

Regularized Rao–Blackwellization

An Extension of a Classical Technique
with Applications to Gibbs Point Process Statistics

DISSERTATION

FOR THE AWARD OF THE DEGREE

Doctor rerum naturalium
(Dr. rer. nat.)

OF THE
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
WITHIN THE
DOCTORAL PROGRAM

Mathematical Sciences

OF THE
GEORG-AUGUST-UNIVERSITY SCHOOL OF SCIENCE
(GAUSS)

SUBMITTED BY

Henning Höllwarth
from Malchin

GÖTTINGEN, 2020

THESIS COMMITTEE

Prof. Dr. Dominic Schuhmacher
Institute for Mathematical Stochastics
University of Göttingen

Prof. Dr. Stephan Huckemann
Institute for Mathematical Stochastics
University of Göttingen

MEMBERS OF THE EXAMINATION BOARD

REVIEWER

Prof. Dr. Dominic Schuhmacher
Institute for Mathematical Stochastics
University of Göttingen

SECOND REVIEWER

Prof. Dr. Lutz Mattner
Professorship for Mathematical Stochastics
University of Trier

FURTHER MEMBERS OF THE EXAMINATION BOARD

Prof. Dr. Thorsten Hohage
Institute for Numerical and Applied Mathematics
University of Göttingen

Prof. Dr. Stephan Huckemann
Institute for Mathematical Stochastics
University of Göttingen

Jun.-Prof. Dr. Daniel Rudolf
Institute for Mathematical Stochastics
University of Göttingen

Prof. Dr. Anja Sturm
Institute for Mathematical Stochastics
University of Göttingen

DATE OF THE ORAL EXAMINATION

December 3, 2020

Preface

How can the concept of Rao–Blackwellization deal with misspecification issues in statistical models?

This question links two apparently quite different areas of statistics. The first, the so-called *Rao–Blackwellization*, concerns the concept of data reduction to the sufficient information. It refers to the classical theory of mathematical statistics and may rather appeal to theoretically minded statisticians. The second, namely misspecification issues of statistical models, is of practical concern and deals with building and justifying statistical models for complex dependent data.

Statisticians with some background in classical statistical theory might simply reply to the initially stated question as follows: Rao–Blackwellizations are not affected by the model misspecification, since the data reduction to sufficient information even has to be independent of the parameter of interest. The question, however, aims at a different aspect. Due to the misspecification, parameter values can be interpreted less reliably. Hence, not every source of information, even within the so-called minimal sufficient information, has a reasonable interpretation. In other words, there is still some kind of data which does not help toward making inference on the parameter. The present thesis addresses this issue and consequently aims at a generalized Rao–Blackwellization concept through regularizing an ill-posed inverse problem.

This thesis was written in the inspiring and motivating atmosphere of the Institute for Mathematical Stochastics (IMS) at the University of Göttingen. Many people have accompanied and supported me in writing this thesis, either by sharing their knowledge, by providing useful feedback, or by encouraging me when needed. I shall be more specific about this, with apologies to those I neglect to mention explicitly.

To begin with, I would like to thank my first supervisor Prof. Dr. Dominic Schuhmacher for introducing me to the field of Gibbs point process statistics. I am most grateful for the time and effort he spent in numerous

discussions and also for many stimulating suggestions, comments, and remarks. With his support I have been able to complete this thesis.

My thanks go also to my second supervisor Prof. Dr. Stephan Huckemann for providing full and kind support for all of my intentions in this PhD project.

I would like to express my gratitude to Prof. Dr. Lutz Mattner for fruitful discussions, comments, and suggestions. I am very appreciative of invaluable literature references about classical mathematical statistics he gave me. Especially, he called my attention to the conference papers of Schmetterer (1977, 1978) which indicate an early attempt to benefit from regularizing the ill-posed problem that is inherent in the theory of unbiased estimation. I am very pleased about his commitment for conducting the second review.

Furthermore, I am very grateful to Jun.-Prof. Dr. Daniel Rudolf for his encouragement and kind support, in particular for useful feedback on early drafts of different parts. All of this was frequently accompanied by inspiring discussions on a broad range of topics, which I always enjoyed.

I also want to thank Prof. Dr. Thorsten Hohage and Prof. Dr. Frank Werner for valuable conversations on the theory of regularization of ill-posed problems and for helpful remarks and hints on the literature.

For the simulation study in the present thesis, I have greatly benefited from unpublished R codes which were kindly provided by Prof. Dr. Adrian Baddeley (implementation of the variational estimator, see also Baddeley & Dereudre (2013)) and by Philipp Möller and Prof. Dr. Dominic Schuhmacher (implementation of the maximum likelihood estimator).

I am indebted to Christian Böhm and Dr. Carsten Gottschlich for tackling and solving technical difficulties immediately and in a very straightforward and uncomplicated way.

Many thanks go to all of my (former) colleagues from the IMS for providing an encouraging, supporting, and friendly working place. In particular, I would like to mention Dr. Merle Behr, Dr. Anne Hobert, Dr. Claudia König, Dr. Fabian Kück, Raoul Müller, Viacheslav Natarovskii, Ass.-Prof. Dr. Katharina Proksch, Dr. Robin Richter, Dr. Laura Fee Schneider, Marco Seiler, Dr. Max Sommerfeld, Jun.-Prof. Dr. Björn Sprungk, Dr. Carla Taming, and Johannes Wieditz.

My absorption in the work on this thesis has also demanded much patience and tolerance from my family. I owe them a very special thank you for their faithful and consistent moral support.

Göttingen, October 2020

Henning Höllwarth

Contents

Abstract	vii
Introduction	ix
I General Statistics	1
1 Fundamentals of Statistics	3
1.1 Basic Notions and Concepts	3
1.2 Sufficiency, Ancillarity, and Completeness	12
1.3 Exponential Families	26
2 Estimation Procedures	29
2.1 Implicitly Defined Estimators	29
2.2 Evaluation of Estimation Procedures	35
2.3 Optimal Estimation Procedures	38
II Regularized Rao–Blackwellization	43
3 The Rao–Blackwell Inverse Problem	49
3.1 Lehmann–Scheffé Topologies	51
3.2 Lehmann–Scheffé Spaces	61
4 Regularized Rao–Blackwellization	69
4.1 Tikhonov’s Regularization Approach	71
4.2 Ivanov’s Regularization Approach	87
III Statistics for Gibbs Point Processes	99
5 Basics on Random Measures	101
5.1 The State Spaces	102

5.2	Random Measures	109
5.3	Intensity of Random Measures	118
5.4	Conditional Point Processes	119
6	Gibbs Point Processes	125
6.1	Toward a Definition of Gibbs Point Processes	126
6.2	First Simple Examples and Results	129
6.3	Gibbs Specifications	141
6.4	Existence of Gibbs Point Processes	151
7	Statistics for Parametric Gibbs Point Process Models	157
7.1	Maximum Likelihood Estimation	158
7.2	Maximum Pseudo-Likelihood Estimation	160
7.3	Variational Estimation	165
7.4	Simulation Study	167
	Discussion and Outlook	187
	Abbreviations and Notations	193
	Figures and Tables	199
	References	201
	Index	211

Abstract

Parametric specifications of statistical models often fail to describe complex dependent data precisely. Inference then cannot go beyond that inaccuracy. Consequently, a statistical procedure that maps the observed data to a statistical statement is required to be stable in the sense that small errors in the parametric specification can be taken into account by at most small changes of the statistical procedure.

The popular likelihood based statistical analysis of parametrized statistical models is, in fact, stable against such misspecifications (invariance property of the maximum likelihood (ML) estimator). However, likelihood approaches often suffer from an intractable normalizing constant. ML estimates and likelihood based confidence sets are hence not known explicitly and even numerically they are hard to compute, if at all.

Therefore, alternative estimation procedures arose, which are simple to compute, however, less efficient. When complete sufficient information is known, Rao–Blackwellizations of these estimators gain central importance. Apparently, however, neither their stability issues were studied nor is it known how to compute Rao–Blackwellizations analytically or numerically. The situation is therefore still quite unsatisfactory, especially for statistical disciplines that deal with complex dependent data such as point patterns.

The present thesis concerns Rao–Blackwellizations in misspecified models. It is demonstrated that this elegant technique to improve estimators fails to be stable. To put it differently, it is shown that the classical Rao–Blackwellization is an ill-posed inverse problem. Regularizations are proposed and, consequently, the concept of *regularized Rao–Blackwellization* is introduced. In classical examples, this leads to new estimators and new interpretations of existing ones. For more complex examples like several ones in Gibbs point processes statistics, regularized Rao–Blackwellizations can be computed at least approximately. A simulation study where we consider the Lennard-Jones model demonstrates the computational feasibility and the benefit from these results, especially in constructing parametric bootstrap confidence regions on the basis of the maximum likelihood estimator.

Introduction

Statistical analysis of complex dependent data meets the difficulty of building and dealing with an appropriate statistical model. An exponential family, for example, may be a suitable modeling approach, which enables a flexible and intuitive description of the *principal* components of the underlying dependence structure. Remaining components in turn have often to be neglected to keep things tractable or due to a lack of knowledge of the dependence mechanism. Consequently, statistical analysis has then to cope with model misspecification, even with slight ones.

To be more explicit, consider some parameter set $\Theta \subset \mathbb{R}^k$ and recall that an exponential family is a Θ -indexed family of distributions $(P_\vartheta)_{\vartheta \in \Theta}$ that is dominated by a common σ -finite reference measure μ on a measurable space $(\mathfrak{X}, \mathcal{A})$. It is assumed that each P_ϑ has a probability density w. r. t. μ which is of the form

$$f_\vartheta(x) \propto \exp(\langle \vartheta, S(x) \rangle) \quad \text{for } x \in \mathfrak{X},$$

where $S : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is a statistic that summarizes the main features in the raw data $x \in \mathfrak{X}$.

Within this context, we then believe that P_{ϑ_0} for some $\vartheta_0 \in \Theta$ describes the stochastic law that we are looking for, may be not exactly, but up to a reasonable approximation in a certain sense. We additionally assume that an approximation can be arbitrarily close to the stochastic law under investigation by increasing the amount of parameters, that is, by increasing k . Furthermore, interpretations of P_{ϑ_0} through ϑ_0 require in some cases a transformation of ϑ_0 , that is, one considers $\kappa(\vartheta_0) \in \mathbb{R}^l$, where $\kappa : \Theta \rightarrow \mathbb{R}^l$ for some $l \in \mathbb{N}$ is a known parameter function of interest.

The Problem

Given an observation $x \in \mathfrak{X}$, we are looking for *accurate statements* about the value of interest $\kappa(\vartheta_0)$ which are also *stable* against misspecifications.

To make that clearer, note that a statement might be given, for example, in terms of a point estimate $\hat{\kappa}(x) \in \kappa(\Theta)$. For a precision statement

as well, a (parameter) set estimate¹ $\widehat{K}_{1-\alpha}(x) \subset \kappa(\Theta)$ might be of interest, where $\widehat{K}_{1-\alpha}$ is required to cover $\kappa(\vartheta_0)$ under P_{ϑ_0} with a preassigned probability $1 - \alpha$. We call $\widehat{\kappa}$ and $\widehat{K}_{1-\alpha}$ *estimator* and $(1 - \alpha)$ -*confidence region*, respectively, and in general any function that maps the data to a statistical statement about the parameter a *(statistical) procedure*.

Of course, we aim at extracting information about the stochastic mechanism behind the data x as much as possible. Regarding an estimator $\widehat{\kappa}$, this means we want the distribution of $\widehat{\kappa}$ under P_{ϑ} to be as close to δ_{ϑ} (for $\vartheta \in \Theta$) as possible. For a confidence region $\widehat{K}_{1-\alpha}$, we correspondingly require that the distribution of $\widehat{K}_{1-\alpha}$ under P_{ϑ} is as close to $\delta_{\{\vartheta\}}$ (for $\vartheta \in \Theta$) as possible. According to the assumed statistical model, information that cannot be described by S does not help toward making inference on the parameter. Hence, we at least look for $\sigma(S)$ -measurable procedures.

Furthermore, the assumed statistical model usually does not account for every source of dependence that underlies the data. The construction method which yields the statistical procedures, such as estimators $\widehat{\kappa}$ or confidence sets $\widehat{K}_{1-\alpha}$, should therefore be stable against such model misspecifications. In other words, small errors in the specification should then be able to be accounted for by small changes of $\widehat{\kappa}$ and $\widehat{K}_{1-\alpha}$, respectively.

Statistical procedures based on the likelihood $\vartheta \mapsto f_{\vartheta}(x)$, especially maximum likelihood estimation, often have this stability property and can often be deemed an accurate procedure. However, the required normalizing constant $c(\vartheta)^{-1} := \int \exp(\langle \vartheta, S(x) \rangle) \mu(dx)$ for $\vartheta \in \Theta$ is in many cases analytically and numerically intractable. This raises the following questions which are still of current research.

Questions How to compute maximum likelihood estimates efficiently and how to construct confidence regions based on the likelihood? ♦

One may also consider the conditional expectation of some inefficient estimator $\widehat{\lambda}$ given S under P_{ϑ} , which is in fact independent of ϑ . This is the so-called Rao–Blackwellization $\mathbf{E}(\widehat{\lambda}|S)$ that leads to the questions:

Questions How to compute the Rao–Blackwellization $\mathbf{E}(\widehat{\lambda}|S)$ at least approximately and how can the concept of Rao–Blackwellization deal with misspecifications? ♦

All of these questions finally aim at getting *accurate* and *stable* statistical procedures. However, the first two seem to have been studied more than the latter two.

Background and State of the Art

For more than forty years, the challenging obstacle of intractable likelihoods has proved to be quite stimulating for statistics. It led to so-

¹Throughout this thesis we write $A \subset B$ iff A is a subset of B , that is, $x \in B$ for all $x \in A$. We are less often faced with *proper* subsets which will be denoted by \subsetneq .

phisticated tools for computing maximum likelihood estimates at least approximately by Markov chain Monte Carlo approaches (see Penttinen, 1984; Geyer & Thompson, 1992). Additionally, several computationally efficient alternative estimation methods were developed, analyzed, and discussed, for example, composite likelihood (Besag, 1974, 1975; Lindsay, 1988) and variational estimation procedures (Almeida & Gidas, 1993). All of this was motivated by a broad range of modern statistical problems. With regard to Varin *et al.* (2011), Larribe & Fearnhead (2011), Almeida & Gidas (1993), and the references therein, image analysis, statistical genetics, and speech recognition may be mentioned as notable applications that gave rise to these developments due to their complex data structures.

Regarding the present thesis, specific attention is paid to Gibbs point process statistics. This well-established subject made in fact extensive use of the above-mentioned developments. Composite likelihood, including pseudo-likelihood, and variational methods were proposed and analyzed with respect to their asymptotic properties and their computational efficient implementation, see for example Ripley (1988); Møller & Waagepetersen (2004); Coeurjolly & Rubak (2013); Baddeley *et al.* (2014) and Baddeley & Dereudre (2013); Coeurjolly & Møller (2014). In contrast, not much is known about the MLE regarding asymptotic normality. Strong consistency of the MLE was shown only recently by Dereudre & Lavancier (2017). However, based on simulation studies (see for example Diggle *et al.*, 1994; Baddeley & Dereudre, 2013) and due to asymptotic results of Mase (1992) for the ML estimation in a restrictive class of Gibbs point processes (see also the argumentation in Dereudre & Lavancier (2017)), there is the common belief that the MLE is, at least asymptotically, more efficient than its surrogates.

Non-asymptotic, that is, fixed sample size results were developed within the theory of unbiased estimating functions. A pioneering paper is Godambe (1960), where it is proved that under some regularity properties the MLE can be derived from an *optimal* unbiased estimating function, the so-called score, where optimality refers to an intuitive efficiency criterion. Due to the intractabilities of the score, one then looks for optimal estimating functions in a certain subclass (see also e. g. Heyde, 1997). Small & McLeish (1988) extended notions like sufficiency, ancillarity, and completeness to estimating functions and proved a corresponding Rao–Blackwell–Lehmann–Scheffé theorem regarding an efficiency criterion which is closely related to that of Godambe, see also Small & McLeish (1994). With respect to point process statistics, we may mention Guan *et al.* (2015), who proposed optimal estimating functions within a certain class. Concerning specifically Gibbs point process models, Coeurjolly *et al.* (2016) proposed at least a “semi-optimal” choice within the class of so-called Takacs–Fiksel estimating functions. This can lead to a more efficient estimating procedure than the well-known

maximum pseudo-likelihood method (e. g. Møller & Waagepetersen, 2004). However, Coeurjolly *et al.* (2016) finally concluded that their approach is computationally demanding and that further improvements are required.

The above-mentioned stability aspect especially with regard to model misspecification appears to be hardly studied. While this may in fact be less relevant for the maximum likelihood estimator, it becomes of more relevance if we deal with Rao–Blackwellizations of alternative estimation methods.

Main Contributions of the Present Thesis

The main contributions of the present thesis belong to non-asymptotic statistics with applications particularly to exponential families.

First, a new generalized concept of Rao–Blackwellization is provided through regularization techniques. This is preceded by an appropriate introduction of the Rao–Blackwell projection and the demonstration that this projection yields the best approximation to the initial estimator w. r. t. “ancillary information” (Proposition 1.34, page 25). To examine the amount of ancillarity (of first order), topological and metrical structures are introduced and analyzed which then lead to an inverse problem that is usually ill-posed (see Chapter 3). We introduce the concept of *regularized Rao–Blackwellization* (see Chapter 4) and, consequently, propose a corresponding *regularized* version of the Rao–Blackwell–Lehmann–Scheffé theorem (see Theorem 4.12, page 86). Furthermore, we get new estimators and also new interpretations of already existing estimators (see for example Example 4.5, page 74 or Example 4.8, page 81).

Second, a new method for computing an approximate *distribution* of the MLE under P_{θ_0} is given that is closely related to the regularized Rao–Blackwellization idea (see Proposition 4.20, page 95). The method is straightforward to implement using standard software that fits linear models. It is also computationally much faster than simulating from the distribution of the MLE (see also Algorithm 4.21). Furthermore, we use our method to propose heuristic construction procedures of confidence regions. Finally, the performance of these results will be demonstrated by simulation studies considering specifically the situation of Gibbs point process statistics (see Part III, Chapter 7, page 157 ff.).

The main contributions are presented in more detail in the introduction to Part II on page 43 ff.

Spatial Distribution of Envelope Spike Proteins

What are the above-mentioned results good for? In order to give an illustrative answer, we take a brief look at an application. Specifically, under the current prevalent of the SARS-CoV-2 pandemic, we consider

some questions in virology, concerning the spatial distribution of so-called envelope spike proteins (see e. g. Zhu *et al.* (2006) and Klein & Bjorkman (2010)). Since the present illustration is intended for the mathematical community, the biological viewpoint is simplified to a minimum. For details references are provided.

First, it should be noted that many viruses are enveloped, for example, Corona-, Herpes-, human immunodeficiency (HI), influenza, and also simian immunodeficiency (SI) viruses, to mention just a few. The envelope consists of a membrane where so-called spike proteins are embedded and which enclose the viral genome (see Figure A for a simplified scheme, see for example Mateu (2013) for detailed descriptions).

Second, since these envelope spike proteins may have several functions with regard to the infection of host cells, they constitute a central object of study for understanding the behavior of a virus and, consequently, for developing vaccines. Zhu *et al.* (2006, page 847 f.) studied the distribution of envelope spikes of HI and SI viruses and concluded that “[envelope] spikes are not free to diffuse in the plane of the membrane” and, apparently, are not arranged through “a geometrically arrayed set of matrix pores”. Furthermore, Klein & Bjorkman (2010, page 1) argued that the spatial distribution of the envelope spike proteins is one strategy of

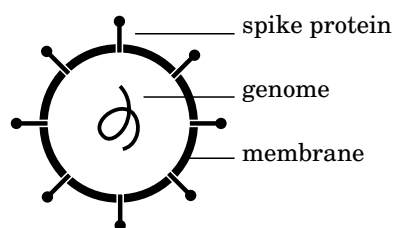


Figure A: Schematic profile of an enveloped virus. The regular arrangement of the spike proteins is oversimplified.

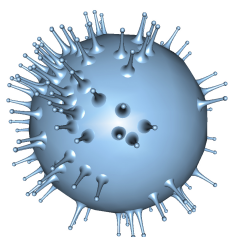


Figure B: Realization of spikes distributed according to a homogeneous Poisson point process on a unit sphere with intensity parameter $\beta = 130$. (Complete spatial randomness)

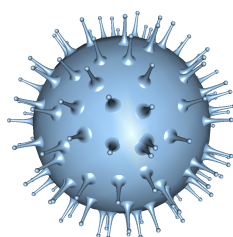


Figure C: Realization of spikes distributed according to a Strauss point process on a unit sphere with interaction range $R = 0.25$, interaction parameter $\gamma = 0.05$ and intensity parameter $\beta = 5000$. (Repulsion between spikes)

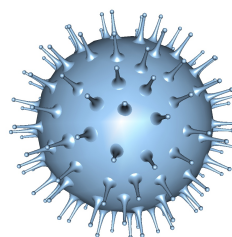


Figure D: Realization of spikes distributed according to a (12,6)-Lennard-Jones point process on a unit sphere with intensity parameter $\beta = 100$ and interaction parameters $\sigma = 0.25$ and $\epsilon = 10$. (Repulsion and attraction between spikes)

the HI virus to evade antibodies successfully. Through mutations, the HI virus can adapt and change the spatial distribution of spike proteins to make use of “an inherent limitation [of] the architecture of an antibody.”

These observations and hypotheses raise several statistical issues. What are suitable statistical models for different kinds of viruses? How to fit these models to the data, that is, how to estimate such statistical models and how to evaluate the precision of these estimates? Also, how to decide if differences in observations of viruses are significant regarding their spatial arrangement of proteins?

Zhu *et al.* (2006) proposed a first, quite elementary statistical model for a certain HI and SI virus. However, since some summary statistics of synthetic data from the fitted model differ from those of the observed data, the proposed model seems to be overly simplistic.

The next step is, therefore, to consider more flexible statistical models for enveloped viruses. Various Gibbs point process models on the sphere meet this requirement. For illustrations, Figures B, C, and D show realizations of spike arrangements according to a Poisson, a Strauss, and a Lennard-Jones point process on the unit sphere, respectively (see also Chapter 6). Statistical inference within such models can be done by the proposed methods in this thesis.

Organization of the Present Thesis

This thesis is divided into three parts. The first part reviews fundamental notions of mathematical statistics, in particular concepts such as sufficiency, ancillarity, and completeness. Essential results are provided – some of them seem to be unknown in the literature – which prepare the subsequent Part II. The latter contains the main contributions of this thesis about the study of the Rao–Blackwellization as an ill-posed inverse problem. Standard examples from classical statistics accompany the presentation in this part, while some of the methods are separately examined for Gibbs point process statistics in Part III. This last part begins with a fundamental account on point process theory emphasizing some peculiarities. After some theory on Gibbs point processes, three established estimation methods for Gibbs point process models are presented with some regard to measurability aspects. This part concludes with a simulation study that illustrates the proposed methods.

More details, especially about the minor and major contributions of this thesis are given in separate introductions to Part I, II, and III.

PART I

General Statistics

This part provides the required mathematical account on statistics and some preliminaries that seems to be unknown in the literature. Starting with a brief decision theoretical introduction, we proceed with the principle notions of sufficiency, ancillarity, and completeness. We demonstrate a close relationship to one of the central objects of this thesis, the *Rao–Blackwell projection* (see page 16 ff.). To study and to benefit from the interaction of statistical and topological-algebraic structures, we propose a functional analytic treatment (see Theorem 1.27, see also Proposition 1.29, 1.30, and Corollary 1.31) which seems to be unknown in the literature and leads in particular to the following statement: the Rao–Blackwell projection yields the best approximation w. r. t. “first-order ancillarity information” (see Proposition 1.34, page 25), an observation which will serve as a starting point for the main contributions of this thesis in Part II.¹ The first chapter is then concluded with a review of basic properties of exponential families, which constitute a rich class of statistical models that is popular in theory and practice and to which the findings of the present thesis can be applied.

Two fundamental decision rules will be of greater interest in this thesis, namely, estimation and confidence procedures. A closer look at estimators, which are often introduced implicitly through estimating equations, is taken in Chapter 2. We introduce the notion of a *stable* estimator construction principle. Confidence regions that are constructed based on the distribution of estimators are considered later (see Section 7.4 in Part III.)

The rather technical style is tailored to what we require in the main contribution. Apart from Section 1.2 where larger parts cannot be found in the literature, the style and the chosen content are mainly guided by Barra (1981), Torgersen (1991), Small & McLeish (1994), Heyde (1997),

¹In Part II we especially introduce norms that measure the “first-order ancillarity information”. Consequently, Proposition 1.34 will then be generalized according to these distances, see Theorem 4.1, page 70.

Song (2007), Godambe & Heyde (2010), and Gaetan & Guyon (2010). For some aspects concerning well-known classical statistics, we refer to books like Witting (1985), Strasser (1985), Pfanzagl (1994), or Lehmann & Casella (1998).

Basically, proofs are given whenever they contribute to a comprehensive presentation or if there is no easy reference. Otherwise, we frequently refer to the mentioned textbooks.

CHAPTER 1

Fundamentals of Statistics

1.1 Basic Notions and Concepts

Statistical investigation consists of making statements about a random observation. To this end, we first introduce fundamental definitions and notions that are used throughout this thesis.

Statistical Spaces

Let all possible outcomes of a random observation of interest be given by a measurable space $(\mathfrak{X}, \mathcal{A})$, which is called *sample space*. A *random sample* is a $(\mathfrak{X}, \mathcal{A})$ -valued random element X on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

To make statements about the distribution of X , prior assumptions and information on¹ \mathbf{P}^X is summarized by a non-empty class of distributions $\mathfrak{P} \subset \text{Prob}(\mathfrak{X}, \mathcal{A})$, where $\text{Prob}(\mathfrak{X}, \mathcal{A})$ denotes the set of all probability measures on $(\mathfrak{X}, \mathcal{A})$.

1.1 Definition A non-empty class of distributions $\mathfrak{P} \subset \text{Prob}(\mathfrak{X}, \mathcal{A})$ is called *statistical model* on the sample space $(\mathfrak{X}, \mathcal{A})$. Furthermore, the triple $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is called *statistical space*. \blacklozenge

A probability space $(\mathfrak{X}, \mathcal{A}, P)$ is a *trivial* example of a statistical space, where the statistical model is $\{P\}$, that is, it contains exactly one element. While probability theory is concerned with statements about a known distribution $P \in \{P\}$, statistical theory is concerned with statements about an unknown distribution $\mathbf{P}^X \in \mathfrak{P}$, where \mathfrak{P} is non-trivial, that is, $|\mathfrak{P}| > 1$.

A map κ from \mathfrak{P} to some space \mathfrak{R} containing characteristics of interests is called \mathfrak{R} -valued *parameter* and $\Theta := \kappa(\mathfrak{P})$ is called κ -*parameter space*

¹We denote the distribution of a random element X under a probability measure P by \mathbf{P}^X , $\mathbf{P}X^{-1}$ or $\mathbf{P} \circ X^{-1}$. More generally, the image measures of a measure μ under a function f is denoted μ^f , μf^{-1} or $\mu \circ f^{-1}$.

of \mathfrak{P} . That parameter κ is said to *parametrizes* \mathfrak{P} iff² κ is injective, that is, there is a map $P : \Theta \rightarrow \mathfrak{P}$ such that $P \circ \kappa = \mathbf{id}_{\mathfrak{P}}$. The map $(P_{\vartheta})_{\vartheta \in \Theta}$ is then called a *parametrization* of \mathfrak{P} or *parametrized statistical model*. Any statistical model \mathfrak{P} can be parametrized by $\kappa := \mathbf{id}_{\mathfrak{P}}$. However, if \mathfrak{P} can be parametrized by an \mathbb{R}^k -valued parameter then \mathfrak{P} is said to be *k-parametric*. Further, we call \mathfrak{P} *non-parametric* if there is no $k \in \mathbb{N}$ and no \mathbb{R}^k -valued parameter that parametrizes \mathfrak{P} .

In what follows we will often write $\mathfrak{P} = \{P_{\vartheta} : \vartheta \in \Theta\}$ to denote the (unparametrized) statistical model $P(\Theta)$ and, simultaneously, to indicate the corresponding parametrized statistical model $(P_{\vartheta})_{\vartheta \in \Theta}$. If no clarification is needed, we write \mathfrak{P} even if the parametrized statistical model is meant.

As already indicated above, a fundamental assumption in statistical theory is that the unknown distribution \mathbf{P}^X that underlies the observation of the investigator is contained in the model \mathfrak{P} . In most cases, however, especially parametric specifications result from simplified assumptions about the random mechanism. Consequently, we may usually assume at most that \mathbf{P}^X can be “well” approximated within \mathfrak{P} , since $\mathbf{P}^X \notin \mathfrak{P}$. The statistical model is then called *misspecified* or *incorrectly specified*.

In a statistical space $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, statements concerning random elements, for example, can usually be made at most in the almost sure sense with respect to every $P \in \mathfrak{P}$. A set $N \in \mathcal{A}$ with $P(N) = 0$ for all $P \in \mathfrak{P}$ is called *\mathfrak{P} -null set*. Correspondingly, a statement about the elements of \mathfrak{X} characterized by an event $A \in \mathcal{A}$ holds *\mathfrak{P} -almost surely* iff A^c is a \mathfrak{P} -null set.

A statistical model \mathfrak{P} is called *dominated* iff there is a σ -finite measure μ on $(\mathfrak{X}, \mathcal{A})$ such that every μ -null set is a \mathfrak{P} -null set. We will then write $\mathfrak{P} \ll \mu$. Dominated models constitute an important class of statistical models. A first pleasant consequence is that we may deal with probability densities rather than probability distributions. Second, dominated models exhibit an important technical property, which is, if \mathfrak{P} is dominated, then \mathfrak{P} -null sets are determined by a countable submodel $\mathfrak{P}_0 \subset \mathfrak{P}$.

1.2 Proposition (Halmos & Savage (1949)) A statistical model \mathfrak{P} is dominated if and only if there is a countable $\mathfrak{P}_0 \subset \mathfrak{P}$ and a distribution π on \mathfrak{P}_0 such that $Q := \int_{\mathfrak{P}_0} P(\cdot) \pi(dP)$ dominates \mathfrak{P} . \blacklozenge

Proof: For the sufficiency of that equivalence we refer to Halmos & Savage (1949, Lemma 7, page 232) which yields the existence of

$$\mathfrak{P}_0 = \{P_k : k \in \mathbb{N}\} \subset \mathfrak{P}$$

such that Q defined by

²We mostly write “iff” instead of “if and only if”.

$$Q(A) := \sum_{k \in \mathbb{N}} 2^k P_k(A) \quad \text{for } A \in \mathcal{A}$$

dominates \mathfrak{P} . The necessity is evident. \blacksquare

The countable structure of dominated models will turn out to be very beneficial, not just within the discussion of the sufficiency concept (see next section and the references therein), but also for purposes of this thesis (see Part II).

Statistical Morphisms

There are several ways and several starting points to formulate prior assumptions about \mathbf{P}^X . Some statistical models are equivalent, some are more, and some are less informative. To compare statistical models in this sense, we introduce the notions of a statistical morphism and isomorphism, for more details see e. g. Torgersen (1991).

1.3 Definitions (Statistical Morphism) Let $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and $(\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$ be two statistical spaces with parametrized statistical models $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ and $\mathfrak{Q} := \{Q_\vartheta : \vartheta \in \Theta\}$. Note that \mathfrak{P} and \mathfrak{Q} have the same parameter space. A Markov kernel K from $(\mathfrak{X}, \mathcal{A})$ to $(\mathfrak{Y}, \mathcal{B})$ is called a *statistical morphism* from $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ to $(\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$ iff

$$P_\vartheta^K := \int_{\mathfrak{X}} K(x, \cdot) P_\vartheta(dx) = Q_\vartheta \quad \text{for all } \vartheta \in \Theta.$$

Furthermore, if K is a Markov kernel on $(\mathfrak{X}, \mathcal{A})$, it is called an *identity* on $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ iff $P_\vartheta^K = P_\vartheta$ holds for all $\vartheta \in \Theta$.

We call K *statistical isomorphism* iff there is a statistical morphism L from $(\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$ to $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ such that $K \circ L$ given by

$$(K \circ L)(x, A) := \int_{\mathfrak{Y}} L(y, A) K(x, dy) \quad \text{for all } x \in \mathfrak{X}, A \in \mathcal{A}$$

and correspondingly $L \circ K$ are identities on $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and $(\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$, respectively. Then, L is called an *inverse* of K and vice versa. Moreover, we say that $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is *isomorphic* to $(\mathfrak{X}, \mathcal{B}, \mathfrak{Q})$ and write $(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) \cong (\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$. \blacklozenge

Whenever the underlying sample spaces are clear, we will use the notions and conventions just for the statistical models instead of the whole statistical spaces. For example, we will just write $\mathfrak{P} \cong \mathfrak{Q}$ instead of $(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) \cong (\mathfrak{Y}, \mathcal{B}, \mathfrak{Q})$ if no clarification is needed.

1.4 Remark (Morphism and Image Measure) A transformation by a morphism is a generalization of the usual imaging of measures. Let $T : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathfrak{Y}, \mathcal{B})$ be a measurable function and consider $K(x, \cdot) := \delta_{T(x)}(\cdot)$ for $x \in \mathfrak{X}$. Then we have indeed $P^T = P^K$ for every distribution $P \in$

$\text{Prob}(\mathfrak{X}, \mathcal{A})$. That means, K is a statistical morphism from $\mathfrak{P} := \text{Prob}(\mathfrak{X}, \mathcal{A})$ to $\mathfrak{Q} := \{P^T : P \in \text{Prob}(\mathfrak{X}, \mathcal{A})\}$. \blacklozenge

1.5 Example Let \mathfrak{P} be a statistical model on $(\mathfrak{X}, \mathcal{A})$ and let Q be an arbitrary distribution on $(\mathfrak{Y}, \mathcal{B})$. Then, the statistical model

$$\mathfrak{P} \otimes Q := \{P \otimes Q : P \in \mathfrak{P}\} \quad (1.1)$$

on $(\mathfrak{X} \times \mathfrak{Y}, \mathcal{A} \otimes \mathcal{B})$ is isomorphic to \mathfrak{P} .

We denote by $\text{pr}_1 : \mathfrak{X} \times \mathfrak{Y} \rightarrow \mathfrak{X}$, $(x, y) \mapsto x$ the projection on the first component. Obviously, $K((x, y), \cdot) := \delta_{\text{pr}_1(x, y)}(\cdot)$ for $(x, y) \in \mathfrak{X} \times \mathfrak{Y}$ defines a morphism from $\mathfrak{P} \otimes Q$ to \mathfrak{P} . Conversely, $L(x, A \times B) := \delta_x(A)Q(B)$ for $x \in \mathfrak{X}$ and $A \in \mathcal{A}$, $B \in \mathcal{B}$ determines uniquely a Markov kernel from $(\mathfrak{X}, \mathcal{A})$ to $(\mathfrak{X} \times \mathfrak{Y}, \mathcal{A} \otimes \mathcal{B})$. Furthermore,

$$P^L(A \times B) = \int L(x, A \times B)P(dx) = P(A)Q(B)$$

for all $A \in \mathcal{A}$, $B \in \mathcal{B}$ and, hence, L is a morphism from \mathfrak{P} to $\mathfrak{P} \otimes Q$. Note, $L \circ K$ and $K \circ L$ are identities on $\mathfrak{P} \otimes Q$ and \mathfrak{P} , respectively. \blacklozenge

Decision Rules

To make statements about \mathbf{P}^X , we consider statistical morphisms that transform $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ to a decision space.

1.6 Definition A measurable space $(\mathfrak{D}, \mathcal{D})$ representing a class of possible statements about \mathbf{P}^X is called *decision space*. A measurable map $\delta : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathfrak{D}, \mathcal{D})$ is called *decision rule* or *decision procedure*. \blacklozenge

The δ -induced class of probability measures $\mathfrak{P}^\delta := \{P^\delta : P \in \mathfrak{P}\}$ on $(\mathfrak{D}, \mathcal{D})$ describes the behavior of the decision rule δ . Therefore, a decision rule yields the transformation

$$(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) \xrightarrow{\delta} (\mathfrak{D}, \mathcal{D}, \mathfrak{P}^\delta).$$

This general concept is illustrated by the following types of decision procedures.

1.7 Example (Test Procedure) We consider the decision space $\mathfrak{D} := \{0, 1\}$, that is, a certain statement about \mathbf{P}^X is true or false. By that the statistical model \mathfrak{P} is divided into the so-called *hypothesis*

$$\mathfrak{P}_0 := \{P \in \mathfrak{P} : \text{statement about } P \text{ is true}\}$$

and the *alternative* $\mathfrak{P}_1 := \mathfrak{P} \setminus \mathfrak{P}_0$. The tuple $(\mathfrak{P}_0, \mathfrak{P}_1)$ is called *test problem*. A measurable map $\psi : \mathfrak{X} \rightarrow \{0, 1\}$ is a so-called *test procedure* or *test* for short. \blacklozenge

Another decision rule aims at estimating the parameter value $\kappa(\mathbf{P}^X)$, where $\kappa : \mathfrak{P} \rightarrow \mathfrak{R}$ is some \mathfrak{R} -valued parameter of \mathfrak{P} .

1.8 Example (Estimation Procedure) Let $\kappa : \mathfrak{P} \rightarrow \mathfrak{R}$ be some parameter, then (\mathfrak{P}, κ) is called *estimation problem*. Let \mathcal{K} be a σ -algebra on \mathfrak{R} , a measurable function $\hat{\kappa} : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathfrak{R}, \mathcal{K})$ is called *estimation procedure* or *estimator*, which is a decision rule with decision space $(\mathfrak{D}, \mathfrak{D}) = (\mathfrak{R}, \mathcal{K})$. ♦

An estimator is sometimes called *point estimator*. This refers to the fact that an estimation procedure picks a point in \mathfrak{R} . If we relax this requirement in favor of a set-valued function $\hat{\mathbf{K}}$ with $\hat{\mathbf{K}}(x) \subset \mathfrak{R}$ for $x \in \mathfrak{X}$ satisfying a “confidence” statement like “ $P(\hat{\mathbf{K}} \ni \kappa(P)) \geq 1 - \alpha$ ” for every $P \in \mathfrak{P}$ and some fixed $\alpha \in [0, 1]$, we arrive at the following decision procedure.

1.9 Example (Confidence Procedure) Consider a parameter $\kappa : \mathfrak{P} \rightarrow \mathfrak{R}$ and the decision space $(\mathfrak{D}, \mathfrak{D})$ given through a set system $\mathfrak{D} \subset 2^{\mathfrak{R}}$ and a σ -algebra \mathfrak{D} on \mathfrak{D} such that

$$\sigma(\{D \in \mathfrak{D} : D \cap \{\eta\} \neq \emptyset\} : \eta \in \mathfrak{R}) \subset \mathfrak{D}. \quad (1.2)$$

Any decision procedure $\hat{\mathbf{K}} : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathfrak{D}, \mathfrak{D})$ is called *confidence procedure* or *confidence region* for (\mathfrak{P}, κ) . By the measurability of $\hat{\mathbf{K}}$ we have that for every $\eta \in \kappa(\mathfrak{P})$

$$\{\hat{\mathbf{K}} \ni \eta\} := \{x \in \mathfrak{X} : \hat{\mathbf{K}}(x) \ni \eta\} = \hat{\mathbf{K}}^{-1}(\{D \in \mathfrak{D} : D \cap \{\eta\} \neq \emptyset\})$$

is an \mathcal{A} -event describing the coverage of the parameter value η . On \mathfrak{P} we may therefore consider the function

$$P \mapsto P(\{\hat{\mathbf{K}} \ni \kappa(P)\}),$$

which is called *coverage map* of $\hat{\mathbf{K}}$. Furthermore, $\inf_{P \in \mathfrak{P}} P(\{\hat{\mathbf{K}} \ni \kappa(P)\})$ is called *effective confidence level*. If $\inf_{P \in \mathfrak{P}} P(\{\hat{\mathbf{K}} \ni \kappa(P)\}) \geq 1 - \alpha$ for some $\alpha \in [0, 1]$, then $\hat{\mathbf{K}}$ is called $(1 - \alpha)$ -*confidence region* and $1 - \alpha$ is called *confidence level* of $\hat{\mathbf{K}}$. ♦

A confidence region is therefore a set-valued random element. The set \mathfrak{D} of the corresponding decision spaces $(\mathfrak{D}, \mathfrak{D})$ may be the system of subsets of \mathfrak{R} that are open, closed, or compact. Interestingly, there are Borel- σ -algebras \mathfrak{D} with (1.2) that come from a so-called *hit and miss topology* \mathcal{T} such that $(\mathfrak{D}, \mathcal{T})$ is a Polish space. For a discussion of such measure spaces and random elements, in particular *random closed sets*, see for example Molchanov (2005) and the references therein.

Spaces of Decision Rules

Each type of decision rule is accompanied with a natural set of corresponding procedures whose structure is determined by the underlying

statistical space $(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$. For the purpose of introducing these sets, a kind of Lebesgue–Bochner space w. r. t. the considered *statistical* space will serve as a starting point. This subsection is devoted to introducing this space. We discuss some technical details which facilitates an exact formulation of statements and proofs later on.

Let $k \in \mathbb{N}$, $p \in [1, \infty]$ and denote by $\mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, P) := \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, P; \mathbb{R}^k)$ the vector space of $(\mathbb{R}^k, \|\cdot\|)$ -valued and \mathfrak{A} -measurable functions φ such that

$$\int \|\varphi\|^p dP < \infty.$$

With respect to a statistical space $(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$, the relevant functions are in

$$\mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}) := \bigcap_{P \in \mathfrak{P}} \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, P).$$

Note that in this space interesting quantities will be described by integrals w. r. t. $P \in \mathfrak{P}$, which do not care about changes on null sets w. r. t. \mathfrak{P} . In what follows, we denote by $[\mathfrak{P}]$ the subspace of $\mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ containing all functions that equal the null function \mathfrak{P} -almost surely. We then define the quotient space

$$L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}) := \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}) / [\mathfrak{P}] := \{\varphi + [\mathfrak{P}] : \varphi \in \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})\}.$$

For some $\varphi \in \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$, we will write $\varphi + [\mathfrak{P}]$ to denote the corresponding $[\mathfrak{P}]$ -equivalence class of φ in $L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$, especially if there is a need to emphasize the difference between φ and $\varphi + [\mathfrak{P}]$. We often deal with elements in $L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ for which we usually write small Latin letters. That is, we write $f \in L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ which is consequently the equivalence class $f = \varphi + [\mathfrak{P}]$ for some $\varphi \in \mathfrak{L}_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$. For any $P \in \mathfrak{P}$, we may then consistently write $\int f dP$ for $\int \varphi dP$ and any $\varphi \in f$. In what follows, we will, for the sake of convenience, call f a function, too, rather than referring to f as equivalence class.

On $L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ we may introduce the following canonical locally convex topological structure given by the semi-norms

$$\|f\|_{P,p} := \sqrt[p]{\int \|f(x)\|^p P(dx)} \quad \text{for } f \in L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$$

and for $P \in \mathfrak{P}$. Hence, we define $\mathcal{T}_{L_p(\mathfrak{P})} = \tau(\|\cdot\|_{P,p} : P \in \mathfrak{P})$, which is referred to the *canonical topology* on $L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$. Note that $(L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})})$ is a topological vector space (TVS).³ Furthermore, note that $\|\cdot\|_{P,p}$ for $P \in \mathfrak{P}$

³A pair (E, \mathcal{T}) consisting of a linear space E (over a field $K \in \{\mathbb{R}, \mathbb{C}\}$) and a topology \mathcal{T} is called *topological vector space* iff the vector space operations, that is, $(f, g) \mapsto f + g$ and $(\alpha, f) \mapsto \alpha f$ are continuous on $(E \times E, \mathcal{T} \otimes \mathcal{T})$ and $(K \times E, \tau(|\cdot|) \otimes \mathcal{T})$, respectively.

usually does not constitute a norm on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, since in general $[P] \neq [\mathfrak{P}]$. However, $\mathcal{T}_{L_p(\mathfrak{P})}$ is a Hausdorff topology on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, which follows easily from, for example, Werner (2018, Lemma VIII.1.4, page 429).

If $\mathcal{C} \subset \mathcal{A}$ is a sub- σ -algebra, then $\mathcal{L}_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ is a linear subspace of $\mathcal{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. Turning to the corresponding quotient spaces, we point out that $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ can at least be identified with a linear subspace of $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. This is due to the fact that $[\mathfrak{P}] \cap \mathcal{L}_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ can be identified with $[\mathfrak{P}]$. In the following, we may therefore just write $[\mathfrak{P}]$ to denote the null element in the quotient space regardless of the considered σ -algebra.

We will use the conditional expectation in the spirit of these considerations, too. Let $P \in \mathfrak{P}$ and let $\mathcal{C} \subset \mathcal{A}$ be a sub- σ -algebra. By definition, the *conditional expectation* of $f \in L_1(\mathfrak{X}, \mathcal{A}, P)$ given \mathcal{C} is the unique $[P]$ -equivalence class $\mathbf{E}_P(f|\mathcal{C}) \in L_1(\mathfrak{X}, \mathcal{C}, P)$ which consists functions $\varphi \in \mathcal{L}_1(\mathfrak{X}, \mathcal{C}, P)$ fulfilling the Radon–Nikodým equations

$$\int_C f \, dP = \int_C \varphi \, dP \quad \text{for } C \in \mathcal{C}.$$

Recall furthermore that $\mathbf{E}_P(\cdot|\mathcal{C})$ is a well-defined continuous linear projection on the subspace $(L_1(\mathfrak{X}, \mathcal{C}, P), \|\cdot\|_{P,1})$ for every sub- σ -algebra $\mathcal{C} \subset \mathcal{A}$ and every single probability measure P . If \mathcal{C} and the model \mathfrak{P} are well-related to each other, there is a statistical analog, the *Rao–Blackwell projection*. This is a linear projection on $L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ which will be made precise in the next section (see Proposition 1.26, page 17).

Evaluation of Decision Rules

The basic problem is to select “good” decision rules, which in particular means that we have to evaluate them first. To this end, we consider the loss that a decision $d \in \mathcal{D}$ yields if $P \in \mathfrak{P}$ is the underlying true distribution. A non-negative measurable function $L : (d, P) \mapsto L(d, P)$ is called *loss function*. The *expected loss* of decision procedure δ under P is $\mathbf{E}_P(L(\delta(\cdot), P))$. The function $P \mapsto \mathbf{E}_P(L(\delta(\cdot), P))$ is called *risk function* of δ .

For estimation procedures there is a popular loss function that is widely used in the literature.

1.10 Example (Mean Squared Error) Consider $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ and some parameter $\kappa : \Theta \rightarrow \mathbb{R}^l$. Let $\hat{\kappa}$ be an estimator for an estimation problem (\mathfrak{P}, κ) . Then the expected loss w. r. t. the squared error loss function $L : \mathbb{R}^l \times \mathbb{R}^k \rightarrow \mathbb{R}_+$ with $(d, \vartheta) \mapsto \|d - \kappa(\vartheta)\|_{\mathbb{R}^l}^2$ is

$$\text{MSE}(\hat{\kappa}, \vartheta) := \int \|\hat{\kappa}(x) - \kappa(\vartheta)\|_{\mathbb{R}^l}^2 P_\vartheta(dx) \quad \text{for } \vartheta \in \Theta,$$

which is called *mean squared error*. ◆

Loss functions for confidence regions appear to be less obvious, compared to the situation for estimators. Within the class of confidence regions with an effective level $1 - \alpha$, we clearly would prefer smaller ones. This can be reflected by the following loss.

1.11 Example (Confidence Region Volume) Consider $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ with $\lambda^l(\Theta) > 0$ and some parameter $\kappa : \Theta \rightarrow \mathbb{R}^l$. Let $\widehat{\mathbf{K}}$ be a confidence region for (\mathfrak{P}, κ) with an effective level $1 - \alpha$. The loss of a realization $\widehat{\mathbf{K}}(x) \in \mathcal{B}(\mathbb{R}^l)$ for some $x \in \mathfrak{X}$ may be represented by $\lambda^l(\widehat{\mathbf{K}}(x))$ if each false parameter $\eta \in \widehat{\mathbf{K}}(x) \setminus \{\kappa(\vartheta)\}$ is associated with the same loss.

The expected loss is therefore the expected volume of $\widehat{\mathbf{K}}$, that is, $\mathbf{E}_\vartheta(\lambda^l(\widehat{\mathbf{K}}))$ for $\vartheta \in \Theta$. Due to Fubini's theorem, we may note furthermore that

$$\mathbf{E}_\vartheta(\lambda^l(\widehat{\mathbf{K}})) = \int P_\vartheta(\widehat{\mathbf{K}} \ni \eta) \lambda^l(d\eta).$$

For confidence regions $\widehat{\mathbf{K}}$ with $P_\vartheta(\widehat{\mathbf{K}} \ni \kappa(\vartheta)) \geq 1 - \alpha$ for all $\vartheta \in \Theta$, it is desired that $P_\vartheta(\widehat{\mathbf{K}} \ni \eta)$ for all $\eta \neq \kappa(\vartheta)$ and all $\vartheta \in \Theta$ is small. \blacklozenge

The remainder of this section provides some further considerations of the loss for multivariate estimators, in particular in the case of prior knowledge. For the sake of completeness, we first note the case for a general decision procedure δ . If there is a prior belief in terms of a distribution π on \mathfrak{P} , then

$$\int \mathbf{E}_P(L(\delta(\cdot), P)) \pi(dP)$$

constitutes the so-called *Bayes risk* of δ w. r. t. prior π . Eventually, we are looking for decision rules with low (Bayes) risk.

Consider specifically a k -parametric statistical model $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$. The *Bayes risk* w. r. t. the (Euclidean) squared error loss of an estimator $\widehat{\kappa} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ for the parameter \mathbf{id}_Θ and w. r. t. π is

$$\mathbf{r}(\widehat{\kappa}, \pi) := \int \int \|\widehat{\kappa}(x) - \vartheta\|^2 P_\vartheta(dx) \pi(d\vartheta).$$

We may generalize this as follows. Recall that any positive semi-definite symmetric matrix W defines a semi-norm which we denote by $\|x\|_W := \sqrt{x^\top W x}$. If W is positive definite, $\|\cdot\|_W$ is even a norm. For $W = \mathbf{I}$ we get the Euclidean norm $\|\cdot\| = \|\cdot\|_{\mathbf{I}}$. The *Bayes risk* w. r. t. the $\|\cdot\|_W$ -weighted squared error loss of an estimator $\widehat{\kappa} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ for \mathbf{id}_Θ and w. r. t. π is

$$\mathbf{r}(\widehat{\kappa}, \pi; W) := \int \int \|\widehat{\kappa}(x) - \vartheta\|_W^2 P_\vartheta(dx) \pi(d\vartheta). \quad (1.3)$$

For $W = \mathbf{I}$ we have $\mathbf{r}(\widehat{\kappa}, \pi; \mathbf{I}) = \mathbf{r}(\widehat{\kappa}, \pi)$. We furthermore may consider the matrix

$$\mathbf{R}(\widehat{\kappa}, \pi) := \int \int (\widehat{\kappa}(x) - \vartheta)(\widehat{\kappa}(x) - \vartheta)^\top P_\vartheta(dx) \pi(d\vartheta), \quad (1.4)$$

which we then compare with respect to the Löwner order⁴ \leq_L .

According to these evaluation concepts, we are interested in corresponding minimizers.

1.12 Definition (Bayes Estimate) An estimator which minimizes the risk (1.3) among all estimators is called *Bayes estimator* w. r. t. the $\|\cdot\|_W$ -weighted squared error loss. An estimator which minimizes the risk (1.4) among all estimators is called *Bayes estimator* w. r. t. \leq_L . ♦

We note a first relation between these two partial orders.

1.13 Proposition (Löwner Order, Weighted Squared Error Loss) Consider two estimators $\hat{\kappa}, \hat{\lambda} \in L_2(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}; \mathbb{R}^k)$ for the canonical parameter \mathbf{id}_Θ . Then, the following statements are equivalent:

(a) $\mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) \leq_L \mathbf{R}(\hat{\lambda}, \boldsymbol{\pi})$;

(b) $\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; W) \leq \mathbf{r}(\hat{\lambda}, \boldsymbol{\pi}; W)$ for every positive semi-definite W . ♦

Proof: By definition of the Löwner order we have $\mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) \leq_L \mathbf{R}(\hat{\lambda}, \boldsymbol{\pi})$ if and only if

$$u^\top \mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) u \leq u^\top \mathbf{R}(\hat{\lambda}, \boldsymbol{\pi}) u \quad \text{for all } u \in \mathbb{R}^k.$$

Note that $u^\top \mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) u$ corresponds to

$$\begin{aligned} u^\top \mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) u &= \int \int (u^\top (\hat{\kappa}(x) - \vartheta))^2 P_\vartheta(dx) \boldsymbol{\pi}(d\vartheta) \\ &= \mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; uu^\top), \end{aligned}$$

which represents a uu^\top -weighted squared error loss. Furthermore, by the spectral decomposition theorem, we have for every positive semi-definite symmetric matrix W vectors $u_1, \dots, u_n \in \mathbb{R}^k$ such that $W = \sum_{i=1}^n u_i u_i^\top$. Hence,

$$\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; W) = \sum_{i=1}^n \mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; u_i u_i^\top) \leq \sum_{i=1}^n \mathbf{r}(\hat{\lambda}, \boldsymbol{\pi}; u_i u_i^\top) = \mathbf{r}(\hat{\lambda}, \boldsymbol{\pi}; W)$$

and thus $\mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) \leq_L \mathbf{R}(\hat{\lambda}, \boldsymbol{\pi})$ if and only if $\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; W) \leq \mathbf{r}(\hat{\lambda}, \boldsymbol{\pi}; W)$ for all positive semi-definite matrices W . ■

Clearly, the Löwner order is stronger than any weighted squared error loss ordering. However, as long as a \leq_L -least element exists, we may restrict ourselves to the usual squared error loss. To show this, we first present the following elementary lemma.

1.14 Lemma (Chandrasekar & Kale (1984)) Let \mathfrak{W} be a set of symmetric, positive semi-definite matrices with existing \leq_L -least element $A \in \mathfrak{W}$. Then the following two statements are equivalent:⁵

⁴The *Löwner order* is a partial order on the set of all symmetric matrices defined by $A \leq_L B$ iff $B - A$ is positive semi-definite.

⁵We denote by $\text{tr}(A)$ the *trace* of a matrix.

(a) $A \leq_L B$ for all $B \in \mathfrak{B}$;

(b) $\text{tr}(A) \leq \text{tr}(B)$ for all $B \in \mathfrak{B}$. \blacklozenge

Proof: Obviously, (a) implies (b), since by definition $A \leq_L B$ means $B - A$ is positive semi-definite and hence

$$\text{tr}(B) - \text{tr}(A) = \text{tr}(B - A) \geq 0$$

for all $B \in \mathfrak{B}$.

To show that (b) implies (a), let $A^* \in \mathfrak{B}$ be the \leq_L -least element. Then $A^* \leq_L A$, that is, $A - A^*$ is positive semi-definite and, hence, we have $\text{tr}(A - A^*) \geq 0$. By assumption, we also have

$$\text{tr}(A - A^*) = \text{tr}(A) - \text{tr}(A^*) \leq 0.$$

Hence, $\text{tr}(A - A^*) = 0$ which means together with the positive semi-definiteness of $A - A^*$ that $A - A^* = \mathbf{0}$, that is, $A^* = A$. \blacksquare

According to Heyde (1997), the following observation was first given by Chandrasekar & Kale (1984) for estimating functions (see next chapter). Here, we consider a corresponding result for estimators.

1.15 Corollary Let \mathfrak{S} be a set of estimators which contains a Bayes estimator w. r. t. \leq_L . Then the following statements are equivalent:

(a) $\mathbf{R}(\hat{\kappa}, \boldsymbol{\pi}) \leq_L \mathbf{R}(\hat{\lambda}, \boldsymbol{\pi})$ for all $\hat{\lambda} \in \mathfrak{S}$, that is, $\hat{\kappa}$ is a Bayes estimator w. r. t. \leq_L in \mathfrak{S} ;

(b) $\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}) \leq \mathbf{r}(\hat{\lambda}, \boldsymbol{\pi})$ for all $\hat{\lambda} \in \mathfrak{S}$, that is, $\hat{\kappa}$ is a Bayes estimator w. r. t. squared error loss in \mathfrak{S} . \blacklozenge

Proof: We just note that for any estimator $\hat{\kappa} \in L_2(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}; \mathbb{R}^k)$ we have $\text{tr}(\mathbf{R}(\hat{\kappa}, \boldsymbol{\pi})) = \mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; \mathbf{I})$. Hence, the claim follows directly by Lemma 1.14. \blacksquare

1.2 Sufficiency, Ancillarity, and Completeness

The performance of a decision procedure highly depends on the information at hand. This concerns not just the amount, but also the quality of the data. Clearly, we only want to take account of the *useful*, that is, the *sufficient* information of our sample. In contrast, the *useless*, that is, the *ancillary* data should be eliminated as much as possible because, roughly speaking, it only adds noise to our decision-making process.

This section is devoted to these notions and corresponding concepts. We start with the classical notions sufficiency, ancillarity and completeness and recall some results that relate these concepts to one another.

We conclude the section with a study of corresponding spaces of decision procedures. These spaces are required to introduce the Rao–Blackwell projection which plays an essential role in this thesis.

Information and σ -Algebras

Intuitively, the information given by a σ -algebra $\mathcal{C} \subset \mathcal{A}$ is “sufficient” for \mathfrak{P} iff any information beyond \mathcal{C} fails to separate distributions of \mathfrak{P} . Mathematically speaking, there should be a common version of conditional probabilities given the sufficient information, that is, we precisely define:⁶

1.16 Definition A σ -algebra $\mathcal{C} \subset \mathcal{A}$ is called *sufficient* for \mathfrak{P} iff for every $A \in \mathcal{A}$

$$\bigcap_{P \in \mathfrak{P}} P(A|\mathcal{C}) \neq \emptyset.$$

A statistic S is called *sufficient* iff $\sigma(S)$ is sufficient. ◆

To argue the other way around, we specify the intuition of useless information. This consists of events that do not help toward making statements about \mathbf{P}^X .

1.17 Definition Any event $D \in \mathcal{A}$ such that $P \mapsto P(D)$ is constant is called *ancillary* for \mathfrak{P} . A set system $\mathcal{D} \subset \mathcal{A}$ is called *ancillary* iff every set in \mathcal{D} is ancillary. ◆

An ancillary set system $\mathcal{D} \subset \mathcal{A}$ is not generally useless toward making statements about \mathbf{P}^X . This is due to the fact that $\sigma(\mathcal{C} \cup \mathcal{D})$, where $\mathcal{C} \subset \mathcal{A}$ is some σ -algebra, may contain more information than \mathcal{C} alone (see for example Pfanzagl, 1994, page 45). The following theorem due to Basu (1955, 1958) shows when it is possible to remove ancillary events and reveals the complementary nature of the two concepts of sufficiency and ancillarity.

1.18 Theorem (Basu) Let \mathfrak{P} be a statistical model without a *singular pair*, that is, $P \not\perp Q$ for any choice $P, Q \in \mathfrak{P}$.⁷ Furthermore, let $\mathcal{C} \subset \mathcal{A}$ be sufficient for \mathfrak{P} . Then the *independent complement* of \mathcal{C} (w. r. t. \mathfrak{P})

$$\mathcal{C}^\perp := \{D \in \mathcal{A} : P(C \cap D) = P(C)P(D) \forall P \in \mathfrak{P}, \forall C \in \mathcal{C}\}$$

is ancillary for \mathfrak{P} . ◆

Proof: See Barra (1981, Theorem 2, page 26). ■

⁶Recall that for a sub- σ -algebra $\mathcal{C} \subset \mathcal{A}$ and $A \in \mathcal{A}$ the conditional probability $P(A|\mathcal{C}) = \mathbf{E}_P(\mathbb{1}_A|\mathcal{C}) \in \mathcal{L}_1(\mathfrak{X}, \mathcal{C}, P)$ is a $[P]$ -equivalence class (see remarks on page 9).

⁷Two measures μ and ν on some measurable space $(\mathfrak{X}, \mathcal{A})$ are called *singular* to each other iff there is an $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$. We then write $\mu \perp \nu$.

In the situation of Theorem 1.18 let $\mathcal{D} \subset \mathcal{C}^\perp$ be an arbitrary σ -algebra. By the statement above, \mathcal{D} is ancillary, that is, $\mathfrak{P}|_{\mathcal{D}} := \{P|_{\mathcal{D}} : P \in \mathfrak{P}\}$ has exactly one element which we denote by Q . Furthermore (see Example 1.5, page 6), we then have⁸

$$(\mathfrak{X}, \mathcal{C} \vee \mathcal{D}, \mathfrak{P}|_{\mathcal{C} \vee \mathcal{D}}) \cong (\mathfrak{X} \times \mathfrak{X}, \mathcal{C} \otimes \mathcal{D}, \mathfrak{P}|_{\mathcal{C}} \otimes Q).$$

Hence, every $P \in \mathfrak{P}|_{\mathcal{C} \vee \mathcal{D}}$ can be considered as the factorization $P|_{\mathcal{C}} \otimes P|_{\mathcal{D}}$ on $(\mathfrak{X} \times \mathfrak{X}, \mathcal{C} \otimes \mathcal{D})$. This decomposition reveals that P consists of the interesting component $P|_{\mathcal{C}}$ plus an additional randomization given by $Q = P|_{\mathcal{D}}$. Since the latter does not depend on $P \in \mathfrak{P}$, it is redundant for the corresponding statistical problem. We may then conclude by Example 1.5 that $\mathfrak{P}|_{\mathcal{C} \vee \mathcal{D}}$ is isomorphic to $\mathfrak{P}|_{\mathcal{C}}$, which reflects the intuitive role of the sufficiency concept. This observation can be extended as follows.⁹

1.19 Proposition Let $(\mathfrak{X}, \mathcal{T})$ be a Polish space¹⁰, let \mathfrak{P} be dominated, and let $\mathcal{C} \subset \mathcal{B}(\mathfrak{X})$ be sufficient for \mathfrak{P} . Then, \mathfrak{P} is isomorphic to $\mathfrak{P}|_{\mathcal{C}}$. \blacklozenge

Proof: A morphism from \mathfrak{P} to $\mathfrak{P}|_{\mathcal{C}}$ is trivially given by $L := \delta_{\text{id}_{\mathfrak{X}}}$, where $\text{id}_{\mathfrak{X}} : (\mathfrak{X}, \mathcal{B}(\mathfrak{X})) \rightarrow (\mathfrak{X}, \mathcal{C})$, see also Remark 1.4 on page 5.

To construct a morphism from $\mathfrak{P}|_{\mathcal{C}}$ to \mathfrak{P} , we note that due the Polishness of $(\mathfrak{X}, \mathcal{T})$, every $P \in \mathfrak{P}$ has a regular conditional distribution $K_P \in P(\cdot | \mathcal{C})$. Furthermore, we use that $\mathcal{B}(\mathfrak{X})$ is generated by a countable algebra \mathcal{R} (see e.g. Bauer, 1991, proof of Satz 44.3, page 397) and since \mathcal{C} is sufficient for \mathfrak{P} there is

$$K_0(\cdot, R) \in \bigcap_{P \in \mathfrak{P}} P(R | \mathcal{C}) \quad (1.5)$$

for every $R \in \mathcal{R}$. Since \mathcal{R} is countable, we have for all $P \in \mathfrak{P}$ that $K_0(\cdot, R) = P(R | \mathcal{C}) = K_P(\cdot, R)$ for all $R \in \mathcal{R}$ almost surely w. r. t. P . Furthermore, \mathfrak{P} is dominated, thus there is a countable submodel $\mathfrak{P}_0 \subset \mathfrak{P}$ such that $N \in \mathcal{B}(\mathfrak{X})$ is a \mathfrak{P}_0 -null set if and only if N is a \mathfrak{P} -null set (see Proposition 1.2). Hence, there is a \mathfrak{P}_0 -null set $N \in \mathcal{C}$ such that $K_0(x, R) = P(R | \mathcal{C})(x) = K_P(x, R)$ holds for all $R \in \mathcal{R}$ and all $x \in N^c$. Recall that \mathcal{R} is an intersection stable generator of $\mathcal{B}(\mathfrak{X})$, we therefore have that $K_P(x, \cdot)$ does not depend on P for all $x \in N^c$. For some $P \in \mathfrak{P}$ and some distribution Q on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, we set

$$K(x, B) := K_P(x, B) \mathbb{1}_{N^c}(x) + Q(B) \mathbb{1}_N(x) \quad x \in \mathfrak{X}, B \in \mathcal{B}(\mathfrak{X}).$$

Then, we have that

$$P|_{\mathcal{C}}^K(B) = \int K(x, B) P|_{\mathcal{C}}(dx) = \int P(B | \mathcal{C}) dP|_{\mathcal{C}} = P(B)$$

⁸We write $\mathcal{C} \vee \mathcal{D}$ for $\sigma(\mathcal{C} \cup \mathcal{D})$.

⁹A broader and more detailed view is given by Strasser (1985, Theorem 24.11 and 24.12, p. 107), see also Pfanzagl (1994, Proposition 1.3.1, p. 8 f.) and references therein.

¹⁰A topological space $(\mathfrak{X}, \mathcal{T})$ is called *Polish* iff there is a metric d on \mathfrak{X} such that \mathcal{T} is generated by d and (\mathfrak{X}, d) is separable and complete.

for all $B \in \mathfrak{B}(\mathfrak{X})$ and $P \in \mathfrak{P}$. Hence, K is a morphism from $\mathfrak{P}|_{\mathcal{C}}$ to \mathfrak{P} . Obviously, $K \circ L$ and $L \circ K$ are identities on \mathfrak{P} and $\mathfrak{P}|_{\mathcal{C}}$, respectively. That means, we have $\mathfrak{P} \cong \mathfrak{P}|_{\mathcal{C}}$. ■

In the above proposition we need two regularity assumptions in terms of a countable structure. The first is included by the assumption that $(\mathfrak{X}, \mathcal{T})$ is Polish, which yields the existence of a regular conditional distribution. The second comes from the assumption that \mathfrak{P} is dominated. Hence, the \mathfrak{P} -null sets are described by a countable sub-model $\mathfrak{P}_0 \subset \mathfrak{P}$, see Proposition 1.2.

By the nature of the sufficiency concept, we go for a “minimal” sufficient σ -algebra \mathcal{C} . In terms of Theorem 1.18, we want to separate the ancillary material as an independent complement as much as possible.

1.20 Definition (Free of Ancillarity) We call a σ -algebra $\mathcal{C} \subset \mathcal{A}$ *free from ancillarity* w. r. t. \mathfrak{P} iff $\{\emptyset, \mathfrak{X}\}$ coincides \mathfrak{P} -almost surely with the system of all \mathcal{C} -measurable ancillary events. ◆

In other words, \mathcal{C} is free from ancillarity iff $C \in \mathcal{C}$ with

$$P \mapsto P(C) = \text{const} \quad (1.6)$$

implies that either $\mathbb{1}_C = \mathbb{1}_{\mathfrak{X}}$ or $\mathbb{1}_C = \mathbb{1}_{\emptyset}$ holds almost surely w. r. t. \mathfrak{P} . Clearly, a sufficient σ -algebra that is independent of every ancillary event needs to be free from ancillarity. This is an observation due to Lehmann (1981, Theorem 2, page 336) which is shown in the following lemma.

1.21 Lemma Let $\mathcal{D} := \{D \in \mathcal{A} : P \mapsto P(D) \equiv \text{const}\}$ be the system of all ancillary events for \mathfrak{P} . Then, any sub- σ -algebra $\mathcal{C} \subset \mathcal{A}$ with

$$\mathcal{D} \subset \mathcal{C}^{\perp} \quad (1.7)$$

is free from ancillarity. ◆

Proof: Any event in $\mathcal{C} \cap \mathcal{D}$ is \mathfrak{P} -trivial, that is, any ancillary event in \mathcal{C} has probability either 0 or 1 w. r. t. every $P \in \mathfrak{P}$. Hence, for any $C \in \mathcal{C} \cap \mathcal{D}$ we have either $\mathbb{1}_C = \mathbb{1}_{\mathfrak{X}}$ or $\mathbb{1}_C = \mathbb{1}_{\emptyset}$ almost surely w. r. t. \mathfrak{P} . ■

In general, a sufficient σ -algebra \mathcal{C} that is free from ancillarity does not imply (1.7), see Lehmann (1981, Example 2, page 337). However, considering a linear extension of these notions will help. That is, instead of considering a set system containing sets D such that $P \mapsto \int \mathbb{1}_D dP$ is constant, we are now going to extend the notion of ancillarity to elements of $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$.

1.22 Definition An element $h \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is called *ancillary of first order* iff $P \mapsto \int h dP$ is constant. The set of all first order ancillary functions in $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is denoted by $A_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. ◆

A σ -algebra \mathcal{C} is said to be “completely” free of ancillarity iff there is no non-trivial \mathcal{C} -measurable ancillary function of first order. By that we arrive at the following notion.¹¹

1.23 Definition A σ -algebra \mathcal{C} is called *(boundedly) complete* for \mathfrak{P} iff every (bounded) \mathcal{C} -measurable and \mathfrak{P} -integrable function h is \mathfrak{P} -a. s. uniquely determined by their coordinates $(\int h dP)_{P \in \mathfrak{P}}$. A model \mathfrak{P} is called *(boundedly) complete*, iff \mathcal{A} is (boundedly) complete. \blacklozenge

Any complete σ -algebra is boundedly complete, and any boundedly complete σ -algebra is free from ancillarity. Conversely, however, bounded completeness is not equivalent to completeness (see Lehmann & Scheffé, 1950). Furthermore, a σ -algebra that is free from ancillarity is in general not boundedly complete (see Lehmann, 1981).¹²

Regarding our initial intention to separate the sufficient information from the ancillary one as well as possible, we note that a bounded complete and sufficient σ -algebra \mathcal{C} fulfills (1.7).

1.24 Theorem (Basu (1958)) Let \mathcal{C} be boundedly complete and sufficient, then any ancillary event is \mathfrak{P} -independent of \mathcal{C} , that is, (1.7) holds. \blacklozenge

Proof: See Barra (1981, Theorem 1, page 26). \blacksquare

1.25 Corollary Let \mathfrak{P} be a statistical model without singular pair and let $\mathcal{C} \subset \mathcal{A}$ be boundedly complete and sufficient for \mathfrak{P} , then

$$\mathcal{C}^\perp = \{D \in \mathcal{A} : D \text{ is ancillary}\}.$$

\blacklozenge

Proof: Let $\mathcal{D} \subset \mathcal{A}$ be the set system of \mathfrak{P} -ancillary events. By Proposition 1.18 we have $\mathcal{C}^\perp \subset \mathcal{D}$ and by Proposition 1.24 we have $\mathcal{D} \subset \mathcal{C}^\perp$. \blacksquare

As we have seen, a complete sufficient σ -algebra \mathcal{C} is free from non-trivial ancillary events. In this sense, \mathcal{C} is “completely” reduced.

Information and Decision Procedures

So far, we have decomposed the information \mathcal{A} , roughly speaking, into sufficient and ancillary events w. r. t. the statistical model \mathfrak{P} . Instead of

¹¹The notion of completeness was systematically introduced and studied by Lehmann & Scheffé (1947, 1950, 1955). In a more general fashion, the notion of completeness can be introduced w. r. t. a class of functions. Lehmann (1981) introduced in particular the notion of *completeness w. r. t. to the class of two-valued functions*, “ F_0 -completeness” in his terms, which corresponds to our notion of a σ -algebra that is free from ancillarity.

¹²Here and in what follows, the notion of completeness could be introduced and considered in a more differentiated way (see for example Schmetterer, 1960).

$(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ we may equivalently¹³ consider the statistical subspace $(\mathfrak{X}, \mathcal{C}, \mathfrak{P}|_{\mathcal{C}})$ if $\mathcal{C} \subset \mathcal{A}$ is sufficient (see Proposition 1.19). If in addition \mathcal{C} is boundedly complete, the decomposition is most successful¹⁴ in the sense of Lemma 1.21 and Corollary 1.25.

However, since we are finally interested in decision procedures, we want to know if these intuitions carry over to decision procedures. The central question for the remainder of this section is, therefore, how do the notions of sufficiency, ancillarity and completeness relate to suitable spaces of decision procedures. We specifically consider the decision space $(\mathfrak{D}, \mathfrak{D}) = (\mathbb{R}^k, \mathfrak{B}(\mathbb{R}^k))$.

Recall that for every $P \in \mathfrak{P}$ and any $\mathcal{C} \subset \mathcal{A}$ the conditional expectation $\mathbf{E}_P(\cdot|\mathcal{C})$ is a continuous linear projection in $(L_1(\mathfrak{X}, \mathcal{A}, P), \|\cdot\|_{P,1})$. There is a statistical analog, if the σ -algebra $\mathcal{C} \subset \mathcal{A}$ is sufficient for \mathfrak{P} .

1.26 Proposition (Rao–Blackwell Projection) Let $\mathcal{C} \subset \mathcal{A}$ be a sufficient sub- σ -algebra for \mathfrak{P} and let $p \in [1, \infty]$, then the map

$$\begin{aligned} \mathbf{E}(\cdot|\mathcal{C}): \quad (L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})}) &\rightarrow (L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})}) \\ f &\mapsto \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C}) \end{aligned}$$

is a well-defined continuous linear projection on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$. \blacklozenge

Proof: Since \mathcal{C} is sufficient, we have $\bigcap_{P \in \mathfrak{P}} P(A|\mathcal{C}) \neq \emptyset$ for all $A \in \mathcal{A}$ by definition. Hence, by the *standard extension argument*, that is, by approximating non-negative and \mathcal{A} -measurable functions f through simple functions from below and by using monotone limits, we have

$$\bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C}) \neq \emptyset. \quad (1.8)$$

Finally, by the linearity of $\mathbf{E}_P(\cdot|\mathcal{C})$ this holds also for any $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ since $f = f^+ - f^-$, where f^+ and f^- denote the positive and the negative part of f , respectively.

Further, (1.8) determines a unique $[\mathfrak{P}]$ -equivalence class in $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$. To show this, consider $\varphi \in \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C})$, that is, we have

$$\int_C f \, dP = \int_C \varphi \, dP \quad \text{for } C \in \mathcal{C}, P \in \mathfrak{P} \quad (1.9)$$

by the definition of $\mathbf{E}_P(f|\mathcal{C})$ for all $P \in \mathfrak{P}$. Since condition (1.9) does not care about \mathfrak{P} -null sets, we have

$$\varphi + [\mathfrak{P}] \subset \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C}).$$

¹³Under regularity assumptions about the statistical space $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ which is: (i) \mathfrak{P} is dominated, and (ii) $(\mathfrak{X}, \mathcal{A}) = (\mathfrak{X}, \mathfrak{B}(\mathfrak{X}, \mathcal{T}))$, where $(\mathfrak{X}, \mathcal{T})$ is a Polish space.

¹⁴Under the regularity assumption that \mathfrak{P} has no singular pair, that is, the existence of $P, Q \in \mathfrak{P}$ with $P \perp Q$ is excluded.

Conversely, for any $\varphi' \notin \varphi + [\mathfrak{F}]$, which is equivalent to $\varphi' - \varphi \notin [\mathfrak{F}]$, there is a $P \in \mathfrak{P}$ such that

$$\int f \, dP = \int \varphi \, dP \neq \int \varphi' \, dP$$

and thus $\varphi' \notin \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C})$. By that we have the stated well-definedness of the map $\mathbf{E}(\cdot|\mathcal{C})$ from $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ to $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$.

Since $\mathbf{E}_P(\cdot|\mathcal{C})$ is a linear projection for every $P \in \mathfrak{P}$, we immediately get that $\mathbf{E}(\cdot|\mathcal{C})$ is a linear projection, too. Furthermore, by Jensen's inequality for conditional expectations, we have

$$\|\mathbf{E}(f|\mathcal{C})\|_{P,p} \leq \sqrt[p]{\int \mathbf{E}(\|f\|_{\mathbb{R}^k}^p|\mathcal{C}) \, dP} = \|f\|_{P,p}$$

for every $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and every $P \in \mathfrak{P}$. Hence, $\mathbf{E}(\cdot|\mathcal{C})$ is continuous w. r. t. the canonical topology

$$\mathcal{T}_{L_p(\mathfrak{P})} = \tau(\|\cdot\|_{P,p} : P \in \mathfrak{P})$$

on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$, respectively (see for example Werner, 2018, Satz VIII.2.3, page 431). \blacksquare

Let $\mathcal{C} \subset \mathcal{A}$ be a sufficient sub- σ -algebra for \mathfrak{P} , we then call $\mathbf{E}(\cdot|\mathcal{C})$ the *Rao–Blackwell projection* given \mathcal{C} and w. r. t. \mathfrak{P} , while $\mathbf{E}(f|\mathcal{C})$ for an $f \in L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is called *Rao–Blackwellization* of f w. r. t. \mathcal{C} . The name is due to studies of Rao (1945) and Blackwell (1947) which will be discussed later (see Theorem 2.17).

The next proposition shows that $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$, that is, the subspace of all Rao–Blackwellizations, is a complemented subspace¹⁵ in $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. In addition, if \mathcal{C} is complete, then the complement is

$$\mathbf{A}_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) := \{g \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) : \int g \, dP = \mathbf{0} \text{ for all } P \in \mathfrak{P}\},$$

that is, $\mathbf{A}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ modulo a constant function.

1.27 Theorem (Rao–Blackwellization, TVS Version) Let $p \in [1, \infty]$ and let \mathcal{C} be sufficient for \mathfrak{P} , then:¹⁶

(a) $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) = L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}) \oplus \ker(\mathbf{E}(\cdot|\mathcal{C}))$,

(b) $\ker(\mathbf{E}(\cdot|\mathcal{C})) \subset \mathbf{A}_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ with equality, if \mathcal{C} is complete. \blacklozenge

¹⁵A closed subspace F of a topological vector space E is said to be *complemented in E* iff there exists a closed subspace G of E such that $E = F + G$ and $F \cap G = \{\mathbf{0}\}$. In this case, we write $E = F \oplus G$ and E is said to be the *direct sum* of F and G , while F , G are called *complementary subspaces* (see e. g. Lang, 1993, page 389).

¹⁶We denote by $\ker(T) := \{e \in E : Te = \mathbf{0}\}$ the *kernel* of an operator $T : E \rightarrow F$.

Proof: Since $\mathbf{E}(\cdot|\mathcal{C})$ is a continuous linear projection in the topological vector space $(L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})})$, see Proposition 1.26, **(a)** follows from (Rudin, 1973, Theorem 5.16, page 126) which yields¹⁷

$$L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}) = \text{ran}(\mathbf{E}(\cdot|\mathcal{C})) \oplus \ker(\mathbf{E}(\cdot|\mathcal{C})). \quad (1.10)$$

By definition of $\mathbf{E}(\cdot|\mathcal{C})$ we have $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}) = \text{ran}(\mathbf{E}(\cdot|\mathcal{C}))$.

To show **(b)**, we note that $f \in \ker(\mathbf{E}(\cdot|\mathcal{C}))$ implies $\int f dP = \mathbf{0}$ for all $P \in \mathfrak{P}$ and, hence, $f \in A_p^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$, that is, $\ker(\mathbf{E}(\cdot|\mathcal{C})) \subset A_p^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$. Conversely, if $f \in A_p^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$, we have that $\int \mathbf{E}(f|\mathcal{C}) dP = \int f dP = \mathbf{0}$ for all $P \in \mathfrak{P}$. Hence, if \mathcal{C} is additionally complete, we can conclude that $\mathbf{E}(f|\mathcal{C}) = \mathbf{0}$, since $\mathbf{E}(f|\mathcal{C})$ is \mathcal{C} -measurable by construction, and consequently that $f \in \ker(\mathbf{E}(\cdot|\mathcal{C}))$. ■

The factorization (1.10) is of statistical interest since it reveals an additive superposition of the useful and the useless components of decision procedures in $L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$. Formally, any $f \in L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ can be represented as $f = \mathbf{E}(f|\mathcal{C}) + \alpha$, where $\alpha \in \ker(\mathbf{E}(\cdot|\mathcal{C})) \subset A_p^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$ does not contribute something on average and can therefore be seen as a noise.

Especially, if $p \geq 2$, we have that the noise α is uncorrelated to $\mathbf{E}(f|\mathcal{C})$ w. r. t. every $P \in \mathfrak{P}$, since $\mathbf{E}(\alpha|\mathcal{C}) = \mathbf{0}$ and therefore

$$\begin{aligned} \text{Cov}_P(\mathbf{E}(f|\mathcal{C}), \alpha) &= \int \mathbf{E}(f|\mathcal{C}) \alpha^\top dP - \mathbf{E}_P f (\mathbf{E}_P \alpha)^\top \\ &= \int \mathbf{E}(f|\mathcal{C}) (\mathbf{E}(\alpha|\mathcal{C}))^\top dP - \mathbf{E}_P f (\mathbf{E}_P \mathbf{E}(\alpha|\mathcal{C}))^\top \\ &= \mathbf{0} \end{aligned}$$

for all $P \in \mathfrak{P}$. Regarding Theorem 1.27, we therefore additionally note:

1.28 Remark If $p \geq 2$ the subspace of Rao–Blackwellizations $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ is orthogonal¹⁸ to $\ker(\mathbf{E}(\cdot|\mathcal{C}))$, that is,

$$L_p(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}) = L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}) \oplus \ker(\mathbf{E}(\cdot|\mathcal{C})), \quad (1.11)$$

where the orthogonality is meant w. r. t. every bilinear form $\langle \cdot, \cdot \rangle_P$ for $P \in \mathfrak{P}$ defined by $(f, g) \mapsto \int \langle f(x), g(x) \rangle_{\mathbb{R}^k} P(dx)$. This is due to the fact that for all $f \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ and all $g \in \ker(\mathbf{E}(\cdot|\mathcal{C}))$ we have for all $P \in \mathfrak{P}$

$$\langle f, g \rangle_P = \int \langle f, g \rangle_{\mathbb{R}^k} dP = \int \langle f, \mathbf{E}(g|\mathcal{C}) \rangle_{\mathbb{R}^k} dP = 0.$$

¹⁷We denote by $\text{ran}(T) := \{Te : e \in E\} \subset F$ the *range* of an operator $T : E \rightarrow F$.

¹⁸Let F, G be complementary subspaces of a topological vector space E , that is, $E = F \oplus G$. Then F is said to be *orthogonal* to G w. r. t. a bilinear form b on E iff $b(f, g) = 0$ for all $f \in F$ and $g \in G$. Furthermore, F is called the *orthogonal complement* of G w. r. t. b and vice versa and we write $E = F \oplus G$ for the so-called *orthogonal sum*.

Furthermore, if \mathcal{C} is complete then $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) = L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}) \oplus A_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. Related factorizations of this kind emerge implicitly in the theory of unbiased estimation when characterizing uniformly minimum variance unbiased estimators, see for example Lehmann & Scheffé (1950, Theorem 5.3) or Rao (1952, Theorem 1). \blacklozenge

Roughly speaking, Theorem 1.27 and Remark 1.28 summarize Basu's Theorems 1.18 and 1.24 in terms of topological vector spaces. While Theorem 1.27 (a) and (b) corresponds to the statement of Theorem 1.18, the supplement in part (b) reveals that $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ is maximally separated from ancillary material if \mathcal{C} is sufficient and complete which is the counterpart to Theorem 1.24 and Corollary 1.25.

A generalization of Remark 1.28 for \mathbb{R} -valued functions reveals a way toward the existence of a maximally complete σ -algebra for an arbitrary statistical model \mathfrak{P} . Before we go into that, we describe the (topological) dual space¹⁹ of the statistical spaces $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})$. The proof of the following proposition is similar to that of Werner (2018, Satz II.2.4, page 65) and was given by the author on MathOverflow (see H. 2020b).

1.29 Proposition Let $p, q \in [1, \infty]$, $p < \infty$ be conjugate exponents, that is, $\frac{1}{p} + \frac{1}{q} = 1$, then the dual space of $(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}), \mathcal{T}_{L_p(\mathfrak{P})})$ is given by

$$L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})' = \text{span} \left(\bigcup_{P \in \mathfrak{P}} \{ \langle \cdot, g \rangle_P |_{L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})} : g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, P) / [\mathfrak{P}] \} \right). \quad (1.12)$$

\blacklozenge

Proof: Consider $P \in \mathfrak{P}$ and $g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, P) / [\mathfrak{P}]$, then by Hölder's inequality

$$| \langle f, g \rangle_P | \leq \int |f(x)g(x)| P(dx) \leq \|f\|_{P,p} \cdot \|g\|_{P,q}.$$

Recall that $\mathcal{T}_{L_p(\mathfrak{P})} = \tau(\| \cdot \|_{P,p} : P \in \mathfrak{P})$ and, hence, $\langle \cdot, g \rangle_P$ is a linear continuous functional on $(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})})$. Furthermore, any linear combination of such linear functionals is continuous and, hence, the right-hand side of (1.12) is contained in the left-hand side.

Conversely, let $f' \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})'$ and define

$$v(A) := f'(\mathbb{1}_A) \quad \text{for } A \in \mathcal{A}.$$

Then, v is a signed measure. Furthermore, since f' is linear and continuous, there is a *finite* set $I \subset \mathfrak{P}$ and a constant $M > 0$ such that

$$|f'(f)| \leq M \cdot \max \{ \|f\|_{P,p} : P \in I \} \quad \text{for } f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})$$

¹⁹Let (E, \mathcal{T}) be a topological vector space. The linear space consisting of all continuous linear functionals on E is called *topological dual space* of E and is denoted by E' .

(Werner, 2018, Satz VIII.2.3, page 431). Hence, ν is dominated by the finite measure $\mu := \sum_{P \in I} P(\cdot)$. According to the theorem of Radon–Nikodým (see e. g. Elstrodt, 2018, Satz 2.3), ν has a density g w. r. t. μ , that is,

$$f'(\mathbb{1}_A) = \nu(A) = \int \mathbb{1}_A g \, d\mu = \sum_{P \in I} \langle \mathbb{1}_A, g \rangle_P \quad \text{for } A \in \mathcal{A}.$$

Obviously, by the linearity of both sides we have

$$f'(\varphi) = \int \varphi g \, d\mu = \sum_{P \in I} \langle \varphi, g \rangle_P \quad \text{for } A \in \mathcal{A} \quad (1.13)$$

for all simple functions φ . Furthermore, the continuity w. r. t. $\|\cdot\|_\infty$ of both sides yields that (1.13) holds on $\mathcal{L}_\infty(\mathfrak{X}, \mathcal{A}, \mu)/[\mathfrak{F}]$.

It remains to show that g can be chosen such that $g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, P)/[\mathfrak{F}]$ for all $P \in I$. To this end, we first treat the case $q < \infty$ and define

$$h := \frac{|g|^q}{g},$$

where $\frac{0}{0} := 0$. Then, $|g|^q = hg = |h|^p$ and for $n \in \mathbb{N}$ we set $A_n := \{|g| \leq n\}$. That means, we have $h\mathbb{1}_{A_n} \in \mathcal{L}_\infty(\mathfrak{X}, \mathcal{A}, \mu)/[\mathfrak{F}]$ and

$$\begin{aligned} \int_{A_n} |g|^q \, d\mu &= \int \mathbb{1}_{A_n} h g \, d\mu = f'(\mathbb{1}_{A_n} h) \leq \|f'\| \|\mathbb{1}_{A_n} h\|_{\mu, p} \\ &= \|f'\| \left(\int_{A_n} |h|^p \, d\mu \right)^{\frac{1}{p}} \\ &= \|f'\| \left(\int_{A_n} |g|^q \, d\mu \right)^{\frac{1}{p}} \end{aligned}$$

and, consequently,

$$\left(\int_{A_n} |g|^q \, d\mu \right)^{\frac{1}{q}} = \left(\int_{A_n} |g|^q \, d\mu \right)^{1 - \frac{1}{p}} \leq \|f'\|$$

for all $n \in \mathbb{N}$. The continuity of f' yields that the right-hand side is finite, and the theorem of monotone convergence yields that $g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, \mu)/[\mathfrak{F}]$. Since $\mu = \sum_{P \in I} P(\cdot)$, we also have $g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, P)/[\mathfrak{F}]$ for all $P \in I$.

For the case $q = \infty$, and hence $p = 1$, we show that $g \in \mathcal{L}_\infty(\mathfrak{X}, \mathcal{A}, P)/[\mathfrak{F}]$ for all $P \in I$. To this end, define $A := \{|g| - \|f'\| > 0\}$ and $h := \mathbb{1}_A \frac{|g|}{g}$. Then, note that $h \in L_\infty(\mu)$, $\mu(A) = \|h\|_{\mu, 1}$, and

$$\begin{aligned} 0 \leq \int_A |g| - \|f'\| \, d\mu &= \int h g \, d\mu - \mu(A) \|f'\| \leq f'(h) - \mu(A) \|f'\| \\ &\leq \|f'\| \|h\|_{\mu, 1} - \mu(A) \|f'\| \\ &= 0, \end{aligned}$$

that is, $\mu(A) = 0$. Hence, $g \in \mathcal{L}_\infty(\mathfrak{X}, \mathcal{A}, \mu)/[\mathfrak{P}]$, that is, $g \in \mathcal{L}_\infty(\mathfrak{X}, \mathcal{A}, P)/[\mathfrak{P}]$ for all $P \in I$.

