymbol{s}_m$ and $\boldsymbol{s}_c$ given our knowledge of $\boldsymbol{s}_n$ by construction of the covariance matrix in 5.28

and 5.29. These covariance matrices also show that $P\left(\boldsymbol{s}_n \boldsymbol{s}_c | M_p\right) = P\left(\boldsymbol{s}_n \boldsymbol{s}_c | M_d\right)$.

7.2.3 Results

Using this method, we were able to correctly identify the source of 55 out of 60 sets of fibers based off of their MSP spectra. The 5 misclassified fibers were also misclassified by the investigator. The misclassification is likely due to the high level of similarity between multiple sets of fibers from different sources. We show the results in table 7.1 where green labels are correct classification and red labels are misclassifications.

7.2.4 Discussion

Implementing the method developed in this research resulted in very satisfactory results for a proof-of-concept. For this dataset, we were able to correctly classify the source of MSP with an objective method at a higher rate than a trained forensic examiner. The model required only eight parameters to be estimated from the scores and and can also be used to assign the probative value of evidence. Once coded, running this algorithm took less than 30 minutes, a significant time improvement over human-comparison.

We designed the kernel to reflect what a forensic technician measures. However, we did not spend time fine-tuning the kernel and it may be possible to increase discrimination with the kernel. We do need to investigate further the rescaling method used.

7.1

Table 7.1: Classification results for proportional posterior probability

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\boldsymbol{x}_{11}$	**492.8**	3677.4	1785.4	9841.3	13415	1121.1	4090.1	2737.5	2572.2	2364.3	2017.1	1192.1	1161	5759.7	3321.3	4952.7	1607.5	3128.2	1825.2	4079.3
$\boldsymbol{x}_{12}$	**309.7**	1274.6	342.3	5438.4	16627.7	701.8	1633.9	719	649	568	424.2	702.6	1162.4	3236.3	1081.9	2216.4	319.3	935.2	355.6	1671
$\boldsymbol{x}_{13}$	**409.1**	2306.5	877.7	7472	14840.4	740.9	2620.5	1536.5	1419.5	1283	1036.3	817.9	1021.5	4393.6	2022.7	3397.7	781.7	1841.4	902	2708.9
$\boldsymbol{x}_{21}$	3049.5	**16.4**	667.7	2333.9	21955.1	2263.5	771.3	209.2	258.1	351.5	511.5	1712.8	2672.8	1437.6	194	529.2	778.7	159	646	277.4
$\boldsymbol{x}_{22}$	2781.5	**16.5**	522.7	2516.8	21442.1	2008.7	755.5	146.1	184.8	255.7	388.6	1530.5	2467.1	1537.3	162.6	615.4	627.1	121.3	503.9	287.2
$\boldsymbol{x}_{23}$	2888.7	**16.6**	580.5	2427.4	21656.1	2110.5	752.1	167.9	210.3	293.5	437.2	1604.8	2547.3	1484.1	174.8	572.3	688.4	132.4	560.8	278.4
$\boldsymbol{x}_{31}$	1577.2	582	**29.7**	4195.1	18201.3	861.8	1150	252.9	222.6	169	119.4	728	1414.5	2567.9	484.5	1516.1	137.8	395.9	114.3	968.5
$\boldsymbol{x}_{32}$	1634.2	495	**21.3**	4025.3	18477.4	915.9	1087.5	203.6	179.1	130.8	96.4	759.9	1478.7	2468.5	414.8	1420.6	135.5	334	102.2	874.7
$\boldsymbol{x}_{33}$	1598.3	549	**26.6**	4130.1	18306.9	881.6	1125.4	233.5	205.3	154	110.1	739.8	1438.4	2530	457.8	1479.4	136.2	371.8	109	933
$\boldsymbol{x}_{41}$	11620.9	3487	6321.8	**769.1**	57936.8	10555	4216.5	4257.3	4418.6	5215	5841.9	8232.9	9477.1	3324.2	4176.6	2182	6758.9	3938.9	6360.7	3722.8
$\boldsymbol{x}_{42}$	5621.1	685	2211.7	780.5	29729.6	4730.1	1345	1112.5	1210.1	1584.5	1922.3	3560	4588	1283.3	1052.3	**400.1**	2419.6	926.7	2200.5	841.9
$\boldsymbol{x}_{43}$	6268.8	868.8	2600.8	778.8	32260.9	5355.3	1560.3	1378.5	1488.9	1896.4	2281.2	4034.4	5125.3	1343.1	1279.4	**462.4**	2842	1166.1	2596.8	1044.7
$\boldsymbol{x}_{51}$	11366.7	18640.7	14104.6	46024.6	**2091.2**	10898.7	18545.6	17400.6	17231.5	15605.1	14652.1	11902.9	10280.8	21938.8	17866.4	21560.4	13433.3	18136.6	14175.7	19025.5
$\boldsymbol{x}_{52}$	11704.5	18731.5	14398.7	38911.2	**1966.6**	11254.3	18658.8	17121.3	16899.9	15845.3	14959	12094.1	10543.4	21842	18049.4	21773.1	13765	18130	14475.5	19148.3
$\boldsymbol{x}_{53}$	13019.6	20025	15784.8	37893.8	**1877.5**	12576.4	20282.7	18478.2	18280	17226.9	16351.3	13351.9	11791	22815.1	19388.8	23223.9	15135.3	19500.6	15881.7	20492.9
$\boldsymbol{x}_{61}$	1353	2136.9	783.7	7189.7	14973.2	**246.2**	2499.5	1415.6	1302.3	1164.7	930.1	755	992.6	4227.8	1872.2	3231.1	691.8	1704.3	809.1	2553.3
$\boldsymbol{x}_{62}$	1340.8	2103.6	770.1	7133	14966.8	**237.2**	2482.8	1392.3	1283.1	1145.5	914.3	732.3	977.2	4188.5	1844.9	3200.2	678.2	1680.2	796.6	2524.9
$\boldsymbol{x}_{63}$	1313.8	1942.3	683.1	6852.8	15147.6	**218.1**	2366.7	1277.4	1172.7	1034.6	816.9	683.4	963.9	4032.3	1699.8	3039.1	598.7	1547.6	710.9	2371.3
$\boldsymbol{x}_{71}$	2550.9	141	450.2	2559.2	21053.3	1785.8	139.7	**133.2**	154.1	254.2	348.4	1371.7	2197.2	1337.6	254.7	623	560.9	130.5	449.7	381.5
$\boldsymbol{x}_{72}$	4047	709.6	1501.4	2309.1	22773.9	3211.3	**141.4**	896.7	943.3	1161	1336	2555.6	3647.5	1277.9	1091	831.7	1798.7	879.4	1609.7	911.1
$\boldsymbol{x}_{73}$	2180	192.5	275	2913.3	19999.9	1441.5	143.9	**81.2**	86	153.4	206.5	1110.9	1915.6	1647.4	249.2	792.1	370.2	122.5	277	469.7
$\boldsymbol{x}_{81}$	2048.4	200	187.8	3079.5	19709.7	1306.1	812.2	**11.3**	46.2	91.1	131.2	1041	1841.6	1964.3	214	928	265.1	112.9	184.8	507.8
$\boldsymbol{x}_{82}$	2264.1	131.1	277.9	2771.8	20288.9	1512.5	750.7	**10.8**	51.9	128.3	197.8	1187.1	2005	1783.1	185	774.1	366.2	79.7	271.2	408.1
$\boldsymbol{x}_{83}$	2258.6	143.6	280.6	2757	20307.5	1508.4	771.6	**10.8**	50.6	134.4	202.5	1186.6	2016.9	1819.6	195.1	784.4	368.2	84.2	274.4	422.8
$\boldsymbol{x}_{91}$	2046.6	209.4	191.8	3061.3	19712.1	1305.4	814.5	47.9	**11.7**	97.6	136.3	1040.1	1845.6	1967.6	223.6	926	270.5	115.1	189.6	516.5
$\boldsymbol{x}_{92}$	2033.7	219.5	189.4	3075.3	19671.1	1293.1	821	50	**12**	99.7	136	1032.6	1834	1985	231	936.4	267.2	119.7	187.5	528.3
$\boldsymbol{x}_{93}$	2235.6	157.8	280.3	2753	20265.1	1487.5	765.8	44.1	**10.5**	140.4	204.7	1174.5	1994.7	1817.8	209	785.7	369.4	88.1	275.5	436.7
$\boldsymbol{x}_{101}$	1810	329.1	106.1	3654.9	19175.2	1082.5	961.2	118.8	107.3	**12**	79.2	870.9	1650.5	2255.6	287.5	1212.9	166.4	219.4	107.9	681.1
$\boldsymbol{x}_{102}$	1852.1	301.7	114.4	3571.7	19307.1	1121.5	932.8	104.2	94.8	**11.8**	81.9	899.2	1688.5	2205.3	267.9	1166.7	178.5	198.4	115	646.4
$\boldsymbol{x}_{103}$	2167	145	221	3112	20184.9	1425	820.3	68.8	79.5	**10.7**	148.2	1112.8	1969.2	1918	169.2	922.9	305.7	116.7	213.3	436.1
$\boldsymbol{x}_{111}$	1642.2	492.1	100.6	4014.1	18473.4	925.1	1079.4	202.5	178.2	133.5	**23.1**	771.7	1470.3	2468.4	411.8	1392.9	140.1	330.2	105.1	862.3
$\boldsymbol{x}_{112}$	1711.7	405.2	97.7	3835	18767.9	991.6	1016.8	156.5	138.6	100.7	**16.2**	812	1544	2362.2	344.7	1296.3	146.3	270.4	100.7	766
$\boldsymbol{x}_{113}$	1834.1	294.2	115.2	3574.8	19207.3	1109.2	931.5	106.2	98	73.2	**10**	889.2	1657.8	2201.6	263	1156.7	176.2	198.2	114.7	633.2
$\boldsymbol{x}_{121}$	1402.9	1949.2	882.8	6245.1	15351	785.4	2345.3	1386.5	1317.9	1186.2	1004	**176.9**	934	3609.3	1760.4	2971.4	784.1	1625.6	915.5	2312.2
$\boldsymbol{x}_{122}$	1278.6	1204.1	396.8	5102.2	16080	624.9	1450	755	703.9	602.8	475.5	**142.6**	925.8	2856.6	1062.6	2149.3	347.5	945.4	422.9	1561.9
$\boldsymbol{x}_{123}$	1299.4	1202.4	423.4	5068.5	16173.5	649.9	1558.4	774.5	726.7	621.7	498.7	**137.2**	930.8	2856.7	1067.5	2155.3	373	959.2	451.1	1560.1
$\boldsymbol{x}_{131}$	1582.9	2867.7	1431.7	7409.1	13947.5	978	3314.3	2130	2036.3	1872.6	1606.8	980.1	**379**	4473	2600.6	3686.9	1271.6	2419.5	1455.8	3042.5
$\boldsymbol{x}_{132}$	1410.4	2233.7	1022.2	6338.6	14513.2	786.5	2692.9	1615.5	1535.5	1388.3	1166.4	774.4	**319.8**	3811	2012	2999.3	897.1	1855.2	1049.3	2433.1
$\boldsymbol{x}_{133}$	1531.9	2662.7	1283.2	7072.9	14184.2	915.4	3095.1	1960.6	1868.5	1703.7	1449.2	935.5	**383.5**	4275.9	2400.7	3462.4	1139.7	2233.7	1309.9	2838.1
$\boldsymbol{x}_{141}$	3340.3	**502.9**	1339.1	2199.2	20244.2	2832.4	1002.3	757.4	814	984	1173.1	2041.7	2794	556.4	739	609.8	1442.8	629.2	1318.5	650
$\boldsymbol{x}_{142}$	5149.8	1362.4	2627.1	2582.2	30598.6	4567.1	1794.5	1829.1	1914.6	2125.9	2395.8	3490	4326.8	**557.5**	1686.4	1192.5	2763.4	1598.8	2592.8	1467.4
$\boldsymbol{x}_{143}$	5051.7	990	2344.6	1815.8	15188.2	4408.9	1264.6	1429.1	1514.9	1824.4	2108.6	3255	4056.2	**557**	1370.3	686.5	2511.2	1198.5	2313	1054.2
$\boldsymbol{x}_{151}$	2860.5	55	549.7	2544.9	21884	2079.9	879.5	170.2	211.4	273.5	410.9	1583	2545.2	1574.9	**30.4**	597.2	649.4	140	517.3	284.2
$\boldsymbol{x}_{152}$	2792.2	55.6	511.3	2579.9	21760.4	2015.2	868.6	151	188.7	248.6	379	1537.4	2491.7	1591.1	**30.7**	613.2	610.7	127.9	480.8	285.1
$\boldsymbol{x}_{153}$	2048.7	176.3	167.9	3245.5	20116.7	1309.1	935.6	68.7	72	67.6	110.4	1043.2	1883.5	1981.1	**24**	930.7	246.1	128.6	157.9	465.3
$\boldsymbol{x}_{161}$	5289.9	700	2209.7	1311.9	27212.8	4477.6	1383.3	1177.5	1273.3	1603.6	1923.7	3446.4	4279	1177.3	1004.8	**92.1**	2340.4	934.2	2138.1	851.1
$\boldsymbol{x}_{162}$	2845.2	107.7	759.2	1945.9	20734	2136.9	794	265.2	309.5	461.3	609	1720.4	2387.9	1263.6	259.7	**99.6**	836.4	190.3	726.9	262.5
$\boldsymbol{x}_{163}$	4244.4	375.1	1547.4	1519.3	24092.5	3470.6	1075.1	737.6	816.1	1062.9	1316.1	2673	3455.3	1136.7	618.1	**97**	1656.2	560	1494.2	525.8
$\boldsymbol{x}_{171}$	1488.1	661.1	128.3	4396.4	17670.7	785	1327.4	312.4	280.2	216.7	151.7	667.9	1275.3	2645	550.6	1569.6	**41.5**	459.5	137.7	1024.3
$\boldsymbol{x}_{172}$	1506.7	585.7	111.1	4265.3	17808.1	804.2	1284.3	272.1	244.4	180.4	126.6	663.5	1296.8	2568.8	489.3	1502.2	**32.1**	409	121.1	946.6
$\boldsymbol{x}_{173}$	1483.1	691.3	136.3	4449.9	17626	779.9	1346.9	330	295.9	232	162.8	671.1	1270.4	2678.2	574.9	1598.5	**45.7**	480.7	145.6	1054.8
$\boldsymbol{x}_{181}$	2581.3	94.2	429	2473.9	21279.5	1810.7	782.8	76.5	97.8	216.7	320.8	1395.9	2246.2	1577	193.1	597.3	524.6	**22.3**	412	331.3
$\boldsymbol{x}_{182}$	2664.1	67.3	453.5	2491.6	21409.5	1889.9	786	92.3	120.2	220.4	335.5	1442.2	2328.1	1539.1	169.9	586	550.9	**23.6**	433.8	299.2
$\boldsymbol{x}_{183}$	2172.1	148.1	226.4	2991.6	20129.7	1419	845.9	43.3	50.1	98.2	155.1	1113.2	1931.8	1825.3	180.3	817.4	307.2	**20.4**	217.5	432.3
$\boldsymbol{x}_{191}$	1614.3	543.5	108.5	4190.9	18347.2	898.9	1217.4	234.5	207.7	155.5	113.9	758.5	1437.1	2534.4	442	1451.3	143.9	366	**32.5**	907.7
$\boldsymbol{x}_{192}$	1600.8	564.3	111.7	4234	18279.1	886.5	1233.9	247.2	219	165	119.8	751.4	1422.1	2558.9	458.9	1474.9	144.9	381.4	**34.4**	929.9
$\boldsymbol{x}_{193}$	1673.4	457.3	100.4	4013.6	18622.6	955.4	1149.4	185.8	164.7	119.9	93.9	792	1498.1	2428.2	374.6	1352.8	144.4	304.8	**24**	816.5
$\boldsymbol{x}_{201}$	3304.6	117.6	998	2277	21930.2	2619.4	880.3	391.8	455.2	609.1	808.1	1999	2699.9	1277.1	310.7	493.6	1089.5	277.8	962.7	**59.7**
$\boldsymbol{x}_{202}$	1966.5	71.2	281.3	3078.5	19352.1	1349.6	826	70.9	88.1	128.9	198.3	1115.6	1695.6	1738.2	122.7	832.4	354.9	80.9	268.5	**46.7**
$\boldsymbol{x}_{203}$	3517.6	158.7	1128.6	2179.4	22349	2824.5	911.6	465.7	534.2	708.9	924.7	2148.7	2851.6	1234.7	371.5	467.3	1222.6	334.5	1084.5	**63.9**

Part III

Discussion and conclusions

8 Discussion

During this research, we identified several potential negative impacts of the continued development of SBFs on the criminal justice system despite their convenience and simplicity to quantify the weight of forensic evidence. The implications on the criminal justice system are twofold: (1) to prevent forensic conclusions which have a statistically sound appearance but are in fact prejudicial to be reported in court; (2) to further explore the validity and properties of KBFs as a more rigorous method for using similarity measures in a way that guarantees a fair and impartial administration of criminal justice. The KBF method developed in this research combines the data reduction power of score-based methods and the statistical rigor of the traditional Bayes factor to objectively quantify the weight of forensic evidence.

Our purpose was to develop a kernel model, in an analogous manner to the Gantz and SSaunders model, that captures the dependency structure existing between scores from hierarchically-sampled objects in order to quantify the weight of forensic evidence. This model allowed us to address major flaws identified with SBF techniques, restated here for convenience:

1. The dimension of the likelihood function of the SBF depends only on the scoring function and not on the number of trace and control samples that may have been observed i.e. it can only handle a single score between a single trace object and a single control object, which unnecesarily limits the use of the model in practice when multiple trace and control objects are recovered;

2. The use of the scoring function to measure the pairwise similarity between a set of objects induces dependencies between the resulting scores. These dependencies are not accounted for in SBF algorithms proposed in literature. This poses two problems:

   (a) SBF techniques assume independence between scores and as a result, the likelihood structures assigned to model them are wrong. Multiple SBF likelihood structures have been proposed in the literature, all providing different values for the weight of a given score [64, 30, 3];

   (b) If the scores were independent, the resulting weights of evidence could be easily combined. However, since they are not independent they cannot be combined and as such there does not exist a way to combine the evidence from multiple objects collected in connection with a crime;

3. SBFs are extremely dependent on the choice of the control material from the suspected source, which may be a problem if there exists a large variability between multiple control objects from a given source (e.g., handwriting, tool marks, footwear impressions);

4. None of the SBFs satisfy basic properties shown by the BF such as the coherency principle, given in definition 2.1;

The KBF method developed in this dissertation has addressed these problems by considering a vector of scores $\boldsymbol{s}$ and its distribution. The dimension of the likelihood for $\boldsymbol{s}$, $N = \binom{nn_o}{2}$, is a function of n sources and n_o objects per source. In this manner, the likelihood function adjusts to the number of trace and control samples being considered.. This is apparent when observing the three subsets of $\boldsymbol{s}$:
$\boldsymbol{s}_m,\ \boldsymbol{s}_n,\ \boldsymbol{s}_c$ where:

1. $\boldsymbol{s}_m = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\boldsymbol{x}_{ij} \in e_{u1}$, $\boldsymbol{x}_{kl} \in e_{u2}$ (this is trace vs. trace)

2. $\boldsymbol{s}_n = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\boldsymbol{x}_{ij} \in \{e_{u1},\, e_{u2}\}$, $\boldsymbol{x}_{kl} \in e_a$ (this is trace vs. controls)

3. $\boldsymbol{s}_c = \kappa\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right)$ s.t. $\left(\boldsymbol{x}_{ij},\, \boldsymbol{x}_{kl}\right) \in e_a$ (this is control vs. control)

4. $\boldsymbol{s} = \begin{pmatrix} \boldsymbol{s}_m \\ \boldsymbol{s}_n \\ \boldsymbol{s}_c \end{pmatrix}$

Accounting for the dependency structure for $\boldsymbol{s}$ allows us to model these scores in a rigorous manner[20]. We have derived the form of the distribution of vectors of scores which: 1) allows us to calculate the weight of evidence for multiple trace and control objects simultaneously; 2) accounts for variability between objects within a source in the dependency structure $\boldsymbol{\Sigma}$ of the model; 3) and enables us to combine the weight of evidence of multiple observation of a given trace. Finally, we note that the Bayes factor in equation 5.30 satisfies the coherency principle.

Our model relies on very few assumptions. The main assumption is the assumption of multivariate normality of the vector-of-scores, $\boldsymbol{s}$. We have demonstrated that the scores will asymptotically converge to a multivariate normal distribution as the intrinsic dimension p grows large under certain conditions, and we have shown empirically that this convergence holds in general. The experiments to study convergence of $\boldsymbol{s}$ to a multivariate normal were expository in their design and as such we never explored formally the criteria for convergence. That said, we were surprised to observe that the rate of convergence to normality was faster for between-source scores than within-source scores. It was found that accepting the null hypothesis for multivariate normality occured at an object dimension of $p = 64$. Through our practical examples, we have shown that even when p is not large enough to satisfy asymptotic normality, it is possible to construct a kernel so that the assumption of normality is approximately satisfied.

[20] Instead of as erroneously assumed independent objects with unknown parameteric distribution.

We did not find an analytical solution to the integrals required to assign the Bayes factor. Therefore, assigning the KBF requires estimating those integrals by MCMC methods. We proposed two algorithms to assign the KBF depending on whether the numbers of objects in the evidence sets are the same, providing that the reference population is studied through a balanced experiment. Our approach enables us to assign posterior distributions to the parameters of our score model from the sufficient statistics of the data in a very efficient manner and thus, to easily obtain large numbers of samples from these posterior distributions which in turn allows us to reduce the MCMC error. Although our method relies on the requirement that the general population is studied using a balanced experiment, this may not always be possible. In these situations, undersampling or jacknife techniques might be used to obtain a balanced number of samples for all the sources in e_a. The effect of these techniques on the posterior distributions of the model parameters might be significant and should be investigated further.

The development of our model has left us with two open concerns. We have not investigated the loss of information induced by the use of a kernel on the original objects. While we believe that by accounting for the dependency between the scores in the KBF to recover some of the information lost by mapping the original feature space to the kernel space, we have not explored the convergence of the KBF to what would be the BF in the original feature space for a given set of evidence.

Another major concern is related to the scalability of the model. While the rules to build the design matrices of the model are simple and the subsequent calculation of the summary statistics of the model parameters are trivial in theory, the practical implimentation of the model is computationally intensive. The dimension of $\mathbf{\Sigma}$ is $\binom{nn_o}{2} \times \binom{nn_o}{2}$ and our calculations require us to perform its eigen decomposition and to invert parts or all of it. Computations were completed in R (3.3.1) on a Windows 8.1 machine with 16 Gb of RAM and 3.8 ghz 4-core CPU. In some testing we did not go beyond $nn_o = 120$ due to computational resources. However, with further work and more efficient code, this number may be able to be increased. As such, it remains open for future research.

9 Conclusion

In this research, we developed a kernel model that considers the dependency structure for the distribution of a vector of pairwise scores calculated between all objects in the sets of evidence e_{u1}, e_{u2}, and e_a. The model requires only eight parameters to be estimated to describe the dependencies, irrespective of n and n_o. Estimating the parameters requires at minimum that e_a is a balanced sample of a relevant population of sources. Once parameters are estimated, we can build the dependency structure for any sort of unbalanced samples of e_{u1} and e_{u2}. This offers a significant amount of flexibility when calculating the weight of forensic evidence as the KBF.

$$KBF = \frac{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_p\right) d\Pi\left(\psi\right)}{\int f\left(\boldsymbol{s}|\psi,\, \mathcal{M}_d\right) d\Pi\left(\psi\right)}$$

In order to estimate the parameters for the KBF, we make use of a very efficient sampler, as outlined in seciton 5.6. The most computationally expensive part of parameters estimation comes from building the design matrices $\mathbf{B},\ \mathbf{D},\ \mathbf{T},\ \mathbf{W}$ and computing the eigenstructure of the sample covariance matrix $\mathbf{\Gamma}$.

Another aspect of the flexibility of the model is that no matter what the original form of evidence was (e.g. toolmarks, spectra, fingermarks, etc.), once the vector of pairwise comparisons $\boldsymbol{s}$ is computed, the computation of the KBF is the same. The user of the model is able to build a kernel that is specific to the evidence being considered, so long as the final form satisfies the condition of definition 3.1. By taking advantage of the closure properties of kernels, we can measure and combine multiple aspects of evidence into a new kernel. Theoretically, any type of data can be compared with the use of these kernels. We used this when applying the model to the MSP dataset by using a cross correlation and squared-Euclidean distance to account for the shape and difference in intensity of the spectra. This was a relatively simple kernel construction and it may be possible to create kernels that account for additional aspects of the data, such as the count of intersections between two spectra to increase discriminative power.

In summary, the major contributions of this research to the field of forensic statistics are:

1. A parametric model that considers the dependency structure for the distribution of a vector of scores $\boldsymbol{s}$;
   (a) A model that can consider multiple pieces of trace evidence and control samples simultaneously;
   (b) A model that can quantify the weight of forensic evidence for multiple trace objects simultaneously;
2. A method that is generalizable to multiple forms of trace evidence (i.e. spectra, VSPs, fingerprints, etc.);
3. A flexible method to use and create new similarity scores to combine measurements of different aspects of evidence (i.e. shape, closeness, counts, etc.);
4. A rigorous method to calculate a kernel Bayes factor for the weight of forensic evidence.

As we have shown, the careful consideration of the dependency structure in this KBF model corrects the shortcomings of the SBF and allows for the continued use of similarity scores in a rigorous manner to quantify the weight of evidence. While the method requires further validation, it is a significant step towards, “unbiased and quantifiable measures of uncertainty in the conclusions of forensic analyses” [42].

9.0.1 Future work

> “The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research.” — Paul A.M. Dirac

While only time and members of the scientific community will tell the true measure of this reseach, we believe have opened many doors for future research and development of this kernel method for the KBF. The generalization and relative ease of application of this method to many forensic evidence types can be extended to data types outside of the interests of the forensic world.

Future areas of research into this method include:

- Issues with scalability with large samples of n and n_o;
- Robustness of the model for parameter estimation when within-source variance differs between sources;
- Unbalanced sampling and its effect on likelihood evaluation and parameter estimation;
- Kernel choice and construction for different data types;
- Loss of information from data reduction;
- Efficient optimization of kernel tuning parameters to ensure correct eigenstructure;
- Convergence rates of $\boldsymbol{s}$ to multivariate normality;
- Development of a specific source score model using a mixed effects model.

Developing and researching these areas would further solidify the foundations of kernel based methods for the quantification of the weight of forensic evidence, validate its use in practice, increase its efficiency, and create guidelines for the use of different classes of kernel for different data types. We plan to continue research in these areas as it is a promising improvement to current score-based techniques for the quantification of the weight of forensic evidence.

Part IV

Appendix

10 Appendix

10.1 Covariance calculations

10.1.1 Sampling-driven model

This appendix section contains the covariance calculations for the sampling-driven model, developed in collaboration with John Miller (George Mason University). Some identities used in these calculations for $X \sim N\left(0,\, \sigma^2\right)$ are:

$$\begin{aligned} E\left[X^2\right] &= Var\left[X\right] = \sigma^2 \\ E\left[X^4\right] &= 3\sigma^4 \\ Var\left[X^2\right] &= E\left[X^4\right] - \left(E\left[X^2\right]\right)^2 \\ &= 2\sigma^4 \end{aligned}$$

$$\begin{aligned} E\left(s_{ijkl}|i=k\right) &= E\left[\left(a_i-a_i\right)^2+\left(r_{ij}-r_{il}\right)^2+2\left(a_i-a_i\right)\left(r_{ij}-r_{il}\right)\right] \\ &= E\left[r_{ij}^2+r_{il}^2-2r_{ij}r_{il}\right] \\ &= 2\sigma_r^2 \end{aligned}$$

$$\begin{aligned} E\left(s_{ijkl}|i\neq k\right) &= E\left[\left(a_i-a_k\right)^2+\left(r_{ij}-r_{kl}\right)^2+2\left(a_i-a_k\right)\left(r_{ij}-r_{kl}\right)\right] \\ &= E\left[\left(a_i-a_k\right)^2+\left(r_{ij}-r_{kl}\right)^2\right] \\ &= E\left[a_i^2+a_k^2-2a_ia_k+r_{ij}^2+r_{kl}^2-2r_{ij}r_{kl}\right] \\ &= 2\sigma_a^2+2\sigma_r^2 \end{aligned}$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{1112}\right)\\
&= Var\left[r_{11}^2 + r_{12}^2 - 2r_{11}r_{12} - 2\sigma_r^2\right]\\
&= 2\sigma_r^4 + 2\sigma_r^4 + 4\sigma_r^4\\
&= 8\sigma_r^4
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{1113}\right)\\
&= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{1113} - 2\sigma_r^2\right)\right]\\
&= E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(r_{11} - r_{13}\right)^2 - 2\sigma_r^2\right)\right]\\
&= E\left[\left(r_{11}^2 + r_{12}^2 - 2r_{11}r_{12} - 2\sigma_r^2\right)\left(r_{11}^2 + r_{13}^2 - 2r_{11}r_{13} - 2\sigma_r^2\right)\right]\\
&= E\left[\begin{array}{c} r_{11}^2\left(r_{11}^2 + r_{13}^2 - 2r_{11}r_{13} - 2\sigma_r^2\right)\\ +r_{12}^2\left(r_{11}^2 + r_{13}^2 - 2r_{11}r_{13} - 2\sigma_r^2\right)\\ -2r_{11}r_{12}\left(r_{11}^2 + r_{13}^2 - 2r_{11}r_{13} - 2\sigma_r^2\right)\\ -2\sigma_r^2\left(r_{11}^2 + r_{13}^2 - 2r_{11}r_{13} - 2\sigma_r^2\right) \end{array}\right]\\
&= E\left[\begin{array}{c} r_{11}^4 + r_{11}^2r_{13}^2 - 2r_{11}^2\sigma_r^2\\ +r_{12}^2r_{11}^2 + r_{12}^2r_{13}^2 - 2r_{12}^2\sigma_r^2\\ -0\\ -2\sigma_r^2r_{11}^2 - 2\sigma_r^2r_{13}^2 + 4\sigma_r^4 \end{array}\right]\\
&= 2\sigma_r^4
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{1314}\right)\\
&= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{1314} - 2\sigma_r^2\right)\right]\\
&= E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(r_{13} - r_{14}\right)^2 - 2\sigma_r^2\right)\right]\\
&= E\left[\left(r_{11}^2 + r_{12}^2 - 2r_{11}r_{12} - 2\sigma_r^2\right)\left(r_{13}^2 + r_{14}^2 - 2r_{13}r_{14} - 2\sigma_r^2\right)\right]\\
&= E\left[\begin{array}{c}
r_{11}^2\left(r_{13}^2 + r_{14}^2 - 2r_{13}r_{14} - 2\sigma_r^2\right)\\
+r_{12}^2\left(r_{13}^2 + r_{14}^2 - 2r_{13}r_{14} - 2\sigma_r^2\right)\\
-2r_{11}r_{12}\left(r_{13}^2 + r_{14}^2 - 2r_{13}r_{14} - 2\sigma_r^2\right)\\
-2\sigma_r^2\left(r_{13}^2 + r_{14}^2 - 2r_{13}r_{14} - 2\sigma_r^2\right)
\end{array}\right]\\
&= 0
\end{aligned}
$$

$$cov\left(s_{1121},\, s_{1121}\right)$$

$$= Var\left[s_{1121}\right]$$

$$= Var\left[a_1^2 + a_2^2 - 2a_1a_2 + r_{11}^2 + r_{21}^2 - 2r_{11}r_{21} + 2a_1r_{11} - 2a_1r_{21} - 2a_2r_{11} + 2a_2r_{21}\right]$$

$$= 2\sigma_a^4 + 2\sigma_a^4 + 4\sigma_a^4 + 2\sigma_r^4 + 2\sigma_r^4 + 4\sigma_r^4 + 4\sigma_a^2\sigma_r^2 + 4\sigma_a^2\sigma_r^2 + 4\sigma_a^2\sigma_r^2 + 4\sigma_a^2\sigma_r^2$$

$$= 8\sigma_a^4 + 8\sigma_r^4 + 16\sigma_a^2\sigma_r^2$$

$$cov\left(s_{1121},\, s_{1221}\right)$$

$$= E\left[\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\left(s_{1221} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E[\Big(\left(a_1 - a_2\right)^2 + \left(r_{11} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)$$

$$\times\ \Big(\left(a_1 - a_2\right)^2 + \left(r_{12} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{12} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)]$$

$$= E\left[\begin{array}{c}
a_1^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+a_2^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1a_2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{11}^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{21}^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2r_{11}r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_1r_{11}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_2r_{11}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_2r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_a^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_r^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{21}^2 + 2a_1r_{12} + 2a_2r_{21} - 2a_1a_2 - 2a_1r_{21} - 2a_2r_{12} - 2r_{12}r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right)
\end{array}\right]$$

$$= E\left[\begin{array}{c}
a_1^4 + a_1^2a_2^2 + a_1^2r_{12}^2 + a_1^2r_{21}^2 - 2\sigma_a^2a_1^2 - 2a_1^2\sigma_r^2 \\
+a_2^2a_1^2 + a_2^4 + a_2^2r_{12}^2 + a_2^2r_{21}^2 - 2a_2^2\sigma_a^2 - 2a_2^2\sigma_r^2 \\
+4a_1^2a_2^2 \\
+r_{11}^2a_1^2 + r_{11}^2a_2^2 + r_{11}^2r_{12}^2 + r_{11}^2r_{21}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\
+r_{21}^2a_1^2 + r_{21}^2a_2^2 + r_{21}^2r_{12}^2 + r_{21}^4 - 2r_{21}^2\sigma_a^2 - 2r_{21}^2\sigma_r^2 \\
+4a_1^2r_{21}^2 \\
+4a_2^2r_{21}^2 \\
-2\sigma_a^2a_1^2 - 2\sigma_a^2a_2^2 - 2\sigma_a^2r_{12}^2 - 2\sigma_a^2r_{21}^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_r^2a_1^2 - 2\sigma_r^2a_2^2 - 2\sigma_r^2r_{12}^2 - 2\sigma_r^2r_{21}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4
\end{array}\right]$$

$$= \quad \begin{array}{c}
3\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+\sigma_a^4 + 3\sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+4\sigma_a^4 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + 3\sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
+4\sigma_a^2\sigma_r^2 \\
+4\sigma_a^2\sigma_r^2 \\
-2\sigma_a^4 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4
\end{array}$$

$$= 8\sigma_a^4 + 8\sigma_a^2\sigma_r^2 + 2\sigma_r^4$$

$cov\left(s_{1121},\, s_{1222}\right)$

$= E\left[\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\left(s_{1222} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$

$= E[\Big((a_1 - a_2)^2 + (r_{11} - r_{21})^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)$

$= \times\left((a_1 - a_2)^2 + (r_{12} - r_{22})^2 + 2\left(a_1 - a_2\right)\left(r_{12} - r_{22}\right) - 2\sigma_a^2 - 2\sigma_r^2\right)]$

$$= E\begin{bmatrix} a_1^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +a_2^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2a_1a_2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +r_{11}^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +r_{21}^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2r_{11}r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +2a_1r_{11}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2a_1r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2a_2r_{11}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +2a_2r_{21}\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2\sigma_a^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2\sigma_r^2\left(a_1^2 + a_2^2 + r_{12}^2 + r_{22}^2 + 2a_1r_{12} + 2a_2r_{22} - 2a_1a_2 - 2a_1r_{22} - 2a_2r_{12} - 2r_{12}r_{22} - 2\sigma_a^2 - 2\sigma_r^2\right) \end{bmatrix}$$

$$= E\begin{bmatrix} a_1^4 + a_1^2a_2^2 + a_1^2r_{12}^2 + a_1^2r_{22}^2 - 2\sigma_a^2a_1^2 - 2a_1^2\sigma_r^2 \\ +a_2^2a_1^2 + a_2^4 + a_2^2r_{12}^2 + a_2^2r_{22}^2 - 2a_2^2\sigma_a^2 - 2a_2^2\sigma_r^2 \\ +4a_1^2a_2^2 \\ +r_{11}^2a_1^2 + r_{11}^2a_2^2 + r_{11}^2r_{12}^2 + r_{11}^2r_{22}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\ +r_{21}^2a_1^2 + r_{21}^2a_2^2 + r_{21}^2r_{12}^2 + r_{21}^2r_{22}^2 - 2r_{21}^2\sigma_a^2 - 2r_{21}^2\sigma_r^2 \\ -2\sigma_a^2a_1^2 - 2\sigma_a^2a_2^2 - 2\sigma_a^2r_{12}^2 - 2\sigma_a^2r_{21}^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\ -2\sigma_r^2a_1^2 - 2\sigma_r^2a_2^2 - 2\sigma_r^2r_{12}^2 - 2\sigma_r^2r_{21}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{bmatrix}$$

$$= \begin{matrix} 3\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\ +\sigma_a^4 + 3\sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\ +4\sigma_a^4 \\ +\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\ +\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\ -2\sigma_a^4 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\ -2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_a^2\sigma_r^2 + 4\sigma_r^4 \end{matrix}$$

