Statistical Problems Related to Excitation Threshold and Reset Value of Membrane Potentials

Dissertation

zur Erlangung des Grades
"Doktor der Naturwissenschaften"
am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität Mainz
von

Patrick Jahn

geboren in Simmern

Mainz, 03. Februar 2009

Tag der mündlichen Prüfung: Montag, 30. März 2009

(D77) Dissertation, Johannes Gutenberg-Universität Mainz

Patrick Jahn

Statistical Problems Related to Excitation Threshold and Reset Value of Membrane Potentials

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Abstract

The present work is motivated by biological questions about the behavior of membrane potentials in neurons. A commonly used model for spiking neurons is to assume that between spikes the membrane potential is given by a diffusion process $X = (X_t)_{t \geq 0}$ which is a solution of an SDE

$$\mathrm{d}X_t = \beta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion. The spiking behavior is usually explained as follows. Whenever X reaches a certain *excitation threshold* S, a spike occurs. Thereafter the potential is set down to a certain *reset value* x_0 again.

In applications it is sometimes possible to observe the diffusion process X between spikes by estimating the drift coefficient $\beta(\cdot)$ and the diffusion coefficient $\sigma(\cdot)$ from real data. Nevertheless, S and x_0 have to be determined in order to fix the model. However, in real data this is not obvious at all.

One possible approach is to view x_0 and S as parameters in a statistical model and to estimate them. In the present work, we discuss four cases for which we assume the diffusion process between spikes is given by a *Brownian motion with drift*, a *geometric Brownian motion*, an *Ornstein-Uhlenbeck* process or a *Cox-Ingersoll-Ross* process. We further assume to observe iid inter spike times interpreted as level crossing times of X from x_0 to S. The first two cases are very similar and one can explicitly compute the *maximum likelihood* estimator. Moreover, we use LAN theory in order to get optimal results. The cases OU and CIR process are treated by a *minimum distance* method based on the comparison of empirical and true Laplace transform with respect to a Hilbert space norm. It will be shown that all estimators are strongly consistent and asymptotically normal. In the last chapter, we will check the performance of the minimum distance estimator by application to simulated data. Moreover, applications to real data sets are given, including a detailed discussion of the results.

Keywords and phrases: statistical inference for stochastic processes - neuronal modeling - diffusion integrate-and-fire models - membrane potential - reset value and excitation threshold - maximum likelihood estimation - LAN - minimum distance estimation

Zusammenfassung

Die vorliegende Arbeit ist motiviert durch biologische Fragestellungen bezüglich des Verhaltens von Membranpotentialen in Neuronen. Ein vielfach betrachtetes Modell für spikende Neuronen ist das Folgende. Zwischen den Spikes verhält sich das Membranpotential wie ein Diffusionsprozess $X = (X_t)_{t\geq 0}$ der durch die SDGL

$$\mathrm{d}X_t = \beta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t$$

gegeben ist, wobei $(B_t)_{t\geq 0}$ eine Standard-Brown'sche Bewegung bezeichnet. Spikes erklärt man wie folgt. Sobald das Potential X eine gewisse Exzitationsschwelle S überschreitet entsteht ein Spike. Danach wird das Potential wieder auf einen bestimmten Wert x_0 zurückgesetzt.

In Anwendungen ist es manchmal möglich, einen Diffusionsprozess X zwischen den Spikes zu beobachten und die Koeffizienten der SDGL $\beta(\cdot)$ und $\sigma(\cdot)$ zu schätzen. Dennoch ist es nötig, die Schwellen x_0 und S zu bestimmen um das Modell festzulegen.

Eine Möglichkeit, dieses Problem anzugehen, ist x_0 und S als Parameter eines statistischen Modells aufzufassen und diese zu schätzen. In der vorliegenden Arbeit werden vier verschiedene Fälle diskutiert, in denen wir jeweils annehmen, dass das Membranpotential X zwischen den Spikes eine *Brown'sche Bewegung mit Drift*, eine *geometrische Brown'sche Bewegung*, ein *Ornstein-Uhlenbeck* Prozess oder ein *Cox-Ingersoll-Ross* Prozess ist. Darüber hinaus beobachten wir die Zeiten zwischen aufeinander folgenden Spikes, die wir als iid Treffzeiten der Schwelle S von X gestartet in x_0 auffassen. Die ersten beiden Fälle ähneln sich sehr und man kann jeweils den Maximum-Likelihood-Schätzer explizit angeben. Darüber hinaus wird, unter Verwendung der LAN-Theorie, die Optimalität dieser Schätzer gezeigt. In den Fällen OU- und CIR-Prozess wählen wir eine Minimum-Distanz-Methode, die auf dem Vergleich von empirischer und wahrer Laplace-Transformation bezüglich einer Hilbertraumnorm beruht. Wir werden beweisen, dass alle Schätzer stark konsistent und asymptotisch normalverteilt sind. Im letzten Kapitel werden wir die Effizienz der Minimum-Distanz-Schätzer anhand simulierter Daten überprüfen. Ferner, werden Anwendungen auf reale Datensätze und deren Resultate ausführlich diskutiert.

Contents

Abbreviations

BMD	Brownian motion with drift
cadlag	continuité à droite et limites à gauche (right continuous and left limits)
CIR	Cox-Ingersoll-Ross
GBM	geometric Brownian motion
iid	independent, identically distributed
ISI	inter spike interval
LT	Laplace transform
MDE	minimum distance estimator
MLE	maximum likelihood estimator
OU	Ornstein-Uhlenbeck
RV	random variable
SDE	stochastic differential equation
SLLN	strong law of large numbers

Conventions

$$\sum_{i=n+1}^{n} \cdots \equiv 0$$

$$\prod_{i=n+1}^{n} \cdots \equiv 1$$

Symbols and Terms

x_0	reset value
S	excitation threshold
$\mathbb{N}$	natural numbers
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\mathbb{Z}$	integers
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	$= [0, \infty)$
$\bar{\mathbb{A}}$	closure of $\mathbb{A}$
x^+	$= \max\{0, x\}$ the positive part of x
$[r]$	$= \max\{k \in \mathbb{Z} \| k \leq r\}$
$(\alpha)_k$	$= \prod_{i=0}^{k-1}(\alpha + i)$ the Pochhammer symbol, where $(\alpha)_0 = 1$
$L^p(\Omega, \mathfrak{A}, \mu)$	$= \{f : \Omega \to \mathbb{R},\ \mathfrak{A}$-measurable and $\\|f\\|_p < \infty\}$, $p \in [1, \infty]$
$\\|f\\|_p$	$= (\int_\Omega \|f\|^p \mathrm{d}\mu)^{1/p}$, for $p \in [1, \infty)$
$\\|f\\|_\infty$	$= \sup_{x \in \Omega} \|f(x)\|$
$C^n(E)$	$= \{f : E \subset \mathbb{R} \to \mathbb{R},\ n$-times continuously differentiable$\}$
$\mathscr{B}(E)$	Borel σ-algebra of a metric space E
$\mathbb{1}_{\mathbb{M}}$	indicator function of the set $\mathbb{M}$
$\mathbb{M}^c$	complement of the set $\mathbb{M}$
$B_\varepsilon(x)$	open ball of radius $\varepsilon > 0$ centered at $x \in \mathbb{R}^d$
$\bar{B}_\varepsilon(x)$	closed ball of radius $\varepsilon > 0$ centered at $x \in \mathbb{R}^d$
$\xrightarrow[n\to\infty]{\mathcal{L}}$	weak-convergence as $n \to \infty$
$\xrightarrow[n\to\infty]{\mathbf{P}}$	$\mathbf{P}$-stochastic convergence as $n \to \infty$
$\xrightarrow[n\to\infty]{\mathbf{P}\text{-a.s.}}$	$\mathbf{P}$-almost sure convergence as $n \to \infty$
$\mathcal{L}(Z\|\mathbf{P})$	$= \mathbf{P}^Z$, the distribution of the RV Z under the probability measure $\mathbf{P}$
$Z \sim \mathbf{P}$	the distribution of Z is given by the probability measure $\mathbf{P}$
$\bar{x}_n^h$	$:= \left(\frac{1}{n}\sum_{i=1}^n x_i^{-1}\right)^{-1}$ the harmonic mean of the vector $(x_1, \ldots, x_n)$
$\langle X \rangle$	the quadratic variation process of X
$A^\top$	transpose of the vector or matrix A.

Chapter 1

Introduction

1.1 Biomathematical Background and Problem Definition

To understand the human or animal brain one is interested in the functionality of neurons in a neuronal network. The neurons communicate by sending out electric pulses caused by spikes. This is also called “firing”. A single neuron receives lots of electric pulses from other neurons in the network. This information is added up in the so-called *membrane potential*. Biologists are able to measure the membrane potential inside a single neuron. An example of such an observation is shown in figure 1.1. One observes that the

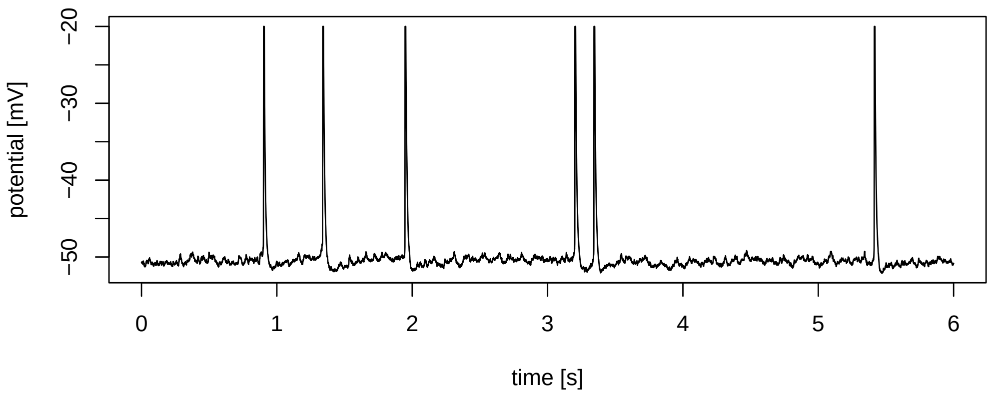

Figure 1.1: This membrane potential was recorded in vitro (by W. Kilb, Institute of Physiology, Uni-Mainz) from a neuron belonging to a cortical slice preparation from a 6 week old mouse. The network in this cortical slice was stimulated by a potassium solution (level: 5mM K^+). The top of each spike typically reaches positive voltage (ca. +20mV). For illustration the potential has been truncated at −20mV.

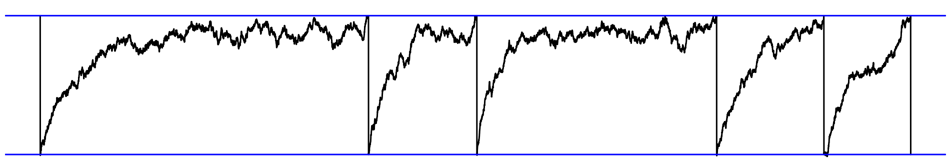

Figure 1.2: Simulated data mimicking the behavior of membrane potentials. The vertical lines represent the spike times. x_0 and S are fixed.

membrane potential of a neuron spikes from time to time. Whenever this happens the neuron itself sends out information to other neurons. Biologists are very interested in modeling this spiking behavior in order to understand the neuronal activity. The time intervals between spikes are called inter spike intervals (ISIs). The trajectory of the membrane potential between spikes looks like a stochastic process. On the other hand, the shape of a spike looks very deterministic and takes only a very short time compared to an ISI. For this reason, an easy way to model the behavior of a membrane potential is to assume that between spikes the potential is driven by a diffusion process $X = (X_t)_{t\geq 0}$ a solution of the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \beta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \tag{1.1.1}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion, $\beta(\cdot)$ the drift coefficient and $\sigma(\cdot)$ the diffusion coefficient. The spiking behavior is usually explained as follows. Whenever a certain *excitation threshold* S is reached, a spike of deterministic shape occurs. Thereafter the potential is set down to a certain *reset value* x_0 again. If time duration of a spike is negligible compared to the inter spike time, each spike can be modeled simply by a single point on the time axis. Otherwise the spike is modeled by some deterministic curve. Figure 1.2 illustrates such a model for a membrane potential.

The most commonly used neuronal models in the literature, based on the idea described above, are those where the process $X = (X_t)_{t\geq 0}$ between spikes is assumed to be one of the following stochastic processes. Let $\sigma, b > 0$, $a \geq 0$ and $(B_t)_{t\geq 0}$ be a standard Brownian motion.

(BMD): The *Brownian motion with drift* starting in $X_0 = x_0$ is defined by

$$X_t := x_0 + at + \sigma B_t, \quad t \geq 0,$$

(GBM): The *geometric Brownian motion* is given by the strong solution of the SDE

$$\mathrm{d}X_t = bX_t\mathrm{d}t + \sigma X_t\mathrm{d}B_t, \quad X_0 = x_0.$$

(OU): The *Ornstein-Uhlenbeck process* is given by the strong solution of the SDE

$$\mathrm{d}X_t = (a - bX_t)\mathrm{d}t + \sigma \mathrm{d}B_t, \quad X_0 = x_0,$$

where $a \in \mathbb{R}$. The Ornstein-Uhlenbeck neuronal model is also often called the *diffusion leaky integrate-and-fire neuronal model*, known to be the stochastic version of the deterministic *leaky integrate-and-fire neuronal model*.

(CIR): The *Cox-Ingersoll-Ross process* is given by the strong solution of the SDE

$$\mathrm{d}X_t = (a - b(X_t - c))\mathrm{d}t + \sigma\sqrt{(X_t - c)^+}\mathrm{d}B_t, \quad X_0 = x_0 \geq c,$$

where $c \leq 0$ is a constant and $(X - c)^+ := \max\{0, X - c\}$ denotes the positive part of $X - c$. The Cox-Ingersoll-Ross process is also called Feller process or square-root process. In neuronal framework this model is also called *Feller neuronal model*.

In the literature, the OU neuronal model is the most important of the models mentioned above. It has been extensively studied, see e.g. [LL 87], [RS 88], [LST 95], [LS 01], [DL 05], [LSH 06], [DL 07] and the literature quoted there. The OU model maps very well the property of exponential decay of a membrane potential. However, synaptic input from a large number of stochastically active sources should imply that incremental variances of the membrane potential are proportional to the present state. Certainly, this is realized by the CIR process. Concerning neuronal models, the CIR process was investigated in [LL 87], [GLNR 88], [LST 95], [DL 06], [HB 06] and [DL 07]. For both OU and CIR model, one often distinguishes between different regimes of parameter configuration. The asymptotic mean of the OU and CIR process (with $c = 0$) is a/b. So wherever the process starts it will run exponentially fast to the vicinity of this value and oscillate around there. Hence we have a more regular spiking behavior if $a/b > S$, which is called the *suprathreshold* regime. The *subthreshold* regime is the case if $a/b < S$ where spiking is a "rare" event. This is because the deterministic drift in the SDE becomes negative for $X > a/b$ and spikes can only be initiated by random fluctuations. If $a/b = S$ or even $a/b \approx S$ this is called *threshold* regime.

In the literature more complex models do exist such as the Fitzhugh-Nagumo or the Hodgkin-Huxley model. They are described in the book of Tuckwell [T 89] for example. These models might be more realistic than the models considered in this work, but they involve a large number of parameters. Hence statistical identifiability problems arise such that a parameter estimation from observation of the membrane potential is impossible.

However, this is not the only reason why the simple models described above are relevant for applications. In fact, between spikes of real data it is sometimes possible to observe a diffusion process by estimating the coefficients $\beta(\cdot)$ and $\sigma(\cdot)$ in (1.1.1). This was done by Höpfner [H 07], who proposed a nonparametric method based on estimators investigated by Florens-Zmirou [FZ 93]. His estimation method is independent of the values x_0 and S. However, we have to determine x_0 and S in order to fix the model.

Important papers about parameter estimation from discrete observed membrane potentials and from ISI data in the framework mentioned above are [DL 05], [DL 06], [DL 07] by Ditlevsen and Lansky and [LSH 06] by Lansky, Sanda and He. In [DL 05] Ditlevsen and Lansky considered the OU model and proposed different parametric estimators for the parameters a and σ according to the regime. These are based on the observation of iid level crossing times (ISIs)

$$\tau_i := \inf\{t \geq 0 \mid X_t = S\}, \quad i = 1, \ldots, n.$$

For the subthreshold and threshold regime they applied maximum likelihood estimation. Concerning the suprathreshold regime they used the Laplace transform of τ to define moment estimators. For the CIR model treated in [DL 06] they obtained corresponding results. In [DL 07] they considered again OU and CIR model and proposed a different estimation method based on the so called *Fortet's integral equation*. This method can be regarded as a kind of minimum distance estimation (MDE) investigated by Millar [M 84], which will be described in section 2.2. In [LSH 06] Lansky, Sanda and He also proposed estimators for the parameters a, b and σ but based on discrete observations of ISI data in the OU case. Beside this neuronal framework, we should mention the paper of Overbeck an Ryden [OR 97] about parameter estimation for the CIR process from equidistant observations and the work of Hammer [Ha 05] who reinvestigated this problem in a more general setting.

For all the results in the papers mentioned above, reset value x_0 and excitation threshold S are assumed to be known and fixed. Of course from a naive point of view one could say that S and x_0 are just beginning and end of a spike. But in real data this observation is not obvious at all. Even if we record the whole trajectory of a membrane potential, it is not evident where a spike starts and ends. In fact, right before and after the spike fluctuations are visible, disappearing more and more while the upturn of the spike. Unfortunately, one cannot catch where exactly this upturn was initiated. How is one supposed to decide whether the observed data behaves basically deterministic or like a stochastic process? One only could have an idea of a reasonable threshold for a single spike, but this wont be suitable for

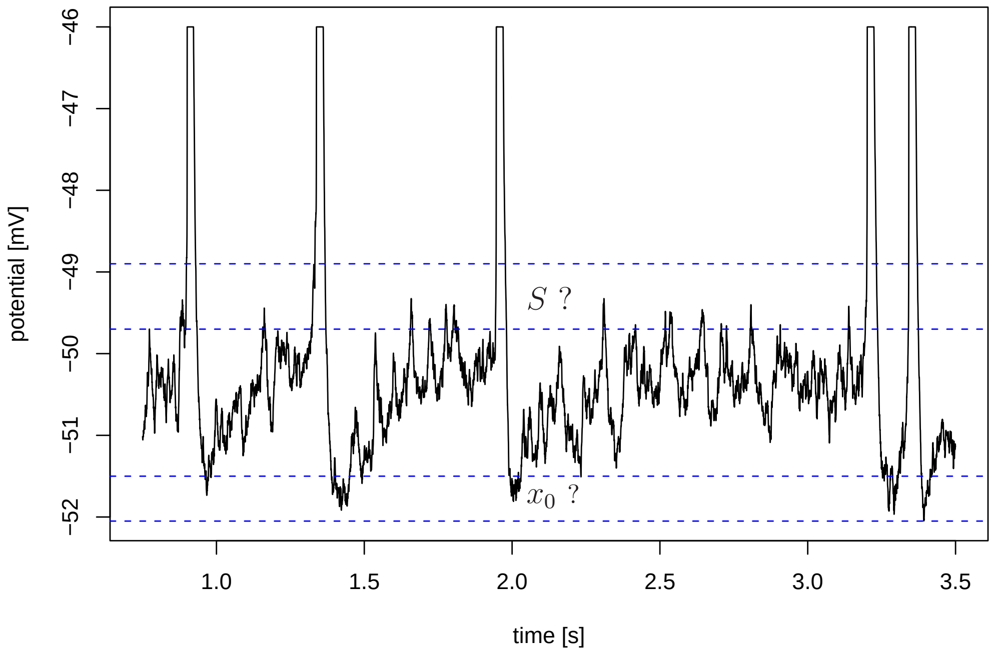

Figure 1.3: A zoom into figure 1.1 yields this picture. The upper dashed line might be a possible level for S if we only look on the 2nd spike that has its last visible fluctuation right under this level. Between the 1st and the 2nd dashed line, we have a band of possible levels for S if we consider the other spikes. Analogously, the lower dashed lines mark the band of possible values for x_0 corresponding to each spikes. Note that fixing any concrete level for S or x_0 from the picture is only conjectural.

most of the other spikes in the membrane potential. Concerning the reset value, the same problems arise if we examine the downturn of the spike. Consequently, it is not possible to observe the positions of S and x_0 by visual inspection of the trajectory. This fact is visualized in figure 1.3.

Certainly, the model described above is not reality and one could have doubts concerning the intrinsic existence of a reset value and a excitation threshold. Anyway, this model is popular due to its simplicity allowing for statistical inference. Further, the investigations of Höpfner in [H 07] support the idea that the membrane potential between spikes is a diffusion process, because he was able to estimate drift and diffusion coefficients. Since x_0 and S are important parameters of this model in order to mimic spiking behavior, one has to search for reasonable values of x_0 and S in order to perform a model calibration. This leads to a statistical problem.

In the paper [LSH 06] Lansky, Sanda and He seemed to be the first who drew attention to this issue and tried to determine x_0 and S. Their approach, described in [LSH 06] p. 215, was ad hoc and not a parameter estimation from a statistical model. To summarize their results, they searched for the minimum of the potential in a certain interval right after the spike and defined the median value of all these minima to be x_0. The excitation threshold S was defined as the median value of all levels of last decrease right before a spike in the trajectory of the membrane potential. In fact, this approach did not care about any model assumptions and is a kind of nonparametric. Further this method, especially for the value of S, is influenced very much by accuracy of voltage measurement and time frequency of the data. Measuring with higher time frequency might visualize more fluctuations and the method of Lansky, Sanda and He would find a higher value for S.

Unlike Lansky, Sanda and He, in the present work we view reset value x_0 and excitation threshold S as parameters of a statistical experiment. Assuming to know the law of the diffusion process X between spikes, we develop different parameter estimation methods in order to estimate x_0 and S from given sets of iid inter spike times representing level crossing times of X starting in x_0 and hitting S (cf. figure 1.4). The observation of inter spike times (ISIs) from real data is very easy and robust. In application we will simply use the times between the maxima of consecutive spikes taking advantage of the fact that we can observe these times very accurately and we do not have to care about the problem of beginning and ending of a spike. Further, these times can only be slightly different from the "true" inter spike times because of the short time extension of the spike trajectory compared to the ISI.

Genon-Catalot and Laredo investigated in [GL 87] an [GL 90] the problem of parameter estimation by observing level crossing times in a general framework for diffusion processes. However, they considered the estimation problem of a parameter θ in the drift function $\beta(\cdot) = \beta_\theta(\cdot)$, cf. (1.1.1). Since x_0 and S are not parameters of the diffusion process, the results of Genon-Catalot and Laredo are not suitable for our purposes.

For the mathematical and statistical methods used in this work, we refer to the books of Kutoyants [Ky 94], [Ky 04] and Prakasa Rao [PR 99]. Further, we refer to the books of Liptser & Shiryaev [LiS 01] and Jacod & Shiryaev [JS 03]. The reader finds all theoretical background there.

1.2 The Statistical Experiment

In the following, we will describe the general assumptions and the statistical model in order to estimate the parameters $\theta_1 := x_0$ and $\theta_2 := S$ by observing the inter spike times. An illustration is given in figure 1.4. Assume the membrane potential $X = (X_t)_{t\geq 0}$ between spikes follows an SDE

$$\mathrm{d}X_t = \beta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t \tag{1.2.1}$$

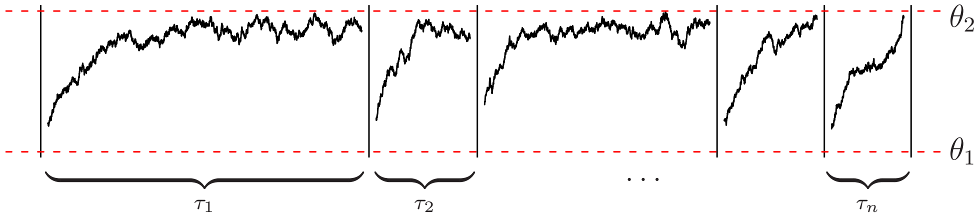

Figure 1.4: We assume to know the law of the process between successive spikes, but not the trajectory itself. Observing ISI data representing the iid level crossing times τ_i, $i = 1, \ldots, n$ from some unknown value $\theta_1 = x_0$ to some unknown threshold $\theta_2 = S$ ($\theta_1 < \theta_2$), the aim is to find probable values for θ_1 and θ_2.

where $(B_t)_{t\geq 0}$ is a standard Brownian motion and let the coefficients $\beta(\cdot)$ and $\sigma(\cdot)$ be known. Let $X^{(\theta_1)}$ denote X starting in $X_0 = x_0 = \theta_1$. In this experiment we do not claim to observe the trajectory of X itself. We only assume to observe $n \in \mathbb{N}$ inter spike times τ_i, $i = 1, \ldots, n$ which are iid copies of τ the level crossing time to the unknown level $\theta_2 = S$, defined by

$$\tau := \inf \left\{ t \geq 0 \mid X_t^{(\theta_1)} = \theta_2 \right\} \tag{1.2.2}$$

(cf. fig 1.4). For biological reasons we claim $x_0 < S$ or $\theta_1 < \theta_2$ respectively. To specify the statistical experiment, our parameter space is defined by

$$\Theta := \{(\theta_1, \theta_2) \in \mathbb{R}^2 | \, \theta_1 < \theta_2\}.$$

We will consider the four different cases defined above in which X is given by a Brownian motion with drift (BMD), a geometric Brownian motion (GBM), an Ornstein-Uhlenbeck (OU) or a Cox-Ingersoll-Ross (CIR) process. For the cases Brownian motion with drift and geometric Brownian motion, it will turn out that the joint estimation of reset value and excitation threshold is impossible. One has to know one of them to estimate the other. However, this is not an essential restriction. If we fix a value y between reset value x_0 and excitation threshold S then it is possible to observe the level crossing times between x_0 and y or y and S respectively from the trajectory of the membrane potential. This is visualized in figure 1.5. Since y is known, we can apply our methods to estimate x_0 and S respectively in separated one-parametric experiments. If we estimate θ_1, the parameter space is given by

$$\Theta_1 := (-\infty, y),$$

where y is any known upper level. If we estimate θ_2, the parameter space is given by

$$\Theta_2 := (y, \infty),$$

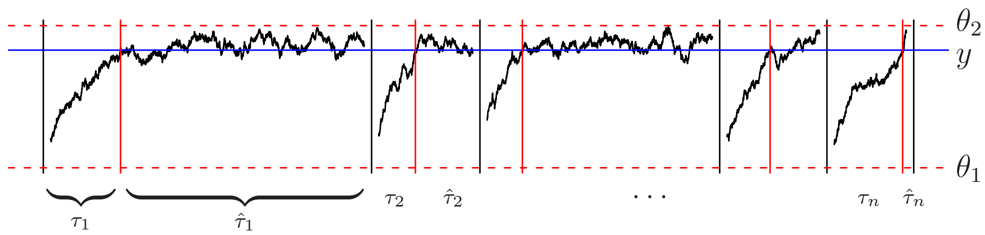

Figure 1.5: In this picture y was chosen to be the median of the membrane potential. For a first experiment the level crossing times τ_i, $i = 1, \ldots, n$ from $\theta_1 = x_0$ to y will help to find θ_1 and for a second experiment the level crossing times $\hat{\tau}_i$, $i = 1, \ldots, n$ from y to $\theta_2 = S$ will help to find θ_2. Since y is a known threshold, both experiments are only one-parametric.

where y is any known lower level. In applications a natural candidate for the choice of y would be the median value of the membrane potential (cf. figure 1.5) because in contrast to the mean value, it is very robust concerning spikes.

1.3 Outline

In the cases Brownian motion with drift and geometric Brownian motion, the Lebesgue-density of the level crossing time τ is well-known and we can apply the *maximum likelihood* approach. Using LAN (local asymptotic normality) theory, we will show that these estimators are "optimal". In contrast to that, the Lebesgue-density of the level crossing time τ is not given explicitly in the cases of OU and CIR process. We only know the Laplace transform of the distribution of τ. Hence we suggest the method of *minimum distance estimation* (MDE) instead of a maximum likelihood approach. For this purpose, the empirical Laplace transform is compared to a family of possible true Laplace transforms parametrized by $\theta \in \Theta$. The *minimum distance estimator* (MDE) takes a certain value θ^* such that the corresponding Laplace transform fits the empirical Laplace transform best in the sense of a Hilbert space norm.

In the second chapter of this work, a short introduction to LAN theory and a concept of "optimal" estimators is given. Further, the minimum distance estimation theory by Millar [M 84] is introduced. The following chapters are based on these theoretical preliminaries.

The third chapter discusses maximum likelihood estimation in the cases of Brownian motion with drift (BMD) and geometric Brownian motion (GBM), where the density of the level crossing time is given explicitly. In the beginning, a generalized experiment is considered which covers the cases just mentioned. Accordingly, the main results giving strong consistency, asymptotic normality and "optimality" are just corollaries.

The fourth chapter considers the cases Ornstein-Uhlenbeck (OU) and Cox-Ingersoll-Ross (CIR) process. The MDE method works very similarly in these two cases. The corresponding main results will show strong consistency and asymptotic normality of the MDE.

The last chapter contains first a description of the strategy how to apply our methods to real or simulated data. Accordingly, we will apply the MDE to simulated data in order to check the performance of our methods. It will turn out that in the subthreshold case it is possible to estimate S very accurately with a small sample size (< 50), in contrast to x_0 where we need a huge sample size which is unrealistic for applications. This is due to the fact that in the subthreshold case the Laplace transform of the level crossing time changes very little if vary x_0. However, this also means that x_0 is a parameter of secondary importance for the model in the subthreshold regime because it is not necessary to determine exactly the value of x_0 in order to reproduce the typical spiking behavior. In this case, the crucial parameter is the excitation threshold S.

The following sections deal with the concrete applications to real data sets. The first dataset of our consideration was recorded by Werner Kilb from the Institute of Physiology and Pathophysiology, University of Mainz. The second dataset was provided by Petr Lansky, Pavel Sanda and Jufang He who investigated this data in [LSH 06]. Both data sets are non regular spiking or "subthreshold" respectively. This is important in order to have long inter spike intervals. Only in this situation we are able to apply first of all the nonparametric methods proposed by Höpfner [H 07] in order to estimate $\beta(\cdot)$ and $\sigma(\cdot)$ in (1.1.1) which determine the case (OU, CIR,...) of our statistical experiment described above. A comprehensive description of our data analysis and a detailed discussion of the estimation results is given.

It will turn out that between spikes the data set provided by Kilb behaves like a CIR process and the data set provided by Lansky, Sanda and He seem to behave like an OU processes. Based on these insights the MDE finds reasonable values for S which should be reliable. Due to the sample size the results for x_0 are rather not reliable but of secondary importance for the model anyway. Further, for all our MDE estimates the Laplace transform corresponding to the estimated value θ^* fits the empirical Laplace transform very well. For this reason, the model in the setting $x_0 = \theta_1^*$ and $S = \theta_2^*$ is able to reproduce very well the characteristics of spiking behavior as it is observed in the real data. Since mimicking of spiking behavior by using a manageable model is a main goal of biologists, this is an important conclusion.

Chapter 2

Theoretical Preliminaries

2.1 A Recap of LAN

This section gives an overview about the LAN (local asymptotic normality) theory and optimality of estimators. It is based on the script of Höpfner [H 0708] Kap. V + VII. We also refer to the book of Le Cam and Yang [LY 00] Chapters 6 and 7. Two important papers developing this theory are [LC 69] by Le Cam and [D 85] by Davis.

The LAN property of a statistical experiment allows us to give asymptotic lower bounds for the estimation errors. This will be clear at the end of this section were we will formulate the LAN minimax theorem. If an estimator attains this lower bound it is called "optimal" or LAM (locally asymptotically minimax). In order to define the LAN property we first have to define the nicest experiment a statistician can imagine, the *Gaussian Shift Experiment*.

Definition 2.1.1 *A statistical experiment* $\mathscr{E} := \left(\Omega, \mathscr{F}, \left\{\mathbf{P}_h : h \in \mathbb{R}^d\right\}\right)$ *is called* Gaussian Shift Experiment *if there exists a positive definite symmetric matrix* $J \in \mathbb{R}^{d\times d}$ *and a* $\mathscr{F} - \mathscr{B}(\mathbb{R}^d)$*-measurable random variable* S *such that for* $h \in \mathbb{R}^d$ *the likelihood ratio is given by*

$$\frac{\mathrm{d}\mathbf{P}_h}{\mathrm{d}\mathbf{P}_0} = \exp\left(h^\top S - \frac{1}{2}h^\top J h\right).$$

Remark:

- $Z := J^{-1}S$ is the MLE of $\mathscr{E}$.
- $\mathcal{L}(Z - h|\mathbf{P}_h) = \mathcal{N}\left(0, J^{-1}\right)$ for all $h \in \mathbb{R}^d$. This means that Z is an *equivariant* estimator, working with the same performance at all points of the parameter space.

- For every positive definite symmetric matrix $J \in \mathbb{R}^{d\times d}$ there exists a Gaussian Shift Experiment.