Finally, since $\sum_{P \in I} \langle \cdot, g \rangle_P$ is a linear continuous functional on $L_\infty(\mu)$, and since $L_\infty(\mu)$ is a dense subspace of $L_1(\mu)$, we can uniquely extend that functional on $L_1(\mu)$. Furthermore, $\sum_{P \in I} \langle \cdot, g \rangle_P$ is then also a linear continuous functional on $(\mathcal{L}_1(\mathfrak{X}, \mathcal{A}, \mu; \mathbb{R})/[\mathfrak{P}], \mathcal{T}_{L_1(\mathfrak{P})})$ and

$$\sum_{P \in I} \langle \cdot, g \rangle_P|_{L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})} = f',$$

which yields the claim. \blacksquare

1.30 Proposition Let $\mathcal{C} \subset \mathcal{A}$ be complete and sufficient, and let $p, q \in [1, \infty]$ with $p < \infty$ be conjugate exponents, then:

(a) For the annihilator²⁰ of $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R})$ in $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})'$ we have

$$L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R})^\perp = \text{span} \{ \langle \cdot, g \rangle_P : g \in A_q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}), P \in \mathfrak{P} \}.$$

(b) Furthermore,

$$L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}) = \bigcap \{ \ker \langle \cdot, g \rangle_P : g \in A_q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}), P \in \mathfrak{P} \}.$$

\blacklozenge

Proof: To show (a), let $g \in A_q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})$ and let $P \in \mathfrak{P}$, then $\langle \cdot, g \rangle_P$ is a linear continuous functional on $(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathcal{T}_{L_p(\mathfrak{P})})$ according to Proposition 1.29. Furthermore, since \mathcal{C} is complete and sufficient, we have $A_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}) = \ker(\mathbf{E}(\cdot|\mathcal{C}))$ according to Theorem 1.27 (b). Consequently, we have $\mathbf{E}(g|\mathcal{C}) = \mathbf{0}$ and hence

$$\int f g dP = \int f \mathbf{E}(g|\mathcal{C}) dP = 0 \quad \text{for all } f \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}).$$

That means $\langle \cdot, g \rangle_P \in L_q(\mathfrak{X}, \mathcal{C}, \mathfrak{P})^\perp$ and thus

$$L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R})^\perp \supset \text{span} \{ \langle \cdot, g \rangle_P : g \in A_q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}), P \in \mathfrak{P} \}.$$

Conversely, let $f' \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R})^\perp$ then there is a finite set $I \subset \mathfrak{P}$ and a function $g \in \mathcal{L}_q(\mathfrak{X}, \mathcal{A}, \sum_{P \in I} P)/[\mathfrak{P}]$ such that (see Proposition 1.29 and its proof)

$$f' = \sum_{P \in I} \langle \cdot, g \rangle_P = |I| \langle \cdot, g \rangle_Q$$

with $Q := \frac{1}{|I|} \sum_{P \in I} P$. Furthermore, $f'(f) = 0$ for all $f \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ implies

$$0 = \langle f, g \rangle_Q = \int f g dQ = \int f \mathbf{E}_Q(g|\mathcal{C}) dQ$$

²⁰Let (E, \mathcal{T}) be a topological vector space and let F be a subspace of E . The subspace of the dual space E' defined by $F^\perp := \{f' \in E' : f'(f) = 0 \quad \forall f \in F\}$ is called *annihilator* of F in E' .

for all $f \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ and hence $\mathbf{E}_Q(g|\mathcal{C}) = \mathbf{0}$. Since \mathcal{C} is also sufficient and complete for $\mathfrak{P} \cup \{Q\}$, we have thus $g \in \ker(\mathbf{E}(\cdot|\mathcal{C})) = A_Q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})$.

To show **(b)**, we note that by the theorem of Hahn–Banach (see e. g. Werner, 2018, Theorem VIII.2.11, page 437) the closure²¹ of $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ w. r. t. the $\mathcal{T}_{L_p(\mathfrak{P})}$ can be described through the continuous linear functionals as follows

$$\begin{aligned} \text{cl}(L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}))_{\mathcal{T}_{L_p(\mathfrak{P})}} &= \bigcap \{ \ker(g') : g' \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})', g'|_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})} = \mathbf{0} \} \\ &= \bigcap \{ \ker(g') : g' \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})^\perp \}. \end{aligned}$$

Since $\mathbf{I} - \mathbf{E}(\cdot|\mathcal{C})$ is continuous w. r. t. $\mathcal{T}_{L_p(\mathfrak{P})}$, we have that

$$L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}) = \text{ran}(\mathbf{E}(\cdot|\mathcal{C})) = \ker(\mathbf{I} - \mathbf{E}(\cdot|\mathcal{C}))$$

is closed w. r. t. $\mathcal{T}_{L_p(\mathfrak{P})}$ and the claim follows by part (a). \blacksquare

1.31 Corollary Let $\mathcal{C} \subset \mathcal{A}$ be complete and sufficient and let $p, q \in [1, \infty]$ with $p < \infty$ be conjugate exponents. Then, $f \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R})$ if and only if $\langle f, g \rangle_P = 0$ for all $g \in A_q^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R})$ and $P \in \mathfrak{P}$. \blacklozenge

Proof: The claim follows directly by Proposition 1.30 (b). \blacksquare

In correspondence with Proposition 1.30 (b) and Corollary 1.31, we have an explicit description of a maximally complete σ -algebra (see Proposition 1.32 (a) and (b) below). Bahadur (1957) seems to be the first who mentioned the σ -algebra (1.14) below and, to the best of the author's knowledge, the maximality was first noticed by Schmetterer & Strasser (1974), see also Kagan *et al.* (2014).

1.32 Proposition Let $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ be a statistical space and let $p \in [1, \infty[$.

(a) The set system

$$\mathcal{V}(\mathfrak{P}) := \{ A \in \mathcal{A} : \langle \mathbb{1}_A, g \rangle_P = 0 \ \forall g \in A_1^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}), \ \forall P \in \mathfrak{P} \} \quad (1.14)$$

is a complete σ -algebra for \mathfrak{P} .

(b) Let $\mathcal{W} \subset \mathcal{A}$ be complete for \mathfrak{P} , then

$$L_p(\mathfrak{X}, \mathcal{W}, \mathfrak{P}) \subset L_p(\mathfrak{X}, \mathcal{V}(\mathfrak{P}), \mathfrak{P})$$

and, accordingly, $\mathcal{V}(\mathfrak{P})$ is the maximal complete σ -algebra for \mathfrak{P} .

(c) Let $\mathcal{C} \subset \mathcal{A}$ be sufficient for \mathfrak{P} , then

$$L_p(\mathfrak{X}, \mathcal{V}(\mathfrak{P}), \mathfrak{P}) \subset L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$$

with equality, if \mathcal{C} is also complete. Therefore, a σ -algebra that is complete and sufficient is \mathfrak{P} -almost surely unique and minimally sufficient. \blacklozenge

²¹Let $(\mathfrak{X}, \mathcal{T})$ be a topological space and let $M \subset \mathfrak{X}$ be a subset. Then, we denote the closure of M w. r. t. \mathcal{T} by $\text{cl}(M)_{\mathcal{T}}$.

Proof: For part (a), we note that $\mathcal{V}(\mathfrak{F})$ is an intersection stable Dynkin system and, hence, a σ -algebra. The completeness follows by the explicit definition of $\mathcal{V}(\mathfrak{F})$ (see also Kagan *et al.*, 2014, Theorem 3.1).

(b) Let $\mathcal{W} \subset \mathcal{A}$ be an arbitrary complete σ -algebra for \mathfrak{F} . By the completeness of \mathcal{W} , we then get

$$\mathbf{A}_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{F}; \mathbb{R}) \subset \bigcap_{P \in \mathfrak{F}} \ker(\mathbf{E}_P(\cdot | \mathcal{W})).$$

Hence, we have for any $W \in \mathcal{W}$

$$\langle \mathbb{1}_W, g \rangle_P = \int \mathbb{1}_W \underbrace{\mathbf{E}_P(g | \mathcal{W})}_{=0} dP = 0 \quad \text{for all } g \in \mathbf{A}_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{F}; \mathbb{R}), P \in \mathfrak{F}$$

that is, $W \in \mathcal{V}(\mathfrak{F})$ and consequently $\mathcal{W} \subset \mathcal{V}(\mathfrak{F})$. Clearly, this implies that

$$\mathbf{L}_p(\mathfrak{X}, \mathcal{W}, \mathfrak{F}; \mathbb{R}^k) \subset \mathbf{L}_p(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F}; \mathbb{R}^k).$$

(c) Let $\mathcal{C} \subset \mathcal{A}$ be an arbitrary sufficient σ -algebra for \mathfrak{F} . Then by Theorem 1.27 (b), we have $\ker(\mathbf{E}(\cdot | \mathcal{C})) \subset \mathbf{A}_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{F}; \mathbb{R})$. By definition of $\mathcal{V}(\mathfrak{F})$, we have that any $V \in \mathcal{V}(\mathfrak{F})$ fulfills $\langle \mathbb{1}_V, g \rangle_P = 0$ for all $g \in \mathbf{A}_1^0(\mathfrak{X}, \mathcal{A}, \mathfrak{F}; \mathbb{R})$ and all $P \in \mathfrak{F}$, in particular,

$$\langle \mathbb{1}_V, g \rangle_P = 0 \quad \text{for all } g \in \ker(\mathbf{E}(\cdot | \mathcal{C}))$$

and all $P \in \mathfrak{F}$. Hence, the theorem of Hahn–Banach (see also Theorem 1.27 and the proof to Proposition 1.30 (b)) implies that

$$\mathbb{1}_V + [\mathfrak{F}] \in \mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}; \mathbb{R}) = \bigcap \{ \ker \langle \cdot, g \rangle_P : g \in \ker(\mathbf{E}(\cdot | \mathcal{C})) \}.$$

Consequently, there is a $V' \in \mathcal{C}$ such that $\mathbb{1}_V = \mathbb{1}_{V'}$ almost surely w. r. t. \mathfrak{F} , that is, $\mathcal{V}(\mathfrak{F}) \subset \mathcal{C}$ almost surely w. r. t. \mathfrak{F} . This corresponds to

$$\mathbf{L}_p(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F}; \mathbb{R}^k) \subset \mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}; \mathbb{R}^k).$$

If \mathcal{C} is complete as well, we also have

$$\mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}; \mathbb{R}^k) \subset \mathbf{L}_p(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F}; \mathbb{R}^k),$$

due to (b), and finally the claimed equality and uniqueness. \blacksquare

For $p = 2$ the Rao–Blackwellization w. r. t. an arbitrary sufficient sub- σ -algebra $\mathcal{C} \subset \mathcal{A}$ has the following well-known geometric interpretation.

1.33 Proposition (Best Approximation w. r. t. Least Squares Errors) Let $\mathcal{C} \subset \mathcal{A}$ be a sufficient sub- σ -algebra for \mathfrak{F} and let $f \in \mathbf{L}_2(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$. Then the Rao–Blackwellization $\mathbf{E}(f | \mathcal{C})$ is the unique solution of

$$\operatorname{argmin} \left\{ \sup \{ \|f - Z\|_{P,2}^2 : P \in \mathfrak{F} \} : Z \in \mathbf{L}_2(\mathfrak{X}, \mathcal{C}, \mathfrak{F}) \right\}$$

in $\mathbf{L}_2(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$. \blacklozenge

Proof: To begin with, we recall that $\mathbf{E}(f|\mathcal{C}) = \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C})$. For $\varphi \in \mathbf{E}(f|\mathcal{C})$ we note that $\varphi + [P] = \mathbf{E}_P(f|\mathcal{C})$ is the unique solution of

$$\operatorname{argmin}\{\|f - Z\|_{P,2}^2 : Z \in \mathfrak{L}_2(\mathfrak{X}, \mathcal{C}, P)/[P]\},$$

where $P \in \mathfrak{P}$. Hence, $\mathbf{E}(f|\mathcal{C})$ is the common solution of

$$\operatorname{argmin}\{\|f - Z\|_{P,2}^2 : Z \in \mathfrak{L}_2(\mathfrak{X}, \mathcal{C}, P)/[\mathfrak{P}]\}$$

for every $P \in \mathfrak{P}$, which yields the claim. \blacksquare

The Rao–Blackwellization of some $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ is defined as the sufficient part of the decision procedure f . This statistical point of view was extended by Theorem 1.27 which yields that $f - \mathbf{E}(f|\mathcal{C})$ is ancillary of first order. The following proposition interprets that component of f as a corresponding best approximation w. r. t. ancillarity of first order.

1.34 Proposition (Best Approximation w. r. t. Ancillarity of First Order) Let $\mathcal{C} \subset \mathcal{A}$ be the complete and sufficient sub- σ -algebra for \mathfrak{P} , then the Rao–Blackwellization of $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ given \mathcal{C} is the unique solution of

$$\operatorname{argmin}\left\{\sup\{\|\int f - Z \, dP\|_{\mathbb{R}^k} : P \in \mathfrak{P}\} : Z \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})\right\} \quad (1.15)$$

in $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$. \blacklozenge

Proof: Note that $\mathbf{E}(f|\mathcal{C}) = \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C})$ is the unique element in the space $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ that fulfills

$$\int f \, dP = \int \mathbf{E}(f|\mathcal{C}) \, dP \quad \text{for all } P \in \mathfrak{P}$$

and, thus,

$$\sup\{\|\int f - \mathbf{E}(f|\mathcal{C}) \, dP\|_{\mathbb{R}^k} : P \in \mathfrak{P}\} = 0.$$

Hence, $\mathbf{E}(f|\mathcal{C})$ is the unique solution to (1.15). \blacksquare

With regard to Proposition 1.33 and 1.34, we note the following: Let $\mathfrak{B}(\mathfrak{X}, \mathcal{A})$ be the set of all bounded and measurable functions $f : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, then we have that

$$\|f\|_{\mathfrak{B}(\mathfrak{P}),\infty} := \sup\{\|f\|_{P,2} : P \in \mathfrak{P}\} \quad \text{for } f \in \mathfrak{B}(\mathfrak{X}, \mathcal{A})/[\mathfrak{P}]$$

is a norm, while

$$\|f\|_{\mathbf{LS},\mathfrak{B}(\mathfrak{P}),\infty} := \sup\{\|\int f \, dP\| : P \in \mathfrak{P}\} \quad \text{for } f \in \mathfrak{B}(\mathfrak{X}, \mathcal{A})/[\mathfrak{P}]$$

is in general not a norm, but just a semi-norm on $\mathfrak{B}(\mathfrak{X}, \mathcal{A})/[\mathfrak{P}]$. The latter, however, is a norm on $\mathfrak{B}(\mathfrak{X}, \mathcal{C})/[\mathfrak{P}]$ if $\mathcal{C} \subset \mathcal{A}$ is complete. In Chapter 3, we

will introduce larger normed spaces which yields generalizations of these approximation statements.

Proposition 1.34 offers a new perspective which plays a central role in Part II. This is explained in the following concluding summary of this section, which is numbered for the purpose of easy reference.

1.35 Summary and Conclusion Let $(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ be a statistical space and assume that there is a complete sufficient σ -algebra $\mathcal{C} \subset \mathcal{A}$. The Rao–Blackwell projection $\mathbf{E}(\cdot|\mathcal{C})$ projects decision procedures onto $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$, which is complemented in $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ by $A_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, see Proposition 1.26 and 1.27. This means that $\mathbf{E}(\cdot|\mathcal{C})$ reduces the ancillary information to a minimum, since \mathcal{C} is complete and sufficient (see Proposition 1.32).

The theory on Rao–Blackwellization presented so far considers a sharp separation between sufficient and ancillary information. This, however, also means, if $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ has components $h \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ which are *non-trivially “almost” ancillary of first order*, then so has $\mathbf{E}(f|\mathcal{C})$. By non-trivially “almost” ancillary of first order, it is meant that for all $P \in \mathfrak{P}$

$$\left\| \int h \, dP \right\| \approx 0 \quad (1.16)$$

which is understood to be “small” compared to

$$\int \|h\| \, dP > 0.$$

Hence, in spite of (1.16), h still differs from $\mathbf{0}$ “significantly”. Therefore, $\mathbf{E}(f|\mathcal{C})$ might still have some useless noise in terms of h if f has such components.

In this thesis we propose a new, more general kind of Rao–Blackwellization that also accounts for a reduction of “almost” ancillary material through a regularization approach. For example, the optimization problem (1.15) is extended to take the noise of Z into account. We refer to Part II, page 43 ff., for details and a discussion of similar approaches. \blacklozenge

1.3 Exponential Families

Exponential families constitute a very important class of statistical models. This class is flexible enough for a broad range of situations in practice and much is known about sufficiency, ancillarity, and completeness for these models.

The goal of this section is to recall and provide some results on exponential families for easy reference. The presentation follows mainly Pfanzagl (1994, page 22 ff.). We start with the central definition.

1.36 Definition A statistical model \mathfrak{P} is called *exponential family* in $a_1, \dots, a_k : \mathfrak{P} \rightarrow \mathbb{R}$ and measurable $S_1, \dots, S_k : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ w. r. t. a σ -finite reference measure μ iff \mathfrak{P} has μ -densities of the form

$$x \mapsto c(P)g(x) \exp\left(\sum_{i=1}^k a_i(P)S_i(x)\right) \quad \text{for all } P \in \mathfrak{P},$$

where g is some non-negative measurable function on $(\mathfrak{X}, \mathcal{A})$. In this case a_1, \dots, a_k are called *canonical parameters*. \blacklozenge

In the following we may assume without loss of generality that the reference measure μ is equivalent to \mathfrak{P} . We also assume that a_1, \dots, a_k are affinely independent, that is,

$$\sum_{i=1}^k c_i a_i(P) = c_0 \quad \text{for all } P \in \mathfrak{P}$$

implies $c_i = 0$ for $i = 1, \dots, k$. Furthermore, we assume that the functions S_1, \dots, S_k are affinely μ -independent, that is,

$$\sum_{i=1}^k c_i S_i(x) = c_0 \quad \text{for } \mu\text{-almost all } x \in \mathfrak{X}$$

implies $c_i = 0$ for $i = 1, \dots, k$. In this case the exponential family is called *k-parametric* and furthermore it is said to have *full rank* if $a(\mathfrak{P}) \subset \mathbb{R}^k$ has a non-empty interior.

An exponential family \mathfrak{P} is by definition a parametrized statistical model. We consider the *greatest or natural parameter space* which is given by

$$\Theta := \left\{ a \in \mathbb{R}^k : \int g(x) \exp\left(\sum_{i=1}^k a_i S_i(x)\right) \mu(dx) < \infty \right\}.$$

The statistic $S := (S_1, \dots, S_p)$ is in fact sufficient.

1.37 Proposition Let \mathfrak{P} be an exponential family in $a = (a_1, \dots, a_k)$ and $S = (S_1, \dots, S_k)$. Then, S is sufficient for \mathfrak{P} . \blacklozenge

Proof: See Pfanzagl (1994, Theorem 1.6.9, page 25 f.). \blacksquare

In regular cases, S is complete and therefore the minimal sufficient statistic for \mathfrak{P} . A proof of the completeness of S uses the following lemma.

1.38 Lemma Let μ be a measure on $(\mathfrak{X}, \mathcal{A})$, $f : \mathfrak{X} \rightarrow \mathbb{C}$ and $T : (\mathfrak{X}, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable. Furthermore, define

$$A := \left\{ \alpha \in \mathbb{R} : \int |f(x)| \exp(\alpha T(x)) \mu(dx) < \infty \right\}$$

and assume that A has a non-empty interior, that is, $\text{int}(A) \neq \emptyset$. Then, the function defined by

$$\begin{aligned} \varphi : \text{int}(A) + i\mathbb{R} &\rightarrow \mathbb{C} \\ z &\mapsto \int f(x) \exp(zT(x)) \mu(dx) \end{aligned}$$

is holomorphic with

$$\frac{\partial}{\partial z} \int f(x) \exp(zT(x)) \mu(dx) = \int T(x) f(x) \exp(zT(x)) \mu(dx).$$

◆

Proof: See Pfanzagl (1994, Lemma 1.6.6, page 24). ■

1.39 Proposition Let \mathfrak{F} be an exponential family in $a = (a_1, \dots, a_k)$ and $S = (S_1, \dots, S_k)$. Furthermore, assume that $\Theta := \{a(P) : P \in \mathfrak{F}\} \subset \mathbb{R}^k$ has a non-empty interior. Then, S is complete for \mathfrak{F} . ◆

Proof: Use Lemma 1.38, induction and the uniqueness theorem for holomorphic functions (see also Pfanzagl, 1994, page 26). ■

CHAPTER 2

Estimation Procedures

The purpose of this chapter is to provide necessary parts of estimation theory. Often estimators are implicitly motivated by comparing the data hypothetically with distributions of the statistical model. This is considered in the first section. We then proceed with the evaluation of these estimators.

Throughout this section we consider a parametrized statistical model $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ on the sample space $(\mathfrak{X}, \mathcal{A})$, where $\Theta \subset \mathbb{R}^k$ is a collection of finite-dimensional parameters.

2.1 Implicitly Defined Estimators

Suppose we are able to measure some sort of discrepancy of our observation X to the possible underlying parameters $\vartheta \in \Theta$. By describing this measurement through a function $A : \mathfrak{X} \times \Theta \rightarrow \overline{\mathbb{R}}$, we then naturally look for a parameter which fits best to the data X . Let¹ $A(\cdot, \eta)$ be integrable w. r. t. P_ϑ for all $\vartheta, \eta \in \Theta$, we may then consider the expected discrepancy

$$\begin{aligned} \alpha : \quad \Theta \times \Theta &\rightarrow \mathbb{R} \\ (\vartheta, \eta) &\mapsto \mathbf{E}_\vartheta A(\cdot, \eta). \end{aligned}$$

By that α compares or *contrasts* the choice η with the underlying true parameter ϑ . If $\alpha(\vartheta, \cdot)$ is uniquely minimized for $\vartheta = \eta$, we call A a *contrast function* and α is called *expected contrast* of A . For $x \in \mathfrak{X}$, minimizers of $\vartheta \mapsto A(x, \vartheta)$, that is, elements of

$$\operatorname{argmin}\{A(x, \vartheta) : \vartheta \in \Theta\}$$

are called *minimum contrast estimates* of x . An estimator $\hat{\kappa} : (\mathfrak{X}, \mathcal{A}) \rightarrow (\Theta, \mathcal{B}(\Theta))$ such that $\hat{\kappa}(x) \in \operatorname{argmin}\{A(x, \vartheta) : \vartheta \in \Theta\}$ for all $x \in \mathfrak{X}$ is called *minimum contrast estimator* according to the contrast function A .

¹In what follows, we implicitly assume that A is measurable in the first argument.

This approach is the so-called *minimum contrast* approach and its notions are closely related to the decision theoretic ones.

2.1 Remark (Minimum Contrast and Decision Theory) Let $L : \Theta \times \Theta \rightarrow \mathbb{R}_+$ be a loss function with $L(\vartheta, \eta) = 0$ if and only if $\vartheta = \eta$. For an explicitly known estimator $\hat{\kappa}$, we have that $A : (x, \vartheta) \mapsto L(\hat{\kappa}(x), \vartheta)$ is a contrast function. The expected contrast is $\alpha : (\vartheta, \eta) \mapsto \mathbf{E}_\vartheta(L(\hat{\kappa}, \eta))$. Let $D := \{(\vartheta, \vartheta) : \vartheta \in \Theta\}$ be the diagonal in $\Theta \times \Theta$ then the restriction $\alpha|_D$ is the risk function. \blacklozenge

Alternatively, we may relax the requirement of the contrast function A to map in a totally ordered image space. Instead, we just consider a function $h : \mathfrak{X} \times \Theta \rightarrow \mathbb{R}^k$ which is assumed to be \mathcal{A} -measurable in the first component. Based on our observation $x \in \mathfrak{X}$, we obtain estimates as roots of $\vartheta \mapsto h(x, \vartheta)$, that is, by solving

$$h(x, \vartheta) = \mathbf{0} \quad (2.1)$$

for given x . The equation (2.1) is called *estimating equation*, the function h is called *inference function* or *estimating function*. An estimator $\hat{\kappa} : (\mathfrak{X}, \mathcal{A}) \rightarrow (\Theta, \mathcal{B}(\Theta))$ such that $h(x, \hat{\kappa}(x)) = \mathbf{0}$ for every $x \in \mathfrak{X}$ is called *estimator according to (the inference function) h* .

In the remaining part, we deal with the following class of estimating functions and estimators, respectively.

2.2 Definition (Unbiasedness) An estimating function h is said to be *unbiased* iff $\mathbf{E}_\vartheta h(\cdot, \vartheta) = \mathbf{0}$ for all $\vartheta \in \Theta$. An estimator $\hat{\kappa}$ is called *unbiased* for the parameter $\kappa : \Theta \rightarrow \mathbb{R}^k$ iff $(x, \vartheta) \mapsto (\hat{\kappa}(x) - \kappa(\vartheta))$ is an unbiased estimating function. \blacklozenge

Obviously, unbiased estimators for the canonical parameter \mathbf{id}_Θ can be represented in terms of unbiased estimating functions which are linear in \mathbf{id}_Θ . Conversely, we can see that estimators according to a “smooth” unbiased estimating function are at least approximately unbiased.

2.3 Remark (Approximately Unbiased Estimators) Let h be an unbiased estimating function such that $\vartheta \mapsto h(x, \vartheta)$ is continuously differentiable and let $\hat{\vartheta}$ be an estimator according to h . Then, the first order Taylor series expansion of $\vartheta \mapsto h(x, \vartheta)$ at $\hat{\vartheta}(x)$ yields that for $x \in \mathfrak{X}$

$$h(x, \vartheta) = h(x, \hat{\vartheta}(x)) + \mathbf{D}h(x, \hat{\vartheta}(x))(\vartheta - \hat{\vartheta}(x)) + \text{remainder}, \quad (2.2)$$

where $\mathbf{D}h(x, \hat{\vartheta}(x))$ denotes the Jacobi-matrix of $h(x, \cdot)$ at $\hat{\vartheta}(x)$. If we assume that $\mathbf{D}h(\cdot, \hat{\vartheta}(\cdot))$ and $\hat{\vartheta}$ are uncorrelated under each $\vartheta \in \Theta$, that is,

$$\mathbf{E}_\vartheta(\mathbf{D}h(\cdot, \hat{\vartheta}(\cdot))(\vartheta - \hat{\vartheta}(\cdot))) = \mathbf{E}_\vartheta(\mathbf{D}h(\cdot, \hat{\vartheta}(\cdot)))\mathbf{E}_\vartheta(\vartheta - \hat{\vartheta})$$

holds for all $\vartheta \in \Theta$, we have that (2.2) yields

$$\mathbf{0} = \mathbf{E}_\vartheta h(\cdot, \vartheta) \approx \mathbf{E}_\vartheta \mathbf{D}h(\cdot, \hat{\vartheta}(\cdot))\mathbf{E}_\vartheta(\vartheta - \hat{\vartheta}), \quad (2.3)$$

since $h(x, \hat{\vartheta}(x)) = \mathbf{0}$ for all $x \in \mathfrak{X}$ by definition of $\hat{\vartheta}$. The matrix $\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta)$ is called *sensitivity matrix* of h . If we assume furthermore that $\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta)$ is non-singular for all ϑ , we get by equation (2.3) that $\mathbf{E}_\vartheta(\vartheta - \hat{\vartheta}) \approx \mathbf{0}$ for all ϑ which means that $\hat{\vartheta}$ is approximately unbiased for \mathbf{id}_Θ . \blacklozenge

The minimum contrast and the estimating function approach may be related which we illustrate by the following example.

2.4 Example (Maximum Likelihood Estimation) For some non-empty, open parameter set $\Theta \subset \mathbb{R}^k$, consider a statistical model $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ that is dominated by a σ -finite measure μ . Let f_ϑ denote the corresponding density of P_ϑ w. r. t. μ for $\vartheta \in \Theta$. To a given observation $x \in \mathfrak{X}$, Fisher (1912, 1922) proposed to choose a parameter $\hat{\vartheta}_{\text{ML}}(x) \in \Theta$ which is most likely in the sense that $f_\vartheta(x) \leq f_{\hat{\vartheta}_{\text{ML}}(x)}(x)$ for all $\vartheta \in \Theta$. Given data $x \in \mathfrak{X}$, we therefore look for a maximizer of the so-called *likelihood*

$$\vartheta \mapsto L(x, \vartheta) := f_\vartheta(x)$$

or equivalently a maximizer of the *log-likelihood*

$$\vartheta \mapsto \ell(x, \vartheta) := \log f_\vartheta(x). \quad (2.4)$$

Each $\hat{\vartheta}_{\text{ML}}(x) \in \operatorname{argmax}\{L(x, \vartheta) : \vartheta \in \Theta\}$ is called *maximum likelihood estimate* for the data x , in the case of existence. If $\hat{\vartheta}_{\text{ML}}(x)$ exists for all $x \in \mathfrak{X}$ and if

$$\hat{\vartheta}_{\text{ML}} : (\mathfrak{X}, \mathcal{A}) \rightarrow (\Theta, \mathcal{B}(\Theta))$$

is measurable, the function $\hat{\vartheta}_{\text{ML}}$ is called *maximum likelihood estimator* (MLE) for \mathbf{id}_Θ . Under regularity conditions, the MLE can be obtained as a minimum contrast estimate as well as a root of an estimating equation. Here, we give a sketch of both approaches. Concerning the regularity conditions especially with regard to the existence and uniqueness of an MLE, we refer to the subsequent Proposition 2.5.

MLE as a Minimum Contrast Estimate. Defining $A(x, \vartheta) := -\ell(x, \vartheta)$ for $(x, \vartheta) \in \mathfrak{X} \times \Theta$ yields a contrast function, provided that $A(\cdot, \vartheta)$ is \mathfrak{P} -integrable for all $\vartheta \in \Theta$. The expected contrast function is

$$\alpha(\vartheta, \eta) := -\mathbb{E}_\vartheta \log f_\eta \quad \text{for } \vartheta, \eta \in \Theta.$$

Note that $\alpha(\vartheta, \eta)$ is minimized if and only if $\vartheta = \eta$. This is due to

$$\begin{aligned} \alpha(\vartheta, \eta) - \alpha(\vartheta, \vartheta) &= -(\mathbb{E}_\vartheta \log f_\eta - \mathbb{E}_\vartheta \log f_\vartheta) \\ &= -\mathbb{E}_\vartheta \log \frac{f_\eta}{f_\vartheta} \\ &\geq -\log \mathbb{E}_\vartheta \frac{f_\eta}{f_\vartheta} \\ &= 0, \end{aligned}$$

where equality holds if and only if $\vartheta = \eta$. The left-hand side is also called *Kullback-Leibler divergence* or *Kullback-Leibler information* which measures how different distributions are from another.

MLE as a Root of an Estimating Function. According to the initial motivation, we are looking for maximizer of (2.4). Hence, assuming that $\ell(x, \cdot)$ is differentiable for \mathfrak{P} -almost all $x \in \mathfrak{X}$, we are looking for roots of the so-called *score*

$$s(x, \vartheta) := \frac{\partial}{\partial \vartheta} \ell(x, \vartheta) \quad (2.5)$$

as a function in $\vartheta \in \Theta$ given the data $x \in \mathfrak{X}$. If we additionally assume that the likelihood is strictly concave, there is at most one root of (2.5) and hence at most one MLE. The score $s : \mathfrak{X} \times \Theta \rightarrow \mathbb{R}^k$ is therefore an estimating function that may be used to define an MLE. \blacklozenge

The existence of MLEs is generally not ensured, especially if the parameter space is not compact (see for example Liese & Miescke, 2008, Example 7.90, page 335). For the existence and computation of the MLE in exponential families, we note the following proposition.

2.5 Proposition (Exponential Family and MLE) Let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be an exponential family of full rank with canonical parameter \mathbf{id}_Θ , sufficient statistic $S = (S_1, \dots, S_k) \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, and let $\text{supp } \mathfrak{P}^S$ be the common support of \mathfrak{P}^S . Then, the log-likelihood is twice differentiable and

(a) the score is²

$$\frac{\partial}{\partial \vartheta} \ell(x, \vartheta) = S(x) - \mathbf{E}_\vartheta S \quad \text{for } \vartheta \in \text{int}(\Theta)$$

and, hence, an unbiased estimating function.

(b) The so-called (*observed*) Fisher information is

$$-\frac{\partial^2}{\partial \vartheta \partial \vartheta^\top} \ell(x, \vartheta) = \text{Var}_\vartheta S \quad \text{for } \vartheta \in \text{int}(\Theta).$$

(c) For $x \in \mathfrak{X}$ such that

$$S(x) \in \text{int}(\text{cl}(\text{conv}(\text{supp } \mathfrak{P}^S))) \quad (2.6)$$

there is a unique MLE that is given as the root of the score. \blacklozenge

Proof: Barndorff-Nielsen (1978, Section 9.3, in particular Theorem 9.13, page 150 f.). \blacksquare

²For a subset M of \mathbb{R}^k , we recall that $\text{int}(M)$, $\text{cl}(M)$, and $\text{conv}(M)$ denote the interior, the closure, and the convex hull of M .

2.6 Remark Assume that (2.6) holds for \mathfrak{P} -almost all $x \in \mathfrak{X}$. Then, the MLE $\hat{\vartheta}_{\text{ML}}$ is defined except on the \mathfrak{P} -null set

$$\{S \text{ does not fulfill (2.6)}\} \in \mathcal{A}.$$

Furthermore, $\hat{\vartheta}_{\text{ML}}$ is $\sigma(S)$ -measurable. This is due to the fact that the set $\{S \text{ does not fulfill (2.6)}\}$ is obviously $\sigma(S)$ -measurable and for all

$$x \in \{S \text{ fulfills (2.6)}\}$$

the MLE exists uniquely as a root of the score, that is, $\hat{\vartheta}_{\text{ML}}(x)$ is given such that

$$\mathbf{E}_{\hat{\vartheta}_{\text{ML}}(x)} S = S(x).$$

Hence, $\hat{\vartheta}_{\text{ML}} \mathbb{1}_{\{S \text{ fulfills (2.6)}\}}$ is measurable w. r. t. $\sigma(S)$. \blacklozenge

2.7 Remark (MLE, Invariance Property) Consider a parametrized statistical model $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ and assume that the MLE $\hat{\vartheta}_{\text{ML}}$ for the canonical parameter exists. Furthermore, for some $H \subset \mathbb{R}^l$ let $\kappa : \Theta \rightarrow H$ be a measurable function. Then, $\kappa(\hat{\vartheta}_{\text{ML}})$ is the MLE for the transformed parameter $\kappa(\mathbf{id}_\Theta)$. In the literature this is known as the *invariance property* of the MLE. \blacklozenge

Consider a statistical model \mathfrak{P} which is misspecified, that is, \mathfrak{P} does not contain the distribution of our observation X . Hence, an estimator $\kappa(\hat{\vartheta}(X))$ of the parameter value of interest³ $\kappa(\mathbf{P}^X)$ may have some slight, however, still systematic error due to the misspecification. Assume that $\varphi : H \rightarrow H$ denotes a slight correction⁴ that eliminates this error, meaning on the one hand $\kappa(\mathbf{P}^X) = \varphi(\kappa(\vartheta))$ for some $\vartheta \in \Theta$ and on the other hand we have that, for example,

$$\|\eta - \varphi(\eta)\|_{\mathbb{R}^l} \quad \text{for } \eta \in H$$

is small, say. It is then desirable that our construction principle for the estimator $\hat{\kappa}$ is not much affected by such slight, hypothetical parameter correction. That means, our estimation method yields an estimator $\hat{\kappa}^\varphi$ for the corrected parameter function $\varphi \circ \kappa$ such that, for example,

$$\int \|\hat{\kappa} - \hat{\kappa}^\varphi\|_{\mathbb{R}^l} dP_\vartheta \quad \text{for } \vartheta \in \Theta$$

is small, too. In this sense the construction principle is desired to exhibit a stability. Since this notion of stability is crucial for the present thesis we specify the concept in the following definition in a rather general context.

³We assume here that the parameter function $\kappa : \Theta \rightarrow H$ for some $H \subset \mathbb{R}^l$ has an extension that is also denoted by κ such that $\kappa(\mathbf{P}^X)$ can be considered.

⁴In practice, this correction is typically unknown.

2.8 Definition (Stability of Estimator Constructions) Let \mathfrak{K} be a set of parameter functions $\kappa : \Theta \rightarrow \mathbb{R}^l$ and let \mathcal{K} be a topology on \mathfrak{K} . Furthermore, consider some topology \mathcal{T} on the space of estimators $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$.⁵ We call a construction principle of estimators as a map from $(\mathfrak{K}, \mathcal{K})$ to $(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathcal{T})$ (weakly) *stable* (w. r. t. \mathcal{K} and \mathcal{T}) iff it is (sequentially) continuous. \blacklozenge

The invariance property of the maximum likelihood principle is (weakly) stable for corresponding choices of $(\mathfrak{K}, \mathcal{K})$.

2.9 Example (MLE, Stability) Consider the situation of Remark 2.7 and a set of parameter functions $\mathfrak{K} \subset L_1(\Theta, \mathcal{B}(\Theta), \mathfrak{P}^{\hat{\vartheta}_{\text{ML}}})$ such that $\|\kappa\| \leq m$ for all $\kappa \in \mathfrak{K}$ and a $\mathfrak{P}^{\hat{\vartheta}_{\text{ML}}}$ -integrable measurable function m . Consider the topology of pointwise convergence⁶ $\mathcal{K} := \mathcal{T}_{\mathfrak{p}}(\mathfrak{K})$ on \mathfrak{K} . According to Remark 2.7 the MLE construction principle is given by the map

$$\begin{aligned} (\mathfrak{K}, \mathcal{K}) &\rightarrow (L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathcal{T}_{L_1(\mathfrak{P})}) \\ \kappa &\mapsto \kappa \circ \hat{\vartheta}_{\text{ML}}. \end{aligned} \quad (2.7)$$

One can easily see that due to Lebesgue's dominated convergence theorem (2.7) is sequentially continuous and hence the maximum likelihood principle is weakly stable.

Let $(\mathfrak{K}, \mathcal{K}) := (\mathcal{B}(\Theta, \mathcal{B}(\Theta)), \|\cdot\|_{\infty})$, where $\mathcal{B}(\Theta, \mathcal{B}(\Theta))$ denotes the set of all measurable, bounded functions from Θ to \mathbb{H} and $\|\cdot\|_{\infty}$ denotes the sup norm. Then, (2.7) is continuous, since

$$\int \|\kappa'(\hat{\vartheta}_{\text{ML}}(x)) - \kappa(\hat{\vartheta}_{\text{ML}}(x))\| P_{\vartheta}(\mathrm{d}x) \leq \|\kappa' - \kappa\|_{\infty}$$

for all $\vartheta \in \Theta$. \blacklozenge

In Part II the set \mathfrak{K} will consist of unbiasedly estimable parameter functions. Furthermore, we will see that the method of unbiased estimation is in general not (weakly) stable, neither w. r. t. the topology of pointwise convergence nor w. r. t. the norm topology $\tau(\|\cdot\|_{\infty})$.

If an MLE exists, it is in most cases not explicitly known. Beyond that, the numerical computation of this estimator involves some intractabilities and is therefore quite cumbersome.

This is often the case in the situation of a Gibbs point process model which we discuss in Part III. For this reason, alternatives like the *max-*

⁵As an example we may mention $\mathcal{T} = \mathcal{T}_{L_p(\mathfrak{P})}$. In the next part, we will also introduce and consider norm topologies for \mathcal{T} .

⁶Let F be a space of functions with values in a normed space $(E, \|\cdot\|_E)$ and with the domain \mathfrak{X} . We denote by $\mathcal{T}_{\mathfrak{p}}(F) = \tau(f \mapsto \|f(x)\|_E : x \in \mathfrak{X})$ the *topology of pointwise convergence*.

imum pseudo-likelihood, the *logistic regression*, and the *variational estimator* were introduced through an unbiased estimating function, see Chapter 7.

2.2 Evaluation of Estimation Procedures

By the nature of an estimation problem (\mathfrak{X}, κ) we are interested in a procedure that yields an estimator $\hat{\kappa}$ such that the distribution $P_{\vartheta}^{\hat{\kappa}}$ is concentrated around $\kappa(\vartheta)$ as well as possible, uniformly for every $\vartheta \in \Theta$. This is clearly a challenging task, not to mention that the notion “concentration” has to be formalized in terms of a distributional property of the estimator $\hat{\kappa}$.

To this end, we consider the above-mentioned concentration aspect in terms of the decision theoretic situation given on page 9.

2.10 Definition Let $\hat{\kappa}, \hat{\lambda} \in L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ be two estimators and let \mathcal{L} be a set of loss functions that are convex in the first argument. Then, $\hat{\kappa}$ is preferred to $\hat{\lambda}$ w. r. t. \mathcal{L} iff

$$\mathbf{E}_{\vartheta}(L(\hat{\kappa}(\cdot), \vartheta)) \leq \mathbf{E}_{\vartheta}(L(\hat{\lambda}(\cdot), \vartheta)) \quad \text{for all } \vartheta \in \Theta, L \in \mathcal{L}.$$

In this case we write $\hat{\kappa} \leq_{\mathcal{L}} \hat{\lambda}$. ◆

In practice, there might be a “true” loss function which is, however, often unknown. It is therefore reasonable to have an ordering w. r. t. a class of loss function as it is introduced in the above definition. The following example yields a popular measure of concentration.

2.11 Example (Weighted Mean Squared Error) Let $\hat{\kappa}, \hat{\lambda} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ be two estimators. We define for each $u \in \mathbb{R}^k$ the convex loss function

$$L_u(x, \vartheta) := (u^{\top}(x - \kappa(\vartheta)))^2 \quad \text{for } x \in \mathbb{R}^k, \vartheta \in \Theta.$$

By $L_u(x, \vartheta)$ we measure the u -weighted mean squared error. Set $\mathcal{L} := \{L_u : u \in \mathbb{R}^k\}$, then $\hat{\kappa} \leq_{\mathcal{L}} \hat{\lambda}$ if and only if

$$\begin{aligned} \mathbf{E}_{\vartheta}(L_u(\hat{\kappa}(\cdot), \vartheta)) &= u^{\top} \left(\int (\hat{\kappa}(x) - \kappa(\vartheta))(\hat{\kappa}(x) - \kappa(\vartheta))^{\top} P_{\vartheta}(dx) \right) u \\ &\leq u^{\top} \left(\int (\hat{\lambda}(x) - \kappa(\vartheta))(\hat{\lambda}(x) - \kappa(\vartheta))^{\top} P_{\vartheta}(dx) \right) u \\ &= \mathbf{E}_{\vartheta}(L_u(\hat{\lambda}(\cdot), \vartheta)) \end{aligned}$$

holds for all $u \in \mathbb{R}^k$ and $\vartheta \in \Theta$. Then, the $k \times k$ -matrix

$$\text{MSE}(\hat{\kappa}, \vartheta) := \int (\hat{\kappa}(x) - \kappa(\vartheta))(\hat{\kappa}(x) - \kappa(\vartheta))^{\top} P_{\vartheta}(dx) \quad (2.8)$$

is called (*multivariate*) *mean squared error* (MSE) of $\hat{\kappa}$ under $\vartheta \in \Theta$. Furthermore, $\hat{\kappa} \leq_{\mathcal{L}} \hat{\lambda}$ holds if and only if $\text{MSE}(\hat{\kappa}, \vartheta) \leq_{\mathbb{L}} \text{MSE}(\hat{\lambda}, \vartheta)$ for all $\vartheta \in \Theta$, where $\leq_{\mathbb{L}}$ denotes the *Löwner order*⁷. \blacklozenge

The MSE makes use of the estimator in an explicit form. This is somewhat unfortunate, since estimators are often defined implicitly (see also previous section). For that reason a generalized concept of the MSE for estimating functions are needed. Instead of using the second moment of the estimating function $(x, \vartheta) \mapsto (\hat{\kappa}(x) - \kappa(\vartheta))$, we henceforth generally consider

$$\mathbf{E}_{\vartheta} h(\cdot, \vartheta) h(\cdot, \vartheta)^{\top} = \int h(x, \vartheta) h(x, \vartheta)^{\top} P_{\vartheta}(\mathrm{d}x), \quad (2.9)$$

where h is the estimating function which determines the estimator $\hat{\kappa}$. A comparison of estimating procedures according to (2.9) is reasonable as long as we take the sensitivity of the estimating function in ϑ into account. We assume that $\vartheta \mapsto h(x, \vartheta)$ is differentiable for all $x \in \mathfrak{X}$ and we require that (2.9) is normalized by

$$\mathbf{E}_{\vartheta} \mathbf{D} h(\cdot, \vartheta) := \left(\mathbf{E}_{\vartheta} \frac{\partial}{\partial \vartheta_i} h_j(\cdot, \vartheta) \right)_{1 \leq i, j \leq k}, \quad (2.10)$$

where we also assume that (2.10) exists. For a further discussion of an evaluation concept for estimating function, we consider the following class of regular ones. To this end, we especially consider a parametrized statistical model $\mathfrak{P} = \{P_{\vartheta} : \vartheta \in \Theta\}$ which is dominated by μ . The corresponding densities of P_{ϑ} w. r. t. μ are denoted by f_{ϑ} for $\vartheta \in \Theta$.

2.12 Definition (Regular Estimating Functions) An unbiased estimating function h is called *regular* iff

- (a) $\vartheta \mapsto h(x, \vartheta)$ is differentiable for all $x \in \mathfrak{X}$;
- (b) the order of integration and differentiation can be interchanged, that is,

$$\frac{\partial}{\partial \vartheta_j} \int \varphi(x) h(x, \vartheta) f_{\vartheta}(x) \mu(\mathrm{d}x) = \int \varphi(x) \frac{\partial}{\partial \vartheta_j} (h(x, \vartheta) f_{\vartheta}(x)) \mu(\mathrm{d}x)$$

for $j = 1, \dots, k$;

- (c) the matrices $\mathbf{E}_{\vartheta} \mathbf{D} h(\cdot, \vartheta)$ and $\mathbf{E}_{\vartheta} h(\cdot, \vartheta) h(\cdot, \vartheta)^{\top}$ exist and are non-singular for all $\vartheta \in \Theta$.

The set of all regular estimating functions is denoted by $\mathfrak{H} = \mathfrak{H}(\mathfrak{P})$. \blacklozenge

⁷For a definition see Footnote 4 of Chapter 1 on page 11.

2.13 Example (Score Function, Exponential Family) Let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be an exponential family of full rank, with sufficient statistic $S \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, and open parameter set $\Theta \subset \mathbb{R}^k$. Then, the score function

$$s(x, \vartheta) = \frac{\partial}{\partial \vartheta} \ell(x, \vartheta) = S(x) - \mathbf{E}_\vartheta S \quad \text{for } x \in \mathfrak{X}, \vartheta \in \Theta$$

is a regular unbiased estimating function, since: according to Proposition 2.5 we have that $\mathbf{E}_\vartheta s(\cdot, \vartheta) = 0$ for all $\vartheta \in \Theta$ and that $s(x, \cdot)$ is differentiable for \mathfrak{P} -almost all $x \in \mathfrak{X}$. Furthermore, we have

$$\begin{aligned} \mathbf{E}_\vartheta s(\cdot, \vartheta) s(\cdot, \vartheta)^\top &= \mathbf{E}_\vartheta (S - \mathbf{E}_\vartheta S)(S - \mathbf{E}_\vartheta S)^\top \\ &= \text{Var}_\vartheta S \end{aligned}$$

and Proposition 2.5 (c) directly yields $\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta) = \text{Var}_\vartheta S$. Note that $\text{Var}_\vartheta S$ is positive definite for all $\vartheta \in \Theta$ since the exponential family is of full rank. Finally, we note that (see also Lemma 1.38)

$$\begin{aligned} \frac{\partial}{\partial \vartheta_j} \int \varphi(x) (\langle \vartheta, S \rangle - \log c(\vartheta)) f_\vartheta(x) \mu(dx) \\ = \int \varphi(x) \frac{\partial}{\partial \vartheta_j} ((\langle \vartheta, S \rangle - \log c(\vartheta)) f_\vartheta(x)) \mu(dx) \end{aligned}$$

for $j = 1, \dots, k$, that is, the order of integration and differentiation can be interchanged. \blacklozenge

For a regular estimating function $h \in \mathfrak{H}$, we define

$$h^{(s)} := -(\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta))^\top \left(\mathbf{E}_\vartheta h(\cdot, \vartheta) h(\cdot, \vartheta)^\top \right)^{-1} h$$

which is called *standardization* of h .

The following traces back to Godambe (1960).

2.14 Definition For $h \in \mathfrak{H}$ we set for $\vartheta \in \Theta$

$$\mathbf{G}_\vartheta(h) := (\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta))^\top \left(\mathbf{E}_\vartheta h(\cdot, \vartheta) h(\cdot, \vartheta)^\top \right)^{-1} (\mathbf{E}_\vartheta \mathbf{D}h(\cdot, \vartheta)) \quad (2.11)$$

which is called the (*generalized*) *Godambe information* of h . \blacklozenge

2.15 Example (Godambe's Efficiency Criterion) Let $h, g \in \mathfrak{H}$ be two regular estimating functions. Similar to Example 2.11 we consider for each $u \in \mathbb{R}^k$ the loss of h given by $L_u(h(x, \vartheta)) := (u^\top h(x, \vartheta))^2$ for $x \in \mathbb{R}^k$ and $\vartheta \in \Theta$. Set $\mathcal{L} := \{L_u : u \in \mathbb{R}^k\}$, we then write $h \leq_{\mathcal{L}} g$ iff

$$\mathbf{E}_\vartheta (L_u(h^{(s)}(\cdot, \vartheta))) \geq \mathbf{E}_\vartheta (L_u(g^{(s)}(\cdot, \vartheta)))$$

for the standardized estimating functions $h^{(s)}$ and $g^{(s)}$. That also means

$$\begin{aligned} u^\top \mathbf{G}_\vartheta(h)u &= u^\top \mathbf{E}_\vartheta \left(h^{(s)}(\cdot, \vartheta) h^{(s)}(\cdot, \vartheta)^\top \right) u \\ &= \mathbf{E}_\vartheta \left(L_u(h^{(s)}(\cdot, \vartheta)) \right) \\ &\geq \mathbf{E}_\vartheta \left(L_u(g^{(s)}(\cdot, \vartheta)) \right) \\ &= u^\top \mathbf{E}_\vartheta \left(g^{(s)}(\cdot, \vartheta) g^{(s)}(\cdot, \vartheta)^\top \right) u \\ &= u^\top \mathbf{G}_\vartheta(g)u \end{aligned}$$

for all $u \in \mathbb{R}^k$ and hence we have $\mathbf{G}_\vartheta(g) \leq_L \mathbf{G}_\vartheta(h)$ for all $\vartheta \in \Theta$. \blacklozenge

We look for an estimating function h with large Godambe information $\mathbf{G}_\vartheta(h)$ for $\vartheta \in \Theta$. Therefore, we arrive at the following optimality criterion.

2.16 Definition Let $\mathfrak{G} \subset \mathfrak{H}$ be some subset of regular estimating functions. We call $h \in \mathfrak{G}$ *optimal* in \mathfrak{G} iff $\mathbf{G}_\vartheta(g) \leq_L \mathbf{G}_\vartheta(h)$ holds for all $\vartheta \in \Theta$ and all $g \in \mathfrak{G}$. \blacklozenge

2.3 Optimal Estimation Procedures

In the previous section we have introduced evaluation concepts of estimation procedures. This section is concerned with the optimality with respect to the introduced partial orders.

To this end, we first state a popular result concerning the improvement of estimators that is due to Rao (1945) and Blackwell (1947). The result was extended to the multivariate case and arbitrary convex loss functions by Hodges & Lehmann (1950).

2.17 Theorem (Rao–Blackwell) Let $\hat{\kappa} \in L_1(\mathfrak{X}, \mathfrak{A}, \mathfrak{P}; \mathbb{R}^k)$ be an estimator for (\mathfrak{Y}, κ) and let $\mathcal{C} \subset \mathfrak{A}$ be sufficient for \mathfrak{P} . Furthermore, let \mathcal{L} be the set of all convex loss functions. Then,

$$\mathbf{E}(\hat{\kappa}|\mathcal{C}) \leq_{\mathcal{L}} \hat{\kappa},$$

that is, the Rao–Blackwellization⁸ of $\hat{\kappa}$ yields an improvement with respect to $\leq_{\mathcal{L}}$. \blacklozenge

Proof: Since \mathcal{C} is sufficient we may consider the Rao–Blackwellization of $\hat{\kappa}$, that is,

$$\mathbf{E}(\hat{\kappa}|\mathcal{C}) = \bigcap_{\vartheta \in \Theta} \mathbf{E}_\vartheta(\hat{\kappa}|\mathcal{C})$$

⁸See Proposition 1.26 on page 17 f. for the Rao–Blackwell projection $\mathbf{E}(\cdot|\mathcal{C})$, in particular for the notion Rao-Blackwellization of an estimator.

which is an element in $L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$, see Proposition 1.26 on page 17. By Jensen's inequality we obtain

$$\mathbf{E}_\vartheta(L(\mathbf{E}(\widehat{\kappa}|\mathcal{C}), \vartheta)) \leq \int \mathbf{E}(L(\widehat{\kappa}, \vartheta)|\mathcal{C}) dP_\vartheta = \mathbf{E}_\vartheta L(\widehat{\kappa}, \vartheta)$$

for every $\vartheta \in \Theta$ and every $L \in \mathcal{L}$. \blacksquare

By the Rao–Blackwell Theorem, we obtain in particular for an arbitrary estimator $\widehat{\kappa}$

$$\text{MSE}(\mathbf{E}(\widehat{\kappa}|\mathcal{C}), \vartheta) \leq_L \text{MSE}(\widehat{\kappa}, \vartheta)$$

for all $\vartheta \in \Theta$ (see also Example 2.11).

Optimal Unbiased Estimators

A restriction to the class of unbiased estimators yields an optimality result that is due to Lehmann & Scheffé (1950).

2.18 Theorem (Lehmann–Scheffé) Let $\widehat{\kappa} \in L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ be an unbiased estimator for (\mathfrak{P}, κ) and let $\mathcal{C} \subset \mathcal{A}$ be complete and sufficient for \mathfrak{P} . Furthermore, let \mathcal{L} be the set of all convex loss functions. Then, the Rao–Blackwellization $\mathbf{E}(\widehat{\kappa}|\mathcal{C})$ is the optimal unbiased estimator w. r. t. $\leq_{\mathcal{L}}$, that is,

$$\mathbf{E}(\widehat{\kappa}|\mathcal{C}) \leq_{\mathcal{L}} \widehat{\lambda}$$

for any estimator $\widehat{\lambda}$ which is unbiased for (\mathfrak{P}, κ) . \blacklozenge

Proof: Since \mathcal{C} is sufficient, we have $\mathbf{E}(\widehat{\lambda}|\mathcal{C}) \leq_{\mathcal{L}} \widehat{\lambda}$ for every estimator, in particular for every unbiased estimator (see Theorem 2.17). Furthermore, there is at most one \mathcal{C} -measurable unbiased estimator, since \mathcal{C} is complete. Hence, $\mathbf{E}(\widehat{\kappa}|\mathcal{C}) = \mathbf{E}(\widehat{\lambda}|\mathcal{C})$ for every choice of (\mathfrak{P}, κ) -unbiased estimators $\widehat{\kappa}$ and $\widehat{\lambda}$. \blacksquare

Let $\mathcal{C} \subset \mathcal{A}$ be complete and sufficient, then the theorem of Lehmann–Scheffé states that $L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}^k)$ contains the optimal unbiased estimator of any unbiasedly estimable parameter. The converse does also hold which we want to show in the following. To this end, we consider the *expectation operator*

$$\begin{aligned} \mathbf{E} \quad L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) &\rightarrow (\mathbb{R}^k)^\Theta \\ \widehat{\kappa} &\rightarrow (\mathbf{E}_\vartheta \widehat{\kappa})_{\vartheta \in \Theta} = \left(\int \widehat{\kappa} dP_\vartheta \right)_{\vartheta \in \Theta} \end{aligned} \quad (2.12)$$

and state the following proposition that is due to Schmetterer & Strasser (1974).

2.19 Proposition For an estimator $\widehat{\kappa} \in L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ the following two statements are equivalent:

(a) $\widehat{\kappa} \in L_1(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F})$,

(b) $\widehat{\kappa}$ is an optimal unbiased estimator for $(\mathfrak{F}, \mathbf{E}\widehat{\kappa})$. \blacklozenge

Proof: The implication that (b) follows from (a) is shown in Schmetterer & Strasser (1974, Satz 1, page 3). Furthermore, by Schmetterer & Strasser (1974, Satz 2, page 3 f.) we have that (b) implies (a). \blacksquare

The following theorem is due to Bahadur (1957) and concerns, implicitly, the inversion of the expectation operator (2.12). We also refer to Schmetterer & Strasser (1974, Satz 5, page 7).

2.20 Theorem (Bahadur) Assume that any parameter $\kappa \in \text{ran}(\mathbf{E})$ has an optimal unbiased estimator $\widehat{\kappa}$. Then, there is a complete sufficient σ -algebra $\mathcal{C} \subset \mathcal{A}$ and the restricted expectation operator $\mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{F})}$ is invertible. \blacklozenge

Proof: According to the prerequisite, there is an operator

$$T : \text{ran}(\mathbf{E}) \rightarrow L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$$

that maps every parameter $\kappa \in \text{ran}(\mathbf{E})$ to the existing optimal unbiased estimator $\widehat{\kappa} := T\kappa$. We show that T is the inverse operator to a restriction of the expectation operator \mathbf{E} .

First we show that T is linear. To this end, we note that $\mathbf{E}T\kappa = \kappa$ by the required unbiasedness of $T\kappa$. Consider $\kappa, \kappa' \in \text{ran}(\mathbf{E})$ then $T\kappa + T\kappa'$ is the optimal unbiased estimator for $\kappa + \kappa'$, since

$$\mathbf{E}(T\kappa + T\kappa') = \mathbf{E}T\kappa + \mathbf{E}T\kappa' = \kappa + \kappa'$$

and $T\kappa + T\kappa'$ is measurable w. r. t. the complete σ -algebra $\mathcal{V}(\mathfrak{F})$ (see Proposition 2.19). In the same way we obtain $T(a\kappa) = aT\kappa$ for a scalar a and $\kappa \in \text{ran}(\mathbf{E})$. Hence, T is linear.

Proposition 2.19 and the prerequisite yield $\text{ran}(T) = L_1(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F})$. By Schmetterer & Strasser (1974, Satz 1, page 3) we also have that $\mathcal{V}(\mathfrak{F})$ is sufficient.

Finally, we have $T\mathbf{E}\widehat{\kappa} = \widehat{\kappa}$, since each $\widehat{\kappa} \in L_1(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F})$ is an optimal unbiased estimator for $(\mathfrak{F}, \mathbf{E}\widehat{\kappa})$, see again Proposition 2.19. That means, $T = (\mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{V}(\mathfrak{F}), \mathfrak{F})})^{-1}$. \blacksquare

For a complete and sufficient sub- σ -algebra $\mathcal{C} \subset \mathcal{A}$, we may consider the *construction principle of unbiased estimators* through the inverse expectation operator $\mathbf{E}_{\mathcal{C}}^{-1} := \mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{F})}^{-1}$. That is, each unbiased estimable parameter is mapped to an optimal unbiased estimator. In Part II we show that this construction principle is in general not stable. In other words, we will see that for $\kappa \in \text{ran}(\mathbf{E})$

$$\mathbf{E}_{\mathcal{C}}\widehat{\kappa} = \kappa \tag{2.13}$$

is usually an ill-posed problem. A slight change of the parameter function $\kappa \in \text{ran}(\mathbf{E})$ may have a large impact on the optimal unbiased estimator for the modified estimation problem. To the author's best knowledge, this was only noticed by Schmetterer (1977, page 313) and Schmetterer (1978, Section 3, page 493 f.) specifically for the case $(\mathfrak{R}, \mathcal{K}) := (\text{ran}(\mathbf{E}), \mathcal{T}_{\mathbf{p}})$.

In the present situation, the construction principle of optimal unbiased estimators is obviously an inverse problem, which we may call *inverse problem of the unbiased estimation theory*. Note that Proposition 1.34 (page 25) yields $\mathbf{E}(\cdot|\mathcal{C}) = \mathbf{E}_{\mathcal{C}}^{-1}\mathbf{E}$, that is, composing the right-hand side of (2.13) leads to the *Rao–Blackwell inverse problem* which will be considered systematically in the next part. Figure 2.1 summarizes these results concerning the theory of optimal unbiased estimation theory and also the relation to the Rao–Blackwell projection through a commutative diagram. The present thesis discusses the ill-posedness of the Rao–Blackwell inverse problem by studying the ill-posedness of the inverse problem of unbiased estimation. Regularizations of $\mathbf{E}(\cdot|\mathcal{C})$ will hence lead to a new concept of Rao–Blackwellization.

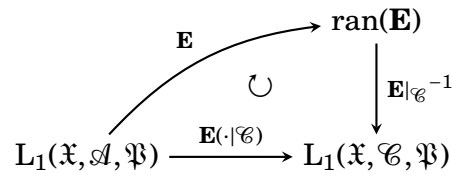


Figure 2.1: The theory of optimal unbiased estimation can be illustrated by this diagram. The arrow \circlearrowright indicates the commutativity of the diagram which follows by Proposition 1.34, page 25. Theorem 1.27 on page 18 concerns the Rao–Blackwellization $\mathbf{E}(\cdot|\mathcal{C})$ w. r. t. the complete and sufficient σ -algebra \mathcal{C} . Bahadur's converse theorem (Theorem 2.20) concerns the construction principle of unbiased estimators through the inversion of the restricted expectation operator $\mathbf{E}_{\mathcal{C}} := \mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{B})}$.

Optimal Unbiased Estimating Functions

The optimality considered so far was achieved just within the set of all unbiased estimators for a certain parameter κ . We now turn to an optimality result for unbiased estimating functions (see Godambe (1960) and Bhapkar (1972)). It gives a non-asymptotic, fixed sample size justification of the MLE for the canonical parameter \mathbf{id}_{Θ} .

2.21 Theorem Let $\mathfrak{B} = \{P_{\vartheta} : \vartheta \in \Theta\}$ be an exponential family of full rank and with sufficient statistic $S \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{B})$. The score s is the optimal estimating function within $\mathfrak{H}(\mathfrak{B})$, that is, $s \in \mathfrak{H}(\mathfrak{B})$ and

$$\mathbf{G}_{\vartheta}(h) \leq_L \mathbf{G}_{\vartheta}(s) \quad \text{for all } \vartheta \in \Theta$$

holds for all $h \in \mathfrak{H}(\mathfrak{P})$. ◆

Proof: See Song (2007, Theorem 3.12). ■

Theorem 2.21 states that the MLE $\hat{\vartheta}_{\text{ML}}$ for the canonical parameter is optimal with respect to the Godmabe information of its estimating function. Furthermore, if \mathcal{C} is complete and sufficient for \mathfrak{P} the present chapter yields also that $\mathbf{E}(\hat{\lambda}|\mathcal{C})$ is the *optimal* unbiased estimator for $(\mathfrak{P}, \mathbf{E}\hat{\vartheta}_{\text{ML}})$, where $\hat{\lambda}$ is any unbiased estimator for $(\mathfrak{P}, \mathbf{E}\hat{\vartheta}_{\text{ML}})$. In general, we have $\hat{\vartheta}_{\text{ML}} \neq \mathbf{E}(\hat{\lambda}|\mathcal{C})$ and, in fact, $\hat{\vartheta}_{\text{ML}}$ and $\mathbf{E}(\hat{\lambda}|\mathcal{C})$ can be quite different, although $\hat{\vartheta}_{\text{ML}}$ is often at least approximately unbiased (see also Remark 2.3). In this sense both optimality concepts are only loosely related. This is due to the ill-posedness of the *Rao–Blackwell inverse problem* which we examine in the next part.

PART II

Regularized Rao–Blackwellization

Parametric specifications of statistical models usually result from simplified assumptions about the underlying stochastic mechanism. That means, we often deal with misspecified models and statistical inference cannot go beyond the inaccuracy that comes from statistical modeling. We therefore want to make sure that the problem of constructing “good” estimators for a given parameter is stable. That means small errors in the parametric specification can be captured by at most small changes of the estimator (see also Definition 2.8). We may put it differently, by saying that our ambition is the well-posedness of the inverse problem inherently considered in unbiased estimation theory (see Section 2.3). However, this well-posedness does usually not hold.

Analyzing and tackling this problem constitutes the main contribution of the present thesis and is the subject of this part. First, we introduce a framework for demonstrating and analyzing the ill-posedness of the inverse problem that is given by (unbiased) estimation problems. Second, we present two regularizations through a Tikhonov and an Ivanov approach. In both cases, estimators are constructed by accounting for prior knowledge. This new *regularization* view point does not just concern the issues of model misspecification and computational accessibility of the estimators, but also supplements the discussion of optimal estimation procedures (see last paragraph of Section 2.3).

To clarify the intention and the role of this part, a more detailed outline of the general *problem*, the *state of the art*, and the *contribution* is given at this point. For details, references to the relevant passages in the subsequent chapters are provided.

Problem and State of the Art

Let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be a parametrized statistical model on a sample space $(\mathfrak{X}, \mathcal{A})$ and let $\mathcal{C} \subset \mathcal{A}$ be a sufficient and complete σ -algebra for \mathfrak{P} . Given an observation $x \in \mathfrak{X}$, we want to estimate an \mathbb{R}^k -valued parameter $\kappa : \Theta \rightarrow \mathbb{R}^k$, for example, the canonical parameter \mathbf{id}_Θ of the underlying

distribution P_ϑ . In fact, there are typically several estimators for κ , but “good” ones with “small” mean squared errors are wanted.

To this end, we first look at unbiased estimators. The theory of unbiased estimation concerns the inverse problem

$$\mathbf{E}_\vartheta \hat{\lambda} = \kappa(\vartheta) \quad \text{for all } \vartheta \in \Theta \quad (2.14)$$

that is usually ill-posed according to Hadamard’s well-known criteria of *existence*, *uniqueness*, and *continuous dependence on the parameter function* κ (see Hadamard (1923) and also *Hadamard (1902) according to e. g. Baumeister (1987, page 14)).

Existence of $\hat{\lambda}$ in (2.14), that is, existence of an unbiased estimator for κ , is usually not guaranteed (see e. g. Kolmogorov (1950) for a simple example regarding the Binomial model). Hence, we restrict ourself to $\kappa \in \text{ran}(\mathbf{E})$, that is, to unbiasedly estimable parameters. Then, however, (2.14) still has usually no unique solution, so we may take additional requirements into account. Since our interest concerns “good” estimators, for example, in terms of the mean squared error, the theorem of Rao–Blackwell and that of Lehmann–Scheffé tell us to look for a \mathcal{C} -measurable solution $\mathbf{E}(\hat{\lambda}|\mathcal{C})$ (see Rao (1945), Halmos (1946), Blackwell (1947), Lehmann & Scheffé (1950), Rao (1952), Bahadur (1957) for pioneering concepts and results, see also for example Barankin (1950), Hodges & Lehmann (1950) and Schmetterer & Strasser (1974) for extensions).

While these two problems, existence and uniqueness, were extensively discussed and studied by several authors long time ago, the problem of continuous dependence on the parameter κ was, to the best of the author’s knowledge, just mentioned in two conference contributions of Leopold Schmetterer (see Schmetterer (1977, 1978)). Probably, due to a lack of notions, concepts and, finally, interpretable results, regularizing (2.14) was not further examined as it appears. This is even more surprising in the light of the following problem we are concerned with in practice.

Question How to compute the Rao–Blackwellization $\mathbf{E}(\hat{\lambda}|\mathcal{C})$? ◆

In most cases an exact computation is not just analytically, but also numerically intractable, which is due to the instability of (2.14).¹ More precisely, $\mathbf{E}(\hat{\lambda}|\mathcal{C})$ does not depend continuously on the parameter $(\mathbf{E}_\vartheta \hat{\lambda})_{\vartheta \in \Theta}$, which

¹See also Fu & Li (1992, 1998) for an iterative Monte Carlo approach to compute the Rao–Blackwellization of an initial estimator w. r. t. a complete sufficient statistic numerically. The number of required samples depend implicitly on the parameter of interest. Due to the described instability, that means that arbitrarily many samples may be required to achieve a certain predetermined accuracy. However, sampling from distributions which model complex dependences, for example, certain Gibbs point process distributions (see Chapter 6), is computationally expensive. In addition, there is a need for discretization if the distribution of the complete sufficient statistic is continuous.

is quite unfortunate in view of the misspecification of the statistical model that we usually have to deal with. We therefore look for a substitute, that is, a *regularized Rao–Blackwellization*.

Main Contributions

In Chapter 3, we first of all introduce an appropriate topological and analytical structure for discussing the inverse problem we are concerned with, which we call *Rao–Blackwell inverse problem*. Since the required structure is closely related to the notion of a complete σ -algebra introduced by Lehmann & Scheffé (1950), we will speak of *Lehmann–Scheffé topologies* and *spaces*. We show that the Lehmann–Scheffé topologies are related to the well-known weak topologies, especially to these on Lebesgue spaces (see for example Theorem 3.9 and 3.10). Furthermore, we show that the canonical topology $\mathcal{T}_{L_p(\mathfrak{P})}$ rather relates to norm topologies (see Theorem 3.11). The corresponding ill-posedness of the inverse problem is demonstrated by examples.

These considerations then provide the basis for the regularization approaches used in Chapter 4, that lead to *regularized Rao–Blackwellizations*. Here, the crucial starting point is given by Theorem 4.1 on page 70, which essentially states for $p \in [1, \infty]$ and some sufficiently integrable $\hat{\lambda}$

$$\mathbf{E}(\hat{\lambda}|\mathcal{C}) = \mathbf{E}_{\mathcal{C}}^{-1} \mathbf{E} \hat{\lambda} = \operatorname{argmin}\{\|\mathbf{E}(\hat{\lambda} - g)\|_{\pi,p} : g \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})\}, \quad (2.15)$$

where π is some (probability) measure on $(\Theta, \mathcal{B}(\Theta))$ and $\mathcal{C} \subset \mathcal{A}$ fulfills a technical condition which goes slightly beyond complete sufficiency. We may then include prior information about the estimator we have in mind by adapting the variational problem in (2.15) as described in the two following approaches.

Tikhonov’s regularization approach accounts for regularity conditions through semi-norms, in addition to the minimization problem (2.15). More precisely, we take the oscillation or variability of a solution into account through a variation operator \mathbf{V} . By Theorem 4.3 on page 71 a Tikhonov regularized Rao–Blackwellization is provided, given by

$$\begin{aligned} \mathbf{E}^{(\alpha)}(\hat{\lambda}|\mathcal{C}) &= (\alpha \mathbf{V}' \mathbf{V} + \mathbf{E}' \mathbf{E})^{-1} \mathbf{E}' \mathbf{E} \hat{\lambda} \\ &= \operatorname{argmin}\{\|\mathbf{E}(\hat{\lambda} - g)\|_{\pi,2}^2 + \alpha \|\mathbf{V}g\|_{\pi \otimes \mathbf{id}_{\mathfrak{P}},2}^2 : g \in L_2(\mathfrak{X}, \mathcal{C}, \mathfrak{P})\}, \end{aligned}$$

where $\alpha > 0$ is some regularization parameter that accounts for the model misspecification.

This is examined for the i. i. d. Gaussian model $\mathfrak{P} = \{\mathbf{N}(a, \sigma^2)^{\otimes n} : a \in \mathbb{R}\}$ where $\sigma^2 > 0$ is fixed and the canonical parameter $\mathbf{id}_{\mathbb{R}}$ is of interest. Considering a Gaussian prior $\pi = \mathbf{N}(b, \tau^2)$ and the complete sufficient statistic $S(x) := \sum_{i=1}^n x_i$ for $x \in \mathbb{R}^n$, the regularized Rao–Blackwellization

of the median $\text{Med}(X)$ can be computed (see Example 4.5 and 4.6, page 74 ff.) as

$$\mathbf{E}^{(\alpha)}(\text{Med}(X)|S) = \frac{\tau^2}{\alpha \frac{\sigma^2}{n} + \tau^2} \left(\frac{1}{n}S\right) + \frac{\alpha \frac{\sigma^2}{n}}{\alpha \frac{\sigma^2}{n} + \tau^2} b.$$

Hence, as the model misspecification increases, the prior gets more involved in the regularized Rao–Blackwell projection process. Specifically, setting $\alpha = 0$ we obtain the maximum likelihood estimator, while $\alpha = 1$ leads to the Bayes estimator (w. r. t. squared error loss) for $\mathbf{id}_{\mathbb{R}}$.

Generally, the Tikhonov regularized Rao–Blackwellization for $\alpha = 1$ has a close relation to the Bayes estimator $\hat{\vartheta}_{\text{Bayes}}$ (w. r. t. squared error loss) for the canonical parameter. It is shown that (see Theorem 4.7 on page 80 for technical details)

$$\mathbf{E}^{(1)}(\hat{\lambda}|S) = \int \mathbf{E}_{\vartheta} \hat{\lambda} \Pi_{\pi}(S, dP) \quad \text{for } \hat{\lambda} \in L_2(\mathfrak{X}, \mathfrak{A}, Q),$$

where S is a sufficient statistic that fulfills a completeness condition, $\Pi(S, \cdot)$ is the posterior distribution and $Q := \int P_{\vartheta}(\cdot) \pi(d\vartheta)$. Furthermore, we have (see Theorem 4.10 on page 84)

$$\|\mathbf{E}^{(1)}(\hat{\lambda}|S) - \hat{\vartheta}_{\text{Bayes}}\|_{Q,2} \leq \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \pi(d\vartheta)}.$$

More generally, the Bayes risk of the Tikhonov regularized Rao–Blackwellization $\mathbf{E}^{(\alpha)}(\hat{\lambda}|S)$ is shown to be continuous in the bias of the estimator $\hat{\lambda}$ and, furthermore,

$$\mathbf{r}(\mathbf{E}^{(\alpha)}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\mathbf{E}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\hat{\lambda}, \pi),$$

if $\hat{\lambda}$ is unbiased for the canonical parameter. This result is a Tikhonov regularized version of the Rao–Blackwell and Lehmann–Scheffé theorem (see Theorem 4.12, page 86).

Ivanov’s regularization approach establishes the regularity of a pseudosolution of (2.15) by prespecifying a set of admissible solutions. Let $\mathcal{C} \subset \mathfrak{A}$ be sufficient for \mathfrak{B} with a certain completeness condition. We especially study the case of a finite-dimensional space $H \subset L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{B})$ and introduce an Ivanov regularized Rao–Blackwellization, that is,

$$\mathbf{E}^{(H)}(\hat{\lambda}|\mathcal{C}) = \text{argmin}\{\|\mathbf{E}(\hat{\lambda} - g)\|_{\pi,2} : g \in H\}$$

(see Theorem 4.14 on page 88). Concerning the parameter $(\mathbf{E}_{\vartheta} \hat{\lambda})_{\vartheta \in \Theta}$, we are looking for a \mathcal{C} -measurable estimator that is assumed to have a “reasonable” discrete description in terms of basis elements of H .

With regard to the computation of such regularized Rao–Blackwellizations, we aim at finite-dimensional linear subspaces H that are spanned

by “regular” eigenfunctions to eigenvalues close to 1 of a tractable Markov operator R .² For $Q := \int P_{\vartheta}(\cdot) \boldsymbol{\pi}(d\vartheta)$, $p \in]1, \infty[$ and $\|(I - R)|_H\| \leq \varepsilon$, Theorem 4.16 yields for all $f \in L_p(\mathfrak{X}, \mathfrak{A}, Q)$ that

$$\begin{aligned} \|\mathbf{E}^{(H)}(f|\mathcal{C}) - \int f(y)R(\cdot, dy)\|_{Q,p} \\ \leq \|R\| \inf\{\|\mathbf{E}(f - g)\|_{\pi,p} : g \in H\} + \varepsilon \cdot \|\mathbf{E}^{(H)}(f|\mathcal{C})\|_{Q,p}, \end{aligned}$$

where $\|R\| := \sup\{\|Rf\|_{Q,p} : \|\mathbf{E}f\|_{\pi,p} \leq 1\}$. As an example for such a Markov operator R , we consider, for example,

$$R : \hat{\lambda} \mapsto \int \hat{\lambda}(x) P_{\hat{\vartheta}_{\text{ML}}(\cdot)}(dx), \quad (2.16)$$

where $\hat{\vartheta}_{\text{ML}}$ is the maximum likelihood estimator for the canonical parameter \mathbf{id}_{Θ} , see Example 4.17 and 4.18.

Furthermore, we can use (2.16) to compute approximates of the maximum likelihood estimator. To this end, let $\mathfrak{P} = \{P_{\vartheta} : \vartheta \in \Theta\}$ be an exponential model of full rank and with sufficient statistic S . We may then consider the factorization of the MLE $\hat{\vartheta}_{\text{ML}}$ given by

$$\hat{\vartheta}_{\text{ML}} =: \hat{\theta}_{\text{ML}} \circ S.$$

In Theorem 4.20 we show that

$$\begin{aligned} \hat{\theta}_{\text{ML}}(s) = \mathbf{E}_{\hat{\vartheta}_{\text{ML}}(x)}(\hat{\lambda}) + \text{Var}_{\hat{\vartheta}_{\text{ML}}(x)}^{-1}(S) \text{Cov}_{\hat{\vartheta}_{\text{ML}}(x)}(S, \hat{\lambda})(s - S(x)) \\ + \text{Bias}(\hat{\lambda}, \hat{\theta}_{\text{ML}}(s)) + o(\|s - S(x)\|) \end{aligned}$$

for values s in a neighborhood of the observed sufficient statistic value $S(x)$. Hence, an approximation of the maximum likelihood estimator in a neighborhood of $S(x)$ can essentially be computed by a linear regression of $\hat{\lambda}$ estimates at sufficient statistic values of realizations simulated from $\hat{\vartheta}_{\text{ML}}(x)$. The computation of an approximate MLE distribution can be realized as follows (see also Algorithm 4.21):

Computing an Approximate Distribution of the MLE. Let $(P_{\vartheta})_{\vartheta \in \Theta}$ be an exponential model with sufficient statistic S . By $\hat{\vartheta}_{\text{ML}}$ we denote the maximum likelihood estimator for the canonical parameter \mathbf{id}_{Θ} and by $\hat{\lambda}$ we denote a simple alternative estimator. For our observed data $x \in \mathfrak{X}$, we proceed as follows:

- (1) Compute $\hat{\vartheta}_{\text{ML}}(x)$;
- (2) Simulate $Y_1, \dots, Y_N \sim P_{\hat{\vartheta}_{\text{ML}}(x)}$ for a large number N ;
- (3) Compute $\hat{\lambda}(Y_1), \dots, \hat{\lambda}(Y_N)$ and sufficient statistic values $S(Y_1), \dots, S(Y_N)$;
- (4) Perform a linear regression of $(\hat{\lambda}(Y_i) : i = 1, \dots, N)$ on the sufficient statistic values $(S(Y_i) : i = 1, \dots, N)$.

²Here, R substitutes the conditional distribution $P(\cdot|\mathcal{C})$. Recall, that $P(\cdot|\mathcal{C})$ is a projection and, hence, the corresponding spectrum consists of 1 and 0 only.

Then, the fitted values constitute an approximation to the distribution of $\hat{\vartheta}_{\text{ML}}$ with underlying true distribution P_{ϑ} for some $\vartheta \in \Theta$ (see Theorem 4.20). ♦

In Part III we demonstrate the benefits of the above results and the performance of the algorithm through a simulation study for the situation of Gibbs point process statistics (see Chapter 7, page 157).

CHAPTER 3

The Rao–Blackwell Inverse Problem

Let $(\mathfrak{X}, \mathcal{A})$ be a measurable space with Borel- σ -algebra \mathcal{A} according to an underlying Hausdorff-topology \mathcal{T} . Furthermore, let $\mathfrak{P} \subset \text{Prob}(\mathfrak{X}, \mathcal{A})$ be a statistical model. The problem is to estimate an \mathbb{R}^k -valued parameter $\kappa : \mathfrak{P} \rightarrow \mathbb{R}^k$.

To begin with, we briefly recall some notations from Chapter 1 (see also page 8 f.). For some $p \in [1, \infty]$ we consider estimators $\hat{\kappa} : \mathfrak{X} \rightarrow \mathbb{R}^k$ that are p th power integrable w. r. t. every $P \in \mathfrak{P}$, that is,

$$\mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k) := \bigcap_{P \in \mathfrak{P}} \mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, P; \mathbb{R}^k).$$

We identify estimators according to the \mathfrak{P} -almost sure equivalence relation. That means, let $[\mathfrak{P}]$ be the set of all functions in $\mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ that equal the null function \mathfrak{P} -almost surely, we then consider

$$L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k) := \mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k) / [\mathfrak{P}],$$

containing all equivalence classes (w. r. t. $[\mathfrak{P}]$) of functions with values in \mathbb{R}^k . We will just write $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ to emphasize the considered σ -algebra and use $L_p(\mathfrak{P})$ for short if no clarification is needed.

Assume that \mathfrak{P} admits a complete and sufficient σ -algebra $\mathcal{C} \subset \mathcal{A}$. A crucial item for the Rao–Blackwell inverse problem is the *expectation operator*

$$\mathbf{E} : \begin{array}{ccc} L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) & \rightarrow & (\mathbb{R}^k)^{\mathfrak{P}} \\ f & \mapsto & (\int f \, dP)_{P \in \mathfrak{P}} \end{array}$$

which determines \mathcal{C} -measurable and \mathfrak{P} -integrable functions uniquely due to the completeness of \mathcal{C} for \mathfrak{P} . Denoting by $\text{ran}(\mathbf{E})$ the range of \mathbf{E} and noting that

$$\text{ran}(\mathbf{E}) = \mathbf{E}(L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})) \subset (\mathbb{R}^k)^{\mathfrak{P}},$$

we finally get the commutative diagram (see also Figure 2.1, page 41, and recall that $\mathbf{E}_{\mathcal{C}} := \mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}$)

$$\begin{array}{ccc}
& & \text{ran}(\mathbf{E}) \\
& \nearrow \mathbf{E} & \downarrow \mathbf{E}|_{\mathcal{C}}^{-1} \\
L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) & \xrightarrow{\mathbf{E}(\cdot|\mathcal{C})} & L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P}).
\end{array}$$

The *classical Rao–Blackwellization* corresponds, therefore, to the following what we call *Rao–Blackwell inverse problem*.¹ For any $f \in L_1(\mathfrak{P})$ we are looking for $g \in L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ such that

$$\mathbf{E}_P f = \mathbf{E}_P g \quad \text{for all } P \in \mathfrak{P}, \quad (3.1)$$

which has in $L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ the unique solution $\mathbf{E}(f|\mathcal{C}) = \bigcap_{P \in \mathfrak{P}} \mathbf{E}_P(f|\mathcal{C})$, since \mathcal{C} is sufficient and complete.² We call (3.1) *Radon–Nikodým equations* in accordance with the following example.

3.1 Remark (Conditional Expectation) Let P be an arbitrary probability measure on $(\mathfrak{X}, \mathcal{A})$ and let $\mathcal{C} \subset \mathcal{A}$ be a sub- σ -algebra. Define

$$\mathfrak{P} := \{P(\cdot|C) : C \in \mathcal{C}, P(C) > 0\},$$

then \mathcal{C} is complete and sufficient for \mathfrak{P} . A more general statement and a proof is given in Kagan *et al.* (2014, Lemma 4.1, see also Example 4.5).

The conditional expectation of $X \in L_1(P)$ is any \mathcal{C} -measurable random variable $\mathbf{E}(X|\mathcal{C})$ which fulfills

$$\int_C X \, dP = \int_C \mathbf{E}(X|\mathcal{C}) \, dP \quad \forall C \in \mathcal{C}. \quad (3.2)$$

Condition (3.2) corresponds to (3.1) with \mathfrak{P} considered in this example. Furthermore, a standard lifting argument obviously yields that (3.2) is equivalent to

$$\int X \varphi \, dP = \int \mathbf{E}(X|\mathcal{C}) \varphi \, dP \quad (3.3)$$

for all \mathcal{C} -measurable, non-negative functions φ . Here, (3.3) corresponds to (3.1) with the statistical model $\mathfrak{P}' := \{Q : Q \ll P|_{\mathcal{C}}\}$. Note also that \mathfrak{P}' can be reduced to a class of distributions that separates \mathcal{C} -measurable functions. \blacklozenge

In most cases, it is hard or even impossible to solve the Radon–Nikodým equations exactly, that is, to compute $\mathbf{E}(f|\mathcal{C})$ for given $(\mathbf{E}_P f)_{P \in \mathfrak{P}}$. In fact, there is usually no need to solve these equations exacty due to model

¹The closely related observation that the unbiased estimation theory belongs to the theory of (ill-posed) integral equations of the first kind is, to the best knowledge of the author, only given in Schmetterer (1977, page 313) and Schmetterer (1978, Section 3, page 493 f.).

²See Proposition 1.26.

misspecifications that often occur in practice. The parameter function $(\mathbf{E}_P f)_{P \in \mathfrak{P}}$ can then only be interpreted up to slight (unknown) corrections.

The best we can do is to compute a $\mathbf{E}^{(*)}(f|\mathcal{C})$ which fulfills (3.1) in some sense approximately. Our concern is therefore to analyze the stability of the Rao–Blackwell inverse problem, that is, if and in what sense the solution $\mathbf{E}^{(*)}(f|\mathcal{C})$ depends continuously on the information $(\mathbf{E}_P f)_{P \in \mathfrak{P}}$.

This question of continuity of operators from $\text{ran}(\mathbf{E})$ to $L_p(\mathfrak{P})$ depends evidently on the principal questions of how to compare the mean functionals in $\text{ran}(\mathbf{E})$ and how to compare the Rao–Blackwellizations in $L_p(\mathfrak{P})$. However, since we are interested in the Rao–Blackwell operator $\mathbf{E}(\cdot|\mathcal{C})$, the choice of the topological structure on $\text{ran}(\mathbf{E})$ should also account for the continuity of the expectation operator \mathbf{E} from $L_p(\mathfrak{P})$ to $\text{ran}(\mathbf{E})$ (see also diagram on page 50).

In the next section, we introduce what we call *Lehmann–Scheffé topologies* on $L_p(\mathfrak{P})$ by borrowing canonical topological structures from $\text{ran}(\mathbf{E})$ via the expectation operator \mathbf{E} . Subsequently, we introduce associated normed spaces which we call *Lehmann–Scheffé spaces*. Based on these topological structures, we give various results and examples regarding the described continuity problem.

3.1 Lehmann–Scheffé Topologies

Let $p \in [1, \infty]$. A statistical model \mathfrak{P} induces a topological structure on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ based on the pointwise behavior of the expectation operator \mathbf{E} by way of the semi-norms

$$f \mapsto q_P(f) := \left\| \int f \, dP \right\| \quad \text{for } P \in \mathfrak{P}. \quad (3.4)$$

Consequently, we consider the locally convex topology³

$$\mathcal{T}_{\text{LS}} := \mathcal{T}_{\text{LS}}(\mathfrak{P}) := \mathcal{T}_{\text{LS}}(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathfrak{P}) := \tau(q_P : P \in \mathfrak{P}) \quad (3.5)$$

on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, that is, the smallest topology such that all semi-norms q_P for $P \in \mathfrak{P}$ are continuous. We call (3.5) the *Lehmann–Scheffé (LS) topology*, referring to Lehmann & Scheffé (1947, 1950, 1955, 1956) who introduced the notion of completeness⁴ of a σ -algebra for a statistical model and demonstrated the utility of this general concept for statistical theory.

In fact, if $\mathcal{C} \subset \mathcal{A}$ is a complete σ -algebra for \mathfrak{P} the Lehmann–Scheffé topology $\mathcal{T}_{\text{LS}}(L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}), \mathfrak{P})$ on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ is a Hausdorff topology which follows easily from e. g. Werner (2018, Lemma VIII.1.4, page 429). In the

³Let $(f_\alpha)_{\alpha \in I}$ be a family of functions from \mathfrak{X} to a topological space $(\mathfrak{Y}, \mathcal{C})$. We denote by $\tau(f_\alpha : \alpha \in I)$ the coarsest topology on \mathfrak{X} such that all f_α are continuous.

⁴See Definition 1.23 on page 16.

notation of the LS topology we will often suppress the space $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ and just write $\mathcal{T}_{\text{LS}}(\mathfrak{P})$, since one can easily show that

$$\mathcal{T}_{\text{LS}}(L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}), \mathfrak{P}) = \mathcal{T}_{\text{LS}}(L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}), \mathfrak{P}) \cap L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}). \quad (3.6)$$

So far we noted that a statistical model \mathfrak{P} induces a topological structure on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}^k)$. Conversely, $L_p(\mathfrak{P})$ also generates “canonical” topological structure on \mathfrak{P} given by

$$\mathcal{T}_{\text{LS}}^*(L_p(\mathfrak{P})) := \mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_p(\mathfrak{P})) := \tau(P \mapsto \int f \, dP : f \in L_p(\mathfrak{P}))$$

which we call *Lehmann–Scheffé-*topology*. We discuss this notion and the relation to the weak topologies in the following remark.