$= 8\sigma_a^4$

$cov\left(s_{1112},\, s_{1121}\right)$

$$= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(a_1 - a_2\right)^2 + \left(r_{11} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\begin{array}{c} r_{11}^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{11}^2 + r_{21}^2 - 2r_{11}r_{21} + 2a_1r_{11} - 2a_1r_{21} - 2a_2r_{11} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +r_{12}^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{11}^2 + r_{21}^2 - 2r_{11}r_{21} + 2a_1r_{11} - 2a_1r_{21} - 2a_2r_{11} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2r_{11}r_{12}\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{11}^2 + r_{21}^2 - 2r_{11}r_{21} + 2a_1r_{11} - 2a_1r_{21} - 2a_2r_{11} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2\sigma_r^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{11}^2 + r_{21}^2 - 2r_{11}r_{21} + 2a_1r_{11} - 2a_1r_{21} - 2a_2r_{11} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \end{array}\right]$$

$$= E\left[\begin{array}{c} r_{11}^2a_1^2 + r_{11}^2a_2^2 + r_{11}^4 + r_{11}^2r_{21}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\ +r_{12}^2a_1^2 + r_{12}^2a_2^2 + r_{12}^2r_{11}^2 + r_{12}^2r_{21}^2 - 2r_{12}^2\sigma_a^2 - 2r_{12}^2\sigma_r^2 \\ -2\sigma_r^2a_1^2 - 2\sigma_r^2a_2^2 - 2\sigma_r^2r_{11}^2 - 2\sigma_r^2r_{21}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}\right]$$

$$= \begin{array}{c} \sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + 3\sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ +\sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ -2\sigma_r^2\sigma_a^2 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}$$

$$= 2\sigma_r^4$$

$$cov\left(s_{1112},\, s_{1321}\right)$$

$$= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{1321} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(a_1 - a_2\right)^2 + \left(r_{13} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{13} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\begin{array}{c} r_{11}^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{13}^2 + r_{21}^2 - 2r_{13}r_{21} + 2a_1r_{13} - 2a_1r_{21} - 2a_2r_{13} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +r_{12}^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{13}^2 + r_{21}^2 - 2r_{13}r_{21} + 2a_1r_{13} - 2a_1r_{21} - 2a_2r_{13} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2r_{11}r_{12}\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{13}^2 + r_{21}^2 - 2r_{13}r_{21} + 2a_1r_{13} - 2a_1r_{21} - 2a_2r_{13} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2\sigma_r^2\left(a_1^2 + a_2^2 - 2a_1a_2 + r_{13}^2 + r_{21}^2 - 2r_{13}r_{21} + 2a_1r_{13} - 2a_1r_{21} - 2a_2r_{13} + 2a_2r_{21} - 2\sigma_a^2 - 2\sigma_r^2\right) \end{array}\right]$$

$$= E\left[\begin{array}{c} r_{11}^2a_1^2 + r_{11}^2a_2^2 + r_{11}^2r_{13}^2 + r_{11}^2r_{21}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\ +r_{12}^2a_1^2 + r_{12}^2a_2^2 + r_{12}^2r_{13}^2 + r_{12}^2r_{21}^2 - 2r_{12}^2\sigma_a^2 - 2r_{12}^2\sigma_r^2 \\ -2\sigma_r^2a_1^2 - 2\sigma_r^2a_2^2 - 2\sigma_r^2r_{13}^2 - 2\sigma_r^2r_{21}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}\right]$$

$$= \begin{array}{c} \sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ +\sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ -2\sigma_r^2\sigma_a^2 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}$$

$$= 0$$

$$cov\left(s_{1121},\, s_{1131}\right)$$

$$= E\left[\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\left(s_{1131} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E[\Big(\left(a_1 - a_2\right)^2 + \left(r_{11} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)$$

$$\times\Big(\left(a_1 - a_3\right)^2 + \left(r_{11} - r_{31}\right)^2 + 2\left(a_1 - a_3\right)\left(r_{11} - r_{31}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)]$$

$$= E\left[\begin{array}{c}
a_1^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
+a_2^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2a_1a_2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
+r_{11}^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
+r_{21}^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2r_{11}r_{21}\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
+2a_1r_{11}\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2a_1r_{21}\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2a_2r_{11}\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
+2a_2r_{21}\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2\sigma_a^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)\\
-2\sigma_r^2\left(a_1^2 + a_3^2 + r_{11}^2 + r_{31}^2 + 2a_1r_{11} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{11} - 2r_{11}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)
\end{array}\right]$$

$$= E\left[\begin{array}{c}
a_1^4 + a_1^2a_3^2 + a_1^2r_{11}^2 + a_1^2r_{31}^2 - 2a_1^2\sigma_a^2 - 2a_1^2\sigma_r^2\\
+a_2^2a_1^2 + a_2^2a_3^2 + a_2^2r_{11}^2 + a_2^2r_{31}^2 - 2a_2^2\sigma_a^2 - 2a_2^2\sigma_r^2\\
+r_{11}^2a_1^2 + r_{11}^2a_3^2 + r_{11}^4 + r_{11}^2r_{31}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2\\
+r_{21}^2a_1^2 + r_{21}^2a_3^2 + r_{21}^2r_{11}^2 + r_{21}^2r_{31}^2 - 2r_{21}^2\sigma_a^2 - 2r_{21}^2\sigma_r^2\\
+4a_1^2r_{11}^2\\
-2\sigma_a^2a_1^2 - 2\sigma_a^2a_3^2 - 2\sigma_a^2r_{11}^2 - 2\sigma_a^2r_{31}^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2\\
-2\sigma_r^2a_1^2 - 2\sigma_r^2a_3^2 - 2\sigma_r^2r_{11}^2 - 2\sigma_r^2r_{31}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4
\end{array}\right]$$

$$= \begin{array}{c}
3\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2\\
+\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2\\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + 3\sigma_4^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4\\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4\\
+4\sigma_a^2\sigma_r^2\\
-2\sigma_a^4 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2\\
-2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_a^2\sigma_r^2 + 4\sigma_r^4
\end{array}$$

$$= 2\sigma_a^4 + 4\sigma_a^2\sigma_r^2 + 2\sigma_r^4$$

$$
\begin{aligned}
& cov\left(s_{1112},\, s_{2122}\right) \\
& = E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{2122} - 2\sigma_r^2\right)\right] \\
& = E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(r_{21} - r_{22}\right)^2 - 2\sigma_r^2\right)\right] \\
& = E\left[\begin{array}{c}
r_{11}^2\left(r_{21}^2 + r_{22}^2 - 2r_{21}r_{22} - 2\sigma_r^2\right) \\
+r_{12}^2\left(r_{21}^2 + r_{22}^2 - 2r_{21}r_{22} - 2\sigma_r^2\right) \\
-2r_{11}r_{12}\left(r_{21}^2 + r_{22}^2 - 2r_{21}r_{22} - 2\sigma_r^2\right) \\
-2\sigma_r^2\left(r_{21}^2 + r_{22}^2 - 2r_{21}r_{22} - 2\sigma_r^2\right)
\end{array}\right] \\
& = E\left[\begin{array}{c}
r_{11}^2 r_{21}^2 + r_{11}^2 r_{22}^2 - 2r_{11}^2\sigma_r^2 \\
+r_{12}^2 r_{21}^2 + r_{12}^2 r_{22}^2 - 2r_{12}^2\sigma_r^2 \\
-2\sigma_r^2 r_{21}^2 - 2\sigma_r^2 r_{22}^2 + 4\sigma_r^4
\end{array}\right] \\
& = \begin{array}{c}
\sigma_r^4 + \sigma_r^4 - 2\sigma_r^4 \\
+\sigma_r^4 + \sigma_r^4 - 2\sigma_r^4 \\
-2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_r^4
\end{array} \\
& = 0
\end{aligned}
$$

$$cov\left(s_{1121},\, s_{1231}\right)$$

$$= E\left[\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\left(s_{1231} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E[\Big(\left(a_1 - a_2\right)^2 + \left(r_{11} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)$$

$$\times\Big(\left(a_1 - a_3\right)^2 + \left(r_{12} - r_{31}\right)^2 + 2\left(a_1 - a_3\right)\left(r_{12} - r_{31}\right) - 2\sigma_a^2 - 2\sigma_r^2\Big)]$$

$$= E\begin{bmatrix}
a_1^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+a_2^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1a_2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{11}^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{21}^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2r_{11}r_{21}\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_1r_{11}\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1r_{21}\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_2r_{11}\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_2r_{21}\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_a^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_r^2\left(a_1^2 + a_3^2 + r_{12}^2 + r_{31}^2 + 2a_1r_{12} + 2a_3r_{31} - 2a_1a_3 - 2a_1r_{31} - 2a_3r_{12} - 2r_{12}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right)
\end{bmatrix}$$

$$= E\begin{bmatrix}
a_1^4 + a_1^2a_3^2 + a_1^2r_{12}^2 + a_1^2r_{31}^2 - 2a_1^2\sigma_a^2 - 2a_1^2\sigma_r^2 \\
+a_2^2a_1^2 + a_2^2a_3^2 + a_2^2r_{12}^2 + a_2^2r_{31}^2 - 2a_2^2\sigma_a^2 - 2a_2^2\sigma_r^2 \\
+r_{11}^2a_1^2 + r_{11}^2a_3^2 + r_{11}^2r_{12}^2 + r_{11}^2r_{31}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\
+r_{21}^2a_1^2 + r_{21}^2a_3^2 + r_{21}^2r_{12}^2 + r_{21}^2r_{31}^2 - 2r_{21}^2\sigma_a^2 - 2r_{21}^2\sigma_r^2 \\
-2\sigma_a^2a_1^2 - 2\sigma_a^2a_3^2 - 2\sigma_a^2r_{12}^2 - 2\sigma_a^2r_{31}^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_r^2a_1^2 - 2\sigma_r^2a_3^2 - 2\sigma_r^2r_{12}^2 - 2\sigma_r^2r_{31}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4
\end{bmatrix}$$

$$= \begin{matrix}
3\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
-2\sigma_a^4 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_a^2\sigma_r^2 + 4\sigma_r^4
\end{matrix}$$

$$= 2\sigma_a^4$$

$$cov\left(s_{1112},\, s_{2131}\right)$$

$$= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{2131} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\left(\left(r_{11} - r_{12}\right)^2 - 2\sigma_r^2\right)\left(\left(a_2 - a_3\right)^2 + \left(r_{21} - r_{31}\right)^2 + 2\left(a_2 - a_3\right)\left(r_{21} - r_{31}\right)^2 - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\begin{array}{c} r_{11}^2\left(a_2^2 + a_3^2 + r_{21}^2 + r_{31}^2 + 2a_2r_{21} + 2a_3r_{31} - 2a_2a_3 - 2a_2r_{31} - 2a_3r_{21} - 2r_{21}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ +r_{12}^2\left(a_2^2 + a_3^2 + r_{21}^2 + r_{31}^2 + 2a_2r_{21} + 2a_3r_{31} - 2a_2a_3 - 2a_2r_{31} - 2a_3r_{21} - 2r_{21}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2r_{11}r_{12}\left(a_2^2 + a_3^2 + r_{21}^2 + r_{31}^2 + 2a_2r_{21} + 2a_3r_{31} - 2a_2a_3 - 2a_2r_{31} - 2a_3r_{21} - 2r_{21}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \\ -2\sigma_r^2\left(a_2^2 + a_3^2 + r_{21}^2 + r_{31}^2 + 2a_2r_{21} + 2a_3r_{31} - 2a_2a_3 - 2a_2r_{31} - 2a_3r_{21} - 2r_{21}r_{31} - 2\sigma_a^2 - 2\sigma_r^2\right) \end{array}\right]$$

$$= E\left[\begin{array}{c} r_{11}^2a_2^2 + r_{11}^2a_3^2 + r_{11}^2r_{21}^2 + r_{11}^2r_{31}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\ +r_{12}^2a_2^2 + r_{12}^2a_3^2 + r_{12}^2r_{21}^2 + r_{12}^2r_{31}^2 - 2r_{12}^2\sigma_a^2 - 2r_{12}^2\sigma_r^2 \\ -2\sigma_r^2a_2^2 - 2\sigma_r^2a_3^2 - 2\sigma_r^2r_{21}^2 - 2\sigma_r^2r_{31}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}\right]$$

$$= \begin{array}{c} \sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ +\sigma_r^2\sigma_a^2 + \sigma_r^2\sigma_a^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 \\ -2\sigma_r^2\sigma_a^2 - 2\sigma_r^2\sigma_a^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4 \end{array}$$

$$= 0$$

$$cov\left(s_{1121},\, s_{3141}\right)$$

$$= E\left[\left(s_{1121} - 2\sigma_a^2 - 2\sigma_r^2\right)\left(s_{3141} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E[\Big(\left(a_1 - a_2\right)^2 + \left(r_{11} - r_{21}\right)^2 + 2\left(a_1 - a_2\right)\left(r_{11} - r_{21}\right)^2 - 2\sigma_a^2 - 2\sigma_r^2\Big)$$

$$\times \Big(\left(a_3 - a_4\right)^2 + \left(r_{31} - r_{41}\right)^2 + 2\left(a_3 - a_4\right)\left(r_{31} - r_{41}\right)^2 - 2\sigma_a^2 - 2\sigma_r^2\Big)]$$

$$= E\left[\begin{array}{c}
a_1^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+a_2^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1a_2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{11}^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+r_{21}^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2r_{11}r_{21}\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_1r_{11}\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_1r_{21}\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2a_2r_{11}\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
+2a_2r_{21}\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_a^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right) \\
-2\sigma_r^2\left(a_3^2 + a_4^2 + r_{31}^2 + r_{41}^2 + 2a_3r_{31} + 2a_4r_{41} - 2a_3a_4 - 2a_3r_{41} - 2a_4r_{31} - 2r_{31}r_{41} - 2\sigma_a^2 - 2\sigma_r^2\right)
\end{array}\right]$$

$$= E\left[\begin{array}{c}
a_1^2a_3^2 + a_1^2a_4^2 + a_1^2r_{31}^2 + a_1^2r_{41}^2 - 2a_1^2\sigma_a^2 - 2a_1^2\sigma_r^2 \\
+a_2^2a_3^2 + a_2^2a_4^2 + a_2^2r_{31}^2 + a_2^2r_{41}^2 - 2a_2^2\sigma_a^2 - 2a_2^2\sigma_r^2 \\
+r_{11}^2a_3^2 + r_{11}^2a_4^2 + r_{11}^2r_{31}^2 + r_{11}^2r_{41}^2 - 2r_{11}^2\sigma_a^2 - 2r_{11}^2\sigma_r^2 \\
+r_{21}^2a_3^2 + r_{21}^2a_4^2 + r_{21}^2r_{31}^2 + r_{21}^2r_{41}^2 - 2r_{21}^2\sigma_a^2 - 2r_{21}^2\sigma_r^2 \\
-2\sigma_a^2a_3^2 - 2\sigma_a^2a_4^2 - 2\sigma_a^2r_{31}^2 - 2\sigma_a^2r_{41}^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_r^2a_3^2 - 2\sigma_r^2a_4^2 - 2\sigma_r^2r_{31}^2 - 2\sigma_r^2r_{41}^2 + 4\sigma_r^2\sigma_a^2 + 4\sigma_r^4
\end{array}\right]$$

$$= \begin{array}{c}
\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+\sigma_a^4 + \sigma_a^4 + \sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
+\sigma_a^2\sigma_r^2 + \sigma_a^2\sigma_r^2 + \sigma_r^4 + \sigma_r^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 \\
-2\sigma_a^4 - 2\sigma_a^4 - 2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 + 4\sigma_a^4 + 4\sigma_a^2\sigma_r^2 \\
-2\sigma_a^2\sigma_r^2 - 2\sigma_a^2\sigma_r^2 - 2\sigma_r^4 - 2\sigma_r^4 + 4\sigma_a^2\sigma_r^2 + 4\sigma_r^4
\end{array}$$

$$= 0$$

10.1.2 Score model covariance calculations

The expectations and covariances for the score model is given here. We restate the score model for conveinence

$$s_{ijkl} = \begin{cases} \theta_b + b_i + b_k + d_{ik} + t_{i:ij} + t_{i:kl} + t_{k:ij} + t_{k:kl} + w_{ij} + w_{kl} + e^b_{ijkl} & if\ i \neq k \\ \theta_w + w_{ij} + w_{kl} + e^w_{ijkl} & if\ i = k \end{cases}$$

where $b_i \sim N\left(0,\, \sigma_b^2\right)$, $d_{ik} \sim N\left(0,\, \sigma_d^2\right)$, $w_{ij} \sim N\left(0,\, \sigma_w^2\right)$, $t_{i:ij} \sim N\left(0,\, \sigma_t^2\right)$, $e^b_{ijkl} \sim N\left(0,\, \sigma^2_{e^b}\right)$, and $e^w_{ijkl} \sim N\left(0,\, \sigma^2_{e^w}\right)$.

$$\begin{aligned} E\left(s_{ijkl} | i = k\right) &= E\left[\theta_w + w_{ij} + w_{kl} + e^w_{ijkl}\right] \\ &= \theta_w \end{aligned}$$

$$\begin{aligned} E\left(s_{ijkl} | i \neq k\right) &= E\left[\theta_b + b_i + b_k + d_{ik} + t_{i:ij} + t_{i:kl} + t_{k:ij} + t_{k:kl} + w_{ij} + w_{kl} + e^b_{ijkl}\right] \\ &= \theta_b \end{aligned}$$

$$\begin{aligned} & cov\left(s_{1112},\, s_{1112}\right) \\ & = Var\left[\theta_w + w_{11} + w_{12} + e^w_{1112}\right] \\ & = 2\sigma_w^2 + \sigma^2_{e^w} \end{aligned}$$

$$
\begin{aligned}
&cov\left(s_{1112},\ s_{1113}\right) \\
&= E\left[\left(s_{1112}-\theta_{w}\right)\left(s_{1113}-\theta_{w}\right)\right] \\
&= E\left[\left(w_{11}+w_{12}+e^{w}_{1112}\right)\left(w_{11}+w_{13}+e^{w}_{1113}\right)\right] \\
&= E\left[w_{11}^{2}\right] \\
&= \sigma_{w}^{2}
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1112},\ s_{1314}\right) \\
&= E\left[\left(s_{1112}-\theta_{w}\right)\left(s_{1314}-\theta_{w}\right)\right] \\
&= E\left[\left(w_{11}+w_{12}+e^{w}_{1112}\right)\left(w_{13}+w_{14}+e^{w}_{1314}\right)\right] \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1121},\ s_{1121}\right) \\
&= Var\left[s_{1121}\right] \\
&= Var\left[\theta_{b}+b_{i}+b_{k}+d_{ik}+t_{i:ij}+t_{i:kl}+t_{k:ij}+t_{k:kl}+w_{ij}+w_{kl}+e^{b}_{ijkl}\right] \\
&= 2\sigma_{b}^{2}+\sigma_{d}^{2}+4\sigma_{t}^{2}+2\sigma_{w}^{2}+\sigma_{e^{b}}^{2}
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1121},\, s_{1221}\right)\\
&= E\left[\left(s_{1121}-\theta_b\right)\left(s_{1221}-\theta_b\right)\right]\\
&= E\big[\big(b_1+b_2+d_{12}+t_{1:11}+t_{1:21}+t_{2:11}+t_{2:21}+w_{11}+w_{21}+e^b_{1121}\big)\\
&\times\big(b_1+b_2+d_{12}+t_{1:12}+t_{1:21}+t_{2:12}+t_{2:21}+w_{12}+w_{21}+e^b_{1221}\big)\big]\\
&= E\left[b_1^2+b_2^2+d_{12}^2+t_{1:21}^2+t_{2:21}^2+w_{21}^2\right]\\
&= 2\sigma_b^2+\sigma_d^2+2\sigma_t^2+\sigma_w^2
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1121},\, s_{1222}\right)\\
&= E\left[\left(s_{1121}-\theta_b\right)\left(s_{1222}-\theta_b\right)\right]\\
&= E\big[\big(b_1+b_2+d_{12}+t_{1:11}+t_{1:21}+t_{2:11}+t_{2:21}+w_{11}+w_{21}+e^b_{1121}\big)\\
&\times\big(b_1+b_2+d_{12}+t_{1:12}+t_{1:22}+t_{2:12}+t_{2:22}+w_{12}+w_{22}+e^b_{1222}\big)\big]\\
&= E\left[b_1^2+b_2^2+d_{12}^2\right]\\
&= 2\sigma_b^2+\sigma_d^2
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{1121}\right)\\
&= E\left[\left(s_{1112}-\theta_w\right)\left(s_{1121}-\theta_b\right)\right]\\
&= E\left[\left(w_{11}+w_{12}+e^w_{1112}\right)\big(b_1+b_2+d_{12}+t_{1:11}+t_{1:21}+t_{2:11}+t_{2:21}+w_{11}+w_{21}+e^b_{1121}\big)\right]\\
&= E\left[w_{11}^2\right]\\
&= \sigma_w^2
\end{aligned}
$$

$$cov\left(s_{1112},\, s_{1321}\right)$$

$$= E\left[\left(s_{1112} - 2\sigma_r^2\right)\left(s_{1321} - 2\sigma_a^2 - 2\sigma_r^2\right)\right]$$

$$= E\left[\left(w_{11} + w_{12} + e^w_{1112}\right)\left(b_1 + b_2 + d_{12} + t_{1:13} + t_{1:21} + t_{2:13} + t_{2:21} + w_{13} + w_{21} + e^b_{1321}\right)\right]$$

$$= 0$$

$$cov\left(s_{1121},\, s_{1131}\right)$$

$$= E\left[\left(s_{1121} - \theta_b\right)\left(s_{1131} - \theta_b\right)\right]$$

$$= E\big[\big(b_1 + b_2 + d_{12} + t_{1:11} + t_{1:21} + t_{2:11} + t_{2:21} + w_{11} + w_{21} + e^b_{1121}\big)$$

$$\times \big(b_1 + b_3 + d_{13} + t_{1:11} + t_{1:31} + t_{3:11} + t_{3:31} + w_{11} + w_{31} + e^b_{1131}\big)\big]$$

$$= E\left[b_1^2 + t_{1:11}^2 w_{11}^2\right]$$

$$= \sigma_b^2 + \sigma_t^2 + \sigma_w^2$$

$$cov\left(s_{1121},\, s_{1231}\right)$$

$$= E\left[\left(s_{1121} - \theta_b\right)\left(s_{1231} - \theta_b\right)\right]$$

$$= E\big[\big(b_1 + b_2 + d_{12} + t_{1:11} + t_{1:21} + t_{2:11} + t_{2:21} + w_{11} + w_{21} + e^b_{1121}\big)$$

$$\times \big(b_1 + b_3 + d_{13} + t_{1:12} + t_{1:31} + t_{3:12} + t_{3:31} + w_{12} + w_{31} + e^b_{1231}\big)\big]$$

$$= E\left[b_1^2\right]$$

$$= \sigma_b^2$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{2122}\right)\\
&= E\left[\left(s_{1112}-\theta_w\right)\left(s_{2122}-\theta_w\right)\right]\\
&= E\left[\left(w_{11}+w_{12}+e^{w}_{1112}\right)\left(w_{21}+w_{22}+e^{w}_{2122}\right)\right]\\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1112},\, s_{2131}\right)\\
&= E\left[\left(s_{1112}-\theta_w\right)\left(s_{2131}-\theta_b\right)\right]\\
&= E\left[\left(w_{11}+w_{12}+e^{w}_{1112}\right)\left(b_2+b_3+d_{23}+t_{2:21}+t_{2:31}+t_{3:21}+t_{3:31}+w_{21}+w_{31}+e^{b}_{2131}\right)\right]\\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
&cov\left(s_{1121},\, s_{3141}\right)\\
&= E\left[\left(s_{1121}-\theta_b\right)\left(s_{3141}-\theta_b\right)\right]\\
&= E[\left(b_1+b_2+d_{12}+t_{1:11}+t_{1:21}+t_{2:11}+t_{2:21}+w_{11}+w_{21}+e^{b}_{1121}\right)\\
&\times\left(b_3+b_4+d_{34}+t_{3:31}+t_{3:41}+t_{4:31}+t_{4:41}+w_{31}+w_{41}+e^{b}_{3141}\right)]\\
&= 0
\end{aligned}
$$

10.2 Supporting theorems and definitions

The theorems and definitions in this section were not used in the research. However, they are important theorems in kernel-based methods and as such deserve to be mentioned in this document. Their relevance may become important in future research into KBF methods

Mercer's theorem [56] allows the expression of a kernel in terms of its eigen-decomposition. Let T be a linear operator with respect to κ, L_∞ the function space of bounded measurable functions, and L_2 a Hilbert space.

Theorem 10.1. *Mercer's theorem*

Let $(X,\, \mu)$ be a finite measure space and $\kappa \in L_\infty\left(X^2,\, \mu^2\right)$ be a kernel such that, $T_\kappa :\ L_2\left(X,\, \mu\right) \to L_2\left(X,\, \mu\right)$ is positive definite. Let $\phi_i \in L_2\left(X,\, \mu\right)$ be the normalized eigenfunctions of T_κ associated with eigenvalues $\lambda_i > 0$. Then:

1. *The eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ are absolutely summable*
2. $$\kappa\left(\mathbf{x}\,\mathbf{x}'\right) = \sum_{i=1}^{\infty} \lambda_i \phi_i\left(\mathbf{x}\right) \phi_i\left(\mathbf{x}'\right)$$

   where the convergence is absolute and uniform.

It is assumed in general that there exist an infinite number of eigenfunctions $\phi_1\left(\mathbf{x}\right),\ \phi_2\left(\mathbf{x}\right), ...$ which are ranked by decreasing eigenvalues $\lambda_1,\ \lambda_2, ...$ and are orthogonal to each other. Mercer's theorem was suspected to be relevant when proving the distribution of all pairwise scores converges to a multivariate normal distribution.

Theorem 10.2. *Lindeberg-Feller Theorem [71] (CLT)*

For each n let $Y_{n,1}, ..., Y_{n,k_n}$ be independent random vectors with finite variances such that

$$\sum_{i}^{k_n} E\left\|Y_{n,i}\right\|^2 1\left\{\left\|Y_{n,i}\right\| > \varepsilon\right\} \quad \to 0, \quad \textit{every } \varepsilon > 0,$$

$$\sum_{i}^{k_n} Cov\, Y_{n,i} \quad \to \Sigma.$$

Then the sequence $\sum_{i=1}^{k_n}\left(Y_{n,i} - EY_{n,i}\right)$ converges in distribution to a normal $N\left(\mathbf{0},\ \Sigma\right)$ distribution.

Theorem 10.3. *Schoenberg's Basis for Smooth Functions [66]*

For a class of basis functions Ω_d in $\mathbb{R}^d$ and an isotropic stationary kernel, κ_I, as $d \to \infty$, $\Omega_d \to e^{-x^2}$ thus imposing a smoothness condition on the basis functions. As a result, any basis used for κ with high-dimensional data should have the form e^{-x^2}.

The results of Schoenberg says that as the dimension of data increases, the number of useful kernel families decreases. However, the size of the dimensions is not suggested and requires testing.

10.3 Appendix of design matrices

10.3.1 Design matrix A for sampling-driven model

Table 10.1: Example design matrix for **A** for $n = 3$ and $n_o = 3$

	1112	1113	1121	1122	1123	1131	1132	1133	1213	1221	1222	1223	1231	1232	1233	1321	1322	1323	1331	1332	1333	2122	2123	2131	2132	2133	2223	2231	2232	2233	2331	2332	2333	3132	3133	3233
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1122	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1123	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1131	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1132	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1133	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1213	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1221	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1222	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1223	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1231	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1232	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1233	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1321	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1322	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1323	0	0	8	8	8	2	2	2	0	8	8	8	2	2	2	8	8	8	2	2	2	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1331	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1332	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
1333	0	0	2	2	2	8	8	8	0	2	2	2	8	8	8	2	2	2	8	8	8	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2132	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2133	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2223	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2231	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2232	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2233	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2331	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2332	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
2333	0	0	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	8	8	8	0	8	8	8	8	8	8	0	0	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3133	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3233	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.3.2 Design matrix R for sampling-driven model

Table 10.2: Example design matrix for **R** for $n = 3$ and $n_o = 3$

	1112	1113	1121	1122	1123	1131	1132	1133	1213	1221	1222	1223	1231	1232	1233	1321	1322	1323	1331	1332	1333	2122	2123	2131	2132	2133	2223	2231	2232	2233	2331	2332	2333	3132	3133	3233
1112	8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1113	2	8	2	2	2	2	2	2	2	0	0	0	0	0	0	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	2	2	8	2	2	2	2	2	0	2	0	0	0	0	0	2	0	0	0	0	0	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
1122	2	2	2	8	2	2	2	2	0	0	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	2	2	2	2	0	0	0	0	0	0
1123	2	2	2	2	8	2	2	2	0	0	0	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	0	2	2	2	0	0	0
1131	2	2	2	2	2	8	2	2	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	0	2	2	0
1132	2	2	2	2	2	2	8	2	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	2	0	2
1133	2	2	2	2	2	2	2	8	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	2	2
1213	2	2	0	0	0	0	0	0	8	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1221	2	0	2	0	0	0	0	0	2	8	2	2	2	2	2	2	0	0	0	0	0	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
1222	2	0	0	2	0	0	0	0	2	2	8	2	2	2	2	0	2	0	0	0	0	2	0	0	0	0	2	2	2	2	0	0	0	0	0	0
1223	2	0	0	0	2	0	0	0	2	2	2	8	2	2	2	0	0	2	0	0	0	0	2	0	0	0	2	0	0	0	2	2	2	0	0	0
1231	2	0	0	0	0	2	0	0	2	2	2	2	8	2	2	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	0	2	2	0
1232	2	0	0	0	0	0	2	0	2	2	2	2	2	8	2	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	2	0	2
1233	2	0	0	0	0	0	0	2	2	2	2	2	2	2	8	0	0	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	2	0	2	2
1321	0	2	2	0	0	0	0	0	2	2	0	0	0	0	0	8	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0
1322	0	2	0	2	0	0	0	0	2	0	2	0	0	0	0	2	8	2	2	2	2	2	0	0	0	0	2	2	2	2	0	0	0	0	0	0
1323	0	2	0	0	2	0	0	0	2	0	0	2	0	0	0	2	2	8	2	2	2	0	2	0	0	0	2	0	0	0	2	2	2	0	0	0
1331	0	2	0	0	0	2	0	0	2	0	0	0	2	0	0	2	2	2	8	2	2	0	0	2	0	0	0	2	0	0	2	0	0	2	2	0
1332	0	2	0	0	0	0	2	0	2	0	0	0	0	2	0	2	2	2	2	8	2	0	0	0	2	0	0	0	2	0	0	2	0	2	0	2
1333	0	2	0	0	0	0	0	2	2	0	0	0	0	0	2	2	2	2	2	2	8	0	0	0	0	2	0	0	0	2	0	0	2	0	2	2
2122	0	0	2	2	0	0	0	0	0	2	2	0	0	0	0	2	2	0	0	0	0	8	2	2	2	2	2	2	2	2	0	0	0	0	0	0
2123	0	0	2	0	2	0	0	0	0	2	0	2	0	0	0	2	0	2	0	0	0	2	8	2	2	2	2	0	0	0	2	2	2	0	0	0
2131	0	0	2	0	0	2	0	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2	2	8	2	2	0	2	0	0	2	0	0	2	2	0
2132	0	0	2	0	0	0	2	0	0	2	0	0	0	2	0	2	0	0	0	2	0	2	2	2	8	2	0	0	2	0	0	2	0	2	0	2
2133	0	0	2	0	0	0	0	2	0	2	0	0	0	0	2	2	0	0	0	0	2	2	2	2	2	8	0	0	0	2	0	0	2	0	2	2
2223	0	0	0	2	2	0	0	0	0	0	2	2	0	0	0	0	2	2	0	0	0	2	2	0	0	0	8	2	2	2	2	2	2	0	0	0
2231	0	0	0	2	0	2	0	0	0	0	2	0	2	0	0	0	2	0	2	0	0	2	0	2	0	0	2	8	2	2	2	0	0	2	2	0
2232	0	0	0	2	0	0	2	0	0	0	2	0	0	2	0	0	2	0	0	2	0	2	0	0	2	0	2	2	8	2	0	2	0	2	0	2
2233	0	0	0	2	0	0	0	2	0	0	2	0	0	0	2	0	2	0	0	0	2	2	0	0	0	2	2	2	2	8	0	0	2	0	2	2
2331	0	0	0	0	2	2	0	0	0	0	0	2	2	0	0	0	0	2	2	0	0	0	2	2	0	0	2	2	0	0	8	2	2	2	2	0
2332	0	0	0	0	2	0	2	0	0	0	0	2	0	2	0	0	0	2	0	2	0	0	2	0	2	0	2	0	2	0	2	8	2	2	0	2
2333	0	0	0	0	2	0	0	2	0	0	0	2	0	0	2	0	0	2	0	0	2	0	2	0	0	2	2	0	0	2	2	2	8	0	2	2
3132	0	0	0	0	0	2	2	0	0	0	0	0	2	2	0	0	0	0	2	2	0	0	0	2	2	0	0	2	2	0	2	2	0	8	2	2
3133	0	0	0	0	0	2	0	2	0	0	0	0	2	0	2	0	0	0	2	0	2	0	0	2	0	2	0	2	0	2	2	0	2	2	8	2
3233	0	0	0	0	0	0	2	2	0	0	0	0	0	2	2	0	0	0	0	2	2	0	0	0	2	2	0	0	2	2	0	2	2	2	2	8

10.3.3 Design matrix $\mathbf{QQ}^t$ for sampling-driven model

Table 10.3: $\mathbf{QQ}^t$ for $n = 3$ and $n_o = 3$

	1112	1113	1121	1122	1123	1131	1132	1133	1213	1221	1222	1223	1231	1232	1233	1321	1322	1323	1331	1332	1333	2122	2123	2131	2132	2133	2223	2231	2232	2233	2331	2332	2333	3132	3133	3233
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1122	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1123	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1131	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1132	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1133	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1213	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1221	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1222	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1223	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1231	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1232	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1233	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1321	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1322	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1323	0	0	2	2	2	1	1	1	0	2	2	2	1	1	1	2	2	2	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1331	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1332	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
1333	0	0	1	1	1	2	2	2	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2132	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2133	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2223	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2231	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2232	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2233	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2331	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2332	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
2333	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2	2	2	0	2	2	2	2	2	2	0	0	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3133	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3233	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.3.4 Design matrix C for sampling-driven model

Table 10.4: Example design matrix for **C** for $n = 3$ and $n_o = 3$

	1112	1113	1121	1122	1123	1131	1132	1133	1213	1221	1222	1223	1231	1232	1233	1321	1322	1323	1331	1332	1333	2122	2123	2131	2132	2133	2223	2231	2232	2233	2331	2332	2333	3132	3133	3233
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	0	16	8	8	4	4	4	0	8	0	0	0	0	0	8	0	0	0	0	0	0	0	4	4	4	0	0	0	0	0	0	0	0	0	0
1122	0	0	8	16	8	4	4	4	0	0	8	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	4	4	4	0	0	0	0	0	0
1123	0	0	8	8	16	4	4	4	0	0	0	8	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	0	0	0
1131	0	0	4	4	4	16	8	8	0	0	0	0	8	0	0	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0	0	0
1132	0	0	4	4	4	8	16	8	0	0	0	0	0	8	0	0	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0	0
1133	0	0	4	4	4	8	8	16	0	0	0	0	0	0	8	0	0	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0
1213	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1221	0	0	8	0	0	0	0	0	0	16	8	8	4	4	4	8	0	0	0	0	0	0	0	4	4	4	0	0	0	0	0	0	0	0	0	0
1222	0	0	0	8	0	0	0	0	0	8	16	8	4	4	4	0	8	0	0	0	0	0	0	0	0	0	0	4	4	4	0	0	0	0	0	0
1223	0	0	0	0	8	0	0	0	0	8	8	16	4	4	4	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	0	0	0
1231	0	0	0	0	0	8	0	0	0	4	4	4	16	8	8	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0	0	0
1232	0	0	0	0	0	0	8	0	0	4	4	4	8	16	8	0	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0	0
1233	0	0	0	0	0	0	0	8	0	4	4	4	8	8	16	0	0	0	0	0	8	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0
1321	0	0	8	0	0	0	0	0	0	8	0	0	0	0	0	16	8	8	4	4	4	0	0	4	4	4	0	0	0	0	0	0	0	0	0	0
1322	0	0	0	8	0	0	0	0	0	0	8	0	0	0	0	8	16	8	4	4	4	0	0	0	0	0	0	4	4	4	0	0	0	0	0	0
1323	0	0	0	0	8	0	0	0	0	0	0	8	0	0	0	8	8	16	4	4	4	0	0	0	0	0	0	0	0	0	4	4	4	0	0	0
1331	0	0	0	0	0	8	0	0	0	0	0	0	8	0	0	4	4	4	16	8	8	0	0	4	0	0	0	4	0	0	4	0	0	0	0	0
1332	0	0	0	0	0	0	8	0	0	0	0	0	0	8	0	4	4	4	8	16	8	0	0	0	4	0	0	0	4	0	0	4	0	0	0	0
1333	0	0	0	0	0	0	0	8	0	0	0	0	0	0	8	4	4	4	8	8	16	0	0	0	0	4	0	0	0	4	0	0	4	0	0	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2123	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	0	4	0	0	4	0	0	0	4	0	0	4	0	0	4	0	0	4	0	0	0	0	16	8	8	0	8	0	0	8	0	0	0	0	0
2132	0	0	4	0	0	0	4	0	0	4	0	0	0	4	0	4	0	0	0	4	0	0	0	8	16	8	0	0	8	0	0	8	0	0	0	0
2133	0	0	4	0	0	0	0	4	0	4	0	0	0	0	4	4	0	0	0	0	4	0	0	8	8	16	0	0	0	8	0	0	8	0	0	0
2223	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2231	0	0	0	4	0	4	0	0	0	0	4	0	4	0	0	0	4	0	4	0	0	0	0	8	0	0	0	16	8	8	8	0	0	0	0	0
2232	0	0	0	4	0	0	4	0	0	0	4	0	0	4	0	0	4	0	0	4	0	0	0	0	8	0	0	8	16	8	0	8	0	0	0	0
2233	0	0	0	4	0	0	0	4	0	0	4	0	0	0	4	0	4	0	0	0	4	0	0	0	0	8	0	8	8	16	0	0	8	0	0	0
2331	0	0	0	0	4	4	0	0	0	0	0	4	4	0	0	0	0	4	4	0	0	0	0	8	0	0	0	8	0	0	16	8	8	0	0	0
2332	0	0	0	0	4	0	4	0	0	0	0	4	0	4	0	0	0	4	0	4	0	0	0	0	8	0	0	0	8	0	8	16	8	0	0	0
2333	0	0	0	0	4	0	0	4	0	0	0	4	0	0	4	0	0	4	0	0	4	0	0	0	0	8	0	0	0	8	8	8	16	0	0	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3133	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3233	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.3.5 $\mathbf{BB}^t$

