Climate Risk Assessment and Management

Consolidated Lecture Notes

James Doss-Gollin

Table of contents

Welcome

Welcome to Climate Risk Assessment and Management, an online textbook **under construction** by James Doss-Gollin.

> **!** Under Construction
>
> This textbook is a work in progress. Currently it's largely a brain dump, but I am building it out incrementally for use in my own classes. As I add and organize content, I will update the chapter status codes: (*planning*), (*draft*), (*revision*), and (*ready*). Contributions are welcome!

Motivation and Scope

History

This project emerged from two courses taught at Rice University by James Doss-Gollin: CEVE 543 focused on climate hazard and extremes and CEVE 421/521 focused on risk management.

Aim

The book is motivated by questions like

- What is the probability distribution of wind speeds that a building structure might experience?
- What will the probability distribution of extreme rainfall be in 2050, and what drives uncertainty in this estimate?
- What is the probability distribution of tropical cyclone losses across a regional portfolio?
- When, and how high, should a house be elevated to proactively manage future flood risk?
- What are robust, efficient, and equitable strategies for reducing flood risk in an urban area?

These questions span scales and sectors, yet they share fundamental challenges: characterizing extreme events, quantifying uncertainty, assessing risks, and making robust decisions when probability distributions are unknown or contested. Moreover, there is not a single correct answer to these questions, or a single method that will incontrovertibly answer them.

How to Use This Resource

The book is designed to be useful for practitioners, students, and teachers. Teachers may use individual chapters in their courses. Students may use it as a class text or reference. Practitioners may focus on specific chapters relevant to their work. Each chapter includes learning objectives and can be read independently, though some chapters build on concepts introduced in others.

Structure

- The **Preface** introduces the book's motivation and frames key challenges
- **Part 1** introduces key topics in probability, inference, Bayesian methods, optimization, machine learning, and Earth science. Rather than providing a comprehensive treatment, this part focuses on essential concepts and links to further resources.
- **Part 2** focuses on hazard **assessment**, namely modeling climate hazards and extremes. Material is organized around thematic applications and predictive tasks. The foundational idea is integrating information from noisy and/or biased sources to estimate the joint probability distribution of relevant hydroclimatic variables.
- **Part 3** risk management, which involves both mapping hazard to risk and designing interventions to manage these risks. Key ideas include the sequential nature of decisions, the pursuit of unclear and/or contested objectives, and the need to account for the sensitivity of estimated probability distributions (of hazard and of other relevant physical, social, and economic variables) to underlying models and assumptions.
- **Computational notebooks** written in Julia illustrate and complement the methods and concepts discussed in the text. While notebooks are referenced in the text, they are designed as standalone and self-contained resources.

Prerequisites

Basic probability and multivariate calculus, along with linear algebra, are sufficient mathematical foundations for this textbook. Some exposure to Earth science, hydrology, water resources, or related topics is strongly encouraged for context, though not strictly necessary for understanding methods. This book builds on a wide range of topics and methods in statistics, machine learning, optimization, and Earth science, and expertise in any of these areas may deepen your understanding, but is not necessary. No programming is required to read the book, but going through computational examples and applying methods to your own problems, which can substantially strengthen your understanding, does require programming.

Part I

About this book

License

Contributing

This textbook is a work in progress, and we welcome your contributions. Whether it's fixing a typo or proposing a new module, every suggestion helps. The easiest way to contribute is to fork the repository and submit a pull request. If you're not comfortable with that workflow, please open an issue on GitHub.

Citing

Please cite this resource as

```
@book{doss-gollin_textbook:2025,
  author = {Doss-Gollin, James},
  title = {Climate Risk Management},
  year = {2025},
  url = {https://jdossgollin.github.io/climate-risk-book},
}
```

In the future, we will move to stable releases with numbered versions.

Further Reading

Climate risk assessment and management are complex and interdisciplinary topics, and we are by no means comprehensive here. This page provides some helpful resources (textbooks, detailed online tutorials, and class websites) for your continued and supplementary study.

Inspiration

This textbook draws inspiration and content from several courses and lecture notes, and I am grateful to the instructors who have shared their materials with me.

- Upmanu Lall's Environmental Data Analysis course at Columbia
- Vivek Srikrishnan's Environmental Systems Analysis and Climate Risk Analysis classes at Cornell
- R. Balaji's Advanced Data Analysis Techniques (Statistical Learning Techniques for Engineering and Science) course at CU Boulder
- Alberto Montanari's collection of open course notes and lectures
- **Applegate and Keller (2015)** motivates this project and demonstrates problem-based learning.

Stats + ML basics

This book assumes familiarity with these topics, but these resources may be helpful as a refresher.

- **Blitzstein and Hwang (2019)** provides a thorough introduction to key concepts and ideas in probability. The book accompanies a free online course, Stat 110, which is a great resource for learning probability and statistics. Practice problems and solutions, handouts, and lecture videos are all available online.
- **Downey (2021)** offers an introduction to Bayesian statistics using computational methods. It's not environment focused but provides code and a clear explanation of core concepts.
- **Gelman (2021)** is a textbook designed for a first course on applied statistics. Clear and well-worked examples underpin discussion of fundamental ideas in statistical analysis and thinking about data.

Applications

There are lots of related books on catastrophe modeling, water resources research, geostats, statistical hydrology and related topics. Here is an incomplete list of some core references.

- **Naghettini (2017)** is a textbook on statistical hydrology that covers many of the same topics as this course. The statistical hydrology literature often obfuscates key ideas with complex notation and terminology, but this book is a helpful introduction to the field.
- **Helsel et al. (2020)** is a comprehensive introduction to water resources and hydrology, focusing on statistical methods for analyzing hydrologic data. Its methods are traditional, with less emphasis on machine learning or Bayesian methods and more attention to null hypothesis significance testing, but its case studies are well-worked and thoughtfully described.
- **Abernathey (2024)** is an excellent resource covering introductory topics in Earth and climate data science using Python, with an emphasis on foundational computations. These core computational concepts serves as a recommended prerequisite for more advanced material in this book.
- **Pyrcz (2024)** is a textbook focused on applied machine learning, with a particular focus on geostatistics. There's less focus on extremes, hydroclimate, and decision-making, but it provides very clear and interpretable explanations of many machine learning methods, including some that are not directly covered in this book.
- **Mignan (2024)** is a modern introduction to catastrophe risk modeling that covers a wide range of hazards, including hydroclimatic extremes, from a physics-based perspective. It provides a structured framework for quantifying hazard, exposure, and vulnerability, following industry-standard CAT modeling approaches. While broader in scope and more introductory in level, it complements this book's focus by illustrating foundational principles of probabilistic risk modeling in practice.

More Stats + ML

This book covers a broad set of topics in statistics, machine learning, and optimization. Most chapters could be a textbook of their own, and in fact many exist.

- **Friedman, Hastie, and Tibshirani (2001)** is a classic introduction to machine learning, which complements the Bayesian perspective nicely.
- **Jaynes (2003)** is a classic text on probability theory that you should read if you're interested in questions like "what is probability?"
- **Gelman et al. (2014)** and **McElreath (2020)** are the classic textbooks on Bayesian inference and provide a wealth of insight and detail. The Gelman textbook is a bit more dense while the McElreath book has a more conversational tone, but both cover similar topics.
- **Cressie and Wikle (2011)** provides a detailed exploration of hierarchical space-time models. There have been some computational advances since then that are worth keeping in mind before you apply these models directly, but it's a clearly written and overview.
- **Thuerey et al. (2024)** is a new textbook on physics-based deep learning, which is a rapidly growing area of research. It provides a comprehensive overview of the field, including theoretical foundations and practical applications. It covers topics, including neural operators and diffusion models, that are not covered in this course, but which are increasingly used in the climate risk space.
- **Bishop and Bishop (2024)** is a comprehensive, modern, and accessible start-to-finish textbook covering machine learning from basic probability through diffusion models.
- Michael Betancourt's writing page has detailed and mathematically rigorous explanations of many topics in Bayesian data analysis and probabilistic modeling.

Preface

What is climate risk?

Climate risks arise at the intersection of climate hazards, exposed systems, and vulnerability. They manifest when extreme or changing climate conditions—floods, droughts, extreme temperatures, sea-level rise, or shifting precipitation patterns—impact human and natural systems that are exposed and vulnerable to these conditions. The financial sector terms these "physical risks" to distinguish them from transition risks related to policy and market changes.

Climate risks span scales from the hyperlocal (a single building's flood exposure) to the global (climate impacts on agricultural productivity). They encompass immediate acute risks from individual extreme events and longer-term chronic risks from gradual climate changes. Crucially, climate risks are not solely natural phenomena but emerge from the complex interactions between climate hazards and the human systems—infrastructure, institutions, communities, and economies—that experience their impacts.

Climate risk is often defined as the product of hazard (probability that something will happen) and consequences (exposure and vulnerability). However, it's often helpful to start with the decisions we care about.

Risk management

The goal of assessing climate risks is to manage them, as is the focus of Part III. We manage climate risks by

- **building infrastructure**, such as seawalls, stormwater pipes, oyster beds, green roofs, dams
- **designing policy**, such as water pricing, land-use regulations, building codes
- **responding** to climate disasters through disaster response and recovery. While emergency management is beyond the scope of the book, disaster prevention (through infrastructure, policy, etc) and preparation (planning evacuation routes, assessing resource needs, etc) are problems that the tools of this class can inform.

A key insight from considering these applications is that climate risks are not natural phenomena, but occur at the intersection of natural and human systems. A second insight is that decisions about how to manage climate risks do not depend only on climate hazard, but also on human systems and values.

Exposure and vulnerability

Hazards do not create consequences by themselves. Hazards affect things that we care about, whether natural ecosystems, human homes, infrastructure systems, or something else. Quantitatively these are often described as exposure and vulnerability. However, this is not always a helpful framing because everything is exposed, to at least some degree, to climate hazards.