3.2 Remark ($L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}^k)$ - \mathfrak{P} -Duality) **(a) Bilinear Map.** The statistical model \mathfrak{P} technically in terms of the generated linear space⁵

$$M := \text{span } \mathfrak{P} \subset \mathfrak{S}(\mathfrak{X}, \mathcal{A})$$

is linked to the estimator space $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}^k)$ by the bilinear map

$$\begin{aligned} [\cdot, \cdot] : L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P}; \mathbb{R}^k) \times M &\rightarrow \mathbb{R}^k \\ (f, \mu) &\mapsto \int f \, d\mu. \end{aligned} \quad (3.7)$$

Note that $[f, \mu] = \mathbf{0}$ for all $f \in L_p(\mathfrak{P})$ implies $\mu = \mathbf{0}$. Conversely, the completeness of \mathcal{C} for \mathfrak{P} yields that $[f, \mu] = \mathbf{0}$ for all $\mu \in M$ implies $f = \mathbf{0}$.

(b) Linear Functionals and Dual Subspaces. For $\alpha \in \mathbb{R}^k$ we define the bilinear form

$$\begin{aligned} [\cdot, \cdot]_\alpha : L_p(\mathfrak{P}) \times M &\rightarrow \mathbb{R} \\ (f, \mu) &\mapsto \int \langle \alpha, f \rangle \, d\mu. \end{aligned}$$

Furthermore, $f \mapsto [f, \mu]_\alpha$ is a linear functional on $L_p(\mathfrak{P})$ for every $\mu \in M$ and every $\alpha \in \mathbb{R}^k$. The vector space $M \times \mathbb{R}^k$ can therefore be considered as a subspace of⁶ $L_p(\mathfrak{P})^*$. Analogously, we may conclude that $L_p(\mathfrak{P}) \times \mathbb{R}^k$ represents a subspace of M^* .

(c) Linear Functionals, LS- and Weak Topologies. For the Lehmann–Scheffé topology we have

$$\mathcal{T}_{\text{LS}}(L_p(\mathfrak{P}), \mathfrak{P}) = \tau([\cdot, \mu]_\alpha : \mu \in M, \alpha \in \mathbb{R}^k).$$

To show this, we first note that for $\mu \in M$ the function $[\cdot, \mu]$ is continuous w. r. t. $\mathcal{T}_{\text{LS}}(\mathfrak{P})$. Hence, $[\cdot, \mu]_\alpha$ is continuous w. r. t. $\mathcal{T}_{\text{LS}}(\mathfrak{P})$ for all $\alpha \in \mathbb{R}^k$. That

⁵We denote by $\mathfrak{S}(\mathfrak{X}, \mathcal{A})$ the set of finite signed measures on $(\mathfrak{X}, \mathcal{A})$.

⁶Let E be a vector space. The linear space consisting of all linear functionals on E is called *algebraic dual space* of E and denoted by E^* .

means, the right-hand side is coarser than the left-hand side. To see the converse inclusion, we note that $[\cdot, P]$ is continuous if and only if $[\cdot, P]_\alpha$ is continuous for all $\alpha \in \mathbb{R}^k$ and that $\mathfrak{P} \subset M$.

For the Lehmann–Scheffé- $*$ -topology it can be shown that

$$\mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_p(\mathfrak{P})) = \tau([f, \cdot]_\alpha : f \in L_p(\mathfrak{P}), \alpha \in \mathbb{R}^k) \cap \mathfrak{P}.$$

Hence, the LS- and the LS- $*$ -topology are the coarsest topologies such that the linear functionals essentially determined by \mathfrak{P} and $L_p(\mathfrak{P})$, respectively, are continuous.

In fact, for the case $k = 1$ the Lehmann–Scheffé topology coincides with the usual *weak topology* according to the *dual pair* $(L_p(\mathfrak{P}), M, [\cdot, \cdot])$ (see Schaefer (1971, page 123)), that is, the coarsest topology on $L_p(\mathfrak{P})$, such that all linear functionals $f \mapsto [f, \mu]$ for $\mu \in M$ are continuous, which is formally given by

$$\sigma_w(L_p(\mathfrak{P}), M, [\cdot, \cdot]) := \tau([\cdot, \mu] : \mu \in M).$$

Conversely, the *weak- $*$ -topology* w.r.t. the dual pair $(L_p(\mathfrak{P}), M, [\cdot, \cdot])$ is defined by

$$\sigma_w(M, L_p(\mathfrak{P}), [\cdot, \cdot]) := \tau([f, \cdot] : f \in L_p(\mathfrak{P})).$$

This is the weakest topology on M such that all linear functionals $\mu \mapsto [f, \mu]$ for $f \in L_p(\mathfrak{P})$ are continuous. Hence, we have a topological structure on M and consequently on \mathfrak{P} . \blacklozenge

Rao–Blackwellizations that we (approximately) compute with regard to the Radon–Nikodým equations (3.1) should be compared, for example, w.r.t. the *canonical topology* $\mathcal{T}_{L_p(\mathfrak{P})}$ for $p \in [1, \infty[$. This topological structure is stronger than that of the Lehmann–Scheffé topology $\mathcal{T}_{\text{LS}}(\mathfrak{P})$. We recall from Chapter 1 that the topology $\mathcal{T}_{L_p(\mathfrak{P})}$ on $L_p(\mathfrak{P})$ is generated by the \mathfrak{P} -indexed family of semi-norms

$$f \mapsto \|f\|_{P,p} := \sqrt[p]{\int \|f(x)\|^p P(dx)} \quad \text{for } P \in \mathfrak{P}.$$

With this structure we simultaneously compare estimators in $L_p(P)$ for every $P \in \mathfrak{P}$. Furthermore, for $p = \infty$ we note that the vector space $L_\infty(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ contains the \mathfrak{P} -equivalence classes that are essentially bounded. That is, a measurable f is an element of $L_\infty(\mathfrak{P})$ iff there is a number $\alpha \in \mathbb{R}$ such that

$$\{x \in \mathfrak{X} : \|f(x)\|_{\mathbb{R}^k} > \alpha\}$$

is a \mathfrak{P} -null set. By setting

$$\|f\|_{L_\infty(\mathfrak{P}), \infty} := \inf \left\{ \sup \{ \|f(x)\|_{\mathbb{R}^k} : x \in N^c \} : N \in \mathcal{A}, P(N) = 0 \forall P \in \mathfrak{P} \right\}$$

for $f \in L_\infty(\mathfrak{F})$, we obtain a norm on $L_\infty(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$. With regard to the Radon–Nikodým equations, we are concerned with

$$\|f\|_{\mathbf{LS},\infty} := \sup \left\{ \left\| \int f \, dP \right\|_{\mathbb{R}^k} : P \in \mathfrak{F} \right\} \quad \text{for } f \in L_\infty(\mathfrak{F})$$

on the one hand. Note that $\|\cdot\|_{\mathbf{LS},\infty} \leq \|\cdot\|_{L_\infty(\mathfrak{F}),\infty}$ and $\|\cdot\|_{\mathbf{LS},\infty}$ is a seminorm on $L_\infty(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$ and even a norm on $L_\infty(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$. On the other hand, however, the comparison of Rao–Blackwellizations should at least be considered w. r. t.

$$\|f\|_{\mathbf{LSI},\infty} := \sup \left\{ \int \|f(x)\|_{\mathbb{R}^k} P(dx) : P \in \mathfrak{F} \right\} \leq \|f\|_{L_\infty(\mathfrak{F}),\infty}.$$

In general we have $\|\cdot\|_{\mathbf{LSI},\infty} \neq \|\cdot\|_{L_\infty(\mathfrak{F}),\infty}$. Note, however, that $\|\cdot\|_{\mathbf{LSI},\infty}$ constitutes also a norm on $L_\infty(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$.

Concerning misspecified statistical models and with regard to the (approximate) computation of Rao–Blackwellizations, our intention is to analyze the well- or ill-posedness of the Rao–Blackwell inverse problem. According to the above preparations this corresponds to analyzing the continuity of the Rao–Blackwell projection considered as⁷

$$\begin{aligned} \mathbf{E}(\cdot|\mathcal{C}) : (L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{F}), \mathcal{T}_{\mathbf{LS}}(\mathfrak{F})) &\rightarrow (L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}), \mathcal{T}_{L_p(\mathfrak{F})}) \\ f &\mapsto \mathbf{E}(f|\mathcal{C}), \end{aligned} \quad (3.8)$$

if $1 \leq p < \infty$, and as

$$\begin{aligned} \mathbf{E}(\cdot|\mathcal{C}) : (L_\infty(\mathfrak{X}, \mathcal{A}, \mathfrak{F}), \|\cdot\|_{\mathbf{LS},\infty}) &\rightarrow (L_\infty(\mathfrak{X}, \mathcal{C}, \mathfrak{F}), \|\cdot\|_{\mathbf{LSI},\infty}) \\ f &\mapsto \mathbf{E}(f|\mathcal{C}), \end{aligned}$$

if $p = \infty$. In the remaining of this thesis we will often skip the latter case. A characterization of the continuity of (3.8) is given in the following lemma.

3.3 Lemma The Rao–Blackwell projection in (3.8) is continuous if and only if $\mathcal{T}_{\mathbf{LS}}(\mathfrak{F}) = \mathcal{T}_{L_p(\mathfrak{F})}$ holds on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$. \blacklozenge

Proof: We decompose $\mathbf{E}(\cdot|\mathcal{C})$ in (3.8) in a projection and a topological part, that is, in

$$\begin{aligned} \text{pr}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F})} : (L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{F}), \mathcal{T}_{\mathbf{LS}}(\mathfrak{F})) &\rightarrow (L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}), \mathcal{T}_{\mathbf{LS}}(\mathfrak{F})) \\ f &\mapsto \mathbf{E}(f|\mathcal{C}) \end{aligned}$$

and the identity

$$\mathbf{id}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F})} : (L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}), \mathcal{T}_{\mathbf{LS}}(\mathfrak{F})) \rightarrow (L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F}), \mathcal{T}_{L_p(\mathfrak{F})}). \quad (3.9)$$

⁷See also Proposition 1.26 on page 17 where we have shown that the Rao–Blackwell projection is continuous w. r. t. the canonical topology.

Note that $\mathbf{E}(\cdot|\mathcal{C}) = \mathbf{id}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})} \circ \text{pr}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}$, where $\text{pr}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}$ is continuous, since: for all $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$

$$\| \int \mathbf{E}(f|\mathcal{C}) dP \|_{\mathbb{R}^k} = \| \int f dP \|_{\mathbb{R}^k}$$

for all $P \in \mathfrak{P}$. Hence, if \mathcal{C} is a generator of $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ we have that $\text{pr}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}^{-1}(\mathcal{C})$ is the generator of $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. Note also that $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) \subset \mathcal{T}_{L_p(\mathfrak{P})}$, since for a constant $C > 0$, all $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and all $P \in \mathfrak{P}$

$$\| \int f dP \|_{\mathbb{R}^k} \leq \int \| f \|_{\mathbb{R}^k} dP \leq C \| f \|_{P,p}.$$

Therefore, $\mathbf{E}(\cdot|\mathcal{C})$ is continuous if $\mathbf{id}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}$ in (3.9) is continuous, that is, $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) = \mathcal{T}_{L_p(\mathfrak{P})}$ on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$.

Conversely, the continuity of $\mathbf{E}(\cdot|\mathcal{C})$ yields the continuity of

$$\mathbf{E}(\cdot|\mathcal{C})|_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})} = \mathbf{id}_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})},$$

since $\mathbf{E}(\cdot|\mathcal{C})^{-1}(\mathcal{T}_{L_p(\mathfrak{P})}) \subset \mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ yields

$$\mathbf{E}(\cdot|\mathcal{C})|_{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}^{-1}(\mathcal{T}_{L_p(\mathfrak{P})}) \subset \mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) \cap L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$$

and the right-hand side corresponds to $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$ (see also (3.6)). Hence, $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) = \mathcal{T}_{L_p(\mathfrak{P})}$ on $L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})$. \blacksquare

The following examples show that in some cases $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ and $\mathcal{T}_{L_p(\mathfrak{P})}$ are already well known to us if $(\mathfrak{X}, \mathcal{A})$ and \mathfrak{P} are suitable chosen. They allow us to obtain an impression of the nature of the topologies we are concerned with, depending on the statistical model \mathfrak{P} and the underlying sample space $(\mathfrak{X}, \mathcal{A})$.

3.4 Example (Topology of Pointwise Convergence) Let $(\mathfrak{X}, \mathcal{A})$ be an arbitrary measurable space and consider the family of Dirac measures $\mathfrak{P} = (\delta_x : x \in \mathfrak{X})$. Obviously, \mathfrak{P} is complete. Furthermore, we note for the spaces of estimators

$$L_1(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k) = L(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k),$$

where the latter denotes the vector space of $[\mathfrak{P}]$ -equivalence classes of \mathcal{A} -measurable \mathbb{R}^k -valued functions. The corresponding LS topology $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ as well as $\mathcal{T}_{L_1(\mathfrak{P})}$ are generated by the same family of semi-norms $f \mapsto \|f(x)\|$ for $x \in \mathfrak{X}$. That is, $\mathcal{T}_{\mathbf{LS}}$ and $\mathcal{T}_{L_1(\mathfrak{P})}$ coincide with the topology of pointwise convergence and we have

$$(L(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k), \mathcal{T}_{\mathbf{LS}}(\mathfrak{P})) = (L(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k), \mathcal{T}_{L_1(\mathfrak{P})}).$$

Furthermore, $\mathcal{T}_{\mathbf{LS}}$ cannot be generated by a metric, if \mathfrak{X} is uncountable (see e. g. Werner (2018)). \blacklozenge

3.5 Example (Binomial Model) For $n \in \mathbb{N}$ consider the statistical model $\mathfrak{P} = (\mathbf{B}(n, \pi) : \pi \in]0, 1[)$, where $\mathfrak{X} := \{0, 1, \dots, n\}$ is our sample space. The statistic $\mathbf{id}_{\{0, \dots, n\}}$ is complete and sufficient for \mathfrak{P} . Note that $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}) = \mathbb{R}^{\{0, \dots, n\}}$ holds for any $p \in [1, \infty]$. Both, $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ and $\mathcal{T}_{L_1(\mathfrak{P})}$ make $\mathbb{R}^{\{0, \dots, n\}}$ a finite-dimensional Hausdorff topological vector space, whose topology is uniquely determined. Hence, $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) = \mathcal{T}_{L_1(\mathfrak{P})}$ and furthermore $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$, $\mathcal{T}_{L_1(\mathfrak{P})}$ coincide with the Euclidean topology, that is, they are in particular generated by a norm on \mathbb{R}^n . \blacklozenge

In general, however, $\mathcal{T}_{\mathbf{LS}}$ is usually strictly weaker than $\mathcal{T}_{L_1(\mathfrak{P})}$ as we will see in the following example. According to Lemma 3.3 this means that in general the Rao–Blackwell projection $\mathbf{E}(\cdot | \mathcal{C})$ given in (3.8) is not continuous and hence the Rao–Blackwell inverse problem is usually ill-posed.

3.6 Example (Weak vs Norm Topology, $k = 1$) Let Q be a probability measure on $(\mathfrak{X}, \mathcal{A})$. For some conjugate exponents $1 \leq p \leq q \leq \infty$, that is, $\frac{1}{p} + \frac{1}{q} = 1$, we consider the Q -dominated, non-parametric statistical model

$$\mathfrak{P} := \left\{ P : P \in \text{Prob}(\mathfrak{X}, \mathcal{A}), P \ll Q, \frac{dP}{dQ} \in L_q(Q) \right\}. \quad (3.10)$$

Then, by definition of \mathfrak{P} we have $L_p(Q) \subset L_1(\mathfrak{P})$, since $[\mathfrak{P}] = [Q]$ and for every $f \in L_p(Q)$

$$\int |f| dP = \int |f| \frac{dP}{dQ} dQ < \infty,$$

hence, $f \in L_1(\mathfrak{P})$. Furthermore, the LS topology $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ on $L_p(Q)$ equals the weak topology of $L_p(Q)$

$$\sigma_w(L_p(Q), L_q(Q)) = \tau(f \mapsto |\int f g dQ| : g \in L_q(Q)),$$

see for example Werner (2018).

Obviously, we have $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P}) \subset \sigma_w(L_p(Q), L_q(Q))$ by the definition of \mathfrak{P} and $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$. To show the converse let $g \in L_q(Q)$ with⁸ $g^+, g^- \neq \mathbf{0}$, then

$$\left| \int f g dQ \right| \leq \|g^+\|_{Q,1} \left| \int f dP_{g^+} \right| + \|g^-\|_{Q,1} \left| \int f dP_{g^-} \right|,$$

where $P_{g^+}, P_{g^-} \in \mathfrak{P}$ are distributions with densities proportional to g^+ and g^- (w. r. t. Q), respectively. Hence, $f \mapsto |\int f g dQ|$ is $\mathcal{T}_{\mathbf{LS}}(\mathfrak{P})$ -continuous.

If $p = 1$ the absolute LS topology $\mathcal{T}_{L_1(\mathfrak{P})}$ is exactly the $\|\cdot\|_Q$ -norm topology on $L_p(Q)$, which is seen as follows: First, we have $Q \in \mathfrak{P}$, thus $\tau(\|\cdot\|_Q) \subset \mathcal{T}_{L_1(\mathfrak{P})}$. Second, any $P \in \mathfrak{P}$ has a density w. r. t. Q which is an element in $L_\infty(Q)$, hence $\|f\|_P \leq C_P \cdot \|f\|_Q$ for some $C_P > 0$. Thus, $\|\cdot\|_P$ is continuous for every $P \in \mathfrak{P}$, that is, we have $\mathcal{T}_{L_1(\mathfrak{P})} \subset \tau(\|\cdot\|_Q)$. \blacklozenge

⁸For a function $f : \mathfrak{X} \rightarrow \mathbb{R}$ the *positive* and *negative part* are defined by $f^+(x) := \max\{f(x), x\}$ and $f^-(x) := \max\{-f(x), x\}$ for $x \in \mathfrak{X}$, respectively.

As the example above indicates, $\mathcal{T}_{\mathbf{LS}}(\mathfrak{F})$ is in general strictly weaker than $\mathcal{T}_{L_p(\mathfrak{F})}$. The following well known example illustrates what can go wrong, that is, that $\mathcal{T}_{\mathbf{LS}}(\mathfrak{F})$ may fail to detect “high frequency components” of a function.

3.7 Example (Rademacher Sequence) Consider the situation of Example 3.6 for $p = 1$ and $q = \infty$ with $(\mathfrak{X}, \mathcal{A}, \mathcal{Q}) := ([0, 1], \mathcal{B}([0, 1]), \lambda)$, where λ denotes the Lebesgue measure. According to Example 3.6 we have $\mathcal{T}_{\mathbf{LS}}(\mathfrak{F}) = \sigma_w(L_1(\mathcal{Q}), L_\infty(\mathcal{Q}))$ and $\mathcal{T}_{L_1(\mathfrak{F})} = \tau(\|\cdot\|_{\mathcal{Q}})$. Furthermore, it is well known that $\sigma_w(L_p(\mathcal{Q}), L_q(\mathcal{Q}))$ is strictly weaker than $\tau(\|\cdot\|_{\mathcal{Q}})$. This can be illustrated by the *Rademacher sequence*

$$f_n(x) := \text{sign}(\sin(2^n \pi x)) \quad \text{for } x \in [0, 1], n \in \mathbb{N},$$

which is well known to converge weakly to the zero function, that is, we have convergence w. r. t. $\mathcal{T}_{\mathbf{LS}}(\mathfrak{F})$. However, it obviously does not converge w. r. t. $\mathcal{T}_{L_1(\mathfrak{F})} = \tau(\|\cdot\|_{\mathcal{Q}})$, since $\|f_n\|_P = 1$ for every $n \in \mathbb{N}$ and all $P \in \mathfrak{F}$. ♦

From the above example, we can easily deduce consequences for statistical more relevant situations.

3.8 Example Consider the situation of Example 3.6 for $p = 1$ and $q = \infty$. For any countable complete submodel \mathfrak{Q} of \mathfrak{F} , e. g., $\mathfrak{Q} = (\mathbf{N}(\mu, \sigma^2) : \mu \in \mathcal{Q})$, we have

$$\mathcal{T}_{\mathbf{LS}}(\mathfrak{Q}) \subsetneq \sigma_w(L_1(\lambda), L_\infty(\lambda)).$$

Note that the generator of $\mathcal{T}_{\mathbf{LS}}(\mathfrak{Q})$ is contained in the weak topology on the right-hand side. Furthermore, since $\mathcal{T}_{\mathbf{LS}}(\mathfrak{Q})$ is metrizable, it has to be different and hence strictly weaker than the non-metrizable weak topology $\sigma_w(L_1(\lambda), L_\infty(\lambda))$. Furthermore, we also have

$$\mathcal{T}_{L_1(\mathfrak{Q})} \subsetneq \tau(\|\cdot\|_\lambda)$$

since: $\|\cdot\|_P \leq C\|\cdot\|_\lambda$ for some $C > 0$ and for all $P \in \mathfrak{Q}$, that is, $\mathcal{T}_{L_1(\mathfrak{Q})} \subset \tau(\|\cdot\|_\lambda)$. Note also that $\|\cdot\|_\lambda$ cannot be dominated by finitely many semi-norms $\|\cdot\|_Q$ for $Q \in \mathfrak{Q}$. That means $\|\cdot\|_\lambda$ is not $\mathcal{T}_{L_1(\mathfrak{Q})}$ -continuous and thus $\mathcal{T}_{L_1(\mathfrak{Q})}$ is strictly weaker than $\tau(\|\cdot\|_\lambda)$.

Extend the Rademacher sequence $(f_n)_{n \in \mathbb{N}}$ of Example 3.7 periodically on \mathbb{R} . With the same arguments as in Example 3.7 one can show that this sequence is $\mathcal{T}_{\mathbf{LS}}(\mathfrak{Q})$ -convergent, but does not converge w. r. t. $\mathcal{T}_{L_1(\mathfrak{Q})}$. ♦

More generally, let \mathfrak{F} be a statistical model that is dominated by a σ -finite measure μ . Then, the relationship between $(L_1(\mathfrak{F}), \mathcal{T}_{\mathbf{LS}}(\mathfrak{F}))$ and $(L_p(\mu), \sigma_w(L_p(\mu), L_q(\mu)))$ is as follows (we consider the case $k = 1$).

3.9 Theorem Let $p, q \in [1, \infty]$ be conjugate exponents and let \mathfrak{F} be dominated by a σ -finite measure μ .

Then $L_p(\mu) \subset L_1(\mathfrak{P})$ and $\mathcal{T}_{\text{LS}}(\mathfrak{P}) \subset \sigma_w(L_p(\mu), L_q(\mu))$ on $L_p(\mu)$ if and only if \mathfrak{P} has a μ -density choice⁹ in $L_q(\mu)$. \blacklozenge

Proof: The case $p = \infty$, $q = 1$ is trivial by Riesz's theorem. For $p < \infty$ and $q > 1$ we first show the necessity. Since \mathfrak{P} has by assumption a μ -density choice $(\varphi_P)_{P \in \mathfrak{P}} \in L_q(\mu)^{\mathfrak{P}}$ we have $\int |f| dP = \int |f| \varphi_P d\mu < \infty$ for all $f \in L_p(\mu)$. Hence, $f \in L_1(\mathfrak{P})$. Furthermore, $\sigma_w(L_p(\mu), L_q(\mu))$ is by definition the weakest topology on $L_p(\mu)$ such that all linear functionals are continuous, that is,

$$f \mapsto \int f g d\mu$$

is continuous for every $g \in L_q(\mu)$. By assumption $P \in \mathfrak{P}$ has a μ -density φ_P in $L_q(\mu)$ and thus

$$f \mapsto \int f \varphi_P d\mu$$

is continuous w. r. t. $\sigma_w(L_p(\mu), L_q(\mu))$. Hence, $\mathcal{T}_{\text{LS}}(\mathfrak{P}) \subset \sigma_w(L_p(\mu), L_q(\mu))$.

Conversely, let $L_p(\mu) \subset L_1(\mathfrak{P})$ and $\mathcal{T}_{\text{LS}}(\mathfrak{P}) \subset \sigma_w(L_p(\mu), L_q(\mu))$, then for every $P \in \mathfrak{P}$ the map $f \mapsto \int f dP$ on $L_p(\mu)$ is $\sigma_w(L_p(\mu), L_q(\mu))$ -continuous, that is, a continuous linear functional on $L_p(\mu)$. By Riesz's theorem there is a $g \in L_q(\mu)$ such that

$$\int f dP = \int f g d\mu \quad \text{for all } f \in L_p(\mu).$$

Thus, P has a μ -density in $L_q(\mu)$. \blacksquare

Depending on the “quality” of a dominated statistical model \mathfrak{P} , Theorem 3.9 explains when the weak topology constitutes an upper bound for the Lehmann–Scheffé topology $\mathcal{T}_{\text{LS}}(\mathfrak{P})$. In fact, the latter is usually strictly weaker than the weak topology (see also Example 3.8).

To conclude from the Lehmann–Scheffé topology to its upper bound, we can use statements about the relative weak compactness in $L_p(\mu)$. The following statement is given for the dimension $k = 1$. A similar result was noted before in Schmetterer (1978, Satz 3.2, page 496 f.).

3.10 Theorem Let $p, q \in]1, \infty[$, $p \leq q$ be conjugate exponents and let \mathfrak{P} be dominated and complete. We denote by Q a \mathfrak{P} -equivalent distribution.¹⁰ Assume furthermore that \mathfrak{P} has a Q -density choice in $L_q(Q)$.

(a) The LS topology $\mathcal{T}_{\text{LS}}(\mathfrak{P})$ and the weak topology $\sigma_w(L_p(Q), L_q(Q))$ coincide on any bounded subset M of $L_q(Q)$.

⁹Let \mathfrak{P} be a statistical model that is dominated by a σ -finite measure μ . A family $(\varphi_P)_{P \in \mathfrak{P}}$ is called μ -density choice of \mathfrak{P} w. r. t. μ iff for every $P \in \mathfrak{P}$ the function φ_P is a density of P w. r. t. μ .

¹⁰If \mathfrak{P} is a dominated statistical model, then there is a dominating distribution Q such that Q is *equivalent* to \mathfrak{P} , that is, any $N \in \mathcal{A}$ is a \mathfrak{P} -null set if and only if N is a Q -null set (see also Witting, 1985, Satz 1.136, page 133).

(b) Let $M \subset L_1(Q)$ be bounded and uniformly integrable w. r. t. Q . Then $(M, \mathcal{T}_{\text{LS}}(\mathfrak{F}) \cap M) = (M, \sigma_w(L_1(Q), L_\infty(Q)) \cap M)$. \blacklozenge

Proof: (a) Since $p \leq q$, we note that $M \subset L_q(Q) \subset L_p(Q)$. According to Theorem 3.9 we have $\mathcal{T}_{\text{LS}}(\mathfrak{F}) \subset \sigma_w(L_p(Q), L_q(Q))$, so this holds in particular on any bounded set.

For the converse direction, let $M \subset L_q(Q)$ be a bounded set, that is,

$$\sup\{\|f\|_{Q,q} : f \in M\} < \infty. \quad (3.11)$$

Then, according to the theorem of Banach–Alaoglu M is relatively weakly- $*$ -compact in $(L_q(Q))' = L_p(Q)$, that is, relatively weakly compact in $L_p(Q)$. Since $\mathcal{T}_{\text{LS}}(\mathfrak{F}) \subset \sigma_w(L_p(Q), L_q(Q))$, we have that M is relatively compact w. r. t. the Hausdorff topology $\mathcal{T}_{\text{LS}}(\mathfrak{F})$, too. Hence

$$(M, \mathcal{T}_{\text{LS}}(\mathfrak{F}) \cap M) = (M, \sigma_w(L_p(Q), L_q(Q)) \cap M),$$

since within any chain of topologies there is at most one compact Hausdorff-topology.¹¹

Part (b) follows similarly by Werner (2018, Satz VIII.6.9, page 473). \blacksquare

Consider a statistical model \mathfrak{F} that is dominated by a \mathfrak{F} -equivalent distribution Q . Assume that \mathfrak{F} has a Q -density choice in $L_q(Q)$, then we denote by $\mathfrak{F}_\infty := \mathfrak{F}_\infty^{Q,q}$ the extended statistical model of \mathfrak{F} , that is,

$$\mathfrak{F}_\infty = \{P : P \text{ has a density w. r. t. } Q \text{ in } L_q(Q)\}.$$

In the following theorem we are concerned with Lehmann–Scheffé topologies w. r. t. such extended models. To this end, we also recall that a dominated statistical model \mathfrak{F} contains a countable submodel $\mathfrak{F}_0 \subset \mathfrak{F}$ and a distribution $Q := \sum_{P \in \mathfrak{F}_0} \omega_P P$ that dominates \mathfrak{F} , where $(\omega_P)_{P \in \mathfrak{F}_0}$ is strictly positive with $\sum_{P \in \mathfrak{F}_0} \omega_P = 1$, see also Proposition 1.2 on page 4.

3.11 Theorem Let \mathfrak{F} be a dominated and complete statistical model. Let Q be a \mathfrak{F} -equivalent distribution as given above and assume that \mathfrak{F} has a density choice w. r. t. Q in $L_q(Q)$.

Then, $\mathcal{T}_{L_1(\mathfrak{F}_\infty)}$, $\mathcal{T}_{L_1(\mathfrak{F})}$, $\mathcal{T}_{L_1(\mathfrak{F}_0)}$, and the norm topology $\tau(\|\cdot\|_Q)$ coincide on any order bounded¹² subset of $L_p(Q)$. \blacklozenge

Proof: Since $\mathfrak{F}_0 \subset \mathfrak{F} \subset \mathfrak{F}_\infty$ we have that $\mathcal{T}_{L_1(\mathfrak{F}_0)} \subset \mathcal{T}_{L_1(\mathfrak{F})} \subset \mathcal{T}_{L_1(\mathfrak{F}_\infty)}$. Furthermore, we have obviously

$$(L_p(Q), \mathcal{T}_{L_1(\mathfrak{F}_0)}) = (L_p(Q), \tau(\|\cdot\|_Q)).$$

¹¹A topology \mathcal{T} on \mathfrak{X} is called *compact Hausdorff* iff $(\mathfrak{X}, \mathcal{T})$ is a compact Hausdorff space. For the statement see also Kelley *et al.* (1976, Exercise A (a), page 32).

¹²Let $(E, \|\cdot\|_E)$ be a normed space and let $H \subset E^{\mathfrak{X}}$ be a function space. A subset M of H is called *order bounded* (in H) iff there is an $h \in H$ such that $\|f(x)\|_E \leq \|h(x)\|_E$ for all $x \in \mathfrak{X}$ and all $f \in M$.

It remains to show that $\mathcal{T}_{L_1(\mathfrak{P}_\infty)} \subset \tau(\|\cdot\|_Q)$ on order bounded subsets. To this end, let M be an order bounded subset of $L_p(Q)$ and let $(f_\alpha)_{\alpha \in D} \in M^D$ be a net¹³ that converges to $f \in L_p(Q)$ w. r. t. the $\|\cdot\|_Q$ -norm, that is,

$$\lim_{\alpha \in D} \int |f_\alpha - f| dQ = 0.$$

We then also have that

$$\lim_{\alpha \in D} \int |f_\alpha - f| g dQ = 0 \quad \text{for all } g \in L_\infty(Q).$$

Furthermore, let $\varphi \in L_q(Q)$ be a probability density w. r. t. Q we then have

$$\lim_{\alpha \in D} \int |f_\alpha - f| \varphi dQ = 0 \quad (3.12)$$

as well, since $L_\infty(Q)$ is $\|\cdot\|_{Q,q}$ -dense in $L_q(Q)$. Hence, we have

$$\lim_{\alpha \in D} \int |f_\alpha - f| dP = 0 \quad \text{for all } P \in \mathfrak{P}_\infty.$$

That means, the arbitrarily chosen $\tau(\|\cdot\|_Q)$ -convergent net $(f_\alpha)_{\alpha \in D}$ converges also w. r. t. $\mathcal{T}_{L_1(\mathfrak{P}_\infty)}$ and hence the identity \mathbf{id}_M from $(M, \tau(\|\cdot\|_Q) \cap M)$ to $(M, \mathcal{T}_{L_1(\mathfrak{P}_\infty)} \cap M)$ is continuous (see e. g. Kelley, 1975, page 86), that is,

$$\mathcal{T}_{L_1(\mathfrak{P}_\infty)} \cap M \subset \tau(\|\cdot\|_Q) \cap M,$$

which yields the claim. ■

Another important question is that of metrizability.

3.12 Remark (Metrizability) Let \mathfrak{P} be any statistical model on $(\mathfrak{X}, \mathcal{A})$.

- (a) The LS topology is metrizable if and only if there is a countable $\mathfrak{P}_0 \subset \mathfrak{P}$ such that $\mathcal{T}_{\text{LS}}(\mathfrak{P}_0) = \mathcal{T}_{\text{LS}}(\mathfrak{P})$, see also Werner (2018, page 486). The same holds for the canonical topology $\mathcal{T}_{L_p(\mathfrak{P})}$.
- (b) If $\mathcal{T}_{\text{LS}}(\mathfrak{P})$ is metrizable, then \mathcal{A} is complete for the submodel \mathfrak{P}_0 (see part (a)) and also for \mathfrak{P} .
- (c) If $\mathcal{T}_{\text{LS}}(\mathfrak{P})$ is metrizable and consider M from Theorem 3.10, then

$$(M, \mathcal{T}_{\text{LS}}(\mathfrak{P}_0) \cap M) = (M, \sigma_w(L_p(Q), L_q(Q)) \cap M)$$

is metrizable, too. Note that $\sigma_w(L_p(Q), L_q(Q))$ itself is in general not metrizable. ◆

¹³Let D be a set *directed* by a partial order \leq , that is, for every $\alpha, \beta \in D$ there is a $\gamma \in D$ such that $\alpha \leq \gamma$ and $\beta \leq \gamma$. Then, a map from D to a topological space $(\mathfrak{X}, \mathcal{T})$ is called a *net*. A net $(x_\alpha)_{\alpha \in D}$ is said to *converge* to a point $x_0 \in \mathfrak{X}$ iff for all $U \in \mathcal{T}$ with $x_0 \in U$ there is a $\alpha_U \in D$ such that $x_\alpha \in U$ for all $\alpha \geq \alpha_U$. For these notions and concepts see e. g. Kelley (1975, page 65 ff.).

So far we have not assumed any structure of the model. However, \mathfrak{P} often exhibits a (canonical) measure theoretic structure, especially in the case where \mathfrak{P} is a parametrized model with parameter space $\Theta \subset \mathbb{R}^k$. The conception of continuity w. r. t. the Lehmann–Scheffé topology, as it was intended above (see also Lemma 3.3), may then lead to sequential continuity in terms of almost sure pointwise convergence. In fact, this kind of convergence cannot be represented by a topology. A more general concept is needed, the so-called *limit spaces*, which we, however, do not want to consider here. For the corresponding concepts the interested reader is referred to, for example, Kowalsky (1954) and Fischer (1959).

Let $(\mathfrak{P}, \mathcal{F}, \pi)$ denote the measure theoretic structure of the model \mathfrak{P} . Then the implicitly given Lebesgue space of functions on \mathfrak{P} consequently leads to reasonable semi-norms and norms on estimator spaces. In the following section we therefore introduce corresponding spaces which we call *Lehmann–Scheffé spaces*.

3.2 Lehmann–Scheffé Spaces

The goal of this section is to introduce appropriate normed spaces, by which we can describe the continuity issues of the Rao–Blackwell inverse problem w. r. t. ancillarity of first order. Henceforth, we assume that a *statistical model* $\mathfrak{P} \subset \text{Prob}(\mathfrak{X}, \mathcal{A})$ is given in terms of a σ -finite measure space $(\mathfrak{P}, \mathcal{F}, \pi)$.¹⁴ Our intention is then to mimic the Lehmann–Scheffé topologies.

Let $p \in [1, \infty]$. Then, by $\text{LS}_{\pi, p}(\mathfrak{P}) = \text{LS}_{\pi, p}(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ we denote the set of all $[\mathfrak{P}]$ -equivalence classes of measurable functions f such that $P \mapsto \int f dP$ is \mathcal{F} -measurable and¹⁵

$$\|f\|_{\text{LS}, p} := \|\mathbf{E}f\|_{\pi, p} = \sqrt[p]{\int q_P(f)^p \pi(dP)} < \infty, \quad (3.13)$$

if $p < \infty$, and¹⁶

$$\|f\|_{\text{LS}, \infty} := \|\mathbf{E}f\|_{\pi, \infty} = \inf \left\{ \sup \left\{ \left\| \int f dP \right\| : P \in N^c \right\} : N \in \mathcal{F}, \pi(N) = 0 \right\} < \infty,$$

if $p = \infty$, respectively. We call $(\text{LS}_{\pi, p}(\mathfrak{P}), \|\cdot\|_{\text{LS}, p})$ *Lehmann–Scheffé (LS) space*, which is a semi-normed vector space.

3.13 Proposition The set $\text{LS}_{\pi, p}(\mathfrak{P})$ with the map $\|\cdot\|_{\text{LS}, p}$ in (3.13) is a semi-normed vector space. \blacklozenge

¹⁴If π is a probability measure, it can be interpreted as *prior distribution*.

¹⁵Recall, $q_P(f) := \left\| \int f dP \right\|$, see (3.4) on page 51.

¹⁶By $\|f\|_{\mu, \infty}$ we denote the essential supremum of f w. r. t. the μ -null sets.

Proof: Note that \mathbf{E} is obviously a well-defined linear operator from $\text{LS}_{\pi,p}(\mathfrak{F})$ to $\text{L}_p(\mathfrak{F}, \mathcal{F}, \pi)$ and $\|\cdot\|_{\text{LS},p} = \|\mathbf{E}(\cdot)\|_{\pi,p}$. Since $\|\cdot\|_{\pi,p}$ is a norm, the claim follows easily by the linearity of \mathbf{E} . ■

Even if we assume that \mathcal{A} is complete for \mathfrak{F} , the Lehmann–Scheffé space $(\text{LS}_{\pi,p}(\mathfrak{F}), \|\cdot\|_{\text{LS},p})$ is in general not a normed space as the following example shows.

3.14 Example (Lebesgue Space) Let $\mathfrak{F} = (\delta_x : x \in \mathfrak{X})$, then \mathfrak{F} can be identified with $(\mathfrak{X}, \mathcal{A})$. Furthermore, let π be a measure on $(\mathfrak{X}, \mathcal{A})$. In this case, $\text{LS}_{\pi,p}(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$ turns out to be the usual semi-normed Lebesgue–Bochner space $\mathcal{L}_p(\mathfrak{X}, \mathcal{A}, \pi)$. In fact, if \mathfrak{X} is uncountable

$$\|f\|_{\text{LS},p} := \sqrt[p]{\int \|f\|^p \pi(dx)} \quad \text{for } f \in \mathcal{L}_p(\mathfrak{X}, \mathcal{A}, \pi)$$

is merely a semi-norm, although \mathcal{A} is complete for \mathfrak{F} . This is due to the fact that $[\mathfrak{F}] = \{\mathbf{0}\} \neq [\pi]$, in particular the \mathfrak{F} -null sets are not equivalent to the π -null sets. ◆

Semi-normed spaces fail to identify their elements by the semi-norm. To get a normed vector space $\text{LS}_{\pi,p}(\mathfrak{F})$, Example 3.14 explains that $f(x) = 0$ for π -almost all $x \in \mathfrak{X}$ should be sufficient to identify f with $\mathbf{0}$. We therefore introduce the following notion of completeness.

3.15 Definition Let π be a measure on a statistical model \mathfrak{F} . A σ -algebra $\mathcal{C} \subset \mathcal{A}$ is called π -complete for \mathfrak{F} iff $\int f dP = 0$ for π -almost all $P \in \mathfrak{F}$ is within $\text{L}_1(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$ only fulfilled by $f = \mathbf{0}$. Furthermore, \mathfrak{F} is called π -complete iff \mathcal{A} is π -complete. ◆

The notion of π -completeness of the underlying model \mathfrak{F} is the correct ingredient to conclude that $\|\cdot\|_{\text{LS},p}$ is a norm.

3.16 Proposition The vector space $\text{LS}_{\pi,p}(\mathfrak{X}, \mathcal{A}, \mathfrak{F})$ with $\|\cdot\|_{\text{LS},p}$ is a normed space if and only if \mathfrak{F} is π -complete. ◆

Proof: Let \mathfrak{F} be π -complete and consider $f \in \text{LS}_{\pi,p}$ with $\|f\|_{\text{LS},p} = 0$. This implies $\|\int f dP\| = 0$ for π -almost every $P \in \mathfrak{F}$. By the π -completeness we may conclude that $f = \mathbf{0}$ almost surely w. r. t. \mathfrak{F} and hence definiteness of $\|\cdot\|_{\text{LS},p}$ holds.

Let $\|\cdot\|_{\text{LS},p}$ be a norm. Then, $\int f dP = 0$ for π -almost all $P \in \mathfrak{F}$ implies $\|f\|_{\text{LS},p} = 0$. By the definiteness of a norm we have $f = \mathbf{0}$ almost surely w. r. t. \mathfrak{F} . This implies π -completeness of \mathfrak{F} . ■

In Example 3.14 we have that \mathfrak{F} is complete, however, if \mathfrak{X} is uncountable, there is no σ -finite measure π such that \mathfrak{F} is π -complete. Clearly, π -completeness implies completeness and, hence, is a stronger notion. For

the converse direction, π has to be consistent with a topological structure. That is, we need that π has full support w. r. t. the Lehmann-Scheffé- $*$ -topology

$$\mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F})) = \tau(P \mapsto \int f \, dP : f \in L_p(\mathfrak{F})),$$

which we introduced in Section 3.1, page 51. We henceforth consider the statistical model $(\mathfrak{F}, \mathfrak{B}, \pi)$ with the Borel- σ -algebra

$$\mathfrak{B} = \mathfrak{B}(\mathfrak{F}, \mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F}))).$$

For a statement concerning the above-mentioned converse direction, we need the following lemma.

3.17 Lemma Let $(\mathfrak{X}, \mathcal{T})$ be a topological space and let μ be a measure on \mathfrak{X} with full support w. r. t. \mathcal{T} , that is, $\text{supp } \mu = \mathfrak{X}$. If $f : (\mathfrak{X}, \mathcal{T}) \rightarrow \mathbb{R}^k$ is continuous, then $\mu(\{f \neq \mathbf{0}\}) = 0$ implies $f = \mathbf{0}$. \blacklozenge

Proof: Consider the *open* set $V := \mathbb{R}^k \setminus \{\mathbf{0}\}$. Since f is continuous, $f^{-1}(V)$ is open. Furthermore, $f^{-1}(V) \subset \{f \neq \mathbf{0}\}$, which yields

$$\mu(f^{-1}(V)) \leq \mu(\{f \neq \mathbf{0}\}) = 0.$$

Finally, $\mu(f^{-1}(V)) = 0$ and $\text{supp } \mu = \mathfrak{X}$ imply $f^{-1}(V) = \emptyset$, that is, $f(x) = \mathbf{0}$ for all $x \in \mathfrak{X}$. \blacksquare

3.18 Lemma Let \mathfrak{F} be complete and assume π has full support w. r. t. the Lehmann–Scheffé- $*$ -topology $\mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_1(\mathfrak{F}))$, then \mathfrak{F} is π -complete. \blacklozenge

Proof: For any $f \in L_1(\mathfrak{F})$ with

$$\int f \, dP = \mathbf{0} \quad \text{for } \pi\text{-almost all } P \in \mathfrak{F},$$

we have by the $\mathcal{T}_{\mathbf{LS}}^*$ -continuity and by Lemma 3.17 that $\int f \, dP = \mathbf{0}$ holds for all $P \in \mathfrak{F}$. Since \mathfrak{F} is complete, $f = \mathbf{0}$ almost surely w. r. t. \mathfrak{F} . \blacksquare

3.19 Example (Lebesgue Space, Ctd.) In Example 3.14 the topology $\mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F}))$ equals the power set $2^{\mathfrak{X}}$. To show this, we recall from the very beginning of this chapter that $(\mathfrak{X}, \mathcal{T})$ is assumed to be a Hausdorff space and, hence, all singletons are \mathcal{A} -measurable. Furthermore, for any $A \in \mathcal{A}$ the function $x \mapsto \int \mathbb{1}_A(y) \delta_x(dy)$ is $\mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F}))$ -continuous by definition. Hence, every $A \in \mathcal{A}$ and in particular every singleton is $\mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F}))$ -open, thus $2^{\mathfrak{X}} = \mathcal{T}_{\mathbf{LS}}^*(\mathfrak{F}, L_p(\mathfrak{F}))$.

Hence, if \mathfrak{X} is uncountable no σ -finite measure π on \mathfrak{X} has full support w. r. t. the weak- $*$ -topology $2^{\mathfrak{X}}$.¹⁷ \blacklozenge

¹⁷Let $(\mathfrak{X}, \mathcal{A}, \mu)$ be a σ -finite measure space. Then, any set system $\mathcal{D} \subset \mathcal{A}$ consisting of disjoint sets $D \in \mathcal{D}$ with positive μ -measure is countable.

3.20 Example (Continuous Uniform Model) Let $n \in \mathbb{N}$, let $\Theta := [0, t]$ for some $t > 0$ and consider¹⁸ $\mathfrak{P} := \{\mathbf{U}_{[0, \vartheta]}^{\otimes n} : \vartheta \in \Theta\}$. Let π be a distribution on Θ such that the support of π w. r. t. the Euclidean topology on Θ satisfies

$$\text{supp}_{\tau(\cdot|\cdot)}(\pi) = \Theta.$$

Note that $S(x) := \max(x)$ for $x \in \mathbb{R}_+^n$ is complete and sufficient for \mathfrak{P} , see for example Lehmann & Casella (1998, Example 6.23, page 42).

Since $\vartheta \mapsto \mathbf{U}_{[0, \vartheta]}^{\otimes n} h$ is continuous for every $h \in L_1(\mathfrak{X}, \sigma(S), \mathcal{Q})$ we have that $\mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_1(\mathfrak{P})) \subset \tau(\cdot|\cdot)$, hence,

$$\text{supp}_{\mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_1(\mathfrak{P}))}(\pi) \supset \text{supp}_{\tau(\cdot|\cdot)}(\pi) = \Theta,$$

that is, π has also full support w. r. t. the Lehmann–Scheffé-* -topology. Consequently, by Lemma 3.18 it follows that $\sigma(S)$ is π -complete for \mathfrak{P} . ♦

Furthermore, π -completeness can be generally shown for an important class of parametrized statistical models, the exponential families. This is formulated in the following proposition.

3.21 Proposition (Exponential Family and π -Completeness) For some non-empty, open parameter space $\Theta \subset \mathbb{R}^k$ let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be an exponential family with sufficient statistic S and reference measure ν . That means, for every $\vartheta \in \Theta$, the distribution P_ϑ has a ν -density of the form¹⁹

$$f_\vartheta := \frac{dP_\vartheta}{d\nu} \propto x \mapsto g(x) \exp(\langle \vartheta, S(x) \rangle),$$

for some non-negative, measurable g . Then, for every measure π on Θ with full support w. r. t. $\tau(\|\cdot\|_2)$ the σ -algebra $\sigma(S)$ is π -complete for \mathfrak{P} . ♦

Proof: We have $\mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_1(\mathfrak{X}, \sigma(S), \mathfrak{P})) \subset \tau(\|\cdot\|_2)$, since for $h \in L_1(\mathfrak{X}, \sigma(S), \mathfrak{P})$ the function $\vartheta \mapsto \int h(x) f_\vartheta(x) \nu(dx)$ is continuous (see Pfanzagl, 1994, Lemma 1.6.6, page 24). Consequently, any measure π on Θ with $\text{supp } \pi = \Theta$ w. r. t. the Euclidean topology $\tau(\|\cdot\|_2)$ has also a full support w. r. t. the LS-* -topology. Finally, by the well-known fact that $\sigma(S)$ is complete for \mathfrak{P} (see for example Pfanzagl, 1994, Theorem 1.6.10, page 26) the claim follows by Lemma 3.18. ■

If \mathfrak{P} is π -complete then $(\text{LS}_{\pi, p}(\mathfrak{P}), \|\cdot\|_{\text{LS}, p})$ is a normed space. In addition, we check for norm completeness. In general $(\text{LS}_{\pi, p}(\mathfrak{P}), \|\cdot\|_{\text{LS}, p})$ is not complete as the following simple example shows.

3.22 Example Consider $\mathfrak{P} := (\mathbf{N}(\mu, \sigma^2) : \mu \in [0, 1] \cap \mathbb{Q})$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let ν be a distribution on $[0, 1] \cap \mathbb{Q}$ with $\text{supp } \nu = [0, 1] \cap \mathbb{Q}$. By \mathcal{Q} we denote

¹⁸ Let $H \in \mathcal{H}(\mathbb{R}^k)$ with positive but finite Lebesgue measure $\lambda(H)$, then the uniform distribution on H is denote $\mathbf{U}_H := \frac{1}{\lambda(H)} \lambda|_H$.

¹⁹ See also Definition 1.36 on page 27.

the ν -mixture of \mathfrak{F} , that is, $Q(B) := \int \mathbf{N}(\mu, \sigma^2)(B) \nu(d\mu)$ for $B \in \mathfrak{B}(\mathbb{R})$. Then, for any $f \in L_1(Q) = L_1(\mathfrak{F})$ we have

$$\begin{aligned} \|f\|_{\mathbf{LS},1} &= \int_{[0,1]} \left| \int f \, d\mathbf{N}(\mu, \sigma^2) \right| \nu(d\mu) \\ &\leq \int |f| \, dQ = \|f\|_Q. \end{aligned}$$

Hence, the $\|\cdot\|_{\mathbf{LS},1}$ -norm is weaker than the $\|\cdot\|_Q$ -norm, in particular $L_1(Q) \subset \mathbf{LS}_1(\mathfrak{F})$. Since $\mathbf{LS}_1(\mathfrak{F}) \subset L_1(\mathfrak{F}) = L_1(Q)$, we have $\mathbf{LS}_1(\mathfrak{F}) = L_1(Q)$. Note that $(L_1(Q), \|\cdot\|_Q)$ is complete. If $(L_1(Q), \|\cdot\|_{\mathbf{LS},1})$ would be a Banach space, the two norms are equivalent.²⁰ However, this does not hold as can be shown again by the Rademacher sequence.

For the Rademacher sequence $(f_n)_{n \in \mathbb{N}} \in L_1(Q)^{\mathbb{N}}$ we have for every $\mu \in \mathbb{R}$ (see Example 3.7 and 3.8)

$$\left| \int f_n \, d\mathbf{N}(\mu, \sigma^2) \right| \leq 1 \text{ and } \int f_n \, d\mathbf{N}(\mu, \sigma^2) \rightarrow 0 \text{ for } n \rightarrow \infty.$$

Hence, by Lebesgue's theorem, we have

$$\lim_{n \rightarrow \infty} \|f_n\|_{\mathbf{LS},1} = \int_{[0,1]} \lim_{n \rightarrow \infty} \left| \int f_n \, d\mathbf{N}(\mu, \sigma^2) \right| \nu(d\mu) = 0.$$

However, $\|f_n\|_Q = Q([0,1]) > 0$ for all $n \in \mathbb{N}$, that is, $f_n \not\rightarrow \mathbf{0}$ for $n \rightarrow \infty$ in $L_1(Q)$. Hence, $(\mathbf{LS}_{\pi,1}(\mathfrak{F}), \|\cdot\|_{\mathbf{LS},1})$ cannot be a Banach space. \blacklozenge

3.23 Remark (Norm Completion) Let \mathfrak{F} be π -complete. To give at least an completion of $\mathbf{LS}_{\pi,p}$, we consider any $f \in \mathbf{LS}_{\pi,p}(\mathfrak{F})$ under the *extended* expectation operator (coordinate transformation)

$$\begin{aligned} \mathbf{E}: \quad \mathbf{LS}_{\pi,p}(\mathfrak{F}) &\rightarrow L_p(\mathfrak{F}, \mathfrak{B}, \pi) \\ f &\mapsto (\mathbf{E}_P f)_{P \in \mathfrak{F}}, \end{aligned}$$

where $\mathbf{E}_P f = \int f \, dP$ exists for π -almost all P . By the π -completeness of \mathfrak{F} , we can directly conclude that \mathbf{E} is injective. Furthermore, \mathbf{E} respects the norm, since we have by definition

$$\|\mathbf{E}(f)\|_{L_p(\pi)}^p = \int \left\| \int f \, dP \right\|^p \pi(dP) = \|f\|_{\mathbf{LS}_{\pi,p}}^p$$

for every $f \in \mathbf{LS}_{\pi,p}$. Hence, the closure of the range $\mathbf{E}(\mathbf{LS}_{\pi,p})$ in $L_p(\pi)$ can be identified with a completion of $\mathbf{LS}_{\pi,p}(\mathfrak{F})$. \blacklozenge

²⁰Let \mathfrak{X} be a linear space such that $\|\cdot\|_1$ and $\|\cdot\|_2$ make \mathfrak{X} a Banach space and assume that $\|\cdot\|_1 \leq K \cdot \|\cdot\|_2$ for some $K > 0$. Then, $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent due to the open mapping theorem (see e. g. Werner, 2018, Theorem IV.3.3, page 168).

Finally, we want to point out that $(\text{LS}_{\pi,p}(\mathfrak{P}), \|\cdot\|_{\text{LS},p})$ for $p = 2$ turns out to be an inner product space.

3.24 Remark (Inner Product Space) Note that $(\text{LS}_{\pi,p}(\mathfrak{P}), \|\cdot\|_{\text{LS},p})$ for $p = 2$ is a inner product space with inner product

$$(f, g) \mapsto \langle f, g \rangle_{\text{LS}} := \int \left\langle \int f \, dP, \int g \, dP \right\rangle \pi(dP).$$

A corresponding completion can be achieved as shown in Remark 3.23. \blacklozenge

In the above example we have already compared the $\|\cdot\|_{\text{LS},1}$ with $\|\cdot\|_Q$, where Q is a mixture of \mathfrak{P} . In what follows, we relate this to the canonical topology $\mathcal{T}_{L_1(\mathfrak{P})}$.

3.25 Remark Let \mathcal{A} be π -complete for \mathfrak{P} . We define²¹

$$Q(A) := \pi \cdot \mathbf{id}_{\mathfrak{P}} = \int_{\mathfrak{P}} P(A) \pi(dP) \quad \text{for } A \in \mathcal{A}$$

and consider the canonical topology $\mathcal{T}_{L_p(\mathfrak{P})}$ on $L_p(\mathfrak{P})$, which is generated by the semi-norms $\|\cdot\|_P$ for $P \in \mathfrak{P}$ (see also Chapter 1 on page 8).

Analogously to the Lehmann–Scheffé space, we then consider all functions $f \in L_p(\mathfrak{P})$ such that $P \mapsto \|f\|_{P,p}$ is \mathfrak{B} -measurable and

$$\|f\|_{Q,p} = \sqrt[p]{\int \|f\|^p \, dQ} = \sqrt[p]{\int \int \|f\|^p \, dP \, \pi(dP)} = \sqrt[p]{\int \|f\|_{P,p}^p \, \pi(dP)},$$

is finite. Hence, introducing a normed space with regard to the canonical topology leads to the usual Lebesgue space $L_p(Q)$. Furthermore, for $f \in L_p(Q)$ we have with Jensen's inequality²²

$$\|f\|_{\text{LS},p}^p = \int \|\mathbf{E}_P f\|^p \, \pi(dP) \leq \int \mathbf{E}_P \|f\|^p \, \pi(dP) = \|f\|_{Q,p}^p,$$

that is, $L_p(Q) \subset \text{LS}_{\pi,p}(\mathfrak{P})$ and $\|\cdot\|_{\text{LS},p}$ is generally weaker than $\|\cdot\|_{Q,p}$. \blacklozenge

In terms of these normed spaces, the well- or ill-posedness of the Rao–Blackwell inverse problem corresponds to the question under what conditions

$$\mathbf{E}(\cdot|\mathcal{C}) : (L_p(\mathfrak{X}, \mathcal{A}, Q), \|\cdot\|_{\text{LS},p}) \rightarrow (L_p(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q,p})$$

is continuous (see also Lemma 3.3). As we have seen, $\|\cdot\|_{\text{LS},p}$ is generally weaker than $\|\cdot\|_{Q,p}$, since it basically measures the location of a function.

²¹Note that $\mathbf{id}_{\mathfrak{P}}$ is a Markov kernel from $(\mathfrak{P}, \mathfrak{B})$ to $(\mathfrak{X}, \mathcal{A})$, since $\mathbf{id}_{\mathfrak{P}}(P) = P(\cdot)$ is a distribution for any $P \in \mathfrak{P}$. Furthermore, $P \mapsto P(A) = \mathbf{id}_{\mathfrak{P}}(P)(A)$ is \mathfrak{B} -measurable for every $A \in \mathcal{A}$ (see also page 63 regarding the considered σ -algebra \mathfrak{B}).

²²Note that $P \in \mathfrak{P}$ are probability measures.

This drawback can be compensated by measuring the oscillation as well. To this end, we consider the linear operator

$$\begin{aligned} \mathbf{V} : L_p(\mathfrak{X}, \mathcal{A}, Q) &\rightarrow L_p(\mathfrak{X} \times \mathfrak{P}, \mathcal{A} \otimes \mathcal{B}, \pi \otimes \mathbf{id}_{\mathfrak{P}}) \\ f &\mapsto (f(x) - \mathbf{E}_P f)_{(x,P) \in \mathfrak{X} \times \mathfrak{P}} \end{aligned}$$

which we call *oscillation operator*. It is clear that

$$\|f\|_{\mathbf{OS},p} := \|\mathbf{V}f\|_{\pi \otimes \mathbf{id}_{\mathfrak{P}},p} = \sqrt[p]{\int \mathbf{E}_P \|f - \mathbf{E}_P f\|^p \pi(dP)} \quad \text{for } f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$$

is a semi-norm. To get a norm, we first of all consider the *null space* of that semi-norm

$$\mathbf{N}(\|\cdot\|_{\mathbf{OS},p}) := \{f : \|f\|_{\mathbf{OS},p} = 0\}.$$

Then, $\|\cdot\|_{\mathbf{OS},p}$ is a norm on the quotient space

$$L_p(Q) / \mathbf{N}(\|\cdot\|_{\mathbf{OS},p}) = \{f + \mathbf{N}(\|\cdot\|_{\mathbf{OS},p}) : f \in L_p(Q)\}.$$

We therefore take a closer look at the null space \mathbf{N} .

3.26 Example Consider $\mathfrak{P} = (\delta_x : x \in \mathfrak{X})$. Then $\|\cdot\|_{\mathbf{OS},p} = \mathbf{0}$, that is, $\|\cdot\|_{\mathbf{OS},p}$ is the null semi-norm and thus $\mathbf{N}(\|\cdot\|_{\mathbf{OS},p}) = L_p(Q)$. Hence, the quotient space $L_p(Q) / \mathbf{N}(\|\cdot\|_{\mathbf{OS},p})$ is a singleton which is a normed space with the null norm. \blacklozenge

3.27 Example Let $\mathfrak{P} = (\mathbf{N}(\mu, \sigma^2) : \mu \in \mathbb{R})$. Then

$$\mathbf{N}(\|\cdot\|_{\mathbf{OS},p}) = \{f : f \text{ is constant } \lambda\text{-almost surely}\},$$

since $\|f\|_{\mathbf{OS},p} = 0$ implies $\mathbf{E}_P \|f - \mathbf{E}_P f\|^p = 0$ for π -almost all $P \in \mathfrak{P}$. Hence, we have

$$f = \mathbf{E}_P f \quad \text{almost surely w. r. t. } P \text{ and for } \pi\text{-almost all } P \in \mathfrak{P}$$

which holds if and only if f is λ -almost surely constant. \blacklozenge

The oscillation semi-norm measures the oscillation, however, it fails to measure the location. Hence, it seems to be the right counterpart to the Lehmann–Scheffé norm. By introducing the norm

$$\|\cdot\|_* := \|\cdot\|_{\mathbf{LS},p} + \|\cdot\|_{\mathbf{OS},p},$$

we now measure both and can show in the case $p = 1$ that, in fact, $\|\cdot\|_*$ is equivalent to $\|\cdot\|_Q$.

3.28 Remark Obviously, we have for every P that $\|\mathbf{E}_P f\| \leq \mathbf{E}_P \|f\|$ and

$$\mathbf{E}_P \|f - \mathbf{E}_P f\| \leq \mathbf{E}_P \|f\| + \|\mathbf{E}_P f\| \leq 2\mathbf{E}_P \|f\|,$$

that is, $\|\mathbf{E}_P f\| + \mathbf{E}_P \|f - \mathbf{E}_P f\| \leq 3 \mathbf{E}_P \|f\|$. Hence, by the monotonicity of the integral we have $\|\cdot\|_* \leq 3 \|\cdot\|_{Q,1}$. Conversely, we can conclude that

$$\mathbf{E}_P \|f\| \leq \mathbf{E}_P \|f - \mathbf{E}_P f\| + \|\mathbf{E}_P f\|$$

and hence again by the monotonicity of the integral $\|\cdot\|_{Q,1} \leq \|\cdot\|_*$. \blacklozenge

Often, the Rao–Blackwell projection

$$\mathbf{E}(\cdot|\mathcal{C}) : (L_p(\mathfrak{X}, \mathcal{A}, Q), \|\cdot\|_{\mathbf{LS},p}) \rightarrow (L_p(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q,p}) \quad (3.14)$$

is not continuous or, correspondingly, the Rao–Blackwell inverse problem, that is, to look for a $g \in L_1(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$ such that

$$\mathbf{E}_P f = \mathbf{E}_P g \quad \text{for all } P \in \mathfrak{F}, \quad (3.15)$$

for any $f \in L_1(\mathfrak{F})$ is ill-posed. In the next section, we are interested in approximating (3.15) through a family of well-posed problems and, consequently, in approximating the Rao–Blackwell projection (3.14) through a family of continuous operators.

To this end, we will define $\mathbf{E}^{(*)}(f|\mathcal{C})$ not just through (3.15), but also by taking the oscillation into account. One strategy will be to include the oscillation norm $\|\cdot\|_{\mathbf{OS},p}$ in (3.15) as a penalty term. Another strategy may be to restrict the image space of $\mathbf{E}(\cdot|\mathcal{C})$, for example, to a finite-dimensional space by using a pseudo-solution approach.

CHAPTER 4

Regularized Rao–Blackwellization

In the previous chapter, we came across the Rao–Blackwell inverse problem. We have seen that the Rao–Blackwell projection may be discontinuous (see e. g. Example 3.6 and 3.8) and hence the RB inverse problem may be ill-posed. That means, a slight change of the parameter function can have a large impact on the corresponding Rao–Blackwellization. Since statistical models usually result from simplifications and, therefore, are inherently misspecified, the ill-posedness of the Rao–Blackwell inverse problem is quite unfortunate.

This chapter is devoted to regularization schemes¹ for such ill-posed Rao–Blackwell inverse problems. This leads directly to the notion of “regularized Rao–Blackwellization”, which we are now going to explain. To this end, suppose that a statistical model is given in terms of a finite Borel measure space

$$\left(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \mathcal{T}_{\mathbf{LS}}^*, \boldsymbol{\pi}\right),$$

where we recall that $\mathcal{T}_{\mathbf{LS}}^*(L_p(\mathfrak{X}))$ is the Lehmann–Scheffé-*-topology (see also Section 3.1 on page 52).

Let \mathcal{C} be complete and sufficient for \mathfrak{X} and let $\mathbf{E}_{\mathcal{C}} := \mathbf{E}|_{L_1(\mathfrak{X}, \mathcal{C}, \boldsymbol{\pi})}$. Then, we recall that $\mathbf{E}(f|\mathcal{C})$ for some $f \in L_p(\mathfrak{X}, \mathcal{A}, \boldsymbol{\pi})$ is the unique solution $g \in L_p(\mathfrak{X}, \mathcal{C}, \boldsymbol{\pi})$ of the following operator equation

$$\mathbf{E}_{\mathcal{C}} g = \mathbf{E}f. \tag{4.1}$$

¹Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be normed vector spaces, let $A : E \rightarrow F$ be an injective bounded linear operator, and consider some set $M \subset E$. Then, a family $(B_\alpha)_{\alpha>0}$ of bounded (linear) operators from F to E with

$$\lim_{\alpha \rightarrow 0} \|B_\alpha A e - e\|_E = 0 \quad \text{for all } e \in M$$

is called *regularization scheme* for the operator A on M . The parameter α is called *regularization parameter* (see also Kress, 1999; Ivanov *et al.*, 2002).

If, in addition, \mathcal{C} is π -complete for \mathfrak{F} , then $\mathbf{E}(f|\mathcal{C})$ is given by a solution of (4.1) w. r. t. $\|\cdot\|_{\pi,p}$ by which we mean a solution g such that

$$\|\mathbf{E}_{\mathcal{C}}g - \mathbf{E}f\|_{\pi,p} = \|f - g\|_{\mathbf{LS},p} = 0. \quad (4.2)$$

In fact, for any $g \in L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{F})$ that fulfills (4.2) we have

$$\int g \, dP = \int f \, dP = \int \mathbf{E}(f|\mathcal{C}) \, dP \quad \text{for } \pi\text{-almost all } P \in \mathfrak{F},$$

which is only true for $g = \mathbf{E}(f|\mathcal{C})$ due to the π -completeness of \mathcal{C} . This proves the following theorem² that constitutes the starting point for regularizations.

In the present Chapter we use as before

$$Q(A) := \pi \cdot \mathbf{id}_{\mathfrak{F}}(A) = \int P(A) \pi(dP) \quad \text{for } A \in \mathcal{A},$$

see also Footnote 21 on page 66.

4.1 Theorem (Best Approximation w. r. t. Ancillarity of First Order, II) Let $\mathcal{C} \subset \mathcal{A}$ be π -complete and sufficient for \mathfrak{F} , let $Q := \pi \cdot \mathbf{id}_{\mathfrak{F}}$, and let $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$. Then, we have

$$\mathbf{E}(f|\mathcal{C}) = \mathbf{E}_{\mathcal{C}}^{-1} \mathbf{E}f = \operatorname{argmin}\{\|f - g\|_{\mathbf{LS},p} : g \in L_p(\mathfrak{X}, \mathcal{C}, Q)\}. \quad (4.3)$$

◆

Proof: Follows directly by the previous paragraph. ■

By the nature of the Rao–Blackwellization, we are dealing with a projection and, hence, with a non-injective operator. To be precise, the notion of a *regularized* Rao–Blackwellization has to be clarified.

4.2 Definition (Regularized Rao–Blackwellization) Let $p \in [1, \infty]$ and let \mathcal{C} be sufficient and π -complete for \mathfrak{F} . Let³ $Q := \pi \cdot \mathbf{id}_{\mathfrak{F}}$ and let $(B_\alpha)_{\alpha>0}$ be a regularization scheme of

$$\mathbf{E}_{\mathcal{C}} : (L_p(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q,p}) \rightarrow (L_p(\mathfrak{F}, \mathfrak{B}, \pi), \|\cdot\|_{\pi,p}). \quad (4.4)$$

We then call $\mathbf{E}^{(\alpha)}(\cdot|\mathcal{C}) := B_\alpha \circ \mathbf{E}$ the α -regularized Rao–Blackwellization given by $(B_\alpha)_{\alpha>0}$ (see Figure 4.1). ◆

The discontinuity of $\mathbf{E}(\cdot|\mathcal{C})$ w. r. t. the Lehmann–Scheffé norm $\|\cdot\|_{\mathbf{LS},p}$ or the ill-posedness of the Rao–Blackwellization can be understood as an overfitting problem w. r. t. the $\|\cdot\|_{\mathbf{LS},p}$ -norm. To avoid this overfitting, quantitative and qualitative prior knowledge about the (pseudo-)solution has to be included in (4.3). On the one hand, this leads to the regularization approach due to Tikhonov, which we present in the following section. On the other hand, we arrive at regularizations by what we call Ivanov’s approach presented in the subsequent section.

²See also Proposition 1.34 on page 25.

³We write $\pi \cdot \mathbf{id}_{\mathfrak{F}}$ for the mixture distribution $A \mapsto \int P(A) \pi(dP)$ on $(\mathfrak{X}, \mathcal{A})$.

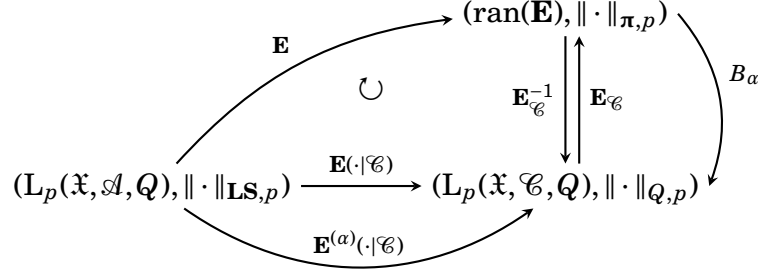


Figure 4.1: Illustration of the definition of the α -regularized Rao–Blackwellization. The arrow \circlearrowleft indicates that the corresponding inner part of the diagram is commutative, that is, $\mathbf{E}(\cdot|\mathcal{C}) = \mathbf{E}_{\mathcal{C}}^{-1} \circ \mathbf{E}$. The α -regularized Rao–Blackwellization is defined by $\mathbf{E}^{(\alpha)}(\cdot|\mathcal{C}) = B_{\alpha} \circ \mathbf{E}$, that is, the outer part of the diagram is by definition commutative as well.

4.1 Tikhonov's Regularization Approach

Basically, classical Rao–Blackwellization aims at smoothing an estimator by data reduction. Now, by the Tikhonov's regularization approach, we also account for another smoothing, that is, for a regularity condition through the oscillation semi-norm $\|\cdot\|_{\text{OS},2}$, in addition to the minimization problem (4.3). For $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$ and $p = n = 2$ the minimization problem in terms of a Tikhonov regularization then reads

$$\arg \inf \{ \|g - f\|_{\text{LS},2}^2 + \alpha \|g\|_{\text{OS},2}^2 : g \in L_2(\mathfrak{X}, \mathcal{C}, Q) \}, \quad (4.5)$$

where $\alpha > 0$. As seen in Example 3.22 ($L_2(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{\text{LS},2}$) is generally not a Hilbert space. However, the minimization problem (4.5) can be traced back to Hilbert spaces which we can use to show the following.

4.3 Theorem (Tikhonov Regularized Rao–Blackwellization) For every $\alpha > 0$ and every $f \in L_2(\mathfrak{X}, \mathcal{A}, Q)$ the minimization problem (4.5) has a unique solution given by⁴

$$\mathbf{E}^{(\alpha)}(f|\mathcal{C}) := (\alpha \cdot \mathbf{V}'\mathbf{V} + \mathbf{E}'\mathbf{E})^{-1} \mathbf{E}'\mathbf{E}f, \quad (4.6)$$

which is a α -regularized Rao–Blackwellization of f . \blacklozenge

Proof:⁵ By definition we have first of all $\|g - f\|_{\text{LS},2}^2 = \|\mathbf{E}(g - f)\|_{\pi,2}^2$ and $\|g\|_{\text{OS},2} = \|\mathbf{V}g\|_{\pi \otimes \text{id}_{\mathfrak{q}},2}$. Therefore, the minimization problem (4.5) corresponds to looking for a minimizer of the functional

$$\phi_m(g) := \|\mathbf{E}g - m\|_{\pi,2}^2 + \alpha \|\mathbf{V}g\|_{\pi \otimes \text{id}_{\mathfrak{q}},2}^2 \quad \text{for } g \in L_2(\mathfrak{X}, \mathcal{C}, Q) \quad (4.7)$$

⁴Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed spaces and let $T : E \rightarrow F$ be a linear operator. Then $T' : F' \rightarrow E'$ defined by $T'(f') := f'(T(\cdot))$ for $f' \in F'$ is called *adjoint operator* of T .

⁵The proof of Theorem 4.3 is given in detail for the stochastics community. The main arguments can be found in the literature on regularization theory (see e. g. Engl *et al.*, 1996). Essentially, we only have to show that $\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V}$ is strictly coercive.

where $m = \mathbf{E}f \in L_2(\boldsymbol{\pi})$. We show that $\phi_m(\cdot)$ has a unique minimum which is continuous in $m \in L_2(\boldsymbol{\pi})$. Since

$$\phi_m(g+h) - \phi_m(g) = 2\langle \mathbf{E}'(\mathbf{E}g - m) + \alpha \mathbf{V}'\mathbf{V}g, h \rangle_{L_2(Q)} + \|\mathbf{E}h\|_{\boldsymbol{\pi},2}^2 + \alpha \|\mathbf{V}h\|_{\boldsymbol{\pi} \otimes \text{id}_{\mathfrak{P}},2}^2$$

for all $g, h \in L_2(Q)$, we first of all get the Fréchet differentiability of ϕ_m with

$$\mathbf{D}\phi_m(g)(h) = 2\langle \mathbf{E}'(\mathbf{E}g - m) + \alpha \mathbf{V}'\mathbf{V}g, h \rangle_{L_2(Q)} \quad \text{for } h \in L_2(Q)$$

and furthermore

$$\mathbf{D}^2\phi_m(g)(h, k) = 2\langle (\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V})h, k \rangle_{L_2(Q)} \quad \text{for } h, k \in L_2(Q).$$

For $\alpha \in]0, 1]$ we note that $\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V}$ is strictly coercive⁶, because we have for all $f \in L_2(Q)$

$$\begin{aligned} \alpha \|f\|_{Q,2}^2 &= \alpha \int \int f^2 dP \boldsymbol{\pi}(dP) \\ &= \alpha \int \int (f - \mathbf{E}_P f + \mathbf{E}_P f)^2 dP \boldsymbol{\pi}(dP) \\ &= \alpha \int \int (f - \mathbf{E}_P f)^2 dP \boldsymbol{\pi}(dP) + \alpha \int (\mathbf{E}_P f)^2 \boldsymbol{\pi}(dP) \\ &\leq \|\mathbf{E}f\|_{\boldsymbol{\pi},2}^2 + \alpha \|\mathbf{V}f\|_{\boldsymbol{\pi} \otimes \text{id}_{\mathfrak{P}},2}^2 \\ &= \langle (\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V})f, f \rangle. \end{aligned}$$

Hence, there is a unique minimizer e of ϕ_m , which is implicitly given by

$$\mathbf{D}\phi_m(e)(h) = 2\langle \mathbf{E}'(\mathbf{E}e - m) + \alpha \mathbf{V}'\mathbf{V}e, h \rangle = 0 \quad \text{for all } h \in L_2(Q)$$

or equivalently

$$(\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V})e = \mathbf{E}'m.$$

The theorem of Lax–Milgram⁷ yields that $(\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V})^{-1}$ exists and is bounded. Note that $(B_\alpha)_{\alpha>0}$ given by $B_\alpha := (\mathbf{E}'\mathbf{E} + \alpha \mathbf{V}'\mathbf{V})^{-1}\mathbf{E}'$ is a regularization scheme for $\mathbf{E}_{\mathcal{C}}$, because for $f \in L_2(\mathfrak{X}, \mathcal{C}, Q)$ we have by Engl *et al.* (1996, Theorem 5.2, page 118)

$$\lim_{\alpha \rightarrow 0} \|f - B_\alpha \mathbf{E}f\|_{Q,2}^2 = 0.$$

In summary, we have that

$$\mathbf{E}^{(\alpha)}(f|\mathcal{C}) := (\alpha \cdot \mathbf{V}'\mathbf{V} + \mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{E}f \quad (4.8)$$

⁶Let E be a Hilbert space. An operator $T: E \rightarrow E$ is called *strictly coercive* iff $\langle Tx, x \rangle \geq c \cdot \|x\|^2$ for all $x \in E$ and some $c > 0$.

⁷Lax–Milgram Theorem: Let E be a Hilbert space. Then, a bounded and strictly coercive operator $T: E \rightarrow E$ has a bounded inverse operator $T^{-1}: E \rightarrow E$ (see for example Werner, 2018, page 266).

is the unique solution of (4.5). Furthermore, since

$$\mathbf{E} : (\mathbf{L}_2(\mathfrak{X}, \mathfrak{A}, \mathbf{Q}), \|\cdot\|_{\mathbf{LS},2}) \rightarrow (\mathbf{L}_2(\mathfrak{Y}, \mathfrak{F}, \boldsymbol{\pi}), \|\cdot\|_{\mathbf{L}_2(\boldsymbol{\pi},2)})$$

is continuous by definition of $\|\cdot\|_{\mathbf{LS},2}$, we have that

$$\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C}) : (\mathbf{L}_2(\mathfrak{X}, \mathfrak{A}, \mathbf{Q}), \|\cdot\|_{\mathbf{LS},2}) \rightarrow (\mathbf{L}_2(\mathfrak{X}, \mathfrak{C}, \mathbf{Q}), \|\cdot\|_{\mathbf{Q},2})$$

given by (4.8) is continuous. \blacksquare

4.4 Remark (Singular Value Decomposition) **(a)** Assume that $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})$ is a compact linear operator from $\mathbf{L}_2(\mathfrak{X}, \mathfrak{A}, \mathbf{Q})$ to $\mathbf{L}_2(\mathfrak{X}, \mathfrak{C}, \mathbf{Q})$. To compute the α -regularized Rao–Blackwellization, we note that there is a singular value decomposition $(\lambda_n, u_n, v_n)_{n \in \mathbb{N}}$ of $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})$, that is, $(u_n)_{n \in \mathbb{N}}$ and $(v_n)_{n \in \mathbb{N}}$ are orthonormal bases of $(\mathbf{L}_2(\mathfrak{X}, \mathfrak{A}, \mathbf{Q}), \|\cdot\|_{\mathbf{Q},2})$ and $(\mathbf{L}_2(\mathfrak{X}, \mathfrak{C}, \mathbf{Q}), \|\cdot\|_{\mathbf{Q},2})$, respectively, and λ_n^2 , $n \in \mathbb{N}$ are eigenvalues of $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})' \mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})$ such that

$$\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})u_n = \lambda_n v_n, \quad \mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})'v_n = \lambda_n u_n$$

for all $n \in \mathbb{N}$ (see for example Kress, 1999, page 277). Then, a simple computation yields

$$\mathbf{E}^{(\alpha)}(f|\mathfrak{C}) = \sum_{n \in \mathbb{N}} \lambda_n \langle f, u_n \rangle_{\mathbf{L}_2(\mathbf{Q})} v_n \quad \text{for } f \in \mathbf{L}_2(\mathfrak{X}, \mathfrak{A}, \mathbf{Q}).$$

(b) Assume that u_1, \dots, u_n are orthonormal eigenfunctions to corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ of $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})$, that is,

$$\mathbf{E}^{(\alpha)}(u_i|\mathfrak{C}) = \lambda_i u_i \quad \text{for } i = 1, \dots, n.$$

We then have that

$$H := \text{span}\{u_1, \dots, u_n\} \subset \mathbf{L}_2(\mathfrak{X}, \mathfrak{C}, \mathbf{Q})$$

has an orthogonal complement in $\mathbf{L}_2(\mathfrak{X}, \mathfrak{C}, \mathbf{Q})$ and, furthermore, that H is invariant w. r. t. $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})$, since

$$\langle \mathbf{E}^{(\alpha)}(u_i|\mathfrak{C}), f \rangle_{\mathbf{L}_2(\mathbf{Q})} = \lambda_i \langle u_i, f \rangle_{\mathbf{L}_2(\mathbf{Q})} = 0 \quad \text{for all } f \in H^\perp, i = 1, \dots, n.$$

Hence, $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})|_H$ is a compact operator from H to H . Furthermore, we have the representation

$$\mathbf{E}^{(\alpha)}(h|\mathfrak{C})|_H = \sum_{i=1}^n \lambda_i \langle h, u_i \rangle_{\mathbf{L}_2(\mathbf{Q})} u_i \quad \text{for } h \in H \quad (4.9)$$

due to the orthonormality of u_1, \dots, u_n . Since $\mathbf{A}_2^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{Y}) \subset \ker(\mathbf{E})$ and according to Theorem 4.3, we have that

$$\mathbf{A}_2^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{Y}) \subset \ker(\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})),$$

and hence we have that the representation (4.9) holds also for the extension of $\mathbf{E}^{(\alpha)}(\cdot|\mathfrak{C})|_H$ on $H \oplus \mathbf{A}_2^0(\mathfrak{X}, \mathfrak{A}, \mathfrak{Y})$. \blacklozenge

We follow Remark 4.4 (b) to compute the regularized Rao–Blackwellization operator of Theorem 4.3 for the Gaussian model at least on a corresponding subspace.

4.5 Example (Gaussian Model, Tikhonov Regularized RB, Act I) For $\Theta = \mathbb{R}$ and some known $\sigma^2 > 0$ consider the Gaussian location model

$$\mathfrak{B} := \{\mathbf{N}(a, \sigma^2) : a \in \mathbb{R}\}.$$

For some $b \in \mathbb{R}$ and $\tau^2 > 0$, we consider the weighting measure $\pi := \mathbf{N}(b, \tau^2)$ and set $Q := \pi \cdot \mathbf{id}_{\mathfrak{B}} = \mathbf{N}(b, \sigma^2 + \tau^2)$. Recall that $S = \mathbf{id}_{\mathbb{R}}$ is sufficient and π -complete for \mathfrak{B} , see also Proposition 3.21 on 64.

We compute $\mathbf{E}^{(\alpha)}(\cdot|S)$ on $H := \text{span}\{\mathbf{1}_{\mathbb{R}}, \mathbf{id}_{\mathbb{R}}\}$ in terms of the representation (4.9) given in Remark 4.4 (b). To this end, we first show that

$$u_1 := \mathbf{1}_{\mathbb{R}} \text{ and } u_2 := \frac{1}{\sqrt{\sigma^2 + \tau^2}}(\mathbf{id}_{\mathbb{R}} - b)$$

are orthonormal in $L_2(\mathfrak{X}, \sigma(S), Q)$. Furthermore, we show that

$$\lambda_1 = 1 \text{ and } \lambda_2 = \frac{\tau^2}{\alpha\sigma^2 + \tau^2}$$

are corresponding eigenvalues, that is, we have

$$\mathbf{E}^{(\alpha)}(u_i|S) = \lambda_i u_i \quad \text{for } i = 1, 2.$$

To show orthonormality of u_1 and u_2 in $L_2(Q)$, we note first that we obviously have $\|u_1\|_{Q,2} = \|u_2\|_{Q,2} = 1$ and compute

$$\langle u_1, u_2 \rangle_{L_2(Q)} = \frac{1}{\sqrt{\sigma^2 + \tau^2}} \cdot \int (\mathbf{id}_{\mathbb{R}} - b) d\mathbf{N}(b, \sigma^2 + \tau^2) = 0,$$

that is, $u_1 \perp u_2$.

To show that λ_1, λ_2 are eigenvalues of $\mathbf{E}^{(\alpha)}(u_i|S)$ to eigenfunctions u_1, u_2 , we show that (λ_1, u_1) and (λ_2, u_2) constitute eigenpairs of $\mathbf{E}'\mathbf{E}$ and $\mathbf{V}'\mathbf{V}$. We compute for the constant function u_1

$$\begin{aligned} \langle \mathbf{E}'\mathbf{E}u_1, f \rangle_{L_2(Q)} &= \langle \mathbf{E}u_1, \mathbf{E}f \rangle_{L_2(\pi)} \\ &= \int \int u_1 f d\mathbf{N}(a, \sigma^2) \mathbf{N}(b, \tau^2)(da) \\ &= \langle u_1, f \rangle_{L_2(Q)}, \end{aligned}$$

for all $f \in L_2(Q)$, that is, $\mathbf{E}'\mathbf{E}u_1 = u_1$. Hence, $(1, u_1)$ is an eigenpair of $\mathbf{E}'\mathbf{E}$. In the following computation we consider $f \in L_\infty(Q)$ and denote by $\mathbf{E}_a f$ the expectation of f w. r. t. $\mathbf{N}(a, \sigma^2)$. For $g := (\sqrt{\sigma^2 + \tau^2})u_2$ we compute

$$\begin{aligned} \langle \mathbf{E}'\mathbf{E}g, f \rangle_{L_2(Q)} &= \langle \mathbf{E}g, \mathbf{E}f \rangle_{L_2(\pi)} \\ &= \int \mathbf{E}_a g \mathbf{E}_a f \mathbf{N}(b, \tau^2)(da) \\ &= \int g(a) \mathbf{E}_a f \mathbf{N}(b, \tau^2)(da), \end{aligned}$$

where the latter equation follows by the definition of g and the fact that $\mathbf{E}_a g = a - b = g(a)$ for all $a \in \mathbb{R}$. Denoting by φ_{a,σ^2} the Lebesgue density of $\mathbf{N}(a, \sigma^2)$, then

$$\frac{\partial}{\partial a} \varphi_{b,\tau^2}(a) = -\frac{(a-b)}{\tau^2} \varphi_{b,\tau^2}(a) = -\frac{1}{\tau^2} g(a) \varphi_{b,\tau^2}(a) \quad \text{for } a \in \mathbb{R}$$

and

$$\frac{\partial}{\partial a} \mathbf{E}_a f = \int \frac{(x-a)}{\sigma^2} f(x) \varphi_{a,\sigma^2}(x) \lambda(dx) \quad \text{for } a \in \mathbb{R}.$$

Using this and integration by parts, we obtain (recall that $g = \mathbf{id}_{\mathbb{R}} - b$ and that $f \in L_\infty(Q)$)

$$\begin{aligned} & \int g(a) \varphi_{b,\tau^2}(a) \mathbf{E}_a f \lambda(da) \\ &= \lim_{c \rightarrow -\infty} [-\tau^2 \varphi_{b,\tau^2}(a) \mathbf{E}_a f]_{-c}^c \\ & \quad + \int \int \tau^2 \varphi_{b,\tau^2}(a) \frac{(x-a)}{\sigma^2} f(x) \varphi_{a,\sigma^2}(x) \lambda(dx) \lambda(da) \\ &= \frac{\tau^2}{\sigma^2} \int \int \varphi_{b,\tau^2}(a) ((x-b) + (b-a)) f(x) \varphi_{a,\sigma^2}(x) \lambda(dx) \lambda(da) \\ &= \frac{\tau^2}{\sigma^2} \int \int \varphi_{b,\tau^2}(a) (g(x) - g(a)) f(x) \varphi_{a,\sigma^2}(x) \lambda(dx) \lambda(da) \\ &= \frac{\tau^2}{\sigma^2} \int g(x) f(x) Q(dx) - \frac{\tau^2}{\sigma^2} \int g(a) \varphi_{b,\tau^2}(a) \mathbf{E}_a f \lambda(da), \end{aligned}$$

which is equivalent to

$$\left(1 + \frac{\tau^2}{\sigma^2}\right) \int g(a) \varphi_{b,\tau^2}(a) \mathbf{E}_a f \lambda(da) = \frac{\tau^2}{\sigma^2} \langle g, f \rangle_{L_2(Q)}.$$

Recall from above that $\langle \mathbf{E}' \mathbf{E} g, f \rangle_{L_2(Q)} = \int g(a) \varphi_{b,\tau^2}(a) \mathbf{E}_a f \lambda(da)$, we then finally have

$$\langle \mathbf{E}' \mathbf{E} g, f \rangle_{L_2(Q)} = \frac{\tau^2}{\tau^2 + \sigma^2} \langle g, f \rangle_{L_2(Q)}, \quad (4.10)$$

hence $(\frac{\tau^2}{\tau^2 + \sigma^2}, g)$ is an eigenpair of $\mathbf{E}' \mathbf{E}$, since (4.10) holds for all $f \in L_\infty(Q)$.

By using (4.10) we compute for the variance operator

$$\begin{aligned} \langle \mathbf{V}' \mathbf{V} g, f \rangle_{L_2(Q)} &= \langle \mathbf{V} g, \mathbf{V} f \rangle_{L_2(\pi \otimes \mathbf{id}_{\mathbb{R}})} \\ &= \langle g, f \rangle_{L_2(Q)} - \langle \mathbf{E}' \mathbf{E} g, f \rangle_{L_2(Q)} \\ &= \langle g, f \rangle_{L_2(Q)} - \frac{\tau^2}{\tau^2 + \sigma^2} \langle g, f \rangle_{L_2(Q)} \\ &= \left(1 - \frac{\tau^2}{\tau^2 + \sigma^2}\right) \langle g, f \rangle_{L_2(Q)} \end{aligned}$$

for all $f \in L_\infty(Q)$ and hence $(\frac{\sigma^2}{\tau^2 + \sigma^2}, g)$ is a eigenpair of $\mathbf{V}'\mathbf{V}$. Note that $\langle \mathbf{V}'\mathbf{V}u_1, f \rangle_{L_2(Q)} = \langle \mathbf{V}u_1, \mathbf{V}f \rangle_{L_2(\pi \otimes \text{id}_{\mathfrak{F}})} = 0$ for all $f \in L_\infty(Q)$, since $\mathbf{V}u_1 = \mathbf{0}$. Hence, $(0, u_1)$ is also an eigenpair of $\mathbf{V}'\mathbf{V}$.

