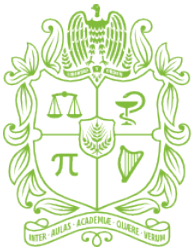
El algoritmo Metropolis-Hastings: problemas inversos, calibración de modelos y el modelo TOMGRO

The Metropolis-Hastings Algorithm: inverse problems, model calibration, and the TOMGRO model

Carlos Enrique Nosa Guzman



UNIVERSIDAD NACIONAL DE COLOMBIA

Advisors: PhD Juan Galvis
PhD Rodrigo Gil Castañeda
Collaborators: PhD Francisco Gomez
PhD Freddy Hernandez
Department of Mathematics. Faculty of Sciences
Universidad Nacional de Colombia

March, 2025

El algoritmo Metropolis-Hastings: problemas inversos, calibración de modelos y el modelo TOMGRO

Abstract

Este trabajo investigativo aborda el estudio y análisis del algoritmo Metropolis-Hastings (MH), un método numérico perteneciente a la familia de algoritmos Markov Chain Monte Carlo (MCMC), ampliamente utilizado en la inferencia bayesiana. Se exploran sus fundamentos teóricos y aplicaciones en problemas inversos, destacando su capacidad para estimar parámetros desconocidos y capturar incertidumbres en modelos matemáticos. Además, se analizan las dificultades asociadas a su implementación, como la selección de distribuciones a priori y la evaluación de la convergencia. Este enfoque se ilustra con ejemplos prácticos que muestran la eficacia del algoritmo en el contexto de la calibración de modelos, teniendo como principal objetivo la calibración del modelo TOMGRO, un sistema dinámico destinado al estudio del crecimiento y desarrollo de una planta de tomate.

The Metropolis-Hastings Algorithm: inverse problems, model calibration, and the TOMGRO model

Abstract

This research work addresses the study and analysis of the Metropolis-Hastings (MH) algorithm, a numerical method belonging to the family of Markov Chain Monte Carlo (MCMC) algorithms, widely used in Bayesian inference. Its theoretical foundations and applications in inverse problems are explored, highlighting its ability to estimate unknown parameters and capture uncertainties in mathematical models. Additionally, the challenges associated with its implementation, such as the selection of prior distributions and the evaluation of convergence, are analyzed. This approach is illustrated with practical examples that demonstrate the algorithm's effectiveness in the context of model calibration, with a primary focus on calibrating the TOMGRO model, a dynamic system designed for studying the growth and development of a tomato plant.

Advisors: PhD Juan Galvis, PhD Rodrigo Gil Castañeda

Contents

List of Figures

List of Tables

Nomenclature

MH Metropolis Hastings

MCMC Markov Chain Monte Carlo

ODE Ordinary differential equation

CDF Cumulative distribution function

PDF Probability density function

CLT Central limit theorem

TOMGRO Tomato growth

Chapter 1

Introduction

Inverse problems constitute a fundamental area of research in contemporary science due to their relevance in various disciplines, including the exact and natural sciences (consult [10]). These problems are characterized by their main objective: to reconstruct, infer or estimate unknown information from available observations, thus allowing a more comprehensive understanding of the phenomena studied (as can be found in [5,10]) . Their application is almost ubiquitous, from the reconstruction of medical images to the interpretation of geophysical signals or the modeling of climate systems. However, its solution presents multiple challenges, including the inherent uncertainty in the available data, the sensitivity to perturbations in the initial conditions and the numerical stability of the methods used. Addressing these problems requires not only a solid theoretical framework, but also robust and versatile computational strategies (see [1,10,11]).

This research work focuses on the study of the Metropolis-Hastings (MH) algorithm, one of the most prominent methods within the Markov Chain Monte Carlo (MCMC) approach for sampling probability distributions. This algorithm, in particular, has proven to be a powerful tool for addressing the challenges associated with inverse problems. Its ability to capture uncertainty in parameter estimation makes it an essential resource for the calibration of mathematical models, especially in contexts where the available information is limited or in the presence of measurement error (consult [5,10]). In this framework, the present work explores how the Metropolis-Hastings algorithm allows improving the accuracy and reliability of the estimates, while providing a solid theoretical framework that supports the validity of the results obtained.

Beyond its theoretical importance, the Metropolis-Hastings algorithm finds applications across various domains. These include classical techniques like least

squares for data fitting and naive optimization methods for parameter estimation. Its utility extends to parameter estimation in ordinary differential equations (ODEs), where it enhances the precision of model calibration (see [5,6,7,10]).

An important application of this approach is in the calibration of the TOMGRO model, a mathematical model designed to simulate tomato crop growth and production. This model combines ordinary differential equations with observational data to realistically represent the state of the plants and their evolution over time. The implementation of the Metropolis-Hastings algorithm in this context not only allows to adjust key parameters of the model, but also to validate its predictive capacity and its usefulness as a tool for decision making in agriculture. This case study not only illustrates the applicability of the method, but also highlights its potential for addressing inverse problems in complex systems.

In order to provide a comprehensive framework, this paper sets as main objectives: to review the theoretical foundations of MCMC methods, to study the Metropolis-Hastings algorithm in depth, to analyze variants such as the Metropolis-Hastings Random Walk and the t-walk, and to explore its applications in the calibration of mathematical models through the estimation of scalar and functional parameters. In addition, the fundamental concepts of dynamical systems necessary to understand the TOMGRO model are included, thus laying the groundwork for effective implementation and proper interpretation of the results.

The structure of the thesis is designed to guide the reader from fundamentals to practical applications. In Chapter 2, the basic concepts necessary to understand inverse problems, dynamical systems and Markov chains are presented. Chapter 3 introduces a statistical approach to inverse problems and describes the Bayesian framework underlying the Metropolis-Hastings algorithm. In Chapter 4, MCMC methods are explained and the operation of the Metropolis-Hastings algorithm is detailed, along with some of its variants such as Gibbs sampler y the Two-walk algorithm. Chapter 5 is devoted to practical applications, with emphasis on the calibration of the TOMGRO model. Finally, in section 6, conclusions are presented and possible lines of future work are discussed. Additionally, the document includes a list of figures, tables, nomenclature and appendices that provide technical and theoretical support for the topics discussed. The ***repository*** of this thesis is composed of several notebooks in the Julia programming language and in the Pluto or Jupyter environment and all the graphics created are extracted from such notebooks (please refer to the reference [21]).

Chapter 2

Fundamental concepts

In this chapter a short review of fundamental concepts related to the topic of study is made. In particular, the definitions and results of inverse problems, deterministic dynamical systems (ordinary differential equations) and Markov chains are reviewed. This has the purpose of laying the theoretical foundations necessary to address the main topic of this thesis. The material presented here is extracted from [10].

2.1 Inverse problems

Inverse problems occur when the objective is to determine the underlying causes or parameters of a system based on observed data, essentially working backwards from the outcomes to the model. These problems are often ill-posed, meaning that solutions may not exist, may not be unique, or may not depend continuously on the data. This section begins by exploring classical regularization methods, which offer stable approximations when exact solutions are not feasible (this section is inspired in the first chapter of [10]).

In this thesis, the following definition of *inverse problems* will be used.

Definition 2.1. (*Inverse problem*, [5,10]) A inverse problem is one where an effect can be measured, and the goal is to determine a cause. These problems arise when one seeks to infer quantities of interest that cannot be measured directly in physical systems.

Now, given the definition of an inverse problem, it is important to highlight when a problem is well-posed. According to Hadamard, a problem is considered well-posed if, for any given data:

1. A solution exists.
2. The solution is unique.
3. The solution depends continuously on the data.

These conditions are crucial in ensuring that the problem is both theoretically sound and practically solvable, particularly where stability and uniqueness of the solution are key factors. A problem that does not satisfy any of Hadamard's conditions is called **ill-posed**. The mathematical formulation of an inverse problem typically leads to an ill-posed problem.

In a mathematical formulation, inverse problems are typically expressed as either linear or nonlinear equation in the following form

$$y = F(x)$$

where y corresponds to the measured or observable quantities, F is the operator that usually represents the model and x are the unknown quantities of interest, also called parameters. According to the Hadammard's conditions for well-posed problems F has to be onto (existence), one-to-one (uniqueness) and F^{-1} must exist and be continuous (stability).

To present the classical and statistical approaches to inverse problems, a relevant example will be used, where the advantages and disadvantages of all the methods addressed will be discussed. One of the objectives of applying different methods to the same example is the ability to compare them.

Example 2.2. (*Inverse linear problem*) Consider the matrix A in $\mathbb{R}^{3\times 3}$ given by

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.01 \end{pmatrix}.$$

The inverse linear problem is to find a vector $x \in \mathbb{R}^3$ such that

$$y = Ax \tag{2.1}$$

where $y \in \mathbb{R}^3$ is given by $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. It is not difficult to see that for any matrix A:

1. the solution exists if and only if y belongs to the range of the matrix A, and
2. the solution is unique if and only if the kernel of matrix A only has the null vector as element.

The previous conditions must be satisfied in order to ensure the problem has a unique solution. The requirements over the kernel and range of the matrix does not assure a useful solution. From the practical point of view the vector y typically represents the data which is contaminated by errors, then the equation (2.1) is approximated by $Ax \approx y$ and, although the inverse of the matriz A exists, small errors in y may cause errors of arbitrary size in x, in the solution to the inverse problems.

The exact inverse matrix of A is

$$A^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 100 \end{pmatrix}.$$

Then,

$$x = A^{-1}y = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 100 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ 10y_2 \\ 100y_3 \end{pmatrix},$$

this indicates that, if y_3 is contaminated by an additive error then x_3 will be perturbed by an error 100 times larger than the error of y_3. □

2.1.1 Classical regularization methods

The basic idea behind regularization methods is that, instead of attempting to solve equation (2.1) exactly, the goal is to find a nearby problem that is uniquely solvable. This approach ensures robustness, meaning that is expected that small errors in the data do not excessively corrupt the approximate solution.

2.1.1.1 Truncated singular value decomposition (TSVD)

Let $A \in \mathbb{R}^{m\times n}$, $A \neq 0$ be a real matrix defining a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ and consider the matrix equation $Ax = y$. The starting point in this part is the following proposition.

Proposition 2.3. Let A be as above, and let $A^\top$ be the transpose matrix of A. Then

1. The spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, allow orthogonal decompositions:

$$\mathbb{R}^n = \ker(A) \oplus \ker(A)^\perp = \ker(A) \oplus \mathrm{Ran}(A^\top),$$

$$\mathbb{R}^m = \mathrm{Ran}(A) \oplus \mathrm{Ran}(A)^\perp = \mathrm{Ran}(A) \oplus \ker(A^\top).$$

2. There exist orthonormal sets of vectors $(v_k) \in \mathbb{R}^n$, $(u_k) \in \mathbb{R}^m$, and a finite sequence (σ_k) of positive numbers, $\sigma_k > 0$, such that

$$\mathrm{Ran}(A) = \mathrm{span}\{u_k \mid k \in \{1, \ldots, p\}\},$$

$$\ker(A)^\perp = \mathrm{span}\{v_k \mid k \in \{1, \ldots, p\}\},$$

   and the operator A can be represented as

$$Ax = \sum_{k=1}^{p} \sigma_k \langle x, v_k \rangle u_k,$$

   where p is the range of the matrix A. The system (v_k, u_k, σ_k) is called the singular system of the matrix A and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{R}^{\cdot}$.

3. The equation $Ax = y$ has a solution if and only if

$$y = \sum_{k=1}^{p} \langle y, u_k \rangle u_k.$$

   In this case, a solution is of the form

$$x = x_0 + \sum_{k=1}^{p} \frac{1}{\sigma_k} \langle y, u_k \rangle v_k,$$

   where $x_0 \in \ker(A)$ can be chosen arbitrarily.

The previous result leads to the **singular value decomposition** of the matrix A, that is,

$$A = U \Sigma V^\top$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrix, that is, $U^\top = U^{-1}$, $V^\top = V^{-1}$ and $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix with diagonal elements

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p > \ldots \geq \sigma_{min\{m,n\}} \geq 0.$$

The representation of the operator A as $U\Sigma V^\top$ allows to remove one of the possible problems in the inverse problem $Ax = y$, that is $y \notin Ran(A)$. Consider P as the orthogonal projection in the range of A defined as

$$P : \mathbb{R}^m \to Ran(A) \subseteq \mathbb{R}^m : y \mapsto \sum_{k=1}^{p} \langle y, u_k \rangle u_k.$$

It follows that for any $x \in \mathbb{R}^n$,

$$\|Ax - y\|^2 = \|Ax - Py\|^2 + \|(1 - P)y\|^2 \geq \|(1 - P)y\|^2$$

due to the orthogonality of $Ax - Py$ and $(1 - P)y$ as is shown in Figure 2.1. The previous inequality means that if y has a nonzero component in the subspace orthogonal to the range of A then the equation $Ax = y$ cannot be satisfied exactly. Therefore, the best adjustment to overcome the nonsolvability is to solve the projected equation

$$Ax = PAx = Py.$$

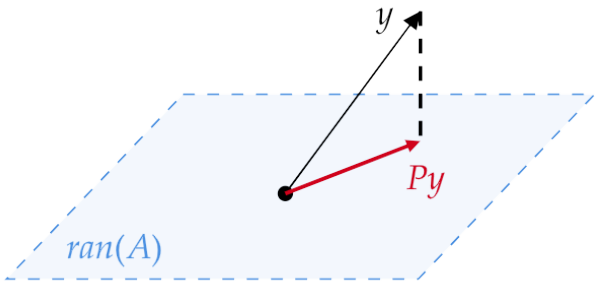


Figure 2.1: Orthogonal projection of a vector y in the range of a matrix A represented as Py. Figure elaborated by the author.

Let P_r denote the *truncated* orthogonal projection defined as

$$P_r : \mathbb{R}^m \to Ran(A) \subseteq \mathbb{R}^m : y \mapsto \sum_{k=1}^{r} \langle y, u_k \rangle u_k,$$

where $1 \leq r \leq p$. Thus, instead of the equation $Ax = Py$ it is considered the truncated projected equation

$$Ax = P_r y.$$

This equation is solvable and the solution lives in the subspace generated by the first r columns of the matrix U and is given as

$$x_r = x_0 + \sum_{k=1}^{r} \frac{1}{\sigma_k} \langle y, u_k \rangle u_k$$

where $x_0 \in \ker(A)$ is chosen arbitrarily, for convenience $x_0 = 0$ (this could be justified chosen the solution x_k that minimizes the norm).

These considerations lead to the following proposition:

Proposition 2.4. ([10]) The problem $Ax = P_r y$ mentioned above has unique solution x_k, called the truncated SVD solution, which is

$$x_r = \sum_{k=1}^{r} \frac{1}{\sigma_k} \langle y, u_k \rangle u_k.$$

When $r = p$, the propositions establishes that

$$x_p = A^+ y,$$

where the matrix A^+ represents the *pseudoinverse* or *Moore-Penrose inverse* of A, and it is defined as $A^+ = V\Sigma^+ U^\top$ where

$$\Sigma^+ = \begin{pmatrix} 1/\sigma_1 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & \dots & 0 \\ & \ddots & & \\ \vdots & & 1/\sigma_p & \vdots \\ & & & \ddots \\ 0 & \dots & & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

There is a natural question that arises in this regularization: How to choose the truncation index r? There exists a cualitative rule that is often referred to as *discrepancy principle* or **Morozov's principle** (see [5,10]). Assume that the data vector y is a noisy approximation of a noiseless vector y_0. While y_0 is unknown, it can be possible to have an estimation of the noise level, i.e., $\|y - y_0\| \approx \epsilon$ for som $\epsilon > 0$. This discrepancy principle states that we cannot expect the approximate solution to yield the smaller residual error than the measurement error, since otherwise the solution would be adjusted to the noise. This principle leads to the following selection criterion for r: choose r, $1 \leq r \leq p$, as the largest index that satisfies $\|y - Ax_k\| = \|y - P_k y\| \leq \epsilon$.

Example 2.5. In the Example 2.2, it can be shown that

- for $r = 1$, $x_1^\top = (y_1\ 0\ 0)$,
- for $r = 2$, $x_2^\top = (y_1\ 10y_2\ 0)$, and,
- for $r = 3$, $x_3^\top = (y_1\ 10y_2\ 100y_3)$.

2.1.1.2 Tikhonov regularization

The previous regularization shows that problems occur when singular values of the operator are near to zero, causing the norm of the approximate solution x_k increases in norm when r becomes larger. The idea of this regularization is to control simultaneously the norm of the residual $Ax - y$ and the norm of the approximate solution x. Let δ be a given positive number. The Tikhonov regularized solution $x_\delta \in \mathbb{R}^n$ is the minimizer of the functional

$$F_\delta(x) = \|Ax - y\|^2 + \delta\|x\|^2,$$

provided that a minimizer exists. The parameter δ is called the regularization parameter.

It is easy to prove see that the functional can be written as

$$F_\delta(x) = \left\| \begin{pmatrix} A \\ \sqrt{\delta} I \end{pmatrix} - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2,$$

therefore, the problem becomes a problem of least squares, and, the minimizer to this kind of problems is given by*

$$x_\delta = (A^\top A + \delta I)^{-1} A^\top y,$$

or in terms of the singular value decomposition of A,

$$x_\delta = \sum_{k=1}^{p} \frac{\sigma_k}{\sigma_k^2 + \delta} \langle y, u_k \rangle v_k.$$

Assume an estimate $\epsilon > 0$ of the norm of the error in the data vector. Consider the function $f : [0, \infty) \to [0, \infty)$ such that $f(\delta) = \|Ax_\delta - y\|$ representing the discrepancy asociated with the parameter δ. The Morozov's principle states that the regularization parameter δ should be chosen from the condition $f(\delta) \leq \epsilon$ if is possible, that is, the regularization solution should no try to satisfy the data more accurately than up the noise level.

Observe in the previous proposition, if $\delta = 0$ the solution given by the *pseudo inverse* A^+ is recover. This highlights how this regularization addresses the issue of singular values close to zero, providing stability to the solution and mitigating the effects of ill-conditioning in the matrix A.

*See section 5.2.

Example 2.6. Differentiation of data as an inverse problem. (*Example from [6]*) One of the simplest ill-posed problems is numerical differentiation of noisy functions, a task that is faced in many applications. Assume that it is needed to compute the derivative of a function which includes additive noise, instead of the exact function f, it is only known a function f^α with

$$f^\alpha(x) = f(x) + \sqrt{2}\alpha \sin(2\pi k x)$$

for some $k \in \mathbb{N}$, $\alpha > 0$, where $\sqrt{2}\alpha \sin(2\pi k x)$ represents the additive noise. Note that the distance in $L^2([0,1])$ is

$$\|f^\alpha - f\|_{L^2([0,1])} = \left(\int_0^1 |f^\alpha(x) - f(x)|^2 dx\right)^{1/2} = \left(\int_0^1 |\sqrt{2}\alpha \sin(2\pi k x)|^2 dx\right)^{1/2} = \alpha.$$

On the other hand,

$$\frac{df^\alpha}{dx}(x) = \frac{df}{dx}(x) + 2\pi k\sqrt{2}\ \alpha cos(2\pi k x),$$

and the $L^2([0,1])$ distance of the derivates is

$$\left\|\frac{df^\alpha}{dx} - \frac{df}{dx}\right\|_{L^2([0,1])} = \left(\int_0^1 |2\pi k\sqrt{2}\ \alpha cos(2\pi k x)|^2 dx\right)^{1/2} = 2\pi k\alpha.$$

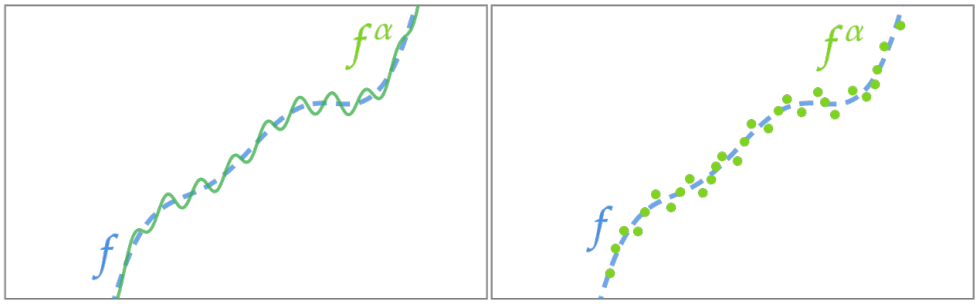


Figure 2.2: Differentiation of data: In the left is shown a data with differentiable noise and in the right is illustrated a data with a non differentiable noise. Figure elaborated by the author and inspired in [6].

Although α could be small or near of zero, the term $2\pi k\alpha$ can be large since it depends on the value of $k \in \mathbb{N}$, that means, despite of the proximity of the functions

f^α and f, the derivatives $\frac{df^\alpha}{dx}$ and $\frac{df}{dx}$ can be far away in the norm $L^2([0,1])$. This demonstrates that differentiation of data is a ill-posed problem because there is no continuity with respect to data of the problem.

In order to solve this inverse problem, one possible option is to replace the direct differentiation $\frac{f^\alpha(x+h)-f^\alpha(x)}{h}$ to a problem derived from Tikhonov regularization:

Find the function f_λ that is the minimizer of the functional
$$\|f_\lambda - f^\alpha\|^2_{L^2([0,1])} + \lambda \|\frac{df_\lambda}{dx}\|^2_{L^2([0,1])}$$

With this regularization is controlled the norm of the derivative and the proximity to the data f^α. The implications of differential treatment of the optimization problem can be found in [6].

2.2 Deterministic dynamical systems

Deterministic dynamical systems provide a powerful framework for modeling and understanding the evolution of processes governed by explicit and predictable rules. Foundational in disciplines such as physics, biology, and engineering, these systems describe how a system's state evolves over time, entirely determined by its current state and governing equations. From differential equations to iterative maps, the study of these systems reveals insights into stability, behavior, and their wide-ranging applications in science and technology. This section is extracted from the first and second chapters of [14].

In this context, inverse problems emerge as a compelling challenge, focusing on uncovering the hidden rules or parameters that drive a system based on observed data. Unlike forward problems, which predict future states from known dynamics, inverse problems reconstruct the systems governing equations from limited or noisy measurements. These problems are critical in real-world scenarios, where direct observation is often impractical, requiring advanced mathematical techniques to overcome their inherent ill-posedness. Together, the study of deterministic dynamical systems and inverse problems forms a bridge between theoretical modeling and practical understanding, highlighting the intricate dance between predictability and discovery (see the second chapter of [10]).

This section explores deterministic dynamical systems through the following key topics: important theorems about the existence and uniqueness of solutions, an introduction to the phase plane as a tool for analyzing one or two dimensional systems, the study of bifurcations to understand qualitative changes in system behavior, and an overview of numerical methods for solving dynamical systems when analytical solutions are unattainable (a more comprenhensive treatment of these topics can be found in [13,14]).

Definition 2.7. (*Dynamical system*) A dynamical system is a tuple (T, M, ψ), where:

- T is the time set, usually taken as $\mathbb{R}$ or $\mathbb{Z}$, representing continuous or discrete time.
- M is the state space, which is a set of all possible states that the system can occupy.
- $\psi : T \times M \to M$ is the evolution map, which describes how the system evolves over time. For each $t \in T$ and $m \in M$, $\psi(t, m)$ gives the state of the system at time t, starting from the initial state m.

2.2.1 Ordinary differential equations

Ordinary differential equations (ODEs) are a pillar in the study of deterministic dynamical systems, providing a mathematical framework for describing the continuous evolution of a system over time. An ODE is an equation involving a function and its derivatives, where the independent variable is typically time. These equations model a vast array of phenomena in science and engineering, from the motion of celestial bodies and the dynamics of chemical reactions to population growth and fluid flow. For a more extensive discussion of this topic, see [13] and [14].

Let Ω be an open set contained in $\mathbb{R} \times \mathbb{R}^n$. A point of $\mathbb{R} \times \mathbb{R}^n$ will be denoted by (t, x), $t \in \mathbb{R}$ and $x = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$. Let $f : \Omega \to \mathbb{R}^n$ be a continuous function and let I be a non degenerated interval, i.e., a conected subset of $\mathbb{R}$ not reduced to a single point.

Definition 2.8. (**Solution to an ODE**, [13]) A differentiable function $\phi : I \to \mathbb{R}^n$ is called a *solution* of the equation

$$\frac{dx}{dt} = f(t, x)$$

in the interval I if:

- $\{(t, \phi(t))|t \in I\}$ is contained in Ω, and,
- $\frac{d\phi}{dt}(t) = f(t, \phi(t))$ for all $t \in I$. If t is an extreme of the interval, then a lateral derivate is replaced in the previous equation.

This work mainly uses the type of equations defined below.

Definition 2.9. ([13]) A equation $\frac{dx}{dt} = f(t, x)$ is called a *ordinary differential equation of first order* and can be abbreviated as $x' = f(t, x)$.

As stated above, an ordinary differential equation can be seen as a dynamical system considering the time set as the interval I, the state spaces as $\mathbb{R}^n$, and finally $\psi = f$ as the evolution map.

Example 2.10. The **exponential growth** model is a fundamental concept in ordinary differential equations, often used to describe the unrestricted growth of populations, capital investments, or other quantities over time. This model is governed by the differential equation:

$$\frac{dx}{dt} = rx,$$

where $x(t)$ represents the quantity of interest at time t, and $r > 0$ is the growth rate. The solution to this equation, given an initial value $x(0) = x_0$, is:

$$x(t) = x_0 e^{rt}.$$

The exponential growth model assumes constant proportional growth, implying that the rate of change of $x(t)$ is directly proportional to its current value. While idealized and simple, this model serves as the basis for more complex systems and has widespread applications in biology, economics, and physics, among other fields.

Example 2.11. (*Example from [13]*) Let $\Omega = \mathbb{R}^2$ and $f(t, x) = 3x^{2/3}$. For all $c \in \mathbb{R}$ a function $\phi_c : \mathbb{R} \to \mathbb{R}$ defined as

$$\phi_c(t) = \begin{cases} (t - c)^3, & t \geq c, \\ 0, & t \leq c, \end{cases}$$

is a solution of the differential equation $x' = 3x^{2/3}$ over $\mathbb{R}$ satisfying both conditions in order to be a solution. This example illustrates the fact that a ordinary differential equation can have an infinity of solutions. □

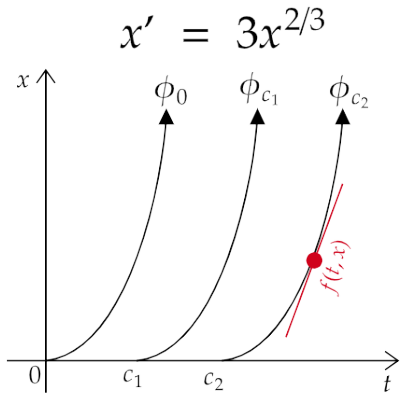


Figure 2.3: Non uniqueness in the solution of a ordinary differential equation. Figure elaborated by the author and inspired in [13].

The next part of this chapter focuses on reviewing what are the sufficient conditions to prove the existence and uniqueness of a solution of a first order ordinary differential equation.

2.2.1.1 Existence and uniqueness of solutions

The proof of existence and uniqueness of a solution of ordinary differential equation is based on the following theorem.

Theorem 2.12. (**Banach fixed-point theorem**, [10,13]) Let (X, d) be a non empty complete metric space with a contraction mapping $T : X \to X$. Then T admits a unique fixed point x^* in X, that is, $T(x^*) = x^*$. Furthermore, x^* is an attractor of T, that is, starting with an arbitrary point $x_0 \in X$ and defining the sequence $x_{n+1} = T(x_n)$ for $n \in \mathbb{N}$, then $\lim_{n\to\infty} x_n = x^*$.

The proof of the previous theorem can be found in the appendixes. The following theorem and demostration is based in [13].

Theorem 2.13. Picard's Theorem. Let $f(t, x)$ be a continuous function and Lipschitz with respect to the second variable x over $\Omega = I_a \times B_b$ where $I_a = \{t : |t - t_0| \leq a\}$ and $B_b = \{x : \|x - x_0\| \leq b\}$. Suppose $|f| \leq M$ in Ω, then, there

exists an unique solution of the Cauchy problem $x' = f(t, x)$, $x(t_0) = x_0$ in I_α where $\alpha = \min\{a, b/M\}$.

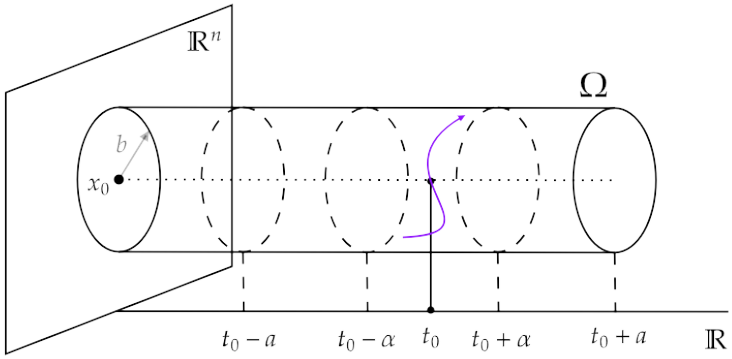


Figure 2.4: Picard's Theorem. Figure elaborated by the autor and inspired in [13].

Proof. (*Proof taken from [13]*) Let $X = C(I_a, B_b)$ be the metric space of continuous function $\phi : I_a \to B_b$ with the metric

$$d(\phi_1, \phi_2) = \sup_{t \in I_a} |\phi_1(t) - \phi_2(t)|.$$

For $\phi \in X$, let $F(\phi) : I_a \to \mathbb{R}$ defined by

$$F(\phi)(t) = x_0 + \int_{t_0}^{t} f(s, \phi(s))ds, \ t \in I_a.$$

The correspondence $\phi \to F(\phi)$ define a function F with the next properties

1. $F(X) \subseteq X$,
2. $F^n = \underbrace{F \circ F \circ \ldots \circ F}_{n \text{ times}}$ is a contraction for n large enough.