Climate hazard

Climate hazards have several key characteristics:

- **Location-specific impacts**: Specific weather patterns cause different things in different places—tropical cyclones cause extreme winds on the Gulf Coast, while persistent intense rainfall causes flooding in major rivers
- **Require Earth science and data**: Understanding hazards requires both physical process knowledge and empirical data
- **Variable focus on extremes**: Some applications care about extremes, but others (e.g., water management) care about shifts in the whole distribution
- **Multi-scale variability**: Characterized by variability across multiple spatial and temporal scales

What are good strategies?

The simple story

In principle, managing climate risks should be straightforward. If we had clear objectives and well-characterized uncertainty, there are **established mathematical formalisms for decision-making under uncertainty**. Notably, Bayesian Decision Theory provides an elegant framework: find the action a that maximizes expected utility

$$\mathbb{E}[U(a)] = \int U(a, s) p(s) ds,$$

where $U(a, s)$ is the utility of action a given s, and $p(s)$ is the over states of the world. The $\mathbb{E}[U(a)]$ represents the average utility we would expect from action a across all possible future states, weighted by their probabilities (see Chapter on Probability and Statistics for mathematical foundations).

With this framework and modern advances in operations research and optimization, we could frame climate risk management as a large-scale optimization problem. This might still be a challenging problem, requiring sophisticated optimization methods, large-ensemble Monte Carlo simulation, high-performance computing, and more, but fundamentally **there would be a right answer** that we could identify, at least seek to approximate.

Why this isn't enough

In practice, climate risk management defies this idealized approach for several fundamental reasons:

1. **Deep uncertainty**: Unlike textbook optimization problems, we rarely have well-defined probability distributions over future states. Climate risks involve poorly characterized, multiple, and interacting uncertainties spanning physical processes (climate projections), socioeconomic factors (development patterns, institutional capacity, human behavior), and their complex dependencies. The probability distributions we need span climate hazards, exposure patterns, vulnerability functions, and policy effectiveness—all evolving in ways that resist precise characterization.
2. **Large and poorly defined decision spaces**: The solution space includes not just individual projects but entire systems: infrastructure networks, policy portfolios, risk transfer arrangements, and adaptive management sequences. These decisions interact across scales, sectors, and time horizons in ways that resist comprehensive optimization.
3. **Contested objectives**: Different stakeholders hold different values about what we should optimize for—economic efficiency, equity, robustness, or flexibility. These objectives often conflict, and their relative importance is itself contested and evolving.

This brings us to a crucial insight: **we cannot simply frame climate risk management as a big optimization problem**. The field has witnessed an explosion of computational tools—climate models with ever-finer resolution, machine learning algorithms for processing vast datasets, and sophisticated visualization platforms for rendering complex projections. While these advances represent genuine progress, their proliferation has created new challenges for practitioners seeking to manage real-world climate risks.

The abundance of available tools does not automatically translate to better decisions. Indeed, the sophistication of modern computational approaches can obscure fundamental questions about problem framing, uncertainty characterization, and appropriate methods selection. Without solid conceptual foundations, practitioners may find themselves applying powerful tools inappropriately or mistaking methodological novelty for substantive insight.

The stakes of getting it wrong

The consequences of inadequate climate risk management are severe and diverse. **Infrastructure failures** occur when designs based on historical extremes prove insufficient for future conditions—leading to flooded neighborhoods when storm drains are undersized, or to costly over-design when extreme projections are treated as certainties. **Policy mistakes** compound these problems: development policies that ignore flood risks concentrate vulnerable populations in harm's way, while overly conservative regulations can stifle economic development without commensurate risk reduction benefits.

Financial miscalculations affect both public and private sectors. Insurance companies that underestimate climate risks face catastrophic losses, while those that overestimate risks price themselves out of markets. Infrastructure investors struggle to balance climate resilience against cost constraints, often erring toward solutions that prove either inadequate or prohibitively expensive. These failures cascade across scales: a poorly designed local drainage system contributes to regional flood management challenges, while flawed national climate risk assessments misguide infrastructure investment priorities across entire countries.

This book

This book develops both the technical tools and conceptual frameworks needed for climate risk management:

- **Part I** provides the statistical, optimization, and machine learning foundations that enable rigorous analysis of climate risks and decision alternatives
- **Part II** focuses on characterizing climate hazards and their uncertainties, emphasizing the integration of multiple imperfect information sources
- **Part III** addresses the transition from hazard to risk and the design of management strategies under deep uncertainty

Throughout, we emphasize that technical sophistication must be coupled with conceptual clarity about the nature of climate risks and the limits of optimization approaches. The goal is not to abandon quantitative analysis, but to use it more wisely—focusing computational power where it adds most value while acknowledging the irreducible uncertainties that require adaptive, robust approaches to climate risk management.

This book aims to teach readers how to **apply** tools from applied mathematics, statistics, and machine learning to answer questions such as

- What is the probability distribution of some relevant hazards or variables, such as (rainfall, wind, flood, temperature, streamflows) at a specific location?
- How do these probability distributions change in the next 50 years?
- How uncertain are these estimates and what specific mechanisms drive these uncertainties?
- What is the distribution of annual losses of a portfolio of assets exposed to one or many climate risks?
- What are trade-offs between up-front costs and future damages for decisions like how high to elevate a house?
- What are robust strategies for sequentially hardening infrastructure against climate risks?
- What are trade-offs between flood and drought protection for managing a reservoir?

While Part I does provide building blocks, they are intended to be self-contained references rather than a comprehensive overview to applied math, statistics, computer science, machine learning, and operations research. Instead, it aims to give you "just enough" context to think carefully about how to apply tools from these fields to climate risk management challenges.

What this book is not

This book focuses on the technical foundations of climate risk assessment and quantitative decision-making under uncertainty. While we address design requirements, social dimensions, and stakeholder considerations throughout—recognizing that technical tools can significantly inform these challenges—there are important aspects of climate risk management that require specialized expertise beyond our scope.

This book will **not** primarily teach you how to:

- **Manage reputational and transition risks**: While we focus on physical climate risks and their quantitative assessment, organizations also face complex risks from changing policies, markets, and stakeholder expectations that require specialized risk management expertise
- **Design and implement adaptive organizations**: While we cover adaptive management strategies and robust decision-making frameworks, the organizational design and management expertise needed to implement these approaches in practice requires additional specialized knowledge
- **Facilitate stakeholder processes**: While the quantitative tools we teach can strongly support consensus building by clarifying trade-offs and uncertainties, the facilitation, negotiation, and collaborative governance skills needed to lead stakeholder processes require specialized training
- **Develop communication strategies**: While we emphasize how to interpret and present quantitative risk assessments, developing effective communication strategies for diverse audiences—policymakers, communities, investors—requires specialized expertise in science communication and public engagement
- **Navigate implementation challenges**: While we address policy design and infrastructure planning from an analytical perspective, the practical challenges of construction management, regulatory processes, and community engagement require domain-specific expertise

This is an interdisciplinary text that draws insights from multiple fields and acknowledges the social, political, and institutional contexts that shape climate risk management. However, our primary focus remains on the quantitative and analytical foundations that can inform—but not replace—the broader expertise needed for effective practice.

Part II

I: Foundations

1 Climate science

Learning objectives

After reading this chapter, you should be able to:

- Understand why climate science is essential for risk assessment
- Identify key climate science topics you should study for effective risk management
- Navigate to appropriate resources for learning climate science fundamentals

1.1 Why climate science matters for risk assessment

Climate hazards don't occur in isolation. A hurricane's intensity depends on sea surface temperatures and atmospheric conditions. Droughts emerge from large-scale circulation patterns and ocean-atmosphere interactions. Floods reflect not just local rainfall but also broader weather systems and seasonal cycles.

Understanding these connections is crucial for risk assessment because climate science helps us answer three fundamental questions:

1. **What physical processes create hazardous weather patterns?** Understanding the mechanisms behind hurricanes, droughts, heat waves, and floods helps us identify when and where they're most likely to occur.

2. **How do these patterns vary naturally over time?** Climate systems exhibit variability on multiple timescales—from seasonal cycles to multi-decadal oscillations—that affect the frequency and intensity of extreme events.

3. **How might climate change alter these patterns?** As greenhouse gas concentrations rise, the statistical properties of weather and climate are shifting, requiring us to account for non-stationarity in our risk assessments.

1.2 Essential climate science topics

1. Climate models and modeling
2. Multi-scale variability
3. Specific weather patterns
4. Climate change and sensitivity
5. Hydrologic cycle

Further reading

- Mudelsee (2020): statistical approaches to climate extremes
- Merz et al. (2014): flood risk methods connecting climate to impacts
- Ghil et al. (2011): physical processes and extreme value behavior

2 Probability and inference

This chapter covers the fundamental concepts of probabilistic modeling and statistical inference. Data analysis proceeds by building a generative model—a formal, probabilistic hypothesis about how data are created. Inference is the inverse problem of using observed data to learn about model parameters. Computational methods make complex inference problems tractable.

Learning objectives

After reading this chapter, you should be able to:

- Build generative models using probability distributions.
- Apply maximum likelihood and Bayesian inference to estimate parameters.
- Use Monte Carlo methods when analytical solutions don't exist.
- Recognize when computational methods are required versus analytical approaches.

2.1 Probability theory

The concepts in this section provide the mathematical language for describing uncertainty. These mathematical tools support practical modeling applications.

2.1.1 Basic concepts

2.1.1.1 Random variables

A random variable is a function that assigns numerical values to the outcomes of a random experiment. Random variables provide the mathematical foundation for describing uncertainty.

- Discrete random variables take on countable values (e.g., number of floods per year)
- Continuous random variables take on uncountable values (e.g., temperature, precipitation amount)

2.1.1.2 Notation conventions

- Random variables: Capital letters (X, Y, Z)
- Realizations (specific values): Lowercase letters (x, y, z)
- Parameters: Greek letters (θ, μ, σ)
- Observed data: y (following Bayesian convention)
- Predictions: $\tilde{y}$ (y-tilde)

2.1.2 Distribution functions

The foundation of probability theory rests on three fundamental functions that describe random variables.