For the α -regularized Rao–Blackwellization of Theorem 4.3, we consequently get

$$\begin{aligned} \mathbf{E}^{(\alpha)}(u_1|S) &= (\alpha\mathbf{V}'\mathbf{V} + \mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{E}u_1 \\ &= u_1 \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}^{(\alpha)}(u_2|S) &= (\alpha\mathbf{V}'\mathbf{V} + \mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{E}u_2 \\ &= \frac{\tau^2}{\alpha\sigma^2 + \tau^2}u_2. \end{aligned}$$

Hence, with (λ_i, u_i) for $i = 1, 2$ we have a representation of $\mathbf{E}^{(\alpha)}(\cdot|S)$ on H as

$$\mathbf{E}^{(\alpha)}(\cdot|S)|_H = \lambda_1 \langle \cdot, u_1 \rangle_{L_2(Q)} u_1 + \lambda_2 \langle \cdot, u_2 \rangle_{L_2(Q)} u_2. \quad (4.11)$$

Now we can use (4.11) to compute the α -regularized Rao–Blackwellization of the MLE $\mathbf{id}_{\mathbb{R}} \in H$, which is

$$\begin{aligned} \mathbf{E}^{(\alpha)}(\mathbf{id}_{\mathbb{R}}|S) &= \mathbf{E}^{(\alpha)}(\mathbf{id}_{\mathbb{R}} - b|S) + \mathbf{E}^{(\alpha)}(b|S) \\ &= \frac{\tau^2}{\alpha\sigma^2 + \tau^2}(\mathbf{id}_{\mathbb{R}} - b) + b \\ &= \frac{\tau^2}{\alpha\sigma^2 + \tau^2}\mathbf{id}_{\mathbb{R}} + \frac{\alpha\sigma^2}{\alpha\sigma^2 + \tau^2}b. \end{aligned}$$

For $\alpha = 1$, this corresponds to the Bayes estimator with prior $\pi = \mathbf{N}(b, \tau^2)$ and squared error loss (see e. g. Lehmann & Casella, 1998, Example 2.2, page 233). Setting $\alpha = 0$ we obtain the MLE. On the one hand, the regularization parameter α can be interpreted as a “degree of belief” in the prior distribution π and, therefore, connects the Bayesian and the frequentist approach continuously (see also Figure 4.2). On the other hand, with regard to a misspecification of \mathfrak{F} , the parameter α can also be interpreted as a “degree of belief” in \mathfrak{F} . As the model misspecification increases, the prior gets more important. \blacklozenge

In the above example we just considered a one-dimensional sample. For the extension of the general n -dimensional i. i. d. case, we note the following.

4.6 Example (Gaussian Model, Tikhonov Regularized RB, Act II) For $\Theta = \mathbb{R}$ and some known $\sigma^2 > 0$ consider the n -dimensional i. i. d. Gaussian location model

$$\mathfrak{F} := \{\mathbf{N}^{\otimes n}(a, \sigma^2) : a \in \mathbb{R}\}.$$

For some $b \in \mathbb{R}$ and $\tau^2 > 0$, we consider the weighting measure $\pi := \mathbf{N}(b, \tau^2)$. Recall that $S(x) = \sum_{i=1}^n x_i$ for $x \in \mathbb{R}^n$ is sufficient and π -complete

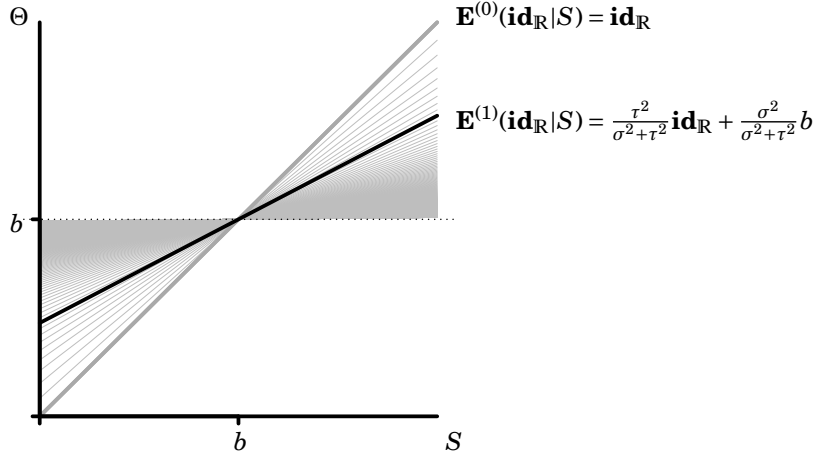


Figure 4.2: Graphs of the Tikhonov regularized Rao–Blackwellizations of the MLE $\mathbf{id}_{\mathbb{R}}$ for the Gaussian model $(\mathbf{N}(\mu, \sigma^2) : \mu \in \mathbb{R})$ with prior distribution $\pi := \mathbf{N}(b, \tau^2)$ and regularization parameter $\alpha \in \{\frac{k}{10} : k = 0, \dots, 1000\}$ (see also Example 4.5). The family includes the MLE ($\alpha = 0$) and the Bayes estimator w. r. t. the squared error loss ($\alpha = 1$).

for \mathfrak{B} , see also Proposition 3.21 on 64. To compute the regularized Rao–Blackwell projection on a subspace, we will also consider the transformed model on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ given by

$$\mathfrak{B}^S := \{\mathbf{N}^{\otimes n}(a, \sigma^2)^S : a \in \mathbb{R}\} = \{\mathbf{N}(na, n \cdot \sigma^2) : a \in \mathbb{R}\}.$$

to make use of the results of Example 4.5. Furthermore, we set $R := \pi \cdot \mathbf{id}_{\mathfrak{B}}$ for the mixture distribution on $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ and also

$$\begin{aligned} Q := \pi \cdot \mathbf{id}_{\mathfrak{B}^S} &= \int \mathbf{N}(na, n\sigma^2)(\cdot) \mathbf{N}(b, \tau^2)(da) \\ &= \int \mathbf{N}(z, n\sigma^2)(\cdot) \mathbf{N}(b, \tau^2)^{a \rightarrow n \cdot a}(dz) \\ &= \int \mathbf{N}(z, n\sigma^2)(\cdot) \mathbf{N}(nb, n^2 \tau^2)(dz) \\ &= \mathbf{N}(nb, n\sigma^2 + n^2 \tau^2)(\cdot), \end{aligned}$$

which is a distribution on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. We write \mathbf{E}^S for the expectation operator on $L_2(Q)$ w. r. t. \mathfrak{B}^S , that is,

$$\begin{aligned} \mathbf{E}^S : L_2(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Q) &\rightarrow L_2(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \pi) \\ f &\mapsto \left(\int f d\mathbf{N}^{\otimes n}(a, \sigma^2)^S \right)_{a \in \mathbb{R}}. \end{aligned}$$

Within this setting we compute $\mathbf{E}^{(\alpha)}(\cdot | S)$ on

$$H := \text{span}\{\mathbf{1}_{\mathbb{R}^n}, S\} \subset L_2(\mathbb{R}^n, \sigma(S), R)$$

in terms of a SVD of $\mathbf{E}^{(\alpha)}(\cdot|S)|_H$, that is, we show

$$u_1 := \mathbf{1}_{\mathbb{R}^n} \text{ and } u_2 := \frac{1}{\sqrt{n\sigma^2 + n^2\tau^2}}(S - nb\mathbf{1}_{\mathbb{R}^n}) \quad (4.12)$$

are orthonormal eigenfunctions of $\mathbf{E}^{(\alpha)}(\cdot|S)$ in $L_2(\mathfrak{X}, \sigma(S), R)$ to the eigenvalues

$$\lambda_1 = 1 \text{ and } \lambda_2 = \frac{n^2\tau^2}{n\sigma^2 + n^2\tau^2}, \quad (4.13)$$

that is, we have

$$\mathbf{E}^{(\alpha)}(u_i|S) = \lambda_i u_i \quad \text{for } i = 1, 2. \quad (4.14)$$

Orthonormality of u_1 and u_2 in $L_2(\mathbb{R}^n, \sigma(S), R)$ and furthermore that $\mathbf{E}^{(\alpha)}(u_1|S) = \lambda_1 u_1$ is easy to show. Therefore, we just show (4.14). We compute for $g := (\sqrt{n\sigma^2 + n^2\tau^2})(\mathbf{id}_{\mathbb{R}} - nb)$ and $f \in L_\infty(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{Q})$

$$\begin{aligned} & \langle \mathbf{E}' \mathbf{E} g \circ S, f \circ S \rangle_{L_2(R)} \\ &= \langle \mathbf{E} g \circ S, \mathbf{E} f \circ S \rangle_{L_2(\pi)} \\ &= \langle \mathbf{E}^S g, \mathbf{E}^S f \rangle_{L_2(\pi)} \\ &= \int \left(\int g \, d\mathbf{N}(na, n\sigma^2) \right) \left(\int f \, d\mathbf{N}(na, n\sigma^2) \right) \mathbf{N}(b, \tau^2)(da) \\ &= \int \left(\int g \, d\mathbf{N}(a, n\sigma^2) \right) \left(\int f \, d\mathbf{N}(a, n\sigma^2) \right) \mathbf{N}(b, n^2\tau^2)(da) \\ &= \frac{n^2\tau^2}{n\sigma^2 + n^2\tau^2} \langle g, f \rangle_{L_2(Q)}, \end{aligned}$$

where the latter equation is according to Example 4.5 with $n\sigma^2$ instead of σ^2 and $n^2\tau^2$ instead of τ^2 . Furthermore, we have

$$\begin{aligned} \langle g, f \rangle_{L_2(Q)} &= \int \int g(z) f(z) \mathbf{N}(na, n\sigma^2)(dz) \mathbf{N}(b, \tau^2)(da) \\ &= \int \int g(S(x)) f(S(x)) \mathbf{N}^{\otimes n}(a, \sigma^2)(dx) \mathbf{N}(b, \tau^2)(da) \\ &= \langle g \circ S, f \circ S \rangle_{L_2(R)} \end{aligned}$$

and consequently

$$\langle \mathbf{E}' \mathbf{E} g \circ S, f \circ S \rangle_{L_2(R)} = \frac{n^2\tau^2}{n\sigma^2 + n^2\tau^2} \langle g \circ S, f \circ S \rangle_{L_2(R)}.$$

With g and f as above and with the same arguments, we compute for the

variance operator

$$\begin{aligned}
& \langle \mathbf{V}^{\mathbf{V}} \mathbf{V} g \circ S, f \circ S \rangle_{L_2(R)} \\
&= \langle \mathbf{V} g \circ S, \mathbf{V} f \circ S \rangle_{L_2(\pi \otimes \mathbf{id}_{\mathfrak{F}})} \\
&= \int \int (g(S(x)) - \mathbf{E}_a g(S(\cdot))) (g(S(x)) - \mathbf{E}_a g(S(\cdot))) \\
&\quad \mathbf{N}^{\otimes n}(a, \sigma^2)(dx) \mathbf{N}(b, \tau^2)(da) \\
&= \int \int (g(z) - \mathbf{E}_a^S g)(g(z) - \mathbf{E}_a^S g) \mathbf{N}(na, n\sigma^2)(dz) \mathbf{N}(b, \tau^2)(da) \\
&= \left(1 - \frac{n^2 \tau^2}{n\sigma^2 + n^2 \tau^2}\right) \langle g, f \rangle_{L_2(Q)},
\end{aligned}$$

where the latter equation follows according to Example 4.5 with $n\sigma^2$ instead of σ^2 , with $n^2\tau^2$ instead of τ^2 , and due to the fact that

$$\mathbf{E}_a^S g = \int g(y) \mathbf{N}(na, n\sigma^2)(dy) \quad \text{for } a \in \mathbb{R}.$$

Since $\langle g, f \rangle_{L_2(Q)} = \langle g \circ S, f \circ S \rangle_{L_2(R)}$, we consequently have

$$\langle \mathbf{V}^{\mathbf{V}} \mathbf{V} g \circ S, f \circ S \rangle_{L_2(R)} = \left(1 - \frac{n^2 \tau^2}{n\sigma^2 + n^2 \tau^2}\right) \langle g \circ S, f \circ S \rangle_{L_2(R)}.$$

With these computations and with (λ_1, u_1) and (λ_2, u_2) given in (4.12) and (4.13) one can easily show that

$$\mathbf{E}^{(\alpha)}(\cdot | S)|_H = \lambda_1 \langle \cdot, u_1 \rangle_{L_2(R)} u_1 + \lambda_2 \langle \cdot, u_2 \rangle_{L_2(R)} u_2. \quad (4.15)$$

Furthermore, since $\mathbf{E}^{(\alpha)}(h | S) = \mathbf{0}$ for all $h \in A_2^0(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathfrak{F})$ we may consider (4.15) on $G := H \oplus A_2^0(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathfrak{F})$, that is,

$$\mathbf{E}^{(\alpha)}(\cdot | S)|_G = \lambda_1 \langle \cdot, u_1 \rangle_{L_2(R)} u_1 + \lambda_2 \langle \cdot, u_2 \rangle_{L_2(R)} u_2.$$

Indeed, $A_2^0 \neq \{\mathbf{0}\}$ if $n \geq 2$. Consider for example $\frac{1}{n}S - \text{pr}_1 \in A_2^0(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \mathfrak{F})$, where obviously $\frac{1}{n}S - \text{pr}_1 \neq \mathbf{0}$.

Now we can use (4.15) to compute the α -regularized Rao–Blackwellization of estimators in G . Consider, for instance, the sample median $\text{Med}(X) \in G$, where $X := \mathbf{id}_{\mathbb{R}^n}$ and denote by $\bar{X} := \frac{1}{n}S(X)$ the MLE for \mathbf{id}_{Θ} . Then, using $\text{Med}(X) - \bar{X} \in A_1^0$ which yields that $\mathbf{E}^{(\alpha)}(\text{Med}(X) - \bar{X} | S) = \mathbf{0}$,

we obtain

$$\begin{aligned}
\mathbf{E}^{(\alpha)}(\text{Med}(X)|S) &= \mathbf{E}^{(\alpha)}(\bar{X}|S) + \mathbf{E}^{(\alpha)}(\text{Med}(X) - \bar{X}|S) \\
&= \mathbf{E}^{(\alpha)}(b\mathbf{1}_{\mathbb{R}^n}|S) + \mathbf{E}^{(\alpha)}(\bar{X} - b\mathbf{1}_{\mathbb{R}^n}|S) \\
&= \mathbf{E}^{(\alpha)}(b\mathbf{1}_{\mathbb{R}^n}|S) + \frac{1}{n}\mathbf{E}^{(\alpha)}(S(X) - nb\mathbf{1}_{\mathbb{R}^n}|S) \\
&= \lambda_1 \langle b\mathbf{1}_{\mathbb{R}^n}, u_1 \rangle_{L_2(R)} u_1 + \frac{1}{n} \lambda_2 \langle S(X) - nb\mathbf{1}_{\mathbb{R}^n}, u_2 \rangle_{L_2(R)} u_2 \\
&= \frac{1}{n} \lambda_1 \langle b\mathbf{1}_{\mathbb{R}^n}, u_1 \rangle_{L_2(R)} u_1 + \frac{1}{n} \lambda_2 (\sqrt{n\sigma^2 + n^2\tau^2}) u_2 \\
&= b + \frac{1}{n} \lambda_2 (S(X) - nb\mathbf{1}_{\mathbb{R}^n}) \\
&= \frac{n^2\tau^2}{\alpha n\sigma^2 + n^2\tau^2} \bar{X} + \frac{\alpha n\sigma^2}{\alpha n\sigma^2 + n^2\tau^2} b\mathbf{1}_{\mathbb{R}^n}.
\end{aligned}$$

Of course, we also have that $\mathbf{E}^{(\alpha)}(\bar{X}|S) = \mathbf{E}^{(\alpha)}(\text{pr}_k(X)|S) = \mathbf{E}^{(\alpha)}(\text{Med}(X)|S)$, where $k \in \{1, \dots, n\}$. \blacklozenge

Compared to the classical Rao–Blackwellization, the concept of the regularized Rao–Blackwellization presented in this thesis is, generally, a concept that goes beyond mere data reduction. That is, in addition to a reduction to the complete sufficient information, prior information (π and α) is used to account for model misspecification. The latter is quantified through the regularization parameter $\alpha > 0$. In the Gaussian case (see Examples 4.5 and 4.6) one can see, for example, that in estimating the canonical parameter \mathbf{id}_{Θ} the prior distribution π gets more involved as the misspecification increases. In particular, $\alpha = 0$ (no misspecification) and $\alpha = 1$ yield to the maximum likelihood and the Bayes estimate (w. r. t. the squared error loss), respectively.

The computation of the Tikhonov regularized Rao–Blackwellization for an arbitrary regularization parameter α can be quite laborious, see for instance the previous Examples 4.5 and 4.6. For $\alpha = 1$ the computation of $\mathbf{E}^{(1)}(\cdot|S)$ is directly accessible in a broad range of situations, in a way closely related to Bayesian statistics.

4.7 Theorem Let $(\mathfrak{F}, \mathfrak{B}, \pi)$ be a statistical model, let S be a sufficient and π -complete statistic for \mathfrak{F} and define $\Pi_{\pi}(S, \cdot) := (\pi \otimes \mathbf{id}_{\mathfrak{F}})(\cdot|\sigma(S) \otimes \{\emptyset, \mathfrak{F}\})$. Then, the Tikhonov regularized Rao–Blackwellization $\mathbf{E}^{(1)}(\cdot|S)$ is given by

$$\mathbf{E}^{(1)}(f|S) = \int \mathbf{E}_P f \Pi_{\pi}(S, dP) \quad \text{for } f \in L_2(Q). \quad (4.16)$$

In particular, $\mathbf{E}^{(1)}(f|S)$ equals the Bayes estimator (w. r. t. the squared error loss) if f is unbiased. \blacklozenge

Proof: For $f \in L_2(Q)$ we have that the map $(x, P) \mapsto (\mathbf{E}_P f) \mathbb{1}_{\mathfrak{X}}(x)$ on $\mathfrak{X} \times \mathfrak{P}$ is in $L_2(\mathfrak{X} \times \mathfrak{P}, \mathcal{A} \otimes \mathcal{B}, \boldsymbol{\pi} \otimes \mathbf{id}_{\mathfrak{P}})$. Furthermore, a simple calculation yields

$$\|g - f\|_{\mathbf{L}\mathbf{S},2}^2 + \|g\|_{\mathbf{O}\mathbf{S},2}^2 = \int \int (g(x) - \mathbf{E}_P(f))^2 P(dx) \boldsymbol{\pi}(dP). \quad (4.17)$$

Note that $\mathbf{E}^{(1)}(f|S)$ is defined to be the minimizer of (4.17) among all $\sigma(S)$ -measurable $g \in L_2(Q)$. Since $L_2(\mathfrak{X}, \sigma(S), Q)$ can be identified with $L_2(\mathfrak{X} \times \mathfrak{P}, \sigma(S) \otimes \{\emptyset, \mathfrak{P}\}, \boldsymbol{\pi} \otimes \mathbf{id}_{\mathfrak{P}})$, the regularization $\mathbf{E}^{(1)}(f|S)$ corresponds to a solution of

$$\operatorname{argmin}\{\|h - \mathbf{E}(f) \mathbb{1}_{\mathfrak{X}}\|_{\boldsymbol{\pi} \otimes \mathbf{id}_{\mathfrak{P}},2}^2 : h \in L_2(\mathfrak{X} \times \mathfrak{P}, \sigma(S) \otimes \{\emptyset, \mathfrak{P}\}, \boldsymbol{\pi} \otimes \mathbf{id}_{\mathfrak{P}})\}, \quad (4.18)$$

which is given by the conditional expectation of $(x, P) \mapsto \mathbf{E}_P f \mathbb{1}_{\mathfrak{X}}(x)$ given $\sigma(S) \otimes \{\emptyset, \mathfrak{P}\}$, that is,

$$\mathbf{E}^{(1)}(f|S) = \mathbf{E}(\mathbf{E}(f \mathbb{1}_{\mathfrak{X}} | \sigma(S) \otimes \{\emptyset, \mathfrak{X}\})) = \int \int f(x) P(dx) \Pi_{\boldsymbol{\pi}}(S, dP),$$

which yields the claim. ■

In many cases the kernel $\Pi_{\boldsymbol{\pi}}(S, \cdot)$ in Theorem (4.7) can be computed explicitly, especially if $\boldsymbol{\pi}$ is chosen as a so-called conjugate prior.

4.8 Example (Exponential Distribution Model) For some $n \in \mathbb{N}$ consider the exponential model $\mathfrak{P} = \{\mathbf{Exp}(\vartheta)^{\otimes n} : \vartheta \in]0, \infty[\}$, that is, $\mathbf{Exp}(\vartheta)$ has the $\boldsymbol{\lambda}^{\otimes n}$ -density

$$\vartheta^n \exp\left(-\vartheta \sum_{i=1}^n x_i\right) \mathbb{1}_{[0, \infty[}(x) \quad \text{for } x \in \mathbb{R}^n.$$

Let $\boldsymbol{\pi} := \mathbf{G}(a, b)$ be the *Gamma distribution* with $\boldsymbol{\lambda}$ -density

$$f_{a,b}(x) := \frac{1}{\Gamma(a)} b^a x^{a-1} \exp(-bx) \mathbb{1}_{[0, \infty[}(x) \quad \text{for } x \in \mathbb{R}^n.$$

Recall that $S(x) := \sum_{i=1}^n x_i$ for $x \in \mathbb{R}^n$ is sufficient and $\boldsymbol{\pi}$ -complete for \mathfrak{P} , see also Proposition 3.21 on 64. Furthermore, the posterior distribution is given by

$$\Pi_{\boldsymbol{\pi}}(S, \cdot) = \mathbf{G}(a + n, b + S).$$

By Theorem 4.7 we have that the 1-Tikhonov regularized Rao–Blackwellization of f given S is

$$\mathbf{E}^{(1)}(f|S) = \int \mathbf{E}_{\vartheta} f \mathbf{G}(a + n, b + S)(d\vartheta).$$

Canonical Parameter. For $n \geq 2$ consider the MLE for the canonical parameter $\mathbf{id}_{]0, \infty[}$, that is, $\frac{n}{S}$. Since $\mathbf{Exp}(\vartheta)^{\otimes n} \circ S^{-1} = \mathbf{G}(n, \frac{1}{\vartheta})$, we obtain for

the corresponding expectation functional

$$\begin{aligned}\mathbf{E}_\vartheta\left(\frac{n}{S}\right) &= n \int_0^\infty \frac{1}{y} y^{n-1} \frac{1}{\Gamma(n)} \vartheta^n \exp(-\vartheta y) \lambda(dy) \\ &= n \int_0^\infty y^{n-2} \frac{1}{(n-1)\Gamma(n-1)} \vartheta^n \exp(-\vartheta y) \lambda(dy) \\ &= \frac{n}{n-1} \vartheta.\end{aligned}$$

Hence, we have that $\frac{n-1}{S}$ is unbiased⁸ for the canonical parameter, which yields

$$\mathbf{E}^{(1)}\left(\frac{n-1}{S} \middle| S\right) = \int \vartheta \mathbf{G}(a+n, b+S)(d\vartheta) = \frac{a+n}{b+S}.$$

Furthermore,

$$\mathbf{E}^{(1)}\left(\frac{n}{S} \middle| S\right) = \frac{n}{n-1} \frac{a+n}{b+S}$$

is the 1-Tikhonov regularized Rao–Blackwellization of the MLE $\frac{n}{S}$.

Inverse Canonical Parameter. The MLE and the UMVUE for the parameter $\vartheta \mapsto \frac{1}{\vartheta}$ for $\vartheta \in]0, \infty[$ coincide and is $\frac{1}{n}S$. This is due to the invariance property of the MLE (see also Remark 2.7), the unbiasedness can be shown easily. The 1-Tikhonov regularized Rao–Blackwellization of $\frac{1}{n}S$ is due to the unbiasedness just the Bayes estimator

$$\mathbf{E}^{(1)}\left(\frac{1}{n}S \middle| S\right) = \frac{1}{n+a-1} \cdot S + \frac{b}{a+n-1}.$$

Square Root of the Canonical Parameter. The MLE for $\vartheta \mapsto \sqrt{\vartheta}$ for $\vartheta \in]0, \infty[$ is $\sqrt{\frac{n}{S}}$ while the UMVUE is $\sqrt{\frac{n}{\pi S}}$. Hence, we have

$$\begin{aligned}\mathbf{E}^{(1)}\left(\sqrt{\frac{n}{\pi S}} \middle| S\right) &= \int \mathbf{E}_\vartheta\left(\sqrt{\frac{n}{\pi S}}\right) \mathbf{G}(a+n, b+S)(d\vartheta) \\ &= \int \sqrt{\vartheta} \mathbf{G}(a+n, b+S)(d\vartheta) \\ &= \left(\frac{1}{2}\right)^{2(a+n)+1} \frac{2 \cdot (a+n)!}{\Gamma(a+n)} \sqrt{\frac{\pi}{b+S}}\end{aligned}$$

for the UMVUE. Correspondingly, we obtain

$$\mathbf{E}^{(1)}\left(\sqrt{\frac{n}{S}} \middle| S\right) = \sqrt{\pi} \cdot \mathbf{E}^{(1)}\left(\sqrt{\frac{n}{\pi S}} \middle| S\right) = \left(\frac{1}{2}\right)^{2(a+n)+1} \frac{(2 \cdot (a+n)!}{\Gamma(a+n)} \pi \cdot \sqrt{\frac{1}{b+S}}$$

for the MLE. ◆

⁸Note that $\frac{n-1}{S}$ is even the uniformly minimum variance unbiased estimator (UMVUE) estimator.

4.9 Example (Continuous Uniform Model) For some $n \in \mathbb{N}$ consider the continuous uniform model $\mathfrak{F} = \{\mathbf{U}_{[0,\vartheta]}^{\otimes n} : \vartheta \in [0, 1]\}$ and let $\boldsymbol{\pi} := \mathbf{U}_{[0,1]}$. Note that $S(x) := \max(x)$ for $x \in [0, 1]^n$ is sufficient and $\boldsymbol{\pi}$ -complete for \mathfrak{F} , see Example 3.20 on page 64. The posterior distribution $\Pi_{\boldsymbol{\pi}}(S, \cdot)$ can be computed and has the $\boldsymbol{\lambda}$ -density

$$\frac{S^{n-1}(n-1)}{1-S^{n-1}} \frac{1}{\vartheta^n} \mathbb{1}_{]S,1]}(\vartheta) \quad \text{for } \vartheta \in \mathbb{R}$$

and, hence, the 1-Tikhonov regularized Rao–Blackwellization of f given S is

$$\mathbf{E}^{(1)}(f|S) = \int \mathbf{E}_{\vartheta} f \Pi_{\boldsymbol{\pi}}(S, d\vartheta).$$

Due to the biasedness of the MLE S , the 1-Tikhonov regularized Rao–Blackwellization of S is

$$\begin{aligned} \mathbf{E}^{(1)}(S|S) &= \int \mathbf{E}_{\vartheta} S \Pi_{\boldsymbol{\pi}}(S, d\vartheta) \\ &= \frac{n}{n+1} \int \mathbf{E}_{\vartheta} \frac{n+1}{n} S \Pi_{\boldsymbol{\pi}}(S, d\vartheta) \\ &= \frac{n}{n+1} \int \vartheta \Pi_{\boldsymbol{\pi}}(S, d\vartheta) \\ &= \frac{n}{n+1} \int \frac{S^{n-1}(n-1)}{1-S^{n-1}} \frac{1}{\vartheta^{n-1}} \mathbb{1}_{]S,1]}(\vartheta) \boldsymbol{\lambda}(d\vartheta) \\ &= \frac{n}{n+1} \frac{(n-1)}{(n-2)} S \cdot \left(\frac{1-S^{n-2}}{1-S^{n-1}} \right), \end{aligned}$$

which is *not* the Bayes estimator. However, we obtain

$$\mathbf{E}^{(1)}\left(\frac{n+1}{n} S \middle| S\right) = \frac{(n-1)}{(n-2)} S \cdot \left(\frac{1-S^{n-2}}{1-S^{n-1}} \right)$$

for the regularized RB of the UMVUE $\frac{n+1}{n} S$ and any other unbiased estimator, such as $x \mapsto \frac{2}{n} \sum_{i=1}^n x_i$ (see also Figure 4.4)

Note that the 1-Tikhonov regularized Rao–Blackwellizations are not affine functions of the complete sufficient statistic S and, hence, S is not an eigenfunction of $\mathbf{E}^{(1)}(\cdot|S) \mathbf{E}^{(1)}(\cdot|S)$. An illustration of the graphs of the MLE, the UMVUE, and their regularized Rao–Blackwellizations are given in Figure 4.3. \blacklozenge

Due to the continuity of the regularized Rao–Blackwellization, we may expect that $\mathbf{E}^{(1)}(\hat{\lambda}|S)$ for an “approximately” unbiased estimator $\hat{\lambda}$ is “close” to the Bayes estimator $\hat{\vartheta}_{\text{Bayes}}$ w. r. t. the squared error loss. That is made precise in the following theorem.

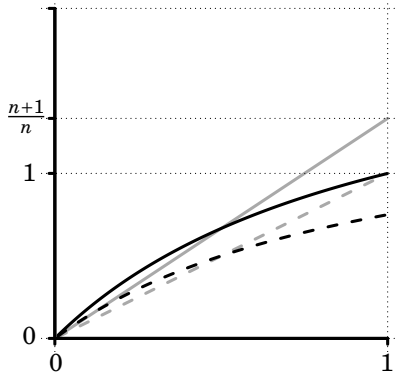


Figure 4.3: Graphs of the MLE (gray, dashed), the UMVUE (gray, solid), and the corresponding 1-Tikhonov regularized Rao–Blackwellizations (black, dashed) and (black, solid), respectively.

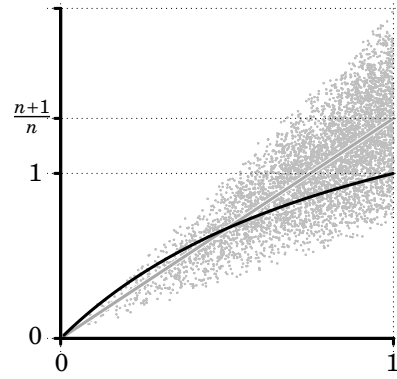


Figure 4.4: Graphs of the 1-Tikhonov regularized Rao–Blackwellizations (black, solid) of *any* unbiased estimator for \mathbf{id}_Θ . The behavior of $2\bar{X}$ is demonstrated by the gray point cloud that represents 5000 realizations of $\pi \otimes \mathbf{id}_\mathfrak{F}(S, 2\bar{X})^{-1}$. The graph of the UMVUE is shown by the solid gray line.

4.10 Theorem For a statistical model $\mathfrak{P} := \{P_\vartheta : \vartheta \in \Theta\}$ let S be a sufficient and π -complete statistic. Furthermore, let $\hat{\vartheta}_{\text{Bayes}}$ be the Bayes estimator w. r. t. the squared error loss and let $\hat{\lambda} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$. Then, we have

$$\|\mathbf{E}^{(1)}(\hat{\lambda}|S) - \hat{\vartheta}_{\text{Bayes}}\|_{Q,2} \leq \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \pi(d\vartheta)}. \quad (4.19)$$

◆

Proof: Let ϕ be the map from (4.7) (see also page 71), then

$$\begin{aligned} R_1 : \quad (L_2(\Theta, \mathfrak{B}, \pi), \|\cdot\|_{\pi,2}) &\rightarrow (L_2(\mathfrak{X}, \sigma(S), Q), \|\cdot\|_{Q,2}) \\ m &\rightarrow \arg \inf \{\phi_m(g) : g \in L_2(\mathfrak{X}, \sigma(S), Q)\} \end{aligned}$$

is well-defined, continuous, and has an explicit form given by (see also proof of Theorem 4.3)

$$R_1 m = (\mathbf{E}'\mathbf{E} + \mathbf{V}'\mathbf{V})^{-1}\mathbf{E}'m \quad \text{for } m \in L_2(\pi).$$

For R_1 we show $\|R_1\| \leq 1$. To this end, we use Jensen's inequality to obtain

$$\|\mathbf{E}g\|_{\pi,2}^2 = \int \|\mathbf{E}_\vartheta g\|^2 \pi(d\vartheta) \leq \int \int \|g(x)\|^2 P_\vartheta(dx) \pi(d\vartheta) = \|g\|_{Q,2}^2$$

and thus we may follow that

$$\|\mathbf{E}\| = \sup \{\|\mathbf{E}g\|_{\pi,2} : g \in L_2(Q), \|g\|_{Q,2} \leq 1\} \leq 1$$

and consequently we have $\|\mathbf{E}'\| = \|\mathbf{E}\| \leq 1$ as well (see e. g. Werner, 2018, Satz III.4.2 (a), page 123). We also recall from the proof of Theorem 4.3 on page 72 that

$$\alpha \|f\|_{Q,2}^2 \leq \langle (\mathbf{E}'\mathbf{E} + \alpha\mathbf{V}'\mathbf{V})f, f \rangle \quad \text{for all } f \in L_2(Q).$$

Hence, the theorem of Lax–Milgram (see e. g. Werner, 2018, page 266 f.) yields that

$$\|(\mathbf{E}'\mathbf{E} + \alpha\mathbf{V}'\mathbf{V})^{-1}\| \leq \frac{1}{\alpha}.$$

Consequently, with $\alpha = 1$ we obtain

$$\|R_1\| \leq \|(\mathbf{E}'\mathbf{E} + \alpha\mathbf{V}'\mathbf{V})^{-1}\| \|\mathbf{E}'\| \leq 1.$$

Finally, to show the stated inequality, we note for $\mathbf{id}_\Theta \in L_2(\Theta, \mathcal{B}, \boldsymbol{\pi})$ that $R_1(\mathbf{id}_\Theta) = \hat{\vartheta}_{\text{Bayes}}$. Hence, we have

$$\begin{aligned} \|\mathbf{E}^{(1)}(\hat{\lambda}|S) - \hat{\vartheta}_{\text{Bayes}}\|_{Q,2} &= \|R_1(\mathbf{E}\hat{\lambda} - \mathbf{id}_\Theta)\|_{Q,2} \\ &\leq \|R_1\| \|(\mathbf{E}\hat{\lambda} - \mathbf{id}_\Theta)\|_{Q,2} \\ &\leq \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \boldsymbol{\pi}(d\vartheta)}, \end{aligned}$$

where we use $\|R_1\| \leq 1$ in the latter step. ■

4.11 Example (Continuous Uniform Model) In continuation of Example 4.9 on page 83 we recall that $\mathbf{E}^{(1)}(\frac{n+1}{n}S|S) = \hat{\vartheta}_{\text{Bayes}}$ which holds due to the unbiasedness of $\frac{n+1}{n}S$. For the biased MLE $\hat{\vartheta}_{\text{ML}}(x) = S(x)$ for $x \in \mathbb{R}_+^n$, we first note that

$$\text{Bias}(\hat{\vartheta}_{\text{ML}}, \vartheta) = -\frac{\vartheta}{n+1}$$

and, hence, by Theorem 4.19

$$\|\mathbf{E}^{(1)}(\hat{\vartheta}_{\text{ML}}|S) - \hat{\vartheta}_{\text{Bayes}}\|_{Q,2} \leq \frac{1}{n+1} \sqrt{\int_0^1 \vartheta^2 \boldsymbol{\lambda}(d\vartheta)} = \frac{1}{\sqrt{3} \cdot (n+1)}.$$

Furthermore, using $\mathbf{E}^{(1)}(\hat{\vartheta}_{\text{ML}}|S) = \frac{n}{n+1}\hat{\vartheta}_{\text{Bayes}}$ from Example 4.9, an exact computation of the left-hand side yields

$$\|\mathbf{E}^{(1)}(\hat{\vartheta}_{\text{ML}}|S) - \hat{\vartheta}_{\text{Bayes}}\|_{Q,2} = \frac{1}{\sqrt{3} \cdot (n+1)} \cdot \sqrt{\frac{2n(n-1)}{(n+1)(2n-1)}}.$$

◆

Finally, we propose a *Tikhonov regularized* version of the Rao–Blackwell–Lehmann–Scheffé (RBLs) theorem (see also Theorem 2.17, page 38, and Theorem 2.18, page 39).

Recall that the classical Rao–Blackwellization aims at reducing the ancillary information and therefore is a concept for improving estimation procedures. In addition, the presented regularization approach looks for approximate *regular* substitutes of the Rao–Blackwell projection. Consequently, we may expect a regularized version of the Rao–Blackwell–Lehmann–Scheffé Theorem as follows.

4.12 Theorem (Tikhonov Regularized RBLS) For a statistical model $\mathfrak{P} := \{P_\vartheta : \vartheta \in \Theta\}$ let S be a sufficient and π -complete statistic for \mathfrak{P} . Furthermore, let $\hat{\lambda} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$. Then, for every $\alpha \in]0, 1]$ and $\kappa = \mathbf{id}_\Theta$ we have

$$\begin{aligned} \mathbf{r}(\mathbf{E}^{(\alpha)}(\hat{\lambda}|S), \pi) - 2 \cdot \|\mathbf{E}^{(\alpha)}(\hat{\lambda}|S) - \hat{\lambda}\|_{\mathbf{LS},2} \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \pi(d\vartheta)} \\ \leq \mathbf{r}(\mathbf{E}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\hat{\lambda}, \pi). \end{aligned}$$

and in particular

$$\mathbf{r}(\mathbf{E}^{(\alpha)}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\mathbf{E}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\hat{\lambda}, \pi),$$

if $\hat{\lambda}$ is unbiased. ◆

Proof: The part $\mathbf{r}(\mathbf{E}(\hat{\lambda}|S), \pi) \leq \mathbf{r}(\hat{\lambda}, \pi)$ follows by the classical theorem of Rao and Blackwell (see Theorem 2.17). To show the remaining inequality for the Bayes risk, we first note that for $\alpha \in]0, 1]$ by definition of $\hat{\lambda}^* := \mathbf{E}^{(\alpha)}(\hat{\lambda}|S)$ and since $\mathbf{E}(\mathbf{E}(\hat{\lambda}|S)) = \mathbf{E}\hat{\lambda}$

$$\|\hat{\lambda}^*\|_{\mathbf{OS},2}^2 + \|\hat{\lambda}^* - \hat{\lambda}\|_{\mathbf{LS},2}^2 \leq \|\hat{\lambda}^*\|_{\mathbf{OS},2}^2 + \frac{1}{\alpha} \|\hat{\lambda}^* - \hat{\lambda}\|_{\mathbf{LS},2}^2 \leq \|\mathbf{E}(\hat{\lambda}|S)\|_{\mathbf{OS},2}^2.$$

Hence, we may compute for the corresponding Bayes risk of $\hat{\lambda}^*$

$$\begin{aligned} \mathbf{r}(\hat{\lambda}^*, \pi) &= \int \text{Var}_\vartheta(\hat{\lambda}^*) \pi(d\vartheta) + \int \text{Bias}(\hat{\lambda}^*, \vartheta)^2 \pi(d\vartheta) \\ &\leq \|\hat{\lambda}^*\|_{\mathbf{OS},2}^2 + \|\hat{\lambda}^* - \hat{\lambda}\|_{\mathbf{LS},2}^2 + 2\|\hat{\lambda}^* - \hat{\lambda}\|_{\mathbf{LS},2} \cdot \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \pi(d\vartheta)} \\ &\leq \mathbf{r}(\mathbf{E}(\hat{\lambda}|S), \pi) + 2\|\hat{\lambda}^* - \hat{\lambda}\|_{\mathbf{LS},2} \cdot \sqrt{\int \text{Bias}(\hat{\lambda}, \vartheta)^2 \pi(d\vartheta)}, \end{aligned}$$

which is equivalent to the stated inequality. ■

4.13 Example (Continuous Uniform Model) To measure the performance of the estimators given in Example 4.9 on page 83 f., we consider realizations $(X_1, \theta_1), \dots, (X_N, \theta_N)$ simulated from

$$\pi \otimes \mathbf{id}_\mathfrak{P} = \mathbf{U}_{[0,1]} \otimes \mathbf{U}_{[0,1]}^n,$$

where $n = 10$ and $N = 10^6$. The empirical distributions of

$$(\hat{\kappa}(X_i) - \kappa(\theta_i)) : i = 1, \dots, N$$

for estimators

$$\widehat{\kappa} \in \left\{ S, \mathbf{E}^{(1)}(S|S), 2\overline{X}, \frac{n+1}{n}S, \mathbf{E}^{(1)}\left(\frac{n+1}{n}S|S\right) \right\}$$

are shown in Figure 4.5. In accordance with Theorem 4.12, one can see that the 1-Tikhonov regularized Rao–Blackwellization $\mathbf{E}^{(1)}(\cdot|S)$ improves the unbiased estimators $2\overline{X}$ and $\frac{n+1}{n}S$ w. r. t. the Bayes risk

$$\mathbf{r}(\widehat{\lambda}, \mathbf{U}_{[0,1]}) = \int \|\widehat{\lambda}(x) - \vartheta\|^2 \mathbf{U}_{[0,\vartheta]}(dx) \mathbf{U}_{[0,1]}(d\vartheta).$$

Furthermore, note that $\mathbf{E}^{(1)}(S|S)$ is worse than S , which is due to the bias of S . However, due to the continuity of $\mathbf{E}^{(1)}(\cdot|S)$, improvements can be expected at least in an $\|\cdot\|_{\mathbf{LS},2}$ -neighborhood of unbiased estimators. \blacklozenge

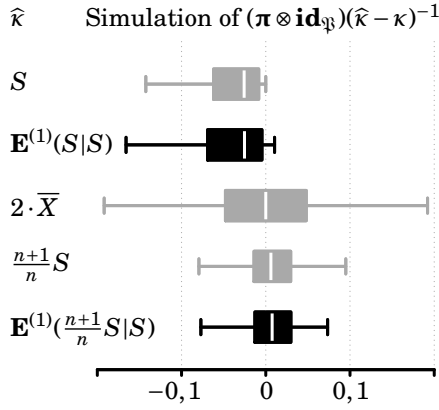


Figure 4.5: Boxplots of 10^6 realizations of $\pi \otimes \mathbf{id}_{\mathfrak{P}}$ under $\widehat{\kappa} - \widehat{\kappa}(\cdot)$. While gray colored boxplots correspond to conventional estimators (MLE, $2 \cdot \overline{X}$, and UMVUE) for $\kappa = \mathbf{id}_{\Theta}$, the black colored boxplots correspond to their 1-Tikhonov regularized Rao–Blackwellizations. The boxplots illustrate that $\mathbf{E}^{(1)}(\cdot|S)$ improves the unbiased estimators $2 \cdot \overline{X}$ and $\frac{n+1}{n}S$, however, $\mathbf{E}^{(1)}(S|S)$ is worse than the MLE S .

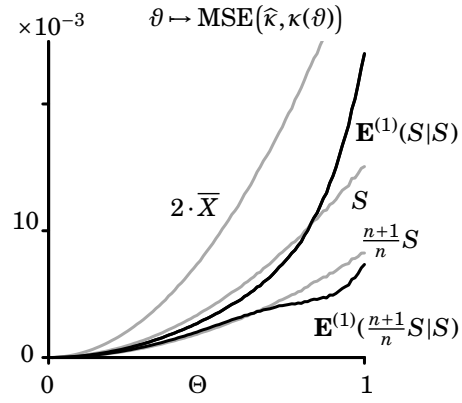


Figure 4.6: Graphs of $\vartheta \mapsto \text{MSE}(\widehat{\kappa}, \kappa(\vartheta))$. While the gray colored graphs correspond to the conventional estimators (MLE, $2 \cdot \overline{X}$, and UMVUE) for $\kappa = \mathbf{id}_{\Theta}$, the black colored graphs correspond to the 1-Tikhonov regularized Rao–Blackwellizations.

4.2 Ivanov's Regularization Approach

To avoid an overfitting in (see also (4.3))

$$\mathbf{E}(f|\mathcal{C}) = \mathbf{E}_{\mathcal{C}}^{-1} \mathbf{E}f = \text{argmin} \{ \|f - g\|_{\mathbf{LS},p} : g \in L_p(\mathfrak{X}, \mathcal{C}, Q) \}.$$

for some $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$ and hence to regularize $\mathbf{E}_{\mathcal{C}}$, we may also take qualitative prior knowledge of the solutions into account. We do this by

narrowing the class of possible solutions to a subset $G \subset L_p(\mathfrak{X}, \mathcal{C}, Q)$ such that for every $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$

$$\operatorname{argmin}\{\|f - g\|_{\mathbf{LS}, p} : g \in G\} \quad (4.20)$$

is well-posed, that is, (4.20) has a *unique* solution that depends *continuously* on f and which is then also called *quasi solution*. Consequently, we write $\mathbf{E}^{(G)}(f|\mathcal{C})$ for the solution of (4.20), which we call *G-quasi Rao–Blackwellization* of f given \mathcal{C} .

Furthermore, $\mathbf{E}^{(G)}(\cdot|\mathcal{C})$ is by definition also a regularized Rao–Blackwellization. To see this, we define the map

$$B_G : L_p(\mathfrak{F}, \mathfrak{B}, \pi) \rightarrow G \\ m \mapsto \operatorname{argmin}\{\|m - \mathbf{E}g\|_{\pi, p} : g \in G\}.$$

Then B_G constitutes, as a constant family, a regularization scheme for $\mathbf{E}_{\mathcal{C}}$ on G and satisfies $\mathbf{E}^{(G)}(\cdot|\mathcal{C}) = B_G \circ \mathbf{E}$. According to Kress (1999) this regularization approach of ill-posed problems traces back to *Ivanov (1962).

Concerning the existence of subsets G such that (4.20) yields a well-defined regularized Rao–Blackwellization $\mathbf{E}^{(G)}(\cdot|\mathcal{C})$ in the sense of Ivanov, we note the following.

4.14 Theorem (Ivanov Regularized Rao–Blackwellization) Let $p \in]1, \infty[$ and let G be an absolutely convex, closed, and boundedly compact⁹ subset of $(L_p(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q, p})$. Then, for each $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$

$$\operatorname{argmin}\{\|f - g\|_{\mathbf{LS}, p} : g \in G\} \quad (4.21)$$

is well-posed. Thus, the G -quasi Rao–Blackwellization $\mathbf{E}^{(G)}(f|\mathcal{C})$ is well-defined. \blacklozenge

Proof: Instead of (4.21), we consider first of all

$$\operatorname{argmin}\{\|\mathbf{E}_{\mathcal{C}}(f - g)\|_{\pi, p} : g \in G\}$$

with the linear continuous invertible expectation operator $\mathbf{E}_{\mathcal{C}}$ acting between the two Banach spaces $L_p(\mathfrak{X}, \mathcal{C}, Q)$ and $L_p(\mathfrak{F}, \mathfrak{B}, \pi)$, where we recall that, in fact,

$$\int \|\mathbf{E}_P f\|_{\mathbb{R}^k}^p \pi(dP) \leq \int \int \|f(x)\|_{\mathbb{R}^k}^p Q(dx) < \infty \quad (4.22)$$

for all $f \in L_p(\mathfrak{X}, \mathcal{C}, Q)$. Furthermore, $\|\cdot\|_{\pi, p}$ is strictly convex for all $p \in]1, \infty[$. Then, Vasin (2008, Theorem 2.2 and Corollary 2.7) yields that

⁹A subset M of a metric space (\mathfrak{X}, d) is called *boundedly compact* iff every closed (d) -metrically bounded subset of M is compact.

(4.22) has a unique solution $\mathbf{E}^{(G)}(f|\mathcal{C}) \in G \subset L_p(\mathfrak{X}, \mathcal{C}, Q)$ which depends continuously on $(\mathbf{E}_P f)_{P \in \mathfrak{P}} \in L_p(\mathfrak{Y}, \mathfrak{B}, \boldsymbol{\pi})$.

Note that $\mathbf{E}(\mathbf{E}(f|\mathcal{C})) = \mathbf{E}f$ for all $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$. Hence, we may continuously extend $\mathbf{E}^{(G)}(\cdot|\mathcal{C})$ on $L_p(\mathfrak{X}, \mathcal{A}, Q)$, that is,

$$\mathbf{E}^{(G)}(\cdot|\mathcal{C}) : (L_p(\mathfrak{X}, \mathcal{A}, Q), \|\cdot\|_{\mathbf{LS}, p}) \rightarrow (L_p(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q, p})$$

and $\mathbf{E}^{(G)}(f|\mathcal{C})$ is the unique solution of (4.21) for $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$. Note also that $\mathbf{E}^{(G)}(\cdot|\mathcal{C})|_G = \mathbf{id}|_G$. Hence, $\mathbf{E}^{(G)}(\cdot|\mathcal{C})$ is a regularized Rao–Blackwellization. \blacksquare

The idea of a quasi-solution naturally raises the question of constructing a regularization scheme for $\mathbf{E}_{\mathcal{C}}$ on the whole space L_p by considering a suitable family $(G)_{G \in \mathcal{G}}$ of absolutely convex, closed, and boundedly compact sets. A classical choice might be a family of finite-dimensional spaces, which we briefly consider in the following remark.

4.15 Remark (Ritz–Galerkin Solutions) Let $p = 2$ and let $(G_n)_{n \in \mathbb{N}}$ be an increasing sequence of finite-dimensional subspaces of $L_2(\mathfrak{X}, \mathcal{C}, Q)$. Theorem 4.14 yields then that $\mathbf{E}^{(G_n)}(\cdot|\mathcal{C})$ is well-defined through

$$\operatorname{argmin}\{\|f - g\|_{\mathbf{LS}, 2}^2 : g \in G_n\}.$$

Note that

$$\begin{aligned} \|f - g\|_{\mathbf{LS}, 2}^2 &= \langle f - g, f - g \rangle_{\mathbf{LS}} \\ &= \langle \mathbf{E}(f - g), \mathbf{E}(f - g) \rangle_{L_2(\boldsymbol{\pi})} \\ &= \langle \mathbf{E}'\mathbf{E}(f - g), (f - g) \rangle_{L_2(Q)}. \end{aligned}$$

Consequently, $\|\cdot\|_{\mathbf{LS}, 2}$ is the so-called *energy norm* that corresponds to the operator equation $\mathbf{E}'\mathbf{E}g = f$ and $\mathbf{E}^{(G_n)}(\cdot|\mathcal{C})$ is the (*Ritz–Galerkin solution* of $\mathbf{E}'\mathbf{E}g = f$ in G_n , see also Kress (1999, Section 13.5, page 240 ff.).

Even if $\bigcup_{n \in \mathbb{N}} G_n$ is dense in $(L_2(\mathfrak{X}, \mathcal{C}, Q), \|\cdot\|_{Q, 2})$, Galerkin solutions generally do not converge to the solution in $L_2(\mathfrak{X}, \mathcal{C}, Q)$. However, if we consider

$$(\mathbf{E}'\mathbf{E} + \alpha\mathbf{V}'\mathbf{V})g = f$$

for some $\alpha > 0$ instead of $\mathbf{E}'\mathbf{E}g = f$ and if we denote the corresponding Galerkin solution by $\mathbf{E}^{(G_n, \alpha)}(f|\mathcal{C})$, we have by Cea's Lemma (see e.g. Braess, 2003, Lemma 4.2, page 53)

$$\|\mathbf{E}^{(G_n, \alpha)}(f|\mathcal{C}) - \mathbf{E}(f|\mathcal{C})\|_{Q, 2} \leq \frac{1}{\alpha} \inf\{\|f - g\|_{\mathbf{LS}, 2} + \alpha\|f - g\|_{\mathbf{OS}, 2} : g \in G_n\},$$

since $\mathbf{E}'\mathbf{E} + \alpha\mathbf{V}'\mathbf{V}$ is strictly coercive for every positive α (see proof of Theorem 4.3, page 72). Consequently,

$$\lim_{n \rightarrow \infty} \|\mathbf{E}^{(G_n, \alpha)}(f|\mathcal{C}) - \mathbf{E}(f|\mathcal{C})\|_{Q, 2} = 0,$$

see also Kress (1999, Theorem 13.27, page 242 f.). Note that $\mathbf{E}^{(G_n, \alpha)}(f|\mathcal{C})$ is still different from the Rao–Blackwellization $\mathbf{E}^{(\alpha)}(f|\mathcal{C})$ obtained by Tikhonov’s regularization. Note that for $\alpha = 1$ we have that $\mathbf{E}^{(G_n, 1)}(f|\mathcal{C})$ is given by

$$\operatorname{argmin}\{\|f - g\|_{Q,2}^2 : g \in G_n\},$$

since the energy norm of $\mathbf{E}'\mathbf{E} + \mathbf{V}'\mathbf{V}$ is equivalent to $\|\cdot\|_{Q,2}$ and both are strictly convex. By this we have shown how regularized Rao–Blackwellizations can be computed (approximately). \blacklozenge

Theorem 4.14 yields a sufficient condition for a set G of prescribed solution characteristics that leads to a regularized Rao–Blackwellization. In Remark 4.15 we have specifically considered finite-dimensional subspaces G , that might be spanned by simple basis functions. For computational reasons, however, G might be chosen implicitly according to “representation properties” of the corresponding regularized Rao–Blackwellizations.

To this end, let \mathcal{C} be a complete sufficient σ -algebra for \mathfrak{F} and consider the Rao–Blackwell operator as the integral operator

$$\begin{aligned} \mathbf{E}(\cdot|\mathcal{C}) : \quad & (\mathbf{L}_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q}), \|\cdot\|_{\mathbf{LS},p}) \rightarrow (\mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathbf{Q}), \|\cdot\|_{Q,p}) \\ & f \quad \mapsto \quad \int f(x)P(dx|\mathcal{C}) \end{aligned} \quad (4.23)$$

for some $P \in \mathfrak{P}$ and some power $p \in [1, \infty]$. Substituting $P(\cdot|\mathcal{C})$ by another Markov kernel R from $(\mathfrak{X}, \mathcal{C})$ to $(\mathfrak{X}, \mathcal{A})$ such that

$$\begin{aligned} R : \quad & (\mathbf{L}_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q}), \|\cdot\|_{\mathbf{LS},p}) \rightarrow (\mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathbf{Q}), \|\cdot\|_{Q,p}) \\ & f \quad \mapsto \quad \int f(y)R(\cdot, dy) \end{aligned} \quad (4.24)$$

is a well-defined continuous operator leads then to an Ivanov regularized Rao–Blackwellization. This regularization is interesting because for an appropriately chosen subspace $G_R \subset \mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathbf{Q})$ the regularization $\mathbf{E}^{(G_R)}(\cdot|\mathcal{C})$ can be approximately computed by R .

4.16 Theorem Let $p \in]1, \infty[$ and let R be a Markov kernel from $(\mathfrak{X}, \mathcal{C})$ to $(\mathfrak{X}, \mathcal{A})$ such that (4.24) is well-defined and continuous. Furthermore, assume $[\mathfrak{P}] \subset \ker(R)$ and let $G_R \subset \mathbf{L}_p(\mathfrak{X}, \mathcal{C}, \mathbf{Q})$ be a finite-dimensional subspace such that $\|(I - R)|_{G_R}\| \leq \varepsilon$. Define

$$\|R\| := \sup\{\|Ru\|_{Q,p} : u \in \mathbf{L}_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q}), \|u\|_{\mathbf{LS},p} \leq 1\},$$

that is, the operator norm w. r. t. $\|\cdot\|_{\mathbf{LS},p}$, then

$$\begin{aligned} & \|\mathbf{E}^{(G_R)}(f|\mathcal{C}) - \int f(y)R(\cdot, dy)\|_{Q,p} \\ & \leq \|R\| \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\} + \varepsilon \cdot \|\mathbf{E}^{(G_R)}(f|\mathcal{C})\|_{Q,p} \end{aligned} \quad (4.25)$$

for all $f \in \mathbf{L}_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q})$. \blacklozenge

Proof: Let $f \in L_p(\mathfrak{X}, \mathcal{A}, Q)$ be arbitrary. According to Theorem 4.14, we first have that

$$\operatorname{argmin}\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\}$$

yields the well-defined Ivanov regularized Rao–Blackwellization of f given \mathcal{C} , which is denoted by $\mathbf{E}^{(G_R)}(f|\mathcal{C})$. Furthermore, R is a linear operator that is continuous w. r. t. $\|\cdot\|_{\mathbf{LS},p}$, that is, for all $f, g \in L_2(\mathfrak{X}, \mathcal{A}, Q)$ we have

$$\|Rf - Rg\|_{Q,p} \leq \|R\| \|f - g\|_{\mathbf{LS},p}.$$

Then, we obtain

$$\begin{aligned} & \|\mathbf{E}^{(G_R)}(f|\mathcal{C}) - \int f(y)R(\cdot, dy)\|_{Q,p} \\ & \leq \|\int \mathbf{E}^{(G_R)}(f|\mathcal{C})(y) - f(y)R(\cdot, dy)\|_{Q,p} \\ & \quad + \|\mathbf{E}^{(G_R)}(f|\mathcal{C}) - \int \mathbf{E}^{(G_R)}(f|\mathcal{C})(y)R(\cdot, dy)\|_{Q,p} \\ & = \|R(\mathbf{E}^{(G_R)}(f|\mathcal{C}) - f)\|_{Q,p} + \|(I - R)|_{G_R}(\mathbf{E}^{(G_R)}(f|\mathcal{C}))\|_{Q,p} \\ & \leq \|R\| \|\mathbf{E}^{(G_R)}(f|\mathcal{C}) - f\|_{\mathbf{LS},p} + \|(I - R)|_{G_R}\| \|\mathbf{E}^{(G_R)}(f|\mathcal{C})\|_{Q,p}. \end{aligned}$$

Finally, we have by definition of $\mathbf{E}^{(G_R)}(f|\mathcal{C})$ that

$$\|f - \mathbf{E}^{(G_R)}(f|\mathcal{C})\|_{\mathbf{LS},p} = \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\},$$

which yields the claim. \blacksquare

Roughly speaking, computing $\mathbf{E}^{(G_R)}(f|\mathcal{C})$ can be replaced by computing $\int f(y)R(\cdot, dy)$, if f is “close” to G_R w. r. t. $\|\cdot\|_{\mathbf{LS},p}$ and if R is “almost” idempotent, that is, $\|(I - R)|_{G_R}\|$ is “small”.

4.17 Example (Gaussian Model, Ivanov Regularized RB of the MLE) As in Example 4.5, page 74, let $\mathfrak{B} := \{\mathbf{N}(a, \sigma^2) : a \in \mathbb{R}\}$, let $\boldsymbol{\pi} := \mathbf{N}(b, \tau^2)$, and define $Q := \boldsymbol{\pi} \cdot \mathbf{id}_{\mathfrak{B}} = \mathbf{N}(b, \sigma^2 + \tau^2)$. Then, we first note that

$$\begin{aligned} R : (L_p(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Q), \|\cdot\|_{\mathbf{LS},p}) & \rightarrow (L_p(\mathbb{R}, \mathfrak{B}(\mathbb{R}), Q), \|\cdot\|_{Q,p}) \\ f & \mapsto \int f(x)\mathbf{N}(\cdot, \sigma^2)(dx) \end{aligned}$$

is well-defined and continuous. This is due to $\mathbf{N}(b, \sigma^2 + \tau^2) = \mathbf{N}(b, \tau^2)^T$ for some affine function T . By substitution, one can consequently see that there is an $M > 0$ such that

$$\begin{aligned} \|Rf - Rg\|_{Q,p}^p & \leq \int \left\| \int (f - g)(y)\mathbf{N}(x, \sigma^2)(dy) \right\|^p \mathbf{N}(b, \sigma^2 + \tau^2)(dx) \\ & \leq M \int \left\| \int (f - g)(y)\mathbf{N}(x, \sigma^2)(dy) \right\|^p \mathbf{N}(b, \tau^2)(dx) \\ & = M \cdot \|f - g\|_{\mathbf{LS},p}^p. \end{aligned}$$

Second, we obviously have that $f \in [\mathfrak{F}]$ implies $f \in \ker(R)$. Third, note that $G_R := \text{span}\{\mathbf{1}_{\mathbb{R}}, \mathbf{id}_{\mathbb{R}}\}$ is a (finite-dimensional) subspace of the eigenspace $\ker(\mathbf{I} - R)^{10}$ and, hence, $\|(\mathbf{I} - R)|_{G_R}\| = 0$. In total, Theorem 4.16 yields for the G_R -quasi Rao–Blackwellization that

$$\|\mathbf{E}^{(G_R)}(f|\mathcal{C}) - \int f(y)R(\cdot, dy)\|_{Q,p} \leq M \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\}$$

holds for general $f \in L_p(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \mathcal{Q})$ and in particular for the MLE

$$\mathbf{E}^{(G_R)}(\mathbf{id}_{\mathbb{R}}|\mathcal{C}) = \mathbf{id}_{\mathbb{R}},$$

which is different from the Tikhonov regularized Rao–Blackwellization (see Example 4.5, page 74). \blacklozenge

4.18 Example (Continuous Uniform Model) Let $p \in]1, \infty[$, $n \in \mathbb{N}$, consider the statistical model $\mathfrak{F} := \{\mathbf{U}_{[0,\vartheta]}^{\otimes n} : \vartheta \in [0, 1]\}$ on $([0, 1]^n, \mathfrak{B}([0, 1]^n))$, and set $\boldsymbol{\pi} := \mathbf{U}_{[0,1]}$. Then, a complete sufficient statistic for \mathfrak{F} is given by

$$\begin{aligned} S : [0, 1]^n &\rightarrow [0, 1] \\ x &\mapsto \max(x), \end{aligned}$$

which is also the MLE for the canonical parameter $\mathbf{id}_{[0,1]}$. By defining the map $R : (x, A) \mapsto \mathbf{U}_{[0,S(x)]}(A)$, we have a Markov kernel from $(\mathfrak{X}, \sigma(S))$ to $(\mathfrak{X}, \mathcal{A})$. Note that the corresponding Markov operator

$$\begin{aligned} R : (L_p(\mathfrak{X}, \mathcal{A}, \mathcal{Q}), \|\cdot\|_{\mathbf{LS},p}) &\rightarrow (L_p(\mathfrak{X}, \sigma(S), \mathcal{Q}), \|\cdot\|_{Q,p}) \\ f &\mapsto \int f(y)R(\cdot, dy) \end{aligned}$$

is continuous, since: $\mathcal{Q} := \boldsymbol{\pi} \cdot \mathfrak{F}$ has a density w. r. t. $\mathbf{U}_{[0,1]^n}$ which is

$$x \mapsto \frac{1}{n-1} \left(\frac{1}{S^{n-1}(x)} - 1 \right).$$

Furthermore, \mathcal{Q}^S has the density $\varphi : s \mapsto \frac{n}{n-1} (s^{n-1} - 1)$ w. r. t. $\boldsymbol{\pi}$, which is in particular bounded on $[0, 1]$, more precisely, $\sup \varphi = \frac{n}{n-1}$. Consequently,

$$\begin{aligned} \|Rf\|_{Q,p}^p &= \int \left\| \int f(y) \mathbf{U}_{[0,S(x)]}(dy) \right\|^p \mathcal{Q}(dx) \\ &= \int \left\| \int f(y) \mathbf{U}_{[0,\vartheta]}(dy) \right\|^p \mathcal{Q}^S(d\vartheta) \\ &\leq \frac{n}{n-1} \int \left\| \int f(y) \mathbf{U}_{[0,\vartheta]}(dy) \right\|^p \boldsymbol{\pi}(d\vartheta) \\ &\leq \frac{n}{n-1} \|f\|_{\mathbf{LS},p}^p \end{aligned}$$

¹⁰If R is restricted to tempered distributions, that is, to continuous linear functionals on the Schwartz space on \mathbb{R} , one can in fact show that $G_R = \ker(\mathbf{I} - R)$, see also Kwaśnicki (2020) and Vladimirov (2002).

for all $f \in L_p(\mathfrak{X}, \mathfrak{A}, \mathcal{Q})$ and, hence, R is continuous. Furthermore, due to

$$\frac{n}{n+1}S = \int S(y)R(\cdot, dy)$$

the Markov operator R has eigenfunctions $\mathbb{1}_{\mathbb{R}}$ and S with corresponding eigenvalues 1 and $\frac{n}{n+1}$. For $G_R := \text{span}\{\mathbb{1}_{\mathbb{R}}, S\}$ we may compute $\|(I-R)|_{G_R}\|$ through the maximal eigenvalue of $(I-R)|_{G_R}$, that is, $1 - \frac{n}{n+1}$ and, hence,

$$\|(I-R)|_{G_R}\| \leq M_p \cdot \left(1 - \frac{n}{n+1}\right) = \frac{M_p}{n+1}$$

for some constant M_p that depends on the chosen norm¹¹ $\|\cdot\|_p$. By Theorem 4.16 we then have

$$\begin{aligned} & \|\mathbf{E}^{(G_R)}(f|S) - \int f(y)R(\cdot, dy)\|_{Q,p} \\ & \leq \sqrt[p]{\frac{n}{n-1}} \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\} + \frac{M_p}{n+1} \|\mathbf{E}^{(G_R)}(f|S)\|_{Q,p}. \end{aligned}$$

By defining $\tilde{R} : (x, A) \mapsto \mathbf{U}_{[0, \frac{n+1}{n}S(x)]}(A)$, we have another Markov kernel from $(\mathfrak{X}, \sigma(S))$ to $(\mathfrak{X}, \mathfrak{A})$. The continuity of the corresponding Markov operator w. r. t. $\|\cdot\|_{\mathbf{LS},p}$ can be shown analogously with some constant $C > 0$ and we have

$$S = \int S(y)\tilde{R}(\cdot, dy),$$

that is, we have eigenfunctions $\mathbb{1}_{\mathbb{R}}$ and S to the same eigenvalue 1. Hence, $\|(I-\tilde{R})|_{G_R}\| = 0$ and, for the computation of the Ivanov regularized Rao-Blackwellization, we may note that

$$\|\mathbf{E}^{(G_R)}(f|S) - \int f(y)\tilde{R}(\cdot, dy)\|_{Q,p}^p \leq C \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\}. \quad (4.26)$$

Note that (4.26) in particular says that the Ivanov regularized Rao-Blackwellization $\mathbf{E}^{(G_R)}(\hat{\kappa}|\mathcal{C})$ of approximately unbiased estimators $\hat{\kappa}$ are close to the UMVUE $\frac{n+1}{n}S$. \blacklozenge

The arguments of the above example can also be used for a general exponential family.

4.19 Example (Exponential Family) Consider $p \in]0, 1[$ and let $\mathfrak{P} := \{P_{\vartheta} : \vartheta \in \Theta\}$ be a k -parametric exponential model on $(\mathfrak{X}, \mathfrak{A})$ with complete sufficient statistic S . Furthermore, for $\hat{\kappa} : (\mathfrak{X}, \sigma(S)) \rightarrow (\Theta, \mathcal{B}(\Theta))$, we set $\pi := P_{\vartheta_0}^{\hat{\kappa}}$ and assume that (1) $\text{supp } \pi = \Theta$ and, hence, S is π -complete; (2) $Q^{\hat{\kappa}}$ has a bounded density w. r. t. π , which we denote by φ . Then,

$$\begin{aligned} R : (L_p(\mathfrak{X}, \mathfrak{A}, \mathcal{Q}), \|\cdot\|_{\mathbf{LS},p}) & \rightarrow (L_p(\mathfrak{X}, \sigma(S), \mathcal{Q}), \|\cdot\|_{Q,p}) \\ f & \mapsto \int f(x)P_{\hat{\kappa}(\cdot)}(dx), \end{aligned} \quad (4.27)$$

¹¹In the Euclidean case $p = 2$ we may choose $M_2 = 1$.

is continuous. This is because, for $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q})$, we have with $M := \sup \varphi$

$$\begin{aligned} \|Rf\|_{\mathbf{Q},p}^p &= \int \left\| \int f(y) P_{\hat{\kappa}(x)}(\mathrm{d}y) \right\|_{\mathbf{Q}}^p \mathbf{Q}(\mathrm{d}x) \\ &= \int \left\| \int f(y) P_{\vartheta}(\mathrm{d}y) \right\|_{\mathbf{Q}^{\hat{\kappa}}}^p \mathbf{Q}^{\hat{\kappa}}(\mathrm{d}\vartheta) \\ &\leq M \int \left\| \int f(y) P_{\vartheta}(\mathrm{d}y) \right\|_{\mathbf{Q}}^p \boldsymbol{\pi}(\mathrm{d}\vartheta) \\ &= M \|f\|_{\mathbf{LS},p}^p. \end{aligned}$$

Assume that $\hat{\kappa}$ is an approximately unbiased estimator and that this can be quantified via

$$\left\| \hat{\kappa} - \int \hat{\kappa}(y) P_{\hat{\kappa}(\cdot)}(\mathrm{d}y) \right\|_{\mathbf{Q},p} < \frac{\varepsilon}{\|\hat{\kappa}\|_{\mathbf{Q},p}}$$

for some $\varepsilon > 0$. Let $G_R := \text{span}\{\mathbf{1}_{\mathbb{R}}, \hat{\kappa}\}$, then for every $h \in G_R$ there are $a, b \in \mathbb{R}$ such that $h = a\mathbf{1}_{\mathbb{R}} + b\hat{\kappa}$. Furthermore,

$$\|(\mathbf{I} - R)|_{G_R} h\|_{\mathbf{Q},p} \leq |a| \cdot \|(\mathbf{I} - R)|_{G_R} \mathbf{1}_{\mathbb{R}}\|_{\mathbf{Q},p} + |b| \cdot \|(\mathbf{I} - R)|_{G_R} \hat{\kappa}\|_{\mathbf{Q},p} \leq |b| \cdot \varepsilon.$$

Hence, $\|(\mathbf{I} - R)|_{G_R}\| \leq \varepsilon$ and the statement of Theorem 4.16 is in these terms

$$\begin{aligned} &\|\mathbf{E}^{(G_R)}(f|S) - \int f(y) R(\cdot, \mathrm{d}y)\|_{\mathbf{Q},p} \\ &\leq \sqrt[p]{M} \inf\{\|f - g\|_{\mathbf{LS},p} : g \in G_R\} + \varepsilon \|\mathbf{E}^{(G_R)}(f|S)\|_{\mathbf{Q},p} \end{aligned}$$

for all $f \in L_p(\mathfrak{X}, \mathcal{A}, \mathbf{Q})$. ◆

The regularized Rao–Blackwellization $\mathbf{E}^{(G_R)}(\cdot|S)$ due to Ivanov’s approach reduces the ancillary material of estimators w. r. t. the prior knowledge that is determined by G_R . Favorable or possibly ideal estimators in G_R remain unchanged if $\mathbf{E}^{(G_R)}(\cdot|S)$ is applied to these estimators. Theorem 4.16 shows that the Markov operator $R : f \mapsto \int f(y) R(\cdot, \mathrm{d}y)$ reduces the “almost” ancillary material of G_R -close estimators similarly as $\mathbf{E}^{(G_R)}(\cdot|S)$.

Regarding the above Example 4.19, the choice $\hat{\kappa} = \hat{\vartheta}_{\text{ML}}$ will be of special interest in the remainder of this section. By this we are interested in approximating the MLE for the canonical parameter \mathbf{id}_{Θ} through the corresponding Markov operator (4.27) considered in the above example. To this end, note that in this case $\hat{\vartheta}_{\text{ML}} \in G_R$ and, hence, $\mathbf{E}^{(G_R)}(\hat{\vartheta}_{\text{ML}}|S) = \hat{\vartheta}_{\text{ML}}$. Furthermore, we point out that

$$\begin{aligned} &\|\mathbf{E}^{(G_R)}(\hat{\vartheta}_{\text{ML}}|S) - \int \hat{\lambda}(y) R(\cdot, \mathrm{d}y)\|_{\mathbf{Q},p} \\ &= \|\hat{\vartheta}_{\text{ML}} - \int \hat{\lambda}(y) P_{\hat{\vartheta}_{\text{ML}}(\cdot)}(\mathrm{d}y)\|_{\mathbf{Q},p} = \|\text{Bias}(\hat{\lambda}, \hat{\vartheta}_{\text{ML}}(\cdot))\|_{\mathbf{Q},p} \end{aligned}$$

for every $\hat{\lambda} \in L_p(\mathfrak{X}, \mathcal{A}, \mathcal{Q})$. That means, the more unbiased $\hat{\lambda}$ is the better we approximate $\hat{\vartheta}_{\text{ML}}$ through $R\hat{\lambda}$. We are therefore interested in computing the latter which is the concern of the following theorem.

4.20 Theorem Let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be a k -parameter exponential family (full rank) with sufficient statistic $S \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$. Then, for an observation $x \in \mathfrak{X}$ with $S(x) \in \{\mathbf{E}_\vartheta S : \vartheta \in \text{int}(\Theta)\}$, the maximum likelihood estimator for the canonical parameter \mathbf{id}_Θ has the factorization $\hat{\vartheta}_{\text{ML}} = \hat{\theta}_{\text{ML}} \circ S$ on a neighborhood of $S(x)$ and we obtain the following approximation of $\hat{\theta}_{\text{ML}}$

$$\begin{aligned} \hat{\theta}_{\text{ML}}(s) = & \mathbf{E}_{\hat{\vartheta}_{\text{ML}}(x)}(\hat{\lambda}) + \text{Var}_{\hat{\vartheta}_{\text{ML}}(x)}^{-1}(S) \text{Cov}_{\hat{\vartheta}_{\text{ML}}(x)}(S, \hat{\lambda})(s - S(x)) \\ & + \text{Bias}(\hat{\lambda}, \hat{\theta}_{\text{ML}}(s)) + o(\|s - S(x)\|) \end{aligned} \quad (4.28)$$

where $\hat{\lambda} \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and s lies in a neighborhood of $S(x)$. \blacklozenge

Proof: First, we note that the MLE $\hat{\vartheta}_{\text{ML}}$ is measurable w. r. t. the sufficient statistic S on a neighborhood of $S(x)$ (see also Proposition 2.5 (a), page 32). Hence, we may consider the factorization $\hat{\vartheta}_{\text{ML}} = \hat{\theta}_{\text{ML}} \circ S$.

Second, we note that

$$\hat{\theta}_{\text{ML}} = \text{Bias}(\hat{\lambda}, \hat{\theta}_{\text{ML}}(\cdot)) + \int \hat{\lambda}(y) P_{\hat{\theta}_{\text{ML}}(\cdot)}(dy).$$

Third, we compute the linear Taylor approximation of the second term, that is, $\mathbf{E}_{\hat{\theta}_{\text{ML}}(\cdot)} \hat{\lambda} : s \mapsto \int \hat{\lambda}(y) P_{\hat{\theta}_{\text{ML}}(s)}(dy)$. To this end, we note that

$$\frac{\partial}{\partial s} \mathbf{E}_{\hat{\theta}_{\text{ML}}(s)} \hat{\lambda} = \frac{\partial}{\partial \hat{\theta}_{\text{ML}}(s)} \mathbf{E}_{\hat{\theta}_{\text{ML}}(s)} \hat{\lambda} \cdot \frac{\partial}{\partial s} \hat{\theta}_{\text{ML}}(s) \quad (4.29)$$

can be computed by using the prerequisite that \mathfrak{P} is an exponential family, see also Lemma 1.38, page 27 for the exchange of integration and differentiation. On the one hand, we may compute the Jacobian $\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\lambda} \in \mathbb{R}^{k \times k}$ as

$$\begin{aligned} \frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\lambda} &= \int \hat{\lambda}(y) \left(\frac{\partial}{\partial \vartheta} c(\vartheta) \right) \exp(\langle \vartheta, S(y) \rangle) \mu(dy) \\ &\quad + \int \hat{\lambda}(y) c(\vartheta) \left(\frac{\partial}{\partial \vartheta} \exp(\langle \vartheta, S(y) \rangle) \right) \mu(dy), \end{aligned}$$

where $c(\vartheta)^{-1} = \int \exp(\langle \vartheta, S(y) \rangle) \mu(dy)$ is the normalizing constant. Since

$$\frac{\partial}{\partial \vartheta} c(\vartheta) = -c(\vartheta)^2 \cdot \int S^\top(y) \exp(\langle \vartheta, S(y) \rangle) \mu(dy) = -c(\vartheta) \cdot \mathbf{E}_\vartheta S^\top,$$

we have in total

$$\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\lambda} = -\mathbf{E}_\vartheta \hat{\lambda} \cdot \mathbf{E}_\vartheta S^\top + \mathbf{E}_\vartheta \hat{\lambda} S^\top = \text{Cov}_\vartheta(S, \hat{\lambda}) \in \mathbb{R}^{k \times k}. \quad (4.30)$$

On the other hand, to compute $\frac{\partial}{\partial s} \hat{\theta}_{\text{ML}}(s) \in \mathbb{R}^{k \times k}$, we first recall that $\hat{\theta}_{\text{ML}}$ is implicitly defined by $\varphi(\theta(s), s) = 0$, where

$$\varphi : (\vartheta, s) \mapsto \mathbf{E}_{\vartheta} \mathbf{S} - s,$$

see Proposition 2.5 (a). By the theorem on implicit functions, we have that

$$\frac{\partial}{\partial s} \hat{\theta}_{\text{ML}}(s) = - \left[\frac{\partial}{\partial \vartheta} \varphi(\vartheta, s) \right]_{\vartheta = \hat{\theta}_{\text{ML}}(s)}^{-1} \left[\frac{\partial}{\partial s} \varphi(\vartheta, s) \right], \quad (4.31)$$

where we note that $\frac{\partial}{\partial s} \varphi(\vartheta, s) = -\mathbf{I}_k$ and

$$\left[\frac{\partial}{\partial \vartheta} \varphi(\vartheta, s) \right]_{\vartheta = \hat{\theta}_{\text{ML}}(s)}^{-1} = (\text{Var}_{\hat{\theta}_{\text{ML}}(s)} \mathbf{S})^{-1}, \quad (4.32)$$

which is due to (4.30). Finally, using (4.30), (4.31), and (4.32) the Jacobian in (4.29) turns out to be

$$\frac{\partial}{\partial s} \mathbf{E}_{\hat{\theta}_{\text{ML}}(s)} \hat{\lambda} = \text{Var}_{\hat{\theta}_{\text{ML}}(s)} \mathbf{S}^{-1} \text{Cov}_{\hat{\theta}_{\text{ML}}(s)}(\mathbf{S}, \hat{\lambda})$$

and, hence, yields the linear Taylor approximation of $\mathbf{E}_{\hat{\theta}_{\text{ML}}(\cdot)} \hat{\lambda}$ with the remaining term $o(\|s - S(x)\|)$. The statement follows then together with the second part. \blacksquare

The formula (4.28) paves a way toward obtaining an empirical distribution of the MLE by performing a linear regression based on simulated data from some (approximately) unbiased estimator (dependent variable) and the sufficient statistic S (covariate). This leads to the following practical procedure.

4.21 Algorithm (Empirical Distribution of the MLE) Let $\mathfrak{P} = \{P_{\vartheta} : \vartheta \in \Theta\}$ be a k -parametric exponential family with sufficient statistic $S \in L_2(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ and reference measure μ . Assume that $\hat{\theta}_{\text{ML}} = \hat{\vartheta}_{\text{ML}} \circ S$ is a \mathfrak{P} -integrable maximum likelihood estimator for the canonical parameter \mathbf{id}_{Θ} . Let $\hat{\lambda}$ denote an alternative estimator which is approximately unbiased for \mathbf{id}_{Θ} and easier to compute than the MLE. Given the observation $x \in \mathfrak{X}$, proceed as follows:

- (1) Compute $\hat{\vartheta}_{\text{ML}}(x)$;
- (2) Simulate $X_1, \dots, X_N \sim P_{\hat{\vartheta}_{\text{ML}}(x)}$;
- (3) Compute $(\hat{\lambda}(X_i))_{1 \leq i \leq N}$ and the sufficient statistic values $(S(X_i))_{1 \leq i \leq N}$;
- (4) Perform a linear regression of $(\hat{\lambda}(X_i))_{1 \leq i \leq N}$ on $(S(X_i))_{1 \leq i \leq N}$.

According to Theorem 4.20 the fitted regression line approximates the MLE $\hat{\theta}_{\text{ML}}$ locally, that is, in a neighborhood of $S(x)$, assuming also that N is large.¹² In this context, the fitted values can be seen as realizations

¹²Note that this rough conclusion makes use of the relation $\mathbf{E}_{\hat{\vartheta}_{\text{ML}}(x)} \mathbf{S} = S(x)$, that is, we may also assume that $S(X_1), \dots, S(X_N) \sim P_{\hat{\vartheta}_{\text{ML}}(x)}$ are “expected” to be in neighborhood of $S(x)$.

of $\hat{\vartheta}_{\text{ML}}$ under the empirical prior P_{ϑ_0} , where ϑ_0 is the unknown “true” parameter from Θ . Therefore, by the fitted values we have an approximate empirical distribution of $P_{\hat{\vartheta}_{\text{ML}}}$. \blacklozenge

We consider as a first simple example the MLE of the canonical parameter in the exponential model. For a more interesting model, the Lennard-Jones point process model, we refer to the simulation study in Chapter 7 (see Section 7.4 on page 167 f.). In the following we will write $\mathbf{E}^{(\text{LR})}(\hat{\lambda}|S)$ for the (empirical) linear regression line as a first-order Taylor approximation of $\mathbf{E}_{\hat{\vartheta}_{\text{ML}}(\cdot)}\hat{\lambda}$ and simply speak of a (Ivanov) regularized Rao–Blackwellization of $\hat{\lambda}$ given S . This is due to Theorem 4.16 and the intuitive role that $\mathbf{E}_{\hat{\vartheta}_{\text{ML}}(\cdot)}\hat{\lambda}$ is a (regularized) Rao–Blackwellization of $\hat{\lambda}$ given S .

4.22 Example (Exponential Distribution Model) For some $n \in \mathbb{N}$ and $\Theta :=]0, \infty[$ we consider the statistical model $\mathfrak{P} := (\mathbf{Exp}(\vartheta)^{\otimes n} : \vartheta \in \Theta)$. Recall that $S(x) = \sum_{i=1}^n x_i$ for $x \in \mathbb{R}^n$ is complete and sufficient for \mathfrak{P} . The MLE is known to be $\hat{\vartheta}_{\text{ML}}(x) = \frac{n}{S(x)}$ for $x \in \mathbb{R}^n$.

For an illustration of Theorem 4.20, we consider for some $m \leq n$ the estimator

$$\hat{\kappa}'_m(x) := \left(\frac{1}{m} \sum_{i=1}^m x_i \right)^{-1} \quad \text{for } x \in]0, \infty[^m.$$

Figure 4.7 shows the result of a simulation study. We simulate $N = 10000$ times from $\mathbf{Exp}(2)^{\otimes 50}$ and compute $(S, \hat{\kappa}'_{25})$. The linear regression of $\hat{\kappa}'_{25}$ on

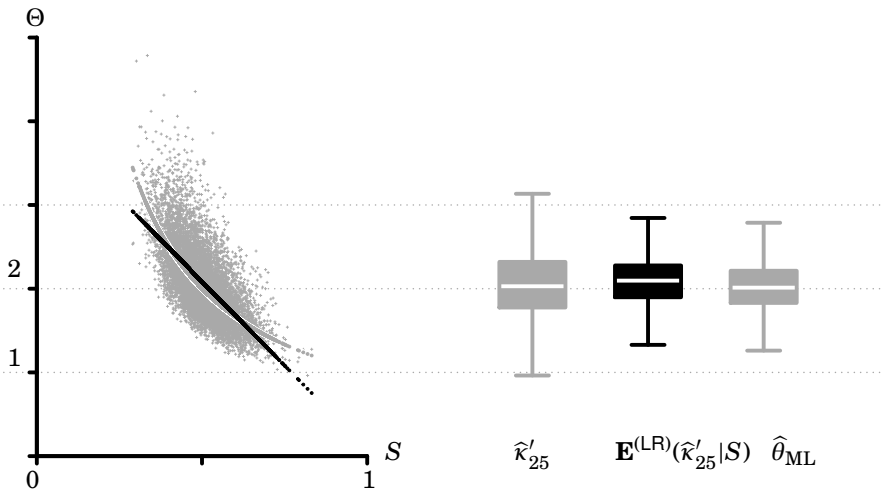


Figure 4.7: Joint empirical distribution of the sufficient statistic S together with the estimator $\hat{\kappa}'_{25}$ (gray point cloud), the MLE $\hat{\kappa}_{\text{ML}}$ (gray line; formed by points) and the fitted values of the linear regression of $\hat{\kappa}'$ on S (black line; formed by points) under P_2 . Corresponding boxplots are shown on the right.

S yields only a rough approximation of the MLE. However, distributions of $\mathbf{E}^{(\text{LR})}(\hat{\kappa}'_{25}|S)$ and the MLE seems to be fairly similar. ♦

In Chapter 7 we discuss Gibbs point process models that constitute more interesting situations, where the MLE is hard to compute. Therefore, Theorem 4.20 provides an access to compute even an empirical distribution of the MLE.

PART III

Statistics for Gibbs Point Processes

The overall aim of this part is an application of the regularized Rao–Blackwellization in the context of statistical inference for Gibbs point processes. To this end, we consider three chapters.

In Chapter 5 we first give an introduction to the theory of random measures, putting strong emphasis on point processes. We provide a motivation by a counterexample for the required technical assumptions which are often used in the literature (see Remark 5.1). For a better understanding of the object of interest, that is point processes, we also take a look at similar random elements that are known from the theory of stochastic processes (see e. g. Example 5.33). To show the existence of random measures, we provide a Kolmogorov version (see Proposition 5.25) which we then use to introduce the class of Poisson point process distributions. This preliminary chapter will be closed by a brief survey of basic notions and definitions regarding the so-called *interior* and *exterior conditioning* of point processes.

In Chapter 6 we are concerned with introducing distributions of Gibbs point processes, which is a class of point processes that account for interactions. This will be done through certain consistent families of conditional distributions, the so-called *specifications*. By this approach we are able to introduce the so-called *Gibbs distributions* even on non-compact spaces such as $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, which we need, for example, to consider stationary¹ distributions. However, in contrast to the Kolmogorov approach, questions concerning the existence and uniqueness are much more challenging. We provide an example of a specification that has no Gibbs point process distribution (see Example 6.6, see also Georgii (2011, page 66) for a version in the lattice case).

¹The concept of stationarity will be required for introducing the variational estimator in Chapter 7, for example.

A systematic treatment regarding the question of existence in general was given by Preston (1976) which later on was revised and modified in Preston (2005). Although Preston himself regarded the first presentation as “fairly abstract” (see Preston, 1976, preface) and the second as “not [...] suitable for learning about Gibbs [point processes]” (see Preston, 2005, preface), Chapter 6 also aims at a comprehensive survey of some key issues, mainly guided by a merge of the two mentioned monographs of Preston. To this end, we also embed some known results from the theory of Markov chains in this context of specifications and Gibbs measures. We also provide examples. The chapter on Gibbs point processes will be closed with a variational equation for stationary Gibbs point processes.²

The present part culminates in Chapter 7 on statistical inference for parametric Gibbs point process models. We provide a brief survey of estimation methods in a Gibbs point process setting which contains the *maximum likelihood estimator (MLE)*, the so-called *maximum pseudo-likelihood estimator (MPLE)*, and the *variational estimator (VARE)*. In line with the overall topic of this thesis, some emphasis is put on measurability aspects of these estimators. It is shown that the measurability of the MLE is less dependent on the boundary condition (see Proposition 7.1). Furthermore, we demonstrate that the MPLE is in general not measurable w. r. t. the complete sufficient statistic (see Remark 7.6). This chapter is closed with a simulation study for the (cut-off) Lennard-Jones point process model. Here, we illustrate aspects of the (Ivanov) regularized Rao–Blackwellization for different parameters and different boundary conditions. Furthermore, we propose constructions of confidence sets for the canonical parameter (see Methods 7.9, 7.10, and 7.11), which make use of the regularization idea. Then, these methods are also compared w. r. t. different boundary conditions.

²This variational equation motivates an unbiased estimating equation that leads to the variational estimator in Chapter 7.

CHAPTER 5

Basics on Random Measures

Random measures and specifically point processes appear frequently in probability theory and mathematical statistics. The theory of random measures can be used to present a general account of a sampling theory (see H. 2015). Random measures themselves serve as a probabilistic model for (spatial) point patterns arising in, for example, astronomy, biology, forestry, geology, and notably statistical physics. The latter has a high interest in studying models for interacting particles and has had a fundamental impact in developing the theory of so-called Gibbs point processes in the last century.

This chapter gives an introduction to the theory of random measures. Strong emphasis is put on point processes to prepare the reader for the subsequent chapters on Gibbs point processes and their statistical theory. Random measure theory shares some fundamental analogies with the theory of stochastic processes and their analysis. The present introduction also aims at this issue and, hence, will not just be guided by corresponding textbooks, such as Schneider & Weil (2000), Daley & Vere-Jones (2008), and Kallenberg (2017). Basic knowledge on probability theory is a prerequisite. We frequently make use of elementary results which can be found in standard textbooks, for example, Bauer (1991, 2001) and Kallenberg (2002). As usual, we refer to the literature where possible. However, proofs are presented if there is no suitable reference, or no reference at all. This is especially the case for auxiliary results which explain and motivate commonly used technical assumptions.