Definition 2.1.2 (LAN) *Let $\Theta \subset \mathbb{R}^d$ be an open subset. For fixed $\theta \in \Theta$ a sequence of statistical experiments*

$$\mathscr{E}_n := \left(\Omega_n, \mathscr{F}_n, \{\mathbf{P}_{n,\theta} : \theta \in \Theta\}\right), \quad n \in \mathbb{N},$$

is called locally asymptotically normal *(LAN) at θ if there exists a real sequence $(\delta_n)_{n\in\mathbb{N}} = (\delta_n(\theta))_{n\in\mathbb{N}}$, $\delta_n \to 0$, a positive definite symmetric matrix $J = J_\theta \in \mathbb{R}^{d\times d}$ and a $\mathscr{F}_n - \mathscr{B}(\mathbb{R}^d)$-measurable random variables $S_n = S_n(\theta)$, $n \in \mathbb{N}$ such that the following conditions hold:*

i) For every bounded sequence $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$

$$\log \frac{\mathrm{d}\mathbf{P}_{n,\theta+\delta_n h_n}}{\mathrm{d}\mathbf{P}_{n,\theta}} = h_n^\top S_n - \frac{1}{2} h_n^\top J h_n + o_{(\mathbf{P}_{n,\theta})}(1), \quad n \to \infty$$

ii)

$$\mathcal{L}(S_n|\mathbf{P}_{n,\theta}) \xrightarrow[n\to\infty]{} \mathcal{N}(0, J), \text{ weak convergence in } \mathbb{R}^d.$$

Remark: Since we can construct a Gaussian shift experiment

$$\mathscr{E}_\infty = \mathscr{E}_\infty(S(\theta), J_\theta) := \left(\Omega, \mathscr{F}, \left\{\mathbf{P}_h : h \in \mathbb{R}^d\right\}\right)$$

for a given positive definite symmetric matrix $J_\theta \in \mathbb{R}^{d\times d}$, the LAN property implies

$$\mathcal{L}(S_n|\mathbf{P}_{n,\theta}) \xrightarrow[n\to\infty]{} \mathcal{L}(S|\mathbf{P}_0)$$

and also the convergence of the likelihood ratios to the likelihood ratio of the Gaussian shift experiment if $h_n \to h$. Hence the LAN property means just that a sequence of experiments, locally scaled by δ_n at a fixed point $\theta \in \Theta$, is "approximately" a *Gaussian Shift* experiment with local parameter h.

Definition 2.1.3 (Differentiability in Quadratic Mean) *Let $\Theta \subset \mathbb{R}^d$ be an open subset. A family of probability measures $\{\mathbf{P}_\theta : \theta \in \Theta\}$ on $(\Omega, \mathscr{F})$ is called* differentiable in quadratic mean *at θ with derivative (Score) $V_\theta \in \mathbb{R}^d$ if the following conditions hold:*

i) The components of V_θ are $V_{\theta,1}, \ldots, V_{\theta,d} \in L^2(\Omega, \mathscr{F}, \mathbf{P}_\theta)$.

ii)

$$\frac{1}{|\xi - \theta|^2} \mathbf{P}_\xi \left[\frac{\mathrm{d}\mathbf{P}_\xi}{\mathrm{d}\mathbf{P}_\theta} = \infty \right] \xrightarrow{\xi\to\theta} 0$$

iii)

$$\frac{1}{|\xi-\theta|^2}\int_\Omega \left| \sqrt{\frac{\mathrm{d}\mathbf{P}_\xi}{\mathrm{d}\mathbf{P}_\theta}} - 1 - \frac{1}{2}(\xi-\theta)^\top V_\theta \right|^2 \mathrm{d}\mathbf{P}_\theta \xrightarrow{\xi\to\theta} 0$$

Remark:

- Differentiability in quadratic mean implies that the score is a centered random variable, i.e. $\mathbf{E}_\theta[V_\theta] = 0$.
- Note that LAN and Differentiability in Quadratic Mean are local properties at $\theta \in \Theta$. In the next step, Le Cam's 2nd Lemma shows that differentiability in quadratic mean of a single experiment implies the LAN property with local scale $\delta_n = n^{-1/2}$ for the corresponding product experiment.

Lemma 2.1.4 (Le Cam's 2nd Lemma) *Let*

$$\mathscr{E} := (\Omega, \mathscr{F}, \{\mathbf{P}_\theta : \theta \in \Theta\})$$

be a single experiment and let $\{\mathbf{P}_\theta : \theta \in \Theta\}$ be differentiable in quadratic mean at θ with derivative V_θ. Then the sequence of the n-fold product experiments of $\mathscr{E}$

$$\mathscr{E}_n := \left(\Omega_n := \bigtimes_{j=1}^n \Omega, \mathscr{F}_n := \bigotimes_{j=1}^n \mathscr{F}, \left\{\mathbf{P}_{n,\theta} := \bigotimes_{j=1}^n \mathbf{P}_\theta : \theta \in \Theta\right\}\right), \quad n \in \mathbb{N}$$

is LAN at θ. More precisely, let $(V_\theta^i)_{i\in\mathbb{N}}$ iid copies of V_θ, $J := \mathbf{E}_\theta[V_\theta V_\theta^\top]$ then

$$S_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n V_\theta^i \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, J)$$

and for any real, bounded sequence $(h_n)_{n\in\mathbb{N}}$

$$\log\left\{\frac{\mathrm{d}\mathbf{P}_{n,\theta+h_n/\sqrt{n}}}{\mathrm{d}\mathbf{P}_{n,\theta}}\right\} - \left\{h_n S_n - \frac{1}{2}h_n^\top J h_n\right\} \xrightarrow[n\to\infty]{\mathbf{P}_{n,\theta}} 0.$$

Proof: cf. [H 0708] Kap.IV 4.10, 4.11 or [LY 00] ch. 7.2 prop. 1, pp.180-181. □

For the rest of this section we assume that the sequence of statistical experiments

$$\mathscr{E}_n := \left(\Omega_n, \mathscr{F}_n, \{\mathbf{P}_{n,\theta} : \theta \in \Theta\}\right), \quad n \in \mathbb{N}, \quad \Theta \subset \mathbb{R}^d$$

satisfies the LAN condition at a fixed point $\theta \in \Theta$.

Definition 2.1.5 *In the case of LAN at θ, a sequence of estimators κ_n for the unknown parameter in $\mathscr{E}_n$, $n \in \mathbb{N}$ is called* regular *at θ if there exists a probability measure μ on $\mathbb{R}^d$ such that for every $h \in \mathbb{R}^d$*

$$\mathcal{L}\left(\delta_n^{-1}(\kappa_n - (\theta + \delta_n h))|\mathbf{P}_{n,\theta+\delta_n h}\right) \xrightarrow[n\to\infty]{w} \mu,$$

where μ is independent of the local parameter h. This definition is a local asymptotic analogon to the definition of equivariance.

Theorem 2.1.6 (Hájek, 1970) *Assume LAN holds at θ. For every sequence of estimators $(\kappa_n)_{n\in\mathbb{N}}$ regular at θ, where*

$$\mathcal{L}\left(\delta_n^{-1}(\kappa_n - (\theta + \delta_n h))|\mathbf{P}_{n,\theta+\delta_n h}\right) \xrightarrow[n\to\infty]{w} \mu,$$

μ can be described as

$$\mu = \mathcal{N}\left(0, J_\theta^{-1}\right) * Q,$$

where Q is a suitable probability measure on $\mathbb{R}^d$.

Proof: cf. [H 0708] Kap.VII Thm. 7.8 or [LY 00] ch. 6.6 Thm. 3, pp.157-159. □

Definition 2.1.7 *If the measure Q in theorem 2.1.6 is equal to the Dirac measure at* 0 *we call $(\kappa_n)_{n\in\mathbb{N}}$* efficient *at θ.*

Remark: The previous definition is very natural because in this case the limit distribution in Thm. 2.1.6 is best concentrated. Note that Thm. 2.1.6 only considers regular estimators. Non-regular estimators might have a more concentrated asymptotic distribution. Thm. 2.1.9 will give an answer to this question. But first we will give a characterization of regularity and efficiency.

Lemma 2.1.8 *In the case of LAN at θ, for every sequence of estimators $(\kappa_n)_{n\in\mathbb{N}}$ for the unknown parameter in $(\mathscr{E}_n)_{n\in\mathbb{N}}$ the following statements are equivalent:*

- *$(\kappa_n)_{n\in\mathbb{N}}$ is regular and efficient.*
- $\delta_n^{-1}(\kappa_n - \theta) - Z_n(\theta) \xrightarrow[n\to\infty]{\mathbf{P}_{n,\theta}} 0$, *where* $Z_n(\theta) := J_\theta^{-1} S_n(\theta)$

Proof: cf. [H 0708] Kap.VII Satz 7.12. □

Theorem 2.1.9 (MiniMax) *Assume LAN holds at* θ. *Define* $Z_n(\theta) := J_\theta^{-1} S_n(\theta)$ *and* $Z(\theta) := J_\theta^{-1} S(\theta)$. *Let* $\ell : \mathbb{R}^d \to [0, \infty)$ *be a bounded, continuous, "bowl-shaped" loss function, i.e. the sublevel sets* $\{x \in \mathbb{R}^d : \ell(x) \leq \alpha\}$, $\alpha \geq 0$ *are convex and symmetric. Further, let* $(\kappa_n)_{n\in\mathbb{N}}$ *be a sequence of estimators for the unknown parameter in* $(\mathscr{E}_n)_{n\in\mathbb{N}}$.

a) If $\mathcal{L}\left(\delta_n^{-1}(\kappa_n - \theta)|\mathbf{P}_{n,\theta}\right)$, $n \in \mathbb{N}$ *is tight, then*

$$\sup_{c>0} \liminf_{n\to\infty} \sup_{|h|\leq c} \mathbf{E}_{\theta+\delta_n h}\left[\ell\left(\delta_n^{-1}(\kappa_n - (\theta + \delta_n h))\right)\right] \geq \mathbf{E}_0[\ell(Z)]$$

holds.

b) If $\delta_n^{-1}(\kappa_n - \theta) - Z_n(\theta) \xrightarrow[n\to\infty]{\mathbf{P}_{n,\theta}} 0$, $(\kappa_n)_{n\in\mathbb{N}}$ *attains the lower bound, i.e. for any* $c > 0$

$$\lim_{n\to\infty} \sup_{|h|\leq c} \mathbf{E}_{\theta+\delta_n h}\left[\ell\left(\delta_n^{-1}(\kappa_n - (\theta + \delta_n h))\right)\right] = \mathbf{E}_0[\ell(Z)]$$

holds. In this case $(\kappa_n)_{n\in\mathbb{N}}$ *is called* locally asymptotically minimax *(LAM) at* θ.

Proof: cf. [H 0708] Kap.VII Satz 7.13. or [LY 00] ch. 6.6 Lem. 5 + Thm. 1, pp.153-155. □

Remark: If the LAN property holds we know that $\mathcal{L}(Z|\mathbf{P}_0) = \mathcal{N}(0, J^{-1})$. This gives a connection between Thm. 2.1.6 and Thm. 2.1.9. In addition, assertion a) of Thm. 2.1.9 gives an asymptotic lower bound for the estimation errors *for all* estimators. This is a type of Cramér-Rao bound, where in a local scale the parameters are taken uniformly into account. This is important in order to punish the estimator $\kappa_n \equiv \theta$ that just guesses θ. Together with assertion b) this shows that estimators attain this lower bound if their estimation errors are stochastically equivalent to the centering sequence $Z_n(\theta)$, $n \in \mathbb{N}$. These LAM estimators are "optimal" in this sense.

Note that θ also in Thm. 2.1.9 is fixed. Hence an optimal estimator for the unknown parameter of an experiment where LAN holds for all $\theta \in \Theta$ has to be LAM for all $\theta \in \Theta$.

2.2 Minimum Distance Estimation

In this section, we will describe the minimum distance estimation method. It is a very general method of parameter estimation and was basically and essentially investigated by Millar [M 84]. A big advantage of this method is that quite mild conditions lead to strong consistency and asymptotic normality. In most cases the estimators are not given explicitly in a closed analytic form and for applications we need a numerical evaluation. The following introduction is based again on the script of Höpfner [H 0708] Kap. II. The results of Millar [M 84] are given in a slightly more general context. So we also refer to the book of Kutoyants [Ky 94] ch. 7.2 pp. 221-222 who gives a short summary of these results.

Let the parameter space $\Theta \subset \mathbb{R}^d$ be an open subset and consider the sequence of experiments

$$\mathscr{E}_n := \big(\Omega, \mathscr{F}_n, \{\mathbf{P}_{n,\theta} : \theta \in \Theta\}\big), \quad n \in \mathbb{N}.$$

Further, let $\mathcal{H}$ be a Hilbert space with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. Define $\hat{\Psi}_n$, $n \in \mathbb{N}$ a sequence of $\mathcal{H}$-valued $\mathscr{F}_n$-measurable random variables

$$\hat{\Psi}_n : (\Omega, \mathscr{F}_n) \longrightarrow (\mathcal{H}, \mathscr{B}(\mathcal{H})), \quad n \in \mathbb{N}.$$

$\hat{\Psi}_n$ is a so-called *empirical value* observable after n steps in $\mathscr{E}_n$. Furthermore, let $\{\Psi_\theta : \theta \in \Theta\} \subset \mathcal{H}$ be a continuously parametrized subset of $\mathcal{H}$ with respect to θ, i.e.

$$\Theta \ni \theta \mapsto \Psi_\theta \in \mathcal{H} \quad \text{is continuous.}$$

Every Ψ_θ corresponds to $\mathbf{P}_{n,\theta}$ and is a so-called *reference value* for $\hat{\Psi}_n$. If we compare Ψ_θ and $\hat{\Psi}_n$ in $\mathcal{H}$ we can define a minimum distance estimator (MDE) θ_n^* by

$$\theta_n^* := \arg\inf_{\xi \in \Theta} \|\hat{\Psi}_n - \Psi_\xi\|_{\mathcal{H}}, \quad n \in \mathbb{N}, \tag{2.2.1}$$

provided this arg inf exists. If the arg inf does not exist, we simply set $\theta_n^* = \theta_0$ where $\theta_0 \in \Theta$ is arbitrary and may depend on n. An MDE takes the value of a parameter such that the reference value fits the empirical value best in $\mathcal{H}$. Note that this definition is not necessarily unique. Certainly, we need conditions for $\hat{\Psi}_n$ and Ψ_θ such that the "arg inf" exists asymptotically. These will lead to consistency and further conditions will ensure asymptotic normality of θ_n^*, $n \in \mathbb{N}$.

Conditions for MDE:

- The strong law of large numbers **SLLN(θ)** is fulfilled at $\theta \in \Theta$ if
$$\|\Psi_\theta - \hat{\Psi}_n\|_{\mathcal{H}} \xrightarrow[n\to\infty]{\mathbf{P}_\theta - a.s.} 0. \tag{2.2.2}$$

- The identifiability condition **I(θ)** is fulfilled at $\theta \in \Theta$ if
$$\inf_{\xi\in\Theta, |\theta-\xi|\geq\delta} \|\Psi_\theta - \Psi_\xi\|_{\mathcal{H}} > 0, \quad \forall \delta > 0.$$

- The differentiability condition **D(θ)** is fulfilled at $\theta \in \Theta$ if
$$\Theta \ni \xi \mapsto \Psi_\xi \in \mathcal{H}$$
is Fréchet-differentiable at θ with derivative $D\Psi_\theta = (D_1\Psi_\theta, \dots, D_d\Psi_\theta)$ whose components are linearly independent in $\mathcal{H}$.

- The tightness condition **T(θ)** is fulfilled at $\theta \in \Theta$ if there exists a real sequence $\varphi_n = \varphi_n(\theta) \uparrow \infty$ such that the family
$$\mathcal{L}\Big(\varphi_n \|\hat{\Psi}_n - \Psi_\theta\|_{\mathcal{H}} \,\big|\, \mathbf{P}_{\theta,n}\Big), \quad n \in \mathbb{N}$$
is tight in $\mathbb{R}$.

Theorem 2.2.1 *If the conditions **SLLN(θ)** and **I(θ)** are satisfied for every $\theta \in \Theta$, then every sequence of MDEs θ_n^*, $n \in \mathbb{N}$ for θ is strongly consistent.*

Proof: This was proved by Höpfner ([H 0708], Thm. 2.11). □

Theorem 2.2.2 (Millar, 1984) *If the conditions **SLLN(θ)**, **I(θ)**, **D(θ)** and **T(θ)** are satisfied for every $\theta \in \Theta$, then for every sequence of MDEs θ_n^*, $n \in \mathbb{N}$, defined in* (2.2.1)*, the rescaled estimation errors have the representation*
$$\varphi_n\,(\theta_n^* - \theta) = \Pi\left(\varphi_n(\hat{\Psi}_n - \Psi_\theta)\right) + o_{\mathbf{P}_{\theta,n}}(1), \quad n \to \infty$$
where $\Pi : \mathcal{H} \to \mathbb{R}$ is a linear map given by
$$\Pi(\cdot) := \Lambda_\theta^{-1} \begin{pmatrix} \langle D_1\Psi_\theta, \,\cdot\, \rangle_{\mathcal{H}} \\ \cdots \\ \langle D_d\Psi_\theta, \,\cdot\, \rangle_{\mathcal{H}} \end{pmatrix}, \quad \Lambda_\theta := \big(\langle D_i\Psi_\theta, D_j\Psi_\theta\rangle_{\mathcal{H}}\big)_{1\leq i,j\leq d}.$$

Proof: This was proved by Millar [M 84], even if condition (2.2.2) is weakened to stochastic convergence (WLLN). Thus weak consistency would follow easily from this result. We also refer to [H 0708] (Thm. 2.14) who gives a much more detailed proof. □

Remark: Evidently, $\Pi : \mathcal{H} \to \mathbb{R}$ is continuous. Provided we know the asymptotic distribution of $\varphi_n(\hat{\Psi}_n - \Psi_\theta)$, theorem 2.2.2 and the continuous mapping theorem give the asymptotic distribution of the rescaled estimation errors of θ_n^*, $n \in \mathbb{N}$. Since $\Pi : \mathcal{H} \to \mathbb{R}$ is linear, one only has to show that $\varphi_n(\hat{\Psi}_n - \Psi_\theta)$ is asymptotically Gaussian to get asymptotic normality for the MDE. Hence we will now introduce some more assumptions and give a further condition on the $\hat{\Psi}_n$ and Ψ_θ.

Let $T \subset \mathbb{R}^k$ be a k-dimensional interval with σ-algebra $\mathscr{B}(T)$ and μ be a finite measure absolutely continuous with respect to $\lambda\!\!\lambda_{|T}$, the Lebesgue measure restricted on T. In particular, let $\mathcal{H} = L^2(T, \mathscr{B}(T), \mu)$ and further let $\hat{\Psi}_n$ be a measurable real valued process

$$\hat{\Psi}_n : (\Omega \times T, \mathscr{F}_n \otimes \mathscr{B}(T)) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

with paths in $\mathcal{H} = L^2(T, \mathscr{B}(T), \mu)$.

A further Condition:

- The asymptotic normality condition $\mathbf{AN}(\theta)$ is fulfilled at $\theta \in \Theta$ if the processes $W^n_\bullet$, $n \in \mathbb{N}$ defined by $W^n_\alpha := \sqrt{n}\left(\hat{\Psi}_n(\alpha) - \Psi_\theta(\alpha)\right)$ converge weakly in $\mathcal{H} = L^2(T, \mathscr{B}(T), \mu)$ to a Gaussian process $W = W(\theta)$ with covariance function $K(\cdot,\cdot)$, i.e.

$$W^n \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} W \quad \text{in } \mathcal{H} = L^2(T, \mathscr{B}(T), \mu).$$

  Obviously, condition $\mathbf{AN}(\theta)$ implies condition $\mathbf{T}(\theta)$ with $\varphi_n := \sqrt{n}$.

The next theorem by Cremers and Kadelka [CK 86] gives a possibility to prove the weak convergence of stochastic processes whose paths belong to $L^p(T, \mathscr{B}(T), \mu)$. This will help us to check the previous condition $\mathbf{AN}(\theta)$.

Theorem 2.2.3 (Cremers / Kadelka, 1986) *Let $\xi_n(\cdot)$, $n \in \mathbb{N}_0$ be a sequence of stochastic processes with paths in $L^p(T, \mathscr{B}(T), \mu)$, $p \in [1, \infty)$. If the finite dimensional distributions of ξ_n converge weakly to those of ξ_0 and if further*

$$\limsup_{n\to\infty} \int |\xi_n|^p \mathrm{d}\mathbf{P} \otimes \mu \le \int |\xi_0|^p \mathrm{d}\mathbf{P} \otimes \mu < \infty$$

holds, then

$$\xi_n \xrightarrow[n\to\infty]{\mathcal{L}} \xi_0 \quad \text{in } L^p(T, \mathscr{B}(T), \mu).$$

Proof: This can be found in [CK 86] Thm. 2 or [H 0708] (Kap. II Thm. 2.4 + Satz 2.6).

In fact, Cremers and Kadelka in [CK 86] give two more possibilities to prove this weak convergence, but for our purposes these are not needed. They even show these results for $L^p(T, \mathscr{B}(T), \mu)$ replaced by $L_p^E(T, \mathscr{B}(T), \mu)$, where E is a Banach space and $(T, \mathscr{B}(T), \mu)$ an arbitrary $\sigma-$finite measure space. $L_p^E(T, \mathscr{B}(T), \mu)$ denotes the space of p-integrable, E-valued functions. In our context, E is just $\mathbb{R}$. $\square$

Under condition **AN(θ)**, the asymptotic normality of the MDE is an obvious corollary of theorem 2.2.2. We formulate all the results stated above in one main theorem:

Theorem 2.2.4 (Millar, 1984) *If the conditions **SLLN(θ)** and **I(θ)** are fulfilled for every $\theta \in \Theta$, then every sequence of MDEs θ_n^*, $n \in \mathbb{N}$ for θ, defined in* (2.2.1)*, is strongly consistent. If further **D(θ)** and **AN(θ)** hold for every $\theta \in \Theta$, the estimation errors are asymptotically normal:*

$$\sqrt{n}\,(\theta_n^* - \theta) \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \Pi(W(\theta)) \sim \mathcal{N}\left(0, \Lambda_\theta^{-1} V_\theta \Lambda_\theta^{-1}\right),$$

where Π is defined as in 2.2.2,

$$\Lambda_\theta := \left(\langle D_i\Psi_\theta, D_j\Psi_\theta\rangle_{\mathcal{H}}\right)_{1\le i,j\le d}$$

and $V \in \mathbb{R}^{d\times d}$ with

$$(V_\theta)_{i,j} := \int_T \int_T D_i\Psi_\theta(\alpha_1) K(\alpha_1, \alpha_2) D_j\Psi_\theta(\alpha_2) \mu(\mathrm{d}\alpha_1)\mu(\mathrm{d}\alpha_2).$$

Proof: [H 0708] Kap.II C Thm 2.23. $\square$

Chapter 3

Brownian Motion with Drift & Geometric BM

In this chapter, we consider our estimation problem described in chapter 1, assuming the process starting at θ_1 and hitting the level θ_2 is either a Brownian motion with drift (BMD) or a geometric Brownian motion (GBM). Our observations are iid level crossing times from θ_1 to θ_2. We will see that these two cases (BMD and GBM) lead to very similar statistical experiments. Hence the first section will discuss maximum likelihood estimation with respect to an experiment which is a generalization of these particular experiments. We will see that the maximum likelihood estimator (MLE) is strongly consistent and asymptotically normal. Further, we will prove the LAN property for the general experiment. This leads to an optimality result for the MLE which is shown to be LAM. Accordingly, the particular results for the cases BMD and GBM, considered in subsequent sections, will be simple corollaries to the results stated in the first section.

3.1 Maximum Likelihood Estimation

In the following, we will define a statistical experiment and compute the MLE. Let $\sigma > 0$, $a \geq 0$, $\Theta \subset \mathbb{R}$ be an open interval and $\phi : \Theta \to (0, \infty)$ be a positive C^2-diffeomorphism. We define the following single experiment

$$\mathscr{E} := ((0,\infty), \mathscr{B}((0,\infty)), \{\mathbf{P}_\theta : \theta \in \Theta\}), \tag{3.1.1}$$

where $\mathbf{P}_\theta$ has the Lebesgue-density

$$\frac{\mathrm{d}\mathbf{P}_\theta}{\mathrm{d}t} = f(t;\theta) := \frac{\phi(\theta)}{\sqrt{2\pi}\sigma} t^{-3/2} \exp\left\{-\frac{(\phi(\theta) - at)^2}{2\sigma^2 t}\right\}. \tag{3.1.2}$$

For every $\theta \in \Theta$, $f(\cdot;\theta)$ is the probability density of a level crossing time of a Brownian motion with drift, cf. [BS 02] II.2.2 p.295. The diffeomorphism ϕ will be specified in the next sections in order to derive the corresponding experiments for BMD and GBM. Now we consider the sequence of statistical experiments

$$\mathscr{E}_n := \big((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta} : \theta \in \Theta\}\big), \quad n \in \mathbb{N}, \tag{3.1.3}$$

where $\mathscr{E}_n$ is the n-fold product experiment of $\mathscr{E}$. The corresponding probability measure is given by

$$\mathbf{P}_{n,\theta} := \bigotimes_{j=1}^{n} \mathbf{P}_\theta,$$

with density

$$\begin{aligned} \frac{\mathrm{d}\mathbf{P}_{n,\theta}}{\mathrm{d}(t_1,\ldots,t_n)} &= f_n(t_1,\ldots,t_n;\theta) := \prod_{j=1}^{n} f(t_j,\theta) \\ &= \left(\frac{\phi(\theta)}{\sqrt{2\pi}\sigma}\right)^n \prod_{j=1}^{n} t_j^{-3/2} \exp\left\{-\sum_{i=1}^{n} \frac{(\phi(\theta)-at_j)^2}{2\sigma^2 t_j}\right\}. \end{aligned}$$

Our aim is to estimate the parameter $\theta \in \Theta$ from observations of iid random variables $\tau_1,\ldots,\tau_n$ where $\tau_1 \sim \mathbf{P}_\theta$ and $(\tau_1,\ldots,\tau_n) \sim \mathbf{P}_{n,\theta}$ respectively. Since the density is given explicitly, it is easy to compute the MLE for θ, which is shown by the following lemma.

Lemma 3.1.1 *The MLE for the parameter θ of the experiment $\mathscr{E}_n$ defined above in* (3.1.3) *is given by*

$$\hat{\theta}_n := \phi^{-1}\left(\frac{a}{2}\bar{\tau}_n^h + \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h}\right), \tag{3.1.4}$$

where

$$\bar{\tau}_n^h := \left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\tau_j}\right)^{-1}$$

denotes the harmonic mean of $\tau_1,\ldots,\tau_n$.

Proof: We make the maximum likelihood ansatz in order to estimate θ and compute the derivative of the log-likelihood function with respect to θ,

$$\begin{aligned}
&\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f_n\right)(\tau_1,\dots\tau_n;\theta)\\
&=\frac{\mathrm{d}\phi(\theta)}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}\phi(\theta)}\left[n\log\left(\frac{\phi(\theta)}{\sqrt{2\pi}\sigma}\right)+\sum_{j=1}^{n}\left\{-\frac{3}{2}\log(\tau_j)-\frac{(\phi(\theta)-a\tau_j)^2}{2\sigma^2\tau_j}\right\}\right]\\
&=\phi'(\theta)\left(\frac{n}{\phi(\theta)}+\sum_{j=1}^{n}-\frac{\phi(\theta)-a\tau_j}{\sigma^2\tau_j}\right)\\
&=\phi'(\theta)\left(\frac{n}{\phi(\theta)}+\frac{na}{\sigma^2}-\frac{\phi(\theta)}{\sigma^2}\sum_{j=1}^{n}\frac{1}{\tau_j}\right).
\end{aligned}\tag{3.1.5}$$

Since ϕ is a C^2-diffeomorphism, $\phi'(\theta)\neq 0$ holds for all $\theta\in\Theta$. If we multiply the last term of (3.1.5) by the factor

$$-\frac{\sigma^2\phi(\theta)}{\phi'(\theta)}\left(\sum_{j=1}^{n}\frac{1}{\tau_j}\right)^{-1}<0,$$

we get

$$\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log f_n\right)(\tau_1,\dots\tau_n;\theta)=0 \quad&\iff\quad \phi^2(\theta)-\phi(\theta)a\bar{\tau}_n^h-\sigma^2\bar{\tau}_n^h=0\\
&\iff\quad \phi(\theta)=\frac{a}{2}\bar{\tau}_n^h+\sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2+\sigma^2\bar{\tau}_n^h}
\end{aligned}\tag{3.1.6}$$

because $\phi>0$. Since ϕ is bijective, this shows that we have a unique critical point for $\theta\mapsto f_n(\tau_1,\dots\tau_n;\theta)$, given by

$$\hat{\theta}_n:=\phi^{-1}\left(\frac{a}{2}\bar{\tau}_n^h+\sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2+\sigma^2\bar{\tau}_n^h}\right).\tag{3.1.7}$$

In fact $\hat{\theta}_n$ yields a maximum of $f_n(t;\theta)$ in $\theta\in\Theta$ because

$$f_n(t;\theta)\to 0$$

as $\phi(\theta)\to 0$ or $\phi(\theta)\to\infty$. Hence $\hat{\theta}_n$ is the MLE for θ. □

In order to give asymptotic results for the MLE $\hat{\theta}_n$ defined in (3.1.4) the following lemmas are needed. We will prove the LAN property for the current experiment by showing differentiability in quadratic mean for the single experiment. Accordingly, the LAN property will lead to LAM for $\hat{\theta}_n$. But first we have to care about the negative moments of τ to compute score and Fisher information of the experiment.

Lemma 3.1.2 *Let $\tau \sim \mathbf{P}_\theta$, defined in* (3.1.2)*, then all negative moments of τ are finite, i.e. $\mathbf{E}[\tau^{-k}] < \infty$ for all $k \in \mathbb{N}$. Further, the first and second negative moments are given by*

$$\mathbf{E}[\tau^{-1}] = \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)}$$

and

$$\mathbf{E}[\tau^{-2}] = \frac{3\sigma^4 + 3\sigma^2 a\phi(\theta) + a^2\phi^2(\theta)}{\phi^4(\theta)}.$$

Proof: Let $k \in \mathbb{N}$. Obviously

$$\lim_{t\to 0,\infty} t^{-k} f(t;\theta) = \lim_{t\to 0,\infty} \frac{\phi(\theta)}{\sqrt{2\pi}\sigma} t^{-\frac{3}{2}-k} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} = 0$$

holds. Since $t \mapsto t^{-k} f(t;\theta)$ is continuous on $(0,\infty)$ this implies that $t^{-k}f(t;\theta)$ is bounded on $(0,\infty)$. Further, $t^{-k}f(t;\theta) = O(t^{-3/2})$, as $t \to \infty$. Hence we deduce

$$\mathbf{E}[\tau^{-k}] = \int_0^\infty t^{-k} f(t;\theta) \mathrm{d}t < \infty.$$

Now we compute the first negative moment of τ. Since

$$\lim_{t\to 0,\infty} t^{-1/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} = 0$$

and f, defined in (3.1.2), is a probability density, integration by parts leads to

$$\begin{aligned}
\frac{\sqrt{2\pi}\sigma}{\phi(\theta)} &= \int_0^\infty t^{-3/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= -\int_0^\infty -2t^{-1/2} \left(\frac{\phi^2(\theta)}{2\sigma^2 t^2} - \frac{a^2}{2\sigma^2}\right) \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= \frac{\phi^2(\theta)}{\sigma^2} \int_0^\infty t^{-5/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&\quad - \frac{a^2}{\sigma^2} \int_0^\infty t^{-1/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t.
\end{aligned} \tag{3.1.8}$$

Further, with the substitution $t \mapsto s^{-1}$,

$$\int_0^\infty t^{-1/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t = \int_0^\infty s^{-3/2} \exp\left\{-\frac{(\phi(\theta)s-a)^2}{2\sigma^2 s}\right\} \mathrm{d}s$$
$$= \frac{\sqrt{2\pi}\sigma}{a}$$

holds for $a > 0$. So we deduce from (3.1.8) that

$$\begin{aligned}
\mathbf{E}[\tau^{-1}] &= \frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-5/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= \left(1 + \frac{a^2}{\sigma^2}\frac{\sqrt{2\pi}\sigma}{a}\frac{\phi(\theta)}{\sqrt{2\pi}\sigma}\right)\frac{\sigma^2}{\phi^2(\theta)} \\
&= \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)}.
\end{aligned}$$

Note that this is also true for $a = 0$ because in this case the last term of the last line in (3.1.8) is equal to 0. Using the same arguments as above, one computes the second negative moment. Integration by parts yields

$$\begin{aligned}
\frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)} &= \frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-5/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= -\frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty -\frac{2}{3}t^{-3/2}\left(\frac{\phi^2(\theta)}{2t^2\sigma^2} - \frac{a^2}{2\sigma^2}\right) \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= \frac{\phi^2(\theta)}{3\sigma^2}\frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-7/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&\quad - \frac{a^2}{3\sigma^2}\frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-3/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= \frac{\phi^2(\theta)}{3\sigma^2}\frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-7/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t - \frac{a^2}{3\sigma^2}.
\end{aligned}$$

So we conclude

$$\begin{aligned}
\mathbf{E}[\tau^{-2}] &= \frac{\phi(\theta)}{\sqrt{2\pi}\sigma} \int_0^\infty t^{-7/2} \exp\left\{-\frac{(\phi(\theta)-at)^2}{2\sigma^2 t}\right\} \mathrm{d}t \\
&= \left(\frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)} + \frac{a^2}{3\sigma^2}\right)\frac{3\sigma^2}{\phi^2(\theta)} \\
&= \frac{3\sigma^4 + 3\sigma^2 a\phi(\theta) + a^2\phi^2(\theta)}{\phi^4(\theta)}.
\end{aligned}$$

$\square$

Lemma 3.1.3 *The family of probability measures*

$$\{\mathbf{P}_\theta : \theta \in \Theta\}$$

defined in (3.1.2) *is* differentiable in quadratic mean *for all* $\theta \in \Theta$ *with derivative (or score)*

$$V_\theta = \phi'(\theta) \cdot \left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} - \frac{\phi(\theta)}{\sigma^2 \tau} \right).$$

Further, the score V_θ *is a centered random variable under* $\mathbf{P}_\theta$*, i.e.*

$$\mathbf{E}_\theta[V_\theta] = 0.$$

Proof: From lemma 3.1.2 we have $\mathbf{E}_\theta[\tau^{-2}] < \infty$ implying $V_\theta \in L^2(\mathbf{P}_\theta)$ for all $\theta \in \Theta$. It is well-known that for the derivative in quadratic mean of a family of probability measures $\mathbf{E}_\theta[V_\theta] = 0$ holds. Anyway, with lemma 3.1.2 we verify

$$\begin{aligned} \mathbf{E}_\theta[V_\theta] &= \phi'(\theta) \left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} - \frac{\phi(\theta)}{\sigma^2} \mathbf{E}_\theta \left[\tau^{-1} \right] \right) \\ &= \phi'(\theta) \left(\frac{\sigma^2 + a\phi(\theta)}{\sigma^2 \phi(\theta)} - \frac{\phi(\theta)}{\sigma^2} \cdot \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)} \right) = 0. \end{aligned}$$

Since all $\mathbf{P}_\theta$, $\theta \in \Theta$ are equivalent by definition, the second condition in 2.1.3 is also fulfilled. Hence we only have to show

$$R(\delta) := \frac{1}{\delta^2} \int_0^\infty \left| \sqrt{\frac{\mathrm{d}\mathbf{P}_{\theta+\delta}}{\mathrm{d}\mathbf{P}_\theta}} - 1 - \delta \frac{1}{2} V_\theta \right|^2 \mathrm{d}\mathbf{P}_\theta \xrightarrow{\delta \to 0} 0 \tag{3.1.9}$$

for all $\theta \in \Theta$, to complete the proof.