2.1.2.1 Probability Mass Function (PMF)

For discrete random variables, $P(X = x)$ gives the probability that the variable takes on a specific value x. The PMF satisfies $\sum_x P(X = x) = 1$.

2.1.2.2 Probability Density Function (PDF)

For continuous random variables, $p(x)$ describes the relative likelihood of different values.

PDF $p(x)$ is not a probability but a **density**. Since probability is density multiplied by a (potentially very small) interval of x, the value of $p(x)$ itself can exceed 1 without violating the laws of probability. Probabilities are areas under the curve: $P(a \leq X \leq b) = \int_a^b p(x)\,dx$. PDFs are sometimes written as $f(x)$ or $f_X(x)$ and must satisfy $\int_{-\infty}^{\infty} p(x)\,dx = 1$.

2.1.2.3 Cumulative Distribution Function (CDF)

$F(x) = P(X \leq x)$ gives the probability that a random variable is less than or equal to x. The CDF is the most fundamental descriptor, defined for all random variables (discrete and continuous). It unifies probability concepts and is essential for quantiles and return periods.

2.1.2.4 Quantile function

The quantile function $Q(p) = F^{-1}(p)$ is the inverse of the CDF. It takes a probability $p \in [0, 1]$ and returns the value x such that $P(X \leq x) = p$.

For example, the median is $Q(0.5)$. Then 99th percentile is $Q(0.99)$.

2.1.2.5 Examples of key distribution functions

Understanding these concepts requires working through concrete examples. The Distribution Examples notebook demonstrates these relationships using normal and Poisson distributions.

These examples show both forward operations (finding probabilities from values) and inverse operations (finding values from probabilities), illustrating how the three fundamental functions work together to characterize uncertainty.

2.1.3 Multiple variables

2.1.3.1 Joint, marginal, and conditional distributions

Real systems involve multiple random variables, requiring tools to describe their relationships. This machinery allows construction of complex models from simpler components.

The **joint distribution** $p(x, y)$ for continuous or $P(X = x, Y = y)$ for discrete gives the probability of events occurring together.

The **marginal distribution** $p(x)$ or $P(X = x)$ gives the probability of an event, irrespective of other variables. Calculated by summing or integrating over the other variables: $p(x) = \int p(x, y)\, dy$.

The **conditional distribution** $p(y \mid x)$ or $P(Y = y \mid X = x)$ gives the probability of an event given that another event has occurred. Conditional distributions describe how variables depend on each other.

2.1.3.2 Visualizing joint, marginal, and conditional distributions

Understanding the relationships between joint, marginal, and conditional distributions becomes clearer with visualization. The Distribution Examples notebook includes a comprehensive multivariate example showing how joint distributions decompose into marginal and conditional components.

2.1.3.3 Independence

Two random variables X and Y are independent if their joint distribution is the product of their marginal distributions: $p(x, y) = p(x)p(y)$ for continuous variables, and $P(X = x, Y = y) = P(X = x)P(Y = y)$ for discrete variables. For example, temperature and rainfall on a given day are typically **not** independent—hot days often have lower rainfall probability.

2.1.3.4 IID (Independent and Identically Distributed)

A sequence of random variables that are independent and have the same distribution. Many statistical models assume IID data points, which enables powerful analytical and computational techniques.

2.1.3.5 Bayes' rule

The mechanical relationship between joint, marginal, and conditional distributions:

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)}$$

Bayes' rule is a consequence of the definition of conditional probability. It becomes a tool for inference when interpreted probabilistically.

2.2 Statistical foundations

This section provides the mathematical toolkit that underlies statistical inference. These results justify why statistical methods work and enable practical computation.

2.2.1 Summary statistics

2.2.1.1 Expectation (Expected Value)

The expectation is the formal definition of the quantity we approximate with the sample mean in a Monte Carlo simulation. The expectation of a function $g(X)$ is:

$$\mathbb{E}[g(X)] = \int g(x)p(x)\,dx$$

for continuous variables, or

$$\mathbb{E}[g(X)] = \sum_x g(x)P(X = x)$$

for discrete variables.

2.2.1.2 Moments of a distribution

Probability distributions are completely described by their PDF/PMF and CDF, but we often need summary statistics that capture essential properties.

- Mean: $\mu = \mathbb{E}[X]$ measures central tendency
- Variance: $\sigma^2 = \mathbb{E}[(X - \mu)^2]$ measures spread or scale; standard deviation is $\sigma = \sqrt{\sigma^2}$
- Higher-order moments: Skewness measures asymmetry, kurtosis measures tail weight

For certain heavy-tailed distributions, some higher-order moments (or even the variance) may not exist because the defining integrals diverge.

2.2.2 Fundamental theorems

Two fundamental theorems provide the mathematical foundation for both statistical estimation and computational methods. These results justify why statistical methods work and when we can trust their results.

2.2.2.1 Law of Large Numbers

Subject to mild conditions, the sample mean converges to the expected value as the number of samples increases:

$$\frac{1}{N}\sum_{i=1}^{N} X_i \to \mathbb{E}[X]$$

This theorem underlies both parameter estimation and Monte Carlo simulation. It guarantees that maximum likelihood estimates become accurate with sufficient data, and that Monte Carlo approximations become precise with enough samples.

2.2.2.2 Central Limit Theorem

The distribution of a sample mean approaches a Normal distribution as the sample size increases, regardless of the underlying distribution shape:

$$\frac{\sqrt{N}(\bar{X}-\mu)}{\sigma} \to N(0,1)$$

where $\bar{X}$ is the sample mean, $\mu = \mathbb{E}[X]$, and $\sigma^2 = \mathrm{Var}(X)$.

This enables uncertainty quantification through confidence intervals and justifies the widespread use of normal approximations in statistical inference. The CLT explains why many phenomena follow normal distributions—they arise from sums of many small, independent effects.

2.2.3 Monte Carlo Expectations

Most decision-relevant quantities can be expressed as expectations. When analytical calculation is impossible, simulation provides a practical approximation method.

The basic Monte Carlo estimate of an expectation is:

$$\mathbb{E}[g(X)] \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)$$

where $x_1, x_2, \ldots, x_N$ are independent samples from the distribution of X. More sophisticated methods (importance sampling, MCMC) exist for drawing samples from more complex distributions.

The Law of Large Numbers guarantees convergence, while the Central Limit Theorem provides the convergence rate. The Monte Carlo standard error is approximately $\sigma/\sqrt{N}$, where σ is the standard deviation of $g(X)$. This means that to halve the error, we need four times as many samples.

Monte Carlo methods become essential when dealing with high-dimensional integrals that arise in Bayesian inference and uncertainty propagation through complex models.

2.2.4 Transformation of variables

A core task in probabilistic modeling is understanding how randomness propagates through a system. If we have a random variable X with PDF $p_X(x)$ and create $Y = g(X)$, what is the PDF of Y?

Simply substituting $x = g^{-1}(y)$ in the original PDF is incorrect because functions can stretch or compress the probability space. We must account for this distortion.

The general change of variables formula is:

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

The term $\left| \frac{d}{dy} g^{-1}(y) \right|$ is the Jacobian—a "stretching factor" ensuring probability mass is conserved. When a function stretches a region, density decreases proportionally to keep total probability equal to 1. When it compresses a region, density increases.

This formula derives from working through CDFs and applying the chain rule, but the key insight is that transformations distort the coordinate system and we must adjust densities accordingly.

2.3 Likelihood and maximum likelihood estimation

The probability theory and statistical foundations we've covered provide the mathematical language for uncertainty and the tools for computation. We now turn from describing uncertainty to learning from data. The first major approach to statistical inference connects data to parameters through likelihood functions and optimization.

2.3.1 The likelihood function

The central tool for connecting data to parameters is the likelihood function. The likelihood is the conditional probability $p(y \mid \theta)$, where y represents our observed data.

2.3.1.1 Definition

The likelihood tells us how likely we are to see the observed data y for some value of the parameters θ.

2.3.1.2 Crucial distinction

The likelihood is not the probability of the parameters. It's the probability (or probability density) of the data given the parameters.

This confusion is common: $p(\text{data}|\text{parameters})$ tells us about data likelihood, not parameter probability. `MLE` provides point estimates of the most likely parameter values, while Bayesian inference provides probability distributions over parameters. Only Bayesian inference gives us $p(\text{parameters}|\text{data})$.

For continuous variables, since we're dealing with a density, the probability of getting exactly that value is zero, but the probability of getting near it is the integral of the PDF over a small interval.

2.3.1.3 Independence and the product form

If we assume our data points are independent and identically distributed (IID), then by the definition of independence:

$$p(y_1, y_2, \dots, y_n \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$$

2.3.1.4 The log-likelihood

Products are numerically unstable and difficult to work with. Since the logarithm is monotonic,

$$\arg\max_{\theta} p(y \mid \theta) = \arg\max_{\theta} \log p(y \mid \theta)$$

although

$$\max_{\theta} p(y \mid \theta) \neq \max_{\theta} \log p(y \mid \theta)$$

in general.

For independent data, this gives us:

$$\log p(y \mid \theta) = \sum_{i=1}^{n} \log p(y_i \mid \theta)$$

2.3.2 Maximum likelihood estimation

Maximum likelihood estimation (MLE) finds the parameter values that maximize the likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} p(y \mid \theta)$$

2.3.2.1 Why maximum likelihood makes sense

The likelihood function $p(y \mid \theta)$ gives the probability of observed data under different parameter values. Maximum likelihood estimation finds the parameter values that maximize the probability of the observed data. The approach selects parameters that best explain the observations.

In practical applications, `MLE` estimates parameters of distributions describing observed phenomena by finding values that maximize the probability of historical observations. The estimates inform subsequent analysis and decision-making.

2.3.2.2 Implementation

We find the actual parameter values using optimization approaches. This may involve analytical differentiation (setting derivatives to zero) or numerical optimization methods when closed-form solutions don't exist.