Table 10.5: $\mathbf{BB}^t$ for $n = 4$, $n_o = 2$

	1112	1121	1122	1131	1132	1141	1142	1221	1222	1231	1232	1241	1242	2122	2131	2132	2141	2142	2231	2232	2241	2242	3132	3141	3142	3241	3242	4142
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	2	2	1	1	1	1	2	2	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1122	0	2	2	1	1	1	1	2	2	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1131	0	1	1	2	2	1	1	1	1	2	2	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0
1132	0	1	1	2	2	1	1	1	1	2	2	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0
1141	0	1	1	1	1	2	2	1	1	1	1	2	2	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0
1142	0	1	1	1	1	2	2	1	1	1	1	2	2	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0
1221	0	2	2	1	1	1	1	2	2	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1222	0	2	2	1	1	1	1	2	2	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1231	0	1	1	2	2	1	1	1	1	2	2	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0
1232	0	1	1	2	2	1	1	1	1	2	2	1	1	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0
1241	0	1	1	1	1	2	2	1	1	1	1	2	2	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0
1242	0	1	1	1	1	2	2	1	1	1	1	2	2	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	1	1	1	1	0	0	1	1	1	1	0	0	0	2	2	1	1	2	2	1	1	0	1	1	1	1	0
2132	0	1	1	1	1	0	0	1	1	1	1	0	0	0	2	2	1	1	2	2	1	1	0	1	1	1	1	0
2141	0	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	2	2	1	1	2	2	0	1	1	1	1	0
2142	0	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	2	2	1	1	2	2	0	1	1	1	1	0
2231	0	1	1	1	1	0	0	1	1	1	1	0	0	0	2	2	1	1	2	2	1	1	0	1	1	1	1	0
2232	0	1	1	1	1	0	0	1	1	1	1	0	0	0	2	2	1	1	2	2	1	1	0	1	1	1	1	0
2241	0	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	2	2	1	1	2	2	0	1	1	1	1	0
2242	0	1	1	0	0	1	1	1	1	0	0	1	1	0	1	1	2	2	1	1	2	2	0	1	1	1	1	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3141	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	2	2	2	2	0
3142	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	2	2	2	2	0
3241	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	2	2	2	2	0
3242	0	0	0	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	2	2	2	2	0
4142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.3.6 $\mathbf{DD}^t$

Table 10.6: $\mathbf{DD}^t$ for $n = 4$, $n_o = 2$

	1112	1121	1122	1131	1132	1141	1142	1221	1222	1231	1232	1241	1242	2122	2131	2132	2141	2142	2231	2232	2241	2242	3132	3141	3142	3241	3242	4142
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1122	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1131	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1132	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1141	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1142	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1221	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1222	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1231	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1232	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1241	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1242	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
2132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
2141	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0
2142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0
2231	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
2232	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
2241	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0
2242	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3141	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
3142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
3241	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
3242	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0
4142	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

10.3.7 $\mathbf{TT}^t$

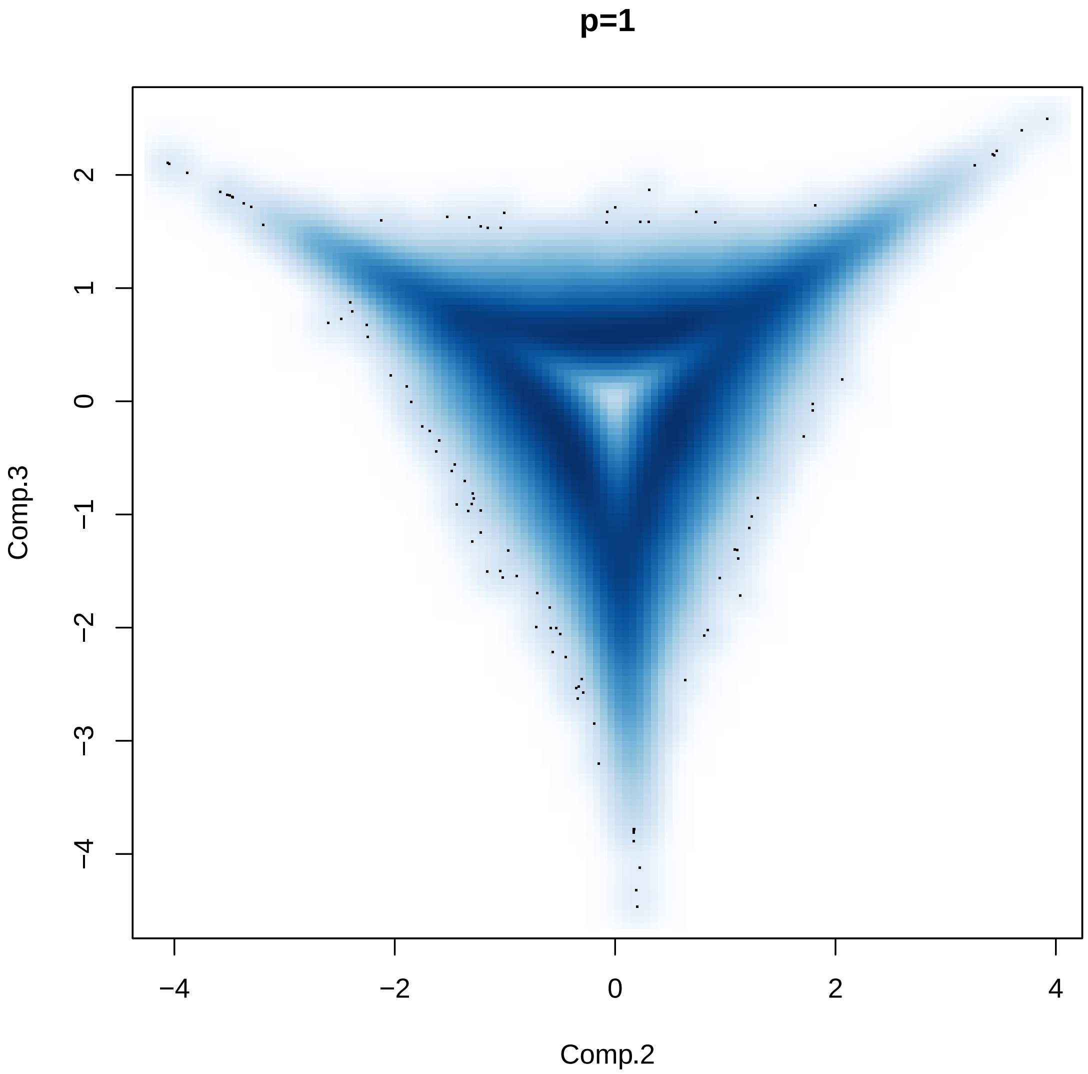
Table 10.7: $\mathbf{TT}^t$ for $n = 3$, $n_o = 2$

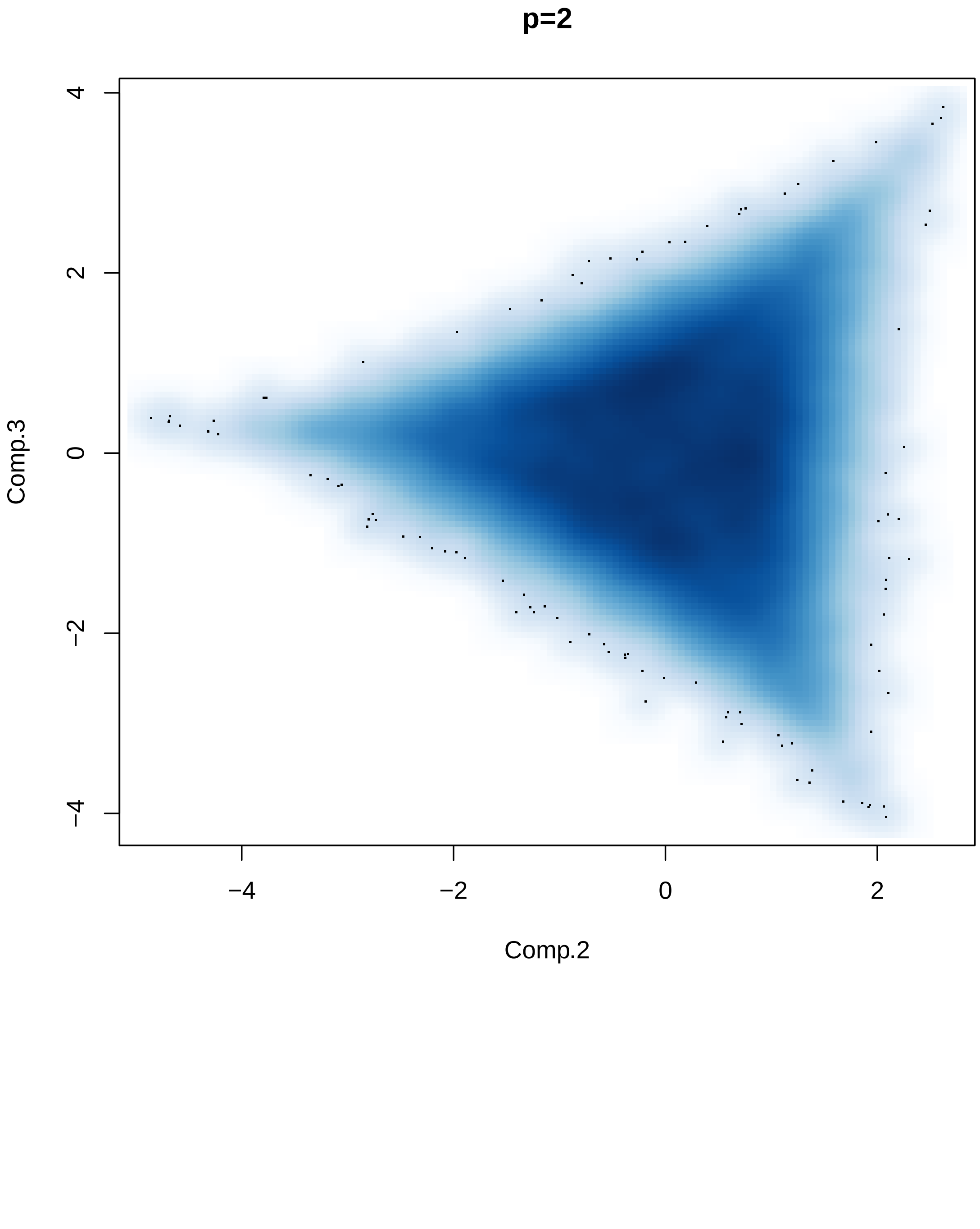
	1112	1121	1122	1131	1132	1221	1222	1231	1232	2122	2131	2132	2231	2232	3132
1112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1121	0	4	2	1	1	2	0	0	0	0	1	1	0	0	0
1122	0	2	4	1	1	0	2	0	0	0	0	0	1	1	0
1131	0	1	1	4	2	0	0	2	0	0	1	0	1	0	0
1132	0	1	1	2	4	0	0	0	2	0	0	1	0	1	0
1221	0	2	0	0	0	4	2	1	1	0	1	1	0	0	0
1222	0	0	2	0	0	2	4	1	1	0	0	0	1	1	0
1231	0	0	0	2	0	1	1	4	2	0	1	0	1	0	0
1232	0	0	0	0	2	1	1	2	4	0	0	1	0	1	0
2122	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2131	0	1	0	1	0	1	0	1	0	0	4	2	2	0	0
2132	0	1	0	0	1	1	0	0	1	0	2	4	0	2	0
2231	0	0	1	1	0	0	1	1	0	0	2	0	4	2	0
2232	0	0	1	0	1	0	1	0	1	0	0	2	2	4	0
3132	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

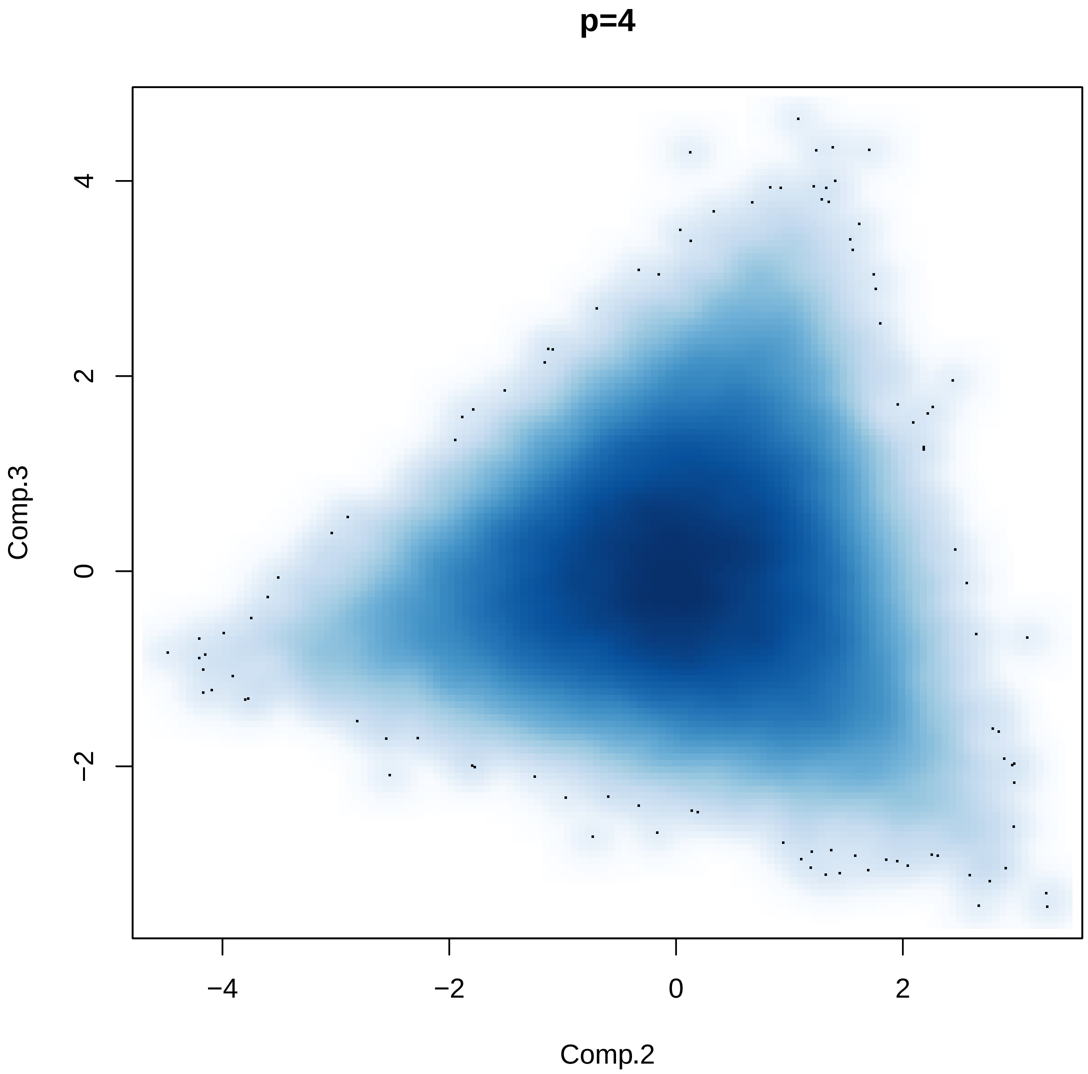
10.4 Convergence plots

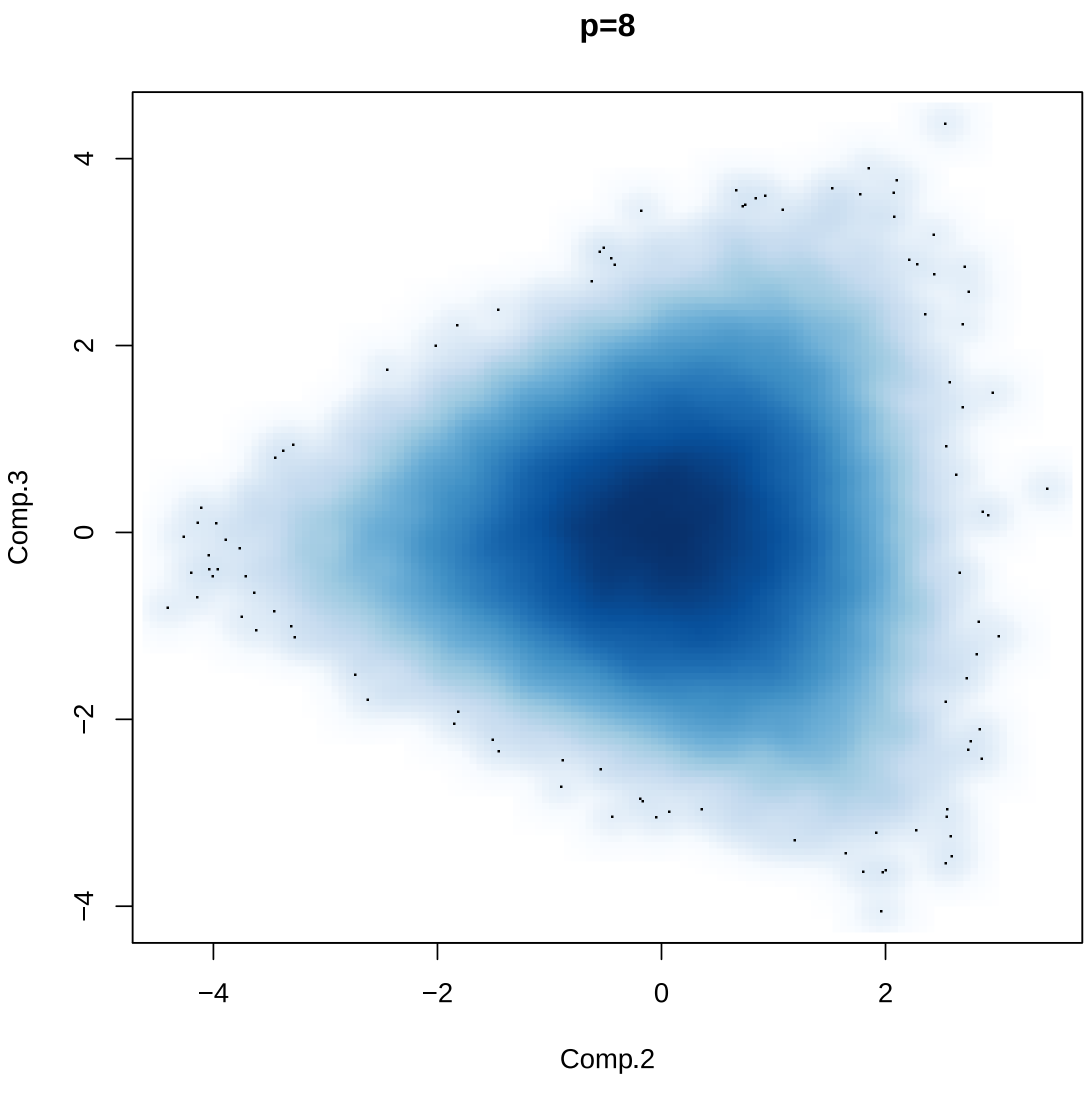
10.4.1 “Hollow-triangle” convergence plots for Theorem 5.4

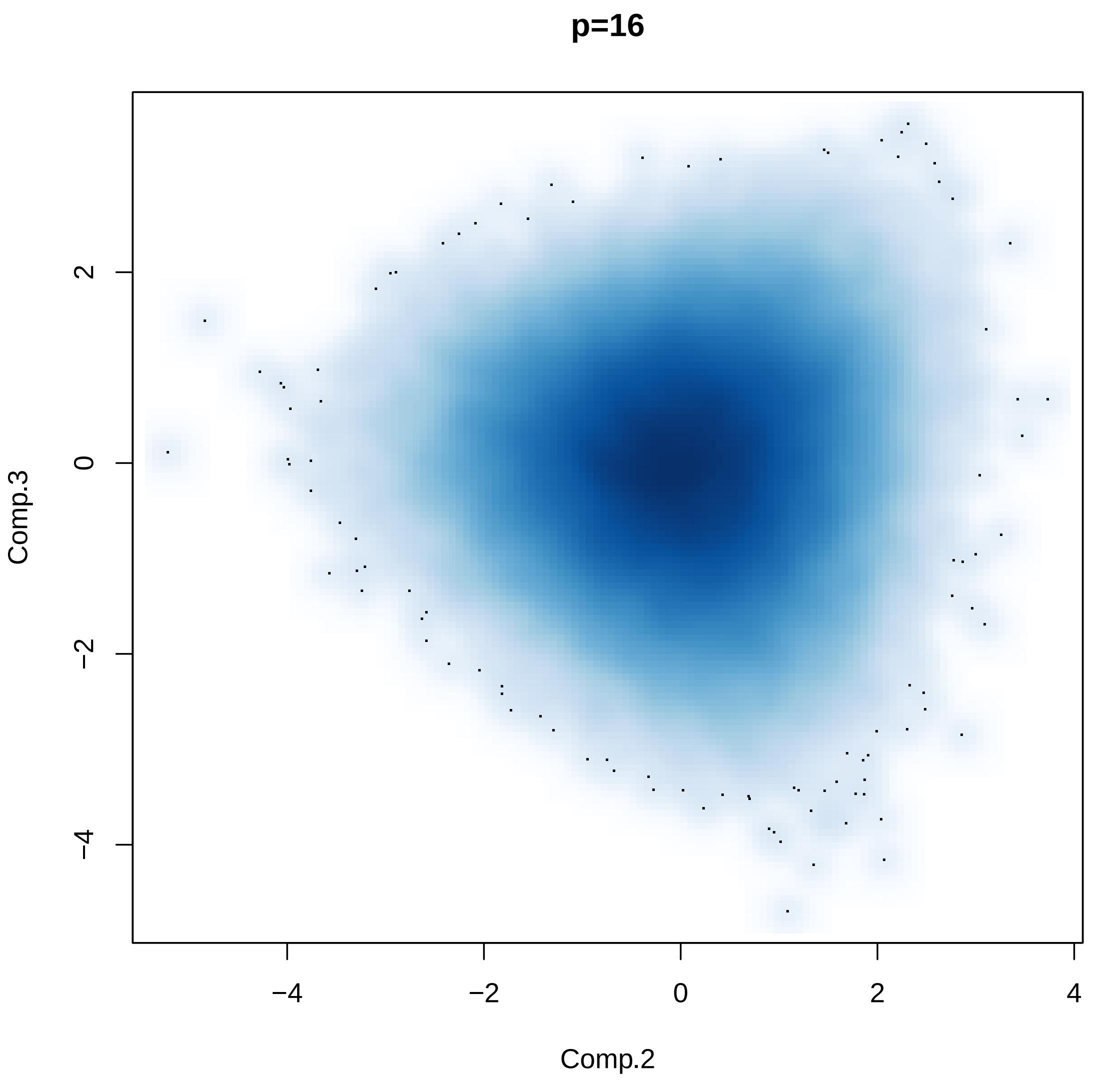
These convergence plots are the bivariate distributions of the 2^{nd} and 3^{rd} principal scores from the experiment 1 for the simulation study of Theorem 5.4.

