July 2018

Asymptotic behavior of the Random Logistic Model and of parallel Bayesian logspline density estimators

Konstandinos Kotsiopoulos
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Dynamical Systems Commons, Probability Commons, and the Statistical Theory Commons

Recommended Citation

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
Asymptotic behavior of the Random Logistic Model and of parallel Bayesian logspline density estimators

A Dissertation Presented
by
KONSTANDINOS KOTSIOPOULOS

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2018

Department of Mathematics and Statistics
Asymptotic behavior of the Random Logistic Model and of parallel Bayesian logspline density estimators

A Dissertation Presented
by
KONSTANDINOS KOTSIOPOULOS

Approved as to style and content by:

Erin Conlon, Chair

Luc Rey-Bellet, Member

Markos Katsoulakis, Member

Jonathan Machta, Member

Nathaniel Whitaker, Department Head
Mathematics & Statistics
DEDICATION

To my loving parents, Christos and Dimitra, and my sister Panagiota
ACKNOWLEDGMENTS

The work presented in this dissertation would not have been possible without the guidance of Professors Richard S. Ellis, Erin Conlon, and Alexey Miroshnikov. I’m extremely grateful to them for their support and professional counsel throughout my years as a graduate student and I cannot thank them enough. Professor Ellis, my primary advisor, has taught me a great deal, both in my professional field and life in general. His knowledge and friendship have shown me not only what an excellent mathematician should be, but how to be a better person. Professor Conlon, whom I’m indebted to for choosing to be my dissertation committee chair, has provided me with far greater insight into the field of statistics. I greatly appreciate all the advice she has given me. Last, but not least, I’d like to thank Alexey Miroshnikov, whose guidance and determination have allowed me to hone my skills and grant me the persistence and fortitude necessary to tackle any problem.

I would also like to thank the rest of my committee members, Professors Markos Katsoulakis, Luc Rey-Bellet and Jonathan Machta, who have all played a role in my professional career through the courses they have taught and their frequent advice. I’m especially grateful to Professor Machta, since half my dissertation would not have been possible without him.

At this point I’d like to thank other professors in the department, whose advice and teaching has been appreciated. Professors Tom Braden, Krista J. Gile, Nestor Guillen, Panos Kevrekidis, Michael Lavine, Bruce Turkington, Siman Wong, Robin Young, and especially my dear friend Stathis Charalampidis, who has been with me every step of the way and supported me all these years. Other members in the department I’d like to thank for their hard work and dedication are Ilona Trousdale and Jacob Lagerstrom.

Finally, a big thank you to my family and friends, since they are the reason that made this pursuit meaningful. My parents, Christos and Dimitra, my sister Panagiota, my nephew George, my grandmother Alexandra, my uncle Tony and his family, my uncle Chris and his family, my friends and colleagues Michael Boratko, Kostis Gourgoulias, Weilong Hu, Tangxin Jin, Shuaimin Kang, Vicky Laina, Jon Maack, Dan Nichols, Giorgos Tsolias, Tassos Vogiannou and Joy Yu. Last, but not least, I’m grateful to Kathryn Aloisio for believing in me and standing side-by-side with me while conducting this work. Her input and advice has meant the world to me.
ABSTRACT

ASYMPTOTIC BEHAVIOR OF THE RANDOM LOGISTIC MODEL AND OF PARALLEL BAYESIAN LOGSPLINE DENSITY ESTIMATORS

MAY 2018

KONSTANDINOS KOTSIOPOULOS
B.S., UNIVERSITY OF CRETE, GREECE
M.S., UNIVERSITY OF MASSACHUSETTS AMHERST
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professors Richard S. Ellis, Erin Conlon and Alexey Miroshnikov

This dissertation is comprised of two separate projects. The first concerns a Markov chain called the Random Logistic Model. For $r \in (0, 4]$ and $x \in [0, 1]$ the logistic map $f_r(x) = rx(1-x)$ defines, for $t \in \mathbb{N}$, the dynamical system $x_r(t+1) = f(x_r(t))$ on $[0, 1]$, where $x_r(1) = x$. The interplay between this dynamical system and the Markov chain $x_{r,N}(t)$ defined by perturbing the logistic map by truncated Gaussian noise scaled by $1/\sqrt{N}$, where $N \to \infty$, is studied. A natural question is whether one can quantify this interplay via probabilistic limit theorems for $x_{r,N}(t)$.

There are two possible limits: the vanishing-noise limit $N \to \infty$ for fixed $t \in \mathbb{N}$, then taking $t \to \infty$, and the ergodic limit $t \to \infty$ followed by the vanishing-noise limit $N \to \infty$. Both lead to a set of probabilistic limit theorems where the underlying deterministic dynamics take over. A particular case of interest is for $r = 4$.

In the second project we perform an asymptotic analysis of Bayesian parallel density estimators which are based on logspline density estimation presented in [25]. The parallel estimator we introduce is in the spirit of the kernel density estimator presented by Neiswanger, Wang and Xing [17]. We provide a numerical procedure that produces the density estimator itself in place of the sampling algorithm. We derive an error bound for the mean integrated squared error for the full data posterior estimator and investigate the parameters that arise from the logspline density estimation and the numerical approximation procedure. Our investigation leads to the choice of parameters that result in the error bound scaling appropriately in relation to them.
# CONTENTS

ACKNOWLEDGMENTS ...................................................................................... v

ABSTRACT ....................................................................................................... vi

CHAPTER

1. INTRODUCTION ......................................................................................... 1

1.1 Random Logistic Model .......................................................................... 1

1.2 Parallel Bayesian logspline estimators ................................................... 6

2. PROPERTIES OF THE LOGISTIC MODEL MARKOV CHAIN ................. 10

3. PROPERTIES OF THE IN Variant DENSITY \( h_{r,N} \) .............................. 16

4. FIRST RESULTS: CONVERGENCE ALMOST SURELY AND IN PROBABILITY .... 28

4.1 Almost sure convergence ......................................................................... 28

4.2 Convergence rates in probability ............................................................ 32

5. PROOF OF THE MAIN THEOREM: WEAK CONVERGENCE ............... 38

5.1 Generalized Frobenius-Perron operator ............................................... 38

5.2 Main result .............................................................................................. 47

6. PARALLEL BAYESIAN LOGSPINE ESTIMATORS: NOTATION & HYPOTHESES . 79

7. ANALYSIS OF MISE FOR \( \hat{p} \) .................................................................. 82

7.1 Error analysis for unnormalized estimator ............................................. 82

7.2 Analysis for renormalization constant ..................................................... 86

8. NUMERICAL ERROR .................................................................................. 89

8.1 Interpolation of an estimator: preliminaries ........................................... 89

8.2 Numerical approximation of the renormalization constant \( \hat{c} = \hat{\lambda}^{-1} \) .... 93

9. NUMERICAL EXPERIMENTS ................................................................... 100
The work presented in this dissertation consists of two separate projects. The first can be categorized under the fields of probability and dynamical systems, where the model of interest is a Markov chain we call the Random Logistic Model. The second is placed under the field of mathematical statistics, where the focus is on the model found in [17] and an algorithm, the direct density product method, which was presented in [16]. Although unrelated, the connecting thread between the two projects is the asymptotic analysis. For the Random Logistic Model, we obtain a set of probabilistic limits that show the interplay between the stochastic model and the underlying deterministic dynamics in the ergodic limit and the vanishing noise limit. For the direct density product method, we provide the framework and obtain error estimates for the algorithm and how that scales for large number of samples. Below we introduce the two projects in detail and present our main results.

1.1 Random Logistic Model

Ever since the biologist Robert May popularized the logistic map in 1976 [15], its wide usage in various fields has grown tremendously. Starting as a model for population dynamics, it has been successfully applied to model turbulent flows [19] and has been used as a tool in cryptography [29] and as a model of critical phenomena in statistical mechanics [20], just to name a few. The logistic map is defined as \( f_r(x) = rx(1-x) \), where \( r \) is a parameter whose values are restricted in the interval \((0, 4]\). For \( r \in (0, 4] \) \( f_r \) maps \([0,1]\) into \([0,1]\) and exhibits a wide variety of complicated behaviors on that interval, which includes an infinite sequence of bifurcations of the set of fixed points of \( f_r \) as \( r \) increases. This behavior is summarized at the start of the next chapter. The logistic map as a dynamical system is then defined for \( t \in \mathbb{N} \) and \( x \in [0,1] \) by

\[
x_r(t+1) = f_r(x_r(t)), \quad x_r(1) = x.
\]

Clearly, for \( t \in \mathbb{N} \) satisfying \( t \geq 2 \), \( x_r(t) = f^{(t-1)}(x) \).

The work presented in this dissertation focuses on the interplay between this dynamical system and a random system defined by perturbing the logistic map \( f_r \) by suitably scaled, additive white noise. The study of such random systems has been carried out in many papers in the literature in-
cluding [9, 30]. The main idea in these papers is to study the asymptotic behavior of the system on the interval [0, 1] in the presence of this noise while at the same time trying to maintain control over the iterates $x_r(t)$ so that they never escape from the interval. In most such studies the assumption is that the intensity of the noise is weak enough to guarantee that control. However, from a mathematically rigorous perspective, this assumption is flawed since the additive white noise is a Gaussian random variable. Therefore there is always a positive probability that an iterate escapes the interval \([0, 1]\) no matter how weak the noise is.

To correct this flaw, we construct a process by perturbing the logistic map $f_r$ by an additive sequence $\xi_t$ of truncated Gaussian random variables scaled by $1/\sqrt{N}$. This process is a Markov chain \(\{x_{r,N}(t,\omega)\}\), called the random logistic model and indexed by $r \in \mathbb{N}$, $N \in \mathbb{N}$, $t \in \mathbb{N}$, $x \in [0, 1]$, and $\omega$ lying in a probability space that can be taken to be $\mathbb{R}^N$. The random logistic model is defined by the formula

$$x_{r,N}(t+1,\omega) = f_r(x_{r,N}(t,\omega)) + \frac{1}{\sqrt{N}}\xi_t(f_r(x_{r,N}(t,\omega))), \quad x_{r,N}(1) = x. \quad (1.2)$$

Because of the properties of the truncated Gaussian noise, $x_{r,N}(t)$ lies in $[0, 1]$ for all $t \in \mathbb{N}$. The concept of adding such a noise was proposed in [11]. We give complete details showing that the random logistic model Markov chain is rigorously defined.

A natural question is whether one can quantify the interplay between $x_{r,N}(t)$ and $x_r(t)$ via probabilistic limit theorems for the random logistic model Markov chain. There are two possible limits of the Markov chain: the vanishing-noise limit $N \to \infty$ followed by the ergodic limit $t \to \infty$ and the ergodic limit $t \to \infty$ followed by the vanishing-noise limit $N \to \infty$. Part (a) of our first theorem focuses on the $N \to \infty$ limit of the Markov chain while part (b) considers the double limit $N \to \infty$, then $t \to \infty$. According to part (a), as $N \to \infty$, the almost sure limit of the Markov chain $x_{r,N}(t)$ equals $f_r^{(t-1)}(x)$, which coincides with the deterministic value $x_r(t)$ of the dynamical system at time $t$. This reduction can be seen by formally taking the limit $N \to \infty$ in the definition (1.2) of the Markov chain, yielding an equation having the identical form of the dynamical system in (1.1). For the following let $m$ denote Lebesgue measure on $[0, 1]$.

**Theorem 1.1.** The Markov chain $x_{r,N}(t)$ defined in (1.2) has the following properties.
(a) For fixed \( t \in \mathbb{N} \) satisfying \( t \geq 2 \)

\[
\lim_{N \to \infty} x_{r,N}(t) = f_r^{(t-1)}(x) \text{ P-almost surely for } \omega \in \mathbb{R}^N.
\]

(b) \( \lim_{t \to \infty} \lim_{N \to \infty} x_{r,N}(t) = \lim_{t \to \infty} f_r^{(t-1)}(x) \text{ P-almost surely for } \omega \in \mathbb{R}^N. \)

Theorem 1.1 is based on an inductive argument and the repeated application of standard inequalities that are tight enough to show that for any \( \varepsilon > 0 \)

\[
\sum_{N=1}^{\infty} P(|x_{r,N}(t) - f_r^{(t-1)}(x)| > \varepsilon) < \infty.
\]

This summability condition yields the conclusion of part (a). Part (b) is then answered by considering the limiting behavior of \( x_r(t) \) as \( t \to \infty \) for the various values of \( r \) as outlined in Chapter 2.

A particular case of interest is the case \( r = 4 \), for which the logistic map \( f_4 \) exhibits chaotic behavior in the limit \( t \to \infty \). The chaotic behavior of \( f_4 \) includes the following features.

These are presented in detail in the first section of [6].

- \( f_4 \) has periodic cycles of every integer order.

- For \( m \)-almost every initial value \( x \in [0, 1] \) the limit set of the sequence \( x_4(t) = f_4^{(t-1)}(x) \) as \( t \to \infty \) is the interval \([0, 1]\), i.e. for \( m \)-almost every initial value \( x \in [0, 1] \) and any \( y \in [0, 1] \)

  there exists a subsequence \( x_4(t_j) = f_4^{(t_j-1)}(x) \) for \( j \in \mathbb{N} \) such that \( x_4(t_j) \to y \) as \( j \to \infty \).

- Given \( \varepsilon > 0 \) small enough, for \( m \)-almost any \( x, y \in [0, 1] \) such that \( |x - y| < \varepsilon \), there exists a \( j \in \mathbb{N} \) such that \( |f_4^{(j)}(x) - f_4^{(j)}(y)| > \varepsilon \).

The limit \( \lim_{t \to \infty} f_4^{(t-1)}(x) \) in part (b) of Theorem 1.1 involves the chaotic behavior of the logistic map dynamical system as summarized in the three bullets preceding the statement of the theorem. Clearly, part (b) of Theorem 1.1 does not quantify the chaotic behavior of this dynamical system.

Interestingly, this quantification is the content of part (b) of Theorem 1.2, which studies the asymptotic behavior of the Markov chain \( x_{4,N}(t) \) in the reverse double limit \( t \to \infty \) followed by \( N \to \infty \). The statement of the theorem requires that we fill in some background. In Theorem 2.2 we prove that there exists a probability measure \( \sigma_{r,N} \) on \([0, 1]\) with the following properties:
• $\sigma_{r,N}$ is the unique invariant measure of $x_{r,N}(t)$.

• As $t \to \infty$, $x_{r,N}(t)$ converges in distribution to $\sigma_{r,N}$; i.e., for any continuous function $g$ mapping $[0, 1]$ into $\mathbb{R}$ and any $x \in [0, 1]$

$$\lim_{t \to \infty} E_x(g(x_{r,N}(t))) = \int_{[0,1]} g(x)\sigma_{r,N}(dx).$$

In this formula $E_x$ denotes expectation conditioned on $x_{r,N}(1) = x$. The limit in the second bullet is the content of part (a) of Theorem 1.2 for $r = 4$. Part (b) of Theorem 1.2 involves a new quantity $\sigma^*$, a probability measure on $[0, 1]$ that is the unique invariant measure for the logistic map $f_4$ that is absolutely continuous with respect to Lebesgue measure on $[0, 1]$; i.e., for any Borel subset $A$ of $[0, 1]$, $\sigma^*(f_4^{-1}(A)) = \sigma^*(A)$. The probability measure $\sigma^*$, which was discovered by Ulam and von Neumann [27], has the density $1/\left[\pi \sqrt{x(1-x)}\right]$ on $[0, 1]$. In part (b) we prove that the sequence $\sigma_{4,N}$ converges weakly to $\sigma^*$ as $N \to \infty$. The proof of part (b) is given in Chapter 5 and involves the theory of Frobenius-Perron operators as well as Birkhoff’s pointwise ergodic theorem. In part (c) of the theorem we combine the limits in parts (a) and (b) to show that in the limit $t \to \infty$ followed by $N \to \infty$ the Ulam-von Neumann Markov chain $x_{4,N}(t)$ converges in distribution to the Ulam-von Neumann invariant measure $\sigma^*$.

**Theorem 1.2.** For $N \in \mathbb{N}$, $\sigma_{4,N}$ denotes the unique invariant measure of the Markov chain $x_{4,N}(t)$, the existence of which is proved in Theorem 2.2. The quantity $\sigma^*$ is the Ulam-von Neumann invariant measure for $f_4$. Then for any continuous function $g$ mapping $[0, 1]$ into $\mathbb{R}$ the following conclusions hold.

(a) $\lim_{t \to \infty} E_x(g(x_{4,N}(t))) = \int_{[0,1]} g(x)\sigma_{4,N}(dx)$.

(b) $\lim_{N \to \infty} \int_{[0,1]} g(x)\sigma_{4,N}(dx) = \int_{[0,1]} g(x)\sigma^*(dx)$.

(c) It follows that

$$\lim_{N \to \infty} \lim_{t \to \infty} E_x(g(x_{4,N}(t))) = \int_{[0,1]} g(x)\sigma^*(dx).$$

A similar result was obtained by Katok and Kifer in [12]. They considered random perturbations of the logistic map and obtained a set of weak limits for values of $r$ close to 4, since they were
interested in the case where the limit set of the logistic map contained an interval. In this work, we
apply a technique they used for \( r = 4 \) and present a modified proof for our model. Furthermore,
we consider other values of \( r \) that were not presented in their paper. The weak limit in those cases
expresses the limiting behavior of the deterministic system. There are three other cases considered:

(a) \( r \in (0, 1] \), where \( \lim_{t \to \infty} x_r(t) = 0 \) for \( x_r(1) = x \in [0, 1] \).

(b) \( r \in (1, 3] \), where \( \lim_{t \to \infty} x_r(t) = 1 - \frac{1}{r} \) for \( x_r(1) = x \in (0, 1) \) and \( x_r(t) = 0 \) for all \( t \) when
\( x_r(1) = x \in \{0, 1\} \).

(c) \( x_r(t) \) has a stable \( 2^k \)-cycle \( \{p^{(1)}_{r,k}, \ldots, p^{(2^k)}_{r,k}\} \) when \( r \in (3, r_\infty) \) and \( k = k(r) \) is a positive
integer that depends on \( r \). The constant \( r_\infty \) is called the Feigenbaum number and is approximately
equal to 3.561547...

Specifically for case (c), the \( 2^k \)-cycle satisfies \( f_r \left( p^{(i)}_{r,k} \right) = p^{(i+1)}_{r,k}, \ i = 1, \ldots, 2^k-1 \), and \( f_r \left( p^{(2^k)}_{r,k} \right) =
p^{(1)}_{r,k} \). As for the values of \( k \), there exists a sequence \( \{r_n\}_{n=0}^{\infty} \) such that

- \( k = 1 \) for \( r_0 = 3 < r \leq r_1 = 1 + \sqrt{6} \).
- \( k = n \) for \( r_{n-1} < r \leq r_n, \ n \geq 2 \).
- \( \lim_{n \to \infty} r_n = r_\infty \).

Considering the cases (a), (b) and (c), we obtain the following analogue to Theorem 1.2.

**Theorem 1.3.** For any continuous function \( g \) mapping \([0, 1]\) into \( \mathbb{R} \) and \( x \in [0, 1] \) the following
conclusions hold.

(a) For \( r \in (0, 1] \), \( \lim_{N \to \infty} \lim_{t \to \infty} E_x(g(x_{r,N}(t))) = g(0) \).

(b) For \( r \in (1, 3] \), \( \lim_{N \to \infty} \lim_{t \to \infty} E_x(g(x_{r,N}(t))) = g \left( 1 - \frac{1}{r} \right) \).

(c) For \( r \in (3, r_\infty) \), \( \lim_{N \to \infty} \lim_{t \to \infty} E_x(g(x_{r,N}(t))) = \frac{1}{2^k} \sum_{i=1}^{2^k} g \left( p^{(i)}_{r,k} \right) \).

In Chapter 2 we introduce the random logistic model and prove the existence and uniquenesses
of the invariant measure \( \sigma_{r,N} \) of the associated Markov chain. In Chapter 3 several properties
of the invariant density \( h_{r,N} \) of \( \sigma_{r,N} \) are proved. Chapter 4 presents the details of the almost sure
convergence result as stated in Theorem 1.1, as well as convergence rates for the convergence in
probability. Chapter 5 presents the proof of the weak convergence result in part (b) of Theorem 1.2.
Finally, in the appendices we give full details concerning the measurability of the random logistic
model and present all the tools needed from ergodic theory required to prove part (b) of Theorem 1.2.

1.2 Parallel Bayesian logspline estimators

The recent advances in data science and big data research have brought challenges in analyzing large data sets in full. These massive data sets may be too large to read into a computer’s memory in full, and data sets may be located on different machines. In addition, there is a lengthy time needed to process these data sets. To alleviate these difficulties, many parallel computing methods have recently been developed. One such approach partitions large data sets into subsets, where each subset is analyzed on a separate machine using parallel Markov chain Monte Carlo (MCMC) methods [14, 18, 22]; here, communication between machines is required for each MCMC iteration, increasing computation time.

Due to the limitations of methods requiring communication between machines, a number of alternative communication-free parallel MCMC methods have been developed for Bayesian analysis of big data [16, 17]. For these approaches, Bayesian MCMC analysis is performed on each subset independently, and the subset posterior samples are combined to estimate the full data posterior distributions. Neiswanger, Wang and Xing [17] introduced a parallel kernel density estimator that first approximates each subset posterior density and then estimates the full data posterior by multiplying together the subset posterior estimators. The authors of [17] show that the estimator they use is asymptotically exact; they then develop an algorithm that generates samples from the posterior distribution approximating the full data posterior estimator. Though the estimator is asymptotically exact, the algorithm of [17] does not perform well for posteriors that have non-Gaussian shape. This under-performance is attributed to the method of construction of the subset posterior densities; this method produces near-Gaussian posteriors even if the true underlying distribution is non-Gaussian. Another limitation of the method of Neiswanger, Wang and Xing is its use in high-dimensional parameter spaces, since it becomes impractical to carry out this method when the number of model parameters increases.

Miroshnikov and Conlon [16] introduced a new approach for parallel MCMC that addresses the limitations of [17]. Their method performs well for non-Gaussian posterior distributions and only analyzes densities marginally for each parameter, so that the size of the parameter space is not a
limitation. The authors use logspline density estimation for each subset posterior, and the subsets are combined by a direct numeric product of the subset posterior estimates. However, note that this technique does not produce joint posterior estimates, as in [17].

The estimator introduced in [16] follows the ideas of Neiswanger et al. [17]. Specifically, let $p(x|\theta)$ be the likelihood of the full data given the parameter $\theta \in \mathbb{R}$. We partition $x$ into $M$ disjoint subsets $x_m$, with $m \in \{1, 2, \ldots, M\}$. For each subset we draw $N$ samples $\theta_1^m, \theta_2^m, \ldots, \theta_N^m$ whose distribution is given by the subset posterior density $p(\theta|x_m)$. Given prior $p(\theta)$, the datasets $x_1, x_2, \ldots, x_M$ and assuming that they are independent from each other, then the posterior density, see [17], is expressed by

$$
p(\theta|x) \propto p(\theta) \prod_{m=1}^{M} p(x_m|\theta) = \prod_{m=1}^{M} p_m(\theta) =: p^*(\theta),
$$

where $p(\theta|x_m) := p_m(\theta) = p(x_m|\theta)p(\theta)^{1/M}$.

We investigate the properties of the estimator $\hat{p}(\theta|x)$, defined in [16], that has the form

$$
\hat{p}(\theta|x) \propto \prod_{m=1}^{M} \hat{p}_m(\theta) =: \hat{p}^*(\theta),
$$

where $\hat{p}_m(\theta)$ is the logspline density estimator of $p_m(\theta)$ and where we suppressed the information about the data $x$.

The estimated product $\hat{p}^*$ of the subset posterior densities is, in general, unnormalized. This motivates us to define the normalization constant $\hat{c}$ for the estimated product $\hat{p}^*$. Thus, the normalized density $\hat{p}$, one of the main points of interest in our work, is given by

$$\hat{p}(\theta) = \hat{c}^{-1}\hat{p}^*(\theta), \quad \text{where} \quad \hat{c} = \int \hat{p}^*(\theta) \, d\theta.$$ 

Computing the normalization constant analytically is a difficult task since the subset posterior densities are not explicitly calculated, with the exception of a finite number of points $(\theta_i, \hat{p}_m^*(\theta_i))$, where $i \in \{1, \ldots, n\}$. By taking the product of these values for each $i$ we obtain the value of $\hat{p}^*(\theta_i)$. This allows us to numerically approximate the unnormalized product $\hat{p}^*$ by using Lagrange interpolation. The approximating polynomial is denoted by $\tilde{p}^*$. Then we approximate the constant
by numerically integrating $\tilde{p}^*$. The approximation of the normalization constant $\tilde{c}$ is denoted by $\hat{c}$, given by
\[
\hat{c} = \int \tilde{p}^*(\theta) \, d\theta, \text{ and we set } \hat{p}(\theta) := \hat{c}^{-1} \tilde{p}^*(\theta).
\]
The newly defined density $\hat{p}$ acts as the estimator for the full-data posterior $p$.

We establish error estimates between the three densities via the mean integrated squared error (MISE), defined for two functions $f,g$ as
\[
\text{MISE}(f,g) := \mathbb{E} \int (f(\theta) - g(\theta))^2 \, d\theta.
\]
(1.5)

Thus, our work involves two types of approximations: 1) the construction of $\hat{p}^*$ using logspline density estimators and 2) the construction of the interpolation polynomial $\tilde{p}^*$. The methodology of logspline density estimation was introduced in [25] and corresponding error estimates between the estimator and the density it is approximating are presented in [23, 24]. These error estimates depend on three factors: i) the $N_m$ number of samples drawn from the subset posterior density, ii) the $K_m + 1$ number of knots used to create the $k$-order B-splines, and iii) the step-size of those knots, which we denote by $h_m$.

We estimate the MISE between the functions $\hat{p}^*$ and $p^*$ by adapting the estimation techniques introduced in [23, 24]. We then utilize this analysis to establish a similar estimate for the normalized densities $\hat{p}$ and $p$,
\[
\text{MISE}(p^*, \hat{p}^*) = O \left[ \exp \left( \sum_{m=1}^{M} \frac{K_m + 1 - k}{N_m^{1/2}} + h_{\text{max}}^{j+1} \sum_{m=1}^{M} \left\| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \right\|_{\infty} \right) - 1 \right]^2,
\]
where $h_{\text{max}} = \max_m \{h_m\}$ and $j + 1$ is the number of continuous derivatives of $p$. Notice that the exponential contains two terms, where the first depends on the number of samples and the number of knots and the other depends on the placement of the spline knots. Both terms converge to zero and for MISE to scale optimally both terms must converge at the same rate. To this end, we choose $h_{\text{max}}$ and each $K_m$ to be functions of the vector $\mathbf{N} = \{N_1, \ldots, N_M\}$ and scale appropriately with the norm $||\mathbf{N}||$. This simplifies the above estimate to
\[
\text{MISE}(p^*, \hat{p}^*) = O \left( M^{2-2\beta} ||\mathbf{N}||^{-2\beta} \right)
\]
where the parameter $\beta \in (0, 1/2)$ is related to the convergence of the logspline density estimators.

The estimate for MISE between $\hat{p}^*$ and $\tilde{p}^*$ is obtained in a similar way by utilizing Lagrange interpolation error bounds, as described in [1]. This error depends on two factors: i) the step-size $\Delta x$ of the grid points chosen to construct the polynomial, where the grid points correspond to the coordinates $(\theta_i, \tilde{p}_{m}(\theta_i))$ discussed earlier, and ii) the degree $l$ of the Lagrange polynomial. The estimate obtained is also shown to hold for the normalized densities $\tilde{p}$ and $\hat{p}$.

\[
MISE(\hat{p}^*, \tilde{p}^*) = O \left( \left( \frac{\Delta x}{h_{\text{min}}(N)} M \right)^{2(l+1)} \right),
\]

where $h_{\text{min}}(N)$ is the minimal distance between the spline knots and is chosen to asymptotically scale with the norm of the vector of samples $N$, see Chapter 6.

We then combine both estimates to obtain a bound for MISE for the densities $p$ and $\tilde{p}$. We obtain

\[
MISE(p, \tilde{p}) = O \left( M^{2 - 2\beta} \|N\|^{-2\beta} + \left( \frac{\Delta x}{h_{\text{min}}(N)} M \right)^{2(l+1)} \right).
\]

In order for MISE to scale optimally the two terms in the sum must converge to zero at the same rate. As before with the distance between $\hat{p}^*$ and $p^*$, we choose $\Delta x$ to scale appropriately with the norm of the vector $N$. This leads to the optimum error bound for the distance between the estimator $\tilde{p}$ and the density $p$,

\[
MISE(p, \tilde{p}) = O \left( \|N\|^{-2\beta} \right) \text{ where we choose } \Delta x = O \left( \|N\|^{-\beta \left( \frac{1}{m+1} + \frac{1}{n+1} \right)} \right). \tag{1.6}
\]

The arrangement is as follows. In Chapter 6 we set notation and hypotheses that form the foundation of the analysis. In Chapter 7 we derive an asymptotic expansion for MISE of the non-normalized estimator, which are central to the analysis performed in subsequent sections. We also perform there the analysis of MISE for the full data set posterior density estimator $\hat{p}$. In Chapter 8, we perform the analysis for the numerical estimator $\tilde{p}$. In Chapter 9 we showcase our simulated experiments and discuss the results. Finally, in the appendices we provide supplementary lemmas and theorems employed in Chapters 7 and 8.
CHAPTER 2
PROPERTIES OF THE LOGISTIC MODEL MARKOV CHAIN

The Markov chain that we analyze is associated with a randomization of the well known logistic map, some of the properties of which we now summarize [26, §10.3], [6, 1.1]. For $r \in (0, 4]$ and $y \in [0, 1]$ we define the logistic map
\[ f_r(y) = ry(1 - y), \quad (2.1) \]
and for $t \in \mathbb{N}$ we consider the associated dynamical system $y_r(t + 1) = f_r(y_r(t))$, where $y_r(1) = y$.

For these values of $r$, the dynamical system \( \{y_r(t), t \in \mathbb{N}\} \) is well defined because $f_r$ maps $[0, 1]$ onto $[f_r(0), f_r(1/2)] = [0, r/4]$, which for $r \in (0, 4]$ is a proper subset of $[0, 1]$ and for $r = 4$ equals $[0, 1]$.

As is well known, the structure of the set of fixed points of $f_r$ changes as $r$ increases. For all $r \in (0, 4]$ the point $x = 0$ is a fixed point of the logistic map; it is stable for $0 < r \leq 1$ and unstable for $1 < r \leq 4$. At $r = 1$ the logistic map undergoes a transcritical bifurcation as the fixed point $x_r = 1 - 1/r$ bifurcates from the 0 solution; this new solution is stable for $1 < r \leq 3$ and unstable for $3 < r \leq 4$.

At $r = 3$ the logistic map undergoes a flip bifurcation as a 2-cycle bifurcates from the fixed point $x_3 = 2/3$; this 2-cycle consists of a pair of points $p_r$ and $q_r$ that are given explicitly in terms of $r$ and satisfy $f_r(p_r) = q_r$ and $f_r(q_r) = p_r$. It follows that $p_r$ and $q_r$ are each fixed points of the second-iterate map $f_r^{(2)}(x) = f_r(f_r(x))$. For $3 < r < 1 + \sqrt{6}$ both $p_r$ and $q_r$ are stable fixed points of $f_r^{(2)}$. When $r$ is just over $1 + \sqrt{6}$ the 2-cycle $\{p_r, q_r\}$ becomes repelling and an attracting 4-cycle appears. In general, the behavior becomes much more complex as $r$ increases to 4. There exists an increasing sequence $\{r_n\} \subset (3, 4)$ such that for $r \in (r_{n-1}, r_n]$, $f_r$ has one repelling $2^k$-cycle for $k = 1, \ldots, n - 1$ and one attracting $2^n$-cycle. The sequence $\{r_n\}$ has a limit $r_\infty = 3.561547\ldots$ which is known as the Feigenbaum number. For $r \in (r_\infty, 4)$ the behavior is extremely irregular and we will not go into much detail. The case $r = 4$ is worth special attention for a couple of reasons. First of all, $f_4$ has periodic cycles of any order. Secondly, for $m$-almost every $y \in [0, 1]$ the limit set of $\{y_4(t)\}$ is the interval $[0, 1]$, i.e. for any $y_0$ and for $m$-almost every $y$ in the unit interval there exists a subsequence $\{y_4(t_j)\}$ such that $y_4(t_j) \to y_0$ as $j \to \infty$. Last but not least, $f_4$ is chaotic and it has an invariant measure $\sigma^*$ such that $\sigma^* \ll m$ given in (B.4).
We rigorously construct the random logistic model and study it for \(0 < r \leq 4\), while the weak limit theorems obtained in Chapter 5 are for \(r \in (0, r_\infty) \cup \{4\}\) due to the well-defined nature of the map. The method employed to obtain those results can be applied to isolated values of \(r \in (r_\infty, 4)\). The random logistic model is defined by perturbing the logistic map with Gaussian noise parametrized by a quantity \(N \in \mathbb{N}\); the limit \(N \to \infty\) defines the vanishing-noise limit. This model is defined in terms of a Markov chain \(\{x_{r,N}(t), t \in \mathbb{N}\}\) introduced in (2.3).

The definition of this Markov chain is based on a sequence \(\{G_t, t \in \mathbb{N}\}\) of independent \(N(0, 1)\) normal random variables that have mean 0 and variance 1 and are defined on a probability space \((\Omega, F, P)\). It is convenient to use the coordinate representation for this sequence [21, §II.9.2, Remark 1], defining it on the canonical probability space \((\mathbb{R}^N, B(\mathbb{R}^N), P)\), where \(\mathbb{R}^N\) consists of all sequences \(\omega = \{\omega_t, t \in \mathbb{N}\}\) with each \(\omega_t \in \mathbb{R}\), \(B(\mathbb{R}^N)\) is the Borel \(\sigma\)-algebra of subsets of \(\mathbb{R}^N\) generated by the algebra of finite-dimensional cylinder sets [21, §II.2.4], and \(P\) is the infinite product measure \(\prod_{t \in \mathbb{N}} \rho_t(d\omega_t)\) on \(B(\mathbb{R}^N)\) having identical one-dimensional marginals \(\rho_t(d\omega_t) = (2\pi)^{-1/2} \exp[-\omega_t^2/2] d\omega_t\). The finite-dimensional cylinder sets have the form \(\{\omega \in \mathbb{R}^N : (\omega_1, \omega_2, \ldots, \omega_n) \in B\}\) for \(n \in \mathbb{N}\) and \(B\) a Borel subset of \(\mathbb{R}^n\). For \(t \in \mathbb{N}\) and \(\omega = (\omega_1, \omega_2, \ldots) \in \mathbb{R}^N\) the sequence of independent \(N(0, 1)\) random variables is defined by \(G_t(\omega) = \omega_t\), the projection of \(\omega\) upon the coordinate \(\omega_t\).

In terms of this i.i.d. sequence we define a family of random variables \(\xi_t(f_r(x), N)\) that represent the noise in the random logistic model and assure that the Markov chain defined in (2.3) takes values in \([0, 1]\). Such a cutoff is suggested in [11]. For \(t \in \mathbb{N}\), \(x \in [0, 1]\), and \(\omega \in \Omega\) we define \(\xi_t(f_r(x), N)(\omega)\) to equal \(G_t(\omega)\) conditioned on \(G_t \in [-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x))]\). For \(\Gamma\) a Borel subset of \(\mathbb{R}\), \(\xi_t(f_r(x), N)\) has the distribution

\[
P(\xi_t(f_r(x), N) \in \Gamma) = P(G_t \in \Gamma \mid G_t \in [-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x))])
\]

\[
= \frac{P(G_t \in \Gamma \cap [-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x))])}{P(G_t \in [-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x))])}
\]

\[
= \frac{1}{\alpha_{r,N}(x)} \cdot \int_{\Gamma \cap [-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x))]} \exp\left[-\frac{1}{2} w^2\right] dw,
\]
where \( \alpha_{r,N}(x) \) is the normalization

\[
\alpha_{r,N}(x) = \int_{[-\sqrt{N} f_r(x), \sqrt{N}(1-f_r(x))]} \exp \left[ -\frac{1}{2} w^2 \right] dw.
\]

Fix \( r \in (0, 4] \). For \( x \in [0, 1] \) define \( x(1) = x \); although we do not assume this, \( x(1) \) can also be a random variable whose distribution is a nontrivial probability measure on \( [0, 1] \). For \( t \in \mathbb{N} \) and \( \omega \in \mathbb{R}^\mathbb{N} \) the random logistic model is defined as the Markov chain \( x_{r,N}(t) = x_{r,N}(t, \omega) \) satisfying the iteration

\[
x_{r,N}(t + 1, \omega) = f_r(x_{r,N}(t, \omega)) + \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t, \omega))),
\]

where \( f_r(y) \) is the logistic map \( ry(1-y) \) for \( y \in [0, 1] \). In order to simplify the notation, we write \( x(t) = x_{r,N}(t) \), suppressing the dependence on \( r \) and \( N \). This Markov chain is stationary because its transition probability function does not depend on \( t \). The explicit form of this transition probability function is given in Theorem 2.1 where we use the independence proved in part (c) of Lemma A.1.

**Theorem 2.1.** Fix \( r \in (0, 4] \) and \( N \in \mathbb{N} \). Let \( \{x(t), t \in \mathbb{N}\} = \{x_{r,N}(t), t \in \mathbb{N}\} \) be the stationary Markov chain on \([0, 1]\) defined in (2.3). For \( x \in [0, 1] \) and \( \Gamma \) a Borel subset of \([0, 1]\) the transition probability function \( K_{r,N}(x, \Gamma) = P(x(t + 1) \in \Gamma | x(t) = x) \) has the form

\[
K_{r,N}(x, \Gamma) = \int_{\Gamma} q_{r,N}(x, y) dy.
\]

The transition probability density \( q_{r,N}(x, y) \) is given by

\[
q_{r,N}(x, y) = \frac{1}{b_{r,N}(x)} \cdot \exp \left[ -\frac{N}{2} (y - f_r(x))^2 \right],
\]

(2.4)

where \( b_{r,N}(x) \) is the normalization

\[
b_{r,N}(x) = \int_{[0,1]} \exp \left[ -\frac{N}{2} (y - f_r(x))^2 \right] dy.
\]
Proof. For any Borel subset $\Gamma$ of $[0, 1]$

$$K_{r,N}(x, \Gamma) = P(x(t + 1) \in \Gamma \mid x(t) = x)$$

$$= P(f_r(x(t)) + N^{-1/2}\xi_t(f_r(x(t)), N) \in \Gamma \mid x(t) = x)$$

$$= P(f_r(x) + N^{-1/2}\xi_t(f_r(x), N) \in \Gamma \mid x(t) = x)$$

$$= P(N^{-1/2}\xi_t(f_r(x), N) \in \Gamma - f_r(x))$$

$$= P(N^{-1/2}G_t \in \Gamma - f_r(x) \mid N^{-1/2}G_t \in [-f_r(x), 1 - f_r(x)])$$

$$= \frac{P(N^{-1/2}G_t \in (\Gamma - f_r(x)) \cap [-f_r(x), 1 - f_r(x)])}{P(N^{-1/2}G_t \in [-f_r(x), 1 - f_r(x)])}$$

$$= \frac{1}{b_{r,N}(x)} \cdot \int_{\Gamma} \exp\left[-\frac{N}{2}(y - f_r(x))^2\right] dy.$$

The first two equalities follow from the definitions of $K(x, \Gamma)$ and $x_{t+1}$, the third equality by conditioning on $x(t) = x$, the fourth equality from the independence of $\xi_t(f_r(x), N)$ and $x(t)$ [Lemma A.1(c)], the fifth through seventh equalities from the fact that $N^{-1/2}\xi_t(f_r(x), N)$ equals $N^{-1/2}G_t$ conditioned on $N^{-1/2}G_t \in [-f_r(x), 1 - f_r(x)]$, and the last equality from the change of variables $y = w + f_r(x)$. In the last line of the display $b_{r,N}(x)$ is the normalization

$$b_{r,N}(x) = \int_{[0, 1]} \exp\left[-\frac{N}{2}(y - f_r(x))^2\right] dy.$$

It follows that $K_{r,N}(x, \Gamma) = \int_{\Gamma} q_{r,N}(x, y) dy$, where $q_{r,N}(x, y)$ is defined in (2.4). This completes the proof of the theorem. \qed

In Theorem 2.1 we show that the Markov chain $x(t)$ on $[0, 1]$ has the transition probability density $q_{r,N}(x, y)$ defined in (2.4). For all $r \in (0, 4]$ and $N \in \mathbb{N}$, $q_{r,N}(x, y)$ is continuous and strictly positive for $(x, y) \in [0, 1] \times [0, 1]$. Hence we have the following result, which is a consequence of standard ergodic theory for Markov chains [8, §VIII.7], [10, Thm. 1.2]. For $m \in \mathbb{N}$, $m \geq 2$ and $\Gamma$ a Borel subset of $[0, 1]$ we denote by $K_{r,N}^m(x, \Gamma)$ the $m$-step transition probability function defined
iteratively by

\[ K_{r,N}^m(x, \Gamma) = \int_{[0,1]} K_{r,N}(y, \Gamma) K_{r,N}^{m-1}(x, dy). \]

**Theorem 2.2.** Fix \( r \in (0, 4] \) and \( N \in \mathbb{N} \). Let \( \{x(t), t \in \mathbb{N}\} = \{x_{r,N}(t), t \in \mathbb{N}\} \) be the stationary Markov chain on \([0,1]\) defined in (2.3). The transition probability density \( q_{r,N}(x,y) \) is defined in (2.4). For all \( N \in \mathbb{N} \) the following conclusions hold.

(a) The Markov chain \( \{x(t), t \in \mathbb{N}\} \) is ergodic in the sense that there exists a probability measure \( \sigma_{r,N} \) on \([0,1]\) such that for all \( x \in [0,1] \) the following weak limit holds: for any continuous function \( \theta \) mapping \([0,1]\) into \( \mathbb{R} \)

\[
\lim_{m \to \infty} \int_{[0,1]} \theta(y) K_{r,N}^m(x, dy) = \int_{[0,1]} \theta(y) \sigma_{r,N}(dy). \tag{2.5}
\]

This weak limit is summarized by writing \( K_{r,N}^m(x,\cdot) \Rightarrow \sigma_{r,N} \).

(b) The measure \( \sigma_{r,N} \) is the unique invariant measure of the Markov chain; that is, \( \sigma_{r,N} \) is the unique probability measure on \([0,1]\) satisfying for any Borel subset \( \Gamma \) of \([0,1]\)

\[
\sigma_{r,N}(\Gamma) = \int_{[0,1]} K_{r,N}(x, \Gamma) \sigma_{r,N}(dx). \tag{2.6}
\]

The invariant measure \( \sigma_{r,N} \) has a strictly positive, continuous density \( h_{r,N} \) on \([0,1]\) satisfying for all \( y \in [0,1] \)

\[
h_{r,N}(y) = \int_{[0,1]} h_{r,N}(x) q_{r,N}(x,y) dx. \tag{2.7}
\]

**Proof.** (a) The existence of a probability measure \( \sigma_{r,N} \) satisfying a limit that implies the weak limit (2.5) is proved in Theorem 1.2 in [10]. Assumption 1 of that theorem is satisfied with \( V(x) = 1 \) for all \( x \in [0,1] \) and \( K = 1 \). Since the density \( q_{r,N}(x,y) \) is strictly positive for \((x,y) \in [0,1] \times [0,1]\), Assumption 2 is satisfied by choosing \( \alpha = \frac{1}{\sqrt{2\pi}} \sqrt{Ne^{-N/2}} \), which is less than 1 for any \( N \), and \( \nu \) the Lebesgue measure on \([0,1]\). The same limit also follows from Theorems 1 and 2 in Section VIII.7 in [8]. As shown in Example (b) on page 265 of [8], these theorems are applicable since the transition probability density \( q_{r,N}(x,y) \) is strictly positive for \((x,y) \in [0,1] \times [0,1]\), making the transition probability function \( K_{r,N}(x,\Gamma) \) a strictly positive regular kernel on the closed bounded interval \([0,1]\) in the sense of Definitions 1 and 2 on page 264. In appendix C the proof of the
ergodicity stated in [8] is given in detail.