This reframes the statistical problem of inference as a numerical problem of optimization.

2.3.2.3 Properties and theoretical foundations

Understanding when and why MLE works requires defining estimator quality. An estimator should be consistent (converge to the true value as sample size increases), efficient (achieve low variance), and unbiased (correct on average).

Under regularity conditions, MLE estimators have desirable asymptotic properties. As the sample size n grows large, the MLE estimator $\widehat{\theta}_{\text{MLE}}$ becomes consistent—it converges to the true parameter value θ_0.

2.3.2.4 Computational considerations

Finding maximum likelihood estimates requires different approaches depending on the complexity of the model.

Analytical solutions exist when we can solve

$$\frac{d}{d\theta} \log p(y \mid \theta) = 0$$

in closed form. This works for simple models like Normal distributions with known variance, or the coin flip example we examine below. These cases provide valuable intuition and serve as building blocks for more complex problems.

Numerical optimization becomes necessary when no closed-form solution exists. Practical challenges arise in applications. The likelihood surface may contain multiple local maxima, requiring different starting values to find the global optimum. Numerical stability requires working with log-likelihoods rather than products of small probabilities. Flat likelihood surfaces indicate that data contain limited information about parameters. All methods assume correct model specification – poor model approximations yield misleading results regardless of optimization quality.

2.3.3 Example: coin flip maximum likelihood estimation

The coin flip example provides a pedagogically clean introduction to maximum likelihood estimation. The Coin Flip Inference notebook demonstrates both the analytical derivation and computational implementation of MLE.

This example shows how MLE connects intuitive parameter estimates (the observed proportion) with formal statistical theory. The same principles apply whether estimating coin bias or the frequency of extreme climate events from historical data.

2.3.4 Linear regression example

Linear regression extends inference to multiple parameters, demonstrating the connections between curve fitting, maximum likelihood estimation, and Bayesian approaches.

The Linear Regression Examples notebook shows how the same statistical problem can be approached from three different philosophical perspectives, each providing different insights into parameter estimation and uncertainty quantification.

This example illustrates how probabilistic models connect optimization-based and fully Bayesian approaches, with implications for modeling relationships between climate variables and their impacts.

2.4 Bayesian inference

Maximum likelihood estimation provides point estimates of parameters by finding values that maximize the probability of observed data. Bayesian inference takes a fundamentally different approach: it treats parameters as random variables and computes full probability distributions that quantify uncertainty. In other words, rather than searching for $\hat{\theta}$ that maximizes the likelihood, Bayesian inference seeks to estimate the entire distribution $p(\theta \mid y)$.

2.4.1 Motivation and overview

Real-world risk assessment relies on multiple, imperfect data sources: short instrumental records, longer regional records, qualitative historical accounts, and physical constraints from models. Traditional statistical methods often struggle to formally integrate these different types of information into a single analysis. The Bayesian framework provides a principled solution by treating all knowledge—both prior beliefs and new data—as probability distributions that can be mathematically combined.

The core relationship is Bayes' rule:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

This deceptively simple equation describes how we update our beliefs:

- **Prior distribution** $p(\theta)$: Quantifies existing knowledge about parameters before analyzing the current dataset
- **Likelihood function** $p(y \mid \theta)$: The engine for learning from data (identical to the likelihood used in maximum likelihood estimation)
- **Posterior distribution** $p(\theta \mid y)$: Our updated beliefs after combining prior knowledge with observed data
- **Marginal likelihood** $p(y)$: A normalizing constant ensuring the posterior integrates to 1

Since $p(y)$ doesn't depend on θ for a fixed dataset, we often write:

$$p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$$

The result is fundamentally different from maximum likelihood estimation: instead of a single "best" parameter estimate, we obtain a full probability distribution that naturally quantifies uncertainty. This enables direct probabilistic statements like "there is a 95% probability that the parameter lies between these values."

2.4.2 Maximum A Posteriori: a bridge to optimization

Before exploring full Bayesian inference, we can find the single most probable parameter value given the data. Maximum A Posteriori (MAP) estimation finds the mode of the posterior distribution:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p(\theta \mid y) = \arg\max_{\theta} p(y \mid \theta)p(\theta)$$

Taking logarithms (since log is monotonic):

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta}[\log p(y \mid \theta) + \log p(\theta)]$$

This reveals an elegant connection to machine learning and optimization. The log-posterior decomposes into the familiar log-likelihood plus a log-prior term. The log-prior acts as a regularization penalty, preventing overfitting by favoring certain parameter values.

When the prior is uniform (non-informative), the log-prior is constant and MAP reduces to maximum likelihood estimation. When the prior is informative, it regularizes the estimate by pulling it toward prior beliefs. This is mathematically identical to penalized likelihood methods like Ridge regression (with Gaussian priors) or Lasso regression (with Laplace priors).

However, MAP provides only a point estimate and discards uncertainty information. To fully leverage the Bayesian framework, we need the entire posterior distribution.

2.4.3 Analytic solutions and conjugate priors

In special cases, we can compute the posterior distribution analytically using conjugate priors. A prior is conjugate to a likelihood if the posterior has the same functional form as the prior. This mathematical convenience allows us to update our beliefs with a simple algebraic formula.

The coin flip example demonstrates this perfectly. For the Binomial likelihood, the Beta distribution is conjugate. To see why, let's work through the mathematics.

Starting with our prior and likelihood:

- **Prior**: $\theta \sim \text{Beta}(\alpha, \beta)$ with PDF $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$
- **Likelihood**: $y \mid \theta \sim \text{Binomial}(n, \theta)$ with $p(y \mid \theta) \propto \theta^{y}(1-\theta)^{n-y}$

The posterior is proportional to the product of prior and likelihood:

$$p(\theta \mid y) \propto p(y \mid \theta) \cdot p(\theta) \propto \theta^{y}(1-\theta)^{n-y} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Combining the powers:

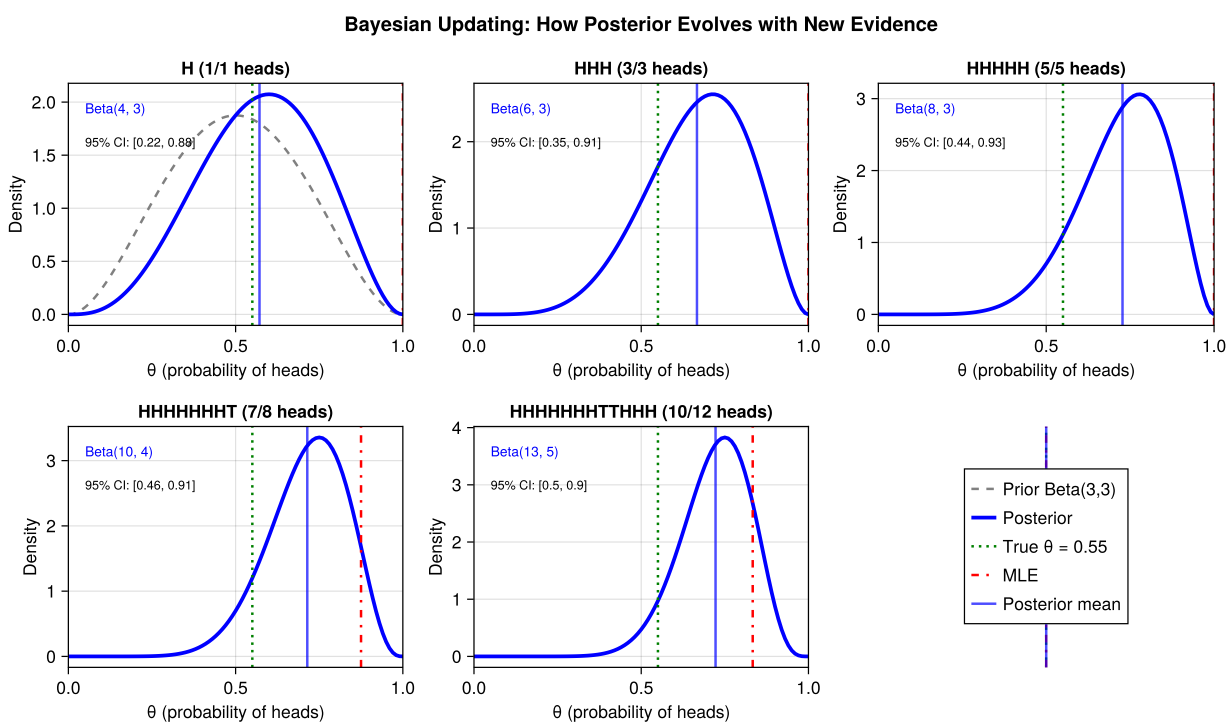
$$p(\theta \mid y) \propto \theta^{(\alpha+y)-1}(1-\theta)^{(\beta+n-y)-1}$$

This is exactly the kernel of a $\text{Beta}(\alpha + y, \beta + n - y)$ distribution! Therefore:

$$\theta \mid y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

The posterior parameters are intuitive: prior "successes" (α) plus observed successes (y), and prior "failures" (β) plus observed failures ($n - y$). This demonstrates the core Bayesian principle: new data updates our beliefs in a mathematically principled way.

The following example shows this updating process in action, demonstrating how the posterior distribution evolves with each new coin flip:



As data accumulate, the influence of the prior diminishes relative to that of the likelihood. With sufficient data, Bayesian and maximum likelihood estimates converge regardless of the prior choice.

In practice, conjugate priors exist for only a limited set of models.

2.4.4 Markov Chain Monte Carlo

For any non-trivial model, analytically computing the posterior distribution becomes mathematically intractable. A brute-force approach of evaluating the posterior on a grid fails catastrophically: with k parameters and n grid points per parameter, we need n^k evaluations. For even modest problems (say, 10 parameters with 100 grid points each), this requires $100^{10} = 10^{20}$ calculations—computationally impossible.

This "curse of dimensionality" means that analytical approaches work only for the simplest models. Real scientific applications require computational methods.