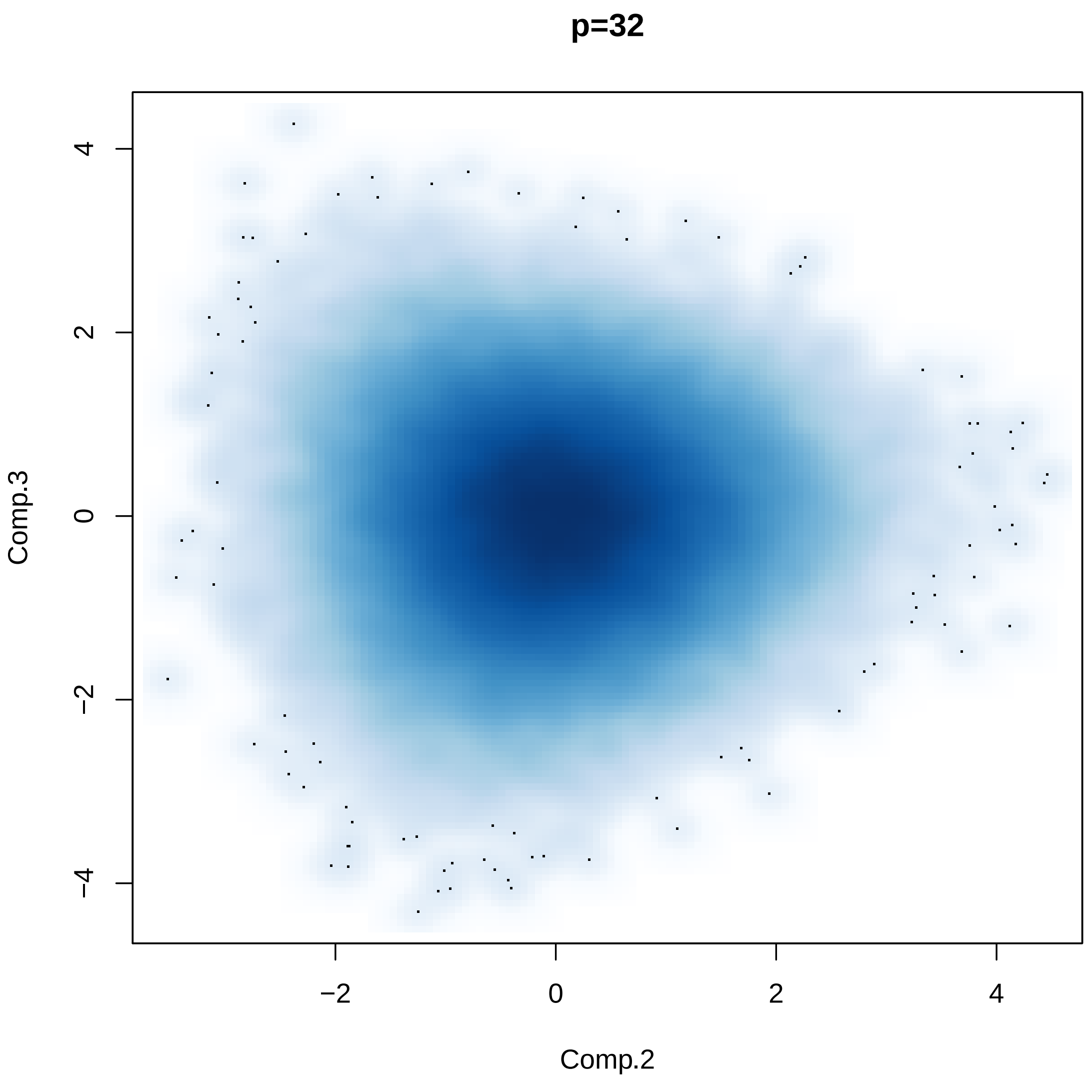

p=1
Comp.3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4
Comp.2

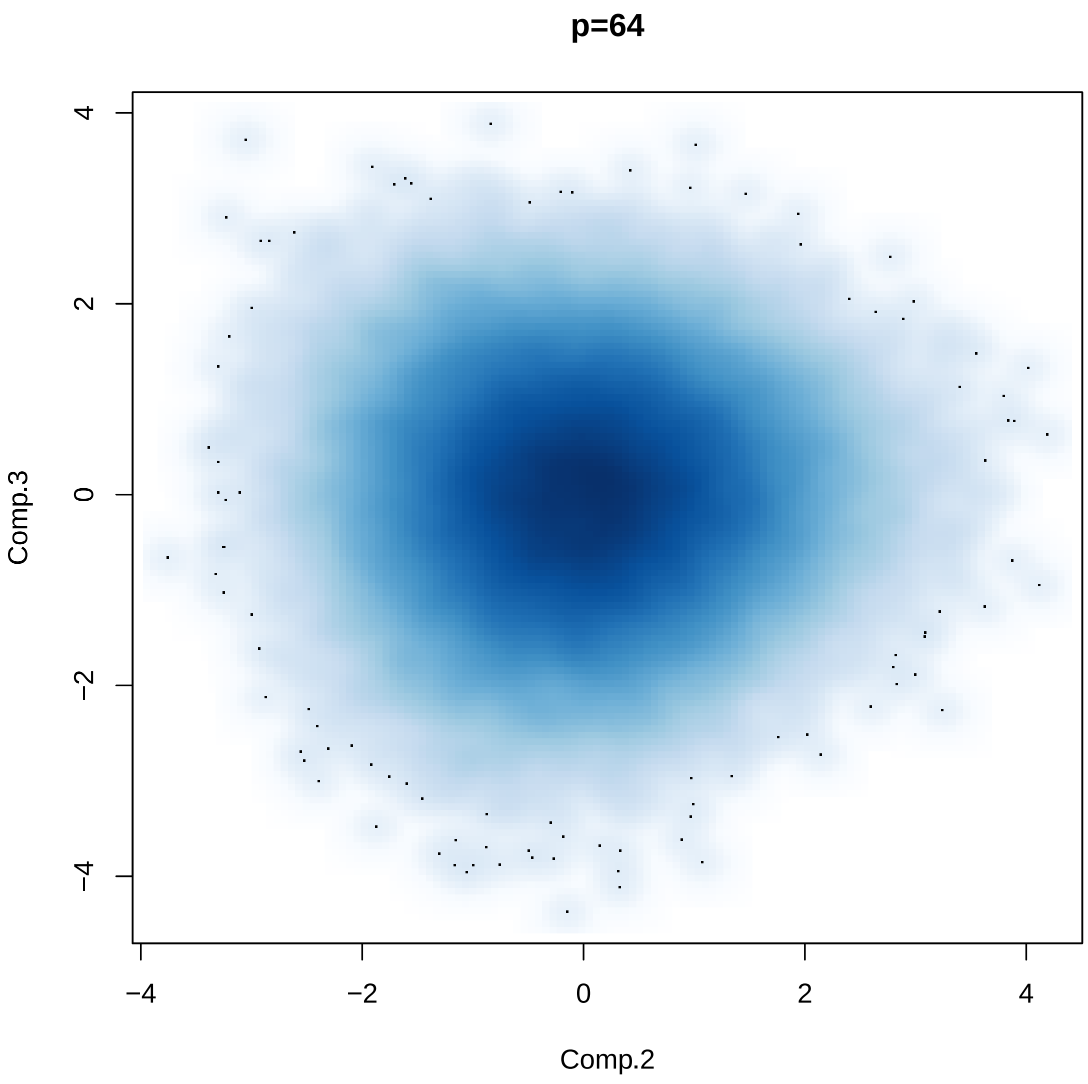

p=2
Comp.3
Comp.2
4
2
0
−2
−4
−4
−2
0
2

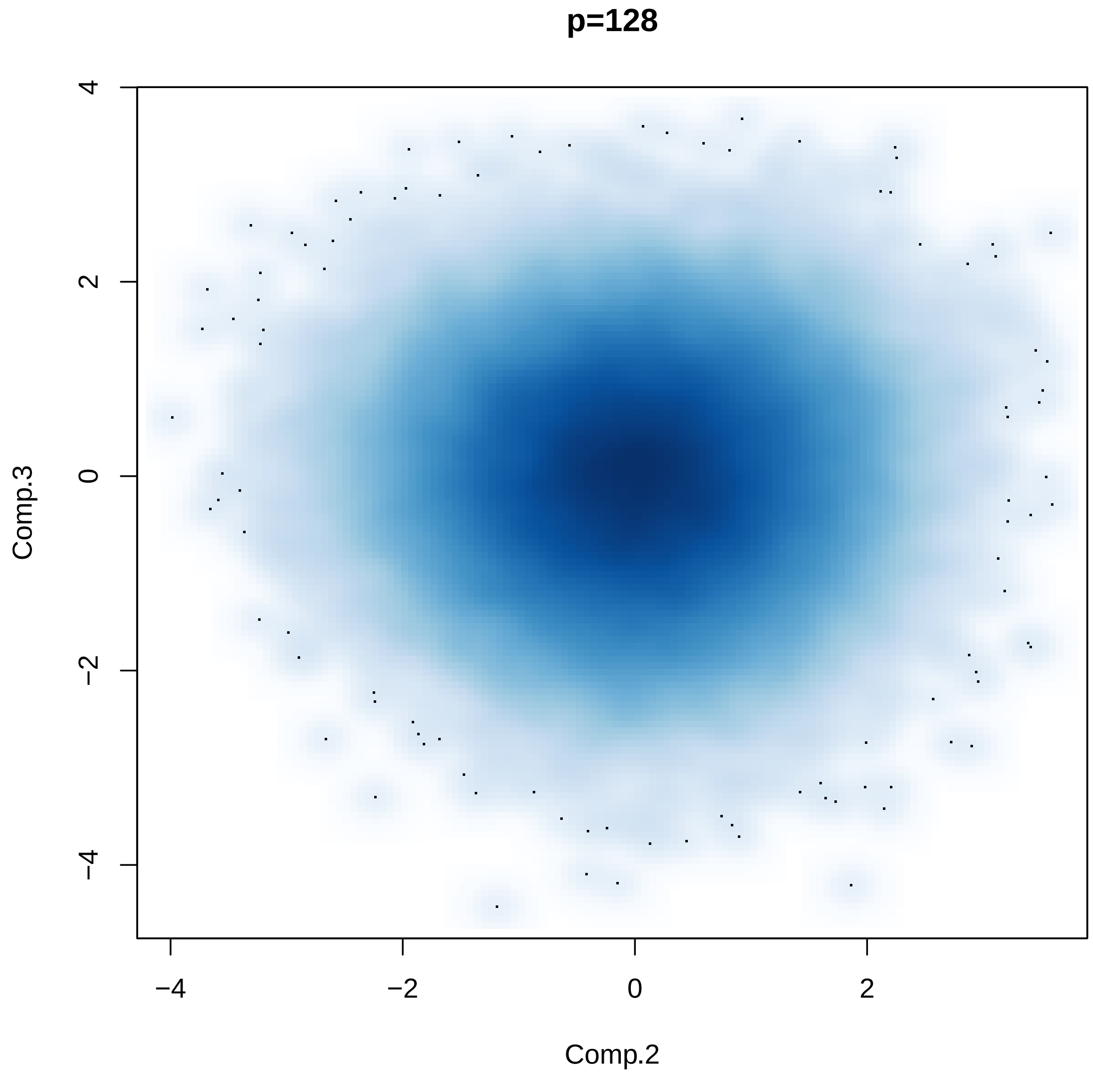

p=4
Comp.3
Comp.2
4
2
0
−2
−4
−2
0
2

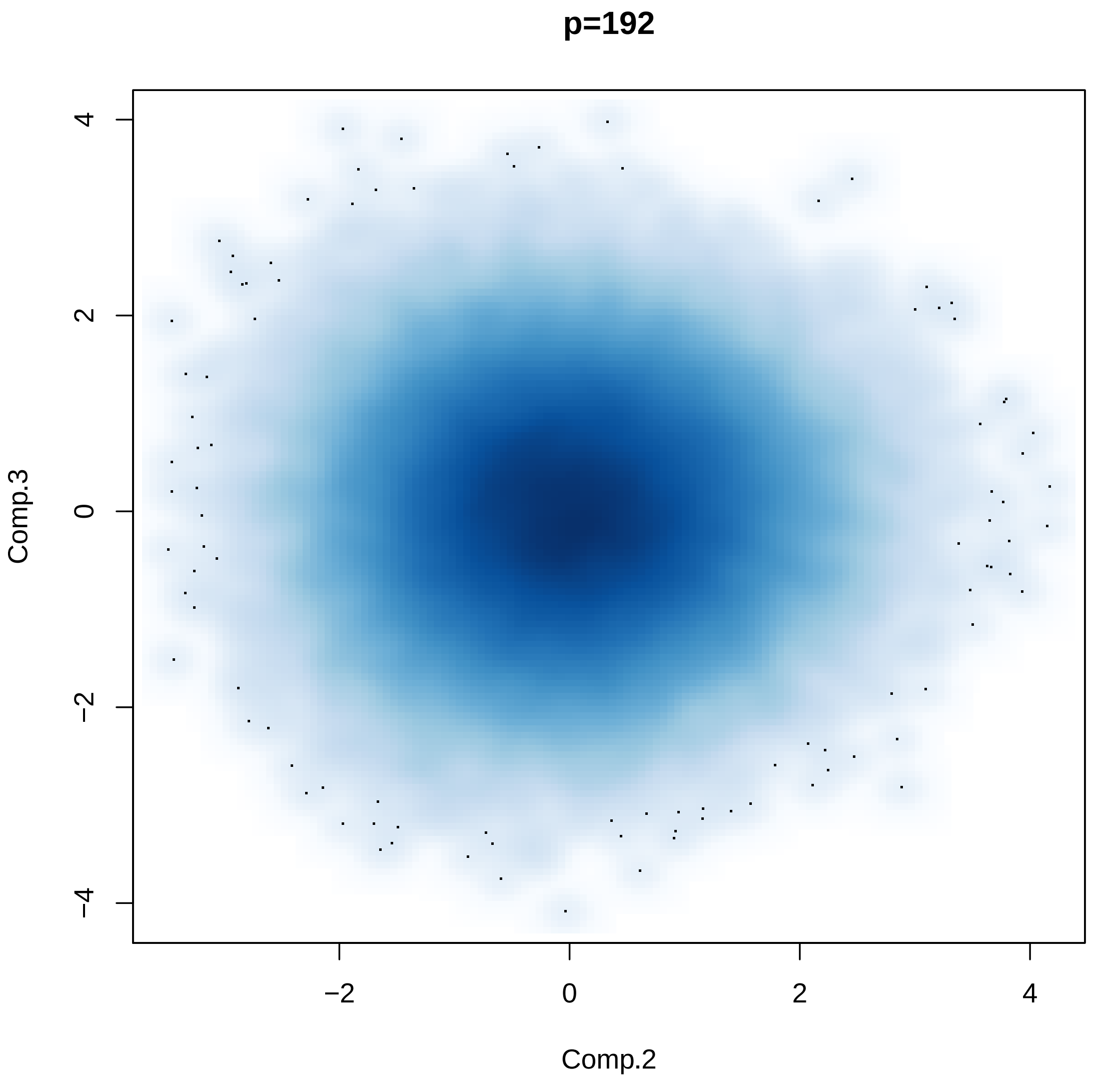

p=8
Comp.3
Comp.2
4
2
0
−2
−4
−4
−2
0
2

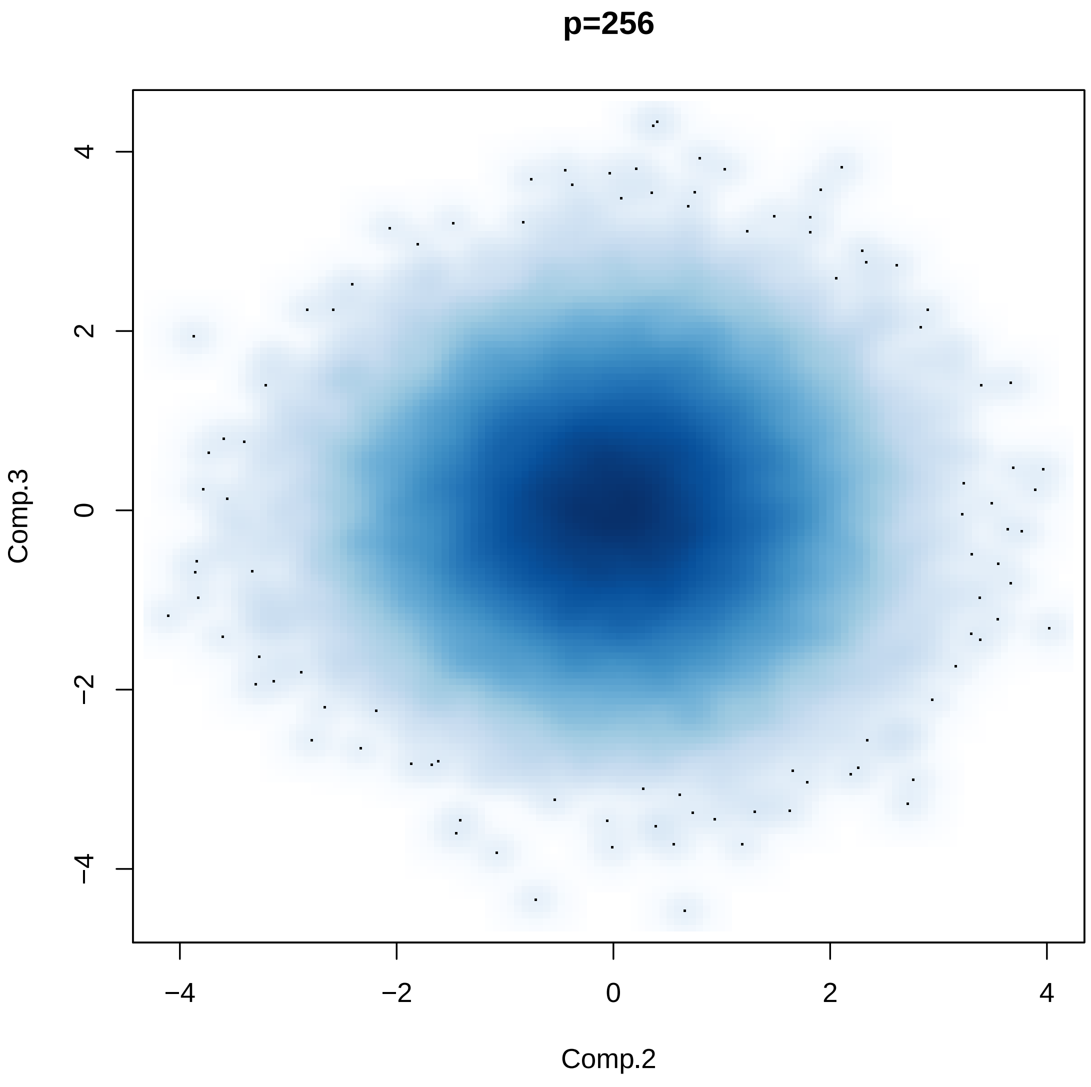

p=16
Comp.3
Comp.2
2
0
−2
−4
−4
−2
0
2
4


p=32
Comp.3
Comp.2
4
2
0
−2
−4
−2
0
2
4

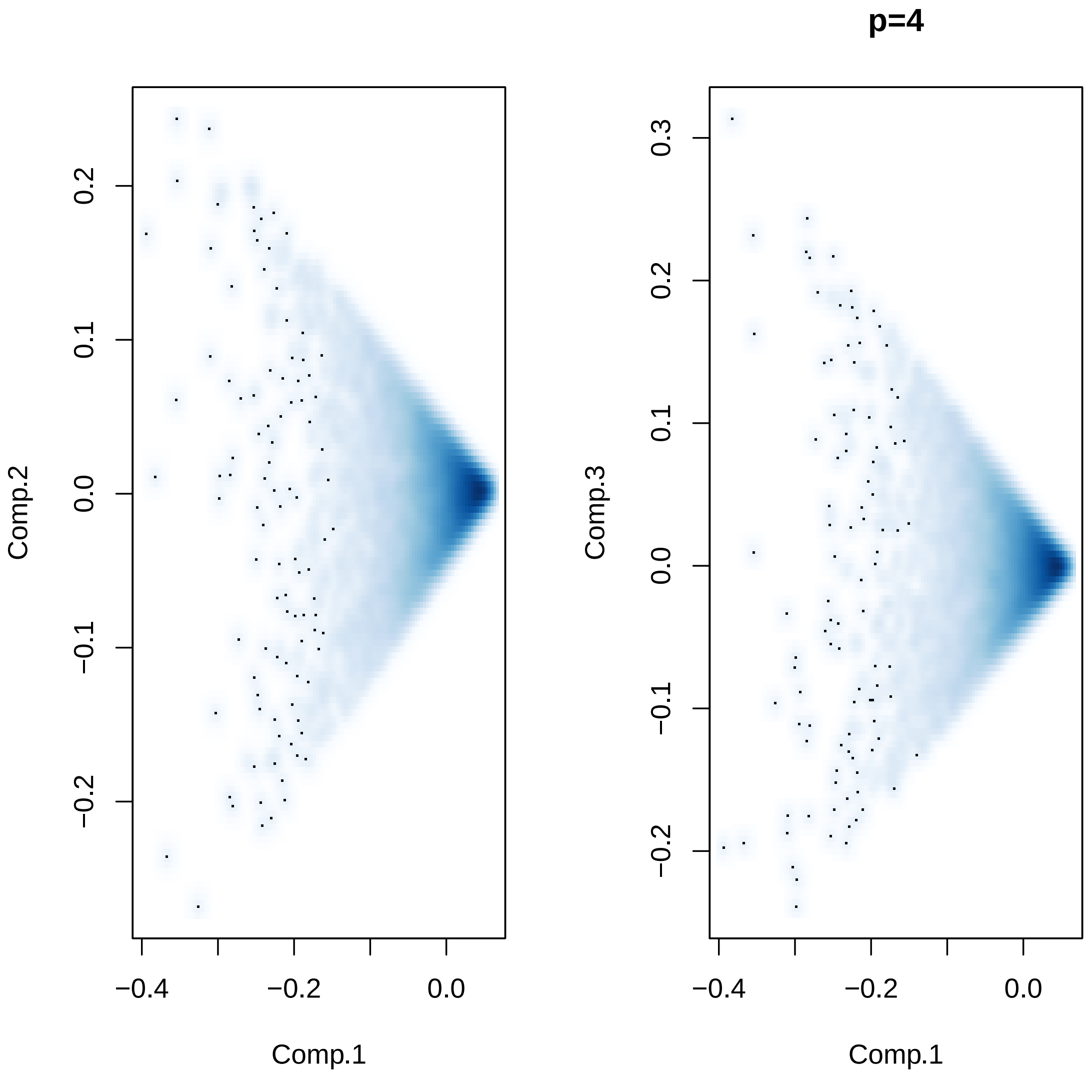

p=64
Comp.3
Comp.2
4
2
0
−2
−4
−4
−2
0
2
4

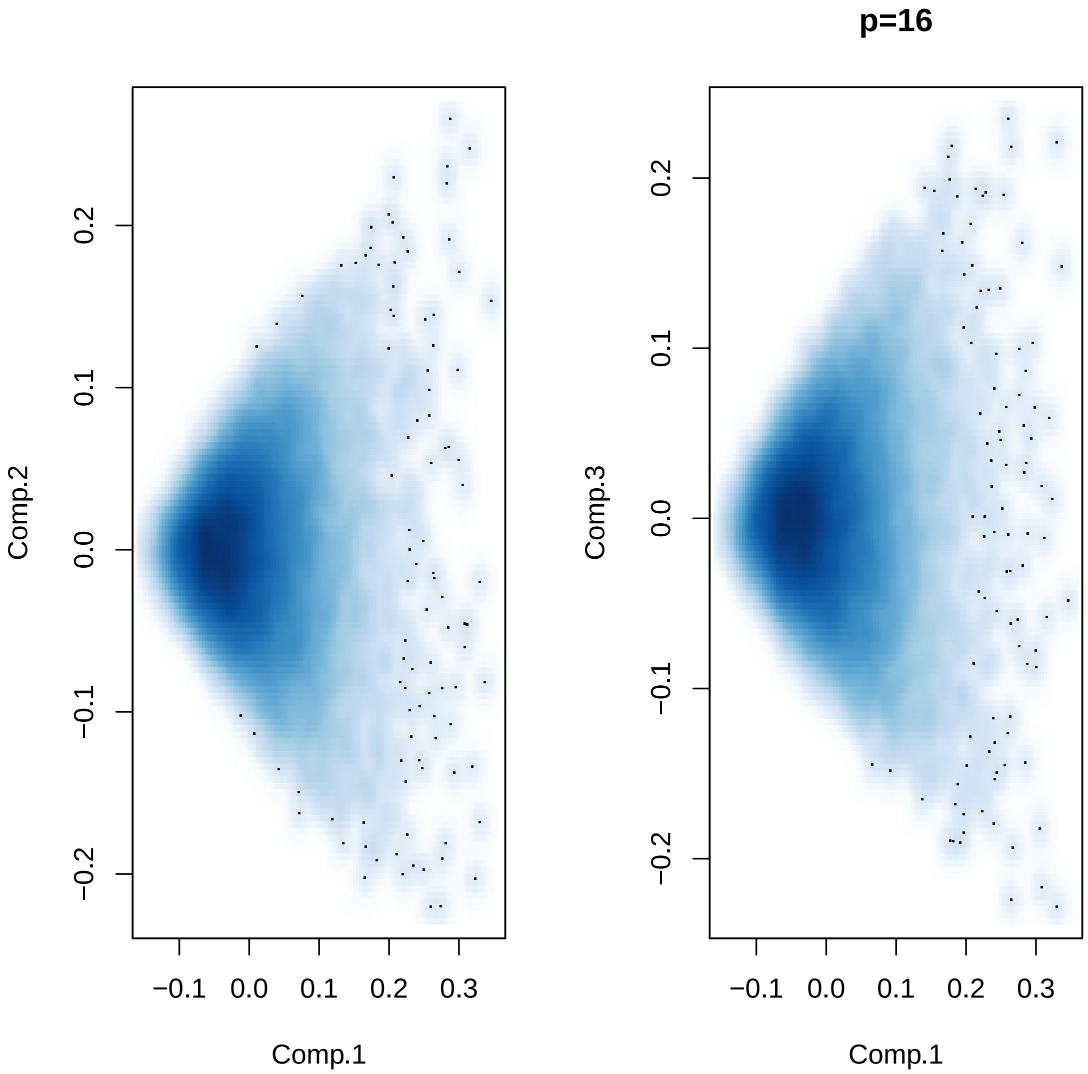

p=128
Comp.3
Comp.2
4
2
0
−2
−4
−4
−2
0
2

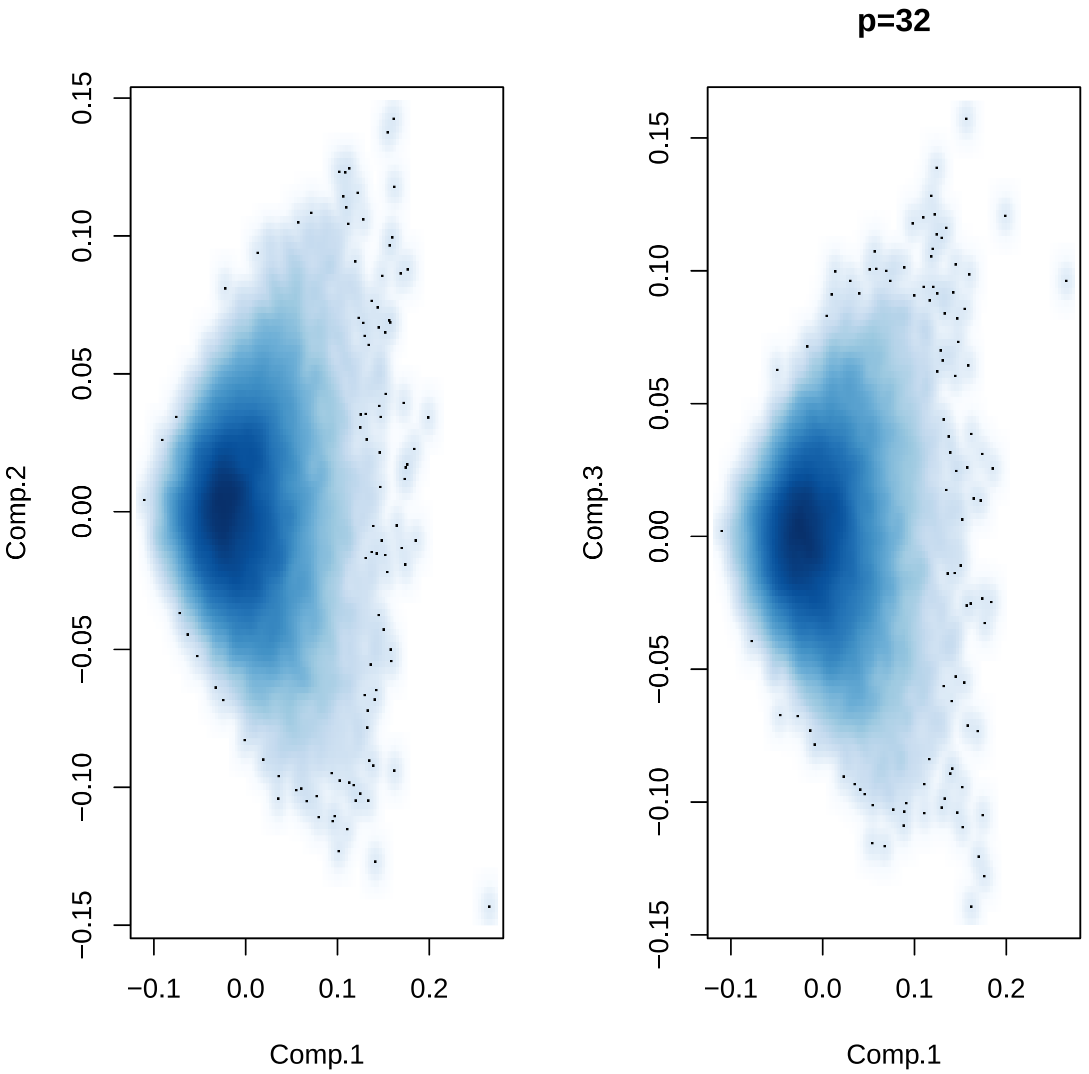

p=192
Comp.3
Comp.2
4
2
0
−2
−4
−2
0
2
4

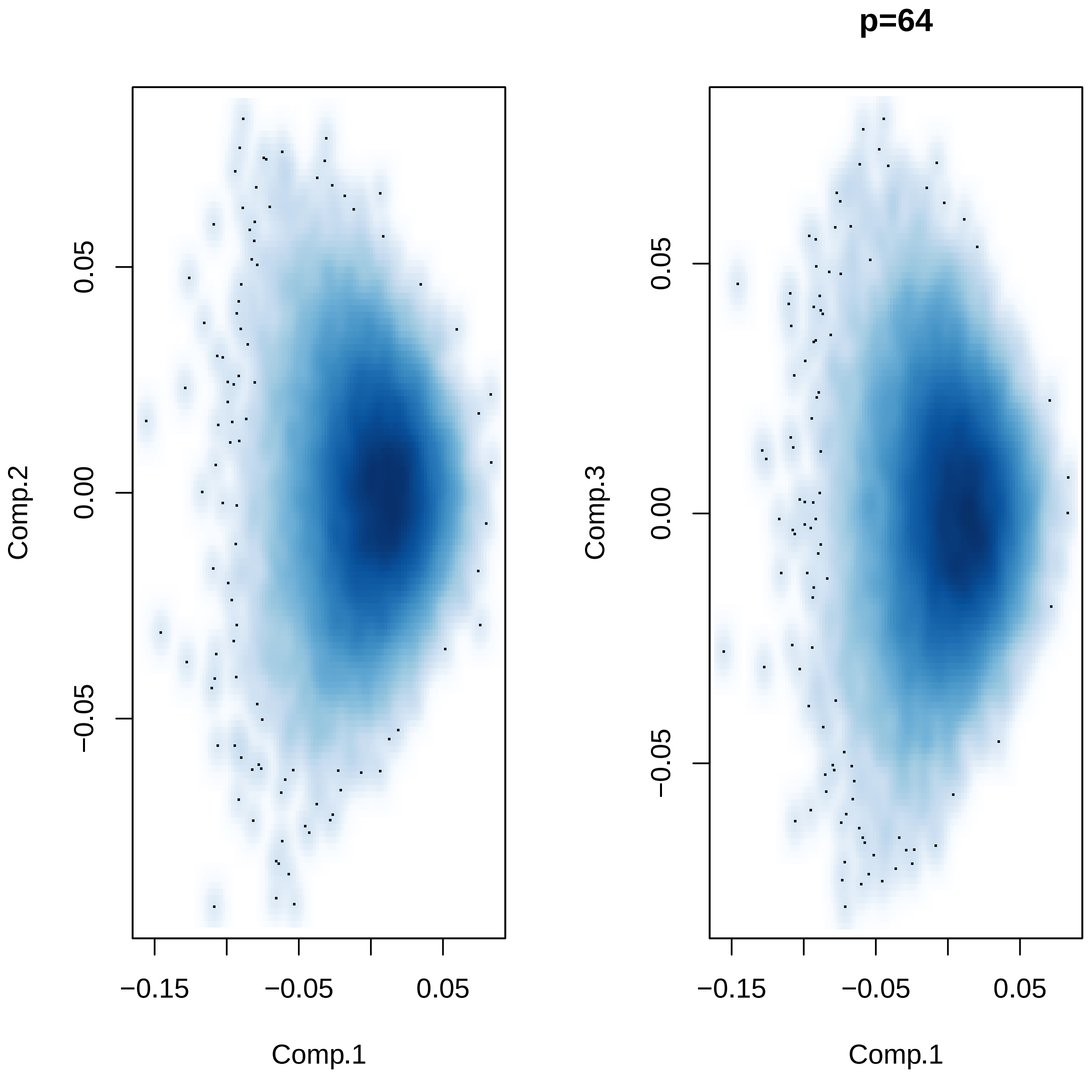

p=256
Comp.3
Comp.2
4
2
0
−2
−4
−4
−2
0
2
4

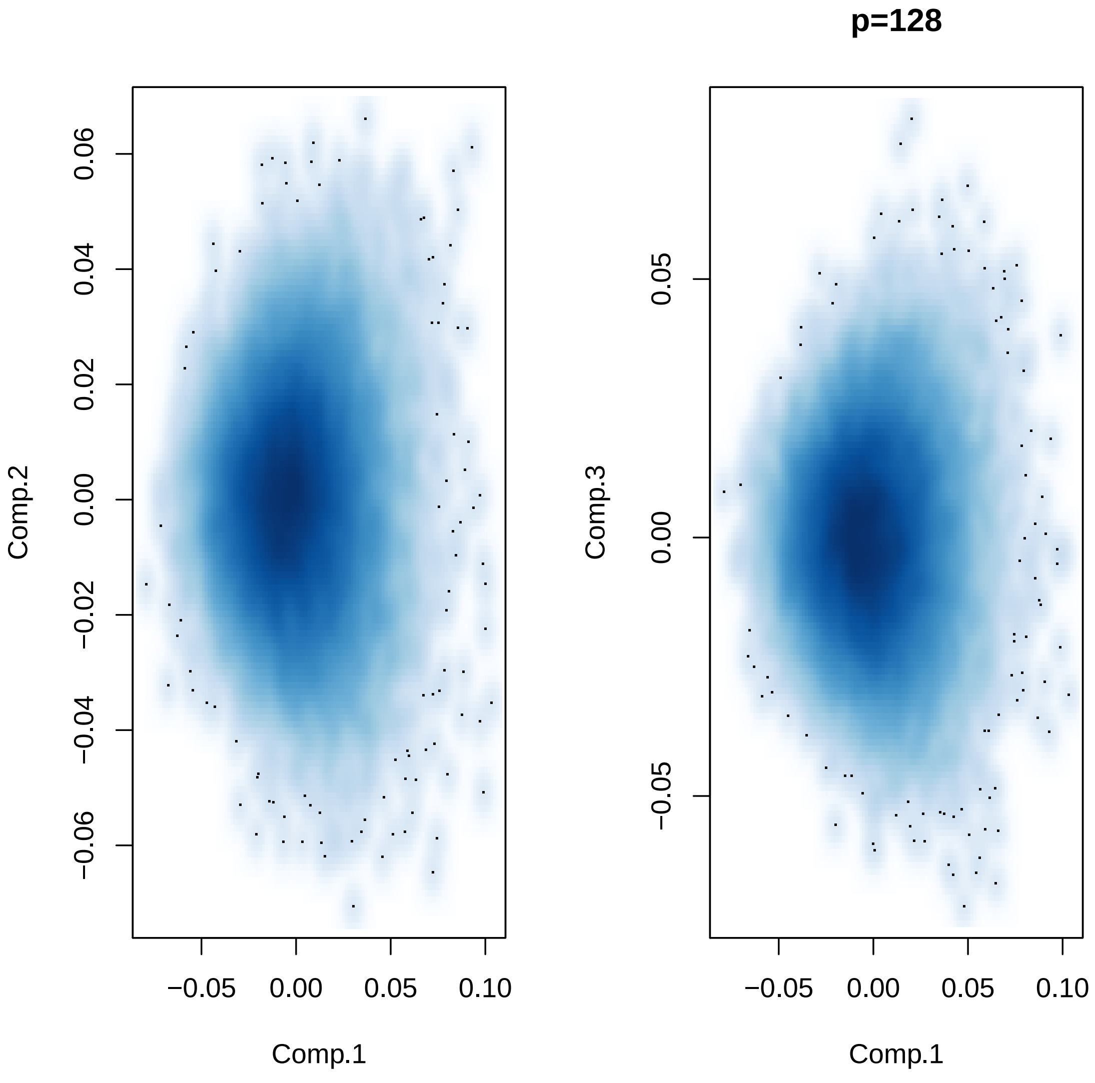
10.4.2 Fourier basis with cross correlation metric convergence plots

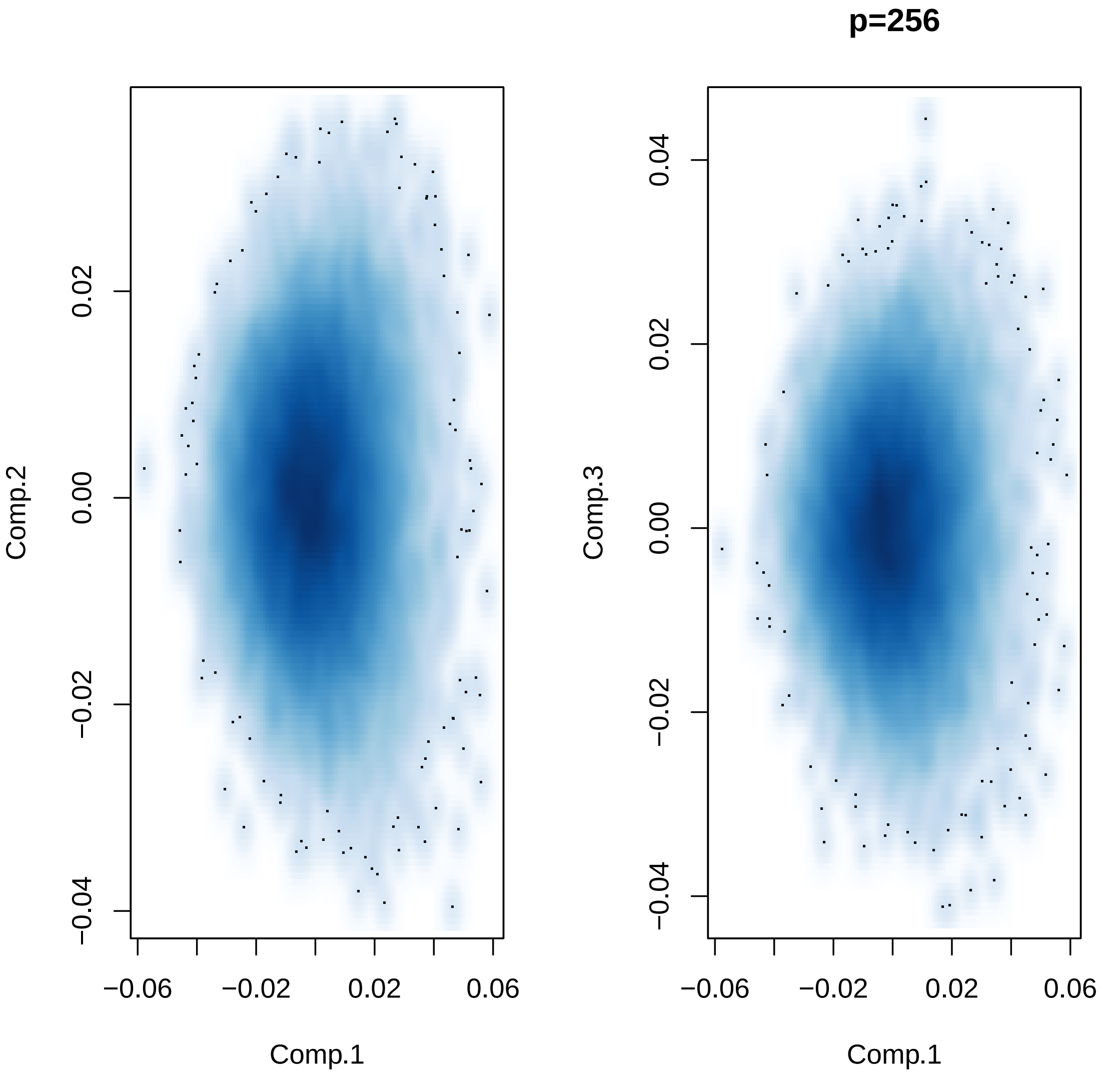

p=4
Comp.2
0.2
0.1
0.0
−0.1
−0.2
−0.4
−0.2
0.0
Comp.1
Comp.3
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.4
−0.2
0.0
Comp.1

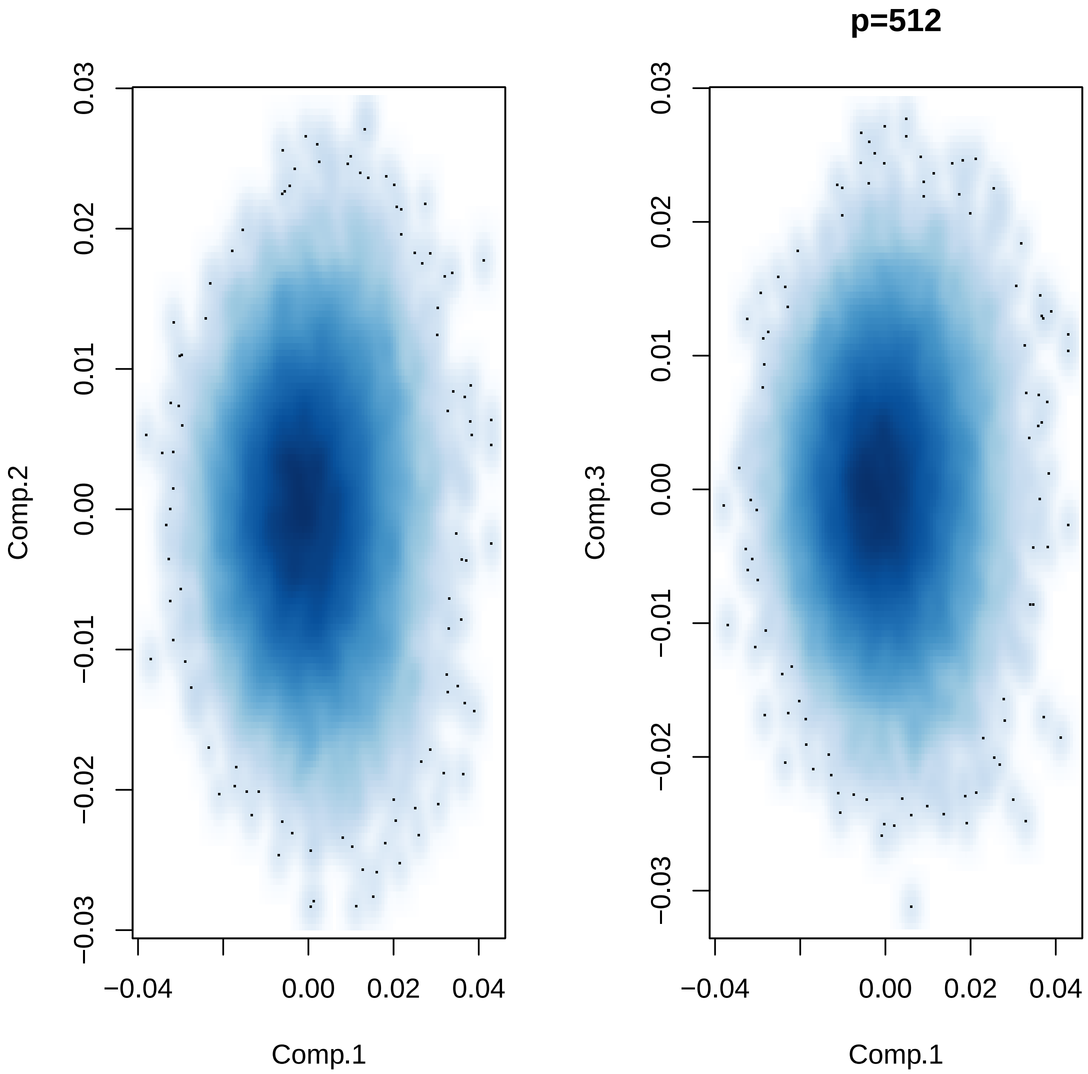

p=16
Comp.2
0.2
0.1
0.0
−0.1
−0.2
−0.1
0.0
0.1
0.2
0.3
Comp.1
Comp.3
0.2
0.1
0.0
−0.1
−0.2
−0.1
0.0
0.1
0.2
0.3
Comp.1


p=32
Comp.2
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.1
0.0
0.1
0.2
Comp.1
Comp.3
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.1
0.0
0.1
0.2
Comp.1

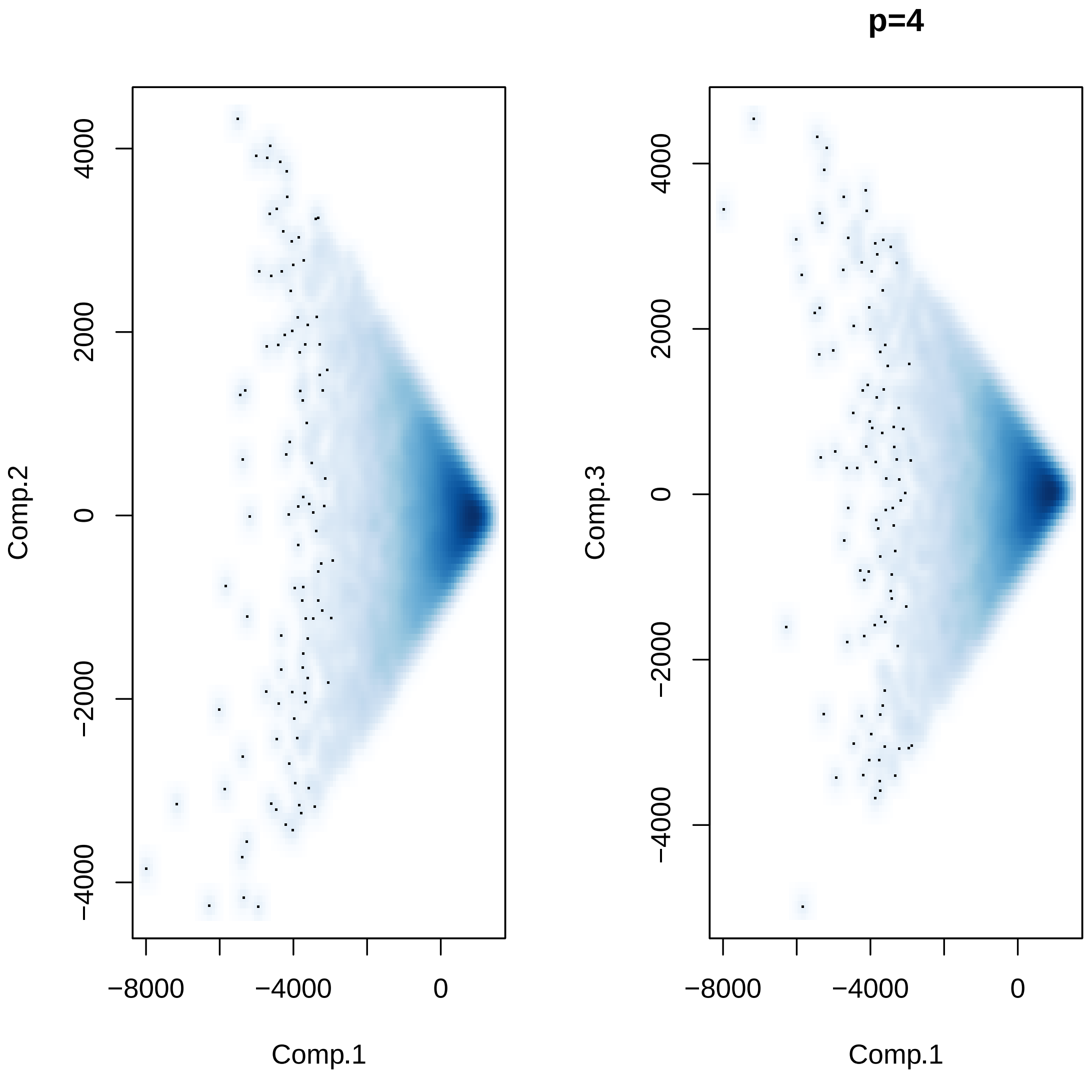

p=64
Comp.2
Comp.3
Comp.1
Comp.1
0.05
0.00
−0.05
0.05
0.00
−0.05
−0.15
−0.05
0.05
−0.15
−0.05
0.05

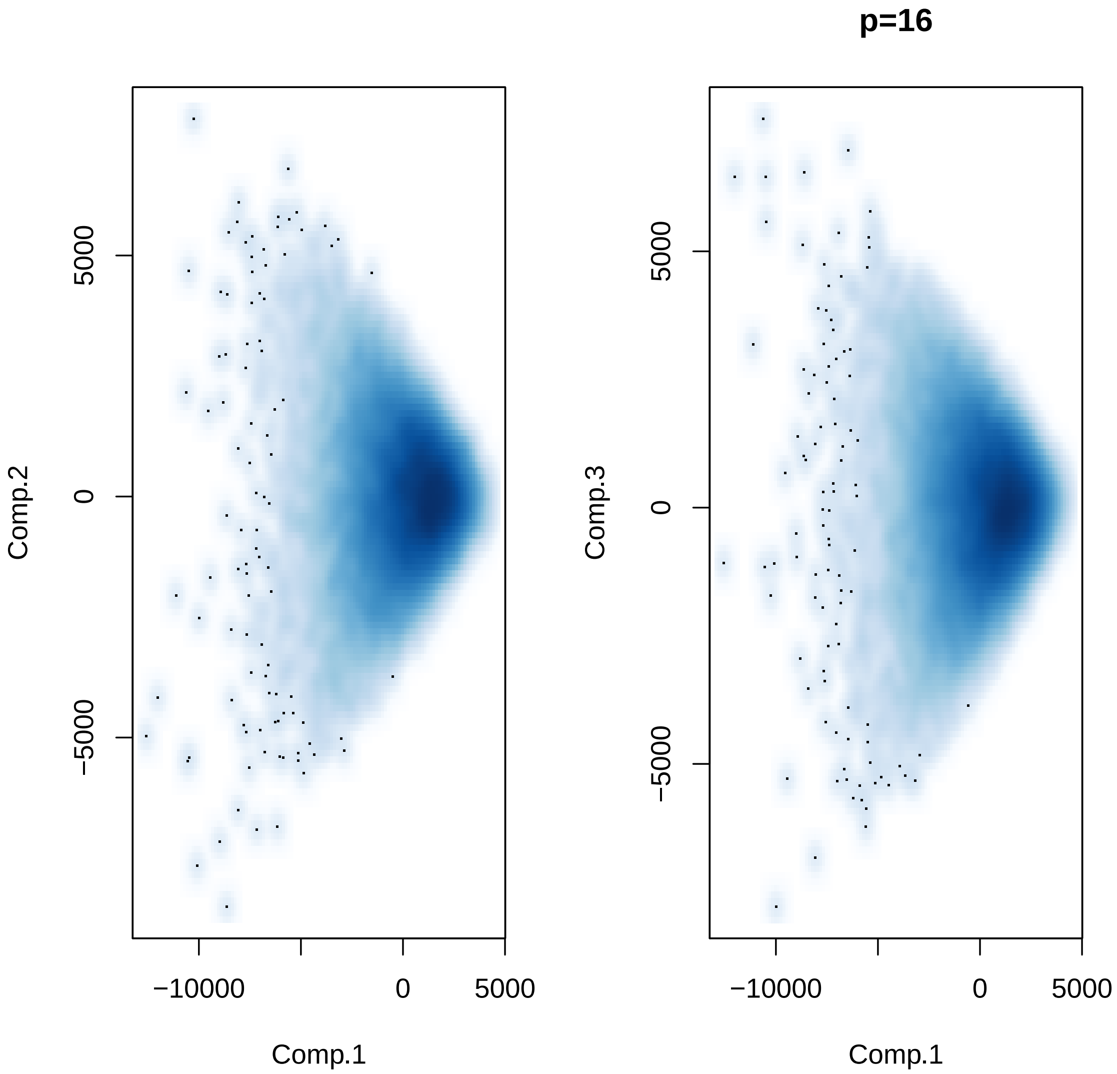

p=128
Comp.2
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.05
0.00
0.05
0.10
Comp.1
Comp.3
0.05
0.00
−0.05
−0.05
0.00
0.05
0.10
Comp.1