(b) The fact that $\sigma_{r,N}$ is the unique invariant measure of the Markov chain $x(t)$ is a standard consequence of the existence of the limit (2.5) for any continuous function $\theta$ mapping $[0, 1]$ into $\mathbb{R}$. To show that $\sigma_{r,N}$ has a strictly positive, continuous density, we substitute into (2.6) the following formula, proved in Theorem 2.1 and valid for $x \in [0, 1]$ and $\Gamma$ a Borel subset of $[0, 1]$: \[
K_{r,N}(x, \Gamma) = \int_{\Gamma} q_{r,N}(x, y) \, dy.
\]

An application of the Fubini-Tonelli Theorem allows us to write

$$
\sigma_{r,N}(\Gamma) = \int_{\Gamma} \left( \int_{[0,1]} q_{r,N}(x, y) \sigma_{r,N}(dx) \right) dy,
$$

which identifies

$$
h_{r,N}(y) = \int_{[0,1]} q_{r,N}(x, y) \sigma_{r,N}(dx)
$$

as the density of $\sigma_{r,N}$. Since $q_{r,N}(x, y)$ is a strictly positive, continuous function on $[0, 1] \times [0, 1]$, the strict positivity and continuity of $h_{r,N}(y)$ for $y \in [0, 1]$ follow. In turn, since $\sigma_{r,N}$ has the density $h_{r,N}$, the last display implies that for all $y \in [0, 1]$

$$
h_{r,N}(y) = \int_{[0,1]} q_{r,N}(x, y) h_{r,N}(x) \, dx,
$$

which is equation (2.7). This completes the proof of the theorem.

In the next chapter we discuss the properties of the density $h_{r,N}$ of the invariant measure $\sigma_{r,N}$ of the Markov chain $x(t)$, where we drop the subscripts $r, N$ for convenience.
CHAPTER 3
PROPERTIES OF THE INVARIANT DENSITY \( h_{r,N} \)

In this chapter we study additional properties of the density \( h_{r,N} \) of the invariant measure \( \sigma_{r,N} \) of the Markov chain \( x(t) \). We refer to \( h_{r,N} \) as the invariant density. The existence of \( h_{r,N} \), its continuity, and strict positivity are proved in part (b) of Theorem 2.2. The properties of \( h_{r,N} \) will be applied in subsequent chapters where we determine the asymptotic behavior of \( h_{r,N} \) and of \( x_{r,N}(t) \).

We start by studying the asymptotic behavior of the normalization \( b_{r,N} \) of the transition probability density of the logistic Markov chain, the form of which is given in Theorem 2.1. For \( x \in [0,1] \) this normalization is defined by

\[
b_{r,N}(x) = \int_{[0,1]} \exp \left[ -\frac{N}{2} (y - f_r(x))^2 \right] dy.
\]

For \( r \in (0,4] \) this asymptotic behavior is a key component in the proof, given in part (b) of Proposition 3.2, of the uniform positivity of \( b_{r,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [0,1] \). In turn, the uniform positivity of \( b_{r,N}(x) \) is used to prove the estimates on the invariant density \( h_{r,N} \) in parts (a) and (b) of Theorem 3.3 and is required in several crucial places in the proofs of the weak convergence limits for \( h_{r,N} \) in Chapter 5. The normalization \( b_{r,N} \) exhibits a slight change in behavior for \( r = 4 \). Therefore, we will consider two cases in the following analysis, one where \( r \in (0,4) \) and one where \( r = 4 \). This difference is attributed to the fact that \( f_4 \) maps the interval \([0,1]\) to \([0,1]\) with \( f_4(1/2) = 1 \).

The change of variables \( z = \sqrt{N} (y - f_r(x)) \) in the definition of \( b_{r,N} \) gives

\[
\sqrt{N} b_{r,N}(x) = \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} \exp \left[ -\frac{1}{2} z^2 \right] dz.
\]

For \( r \in (0,4) \) and \( x \in (0,1) \), we have \( f_r(x) > 0 \) and \( 1 - f_r(x) > 0 \), and thus

\[
\lim_{N \to \infty} \sqrt{N} b_{r,N}(x) = \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz = \sqrt{2\pi}.
\]
On the other hand, if \( x = 0 \) or \( x = 1 \), then \( f_r(x) = 0 \), and therefore

\[
\lim_{N \to \infty} \sqrt{N} b_{r,N}(0) = \lim_{N \to \infty} \sqrt{N} b_{r,N}(1) = \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} z^2\right] \, dz = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}.
\]

For \( r = 4 \) and \( x \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1) \), we again have \( f_r(x) > 0 \) and \( 1 - f_r(x) > 0 \) and thus

\[
\lim_{N \to \infty} \sqrt{N} b_{r,N}(x) = \sqrt{2\pi}.
\]

The cases \( x = 0 \) and \( x = 1 \) yield the same result when \( r \in (0, 4) \).

Now let’s consider \( x = \frac{1}{2} \). Then \( f_r(\frac{1}{2}) = 1 \) and thus

\[
\lim_{N \to \infty} \sqrt{N} b_{r,N}(1/2) = \int_{-\infty}^{0} \exp\left[-\frac{1}{2} z^2\right] \, dz = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}.
\]

These simple calculations reveal a singularity in the limiting behavior of \( \sqrt{N} b_{r,N}(x) \). For \( r \in (0, 4) \), \( \sqrt{N} b_{r,N} \) converges to \( \sqrt{2\pi} \) for \( x \in (0, 1) \) and to \( \sqrt{\pi/2} \) for \( x = 0 \) and \( x = 1 \). For \( r = 4 \), \( \sqrt{N} b_{r,N} \) converges to \( \sqrt{2\pi} \) for \( x \in (0, 1/2) \cup (1/2, 1) \) and to \( \sqrt{\pi/2} \) for \( x \in \{0, 1/2, 1\} \). Because of this asymptotic singularity, the proof in Proposition 3.2 of the uniform positivity of \( b_{r,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [0, 1] \) is subtle. It depends on the next proposition, in part (a) of which we show that for \( r \in (0, 4) \) and any \( \varepsilon \in (0, 1/2) \) the convergence of \( \sqrt{N} b_{r,N}(x) \) to \( \sqrt{2\pi} \) is uniform for \( x \in [\varepsilon, 1 - \varepsilon] \). The uniform positivity of \( b_{r,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [\varepsilon, 1 - \varepsilon] \), stated in part (b), follows immediately. In part (c) we prove for \( r = 4 \) and any \( \varepsilon \in (0, 1/4) \) that \( \sqrt{N} b_{4,N}(x) \) converges uniformly to \( \sqrt{2\pi} \) for \( x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon] \) and part (d) states the uniform positivity of \( b_{4,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon] \).

**Proposition 3.1.** The normalization \( b_{r,N}(x) \) defined in (3.1) has the following properties.

(a) Fix \( r \in (0, 4) \) and \( \varepsilon \in (0, 1/2) \). We have the uniform limit

\[
\lim_{N \to \infty} \sup_{x \in [\varepsilon, 1 - \varepsilon]} |\sqrt{N} b_{r,N}(x) - \sqrt{2\pi}| = 0.
\]

(b) With the same conditions as in part (a), there exists \( \alpha_1 > 0 \) such that

\[
\inf_{N \in \mathbb{N}} \inf_{x \in [\varepsilon, 1 - \varepsilon]} \sqrt{N} b_{r,N}(x) \geq \alpha_1 > 0.
\]
(c) Fix $r = 4$ and $\varepsilon \in (0, 1/4)$. We have the uniform limit

$$\lim_{N \to \infty} \sup_{x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon]} |\sqrt{N}b_{4,N}(x) - \sqrt{2\pi}| = 0.$$ 

(d) With the same conditions as in part (c), there exists $\alpha_2 > 0$ such that

$$\inf_{N \in \mathbb{N}} \inf_{x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon]} \sqrt{N}b_{4,N}(x) \geq \alpha_2 > 0.$$

**Proof.** We use the fact that for any $a \in (0, \infty)$

$$\int_{a}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz \leq \int_{a}^{\infty} \frac{z}{a} \exp \left[ -\frac{1}{2} z^2 \right] dz = \frac{1}{a} \exp \left[ -\frac{1}{2} a^2 \right].$$

The inequality holds since $z \geq a$ in the integral while the equality holds by integrating by parts. It follows that for all $N \in \mathbb{N}$ and any $x \in [\varepsilon, 1 - \varepsilon]$

$$0 < \sqrt{2\pi} - \sqrt{N}b_{r,N}(x)$$

$$= \int_{-\infty}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz - \int_{-\infty}^{\sqrt{N}(1 - f_r(x))} \exp \left[ -\frac{1}{2} z^2 \right] dz$$

$$= \int_{\sqrt{N}f_r(x)}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz + \int_{\sqrt{N}(1 - f_r(x))}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz$$

$$\leq \int_{\sqrt{N}f_r(\varepsilon)}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz + \int_{\sqrt{N}(1 - r/4)}^{\infty} \exp \left[ -\frac{1}{2} z^2 \right] dz$$

$$\leq \frac{1}{\sqrt{N}f_r(\varepsilon)} \exp \left[ -\frac{N}{2} (f_r(\varepsilon))^2 \right] + \frac{1}{\sqrt{N}(1 - r/4)} \exp \left[ -\frac{N}{2} (1 - r/4)^2 \right].$$

We now choose any positive number $c$ satisfying $0 < c < \min(f_r(\varepsilon), 1 - r/4)$; since $r \in (0, 4)$, this minimum is positive. Then for all $N$ the following uniform bound holds:

$$0 < \sup_{x \in [\varepsilon, 1 - \varepsilon]} \left( \sqrt{2\pi} - \sqrt{N}b_{N}(x) \right) \leq \frac{2}{\sqrt{Nc}} \exp \left[ -\frac{N}{2} c^2 \right].$$

This bound shows that $\sqrt{N}b_{r,N}(x)$ converges to $\sqrt{2\pi}$ uniformly for $x \in [\varepsilon, 1 - \varepsilon]$. The proof of part (a) is complete.
(b) Part (a) implies that there exists $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and all $x \in [\varepsilon, 1 - \varepsilon]$, \( \sqrt{N} b_{r,N}(x) \geq \sqrt{2\pi/2} = \sqrt{\pi/2} \). Since each of the functions $\sqrt{N} b_{r,N}(x)$ for $N \in \{1, 2, \ldots, N_0 - 1\}$ is strictly positive, there exists a constant $\alpha_1 > 0$ such that $\sqrt{N} b_{r,N}(x) \geq \alpha_1$ for all $N \in \mathbb{N}$ and all $x \in [\varepsilon, 1 - \varepsilon]$. This proves part (b) of the proposition.

(c) The proof is the same as in part (a) with the difference that $r/4$ is replaced with $f(1/2 - \varepsilon)$. This yields $1 - f(1/2 - \varepsilon) > 0$ which, in turn, validates the proof.

(d) The proof is the same as in part (b) with the interval $[\varepsilon, 1 - \varepsilon]$ replaced with $[\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon]$. $lacksquare$

For $r \in (0, 4)$ part (b) of Proposition 3.1 proves the uniform positivity of $b_{r,N}(x)$ for $N \in \mathbb{N}$ and $x \in [\varepsilon, 1 - \varepsilon]$ for any $\varepsilon \in (0, 1/2)$. For $r = 4$ part (d) proves the uniform positivity of $b_{r,N}(x)$ for $N \in \mathbb{N}$ and $x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1 - \varepsilon]$ for any $\varepsilon \in (0, 1/4)$. We next give the extension of parts (b) and (d) of Proposition 3.1 to show the uniform positivity of $b_{r,N}(x)$ for $N \in \mathbb{N}$ and $x \in [0, 1]$.

**Proposition 3.2.** Fix $r \in (0, 4]$. The normalization $b_{r,N}(x)$ defined in (3.1) has the following property: there exists $\beta > 0$ such that

$$\inf_{N \in \mathbb{N}} \inf_{x \in [0,1]} \sqrt{N} b_{r,N}(x) \geq \beta > 0.$$  

**Proof.** We prove the proposition by contradiction, assuming that it is false; thus that

$$\inf_{N \in \mathbb{N}} \inf_{x \in [0,1]} \sqrt{N} b_{r,N}(x) = 0.$$  

It follows that there exists an increasing subsequence $\{N_k, k \in \mathbb{N}\}$ in $\mathbb{N}$ and a sequence of points $\{x_{N_k}, k \in \mathbb{N}\}$ in $[0, 1]$ such that

$$\lim_{k \to \infty} \sqrt{N_k} b_{r,N_k}(x_{N_k}) = 0.$$  

(3.2)

Since $[0, 1]$ is compact, there exists an increasing subsequence $\{N_{k'}, k' \in \mathbb{N}\}$ of $\{N_k, k \in \mathbb{N}\}$
and \( \bar{x} \in [0,1] \) such that

\[
\lim_{k' \to \infty} x_{N_{k'}} = \bar{x} \quad \text{and} \quad \lim_{k' \to \infty} \sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) = 0. \tag{3.3}
\]

First we will consider the case \( r \in (0,4) \). There are three possibilities: either \( \bar{x} \in (0,1) \) or \( \bar{x} = 0 \) or \( \bar{x} = 1 \). If \( \bar{x} \in (0,1) \), then the second limit in (3.3) contradicts part (b) of Proposition 3.1, which asserts the uniform positivity of \( b_{r,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [\varepsilon, 1-\varepsilon] \) for any positive number \( \varepsilon \) satisfying \( 0 < \varepsilon < \min(\bar{x}/2, 1-\bar{x}/2) \). It follows that the case \( \bar{x} \in (0,1) \) cannot arise.

We now consider the case \( \bar{x} = 0 \). By a change of variables

\[
\sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) = \sqrt{N_{k'}} \int_{[0,1]} \exp \left( -\frac{N_{k'}}{2}(y-f_r(x_{N_{k'}}))^2 \right) dy
\]

\[
= \int_{\sqrt{-N_{k'}} f_r(x_{N_{k'}})}^{\sqrt{N_{k'}} f_r(x_{N_{k'}})} \exp \left( -\frac{1}{2} z^2 \right) dz.
\]

Since by assumption \( x_{N_{k'}} \to \bar{x} = 0 \) as \( k' \to \infty \), in the upper limit of the integral \( f_r(x_{N_{k'}}) \to 0 \) as \( k' \to \infty \). Since in the lower limit of the integral \( f_r(x_{N_{k'}}) \geq 0 \), there exists \( k_0 \in \mathbb{N} \) such that for all \( k' \geq k_0 \)

\[
\sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) \geq \int_0^{\sqrt{N_{k'}}/2} \exp \left( -\frac{1}{2} z^2 \right) dz.
\]

It follows that there exists \( k_1 \in \mathbb{N} \) such that for all \( k' \geq k_1 \)

\[
\sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) \geq \frac{1}{2} \int_0^{\infty} \exp \left( -\frac{1}{2} z^2 \right) dz = \sqrt{\frac{\pi}{8}}.
\]

Since \( \sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) > 0 \) for \( k' \in \{1,2,\ldots,k_1-1\} \), we conclude that there exists \( \beta > 0 \) such that \( \sqrt{N_{k'}} b_{r,N_{k'}}(x_{N_{k'}}) \geq \beta > 0 \) for all \( k' \in \mathbb{N} \). However, this assertion contradicts the limit (3.3), proving that the case \( \bar{x} = 0 \) cannot arise.

By the same proof we show that the case \( \bar{x} = 1 \) also cannot arise.

Now we will fix \( r = 4 \). For this value of \( r \) we will investigate four possibilities: either \( \bar{x} \in (0,1/2) \cup (1/2,1) \) or \( \bar{x} = 0 \) or \( \bar{x} = 1/2 \) or \( \bar{x} = 1 \). If \( \bar{x} \in (0,1/2) \cup (1/2,1) \), then the second limit in (3.3) contradicts part (d) of Proposition 3.1, which asserts the uniform positivity of \( b_{4,N}(x) \) for \( N \in \mathbb{N} \) and \( x \in [\varepsilon, 1/2 - \varepsilon] \cup [1/2 + \varepsilon, 1-\varepsilon] \) for any positive number \( \varepsilon \) satisfying \( 0 < \varepsilon < \).
min(\(\bar{x}/4, 1 - \bar{x}/4\)). It follows that the case \(\bar{x} \in (0, 1/2) \cup (1/2, 1)\) cannot arise.

We now consider the case \(\bar{x} = 1/2\). By a change of variables

\[
\sqrt{N_{k'}} b_{4,N_{k'}}(x_{N_{k'}}) = \sqrt{N_{k'}} \int_{[0,1]} \exp \left[ - \frac{N_{k'}}{2} (y - f_4(x_{N_{k'}}))^2 \right] dy
\]

\[
= \int_{-\sqrt{N_{k'}} f_4(x_{N_{k'}})}^{\sqrt{N_{k'}} (1 - f_4(x_{N_{k'}}))} \exp \left[- \frac{1}{2} z^2 \right] dz.
\]

Since by assumption \(x_{N_{k'}} \to \bar{x} = 1/2\) as \(k' \to \infty\), in the lower limit of the integral \(f_4(x_{N_{k'}}) \to 1\) as \(k' \to \infty\). Since in the upper limit of the integral \(1 - f_4(x_{N_{k'}}) \geq 0\), there exists \(k_0 \in \mathbb{N}\) such that for all \(k' \geq k_0\)

\[
\sqrt{N_{k'}} b_{4,N_{k'}}(x_{N_{k'}}) \geq \int_{-\sqrt{N_{k'}}/2}^{0} \exp \left[- \frac{1}{2} z^2 \right] dz.
\]

It follows that there exists \(k_1 \in \mathbb{N}\) such that for all \(k' \geq k_1\)

\[
\sqrt{N_{k'}} b_{4,N_{k'}}(x_{N_{k'}}) \geq \frac{1}{2} \int_{-\infty}^{0} \exp \left[- \frac{1}{2} z^2 \right] dz = \sqrt{\frac{\pi}{8}}.
\]

Since \(\sqrt{N_{k'}} b_{4,N_{k'}}(x_{N_{k'}}) > 0\) for \(k' \in \{1, 2, \ldots, k_1 - 1\}\), we conclude that there exists \(\beta > 0\) such that \(\sqrt{N_{k'}} b_{4,N_{k'}}(x_{N_{k'}}) \geq \beta > 0\) for all \(k' \in \mathbb{N}\). However, this assertion contradicts the limit (3.3), proving that the case \(\bar{x} = 1/2\) cannot arise.

The cases \(\bar{x} = 0\) and \(\bar{x} = 1\) are exactly the same as with the proof done for the values of \(r \in (0, 4)\).

The proof of the proposition is complete. \(\blacksquare\)

The next theorem proves a number of properties of the invariant density \(h_{r,N}\) needed in the analysis of the asymptotic behavior of this quantity, carried out in later sections. In parts (a) and (b) we obtain bounds on \(h_{r,N}\). The proof of the theorem uses the fact, proved in part (b) of Theorem 2.2, that for all \(y \in [0, 1]\)

\[
h_{r,N}(y) = \int_{[0,1]} h_{r,N}(x) q_{r,N}(x, y) dx.
\]
The transition probability density \( k_{r,N}(x,y) \) is given by

\[
k_{r,N}(x,y) = \frac{1}{b_{r,N}(x)} \cdot \exp\left[ -\frac{N}{2}(y - f_r(x))^2 \right],
\]

where

\[
b_{r,N}(x) = \int_{[0,1]} \exp\left[ -\frac{N}{2}(y - f_r(x))^2 \right] dy.
\]

Substituting the formula for \( k_{r,N}(x,y) \) into (3.4) gives

\[
h_{r,N}(y) = \int_{[0,1]} h_r,N(x) \cdot \frac{1}{b_{r,N}(x)} \cdot \exp\left[ -\frac{N}{2}(y - f_r(x))^2 \right] dx.
\]  (3.5)

In part (b) of the theorem the notation \( \text{Ran} f_r \) denotes the range of \( f_r \) on \([0, 1]\), which equals the closed interval \([f_r(0), f_r(1/2)] = [0, r/4]\).

**Theorem 3.3.** Fix \( r \in (0, 4) \). The invariant density \( h_{r,N} \) has the following properties.

(a) There exists a positive constant \( \gamma \in (0, \infty) \) such that

\[
0 < \sup_{y \in [0,1]} h_{r,N}(y) \leq \gamma \sqrt{N}.
\]

The constant \( \gamma \) equals \( 1/\beta \), where \( \beta \in (0, \infty) \) is the constant in Proposition 3.2.

(b) Let \( y \) be any point in \([0, 1] \setminus \text{Ran} f_r = (r/4, 1] \). Then \( (y - r/4)^2 > 0 \) and

\[
0 < h_{r,N}(y) \leq \gamma \sqrt{N} \exp\left[ -\frac{N}{2}(y - r/4)^2 \right],
\]

where \( \gamma \) is the constant in part (a).

(c) For fixed \( N \in \mathbb{N} \), the invariant density \( h_{r,N} \) is an analytic function.

**Proof.** (a) In Proposition 3.2 we prove that there exists \( \beta > 0 \) such that

\[
\inf_{N \in \mathbb{N}} \inf_{x \in [0,1]} \sqrt{N} b_{r,N}(x) \geq \beta > 0.
\]

Since for all \( x \) and \( y \) in \([0, 1] \) we have \( \exp\left[ -\frac{N}{2}(y - f_r(x))^2 \right] \leq 1 \), it follows from (3.5) that

\[
0 < h_{r,N}(y) = \sqrt{N} \int_{[0,1]} h_r,N(x) \cdot \frac{1}{\sqrt{N}b_{r,N}(x)} \cdot \exp\left[ -\frac{N}{2}(y - f_r(x))^2 \right] dx
\]
where \( \gamma = 1/\beta < \infty \). The last equality follows from the fact that \( h_{r,N} \) is a density and so has integral 1 over \([0, 1]\). The proof of part (a) is complete.

(b) For all \( x \in [0, 1] \), \((y - f_r(x))^2 \geq (y - r/4)^2 > 0\). Hence by (3.5) and Proposition 3.2

\[
0 < h_{r,N}(y) = \sqrt{N} \int_{[0, 1]} h_{r,N}(x) \cdot \frac{1}{\sqrt{N} b_{r,N}(x)} \cdot \exp \left[ -\frac{N}{2} (y - f_r(x))^2 \right] dx
\leq \frac{1}{\beta} \sqrt{N} \cdot \exp \left[ -\frac{N}{2} (y - r/4)^2 \right] \cdot \int_{[0, 1]} h_{r,N}(x) dx
\leq \frac{1}{\beta} \sqrt{N} \cdot \exp \left[ -\frac{N}{2} (y - r/4)^2 \right].
\]

This gives the inequality in part (b) with \( \gamma = 1/\beta \). The proof of part (b) is complete.

(c) For \( h_N = h_{r,N} \) to be analytic, we would like to have a positive radius of convergence around \( x \) so we can write \( h_N(y) = \sum_{k=0}^{\infty} h^{(k)}_{N}(x)(y - x)^k \) for some \( x \in (0, 1) \). That means having control over the derivatives.

Let \( y \in (0, 1) \). Since \( h_N(y) = \int_{[0, 1]} h_N(x) \frac{1}{b_N(x)} \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx \), differentiating under the integral sign we have

\[
h^{(k)}_{N}(y) = \int_{[0, 1]} h_N(x) \frac{1}{b_N(x)} \frac{d^k}{dy^k} \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx
\]

\[
= \int_{[0, 1]} h_N(x) \frac{1}{b_N(x)} \left[ (-1)^k N^{k/2} \Gamma_k \left( \sqrt{N} |y - f(x)| \right) \right] \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx,
\]

where \( \Gamma_k(x) \) is the \( k \)-th Hermite polynomial and is given explicitly by

\[
\Gamma_k(x) = k! \sum_{n=0}^{\lfloor k/2 \rfloor} \frac{(-1)^n x^{k-2n}}{n!(k-2n)!} 2^n.
\]
If $|x| \leq C$ for some $C$, then a bound for the $k$-th Hermite polynomial is

$$|\Gamma_k(x)| \leq k! \sum_{n=0}^{[\frac{k}{2}]} \frac{(-1)^n}{n!(k-2n)!} \frac{|x|^{k-2n}}{2^n} \leq k! \sum_{n=0}^{[\frac{k}{2}]} \frac{1}{n!(k-2n)!} \frac{C^{k-2n}}{2^n},$$

where the bound is independent of the variable $x$.

Back to the derivatives of $h_N$. Taking absolute value we have

$$|h_N^{(k)}(y)| \leq \int_{[0,1]} h_N(x) \frac{1}{b_N(x)} \left| (-1)^k N^{\frac{k}{2}} \Gamma_k \left( \sqrt{N} |y - f(x)| \right) \right| \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx$$

$$= \int_{[0,1]} h_N(x) \frac{1}{b_N(x)} N^{\frac{k}{2}} \left| \Gamma_k \left( \sqrt{N} |y - f(x)| \right) \right| \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx$$

$$\leq N^{\frac{k}{2}} \sup_{x,y \in [0,1]} \left| \Gamma_k \left( \sqrt{N} |y - f(x)| \right) \right| \int_{[0,1]} h_N(x) \frac{1}{b_N(x)} \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx.$$

Since $\exp \left[ -\frac{N}{2} (y - f(x))^2 \right] \leq 1$ and $\frac{1}{b_N(x)} \leq \gamma \sqrt{N}$ from part (a) and the fact that $h_N$ is a probability density on $[0,1]$ yield

$$\int_{[0,1]} h_N(x) \frac{1}{b_N(x)} \exp \left[ -\frac{N}{2} (y - f(x))^2 \right] dx \leq \gamma \sqrt{N} \int_{[0,1]} h_N(x) dx = \gamma \sqrt{N}.$$

Thus, we have that the $k$-th derivative of $h_N$ is bounded by

$$|h_N^{(k)}(y)| \leq \gamma \sqrt{N} N^{\frac{k}{2}} \sup_{x,y \in [0,1]} \left| \Gamma_k \left( \sqrt{N} |y - f(x)| \right) \right|.$$

Since $|y - f(x)| \leq 1$ and thus $\sqrt{N} |y - f(x)| \leq \sqrt{N}$, we have from the bound we calculated above for $\Gamma_k$ that

$$|h_N^{(k)}(y)| \leq \gamma \sqrt{N} N^{\frac{k}{2}} k! \sum_{n=0}^{[\frac{k}{2}]} \frac{1}{n!(k-2n)!} \frac{(\sqrt{N})^{k-2n}}{2^n}.$$
\[\sum_{n=0}^{\left\lfloor \frac{k}{2} \right\rfloor} \frac{1}{n!(k-2n)!} \frac{N^{\frac{k}{2}-n}}{2^n}\]

\[= \gamma \sqrt{N} N^k k! \sum_{n=0}^{\left\lfloor \frac{k}{2} \right\rfloor} \frac{1}{n!(k-2n)!} (2N)^n\]

\[\leq \gamma \sqrt{N} N^k k! \sum_{n=0}^{\left\lfloor \frac{k}{2} \right\rfloor} \frac{1}{n!(k-2n)!} 2^n\]

\[= \gamma \sqrt{N} N^k T(k),\]

where the last inequality is due to the fact that \(\frac{1}{N^n} \leq 1\) and \(T(k)\) is the \(k\)-th telephone number. For more details on that sequence of numbers, the reader is referred to [2, 5].

At this point let’s write down the Taylor series for \(h_N\) expanded around \(y = x\):

\[h_N(y) = \sum_{k=0}^{\infty} \frac{h_N^{(k)}(x)}{k!} (y - x)^k.\]

Taking absolute values we get

\[\left| \sum_{k=0}^{\infty} \frac{h_N^{(k)}(x)}{k!} (y - x)^k \right| \leq \sum_{k=0}^{\infty} \left| \frac{h_N^{(k)}(x)}{k!} \right| |y - x|^k\]

\[\leq \sum_{k=0}^{\infty} \gamma \sqrt{N} N^k T(k) \frac{|y - x|^k}{k!}\]

\[= \gamma \sqrt{N} \sum_{k=0}^{\infty} \frac{T(k)}{k!} (N|y - x|)^k\]

\[= \gamma \sqrt{N} \exp \left( \frac{(N|y - x|)^2}{2} + N|y - x| \right),\]

where the last result is bounded for fixed \(N\) and \(|y - x| \leq 1\). It is due to the fact that

\[\sum_{k=0}^{\infty} \frac{T(k)}{k!} y^k = \exp \left( \frac{y^2}{2} + y \right), \quad y \in \mathbb{R},\]

and thus

\[\sum_{k=0}^{\infty} \frac{h_N^{(k)}(x)}{k!} (y - x)^k\]
converges absolutely. The next and final step is to show that the series converges to $h_N$ itself.

To show convergence of the Taylor series to $h_N$, we need to show that the remainder $R_k(y)$ converges to 0 as $k \to \infty$. Let $y \in (0, 1)$. Then we have

$$0 \leq |R_k(y)| = \left| \int_0^y \frac{h_N^{(k+1)}(t)}{k!} (y-t)^k dt \right|$$

$$\leq \int_0^y \frac{|h_N^{(k+1)}(t)|}{k!} (y-t)^k dt.$$ 

Since $0 \leq y-t \leq 1$ and $|h_N^{(k+1)}(t)| \leq \gamma \sqrt{NN^{k+1}T(k+1)}$, we have that

$$|R_k(y)| \leq \int_0^y \gamma \sqrt{NN^{k+1}T(k+1)} \frac{1}{k!} dt$$

$$= \frac{\gamma \sqrt{NN^{k+1}T(k+1)}}{k!} y$$

$$\leq \frac{\gamma \sqrt{NN^{k+1}T(k+1)}}{k!}.$$ 

Now, using the fact that $T(k+1) = T(k) + kT(k-1) \leq (k+1)T(k)$ we have that

$$\sum_{k=0}^{\infty} \frac{\gamma \sqrt{NN^{k+1}T(k+1)}}{k!} \leq \sum_{k=0}^{\infty} \frac{\gamma \sqrt{NN^{k+1}(k+1)T(k)}}{k!}$$

$$\leq \sum_{k=0}^{\infty} \frac{\gamma \sqrt{NN^{k+1}2^{k+1}T(k)}}{k!}$$

$$= \gamma \sqrt{N} 2N \sum_{k=0}^{\infty} \frac{(2N)^k T(k)}{k!}$$

$$= 2\gamma N^{3/2} \sum_{k=0}^{\infty} \frac{T(k)}{k!} (2N)^k$$

$$= 2\gamma N^{3/2} \exp \left( \frac{4N^2}{2} + 2N \right)$$

$$= 2\gamma N^{3/2} \exp(2N^2 + 2N).$$
Therefore, the series \( \sum_{k=0}^{\infty} \frac{\gamma \sqrt{N} N^{k+1} T(k+1)}{k!} \) converges which means that

\[
\frac{\gamma \sqrt{N} N^{k+1} T(k+1)}{k!} \rightarrow 0 \quad \text{as} \quad k \rightarrow \infty.
\]

Thus, we conclude that \( R_k(y) \rightarrow 0 \) as \( k \rightarrow \infty \) which implies that for any \( x \in (0, 1) \)

\[
h_N(y) = \sum_{k=0}^{\infty} \frac{h_N^{(k)}(x)}{k!} (y-x)^k, \quad y \in (0,1),
\]

and thus \( h_N \) is an analytic function.
CHAPTER 4
FIRST RESULTS: CONVERGENCE ALMOST SURELY AND IN PROBABILITY

4.1 Almost sure convergence

The goal is to show that the Random Logistic Model reduces to the deterministic logistic map almost surely as the noise vanishes. Therefore, we need to show that for fixed \( t \in \mathbb{N} \) we have

\[
\lim_{N \to \infty} x_{r,N}(t) = f_r^{(t-1)}(y) \text{ with probability 1},
\]

where \( x(1) = y \in [0, 1] \) fixed. The proof will be done by induction. First we prove that

\[
\lim_{N \to \infty} x_{r,N}(2) = f_r(y) \text{ with probability 1}
\]

by using a known sufficient condition for almost sure convergence, which is stated in the following lemma:

**Lemma 4.1.** Consider a sequence of random variables \( X_1, X_2, \ldots \). If for all \( \varepsilon > 0 \) we have

\[
\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty
\]

then \( X_n \to X \) \( P \)-almost surely.

Now that we stated the above lemma, we are ready to state and prove the first of the two asymptotic results.

**Theorem 4.2.** For fixed \( r \in (0, 4) \) and \( x_{r,N}(1) = y \in [0, 1] \), consider the random logistic model defined by the Markov Chain \( \{x_{r,N}(t, \omega)\} \) which satisfies the iteration given in 2.3. Letting \( N \to \infty \), causing the noise to vanish, and fixing \( t \in \mathbb{N}, t > 1 \), we obtain the following result:

\[
P \left( \omega \in \mathbb{R}^\mathbb{N} : \lim_{N \to \infty} x_{r,N}(t, \omega) = f_r^{(t-1)}(y) \right) = 1.
\]

**Proof.** The proof is done by induction. First we show that the noise term \( \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t, \omega))) \) converges with probability 1 to 0 as \( N \to \infty \) for any \( t \in \mathbb{N} \). This is shown by applying Lemma 4.1.
We need to verify that for any $\varepsilon > 0$
\[
\sum_{N=1}^{\infty} P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_r,N(t,\omega))) \right| > \varepsilon \right) < \infty.
\]

Pick $\varepsilon > 0$. By (2.2) we have that

\[
P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_r,N(t,\omega))) \right| > \varepsilon \right) = P \left( |\xi_t(f_r(x_r,N(t,\omega)))| > \sqrt{N}\varepsilon \right)
\]

\[
= P \left( |\xi_t(f_r(x_r,N(t,\omega)))| \in (-\infty, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty) \right)
\]

\[
\leq \sup_{x \in [0,1]} P \left( |\xi_t(f_r(x))| \in (-\infty, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty) \right)
\]

\[
= \sup_{x \in [0,1]} \frac{1}{\alpha_{r,N}(x)} \int_{[-\infty, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty)} \exp\left(-\frac{1}{2}w^2\right) dw
\]

\[
\leq \frac{1}{\beta} \int_{[-\infty, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty)} \exp\left(-\frac{1}{2}w^2\right) dw,
\]

where the last step is justified due the fact that $f_r(x) \in [0,1]$ and Proposition 3.2, since $\alpha_{r,N}(x) = \sqrt{N}b_{r,N}(x)$ by a change of variables. There are two cases to take into account here. One for $\varepsilon \geq 1$ and the other for $0 < \varepsilon < 1$.

If $\varepsilon \geq 1$, then $[-\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty)] \cap [-\sqrt{N}, \sqrt{N}] = \emptyset$, which implies

\[
P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_r,N(t,\omega))) \right| > \varepsilon \right) = 0
\]

for all $N \in \mathbb{N}$ and thus

\[
\sum_{N=1}^{\infty} P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_r,N(t,\omega))) \right| > \varepsilon \right) = 0 < \infty.
\]

Next we consider the case $0 < \varepsilon < 1$. Then we have that

\[
([(-\infty, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \infty)] \cap [-\sqrt{N}, \sqrt{N}] = [-\sqrt{N}, -\sqrt{N}\varepsilon) \cup (\sqrt{N}\varepsilon, \sqrt{N}].
\]

29
Therefore,

\[
P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) \leq \frac{1}{\beta} \int_{[-\sqrt{N}, \sqrt{N}]} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
= \frac{1}{\beta} \int_{[-\sqrt{N}, -\sqrt{N} \varepsilon] \cup (\sqrt{N} \varepsilon, \sqrt{N})} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
= \frac{1}{\beta} \int_{-\sqrt{N} \varepsilon}^{\sqrt{N}} \exp \left[ -\frac{1}{2} w^2 \right] dw + \frac{1}{\beta} \int_{\sqrt{N} \varepsilon}^{\sqrt{N}} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
= \frac{2}{\beta} \int_{\sqrt{N} \varepsilon}^{\sqrt{N}} \exp \left[ -\frac{1}{2} w^2 \right] dw.
\]

Replacing the upper limit in the last equality above with \( \infty \) and applying the same technique we used in the proof of Proposition 3.1, we have

\[
P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) \leq \frac{2}{\beta} \int_{\sqrt{N} \varepsilon}^{\infty} \exp \left[ -\frac{1}{2} w^2 \right] dw \leq \frac{2}{\beta \sqrt{N} \varepsilon} \exp \left[ -\frac{1}{2} N \varepsilon^2 \right].
\]

Now, pick any integer \( N' > \left( \frac{2}{\beta \varepsilon} \right)^2 \). Then, for all \( N \geq N' \) we have that \( \frac{2}{\beta \sqrt{N} \varepsilon} < 1 \).

Thus,

\[
P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) \leq \exp \left[ -\frac{1}{2} N \varepsilon^2 \right].
\]

By summing from \( N' \) to \( \infty \) we have

\[
\sum_{N=N'}^{\infty} P \left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) \leq \sum_{N=N'}^{\infty} \exp \left[ -\frac{1}{2} N \varepsilon^2 \right]
\]

\[
= \sum_{N=N'}^{\infty} \left( \exp \left[ -\frac{1}{2} \varepsilon^2 \right] \right)^N
\]

\[
\leq \frac{1}{1 - \exp \left[ -\frac{1}{2} \varepsilon^2 \right]}.
\]
The convergence is justified since \( 0 < \exp\left(-\frac{1}{2} \varepsilon^2\right) < 1 \). An immediate consequence is that

\[
\sum_{N=1}^{\infty} P\left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) < \infty \quad \text{for } \varepsilon \in (0, 1).
\]

In conclusion, for any \( \varepsilon > 0 \) we have that

\[
\sum_{N=1}^{\infty} P\left( \left| \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \right| > \varepsilon \right) < \infty,
\]

which implies the almost sure convergence of \( \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(t,\omega))) \) to 0 as \( N \to \infty \), for any \( t \in \mathbb{N}, \omega \in \mathbb{R}^N \).

The next step is to perform the induction and prove that \( \lim_{N \to \infty} x_{r,N}(t) = f_r^{(t-1)}(y) \) with probability 1 for fixed \( x_{r,N}(1) = y \in [0, 1] \).

We start with \( t = 2 \). By definition of the Markov Chain and equation (2.3), we have

\[
x_{r,N}(2) = f_r(x_{r,N}(1)) + \frac{1}{\sqrt{N}} \xi_t(f_r(x_{r,N}(1))) = f_r(y) + \frac{1}{\sqrt{N}} \xi_t(f_r(y)).
\]

Since \( f_r(y) \) is constant and \( \frac{1}{\sqrt{N}} \xi_t(f_r(y)) \) converges to 0 P-almost surely, we have

\[
\lim_{N \to \infty} x_{r,N}(2) = f_r(y) \quad \text{P-almost surely.}
\]

This proves that for \( t = 2 \), the random variable \( x_{r,N}(2) \), as the noise vanishes, with probability 1 it is reduced to the deterministic logistic map.

Now for the induction step. Suppose that \( x_{r,N}(t) \) converges with probability 1 to \( f_r^{(t-1)}(y) \) for some \( t \in \mathbb{N} \) as \( N \to \infty \) and \( y \in [0, 1] \). We show it is true for \( t + 1 \). Again by equation (2.3) we
have
\[ x_{r,N}(t + 1) = f_r(x_{r,N}(t)) + \frac{1}{\sqrt{N}}\xi_t(f_r(x_{r,N}(t))). \]

Let’s investigate the two terms on the RHS separately as \( N \to \infty \).

We have already proven that \( \frac{1}{\sqrt{N}}\xi_t(f_r(x_{r,N}(t))) \to 0 \) \( \text{P-almost surely} \).

Now, by hypothesis of the induction step we have that \( x_{r,N}(t) \to f_r^{(t-1)}(y) \) \( \text{P-almost surely} \). Also, \( f_r \) is a continuous function on \([0,1]\). Thus, by the continuous mapping theorem, we have that
\[
\lim_{N \to \infty} f_r(x_{r,N}(t)) = f_r \left( f_r^{(t-1)}(y) \right) \text{ P-almost surely.}
\]

Therefore, we conclude that
\[
\lim_{N \to \infty} x_{r,N}(t + 1) = f_r \left( f_r^{(t-1)}(y) \right) = f_r^{(t)}(y) \text{ P-almost surely.}
\]

This concludes the induction and we have proven that for any \( t \in \mathbb{N} \) and any initial value \( x_{r,N}(1) = y \in [0,1] \), with probability 1 the random logistic model is reduced to the dynamics of the logistic map \( f_r \) as the noise vanishes.

A consequence of Theorem 4.2 is that if we take the limit of \( t \to \infty \) after the vanishing-noise limit, the behavior is that of the logistic map as a dynamical system outlined in the beginning of Chapter 2.

4.2 Convergence rates in probability

In Theorem 4.2 we proved the almost sure convergence result using the tightness argument. The almost sure convergence implies convergence in probability and for \( r \in (0,4] \) the convergence rates are stated below.

**Theorem 4.3.** Given \( \varepsilon > 0 \) and \( r \in (0,4] \), the following convergence rates hold
\[
P \left\{ \left| x_{r,N}(t + 1) - f_r^{(t)}(x) \right| \geq \varepsilon \right\} \leq \frac{t^2 \gamma}{\sqrt{N} \varepsilon} \exp \left[ -N \frac{\varepsilon^2}{2^{2r-1}} \right], \text{ when } r \in (0,1),
\]
\( P \left\{ \left| x_{r,N}(t+1) - f^{(t)}_r(x) \right| \geq \epsilon \right\} \leq \frac{2\gamma t (2r)^{t-1}}{\sqrt{N} \epsilon} \exp \left[ -\frac{N}{2} \frac{\epsilon^2}{(2r)^{2t-2}} \right], \) when \( r \in [1,4] \).

The proof will be based on the following lemma:

**Lemma 4.4.** For two random variables \( X,Y \) defined on a probability space \( (\Omega, \mathcal{F}, P) \) and given \( \epsilon > 0 \) we have

\[
P(|X| + |Y| \geq \epsilon) \leq P(|X| \geq \frac{\epsilon}{2}) + P(|Y| \geq \frac{\epsilon}{2}).
\]

In general, for a sequence of random variables \( X_1, \ldots, X_n \) and given \( \epsilon > 0 \) we have that

\[
P \left( \sum_{i=1}^{n} |X_i| \geq \epsilon \right) \leq \sum_{i=1}^{n} P \left( |X_i| \geq \frac{\epsilon}{2^{n-1}} \right).
\]

**Proof.** Fix \( x_1 = x \in [0,1] \), pick \( \epsilon > 0 \) and let \( \lambda = 1/\sqrt{N} \). The first step is to find a bound for the term \( |x_{t+1} - f^{(t)}(x)| \). By the Mean Value Theorem there exists some \( c \in (0,1) \) such that \( f(b) - f(a) = f'(c)(b - a) \). Thus, \( |f(b) - f(a)| \leq \sup_{y \in (0,1)} |f'(y)||b - a| \). Since \( f'(y) = r - 2ry \), we have that \( \sup_{y \in (0,1)} |f'(y)| = r \).