Let $\theta \in \Theta$ be arbitrary and fix. Choose $\delta_0 > 0$ such that $\bar{B}_{\delta_0}(\theta) \subset \Theta$. For $|\delta| < \delta_0$ define

$$A_\tau(\delta) := \frac{(\phi(\theta) - a\tau)^2 - (\phi(\theta + \delta) - a\tau)^2}{4\sigma^2 \tau}$$

and

$$g_\tau(\delta) := \sqrt{\frac{\mathrm{d}\mathbf{P}_{\theta+\delta}}{\mathrm{d}\mathbf{P}_\theta}} = \sqrt{\frac{f(\tau; \theta + \delta)}{f(\tau, \theta)}} = \sqrt{\frac{\phi(\theta + \delta)}{\phi(\theta)}} \exp\{A_\tau(\delta)\}.$$

Now we rewrite (3.1.9) as

$$R(\delta) = \delta^2 \int_0^\infty \delta^{-4} \left| g_t(\delta) - 1 - \delta \frac{1}{2} V_\theta(t) \right|^2 f(t;\theta) \mathrm{d}t. \tag{3.1.10}$$

If we can show that the integral term above is bounded in δ, the proof is complete. For this purpose, we have to compute the first and second derivative of the function $g_t(\delta)$ with respect to δ. These are given by

$$\begin{aligned}
g_t'(\delta) =& \frac{\mathrm{d}\phi(\theta+\delta)}{\mathrm{d}\delta} \cdot \frac{\mathrm{d}g_t}{\mathrm{d}\phi(\theta+\delta)} \\
=& \phi'(\theta+\delta) \left(\frac{1}{\phi(\theta)} \frac{1}{2} \left(\frac{\phi(\theta+\delta)}{\phi(\theta)} \right)^{-1/2} \exp\left\{ A_t(\delta) \right\} - \frac{\phi(\theta+\delta) - at}{2\sigma^2 t} g_t(\delta) \right) \\
=& \phi'(\theta+\delta) \left(\frac{1}{\phi(\theta)} \frac{1}{2} \left(\frac{\phi(\theta+\delta)}{\phi(\theta)} \right)^{-1} g_t(\delta) - \frac{\phi(\theta+\delta) - at}{2\sigma^2 t} g_t(\delta) \right) \\
=& \phi'(\theta+\delta) \cdot \frac{g_t(\delta)}{2} \cdot \left(\frac{1}{\phi(\theta+\delta)} - \frac{\phi(\theta+\delta) - at}{\sigma^2 t} \right) \\
=& g_t(\delta) \frac{1}{2} V_{\theta+\delta}(t)
\end{aligned} \tag{3.1.11}$$

and

$$\begin{aligned}
g_t''(\delta) =& \big(g_t'(\delta)\phi'(\theta+\delta) + g_t(\delta)\phi''(\theta+\delta) \big) \frac{1}{2} \left(\frac{1}{\phi(\theta+\delta)} - \frac{\phi(\theta+\delta) - at}{\sigma^2 t} \right) \\
&+ g_t(\delta)\phi'(\theta+\delta) \frac{1}{2} \left(- \frac{\phi'(\theta+\delta)}{\phi^2(\theta+\delta)} - \frac{\phi'(\theta+\delta)}{\sigma^2 t} \right) \\
=& g_t(\delta) \frac{1}{2} \Bigg[(\phi'(\theta+\delta))^2 \frac{1}{2} \left(\frac{1}{\phi(\theta+\delta)} - \frac{\phi(\theta+\delta) - at}{\sigma^2 t} \right)^2 \\
&\quad + \phi''(\theta+\delta) \left(\frac{1}{\phi(\theta+\delta)} - \frac{\phi(\theta+\delta) - at}{\sigma^2 t} \right) \\
&\quad - (\phi'(\theta+\delta))^2 \left(\frac{1}{\phi^2(\theta+\delta)} + \frac{1}{\sigma^2 t} \right) \Bigg].
\end{aligned} \tag{3.1.12}$$

By Taylor expansion of g with respect to δ at 0, there exists a $\xi = \xi(t)$ between 0 and δ such that

$$\begin{aligned}
\delta^{-4}\left|g_t(\delta) - 1 - \delta\frac{1}{2}V_\theta(t)\right|^2 &= \delta^{-4}\left|g_t(\delta) - g_t(0) - \delta g_t'(0)\right|^2 \\
&= \delta^{-4}\left|\frac{\delta^2}{2}g_t''(\xi(t))\right|^2 \\
&\le \left|g_t''(\xi(t))\right|^2 .
\end{aligned} \tag{3.1.13}$$

Note that ϕ, ϕ' and ϕ'' are continuous and bounded on $\bar{B}_{\delta_0}(\theta)$. Further, ϕ is bounded away from 0 on $\bar{B}_{\delta_0}(\theta)$. Hence the last representation of g_t'' in (3.1.12) has an upper bound such that

$$\left|g_t''(\delta)\right|^2 \le (g_t(\delta))^2 \tilde{K}\left(1 + \frac{1}{t} + \frac{1}{t^2} + \frac{1}{t^3} + \frac{1}{t^4}\right)$$

for all $|\delta| \le \delta_0$ and $t \in (0, \infty)$, where $0 \le \tilde{K} < \infty$ is a suitable constant independent of t and δ. Moreover, there exists a constant $0 \le K < \infty$ such that

$$\left|g_t''(\delta)\right|^2 \le (g_t(\delta))^2 K\left(1 + \frac{1}{t^4}\right) \quad \text{for all } |\delta| \le \delta_0 \text{ and } t \in (0, \infty). \tag{3.1.14}$$

Since $|\delta| \le \delta_0$ and $\xi = \xi(t)$ is situated between 0 and δ, we have $|\xi(t)| \le \delta_0$. Consequently, (3.1.14) leads to

$$\begin{aligned}
&\left|g_t''(\xi(t))\right|^2 f(t;\theta) \\
&\le K\left(1 + \frac{1}{t^4}\right)(g_t(\xi(t)))^2 f(t;\theta) \\
&= K\left(1 + \frac{1}{t^4}\right)\frac{\phi(\theta + \xi(t))}{\phi(\theta)}\exp\left\{\frac{(\phi(\theta) - at)^2 - (\phi(\theta + \xi(t)) - at)^2}{2\sigma^2 t}\right\} \\
&\quad \cdot \frac{\phi(\theta)}{\sqrt{2\pi}\sigma}t^{-3/2}\exp\left\{-\frac{(\phi(\theta) - at)^2}{2\sigma^2 t}\right\} \\
&= K\left(\frac{1}{t^{3/2}} + \frac{1}{t^{11/2}}\right)\frac{\phi(\theta + \xi(t))}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\phi(\theta + \xi(t)) - at)^2}{2\sigma^2 t}\right\}
\end{aligned} \tag{3.1.15}$$

Since ϕ is positive and monotone, we have for all $t \in (0, \infty)$

$$\begin{aligned}
0 < \phi_* :&= \min\{\phi(\theta + \delta_0), \phi(\theta - \delta_0)\} \\
&\le \phi(\theta + \xi(t)) \\
&\le \max\{\phi(\theta + \delta_0), \phi(\theta - \delta_0)\} =: \phi^* < \infty.
\end{aligned}$$

Thus, the following upper bound for the last term in (3.1.15)

$$K\left(\frac{1}{t^{3/2}}+\frac{1}{t^{11/2}}\right)\frac{\phi(\theta+\xi(t))}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\phi(\theta+\xi(t))-at)^2}{2\sigma^2 t}\right\}$$
$$\leq G(t) := \left(\frac{1}{t^{3/2}}+\frac{1}{t^{11/2}}\right)\frac{K\phi^*}{\sqrt{2\pi}\sigma}\cdot\begin{cases}\exp\left\{-\frac{(\phi_*/2)^2}{2\sigma^2 t}\right\} & \text{if } at < \phi_*/2\\ 1 & \text{else}\end{cases} \tag{3.1.16}$$

holds for all $|\delta| \leq \delta_0$ and $t \in (0,\infty)$. If we combine (3.1.13), (3.1.15) and (3.1.16), we get

$$\delta^{-4}\left|g_t(\delta)-1-\delta\frac{1}{2}V_\theta(t)\right|^2 f(t;\theta) \leq G(t) \quad \text{for all } |\delta| \leq \delta_0 \text{ and } t \in (0,\infty).$$

Together with (3.1.10), we have for $|\delta| \leq \delta_0$ that

$$R(\delta) \leq \delta^2 \int_0^\infty G(t)\mathrm{d}t. \tag{3.1.17}$$

Note that G is independent from δ. Clearly, $G(t) \to 0$ as $t \to 0$ or $t \to \infty$. Further, $G(\cdot)$ is continuous except for one point and cadlag by definition. Consequently G is bounded on $(0,\infty)$ and obviously $G(t) = O(t^{-3/2})$ as $t \to \infty$. Hence $\int_0^\infty G(t)\mathrm{d}t < \infty$. With (3.1.17) in mind, this shows finally $R(\delta) \to 0$, as $\delta \to 0$. □

Lemma 3.1.4 *Consider again the single experiment*

$$\mathscr{E} := ((0,\infty), \mathscr{B}((0,\infty)), \{\mathbf{P}_\theta : \theta \in \Theta\})$$

defined in (3.1.1)*. The Fisher information* $J_\theta = \mathbf{E}_\theta[V_\theta^2]$ *of* $\mathscr{E}$ *is given by*

$$J_\theta = (\phi'(\theta))^2\,\frac{2\sigma^2 + a\phi(\theta)}{\phi^2(\theta)\sigma^2}.$$

Proof: From lemma 3.1.3 we know that $\mathscr{E}$ is differentiable in quadratic mean for all $\theta \in \Theta$ with derivative $V_\theta \in L^2(\mathbf{P}_\theta)$ and $\mathbf{E}_\theta[V_\theta] = 0$. Hence the Fisher information is well defined by $J_\theta = \mathbf{E}_\theta[V_\theta^2]$. To compute J_θ we apply again lemma 3.1.2 which yields

$$
\begin{aligned}
J_\theta &= \mathbf{E}_\theta \left[(\phi'(\theta))^2 \left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} - \frac{\phi(\theta)}{\sigma^2 \tau} \right)^2 \right] \\
&= (\phi'(\theta))^2 \left[\left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} \right)^2 + \frac{\phi^2(\theta)}{\sigma^4} \mathbf{E}[\tau^{-2}] - 2 \left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} \right) \frac{\phi(\theta)}{\sigma^2} \mathbf{E}[\tau^{-1}] \right] \\
&= (\phi'(\theta))^2 \left[\frac{1}{\phi^2(\theta)} + 2 \frac{a}{\phi(\theta)\sigma^2} + \frac{a^2}{\sigma^4} + \frac{\phi^2(\theta)}{\sigma^4} \cdot \frac{3\sigma^4 + 3\sigma^2 a \phi(\theta) + a^2 \phi^2(\theta)}{\phi^4(\theta)} \right. \\
&\qquad \left. - \left(\frac{2}{\sigma^2} + \frac{2a\phi(\theta)}{\sigma^4} \right) \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)} \right] \\
&= (\phi'(\theta))^2 \left[\frac{1}{\phi^2(\theta)} + \frac{2a}{\phi(\theta)\sigma^2} + \frac{a^2}{\sigma^4} + \frac{3}{\phi^2(\theta)} + \frac{3a}{\sigma^2 \phi(\theta)} + \frac{a^2}{\sigma^4} \right. \\
&\qquad \left. - \left(\frac{2}{\phi^2(\theta)} + \frac{2a}{\phi(\theta)\sigma^2} + \frac{2a}{\phi(\theta)\sigma^2} + \frac{2a^2}{\sigma^4} \right) \right] \\
&= (\phi'(\theta))^2 \left[\frac{2}{\phi^2(\theta)} + \frac{a}{\phi(\theta)\sigma^2} \right]
\end{aligned}
$$

and the assertion is proved. □

Now we come to the main result of this section showing the MLE defined in (3.1.4) to be asymptotically the best of all estimators one can construct in the current experiment. (cf. section 2.1)

Theorem 3.1.5 *The sequence of statistical experiments $\mathscr{E}_n$, $n \in \mathbb{N}$ defined in* (3.1.3) *satisfies the LAN condition for all $\theta \in \Theta$. Further, the MLE $\hat{\theta}_n$, defined in* (3.1.4) *is strongly consistent, regular, efficient and LAM for all $\theta \in \Theta$, which implies*

$$
\sqrt{n} \left(\hat{\theta}_n - \theta \right) \xrightarrow[n \to \infty]{\mathcal{L}(\theta)} \mathcal{N} \left(0, J_\theta^{-1} \right),
$$

where $J_\theta = (\phi'(\theta))^2 \dfrac{2\sigma^2 + a\phi(\theta)}{\phi^2(\theta)\sigma^2}$.

Proof: 1) The LAN property: From lemma 3.1.3 we know that the single experiment $\mathscr{E}$ is differentiable in quadratic mean for all $\theta \in \Theta$ with derivative V_θ. Hence Le Cam's 2nd Lemma 2.1.4 yields the LAN property for the sequence of product experiments $\mathscr{E}_n$, $n \in \mathbb{N}$.
2) The strong consistency of $\hat{\theta}_n$ is easy to see if we apply the SLLN and

lemma 3.1.2. We have

$$(\bar{\tau}_n^h)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tau_i} \xrightarrow[n\to\infty]{\mathbf{P}_\theta\text{-a.s.}} \mathbf{E}[\tau^{-1}] = \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)},$$

and accordingly

$$\bar{\tau}_n^h \xrightarrow[n\to\infty]{\mathbf{P}_\theta\text{-a.s.}} \frac{\phi^2(\theta)}{\sigma^2 + a\phi(\theta)}. \tag{3.1.18}$$

Since ϕ^{-1} is continuous, we deduce

$$\begin{aligned}
\lim_{n\to\infty}\hat{\theta}_n &= \lim_{n\to\infty}\phi^{-1}\left(\frac{a}{2}\bar{\tau}_n^h + \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h}\right)\\
&\overset{\text{a.s.}}{=} \phi^{-1}\left(\frac{a}{2}\cdot\frac{\phi^2(\theta)}{\sigma^2 + a\phi(\theta)} + \sqrt{\left(\frac{a}{2}\cdot\frac{\phi^2(\theta)}{\sigma^2 + a\phi(\theta)}\right)^2 + \sigma^2\frac{\phi^2(\theta)}{\sigma^2 + a\phi(\theta)}}\right)\\
&= \phi^{-1}\left(\frac{a\phi^2(\theta)}{2(a\phi(\theta)+\sigma^2)} + \sqrt{\frac{\phi^2(\theta)\cdot[a^2\phi^2(\theta) + 4a\phi(\theta)\sigma^2 + 4\sigma^4]}{4(a\phi(\theta)+\sigma^2)^2}}\right)\\
&= \phi^{-1}\left(\frac{a\phi^2(\theta) + \phi(\theta)(a\phi(\theta) + 2\sigma^2)}{2(a\phi(\theta)+\sigma^2)}\right)\\
&= \phi^{-1}\left(\phi(\theta)\frac{a\phi(\theta) + a\phi(\theta) + 2\sigma^2}{2(a\phi(\theta)+\sigma^2)}\right)\\
&= \phi^{-1}\left(\phi(\theta)\right)\\
&= \theta.
\end{aligned} \tag{3.1.19}$$

3) From now on we fix an arbitrary $\theta \in \Theta$. Let us recall some results from previous lemmas and give some notation. Define $(V_\theta^i)_{i\in\mathbb{N}}$ by

$$V_\theta^i := \phi'(\theta)\left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} - \frac{\phi(\theta)}{\sigma^2\tau_i}\right), \quad i \in \mathbb{N},$$

a sequence of iid copies of V_θ computed in 3.1.3. Further, define

$$S_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_\theta^i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi'(\theta)\left(\frac{1}{\phi(\theta)} + \frac{a}{\sigma^2} - \frac{\phi(\theta)}{\sigma^2\tau_i}\right), \quad n \in \mathbb{N}, \tag{3.1.20}$$

as in 2.1.4. From the central limit theorem we know that

$$S_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, J), \tag{3.1.21}$$

where

$$J = J_\theta = \mathbf{Var}_\theta[V_\theta] = (\phi'(\theta))^2 \, \frac{2\sigma^2 + a\phi(\theta)}{\phi^2(\theta)\sigma^2}$$

is computed in lemma 3.1.4. If we define $Z_n := J_\theta^{-1} S_n(\theta)$ this implies

$$Z_n \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\left(0, J_\theta^{-1}\right).$$

In the situation of LAN, lemma 2.1.8 characterizes regularity and efficiency for $\hat{\theta}_n$, $n \in \mathbb{N}$ by

$$\sqrt{n}(\hat{\theta}_n - \theta) - Z_n \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0. \tag{3.1.22}$$

Further, the minimax Thm. 2.1.9 shows that (3.1.22) also implies the LAM property for $\hat{\theta}_n$. Thus we only have to show (3.1.22) to complete the proof. 4) This coupling will be done by a Taylor expansion, where we express $\hat{\theta}_n$ in terms of $1/\tau$. Define the real function $g : (0, \infty) \to \mathbb{R}$ by

$$g(x) := \phi^{-1}\left(\frac{a}{2}\frac{1}{x} + \sqrt{\left(\frac{a}{2}\frac{1}{x}\right)^2 + \sigma^2\frac{1}{x}}\right),$$

then $\hat{\theta}_n = g\left(\left(\bar{\tau}_n^h\right)^{-1}\right)$, cf. (3.1.4). In order to apply the Taylor expansion, note that g is C^2 on $(0, \infty)$ and

$$\begin{aligned} g'(x) &= \left(-\frac{a}{2x^2} - \frac{\frac{a^2}{2x^3} + \frac{\sigma^2}{x^2}}{2\sqrt{(\frac{a}{2}\frac{1}{x})^2 + \sigma^2\frac{1}{x}}}\right) \cdot \frac{1}{\phi'(g(x))} \\ &= \frac{1}{x^2 \cdot \phi'(g(x))}\left(-\frac{a}{2} - \frac{\frac{a^2}{2x} + \sigma^2}{2\sqrt{(\frac{a}{2}\frac{1}{x})^2 + \sigma^2\frac{1}{x}}}\right). \end{aligned}$$

For $x^* := \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)} > 0$ we have by the same computation as in (3.1.19) that

$$g(x^*) = \theta. \tag{3.1.23}$$

Further,

$$
\begin{aligned}
g'(x^*) &= \frac{\phi^4(\theta)}{(\sigma^2 + a\phi(\theta))^2 \cdot \phi'(\theta)} \left(-\frac{a}{2} - \frac{\frac{a^2}{2}\frac{\phi^2(\theta)}{\sigma^2+a\phi(\theta)} + \sigma^2}{2\sqrt{\left(\frac{a}{2}\frac{\phi^2(\theta)}{\sigma^2+a\phi(\theta)}\right)^2 + \sigma^2 \frac{\phi^2(\theta)}{\sigma^2+a\phi(\theta)}}} \right) \\
&= \frac{\phi^4(\theta)}{(\sigma^2 + a\phi(\theta))^2 \cdot \phi'(\theta)} \left(-\frac{a}{2} - \frac{\frac{a^2}{2}\frac{\phi^2(\theta)}{\sigma^2+a\phi(\theta)} + \sigma^2}{2\sqrt{\frac{\phi^2(\theta)\cdot(a^2\phi^2(\theta)+4\sigma^4+4\sigma^2 a\phi(\theta))}{4(\sigma^2+a\phi(\theta))^2}}} \right) \\
&= \frac{\phi^4(\theta)}{(\sigma^2 + a\phi(\theta))^2 \cdot \phi'(\theta)} \left(-\frac{a}{2} - \frac{\frac{a^2}{2}\frac{\phi^2(\theta)}{\sigma^2+a\phi(\theta)} + \sigma^2}{\frac{\phi(\theta)\cdot(a\phi(\theta)+2\sigma^2)}{a\phi(\theta)+\sigma^2}} \right) \\
&= \frac{\phi^4(\theta)}{(\sigma^2 + a\phi(\theta))^2 \cdot \phi'(\theta)} \left(-\frac{a}{2} - \frac{a^2\phi^2(\theta) + 2\sigma^2(a\phi(\theta)+\sigma^2)}{2\phi(\theta)\cdot(a\phi(\theta)+2\sigma^2)} \right) \\
&= \frac{\phi^4(\theta)}{(\sigma^2 + a\phi(\theta))^2 \cdot \phi'(\theta)} \cdot \frac{-2a^2\phi^2(\theta) - 4a\phi(\theta)\sigma^2 - 2\sigma^4}{2\phi(\theta)\cdot(a\phi(\theta)+2\sigma^2)} \\
&= -\frac{\phi^3(\theta)}{\phi'(\theta)\cdot(a\phi(\theta)+2\sigma^2)}
\end{aligned}
\tag{3.1.24}
$$

holds. Hence the Taylor expansion of g at x^* leads to

$$
g(x) = g(x^*) + g'(x^*)\cdot(x - x^*) + \frac{g''(\xi)}{2}(x - x^*)^2
$$

where $x > 0$ and ξ is a suitable value between x and x^*. If we set $x = x_n := (\bar{\tau}_n^h)^{-1} = \frac{1}{n}\sum_{i=1}^n \frac{1}{\tau_i} > 0$, by (3.1.23) and (3.1.24) there exists a random point ξ_n between the random point x_n and the fixed point x^* such that

$$
\hat{\theta}_n = g(x_n) = \theta - \frac{\phi^3(\theta)}{\phi'(\theta)\cdot(a\phi(\theta)+2\sigma^2)}(x_n - x^*) + \frac{g''(\xi_n)}{2}(x_n - x^*)^2.
\tag{3.1.25}
$$

Note from (3.1.18) that $x_n \xrightarrow[n\to\infty]{\mathbf{P}_\theta\text{-a.s.}} x^*$ and consequently $\xi_n \xrightarrow[n\to\infty]{\mathbf{P}_\theta\text{-a.s.}} x^*$. Since $x \mapsto g''(x)$ is continuous, this implies

$$
g''(\xi_n) \xrightarrow[n\to\infty]{\mathbf{P}_\theta\text{-a.s.}} g''(x^*).
\tag{3.1.26}
$$

Further, with S_n defined in (3.1.20), $x_n = (\bar{\tau}_n^h)^{-1}$ and $x^* = \frac{\sigma^2 + a\phi(\theta)}{\phi^2(\theta)}$ we have

$$
\begin{aligned}
-\frac{\sigma^2}{\phi'(\theta)\phi(\theta)} n^{-1/2} S_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{\sigma^2}{\phi^2(\theta)} - \frac{a}{\phi(\theta)} + \frac{1}{\tau_i}\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tau_i} - \left(\frac{\sigma^2}{\phi^2(\theta)} + \frac{a}{\phi(\theta)}\right) \\
&= (x_n - x^*).
\end{aligned}
$$

If we put this into (3.1.25), we get

$$
\begin{aligned}
\hat{\theta}_n = \theta &- \frac{\phi^3(\theta)}{\phi'(\theta)\cdot(a\phi(\theta)+2\sigma^2)} \cdot \left(-\frac{\sigma^2}{\phi'(\theta)\phi(\theta)} n^{-1/2} S_n(\theta)\right) \\
&+ \frac{g''(\xi_n)}{2}\left(\frac{\sigma^2}{\phi'(\theta)\phi(\theta)} n^{-1/2} S_n(\theta)\right)^2 \\
= \theta &+ n^{-1/2} J_\theta^{-1} S_n(\theta) + n^{-1} S_n^2(\theta) \frac{g''(\xi_n)}{2} \frac{\sigma^4}{(\phi'(\theta)\phi(\theta))^2}
\end{aligned}
$$

and hence with $Z_n = J_\theta^{-1} S_n(\theta)$

$$
\sqrt{n}(\hat{\theta}_n - \theta) - Z_n = n^{-1/2} S_n^2(\theta) \frac{g''(\xi_n)}{2} \frac{\sigma^4}{(\phi'(\theta)\phi(\theta))^2} \tag{3.1.27}
$$

holds. From (3.1.21) and the continuous mapping theorem we know that $\{S_n^2 : n \in \mathbb{N}\}$ converges in distribution. Together with (3.1.26) and Slutsky's Lemma, this implies

$$
n^{-1/2} S_n^2(\theta) \frac{g''(\xi_n)}{2} \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0.
$$

So with (3.1.27) in mind we finally conclude that (3.1.22) holds. Since $\theta \in \Theta$ was arbitrary, the proof is complete. $\square$

3.2 Consequences for Brownian Motion with Drift

In our estimation problem, described in chapter 1, we aim to estimate θ_1 and θ_2 from iid observations of the level crossing times from θ_1 to θ_2. In this section, we assume that the process X starting in θ_1 and hitting the level θ_2 is a Brownian motion with drift (BMD) which is defined in the following. Let $\sigma > 0$, $a \geq 0$ and $(B_t)_{t\geq 0}$ be a standard Brownian motion. Define $X = (X_t)_{t\geq 0}$, a Brownian motion with nonnegative drift, starting in $X_0 = \theta_1$, by

$$X_t := \theta_1 + at + \sigma B_t, \quad t \geq 0.$$

Referring to (1.2.2), we denote the level crossing time of the Brownian motion with drift by

$$\begin{aligned} \tau^{BD} &= \inf\left\{t \geq 0 \,\middle|\, \theta_1 + at + \sigma B_t = \theta_2\right\} \\ &= \inf\left\{t \geq 0 \,\middle|\, \frac{\theta_1}{\sigma} + \frac{a}{\sigma}t + B_t = \frac{\theta_2}{\sigma}\right\}. \end{aligned} \tag{3.2.1}$$

To ensure $\tau < \infty$ a.s., which is reasonable in order to model spiking behavior, it is important to allow only nonnegative drift. We know from [BS 02] II.2.2 p.295 that the distribution of τ^{BD} has an explicit Lebesgue density given by

$$f^{BD}(t) := \frac{(\theta_2 - \theta_1)}{\sqrt{2\pi}\sigma} t^{-3/2} \exp\left\{-\frac{((\theta_2 - \theta_1) - at)^2}{2\sigma^2 t}\right\}, \qquad t \in (0, \infty). \tag{3.2.2}$$

From (3.2.2) we realize that the joint estimation of θ_1 and θ_2 by observing τ^{BD} is impossible because we have a problem of identifiability. If $\theta_2 - \theta_1 = \theta_2' - \theta_1'$, different parameters induce the same density. This is due to the fact that the Brownian motion with drift is homogeneous in space.

So we consider two statistical experiments in order to estimate θ_1 and θ_2 respectively. Comparing (3.2.2) and (3.1.2), we further see that both statistical experiments are special cases of the estimation problem in section 3.1. To estimate θ_1 we claim to know θ_2. Accordingly, in 3.1 we choose $\Theta = \Theta_1 = (-\infty, \theta_2)$ and $\phi(\theta_1) = \theta_2 - \theta_1$ such that we obtain the statistical experiment

$$\mathscr{E}_n^1 := \big((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_1} : \theta_1 \in \Theta_1\}\big), \quad n \in \mathbb{N},$$

where

$$\frac{\mathrm{d}\mathbf{P}_{n,\theta_1}}{\mathrm{d}(t_1,\ldots,t_n)} = f_n^{BD}(t_1,\ldots,t_n;\theta_1) := \prod_{j=1}^n f^{BD}(t_j,\theta_1).$$

Conversely, to estimate θ_2 we claim to know θ_1 and in 3.1 we choose $\Theta = \Theta_2 = (\theta_1, \infty)$ and $\phi(\theta_2) = \theta_2 - \theta_1$. Hence the corresponding statistical experiment is

$$\mathscr{E}_n^2 := \left((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_2} : \theta_1 \in \Theta_2\}\right), \quad n \in \mathbb{N},$$

where

$$\frac{\mathrm{d}\mathbf{P}_{n,\theta_2}}{\mathrm{d}(t_1,\ldots,t_n)} = f_n^{BD}(t_1,\ldots,t_n;\theta_2) := \prod_{j=1}^{n} f^{BD}(t_j,\theta_2).$$

Corollary 3.2.1 *Let $\bar{\tau}_n^h$ denote the harmonic mean of the iid level crossing times $\left(\tau_1^{BD},\ldots,\tau_n^{BD}\right)$. Both sequences of statistical experiments*

$$\mathscr{E}_n^i := \left((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_i} : \theta_i \in \Theta_i\}\right), \quad n \in \mathbb{N}, \quad i = 1,2$$

satisfy the LAN condition for all $\theta_1 \in \Theta_1 = (-\infty, \theta_2)$ and $\theta_2 \in \Theta_2 = (\theta_1, \infty)$ respectively. The corresponding MLEs

$$\begin{aligned} \hat{\theta}_{1,n} &:= \theta_2 - \frac{a}{2}\bar{\tau}_n^h - \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h}, \\ \hat{\theta}_{2,n} &:= \theta_1 + \frac{a}{2}\bar{\tau}_n^h + \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h} \end{aligned} \tag{3.2.3}$$

are strongly consistent, regular, efficient and LAM for all $\theta_i \in \Theta_i$, $i = 1,2$ such that

$$\sqrt{n}\left(\hat{\theta}_{i,n} - \theta_i\right) \xrightarrow[n\to\infty]{\mathcal{L}(\theta_i)} \mathcal{N}\left(0, J_{\theta_i}^{-1}\right), \quad i = 1,2,$$

where

$$J_{\theta_i}^{-1} = \frac{(\theta_2-\theta_1)^2\sigma^2}{2\sigma^2 + a(\theta_2-\theta_1)}, \quad i = 1,2.$$

Proof: To estimate θ_1, we choose $\phi(\theta_1) = \theta_2 - \theta_1$. Then $\phi'(\theta_1) = -1$ and Thm. 3.1.5 implies all the assertions for $\mathscr{E}_n^1$, $n \in \mathbb{N}$ and its MLE $\hat{\theta}_{1,n}$ stated above. Analogously, for θ_2 we choose $\phi(\theta_2) = \theta_2 - \theta_1$. Hence $\phi'(\theta_2) = 1$ and again Thm. 3.1.5 implies the assertions for $\mathscr{E}_n^2$, $n \in \mathbb{N}$ and its MLE $\hat{\theta}_{2,n}$ stated above. □

Remark:

- $\hat{\theta}_{1,n}$ and $\hat{\theta}_{2,n}$ are connected via

$$\hat{\theta}_{2,n} - \theta_1 = \theta_2 - \hat{\theta}_{1,n}.$$

- The MLEs given in (3.2.3) are LAM and attain the local asymptotic minimax bound (cf. section 2.1 and Thm. 2.1.9). Hence, for the current experiments one cannot construct estimators asymptotically better than the MLEs.

- In the following, we analyze the boundary behavior of $J_{\theta_i}^{-1}$, the variance of the limit distribution, for the parameters $a, \sigma \geq 0$. The cases $a \longrightarrow \infty$ or $\sigma \longrightarrow 0$ represent a more deterministic behavior of the process, where drift is much stronger than fluctuations. Hence, one would expect that the variance vanishes. In fact we have

$$J_{\theta_i}^{-1} = \frac{(\theta_2 - \theta_1)^2\sigma^2}{2\sigma^2 + a(\theta_2 - \theta_1)} \longrightarrow 0, \quad a \longrightarrow \infty \text{ or } \sigma \longrightarrow 0.$$

  On the other hand

$$J_{\theta_i}^{-1} = \frac{(\theta_2 - \theta_1)^2\sigma^2}{2\sigma^2 + a(\theta_2 - \theta_1)} \longrightarrow \frac{(\theta_2 - \theta_1)^2}{2}, \quad a \longrightarrow 0 \text{ or } \sigma \longrightarrow \infty.$$

  Note that this limit is also an upper bound for $J_{\theta_i}^{-1}$ for all $a, \sigma \geq 0$.

- In the case of standard Brownian motion, the MLEs for θ_i, $i = 1, 2$ have the simple expression

$$\hat{\theta}_{1,n} := \theta_2 - \sqrt{\bar{\tau}_n^h}, \quad \text{and} \quad \hat{\theta}_{2,n} := \theta_1 + \sqrt{\bar{\tau}_n^h}$$

  respectively. So in this case the distance $\theta_2 - \theta_1$ between θ_1 and θ_2 is estimated simply by the square root of the harmonic mean of $(\tau_1, \ldots, \tau_n)$.