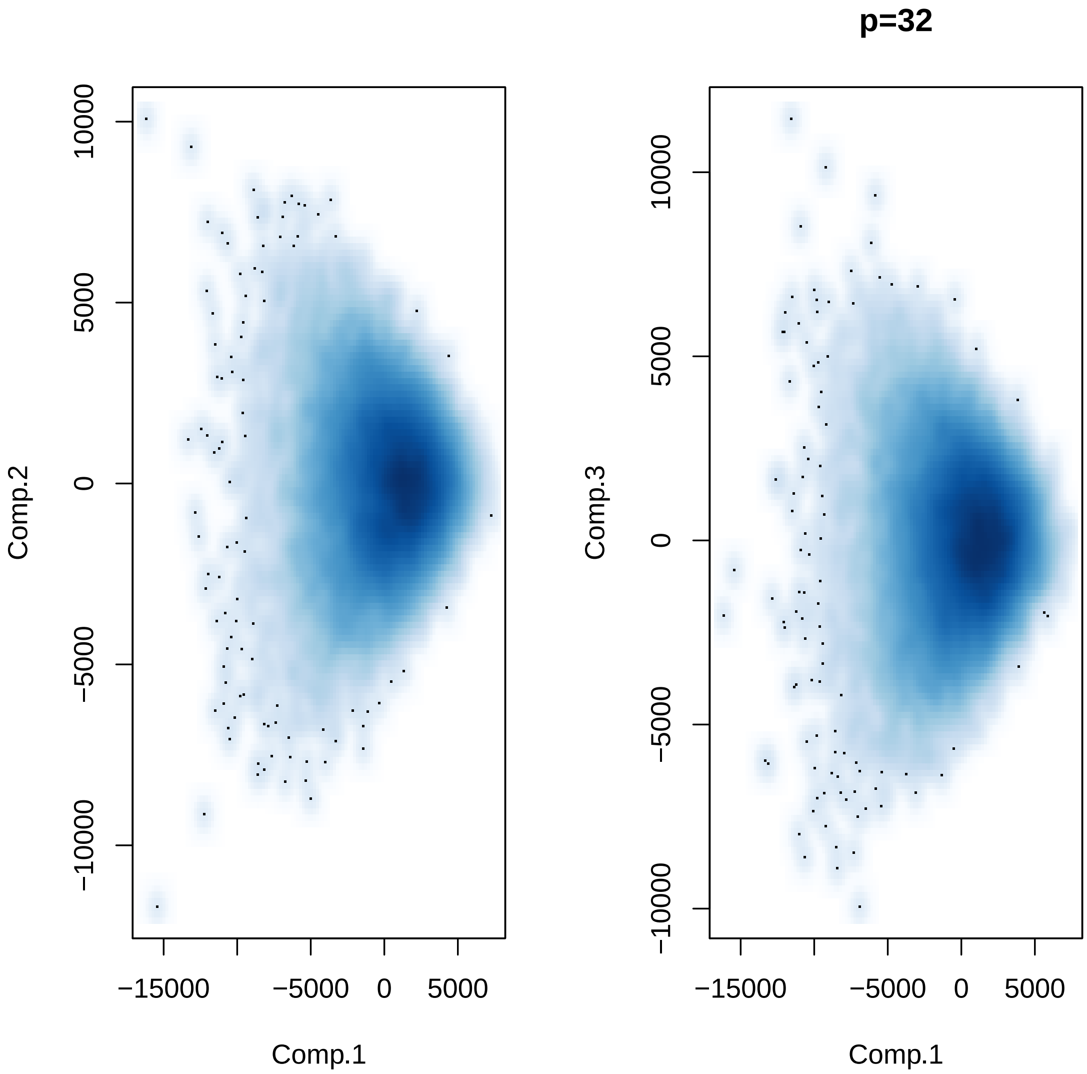

p=256
Comp.2
0.02
0.00
−0.02
−0.04
−0.06
−0.02
0.02
0.06
Comp.1
Comp.3
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.02
0.02
0.06
Comp.1

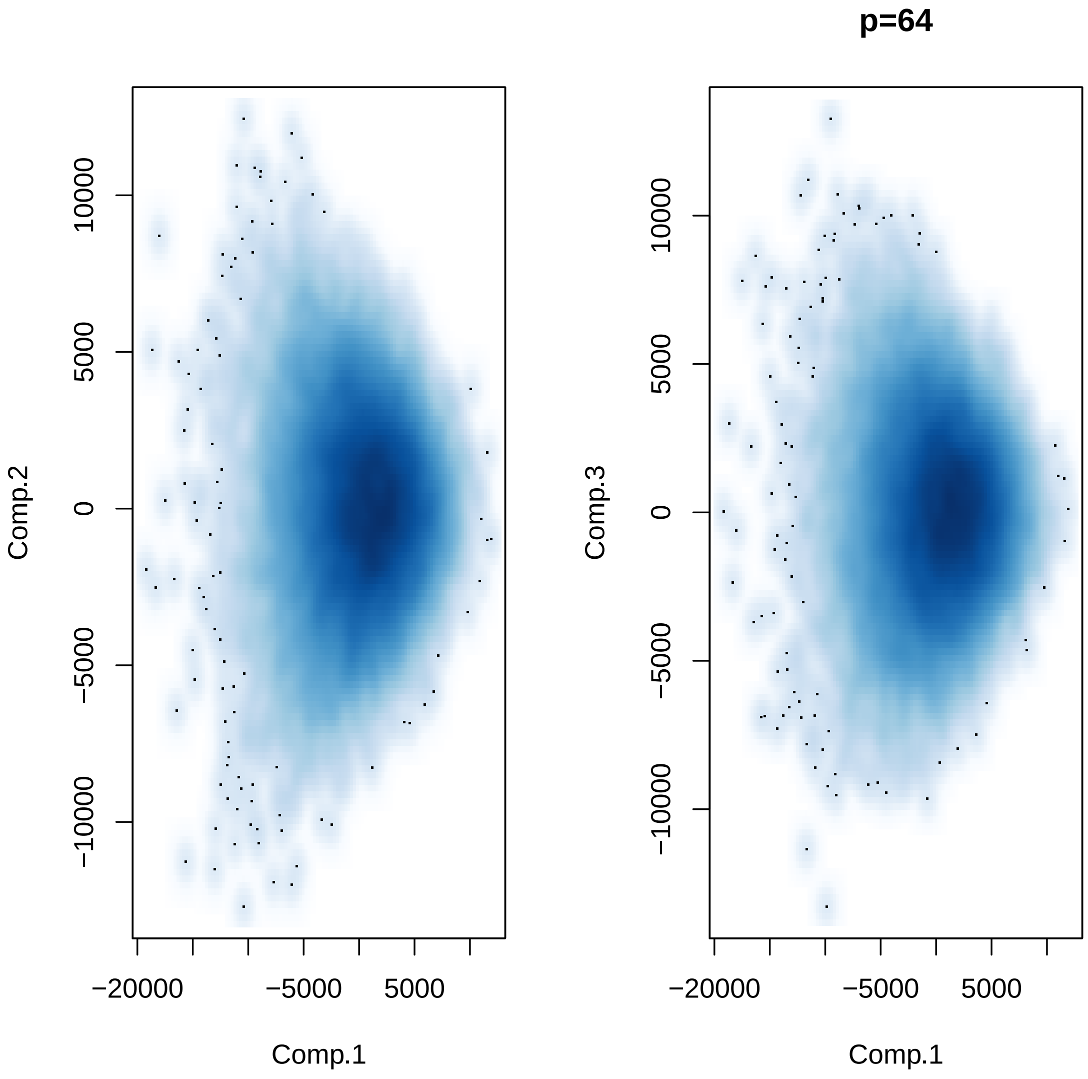

p=512
Comp.2
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
0.00
0.02
0.04
Comp.1
Comp.3
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
0.00
0.02
0.04
Comp.1

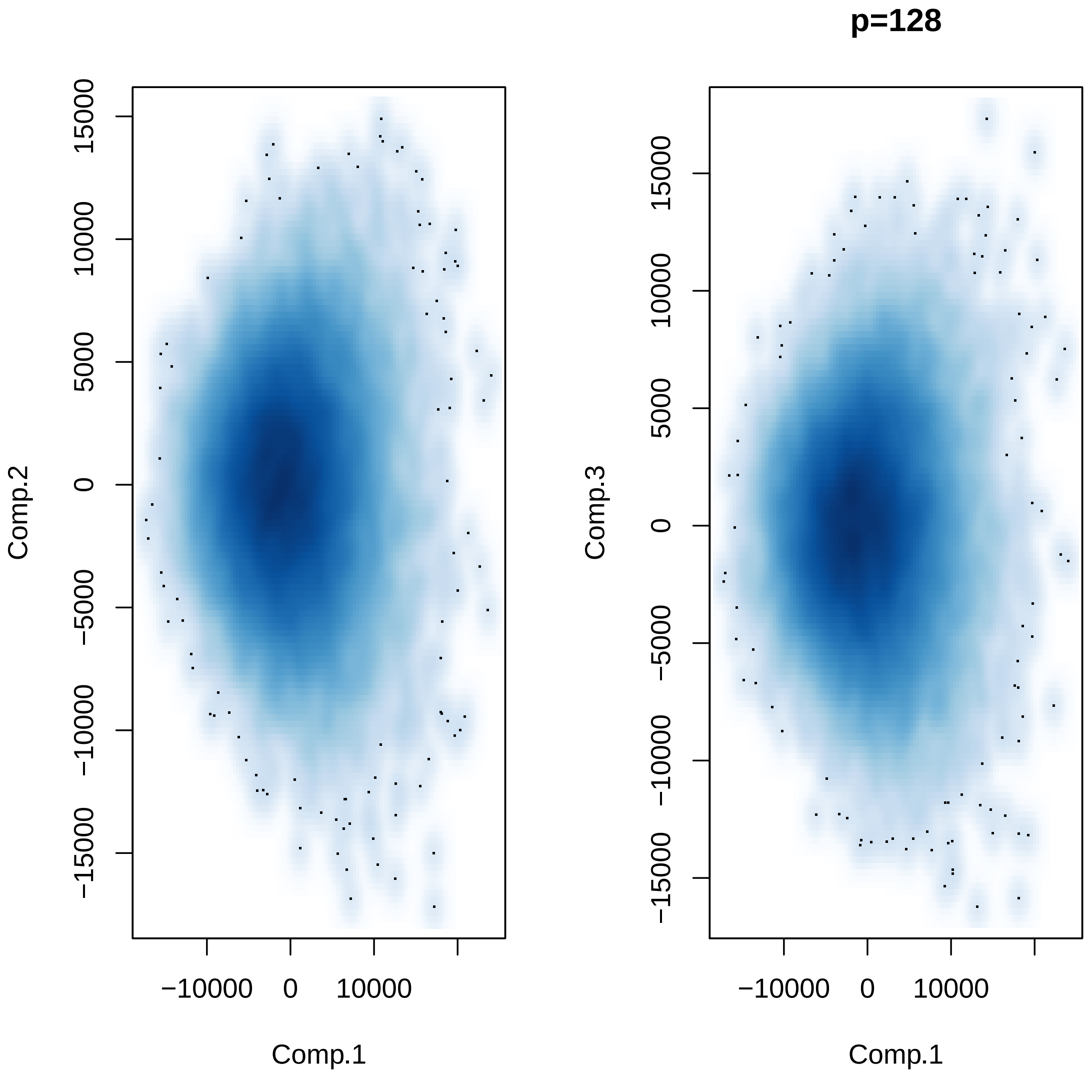
10.4.3 Fourier basis Euclidean metric convergence plots

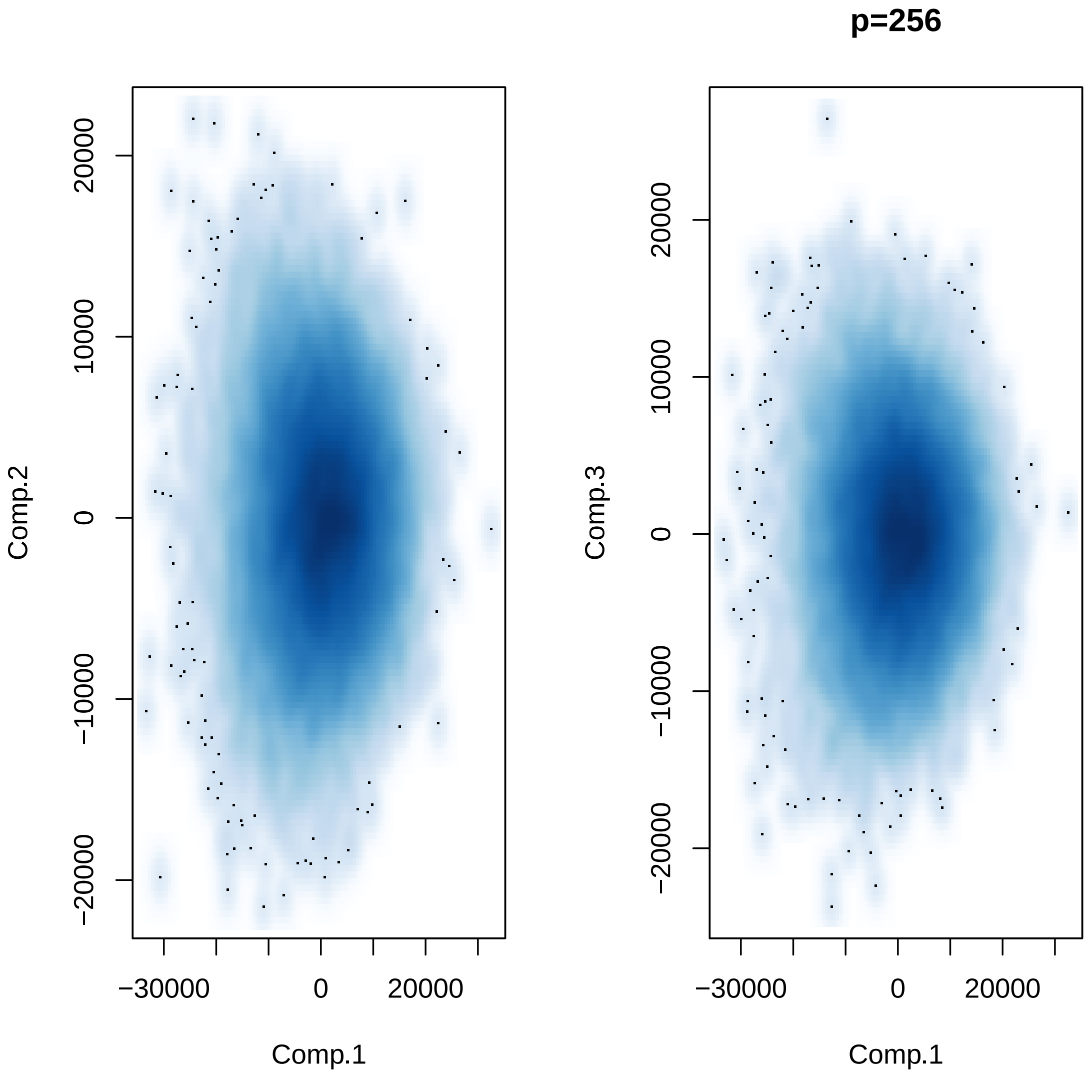

p=4
Comp.2
4000
2000
0
−2000
−4000
−8000
−4000
0
Comp.1
Comp.3
4000
2000
0
−2000
−4000
−8000
−4000
0
Comp.1

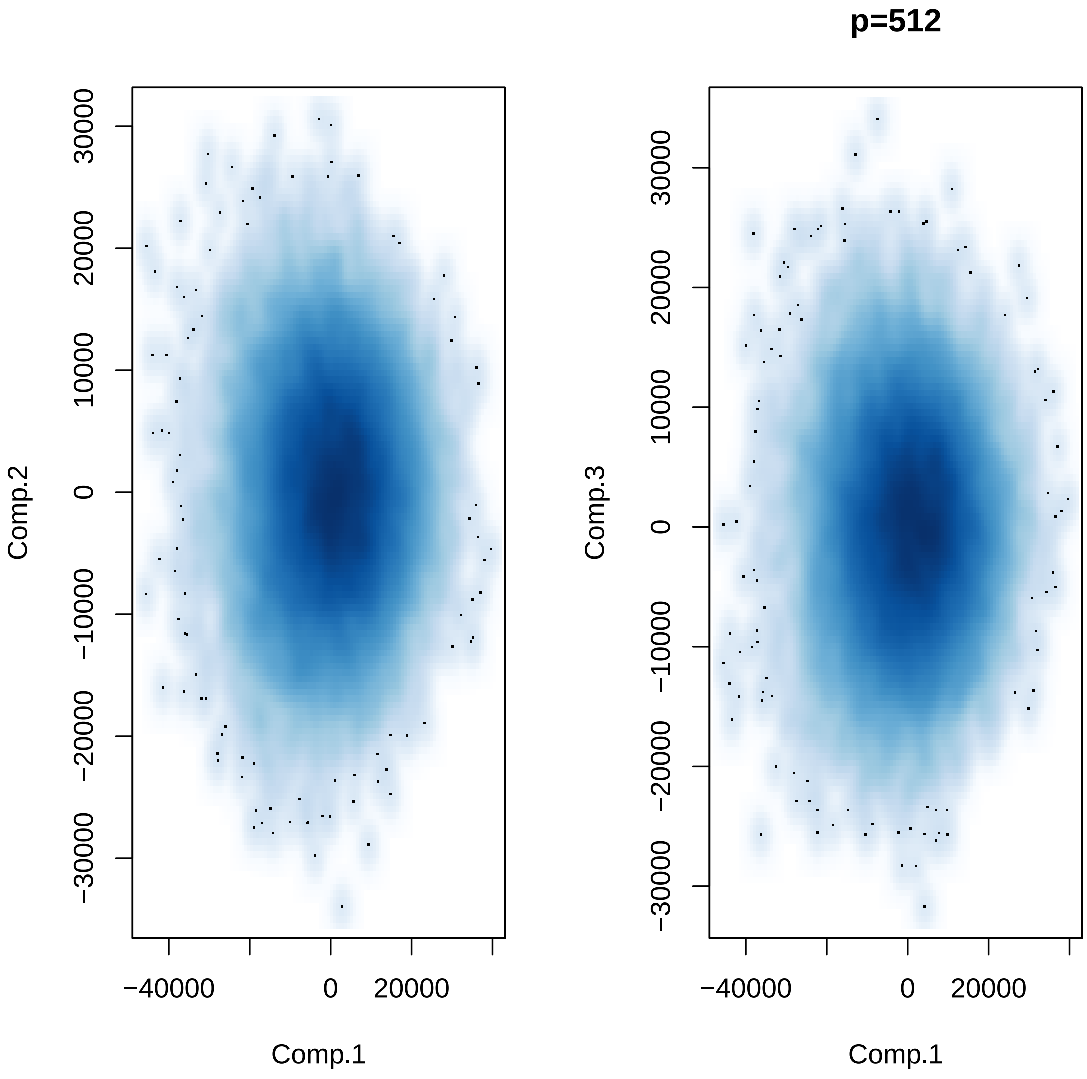

p=16
Comp.2
Comp.3
Comp.1
Comp.1
5000
0
−5000
5000
0
−5000
−10000
0
5000
−10000
0
5000


p=32
Comp.2
10000
5000
0
−5000
−10000
−15000
−5000
0
5000
Comp.1
Comp.3
10000
5000
0
−5000
−10000
−15000
−5000
0
5000
Comp.1

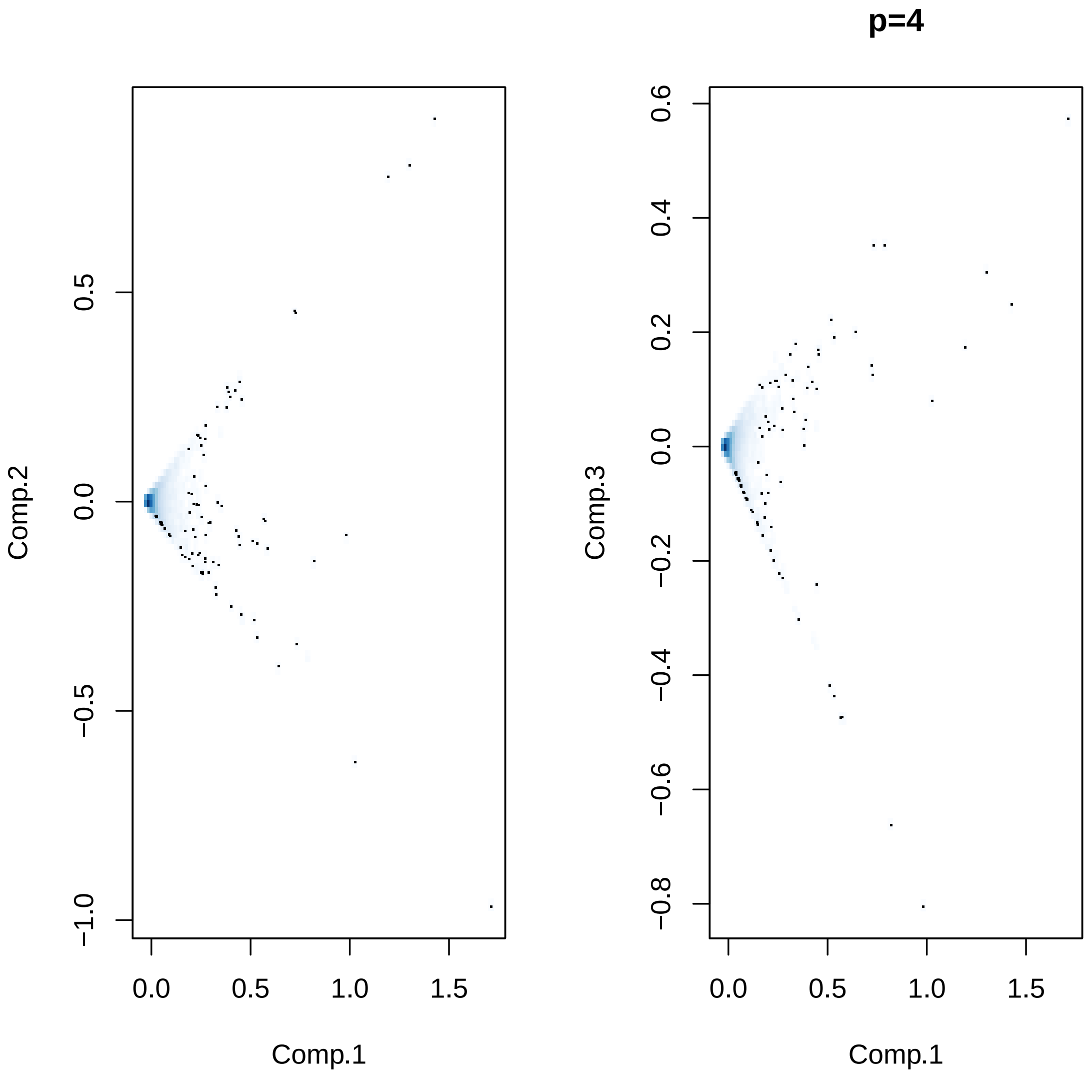

p=64
Comp.2
10000
5000
0
−5000
−10000
−20000
−5000
5000
Comp.1
Comp.3
10000
5000
0
−5000
−10000
−20000
−5000
5000
Comp.1

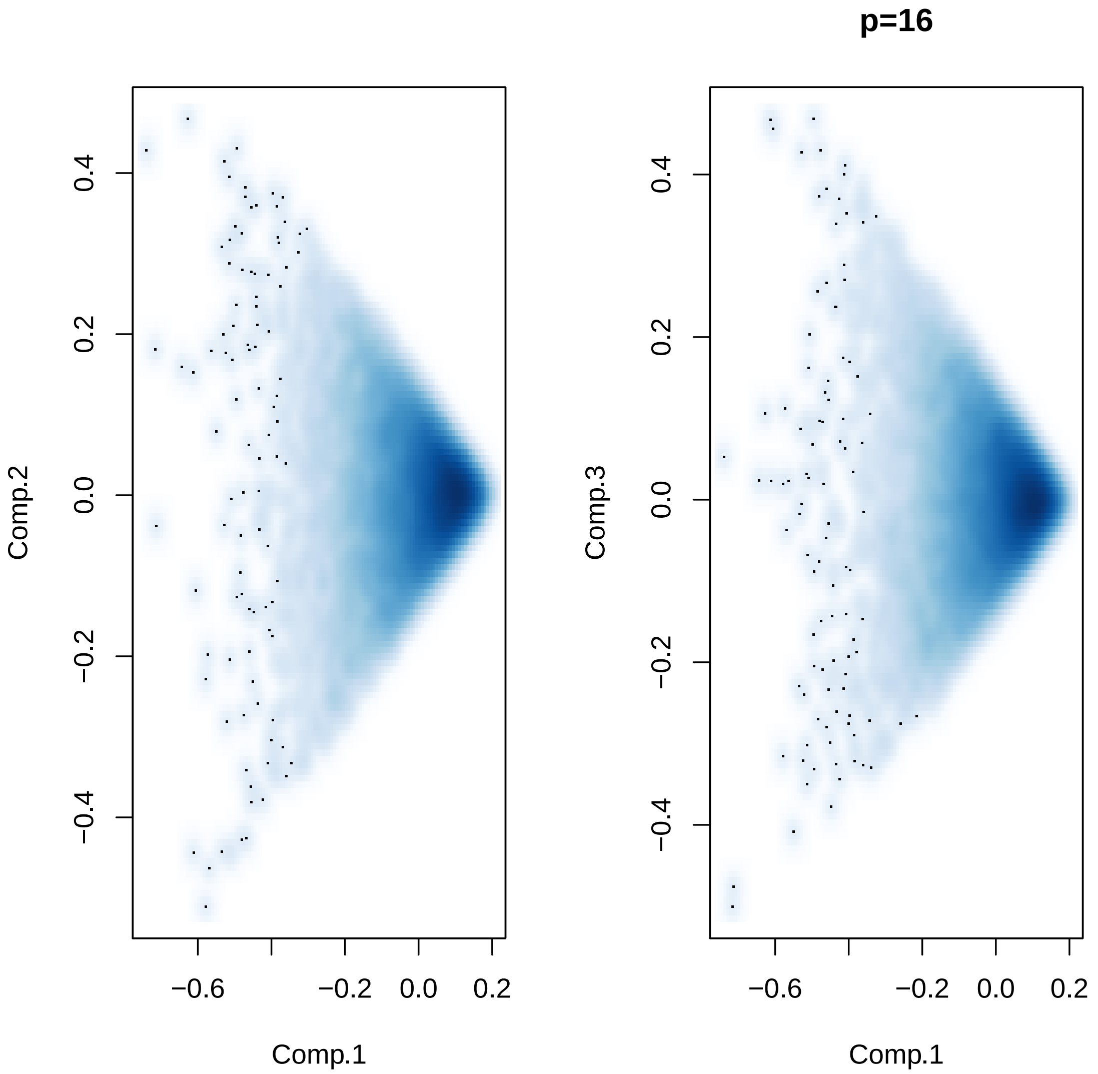

p=128
Comp.2
15000
10000
5000
0
−5000
−10000
−15000
−10000
0
10000
Comp.1
Comp.3
15000
10000
5000
0
−5000
−10000
−15000
−10000
0
10000
Comp.1

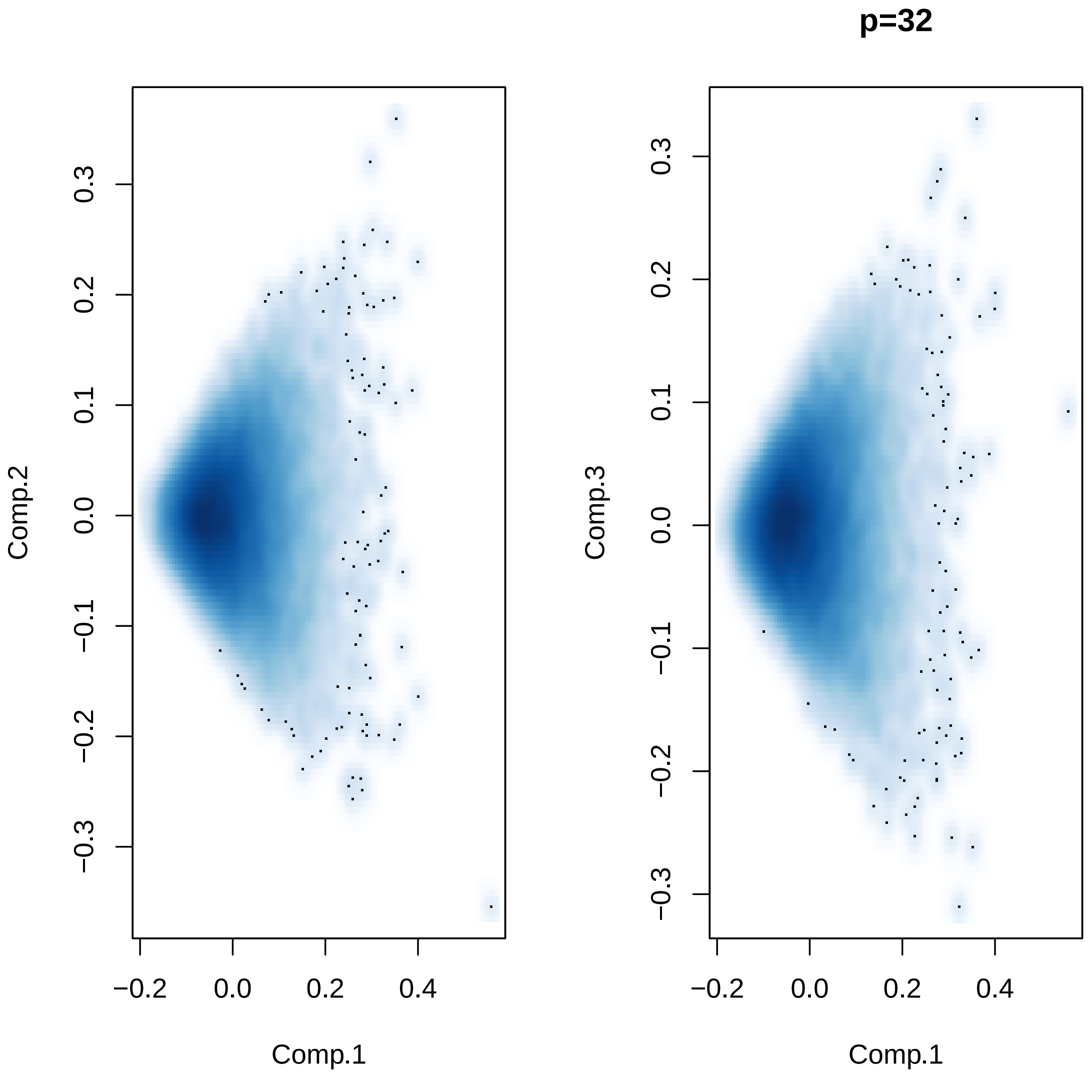

p=256
Comp.2
20000
10000
0
−10000
−20000
−30000
0
20000
Comp.1
Comp.3
20000
10000
0
−10000
−20000
−30000
0
20000
Comp.1

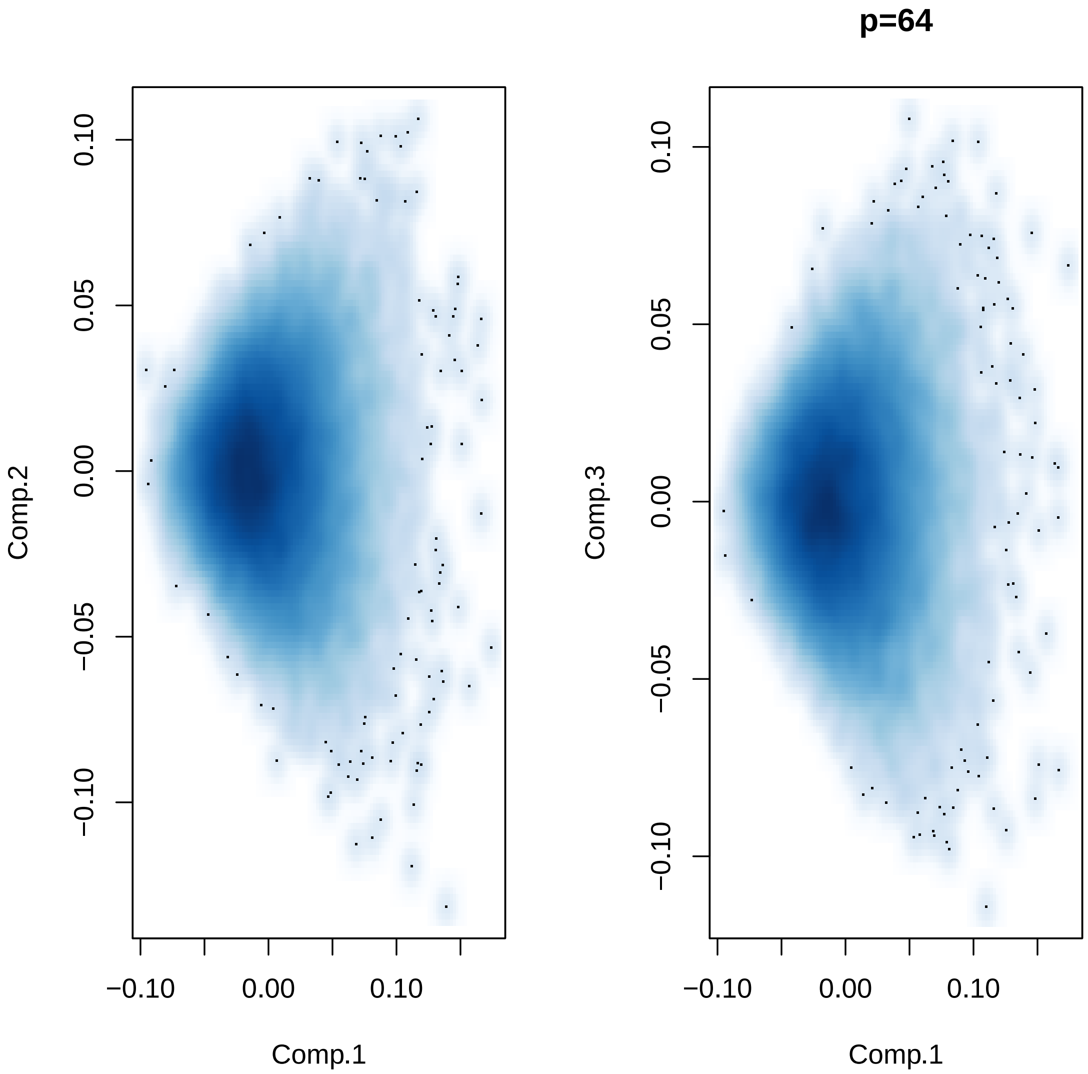

p=512
Comp.2
30000
20000
10000
0
−10000
−20000
−30000
−40000
0
20000
Comp.1
Comp.3
30000
20000
10000
0
−10000
−20000
−30000
−40000
0
20000
Comp.1

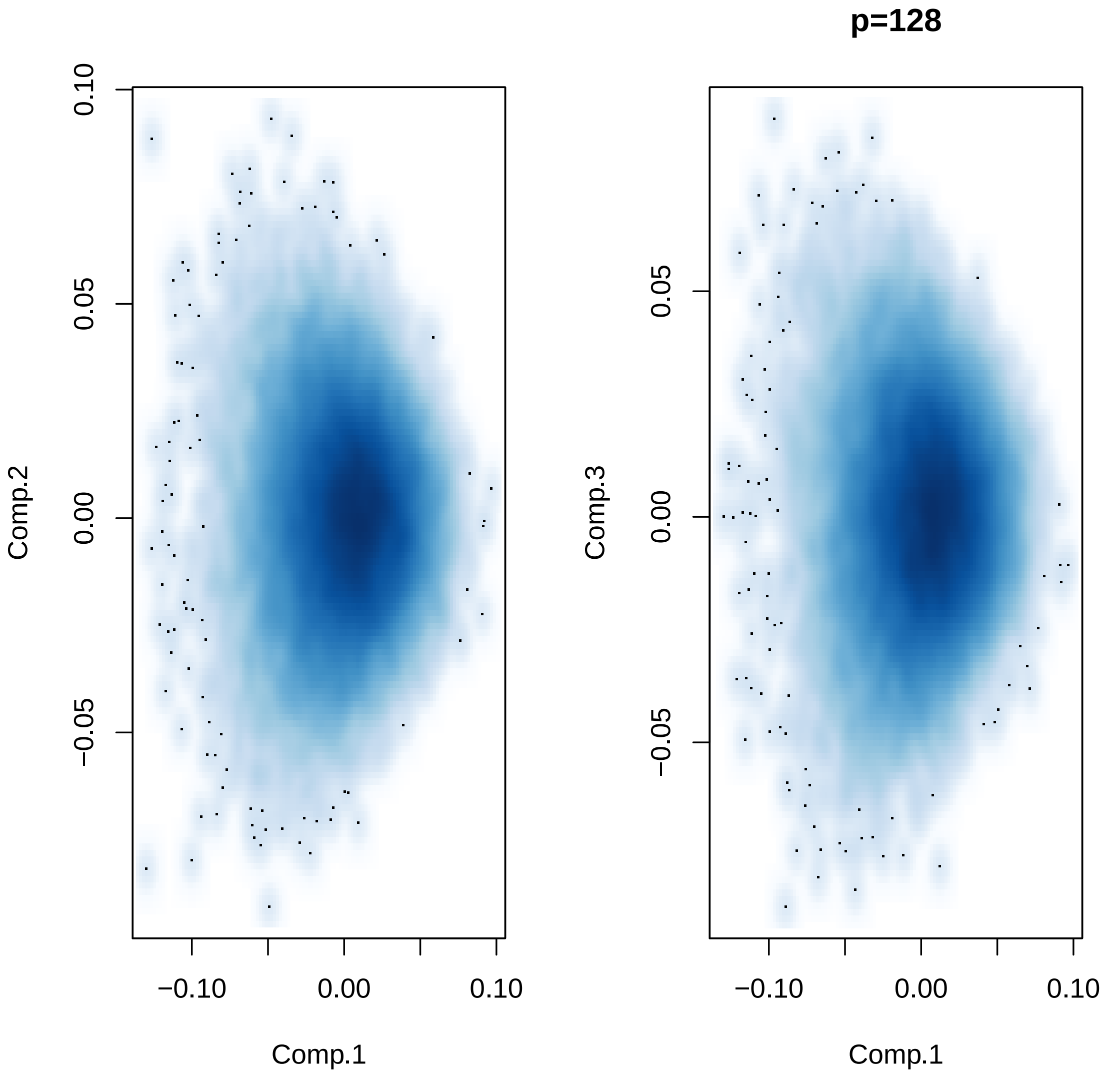
10.4.4 B-spline basis with cross correlation metric convergence plots