Therefore

\[
|f(x_t) - f^{(t)}(x)| \leq r|x_t - f^{(t-1)}(x)|.
\] (4.1)

We will prove by induction for fixed \( t \in \mathbb{N} \) and \( m = 0,1,\ldots,t-1 \) the following result:

\[
|x_{t+1} - f^{(t)}(x)| \leq r^m |f(x_{t-m}) - f^{(t-m)}(x)| + \sum_{i=0}^{m} r^i \lambda |\xi_{t-i}|.
\] (4.2)

For \( m = 0 \) it is true:

\[
|x_{t+1} - f^{(t)}(x)| = |f(x_t) + \lambda \xi_t - f^{(t)}(x)| \leq |f(x_t) - f^{(t)}(x)| + \lambda |\xi_t|
\]
and for $m = 1$ it is true:

$$|x_{t+1} - f(t)(x)| \leq |f(x_t) - f(t)(x)| + \lambda|\xi_t|$$

$$\leq r|x_t - f^{(t-1)}(x)| + \lambda|\xi_t|$$

$$= r|f(x_{t-1}) + \lambda\xi_{t-1} - f^{(t-1)}(x)| + \lambda|\xi_t|$$

$$\leq r|f(x_{t-1}) - f^{(t-1)}(x)| + r\lambda|\xi_{t-1}| + \lambda|\xi_t|$$

$$= r|f(x_{t-1}) - f^{(t-1)}(x)| + \sum_{i=0}^{1} r^i \lambda|\xi_{t-i}|.$$  

Now assume the result is true for $k \in \{0, 1, \ldots, t-2\}$. We will show it holds for $k + 1$.

$$|x_{t+1} - f(t)(x)| \leq r^k|f(x_{t-k}) - f^{(t-k)}(x)| + \sum_{i=0}^{k} r^i \lambda|\xi_{t-i}|$$

$$\leq r^k r|x_{t-k} - f^{(t-k-1)}(x)| + \sum_{i=0}^{k} r^i \lambda|\xi_{t-i}|$$

$$= r^{k+1}|f(x_{t-k-1}) + \lambda\xi_{t-k-1} - f^{(t-k-1)}(x)| + \sum_{i=0}^{k} r^i \lambda|\xi_{t-i}|$$

$$\leq r^{k+1}|f(x_{t-k-1}) - f^{(t-k-1)}(x)| + r^{k+1}\lambda|\xi_{t-k-1}| + \sum_{i=0}^{k} r^i \lambda|\xi_{t-i}|$$

$$= r^{k+1}|f(x_{t-k-1}) - f^{(t-k-1)}(x)| + \sum_{i=0}^{k+1} r^i \lambda|\xi_{t-i}|.$$  

Thus, the result holds for $k + 1$ and we have concluded the induction. Setting $m = t - 1$ yields the following

$$|x_{t+1} - f(t)(x)| \leq r^{t-1}|f(x_{t-t+1}) - f^{(t-t+1)}(x)| + \sum_{i=0}^{t-1} r^i \lambda|\xi_{t-i}|$$

$$= r^{t-1}|f(x_1) - f(x)| + \sum_{i=0}^{t-1} r^i \lambda|\xi_{t-i}|$$

34
\[ t_{t+1} − \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}|, \]

since \( t_1 = x \). Therefore, we have

\[ |x_{t+1} − f^{(f)}(x)| \leq \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}|. \] \hfill (4.3)

Assume that \( r \geq 1 \).

\[ P \left( |x_{t+1} − f^{(f)}(x)| \geq \varepsilon \right) \leq P \left( \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}| \geq \varepsilon \right), \]

and then by applying Lemma 4.4 we have

\[ P \left( \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}| \geq \varepsilon \right) = P \left( \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}| \geq \sqrt{N} \varepsilon \right) \]

\[ \leq \sum_{i=0}^{t-1} P \left( r^i \lambda |\xi_{t-i}| \geq \sqrt{N} \frac{\varepsilon}{2^{t-1}} \right) \]

\[ = \sum_{i=0}^{t-1} P \left( |\xi_{t-i}| \geq \sqrt{N} \frac{\varepsilon}{2^{t-1} r^i} \right) \]

\[ \leq \sum_{i=0}^{t-1} \sup_{y \in [0,1]} P \left( |\xi_{t-i}(f(y), \lambda)| \geq \sqrt{N} \frac{\varepsilon}{2^{t-1} r^i} \right). \]

Let’s work with each probability in the sum individually and set \( \varepsilon_t = \frac{\varepsilon}{(2r)^{t-1}} \). We then have

\[ P \left( |\xi_{t-i}(f(y), \lambda)| \geq \sqrt{N} \varepsilon_t \right) = \frac{1}{a(y, \lambda)} \int_{[(-\infty, -\sqrt{N} \varepsilon_t] \cup [\sqrt{N} \varepsilon_t, \infty] \cap B(y)}} \exp \left( -\frac{1}{2} w^2 \right) \, dw, \]

where

\[ a(y, \lambda) = \int_{B(y)} \exp \left( -\frac{1}{2} w^2 \right) \, dw \quad \text{and} \quad B(y) = [-\sqrt{N} f(y), \sqrt{N} (1 - f(y))]. \]
Notice that

\[
\left( -\infty, -\sqrt{N \epsilon_t} \right] \cup \left[ \sqrt{N \epsilon_t}, \infty \right) \cap B(y) \subset \left( -\infty, -\sqrt{N \epsilon_t} \right] \cup \left( \sqrt{N \epsilon_t}, \infty \right).
\]

So we have

\[
P\left( |\xi_{t-i}(f(y), \lambda)| \geq \sqrt{N \epsilon_t} \right) = \frac{1}{a(y, \lambda)} \int_{\left( -\infty, -\sqrt{N \epsilon_t} \right] \cup \left( \sqrt{N \epsilon_t}, \infty \right) \cap B(y)} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
\leq \frac{1}{a(y, \lambda)} \int_{\left( -\infty, -\sqrt{N \epsilon_t} \right] \cup \left( \sqrt{N \epsilon_t}, \infty \right)} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
= \frac{2}{a(y, \lambda)} \int_{\sqrt{N \epsilon_t}}^{\infty} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
\leq \frac{2}{a(y, \lambda)} \int_{\sqrt{N \epsilon_t}}^{\infty} \frac{w}{\sqrt{N \epsilon_t}} \exp \left[ -\frac{1}{2} w^2 \right] dw
\]

\[
= \frac{2}{a(y, \lambda)} \exp \left[ -\frac{1}{2} N \epsilon_t^2 \right].
\]

Since we know that \( a(y, \lambda) \geq \beta \) for any \( N \) we finally have that

\[
P\left( |x_{t+1} - f^{(t)}(x)| \geq \epsilon \right) \leq P\left( \sum_{i=0}^{t-1} x_i^i \lambda |\xi_{t-i}| \geq \epsilon \right)
\]

\[
\leq \sum_{i=0}^{t-1} \sup_{y \in [0,1]} P\left( |\xi_{t-i}(f(y), \lambda)| \geq \sqrt{N \epsilon_t} \right)
\]

\[
\leq \sum_{i=0}^{t-1} 2\gamma \exp \left[ -\frac{1}{2} N \epsilon_t^2 \right]
\]

\[
\leq \sum_{i=0}^{t-1} (2r)^{t-i-1} 2\gamma \exp \left[ -\frac{1}{2} N \frac{\epsilon^2}{(2r)^{2t-2}} \right]
\]

\[
= t(2r)^{t-1} 2\gamma \exp \left[ -\frac{1}{2} N \frac{\epsilon^2}{(2r)^{2t-2}} \right].
\]
Similarly it is done for \( r < 1 \) where instead we obtain

\[
P \left( \sum_{i=0}^{t-1} r^i \lambda |\xi_{t-i}| \geq \epsilon \right) = P \left( \sum_{i=0}^{t-1} r^i |\xi_{t-i}| \geq \sqrt{N} \epsilon \right)
\]

\[
\leq \sum_{i=0}^{t-1} P \left( r^i |\xi_{t-i}| \geq \sqrt{N} \frac{\epsilon}{2^{t-1}} \right)
\]

\[
= \sum_{i=0}^{t-1} P \left( |\xi_{t-i}| \geq \sqrt{N} \frac{\epsilon}{2^{t-1} r^i} \right)
\]

\[
\leq \sum_{i=0}^{t-1} \sup_{y \in [0,1]} P \left( |\xi_{t-i}(y, \lambda)| \geq \sqrt{N} \frac{\epsilon}{2^{t-1} r^i} \right),
\]

and the bound is the one stated in the Theorem.

\[\blacksquare\]

**Remark 4.5.** Notice that if \( t \) scales as a function of \( N \), and specifically for \( c \in (0, 1/2) \),

\[
t = c \log_2 (N), \quad \text{for } r \in (0, 1)
\]

and

\[
t = c \log_{2r} (N), \quad \text{for } r \in [1, 4],
\]

then as \( N \to \infty \) we have that the bounds in Theorem 4.3 vanish, which implies that,

\[
\left| x_{r,N}(t(N) + 1) - f_r^{t(N)}(x) \right| \to 0 \quad \text{in probability.}
\]
CHAPTER 5

PROOF OF THE MAIN THEOREM: WEAK CONVERGENCE

Theorem 1.2, stated in the Introduction, consists of three parts. Part (a) states that as \( t \to \infty \), \( x_{4,N}(t) \) converges in distribution to the probability measure \( \sigma_{4,N} \) on \([0,1]\), which is the unique invariant measure of the Markov chain. Part (a) is proved in Theorem 2.2. Part (b) of Theorem 1.2 states that as \( N \to \infty \) the sequence \( \sigma_{4,N} \) converges weakly to the Ulam-von Neumann invariant measure \( \sigma^* \), which is absolutely continuous with respect to Lebesgue measure on \([0,1]\) and has the density \( 1/\pi \sqrt{x(1-x)} \). In this chapter we prove this convergence along with the related convergence results for \( r \in (0, r_\infty) \) stated in Theorem 1.3. Part (c) of Theorem 1.2 combines the limits in parts (a) and (b), deducing that the Ulam-von Neumann Markov chain converges in distribution to \( \sigma^* \).

The proofs of Theorems 1.2 and 1.3 consist of the following three steps.

1. For any subsequence \( \sigma_{r,N'} \) of \( \sigma_{r,N} \) there exists a subsubsequence \( \sigma_{r,N''} \) and a probability measure \( \sigma \) on \([0,1]\) such that the subsubsequence \( \sigma_{r,N''} \) converges weakly to \( \sigma \). This property follows from Prohorov’s Theorem and the compactness of \([0,1]\).

2. The measure \( \sigma \) is an invariant measure of the map \( f_r \). This is implied by Lemma 5.5, which is proven using the operator approach described in the next section. The operators \( P_N \) and its dual \( U_N \) that we define are viewed, respectively, as the generalized Frobenius-Perron operator and its dual of the Markov chain \( x_{r,N} \). As shown in Lemma 5.3 and Theorem B.7, \( P_N \) and \( U_N \) converge to the Frobenius-Perron operator of \( f_4 \) and its dual as \( N \to \infty \).

5.1 Generalized Frobenius-Perron operator

We begin the set-up by shifting our viewpoint to operators, specifically Markov operators and their duals. For \( r \in (0, 4], N \in \mathbb{N} \), define an operator \( P_{r,N} \) on \( L^1([0,1]) \) such that for \( g \in L^1([0,1]) \)

\[
P_{r,N}[g(y)] = \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dx.
\] (5.1)

Our first goal is to show that \( P_{r,N} \) is a Markov operator.

**Definition 5.1.** A linear operator \( P : L^1 \to L^1 \) is called a Markov operator if for any given density \( g \in L^1 \), \( P[g] \) is also a density. Thus, \( P \) is a Markov operator if \( P[g] \geq 0 \) and \( \|P[g]\|_1 = 1 \).
Lemma 5.2. For the operator $P_{r,N}$ defined in equation (5.1) we have that $P_{r,N}$ is a Markov operator.

Proof. We will show that $P_{r,N}$ satisfies all the properties:

(a) $P_{r,N}$ is a linear operator on $L^1([0, 1])$. Suppose that $g, h \in L^1([0, 1])$ and $c, d$ are constants.

$$P_{r,N}[cg(y) + dh(y)] = \int_0^1 (cg(x) + dh(x)) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dx$$

$$= \int_0^1 cg(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dx + \int_0^1 dh(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dx$$

$$= cP_{r,N}[g(y)] + dP_{r,N}[h(y)].$$

Thus, $P_{r,N}$ is linear.

(b) Take $g \in L^1([0, 1])$. We will show that $P_{r,N}[g] \in L^1([0, 1])$.

$$\|P_{r,N}[g]\|_1 = \int_0^1 |P_{r,N}[g(y)]| dy$$

$$= \int_0^1 \left| \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dx \right| dy$$

$$\leq \int_0^1 \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dxdy. $$

If we reverse the order of integration in the last step, we have

$$\int_0^1 \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dydx = \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} \left( \int_0^1 e^{-\frac{N}{2}(y - f_r(x))^2} dy \right) dx$$

$$= \int_0^1 |g(x)| dx$$

$$= \|g\|_1,$$

where $\|g\|_1 < \infty$ by assumption. Therefore, by Fubini-Tonelli we have

$$\|P_{r,N}[g]\|_1 \leq \int_0^1 \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y - f_r(x))^2} dxdy$$
Thus, \( \| P_{r,N}[g] \|_1 < \infty \) and so \( P_{r,N}[g] \in L^1([0,1]) \).

(c) Suppose \( g \in L^1([0,1]) \) is a density. Thus, \( g \geq 0 \) and \( \| g \|_1 = 1 \). We will show that \( P_{r,N}[g] \) is a density.

Since \( g \geq 0 \) and \( \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} > 0 \) for any \( x, y \in [0,1] \), then we have

\[
P_{r,N}[g(y)] = \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} dx \geq 0
\]

for any \( y \in [0,1] \). Furthermore,

\[
\| P_{r,N}[g] \|_1 = \int_0^1 \left| \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} dx \right| dy = \int_0^1 \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} dxdy.
\]

Since \( g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \) is non-negative, by Fubini-Tonelli we have

\[
\int_0^1 \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} dxdy = \int_0^1 \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} dxdy = \int_0^1 g(x) dx = 1.
\]

Thus, \( \| P_{r,N}[g] \|_1 = 1 \) and we have that \( P_{r,N}[g] \) is a density on \([0,1]\).

Parts (a), (b) and (c) yield the fact that \( P_{r,N} \) is a Markov operator.

Now that we have shown that \( P_{r,N} \) is a Markov operator, it is interesting to note that the density \( h_{r,N} \) of the invariant measure \( \sigma_{r,N} \) is a fixed point of the operator \( P_{r,N} \). We can then rewrite equation
At this point we will define the dual operator $U_{r,N}$ of $P_{r,N}$. Assume $g \in L^1([0,1])$ and $s$ is a bounded function on $[0,1]$, i.e. $s \in L^\infty([0,1])$. Set

$$\langle P_{r,N}[g], s \rangle = \int_0^1 s(y) P_{r,N}[g(y)] \, dy$$  \quad (5.3)$$

which is well-defined since

$$\left| \int_0^1 s(y) P_{r,N}[g(y)] \, dy \right| \leq \int_0^1 |s(y)||P_{r,N}[g(y)]| \, dy$$

$$\leq \int_0^1 \|s\|_\infty |P_{r,N}[g(y)]| \, dy$$

$$= \|s\|_\infty \|P_{r,N}[g]\|_1 < \infty.$$  

The operator $U_{r,N}$ will be such that for any $g \in L^1([0,1]), s \in L^\infty([0,1])$

$$\langle P_{r,N}[g], s \rangle = \langle g, U_{r,N}[s] \rangle.$$  \quad (5.4)$$

Let’s calculate $U_{r,N}$. Assume $g \in L^1([0,1]), s \in L^\infty([0,1])$.

$$\langle P_{r,N}[g], s \rangle = \int_0^1 s(y) P_{r,N}[g(y)] \, dy$$

$$= \int_0^1 s(y) \left( \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2} (y-f_r(x))^2} \, dx \right) \, dy.$$  

By substituting $|s|$ and $|g|$ for $s$ and $g$ in the above, we have by Fubini-Tonelli

$$\int_0^1 |s(y)| \left( \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2} (y-f_r(x))^2} \, dx \right) \, dy$$
\[ \leq \|s\|_\infty \int_0^1 \left( \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dx \right) \, dy \]
\[ = \|s\|_\infty \int_0^1 \left( \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \right) \, dx \]
\[ = \|s\|_\infty \int_0^1 |g(x)| \frac{1}{b_{r,N}(x)} \left( \int_0^1 e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \right) \, dx \]
\[ = \|s\|_\infty \|g\|_1 < \infty. \]

Therefore, again by Fubini-Tonelli we obtain

\[ \langle P_{r,N}[g], s \rangle = \int_0^1 s(y) \left( \int_0^1 g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dx \right) \, dy \]
\[ = \int_0^1 \left( \int_0^1 s(y) g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dx \right) \, dy \]
\[ = \int_0^1 \left( \int_0^1 s(y) g(x) \frac{1}{b_{r,N}(x)} e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \right) \, dx \]
\[ = \int_0^1 g(x) \frac{1}{b_{r,N}(x)} \left( \int_0^1 s(y) e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \right) \, dx \]
\[ = \langle g, U_{r,N}[s] \rangle, \]

where the operator \( U_{r,N} \) is defined as

\[ U_{r,N}[s(x)] = \frac{1}{b_{r,N}(x)} \int_0^1 s(y) e^{-\frac{N}{2}(y-f_r(x))^2} \, dy. \] (5.5)

Thus, \( U_{r,N} : L^\infty([0, 1]) \to L^\infty([0, 1]) \) since

\[ |U_{r,N}[s(x)]| = \left| \frac{1}{b_{r,N}(x)} \int_0^1 s(y) e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \right| \]
\[ \leq \frac{1}{b_{r,N}(x)} \int_0^1 |s(y)| e^{-\frac{N}{2}(y-f_r(x))^2} \, dy \]
\[
\begin{align*}
&\leq \frac{1}{b_{r,N}(x)} \int_0^1 \|s\|_\infty e^{-\frac{N}{2}(y-f_r(x))^2} dy \\
&= \|s\|_\infty \frac{1}{b_{r,N}(x)} \int_0^1 e^{-\frac{N}{2}(y-f_r(x))^2} dy \\
&= \|s\|_\infty < \infty.
\end{align*}
\]

So \( U_{r,N} \) is a well-defined operator.

Now that we have defined the dual operator \( U_{r,N} \) of \( P_{r,N} \), we will investigate the behavior of the dual operator on continuous functions as \( N \to \infty \). The behavior is stated in the following lemma which we will subsequently prove.

**Lemma 5.3.** Fix \( r \in (0, 4] \). Suppose \( U_{r,N} : L^\infty([0,1]) \to L^\infty([0,1]) \) is defined as in equation (5.5) and \( s \in C([0,1]) \). Then

\[
\lim_{N \to \infty} U_{r,N}[s(x)] = s(f_r(x)) \text{ uniformly.}
\]

**Proof.** Since \( s \) is continuous on \([0,1]\), a compact interval, then \( s \) is bounded and thus \( U_{r,N}[s] \) is defined. We will first prove the result for \( s_1 \in C^1([0,1]) \),

\[
U_{r,N}[s_1(x)] = \frac{1}{b_{r,N}(x)} \int_0^1 s_1(y)e^{-\frac{N}{2}(y-f_r(x))^2} dy.
\]

Set \( z = \sqrt{N}(y - f_r(x)) \), \( dz = \sqrt{N} dy \). We then rewrite \( U_{r,N}[s_1(x)] \) as

\[
U_{r,N}[s_1(x)] = \frac{1}{b_{r,N}(x)} \int_{-\sqrt{N}f_r(x)}^{\sqrt{N}(1-f_r(x))} s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) e^{-\frac{z^2}{2}} \frac{1}{\sqrt{N}} dz.
\]

For an arbitrary \( x \in [0,1] \), set \( v(z) = s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) \), where \( z \in [-\sqrt{N}f_r(x), \sqrt{N}(1-f_r(x))] \).
Since $s_1$ is $C^1$, we can define

$$v'(z) = \left( s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) \right)' = \frac{1}{\sqrt{N}} s_1' \left( f_r(x) + \frac{z}{\sqrt{N}} \right).$$

Thus, for any $z \in \left[-\sqrt{N}f_r(x), \sqrt{N}(1 - f_r(x))\right]$ we can write

$$v(z) - v(0) = \int_0^z v'(t) \, dt$$

and

$$s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) = \int_0^z \frac{1}{\sqrt{N}} s_1' \left( f_r(x) + \frac{t}{\sqrt{N}} \right) \, dt.$$

By taking absolute value on both sides we get

$$\left| s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) \right| = \left| \int_0^z \frac{1}{\sqrt{N}} s_1' \left( f_r(x) + \frac{t}{\sqrt{N}} \right) \, dt \right|$$

$$\leq \frac{1}{\sqrt{N}} \|s_1'\|_{\infty} \left| \int_0^z dt \right|$$

$$= \frac{1}{\sqrt{N}} \|s_1'\|_{\infty} |z|,$$

which is finite since $s_1'$ is continuous on a compact interval.

Therefore, we have

$$U_{r,N}[s_1(x)] = \frac{1}{b_{r,N}(x)} \int_0^1 s_1(y) e^{-\frac{N}{2}(y-f_r(x))^2} \, dy$$

$$= \frac{1}{\sqrt{Nb_{r,N}(x)}} \int_{\sqrt{N}f_r(x)}^{\sqrt{N}(1-f_r(x))} s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) e^{-\frac{z^2}{2}} \, dz$$

$$= \frac{1}{\sqrt{Nb_{r,N}(x)}} \int_{-\sqrt{N}f_r(x)}^{\sqrt{N}(1-f_r(x))} \left[ s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) + s_1(f_r(x)) \right] e^{-\frac{z^2}{2}} \, dz$$

$$= \frac{1}{\sqrt{Nb_{r,N}(x)}} \int_{-\sqrt{N}f_r(x)}^{\sqrt{N}(1-f_r(x))} \left[ s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) \right] e^{-\frac{z^2}{2}} \, dz$$

$$+ \frac{1}{\sqrt{Nb_{r,N}(x)}} \int_{-\sqrt{N}f_r(x)}^{\sqrt{N}(1-f_r(x))} s_1(f_r(x)) e^{-\frac{z^2}{2}} \, dz.$$
\[
U_{r,N} = \frac{1}{\sqrt{N} b_{r,N}(x)} \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} \left[ s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) \right] e^{-\frac{z^2}{2}} \, dz + s_1(f_r(x)).
\]

Placing \( s_1(f_r(x)) \) on the left-hand side and taking absolute value, we get

\[
|U_{r,N}[s_1(x)] - s_1(f(x))| = \left| \frac{1}{\sqrt{N} b_{r,N}(x)} \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} \left[ s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) \right] e^{-\frac{z^2}{2}} \, dz \right|
\]

\[
\leq \frac{1}{\sqrt{N} b_{r,N}(x)} \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} \left| s_1 \left( f_r(x) + \frac{z}{\sqrt{N}} \right) - s_1(f_r(x)) \right| e^{-\frac{z^2}{2}} \, dz
\]

\[
\leq \frac{1}{\sqrt{N} b_{r,N}(x)} \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} \frac{1}{\sqrt{N}} \|s_1\|_{\infty} |z| e^{-\frac{z^2}{2}} \, dz
\]

\[
= \frac{\sqrt{N}}{\sqrt{N} b_{r,N}(x)} \frac{1}{\sqrt{N}} \|s_1\|_{\infty} \int_{-\sqrt{N} f_r(x)}^{\sqrt{N}(1-f_r(x))} |z| e^{-\frac{z^2}{2}} \, dz
\]

\[
\leq \frac{1}{\beta \sqrt{N}} \|s_1\|_{\infty} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} \, dz,
\]

where the \( \beta \) is the same bound given in Proposition 3.2. Let’s take a look at the terms separately.

As \( N \to \infty \) we have the following:

\[
\frac{1}{\beta \sqrt{N}} \|s_1\|_{\infty} \to 0
\]

and

\[
\int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} \, dz = 2 \int_{0}^{\infty} z e^{-\frac{z^2}{2}} \, dz = -2 \left[ e^{-\frac{z^2}{2}} \right]_{0}^{\infty} = 2.
\]

Therefore, the above yield

\[
\frac{1}{\beta \sqrt{N}} \|s_1\|_{\infty} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} \, dz = \frac{2}{\beta \sqrt{N}} \|s_1\|_{\infty} \to 0
\]

as \( N \to \infty \), independent of the variable \( x \) which in turn implies

\[
\lim_{N \to \infty} U_{r,N}[s_1(x)] = s_1(f_r(x)) \text{ uniformly.} \tag{5.6}
\]
To show that the result holds for any \( s \in C([0, 1]) \), we use the fact that \( C^1([0, 1]) \) is a dense subset of \( C([0, 1]) \) when equipped with the supremum norm. So, given \( s \in C([0, 1]) \) and \( \varepsilon > 0 \), there exists an \( s_1 \in C^1([0, 1]) \) such that
\[
\|s - s_1\|_{\infty} < \varepsilon.
\]

Thus,
\[
|U_{r,N}[s(x)] - s(f_r(x))| = |U_{r,N}[s(x)] - U_{r,N}[s_1(x)] + U_{r,N}[s_1(x)] - s_1(f_r(x)) + s_1(f_r(x)) - s(f_r(x))| \leq |U_{r,N}[s(x)] - U_{r,N}[s_1(x)]| + |U_{r,N}[s_1(x)] - s_1(f_r(x))| + |s_1(f_r(x)) - s(f_r(x))|
\]

\[= I_1 + I_2 + I_3.\]

Let’s investigate each term separately

\[
I_1 = |U_{r,N}[s(x)] - U_{r,N}[s_1(x)]| = \left| \frac{1}{b_{r,N}(x)} \int_0^1 s(y)e^{-\frac{N}{2}(y-f_r(x))^2} dy - \frac{1}{b_{r,N}(x)} \int_0^1 s_1(y)e^{-\frac{N}{2}(y-f_r(x))^2} dy \right|
\]

\[= \left| \frac{1}{b_{r,N}(x)} \int_0^1 (s(y) - s_1(y)) e^{-\frac{N}{2}(y-f_r(x))^2} dy \right| \leq \frac{1}{b_{r,N}(x)} \int_0^1 |s(y) - s_1(y)| e^{-\frac{N}{2}(y-f_r(x))^2} dy \]

\[\leq \frac{1}{b_{r,N}(x)} \int_0^1 \|s - s_1\|_{\infty} e^{-\frac{N}{2}(y-f_r(x))^2} dy \leq \|s - s_1\|_{\infty} \frac{1}{b_{r,N}(x)} \int_0^1 e^{-\frac{N}{2}(y-f_r(x))^2} dy \]

\[= \|s - s_1\|_{\infty} < \varepsilon.\]

\[
I_2 = |U_{r,N}[s_1(x)] - s_1(f_r(x))| \to 0 \text{ as } N \to \infty \text{ due to (5.6). Thus, there exists } N_0 \text{ large}
\]
enough such that $I_2 < \varepsilon$ for $N > N_0$. Finally,

$$I_3 = |s_1(f_r(x)) - s(f_r(x))| \leq \|s - s_1\|_\infty < \varepsilon.$$ 

Therefore, for $N > N_0$, we have that

$$[U_{r,N}[s(x)] - s(f_r(x))] \leq \varepsilon + \varepsilon + \varepsilon = 3\varepsilon$$

and we conclude that for any $s \in C(0,1)$,

$$\lim_{N \to \infty} U_{r,N}[s(x)] = s(f_r(x))$$

uniformly for $x \in [0,1]$. \hfill \Box

This is where our preliminary work ends towards the second asymptotic result. We are now ready to state and prove it.

### 5.2 Main result

**Theorem 5.4.** Let $r \in (0,4]$. Suppose $\{\sigma_{r,N}\}_{N \in \mathbb{N}}$ is a sequence of measures given by $\sigma_{r,N}(dy) = h_{r,N}(y)dy$. Therefore, $\sigma_{r,N}$ is the invariant measure for the Random Logistic Model. As $N \to \infty$, depending on the value of $r$, we have the following results:

(i) For $r \in (0,1]$ we have $\sigma_{r,N} \Rightarrow \delta_0$.

(ii) For $r \in (1,3]$ we have $\sigma_{r,N} \Rightarrow \delta_{1-1/r}$.

(iii) For $r \in (3,r_\infty)$ we have $\sigma_{r,N} \Rightarrow \frac{1}{2^k} \sum_{i=1}^{2^k} \delta_{p_{r,k}^{(i)}}$, where $k = k(r)$.

(iv) For $r = 4$ we have $\sigma_{r,N} \Rightarrow \sigma^*$, where $\sigma^*$ is a measure with density $h(y) = \frac{1}{\pi \sqrt{y(1-y)}}$.

In the above results, $\Rightarrow$ denotes weak convergence and $\delta_x$ denotes the Dirac measure centered at $x$.

Before we begin the proof we will make an observation and state and prove a lemma which will be utilized in the proof. Since $[0,1]$ is a compact interval, then any sequence of measures defined
on it is tight. Thus, take any subsequence \( \{\sigma_{r,N'}\} \) of \( \{\sigma_{r,N}\} \). Since \( \{\sigma_{r,N'}\} \) is tight, there exists a subsequence \( \{\sigma_{r,N''}\} \) of \( \{\sigma_{r,N'}\} \) such that \( \sigma_{r,N''} \Rightarrow \sigma_r \) for some measure \( \sigma_r \).

**Lemma 5.5.** Suppose \( \{\sigma_{r,N''}\} \) is defined as above, which converges weakly to some measure \( \sigma_r \) on \([0,1] \). Then, for any \( t \in \mathbb{N} \) and \( g \in C([0,1]) \) we have

\[
\lim_{N \to \infty} \int_0^1 g(y) \, d\sigma_{r,N''}(y) = \int_0^1 g(f_r^{(t)}(y)) \, d\sigma_r(y). \tag{5.7}
\]

**Proof.** We will prove (5.7) by induction. First, let’s show it for \( t = 1 \):

Pick \( \epsilon > 0 \). Since \( g \) and \( f_r \) are continuous functions on \([0,1] \) and \( f_r([0,1]) \subset [0,1] \), then the composition \( g \circ f_r \) is defined, is continuous and bounded. Thus,

\[
\lim_{N \to \infty} \int_0^1 g(f_r(y)) \, d\sigma_{r,N''}(y) = \int_0^1 g(f_r(y)) \, d\sigma_r(y).
\]

Now,

\[
\int_0^1 g(y) \, d\sigma_{r,N''}(y) = \int_0^1 g(y)h_{r,N''}(y) \, dy
\]

and by (5.2) and (5.4) we have

\[
\int_0^1 g(y)h_{r,N''}(y) \, dy = \int_0^1 g(y)P_{r,N}[h_{r,N''}(y)] \, dy = \int_0^1 U_{r,N}[g(y)]h_{r,N''}(y) \, dy.
\]

By Lemma 5.3 we have that \( U_{r,N}[g(y)] \) converges to \( g(f_r(y)) \) uniformly.

Thus, there exists \( N_1 \) large enough such that for \( N'' > N_1 \)

\[
\left| \int_0^1 g(f_r(y)) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) \, d\sigma_r(y) \right| < \epsilon
\]

and there exists \( N_2 \) large enough such that for \( N'' > N_2 \)

\[
|U_{r,N''}[g(y)] - g(f_r(y))| < \epsilon \text{ for any } y \in [0,1],
\]

48
and therefore,

\[
\left| \int_0^1 g(y) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_{r,N''}(y) \right|
\]

\[
= \left| \int_0^1 U_{r,N''}[g(y)] h_{r,N''}(y) dy - \int_0^1 g(f_r(y)) h_{r,N''}(y) dy \right|
\]

\[
= \left| \int_0^1 (U_{r,N''}[g(y)] - g(f_r(y))) h_{r,N''}(y) dy \right|
\]

\[
\leq \int_0^1 |U_{r,N''}[g(y)] - g(f_r(y))| h_{r,N''}(y) dy
\]

\[
\leq \int_0^1 \varepsilon h_{r,N''}(y) dy = \varepsilon.
\]

So, assign \( N_0 = \max\{N_1, N_2\} \). Thus, for \( N'' > N_0 \) we have

\[
\left| \int_0^1 g(y) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_r(y) \right|
\]

\[
= \left| \int_0^1 g(y) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_{r,N''}(y) + \int_0^1 g(f_r(y)) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_r(y) \right|
\]

\[
\leq \left| \int_0^1 g(y) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_{r,N''}(y) \right| + \left| \int_0^1 g(f_r(y)) d\sigma_{r,N''}(y) - \int_0^1 g(f_r(y)) d\sigma_r(y) \right|
\]

\[
\leq 2\varepsilon
\]

which implies the result for \( t = 1 \).

Now suppose (5.7) holds for \( t = k \). We will show it holds for \( t = k + 1 \). Pick \( \varepsilon > 0 \). Therefore, by assumption there exists \( N_1 \) large enough such that for \( N'' > N_1 \)

\[
\left| \int_0^1 g(y) d\sigma_{r,N''}(y) - \int_0^1 g(f_r^{(k)}(y)) d\sigma_r(y) \right| < \varepsilon.
\]

Now, as before, since compositions of \( f_r \) with itself are continuous functions, then \( g \circ f_r^k \) and \( g \circ f_r^{k+1} \)
are continuous functions on \([0, 1]\) and thus bounded. So, by definition of weak convergence,

\[
\lim_{N \to \infty} \int_0^1 g(f_r^{(k)}(y)) \, d\sigma_{r,N''}(y) = \int_0^1 g(f_r^{(k)}(y)) \, d\sigma_r(y)
\]

and

\[
\lim_{N \to \infty} \int_0^1 g(f_r^{(k+1)}(y)) \, d\sigma_{r,N''}(y) = \int_0^1 g(f_r^{(k+1)}(y)) \, d\sigma_r(y).
\]

Thus there exists \(N_2\) large enough such that for \(N'' > N_2\)

\[
\left| \int_0^1 g(f_r^{(k)}(y)) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r^{(k)}(y)) \, d\sigma_r(y) \right| < \epsilon
\]

and

\[
\left| \int_0^1 g(f_r^{(k+1)}(y)) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r^{(k+1)}(y)) \, d\sigma_r(y) \right| < \epsilon.
\]

Finally, since Lemma 5.3 holds for any continuous function on \([0, 1]\) and we have that \(g \circ f_r^{(k)} \in C([0, 1])\), then

\[
\lim_{N \to \infty} U_{r,N}[g(f_r^{(k)}(y))] = g(f_r^{(k)}(f_r(y))) = g(f_r^{(k+1)}(y)) \text{ uniformly},
\]

and thus there exists \(N_3\) large enough such that for any \(N'' > N_3\), we have

\[
|U_{r,N''}[g(f_r^{(k)}(y))] - g(f_r^{(k+1)}(y))| < \epsilon \text{ for any } y \in [0, 1].
\]

Therefore,

\[
\left| \int_0^1 g(f_r^{(k)}(y)) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r^{(k+1)}(y)) \, d\sigma_{r,N''}(y) \right|
\]

\[
= \left| \int_0^1 U_{r,N''}[g(f_r^{(k)}(y))] h_{r,N''}(y) \, dy - \int_0^1 g(f_r^{(k+1)}(y)) h_{r,N''}(y) \, dy \right|
\]

\[
= \left| \int_0^1 (U_{r,N''}[g(f_r^{(k)}(y))] - g(f_r^{(k+1)}(y))) h_{r,N''}(y) \, dy \right|
\]
≤ \int_0^1 \left[ U_{r,N''}[g(f_r^{(k)}(y))] - g(f_r^{(k+1)}(y))\right] h_{r,N''}(y) \, dy \\
≤ \int_0^1 \epsilon h_{r,N''}(y) \, dy \\
= \epsilon.

Thus, by choosing \( N_0 = \max\{N_1, N_2, N_3\} \), for \( N'' > N_0 \) we have

\[
\left| \int_0^1 g(y) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r^k(y)) \, d\sigma_r(y) \right| \\
\leq \left| \int_0^1 g(y) \, d\sigma_{r,N''}(y) - \int_0^1 g(f_r^k(y)) \, d\sigma_r(y) \right| \\
+ \left| \int_0^1 g(f_r^k(y)) \, d\sigma_r(y) - \int_0^1 g(f_r^{k+1}(y)) \, d\sigma_{r,N''}(y) \right| \\
+ \left| \int_0^1 g(f_r^{k+1}(y)) \, d\sigma_{r,N''}(y) - \int_0^1 g^2(f_r^k(y)) \, d\sigma_r(y) \right| \\
\leq 4\epsilon.
\]

Thus, (5.7) holds for \( t = k + 1 \) which concludes the induction and the proof of the lemma.

Now let’s prove Theorem 5.4:

**Proof.** From the dynamics of the logistic map \( f_r(x) = rx(1-x) \), we know that \( f_r^{(t)}(x) \to 0 \) as \( t \to \infty \), \( r \in (0, 1] \) and \( x \in [0, 1] \). Furthermore, for \( r \in (1, 3] \) we know that \( f_r^{(t)}(x) \to 1 - 1/r \) as \( t \to \infty \) and \( x \in (0, 1) \). Take any \( g \in C([0, 1]) \). By continuity, we have

\[
\lim_{t \to \infty} g(f_r^{(t)}(x)) = g(0), \quad x \in [0, 1], \quad r \in (0, 1],
\]

\[
\lim_{t \to \infty} g(f_r^{(t)}(x)) = g(1 - 1/r), \quad x \in (0, 1), \quad r \in (1, 3].
\]

(i) Since \( g \) is bounded on \([0, 1]\) we can apply the Bounded Convergence Theorem to \( g \circ f_r^{(t)} \) and for
$r \in (0, 1]$ we have

$$\lim_{t \to \infty} \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) = \int_{[0,1]} g(0) d\sigma_r(y)$$

$$= g(0) \int_{[0,1]} d\sigma_r(y)$$

$$= g(0).$$

Thus, pick $\epsilon > 0$. There exists $t_0$ large enough such that for $t > t_0$ we have

$$\left| \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) - g(0) \right| < \epsilon$$

and applying Lemma 5.5 on $g$ we have that there exists $N_0$ large enough such that

$$\left| \int_{[0,1]} g(y) d\sigma_{r,N''}(y) - \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) \right| < \epsilon.$$

Thus,

$$\left| \int_{[0,1]} g(y) d\sigma_{r,N''}(y) - g(0) \right|$$

$$= \left| \int_{[0,1]} g(y) d\sigma_{r,N''}(y) - \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) + \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) - g(0) \right|$$

$$\leq \left| \int_{[0,1]} g(y) d\sigma_{r,N''}(y) - \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) \right| + \left| \int_{[0,1]} g(f_t^r(y)) d\sigma_r(y) - g(0) \right|$$

$$\leq 2\epsilon.$$

Since

$$g(0) = \int_{[0,1]} g(y) d\delta_0(y),$$

we have that $\sigma_{r,N''} \Rightarrow \delta_0$. Now, since this holds for an arbitrary subsequence of a subsequence, we conclude that the sequence itself must converge to the same limit and thus $\sigma_{r,N} \Rightarrow \delta_0$ which concludes the proof of Theorem 5.4 (i).
(ii) Suppose \( r \in (1, 3] \) and \( g \in C[0, 1] \). Since \( \lim_{t \to \infty} g(f^{(t)}(x)) = g(1 - 1/r), \quad x \in (0, 1), \ r \in (1, 3] \) and \( g(f^{(t)}(0)) = g(f^{(t)}(1)) = 0 \), by applying the Bounded Convergence Theorem we obtain

\[
\lim_{t \to \infty} \int_{[0,1]} g(f^{(t)}(y)) \, d\sigma_r(y) = \int_{[0,1]} \left[ g\left(1 - 1/r\right) \cdot \mathbb{1}_{(0,1)}(y) + g(0) \cdot \mathbb{1}_{\{0,1\}} \right] \, d\sigma_r(y)
\]

\[
= g\left(1 - 1/r\right) \sigma_r((0,1)) + g(0) \sigma_r(\{0,1\}).
\]

At this point we can determine that \( \sigma_r(\{1\}) = 0 \). Pick \( \varepsilon > 0 \) small enough. By Theorem 3.3 we conclude that

\[
\lim_{t \to \infty} \int_{[0,1]} g(f^{(t)}(y)) \, d\sigma_r(y) = g\left(1 - 1/r\right) \sigma_r((0,1)) + g(0) \sigma_r(\{0,1\}).
\]

Continuing as in part (i), we obtain the form for the weak limit of \( \sigma_{r,N''} \) to be

\[
\sigma_r = \alpha_1 \delta_0 + \alpha_2 \delta_{1-1/r} \quad \text{where} \quad \alpha_1 = \sigma_r(\{0\}), \ \alpha_2 = \sigma_r((0,1)), \ \alpha_1 + \alpha_2 = 1.
\]

To conclude that \( \sigma_r = \delta_{1-1/r} \), we show that \( \sigma_r(\{0\}) = 0 \). To this end, we state and prove the following claim:

Pick \( \delta > 0 \) small. For any positive integer \( n \) and \( r \in (1, 3] \), we have
\[
\sigma_{r,N}([0, \delta]) \leq \int_{A_{\delta}(n)} \int_{A_{\delta}(n-1)} \cdots \int_{A_{\delta}(0)} h_{r,N}(x_n) q_N(x_n, x_{n-1}) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_n \\
+ \gamma \sqrt{N\delta} \sum_{i=0}^{n-1} \left( \frac{2}{1+r} \right)^i \exp \left\{ -\frac{N}{2} \left( \frac{2}{1+r} \right)^i \delta \right\} \left( r - 1 \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \left( \frac{2}{1+r} \right)^i \delta \right)^2 \right\} \\
+ \gamma \sqrt{N\delta} \exp \left\{ -\frac{N}{2} \right\} \sum_{i=1}^{n} \left( \frac{2}{1+r} \right)^i ,
\]

where the sets \( A_{\delta}(n) \) are defined as

\[
A_{\delta}(n) = \left[ 0, \left( \frac{2}{1+r} \right)^n \delta \right], \quad n \geq 0,
\]

and the densities

\[
q_N(x, y) = \frac{1}{b_{r,N}(x)} \exp \left\{ -\frac{N}{2} (y - f_r(x))^2 \right\}.
\]

Proof of claim:

We show this by induction. Before we begin, we need to explain why we chose the sets above the way we did. Since \( 1 < r \leq 3 \), notice that \( 2/(1+r) < 1 \) and, thus, \( A_{\delta}(n) \subset A_{\delta}(n-1) \).

Furthermore, compare the line \( y = [(r + 1)/2]x \) to the curve \( y = f_r(x) \). We have that \( f_r(x) = [(r + 1)/2]x \) for \( x = 0 \) and \( x = 1/2 - 1/2r \). On the interval \((0, 1/2 - 1/2r)\) we have that \( f_r(x) > [(r + 1)/2]x \). Therefore, for \( \delta < 1/2 - 1/2r \) we have that \( f_r^{-1}(A_{\delta}(n-1)) \subset A_{\delta}(n) \).

These motivated us to define the sets. Now, we start with the base case, \( n = 1 \). Let \( I = [0, 1] \).

Applying (2.7) to \( \sigma_{r,N}([0, \delta]) \), we have

\[
\sigma_{r,N}([0, \delta]) = \int_{A_{\delta}(0)} h_{r,N}(y) \, dy \\
= \int_I \int_{A_{\delta}(0)} h_{r,N}(x) q_N(x, y) \, dy \, dx
\]

54
\[= \int_{I_1 \setminus [A_\delta(1) \cup (1 - A_\delta(1))] \cap A_\delta(0)} h_{r,N}(x)q_N(x,y) \, dy \, dx + \int_{A_\delta(1)} \int_{A_\delta(0)} h_{r,N}(x)q_N(x,y) \, dy \, dx + \int_{1 - A_\delta(1)} \int_{A_\delta(0)} h_{r,N}(x)q_N(x,y) \, dy \, dx\]

\[= I_1 + I_2 + I_3,\]

where \(1 - A_\delta(n) := \{1 - x \mid x \in A_\delta(n)\}\). Notice that the term \(I_2\) is the first term in (5.8) for \(n = 1\).

Now, by choosing \(\delta\) small so that \(1 - \delta > \frac{r}{4} + \delta\), we have from Theorem 3.3 that

\[I_3 \leq \int_{1 - A_\delta(1)} h_{r,N}(x) \, dx \leq \gamma \sqrt{N} \exp \left( -\frac{N}{2} \delta^2 \right) m(1 - A_\delta(1)) = \gamma \sqrt{N} \exp \left( -\frac{N}{2} \delta^2 \right) \frac{2}{1 + r},\]

which covers the third term of (5.8) for \(n = 1\). Finally, for \(I_1\) observe that \(x \not\in A_\delta(1) \cup (1 - A_\delta(1))\), which implies

\[|y - f_r(x)| \geq f_r \left( \frac{2}{1 + r} \delta \right) - \delta = r \frac{2}{1 + r} \delta \left( 1 - \frac{2}{1 + r} \delta \right) - \delta = \frac{r - 1}{r + 1} \delta - r \frac{4}{(1 + r)^2} \delta^2 = \frac{r - 1}{r + 1} \delta \left( 1 - \frac{4r}{r^2 - 1} \delta \right) > 0.\]

This implies

\[q_N(x,y) \leq \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \left( \frac{r - 1}{r + 1} \delta \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \delta \right)^2 \right\}.