2.4.4.1 Example: A Metropolis-Hastings Sampler from Scratch

Modern MCMC algorithms differ in how they propose new parameter values. A classical, foundational approach is the **Metropolis-Hastings** algorithm, which gives us a powerful intuition for how MCMC works.

The algorithm allows us to sample from a **target** distribution that we can't directly sample from, as long as we can evaluate its density up to a normalizing constant. We use a separate **proposal** distribution to suggest new parameter values, and then we probabilistically accept or reject these proposals based on how well they align with the target.

(Shout out to Danielle Navarro for a clear explanation and the working example that inspired this section).

The algorithm proceeds as follows:

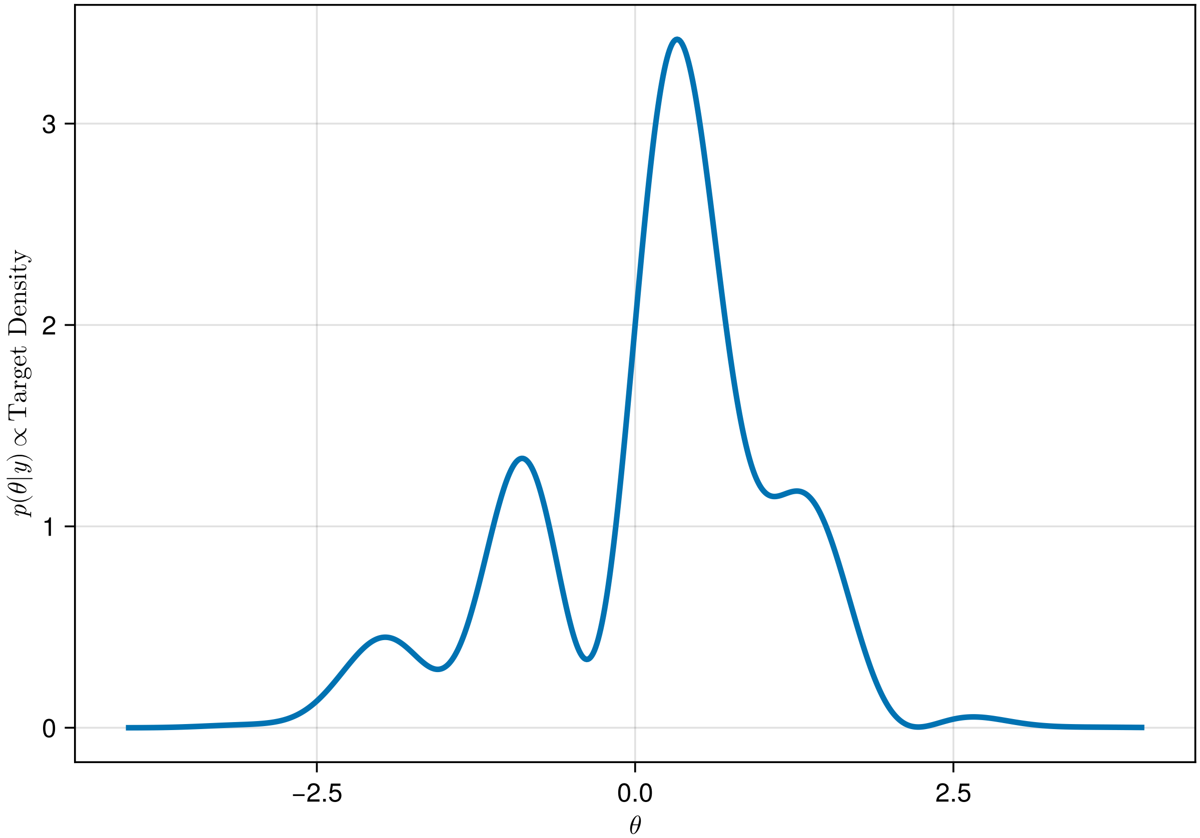
1. Start with an initial parameter value $\theta^{(0)}$.
2. For each iteration i:
    - Propose a new value θ^* from a proposal distribution $q(\theta^* \mid \theta^{(i-1)})$.
    - Compute the acceptance ratio:

    $$r = \frac{p(\theta^* \mid y)\, q(\theta^{(i-1)} \mid \theta^*)}{p(\theta^{(i-1)} \mid y)\, q(\theta^* \mid \theta^{(i-1)})}$$

    - Accept the proposal with probability $\min(1, r)$:
        - If a random number is less than r, set $\theta^{(i)} = \theta^*$.
        - Otherwise, reject the proposal and set $\theta^{(i)} = \theta^{(i-1)}$.
3. Repeat for many iterations to generate a chain of samples that approximates the posterior distribution $p(\theta \mid y)$.

2.4.4.1.1 Step 1: Target Distribution

First, let’s define the distribution we want to sample from. To make it interesting, we’ll use a “squiggly” function that isn’t a standard probability distribution. This is our unnormalized target density, which corresponds to $p(\theta^* \mid y)$ in the ratio above.



Our goal is to generate samples whose histogram matches the shape of this distribution.

2.4.4.1.2 Step 2: Proposal Distribution

Next, we need a way to propose new steps. For this example, we’ll use a **Normal distribution** centered on our current position. This is like telling our sampler to take a random “hop” left or right.

This choice is purely for didactic purposes. A Normal distribution is **symmetric**, meaning the probability of hopping from θ_A to θ_B is the same as hopping from θ_B to θ_A. This simplifies the acceptance ratio, as the proposal terms $q(\theta^{(i-1)} \mid \theta^*)$ and $q(\theta^* \mid \theta^{(i-1)})$ cancel out, leaving us with the simpler Metropolis algorithm.

The standard deviation, , controls the “hop size.” A larger will propose more distant jumps, while a smaller will explore the local area more finely.

2.4.4.1.3 Step 3: Sampler

Now we can write the core function that implements the Metropolis-Hastings logic. The function will take the number of samples, a starting value, and the proposal hop size as inputs. It will loop through the algorithm, building the chain one sample at a time.

2.4.4.1.4 Step 4: Implement

Finally, we run the sampler and plot the results. We need to check two things:

1. **The Trace Plot**: This shows the value of the parameter at each iteration. We are looking for a "fuzzy caterpillar" pattern, which indicates the chain is exploring the distribution well and is stationary.
2. **The Histogram**: This shows the distribution of the collected samples. If the sampler worked, its shape should match our target density.

```
Acceptance rate: 0.594
```

The acceptance rate is a useful diagnostic. A very high rate (> 0.8) might mean the hop size is too small, and the chain isn't exploring efficiently. A very low rate (< 0.1) might mean the hop size is too large, and most proposals are being rejected. Rates between 0.2 and 0.5 are often considered good for this type of sampler.

As we can see, the final histogram of samples beautifully recovers the shape of the target distribution, demonstrating that this simple algorithm successfully turned a function we can evaluate into a distribution we can sample from.

2.4.4.2 Modern MCMC algorithms

In practice, Metropolis-Hastings is rarely used due to its inefficiency and because it scales poorly to high-dimensional spaces where the acceptance rate can become very low. Modern MCMC algorithms, such as Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS), leverage gradient information to propose more informed jumps in parameter space. These methods explore the posterior distribution more efficiently, especially in high-dimensional problems common in climate modeling and other scientific applications.

WHen using these advanced samplers, several best practices ensure robust and reliable inference:

1. Running multiple independent chains helps assess convergence and identify multimodal posteriors. You will often see four chains as a default, but that's mainly because laptops ten years ago mainly had four cores. More is often better.
2. Initial iterations allow the sampler to adapt to the posterior geometry and find the typical set. We typically discard these "warm-up" samples before final inference.
3. Most samplers have hyperparameters (e.g., step size, number of leapfrog steps) that require tuning for optimal performance. Many software packages include automatic tuning during a warm-up phase.
4. Diagnostics are essential to ensure the sampler has converged and is exploring the posterior adequately.