3.3 Consequences for Geometric Brownian Motion

Analogously to the previous section, we consider the estimation problem described in chapter 1 but assuming that the process X, starting in θ_1 and hitting the level θ_2, is a geometric Brownian motion (GBM). For $\sigma > 0$, $b \in \mathbb{R}$ and $(B_t)_{t\geq 0}$, a standard Brownian motion, the geometric Brownian motion $X = (X_t)_{t\geq 0}$ is defined as the strong solution of the SDE

$$\mathrm{d}X_t = bX_t\mathrm{d}t + \sigma X_t\mathrm{d}B_t, \quad X_0 = \theta_1 > 0,$$

given by

$$X_t = \theta_1 \exp\left[\sigma B_t + \left(b - \frac{\sigma^2}{2}\right)t\right], \quad t \geq 0.$$

This can be seen by Itô's Formula (cf. e.g. [K 06], example 26.6). In our biological context, X describes a membrane potential which is not asymptotically decreasing. This suggests the assumption that

$$a := b - \frac{\sigma^2}{2} \geq 0.$$

Further, X is positive with lower bound 0 which is not reasonable from a biological point of view. However, a shift $(X_t - c)_{t\geq 0}$, $c \in \mathbb{R}$ easily resolves this problem, so we omit this c. Hence the GBM is just the exponential of a Brownian motion with positive drift. The level crossing time of the geometric Brownian motion is denoted by

$$\begin{aligned}\tau^{GB} &= \inf\left\{t \geq 0 \;\middle|\; \theta_1 \exp\left[\sigma B_t + at\right] = \theta_2\right\}\\ &= \inf\left\{t \geq 0 \;\middle|\; \log(\theta_1) + at + \sigma B_t = \log(\theta_2)\right\}.\end{aligned} \tag{3.3.1}$$

Comparing (3.2.1) and (3.3.1), we realize that τ^{GB} differs from τ^{BD} only by the logarithmic scale in θ_1 and θ_2. Consequently, with (3.2.2) in mind, the density of $\mathcal{L}(\tau^{GB})$ is given by

$$f^{GB}(t) := \frac{\log(\theta_2/\theta_1)}{\sqrt{2\pi}\sigma} t^{-3/2} \exp\left\{-\frac{(\log(\theta_2/\theta_1) - at)^2}{2\sigma^2 t}\right\}, \qquad t \in (0, \infty). \tag{3.3.2}$$

Also in this case we have a problem of identifiability because different parameters induce the same density, e.g. $\theta_2/\theta_1 = \theta_2'/\theta_1'$. Thus the joint estimation of θ_1 and θ_2, by observing an iid sequence τ_i^{GB}, $i = 1, \ldots, n$, is impossible.

Again, we have two statistical experiments in order to estimate θ_1 and θ_2 respectively. Further, comparing (3.3.2) and (3.1.2) we recognize again that both statistical experiments are special cases of the estimation problem treated in section 3.1. To estimate θ_1, we choose $\Theta = \Theta_1 = (0, \theta_2)$ and $\phi(\theta_1) = \log(\theta_2/\theta_1)$ in 3.1, where θ_2 is assumed to be known. From this we obtain the statistical experiment

$$\mathscr{E}_n^1 := \big((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_1} : \theta_1 \in \Theta_1\}\big), \quad n \in \mathbb{N},$$

where

$$\frac{\mathrm{d}\mathbf{P}_{n,\theta_1}}{\mathrm{d}(t_1,\ldots,t_n)} = f_n^{GB}(t_1,\ldots,t_n;\theta_1) := \prod_{j=1}^n f^{GB}(t_j,\theta_1).$$

To estimate θ_2, we choose $\Theta = \Theta_2 = (\theta_1, \infty)$ and $\phi(\theta_2) = \log(\theta_2/\theta_1)$ in 3.1, where θ_1 is assumed to be known. Hence the corresponding statistical experiment is

$$\mathscr{E}_n^2 := \big((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_2} : \theta_1 \in \Theta_2\}\big), \quad n \in \mathbb{N},$$

where

$$\frac{\mathrm{d}\mathbf{P}_{n,\theta_2}}{\mathrm{d}(t_1,\ldots,t_n)} = f_n^{GB}(t_1,\ldots,t_n;\theta_2) := \prod_{j=1}^n f^{GB}(t_j,\theta_2).$$

Corollary 3.3.1 *Let $\bar{\tau}_n^h$ denote the harmonic mean of the iid level crossing times $\big(\tau_1^{GB},\ldots,\tau_n^{GB}\big)$. Both sequences of statistical experiments*

$$\mathscr{E}_n^i := \big((0,\infty)^n, \mathscr{B}((0,\infty)^n), \{\mathbf{P}_{n,\theta_i} : \theta_i \in \Theta_i\}\big), \quad n \in \mathbb{N}, \quad i = 1,2$$

satisfy the LAN condition for all $\theta_1 \in \Theta_1 = (0,\theta_2)$ and $\theta_2 \in \Theta_2 = (\theta_1,\infty)$ respectively. The corresponding MLEs

$$\begin{aligned} \hat{\theta}_{1,n} &:= \theta_2 \exp\left\{-\frac{a}{2}\bar{\tau}_n^h - \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h}\right\}, \\ \hat{\theta}_{2,n} &:= \theta_1 \exp\left\{\frac{a}{2}\bar{\tau}_n^h + \sqrt{\left(\frac{a}{2}\bar{\tau}_n^h\right)^2 + \sigma^2\bar{\tau}_n^h}\right\} \end{aligned} \tag{3.3.3}$$

are strongly consistent, regular, efficient and LAM for all $\theta_i \in \Theta_i$, $i = 1,2$ such that

$$\sqrt{n}\left(\hat{\theta}_{i,n} - \theta_i\right) \xrightarrow[n\to\infty]{\mathcal{L}(\theta_i)} \mathcal{N}\left(0, J_{\theta_i}^{-1}\right), \quad i = 1,2,$$

where

$$J_{\theta_1}^{-1} = \frac{(\theta_1\sigma\log(\theta_2/\theta_1))^2}{2\sigma^2 + a\log(\theta_2/\theta_1)}$$

and

$$J_{\theta_2}^{-1} = \frac{(\theta_2\sigma\log(\theta_2/\theta_1))^2}{2\sigma^2 + a\log(\theta_2/\theta_1)}.$$

Proof: To estimate θ_1, we choose $\phi(\theta_1) = \log(\theta_2/\theta_1)$. Then $\phi'(\theta_1) = -1/\theta_1$ and Thm. 3.1.5 implies all the assertions for $\mathscr{E}_n^1$, $n \in \mathbb{N}$ and its MLE $\hat{\theta}_{1,n}$ stated above. Analogously, for θ_2 we choose $\phi(\theta_2) = \log(\theta_2/\theta_1)$. Hence $\phi'(\theta_2) = 1/\theta_2$ and Thm. 3.1.5 again implies the assertions for $\mathscr{E}_n^2$, $n \in \mathbb{N}$ and its MLE $\hat{\theta}_{2,n}$ stated above. □

Remark:

- $\hat{\theta}_{1,n}$ and $\hat{\theta}_{2,n}$ are connected via
$$\hat{\theta}_{2,n}/\theta_1 = \theta_2/\hat{\theta}_{1,n}.$$
- As in the case of BMD, also for GBM the MLEs given in (3.3.3) are LAM and attain the local asymptotic minimax bound (cf. section 2.1 and Thm. 2.1.9). Hence, also for the experiments in this section, one cannot construct estimators asymptotically better than the MLEs.
- Again, we analyze the boundary behavior of $J_{\theta_i}^{-1}$, the variance of the limit distribution, for the parameters $a, \sigma \geq 0$. As mentioned in the previous section, the cases $a \to \infty$ or $\sigma \to 0$ represent a more deterministic behavior of the process, where drift is much stronger than fluctuations. For this reason, as in the previous section, the variance vanishes or formally
$$J_{\theta_i}^{-1} = \frac{(\theta_i \sigma \log(\theta_2/\theta_1))^2}{2\sigma^2 + a\log(\theta_2/\theta_1)} \longrightarrow 0, \quad a \to \infty \text{ or } \sigma \to 0.$$
Further,
$$J_{\theta_i}^{-1} = \frac{(\theta_i \sigma \log(\theta_2/\theta_1))^2}{2\sigma^2 + a\log(\theta_2/\theta_1)} \longrightarrow \frac{\theta_i^2 \log^2(\theta_2/\theta_1)}{2}, \quad a \to 0 \text{ or } \sigma \to \infty$$
holds and again the limit is also an upper bound for $J_{\theta_i}^{-1}$ for all $a, \sigma \geq 0$.

Chapter 4

Ornstein-Uhlenbeck and CIR processes

In this chapter, we will consider the cases where the membrane potential between spikes is given by an Ornstein-Uhlenbeck process (OU) or a Cox-Ingersoll-Ross process (CIR). To perform a maximum likelihood estimation, we need the densities of the level crossing times τ. In fact, Alili, Patie and Pedersen [APP 05] gave some representations for the density of the level crossing time of OU processes. Even much earlier Ricciardi and Sato [RS 88] investigated the asymptotic behavior of this density function and its moments. Also for the CIR process Giorno, Lansky, Nobile, Ricciardi [GLNR 88] computed the moments of τ and investigated its density function numerically. Further, Linetsky [L 04] found an expression for the density of τ. Moreover, Lansky Sacerdote and Tomassetti [LST 95] compared the distribution of τ in the OU case to the distribution of τ in the CIR case, with the result that for some parameter configuration the distributions are very similar.

Unfortunately, all the representations for the densities of τ proposed by the authors mentioned above are not given explicitly in a closed analytic form and too complicated to handle in order to compute the MLE. Only in the case of the threshold regime for the OU process, where $S = a/b$, an explicit expression for the density of the level crossing time τ is known (cf. Keilson and Ross [KR 75]). However, since $S = \theta_2$ is a parameter of our estimation problem, this is not interesting for our purposes.

Anyway, in both cases OU and CIR we know the Laplace transform of the level crossing time given by a ratio of two well-known special functions. The Laplace transform of τ for the OU process was found first by Roy and Smith [RS 69]. Keilson and Ross [KR 75] gave another discussion of the OU first passage time problem. The Laplace transform of τ for the CIR process was given by Göing-Jaeschke and Yor [GY 03] within the consideration of radial Ornstein-Uhlenbeck processes.

In the OU case the Laplace transform of τ is a ratio of two Hermite functions and in the CIR case we have a ratio of two confluent hypergeometric functions. As a matter of fact, in general these functions do not have an explicit closed analytic form either, but their properties are well studied.

Hence we are going to develop an estimation method that only considers the Laplace transform of the level crossing times. Ditlevsen and Lansky, considering the OU case in [DL 05], also used the Laplace transform of τ in order to define moment estimators. In contrast to that, we will apply a minimum distance approach (cf. section 2.2) that compares empirical and true Laplace transform of the level crossing times τ as functions with respect to a Hilbert space norm. Again, we just infer from iid observations $(\tau_1, \dots, \tau_n)$ in order to compute the empirical Laplace transform which leads to the estimator. The results of the following section are formulated in a slightly more general context in order to apply them for both cases OU and CIR, treated in the subsequent sections.

4.1 MDE Based on the Laplace Transform

In this section, we investigate an MDE (cf. section 2.2) that is based on comparing empirical and true Laplace transform of a positive random variable τ. First, we give some notation and a description of the mathematical framework.

Let $\Theta \subset \mathbb{R}^d$ be an open subset and let $(\Omega, \mathcal{A}, \{\mathbf{P}_\theta : \theta \in \Theta\})$ be a statistical experiment. In addition, let $\mathbf{P}_\theta$ be well defined as a probability measure for all $\theta \in \bar{\Theta}$, where $\bar{\Theta}$ denotes the closure of Θ in $\mathbb{R}^d$. In the subsequent sections we will have to assume that Θ is bounded, but for now we omit this assumption. Further, let τ_n, $n \in \mathbb{N}$ be iid copies of a nonnegative random variable τ. If we set $\mathscr{F}_n := \sigma(\tau_1, \dots, \tau_n) \subset \mathcal{A}$ and $\mathbf{P}_{n,\theta} := \mathbf{P}_{\theta|\mathscr{F}_n}$, the restricted measure on $\mathscr{F}_n$, we can define the following sequence of statistical experiments

$$\mathscr{E}_n := (\Omega, \mathscr{F}_n, \{\mathbf{P}_{n,\theta} : \theta \in \Theta\}), \quad n \in \mathbb{N}. \tag{4.1.1}$$

Moreover, for $\alpha \geq 0$ we define

$$\hat{\mathfrak{L}}_n(\alpha) := \frac{1}{n} \sum_{i=1}^{n} e^{-\alpha \tau_i} \qquad \text{and} \qquad \mathfrak{L}_\theta(\alpha) := \mathbf{E}_\theta[e^{-\alpha \tau}]$$

the empirical Laplace transform of τ after n observations and its true Laplace transform under $\mathbf{P}_\theta$ respectively. To apply the MDE-method, we have to introduce a Hilbert space $\mathcal{H}$, in which these two functions will be compared. This will be $\mathcal{H} := L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mu)$, where $\mu \not\equiv 0$ is a finite measure on $\mathbb{R}_+ := [0, \infty)$, with piecewise continuous Lebesgue density h such that

$$\mu(\mathbb{A}) = \int_{\mathbb{A}} h(\alpha)\lambda\!\!\lambda(\mathrm{d}\alpha), \quad \mathbb{A} \in \mathscr{B}(\mathbb{R}_+).$$

We denote scalar product and norm on $\mathcal{H}$ by $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ respectively. Hence in our framework, the minimum distance estimator (MDE) θ_n^*, $n \in \mathbb{N}$ is defined by

$$\theta_n^* := \arg\inf_{\xi\in\Theta} \|\hat{\mathfrak{L}}_n - \mathfrak{L}_\xi\|_{\mathcal{H}},$$

provided this arg inf exists (cf. (2.2.1)).

In the next step, we have to check the assumptions and conditions of section 2.2 to apply the results for MDE. Some of them can be checked without explicit knowledge of the parametrization of the Laplace transform in θ. These are treated in the following lemmas. To verify the remaining conditions we need to specify τ or its Laplace transform respectively. This will be done in the subsequent sections, where τ is assumed to be a level crossing time of an OU or a CIR process. Anyway, in the following we will also give some lemmas that will at least help us to check these remaining conditions.

Lemma 4.1.1 $(\hat{\mathfrak{L}}_n)_{n\in\mathbb{N}}$ *is a sequence of* $\mathcal{H}$*-valued* $\mathscr{F}_n$*-measurable random variables*

$$\hat{\mathfrak{L}}_n : (\Omega, \mathscr{F}_n) \longrightarrow (\mathcal{H}, \mathscr{B}(\mathcal{H})), \quad n \in \mathbb{N}.$$

Proof: Let $n \in \mathbb{N}$ and define the function $G : [0,\infty)^n \longrightarrow \mathcal{H}$,

$$G(x_1,\ldots,x_n;\alpha) := \frac{1}{n}\sum_{i=1}^{n} e^{-\alpha x_i}, \quad \alpha \in \mathbb{R}_+.$$

Since $\alpha, x_1, \ldots, x_n \geq 0$, G is bounded by 1. Thus $G(x_1,\ldots,x_n;\cdot) \in \mathcal{H}$ because μ is a finite measure. Furthermore, $G : [0,\infty)^n \longrightarrow \mathcal{H}$ is continuous because for arbitrary $y \in [0,\infty)^n$ dominated convergence implies that

$$\begin{aligned}
\lim_{x\to y} \|G(x) - G(y)\|_{\mathcal{H}}^2 &= \lim_{x\to y} \int_{\mathbb{R}_+} |G(x_1,\ldots,x_n;\alpha) - G(y_1,\ldots,y_n;\alpha)|^2 \mu(\mathrm{d}\alpha) \\
&= \int_{\mathbb{R}_+} \lim_{x\to y} |G(x_1,\ldots,x_n;\alpha) - G(y_1,\ldots,y_n;\alpha)|^2 \mu(\mathrm{d}\alpha) \\
&= 0
\end{aligned}$$

holds. Hence $G : ([0,\infty)^n, \mathscr{B}([0,\infty)^n)) \longrightarrow (\mathcal{H}, \mathscr{B}(\mathcal{H}))$ is measurable. Obviously, by definition

$$(\tau_1, \dots \tau_n) : (\Omega, \mathscr{F}_n) \longrightarrow ([0,\infty)^n, \mathscr{B}([0,\infty)^n))$$

is measurable. Thus,

$$\hat{\mathfrak{L}}_n = G(\tau_1, \dots \tau_n; \cdot\,) : (\Omega, \mathscr{F}_n) \longrightarrow (\mathcal{H}, \mathscr{B}(\mathcal{H}))$$

is measurable as a composition of measurable functions. □

Lemma 4.1.2 $(\hat{\mathfrak{L}}_n)_{n\in\mathbb{N}}$ *is a family of* $[0,1]$*-valued measurable processes*

$$\hat{\mathfrak{L}}_n : (\mathbb{R}_+ \times \Omega\,,\, \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{F}_n) \longrightarrow ([0,1], \mathscr{B}([0,1])), \quad n \in \mathbb{N}$$

whose paths are continuous and belong to $\mathcal{H}$.

Proof: Let $n \in \mathbb{N}$. Obviously, $\hat{\mathfrak{L}}_n(\alpha) : \Omega \to [0,1]$ is $\mathscr{F}_n$-measurable for every $\alpha \in \mathbb{R}_+$ and $\hat{\mathfrak{L}}_n(\cdot, \omega) : \mathbb{R}_+ \to [0,1]$ is continuous for every $\omega \in \Omega$. Since μ is finite, we consequently have that $\hat{\mathfrak{L}}_n(\cdot,\omega) \in \mathcal{H}$ for all $\omega \in \Omega$. Further, the continuity in α implies that $\hat{\mathfrak{L}}_n$ is a measurable process. □

Lemma 4.1.3 *If* $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta(\alpha) \in [0,1]$ *is continuous for every* $\alpha \in \mathbb{R}_+$, *then the parametrization* $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ *is continuous.*

Proof: Let $\xi \in \Theta$ be arbitrary and $(\xi_n)_{n\in\mathbb{N}} \subset \Theta$ be any sequence such that $\lim_{n\to\infty} \xi_n = \xi$. Hence we have by assumption that

$$\lim_{n\to\infty} \mathfrak{L}_{\xi_n}(\alpha) = \mathfrak{L}_\xi(\alpha)$$

for all $\alpha \in \mathbb{R}_+$. Since τ is nonnegative, $\mathfrak{L}_\xi$ and $\mathfrak{L}_{\xi_n}$, $n \in \mathbb{N}$ are uniformly bounded by 1. As μ is finite, we conclude by dominated convergence that

$$\begin{aligned}
\lim_{n\to\infty} \|\mathfrak{L}_\xi - \mathfrak{L}_{\xi_n}\|_{\mathcal{H}}^2 &= \lim_{n\to\infty} \int_{\mathbb{R}_+} |\mathfrak{L}_\xi(\alpha) - \mathfrak{L}_{\xi_n}(\alpha)|^2 \mu(\mathrm{d}\alpha) \\
&= \int_{\mathbb{R}_+} \lim_{n\to\infty} |\mathfrak{L}_\xi(\alpha) - \mathfrak{L}_{\xi_n}(\alpha)|^2 \mu(\mathrm{d}\alpha) \\
&= 0
\end{aligned}$$

holds. Given that ξ and $(\xi_n)_{n\in\mathbb{N}}$ were arbitrary, this proves the assertion. □

Lemma 4.1.4 *The strong law of large numbers **SLLN(θ)** holds for every $\theta \in \Theta$, i.e.*

$$\|\hat{\mathfrak{L}}_n - \mathfrak{L}_\theta\|_{\mathcal{H}} \xrightarrow[n\to\infty]{\mathbf{P}_\theta - a.s.} 0.$$

Proof: Let $\theta \in \Theta$ be fix. Define $\mathbb{A} \subset \Omega$ by

$$\mathbb{A} := \bigcap_{q \in \mathbb{Q}} \left\{ \omega \in \Omega \;\middle|\; \hat{\mathfrak{L}}_n(q;\omega) \xrightarrow[n\to\infty]{} \mathfrak{L}_\theta(q;\omega) \right\}.$$

Evidently, $\mathbb{A} \in \mathcal{A}$ and by the SLLN for iid random variables $\mathbb{A}$ is an intersection of countably many sets with probability 1. Hence we have $\mathbf{P}_\theta(\mathbb{A}) = 1$.

Let $\omega \in \mathbb{A}$ be arbitrary. Since $\alpha \mapsto \mathfrak{L}_\theta(\alpha)$ and $\alpha \mapsto \hat{\mathfrak{L}}_n(\alpha)$ are continuous, we deduce that

$$\lim_{n\to\infty} \hat{\mathfrak{L}}_n(\alpha;\omega) = \mathfrak{L}_\theta(\alpha;\omega)$$

holds for every $\alpha \in \mathbb{R}_+$. Since τ_i, $i \in \mathbb{N}$ are nonnegative, $\mathfrak{L}_\theta$ and $\hat{\mathfrak{L}}_n$ are bounded by 1. Given that μ is a finite measure, we conclude by dominated convergence that

$$\begin{aligned}
\lim_{n\to\infty} \|\hat{\mathfrak{L}}_n(\,\cdot\,;\omega) - \mathfrak{L}_\theta(\,\cdot\,;\omega)\|_{\mathcal{H}}^2 &= \lim_{n\to\infty} \int_{\mathbb{R}_+} |\hat{\mathfrak{L}}_n(\alpha;\omega) - \mathfrak{L}_\theta(\alpha;\omega)|^2 \mu(\mathrm{d}\alpha) \\
&= \int_{\mathbb{R}_+} \lim_{n\to\infty} |\hat{\mathfrak{L}}_n(\alpha;\omega) - \mathfrak{L}_\theta(\alpha;\omega)|^2 \mu(\mathrm{d}\alpha) \\
&= 0.
\end{aligned}$$

This holds for every $\omega \in \mathbb{A}$ with $\mathbf{P}_\theta(\mathbb{A}) = 1$, which proves the assertion. □

Lemma 4.1.5 *The asymptotic normality condition **AN(θ)** is fulfilled for all $\theta \in \Theta$. More precisely: Let $\theta \in \Theta$ be arbitrary. Define $K(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ a symmetric, continuous and bounded covariance function by*

$$K(\alpha_1, \alpha_2) := \mathfrak{L}_\theta(\alpha_1 + \alpha_2) - \mathfrak{L}_\theta(\alpha_1)\mathfrak{L}_\theta(\alpha_2) = \mathbf{Cov}_\theta\left[e^{-\alpha_1 \tau}, e^{-\alpha_2 \tau}\right].$$

Further, define $W^n_\bullet$, $n \in \mathbb{N}$ by

$$W^n_\alpha := \sqrt{n}\left(\hat{\mathfrak{L}}_n(\alpha) - \mathfrak{L}_\theta(\alpha)\right).$$

Then the following holds:

1) $\mathbf{E}_\theta\left[(W^n_\alpha)^2\right] = K(\alpha, \alpha), \quad \forall n \in \mathbb{N},\ \alpha \in \mathbb{R}_+$

2) For arbitrary $\beta_1, \dots \beta_l \in \mathbb{R}$ and $\alpha_1, \dots \alpha_l \in \mathbb{R}_+$, $l \in \mathbb{N}$,

$$\sum_{k=1}^{l} \beta_k W_{\alpha_k}^n \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\left(0, \sum_{k,j=1}^{l} \beta_k K(\alpha_k, \alpha_j)\beta_j\right).$$

3) Let $W = W(\theta)$ be the Gaussian process corresponding to the covariance function $K(\cdot,\cdot)$, then

$$W^n \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} W \quad \text{in } \mathcal{H}.$$

In particular the family

$$\mathcal{L}\left(\sqrt{n}\left\|\hat{\mathfrak{L}}_n(\alpha) - \mathfrak{L}_\theta(\alpha)\right\|_{\mathcal{H}} \middle| \mathbf{P}_\theta\right), \quad n \in \mathbb{N}$$

is tight in $\mathbb{R}$.

Proof: By definition, $K(\cdot,\cdot)$ is a symmetric covariance function. Since $\mathfrak{L}_\theta(\cdot)$ is a Laplace transform of a nonnegative RV, it is bounded by 1 and $\alpha \mapsto \mathfrak{L}_\theta(\alpha)$ is continuous. Hence $K(\cdot,\cdot)$ is also bounded and continuous. Note that there exists a Gaussian process W with covariance function $K(\cdot,\cdot)$, where the boundedness of K ensures that the paths of W belong to $\mathcal{H}$. This can be found in [L 78] §37.1 Thm. A, §37.5 Thm. B + C. We also refer to [H 0708] (Kap II.C Satz 2.20).

Since $\{\tau_i : i =\in \mathbb{N}\}$ are iid, we show 1) by the computation

$$\begin{aligned}
\mathbf{E}_\theta\left[(W_\alpha^n)^2\right] &= \mathbf{E}_\theta\left[n\left(\frac{1}{n}\sum_{i=1}^{n} e^{-\alpha\tau_i} - \mathbf{E}_\theta[e^{-\alpha\tau}]\right)^2\right] \\
&= \frac{1}{n}\mathbf{Var}_\theta\left[\sum_{i=1}^{n} e^{-\alpha\tau_i}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{Var}_\theta[e^{-\alpha\tau_i}] \\
&= K(\alpha,\alpha).
\end{aligned}$$

To show 2), we define iid RVs

$$A_i := \sum_{k=1}^{l} \beta_k \left(e^{-\alpha_k\tau_i} - \mathbf{E}_\theta[e^{-\alpha_k\tau_i}]\right) \quad i \in \mathbb{N}.$$

By definition, A_i is bounded and we have $\mathbf{E}_\theta[A_i] = 0$ for all $i \in \mathbb{N}$. Further,

$$\mathbf{Var}_\theta[A_i] = \sum_{k,j=1}^{l} \beta_k \beta_j \mathbf{Cov}_\theta[e^{-\alpha_k \tau_i}, e^{-\alpha_j \tau_i}] = \sum_{k,j=1}^{l} \beta_k \beta_j K(\alpha_k, \alpha_j)$$

holds for all $i \in \mathbb{N}$. Applying the central limit theorem for iid RVs, we conclude

$$\begin{aligned}\sum_{k=1}^{l} \beta_k W_{\alpha_k}^n =& \sqrt{n} \sum_{k=1}^{l} \beta_k \left(\frac{1}{n} \sum_{i=1}^{n} e^{-\alpha_k \tau_i} - \mathbf{E}_\theta[e^{-\alpha_k \tau}] \right) \\ =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i \\ & \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\left(0, \sum_{k,j=1}^{l} \beta_k K(\alpha_k, \alpha_j) \beta_j \right).\end{aligned}$$

To show 3) we first prove the weak convergence in $\mathcal{H}$. For this purpose, we will use the criterion by Cremers and Kadelka in theorem 2.2.3. With the Cramér-Wold device, 2) implies the convergence of the marginal distributions, i.e.

$$\left(W_{\alpha_1}^n, \ldots, W_{\alpha_l}^n\right) \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} (W_{\alpha_1}, \ldots, W_{\alpha_l})$$

for arbitrary $l \in \mathbb{N}$ and $\alpha_1, \ldots \alpha_l \in \mathbb{R}_+$. Since K is bounded and μ is finite,

$$\int_{\mathbb{R}_+} K(\alpha, \alpha) \mu(\mathrm{d}\alpha) < \infty$$

holds. Moreover, from 1) and Fubini we have

$$\begin{aligned}\limsup_{n\to\infty} \int |W^n|^2 \mathrm{d}(\mathbf{P}_\theta \otimes \mu) &= \limsup_{n\to\infty} \int_{\mathbb{R}_+} \int_{\Omega} |W^n|^2 \mathrm{d}\mathbf{P}_\theta \mathrm{d}\mu \\ &= \int_{\mathbb{R}_+} K(\alpha, \alpha) \mu(\mathrm{d}\alpha) \\ &= \int |W|^2 \mathrm{d}(\mathbf{P}_\theta \otimes \mu) < \infty.\end{aligned}$$

Hence the conditions of theorem 2.2.3 are fulfilled, which proves the weak convergence of $(W_n)_{n\in\mathbb{N}}$ to W in $\mathcal{H}$. Finally, since the norm $\|\cdot\|_{\mathcal{H}}$ is continuous, the last assertion follows from the continuous mapping theorem.

□

Lemma 4.1.6 *Let $\Theta \subset \mathbb{R}^d$ be bounded and the parametrization $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ be continuous. Moreover, for $\xi \in \bar{\Theta}$ let $\mathfrak{L}_\theta \not\equiv \mathfrak{L}_\xi$ if $\xi \neq \theta$. Then the identifiability condition* ***I(θ)*** *holds for every $\theta \in \Theta$, i.e.*

$$\inf_{\xi \in \Theta, |\theta - \xi| \geq \delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} > 0 \qquad \forall \delta > 0.$$

Proof: Let $\delta > 0$ be arbitrary. Since $\bar{\Theta}$ is compact and $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ is continuous, the continuity of the norm $\|\cdot\|_{\mathcal{H}}$ yields the existence of a $\xi^* \in \{\xi \in \bar{\Theta} : |\xi - \theta| \geq \delta\}$ such that

$$\inf_{\xi \in \bar{\Theta}, |\theta - \xi| \geq \delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} = \|\mathfrak{L}_\theta - \mathfrak{L}_{\xi^*}\|_{\mathcal{H}}. \tag{4.1.2}$$

Define $\mathbb{C}_{>0} := \{z \in \mathbb{C} : \mathrm{Re} z > 0\}$. Since τ is positive, it is well-known that $\mathbb{C}_{>0} \ni \alpha \mapsto \mathfrak{L}_\xi(\alpha)$ is an analytic function for all $\xi \in \Theta$ (cf. e.g. [B 81] Ch.10.1 theorem 2). Further, it is well-known that if two analytic functions on $\mathbb{C}_{>0}$ coincide on an open interval of $\mathbb{R}_+$, they coincide on $\mathbb{C}_{>0}$ (cf. e.g. [B 81] Ch.10.1 theorem 1). Hence by contraposition, the assumptions above imply that on every open subset $\mathbb{A} \subset \mathbb{R}_+$, $\mathfrak{L}_\theta \not\equiv \mathfrak{L}_\xi$. Since μ has a piecewise continuous Lebesgue density, there exists an $\mathbb{A} \subset \mathbb{R}_+$ open such that μ and the Lebesgue measure are equivalent on $\mathbb{A}$. Moreover, there exists an $\tilde{\alpha} \in \mathbb{A}$ such that $\mathfrak{L}_\theta(\tilde{\alpha}) \neq \mathfrak{L}_{\xi^*}(\tilde{\alpha})$. Since $\alpha \mapsto \mathfrak{L}_\xi(\alpha)$ is continuous, there exists an $\varepsilon > 0$ such that

$$\mathfrak{L}_\theta(\alpha) \neq \mathfrak{L}_{\xi^*}(\alpha) \quad \forall \alpha \in \bar{B}_\varepsilon(\tilde{\alpha}) \subset \mathbb{A}.$$

Further,

$$\alpha^* := \arg\min_{\alpha \in \bar{B}_\varepsilon(\tilde{\alpha})} |\mathfrak{L}_\theta(\alpha) - \mathfrak{L}_{\xi^*}(\alpha)|$$

exists. Hence

$$\begin{aligned} \|\mathfrak{L}_\theta - \mathfrak{L}_{\xi^*}\|_{\mathcal{H}}^2 &\geq \int_{\bar{B}_\varepsilon(\tilde{\alpha})} |\mathfrak{L}_\theta(\alpha) - \mathfrak{L}_{\xi^*}(\alpha)|^2 \mu(\mathrm{d}\alpha) \\ &\geq \int_{\bar{B}_\varepsilon(\tilde{\alpha})} |\mathfrak{L}_\theta(\alpha^*) - \mathfrak{L}_{\xi^*}(\alpha^*)|^2 \mu(\mathrm{d}\alpha) \\ &= |\mathfrak{L}_\theta(\alpha^*) - \mathfrak{L}_{\xi^*}(\alpha^*)|^2 \cdot \mu(\bar{B}_\varepsilon(\tilde{\alpha})) \\ &> 0. \end{aligned} \tag{4.1.3}$$

Thus, from (4.1.2) and (4.1.3) together we conclude

$$\inf_{\xi \in \Theta, |\theta - \xi| \geq \delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} \geq \inf_{\xi \in \bar{\Theta}, |\theta - \xi| \geq \delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} > 0$$

which shows the assertion. $\square$

Lemma 4.1.7 *Let* $\mathbb{R}_+ \times \Theta \ni (\alpha; \theta_1, \theta_2, \ldots, \theta_d) \mapsto \mathfrak{L}_\theta(\alpha) \in \mathbb{R}$ *be twice continuously differentiable. Assume further that* μ *has compact support. Then* $\frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathfrak{L}_\theta, \frac{\mathrm{d}^2}{\mathrm{d}\theta_i \mathrm{d}\theta_j}\mathfrak{L}_\theta \in \mathcal{H}$ *for* $i, j \in \{1, \ldots, d\}$ *and the function*

$$\Theta \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$$

is Fréchet-differentiable with derivative

$$D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta, \ldots, \frac{\mathrm{d}}{\mathrm{d}\theta_d}\mathfrak{L}_\theta \right)^\top.$$

Proof: Particularly, we have by assumption that

$$\mathbb{R}_+ \times \Theta \ni (\alpha, \theta) \mapsto \frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathfrak{L}_\theta(\alpha) \in \mathbb{R}, \quad i \in \{1, \ldots, d\}$$

and

$$\mathbb{R}_+ \times \Theta \ni (\alpha, \theta) \mapsto \frac{\mathrm{d}^2}{\mathrm{d}\theta_i \mathrm{d}\theta_j}\mathfrak{L}_\theta(\alpha) \in \mathbb{R}, \quad i, j \in \{1, \ldots, d\} \tag{4.1.4}$$

are continuous functions. Therefore they are locally bounded in α for fixed θ. Since μ is a finite measure with compact support, this leads to

$$\left\| \frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathfrak{L}_\theta \right\|_{\mathcal{H}} < \infty \quad \text{and} \quad \left\| \frac{\mathrm{d}^2}{\mathrm{d}\theta_i \mathrm{d}\theta_j}\mathfrak{L}_\theta \right\|_{\mathcal{H}} < \infty, \quad i, j \in \{1, \ldots, d\}$$

which implies $\frac{\mathrm{d}}{\mathrm{d}\theta_i}\mathfrak{L}_\theta, \frac{\mathrm{d}^2}{\mathrm{d}\theta_i \mathrm{d}\theta_j}\mathfrak{L}_\theta \in \mathcal{H}$, for all $\theta \in \Theta$.