55
and we conclude

\[ I_1 \leq \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \left( \frac{r-1}{r+1} \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \delta \right)^2 \right\} m(A_\delta(0)) \int_{\Gamma \setminus \{A_\delta(1) \cup (1-A_\delta(1))\}} h_{r,N}(x) \, dx \]

\[ \leq \gamma \sqrt{N} \delta \exp \left\{ -\frac{N}{2} \left( \frac{r-1}{r+1} \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \delta \right)^2 \right\}. \]

This ends the base case. For the induction step, assume (5.8) holds for \( n \) and we will show it holds for \( n + 1 \). Consider the first term of that inequality and apply (2.7). This yields

\[
\int_{A_\delta(n)} \cdots \int_{A_\delta(0)} h_{r,N}(x_n) q_N(x_n, x_{n-1}) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_n = \\
= \int I \int_{A_\delta(n)} \cdots \int_{A_\delta(0)} h_{r,N}(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[
= \int I \int_{A_\delta(n)} \cdots \int_{A_\delta(0)} h_{r,N}(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[
+ \int_{A_\delta(n+1)} \int_{A_\delta(n)} \cdots \int_{A_\delta(0)} h_{r,N}(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[
+ \int_{1-A_\delta(n+1)} \int_{A_\delta(n)} \cdots \int_{A_\delta(0)} h_{r,N}(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[ = I'_1 + I'_2 + I'_3. \]

As in the base case, notice that \( I'_2 \) is the first term of (5.8) for \( n + 1 \). Next, \( I'_3 \) can be bounded as follows by Theorem 3.3.

\[ I'_3 \leq \int_{1-A_\delta(n+1)} h_{r,N}(x_{n+1}) \, dx_{n+1} \]

\[ \leq \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \delta^2 \right\} m (1 - A_\delta(n+1)) \]

\[ = \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \delta^2 \right\} \left( \frac{2}{1+r} \right)^{n+1} \delta, \]

which is then added to the third term of (5.8), and the exponential is justified by
\[ 1 - \left( \frac{2}{1 + r} \right)^{n+1} \delta > 1 - \delta > \delta + r/4. \]

Finally, the term \( I'_1 \) is bounded as follows,

\[ I'_1 \leq \int_{I_1(A_n+1) \cup (1-A_n+1))} \int_{A_n(n)} h_{r,N}(x_{n+1})q_N(x_{n+1}, x_n) dx_n \, dx_{n+1}. \]

Observe that \( x_{n+1} \notin A_n(n + 1) \cup (1 - A_n(n + 1)) \), which implies

\[ |x_n - x_{n+1}| \geq f_r \left( \left( \frac{2}{1 + r} \right)^{n+1} \delta \right) - \left( \frac{2}{1 + r} \right)^n \delta \]

\[ = r \left( \frac{2}{1 + r} \right)^{n+1} \delta \left( 1 - \left( \frac{2}{1 + r} \right)^{n+1} \right) - \left( \frac{2}{1 + r} \right)^n \delta \]

\[ = \frac{r - 1}{r + 1} \left( \frac{2}{1 + r} \right)^n \delta - \frac{2r}{1 + r} \left( \frac{2}{1 + r} \right)^n \left( \frac{2}{1 + r} \right)^{n+1} \delta^2 \]

\[ = \frac{r - 1}{r + 1} \left( \frac{2}{1 + r} \right)^n \delta \left( 1 - \frac{2r}{r - 1} \left( \frac{2}{1 + r} \right)^n \delta \right) \]

\[ = \frac{r - 1}{r + 1} \left( \frac{2}{1 + r} \right)^n \delta \left( 1 - \frac{4r}{r^2 - 1} \left( \frac{2}{1 + r} \right)^n \delta \right). \]

The last expression above yields

\[ I'_1 \leq \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \left( \frac{r - 1}{r + 1} \right)^2 \left( \frac{2}{1 + r} \right)^{2n} \left( 1 - \frac{4r}{r^2 - 1} \left( \frac{2}{1 + r} \right)^n \delta \right)^2 \right\} m(A_n(n)) \]

\[ = \gamma \sqrt{N} \left( \frac{2}{1 + r} \right)^n \delta \exp \left\{ -\frac{N}{2} \left( \frac{r - 1}{r + 1} \right)^2 \left( \frac{2}{1 + r} \right)^{2n} \left( 1 - \frac{4r}{r^2 - 1} \left( \frac{2}{1 + r} \right)^n \delta \right)^2 \right\}, \]

which is then added to the second term of (5.8) and, in turn, concludes the induction step and with it the proof of the claim.

The next step is to show that the bound in (5.8) is asymptotically bounded by \( 2\gamma \delta \), which will guarantee that \( \sigma_r(\{0\}) = 0 \). Let’s start with the first term. Notice that it can be bounded as follows,
The last expression above dictates that we should choose \( n \) to be a function of \( N \) so that the expression is bounded as \( N \to \infty \). We take \( n = n(N) \) such that

\[
n = \left[ \log_{1+2r}(\sqrt{N}) \right] + 1. \tag{5.9}
\]

Thus, with this choice we have

\[
\gamma \sqrt{N} \left( \frac{2}{1 + r} \right)^n \delta \leq \gamma \sqrt{N} \frac{1}{\sqrt{N}} \delta = \gamma \delta.
\]

Next, consider the third term of (5.8). We immediately have

\[
\gamma \sqrt{N} \delta \exp \left[ -\frac{N}{2} \delta^2 \right] \sum_{i=1}^{n} \left( \frac{2}{1 + r} \right)^i \leq \gamma \sqrt{N} \delta \exp \left[ -\frac{N}{2} \delta^2 \right] \sum_{i=1}^{\infty} \left( \frac{2}{1 + r} \right)^i
\]

\[
= \gamma \sqrt{N} \delta \exp \left[ -\frac{N}{2} \delta^2 \right] \frac{2/(1 + r)}{1 - 2/(1 + r)} \to 0 \quad \text{as} \quad N \to \infty.
\]

Now for the second term of (5.8). Consider it without the \( \gamma \delta \), so we have

\[
\sum_{i=0}^{n-1} \sqrt{N} \left( \frac{2}{1 + r} \right)^i \exp \left\{ -\frac{N}{2} \left( \frac{2}{1 + r} \right)^{2i} \left( \delta - r - 1 \right)^{2} \left( 1 - 4r \right) \right\}.
\]

It is a two-step process to show that it is asymptotically bounded by 1. First we bound the sum using
the following inequality. For any $x \geq 0$, it holds that

$$\sqrt{x} \exp\{-x\} \leq \exp\{-\sqrt{x}\},$$

(5.10)

with equality only when $x = 1$.

$$\sum_{i=0}^{n-1} \sqrt{N} \left( \frac{2}{1+r} \right)^i \exp\left\{ -\frac{N}{2} \left( \frac{2}{1+r} \right)^{2i} \left( \frac{\delta - 1}{r+1} \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \right)^2 \right\}$$

$$\leq \sum_{i=0}^{n-1} \sqrt{N} \left( \frac{2}{1+r} \right)^i \exp\left\{ -\frac{N}{2} \left( \frac{2}{1+r} \right)^{2i} \left( \frac{\delta - 1}{r+1} \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \right)^2 \right\}$$

$$\leq E(\delta, r)^{-1} \sum_{i=0}^{n-1} \sqrt{N} \left( \frac{2}{1+r} \right)^{2i} E(\delta, r)^2 \exp\left\{ -\frac{N}{2} \left( \frac{2}{1+r} \right)^{2i} \left( \frac{\delta - 1}{r+1} \right)^2 \left( 1 - \frac{4r}{r^2 - 1} \right)^2 \right\}$$

$$\leq E(\delta, r)^{-1} \sum_{i=0}^{n-1} \exp\left\{ -\sqrt{N} \left( \frac{2}{1+r} \right)^i E(\delta, r) \right\}$$

$$\leq E(\delta, r)^{-3/2} \sum_{i=0}^{n-1} N^{-1/4} \left( \frac{2}{1+r} \right)^{-i/2} \exp\left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{i/2} E(\delta, r)^{1/2} \right\},$$

by applying (5.10) twice, where

$$E(\delta, r) = \frac{\delta - 1}{\sqrt{2}r + 1} \left( 1 - \frac{4r}{r^2 - 1} \delta \right).$$

This concludes the first step. For the next step, let

$$\phi(x) = |x|^{((1+r)/2)^n},$$

which is a convex function. Thus, by applying Jensen’s inequality, we obtain

$$\phi \left( \sum_{i=0}^{n-1} E(\delta, r)^{-3/2} N^{-1/4} \left( \frac{2}{1+r} \right)^{-i/2} \exp\left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{i/2} E(\delta, r)^{1/2} \right\} \right)$$
\[
\phi \left( \frac{1}{n} \sum_{i=0}^{n-1} nE(\delta, r)^{-3/2}N^{-1/4} \left( \frac{2}{1+r} \right)^{-i/2} \exp \left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{i/2} E(\delta, r)^{1/2} \right\} \right) \\
\leq \frac{1}{n} \sum_{i=0}^{n-1} n^{((1+r)/2)^{n}} nE(\delta, r)^{-3/2((1+r)/2)^{n}} N^{-1/4((1+r)/2)^{n}} \left( \frac{2}{1+r} \right)^{-(i/2)((1+r)/2)^{n}} \\
\times \exp \left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{i/2-N} E(\delta, r)^{1/2} \right\} \\
= I_4.
\]

By taking into account (5.9) (and dropping the base for convenience), \( I_4 \) turns into

\[
I_4 \leq \frac{1}{\log\left(\sqrt{N}\right)} \sum_{i=0}^{\log(\sqrt{N})} \left(N^{-1/4} E(\delta, r)^{-3/2} \log\left(\sqrt{N}\right)\right)^{\sqrt{N}} \\
\times \exp \left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{\log(\sqrt{N})/2-\log(\sqrt{N})} E(\delta, r)^{1/2} \right\} \\
\leq \left(N^{-1/4} E(\delta, r)^{-3/2} \log\left(\sqrt{N}\right)\right)^{\sqrt{N}} \exp \left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{\log(\sqrt{N})/2-\log(\sqrt{N})} E(\delta, r)^{1/2} \right\} \\
= \left(N^{-1/4} E(\delta, r)^{-3/2} \log\left(\sqrt{N}\right)\right)^{\sqrt{N}} \exp \left\{ -N^{1/4} N^{-1/4} E(\delta, r)^{1/2} \right\} \\
\leq \left(N^{-1/4} E(\delta, r)^{-3/2} \log\left(\sqrt{N}\right)\right)^{\sqrt{N}} \to 0 \quad \text{as} \quad N \to \infty.
\]

Therefore,

\[
\phi \left( \sum_{i=0}^{n-1} N^{-1/4} \left( \frac{2}{1+r} \right)^{-i/2} \exp \left\{ -N^{1/4} \left( \frac{2}{1+r} \right)^{i/2} E(\delta, r)^{1/2} \right\} \right) \to 0,
\]

and since \( \phi(x) \to 0 \iff |x| \leq 1 \), we finally conclude that the second term of (5.8) is bounded by \( \gamma \delta \) as \( N \to \infty \).

Now, if it were true that \( \sigma_r(\{0\}) = \alpha_1 > 0 \), then for \( \varepsilon > 0 \) there exists \( N_0 \) large such that we would have \( \sigma_{4,N''}(\{0, \delta\}) > \alpha_1 - \varepsilon \) for all \( N'' > N_0 \). However, by picking \( \delta \) small so that
which is a contradiction. Thus, we show that there can be no accumulation of probability mass at
\( x = 0 \), which implies
\[ \sigma_r = \delta_{1-1/r}. \]

(iii) To prove this part first observe that the proof of \( \sigma_r(\{0\}) = 0 \) in the previous part can be extended
to values of \( r \in (3, 4) \). Thus, in the case of the \( 2^k \)-cycles, it still holds that \( \sigma_r(\{0\}) \) vanishes.
Now, let \( p_1, p_2, \ldots, p_{2^k} \) be the elements in the stable \( 2^k \)-cycle and \( B_1, B_2, \ldots, B_{2^k} \) their basins of
attraction, respectively. Also note that \( f_r(p_i) = p_{i+1} \) for \( i = 1, \ldots, 2^k - 1 \) and \( f_r(p_{2^k}) = p_1 \). We
will prove that
\[ \sigma_r = \frac{1}{2^k} \sum_{i=1}^{2^k} \delta_{p_i} \]
using the above together with Lemma 5.5. Now, from that Lemma we obtain the following for
\( g \in C[0, 1] \) and \( t \) any positive integer.

\[ \lim_{N'' \to \infty} \int_I g(x) \, d\sigma_{r,N''}(x) = \int_I g(f^{(jt)}(x)) \, d\sigma_r(x), \]
for \( k \in \{1, 2, \ldots, 2^n\} \). Thus, by summing both sides from 1 to \( 2^n \), we obtain

\[ \lim_{N'' \to \infty} \int_I g(x) \, d\sigma_{r,N''}(x) = \frac{1}{2^k} \int_I \left( \sum_{j=1}^{2^k} g(f^{(jt)}(x)) \right) \, d\sigma_r(x). \]

Using the basins of attraction and working in a manner as in the previous case, we can conclude the
result.

(iv) Pick any \( g \in C[0, 1] \). To prove this weak convergence we first make an observation concerning
Lemma 5.5. Since the limit holds, independent of the number of compositions of \( f_4 \) with itself on
the right-hand side, we have that

\[ \lim_{N'' \to \infty} \int_0^1 g(y) \, d\sigma_{4,N''}(y) = \int_0^1 g(f_4^{(k)}(y)) \, d\sigma(y) \]
for any \( k \in \{0, 1, \ldots, M\} \), where \( M \) is a positive integer and \( k = 0 \) yields the identity function \( f_4^{(0)}(y) = y \). By summing we obtain

\[
(M + 1) \cdot \lim_{N'' \to \infty} \int_0^1 g(y) \, d\sigma_4,N''(y) = \sum_{k=0}^{M} \int_0^1 g(f_4^{(k)}(y)) \, d\sigma(y)
\]

(5.11)

\[
\lim_{N'' \to \infty} \int_0^1 g(y) \, d\sigma_4,N''(y) = \frac{1}{M + 1} \sum_{k=0}^{M} \int_0^1 g(f_4^{(k)}(y)) \, d\sigma(y)
\]

The sum obtained above is true for any \( M \) positive which means it will still be true as \( M \to \infty \). At this point, notice that the distance

\[
\left| \int_0^1 g(y) \, d\sigma_4,N''(y) - \int_0^1 g \, d\sigma^* \right|
\]

can be bounded in the following manner:

\[
\left| \int_0^1 g(y) \, d\sigma_4,N''(y) - \int_0^1 g \, d\sigma^* \right| \leq \left| \int_0^1 g(y) \, d\sigma_4,N''(y) \right| - \left| \int_0^1 \frac{1}{M + 1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) \, d\sigma(y) \right| + \left| \int_0^1 \frac{1}{M + 1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) \, d\sigma(y) - \int_0^1 g(y) \, d\sigma^*(y) \right|
\]

Therefore, to prove our main result, we require that the two terms above can be made arbitrarily small. This is true for the first term by equation (5.11). However, for the second term we need to do some more work. By applying Birkhoff’s pointwise ergodic theorem and by equation (B.5) and Remark B.8, we have that for \( m \)-almost every \( x \in [0, 1] \)

\[
\lim_{M \to \infty} \frac{1}{M + 1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) = \int_0^1 g \, d\sigma^*,
\]

where \( \sigma^* \) is the invariant measure for the map \( f_4 \). The Bounded Convergence Theorem could be applied in the second term to show that it can be made small if the measure we were integrating with respect to was the Lebesgue measure \( m \) or, in general, a measure that is absolutely continuous with respect to \( m \). The latter is what we will show for \( \sigma \). Specifically, we will show that for an open
interval $Q \subset [0, 1] = I$ there exists a constant $D$ such that

$$\sigma(Q) \leq D m(Q).$$

Pick $\delta > 0$ small enough and take an open interval $Q \subset [f_4(\delta), f_4(1/2 - \delta)]$. Clearly, we have

$$\sigma_{4,N}(Q) = \int_Q \frac{1}{\sqrt{N}b_N(x)} \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} dz \right) dx,$$

where

$$Q_N(x) = \sqrt{N}(Q - f_4(x)) = \left\{ \sqrt{N}(y - f_4(x)) : y \in Q \right\}.$$

Now, let $U_\rho(A)$ be the open $\rho$-neighborhood of a set $A$ for some $\rho > 0$, meaning

$$U_\rho(A) = \{ x : \text{dist}(x, A) < \rho \}.$$

Using this we can write the right-hand side of (5.12) for $N$ large enough as follows:

$$\sigma_{4,N}(Q) = \int_{I \setminus U_{N^{-\frac{1}{2}}(f_4^{-1}(Q))}} h_N(x) \frac{1}{\sqrt{N}b_N(x)} \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} dz \right) dx$$

$$+ \int_{U_{N^{-\frac{1}{2}}(f_4^{-1}(Q))}} h_N(x) \frac{1}{\sqrt{N}b_N(x)} \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} dz \right) dx$$

$$= I_1 + I_2.$$

Let’s take a look at the two terms separately.
For the term $I_1$ we have that $x \in I \setminus U_{N^{-\frac{1}{4}}}(f_4^{-1}(Q))$. First, notice that

$$f_4^{-1}(Q) \cap \left([0, \delta] \cup [1/2 - \delta, 1/2 + \delta] \cup [1 - \delta, 1]\right) = \emptyset$$

since $Q \cap \left([0, f_4(\delta)] \cup [f_4(1/2 - \delta), 1]\right) = \emptyset$

and $f_4^{-1}\left([0, f_4(\delta)] \cup [f_4(1/2 - \delta), 1]\right) = [0, \delta] \cup [1/2 - \delta, 1/2 + \delta] \cup [1 - \delta, 1]$.

Therefore, it is clear that

$$U_{N^{-\frac{1}{4}}}(f_4^{-1}(Q)) \cap \left([0, \delta - N^{-\frac{1}{4}}] \cup [1/2 - \delta + N^{-\frac{1}{4}}, 1/2 + \delta - N^{-\frac{1}{4}}] \cup [1 - \delta + N^{-\frac{1}{4}}, 1]\right) = \emptyset.$$

We take $N$ large enough so that $N^{-\frac{1}{4}} < \delta$. We will consider two cases for $I_1$, one where

$x \in \left(1/2 - \delta + N^{-\frac{1}{4}}, 1/2 + \delta - N^{-\frac{1}{4}}\right) = U_{\delta - N^{-\frac{1}{4}}}({1/2})$ and the case where $x \notin U_{\delta - N^{-\frac{1}{4}}}({1/2})$.

If $x \in U_{\delta - N^{-\frac{1}{4}}}({1/2})$ then $f_4(x) \in \left(f_4(1/2 - \delta + N^{-\frac{1}{4}}), 1\right)$ and since $y \in Q$, we have

$$|y - f_4(x)| \geq |f_4(1/2 - \delta) - f_4(1/2 - \delta + N^{-\frac{1}{4}})|$$

$$= f_4(1/2 - \delta + N^{-\frac{1}{4}}) - f_4(1/2 - \delta)$$

$$= 1 - 4(\delta - N^{-\frac{1}{4}})^2 - 1 + 4\delta^2$$

$$= 8\delta N^{-\frac{1}{4}} - N^{-\frac{1}{2}},$$

which then implies

$$\frac{N}{2} (y - f_4(x))^2 \geq \frac{N}{2} (8\delta N^{-\frac{1}{4}} - N^{-\frac{1}{2}})^2 = \frac{\sqrt{N}}{2} (8\delta - N^{-\frac{1}{4}})^2. \quad (5.13)$$

If $x \notin U_{\delta - N^{-\frac{1}{4}}}({1/2})$ and $y \in Q$, then by the Mean Value Theorem we have

$$|y - f_4(x)| \geq \min_{z \in I \setminus U_{\delta - N^{-\frac{1}{4}}}({1/2})} |f_4'(z)| \cdot |f_4^{-1}(y) - x|$$
\[ \geq f_4'(1/2 - \delta + N^{-\frac{1}{4}}) \cdot N^{-\frac{1}{4}} \]

\[ = [4 - 8(1/2 - \delta + N^{-\frac{1}{4}})]N^{-\frac{1}{4}} \]

\[ = (8\delta - 8N^{-\frac{1}{4}})N^{-\frac{1}{4}}. \]

Thus, we have

\[ \frac{N}{2} (y - f_4(x))^2 \geq \frac{\sqrt{N}}{2} (8\delta - 8N^{-\frac{1}{4}})^2. \quad (5.14) \]

Now, by Proposition 3.2 observe that the term \( I_1 \) can be bounded in the following way:

\[ I_1 \leq \gamma \int_{I \cup U_{\delta - N^{-\frac{1}{4}}}} h_N(x) \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} \, dz \right) \, dx \leq \]

\[ = \gamma \int_{I \cup U_{\delta - N^{-\frac{1}{4}}}} h_N(x) \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} \, dz \right) \, dx + \gamma \int_{U_{\delta - N^{-\frac{1}{4}}}} h_N(x) \left( \int_{Q_N(x)} \exp \left\{ -\frac{z^2}{2} \right\} \, dz \right) \, dx \]

and thus by equations (5.13) and (5.14), we obtain

\[ I_1 \leq \gamma e^{-\frac{\sqrt{N}}{2} (8\delta - 8N^{-\frac{1}{4}})^2} \int_{I \cup U_{\delta - N^{-\frac{1}{4}}}} h_N(x) \left( \int_{Q_N(x)} \, dz \right) \, dx \]

\[ + \gamma e^{-\frac{\sqrt{N}}{2} (8\delta - 8N^{-\frac{1}{4}})^2} \int_{U_{\delta - N^{-\frac{1}{4}}}} h_N(x) \left( \int_{Q_N(x)} \, dz \right) \, dx \quad (5.15) \]

\[ \leq 2\gamma \sqrt{N} e^{-32\sqrt{N} (\delta - N^{-\frac{1}{4}})^2} m(Q). \]

Now we investigate the term \( I_2 \). Suppose that \( Q = (a, b) \). Then we have

\[ f_4^{-1}(Q) = \left( f^{-1}_4(a), f^{-1}_4(b) \right) \cup \left( f^{-1}_4(b), f^{-1}_4(a) \right) \]
and thus for any $c, \eta > 0$ we have

$$U_{cN^{-\eta}}(f_4^{-1}(Q)) = \left( f_4^{-1}(a) - cN^{-\eta}, f_4^{-1}(b) + cN^{-\eta} \right) \cup \left( f_4^{-1}(b) - cN^{-\eta}, f_4^{-1}(a) + cN^{-\eta} \right).$$

By taking the image of the closure of the above set, we obtain

$$f_4 \left[ U_{cN^{-\eta}}(f_4^{-1}(Q)) \right] = \left[ f_4(f_4^{-1}(a) - cN^{-\eta}), f_4(f_4^{-1}(b) + cN^{-\eta}) \right] \subset U_{4cN^{-\eta}}(Q) \quad (5.16)$$

since

$$f_4(f_4^{-1}(a) - cN^{-\eta}) = 4(f_4^{-1}(a) - cN^{-\eta})(1 - f_4^{-1}(a) + cN^{-\eta})$$

$$= f_4(f_4^{-1}(a)) + 8f_4^{-1}(a)cN^{-\eta} - 4cN^{-\eta}(1 + cN^{-\eta})$$

$$> a - 4cN^{-\eta}$$

and

$$f_4(f_4^{-1}(b) + cN^{-\eta}) = 4(f_4^{-1}(b) + cN^{-\eta})(1 - f_4^{-1}(b) - cN^{-\eta})$$

$$= f_4(f_4^{-1}(b)) - 8f_4^{-1}(b)cN^{-\eta} + 4cN^{-\eta}(1 - cN^{-\eta})$$

$$< b + 4cN^{-\eta}.$$

This means that for any $x \in U_{cN^{-\eta}}(f_4^{-1}(Q))$, $f_4(x)$ is strictly within the set $U_{4cN^{-\eta}}(Q)$. Furthermore, notice that the following is true for large $N$:

If $x \notin U_{cN^{-\eta}}(f_4^{-1}(Q))$, then $f_4(x) \notin U_{4\delta cN^{-\eta}}(Q)$, \quad (5.17)

which is shown by the calculation below.

$$f_4(f_4^{-1}(a) - cN^{-\eta}) = f_4(f_4^{-1}(a)) + 8f_4^{-1}(a)cN^{-\eta} - 4cN^{-\eta}(1 + cN^{-\eta})$$

$$= a + 8 \left( \frac{1}{2} - \frac{1}{2} \sqrt{1 - a} \right) cN^{-\eta} - 4cN^{-\eta}(1 + cN^{-\eta})$$

$$= a - 4cN^{-\eta}(\sqrt{1 - a} + cN^{-\eta})$$
\[ \leq a - 4cN^{-\eta}\sqrt{1 - f_4(1/2 - \delta)} \]
\[ = a - 8\delta cN^{-\eta}, \]

and \[ f_4(f_{-1}^{-1}(b) + cN^{-\eta}) = f_4(f_{-1}^{-1}(b)) - 8f_{-1}^{-1}(b)cN^{-\eta} + 4cN^{-\eta}(1 - cN^{-\eta}) \]
\[ = b - 8 \left( \frac{1}{2} - \frac{1}{2}\sqrt{1 - b} \right) cN^{-\eta} + 4cN^{-\eta}(1 - cN^{-\eta}) \]
\[ = b + 4cN^{-\eta}\sqrt{1 - b} - 4cN^{-2\eta} \]
\[ = b + 4cN^{-\eta}(\sqrt{1 - b} - cN^{-\eta}) \]
\[ \geq b + 2cN^{-\eta}\sqrt{1 - f_4(1/2 - \delta)} \]
\[ = b + 4\delta cN^{-\eta}, \]

where the two inequalities are justified by the fact that \( a < b \leq f_4(1/2 - \delta) = 1 - 4\delta^2 \) and \( N \) is taken large enough so that

\[ \sqrt{1 - b} - cN^{-\eta} \geq \sqrt{1 - b}/2. \]

Now, in order for us to bound the term \( I_2 \), we will utilize the fact that the interval \( Q \) is totally bounded. This technique has also been utilized in [12] by Katok and Kifer. Given \( N \) large enough there exist \( k_N \) points \( q_1, q_2, \ldots, q_{k_N} \in Q \) such that

\[ Q \subset \bigcup_{i=1}^{k_N} U_{N^{-1/2}}(\{q_i\}) \quad \text{and} \quad m(Q) > \frac{1}{2} \sum_{i=1}^{k_N} m(U_{N^{-1/2}}(\{q_i\})) = N^{-1/2}k_N. \quad (5.18) \]

The term \( I_2 \) is then bounded as follows.
\[I_2 = \int_{U_{N^{-1/2} (f_4^{-1}(Q))}} h_N(x) \left( \int_{Q} \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} dy \right) dx\]

\[\leq \sum_{i=1}^{k_N} \int_{U_{N^{-1/2} (f_4^{-1}(Q))}} h_N(x) \left( \int_{U_{N^{-1/2} (\{q_i\})}} \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} dy \right) dx \] (5.19)

\[= \sum_{i=1}^{k_N} I_{2,i}\]

We make the following claim for each term \(I_{2,i}\). For \(\varepsilon > 0\) small and large \(n\), we have

\[I_{2,i} \leq 2\gamma N^{-1/2} + 2\gamma \sum_{k=1}^{n} (2\delta)^{k-1} N^{(1-2^{-k+1})} \exp \left\{ -\frac{N^{(2-2^{-k+1})} \varepsilon}{2} (4\delta^{k+1} - \delta^{k-1} N^{-\varepsilon/2^k})^2 \right\}. \] (5.20)

Proof of claim:

Pick \(\varepsilon > 0\). We first prove by induction that for any positive integer \(n\) we have

\[I_{2,i} \leq \int_{A(n, \varepsilon, i)} \int_{A(n-1, \varepsilon, i)} \cdots \int_{A(0, \varepsilon, i)} h_N(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) dy \, dx_1 \cdots dx_{n+1} + 2\gamma \sum_{k=1}^{n} (2\delta)^{k-1} N^{(1-2^{-k+1})} \exp \left\{ -\frac{N^{(2-2^{-k+1})} \varepsilon}{2} (4\delta^{k+1} - \delta^{k-1} N^{-\varepsilon/2^k})^2 \right\}\]

where

\[A(k, \varepsilon, i) = U_{\delta^{k} N^{-1/2} + \sum_{j=1}^{k} \varepsilon/2^j} (f_4^{-k}(\{q_i\})) \text{ and we set } A(0, \varepsilon, i) = U_{N^{-1/2} (\{q_i\})},\]

and we have the density

\[q_N(x, y) = \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} \].

For \(n = 1\) we have

\[I_{2,i} = \int_{U_{N^{-1/2} (f_4^{-1}(Q))}} h_N(x) \left( \int_{U_{N^{-1/2} (\{q_i\})}} \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} dy \right) dx\]

68
\[
= \int_{U_{N-1/2}^{-1}(f^{-1}(Q)) \setminus A(1, \varepsilon, i)} h_N(x) \left( \int_{U_{N-1/2}^{-1}(\{q_i\})} \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} dy \right) dx \\
+ \int_{A(1, \varepsilon, i)} h_N(x) \left( \int_{U_{N-1/2}^{-1}(\{q_i\})} \frac{1}{b_N(x)} \exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} dy \right) dx
\]

For the first integral in the sum where \( x \notin A(1, \varepsilon, i) \), by (5.17) we have

\[
f_4(x) \notin U_{4\delta^2 N^{-1/2+\varepsilon/2}}(\{q_i\}).
\]

This implies

\[
|y - f_4(x)| \geq 4\delta^2 N^{-1/2+\varepsilon/2} - N^{-1/2} = N^{-1/2+\varepsilon/2} (4\delta^2 - N^{-\varepsilon/2})
\]

and thus

\[
\exp \left\{ -\frac{N}{2} (y - f_4(x))^2 \right\} \leq \exp \left\{ -\frac{N}{2} N^{-1+\varepsilon} (4\delta^2 - N^{-\varepsilon/2})^2 \right\} = \exp \left\{ -\frac{N^\varepsilon}{2} (4\delta^2 - N^{-\varepsilon/2})^2 \right\}.
\]

Finally, by applying equation (2.7) to the second integral we have

\[
I_{2,i} \leq \int_I \int_{A(1, \varepsilon, i)} \int_{U_{N-1/2}^{-1}(\{q_i\})} h_N(z) q_N(z, x) q_N(x, y) dy \, dx \, dz + 2\gamma \exp \left\{ -\frac{N^\varepsilon}{2} (4\delta^2 - N^{-\varepsilon/2})^2 \right\},
\]

which concludes the base case. We move on to the induction step. Assume it is true for \( n \), we will show it holds for \( n + 1 \). From above, we write the \((n + 2)\)-tuple integral as

\[
\int_I \int_{A(n, \varepsilon, i)} \int_{A(n, \varepsilon, i)} \cdots \int_{A(0, \varepsilon, i)} h_N(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[
= \int_{I \setminus A(n+1, \varepsilon, i)} \int_{A(n, \varepsilon, i)} \int_{A(n, \varepsilon, i)} \cdots \int_{A(0, \varepsilon, i)} h_N(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}
\]

\[
+ \int_{A(n+1, \varepsilon, i)} \int_{A(n, \varepsilon, i)} \int_{A(n, \varepsilon, i)} \cdots \int_{A(0, \varepsilon, i)} h_N(x_{n+1}) q_N(x_{n+1}, x_n) \cdots q_N(x_1, y) \, dy \, dx_1 \cdots dx_{n+1}.
\]

Again, for the first integral in the sum notice that \( x_{n+1} \notin A(n+1, \varepsilon, i) \), which implies from (5.17)

69
that
\[ f_4(x_{n+1}) \notin U_{\frac{4\delta_{n+2}N^{-1/2+\sum_{j=1}^{n+1} \varepsilon/2^j}}{4\delta_{n+2}N^{-1/2+\sum_{j=1}^{n+1} \varepsilon/2^j}}} (f_4^{-n}(\{q_i\})). \]

As with the base case, we have
\[
\exp\left\{ -\frac{N}{2} (x_n - f_4(x_{n+1}))^2 \right\} \leq \exp\left\{ -\frac{N}{2} N^{-1+\sum_{j=1}^{n+1} \varepsilon/2^{j-1}} (4\delta_{n+2} - \delta^n N^{-\varepsilon/2^{(n+1)}})^2 \right\}
\]
\[
= \exp\left\{ -\frac{N(2-2^{-n})\varepsilon}{2} (4\delta_{n+2} - \delta^n N^{-\varepsilon/2^{(n+1)}})^2 \right\}.
\]

From the previous inequality and by applying equation (2.7) to the second integral we have
\[
I_{2,i} \leq \int_I A(n+1,\varepsilon,i) \int_A(n,\varepsilon,i) \int_A(n-1,\varepsilon,i) \cdots \int_A(0,\varepsilon,i) h_N(x_{n+2})q_N(x_{n+2},x_{n+1}) \cdots q_N(x_1,y) \, dy \, dx_1 \cdots dx_{n+2}
\]
\[
+ \gamma \sqrt{N} \exp\left\{ -\frac{N(2-2^{-n})\varepsilon}{2} (4\delta_{n+2} - \delta^n N^{-\varepsilon/2^{(n+1)}})^2 \right\} m(A(n,\varepsilon,i))
\]
\[
+ 2\gamma \sum_{k=1}^{n} (2\delta)^{k-1} N^{(1-2^{-k+1})} \varepsilon \exp\left\{ -\frac{N(2-2^{-k+1})\varepsilon}{2} (4\delta^k - \delta^{k-1} N^{-\varepsilon/2^k})^2 \right\}.
\]

Since \( m(A(n,\varepsilon,i)) = 2^n 2\delta^n N^{-1/2+\sum_{j=1}^{n} \varepsilon/2^j} \), this concludes the induction step.

Now, to prove (5.20), notice that the \((n+2)\)-tuple integral is a \((n+2)\)-step transition probability and can be bounded as shown below.
\[
\int_I A(n,\varepsilon,i) \int_A(n-1,\varepsilon,i) \cdots \int_A(0,\varepsilon,i) h_N(x_{n+1})q_N(x_{n+1},x_n) \cdots q_N(x_1,y) \, dy \, dx_1 \cdots dx_{n+1}
\]
\[
\leq \int_I A(n,\varepsilon,i) h_N(x_{n+1})q_N(x_{n+1},x_n) \, dx_n \, dx_{n+1}
\]
\[
\leq \gamma \sqrt{N} m(A(n,\varepsilon,i))
\]
\[
= \gamma \sqrt{N} 2^{n+1} \delta^n N^{-1/2+(1-2^{-n})\varepsilon}
\]
\[
= 2\gamma (2\delta)^n N^{(1-2^{-n})\varepsilon}
\]
Since \(2\delta\) can be made arbitrarily small, we can take \(n\) large enough so that \((2\delta)^n N^{(1-2^{-n})\varepsilon} \leq N^{-1/2}\). As with case (ii), we consider \(n = n(N)\) such that
\[
 n = \left\lceil \log_{2\delta} \left( N^{-1/2-\varepsilon} \right) \right\rceil + 1.
\]
This finally proves the claim.

Therefore, by putting together equations (5.15), (5.18), (5.19) and (5.20), we obtain
\[
\sigma_{4,N}(Q) \leq \left( \sqrt{N} \sum_{k=1}^{n} (2\delta)^{k-1} N^{(1-2^{-k+1})\varepsilon} \exp \left\{ -\frac{N^{(2-2^{-k+1})\varepsilon}}{2} (4\delta^{k+1} - \delta^{k-1} N^{-\varepsilon/2^k})^2 \right\} \right. 
+ \sqrt{N} e^{-32\sqrt{N}(\delta-N^{-\frac{1}{2}})^2 + 1} \left. \right) 2\gamma m(Q). \tag{5.21}
\]
This immediately yields the next result for the open interval \(Q \subset [f_4(\delta), f_4(1/2 - \delta)]\),
\[
\sigma(Q) \leq \lim inf \sigma_{4,N}(Q) = 4\gamma m(Q). \tag{5.22}
\]
The first term in the sum of (5.21) can be shown to be bounded in a similar manner as with the term in (5.8) of case (ii) by using (5.10) and applying Jensen’s inequality with the convex function \(\phi(x) = |x|^{[1/(2\delta)]^n}\).

Next we consider the case where \(Q = (f_4(1/2 - \delta), 1]\). This is resolved by taking \(\sigma_{4,N}(Q)\) and applying (2.7) twice. By doing this, the result will be given by falling back to the previous case.
\[
\sigma_{4,N}(Q) = \int_Q h_N(y) \ dy 
= \int_I h_N(x) \left( \int_Q q_N(x,y) \ dy \right) \ dx 
= \int_{I \setminus U_{N-1/4}(f_4^{-1}(Q))} h_N(x) \left( \int_Q q_N(x,y) \ dy \right) \ dx + \int_{U_{N-1/4}(f_4^{-1}(Q))} h_N(x) \left( \int_Q q_N(x,y) \ dy \right) \ dx
\]
71
Observe that the preimage of $Q$ is the open interval $f_{\frac{1}{4}}^{-1}(Q) = (f_{\frac{1}{4}}^{-1}(f_{\frac{1}{4}}(1/2 - \delta)), f_{\frac{1}{4}}^{-1}(f_{\frac{1}{4}}(1/2 - \delta))) = (1/2 - \delta, 1/2 + \delta)$. Thus, if

$$x \notin U_{N-\frac{1}{4}}(f_{\frac{1}{4}}^{-1}(Q)) = U_{\delta + N-\frac{1}{4}}(\{1/2\}),$$

then

$$|y - f_{\frac{1}{4}}(x)| \geq f_{\frac{1}{4}}(1/2 - \delta) - f_{\frac{1}{4}}(1/2 - \delta - N^{-1/4})$$

$$= f_{\frac{1}{4}}(1/2 - \delta) - [f_{\frac{1}{4}}(1/2 - \delta) - 4N^{-1/4}(\sqrt{1 - f_{\frac{1}{4}}(1/2 - \delta)} + N^{-1/4})]$$

$$= 4N^{-1/4}(\sqrt{4\delta^2} + N^{-1/4})$$

$$\geq 8\delta N^{-1/4}.$$  

Thus, for the first integral in the sum above we have

$$\int_{I \setminus U_{N-\frac{1}{4}}(f_{\frac{1}{4}}^{-1}(Q))} h_{N}(x) \left( \int_{Q} q_{N}(x, y) \, dy \right) \, dx \leq \gamma \sqrt{N} m(Q) \exp\left\{ -N^2 (8\delta N^{-1/4})^2 \right\}$$

$$\leq \gamma \sqrt{N} (1 - f_{\frac{1}{4}}(1/2 - \delta)) \exp\left\{ -N \frac{64\delta^2 N^{-1/2}}{2} \right\}$$

$$= 4\delta^2 \gamma \sqrt{N} \exp\left\{ -32\delta^2 \sqrt{N} \right\}.$$  

Now, the second integral is written as follows

$$\int_{U_{N-\frac{1}{4}}(f_{\frac{1}{4}}^{-1}(Q))} h_{N}(x) \left( \int_{Q} q_{N}(x, y) \, dy \right) \, dx$$

$$= \int_{I} h_{N}(z) \left( \int_{U_{N-\frac{1}{4}}(f_{\frac{1}{4}}^{-1}(Q))} q_{N}(z, x) \left( \int_{Q} q_{N}(x, y) \, dy \right) \, dx \right) \, dz$$

$$\leq \int_{I} h_{N}(z) \left( \int_{U_{N-\frac{1}{4}}(f_{\frac{1}{4}}^{-1}(Q))} q_{N}(z, x) \, dx \right) \, dz.$$  

Compare the last expression in the inequality above with the integral in (5.12) from the previous
case. Also, notice that
\[ U_{N^{-1/4}}(f_4^{-1}(Q)) \subset [f_4(\delta), f_4(1/2 - \delta)]. \]

Therefore, the above integral falls back to the previous case. So (5.22) implies that
\[
\sigma(Q) \leq \lim \inf \left(4\delta^2 \gamma \sqrt{N} \exp\left\{-32\delta^2 \sqrt{N}\right\} + \int_I h_N(z) \left(\int_{U_{N^{-1/4}}(f_4^{-1}(Q))} q_N(z,x) \, dx\right) \, dz\right)
\]
\[
= \lim \inf \left(4\gamma m(U_{N^{-1/4}}(f_4^{-1}(Q)))\right)
\]
\[
= \lim \inf \left(4\gamma m(f_4^{-1}(Q)) + 8\gamma N^{-1/4}\right)
\]
\[
= 4\gamma m(f_4^{-1}(Q)).
\]

In conclusion, for the case where \( Q = (f_4(1/2 - \delta), 1], \) we have
\[
\sigma(Q) \leq 4\gamma \sqrt{m(Q)} = 8\gamma \delta. \tag{5.23}
\]

We now consider the final case, \( Q = [0, f_4(\delta)). \) First, notice that
\[
f_4^{-1}(Q) = [0, \delta) \cup (1 - \delta, 1] .
\]

Secondly, for any nonnegative integer \( n \) we have
\[
f_4^{-1}\left(\frac{\delta}{2^n}\right) = \frac{1}{2} - \frac{1}{2} \sqrt{1 - \frac{\delta}{2^n}} \leq \frac{1}{2} - \frac{1}{2} \left(1 - \frac{\delta}{2^n}\right) = \frac{\delta}{2^{n+1}}
\]
and similarly
\[
f_4^{-1}\left(\frac{\delta}{2^n}\right) \geq 1 - \frac{\delta}{2^{n+1}} .
\]

The above two inequalities yield
\[
f_4^{-1}(0, 2^{-n}\delta) \subset [0, 2^{-(n+1)}\delta) \cup (1 - 2^{-(n+1)}\delta, 1], \tag{5.24}
\]

73
which will be useful when calculating the bound for $\sigma(Q)$, especially since $m\left(f_4^{-1}([0, 2^{-n}\delta])\right)$ vanishes as $n \to \infty$. Now we make the following claim: for any integer $n \geq 1$, we have

$$\sigma_{4,N}(Q) \leq \int_Q h_N(y) dy = \int_I h_N(x) \left( \int_Q q_N(x, y) dy \right) dx$$

$$= \int_{I \setminus [B_{\delta}(1) \cup (1 - B_{\delta}(1))]} h_N(x) \left( \int_Q q_N(x, y) dy \right) dx + \int_{B_{\delta}(1)} h_N(x) \left( \int_Q q_N(x, y) dy \right) dx + \int_{1 - B_{\delta}(1)} h_N(x) \left( \int_Q q_N(x, y) dy \right) dx$$

$$= I_1' + I_2 + I_3'.$$

Observe that $f_4^{-1}(Q) \subset B_{\delta}(1) \cup (1 - B_{\delta}(1))$. Thus, for the term $I_1'$ we have

$$|y - f_4(x)| \geq f_4(2\delta) - f_4(\delta)$$

$$= 8\delta(1 - 2\delta) - 4\delta(1 - \delta)$$

$$= 4\delta(1 - 3\delta)$$

$$\geq 4\delta(1 - 4\delta) > 0.$$
Therefore,

\[ I'_1 \leq \gamma \sqrt{N} \exp \left\{-\frac{N}{2} [4\delta(1 - 4\delta)]^2 \right\} m(Q) \int_{I_{\setminus [B_\delta(1)\cup(1-B_\delta(1))]} h_N(x) \, dx \]

\[ \leq \gamma \sqrt{N} f_4(\delta) \exp \left\{-\frac{N}{2} [4\delta(1 - 4\delta)]^2 \right\}, \]

which covers the third term in (5.25). By reapplying (2.7) to the terms \( I'_2 \) and \( I'_3 \), we obtain the other terms in (5.25) for \( n = 1 \).

\[ I'_2 = \int_I h_N(z) \left[ \int_{B_\delta(1)} q_N(z, x) \left( \int_{Q} q_N(x, y) \, dy \right) \, dx \right] \, dz \]

\[ \leq \int_I h_N(z) \left( \int_{B_\delta(1)} q_N(z, x) \, dx \right) \, dz, \]

and similarly for \( I'_3 \), which concludes the base case.