Indeed, for all $t \in I_a$,

$$|F(\phi)(t) - x_0| = \left| \int_{t_0}^{t} f(s, \phi(s))ds \right| \leq M\alpha \leq b.$$

This proves (1). With respect to (2), for all $\phi_1, \phi_2 \in X$ and for all $n \geq 0$,

$$|F^n(\phi_1)(t) - F^n(\phi_2)(t)| \leq \frac{K^n |t - t_0|^n}{n!} d(\phi_1, \phi_2), \ t \in I_\alpha,$$

where K is the Lipschitz constant of f. For $n = 0$, since $F^0 = I_X$ is the identity over X then the property is obvious. Suppose that the property is valid to k. Therefore,

$$\begin{aligned}
|F^{k+1}(\phi_1)(t) - F^{k+1}(\phi_2)(t)| &= |F(F^k(\phi_1))(t) - F(F^k(\phi_2))(t)| \\
&\leq \left| \int_{t_0}^{t} |f(s, F^k(\phi_1)(s)) - f(s, F^k(\phi_2)(s))| ds \right| \\
&\leq \left| \int_{t_0}^{t} K |F^k(\phi_1)(s) - F^k(\phi_2)(s)| \right| \\
&\leq K \left| \int_{t_0}^{t} \frac{K^k |s - t_0|^k}{k!} d(\phi_1, \phi_2) \right| \\
&= \frac{K^{k+1} |t - t_0|^{k+1}}{(k+1)!} d(\phi_1, \phi_2).
\end{aligned}$$

Thus, $d(F^n(\phi_1), F^n(\phi_2)) \leq \frac{K^n \alpha^n}{n!} d(\phi_1, \phi_2)$ and for n large enough $\frac{K^n \alpha^n}{n!} < 1$ since $\frac{K^n \alpha^n}{n!}$ goes to 0 when n tends to infinity. By the *Banach fixed-point theorem* this application F admits one and only one $\phi \in X$ such that $F(\phi) = \phi$. By the Fundamental Theorem of Calculus this fixed point belongs to $C^1(I_a, B_b)$, which concludes the proof of the theorem. □

2.2.1.2 Systems

Let Ω be a subset of $\mathbb{R} \times \mathbb{R}^{d_1 \cdot d_1 \cdot \ldots \cdot d_m}$. Consider $f_i : \Omega \to \mathbb{R}^{d_i}$, $i = 1, \ldots, m$ continuous functions and a family $\{\phi_1, \ldots, \phi_m\}$ where each $\phi_i : I \to \mathbb{R}^{d_i}$, $i = 1, \ldots, m$ is a differentiable function over the interval $I \subseteq \mathbb{R}$.

Definition 2.14. It is said that $\phi(t) = (\phi_1(t), \ldots, \phi_m(t))$ is a *solution* of the system of ordinary differential equation

$$\begin{cases}
\frac{dx_1}{dt} &= f_1(t, x_1, \ldots, x_m), \\
\frac{dx_2}{dt} &= f_2(t, x_1, \ldots, x_m), \\
&\vdots \\
\frac{dx_m}{dt} &= f_m(t, x_1, \ldots, x_m),
\end{cases}$$

in the interval I if:

- for all $t \in I$, $(t, \phi(t))$ belongs to Ω, and,
- for all $i = 1, 2, \ldots, m$,

$$\frac{d\phi_i}{dt} = f_2(t, \phi_1(t), \ldots, \phi_m(t)),$$

for all $t \in I$.

Example 2.15. The Lotka–Volterra equations, also known as the predator–prey model, describe the dynamics of two interacting species, one as a predator and the other as prey. This first-order nonlinear system of differential equations is given by

$$\frac{dx}{dt} = \alpha x - \beta xy,$$
$$\frac{dy}{dt} = -\gamma y + \delta xy,$$

where

- $x(t)$ represents the prey population at time t,
- $y(t)$ represents the predator population at time t,
- $\alpha > 0$ is the natural growth rate of the prey in the absence of predators,
- $\beta > 0$ is the predation rate coefficient,
- $\gamma > 0$ is the natural death rate of the predator in the absence of prey,
- $\delta > 0$ is the growth rate of predators per prey consumed.

2.2.1.3 Further analysis

The study of dynamical systems provides a systematic approach to understanding the evolution of systems over time through mathematical modeling. Key concepts such as fixed points, phase planes, and bifurcations play a central role in this qualitative analysis. Fixed points, or equilibrium points, represent states where the system remains unchanged over time, offering insights into the system's stability. The phase plane provides a geometric representation of the trajectories of two-dimensional systems, enabling visualization of behaviors such as oscillations, convergence, or divergence. Bifurcations, on the other hand, describe qualitative changes in the system's dynamics as parameters are varied, illustrating phenomena such as the transition from stability to chaos (consult [14]).

So far, a purely quantitative treatment of ordinary differential equations has been carried out, but here is an example that illustrates a basic technique in the qualitative analysis of dynamical systems: *interpreting a differential equation as a vector field.*

Example 2.16. (*Example from [14]*) Consider the ODE $x' = x^2 - 1$. If x' is plotted against x, the resulting graph is shown in the figure 2.5.

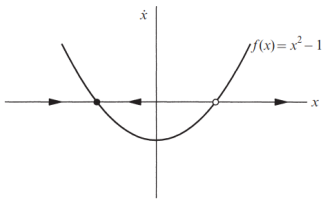


Figure 2.5: Example of qualitative analysis in one dimension differential equation. Figure from [14].

Note that a solution of the ODE is the function $\phi(t) = 1$ for all $t \in \mathbb{R}$; this happens since the function $f(t, x) = x^2 - 1$ is equal to zero when $x = 1$ for all t. The points x^* such that $f(t, x^*) = 0$ for all t are called **equilibrium points** or **fixed points**.

If $f(x) > 0$ over a set $I \subseteq \mathbb{R}$ then the resulting solution of the ODE will increase over I, in the same way, if $f(x) < 0$ then the solution has to be decreasing. In the figure 2.5 these characteristics are represented by arrows pointing to the right if f is positive and to the left if f is negative.

It can be observed that for the point $x = -1$, the vector field points towards it in the surrounding region, indicating that small perturbations around $x = -1$ will evolve and eventually converge at this point; this type of equilibrium is called a **stable equilibrium**. On the other hand, for the point $x = 1$, the vector field moves away from it, indicating an **unstable equilibrium**. □

Example 2.17. (*Example from [14]*) Consider the logistic equation $x' = rx - x^2$ where r is a unknown parameter of the differential equation. The following figure illustrates the effect of r in the vector field of the equation.

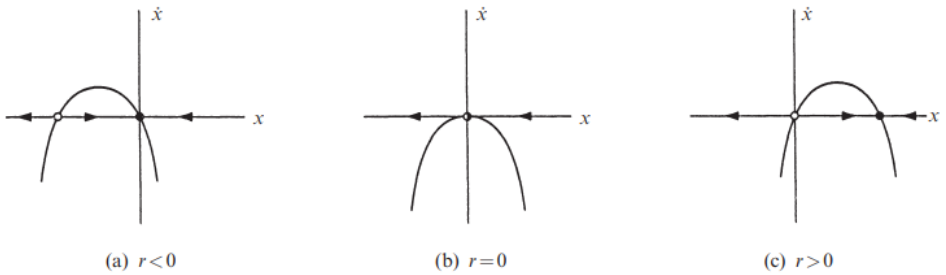


Figure 2.6: Effect of the parameter r in the vector field associated to the differential equation. Figure from [14].

The fixed points of this equation are $\{0, r\}$; this implies a fixed point may change its stability as the parameter is varied. Note that $x^* = 0$ is an equilibrium point independently the value of r. Nonetheless, for $r < 0$, $x^* = r$ is an unstable fixed point. As r increases, the unstable fixed point approaches to the origin, and joins with it when $r = 0$. Finally, when $r > 0$, the origin has become unstable and $x^* = r$ is now stable.

The following figure shows the **bifurcation** diagram for this equation: the parameter r is regarded as the independent variable, and the fixed points $x^* = 0$ and $x^* = r$ are shown as dependent variables.

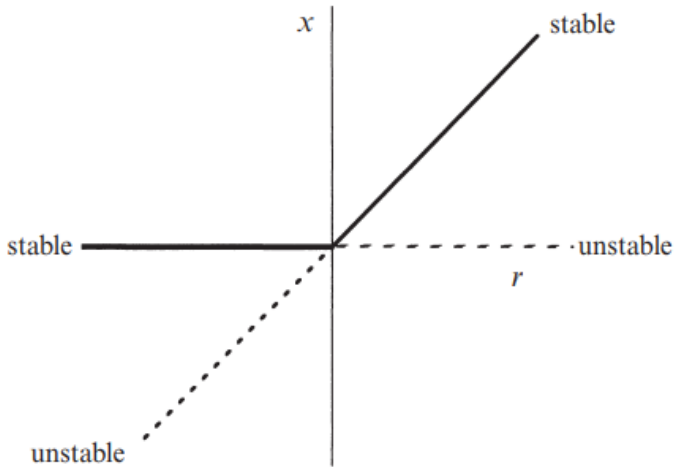


Figure 2.7: Bifurcation diagram of $x' = rx - x^2$. Figure from [14].

Note that the two fixed points change its stability when $r = 0$.

Example 2.18. Continuing with example about predator-prey model, it is easy to see that the points $\{(0,0), \left(\frac{\alpha}{\beta}, \frac{\gamma}{\delta}\right)\}$ are the only equilibrium points. The first fixed point effectively represents the extinction of both species. If both populations are at 0, then they will continue to be so indefinitely. The second fixed point represents a fixed point at which both populations sustain their current, non-zero numbers, and, in the simplified model, do so indefinitely. The following figure shows the phase portrait for the system of differential equations of the predator-prey model when all parameters are taken equal to 1. A phase portrait is a graphical representation showing how the solutions of a dynamical system evolve in phase space. The axes correspond to the system variables and the trajectories show how these variables change with time. It helps to identify qualitative behaviors such as equilibrium points, limit cycles and stability.

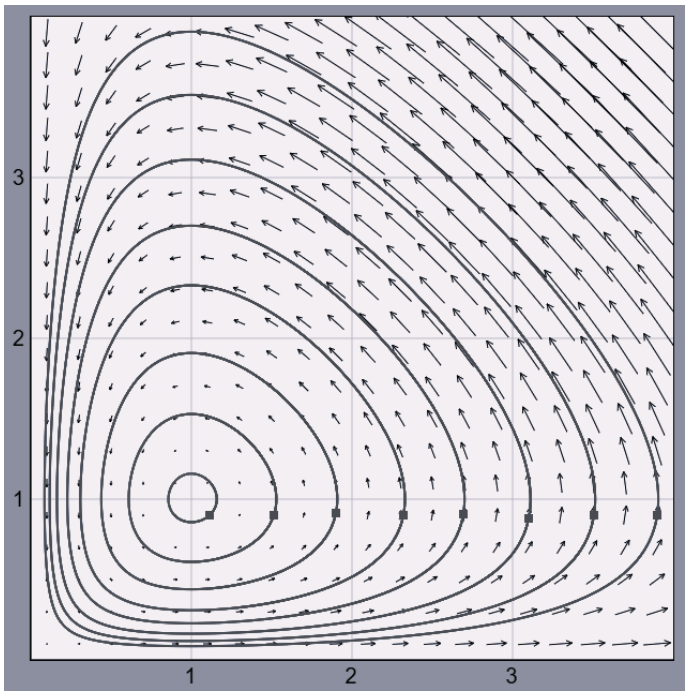


Figure 2.8: Phase portrait of the predator-prey model with trayectories and the vector field. Figure elaborated by the author used the software MATLAB provided by [22].

2.2.1.4 Numerical methods for ODE

In this section, numerical methods for solving ordinary differential equations (ODEs) are explored. The focus is on two methods: the Euler method and the Picard method. These methods are widely used to approximate the solutions to ODEs when analytical solutions are difficult or impossible to obtain. This section is based on [11,14].

Euler Method

The Euler method is one of the simplest and most commonly used numerical methods for solving first-order ODEs. Given an ODE of the form

$$y'(t) = f(t, y(t)),$$

with an initial condition $y(t_0) = y_0$, the Euler method approximates the solution at discrete time steps using the approximation of first order of the function f.

Definition 2.19. (**Euler Method**, [11]) The Euler method for solving the ODE is

given by the recurrence relation:

$$y_{n+1} = y_n + h \cdot f(t_n, y_n),$$

where:

- y_n is the approximation of the solution at time t_n,
- h is the step size,
- $f(t_n, y_n)$ is the value of the derivative of y at t_n.

The method uses the current value y_n to compute the next value y_{n+1}. It is straightforward to implement, but it is conditionally stable, meaning that it can become unstable if the step size h is too large.

Picard Method

The Picard method is an iterative approach for solving initial value problems of the form:

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0.$$

It is based on the idea of approximating the solution by iterating on the integral form of the ODE, such as the proof of Picard's Theorem.

Definition 2.20. (**Picard's iteration.**, [11]) The Picard method generates a sequence of functions $y_n(t)$ that converge to the solution of the ODE. The method is defined by the following recurrence relation:

$$y_{n+1}(t) = y_0 + \int_{t_0}^{t} f(s, y_n(s))\, ds,$$

where y_0 is the initial condition and $y_n(t)$ is the approximation of the solution at the n-th iteration. The Picard iteration is repeated until the difference between successive approximations becomes small. It is important to remark that the iterations of the Picard's method corresponds to polinomials approximating the real solution, then, the approximation is local.

Example 2.21. In order to illustrate both numerical methods, consider the initial value problem:

$$\frac{dy}{dt}(t) = -ty(t), \quad y(-\sqrt{2}) = e^{-1}.$$

The analytic solution to this problem is $y(t) = e^{-\frac{1}{2}t^2}$. In the following two figures are shown the results of apply the methods in the solution of this Cauchy problem.

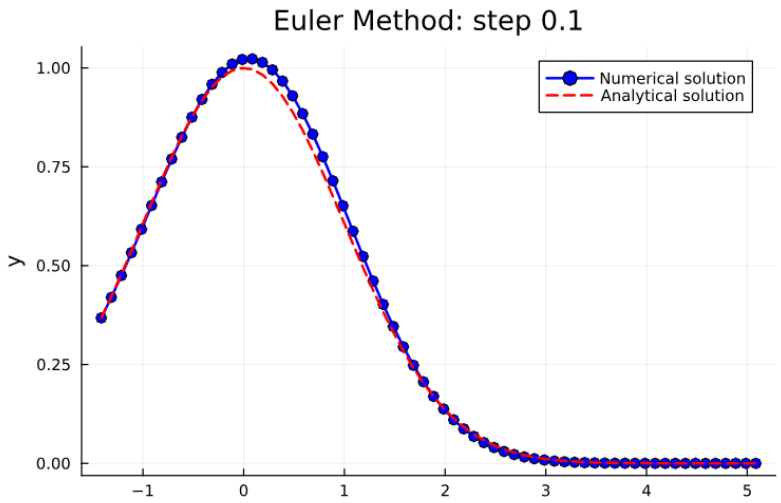


Figure 2.9: Euler method compared with the analytical solution. Figure elaborated by the author in Julia.

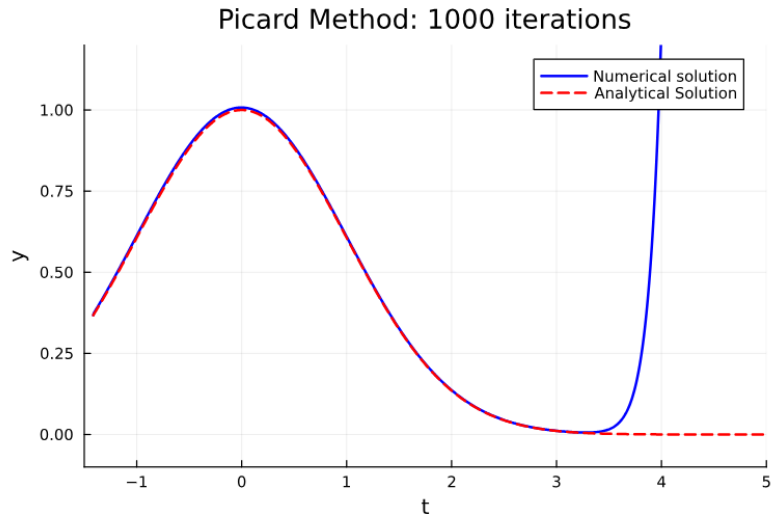


Figure 2.10: Picard method compared with the analytical solution. Figure elaborated by the author in Julia.

2.3 Markov chains

Markov chains offer a mathematical framework for modeling systems that transition between states, where the probability of moving to a new state depends only on the current state, not on the sequence of events that preceded it. This section will cover essential concepts such as transition matrices, steady-state distributions, and state classifications. Mastery of these concepts will be crucial for understanding the most

important theorem in this thesis, which leverages the principles of Markov chains to analyze complex dynamical systems and their long-term behavior. The content of this section is taken from [8].

Definition 2.22. ([8]) A **Markov chain** is a stochastic process that satisfies the **Markov property**, which states that the future state depends only on the present state and not on the sequence of past states. Formally, a discrete-time Markov Chain is a sequence of random variables $\{X_n\}_{n=0}^{\infty}$ defined on a state space S such that for all $n \geq 0$ and all $x_0, x_1, \ldots, x_n, x_{n+1} \in S$,

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

The *transition probabilities* are given by

$$P_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad \text{for } i, j \in S,$$

where P_{ij} represents the probability of transitioning from state i to state j in one step.

The *transition matrix* of the Markov Chain is

$$P = \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1m} \\ P_{21} & P_{22} & \ldots & P_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \ldots & P_{mm} \end{bmatrix},$$

where $m = |S|$ is the size of the state space.

A Markov chain is said to be **time-homogeneous** if P_{ij} does not depend on n, i.e., the transition probabilities are constant over time. This chapter it will consider only time-homogeneous Markov chains.

Example 2.23. A first Markov Chain. Consider the Markov chain $(X_0, X_1, \ldots)$ with state space $S = \{a, b, c, d\}$ and the transition matrix

$$P = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

This could be represented in a graph called transition graph as follows.

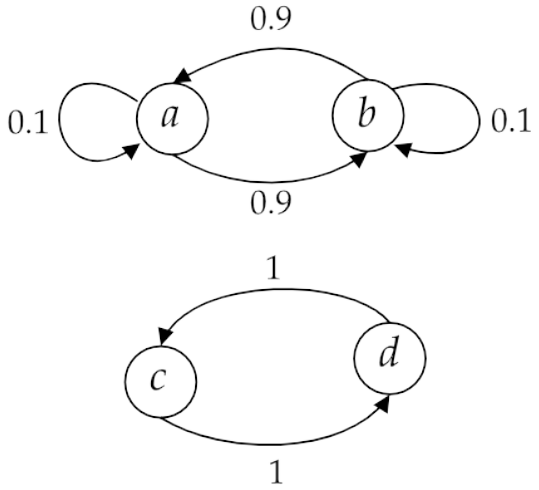


Figure 2.11: Graph representation of a Markov chain with four states. Figure elaborated by the autor.

Note that the four states are represented by nodes and the arrows betwenn nodes depict the possible transition with its respective transition probability.

2.3.1 Irreducible and aperiodic Markov Chains

For many results and applications in Markov theory, certain assumptions about the Markov chains are required. These conditions are crucial for studying stationary distributions and understanding the Markov Chain Monte Carlo (MCMC) methods. In this subsection, it is assumed that all Markov chains under consideration are homogeneous.

The main idea when a Markov chain is **irreducible** is centered in the property that "all states of the chain can be reached from all others". To make this more precise, consider the following definitions.

Definition 2.24. (**Communication**, [8]) Let $(X_0, X_1, \ldots)$ a Markov chain with finite state space $S = \{s_1, s_2, \ldots, s_k\}$ and transition matrix $P \in [0,1]^{k\times k}$. A state s_i communicates with another state s_j, written as $s_i \to s_j$ if the chain has positive probability of ever reaching the state s_j when we start from the state s_i. In other words, $s_i \to s_j$ if there exists an $n \in \mathbb{N}$ such that

$$\mathbb{P}(X_{m+n} = s_j |\ X_m = s_i) > 0.$$

Definition 2.25. (**Intercommunication**, [8]) Let $(X_0, X_1, \ldots)$ a Markov chain with finite state space $S = \{s_1, s_2, \ldots, s_k\}$ and transition matrix $P \in [0,1]^{k\times k}$. A state s_i intercommunicates with another state s_j, written as $s_i \leftrightarrow s_j$ if $s_i \to s_j$ and $s_j \to s_i$.

The definition of intercommunication between states directly implies the notion of irreducibility in a Markov chain. First, it is easy to prove that this corresponds to an equivalence relation on the set of states of the chain.

Proposition 2.26. The relation of intercommunication defines an equivalence relation over the state space of a Markov Chain.

Definition 2.27. (**Irreducibility**, [8]) A Markov chain $(X_0, X_1, \ldots)$ with state space $S = \{s_1, s_2, \ldots\}$ and transition matrix P is said to be **irreducible** if for all $s_i, s_j \in S$ the states are intercommunicated. In other words, the equivalence relation $\leftrightarrow$ only has one equivalence class.

In the transition graph of example 2.23, immediately it could be seen that if the chain starts in the state a or b, then it is restricted to this states forever, as well as if the initial state is c or d. This means that the equivalence class of of a is $\{a, b\}$ and the class of c is $\{c, d\}$.

Definition 2.28. (**Aperiodicity**, [8]) A Markov chain is said to be **aperiodic** if all its states are aperiodic: for all $s_i \in S$, $gcd(\{n \geq 1 : (P^n)_{i,i} > 0\}) = 1$. Otherwise the chain is said to be **periodic**.

In the example 2.23, it is remarkable that if the chain starts in the state c then the chain will return every two steps to the same state, therefore, that Markov chain is periodic.

2.3.2 Stationary distributions

Definition 2.29. (**Stationary Distribution**, [8]) Let P be a transition matrix of a Markov chain with state space S. A probability distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ ($\sum_{i=1}^{k} \pi_i = 1$ and $\pi_i \geq 0$) is called a stationary distribution if it satisfies the equation:

$$\pi P = \pi,$$

meaning that $\sum_{i=1}^{k} \pi_i P_{i,j} = \pi_j$ for $j = 1, \ldots, k$. In other words, the stationary distribution is a distribution that remains unchanged after applying the transition

matrix P. This means that the probability distribution over the states does not evolve over time.

Theorem 2.30. (**Existence of Stationary Distribution**, [8]) Let P be a transition matrix of a finite Markov chain. If the chain is irreducible and aperiodic then there exists a stationary distribution π. Specifically, there exists a probability vector π that satisfies:

$$\pi P = \pi.$$

Furthermore, the stationary distribution is unique, and the chain converges to this distribution regardless of the initial state distribution.

Proof. See chapter 5 of [8].

Now, it is important to consider the asymptotic behavior of the probability distribution vector $\mu^{(n)}$ in the step $n \in \mathbb{N}$ when the chain starts with arbitrary initial distribution $\mu^{(0)}$. That is why it is necessary to define a metric.

Definition 2.31. (*Total variation distance*, [8]) If $\nu^{(1)} = (\nu_1^{(1)}, \ldots, \nu_k^{(1)})$ and $\nu^{(2)} = (\nu_1^{(2)}, \ldots, \nu_k^{(2)})$ are probability distributions on $S = \{s_1, \ldots, s_k\}$, then we define the **total variation distance** between $\nu^{(1)}$ and $\nu^{(2)}$ as

$$d_{\mathrm{TV}}(\nu^{(1)}, \nu^{(2)}) = \frac{1}{2} \sum_{i=1}^{k} \left| \nu_i^{(1)} - \nu_i^{(2)} \right|.$$

If $\nu^{(1)}, \nu^{(2)}, \ldots$ and ν are probability distributions on S, then we say that $\nu^{(n)}$ converges to ν in total variation as $n \to \infty$, writing $\nu^{(n)} \xrightarrow{\mathrm{TV}} \nu$, if

$$\lim_{n \to \infty} d_{\mathrm{TV}}(\nu^{(n)}, \nu) = 0.$$

The constant $\frac{1}{2}$ in the definition of the metric is designed to make the total variation distance d_{TV} take values between 0 and 1.

The main result about convergence to stationarity is stated here below.

Theorem 2.32. (**Convergence to stationary distribution**, [8,10]) If the Markov chain is irreducible and aperiodic, then for any initial distribution μ, the distribution of the chain converges to the stationary distribution π as $n \to \infty$, that is,

$$\mu^{(n)} \xrightarrow{\mathrm{TV}} \pi.$$

This means that, in the long run, the system forgets its initial state and settles into the stationary distribution.

Proof. Consult Chapter 5 of [8].

A collorary of the convergence theorem is the uniqueness of the stationary distribution.

Theorem 2.33. (Uniqueness of the stationary distribution). Any irreducible and aperiodic Markov chain has exactly one stationary distribution.

2.3.3 Reversible Markov Chains

Definition 2.34. (Reversible Markov Chain). A Markov chain with transition matrix P is said to be reversible with respect to a probability distribution π if it satisfies the **detailed balance equations**

$$\pi_i P_{i,j} = \pi_j P_{j,i}, \quad \text{for all } i, j \in S.$$

In other words, for any pair of states i and j, the probability of transitioning from i to j is the same as transitioning from j to i, weighted by the distribution π.

The Markov Chain is said to be reversible if there exists a reversible distribution for it.

Theorem 2.35. If a Markov chain is reversible with respect to a probability distribution π, then π is a stationary distribution. Specifically, the detailed balance equations imply that

$$\pi P = \pi.$$

Thus, reversibility guarantees the existence of a stationary distribution for the chain.

Proof. The property that $\pi_i \geq 0$ for all i is already satisfied. On the other hand,

$$\pi_j = \sum_{i=1}^{k} P_{j,i} = \sum_{i=1}^{k} \pi_j P_{j,i} = \sum_{i=1}^{k} \pi_i P_{i,j}.$$

Then, π is a stationary distribution of the Markov chain.

Example 2.36. Consider the Markov chain defined by the graph in the Figure 2.12,

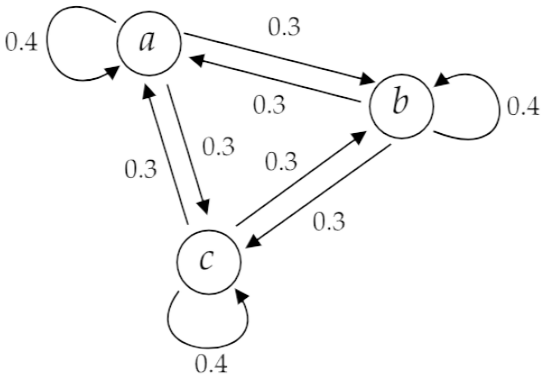


Figure 2.12: The transition graph of a reversible Markov chain. Figure elaborated by the autor.

with transition matrix,

$$P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}.$$

It is easy to see that the transition matrix of this Markov chain is symmetric and the stationary distribution is $\pi = (1/3, 1/3, 1/3)$, these two facts implies that the Markov chain is reversible.

Chapter 3

Statistical approach to inverse problems

Statistical inversion methods reframe inverse problems as a quest for information within a statistical framework. These methods distinguish between directly observable quantities and unobservable variables, focusing primarily on the latter when they are of key interest. Unobservable quantities are interconnected through models, and the aim of statistical inversion is to extract information about these variables while quantifying the uncertainty associated with them. This is achieved by incorporating knowledge of the measurement process and prior information about the unknowns (consult [5,6,7,10]).

The statistical inversion approach relies on four fundamental principles: all variables in the model are treated as random variables, randomness represents the degree of information about their realizations, this information is encoded in probability distributions, and the solution to the inverse problem is the posterior probability distribution. This framework provides a robust method for integrating uncertainty into the analysis of inverse problems (see [10]).

As stated above, the main objective in solving an inverse problem is to infer the parameters of a model based on observed data. Unlike forward problems, which are typically well-posed and have a unique solution due to the principle of causality, inverse problems often face multiple solutions or no solution at all. The uncertainty associated with inverse problems stems from the fact that different models can predict similar observations, or the data may not be consistent, making the solution highly non-unique. This non-uniqueness creates a major challenge in extracting

meaningful conclusions from the data.

One of the critical aspects of dealing with inverse problems is the handling of errors in the data. For instance, in classical approaches such as the least-squares method, aim to minimize the sum of squared residuals between the model's predictions and the observations. This approach assumes that errors in the data follow a Gaussian distribution. On the other hand, the least-absolute-values method assumes that the errors follow a Laplacian distribution. While the least-squares method is widely used because of its simplicity and computational efficiency, it is less robust in the presence of large outliers or uncontrolled errors. In contrast, the least-absolute-values method is more robust to outliers, although it is computationally more complex (see [6]).

In recent years, the Bayesian framework has emerged as a powerful tool for solving inverse problems, offering a more comprehensive approach to error handling. Bayesian inference allows the incorporation of prior information about the model parameters and the uncertainty associated with the data. Rather than seeking a single "best" solution to the inverse problem, Bayesian methods estimate the probability distribution of the parameters, given the observed data. This approach provides a more nuanced understanding of the solution, acknowledging the inherent uncertainty and the possibility of multiple models that could explain the observations (a deeper treatment of this topic can be found in [15]).