Now let $\theta, \xi \in \Theta$ and let $[\xi, \theta]$ denote the connecting line between ξ and θ. W.l.o.g. we assume $[\xi, \theta] \subset \Theta$. By Taylor expansion, there exists a $\zeta \in [\xi, \theta]$ such that

$$\begin{aligned} \|\mathfrak{L}_\xi - \mathfrak{L}_\theta - (\xi - \theta)^\top D\mathfrak{L}_\theta\|_{\mathcal{H}} &= \left\| \sum_{i,j=1}^{d} \left(\frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j}\mathfrak{L}_\zeta \right) \cdot (\xi_i - \theta_i) \cdot (\xi_j - \theta_j) \right\|_{\mathcal{H}} \\ &\leq \sum_{i,j=1}^{d} \left\| \left(\frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j}\mathfrak{L}_\zeta \right) \right\|_{\mathcal{H}} \cdot |\xi_i - \theta_i||\xi_j - \theta_j| \\ &\leq \sum_{i,j=1}^{d} \sup_{\zeta \in [\xi, \theta]} \left\| \left(\frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j}\mathfrak{L}_\zeta \right) \right\|_{\mathcal{H}} \cdot |\xi_i - \theta_i||\xi_j - \theta_j|. \end{aligned} \tag{4.1.5}$$

Since μ is a finite measure with compact support $\mathbb{A} \subset \mathbb{R}_+$ and the function in (4.1.4) is continuous, we deduce

$$C := \max_{i,j \le d} \sup_{\zeta \in [\xi,\theta], \alpha \in \mathbb{A}} \left| \frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j} \mathfrak{L}_\zeta(\alpha) \right|^2 \cdot \mu(\mathbb{A}) < \infty.$$

Further,

$$\sup_{\zeta \in [\xi,\theta]} \left\| \left(\frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j} \mathfrak{L}_\zeta \right) \right\|_{\mathcal{H}}^2 \le \int_{\mathbb{R}_+} \sup_{\zeta \in [\xi,\theta], \alpha \in \mathbb{A}} \left| \frac{\mathrm{d}^2}{\mathrm{d}\zeta_i \mathrm{d}\zeta_j} \mathfrak{L}_\zeta(\alpha) \right|^2 \mu(\mathrm{d}\alpha) \le C$$

for all $i, j \in \{1, \ldots, d\}$. With (4.1.5) we finally conclude that

$$\frac{\|\mathfrak{L}_\xi - \mathfrak{L}_\theta - (\xi - \theta)^\top D\mathfrak{L}_\theta\|_{\mathcal{H}}}{\|\xi - \theta\|} \le C^{1/2} \sum_{i,j=1}^{d} \frac{|\xi_i - \theta_i| \cdot |\xi_j - \theta_j|}{\|\xi - \theta\|} \quad \xrightarrow{\xi \to \theta} 0$$

and the Fréchet-differentiability of $\theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ is shown. □

4.2 The Ornstein-Uhlenbeck Case

In this section, we assume that the membrane potential between spikes is given by an OU process. To solve the estimation problem described in chapter 1 and recalled in the following, we use an MDE method that compares empirical and true Laplace transform of the inter spike or level crossing time respectively. In section 4.1 we already checked some of the conditions that lead to strong consistency and asymptotic normality for the MDE. To check the remaining conditions, we need to specify the estimation problem and the Laplace transform of the level crossing time of an OU process, defined as follows.

Let $\sigma, b > 0$, $a \in \mathbb{R}$ and $(B_t)_{t \ge 0}$ be a standard Brownian motion. The OU process $X = \big(X_t^{(\theta_1)}\big)_{t \ge 0}$ is defined as the strong solution of the SDE

$$\mathrm{d}X_t = (a - bX_t)\mathrm{d}t + \sigma \mathrm{d}B_t, \quad X_0 = x_0 = \theta_1. \tag{4.2.1}$$

Again, our aim is to estimate the parameters θ_1 and θ_2 from iid observations $(\tau_1, \ldots, \tau_n)$, where

$$\tau = \inf\left\{ t \ge 0 \mid X_t^{(\theta_1)} = \theta_2 \right\} \tag{4.2.2}$$

denotes the level crossing time of the OU process starting in θ_1 and hitting θ_2. Since the OU process is not translation invariant like the Brownian motion with drift, it will turn out that it is possible to estimate θ_1 and θ_2 jointly. This we will prove later by checking the identifiability condition $\mathbf{I}(\theta)$ from the MDE theory (cf. section 2.2). The notions from section 4.1 remain valid, but we further assume for $\mu \not\equiv 0$, a finite measure on $\mathbb{R}_+$ with piecewise continuous Lebesgue density, to have compact support. This is needed to ensure integrability of the derivatives for condition $\mathbf{D}(\theta)$ treated in lemma 4.2.4. The Hilbert space is again $\mathcal{H} := L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mu)$ with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\| \cdot \|_{\mathcal{H}}$. Let $\theta := (\theta_1, \theta_2) \in \Theta$ denote the parameter where the parameter space Θ is an arbitrary subset

$$\Theta \subset \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 < \theta_2\} \text{ open and bounded.}$$

For application the boundedness of Θ is no restriction. However, we need this restriction for the identifiability condition $\mathbf{I}(\theta)$ treated in lemma 4.2.3 to control the boundaries of the parameter space. Note that for $\theta_1 = \theta_2$ we have $\mathcal{L}(\tau | \mathbf{P}_\theta) = \delta_0$ the Dirac measure at 0. Hence $\mathbf{P}_\theta$ is well defined for all $\theta \in \bar{\Theta}$. Let us recall the definition of the minimum distance estimator (MDE) θ_n^*, $n \in \mathbb{N}$,

$$\theta_n^* := \arg \inf_{\xi \in \Theta} \| \hat{\mathfrak{L}}_n - \mathfrak{L}_\xi \|_{\mathcal{H}} \tag{4.2.3}$$

(cf. section 4.1) which takes the value θ_n^* if the "true" Laplace transform corresponding to $\mathbf{P}_{\theta_n^*}$ fits the empirical Laplace transform of τ after n observations best in $\mathcal{H}$.

The Laplace transform of τ is well-known and was found first by Roy and Smith [RS 69]. The next lemma will specify the Laplace transform of τ and give references and explanatory statements instead of a detailed proof.

Lemma 4.2.1 *Let $\theta_1 \leq \theta_2$. The level-crossing time τ from θ_1 to θ_2 of an OU process with parameters a, b, σ as in* (4.2.1) *has the Laplace transform*

$$\mathfrak{L}_\theta(\alpha) := \mathbf{E}_\theta[e^{-\alpha\tau}] = \frac{H_{-\alpha/b}\left(-\left(\theta_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)}{H_{-\alpha/b}\left(-\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)}, \quad \alpha \geq 0,$$

where H is the Hermite function, defined in A.2.1.

Proof: We can rewrite (4.2.1) as

$$\mathrm{d}\tilde{X}_t = -b\tilde{X}_t \mathrm{d}t + \sigma \mathrm{d}B_t, \quad \tilde{X}_0 = \tilde{x}_0$$

where $\tilde{X}_t := X_t - a/b$ and $\tilde{x}_0 := \theta_1 - a/b$. The solution of this SDE is given by

$$\tilde{X}_t = e^{-bt} \cdot \left(\tilde{x}_0 + \sigma \int_0^t e^{bs} \mathrm{d}B_s \right),$$

which can be proved by Itô's formula (cf. e.g. [K 06] 26.5). Since $\int_0^t e^{bs}\mathrm{d}B_s$ is a continuous local martingale, by the theorem of Dubins and Schwarz we have the representation

$$\tilde{X}_t = e^{-bt} \cdot \left(\tilde{x}_0 + \frac{\sigma}{\sqrt{2b}} W_{(e^{2bt}-1)}\right),$$

for a standard Brownian motion W. If we rewrite

$$\tau = \inf\{t > 0 \mid \tilde{X}_t = \theta_2 - a/b\},$$

[BS 02] II.7.2 p.542 gives the following expression for its Laplace transform

$$\mathbf{E}[e^{-\alpha\tau}] = \frac{\exp\left(\frac{2b\tilde{x}_0^2}{4\sigma^2}\right) D_{-\alpha/b}\left(-\tilde{x}_0\frac{\sqrt{2b}}{\sigma}\right)}{\exp\left(\frac{2b(\theta_2-a/b)^2}{4\sigma^2}\right) D_{-\alpha/b}\left(-(\theta_2 - a/b)\frac{\sqrt{2b}}{\sigma}\right)}, \quad \alpha \geq 0,$$

where D is the parabolic cylinder function given in A.2.2. Further, from A.2.2 we know that the parabolic cylinder function and the Hermite function are connected by

$$\exp(z^2/4)D_\nu(z) = 2^{-\nu/2}H_\nu(z/\sqrt{2}),$$

for all $z, \nu \in \mathbb{C}$ and since $\tilde{x}_0 = \theta_1 - a/b$, this implies the assertion. □

Lemma 4.2.2 *The parametrization $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ is continuous.*

Proof: Lemma 4.2.1 and (A.2.1) even show that

$$\mathbb{R}^2 \ni \theta \mapsto \mathfrak{L}_\theta(\alpha) \in [0, 1]$$

is analytic for all $\alpha \in \mathbb{R}_+$. Hence, the assertion follows immediately from lemma 4.1.3. □

Lemma 4.2.3 *The identifiability condition **I(θ)** holds for every $\theta \in \Theta$, i.e.*

$$\inf_{\xi\in\Theta, |\theta-\xi|\geq\delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} > 0 \qquad \forall\delta > 0.$$

Proof: Fix $\theta \in \Theta$. To apply lemma 4.1.6, we only need the boundedness of Θ, Lemma 4.2.2 and the following:

$$\text{Let } \xi \in \bar{\Theta} \text{ and } \xi \neq \theta, \text{ then } \mathfrak{L}_\theta \not\equiv \mathfrak{L}_\xi. \tag{4.2.4}$$

Hence, we only have to show (4.2.4), then lemma 4.1.6 completes the proof. So let $\xi \in \bar{\Theta}$ and $\xi \neq \theta$. Define $z_1, z_2, z_3, z_4 \in \mathbb{R}$ by

$$z_1 := -\left(\theta_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}, \quad z_2 := -\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}$$

and

$$z_3 := -\left(\xi_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}, \quad z_4 := -\left(\xi_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}.$$

Suppose $\mathfrak{L}_\theta \equiv \mathfrak{L}_\xi$, then for every $\nu := -\alpha/b \leq 0$

$$\frac{H_\nu(z_1)}{H_\nu(z_2)} = \frac{H_\nu(z_3)}{H_\nu(z_4)}$$

holds. The asymptotic behavior of H in ν described in (A.2.3) implies

$$\begin{aligned} 1 &= \lim_{\nu \to -\infty} \frac{H_\nu(z_1)H_\nu(z_4)}{H_\nu(z_2)H_\nu(z_3)} \\ &= \lim_{\nu \to -\infty} \exp\left[\frac{z_1^2}{2} - \frac{z_2^2}{2} + \frac{z_4^2}{2} - \frac{z_3^2}{2} - (z_1 - z_2 + z_4 - z_3)\sqrt{-2\nu}\right]. \end{aligned}$$

However, this only holds if

$$z_1 - z_2 + z_4 - z_3 = 0 \qquad \text{and} \qquad z_1^2 - z_2^2 + z_4^2 - z_3^2 = 0. \tag{4.2.5}$$

Since $\xi \neq \theta$, (4.2.5) implies that $z_1 \neq z_3$ and $z_2 \neq z_4$ and further $z_1 \neq z_2$ because $\theta \in \Theta$. Now we consider two cases, $z_1 = -z_3$ and $z_1 \neq -z_3$.

$z_1 = -z_3$: The right side of (4.2.5) implies that $z_2^2 = z_4^2$ and hence $z_2 = -z_4$. If we put this into the left side of (4.2.5) we have $2z_1 - 2z_2 = 0$, but from the definition of z_1 and z_2, this is contradictory to the fact that $\theta_1 < \theta_2$.

$z_1 \neq -z_3$: Equations (4.2.5) give

$$\frac{z_2 - z_4}{z_1 - z_3} = 1 = \frac{z_2^2 - z_4^2}{z_1^2 - z_3^2} = \frac{(z_2 - z_4)(z_2 + z_4)}{(z_1 - z_3)(z_1 + z_3)} = \frac{z_2 + z_4}{z_1 + z_3}.$$

Hence

$$z_1 + z_3 - z_2 - z_4 = 0.$$

If we add the left equation of (4.2.5) to the equation above, we get $2z_1 - 2z_2 = 0$ which is again contradictory to the fact that $\theta_1 < \theta_2$.

Thus both cases expose the assumption $\mathfrak{L}_\theta \equiv \mathfrak{L}_\xi$ to be wrong and assertion (4.2.4) is proved. $\square$

Lemma 4.2.4 *The differentiability condition* $\boldsymbol{D(\theta)}$ *is fulfilled for every* $\theta \in \Theta$. *More precisely, the function* $\theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ *is Fréchet-differentiable in every* $\theta = (\theta_1, \theta_2) \in \Theta$ *with derivative* $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$ *given by*

$$\begin{pmatrix} \dfrac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta(\alpha) \\ \dfrac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta(\alpha) \end{pmatrix} = \frac{2\alpha}{\sigma\sqrt{b}} \cdot \begin{pmatrix} \dfrac{H_{-(\alpha/b+1)}\left(-\left(\theta_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)}{H_{-\alpha/b}\left(-\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)} \\ \dfrac{H_{-(\alpha/b+1)}\left(-\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right) H_{-\alpha/b}\left(-\left(\theta_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)}{-\left[H_{-\alpha/b}\left(-\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)\right]^2} \end{pmatrix}$$

where H *is the Hermite function. Further, the partial derivatives* $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ *and* $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ *are linearly independent in* $\mathcal{H}$.

Proof: 1) From (A.2.1) we know that $\mathbb{C}^2 \ni (\nu, z) \mapsto H_\nu(z)$ is an entire function and further $H_\nu(z) > 0$ for all $\nu \le 0$ and $z \in \mathbb{R}$. Hence the function $\mathfrak{L} : \mathbb{R}^+ \times \Theta \longrightarrow \mathbb{R}$,

$$(\alpha, \theta_1, \theta_2) \mapsto \mathfrak{L}_\theta(\alpha) = \frac{H_{-\alpha/b}\left(-\left(\theta_1 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)}{H_{-\alpha/b}\left(-\left(\theta_2 - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)} \tag{4.2.6}$$

computed in lemma 4.2.1 is a composition of C^∞ functions in each component and therefore C^∞ itself. Since μ has compact support, lemma 4.1.7 shows that for all $\theta \in \Theta$, $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta \in \mathcal{H}$ and

$$\Theta \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$$

is Fréchet-differentiable with derivative $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$.

Now we have to compute the partial derivatives of $\theta \mapsto \mathfrak{L}_\theta$. We point out the derivation rule (A.2.4),

$$\frac{\mathrm{d}}{\mathrm{d}z}H_\nu(z) = 2\nu H_{\nu-1}(z)$$

and define

$$\tilde{H}_\alpha(z) := H_{-\alpha/b}\left(-\left(z - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)$$

for $z \in \mathbb{R}$ and $\alpha \in \mathbb{R}_+$. Hence,

$$\tilde{H}'_\alpha(z) := \frac{\mathrm{d}}{\mathrm{d}z}\tilde{H}_\alpha(z) = \frac{2\alpha}{\sigma\sqrt{b}} H_{-\alpha/b-1}\left(-\left(z - \frac{a}{b}\right)\frac{\sqrt{b}}{\sigma}\right)$$

holds. Note from (4.2.6) that $\mathfrak{L}_\theta(\alpha) = \dfrac{\tilde{H}_\alpha(\theta_1)}{\tilde{H}_\alpha(\theta_2)}$. Evidently, we only have to use the chain rule to get the partial derivatives of $\mathfrak{L}_\theta$ given by

$$\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta = \frac{\tilde{H}'_\alpha(\theta_1)}{\tilde{H}_\alpha(\theta_2)}$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta = -\frac{\tilde{H}_\alpha(\theta_1)\cdot\tilde{H}'_\alpha(\theta_2)}{\left(\tilde{H}_\alpha(\theta_2)\right)^2}$$

Note that $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$ has the form stated in the assertion.
2) It remains to show the linear independence of $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ and $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ in $\mathcal{H}$. Suppose that they are linearly dependent for some $\theta \in \Theta$. Then there exists a constant $c \in \mathbb{R}\setminus\{0\}$ such that

$$\left\|\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta + c\cdot\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right\|_{\mathcal{H}} = 0. \tag{4.2.7}$$

Define $F \in \mathcal{H}$ by

$$F(\alpha) := \frac{\tilde{H}'_\alpha(\theta_2)}{\tilde{H}_\alpha(\theta_2)}.$$

Since

$$\mathfrak{L}_\theta(\alpha+b) = \frac{\tilde{H}_{\alpha+b}(\theta_1)}{\tilde{H}_{\alpha+b}(\theta_2)} = \frac{\tilde{H}'_\alpha(\theta_1)}{\tilde{H}'_\alpha(\theta_2)}, \qquad \text{for all } \alpha \in \mathbb{R}_+,$$

we deduce

$$\begin{aligned}\int_{\mathbb{R}_+} F^2(\alpha)\cdot\left|\mathfrak{L}_\theta(\alpha+b) - c\cdot\mathfrak{L}_\theta(\alpha)\right|^2\mu(\mathrm{d}\alpha) &= \left\|F\cdot\{\mathfrak{L}_\theta(\cdot+b) - c\cdot\mathfrak{L}_\theta\}\right\|^2_{\mathcal{H}} \\ &= \left\|F\cdot\left\{\frac{\tilde{H}'_\cdot(\theta_1)}{\tilde{H}'_\cdot(\theta_2)} - c\cdot\frac{\tilde{H}_\cdot(\theta_1)}{\tilde{H}_\cdot(\theta_2)}\right\}\right\|^2_{\mathcal{H}} \\ &= \left\|\frac{\tilde{H}'_\cdot(\theta_1)}{\tilde{H}_\cdot(\theta_2)} - c\cdot\frac{\tilde{H}_\cdot(\theta_1)\cdot\tilde{H}'_\cdot(\theta_2)}{\left(\tilde{H}_\cdot(\theta_2)\right)^2}\right\|^2_{\mathcal{H}} \\ &= \left\|\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta + c\cdot\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right\|^2_{\mathcal{H}}\end{aligned}$$

such that (4.2.7) implies

$$\int_{\mathbb{R}_+} F^2(\alpha) \cdot |\mathfrak{L}_\theta(\alpha+b) - c \cdot \mathfrak{L}_\theta(\alpha)|^2 \, \mu(\mathrm{d}\alpha) = 0.$$

From this it follows that

$$F^2(\alpha) \cdot |\mathfrak{L}_\theta(\alpha+b) - c \cdot \mathfrak{L}_\theta(\alpha)|^2 = 0, \quad \text{for } \mu \text{ almost every } \alpha \in \mathbb{R}_+.$$

Further, $F(\alpha) > 0$ holds for all $\alpha \geq 0$. This is due to the fact that $H_\nu(z) > 0$ for all $z \in \mathbb{R}$, $\nu \leq 0$ (cf. (A.2.1)). Hence

$$|\mathfrak{L}_\theta(\alpha+b) - c \cdot \mathfrak{L}_\theta(\alpha)| = 0, \quad \text{for } \mu \text{ almost every } \alpha \in \mathbb{R}_+.$$

Since $\mathbb{R}_+ \ni \alpha \mapsto \mathfrak{L}_\theta(\alpha) \in [0,1]$ is continuous and μ has a piecewise continuous Lebesgue density, there exists at least one open interval $I \subset \mathbb{R}_+$ such that

$$\mathfrak{L}_\theta(\alpha+b) = c \cdot \mathfrak{L}_\theta(\alpha) \quad \forall \alpha \in I.$$

However, if two analytic functions on $\mathbb{C}_{>0} := \{z \in \mathbb{C} : \mathrm{Re} z > 0\}$ coincide on an open interval of $\mathbb{R}_+$, they coincide on $\mathbb{C}_{>0}$ (see e.g. [B 81] Ch.10.1 Thm.1, Thm.2). Therefore

$$\mathfrak{L}_\theta(\alpha+b) = c \cdot \mathfrak{L}_\theta(\alpha) \quad \forall \alpha \in \mathbb{R}_+$$

holds. In addition, the differentiability allows to apply an inversion of the Laplace transform (cf. [F 66] Ch. VII.6, p.232) which yields

$$e^{-bt} \cdot f^\tau(t;\theta) = c \cdot f^\tau(t;\theta), \quad t > 0,$$

where $f^\tau(\cdot\,;\theta)$ is the Lebesgue density of $\mathcal{L}(\tau|\mathbf{P}_\theta)$, $\theta \in \Theta$. However, since $b > 0$ this is contradictory to the fact that c is a constant. So we conclude that $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ and $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ must be linearly independent. $\square$

Now we come to the main result of this section.

Theorem 4.2.5 *Let* θ_n^*, $n \in \mathbb{N}$ *be any MDE defined by* (4.2.3), $\mathfrak{L}_\theta$ *as given in 4.2.1 and let*

$$D_1\mathfrak{L}_\theta := \frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta \quad \textit{and} \quad D_2\mathfrak{L}_\theta := \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$$

be as in lemma 4.2.4. Further, define $U_\theta, V_\theta \in \mathbb{R}^{2\times 2}$ *by*

$$(U_\theta)_{i,j} := \langle D_i\mathfrak{L}_\theta, D_j\mathfrak{L}_\theta\rangle_{\mathcal{H}}, \quad 1 \leq i,j \leq 2$$

and

$$(V_\theta)_{i,j} := \int_0^\infty \int_0^\infty D_i \mathfrak{L}_\theta(\alpha_1) K(\alpha_1, \alpha_2) D_j \mathfrak{L}_\theta(\alpha_2) \mu(\mathrm{d}\alpha_1)\mu(\mathrm{d}\alpha_2), \quad 1 \le i, j \le 2,$$

where $K(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ *is given by*

$$K(\alpha_1, \alpha_2) := \mathfrak{L}_\theta(\alpha_1 + \alpha_2) - \mathfrak{L}_\theta(\alpha_1)\mathfrak{L}_\theta(\alpha_2).$$

Then $(\theta_n^*)_{n\in\mathbb{N}}$ *is strongly consistent and asymptotically normal, i.e.*

$$\sqrt{n}\,(\theta_n^* - \theta) \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\left(0, U_\theta^{-1} V_\theta U_\theta^{-1}\right).$$

Proof: Since we are in the setting of section 4.1, lemmas 4.1.1, 4.1.2, 4.1.4 and 4.1.5 remain valid. Together with lemmas 4.2.2, 4.2.3 and 4.2.4, all assumptions and conditions of section 2.2 are fulfilled. Finally, we can apply theorem 2.2.4 which completes the proof. □

Remark: Evidently, all results of this section remain valid for the one-dimensional sub-experiments, i.e. for unknown θ_1 the parameter space is given by

$$\Theta_1 := (-\infty, y),$$

where y is any known upper level. For unknown θ_2 the parameter space is given by

$$\Theta_2 := (y, \infty),$$

where y is any known lower level. All proofs are literally the same or even easier.

4.3 The Cox-Ingersoll-Ross Case

We now come to the last case considered in this work, where the membrane potential between spikes is assumed to be a Cox-Ingersoll-Ross (CIR) process. The strategy of this section is the same as in the previous section and some proofs are very similar. All notations introduced in section 4.2 remain valid for this section. The only difference is that X now denotes a CIR process defined as follows.

Let $a, b, \sigma > 0$ and $(B_t)_{t\ge0}$ be a standard Brownian motion. The CIR process $X = \left(X_t^{(\theta_1)}\right)_{t\ge0}$ is defined as the strong solution of the SDE

$$\mathrm{d}X_t = (a - bX_t)\mathrm{d}t + \sigma\sqrt{X_t^+}\mathrm{d}B_t, \quad X_0 = x_0 = \theta_1 \ge 0, \tag{4.3.1}$$

where $X^+ = \max\{0, X\}$ denotes the positive part of X. In fact, X is known to be nonnegative (cf. [IW 89] Ch.4 example 8.2), so we may omit the superscript $+$. In biological context we have a "more general" situation such that for some $c \leq 0$ the membrane potential X is the solution of

$$\mathrm{d}X_t = (a - b(X_t - c))\mathrm{d}t + \sigma\sqrt{(X_t - c)^+}\mathrm{d}B_t, \quad X_0 = x_0 \geq c. \tag{4.3.2}$$

But this is an easy transformation by adding c, so we only consider X as a solution of (4.3.1).

To solve our estimation problem described in chapter 1, we use again the MDE-method in order to estimate θ_1 and θ_2 from iid observations $(\tau_1, \ldots, \tau_n)$, where

$$\tau = \inf\left\{t \geq 0 \mid X_t^{(\theta_1)} = \theta_2\right\} \tag{4.3.3}$$

now denotes the level crossing time of a CIR process starting in θ_1 and hitting θ_2. Like the OU process, the CIR process is not translation invariant. So also in this case it is possible to estimate θ_1 and θ_2 jointly. Now we recall some notation already used in the previous sections 4.1 and 4.2. Let $\mu \not\equiv 0$ be a finite measure on $\mathbb{R}_+$ with piecewise continuous Lebesgue density and compact support and $\mathcal{H} := L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mu)$ with scalar product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$. Further, let $\theta := (\theta_1, \theta_2) \in \Theta$ denote the parameter where the parameter space Θ is an arbitrary subset

$$\Theta \subset \{(\theta_1, \theta_2) \in [0, \infty)^2 : \theta_1 < \theta_2\} \text{ open and bounded.}$$

As in the previous sections the MDE method compares empirical and true Laplace transform of τ such that the minimum distance estimator (MDE) θ_n^*, $n \in \mathbb{N}$ is defined by

$$\theta_n^* := \arg\inf_{\xi\in\Theta} \|\hat{\mathfrak{L}}_n - \mathfrak{L}_\xi\|_{\mathcal{H}} \tag{4.3.4}$$

(cf. section 4.1 and 4.2). Hence, we are again in the setting of section 4.1. There we already checked some of the conditions that lead to strong consistency and asymptotic normality for the MDE. To check the remaining conditions we again have to specify the Laplace transform of τ, now the level-crossing time of a CIR process. This Laplace transform was found by Göing-Jaeschke and Yor [GY 03] and is given as follows.

Lemma 4.3.1 *Let $0 \leq \theta_1 \leq \theta_2$. The level crossing time τ from θ_1 to θ_2 defined in* (4.3.3) *of a CIR Process represented by* (4.3.1) *has the Laplace transform*

$$\mathfrak{L}_\theta(\alpha) := \mathbf{E}_\theta[e^{-\alpha\tau}] = \frac{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)}, \quad \alpha \geq 0,$$

where ϕ is the confluent hypergeometric function defined in A.1.1.

To prove the lemma we will use the results of Göing-Jaeschke and Yor in [GY 03], formulated in the setting of radial Ornstein-Uhlenbeck processes. To link their framework to ours, it is necessary to recall the definition of the radial Ornstein-Uhlenbeck process and the squared radial Ornstein-Uhlenbeck process respectively. Accordingly, we will show how they are connected to the CIR process.

Definition 4.3.2 *Let $\delta \geq 0$, $\lambda \in \mathbb{R}$ and $(\hat{B}_t)_{t\geq 0}$ be a standard Brownian motion. The strong solution $Y = (Y_t)_{t\geq 0}$ of the SDE*

$$\mathrm{d}Y_t = (\delta - 2\lambda Y_t)\,\mathrm{d}t + 2\sqrt{Y_t^+}\mathrm{d}\hat{B}_t, \quad Y_0 = y_0 \geq 0, \tag{4.3.5}$$

is called squared radial Ornstein-Uhlenbeck process *with parameters δ and λ. In fact this strong solution exists (cf. [RY 91] Chapter IX §3). Since Y is known to be a nonnegative Markov process, the square root of Y is a well defined Markov process and $R = (R_t)_{t\geq 0}$ defined by*

$$R_t := \sqrt{Y_t}, \quad t \geq 0$$

is called radial Ornstein-Uhlenbeck process *with parameters δ and λ.*

Remark: Originally, the squared radial Ornstein-Uhlenbeck process Y defined above simply comes from the sum of $\delta \in \mathbb{N}$ squared and independent OU processes X^i, $i = 1, \ldots, \delta$ with drift parameter λ given by

$$\mathrm{d}X_t^i = -\lambda X_t^i \mathrm{d}t + \mathrm{d}B_t^i, \quad t \geq 0.$$

One only has to apply Itô's formula to the following

$$Y_t = \sum_{i=1}^{\delta} \left(X_t^i\right)^2 = \|(X_t^1, \ldots, X_t^\delta)\|^2, \quad t \geq 0$$

and SDE (4.3.5) emerges. For this reason, δ is also called "dimension" parameter of Y. Evidently,

$$R_t = \|(X_t^1, \ldots, X_t^\delta)\|, \quad t \geq 0$$

yields the radial Ornstein-Uhlenbeck process R in dimension δ. Definition 4.3.2 is just a generalization for arbitrary $\delta \geq 0$.

Lemma 4.3.3 *Let $X = (X_t)_{t\geq 0}$ be a CIR process with parameters a, b and σ, starting in x_0 as defined (4.3.1). Moreover, set $g(t) := \frac{4}{\sigma^2}t$. Then $Y = (Y_t)_{t\geq 0}$ defined by*

$$Y_t := X_{g(t)} \quad t \geq 0$$

is a squared radial Ornstein-Uhlenbeck process with parameters $\delta := \frac{4a}{\sigma^2}$ and $\lambda := \frac{2b}{\sigma^2}$, starting in $Y_0 = x_0$. In other words, the CIR process is connected to a squared radial Ornstein-Uhlenbeck process just by a linear transformation in time.

Proof: Consider the Itô-integral term of

$$X_{g(t)} = x_0 + \int_0^{g(t)} (a - bX_s)\mathrm{d}s + \sigma \int_0^{g(t)} \sqrt{X_s}\mathrm{d}B_s.$$

Obviously, $\hat{B}$ defined by $\hat{B}_t := \frac{\sigma}{2}B_{g(t)}$ is again a standard Brownian motion. We will show now that

$$\sigma \int_0^{g(t)} \sqrt{X_s}\mathrm{d}B_s = 2\int_0^t \sqrt{X_{g(s)}}\mathrm{d}\hat{B}_t \quad \text{for all } t \geq 0 \tag{4.3.6}$$

holds. Since both expressions in (4.3.6) are continuous local martingales, (4.3.6) is a direct consequence of the coincidence of their quadratic variation processes given by

$$\begin{aligned} \left\langle \sigma \int \sqrt{X_s}\mathrm{d}B_s \right\rangle_{g(t)} &= \sigma^2 \int_0^{g(t)} X_s \mathrm{d}s \\ &= \sigma^2 \int_0^t X_{g(s)} g'(s)\mathrm{d}s \\ &= \left\langle 2\int \sqrt{X_{g(s)}}\mathrm{d}\hat{B}_s \right\rangle_t. \end{aligned}$$

Further, with $\delta = \frac{4a}{\sigma^2}$ and $\lambda = \frac{2b}{\sigma^2}$ we have from an ordinary integration by substitution that

$$\int_0^{g(t)} (a - bX_s)\mathrm{d}s = \int_0^t g'(s)\left(a - bX_{g(s)}\right)\mathrm{d}s = \int_0^t \left(\delta - 2\lambda X_{g(s)}\right)\mathrm{d}s. \tag{4.3.7}$$

Since Y was defined by $Y_t = X_{g(t)}$, $t \geq 0$, we conclude from (4.3.6) and (4.3.7) that

$$\begin{aligned} Y_t &= x_0 + \int_0^{g(t)} (a - bX_s)\mathrm{d}s + \sigma \int_0^{g(t)} \sqrt{X_s}\mathrm{d}B_s. \\ &= x_0 + \int_0^t (\delta - 2\lambda Y_s)\,\mathrm{d}s + 2\int_0^t \sqrt{Y_s}\mathrm{d}\hat{B}_s \end{aligned}$$

holds. Thus, we have identified Y as a squared radial Ornstein-Uhlenbeck process starting in x_0. □

Proof of Lemma 4.3.1: To proof the lemma, we only have to transform the level crossing time of a CIR process into a level crossing time of a radial Ornstein-Uhlenbeck process such that we are able to apply the results of Göing-Jaeschke and Yor [GY 03].