We start with elementary notions and definitions and investigate some peculiarities of the state spaces used for random measures and point processes. The presentation comes with elementary examples. We also provide crucial statements for easy reference.

5.1 The State Spaces

We will introduce the random objects in question, random measures and point processes, as a random element in appropriate measurable *state spaces*. This section is devoted to provide some preliminaries.

Let $(\mathfrak{X}, \mathcal{A})$ be a measurable space and denote by $\mathfrak{M}^* := \mathfrak{M}^*(\mathfrak{X}, \mathcal{A})$ the set of all measures on $(\mathfrak{X}, \mathcal{A})$. To deal with $\mathfrak{M}^*(\mathfrak{X}, \mathcal{A})$ -valued random elements, a measurable structure is required. That is, we have to equip $\mathfrak{M}^*(\mathfrak{X}, \mathcal{A})$ with a σ -algebra. We may consider the canonical σ -algebra

$$\mathcal{M}^* := \mathcal{M}^*(\mathfrak{X}, \mathcal{A}) := \sigma(\mu \mapsto \mu(A) : A \in \mathcal{A}).$$

A kernel K from a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ to $(\mathfrak{X}, \mathcal{A})$ can then be considered as an $(\mathfrak{M}^*, \mathcal{M}^*)$ -valued random element on $(\Omega, \mathcal{F}, \mathbf{P})$ and vice versa.

Note, however, that $\mathcal{M}^*(\mathfrak{X}, \mathcal{A})$ is the product σ -algebra on the function space \mathfrak{M}^* of σ -additive set functions $\mu : \mathcal{A} \rightarrow \mathbb{R}_+ \cup \{\infty\}$ and, hence, events are *countably determined*. That is, for every $F \in \mathcal{M}^*$ there is a countable $\mathcal{C} \subset \mathcal{A}$ such that $F = \text{pr}_{\mathcal{C}}^{-1}(\text{pr}_{\mathcal{C}}(F))$, where $\text{pr}_{\mathcal{C}} : \mathfrak{M}^* \rightarrow [0, \infty]^{\mathcal{C}}$ is the projection on set functions restricted to \mathcal{C} . Consequently, we encounter improper situations if we just consider $(\mathfrak{M}^*(\mathfrak{X}), \mathcal{M}^*(\mathfrak{X}))$ for a general measurable space $(\mathfrak{X}, \mathcal{A})$. By improper we mean the defect that the set of all Dirac measures is in general *not* measurable and hence does in general *not* constitute an event.¹

5.1 Remark (Improper Point Measures) Consider $\mathfrak{X} := [0, 1]$ and let \mathcal{A} be the σ -algebra generated by all singletons, that is, $\mathcal{A} := \sigma(\{x\} : x \in \mathfrak{X})$, the so-called *countable-cocountable* σ -algebra. Then, the set D consisting of all Dirac measures on \mathfrak{X} is not $\mathcal{M}^*(\mathfrak{X})$ -measurable.

We show that D is not countably determined. To this end, we define

$$\mu(A) := \begin{cases} 1 & \text{if } A^c \text{ is countable,} \\ 0 & \text{else} \end{cases}$$

which is a measure on $(\mathfrak{X}, \mathcal{A})$, that is, $\mu \in \mathfrak{M}^*(\mathfrak{X}, \mathcal{A})$. Then, for all countable $\mathcal{C} \subset \mathcal{A}$, we have

$$\mu \in \text{pr}_{\mathcal{C}}^{-1}(\text{pr}_{\mathcal{C}}(D)),$$

for the following reason: Let $A := \cap \{E \in \mathcal{C} : E^c \text{ countable}\}$ and $B := \cup \{E \in \mathcal{C} : E \text{ countable}\}$, then $A \setminus B$ is not empty and for every $x \in A \setminus B$ we have $\mu|_{\mathcal{C}} = \delta_x|_{\mathcal{C}}$. However, there is no $x \in \mathfrak{X}$ such that $\mu = \delta_x$ on \mathcal{A} , that is, $D \neq \text{pr}_{\mathcal{C}}^{-1}(\text{pr}_{\mathcal{C}}(D))$. That means, D is not countably determined and, hence, cannot be \mathcal{M}^* -measurable.

¹Remark 5.1 arose from a discussion on MathOverflow of Newman (2020) with the author, see H. (2020a).

We conclude this remark by some notes on the underlying topological structure of $(\mathfrak{X}, \mathcal{A})$. In fact, \mathcal{A} is the Borel- σ -algebra of the so-called *cocountable topology* $\mathcal{T} := \{U \subset \mathfrak{X}, : U = \emptyset \text{ or } U^c \text{ countable}\}$, that is, $\mathcal{A} = \mathcal{B}(\mathfrak{X}, \mathcal{T})$.² Limits of convergent sequences in $(\mathfrak{X}, \mathcal{T})$ are unique and compact sets in $(\mathfrak{X}, \mathcal{T})$ are measurable, although the intersection of any two non-empty \mathcal{T} -open sets is non-empty again, that is, $(\mathfrak{X}, \mathcal{T})$ is not a Hausdorff space. We, furthermore, emphasize the crucial point that the topology does *not* have a countable base (see also Definition 5.3 below). \blacklozenge

The message of Remark 5.1 is twofold. From a probabilistic point of view, it says first of all that the set of all Dirac measure does in general *not* constitute an event in a probability space of the form $(\mathfrak{M}^*, \mathcal{M}^*, P)$ for some distribution P . In fact, the same holds true for

$$\{\sum_{x \in M} \delta_x : M \text{ countable subset of } \mathfrak{X}\}.$$

These events are quite important and natural since their elements relate to *point patterns* through their support. With regard to point processes, a theory of random measures on mere abstract measure spaces appears therefore to be deficient.

Second, in order to avoid non-measurable sets that are desired to be events, Remark 5.1 also indicates to look for situations where measures on $(\mathfrak{X}, \mathcal{A})$ are countably determined.³

The theory of random *locally finite* measures on *locally compact second countable Hausdorff spaces* provides a sufficient and commonly used framework to cover point processes. We recall some basic definitions.

5.2 Definition Let $(\mathfrak{X}, \mathcal{T})$ be a topological space. A measure μ on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ is called *locally finite* if and only if for every $x \in \mathfrak{X}$ there is a $U \in \mathcal{T}$ such that $x \in U$ and $\mu(U) < \infty$. The set of all locally finite measures on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ is denoted by $\mathfrak{M} := \mathfrak{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. \blacklozenge

5.3 Definition A topological space $(\mathfrak{X}, \mathcal{T})$ is called *locally compact* iff every $x \in \mathfrak{X}$ has a compact neighborhood, that is, there is a compact set $K \subset \mathfrak{X}$ and an open set $U \in \mathcal{T}$ such that $x \in U \subset K$.

A topological space $(\mathfrak{X}, \mathcal{T})$ is called *second countable* if and only if there is a countable base $\mathcal{U} \subset \mathcal{T}$ of \mathcal{T} , that is, for every $V \in \mathcal{T}$ there is a $\mathcal{V} \subset \mathcal{U}$ such that $V = \bigcup \mathcal{V}$. \blacklozenge

²Recall that $\mathcal{B}(\mathfrak{X}) := \mathcal{B}(\mathfrak{X}, \mathcal{T}) := \sigma(\mathcal{T})$ denotes the Borel- σ -algebra that corresponds to the topological space $(\mathfrak{X}, \mathcal{T})$.

³In the theory of stochastic processes similar problems occur. A stochastic process $X = (X_t)_{t \in T}$ is an $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$ -random element. Naturally, one is interested in “events” like $\{X \text{ has continuous paths}\}$. However, the set of continuous paths $\mathbf{C}(T)$ is in general not $\mathcal{B}(\mathbb{R}^T)$ -measurable. To evade these problems, one considers so-called *separable stochastic processes*, that is, processes $X = (X_t)_{t \in T}$, where paths are countably determined. Here, one usually needs a separable metric space (T, d) which corresponds to our countably generated Borel- σ -algebra.

Events consisting of locally finite measures on locally compact second countable Hausdorff spaces are countably determined. This is the message of the following lemma.

5.4 Lemma Let $(\mathfrak{X}, \mathcal{T})$ be a locally compact second countable Hausdorff space. Then, there is a countable $\mathcal{E} \subset \mathcal{B}(\mathcal{T})$ that determines locally finite measures uniquely. \blacklozenge

Proof: By definition there is a countable basis \mathcal{U} of \mathcal{T} . Furthermore, there is a sequence $G = (G_k)_{k \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ of open, relatively compact sets such that $G_k \uparrow \mathfrak{X}$ for $k \rightarrow \infty$ (see e. g. Schneider & Weil, 2000). Then

$$\mathcal{E} := \left\{ \bigcap \mathcal{O} : \mathcal{O} \subset \mathcal{U} \cup G(\mathbb{N}), |\mathcal{O}| < \infty \right\}$$

is countable and generates $\mathcal{B}(\mathfrak{X}, \mathcal{T})$. Let μ be a locally finite measure on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}, \mathcal{T}))$. Then $\mu(G_k) < \infty$ for all k since G_k are relatively compact. Furthermore, by the measure uniqueness theorem we have that μ is uniquely determined by $\mu|_{\mathcal{E}}$. \blacksquare

For the remaining chapter, we assume that $(\mathfrak{X}, \mathcal{T})$ is a locally compact second countable Hausdorff space. We denote by $\mathcal{H} = \mathcal{H}(\mathfrak{X})$ the system of relatively compact Borel measurable subsets of \mathfrak{X} . Note that compact subsets of \mathfrak{X} are closed and, hence, Borel measurable. Therefore, such sets are elements in $\mathcal{H}(\mathfrak{X})$. We may call sets in \mathcal{H} *bounded* due to the fact that there is a \mathcal{T} -compatible metric d such that (\mathfrak{X}, d) is boundedly compact (see Vaughan, 1937).⁴

One can easily show that a measure μ on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ is locally finite if and only if $\mu(B) < \infty$ for all $B \in \mathcal{H}(\mathfrak{X})$. We correspondingly define:

5.5 Definition A locally finite measure $\xi \in \mathfrak{M}(\mathfrak{X})$ is called *point measure* if and only if $\xi(B) \in \mathbb{N}_0$ for all $B \in \mathcal{H}$. The set of all locally finite point measures is denoted by $\mathfrak{N}^*(\mathfrak{X}) := \mathfrak{N}^*(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$.

Furthermore, a locally finite point measure $\xi \in \mathfrak{N}^*(\mathfrak{X})$ is called *simple* if and only if $\xi(\{x\}) \leq 1$ for all $x \in \mathfrak{X}$. The set of all *simple* locally finite point measures is denoted by $\mathfrak{N}(\mathfrak{X}) := \mathfrak{N}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. \blacklozenge

We are now going to equip $\mathfrak{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ with the σ -algebra

$$\mathcal{M} := \mathcal{M}(\mathfrak{X}) := \mathcal{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X})) := \sigma(\mu \mapsto \mu(B) : B \in \mathcal{H})$$

and note the following.

5.6 Lemma In the above situation we have $\mathfrak{N}(\mathfrak{X}), \mathfrak{N}^*(\mathfrak{X}) \in \mathcal{M}(\mathfrak{X})$. \blacklozenge

⁴In an arbitrary metrizable topological space $(\mathfrak{X}, \mathcal{T})$ relatively compact sets have a bounded diameter w. r. t. any \mathcal{T} -compatible metric d , that is, they are *(d)-metrically bounded*. However, metrically bounded sets do not need to be relatively compact. If every d -metrically bounded subset is relatively compact, we call the metric space (\mathfrak{X}, d) *boundedly compact*.

Proof: See for example Kallenberg (1983). ■

On $\mathfrak{N}(\mathfrak{X})$ and $\mathfrak{N}^*(\mathfrak{X})$ we consider the trace- σ -algebras

$$\mathcal{N}^* := \mathcal{N}^*(\mathfrak{X}) := \mathcal{N}^*(\mathfrak{X}, \mathcal{B}(\mathfrak{X})) := \mathcal{M}(\mathfrak{X}) \cap \mathfrak{N}^*(\mathfrak{X})$$

and

$$\mathcal{N} := \mathcal{N}(\mathfrak{X}) := \mathcal{N}(\mathfrak{X}, \mathcal{B}(\mathfrak{X})) := \mathcal{M}(\mathfrak{X}) \cap \mathfrak{N}(\mathfrak{X}).$$

According to Lemma 5.6, we immediately have that $(\mathfrak{N}, \mathcal{N})$ and $(\mathfrak{N}^*, \mathcal{N}^*)$ are measurable subspaces of $(\mathfrak{M}, \mathcal{M})$. Hence, the state spaces for random measures and point processes will be $(\mathfrak{M}, \mathcal{M})$ and $(\mathfrak{N}, \mathcal{N})$, respectively

A simple point measure serves as a formalization of a point pattern. Another and possibly more appealing formal description of a point pattern might be given by a locally finite subset of \mathfrak{X} . In fact, these two formalisms coincide.

5.7 Remark and Definition (Locally Finite Sets vs Simple Point Measures) A set $F \subset \mathfrak{X}$ is called *locally finite* iff $|F \cap B| < \infty$ holds for all $B \in \mathcal{H}(\mathfrak{X})$. The set of all locally finite subsets of \mathfrak{X} is denoted by $\mathfrak{A}_{\text{lf}} := \mathfrak{A}_{\text{lf}}(\mathfrak{X}, \mathcal{T})$. On \mathfrak{A}_{lf} we may consider the σ -algebra

$$\mathcal{E}_{\text{lf}} := \mathcal{E}_{\text{lf}}(\mathfrak{X}, \mathcal{T}) := \sigma(\{F \in \mathfrak{A}_{\text{lf}} : F \cap G \neq \emptyset : G \in \mathcal{T}\}),$$

where we additionally note that there are several generators of \mathcal{E}_{lf} , see for example H. (2015, Section A.2) for a brief overview.

In fact, $(\mathfrak{A}_{\text{lf}}, \mathcal{E}_{\text{lf}})$ is measurably equivalent to $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$, that is, there is a bijective map $\varphi : (\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X})) \rightarrow (\mathfrak{A}_{\text{lf}}, \mathcal{E}_{\text{lf}})$ such that φ and φ^{-1} are measurable (see Schneider & Weil, 2000, Satz 3.1.2, page 63 f.). ◆

The following convention concerning the use of set theoretical operations and relations for simple point measures will be quite useful.

5.8 Remark (Set Theoretical Operations on \mathfrak{N}) According to Remark and Definition 5.7, we do not have to distinguish between $(\mathfrak{A}_{\text{lf}}(\mathfrak{X}), \mathcal{E}_{\text{lf}}(\mathfrak{X}))$ and $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$. Binary relations and operations in one space are well-defined in the other and vice versa. By the isomorphism φ from Remark and Definition 5.7, we may define counterparts of the set theoretical operations on $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$, for example, for the union \cup we set

$$\begin{aligned} \cup : (\mathfrak{N} \times \mathfrak{N}, \mathcal{N} \otimes \mathcal{N}) &\rightarrow (\mathfrak{N}, \mathcal{N}) \\ (\xi, \eta) &\mapsto \varphi^{-1}(\varphi(\xi) \cup \varphi(\eta)) \end{aligned}$$

and write as usual $\xi \cup \eta$ instead of $\cup(\xi, \eta)$. Analogously, we define the intersection $(\xi, \eta) \mapsto \xi \cap \eta$ and the set theoretical difference $(\xi, \eta) \mapsto \xi \setminus \eta$.

Two simple point measures are said to be *disjoint*, denoted by $\xi \cap \eta = \emptyset$, iff $\varphi(\xi)$ and $\varphi(\eta)$ are disjoint. The binary statement $x \in \xi$ is considered to be

true or false iff $x \in \varphi(\xi)$ is true or false, respectively. For the measurability of these binary relations we refer to Schneider & Weil (2000).

Note that the mentioned maps and relations extend the already existing maps and relations on \mathfrak{N} . For example, \cup and $+$ coincide as long as the arguments are disjoint. \blacklozenge

The state spaces $(\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ and $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ can be identified with subspaces of the product space

$$\left(\prod_{B \in \mathcal{H}(\mathfrak{X})} \mathbb{R}_+, \otimes_{B \in \mathcal{H}(\mathfrak{X})} \mathcal{B}(\mathbb{R}_+)\right).$$

This is due to the fact that $\mathfrak{M}(\mathfrak{X})|_{\mathcal{H}(\mathfrak{X})}$ and $\mathfrak{N}(\mathfrak{X})|_{\mathcal{H}(\mathfrak{X})}$ are a subspaces of $\prod_{B \in \mathcal{H}} \mathbb{R}_+$, due to the definition of $\mathcal{M}(\mathfrak{X})$ and $\mathcal{N}(\mathfrak{X})$,⁵ and due to the measure uniqueness theorem as well.

We are now going to show that $(\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ and $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ exhibit another product space representation. For that reason, we introduce the following notions.

5.9 Remark and Definition Let $B \in \mathcal{B}_{\mathfrak{X}}$ and define

$$\mathfrak{M}_B := \mathfrak{M}_B(\mathfrak{X}) := \{\xi \in \mathfrak{M}(\mathfrak{X}) : \xi(\cdot \cap B^c) = \mathbf{0}\},$$

which is a subspace of $\mathfrak{M}(\mathfrak{X})$. Analogously, we define $\mathfrak{N}_B := \mathfrak{N}_B(\mathfrak{X})$ which is then a subspace of $\mathfrak{N}(\mathfrak{X})$. Furthermore, consider on \mathfrak{M}_B the σ -algebra

$$\mathcal{M}_B := \mathcal{M}_B(\mathfrak{X}) := \sigma(\xi \mapsto \xi(F) : F \in \mathcal{H}(B)),$$

and correspondingly $\mathcal{N}_B := \mathcal{N}_B(\mathfrak{X})$ on \mathfrak{N}_B . Let $B, C \in \mathcal{B}_{\mathfrak{X}}$ be such that $C \subset B$, then

$$\begin{array}{ccc} \text{pr}_C^B : & (\mathfrak{M}_B, \mathcal{M}_B) & \rightarrow & (\mathfrak{M}_C, \mathcal{M}_C) \\ & \xi & \mapsto & \xi(\cdot \cap C) \end{array}$$

is a projection which is measurable, because we have

$$\xi(F) = \xi(F \cap C) = \text{pr}_C^B(\xi)(F) \quad \text{for every } F \in \mathcal{B}_{\mathfrak{X}} \cap C.$$

Hence, according to the definition of \mathcal{M}_B , we have for every bounded $F \in \mathcal{B}_{\mathfrak{X}} \cap C$ that $\xi \mapsto \text{pr}_C^B(\xi)(F)$ is measurable w. r. t. \mathcal{N}_B , which implies $(\text{pr}_C^B)^{-1}(\mathcal{N}_C) \subset \mathcal{N}_B$. We just write pr_B for $\text{pr}_B^{\mathfrak{X}}$. Furthermore, we write ξ_B for $\text{pr}_B(\xi)$ if we consider a specific point pattern ξ . \blacklozenge

With the projection on sub-point patterns, we gain a different point of view on locally finite measures. We are now going to show the following product space representation of the space of locally finite point measures which will be frequently used in what follows. The same arguments yields the product space representation of $(\mathfrak{M}, \mathcal{M})$.

⁵Let \mathfrak{Y} be an arbitrary set. Then, for some $\mathfrak{Y}_0 \subset \mathfrak{Y}$ and some set system $\mathcal{E} \subset 2^{\mathfrak{Y}}$ it holds that $\sigma_{\mathfrak{Y}_0}(\mathcal{E} \cap \mathfrak{Y}_0) = \sigma_{\mathfrak{Y}}(\mathcal{E}) \cap \mathfrak{Y}_0$, where $\sigma_{\mathfrak{Y}}(\mathcal{E})$ denotes the σ -algebra on \mathfrak{Y} that is generated by \mathcal{E} .

5.10 Proposition Let \mathcal{D} be a countable partition of \mathfrak{X} . Then, we have

$$(\mathfrak{N}, \mathcal{N}) \cong \left(\prod_{D \in \mathcal{D}} \mathfrak{N}_D, \otimes_{D \in \mathcal{D}} \mathcal{N}_D \right),$$

that is, these measure spaces are measurably isomorphic. \blacklozenge

Proof: The map

$$\begin{aligned} \varphi : \mathfrak{N} &\rightarrow \prod_{D \in \mathcal{D}} \mathfrak{N}_D \\ \xi &\mapsto (\text{pr}_D(\xi))_{D \in \mathcal{D}} \end{aligned}$$

is obviously well-defined and bijective. Furthermore, if $\text{pr}_{\{D\}}$ denotes the projection of $(\xi_D)_{D \in \mathcal{D}}$ on the D -th coordinate, we see that the measurability of φ follows simply from the fact that for any $D \in \mathcal{D}$ the projection $\text{pr}_D = \text{pr}_{\{D\}} \circ \varphi$ is measurable. We obtain the measurability of φ^{-1} , that is, the measurability of the function $(\xi_D)_{D \in \mathcal{D}} \mapsto \sum_{D \in \mathcal{D}} \xi_D$, from the fact that, for every $F \in \mathcal{H}$, the function

$$(\xi_D)_{D \in \mathcal{D}} \mapsto \sum_{D \in \mathcal{D}} \xi_D(F)$$

is measurable as a pointwise limit of measurable functions. \blacksquare

5.11 Remark (Compatibility of Projections) In Remark 5.9 we have introduced the projection $\text{pr}_B : \mathfrak{N} \rightarrow \mathfrak{N}_B$ defined by $\xi \mapsto \xi(\cdot \cap B)$ for some measurable set $B \in \mathcal{B}(\mathfrak{X})$. In the situation of Proposition 5.10 and especially regarding the product space representation in the previous proposition, we note that for some $\mathcal{E} \subset \mathcal{D}$ the projection

$$\begin{aligned} \text{pr}_{\mathcal{E}} : \prod_{D \in \mathcal{D}} \mathfrak{N}_D &\rightarrow \prod_{E \in \mathcal{E}} \mathfrak{N}_E \\ (\xi_D)_{D \in \mathcal{D}} &\mapsto (\xi_E)_{E \in \mathcal{E}} \end{aligned}$$

on a sub-product space is compatible with the corresponding projection $\text{pr}_{\cup \mathcal{E}}$ on the subset $\cup \mathcal{E}$. This situation is presented by the following commutative diagram, where φ is the corresponding bijective map from the proof of Proposition 5.10:

$$\begin{array}{ccc} (\mathfrak{N}, \mathcal{N}) & \cong_{\varphi} & \left(\prod_{D \in \mathcal{D}} \mathfrak{N}_D, \otimes_{D \in \mathcal{D}} \mathcal{N}_D \right) \\ \downarrow \text{pr}_{\cup \mathcal{E}} & & \downarrow \text{pr}_{\mathcal{E}} \\ (\mathfrak{N}_{\cup \mathcal{E}}, \mathcal{N}_{\cup \mathcal{E}}) & \cong_{\varphi} & \left(\prod_{E \in \mathcal{E}} \mathfrak{N}_E, \otimes_{E \in \mathcal{E}} \mathcal{N}_E \right) \end{array}$$

That means, it does not matter which path we follow. Hence, we have, for example, that $\text{pr}_{\cup \mathcal{E}} = \varphi^{-1} \circ \text{pr}_{\mathcal{E}} \circ \varphi$. \blacklozenge

Doing probability theory on the space $(\mathfrak{N}, \mathcal{M})$ presupposes the existence of distributions that are worth studying. Clearly, trivial distributions

like Dirac measures always exist on any measurable space. Non-trivial distributions such as non-atomic ones, however, do not always exist (see e. g. Billingsley, 1995, page 46). In addition, the support of certain point process distributions will be of special interests, in particular for aspects of statistical modeling in point process statistics. With regard to such issues, we therefore close this section with some notes about the underlying topological structure of $(\mathfrak{M}, \mathcal{M})$.

We define on $\mathfrak{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ the so-called *vague topology*⁶

$$\mathcal{T}_v := \tau(\mu \mapsto \int f d\mu : f \in \mathbf{C}(\mathfrak{X}, \mathcal{T}), \text{supp } f \text{ compact}),$$

that is, the coarsest topology such that $\mu \mapsto \int f d\mu$ is continuous for all real-valued continuous functions f on $(\mathfrak{X}, \mathcal{T})$ with compact support.

The next proposition 5.12 states that \mathcal{M} is a Borel- σ -algebra to the Polish space $(\mathfrak{M}, \mathcal{T}_v)$.

5.12 Proposition The topological space $(\mathfrak{M}, \mathcal{T}_v)$ is a Polish space and \mathcal{M} is generated by \mathcal{T}_v , that is, $\mathcal{M} = \mathcal{B}(\mathfrak{M}, \mathcal{T}_v)$. \blacklozenge

Proof: See Kallenberg (2017, Theorem 4.2). \blacksquare

Proposition 5.12 yields also that $\mathcal{N}^*(\mathfrak{X})$ and $\mathcal{N}(\mathfrak{X})$ are Borel- σ -algebras on the set of locally finite and the set of *simple* locally finite point measures on \mathfrak{X} , since $\mathcal{N}^*(\mathfrak{X}) = \mathcal{B}(\mathfrak{M}, \mathcal{T}_v) \cap \mathfrak{N}^*(\mathfrak{X})$ and $\mathcal{N}(\mathfrak{X}) = \mathcal{B}(\mathfrak{M}, \mathcal{T}_v) \cap \mathfrak{N}(\mathfrak{X})$. Furthermore, we note the following well-known result.

5.13 Proposition The set \mathfrak{N}^* is closed in $(\mathfrak{M}, \mathcal{T}_v)$ and $(\mathfrak{N}^*, \mathcal{T}_v \cap \mathfrak{N}^*)$ is a Polish space. \blacklozenge

Proof: See for example Matthes *et al.* (1978, page 162) for the statement that \mathfrak{N}^* is closed in $(\mathfrak{M}, \mathcal{T}_v)$. The Polishness of $(\mathfrak{N}^*, \mathcal{T}_v \cap \mathfrak{N}^*)$ follows from Proposition 5.12 and the closeness of \mathfrak{N}^* w. r. t. \mathcal{T}_v (see also Bauer, 1992, page 179). \blacksquare

Another topological structure that fits to the measurable space $(\mathfrak{N}, \mathcal{N})$ is given by the following hit- and miss topology.

5.14 Proposition The σ -algebra $\mathcal{N}(\mathfrak{X})$ on $\mathfrak{N}(\mathfrak{X})$ is also generated by

$$\mathcal{T}_F = \tau(\{\xi : \xi(G) > 0\}, \{\xi : \xi(C) = 0\} : G \in \mathcal{H} \text{ open}, C \in \mathcal{H} \text{ compact}),$$

which is called *Fell topology*. \blacklozenge

Proof: Follows by H. (2015, Satz A.21), a corresponding restriction to locally finite sets, and finally by Remark and Definition 5.7. \blacksquare

⁶We denote by $\mathbf{C}(\mathfrak{X}, \mathcal{T})$ the set of real-valued continuous functions on a topological space $(\mathfrak{X}, \mathcal{T})$.

5.2 Random Measures

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and recall that in the remaining chapter $(\mathfrak{X}, \mathcal{T})$ is assumed to be a locally compact second countable Hausdorff space. We start with the principal definition of this section.

5.15 Definition A random element $\Xi : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ is called *random measure* on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. A random measure Ξ on \mathfrak{X} such that $\Xi \in \mathfrak{N}$ holds \mathbf{P} -almost surely is called *point process*. \blacklozenge

Note that the term “random measure” is not always used in the same way in the literature. Here, a random measure \mathbf{M} is required to be a measure for every realization $\omega \in \Omega$. In Applebaum (2009, page 89) the definition of a random measure already deviates by the requirement that the σ -additivity only has to hold \mathbf{P} -almost surely.

The measurability w. r. t. \mathcal{M} can be characterized as follows.

5.16 Lemma A map $\Xi : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ is measurable iff

$$\Xi(B) : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$$

is measurable for all B from a semi-ring \mathcal{C} of bounded sets such that \mathcal{C} generates $\mathcal{B}(\mathfrak{X})$. \blacklozenge

Proof: See Daley & Vere-Jones (2008, Proposition 9.1.VIII, page 8). \blacksquare

The random object we are studying has several useful interpretations. By definition and Lemma 5.16, a random measure is a kernel from a *probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ to $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. Hence, for some \mathfrak{X} -valued random element X , an immediate example is given by $(\omega, B) \mapsto \delta_{X(\omega)}(B)$ which we may call *Bernoulli point process*. Correspondingly, we obtain the so-called binomial point process.

5.17 Example (Binomial Point Process) For $n \in \mathbb{N}$ let X_1, \dots, X_n be \mathfrak{X} -valued i. i. d. random elements, then

$$\Xi := \sum_{k=1}^n \delta_{X_k}$$

is a point process, which is called *binomial point process*. \blacklozenge

5.18 Example (Bayesian Statistical Model) Let $(P_\vartheta)_{\vartheta \in \Theta}$ be a dominated parametrized statistical model on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ with parameter space $\Theta \subset \mathbb{R}^k$. For $\vartheta \in \Theta$ we denote by f_ϑ the density of P_ϑ w. r. t. the dominating measure. Assume that $(\vartheta, x) \mapsto f_\vartheta(x)$ is $\mathcal{B}(\Theta) \otimes \mathcal{B}(\mathfrak{X})$ -measurable.

Furthermore, assume that information or “some belief” is a priori given in terms of a prior distribution of a $(\Theta, \mathcal{B}(\Theta))$ -valued random element θ . Then, P_θ is a random measure on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, since $\omega \mapsto P_{\theta(\omega)}(B)$ is measurable for all $B \in \mathcal{B}(\mathfrak{X})$. \blacklozenge

A random measure may be interpreted as a $\mathcal{B}(\mathfrak{X})$ -indexed stochastic process such that the paths are required to be locally finite measures.

5.19 Remark (Set-Indexed Stochastic Processes) Every random measure Ξ on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ is obviously a $\mathcal{B}(\mathfrak{X})$ -indexed stochastic process, that is, a $\mathcal{B}(\mathfrak{X})$ -indexed family of random variables. Conversely, let $(X_B)_{B \in \mathcal{B}(\mathfrak{X})}$ be an $\overline{\mathbb{R}}_+$ -valued stochastic process such that for \mathbf{P} -almost all $\omega \in \Omega$

$$B \mapsto X_B(\omega)$$

is a locally finite measure. Then, there is a random measure Ξ on \mathfrak{X} with $\Xi = X$ almost surely w. r. t. \mathbf{P} . To see this, define

$$\Xi(\omega, B) := X_B(\omega) \mathbb{1}\{\omega' \in \Omega : X(\omega', \cdot) \text{ is a measure}\}$$

for all $B \in \mathcal{B}(\mathfrak{X})$. Then, $\Xi(\cdot, B)$ is measurable for all $B \in \mathcal{B}(\mathfrak{X})$. Hence, Ξ is an \mathfrak{M} -valued random element. \blacklozenge

Random measures on some $T \subset \mathbb{R}_+$ correspond to stochastic processes with increasing⁷ paths that are continuous from the right and, therefore, provide a generalization of these processes.

5.20 Proposition Let $T \subset \mathbb{R}_+$ and let $X = (X_t)_{t \in T}$ be an \mathbb{R} -valued stochastic process such that \mathbf{P} -almost all paths are increasing and continuous from the right. Then, there is a unique random measure Ξ on T such that $\Xi([s, t]) = X_t - X_s$ for all $s, t \in T$, $s \leq t$ almost surely w. r. t. \mathbf{P} . \blacklozenge

Proof: We define

$$\Omega' := \{\omega \in \Omega : X(\omega, \cdot) \text{ increasing, continuous from the right}\}.$$

Then, it is well-known (see e. g. Bauer, 1992, Satz 6.5, page 36) that for every $\omega \in \Omega'$ there exists exactly one measure $\Xi(\omega, \cdot)$ on T such that

$$\Xi(\omega, [s, t]) := X_t(\omega) - X_s(\omega)$$

for all $s, t \in T$ with $s \leq t$. Hence, we have that $\Xi(\cdot, J)$ is measurable for all semi-open intervals J and finally Ξ is a random measure (see Lemma 5.16). \blacksquare

According to this proposition, a well-known and popular stochastic process yields the following example.

5.21 Example (Poisson Process on \mathbb{R}_+) For some $\alpha > 0$ let $(Z_k)_{k \in \mathbb{N}}$ be an i. i. d. sequence of exponentially distributed random variables with parameter α , that is, $Z_k \sim \mathbf{Exp}(\alpha)$ for $k \in \mathbb{N}$. Defining $N = (N_t)_{t \geq \mathbb{R}_+}$ as

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{[0, t]}(\sum_{j=1}^k Z_j) \quad \text{for } t \in \mathbb{R}_+$$

⁷We call a function *increasing* iff it is non-decreasing.

we get an \mathbb{N}_0 -valued stochastic process N . Then, N has *stationary increments*, that is, for $s, t \in \mathbb{R}_+$ with $s \leq t$ we have that $N_t - N_s \sim \mathbf{Poi}(\alpha(t-s))$. Furthermore, N has *independent increments*, i. e., for $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \dots < t_n$ we have that $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ is independent. This process is called homogeneous *Poisson process* with intensity $\alpha > 0$.

Note that the paths of N are increasing and continuous from the right. Hence, by Proposition 5.20 there is a random measure Ξ on \mathbb{R}_+ with $\Xi([s, t]) = N_t - N_s$ for $s, t \in \mathbb{R}_+$ with $s \leq t$.

Since $(Z_k)_{k \in \mathbb{N}}$ are \mathbf{P} -almost surely pairwise different, N has \mathbf{P} -almost surely jumps of size 1 and, hence, $\Xi \in \mathfrak{N}(\mathfrak{X})$ almost surely w. r. t. \mathbf{P} , that is, Ξ is a simple point process on \mathbb{R}_+ . \blacklozenge

To introduce more general point processes, we need some results on the existence and uniqueness of point process distributions.⁸ Concerning the uniqueness and considering a random measure as a set-indexed stochastic process, we show the following.

5.22 Theorem (Uniqueness of Random Measure Distributions I) Let M and N be two random measures on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, then the following statements are equivalent:

- (a) $M \stackrel{d}{=} N$;
- (b) all finite-dimensional distributions of M and N agree, that is,

$$(M(B_1), \dots, M(B_n)) \stackrel{d}{=} (N(B_1), \dots, N(B_n)) \quad (5.1)$$

for all $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{H}(\mathfrak{X})$;

- (c) all *disjoint* finite-dimensional distributions of M and N agree, i. e.,

$$(M(B_1), \dots, M(B_n)) \stackrel{d}{=} (N(B_1), \dots, N(B_n)) \quad (5.2)$$

for all $n \in \mathbb{N}$ and pairwise disjoint $B_1, \dots, B_n \in \mathcal{H}(\mathfrak{X})$. \blacklozenge

Proof: From (a) we can directly follow (c). From (c) follows (b), since for any $n \in \mathbb{N}$ and any choice $B_1, \dots, B_n \in \mathcal{H}(\mathfrak{X})$ there is a finite partition \mathcal{D} of $\bigcup_{i=1}^n B_i$ such that $B_i = \bigcup_{D \in \mathcal{D}, D \subset B_i} D$ for all $i = 1, \dots, n$. Furthermore, the paths of M and N are additive, thus,

$$\left(\sum_{D \in \mathcal{D}, D \subset B_1} M(D), \dots, \sum_{D \in \mathcal{D}, D \subset B_n} M(D) \right) = (M(B_1), \dots, M(B_n))$$

and the corresponding equality holds for N . That means, $(M(B_1), \dots, M(B_n))$ is a measurable function of $(M(D) : D \in \mathcal{D})$. Since $n \in \mathbb{N}$ and $B_1, \dots, B_n \in \mathcal{H}$ are arbitrarily chosen, and since (c) implies

$$(M(D) : D \in \mathcal{D}) \stackrel{d}{=} (N(D) : D \in \mathcal{D}),$$

⁸We write $X \stackrel{d}{=} Y$ for two equally distributed random elements X and Y .

we have

$$(M(B_1), \dots, M(B_n)) \stackrel{d}{=} (N(B_1), \dots, N(B_n)),$$

that is, (b). Clearly, (a) follows from (b) since (5.1) says that M and N agree on an \cap -stable generator of \mathcal{M} . ■

Note that in the above proposition, we crucially use the additivity of the paths, which then yields the stronger equivalence of (a) and (c). Furthermore, considering the product space representation of $(\mathfrak{M}, \mathcal{M})$ (see Remark 5.10, page 107), we can state a different uniqueness result. To this end, we may call $\text{pr}_F \circ M$ for some $F \in \mathcal{B}(\mathfrak{X})$ the *F-component* of the random measure μ , while the corresponding distribution may be called *marginal distribution on F*.

Note that $\text{pr}_F \circ M$ is an \mathfrak{M}_F -valued random measure, while $M(F)$ is a random variable, that is, an $\overline{\mathbb{R}}_+$ -valued random element. With these preparations we can state the following uniqueness result.

5.23 Theorem (Uniqueness of Random Measure Distributions II) Let M and N be two random measures on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, then the following statements are equivalent:

- (a) $M \stackrel{d}{=} N$;
- (b) M and N have the same marginal distributions on sets of an increasing sequence of bounded sets, that is, for a sequence $(G_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ with $G_n \uparrow \mathfrak{X}$ for $n \rightarrow \infty$ we have $\text{pr}_{G_n}(M) \stackrel{d}{=} \text{pr}_{G_n}(N)$ for all $n \in \mathbb{N}$;
- (c) M and N have the same marginal distributions on any bounded set, that is, for all $F \in \mathcal{H}$ we have $\text{pr}_F(M) \stackrel{d}{=} \text{pr}_F(N)$. ◆

Proof: First, (b) follows immediately from (a). Second, (c) follows from (b), since for any $F \in \mathcal{H}$ we have $F \subset G_k$ for some k large enough. Furthermore, we have $\text{pr}_F = \text{pr}_F^{G_k} \circ \text{pr}_{G_k}$ and thus $\text{pr}_F(M) \stackrel{d}{=} \text{pr}_F(N)$. Finally, (a) follows from (c), since $\bigcup_{F \in \mathcal{H}} \text{pr}_F^{-1}(\mathcal{N}_F)$ is an \cap -stable generator of $\mathcal{N}(\mathfrak{X})$. ■

In addition to the uniqueness, we are also interested in an existence result. To this end, we recall that according to Theorem 5.23 a point process distribution is uniquely determined by the family of all marginal distributions on bounded subsets of \mathfrak{X} . Now we are looking for a point process distribution with prescribed marginal distributions. Obviously, such a family of distributions has to be consistent as follows.

5.24 Definition A family $(P_H)_{H \in \mathcal{H}}$ of point process distributions is called *projective* iff for every $G, H \in \mathcal{H}$ with $G \subset H$ we have $P_H^{\text{pr}_G^H} = P_G$. ◆

Introducing distributions by preassigned marginal distributions is due to Kolmogorov (1933). The following proposition concerns the case of random measure distributions.

5.25 Proposition (Kolmogorov Construction) For any projective family $(P_H)_{H \in \mathcal{H}}$ of random measure distributions there is a unique distribution $P_{\mathfrak{X}}$ on $(\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ such that

$$P_{\mathfrak{X}}^{\text{pr}H} = P_H \quad (5.3)$$

holds for every $H \in \mathcal{H}$. \blacklozenge

Proof: Let $\mathcal{D} \subset \mathcal{H}$ be a countable partition of \mathfrak{X} and let φ be the measurable isomorphism corresponding to

$$(\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X})) \cong_{\varphi} \left(\prod_{D \in \mathcal{D}} \mathfrak{M}_D(\mathfrak{X}), \otimes_{D \in \mathcal{D}} \mathcal{M}_D(\mathfrak{X}) \right),$$

see Remark 5.10. Then, a projective family of random measure distributions is given by

$$(P_{\bigcup \mathcal{E}}^{\varphi})_{\mathcal{E} \subset \mathcal{D}, |\mathcal{E}| < \infty}$$

on the corresponding product spaces

$$\left(\prod_{E \in \mathcal{E}} \mathfrak{M}_E(\mathfrak{X}), \otimes_{E \in \mathcal{E}} \mathcal{M}_E(\mathfrak{X}) \right)$$

for finite $\mathcal{E} \subset \mathcal{D}$. Hence, by the Kolmogorov existence theorem (see for example Bauer (1991, Satz 35.3, page 307), see also Proposition 5.12 for the Polishness of the component spaces $(\mathfrak{M}_E(\mathfrak{X}), \mathcal{M}_E(\mathfrak{X}))$ for $E \in \mathcal{E}$), we obtain that there is a unique distribution $\tilde{P}_{\mathcal{D}}$ on $(\prod_{D \in \mathcal{D}} \mathfrak{M}_D(\mathfrak{X}), \otimes_{D \in \mathcal{D}} \mathcal{M}_D(\mathfrak{X}))$ and, consequently, a unique distribution on $(\mathfrak{M}(\mathfrak{X}), \mathcal{M}(\mathfrak{X}))$ via $P_{\mathcal{D}} := \tilde{P}_{\mathcal{D}}^{\varphi^{-1}}$.

Furthermore, $P_{\mathcal{D}}$ is independent of the initially chosen partition \mathcal{D} by the following argument: Let $\mathcal{D}_1, \mathcal{D}_2 \subset \mathcal{H}$ be two countable partitions of \mathfrak{X} with corresponding distributions P_1 and P_2 . By defining

$$\mathcal{D} := \{D_1 \cap D_2 : D_i \in \mathcal{D}_i \text{ for } i = 1, 2\},$$

we get a countable partition of \mathfrak{X} that is nested in \mathcal{D}_1 and \mathcal{D}_2 . Since $(P_B)_{B \in \mathcal{H}}$ is projective and according to the compatibility of the projections (see Remark 5.11, page 107), we have

$$(P_1^{\text{pr}D})_{D \in \mathcal{D}} = (P_D)_{D \in \mathcal{D}} = (P_2^{\text{pr}D})_{D \in \mathcal{D}}$$

and thus $P_1 = P_2$. Finally, $P := P_{\mathcal{D}}$ as defined above also fulfills (5.3). \blacksquare

Proposition 5.25 allows us to define point process distributions on an unbounded state space \mathfrak{X} via a projective family of marginal distributions on bounded subsets of \mathfrak{X} . This will be very helpful to show the existence of Poisson point process distributions generally even on unbounded state spaces.

We get another formulation of Proposition 5.25 when a distribution P_H on $(\mathfrak{N}_H, \mathcal{N}_H)$ is interpreted as a distribution Q_H on $(\mathfrak{N}, \mathcal{F}_H)$, where

$\mathcal{F}_H := \text{pr}_H^{-1}(\mathcal{N}_H)$. Then, if $(P_H)_{H \in \mathcal{H}}$ is a projectiv family, the corresponding family $(Q_H)_{H \in \mathcal{H}}$ is *consistent* in the sense, that for every choice $G, H \in \mathcal{H}$ with $G \subset H$ we have $Q_H|_{\mathcal{F}_G} = Q_G$. In this framework Proposition 5.25 says:

5.26 Corollary For any consistent family $(Q_H)_{H \in \mathcal{H}}$, there is a unique distribution Q on \mathfrak{N} such that $Q|_{\mathcal{F}_H} = Q_H$ for all $H \in \mathcal{H}$. \blacklozenge

Poisson Point Processes

Motivated by Example 5.21 we may define a general Poisson point process as follows.

5.27 Definition Let $\lambda \in \mathfrak{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. A point process Ξ on \mathfrak{X} is called a *Poisson point process* with parameter measure λ iff

- (a) Ξ has Poisson marginal distributions, that is, $\Xi(B) \sim \mathbf{Poi}(\lambda(B))$ for all $B \in \mathcal{H}(\mathfrak{X})$;
- (b) Ξ has *independent increments*, that is, for $n \in \mathbb{N}$ and for pairwise disjoint $B_1, \dots, B_n \in \mathcal{H}(\mathfrak{X})$ the random variables $\Xi(B_1), \dots, \Xi(B_n)$ are independent. \blacklozenge

If $\mathfrak{X} \subset \mathbb{R}^k$ and $\lambda = a \cdot \lambda^k$, then Ξ is called *homogeneous*.

A Poisson point process is a well-defined object according to the following theorem.

5.28 Theorem For every $\lambda \in \mathfrak{M}(\mathfrak{X})$ there is a Poisson point process with unique distribution, which we denote by $\mathbf{Pop} = \mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \lambda)$.⁹ \blacklozenge

Proof: If $\lambda = \mathbf{0}$, then $\Xi \equiv \mathbf{0}$ trivially fulfills (a) and (b) of the definition above. Thus, we assume $\lambda \neq \mathbf{0}$. Following Proposition 5.25 we construct a Poisson point process P_B on a bounded set $B \in \mathcal{H}(\mathfrak{X})$ and show that $(P_B)_{B \in \mathcal{H}}$ is a projective family.

Construction on \mathfrak{X} with $\lambda(\mathfrak{X}) < \infty$: Let N, X_1, X_2, \dots be independent random variables such that $N \sim \mathbf{Poi}(\lambda(\mathfrak{X}))$ and

$$X_k \sim \frac{1}{\lambda(\mathfrak{X})} \lambda(\cdot) \quad \text{for } k \in \mathbb{N}.$$

We set

$$\Xi := \sum_{i=1}^N \delta_{X_i} \tag{5.4}$$

and show that Ξ has (a) Poisson marginals and (b) independent increments by the following computation. Let $B \in \mathcal{H}$ and let $k, l \in \mathbb{Z}_+$. Then we

⁹We will just write \mathbf{Pop} if no clarification concerning the underlying measure state space is needed. Furthermore, we also write $\mathbf{Pop}_{\mathfrak{X}}$ and \mathbf{Pop}_{λ} to emphasize only a certain aspect of the measure space if it does not cause confusion.

have¹⁰

$$\begin{aligned} \mathbf{P}(\Xi(B) = k, \Xi(B^c) = l) &= \underbrace{\mathbf{Poi}(\lambda(\mathfrak{X}))(\{k+l\})}_{=\exp(-\lambda(\mathfrak{X}))\frac{\lambda(\mathfrak{X})^{k+l}}{(k+l)!}} \underbrace{\binom{k+l}{k}}_{=\frac{(k+l)!}{k!l!}} \left(\frac{\lambda(B)}{\lambda(\mathfrak{X})}\right)^k \left(\frac{\lambda(B^c)}{\lambda(\mathfrak{X})}\right)^l \\ &= \exp(-\lambda(B))\frac{\lambda(B)^k}{k!} \cdot \exp(-\lambda(B^c))\frac{\lambda(B^c)^l}{l!}. \end{aligned}$$

Hence, $\Xi(B) \sim \mathbf{Poi}(\lambda(B))$ and $\Xi(B)$ is independent of $\Xi(B^c)$. Note that $B \in \mathcal{H}$ was arbitrary. Proposition 5.10 and induction yields that for every $n \in \mathbb{N}$ and pairwise disjoint $B_1, \dots, B_n \in \mathcal{H}$ we have that $\Xi(B_1), \dots, \Xi(B_n)$ are independent. Thus, Ξ is a Poisson process with parameter measure λ .

Extension to the case $\lambda(\mathfrak{X}) = \infty$ and uniqueness: For every bounded set $B \in \mathcal{H}$, we have $\lambda(B) < \infty$, since $\lambda \in \mathfrak{M}(\mathfrak{X})$. We denote the corresponding construction (5.4) with parameter measure $\lambda|_B$ by Ξ_B . Defining $P_B := \mathbf{P}(\Xi_B \in \cdot)$ for every $B \in \mathcal{H}$ yields a projective family $(P_B)_{B \in \mathcal{H}}$ by the following argument: Let $B, C \in \mathcal{H}$ such that $C \subset B$. Then:¹¹

$$P_B^{\text{pr}^B} = \left((P_C \otimes P_{B \setminus C})^{\varphi^{-1}} \right)^{\text{pr}^B} = (P_C \otimes P_{B \setminus C})^{\text{pr}^1} = P_C,$$

where we use the independence of $\Xi_B(C)$ and $\Xi_B(B \setminus C)$ in the first equality. Finally, the claim follows by Proposition 5.25. \blacksquare

5.29 Proposition (Simple Poisson Point Processes) A Poisson point process Ξ with parameter measure $\lambda \in \mathfrak{M}(\mathfrak{X})$ is simple if and only if λ is *diffuse*, that is, $\lambda(\{x\}) = 0$ for all $x \in \mathfrak{X}$. \blacklozenge

Proof: See Schneider & Weil (2000, Lemma 3.2.1, page 72) \blacksquare

To model dependences, we consider densities w. r. t. the distribution of a (homogeneous) Poisson point process. This seems reasonable due to two properties of that distribution, one practical and one technical. That is, (1) the homogeneous Poisson point process serves as a model for complete spatial randomness. Hence, densities and the corresponding distributions are interpretable. (2) The Poisson point process distribution has mass all over the space $\mathfrak{N}(\mathfrak{X})$.

While the first is clear due to the definition, we show the latter in Proposition 5.31 (see below) with the help of the following property.

¹⁰Given the event “observing $k+l$ points in \mathfrak{X} ”, which occurs with probability $\mathbf{Poi}(\lambda(\mathfrak{X}))(\{k+l\})$, we note that $\Xi(B)$ is binomially distributed with probability of success $p = \lambda(B)/\lambda(\mathfrak{X})$ and length $k+l$.

¹¹Here we use the isomorphism φ between $(\mathfrak{N}, \mathcal{N})$ and the corresponding product space given in the proof of Proposition 5.10.

5.30 Proposition (Void Sets of Poisson Point Processes) Let $\lambda \in \mathfrak{M}(\mathfrak{X})$ and let $\Xi \sim \mathbf{Pop}(\lambda)$, then

$$\mathbf{P}(\Xi(B) = 0) = \exp(-\lambda(B))$$

for all $B \in \mathcal{H}(\mathfrak{X})$. ◆

Proof: The statement is obvious due to $\Xi(B) \sim \mathbf{Poi}(\lambda(B))$ for $B \in \mathcal{H}$, see Definition 5.27. ■

5.31 Proposition (Full Support) Let $\lambda \in \mathfrak{M}(\mathfrak{X})$ be diffuse and such that $\text{supp}_{\mathcal{T}} \lambda = \mathfrak{X}$. Then we have $\text{supp}_{\mathcal{T}_{\mathbf{F}}} \mathbf{Pop}(\lambda) = \mathfrak{N}(\mathfrak{X})$. ◆

Proof: To show the claim, we first recall from Proposition 5.14 that $\mathcal{N}(\mathfrak{X})$ is generated by the Fell topology on $\mathfrak{N}(\mathfrak{X})$, that is, by

$$\mathcal{T}_{\mathbf{F}} = \tau(\{\xi : \xi(G) > 0\}, \{\xi : \xi(C) = 0\} : G \in \mathcal{H} \text{ open}, C \in \mathcal{H} \text{ compact}).$$

Then, the claim follows by $\mathbf{Pop}_{\lambda}(U) > 0$ for all $U \in \mathcal{T}_{\mathbf{F}}$ which we are going to show now.

According to Proposition 5.30, we have for every compact set $C \in \mathcal{H}(\mathfrak{X})$

$$\mathbf{Pop}_{\lambda}(\{\xi : \xi(C) = 0\}) = \exp(-\lambda(C)) > 0.$$

Furthermore, we have for all open $G \in \mathcal{H}(\mathfrak{X})$

$$\mathbf{Pop}_{\lambda}(\{\xi : \xi(G) > 0\}) = 1 - \exp(-\lambda(G)) > 0,$$

since $\text{supp} \lambda = \mathfrak{X}$. Hence, any element of the subbasis has positive mass w. r. t. \mathbf{Pop}_{λ} . Let $U^C := \{\xi : \xi(C) = 0\}$ and $U_G := \{\xi : \xi(G) > 0\}$ denote the corresponding sets of the subbasis. Then we have for compact subsets C and D of \mathfrak{X} that

$$\mathbf{Pop}_{\lambda}(U^C \cap U^D) = \mathbf{Pop}_{\lambda}(U^{C \cup D}) = \exp(-\lambda(C \cup D)) > 0$$

and, for two open $G, F \in \mathcal{H}(\mathfrak{X})$ with $G \cap F \neq \emptyset$, we have that

$$\mathbf{Pop}_{\lambda}(U_G \cap U_F) \geq \mathbf{Pop}_{\lambda}(U_{G \cap F}) = 1 - \exp(-\lambda(G \cap F)) > 0$$

as well. Finally, for open $G \in \mathcal{H}$ and compact $C \in \mathcal{H}$ with $G \not\subset C$, we have that $G \cap C^c \neq \emptyset$ is open and

$$\mathbf{Pop}_{\lambda}(U_G \cap U^C) = \mathbf{Pop}_{\lambda}(U_{G \setminus C} \cap U^C) = (1 - \exp(-\lambda(G \setminus C))) \cdot \exp(-\lambda(C)) > 0.$$

Hence, any non-empty finite intersection of sets of the subbasis has positive mass w. r. t. \mathbf{Pop} . Thus, $\mathbf{Pop}(\lambda)$ assigns positive mass to any non-empty $\mathcal{T}_{\mathbf{F}}$ -open subset of \mathfrak{N} . That means, $\text{supp}_{\mathcal{T}_{\mathbf{F}}} \mathbf{Pop}_{\lambda} = \mathfrak{N}(\mathfrak{X})$. ■

The full support property of the Poisson point process distribution \mathbf{Pop}_λ is at least one aspect that may be used to justify the introduction of Gibbs point process distributions in Chapter 6 through densities w. r. t. \mathbf{Pop}_λ . Here, we at least point out that our topological considerations seem to be closely related to the so-called Σ -condition of a distribution $P \in \text{Prob}(\mathfrak{N}, \mathcal{N})$, which means that each $F \in \mathcal{N}$ with $P(F) > 0$ implies that $P(\{\xi : \xi(G) = 0\} \cap F) > 0$ for all $G \in \mathcal{H}$. This condition was studied in the literature in particular for the converse of the above-mentioned question, that is, which point process distribution has a density w. r. t. \mathbf{Pop}_λ (see e. g. Rauchenschwandtner (1978); Matthes *et al.* (1979); Glötzl (1980)). The present thesis does not discuss these relations and questions.

Random Signed Measures and Brownian Motion

So far, the classical Poisson process on \mathbb{R}_+ was shown to be a point process and, consequently, was extended to the notion of a Poisson point process on general state spaces. Considering another important stochastic process, Brownian motion, immediately leads to the subsequent question: In what sense is Brownian motion a random signed measure? In fact, considering the obvious extension of the random measure concept to a random signed measure is not sufficient. We therefore close this section with a brief discussion of this issue. This also points out another peculiarity of random measure theory.

Denote by $\mathfrak{S}(\mathfrak{X}) = \mathfrak{S}(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ the set of all finite signed measures on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$, that is, the set of all set-functions $\mu : \mathcal{A} \rightarrow \mathbb{R}$ that are the difference of two finite measures on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$. Then, with respect to the definition of random measures (see Definition 5.15), we may define:

5.32 Definition A $\mathfrak{S}(\mathfrak{X})$ -random element Λ on $(\Omega, \mathcal{F}, \mathbf{P})$ is called *random signed measure* on \mathfrak{X} iff $\Lambda(B)$ is a random variable for all $B \in \mathfrak{B}(\mathfrak{X})$. \blacklozenge

Let $T := [0, L]$ for some $L > 0$. We recall from the literature that a *Brownian motion* (with finite time horizon) is a stochastic process $W = (W_t)_{t \in T}$ with \mathbf{P} -almost surely continuous paths which has independent and stationary increments such that $W_t - W_s \sim \mathbf{N}(0, t-s)$ for all $s, t \in T, s \leq t$.¹²

5.33 Example (Discretely Sampled Brownian Motion) Let $W = (W_t)_{t \in T}$ be a Brownian motion with T as above. Let $n \in \mathbb{N}$ and let $t \in T^n$ be an increasing vector of n indices. Define

$$X_s(\omega) := \sum_{j=1}^{n-1} W_{t_j}(\omega) \mathbb{1}_{[t_j, t_{j+1}[}(s) \quad \text{for } s \in T, \omega \in \Omega,$$

¹²We set $\mathbf{N}(0, 0) := \delta_0$.

then $s \mapsto X_s(\omega)$ is continuous from the right and, furthermore, of bounded variation. Therefore, $s \mapsto X_s(\omega)$ is the difference of two increasing functions and by Proposition 5.20 there are random measures Ξ_t^+ and Ξ_t^- such that

$$\Xi_t^+([0, s]) - \Xi_t^-([0, s]) = X_s \quad \text{for } s \in T.$$

By $\Lambda_t := \Xi_t^+ - \Xi_t^-$ we have a random signed measure and, furthermore, $\Lambda_t(\cdot, [0, t_j]) = W_{t_j}$ for $j = 1, \dots, n$. \blacklozenge

The discretization in Example 5.33 is necessary, since \mathbf{P} -almost all paths of a Brownian motion $W = (W_t)_{t \in T}$ are of unbounded variation (see e. g. Kallenberg, 2002, Corollary 13.10). Hence, W does not determine a random signed measure in the sense of Definition 5.32, because we do not have σ -additivity ω -wise. However, we have a σ -additivity in L_2 .

5.34 Remark Consider a Brownian motion $W = (W_t)_{t \in T}$ as a random element with values in the Hilbert space $L_2(T, \mathcal{B}(T), \lambda)$. Then, for all $\varphi \in L_2(T, \mathcal{B}(T), \lambda)$ the random variable $\langle W, \varphi \rangle_{L_2}$ on $(\Omega, \mathcal{F}, \mathbf{P})$ is a one-dimensional Gaussian random variable with

$$\langle W, \varphi \rangle_{L_2} \sim \mathbf{N}(0, \int \varphi^2(s) \lambda(ds))$$

and for $\varphi, \psi \in L_2(T, \mathcal{B}(T), \lambda)$ we have

$$\text{Cov}(\langle W, \varphi \rangle_{L_2}, \langle W, \psi \rangle_{L_2}) = \int \varphi(s) \psi(s) \lambda(ds)$$

(see e. g. Bogachev, 1998). Defining $X_A := \langle W, \mathbb{1}_A \rangle_{L_2}$ for $A \in \mathcal{B}(T)$, we have first of all that $X_\emptyset = 0$. Second, according to the above, we also have $X_A \sim \mathbf{N}(0, \lambda(A))$ and $\text{Cov}(X_A, X_B) = \lambda(A \cap B)$ for $A, B \in \mathcal{B}(T)$. With these properties one can simply compute for $A, B \in \mathcal{B}(T)$ with $A \cap B = \emptyset$ that

$$\mathbf{E}_{\mathbf{P}}(X_{A \cup B} - (X_A + X_B))^2 = 0,$$

that is, $X_{A \cup B} = X_A + X_B$ in $L_2(\Omega, \mathcal{F}, \mathbf{P})$. Similarly, one obtains for $A_n, A \in \mathcal{B}(T)$ for $n \in \mathbb{N}$ with $A_n \uparrow A$ for $n \rightarrow \infty$ that $X_{A_n} \rightarrow X_A$ in $L_2(\Omega, \mathcal{F}, \mathbf{P})$ for $n \rightarrow \infty$. Hence, for a pairwise disjoint sequence $(A_n)_{n \in \mathbb{N}} \in \mathcal{B}(T)^{\mathbb{N}}$ we have

$$X_{\bigcup_{n \in \mathbb{N}} A_n} = \sum_{n \in \mathbb{N}} X_{A_n} \quad \text{in } L_2(\Omega, \mathcal{F}, \mathbf{P}).$$

Consequently, the map $A \mapsto X_A$ behaves σ -additively in $L_2(\Omega, \mathcal{F}, \mathbf{P})$. \blacklozenge

5.3 Intensity of Random Measures

Important characteristics of random measures are given by their moment measures. Of special concern will be the intensity measure.

5.35 Definition For a random measure M on a measure space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \mu)$ we define

$$\mathbf{M}_M(A) := \mathbf{E}M(A) := \int M(A) d\mathbf{P} \quad \text{for } A \in \mathcal{B}(\mathfrak{X})$$

and call \mathbf{M}_M *intensity measure* of M , which is a measure on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. A density ρ of \mathbf{M}_M w. r. t. μ is called *intensity function* of \mathbf{M}_M w. r. t. μ . \blacklozenge

5.36 Example (Intensity Measure of Poisson point processes) Let Ξ be a Poisson point process on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ with parameter measure $\lambda \in \mathfrak{M}(\mathfrak{X})$. Then, $\mathbf{M}_\Xi = \lambda$. \blacklozenge

From the definition of the intensity measure we directly get the following important relation.

5.37 Theorem (Campbell) Let M be a random measure on $(\mathfrak{X}, \mathcal{A})$ and let $f : (\mathfrak{X}, \mathcal{B}(\mathfrak{X})) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ be measurable, then

$$\int \int f(x) M(\omega, dx) \mathbf{P}(d\omega) = \int f(x) \mathbf{M}_M(dx).$$

\blacklozenge

Proof: By a standard lifting argument, that is, deducing the statement from indicators to simple and, finally, to non-negative functions, yields the statement. \blacksquare

Campbell's Theorem has a common concern with Fubini's Theorem, which is to exchange integrals. However, from a technical point of view, they are able to consider different situations.

5.4 Conditional Point Processes

In view of our probabilistic and statistical concerns about point processes, we would like to describe and estimate their distributions. This naturally involves the dependence of points or, more generally speaking, of marginal point patterns on bounded sets $H \in \mathcal{H}$.

To this end, a promising approach is to consider conditional distributions. That means, we are interested in the local behavior of a point process Ξ given a neighborhood. Mathematically speaking, we are interested in the distribution of Ξ on an arbitrary set $H \in \mathcal{H}$, given a realization of Ξ on the outside H^c . This so-called *exterior conditioning* is in contrast to the *interior conditioning*, where we consider the distribution of Ξ on the outside H^c given the corresponding local behavior of Ξ , that is, a realization of Ξ_H .

Generally speaking, we are therefore interested in the behavior of the pair (Ξ_H, Ξ_{H^c}) and, in particular, in disintegrations of this joint distribution. For this we consider conditional distributions given interior

conditioning of the form $\mathbf{P}(\Xi_{H^c} \in \cdot | \Xi_H)$ and exterior conditioning of the form $\mathbf{P}(\Xi_H \in \cdot | \Xi_{H^c})$.

Campbell Measure

We start by a degenerate case, that is, by considering interior and exterior conditioning of a point process w. r. t. singletons $H = \{x\}$ for some $x \in \mathfrak{X}$. To this end, we define the following measure.

5.38 Definition (Campbell Measure) Let Ξ be a point process on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. The *Campbell measure* of Ξ is the measure on $\mathfrak{X} \times \mathfrak{N}(\mathfrak{X})$ defined through

$$\mathbf{C}_{\Xi}(A \times B) := \int \int \mathbb{1}_{A \times B}(x, \xi) \xi(dx) \mathbf{P}^{\Xi}(d\xi)$$

for $A \in \mathcal{B}(\mathfrak{X})$ and $B \in \mathcal{N}$. \blacklozenge

For fixed $A \in \mathcal{B}(\mathfrak{X})$ and $B \in \mathcal{N}$, those pairs (x, ξ) contribute “mass” to $\mathbf{C}_{\Xi}(A \times B)$ which satisfy $x \in \xi$, $x \in A$, $\xi \in B$. Hence, by \mathbf{C}_{Ξ} we measure the joint occurrence of a point x and a point pattern ξ .

Furthermore, we note that $\mathbf{C}_{\Xi}(\cdot \times \mathfrak{N}(\mathfrak{X})) = \mathbf{M}_{\Xi}$ and $\mathbf{C}_{\Xi}(\mathfrak{X} \times \cdot) = \mathbf{P}^{\Xi}$, that is, the first and the second marginal distribution of \mathbf{C}_{Ξ} is the intensity measure of Ξ and the point process distribution, respectively. In what follows, we assume that \mathbf{M}_{Ξ} is σ -finite.

We are now going to disintegrate \mathbf{C}_{Ξ} with respect to either of its marginal measures. This corresponds to the earlier mentioned interior and exterior conditioning regarding singletons $H = \{x\}$ for $x \in \mathfrak{X}$ and leads to the so-called *Palm* and *Papangelou kernels*, respectively.

5.39 Remark and Definition (Palm Kernel) Let Ξ be a point process on \mathfrak{X} with σ -finite intensity measure \mathbf{M}_{Ξ} . Then, there is a Markov kernel \mathbf{Pa} from $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ to $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ such that

$$\mathbf{C}_{\Xi}(A \times B) = \mathbf{M}_{\Xi} \otimes \mathbf{Pa}(A \times B) = \int_A \mathbf{Pa}(x, B) \mathbf{M}_{\Xi}(dx)$$

for $A \in \mathcal{B}(\mathfrak{X})$ and $B \in \mathcal{N}$ (see e. g. Kallenberg, 2017, Theorem 1.23, page 37). The Kernel \mathbf{Pa} is called *Palm kernel*, the distribution $\mathbf{Pa}_x := \mathbf{Pa}(x, \cdot)$ for $x \in \mathfrak{X}$ is called *Palm distribution* of Ξ . \blacklozenge

According to the definition of the Palm kernel \mathbf{Pa} through disintegration \mathbf{C}_{Ξ} w. r. t. the intensity measure \mathbf{M}_{Ξ} , the distribution \mathbf{Pa}_x for some $x \in \mathfrak{X}$ can heuristically be interpreted as the distribution of Ξ given the event $\{x \in \Xi\}$. This is consistent with the following observation.

5.40 Proposition Let \mathbf{Pa}_x^{Ξ} be a Palm kernel to a point process Ξ on \mathfrak{X} . Then

$$\mathbf{Pa}_x^{\Xi}(\{\xi : \xi(\{x\}) > 1\}) = 1$$

for \mathbf{M}_{Ξ} -almost all $x \in \mathfrak{X}$. \blacklozenge

Proof: See Kallenberg (2017, Lemma 6.2, page 216). ■

According to this proposition, \mathbf{Pa}_x has a deterministic unit mass at x . We would like to omit this mass and focus on the more interesting random part. This leads to the reduced Campbell Measure.

5.41 Definition (Reduced Campbell Measure) Let Ξ be a point process on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$. The *reduced Campbell measure* of Ξ is the measure on $\mathfrak{X} \times \mathfrak{N}(\mathfrak{X})$ defined through

$$\mathbf{C}_{\Xi}^{\downarrow}(A \times B) := \int \int \mathbb{1}_{A \times B}(x, \xi - \delta_x) \xi(dx) \mathbf{P}^{\Xi}(d\xi)$$

for $A \in \mathcal{B}(\mathfrak{X})$ and $B \in \mathcal{N}$. ◆

By that, we can now introduce the so-called reduced Palm kernel by disintegration of the reduced Campbell measure w. r. t. the intensity measure in the same way as in Remark 5.39.

5.42 Remark and Definition (Reduced Palm Distribution) Let Ξ be a point process on \mathfrak{X} with σ -finite intensity measure \mathbf{M}_{Ξ} . Then, there is a Markov kernel \mathbf{Pa}^{\downarrow} from $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ to $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ such that

$$\mathbf{C}_{\Xi}^{\downarrow}(A \times B) = \mathbf{M}_{\Xi} \otimes \mathbf{Pa}^{\downarrow}(A \times B) = \int_A \mathbf{Pa}^{\downarrow}(x, B) \mathbf{M}_{\Xi}(dx)$$

for $A \in \mathcal{B}(\mathfrak{X})$ and $B \in \mathcal{N}$. The Kernel \mathbf{Pa}^{\downarrow} is called *reduced Palm kernel*, the distribution $\mathbf{Pa}_x^{\downarrow} := \mathbf{Pa}^{\downarrow}(x, \cdot)$ for $x \in \mathfrak{X}$ is called *reduced Palm distribution* of the point process Ξ . ◆

We add a remark on reduced Campbell measures and Palm distributions for stationary point processes.

5.43 Remark and Definition (Palm Distribution and Stationarity) Consider the case $\mathfrak{X} = \mathbb{R}^d$ and introduce the *translation* $\mathbf{t}_u : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $x \mapsto x + u$ for $u \in \mathbb{R}^d$. A point process Ξ is called *stationary* iff $\Xi \circ \mathbf{t}_u^{-1} \stackrel{d}{=} \Xi$ for all $u \in \mathbb{R}^d$. Let Ξ be a stationary point process. Then, one can show (see for example Kallenberg (2017))

- (a) for the intensity measure that $\mathbf{M}_{\Xi}(B) = c(\mathbf{P}^{\Xi}) \cdot \lambda^d(B)$ for $B \in \mathcal{B}(\mathbb{R}^d)$;
- (b) there is a distribution $\mathbf{Pa}^{\downarrow 0}$ on $(\mathfrak{N}(\mathbb{R}^d), \mathcal{N}(\mathbb{R}^d))$ such that reduced Campbell measure satisfies

$$\mathbf{C}^{\downarrow}(A \times B) = c(\mathbf{P}^{\Xi}) \cdot \int \int \mathbb{1}_A(x) \mathbb{1}_B(\xi \circ \mathbf{t}_x) \mathbf{Pa}^{\downarrow 0}(d\xi) \lambda^d(dx).$$

These facts will be used in Chapter 6 and 7 to formulate an identity and an unbiased estimating equation, respectively. ◆

We now come to the exterior conditioning. To this end, we consider disintegration of the reduced Campbell measure w. r. t. \mathbf{P}^{Ξ} .

5.44 Remark and Definition (Papangelou Kernel) Let Ξ be a point process on \mathfrak{X} and let $\mathbf{C}_{\Xi}^{\downarrow}(A \times \cdot) \ll \mathbf{P}^{\Xi}$ for all relatively compact Borel sets $A \in \mathfrak{B}(\mathfrak{X})$. Then, there is a Markov kernel Λ from $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ to $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ such that

$$\mathbf{C}_{\Xi}^{\downarrow}(A \times B) = \mathbf{P}^{\Xi} \otimes \Lambda(A \times B) = \int_B \Lambda(\xi, A) \mathbf{P}^{\Xi}(d\xi)$$

for $A \in \mathfrak{B}(\mathfrak{X})$ and $B \in \mathcal{N}$. The kernel Λ is called *Papangelou kernel* of Ξ . \blacklozenge

Consider a reference measure μ on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ and, furthermore, a point process Ξ with a Papangelou kernel $\Lambda(\xi, \cdot)$ which is absolutely continuous w. r. t. μ . Then, $\Lambda(\xi, \cdot)$ has a density $\lambda(\cdot|\xi)$ w. r. t. μ which is called *Papangelou conditional intensity* of Ξ .

5.45 Example (Poisson Point Process) Consider a Poisson point process $\Xi \sim \mathbf{Pop}(\nu)$ with parameter measure $\nu \in \mathfrak{M}(\mathfrak{X})$ that has a density φ w. r. t. a reference measure μ on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$. Then, the Papangelou conditional intensity of Ξ given $\xi \in \mathfrak{N}(\mathfrak{X})$ (w. r. t. μ) is $\lambda(\cdot|\xi) = \varphi$. \blacklozenge

Similarly to Campbell's Theorem we get by the definition of the reduced Campbell measure and the disintegration w. r. t. \mathbf{P}^{Ξ} the so-called *Georgii–Nguyen–Zessin (GNZ) equation*.

5.46 Theorem (GNZ Equation) Let Ξ be a point process on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$ with Papangelou conditional intensity λ and let h be a measurable map from $(\mathfrak{X} \times \mathfrak{N}(\mathfrak{X}), \mathfrak{B}(\mathfrak{X}) \otimes \mathcal{N}(\mathfrak{X}))$ to $(\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+))$. Then we have

$$\int_{\mathfrak{N}} \int_{\mathfrak{X}} h(x, \xi - \delta_x) \xi(dx) \mathbf{P}^{\Xi}(d\xi) = \int_{\mathfrak{X}} \int_{\mathfrak{N}} h(x, \xi) \lambda(x|\xi) \mathbf{P}^{\Xi}(d\xi) \mu(dx).$$

\blacklozenge

Proof: The statement follows by the definition of the reduced Campbell measure, standard lifting arguments (from indicators to simple functions and then to non-negative measurable functions) and Remark and Definition 5.44. \blacksquare

The GNZ equation will be useful to show the unbiasedness of certain estimating functions, see Remark 7.8. For the Gibbs point processes (see next Chapter), we will present another interesting equation which is also linked to the reduced Campbell measure. This equation is called *variational equation* given by Baddeley & Dereudre (2013, Proposition 1) and will be used in Chapter 7 to define an unbiased estimating function.

Compound Campbell Measure

So far, the conditional distribution $\mathbf{P}(\Xi_H \in \cdot | \Xi_{H^c})$ was considered for the degenerate case $H = \{x\}$ for some $x \in \mathfrak{X}$ in terms of the Papangelou kernel. In fact, one can prove (see Kallenberg, 2017, Chapter 8) that all

conditional distributions of the form $\mathbf{P}(\Xi_H \in \cdot | \Xi_{H^c})$ can be constructed by a single Markov kernel \mathbf{G} from¹³ $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ to $(\mathfrak{N}_f(\mathfrak{X}), \mathcal{N}_f(\mathfrak{X}))$, which is called *Gibbs kernel*. This kernel can again be obtained via disintegration, now from the so-called *compound Campbell measure*. We close this section with some notes on this generalization.

5.47 Definition Let Ξ be a point process on $(\mathfrak{X}, \mathcal{A})$. The *compound Campbell measure* of Ξ is the measure on $\mathfrak{N}_f(\mathfrak{X}) \times \mathfrak{N}(\mathfrak{X})$ defined through

$$\mathbf{Co}(A \times B) := \mathbf{Co}_\Xi(A \times B) := \int \sum_{\eta \leq \xi} \mathbb{1}_{A \times B}(\eta, \xi - \eta) \mathbf{P}^\Xi(d\xi)$$

for $A \in \mathcal{N}_f$ and $B \in \mathcal{N}$. ◆

Let $A \in \mathcal{N}_f$ and $B \in \mathcal{N}$. Then $\mathbf{Co}(A \times B)$ represents the total mass of the partition pairs (η, ζ) of ξ , which means $\eta + \zeta = \xi$, such that $(\eta, \zeta) \in A \times B$.

The joint distribution of (Ξ_H, Ξ_{H^c}) for any $H \in \mathcal{H}$ can be represented through the compound Campbell measure. This is the statement of the following theorem.

5.48 Theorem Let Ξ be a point process and let $H \in \mathcal{H}$. Then for $A \in \mathcal{N}_f$ and $B \in \mathcal{N}$ we have

$$\mathbf{P}^{(\Xi_H, \Xi_{H^c})}(A \times B) = \int \mathbb{1}_{\mathfrak{N}_{f,H}}(\xi) \mathbb{1}_{\mathfrak{N}_{H^c}}(\eta) \mathbb{1}_{A \times B}(\xi, \eta) \mathbf{Co}_\Xi(d(\xi, \eta)).$$

◆

Proof: See Last (1990, page 130). ■

5.49 Remark and Definition (Gibbs Kernel) We assume $\mathbf{Co}(\mathfrak{N} \times \cdot) \ll \mathbf{P}^\Xi$. Then there is a kernel \mathbf{G} from $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ to $(\mathfrak{N}_f(\mathfrak{X}), \mathcal{N}_f(\mathfrak{X}))$ such that

$$\mathbf{Co}(A \times B) = \int_B \mathbf{G}(\xi, A) \mathbf{P}^\Xi(d\xi)$$

for $A \in \mathcal{N}_f$ and $B \in \mathcal{N}$. The kernel \mathbf{G} is called *Gibbs kernel*. ◆

Finally, we have the following formula.

5.50 Theorem Let \mathbf{G} be the Gibbs kernel of a point process Ξ . Then, we obtain for any $H \in \mathcal{H}$

$$\mathbf{P}(\Xi_H \in B | \Xi_{H^c}) = \frac{\mathbf{G}(\Xi, B \cap \mathfrak{N}_H)}{\mathbf{G}(\Xi, \mathfrak{N}_H)} \quad \text{for all } B \in \mathcal{N}_f$$

which holds almost surely on $\{\Xi(H) = 0\}$. ◆

¹³We denote by $\mathfrak{N}_f(\mathfrak{X})$ the set of all *finite* point patterns $\xi \in \mathfrak{N}(\mathfrak{X})$, that is, $\xi(\mathfrak{X}) < \infty$. Furthermore, we will write $\mathfrak{N}_{f,B} = \mathfrak{N}_{f,B}(\mathfrak{X})$ for the set of all $\xi \in \mathfrak{N}_f(\mathfrak{X})$ such that $\xi(B^c) = 0$, where $B \in \mathcal{B}(\mathfrak{X})$.

Proof: See Kallenberg (2017, page 314). ■

In the next chapter, we are interested in introducing point processes that exhibit dependences. To this end, we do *not* model point processes through a family of joint distributions (Ξ_H, Ξ_{H^c}) or the compound Campbell measure. Instead and in accordance with the latter result, we will define a point process distribution \mathbf{P}^Ξ through a prespecified *consistent* family of conditional distributions that corresponds to the single Gibbs kernel \mathbf{G} .

CHAPTER 6

Gibbs Point Processes

In practice, we often encounter point processes that exhibit some form of dependence. The observation of a point at a certain position is usually not independent of points that occur in a neighborhood. Arrangements of interacting atoms or molecules or the location of trees in a forest may be mentioned as an example of such point process outcomes.

The purpose of this chapter is to introduce *Gibbs point processes*¹, a class of point processes that account for interactions. Here, the joint distribution of such a process (macroscopic state) is motivated and derived from its local behavior, that is, microscopic states. More precisely, the joint distribution is implicitly defined through a consistent family of conditional distributions given an observation in the complement of bounded regions. Such a family is called (*local specification*). In contrast to Kolmogorov's approach which uses projective families (see Proposition 5.25, page 113), the Gibbsian way has fundamentally different consequences. In general, local specifications do not determine distributions uniquely, such a distribution might not even exist at all.

Specifying the dependence structure of a random field by a suitable family of conditional distributions is an approach due to Dobrushin (1968a,b) and Lanford III & Ruelle (1969). A systematic treatment was then given by Preston (1976), which was later on revised and expanded (see Preston, 2005). Although Preston himself regarded the first presentation as "fairly abstract" (see Preston, 1976, preface) and the second as "not [...] suitable for learning about Gibbs [point processes]" (see Preston, 2005, preface), the present chapter aims at a comprehensive survey of some key issues that are mainly guided by the two mentioned monographs of Preston.²

¹The name is due to Gibbs (1902) who introduced such processes and corresponding distribution, so-called *equilibrium states*, through a maximum entropy principle (see also Nguyen & Zessin (1976)).

²Occasionally, we will also refer to Georgii (2011) who, however, does not consider

The formal definition of a *Gibbs measure*, that is, the distribution of a Gibbs point process, as well as of related objects are given in the first section. There we also highlight some relations to the theory of stochastic processes. In preparation of the discussion regarding existence (Section 6.4) and uniqueness, we take a closer look at some first simple examples and situations in advance (Section 6.2). Two concrete and non-trivial examples of Gibbs point process models are then given through certain specifications in Section 6.3.

Throughout this chapter we consider a locally compact second countable Hausdorff space $(\mathfrak{X}, \mathcal{T})$ and a metric d that induces \mathcal{T} and such that (\mathfrak{X}, d) is boundedly compact.³ We recall from Chapter 5 that $\mathcal{H} = \mathcal{H}(\mathfrak{X})$ denotes the system of bounded Borel measurable subsets of \mathfrak{X} .

6.1 Toward a Definition

We are going to introduce point process distributions on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}, \mathcal{T}))$ by prescribing conditional distributions. To this end, we define for any $H \in \mathcal{H}$ the σ -algebras $\mathcal{F}_H := \text{pr}_H^{-1}(\mathcal{N}_H)$ and $\mathcal{F}^H := \text{pr}_{H^c}^{-1}(\mathcal{N}_{H^c})$, that is, the σ -algebra of events observable from the inside and the σ -algebra of events observable from the outside of H . Corresponding names are used for the filtrations $(\mathcal{F}_H)_{H \in \mathcal{H}}$ and $(\mathcal{F}^H)_{H \in \mathcal{H}}$.

Given a family of Markov kernels $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ such that Γ_H is a kernel from $(\mathfrak{N}, \mathcal{F}^H)$ to $(\mathfrak{N}, \mathcal{N})$ for every $H \in \mathcal{H}$, we are looking for a point process Ξ on \mathfrak{X} , which satisfies for all $A \in \mathcal{N}(\mathfrak{X})$ and $H \in \mathcal{H}(\mathfrak{X})$

$$\mathbf{P}^\Xi(A | \mathcal{F}^H) = \Gamma_H(\cdot, A) \quad \text{almost surely w. r. t. } \mathbf{P}^\Xi. \quad (6.1)$$

By doing this, we simultaneously require Γ to be consistent in a way that is shown in the upcoming remark.

6.1 Remark Let Γ be a family of Markov kernels that fulfill (6.1) and let $G, H \in \mathcal{H}$, with $G \subset H$. Then, for $A \in \mathcal{N}$ we have the so-called *Chapman–Kolmogorov equation*

$$\begin{aligned} \Gamma_H \Gamma_G(\cdot, A) &= \int_{\mathfrak{N}(\mathfrak{X})} \Gamma_G(\eta, A) \Gamma_H(\cdot, d\eta) \\ &= \Gamma_H(\cdot, A) \quad \text{almost surely w. r. t. } \mathbf{P}^\Xi. \end{aligned}$$

Furthermore, Γ_H is *proper* for any $H \in \mathcal{H}$, that is, for every $B \in \mathcal{F}^H$ we have

$$\Gamma_H(\cdot, A \cap B) = \Gamma_H(\cdot, A) \mathbf{1}_B \quad \text{almost surely w. r. t. } \mathbf{P}^\Xi.$$

point processes, but the lattice case.

³Note that any locally compact second countable Hausdorff space $(\mathfrak{X}, \mathcal{T})$ has a compatible metric d such that (\mathfrak{X}, d) is boundedly compact, see Vaughan (1937).

In particular, we then have *normality*, that is,

$$\Gamma_H(\xi, \cdot)|_{\mathcal{F}^H} = \delta_\xi(\cdot)|_{\mathcal{F}^H}$$

for \mathbf{P}^Ξ -almost all $\xi \in \mathfrak{X}$. ◆

Suitable prescriptions are obviously those which fulfill the necessities mentioned in the remark above. Additionally, we point out that these requirements are stated in an almost sure sense w. r. t. the distribution \mathbf{P}^Ξ we are looking for. Such a dependence on \mathbf{P}^Ξ is undesirable, which leads to the following definition.

6.2 Definitions A kernel $K: \mathfrak{N}(\mathfrak{X}) \times \mathcal{N}(\mathfrak{X}) \rightarrow [0, 1]$ is called *quasi-Markov kernel* or *quasi-probability kernel* iff $K(x, \cdot)$ for $x \in \mathfrak{X}$ is a probability measure or the null measure $\mathbf{0}$.

Consider $H \in \mathcal{H}(\mathfrak{X})$. A quasi-Markov kernel K from $(\mathfrak{N}(\mathfrak{X}), \mathcal{F}^H)$ to $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ is called *strict* iff for every $A \in \mathcal{F}_H$ and every $B \in \mathcal{F}^H$ we have $K(\cdot, A \cap B) = K(\cdot, A)\mathbb{1}_B$ on $\mathfrak{N}(\mathfrak{X})$.

Consider $\mathcal{J} \subset \mathcal{H}$. A family of kernels $\Gamma = (\Gamma_H)_{H \in \mathcal{J}}$ is called *specification* w. r. t. $(\mathcal{F}^H)_{H \in \mathcal{J}}$ iff (1) \mathcal{J} contains a sequence $(H_n)_{n \in \mathbb{N}}$ such that $H_n \uparrow \mathfrak{X}$ for $n \rightarrow \infty$, (2) for each $H \in \mathcal{J}$ we have that Γ_H is a strict quasi-Markov kernel from $(\mathfrak{N}(\mathfrak{X}), \mathcal{F}^H)$ to $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$, and (3) for all $G, H \in \mathcal{J}$ with $G \subset H$ we have $\Gamma_H \Gamma_G = \Gamma_H$. ◆

Note that the equations above do not only hold almost surely but on the whole space $\mathfrak{N}(\mathfrak{X})$. Unless otherwise stated we consider $\mathcal{J} = \mathcal{H}$. We are now prepared to give the fundamental definition of this chapter.

6.3 Definition (Gibbs Distribution and Point Process) Let $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ be a specification. A distribution $P \in \text{Prob}(\mathfrak{N})$ is called *Gibbs distribution* w. r. t. Γ iff for all $H \in \mathcal{H}$, $A \in \mathcal{N}$

$$P(A|\mathcal{F}^H) = \Gamma_H(\cdot, A) \quad \text{almost surely w. r. t. } \mathbf{P}^\Xi. \tag{6.2}$$

A point process $\Xi \sim P$ is called *Gibbs point process* w. r. t. Γ . The set of all Gibbs distributions w. r. t. Γ is denoted by **Gibbs**(Γ). ◆

Often, Gibbs distributions are also called *Gibbs measures*, *Gibbs states*, or *Gibbs ensembles*. In addition, we use the phrase *finite* or *infinite volume Gibbs distribution* to emphasize the boundedness or unboundedness of the underlying space $(\mathfrak{X}, \mathcal{T})$, respectively.

In fact, the strict properness of a specification Γ leads to a simplified sufficient and necessary condition for P being a Gibbs measure for Γ . This is given in the following remark.

6.4 Remark (Dobrushin–Lanford–Ruelle Equations) Let $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ be a specification. Then $P \in \mathbf{Gibbs}(\Gamma)$ if and only if

$$P = \int \Gamma_H(\xi, \cdot) P(d\xi) \tag{6.3}$$

holds for all $H \in \mathcal{H}$. Since (6.3) is an obvious consequence of $P \in \mathbf{Gibbs}(\Gamma)$, we just have to show the converse. To this end, consider a P that fulfills (6.3) and let $H \in \mathcal{H}$. Then, by the prerequisite we have for all $A \in \mathcal{N}$ and $B \in \mathcal{F}^H$

$$\begin{aligned} P(A \cap B) &= \int \Gamma_H(\eta, A \cap B) P(d\eta) \\ &= \int \Gamma_H(\eta, A) \mathbb{1}_B(\eta) P(d\eta), \end{aligned}$$

where the second equality is due to the strictness of Γ_H which in particular yields that $\{\Gamma_H(\cdot, A) \mathbb{1}_B(\cdot) \neq \Gamma_H(\cdot, A \cap B)\} = \emptyset$ is a P -null set. We therefore obtain $P(A | \mathcal{F}^H) = \Gamma_H(\cdot, A)$ almost surely w. r. t. P .

Introducing distributions via (6.3) for a given specification was first considered by Dobrushin (1968a,b) and Lanford III & Ruelle (1969). These equations are therefore called *Dobrushin–Lanford–Ruelle (DLR) equations* and Gibbs distributions are also called *DLR states*. \blacklozenge

Along these first steps, the reader might have noticed a certain relation to the theory of Markov processes (indexed by time) or Markov random fields (indexed by a countable set). In fact, so-called finite interaction Gibbs point processes (see below, Example 6.14, page 142 ff.) exhibit a canonical form of Markovianity if a corresponding equivalence relation is used that defines neighboring areas.⁴ Therefore, these Gibbs point processes are also commonly called *Markov point processes* (see Ripley & Kelly (1977) who introduced this notion; see also van Lieshout (2000) and Møller & Waagepetersen (2004)).

To outline another analogy, we take a closer look at the nature of a given specification Γ . For any chain⁵ $\mathcal{C} \subset \mathcal{H}(\mathfrak{X})$ w. r. t. the partial order \subset it then turns out that the restriction $\Gamma|_{\mathcal{C}} = (\Gamma_C)_{C \in \mathcal{C}}$ is a semigroup of quasi-Markov kernels, since $\Gamma_D \Gamma_C = \Gamma_{C \cup D}$ holds for all $C, D \in \mathcal{C}$.⁶ If in addition $\emptyset \in \mathcal{C}$, we have that the semigroup $\Gamma|_{\mathcal{C}}$ is normal due to $\Gamma_{\emptyset}(\xi, \cdot) = \delta_{\xi}$ for all $\xi \in \mathfrak{X}$. This is the counterpart to the (time indexed) *Markov semigroup*. However, while the latter leads canonically to a projective family of marginal distributions and, consequently, to a unique distribution of a Markov process if an initial distribution is given and if

⁴The lattice process interpretation of Gibbs point processes is often used to study their “infinite volume” properties such as existence and uniqueness of Gibbs measures, see Section 6.4, and also asymptotic properties of estimators for Gibbs point process parameters, see for example Baddeley *et al.* (2014) or Dereudre & Lavancier (2017) and the references therein.

⁵Let (\mathfrak{X}, \preceq) be a partially ordered set. A subset $M \subset \mathfrak{X}$ is called *chain* (w. r. t. \preceq) iff (M, \preceq) is totally ordered, that is, for all $x, y \in M$ we have either $x \preceq y$ or $y \preceq x$.