For the induction step, assume the claim holds for \( n \), we will show it holds for \( n + 1 \). We write the first term of (5.25) in the following manner,

\[ \int_I \int_{B_\delta(n)} h_N(x)q_N(x, y) \, dy \, dx = \int_I \int_{[B_\delta(n+1)\cup(1-B_\delta(n+1))]} h_N(x)q_N(x, y) \, dy \, dx \]

\[ + \int_{B_\delta(n+1)} \int_{B_\delta(n)} h_N(x)q_N(x, y) \, dy \, dx \]

\[ + \int_{1-B_\delta(n+1)} \int_{B_\delta(n)} h_N(x)q_N(x, y) \, dy \, dx \]

\[ = J'_1 + J'_2 + J'_3. \]

For the term \( J'_1 \), since \( x \notin B_\delta(n + 1) \cup (1 - B_\delta(n + 1)) \) and \( y \in B_\delta(n) \), we have

\[ |y - f_\delta(x)| \geq f_4(\delta 2^{-n+1}) - \delta 2^{-n+2} \]

\[ = 4\delta 2^{-n+1}(1 - \delta 2^{-n+1}) - \delta 2^{-n+2} \]

\[ = \delta 2^{-n+2} - \delta^2 2^{-2n+4} \]
\[ \delta^2 - n + 2 \geq \delta^2 - n + 2 \]

This implies

\[ J_1' \leq \gamma \sqrt{N} \exp \left\{ -\frac{N}{2} \left[ \delta^2 - n + 2 \right] ^2 \right\} m(B_\delta(n)) \int_{I \setminus \left[ B_\delta(n+1,N) \cup (1-B_\delta(n+1,N)) \right]} h_N(x) \, dx \]

\[ \leq \gamma \sqrt{N} \delta^2 - (n+1)^3 \exp \left\{ -N2^{-2(n+1)+6} \left[ \delta(1 - 4\delta) \right]^2 \right\} \]

\[ \leq \gamma \sqrt{N} \delta^2 - (n+1)^3 \exp \left\{ -N2^{-2(n+1)} \left[ \delta(1 - 4\delta) \right]^2 \right\}, \]

which is added to the third term in (5.25). Next, the first and second terms of (5.25) for \( n + 1 \) are obtained by applying (2.7) to \( J_2' \) and \( J_3' \).

\[ J_2' = \int_{I} \int_{B_\delta(n+1)} h_N(z)q_N(z,x)q_N(x,y) \, dy \, dx \, dz \]

\[ \leq \int_{I} \int_{B_\delta(n+1)} h_N(z)q_N(z,x) \, dx \, dz, \]

and similarly for \( J_3' \) which is then added to the finite sum in (5.25). This concludes the induction step and the proof of the claim. Immediately, we have

\[ \sigma_{4,N}(Q) \leq 4\gamma \sqrt{N} \delta^2 - n + \sum_{i=1}^{n} \int_{I} \int_{1-B_\delta(i)} h_N(x)q_N(x,y) \, dy \, dx \]

\[ + \gamma \sqrt{N} \left( f_4(\delta) \exp \left\{ -\frac{N}{2} \left[ 4\delta(1 - 4\delta) \right]^2 \right\} + 8\delta \sum_{i=2}^{n} 2^{-i} \exp \left\{ -N2^{-2i} \left[ \delta(1 - 4\delta) \right]^2 \right\} \right), \]

where each integral in the second term’s sum falls back to the previous case since each interval \( 1 - B_\delta(i, N) \) is of the form \(( f_4(1/2 - \delta), 1 \). Thus, taking \( n \) large enough, specifically \( n = \left\lceil \log_2(\sqrt{N}) \right\rceil + 1 \), we will have

\[ \sigma(Q) \leq 12\gamma \delta + 4\gamma \frac{\sqrt{2}}{\sqrt{2} - 1} \sqrt{\delta} \]

(5.26)
for \( Q = [0, f_4(\delta)) \), which concludes the third and final case. The sum in the third term

\[
\sum_{i=2}^{n} \sqrt{N} 2^{-i} \exp\{-N2^{-2i}[\delta(1 - 4\delta)]^2\}
\]

again can be shown to be asymptotically bounded by 1 as with the term in (5.8) using (5.10) and applying Jensen’s inequality with the convex function

\[ \phi(x) = |x|^{2n}. \]

Therefore, the inequalities (5.22), (5.23) and (5.26) yield the fact that \( \sigma \) is an absolutely continuous measure. We now return to the time average (5.11) and continue with the proof of showing that \( \sigma = \sigma^* \). We view the time average as a term in a sequence of functions \( \{g_M(y)\}_{M=0}^{\infty} \), where

\[ g_M(y) = \frac{1}{M+1} \sum_{k=0}^{M} g(f_4^{(k)}(y)), \quad y \in [0,1]. \]

Since \( \sup_{y \in [0,1]} |g_M(y)| \leq \|g\|_{\infty} \) for all \( M \geq 0 \), we apply the Bounded Convergence Theorem to the sequence on the measure space \( ([0,1], \mathcal{B}[0,1], \sigma) \) and we have

\[
\lim_{M \to \infty} \int_{0}^{1} \frac{1}{M+1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) d\sigma(y) = \int_{0}^{1} \left( \int_{0}^{1} g \, d\sigma^* \right) d\sigma(y) = \int_{0}^{1} g \, d\sigma^* .
\]

We can say this now due to the fact that \( \sigma \ll m \). Therefore, we have

\[
\left| \int_{0}^{1} g(y) \, d\sigma_{4,N''}(y) - \int_{0}^{1} g \, d\sigma^* \right| \leq \left| \int_{0}^{1} g(y) \, d\sigma_{4,N''}(y) - \int_{0}^{1} \frac{1}{M+1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) \, d\sigma(y) \right| + \left| \int_{0}^{1} \frac{1}{M+1} \sum_{k=0}^{M} g(f_4^{(k)}(y)) \, d\sigma(y) - \int_{0}^{1} g(y) \, d\sigma^*(y) \right|
\]

where by equations (5.11) and (5.27) the two terms above can be made arbitrarily small. In conclusion,

\[
\lim_{N'' \to \infty} \int_{0}^{1} g(y) \, d\sigma_{4,N''}(y) = \int_{0}^{1} g \, d\sigma^* .
\]
Thus, we have $\sigma_{4,N''}$ converging weakly to $\sigma^*$, which implies the result for the sequence of measures $\{\sigma_{4,N}\}$ since $\{\sigma_{4,N''}\}$ is a subsubsequence of an arbitrary subsequence. This concludes the proof of our main theorem.

This is where the main part of the first project ends. Supplementary material can be found in the appendices. In the next chapter we begin the discussion of the work on parallel Bayesian logspline estimators by introducing notation and hypotheses.
CHAPTER 6
PARALLEL BAYESIAN LOGSPLINE ESTIMATORS: NOTATION & HYPOTHESES

For the convenience of the reader we collect in this section all hypotheses and results relevant to our analysis and present the notation that is utilized in the second project detailed in this dissertation.

(H1) Motivated by the form of the posterior density at Neiswanger et al. [17] we consider the probability density function of the form

\[ p(\theta) \propto p^*(\theta) \quad \text{where} \quad p^*(\theta) := \prod_{m=1}^{M} p_m(\theta) \quad (6.1) \]

where we assume that \( p_m(\theta), m \in \{1, \ldots, M\} \) have compact support on the interval \([a, b]\).

(H2) For each \( m \in \{1, \ldots, M\} \) \( p_m(\theta) \) is a probability density function. We consider the estimator of \( p \) in the form

\[ \hat{p}(\theta) \propto \hat{p}^*(\theta) \quad \text{where} \quad \hat{p}^*(\theta) := \prod_{m=1}^{M} \hat{p}_m(\theta) \quad (H2-a) \]

and for each \( m \in \{1, \ldots, M\} \) \( \hat{p}_m(\theta) \) is the logspline density estimator of the probability density \( p_m(\theta) \) that has the form

\[ \hat{p}_m : \mathbb{R} \times \Omega_{nm}^m \quad \text{defined by} \quad \hat{p}_m(\theta, \omega) = f_m(\theta, \hat{y}(\theta_1^m, \ldots, \theta_{nm}^m)), \omega \in \Omega_{nm}^m \quad (H2-b) \]

We also consider the additional estimators \( \bar{p}_m \) of \( p_m \) as defined in (G.12) and

\[ \bar{p}^*(\theta) := \prod_{m=1}^{M} \bar{p}_m(\theta). \]

Here \( \theta_1^m, \theta_2^m, \ldots, \theta_{nm}^m \sim p_m(x) \) are independent identically distributed random variables and \( f_m \) is the logspline density estimate introduced in Definition (G.1) with \( N_m \) number of knots and the order of the B-splines is \( k_m \).

\[ \Omega_{nm}^m = \left\{ \omega \in \Omega : \hat{y} = \hat{y}(\theta_1^m, \ldots, \theta_{nm}^m) \in \mathbb{R}^{L_m+1} \text{ exists} \right\}. \quad (6.2) \]

where \( L_m := N_m - k_m \).
The mean integrated square error of the estimator \( \hat{p}^* \) of the product \( p^* \) is defined by

\[
\text{MISE}[N] := \text{MISE}(p^*, \hat{p}^*) = \mathbb{E} \int (\hat{p}^*(\theta; \omega) - p^*(\theta))^2 \, d\theta
\]  

(6.3)

where we use the notation \( N = (N_m)_{m=1}^N \).

We assume that the probability densities functions \( p_1, \ldots, p_M \) satisfy the following hypotheses:

**H3** The number of samples for each subset are parameterized by a governing parameter \( n \) as follows:

\[
N(n) = \{N_1(n), N_2(n), N_3(n), \ldots, N_M(n)\} : \mathbb{N} \rightarrow \mathbb{N}^M
\]

such that for all \( m \in \{1, 2, \ldots, M\} \)

\[
D_1 \leq \frac{N_m}{n} \leq D_2,
\]

\[
\lim_{n \rightarrow \infty} N_m(n) = \infty.
\]  

(6.4)

Note that \( C_1 \|N(n)\| \leq N_m(n) \leq C_2 \|N(n)\| \).

**H4** For each \( m \in \{1, \ldots, M\}, \) \( k_1 = k_2 = \cdots = k_M = k \) for some fixed \( k \) in \( \mathbb{N} \). For the number of knots for each \( m \) are parameterized by \( n \) as follows:

\[
K(n) = \{K_1(n), K_2(n), K_3(n), \ldots, K_M(n)\} : \mathbb{N} \rightarrow \mathbb{N}^M
\]  

(6.5)

where \( K_m(n) + 1 \) is the number of knots for B-splines on the interval \([a, b]\) and thus

\[
L(n) = \{L_1(n), L_2(n), L_3(n), \ldots, L_M(n)\} : \mathbb{N} \rightarrow \mathbb{N}^M \quad \text{with} \quad L_m(n) = K_m(n) - k
\]

and we require

\[
\lim_{n \rightarrow \infty} L_m(n) = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{L_m(n)}{N_m(n)^{1/2-\beta}} = 0, \quad 0 < \beta < \frac{1}{2}.
\]
(H5) For the knots $T_{K_m(n)} = (t^m_{i})_{i=0}^{K_m(n)}$, we write

$$\bar{h}_m = \max_{k-1 \leq i \leq K_m(n)-k} (t^m_{i+1} - t^m_i)$$

and

$$\underline{h}_m = \min_{k-1 \leq i \leq K_m(n)-k} (t^m_{i+1} - t^m_i). \quad (6.6)$$

(H6) For each $m \in \{1, \ldots, M\}$, $j \in \{0, \ldots, k - 1\}$ and density $p_m \in C^{j+1}([a, b])$ there exists $C_{m,s} \geq 0$ such that

$$\left| \frac{d^{j+1} \log (p_m(\theta))}{d\theta^{j+1}} \right| < C_{m,s} \quad \text{for all} \quad x. \quad (6.7)$$

(H7) Let $\| \cdot \|_2$ denote the $L^2$-norm on $[a, b]$. For $p^*$ defined as in H1, there exists $C^* \geq 0$ such that

$$\| p^* \|^2_2 = \int (p^*(\theta))^2 \, d\theta < C^*. \quad (6.8)$$

(H8) For each subset $x_m$, the B-splines are created by choosing a uniform knot sequence. Thus,

$$\bar{h}_m = \underline{h}_m = h_m, \quad \text{for} \quad m \in \{1, \ldots, M\}. \quad (6.9)$$

Let

$$h_{\min} = \min_{1 \leq m \leq M} \{h_m\} \quad \text{and} \quad h_{\max} = \max_{1 \leq m \leq M} \{h_m\}. \quad (6.10)$$

We assume that $h_{\min}, h_{\max}$ scale in a similar way to the number of samples, i.e

$$c_1 \| N(n) \|^{-\beta} \leq h^{j+1}_{\min}(n) \leq h^{j+1}_{\max}(n) \leq c_2 \| N(n) \|^{-\beta},$$

where $j \in \{0, \ldots, k - 1\}$ is the same as in hypothesis (H6).
7.1 Error analysis for unnormalized estimator

Suppose we are given a data set \( x \) and it is partitioned into \( M \geq 1 \) disjoint subsets \( x_m, m \in \{1, \ldots, M\} \). We are interested in the subset posterior densities \( p_m(\theta) = p(\theta|x_m) \). For each such density we apply the analysis from before. Let \( \hat{p}_m \) and \( \bar{p}_m, m \in \{1, \ldots, M\} \) be the corresponding logspline estimators as defined in (G.11) and (G.12) respectively. By definition of \( \hat{p}_m \), that is equal to the logspline density estimate on \( \Omega_{n_m} \subset \Omega \), where \( \Omega_{n_m} \) is the set defined in (G.10) for \( \hat{p}_m \).

**Definition 7.1.** For \( m \in \{1, \ldots, M\} \), let \( \Omega_{n_m} \) be the set defined in (6.2). We then set

\[
\Omega^{M,N} := \bigcap_{m=1}^{M} \Omega_{n_m} \quad \text{where} \quad N = (n_1, \ldots, n_m)
\]

which is the set where the maximizer for the log-likelihood exists given each data subset and thus all logspline density estimators \( \hat{p}_m \) exist.

**Lemma 7.2.** Suppose the conditions in (H3) and (H4) hold. Given the previous definition, we have that

\[
\lim_{n \to \infty} \mathbb{P}\left( \Omega^{M,N(n)} \right) = 1.
\]

**Proof.** By Theorem H.13 we have that

\[
\mathbb{P}\left( \Omega \setminus \Omega^{M,N(n)} \right) = \mathbb{P}\left( \bigcup_{m=1}^{M} (\Omega_{n_m}^{m})^c \right) \leq \sum_{m=1}^{M} \mathbb{P}\left( (\Omega_{n_m}^{m})^c \right) \leq \sum_{m=1}^{M} 2e^{-N_m(n)^2(L_m(n)+1)\delta_m(D)}
\]

and the result follows by taking \( n \) to infinity.

Since the probability of the set where the estimators \( \hat{p}_m \) exist for all \( m \in \{1, \ldots, M\} \) tends to 1, it makes sense to do our analysis for a conditional MISE on the set \( \Omega^{M,N(n)} \). Considering the practical aspect, we will never encounter the set where the maximizer of the log-likelihood doesn’t exist.

At this point, let’s state a bound for \( |\hat{p}^*(\theta; \omega) - p^*(\theta)| \) which will be essential in our analysis of MISE.
Lemma 7.3. Suppose the hypotheses (H1)-(H7) hold and that we are restricted to the sample subspace $\Omega^{M,N(n)}$. We then have the following:

(a) There exists a positive constant $R_1 = R_1(M)$ such that

$$
\| \log(\hat{p}^*(\cdot, \omega)) - \log(\bar{p}^*(\cdot)) \|_\infty \leq R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}}.
$$

(b) There exists a positive constant $R_2 = R_2(M, k, j, F_p, \gamma(T_{K_1(n)}), \ldots, \gamma(T_{K_M(n)}))$ such that

$$
\| \log (p^*) - \log (\bar{p}^*) \|_\infty \leq R_2 \tilde{h}^{j+1}_{\text{max}} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \text{ where } \tilde{h}^{\text{max}} = \max_{m} \{ \tilde{h}_m \}.
$$

(c) Using the bounds from (a) and (b) we have

$$
|\hat{p}^*(\theta; \omega) - p^*(\theta)| \leq \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}^{j+1}_{\text{max}} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \right\} - 1 \right) p^*(\theta).
$$

Proof.

(a) The bound can be shown by writing

$$
\| \log(\hat{p}^*(\cdot, \omega)) - \log(\bar{p}^*(\cdot)) \|_\infty = \| \log \left( \prod_{m=1}^{M} \hat{p}_m(\cdot; \omega) \right) - \log \left( \prod_{m=1}^{M} \bar{p}_m(\cdot) \right) \|_\infty
$$

$$
= \| \sum_{m=1}^{M} \log(\hat{p}_m(\cdot; \omega)) - \sum_{m=1}^{M} \log(\bar{p}_m(\cdot)) \|_\infty
$$

$$
\leq \sum_{m=1}^{M} \| \log(\hat{p}_m(\cdot; \omega)) - \log(\bar{p}_m(\cdot)) \|_\infty
$$

and then applying Theorem H.16. For each $m \in \{1, \ldots, M\}$ there will be an $M^m_3$ appearing in the bound and we can take $R_1 = \max_{m} \{ M^m_3 \}$.

(b) Similar to part (a) we can write

$$
\| \log(p^*(\cdot)) - \log(\bar{p}^*(\cdot)) \|_\infty = \| \log \left( \prod_{m=1}^{M} p_m(\cdot) \right) - \log \left( \prod_{m=1}^{M} \bar{p}_m(\cdot) \right) \|_\infty
$$
\[ = \sum_{m=1}^{M} \log(p_m(\cdot)) - \sum_{m=1}^{M} \log(\bar{p}_m(\cdot)) \|_{\infty} \]

\[ \leq \sum_{m=1}^{M} \| \log(p_m(\cdot)) - \log(\bar{p}_m(\cdot)) \|_{\infty} \]

and then we apply Lemma H.7. For each \( m \in \{1, \ldots, M\} \) there will be constants \( M'_m \) and \( C_m(k,j) \) appearing and we can take \( R_2 = \max_{m} \{ M'_m C_m(k,j) \} \).

(c) To see why this is true, we write

\[ |\hat{p}^*(\theta; \omega) - p^*(\theta)| = p^*(\theta) \left| \frac{\hat{p}^*(\theta; \omega)}{p^*(\theta)} - 1 \right| = p^*(\theta) \left| \exp\{\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))\} - 1 \right|. \]

If \( \hat{p}^*(\theta; \omega) \geq p^*(\theta) \) then

\[ |\exp\{\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))\} - 1| = \exp\{\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))\} - 1. \]

If \( \hat{p}^*(\theta; \omega) < p^*(\theta) \) then

\[ |\exp\{\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))\} - 1| = 1 - \exp\{-[\log(p^*(\theta)) - \log(\hat{p}^*(\theta; \omega))]\} \]

\[ \leq \exp\{\log(p^*(\theta)) - \log(\hat{p}^*(\theta; \omega))\} - 1 \]

where the last step is justified by the fact that \( 1 - e^{-x} \leq e^x - 1 \), for any \( x \geq 0 \). This implies

\[ |\hat{p}^*(\theta; \omega) - p^*(\theta)| \leq p^*(\theta) (\exp\{|\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))|\} - 1) \]

\[ \leq p^*(\theta) (\exp\{|\log(\hat{p}^*(\theta; \omega)) - \log(p^*(\theta))| + |\log(p^*(\theta)) - \log(p^*(\theta))|\} - 1) \]

and then we apply the bounds from the previous two parts. \[\square\]

This leads us directly to the theorem for the conditional MISE of the unnormalized densities \( p^* \) and \( \hat{p}^* \).
Theorem 7.4. Assume the conditions (H1)-(H7) hold. Given \( M \geq 1 \) we have

\[
\text{MISE}(p^*, \hat{p}^* | \Omega^{M,N(n)}) \leq \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}_{j_{\max}}^{j+1} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \right\} - 1 \right)^2 \| p^* \|_2^2
\]

where \( R_1, R_2 \) are as in Lemma 7.3.

In addition, if (H8) holds, then MISE scales optimally in regards to the number of samples,

\[
\sqrt{\text{MISE}(p^*, \hat{p}^*)} = O(Mn^{-\beta}) = O(M^{1-\beta} \| N(n) \|^{-\beta})
\]

Proof. By definition of the conditional MISE and Lemma 7.3, we have

\[
\text{MISE}(p^*, \hat{p}^* | \Omega^{M,N(n)}) = \mathbb{E}_{\Omega^{M,N(n)}} [(\hat{p}^*(\theta; \omega) - p^*(\theta))^2]
\]

\[
\leq \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}_{j_{\max}}^{j+1} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \right\} - 1 \right) p^*(\theta) \right]^2 d\theta
\]

\[
= \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}_{j_{\max}}^{j+1} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \right\} - 1 \right)^2 \mathbb{E}_{\Omega^{M,N(n)}} [(p^*(\theta))^2] d\theta
\]

\[
= \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}_{j_{\max}}^{j+1} \sum_{m=1}^{M} \| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \|_\infty \right\} - 1 \right)^2 \| p^* \|_2^2
\]

which concludes the proof for (7.1). Next, if (H8) holds, then (7.2) follows directly.

Remark 7.5. It’s interesting to note how the number of knots, their placement and the number of samples all play a role in the above bound. If we want to be accurate, all of the parameters \( L_m(n), N_m(n) \) and \( \tilde{h}_{\max} \) must be chosen appropriately. For instance, if the knots are not placed correctly, no matter how large of a number of samples we take for each subset, the error will be substantial since the second term in the exponential will not be small.
7.2 Analysis for renormalization constant

We will now consider the error that arises for MISE when one renormalizes the product of the estimators so it can be a probability density. The renormalization can affect the error since \( p^* \) and \( \hat{p}^* \) are rescaled. We define the renormalization constant and its estimator to be

\[
\lambda = \int p^*(\theta) \, d\theta \quad \text{and} \quad \hat{\lambda} = \hat{\lambda}(\omega) = \int \hat{p}^*(\theta; \omega) \, d\theta.
\]

(7.3)

Therefore, we are interested in analyzing

\[
\text{MISE}(p, \hat{p}) = \text{MISE}(cp^*, \hat{c}\hat{p}^*), \quad \text{where} \quad c = \lambda^{-1}, \ \hat{c} = \hat{\lambda}^{-1}.
\]

We first state the following lemma for \( \lambda \) and \( \hat{\lambda}(\omega) \).

**Lemma 7.6.** Let \( \lambda \) and \( \hat{\lambda}(\omega) \) be defined as in (7.3). Suppose that (H8) holds and we are restricted to the sample subspace \( \Omega_{M,N(n)} \). Then we have

\[
\left| \frac{\hat{\lambda}(\omega)}{\lambda} - 1 \right| = O(M^{1-\beta} N(n)^{-\beta})
\]

(7.4)

**Proof.** By definition of \( \lambda \) and \( \hat{\lambda}(\omega) \), we have

\[
|\lambda - \hat{\lambda}(\omega)| = \left| \int p^*(\theta) \, d\theta - \int \hat{p}^*(\theta; \omega) \, d\theta \right|
\]

\[
\leq \int |p^*(\theta) - \hat{p}^*(\theta; \omega)| \, d\theta
\]

\[
\leq \left( \exp \left\{ R_1 \sum_{m=1}^{M} \frac{L_m(n) + 1}{\sqrt{N_m(n)}} + R_2 \tilde{h}_{j+1} \sum_{m=1}^{M} \left\| \frac{d^{j+1} \log(p_m)}{d\theta^{j+1}} \right\|_{\infty} \right\} - 1 \right) \lambda
\]

where the second inequality is justified by Lemma 7.3(c). Dividing by \( \lambda \) the result then follows by hypothesis (H8).

So what the above lemma suggests is that when restricted to the sample subspace \( \Omega_{M,N(n)} \), the space
where the logspline density estimators \( \hat{p}_m, m \in \{1, \ldots, M\} \) are all defined, the renormalization constant \( \hat{c} \) of the product of the estimators approximates the true renormalization constant \( c \).

Knowing now how \( \hat{\lambda}(\omega) \) scales we can start analyzing MISE\((p, \hat{p})\) on the sample subspace. However, to make the analysis slightly easier we introduce a new functional, called \( \overline{\text{MISE}} \). This new functional is asymptotically equivalent to MISE as we will show, thus providing us with the means to view how MISE scales without having to directly analyze it.

**Definition 7.7.** Suppose \( M \geq 1 \) and hypotheses \((\text{H1})-(\text{H2})\) hold. Given the sample subspace \( \Omega^{M,N(n)} \) we define the functional

\[
\overline{\text{MISE}}(p, \hat{p} | \Omega^{M,N(n)}) = \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \frac{\hat{\lambda}(\omega)}{\lambda} - 1 + 1 \right)^2 \int (\hat{p}(\theta; \omega) - p(\theta))^2 d\theta \right].
\]

**Proposition 7.8.** The functional \( \overline{\text{MISE}} \) is asymptotically equivalent to MISE on \( \Omega^{M,N(n)} \), in the sense that

\[
\lim_{||N(n)|| \to \infty} \frac{\text{MISE}(p, \hat{p} | \Omega^{M,N(n)})}{\overline{\text{MISE}}(p, \hat{p} | \Omega^{M,N(n)})} = 1.
\]

**Proof.** Notice that \( \overline{\text{MISE}} \) can be written as

\[
\overline{\text{MISE}}(p, \hat{p} | \Omega^{M,N(n)}) = \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \frac{\hat{\lambda}}{\lambda} - 1 \right)^2 \int (\hat{p}(\theta; \omega) - p(\theta))^2 d\theta \right] = \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \frac{\hat{\lambda}}{\lambda} - 1 \right)^2 + 2 \left( \frac{\hat{\lambda}}{\lambda} - 1 \right) + 1 \right] \int (\hat{p}(\theta; \omega) - p(\theta))^2 d\theta,
\]

and thus by Lemma 7.6

\[
\overline{\text{MISE}}(p, \hat{p} | \Omega^{M,N(n)}) = (1 + \mathcal{E}(n))\text{MISE}(p, \hat{p} | \Omega^{M,N(n)}),
\]

where \( \mathcal{E}(n) = O(M^{1-\beta}||N(n)||^{-\beta}) \)

which then implies the result.

We conclude our analysis with the next theorem, which states how MISE scales for the renormalized estimators.
**Theorem 7.9.** Let \( M \geq 1 \). Assume the conditions (H1)-(H8) hold. Then

\[
\text{MISE} \left( p, \hat{p} \mid \Omega^{M,N(n)} \right) = O(M^{2-2\beta} \| N(n) \|^{-2\beta}).
\]  

**Proof.** We will do the work for \( \text{MISE} \) and the result will follow from Proposition 7.8. Notice that \( \text{MISE} \) can be written as below. Also, let \( E_n(\cdot) = \mathbb{E}(\cdot | \Omega^{M,N(n)}) \)

\[
\text{MISE} \left( p, \hat{p} \mid \Omega^{M,N(n)} \right) = \mathbb{E}_n \left[ \left( \frac{\hat{\lambda}}{\lambda} \right)^2 \int (p - \hat{p})^2 \, d\theta \right]
\]

\[
= \mathbb{E}_n \int \left( \lambda^{-1}(\hat{\lambda} - \lambda)p - \lambda^{-1}(\hat{p}^* - p^*) \right)^2 \, d\theta
\]

\[
= ||p||^2_2 \mathbb{E}_n \left[ \left( \frac{\hat{\lambda}}{\lambda} - 1 \right)^2 \right] + \lambda^{-2} \text{MISE}_n(p^*, \hat{p}^*) - 2\lambda^{-1} \mathbb{E}_n \int \left( \frac{\hat{\lambda}}{\lambda} - 1 \right) (\hat{p}^* - p^*) p \, d\theta
\]

\[
= J_1 + J_2 + J_3.
\]

We now determine how each of the \( J_i, i \in \{1, 2, 3\} \) scale. For \( J_1 \) by Lemma 7.6 we have

\[
J_1 = O(M^{2-2\beta} \| N(n) \|^{-2\beta}),
\]

for \( J_2 \) we have from (H8)

\[
J_2 = O(M^{2-2\beta} \| N(n) \|^{-2\beta})
\]

and for \( J_3 \) we have from Lemmas 7.3(c) and 7.6

\[
|J_3|^2 \leq 4\lambda^{-2} \left( \mathbb{E}_n \int \left( \frac{\hat{\lambda}}{\lambda} - 1 \right) |\hat{p}^* - p^*| p \, d\theta \right)^2
\]

\[
\leq 4\lambda^{-2} \mathbb{E}_n \left[ \left( \frac{\hat{\lambda}}{\lambda} - 1 \right)^2 \int p^2 \, d\theta \right] \cdot \text{MISE}_n(p^*, \hat{p}^*).
\]

Thus, by hypotheses (H7)-(H8),

\[
|J_3| = O(M^{2-2\beta} \| N(n) \|^{-2\beta}).
\]
CHAPTER 8
NUMERICAL ERROR

So far we have estimated the error that arises between the unknown density \( p \) and the full-data estimator \( \hat{p} \). However, in practice it is difficult to evaluate the renormalization constant

\[
\hat{\lambda}(\omega) = \int \hat{p}^*(\theta) \, d\theta = \int \prod_{m=1}^{M} \hat{p}_m(\theta) \, d\theta
\]

defined in (7.3). The difficulty is due to the process of generating MCMC samples and thus \( \hat{p}^* \) is not explicitly known. In order to circumvent this issue, our idea is to approximate the integral above numerically. To accomplish this, we interpolate \( \hat{p}^* \) using Lagrange polynomials. This procedure leads to the construction of an interpolant estimator \( \tilde{p}^* \) which we then integrate numerically. We then normalize \( \tilde{p}^* \) and use that as a density estimator for \( p \). Unfortunately, to estimate the error by considering that kind of approximation given an arbitrary grid of points for Lagrange polynomials, independent of the set of knots \( (t_i) \) for B-splines gives a stringent condition on the smoothness of B-splines we incorporate. It turns out that we have to utilize B-splines of order at least \( k = 4 \). For this reason we consider using Lagrange polynomials of order \( l + 1 \) which satisfy \( l < k - 2 \).

8.1 Interpolation of an estimator: preliminaries

We remind the reader the model we deal with throughout our work. We recall that the (marginal) posterior of the parameter \( \theta \in \mathbb{R} \) (which is a component of a multidimensional parameter \( \theta \in \mathbb{R}^d \)) given the data

\[
x = \{x_1, x_2, \ldots, x_M\}
\]

partitioned into \( M \) disjoint sets \( x_m, m = 1, \ldots, M \) is assumed to have the form

\[
p(\theta|x) \propto \prod_{m=1}^{M} p_m(\theta)
\]  

(8.1)

with \( p(\theta|x_m) \) denoting the (marginal) posterior density of \( \theta \) given data \( x_m \).
The estimator \( \hat{p}(\theta|x) \) of the posterior \( p(\theta|x) \) is taken to be

\[
\hat{p}(\theta|x) \propto \prod_{m=1}^{M} \hat{p}_m(\theta)
\]

(8.2)

where \( \hat{p}_m(\theta) \) stands for the logspline density estimator of the sub-posterior density \( p_m(\theta) \). From Definition G.1 and hypotheses (H1)-(H5) we have that for each \( m \in \{1, \ldots, M\} \), the estimator \( \hat{p}_m \) has the form

\[
\hat{p}_m(\theta) = \exp \left( B_m(\theta; \hat{y}^m) - c(\hat{y}^m) \right)
\]

(8.3)

where

\[
B_m(\theta; \hat{y}^m) = \sum_{j=0}^{L_{m}(n)} \hat{y}_j^m B_{j,k,T}(\theta)
\]

and \( c(\hat{y}^m) = \log \left( \int \exp \left( B_m(\theta; \hat{y}^m) \right) d\theta \right) \).

The vector \( \hat{y}^m = (\hat{y}_1^m, \ldots, \hat{y}_{L_{m}(n)}^m) \) is the argument that maximizes the log-likelihood, as described in (G.6) and note that this maximizer exists for all \( m \in \{1, \ldots, M\} \) as we carry out our analysis on the sample subspace \( \Omega_{M,N(n)} \).

Together with the hypotheses stated in Chapter 6, we now add the next proposition which will be necessary for our work later on.

**Proposition 8.1.** Suppose hypotheses (H1)-(H8) hold. Given the space \( \Omega_{M,N(n)} \), we have that the estimator \( \hat{p}_m \) is bounded and its derivatives of all orders satisfy

\[
\left| \hat{p}_m^{(\alpha)}(\theta) \right| \leq C(\alpha, k, p_m)\|N(n)\|^{{\alpha\beta}/(j+1)} \text{ for } \theta \in (a, b) \text{ and } \alpha < k - 1
\]

where the constant \( C(\alpha, k, p_m) \) depends on the order \( k \) of the B-splines, the order \( \alpha \) of the derivative and the density \( p_m \).

**Proof.** Observe that the estimator \( \hat{p}_m \) can be expressed as

\[
\hat{p}_m(\theta) = \exp \left\{ \sum_{j=0}^{L_{m}(n)} \hat{y}_j^m B_{j,k}(\theta) - c(\hat{y}^m) \right\} = \exp \left\{ \sum_{j=0}^{L_{m}(n)} (\hat{y}_j^m - c(\hat{y}^m)) B_{j,k}(\theta) \right\}
\]
Then, applying Faa di Bruno’s formula, we obtain

\[ |\hat{p}_{m}^{(\alpha)}(\theta)| \leq \hat{p}_{m}(\theta) \sum_{k_1+2k_2+\ldots+\alpha k_\alpha = \alpha} \frac{\alpha!}{k_1!k_2!\ldots k_\alpha!} \prod_{i=1}^{\alpha} \left( \frac{d^i \sum_{j=0}^{L_m(n)} (\hat{y}_j^m - c(\hat{y}_j^m)) B_{j,k}(\theta)}{i!} \right)^{k_i}, \]

for \( \theta \in [t_i, t_{i+1}] \), where \( k_1, \ldots, k_\alpha \) are nonnegative integers and if \( k_i > 0 \) with \( i \geq k \) then that term in the sum above will be zero since almost everywhere \( B_{j,k}^{(i)}(\theta) = 0 \). By De Boor’s formula [3, p.132], we can estimate the derivative of a spline as follows

\[ \left| \frac{d^i}{d\theta^i} \sum_{j=0}^{L_m(n)} (\hat{y}_j^m - c(\hat{y}_j^m)) B_{j,k}(\theta) \right| = \left| \frac{d^i}{d\theta^i} \log \hat{p}_{m}(\theta) \right| \leq C \| \log \hat{p}_{m} \|_{\infty} h_m^i, \]

where the constant \( C \) depends only on the order \( k \) of the B-splines. Therefore, we can bound \( |\hat{p}_{m}^{(\alpha)}(\theta)| \) as follows

\[ |\hat{p}_{m}^{(\alpha)}(\theta)| \leq \hat{p}_{m}(\theta) \sum_{k_1+2k_2+\ldots+\alpha k_\alpha = \alpha} \frac{\alpha!}{k_1!k_2!\ldots k_\alpha!} \prod_{i=1}^{\alpha} \left( C \| \log \hat{p}_{m} \|_{\infty} \right)^{k_i} \frac{1}{h_m^i} \]

\[ \leq \hat{p}_{m}(\theta) \left( 1 + \frac{C^\alpha \| \log \hat{p}_{m} \|_{\infty}^\alpha}{h_m^\alpha} \right) \sum_{k_1+2k_2+\ldots+\alpha k_\alpha = \alpha} \frac{\alpha!}{k_1!k_2!\ldots k_\alpha!}. \]

The above leads to the following bound:

\[ |\hat{p}_{m}^{(\alpha)}(\theta)| \leq \hat{p}_{m}(\theta) \frac{1 + C^\alpha \| \log \hat{p}_{m} \|_{\infty}^\alpha}{h_m^\alpha} \sum_{\zeta=1}^{\alpha} \frac{\alpha!}{\zeta! (\alpha - \zeta + 1)^\zeta} \]

\[ \leq C(k, \alpha) \hat{p}_{m}(\theta) \frac{1 + \| \log \hat{p}_{m} \|_{\infty}^\alpha}{h_m^\alpha} \]

where \( C(k, \alpha) \) is a constant that depends on the order \( k \) and the \( \alpha \). Next, recalling the hypotheses (H3), (H4),(H6) and (H8), we obtain

\[ \hat{p}_{m}(\theta) \leq |\hat{p}_{m}(\theta) - p_{m}(\theta)| + p_{m}(\theta) \leq \| p_{m} \|_{\infty} (1 + c \| N(n) \|^{-\beta}) \]

and
\[
\| \log \hat{p}_m \|_\infty \leq \| \log \hat{p}_m - \log \bar{p}_m \|_\infty + \| \log \bar{p}_m - \log p_m \|_\infty + \| \log p_m \|_\infty \\
\leq c \| N(n) \|^{-\beta} + \| \log p_m \|_\infty
\]

where we also used Lemma 7.3, Lemma H.7 and Theorem H.16. Therefore,

\[
\left| \hat{p}_m^{(\alpha)}(\theta) \right| \leq C(k, \alpha) \| p_m \|_\infty (1 + \| N(n) \|^{-\beta}) \frac{1 + \| N(n) \|^{-\alpha \beta} + \| \log p_m \|_\infty}{L_m} \]

\[
\leq C(\alpha, k, p_m) \frac{1}{L_m} \]

\[
= C(\alpha, k, p_m) (L_m^{j+1})^{-\alpha/(j+1)} \sim C(\alpha, k, p_m) \| N(n) \|^{\alpha \beta/(j+1)}. \]

The final result follows immediately and since the index \( i \) was chosen arbitrarily and that all interior knots are simple, this concludes the proof. \( \square \)

**Remark 8.2.** Remark E.8, in Appendix E, allowed us to extend the bound for all \( \theta \in (a, b) \) in the proof above. In reality, we can also extend the bound to the closed interval \([a, b]\). Since \( a = t_0 \) and \( b = t_{K_m(n)} \) are knots with multiplicity \( k \), any B-spline that isn’t continuous at those knots will just be a polynomial that has been cut off, which means there is no blow-up. Thus, we can extend the bound by considering right-hand and left-hand limits of derivatives at \( a \) and \( b \), respectively. From this point on we consider the bound in Proposition 8.1 holds for all \( \theta \in [a, b] \).

**Lemma 8.3.** Assume hypotheses (H1)-(H8) hold. Suppose that for each \( m = 1, \ldots, M \) the sub-posterior estimator \( \hat{p}_m(\theta) \) is \( \alpha \)-times differentiable on \([a, b]\) for some positive integer \( \alpha < k - 1 \).

Then, the estimator \( \hat{p}^* \) satisfies

\[
\left| \frac{d^\alpha}{d\theta^\alpha} \hat{p}^*(\theta) \right| = \left| (\hat{p}_1 \ldots \hat{p}_M)^{(\alpha)}(\theta) \right| \leq C(\alpha, k, p_1, \ldots, p_M) \| N(n) \|^{\alpha \beta/(j+1)} M^\alpha \quad (8.4)
\]

for \( \theta \in [a, b] \), where \( C(\alpha, k, p_1, \ldots, p_M) \) depends on the order \( k \) of the B-splines, the order \( \alpha \) of the derivative and the densities \( p_1, \ldots, p_M \).
Proof. Let $\theta \in [a, b]$. By Proposition (8.1) we have

$$|\hat{p}_m^{(\alpha)}(\theta)| \leq C(\alpha, k, p_m)\|N(n)\|^{\alpha\beta/\left(j+1\right)}.$$ 

Then, using the general Leibnitz rule and employing the above inequality we obtain

$$\left| \frac{d^\alpha}{d\theta^\alpha} \hat{p}^*(\theta) \right| = \left| \left( \hat{p}_1 \ldots \hat{p}_M \right)^{(\alpha)}(\theta) \right| =$$

$$\leq \sum_{i_1 + \ldots + i_M = \alpha} \frac{\alpha!}{i_1! \ldots i_M!} |\hat{p}_1^{(i_1)} \ldots \hat{p}_M^{(i_M)}|$$

$$\leq \sum_{i_1 + \ldots + i_M = \alpha} \frac{\alpha!}{i_1! \ldots i_M!} C(i_1, k, p_1)\|N(n)\|^{i_1\beta/(j+1)} \ldots C(i_M, k, p_M)\|N(n)\|^{i_M\beta/(j+1)}$$

$$= \|N(n)\|^{\alpha\beta/\left(j+1\right)} \sum_{i_1 + \ldots + i_M = \alpha} \frac{\alpha!}{i_1! \ldots i_M!} C(i_1, k, p_1) \ldots C(i_M, k, p_M).$$

From the proof of Proposition 8.1, notice that $C(i, k, p_m) \leq C(j, k, p_m)$ for positive integers $i \leq j$. Therefore, we have

$$|\hat{p}_m^{(\alpha)}(\theta)| \leq C(\alpha, k, p_1, \ldots, p_M)\|N(n)\|^{\alpha\beta/\left(j+1\right)} \sum_{i_1 + \ldots + i_M = \alpha} \frac{\alpha!}{i_1! \ldots i_M!}$$

where $C(\alpha, k, p_1, \ldots, p_M) = C(\alpha, k, p_1) \ldots C(\alpha, k, p_M)$ and the result follows from the multinomial theorem. This concludes the proof.

8.2 Numerical approximation of the renormalization constant $\hat{c} = \hat{\lambda}^{-1}$

By Remark E.8, in Appendix E, we have that B-splines of order $k$, and therefore any splines that arise from these, will have $k - 2$ continuous derivatives on $(a, b)$. Thus, in order to utilize Lemma I.3, we must have that the order of the Lagrange polynomials be at most $k - 2$, i.e. $l \leq k - 3$. Since $l \geq 1$ this implies that the B-splines used in the construction of the logspline estimators be at least cubic. Thus, assume $k \geq 4$ and let $1 \leq l \leq k - 3$ be a positive integer that denotes the degree of
the interpolating polynomials. Let \( N \in \mathbb{N} \) be the number of sub-intervals of \([a, b]\) on each of which we will interpolate the product of estimators by the polynomial of degree \( l \). Thus each sub-interval has to be further subdivided into \( l \) intervals. Define the partition \( \mathcal{X} \) of \([a, b]\) such that

\[
\mathcal{X} = \{a = x_0 < x_1 < x_2 < \cdots < x_{Nl} = b\} \quad \text{and} \quad x_{i+1} - x_i = \frac{b - a}{Nl} = \Delta x. \tag{8.5}
\]

For each \( i = 0, \ldots, N - 1 \), recalling the formula (I.1), we define the (random) Lagrange polynomial

\[
\hat{q}_i(\theta) := \sum_{\tau=0}^{l} \hat{p}^*(x_{il+\tau}) l_{\tau,i}(\theta) \quad \text{with} \quad l_{\tau,i}(\theta) := \prod_{j \in \{0, \ldots, l\} \setminus \{\tau\}} \left( \frac{\theta - x_{il+j}}{x_{il+\tau} - x_{il+j}} \right), \tag{8.6}
\]

which is a polynomial that interpolates the estimator \( \hat{p}^*(\theta) \) on the interval \([x_{il}, x_{(i+1)l}]\). We next define an interpolant estimator \( \tilde{p}^* \) to be a random composite polynomial given by

\[
\tilde{p}^*(\theta) := \begin{cases} 
0, & \theta \in \mathbb{R} \setminus [a, b] \\
\hat{q}_i(\theta), & \theta \in [x_{il}, x_{(i+1)l}] 
\end{cases} \tag{8.7}
\]

which approximates the estimator \( \hat{p}^* \) on the whole interval \([a, b]\).

We are now ready to estimate the mean integrated squared error given by

\[
\text{MISE}(\hat{p}^*, \tilde{p}^* \mid \Omega^{M,N(n)}) = \mathbb{E} \int (\hat{p}^*(\theta) - \tilde{p}^*(\theta))^2 \, d\theta. \tag{8.8}
\]

**Lemma 8.4.** Assume that hypotheses (H1)-(H8) hold and \( \hat{p}^* \) is the estimator of \( \hat{p}^* \) as defined in (8.7) given the partition \( \mathcal{X} \) from (8.5) respectively. The following estimate holds provided \( 1 \leq l \leq k - 3 \).

\[
\text{MISE}(\hat{p}^*, \tilde{p}^* \mid \Omega^{M,N(n)}) = \mathbb{E} \int_a^b (\hat{p}^*(\theta) - \tilde{p}^*(\theta))^2 \, d\theta 
\leq \left( \frac{(\Delta x)^{l+1}}{4(l+1)} \|N(n)\|^{(l+1)/j+1} \frac{M^{l+1}}{)} \right)^2 C(l + 1, k, p_1, \ldots, p_M, (a, b)) \tag{8.9}
\]

where the constant \( C(l + 1, k, p_1, \ldots, p_M, (a, b)) \) depends on the order \( l + 1 \) of the Lagrange polynomials, the order \( k \) of the B-splines, the densities \( p_1, \ldots, p_M \) and the length of the interval \((a, b)\).