Let X be a CIR process as in (4.3.1) with parameters a, b and σ, starting in $X_0 = \theta_1$. Further, set $g(t) = \frac{4}{\sigma^2}t$. From lemma 4.3.3 we know that $Y = (Y_t)_{t \geq 0}$ defined by

$$Y_t := X_{g(t)} \quad t \geq 0$$

is a squared radial Ornstein-Uhlenbeck process with parameters $\delta := \frac{4a}{\sigma^2}$ and $\lambda := \frac{2b}{\sigma^2}$ starting in θ_1. Hence $R = (R_t)_{t \geq 0}$ defined by $R_t := \sqrt{Y_t}$ is a radial Ornstein-Uhlenbeck process with parameters δ and λ starting in $\sqrt{\theta_1}$. Further, we define the level-crossing time of a radial Ornstein-Uhlenbeck process starting in $\sqrt{\theta_1}$ hitting the level $\sqrt{\theta_2}$ by

$$\tilde{\tau} := \inf \left\{ t \geq 0 \mid R_t^{(\sqrt{\theta_1})} = \sqrt{\theta_2} \right\}$$

whose Laplace transform was investigated in [GY 03]. Since $g : \mathbb{R} \to \mathbb{R}$ is linear and strictly increasing, we have the following connection between τ and $\tilde{\tau}$,

$$\begin{aligned}
\tau &= \inf \left\{ t \geq 0 \mid X_t^{(\theta_1)} = \theta_2 \right\} \\
&= \inf \left\{ t \geq 0 \mid Y_{g^{-1}(t)}^{(\theta_1)} = \theta_2 \right\} \\
&= g \left(\inf \left\{ s \geq 0 \mid Y_s^{(\theta_1)} = \theta_2 \right\} \right) \qquad (4.3.8) \\
&= g \left(\inf \left\{ s \geq 0 \mid R_s^{(\sqrt{\theta_1})} = \sqrt{\theta_2} \right\} \right) \\
&= \frac{4}{\sigma^2} \tilde{\tau}.
\end{aligned}$$

Theorem 1 and corollary 3 of [GY 03] with $\nu := \delta/2 - 1 > -1$ yield the Laplace transform of $\tilde{\tau}$

$$\mathbf{E}_\theta \left[e^{-\alpha \tilde{\tau}} \right] = \frac{\phi\left(\frac{\alpha}{2\lambda}, \nu + 1; \lambda\theta_1 \right)}{\phi\left(\frac{\alpha}{2\lambda}, \nu + 1; \lambda\theta_2 \right)}.$$

Finally, since $\delta = \frac{4a}{\sigma^2}$ and $\lambda = \frac{2b}{\sigma^2}$ we conclude by (4.3.8) that

$$\mathbf{E}_\theta[e^{-\alpha\tau}] = \mathbf{E}_\theta\left[e^{-\frac{4\alpha}{\sigma^2}\tilde{\tau}}\right] = \frac{\phi\left(\frac{4\alpha}{\sigma^2 2\lambda}, \nu+1; \lambda\theta_1\right)}{\phi\left(\frac{4\alpha}{\sigma^2 2\lambda}, \nu+1; \lambda\theta_2\right)} = \frac{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)}$$

which proves the assertion of lemma 4.3.1. $\square$

Lemma 4.3.4 *The parametrization $\bar{\Theta} \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ is continuous.*

Proof: Lemma 4.3.1 and (A.1.1) even show that for every fixed $\alpha \in \mathbb{R}_+$, the function

$$\theta \mapsto \mathfrak{L}_\theta(\alpha)$$

is analytic at every $\theta \in [0,\infty)^2$. Hence the assertion follows immediately from lemma 4.1.3.

$\square$

Lemma 4.3.5 *The identifiability condition **I(θ)** holds for every $\theta \in \Theta$, i.e.*

$$\inf_{\xi\in\Theta, |\theta-\xi|\geq\delta} \|\mathfrak{L}_\theta - \mathfrak{L}_\xi\|_{\mathcal{H}} > 0 \qquad \forall\delta > 0.$$

Proof: The strategy of this proof is the same as in the proof of 4.2.3. Fix $\theta \in \Theta$. Again, we only have to show the following to apply lemma 4.1.6.

$$\text{Let } \xi \in \bar{\Theta} \text{ and } \xi \neq \theta, \text{ then } \mathfrak{L}_\theta \not\equiv \mathfrak{L}_\xi. \tag{4.3.9}$$

For this purpose, let $\xi \in \bar{\Theta}$ and $\xi \neq \theta$. Define $\gamma := \frac{2a}{\sigma^2} > 0$ and $z_1, z_2, z_3, z_4 \geq 0$ by

$$z_1 := \frac{2b}{\sigma^2}\theta_1, \quad z_2 := \frac{2b}{\sigma^2}\theta_2$$

and

$$z_3 := \frac{2b}{\sigma^2}\xi_1, \quad z_4 := \frac{2b}{\sigma^2}\xi_2.$$

Suppose $\mathfrak{L}_\theta \equiv \mathfrak{L}_\xi$. Since $b > 0$, we deduce that for every $\alpha \in \mathbb{R}_+$

$$\frac{\phi(\alpha,\gamma,z_1)}{\phi(\alpha,\gamma,z_2)} = \frac{\phi(\alpha,\gamma,z_3)}{\phi(\alpha,\gamma,z_4)} \tag{4.3.10}$$

holds. By definition A.1.1 $\phi(\gamma, \gamma, z) = e^z$ holds for every γ and z. Hence for $\alpha = \gamma$ we have $1 = \exp(z_1 - z_2 - z_3 + z_4)$ which implies

$$z_1 - z_2 - z_3 + z_4 = 0. \tag{4.3.11}$$

Equation (4.3.10) together with the asymptotic property (A.1.2) imply

$$\begin{aligned}
1 &= \lim_{\alpha\to\infty} \frac{\phi(\alpha,\gamma,z_1)\phi(\alpha,\gamma,z_4)}{\phi(\alpha,\gamma,z_2)\phi(\alpha,\gamma,z_3)} \\
&= \lim_{\alpha\to\infty} \frac{(z_1\alpha)^{1/4-\gamma/2}\exp\left(z_1/2 + 2\sqrt{z_1\alpha}\right)(z_4\alpha)^{1/4-\gamma/2}\exp\left(z_4/2 + 2\sqrt{z_4\alpha}\right)}{(z_2\alpha)^{1/4-\gamma/2}\exp\left(z_2/2 + 2\sqrt{z_2\alpha}\right)(z_3\alpha)^{1/4-\gamma/2}\exp\left(z_3/2 + 2\sqrt{z_3\alpha}\right)} \\
&= \left(\frac{z_1 z_4}{z_2 z_3}\right)^{1/4-\gamma/2} \cdot \exp\left(\frac{z_1 - z_2 - z_3 + z_4}{2}\right) \\
&\quad \cdot \lim_{\alpha\to\infty} \exp\left(2\sqrt{\alpha}\left(\sqrt{z_1} - \sqrt{z_2} - \sqrt{z_3} + \sqrt{z_4}\right)\right).
\end{aligned}$$

However, this is only possible if

$$\sqrt{z_1} - \sqrt{z_2} - \sqrt{z_3} + \sqrt{z_4} = 0 \tag{4.3.12}$$

holds. Since $z_1, z_2, z_3, z_4 \geq 0$, equations (4.3.11) and (4.3.12) are a special case of the equations in (4.2.5). Hence the same tedious arguments as for (4.2.5) shows that (4.3.11) and (4.3.12) imply $z_1 = z_2$ and $z_3 = z_4$. However, from the definition of z_1 and z_2, this is contradictory to the fact that $\theta_1 < \theta_2$. So we conclude that $\mathfrak{L}_\theta \not\equiv \mathfrak{L}_\xi$ and (4.3.9) holds. Since Θ is bounded, lemma 4.3.4 and (4.3.9) show that the conditions of lemma 4.1.6 are fulfilled, which completes the proof. □

Lemma 4.3.6 *The differentiability condition* ***D(***θ***)*** *is fulfilled for every* $\theta \in \Theta$. *More precisely, the function* $\theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ *is Fréchet-differentiable in every* $\theta = (\theta_1, \theta_2) \in \Theta$ *with derivative* $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$ *given by*

$$\begin{pmatrix} \dfrac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta(\alpha) \\ \dfrac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta(\alpha) \end{pmatrix} = \frac{\alpha}{a} \cdot \begin{pmatrix} \dfrac{\phi\left(\frac{\alpha}{b}+1, \frac{2a}{\sigma^2}+1; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)} \\ \dfrac{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_1\right)\phi\left(\frac{\alpha}{b}+1, \frac{2a}{\sigma^2}+1; \frac{2b}{\sigma^2}\theta_2\right)}{-\phi^2\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)} \end{pmatrix}$$

where ϕ *is the confluent hypergeometric function. Further, the partial derivatives* $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ *and* $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ *are linearly independent in* $\mathcal{H}$.

Proof: We follow the same reasoning as in the proof of 4.2.4.
1) From (A.1.1) we know that for fixed $\gamma > 0$ $\mathbb{C}^2 \ni (\alpha, z) \mapsto \phi(\alpha, \gamma; z)$ is an entire function and further $\phi(\alpha, \gamma; z) > 0$ for all $\alpha \in \mathbb{R}_+$ and $z \geq 0$. Hence $\mathfrak{L} : \mathbb{R}^+ \times \Theta \longrightarrow \mathbb{R}$,

$$(\alpha, \theta_1, \theta_2) \mapsto \mathfrak{L}_\theta(\alpha) = \frac{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)} \tag{4.3.13}$$

computed in lemma 4.3.1, is a composition of C^∞ functions in each component and therefore C^∞ itself. Since μ has compact support, lemma 4.1.7 shows that for all $\theta \in \Theta$, $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta \in \mathcal{H}$ and

$$\Theta \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$$

is Fréchet-differentiable with derivative $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$.

Now we have to compute the partial derivatives of $\theta \mapsto \mathfrak{L}_\theta$. We point out the derivation rule (A.1.3),

$$\frac{\mathrm{d}}{\mathrm{d}z}\phi(\alpha, \gamma; z) = \frac{\alpha}{\gamma}\phi(\alpha + 1, \gamma + 1; z)$$

and define

$$\tilde{\phi}(\alpha, z) := \phi\left(\frac{\alpha}{b}, \frac{2a}{\sigma^2}; \frac{2b}{\sigma^2}z\right),$$

for $z \in \mathbb{R}$ and $\alpha \in \mathbb{R}_+$. Hence

$$\begin{aligned}\tilde{\phi}'(\alpha, z) :=& \frac{\mathrm{d}}{\mathrm{d}z}\tilde{\phi}(\alpha, z) = \frac{2b}{\sigma^2} \cdot \frac{\alpha}{b} \cdot \frac{\sigma^2}{2a}\phi\left(\frac{\alpha}{b} + 1, \frac{2a}{\sigma^2} + 1; \frac{2b}{\sigma^2}z\right) \\ =& \frac{\alpha}{a} \cdot \phi\left(\frac{\alpha}{b} + 1, \frac{2a}{\sigma^2} + 1; \frac{2b}{\sigma^2}z\right)\end{aligned}$$

holds. Note from (4.3.13) that $\mathfrak{L}_\theta(\alpha) = \dfrac{\tilde{\phi}(\alpha, \theta_1)}{\tilde{\phi}(\alpha, \theta_2)}$. Evidently, we only have to use the chain rule to get the partial derivatives of $\mathfrak{L}_\theta$ given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta &= \frac{\tilde{\phi}'(\alpha, \theta_1)}{\tilde{\phi}(\alpha, \theta_2)} \\ \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta &= -\frac{\tilde{\phi}(\alpha, \theta_1) \cdot \tilde{\phi}'(\alpha, \theta_2)}{\left(\tilde{\phi}(\alpha, \theta_2)\right)^2}.\end{aligned} \tag{4.3.14}$$

Note that $D\mathfrak{L}_\theta := \left(\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta, \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta\right)^\top$ has the form stated in the assertion.
2) To show the linear independence of $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ and $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ in $\mathcal{H}$ by contradiction, we suppose now that they are linearly dependent. Hence there exists a constant $c \in \mathbb{R} \setminus \{0\}$ such that

$$\left\| \frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta + c \cdot \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta \right\|_{\mathcal{H}} = 0. \tag{4.3.15}$$

Define $F \in \mathcal{H}$ by

$$F(\alpha) := \frac{\tilde{\phi}'(\alpha, \theta_2)}{\tilde{\phi}(\alpha, \theta_2)}$$

and let $\tilde{\mathfrak{L}}_\theta(\cdot)$ be the Laplace transform of the level-crossing time of a CIR process with parameter $\tilde{a} := a + \frac{\sigma^2}{2}$, b and σ,

$$\tilde{\mathfrak{L}}_\theta(\alpha) := \frac{\phi\left(\frac{\alpha}{b}, \frac{2\tilde{a}}{\sigma^2}; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b}, \frac{2\tilde{a}}{\sigma^2}; \frac{2b}{\sigma^2}\theta_2\right)}, \quad \alpha \geq 0.$$

Then

$$\tilde{\mathfrak{L}}_\theta(\alpha + b) = \frac{\phi\left(\frac{\alpha}{b} + 1, \frac{2a}{\sigma^2} + 1; \frac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\frac{\alpha}{b} + 1, \frac{2a}{\sigma^2} + 1; \frac{2b}{\sigma^2}\theta_2\right)} = \frac{\tilde{\phi}'(\alpha, \theta_1)}{\tilde{\phi}'(\alpha, \theta_2)}, \quad \alpha \geq 0 \tag{4.3.16}$$

holds and with $\mathfrak{L}_\theta(\alpha) = \dfrac{\tilde{\phi}(\alpha, \theta_1)}{\tilde{\phi}(\alpha, \theta_2)}$, (4.3.14) and (4.3.15) it follows that

$$\begin{aligned}
\int_{\mathbb{R}_+} F^2(\alpha) \cdot \left| \tilde{\mathfrak{L}}_\theta(\alpha + b) - c \cdot \mathfrak{L}_\theta(\alpha) \right|^2 \mu(\mathrm{d}\alpha) &= \left\| F \cdot \left\{ \tilde{\mathfrak{L}}_\theta(\cdot + b) - c \cdot \mathfrak{L}_\theta \right\} \right\|_{\mathcal{H}}^2 \\
&= \left\| \frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta + c \cdot \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta \right\|_{\mathcal{H}}^2 \\
&= 0.
\end{aligned}$$

Hence

$$F^2(\alpha) \cdot \left| \tilde{\mathfrak{L}}_\theta(\alpha + b) - c \cdot \mathfrak{L}_\theta(\alpha) \right|^2 = 0, \quad \text{for } \mu \text{ almost every } \alpha \in \mathbb{R}_+.$$

Moreover, $F(\alpha) > 0$ holds for all $\alpha \geq 0$. This is due to the fact that $\phi(\alpha, \gamma; z) > 0$ for all $\alpha \in \mathbb{R}_+$, $z \geq 0$ and $\gamma > 0$ (cf. (A.1.1)). So we deduce

$$\left| \tilde{\mathfrak{L}}_\theta(\alpha + b) - c \cdot \mathfrak{L}_\theta(\alpha) \right| = 0, \quad \text{for } \mu \text{ almost every } \alpha \in \mathbb{R}_+.$$

Since $\mathbb{R}_+ \ni \alpha \mapsto \mathfrak{L}_\theta(\alpha) \in [0,1]$ is continuous and μ has a piecewise continuous Lebesgue density, there exists at least one open interval $I \subset \mathbb{R}_+$ such that

$$\tilde{\mathfrak{L}}_\theta(\alpha + b) = c \cdot \mathfrak{L}_\theta(\alpha) \quad \forall \alpha \in I.$$

However, if two analytic functions on $\mathbb{C}_{>0} := \{z \in \mathbb{C} : \mathrm{Re} z \geq 0\}$ coincide on an open interval $I \subset \mathbb{R}_+$, they coincide on $\mathbb{C}_{>0}$ (see e.g. [B 81] Ch.10.1 Thm.1, Thm.2) and so

$$\tilde{\mathfrak{L}}_\theta(\alpha + b) = c \cdot \mathfrak{L}_\theta(\alpha) \quad \forall \alpha \in \mathbb{R}_+. \tag{4.3.17}$$

holds. By definition A.1.1 $\phi(\alpha, \alpha, z) = e^z$ holds for every α and z. If we put $\tilde{\alpha} := \frac{2ab}{\sigma^2} > 0$ into (4.3.16), we deduce

$$\begin{aligned}
\tilde{\mathfrak{L}}_\theta(\tilde{\alpha} + b) &= \frac{\phi\left(\dfrac{2a}{\sigma^2} + 1, \dfrac{2a}{\sigma^2} + 1; \dfrac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\dfrac{2a}{\sigma^2} + 1, \dfrac{2a}{\sigma^2} + 1; \dfrac{2b}{\sigma^2}\theta_2\right)} \\
&= \frac{\exp\left(\frac{2b}{\sigma^2}\theta_1\right)}{\exp\left(\frac{2b}{\sigma^2}\theta_2\right)} \\
&= \frac{\phi\left(\dfrac{2a}{\sigma^2}, \dfrac{2a}{\sigma^2}; \dfrac{2b}{\sigma^2}\theta_1\right)}{\phi\left(\dfrac{2a}{\sigma^2}, \dfrac{2a}{\sigma^2}; \dfrac{2b}{\sigma^2}\theta_2\right)} \\
&= \mathfrak{L}_\theta(\tilde{\alpha}).
\end{aligned}$$

With (4.3.17) in mind, this implies that $c = 1$ and in particular

$$\tilde{\mathfrak{L}}_\theta(b) = \mathfrak{L}_\theta(0) = 1.$$

However, since $b > 0$ this is contradictory to the fact that $\tilde{\mathfrak{L}}_\theta$ is the Laplace transform of a non-degenerate distribution. So we finally conclude that $\frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta$ and $\frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$ must be linearly independent. $\square$

Also in the CIR case we have a main result corresponding to the one in the OU case.

Theorem 4.3.7 *Let θ_n^*, $n \in \mathbb{N}$ be any MDE defined in* (4.3.4) *and $\mathfrak{L}_\theta$ as given in 4.3.1. Let further*

$$D_1\mathfrak{L}_\theta := \frac{\mathrm{d}}{\mathrm{d}\theta_1}\mathfrak{L}_\theta \quad \textit{and} \quad D_2\mathfrak{L}_\theta := \frac{\mathrm{d}}{\mathrm{d}\theta_2}\mathfrak{L}_\theta$$

be as in lemma 4.3.6. Define $U_\theta, V_\theta \in \mathbb{R}^{2\times 2}$ by

$$(U_\theta)_{i,j} := \langle D_i\mathfrak{L}_\theta, D_j\mathfrak{L}_\theta\rangle_{\mathcal{H}}, \quad 1 \leq i,j \leq 2$$

and

$$(V_\theta)_{i,j} := \int_0^\infty \int_0^\infty D_i\mathfrak{L}_\theta(\alpha_1)K(\alpha_1,\alpha_2)D_j\mathfrak{L}_\theta(\alpha_2)\mu(\mathrm{d}\alpha_1)\mu(\mathrm{d}\alpha_2), \quad 1 \leq i,j \leq 2,$$

where $K(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$K(\alpha_1,\alpha_2) := \mathfrak{L}_\theta(\alpha_1+\alpha_2) - \mathfrak{L}_\theta(\alpha_1)\mathfrak{L}_\theta(\alpha_2) = \mathbf{Cov}_\theta\left[e^{-\alpha_1\tau}, e^{-\alpha_2\tau}\right].$$

Then $(\theta_n^*)_{n\in\mathbb{N}}$ is strongly consistent and asymptotically normal, i.e.

$$\sqrt{n}\left(\theta_n^* - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\left(0, U_\theta^{-1}V_\theta U_\theta^{-1}\right).$$

Proof: Since we are in the setting of section 4.1 lemmas 4.1.1, 4.1.2, 4.1.4 and 4.1.5 remain valid. Together with lemmas 4.3.4, 4.3.5 and 4.3.6, all assumptions and conditions of section 2.2 are fulfilled. Finally, we can apply theorem 2.2.4 which completes the proof. □

Remark 1: As in the previous section, all results of this section remain valid for the one-dimensional sub-experiments, i.e. for unknown θ_1 the parameter space is given by

$$\Theta_1 := (0, y),$$

where y is any known upper level. For unknown θ_2 the parameter space is given by

$$\Theta_2 := (y, \infty),$$

where y is any known lower level. All proofs are literally the same or even easier.

Remark 2: Theorem 4.2.5 and 4.3.7 show asymptotic normality and strong consistency for the MDE estimating $\theta = (\theta_1, \theta_2)$ in the OU and CIR case. However, their is no evidence that the asymptotic variance of the MDE attains a Cramér-Rao bound. If we could show the LAN or LAMN (local asymptotic mixed normality) property for the corresponding sequences of experiments, we could perform a one step correction proposed by Le Cam in order to construct asymptotically best estimators having the LAM property (cf. [LY 00] and [D 85]). For this purpose, we would have to compute empirical score and Fisher information from inverse Laplace transforms. These are not given explicitly of course. Hence for applications this might be too complicated. Moreover, up to now there is no indication of whether this is a LAN, LAMN, LAQ (locally asymptotically quadratic) or even more general situation. Nevertheless, an investigation of this problem would be an interesting topic for a further work.

Chapter 5

Applications

5.1 Explanatory Notes on Tools and Strategy

To apply our theoretical results to real data in order to find x_0 and S, we have to determine the statistical experiment (OU, CIR, ...) first. This will be done by estimating diffusion and drift coefficient of the diffusion process between spikes. Höpfner [H 07] introduced a nonparametric estimation method based on estimators suggested by Florens-Zmirou [FZ 93]. This method is described in the following subsection. For a detailed explanation we refer to [H 07] p.279,280 + pp. 296-299. The second subsection describes the application of the MDE method developed in chapter 4.

5.1.1 Nonparametric Estimation of Diffusion and Drift Coefficients

We now consider a process X defined by an SDE

$$\mathrm{d}X_t = \beta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t,$$

as introduced in (1.1.1). To estimate the drift coefficient

$$x \mapsto \beta(x)$$

and the squared diffusion coefficient

$$x \mapsto \sigma^2(x),$$

we infer from discrete time observations

$$X_{i\Delta}, \quad i_0 \le i \le i_1$$

where $i_0, i_1 \in \mathbb{N}$ and $\Delta > 0$. The kernel estimators are given by

$$\widehat{\beta}(x) := \widehat{\beta}_{(\Delta,M,h)}(x) = \frac{\sum_{i=i_0}^{i_1-M} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{X_{(i+M)\Delta}-X_{i\Delta}}{\Delta M}\right)}{\sum_{i=i_0}^{i_1-M} K\left(\frac{X_{i\Delta}-x}{h}\right)} \tag{5.1.1}$$

and

$$\widehat{\sigma^2}(x) := \widehat{\sigma^2}_{(\Delta,M,h)}(x) = \frac{\sum_{i=i_0}^{i_1-M} K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{X_{(i+M)\Delta}-X_{i\Delta}}{\sqrt{\Delta M}}\right)^2}{\sum_{i=i_0}^{i_1-M} K\left(\frac{X_{i\Delta}-x}{h}\right)}, \tag{5.1.2}$$

where K is either a rectangular or a triangular kernel

$$K(y) = \frac{1}{2}\mathbb{1}_{(-1,1)}(y), \qquad K(y) = (1-|y|)\mathbb{1}_{(-1,1)}(y). \tag{5.1.3}$$

The bandwidth $h > 0$ and a number $M \in \mathbb{N}$ of Δ-time steps are suitable constants. Evidently, the choice of h depends on the accuracy of the data. To be in the scope of consistency, we estimate drift and diffusion coefficients only at points x visited many times by X. This means, we are only concerned about points x such that

$$OT(x) := \sum_{i=i_0}^{i_1} \mathbb{1}_{[x-\frac{h}{2}, x+\frac{h}{2}]}(X_{i\Delta})$$

is 'large' (e.g. > 1000). We call $OT(x)$ *occupation time* at x because it measures how often the process visits a vicinity of x. Certainly, this quantity should be as large as possible. On the other hand, we also would like to consider as many points as possible. Consequently, a suitable lower bound on $OT(x)$ depends on the data.

The choice of M further depends on Δ which is given by data. A suitable choice can help to compensate measurement errors and microstructure noise. If accuracy of measurement in the data set is poor but Δ very small, the estimator will produce huge values if M is not suitably large. If M is too large, the time scale might be too large to determine infinitesimal behavior. So we have to determine empirically from the data in which range of M the results remain stable under variation of M. Since

$$\begin{aligned} \sigma^2(X_t) &= \lim_{\delta\to 0} \frac{1}{\delta}\mathbf{E}[(X_{t+\delta}-X_t)^2|X_t], \\ \beta(X_t) &= \lim_{\delta\to 0} \frac{1}{\delta}\mathbf{E}[(X_{t+\delta}-X_t)|X_t] \end{aligned} \tag{5.1.4}$$

holds, a moderate variation of M should not change the result of the estimator too much. This will also be visualized later on with the help of simulated data. Anyway, since we assume $\beta(\cdot), \sigma(\cdot) \in C^1(\mathbb{R})$, from the theory of nonparametric estimation it is necessary to choose M and h such that the quantities

$$\Delta M \quad \text{and} \quad \frac{h}{(\Delta M)^{1/3}}$$

are small. This fact is also discussed in [H 07] p.279,280 + pp. 296-299.

To apply these kernel estimators we have to ensure that the discrete observations of the membrane potential which are taken into account are not influenced by a spike. This is necessary because a spike is of almost deterministic shape, whereas the kernel estimators require the observation of a diffusion process. Hence, data from a spike trajectory would distort the estimation results. Further, the model assumes the diffusion process only between spikes. For these reasons, we cut out a fixed broad neighborhood around every spike such that we are sure the resulting segments are nowhere belonging to a spike trajectory. However, this works properly only if the intervals between spikes are long enough, as in the case of the subthreshold regime. These segments contain the data used for the evaluation of the kernel estimators. For every segment nominator and denominator of (5.1.1) and (5.1.2) are computed separately. Finally, the corresponding terms are added up before the ratio is computed in order to get the estimate.

5.1.2 Application of the MDE Method in Order to Estimate x_0 and S

The results of (5.1.1) and (5.1.2) determine the model described in chapter 1 except reset value and excitation threshold. Hence, we are in the position to consider the statistical experiment described in 1.2 in order to find reasonable values for these quantities. In the previous chapters an estimation theory was developed concerning the cases BMD, GBM, OU and CIR. The MLE method in the cases BMD and GBM is easy to apply because the estimators are given explicitly. So we will now explain how to apply the MDE method in OU and CIR cases.

Since in our considerations the MDE is not given explicitly, we have to use a computer for evaluation. For this purpose, we have to compare the empirical Laplace transform of τ with a parametrized family of possible true Laplace transforms of τ in the Hilbert space $\mathcal{H} := L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mu)$ with respect to its norm $\|\cdot\|_{\mathcal{H}} := \int_{\mathbb{R}_+} |\cdot| \mathrm{d}\mu$. From section 2.2 we have the condition that the comparison measure μ should be absolutely continuous with respect to the Lebesgue measure. Certainly, this never holds as long as we use the computer who is only able to compare functions at finitely many points. Obviously, it is very easy to evaluate the empirical Laplace transform $\hat{\mathfrak{L}}_n : \mathbb{R}_+ \longrightarrow [0,1]$

$$\hat{\mathfrak{L}}_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} e^{-\alpha\tau_i}$$

at certain points $\alpha \in \mathbb{R}_+$. In the cases OU or CIR lemmas 4.2.1 and 4.3.1 show that true Laplace transforms are ratios of Hermite and confluent hypergeometric functions respectively, which are not given explicitly in general. For this reason, we use the integral representations (A.1.4) and (A.2.5) for Hermite and confluent hypergeometric functions respectively to evaluate

$$\mathfrak{L}_\theta(\alpha) := \mathbf{E}_\theta[e^{-\alpha\tau}]$$

numerically for certain values of $\alpha \in \mathbb{R}_+$ and $\theta \in \Theta$. Anyway, we have to think about the choice of μ or the choice of points that μ should take into account. Since μ is crucial to distinguish $\hat{\mathfrak{L}}_n$ from other functions and the computer only has finite accuracy, μ should consider only points of the domain where $\hat{\mathfrak{L}}_n$ is most distinguishable, i.e. where slope and curvature of $\hat{\mathfrak{L}}_n$ are substantial. Since $\hat{\mathfrak{L}}_n$ is convex and strictly monotonically decreasing to 0, μ may disregard all points of the domain where $\hat{\mathfrak{L}}_n$ takes small values. In addition, μ should consider more points near 0 because there the absolute value of the slope of $\hat{\mathfrak{L}}_n$ is maximal. By the way, the values of the Laplace transform on the vicinity of 0 also determine its derivatives in 0 and thereby the moments of the corresponding distribution. In the following, we describe how to find a μ that satisfies our requests. We divide $(0,1]$, the range of $\hat{\mathfrak{L}}_n$, into m parts, $\left(\frac{j-1}{m}, \frac{j}{m}\right]$, $j = 1\ldots, m$ and choose points α_j such that $\hat{\mathfrak{L}}_n(\alpha_j) \in \left(\frac{j-1}{m}, \frac{j}{m}\right]$, $j = 1\ldots, m$. This is easy work for the computer. From this we define the comparison measure μ for some suitable integer $m \in \mathbb{N}$,

$$\mu := \sum_{j=1}^{m} \delta_{\alpha_j} \tag{5.1.5}$$

where δ_α denotes the Dirac measure at the point α. This procedure is adaptive in a sense. The greater the absolute value of the slope of $\hat{\mathfrak{L}}_n$, the more points are taken into account. Note that μ ignores the interval (α_m, ∞) completely. However, this is not crucial. Since $\hat{\mathfrak{L}}_n$ is bounded by 0 and $\frac{1}{m}$ on (α_m, ∞), convexity and monotonicity ensure that the absolute value of the slope is relatively small on this interval. This now defines the Hilbert space $\mathcal{H} := L^2(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), \mu)$ with norm $\|\cdot\|_{\mathcal{H}}$, in order to compare empirical and true LT. On the one hand m should be large such that μ has approximately a

Lebesgue density in the weak sense. On the other hand m should not be too large in order to save computing costs. In the applications of this chapter we choose $m = 100$. Note that our choice of μ depends on $\hat{\mathfrak{L}}_n$ and so it depends on the data. However, this is not a problem because excepting some restrictions from theory (cf. section 2.2) we are free to choose a reasonable measure μ.

There is one more problem we have to consider. To evaluate the true Laplace transform we also have to choose finitely many points of $\Theta \subset \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 < \theta_2\}$. For this reason, we define a finite lattice $\Theta^L \subset \Theta$. Since the parametrization $\Theta \ni \theta \mapsto \mathfrak{L}_\theta \in \mathcal{H}$ has to be continuous, a refinement of the lattice Θ^L should realize a good approximation of the MDE and accordingly

$$\theta_n^* := \arg\inf_{\xi \in \Theta} \|\hat{\mathfrak{L}}_n - \mathfrak{L}_\xi\|_{\mathcal{H}} \quad \approx \quad \arg\min_{\xi \in \Theta^L} \|\hat{\mathfrak{L}}_n - \mathfrak{L}_\xi\|_{\mathcal{H}}. \tag{5.1.6}$$

The right hand side of (5.1.6) can be evaluated by the computer. However, we have to care about the size of Θ^L concerning the computing costs of course. A strategy to save computing costs is the following. First we consider a rough lattice to find some points $\xi^1, \xi^2, \dots \in \Theta^L$ such that the values $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\xi^i}\|_{\mathcal{H}}$ are small. The next step is to refine Θ^L only around these points $\xi^1, \xi^2, \dots$ and to evaluate the MDE again, which leads to new "local minima" $\hat{\xi}^1, \hat{\xi}^2, \dots,$ and so on.

Nevertheless, the number of observations $n \in \mathbb{N}$ has to be large in order to be in the scope of the SLLN and consistency. Otherwise, the refinement of Θ^L makes no sense. In fact, realizing a high value of n is a big problem in real data. This is due to the fact that a neuron gets tired while firing for a long time such that the ISIs become longer. However, this behavior does not justify our iid assumption for the inter spike or level crossing times respectively. So n might be rather small in order to satisfy the iid assumption. We will return to this problem within the next sections.

5.2 MDE Application to Simulated Data

To check the performance of our MDE method, we are now concerned about simulated data where we know what the MDE result should be. Exemplarily, we consider simulated CIR inter spike times (or level crossing times). We recall again the SDE for the CIR process given in (4.3.2)

$$\mathrm{d}X_t = (a - b(X_t - c))\mathrm{d}t + \sigma\sqrt{X_t - c}\,\mathrm{d}B_t, \quad X_0 = x_0 \geq c.$$

To simulate an approximate solution of this SDE we applied the *Milstein Scheme* (cf. Kloeden & Platen [KP 99]) which has the following representation for the CIR process,

$$X_{\Delta(n+1)} = X_{\Delta n} + (a - b(X_{\Delta n} - c)) \cdot \Delta + \sigma\sqrt{X_{\Delta n} - c} \cdot W_{\Delta,n} + \frac{\sigma^2}{4} \cdot (W_{\Delta,n}^2 - \Delta),$$

$n \in \mathbb{N}_0$ where $W_{\Delta,n} \sim \mathcal{N}(0, \Delta)$ are iid random variables.

Remark: Certainly, there are better methods to simulate solutions of SDEs discussed in the book of Kloeden & Platen [KP 99]. Especially, even the simulation of level crossing times was considered by Giraudo, Sacerdote and Zucca in [GSZ 01] who were concerned about the detection of level crossing events. We also could have taken the transition probability of the CIR process in order to evaluate the process at discrete points of time. However, our simulations were only for checking the applicability of the MDE and in the first instance we want to apply this MDE to real data. So for simplicity, we did not consider these methods.