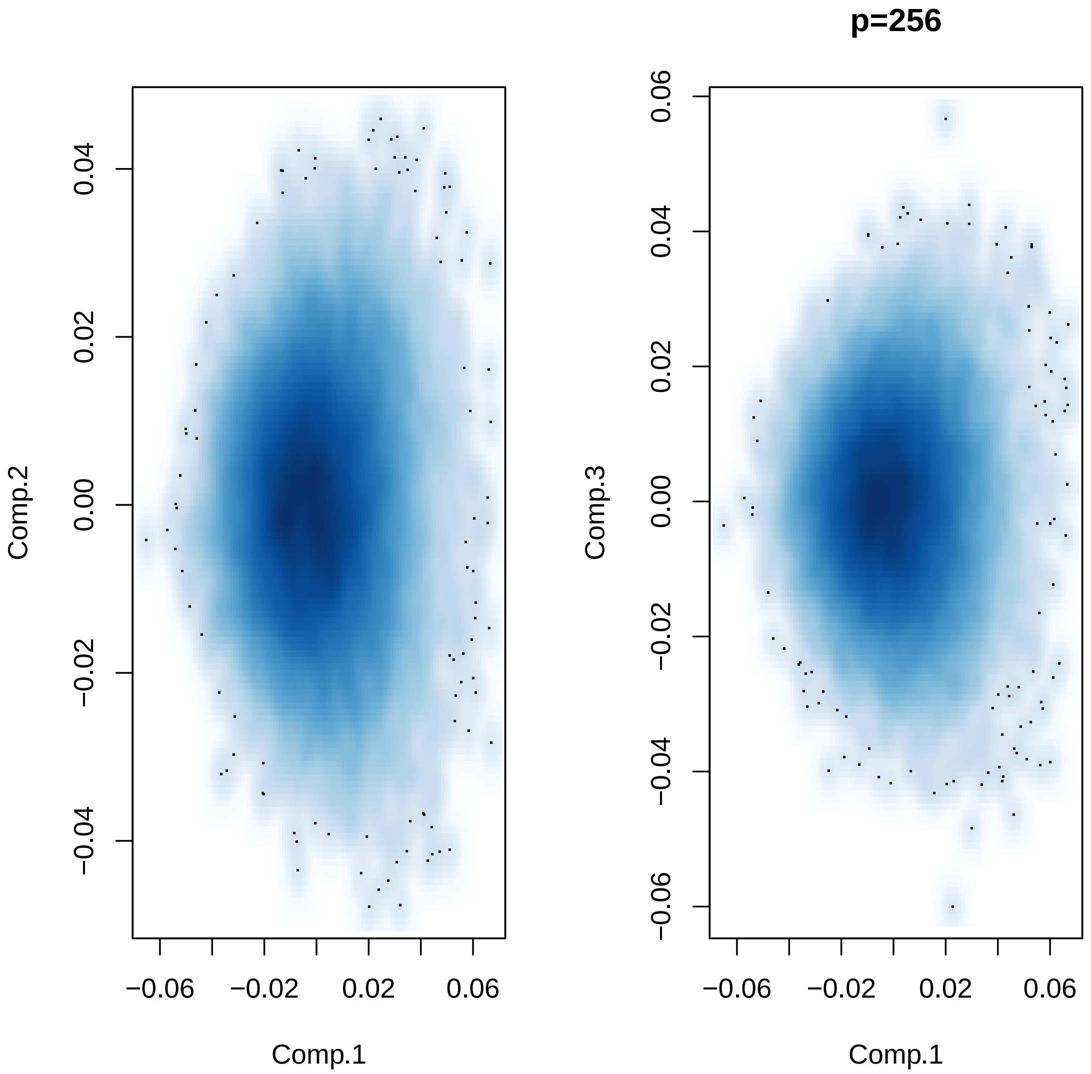

p=4
Comp.2
Comp.1
0.5
0.0
−0.5
−1.0
0.0
0.5
1.0
1.5
Comp.3
Comp.1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0.0
0.5
1.0
1.5

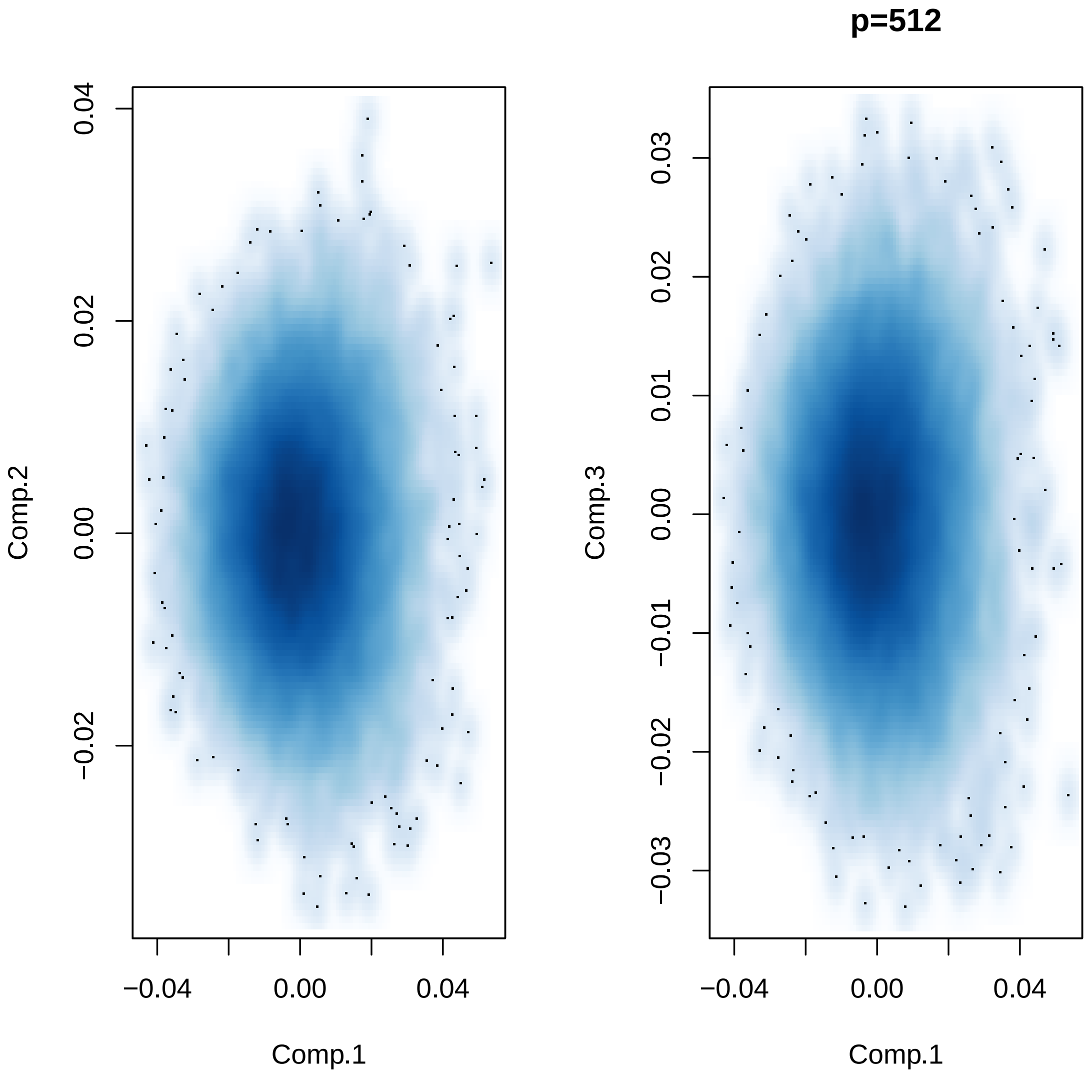

p=16
Comp.2
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.2
0.0
0.2
Comp.1
Comp.3
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.2
0.0
0.2
Comp.1


p=32
Comp.2
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
0.0
0.2
0.4
Comp.1
Comp.3
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
0.0
0.2
0.4
Comp.1

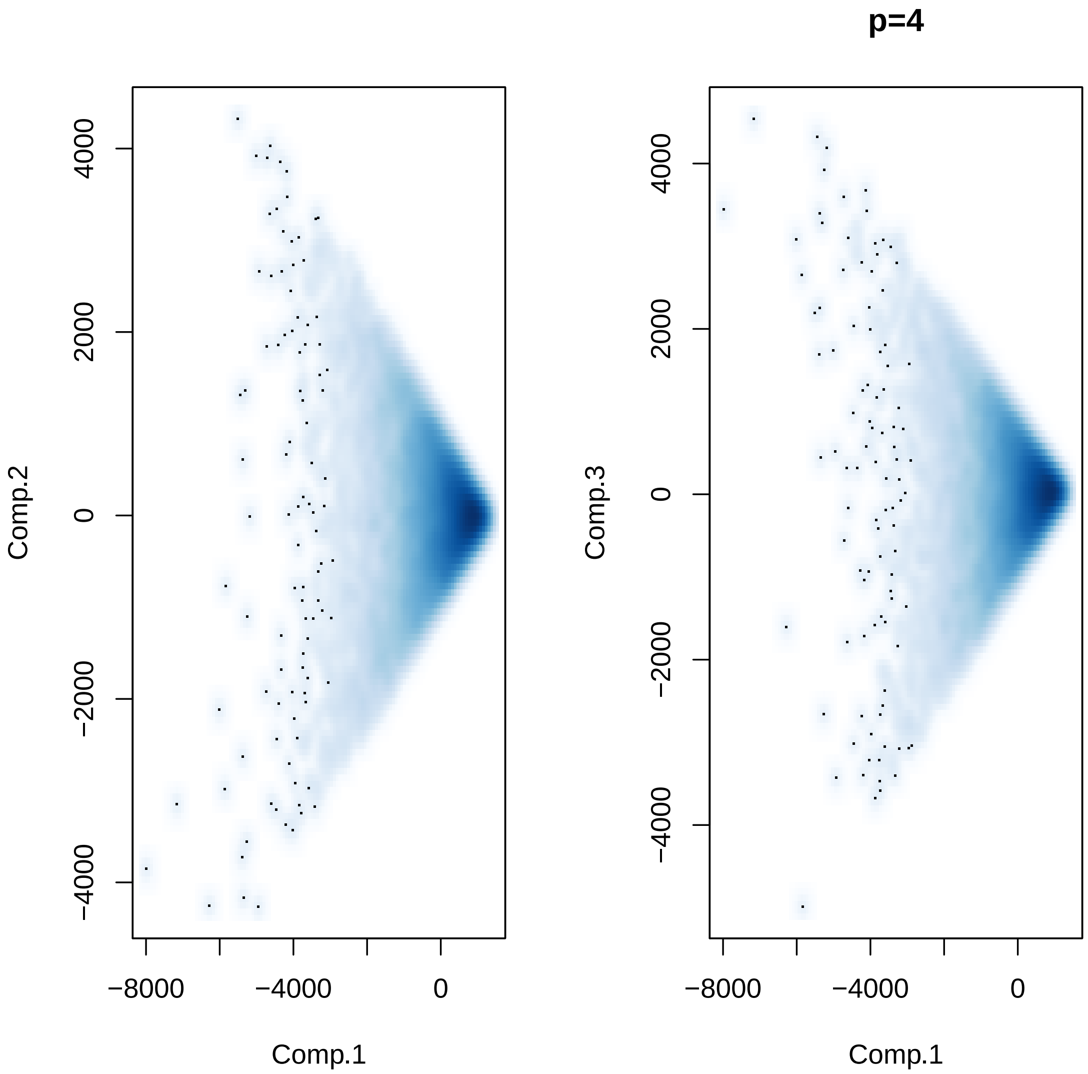

p=64
Comp.2
Comp.3
Comp.1
Comp.1
−0.10
0.00
0.10
0.10
0.05
0.00
−0.05
−0.10

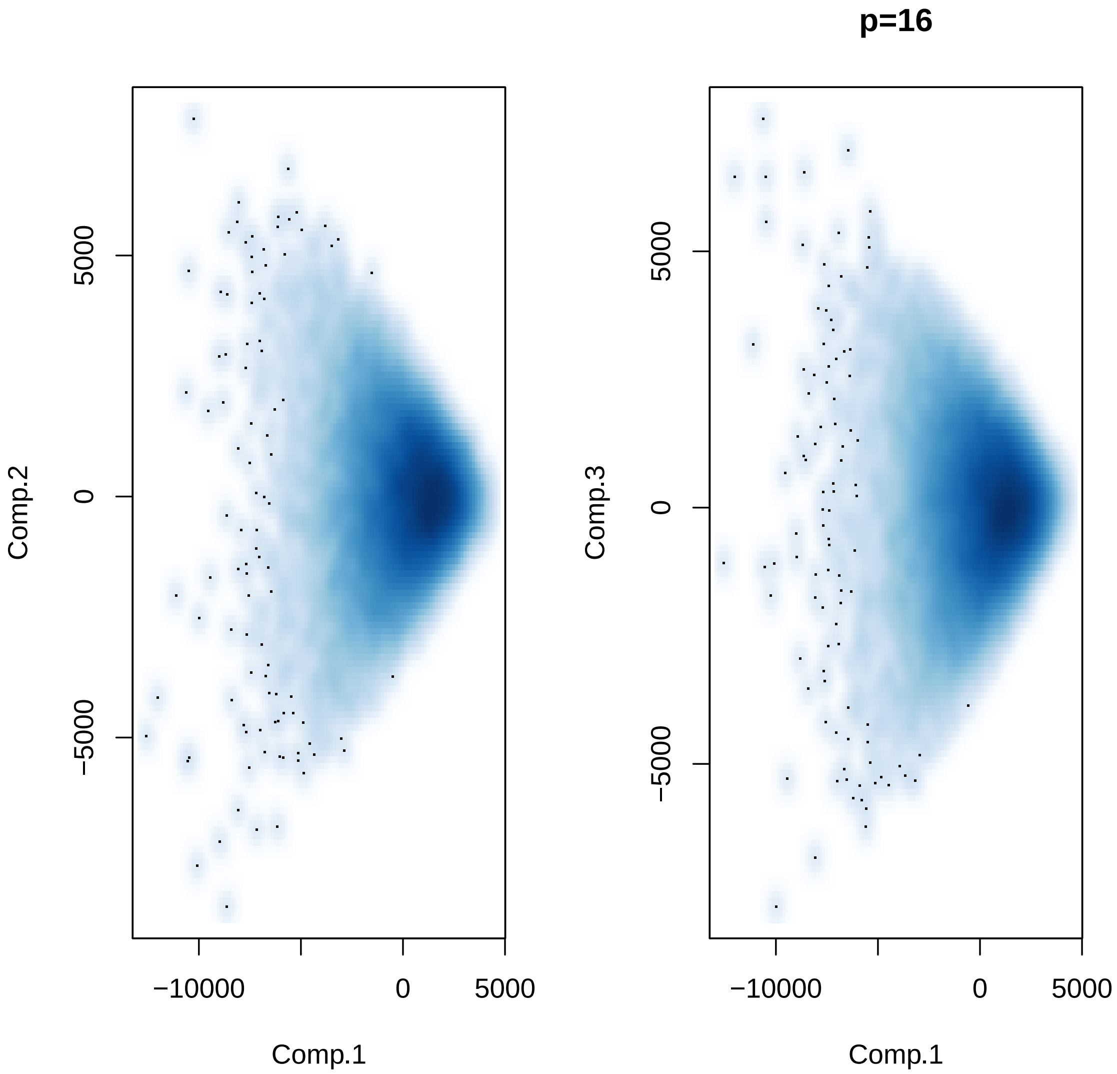

p=128
Comp.2
0.10
0.05
0.00
−0.05
−0.10
0.00
0.10
Comp.1
Comp.3
0.05
0.00
−0.05
−0.10
0.00
0.10
Comp.1

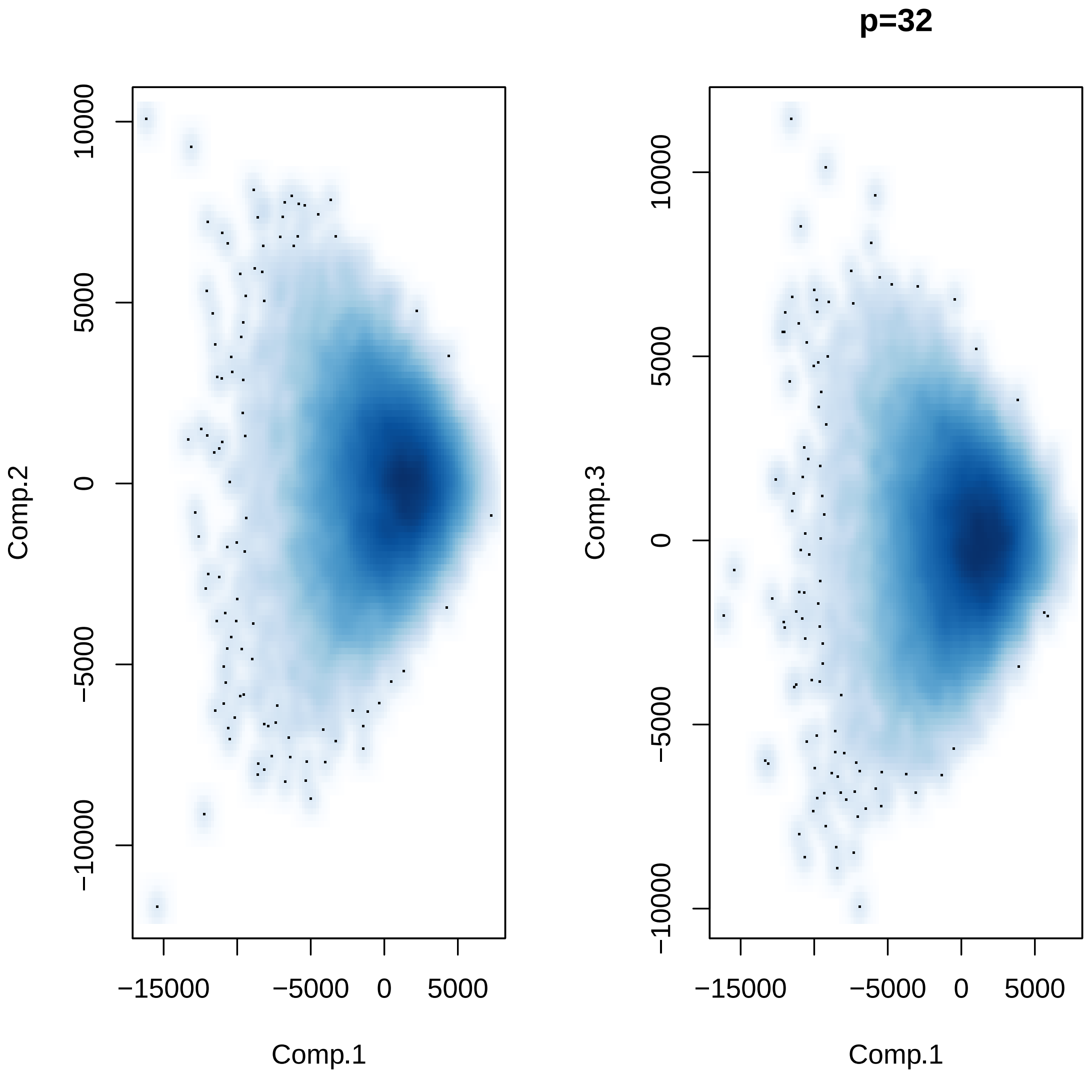

p=256
Comp.2
0.04
0.02
0.00
−0.02
−0.04
Comp.1
−0.06
−0.02
0.02
0.06
Comp.3
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
Comp.1
−0.06
−0.02
0.02
0.06

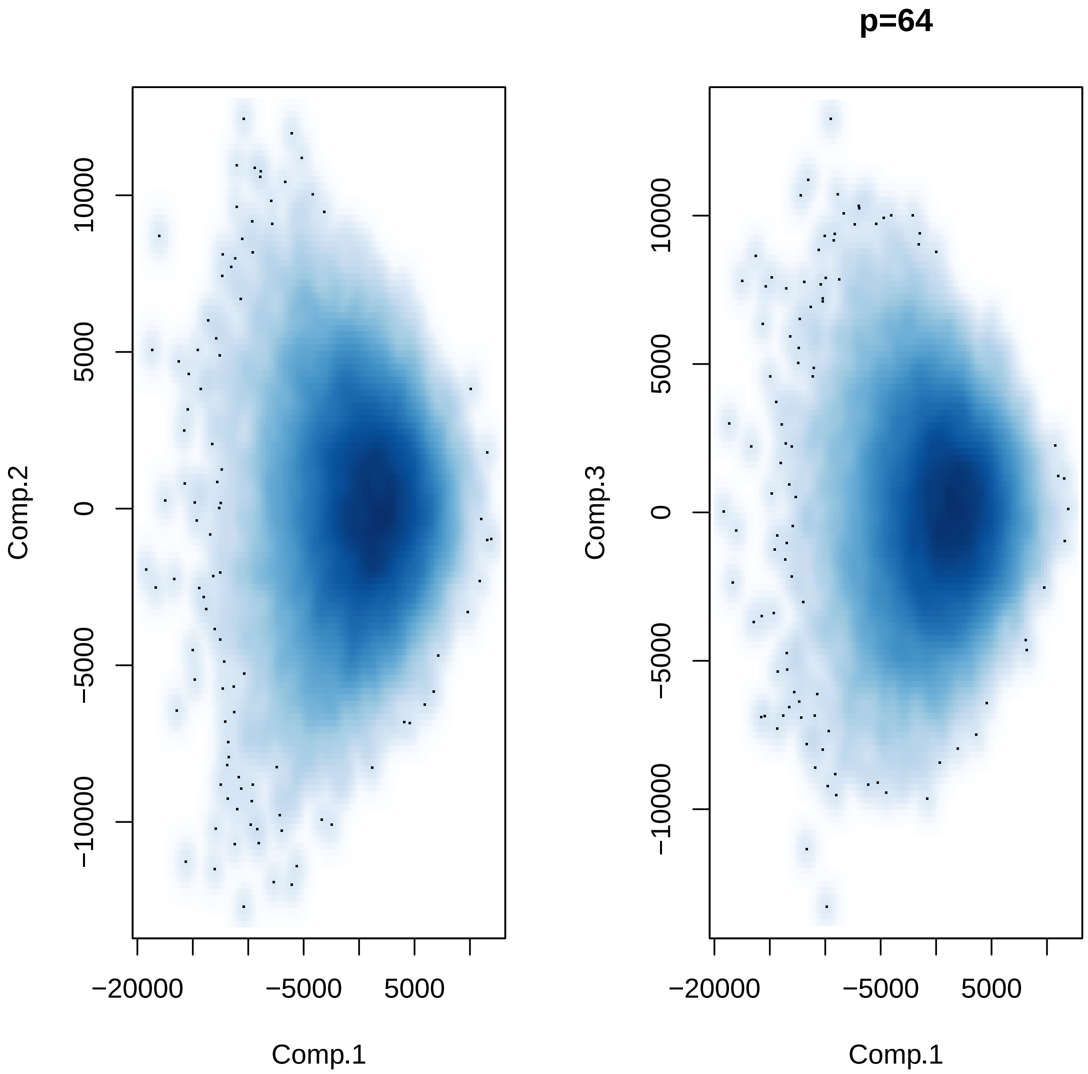

p=512
Comp.2
0.04
0.02
0.00
−0.02
Comp.1
−0.04
0.00
0.04
Comp.3
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03
Comp.1
−0.04
0.00
0.04

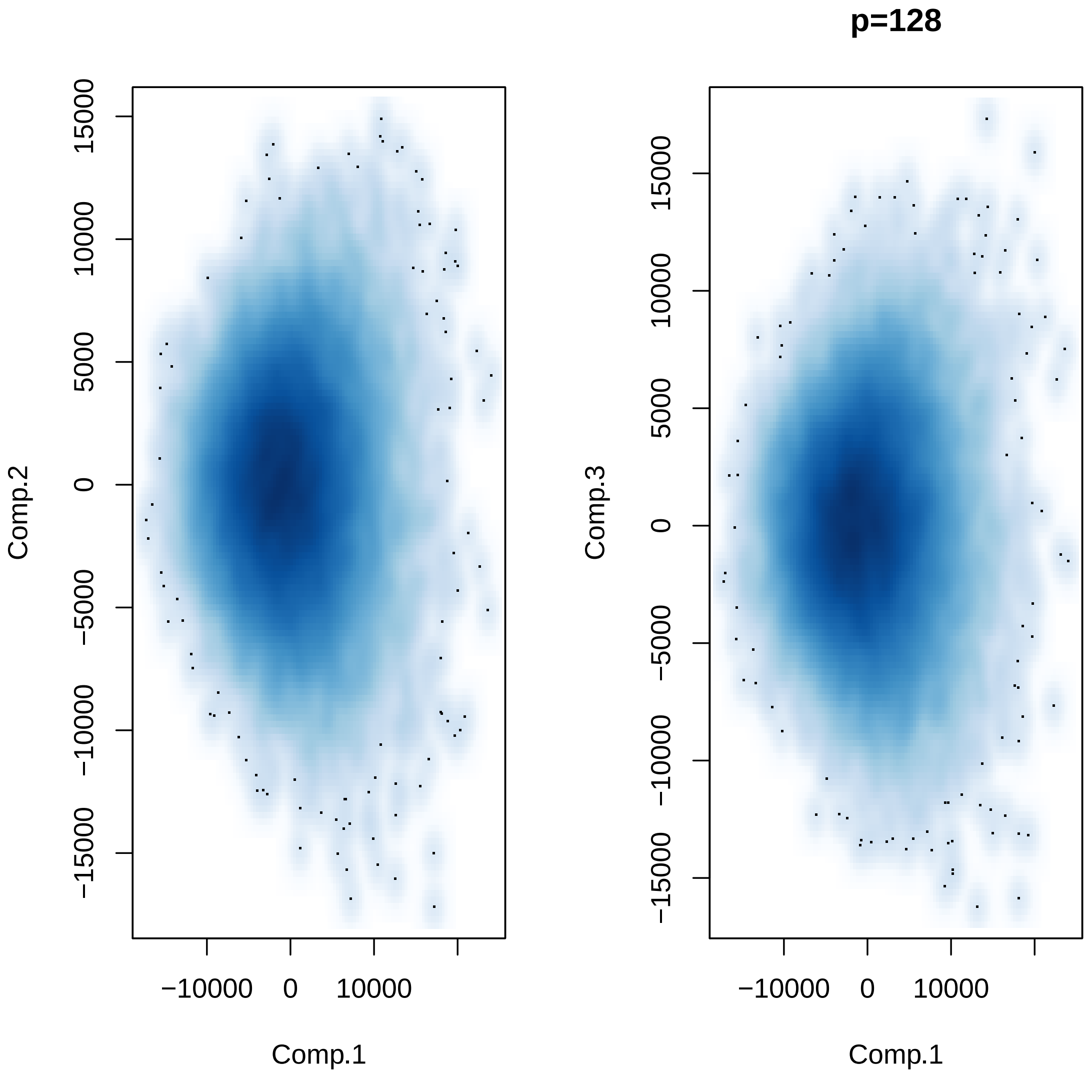
10.4.5 B-spline basis with Euclidean metric convergence plots

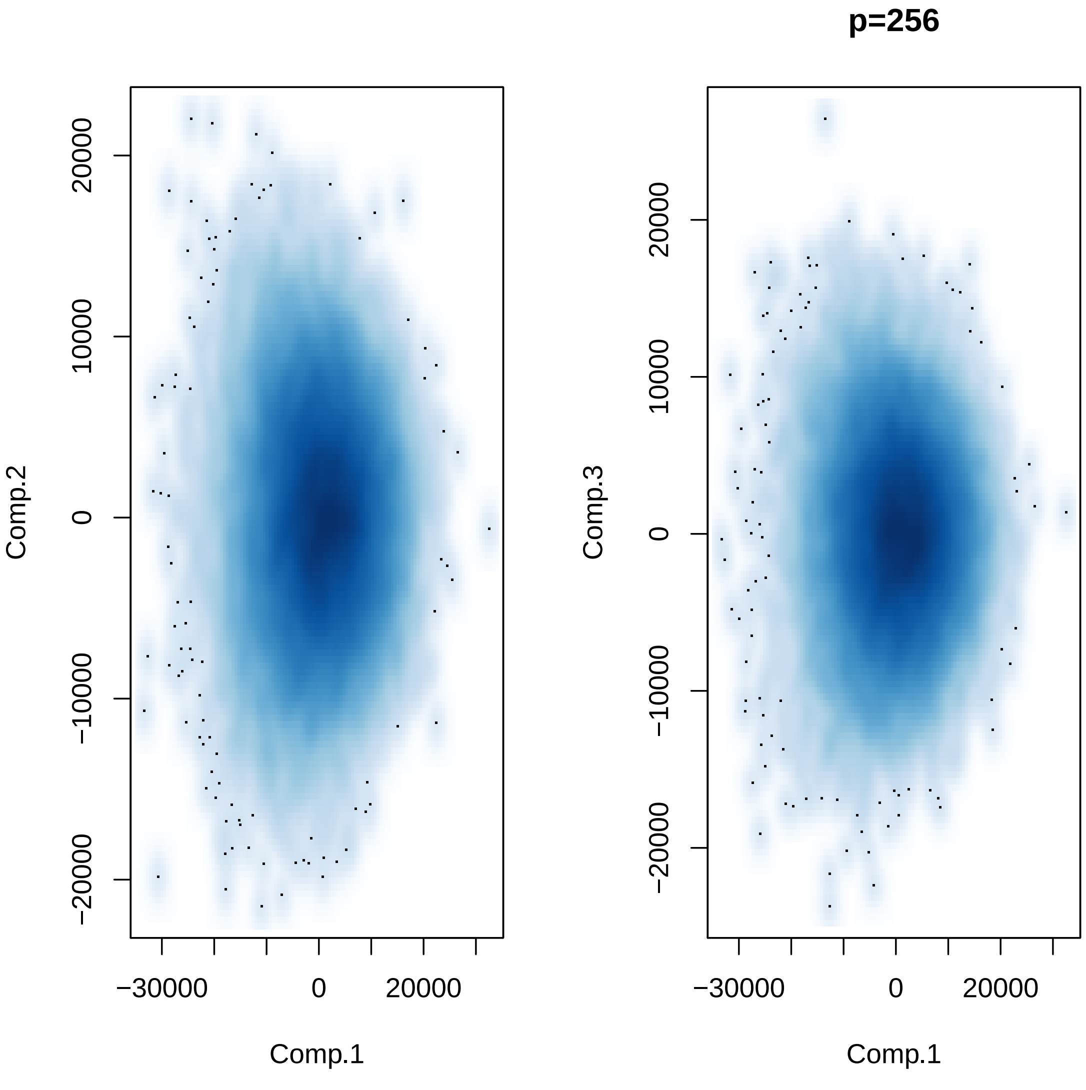

p=4
Comp.2
4000
2000
0
−2000
−4000
−8000
−4000
0
Comp.1
Comp.3
4000
2000
0
−2000
−4000
−8000
−4000
0
Comp.1

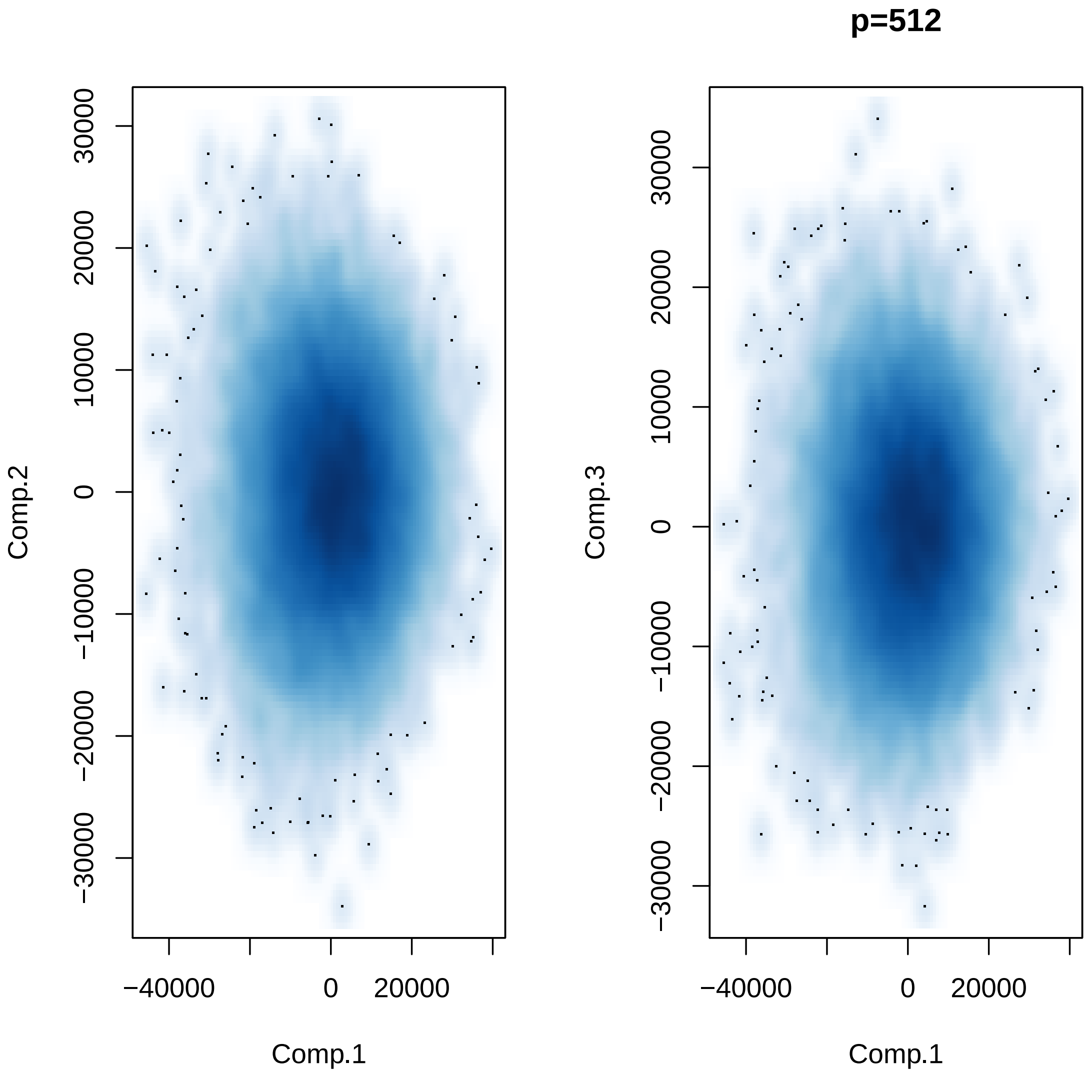

p=16
Comp.2
5000
0
−5000
−10000
0
5000
Comp.1
Comp.3
5000
0
−5000
−10000
0
5000
Comp.1


p=32
Comp.2
10000
5000
0
−5000
−10000
−15000
−5000
0
5000
Comp.1
Comp.3
10000
5000
0
−5000
−10000
−15000
−5000
0
5000
Comp.1


p=64
Comp.2
10000
5000
0
−5000
−10000
−20000
−5000
5000
Comp.1
Comp.3
10000
5000
0
−5000
−10000
−20000
−5000
5000
Comp.1


p=128
Comp.2
15000
10000
5000
0
−5000
−10000
−15000
Comp.1
−10000
0
10000
Comp.3
15000
10000
5000
0
−5000
−10000
−15000
Comp.1
−10000
0
10000


p=256
Comp.2
20000
10000
0
−10000
−20000
−30000
0
20000
Comp.1
Comp.3
20000
10000
0
−10000
−20000
−30000
0
20000
Comp.1


p=512
Comp.2
30000
20000
10000
0
−10000
−20000
−30000
−40000
0
20000
Comp.1
Comp.3
30000
20000
10000
0
−10000
−20000
−30000
−40000
0
20000
Comp.1

10.5 Manuscripts

10.5.1 Draft manuscript for SBF paper

Foundational properties of numerical methods for the quantification of the weight of forensic evidence

by

Ausdemore, M.A., Armstrong, D.E, Ommen, D.M., Neumann, C., Saunders, C.P,
Henricks, J.H., Bayer, D.M , Leegwater, J., Huang, W.

1. Introduction

This paper is concerned with the quantification of the weight of forensic evidence[1], and more specifically with the validation of numerical methods used to calculate the weight of complex and high-dimension form of evidence, such as pattern evidence (e.g., fingerprints, firearms, footwear). The validation of a model involves verifying that its assumptions are reasonable and determining its range of application. This may include attesting that the information contained in forensic traces and control material is characterized using appropriate and robust measurements, determining the number of observations that are necessary to define within-object and between-objects variations, and confirming that the assumptions used to model the probability distributions[2] of these characteristics are sound.

The main issue is that. for most evidence types, and in particular the ones considered in this paper, the output of numerical methods aimed at quantifying the weight of the evidence cannot be compared against an analytical or empirical *gold standard* for the probative value of a given trace. A second issue is that numerical methods rely on a series of assumptions aimed at making the problem of interpreting complex form of evidence tractable. It is often not possible to test the robustness of each assumption independently and the numerical method has to be considered as whole: all assumptions are tested jointly.