**Proof.** Let \( i \in \{0, \ldots, N - 1\} \). By Lemma I.3, Lemma 8.3, and (8.7) for any \( \theta \in [x_{il}, x_{(i+1)l}] \) we
have

\[ |\hat{p}^*(\theta) - \tilde{p}^*(\theta)| = |\hat{p}^*(\theta) - \hat{q}_i(\theta)| \]

\[ \leq \left( \sup_{\theta \in [x_i \cdots x_{i+1}]} \left| \frac{d}{d\theta} \hat{p}^*(\theta) \right| \right) \frac{(\Delta x)^{l+1}}{4(l+1)} \]

\[ \leq \frac{(\Delta x)^{l+1}}{4(l+1)} C(l + 1, k, p_1, \ldots, p_M) \|N(n)\|^{(l+1)\beta/(j+1)} M^{l+1}. \]

Thus we conclude that

\[ \mathbb{E} \int_a^b (\hat{p}^*(\theta) - \tilde{p}^*(\theta))^2 d\theta = N-1 \sum_{i=0}^{N-1} \mathbb{E} \int_{x_i}^{x_{i+1}} (\hat{p}^*(\theta) - \hat{q}_i(\theta))^2 d\theta \]

\[ \leq \left( \frac{(\Delta x)^{l+1}}{4(l+1)} \|N(n)\|^{(l+1)\beta/(j+1)} M^{l+1} \right)^2 C(l + 1, k, p_1, \ldots, p_M, (a, b)). \]

where \( C(l + 1, k, p_1, \ldots, p_M, (a, b)) = C^2(l + 1, k, p_1, \ldots, p_M)(b - a). \)

Now that we have bounded the error between \( \hat{p}^* \) and \( \tilde{p}^* \), we define the renormalization constant \( \tilde{c} \) and the density estimator \( \tilde{p} \) of \( \tilde{p} \).

\[ \frac{1}{\tilde{c}} = \tilde{\lambda} = \int_a^b \tilde{p}^*(\theta) d\theta \quad \text{and} \quad \tilde{p} := \tilde{c} \tilde{p}^* \]

Now the question is, how close is \( \tilde{\lambda} \) to \( \hat{\lambda} \). This is answered in the following lemma.

**Lemma 8.5.** Given the definitions of \( \hat{\lambda} \) and \( \tilde{\lambda} \) in (7.3) and (8.11) respectively, we have that the distance between the two renormalization constants is bounded by

\[ |\hat{\lambda} - \tilde{\lambda}| \leq \left( \frac{(\Delta x)^{l+1}}{4(l+1)} \|N(n)\|^{(l+1)\beta/(j+1)} M^{l+1} \right) \mathcal{R}(l + 1, k, p_1, \ldots, p_M, (a, b)) \]

where the constant \( \mathcal{R}(l + 1, k, p_1, \ldots, p_M, (a, b)) = C(l + 1, k, p_1, \ldots, p_M)(b - a). \)

**Proof.** We write

\[ |\hat{\lambda} - \tilde{\lambda}| \leq \int_a^b |\hat{p}^*(\theta) - \tilde{p}^*(\theta)| d\theta \]
and then we just apply the Lagrange interpolation error from Lemma I.3.

We will continue by following the same steps as in Chapter 7. The idea is to introduce a functional that will scale the same as $\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)})$.

**Definition 8.6.** Suppose $M \geq 1$ and hypotheses (H1),(H2) and (H8) hold. Given the sample subspace $\Omega^{M,N(n)}$ we define the functional

$$\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)}) = \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \frac{\tilde{\lambda}}{\lambda(\omega)} \right)^2 \int (\hat{p}(\theta; \omega) - \tilde{p}(\theta))^2 d\theta \right]. \quad (8.13)$$

**Proposition 8.7.** The functional $\text{MISE}$ is asymptotically equivalent to $\text{MISE}$ on $\Omega^{M,N(n)}$, in the sense that

$$\lim_{\Delta x \to 0} \frac{\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)})}{\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)})} = 1. \quad (8.14)$$

**Proof.** Notice that $\text{MISE}$ can be written as

$$\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)}) = \mathbb{E}_{\Omega^{M,N(n)}} \left[ \left( \frac{\tilde{\lambda}}{\lambda} - 1 + 1 \right)^2 \int (\hat{p}(\theta; \omega) - \tilde{p}(\theta))^2 d\theta \right]$$

$$= \mathbb{E}_{\Omega^{M,N(n)}} \left[ \lambda^{-2} \left( \frac{\lambda}{\tilde{\lambda}} \right)^2 \left( \tilde{\lambda} - \lambda \right)^2 + 2\lambda^{-1} \frac{\lambda}{\tilde{\lambda}} \left( \tilde{\lambda} - \lambda \right) + 1 \right] \int (\hat{p}(\theta; \omega) - \tilde{p}(\theta))^2 d\theta \right].$$

Thus, by Lemmas 7.6 and 8.5, where the former implies

$$\frac{\lambda}{\tilde{\lambda}} \leq \frac{1}{1 - C M^{1-\beta} \|N(n)\|^{-\beta}},$$

and for large enough $n$ for which $1 - C M^{1-\beta} \|N(n)\|^{-\beta} > 0$, we have

$$\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)}) = (1 + \mathcal{E}(n))\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)})$$

with $\mathcal{E}(n) = O(M^{l+1}(\Delta x)^{l+1})$. This then implies the result.

**Theorem 8.8.** Let $M \geq 1$. Assume the conditions (H1)-(H8) hold. Then

$$\text{MISE}(\hat{p}, \tilde{p} \mid \Omega^{M,N(n)}) = O \left[ \left( \|N(n)\|^{\beta/(j+1)}(\Delta x)M \right)^{(l+1)} \right]. \quad (8.15)$$
**Proof.** We will do the work for \textbf{MISE} and the result will follow from Proposition 8.7. Notice that \textbf{MISE} can be written as below. Also, let \( \mathbb{E}_n(\cdot) = \mathbb{E}(\cdot | \Omega M, N(n)) \)

\[
\text{MISE} \left( \hat{p}, \tilde{p} \mid \Omega M \right) = \mathbb{E}_n \left[ \left( \frac{\tilde{\lambda}}{\lambda} \right)^2 \int (\hat{p}(\theta; \omega) - \tilde{p}(\theta))^2 \, d\theta \right]
\]

\[
= \mathbb{E}_n \int \left( \frac{\tilde{\lambda}}{\lambda} \hat{p} - \frac{1}{\lambda} \hat{p}^* \right)^2 \, d\theta
\]

\[
= \mathbb{E}_n \int \left( \frac{\tilde{\lambda}}{\lambda} \hat{p} - \frac{1}{\lambda} \hat{p}^* - \hat{p} + \hat{p} \right)^2 \, d\theta
\]

\[
\leq \frac{\lambda^{-1}}{1 - C M^{1-\beta} \|N(n)\|^{-\beta}} \mathbb{E}_n \int \left( (\tilde{\lambda} - \lambda)(\hat{p} - p) + (\tilde{\lambda} - \lambda)p + (\hat{p}^* - \hat{p}^*) \right)^2 \, d\theta
\]

\[
\leq \frac{\lambda^{-1}}{1 - C M^{1-\beta} \|N(n)\|^{-\beta}}(J_1 + J_2 + J_3 + J_4 + J_5 + J_6)
\]

where

\[
J_1 = \mathbb{E}_n \int (\tilde{\lambda} - \lambda)^2 (\hat{p} - p)^2 \, d\theta, \quad J_2 = \mathbb{E}_n \int (\tilde{\lambda} - \lambda)^2 p^2 \, d\theta,
\]

\[
J_3 = \mathbb{E}_n \int (\hat{p}^* - \hat{p}^*)^2 \, d\theta, \quad J_4 = 2 \mathbb{E}_n \int (\tilde{\lambda} - \lambda)^2 (\hat{p} - p)p \, d\theta,
\]

\[
J_5 = 2 \mathbb{E}_n \int (\tilde{\lambda} - \lambda)(\hat{p} - p)(\hat{p}^* - \hat{p}^*) \, d\theta, \quad J_6 = 2 \mathbb{E}_n \int (\tilde{\lambda} - \lambda)(\hat{p}^* - \hat{p}^*)p \, d\theta.
\]

and by hypotheses (H1)-(H8) and Lemmas 7.9, 8.4 and 8.5, we obtain

\[
J_1 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right. \cdot M^{2-2\beta} \|N(n)\|^{-2\beta}
\]

\[
J_2 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right)
\]

\[
J_3 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right)
\]

\[
J_4 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right) \cdot M^{1-\beta} \|N(n)\|^{\beta}
\]

\[
J_5 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right) \cdot M^{1-\beta} \|N(n)\|^{\beta}
\]

\[
J_6 = O \left( \|N(n)\|^{\beta(j+1) \Delta x} M^{2(l+1)} \right)
\]
which for large \( n \) implies the result.

**Theorem 8.9.** Assume that hypotheses (H1)-(H8) hold. Let \( \hat{p} \) be the polynomial that interpolates \( \hat{p} \) as defined in (8.7), given the partition \( \mathcal{X} \). We then have the estimate

\[
\text{MISE}(p, \hat{p} | \Omega^{M,N(n)}) = \mathbb{E} \int_a^b (p(\theta) - \hat{p}(\theta))^2 d\theta \\
\leq C \left[ M^{2-2\beta} \|N(n)\|^{-2\beta} + \left( (\Delta x) \|N(n)\|^{\beta/(j+1)} M \right)^{2(l+1)} \right] \tag{8.16}
\]

where the constant \( C \) depends on the order \( k \) of the B-splines, the degree \( l \) of the interpolating polynomial, the densities \( p_1, \ldots, p_M \) and the length of the interval \((a, b)\). Furthermore, assuming that \( \Delta x \) is a function of the number of samples \( N(n) \), then MISE scales optimally with respect to \( N(n) \) such that

\[
\text{MISE}(p, \hat{p} | \Omega^{M,N(n)}) \leq C \|N(n)\|^{-2\beta} \quad \text{when} \quad \Delta x = O \left( \|N(n)\|^{-\beta \left( \frac{1}{j+1} + \frac{1}{l+1} \right)} \right) . \tag{8.17}
\]

**Proof.** Observe that

\[
\text{MISE}(p, \hat{p} | \Omega^{M,N(n)}) = \mathbb{E} \int_a^b (p(\theta) - \hat{p}(\theta) + \hat{p}(\theta) - \tilde{p}(\theta))^2 d\theta \\
\leq \mathbb{E} \int_a^b (p(\theta) - \hat{p}(\theta))^2 d\theta + \mathbb{E} \int_a^b (\hat{p}(\theta) - \tilde{p}(\theta))^2 d\theta \\
=: I_1 + I_2.
\]

(8.16) then follows from Theorem 7.9 and Theorem 8.8. Using that estimate we can ask the following question. Suppose that we chose \( \Delta x \) to be a function of the number of samples so that

\[
c_1 \|N(n)\|^{-\alpha} \leq \Delta x(n) \leq c_2 \|N(n)\|^{-\alpha} \tag{8.18}
\]

for some constants \( c_1, c_2 \) and \( \alpha \). Clearly, one would not like \( \Delta x \) to be excessively small in order to avoid difficulties that appear with round-off error when computing. On the other hand one would like the error to converge to zero as fast as possible. Thus, let us find the smallest rate \( \alpha \) for which
the asymptotic rate achieves its maximum. To this end we define the function

$$R(\alpha) := -\lim_{\|N(n)\| \to \infty} \log\|N(n)\| \cdot \text{MISE}(p, \tilde{p} \mid \Omega^{M,N(n)})$$

that describes the asymptotic rate of convergence of the mean integrated squared error. By (8.16) we have

$$R(\alpha) = \begin{cases} 
2\beta, & \alpha \geq \beta \left(\frac{1}{l+1} + \frac{1}{j+1}\right) \\
\left(\alpha - \frac{\beta}{j+1}\right)2(l+1), & \alpha < \beta \left(\frac{1}{l+1} + \frac{1}{j+1}\right)
\end{cases}$$

It is obvious that the smallest rate for which the function $R(\alpha)$ achieves its maximum value of $2\beta$ is given by $\alpha = \beta \left(\frac{1}{l+1} + \frac{1}{j+1}\right)$. This concludes the proof. □
9.1 Numerical experiment with normal subset posterior densities

9.1.1 Description of experiment

This numerical experiment, as well as the following, is designed to investigate the relationship between the approximated value of $\text{MISE}(p, \hat{p} | \Omega^{M, N(n)})$ and the bound given by (8.17). One iteration of the experiment generates $M = 3$ subsets of a predetermined number of MCMC samples with $\hat{p}_m \sim N(2, 1), m = 1, 2, 3$. Then for each iteration the Lagrange polynomial $\tilde{p}$ is computed a hundred times by re-sampling in order to obtain an approximation to MISE and its standard deviation. For this specific example, we perform ten iterations starting with 20,000 samples and increasing that number by 10,000 for each experiment. In the experiments we ran, we chose the parameters so that the optimal rate of convergence for MISE was obtained. Thus, $\beta = 1/2$ was chosen. The logspline density estimation that was implemented utilized cubic B-splines (thus, order $k = 4$), which implies $l = 1$ in (8.17). Furthermore, we chose $j = 1$. This yields the rate $C\|N\|^{-1}$ as the upper bound for the convergence rate of MISE.

9.1.2 Numerical results

![Figure 1: The full data posterior (black line) is shown with the 3 subset posterior densities (red, blue, green) for one iteration of 110,000 samples.](image)

Figure 1: The full data posterior (black line) is shown with the 3 subset posterior densities (red, blue, green) for one iteration of 110,000 samples.
Figure 2: The full data posterior (black line) is shown with the combined subset posterior density (blue points) for one iteration of 110,000 samples.

Figure 3: The average MISE estimate is depicted for the ten experiments along with standard deviation bars (black). The red line is the upper bound of (8.17) as calculated for the different number of samples.
Figure 4: The ratio between the average MISE estimate with standard deviation error over the value 
$C\|N\|^{-1}$ is depicted for the different sample sizes.

Notice in Figure 4 how the value of the ratio seems to remain constant. This implies that the rate of 
the numerically computed error and the theoretical bound from (8.17) is similar, which is what we 
wanted.

## 9.2 Numerical experiment with gamma subset posterior densities

### 9.2.1 Description of experiment

This experiment mimics the previous one with the normally distributed generated samples, with 
the difference now that they are generated by a $Gamma(1, 1)$ and the number of samples increases 
from 40,000 to 130,000 by an increment of 10,000 for each iteration. Furthermore, $M = 5$ subsets 
are now created.
9.2.2 Numerical results

Figure 5: The full data posterior (black line) is shown with the 5 subset posterior densities (red, blue, green, purple, gray) for one iteration of 130,000 samples.

Figure 6: The full data posterior (black line) is shown with the combined subset posterior density (blue points) for one iteration of 130,000 samples.
Figure 7: The averaged MISE is depicted for the ten experiments along with standard deviation bars (black). The red line is the upper bound of (8.17) as calculated for the different number of samples.

Figure 8: The ratio between the averaged MISE with standard deviation error over the value $C\|N\|^{-1}$ is depicted for the different sample sizes.

As with the previous example, notice how in Figure 8 how the ratio between the numerically computed error and the theoretical bound from (8.17) seems to remain constant, which again is what we wanted.
CHAPTER 10
CONCLUSIONS

Now that we have finished with the main body of work, we state in this section the conclusions for both projects and reiterate the connection between the two.

For the Random Logistic Model, we expected by taking the two limits, the ergodic and the vanishing noise limit, that the deterministic behavior of the logistic map as a dynamical system would appear in the results. From Theorems 4.2 and 5.4 we see exactly that. In Theorem 4.2 the model just reduces to the iterates of the logistic map, which is formally seen by disregarding the additive noise in (2.3). The weak limits stated in Theorem (5.4) are just the invariant measures of the logistic map.

For the error estimate presented in Theorem 8.9, we see now what the choice of the parameters must be in relation to the number of samples, so that MISE scales optimally with respect to them. The figures in the previous chapter demonstrate how the bound presented in (8.17) is numerically justified.

At this point, notice below some results from the two projects.

From Theorem 4.3:

\[ P \left\{ \left| x_{r,N}(t+1) - f_r^{(t)}(x) \right| \geq \epsilon \right\} \leq \frac{2\gamma t (2r)^{t-1}}{\sqrt{N} \epsilon} \exp \left[ -\frac{N}{2} \frac{\epsilon^2}{(2r)^{2t-2}} \right], \text{ when } r \in [1,4], \]

and from Theorem 8.9:

\[ \text{MISE}(p, \tilde{p} \mid \Omega^{M,N(n)}) \leq C \left[ M^{2-2\beta} \| N(n) \|^{-2\beta} + \left( \epsilon \| N(n) \|^{3j+1} M \right)^{2(l+1)} \right]. \]

This is the depiction of the connecting thread between the two projects. The models are not related, but in the end we were interested in how certain quantities behave asymptotically and to establish rates for those quantities.
The analysis of the iteration (2.3) involves a subtle but crucial measurability question. How do we know that the quantities \( x(t, \omega) \) are measurable functions of \( \omega \in \mathbb{R}^N \)? We prove this in part (b) of Lemma A.1 by studying, in part (a) of the lemma, measurability properties of \( \xi_t(f_r(x), N)(\omega) \).

In part (c) of the lemma we prove an independence property of the Markov chain that is needed in order to calculate its transition probability function in Theorem 2.1.

Before stating the lemma, we use induction to show that for \( t \in \mathbb{N} \) the range of \( x(t) \) is a subset of \([0, 1]\). This is obviously true for \( t = 1 \). If it is true for \( t = t^* \in \mathbb{N} \), then since \( x(t^*) \in [0, 1] \) and \( f \) maps \([0, 1]\) into a subset of \([0, 1]\), \( \xi_t(f(x(t^*)), N) \) is well defined and \( N^{-1/2} \xi_t(f_r(x(t^*)), N) \in [-f_r(x(t^*)), 1 - f_r(x(t^*))] \). It follows that

\[
x(t^* + 1) \in [-f_r(x(t^*)), 1 - f_r(x(t^*))] + f_r(x(t^*)) = [0, 1].
\]

In fact, for any \( t \in \mathbb{N} \) the support of \( x(t^* + 1) \) is all of \([0, 1]\). This follows from the fact that the support of \( \xi_t(f_r(x(t)), N) \) equals the support of \( G_t \) conditioned on \( G_t \in [-\sqrt{N} f_r(x(t)), \sqrt{N}(1 - f_r(x(t))) \) which is all of \([-\sqrt{N} f_r(x(t)), \sqrt{N}(1 - f_r(x(t))) \].

This calculation makes it clear that if the Gaussian noise \( \xi_t(f_r(x), N) \) had been defined without being restricted to the interval \([-\sqrt{N} f_r(x), \sqrt{N}(1 - f_r(x)) \] then the random model would not be defined because the support of \( G_t \) is all of \( \mathbb{R} \); with positive probability \( x(t) \) would lie outside \([0, 1]\) for \( t \in \mathbb{N} \), \( t \geq 2 \), and if \( x(t) \not\in [0, 1] \), then in the next iteration \( f_r(x(t + 1)) \) would not be defined.

In the next lemma we prove that \( x(t) \) is well defined as a random variable. We also prove that \( x(t) \) and \( \xi_t(f_r(x), N) \) are independent, a property needed in order to determine the form of the transition probability function of \( x(t) \) in Theorem 2.1. Intuitively, this property follows from the following observation, proved in parts (a) and (b) of the lemma: for each \( t \in \mathbb{N} \) satisfying \( t \geq 2 \), \( \xi_t(f_r(x), N)(\omega) \) depends on \( \omega \in \mathbb{R}^N \) only through \( \omega_t \) and is a measurable function of \( \omega_t = G_t(\omega) \) while \( x(t, \omega) \) depends on \( \omega \in \mathbb{R}^N \) only through \( \omega_1, \omega_2, \ldots, \omega_{t-1} \) and is a measurable function of \((\omega_1, \omega_2, \ldots, \omega_{t-1}) = (G_1(\omega), G_2(\omega), \ldots, G_{t-1}(\omega)) \). The independence of \( x(t) \) and
\(\xi_t(f_r(x), N)\) is then a consequence of the fact that the random vector \((G_1(\omega), G_2(\omega), \ldots, G_{t-1}(\omega))\) is independent of \(G_t\).

**Lemma A.1.** Fix \(r \in (0, 4]\) and \(N \in \mathbb{N}\). For \(x \in [0, 1]\) define \(x(1) = x\) and let \(\{x(t), t \in \mathbb{N}, t \geq 2\} = \{x_{r,N}(t), t \in \mathbb{N}, t \geq 2\}\) be defined in (2.3).

(a) For all \(t \in \mathbb{N}\) the following properties hold.

   (i) \(\xi_t(f_r(x), N)(\omega)\) depends on \(\omega \in \mathbb{R}^N\) only through \(\omega_t \in \mathbb{R}\); thus \(\xi_t(f_r(x), N)(\omega) = \xi_t(f_r(x), N)(\omega_t)\).

   (ii) \(\xi_t(f_r(x), N)(\omega) = \xi_t(f_r(x), N)(\omega_t)\) is a measurable function of \((x, \omega_t) \in [0, 1] \times \mathbb{R}\).

(b) For all \(t \in \mathbb{N}\) satisfying \(t \geq 2\) the following properties hold. To ease the notation we write \(\omega[1,t-1] \) for \((\omega_1, \omega_2, \ldots, \omega_{t-1}) \in \mathbb{R}^{t-1}\) and \(\omega[1,t] \) for \((\omega_1, \omega_2, \ldots, \omega_t) \in \mathbb{R}^t\).

   (i) \(x(t, \omega)\) depends on \(\omega \in \mathbb{R}^N\) only through \(\omega[1,t-1] \in \mathbb{R}^{t-1}\). Thus

   \[x(t, \omega) = x(t, \omega[1,t-1]).\]

   (ii) \(x(t, \omega) = x(t, \omega[1,t-1])\) is a measurable function of \(\omega[1,t-1] \in \mathbb{R}^{t-1}\).

(c) For all \(t \in \mathbb{N}\) satisfying \(t \geq 2\), \(x \in [0, 1]\), and \(N \in \mathbb{N}\) the random variables \(x(t)\) and \(\xi_t(f_r(x), N)\) are independent; that is, if \(A\) and \(B\) are any Borel subsets of \(\mathbb{R}\), then with respect to the product measure \(P\) on \(\mathbb{R}^N\)

\[P(\{x(t) \in A\} \cap \{\xi_t(f_r(x), N) \in B\}) = P(x(t) \in A) \cdot P(\xi_t(f_r(x), N) \in B);\]

Since \(x(1) = x\) is a constant in \([0, 1]\), \(x(1)\) and \(\xi_t(f_r(x), N)\) are trivially independent.

**Proof.** (a) (i) By definition, for \(\omega \in \mathbb{R}^N\), \(\xi_t(f_r(x), N)(\omega)\) equals \(G_t(\omega)\) conditioned on \(G_t(\omega) \in \left[-\sqrt{N}f_r(x), \sqrt{N}(1 - f_r(x))\right]\). Since \(G_t(\omega) = \omega_t\), part (a) (i) follows.

(a) (ii) As we just showed, for \(\omega \in \mathbb{R}^N\), \(\xi_t(f_r(x), N)(\omega) = \xi_t(f_r(x), N)(\omega_t)\). For \(x \in [0, 1]\) the range of \(f_r(x)\) is \(f([0, 1]) = [0, r/4]\). We start the proof of part (a) (ii) by replacing \(f_r(x)\) in
the notation for $\xi_t(f_r(x), N)$ by $\xi$. The quantity $\xi_t(z, N)$ equals the $N(0, 1)$ random variable $G_t$ conditioned on $G_t \in A_{N,z}$, where $A_{N,z}$ denotes the closed interval $[−\sqrt{N}z, \sqrt{N}(1 − z)]$. We first show that $\xi_t(z, N)(\omega_t)$ is a measurable function of $(z, \omega_t) \in [0, r/4] \times \mathbb{R}$. Since $f$ is a continuous function on $[0, 1]$, it will then follow that $\xi_t(f_r(x), N)(\omega_t)$, as the composition of $\xi_t(z, N)(\omega_t)$ and $z = f_r(x)$, is a measurable function of $(x, \omega_t) \in [0, 1] \times \mathbb{R}$. This will complete the proof of part (a) (ii).

For each $\omega_t \in \mathbb{R}$, as a function of $z \in [0, r/4]$ the domain of $\xi_t(z, N)(\omega_t)$ is $[0, r/4]$, and, for each $z \in [0, r/4]$, as a function of $\omega_t \in \mathbb{R}$ the domain of $\xi_t(z, N)(\omega_t)$ is the closed interval $A_{N,z}$ as is the range of $\xi_t(f_r(x), N)(\omega_t)$. To prove that $\xi_t(z, N)(\omega_t)$ is a measurable function of $(z, \omega_t) \in [0, r/4] \times \mathbb{R}$, it suffices to show that for each closed interval $[a, b]$ in $\mathbb{R}$ the set

$$((\xi_t(\cdot, N)(\cdot))^{-1}([a, b]) = \{(z, \omega_t) \in [0, r/4] \times \mathbb{R} : \xi_t(z, N)(\omega_t) \in [a, b]\} \quad (A.1)$$

is a Borel subset of $[0, r/4] \times \mathbb{R}^N$. There are two cases to consider depending on whether $[a, b] \cap A_{N,z}$ is empty or nonempty for $z \in [0, r/4]$.

**Case 1.** For all $z \in [0, r/4]$, $[a, b] \cap A_{N,z} = \emptyset$. In this case $((\xi_t(\cdot, N)(\cdot))^{-1}([a, b]) = \emptyset \times \emptyset$, which is a Borel subset of $\mathbb{R} \times \mathbb{R}^N$.

**Case 2.** There exists a closed interval $[c, d] \subset [0, r/4]$ with $c \leq d$ such that $[a, b] \cap A_{N,z} \neq \emptyset$ for all $z \in [c, d]$. In this case

$$\xi_t(\cdot, N)(\cdot)^{-1}([a, b]) = \bigcup_{z \in [c, d]} \{z\} \times ([a, b] \cap A_{N,z}).$$

We now show that in case 2 $((\xi_t(\cdot, N)(\cdot))^{-1}([a, b])$ is a closed subset of $[0, r/4] \times \mathbb{R}$. Let $\{(z_n, \omega_{t,n}), n \in \mathbb{N}\}$ be a sequence in $((\xi_t(\cdot, N)(\cdot))^{-1}([a, b])$ converging to some $(z, \omega_t) \in [0, r/4] \times \mathbb{R}$. Thus $z_n \in [c, d]$ and $\omega_{t,n} \in [a, b] \cap A_{z_n,N}$, and so

$$c \leq z_n \leq d \quad \text{and} \quad -\sqrt{N}z_n \leq \omega_{t,n} \leq \sqrt{N}(1 - z_n).$$

Taking $n \to \infty$, we see that $(z, \omega_t) \in [c, d] \times A_{N,z}$. This proves that $((\xi_t(\cdot, N)(\cdot))^{-1}([a, b])$ is a closed subset of $[0, r/4] \times \mathbb{R}$ and hence is a Borel subset of $[0, r/4] \times \mathbb{R}$. This completes the proof.
of part (a) (ii).

(b) (i) We proceed by induction. Since \( x(1) = x \in [0, 1] \), we have for \( \omega \in \mathbb{R}^N \)

\[
x(2, \omega) = f(x(1, \omega)) + \frac{1}{\sqrt{N}} \xi_1(f(x(1, \omega)), \lambda) = f_r(x) + \frac{1}{\sqrt{N}} \xi_1(f_r(x), N) \omega.
\]  

(A.2)

By part (a) (i) of the lemma, \( \xi_1(f_r(x), N) \omega \) depends on \( \omega \) only through \( \omega_1 \in \mathbb{R} \). Hence \( x(2, \omega) \) depends on \( \omega \) only through \( \omega_1 \in \mathbb{R} \). We now assume that for \( t^* \in \mathbb{N} \) satisfying \( t^* \geq 2 \), \( x(t^*, \omega) \) depends on \( \omega \) only through \( \omega_{1,t^* - 1} \). Since by part (a) (i) of this lemma \( \xi_{t^*}(f_r(x), N) \omega \) depends on \( \omega \in \mathbb{R}^N \) only through \( \omega_{t^*} \), it follows that

\[
x(t^* + 1, \omega) = f(x(t^*, \omega)) + \frac{1}{\sqrt{N}} \xi_{t^*}(f(x(t^*, \omega)), N) \omega
\]

depends on \( \omega \) only through \( \omega_{1,t^*} \). This completes the proof of part (b) (i).

(b) (ii) Again we proceed by induction. As shown in parts (a) (i) and (b) (i), \( x(2, \omega) \) and \( \xi_1(f_r(x), N) \omega \) both depend on \( \omega \in \mathbb{R}^N \) only through \( \omega_1 \). Hence by (A.2)

\[
x(2, \omega_1) = f_r(x) + \frac{1}{\sqrt{N}} \xi_1(f_r(x), N) \omega_1.
\]

Since by part (a) (ii) of this lemma \( \xi_1(f_r(x), N) \omega_1 \) is a measurable function of \( \omega_1 \in \mathbb{R} \), it follows that \( x(2, \omega_1) \) is a measurable function of \( \omega_1 \in \mathbb{R} \). We now assume that for \( t^* \in \mathbb{N} \) satisfying \( t^* \geq 2 \), \( x(t^*, \omega_{1,t^* - 1}) \) is a measurable function of \( \omega_{1,t^* - 1} \in \mathbb{R}^{t^* - 1} \). By parts (a) (i) and (a) (ii) of this lemma \( \xi_{t^*}(f_r(x), N) \omega \) depends on \( \omega \) only through \( \omega_{t^*} \) and \( \xi_{t^*}(f_r(x), N) \omega_{t^*} \) is a measurable function of \( (x, \omega_{t^*}) \in [0, 1] \times \mathbb{R} \); hence it is a measurable function of \( (x, \omega_{1,t^*}) \in [0, 1] \times \mathbb{R}^{t^*} \). We now consider

\[
x(t^* + 1, \omega_{1,t^*}) = f(x(t^*, \omega_{1,t^* - 1})) + \frac{1}{\sqrt{N}} \xi_{t^*}(f(x(t^*, \omega_{1,t^* - 1})), N) \omega_{t^*}.
\]  

(A.3)

By the inductive hypothesis \( x(t^*, \omega_{1,t^* - 1}) \) is a measurable function of \( \omega_{1,t^* - 1} \in \mathbb{R}^{t^* - 1} \) and hence a measurable function of \( \omega_{1,t^*} \in \mathbb{R}^{t^*} \). By the continuity of \( f \), the quantity \( f(x(t^*, \omega_{1,t^* - 1})) \) is
also a measurable function of $\omega_{[1,t^*]} \in \mathbb{R}^{t^*}$. We now consider the second term on the right side of (A.3). The function $\xi_{t^*}(f(x(t^*,\omega_{[1,t^*-1]})),N)(\omega_{t^*})$ equals the composition of the following two functions:

- $\xi_t(f_r(x),N)(\omega_{t^*})$, which is a measurable function of $(x,\omega_{[1,t^*]}) \in [0,1] \times \mathbb{R}^{t^*}$;
- $x = x(t^*,\omega_{[1,t^*-1]})$, which is a measurable function of $\omega_{[1,t^*]} \in \mathbb{R}^{t^*}$.

It follows that $\xi_{t^*}(f(x(t^*,\omega_{[1,t^*-1]})),N)(\omega_{t^*})$ is a measurable function of $\omega_{[1,t^*]} \in \mathbb{R}^{t^*}$. Formula (A.3) now exhibits $x(t^* + 1,\omega_{[1,t^*]})$ as a measurable function of $\omega_{[1,t^*]} \in \mathbb{R}^{t^*}$. This completes the proof of part (b) (ii).

(c) For all $t \in \mathbb{N}$ satisfying $t \geq 2$ part (b) (ii) of this lemma proves that $x(t)(\omega)$ is a measurable function of $(\omega_1, \omega_2, \ldots, \omega_{t-1}) = (G_1(\omega),G_2(\omega),\ldots,G_{t-1}(\omega))$, and part (a) (ii) of this lemma proves that $\xi_t(f_r(x),N)(\omega)$ is a measurable function of $\omega_t = G_t(\omega)$. Let $A$ and $B$ be Borel subsets of $\mathbb{R}$. Since $x(t)^{-1}(A)$ is a Borel subset of $\mathbb{R}^{t-1}$ and $(\xi_t(f_r(x),N))^{-1}(B)$ is a Borel subset of $\mathbb{R}$, we have by the independence of the random vector $(G_1(\omega),G_2(\omega),\ldots,G_{t-1}(\omega))$ and the random variable $G_t$

$$P(\{x(t) \in A\} \cap \{\xi_t(f_r(x),N) \in B\})$$

$$= P(\{x(t)(G_1,G_2,\ldots,G_{t-1}) \in A\} \cap \{\xi_t(f_r(x),N)(G_t) \in B\})$$

$$= P(\{(G_1,G_2,\ldots,G_{t-1}) \in x(t)^{-1}(A)\} \cap \{G_t \in (\xi_t(f_r(x),N))^{-1}(B)\})$$

$$= P((G_1,G_2,\ldots,G_{t-1}) \in x(t)^{-1}(A)) \cdot P(G_t \in (\xi_t(f_r(x),N))^{-1}(B))$$

$$= P(x(t)(G_1,G_2,\ldots,G_{t-1}) \in A) \cdot P(\xi_t(f_r(x),N)(G_t) \in B)$$

$$= P(x(t) \in A) \cdot P(\xi_t(f_r(x),N) \in B).$$

This shows that $x(t)$ and $\xi_t(f_r(x),N)$ are independent random variables. The proof of part (c) is complete as is the proof of the lemma.

\[\square\]
APPENDIX B
IN Variant Measures AND Frobenius-Perron OPERATORS

General Theory

The following definitions and tools are essential in the study of the random logistic model. For reference we point the reader to [6], [28] and [27]. First we introduce the concept of an invariant measure of a dynamical system.

Definition B.1. Let \((\Omega, A, \mu)\) be a measure space and let \(f\) be a map from \(\Omega\) into itself. We say that \(\mu\) is an invariant measure of \(f\) if for any \(A \in A\) we have

\[
\mu(A) = \mu(f^{-1}(A)).
\]

In a sense, this means that the ”size” of a \(\mu\)-measurable set does not change under \(f\).

Invariant measures of dynamical systems have been widely studied and there are theorems that guarantee the existence of such measures for continuous maps. Invariant measures are closely related to what is known as a time average of an integrable function \(g : \Omega \to \mathbb{R}\) under \(f\). A time average is a sum given by

\[
\frac{1}{n} \sum_{k=0}^{n-1} g(f^k(x)),
\]

whose limit as \(n \to \infty\) exists under certain conditions which are stated in Birkhoff’s pointwise ergodic theorem.

Definition B.2. Let \(\mu, f\) be as in Definition B.1. We say that \(f\) is ergodic for \(\mu\) if for any measurable set \(A\) where \(f^{-1}(A) = A\) yields \(\mu(A) \in \{0, 1\}\).

So, ergodicity of a map implies that if the preimage of a set remains the same, then \(\mu\)-almost everywhere that set is either \(\emptyset\) or \(\Omega\).

Theorem B.3. Let \(\mu\) be a probability measure on \(\Omega\) which is invariant under \(f : \Omega \to \Omega\). Assume also that \(f\) is ergodic for \(\mu\). Then for any integrable function \(g : \Omega \to \mathbb{R}\) and \(\mu\)-almost all \(x \in \Omega\)
we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(f^{(k)}(x)) = \int_{\Omega} g \, d\mu.
\]

Theorem B.3 is used to obtain one of the main results.

Now, invariant measures of deterministic dynamical systems can arise from Frobenius-Perron operators. They are a subclass of Markov operators and are the centerpoint of modern ergodic theory. To motivate the definition of these operators let us first start with the following:

**Definition B.4.** Let \((\Omega, A, \mu)\) be a probability space and \(f : \Omega \to \Omega\) a measurable function. We say that \(f\) is nonsingular if for any \(A \in A\) such that \(\mu(A) = 0\) we have \(\mu(f^{-1}(A)) = 0\).

Clearly, dynamical systems are nonsingular with respect to their invariant measures. The concept of nonsingular maps with respect to a measure \(\mu\) is useful because it enables us to construct absolutely continuous measures with respect to \(\mu\) on \(\Omega\). This idea leads us to Frobenius-Perron operators. The construction is given below.

In the probability space \((\Omega, A, \mu)\), for a given \(g \in L^1(\mu)\) let’s define the measure \(\mu_g\) by

\[
\mu_g(A) = \int_{f^{-1}(A)} g \, d\mu, \quad \forall A \in A.
\]

Assume that \(f\) is nonsingular with respect to \(\mu\). That means for any \(A \in A\) with \(\mu(A) = 0\) we have \(\mu(f^{-1}(A)) = 0\), which in turn implies that \(\mu_g(A) = 0\). Thus, \(\mu_g\) is absolutely continuous with respect to \(\mu\). By the Radon-Nikodym theorem there exists a unique function \(\tilde{g} \in L^1(\mu)\) such that

\[
\mu_g(A) = \int_A \tilde{g} \, d\mu, \quad \forall A \in A.
\]

We denote the function \(\tilde{g}\) by \(Pg\). This leads us to the next definition.

**Definition B.5.** The operator \(P : L^1(\mu) \to L^1(\mu)\) defined by

\[
\int_A Pg \, d\mu = \int_{f^{-1}(A)} g \, d\mu, \quad \forall A \in A, \quad \forall g \in L^1(\mu)
\]

is called the **Frobenius-Perron operator** associated with \(f\).
The fascinating part about these operators is the correspondence between invariant measures of deterministic dynamical systems and fixed points of Frobenius-Perron operators associated with the same systems. Notice the following. Suppose \( g \in L^1(\mu) \) is a fixed point of the operator \( P \) defined in Definition B.5. We therefore have

\[
P g = g.
\]

Take any set \( A \in \mathcal{A} \). We then have

\[
\int_A g \, d\mu = \int_{f^{-1}(A)} g \, d\mu.
\]

Thus, the measure \( \mu_* \) defined by \( \mu_*(A) = \int_A g \, d\mu \) is an invariant measure of \( f \) since

\[
\mu_*(A) = \int_A g \, d\mu = \int_{f^{-1}(A)} g \, d\mu = \mu_*(f^{-1}(A)).
\]

**Application to the logistic map**

The next step is to apply the material from the previous section to the dynamical system \( f_4(x) = 4x(1 - x) \) on \([0, 1]\). A fixed point of the Frobenius-Perron operator \( P_{f_4} \) associated with \( f_4 \) and consequently an invariant measure of \( f_4 \) was discovered by Stanislaw Ulam and Jon von Neumann in 1947 [27]. Choosing \( \mu \) to be Lebesgue measure \( m \) on \([0, 1]\), Ulam and von Neumann worked in the probability space \(([0, 1], \mathcal{B}[0, 1], m)\). We give those details of the calculation that we could not find in the literature.

The first step in defining \( P_{f_4} \) is to show that \( f_4 \) is nonsingular with respect to \( m \).

**Lemma B.6.** The logistic map \( f_4(x) = 4x(1 - x) \) is nonsingular with respect to Lebesgue measure \( m \) on \([0, 1]\).

**Proof.** We prove that if \( A \in \mathcal{B}[0, 1] \) satisfies \( m(A) = 0 \), then \( m(f_4^{-1}(A)) = 0 \). The proof starts with the observation that \( f_4 \) has two inverses, \( f_4^{-1} \) on \([0, 1/2]\) and \( f_4^{-1} \) on \([1/2, 1]\), given by

\[
f_4^{-1}(x) = \frac{1}{2} - \frac{1}{2} \sqrt{1 - x} \quad \text{and} \quad f_4^{-1}(x) = \frac{1}{2} + \frac{1}{2} \sqrt{1 - x}.
\]

113
Since \( m(A) = 0 \), for any \( \epsilon > 0 \) there exists an open set \( U \subset [0, 1] \) such that

\[
A \subset U \quad \text{and} \quad m(U \setminus A) = m(U) < \epsilon.
\]

We write

\[
U = U \cap [0, 1] = U \cap (\{0\} \cup (0, 1 - \epsilon) \cup [1 - \epsilon, 1]) 
\subset \{0\} \cup (U \cap (0, 1 - \epsilon)) \cup [1 - \epsilon, 1].
\]

Since \( U \cap (0, 1 - \epsilon) \) is an open subset of \( \mathbb{R} \), \( U \cap (0, 1 - \epsilon) \) can be written as the countable disjoint union of open intervals in \( \mathbb{R} \); we write

\[
U \cap (0, 1 - \epsilon) = \bigcup_{i \in \mathbb{N}} I_i, \text{ where } I_i = (a_i, b_i) \text{ and } I_i \cap I_j = \emptyset \text{ for } i \neq j.
\]

It follows that

\[
A \subset U \subset \{0\} \cup \bigcup_{i \in \mathbb{N}} I_i \cup [1 - \epsilon, 1], \text{ where } m\left( \bigcup_{i \in \mathbb{N}} I_i \right) \leq m(U) \leq \epsilon. \tag{B.1}
\]

For each \( i \in \mathbb{N} \)

\[
f_4^{-1}(I_i) = f_4^{-1}((a_i, b_i)) = (f_4^{-1}(a_i), f_4^{-1}(b_i)) \cup (f_4^{-1}(b_i), f_4^{-1}(a_i)),
\]

and therefore

\[
m(f_4^{-1}(I_i)) = m(f_4^{-1}(a_i, b_i)) = [f_4^{-1}(b_i)) - f_4^{-1}(a_i)] + [f_4^{-1}(a_i) - f_4^{-1}(b_i)].
\]

The derivatives of \( f_4^{-1} \) and \( f_4^{+1} \) are bounded in absolute value by \( 1/[4\sqrt{\epsilon}] \) on \( [0, 1 - \epsilon] \). Since \( I_i = (a_i, b_i) \subset U \cap (0, 1 - \epsilon) \subset (0, 1 - \epsilon) \), it follows that \( b_1 < 1 - \epsilon \) and thus that

\[
f_4^{-1}(b_i) - f_4^{-1}(a_i) \leq \frac{1}{4\sqrt{\epsilon}}(b_i - a_i) \quad \text{and} \quad f_4^{+1}(a_i) - f_4^{-1}(b_i) \leq \frac{1}{4\sqrt{\epsilon}}(b_i - a_i).
\]
We conclude that

\[ m(f_4^{-1}(I_i)) = m(f_4^{-1}((a_i, b_i))) \leq \frac{1}{2\sqrt{\epsilon}} (b_i - a_i) = \frac{1}{2\sqrt{\epsilon}} m((a_i, b_i)) = \frac{1}{2\sqrt{\epsilon}} m(I_i). \quad \text{(B.2)} \]

By (B.1) \( A \subset U \subset \{0\} \cup \bigcup_{i \in \mathbb{N}} I_i \cup [1 - \epsilon, 1] \). By taking the preimage, we obtain

\[
f_4^{-1}(A) \subset f_4^{-1} \left( \{0\} \cup \bigcup_{i \in \mathbb{N}} I_i \cup [1 - \epsilon, 1] \right)
= f_4^{-1}(\{0\}) \cup f_4^{-1} \left( \bigcup_{i \in \mathbb{N}} I_i \right) \cup f_4^{-1}([1 - \epsilon, 1])
= \{0, 1\} \cup \left( \bigcup_{i \in \mathbb{N}} f_4^{-1}(I_i) \right) \cup f_4^{-1}([1 - \epsilon, 1]).
\]

We now calculate the Lebesgue measure of each term in the last line of this display. We have

\[ m(\{0, 1\}) = 0. \]

It follows from (B.1) and (B.2) that

\[
m \left( \bigcup_{i \in \mathbb{N}} f_4^{-1}(I_i) \right) = \sum_{i \in \mathbb{N}} m(f_4^{-1}(I_i)) \leq \frac{1}{2\sqrt{\epsilon}} \sum_{i \in \mathbb{N}} m(I_i)
= \frac{1}{2\sqrt{\epsilon}} m \left( \bigcup_{i \in \mathbb{N}} I_i \right) \leq \frac{1}{2\sqrt{\epsilon}} \epsilon = \frac{1}{2} \sqrt{\epsilon}.
\]

Finally

\[
m(f_4^{-1}([1 - \epsilon, 1])) = [f^{-1}(1) - f^{-1}(1 - \epsilon)] + [f_4^{-1}(1 - \epsilon) - f_4^{-1}(1)]
= \left[ \frac{1}{2} - \frac{1}{2} + \frac{1}{2} \sqrt{\epsilon} \right] + \left[ \frac{1}{2} + \frac{1}{2} \sqrt{\epsilon} - \frac{1}{2} \right]
= \sqrt{\epsilon}.
\]

These estimates yield that

\[
m(f_4^{-1}(A)) \leq m(\{0, 1\}) + m \left( \bigcup_{i \in \mathbb{N}} f_4^{-1}(I_i) \right) + m(f_4^{-1}([1 - \epsilon, 1])) = \frac{3}{2} \sqrt{\epsilon},
\]

which implies, since \( \epsilon > 0 \) is arbitrary, that \( m(f_4^{-1}(A)) = 0 \). This completes the proof that \( f_4 \) is...
nonsingular with respect to \( m \) on \([0, 1]\).