For our simulations we chose the following parameters:

- $\Delta = 0.001$[ms]$= 0.000001$[s]
- $a = 93.1$[mV/s]
- $b = 28.7$[1/s]
- $\sigma = 1.8[\sqrt{\mathrm{mV/s}}]$
- $c = -53.7$[mV]
- $x_0 = -52.7$[mV]
- $S = -49.3$[mV]

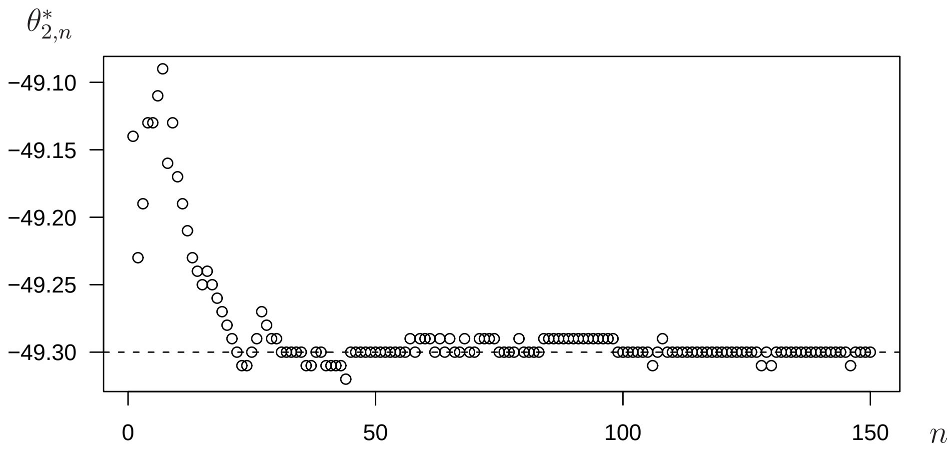

Figure 5.1: Trajectory of the MDE for θ_2. The dashed line is the true simulation parameter $S = -49.3$.

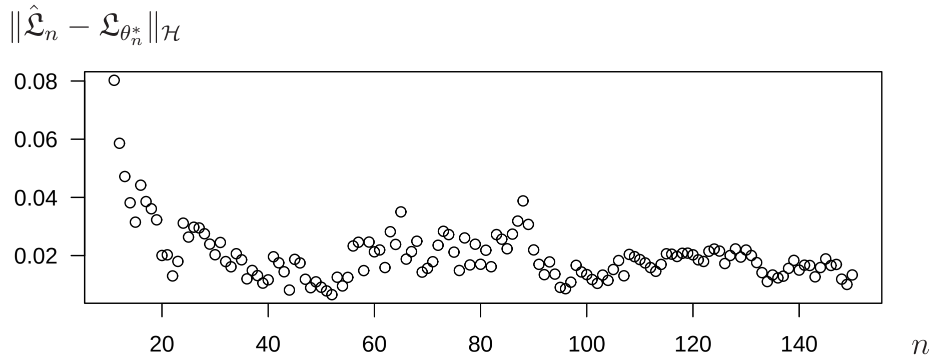

Figure 5.2: Depending on n, the distance of *estimated* and empirical LT in $\mathcal{H}$ is plotted.

Except for x_0 and S, our estimation method uses all parameters listed above. These parameters are close to what one observes in real data (cf. section 5.3). We simulated 100000 CIR processes starting at x_0 and stopping at hitting or crossing the level S. Then we stored the time length of these intervals. These 100000 "inter spike times" were successively used to compute the empirical Laplace transform.

To evaluate the MDE we define a finite lattice $\Theta^L \subset \Theta$ on the parameter space

$$\Theta^L := \{-53.7, -53.69, -53.68, \dots, -50\} \times \{-51, -50.99, \dots, -47\} \cap \Theta.$$

The measure μ is chosen as described in (5.1.5) with $m = 100$.

In figures 5.1 and 5.3 the trajectory of the MDE is plotted according to its components $\theta_1 = x_0$ and $\theta_2 = S$. Obviously, the MDE converges very fast in $\theta^*_{2,n}$. Only 50 observations of τ are enough to be close to the real parameter.

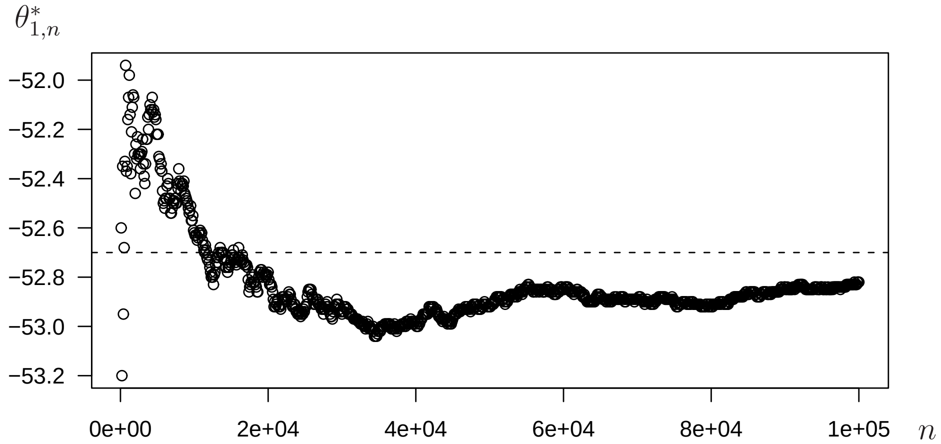

Figure 5.3: Trajectory of the MDE for $\theta_1 = x_0$. Only every hundredth step is plotted. The dashed line is the true simulation parameter $x_0 = -52.7$.

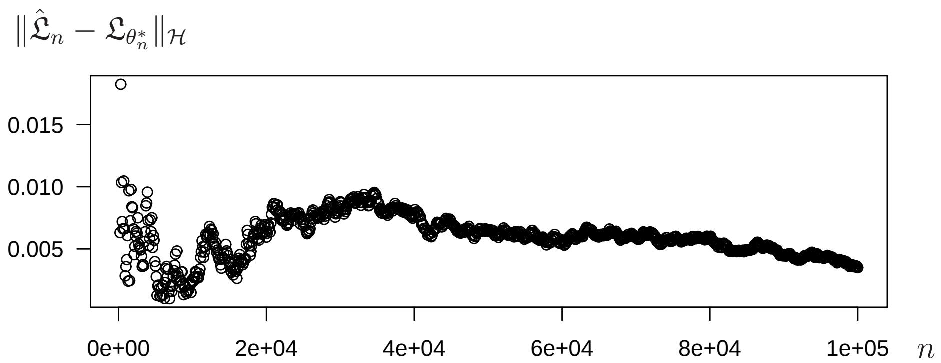

Figure 5.4: A continuation of figure 5.2 with another scaling. Depending on n, the distance of *estimated* and empirical LT in $\mathcal{H}$ of every hundredth step is plotted.

Certainly, the consistency of θ_n^* should take advantage of the fact that Θ has a natural lower bound $c = -53.7$ due to the CIR model. However, the MDE converges very slowly to the true value in $\theta_{1,n}^*$. 100000 observations are needed in order to visualize somehow the consistency of $\theta_{1,n}^*$.

For this reason, now we are interested in the behavior of the asymptotic variance of θ_n^* for different points $\theta \in \Theta$. From theorem 4.3.7 we know that

$$\sqrt{n}\,(\theta_n^* - \theta) \xrightarrow[n\to\infty]{\mathcal{L}(\theta)} \mathcal{N}\,(0, \Sigma_\theta)$$

where $\Sigma_\theta := U_\theta^{-1} V_\theta U_\theta^{-1}$, U the Gramian matrix of the derivatives $D_1\mathfrak{L}_\theta, D_2\mathfrak{L}_\theta$ and

$$(V_\theta)_{i,j} := \int_0^\infty \int_0^\infty D_i\mathfrak{L}_\theta(\alpha)K(\alpha,\tilde{\alpha})D_j\mathfrak{L}_\theta(\tilde{\alpha})\mu(\mathrm{d}\alpha)\mu(\mathrm{d}\tilde{\alpha}) \quad 1 \le i,j \le 2,$$

with

$$K(\alpha,\tilde{\alpha}) := \mathfrak{L}_\theta(\alpha+\tilde{\alpha}) - \mathfrak{L}_\theta(\alpha)\mathfrak{L}_\theta(\tilde{\alpha}).$$

We are able to evaluate all these quantities numerically. Since μ only supports the points α_j, $j = 1, \ldots, m$, the scalar product and the integration with respect to μ become matrix operations for the computer. Further, the derivatives are again ratios of confluent hypergeometric functions which will be evaluated using (A.1.4). The results of the diagonal components of $\Theta \ni \theta \mapsto (\Sigma_\theta)_{i,i}$, $i = 1, 2$ are shown in figures 5.6 and 5.5. Figure 5.7 shows the asymptotic correlation coefficient between $\theta_{1,n}^*$ and $\theta_{2,n}^*$ given by

$$\mathbf{Cor}_\theta(\theta_\infty^*) := \frac{(\Sigma_\theta)_{1,2}}{\sqrt{(\Sigma_\theta)_{1,1}\cdot(\Sigma_\theta)_{2,2}}}.$$

Note that all results are computed for the fixed values a, b, c, σ from above concerning the CIR model (4.3.2). Evidently, $a/b + c \approx -50.46$ is the mean value of the stationary distribution of the process X. Hence, for biological reasons, we only consider the parameters $\theta_1 \le -50.46$ and $\theta_1 \le \theta_2 - 0.1$[mV].

From figure 5.5 we notice that we do not have to worry about the asymptotic variance of $\theta_{2,n}^*$, as long as θ_2 and θ_1 are not too close. However, from a biological point of view a distance $\theta_2 - \theta_1 \ge 0.5$[mV] should be self-evident. For the simulation parameter $\theta = (-52.7, -49.3)$ we computed $(\Sigma_\theta)_{2,2} \approx 0.067$ which explains figure 5.1.

In figure 5.6 we notice from the logarithmic scale that there is only a small promising area in Θ where it might be possible to estimate θ_1 with a realistic amount of observations concerning real data. Only if $\theta_2 - \theta_1 < 1$[mV] the asymptotic variance of $\theta_{1,n}^*$ is smaller than 10. For the simulation parameter

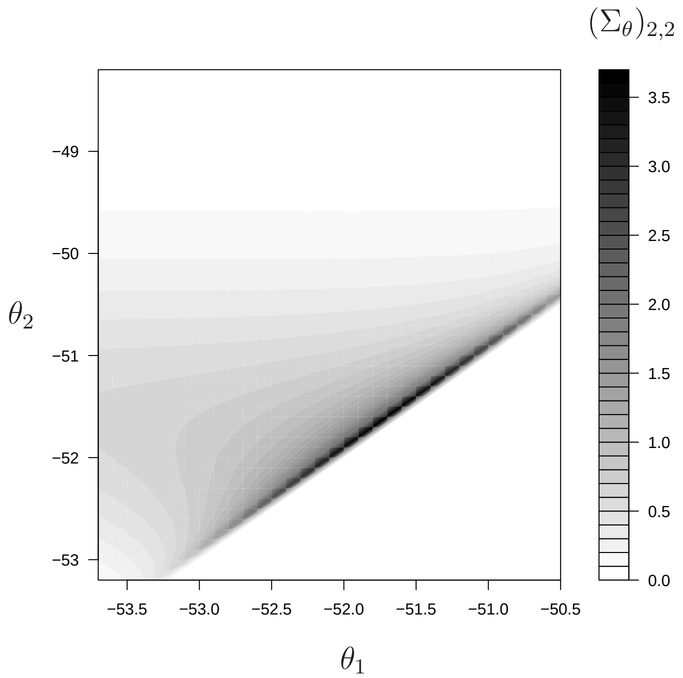

Figure 5.5: The asymptotic variance of $\theta^*_{2,n}$ depending on the true value $\theta \in \Theta$. The mean value of the stationary distribution of X is $a/b + c \approx -50.5$.

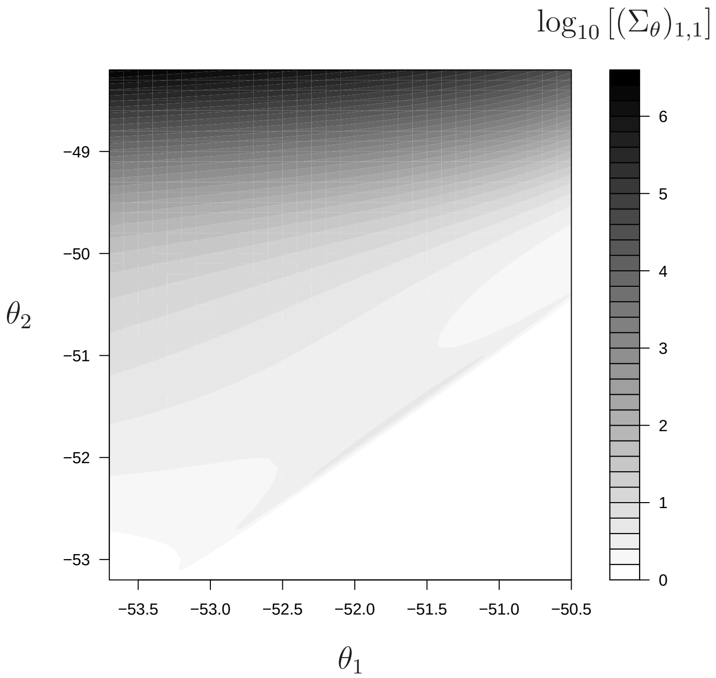

Figure 5.6: The asymptotic variance of $\theta^*_{1,n}$ in logarithmic scale depending on the true value $\theta \in \Theta$. $a/b + c \approx -50.5$.

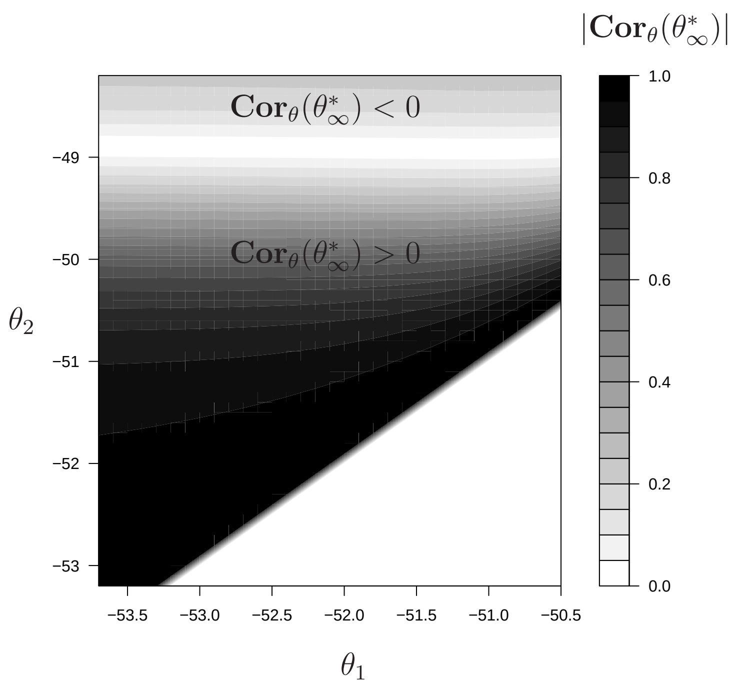

Figure 5.7: The asymptotic correlation coefficient between $\theta^*_{1,n}$ and $\theta^*_{2,n}$. $a/b+c \approx -50.5$.

$\theta = (-52.7, -49.3)$ we computed $(\Sigma_\theta)_{1,1} \approx 383$ which explains figure 5.3, and for nearly all parameters of the subthreshold regime ($\theta_2 > -50.46$) the asymptotic variance for $\theta^*_{1,n}$ is extremely high. This is due to the fact that in the subthreshold case the Laplace transform of the level crossing time changes very little if we vary θ_1, which can be explained in general as follows. Assume the subthreshold case ($\theta_2 > a/b+c$) and θ_1 to be far below $a/b+c$. In these regions, far away from $a/b+c$, drift is much stronger than fluctuations of the process. The behavior is more deterministic and starting in θ_1 the process runs exponentially fast to a vicinity of $a/b+c$ where it may oscillate for quite a long time until a crossing of θ_2 happens. Anyway, the level crossing times are comparatively much longer than the time it takes the process to reach a vicinity of $a/b+c$. So whether θ_1 is near or far below $a/b+c$ does not essentially affect the typical behavior of the level crossing times.

Further, this also means that $\theta_1 = x_0$ is a parameter of secondary importance for the model in the subthreshold regime. It is not necessary to determine exactly the value of x_0 in order to reproduce the typical spiking behavior. The crucial parameter is $\theta_2 = S$. Certainly, this is true only as long as S is greater than the mean value of the stationary distribution.

From figure 5.7 we notice a band in Θ where the components of the asymptotic distribution of θ^*_n are almost uncorrelated. For the simulation parameter $\theta = (-52.7, -49.3)$ we computed $\mathbf{Cor}_\theta(\theta^*_\infty) \approx 0.21$. In fact, for the reasons mentioned above, it is not surprising that in the subthreshold case the correlation between $\theta^*_{1,n}$ and $\theta^*_{2,n}$ is low. This fact is also supported by figures 5.1 and 5.3 that show the very fast convergence of $\theta^*_{2,n}$ along with the very slow convergence of $\theta^*_{1,n}$. The strong correlation between $\theta^*_{1,n}$ and $\theta^*_{2,n}$ in the suprathreshold regime is self-evident, since the process behaves more deterministically in this case.

A further interesting observation is the following. If we take a look at figures 5.4 and 5.3 together we observe that whenever the estimator is near the true value, the empirical LT $\hat{\mathfrak{L}}_n$ fits very well the *estimated* LT $\mathfrak{L}_{\theta^*_n}$. This is not very surprising. On the other hand, if the estimator is far away from the true value, $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*_n}\|_{\mathcal{H}}$ yields bigger values and therefore a worse fit. This is indeed a strange fact, since we consider the fit to the estimated LT $\mathfrak{L}_{\theta^*_n}$ and not the fit to the true LT $\mathfrak{L}_\theta$. The estimator seems to know that it is wrong but in these cases it is not able to find a better value for θ.

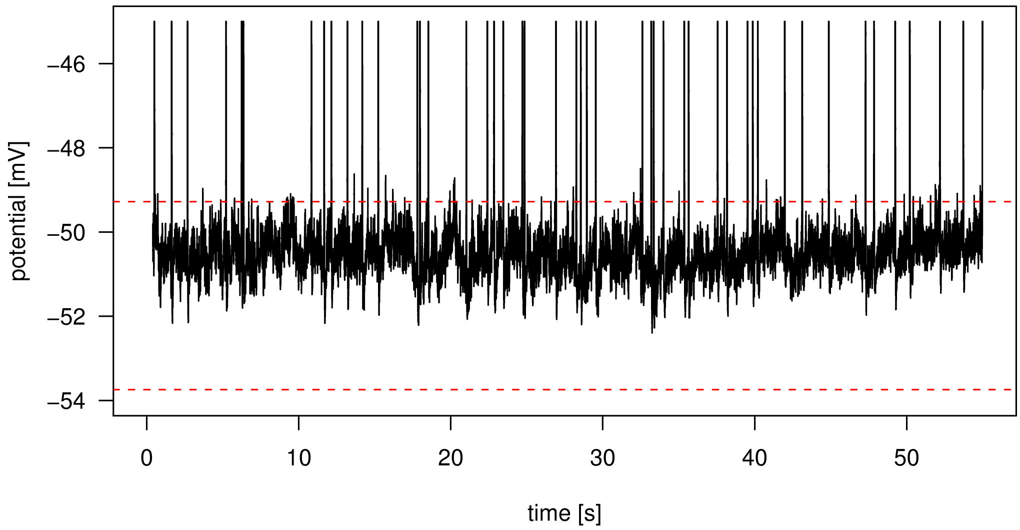

Figure 5.8: This membrane potential including 47 spikes (46 ISIs) was recorded in vitro (by W. Kilb, Institute of Physiology, Uni-Mainz) from a neuron belonging to a cortical slice preparation from a 6 week old mouse. The network in this cortical slice was stimulated by a potassium solution (level: 5mM K^+). For illustration the potential has been truncated at −45mV. The dashed lines are $\theta^* = (-53.74, -49.28)$ which is the result of the MDE for x_0 and S after the analysis described in section 5.3.

5.3 A Real Data Set

Figure 5.8 shows the trajectory of a membrane potential from a homogeneously spiking neuron such that we may assume to observe iid ISIs. We have already seen a part of this data set in figures 1.3 and 1.1 at the beginning of chapter 1. We are very grateful to Werner Kilb from the Institute of Physiology and Pathophysiology, University of Mainz, for providing this beautiful data set. The ISIs are long enough such that we are able to apply the nonparametric estimators for the diffusion coefficients. Unfortunately, only 46 ISIs are available. Nevertheless, we want to apply the methods provided above to the data in order to show their relevance, and from the insights of section 5.2 we may expect at least some reliable result for $\theta_2 = S$. The structure of the data is the following:

- time frequency: $\Delta = 2 \cdot 10^{-4}\text{s} = 0.2\text{ms}$
- time period: $0 - 54.6\text{s}$ including 46 ISIs
- accuracy of voltage measurement: 0.001mV
- potential minimum and maximum: -52.396mV, -1.1879mV

This membrane potential was recorded in vitro from a neuron belonging to a cortical slice preparation from a 6 week old mouse. The network in this cortical slice was stimulated by a potassium solution (level: 5mM K^+). The potential was electrically stabilized such that the potential between spikes sticks to some region. This is important to fulfill the assumptions of homogeneity for the diffusion process (1.1.1). In many data sets, homogeneity in time does not hold and every estimation method fails. This stabilization only effects movements of the potential that last some seconds. The infinitesimal movements of the potential are not affected. In fact, we will see that this stabilization does not destroy the characteristics of the process. We will show the data to have a very nice diffusion behavior.

5.3.1 Nonparametric Estimation

For the nonparametric estimation of diffusion and drift coefficient, as described in section 5.1, we choose $h = 0.01$ and consider the estimates only at points x where $OT(x) \geq 2000$. So we may expect reliable results at these points. Figure 5.9 shows the results of $\widehat{\sigma^2}$ for the diffusion coefficient and figure 5.10 the results of $\widehat{\beta}$ for the drift coefficient, with varying $M = 10, 15, 20, 25, 30$. The solid diamonds represent the results with the triangular kernel and the transparent diamonds represent the results with the rectangular kernel defined in (5.1.3). For different M and different kernels we observe nearly identical results and very nicely clustered points. Note that there is a large variability in M. So with (5.1.4) in mind, these results strongly support the diffusion hypothesis. The results for $M = 15$ represented by the red diamonds, seem to be stable and representative of the whole cluster. So we choose $M = 15$ and perform a linear regression for these points in order to determine the coefficients. The results are

$$\widetilde{\sigma^2}(x) = 3.09 \cdot (x + 53.74)$$
$$\widetilde{\beta}(x) = -28.77 \cdot (x + 53.74) + 94.34$$

where the coefficient of determination for $\widetilde{\sigma^2}(\cdot)$ is $\rho^2 = 0.83$. For $\widetilde{\beta}(\cdot)$ it is $\rho^2 = 0.93$. Evidently, we propose the CIR model (4.3.2) to be the right model, which leads to the following values for the parameters: $\hat{c} = -53.74$[mV], $\hat{a} = 94.34$[mV/s], $\hat{b} = 28.77$[1/s], $\hat{\sigma} = 1.76[\text{mV}/\sqrt{\text{s}}]$.

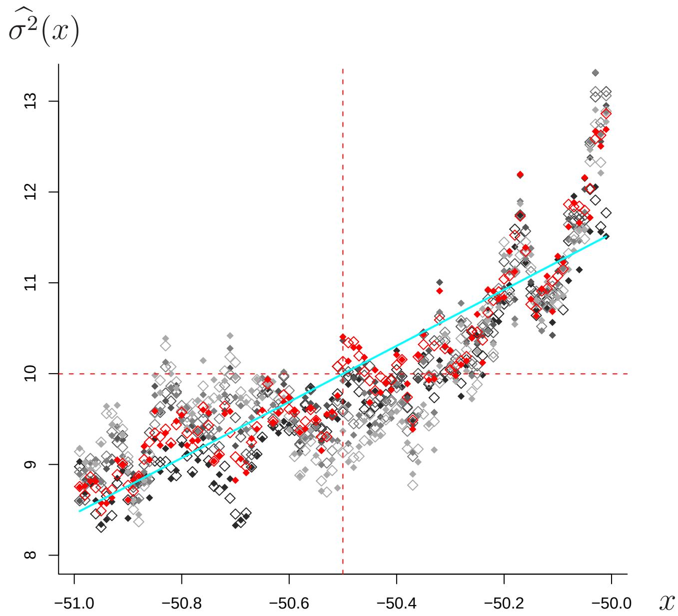

Figure 5.9: Results of the nonparametric estimator $\widehat{\sigma^2}$ for different $M = 10, 15, 20, 25, 30$ plotted in different levels of gray. The red diamonds are the results for $M = 15$ for which a linear fit is given ($\rho^2 = 0.83$).

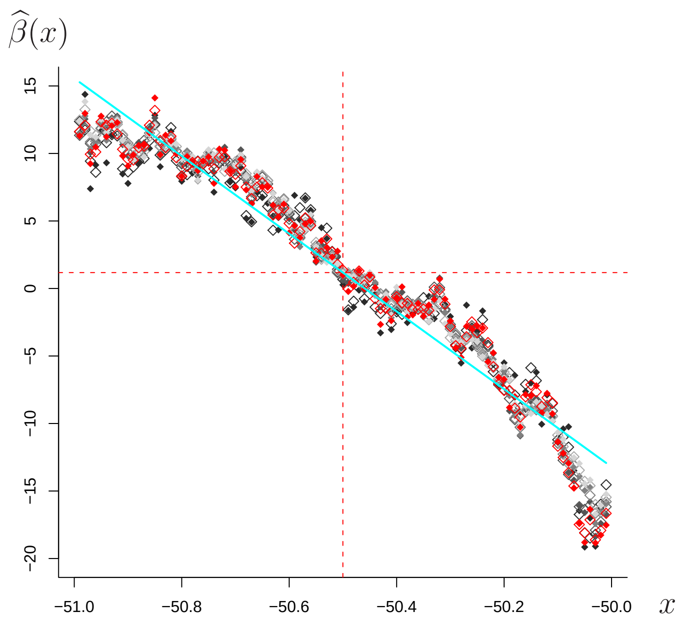

Figure 5.10: Results of the nonparametric estimator $\widehat{\beta}$ for different $M = 10, 15, 20, 25, 30$ plotted in different levels of gray. The red diamonds are the results for $M = 15$ for which a linear fit is given ($\rho^2 = 0.93$).

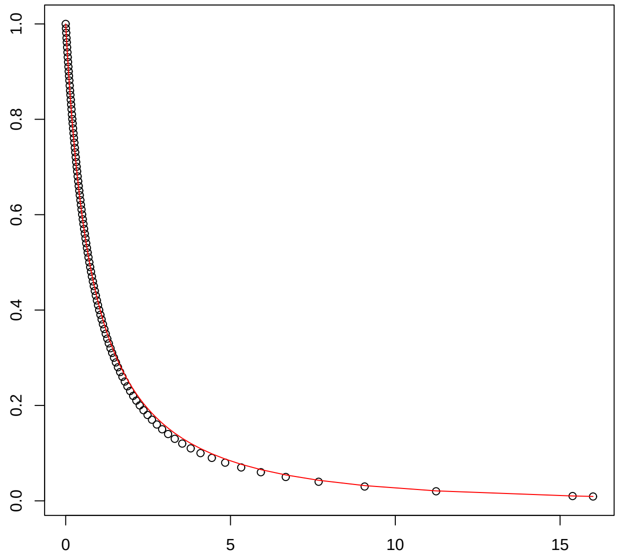

Figure 5.11: The black rings are the values of the empirical LT $\hat{\mathfrak{L}}_n$ for $n = 46$ at the points α_j, $j = 1, \ldots, 100$ which are the mass points of μ. The red line is the estimated LT $\mathfrak{L}_{\theta^*}$ for $\theta^* = (-53.74, -49.28)$ that minimizes the distance of these two functions in $\mathcal{H}$, where $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.081$.

5.3.2 MDE Results for x_0 and S

Now the statistical experiment is determined and we are able to apply the MDE method for the CIR case. The ISI data is simply represented by the times between consecutive spike maxima. The size of our data set is $n = 46$. Again we define a finite lattice $\Theta^L \subset \Theta$ on the parameter space

$$\Theta^L := \{-53.74, -53.73, -53.72, \ldots, -50\} \times \{-51, -50.99, \ldots, -47\} \cap \Theta.$$

Note that $\hat{c} = -53.74$ is the lower bound of the state space of X and consequently it is a natural lower bound for the parameter space Θ. The measure μ is chosen as described in (5.1.5) with $m = 100$. The result of the MDE after 46 ISI observations is

$$\theta^* = (-53.74, -49.28)$$

which is visualized by the dashed lines in figure 5.8. For the excitation threshold the MDE finds a reasonable value $S = \theta_2 = -49.28$ and by reason of the performance analysis in section 5.2, 46 ISIs should be enough to get some reliable result. However, the estimated reset value $x_0 = \theta_1 = -53.74$ is surely not reliable. Moreover, it is just the lower bound of the state space of X. Nevertheless, this might be due to the fact that the *refractory period* after each spike affect the MDE for $\theta_1 = x_0$. Since we just took the times between spike maxima representing the ISIs we have not considered the refractory periods. Hence, the MDE should try to compensate this by looking for a lower value of x_0.

Anyway, recalling that μ takes $m = 100$ points into account and only 46 observations are given, the fit of the estimated LT to the empirical LT is very good and visualized in figure 5.11. The $\mathcal{H}$-distance between these functions is $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.081$. Certainly, using simulated data in section 5.2 we achieved closer fits, but this is not surprising since simulated data represents our model very exactly. However, the model is only an approximation to nature. For this reason, we may not expect better results.

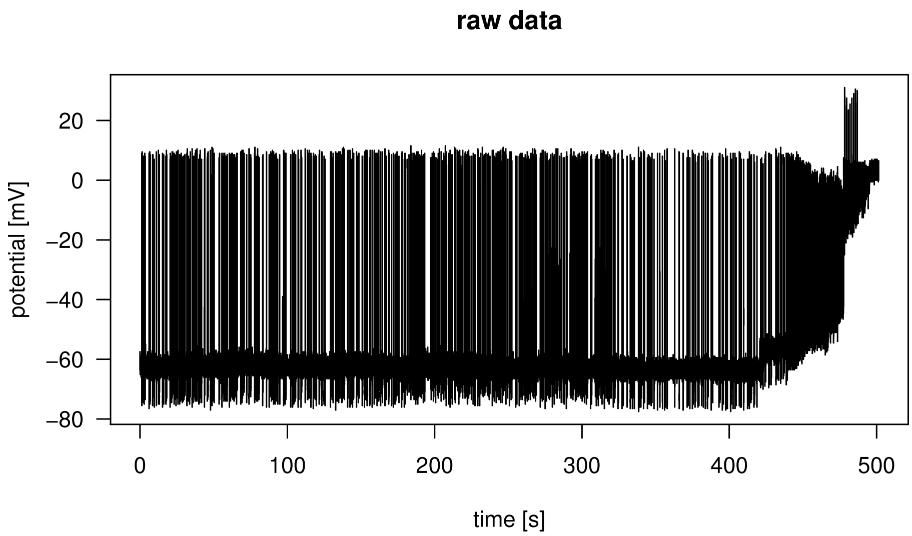

Figure 5.12: In vivo membrane potential data of an astonied Guinea pig. This data set was provided by Petr Lansky, Pavel Sanda and Jufang He and investigated in [LSH 06].

5.4 On a Data Set Considered by Lansky, Sanda and He

Figure 5.12 shows in vivo membrane potential data of an astonied Guinea pig. We are very grateful to Petr Lansky and Pavel Sanda from the Institute of Physiology, Academy of Sciences of the Czech Republic and to Jufang He from the Department of Rehabilitation Sciences, Hong Kong Polytechnic University for providing this nice huge data set including lots of ISIs. The structure of the data set is the following:

- time frequency: $\Delta = 1.5 \cdot 10^{-4}\text{s} = 0.15\text{ms}$ (high frequency)
- time period: $0 - 501.4312\text{s}$
- accuracy of voltage measurement: $0.0005\text{V} = 0.5\text{mV}$
- potential minimum and maximum: -77.5mV, 31mV

During the recording, the neuron has been stimulated by a classical injection of current. This stimulus is shown in figure 5.13. In section 5.1 we already mentioned the problem of ensuring the iid assumptions for real data. The neuron gets "tired" after some time and the spike frequency decreases such that the iid assumption for the ISIs is not fulfilled. Somehow the strong stimulus in the middle and at the end of recording seems to be a good possibility to keep the neuron "awake" because in absence of the strong stimulus

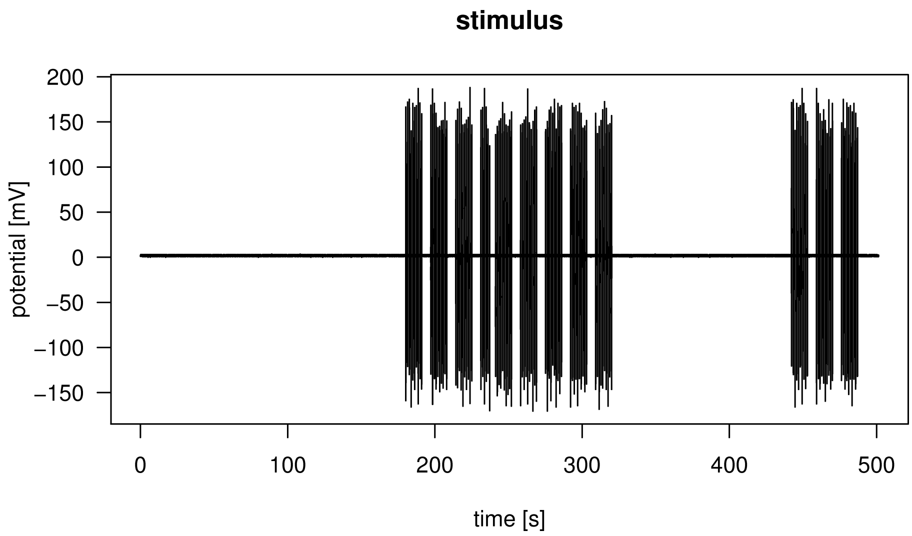

Figure 5.13: In order to stimulate the neuron, the voltage shown above was internally applied during the recording of its membrane potential shown in figure 5.12.

on the intervals $[0, 170]$s and $[330, 410]$s in figure 5.12 the data looks quite homogeneous such that we may assume iid ISIs.