A key strength of the Bayesian approach is its ability to quantify uncertainty in a rigorous way. Instead of relying on a single point estimate for the parameters, Bayesian methods generate a posterior distribution that encapsulates all the possible values of the parameters, weighted by their likelihood given the data. This distribution can then be used to make probabilistic predictions and assess the confidence in the inferred parameters. In situations where the data is noisy or incomplete, Bayesian methods offer a more reliable framework for drawing conclusions about the model.

The application of the Bayesian framework to inverse problems also allows for the inclusion of prior knowledge, which can help constrain the solution and reduce the non-uniqueness of the problem. In the absence of such prior information, the solution space may be vast and ambiguous, but Bayesian methods can help focus the search for solutions in more plausible regions of the parameter space. Moreover, the use of Bayesian methods in inverse problems has been recognized as a unifying framework for predictive science, as it bridges the gap between model selection, uncertainty

quantification, and data interpretation (consult [15]).

One of the most significant challenges in the Bayesian approach to inverse problems is the computational complexity involved in evaluating the posterior distribution, particularly in high-dimensional parameter spaces. The need to explore all possible parameter combinations requires sophisticated sampling techniques, such as Markov Chain Monte Carlo (MCMC) methods or variational inference, which can be computationally expensive. However, the insights gained from the Bayesian framework often outweigh the computational cost, especially when dealing with complex systems where understanding uncertainty is crucial (go to [10]).

In conclusion, the handling of error and uncertainty in inverse problems is one of the key challenges in scientific modeling. The traditional methods, such as least-squares and least-absolute-values, provide valuable tools but often fail to account for the complexities of real-world data, including outliers and non-uniqueness in the solutions. The Bayesian approach, however, offers a more robust and comprehensive framework for addressing these issues. By treating the parameters as probability distributions and incorporating prior knowledge, Bayesian inference allows for more accurate and reliable solutions to inverse problems, providing a powerful tool for scientific discovery. This approach not only enhances the interpretation of data but also enables the inclusion of uncertainty in decision-making, making it a crucial tool in science. Consult the references [5,6,7,10,15].

3.1 Bayesian Formulation

This section introduces the concept of a data assimilation window, followed by the dynamical model and its uncertain quantities, and concludes with the definitions of the model state and the state vector.

The discussion begins with the concept of data assimilation (this section is motivated by [5]).

Definition 3.1. Data assimilation refers to the process of integrating observational data into computational models to enhance their accuracy and predictive capabilities.

To contextualize this concept, it is essential to understand the study space: data assimilation operates sequentially over defined time intervals known as **assimilation**

windows. The length and structure of these windows vary depending on the methodology. Some methods update the model's solution throughout the entire window, while others focus on specific time points. Furthermore, certain approaches treat each assimilation window independently, whereas others propagate information between consecutive windows, enabling diverse applications and formulations, this part is extracted from [5].

3.1.1 Model with uncertain inputs

The dynamic process under analysis is represented by a forward model that incorporates uncertainty within an assimilation window:

$$\begin{aligned} z_0 &= \hat{z}_0 + z_0', \\ \theta &= \hat{\theta} + \theta', \\ \varepsilon_k &= \overline{\varepsilon} + \varepsilon_k', \\ z_k &= m(\{z_0, \ldots, z_{k-1}\}, \theta, q_k). \end{aligned}$$

In this model, z_0 represents the initial conditions, comprising noise-free values $\hat{z}_0$ and associated uncertainty z_0'. The vector θ includes uncertain model parameters with exact values $\hat{\theta}$ and uncertainties θ'. Similarly, the model error ε accounts for discrepancies resulting from unrepresented phenomena or numerical discretization, with uncertainty ε'. The system evolves according to the model equation:

$$z_k = m(\{z_0, \ldots, z_{k-1}\}, \theta, \varepsilon_k),$$

over K time steps within the assimilation window. For simplicity, boundary conditions and their uncertainties are excluded to avoid additional constraints on the system.

Definition 3.2. The *model state*, denoted by $z^T = (z_0^T, z_1^T, \ldots, z_K^T)$, refers to the sequence of state vectors predicted by the model within an assimilation window. Each z_k^T represents the model state at time step k. The model operator m predicts the model state across the entire assimilation window based on the initial state x_0, model parameters θ, and model error ε, as follows:

$$z = m(\{z_0, \ldots, z_{k-1}\}, \theta, \varepsilon).$$

The model state is specifically associated with the variables predicted by the model's equations and may differ from the general state vector employed in data assimilation.

Definition 3.3. The state vector x of the data-assimilation problem is defined containing all the uncertain quantities that are to be estimated. In the model-state formulation, the state vector x includes the model state z or a subset (the model state at the final time step) and the model parameters. In this formulation, the state vector is expressed as:

$$x^T = (z^T, \theta^T),$$

where the model error ε is excluded. This approach centers on directly updating the state of the model. In some cases, the state vector targets only the model state and in other cases it focuses just the model parameters. In addition to the elements mentioned above, there is a *vector of measurements* y, which may exist in different assimilation windows and represents the measurements taken from the dynamic system under consideration.

Once the elements involved in a dynamic system have been defined, it is possible to move on to the relationship with inverse problems and Bayesian inference.

3.1.2 Inverse problems and Bayesian inference

In the context of inverse problems, as stated in the section 2.1.1 of classical regularization, the objective is to obtain information about a quantity $x \in \mathbb{R}^n$ by measuring another quantity $y \in \mathbb{R}^m$. This process requires a model for the dependency between these quantities, typically expressed as (this section corresponds to the treatment performed in [10]):

$$y = f(x, \varepsilon),$$

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ represents the model function or measurement operator, a potentially nonlinear function that maps the model state vector x into measurement space, and $\varepsilon \in \mathbb{R}^k$ encapsulates the unknown measurement noise. Since the measured quantity y always contains noise, calibration measurements or regularization methods are often employed to address this challenge.

In statistical inverse problems, all parameters are treated as random variables, denoted by capital letters. Consequently, the model becomes:

$$Y = f(X, E),$$

where Y, X, and E are random variables. Their associated probability distributions are interdependent, forming the foundation of the statistical inversion approach.

Unlike traditional methods, this approach relies on the probability distributions of the random variables rather than deterministic values. The approach is based on relations between probability distributions (consult [10,17,18]).

The measured quantity Y is referred to as the *observable*, while its realization, $Y = y_{\text{observed}}$, corresponds to the actual data obtained. Conversely, the quantity X, which is the primary focus, is termed the *unknown*.

Before measuring Y, it is assumed that some information about X is available. In Bayesian theory, this is known as the a priori information of unknown quantities and is represented as a probability distribution $x \to \pi_{pr}(x)$ called a *priori density.* On the other hand, considering that X and Y represent random variables, then they have a joint density function $\pi(x, y)$ that is unknown in principle. However, given that there is a previous knowledge of X encoded in the a priori density, it must be satisfied that

$$\int_{\mathbb{R}^m} \pi(x, y)dy = \pi_{pr}(x).$$

If the value of the *unknown* is known, i.e., $X = x$, the conditional probability density of Y given this information is

$$\pi(y|x) = \frac{\pi(x, y)}{\pi_{\text{pr}}(x)}, \quad \text{if} \quad \pi_{\text{pr}}(x) \neq 0.$$

The conditional probability of Y is called the likelihood function, as it expresses the likelihood of different measurement outcomes with $X = x$ given.

Consider finally that the measurement data $Y = y_{\text{observed}}$ is given. The conditional probability distribution

$$\pi(x|y_{\text{observed}}) = \frac{\pi(x, y_{\text{observed}})}{\pi(y_{\text{observed}})}, \quad \text{if} \quad \pi(y_{\text{observed}}) = \int_{\mathbb{R}^n} \pi(x, y_{\text{observed}})\, dx \neq 0,$$

is called the posterior distribution of X, $\pi_{post}(x)$. This distribution expresses what is known about X after the realized observation $Y = y_{\text{observed}}$.

In the Bayesian framework, the inverse problem is expressed as follows:

Given the data $Y = y_{observed}$, find the conditional probability distribution $\pi(x|y_{observed})$ of the variable X is sought.

The notations and results are summarized in the following theorem, referred to as Bayes' theorem of inverse problems.

Theorem 3.4. ([5,10]) Assume that the random variable $X \in \mathbb{R}^n$ has a known prior probability density $\pi_{\text{pr}}(x)$ and the data consist of the observed value y_{observed} of an observable random variable $Y \in \mathbb{R}^m$, such that $\pi(y_{\text{observed}}) > 0$. Then, the posterior probability distribution of X, given the data y_{observed}, is

$$\pi_{\text{post}}(x) = \pi(x|y_{\text{observed}}) = \frac{\pi_{\text{pr}}(x)\pi(y_{\text{observed}}|x)}{\pi(y_{\text{observed}})}.$$

In the sequel, $y = y_{\text{observed}}$ will be written, and it is understood that when the posterior probability density is evaluated, the observed value of y is used.

In the previous equation, the marginal density

$$\pi(y) = \int_{\mathbb{R}^n} \pi(x, y)\, dx = \int_{\mathbb{R}^n} \pi(y|x)\pi_{\text{pr}}(x)\, dx$$

acts as a normalizing constant and is generally of little importance (as will be discussed later). It is important to note that, in principle, $\pi(y) = 0$ is possible, meaning that measurement data could have zero probability. Although this situation is rare in practice, it would indicate that the underlying models are inconsistent with reality.

In summary, examining Bayes' formula, solving an inverse problem involves several important tasks. First, based on prior information about the unknown X, a prior probability density π_{pr} that adequately reflects this information must be found. Next, the likelihood function $\pi(y|x)$, which describes the relationship between the observation and the unknown, must be determined. Finally, methods must be developed to explore the posterior probability density. Each of these steps can present its own challenges.

Before addressing these issues in more detail, it is important to refer to the classical inversion methods of how the statistical solution of an inverse problem can be related to producing unique estimates.

3.1.3 Estimators

The solution of the inverse problem is defined as the posterior distribution. When the unknown consists of a random variable with few components, the posterior probability density can be visualized as a nonnegative function of these variables. However, in most real-world inverse problems, the dimensionality is significantly large, rendering direct visualization of the posterior distribution impossible. Despite

this, a known posterior distribution allows for the calculation of various estimates. Point estimates determine the most probable value of the unknown or the expectation of the posterior distribution, given the data and prior information. Interval estimates provide the range in which the unknown lies with a specific probability, such as 90 %, based on the prior information and the data. Consult [5,7,10] for further analysis.

Definition 3.5. (*Maximum a posteriori*, [10]) Given the posterior distribution and its probability density $\pi(x|y)$ of the unknown $X \in \mathbb{R}^n$ the maximum a posteriori estimation x_{MAP} satisfies

$$x_{MAP} = argmax_{x \in \mathbb{R}^\times} \pi(x|y),$$

assuming that such maximizer exists. Note that nothing ensures the uniqueness of this estimator and this could be a problem.

Definition 3.6. (*Conditional mean*, [10]) The conditional mean estimator x_{CM} of the unknown X conditioned on the data y is defined as

$$x_{CM} = \mathbb{E}(X|y) = \int_{\mathbb{R}^n} x\pi(x|y)dx,$$

provided that this integral converges. Find the conditional mean is generally seen as an integration problem. This leads to *Monte Carlo* methods for calculate in a numerical way the result of the integral.

Example 3.7. (*Example taken from [6]*) Consider the random variable H with the probability density function $\pi(h)$ given by the mixture of two normal distribution

$$\pi(h) = \frac{4}{10}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(h+3)^2}{2}}\right) + \frac{6}{10}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{(h-3)^2}{2}}\right).$$

The density of H is shown in the Figure 3.1.

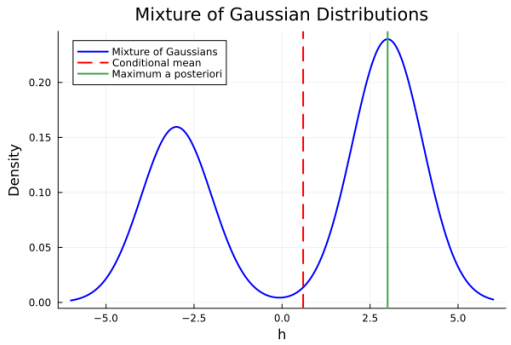


Figure 3.1: Mixture of two normal distribution and its mean and its maximizer value. Figure elaborated by the author using Julia.

As can be seen, the mean of H is $\mathbb{E}(H) = \frac{4}{10}(-3) + \frac{6}{10}(3) = \frac{6}{10} = 0.6$ and the maximizer of the function $\pi(h)$ is $h^* = 3$. In the Bayesian framework the former corresponds to the x_{MAP} estimator while the latter represents the x_{CM} estimator. It is clear that sometimes these point estimations do not represents adequately the information of the posterior distribution, for instance, the maximum a posteriori loses the information about the two modes of the distribution and the conditional mean does not tell anything about the density.

Definition 3.8. Maximum likelihood. The maximum likelihood estimator x_{ML} satisfies

$$x_{ML} = argmax_{x \in \mathbb{R}^\times} \pi(y|x).$$

This estimator seeks to estimate the value of the unknown which is most likely to produce the data y. This is a non Bayesian (frequentist) estimator and in the context of inverse problems in quite useless: often it corresponds to solve the inverse problem without any kind of regularization.

Another approach to estimation is based on intervals. Similar to the frequentist perspective, these intervals are constructed using a point estimate that defines the center of the interval and a radius selected to capture a desired level of probability mass. With a known posterior distribution is easy to construct the interval estimator in exact way, opposed to frequentist perspective that build this estimators with asymptotic assumptions (consult [17]).

3.1.4 Likelihood function

The likelihood function contains the forward model used in classical inversion techniques as well as information about the noise and other measurement and modelling uncertainties.

3.1.4.1 Modelling noise

Consider the random model

$$Y = f(X, E)$$

where X is the unknown value and E denotes the noise, suppose that X and E are mutually independent, and the model f allows the noise E to be expressed as a

function g of Y and X in the following way

$$E = g(X, Y),$$

where g has a strong relation with f and the random variables X and Y take values in the domain of the function g.

Assume that the probability distribution of the noise E is known, that is,

$$\mu_E(B) = \mathbb{P}(E \in B) = \int_B \pi_{noise}(e)de,$$

where B is a Borel measurable set.

If $X = x$ is fixed, the assumption of mutual independence between X and E ensures that the probability density of E remains unchanged when conditioned on $X = x$. Therefore, it can be deduced that Y, conditioned on $X = x$, is distributed like E, with the probability density being modified by $g(\cdot, \cdot)$. In other words, the likelihood function is

$$\pi(y|x) = \pi_{noise}(g(x, y)).$$

Thus, if the prior distribution of X is $\pi_{pr}(x)$, then, according to the Bayes' theorem,

$$\pi(x|y) \propto \pi_{noise}(g(x, y))\pi_{pr}(x).$$

A slightly more complicated situation arises when the unknown X and the noise E are not mutually independent. In this case, the conditional density of the noise must be known, which is given by

$$\mu_E(B \mid x) = \int_B \pi_{\text{noise}}(e \mid x)\, de.$$

In this scenario, the following can be written:

$$\pi(y \mid x) = \int_{\mathbb{R}^m} \pi(y \mid x, e)\pi_{\text{noise}}(e \mid x)\, de.$$

When both $X = x$ and $E = e$ are fixed, Y is completely determined as $Y = y = f(x, e)$. Hence,

$$\pi(y \mid x, e) = \delta(y - f(x, e)),$$

where δ represents the Delta Dirac distribution. Substituting this into the previous formula results in

$$\pi(y \mid x) = \pi_{\text{noise}}(g(x, y) \mid x),$$

and therefore

$$\pi(x \mid y) \propto \pi_{\text{pr}}(x)\pi_{\text{noise}}(g(x, y) \mid x).$$

Example 3.9. (*Additive noise, example from [10]*) If the model is establised as

$$Y = f(X) + E,$$

then the model contains additive noise. In this case, the function g is $g(x, y) = y - f(x)$. If it is assumed the independence between X and E then the likelihood function is

$$\pi(y|x) = \pi_{noise}(y - f(x)).$$

Example 3.10. (*Multiplicative noise, example inspired in [6,10]*) Consider a simple real-valued measurement, where the observation model includes multiplicative noise that is mutually independent of the unknown variable. The stochastic model relating $X \in \mathbb{R}$ and $Y \in \mathbb{R}$ is given by

$$Y = Ef(X),$$

where E is a real-valued noise term, and $f : \mathbb{R} \to \mathbb{R}$. If π_{noise} represents the probability density of E, the likelihood function can be expressed as

$$\begin{aligned}\pi(y \mid x) &= \int_{\mathbb{R}} \delta(y - ef(x))\pi_{\text{noise}}(e)\, de \\ &= \frac{1}{f(x)} \int_{\mathbb{R}} \delta(y - \nu)\pi_{\text{noise}}\left(\frac{\nu}{f(x)}\right) d\nu \\ &= \frac{1}{f(x)}\pi_{\text{noise}}\left(\frac{y}{f(x)}\right),\end{aligned}$$

where δ denotes the Dirac delta function.

Example 3.11. (*Example based in [10]*) Consider now the case in which a noisy measurement is given with an incompletely known forward model. Let $A(v) \in \mathbb{R}^{m\times n}$ denote a matrix that depends on a parameter vector $v \in \mathbb{R}^k$, and assume that the deterministic model without measurement noise is $y = A(v)x$, where $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$. Furthermore, assume that the actual measurement is corrupted by additive noise that is mutually independent of the unknown X and the parameter V. Thus, the statistical model in this case becomes

$$Y = A(V)X + E.$$

If π_{noise} is the probability density of E, which is mutually independent of X and V, then

$$\pi(y \mid x, v) = \pi_{\text{noise}}(y - A(v)x).$$

Additionally, assuming that V and X are mutually independent and that V has density π_{param}, the likelihood density is obtained as

$$\pi(y \mid x) = \int_{\mathbb{R}^k} \pi(y \mid x, v)\pi_{\text{param}}(v)\, dv = \int_{\mathbb{R}^k} \pi_{\text{noise}}(y - A(v)x)\pi_{\text{param}}(v)\, dv.$$

Example 3.12. (*Logistic growth model, example from the reference [6]*) Consider the logistic growth model

$$x'(t) = rx\left(1 - \frac{x}{K}\right), \qquad x(t_0) = x_0,$$

where t_0 and x_0 are known values and r and K are the unknown parameters. The analytic solution of the previous Cauchy problem is

$$x(t) = x(t; r, K) = \frac{Kx_0e^{r(t-t_0)}}{K + x_0\left(e^{r(t-t_0)} - 1\right)}.$$

If it is assumed that data is available $\{(t_i, y_i)|i = 1, \ldots, m\}$ where t_i represents the time and y_i denotes the observation of the state variable x, and these corresponds with a Poisson model that depends on the mean given by $\overline{y_i} = x(t_i; r, K)$, then the model can be written as

$$Y_t = Poisson(x(t)),$$

for each time t. Therefore if it is assumed that each component has mutually independent in fluctuation,

$$\begin{aligned}\pi(y|x) &= \prod_{i=1}^{m} \frac{x(t_i)^{y_i}}{y_i!}e^{-x(t_i)} \\ &\propto \exp(y^\top log(x) - \mathbf{1}^\top x)\end{aligned}$$

where $log(x) = [log(x(t_1)), \ldots, log(x(t_m))]^\top$ and $\mathbf{1}^\top x = x(t_1) + \ldots + x(t_m)$.

3.1.5 Prior models

In the statistical theory of inverse problems, constructing the prior density is often considered the most critical and challenging aspect of the solution. The primary difficulty lies in the nature of the prior information, which is frequently qualitative rather than quantitative. The challenge, therefore, involves converting qualitative knowledge into a quantitative form that can be incorporated into the prior density. For instance, in subsurface electromagnetic studies, a geophysicist may anticipate layered structures with nonlayered inclusions or cracks. While a layered model may

capture some of these expectations, it can be overly restrictive, excluding other relevant structures. Similarly, in medical imaging, a radiologist might search for a well-localized cancer with a distinctive surface structure, as identified by trained expertise. These qualitative descriptions provide valuable insights but are difficult to express within the framework of probability densities (refer to [6,10,17,18]).

The general objective in the design of priors is to specify a density $\pi_{\text{pr}}(x)$ with the following property. If E represents a collection of expectable vectors x that correspond to possible realizations of the unknown X, and U represents a collection of unexpectable ones, the condition should hold that

$$\pi_{\text{pr}}(x) \gg \pi_{\text{pr}}(x') \quad \text{when} \quad x \in E,\, x' \in U.$$

In this way, the prior probability distribution should be concentrated on the values of x that are expected to be observed, assigning a significantly higher probability to these values than to those that are not expected to be observed.

In the Table 3.1, a summary of the possible distributions used to select the prior distribution is presented.

3.1.6 Posterior distribution

The third step in the statistical inversion technique involves developing methods to explore posterior probability densities. Before delving into these tools in the next chapter, it is important to consider how the posterior density should be interpreted in the context of inverse problems. The interpretation of the posterior distribution as the solution to an inverse problem is a nuanced matter. Fully understanding its meaning requires a clear grasp of credibility sets and intervals, also, a deep knowledge of the phenomena. (consult [10,17])

The example 3.13 is one that develops some intuition of the interpretation of the a posteriori distribution and also shows how estimation errors can occur when an a priori distribution is chosen.

Example 3.13. (*Example extracted of [10]*) Consider the following trivial inverse problem: determine the value of $x \in \mathbb{R}$ by directly measuring x with some additive, mutually independent noise. The statistical model is given by

$$Y = X + E.$$

Table 3.1: Summary of distributions commonly used as a priori distributions. Table elaborated by the autor and inspired in [18].

	Support	**Parameters**	**Mean**	**Variance**	**Use**
Binomial	$\{0, 1, \ldots, n\}$	n, p	np	$np(1-p)$	Discrete data modeling
Poisson	$\{0, 1, 2, \ldots\}$	$\lambda > 0$	λ	λ	Counting
Uniform	$[a, b]$	$a, b \in \mathbb{R}, a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	Non informative priors
Normal	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma^2 > 0$	μ	σ^2	Modeling continuous data
Lognormal	$(0, \infty)$	$\mu, \sigma^2 > 0$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	Skewed continuous data
Cauchy	$\mathbb{R}$	$x_0 \in \mathbb{R}, \gamma > 0$	Undefined	Undefined	Robust modeling
Laplace	$\mathbb{R}$	$\mu \in \mathbb{R}, b > 0$	μ	$2b^2$	Sharp peak data
Gamma	$(0, \infty)$	$\alpha > 0, \beta > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	Modeling rates or waiting times
Beta	$[0, 1]$	$\alpha > 0, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Modeling proportions
Exponential	$[0, \infty)$	$\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Modeling waiting times

Assume that the prior probability density of X is zero-mean Gaussian with unit variance, while the density of the noise E is zero-mean Gaussian with variance σ^2. Consequently, the posterior probability density of X is

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}x^2 - \frac{1}{2\sigma^2}(y-x)^2\right).$$

This density is Gaussian with respect to x, as will be discussed below. By completing the square and ignoring factors that depend only on y, it follows that for given data y,

$$\pi(x \mid y) \propto \exp\left(-\frac{1+\sigma^2}{2\sigma^2}\left(x - \frac{y}{1+\sigma^2}\right)^2\right).$$

From this expression, the point estimators x_{CM} and x_{MAP}, and variance γ^2 of X can be directly identified as

$$x_{MAP} = x_{CM} = \frac{y}{1+\sigma^2}, \quad \gamma^2 = \frac{\sigma^2}{1+\sigma^2}.$$

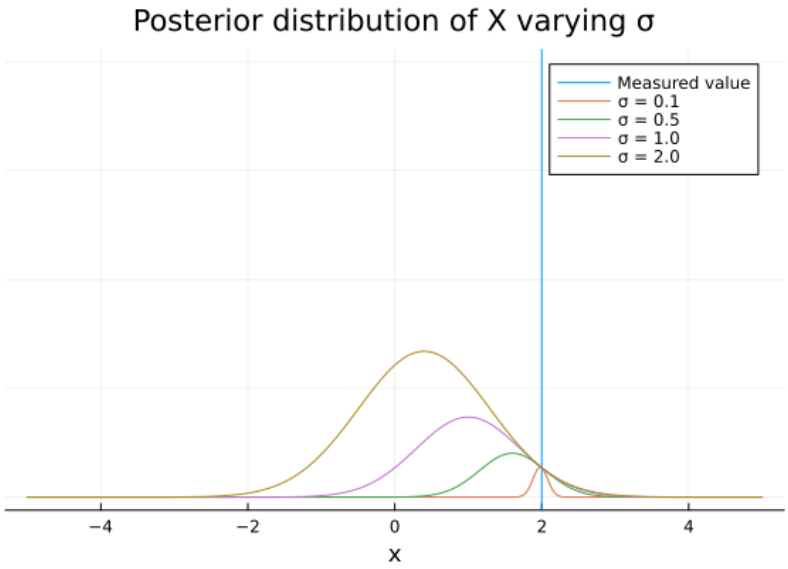


Figure 3.2: Posterior distribution of the unknown X given different values of σ with measured value $y = 2$. Figure elaborated by the author.

It is noteworthy that the posterior distribution is influenced by the value of the parameter σ in the prior distribution. As the value of this parameter increases, the distance between the point estimations and the measured data becomes larger, and the variance also increases. Note that the point estimations move forward the mean of the a priori distribution when σ becomes larger (consult the references [10,17]).

3.1.7 Gaussian environment

Gaussian probability densities have a special role in statistical inversion theory. First, they are relatively easy to handle and therefore they provide a rich source of tractable examples. But more importantly, due to the central limit theorem (consult appendixes and the references [10,18]), the Gaussian densities are often very good approximations to inherently non-Gaussian distributions when the observation is physically based on a large number of mutually independent random events.

Definition 3.14. Let $x_0 \in \mathbb{R}^n$ and $\Gamma \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix, denoted by $\Gamma > 0$ in the sequel. A Gaussian n-variate random variable X with mean x_0 and covariance Γ is a random variable with the probability density

$$\pi(x) = \frac{1}{(2\pi)^{n/2}|\Gamma|^{1/2}} \exp\left(-\frac{1}{2}(x - x_0)^T\Gamma^{-1}(x - x_0)\right),$$

where $|\Gamma| = \det(\Gamma)$. In such a case, we use the notation $X \sim N(x_0, \Gamma)$.

The following result concerning the conditional probability densities of Gaussian random variables will be stated.

Theorem 3.15. ([17,18]) Let $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^k$ be two Gaussian random variables whose joint probability density $\pi : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}_+$ is of the form

$$\pi(x, y) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}^\top \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}\right),$$

where $\Gamma_{11} \in \mathbb{R}^{n\times n}$ and $\Gamma_{22} \in \mathbb{R}^{k\times k}$. Then the probability distribution of X conditioned on $Y = y$, $\pi(x \mid y) : \mathbb{R}^n \to \mathbb{R}_+$, is of the form

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x - \bar{x})^\top \tilde{\Gamma}_{22}^{-1}(x - \bar{x})\right),$$

where

$$\bar{x} = x_0 + \Gamma_{12}\Gamma_{22}^{-1}(y - y_0),$$

and

$$\tilde{\Gamma}_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21},$$

is the Schur complement of Γ_{22} with respect to Γ.

Now, a result on the marginal distribution could be enunciated taking into account the above theorem.

Theorem 3.16. ([17,18]) Let X and Y be Gaussian random variables with joint probability density given by

$$\pi(x, y) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}^\top \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}\right).$$

Then the marginal density of X is

$$\pi(x) = \int_{\mathbb{R}^k} \pi(x, y)\, dy \propto \exp\left(-\frac{1}{2}(x - x_0)^\top \Gamma_{11}^{-1}(x - x_0)\right).$$

Example 3.17. Inverse linear problem. Consider the linear model with additive noise,

$$Y = AX + E$$

where $A \in \mathbb{R}^{m\times n}$ is a known matrix, and $X : \Omega \to \mathbb{R}^n$, $Y, E : \Omega \to \mathbb{R}^m$ are random variables. It is assumed that X and E are mutually independent Gaussian variables with probability densities

$$\pi_{\text{pr}}(x) \propto \exp\left(-\frac{1}{2}(x - x_0)^T \Gamma_{\text{pr}}^{-1}(x - x_0)\right),$$

and

$$\pi_{\text{noise}}(e) \propto \exp\left(-\frac{1}{2}(e-e_0)^T\Gamma_{\text{noise}}^{-1}(e-e_0)\right).$$

Using this information, Bayes' formula yields that the posterior distribution of x conditioned on y is

$$\pi(x \mid y) \propto \pi_{\text{pr}}(x)\pi_{\text{noise}}(y-Ax)$$
$$\propto \exp\left(-\frac{1}{2}(x-x_0)^T\Gamma_{\text{pr}}^{-1}(x-x_0) - \frac{1}{2}(y-Ax-e_0)^T\Gamma_{\text{noise}}^{-1}(y-Ax-e_0)\right).$$

The explicit form of the posterior distribution can be calculated from this expression. However, the factorization approach derived in the previous theorem avoids the tedious matrix manipulations required by a brute-force method. Since X and E are Gaussian, Y is also Gaussian, with

$$\mathbb{E}\begin{bmatrix}X\\Y\end{bmatrix} = \begin{bmatrix}x_0\\y_0\end{bmatrix}, \quad y_0 = Ax_0 + e_0.$$

Additionally, since

$$\mathbb{E}\left[(X-x_0)(X-x_0)^T\right] = \Gamma_{\text{pr}},$$

$$\begin{aligned}\mathbb{E}\left[(Y-y_0)(Y-y_0)^T\right] &= \mathbb{E}\left[(A(X-x_0)+(E-e_0))(A(X-x_0)+(E-e_0))^T\right]\\ &= A\Gamma_{\text{pr}}A^T + \Gamma_{\text{noise}},\end{aligned}$$

and furthermore,

$$\mathbb{E}\left[(X-x_0)(Y-y_0)^T\right] = \mathbb{E}\left[(X-x_0)(A(X-x_0)+(E-e_0))^T\right] = \Gamma_{\text{pr}}A^T,$$

it follows that

$$\operatorname{cov}\begin{bmatrix}X\\Y\end{bmatrix} = \mathbb{E}\left[\begin{bmatrix}X-x_0\\Y-y_0\end{bmatrix}\begin{bmatrix}X-x_0\\Y-y_0\end{bmatrix}^T\right] = \begin{bmatrix}\Gamma_{\text{pr}} & \Gamma_{\text{pr}}A^T\\ A\Gamma_{\text{pr}} & A\Gamma_{\text{pr}}A^T+\Gamma_{\text{noise}}\end{bmatrix}.$$

Therefore, the joint probability density of X and Y is given by

$$\pi(x,y) \propto \exp\left(-\frac{1}{2}\begin{bmatrix}x-x_0\\y-y_0\end{bmatrix}^T\begin{bmatrix}\Gamma_{\text{pr}} & \Gamma_{\text{pr}}A^T\\ A\Gamma_{\text{pr}} & A\Gamma_{\text{pr}}A^T+\Gamma_{\text{noise}}\end{bmatrix}^{-1}\begin{bmatrix}x-x_0\\y-y_0\end{bmatrix}\right).$$

Inspired by the previous example, the following result is mentioned.