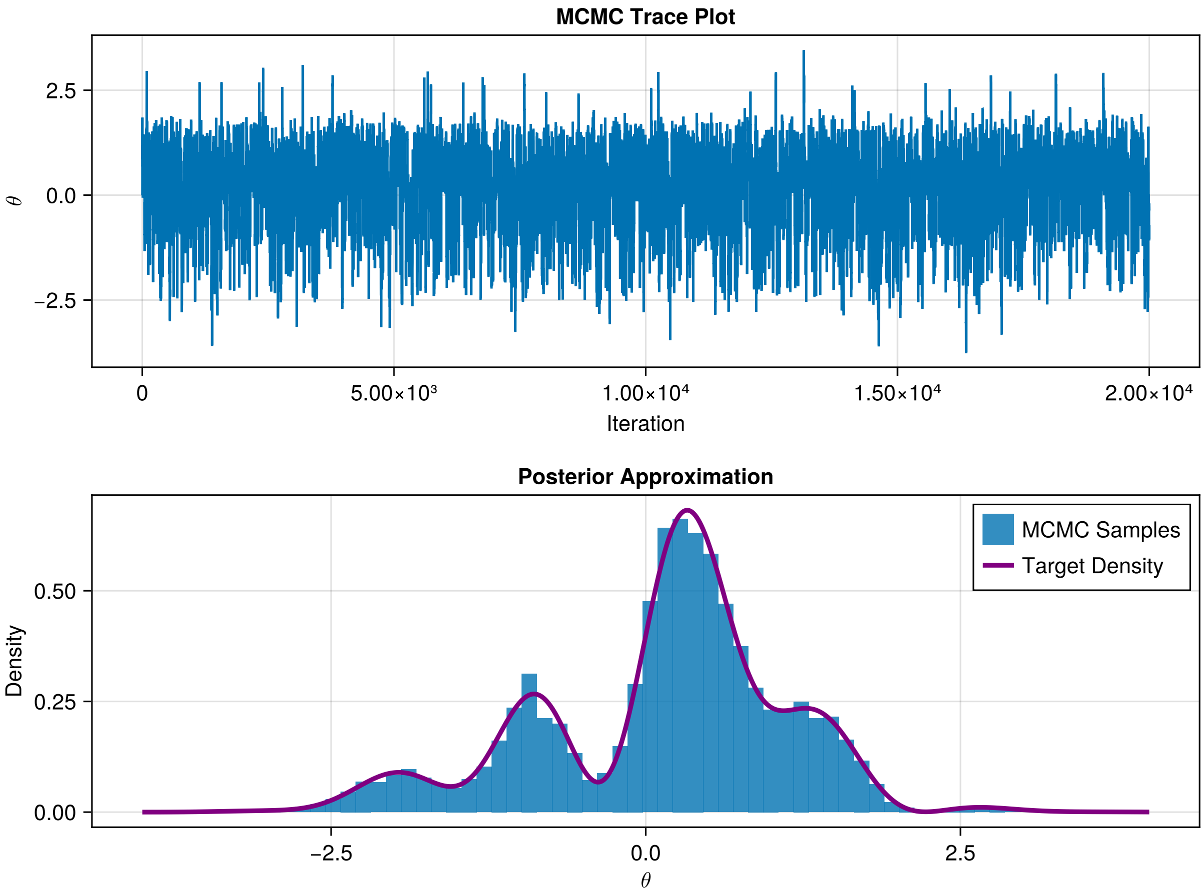


Figure 2.1: Results from our Metropolis-Hastings sampler. The trace plot (top) shows the sampler exploring the parameter space. The histogram of the MCMC samples (bottom) closely matches the shape of the unnormalized target density.

- Trace plots show the parameter values over iterations. We are looking for a "fuzzy caterpillar" pattern, which indicates the chain is exploring the distribution well and is stationary. If you see trends or drifts, the chain has not converged.
- Trace plots for multiple chains should overlap and mix well, indicating they are sampling from the same distribution.
- Autocorrelation plots help assess the degree of correlation between samples. Rapidly decreasing autocorrelation suggests good mixing, while high autocorrelation indicates the chain is stuck in local modes.
- Quantitative metrics, such as the $\hat{R}$ statistic and effective sample size, provide numerical assessments of convergence and sampling efficiency.

In all cases, we can rule out some kinds of convergence problems, but actually proving convergence is impossible, so we must always include other checks. Avoid the tendency common in applied work to state without reflection or critique that because $\hat{R} < 1.1$ no further thought is needed. For tricky problems, work with a computational statistician (or take Bayesian stats courses beyond the scope of these lecture notes).

2.4.4.3 From samples to inference

After running the chain for thousands of iterations, the resulting samples serve as a high-fidelity numerical approximation of the true posterior distribution. We can use these samples to approximate any posterior quantity:

$$\mathbb{E}[g(\theta) \mid y] \approx \frac{1}{N} \sum_{i=1}^{N} g(\theta^{(i)})$$

where $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)}$ are our `MCMC` samples after discarding warm-up iterations.

This Monte Carlo approximation enables: - Parameter estimates and credible intervals - Posterior predictive distributions for forecasting - Model comparison through marginal likelihood estimation - Decision analysis under parameter uncertainty

The quality of these approximations depends on sample size, chain mixing, and the specific posterior quantity being estimated. Tail probabilities and extreme quantiles require more samples than central tendencies.

2.4.5 Example: coin flip Bayesian analysis

The coin flip example demonstrates the complete Bayesian workflow, from prior specification through posterior computation and interpretation.

The Coin Flip Inference notebook shows how treating parameters as random variables enables full uncertainty quantification. This example compares analytical solutions (using conjugate priors) with computational approaches, validating MCMC methods against known exact results.

The example illustrates key Bayesian concepts: how priors influence posterior distributions, the role of conjugacy in analytical tractability, and the use of MCMC when analytical solutions don't exist.

2.4.6 Example: linear regression Bayesian analysis

Linear regression demonstrates Bayesian inference for multivariate problems requiring computational methods.

The Linear Regression Examples notebook provides a comprehensive comparison of least squares, maximum likelihood, and Bayesian approaches to the same problem. This comparison illuminates the philosophical differences between methods and their practical implications for uncertainty quantification.

The Bayesian approach excels when parameter uncertainty must be propagated through subsequent calculations or decision processes—a crucial capability for climate risk assessment and decision-making under uncertainty.

2.4.7 Posterior predictive distribution and model checking

The Bayesian workflow extends beyond parameter estimation to include systematic model validation through posterior predictive checking. This approach uses the fitted model to generate simulated data, then compares these simulations to observed data on relevant metrics.

The posterior predictive distribution represents our beliefs about new data given what we've observed:

$$p(y^{\text{rep}}|y^{\text{obs}}) = \int p(y^{\text{rep}}|\theta)p(\theta|y^{\text{obs}})d\theta$$

where y^{rep} denotes replicated (simulated) data and y^{obs} denotes observed data.

2.4.7.1 The posterior predictive checking workflow

The systematic approach follows these steps:

1. Use MCMC to obtain samples from $p(\theta|y^{\text{obs}})$
2. For each posterior sample $\theta^{(i)}$, draw $y^{\text{rep}(i)} \sim p(y|\theta^{(i)})$
3. Calculate relevant "test statistics" or metrics $T(y^{\text{rep}(i)})$ and $T(y^{\text{obs}})$
4. Assess whether $T(y^{\text{obs}})$ falls within the distribution of $T(y^{\text{rep}})$

This is related to the frequentist idea of p-values explored in the mosquitos case study.

Further reading

For deeper study of probability and statistics:

- Blitzstein and Hwang (2019) provides excellent intuition with computational examples
- Downey (2021) emphasizes Bayesian thinking with practical applications
- Gelman (2021) connects regression to broader statistical modeling
- Gelman et al. (2014) comprehensive treatment of Bayesian computation

This chapter's concepts are also demonstrated through focused computational notebooks:

- Distribution Examples: Understanding probability distributions and their relationships
- Coin Flip Inference: Introduction to statistical inference with likelihood and Bayesian methods
- Linear Regression Examples: Comparing curve fitting, MLE, and Bayesian approaches

3 Machine learning and nonparametric methods

Learning objectives

After reading this chapter, you should be able to:

- Apply fundamental supervised learning concepts to climate hazard assessment problems
- Understand nonparametric and semiparametric methods for flexible modeling
- Critically evaluate machine learning applications in climate risk literature
- Understand bias-variance tradeoffs and model validation approaches

3.1 Essential machine learning concepts

1. Supervised and unsupervised learning paradigms
2. Parametric vs nonparametric methods
3. Bias-variance tradeoff and regularization
4. Cross-validation and model selection
5. Tree-based methods and ensemble learning

Further reading

4 Correlation and dimensionality

Learning objectives

After reading this chapter, you should be able to:

- Model and interpret spatial dependence in climate fields
- Apply time series analysis methods to detect trends and patterns in climate data
- Use dimension reduction techniques for high-dimensional climate datasets
- Integrate spatial and temporal methods for spatiotemporal climate analysis

4.1 Essential concepts

1. Spatial statistics and geostatistical methods
2. Time series analysis and trend detection
3. Principal component analysis and empirical orthogonal functions
4. High-dimensional methods for climate data
5. Spatiotemporal integration approaches

Further reading

For spatial and temporal analysis in climate science:

- Cressie and Wikle (2011): Comprehensive treatment of spatial statistics

5 Model validation and comparison

Learning objectives

After reading this chapter, you should be able to:

- Apply graphical diagnostic methods to assess model fit quality
- Use information criteria (AIC, DIC, BIC) for quantitative model comparison
- Understand the relationship between model selection and predictive accuracy
- Recognize the subjective nature of model selection and make transparent choices

5.1 Essential model validation concepts

1. Graphical diagnostic methods and model checking
2. Information criteria for model comparison
3. Predictive accuracy and cross-validation
4. Model selection philosophy and transparency
5. Ensemble methods and model averaging

Further reading

- Piironen and Vehtari (2017): Technical treatment of predictive accuracy

Part III

II: Hazard Assessment

6 Extreme value theory

See first

This chapter builds on concepts from Probability and Statistics.

Learning objectives

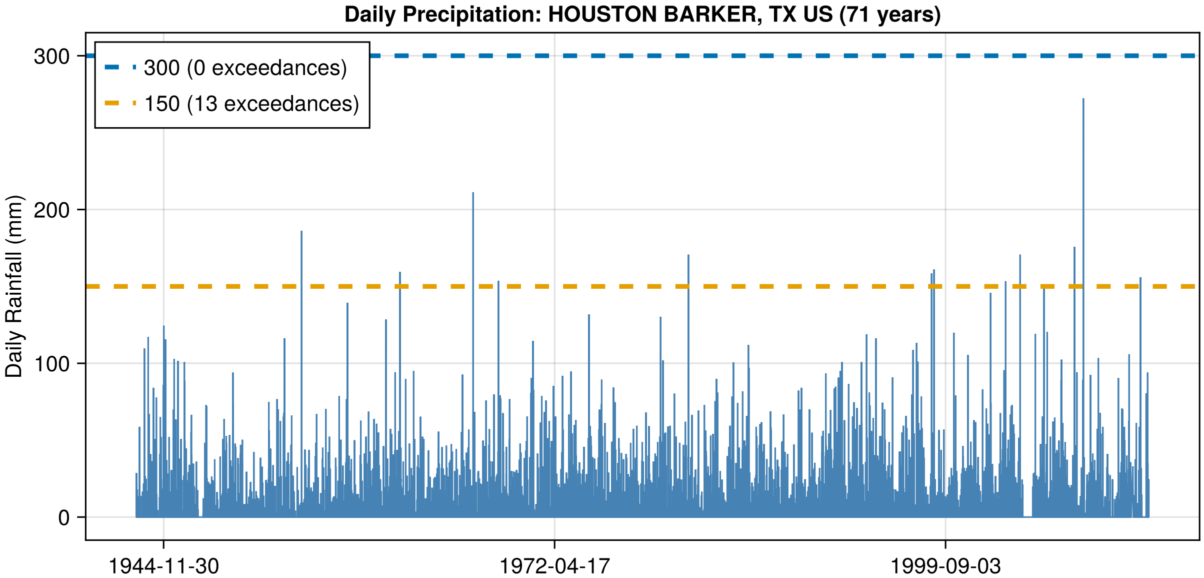
After reading this chapter, you should be able to:

- Understand the motivation for extreme value theory in climate risk assessment.
- Distinguish between block maxima and peak-over-threshold approaches.
- Select and fit appropriate extreme value distributions (GEV, GPD).
- Calculate return periods, return levels, and exceedance probabilities.
- Quantify and interpret different sources of uncertainty in extreme value analysis.
- Recognize the challenges posed by non-stationarity and climate change.
- Apply extreme value methods to real-world case studies.