⁶Consider $C \subset D$, then $\Gamma_D \Gamma_C = \Gamma_D = \Gamma_{C \cup D}$ holds by the Chapman–Kolmogorov equations. For the converse case $C \supset D$, we note that $(\mathcal{F}^C)_{C \in \mathcal{C}}$ is a decreasing filtration, hence, $\mathcal{F}^C \subset \mathcal{F}^D$ and, consequently, by the strictness of Γ_D we have $\Gamma_D \Gamma_C = \Gamma_C = \Gamma_{C \cup D}$.

the state space is Polish (Bauer, 1991, Satz 36.4, page 320), the situation is different when constructing Gibbs point processes w. r. t. $\Gamma|_{\mathcal{C}}$.

The subsequent sections are concerned with the question of the existence of Gibbs measures for a given specification Γ . To some extent, we also consider uniqueness aspects. In fact, specifications Γ in general do not define Gibbs measures uniquely if there is a corresponding distribution at all. The theory of Gibbs measures is, therefore, mainly devoted to the investigation of the set $\mathbf{Gibbs}(\Gamma)$ with respect to its size and shape, instead of looking at specific Gibbs distributions. To get familiar with some of the peculiarities, we present a few examples and results in the next section. These pave the way to some of the challenging issues mentioned above.

6.2 First Simple Examples and Results

Poisson point processes provide a first simple example of Gibbs point processes. This is shown in the following proposition, by which we also get an impression of how the Poisson case fits to the Gibbs framework.⁷

6.5 Proposition (Poisson Specification) Let $\alpha \in \mathfrak{M}(\mathfrak{X})$. Then, for every $H \in \mathcal{B}(\mathfrak{X})$ let $P_H := \mathbf{Pop}(H, \mathcal{B}(H), \alpha|_H)$ be the Poisson point process distribution on $(\mathfrak{N}_H, \mathcal{N}_H)$. We set

$$\Gamma_H(\cdot, A \cap B) := \mathbb{1}_B(\cdot) P_H(\text{pr}_H(A)) \quad \text{for } A \in \mathcal{F}_H, B \in \mathcal{F}^H \quad (6.4)$$

and obtain the following two statements:

- (a) The family $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ has a unique extension to a specification which we call *Poisson specification*.
- (b) We have $\mathbf{Gibbs}(\Gamma) = \{\mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \alpha)\}$, that is, there exists a unique distribution w. r. t. Γ , which is the Poisson point process distribution on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$ with parameter measure α . \blacklozenge

Proof: (a) Let $H \in \mathcal{H}$. Due to the product space representation

$$(\mathfrak{N}, \mathcal{N}) \cong (\mathfrak{N}_H \times \mathfrak{N}_{H^c}, \mathcal{N}_H \otimes \mathcal{N}_{H^c})$$

(see Proposition 5.10, page 107), we may interpret Γ_H as the product kernel from $(\mathfrak{N}, \mathcal{F}^H)$ to $(\mathfrak{N}_H \times \mathfrak{N}_{H^c}, \mathcal{N}_H \otimes \mathcal{N}_{H^c})$, given by

$$\Gamma_H(\xi, A \times B) = \delta_{\text{pr}_{H^c}(\xi)}(B) \cdot P_H(A)$$

for $\xi \in \mathfrak{N}$, $A \in \mathcal{N}_H$ and $B \in \mathcal{N}_{H^c}$. Furthermore, this Markov kernel is strict by construction. To show the consistency condition, we use again a corresponding product representation of $(\mathfrak{N}, \mathcal{N})$. Let $G, H \in \mathcal{H}$ such that $G \subset H$

⁷In the theory of (time-indexed) stochastic processes it is known that a Lévy process $X = (X_t)_{t \geq 0}$ is a Markov process with transition kernels $R_t(x, \cdot) = P^{X_t} * \delta_x(\cdot)$ for $t \geq 0$ and $x \in \mathfrak{X}$. In this spirit, the statement of Proposition 6.5 is an analog result.

and let $A \in \mathcal{N}_G$, $B \in \mathcal{N}_{H \setminus G}$ as well as $C \in \mathcal{N}_{H^c}$. Then, we have

$$\begin{aligned}
& \Gamma_H \Gamma_G(\xi, A \times B \times C) \\
&= \int \Gamma_G(\eta, A \times B \times C) \Gamma_H(\xi, d\eta) \\
&= \int P_G(A) \delta_{\text{pr}_G^c(\eta)}(B \times C) \Gamma_H(\xi, d\eta) \\
&= \int P_H((\text{pr}_G^H)^{-1}(A)) \delta_\eta((\text{pr}_{H \setminus G}^H)^{-1}(B) \times C) \Gamma_H(\xi, d\eta) \\
&= P_H((\text{pr}_G^H)^{-1}(A)) \Gamma_H(\xi, (\text{pr}_{H \setminus G}^H)^{-1}(B) \times C) \\
&= P_H((\text{pr}_G^H)^{-1}(A)) P_H((\text{pr}_{H \setminus G}^H)^{-1}(B)) \delta_{\text{pr}_{H^c}(\xi)}(C),
\end{aligned}$$

where we used the projectivity of $(P_H)_{H \in \mathcal{H}}$ and the definition of the Γ_H . Furthermore, by the fact that $(\text{pr}_G^H)^{-1}(A)$ is independent of $(\text{pr}_{H \setminus G}^H)^{-1}(B)$ with respect to P_H , that is,

$$P_H((\text{pr}_G^H)^{-1}(A)) P_H((\text{pr}_{H \setminus G}^H)^{-1}(B)) = P_H((\text{pr}_G^H)^{-1}(A) \cap (\text{pr}_{H \setminus G}^H)^{-1}(B)),$$

we finally get

$$\begin{aligned}
\Gamma_H \Gamma_G(\xi, A \times B \times C) &= P_H((\text{pr}_G^H)^{-1}(A) \cap (\text{pr}_{H \setminus G}^H)^{-1}(B)) \delta_{\text{pr}_{H^c}(\xi)}(C) \\
&= P_H(A \times B) \delta_{\text{pr}_{H^c}(\xi)}(C) \\
&= \Gamma_H(\xi, A \times B \times C).
\end{aligned}$$

In total we obtain that (6.4) defines a specification.

(b) To show that $\mathbf{Pop}_\alpha = \mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \alpha)$ is the unique Gibbs distribution fitting to the specification Γ , that is, $\mathbf{Gibbs}(\Gamma) = \{\mathbf{Pop}_\alpha\}$, we show that \mathbf{Pop}_α satisfies the DLR equations (6.3). To this end, consider $A \in \mathcal{F}_H$ and $B \in \mathcal{F}^H$ and let $F_A \in \mathcal{N}_H$ and $F_B \in \mathcal{N}_{H^c}$ such that $A = \text{pr}_H(F_A)$ and $B = \text{pr}_{H^c}(F_B)$. Then, due to the definition of Γ_H (see (6.4)), we have

$$\begin{aligned}
\mathbf{Pop}_\alpha(A \cap B) = P_H(F_A) P_{H^c}(F_B) &= \int P_H(F_A) \mathbb{1}_B(\xi) \mathbf{Pop}_\alpha(d\xi) \\
&= \int \Gamma_H(\xi, A \cap B) \mathbf{Pop}_\alpha(d\xi),
\end{aligned}$$

which means $\mathbf{Pop}(\alpha) \in \mathbf{Gibbs}(\Gamma)$, see Remark 6.4. Furthermore, for any $Q \in \mathbf{Gibbs}(\Gamma)$, $H \in \mathcal{H}$, and $A \in \mathfrak{N}_H$, we have

$$\begin{aligned}
Q^{\text{pr}_H}(A) = \int \mathbb{1}_{\text{pr}_H^{-1}(A)} dQ &= \int \Gamma_H(\cdot, \text{pr}_H^{-1}(A)) dQ \\
&= \int \delta_\xi(\text{pr}_{H^c}^{-1}(\mathfrak{N}_{H^c})) \cdot P_H(A) Q(d\xi) \\
&= P_H(A).
\end{aligned}$$

Thus, by the uniqueness theorem (see Theorem 5.23, page 112) we get $Q = \mathbf{Pop}_\alpha$. \blacksquare

The proposition as well as the corresponding proof reveal some interesting facts. First, a Poisson specification admits a Gibbs measure and this is unique. Second, the projectivity of $(\mathbf{Poi}(H, \mathcal{B}(H), \alpha|_H))_{H \in \mathcal{H}}$ yields the Chapman–Kolmogorov equation. The latter can therefore be seen as a generalization of the projectivity and, eventually, as the correct “consistency” condition.

The second example concerns a simple specification that does not lead to a Gibbs measure. A corresponding *lattice gas*⁸ version was already analyzed in Georgii (2011, page 66) serving as a model for a single particle that is placed completely at random in space.

6.6 Proposition (Bernoulli Specification) Let $(\mathfrak{X}, \mathcal{B}(\mathfrak{X})) := (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. We set $\mathcal{F} := \{H \in \mathcal{H}(\mathbb{R}^k) : \lambda^k(H) > 0\}$ and for each $H \in \mathcal{F}$, $\xi \in \mathfrak{N}$, and $A \in \mathcal{N}$, we define⁹

$$\Gamma_H(\xi, A) := \begin{cases} \int_H \mathbf{1}_A(\delta_x) \mathbf{U}_H(dx) & \text{if } \xi_{H^c} = \mathbf{0}, \\ \mathbf{1}_A(\xi_{H^c}) & \text{otherwise} \end{cases} \quad (6.5)$$

and furthermore $\Gamma_\emptyset(\xi, \cdot) := \delta_\xi(\cdot)$ for all $\xi \in \mathfrak{N}$. Then:

- (a) The family $\Gamma = (\Gamma_H)_{H \in \mathcal{F}}$ has a unique extension to a specification which we call *Bernoulli specification*.
- (b) We have $\mathbf{Gibbs}(\Gamma) = \emptyset$, that is, there is no Gibbs measure w. r. t. the specification Γ . \blacklozenge

Proof: (a) Obviously, \mathcal{F} contains an increasing set sequence $(H_n)_{n \in \mathbb{N}}$ such that $H_n \uparrow \mathbb{R}^k$ for $n \rightarrow \infty$. To show that Γ defines a specification, we therefore begin with the Chapman–Kolmogorov equation. Consider $G, H \in \mathcal{F}$ such that $G \subset H$ and let $A \in \mathcal{N}$ and $\xi \in \mathfrak{N}$. For the case $\xi_{H^c} = \mathbf{0}$, we compute

$$\begin{aligned} \Gamma_H \Gamma_G(\xi, A) &= \int_{\mathfrak{N}} \Gamma_G(\eta, A) \Gamma_H(\xi, d\eta) \\ &= \int_H \Gamma_G(\delta_x, A) \mathbf{U}_H(dx) \\ &= \int_G \int_G \mathbf{1}_A(\delta_y) \mathbf{U}_G(dy) \mathbf{U}_H(dx) + \int_{H \setminus G} \mathbf{1}_A(\delta_x) \mathbf{U}_H(dx) \\ &= \int_G \mathbf{1}_A(\delta_y) \mathbf{U}_H(dy) + \int_{H \setminus G} \mathbf{1}_A(\delta_x) \mathbf{U}_H(dx) \\ &= \Gamma_H(\xi, A). \end{aligned}$$

⁸A family of random elements that is indexed by a countably (infinite) set is called *lattice gas*, which might represent a simplified picture of a gas (Georgii, 2011, page 2).

⁹Recall that \mathbf{U}_H denotes the uniform distribution on a set $H \in \mathcal{H}(\mathbb{R}^k)$ with positive but finite Lebesgue measure $\lambda^k(H)$.

If $\xi_{H^c} \neq \mathbf{0}$, then we have in particular $\xi_{G^c} \neq \mathbf{0}$ and consequently

$$\begin{aligned}\Gamma_H \Gamma_G(\xi, A) &= \int \Gamma_G(\eta, A) \delta_{\xi_{H^c}}(d\eta) \\ &= \Gamma_G(\xi_{H^c}, A) \\ &= \mathbf{1}_A(\xi_{H^c}) \\ &= \Gamma_H(\xi, A).\end{aligned}$$

Furthermore, $\Gamma_H(\cdot, A)$ is obviously \mathcal{F}^H -measurable. To show the strictness, we consider $A \in \mathcal{F}_H$, $B \in \mathcal{F}^H$ and obtain indeed $\Gamma_G(\cdot, A \cap B) = \Gamma_G(\cdot, A) \mathbf{1}_B$. Hence, $\Gamma = (\Gamma_H)_{H \in \mathcal{J}}$ is a specification.

(b) We show that $P\Gamma_H = P$ for any $H \in \mathcal{J}$ has the unique solution $P = \mathbf{0}$, that is, P is the zero measure and, consequently, $\mathbf{Gibbs}(\Gamma) = \emptyset$.

To this end, let $\{D_n : n \in \mathbb{N}\}$ be a partition of \mathfrak{X} consisting of bounded sets. Assume that P fulfills $P\Gamma_H = P$ for every $H \in \mathcal{J}$ and let $\Xi := \mathbf{id}_{\mathfrak{X}}$. Then, we have

$$P(\Xi(\mathfrak{X}) > 1) \leq \sum_{n \in \mathbb{N}} P(\Xi(D_n) > 1) = \sum_{n \in \mathbb{N}} \int \Gamma_{D_n}(\xi, \Xi(D_n) > 1) P(d\xi) = 0,$$

since $\Gamma_{D_n}(\xi, \Xi(D_n) > 1) = 0$ for all $n \in \mathbb{N}$. Furthermore, for every $H \in \mathcal{J}$ we have $\Gamma_H(\xi, \Xi(\mathfrak{X}) = 0) = 0$ for all $\xi \in \mathfrak{N}(\mathfrak{X})$ by definition of Γ_H and, hence,

$$P(\Xi(\mathfrak{X}) = 0) = \int \Gamma_H(\xi, \Xi(\mathfrak{X}) = 0) P(d\xi) = 0.$$

Finally, we note

$$P(\Xi(\mathfrak{X}) = 1) = P\left(\underbrace{\bigcup_{n \in \mathbb{N}} \{\Xi(D_n) = 1\}}_{=: A_n} \cap \{\Xi(\mathfrak{X}) = 1\}\right) = \sum_{n \in \mathbb{N}} P(A_n). \quad (6.6)$$

In fact, $P(A_n) = 0$ for all $n \in \mathbb{N}$, since for every $H \in \mathcal{J}$ with $D_n \subset H$ we have

$$P(A_n) = \int \underbrace{\Gamma_H(\xi, A_n)}_{=: \mathbf{U}_H(D_n)} P(d\xi) = \mathbf{U}_H(D_n).$$

By choosing a sufficiently large $H \in \mathcal{H}$, the right-hand side gets arbitrarily small. From (6.6) we thus obtain $P(\Xi(\mathfrak{X}) = 1) = 0$.

In total, $P = \mathbf{0} \notin \mathbf{Gibbs}(\Gamma)$ and therefore $\mathbf{Gibbs}(\Gamma) = \emptyset$. \blacksquare

By these two examples, we have in particular shown that specifications exist. However, while the Poisson specification admits a unique Gibbs measure on the whole space, the Bernoulli specification does not.

We now take a closer look at the situation when the state space $(\mathfrak{X}, \mathcal{T})$ is compact. Regarding the question of existence and uniqueness, things are much simpler as the following proposition shows.

6.7 Proposition (Finite Volume Gibbs Measures) Let $(\mathfrak{X}, \mathcal{T})$ be a compact space and let $\Gamma = (\Gamma_H)_{H \in \mathcal{H}(\mathfrak{X})}$ be a specification. Then, there exists a unique Gibbs distribution, that is, $\mathbf{Gibbs}(\Gamma) = \{P\}$ and $P = \Gamma_{\mathfrak{X}}(\xi, \cdot)$ for every $\xi \in \mathfrak{N}(\mathfrak{X})$. \blacklozenge

Proof: Note that for any $A \in \mathcal{N}(\mathfrak{X}) = \mathcal{F}_{\mathfrak{X}}$ the map $\xi \mapsto \Gamma_{\mathfrak{X}}(\xi, A)$ is measurable w. r. t. $\mathcal{F}^{\mathfrak{X}} = \{\emptyset, \mathfrak{N}\}$. Hence, $\Gamma_{\mathfrak{X}}(\cdot, A)$ is constant for every A and we set $P := \Gamma_{\mathfrak{X}}(\xi, \cdot)$ for some ξ . Furthermore, for any $H \in \mathcal{H}$ we have $P\Gamma_H = P$ and hence $P(\cdot | \mathcal{F}^H) = \Gamma_H$, since Γ is a specification. This implies $P \in \mathbf{Gibbs}(\Gamma)$. For any $Q \in \mathbf{Gibbs}(\Gamma)$ we have for all H and $A \in \mathcal{B}(\mathfrak{X})$

$$Q(A) = \int \Gamma_H(\xi, A) Q(d\xi),$$

in particular $Q(A) = \Gamma_{\mathfrak{X}}(\cdot, A)$ for all A , since $\xi \mapsto \Gamma_{\mathfrak{X}}(\xi, A)$ is measurable w. r. t. $\mathcal{F}^{\mathfrak{X}} = \{\emptyset, \mathfrak{N}\}$. This yields $Q = P$ and hence $\mathbf{Gibbs}(\Gamma) = \{P\}$. \blacksquare

Proposition 6.7 addresses the question of existence and uniqueness of finite Gibbs measures to a given specification. Now we turn to the converse questions that also arises quite naturally and is non-trivial as well. Recall that a specification is more than just a family of conditional distributions. We therefore ask: Which distributions P on $\mathfrak{N}(\mathfrak{X})$ can be seen as a (finite) Gibbs measure? In other words, for which distributions does a specification exist?

Indeed, we can specify a large class of finite Gibbs measures P on $\mathfrak{N}(\mathfrak{X})$ by requiring that P has a so-called *hereditary density* f w. r. t. the Poisson point process distribution $\mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \lambda)$ for some $\lambda \in \mathfrak{M}(\mathfrak{X})$ (see also Proposition 5.31 on page 116, where we studied the support of the Poisson point process distribution).

6.8 Definition A function $h : \mathfrak{N} \rightarrow \mathbb{R}_+$ is called *hereditary* iff $h(\xi) > 0$ implies $h(\eta) > 0$ for all sub patterns $\eta \leq \xi$. \blacklozenge

6.9 Proposition (Finite Volume Gibbs Measures) Let $(\mathfrak{X}, \mathcal{T})$ be a compact space and let P be any point process distribution with a hereditary density f w. r. t. $\mathbf{Pop}_{\mathfrak{X}}$. Define for any $G \in \mathcal{H}$

$$f_G(\xi) := \frac{f(\xi)}{\int f(\eta + \xi_{G^c}) \mathbf{Pop}_G(d\eta)} \quad \text{for } \xi \in \mathfrak{N}(\mathfrak{X})$$

and

$$\Gamma_G(\xi, A) := \int \mathbb{1}_A(\eta + \xi_{G^c}) f_G(\eta + \xi_{G^c}) \mathbf{Pop}_G(d\eta) \quad \text{for } \xi \in \mathfrak{N}(\mathfrak{X}), A \in \mathcal{N}(\mathfrak{X}).$$

Then, $\Gamma = (\Gamma_G)_{G \in \mathcal{H}}$ is a specification and $\mathbf{Gibbs}(\Gamma) = \{P\}$. \blacklozenge

Proof: First, we show that Γ_G is a strict \mathcal{F}^G -Markov kernel. By definition we obviously have for any $\xi \in \mathfrak{N}(\mathfrak{X})$ that $\Gamma_G(\xi, \cdot)$ is a distribution. The

\mathcal{F}^G -mesurability of $\Gamma_G(\cdot, A)$ follows from Tonelli's theorem. Furthermore, it is easy to see from the definition that for any $A \in \mathcal{F}_G$ and $B \in \mathcal{F}^G$ we have $\Gamma_G(\cdot, A \cap B) = \Gamma_G(\cdot, A)\mathbb{1}_B(\cdot)$.

Second, we show the Chapman–Kolmogorov equations. To this end, consider $F, G \in \mathcal{H}$ with $F \subset G$, $A \in \mathcal{N}$ and $\xi \in \mathfrak{N}$, then

$$\begin{aligned} (\Gamma_G \Gamma_F)(\xi, A) &= \int \Gamma_F(\eta, A) \Gamma_G(\xi, d\eta) \\ &= \int \left[\int \mathbb{1}_A(\zeta + \eta_{G \setminus F} + \xi_{G^c}) f_F(\zeta + \eta_{G \setminus F} + \xi_{G^c}) \mathbf{Pop}_F(d\zeta) \right] \\ &\quad f_G(\eta + \xi_{G^c}) \mathbf{Pop}_G(d\eta). \end{aligned}$$

Recall that

$$\mathbf{Pop}_G = \mathbf{Pop}_F * \mathbf{Pop}_{G \setminus F} := (\mathbf{Pop}_F \otimes \mathbf{Pop}_{G \setminus F}) \circ ((\xi, \eta) \mapsto \xi + \eta)^{-1},$$

which leads to

$$\begin{aligned} (\Gamma_G \Gamma_F)(\xi, A) &= \int \int \int \mathbb{1}_A(\zeta + \eta_2 + \xi_{G^c}) f_F(\zeta + \eta_2 + \xi_{G^c}) f_G(\eta_1 + \eta_2 + \xi_{G^c}) \\ &\quad \mathbf{Pop}_F(d\eta_1) \mathbf{Pop}_{G \setminus F}(d\eta_2) \mathbf{Pop}_F(d\zeta) \\ &= \int \mathbb{1}_A(\zeta + \eta_2 + \xi_{G^c}) f_F(\zeta + \eta_2 + \xi_{G^c}) \\ &\quad \int f_G(\eta_1 + \eta_2 + \xi_{G^c}) \mathbf{Pop}_F(d\eta_1) (\mathbf{Pop}_F \otimes \mathbf{Pop}_{G \setminus F})(d(\zeta, \eta_2)). \end{aligned}$$

By using the definitions of f_G and f_F , we obtain the relation

$$\begin{aligned} &f_F(\zeta + \eta_2 + \xi_{G^c}) \cdot \int f_G(\eta_1 + \eta_2 + \xi_{G^c}) \mathbf{Pop}_F(d\eta_1) \\ &= f_F(\zeta + \eta_2 + \xi_{G^c}) \cdot \int \frac{f(\eta_1 + \eta_2 + \xi_{G^c})}{\int f(\eta' + \xi_{G^c}) \mathbf{Pop}_G(d\eta')} \mathbf{Pop}_F(d\eta_1) \\ &= f(\zeta + \eta_2 + \xi_{G^c}) \frac{1}{\int f(\eta' + \xi_{G^c}) \mathbf{Pop}_G(d\eta')} \\ &= f_G(\zeta + \eta_2 + \xi_{G^c}). \end{aligned}$$

In total we have

$$(\Gamma_G \Gamma_F)(\xi, A) = \int \mathbb{1}_A(\eta + \xi_{G^c}) f_G(\eta + \xi_{G^c}) \mathbf{Pop}_G(d\eta).$$

Third, according to Proposition 6.7, we have that $P = \Gamma_{\mathfrak{X}}(\xi, \cdot)$ holds for every $\xi \in \mathfrak{N}(\mathfrak{X})$, which is the unique Gibbs distribution to Γ . \blacksquare

The previous proposition states that any point process distribution with hereditary density w. r. t. \mathbf{Pop} is a (*finite volume*) Gibbs measure.

Authors commonly define Gibbs point processes on compact spaces exactly in this way, since it provides a comfortable and intuitive approach for introducing point process models exhibiting dependences (see for example Møller & Waagepetersen (2004), Baddeley *et al.* (2013)). To this end, certain point patterns are either weighted down or up in a density w. r. t. **Pop**. Concrete examples are given by the *Strauss* and the *Lennard-Jones point process model*, which will be introduced and discussed in a broader setting in Section 6.3.

Note that the Poisson point process distribution as a reference measure is a key factor in this modeling approach. First, $\mathbf{Pop}_x = \mathbf{Pop}(x, \mathcal{B}(x), \alpha)$ with $\text{supp } \alpha = x$ has probability mass all over the space $\mathfrak{N}(x)$, see Proposition 5.31, page 116. Hence, we do not exclude any kind of point pattern by choosing \mathbf{Pop}_x as the reference measure. Second, \mathbf{Pop}_x can be easily decomposed for each $G \in \mathcal{H}$ as $\mathbf{Pop}_x = \mathbf{Pop}_G * \mathbf{Pop}_{G^c}$, which is used to compute the conditional densities.

Roughly speaking, Propositions 6.7 and 6.9 above show that a large class of point process distributions on compact spaces possess a Gibbsian description, that is, a description via a specification. And, conversely, the corresponding class of specifications determines a point process distribution, which consequently means that there is a one-to-one correspondence within this class.

For unbounded spaces (x, \mathcal{T}) with $\alpha(x) = \infty$ for some $\alpha \in \mathfrak{M}(x)$, there is no such one-to-one correspondence. However, we are still interested in the question if specifications that are given in the spirit of Proposition 6.9 define a Gibbs point process distribution on $(\mathfrak{N}(x), \mathcal{N}(x))$. These so-called *Gibbs specification* are considered in Section 6.3.

In the remainder of this section, we consider the comparatively simple non-compact case $(x, \mathcal{T}) := (\mathbb{Z}, 2^{\mathbb{Z}})$. In fact, questions regarding the existence and uniqueness of Gibbs measures on $(\mathfrak{N}(\mathbb{Z}), \mathcal{N}(\mathbb{Z}))$ might be already known to the reader. This refers to a certain part of the theory of (stationary, homogeneous) *Markov chains*. In what follows, our goal is to use this extensively studied case as an example, which might then be helpful as a guide for proofs in the general context. To this end, we start with a certain specification on \mathbb{Z} , modeling the nearest-neighbor dependency, and discuss the existence and uniqueness of a corresponding Gibbs measure. This turns out to be the distribution of a Markov chain.

By considering Markov chains from the Gibbs point process perspective, we are dealing with \mathbb{Z} -indexed stochastic processes with binary states, say 0 and 1. A path of such a Markov chain is an element in $\mathfrak{N}(\mathbb{Z})$, that is, a simple point measure on \mathbb{Z} . The states 1 and 0 indicate the presence and non-presence of mass of that point measure, respectively. Furthermore, our observation of this random element on a bounded and

hence finite set $G \subset \mathbb{Z}$ depend on the nearest-neighbor *periphery* of G , that is,

$$\Delta G := \{k \in G^c : |k - l| = 1 \text{ for some } l \in G\}.$$

A natural way to define a specification concerning the nearest-neighbor relationship is essentially given by the product of transition probabilities. That means, considering a transition matrix $R = (R(x, y))_{x, y \in \{0, 1\}}$, an interval $I := \{k, \dots, l\} \subset \mathbb{Z}$ and $\xi \in \mathfrak{N}(\mathbb{Z})$, we define $\Gamma_I(\xi, \cdot)$ by the probability density $f_{I, \xi}$ w. r. t. the *counting measure*¹⁰ ζ_I on $\{0, 1\}^I = \mathfrak{N}(I)$ given by

$$f_{I, \xi} := \prod_{k \in I} R(\xi_{k-1}, \eta_k) \cdot \dots \cdot R(\eta_l, \xi_{l+1}).$$

Furthermore, in the case of two *disconnected* sets $G, H \in \mathcal{H}(\mathbb{Z})$, that is, $G \cap H = \emptyset$ as well as $\Delta H \cap G = H \cap \Delta G = \emptyset$, we consider

$$f_{G \cup H, \xi} := f_{G, \xi} \cdot f_{H, \xi}.$$

Note that every finite $H \subset \mathbb{Z}$ has a finite decomposition in disconnected intervals. For $H \in \mathcal{H}$, $\xi \in \mathfrak{N}(\mathbb{Z})$, and $A \in \mathcal{N}(\mathbb{Z})$, we set

$$\Gamma_H(\xi, A) := \int \mathbf{1}_A(\eta + \xi|_{H^c}) f_{H, \xi}(\eta) \zeta_H(d\eta). \quad (6.7)$$

In fact, $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ is a specification, which is shown by the upcoming proposition.

6.10 Proposition (Markov Specification) Let $(\mathfrak{X}, \mathcal{T}) = (\mathbb{Z}, 2^{\mathbb{Z}})$ and consider the transition matrix

$$R := \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

with $p, q \in]0, 1[$, that is, non-zero entries. We define for an intervals $I := \{k, k+1, \dots, l\} \subset \mathbb{Z}$

$$f_{I, \xi}(\eta) := \frac{R(\xi_{k-1}, \eta_k) \cdot R(\eta_k, \eta_{k+1}) \cdot \dots \cdot R(\eta_l, \xi_{l+1})}{R^{l-k+2}(\xi_{k-1}, \xi_{l+1})} \quad \text{for } \xi, \eta \in \mathfrak{N}(\mathbb{Z}) \quad (6.8)$$

and, furthermore, for $G, H \in \mathcal{H}$ with $G \cap H = \emptyset$ and $\Delta G \cap H = G \cap \Delta H = \emptyset$ we set

$$f_{G \cup H, \xi} := f_{G, \xi} \cdot f_{H, \xi}. \quad (6.9)$$

Then, (6.8) and (6.9) are probability densities w. r. t. ζ_I and $\zeta_{G \cup H}$, respectively. Furthermore, $\Gamma := (\Gamma_H)_{H \in \mathcal{H}}$ given through (6.7) is a specification, the so-called *Markov specification* with transition matrix R . \blacklozenge

¹⁰Let \mathfrak{X} be an arbitrary set. The set function ζ that maps subsets M of \mathfrak{X} to their cardinality $|M|$, if M is finite and ∞ else, is a measure, the so-called *counting measure* on \mathfrak{X} .

Proof: First, we show that Γ_H for each $H \in \mathcal{H}$ is a Markov kernel from $(\mathfrak{N}, \mathcal{F}^H)$ to $(\mathfrak{N}, \mathcal{N})$. To this end, consider any $\xi \in \mathfrak{N}(\mathbb{Z})$ and let $I \subset \mathbb{Z}$ be an interval. Then, note that for $x, y \in E := \{0, 1\}$ and $n \in \mathbb{N}$ we have

$$R^n(x, y) = \sum_{i=1}^{n-1} \sum_{z_i \in E} R(x, z_1) \cdots R(z_{n-1}, y),$$

which yields that (6.8) and (6.9) are probability densities. Hence, Γ_H defined by (6.7) is a Markov kernel. Furthermore, we note that for every $\eta \in \mathfrak{N}$ the map $\xi \mapsto f_{H, \xi}(\eta)$ is $\mathcal{F}_{\Delta H}$ -measurable. Consequently, for every $A \in \mathcal{N}$ the map $\xi \mapsto \Gamma_H(\xi, A)$ is $\mathcal{F}_{\Delta H}$ -measurable and, thus, in particular \mathcal{F}^{H^c} -measurable.

Second, to show the strictness of Γ_H for any $H \in \mathcal{H}$, we consider $A \in \mathcal{N}$ and $B \in \mathcal{F}^H$ and compute

$$\begin{aligned} \Gamma_H(\xi, A \cap B) &= \sum_{\eta \in \mathfrak{N}} \mathbf{1}_A(\eta) \mathbf{1}_B(\eta) \Gamma_H(\xi, \{\eta\}) \\ &= \begin{cases} \Gamma_H(\xi, A) & \text{if } \xi \in B, \\ 0 & \text{otherwise} \end{cases} \\ &= \Gamma_H(\xi, A) \mathbf{1}_B(\xi). \end{aligned}$$

Third, we show the Chapman–Kolmogorov equations. We define for any $\eta \in \mathfrak{N}$ and any $H \in \mathcal{H}$ the set $\{\eta\}_H := \text{pr}_H^{-1}(\text{pr}_H(\eta))$, that is, the set of all simple point measures on \mathbb{Z} , which equals η on H . Then, we have $\{\eta\}_H = \{\eta\}_F \cap \{\eta\}_{H \setminus F}$. For the sake of simplicity, we show the statement for two intervals

$$I := \{k, \dots, l\} \subset \{j, \dots, m\} =: J.$$

We then compute that

$$\begin{aligned} &\Gamma_J \Gamma_I(\xi, \{\eta\}_J) \\ &= \int \Gamma_I(\zeta, \{\eta\}_I \cap \{\eta\}_{J \setminus I}) \Gamma_J(\xi, d\zeta) \\ &= \int \Gamma_I(\zeta, \{\eta\}_I) \mathbf{1}_{\{\eta\}_{J \setminus I}}(\zeta) \Gamma_J(\xi, d\zeta) \\ &= \frac{R(\eta_{k-1}, \eta_k) \cdots R(\eta_l, \eta_{l+1})}{R^{l-k+2}(\eta_{k-1}, \eta_{l+1})} \cdot \sum_{\zeta} \frac{R(\xi_{j-1}, \zeta_j) \cdots R(\zeta_m, \xi_{m+1})}{R^{m-j+2}(\xi_{j-1}, \xi_{m+1})} \mathbf{1}_{\{\eta\}_{J \setminus I}}(\zeta) \\ &= \frac{R(\xi_{j-1}, \eta_j) \cdots R(\eta_m, \xi_{m+1})}{R^{m-j+2}(\xi_{j-1}, \xi_{m+1})} \cdot \underbrace{\sum_{\zeta} \frac{R(\eta_{k-1}, \zeta_k) \cdots R(\zeta_l, \eta_{l+1})}{R^{l-k+2}(\eta_{k-1}, \eta_{l+1})}}_{=1} \\ &= \Gamma_J(\xi, \{\eta\}_J). \end{aligned}$$

Hence, for all $A \in \mathcal{F}_J$ we have $\Gamma_J \Gamma_I(\xi, A) = \Gamma_J(\xi, A)$. Furthermore, consider any $B \in \mathcal{F}^J$ in addition, we then have that

$$\begin{aligned} \Gamma_J \Gamma_I(\xi, A \cap B) &= \int \Gamma_I(\zeta, A) \mathbf{1}_B(\zeta) \Gamma_J(\xi, d\zeta) \\ &= \Gamma_J \Gamma_I(\xi, A) \cdot \mathbf{1}_B(\xi) \\ &= \Gamma_J(\xi, A) \cdot \mathbf{1}_B(\xi). \end{aligned}$$

Finally, $\Gamma_J \Gamma_I(\xi, A) = \Gamma_J(\xi, A)$ holds for all $A \in \mathcal{N}$. \blacksquare

By the next proposition, we show that each Markov specification with transition matrix R specifies uniquely a Gibbs distribution, which is the distribution of a stationary (binary) Markov chain.¹¹ Beyond that, the crucial message lies in the organization of the corresponding proof, which is designed to prepare existence proofs for general Gibbs point processes in the next section.

6.11 Proposition (Binary Markov Chain on \mathbb{Z}) To each Markov specification Γ with transition matrix R as in Proposition 6.10 there exists a unique Gibbs measure P , that is, $\mathbf{Gibbs}(\Gamma) = \{P\}$. \blacklozenge

Proof: Essentially, we will show for an increasing sequence $(H_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ and any boundary condition ξ as well, that $\Gamma_{H_n}(\xi, \cdot)$ converges to a distribution P , independent of ξ . Furthermore, this measure is the “correct” one, since for $n \rightarrow \infty$ we have

$$P \leftarrow \Gamma_{H_n}(\xi, \cdot) = \Gamma_{H_n} \Gamma_G(\xi, \cdot) \rightarrow \int \Gamma_G(\eta, \cdot) P(d\eta),$$

which implies that P has the “correct” conditional distributions, that is, those given by $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$.

To show the statement, we first note that for a transition matrix

$$R = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

with $p, q \in]0, 1[$ we have that

$$R^n \rightarrow \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix} \quad \text{as } n \rightarrow \infty. \quad (6.10)$$

Let $\xi \in \mathfrak{N}(\mathbb{Z})$ and consider a sequence $(H_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ of intervals with $H_n \uparrow \mathbb{Z}$ for $n \rightarrow \infty$. By $l(n) := \min H_n$ and $u(n) := \max H_n$ we denote the lower and upper bounds of H_n , respectively. That is, we have in particular

¹¹See Georgii (2011, Chapter 3) for the more general case of Markov chains with finite states.

$\Delta H_n = \{l(n) - 1, u(n) + 1\}$ for all $n \in \mathbb{N}$. Additionally, consider some $\eta \in \mathfrak{N}(\mathbb{Z})$ and $G := \{k, \dots, l\} \in \mathcal{H}$. Recall that $\{\eta\}_G := \text{pr}_G^{-1}(\text{pr}_G(\eta))$, we then compute

$$\begin{aligned} & \Gamma_{H_n}(\xi, \{\eta\}_G) \\ &= \sum_{\xi} \frac{R(\xi_{l(n)-1}, \zeta_{l(n)}) \cdots R(\zeta_{k-1}, \eta_k) \cdots R(\eta_l, \zeta_{l+1}) \cdots R(\zeta_{u(n)}, \xi_{u(n)+1})}{R^{u(n)-l(n)+2}(\xi_{l(n)-1}, \xi_{u(n)-1})} \\ &= R(\eta_k, \eta_{k+1}) \cdots R(\eta_{l-1}, \eta_l) \cdot \frac{R^{k-l(n)}(\xi_{l(n)-1}, \eta_k) R^{u(n)-l}(\eta_l, \xi_{u(n)+1})}{R^{u(n)-l(n)+2}(\xi_{l(n)-1}, \xi_{u(n)+1})}. \end{aligned}$$

Due to 6.10 we have $\lim_{n \rightarrow \infty} R^{k-l(n)}(\xi_{l(n)-1}, \eta_k) =: \mu(\eta_k)$ and since

$$\lim_{n \rightarrow \infty} \frac{R^{u(n)-l}(\eta_l, \xi_{u(n)+1})}{R^{u(n)-l(n)+2}(\xi_{l(n)-1}, \xi_{u(n)+1})} = 1,$$

we have in total and independent of $\xi \in \mathfrak{N}(\mathbb{Z})$ that

$$\lim_{n \rightarrow \infty} \Gamma_{H_n}(\xi, \{\eta\}_G) = \mu(\eta_k) R(\eta_k, \eta_{k+1}) \cdots R(\eta_{l-1}, \eta_l) \quad \text{for } \eta \in \mathfrak{N}(\mathbb{Z}). \quad (6.11)$$

Consequently,

$$\Gamma_{H_n}(\xi, \cdot)|_{\mathcal{F}_G} \rightarrow P_G|_{\mathcal{F}_G} \quad \text{as } n \rightarrow \infty,$$

where P_G is defined through the right-hand side of (6.11). Furthermore, let \mathcal{G} be the set of all intervals of \mathbb{Z} . Then $(P_G)_{G \in \mathcal{G}}$ is consistent and by Corollary 5.26 on page 114 there is a distribution P on $\mathfrak{N}(\mathbb{Z})$ such that $P|_{\mathcal{F}_G} = P_G$.

Since $\Gamma_{H_n}(\xi, \cdot)$ converges weakly to P as $n \rightarrow \infty$ and since $\eta \mapsto \Gamma_G(\eta, A)$ is trivially continuous for any $G \in \mathcal{H}$, we have for every $A \in \mathcal{N}$

$$P(A) \leftarrow \Gamma_{H_n}(\xi, A) = \Gamma_{H_n} \Gamma_G(\xi, A) \rightarrow \int \Gamma_G(\eta, A) P(d\eta) \quad \text{as } n \rightarrow \infty.$$

In addition, for any $B \in \mathcal{F}^G$ we have

$$P(A \cap B) = \int \Gamma_G(\eta, A) \mathbb{1}_B(\eta) P(d\eta),$$

and therefore $P(\cdot | \mathcal{F}^G) = \Gamma_G$ and consequently $P \in \mathbf{Gibbs}(\Gamma)$.

To show the uniqueness of a Gibbs measure specified through Γ , we consider $Q \in \mathbf{Gibbs}(\Gamma)$ and note for every $\xi \in \mathfrak{N}(\mathbb{Z})$ and $G, H \in \mathcal{H}$ that

$$Q(\{\xi\}_G) = \int \Gamma_H(\eta, \{\xi\}_G) Q(d\eta).$$

Considering especially $G := \{k, \dots, l\}$, we have for an increasing sequence of intervals $(H_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ and for all $\xi \in \mathfrak{N}(\mathbb{Z})$

$$\begin{aligned} \mathcal{Q}(\{\xi\}_G) &= \int \Gamma_{H_n}(\eta, \{\xi\}_G) \mathcal{Q}(d\eta) \\ &= \lim_{n \rightarrow \infty} \int \Gamma_{H_n}(\eta, \{\xi\}_G) \mathcal{Q}(d\eta) \\ &= \int \lim_{n \rightarrow \infty} \Gamma_{H_n}(\eta, \{\xi\}_G) \mathcal{Q}(d\eta) \\ &= \mu(\xi_k) \cdot R(\xi_k, \xi_{k+1}) \cdot \dots \cdot R(\xi_{l-1}, \xi_l) \\ &= P(\{\xi\}_G). \end{aligned}$$

Hence, \mathcal{Q} equals P and consequently $\mathbf{Gibbs}(\Gamma) = \{P\}$. \blacksquare

Binary Markov chains on \mathbb{Z} as considered above emerge, for instance, in statistical physics by the so-called *Ising model* to study ferromagnetic or antiferromagnetic substances (see e. g. Georgii (2011, page 49) for a mathematical analysis, some explanation and references on the Ising model). The following example also provides an impression of introducing the so-called *Gibbs specification* for continuous models in Section 6.3.

6.12 Example (Ising Model in One Dimension) We set up the *one-dimensional Ising model* as follows: consider two states, for simplicity say -1 and 1 , representing the orientations of so-called *spins*. The positions of these spins are given by \mathbb{Z} . The likelihood of an outcome $\xi \in \mathfrak{N}(\mathbb{Z})$ is then modeled in terms of the *Ising potential*, that is, for $x, y \in \mathbb{Z}$

$$\phi_{\{x,y\}}(\xi) := \begin{cases} -J(\mathbb{1}_{\xi \times \xi}(x, y) - \mathbb{1}_{\xi \times \xi^c}(x, y) + \mathbb{1}_{\xi^c \times \xi^c}(x, y)) & \text{if } |x - y| = 1, \\ -h(\mathbb{1}_{\xi}(x) - \mathbb{1}_{\xi^c}(x)) & \text{if } x = y, \\ 0 & \text{else.} \end{cases}$$

Here, $J \in \mathbb{R}$ is a coupling constant and $h \in \mathbb{R}$ describes the action of an external field. For $H \in \mathcal{H}(\mathbb{Z})$ let

$$U_H(\xi) := \sum_{\substack{x, y \in \mathbb{Z} \\ \{x, y\} \cap H \neq \emptyset}} \phi_{\{x, y\}}(\xi)$$

be the so-called *Ising potential energy* of ξ in H . Then, for $H \in \mathcal{H}$, $\xi \in \mathfrak{N}(\mathbb{Z})$, and $A \in \mathcal{N}(\mathbb{Z})$, we set

$$\Gamma_H^t(\xi, A) := \int \mathbb{1}_A(\eta + \xi_{H^c}) c_t(H, \xi) \exp(-t \cdot U_H(\eta + \xi_{H^c})) \zeta(d\eta),$$

where $t > 0$ is the *inverse (absolute) temperature* and $c_t(H, \xi)$ is the normalizing constant

$$c_t(H, \xi) := \left(\int \exp(-t \cdot U_H(\eta + \xi_{H^c})) \zeta(d\eta) \right)^{-1}.$$

To see that $\Gamma^t = (\Gamma_H^t)_{H \in \mathcal{H}}$ constitutes a Markov specification, in particular for the computation of the corresponding transition matrix R , we refer to Georgii (2011, Chapter 3). According to Proposition 6.11, we have that Γ^t determines a unique Gibbs distribution P_t for each $t > 0$.

In the context of this example, a classical question of statistical physics concerns the behavior of $\mathbf{Gibbs}(\Gamma^t) = \{P_t\}$ at *low temperature*, that is, as t goes to infinity. Assuming $J = 1$, for example, one can show that $\lim_{t \rightarrow \infty} P_t$ is the Dirac distribution δ_ζ , where ζ is the counting measure on \mathbb{Z} (see for example Georgii, 2011, Chapter 3). \blacklozenge

We conclude this section by a brief remark on the (non-)uniqueness of Gibbs distributions. The Ising model presented in Example 6.12 has an obvious extension to the two dimensional space $\mathfrak{X} = \mathbb{Z}^2$. There the behavior at low temperature is different. One can show that, for t large enough, $\mathbf{Gibbs}(\Gamma^t)$ is no longer a singleton (see Georgii, 2011, Chapter 6). To put it differently, the macroscopic behavior described by $\mathbf{Gibbs}(\Gamma^t)$ may change “non-smoothly”, if t is changed “smoothly”. Such a phenomenon is called *phase transition*.

For most specifications $\Gamma^t = (\Gamma_H^t)_{H \in \mathcal{H}}$ controlled by some activity parameter $t > 0$ (still called inverse temperature), proving that phase transition occurs and determining the critical inverse temperature t_c at which it occurs are topics of ongoing research. According to Dereudre (2019, page 27) the “common belief claims that the Gibbs measures are unique when [...] the inverse temperature [t is] small enough”. Intuitively, this is due to the belief that $(\Gamma^t)_{t > 0}$ converges to the Poisson specification for $t \rightarrow \infty$. Further results, notes, and references are given in Dereudre (2019). In the present thesis, the uniqueness problem is not discussed any further.¹²

6.3 Gibbs Specifications

Roughly speaking, *Gibbs specifications* are specifications $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ such that for each $H \in \mathcal{H}(\mathfrak{X})$ and each $\xi \in \mathfrak{N}$ the measure $\Gamma_H(\xi, \cdot)$ has a descriptive construction by a density w. r. t. $\mathbf{Pop}(H, \mathcal{B}(H), \alpha|_H)$ in terms of so-called *energy functions* and *potentials*. This terminology is due to the extensive use of Gibbs specifications in statistical physics for studying (large) particle systems. In addition, these physical notions also provide an illustrative account to describe the underlying processes involved. Therefore, our use of this terminology comes with references to physics, however, from a simplified point of view.

¹²In the simulation study of the present thesis (see Section 7.4, page 167 ff.) we examine statistical procedures, in particular the proposed concept of regularized Rao–Blackwellization, for the so-called *Lennard-Jones model* (see next section) for parameters describing a low, a medium, and a high temperature regime.

In order to construct densities w. r. t. \mathbf{Pop}_H , we assign to each pattern $\xi \in \mathfrak{N}$ a cost relative to a reference state ξ_0 . This is what physicists denote as the *potential energy* of ξ (with respect to ξ_0). In our situation the reference state ξ_0 will be ξ_{H^c} , that is, the configuration outside of H . By doing so we arrive at the following definition.

6.13 Definition (Potential Energy) A family $U = (U_H)_{H \in \mathcal{H}}$ of measurable functions $U_H : \mathfrak{N}_f \rightarrow]-\infty, \infty]$ is called *energy*, *potential energy*, or *Hamiltonian*¹³ iff for all $G, H \in \mathcal{H}$ with $G \subset H$ there is an \mathcal{F}^G -measurable function $\varphi_{G,H} : \mathfrak{N}_f \rightarrow]-\infty, \infty]$ such that $U_H(\xi) = U_G(\xi) + \varphi_{G,H}(\xi)$. \blacklozenge

For a potential energy $(U_H)_{H \in \mathcal{H}}$ and a point pattern $\xi \in \mathfrak{N}_f$, the value $U_H(\xi)$ can be interpreted as the energy held by the particle system ξ in H given the reference state ξ_{H^c} . Loosely speaking, the value $U_H(\xi)$ may be interpreted as the total work necessary to change from the reference configuration ξ_{H^c} to the configuration ξ . The positive or negative sign of $U_H(\xi)$ tells us if work has to be *put into* or *can be gained by* the system ξ .

Note also that the potential energy does not depend on the actual trajectory of the work done, but just depends on the resulting state ξ and the reference state ξ_{H^c} . Nevertheless, for illustrative and modeling purposes, the potential energy $U_H(\xi)$ may be described by interacting forces of all or certain leading sub-configurations. Hence, we define the following terminology.

6.14 Example (Energy and Interaction Potential) Consider a non-empty $I \subset \mathbb{N} \cup \{\infty\}$ and an I -indexed family of measurable functions $\phi_k : \mathfrak{N}_k \rightarrow]-\infty, \infty]$ with $\phi_k \neq \mathbf{0}$ for $k \in I$, where $\mathfrak{N}_k = \mathfrak{N}_k(\mathfrak{X})$ denotes the set of all simple point measures ξ such that $\xi(\mathfrak{X}) = k$. Then $(U_H)_{H \in \mathcal{H}}$ defined by

$$U_H(\xi) := \sum_{k \in I} \sum_{\substack{\eta \in \mathfrak{N}_k \\ \eta \leq \xi, \eta(H) > 0}} \phi_k(\eta) \quad \text{for } \xi \in \mathfrak{N}_f \quad (6.12)$$

is a Hamiltonian. To show this, let $G, H \in \mathcal{H}$ with $G \subset H$ and note that

$$\begin{aligned} U_H(\xi) &= \sum_{k \in I} \sum_{\substack{\eta \in \mathfrak{N}_k \\ \eta \leq \xi, \eta(H) > 0}} \phi_k(\eta) \\ &= \sum_{k \in I} \sum_{\substack{\eta \in \mathfrak{N}_k \\ \eta \leq \xi, \eta(G) > 0}} \phi_k(\eta) + \varphi_{G,H}(\xi), \end{aligned}$$

where

$$\varphi_{G,H}(\xi) := \sum_{k \in I} \sum_{\substack{\eta \in \mathfrak{N}_k, \eta \leq \xi \\ \eta(G) = 0, \eta(H) > 0}} \phi_k(\eta) \quad \text{for } \xi \in \mathfrak{N}_f$$

¹³This name comes from the Irish mathematician, physicist, and astronomer *William Rowan Hamilton* (1805 – 1865).

is obviously \mathcal{F}^G -measurable.

For $k \in I$ we call ϕ_k (*interaction potential of order k* or *k -body (interaction) potential*). In particular, ϕ_2 is often called *pair (interaction) potential*. Furthermore, $\sup I$ is called the *interaction order* and¹⁴

$$\sup \{ \text{diam}(\eta) : \phi_k(\eta) \neq 0, \eta \in \mathfrak{N}_k(\mathfrak{X}), k \in I \}$$

is called the *interaction range* of the potential energy U . \blacklozenge

We provide two well-known examples of pair interaction models. In fact, physical models of simple particle systems often assume that the potential energy is described by two-body forces, while three- and higher-body forces contribute just smaller, even negligible corrections.

6.15 Example (Strauss Interaction Potential) To model repulsion of particles, a simple approach is to count point pairs that occur at a distance less or equal to a certain radius $R > 0$, which we call R -close. The *Strauss interaction potential*¹⁵ is the two parameter pair potential

$$\phi_2^R(\delta_x + \delta_y) := -\log(\gamma) \cdot \mathbb{1}_{[0,R]}(d(x,y)) \quad \text{for } x, y \in \mathfrak{X}, x \neq y,$$

with (regular) interaction parameter $\gamma \in [0, 1]$ (see Figure 6.1). For this parametrization we note that $\gamma = 0$ leads to a model with so-called *hard spheres*, that is, two points cannot be R -close to each other. If $\gamma = 1$ point pairs do not interact. Furthermore,

$$\sup \{ \text{diam}(\eta) : \phi_2^R(\eta) \neq 0, \eta \in \mathfrak{N}_2 \} = R,$$

i. e., the Strauss interaction potential has the finite interaction range R .

Considering the 1-body potential $\phi_1 = \log(\beta) \cdot \mathbb{1}_{\mathfrak{N}_1}$ for some intensity parameter $\beta > 0$, we define the *Strauss potential energy* for $\xi \in \mathfrak{N}_f(\mathfrak{X})$ and $H \in \mathcal{H}$ as

$$\begin{aligned} U_H^{\text{Strauss}}(\xi) &:= \sum_{x \in \xi_H} \phi_1(\delta_x) + \sum_{\substack{\{x,y\} \subset \xi \\ \{x,y\} \cap H \neq \emptyset \\ y \neq x}} \phi_2^R(\delta_x + \delta_y) \\ &= \log(\beta) \cdot \xi_H(\mathfrak{X}) + \log(\gamma) (s_R^{(H)}(\xi) - s_R^{(H)}(\xi_{H^c})), \end{aligned}$$

where $\xi_H(\mathfrak{X})$ is the number particles in H and

$$s_R^{(H)}(\zeta) := \sum_{\substack{\{x,y\} \subset \zeta \\ \{x,y\} \cap H \neq \emptyset \\ y \neq x}} \phi_2(\delta_x + \delta_y)$$

¹⁴Recall that $\text{diam}(M) := \sup\{d(x,y) : x, y \in M\}$ denotes the *diameter* of subset M of a metric space (\mathfrak{X}, d) .

¹⁵This potential is due to Strauss (1975) who aimed at a model for clustering by choosing $\gamma > 1$. However, this parameter choice cannot be used to introduce a point process distribution (see Kelly & Ripley (1976) and below).

is the number of R -close point pairs of $\zeta \in \mathfrak{N}_f(\mathfrak{X})$ in H and at the boundary of H . \blacklozenge

6.16 Example (Lennard-Jones Interaction Potential) The $(12,6)$ -Lennard-Jones (LJ) interaction potential¹⁶ is the two parameter pair potential

$$\phi_2^{LJ}(\delta_x + \delta_y) := 4\varepsilon \left[\left(\frac{\sigma}{d(x,y)} \right)^{12} - \left(\frac{\sigma}{d(x,y)} \right)^6 \right] \quad \text{for } x, y \in \mathfrak{X}, x \neq y,$$

where ε is the *well depth* and σ is the *Lennard-Jones atomic diameter* (see also Figure 6.2). Strong repulsion occurs at short distances, weak attraction is attained at large distances.

Furthermore, considering the 1-body potential $\phi_1 = \log(\beta) \cdot \mathbb{1}_{\mathfrak{N}_1}$ for some intensity parameter $\beta > 0$, we define the LJ potential energy for $\xi \in \mathfrak{N}_f(\mathfrak{X})$ and $H \in \mathcal{H}$ as

$$U_H^{LJ}(\xi) := \sum_{x \in \xi_H} \phi_1(\delta_x) + \sum_{\substack{\eta \in \mathfrak{N}_2 \\ \eta \leq \xi \\ \eta(H) > 0}} \phi_2^{LJ}(\eta).$$

Note that

$$\sup \{ \text{diam}(\eta) : \phi_2^{LJ}(\eta) \neq 0, \eta \in \mathfrak{N}_2 \} = \infty,$$

that is, the LJ interaction potential has an infinite interaction range.

For the statistical purposes that we are concerned with later, we rewrite the LJ potential energy in a way that separates the repulsive and attractive forces and arrive at the following parametric form

$$\begin{aligned} U_H^{LJ}(\xi) &= \log(\beta) \cdot \xi_H(\mathfrak{X}) \\ &+ 4\varepsilon\sigma^{12} \cdot (V(\xi) - V(\xi_{H^c})) + (-4\varepsilon\sigma^6) \cdot (W(\xi) - W(\xi_{H^c})), \end{aligned} \quad (6.13)$$

where $\xi \in \mathfrak{N}_f(\mathfrak{X})$ and

$$V(\xi) := \frac{1}{2} \sum_{x \in \xi} \sum_{\substack{y \in \xi \\ y \neq x}} \frac{1}{d(x,y)^{12}} \quad \text{and} \quad W(\xi) := \frac{1}{2} \sum_{x \in \xi} \sum_{\substack{y \in \xi \\ y \neq x}} \frac{1}{d(x,y)^6}$$

is used to denote the essential component of the forces that describes the repulsion and attraction of $\xi \in \mathfrak{N}_f$, respectively. In this context, β , ε , and σ are called *model parameters*. According to the representation (6.13) of U_H^{LJ} , we consider later (Section 7.4) the so-called *canonical parameters*

$$\vartheta_1 := \log(\beta), \quad \vartheta_2 := 4\varepsilon\sigma^{12}, \quad \text{and} \quad \vartheta_3 := -4\varepsilon\sigma^6$$

and the canonical parameter space $\Theta := \mathbb{R} \times]0, \infty[\times]-\infty, 0[$. \blacklozenge

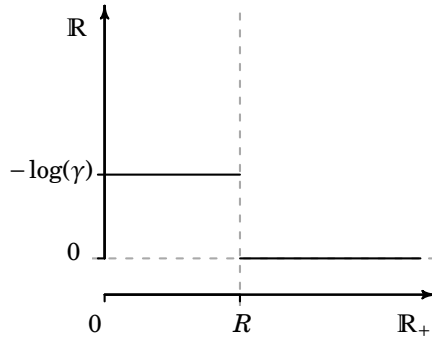


Figure 6.1: Graph of the Strauss pair potential as a function of the particle distance for the interaction range $R > 0$ and (regular) interaction parameter $\gamma \in [0, 1]$. In the case of R -closeness, repulsion is modeled constantly at the level $\log(\gamma)$. There is *no* attraction of particles.

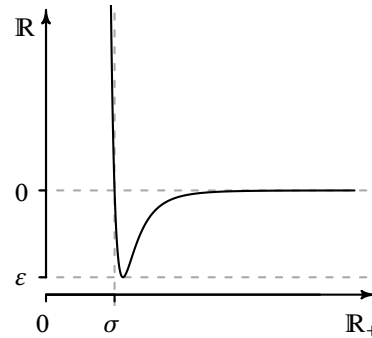


Figure 6.2: Graph of the (12,6)-Lennard-Jones pair potential as a function of the particle distance for the well depth parameter $\varepsilon > 0$ and the atomic diameter $\sigma > 0$. By ε we model attraction, while σ determines the repulsion of the particles. The interaction range is infinite.

We finally aim at constructing specifications $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ based on the potential energy $(U_H)_{H \in \mathcal{H}}$. To begin with, we take a look at the finite volume case, before we consider the general case.

6.17 Example (Finite Volume Gibbs Measures) Assume that $(\mathfrak{X}, \mathcal{T})$ is bounded and consider some diffuse $\alpha \in \mathfrak{M}(\mathfrak{X})$ and let $(U_H)_{H \in \mathcal{H}}$ be a potential energy that is given through a family of interaction potentials $(\phi_k)_{k \in I}$ (see Example 6.14). We consider $\mathbf{Pop}_{\mathfrak{X}} = \mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \alpha)$ and assume that

$$c := \left(\int \exp(-U_{\mathfrak{X}}(\xi)) \mathbf{Pop}_{\mathfrak{X}}(d\xi) \right)^{-1}$$

is finite, we may then define a distribution P on $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$ through

$$P(E) := \int_E c \cdot \exp(-U_{\mathfrak{X}}(\xi)) \mathbf{Pop}_{\mathfrak{X}}(d\xi) \quad \text{for } E \in \mathcal{N}.$$

In words, configurations with low energies w. r. t. $U_{\mathfrak{X}}$ are more likely to be observed under P than those with high energies.

Note that $f := c \cdot \exp(-U_{\mathfrak{X}}(\cdot))$ is a hereditary probability density. Then, according to Proposition 6.9 and since $(U_H)_{H \in \mathcal{H}}$ is a potential energy, we

¹⁶This potential is due to Lennard-Jones (1924), studying the interaction of molecules. It was used to explore the behavior of noble gases (argon, neon etc.), because it captures much of their essential physics according to empirical results. Misspecifications may still emerge, however, in general particle systems, which is due to neglecting three- and higher-body forces (see for example Hammer *et al.* (2013) and the references therein)

have that $\Gamma := (\Gamma_H)_{H \in \mathcal{H}}$ defined by¹⁷

$$\Gamma_H(\xi, E) := \int_{\mathfrak{N}} \mathbb{1}_E(\eta + \xi_{H^c}) c(H, \xi) \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta),$$

where

$$c(H, \xi) := \left(\int \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta) \right)^{-1}$$

is a specification, the so-called *Gibbs specification (on a bounded state space)*. Note that $\mathbf{Gibbs}(\Gamma) = \{P\}$, see Proposition 6.7 on page 133. \blacklozenge

Using Example 6.17 as a reference, we are now going to impose additional conditions on potential energies to be able to construct corresponding Gibbs specifications for *unbounded* state spaces $(\mathfrak{X}, \mathcal{A})$. First, we have to get rid of the finiteness of the reference states ξ_{H^c} in (6.17). This can be accomplished for so-called *tempered* configurations. These contain sub-configurations whose influence on each other w. r. t. U decreases sufficiently fast in their distance. Second, the existence of the normalizing constants $c(H, \xi)$ for $H \in \mathcal{H}$ can be ensured by requiring $(U_H)_{H \in \mathcal{H}}$ to be *stable*. This condition prevents a collapse of infinitely many particles within any bounded region $H \in \mathcal{H}$.

6.18 Definition (Temperedness) Let $U = (U_H)_{H \in \mathcal{H}}$ be a potential energy. For each $H \in \mathcal{H}$ we call U_H *tempered* at $\xi \in \mathfrak{N}$ and ξ *tempered* w. r. t. U_H iff

$$\lim_{G \uparrow \mathfrak{X}} U_H(\xi_G) \text{ exists in } \mathbb{R} \text{ or } U_H(\xi_G) = \infty \text{ for some } G \in \mathcal{H}.$$

We set

$$D_H^{(t)}(U) := \{\xi \in \mathfrak{N} : U_H \text{ tempered at } \xi\},$$

where t stands for tempered. \blacklozenge

6.19 Definition (Stability) Let $U = (U_H)_{H \in \mathcal{H}}$ be a potential energy. For each $H \in \mathcal{H}$ we call U_H (Ruelle) *stable*¹⁸ at $\xi \in \mathfrak{N}(\mathfrak{X})$ iff first $\eta + \xi_{H^c} \in D_H^{(t)}$ holds for all $\eta \in \mathfrak{N}(H)$ and second there is a constant $L > 0$ such that

$$U_H(\eta + \xi_{H^c}) \geq -L\eta(H) \quad \text{for all } \eta \in \mathfrak{N}(H).$$

We define $D_H^{(s)}(U) := \{\xi \in \mathfrak{N} : U_H \text{ is Ruelle stable at } \xi\}$. \blacklozenge

6.20 Lemma Consider a potential energy $U = (U_H)_{H \in \mathcal{H}}$ given via (6.12). Then, for $\xi \in D_H^{(s)}(U)$ we have

$$\exp(-\alpha(H)) \leq \int_{\mathfrak{N}(\mathfrak{X})} \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}(H, \mathcal{B}(H), \alpha)(d\eta) < \infty.$$

\blacklozenge

¹⁷Here, we additionally use that $(U_H)_{H \in \mathcal{H}}$ is a potential energy, that is, we have $U_{\mathfrak{X}} = U_H + \varphi_{H, \mathfrak{X}}$. Since $\varphi_{H, \mathfrak{X}}$ is measurable w. r. t. \mathcal{F}^H the factor $\exp(-\varphi_{H, \mathfrak{X}})$ cancels out.

¹⁸This notion is due to Ruelle (1969).

Proof: The first inequality is due to $U_H(\mathbf{0} + \xi_{H^c}) = 0$, see (6.12). The second inequality is due to the stability of U at ξ (see also Preston, 1976, Lemma 6.3, page 98). ■

6.21 Example (Strauss Potential Energy) Consider the Strauss potential energy $(U_H^{\text{Strauss}})_{H \in \mathcal{H}}$ from Example 6.15 for $\gamma \in [0, 1]$, $\beta > 0$ and recall that the interaction range is R and hence finite. Therefore, we have for each $\xi \in \mathfrak{N}(\mathfrak{X})$ that $U_H^{\text{Strauss}}(\xi_G) = U_H^{\text{Strauss}}(\xi_{H^R})$ for all $G \supset H^R$, where

$$H^R := \{x \in \mathfrak{X} : \text{dist}(x, H) \leq R\}$$

and $\text{dist}(x, H) := \inf\{d(x, y) : y \in H\}$. Consequently, U_H^{Strauss} is tempered at every configuration, that is,

$$D_H^{(t)}(U^{\text{Strauss}}) = \mathfrak{N}(\mathfrak{X}).$$

Furthermore, $(U_H^{\text{Strauss}})_{H \in \mathcal{H}}$ is stable, because

$$U_H^{\text{Strauss}}(\eta) \geq \log(\beta)\eta(H) \quad \text{for all } \eta \in \mathfrak{N}(H),$$

see also Example 6.15. This implies that the normalizing constants exist. If $(\mathfrak{X}, \mathcal{T})$ is compact, $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ given in Example 6.17 is a specification that determines a unique Gibbs distribution P on $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$. Any random element $\Xi \sim P$ is then called *Strauss point process* and specifically *hard-core Strauss point process* if $\gamma = 0$ (see also Example 6.15). Figure 6.3 and 6.5 show realizations of a Strauss point process on the unit square and the unit 2-sphere, respectively. ◆

6.22 Example (Lennard-Jones Potential Energy) Consider the Lennard-Jones potential energy $(U_H^{\text{LJ}})_{H \in \mathcal{H}}$ from Example 6.16 and recall that the interaction range is infinite.

We consider $(\mathfrak{X}, d) := (\mathbb{R}^k, \|\cdot\|_2)$. Let $(a_n)_{n \in \mathbb{N}}$ be a non-negative sequence with $a_n \uparrow \infty$ for $n \rightarrow \infty$. Furthermore, let $(b_n)_{n \in \mathbb{N}}$ be another non-negative sequence which satisfies a summability condition in accordance with the LJ pair potential (see Preston, 1976, Lemma 6.5, page 104).¹⁹ Define $G_0 := \emptyset$ and $G_n := [-a_n, a_n]^k$ for $n \in \mathbb{N}$,

$$B_n := \{\xi \in \mathfrak{N} : \xi(G_n \setminus G_{n-1}) \leq b_n\}$$

and, finally, $B := \bigcap_{n \in \mathbb{N}} B_n$. Preston (1976, Lemma 6.6, page 105) yields that U_H^{LJ} is tempered and stable on B for every H , that is, $B \subset D_H^{(s)}$ for all $H \in \mathcal{H}$.

If $(\mathfrak{X}, \mathcal{T})$ is compact, $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ given in Example 6.17 is a specification that determines a unique Gibbs distribution P on $(\mathfrak{N}(\mathfrak{X}), \mathcal{N}(\mathfrak{X}))$. Any random element $\Xi \sim P$ is then called *Lennard-Jones point process*. Figure 6.4 and 6.6 show corresponding realizations on the unit square and the unit 2-sphere, respectively. ◆

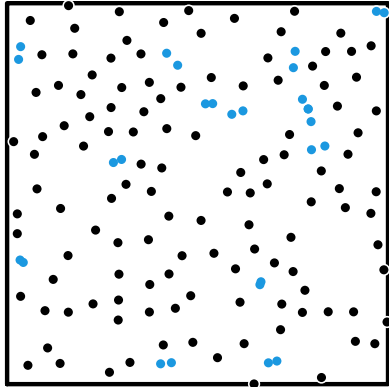


Figure 6.3: Strauss point process realization in the unit square $([0, 1]^2, \|\cdot\|_2)$ w. r. t. parameters $\beta = 500$, $R = 0.05$, and $\gamma = 0.1$. The present point pattern has 14 pairs of R -close points (blue) and 149 points in total (black and blue).²⁰

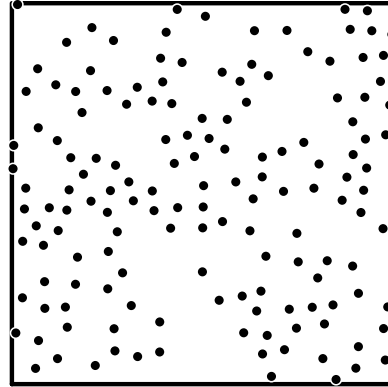
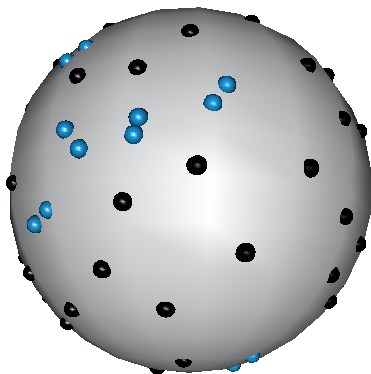
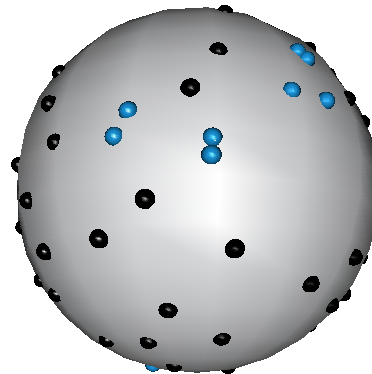


Figure 6.4: Lennard-Jones point process realization observed in the unit square $([0, 1]^2, \|\cdot\|_2)$ w. r. t. parameters $\beta = 100$, $\epsilon = 0.1$, and $\sigma = 0.05$. The present point pattern consists of 150 points.²⁰



Top Half



Bottom Half

Figure 6.5: Strauss point process realization on the unit sphere $\mathbb{S}^2 := \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$ equipped with the great-circle distance. Parameters are chosen to be $R = 0.25$, $\beta = 500$, and $\gamma = 0.1$. The top half of the realization is shown on the left-hand side, the bottom half is shown on the right-hand side. The present realization has 10 pairs of R -close points (blue colored) and 82 points in total (black and blue colored).²¹

¹⁹For a specific choice of the sequences $(a)_{n \in \mathbb{N}}$ and $(b)_{n \in \mathbb{N}}$, we refer to (Preston, 1976, page 107).

²⁰This realization was generated by the Metropolis–Hastings algorithm implemented in the R-package `spatstat` (Baddeley & Turner, 2005).

²¹This realization was generated by the Metropolis–Hastings algorithm with uniform proposal on the sphere which was written by the author.

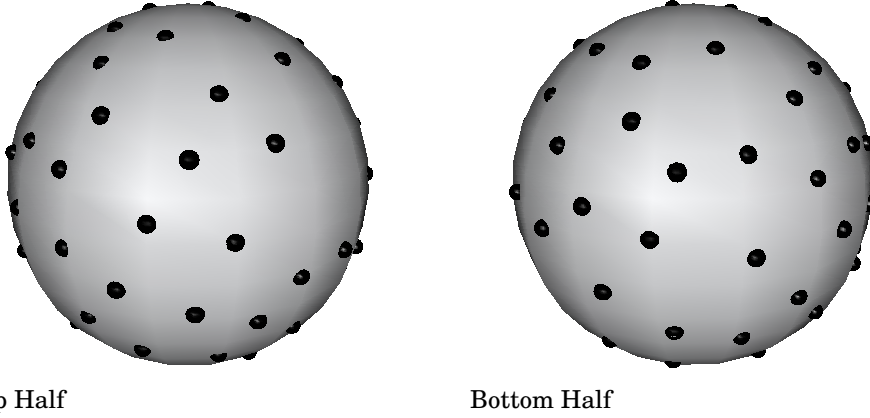


Figure 6.6: Lennard-Jones point process realization on the unit sphere $\mathbb{S}^2 := \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$ equipped with the great-circle distance. Parameters are chosen to be $\beta = 100$, $\epsilon = 5$, and $\sigma = 0.3$. The top half of the realization is shown on the left-hand side, the bottom half is shown on the right-hand side. The present realization has 78 points.²¹

We furthermore define

$$D_H(U) := D_H^{(s)}(U) \cap \{\xi \in \mathfrak{N} : \phi_k(\eta) < \infty \text{ for all } \eta \in \mathfrak{N}_k, \eta \leq \xi_{H^c}, k \in I\}.$$

6.23 Proposition (Gibbs Specification w. r. t. Potential Energy) Let U be a potential energy with tempered and stable configurations, that is, assume $D_H(U) \neq \emptyset$ for every $H \in \mathcal{H}$. With $\mathbf{Pop}_H = \mathbf{Pop}(H, \mathcal{B}(H), \alpha|_H)$ we define

$$c(H, \xi) := \begin{cases} [\int \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta)]^{-1} & \text{for } \xi \in D_H(U), \\ 0 & \text{otherwise} \end{cases}$$

for $\xi \in \mathfrak{N}$ and $H \in \mathcal{H}$. Then, $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ defined through

$$\Gamma_H(\xi, A) := \int_{\mathfrak{N}} \mathbf{1}_A(\eta + \xi_{H^c}) c(H, \xi) \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta)$$

for $\xi \in \mathfrak{N}$ and $A \in \mathcal{N}$ is a specification, called *Gibbs specification* w. r. t. the potential energy U . \blacklozenge

Proof: First, by definition of Γ we have for all $\xi \in D_H(U)$ that $\Gamma_H(\xi, \cdot)$ is a probability distribution on $(\mathfrak{N}, \mathcal{N})$. For $\xi \notin D_H(U)$ we have $\Gamma_H(\xi, \cdot) = \mathbf{0}$. In addition, it is obvious from the definition that $\Gamma_H(\cdot, A)$ is \mathcal{F}^H -measurable and that

$$\Gamma_H(\cdot, A \cap B) = \Gamma_H(\cdot, A) \mathbf{1}_B$$

holds for all $A \in \mathcal{F}_H$ and $B \in \mathcal{F}^H$. Hence, we have that Γ_H is a strict quasi Markov kernel from $(\mathfrak{N}, \mathcal{F}^H)$ to $(\mathfrak{N}, \mathcal{N})$.

Second, to show the Chapman–Kolmogorov equation, we consider $G, H \in \mathcal{H}$ with $G \subset H$. Since U is a potential energy, there is an \mathcal{F}^G -measurable function $\varphi_{G,H}$ such that

$$U_H = U_G + \varphi_{G,H}.$$

Writing for $\xi \in D_H(U)$ and $A \in \mathcal{N}$

$$A(H, \xi) := \{\eta + \xi_{H^c} : \eta \in \mathfrak{N}(H)\},$$

we may compute

$$\begin{aligned} & \Gamma_H \Gamma_G(\xi, A) \\ &= \int_{\mathfrak{N}(H)} \Gamma_G(\eta, A) \Gamma_H(\xi, d\eta) \\ &= \int \Gamma_G(\eta, A) c(H, \xi) \exp(-U_G(\eta + \xi_{H^c}) - \varphi_{G,H}(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta) \\ &= \int \left[\int_{A(G, \eta + \xi_{H^c})} c(G, \eta + \xi_{H^c}) \exp(-U_G(\zeta + \eta_{H \setminus G} + \xi_{H^c})) \mathbf{Pop}_G(d\zeta) \right] \\ & \quad \cdot c(H, \xi) \exp(-U_G(\eta_G + \eta_{H \setminus G} + \xi_{H^c}) - \varphi_{G,H}(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta). \end{aligned}$$

Using the fact

$$\mathbf{Pop}_H = \mathbf{Pop}_G * \mathbf{Pop}_{H \setminus G},$$

we may rewrite the above to

$$\begin{aligned} & \Gamma_H \Gamma_G(\xi, A) \\ &= \int \int \left[\int_{A(G, \eta_2 + \xi_{H^c})} \frac{c(G, \eta_2 + \xi_{H^c})}{\exp(U_G(\zeta + \eta_2 + \xi_{H^c}))} \mathbf{Pop}_G(d\zeta) \right] \\ & \quad \cdot c(H, \xi) \exp(-U_G(\eta_1 + \eta_2 + \xi_{H^c})) \exp(-\varphi_{G,H}(\eta_2 + \xi_{H^c})) \\ & \quad \mathbf{Pop}_G(d\eta_1) \mathbf{Pop}_{H \setminus G}(d\eta_2). \end{aligned}$$

Due to the \mathcal{F}^G -measurability of $\varphi_{G,H}$, we have with $\eta_2 \in \mathfrak{N}(H \setminus G)$

$$\begin{aligned} & \int c(H, \xi) \exp(-U_G(\eta_1 + \eta_2 + \xi_{H^c})) \exp(-\varphi_{G,H}(\eta_2 + \xi_{H^c})) \mathbf{Pop}_G(d\eta_1) \\ &= \frac{c(H, \xi) \exp(-\varphi_{G,H}(\eta_2 + \xi_{H^c}))}{c(G, \eta_2 + \xi_{H^c})}. \end{aligned}$$

and, consequently, in total

$$\begin{aligned} & \Gamma_H \Gamma_G(\xi, A) \\ &= \int \int_{A(G, \eta_2 + \xi_{H^c})} c(H, \xi) \exp(-\varphi_{G,H}(\eta_2 + \xi_{H^c})) \exp(-U_G(\zeta + \eta_2 + \xi_{H^c})) \\ & \quad \mathbf{Pop}_G(d\zeta) \mathbf{Pop}_{H \setminus G}(d\eta_2) \\ &= \int_{A(H, \xi)} c(H, \xi) \exp(-U_H(\eta + \xi_{H^c})) \mathbf{Pop}_H(d\eta) \\ &= \Gamma_H(\xi, A). \end{aligned}$$

Hence, $(\Gamma_H)_{H \in \mathcal{H}}$ is a specification. ■

6.4 Existence of Gibbs Point Processes

In the previous section, we have introduced the important class of Gibbs specifications. For those specifications, we know so far that corresponding Gibbs point processes exist and are even unique if the state space is *bounded*. The purpose of this section is to study the existence of Gibbs point process distributions for a given specification $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ on *unbounded* spaces.

Basically, the idea for proving existence is as simple as that of Proposition 6.11, that is: In the restricted Chapman–Kolmogorov equation

$$(\Gamma_H \Gamma_G)(\xi, \cdot)|_{\mathcal{F}_F} = \Gamma_H(\xi, \cdot)|_{\mathcal{F}_F}$$

with $F, G, H \in \mathcal{H}$ and $G \subset H$, we let H increase to the whole space \mathfrak{X} and hope for convergence. The intuition behind this hope is that we expect that ξ_{H^c} has diminishing influence on $\Gamma_H(\xi, \cdot)|_{\mathcal{F}_F}$ as H goes to \mathfrak{X} . We, therefore, would expect that $\Gamma_H(\xi, \cdot)|_{\mathcal{F}_F}$ “stabilizes” for some ξ , that is, there is a sequence $(\Gamma_{H_n}(\xi, \cdot)|_{\mathcal{F}_F})_{n \in \mathbb{N}}$ which “converges” to some P_F . In doing so, we expect a consistent family $(P_F)_{F \in \mathcal{H}}$ such that we may apply Corollary 5.26 to obtain finally a distribution P on $\mathfrak{N}(\mathfrak{X})$. In order to conclude that P has the correct conditional distributions, we again want to use the Chapman–Kolmogorov equation in combination with taking a suitable (unique) “limit” on both sides.

In this way, however, we encounter some technical difficulties if we consider the case of a general specification rather than just a Markov one and, even more so, in the case of a general state space \mathfrak{X} rather than just the lattice \mathbb{Z}^d . We have to make precise in what sense the concept of “stabilization” and “convergence” may be obtained. To this end, we have to define an appropriate topology on $\text{Prob}(\mathfrak{N}, \mathcal{N})$.

We denote by $\mathfrak{B} = \mathfrak{B}(\mathfrak{N}, \mathcal{N})$ the set of bounded \mathcal{N} -measurable functions $f : \mathfrak{N}(\mathfrak{X}) \rightarrow \mathbb{R}$ and equip it with the supremum norm $\|\cdot\|_\infty$. The intuition that a specification is pointwise stable on \mathcal{F}_H for some $H \in \mathcal{H}$ is originated in the local nature of the events contained in \mathcal{F}_H .²² This notion will be used in a more general sense.

²²A physical motivation and interpretation of these concepts are given in Georgii (2011, page 31): “The motivation for introducing quasilocal specifications arises from the physical idea of a strict separation of microscopic and macroscopic quantities: A microscopic part of a system does not possess any information about the macroscopic state of the system.” Furthermore: “Next we ask for a precise meaning of the term ‘microscopic quantity’. It is natural to say that a function f on \mathfrak{N} is a microscopic quantity if f is arbitrarily close to functions which only depend on finitely many coordinates. There is, however, no canonical interpretation of the word ‘close’. The simplest meaning is ‘close in the uniform norm’.”

6.24 Definition A function $f \in \mathfrak{B}$ is called *local* if and only if it is \mathcal{F}_H -measurable for some $H \in \mathcal{H}$. The set of all local functions is denoted by $\mathfrak{B}_{\text{loc}} = \mathfrak{B}_{\text{loc}}(\mathcal{F})$. \blacklozenge

In fact, locality is too strong for our purposes. We therefore consider test functions which are arbitrarily close to local functions w. r. t. $\|\cdot\|_\infty$, that is, functions in the closure of $\mathfrak{B}_{\text{loc}}$ in $(\mathfrak{B}, \|\cdot\|_\infty)$.

6.25 Definition A function $f : \mathfrak{N} \rightarrow \mathbb{R}$ is called *quasilocal* iff for every $\varepsilon > 0$ there is a $g \in \mathfrak{B}_{\text{loc}}$ such that $\|f - g\|_\infty < \varepsilon$. The set of all quasilocal functions is denoted by $\mathfrak{B}_{\text{qloc}} = \mathfrak{B}_{\text{qloc}}(\mathcal{F})$. \blacklozenge

6.26 Definition The topology on $\text{Prob}(\mathfrak{N}, \mathcal{N})$ generated by the mappings $P \mapsto \int f dP$ for all $f \in \mathfrak{B}_{\text{qloc}}$ is called *topology of local convergence* and is denoted by \mathcal{T}_{loc} . \blacklozenge

Now we are going to transfer these notions to kernels. To this end, consider a bounded kernel K from $(\mathfrak{N}, \mathcal{N})$ to $(\mathfrak{N}, \mathcal{N})$ as an operator applied to bounded measurable functions

$$\begin{aligned} K : \mathfrak{B}(\mathfrak{N}, \mathcal{N}) &\rightarrow \mathfrak{B}(\mathfrak{N}, \mathcal{N}) \\ f &\mapsto \int f(\eta) K(\cdot, d\eta). \end{aligned}$$

The notions local and quasilocal are then used for kernels if they preserve the corresponding property. That means:

6.27 Definition A kernel K from $(\mathfrak{N}, \mathcal{N})$ to $(\mathfrak{N}, \mathcal{N})$ is called *local* iff we have $\int f(\eta) K(\cdot, d\eta) \in \mathfrak{B}_{\text{loc}}$ for all $f \in \mathfrak{B}_{\text{loc}}$. We call K *quasilocal* iff $\int f(\eta) K(\cdot, d\eta) \in \mathfrak{B}_{\text{qloc}}$ for all $f \in \mathfrak{B}_{\text{qloc}}$. \blacklozenge

According to these measure theoretic and topological preparations, we are now able to formulate a general existence result.

6.28 Theorem Let $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ be a *quasilocal* specification, that is, Γ_H is quasilocal for every $H \in \mathcal{H}$. Furthermore, let $(\xi^H)_{H \in \mathcal{H}}$ be a family of point patterns such that $\{\Gamma_H(\xi^H, \cdot) : H \in \mathcal{H}\}$ is relatively sequentially compact w. r. t. the topology \mathcal{T}_{loc} .

Then, **Gibbs**(Γ) $\neq \emptyset$, i. e., there is a Gibbs measure with specification Γ . \blacklozenge

Proof: Consider an increasing sequence of sets $(H_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ such that $H_n \uparrow \mathfrak{X}$ for $n \rightarrow \infty$. By the relative sequential compactness w. r. t. \mathcal{T}_{loc} of

$$\{\Gamma_H(\xi^H, \cdot) : H \in \mathcal{H}\}$$

there is a subsequence $(H_{n_k})_{k \in \mathbb{N}}$ and a distribution P on $(\mathfrak{N}, \mathcal{N})$ such that

$$P = \mathcal{T}_{\text{loc}} - \lim_{k \rightarrow \infty} \Gamma_{H_{n_k}}(\xi^{H_{n_k}}, \cdot).$$

Furthermore, for $F, H \in \mathcal{H}$ we have by the quasilocalness of Γ that in particular $\Gamma_H(\cdot, A) \in \mathfrak{B}_{\text{qloc}}$ holds for all $A \in \mathcal{F}_F$ and therefore

$$\int \Gamma_H(\eta, A) \Gamma_{H_{n_k}}(\xi^{H_{n_k}}, d\eta) \rightarrow \int \Gamma_H(\eta, A) P(d\eta) \quad \text{for } n \rightarrow \infty$$

and for all $A \in \mathcal{F}_F$. Hence, we have

$$\begin{aligned} P\Gamma_H(A) &= \mathcal{T}_{\text{loc}} - \lim_{k \rightarrow \infty} \Gamma_{H_{n_k}} \Gamma_H(\xi^{H_{n_k}}, A) \\ &= \mathcal{T}_{\text{loc}} - \lim_{k \rightarrow \infty} \Gamma_{H_{n_k}}(\xi^{H_{n_k}}, A) \\ &= P(A) \end{aligned}$$

for all $A \in \mathcal{F}_F$. Since F is arbitrary and $\bigcup_{F \in \mathcal{H}} \mathcal{F}_F$ is an \cap -stable generator of \mathcal{N} , we have $P\Gamma_H = P$. Furthermore, H was arbitrarily chosen which yields $P \in \mathbf{Gibbs}(\Gamma)$ due to Remark 6.4 on page 127. \blacksquare

According to this result, we immediately look for a sufficient condition that a set $M \subset \text{Prob}(\mathfrak{N}, \mathcal{N})$ is relatively sequentially compact w. r. t. \mathcal{T}_{loc} .

6.29 Proposition Consider a set $M \subset \text{Prob}(\mathfrak{N}, \mathcal{N})$ and for every $H \in \mathcal{H}$ we define $M|_{\mathcal{F}_H} := \{\mu|_{\mathcal{F}_H} : \mu \in M\}$. Assume, furthermore, that for every $H \in \mathcal{H}$ there is a finite measure $\mu_H \in \text{Prob}(\mathfrak{N}, \mathcal{F}_H)$ such that $M|_{\mathcal{F}_H}$ is *uniformly absolutely continuous* w. r. t. to μ_H , that is, for every $\varepsilon > 0$ there is a $\delta > 0$ such that for each $F \in \mathcal{F}_H$ with $\mu_H(F) < \delta$ we have $\nu(F) < \varepsilon$ for all $\nu \in M|_{\mathcal{F}_H}$. Then, M is relatively sequentially compact w. r. t. \mathcal{T}_{loc} . \blacklozenge

Proof: Consider $(\nu_n)_{n \in \mathbb{N}} \in M^{\mathbb{N}}$ and let $(H_n)_{n \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}}$ be a sequence of bounded sets such that $H_n \uparrow \mathfrak{X}$ for $n \rightarrow \infty$.

Since $M|_{\mathcal{F}_H}$ is uniformly absolutely continuous, we have that $(\nu_n)_{n \in \mathbb{N}}|_{\mathcal{F}_H}$ has a weak convergent subsequence (see e. g. Bogachev, 2007, Theorem 4.7.18, page 285). A diagonal argument leads to a subsequence $(\nu_{n_j})_{j \in \mathbb{N}}$ such that for each $k \in \mathbb{N}$ there is a measure $\lambda_k \in \text{Prob}(\mathfrak{N}, \mathcal{F}_{H_k})$ such that $\lambda_k(A) = \lim_{j \rightarrow \infty} \nu_{n_j}(A)$ for all $A \in \mathcal{F}_{H_k}$.

Obviously, $(\lambda_k)_{k \in \mathbb{N}}$ is consistent, that is, $\lambda_{k+1}|_{\mathcal{F}_{H_k}} = \lambda_k$ for all k . According to Corollary 5.26 on page 114, there is a measure $\lambda \in \text{Prob}(\mathfrak{N}, \mathcal{N})$ such that $\lambda|_{\mathcal{F}_{H_k}} = \lambda_k$ for all $k \in \mathbb{N}$. Hence, $\lambda(A) = \lim_{j \rightarrow \infty} \nu_{n_j}(A)$ for all $A \in \mathcal{F}_H$ and all H . \blacksquare

By Theorem 6.28 and Proposition 6.29, we have already the existence of Gibbs distribution for specifications with a hard-core. As a first non-trivial, inhibitory Gibbs point process, we obtain the hard-core Strauss point process.

6.30 Example (Hard-Core Strauss Point Process) Consider the hard-core Strauss potential energy $(U_H^{\text{Strauss}})_{H \in \mathcal{H}}$ from Example 6.21 and note that

$D_H(U^{\text{Strauss}}) \neq \emptyset$. Hence, the Gibbs specification $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ w. r. t. the Strauss potential energy exists (see Proposition 6.23). Furthermore, one can show that $(\Gamma_H)_{H \in \mathcal{H}}$ is even a quasilocal specification (see Nguyen & Zessin, 1976, Lemma 3) and that $\{\Gamma_H(\mathbf{0}, \cdot) : H \in \mathcal{H}, H \supset G\} |_{\mathcal{F}_G}$ is uniformly absolutely continuous for all $G \in \mathcal{H}$. Hence, $\mathbf{Gibbs}(\Gamma) \neq \emptyset$ follows from Theorem 6.28 and Proposition 6.29. \blacklozenge

The Gibbs specification that corresponds to the Lennard-Jones potential energy is less well-behaved. Hence, the notion of quasilocal specifications and the requirement for relative sequential compactness w. r. t. \mathcal{T}_{loc} are inappropriate for introducing the Lennard-Jones point process on the plane. More general concepts are therefore needed.

Let \mathcal{E}_δ be the system of countable intersections of sets in $\mathcal{E} := \bigcup_{H \in \mathcal{H}} \mathcal{F}_H$. In what follows we consider by $\mathcal{C} \subset \mathcal{E}_\delta$ a countable set-system that is directed upwards w. r. t. \subset such that $\bigcup \mathcal{C} = \mathfrak{N}$. Furthermore, we consider on \mathfrak{B} the \mathcal{C} -indexed family of semi-norms given by

$$\|f\|_{C, \infty} := \|f \mathbf{1}_C\|_\infty \quad \text{for } f \in \mathfrak{B}$$

and for $C \in \mathcal{C}$. By that we can define the so-called \mathcal{C} -local functions.