Theorem 3.18. ([17]) Assume that $X : \Omega \to \mathbb{R}^n$ and $E : \Omega \to \mathbb{R}^m$ are mutually independent Gaussian random variables,

$$X \sim \mathcal{N}(x_0, \Gamma_{\text{pr}}), \quad E \sim \mathcal{N}(e_0, \Gamma_{\text{noise}}),$$

and $\Gamma_{\text{pr}} \in \mathbb{R}^{n\times n}$ and $\Gamma_{\text{noise}} \in \mathbb{R}^{k\times k}$ are positive definite. Consider the linear model

$$Y = AX + E$$

for a noisy measurement Y, where $A \in \mathbb{R}^{k\times n}$ is a known matrix. Then the posterior probability density of X given the measurement $Y = y$ is

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}(x-\bar{x})^\top \Gamma_{\text{post}}^{-1}(x-\bar{x})\right),$$

where

$$\bar{x} = x_0 + \Gamma_{\text{pr}}A^\top(A\Gamma_{\text{pr}}A^\top + \Gamma_{\text{noise}})^{-1}(y - Ax_0 - e_0),$$

and

$$\Gamma_{\text{post}} = \Gamma_{\text{pr}} - \Gamma_{\text{pr}}A^\top(A\Gamma_{\text{pr}}A^\top + \Gamma_{\text{noise}})^{-1}A\Gamma_{\text{pr}}.$$

Note that for assuming a Gaussian prior distribution and a Gaussian noise, then $\overline{x} = x_{MAP} = x_{CM}$ in the linear model, by the behavior of the Gaussian distribution.

Example 3.19. (*Example from [10]*) Consider the simple case in which the prior has covariance proportional to the identity matrix and mean zero, that is, $X \sim N(0, \gamma^2 I)$. This prior is referred to as the Gaussian white noise prior. Similarly, assume that the noise is white noise, $E \sim N(0, \sigma^2 I)$. In this specific case, the following holds:

$$x = \gamma^2 A^T(\gamma^2 AA^T + \sigma^2)^{-1}y = A^T(AA^T + \alpha I)^{-1}y,$$

where $\alpha = \frac{\sigma^2}{\gamma^2}$. By using the singular value decomposition of the matrix A, it is straightforward to verify that

$$x = A^T(AA^T + \alpha I)^{-1}y = (A^TA + \alpha I)^{-1}A^Ty.$$

Thus, the centerpoint of the posterior distribution corresponds to the Tikhonov regularized solution of the equation $Ax = y$, with regularization parameter α, see the section 2.1.1.2. This re-interpretation provides additional insight into the choice of the parameter α: it is the ratio of the noise variance to the prior variance.

Chapter 4

Markov Chain Monte Carlo methods

This chapter explores the use of Markov chain Monte Carlo (MCMC) methods for addressing inverse problems through posterior probability distributions. While abstract definitions of solutions in terms of posterior distributions are foundational, practical applications require tools to effectively explore these distributions. Traditional numerical integration methods, such as quadrature, are computationally infeasible for high-dimensional parameter spaces and rely on prior knowledge of the distribution's support, which is often unavailable. MCMC methods offer an alternative by generating sample points guided by the probability density itself, enabling approximate integration through Monte Carlo techniques. This approach provides a powerful framework for overcoming computational challenges in high-dimensional spaces. In order to see a wide approach, consult [10,17].

4.1 Basic ideas

Let μ be a probability measure defined over $\mathbb{R}^n$. Additionally, consider a scalar or vector-valued measurable function f, which is integrable over $\mathbb{R}^n$ with respect to μ, meaning that $f \in L^1(\mu(\mathrm{d}x))$. The goal is to estimate the integral of f with respect to μ. In numerical quadrature methods, a collection of support points $x_j \in \mathbb{R}^n$, where $1 \leq j \leq N$, is selected along with corresponding weights w_j to approximate the integral:

$$\int_{\mathbb{R}^n} f(x)\mu(\mathrm{d}x) \approx \sum_{j=1}^{N} w_j f(x_j).$$

Quadrature methods are generally designed to achieve high accuracy for functions belonging to a finite-dimensional function space, which is often composed of

polynomials up to a certain degree.

In the case of Monte Carlo integration, the support points x_j are randomly sampled from a probability density, and the weights w_j are determined according to the measure μ. Ideally, these points should be sampled directly from the probability distribution associated with μ. Specifically, let $X \in \mathbb{R}^n$ be a random variable whose probability distribution corresponds to μ. If a random generator capable of producing independent realizations of X were available, it would be possible to generate a representative set of points distributed according to μ. Assuming that $\{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^n$ forms such a sample set, the integral of f can be approximated using the average (consult the references [10,12,18]):

$$\int_{\mathbb{R}^n} f(x)\mu(\mathrm{d}x) = \mathbb{E}[f(X)] \approx \frac{1}{N}\sum_{j=1}^{N} f(x_j).$$

Example 4.1. For instance, consider the integral $\int_0^1 \frac{1}{12}(x-\frac{1}{2})^2 dx = 1$. This integral can be seen as an expectation of the function $f(x) = \frac{1}{12}(x-\frac{1}{2})^2$ over the measure given by the uniform distribution over the interval $(0,1)$. A number n of sample random points x_i with $i = 1, \ldots, n$ are generated from the distribution $U(0,1)$ and then the following expectation is calculated $\mathbb{E}[f(X)]$.

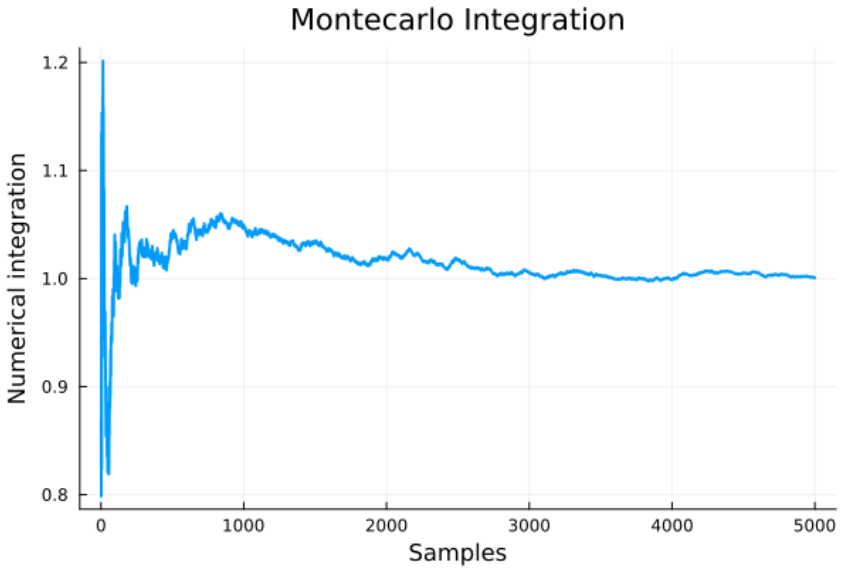


Figure 4.1: Monte Carlo integration for the integral $\int_0^1 \frac{1}{12}(x-\frac{1}{2})^2 dx$. Figure elaborated by the author.

As can be seen in the Figure 4.1, as more points are sampled over the interval $(0,1)$ the integral converges to its true value. This example just illustrates the importance

and ease of use of Monte Carlo methods. The applicability of this type of methods is much more useful in high dimensional problems.

Markov Chain Monte Carlo (MCMC) methods provide a systematic approach to generating a sample ensemble that satisfies this approximation. To achieve this, fundamental concepts from probability theory are required (consult [17]).

Definition 4.2. ([10,17]) Let $\mathcal{S}$ denote a σ-algebra defined over Ω. A function

$$P : \Omega \times \mathcal{S} \to [0, 1]$$

is called a probability transition kernel if:

1. For every $B \in \mathcal{S}$, the mapping $x \mapsto P(x, B)$ is a measurable function over Ω.
2. For each $x \in \Omega$, the function $B \mapsto P(x, B)$ defines a probability distribution.

To get a better intuition of the above definition, consider the following examples.

Example 4.3. Consider the time-homogeneous Markov chain $\{X_n : n \in \mathbb{N}\}$ given by the transition graph shown in the figure 4.2.

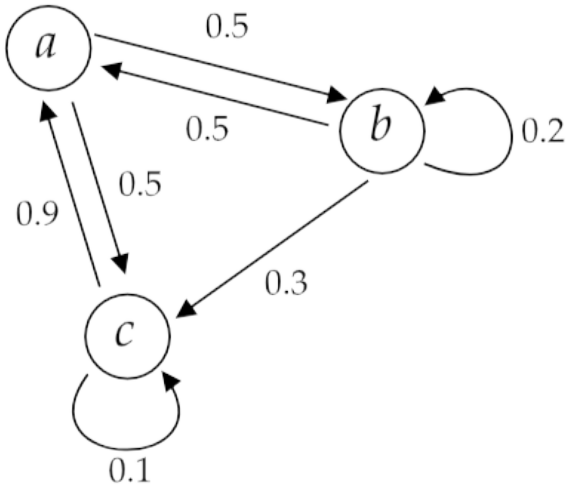


Figure 4.2: Transition graph of a Markov chain with three states. Figure elaborated by the author.

Let $\Omega = \{a, b, c\}$ the state set and $\mathcal{S} := \{X | X \subseteq \Omega\}$ a σ-algebra defined over Ω as the power set. For every $x \in \Omega$ and $B \in \mathcal{S}$, consider

$$P(x, B) = \mathbb{P}(X_{n+1} \in B \mid X_n = x),$$

for any value of $n \in \mathbb{N}$. The function P corresponds to a probability transition kernel:

- If $B = \{a, b\}$ is fixed then the map $x \mapsto P(x, B)$ is a measurable function since:

$$P^{-1}\{[0, x]\} = \begin{cases} \emptyset & 0 \leq x < 0.5 \\ \{a\} & 0.5 \leq x < 0.7 \\ \{a, b\} & 0.7 \leq x < 0.9 \\ \Omega & 0.9 \leq x \leq 1 \end{cases}$$

- If $x = a$ is fixed the the map $B \mapsto P(x, B)$ is a random variable and it is fully described by the following equalities:

$$\begin{aligned} P(a, \emptyset) &= 0, & P(a, \{a, b\}) &= 0.5, \\ P(a, \{a\}) &= 0, & P(a, \{b, c\}) &= 1, \\ P(a, \{b\}) &= 0.5, & P(a, \{a, c\}) &= 0.5 \\ P(a, \{c\}) &= 0.5, & P(a, \Omega) &= 1. \end{aligned}$$

Example 4.4. Consider a time-homogeneous Markov chain $\{X_n : n \in \mathbb{N}\}$ where the property

$$\mathbb{P}(X_{n+1} \in B \mid X_n = x) = \int_B f(x'|x)dx',$$

where B is a Borel set over $\mathbb{R}^n$ and f is a probability density funcion of the Normal distribution with mean $\mu = x$ and variance $\sigma^2 = x^2 + 1$ (in that manner, x can be seen as a parameter). This conditional probability allows to construct a probability transition kernel.

Definition 4.5. A discrete-time stochastic process is an ordered sequence $\{X_j\}_{j=1}^{\infty}$ of real random variables $X_j \in \mathbb{R}^n$. A time-homogeneous Markov chain with transition kernel P is a stochastic process $\{X_j\}_{j=1}^{\infty}$ satisfying the property:

$$\mu_{X_{j+1}}(B_{j+1}|x_1, \ldots, x_j) = \mu_{X_{j+1}}(B_{j+1}|x_j) = P(x_j, B_{j+1}),$$

where μ_X is the measure given by the random variable X. This condition states that the probability of X_{j+1} belonging to B_{j+1}, given prior observations $X_1 = x_1, \ldots, X_j = x_j$, depends only on the most recent state $X_j = x_j$. This property is often summarized by the statement that "the future depends on the past only through the present."

It is still necessary to introduce a few concepts regarding the transition kernels, this objects are satisfies analogues properties and definitions with respect a matrix of a Markov chain with a finite state space. Similar to the section 2.3, given a probability measure μ, the transition kernel P is said to be *irreducible* (with respect to μ) if for each $x \in \mathbb{R}^n$ and $B \in \mathcal{B}$ with $\mu(B) > 0$, there exists an integer k such that $P^{(k)}(x, B) > 0$. This means that, irrespective of the starting point, the Markov chain generated by the transition kernel P will visit any set of positive measure with a positive probability. Let P be an irreducible kernel. It is said that P is *periodic* if, for some integer $m \geq 2$, there exists a set of disjoint nonempty sets $\{E_1, \ldots, E_m\} \subset \mathbb{R}^n$ such that for all $j = 1, \ldots, m$ and all $x \in E_j$, we have $P(x, E_{j+1(\mathrm{mod}\ m)}) = 1$. In other words, a periodic transition kernel generates a Markov chain that remains in a periodic loop indefinitely. A kernel P is termed *aperiodic* if it is not periodic. Consult the section 2.3 for further treatment.

The following result is of significant importance for MCMC methods. The proof of this theorem will be omitted (this proof can be encountered in [10,17]).

Theorem 4.6. Let μ be a probability measure on $\mathbb{R}^n$ and $\{X_j\}$ a time-homogeneous Markov chain with transition kernel P. Assume that μ is an invariant measure of the transition kernel P, and that P is irreducible and aperiodic. Then, for all $x \in \mathbb{R}^n$,

$$\lim_{N\to\infty} P^{(N)}(x, B) = \mu(B) \quad \text{for all } B \in \mathcal{B},$$

where $P^{(N)}$ means the composition of P a number of N times, and for $f \in L^1(\mu(dx))$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} f(X_j) = \int_{\mathbb{R}^n} f(x)\, \mu(dx) \quad \text{almost surely.}$$

This theorem provides a clear indication of how to explore a given probability distribution: it is necessary to construct an invariant, aperiodic, and irreducible transition kernel P, and then draw a sequence of sample points $x_1, x_2, \ldots$ using this kernel. This will be the central part of the following section.

4.2 Metropolis-Hastings algorithm

Let μ denote the target probability distribution in $\mathbb{R}^n$ that the sampling algorithm seeks to explore. To avoid measure-theoretic complexities, it is assumed that

μ is absolutely continuous with respect to the Lebesgue measure, i.e., $\mu(dx) = \pi(x)dx$. The goal is to determine a transition kernel $P(x, B)$ such that μ is its invariant measure. This section is based on [10,12,17].

Let P represent any transition kernel. For a given point $x \in \mathbb{R}^n$, it is postulated that the kernel either proposes a move to another point $y \in \mathbb{R}^n$ or it proposes no move from x. This allows the kernel to be decomposed into two parts:

$$P(x, B) = \int_B K(x, y)\, dy + r(x)\chi_B(x),$$

where χ_B is the characteristic function of the set $B \in \mathcal{B}$. Although $K(x, y) \geq 0$ is a density, $K(x, y)dy$ can be interpreted as the probability of moving from x to the infinitesimal set dy at y, while $r(x) \geq 0$ represents the probability that x remains stationary. The characteristic function $\chi_B(x)$ appears because if $x \notin B$, the only way for x to reach B is by making a move.

The condition $P(x, \mathbb{R}^n) = 1$ implies that

$$r(x) = 1 - \int_{\mathbb{R}^n} K(x, z)\, dz.$$

For $\mu(dx) = \pi(x)dx$ to be an invariant measure of P, the identity

$$\mu P(B) = \mu(B),$$

and that implies

$$\begin{aligned}
\mu P(B) &= \int_{\mathbb{R}^n} P(x, B)\mu(dx) \\
&= \int_{\mathbb{R}^n} P(x, B)\pi(x)dx && (\mu \text{ is absolutely continuous}) \\
&= \int_{\mathbb{R}^n} \left[\int_B K(x, y)\, dy + r(x)\chi_B(x)\right] \pi(x)dx && (\text{decomposition of } P) \\
&= \int_{\mathbb{R}^n} \int_B K(x, y)dydx + \int_{\mathbb{R}^n} r(x)\chi_B(x)\pi(x)dx && (\text{linearity}) \\
&= \int_{\mathbb{R}^n} \int_B K(x, y)dydx + \int_B r(x)\pi(x)dx && (\text{definition of } \chi_B) \\
&= \int_{\mathbb{R}^n} \int_B K(x, y)dydx + \int_B r(y)\pi(y)dy && (\text{change of variable}) \\
&= \int_B \int_{\mathbb{R}^n} K(x, y)dxdy + \int_B r(y)\pi(y)dy, && (\text{Fubini theorem})
\end{aligned}$$

thus

$$\int_B \int_{\mathbb{R}^n} K(x,y)dxdy + \int_B r(y)\pi(y)dy = \int_B \pi(y)dy$$
$$\int_B \int_{\mathbb{R}^n} K(x,y)dxdy = \int_B \pi(y)(1-r(y))dy$$

must hold for all $B \in \mathcal{B}$, implying

$$\pi(y)(1-r(y)) = \int_{\mathbb{R}^n} \pi(x)K(x,y)\,dx.$$

By formula $r(x) = 1 - \int_{\mathbb{R}^n} K(x,z)\,dz$, this is equivalent to

$$\int_{\mathbb{R}^n} \pi(y)K(y,x)\,dx = \int_{\mathbb{R}^n} \pi(x)K(x,y)\,dx.$$

This condition is known as the weak balance equation. In particular, if K satisfies the detailed balance equation

$$\pi(y)K(y,x) = \pi(x)K(x,y)$$

for all pairs $x, y \in \mathbb{R}^n$, then the weak balance equation is automatically satisfied. The decomposition of the transition kernel P and the detailed balance equations form the foundation for constructing the Markov chain transition kernels used in stochastic sampling. In the Metropolis–Hastings algorithm, the goal is to construct a transition kernel K that satisfies the detailed balance equation (according to [10,17]).

Let $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ be a given function with the property that $\int_{\mathbb{R}^n} q(x,y)\,dy = 1$. The kernel q is called the **proposal distribution** or **candidate-generating kernel**, for reasons explained later. Such a function q defines a transition kernel

$$Q(x,A) = \int_A q(x,y)\,dy.$$

If q satisfies the detailed balance equation, then define $K(x,y) = q(x,y)$, $r(x) = 0$, and the task is complete. Otherwise, it is necessary to adjust the kernel by a multiplicative factor and define

$$K(x,y) = \alpha(x,y)q(x,y),$$

where α is a correction term to be determined. Assume that for some $x, y \in \mathbb{R}^n$, instead of detailed balance,

$$\pi(y)q(y,x) < \pi(x)q(x,y).$$

The objective is to choose α so that

$$\pi(y)\alpha(y,x)q(y,x) = \pi(x)\alpha(x,y)q(x,y).$$

This is achieved by setting

$$\alpha(y,x) = 1, \quad \alpha(x,y) = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} < 1.$$

By swapping x and y, it is observed that the kernel K defined by $K(x,y) = \alpha(x,y)q(x,y)$ satisfies the detailed balance equations if α is defined as

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right).$$

This transition kernel is called the **Metropolis–Hastings kernel.**

The above derivation does not provide much insight into the implementation of the method. Fortunately, the algorithm is relatively simple in practice, typically carried out through the following steps (taking the ideas of [10,17]):

1. Pick the initial value $x_1 \in \mathbb{R}^n$ and set $k = 1$.
2. Draw $y \in \mathbb{R}^n$ from the proposal distribution $q(x_k, y)$ and calculate the acceptance ratio
$$\alpha(x_k,y) = \min\left(1, \frac{\pi(y)q(y,x_k)}{\pi(x_k)q(x_k,y)}\right).$$
3. Draw $t \in [0,1]$ from the uniform probability distribution.
4. If $\alpha(x_k,y) \geq t$, set $x_{k+1} = y$, otherwise set $x_{k+1} = x_k$. When $k = K$, the desired sample size is reached, and the process stops. Otherwise, increment k to $k+1$ and return to step 2.

Before presenting some examples, a few remarks are in order. First, if the candidate-generating kernel is symmetric, i.e.,

$$q(x,y) = q(y,x) \quad \text{for all } x, y \in \mathbb{R}^n,$$

then the acceptance ratio simplifies to

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right).$$

Thus, moves that go towards higher probabilities are accepted immediately, while some moves that take us to lower probabilities may also be accepted. When the candidate-generating kernel is symmetric, this kernel is called **Metropolis kernel**.

An important yet challenging issue is the stopping criterion, i.e., how to decide when the sample size is sufficient. This question, as well as convergence issues in general, will be addressed later through an example.

Example 4.7. (*Example inspirated in [10]*) Consider the function

$$f(x,y) = -\frac{1}{20}\left[\left(\frac{1}{10}x^2 - y\right)^2 + \frac{1}{5}(x^2+y^2)\right]$$

and suppose that the distribution $\pi(x,y)$ is proportional to $\exp(f(x,y))$. The contour plot is shown in the following figure.

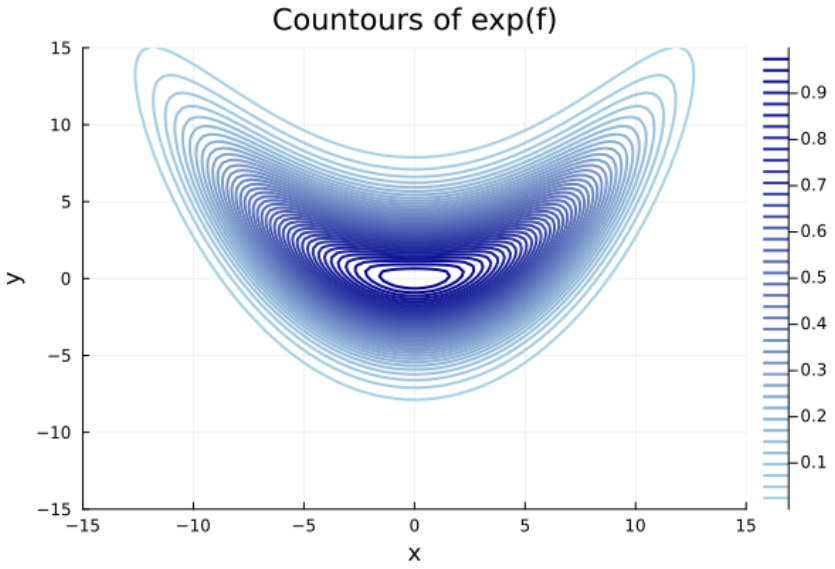


Figure 4.3: Contour plot of the function $\exp(f(x,y))$. Figure elaborated by the author using Julia (consult the repository [21]).

The Metropolis Hastings algorithm will be run with 10000 iterations taking into account that the candidate-generating kernel is selected as a multivariate normal distribution centered on the last accepted sample and with variance equal to a fixed scalar multiple ($\delta = 0.1$) of the identity matrix and, the initial point is $(-15, 15)$. The results are shown in the figures below.

When the candidate-generating kernel is chosen as a normal distribution centered on the last accepted sample and with fixed variance the algorithm is called **Metropolis Hastings Random Walk (MHRW)** (according to [17]).

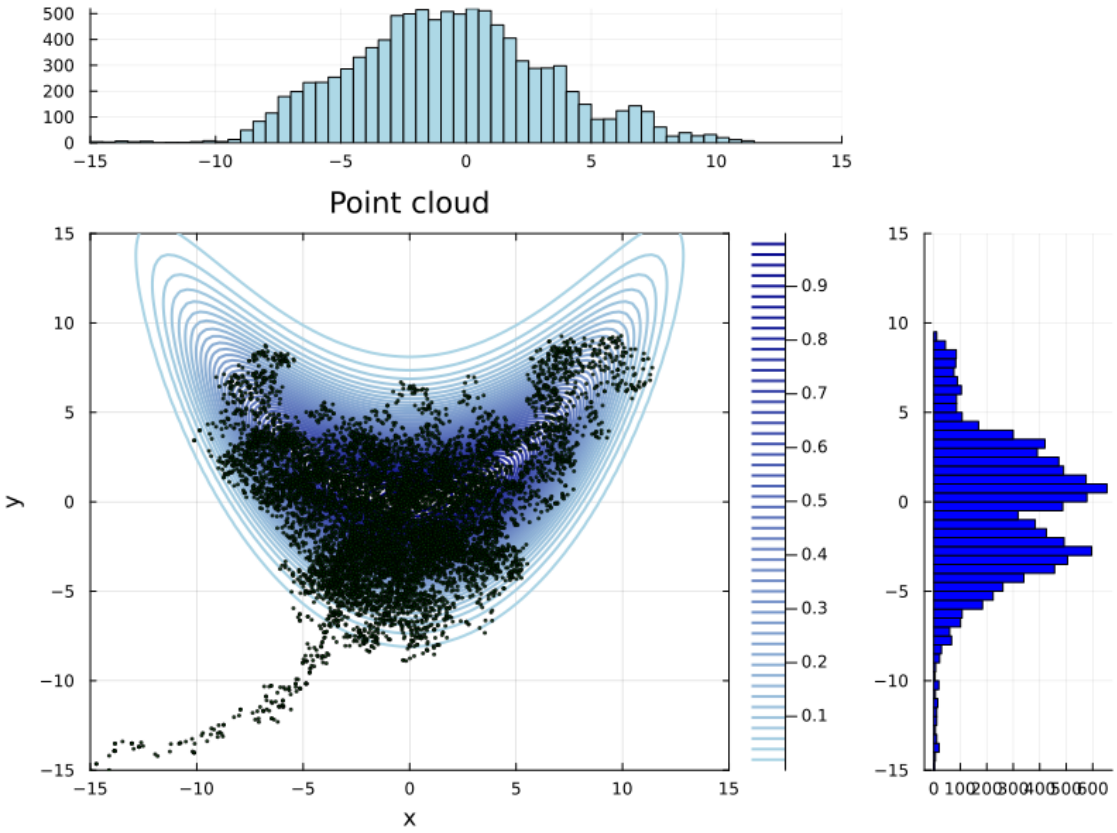


Figure 4.4: Point cloud plot to visualize the samples obtained from the MH execution. Figure elaborated by the author using Julia (consult the repository [21]).

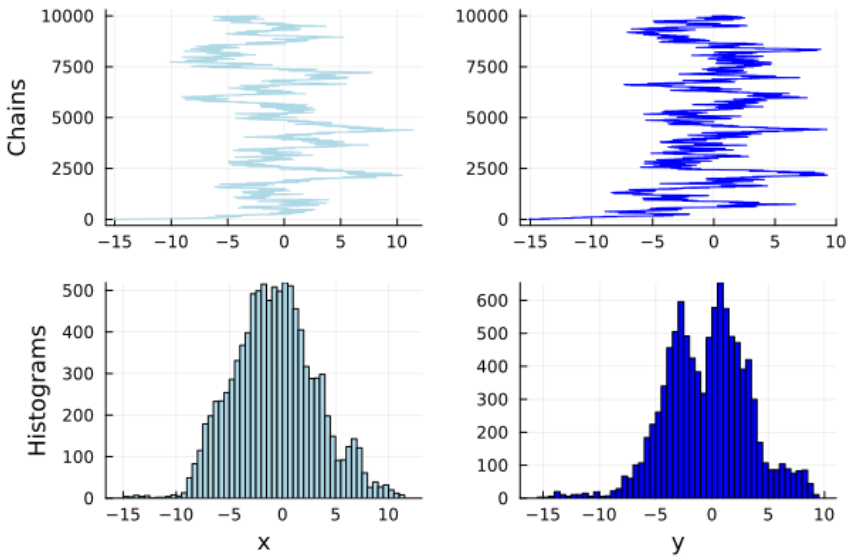


Figure 4.5: Random walks of the variables and theirs respectives histograms. Figure elaborated by the author using Julia (consult the repository [21]).

Note that the figure 4.4 shows that most of the samples are located where the largest values of the function $\exp(f(x, y))$ are found, and the histograms show the marginal distributions of the random variables X and Y. This method refers to random walks since the values of the random variables of interest follow an erratic path when the algorithm is executed, as can be seen in the graphs of the Figure 4.5.