In the next sections, we first discuss some aspects of the quantification of the probative value of forensic evidence using Bayes factors and of the validation of numerical methods; secondly, we present some general properties of Bayes factors[3], which need to be satisfied by numerical methods before they can be declared valid in the forensic and legal contexts; then, we illustrate how the properties can help inform on the appropriateness and range of application of numerical methods during their validation using a series of methods commonly proposed in the forensic literature; finally, we discuss the practical application of these properties for the validation of methods for casework purposes, and the implications of our results in the general forensic and legal contexts.

2. Development and validation of numerical methods

[1] Following Good [14], we define the *weight of evidence* as the logarithm (in our case, base 10) of the *Bayes factor* for the evidence, although we will use both terms interchangeably.

[2] Throughout this paper, we will use probability as a measure of belief.

[3] The Bayes factor is often referred to as the *likelihood ratio* in the forensic science community. We find that there is no consensus on the definition of a *likelihood ratio* in the statistical literature. Hence, we will use the term Bayes factor throughout this paper.

2.1 *Quantification of the weight of forensic evidence*

As early as 1966, Parker [27, 28] reported that the inference process of the source of a trace needs to account for two different elements: the similarity between the trace and control material from the suspected source, and the specificity of the characteristics of the trace in a population of plausible sources. His work led to a seminal paper by Lindley in 1977 [20] describing the use of the Bayes factor as a means to quantify the weight of forensic evidence. In the forensic context, a Bayes factor is the ratio of the probability of observing the forensic evidence under two mutually exclusive propositions, and can be written in general as:

$$V = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}, \tag{1}$$

where:

- E is the forensic evidence. Traditionally, it consists in a series of observations Y made on a trace and observations X made on control material, with $E = \{X, Y\}$;
- H_p is the "prosecution" proposition that the trace originates from the same source as the control material;
- H_d is the "defence" proposition that the trace originates from a different source then the control material;
- I represents a framework of circumstances, or knowledge base, pertaining to the quantification of the weight of E. In particular, I embodies the reasons behind the assumptions and simplifications made during the development of the model.

A Bayes factor larger than one provides support for the proposition that the trace and control material originates from the same source, while a Bayes factor smaller than one provides support for the alternative proposition. Finally, a Bayes factor of 1 indicates that the evidence is not helpful to address the considered propositions. The magnitude of the Bayes factor gives a measure of the strength of the support in favour of the supported proposition: the larger it gets, the stronger the support for H_p; conversely, the closer it gets to 0, the stronger the support for H_d.

2.2 *Quantification of the weight of forensic evidence in practice*

Today, most legal and scientific scholars agree that Bayes factors should be used to report the weight of forensic evidence in court, and legal and forensic communities have started releasing guidelines aiming at standardizing forensic conclusions using Bayes factors [7]. Alas, to this day, very few rigorous and robust statistical approaches have been proposed to calculate Bayes factors for non-DNA forms of forensic evidence, and none has been satisfactorily validated for use in casework.

Over the past three decades, considerable work has been performed to extend Lindley's work. Early models were developed to quantify the weight of glass evidence [8,10] and were subsequently extended to blood typing [9], and forensic DNA profiling [11, for a review of the early work see 19]. Development of these early models, while not trivial, was facilitated by the low dimensional nature of the considered characteristics, and the independence or normality assumptions that were made. For example, early models for glass evidence focused on univariate distributions of refractive index of glass fragments [8, 10], while later models relied on multivariate distributions of a few chemical elements under normality assumptions [1]; similarly, most DNA models proposed in the 1990s and 2000s considered discrete allelic designation and assumed independence between loci [11, 12].

The extension of the work initiated by Parker [27, 28], Lindley [20], and Evett [8, 9, 10, 11] to more complex forms of evidence, such as pattern evidence, is impeded by their natural complexity. In most case, it is not possible to derive analytical models to assign Bayes factors to these forms of evidence and numerical methods have to be used instead. For instance, Neumann et al. for fingerprint [24], or Bozza et al. for the shape of handwritten letters [3] have proposed approaches, which reduce the complexity of the problem by using concise summary of the pattern's features and multiple modelling assumptions.

Other methods suggest to reduce the dimensionality of the problem by modelling the probability distributions of the level of similarity between observations made on trace and control material. This is different than the approaches proposed by Neumann et al. [24] or Bozza et al. [3] in that similarity-based methods use the joint summary of the observations made on trace and control objects. The use of similarity measures has proven convenient, as the terms of the ratio in Eq. (1) can be reduced to univariate continuous distributions. This convenience resulted in the apparition of several *ad hoc* methods for approximating the weight of the evidence [2, 6, 13, 22, 23, 25, 33]. The statistical rigor of these methods has been questioned in the forensic literature [16, 26]. In particular, the appropriateness of using values calculated by such methods as proxy for the weight of forensic evidence is of great concern.

2.3 Related work on the validation of numerical methods for quantifying the weight of forensic evidence

Overall, the validation of any statistical method aimed at quantifying the weight of forensic evidence is an open problem. Some researchers have assessed the performance of the proposed methods by measuring their rates of misleading evidence in large scale simulation settings [15, 23]. Showing that a given method repeatedly supports the correct propositions in laboratory conditions is an important first step; however, it does not imply that the magnitude of the support for a given proposition is even remotely appropriate.

Concentrating on the rates of misleading evidence of a numerical method is arguably equivalent to considering the method as a discrimination tool with know error rates. The use of this type of techniques to infer the source of forensic traces has been explicitly discouraged [7]. Focusing on rates of misleading evidence may be suitable in the biometric context (e.g., access control, database search) where the aim is to discriminate between two propositions, and where reasonable prior odds can be assigned to these propositions. Rates of misleading evidence do not inform on whether a particular method supports a given proposition with the appropriate magnitude. Yet, in the legal context, the magnitude is of critical importance since grossly overestimating the weight of the evidence can seriously distort the fact-finding process and be prejudicial to the accused[4].

In order to provide a more refined measure of the performance of numerical methods and to compare them to some desirable behaviour, some authors have proposed to study the entropy of their outputs. In forensic context, the entropy has been defined as *the average information that the fact-finder still needs, once the evidence under consideration has been analysed, in order to know which proposition is actually true* [17]. The empirical cross-entropy (ECE) proposed by [4, 15, 29, 30] does not directly consider the appropriateness of the magnitude of the weight of a specific piece of evidence reported by a numerical method: ECE assesses it by proxy of the posterior probabilities that could be obtained using weights of evidence calculated by the numerical method in a large scale experiment on training data, and a suitably chosen range of prior odds. ECE does not indicate whether the probative value of a specific piece of evidence is over- or under- evaluated; it measures the general lack of calibration of a method in large scale simulations by considering that, when using a properly *calibrated* method, it should be *x* times more likely to have a weight of evidence of *x* when the proposition considered by the numerator of the Bayes factor is true than when the denominator proposition is true [14]. We plan to return to the relationship between ECE and the general properties proposed in this paper during a future research project.

3. General properties of Bayes factors

To continue the discussion and progress towards the validation of a statistical method for non-DNA evidence, we describe 4 fundamental properties of the Bayes factor, which should be satisfied by any method designed to quantify the weight of forensic evidence. These general properties are applicable to any model designed to assign Bayes factors. In the forensic and legal contexts, we will show that these properties, while not exhaustive, can be used during the validation of a statistical method to assess its appropriateness and its range of application. Importantly, these methods can be used to determine if the weight of any specific piece of evidence is appropriate. The properties are outlined below and are expanded upon later on in the section.

Given two fixed and mutually exclusive propositions:

(1) As the amount of information included in the evidence increases, the Bayes factor for that evidence tends to infinite (or 0);

(2) There exists an upper bound value for the Bayes factor of a particular piece of evidence with limited information;

[4] For example, consider that a partial DNA profile is obtained from some biologic material found at a crime scene. A suspect is found to share similarities (and no dissimilarities) with the DNA profile observed on the trace. A jury will perceive the probative value of the evidence differently and may reach different conclusions if the reported Bayes factor is one thousand, or one billion. Depending on the case circumstances, the defense may be able to argue that the prior odds against the defendant are low enough that a Bayes factor of one thousand is not sufficient to reach a conclusion *beyond reasonable doubt*. A similar argument will be excessively difficult to make if the reported Bayes factor turns out to be one billion.

(3) The expected value of the Bayes factor for a given piece of evidence, when the denominator proposition is true, is equal to 1;

(4) Given a fixed knowledge base, the evidence can only support one of two mutually exclusive propositions (unless the weight of evidence is 1).

A Bayes factor is the ratio between two probabilities. Following Good [14], Jaynes [17], Jeffrey [18], Lindley [21], Savage [31], and many others (for a recent review see [34]), we take the view throughout this paper that probabilities can only represent the degree of belief of an individual about an event. The uncertainty of that individual about the event is influenced by his relationship to the event and by the information that he has about the event. Two individuals possessing different pieces of information about a particular event may very well have different degrees of belief about that event. Thus, probabilities are *subjective* in the sense that they represent the *personal* relationship between the *subject* and the event.

The Bayes factor is not an intrinsic property of the evidence in itself, and we want to be very clear that we do not claim that there is such thing as a *true* or *universal* Bayes factor for a given piece of evidence. Surely, different weights will be assigned to the same evidence if different propositions are considered for the same evidence. Moreover, different scientists may also assign different weights to the evidence, for a fixed pair of alternative propositions, based on their personal handling of the available evidence material:

(1) The evidence form may be characterized using different types of features or measured using different analytical techniques. For example, glass fragments may be characterized by their refractive index, by their elemental composition, or by their chemical structure;

(2) Data may be summarized or organized in different ways. For example, Neumann et al. [24] describes a method to characterize the spatial relationships between fingerprint landmarks (i.e., minutiae) using triangles and used these triangles to assign probability distributions to minutiae constellations. However, it is certainly possible to characterize the spatial relationships between minutiae in many other ways;

(3) Different assumptions may be used to model the distributions of the measured characteristics. Given a set of observations, scientists may choose to rely on normality assumptions, use another parametric model, or use non-parametric models.

Nonetheless, *subjective* or *personal* is not meant to suggest, or justify, that probability can be assigned arbitrarily, or reflect *sloppy thinking*. [21, 34]. Personal probabilities must be coherent and follow the ordinary axioms of probability. The properties presented above are discussed assuming two fixed and mutually exclusive propositions.

First property: As the amount of information included in the evidence increases, the Bayes factor for that evidence tends to infinite (or 0).

The first property stems from Savage [32], who showed that as the amount of information contained in the evidence increases, its weight will tend to infinite (or 0). In other words, if the evidence contains an unlimited amount of information, it will remove all uncertainty with respect to the considered propositions. For instance, if a scientist could measure the refractive indices of an unlimited number of trace and control fragments, the resulting Bayes factor addressing the proposition that the trace and control fragments originate from the same source, H_p, vs. the proposition that the trace and control fragments originate from different sources, H_d, would tend to either infinite or 0. In this case, the explanation of this phenomenon lives in the shrinkage of the variance of the distribution of sample averages as the number of observation increases. As the number of fragments increases, the precision of the mean estimates increases and the two sample averages converge with each other if the two sets of fragments originate from the same window, or diverge if they do not.

This property is not particularly useful in realistic situations where the amount of trace and control material is fixed and limited. However, it enables us to justify the second property that Bayes factors need to satisfy.

Second property: There exists a theoretical upper/lower bound value for the Bayes factor of a particular piece of evidence.

Realistic situations involve limited amount of evidence. This is particularly true for forensic evidence, where recovered traces are often partial or degraded. Based on the first property, it is reasonable to assume that as the amount of information contained in the evidence is reduced, the resulting Bayes factor should tend to 1. This assumption holds for any types of reduction of the information contained in the evidence: whether one considers degraded traces, limited amount of control material, or the use of summary statistics and modelling assumptions to reduce the complexity of the development of a statistical method.

The theoretical upper/lower bound for the Bayes factor corresponds to the weight of the evidence that would be assigned based on the entire amount of information contained in the evidence, and without any modelling assumptions. It follows that any numerical value assigned based on a summary of the information contained in the evidence, using modelling assumptions, or within an incomplete framework of circumstances, should take a value between 1 and the bound. Alternatively, numerical values assigned based on sufficient statistics for the evidence and comprehensive knowledge base should be limited by that bound[5].

A similar property has been described by Cowell et al. [5] in relation to forensic DNA evidence. Cowell et al. show that the probative value of a trace with a mixture of genetic profiles, when the

[5] We wish to emphasise the distinction between *upper/lower bound* and *true value* for the Bayes factor. The *true weight of evidence* represents some intrinsic characteristic of the evidence that it is possible to *estimate*. As we mentioned, it does not exist since probabilities represent measures of personal uncertainty. The *bound* weight of evidence represents the *theoretical* maximum reduction in uncertainty that can be achieved with respect to two propositions by using all of the information contained in the evidence and a complete knowledge base.

proposition that a known individual contributed to the mixture is true, is bounded by the value that would be assigned to a trace with a single profile that is similar to that of the known individual.
In most realistic forensic cases, the upper bound is only theoretical and this property cannot be tested directly. However, it is possible to design situations that are so trivial that most scientists would share the same knowledge base and agree on the models that characterises the evidence. In such situations, the upper bound can be either derived analytically or determined empirically. These situations can be used to test the properties, assumptions, range of application and limitations of numerical methods that are considered for assigning the weight of evidence in more realistic cases. The general argument is that if a numerical method does not behave properly in the trivial situation, it cannot be deemed appropriate in the realistic one.

Third property: The expected value of the Bayes factor for a given piece of evidence, when the denominator proposition is true, is equal to 1.

This property stems from Turing [reported by Good in 14]. The interpretation of this property in the forensic context is that, sometimes, Bayes factors can support the proposition that a considered trace originates from the same source as the control material, when in fact it originates from another source in the population of plausible sources. According to the third property, the rate and magnitude of erroneous support has to be entirely compensated by the rate and magnitude of correct support for the proposition that the trace and control material are from different sources.

Fourth property: Given a fixed knowledge base, the evidence can only support one of two mutually exclusive propositions (unless the weight of evidence is 1).

Contrary to the three properties mentioned above, this property requires a fixed framework of circumstances. In Bayesian decision theory, the coherency principle assumes that degrees of belief obey to the axioms of probability and that consistent decisions can be made based on personal probabilities. In our context, the coherency principle imposes that, given two mutually exclusive propositions, a series of modelling assumptions, a reference population of potential sources and a set of observations made on a trace and a control objects, the numerical method can only support one of the proposition, unless the value returned by the method is 1.

In particular, the numerical method should not be influenced by which proposition is considered first. Mathematically, this property is equivalent to say that:

$$V_{H_p/H_d} = \frac{\Pr(E|H_p,I)}{\Pr(E|H_d,I)} = \frac{1}{V_{H_d/H_p}} = \frac{1}{\frac{\Pr(E|H_d,I)}{\Pr(E|H_p,I)}} \quad (2)$$

4. Application of the properties to test numerical methods

The purpose of this paper is to contribute to the discussion on the validation of the various assumptions involved in the development of any method for quantifying the weight of complex forms of forensic evidence. These assumptions cannot be tested in most realistic scenarios (i.e., for any complex form of evidence type). However, the effect of the modelling assumptions and data reduction techniques used in a numerical method can be tested in trivial situations where the upper bound described in properties 1 and 2 exists. We realise that such trivial situation is abstract. Its power lives in the existence of the upper bound and in our ability to use it to test the assumptions used to develop the numerical method, the argument being that if a numerical method, or a series of assumptions, do not behave properly in the trivial situation, they cannot be trusted in realistic situations.

Our trivial situation assumes that all scientists share the same knowledge base and agree on a set of modelling assumptions that are reduced to a minimum; it also allows for controlling the amount of information contained in the evidence and to ensure that this information is accounted for in its entirety in the Bayes factor. In other words, our chosen trivial situation enables us to derive analytically an *ideal* Bayes factor that can serve as the upper bound described in our second property. Importantly, our trivial situation also allows for deriving analytical solutions to several numerical methods proposed over the years and to compare these solutions to the ideal Bayes factor.

In our situation, we consider a set of observations Y, made on a trace, and X, made on material from a control source, and wish to quantify the weight that these observations provide to the proposition that the trace and the control material originate from the same source. The objective of these methods is to assign the Bayes factor presented in Eq. (1).

We find that the set of alternative propositions addressed in the forensic literature and reported at the beginning of this paper is somewhat vague and can be interpreted in different ways. Our attempt to make the propositions less ambiguous results in the specification of two different pairs of propositions:

$H_{p,SS}$: the trace and the control material come from the suspected source;
$H_{d,SS}$: the trace comes from another, unrelated, unknown source in a population of plausible sources.

$H_{p,CS}$: the trace and the control material come from the same, unspecified, source in the population of plausible sources;
$H_{d,CS}$: the trace and the control material come from two different, unrelated, and unknown sources in a population of plausible sources.

The difference between these two sets of propositions resides in the extent of knowledge about the suspected source. In the first pair of propositions, that we call the *specific source* propositions and denote with the *SS* subscript, the suspected source is known and fixed, and the only source of randomness in the model is associated with the trace. It corresponds to a situation where the suspected source is available and can be studied *ad aeternam*, and where the trace is a random product of one of the sources in the considered population.

The second pair of propositions, that we call the *common source* propositions and denote with the *CS* subscript, implies that there are two sources of randomness in the model: one relative to the source of the trace, and one relative to the source of the control material. It corresponds to a situation where the source(s) of the two samples being compared is(are) not available to be characterized. In practice, this pair of propositions might be used to investigate whether two traces have the same unknown source, or address situations where only very limited sample from the suspected source is available.

In most situations, forensic scientists are interested in addressing the *specific source* pair of propositions. However, we found that most models are constructed using the *common source* pair of propositions[6]. For instance, the Bayes factor introduced by Lindley [20] for glass evidence seems to address $H_{p,CS}$ and $H_{d,CS}$. Lindley's belief is that, in absence of other information, the prior distribution for the common mean of $\bar{X}$ and $\bar{Y}$ under the numerator proposition corresponds to the distribution of refractive indices in the general population. This particular choice corresponds to a situation where the common source of $\bar{X}$ and $\bar{Y}$ is not specified in the numerator proposition and can be any source in the general population. On the contrary, it seems that the method proposed by Evett [10] addresses the first pair of propositions since the probability of $\bar{Y}$ is assigned using a distribution centred on μ_x (estimated by the average of the control samples $\bar{X}$). Similarly, different types of methods based on similarity measures have been proposed, some being *anchored* (i.e., conditioned on a fixed source of observations) and seemingly addressing the *specific source* pair of propositions, and some being *non-anchored* (i.e., the actual source of the observation is unknown, and the method is only evaluating the commonality of the origin the trace and control material) and addressing the latter pair of propositions.

In our trivial situation, we define a population of normally distributed sources centred on μ with between-sources variance τ^2. We assume that all observations within a given source are normally distributed with within-source variance σ^2. Finally, we denote by n the number of observations made in relation to a particular source. The subscripts are used to identify specific sources; for example, n_x represents the number of observations made on source x and μ_x represents its mean.

Using X and Y as above, we have:

$H_{p,SS}$: X and Y are two independent samples from the same distribution characterized by μ_x and σ^2. We have:

$$\binom{X}{Y} \sim MVN\left(\binom{\mu_x}{\mu_x}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right), \text{ and } \binom{\bar{X}}{\bar{Y}} \sim MVN\left(\binom{\mu_x}{\mu_x}, \begin{pmatrix} \sigma^2/n_x & 0 \\ 0 & \sigma^2/n_y \end{pmatrix}\right)$$

$H_{d,SS}$: X and Y are two independent samples from two different distributions. We have:

[6] This is typically the case in DNA [11].

$$\binom{X}{Y} \sim MVN\left(\binom{\mu_x}{\mu}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \tau^2 + \sigma^2 \end{pmatrix}\right), \text{and } \binom{\bar{X}}{\bar{Y}} \sim MVN\left(\binom{\mu_x}{\mu}, \begin{pmatrix} \sigma^2/n_x & 0 \\ 0 & \tau^2 + \sigma^2/n_y \end{pmatrix}\right)$$

$H_{p,CS}$: X and Y are two samples from a common source. Since the source is unspecified, it can be any source in the population of potential sources; however, X and Y are covariate. We have:

$$\binom{X}{Y} \sim MVN\left(\binom{\mu}{\mu}, \begin{pmatrix} \tau^2 + \sigma^2 & \tau^2 \\ \tau^2 & \tau^2 + \sigma^2 \end{pmatrix}\right), \text{ and } \binom{\bar{X}}{\bar{Y}} \sim MVN\left(\binom{\mu}{\mu}, \begin{pmatrix} \tau^2 + \sigma^2/n_x & \tau^2 \\ \tau^2 & \tau^2 + \sigma^2/n_y \end{pmatrix}\right).$$

$H_{d,CS}$: X and Y are from two independent and unspecified sources, and we have:

$$\binom{X}{Y} \sim MVN\left(\binom{\mu}{\mu}, \begin{pmatrix} \tau^2 + \sigma^2 & 0 \\ 0 & \tau^2 + \sigma^2 \end{pmatrix}\right), \text{ and } \binom{\bar{X}}{\bar{Y}} \sim MVN\left(\binom{\mu}{\mu}, \begin{pmatrix} \tau^2 + \sigma^2/n_x & 0 \\ 0 & \tau^2 + \sigma^2/n_y \end{pmatrix}\right).$$

We make the further assumption that μ, σ^2 and τ^2 are known and that we are only interested in providing evidence to support whether $\mu_y = \mu_x$. This assumption is not unrealistic since it is conceivable that one is able to study a reference population well enough to have good estimates of these parameters.

Our abstract situation has the nice property that it is so trivial that we can derive analytically the weight of the observations made on any trace Y under both pairs of propositions. In addition, it also enables us to derive the weight of Y for several methods commonly proposed in the forensic literature.

4.1 Ideal Bayes factor for the specific source set of propositions

When μ_x, μ, σ^2 and τ^2 are known, the observations made on the control material do not contribute to reduce the uncertainty on $\bar{Y}$, and the Bayes factor is only influenced by the observations made on the trace and by their number. We have:

$$V_{SS} = \frac{\Pr(\bar{X}, \bar{Y} | H_{p,SS})}{\Pr(\bar{X}, \bar{Y} | H_{p,SS})} = \frac{\Pr(\bar{Y} | \bar{X}, H_{p,SS})}{\Pr(\bar{Y} | \bar{X}, H_{p,SS})} \frac{\Pr(\bar{X} | H_{p,SS})}{\Pr(\bar{X} | H_{p,SS})} = \frac{\Pr(\bar{Y} | H_{p,SS})}{\Pr(\bar{Y} | H_{p,SS})} = \frac{f(\bar{Y} | \mu_x, \sigma^2/n_y)}{f(\bar{Y} | \mu, \tau^2 + \sigma^2/n_y)}, \tag{3}$$

Thus, for a given set of observations on the trace material, V_{SS} is fixed. Given that the sample average, $\bar{Y}$, is a sufficient statistic for the mean of Y, that no modelling assumptions are required since the distribution of $\bar{Y}$ is given by definition, V_{SS} is the ideal Bayes factor described in our second property. According to this property, any numerical method designed to quantify the weight of evidence of the observations made on Y should result in a value that is between 1 and V_{SS}.

4.2 Bayes factor for the common source set of propositions

In the *common source* situation, μ_x is not known and the *common source* Bayes factor is[7]:

$$V_{CS} = \frac{\Pr(\bar{X},\bar{Y}|H_{p,CS})}{\Pr(\bar{X},\bar{Y}|H_{p,CS})} = \frac{\Pr(\bar{Y}|\bar{X},H_{p,CS})}{\Pr(\bar{Y}|\bar{X},H_{p,CS})}\frac{\Pr(\bar{X}|H_{p,CS})}{\Pr(\bar{X}|H_{p,CS})} = \frac{\Pr(\bar{Y}|\bar{X},H_{p,CS})}{\Pr(\bar{Y}|H_{p,CS})} = \frac{f(\bar{Y}|\mu_{CS},\sigma_{CS}^2)}{f(\bar{Y}|\mu,\tau^2+\sigma^2/n_y)} \qquad (4)$$

where $\mu_{CS} = \mu + \frac{\tau^2(\bar{X}-\mu)}{\tau^2+\sigma^2/n_x}$ and $\sigma_{CS}^2 = \frac{\tau^2\sigma^2}{n_x\tau^2+\sigma^2} + \frac{\sigma^2}{n_y}$

Contrary to the *specific source* situation, we note that $\bar{X}$ and $\bar{Y}$ are not independent under H_p. Hence, V_{CS} is not fixed for a given set of observations made on the trace. However, we see that, as the number of observations made on the control material increases (i.e., as the suspected source become better characterized) $\lim_{n_x\to\infty} \mu_{CS} \to \mu_x$ and $\lim_{n_x\to\infty} \sigma_{CS}^2 \to \sigma^2/n_y$, and the *common source* Bayes factor converges to the ideal Bayes factor.

4.3 Numerical methods to assign Bayes factors

Method 1 - Bayes factor proposed by Lindley

Lindley [20] proposes a method to account for the uncertainty on the common mean of X and Y, when only $\bar{X}$ is available to characterise the suspected source of the trace material. From Eq(3) in Lindley [20], we have that:

$$V_{Lindley} = \frac{\int f(\bar{X}|\mu_u) f(\bar{Y}|\mu_u)\pi(\mu_u)d\mu_u}{\int f(\bar{X}|\mu_u)\pi(\mu_u)d\mu_u \int f(\bar{Y}|\mu_{u'})\pi(\mu_{u'})d\mu_{u'}} \qquad (5)$$

We note that in the numerator, the unknown mean of the suspected source, μ_u, is the same for both $\bar{X}$ and $\bar{Y}$, while under the alternative proposition the means of $\bar{X}$ and $\bar{Y}$ are different and $\mu_u \neq \mu_{u'}$.

We also note that:

$$\frac{\int f(\bar{X}|\mu_u) f(\bar{Y}|\mu_u)\pi(\mu_u)d\mu_u}{\int f(\bar{X}|\mu_u)\pi(\mu_u)d\mu_u} = \int f(\bar{Y}|\mu_u)\pi(\mu_u|\bar{X})d\mu_u \qquad (6)$$

which is the marginal distribution of $\bar{Y}$ given $\bar{X}$. Lindley's development assume that, in absence of any information, μ_u and $\mu_{u'}$ follow the same distribution as the general population. In our case, $\mu_u, \mu_{u'} \sim N(\mu,\tau^2)$. This distribution is a conjugate prior for $\pi(\mu_u|\bar{X})$. Using [31, p121], we have:

$$V_{Lindley} = \frac{f(\bar{Y}|\mu_u,\sigma_u^2)}{f(\bar{Y}|\mu,\tau^2+\sigma^2/n_y)}, \qquad (7)$$

[7] See Appendix 1

with $\mu_u = \mu + \frac{\tau^2(\bar{X}-\mu)}{\tau^2+\sigma^2/n_x}$ and $\sigma_u^2 = \frac{\tau^2\sigma^2}{n_x\tau^2+\sigma^2} + \frac{\sigma^2}{n_y}$.

Interestingly, we note that $V_{Lindley} = V_{CS}$ when using the same priors as Lindley and when the distribution of the sources in the general population is completely characterised. Based on this results, it seems that Lindley was addressing the *common source* pair of propositions.

Method 2 - Bayes factor proposed by Lindley with different priors for the distribution of the parameter of X.

Lindley chose very specific priors for the distribution of the common mean of $\bar{X}$ and $\bar{Y}$, which resulted in a *common source* Bayes factor. It is possible to make the Bayes factor proposed by Lindley more source-specific by using any distribution g as the prior distribution of the mean of the suspected source. In the method below, we kept the assumption of normality for algebraic convenience since it enables us to have an analytical solution for $\pi(\mu_u|\bar{X})$. However, we want to emphasis that g can be any distribution, and that the Bayes factor's properties considered in this paper can be used to determine the appropriateness of the assumptions on g and the number of MCMC samples needed to approximate $\pi(\mu_u|\bar{X})$.

Defining $g(\mu_u) = N(\mu_p, \sigma_p^2)$, using [33, p121], we obtain a Bayes factor for the *specific source* set of propositions, as follows:

$$V_{Gen\ SS} = \frac{f(\bar{Y}|\mu_u, \sigma_u^2)}{f(\bar{Y}|\mu, \tau^2 + \sigma^2/n_y)}, \qquad (8)$$

with $\mu_u = \mu + \frac{\sigma_p^2(\bar{X}-\mu)}{\sigma_p^2+\sigma^2/n_x}$ and $\sigma_u^2 = \frac{\sigma_p^2\sigma^2}{n_x\sigma_p^2+\sigma^2} + \frac{\sigma^2}{n_y}$.

As for V_{CS} and $V_{Lindley}$, when the number of observations made on the control material, n_x, increases, we have that $\lim_{n_x\to\infty} \mu_u \to \mu_x$ and $\lim_{n_x\to\infty} \sigma_u^2 \to \sigma^2/n_y$. Nevertheless, we want to point out that the convergence rates of μ_u to μ_x and σ_u^2 to σ^2/n_y are different than for V_{CS} and $V_{Lindley}$ and that they will be dependent on the choice of g.

Method 3 – "Plug-in" Bayes factor proposed by Evett [10]

In the method proposed by Evett [10], the uncertainty on the mean of the distribution of the observations made on X (and on Y under the numerator proposition) is not integrated out as in the two previous methods. Instead, the sample average, $\bar{X}$, is directly used as an estimate of μ_x as follows:

$$V_{Evett} = \frac{f(\bar{Y}|\bar{X}, \sigma^2/n_y)}{f(\bar{Y}|\mu, \tau^2 + \sigma^2/n_y)} \qquad (9)$$

Interestingly, we note that this method is source specific in its construction.

Method 4 – "Suspect-anchored similarity-based" Bayes factor

This method is one of several methods developed to quantify the weight of complex forms of evidence, such as fingerprint evidence. In general, similarity-based methods rely on a distance function, δ, to reduce the complexity of the modelling of the distributions of X and Y. δ usually returns the level of similarity between X and Y as a continuous univariate measure, which distribution can be modelled using parametric or non-parametric techniques.

The *suspect-anchored similarity-based* Bayes factor is source specific in that all considered similarity measures are conditioned on the observations made on the source considered under H_p. According to this method, the set of observations $E = \{\bar{X}, \bar{Y}\}$ in Eq. (1) is replaced by $E = \{\delta(\bar{X}, \bar{Y}), \bar{X}\}$. Using the latter set of observations, we have:

$$\begin{aligned} V_{SLR\ SA\ SS} &= \frac{\Pr(E \mid H_{p,SS})}{\Pr(E \mid H_{d,SS})} = \frac{\Pr(\delta(\bar{X},\bar{Y}),\bar{X} \mid H_{p,SS})}{\Pr(\delta(\bar{X},\bar{Y}),\bar{X} \mid H_{d,SS})} = \frac{\Pr(\delta(\bar{X},\bar{Y}) \mid \bar{X},H_{p,SS})}{\Pr(\delta(\bar{X},\bar{Y}) \mid \bar{X},H_{d,SS})} \frac{\Pr(\bar{X} \mid H_{p,SS})}{\Pr(\bar{X} \mid H_{d,SS})} \\ &= \frac{\Pr(\delta(\bar{X},\bar{Y}) \mid \bar{X},H_{p,SS})}{\Pr(\delta(\bar{X},\bar{Y}) \mid \bar{X},H_{d,SS})} \end{aligned} \tag{10}$$

Contrary to the first 3 methods presented above, the denominator of the *suspect-anchored similarity-based* Bayes factor is not independent of $\bar{X}$. Calculating $V_{SLR\ SA\ SS}$ involves assigning the density of $\delta(\bar{X}, \bar{Y})$ in a distribution of distances between (a) $\bar{X}$ and pseudo-traces generated by the suspected source in the numerator, and (b) $\bar{X}$ and traces generated by randomly selected sources in the population of potential sources in the denominator. Choosing δ as the Euclidean distance between $\bar{X}$ and $\bar{Y}$, we can show that[8]:

$$\begin{aligned} \left(\bar{Y}-\bar{X}\right)^2 \mid \bar{X},H_{p,SS} &\sim \frac{1}{\sigma^2/n_y}\chi^2\left(\frac{\bar{Y}-\bar{X}}{\sigma^2/n_y}, df=1, \lambda=\frac{\mu_x-\bar{X}}{\sigma^2/n_y}\right) \\ \left(\bar{Y}-\bar{X}\right)^2 \mid \bar{X},H_{d,SS} &\sim \frac{1}{\tau^2+\sigma^2/n_y}\chi^2\left(\frac{\bar{Y}-\bar{X}}{\tau^2+\sigma^2/n_y}, df=1, \lambda=\frac{\mu-\bar{X}}{\tau^2+\sigma^2/n_y}\right) \end{aligned} \tag{11}$$

Method 5 – "Non-anchored similarity-based "Bayes factor

The last method considered in this paper is also similarity-based; however, it is neither conditioned

[8] See Appendix 2

on the observations made on the trace, nor on the suspected source. This method is commonly used in biometry for model selection and is widespread in forensic science [13].

By construction, this method addresses the common source pair of propositions: it involves assigning the probability of $\delta(\bar{X},\bar{Y})$ using (a) a distribution of pairwise comparisons between observations made on trace and control material originating from a same source (multiple randomly selected sources are considered in turn), and (b) a distribution of pairwise comparisons between observations made on trace and control material originating from different, randomly selected, sources.

$$V_{SLR\ NA\ CS} = \frac{\Pr(E|H_{p,CS})}{\Pr(E|H_{d,CS})} = \frac{\Pr(d(\bar{X},\bar{Y})|H_{p,CS})}{\Pr(d(\bar{X},\bar{Y})|H_{d,CS})} \tag{12}$$

Using a similar development as the one presented in appendix 2, we can show that:

$$\begin{aligned} \left(\bar{Y}-\bar{X}\right)^2 | H_{p,CS} &\sim \frac{1}{s^2/n_y} c^2\left(\frac{\left(\bar{Y}-\bar{X}\right)^2}{s^2/n_y}, df=1, / = \frac{\left(m-\bar{X}\right)^2}{s^2/n_y}\right) \\ \left(\bar{Y}-\bar{X}\right)^2 | H_{d,CS} &\sim \frac{1}{t^2+s^2/n_y} c^2\left(\frac{\left(\bar{Y}-\bar{X}\right)^2}{t^2+s^2/n_y}, df=1, / = \frac{\left(m-\bar{X}\right)^2}{t^2+s^2/n_y}\right) \end{aligned} \tag{13}$$

5. Results

By defining μ, τ^2, and σ^2, it is possible to calculate the values of the ideal Bayes Factor in Eq. (3) for any given set of observations made on a trace sample, and to study how the methods describe proposed above in Eqs. (4, 7, 8, 9, 10, 12) behave with respect to the four properties described in section 3. In particular, it is possible to observe the convergence of the values calculated by these methods as a function of μ_x, n_x and n_y. The results presented below were obtained using the same data as in [8, 10].

5.1 First property

Figure 1a presents the behaviour of the ideal value of the Bayes factor for a given pair of trace and control samples originating from the same source. In figure 1a, the common mean of the trace and control samples, μ_x, is fixed, and the number of observations made on the trace, n_y, increases. Figure 1a shows that the resulting value of the ideal Bayes factor V_{SS} increases monotonically. Figure 1b shows that the same results can be obtained for the numerical methods, such as the Lindley Bayes factor, $V_{Lindley}$, when n_x and n_y increase at the same rate.

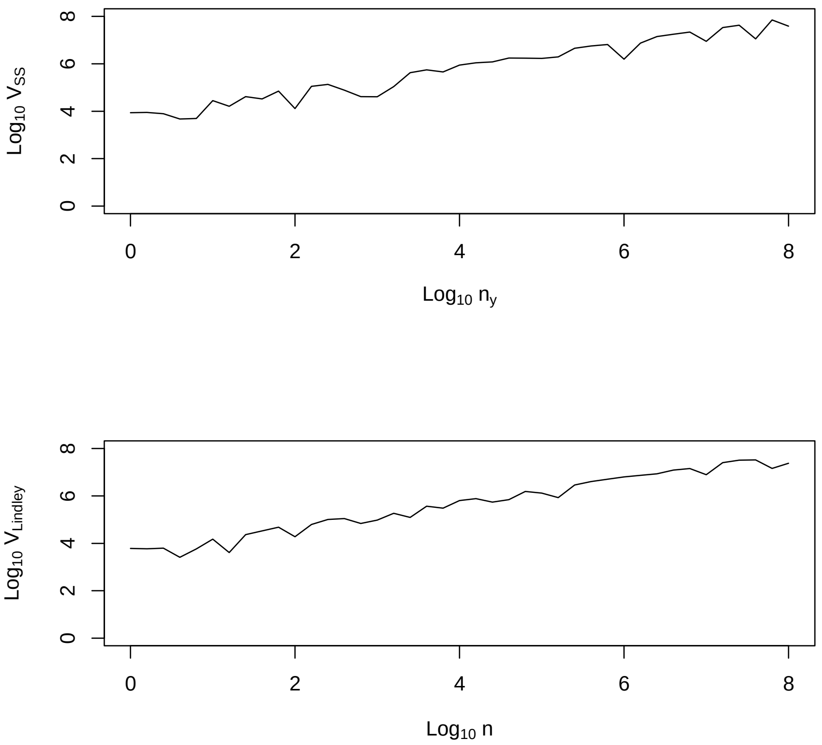


Figure 1: Monotonic increase of the value of the evidence as the number of trace samples (Figure 1a – top) and in the trace and control samples (Figure 1b – bottom) increase. In these simulations, $\mu_x = 1.5302$, $\mu = 1.5182$, $\tau^2 = 1.6 \times 10^{-5}$, $\sigma^2 = 1.6 \times 10^{-9}$.

5.2 Second property

Figures 2-5 show the results of simulations performed for a fixed set of observations made on a trace sample ($n_y = 5$) as the number of observations made on the control sample, n_x, increases. All simulations were performed with $\mu = 1.5182$ and $\tau^2 = 1.6 \times 10^{-5}$. All boxplots in a given figure have been produced by keeping the set of observations made on the trace sample fixed for that figure, and by resampling the source 1,000 times for each n_x.

Figures 2-5 display the convergence of the different numerical methods to the ideal value of the Bayes factor for the 5 observations made on the trace sample (represented by the horizontal line). Figures 2-5 show that the convergence (or lack thereof) of the methods depends mainly on 2 factors: the ratio between the population and source variances, τ^2/σ^2, and the level of rarity of the control material in the population, μ_x.