6.31 Definition A function $f \in \mathfrak{B}$ is called \mathcal{C} -quasilocal iff f is in the closure of $\mathfrak{B}_{\text{loc}}$ w. r. t. $\tau(\|\cdot\|_{C, \infty} : C \in \mathcal{C})$, that is: for all $\varepsilon > 0$ and $C \in \mathcal{C}$ there is a $g \in \mathfrak{B}_{\text{loc}}$ such that $\|f - g\|_{C, \infty} < \varepsilon$. The set of all \mathcal{C} -quasilocal functions is denoted by $\mathfrak{B}_{\text{qloc}}^{\mathcal{C}}$. \blacklozenge

Obviously, we may choose $\mathcal{C} := \{\mathfrak{N}(\mathfrak{X})\}$, which yields $\mathcal{T}_{\text{loc}}^{\mathcal{C}} = \mathcal{T}_{\text{loc}}$ and $\mathfrak{B}_{\text{qloc}}^{\mathcal{C}} = \mathfrak{B}_{\text{qloc}}$. The topology of \mathcal{C} -local convergence is then given by

$$\mathcal{T}_{\text{loc}}^{\mathcal{C}} := \tau(P \mapsto \int f dP : f \in \mathfrak{B}_{\text{qloc}}^{\mathcal{C}}).$$

In an analogous way to the above existence result, one can now state.

6.32 Theorem Let $\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$ be a \mathcal{C} -quasilocal specification, that is, Γ_H maps \mathcal{C} -quasilocal functions to \mathcal{C} -quasilocal functions. Assume, furthermore, that there is a family $(\xi^H)_{H \in \mathcal{H}}$ such that $\{\Gamma_H(\xi^H, \cdot) : H \in \mathcal{H}\}$ is relatively sequentially compact w. r. t. the topology $\mathcal{T}_{\text{loc}}^{\mathcal{C}}$. Then, $\mathbf{Gibbs}(\Gamma) \neq \emptyset$. \blacklozenge

Proof: The proof works in a similar way as for Theorem 6.28. \blacksquare

We finally note the following for the existence of the Lennard-Jones point process in the unbounded case.

6.33 Example (Lennard-Jones Point Process) Consider the Lennard-Jones potential energy $(U_H^{\text{LJ}})_{H \in \mathcal{H}}$ from Example 6.21 and note that we have $D_H(U^{\text{LJ}}) \neq \emptyset$. Hence, the corresponding Gibbs specification $(\Gamma_H)_{H \in \mathcal{H}}$

w. r. t. the Lennard-Jones potential energy exists (see Proposition 6.23). Furthermore, one can show that there is a $\mathcal{C} \subset \mathcal{E}_\delta$ such that $(\Gamma_H)_{H \in \mathcal{H}}$ is a \mathcal{C} -quasilocal specification and that $\{\Gamma_H(\mathbf{0}, \cdot) : H \in \mathcal{H}\}$ is relatively sequentially compact w. r. t. $\mathcal{T}_{\text{loc}}^{\mathcal{C}}$ (see Preston, 1976, Chapter 6). Hence, $\text{Gibbs}(\Gamma) \neq \emptyset$ by Theorem 6.32. \blacklozenge

Variational Equation for Gibbs Point Processes

We close this section on (infinite volume) Gibbs point processes with the variational equation given by Baddeley & Dereudre (2013). This equation concerns the behavior of functionals of point patterns, when the latter are subjected to an infinitesimally perturbation.

To this end, we consider the case $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \alpha) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$ and define the following notion of differentiability.

6.34 Definition (Differentiability at Point Patterns) Let $h : \mathfrak{N}(\mathfrak{X}) \rightarrow \mathbb{R}$ and let $\xi_0 \in \mathfrak{N}(\mathfrak{X})$. We call h *differentiable* at $(x, \xi_0) \in \mathfrak{X} \times \mathfrak{N}(\mathfrak{X})$ iff there is an open neighborhood $V_x \subset \mathfrak{X}$ of x such that

$$h(u, \xi_0) := h((\xi_0 \setminus \{x\}) \cup \{u\}) \quad \text{for } u \in V_x$$

is differentiable at x . In this case we write $\nabla h(x, \xi_0)$ for the gradient of h at (x, ξ_0) . \blacklozenge

6.35 Remark (Differentiability of Potential Energies) Consider a potential energy $U = (U_H)_{H \in \mathcal{H}}$. For some open $G \in \mathcal{H}$, assume that U_G is differentiable at $(x, \xi_0) \in G \times \mathfrak{N}(\mathfrak{X})$. Then for any $H \in \mathcal{H}$ with $G \subset H$, we have that $U_H = U_G + \varphi_{G,H}$ for some \mathcal{F}^G -measurable function $\varphi_{G,H}$, see Definition 6.13 on page 142. Hence, U_H is also differentiable at (x, ξ_0) and furthermore $\nabla U_G(x, \xi_0)$ does not depend on $H \supset G$. We will therefore write $\nabla U(x, \xi_0)$. \blacklozenge

The following notation will be used for convenience. For $u \in \mathbb{R}$ and a vector $x \in \mathbb{R}^d$, we write

$$(u, x_{[i]}) := (u \mathbb{1}_{\{i\}}(j) + x_j \mathbb{1}_{\{i\}^c}(j))_{j \in \{1, \dots, d\}}, \in \mathbb{R}^d$$

where $i \in \{1, \dots, d\}$.

6.36 Definition A measurable function g from $(\mathfrak{X} \times \mathfrak{N}(\mathfrak{X}), \mathcal{B}(\mathfrak{X}) \otimes \mathcal{N}(\mathfrak{X}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called *regularizing* w. r. t. a potential energy $U = (U_H)_{H \in \mathcal{H}}$ iff for every $\xi \in \mathfrak{N}(\mathfrak{X})$

- (a) the map $g(\cdot, \xi)$ is λ^d -almost everywhere differentiable;
- (b) for every compact *rectangle* $H \in \mathcal{H}(\mathfrak{X})$ and for λ^d -almost all $x \in H$, the map

$$u \mapsto g((u, x_{[i]}), \xi) \exp(-U_H(\xi \cup \{(u, x_{[i]})\}))$$

is absolutely continuous²³ as a map from $\text{pr}_i(H)$ to $\mathbb{R} \cup \{\infty\}$ for every $i = 1, \dots, d$.

We denote the class of all functions that are regularizing w. r. t. $(U_H)_{H \in \mathcal{H}}$ by $\mathfrak{R}(U)$. Furthermore, we write $\mathfrak{R}_C(U)$ for the class of all $g \in \mathfrak{R}(U)$ such that $g(\cdot, \xi)$ have a compact support for all ξ . \blacklozenge

The variational equation given in Baddeley & Dereudre (2013, Proposition 1) is an identity for Gibbs point processes, which is a reminiscent of Stein's identity for exponential families on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, see also Lehmann & Casella (1998, Lemma 5.15 in Section 1.5, page 31).

6.37 Proposition (Variational Equation) Let P be a Gibbs point process distribution on \mathfrak{X} w. r. t. the potential energy $U = (U_H)_{H \in \mathcal{H}}$. Then, for every $g \in \mathfrak{R}_C(U)$ such that

$$\int |\nabla g(x, \xi)| + |g(x, \xi) \nabla U(x, \xi)| \mathbf{C}^1(\mathbf{d}(x, \xi)) < \infty$$

we have the identity

$$\int \nabla g(x, \xi) \mathbf{C}^1(\mathbf{d}(x, \xi)) = \int g(x, \xi) \nabla U(x, \xi) \mathbf{C}^1(\mathbf{d}(x, \xi)).$$

\blacklozenge

Proof: With integration by parts, see Baddeley & Dereudre (2013, Proposition 1). \blacksquare

In the stationary case, one obtains the following variational equation.

6.38 Proposition Let P be a stationary Gibbs point process distribution on \mathbb{R}^d for an energy function $U = (U_H)_{H \in \mathcal{H}}$ that is *shift invariant*, that is, $U_H(\xi) = U_{\mathbf{t}_v(H)}(\xi \circ \mathbf{t}_v^{-1})$ for $\xi \in \mathfrak{N}(\mathbb{R}^d)$ and all $v \in \mathbb{R}^d$ and all $H \in \mathcal{H}$.

Then, for $g \in \mathfrak{R}(U)$ such that $g(x, \xi) = g(\mathbf{t}_v(x), \xi \circ \mathbf{t}_v^{-1})$ for all $v \in \mathbb{R}^d$ and all ξ and satisfying

$$\int |\nabla g(\mathbf{0}, \xi)| + |g(\mathbf{0}, \xi)| + |g(\mathbf{0}, \xi) \nabla U(\mathbf{0}, \xi)| \mathbf{Pa}^{10}(\mathbf{d}\xi) < \infty$$

we have

$$\int \nabla g(\mathbf{0}, \xi) \mathbf{Pa}^{10}(\mathbf{d}\xi) = \int g(\mathbf{0}, \xi) \nabla U(\mathbf{0}, \xi) \mathbf{Pa}^{10}(\mathbf{d}\xi).$$

\blacklozenge

Proof: See Baddeley & Dereudre (2013, Proposition 2) \blacksquare

²³Recall that a function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is *absolutely continuous* on $[a, b]$ iff it is differentiable λ^d -almost everywhere on $[a, b]$ and for any $z \in [a, b]$ we have $\varphi(z) - \varphi(a) = \int \varphi'(t) \lambda^d(dt)$, see for example Dudley (2002).

CHAPTER 7

Statistics for Parametric Gibbs Point Process Models

In this chapter we are concerned with statistical inference for Gibbs point process models. To this end, we consider a compact observation window $W \subset \mathfrak{X}$, where (\mathfrak{X}, d) is a metric space which is assumed to be locally compact and second countable. Furthermore, let μ be a σ -finite reference measure on $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}, d))$, where $\mathcal{B}(\mathfrak{X}, d)$ denotes the Borel- σ -algebra w. r. t. the topology generated by the metric d .

For some $\xi \in \mathfrak{N}(\mathfrak{X})$, let $\mathfrak{P}_{W, \xi} = \{P_{\vartheta, W, \xi} : \vartheta \in \Theta\}$ be a family of Gibbs distributions on $(\mathfrak{N}(W), \mathcal{N}(W))$ with a corresponding parametrized potential energy $U_W(\cdot + \xi_{W^c}) := -\langle \vartheta, S^{W, \xi}(\cdot) \rangle$, where $S^{W, \xi}$ is \mathbb{R}^k -valued and each component is given through an $\mathbb{N} \supset I_j$ -indexed family of interaction potentials $(\phi_k^{(j)})_{k \in I_j}$ as

$$S_j^{W, \xi}(\eta) = \sum_{k \in I_j} \sum_{\substack{\zeta \in \mathfrak{N}_k \\ \zeta \leq \eta + \xi_{W^c} \\ \zeta(W) > 0}} \phi_k^{(j)}(\zeta) \quad \text{for } \eta \in \mathfrak{N}(W),$$

for $j = 1, \dots, k$. Hence, $S^{W, \xi}$ depends on ξ through ξ_{W^c} which is a (known) boundary point pattern on W^c . In what follows, we assume that the interaction range is *finite*, that is, in the present situation we assume

$$R := \sup \{ \text{diam}(\eta) : \phi_k^{(j)}(\eta) \neq 0, \eta \in \mathfrak{N}_k(\mathfrak{X}), k \in I_j, 1 \leq j \leq k \} < \infty.$$

In accordance with the previous chapter, we furthermore consider the case where each $\vartheta \in \Theta$ determines a distribution $P_{\vartheta, W, \xi}$ in $\mathfrak{P}_{W, \xi}$ that has a density w. r. t. the Poisson point process distribution of the form

$$f_{\vartheta}(\eta) := c(\vartheta, W) g_{W, \xi}(\eta) \exp \left(\langle \vartheta, S^{W, \xi}(\eta) \rangle \right) \quad \text{for } \eta \in \mathfrak{N}(W). \quad (7.1)$$

The observation window W may relate to the state space \mathfrak{X} in two different ways and, hence, causes a different character of the statistical model. On

the one hand, we consider the case $W = \mathfrak{X}$ the so-called *empty boundary condition*, and on the other hand, we may have $W \subsetneq \mathfrak{X}$, that is, the so-called *non-empty boundary condition*.

In total we have an exponential family $\mathfrak{P}_{W,\xi}$ with complete sufficient statistic $S^{W,\xi}$. In what follows, we furthermore assume that \mathfrak{P} is k -parametric (of full rank, for a definition see page 27) and that $S^{W,\xi} \in L_2$. Hence, we can make use of the results¹ of Part II. We first give a sketch of some estimation methods, namely, the *maximum likelihood*, the *maximum pseudo-likelihood*, and the *variational estimation* procedures.

7.1 Maximum Likelihood Estimation

According to (7.1) the log-likelihood $\vartheta \mapsto \log f_\vartheta$ takes the form

$$\ell(\vartheta) := \ell_{W,\xi}(\eta, \vartheta) := \langle \vartheta, S^{W,\xi}(\eta) \rangle + \log c(\vartheta, W) \quad \text{for } \vartheta \in \Theta. \quad (7.2)$$

The goal is to find maximizers of that log-likelihood which are called *maximum likelihood estimators* (MLE) for the parameter \mathbf{id}_Θ . For the existence and uniqueness of an MLE, see for example Barndorff-Nielsen (1978).

We recall from Section 2.1 that the MLE of the observation $\eta \in \mathfrak{N}(W)$ and w. r. t. the boundary point pattern $\xi_{W^c} \in \mathfrak{N}(W^c)$ is implicitly given as the $\hat{\vartheta}_{\text{ML}}(\eta, \xi_{W^c}) \in \Theta$ such that (see also Proposition 2.5, page 32)

$$\mathbf{E}_{\hat{\vartheta}_{\text{ML}}(\eta, \xi_{W^c})} S^{W,\xi} = S^{W,\xi}(\eta). \quad (7.3)$$

Hence, $\hat{\vartheta}_{\text{ML}}(\cdot, \xi_{W^c})$ is measurable w. r. t. the complete sufficient statistic $S^{W,\xi}$. By the next proposition below, we furthermore note the following interesting fact about the measurability of the MLE. Let

$$W^R := \{x \in \mathfrak{X} : \text{dist}(x, W) \leq R\},$$

then for all $A \in \mathcal{H}(\mathfrak{X})$ with $W \subset A \subset W^R$ the above MLE is also measurable w. r. t. $S^{A,\xi}(\cdot + \xi_{A \cap W^c})$ and, hence, measurable w. r. t.

$$\bigcap_{\substack{A \in \mathcal{H}(\mathfrak{X}) \\ W \subset A \subset W^R}} \sigma(S^{A,\xi}(\cdot + \xi_{A \cap W^c})).$$

By a simulation study in Section 7.4, we will see that this is usually not the case for the regularized Rao–Blackwellization of an alternative estimation procedure.

¹Here, “results” generally refers to the introduced concepts such as π -completeness, which holds in the present situation for suitable π s (see Proposition 3.21, page 64), and considerations about the LS-norms which we introduced in Section 3.2. The goal is to look for regularized Rao–Blackwellizations in terms of Theorems 4.16 and 4.20, see also Example 4.19.

7.1 Proposition Let $\xi \in \mathfrak{N}(\mathfrak{X})$ and let $\mathfrak{P}_{W,\xi} = \{P_{\vartheta,W,\xi} : \vartheta \in \Theta\}$ be a Gibbs point process model with *finite* interaction range R as considered at the beginning of this chapter. Furthermore, we assume that $\mathfrak{X} = W^R$. Then, we have that $\widehat{\vartheta}_{\text{ML}}(\cdot, \xi_{W^c})$ is measurable w. r. t. $\sigma(S^{A,\xi}(\cdot + \xi_{A \cap W^c}))$ for each $A \in \mathcal{H}(\mathfrak{X})$ with $W \subset A \subset \mathfrak{X}$. \blacklozenge

Proof: For $j = 1, \dots, k$ let us define

$$V^{(j)}(\eta) = \sum_{k \in I} \sum_{\substack{\zeta \in \mathfrak{N}_k \\ \zeta \leq \eta}} \phi_k^{(j)}(\zeta) \quad \text{for } \eta \in \mathfrak{N}(\mathfrak{X}).$$

Then, we have for each $A \in \mathcal{H}(\mathfrak{X})$ with $W \subset A \subset \mathfrak{X}$

$$S_j^{A,\xi}(\zeta) = V^{(j)}(\zeta + \xi_{A^c}) - V^{(j)}(\xi_{A^c}) \quad \text{for } \zeta \in \mathfrak{N}(A) \quad (7.4)$$

and $j = 1, \dots, k$. We recall that for $\eta \in \mathfrak{N}(W)$ and the boundary point pattern $\xi_{W^c} \in \mathfrak{N}(W^c)$ the maximum likelihood estimate $\widehat{\vartheta}_{\text{ML}}(\eta, \xi_{W^c})$ is given by solving

$$\mathbf{E}_{\vartheta} S^{W,\xi} = S^{W,\xi}(\eta) \quad (7.5)$$

in $\vartheta \in \Theta$. Equivalently, we get by using (7.4)

$$\mathbf{E}_{\vartheta} V^{(j)}(\cdot + \xi_{W^c}) = V^{(j)}(\eta + \xi_{W^c}) \quad \text{for } j = 1, \dots, k,$$

that is,

$$\mathbf{E}_{\vartheta} V^{(j)}(\cdot + \xi_{A \cap W^c} + \xi_{A^c}) = V^{(j)}(\eta + \xi_{A \cap W^c} + \xi_{A^c}) \quad \text{for } j = 1, \dots, k,$$

and, finally, using (7.4) again

$$\mathbf{E}_{\vartheta} S^{A,\xi}(\cdot + \xi_{A \cap W^c}) = S^{A,\xi}(\eta + \xi_{A \cap W^c}) \quad \text{for } j = 1, \dots, k. \quad (7.6)$$

Hence, $\widehat{\vartheta}_{\text{ML}}(\eta, \xi_{W^c})$ is determined by the right-hand side of (7.6) for each A with $W \subset A \subset \mathfrak{X}$ and, thus, the claim is shown. \blacksquare

In what follows we assume that the complete sufficient statistic satisfies $S^{W,\xi} \in L_2(\mathfrak{N}(W), \mathcal{N}(W), \mathfrak{P})$. Then, the root of the score can be computed by a combination of Monte Carlo methods and the Newton–Raphson algorithm. More precisely, the corresponding gradient and Hessian of the log-likelihood at a certain $\vartheta \in \Theta$ (see Proposition 2.5), that is,

$$\begin{aligned} \frac{\partial}{\partial \vartheta} \ell_{W,\xi}(\eta, \vartheta) &= S^{W,\xi}(\eta) - \mathbf{E}_{\vartheta} S^{W,\xi}, \\ \frac{\partial^2}{\partial \vartheta \partial \vartheta^{\top}} \ell_{W,\xi}(\eta, \vartheta) &= -\text{Var}_{\vartheta} S^{W,\xi}, \end{aligned}$$

are approximated by estimators of the corresponding expectation and variance of the sufficient statistic. To this end, Monte Carlo (MC) samples of

ϑ are generated and, consequently, one may use the empirical expectation and variance of the corresponding sufficient statistic values. Note that because of the Newton–Raphson algorithm, this MC sample generation process has to be performed in each iteration which is computationally quite demanding (see for example Møller & Waagepetersen, 2004).

To reduce the simulation effort, the so-called *importance sampling technique* can be used as long as the parameter iteration remains within a certain neighborhood. For details, we refer again to Møller & Waagepetersen (2004) and references therein. Further improvements that concern more complex sampling strategies which also recycle MC samples from previous iterations are examined in Habeck *et al.* (2020).

7.2 Maximum Pseudo-Likelihood Estimation

In a fundamental paper Besag (1975) propose to analyze the interaction of spatial lattice processes statistically by only accounting for *local* dependence structure instead of examining the overall joint distribution. According to Møller & Waagepetersen (2004, page 171), this idea was first applied by Besag (1977); Besag *et al.* (1982) to the situation of a Strauss point process and leads to a general *pseudo-likelihood* for point processes introduced by Ripley (1988, page 52 ff.). For notational convenience, we consider the special case of the empty boundary condition, the non-empty boundary condition can be considered analogously.

The Pseudo-Likelihood

To motivate the pseudo-likelihood, we follow Jensen & Møller (1991, page 446 f.), see also Møller & Waagepetersen (2004, page 171 f.). To this end, let $W \subset \mathbb{R}^k$ be the sampling window and consider a dissecting system² $(\mathcal{D}_k)_{k \in \mathbb{N}}$ of W . For some $D \subset W$ we denote by $f_{\vartheta, W}(\xi_D | \xi_{D^c})$ the conditional density of $P_{\vartheta, W}$ given \mathcal{F}_{D^c} evaluated at ξ_D and ξ_{D^c} , respectively. We then define

$$\text{PL}_W(\xi, \vartheta) := \exp(-\lambda(W)) \cdot \lim_{k \rightarrow \infty} \prod_{D \in \mathcal{D}_k} f_{\vartheta, W}(\xi_D | \xi_{D^c}) \quad \text{for } \xi \in \mathfrak{R}(W), \vartheta \in \Theta$$

which is called *pseudo-likelihood* w. r. t. \mathfrak{P}_W .

²A *dissecting system* of $W \subset \mathbb{R}^k$ is a sequence of finite partitions of W , denoted by $(\mathcal{D}_k)_{k \in \mathbb{N}}$, such that: (1) If $k \leq l$ and $D \in \mathcal{D}_l$ there is an $E \in \mathcal{D}_k$ with $D \subset E$ and (2) $|\mathcal{D}_k| \rightarrow \infty$ for $k \rightarrow \infty$ and $|\mathcal{D}_k| \cdot \max\{\lambda(D) : D \in \mathcal{D}_k\} \rightarrow 0$ for $k \rightarrow \infty$.

7.2 Theorem (Jensen & Møller (1991)) For almost all $\xi \in \mathfrak{N}(W)$ the pseudo-likelihood w. r. t. \mathfrak{P}_W is well-defined and given by

$$\text{PL}_W(\xi, \vartheta) = \exp\left(-\int_W \lambda_\vartheta(u|\xi) \lambda(du)\right) \cdot \prod_{u \in \xi_W} \lambda_\vartheta(u|\xi \setminus \{u\}) \quad \text{for } \vartheta \in \Theta, \quad (7.7)$$

where λ_ϑ denotes the Papangelou conditional intensity³. ◆

Proof: See Jensen & Møller (1991, Theorem 2.2, page 447 f.) ■

7.3 Example (Poisson Point Process Model) Consider the homogeneous Poisson point process model $\mathfrak{P} = \{\text{Pop}(\vartheta \cdot \lambda) : \vartheta \in \mathbb{R}_+\}$ on $(\mathfrak{N}(W), \mathcal{N}(W))$. According to Example 5.45, we have that $\lambda_\vartheta(u|\xi) = \vartheta$ for every $\xi \in \mathfrak{N}(W)$ and $u \in W$. Hence, we get for $\log \text{PL}_W(\cdot, \xi)$, see also (7.7),

$$\log \text{PL}_W(\xi, \vartheta) = \vartheta \cdot \xi_W(\mathfrak{X}) - \vartheta \cdot \lambda(W) \quad \text{for } \xi \in \mathfrak{N}(W), \vartheta \in \Theta,$$

which corresponds to the likelihood (7.2). ◆

Roughly speaking, the pseudo-likelihood just accounts for dependences only locally and coincides with the likelihood for a Poisson point process (see Example 7.3). If the dependence structure is weak, the pseudo-likelihood is expected to be close to the “real” likelihood (7.2). Furthermore, PL_W deserves the name *pseudo-likelihood*, since it shares some (regular) properties of a real likelihood, see for example the next proposition.

7.4 Proposition (Jensen & Møller (1991)) For $\xi \in \mathfrak{N}(W)$ we have for the *log-pseudo-likelihood* $\text{pl}_W := \log \text{PL}_W$ the following statements:

- (a) The log-pseudo-likelihood function $\vartheta \mapsto \text{pl}_W(\vartheta, \xi)$ is concave for all ξ .
- (b) We have that $\vartheta \mapsto \mathbf{E}_{\vartheta_0}(\text{pl}_W(\Xi_W \cup \Xi_{W^c}) | \Xi_{W^c} = \xi_{W^c})$ is strictly concave and attains its maximum in ϑ_0 if and only if for all $\vartheta \in \Theta \setminus \{\vartheta_0\}$

$$\text{Pop}_W \left(\xi \in \mathfrak{N}(W) : \int \mathbf{1}\{\lambda_\vartheta(u|\xi \cup \eta) \neq \lambda_{\vartheta_0}(u|\xi \cup \eta)\} \lambda(du) > 0 \right) > 0.$$

◆

Proof: See Jensen & Møller (1991, Proposition 2.3, page 449) ■

Given a data point pattern $\xi \in \mathfrak{N}(W)$, we call a maximizer of $\text{PL}_W(\xi, \cdot)$ the *maximum pseudo-likelihood (MPL) estimate* of ξ w. r. t. the model \mathfrak{P}_W which we denote by $\hat{\vartheta}_{\text{MPL}}(\xi)$. For the Poisson point process model, we note the following.

7.5 Example (Poisson Point Process Model) Consider the situation of Example 7.3 where we have seen that the likelihood and pseudo-likelihood

³See Remark and Definition 5.44 on page 122.

function and, consequently, the MLE and MPLE coincide. Furthermore, we have that

$$\widehat{\vartheta}_{\text{ML}}(\xi) = \widehat{\vartheta}_{\text{MPL}}(\xi) = \frac{\xi(W)}{\lambda(W)} \quad \text{for } \xi \in \mathfrak{N}(W),$$

that is, the MLE and MPLE are explicitly known. \blacklozenge

To find an MPLE w.r.t. an observation $\xi \in \mathfrak{N}(W)$, we consider the *pseudo-likelihood estimating equation*

$$\frac{\partial}{\partial \vartheta} \text{pl}_W(\vartheta, \xi) = \mathbf{0}$$

which corresponds to

$$\frac{\partial}{\partial \vartheta} \int_W \lambda_\vartheta(u|\xi) \lambda(du) = \sum_{x \in \xi} \frac{\partial}{\partial \vartheta} \log \lambda_\vartheta(x|\xi \setminus \{x\}). \quad (7.8)$$

Since we are dealing with an exponential family on a compact window W , one can show that (see e. g. van Lieshout, 2000, page 41)

$$\lambda_\vartheta(u|\xi) = g_W(u, \xi) \exp(\langle \vartheta, S^W(u, \xi) \rangle),$$

where

$$\begin{aligned} g_W(u, \xi) &:= g_W(u \cup \xi) / g_W(\xi) \\ S^W(u, \xi) &:= S^W(u \cup \xi) - S^W(\xi) \end{aligned}$$

for $u \in \mathfrak{X}$ and $\xi \in \mathfrak{N}(W)$. That is, (7.8) reads⁴

$$\int_W S^W(u, \xi) \exp(\langle \vartheta, S^W(u, \xi) \rangle) \lambda(du) = \sum_{x \in \xi} S^W(x, \xi \setminus \{x\}). \quad (7.9)$$

Regarding the measurability of the pseudo-likelihood estimating equation (7.9), we note the following.

7.6 Remark In the above situation, consider the case $k = 1$ and specifically a homogeneous Gibbs point process model with a known intensity and a (parametrized) potential energy that is only based on a pair interaction potential ψ .⁵ Then the sufficient statistic reads

$$S^W(\xi) = \sum_{\substack{\{x,y\} \subset \xi \\ x \neq y}} \psi(\{x,y\}) \quad \text{for } \xi \in \mathfrak{N}(W)$$

⁴We refer to Lemma 1.38 for the exchange of integration and differentiation.

⁵We meet this situation, for example, by considering a homogeneous Strauss or a homogeneous Lennard-Jones point process with known intensity parameter β .

and the right-hand side of the pseudo-likelihood estimating equation (7.9)

$$\sum_{x \in \xi} S^W(x, \xi \setminus \{x\})$$

is $\sigma(S)$ -measurable. To show this, we first note for $\xi \in \mathfrak{N}(W)$ and each $x \in \xi$

$$\begin{aligned} S^W(x, \xi \setminus \{x\}) &= S^W(\xi) - S^W(\xi \setminus \{x\}) \\ &= \sum_{\substack{\{u,v\} \subset \xi \\ u \neq v}} \psi(\{u,v\}) - \sum_{\substack{\{u,v\} \subset \xi \setminus \{x\} \\ u \neq v}} \psi(\{u,v\}) \\ &= \sum_{u \in \xi \setminus \{x\}} \psi(\{x,u\}). \end{aligned}$$

Hence,

$$\sum_{x \in \xi} S^W(x, \xi \setminus \{x\}) = \sum_{x \in \xi} \sum_{u \in \xi \setminus \{x\}} \psi(\{x,u\}) = 2 \cdot S^W(\xi).$$

However, we may just remark that the left-hand side in the estimating equation (7.9)

$$\xi \mapsto \int_W S^W(u, \xi) \exp(\langle \vartheta, S^W(u, \xi) \rangle) \lambda(du)$$

is measurable with respect to⁶ $\mathcal{D} := \text{com}(\sigma(S^W(u, \cdot) : u \in W))$, see Mattner (1999, Theorem 3.1, page 274 ff.). Note that \mathcal{D} is, in general, finer than $\sigma(S^W)$. In the case of a homogeneous Strauss point process model with known intensity, for example, $S^W(u, \xi)$ yields the number of points of ξ that are R -close to $u \in W$. Let $W_0 \subset W$ be a countable dense subset. Then,

$$\bigcap_{u \in W_0} \{\xi : S^W(u, \xi) = 0\} = \{\mathbf{0}\}$$

is a set in \mathcal{D} . However, in general $\{\mathbf{0}\} \subsetneq \{\xi : S^W(\xi) = 0\}$ because for a sufficiently large window W , there exists a point pattern $\xi \neq \mathbf{0}$ that have no R -close pairs, which means $S^W(\xi) = 0$. Consequently, we have $\{\mathbf{0}\} \notin \sigma(S^W)$ and thus $\sigma(S^W) \subsetneq \mathcal{D}$. \blacklozenge

Since the MPLE is determined by the estimating equation (7.9) the remark above also demonstrates that the MPLE is in general not measurable w. r. t. the complete sufficient statistic S^W . An improvement with respect to convex loss functions is then, for example, given by the Rao-Blackwellization $\mathbf{E}(\hat{\vartheta}_{\text{MPL}} | S^W) \neq \hat{\vartheta}_{\text{MPL}}$.

⁶Let $(\mathfrak{X}, \mathcal{A}, \mu)$ be a general measure space and let $\mathcal{C} \subset \mathcal{A}$ be a sub- σ -algebra. We then define $\text{com}(\mathcal{C}) := \text{com}(\mathcal{C})_{\mathcal{A}, \mu} := \{A \in \mathcal{A} : \exists C \in \mathcal{C} \text{ with } \mu(A \Delta C) = 0\}$ which is called the *completion* of \mathcal{C} w. r. t. \mathcal{A} and μ .

Computation of the MPLE

Baddeley & Turner (2000) provide a device to compute at least an approximate MPLE. The essential advantage is that we can use standard software which fits generalized linear models.

7.7 Remark (MPLE Computation Device, Poisson Regression) To approximate the integral in the pseudo-likelihood estimating equation (7.8), Baddeley & Turner (2000, Section 4) consider a quadrature rule. Let $\xi \in \mathfrak{N}(W)$ be our observed data point pattern. Independently of ξ , we choose a dummy point pattern in W which we denote by η . Furthermore, we consider weights $w_z > 0$, $z \in \xi \cup \eta$ such that $\sum_{z \in \xi \cup \eta} w_z = \lambda(W)$ and, finally, approximate the integral in (7.8) by

$$\int_W \lambda_\vartheta(u|\xi) \lambda(du) \approx \sum_{z \in \xi \cup \eta} \lambda_\vartheta(z|\xi \setminus \{z\}) w_z.$$

This yields an approximate log-pseudo-likelihood that reads

$$\text{pl}_W(\vartheta, \xi) \approx \sum_{x \in \xi} \log \lambda_\vartheta(x|\xi \setminus \{x\}) - \sum_{z \in \xi \cup \eta} \lambda_\vartheta(z|\xi \setminus \{z\}) w_z,$$

which can be rearranged as

$$\text{pl}_W(\vartheta, \xi) \approx \sum_{z \in \xi \cup \eta} \left(\frac{\mathbf{1}_\xi(z) \log \lambda_\vartheta(z|\xi \setminus \{z\})}{w_z} - \lambda_\vartheta(z|\xi \setminus \{z\}) \right) w_z. \quad (7.10)$$

Note that the right-hand side corresponds to the log-likelihood of a Poisson regression model with independent observations $(\mathbf{1}_\xi(z))_{z \in \xi \cup \eta}$, which depends on the auxiliary information $(\xi \setminus \{z\})_{z \in \xi \cup \eta}$. The vector $(w_z)_{z \in \xi \cup \eta}$ determines the weights for the iteratively reweighted least squares algorithm.

If Ξ is a non-Poisson point process, the conditional intensity

$$u \mapsto \lambda_\vartheta(u|\xi \setminus \{u\})$$

is typically discontinuous at the data points $x \in \xi$. ◆

To interpret the log-pseudo-likelihood as a log-likelihood of a Poisson regression seems to be inappropriate, since the observation is just binary. Using a finer grid of dummy points, this model mismatch can be reduced.

Another approach is given by a logistic regression model, which was proposed and examined by Baddeley *et al.* (2014) for the situation of Gibbs point process models. We follow Baddeley *et al.* (2014, Section 3, page 5 f.) to formulate the next remark.

7.8 Remark (MPLE Computation Device, Logistic Regression) Let H be a dummy point process with intensity function ρ , that is, ρ is a density of

\mathbf{M}_H w. r. t. μ . Due to the Georgii–Nguyen–Zessin equation (see Theorem 5.46, page 122), we have the identity

$$\mathbf{E} \left(\int_W S^W(u, \Xi) \lambda_\vartheta(u|\Xi) \lambda(du) \right) = \mathbf{E} \left(\sum_{u \in \Xi \cup H} S^W(u, \Xi \setminus \{u\}) \frac{\lambda_\vartheta(u|\Xi \setminus \{u\})}{\lambda_\vartheta(u|\Xi \setminus \{u\}) + \rho(u)} \right),$$

which provides an unbiased estimator of the integral in the pseudo-likelihood estimating equation (7.9). According to this, we may consider

$$s_{\text{LR},W}(\vartheta, \xi, \eta) := \sum_{x \in \xi} S^W(x, \xi \setminus \{x\}) - \sum_{u \in \xi \cup \eta} S^W(u, \xi \setminus \{u\}) \frac{\lambda_\vartheta(u|\xi \setminus \{u\})}{\lambda_\vartheta(u|\xi \setminus \{u\}) + \rho(u)} \quad (7.11)$$

for $\xi, \eta \in \mathfrak{N}(W)$ and $\vartheta \in \Theta$, which is still an unbiased estimating equation. In fact, rearranging (7.11) yields

$$s_{\text{LR},W}(\vartheta, \xi, \eta) = \sum_{x \in \xi} \frac{\rho(x) S^W(x, \xi \setminus \{x\})}{\lambda_\vartheta(x|\xi \setminus \{x\}) + \rho(x)} - \sum_{y \in \eta} \frac{S^W(y, \xi \setminus \{y\}) \lambda_\vartheta(y|\xi \setminus \{y\})}{\lambda_\vartheta(y|\xi \setminus \{y\}) + \rho(y)}, \quad (7.12)$$

that is, the score of the logistic log-likelihood

$$\begin{aligned} L_{\text{LR},W}(\vartheta, \xi, \eta) \\ = \sum_{x \in \xi} \log \left(\frac{\lambda_\vartheta(x|\xi \setminus \{x\})}{\lambda_\vartheta(x|\xi \setminus \{x\}) + \rho(x)} \right) - \sum_{y \in \eta} \log \left(\frac{\rho(y)}{\lambda_\vartheta(y|\xi \setminus \{y\}) + \rho(y)} \right). \end{aligned} \quad (7.13)$$

Note that given the joint point pattern $\xi \cup \eta$, the expression (7.13) is the log-likelihood of independent Bernoulli random variables $(Y_u)_{u \in \xi \cup \eta}$, where successes are observed at $x \in \xi$. The success probability at $u \in \xi \cup \eta$ is

$$\mathbf{P}(Y_u = 1) = \frac{\lambda_\vartheta(u|\xi \setminus \{u\})}{\lambda_\vartheta(u|\xi \setminus \{u\}) + \rho(u)} = \frac{\exp(\langle \vartheta, S^W(u, \xi) \rangle + \log H(u, \xi))}{1 + \exp(\langle \vartheta, S^W(u, \xi) \rangle + \log H(u, \xi))},$$

where $H(u, \xi) := g_W(u, \xi)/\rho(u)$ for $u \in W$ and $\xi \in \mathfrak{N}(W)$. Hence, (7.13) is a logistic regression log-likelihood with offset term $H(u, \xi)$. \blacklozenge

Finally, to realize an MPL estimate for a given data point pattern, we have to choose a dummy point process and, in addition, we have to specify the corresponding intensity function ρ . In Section 7.4 we will compute MPL estimates using the R package `spatstat`. To this end, a partition of the window W is considered which consists of n_d^2 squares. The dummy point process is then given by a stratified point process which independently places exactly one dummy point uniformly in each square.

7.3 Variational Estimation

Statistical analysis of point pattern data exhibiting strong dependences is still quite challenging by the methods presented so far. Likelihood-based

methods using Markov Chain Monte Carlo techniques (see for example Geyer & Thompson (1992)) are computationally intensive due to the required simulations from the corresponding distributions. Maximum pseudo-likelihood estimation (see previous section) still requires some kind of numerical integration.

Baddeley & Dereudre (2013) propose an estimation method motivated by variational estimators of Almeida & Gidas (1993). This estimation method is based on a variational equation, which keeps track of infinitesimal perturbations of the so-called local energy using test functionals. This approach seems to be promising, because it does neither require simulation nor numerical integration.

Variational Estimator for Stationary Gibbs Point Processes

Consider the case $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}), \lambda^d) = (\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), \lambda^d)$ and, furthermore, let $\mathfrak{P} = \{P_\vartheta : \vartheta \in \Theta\}$ be a model of stationary Gibbs point process distributions on $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), \lambda^d)$. Consider⁷ $g \in \bigcap_{j=1}^k \mathfrak{R}(S_j^W)$ such that g is *shift invariant*, that is, $g(x, \xi) = g(\mathbf{t}_v(x), \xi)$ for all $x, v \in \mathbb{R}^d$ and $\xi \in \mathfrak{R}(\mathbb{R}^d)$. We assume that

$$\int |\nabla g(\mathbf{0}, \xi)| + |g(\mathbf{0}, \xi) \nabla S_j^W(\mathbf{0}, \xi)| \mathbf{Pa}^{10}(d\xi) < \infty \quad (7.14)$$

for $j = 1, \dots, k$. Writing \mathbf{E}^{10} for the expectation w. r. t. \mathbf{Pa}^{10} , we have by Proposition 6.38

$$\sum_{j=1}^k \vartheta_j \mathbf{E}^{10}(g(\mathbf{0}, \xi) \nabla S_j^W(\mathbf{0}, \xi)) = \mathbf{E}^{10}(\nabla g(\mathbf{0}, \xi)). \quad (7.15)$$

This equation yields a linear system of k equations which, however, are commonly identical. Therefore, we summarize them by the *divergence operator*

$$\operatorname{div} = \sum_{i=1}^d \frac{\partial}{\partial x_i}.$$

The idea is then to consider (7.15) for a choice of k shift invariant functions $g_1, \dots, g_k \in \bigcap_{j=1}^k \mathfrak{R}(S_j^W)$ that fulfill the above integrability condition (7.14). We then get the following linear system of equations

$$A\vartheta = b$$

where $A = (A_{i,j} : i, j \in \{1, \dots, k\})$ with

$$A_{i,j} := \int g_i(\mathbf{0}, \xi) \operatorname{div} S_j^W(\mathbf{0}, \xi) \mathbf{Pa}^{10}(d\xi) \quad \text{for } i, j \in \{1, \dots, k\}$$

⁷Recall that $\mathfrak{R}(U)$ denotes the set of all regularizing functions w. r. t. the potential energy U , see Definition 6.36 on page 155.

and $b = (b_1, \dots, b_k)$ with

$$b_i := \int \operatorname{div} g_i(\mathbf{0}, \xi) \mathbf{Pa}^{l_0}(\mathrm{d}\xi) \quad \text{for } i \in \{1, \dots, k\}.$$

Let $\xi \in \mathfrak{N}(\mathbb{R}^d)$. We then observe ξ_W and estimate A and b by the empirical averages, that is,

$$\widehat{A}_{i,j}(\xi_W) := \frac{1}{\xi_W(\mathfrak{X})} \sum_{x \in \xi_W} g_i(x, \xi) \operatorname{div} S_j^W(x, \xi)$$

and

$$\widehat{b}_i(\xi_W) := \frac{1}{\xi_W(\mathfrak{X})} \sum_{x \in \xi_W} \operatorname{div} g_i(x, \xi),$$

respectively. Assuming that \widehat{A} is invertible, we may then define the *variational estimator*

$$\widehat{\vartheta}_{\text{VAR}}(\xi_W) := (\widehat{A}_{i,j}(\xi_W))^{-1} \widehat{b}_i(\xi_W) \quad \text{for } \xi \in \mathfrak{N}(\mathbb{R}^d).$$

For a stationary pair interaction Gibbs point process model, Baddeley & Dereudre (2013, Section 5 and Section 6.1) propose to choose

$$g_i := \operatorname{div} S_i^W \quad \text{for } i = 1, \dots, k. \quad (7.16)$$

At this point, we mention that Baddeley & Dereudre (2013) also examine a similar estimator the so-called *grid variational estimator*, which uses Proposition 6.37. The stationarity assumption of the underlying Gibbs point process distribution can then be omitted. However, we have to require compactly supported functions g_1, \dots, g_k , that is, we need to adapt (7.16).

The behavior of the estimation methods presented in this section will be illustrated in the next section. To this end, we perform a simulation study for the Lennard-Jones point process model.

7.4 Simulation Study

In this section we illustrate aspects of the Ivanov regularized Rao–Blackwellization from Section 4.2 applied to the presented estimation methods for Gibbs point process models. To this end, we consider three parameter choices for the Lennard-Jones model, which is described at first. In a second step, we take a look at Ivanov regularized Rao–Blackwellizations of the variational and the maximum pseudo-likelihood estimator as an illustration of Theorem 4.20 on page 95. Third, we make use of this result by proposing and checking approaches for constructing parametric

bootstrap confidence regions for the canonical interaction parameters of the Lennard-Jones model.

All point process related simulations in this section are performed with the statistical software R using the package `spatstat` by Baddeley & Turner (2005). For the variational estimator we use an unpublished implementation by Adrian Baddeley. The implementation for the maximum likelihood estimator was provided by Philipp Möller and Dominic Schuhmacher.

The Model

We consider a homogeneous (12, 6)-Lennard-Jones model in the plane $(\mathbb{R}^2, \|\cdot\|_2)$. That is, let $\Gamma^{(\vartheta)}$ be a Gibbs specification with Lennard-Jones pair potential and canonical parameter $\vartheta \in \Theta$, where

$$\Theta :=]0, \infty[{}^2 \times]-\infty, 0[$$

is the canonical parameter space (see also Example 6.16 on page 144), then we denote by⁸

$$(P_\vartheta)_{\vartheta \in \Theta} \in \prod_{\vartheta \in \Theta} \mathbf{Gibbs}(\Gamma^{(\vartheta)})$$

the considered point process distribution model on $\mathfrak{N}(\mathbb{R}^2)$. The quantity of interest $\Xi \sim P_\vartheta$ for some $\vartheta \in \Theta$ is, however, just observed in a bounded observation window $W := [a, b] \times [c, d] \in \mathcal{H}(\mathbb{R}^2)$. Additionally, boundary knowledge is incorporated as a point pattern $\xi \in \mathfrak{N}(W^c)$. Hence, we are faced with the *conditional* model

$$\Gamma_W^{(\vartheta)}(\xi, A) := \int_A c(W, \vartheta, \xi) \exp(-\langle \vartheta, S^{W, \xi}(\eta) \rangle) \mathbf{Pop}_W(d\eta) \quad \text{for } A \in \mathcal{N}(W)$$

and $\vartheta \in \Theta$, where $c(W, \vartheta, t)$ is the normalizing constant and where we have set $S_1^{W, \xi}(\xi) := \eta(W)$ and

$$S_2^{W, \xi}(\eta) := \sum_{\substack{\{x, y\} \subset \eta + \xi_{W^c} \\ \{x, y\} \cap W \neq \emptyset \\ x \neq y}} \frac{1}{\|x - y\|^{12}}, \quad S_3^{W, \xi}(\eta) := \sum_{\substack{\{x, y\} \subset \eta + \xi_{W^c} \\ \{x, y\} \cap W \neq \emptyset \\ x \neq y}} \frac{1}{\|x - y\|^6}$$

for each $\xi \in \mathfrak{N}(\mathbb{R}^2)$ and $\eta \in \mathfrak{N}(W)$. Note that due to computational reasons, we have to approximate $S_2^{W, \xi}$ and $S_3^{W, \xi}$ by cutting off the infinite interaction range. Here, we use the common⁹ cut-off range given by $R := 2.5 \cdot \sigma$, that is, we consider for $\eta \in \mathfrak{N}(W)$

$$\tilde{S}_2^{W, \xi}(\eta) := \sum_{\substack{\{x, y\} \subset \eta + \xi_{W^c} \\ \{x, y\} \cap W \neq \emptyset \\ x \neq y}} \frac{1}{\|x - y\|^{12}} \mathbb{1}_{]0, R]}(\|x - y\|),$$

⁸We note that $\mathbf{Gibbs}(\Gamma^{(\vartheta)}) \neq \emptyset$ referring to Preston (1976). For a sketch of the main arguments therein, which are given in the context of some concepts worked out in Preston (2005), see Section 6.4 in this thesis.

⁹For conventions regarding the cut-off range of the Lennard-Jones pair potential we refer to the literature on simulations of liquids, see for example Allen & Tildesley (2017).

and

$$\tilde{S}_3^{W,\xi}(\eta) := \sum_{\substack{\{x,y\} \subset \eta + \xi_{Wc} \\ \{x,y\} \cap W \neq \emptyset \\ x \neq y}} \frac{1}{\|x-y\|^6} \mathbb{1}_{]0,R[}(\|x-y\|).$$

In terms of $\tilde{S}_2^{W,\xi}$ and $\tilde{S}_3^{W,\xi}$, the corresponding pair potential

$$\phi_2(\{x,y\}) = 4\varepsilon \left[\left(\frac{\sigma}{\|x-y\|} \right)^{12} - \left(\frac{\sigma}{\|x-y\|} \right)^6 \right] \mathbb{1}_{]0,2.5\sigma[}(\|x-y\|) \quad \text{for } x, y \in \mathfrak{X}$$

is called *cut-off Lennard-Jones pair potential*, see also Figure 7.1.

Parameter Choice

We consider three different parameters that follow a path in Θ from weak to strong interaction. For simplicity this path is described by what is called the *inverse temperature* $t > 0$. That is, we consider the fixed model parameters $\beta := 100$ (intensity), $\varepsilon := 1$ (well depth), and $\sigma = 0.1$ (atomic diameter), leading to the fixed canonical parameter $\vartheta_0 := (\log(100), 4 \cdot (0.1)^{12}, -4 \cdot (0.1)^6)$. Then, we define at inverse temperature $t > 0$,

$$\Gamma_{W_t}^{(\vartheta_0,t)}(\xi, A) := \int_A c(W_t, \vartheta_0, \xi) \exp(-t \langle \vartheta_0, S^{\xi, W_t}(\eta) \rangle) \mathbf{Pop}_{W_t}(d\eta)$$

for $A \in \mathcal{N}(W_t)$, where W_t is a scaled observation window to account for the different intensities β^t (see below). In the present simulation study, we examine the cases $t = 0.5$, $t = 0.75$, and $t = 1$, that is, we look at the behavior of a particle system going toward the low temperature regime.

Figure 7.2 shows corresponding realizations in the fixed observation window $W_1 := [0, 2]^2$, when condition on non-empty boundary point patterns in $B_1 := [-0.25, 2.25]^2 \setminus W_1$. In Figure 7.3 we consider the situation with scaled observation windows $W_t := \gamma_t \cdot W_1$ for $t \in \{\frac{1}{2}, \frac{3}{4}, 1\}$ to ensure that the expected number of points is equal in each case.¹⁰ In the following, we denote by B_t the corresponding boundary. Visual inspection of Figure 7.3 (a) suggests only weak interaction, which however is still different from a typical Poisson point process realization according to the repulsion

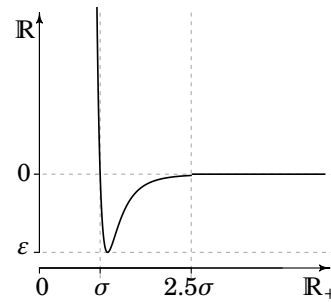


Figure 7.1: Graph of the cut-off (12,6)-Lennard-Jones pair potential as a function of the particle distance for $\varepsilon, \sigma > 0$ (see also Figure 6.2). The interaction range cuts off at $2.5 \cdot \sigma$.

¹⁰We empirically determined $\gamma_{0.5} = 2.557894$ and $\gamma_{0.75} = 1.420784$, in order to obtain $\mathbf{E}_t \Xi_{W_t} \approx 240$ for $t = 0.5$, $t = 0.75$, and $t = 1$, respectively (see also Figure 7.4).

that is obviously inherent (see also Figure 7.2 (a) for a magnification of the point pattern on a subdomain of $W_{0.5}$, which makes the interaction more apparent). In contrast, the point pattern in Figure 7.3 (c) indicates strong interaction in accordance with the choice $t = 1$, that is, the lowest temperature regime that we consider.

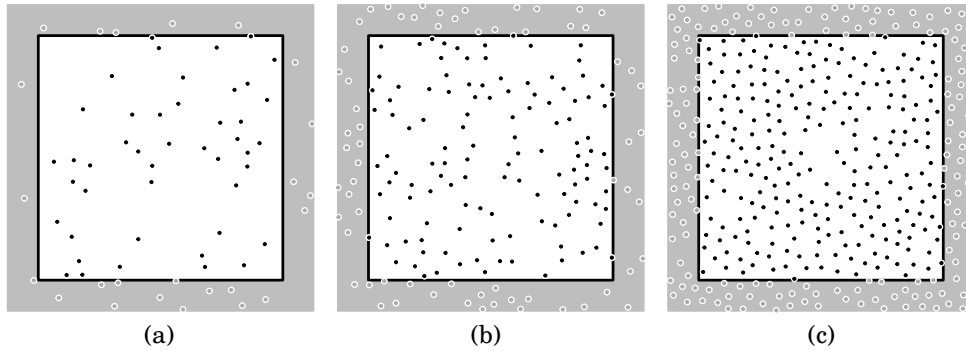


Figure 7.2: Lennard-Jones realizations (parameters: $\beta = 100$, $\sigma = 0.1$, $\varepsilon = 1$) in the window $[0, 2]^2$ at inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$ in (a), (b), and (c), respectively. The gray point pattern in the gray area represents the boundary condition within the *finite* interaction range of the *cut-off* Lennard-Jones potential (see also Figure 7.1).

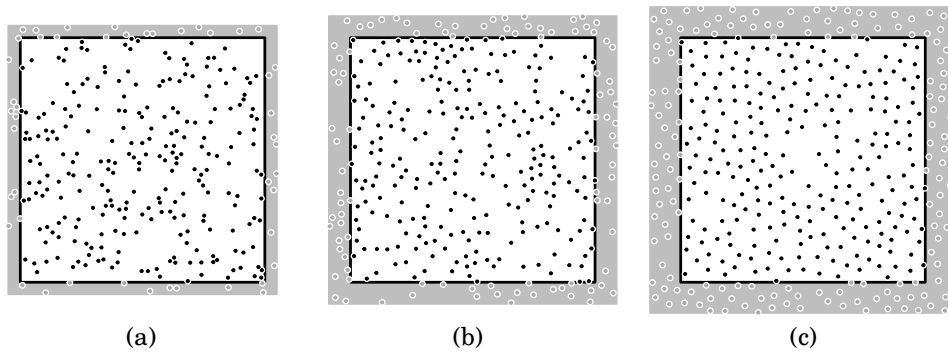


Figure 7.3: Lennard-Jones realizations (parameters: $\beta = 100$, $\sigma = 0.1$, $\varepsilon = 1$) at inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$ in (a), (b), and (c), respectively. In contrast to Figure 7.2, the observation windows in (a) and (b) are increased such that the expected number of points are equal in each case (see also Figure 7.4). The gray point pattern in the gray area represents the boundary condition within the *finite* interaction range of the *cut-off* Lennard-Jones potential (see also Figure 7.1), which decreases in (a) and (b) in relation to the (increased) observation window (see also Figure 7.2 (a) and (b)).

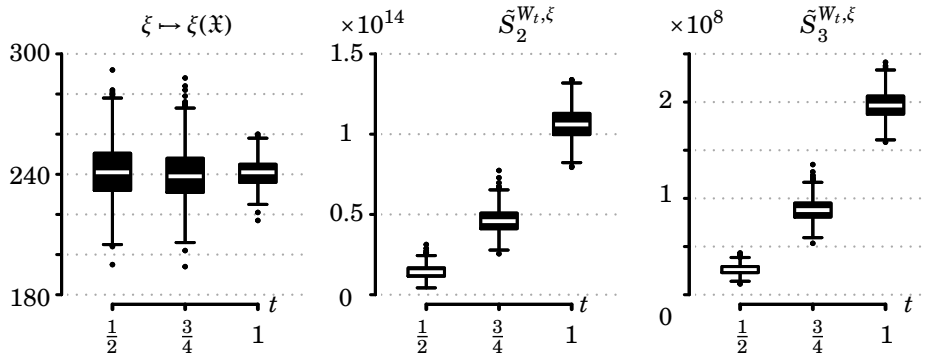


Figure 7.4: Boxplots of the sufficient statistics of 1000 realizations of the (cut-off) Lennard-Jones point process (parameters: $\beta = 100$, $\sigma = 0.1$, $\varepsilon = 1$) at inverse temperatures $t = 0.5$, 0.75 , and 1 . The observation windows are adjusted such that we may expect the same number of points. For $\tilde{S}_2^{W_t, \xi}$ and $\tilde{S}_3^{W_t, \xi}$ we empirically observe that an increase of the inverse temperature from $t = 0.5$ to $t = 1$ leads to a higher level of $\tilde{S}_2^{W_t, \xi}$ (repulsion force) and a higher level of $\tilde{S}_3^{W_t, \xi}$ (attraction force).

(Ivanov) Regularized Rao–Blackwellization

To demonstrate the effect of the regularized Rao–Blackwellization (RRB), we keep on considering the canonical parameter choice

$$\vartheta_0 := (\log(100), 4 \cdot (0.1)^{12}, -4 \cdot (0.1)^6)$$

at inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$. In addition, we also examine the following two boundary conditions.

(1) Non-Empty Boundary Condition. We consider the conditional distributions $\Gamma_{W_t}^{(\vartheta_0, t)}(\xi_{W_t^c}, \cdot)$ for $t \in \{\frac{1}{2}, \frac{3}{4}, 1\}$ with $\xi_{W_t^c}$ shown in Figure 7.3 by the corresponding gray point pattern.

(2) Empty Boundary Condition. We consider the conditional distributions $\Gamma_{W_t}^{(\vartheta_0, t)}(\mathbf{0}, \cdot)$ for $t \in \{\frac{1}{2}, \frac{3}{4}, 1\}$.

In each case we are concerned with distributions that are part of an exponential family with the complete sufficient statistic $\tilde{S}^{W, \xi}$ under the non-empty boundary condition, and $\tilde{S}^{W, \mathbf{0}}$ under the empty boundary condition. Hence, we may apply Theorem 4.20. In what follows, our focus lies on estimating the canonical interaction parameters ϑ_2 and ϑ_3 . To this end, we consider the MLE, the MPLE (with $n_d^2 = 16^2$ and $n_d^2 = 32^2$ dummy points on W_t , enlarging the grid parameter to $n_d^2 = 20^2$ and $n_d^2 = 40^2$, respectively when computing the MPLE of point patterns on $W_t \oplus B_t$)¹¹, and the VARE (see Section 7.1 to 7.3).

¹¹The grid parameter n_d concerns the dummy point process of the logistic regression MPLE computation device, see Remark 7.8 and the subsequent hint on page 164. Here, the grid parameter $n_d = 16$ was specifically chosen for W_1 and the expected number of points ($\mathbf{E}_t \Xi_{W_t} \approx 240$) in accordance with a rule of thumb suggested by Baddeley *et al.* (2014), which was found to lead just to a moderate additional variance due to

Based on one thousand realizations (conditioned on $\xi_{W_t^c}$ and $\mathbf{0}$ for each inverse temperature case), Figure 7.7 shows the corresponding empirical marginal and joint distributions of the MLE, the VARE, and the RRB of the VARE. The latter is computed through a linear regression of the VAR estimates on the sufficient statistic values which are computed from the same one thousand realizations (see also Theorem 4.20). In the low inverse temperature case $t = 0.5$, the distributions of the MLE, the MPLE, and the VARE are skewed due to the restricted support on $\text{pr}_{\{2,3\}}(\Theta) =]0, \infty[\times]-\infty, 0[$. Note, however, that implementations used for these estimation methods still yield inadmissible estimates in a few cases.

Numerically the RRB projection improves the VARE in terms of the MSE in each inverse temperature case for either boundary condition (see Table 7.6). Furthermore, the regularized Rao–Blackwellization also reflects characteristics that are close to those of the MLE. Especially, we can recognize that the regularized Rao–Blackwellization change the correlation of the VARE components toward the correlation of the MLE components. Nevertheless, distributional differences of the MLE and the RRB of the VARE are apparent, which are due to the bias and the boundary condition. First, a (substantial) part of the distributional differences of the MLE and the RRB of the VARE (in the empty boundary condition) can be reduced by shifting the estimators according to their bias. This is in accordance with Theorem 4.20. Second, in the non-empty boundary condition the MLE seems to spread less than the RRB of the VARE due to $\xi_{W_t^c}$ at the boundary,¹² where as in the empty boundary condition, we do not recognize such an effect.

To explain this boundary effect, we recall that W_t and B_t denote the observation window and the boundary zone, respectively. Considering the MLE $\hat{\vartheta}_{\text{ML}}$ as a map on $(\mathfrak{N}(W_t \cup B_t), \mathcal{N}(W_t \cup B_t))$, we note that $\hat{\vartheta}_{\text{ML}}$ is also measurable w. r. t. $\sigma(\tilde{S}^{(W_t \cup B_t), \mathbf{0}})$, see Proposition 7.1. Roughly speaking, $\hat{\vartheta}_{\text{ML}}$ makes use of the boundary point pattern through the complete sufficient statistic of a corresponding enlarged model on $\mathfrak{N}(W_t \cup B_t)$ with empty boundary condition. In contrast, considering the VARE as a map on

$$(\mathfrak{N}(W_t \cup B_t), \mathcal{N}(W_t \cup B_t)) \cong (\mathfrak{N}(W_t) \times \mathfrak{N}(B_t), \mathcal{N}(W_t) \otimes \mathfrak{N}(B_t))$$

(see Proposition 5.10 on page 107 for this identification), we note that the corresponding RRB just performs a data reduction to $\sigma(\tilde{S}^{W_t, \xi})$ in the first component space $\mathfrak{N}(W_t)$. In fact, within the mentioned enlarged model on $\mathfrak{N}(W_t \cup B_t)$ with non-empty boundary condition the RRB still depends on

the randomness that comes from the dummy point process. To study this inherent randomness of the dummy point pattern, we additionally chose $n_d = 32$.

¹²Recall from the above description of the non-empty boundary condition that the condition $\xi_{W_t^c}$ remains unchanged for all realizations. We also point out that $\xi_{W_t^c}$ can be regarded as a model parameter, see also the introduction to this chapter.

reducible data located at the boundary B_t through the mixed pairs of one η -point in W_t and one $\xi_{W_t^c}$ -point in B_t , which visually corresponds to the pairs of one black and one gray point in Figure 7.3.

The regularized Rao–Blackwellization of the MPLE with $n_d^2 = 1024$ dummy points seems to improve the MPLE w. r. t. the MSE significantly in the low inverse temperature regime ($t = 1$), see Figure 7.8. Improvements for the high and middle activity cases $t = 0.5$ and $t = 0.75$ can at least be seen from the Table 7.6. Furthermore, we see a similar boundary effect as for the RRB of the VARE. That is, by comparing the non-empty with the empty boundary condition we conclude that the MLE can use the information from the boundary more efficiently than the MPLE. This boundary effect is more apparent if we consider only $n_d^2 = 256$ dummy points (see Figure 7.9).

Figure 7.5 shows the computation time¹³ of the MLE, the MPLE, and VARE for the non-empty boundary condition. According to these results, the MLE is the most computationally expensive method. In addition, its computation time increases significantly with the inverse temperature. In contrast, the VARE is the least computationally expensive method. The computation time is just slightly affected by an increase in the inverse temperature. The computation time of the MPLE depends on the number of dummy points that is controlled by the grid parameter n_d .

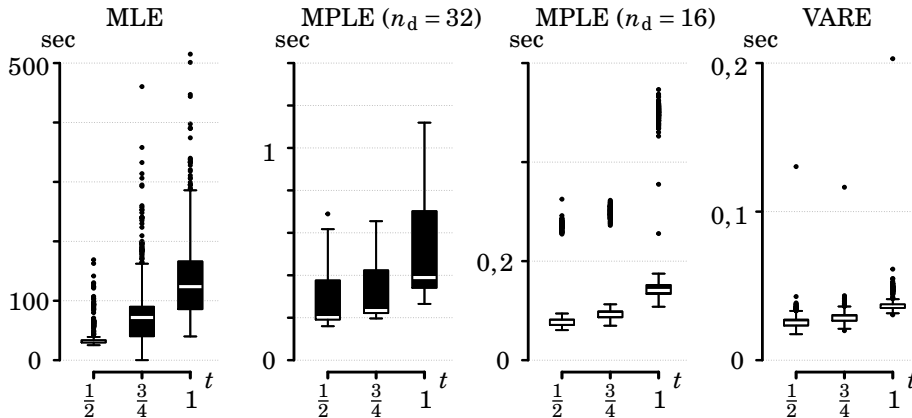


Figure 7.5: Boxplots of the empirical computation time of the MLE, the MPLE, and the VARE for realizations at inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$ and the non-empty boundary condition. These computations were performed using a single core on a Linux server (AMD Opteron Processor 6140 from 2011 with 2.6 GHz).

¹³The computation times refer to the performance using a single core on a Linux server (AMD Opteron Processor 6140 from 2011 with 2.6 GHz).

Inverse Temperature	$t = 0.5$				$t = 0.75$				$t = 1$			
Boundary Condition	Non-Empty		Empty		Non-Empty		Empty		Non-Empty		Empty	
Parameter	ϑ_2	ϑ_3	ϑ_2	ϑ_3	ϑ_2	ϑ_3	ϑ_2	ϑ_3	ϑ_2	ϑ_3	ϑ_2	ϑ_3
$\hat{\vartheta}_{\text{ML}}$	0.707	0.431	0.586	0.421	0.163	0.163	0.267	0.216	0.071	0.063	0.175	0.170
$\hat{\vartheta}_{\text{VAR}}$	7.516	12.050	7.812	12.867	2.226	4.191	2.842	5.023	1.851	4.565	1.801	4.056
$\hat{\vartheta}_{\text{MPL}(32)}$	0.780	0.535	0.706	0.520	0.333	0.312	0.384	0.318	0.678	1.010	0.559	0.652
$\hat{\vartheta}_{\text{MPL}(16)}$	0.992	0.699	0.985	0.782	0.481	0.474	0.555	0.495	4.561	5.248	1.289	1.487
$\mathbf{E}^{(\text{LR})}(\hat{\vartheta}_{\text{VAR}} S)$	3.739	5.059	3.550	5.103	0.635	0.925	1.001	1.429	0.352	0.645	0.363	0.720
$\mathbf{E}^{(\text{LR})}(\hat{\vartheta}_{\text{MPL}(32)} S)$	0.574	0.428	0.532	0.410	0.209	0.160	0.274	0.216	0.179	0.257	0.184	0.190
$\mathbf{E}^{(\text{LR})}(\hat{\vartheta}_{\text{MPL}(16)} S)$	0.635	0.435	0.577	0.428	0.200	0.155	0.288	0.221	0.429	0.477	0.204	0.156

Table 7.6: Empirical mean squared errors of the estimators for the Lennard-Jones interaction parameters with empty and non-empty boundary condition based on one thousand realizations (for inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$). Mean squared error values for ϑ_2 and ϑ_3 are multiplied by 10^{24} and 10^{12} , respectively.

Bootstrap Confidence Regions

To demonstrate the possible benefits of the results above, we propose ad hoc constructions of parametric bootstrap confidence regions¹⁴ for the special case of the canonical parameter \mathbf{id}_Θ . Basically, the idea is to apply computationally efficient estimators, for example, the VARE or the MPLE (see also Figure 7.5) to data simulated from a maximum likelihood estimate of our observation. RRBs of these estimates will then be used to construct parametric bootstrap confidence regions. The validity of this approach depends on the following two qualitative assumptions.

- (a) The corresponding RRB of an alternative estimator $\hat{\vartheta}$ approximates the MLE $\hat{\vartheta}_{\text{ML}}^{\xi_{W_t^c}} \circ S^{W_t, \xi} := \hat{\vartheta}_{\text{ML}}(\cdot, \xi_{W_t^c})$ sufficiently well. To be more precise, that means (see also Theorem 4.20), the error of the first order Taylor approximation of the MLE $\hat{\vartheta}_{\text{ML}}^{\xi_{W_t^c}}$ is small w. r. t. the norm

$$f \mapsto \int \|f(s)\|^p \mu(ds),$$

where

$$\mu(B) := \Gamma_{W_t}^{(\vartheta)}(\xi, S^{W_t, \xi^{-1}}(B)) \quad \text{for } B \in \mathcal{B}(\mathbb{R}^k)$$

is the distribution of the complete sufficient statistic under ϑ . In addition, we assume that $\frac{\partial}{\partial \vartheta}(\mathbf{E}_\vartheta(\hat{\vartheta} - \hat{\vartheta}_{\text{ML}}))$ is small.

- (b) $V(\Xi_{W_t}, \vartheta) := \hat{\vartheta}_{\text{ML}}(\Xi_{W_t}, \xi_{W_t^c}) - \vartheta$ is an approximate pivot element, that is,

$$\vartheta \mapsto \Gamma_{W_t}^{(\vartheta)}(\Xi, \cdot)^{V(\cdot, \vartheta)}$$

is approximately constant.

Studying the validity of this approach is also part of the present simulation study. Justifications of quantified versions of assumptions (a) and (b) are difficult and require further research.

In what follows we consider three methods of constructing parametric bootstrap confidence regions. In every case we make extensive use of the stated qualitative assumptions. We give a common formulation it in terms of Gibbs point processes, especially with regard to the present chapter. The treatment for general parametric models with a complete sufficient statistic is obvious.

The first approach is a natural one that is based on the empirical distribution of an estimator.

7.9 Method (Parametric Bootstrap CR, I) Let $(Q_\vartheta)_{\vartheta \in \Theta}$ be a Gibbs point process model and let $\hat{\kappa}$ be an arbitrary estimator (for \mathbf{id}_Θ).

¹⁴See also Example 1.9 on page 7 for the notion of a confidence region.

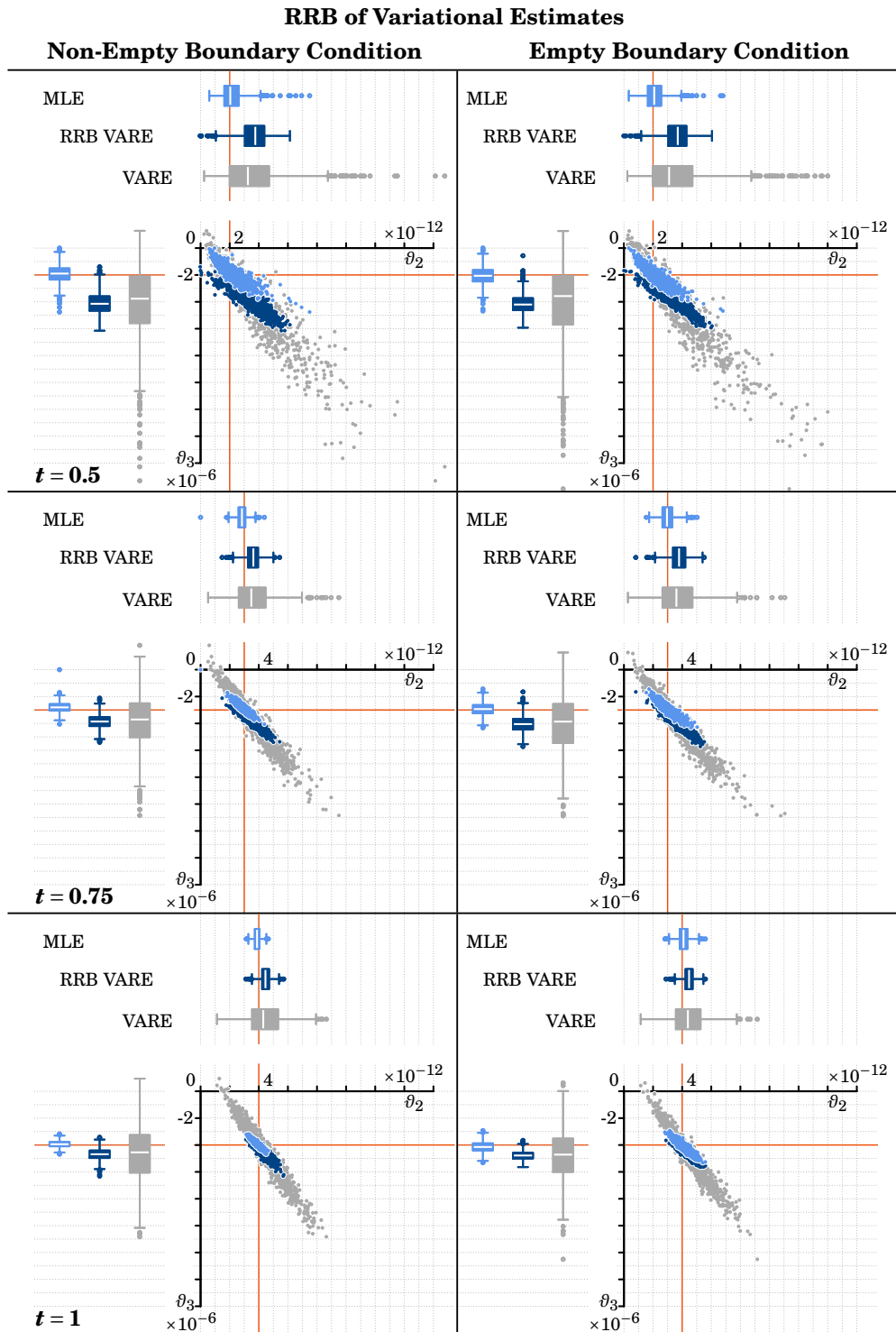


Figure 7.7: Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition of inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$. The gray, dark blue, and light blue point clouds and boxplots correspond to the VARE, the Ivanov regularized Rao–Blackwellization of the VARE, and the MLE, respectively. The intersection point of the two orange lines represents the true parameter.

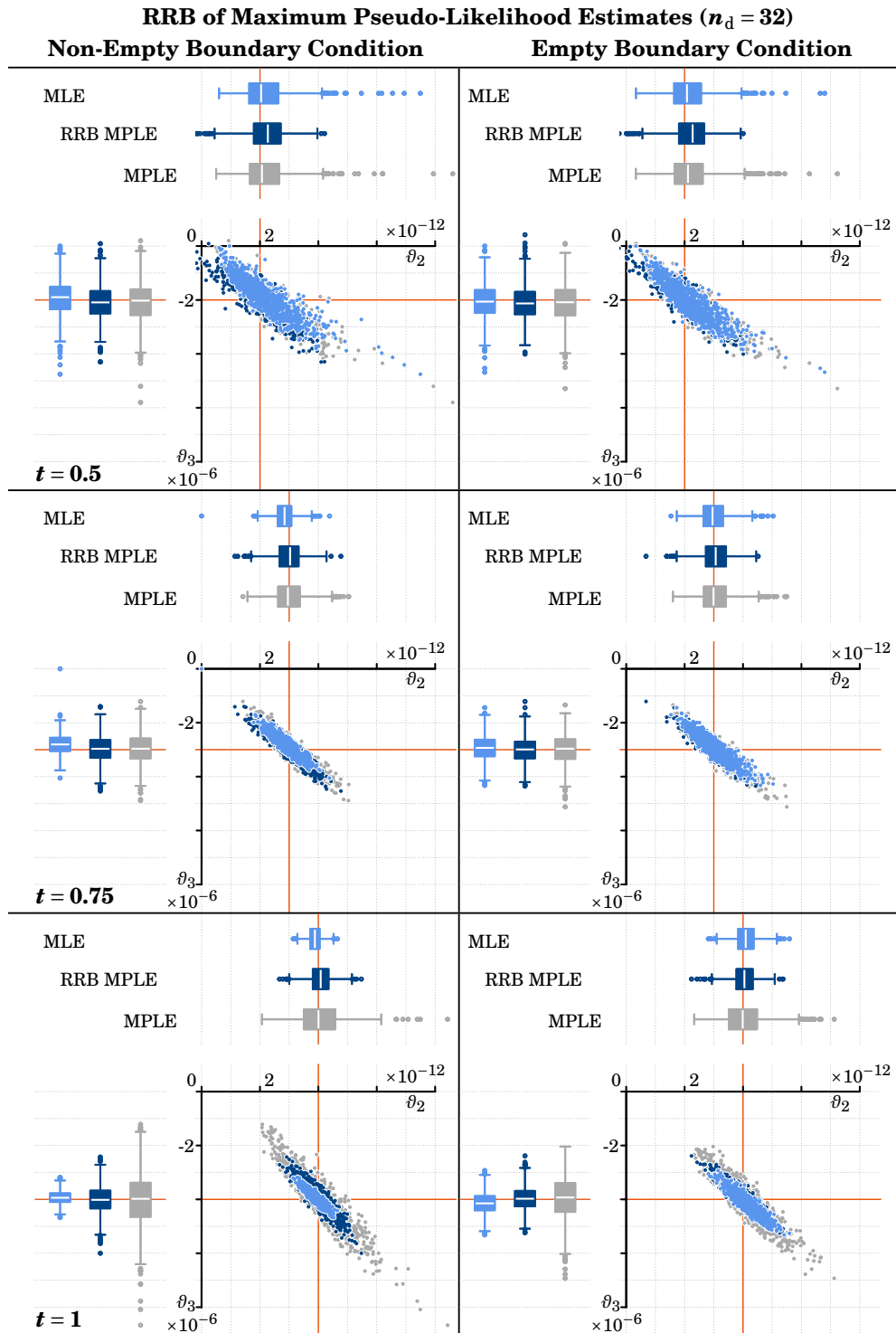


Figure 7.8: Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition based on one thousand realizations (for inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$). The gray, dark blue, and light blue point clouds and boxplots correspond to the MPLE ($n_d = 32$), the Ivanov regularized Rao–Blackwellization of the MPLE, and the MLE, respectively. The intersection point of the two orange lines represents the true parameter.

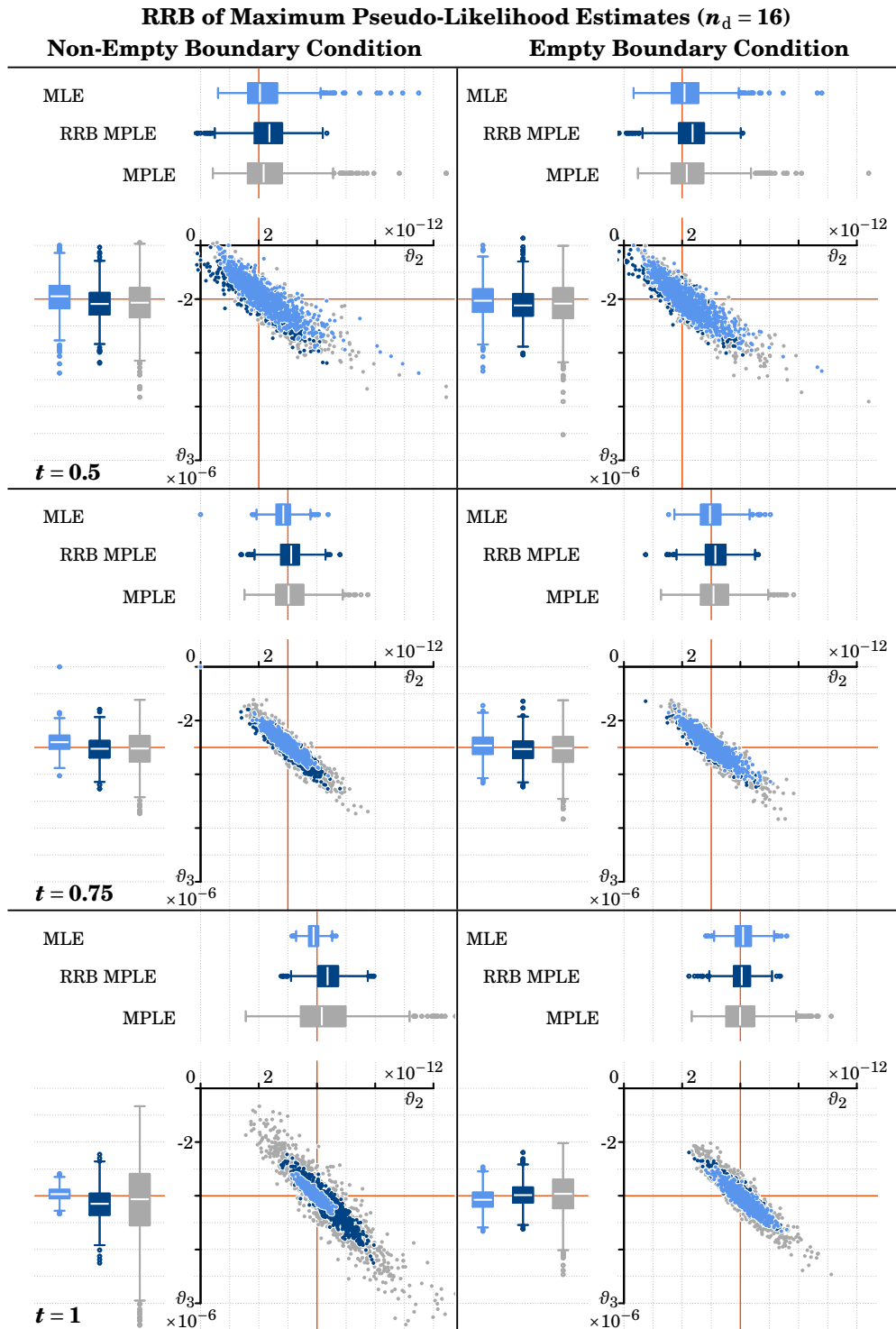


Figure 7.9: Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition based on one thousand realizations (for inverse temperatures $t = 0.5$, $t = 0.75$, and $t = 1$). The gray, dark blue, and light blue point clouds and boxplots correspond to the MPLE ($n_d = 16$), the Ivanov regularized Rao–Blackwellization of the MPLE, and the MLE, respectively. The intersection point of the two orange lines represents the true parameter.

For our observed point pattern data ξ , we aim at a $(1 - \alpha)$ -confidence region for the canonical parameter \mathbf{id}_Θ by proceeding as follows:

- (1) Compute $\widehat{\kappa}(\xi)$;
- (2) Simulate $H_1, \dots, H_N \sim Q_{\widehat{\kappa}(\xi)}$ for large N ;
- (3) Compute $\widehat{\kappa}(H_1), \dots, \widehat{\kappa}(H_N)$;
- (4) Construct a convex set $C(H)$, for example according to Hu & Yang (2013)¹⁵, such that $C(H)$ contains at least $1 - \gamma$ of the mass of the empirical distribution given by $(\widehat{\kappa}(H_i))_{i=1}^N$;
- (5) Set $\widehat{K}_{1-\gamma}(H, \xi) := C(H) + \widehat{\kappa}(\xi)$. ♦

In the above method, we may canonically choose $\gamma := \alpha$ if the assumption (b) is fulfilled. The level $1 - \gamma$ is called *nominal level* of the CR contrasting the (*real*) level $1 - \alpha$ that is required.

Let \widehat{K} and $\widehat{\Lambda}$ be two $(1 - \alpha)$ -confidence regions for a parameter κ , that is, $P_\vartheta(\kappa(\vartheta) \in \widehat{K}), P_\vartheta(\kappa(\vartheta) \in \widehat{\Lambda}) \geq 1 - \alpha$ holds for all $\vartheta \in \Theta$. We then prefer \widehat{K} against $\widehat{\Lambda}$, if $\mathbf{E}_\vartheta(\lambda^k(\widehat{K})) \leq \mathbf{E}_\vartheta(\lambda^k(\widehat{\Lambda}))$ for all $\vartheta \in \Theta$ (see Example 1.11).

The estimator $\widehat{\kappa}$ in the method above may be the MLE or the MPLE. If $\widehat{\kappa}$ is the MLE, however, the method becomes in most situations practically unfeasible, since the computation of N additional MLEs in step (4) is very time consuming.¹⁶ We therefore propose to use the RRB of a computationally efficient estimator to construct the pivot set $C(H)$.

7.10 Method (Parametric Bootstrap CR, II) Let $(Q_\vartheta)_{\vartheta \in \Theta}$ be a Gibbs point process model with sufficient statistic S . Let $\widehat{\vartheta}_{\text{ML}}$ be the MLE for \mathbf{id}_Θ and let $\widehat{\lambda}$ be an alternative estimator that is computationally more efficient.

For our observed point pattern data ξ we aim at a $(1 - \alpha)$ -confidence region for the canonical parameter \mathbf{id}_Θ by proceeding as follows:

- (1) Compute $\widehat{\vartheta}_{\text{ML}}(\xi)$;
- (2) Simulate $H_1, \dots, H_N \sim Q_{\widehat{\vartheta}_{\text{ML}}(\xi)}$ for large N ;
- (3) Compute estimates $(\widehat{\lambda}(H_i))_{i=1}^N$ and sufficient statistic values $(S(H_i))_{i=1}^N$;
- (4) Perform a linear regression of $(\widehat{\lambda}(H_i))_{i=1}^N$ on $(S(H_i))_{i=1}^N$. The fitted values are denoted by $\mathbf{E}^{(\text{LR})}(\widehat{\lambda}|S = S(H_i))$ for $i = 1, \dots, N$;
- (5) Construct a convex set $C(H)$ according to Hu & Yang (2013)¹⁵ such that $C(H)$ contains at least $1 - \gamma$ of the mass of the empirical distribution

¹⁵Hu & Yang (2013) propose a simple distribution-free procedure to construct joint confidence regions for two or more parameters, which is implemented in the R package `distfree.cr`. This method base on reducing the multivariate situation to the one-dimensional case by projecting the data on a family of one-dimensional subspaces. This leads to a corresponding family of cylinder sets. A joint confidence region is then obtained by intersecting these cylinder sets.

¹⁶According to the computation times given in Figure 7.5, computing $N = 5000$ MLEs takes more than six days, see also Figure 7.10 (below, page 183) for the computation times of these parametric bootstrap methods in the present simulation study.

that is given by the fitted values $\mathbf{E}^{(\text{LR})}(\hat{\lambda}|S = S(H_i))$ for $i = 1, \dots, N$;

(6) Set $\hat{\mathbf{K}}_{1-\gamma}(H, \xi) := C(H) + \hat{\vartheta}_{\text{ML}}(\xi)$. \blacklozenge

The following combination of Method 7.9 and 7.10 leads to another interesting procedure which considers the RRB of $\hat{\kappa}$.

7.11 Method (Parametric Bootstrap CR, III) Let $(Q_{\vartheta})_{\vartheta \in \Theta}$ be a Gibbs point process model with sufficient statistic S and let $\hat{\kappa}$ be an arbitrary estimator.

For our observed point pattern data ξ we aim at a $(1 - \alpha)$ -confidence region for the canonical parameter \mathbf{id}_{Θ} by proceeding as follows:

(1) Compute $\hat{\kappa}(\xi)$;

(2) Simulate $H_1, \dots, H_N \sim Q_{\hat{\kappa}(\xi)}$ for large N ;

(3) Compute estimates $(\hat{\kappa}(H_i))_{i=1}^N$ and sufficient statistic values $(S(H_i))_{i=1}^N$;

(4) Perform a linear regression of $(\hat{\kappa}(H_i))_{i=1}^N$ on $(S(H_i))_{i=1}^N$. The fitted values are denoted by $\mathbf{E}^{(\text{LR})}(\hat{\kappa}|S = S(H_i))$ for $i = 1, \dots, N$.

(5) Construct a convex set $C(H)$ according to Hu & Yang (2013)¹⁷ such that $C(H)$ contains at least $1 - \gamma$ of the mass of the empirical distribution that is given by the fitted values $\mathbf{E}^{(\text{LR})}(\hat{\kappa}|S = S(H_i))$ for $i = 1, \dots, N$;

(6) Set $\hat{\mathbf{K}}_{1-\gamma}(H, \xi) := C(H) + \mathbf{E}^{(\text{LR})}(\hat{\kappa}|S = S(\xi))$, where $\mathbf{E}^{(\text{LR})}(\hat{\kappa}|S = S(\xi))$ denotes the predicted value (under the linear model estimated in (4)) at $S(\xi)$. \blacklozenge

We test these methods considering again the Lennard-Jones model from above. As inverse temperature we choose $t = 1$ and as canonical parameter $\vartheta_0 := (\log(100), 4 \cdot 10^{-12}, -4 \cdot 10^{-6})$, a case where we already recognized significant boundary effects. The phenomenon of interest described by a point process $\Xi \sim P_{\vartheta_0}$ on \mathfrak{X} is then observed on $W := W_1 = [0, 2]^2$ in combination with one of three different boundary conditions, which lead to the following statistical models on $(\mathfrak{N}(W), \mathcal{N}(W))$.

(1) **Non-Empty Boundary Condition.** We observe the boundary point pattern $\xi_{W^c} \in \mathfrak{N}(W^c)$ which is considered as a *fixed* known model parameter and therefore motivates the statistical model

$$(\Gamma_W^{(\vartheta)}(\xi_{W^c}, \cdot) : \vartheta \in \Theta)$$

consisting of the conditional distributions given ξ_{W^c} .

(2) **Empty Boundary Condition.** There is *no* boundary point pattern and, therefore, we consider the statistical model

$$(\Gamma_W^{(\vartheta)}(\mathbf{0}, \cdot) : \vartheta \in \Theta)$$

consisting of the conditional distributions with the empty boundary point pattern $\mathbf{0}$.

¹⁷See also Footnote 15.

(3) Random Boundary Condition. We observe the boundary point pattern, which is now considered as a *random* but known model parameter. The observation is therefore assumed to come from

$$(\Gamma_W^{(\vartheta)}(\Xi_{W^c}, \cdot) : \vartheta \in \Theta)$$

where $\Xi_{W^c} \sim P_\vartheta^{\text{Pr}_{W^c}}$ for corresponding $\vartheta \in \Theta$.

In (1) and (3) we account for the first order boundary effects, that is, our realizations are generated on $[0, 2]^2 \oplus [-0.25, 0.25]^2$. We do not account for the boundary effect of realizations in the boundary.

We consider $M = 1000$ realizations Ξ_1, \dots, Ξ_M of each model described above, that is, the non-empty (using ξ_{W^c} from Figure 7.3), the empty and the random boundary condition model. For each realization we construct confidence regions according to Method 7.9 to 7.11 (in what follows we just write Method I, II, and III, respectively) for the required levels $1 - \alpha := 0.50, 0.90, \text{ and } 0.95$. To do so, we consider a broader range of nominal levels $1 - \gamma$ in order to compare the methods more fairly later on. For each CR construction method we consider $N = 5000$ simulations.¹⁸ A construction of 0.50-confidence regions in \mathbb{R}^2 , for example according to Hu & Yang (2013), thus bases roughly on 50^2 inner and 50^2 outer points on average (see also Footnote 15 on page 179). Due to visual inspections this was found to be appropriate.

The results of the simulation study are presented in terms of the absolute coverage counts at the parameter of interest ϑ_0 (see Table 7.13). We deem a construction method at nominal level $1 - \gamma$ admissible for the required level $1 - \alpha$, if the coverage rate is not significantly lower than the required level. To be on the safe side, we reject the coverage hypothesis $P_{\vartheta_0}(\widehat{K}_{1-\gamma}(\mathbf{H}, \Xi) \ni \vartheta_0) \geq 1 - \alpha$ already at a significance level of 0.1 and use a corresponding higher nominal level $1 - \gamma$ for the subsequent comparisons.