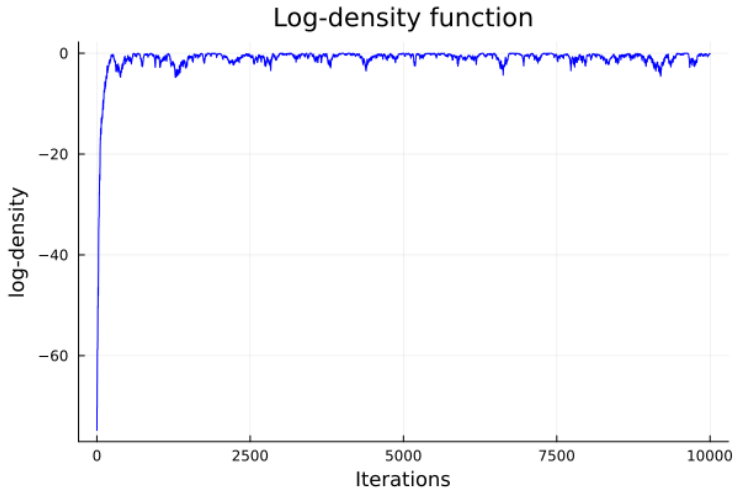


Figure 4.6: Log-density function evaluated in the samples from the distribution $\pi(x, y) \propto \exp(f(x, y))$. Figure elaborated by the author using Julia (consult the repository [21]).

Figure 4.6 shows that the probabilities of the samples as each iteration progresses have a stabilization period, something that could indicate convergence of the MH algorithm (see [17]).

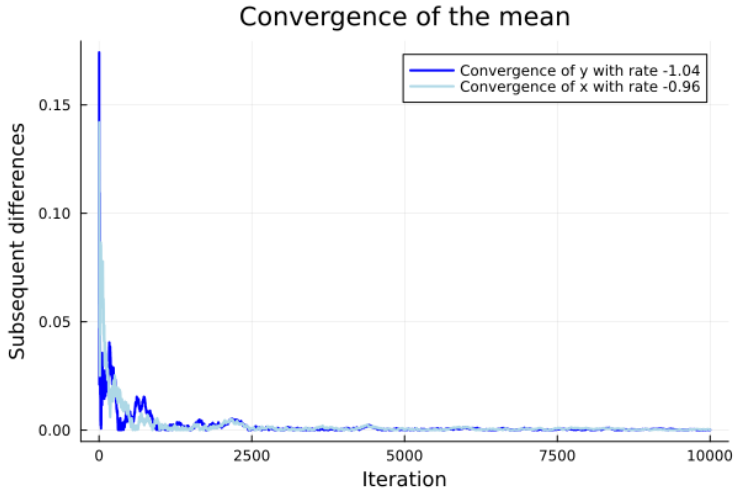


Figure 4.7: Convergence plot of the sampling mean in both random variables. Figure elaborated by the author using Julia (consult the repository [21]).

Finally, the figure 4.7 shows the above theorem 4.6 by calculating the difference between the sample means for consecutive iterations of the algorithm. As can be seen, the convergence of this algorithm is efficient since it does not need too many iterations. If it is fitted an exponential model $Ae^{B\cdot\text{iterations}}$ to the difference between sample means, the parameter B represents the convergence rate of this algorithm, as mentioned in [17].

Another important information to give from the algorithm is the acceptance rate, which was 94.27 % for this algorithm run.

Example 4.8. (**Mixture of Gaussians**) Consider the functions

$$f_1(x,y) = \frac{1}{50}[(x-5)^2 + (y-5)^2] \quad \text{and} \quad f_2(x,y) = \frac{1}{50}[(x+10)^2 + (y+5)^2]$$

and suppose that a density function satisfies $\pi(x,y) \propto \exp(-f_1(x,y)) + \exp(-f_2(x,y))$. The MHRW algorithm will be executed over this density with 10000 iterations, initial point $(20,-20)$ and with the parameter δ taking different values in the set $\{0.01, 0.25, 0.5, 1\}$. The result is shown in the Figure 4.8.

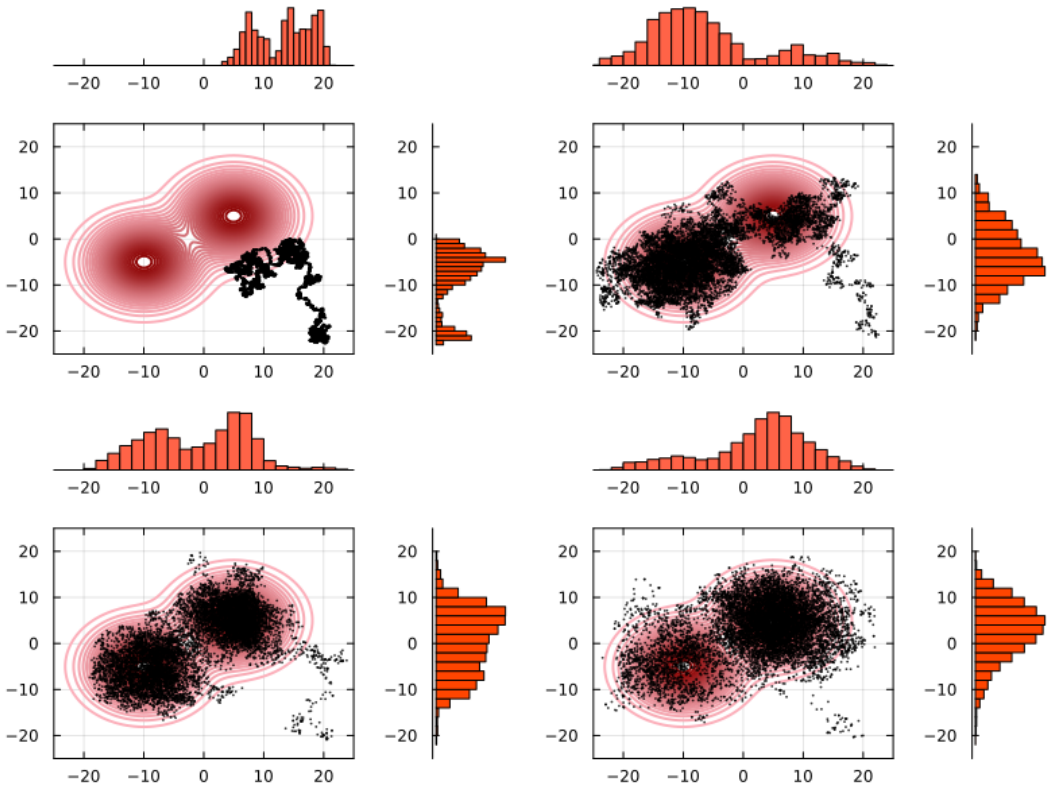


Figure 4.8: MHRW algorithm executed in a mixture of Gaussian with different variances. Figure elaborated by the author using Julia (consult the repository [21]).

As can be seen, different values for the δ parameter affect the motion of the Markov chain by restricting the step length. For example for $\delta = 0.01$ the path has little mobility and this is reflected in samples that do not correspond to the target distribution.

The two examples above show something that is important to highlight. Note that the MH algorithm has a period where it "drifts" in the distribution space until it approaches and stays in the regions with the highest probability density. This period is called the **warm-up (or burning) period** and is usually taken as 10 % of the total iterations and samples obtained during this period are discarded (see [17]).

4.2.1 Gibbs sampler

This description of the method is based on [8,10]. A variant of the sampling algorithm is obtained when the candidate-generating kernel is defined using the density π directly along with a block partitioning of the vectors in $\mathbb{R}^n$. Let $I = 1, 2, \ldots, n$ be the index set of $\mathbb{R}^n$, and consider a partitioning of this set into m disjoint nonempty subsets, denoted by $I = \bigcup_{j=1}^{m} I_j$. Each subset I_j contains k_j elements, such that $k_j = \#I_j$, ensuring that $k_1 + \cdots + k_m = n$. This partitioning allows $\mathbb{R}^n$ to be expressed as the Cartesian product $\mathbb{R}^n = \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m}$. Consequently, any vector $x \in \mathbb{R}^n$ can be written as

$$x = [x_{I_1}; \ldots; x_{I_m}] \in \mathbb{R}^n, \quad x_{I_j} \in \mathbb{R}^{k_j}.$$

The notation x_{-I_j} is introduced to denote the vector x with the elements corresponding to the subset I_j removed

$$x_{-I_j} = [x_{I_1}; \ldots; x_{I_{j-1}}, x_{I_{j+1}}; \ldots; x_{I_m}].$$

This convention simplifies notation when working with conditional distributions. If the random variable $X \in \mathbb{R}^n$ follows the probability density function π, then the conditional probability density of the block X_{I_i} given all other blocks is expressed as

$$\pi(x_{I_i} \mid x_{-I_i}) = C_i \pi(x_{I_1}, \ldots, x_{I_i - 1}, x_{I_i}, x_{I_i + 1}, \ldots, x_{I_m}),$$

where C_i is a normalization constant. Using this formulation, the transition kernel K is defined as

$$K(x, y) = \prod_{i=1}^{m} \pi(y_{I_i} \mid y_{I_1}, \ldots, y_{I_{i-1}}, x_{I_{i+1}}, \ldots, x_{I_m}).$$

This kernel does not generally satisfy the detailed balance condition, but it satisfies a weaker balance condition, which is sufficient for convergence.

To implement the Gibbs sampler algorithm, the following steps are performed:

1. Select an initial value $x_1 \in \mathbb{R}^n$ and set $k = 1$.
2. Set $x = x_k$. For $1 \leq j \leq m$, sample $y_{I_j} \in \mathbb{R}^{k_j}$ from the conditional distribution:
$$\pi(y_{I_j} \mid y_{I_1}, \ldots, y_{I_{j-1}}, x_{I_{j+1}}, \ldots, x_{I_m}).$$
3. Set $x_{k+1} = y$. If $k = K$, the desired sample size is reached, and the procedure terminates. Otherwise, increment $k \to k + 1$ and repeat from step 2.

The fundamental distinction between the Gibbs sampler and the MH algorithm lies in the acceptance criterion: while the latter algorithm requires an acceptance-rejection step, the former always accepts the proposed sample. However, drawing from the conditional distributions in the Gibbs sampler can be computationally intensive, depending on the complexity of π.

Example 4.9. Suppose that $\pi(x, y) \propto exp(-f(x, y))$ where
$$f(x, y) = -\frac{1}{20}(x^2y^2 + x^2 + y^2 - 8x - 8y).$$
Note that the marginal distributions fulfill that
$$\pi(x|y) = g(y) \exp\left(-\frac{y^2+1}{20}\left(x - \frac{4}{1+y^2}\right)^2\right),$$
$$\pi(y|x) = g(x) \exp\left(-\frac{x^2+1}{20}\left(x - \frac{4}{1+x^2}\right)^2\right).$$
These marginal distributions corresponds to normal distributions that are easy to sample:
$$x|y \sim N\left(\mu = \frac{4}{1+y^2}, \sigma = \sqrt{\frac{10}{1+y^2}}\right)$$
$$y|x \sim N\left(\mu = \frac{4}{1+x^2}, \sigma = \sqrt{\frac{10}{1+x^2}}\right).$$
The contour plot of the distribution is shown in Figure 4.9.

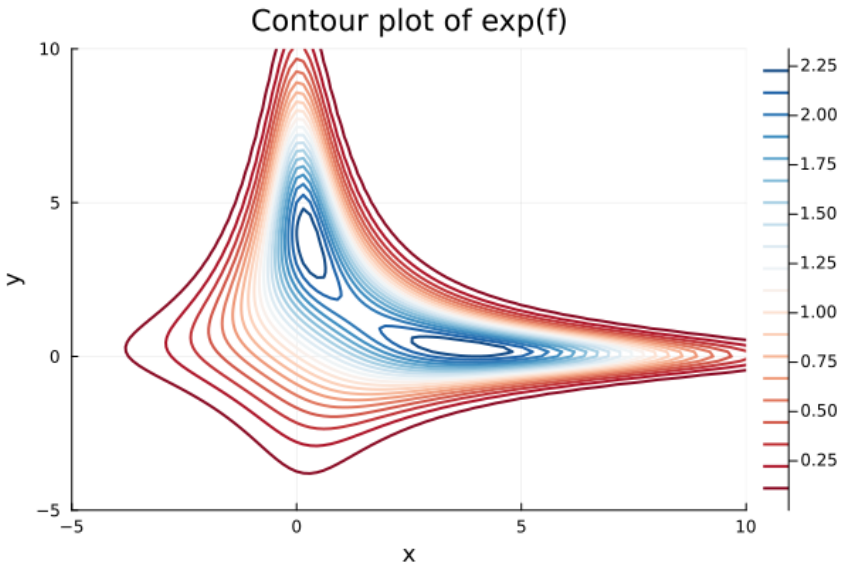


Figure 4.9: Contour plot of the distribution π. Figure elaborated by the author using Julia (consult the repository [21]).

Some of the samples obtained from the algorithm are shown in the Figure 4.10. It is important that the update in each iteration is done component by component, as reflected in the graph. Further comments on this sampling method will be made in the example 4.10.

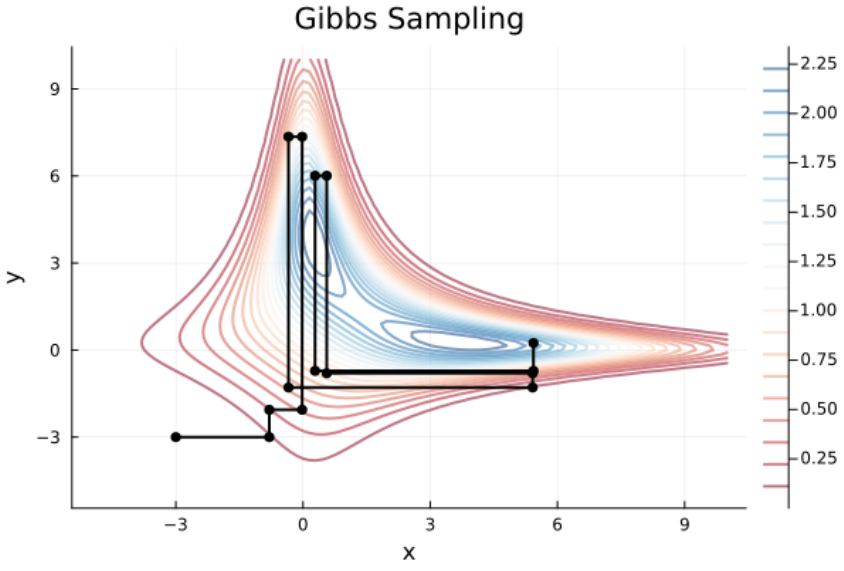


Figure 4.10: Some samples from the Gibbs sampler algorithm. Figure elaborated by the author using Julia (consult the repository [21]).

4.2.2 Two-walk

This method is a modification of the MHRW and is inspired by the t-walk algorithm in [16]. For a given objective function, such as a posterior distribution, denoted by $\pi(x)$ for $x \in X$, where X is an n-dimensional subset of $\mathbb{R}^n$, a new objective function is constructed as $f(x, x^+) = \pi(x)\pi(x^+)$ in the corresponding product space $X \times X$. While a general proposal follows the form $q\{(y, y^+)|(x, x^+)\}$, the approach considers two restricted proposals:

$$(y, y^+) = \begin{cases} (x, h(x^+, x)), & \text{with probability 0.5,} \\ (h(x, x^+), x^+), & \text{with probability 0.5,} \end{cases}$$

where $h(x, x^+)$ represents a random variable used to form the proposal. This implies that, at each step, only one of x or x^+ is modified. However, rather than considering two independent parallel chains in X, the entire process remains within the product space $X \times X$. Four distinct proposals are randomly selected, each characterized by a specific function $h(\cdot, \cdot)$. The procedure involves first selecting an option from the previous equation and then generating the proposal (y, y^+) by sampling from the corresponding h function.

Within the Metropolis-Hastings framework, it is necessary to compute the acceptance ratio. Defining $g(\cdot|x, x^+)$ as the density function of $h(x, x^+)$, the acceptance ratio is given by:

$$q\{(y, y^+)|(x, x^+)\} = \frac{\pi(y^+)}{\pi(x^+)} \frac{g(x^+|y^+, x)}{g(y^+|x^+, x)}$$

for the first case, and

$$q\{(y, y^+)|(x, x^+)\} = \frac{\pi(y)}{\pi(x)} \frac{g(x|y, x^+)}{g(y|x, x^+)}$$

for the second case. Notably, restricting the proposal to h ensures that only a single evaluation of the target density is required in either scenario.

In this method the candidate-generating kernel is the Gaussian distribution with mean the first argument of the function h and variance equal to the distance between the two arguments of the function h plus some fixed parameter δ_0.

Example 4.10. (**Comparison between MCMC methods** (*similar to examples in [10,17]*)) The three MCMC methods are executed for the example 4.9. The following table shows the result of convergence rates for the mean of X, the mean of Y and the mean of the joint distribution of the vector (X, Y).

Table 4.1: Comparison of rate of convergence of the three methods of sampling. Table elaborated by the author.

Rate convergence	Metropolis-Hastings	Gibbs sampler	Two-walk
X	-1.07	-0.99	-0.91
Y	-1.03	-0.97	-1.01
(X, Y)	-1.05	-0.98	-0.96

From the Table 4.1, it can be seen that, in this example, the Metropolis-Hastings algorithm performs better. The following two figures (4.11 and 4.12) first show the logarithm of the probabilities of the selected samples, and then compare the histograms for X and Y obtained from the three methods.

Figure 4.11: Difference of the log-probabilities of the samples in each algorithm. Figure elaborated by the author using Julia (consult the repository [21]).

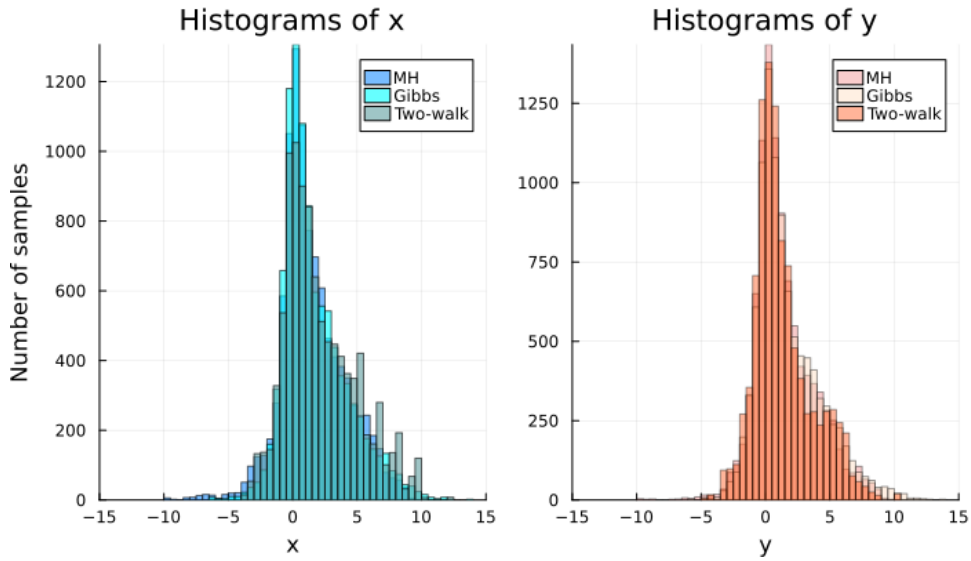


Figure 4.12: Histogram of the samples obtained in each MCMC method. Figure elaborated by the author using Julia (consult the repository [21]).

The parameter δ of the MH algorithm is fixed in $\delta = 0.2$ and the parameter δ_0 of the Two-walk algorithm is taken equal to $\delta_0 = 1$. The acceptance rate of the MH is 82.72 % while the acceptance rate of the Two-walk algorithm is 38.09 %; this difference in acceptance rates indicates that the Two-walk algorithm moves faster to the region with higher probability density compared to the MH algorithm.

Chapter 5

Applications

The use of Markov Chain Monte Carlo (MCMC) methods has proven to be a powerful tool in various applications where parameter space exploration is essential. This chapter presents different problems in which these approaches are particularly useful, including parameter estimation in ordinary differential equation (ODE) models, optimization via exploration, and least squares fitting. Furthermore, the application of these methods in the calibration of the TOMGRO model, which is used to study crop growth, is analyzed, highlighting its description, the data employed, and the obtained results (consult each application for the references).

5.1 Optimization via exploration

Optimization via exploration is a strategy that seeks optimal solutions by iteratively adjusting parameters based on an exploratory search of the parameter space. A key aspect of this method is the adjustment factor α, which introduces a "discrete gradient" that directs the exploration process toward an optimal point. This mechanism allows the algorithm to navigate the parameter space efficiently, balancing exploration and convergence. To illustrate these concepts, examples in both one and two dimensions will be presented, demonstrating how the choice of α influences the optimization trajectory and the final solution. Unlike the other sections, some of the examples presented in this section will not make use of Bayesian inference (this section is proposed by the author and collaborators).

Consider the minimization problem

$$\min_{\boldsymbol{x}\in\Omega\subseteq\mathbb{R}^n} f(\boldsymbol{x})$$

One approach to solving this problem using Markov Chain Monte Carlo (MCMC) methods is to employ the Metropolis algorithm by considering a distribution that satisfies the following proportionality

$$\pi(\boldsymbol{x}) \propto \exp(-\beta f(\boldsymbol{x})),$$

where $\beta > 0$ is a parameter specifically chosen for each function f.

In the Metropolis algorithm, the acceptance probability α is given by

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left(1, \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})}\right).$$

Utilizing the fact that $\pi(\boldsymbol{x}) \propto \exp(-\beta f(\boldsymbol{x}))$, the acceptance probability α can be rewritten as

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left(1, \exp(-\beta[f(\boldsymbol{y}) - f(\boldsymbol{x})])\right).$$

The term $[f(\boldsymbol{y}) - f(\boldsymbol{x})]$ in this expression will be referred to as the "discrete gradient" , as it quantifies the absolute change of the function f between the points $\boldsymbol{x}$ and $\boldsymbol{y}$.

In some cases, the feasible space Ω can be explored using the candidate-generating kernel Q, leading to constrained minimization problems. The following examples illustrate the application of this approach in one and two dimensions, incorporating Bayesian inference in certain cases.

Example 5.1. Consider the function

$$f(x) = \left(\frac{3x}{10}\right)^4 + 10\sin\left(\frac{3\pi}{10}x\right),$$

and consider the problem of minimizing this function over the whole set $\mathbb{R}$. If the distribution $\pi(x) \propto \exp(-\beta f(x))$ is taken with $\beta = \frac{1}{10}$ and, run the Metropolis algorithm with 10000 iterations, initial point $x_0 = 20$ and normal candidate generating kernel with variance $\delta = 5$, obtaining the results shown in the figure below.

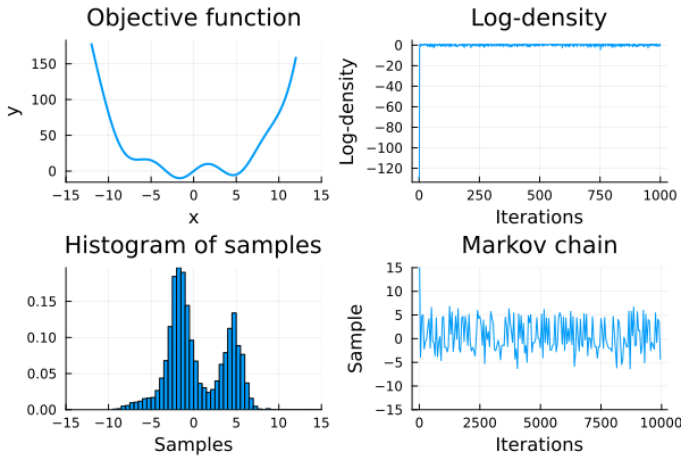


Figure 5.1: Result of the Metropolis algorithm in an optimization problem. Figure elaborated by the author using Julia (consult [21]).

Consider the same minimization problem but now restricting the domain to the positive values of the decision variable using Bayesian inference: consider the *a posteriori* distribution $\pi_{post}(x)$ proportional to the product of the term $\exp(-\beta f(x))$ ($\beta = \frac{1}{10}$) and the *a priori* distribution $\pi_{pr}(x)$ taken as a Gamma distribution with scale parameter 2 and shape parameter 1.

If the Metropolis algorithm is executed with 10000 iterations, initial point $x_0 = 20$ and normal candidate generating kernel with variance $\delta = 5$, the results shown in the figure below are obtained.

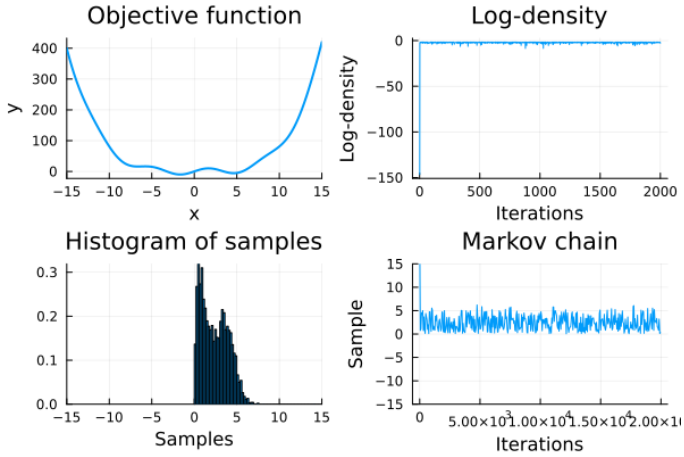


Figure 5.2: Result of the Metropolis algorithm in an optimization problem mixed with Bayesian inference. Figure elaborated by the author using Julia (consult [21]).

In the previous example, the condition that $x > 0$ may have been taken into account by starting from a positive initial point in the algorithm and also using a

candidate-generating kernel that only generates proposals that satisfy the conditions of the feasible set of the optimization problem.

In the following example, a discrete feasible set is taken, i.e., it only takes values in the set $\mathbb{Z}$. The Metropolis algorithm is going to be executed with 10000 iterations and with the candidate-generating kernel

$$\boldsymbol{Q}_x = \mathbb{B} + x - n,$$

where $n \in \mathbb{N}$ (in this case $n = 10$) is a parameter and $\mathbb{B}$ is random variable with distribution Binomial$\left(2n, p = \frac{1}{2}\right)$ and mean $\mathbb{E}(\boldsymbol{Q}_x) = x$, this proposal seeks to generate a symmetrical and unimodal distribution around a fixed point. Note that

$$q(y \mid x) = \mathbb{P}(\boldsymbol{Q}_x = y) = \mathbb{P}(\mathbb{B} = y - x + n) = \mathbb{P}(\boldsymbol{Q}_y = x) = q(x \mid y),$$

then the Metropolis algorithm can be executed over the distribution $\pi(x) \propto \exp(-\beta f(x))$ is taken with $\beta = \frac{1}{10}$. The results are shown in Figure 5.3.

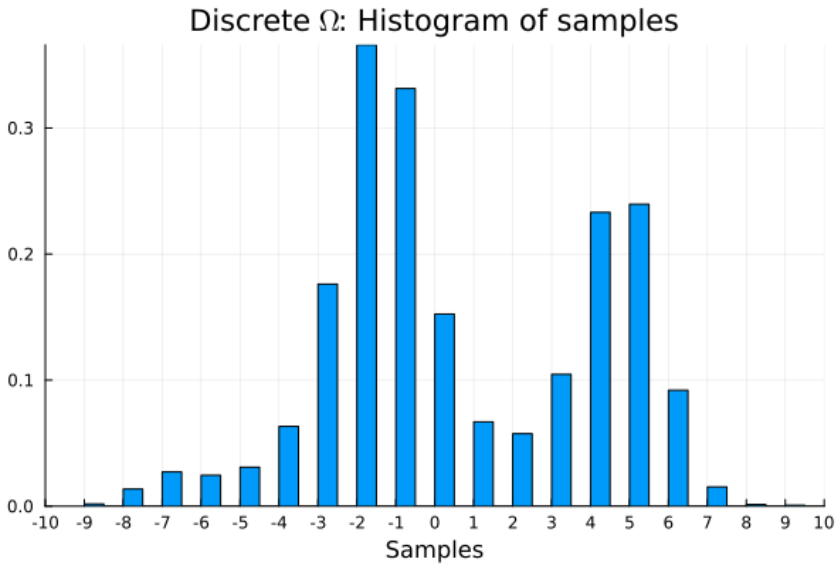


Figure 5.3: Result of the Metropolis algorithm in an optimization problem with restricted domain. Figure elaborated by the author using Julia (consult [21]).

Example 5.2. Consider the minimization problem

$$\min_{x,y \in \mathbb{R}} f(x, y)$$

with

$$f(x, y) = \sin(\pi x/3) + \cos(\pi x/3) + 0.025(x^2 + y^2 - xy) + 1.$$

Running the Metropolis algorithm with 30000 iterations and candidate-generating kernel a bivariate normal distrution with variance $\delta I = I$, it is obtained information about local minima of the function and it is also evident that a greater sampling is done on the global minimum, as shown in Figure 5.4.