In figures 2 and 3, μ_x was chosen to represent a common control source and was taken such that $\mu_x = \mu = 1.5182$. In figures 4 and 5, μ_x was chosen to represent a rare control source and was taken such that $\mu_x = 1.5302$. In figures 2 and 4, the ratio between the between- and within-source variances was chosen to be 10^4 with $\sigma^2 = 1.6 \times 10^{-9}$, while this ratio was 10 for figures 3 and 5 ($\sigma^2 = 1.6 \times 10^{-6}$).

Overall, the data presented in figures 2-5 inform us that the common source Bayes factor, V_{CS} - Eq.(4), the Lindley Bayes factor, $V_{Lindley}$ - Eq. (7), the general specific source Bayes factor, $V_{Gen\ SS}$ - Eq.(8), and the Evett Bayes factor, V_{Evett} – Eq.(8) always converge to the ideal value of the Bayes factor as the number of observations made on the control material increases. That said, the number of observations required to have a reasonable convergence for some of these methods is far larger than the number of observations commonly made in forensic practice. For example, the variance of V_{Evett} when n_x is small appears to be very large.

Our data also show that the *suspect-anchored similarity-based* method, $V_{SLR\ SA\ SS}$ - Eq. (10) does not necessarily converge to the ideal value of the Bayes factor for a given set of observations made on a trace sample. As a general rule, $V_{SLR\ SA\ SS}$ converges to the ideal value of the Bayes factor when the within-source variance, σ^2, is considerably smaller than the between-source variance, τ^2, but may over- (as in figure 3) or under- estimate the ideal Bayes factor. This behaviour is not predictable.

Finally, our data show that the *common source similarity-based* method, $V_{SLR\ NA\ CS}$ - Eq. (12), either over-estimate the ideal Bayes factor when the mean of the material from the suspected source is common with respect to the population of potential sources, and under-estimate the ideal value when the source is rare.

In summary, our trivial situation enables us to study the behaviour of the assumptions made in numerical methods with respect to an ideal Bayes factor for a given set of observations made on a trace sample. Our situation allows us for determining that even in the simplest situation:

(1) The number of observations of the control material routinely made in forensic practice can result in significant variations in the value of the evidence calculated by a given numerical method;

(2) The use of summary statistics, such as similarity measures, to represent the evidence will lead to over- and under-estimation of the value of the evidence depending on the ratio of the between- and within-variance of the sources in the population of potential sources, and the rarity of the suspected source in that population. Unfortunately, while it might be possible to assume that the ratio is very large for evidence types such as fingerprints, it is not possible to assume the same for other evidence types. Furthermore, if it was possible to model the distribution of a particular type of evidence, the use of a similarity measure as means to reduce the complexity of the modelling would not be necessary. In other words, it will never be possible to assess directly if the suspected source has common or rare characteristics.

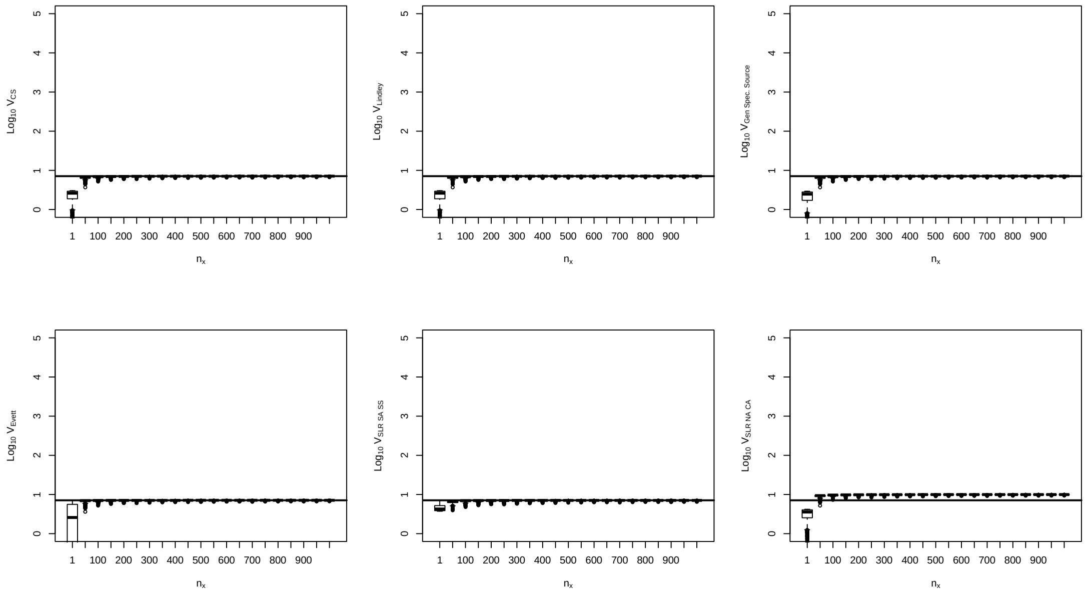


Figure 2. Empirical convergence of Common Source BF - Eq. (4), Lindley BF - Eq. (7) , General SS BF – Eq. (8), Evett BF – Eq.(9), Suspect anchored SLR – Eq. (10) and Non-anchored Source SLR – Eq.(12) for source material with common characteristics $\mu_x = \mu = 1.5182$ and $\tau^2/\sigma^2 = 10$.

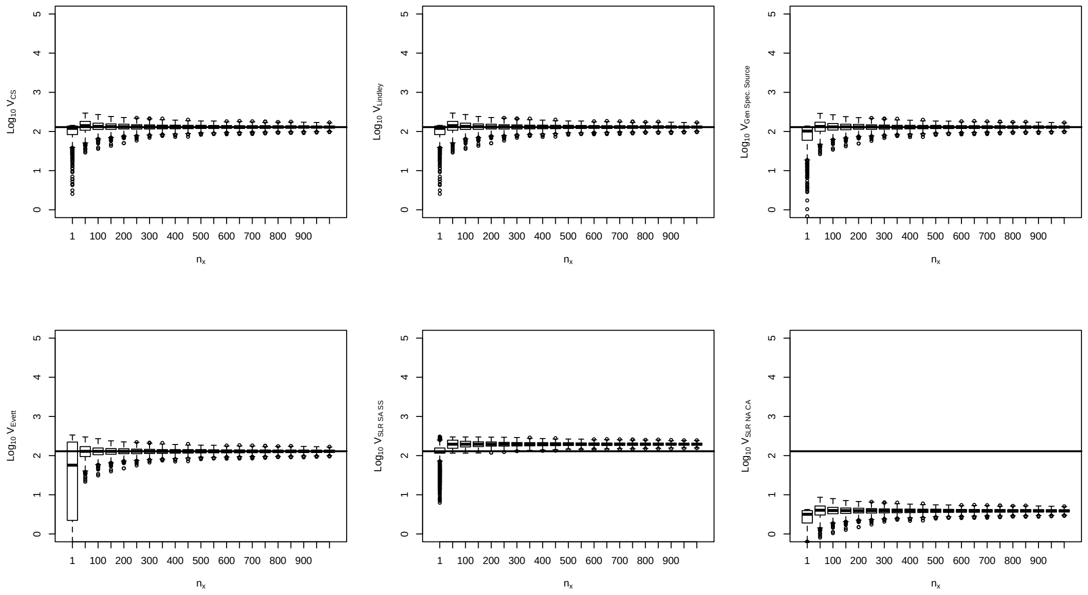


Figure 3. Empirical convergence of Common Source BF - Eq. (4), Lindley BF - Eq. (7) , General SS BF – Eq. (8), Evett BF – Eq.(9), Suspect anchored SLR – Eq. (10) and Non-anchored Source SLR – Eq.(12) for source material with rare characteristics $\mu_x = 1.5302$ and $\tau^2/\sigma^2 = 10$.

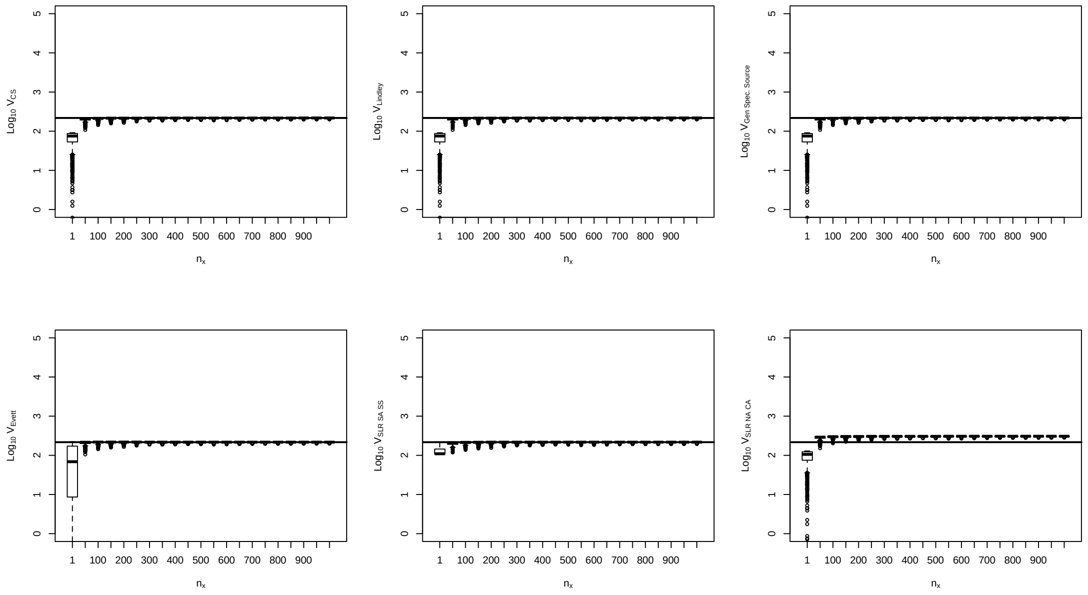


Figure 4. Empirical convergence of Common Source BF - Eq. (4), Lindley BF - Eq. (7) , General SS BF – Eq. (8), Evett BF – Eq.(9), Suspect anchored SLR – Eq. (10) and Non-anchored Source SLR – Eq.(12) for source material with common characteristics $\mu_x = \mu = 1.5182$ and $\tau^2/\sigma^2 = 10{,}000$.

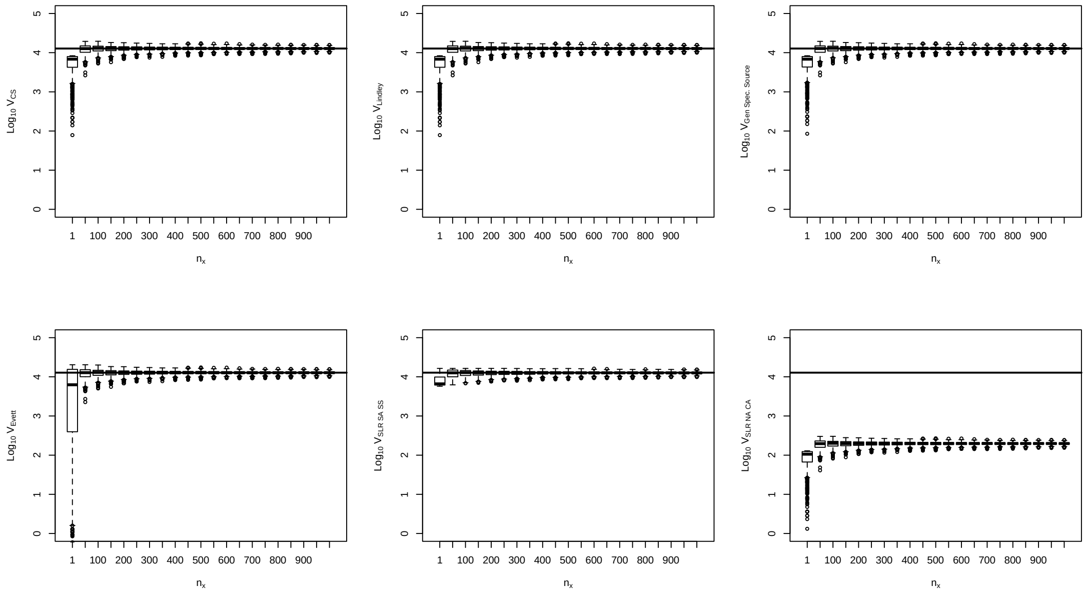


Figure 5. Empirical convergence of Common Source BF - Eq. (4), Lindley BF - Eq. (7) , General SS BF – Eq. (8), Evett BF – Eq.(9), Suspect anchored SLR – Eq. (10) and Non-anchored Source SLR – Eq.(12) for source material with rare characteristics $\mu_x = 1.5302$ and $\tau^2/\sigma^2 = 10{,}000$.

5.3 Third property

Figure 6 shows the results of 100 simulations (for each method) aimed at testing Turing's result that the expected value of the Bayes factor, when the trace and control material do not originate form the same source, is 1. The trace for all simulations was fixed: $n_y = 1$ observation was sampled from a the trace with true mean $\mu_y = 1.5302$. In each simulation, $n_x = 10$ samples were sampled from 100,000 sources in a population of potential sources with mean $\mu = 1.5182$ and between source variance $\tau^2 = 1.6 \times 10^{-5}$. The within-source variance was set at $\sigma^2 = 1.6 \times 10^{-9}$.

Figure 6 shows that all methods satisfy Turing's result, except the *common source similarity-based* method, which has an expected value that is lower than 1. This indicates that this methods, while not well calibrated, is generally more favourable to the defence hypothesis. Similar results can be obtained with values of μ_y representing different levels of rarity of the characteristics of the true source of the trace.

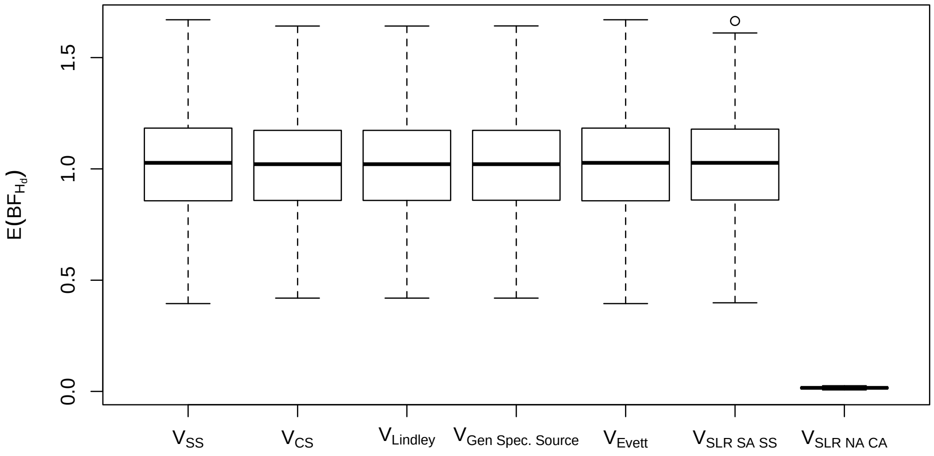


Figure 6: Average weight of evidence assigned to a given trace using different methods when the trace is compared to a randomly selected source from the population of potential sources. Each boxplot represents 100 averages. Each average was calculated by sampling 100,000 sources.

5.4 Fourth property

The coherency property is satisfied by all methods except the two *similarity-based* methods. In these two cases, it is trivial to demonstrate that these methods violate the fourth property. Consider that the population of potential sources has only 2 sources, A and B, and consider a trace Y of unknown origin. At first, we consider that A is the suspected source. By Eq. (10), we have:

$$V_{H_A} = \frac{\Pr(E \mid H_A)}{\Pr(E \mid H_B)} = \frac{\Pr(d(\bar{A},\bar{Y}),\bar{A} \mid H_A)}{\Pr(d(\bar{A},\bar{Y}),\bar{A} \mid H_B)} \quad (14)$$

Now, consider that B is the suspected source. We have

$$V_{H_B} = \frac{\Pr(E\,|\,H_B)}{\Pr(E\,|\,H_A)} = \frac{\Pr(d(\bar{B},\bar{Y}),\bar{B}\,|\,H_B)}{\Pr(d(\bar{B},\bar{Y}),\bar{B}\,|\,H_A)}{}_1\ \frac{1}{\dfrac{\Pr(d(\bar{A},\bar{Y}),\bar{A}\,|\,H_A)}{\Pr(d(\bar{A},\bar{Y}),\bar{A}\,|\,H_B)}} \qquad (15)$$

This implies that for a fixed pair of propositions and a fixed knowledge base, the observations made on the trace may support both hypotheses at the same time, depending on which one is considered in the numerator of the method. This is a clear violation of the coherency principle and, thus, of the fourth property. A similar demonstration can be performed for the *common-source similarity-based* method.

Figure 7 illustrates the effect of the violation of the coherency principle. Figure 7 shows the values of evidence of a trace with varying mean μ_y, for a fixed mean of the suspected source under H_p/H_A and when the considered source under H_d/H_B can be any source in the population of potential sources. In figure 7, $\mu_A = 1.5302$, $\sigma_A^2 = 1.6 \times 10^{-9}$, $\mu_B = 1.5182$ and $\sigma_B^2 = 1.6 \times 10^{-4}$. Figure 7 shows that the ideal Bayes factor will always support the same proposition for any given trace, irrespectively of which proposition is considered first. On the contrary, we can see that this is not the case for the two *similarity-based* methods. Using these methods, it is possible end up in situations where the evidence only supports the hypothesis of a common source between the trace and the control material because that source has been considered first. For example, we see in figures 6b and 6c that for some values of μ_y, the value reported by the *similarity-based* method supports the proposition that the trace sample originates from the source characterised by $\mu_A = 1.5302$, and, at the same time, indicates that it is more likely to observe the trace's characteristics if it were to originate from another randomly selected source in the population (dashed line). Overall, this lack of coherence of similarity-based methods is concerning and can result in miscarriages of justice.

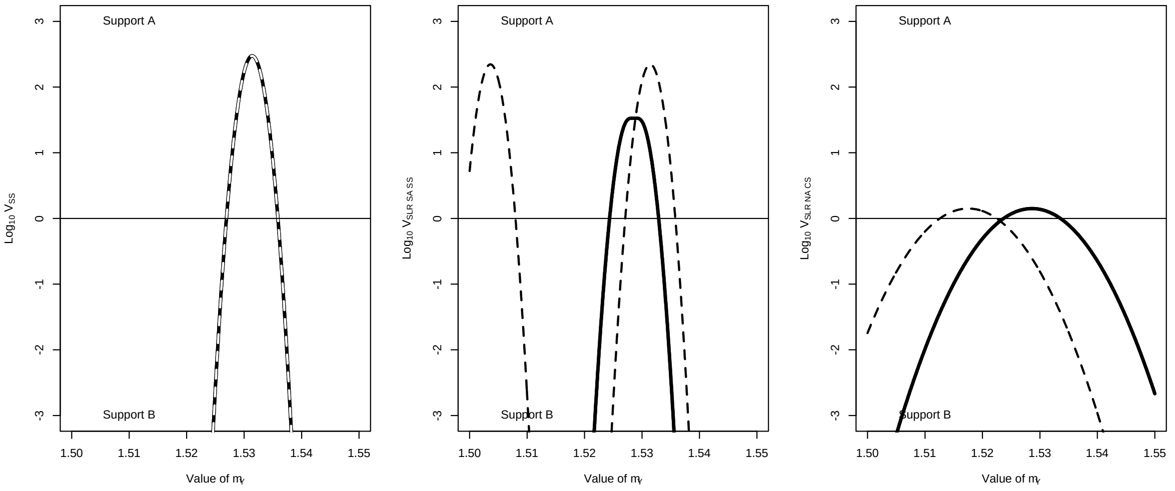


Figure 7: Values of the evidence for a trace with varying mean μ_y, a fixed suspected source $\mu_A = 1.5302, \sigma_A^2 =$

1.6×10^{-9} and a random source from the alternative population in B: $\mu_B = 1.5182$, $\sigma_B^2 = 1.6 \times 10^{-4}$. A vs. B is shown with a plain line; B vs. A is showed with a dashed line. Left panel shows the specific source LR - Eq. (3); middle panel shows the specific source similarity-based method – Eq. (10); the right panel shows the common source similarity-based method – Eq. (12).

6. Discussion of the results and conclusions

Scientific and legal scholars have promoted Bayesian inference as the only rational mean to appropriately account for forensic evidence in the context of a case. There is currently a strong push for reporting the probative value of forensic evidence as a Bayes factor. This push has resulted in the development of a number of ad-hoc methods for quantifying the weight of forensic evidence. The assumptions behind these methods are commonly justified by the subjective nature of probabilities, and their use is supported by large scale experiments in laboratory conditions showing that the rates at which they support the wrong proposition is very low.

This paper presents 4 properties of Bayes factors that directly derive from the foundations of Bayesian statistics. Some of these properties, in particular the first and second properties are quite abstract and cannot be directly applied to numerical methods designed to quantify the weight of complex forms of forensic evidence. On the contrary, the third and fourth properties can readily be tested on any numerical methods using large scale experiment. While not directly applicable to complex forms of evidence types, the first and second properties are extremely useful to test modelling assumptions in a well controlled environment. This paper presents a univariate normal trivial situation; however, it is easily conceivable to design non-normal situations, or multivariate ones. These situations can then be used to test methods of data reduction, rates of convergence of methods aimed at estimating model parameters, or range of applications of these methods.

Our paper provides an example of the application of these properties to various methods commonly encountered in forensic science. In particular, our results show that ad-hoc data reduction techniques, such as the use of similarity measures to quantify the weight of evidence, violate some of the properties. In particular, our simulations show that these methods may overstate the weight of the evidence against a suspected source, and thus, be prejudicial to a defendant. These results do not imply that similarity-based methods cannot be useful, in particular in light of their extremely low error rates; however, they show that similarity-based methods cannot be used as part of the Bayesian inference process advocated by scientific and legal scholars.

Our future work will focus on the relationship between our properties and the empirical-cross-entropy methods discussed above, as well as on the development of trivial situations to test more complex methods designed to quantify the weight of fingerprints, toolmarks and shoeprint evidence.

7. Acknowledgments

The research detailed in this presentation was supported by the National Institute of Justice, Office of Justice Programs, US Department of Justice under Awards No. 2014-IJ-CX-K088 and No. 2015-R2-CX-0028. The opinions and conclusions or recommendations expressed in this presentation are those of the author and do not necessarily represent those of the Department of Justice.

8. Bibliographic references

[1] Aitken, C.G.G., Lucy, D. (2004) Evaluation of trace evidence in the form of multivariate data. Applied Statistics. 53(4) pp. 109-122.

[2] Alberink I., de Jongh, A., Rodriguez, C. (2014) Fingermark evidence evaluation based on Automated Fingerprint Identification System matching scores: the effect of different types of conditioning on Likelihood Ratios, J. For. Sci., 59. 70-81.

[3] Bozza, S., Taroni, F., Marquis, R., Schmittbuhl, M. (2008) Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. J. R. Statistic. Soc. C. 57(3) pp. 329-341.

[4] Brummer N. Measuring, refining and calibrating speaker and language information extracted from speech [Ph.D. thesis]. Stellenbosch, South Africa: School of Electrical Engineering, University of Stellenbosch, 2010; http://sites.google.com/site/nikobrummer/. (accessed June 28, 2013).

[5] Cowell, R.G., Graversen, T., Lauritzen, S.L., Mortera, J. (2015) Analysis of forensic DNA mixtures with artefacts. J. R. Statistic. Soc. C. 64(1) pp. 1-48.

[6] Egli, N., Champod, C., Margot, P. (2006) Evidence evaluation in fingerprint comparison and automated fingerprint identification systems—modeling within finger variability, For. Sci. Int. 176. 189–195.

[7] European Network of Forensic Science Institutes (ENFSI) 2015. ENFSI Guideline for Evaluative Reporting in Forensic Science (Approved Version 3.0, March 8) http://www.enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf

[8] Evett, I.W. (1977) The Interpretation of refractive index measurements. Forensic Science. 9. pp 209-217.

[9] Evett, I.W. (1983) What is the probability that this blood came from that person? A meaningful question? Journal of the Forensic Science Society. 23. pp.35-39.

[10] Evett, I.W. (1986) A Bayesian approach to the problem of interpreting glass evidence in forensic science casework. Journal of the Forensic Science Society. 26. pp. 3-18.

[11] Evett, I.W., Weir, B.S. (1998) Interpreting DNA Evidence. 1st Edition, Sinauer Associates. 278p.

[12] Foreman, L.A., Smith A.F., Evett, I.W. (1997) Bayesian Analysis of DNA profiling data in forensic identification applications. Journal of the Royal Statistical Society, Series A. 3. 429-469.

[13] Gonzalez-Rodriguez, J., Drygajlo, A., Ramos-Castro, D., Garcia-Gomar M., Ortega-Garcia, J. (2005) Bayesian analysis of fingerprint, face and signature evidence with automatic biometrics systems, For. Sci. Int 155. 126-140.

[14] Good, I.J. (1950) Probability and the Weighting of Evdience. Charles Griffin & Co., London. 119p.

[15] Haraksim, R., Ramos, D., Meuwly, D. Berger, C.E.H. (2015) Measuring coherence of computer-assisted likelihood ratio methods. Forensic Science International. 249. pp.123-132.

[16] Hepler, A. B., Saunders, C. P. , Davis, L. J., Buscaglia, J. (2012) Score-based likelihood ratios for handwriting evidence, For. Sci. Int. 219. 129-140.

[17] Jaynes, E.T. (2003) Probability Theory: The Logic of Science. Cambridge University Press. 727p.

[18] Jeffreys, H. (1961) Theory of Probability. Third Edition. Reprinted in 2003. Oxford University Press. 470p.

[19] Kaye, D. H. (1993) DNA evidence: probability ,population genetics and the courts . Harv. J. Law Technol. 7. 101-172.

[20] Lindley, D.V. (1977) A problem in forensic science, Biometrika 64 (1977) 207–213

[21] Lindley, D.V. (2006) Understanding Uncertainty. John Wiley & Sons. 250p.

[22] Meuwly, D. (2001) Reconnaissance de locuteurs en sciences forensiques: L'apport d'une approche automatique, Ph.D. Thesis, University of Lausanne.

[23] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., Bromage-Griffiths, A. (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. J. of Forensic Sci. 52, pp. 54-64.

[24] Neumann, C., Evett, I.W., Skerrett, J. (2012) Quantifying the weight of evidence assigned to a forensic fingerprint comparison: a new paradigm. J. R. Statist. Soc. A. 175. pp. 371-415.

[25] Neumann C., Margot, P. (2009) New perspectives in the use of ink evidence in forensic science. Part III. Operational applications and evaluation. Forensic Sci. Int. 192(1–3). pp. 29–42.

[26] Neumann, C., Saunders, C.P. (2015), On the use of similarity measures in likelihood ratios - Letter to the Editor re Alberink et al. J. For. Sci 59 (2014) 70-81, Journal of Forensic Science, 60(1) pp. 252-256

[27] Parker, J.B. (1966) A statistical treatment of identification problems. Journal of Forensic Science Society. 6. pp. 33-39.

[28] Parker, J.B. (1967) The mathematical evaluation of numerical evidence. Journal of Forensic Science Society. 7(3). Pp. 134-144.

[29] Ramos, D., Gonzalez-Rodriguez, J., Reliable support: Measuring calibration of likelihood ratios, For. Sci. Int., 230 (2013) 156-169.

[30] Ramos, D., Gonzalez-Rodriguez, J., Zadora, G., Aikten, C. (2013) Information-Theoretical Assessment of the Performance of Likelihood Ratio Computation Methods, Journal of Forensic Sciences. 58(6) 1503-1518.

[31] Robert, C. P. (2007) The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation. 2nd Edition. Springer. 603p.

[32] Savage,L.J. (1972) The Foundations of Statistics, Second revised edition, Dover Publications, Inc. New York. 310p.

[33] Srihari, S.N., Cha, S-H., Arora, H., Lee, S. (2002) Individuality of handwriting. J For. Sci. 47. 856-872.

[34] Taroni, F., Bozza, S., Biedermann, A., Aitken, C. (2015) Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. Law, Probability and Risk. IN PRESS. doi:10.1093/lpr/mgv008

References

[1] Abraham, J., Champod, C., Lennard, C., and Roux, C. (2013). Spatial analysis of corresponding fingerprint features from match and close non-match populations. *Forensic Science International*, 230.

[2] Aitken, C. and Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley and Sons, 2nd edition.

[3] Alberink, I., Jongh, A., and Rodriguez, C. (2014). Fingermark evidence evaluation based on automated fingerprint identification system matching scores: The effect of different types of conditioning on likelihood ratios. *Journal of Forensic Sciences*, 59(1):70–81.

[4] Armstrong, D., Neumann, C., Saunders, C., Gantz, D., Miller, J., and Stoney, D. (2016). Kernel-based methods for source identification using very small particles from carpet fibers. *Chemometrics and Intelligent Laboratory Systems*. In Press.

[5] Baiker, M., Keereweer, I., Pieterman, R., Vermeij, E., van der Weerd, J., and Zoon, P. (2014). Quantitative comparison of striated toolmarks. *Forensic Science International*, 242:186–199.

[6] Berger, C. (2013). Objective ink color comparison through image processing and machine learning. *Science & Justice*, 53(1):55–59.

[7] Boender, G. J., Hagenaars, T. J., Bouma, A., Nodelijk, G., Elbers, A. R. W., de Jong, M. C. M., and van Boven, M. (2007). Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Computational Biology*, 3(4).

[8] Bolck, A., Weyermann, C., Dujourdy, L., Esseiva, P., and van den Berg, J. (2009). Different likelihood ratio approaches to evaluate the strength of evidence of mdma tablet comparisons. *Forensic Science International*, 191.

[9] Bozza, S., Taroni, F., Marquis, R., and Schmittbuhl, M. (2008). Probabilistic evaluation of handwriting evidence: likelihood ratio for authorship. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3):329–341.

[10] Brummer, N. and Albert, S. (2014). Bayesian calibration for forensic evidence reporting. *arXiv preprint*.

[11] Champod, C. and Meuwly, D. (2000). The inference of identity in forensic speaker recognition. *Speech Communication*, 31:193–203.

[12] Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.

[13] Chumbley, S. L. (2010). Quantification of toolmarks. Technical report, U.S. Department of Justice.

[14] Cole, S. A. (2005). More than zero: Accounting for error in latent fingerprint identification. 95(3).

[15] Conrad, K. (2014). The minimal polynomial and some applications.

[16] Davis, L., Saunders, C., Hepler, A., and Buscaglia, J. (2012). Using subsampling to estimate the strength of handwriting evidence via score-based likelihood ratios. *Forensic Science International*, 216(1-3):146–157.

[17] de Lázaro, J. M. B., Moreno, A. P., Santiago, O. L., and da Silva Neto, A. J. (2015). Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers & Industrial Engineering*, 87:140–149.

[18] Egli, N. and Champod, C. (2014). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems: modeling between finger variability. *Forensic Science International*, 235.

[19] Egli, N., Champod, C., and Margot, P. (2007). Evidence evaluation in fingerprint comparison and automated fingerprint identification systems-modelling within finger variability. *Forensic Science International*, 167(2-3):189–195.

[20] Evett, I. (1977). The interpretation of refractive index measurements. 9:209–217.

[21] Evett, I. (1983). What is the probability that this blood came from that person? a meaningful question? *Journal of Forensic Science Society*, 23(1):35–39.

[22] Evett, I. (1986). A bayesian approach to the problem of interpreting glass evidence in forensic science casework. *Journal of Forensic Science Society*, 26:3–18.

[23] Finkelstein, M. and Fairley, W. (1970). A bayesian approach to identification evidence. *Harvard Law Review*, 83(3):489–517.

[24] Gantz, D. and Saunders, C. (2014). Quantifying the effects of database size and sample quality on measures of individualization validity and accuracy in forensics.

[25] Genton, M. (2001). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research 2*, pages 299–312.

[26] Giguere, S., Laviolette, F., Marchand, M., Tremblay, D., Moineau, S., Liang, X., Biron, E., and Corbeil, J. (2015). Machine learning assisted design of highly active peptides for drug discovery. *PLOS Computational Biology*, 11(4).

[27] Gonzalez-Rodriguez, J., Drygajlo, A., Ramos, D., Garcia, M., and Ortega, J. (2006). Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. *Computer Speech & Language*, 20(2-3):331–355.

[28] Gonzalez-Rodriguez, J. and Ramos, D. (2007). *Speaker Classification I*, volume 4343, chapter Forensic automatic speaker classification in the "Coming Paradigm Shift", pages 205–217. Springer.

[29] Good, I. (1950). *Probability and the Weighting of Evidence*. Charles Griffin, 1st edition.

[30] Hepler, A., Saunders, C., Davis, L., and Buscaglia, J. (2012). Score-based likelihood ratios for handwriting evidence. *Forensic Science International*, 219(1):129–140.

[31] Hofmann, T., Scholkopf, B., and Smola, A. J. (2008). A review of kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220.

[32] Horn, R. (1985). *Matrix Analysis*. Cambridge University Press.

[33] Izenman, A. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer.

[34] Kirk, P. L. (1963). The ontogeny of criminalistics. 54:235–238.

[35] Kwan, Q. Y. (1977). *Inference of Identity of Source*. PhD thesis.

[36] Lesaffre, E. and Lawson, A. B. (2012). *Bayesian Biostatistics*. Wiley.

[37] Lindley, D. (1977). A problem in forensic science. *Biometrika*, 64(2):207–213.

[38] Liu, Y., Zhang, Z., and Chen, J. (2015). Ensemble local kernel learning for online prediction of distributed product outputsin chemical processes. *Chemical Engineering Science*, 137:140–151.

[39] Lock, A. and Morris, M. (2013). Significance of angle in the statistical comparison of forensic tool marks. *Technometrics*, 55(4):548–561.

[40] Mnookin, J. L. (2003). Fingerprints: Not a gold standard.

[41] Morrison, G. and Kinoshita, Y. (2008). Automatic-type calibration of traditionally derived likelihood ratios: Forensic analysis of australian english /o/ formant trajectories. Interspeech: Special Session: Forensic Speaker Recognition - Traditional and Autmoatic Approaches.

[42] National, R. C. (2009). *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, D.C.

[43] Neumann, C. (2009). New perspectives in the use of ink evidence in forensic science part iii: Operational applications and evaluation. *Forensic Science International*, 192(1-3):29–42.

[44] Neumann, C., Armstrong, D., and Wu, T. (2016). Determination of afis "sufficiency" in friction ridge examination. (263):114–125.

[45] Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., and Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Science*, 52(1):54–64.

[46] Neumann, C., Evett, I., and Skerrett, J. (2012). Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2):371–415.

[47] Neumann, C. and Saunders, C. (2014). Commentary on: Alberink i, de jongh a, rodriguez c. fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. *Journal of Forensic Sciences*, 59(1):70–81.

[48] Neumann, C. and Saunders, C. (2015). Commentary on: Alberink i, de jongh a, rodriguez c. fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. j forensic sci 2014; 59(1):70-81. 60(1).

[49] Ommen, D., Saunders, C., and Neumann, C. (2017). The characterization of monte carlo errors for the quantification of the value of forensic evidence. 87:1608–1643.

[50] Parker, J. (1966). A statistical treatment of identification problems. 6:33–39.

[51] Parker, J. (1967). The mathematical evaluation of numerical evidence. 7:134–144.

[52] Radu T. Ionescu, M. P. (2015). Pq kernel: A rank correlation kernel for visual word histograms. *Pattern Recognition Letters*, 55:51–57.

[53] Ramos, D. and Gonzalez-Rodrigues, J. (2013). Reliable support: Measuring calibration of likelihood ratios. *Forensic Science International*, 230(1-3):159–169.

[54] Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.

[55] Ramsay, J., Wickham, H., Graves, S., and Hooker, G. *Package 'fda'*.

[56] Rasmussen, C. and Williams, C. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

[57] Risinger, D. M., Denbeaux, M. P., and Saks, M. L. (1989). Exorcism of ignornace as a proxy for rational knowledge: The lessons of handwriting identification 'expertise'. 137(731).

[58] Riva, F. and Champod, C. (2014). Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences*, 59(3):637–647.

[59] Saks, M. (1998). Merlin and solomon: Lessons from the law's formative encounters with forensic identification science. *Hastings Law Journal*, 49:1069–1141.

[60] Saks, M. (2003). The legal and scientific evaluation of forensic science (especially fingerprint expert testimony). *Setton Hall Law Review*, 33:1167–1187.

[61] Saks, M. and Koehler, J. (2005). The coming paradigm shift in forensic identification science. *Science*, 309:892–895.

[62] Saks, M. and Koehler, J. (2008). The individualization fallacy in forensic science evidence. *Vanderbilt Law Review*, 61(1):199–219.

[63] Saunders, C. (2013). Understanding individuality of handwriting using score-based likelihood ratios. Measurement Science and Standards in Forensic Handwriting Analysis Conference, National Institute of Standards and Technology.

[64] Saunders, C. and Miller, J. (2013). On desiderata for score-based likelihood ratios for forensic evidence. Joint Statistical Meetings. Topic Contributed.

[65] Saunders, C. and Ommen, D. (2014). Computational and statistical aspects of the forensic identification problem. Leiden, Netherlands. International Conference on Forensic Inference and Statistics.

[66] Schoenberg, I. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, 39(4):811–841.

[67] Shawe-Taylor, J. and Cristianini, N. (2012). *Kernel Methods for Pattern Analysis*. Cambridge University Press, 6 edition.

[68] Stoney, D. A., Neumann, C., Mooney, K. E., Hyatt, J. M., and Stoney, P. L. (2015). Exploitation of very small particles to enhance the probative value of carpet fibers. *Forensic Science International*, 252:52–68.

[69] Tang, Y. and Srihari, S. (2014). Likelihood ratio estimation in forensic identification using similarity and rarity. *Pattern Recognition*, 47:945–958.

[70] Taroni, F., Champod, C., and Margot, P. (1998). Forerunners of bayesianism in early forensic science. 38:183–200.

[71] van der Vaart, A. (2007). *Asymptotic Statistics*. Cambridge University Press.

[72] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

[73] Vapnik, V. N. (1998). *Statistical Learning Theory*. JOHN WILEY & SONS INC.