6.1 Motivation

The design of reliable infrastructure, such as stormwater systems in Harris County, Texas, depends on a quantitative understanding of extreme environmental events. Engineers must characterize the magnitude and frequency of rare phenomena. For example, they might need to determine the expected recurrence interval of a daily rainfall total that exceeds a critical design threshold, such as 300 mm.

An initial, intuitive approach is to analyze the historical record.



Analysis of over 70 years of daily precipitation data from a station in Houston reveals 13 days where rainfall exceeded 150 mm. This allows for a simple empirical estimate of the event's frequency. Thirteen exceedances in 71 years suggests an average recurrence interval of approximately 5.5 years. For the 300 mm threshold, however, the historical record contains zero events. An empirical estimate of the probability is therefore zero, which is uninformative for risk assessment and infrastructure design. We require a method for extrapolating beyond the range of observed data.

A natural extension would be to fit a standard parametric probability distribution—such as a Gamma distribution, which is commonly used for precipitation—to the entire set of daily rainfall observations. One could then use the extreme upper tail of the fitted distribution to estimate the probability of exceeding 300 mm.

This approach, however, is fundamentally unreliable. The goodness-of-fit of a parametric model is driven by its ability to describe the bulk of the data, where observations are plentiful. Minor model misspecifications in this high-density region can translate into substantial errors in the far tails of the distribution, where data are sparse or nonexistent. As stated by Coles (2001), the foundational text on this topic, "very small discrepancies in the estimate of F can lead to substantial discrepancies for F^n", where in our example F is the cumulative distribution function (CDF) of daily rainfall while F^n is the CDF of the maximum daily rainfall over n days.

The models that best describe the central tendency of a process are not necessarily suitable for describing its extremes. This fragility of tail extrapolation necessitates a theoretical framework developed specifically for modeling the behavior of extreme values.

A naive empirical estimator for the exceedance probability of a value x based on n observations X_i is the simple proportion of exceedances:

$$\hat{P}(X > x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i > x)$$

where $\mathbf{1}(\cdot)$ is the indicator function. This is useful when the number of exceedances is large, but fails for extrapolation when the sum is zero, as was the case for the 300mm threshold. This mathematically demonstrates the limitations of simple empirical methods and motivates the specialized approaches that follow.

6.2 Approaches for modeling extreme values

Extreme Value Theory (EVT) provides such a framework. Rather than modeling the entire distribution of a process, EVT focuses on the distribution of its extreme values. There are two primary methods for extracting these values from a time series.

6.2.1 Block maxima

In this method, the period of observation is divided into non-overlapping blocks of equal duration (e.g., years). For each block of size n, we define the block maximum as $M_n = \max\{X_1, \dots, X_n\}$. This yields a new time series of block maxima. A crucial conceptual point is that we are not studying a single maximum possible value of the process. Instead, we are studying the distribution of the sample maximum, M_n, which is itself a random variable. This is why a distribution like the GEV is needed to describe its behavior. This approach is intuitive and directly relates to concepts of annual risk, but it can be inefficient as it discards other potentially extreme events within a block.

6.2.2 Peaks-over-threshold

The peaks-over-threshold (POT) method defines a set of exceedances over a high threshold u as $\{X_i : X_i > u\}$. The variable of interest is the value of the **excesses** themselves, $Y = X_i - u$, for all X_i in the set of exceedances. The GPD is a model for these excess amounts, not the raw values. This approach is more data-efficient than the block maxima method. Its primary challenge lies in selecting an appropriate threshold, which involves a trade-off between bias (if the threshold is too low) and variance (if the threshold is too high, yielding too few data points). In practice, exceedances may occur in clusters (e.g., during a multi-day storm event). Therefore, a declustering algorithm is often applied to the raw exceedances to ensure that the events being modeled are approximately independent.

6.3 Asymptotic theory for extremes

The statistical justification for both the block maxima and POT approaches comes from asymptotic theorems that describe the limiting distributions of extreme values.

6.3.1 Theory for block maxima: The GEV distribution

The theoretical justification for the block maxima approach is the **Extremal Types Theorem**. This theorem states that for a large class of underlying distributions, if a stable, non-degenerate limiting distribution for the block maximum M_n exists, it must be the **Generalized Extreme Value (GEV)** distribution.

A critical point is that the theorem applies to a *linearly rescaled* or *normalized* maximum. Consider the raw maximum, M_n. As the block size n increases, the expected value of M_n will also increase (or stay the same)—its distribution drifts towards larger values and does not converge. A similar issue arises in the Central Limit Theorem (CLT), which describes the convergence of a sample mean. The CLT does not apply to the raw sum of random variables, but to a normalized version of it.

The Extremal Types Theorem is the direct analog of the CLT for maxima. It states that there exist sequences of normalizing constants, a location parameter b_n and a scale parameter $a_n > 0$, such that the distribution of the normalized maximum, $(M_n - b_n)/a_n$, converges to the GEV distribution as $n \to \infty$.

In practice, we do not need to know the specific functional forms of a_n and b_n. Instead, for a fixed, large block size n (e.g., one year of daily data), we fit the GEV distribution directly to the series of block maxima. The GEV's location and scale parameters, μ and σ, effectively absorb and account for the normalization that the theory requires. The GEV is a flexible three-parameter family with a cumulative distribution function (CDF) given by:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

This is defined on the set $\{z : 1 + \xi(z-\mu)/\sigma > 0\}$. The parameters are location (μ), scale ($\sigma > 0$), and shape (ξ). These parameters implicitly depend on the block size n.

6.3.2 Theory for threshold exceedances: The GPD

The corresponding theory for the POT approach is the **Pickands–Balkema–de Haan Theorem**. It states that for a wide range of distributions, the distribution of excesses over a high threshold u can be approximated by the **Generalized Pareto Distribution (GPD)**. Conceptually, if the GEV describes the behavior of the maximum of a large block, the GPD describes the behavior of the distribution's tail that produced that maximum. It is the distribution one would expect to see by "zooming in" on the tail of a distribution above a high threshold.

The GPD is a two-parameter family with a CDF for **excesses** $Y = X - u$ given by:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi}$$

This is defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma_u) > 0\}$. The parameters are the shape, ξ, and a scale parameter σ_u that depends on the threshold u.

If the parent distribution's maxima are GEV-distributed with parameters (μ, σ, ξ), then the GPD scale parameter is given by $\sigma_u = \sigma + \xi(u - \mu)$. This key result shows that the GPD scale parameter is a linear function of the threshold u, a property that is used in advanced diagnostics for threshold selection, and illustrates the links between the GPD and GEV models for extremes.

6.3.3 The shape parameter

The **shape parameter,** ξ, is the most critical parameter in extreme value analysis and has the same interpretation in both the GEV and GPD. It governs the tail behavior of the distribution.

- $\xi = 0$ (**Gumbel-type tail**): The tail decays exponentially (light tail).
- $\xi > 0$ (**Fréchet-type tail**): The tail decays as a polynomial (heavy tail), with no upper bound.
- $\xi < 0$ (**Weibull-type tail**): The distribution has a finite upper bound.

6.3.4 Connection between GEV and GPD

The GEV and GPD models are intrinsically linked. If the block maxima of a process follow a GEV distribution with parameters (μ, σ, ξ), then for a high threshold u, the threshold excesses follow a GPD with the same shape parameter ξ. This provides a theoretical consistency between the two primary modeling approaches.

6.4 Return periods and return levels

6.4.1 Definitions and Calculation

The output of an extreme value analysis is typically expressed in terms of **return periods** and **return levels**. The **N-year return level**, z_N, is the level expected to be exceeded on average once every N years. It corresponds to the quantile of the distribution with an annual exceedance probability of $p = 1/N$.

For the GEV distribution, it is calculated by inverting the CDF:

$$z_N = \mu - \frac{\sigma}{\xi}\left[1 - (-\ln(1-p))^{-\xi}\right]$$

Calculating return levels from a GPD model requires an additional step. If ζ_u is the rate of threshold exceedances (e.g., the average number of exceedances per year), the N-year return level is the value z_N that is exceeded with probability $1/(N\zeta_u)$. It is calculated by adding the threshold back to the corresponding quantile of the GPD:

$$z_N = u + \frac{\sigma_u}{\xi}\left[(N\zeta_u)^{\xi} - 1\right]$$

6.4.2 Return level plots

A standard diagnostic and visualization tool is the return level plot. This plot graphs the estimated return level z_N against the return period N, with N typically plotted on a logarithmic scale. The curvature of the fitted line is a direct visualization of the shape parameter, ξ: a straight line implies $\xi = 0$, a concave curve implies $\xi > 0$, and a convex curve implies $\xi < 0$.

[Placeholder for a Return Level Plot showing a fitted GEV or GPD curve with confidence intervals.]

6.5 Inference

6.5.1 Plotting Positions

To visually assess model fit, the fitted model is plotted against the observed maxima. For this purpose, we require a method to assign a non-exceedance probability (and thus a return period) to each observed maximum. For a sample of n block maxima, let $z_{(1)} < z_{(2)} < \cdots < z_{(n)}$ be the ordered values. We estimate the probability $P(Z \leq z_{(k)})$ using a plotting position formula. These formulas generally take the form:

$$p_k = \frac{k - a}{n + 1 - 2a}$$

where k is the rank of the observation and a is a parameter. Common choices include:

- **Weibull:** $a = 0$, giving $p_k = k/(n+1)$.
- **Gringorten:** $a = 0.44$, giving $p_k = (k - 0.44)/(n + 0.12)$. This is often recommended as an unbiased choice for Gumbel-type distributions.