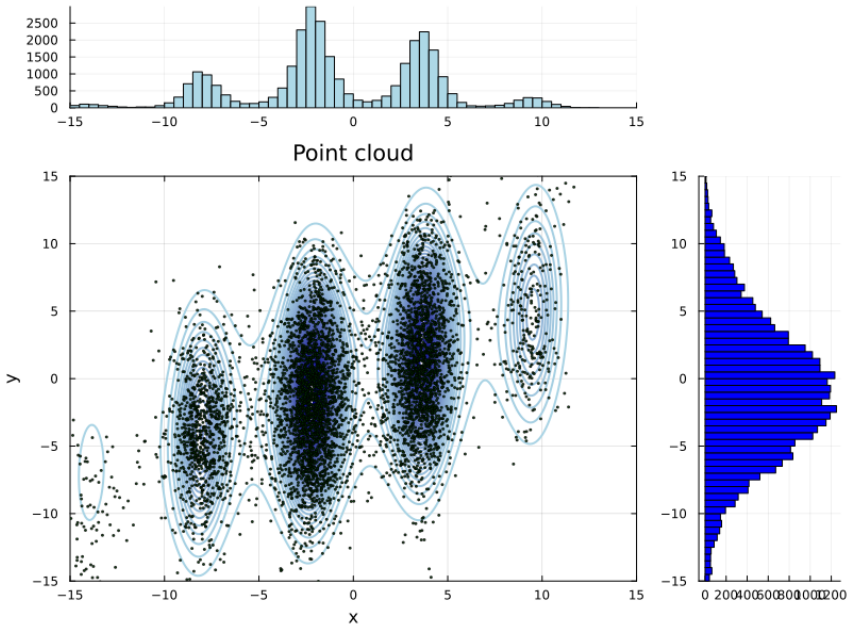


Figure 5.4: Result of the Metropolis algorithm in an optimization problem in two dimensions. Figure elaborated by the author using Julia (consult [21]).

5.2 Least squares

Least squares is a fundamental technique for fitting models to data, particularly in the context of linear models with additive error. This section explores the classical formulation of least squares and its connection to Bayesian inference. Specifically, it is shown that Bayesian estimation naturally leads to a form of regularization equivalent to Tikhonov regularization. Additionally, an example of linear regression is presented to illustrate these concepts and their practical implications (this application comes from [10]).

Formally, given a system represented as:

$$A\mathbf{x} \approx \mathbf{b},$$

where $A \in \mathbb{R}^{m \times n}$ is a known matrix, $\mathbf{b} \in \mathbb{R}^m$ is a known vector, and $\mathbf{x} \in \mathbb{R}^n$ is the unknown vector to be determined, the least squares solution is obtained by

minimizing the objective function:

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2.$$

Theorem 5.3. Suppose A has full column rank. The minimization problem

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2,$$

has a unique solution and is given by $\boldsymbol{x}^* = (A^\top A)^{-1} A^\top \boldsymbol{b}$.

Proof. The least squares problem aims to minimize the function

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2.$$

Expanding the norm, this is equivalent to minimizing

$$f(\mathbf{x}) = (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}).$$

Taking the gradient with respect to $\mathbf{x}$:

$$\nabla f(\mathbf{x}) = 2A^\top (A\mathbf{x} - \mathbf{b}).$$

Setting this equal to zero to find the critical point:

$$A^\top A\mathbf{x} = A^\top \mathbf{b}.$$

Since A has full column rank, the matrix $A^\top A$ is invertible, and solving for $\mathbf{x}$ gives the unique solution:

$$\mathbf{x}^* = (A^\top A)^{-1} A^\top \mathbf{b}.$$

Thus, the least squares problem has a unique solution as stated. □

In the context of Bayesian inference for the least squares problem (see [5,7]), the above minimization problem is considered as a problem of estimating the distribution of variable X given knowledge of a realization of variable Y and assuming an error E with some known distribution, all of these quantities related by the model

$$Y = AX + E,$$

where A is a known and fixed matrix (according to [10]).

As stated in Section 3.1.4.1, the likelihood function is defined as

$$\pi(x|y) = \pi_{\text{noise}}(y - Ax)$$

for this model, and the prior distribution is selected depending on the context of the problem, typically, this prior is either a normal distribution or a uniform distribution.

Example 5.4. Consider the matrix

$$A = \begin{pmatrix} 3.0 & -1.0 & -1.0 \\ -1.0 & 2.0 & 0.0 \\ -1.0 & 0.0 & 2.0 \end{pmatrix}$$

and consider the vector $y_e = y_0 + e$ where $y_e = (-7.0,\ 4.0,\ 2.0)^\top$ and e is a realization of the multivariate normal distribution Normal $\left(\mu = (0.0,\ 0.0,\ 0.0)^\top, \Sigma = 0.2^2\ I\right)$; the unique solution to the problem $Ax = y_0$ is the vector $x_0 = (-2.0,\ 1.0,\ 0.0)^\top$. When $y_e = (-7.025,\ 3.972,\ 1.755)^\top$ the solution to the linear system is $x_e = (-2.080,\ 0.945,\ -0.162)^\top$, this solution can also be obtained by the *maximum likelihood* method assuming normality.

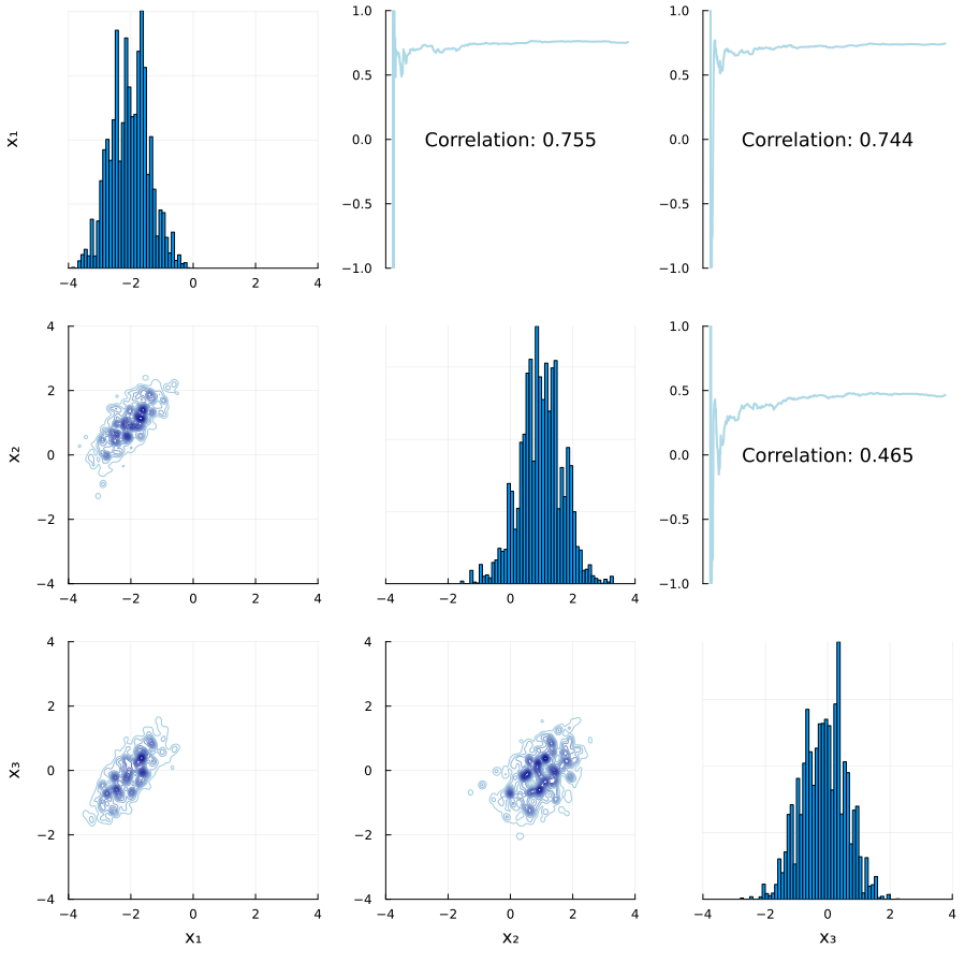


Figure 5.5: Result of the Metropolis algorithm in a linear least squares problem. Figure elaborated by the author using Julia (consult [21]).

The likelihood distribution will be taken as $y \mid x \sim \text{Normal}(\mu = Ax, \Sigma = I)$ and the a priori distribution is going to be $x \sim \text{Normal}(\mu = 0.0,\ 0.0,\ 0.0)^\top, \Sigma =$

$5^2 I$). Applying the Metropolis Hastings algorithm with 25000 iterations and candidate-generating kernel with variance equal to $\delta I = 5I$, the results shown in the figure below are obtained. As stated above in the Example 3.19, theoretically the a posteriori distribution is

$$x \mid y \sim \text{Normal}(\mu = A^\top A + \alpha I)^{-1} A^\top y, \Sigma = \frac{1}{\alpha} I - A^\top (AA^\top + \alpha I)^{-1}) A$$

where $\alpha = \frac{1}{5^2} = \frac{1}{25}$ is the ratio of the noise variance to the prior variance.

In the corner below the main diagonal of the previous graph the densities of the samples obtained for each pair of random variables are shown, in the diagonal the histograms are placed and in the corner above the diagonal the convergence of the sample correlation statistic is shown together with the value obtained when calculating it in all the samples.

The acceptance rate of the algorithm is 89.02 % and some statistics are:

$$\boldsymbol{x}_{MAP} = (-2.080,\ 0.945,\ -0.162)^\top \quad \boldsymbol{x}_{CM} = (-2.037,\ 0.976,\ -0.128)^\top.$$

5.3 Parameter Estimation for Dynamical Systems

This section presents a general overview of parameter estimation in dynamical systems using Bayesian inference combined with Markov Chain Monte Carlo (MCMC) methods. This approach provides a robust framework for model calibration, accounting for measurement errors and incorporating data variability. By considering these uncertainties, the estimated models can be effectively used for prediction while maintaining a realistic representation of the system dynamics (all concepts and methods presented here are based on [6,10]).

To illustrate these concepts, an example using recurrence relations will be presented. Additionally, several case studies involving ordinary differential equations will be explored, employing both simulated and real data.

The procedure for calibrating models given by dynamic systems using some MCMC method is shown in the diagram below (the process described here is inspired in [6]).

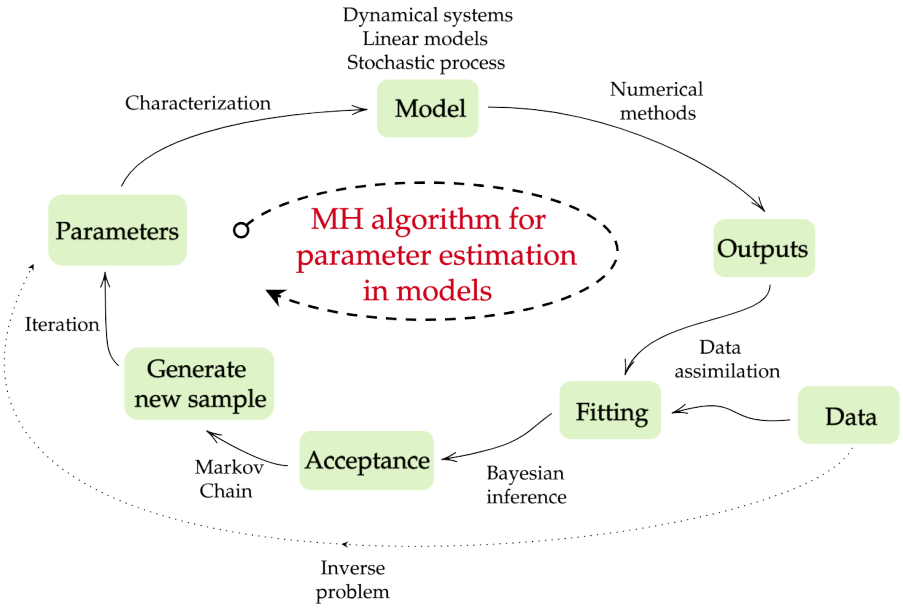


Figure 5.6: Use of the MH algorithm in the estimation of the parameters of a model. Diagram elaborated by the author.

The Figure 5.6 illustrates the implementation of the Metropolis-Hastings (MH) algorithm for parameter estimation in a model. The process begins with the identification of a set of parameters that define the model's behavior. These parameters play a crucial role in characterizing the system and ensuring that the model accurately represents the underlying phenomena. Once the parameters are established, numerical methods—chosen based on their efficiency and computational feasibility—are employed to generate the model outputs. Since many models rely on complex mathematical formulations, selecting appropriate numerical methods is essential to reduce computational costs while maintaining accuracy.

Following the computation of model outputs, a data assimilation process is performed to compare these outputs with observed data, allowing for the evaluation of the model's predictive capability. To refine the estimation process, Bayesian inference is applied to determine the acceptance of newly generated samples. This step provides the algorithm with probabilistic knowledge of the target distribution, ensuring that the generated samples follow the desired statistical properties. If a sample is accepted, it is incorporated into the iterative process; otherwise, a new sample is proposed, and the cycle continues. This iterative procedure is repeated until convergence is achieved, ensuring an optimal estimation of the model parameters and improving the model's overall reliability (consult [6,7,10]).

Example 5.5. (**Logistic growth**) This example is based on [6] and is constructed using simulated data. The logistic population growth model is considered over a time span of 60 years. A total of 60 data points are generated under the assumption that the true parameter values are $r_0 = 0.2$ and $K_0 = 0.75$. These data points are then perturbed by introducing multiplicative noise drawn from a Gamma distribution, specifically $\Gamma(\alpha = 100, \beta = 1/100)$. The objective is to estimate the parameters (r_0, K_0) using the Metropolis-Hastings algorithm, as previously discussed.

Given population growth data over time, it is assumed that the population follows a logistic growth model described by the equation:

$$\frac{dx(t)}{dt} = rx\left(1 - \frac{x}{1000K}\right), \quad x(0) = x_0 = 20$$

where r represents the population growth rate, K is the carrying capacity of the environment in which the population resides, and x_0 is the initial population size. The exact solution to this initial value problem is given by

$$x(t) = \frac{1000x_0K}{x_0 + (1000K - x_0)e^{-rt}}.$$

The inverse problem under consideration aims to estimate the values of r and K based on observations $\{z_\ell\}_{\ell=0}^k$ of the state variable $x(t, \theta)$ at time points $\{t_\ell\}_{\ell=0}^k$.

To apply the Metropolis Hastings Random Walk (MHRW) algorithm to this inverse problem, prior distributions are assigned to the parameters r and K. Specifically, each parameter follows a Gamma distribution, i.e., $r \sim \Gamma(a_r, b_r)$ and $K \sim \Gamma(a_K, b_K)$. Additionally, independence between both parameters is assumed, leading to the joint prior distribution

$$\Theta \sim \Gamma(a_r, b_r) \times \Gamma(a_K, b_K).$$

The Gamma distribution $\Gamma(a, b)$ has a probability density function given by

$$f(x; a, b) = \frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}, \quad x > 0.$$

The data points z_ℓ are observed at times t_ℓ for ℓ ranging from 1 to k. Since these observations represent population counts, it is assumed that

$$z_i \sim \text{Poisson}(x(t_i, \theta)),$$

where $\theta = (r, K)$. Defining $z = (z_1, \ldots, z_k)$, and assuming that the observations are independent and follow a Poisson distribution, the likelihood function $\pi_{Z|\Theta}(z|\theta)$ is

given by the product of Poisson density functions

$$\pi(z|\theta) = \prod_{i=1}^{k} f(z_i; x(t_i, \theta)) = \prod_{i=1}^{k} \frac{e^{-x(t_i,\theta)}(x(t_i,\theta))^{z_i}}{z_i!}.$$

By Bayes' theorem, the posterior distribution is given by

$$\begin{aligned}\pi(\theta|z) &\propto \pi_{Z|\Theta}(z|\theta) \times \pi_{\Theta}(\theta) \\ &= \prod_{i=1}^{k} \frac{e^{-x(t_i,\theta)}(x(t_i,\theta))^{z_i}}{z_i!} \times \Gamma(a_r, b_r)(r) \times \Gamma(a_K, b_K)(K).\end{aligned}$$

The Metropolis Hastings Random Walk algorithm is applied starting from a random point in the set, with a total of 10000 samples and a Gaussian candidate generating kernel with variance equal to $\sigma^2 = 0.1$; as a result, the two figures below are shown.

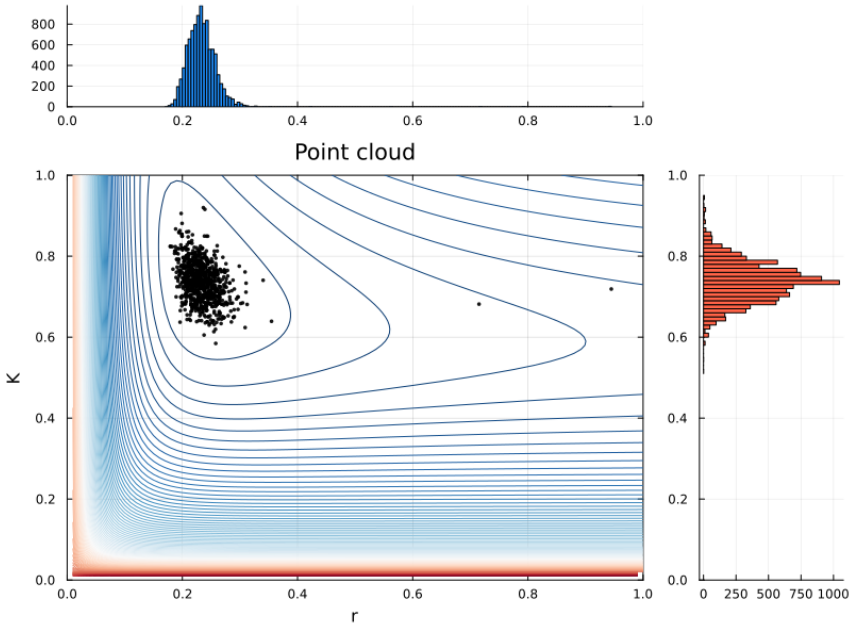


Figure 5.7: Point cloud plot of applying the MH algorithm in parameter estimation of the logistic model. Figure elaborated by the author using Julia (see [21]).

Figure 5.7 presents the scatter plot obtained from the algorithm, with histograms of the estimated parameters r and K displayed along the margins. It is evident that, under this execution of the algorithm and the assumptions made in the Bayesian inference process, the parameter K exhibits greater variability compared to r. Additionally, it is important to note that both distributions are unimodal and

closely centered around the true parameter values, $(r_0, K_0) = (0.2, 0.75)$. One way to verify this proximity to the actual values is by observing that

$$\theta_{MAP} = (0.2283, 0.7318) \text{ and } \theta_{CM} = (0.2325, 0.7372).$$

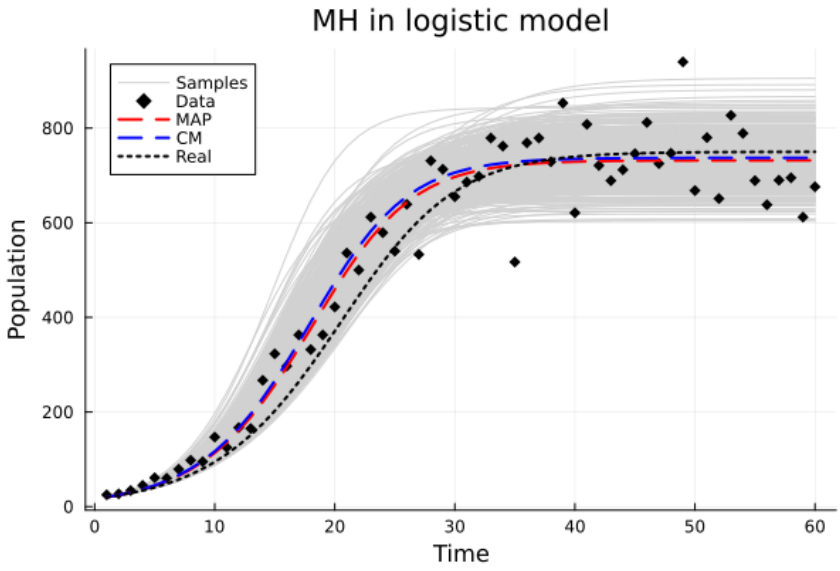


Figure 5.8: Result of applying the MH algorithm in parameter estimation of the logistic model. Figure elaborated by the author using Julia (see [21]).

Figure 5.8 illustrates the results obtained from the execution of the Metropolis-Hastings Random Walk (MHRW) method. In this plot, the black diamonds represent the initial data points. The black dashed curve corresponds to the model with the true parameter values. The blue and red curves depict the outcomes of the point estimates of the conditional mean and the maximum a posteriori, respectively. Additionally, the gray curves represent the solutions of the ordinary differential equation for the parameter samples generated by the algorithm. This method facilitates the assessment of the model's error and variability relative to the data.

Example 5.6. (**Lokta-Volterra model** [14]) Consider the model discussed in Examples 2.15 and 2.18. Similarly to the previous example, 21 data points are simulated over a period of 20 years and modified with multiplicative Gamma noise. The MHRW algorithm is executed under the assumption that all parameters have a uniform prior distribution over the interval $(0, 1)$. The likelihood function follows a normal distribution with variance $\sigma^2 = 0.1$. Additionally, 10,000 iterations are

performed using a Gaussian candidate generating kernel with a variance of 0.05. The results are shown in the Figures 5.9 and 5.10.

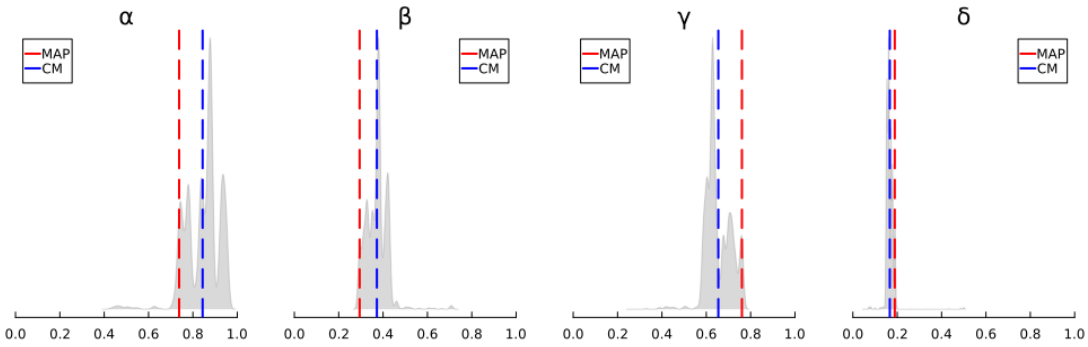


Figure 5.9: Histograms of applying the MH algorithm in parameter estimation of the predator-prey model. Figure elaborated by the author using Julia (see [21]).

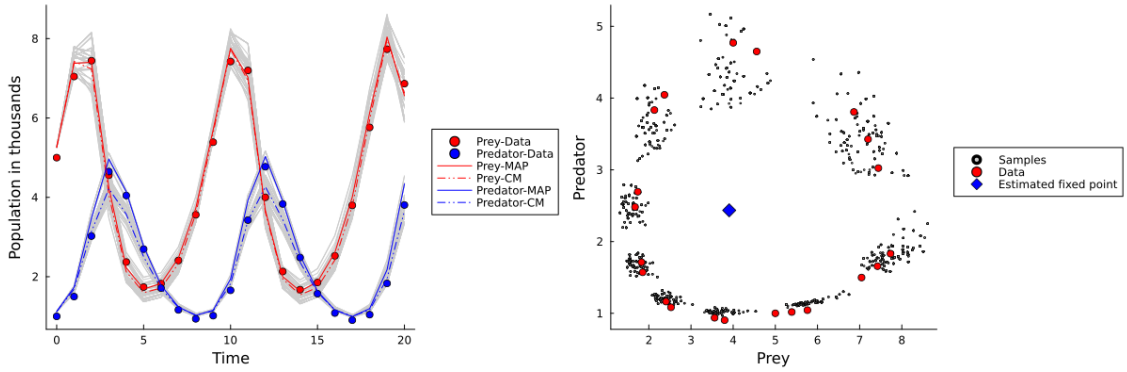


Figure 5.10: Results of applying the MH algorithm in parameter estimation of the predator-prey model. Figure elaborated by the author using Julia (see [21]).

Considering that the parameters used to generate the data were

$$[\alpha, \beta, \gamma, \delta] = [0.7, 0.3, 0.8, 0.2],$$

the obtained point estimates are:

$$[\alpha, \beta, \gamma, \delta]_{MAP} = [0.6420, 0.2657, 0.8616, 0.2079],$$

$$[\alpha, \beta, \gamma, \delta]_{CM} = [0.6625, 0.2840, 0.8354, 0.2060].$$

Additionally, an acceptance rate of 38.03% was obtained.

Example 5.7. (**SIR model**) This example is inspired by [1] and extracts data from that article and the methods from [6]. The spread of the Spanish flu in San

Francisco from 1918 to 1919 is modeled using the following system of differential equations:

$$\begin{aligned}\dot{S} &= \mu N - \beta \frac{I}{N} S - \xi S - \mu S,\\ \dot{I} &= \beta \frac{I}{N} S + \xi S - \gamma I - \mu I,\\ \dot{R} &= \gamma I - \mu R,\end{aligned}$$

where $N = S + I + R$ and $\dot{N} = 0$. The parameters μ, ξ, N, and γ are assumed to be known. The objective of this model is to estimate the basic reproduction number $\mathcal{R}$ as well as the parameters β, $I(0)$, and $R(0)$. The presented model includes both known and unknown parameters, each with a specific epidemiological interpretation. The known parameters are: μ, representing the birth and natural mortality rate, which ensures a constant total population ($\dot{N} = 0$); ξ, denoting the incidence rate of external infections due to contacts with external populations; N, the total population size, assumed to remain constant over time; and γ, the recovery rate, which defines the rate at which infected individuals transition to the recovered category. The unknown parameters to be estimated are: β, the transmission rate, which quantifies the probability of disease spread from an infected individual to a susceptible one; $I(0)$, the initial number of infected individuals at the beginning of the analyzed period, essential for model calibration; and $R(0)$, the initial number of recovered individuals, which represents those who have already gained immunity before the study period. Additionally, the basic reproduction number $\mathcal{R}$ is a key epidemiological metric, defined as $\mathcal{R} = \frac{\beta}{\gamma}$, indicating the average number of secondary infections generated by a single infected individual in a fully susceptible population. If $\mathcal{R} > 1$, the infection tends to spread within the population, whereas if $\mathcal{R} < 1$, the outbreak naturally declines. These unknown parameters must be inferred from observational data using statistical or Bayesian inference methods, such as the Metropolis-Hastings algorithm. The data are shown in Figure 5.11.

The explicit Euler method is used as the numerical approach to solve the SIR model. In [1], a point estimate of the data was obtained using a squared error minimization algorithm, where each estimated value was considered as the mean of a Gamma distribution to define the prior distribution, assuming independence among the unknown parameters. Since the cases represent non-negative integer values resulting from an infection response, the Poisson distribution was selected as the likelihood function for the data given the parameters. The MHRW algorithm was executed with 15000 iterations using a Gaussian kernel with variance $\sigma^2 = 1$.

The point estimates obtained were $\mathcal{R}_{MAP} = 0.0446$ and $\mathcal{R}_{CM} = 0.7848$, with an acceptance rate of 35.29%. The results are presented in Figure 5.12.

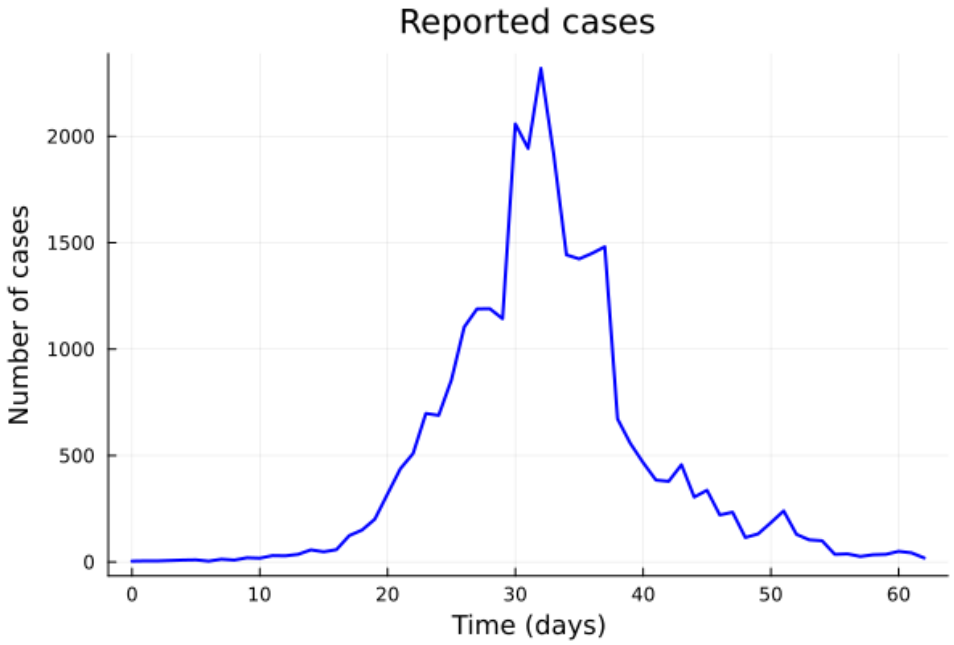


Figure 5.11: Data of reported cases extracted from [1]. Figure elaborated by the author using Julia (see [21]).