In fact, the different boundary conditions considered may lead to either an over or an under estimate of the pivot set $C(\mathbf{H})$ relative to the variabilities of the centers $\widehat{\vartheta}_{\text{MPL}}$, $\widehat{\vartheta}_{\text{ML}}$, and $\mathbf{E}(\widehat{\vartheta}_{\text{MPL}} | \tilde{S}^{W, \xi} = \tilde{S}^{W, \xi}(\cdot))$. Therefore, lower or higher coverage rates are observed. In addition, the empirical coverage rate is crucially determined through the bias of the underlying (centering) estimator. Hence, the required coverage may not be observed at ϑ_0 but at a perturbed parameter $\tilde{\vartheta}_0$ that is driven by the corresponding bias. Here, we will not discuss this issue, although more research has to be done at this point.

¹⁸These realizations were generated by the Metropolis–Hastings algorithm implemented in spatstat, starting from a binomial point process realization with 100 points and setting nburn = 10^6 and nsave = $2.5 \cdot 10^4$.

Furthermore, we compare the confidence regions via their coverage maps $\eta \mapsto P_{\vartheta_0}(\widehat{K}_{1-\gamma} \ni \eta)$ and the distribution of their area, that is, the distribution of $\lambda^2(\widehat{K}_{1-\gamma}(\mathbb{H}, \Xi))$ under ϑ_0 (see Figure 7.11 and 7.12).

In the case of a non-empty (fixed) boundary condition $\xi_{W_t^c}$, we observe that the MPLE based confidence region (Method I) for the nominal levels $1 - \gamma = 0.50, 0.90$, and 0.95 seems to have a coverage rate that is too low. However, a slight increase of the nominal level leads to an admissible coverage. To understand the impact of the (fixed) boundary condition $\xi_{W_t^c}$, we may compare these results with the empty and random boundary condition. The empty boundary condition constitutes an extreme case where the statistical model does not exhibit any artificial boundary parameter. Except for the nominal level $1 - \gamma = 0.95$, all coverage rates are admissible. The significantly higher coverages for 0.50 might be a result of the correlation of the MPLE components that tend to be stronger due to extreme MPL estimates in combination with the empty boundary condition. In contrast, the random boundary condition shows results that are slightly worse than those in the non-empty (fixed) boundary condition.

The parametric bootstrap confidence regions constructed from Method II have a significantly higher coverage rate than required in each of the cases $0.50, 0.90, 0.95$ and for each alternative estimation method $\widehat{\lambda}$. This is due to the fact that the construction of the pivot set is based on the RRB, which just performs a complete sufficient data reduction according to $\sigma(\widetilde{S}^{W,\xi})$. Note that $\widetilde{S}^{W,\xi}$ still depends on reducible data at the boundary. In contrast, the center of that CR is the MLE which is even measurable w. r. t. $\sigma(\widetilde{S}^{W_1} \oplus_{B_1, \mathbf{0}})$ and, therefore, varies less (see Proposition 7.1). The same can be observed in the random boundary condition, although the coverage rates are somewhat lower. In the empty boundary condition the additional variance of the RRB does not occur. The distribution of the RRBs is much closer to that of the MLE. Hence, the coverage rates are much more sensitive to bias effects which might explain the slight coverage loss for the corresponding nominal levels.

Method III shows in the non-empty boundary condition admissible coverage rates for the case $1 - \gamma = 0.50$. For the required rate 0.90 , we need to consider a nominal level of $1 - \gamma = 0.96$ to get an admissible coverage. None of the considered nominal levels $1 - \gamma$ yields an admissible result for a 0.95 -confidence region. In the case of the random boundary condition, the results are mostly worse again, although an admissible 0.95 -confidence region is obtained for a nominal level of 0.99 . For the empty boundary condition, we may observe a convincing agreement of the nominal and the effective coverage level.

Figure 7.11 and 7.12 show the coverage maps of Method I, II, and III for the required rates 0.50 and 0.90 . To make them comparable, we consider the lowest nominal level that yields an admissible coverage.

According to these plots, we may conclude that in every considered case Method II with $\hat{\lambda} = \text{VARE}$ yields a confidence region that is preferred to the remaining two confidence procedures given by Method I and II. In addition, Figure 7.10 shows that also from a computational point of view Method II with $\hat{\lambda} = \text{VARE}$ is computationally faster than the remaining procedures.

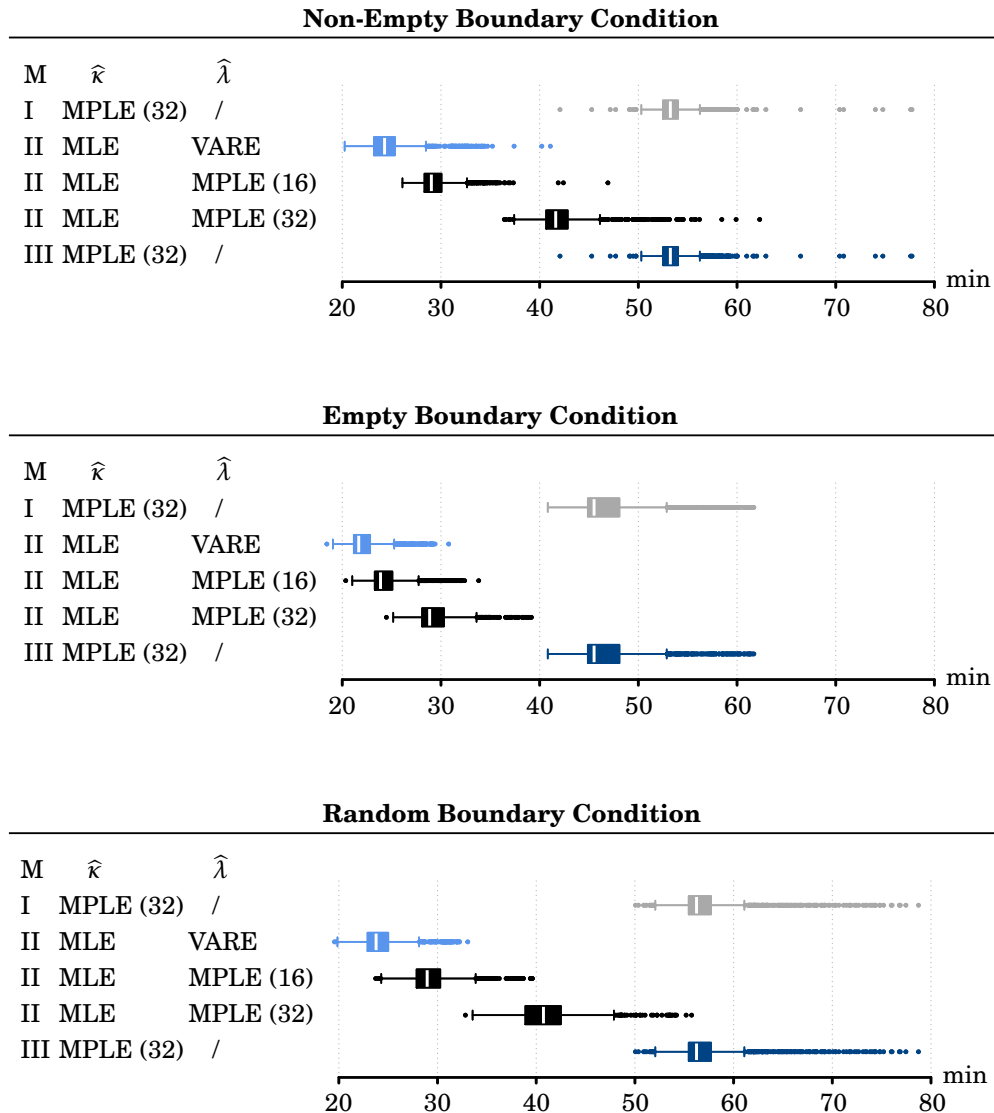


Figure 7.10: . Boxplots of the computation times of the parametric bootstrap confidence regions for all nominal levels $1 - \gamma$ (shown in Table 7.13 combined). Method I and III differ from each other in the additional computation of the sufficient statistics and the fitting of a linear model by the `lm`-function in R. This additional computational effort, however, is in the range of seconds and therefore not reflected in this graphic. The colorings have been chosen in accordance with Figures 7.11 and 7.12 that show the corresponding coverage maps. The computations were performed using a single core on a Linux server (Intel(R) Xeon(R) Silver 4114 with 2.8 GHz).

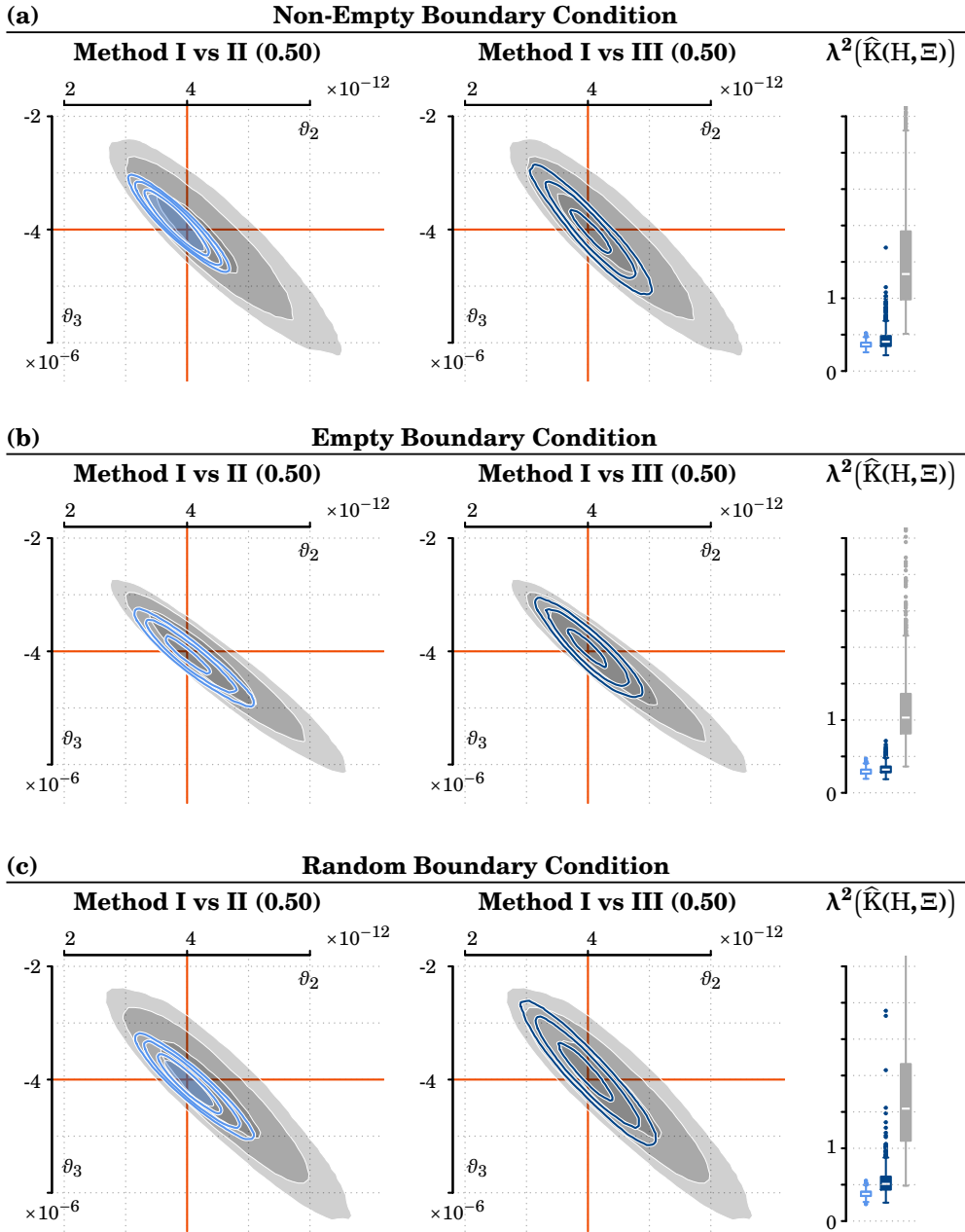


Figure 7.11: Left and middle: Contours (at 0.1, 0.2, 0.4) of the coverage map $\eta \mapsto P_{\theta_0}(\widehat{K}_{0.5}(H, \Xi) \ni \eta)$. Right: distributions of Lebesgue areas of 0.50 parametric bootstrap confidence regions. All three methods (see Method I (gray), II (light blue), and III (dark blue)) and all three boundary conditions (from top to bottom) are considered.

In addition to the contour levels 0.1, 0.2, 0.4, the light blue filled inner area represents the region where 0.50 coverage of Method II holds (visible in the non-empty and random boundary condition).

According to the simulation results (see Table 7.13), the coverage map of Method I for the non-empty and random boundary condition is shown for the nominal level of 0.51 and 0.53, respectively. The coverage map of Method II (case: empty boundary condition) is shown for nominal level of 0.51. Furthermore, the coverage map of Method III (case: random boundary condition) is shown for the nominal level of 0.58.

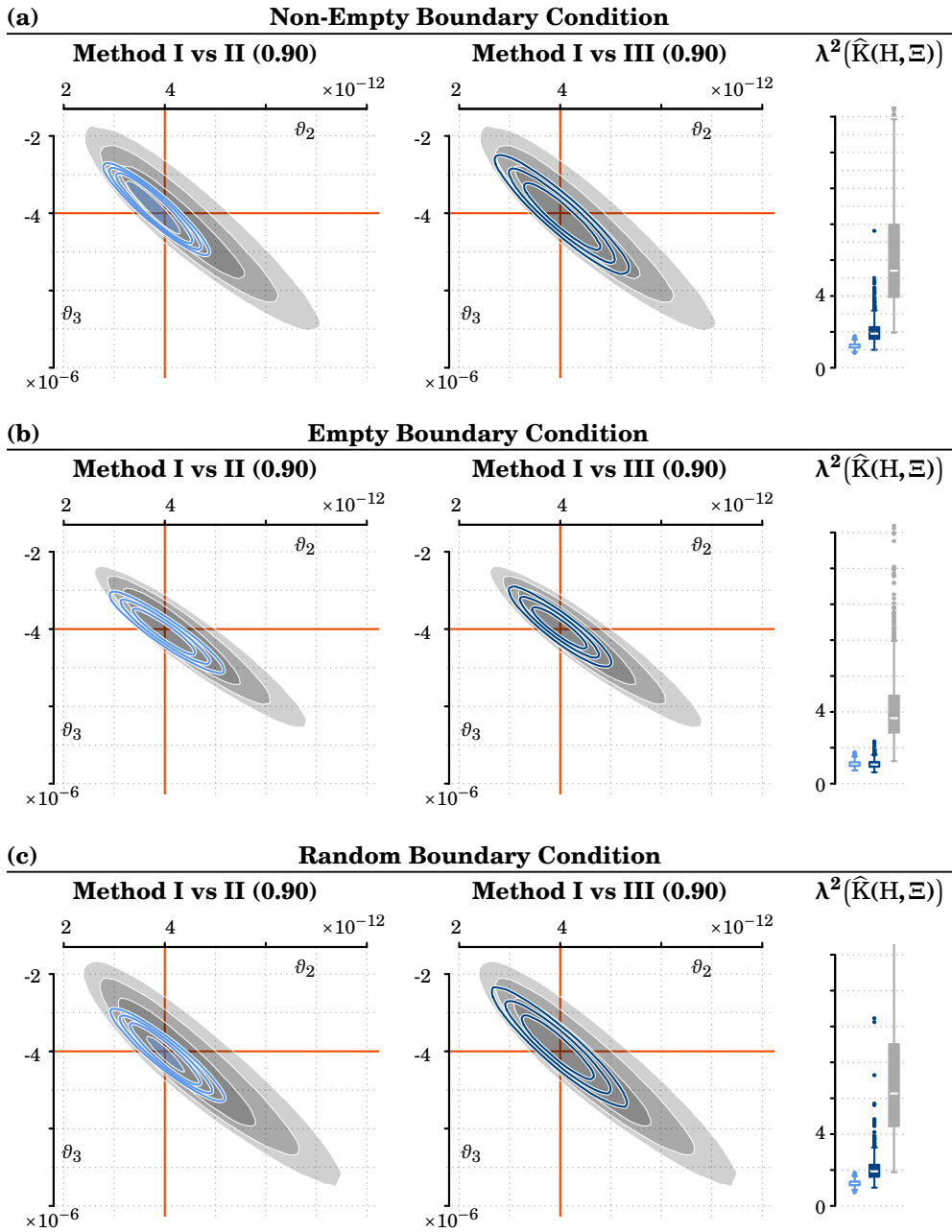


Figure 7.12: Left and middle: Contours (at 0.3, 0.5, 0.7) of the coverage map $\eta \mapsto P_{\theta_0}(\widehat{K}_{0.9}(H, \Xi) \ni \eta)$. Right: distributions of Lebesgue areas of 0.90 parametric bootstrap confidence regions. All three methods (see Method I (gray), II (light blue), and III (dark blue)) and all three boundary conditions (from top to bottom) are considered.

In addition to the contour levels 0.3, 0.5, 0.7, the light blue filled inner area represents the region where 0.90 coverage of Method II holds (visible in the non-empty and random boundary condition).

According to the simulation results (see Table 7.13), the coverage map of Method I for the non-empty and the random boundary condition is shown for the nominal level of 0.93 and 0.90, respectively. The coverage map of Method II (case: empty boundary condition) is shown for $1 - \gamma = 0.93$. Furthermore, the coverage map of Method III for either the non-empty and the random boundary condition is shown for the nominal level of 0.96.

Non-Empty Boundary Condition

Required Rate $1 - \alpha$			0.50						0.90					0.95				
Method	$\hat{\kappa}$	$\hat{\lambda}$	Nominal Level $1 - \gamma$															
			0.50	0.51	0.52	0.53	0.54	0.55	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
I	MPLE (32)	/	480	488	498	512	518	527	853	865	883	894	901	918	931	944	961	974
II	MLE	VARE	734	742	751	757	761	767	960	964	969	975	981	983	984	988	990	996
		MPLE (16)	877	882	886	888	894	897	989	990	991	991	991	993	995	996	998	998
		MPLE (32)	729	737	747	752	755	761	961	965	966	971	977	982	984	985	989	997
III	MPLE (32)	/	486	492	499	507	516	531	834	846	858	865	872	886	900	909	918	932

Empty Boundary Condition

Required Rate $1 - \alpha$			0.50						0.90					0.95				
Method	$\hat{\kappa}$	$\hat{\lambda}$	Nominal Level $1 - \gamma$															
			0.50	0.51	0.52	0.53	0.54	0.55	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
I	MPLE (32)	/	602	609	617	626	631	639	908	912	919	927	934	940	951	966	971	991
II	MLE	VARE	476	487	494	499	504	515	867	878	885	897	905	921	934	943	958	973
		MPLE (16)	572	586	596	603	617	623	922	930	936	946	954	956	964	969	976	983
		MPLE (32)	490	498	509	521	528	534	873	884	893	901	907	917	935	947	960	974
III	MPLE (32)	/	505	511	515	526	535	545	906	916	920	932	944	949	960	967	976	982

Random Boundary Condition

Required Rate $1 - \alpha$			0.50						0.90					0.95				
Method	$\hat{\kappa}$	$\hat{\lambda}$	Nominal Level $1 - \gamma$															
			0.50	0.51	0.52	0.53	0.54	0.55	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
I	MPLE (32)	/	453	462	471	484	493	508	848	854	865	882	899	905	919	937	958	975
II	MLE	VARE	665	673	684	694	704	713	955	960	963	968	973	978	982	987	991	998
		MPLE (16)	850	854	861	863	869	872	975	980	981	981	984	986	989	992	995	997
		MPLE (32)	664	675	682	692	703	711	954	958	963	964	973	976	981	984	989	995
III	MPLE (32)	/	412	425	429	437	448	454	807	822	837	851	862	876	894	907	922	942

Table 7.13: Absolute number of coverages of the canonical interaction parameters $\vartheta_2 = 4 \cdot 10^{-12}$, $\vartheta_3 = -4 \cdot 10^{-6}$ for 1000 bootstrap confidence regions generated according to Method I to III. The inverse temperature is $t = 1$ and three boundary conditions are studied. Gray colored cells mark cases where the coverage rates $P_{\vartheta}(\hat{K}_{1-\alpha} \ni \vartheta)$ are significantly (level 0.1) lower than $1 - \alpha$ for the required rates $\alpha = 0.50$, 0.90 , and 0.95 . The corresponding critical values are 480, 888, and 941, respectively.

Discussion and Outlook

How can the concept of Rao–Blackwellization deal with misspecification issues in statistical models?

In fact, incorrectly specified statistical models require some form of *stability* in Rao–Blackwellizations of estimators. That means, if we consider a Rao–Blackwellization $\mathbf{E}^{(*)}(\hat{\kappa}|S)$ of an estimator $\hat{\kappa}$ to estimate a parameter κ , we allow for at most small changes of $\mathbf{E}^{(*)}(\hat{\kappa}|S)$ to account for a small perturbation of the parameter κ that would lead to the “true”, even if unknown parameter function κ_{true} (see also Definition 2.8, page 34). The classical Rao–Blackwellization is not stable in general. Therefore, the present thesis proposes the concept of regularized Rao–Blackwellization (RRB) to deal with misspecified statistical models.

In the following we discuss aspects and consequences and, subsequently, obstacles and limits of the RRB concept. An outlook of the major and minor related research questions and their applications concludes this thesis.

Aspects and Consequences of the RRB Concept

Tackling Misspecifications We introduced a Lehmann–Scheffé (LS) norm that measures the ancillarity of first-order and paves a way to an understanding of the classical Rao–Blackwellization as the best approximation w. r. t. to that LS norm (see Theorem 4.1). Using regularization techniques, we are able to quantify and reduce “almost” ancillary information by including prior information. In other words, through prior information we consider a data reduction that goes beyond a reduction to the minimal sufficient information. On the one hand, this can be done smoothly through a regularization parameter using Tikhonov regularization, see especially Example 4.5 and 4.6 for the Gaussian location model. The Ivanov approach, on the other hand, includes the prior information by restricting the class of solutions of the Rao–Blackwell inverse problem to a prespecified finite-dimensional subspace. As an example one can consider

the linear space that is spanned by the maximum likelihood estimator (and a constant function). This can be justified by the fact that the MLE is optimal in a certain sense (see Theorem 2.21). Furthermore, the MLE is known to be stable against misspecified statistical models (see Example 2.9).

Relations Between Optimality Concepts We may encounter situations (see e. g. Example 4.8 and 4.9) where for a parameter of interest the maximum likelihood, the optimal unbiased, and the Bayes estimator (w. r. t. squared error loss) are all different, although each of these estimators is optimal in a certain sense (see Definition 1.12 for the Bayes estimator w. r. t. squared error loss; an optimal unbiased estimator is optimal w. r. t. $\leq_{\mathcal{L}}$, see Definition 2.10; see e. g. Theorem 2.21 for an optimal fixed sample size property of the MLE). Basically, these concepts are in fact only loosely connected. For example, the construction principle of optimal unbiased estimators yields estimators that can be arbitrarily different from corresponding MLEs, due to the ill-posedness of the Rao–Blackwell inverse problem.

The concept of RRB provides links between these optimality concepts, concerning the canonical parameter. The Tikhonov regularized Rao–Blackwellization reveals a relation to the Bayes estimator (w. r. t. squared error loss) in terms of the bias of the initial estimator, see Theorem 4.10. More generally, the improvement of the Tikhonov regularized Rao–Blackwellization compared to the classical Rao–Blackwellization depends continuously on the bias of the initial estimator, see Theorem 4.12. Note also that under regularity assumptions the MLE is implicitly defined through an unbiased estimating equation. With regard to Remark 2.3, we may therefore consider the MLE as approximately unbiased. Hence, the 1-Tikhonov regularized Rao-Blackwellization of the MLE is correspondingly close to the Bayes estimator (see also Example 4.8 and 4.9).

The Ivanov regularized Rao-Blackwellization reveals a relation between the MLE and the concept of unbiased estimators via Theorem 4.16 and Theorem 4.20.

Computational Feasibility (Ivanov’s Approach) We provided arguments to understand the Ivanov regularized Rao–Blackwellization approximately as a Markov operator (see Theorem 4.16). Considering the finite-dimensional space that is spanned by the MLE and constant functions, we also computed a first-order Taylor approximation of the corresponding Markov operator applied to an arbitrary estimator. Since this Taylor approximation turns out to be a linear regression formula, it is simple to implement

(see Theorem 4.20). In Section 7.4 we applied these considerations to Gibbs point processes in a simulation study. According to these results the RRB concept seems to provide a competitive approach to compute functionals of the distribution of the MLE.

Obstacles and Limits of the RRB Concept

Computational Feasibility (Tikhonov's Approach) Example 4.5 and 4.6 showed that the computation of the Tikhonov regularized Rao–Blackwell projection even in a very well-behaved model, the Gaussian location model, can be quite complicated. Theorem 4.7 yields the Tikhonov regularized Rao–Blackwell projection for the regularization parameter $\alpha = 1$ only. Due to this analytical obstacle, the Tikhonov regularized Rao–Blackwellization concept seems to be limited to tractable statistical models.

Need of Prior Knowledge Regularization requires prior knowledge or at least a prior guess. This concerns the “amount of misspecification”¹ and hence the choice of the regularization parameter in the Tikhonov regularization approach. The presented Ivanov's approach makes use of prior information through the prespecified finite-dimensional space. With regard to the proposed finite-dimensional space spanned by the MLE, this requirement of prior information leads to the necessity of a single MLE computation. In addition, both approaches use an *a priori distribution* π on the statistical model. This may be given at least implicitly by the distribution of the complete sufficient statistic under the assumed underlying distribution P_{ϑ_0} .

Focus on Fixed Sample Situations To develop the RRB concept, we introduced an analytical framework which depends on the sample size. Therefore, the present thesis focuses only on the situation of fixed sample sizes, that is, the non-asymptotic situation.

Outlook and Related Questions

Tikhonov regularized Rao–Blackwellization (further Examples) In Example 4.5 and 4.6 we computed the α -Tikhonov regularized Rao–Blackwell projection explicitly, at least on a two-dimensional subspace. Further examples would be of interest to understand how the regularization might come in, for example, concerning the relation of the MLE and the Bayes estimator.

¹This corresponds to the so-called *noise level* in the theory of regularization.

Optimality Results The well-known Rao–Blackwell–Lehmann–Scheffé theorem (the combination of Theorems 2.17 and 2.18) yields an optimality result concerning unbiased estimators. Theorem 4.12 of the present thesis already provides a corresponding version for the Tikhonov regularized Rao–Blackwellization where we consider the canonical parameter.

It would be of interest to generalize Theorem 4.12 to a larger class of parameter functions κ . Furthermore, it would also be of interest to establish such a result concerning the performance of the Ivanov regularized Rao–Blackwellization, of course in relation to the optimality properties that are fulfilled by the basis elements of the finite-dimensional subspace.

Computation of Ivanov regularization RB We have shown that the Ivanov regularized Rao–Blackwellization is connected to a linear regression. Further investigations concerning the computation of the Ivanov regularized RB, for example as Galerkin-Solution, would be interesting (see also Remark 4.15).

Asymptotic Properties In the present thesis, we are only concerned with fixed sample size properties and thus with non-asymptotic statistics. For example, the introduced Lehmann–Scheffé topologies and the Lehmann–Scheffé spaces depend on the statistical model and therefore on the sample size. To make use of the asymptotic properties of simple alternative estimators (the strong consistency and the asymptotic normality of the MPLE and the VARE in the point process statistic) for the RRB concept, more general notions in a more general framework have to be developed.

Bounded Mean Oscillation Spaces Girardi (1991) and Balder *et al.* (1994) discuss characterizations of the norm compactness in Lebesgue spaces through weak compactness and, in addition, through a so-called *Bocce criterium*. In Chapter 3 we have shown relations between the Lehmann–Scheffé topologies and the weak topologies on the Lebesgue spaces, and also between the canonical topology $\mathcal{T}_{L_p(\mathbb{B})}$ and the norm topology $\tau(\|\cdot\|_{Q,p})$. An interesting question may be how the characterization results of norm compactness can be generalized to the relation between the Lehmann–Scheffé topology and the canonical topology, especially with regard to regularizations in the sense of Ivanov.

Gibbs Point Process Statistics (Methodology) In the simulation study (see Section 7.4), we considered the Lennard-Jones model in the plane with different boundary conditions. Further investigations concerning other (more) complex models would be of interest. This also includes a study of such models on the sphere.

In the present thesis, we make use of the RRB concept to construct parametric confidence regions. We make use of two qualitative assumptions (see assumption (a) and (b) on page 175) that should be studied more closely with respect to quantitative concerns in concrete models. Furthermore, it would be of interest how the RRB concept can be used to construct (one-sided) statistical tests.

Gibbs Point Process Statistics (Data Examples) With regard to the previous paragraph on the statistical methodology of Gibbs point processes, we also want to see these methods at work. In the introduction we already used research questions concerning the distribution of envelope spike proteins of viruses as a motivation for the methods developed and considered in this thesis. Applications, for examples, to real data of such spike proteins may lead to a further development of more complex Gibbs point process models and consequently to a further development of the methodology considered here.

Applications to Other Areas of Statistics In Section 7.4 we have shown that the concept of regularized Rao–Blackwellization may be interesting and useful for Gibbs point processes. However, we presented the RRB concept in a much more general framework. Since exponential families provide a broad range of applications, it would be interesting to study the advantages and limits of the RRB concept in other fields of statistics where complex dependent data structures occur.

Abbreviations and Notations

The following lists review conventions and notations used throughout this thesis. Some of them are very common. For notation that is less common or that we have introduced, a page number is provided where more details can be found.

Abbreviations

CR	confidence region
DLR	Dobrushin–Lanford–Ruelle, 128
GNZ	Georgii–Nguyen–Zessin, 122
i. i. d.	identically independently distributed
iff	if and only if
LJ	Lennard-Jones, 144
LS	Lehmann–Scheffé, 51
MLE	maximum likelihood estimator
MPLE	maximum pseudo-likelihood estimator
MSE	mean squared error
RRB	regularized Rao–Blackwellization
SVD	singular value decomposition
TVS	topological vector space, 9
UMVUE	uniformly minimum variance unbiased estimator

Analysis, Measure Theory, and Topology

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	natural numbers, integers, rational numbers, real numbers
\mathbb{N}_0	natural numbers including 0
\mathbb{R}_+	non-negative real numbers
$\overline{\mathbb{R}}$	real numbers including $+\infty$ and $-\infty$
$\overline{\mathbb{R}}_+$	non-negative real numbers including $+\infty$
$\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$	general state spaces
\subset, \subsetneq	subset and proper subset relation, x
$\prod_{x \in \mathfrak{X}} \mathfrak{Y}_x$	Cartesian product of \mathfrak{Y}_x for $x \in \mathfrak{X}$
$\mathfrak{Y}^{\mathfrak{X}}$	(\mathfrak{X} -)Cartesian power of \mathfrak{Y} , set of all functions from \mathfrak{X} to \mathfrak{Y}

\mathfrak{X}^n	n th Cartesian power of \mathfrak{X} , for some $n \in \mathbb{N}$; identifiable with e. g. $\mathfrak{X}^{\{1, \dots, n\}}$
$2^{\mathfrak{X}}$	power set of \mathfrak{X} ; identifiable with e. g. $\{0, 1\}^{\mathfrak{X}}$
$\mathbf{id}_{\mathfrak{X}}$	identity map on \mathfrak{X} , i. e., $x \mapsto x$
$\mathbf{1}_M$	indicator function with indicator set M
pr_x	projection on the x th component of elements of $\prod_{x \in \mathfrak{X}} \mathfrak{Y}_t$
f^+, f^-	positive, negative part of the function f
\mathbf{t}_u	translation map $x \mapsto x + u$, where $u \in \mathbb{R}^d$, 121
$\text{diam}(M)$	$:= \sup\{d(x, y) : x, y \in M\}$ diameter of a set M of a metric space (\mathfrak{X}, d) , 143
$\text{dist}(x, M)$	$:= \inf\{d(x, y) : y \in M\}$ the distance of a point x to a subset M of a metric space (\mathfrak{X}, d) , 147
$\text{int}(M), \text{cl}(M)$	interior, closure of a set M , 32
$\ \cdot\ _{P,p}$	Lebesgue space norm of $L_p(\mathfrak{X}, \mathcal{A}, P)$, 8
$\ \cdot\ _P := \ \cdot\ _{P,1}$	Lebesgue space norm of $L_1(\mathfrak{X}, \mathcal{A}, P)$, 8
$\ \cdot\ _{\text{LS},p}$	Lehmann–Scheffé norm, 61
$\ \cdot\ _{\text{os},p}$	oscillation norm, 67
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	σ -algebras
$\sigma(\mathcal{C}) = \sigma_{\mathfrak{X}}(\mathcal{C})$	smallest σ -algebra which contains a set-system $\mathcal{C} \subset 2^{\mathfrak{X}}$ of a state space \mathfrak{X}
$\mathcal{C} \vee \mathcal{D}$	$:= \sigma(\mathcal{C} \cup \mathcal{D})$ maximum of the σ -algebras \mathcal{C} and \mathcal{D} , 14
$\mathcal{B}(\mathfrak{X}) = \mathcal{B}(\mathfrak{X}, \mathcal{T})$	Borel- σ -algebra w. r. t. the topological space $(\mathfrak{X}, \mathcal{T})$
\mathcal{T}, \mathcal{S}	topologies, 51
$\tau(f_\alpha : \alpha \in I)$	smallest topology such that $f_\alpha : \mathfrak{X} \rightarrow (\mathfrak{Y}, \mathcal{S})$ for $\alpha \in I$ are continuous, 51
$\mathcal{T}_{L_p(\mathfrak{P})}$	$:= \tau(\ \cdot\ _{P,p} : P \in \mathfrak{P})$ canonical topology on the space $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, 9
$\mathcal{T}_{\text{LS}}(\mathfrak{P}) := \mathcal{T}_{\text{LS}}(L_p(\mathfrak{P}), \mathfrak{P})$	$:= \tau(f \mapsto \int f dP : P \in \mathfrak{P})$ the Lehmann–Scheffé topology w. r. t. \mathfrak{P} , 51
$\mathcal{T}_{\text{LS}}^*(L_p(\mathfrak{P})) := \mathcal{T}_{\text{LS}}^*(\mathfrak{P}, L_p(\mathfrak{P}))$	$:= \tau(P \mapsto \int f dP : f \in L_p(\mathfrak{P}))$ the Lehmann–Scheffé- $*$ -topology w. r. t. \mathfrak{P} , 52
$\mathcal{T}_{\mathbf{p}} := \mathcal{T}_{\mathbf{p}}(F)$	topology of pointwise convergence on a function space F , 34
$\sigma_{\mathbf{w}}(E, F)$	$:= \tau(f : f \in F)$ weak topology of linear space E w. r. t. the subspace $F \subset E^*$, 53
$\sigma_{\mathbf{w}}(F, E)$	$:= \tau(e : e \in E)$ weak- $*$ -topology on the subspace $F \subset E^*$, where E is a linear space, 53
$\mathcal{H} = \mathcal{H}(\mathfrak{X})$	system of relatively compact Borel measurable subsets of \mathfrak{X} , 104
ζ	counting measure, 136
λ	Lebesgue measure

$\mu \perp \nu$	measure μ is singular to measure ν , 13
$\mu^f, \mu f^{-1}, \mu \circ f^{-1}$	image measures of f under μ
$KL, K \circ L$	$:= \int_{\mathfrak{Y}} L(y, \cdot) K(\cdot, dy)$ for a (Markov-)kernel K from $(\mathfrak{X}, \mathcal{A})$ to $(\mathfrak{Y}, \mathcal{B})$ and a (Markov-)kernel L from $(\mathfrak{Y}, \mathcal{B})$ to $(\mathfrak{Z}, \mathcal{C})$, 5
$\mu \otimes K$	product measure of measure μ with kernel K
$\mu \cdot K, \mu K$	second marginal measure of $\mu \otimes K$
E, F	vector spaces
E/F	quotient space of vector space E w. r. t. a linear subspace $F \subset E$, 8
$\text{span}(M)$	linear span of a set M of a vectors
$\text{conv}(M)$	convex hull of a set M of vectors, 32
pr_F	projection on a sub-space F of E
E^*, E'	algebraic dual space of a vector space E , topological dual space of a TVS (E, \mathcal{T}) , 52
F^\perp	annihilator of the subspace $F \subset E$, where (E, \mathcal{T}) is a TVS, 22
$\ker(T)$	kernel (null space) of an operator T , 18
$\text{ran}(T)$	range of an operator T , 19
T'	adjoint operator of an operator T between two topological spaces, 71
$\mathfrak{L}(\mathfrak{X}, \mathcal{A})$	vector space of \mathcal{A} -measurable functions, 55
$L(P) := L(\mathfrak{X}, \mathcal{A}, P)$	vector space of P -a. s. equivalence classes of \mathcal{A} -measurable functions, 55
$\mathfrak{L}_p(P) := \mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, P; \mathbb{R}^k)$	vector space consisting of p th power P -integrable functions (\mathbb{R}^k -valued) in $\mathfrak{L}(\mathfrak{X}, \mathcal{A})$, 8
$\mathfrak{L}_p(\mathfrak{P}) := \mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$	intersection of $\mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, P; \mathbb{R}^k)$ over all $P \in \mathfrak{P}$, 8
$[\mu], [\mu_\alpha : \alpha \in I]$	set of all (measurable) functions that are almost surely $\mathbf{0}$ w. r. t. the measure μ ; w. r. t. the family of measures $(\mu_\alpha)_{\alpha \in I}$, 49
$L_p(\mathfrak{P}) := L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$	quotient space of $\mathfrak{L}_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P}; \mathbb{R}^k)$ w. r. t. to the subspace $[\mathfrak{P}]$, 8
$LS_{\pi, p}(\mathfrak{P}) = LS_{\pi, p}(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$	Lehmann–Scheffé space, 61
$A_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$	vector subspace of all first order ancillary functions in $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, 15
$A_p^0(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$	elements g of $A_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$ with $\int g dP = \mathbf{0}$ for all $P \in \mathfrak{P}$, 18
$\mathfrak{B} = \mathfrak{B}(\mathfrak{X}, \mathcal{A})$	set of all bounded, measurable functions f from $(\mathfrak{X}, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, 25
$\mathbf{C}(\mathfrak{X}, \mathcal{T})$	set of \mathbb{R} -valued continuous functions on $(\mathfrak{X}, \mathcal{T})$, 108
$\mathfrak{S}(\mathfrak{X}, \mathcal{A})$	set of all finite signed measures on $(\mathfrak{X}, \mathcal{A})$, 52
$\mathfrak{M}^* := \mathfrak{M}^*(\mathfrak{X}) := \mathfrak{M}^*(\mathfrak{X}, \mathcal{A})$	set of all measures on a measurable space $(\mathfrak{X}, \mathcal{A})$, 102

$\mathfrak{M} := \mathfrak{M}(\mathfrak{X}) := \mathfrak{M}(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$	set of all locally finite measures on a topological space $(\mathfrak{X}, \mathcal{T})$, 103
$\mathfrak{M}_B = \mathfrak{M}_B(\mathfrak{X})$	set of locally finite measures on \mathfrak{X} with no mass outside B , 106
$\mathfrak{N}^* := \mathfrak{N}^*(\mathfrak{X}) := \mathfrak{N}^*(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$	set of all locally finite point measures on $(\mathfrak{X}, \mathcal{T})$, 104
$\mathfrak{N} := \mathfrak{N}(\mathfrak{X}) := \mathfrak{N}(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$	set of all simple locally finite point measures on $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$, 104
$\mathfrak{A}_{\text{lf}} := \mathfrak{A}_{\text{lf}}(\mathfrak{X}) := \mathfrak{A}_{\text{lf}}(\mathfrak{X}, \mathcal{T})$	set of all locally finite subsets in a topological space $(\mathfrak{X}, \mathcal{T})$, 105
ξ, η, ζ	point pattern, simple locally finite point measures, 104
$\mathfrak{N}_B = \mathfrak{N}_B(\mathfrak{X})$	set of locally finite point measures on \mathfrak{X} with no mass outside B , 106
ξ_B	$= \text{pr}_B(\xi) = \xi(\cdot \cap B)$, that is, the restriction of a point pattern $\xi \in \mathfrak{N}(\mathfrak{X})$ to $B \in \mathfrak{B}(\mathfrak{X})$, 106
$\mathfrak{N}_k = \mathfrak{N}_k(\mathfrak{X})$	set of simple point patterns on \mathfrak{X} with $\xi(\mathfrak{X}) = k$ for some $k \in \mathbb{N}$, 142
$\mathfrak{N}_f = \mathfrak{N}_f(\mathfrak{X})$	set of simple finite point patterns on \mathfrak{X} , 123
$D_H^{(t)}(U)$	set of all point patterns where the potential energy U is tempered, 146
$D_H^{(s)}(U)$	set of all point patterns where the potential energy U is stable, 146
$\mathcal{M}^* := \mathcal{M}^*(\mathfrak{X}, \mathcal{A})$	$:= \sigma(\mu \mapsto \mu(A) : A \in \mathcal{A})$, 102
$\mathcal{M}_B = \mathcal{M}_B(\mathfrak{X})$	σ -algebra of events observable inside B , 106
$\mathcal{N} = \mathcal{N}(\mathfrak{X})$	trace- σ -algebra $\mathcal{M}(\mathfrak{X}) \cap \mathfrak{N}(\mathfrak{X})$, 105
\mathcal{F}_H	$:= \text{pr}_H^{-1}(\mathcal{N}_H)$; σ -algebra of events observable from the inside of $H \in \mathcal{H}$, 126
\mathcal{F}^H	$:= \text{pr}_{H^c}^{-1}(\mathcal{N}_{H^c})$; σ -algebra of events observable from the outside of $H \in \mathcal{H}$, 126
$\mathcal{T}_{\mathbf{F}}$	Fell topology, 108
\mathcal{T}_{loc}	topology of local convergence, 152
$\mathcal{T}_{\text{loc}}^{\mathcal{C}}$	topology of \mathcal{C} -local convergence, 154
$\mathfrak{B}_{\text{loc}} = \mathfrak{B}_{\text{loc}}(\mathcal{F})$	set of all local functions $f : \mathfrak{N} \rightarrow \mathbb{R}$ w. r. t. filtration \mathcal{F} , 152
$\mathfrak{B}_{\text{qloc}} = \mathfrak{B}_{\text{qloc}}(\mathcal{F})$	set of all quasilocal functions $f : \mathfrak{N} \rightarrow \mathbb{R}$ w. r. t. filtration \mathcal{F} , 152
$\mathfrak{B}_{\text{qloc}}^{\mathcal{C}}$	set of all \mathcal{C} -quasilocal functions, 154

Probability Theory and Mathematical Statistics

$(\Omega, \mathcal{F}, \mathbf{P})$	general underlying probability space
X, Y, Z	random elements
$X \stackrel{d}{=} Y$	identically distributed random elements X, Y

$X \perp\!\!\!\perp Y$	independent random elements X and Y w. r. t. a probability space $(\Omega, \mathcal{F}, \mathbf{P})$
$\mathcal{D} \perp\!\!\!\perp \mathcal{E}$	independent set systems \mathcal{D} and \mathcal{E} w. r. t. an underlying probability space
$\mathbf{P}^X, \mathbf{P}X^{-1}, \mathbf{P} \circ X^{-1}$	distribution of X under P
$\text{Prob}(\mathfrak{X}, \mathcal{A})$	set of all distributions on $(\mathfrak{X}, \mathcal{A})$, 3
δ_x	Dirac distribution in $x \in \mathfrak{X}$, 55
$\mathbf{B}(n, \pi)$	Binomial distribution with parameter $n \in \mathbb{N}$ and $\pi \in [0, 1]$, 56
$\mathbf{Poi}(\alpha)$	Poisson distribution with parameter $\alpha > 0$, 114
\mathbf{U}_H	uniform distribution on $H \in \mathcal{H}(\mathbb{R}^k)$ such that $0 < \lambda(H) < \infty$, 64
$\mathbf{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2 , 57
$\mathbf{G}(a, b)$	Gamma distribution with parameters $a, b > 0$, 81
\mathbf{C}	Campbell measure of a point process Ξ , 120
\mathbf{C}^\dagger	reduced Campbell measure of a point process Ξ , 121
$\mathbf{Pa}, \mathbf{Pa}_x$	Palm kernel of a point process Ξ , Palm distribution of a point process Ξ at x , 120
$\mathbf{Pa}^\dagger, \mathbf{Pa}_x^\dagger$	reduced Palm kernel of a point process Ξ , reduced Palm distribution of a point process Ξ at x , 121
$\mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \lambda)$	Poisson point process distribution with parameter measure $\lambda \in \mathfrak{M}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}))$, 114
$\mathbf{Pop}_{\mathfrak{X}}, \mathbf{Pop}_\lambda$	short notation for $\mathbf{Pop}(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \lambda)$ emphasizing the set \mathfrak{X} or the parameter measure λ , 114
$\mathbf{Pop}_G * \mathbf{Pop}_H$	convolution of two Poisson point process distributions $\mathbf{Pop}_G, \mathbf{Pop}_H$ where $G, H \in \mathcal{H}$ and $G \cap H = \emptyset$, 134
$\Gamma = (\Gamma_H)_{H \in \mathcal{H}}$	specification, that is, a family of quasi-Markov kernels, 127
$U = (U_H)_{H \in \mathcal{H}}$	potential energy, Hamiltonian, 142
$(U_H^{\text{Strauss}})_{H \in \mathcal{H}}$	Strauss potential energy, 143
$(U_H^{\text{LJ}})_{H \in \mathcal{H}}$	Lennard-Jones potential energy, 144
$\mathbf{Gibbs}(\Gamma)$	set of Gibbs distribution to the specification Γ , 127
$\mathfrak{P} \subset \text{Prob}(\mathfrak{X}, \mathcal{A})$	(unparametrized) statistical model, 3
\mathfrak{P}^T	$:= \{P^T : P \in \mathfrak{P}\}$
$(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$	statistical space, 3
$\mathfrak{P} \ll \mu$	statistical model \mathfrak{P} is dominated by μ , 4
$(P_\vartheta)_{\vartheta \in \Theta} \in \text{Prob}(\mathfrak{X}, \mathcal{A})^\Theta$	parametrized statistical model with parameter space Θ , 4

$(\mathfrak{F}, \mathcal{F}, \boldsymbol{\pi})$	statistical model with measure theoretical structure, 61
$\mathbf{E}_P X$	$:= \int X dP$ expectation of X w. r. t. P
$\mathbf{E}_{P_\vartheta} X$	$:= \int X dP_\vartheta$ expectation of X w. r. t. P_ϑ
\mathbf{E}	expectation operator, $f \mapsto (\int f dP)_{P \in \mathfrak{P}}$
$\mathbf{E}_{\mathcal{C}}$	$:= \mathbf{E} _{L_p(\mathfrak{X}, \mathcal{C}, \mathfrak{P})}$ expectation operator restricted to \mathcal{C} -measurable functions, 69
\mathbf{V}	oscillation operator, 67
$\mathbf{E}_P(f \mathcal{C})$	conditional expectation of f given \mathcal{C} w. r. t. P in $L_1(\mathfrak{X}, \mathcal{C}, P)$, 9
$\mathbf{E}(f \mathcal{C})$	Rao–Blackwellization of f given \mathcal{C} , 17
$\mathbf{E}^{(*)}(f \mathcal{C})$	unspecified substitute for the Rao–Blackwellization of f given \mathcal{C} , 51
$\mathbf{E}^{(\alpha)}(f \mathcal{C})$	Tikhonov regularized Rao–Blackwellization of f given \mathcal{C} with regularization parameter $\alpha > 0$, 71
$\mathbf{E}^{(G)}(f \mathcal{C})$	quasi Rao–Blackwellization of f given \mathcal{C} , 88
$\mathbf{E}^{(LR)}(\hat{\lambda} S)$	regularized Rao–Blackwellization of $\hat{\lambda}$ given S ; first-order Taylor approx. of $\mathbf{E}_{\hat{\theta}_{ML}(\cdot)} \hat{\lambda}$, 97
$\hat{\kappa}, \hat{\lambda}$	estimators
$\hat{\mathbf{K}}_{1-\alpha}, \hat{\Lambda}_{1-\alpha}$	$(1 - \alpha)$ -confidence regions
$\hat{\vartheta}_{ML}$	maximum likelihood estimator for the canonical parameter \mathbf{id}_Θ , 31
$\hat{\vartheta}_{ML}(\cdot, \xi_{W^c})$	maximum likelihood estimator w. r. t. the boundary point pattern $\xi_{W^c} \in \mathfrak{R}(W^c)$ (for the canonical parameter \mathbf{id}_Θ), 158
$\hat{\vartheta}_{MPL}$	maximum pseudo-likelihood estimator for the canonical parameter \mathbf{id}_Θ , 161
$\hat{\vartheta}_{VAR}$	variational estimator for the canonical parameter \mathbf{id}_Θ , 167
$\mathfrak{H} = \mathfrak{H}(\mathfrak{F})$	set of all regular estimating functions, 36
$\text{Bias}(\hat{\kappa}, \vartheta)$	$:= \mathbf{E}_\vartheta \hat{\kappa} - \vartheta$ bias of an estimator $\hat{\kappa}$ under P_ϑ
$\text{MSE}(\hat{\kappa}, \vartheta)$	(multivariate) mean squared error of an estimator $\hat{\kappa}$ under P_ϑ
$\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi})$	Bayes risk of $\hat{\kappa}$ w. r. t. the prior $\boldsymbol{\pi}$ (and squared error loss), 10
$\mathbf{r}(\hat{\kappa}, \boldsymbol{\pi}; W)$	Bayes risk w. r. t. the $\ \cdot\ _W$ -weighted squared error loss of an estimator $\hat{\kappa}$, 10
\leq_L	Löwner order, 11
$\leq_{\mathcal{L}}$	partial order for estimators w. r. t. a class of convex loss functions \mathcal{L} , 35

Figures and Tables

A	Schematic profile of an enveloped virus	xiii
B	Realization of spike proteins distributed according to a homogeneous Poisson point process on a unit sphere	xiii
C	Realization of spike proteins distributed according to a Strauss point process on a unit sphere	xiii
D	Realization of spike proteins distributed according to a (12, 6)-Lennard-Jones point process on a unit sphere	xiii
2.1	Illustration of the theory of optimal unbiased estimation through one diagram	41
4.1	Illustration of the definition of the regularized Rao–Blackwellization	71
4.2	Graphs of the Tikhonov regularized Rao–Blackwellizations of the MLE (Gaussian model)	77
4.3	Graphs of the MLE, the UMVUE, and their 1-Tikhonov regularized Rao–Blackwellizations (Uniform model)	84
4.4	Graph of the 1-Tikhonov regularized Rao–Blackwellization of any unbiased estimator for \mathbf{id}_Θ (Uniform model)	84
4.5	Performance comparison of the MLE, the UMVUE, and their 1-Tikhonov regularized Rao–Blackwellization, empirically through box-plots (Uniform model)	87
4.6	Performance comparison of the MLE, the UMVUE, and their 1-Tikhonov regularized Rao–Blackwellization, empirically through MSE graphs (Uniform model)	87
4.7	MLE approximation through an approximate Ivanov regularized Rao–Blackwellization (Exponential model)	97
6.1	Strauss pair potential as a function of the particle distance	145
6.2	Graph of the (12, 6)-Lennard-Jones pair potential as a function of the particle distance	145
6.3	Strauss point process realization in the unit square	148
6.4	Lennard-Jones point process realization in the unit square	148
6.5	Strauss point process realization on the unit sphere	148
6.6	Lennard-Jones point process realization on the unit sphere	149

7.1	Graph of the cut-off (12,6)-Lennard-Jones pair potential as a function of the particle distance	169
7.2	Lennard-Jones realizations at different inverse temperatures (non-scaled)	170
7.3	Lennard-Jones realizations at different inverse temperatures (scaled observation windows)	170
7.4	Boxplots of the sufficient statistics of 1000 realizations of the cut-off Lennard-Jones point process in the simulation study	171
7.5	Boxplots of the empirical computation time of the MLE, MPLE and VARE	173
7.6	Table of empirical mean squared error of the MLE, MPLE, VARE and their RRBs for the Lennard-Jones interaction parameters with (non-)empty boundary condition based on one thousand realizations	174
7.7	Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition (VARE, Ivanov RRB of the VARE, and MLE)	176
7.8	Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition (MPLE ($n_d = 32$), Ivanov RRB of the MPLE, and MLE)	177
7.9	Empirical marginal and joint distributions of the estimators for the Lennard-Jones interaction parameters with (non-)empty boundary condition (MPLE ($n_d = 16$), Ivanov RRB of the MPLE, and MLE)	178
7.10	Boxplots of computation times of the parametric bootstrap confidence regions (Methods I, II and III)	183
7.11	Coverage maps of 0.50 confidence regions for different boundary conditions	184
7.12	Coverage maps of 0.90 confidence regions for different boundary conditions	185
7.13	Table of absolute number of coverages of one thousand parametric bootstrap confidence regions at the canonical parameter	186

References

References marked with an asterisk were found discussed in secondary sources but have not been accessible to the author. To most entries a DOI (digital object identifier) link via <https://www.doi.org> is provided.

- Allen, M. P. & Tildesley, D. J. (2017). *Computer simulation of liquids*. Oxford University Press, New York. DOI: 10.1007/BF00646086
- Almeida, M. P. & Gidas, B. (1993). A variational method for estimating the parameters of MRF from complete or incomplete data. *The Annals of Applied Probability* **3**, 103–136. DOI: 10.1214/AOAP/1177005510
- Applebaum, D. (2009). *Lévy processes and stochastic calculus*. Second edition. Cambridge Studies in Advanced Mathematics, Vol. 116. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511809781
- Baddeley, A., Coeurjolly, J.-F., Rubak, E. & Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika* **101**, 377–392. DOI: 10.1093/BIOMET/AST060
- Baddeley, A. & Dereudre, D. (2013). Variational estimators for the parameters of Gibbs point process models. *Bernoulli* **19**, 905–930. DOI: 10.3150/12-BEJ419
- Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Australian & New Zealand Journal of Statistics* **42**, 283–322. DOI: 10.1111/1467-842X.00128
- Baddeley, A., Turner, R., Mateu, J. & Bevan, A. (2013). Hybrids of Gibbs point process models and their implementation. *Journal of Statistical Software* **55**, 1–43. DOI: 10.18637/JSS.V055.I11
- Baddeley, A. J. & Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42. DOI: 10.18637/JSS.V012.I06
- Bahadur, R. R. (1957). On unbiased estimates of uniformly minimum variance. *Sankhyā* **18**, 211–224.
- Balder, E., Girardi, M. & Jalby, V. (1994). From weak to strong types of L_1 -convergence by the Bocce criterion. *Studia Mathematica* **111**, 241–262.
- Barankin, E. W. (1950). Extension of a theorem of Blackwell. *The Annals of Mathematical Statistics* **21**, 280–284. DOI: 10.1214/AOMS/1177729845
- Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. Wiley Series in Probability and Statistics. John Wiley &

- Sons, Chichester. Reprint, John Wiley & Sons, Chichester, 2014. DOI: 10.1002/9781118857281
- Barra, J.-R. (1981). *Mathematical basis of statistics*. Translation from the French edited by L. Herbach. Academic Press, New York.
- Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15**, 377–380. DOI: 10.1007/978-1-4419-5825-9_14
- (1958). On statistics independent of sufficient statistics. *Sankhyā* **20**, 223–226. DOI: 10.1007/978-1-4419-5825-9_16
- Bauer, H. (1991). *Wahrscheinlichkeitstheorie*. Vierte, völlig überarbeitete und neugestaltete Auflage. De Gruyter-Lehrbuch. Walter de Gruyter, Berlin.
- (1992). *Maß- und Integrationstheorie*. Zweite, überarbeitete Auflage. De Gruyter-Lehrbuch. Walter de Gruyter, Berlin.
- (2001). *Measure and integration theory*. Translation from the German by R. B. Burckel. De Gruyter Studies in Mathematics, Vol. 26. Walter de Gruyter, Berlin. DOI: 10.1515/9783110866209
- Baumeister, J. (1987). *Stable solution of inverse problems*. Advanced Lectures in Mathematics. Friedrich Vieweg & Sons, Braunschweig. DOI: 10.1007/978-3-322-83967-1
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 192–236. DOI: 10.1111/J.2517-6161.1974.TB00999.X
- (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)* **24**, 179–195. DOI: 10.2307/2987782
- (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* **47**, 77–92.
- Besag, J., Milne, R. & Zachary, S. (1982). Point process limits of lattice processes. *Journal of Applied Probability* **19**, 210–216.
- Bhappkar, V. P. (1972). On a measure of efficiency of an estimating equation. *Sankhyā: Series A* **34**, 467–472.
- Billingsley, P. (1995). *Probability and measure*. Tird edition. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.
- Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *The Annals of Mathematical Statistics* **18**, 105–110. DOI: 10.1214/AOMS/1177730497
- Bogachev, V. I. (1998). *Gaussian measures*. Mathematical Surveys and Monographs, Vol. 62. American Mathematical Society, Providence (Rhode Island). DOI: 10.1090/SURV/062
- (2007). *Measure theory, Volume I*. Springer-Verlag, Berlin.
- Braess, D. (2003). *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Dritte, korrigierte und ergänzte Auflage. Springer-Verlag, Berlin.
- Chandrasekar, B. & Kale, B. (1984). Unbiased statistical estimation functions for parameters in presence of nuisance parameters. *Journal of Statistical Planning and Inference* **9**, 45–54. DOI: 10.1016/0378-3758(84)90043-0
- Coeurjolly, J.-F., Guan, Y., Khanmohammadi, M. & Waagepetersen, R. (2016).

- Towards optimal Takacs-Fiksel estimation. *Spatial Statistics* **18**, 396–411. DOI: 10.1016/J.SPASTA.2016.08.002
- Coeurjolly, J.-F. & Møller, J. (2014). Variational approach for spatial point process intensity estimation. *Bernoulli* **20**, 1097–1125. DOI: 10.3150/13-BEJ516
- Coeurjolly, J.-F. & Rubak, E. (2013). Fast covariance estimation for innovations computed from a spatial Gibbs point process. *Scandinavian Journal of Statistics* **40**, 669–684. DOI: 10.1111/SJOS.12017
- Daley, D. J. & Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*. Second edition. Probability and its Applications. Springer-Verlag, New York. DOI: 10.1007/978-0-387-49835-5
- Dereudre, D. (2019). Introduction to the theory of Gibbs point processes. In D. Coupier, ed., *Stochastic geometry: Modern research frontiers*, Lecture Notes in Mathematics, Vol. 2237. Springer-Verlag, Cham, pp. 181–229. DOI: 10.1007/978-3-030-13547-8_5
- Dereudre, D. & Lavancier, F. (2017). Consistency of likelihood estimation for Gibbs point processes. *The Annals of Statistics* **45**, 744–770. DOI: 10.1214/16-AOS1466
- Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. & Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes. *International Statistical Review* **62**, 99–117. DOI: 10.2307/1403548
- Dobrushin, R. L. (1968a). The description of the random field by its conditional distributions and its regularity conditions. *Teoriya Veroyatnostei i ee Primeneniya* **13**, 201–229.
- (1968b). The problem of uniqueness of a Gibbsian random field and the problem of phase transitions. *Functional Analysis and its Applications* **2**, 302–312.
- Dudley, R. M. (2002). *Real analysis and probability*. Revised and extended reprint of [Wadsworth and Brooks, Pacific Grove, California, 1989]. Cambridge Studies in Advanced Mathematics, Vol. 74. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511755347
- Elstrodt, J. (2018). *Maß- und Integrationstheorie*. Achte, erweiterte und aktualisierte Auflage. Springer-Lehrbuch. Springer-Verlag, Berlin. DOI: 10.1007/978-3-662-57939-8
- Engl, H., Hanke, M. & Neubauer, A. (1996). *Regularization of inverse problems*. Mathematics and Its Applications, Vol. 375. Kluwer Academic Publishers, Dordrecht.
- Fischer, H. (1959). Limesräume. *Mathematische Annalen* **137**, 269–303. DOI: 10.1007/BF01360965
- Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* **41**, 155–160.
- (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **222**, 309–368.
- Fu, J. C. & Li, L. A. (1992). Method of Pao-Zhuan Yin-Yu: A method of stochastic

- point estimation. *Statistica Sinica* **2**, 171–188.
- Fu, J. C. & Li, L.-A. (1998). Generalized method of Pao-Zhuan Yin-Yu. *Statistics & Probability Letters* **37**, 321–329. DOI: 10.1016/S0167-7152(97)00127-2
- Gaetan, C. & Guyon, X. (2010). *Spatial statistics and modeling*. Translated from the French by Kevin Bleakley. Springer Series in Statistics. Springer-Verlag, New York. DOI: 10.1007/978-0-387-92257-7
- Georgii, H.-O. (2011). *Gibbs measures and phase transitions*. Second edition. De Gruyter Studies in Mathematics, Vol. 9. Walter de Gruyter. DOI: 10.1515/9783110250329
- Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)* **54**, 657–699. DOI: 10.1111/J.2517-6161.1992.TB01443.X
- Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: Developed with especial reference to the rational foundation of thermodynamics*. Yale University Press, New Haven. Reprint, Dover Publications, New York, 1960.
- Girardi, M. (1991). Weak vs. norm compactness in L_1 : the Bocce criterion. *Studia Mathematica* **98**, 95–97. DOI: 10.4064/SM-98-1-95-97
- Glötzl, E. (1980). Bemerkungen zu einer Arbeit von O. K. Kozlov. *Mathematische Nachrichten* **94**, 277–289. DOI: 10.1002/MANA.19800940116
- Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* **31**, 1208–1211. DOI: 10.1214/AOMS/1177705693
- Godambe, V. P. & Heyde, C. C. (2010). Quasi-likelihood and optimal estimation. In R. Maller, I. Basawa, P. Hall & E. Seneta, eds., *Selected Works of C.C. Heyde*. Springer-Verlag, New York, pp. 386–399. DOI: 10.1007/978-1-4419-5823-5_49
- Guan, Y., Jalilian, A. & Waagepetersen, R. (2015). Quasi-likelihood for spatial point processes. *Journal of the Royal Statistical Society: Series B (Methodological)* **77**, 677–697. DOI: 10.1111/RSSB.12083
- Habek, M., Höllwarth, H. & Schuhmacher, D. (2020). Simulation-based likelihood maximization with multiple importance sampling. Unpublished
- *Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 49–52.
- Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. Yale University Press, New Haven.
- Halmos, P. R. (1946). The theory of unbiased estimation. *The Annals of Mathematical Statistics* **17**, 34–43. DOI: 10.1214/AOMS/1177731020
- Halmos, P. R. & Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *The Annals of Mathematical Statistics* **20**, 225–241. DOI: 10.1214/AOMS/1177730032
- Hammer, H.-W., Nogga, A. & Schwenk, A. (2013). Colloquium: Three-body forces: From cold atoms to nuclei. *Reviews of Modern Physics* **85**, 197–217. DOI: 10.1103/REVMODPHYS.85.197
- Heyde, C. C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation*. Springer Series in Statistics. Springer-Verlag, New York.

- Hodges, J. L., Jr. & Lehmann, E. L. (1950). Some problems in minimax point estimation. *The Annals of Mathematical Statistics* **21**, 182–197. DOI: 10.1214/AOMS/1177729838
- Höllwarth, H. (2015). *Beiträge zur Mathematischen Stichprobentheorie: Statistische Modellbildung mit Stichprobendesigns und anderen Morphismen*. Best-Masters. Springer-Spektrum, Wiesbaden.
- Höllwarth, H. (user Henning). (2020a). *Definition of random measures*. [Online]. MathOverflow. Available from: <https://mathoverflow.net/q/353316> (Accessed on February 22, 2020)
- (2020b). *Dual space of the intersection of locally convex vector spaces*. [Online]. MathOverflow. Available from: <https://mathoverflow.net/q/356955> (Accessed on April 8, 2020)
- Hu, Z. & Yang, R.-C. (2013). A new distribution-free approach to constructing the confidence region for multiple parameters. *PloS ONE* **8**, e81179. DOI: 10.1371/JOURNAL.PONE.0081179
- *Ivanov, V. K. (1962). Integral equations of the first kind and approximate solution of the inverse problem of potential theory. *Doklady Akademii Nauk SSSR* **142**, 998–1000.
- Ivanov, V. K., Vasin, V. V. & Tanana, V. P. (2002). *Theory of linear ill-posed problems and its applications*. Second edition. Inverse and Ill-posed Problems Series. VSP, Utrecht. Translated and revised version of the 1978 Russian original. DOI: 10.1515/9783110944822
- Jensen, J. L. & Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *The Annals of Applied Probability* **1**, 445–461. DOI: 10.1214/AOAP/1177005877
- Kagan, A. M., Malinovsky, Y. & Mattner, L. (2014). Partially complete sufficient statistics are jointly complete. *Teorija verojatnostej i ee primenenija* **59**, 542–561. (2015). *Theory of Probability & Its Applications*, **59**, 359–374. DOI: 10.1137/S0040585X97T987223
- Kallenberg, O. (1983). *Random measures*. Third, revised and enlarged edition. Akademie-Verlag, Berlin.
- (2002). *Foundations of modern probability*. Second edition. Probability and its Applications. Springer-Verlag, New York. DOI: 10.1007/978-1-4757-4015-8
- (2017). *Random measures, theory and applications*. Probability Theory and Stochastic Modelling, Vol. 77. Springer-Verlag, Cham. DOI: 10.1007/978-3-319-41598-7
- Kelley, J. L. (1975). *General topology*. Revised reprint edition [Van Nostrand, New York, 1955]. Graduate Texts in Mathematics, Vol. 27. Springer-Verlag, New York.
- Kelley, J. L., Namioka, I. & eight co-authors (1976). *Linear topological spaces*. Revised reprint edition [Van Nostrand, Princeton, 1963]. Graduate Texts in Mathematics, Vol. 36. Springer-Verlag, Berlin. DOI: 10.1007/978-3-662-41914-4
- Kelly, F. P. & Ripley, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–360.

- Klein, J. S. & Bjorkman, P. J. (2010). Few and far between: how HIV may be evading antibody avidity. *PLoS pathogens* **6**. DOI: 10.1371/JOURNAL.PPAT.1000908
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, Berlin. Reprint, Springer-Verlag, Berlin, 1973. DOI: 10.1007/978-3-642-49888-6
- (1950). Unbiased estimators. *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya* **14**, 303–326. Translation from the Russian by G. Lindquist in: A. N. Shiryaev, ed., *Selected works of A.N. Kolmogorov: Volume II, Probability and mathematical statistics*, Mathematics and its applications, Vol. 26, Kluwer Academic, Dordrecht, 1992, pp. 369–394. DOI: 10.1007/978-94-011-2260-3_38
- Kowalsky, H.-J. (1954). Limesräume und Komplettierung. *Mathematische Nachrichten* **12**, 301–340. DOI: 10.1002/MANA.19540120504
- Kress, R. (1999). *Linear integral equations*. Second edition. Applied Mathematical Sciences, Vol. 82. Springer-Verlag, New York. DOI: 10.1007/978-1-4612-0559-3
- Kwaśnicki, M. (user Mateusz Kwaśnicki). (2020). *Eigenspace of Gaussian Markov operator*. [Online]. MathOverflow. Available from: <https://mathoverflow.net/q/361161> (Accessed on May 23, 2020)
- Lanford III, O. E. & Ruelle, D. (1969). Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics* **13**, 194–215. DOI: 10.1007/BF01645487
- Lang, S. (1993). *Real and functional analysis*. Third edition. Graduate Texts in Mathematics, Vol. 142. Springer-Verlag, New York. DOI: 10.1007/978-1-4612-0897-6
- Larribe, F. & Fearnhead, P. (2011). On composite likelihoods in statistical genetics. *Statistica Sinica* **21**, 43–69.
- Last, G. (1990). Some remarks on conditional distributions for point processes. *Stochastic Processes and their Applications* **34**, 121–135. DOI: 10.1016/0304-4149(90)90059-2
- Lehmann, E. L. (1981). An interpretation of completeness and Basu's theorem. *Journal of the American Statistical Association* **76**, 335–340. DOI: 10.1080/01621459.1981.10477652
- Lehmann, E. L. & Casella, G. (1998). *Theory of point estimation*. Second edition. Springer Texts in Statistics. Springer, New York. DOI: 10.1007/b98854
- Lehmann, E. L. & Scheffé, H. (1947). On the problem of similar regions. *Proceedings of the National Academy of Sciences of the United States of America* **33**, 382–386.
- (1950). Completeness, similar regions and unbiased estimation: Part I. *Sankhyā: The Indian Journal of Statistics* **10**, 305–340.
- (1955). Completeness, similar regions, and unbiased estimation: Part II. *Sankhyā: The Indian Journal of Statistics* **15**, 219–236.
- (1956). Corrigenda: Completeness, similar regions, and unbiased estimation: Part I. *Sankhyā: The Indian Journal of Statistics* **17**, 250.
- Lennard-Jones, J. E. (1924). On the determination of molecular fields. —I. from the variation of the viscosity of a gas with temperature. *Proceedings of the*

- Royal Society (London)* **106**, 463–477.
- Liese, F. & Miescke, K.-J. (2008). *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer Series in Statistics. Springer-Verlag, New York. DOI: 10.1007/978-0-387-73194-0
- Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu, ed., *Statistical inference from stochastic processes: Proceedings of a Summer Research Conference held August 9-15, 1987*, Contemporary Mathematics, Vol. 80. American Mathematical Society, Providence (Rhode Island), pp. 221–239. DOI: 10.1090/CONM/080/999014
- Mase, S. (1992). Uniform LAN condition of planar Gibbsian point processes and optimality of maximum likelihood estimators of soft-core potential functions. *Probability Theory and Related Fields* **92**, 51–67. DOI: 10.1007/BF01205236
- Mateu, M. G. (2013). Introduction: The structural basis of virus function. In M. G. Mateu, ed., *Structure and physics of viruses: An integrated textbook*, Subcellular Biochemistry, Vol. 68. Springer-Verlag, Dordrecht, pp. 3–51. DOI: 10.1007/978-94-007-6552-8_1
- Matthes, K., Kerstan, J. & Mecke, J. (1978). *Infinitely divisible point processes*. Translated from the German by B. Simon. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester.
- Matthes, K., Warmuth, W. & Mecke, J. (1979). Bemerkungen zu einer Arbeit von Nguyen Xuan Xanh und Hans Zessin. *Mathematische Nachrichten* **88**, 117–127. DOI: 10.1002/MANA.19790880110
- Mattner, L. (1999). Product measurability, parameter integrals, and a Fubini counterexample. *L'Enseignement Mathématique* **45**, 271–279. DOI: 10.5169/SEALS-64449
- Molchanov, I. (2005). *Theory of random sets*. Probability and Its Applications. Springer-Verlag, London. DOI: 10.1007/1-84628-150-4
- Møller, J. & Waagepetersen, R. P. (2004). *Statistical inference and simulation for spatial point processes*. Monographs on Statistics and Applied Probability, Vol. 100. Chapman & Hall/CRC, Boca Raton. DOI: 10.1201/9780203496930
- Newman, J. (user Julian Newman). (2020). *Definition of random measures*. [Online]. MathOverflow. Available from: <https://mathoverflow.net/q/353280> (Accessed on February 22, 2020)
- Nguyen, X. & Zessin, H. (1976). Punktprozesse mit Wechselwirkung. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **37**, 91–126.
- Penttinen, A. (1984). Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method. *Jyväskylä Studies in Computer Science, Economics and Statistics* **7**, 1–107.
- Pfanzagl, J. (1994). *Parametric statistical theory*. De Gruyter Textbook. Walter de Gruyter, Berlin. DOI: 10.1515/9783110889765
- Preston, C. (1976). *Random fields*. Lecture Notes in Mathematics, Vol. 534. Springer-Verlag, Berlin. DOI: 10.1007/BFB0080563
- (2005). *Specifications and their Gibbs States*. [Online]. University of Bielefeld. Available from: <https://www.math.uni-bielefeld.de/~preston/rest/gibbs/files/specifications.pdf> (Accessed on July 30, 2020)

- Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* **37**, 81–91. Reprint in S. Kotz & N. L. Johnson, eds., *Breakthroughs in Statistics: Foundations and Basic Theory*. Springer-Verlag, New York, pp. 235–247. DOI: 10.1007/978-1-4612-0919-5_16
- (1952). Some theorems on minimum variance estimation. *Sankhyā* **12**, 27–42.
- Rauchenschwandtner, B. (1978). *Gibbsprozesse und Papangeloukerne*. Dissertationen der Johannes Kepler-Universität Linz, Vol. 17. Verband der wissenschaftlichen Gesellschaften Österreichs, Wien.
- Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511624131
- Ripley, B. D. & Kelly, F. P. (1977). Markov point processes. *Journal of the London Mathematical Society. Second Series* **15**, 188–192. DOI: 10.1112/JLMS/S2-15.1.188
- Rudin, W. (1973). *Functional analysis*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill, New York.
- Ruelle, D. (1969). *Statistical mechanics: Rigorous results*. W. A. Benjamin, New York.
- Schaefer, H. H. (1971). *Topological vector spaces*. Third corrected edition. Graduate Texts in Mathematics, Vol. 3. Springer-Verlag, New York. DOI: 10.1007/978-1-4684-9928-5
- Schmetterer, L. (1960). On unbiased estimation. *The Annals of Mathematical Statistics* **31**, 1154–1163. DOI: 10.1214/AOMS/1177705686
- (1977). On the theory of unbiased estimation. In *Proceedings of the Symposium to Honour Jerzy Neyman*. Polish Scientific Publishers, Warszawa, pp. 313–317.
- (1978). Einige Resultate aus der Theorie der erwartungstreuen Schätzungen. In J. Kozesnik, ed., *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions and the Eighth European Meeting of Statisticians, Volume B*. D. Reidel Publishing Company, Dordrecht, pp. 489–503.
- Schmetterer, L. & Strasser, H. (1974). Zur Theorie der erwartungstreuen Schätzungen. *Anzeiger der Österreichischen Akademie der Wissenschaften. Mathematisch-Naturwissenschaftliche Klasse* **76**, 59–66.
- Schneider, R. & Weil, W. (2000). *Stochastische Geometrie*. Teubner Skripten zur Mathematischen Stochastik. B. G. Teubner, Stuttgart. DOI: 10.1007/978-3-322-80106-7
- Small, C. G. & McLeish, D. L. (1988). Generalizations of ancillarity, completeness and sufficiency in an inference function space. *The Annals of Statistics* **16**, 534–551. DOI: 10.1214/AOS/1176350819
- (1994). *Hilbert space methods in probability and statistical inference*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York. DOI: 10.1002/9781118165522
- Song, P. (2007). *Correlated data analysis: Modeling, analytics, and applications*. Springer Series in Statistics. Springer-Verlag, New York. DOI: 10.1007/978-0-387-71393-9

- Strasser, H. (1985). *Mathematical theory of statistics: Statistical experiments and asymptotic decision theory*. De Gruyter Studies in Mathematics, Vol. 7. Walter de Gruyter, Berlin. DOI: 10.1515/9783110850826
- Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–475. DOI: 10.1093/BIOMET/62.2.467
- Torgersen, E. (1991). *Comparison of statistical experiments*. Encyclopedia of Mathematics and its Applications, Vol. 36. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511666353
- van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. Imperial College Press, London. DOI: 10.1142/9781860949760
- Varin, C., Reid, N. & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* **21**, 5–42.
- Vasin, V. V. (2008). The method of quasi-solutions by Ivanov is the effective method of solving ill-posed problems. *Journal of Inverse and Ill-Posed Problems* **16**, 537–552. DOI: 10.1515/JIIP.2008.028
- Vaughan, H. E. (1937). On locally compact metrisable spaces. *Bulletin of the American Mathematical Society* **43**, 532–535. DOI: 10.1090/S0002-9904-1937-06593-1
- Vladimirov, V. S. (2002). *Methods of the theory of generalized functions*. Analytical Methods and Special Functions, Vol. 6. Taylor & Francis, London. DOI: 10.1201/9781482288162
- Werner, D. (2018). *Funktionalanalysis*. Achte Auflage. Springer-Verlag, Berlin. DOI: 10.1007/978-3-662-55407-4
- Witting, H. (1985). *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.
- Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J. D., Gris e, H., Ofek, G. A., Taylor, K. A. & Roux, K. H. (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847–852. DOI: 10.1038/NATURE04817

Index

- absolute continuity, 156
- adjoint operator, 71
- ancillarity, 13
 - of first order, 15, 25, 70
- annihilator, 22

- Basu's Theorem, 13, 16
- Bayes estimator
 - w.r.t. Löwner order, 11
 - w.r.t. weighted squared error loss, 11
- Bayes risk, 10
- Bernoulli
 - point process, 109
 - specification, 131
- binomial
 - model, 56
 - point process, 109
- boundary
 - condition
 - —, empty, 158
 - —, non-empty, 158
 - effects, 158–186
- bounded
 - kernel, 152
- bounded set
 - , metrically, 88, 104
 - , topologically, 104
- boundedly compact metric space, 88, 104, 126
- Brownian motion, 117

- \mathcal{C} -quasilocal functions, 154
- Campbell
 - measure, 120
 - —, compound, 123
 - —, reduced, 121
- Campbell's Theorem, 119
- canonical parameter, 27
- canonical topology on $L_p(\mathfrak{F})$, 8, 53
- chain, 128
- Chapman–Kolmogorov equation,
126, 131
- complementary subspaces, 18
- complete
 - statistical model, 16
- completeness
 - , w. r. t. a measure, 62
 - , bounded, 16
 - , of a σ -algebra, 16
 - , w. r. t. the class of two-valued functions, 16
- completion
 - of a σ -algebra, 163
- conditional
 - expectation, 9
 - intensity, 161
- confidence
 - level, 7
 - procedure, 7
 - region, x, 7
- construction principle
 - of unbiased estimators, 40
- contrast
 - , expected, 29
 - function, 29
- countably determined events, 102
- counting measure, 136
- coverage map, 7

- decision
 - rule, 6
 - space, 6
- density choice, 58
- derivative
 - , at a point pattern, 155
- diffuse measure, 115
- dissecting system, 160
- distribution
 - , Gamma, 81
 - , Gibbs, *see* Gibbs measure
 - , binomial, 56
 - , exponential, 81
 - , normal, 57
 - , uniform, 64, 131
- DLR
 - equations, 127, 130
 - state, *see* Gibbs measure
- dual space
 - , algebraic, 52
 - , topological, 20
- energy, 142
 - function, 141
- estimate
 - maximum likelihood, 31
- estimating
 - function
 - —, regular, 36
 - equation, 30
 - —, pseudo-likelihood, 162
 - function, 30
 - —, unbiased, xi, 30, 32, 41
- estimation
 - procedure, 7
- estimator, 7
 - , maximum likelihood, 31, 158
 - —, empirical distribution of, 96
 - , maximum pseudo-likelihood, 161
 - , point, 7
 - , unbiased, 30
 - , variational, 167
- expectation operator, 39, 49
- expected loss, 9
- exponential
 - distribution, 81, 97
 - family, ix, 26–28, 32, 64, 156, 158, 171
 - —, k -parametric, 27
 - —, full rank, 27
- exterior conditioning, 119
- F_0 -completeness, *see* completeness
 - w. r. t. to the class of two-valued functions
- Fell topology, 108, 116
- finite volume Gibbs measure, 134
- Fisher information, 32
- function
 - , Rademacher, 57
 - , hereditary, 133
 - , loss, 9
 - , \mathcal{C} -quasilocal, 154
- Gaussian model, 45, 74, 76
- Gibbs
 - distribution, *see* Gibbs measure
 - kernel, 123
 - measure, 99, 126
 - —, finite volume, 127, 133–135, 145
 - —, infinite volume, 127, 151–156
 - point process, 99, 125–127
 - —, statistics for, xi, 157–167
 - specification, 149
- GNZ equation, 122
- Godambe information, 37
- Hadamard's well-posedness criteria, 44
- Hamiltonian, 142
- hard sphere model, 143
- hereditary function, 133
- importance sampling, 160
- increments
 - , independent, 111, 114
 - , stationary, 111
- independent
 - complement of a σ -algebra, 13
 - increments, 111, 114
- inference function, *see* estimating function

- intensity
 - function, 119
 - measure, 119, 120
- interaction
 - potential
 - —, Strauss, 143
 - —, pair, 143
 - range, 143
 - —, finite, 157, 168
 - —, infinite, 144, 168
- interior conditioning, 119
- inverse
 - problem, 44
 - —, Rao–Blackwell, 41, 42, 50–68
 - — of unbiased estimation, 41
 - temperature, 140–141, 169–180
- Ising model, 140
 - , one-dimensional, 140
- Ising potential
 - energy, 140
- kernel
 - , Gibbs, 123
 - , Papangelou, 120, 122
 - of an operator, 18
- Löwner order, 11, 36
- Lax–Milgram Theorem, 72
- Lehmann–Scheffé
 - *-topology, 52
 - Theorem, 39
 - space, 61
 - topology, 51
- Lennard-Jones
 - , cut-off pair potential, 169
 - point process, 135, 147, 167–169
 - potential energy, 144
- likelihood, 31, 158
 - pseudo, 160
- limit space, 61
- local
 - function, 152
 - kernel, 152
- locally finite
 - measure, 103
 - point measure, 104
 - set, 105
 - simple point measure, 104
- log-likelihood, 158
- log-pseudo-likelihood, 161
- loss function, 9
- marginal distribution on a set, 112
- Markov
 - chain, 135
 - kernel
 - —, quasi, 127
 - —, strict, 127
 - point process, 128
 - process, 129
 - semigroup, 128
 - specification, 136
- maximum likelihood
 - , (weak) stability of, 34
 - estimation, 31–34, 158–160
 - estimator, 31, 158
 - —, invariance property, 33
- maximum pseudo-likelihood
 - estimate, 161
 - —, computation of, 164
 - estimation, 160–165
- mean squared error, 9, 36
- measure
 - , Campbell, 120
 - , counting, 136
 - , diffuse, 115
 - , finite
 - —, signed, 52
 - , intensity, 119
 - , locally finite, 103
 - — point, 104
 - , signed
 - —, finite, 117
- minimum contrast
 - approach, 30
 - estimate, 29
- misspecified statistical model, ix, 4, 33, 43
- MLE
 - , invariance property of, 33
- morphism
 - , statistical, 5

- MPLE, 161
- net, 60
 — convergent, 60
 noise level, 189
 null set
 — w. r. t. a statistical model \mathfrak{P} , 4
 null space, 67
- operator
 —, adjoint, 71
 —, coercive, 72
 —, expectation, 39, 49
 —, oscillation, 67
 order bounded, 59
 orthogonal sum, 19
 orthogonality w.r.t. a bilinear form, 19
- π -completeness, 62
 pair (interaction) potential, 143, 162
 —, Lennard-Jones, 144
 — —, cut-off, 169
 —, interaction range of a , 143
- Palm
 — distribution, 120
 — distribution reduced, 121
 — kernel, 120
 — —, reduced, 121
- Papangelou
 — conditional intensity, 122, 161
 — kernel, 120, 122
- parameter space, 3
- point measures
 —, disjoint, 105
 point pattern, 103
 —, differentiability at, 155
 point process, 109
 —, Bernoulli, 109
 —, Gibbs, 125, 127
 —, Lennard-Jones, 135, 147, 154, 167–169
 —, Markov, 128
 —, Poisson, 114
 — —, simple, 115
 —, Strauss, 147
- , binomial, 109
 —, marginal distribution on a set, 112
 —, stationarity of a , 121
 — distribution, 112
 —, F -component of a , 112
- Poisson
 — point process, 114, 129
 — process
 — —, homogeneous, 111
 — specification, 129
- Polish space, 14, 108
- potential, 141
 —, Lennard-Jones, 144
 —, of order k , 143
 —, pair (interaction), 143, 162
 — energy, 142–144, 149
 — —, Ruelle stable, 146
 — —, interaction order of, 143
 — —, parametrized, 157, 162
 — —, tempered, 146
- prior distribution, 10
- pivot element
 —, approximate, 175
- projection
 —, Rao–Blackwell, 9, 18
- projective family, 112
- properness, 126
- pseudo-likelihood, 160
 — estimating equation, 162
 — maximum, 161
- quasi
 — Rao–Blackwellization, 88
- quasilocal
 — function, 152
 — kernel, 152
- quasilocalness
 — w.r.t. a set-system \mathcal{C} , 154
- Rademacher function, 57
- Radon–Nikodým equations, 9, 50
- random
 — measure, 109
 — signed measure, 117
- range

- of an operator, 19
- Rao–Blackwell
 - Theorem, 38
 - inverse problem, 42, 45, 50–69, 187
 - projection, 1, 9, 18, 41
- Rao–Blackwellization, x, 18, 39
 - , α -regularized, 71
 - , classical, 50
 - , quasi, 88
 - , regularized, 45, 70
- reduced
 - Campbell measure, 121
 - Palm kernel, 121
- regular estimating function, 36
- regularization
 - parameter, 69
 - scheme, 69
- regularized
 - Rao–Blackwellization, 71
- risk function, 9
- Ritz–Galerkin solution, 89
- Ruelle stability, 146
- σ -algebra
 - , of events observable from the inside or outside, 126
- sample space, 3
- score, 32, 159
- signed measure
 - , finite, 52, 117
- simple
 - Poisson point process, 115
- singular element of a statistical model, 13
- singular measures, 13
- specification, 99, 127
 - , Bernoulli, 131
 - , Gibbs, 141, 146, 149
 - , Markov, 136
 - , Poisson, 129
 - , local, 125
- stability of estimator constructions, 34
 - , weak, 34
- standard extension argument, 17
- stationarity of point processes, 121
- statistical
 - model, 3, 61
 - —, complete, 16
 - —, dominated, 4
 - —, incorrectly specified, *see* misspecified
 - —, misspecified, ix, 4, 33, 43
 - —, parametrized, 4
 - — without singular element, 13
 - morphism, 5
 - procedure, x
 - space, 3
- stochastic process, 103, 129
 - , set-indexed, 110
 - with stationary and independent increments, 111
- Strauss
 - interaction potential, 143
 - point process, 135, 147
 - —, hard-core, 147, 153
- sufficiency, 13
- temperedness, 146
- test
 - alternative, 6
 - hypothesis, 6
- Theorem
 - , Basu's, 13, 16
 - , Campbell's, 119
 - , Lax–Milgram, 72
 - , Lehmann–Scheffé, 39
 - , Rao–Blackwell, 38
- Tikhonov
 - functional, 71
 - regularization, 45, 71
- topological
 - space
 - —, locally compact, 103
 - —, second countable, 103
 - vector space, 8
- topology
 - , Fell, 108, 116
 - , Lehmann–Scheffé, 51
 - , canonical on $L_p(\mathfrak{X}, \mathcal{A}, \mathfrak{P})$, 8
 - , vague, 108

- , weak, 53, 56
- , weak-*, 53
- of \mathcal{C} -quasilocal convergence, 154
- of local convergence, 152
- of pointwise convergence, 34, 55

- unbiased
 - estimation
 - —, stability of, 34

- uniformly absolutely continuous
 - w. r. t. a measure, 153

- vague topology, 108
- variational
 - equation, 122, 155
 - estimation, 166–167

- weak topology, 53
- weak-* -topology, 53

CURRICULUM VITAE

Name Henning Höllwarth
Date & Place of Birth December 17, 1985 in Malchin (Germany)
E-Mail henning.hoellwarth@math.tu-freiberg.de

EDUCATION

Dec 2020 PhD Thesis Defense (University of Göttingen)
Jul 2014 Diploma in Mathematics (University of Trier)
Jul 2005 Abitur (BSWV Neubrandenburg)

POSITIONS

Since Oct 2020 Research assistant (University of Freiberg)
Feb 2015 – Sep 2020 Research assistant (University of Göttingen)
Oct 2014 – Jan 2015 Research assistant (University of Trier)

TEACHING EXPERIENCE

Exercise Classes Spatial Stochastics (three semester cycle)
accompanying lectures, Stochastic Lab Course with R (block course)
lecture cycles, and Point Processes and Their Statistics
block courses Stochastic Calculus, Random Fields, and Their Statistics
Foundations of Random Processes
Maß- und Wahrscheinlichkeitstheorie

Seminars Markov Chains and Mixing Times
together with Finite Point Processes
Dominic Schuhmacher Statistical Inference for Spatial Point Processes

RESEARCH IN PROGRESS PAPER

with Michael Habeck and Dominic Schuhmacher. *Simulation-Based Likelihood Maximization with Multiple Importance Sampling*. unpublished paper, 2020.

CONTRIBUTIONS TO CONFERENCES

Poster on *Confidence Regions Based on Rao–Blackwellizations*. 20th Workshop on Stochastic Geometry, Stereology and Image Analysis, Sandbjerg Estate near Sønderborg, 2 – 7 June 2019.

Poster on *Parameter Estimation of Gibbs Point Process Models: A Comparative Study*. 12th German Probability and Stochastic Days, Bochum, 1 – 4 March 2016.

PUBLICATION

Beiträge zur Mathematischen Stichprobentheorie: Statistische Modellbildung mit Stichprobendesigns und anderen Morphismen. Springer-Spektrum, Wiesbaden, 2015.