Lemma B.6 enables us to define the Frobenius-Perron operator \( P_{f_4} \) associated with \( f_4 \). This operator maps \( L^1[0, 1] \) into \( L^1[0, 1] \) and has the property that

\[
\int_A P_{f_4} g \, dm = \int_{f_4^{-1}(A)} g \, dm, \quad \forall A \in B[0, 1], \quad \forall g \in L^1[0, 1].
\]

To calculate \( P_{f_4} \) it is convenient to let \( A \) be the interval \((0, x), \ x \in [0, 1]\). The last display and the fact that

\[
f_4^{-1}((0, x)) = \left( f_4^{-1}(0), f_4^{-1}(x) \right) \cup \left( f_4^{-1}(x), f_4^{-1}(0) \right)
\]

\[
= \left( 0, \frac{1}{2} - \frac{1}{2} \sqrt{1-x} \right) \cup \left( \frac{1}{2} + \frac{1}{2} \sqrt{1-x}, 1 \right)
\]

gives the equation

\[
\int_{(0,x)} P_{f_4} g \, dm = \int_{\left(0, \frac{1}{2} - \frac{1}{2} \sqrt{1-x}\right)} g \, dm + \int_{\left(\frac{1}{2} + \frac{1}{2} \sqrt{1-x}, 1\right)} g \, dm.
\]

Taking the derivatives of both sides with respect to \( x \), we obtain the formula for \( P_{f_4} g \).

\[
[P_{f_4} g](x) = \frac{1}{4\sqrt{1-x}} \left[ g \left( \frac{1}{2} - \frac{1}{2} \sqrt{1-x} \right) + g \left( \frac{1}{2} + \frac{1}{2} \sqrt{1-x} \right) \right].
\] \( (B.3) \)

This formula is valid for any \( g \in L^1[0, 1] \).

We take a small detour by presenting the following elegant formula relating the integral operator \( P_N \) defined in (5.1) and the Frobenius-Perron operator \( P_{f_4} \).

**Theorem B.7.** For \( g \in C[0, 1] \) and \( x \in (0, 1) \)

\[
\lim_{N \to \infty} [P_N g](x) = [P_{f_4} g](x).
\]

This formula demonstrates the elegance of the theory presented in this dissertation by exhibiting the close relationship between the two operators \( P_N \) and \( P_{f_4} \). If \( g \in C[0, 1] \) is replaced by \( g \in C^n[0, 1] \) for some \( n \in \mathbb{N} \), then \( P_N g \) can be written as asymptotic expansion in powers of \( 1/\sqrt{N} \) in
which the zeroth order term is $P_{f_4}g$.

**Sketch of the Proof.** Because we do not apply the theorem in this dissertation, we only sketch the proof, which is based on Laplace’s method [4, Chapter 4]. We recall that for $y \in [0, 1]$

$$[P_N g](y) = \int_{[0,1]} g(x) \frac{1}{b_{4,N}(x)} e^{-\frac{N}{2}(y-f_4(x))^2} dx,$$

where $b_{4,N}(x)$ is the normalization that makes

$$q_{4,N}(x, y) = \frac{1}{b_{4,N}(x)} e^{\frac{N}{2}(y-f_4(x))^2}$$

a transition probability density. For fixed $y \in (0, 1)$, $q_{4,N}(, y)$ is a sequence of approximate identities concentrating at those values of $x$ for which the exponent $(y - f_4(x))^2$ attains its minimum value of 0. We make a connection with the Frobenius-Perron operator because these values of $x$ are

$$x_-(y) = f_4^{-1}(y) = \frac{1}{2} - \frac{1}{2} \sqrt{1 - x} \quad \text{and} \quad x_+(y) = f_4^{-1}(y) = \frac{1}{2} + \frac{1}{2} \sqrt{1 - x}.$$

According to Laplace’s method, for all sufficiently small $\varepsilon > 0$

$$\lim_{N \to \infty} [P_N g](y) = \frac{g(x_-(y))}{b_{4,N}(x_-(y))} \lim_{N \to \infty} \int_{(x_-(y)-\varepsilon, x_-(y)+\varepsilon)} \exp^{-\frac{N}{2}(y-f_4(x))^2} dx + \frac{g(x_+(y))}{b_{4,N}(x_+(y))} \lim_{N \to \infty} \int_{(x_+(y)-\varepsilon, x_+(y)+\varepsilon)} \exp^{-\frac{N}{2}(y-f_4(x))^2} dx.$$

Define $\varphi_y(x) = (y - f_4(x))^2$. For the purpose of calculating the $N \to \infty$ limit in the last display, Laplace’s method replaces the exponent $\varphi_y(x)$ in the first integral with the first three terms in the Taylor expansion of this function around $x_-(y)$; a similar replacement is used in the second integral. Since $\varphi_y(x_-(y)) = 0$ and $\varphi_y'(x_-(y)) = 0$, in the first integral these three terms reduce to

$$\frac{1}{2} \varphi_y''(x_-(y))(x - x_-(y))^2 = 8(1 - y)(x - x_-(y))^2.$$

In the second integral the same formula holds with $x_+(y)$ appearing in place of $x_-(y)$. 

117
Since $y \in (0, 1)$, neither $x_-(y)$ nor $x_+(y)$ equals 0, 1/2, or 1. Therefore part (c) of Proposition 3.1 shows that

$$\lim_{N \to \infty} \sqrt{N}b_{4,N}(x_-(y)) = \lim_{N \to \infty} \sqrt{N}b_{4,N}(x_+(y)) = \sqrt{2\pi}.$$

If we make the replacements in the two integrals discussed in the preceding paragraph and calculate the integrals, then in combination with the last display we find that

$$\lim_{N \to \infty} [P_N g](y) = \frac{1}{4\sqrt{1-y}}[g(x_-(y)) + g(x_+(y))]
= \frac{1}{4\sqrt{1-y}}[g \left(\frac{1}{2} - \frac{1}{2}\sqrt{1-y}\right) + g \left(\frac{1}{2} + \frac{1}{2}\sqrt{1-y}\right)]
= [P_{f_4} g](y).$$

This completes the sketch of the proof of the theorem.

We now turn to the discovery by Ulam and von Neumann [27] of an invariant probability measure for $f_4$ that has the density

$$g^*(x) = \frac{1}{\pi \sqrt{x(1-x)}}; \quad x \in [0, 1]; \quad \text{(B.4)}$$

$g^*$ is the density of a Beta distribution with parameters $\alpha = \beta = \frac{1}{2}$. Thus the probability measure

$$\sigma^*(dx) = [\pi \sqrt{x(1-x)}]^{-1} dx$$

satisfies $\sigma^*(f_4^{-1}(A)) = \sigma^*(A)$ for any Borel subset $A$ of $[0, 1]$. To verify that $g^*$ is an invariant density for $f_4$, one checks that $g^*$ is a fixed point of $P_{f_4}$; i.e., that $[P_{f_4} g^*](x) = g^*(x)$ for all $x \in [0, 1]$.

Our final task in this appendix is to outline the proof that $f_4$ is ergodic for the Ulam-von Neumann measure $\sigma^*$. This property allows us to apply the Birkhoff pointwise ergodic theorem on the probability space $([0, 1], \mathcal{B}[0, 1], \sigma^*)$. The proof of the ergodicity of $f_4$ with respect to $\mu^*$ is based on the following three steps.
1. The logistic map is topologically conjugate to the tent map

\[ T(x) = \begin{cases} 
2x, & x \in [0, 1/2] \\
2 - 2x, & x \in (1/2, 1]
\end{cases} \]

by the topological conjugacy \( h(x) = \frac{2}{\pi} \arcsin(\sqrt{x}), x \in [0, 1] \). Specifically, the conjugation is

\[ h^{-1} \circ T \circ h = f_4. \]

2. The tent map is ergodic with respect to Lebesgue measure \( m \) on \([0, 1]\)

**Proof.** Let \( A \subset [0, 1] \) be a \( m \)-measurable set such that \( T^{-1}(A) = A \). We will show that \( m(A) \in \{0, 1\} \). At this point, consider the characteristic function on \( A \) denoted by \( 1_A \). Since \( 1_A \in L^2([0, 1], m) \), it has a unique Fourier series expansion. Thus, we can write

\[ 1_A(x) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx}, \quad x \in [0, 1] \]

Also, notice that

\[ 1_A \circ T = 1_{T^{-1}(A)} = 1_A \]

Therefore, we have

\[ 1_A(x) = 1_A(T(x)) = \begin{cases} 
1_A(2x), & \text{if } x \in [0, 1/2] \\
1_A(2 - 2x), & \text{if } x \in (1/2, 1]
\end{cases} \]

Suppose \( x \in A \cap [0, 1/2] \). Then we have

\[ 1_A(x) = 1_A(2x) \]

\[ \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ik 2x} \]

\[ \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} e^{2\pi ikx} \]
By equating coefficients we have \( c_k = c_k e^{2\pi ikx}, \ k \in \mathbb{Z} \). So \( k \neq 0 \) yields
\[
  c_k (1 - e^{2\pi ikx}) = 0 \quad \text{which implies} \quad c_k = 0.
\]

Now suppose \( x \in A \cap (1/2, 1] \). Then we have
\[
  1_A(x) = 1_A(2 - 2x)
\]
\[
  \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ik(2-2x)}
\]
\[
  \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k e^{4\pi ikx} e^{-4\pi ikx}
\]
\[
  \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k \cdot 1 \cdot e^{(6+2)\pi ikx}
\]
\[
  \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx} = \sum_{k=-\infty}^{\infty} c_k e^{-6\pi ikx} e^{2\pi ikx}
\]

Again, by equating the coefficients we have \( c_k = c_k e^{-6\pi ikx}, \ k \in \mathbb{Z} \). So, \( k \neq 0 \) yields
\[
  c_k (1 - e^{-6\pi ikx}) = 0 \quad \text{which implies} \quad c_k = 0.
\]

Therefore, we conclude that the only coefficient that can be nonzero in the Fourier series expansion of \( 1_A \) is \( c_0 \). This implies that \( 1_A \) is constant \( m \)-almost everywhere on \([0, 1]\). This yields two cases:

- The first case is \( 1_A(x) = 0 \) \( m \)-almost everywhere on \([0,1]\). Thus, \( m(A) = 0 \).
- The second case is \( 1_A(x) = 1 \) \( m \)-almost everywhere on \([0, 1]\). Thus, \( m(A) = 1 \).

So, we have \( m(A) \in \{0, 1\} \) which concludes the proof.

3. Steps 1 and 2 imply that \( f_4 \) is ergodic with respect to the probability measure \( m \circ h \) on \([0, 1]\). We claim that \( m \circ h = \sigma^* \). This is proved by showing that for every open interval \((a, b) \subset [0, 1], m(h((a, b))) = \sigma^*((a, b))\). It follows that \( m \circ h \) and \( \sigma^* \) agree on \( \mathcal{B}[0, 1] \) and thus that \( f_4 \) is ergodic with respect to the Ulam-von Neumann measure \( \sigma^* \).
Steps 1, 2, and 3 allow us to apply Birkhoff’s pointwise ergodic theorem to the function \( f_4 \) on the probability space \(([0,1], \mathcal{B}[0,1], \sigma^*)\). For any \( g \in L^1([0,1], \sigma^*) \) and for \( \sigma^* \)-almost every \( x \in [0,1] \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(f_4^{(k)}(x)) = \int_{[0,1]} g \, d\sigma^*. \tag{B.5}
\]

For use in the proof of Theorem 5.4 we have the following modification of (B.5).

**Lemma B.8.** The limit (B.5) holds for \( m \)-almost every \( x \in [0,1] \).

**Proof.** By construction, the Ulam-von Neumann measure \( \sigma^* \) is absolutely continuous with respect to Lebesgue measure \( m \) on \([0,1]\). In order to prove the lemma, we must show that \( m \) is absolutely continuous with respect to \( \sigma^* \). This follows from the fact that

\[
\min_{x \in [0,1]} \frac{1}{\pi \sqrt{x(1-x)}} = \frac{1}{\pi \sqrt{\frac{1}{2}(1 - \frac{1}{2})}} = \frac{2}{\pi}.
\]

Hence, if \( A \in \mathcal{B}[0,1] \) satisfies \( \sigma^*(A) = 0 \), then

\[
0 = \sigma^*(A) = \int_A \frac{1}{\pi \sqrt{x(1-x)}} \, dm \geq \int_A \frac{2}{\pi} \, dm = \frac{2}{\pi} m(A).
\]

This implies that \( m(A) = 0 \), completing the proof of the lemma. \( \blacksquare \)
APPENDIX C
PROOF OF ERGODICITY FROM FELLER [8]

In this appendix we provide the reader with detailed proofs of two theorems found in [8]. We first start with some basic definitions that are required for the statement of the theorems.

Definition C.1. For the following, let $\alpha$ be a probability measure, $\Omega$ a sample space and $K$ a probability transition kernel.

(i) A measure $\alpha$ is strictly positive in $\Omega$ if $\alpha\{I\} > 0$ for each open interval $I \subset \Omega$. The kernel $K$ is strictly positive if $K(x, I) > 0$ for each $x$ and each open interval $I$ in $\Omega$.

(ii) We say that a sequence of probability distributions $\{p_n\}$ converges weakly to a probability distribution $p$ in $\Omega$ if $E_n[u] \to E[u]$ for every bounded, continuous function $u$ in $\Omega$. $E_n$ and $E$ denote expectations with respect to $p_n$ and $p$ respectively.

(iii) Given a bounded, continuous function $u_0$ defined on $\Omega$, we define the sequence of transforms $\{u_n\}$ by induction as $u_n(x) = \int \Omega K(x, dy)u_{n-1}(y)$. Similarly, we define a transform of measures $\{\gamma_n\}$ given an initial probability distribution $\gamma_0$ and $\Gamma \subset \Omega$ a Borel set as $\gamma_n(\Gamma) = \int \Omega \gamma_{n-1}(dx)K(x, \Gamma)$.

(iv) We say that the kernel $K$ is regular if the sequence of transforms $\{u_n\}$ is equicontinuous whenever $u_0$ is uniformly continuous in $\Omega$.

(v) The kernel $K$ is ergodic if there exists a unique strictly positive probability measure $\alpha$ such that the transforms $\{\gamma_n\}$ of an initial probability distribution $\gamma_0$ converge weakly to $\alpha$, independent of $\gamma_0$.

(vi) We say that a probability distribution $\gamma$ is stationary for $K$ if it is invariant under transformation, i.e. for a Borel set $\Gamma$ we have $\gamma(\Gamma) = \int \Omega \gamma(dx)K(x, \Gamma)$

Theorem C.2 (pg. 265). A strictly positive, regular kernel $K$ in a bounded closed space $\Omega$ is ergodic.
Proof. Consider the sequence of transforms \( \{ u_n \} \) of a continuous function \( u_0 \) on \( \Omega \), where \( u_n(x) = \int_{\Omega} K(x, dy) u_{n-1}(y) \). Since \( \Omega \) is closed, \( u_0 \) is uniformly continuous on \( \Omega \) and since we have that \( K \) is regular, \( \{ u_n \} \) is equicontinuous on \( \Omega \). Also, \( u_0 \) is bounded, so there exists an \( A \) such that \( |u_0| \leq A \) on \( \Omega \). We have

\[
|u_1(x)| = \left| \int_{\Omega} K(x, dy) u_0(y) \right| \leq \int_{\Omega} K(x, dy)|u_0(y)| \leq A \int_{\Omega} K(x, dy) = A
\]

By induction we have \( |u_n| \leq A \) for all \( n \). Therefore the sequence \( \{ u_n \} \) is uniformly bounded.

Thus, by the Arzela-Ascoli theorem there exists a subsequence \( \{ u_{n_k} \} \) that converges uniformly to a continuous function \( v_0 \). The subsequence \( \{ u_{n_k+1} \} \), the transforms of the \( u_{n_k} \)'s, also converges uniformly to the transform \( v_1 \) of \( v_0 \), since

\[
|u_{n_k+1}(x) - v_1(x)| = \left| \int_{\Omega} K(x, dy)(u_{n_k}(y) - v_0(y)) \right| \leq \int_{\Omega} K(x, dy)|u_{n_k}(y) - v_0(y)|.
\]

So, given \( \epsilon > 0 \), we can choose \( k \) large enough so that \( |u_{n_k}(y) - v_0(y)| < \epsilon \). Then,

\[
|u_{n_k+1}(x) - v_1(x)| \leq \int_{\Omega} K(x, dy)\epsilon = \epsilon.
\]

Now consider the sequence \( \{ M_n \} \), where \( M_n = \max_{x \in \Omega} u_n(x) \). We will show that

\[
M_n \to M = \max v_0 = \max v_1.
\]

First, we will show that \( \{ M_n \} \) is a decreasing sequence. Suppose \( u_n(y_0) = \min_{x \in \Omega} u_n(x) \) for some \( y_0 \in \Omega \). Pick \( \delta > 0 \). Then, by continuity, there exists some interval \( I_{y_0} \) containing \( y_0 \), \( I_{y_0} \subseteq \Omega \), such that \( u_n(x) \leq u_n(y_0) + \delta \), \( \forall x \in I_{y_0} \). Thus

\[
M_{n+1} = \max_x u_{n+1}(x) = \max_x \int_{\Omega} K(x, dy) u_n(y)
\]

\[
= \max_x \left( \int_{\Omega \setminus I_{y_0}} K(x, dy) u_n(y) + \int_{I_{y_0}} K(x, dy) u_n(y) \right)
\]

\[
\leq \max_x \left( \int_{\Omega \setminus I_{y_0}} K(x, dy) M_n + \int_{I_{y_0}} K(x, dy)(u_n(y_0) + \delta) \right)
\]

123
\[
\begin{align*}
&= \max_x \left( M_n \int_{\Omega \setminus I_{y_0}} K(x, dy) + (u_n(y_0) + \delta) \int_{I_{y_0}} K(x, dy) \right) \\
&= \max_x [M_n (1 - K(x, I_{y_0})) + (u_n(y_0) + \delta) K(x, I_{y_0})] \\
&= \max_x [M_n - (M_n - u_n(y_0) - \delta) K(x, I_{y_0})] \\
&= M_n - (M_n - u_n(y_0) - \delta) \min_x K(x, I_{y_0}).
\end{align*}
\]

Since \( M_n - u_n(y_0) > 0 \), we can pick \( \delta \) small enough so that \( M_n - u_n(y_0) - \delta > 0 \). Also, suppose that at \( x_0 \in \Omega \) we have that \( K(x_0, I_{y_0}) = \min_x K(x, I_{y_0}) \). By the strict positivity of \( K \), we have \( K(x_0, I_{y_0}) > 0 \). Thus \( (M_n - u_n(y_0) - \delta) K(x_0, I_{y_0}) > 0 \) which implies

\[
M_{n+1} \leq M_n - (M_n - u_n(y_0) - \delta) \min_x K(x, I_{y_0}) \leq M_n.
\]

Therefore \( \{M_n\} \) is a decreasing sequence and is also bounded since \( \{u_n\} \) is bounded.

Thus, \( \{M_n\} \) converges to some \( M \). Now, pick \( \epsilon > 0 \). Then for \( k_0 \) large enough we have

\[
v_0(x) - \epsilon \leq u_{n_k}(x) \leq v_0(x) + \epsilon, \quad \forall x \in \Omega, \forall k \geq k_0,
\]

due to the uniform convergence of \( \{u_{n_k}\} \). So we can pick a sequence \( \{x_k\} \) in \( \Omega \) such that \( u_{n_k}(x_k) = M_{n_k} \), and we have

\[
M_{n_k} \leq v_0(x_k) + \epsilon \leq \max_x v_0 + \epsilon.
\]

Now, there exists some \( x' \in \Omega \) such that \( v_0(x') = \max_x v_0(x) \). Then,

\[
v_0(x') - \epsilon \leq u_{n_k}(x') \leq M_{n_k}.
\]

Therefore, \( \max_x v_0 - \epsilon \leq M_{n_k} \leq \max_x v_0 + \epsilon \) for all \( k \geq k_0 \) which implies \( M_{n_k} \to \max_x v_0 \). Since we know that \( M_n \to M \), necessarily \( M_{n_k} \to M \) which means \( M = \max_x v_0 \).

Also, we have that the subsequence \( \{M_{n_k + 1}\} \) converges to \( M \) and as was shown above for \( \max_x v_0 \), working with the subsequence \( \{u_{n_k + 1}\} \) in the same way we have that \( M = \max_x v_1 \). Thus, \( v_0 \) and \( v_1 \) have the same maximum.
Suppose $v_1(x_1) = M$ at some $x_1 \in \Omega$. Then,

$$M = v_1(x_1) = \int_{\Omega} K(x_1, dy)v_0(y)$$

$$\implies M - \int_{\Omega} K(x_1, dy)v_0(y) = M \int_{\Omega} K(x_1, dy) - \int_{\Omega} K(x_1, dy)v_0(y) = 0$$

$$\implies \int_{\Omega} K(x_1, dy)(M - v_0(y)) = 0.$$

We have $M - v_0(y) \geq 0$. Set $\xi(y) = M - v_0(y)$. Assume $\xi(y)$ is not identically 0. Then there is some $y'$ such that $\xi(y') > 0$. By continuity there is some interval $I \subset \Omega$ such that $y' \in I$ and $\xi(y) > 0 \quad \forall y \in I$. Then,

$$0 < \int_I K(x_1, dy)\xi(y) \leq \int_{\Omega} K(x_1, dy)\xi(y) = 0$$

which is a contradiction. Thus, $v_0(y) = M$ for all $y \in \Omega$. So $v_0$ is constant and therefore $v_1$ is also constant.

We have shown that $M_n \to M$. The same can be proven for the sequence $\{m_n\}$, where $m_n = \min_x u_n(x)$. Suppose $u_n(z) = M_n = \max_x u_n(x)$ for some $z \in \Omega$. Pick $\delta > 0$. There exists some interval $J \subset \Omega$ containing $z$ such that $u_n(x) \geq M_n - \delta$ for all $x \in J$.

$$m_{n+1} = \min_x u_{n+1}(x) = \min_x \int_{\Omega} K(x, dy)u_n(y)$$

$$= \min_x \left( \int_{\Omega \setminus J} K(x, dy)u_n(y) + \int_{J} K(x, dy)u_n(y) \right)$$

$$\geq \min_x \left( \int_{\Omega \setminus J} K(x, dy)m_n + \int_{J} K(x, dy)(M_n - \delta) \right)$$

$$= \min_x \left( m_n \int_{\Omega \setminus J} K(x, dy) + (M_n - \delta) \int_{J} K(x, dy) \right)$$

$$= \min_x \left[ m_n(1 - K(x, J)) + (M_n - \delta)K(x, J) \right]$$

$$= \min_x \left[ m_n + (M_n - m_n - \delta)K(x, J) \right]$$

125
\[ = m_n + (M_n - m_n - \delta) \min_x K(x, J_z). \]

Since \( M_n - m_n > 0 \), we can pick \( \delta \) small enough so that \( M_n - m_n - \delta > 0 \) and due to the strict positivity of \( K \) we have \( \min_x K(x, J_z) > 0 \). Therefore,

\[ m_{n+1} \geq m_n + (M_n - m_n - \delta) \min_x K(x, J_z) \geq m_n \]

so the sequence \( \{m_n\} \) is increasing and is also bounded, thus converges. Since \( u_{n_k}(x) \to M \) for all \( x \), then \( m_{n_k} \to M \) which implies \( m_n \to M \). Finally, we have \( m_n \leq u_n(x) \leq M_n \) and since \( M_n, m_n \) converge to \( M \) independent of \( x \), \( u_n \) converges to the constant \( M \) uniformly.

Now let \( \gamma_0 \) be an arbitrary probability distribution on \( \Omega \) and denote by \( E_n \) the expectation with respect to the transform \( \gamma_n \), where

\[ \gamma_n(\Gamma) = \int_{\Omega} \gamma_{n-1}(dx) K(x, \Gamma) \]

and \( \Gamma \) a Borel subset of \( \Omega \). First we will show that \( E_n[u_k] = E_0[u_{k+n}] \) by induction. Let’s start with \( n = 1 \).

\[ E_1[u_k] = \int_{\Omega} u_k(x) \gamma_1(dx) \]

\[ = \int_{\Omega} u_k(x) \left( \int_{\Omega} \gamma_0(dy) K(y, dx) \right) \]

\[ = \int_{\Omega} \int_{\Omega} u_k(x) \gamma_0(dy) K(y, dx) \]

\[ = \int_{\Omega} \left( \int_{\Omega} u_k(x) K(y, dx) \right) \gamma_0(dy) \]

\[ = \int_{\Omega} u_{k+1}(y) \gamma_0(dy) \]

\[ = E_0[u_{k+1}] \]

by Fubini-Tonelli, since \( E_n[|u_k|] = \int_{\Omega} |u_k(x)| \gamma_n(dx) \leq \int_{\Omega} A \gamma_n(dx) = A \). Now suppose that it
holds for \( n \) and we will show it holds for \( n + 1 \).

\[
E_{n+1}[u_k] = \int_{\Omega} u_k(x) \gamma_{n+1}(dx)
\]

\[
= \int_{\Omega} u_k(x) \left( \int_{\Omega} \gamma_n(dy) K(y, dx) \right)
\]

\[
= \int_{\Omega} \int_{\Omega} u_k(x) \gamma_n(dy) K(y, dx)
\]

\[
= \int_{\Omega} \left( \int_{\Omega} u_k(x) K(y, dx) \right) \gamma_n(dy)
\]

\[
= \int_{\Omega} u_{k+1}(y) \gamma_n(dy)
\]

\[
= E_n[u_{k+1}]
\]

and since it holds for \( n \), we have \( E_n[u_{k+1}] = E_0[u_{k+n+1}] \). Thus, \( E_n[u_k] = E_0[u_{k+n}] \). By the uniform convergence of \( u_n \) to \( M \), we have that

\[
E_n[u_0] = E_0[u_n] \to E_0[M] = M.
\]

This shows that for an arbitrary bounded, continuous function \( u_0 \), \( E_n[u_0] \to M \) and by invoking Theorem 2 in Section 1 of Chapter VIII in [8], which states that for a sequence of probability distributions \( \{\gamma_n\} \) and for an arbitrary bounded, continuous function \( u \), if \( E_n[u] \) converges to some constant \( C \), then there exists some probability distribution \( \gamma \) such that \( \gamma_n \) converges weakly to \( \gamma \), denoted by \( \gamma_n \Rightarrow \gamma \), with \( C = \int_{\Omega} u(x)\gamma(dx) \).

Thus, by that theorem, there exists some probability distribution \( \alpha \) on \( \Omega \) such that \( \gamma_n \Rightarrow \alpha \) and if \( E \) denotes the expectation with respect to \( \alpha \) then \( M = E[u_0] \).

This convergence is independent of the initial distribution \( \gamma_0 \), so to show that the \( n \)-step transition probability \( K^{(n)}(x_0, \ast) \) for some point \( x_0 \) converges weakly to \( \alpha \), we take \( \gamma_0 = \delta_{x_0} \) and a Borel set \( \Gamma \) and we have by induction

\[
\gamma_1(\Gamma) = \int_{\Omega} \gamma_0(dx) K(x, \Gamma) = \int_{\Omega} \delta_{x_0} K(x, \Gamma) = K(x_0, \Gamma),
\]

127
\[ \gamma_2(\Gamma) = \int_\Omega \gamma_1(dx)K(x,\Gamma) = \int_\Omega K(x_0, dx)K(x,\Gamma) = K^{(2)}(x_0, \Gamma). \]

So if it holds for \( n \), for \( n + 1 \) we have

\[ \gamma_{n+1}(\Gamma) = \int_\Omega \gamma_n(dx)K(x,\Gamma) = \int_\Omega K^{(n)}(x_0, dx)K(x,\Gamma) = K^{(n+1)}(x_0, \Gamma). \]

Thus, for \( \gamma_0 = \delta_{x_0} \) we have \( K^{(n)}(x_0, \ast) \Rightarrow \alpha \).

To show that \( \alpha \) is stationary for \( K \) we simply need to show that for any bounded, continuous function \( f_0 \), we have

\[ \int_\Omega f_0(x)\alpha(dx) = \int_\Omega \left( \int_\Omega \alpha(dy)K(y, dx) \right). \]

Let \( f_0 \) be an arbitrary bounded, continuous function on \( \Omega \) and let \( f_1 \) be its transform. By Fubini-Tonelli we have

\[ \int_\Omega f_0(x)\alpha(dx) = \lim_{n \to \infty} \int_\Omega f_0(x)\gamma_{n+1}(dx) \]

\[ = \lim_{n \to \infty} \int_\Omega f_0(x) \left( \int_\Omega \gamma_n(dy)K(y, dx) \right) \]

\[ = \lim_{n \to \infty} \int_\Omega \left( \int_\Omega f_0(x)K(y, dx) \right) \gamma_n(dy) \]

\[ = \lim_{n \to \infty} \int_\Omega f_1(y)\gamma_n(dy) \]

\[ = \int_\Omega f_1(y)\alpha(dy). \]

Also,

\[ \int_\Omega f_0(x) \left( \int_\Omega \alpha(dy)K(y, dx) \right) = \int_\Omega \left( \int_\Omega f_0(x)K(y, dx) \right)\alpha(dy) \]

\[ = \int_\Omega f_1(y)\alpha(dy). \]

Thus, the two measures \( \alpha \) and \( \int_\Omega \alpha(dy)K(y, \ast) \) are equal which implies \( \alpha \) is stationary for \( K \).

Finally, let \( I \subset \Omega \) be an open interval. Then for any \( x \in \Omega \) we have \( K(x, I) > 0 \). Suppose that for some \( x_0 \in \Omega \) we have \( K(x_0, I) = \min_{x \in \Omega} K(x, I) \). Again, \( K(x_0, I) = \epsilon > 0 \). Then, for \( K^{(n)}(x, I) \)
we have

\[ K^{(n)}(x, I) = \int_{\Omega} K^{(n-1)}(x, dy) K(y, I) \geq \int_{\Omega} K^{(n-1)}(x, dy) \epsilon = \epsilon. \]

Since \( K^{(n)}(x, I) \to \alpha(I) \) and \( K^{(n)}(x, I) \geq \epsilon \) for any \( n \), then \( \alpha(I) \geq \epsilon > 0 \). Thus, \( \alpha \) is strictly positive.

\[ \text{Theorem C.3 (pg. 266). A strictly positive, regular kernel } K \text{ is ergodic iff it possesses a strictly positive stationary distribution } \alpha. \]

\textbf{Proof.} Towards the end of the previous theorem we showed that if \( K \) is ergodic with limiting distribution \( \alpha \) then it is a strictly positive stationary distribution. Suppose now that \( K \) has a strictly positive stationary distribution \( \alpha \) and let \( \epsilon \) denote the expectation with respect to \( \alpha \). Take an arbitrary bounded, continuous function \( u_0 \) on \( \Omega \) and let \( \{ u_n \} \) denote the sequence of its transforms with the kernel \( K \). Taking expectations, we have

\[
E[u_1] = \int_{\Omega} u_1(x) \alpha(dx) = \int_{\Omega} \left( \int_{\Omega} K(x, dy) u_0(y) \right) \alpha(dx) = \int_{\Omega} u_0(y) \left( \int_{\Omega} \alpha(dx) K(x, dy) \right) = \int_{\Omega} u_0(y) \alpha(dy) = E[u_0]
\]

by Fubini-Tonelli, since as shown in the proof of the previous theorem the sequence of functions \( \{ u_n \}_{n=0}^{\infty} \) are uniformly bounded. There exists some \( A \) such that \( |u_n| \leq A \) for all \( n \). Thus, \( E[|u_n|] = \int_{\Omega} |u_n(x)| \alpha(dx) \leq \int_{\Omega} A \alpha(dx) = A < \infty \) and also using the fact that \( \alpha \) is stationary for \( K \). So \( E[u_0] = E[u_1] \).
Suppose that holds up until $k$, i.e. $E[u_0] = E[u_1] = \ldots = E[u_k]$. Working as before we have

$$E[u_{k+1}] = \int_\Omega u_{k+1}(x)\alpha(dx)$$

$$= \int_\Omega \left( \int_\Omega K(x, dy)u_k(y) \right) \alpha(dx)$$

$$= \int_\Omega u_k(y) \left( \int_\Omega \alpha(dx)K(x, dy) \right)$$

$$= \int_\Omega u_k(y)\alpha(dy)$$

$$= E[u_k].$$

Therefore, $E[u_n] = C$ for all $n$, where $C$ is some constant. Furthermore, the sequence $\{E[|u_n|]\}$ is decreasing:

$$E[|u_{k+1}|] = \int_\Omega |u_{k+1}(x)|\alpha(dx)$$

$$= \int_\Omega \left| \int_\Omega K(x, dy)u_k(y) \right| \alpha(dx)$$

$$\leq \int_\Omega \left( \int_\Omega K(x, dy)|u_k(y)| \right) \alpha(dx)$$

$$= \int_\Omega |u_k(y)| \left( \int_\Omega \alpha(dx)K(x, dy) \right)$$

$$= \int_\Omega |u_k(y)|\alpha(dy)$$

$$= E[|u_k|]$$

again by applying Fubini-Tonelli and using the stationarity of $\alpha$. Thus,

$$E[|u_{k+1}|] \leq E[|u_k|].$$

Since the sequence $\{E[|u_n|]\}$ is decreasing and that $E[|u_n|] \geq 0$, this implies that $E[|u_n|]$ converges
to some \( m \). From the bounded sequence \( \{u_n\} \) we can choose a subsequence \( \{u_{n_k}\} \) that converges to \( v_0 \). Then, as shown in the proof of the previous theorem, \( u_{n_k+1} \to v_1 \), where \( u_{n_k+1}, v_1 \) are the transforms of \( u_{n_k}, v_0 \) respectively. Since the functions \( \{u_{n_k}\}_{k=1}^{\infty} \) are uniformly bounded, continuous functions on \( \Omega \) and converge to \( v_0 \), applying the Bounded Convergence Theorem yields \( E[u_{n_k}] \to E[v_0] \). Also,

\[
|E[|u_{n_k}|] - E[|v_0|]| = \left| \int_{\Omega} |u_{n_k}(x)| \alpha(dx) - \int_{\Omega} |v_0(x)| \alpha(dx) \right|
\]

\[
\leq \int_{\Omega} ||u_{n_k}(x)| - |v_0(x)|| \alpha(dx)
\]

\[
\leq \int_{\Omega} |u_{n_k}(x) - v_0(x)| \alpha(dx) \to 0
\]

since we already have applied the Bounded convergence Theorem to \( u_{n_k} \). Therefore, \( E[|u_{n_k}|] \to E[|v_0|] \). Since \( E[u_n] = C \) for all \( n \), and since \( E[|u_n|] \) converges to \( m \), thus \( E[|u_{n_k}|] \) and \( E[|u_{n_k+1}|] \) converge to the same limit, we have

\[
C = E[v_0] = E[v_1] \quad \text{and} \quad E[|v_0|] = E[|v_1|] = m.
\]

The last equality actually shows that \( v_0 \) cannot change signs. To prove it by contradiction, let’s assume that \( S^+ = \{x \in \Omega : v_0(x) > 0\} \) and \( S^- = \{x \in \Omega : v_0(x) < 0\} \) such that \( K(x', S^+) > 0 \) and \( K(x', S^-) > 0, \forall x' \in \Omega \). The equality \( E[|v_0|] = E[|v_1|] \) yields

\[
E[|v_1|] = \int_{\Omega} |v_1(x)| \alpha(dx)
\]

\[
= \int_{\Omega} \left| \int_{\Omega} K(x, dy) v_0(y) \right| \alpha(dx)
\]

\[
\leq \int_{\Omega} \left( \int_{\Omega} K(x, dy) |v_0(y)| \right) \alpha(dx),
\]

and by Fubini-Tonelli and the stationarity of \( \alpha \) we have that

\[
\int_{\Omega} \left( \int_{\Omega} K(x, dy) |v_0(y)| \right) \alpha(dx) = \int_{\Omega} |v_0(y)| \left( \int_{\Omega} \alpha(dx) K(x, dy) \right)
\]

131
\[
\int_{\Omega} |v_0(y)| \alpha(dy)
\]

\[
= E[|v_0|] = E[|v_1|].
\]

Thus, the inequality is in fact an equality. Therefore, we have that

\[
\int_{\Omega} \left| \int_{\Omega} K(x, dy)v_0(y) \right| \alpha(dx) = \int_{\Omega} \left( \int_{\Omega} K(x, dy)|v_0(y)| \right) \alpha(dx). \tag{C.1}
\]

Using the sets \(S^+\) and \(S^-\), \(\int_{\Omega} \left| \int_{\Omega} K(x, dy)v_0(y) \right| \alpha(dx)\) can be written as

\[
\int_{\Omega} \left| \int_{S^+} K(x, dy)v_0(y) + \int_{S^-} K(x, dy)v_0(y) \right| \alpha(dx)
\]

and since the kernel \(K\) is non-negative, we have that

\[
\int_{S^+} K(x, dy)v_0(y) > 0 \quad \text{and} \quad \int_{S^-} K(x, dy)v_0(y) < 0.
\]

Now,

\[
\int_{\Omega} \left( \int_{S^+} K(x, dy)v_0(y) + \int_{S^-} K(x, dy)v_0(y) \right) \alpha(dx)
\]

\[
< \int_{\Omega} \left( \left| \int_{S^+} K(x, dy)v_0(y) \right| + \left| \int_{S^-} K(x, dy)v_0(y) \right| \right) \alpha(dx)
\]

which is a strict inequality since \(\left| \int_{S^-} K(x, dy)v_0(y) \right| > 0\). However,

\[
\int_{\Omega} \left( \left| \int_{S^+} K(x, dy)v_0(y) \right| + \left| \int_{S^-} K(x, dy)v_0(y) \right| \right) \alpha(dx)
\]

\[
\leq \int_{\Omega} \left( \int_{S^+} K(x, dy)v_0(y) + \int_{S^-} K(x, dy)v_0(y) \right) \alpha(dx)
\]

\[
= \int_{\Omega} \left( \int K(x, dy)v_0(y) \right) \alpha(dx).
\]
This means that
\[
\int_{\Omega} \left| \int_{\Omega} K(x, dy)v_0(y) \right| \alpha(dx) < \int_{\Omega} \left( \int_{\Omega} |K(x, dy)v_0(y)| \right) \alpha(dx)
\]
which contradicts (C.1). Therefore, \( v_0 \) must preserve its sign.

Since \( u_0 \) was an arbitrarily chosen bounded, continuous function, if we had that \( E[v_0] = 0 \), then since \( v_0 \) preserves its sign we must have that \( v_0(x) = 0 \quad \forall x \in \Omega \). Thus, we can choose \( u_0 - C \) as an initial bounded, continuous function. This choice will give us that
\[
u_{nk} - C \rightarrow v_0 - C
\]
and that \( E[v_0 - C] = E[u_0 - C] = E[u_0] - C = C - C = 0 \). Therefore, \( v_0(x) = E[u_0] = C \quad \forall x \in \Omega \). This also implies that \( v_1 \) is just the constant \( C \). For any \( x \),
\[
v_1(x) = \int_{\Omega} K(x, dy)v_0(y) = \int_{\Omega} K(x, dy)C = C.
\]
Thus, both \( u_{nk} \) and \( u_{nk+1} \) converge to the constant \( C \). Suppose that \( u_{nk+l} \) also converges to \( C \). We will show that \( u_{nk+l+1} \) will also have the same limit. Pick \( \epsilon > 0 \). There exists \( k \) large enough such that \( |u_{nk+l}(x) - C| < \epsilon, \quad \forall x \in \Omega \). We then have
\[
|u_{nk+l+1}(x) - C| = \left| \int_{\Omega} K(x, dy)u_{nk+l}(y) - C \right|
\]
\[
= \left| \int_{\Omega} K(x, dy)u_{nk+l}(y) - \int_{\Omega} K(x, dy)C \right|
\]
\[
\leq \int_{\Omega} K(x, dy)|u_{nk+l}(y) - C|
\]
\[
< \int_{\Omega} K(x, dy)\epsilon
\]
\[
= \epsilon.
\]
Thus, for any \( l \) we have that \( u_{n_k+l} \) converges to \( C \), which implies \( u_n \to C \). Since

\[
   u_n(x) = \int_\Omega K(x, dy) u_{n-1}(y) = \int_\Omega u_0(y) K^{(n)}(x, dy),
\]

we have that

\[
   \int_\Omega u_0(y) K^{(n)}(x, dy) \to C = E[u_0] = \int_\Omega u_0(y) \alpha(dy)
\]

and since \( u_0 \) was an arbitrary bounded, continuous function, we have that \( K^{(n)}(x, \ast) \Rightarrow \alpha \).
APPENDIX D

B-SPLINES

The idea behind logspline density estimation of an unknown density \( p \) is that the logarithm of \( p \) is estimated by a spline function, a piecewise polynomial that interpolates the function to be estimated. Therefore, the family of estimators constructed for the unknown density is a family of functions that are exponentials of splines that are suitably normalized so that they can be densities. Thus, to build up the logspline estimation method, we need to start the theory with the building blocks of splines themselves, the functions we call basis splines or B-splines for short whose linear combination generates the set of splines of a given order.

So, the main question we will answer in this appendix is how we construct B-splines. There are several ways to do this, some less intuitive than others. The approach we will take will be through the use of divided differences. It is a recursive division process that is used to calculate the coefficients of interpolating polynomials written in a specific form called the Newton form.

**Definition D.1.** The \( k \)th divided difference of a function \( g \) at the knots \( t_0, \ldots, t_k \) is the leading coefficient (meaning the coefficient of \( x^k \)) of the interpolating polynomial \( q \) of order \( k+1 \) that agrees with \( g \) at those knots. We denote this number as

\[
[t_0, \ldots, t_k]g.
\]  

(D.1)

Here we use the terminology found in De Boor [3], where a polynomial of order \( k+1 \) is a polynomial of degree less than or equal to \( k \). It’s better to work with the ”order” of a polynomial since all polynomials of a certain order form a vector space, whereas polynomials of a certain degree do not. The term “agree” in the definition means that for the sequence of knots \( (t_i)_{i=0}^k \), if \( \zeta \) appears in the sequence \( m \) times, then for the interpolating polynomial we have

\[
q^{(i-1)}(\zeta) = g^{(i-1)}(\zeta), \quad i = 1, \ldots, m.
\]  

(D.2)

Since the interpolating polynomial depends only on the data points, the order in which the values of \( t_0, \ldots, t_1 \) appear in the notation in (D.1) does not matter. Also, if all the knots are distinct, then the interpolating polynomial is unique.
At this point let’s write down some examples to see how the recursion algorithm pops up. If we want to interpolate a function \( g \) using only one knot, say \( t_0 \), then we will of course have the constant polynomial \( q(x) = g(t_0) \). Thus, since \( g(t_0) \) is the only coefficient, we have

\[
[t_0]g = g(t_0). \tag{D.3}
\]

Now suppose we have two knots, \( t_0, t_1 \).

If \( t_0 \neq t_1 \), then \( q \) is the secant line defined by the two points \((t_0, g(t_0))\) and \((t_1, g(t_1))\). Thus, the interpolating polynomial will be given by

\[
q(x) = g(t_0) + (x - t_0)\frac{g(t_1) - g(t_0)}{t_1 - t_0}. \tag{D.4}
\]

Therefore,

\[
[t_0, t_1]g = \frac{g(t_1) - g(t_0)}{t_1 - t_0} = \frac{[t_1]g - [t_0]g}{t_1 - t_0}. \tag{D.5}
\]

To see what happens when \( t_0 = t_1 \), we can take the limit \( t_1 \to t_0 \) above and thus \([t_0, t_1]g = g'(t_0)\).