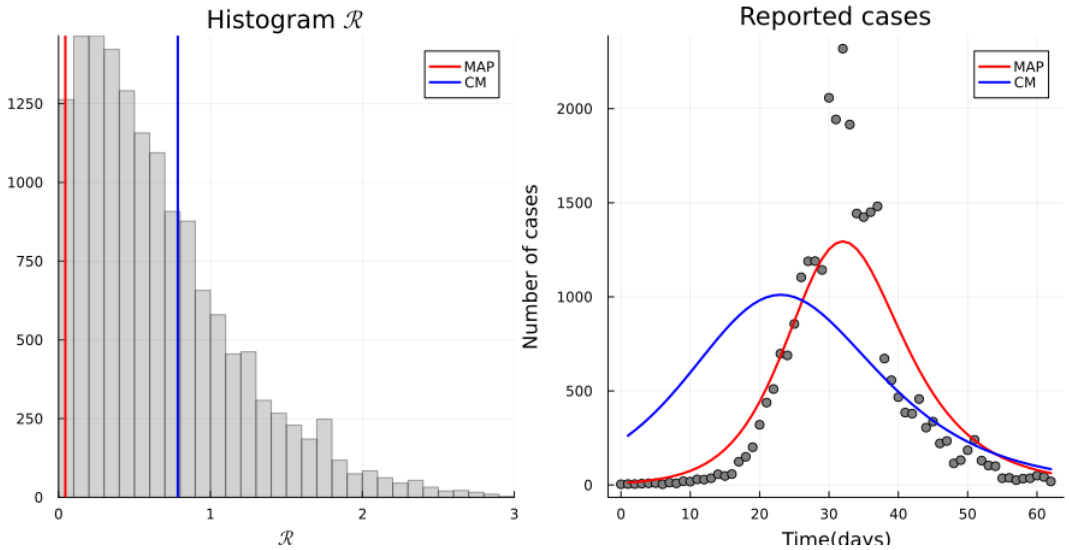


Figure 5.12: Result of applying the MHRW algorithm for estimating the reproduction rate $\mathcal{R}$ in the SIR model. Figure created by the author using Julia (see [21]).

5.4 TOMGRO Model

The TOMGRO model is a mathematical framework designed to simulate the growth dynamics of tomato crops under varying environmental conditions based on a model of ordinary differential equations and recurrence relations. This section provides an overview of its structure, the data used for calibration, and the results obtained from its application. This application is based on [4] and [9].

5.4.1 Model Description

This subsection outlines the fundamental equations governing the TOMGRO model, including the biological and environmental factors that influence plant growth. The key parameters and assumptions underlying the model formulation are also discussed.

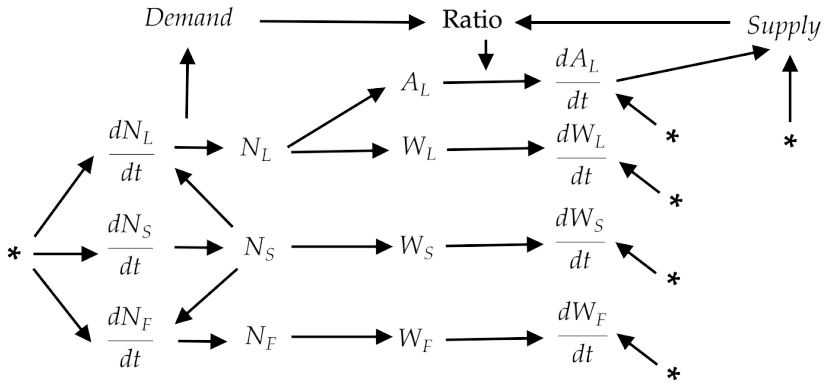


Figure 5.13: Schematic of the overall development and the actual growth of tomato plant state variables in TOMGRO. The direction of the arrows indicates causal effects, and the $*$ indicates an effect of one or more environmental variables on the process rates. Figure elaborated by the author and inspired in [9].

The TOMGRO model is a dynamic simulation model designed to represent the growth and development of tomato plants under greenhouse conditions. It is based on source-sink relationships and incorporates an age-structured representation of state variables. The primary objective of TOMGRO is to predict the growth and yield of tomato plants in response to varying environmental conditions, including temperature, solar radiation, and carbon dioxide concentration.

The model characterizes the tomato plant through seven state variable vectors that include physiological age classes of plant components: number of leaves, number of main stem segments, number of fruits, dry weights of leaves plus petioles, dry weights of main stem segments, dry weights of fruits, and leaf area. Plant growth occurs through changes in these numbers, weights, and areas across different age classes of each component. The model assumes that only the main vegetative stem is considered, with all vegetative branches being removed as soon as they are formed, which is a common practice in greenhouse tomato production systems (see [4,9]). Additionally, the plants are assumed to be well-watered and adequately fertilized to focus solely on the influence of environmental conditions on growth dynamics. The growth of the tomato plant in TOMGRO is determined by the interaction between carbon supply and demand. Carbon supply is derived from photosynthesis, while carbon demand is dictated by the sink strength of different plant organs. The partitioning of assimilates follows a source-sink approach, where different plant organs compete for available resources based on their physiological needs. The demand for carbon is calculated based on the developmental stage of the plant and the relative sink strength of each organ, ensuring a realistic representation of dry matter allocation throughout the plant's life cycle.

The model operates on two time loops. The main loop advances time daily, while a fast loop increments time on an hourly basis or even more frequently if required. Within the fast loop, instantaneous environmental conditions such as temperature, CO_2 concentration, and photosynthetic photon flux density (PPFD) are computed. These environmental variables influence key physiological processes, including development rates, photosynthesis, and respiration. At the end of each day, the accumulated changes in state variables are updated, ensuring that the growth dynamics reflect the cumulative impact of environmental conditions over time (consult [4]).

TOMGRO's structure allows it to respond dynamically to changing environmental conditions, making it a valuable tool for optimizing greenhouse management strategies. By incorporating real-time environmental data, the model can simulate plant responses with high temporal resolution, which is crucial for decision-making in controlled agricultural environments. The explicit representation of plant components in age classes enables accurate predictions of growth stages and biomass accumulation. Moreover, the integration of physiological mechanisms governing carbon assimilation, respiration, and biomass allocation provides a

comprehensive framework for understanding the complex interactions that drive tomato plant growth (consult [4,9]).

The TOMGRO model represents a significant advancement in greenhouse crop modeling by offering a detailed, dynamic, and mechanistic approach to simulating tomato growth. Its ability to incorporate environmental variability and its structured representation of plant components make it a robust tool for researchers and agronomists seeking to improve productivity and resource use efficiency in controlled environments. The model's application extends beyond academic research, as it can be utilized for practical greenhouse management, optimizing environmental control strategies, and enhancing tomato yield predictions. By leveraging TOMGRO, growers can make informed decisions regarding temperature regulation, CO_2 enrichment, and light management to maximize plant performance and overall production efficiency (according to [4]).

The following is a description of the model at the mathematical level. Let m be the number of age classes in the model (typically $m = 20$ according to [4] and [9]). Consider the seven state variables that describe the model

- Number of stems: $N_S = (N_{S,1}, \ldots, N_{S,m})$.
- Number of leaves: $N_L = (N_{L,1}, \ldots, N_{L,m})$.
- Number of fruits: $N_F = (N_{F,1}, \ldots, N_{F,m})$.
- Weight of stems: $W_S = (W_{S,1}, \ldots, W_{S,m})$.
- Weight of leaves: $W_L = (W_{L,1}, \ldots, W_{L,m})$.
- Leaf area: $A_L = (A_{L,1}, \ldots, A_{L,m})$.
- Weight of fruits: $W_F = (W_{F,1}, \ldots, W_{F,m})$.

For the organs stems, leaves, and fruits, the following dynamics are considered:

$$(S)\begin{cases} \frac{dN_{S,1}}{dt} = s_{1,1}N_{S,1} - s_{1,2}N_{S,1}^2 & \\ \frac{dN_{S,i}}{dt} = s_{i-1,2}N_{S,i-1}^2 + s_{i,1}N_{S,i} - s_{i,2}N_{S,i}^2 & \text{for } i = 2, \ldots, m, \\ \frac{dW_{S,1}}{dy} = w_{s,1,1}N_{S,1}W_{S,1} - w_{s,1,2}N_{S,1}W_{S,1}^2 & \\ \frac{dW_{S,i}}{dy} = w_{s,i-1,2}N_{S,i-1}W_{S,i-1}^2 + w_{s,i,1}N_{S,i}W_{S,i} - w_{s,i,2}N_{S,i}W_{S,i}^2 & \text{for } i = 2, \ldots, m, \end{cases}$$

$$(L)\begin{cases} \frac{dN_{L,1}}{dt} = l_{1,1}N_{L,1} - l_{1,2}N_{L,1}^2 & \\ \frac{dN_{L,i}}{dt} = l_{i-1,2}N_{L,i-1}^2 + l_{i,1}N_{L,i} - l_{i,2}N_{L,i}^2 & \text{for } i = 2,\dots,m, \\ \frac{dW_{L,1}}{dy} = w_{l,1,1}N_{L,1}W_{L,1} - w_{l,1,2}N_{L,1}W_{L,1}^2 & \\ \frac{dW_{L,i}}{dy} = w_{l,i-1,2}N_{L,i-1}W_{L,i-1}^2 + w_{l,i,1}N_{L,i}W_{L,i} - w_{l,i,2}N_{L,i}W_{L,i}^2 & \text{for } i = 2,\dots,m, \\ \frac{dA_{L,1}}{dy} = a_{l,1,1}N_{L,1}A_{L,1} - a_{l,1,2}N_{L,1}A_{L,1}^2 & \\ \frac{dA_{L,i}}{dy} = a_{l,i-1,2}N_{L,i-1}A_{L,i-1} + a_{l,i,1}N_{L,i}A_{L,i} - a_{l,i,2}N_{L,i}A_{L,i}^2 & \text{for } i = 2,\dots,,m \end{cases}$$

$$(F)\begin{cases} \frac{dN_{F,1}}{dt} = f_{1,1}N_{F,1} - f_{1,2}N_{F,1}^2 & \\ \frac{dN_{F,i}}{dt} = f_{i-1,2}N_{F,i-1} + f_{i,1}N_{F,i} - f_{i,2}N_{F,i}^2 & \text{for } i = 2,\dots,m, \\ \frac{dW_{F,1}}{dy} = w_{f,1,1}N_{F,1}W_{F,1} - w_{f,1,2}N_{F,1}W_{F,1}^2 & \\ \frac{dW_{F,i}}{dy} = w_{f,i-1,2}N_{F,i-1}W_{F,i-1} + w_{f,i,1}N_{F,i}W_{F,i} - w_{f,i,2}N_{F,i}W_{F,i}^2 & \text{for } i = 2,\dots,m, \end{cases}$$

The expressions $o_{i,k}$, $w_{o,i,k}$, and $a_{o,i,k}$ for $o \in \{s, l, f\}$ are functions that depend on time and some interactions between the state variables; this work will not emphasize these functions since they depend on the model and the interpretation of the phenomenon, including external environmental conditions and the relationship between the supply and demand for resources of the environment and the plant, respectively. These can be found in [4] and [9]. A simplification of this model is to assume all the aforementioned expressions as constants.

To summarize the previous equations, given an organ O of the plant under consideration, let N_O represent the number of occurrences of the organ in the plant, and some biological quantity of interest C_O such as weight, length, volume, or area. The model can thus be described as follows: Given an organ O and a number of age classes m, the dynamics of this organ are governed by:

$$(O)\begin{cases} \frac{dN_{O,1}}{dt} = o_{1,1}N_{O,1} - o_{1,2}N_{O,1}^2 & \\ \frac{dN_{O,i}}{dt} = o_{i-1,2}N_{O,i-1}^2 + o_{i,1}N_{O,i} - o_{i,2}N_{O,i}^2 & \text{for } i = 2,\dots,m, \\ \frac{dC_{O,1}}{dt} = c_{o,1,1}N_{O,1}C_{O,1} - c_{o,1,2}N_{O,1}C_{O,1}^2 & \\ \frac{dC_{O,i}}{dt} = c_{o,i,0}N_{O,i-1}C_{O,i-1}^2 + c_{o,i,1}N_{O,i}C_{O,i} - c_{o,i,2}N_{O,i}C_{O,i}^2 & \text{for } i = 2,\dots,m, \end{cases} \tag{5.1}$$

where the coefficients accompanying the state variables in all the differential equations are functions that depend on time (and potentially other state variables associated with other organs).

5.4.2 Data

The dataset employed for model calibration and validation is described, detailing the sources, measurement techniques, and preprocessing steps. Special attention is given to the treatment of uncertainties and potential sources of error in the recorded data.

In this thesis, two types of data are used, as described below:

Simulated data. Following the dynamics described by Equation 5.1, data for a single organ with 2 age classes is generated using a numerical method that solves the coupled ordinary differential equations and the recurrence relation between both age classes. The model can be described by the following equations:

$$(P)\begin{cases} \frac{dN_{P,1}}{dt} = p_{1,1}N_{P,1} - p_{1,2}N_{P,1}^2 \\ \frac{dN_{P,2}}{dt} = p_{1,2}N_{P,1}^2 + p_{2,1}N_{P,2} - p_{2,2}N_{P,2}^2 \\ \frac{dC_{P,1}}{dt} = c_{p,1,1}N_{P,1}C_{P,1} - c_{p,1,2}N_{P,1}C_{P,1}^2 \\ \frac{dC_{P,2}}{dt} = c_{p,1,2}N_{P,1}C_{P,1}^2 + c_{p,2,1}N_{P,2}C_{P,2} - c_{p,2,2}N_{P,2}C_{P,2}^2 \end{cases}$$

The model can be represented by the Figure 5.14.

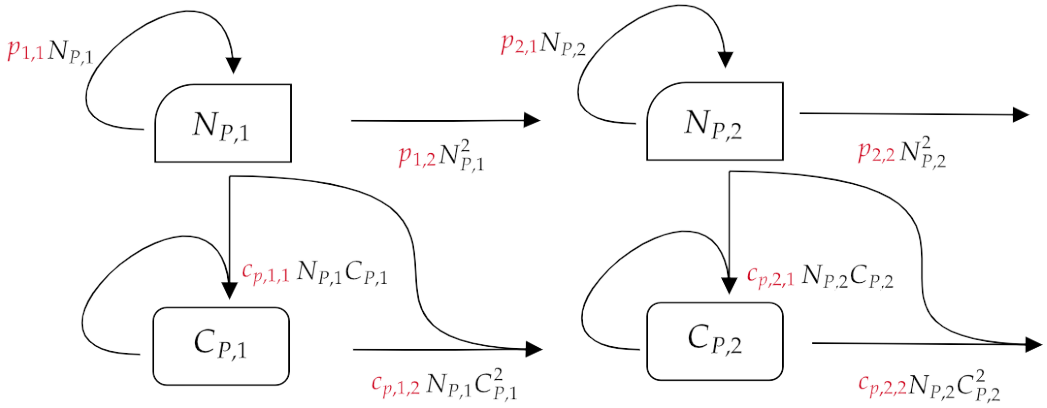


Figure 5.14: Representation in compartments of the TOMGRO model with one organ and two age classes. All parameters are represented in color red. Figure created by the author.

The data are modeled using the following parameter values:

- Time in the interval $t \in [0.0, 3.0]$ with a step size of $dt = 0.02$ for the Euler method.

- Initial values of the state variables:

Variable	Initial Value	Interpretation
$N_{P,1}$	1.15	Initial number of early-stage fruits
$N_{P,2}$	0.55	Initial number of late-stage fruits
$C_{P,1}$	0.15	Initial mass of early-stage fruits (M)
$C_{P,2}$	0.25	Initial mass of late-stage fruits (M)

- Constant parameters in the ODE system:

Parameter	Value	Interpretation	Units (months, M)
$p_{1,1}$	2.0	Rate of early-stage fruit appearance	Fruits / month
$p_{1,2}$	0.5	Transition rate from early to late-stage fruits	1 / month
$p_{2,1}$	0.9	Retention rate of late-stage fruits	Fruits / month
$p_{2,2}$	0.2	Loss rate of late-stage fruits	1 / month
$c_{p,1,1}$	1.2	Biomass accumulation rate in early-stage fruits	M / (fruit $\cdot$ month)
$c_{p,1,2}$	0.8	Biomass transfer rate from early to late-stage fruits	M / (fruit $\cdot$ month)
$c_{p,2,1}$	0.3	Biomass accumulation rate in late-stage fruits	M / (fruit $\cdot$ month)
$c_{p,2,2}$	0.4	Biomass loss rate in late-stage fruits	M / (fruit $\cdot$ month)

By contaminating the Euler method solution of the differential equation system with additive noise from a Gamma distribution $\Gamma(\alpha = 50, \beta = 1/50)$, the data presented in table 5.1 and figure 5.15 are obtained.

t	$N_{P,1}$	$N_{P,2}$	$C_{P,1}$	$C_{P,2}$
0.0	1.15	0.55	0.15	0.25
0.3	1.60842	0.965384	0.202569	0.267582
0.6	3.25006	1.61237	0.413811	0.326375
0.9	2.40355	2.81359	0.537287	0.566242
1.2	2.85006	4.04979	1.00676	1.01483
1.5	4.08752	5.45348	1.28428	1.72505
1.8	2.94283	6.73463	1.30162	1.90091
2.1	4.11737	7.66144	1.26689	1.68079
2.4	3.4853	11.818	1.24081	1.99446
2.7	4.66976	10.2074	1.54236	2.60436
3.0	3.42531	7.73761	1.4128	1.99732

Table 5.1: Simulated values of $N_{P,1}$, $N_{P,2}$, $C_{P,1}$ and $C_{P,2}$ depending on time t. Table created by the author.

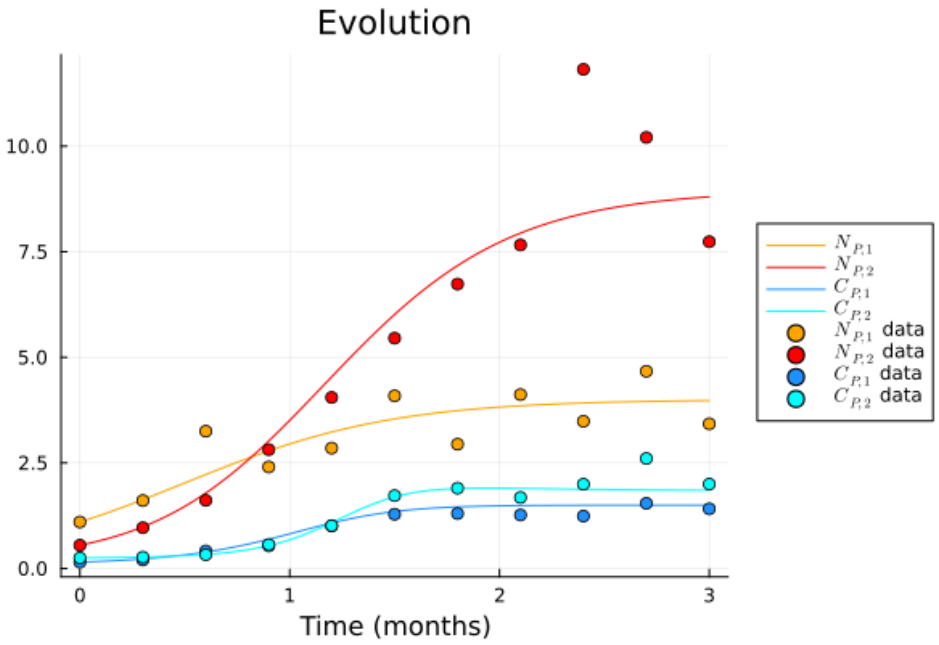


Figure 5.15: Plot of the solution curves of the differential equation system and the simulated data based on these curves. Figure created by the author.

Real data. Real data from the plant's phenology (i.e., some plant measurements) are used, including plant height in centimeters, the number of leaves, the total leaf area of the plant, and the number of clusters on the plant. All of this is measured over a period of 147 days, starting on March 8, 2023, and ending on August 2, 2023, with a total of 88 measurements taken every 22 days. Some of the data are presented in Table 5.2.

Index	Date	Plant height (cm)	Number of leaves	Total leaf area (m^2)	Number of clusters
1	2023-03-08	66.0	15	2.268	0
2	2023-03-08	62.0	15	1.566	0
3	2023-03-08	65.0	17	2.5432	0
4	2023-03-08	68.0	16	1.6128	0
5	2023-03-15	86.0	16	3.168	0
6	2023-03-15	77.0	16	2.9184	0
7	2023-03-15	83.0	19	3.1008	0
8	2023-03-15	88.0	18	3.0528	0
9	2023-03-22	117.0	18	1.548	0
10	2023-03-22	101.0	17	4.6665	0
11	2023-03-22	111.0	20	2.496	0
12	2023-03-22	115.0	20	2.94	0
13	2023-03-29	147.0	21	3.2571	1
⋮	⋮	⋮	⋮	⋮	⋮
77	2023-07-19	435.0	22	1.1594	14
78	2023-07-19	342.0	9	0.3276	9
79	2023-07-19	415.0	34	1.972	15
80	2023-07-19	358.0	21	1.3167	13
81	2023-07-25	449.0	22	2.2572	15
82	2023-07-25	352.0	10	0.54	10
83	2023-07-25	427.0	18	0.6048	15
84	2023-07-25	371.0	21	0.84	13
85	2023-08-02	473.0	23	2.0424	16
86	2023-08-02	361.0	9	0.36	11
87	2023-08-02	445.0	15	0.945	16
88	2023-08-02	389.0	21	1.6632	14

Table 5.2: Plant measurement data obtained from [21,23]. Table created by the author.

As can be seen, several measurements were taken on the same date from different tomato plants of the same "Clanio" family. To account for this information, an average is calculated for the daily measurements of each recorded variable, resulting in the data shown in Figure 5.2.

Given that there is insufficient data to calibrate the TOMGRO model as described in the previous section, the following model is proposed for the four state variables, following the dynamics of Equation 5.1, considering the following state variables:

- Plant height H.
- Number of leaves on the plant N_L.
- Total leaf area A_L.
- Number of clusters N_C.

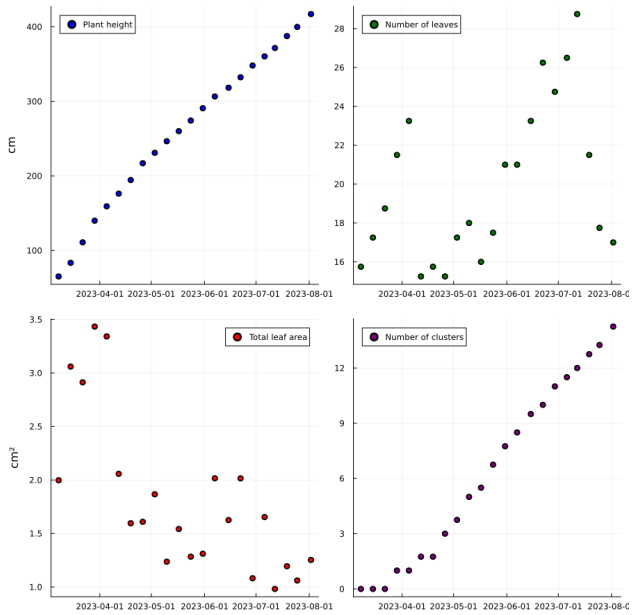


Figure 5.16: Graphs of variables measured on tomato plants over a harvest period. Figure prepared by the author using Julia (consult [21]).

The dynamics governing these variables are as follows

$$\frac{dH}{dt} = h, \quad (L) \begin{cases} \frac{dN_L}{dt} = l \\ \frac{dA_L}{dy} = a \end{cases}, \quad \frac{dN_C}{dt} = c.$$

The parameters in the system of equations represent constants that govern the growth dynamics of different state variables in the model. h is the rate of change of the plant height H over time, capturing how fast the plant grows in height. This parameter reflects the intrinsic growth potential of the plant. l represents the rate at which the number of leaves N_L changes, which is an important factor in determining the leaf production rate. This parameter influences the overall leaf structure of the plant, which is critical for photosynthesis and overall plant development. a is the rate of change of the total leaf area A_L, which quantifies the expansion of the plant's leaf surface area over time. This is a key factor in understanding the plant's capacity to capture sunlight and nutrients. Finally, c indicates the rate of change of the number of clusters N_C, which represents the formation of reproductive structures or clusters within the plant. This parameter helps describe the plant's ability to produce new clusters over time. Together, these parameters define the growth dynamics of the plant in terms of height, number of leaves, leaf area, and reproductive structures,

providing a comprehensive view of the plant's development process (consult the references [4,9]).

5.4.3 Results

The outcomes of the model calibration and its predictive capabilities are presented. The results are analyzed in the context of their agreement with observed data, highlighting the strengths and limitations of the approach.

5.4.3.1 With simulated data

For the likelihood function, a multivariate normal distribution is assumed (which implies additive noise in the model), with mean given by the solution of the system of ODEs for certain parameter values and variance $\Sigma = \gamma^2 I$, where $\gamma = 1/100$. For the prior distributions, uniformity is assumed over the interval $(-2.5, 2.5)$ for all parameters, which in turn implies independence. The MHRW algorithm is executed for these data with a total of 50,000 iterations using a Gaussian candidate-generating kernel with variance $\sigma^2 = 0.01$, taking the origin as the initial point.

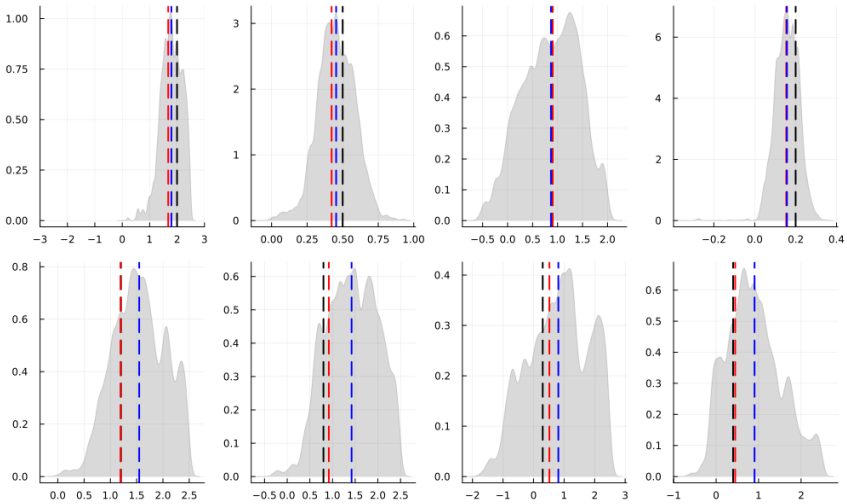


Figure 5.17: Histograms obtained by executing the MHRW algorithm in the model calibration with simulated data. In the histograms, the black line represents the parameter value used to construct the simulation data, the blue line shows the conditional mean value, and the red line indicates the maximum a posteriori value. Figure created by the author using Julia (see [21]).

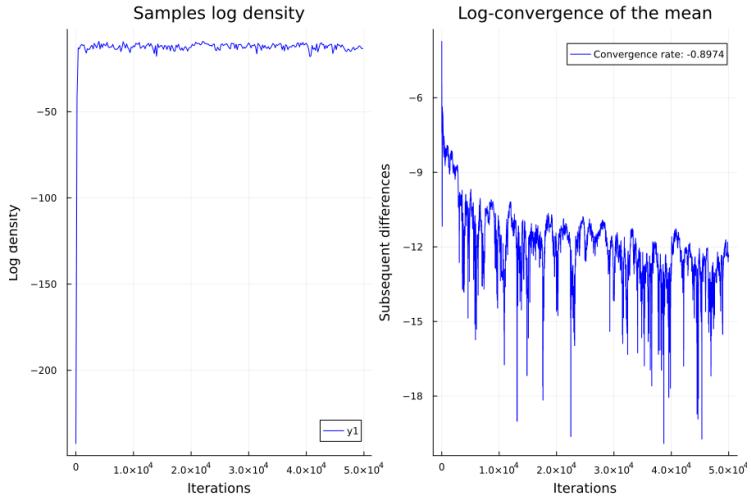


Figure 5.18: On the left, the log-density of the samples is shown, while on the right, the plot of the successive differences of the mean of the estimated parameters in logarithmic scale is displayed. Both graphs depend on the number of iterations. Figure created by the author using Julia (see [21]).

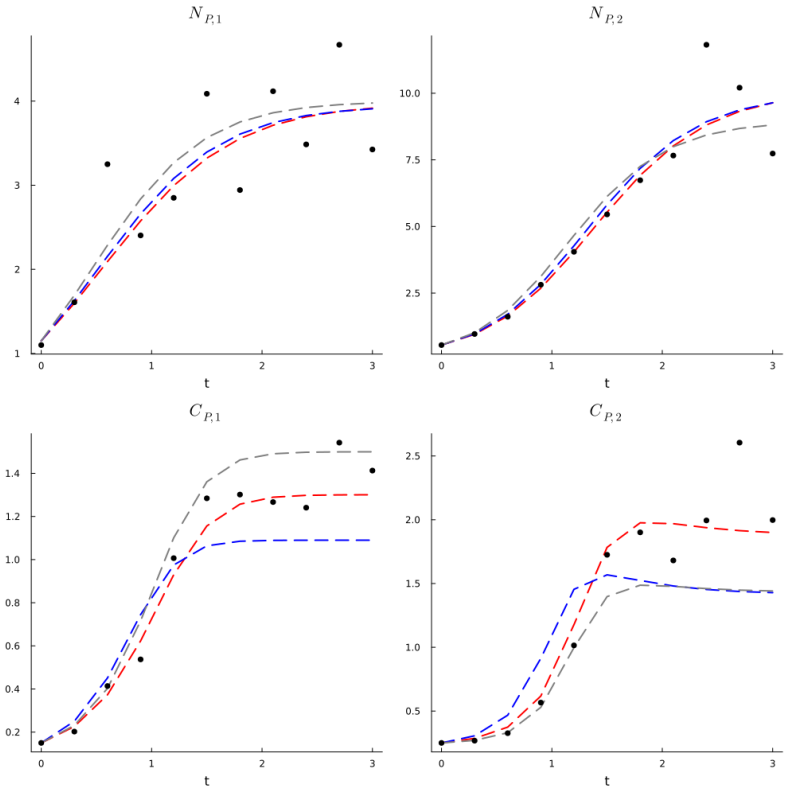


Figure 5.19: This figure presents the dynamics of the point statistics for the data obtained from the model's state variables. The black points represent the data, the gray line corresponds to the solution given by the initial parameters, the red line represents the maximum a posteriori estimator, and the blue line corresponds to the conditional mean estimator. Figure created by the author using Julia (see [21]).