5.4.1 The Raw Data

This data was investigated first by Lansky, Sanda and He in [LSH 06]. For the purpose of their study they selected only the data entirely outside the stimulation period. This is important for estimation methods assuming stationarity of the process. Consequently, also our investigations only consider the data on the intervals $[0, 170]$s and $[330, 410]$s. By reason of the poor accuracy of voltage measurement (0.5mV), the raw data includes crucial rounding errors. Therefore, we may not expect reliable results of the kernel estimators (5.1.1), (5.1.2). Anyway, we are going to apply these estimators. Hence, we have to choose the bandwidth to be $h = 0.5$mV or $h = 1$mV. A smaller h makes no sense due to the accuracy. A larger bandwidth gives results that cannot be interpreted anymore as a trajectory of a diffusion or drift function. The results of the kernel estimators seem to be stable if we vary the step sizes M between 12 and 20. So there is an indication for the data to follow a diffusion process (cf. (5.1.4)). However, it is not an Ornstein-Uhlenbeck process. The result for the diffusion coefficient looks somehow bowl shaped. Also the drift coefficient is quite far away from a linear function and even not monotonically decreasing. So we conclude that this raw data can not be interpreted by any model of our consideration (OU, CIR, BMD, GBM).

5.4.2 The Smoothed Data and the Results of Lansky Sanda and He

To compensate the rounding errors Lansky, Sanda and He smoothed the raw data by a moving average over six time steps. They chose six to be a suitable number of time steps, since for this number their estimation results remained stable (cf. [LSH 06] p.216, Table 1). Thanks to this averaging the smoothed data has a higher "accuracy" in values of voltage. To get separated segments between spikes, the spikes were cut out of the data. For this purpose Lansky, Sanda and He searched for a reset value and an excitation threshold by following a pragmatic procedure, already mentioned in chapter 1, p. 5 and described in [LSH 06] p. 215. Accordingly every segment between these points was cut out and shifted such that the starting point of every ISI is 0. The obtained data has the following structure:

- time frequency: $\Delta = 1.5 \cdot 10^{-4}\text{s} = 0.15\text{ms}$
- time period: $\approx 280\text{s}$ (low stimulated data) including 312 ISIs
- "accuracy" of voltage measurement: 0.001mV
- potential minimum and maximum: -1.5mV, 18.333mV

As mentioned above, Lansky Sanda and He determined for each interval a "rest value" and an "excitation threshold", whereupon they computed the median values given by $x_0 = -73.92[\text{mV}]$ and $S = -61[\text{mV}]$ (cf. [LSH 06] p.219). Consequently, their idea of reset value and excitation threshold for the shifted and smoothed data was $x_0 = 0$ and $S = 12.92$. Further, they assumed this smoothed data to be a discrete observation of an Ornstein-Uhlenbeck process (4.2.1). For each segment they performed a maximum likelihood estimation for a, b and σ. Moreover, an estimator for σ based on squared increments was introduced, but the differences between the results of this estimator and the results of the maximum likelihood estimator were negligible. Further, they used a regression method in order to estimate a and b. Finally, the median value of the corresponding results of each segment was taken which led to the following result: The estimated diffusion coefficient was $\hat{\sigma} = 13.505[\text{mV}/\sqrt{\text{s}}]$. The estimates for the drift parameters based on the regression method were $\hat{a} = 284.6[\text{mV/s}]$ and $\hat{b} = 25.8042[1/\text{s}]$. Further, the maximum likelihood method involved $\hat{a} = 460.6[\text{mV/s}]$ and $\hat{b} = 43.5068[1/\text{s}]$. All these results are given in [LSH 06] p.218.

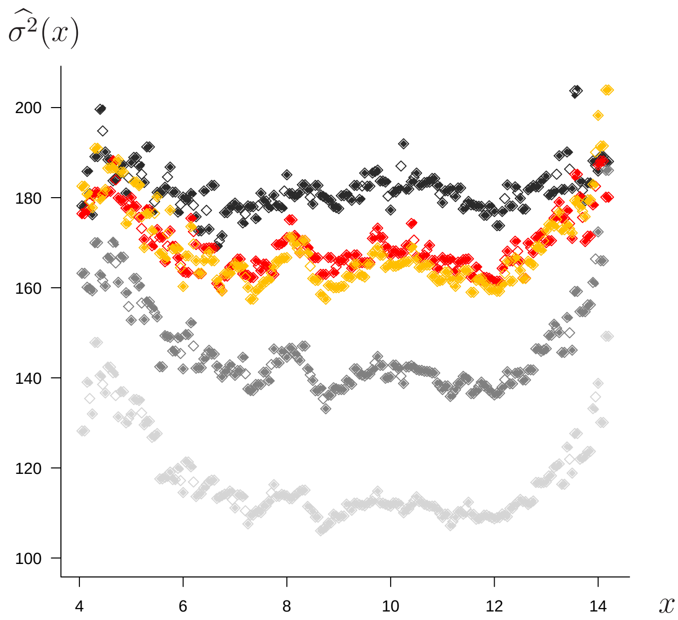

Figure 5.14: Results of the nonparametric estimator $\widehat{\sigma^2}$ top down for $M = 1, 3, 5, 7, 9$. The red diamonds are the results for $M = 3$ and the orange diamonds for $M = 5$.

5.4.3 Nonparametric Estimation Concerning the Smoothed Data

To investigate the smoothed data with the methods provided in section 5.1 and in order to apply in particular the kernel estimators, we have to care again about the bandwidth h. As a matter of fact, by averaging over six values we have vacancies in the range of the data larger than the "accuracy" of 0.001mV. It turns out that $h = 0.05$mV or $h = 0.1$mV are suitable choices and the results are nearly the same whether we use $h = 0.05$mV or $h = 0.1$mV. Further, we only consider the estimates at points x such that $OT(x) \geq 2000$, in order to have results which are as reliable as possible. If we vary $M = 1, 3, 5, 7, 9$ the kernel estimator $\widehat{\sigma^2}$ for the diffusion coefficient gives results as drawn in figure 5.14. Again, in all figures the solid diamonds represent the results with the triangular kernel and the transparent diamonds represent the results with the rectangular kernel, both defined in (5.1.3). Further, in all pictures the results for $M = 1$ are black, the results for $M = 3$ are red, the results for $M = 5$ are orange and $M = 7, 9$ are levels of gray. If we increase the value of $M \geq 10$ the estimates become more and more bowl shaped and tend to 0 in the center. So the estimates are stable only for $3 \leq M \leq 5$. However, if the process between spikes is a diffusion process, the estimates should be stable in some larger region of M. This can

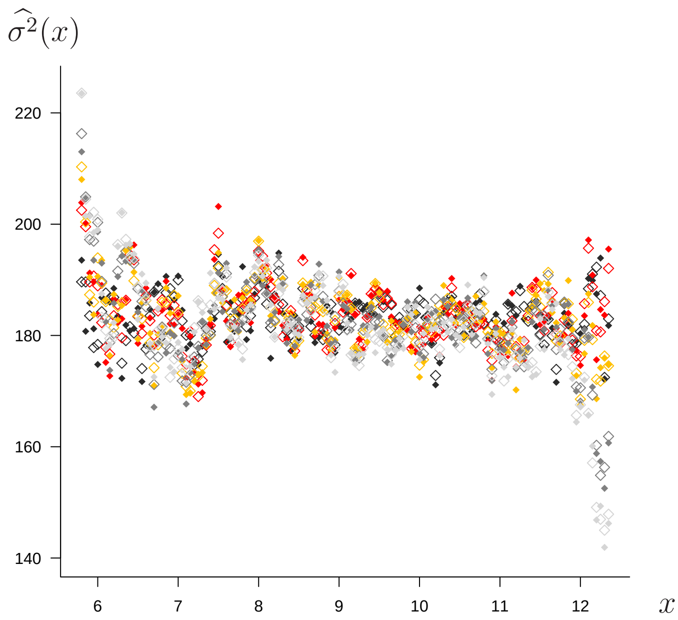

Figure 5.15: Results of the nonparametric estimator $\widehat{\sigma^2}$ for simulated Ornstein-Uhlenbeck data corresponding to the results shown in figure 5.14. The data was simulated with the parameters $\hat{\sigma} = 13.505[\mathrm{mV}/\sqrt{\mathrm{s}}]$, $\hat{a} = 284.6[\mathrm{mV/s}]$, $\hat{b} = 25.8042[1/\mathrm{s}]$, $x_0 = 0[\mathrm{mV}]$ and $S = 12.92[\mathrm{mV}]$, established in [LSH 06].

be visualized by the following.

We simulated an Ornstein-Uhlenbeck neuronal model with the parameters established by Lansky, Sanda and He in [LSH 06] ($\hat{\sigma} = 13.505[\mathrm{mV}/\sqrt{\mathrm{s}}]$, $\hat{a} = 284.6[\mathrm{mV/s}]$, $\hat{b} = 25.8042[1/\mathrm{s}]$, $x_0 = 0[\mathrm{mV}]$, $S = 12.92[\mathrm{mV}]$). Also, the structure of the simulated data (frequency, accuracy) was the same as it is described in subsection 5.4.2 for the smoothed data of Lansky, Sanda and He. Accordingly we followed the same estimation procedure as before to estimate σ^2 with the results shown in figure 5.15. One recognizes that there is essentially no difference between the estimates if we vary $M = 1, 3, 5, 7, 9$. So with (5.1.4) in mind, after comparing figure 5.14 and 5.15 one could doubt the hypothesis that the process between spikes is a diffusion process. Concerning the moving average procedure of Lansky Sanda and He, which destroys diffusion behavior, this is not astonishing.

Anyway, if one still believes in diffusion behavior, the model should have a constant diffusion coefficient. This is due to the fact that in figure 5.14 $\widehat{\sigma^2}$ is nearly flat for $M = 1, 3, 5$. For $M = 1$ the square root of the mean value of $\widehat{\sigma^2}$ is 13.48 which is basically the same value as the estimate of Lansky, Sanda and He. If $M = 3$ the square root of the mean value of $\widehat{\sigma^2}$ is 13.03 and in the case $M = 5$ the corresponding result is 12.99. For these results

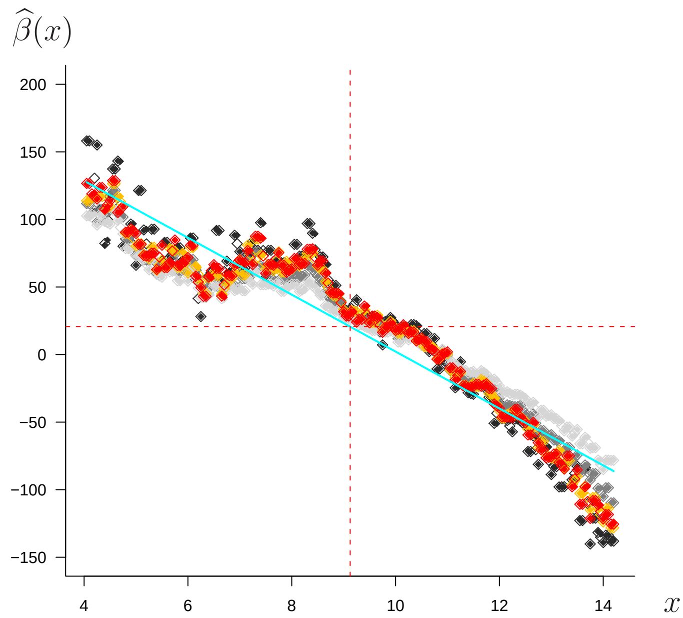

Figure 5.16: Results of the nonparametric estimator $\widehat{\beta}(x)$ for $M = 1, 3, 5, 7, 9$. The red diamonds are the results for $M = 3$ for which a linear fit is given ($\rho^2 = 0.95$).

after rounding it does not make a difference whether the square root of the mean or the mean of the square root is taken. Further, all three results are not very far away from the estimate of Lansky, Sanda and He. Anyway, we propose the results for $M = 3$ to be most stable and representative. For the drift coefficient the kernel estimator $\widehat{\beta}$ gives the results shown in figure 5.16. Obviously, the estimates in all cases $M = 1, 3, 5, 7, 9$ are clustered around the same line. So from this picture we could never reject the diffusion hypothesis. A linear regression for the results of the kernel estimator for $M = 3$ leads to a coefficient of determination of $\rho^2 = 0.95$ and the result

$$\widetilde{\beta}(x) = -21.06 \cdot x + 212.78.$$

By reason of the previous investigation we propose the Ornstein-Uhlenbeck model with the parameters $\sigma = 13.03[\mathrm{mV}/\sqrt{\mathrm{s}}]$, $b = 21.06[1/\mathrm{s}]$ and $a = 212.78[\mathrm{mV}/\mathrm{s}]$ to be most suitable to explain the smoothed data. The mean of the invariant law of the OU process is $a/b = 10.1\mathrm{mV}$. Note that all these estimates are similar to the results of Lansky, Sanda and He in [LSH 06].

5.4.4 MDE Results for x_0 and S

Now the statistical experiment in order to estimate $\theta_1 = x_0$ and $\theta_2 = S$ is determined and it is possible to apply the MDE method for the OU case. The ISI data used for the evaluation of the MDE are the time length of the segments that Lansky, Sanda and He found after cutting out the spikes as it is described above. The size of our data set is $n = 312$ which is enough to expect at least a reliable result for the excitation threshold $S = \theta_2^*$ (cf. 5.2). To evaluate the MDE, we follow the procedure for application of the MDE described in section 5.1 and define again a finite lattice $\Theta^L \subset \Theta$ on the parameter space by

$$\Theta^L = \Theta_1^L \times \Theta_2^L := \{-20, -19.9, \ldots, 9.9, 10\} \times \{5, 5.1, \ldots, 24.9, 25\} \cap \Theta.$$

Further, the measure μ is chosen as described in (5.1.5) with $m = 100$ again.

MDE Results Based on the Parameter Configuration found by the Nonparametric Estimators: We use the parameters based on the results of the kernel estimators from above ($a = 212.78$[mV/s], $b = 21.06$[1/s] and $\sigma = 13.03[\mathrm{mV}/\sqrt{\mathrm{s}}]$) in order to fix the statistical experiment for θ. The corresponding result of the MDE after 312 ISI observations is

$$\theta^* = (2.4, 14.6)$$

This is not a too far away from what Lansky, Sanda and He expected ($x_0 =$ 0[mV] and $S = 12.92$[mV]), since the difference $\theta_1^* - \theta_2^* = 12.2$ is similar to their result. Further, the distance between empirical and estimated LT is $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.019$, which implies a very good fit. This can be seen in figure 5.17.

MDE Results Based on the Parameter Configuration Established in [LSH 06]: If we use the parameters found by Lansky, Sanda and He ($\hat{\sigma} = 13.505[\mathrm{mV}/\sqrt{\mathrm{s}}]$, $\hat{a} = 284.6$[mV/s], $\hat{b} = 25.8042$[1/s]) in order to fix our statistical model, the MDE finds

$$\theta^* = (-4.7, 15.4)$$

which is quite far away from what Lansky, Sanda and He assumed ($x_0 =$ 0[mV] and $S = 12.92$[mV]). The distance between empirical and estimated LT is $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.035$.

If we use instead the drift parameter estimates of Lansky, Sanda and He based on the maximum likelihood method ($\hat{a} = 460.6$[mV/s] and $\hat{b} =$ 43.5068[1/s]), the MDE finds

$$\theta^* = (-4.7, 14.4)$$

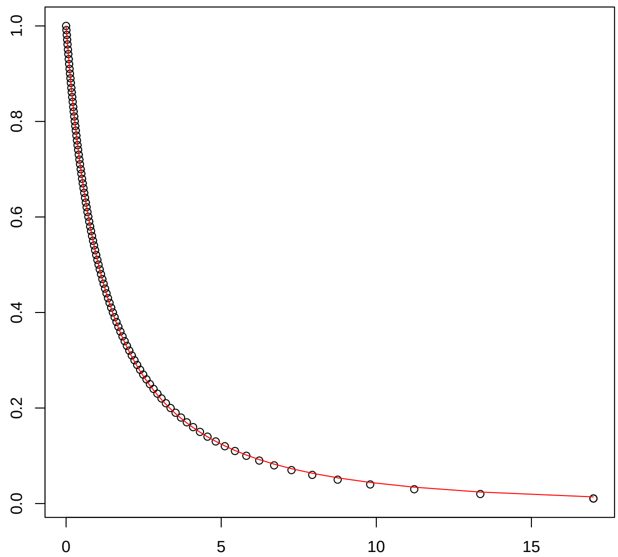

Figure 5.17: The black rings are the values of the empirical LT $\hat{\mathfrak{L}}_n$ for $n = 312$ at the points α_j, $j = 1, \ldots, 100$ which are the mass points of μ. The red line is the estimated LT $\mathfrak{L}_{\theta^*}$ for $\theta^* = (2.4, 14.6)$. This estimate is based on the parameter configuration found by the kernel estimators. θ^* minimizes the distance $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.019$.

which is a little closer to the levels expected by Lansky, Sanda and He. However, the distance $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.051$ is greater than it was for the result in the previous case. From section 5.2 we know that we cannot expect reliable results for θ_1^*. For this reason, the coincidence of both results for the excitation threshold from above is a peculiar fact.

MDE Results Based on the Parameter Configuration proposed by Ditlevsen and Lansky: In [DL 07] Ditlevsen and Lansky reinvestigated this data with a different estimation method based on the so called *Fortet's integral equation*. In this paper they assumed $S - x_0 = 11$[mV] and further $\hat{b} = 25.8042$[1/s] or $1/\hat{b} \approx 0.039$[s] respectively (cf. [DL 07] p.3). Their estimation led to the parameters $\hat{\sigma} = 5[\mathrm{mV}/\sqrt{\mathrm{s}}]$, $\hat{a} = 240$[mV/s]. If we use this configuration for our statistical model, the MDE finds

$$\theta^* = (5.7, 11).$$

The result for $x_0 = \theta_1 = 5.7$ is far away from what Ditlevsen and Lansky expected but maybe not reliable anyway. The estimation $S = \theta_2 = 11$ coincides with the assumption of Ditlevsen and Lansky in this case. Moreover, the distance $\|\hat{\mathfrak{L}}_n - \mathfrak{L}_{\theta^*}\|_{\mathcal{H}} \approx 0.018$ leads to a very good fit.

Conclusion: Since the data might not satisfy the diffusion hypothesis which can be rejected by comparing figure 5.14 and 5.15, we conclude that all results corresponding to the different configurations might not be reliable. If we believe in the diffusion hypothesis we conclude the following.

In every case, $a/b < \theta_2$ holds which implies the subthreshold regime. Hence, from the explanations in section 5.2 p. 79, we know that S is a crucial parameter for the model and can be estimated accurately. Consequently, $n = 312$ should be large enough such that we may assume $S = \theta_2^*$ to be a reliable result for every configuration of parameters given above. $x_0 = \theta_1^*$ is rather not reliable but of secondary importance for the model anyway.

Lansky, Sanda and He in [LSH 06] fixed $x_0 = 0$[mV] and $S = 12.92$[mV] in order to estimate the parameters a, b and σ. Since for this configuration of parameters the MDE finds values different from $x_0 = 0$[mV] and $S = 12.92$[mV], the cross-check of their estimates fails in our framework. For this reason, the pragmatic approach of Lansky, Sanda and He in order to find S might not be the right strategy. Since S is a crucial parameter, a model calibrated by their estimates for a, b, σ, x_0 and S would lead to a spiking behavior essentially different from what we observe in the real data set.

The most important conclusion is that in all cases the MDE finds suitable values for x_0 and S in order to calibrate the model. For all parameter configurations the estimated LT fits the empirical LT very well. For this reason, the model in the setting $x_0 = \theta_1^*$ and $S = \theta_2^*$ is able to reproduce very well the characteristics of spiking behavior as it is observed in real data. Since mimicking of spiking behavior with a manageable model is a main goal of biologists, this is an important insight.

Appendix A

Special Functions

A.1 The Confluent Hypergeometric Function

Definition A.1.1 *For* $\alpha, \gamma, z \in \mathbb{C}$, $\gamma \neq 0, -1, -2, \ldots$ *and* $(\alpha)_k := \prod_{i=0}^{k-1}(\alpha + i)$*, the confluent hypergeometric function* $(\alpha, \gamma, z) \mapsto \phi(\alpha, \gamma; z)$ *is defined by*

$$\phi(\alpha, \gamma, z) = \sum_{k=0}^{\infty} \frac{(\alpha)_k}{(\gamma)_k} \frac{z^k}{k!}.$$

Properties: In the following we give some properties of the confluent hypergeometric function and the corresponding references:

- Let $\gamma > 0$ be fixed. For both variables $\alpha, z \in \mathbb{C}$ the confluent hypergeometric function

$$\begin{aligned} &(\alpha, z) \mapsto \phi(\alpha, \gamma; z) \text{ is an entire function and} \\ &\phi(\alpha, \gamma; z) > 0 \text{ for all } \alpha, z \geq 0. \end{aligned} \tag{A.1.1}$$

  (see e.g. [L 73] Ch.9.9.)

- Let $|\arg(z)| < \pi$, then

$$\phi(\alpha, \gamma, z) = \frac{\Gamma(\gamma)}{2\sqrt{\pi}} (z\alpha)^{1/4-\gamma/2} e^{z/2+2\sqrt{z\alpha}} \cdot \{1 + O(\alpha^{-1/2})\}, \quad \text{as } \alpha \to \infty. \tag{A.1.2}$$

  ([Bh 69] Sec.7.4.(18a) p.98)

- The following rule for the derivative in z holds:

$$\frac{\mathrm{d}}{\mathrm{d}z} \phi(\alpha, \gamma; z) = \frac{\alpha}{\gamma} \phi(\alpha + 1, \gamma + 1; z) \tag{A.1.3}$$

  (see e.g. [L 73] eq.(9.9.4))

- The confluent hypergeometric function has the following integral representation

$$\phi(\alpha, \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} \int_0^1 e^{zt} t^{\alpha-1}(1-t)^{\gamma-\alpha-1} \mathrm{d}t, \quad \operatorname{Re}\gamma > \operatorname{Re}\alpha > 0. \tag{A.1.4}$$

(see e.g. [L 73] eq.(9.11.1))

A.2 The Hermite Function

Definition A.2.1 *(cf.[L 73] §10.2) With the convention* $(\Gamma(-n))^{-1} = 0$ *for* $n \in \mathbb{N}_0$*, the Hermite function* $(z, \nu) \mapsto H_\nu(z)$ *is given by*

$$H_\nu(z) = \frac{2^\nu \Gamma(\frac{1}{2})}{\Gamma\left(\frac{1-\nu}{2}\right)} \phi\left(-\frac{\nu}{2}, \frac{1}{2}; z^2\right) + \frac{2^\nu \Gamma(-\frac{1}{2})}{\Gamma\left(-\frac{\nu}{2}\right)} z\phi\left(\frac{1-\nu}{2}, \frac{3}{2}; z^2\right)$$

where $z, \nu \in \mathbb{C}$ *and* ϕ *is the confluent hypergeometric function defined in A.1.1.*

Definition A.2.2 *(cf.[L 73] §10.2) The parabolic cylinder function* $(z, \nu) \mapsto D_\nu(z)$ *is given by*

$$D_\nu(z) = 2^{-\nu/2} \exp(-z^2/4) H_\nu(z/\sqrt{2}),$$

where $z, \nu \in \mathbb{C}$ *and* H *is the Hermite function defined in A.2.1.*

Properties: Now we give some properties of the Hermite function and the corresponding references:

- For both variables $\nu, z \in \mathbb{C}$ the Hermite function

$$\begin{aligned} &(\nu, z) \mapsto H_\nu(z) \text{ is an entire function and} \\ &H_\nu(z) > 0 \text{ for all } z \in \mathbb{R}, \nu \leq 0. \end{aligned} \tag{A.2.1}$$

  This follows by definition immediately from (A.1.1) and well-known properties of the Γ-function (see e.g. [L 73] Ch.10.2).

- For bounded z and $|\arg(-\nu)| \leq \pi/2$,

$$D_\nu(z) = \frac{1}{\sqrt{2}} \exp\left[\frac{\nu}{2}\log(-\nu) - \frac{\nu}{2} - z\sqrt{-\nu}\right] \cdot \{1 + O(|\nu|^{-1/2})\}, \quad |\nu| \to \infty \tag{A.2.2}$$

  holds. (cf. [MOS 66] ch. 8.1.6, p.332)

- For bounded z and $|\arg(-\nu)| \leq \pi/2$,

$$H_\nu(z) = 2^{(\nu-1)/2} \exp\left[\frac{\nu}{2}\log(-\nu) - \frac{\nu}{2} + \frac{z^2}{2} - z\sqrt{-2\nu}\right] \cdot \{1 + O(|\nu|^{-1/2})\} \tag{A.2.3}$$

  holds as $|\nu| \to \infty$. This follows immediately from (A.2.2) and the connection between H and D in A.2.2.

- The derivative of H with respect to z satisfies the equation

$$\frac{\mathrm{d}}{\mathrm{d}z} H_\nu(z) = 2\nu H_{\nu-1}(z). \tag{A.2.4}$$

  (cf. e.g. [L 73] eq.(10.4.3))

- The Hermite function has the following integral representation

$$H_\nu(z) = \frac{1}{\Gamma(-\nu)} \int_0^\infty e^{-t^2-2tz} t^{-\nu-1} \mathrm{d}t, \quad \mathrm{Re}\, \nu < 0. \tag{A.2.5}$$

  (cf. e.g. [L 73] eq.(10.5.1))

Bibliography

[APP 05] L.Alili, P.Patie, J.L.Pedersen: "Representations of the first hitting time density of an Ornstein-Uhlenbeck process." Stoch. Models 21 (2005), No. 4, pp. 967-980.

[B 81] J.R.Barra: "Mathematical Basis of Statistics". Academic Press (1981).

[BS 02] Borodin, Salminen: "Handbook of Brownian Motion". Second Edition, Birkhäuser (2002).

[Bh 69] Herbert Buchholz: "The Confluent Hypergeometric Function". Springer (1969).

[CK 86] Cremers, Kadelka: "On weak convergence of integral functionals of stochastic processes with application to processes taking paths in L_p^E". Stoch. Proc. Appl. Vol 21 (1986), pp. 305-317.

[D 85] R.Davis: "Asymptotic inference when the amount of information is random". Proc. of the Berkeley conference in honor of J.Neyman and J.Kiefer, Vol. II (Berkeley,1983), 841-864, Wadsworth (1985).

[DL 05] S.Ditlevsen, P.Lansky: "Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model". Phys. Rev. E 71, 011907 (2005).

[DL 06] S.Ditlevsen, P.Lansky: "Estimation of the input parameters in the Feller neuronal model". Phys. Rev. E 73, 061910 (2006).

[DL 07] S.Ditlevsen, P.Lansky: "Parameters of stochastic diffusion processes estimated from observations of first-hitting times: Application to the leaky integrate-and-fire neuronal model". Phys. Rev. E 76, 041906 (2007)

[F 66] W.Feller: "An Introduction to Probability Theory and its Applications". Vol. II. Second Edition, John Wiley & Sons (1966).

[FZ 93] D.Florens-Zmirou: "On estimating the diffusion coefficient from discrete observations". J. Appl. Probab. 30 (1993), No. 4, pp. 790-804.

[GLNR 88] Giorno, Lansky, Nobile, Ricciardi: "Diffusion Approximation and First-Passage-Time Problem for a Model Neuron". Biol. Cybern. Vol. 58 (1988), No. 6, pp. 387-404.

[GL 87] V.Genon-Catalot, C.Laredo: "Limit Theorems for the First Hitting Times Process of a Diffusion and Statistical Applications". Scand. J. Statist., Vol. 14 (1987), No. 2, pp. 143-160.

[GL 90] V.Genon-Catalot, C.Laredo: "An Asymptotic Sufficiency Property of Observations Related to the First Hitting Times of a Diffusion". Statist. Probab. Lett., Vol. 9 (1990), No. 3, pp. 253-258.

[GSZ 01] M.T.Giraudo, L.Sacerdote, C.Zucca: "A Monte Carlo Method for the Simulation of First Passage Times of Diffusion Processes ". Meth. Comp. Appl. Prob., Vol. 3 (2001), No. 2, pp. 215-231.

[GY 03] A.Göing-Jaeschke, M.Yor: "A Survey and Some Generalizations of Bessel Processes". Bernoulli 9 (2003), No. 2, pp. 313-349.

[Ha 05] M.Hammer: "Parameterschätzung in zeitdiskreten ergodischen Markov-Prozessen am Beispiel des Cox-Ingersoll-Ross-Modells". (2005) http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hebis:77-11546

[HB 06] R.Höpfner, K.Brodda: "A Stochastic Model and a Functional Central Limit Theorem for Information Processing in Large Systems of Neurons". J. Math. Biol. 52 (2006) pp. 439-457.

[H 0708] R.Höpfner: "Mathematische Statistik". Vorlesungsskript 07/08: http://joguinf.informatik.uni-mainz.de/~hoepfner/Material-MathStat.html

[H 07] R.Höpfner: "On a set of data for the membrane potential in a neuron". Math. Biosci. 207 (2007), No. 2, pp. 275-301.

[IW 89] Ikeda, Watanabe: "Stochastic Differential Equations and Diffusion Processes". Second Edition, North-Holland (1989).

[JS 03] Jacod, Shiryaev: "Limit Theorems for Stochastic Processes". Second Edition, Springer (2003).

[K 06] A.Klenke: "Wahrscheinlichkeitstheorie". Springer (2006).

[KP 99] P.E.Kloeden, E.Platen: "Numerical Solution of Stochastic Differential Equations". corr. 3. printing, Springer (1999).

[KR 75] J.Keilson, H.F.Ross: "Passage time distributions for Gaussian Markov (Ornstein-Uhlenbeck) statistical processes". Selected Tables in Mathematical Statistics, Vol. III, pp. 233-327. AMS, Providence (1975).

[KS 98] Karatzas, Shreve: "Brownian Motion and Stochastic Calculus". Second Edition, Springer (1998).

[Ky 94] Y.Kutoyants: "Identification of Dynamical Systems with Small Noise". Kluwer Academic Publishers Group (1994).

[Ky 04] Y.Kutoyants: "Statistical Inference for Ergodic Diffusion Processes". Springer (2004).

[L 73] N.N.Lebedew: "Spezielle Funktionen und ihre Anwendung". Bibliographisches Institut AG Zürich (1973).

[L 78] M.Loève: "Probability Theory II". 4th Edition, Springer (1978).

[L 04] V.Linetsky: "Computing Hitting Time Densities for OU and CIR Processes: Applications to Mean-reverting Models". Journal of Computational Finance, Vol. 7 / No. 4 (2004) pp. 1-22.

[LL 87] P.Lansky, V.Lanska: "Diffusion Approximation of the Neuronal Model with Synaptic Reversal Potentials". Biol. Cybern. 56 (1987), pp. 19-26

[LiS 01] Liptser, Shiryaev: "Statistics of Random Processes". Springer (2001).

[LS 01] P.Lansky, L.Sacerdote: "The Ornstein-Uhlenbeck Neuronal Model with Signal-Dependent Noise". Physics Letters A 285 (2001), pp. 132-140.

[LSH 06] P.Lansky, P.Sanda, J.He: "The parameters of the stochastic leaky integrate-and-fire neuronal model". J. Comput. Neurosci. 21 (2006), no. 2, pp. 211-223.

[LST 95] P.Lansky, L.Sacerdote, F.Tomassetti: "On the Comparison of Feller and Ornstein-Uhlenbeck Models for Neural Activity". Biol. Cybern. Vol. 73 (1995), No. 5, pp. 457-465.

[LC 69] L.LeCam: "Théorie asymptotique de la décision statistique". Presses de l'Univ. de Montréal, (1969).

[LY 00] L.LeCam, G.L.Yang: "Asymptotics in Statistics". Second Edition, Springer (2000).

[M 84] P.Millar: "A General Approach to the Optimality of Minimum Distance Estimators". Trans. Amer. Math. Soc., Vol. 286, No. 1. (1984), pp. 377-418.

[MOS 66] Magnus, Oberhettinger, Soni: "Formulas and Theorems for the Special Functions of Mathematical Physics". Third Edition, Springer (1966).

[OR 97] L.Overbeck, T.Ryden: "Estimation in the Cox-Ingersoll-Ross Model". Econometric Theory, Vol. 13, No. 3 (1997), pp. 430-461

[PR 99] B.L.S.Prakasa Rao: "Statistical Inference for Diffusion Type Processes". London, Arnold (1999).

[RS 88] Ricciardi, Sato: "First-Passage-Time Density and Moments of the Ornstein-Uhlenbeck Process". J. Appl. Probab., Vol. 25, No. 1 (1988), pp. 43-57.

[RS 69] Roy, Smith: "Analysis of the Exponential Decay Model of the Neuron Showing Frequency Threshold Effects". Bulletin of Mathematical Biophysics Volume 31, (1969).

[RY 91] Revuz, Yor: "Continuous Martingales and Brownian Motion". Springer (1991).

[T 89] Henry C. Tuckwell: "Stochastic Processes in the Neurosciences". CBMS-NSF Regional Conference Series in Applied Mathematics, 56. (SIAM), Philadelphia, (1989).