By continuing these calculations for more knots yields the following result:

**Lemma D.2.** Given a function \( g \) and a sequence of knots \( (t_i)_{i=0}^k \), the kth divided difference of \( g \) is given by

(a) \([t_0, \ldots, t_k]g = \frac{g^{(k)}(t_0)}{k!}\) when \( t_0 = \cdots = t_k, g \in C^k \), therefore yielding the leading coefficient of the Taylor approximation of order \( k+1 \) to \( g \).

(b) \([t_0, \ldots, t_k]g = \frac{[t_0, \ldots, t_{r-1}, t_{r+1}, \ldots, t_k]g - [t_0, \ldots, t_{s-1}, t_{s+1}, \ldots, t_k]g}{t_s - t_r}, \) where \( t_r \) and \( t_s \) are any two distinct knots in the sequence \((t_i)_{i=0}^k\).

Now that we have defined the kth divided difference of a function, we can easily state what B-splines are. B-splines arise as appropriately scaled divided differences of the positive part of a certain power function and it can be shown that B-splines form a basis of the linear space of splines of some order. Let’s start with the definition.

**Definition D.3.** Let \( t = (t_i)_{i=0}^N \) be a nondecreasing sequence of knots. Let \( 1 \leq k \leq N \). The j-th B-spline of order \( k \), with \( j \in \{0, 1, \ldots, N - k\} \), for the knot sequence \((t_i)_{i=0}^N\) is denoted by \( B_{j,k,t} \).
and is defined by the rule

\[ B_{j,k,t}(x) = (t_{j+k} - t_j)[t_j, \ldots, t_{j+k}](\cdot - x)_{+}^{k-1} \]  \hspace{1cm} (D.6)

where \( (\cdot)_+ \) defines the positive part of a function, i.e. \( (f(x))_+ = \max_x \{ f(x), 0 \} \).

The "placeholder" notation in the above definition says that the kth divided difference of \((\cdot - x)_{+}^{k-1}\) is to be considered for the function \((t - x)_{+}^{k-1}\) as a function of \(t\) and have \(x\) fixed. Of course, in the end the number will vary as \(x\) varies, giving rise to the function \(B_{j,k,t}\). If either \(k\) or \(t\) can be inferred from context then we will usually drop them from the notation and write \(B_j\) instead of \(B_{j,k,t}\). A direct consequence we receive from the above definition is the support of \(B_{j,k,t}\).

**Lemma D.4.** Let \(B_{j,k,t}\) be defined as in D.3. Then the support of the function is contained in the interval \([t_j, t_{j+k})\).

**Proof.** All we need to do is show that if \(x \notin [t_j, t_{j+k})\), then \(B_{j,k,t}(x) = 0\).

Suppose first that \(x \geq t_{j+k}\). Then we will have that \(t_i - x \leq 0\) for \(i = j, \ldots, j + k\) which in turn implies \((t_i - x)_+ = 0\) and finally \([t_j, \ldots, t_{j+k}](\cdot - x)_{+}^{k-1} = 0\).

On the other hand, if \(x < t_j\), then since \((t - x)_{+}^{k-1}\) as a function of \(t\) is a polynomial of order \(k\) and we have \(k + 1\) sites where it agrees with its interpolating polynomial, necessarily they are both the same. This implies \([t_j, \ldots, t_{j+k}](\cdot - x)_{+}^{k-1} = 0\) since the coefficient of \(t^k\) is zero. \(\blacksquare\)
Since we stated the definition of B-splines using divided differences, we can use that to state the recurrence relation for B-splines which will be useful when we will later prove various properties of these functions. We start by stating and proving the Leibniz formula which will be needed in the proof of the recurrence relation.

**Lemma E.1.** Suppose $f, g, h$ are functions such that $f = g \cdot h$, meaning $f(x) = g(x)h(x)$ for all $x$ and let $(t_i)$ be a sequence of knots. Then we have the following formula

$$[t_j, \ldots, t_{j+k}]f = \sum_{r=j}^{j+k}([t_j, \ldots, t_r]g)([t_r, \ldots, t_{j+k}]h), \text{ for some } j, k \in \mathbb{N}. \quad (E.1)$$

**Proof.** First of all, observe that the function

$$\left( g(t_j) + \sum_{r=j+1}^{j+k} (x - t_j) \cdots (x - t_{r-1})[t_j, \ldots, t_r]g \right) \left( h(t_{j+k}) + \sum_{s=j}^{j+k-1} (x - t_{s+1}) \cdots (x - t_{j+k})[t_s, \ldots, t_{j+k}]h \right)$$

agrees with $f$ at the knots $t_j, \ldots, t_{j+k}$ since the first and second factor agree with $g$ and $h$ respectively at those values. Now, observe that if $r > s$ then the above product vanishes at all the knots since the term $(x - t_i)$ for $i = j, \ldots, j + k$ will appear in at least one of the two factors. Thus, the above agrees with $f$ at $t_j, \ldots, t_{j+k}$ when $r \leq s$. But then the product turns into a polynomial of order $k + 1$ whose leading coefficient is

$$\sum_{r=s}^{j+k}([t_j, \ldots, t_r]g)([t_s, \ldots, t_{j+k}]h)$$

and that of course must be equal to

$$[t_j, \ldots, t_{j+k}]f.$$

Now we can state and prove the recurrence relation for B-splines.

**Lemma E.2.** Let $t = (t_i)_{i=0}^N$ be a sequence of knots and let $1 \leq k \leq N$. For $j \in \{0, 1, \ldots, N - k\}$ we can construct the $j$-th B-spline $B_{j,k}$ of order $k$ associated with the knots $t = (t_i)_{i=0}^N$ as follows:
(1) First we have $B_{j,1}$ be the characteristic function on the interval $[t_j, t_{j+1})$

$$B_{j,1}(x) = \begin{cases} 1, & x \in [t_j, t_{j+1}) \\ 0, & x \notin [t_j, t_{j+1}) \end{cases} \tag{E.2}$$

(2) The B-splines of order $k$ for $k > 1$ on $[t_j, t_{j+k})$ are given by

$$B_{j,k}(x) = \frac{x - t_j}{t_{j+k-1} - t_j} B_{j,k-1}(x) + \frac{t_{j+k} - x}{t_{j+k} - t_{j+1}} B_{j+1,k-1}(x) \tag{E.3}$$

**Proof.** (1) easily follows from the definition we gave for B-splines using divided differences in Definition D.3. (2) can be proven using Lemma E.1. Since B-splines were defined using the function $(t - x)^{k-1}_+$ for fixed $x$, we apply the Leibniz formula for the $k$th divided difference to the product

$$(t - x)^{k-1}_+ = (t - x)(t - x)^{k-2}_+.$$ 

This yields

$$[t_j, \ldots, t_{j+k}](\cdot - x)^{k-1}_+ = (t_j - x)[t_j, \ldots, t_{j+k}](\cdot - x)^{k-2}_+ + 1 \cdot [t_{j+1}, \ldots, t_{j+k}](\cdot - x)^{k-2}_+ \tag{E.4}$$

since $[t_j](\cdot - x) = (t_j - x)$, $[t_j, t_{j+1}](\cdot - x) = 1$ and $[t_j, \ldots, t_r](\cdot - x) = 0$ for $r > j + 1$. Now, from Lemma D.2 (b), we have that $(t_j - x)[t_j, \ldots, t_{j+k}](\cdot - x)^{k-2}_+$ can be written as

$$(t_j - x)[t_j, \ldots, t_{j+k}](\cdot - x)^{k-2}_+ = \frac{t_j - x}{t_{j+k} - t_j} ([t_{j+1}, \ldots, t_{j+k}] - [t_j, \ldots, t_{j+k-1}]) \tag{E.5}$$

Thus, by replacing that term in the result (E.4) we obtained by Leibniz, we get

$$[t_j, \ldots, t_{j+k}](\cdot - x)^{k-1}_+ = \frac{x - t_j}{t_{j+k} - t_j} [t_j, \ldots, t_{j+k-1}](\cdot - x)^{k-2}_+ + \frac{t_{j+k} - x}{t_{j+k} - t_j} [t_{j+1}, \ldots, t_{j+k}](\cdot - x)^{k-2}_+ \tag{E.6}$$

The result in (2) follows immediately once we multiply both sides by $(t_{j+k} - t_j)$ and then multiply and divide the first term in the sum on the right hand side by $(t_{j+k-1} - t_j)$ and then multiply and divide the second term by $(t_{j+k} - t_{j+1})$. 

§
From the recurrence relation we acquire information about B-splines that was not clear from
the first definition we gave using divided differences. $B_{j,1}$ is a characteristic function, or otherwise
piecewise constant. By Lemma E.2 (b), since the coefficients of $B_{j, k-1}$ are linear functions of $x$,
we have $B_{j,2}$ is a piecewise linear function on $[t_j, t_{j+2})$. Therefore, inductively we have $B_{j,3}$ is a
piecewise parabolic function on $[t_j, t_{j+3})$, $B_{j,4}$ is a piecewise polynomial of degree 3 on $[t_j, t_{j+4})$
and so on. Below there is a visual representation of B-splines showing how the graph changes as
the order increases.

Since we now have defined what a B-spline is as a function, the next step is to ask what set is
generated when considering linear combinations of these functions. Since B-splines are piecewise
polynomials themselves, we have that this set is a subset of the set of piecewise polynomials with
breaks at the knots $(t_i)$. Something that can be proven though, is that it is exactly the set of piecewise
polynomials with certain break and continuity conditions at the knots and this equality occurs on a
smaller interval, which we call the basic interval, denoted by $I_{k,t}$.

**Definition E.3.** Suppose $t = (t_0, \ldots, t_N)$ is a nondecreasing sequence of knots. Then for the B-
splines of order $k$, with $2k < N + 2$, that arise from these knots, we define $I_{k,t} = [t_{k-1}, t_{N-k+1}]$
and call it the **basic interval**.

**Remark E.4.** In order for this definition to be correct, we need to extend the B-splines and have
them be left continuous at the right endpoint of the basic interval since we are defining it as a closed
interval.
Remark E.5. The basic interval for the $N - k + 1$ B-splines of order $k > 1$ is defined in such a way so that at least two of them are always supported on any subinterval of $I_{k,t}$ and later we will see that the B-splines form a partition of unity on the basic interval. For $k = 1$, by construction the B-splines already form a partition of unity on $I_{1,t} = [t_0, t_N]$.

For example, let $t = (t_i)_{i=0}^6$ be disjoint and $k = 3$. Then there are 4 B-splines, $B_{j,3}$, $j = 0, 1, 2, 3$, of order 3 that arise in this framework. Their supports are $[t_0, t_3)$, $[t_1, t_4)$, $[t_2, t_5)$, $[t_3, t_6)$ respectively. Clearly, on $[t_0, t_1)$ only $B_{0,3}$ is supported and since as a function is non-constant we cannot have $\sum_{j=0}^3 B_{j,3} = B_{0,3}$ on $[t_0, t_1)$ be equal to 1.

The partition of unity is stated and proved in the next lemma together with other properties of the B-splines. The recurrence relation makes the proofs fairly easy compared to using the divided difference definition of the B-splines.

Lemma E.6. Let $B_{j,k,t}$ be the function as given in Definition D.3 for the knot sequence $t = (t_i)_{i=0}^N$. Then the following hold:

(a) $B_{j,k,t}(x) > 0$ for $x \in (t_j, t_{j+k})$.

(b) (Marsden’s Identity) For any $\alpha \in \mathbb{R}$, we have $(x - \alpha)^{k-1} = \sum_j \psi_{j,k}(\alpha)B_{j,k,t}(x)$, where

$$\psi_{j,k}(\alpha) = (t_{j+1} - \alpha) \ldots (t_{j+k-1} - \alpha)$$

and $\psi_{j,1}(\alpha) = 1$.

(c) $\sum_j B_{j,k,t} = 1$ on the basic interval $I_{k,t}$.

Proof. (a) This is a simple induction. For $k = 1$ the hypothesis holds since the B-splines are just characteristic functions on $[t_j, t_{j+1})$ and thus strictly positive in the interior.

For $k = 2$ by the recurrence relation, $B_{j,2,t}$ is a linear combination of $B_{j,1}$, $B_{j+1,1}$ with coefficients the linear functions $\frac{x-t_j}{t_{j+1}-t_j}$, $\frac{t_{j+2}-x}{t_{j+2}-t_{j+1}}$, which is positive on $(t_j, t_{j+2})$.

Assuming the hypothesis holds for $k = r$, we can show it is true for $k = r + 1$ by using the same argument as in the previous case.

(b) Let $\omega_{j,k}(x) = \frac{x-t_j}{t_{j+k-1}-t_j}$. Thus, $\frac{t_{j+k}-t}{t_{j+k}-t_{j+1}} = 1 - \omega_{j+1,k}(x)$. This way we can write the recurrence relation as

$$B_{j,k}(x) = \omega_{j,k}(x)B_{j,k-1}(x) + (1 - \omega_{j+1,k}(x))B_{j+1,k-1} \quad (E.7)$$
Using this we can write $\sum_j \psi_{j,k}(\alpha)B_{j,k,t}(x)$ as

$$
\sum_j \psi_{j,k}(\alpha)B_{j,k,t}(x) = \sum_j [\omega_{j,k}(x)\psi_{j,k}(\alpha) + (1 - \omega_{j,k}(x))\psi_{j-1,k}(\alpha)]B_{j,k-1,t}(x)
$$

$$
= \sum_j \psi_{j,k-1}(\alpha)[\omega_{j,k}(x)(t_{j+k-1} - \alpha) + (1 - \omega_{j,k}(x))(t_j - \alpha)]B_{j,k-1,t}(x)
$$

$$
= \sum_j \psi_{j,k-1}(\alpha)(x - \alpha)B_{j,k-1,t}(x)
$$

(E.8)

since $\omega_{j,k}(x)f(t_{j+k-1}) + (1 - \omega_{j,k}(x))f(t_j)$ is the unique straight line that intersects $f$ at $x = t_j$ and $x = t_{j+k-1}$. Thus,

$$
\omega_{j,k}(x)(t_{j+k-1} - \alpha) + (1 - \omega_{j,k}(x))(t_j - \alpha) = x - \alpha.
$$

Therefore, by induction we have

$$
\sum_j \psi_{j,k}(\alpha)B_{j,k,t}(x) = \sum_j \psi_{j,1}(\alpha)(x - \alpha)^{k-1}B_{j,1,t}(x)
$$

$$
= (x - \alpha)^{k-1} \sum_j \psi_{j,1}(\alpha)B_{j,1,t}(x)
$$

$$
= (x - \alpha)^{k-1}
$$

since $\psi_{j,1}(\alpha) = 1$ and $B_{j,1,t}$ are just characteristic functions.

(c) To prove the partition of unity, we start with Marsden’s Identity and divide both sides by $(k - 1)!$ and differentiate $\nu - 1$ times with respect to $\alpha$ for some positive integer $\nu \leq k - 1$. We then have

$$
\frac{(x - \alpha)^{k-\nu}}{(k - \nu)!} = \sum_j \frac{(-1)^{\nu-1}d^{\nu-1}\psi_{j,k}(\alpha)}{d\alpha^{\nu-1}}B_{j,k,t}(x).
$$

(E.9)

Now, for some polynomial $q$ of order $k$, we can use the Taylor expansion of $q$

$$
q = \sum_{\nu=1}^{k} \frac{(x - \alpha)^{k-\nu}d^{k-\nu}q(\alpha)}{(k - \nu)!}d\alpha^{k-\nu}.
$$

(E.10)
Using this we see that

$$q = \sum_j \lambda_{j,k}[q]B_{j,k,t} \quad \text{where} \quad \lambda_{j,k}[q] = \sum_{\nu=1}^{k} \frac{(-1)^{\nu-1} d^{\nu-1} \psi_{j,k}(\alpha)}{d\alpha^{\nu-1}} \frac{d^{k-\nu}q(\alpha)}{d\alpha^{k-\nu}}$$  \hspace{1cm} (E.11)

which holds only on the basic interval. Now, to show that the B-splines are a partition of unity, we just use this identity for $q = 1$.

**Remark E.7.** Marsden’s Identity says something very important. That all polynomials of order $k$ are contained in the set generated by the B-splines $B_{j,k}$, which is also what makes the step in the proof of (c) viable. Furthermore, we can replace the $(x - \alpha)$ in the identity by $(x - \alpha)_+$ which shows that piecewise polynomials are also contained in the same set.

**Remark E.8.** Another consequence of Marsden’s Identity is the Curry-Schoenberg theorem. We do not explicitly state the theorem as we do not require it, rather we state a simple result from it for B-splines of order $k$ given a sequence of knots $(t_i)_{i=0}^N$, which can be summarized as

$$\text{number of continuity conditions at } t_i + \text{multiplicity of } t_i = k.$$

Therefore, for a simple knot $t_i$, any B-spline of order $k$ there will be continuous and also have $k-2$ continuous derivatives. On the other hand, if $t_i$ has multiplicity $k$, any $k$-th order B-spline will have a discontinuity there.

Below there is a figure which shows the importance of the basic interval as the interval where we have partition of unity.
Remark E.9. When the sequence of $t_i'$s is distinct then the sum of B-splines belongs to $C_0((t_0, t_N))$. However, the sum of B-splines on the basic interval $I_{k, t}$ is equal to 1. To make sure that the sum equals to 1 on the whole interval $(t_0, t_N)$, the assumption of the knots being distinct has to be dropped. It is obvious that we have to take $t_0 = \cdots = t_{k-1}$ and $t_{N-k+1} = \cdots = t_N$.

Definition E.10. Let $(t_i)_{i=0}^N$ be a sequence of knots such that $t_0 = \cdots = t_{k-1}$ and $t_{N-k+1} = \cdots = t_N$, where $1 \leq k \leq N$. Let $B_{j, k, t}$ be the B-splines as defined in Definition D.3 with knot sequence $t = (t_i)_{i=0}^N$. The set generated by the sequence $\{B_{j, k, t} : \text{all } j\}$, denoted by $S_{k, t}$, is the set of splines of order $k$ with knot sequence $t$. In symbols we have

$$S_{k, t} = \left\{ \sum_j a_j B_{j, k, t} : a_j \in \mathbb{R}, \text{ all } j \right\}.$$  \hspace{1cm} (E.12)

Remark E.11. Fix an interval $[a, b]$. Let $T_N = (t_i)_{i=0}^N$ be a sequence as in Definition E.10 with $t_0 = a$ and $t_N = b$, where $N \in \mathbb{N}$. The choice in Definition E.10 implies that

$$\bigcup_{N \in \mathbb{N}} S_{k, T_N} \text{ is dense in } C([a, b]). \hspace{1cm} (E.13)$$
APPENDIX F

DERIVATIVES OF B-SPLINE FUNCTIONS

In this dissertation where we conduct our analysis on MISE, derivatives of spline functions factor in. Since splines are just linear combinations of B-splines we just need to investigate the result of differentiating a B-spline on the interior of its support. The derivative of a \( k \)-th order B-spline is directly associated with B-splines of order \( k - 1 \). To see this we use the recurrence relation which leads us to the following theorem:

**Theorem F.1.** Let \( B_{j,k,t} \) be the function as defined in Definition D.3. The support of \( B_{j,k,t} \) is the interval \([t_j, t_{j+k})\). Then the following equation holds on the open interval \((t_j, t_{j+k})\)

\[
\frac{d}{d\theta} B_{j,k,t}(\theta) = \begin{cases} 
0, & k = 1 \\
(k - 1) \left( B_{j,k-1,t}(\theta) - \frac{B_{j+1,k-1,t}(\theta)}{t_{j+k} - t_{j+1}} \right), & k > 1 \end{cases} \quad (F.1)
\]

**Proof.** The proof is done by induction on \( k \). For \( k = 1 \) it is straightforward since \( B_{j,1,t} \) is a constant on \((t_j, t_{j+1})\) and for \( k > 1 \) we use the recurrence relation described in Lemma E.2.

Using the above formula we can easily obtain bounds for higher derivatives of B-splines. First of all, by construction of the space \( S_{k,t} \), the B-splines we will be working with form a partition of unity on \([t_0, t_N] \) and since they are strictly positive on the interior of their supports, we have that each B-spline is bounded by 1 for all \( \theta \).

\[ B_{j,k,t}(\theta) \leq 1, \quad \forall \theta \in \mathbb{R}. \]

Furthermore, by induction we can prove the following lemma:

**Lemma F.2.** Let \( t = (t_i)_{i=0}^N \) be a sequence of knots as in Definition E.10 and \( B_{j,k,t} \) be the function as defined in Definition D.3. Let \( h_N = \min_{k \leq i \leq N, i+k+1} (t_i - t_{i-1}) \) and \( \alpha \) be a positive integer such that \( \alpha < k - 1 \). Then, on the open interval \((t_j, t_{j+k})\) we have

\[
\sup_{\theta \in (t_j, t_{j+k}) \mid \theta} \left| \frac{d^\alpha}{d\theta^\alpha} B_{j,k,t}(\theta) \right| \leq \frac{2^\alpha (k-1)!}{h_N^\alpha (k-\alpha-1)!}, \quad \text{for any } j.
\]
Proof. We fix $k$ and we do induction on $\alpha$. Let’s start with $\alpha = 1$

\[ \left| \frac{d}{d\theta} B_{j,k,t}(\theta) \right| = \left| (k - 1) \left( \frac{B_{j,k-1,t}(\theta)}{t_{j+k-1} - t_j} - \frac{B_{j+1,k-1,t}(\theta)}{t_{j+k} - t_{j+1}} \right) \right| \]

\[ \leq (k - 1) \left( \frac{1}{t_{j+k-1} - t_j} + \frac{1}{t_{j+k} - t_{j+1}} \right) \]

\[ \leq (k - 1) \frac{2}{h_N} \]

\[ = \frac{2}{h_N} (k - 1)! \]

Thus the inequality holds for $\alpha = 1$.

Now we assume it holds for $\alpha = n$ and we will show it holds for $\alpha = n + 1$.

\[ \left| \frac{d^{n+1}}{d\theta^{n+1}} B_{j,k,t}(\theta) \right| = \left| \frac{d^n}{d\theta^n} (k - 1) \left( \frac{B_{j,k-1,t}(\theta)}{t_{j+k-1} - t_j} - \frac{B_{j+1,k-1,t}(\theta)}{t_{j+k} - t_{j+1}} \right) \right| \]

\[ = (k - 1) \left( \frac{d^n}{d\theta^n} B_{j,k-1,t}(\theta) \frac{d^n}{d\theta^n} B_{j,k-1,t}(\theta) \right) \]

\[ \leq (k - 1) \left( \frac{1}{t_{j+k-1} - t_j} \frac{2^n}{h_N} (k - 2)! + \frac{1}{t_{j+k} - t_{j+1}} \frac{2^n}{h_N} (k - n - 2)! \right) \]

\[ \leq (k - 1)! \left( \frac{1}{(k - n - 2)!} \left( \frac{2^n}{h_N h_N} + \frac{1}{h_N h_N} \right) \right) \]

\[ = \frac{2^{n+1}}{h_N^{n+1}} \frac{(k - 1)!}{(k - (n + 1) - 1)!}. \]

This concludes the proof. ■

Remark F.3. Considering Remark E.8, the bound in Lemma F.2 can be extended to hold on the closed interval $[t_j, t_{j+k}]$ assuming the knots $t_j, \ldots, t_{j+k}$ are simple. Also, it is clear that we need to utilize at least parabolic B-splines in order to have a bound on a continuous derivative.
In this part we will present the method for constructing logspline density estimators using B-splines. Let \( p \) be a continuous probability density function supported on an interval \([a, b]\). Suppose \( p \) is unknown and we would like to construct density estimators for this function. The methodology is as follows

**Definition G.1.** Let \( T_N = (t_i)_{i=0}^N, N \in \mathbb{N} \), be a sequence of knots such that \( t_0 = \cdots = t_{k-1} = a \) and \( t_{N-k+1} = \cdots = t_N = b \), where \( 1 \leq k \leq N \), \( k \) fixed. Thus, the set of splines \( S_{k,T_N} \) of order \( k \) generated by the B-splines \( B_{j,k,T_N} \) can be obtained. We suppress the parameters \( k, T_N \) and just write \( B_j \) instead of \( B_{j,k,T_N} \). Define the spline function

\[
B(\theta; y) = \sum_{j=0}^L y_j B_j(\theta), \quad y = (y_0, \ldots, y_L) \in \mathbb{R}^{L+1} \quad \text{with } L := N - k. \tag{G.1}
\]

and for each \( y \) we set the probability density function

\[
f(\theta; y) = \exp \left( \sum_{j=0}^L y_j B_j(\theta) - c(y) \right) = \exp \left( B(\theta; y) - c(y) \right), \tag{G.2}
\]

where \( c(y) = \log \left( \int_a^b \exp \left( \sum_{j=0}^L y_j B_j(\theta) \right) d\theta \right) < \infty \).

The family of exponential densities \( \{f(\theta; y) : y \in \mathbb{R}^{L+1}\} \) is not identifiable since if \( \beta \) is any constant, then \( c((y_0 + \beta, \ldots, y_L + \beta)) = c(y) + \beta \) and thus

\[
f(\theta; (y_0 + \beta, \ldots, y_L + \beta)) = f(\theta; y)
\]

To make the family identifiable we restrict the vectors \( y \) to the set

\[
Y_0 = \left\{ y \in \mathbb{R}^{L+1} : \sum_{i=0}^L y_i = 0 \right\}. \tag{G.3}
\]

**Remark G.2.** \( Y_0 \) depends only on the number of knots and the order of the B-splines and not the number of samples.
Definition G.3. We define the logspline model as the family of estimators

\[ \mathcal{L} = \{ f(\theta; y) \text{ given by (G.2)} : y \in Y_0 \} \tag{G.4} \]

For any \( f \in \mathcal{L} \)

\[ \log (f) = \sum_{j=0}^{L} y_j B_j(\theta) - c(y) \in S_{k,T_N}. \tag{G.5} \]

Next, let us pick a set of independent, identically distributed random variables

\[ \Theta_n = (\theta_1, \theta_2, \ldots, \theta_n) \in \mathbb{R}^n, \ n \in \mathbb{N} \]

where each \( \theta_i \) is drawn from a distribution that has density \( p(\theta) \).

We next define the log-likelihood function \( l_n : \mathbb{R}^{L+1+n} \rightarrow \mathbb{R} \) corresponding to the logspline model by

\[ l_n(y) = l_n(y; \theta_1, \theta, \ldots, \theta_n) = l_n(y; \Theta_n) \]

\[ = \sum_{i=1}^{n} \log(f(\theta_i; y)) = \sum_{i=1}^{n} \left( \sum_{j=0}^{L} y_j B_j(\theta_i) \right) - nc(y), \ y \in Y_0 \tag{G.6} \]

and the maximizer of the log-likelihood \( l_n(y) \) by

\[ \hat{y}_n = \hat{y}_n(\theta_1, \ldots, \theta_n) = \arg \max_{y \in Y_0} l_n(y) \tag{G.7} \]

whenever this random variable exists. The density \( f(\cdot; \hat{y}_n) \) is called the logspline density estimate of \( p \).

We define the expected log-likelihood function \( \lambda_n(y) \) by

\[ \lambda_n(y) = E[l(y; \theta_1, \ldots, \theta_n)] = \frac{1}{n} \left( -c(y) + \int_a^b \left( \sum_{j=0}^{L} y_j B_j(\theta) \right) p(\theta) \, d\theta \right) < \infty, \tag{G.8} \]

for \( y \in Y_0 \). It follows by a convexity argument that the expected log-likelihood function has a unique maximizing value

\[ \bar{y} = \arg \max_{y \in Y_0} \lambda_n(y) = \arg \max_{y \in Y_0} \frac{\lambda_n(y)}{n} \tag{G.9} \]
which is independent of $n$ but depends on the knots.

Note that the function $\lambda_n(y)$ is bounded above and goes to $-\infty$ as $|y| \to \infty$ within $Y_0$ and therefore, due to Jensen’s Inequality, the constant $\tilde{y}$ is finite; see Stone [24]. The estimator $\hat{y}(\theta_1, \ldots, \theta_n)$, in general does not exist. This motivates us to define the set

$$
\Omega_n = \left\{ \omega \in \Omega : \hat{y} = \hat{y}(\theta_1, \ldots, \theta_n) \in \mathbb{R}^{L+1} \text{ exists} \right\}.
$$

(G.10)

In the next appendix we will show that $\mathbb{P}(\Omega_n) \to 1$ as $n \to \infty$. We also note that due to convexity of $l_n(y)$ and $\lambda_n(y)$ the estimators $\hat{y}$ and $\tilde{y}$ are unique whenever they exist.

We define the logspline estimator $\hat{p}$ of $p$ on the space $\Omega_n$ by

$$
\hat{p} : \mathbb{R} \times \Omega_n \text{ defined by } \hat{p}(\theta, \omega) = f(\theta, \hat{y}(\theta_1, \ldots, \theta_n)), \omega \in \Omega_n
$$

(G.11)

and define the function

$$
\bar{p}(\theta) := f(\theta, \tilde{y}).
$$

(G.12)

**Remark G.4.** In order for the maximum likelihood estimates to be reliable, we require that the modeling error tend to 0 as $n \to \infty$. To this end we state the following hypothesis, which is used in a more general way as hypothesis $(H4)$ in Chapter 6.

$(H1')$ $L = L(n)$ where $n$ is the number of samples and $L$ is as in (G.1). To ensure that the rates of convergence are accurate, we require

$$
\lim_{n \to \infty} L(n) = \infty \quad \text{such that} \quad \lim_{n \to \infty} \frac{L(n) + 1}{n^{1/2 - \beta}} = 0, \quad 0 < \beta < \frac{1}{2}.
$$

(G.13)

So, the above limit suggests that we must have a higher number of samples compared to the number of knots used to construct the logspline family.
It is a well known fact that continuous functions can be approximated by polynomials. Now that we have defined the set of splines $S_{k,t}$ in Definition E.10 and from what we have stated in Remark E.11, that $\bigcup_{N \in \mathbb{N}} S_{k,T_N}$ is dense in the space of continuous functions, there is a question that arises at this point:

Given an arbitrary continuous function $g$ on $[a,b]$, an integer $k \geq 1$ and a set of knots $T_N = (t_i)_{i=0}^{N}$ as in Remark E.11, how close is $g$ to the set $S_{k,T_N}$ of splines of order $k$?

Let’s state this question in a slightly different way. What we would like to do is find a bound for the sup-norm distance between $g \in C[a,b]$ and $S_{k,T_N}$, where this distance is denoted by $\text{dist}(g, S_{k,T_N})$ and is defined as

$$\text{dist}(g, S_{k,T_N}) = \inf_{s \in S_{k,T_N}} \|g - s\|_{\infty}, \quad g \in C[a,b].$$

(H.1)

The answer to our question is given by Jackson’s Theorem found in de Boor [3]. To state it we first need the following definition.

**Definition H.1.** The modulus of continuity $\omega(g; h)$ of some function $g \in C[a,b]$ for some positive number $h$ is defined as

$$\omega(g; h) = \max \{|g(\theta_1) - g(\theta_2)| : \theta_1, \theta_2 \in [a,b], |\theta_1 - \theta_2| \leq h\}.$$  

(H.2)

The bound given by Jackson’s Theorem contains the modulus of continuity of the function whose sup-norm distance we want to estimate from the set of splines. The theorem is stated below.

**Theorem H.2.** Let $T_N = (t_i)_{i=0}^{N}$, $N \in \mathbb{N}$, be a sequence of knots such that $t_0 = \cdots = t_{k-1} = a$ and $b = t_{N-k+1} = \cdots = t_N$, where $1 \leq k \leq N$. Let $S_{k,T_N}$ be the set of splines as in Definition E.10 for the knot sequence $T_N$. For each $j \in \{0, \ldots, k-1\}$, there exists $C = C(k,j)$ such that for $g \in C^j[a,b]$

$$\text{dist}(g, S_{k,T_N}) \leq C h^j \omega \left( \frac{d^j g}{d\theta^j}; |t| \right), \quad \text{where} \quad h = \max_i |t_{i+1} - t_i|.$$  

(H.3)
In particular, from the Mean Value Theorem it follows

\[
\text{dist}(g, S_k, T_N) \leq C h^{j+1} \left\| \frac{d^{j+1}g}{d\theta^{j+1}} \right\|_{\infty}
\]  

(H.4)
in the case that \( g \in C^{j+1}[a, b] \).

**Remark H.3.** Please note that for the approximation the mesh size enters into the bound in (H.4), which dictates the placement for the knots.

Jackson’s Theorem supplies us with an estimate of how good an approximation is contained in the space of splines for a continuous function. However, we are interested in estimates for probability densities, especially since the focus is on logspline density estimates. At this point let’s state results specifically for densities. The following can be found in Stone [23].

Suppose that \( p \) is a continuous probability density supported on some interval \([a, b]\), similar to the set-up when we defined the logspline density estimation method. Define the family \( F_p \) of densities such that

\[
F_p = \left\{ p_{\alpha} : p_{\alpha}(x) = \frac{(p(x))^{\alpha}}{\int (p(y))^{\alpha} \, dy}, \, 0 \leq \alpha \leq 1 \right\}.
\]

(H.5)

It is easy to see that for \( \alpha \in [0, 1] \), \( p_{\alpha} \) is a probability density on \([a,b]\). An interesting consequence from this family is the following

**Lemma H.4.** We define the family of functions

\[
F_p^{\log} = \{ \log(u) : u \in F_p \}.
\]

(H.6)

Then, \( F_p^{\log} \) defines a family of functions that is equicontinuous on the set \( \{ \theta : p(\theta) > 0 \} \).

**Proof.** The proof is simple enough. Pick \( \epsilon > 0 \). There exists \( \delta > 0 \) such that \( |\log(p(x)) - \log(p(y))| < \epsilon \) whenever \( |x - y| < \delta \). Pick any \( \alpha \in [0, 1] \).

If \( \alpha = 0 \) then \( p_0 \) is just a constant and thus \( |\log(p_0(x)) - \log(p_0(y))| = 0 < \epsilon \).

If \( 0 < \alpha < 1 \), then

\[
|\log(p_{\alpha}(x)) - \log(p_{\alpha}(y))| = |\alpha \log(p(x)) - \alpha \log(p(y))| < \alpha \epsilon < \epsilon.
\]

**Remark H.5.** It is practical to work with \( p(x) > 0 \) on the set \([a, b]\) and this is what we assume. In this case, \( \log(p) \in C[a, b] \).
Remark H.6. We will be using the notation $\bar{h} = \max_i |t_{i+1} - t_i|$ and $\underline{h} = \min_i |t_{i+1} - t_i|$, and $\gamma(T_N) = \bar{h}/\underline{h}$.

We can apply the logspline estimation method to $p$. Let $\bar{p}$ be defined as in (G.12), the density estimate given by maximizing the expected log-likelihood. We then have the following lemma:

Lemma H.7. Suppose $p$ is an unknown continuous density function supported on $[a, b]$ and $\bar{p}$ is as in (G.12). Then there exists constant $M' = M'(F_p, k, \gamma(T_N))$ that depends on the family $F_p$, order $k$ and global mesh ratio $\gamma(T_N)$ of $S_{k,T_N}$ such that

$$\| \log (p) - \log (\bar{p}) \|_\infty \leq M' \operatorname{dist}(\log(p), S_{k,T_N})$$

(H.7)

and therefore

$$\| p - \bar{p} \|_\infty \leq \left( \exp\{M' \operatorname{dist}(\log(p), S_{k,T_N})\} - 1\right) \| p \|_\infty.$$  

(H.8)

Moreover, if $\log(p) \in C^{j+1}([a, b])$ for some $j \in \{0, \ldots, k - 1\}$ then by Jackson’s Theorem we obtain

$$\| \log (p) - \log (\bar{p}) \|_\infty \leq M' C(k, j) \bar{h}^{j+1} \left\| \frac{d^{j+1} \log(p)}{d\theta^{j+1}} \right\|_\infty$$

(H.9)

$$\| p - \bar{p} \|_\infty \leq \left( \exp \left\{ M' C(k, j) \bar{h}^{j+1} \left\| \frac{d^{j+1} \log(p)}{d\theta^{j+1}} \right\|_\infty \right\} - 1 \right) \| p \|_\infty.$$  

Remark H.8. Please note that the constant $M$ does not depend on the dimension of $S_{k,T_N}$. For all practical purposes, we will be using uniformly placed knots, thus suppressing the dependence on $\gamma(T_N)$, which will be equal to the constant 1.

Now we will present certain error bounds required in the calculations involving MISE in Chapters 7 and 8. Assume $p, \hat{p}$ and $\bar{p}$ as in Appendix G. Also, assume that $n$ is the number of random samples drawn from $p$.

We will state a series of definitions and theorems that encompass the results from Lemma 5, Lemma 6, Lemma 7, and Lemma 8 in the work of Stone[24][pp.728-729].

Definition H.9. Let $n \geq 1$ and $b > 0$. Let $y \in Y_0$. Let $l_n$ and $\lambda_n$ be defined by (G.6) and (G.8),
respectively. We define

\[ A_{n,b}(y) = \left\{ \omega \in \Omega : |l(y; \Theta_n(\omega)) - l(\bar{y}; \Theta_n(\omega)) - (\lambda_n(y) - \lambda_n(\bar{y}))| < nb\left( \int |\log(f(\theta; y)) - \log(f(\theta; \bar{y}))|^2 d\theta \right)^{1/2} \right\}, \]  

where \( f \) is defined in (G.2) as a function in the logspline family.

**Definition H.10.** Given \( n \geq 1 \) and \( 0 < \epsilon \) we define \( E_{\epsilon,n} \) to be the subset of \( \mathcal{F} = \{ f(\cdot; y) : y \in Y_0 \} \) such that

\[ E_{\epsilon,n} = \left\{ f(\cdot; y) : y \in Y_0 \text{ and } \left( \int |\log(f(\theta; y)) - \log(f(\theta; \bar{y}))|^2 d\theta \right)^{1/2} \leq n^{-\epsilon} \sqrt{\frac{L + 1}{n}} \right\}. \]

**Lemma H.11** (Stone[24][p.728]). For each \( y_1, y_2 \in Y_0 \) and \( \omega \in \Omega \) we have

\[ |l(y_1; \Theta_n(\omega)) - l(y_2; \Theta_n(\omega)) - (\lambda_n(y_1) - \lambda_n(y_2))| \leq 2n\| \log f(\cdot; y_1) - \log f(\cdot; y_2) \|_{\infty}. \]

**Lemma H.12** (Stone[24][p.729]). Let \( n \geq 1 \). Given \( \epsilon > 0 \) and \( \delta > 0 \), there exists an integer \( N = N(n) > 0 \) and sets \( E_j \subset \mathcal{F}, j = 1, \ldots, N \) satisfying

\[ \sup_{f_1, f_2 \in E_j} \| \log f_1 - \log f_2 \|_{\infty} \leq \delta n^{2\epsilon-1}(L + 1) \]

such that \( E_{\epsilon,n} \subset \bigcup_{i=1}^N E_i \).

Combining the above lemmas it leads to the following theorem, which is a result outlined in Lemmas 5 and 8 found in Stone[24].

**Theorem H.13.** Given \( D > 0 \) and \( \epsilon > 0 \), let \( b_n = \frac{n^\epsilon}{\sqrt{n}} \) for each \( n \geq 1 \), and \( 0 < \epsilon < \frac{1}{2} \) and \( \beta = \epsilon \) in (H1'). There exists \( N = N(D) \) such that for all \( n > N \)

\[ A_{n,b_n}(y) \subset \Omega_n \text{ for each } y \in Y_0 \]

(H.13)
and thus
\[ P(\Omega_n) \leq P(A_{n,b_n}(y)) \leq 2e^{-n2e(L+1)\delta(D)}. \] (H.14)

**Remark H.14.** From (H.14) we can see that as number of samples goes to infinity, we have that
\[ P(\Omega_n) \to 1 \text{ as } n \to \infty. \]

**Remark H.15.** The bound (H.14) presented in Theorem H.13 is a consequence of Hoeffding’s inequality which states that for any \( t > 0 \)
\[
P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}X_1 \right| \geq t \right) \leq 2 \exp \left( - \frac{2n^2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)
\]
where \( X_1, \ldots, X_n \) are identically distributed independent random variables with \( P(X_1 \in [a_i, b_i]) = 1 \). To get the bound (H.14) one needs to choose
\[
t = b \left( \int |\log(f(\theta; y)) - \log(f(\theta; \bar{y}))|^2 \, d\theta \right)^{\frac{1}{2}}.
\]

Now that we have defined the set where \( \hat{y} \) exists and showed that the probability of its complement vanishes as \( n \to \infty \) with a specific exponential rate, we will now state certain rates of convergence that only apply on \( \Omega_n \). The following theorem contains results of Theorem 2 and Lemma 12 of Stone[24].

**Theorem H.16.** There exist constants \( M_1, M_2, M_3 \) such that for all \( \omega \in \Omega_n \)
\[
|\hat{y}(\theta_1(\omega), \ldots, \theta_n(\omega)) - \bar{y}| \leq \frac{M_1(L + 1)}{\sqrt{n}},
\]
\[
\|\hat{p}(\cdot, \omega) - \bar{p}(\cdot)\|_2 \leq M_2 \sqrt{\frac{L + 1}{n}},
\] (H.15)
\[
\|\log(\hat{p}(\cdot, \omega)) - \log(\bar{p}(\cdot))\|_{\infty} \leq \frac{M_3(L + 1)}{\sqrt{n}}.
\]
APPENDIX I
LAGRANGE INTERPOLATION

The following two theorems are well-known facts which we cite from [1, p.132, p.134].

**Theorem I.1.** Let \( f : [a, b] \to \mathbb{R} \). Given distinct points \( a = x_0 < x_1 < \ldots < x_l = b \) and \( l + 1 \) ordinates \( y_i = f(x_i), \ i = 0, \ldots, l \) there exists an interpolating polynomial \( q(x) \) of degree at most \( l \) such that \( f(x_i) = q(x_i), \ i = 0, \ldots, l \). This polynomial \( q(x) \) is unique among the set of all polynomials of degree at most \( l \). Moreover, \( q(x) \) is called the Lagrange interpolating polynomial of \( f \) and can be written in the explicit form

\[
q(x) = \sum_{i=0}^{l} y_i l_i(x) \quad \text{with} \quad l_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}, \ i = 0, 1, \ldots, l. \tag{I.1}
\]

**Theorem I.2.** Suppose that \( f : [a, b] \to \mathbb{R} \) has \( l + 1 \) continuous derivatives on \((a, b)\). Let \( a = x_0 < x_1 < \ldots < x_l = b \) and \( y_i = f(x_i), \ i = 0, \ldots, l \). Let \( q(x) \) be the Lagrange interpolating polynomial of \( f \) given by formula (I.1). Then for every \( x \in [a, b] \) there exists \( \xi \in (a, b) \) such that

\[
f(x) - q(x) = \frac{\prod_{i=0}^{l} (x - x_i)}{(l + 1)!} f^{(l+1)}(\xi).
\tag{I.2}
\]

We next prove an elementary lemma that provides the estimate of the interpolation error when information on the derivatives of \( f \) is available. This lemma is used in Theorem 8.9 to compute the bound for the mean integrated squared error.

**Lemma I.3.** Let \( f(x), q(x), \) and \( (x_i, y_i), \ i = 0, \ldots, l, \) with \( l \geq 1 \), be as in Theorem I.2. Suppose that

\[
\sup_{x \in [a, b]} |f^{(l+1)}(x)| \leq C
\]

for some constant \( C \geq 0 \) and \( x_{i+1} - x_i = \frac{b - a}{l} =: \Delta x \) for each \( i = 0, \ldots, l - 1 \). Then

\[
\max_{x \in [a, b]} |f(x) - q(x)| \leq C \frac{(\Delta x)^{l+1}}{4(l + 1)}. \tag{I.3}
\]
Proof. Let \( x \in [a, b] \). Then \( x \in [x_j, x_{j+1}] \) for some \( j \in \{0, \ldots, l-1\} \). Observe that

\[
|(x - x_j)(x - x_{j+1})| \leq \frac{1}{4} (\Delta x)^2
\]

and for \( m \in \{-j, -j+1, \ldots, -1\} \cup \{2, \ldots, l - j\} \) we have \( |x - x_{j+m}| \leq (\Delta x)|m| \). From this it follows that

\[
\prod_{i=0}^{l} |x - x_i| \leq \frac{(\Delta x)^{(l+1)}}{4} j!(l - j)! \leq \frac{(\Delta x)^{(l+1)}l!}{4}.
\]

Then Theorem I.2 together with the above estimate implies (I.3). \(\blacksquare\)
REFERENCES