Figures 5.17, 5.18, and 5.19 illustrate the results of applying the MCMC method to model calibration. The histograms reveal that the point estimates align well with the initial parameter values, indicating a proper execution of the algorithm. Additionally, the unimodal shape of the histograms suggests that the method stabilizes and converges to a specific region in the parameter space.

Moreover, the convergence plots show that the probability of the samples stabilizes as the number of iterations increases. Likewise, the mean of the parameters exhibits convergence, as the successive differences progressively decrease with a convergence rate of -0.8974. This behavior confirms the convergence of the MHRW algorithm. The sample acceptance rate is 70.29%.

Finally, the dynamics of the TOMGRO model are plotted against the point estimates obtained from the sampling process. Overall, the model accurately fits the data.

The prior-weighted likelihood $\pi(\text{Data}|\text{Parameters})\pi(\text{Parameters})$ results in the following values:

$$\frac{\pi(\text{Data}|\text{MAP parameters})\pi(\text{MAP parameters})}{\pi(\text{Data}|\text{Real parameters})\pi(\text{Real parameters})} \approx 8.1575,$$

and,

$$\frac{\pi(\text{Data}|\text{CM parameters})\pi(\text{CM parameters})}{\pi(\text{Data}|\text{Real parameters})\pi(\text{Real parameters})} \approx 3.1915,$$

indicating that both point estimators exhibit higher probability in the posterior distribution of the parameters compared to the true parameter values.

5.4.3.2 With real data

As in the previous case, for the likelihood function, a multivariate normal distribution is assumed (which implies additive noise in the model), with mean given by the solution of the system of ODEs for certain parameter values and variance $\Sigma = \gamma^2 I$, where $\gamma = 1/10$. For the prior distributions, uniformity is assumed over the interval $(-2.5, 2.5)$ for all parameters, which in turn implies independence.

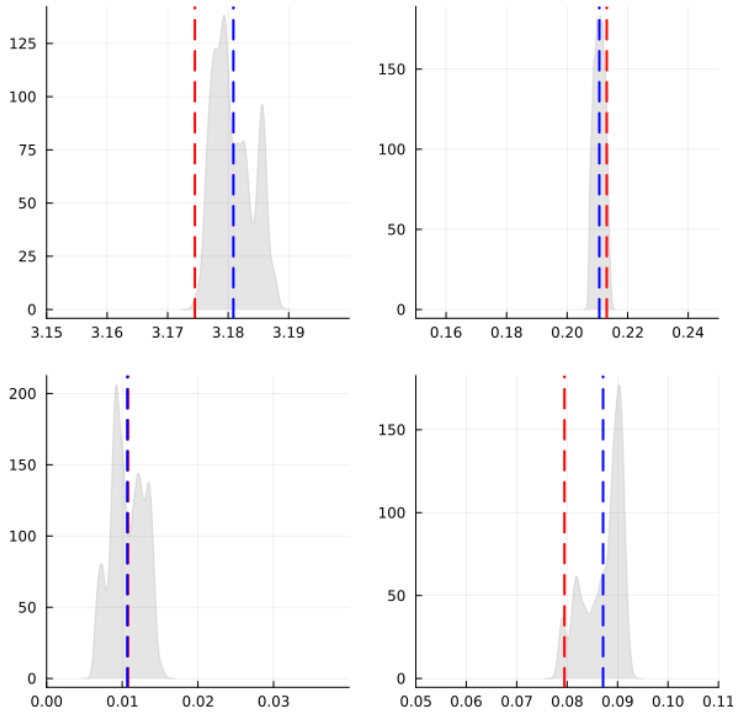


Figure 5.20: Histograms obtained by executing the MHRW algorithm in the model calibration with real data. In the histograms, the black line represents the parameter value used to construct the simulation data, the blue line shows the conditional mean value, and the red line indicates the maximum a posteriori value. Figure created by the author using Julia (see [21]).

The MHRW algorithm is executed for these data with a total of 5000 iterations using a Gaussian candidate-generating kernel with variance $\sigma^2 = 10^{-6}$, taking the solution to the linear system as the initial point. Figures 5.20, 5.21, and 5.22 illustrate the results of applying the MCMC method to model calibration with real data. The acceptance rate is 99.91 % and the convergence rate is -0.0882.

The model seems to be inadequate for simulating the data, or the data may not meet the necessary assumptions to apply the calibration method, as no obvious trends were found between the data and the model results. It is expected to obtain higher quality and more extensive data in order to carry out a proper and consistent calibration of the TOMGRO model.

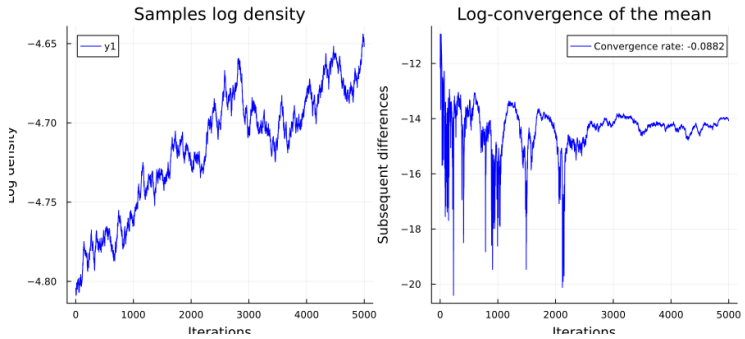


Figure 5.21: On the left, the log-density of the samples is shown, while on the right, the plot of the successive differences of the mean of the estimated parameters in logarithmic scale is displayed. Both graphs depend on the number of iterations. Figure created by the author using Julia (see [21]).

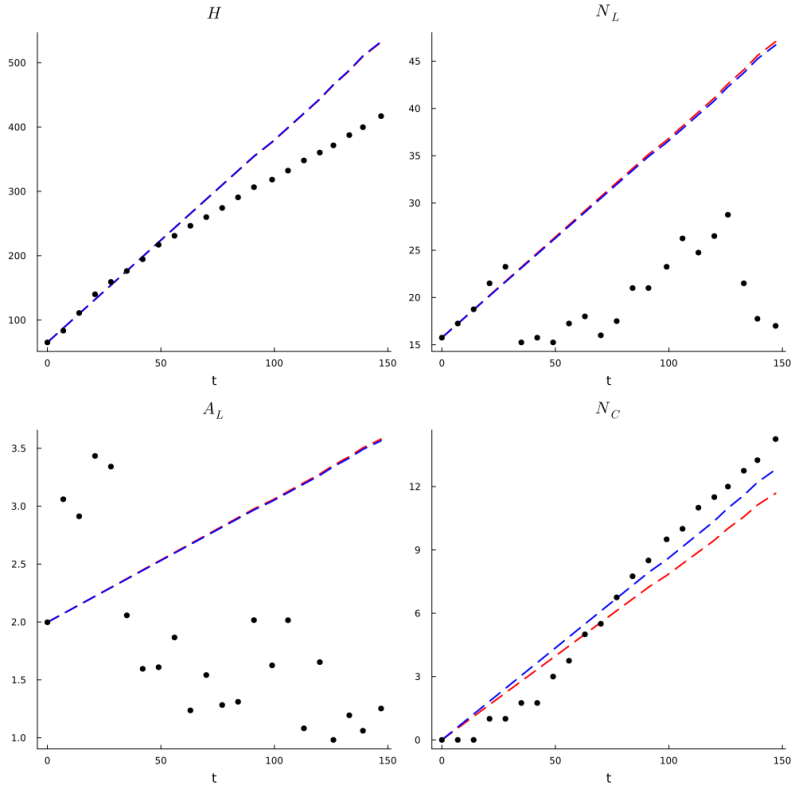


Figure 5.22: This figure presents the dynamics of the point statistics for the data obtained from the model's state variables. The black points represent the data, the red line represents the maximum a posteriori estimator, and the blue line corresponds to the conditional mean estimator. Figure created by the author using Julia (see [21]).

Chapter 6

Conclusions and Future Work

This study focused on analyzing parameter estimation methods within the framework of Markov Chain Monte Carlo (MCMC) algorithms, particularly the Metropolis-Hastings algorithm. The results highlight the effectiveness of these techniques in different scenarios and emphasize the importance of carefully selecting proposal distributions and tuning algorithmic parameters to achieve efficient convergence.

Future research directions could explore various extensions and refinements of MCMC methods. One potential line of study is the application of different variations of the Metropolis-Hastings algorithm that incorporate additional information about the target distribution, potentially improving convergence rates and estimation accuracy.

Another relevant research avenue involves studying the convergence properties of the Metropolis-Hastings algorithm in the context of parameter estimation for both linear models and differential equation models. A rigorous analysis of convergence criteria in these frameworks could provide deeper insights into the theoretical underpinnings of the algorithm and its applicability to complex dynamical systems.

Furthermore, the problem of prediction can be formulated as a non-stationary inverse problem that involves data assimilation techniques. In this regard, methodologies based on Markov models, such as Kalman filters, offer promising approaches to integrating new data dynamically and improving predictive performance.

A particularly interesting extension involves parameter estimation in other types of dynamical systems, including partial differential equations (PDEs) and stochastic

differential equations (SDEs). Investigating the feasibility and limitations of applying the concepts developed in this thesis to stochastic processes could yield valuable theoretical and practical contributions to the field.

These proposed research directions not only extend the scope of the current study but also open new possibilities for enhancing parameter estimation methodologies in various applied contexts.

Bibliography

1. Chowell, G., Nishiura, H., & Bettencourt, L. M. A. (2007). Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. Journal of The Royal Society Interface, 4(12), 155-166. https://doi.org/10.1098/rsif.2006.0161

2. Christen, J. A., & Fox, C. (2010). A general purpose sampling algorithm for continuous distributions (The t-walk). Bayesian Analysis, 5(2). https://doi.org/ 10.1214/10-BA603

3. Cohn, D. L. (2013). Measure theory (Vol. 2). New York: Birkhäuser.

4. Cooman, A. (2002). Feasibility of protected tomato cropping in the high altitude tropics using statistical and system dynamic models for plant growth and development [Doctoral thesis]. Katholieke Universiteit Leuven.

5. Evensen, G., Vossepoel, F. C., & Van Leeuwen, P. J. (2022). Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem. Springer International Publishing. https://doi.org/10.1007/978-3-030-96709-3

6. Galvis, J., & Capistrán, M. (2021). Beyond Research: Asimilación de datos en sistemas complejos [Course].

7. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.

8. Häggström, O. (2002). *Finite Markov chains and algorithmic applications* (Vol. 52). Cambridge University Press.

9. J. W. Jones, E. Dayan, L. H. Allen, H. Van Keulen, & H. Challa. (1991). A dynamic tomato growth and yield model (Tomgro). Transactions of the ASAE, 34(2), 0663-0672. https://doi.org/10.13031/2013.31715

10. Kaipio, J., & Somersalo, E. (2005). Statistical and computational inverse problems. Springer.

11. Kincaid, D. R., & Cheney, E. W. (2009). Numerical analysis: mathematics of scientific computing (Vol. 2). American Mathematical Soc.

12. Roy, V. (2019). Convergence diagnostics for markov chain monte carlo (No. arXiv:1909.11827). arXiv. https://doi.org/10.48550/arXiv.1909.11827

13. Sotomayor, J. (2011). Equações diferenciais ordinárias.

14. Strogatz, S. H. (2018). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC press.

15. Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, *2*(8), 492-494. https://doi.org/10.1038/nphys375

16. The t-walk. (n.d.). Retrieved January 7, 2025, from https://www.cimat.mx/jac/twalk/

17. Tierney, L. (1994). Markov chains for exploring posterior distributions. The Annals of Statistics, 22(4). https://doi.org/10.1214/aos/1176325750

18. Castañeda, Liliana Blanco, et al. Introduction to Probability and Stochastic Processes with Applications. John Wiley & Sons, 2012.

19. Axler, Sheldon. Linear Algebra Done Right. Springer International Publishing, 2024. DOI.org (Crossref), https://doi.org/10.1007/978-3-031-41026-0.

20. Bartle, Robert G., y Donald R. Sherbert. Introduction to Real Analysis. Wiley, 1982.

21. Cnosa. (2025). Cnosa/metropolis-hastings [Jupyter and Julia notebooks]. https://github.com/cnosa/Metropolis-Hastings (Original work published in 2024)

22. Phase plane plotter. (2019). Retrieved March 1, 2025, from https://aeb019.hosted.uark.edu/pplane.html.

23. Yang, L., Liu, H., Tang, X., & Li, L. (2022). Tomato Evapotranspiration, Crop Coefficient and Irrigation Water Use Efficiency in the Winter Period in a Sunken Chinese Solar Greenhouse. Water, 14(15), 2410. https://doi.org/10.3390/w14152410

Appendix A

Probability

This appendix provides a summary of the fundamental concepts in probability theory, focusing on measure theory, random variables, and conditional probabilities. These concepts are essential for understanding statistical inversion theory. All definitions and results are supported in [18].

Definition A.1. (**Probability space**) A probability space is defined as a triple $(\Omega, \mathcal{S}, P)$, where:

- Ω is the sample space, representing all possible outcomes of an experiment.
- $\mathcal{S}$ is a σ-algebra, a collection of subsets of Ω that satisfies the following properties:
  1. $\Omega \in \mathcal{S}$.
  2. If $A \in \mathcal{S}$, then its complement $A^c = \Omega \setminus A$ also belongs to $\mathcal{S}$.
  3. For any countable collection of sets $\{A_i\}_{i=1}^{\infty}$ in $\mathcal{S}$, their union $\bigcup_{i=1}^{\infty} A_i$ is also in $\mathcal{S}$.
- P is a probability measure, a function $P : \mathcal{S} \to [0, 1]$ that satisfies:
  1. $P(A) \geq 0$ for all $A \in \mathcal{S}$.
  2. $P(\Omega) = 1$.
  3. For any countable collection of disjoint sets $\{A_i\}_{i=1}^{\infty}$ in $\mathcal{S}$, $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Definition A.2. (**Borel σ-algebra**) Let (X, d) be a metric space. The *Borel σ-algebra* on X, denoted by $\mathcal{B}(X)$, is the smallest σ-algebra that contains all open sets of X.

Explicitly, $\mathcal{B}(X)$ is the collection of subsets of X that can be formed by countable unions, countable intersections, and relative complements of open sets. That is, $\mathcal{B}(X)$ is the σ-algebra generated by the topology of X.

In the specific case where $X = \mathbb{R}^n$ with the standard Euclidean topology, the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra containing all open subsets of $\mathbb{R}^n$.

Definition A.3. (**Measure**) A measure μ over the tuple $(\Omega, \mathcal{S})$, where Ω is a samaple space and $\mathcal{S}$ is a $\sigma-$algebra defined over Ω, is a function $\mu : \mathcal{S} \to [0, \infty)$ that satisfies:

1. $\mu(\emptyset) = 0$.
2. $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ where $\{A_i\}$ is a countable collection of elements of $\mathcal{S}$, disjoint pairwise.

Definition A.4. A measure μ is called σ-finite if there exists a countable sequence of sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{S}$ such that $\Omega = \bigcup_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$ for all n. A probability measure is a special case where $\mu(\Omega) = 1$.

Definition A.5. (**Random variable**) A random variable X is a measurable function $X : \Omega \to \mathbb{R}^n$, meaning that for every Borel set $B \subset \mathbb{R}^n$, the preimage $X^{-1}(B) \in \mathcal{S}$. The probability distribution of X is the measure μ_X defined on the Borel σ-algebra of $\mathbb{R}^n$ as:

$$\mu_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}^n).$$

The **cumulative distribution function** (CDF) of X is given by:

$$F(x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n), \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

If X is absolutely continuous with respect to the Lebesgue measure, its **probability density function** (PDF) π_X satisfies:

$$\mu_X(B) = \int_B \pi_X(x)\, dx, \quad B \in \mathcal{B}(\mathbb{R}^n).$$

Definition A.6. The expectation of a random variable X is defined as:

$$E\{X\} = \int_{\Omega} X(\omega)\, dP(\omega) = \int_{\mathbb{R}^n} x\, d\mu_X(x).$$

The covariance matrix of X is given by:

$$\operatorname{cov}(X) = E\left\{(X - E\{X\})(X - E\{X\})^T\right\}.$$

For two random variables X_1 and X_2, their joint probability distribution is defined as:

$$\mu_{X_1X_2}(B_1, B_2) = P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)), \quad B_1 \in \mathcal{B}(\mathbb{R}^{n_1}),\, B_2 \in \mathcal{B}(\mathbb{R}^{n_2}).$$

Conditional probability is a fundamental concept in probability theory and statistical inference.

Definition A.7. (**Conditional probability**) For two events $A, B \in \mathcal{S}$ with $P(B) > 0$, the conditional probability of A given B is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

This definition extends to random variables. Let X_1 and X_2 be random variables with joint probability density $\pi(x_1, x_2)$. The conditional probability density of X_1 given $X_2 = x_2$ is:

$$\pi(x_1|x_2) = \frac{\pi(x_1, x_2)}{\pi(x_2)}, \quad \text{where} \quad \pi(x_2) = \int_{\mathbb{R}^{n_1}} \pi(x_1, x_2)\, dx_1 > 0.$$

The joint probability density can then be expressed as:

$$\pi(x_1, x_2) = \pi(x_1|x_2)\pi(x_2) = \pi(x_2|x_1)\pi(x_1).$$

This relationship is known as Bayes' formula and is central to Bayesian inference.

Definition A.8. (**Independence**) Two random variables X_1 and X_2 are independent if their joint probability distribution factorizes as:

$$\mu_{X_1X_2}(B_1, B_2) = \mu_{X_1}(B_1)\mu_{X_2}(B_2), \quad B_1 \in \mathcal{B}(\mathbb{R}^{n_1}),\, B_2 \in \mathcal{B}(\mathbb{R}^{n_2}).$$

For absolutely continuous random variables, independence is equivalent to:

$$\pi(x_1, x_2) = \pi(x_1)\pi(x_2) \quad \text{almost everywhere.}$$

Definition A.9. (Normal or Gaussian distribution) A random variable X follows a *univariate Gaussian distribution* (or normal distribution) with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if its probability density function (PDF) is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

This is denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$.

On the other hand, a random vector $\mathbf{X} \in \mathbb{R}^d$ follows a *multivariate Gaussian distribution* with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ (where $\boldsymbol{\Sigma}$ is symmetric and positive semi-definite) if its probability density function is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d.$$

This is denoted as $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Theorem A.10. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Define the normalized sum:

$$S_n = \frac{1}{n} \sum_{j=1}^{n} X_j.$$

The CLT states that:

$$\lim_{n \to \infty} P\left(\frac{S_n - \mu}{\sigma/\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

In other words, the distribution of S_n converges to a standard normal distribution as $n \to \infty$.

Definition A.11. (**Stochastic process**) A **stochastic process** is defined as a collection of random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra, and P is a probability measure. The random variables, indexed by some set T, all take values in the same measurable space S, which must be measurable with respect to some σ-algebra $\mathcal{S}$.

Formally, a stochastic process is a collection of S-valued random variables, which can be written as:

$$\{X(t) : t \in T\},$$

where each $X(t)$ is a random variable associated with the index $t \in T$. Historically, in many problems from the natural sciences, the index $t \in T$ represented time, and thus $X(t)$ represents a random variable observed at time t. A stochastic process can also be written as:

$$\{X(t, \omega) : t \in T\},$$

which emphasizes that the process is actually a function of two variables, $t \in T$ and $\omega \in \Omega$.

In some alternative formulations, a stochastic process may be interpreted as an S^T-valued random variable, where S^T is the space of all possible functions from the set T into the space S. This alternative definition requires additional regularity assumptions to be well-defined.

Appendix B

Linear algebra

This appendix compiles fundamental results from matrix algebra, focusing on their computational aspects (see [10,19]). A fundamental result in linear algebra is the singular value decomposition (SVD), which is stated as follows.

Theorem B.1. (**SVD decomposition**) Every matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed as

$$A = U\Lambda V^T,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and $\Lambda \in \mathbb{R}^{m\times n}$ is a diagonal matrix with nonnegative diagonal elements λ_j satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\min(m,n)} \geq 0$.

The proof follows an inductive approach based on partial diagonalization using orthogonal transformations.

Proof. Before proceeding with the proof, it is recalled that a diagonal matrix is of the form

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & . & . & \ddots & . \\ . & . & \ddots & \dots & \dots & \ddots & 0 \\ 0 & \dots & \lambda_m & 0 & \dots & 0 & \end{bmatrix} = [\mathrm{diag}(\lambda_1, \dots, \lambda_m), 0],$$

if $m \leq n$, and where 0 denotes a zero matrix of size $m \times (n-m)$. Similarly, if $m > n$, Λ is of the form

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & . \\ . & . & \ddots & \lambda_n \\ 0 & \dots & 0 & . \\ . & . & \dots & . \\ 0 & \dots & 0 & \end{bmatrix} = \begin{bmatrix} \mathrm{diag}(\lambda_1, \dots, \lambda_n) \\ 0 \end{bmatrix},$$

where 0 is a zero matrix of size $(m - n) \times n$. Briefly, Λ is written as $\mathrm{diag}(\lambda_1, \dots, \lambda_{\min(m,n)})$.

Let $\|A\| = \lambda_1$, assuming that $\lambda_1 \neq 0$. Let $x \in \mathbb{R}^n$ be a unit vector such that $\|Ax\| = \lambda_1$. Define $y = (1/\lambda_1)Ax \in \mathbb{R}^m$, meaning that y is also a unit vector. Vectors $v_2, \dots, v_n \in \mathbb{R}^n$ and $u_2, \dots, u_m \in \mathbb{R}^m$ are chosen such that $\{x, v_2, \dots, v_n\}$ forms an orthonormal basis in $\mathbb{R}^n$ and $\{y, u_2, \dots, u_m\}$ forms an orthonormal basis in $\mathbb{R}^m$.

Defining

$$V_1 = [x, v_2, \dots, v_n] \in \mathbb{R}^{n \times n}, \quad U_1 = [y, u_2, \dots, u_m] \in \mathbb{R}^{m \times m},$$

it follows that these matrices are orthogonal. Consequently,

$$A_1 = U_1^T A V_1 = \begin{bmatrix} y^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} [\lambda_1 y, Av_2, \dots, Av_n] = \begin{bmatrix} \lambda_1 & w^T \\ 0 & B \end{bmatrix},$$

where $w \in \mathbb{R}^{(n-1)}$ and $B \in \mathbb{R}^{(m-1)\times(n-1)}$. Since

$$A_1 \begin{bmatrix} \lambda_1 \\ w \end{bmatrix} = \begin{bmatrix} \lambda_1^2 + w^2 \\ Bw \end{bmatrix},$$

it follows that

$$\|A_1\| \geq \lambda_1^2 + \|w\|^2.$$

On the other hand, since an orthogonal transformation preserves the matrix norm,

$$\|A\| = \|A_1\| \geq \lambda_1^2 + \|w\|^2.$$

Therefore, $w = 0$.

Proceeding inductively, let $\lambda_2 = \|B\|$, where $\lambda_2 \leq \|A_1\| = \|A\| = \lambda_1$. If $\lambda_2 = 0$, then $B = 0$ and the proof is complete. If $\lambda_2 > 0$, orthogonal matrices $\tilde{U}_2 \in \mathbb{R}^{(m-1)\times(m-1)}$ and $\tilde{V}_2 \in \mathbb{R}^{(n-1)\times(n-1)}$ are found such that

$$\tilde{U}_2^T B \tilde{V}_2 = \begin{bmatrix} \lambda_2 & 0 \\ 0 & C \end{bmatrix},$$

for some $C \in \mathbb{R}^{(m-2)\times(n-2)}$. Defining

$$U_2 = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{U}_2 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 1 & 0 \\ 0 & \tilde{V}_2 \end{bmatrix},$$

orthogonal matrices are obtained such that

$$U_2^T U_1^T A V_1 V_2 = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & C \end{bmatrix}.$$

By iteratively applying this partial diagonalization through orthogonal transformations, the result follows. □

A key consequence of the singular value decomposition is the characterization of the null space and range of a matrix.

Theorem B.2. Let $A = U\Lambda V^T \in \mathbb{R}^{m\times n}$ be its singular value decomposition, and assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > \lambda_{p+1} = \cdots = \lambda_{\min(m,n)} = 0$. Defining

$$U = [u_1, \ldots, u_m], \quad V = [v_1, \ldots, v_n],$$

the kernel and range relations are given by

$$\begin{aligned} \mathrm{Ker}(A) &= \mathrm{span}\{v_{p+1}, \ldots, v_n\} = \mathrm{Ran}(A^T)^\perp, \\ \mathrm{Ker}(A^T) &= \mathrm{span}\{u_{p+1}, \ldots, u_m\} = \mathrm{Ran}(A)^\perp. \end{aligned}$$

Additionally, the orthogonal projectors onto $\mathrm{Ker}(A)$ and $\mathrm{Ker}(A^T)$ can be expressed in terms of V and U as

$$P = V_2 V_2^T, \quad \tilde{P} = U_2 U_2^T,$$

where V_2 and U_2 correspond to the basis vectors spanning the respective kernels.

These results provide a fundamental basis for further exploration of matrix decompositions and their applications in numerical analysis.

Definition B.3. (**Moore-Penrose pseudoinverse**) Let $A \in \mathbb{R}^{m \times n}$ be a matrix. The *Moore-Penrose pseudoinverse* of A, denoted by A^+, is the unique matrix $A^+ \in \mathbb{R}^{n \times m}$ that satisfies the following four conditions:

1. $AA^+A = A$.
2. $A^+AA^+ = A^+$.
3. $(AA^+)^T = AA^+$.
4. $(A^+A)^T = A^+A$.

The Moore-Penrose pseudoinverse generalizes the concept of the inverse of a matrix and exists for any matrix, even if it is not square or not of full rank.

Definition B.4. (**Condition number**) Let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix. The *condition number* of A, denoted as $\kappa(A)$, is defined as

$$\kappa(A) = \|A\| \|A^{-1}\|,$$

where $\|\cdot\|$ is a chosen matrix norm, typically the spectral norm (induced by the Euclidean norm). In this case,

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)},$$

where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ are the largest and smallest singular values of A, respectively.

The condition number measures the sensitivity of the solution of a linear system $Ax = b$ to perturbations in b or A. A high condition number indicates an ill-conditioned matrix, meaning that small changes in the input can lead to large variations in the output.

Appendix C

Analysis

In this appendix, several fundamental concepts from analysis are presented, including metric spaces, Lipschitz functions, Banach's fixed-point theorem, and function spaces (extracted from [10, 20]).

Definition C.1. A **metric space** (X, d) consists of a set X together with a function $d : X \times X \to \mathbb{R}$, called a **metric**, that satisfies the following properties for all $x, y, z \in X$:

- (Non-negativity) $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$.
- (Symmetry) $d(x, y) = d(y, x)$.
- (Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$.

Definition C.2. A function $f : X \to Y$ between metric spaces (X, d_X) and (Y, d_Y) is said to be *Lipschitz continuous* if there exists a constant $L \geq 0$ such that for all $x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2).$$

If $L < 1$, f is called a *contractive mapping*.

Definition C.3. (**Function spaces**) Let Ω be a subset of a metric space $\mathbb{R}^n$.

- $C(\Omega)$: The space of continuous functions on Ω.
- $C^k(\Omega)$: The space of functions with continuous derivatives up to order k in Ω.
- $C^\infty(\Omega)$: The space of infinitely differentiable functions in Ω.

- $L^p(\Omega)$: The space of measurable functions f such that $\int_\Omega |f(x)|^p dx < \infty$.

Theorem C.4. (**Banach fixed point theorem**) Let (X, d) be a complete metric space, and let $T : X \to X$ be a contractive mapping, i.e., there exists $0 \leq c < 1$ such that

$$d(T(x), T(y)) \leq c\, d(x, y), \quad \text{for all } x, y \in X.$$

Then, T has a unique fixed point $x^* \in X$ such that $T(x^*) = x^*$. Moreover, for any initial point $x_0 \in X$, the sequence defined by $x_{n+1} = T(x_n)$ converges to x^*.

Proof. Let $x_0 \in X$ be an arbitrary initial point, and consider the sequence defined by $x_n = T(x_{n-1})$ for $n \geq 1$. It is now shown that (x_n) is a Cauchy sequence. First, it holds that

$$d(x_n, x_{n+1}) = d(T(x_n), T(x_{n-1})) \leq L d(x_n, x_{n-1}).$$

Applying this inequality iteratively,

$$d(x_n, x_{n+1}) \leq L^n d(x_1, x_0).$$

For any $m > n$, using the triangle inequality,

$$d(x_n, x_m) \leq \sum_{k=n}^{m-1} d(x_k, x_{k+1}) \leq \sum_{k=n}^{m-1} L^k d(x_1, x_0).$$

Since the geometric series $\sum_{k=n}^{\infty} L^k$ converges, it follows that the sequence (x_n) is Cauchy. Given that X is complete, there exists a limit $x^* \in X$ such that $x_n \to x^*$. Taking limits in the recurrence relation $x_n = T(x_{n-1})$ leads to $T(x^*) = x^*$.

To establish uniqueness, suppose there exists another fixed point $y^* \neq x^*$. Then,

$$d(x^*, y^*) = d(T(x^*), T(y^*)) \leq L d(x^*, y^*),$$

which is impossible unless $d(x^*, y^*) = 0$. Thus, $x^* = y^*$, proving the uniqueness of the fixed point. □