The empirical return period for the k-th observation is then estimated as $1/(1 - p_k)$.

6.5.2 Moments

6.5.3 MLE

6.5.4 Bayesian

6.6 Sampling variability

6.7 Regionalization

6.8 Nonstationarity

Further reading

- (Coles 2001): Canonical extreme value textbook with mathematical rigor and practical examples

7 Downscaling and Bias Correction

Learning objectives

After reading this chapter, you should be able to:

- Distinguish between supervised and distributional downscaling approaches
- Understand the motivation for downscaling climate model outputs
- Apply bias correction and quantile-quantile mapping techniques
- Recognize the stationarity assumption and its implications
- Evaluate different downscaling methods for specific applications
- Understand modern machine learning approaches to climate downscaling

Further reading

- Lanzante et al. (2018) for comprehensive review of downscaling challenges
- Lafferty and Sriver (2023)
- Farnham, Doss-Gollin, and Lall (2018)

8 Stochastic weather generators

Learning objectives

After reading this chapter, you should be able to:

- Understand the principles of stochastic weather generation
- Apply statistical models for synthetic weather data
- Use weather generators for downscaling climate model output
- Evaluate the performance of weather generation models

8.1 Essential concepts

1. Hidden Markov models for weather state modeling
2. Multi-site and multi-variable generation
3. Statistical downscaling applications
4. Synthetic data validation and evaluation
5. Integration with physical models

Further reading

9 Physics-based models and calibration

Learning objectives

After reading this chapter, you should be able to:

- Navigate trade-offs between model complexity, interpretability, and computational cost
- Characterize and communicate within- and between-model uncertainty
- Use surrogate models to approximate complex model output
- Apply model calibration techniques for climate applications

9.1 Essential concepts

1. Physics-based vs data-driven modeling spectrum
2. Model chaining and uncertainty propagation
3. Calibration methods and parameter estimation
4. Surrogate modeling for computational efficiency
5. Model structure uncertainty quantification

Further reading

For physics-based modeling in climate applications:

- Rackauckas et al. (2020): Scientific machine learning for differential equations

10 Optimal sampling methods

Learning objectives

After reading this chapter, you should be able to:

- Apply sampling techniques to generate synthetic event sets
- Use importance and stratified sampling to improve efficiency in hazard modeling
- Evaluate how sampling choices affect estimates of extreme risk
- Balance computational cost vs accuracy in climate risk estimation

10.1 Essential sampling concepts

1. Monte Carlo sampling and variance reduction
2. Importance sampling for rare events
3. Stratified sampling strategies
4. Synthetic event generation methods
5. Computational efficiency optimization

Further reading

11 Global sensitivity analysis

Learning objectives

After reading this chapter, you should be able to:

- Understand the role of sensitivity analysis in climate risk modeling
- Apply variance-based sensitivity methods (Sobol indices) to identify key parameters
- Use Morris screening methods for initial parameter importance ranking
- Interpret sensitivity analysis results for model simplification and uncertainty reduction

11.1 Essential sensitivity analysis concepts

1. Global vs local sensitivity analysis approaches
2. Sobol indices and variance decomposition
3. Morris elementary effects for screening
4. Multi-model sensitivity analysis
5. Implementation strategies for complex model chains

Further reading

For global sensitivity analysis:

- Saltelli et al. (2008): Comprehensive introduction to GSA methods
- Herman and Usher (2017): Practical implementation guide with software tools

Part IV

III: Risk Management

12 Exposure and Vulnerability

12.1 Learning objectives

By the end of this chapter, you should be able to:

- Define exposure and vulnerability in the context of climate risk assessment
- Distinguish between different types of vulnerability (physical, social, economic)
- Understand methods for quantifying and mapping exposure
- Apply vulnerability assessment frameworks to real-world scenarios
- Integrate exposure and vulnerability data with hazard information for risk assessment

12.2 Further reading

13 Cost-Benefit Analysis and Net Present Value

Learning objectives

After reading this chapter, you should be able to:

- Understand the theoretical foundation of cost-benefit analysis and Bayesian decision theory
- Apply net present value calculations with appropriate discount rates
- Quantify costs and benefits using utility functions for climate risk decisions
- Handle uncertainty in cost-benefit frameworks using expected value
- Recognize the limitations and appropriate applications of cost-benefit analysis
- Evaluate economic trade-offs over different time horizons and scenarios

Further reading

14 Policy Search & Optimization

15 Risk Transfer

See first

This chapter builds on concepts from: - Exposure and Vulnerability - Expectations and Discounting

Learning objectives

- Explore financial instruments (insurance, reinsurance, catastrophe bonds) for spreading or transferring climate risk.
- Understand parametric insurance triggers and how they differ from indemnity-based approaches.
- Assess the role of public-private partnerships in risk transfer mechanisms.

15.1 Insurance/reinsurance fundamentals

15.2 Parametric vs. indemnity coverage

15.3 Catastrophe bonds, index-based schemes

15.4 Challenges in emerging markets and vulnerable regions

Further reading

16 Deep Uncertainty and Model Structure

See first

This chapter builds on concepts from: - Probability and Statistics - Model Validation and Comparison

Learning objectives

- Distinguish between aleatory and epistemic uncertainty in climate risk assessment
- Understand the challenges posed by structural uncertainty and model disagreement
- Apply Bayesian Model Averaging (BMA) and stacking approaches to combine multiple models
- Recognize when deep uncertainty invalidates traditional decision frameworks
- Identify sources of deep uncertainty in exposure and impact modeling

Further reading

17 Robustness

18 Adaptive Planning and Flexibility

See first

This chapter builds on concepts from: - Deep Uncertainty - Robustness

Learning objectives

- Plan for uncertainty with adaptive management and iterative risk strategies.
- Develop adaptation pathways that evolve with new information (e.g., climate data, impacts).
- Incorporate monitoring and feedback loops into long-term climate policy.

Further reading

19 Working with People: Values, Participation, and Communication

Learning objectives

- Understand and communicate how decision analyses are necessarily influenced by subjective value judgements
- Explain common ethical frameworks for decision analysis
- Be familiar with multiple frameworks for stakeholder engagement and participatory methods
- Design effective approaches for communicating uncertainty in climate risk assessments
- Recognize the social and political dimensions of climate risk management

Further reading

Part V

Computational Case Studies

Overview

This collection of computational notebooks demonstrates the methods and concepts discussed in the main text through practical applications. Each notebook is designed to be standalone and self-contained, using the Julia programming language for all computations.

Understanding probability distributions

This notebook demonstrates the fundamental concepts underlying probability distributions. Understanding these relationships forms the foundation for statistical inference and uncertainty quantification in climate risk assessment.

We explore three essential functions that describe random variables: probability density functions (PDFs) or probability mass functions (PMFs), cumulative distribution functions (CDFs), and quantile functions. These concepts apply whether we're modeling temperature variability, extreme precipitation events, or flood frequencies.

Distribution functions and their relationships

Every probability distribution can be characterized by three related functions. Understanding their relationships helps build intuition for how probability models describe uncertainty.

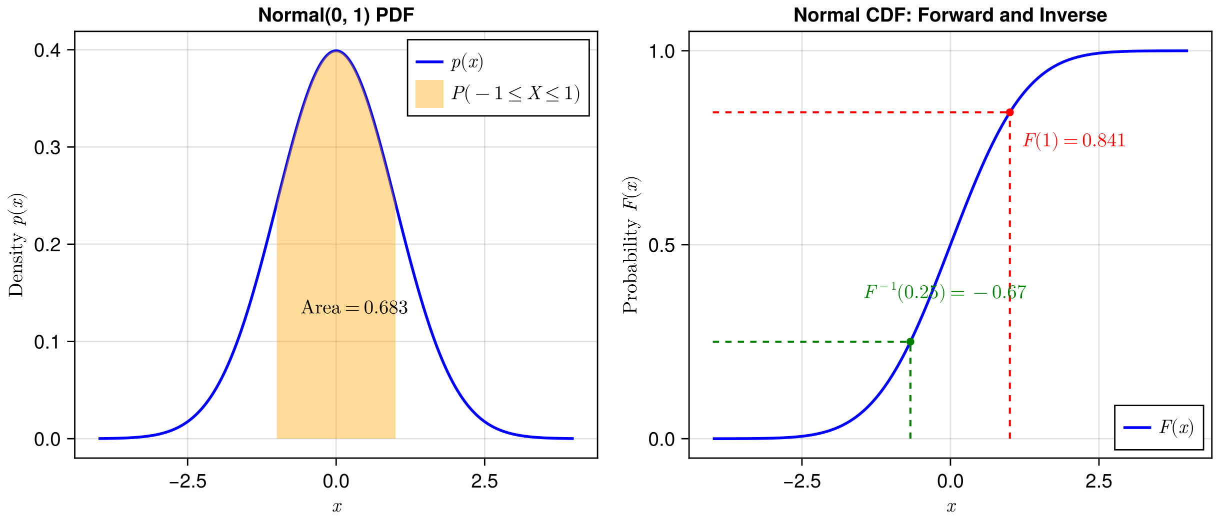
Helper functions for visualization

We start by creating reusable functions for common visualization tasks. This approach keeps our main examples clean while demonstrating good programming practices.

These helper functions encapsulate common visualization patterns. The `add_pdf_area!` function demonstrates how probabilities correspond to areas under density curves. The forward and inverse CDF functions show the relationship between values and cumulative probabilities.

Normal distribution example

The normal distribution illustrates these concepts for continuous random variables. Its smooth curves and well-known properties make it ideal for understanding probability fundamentals.



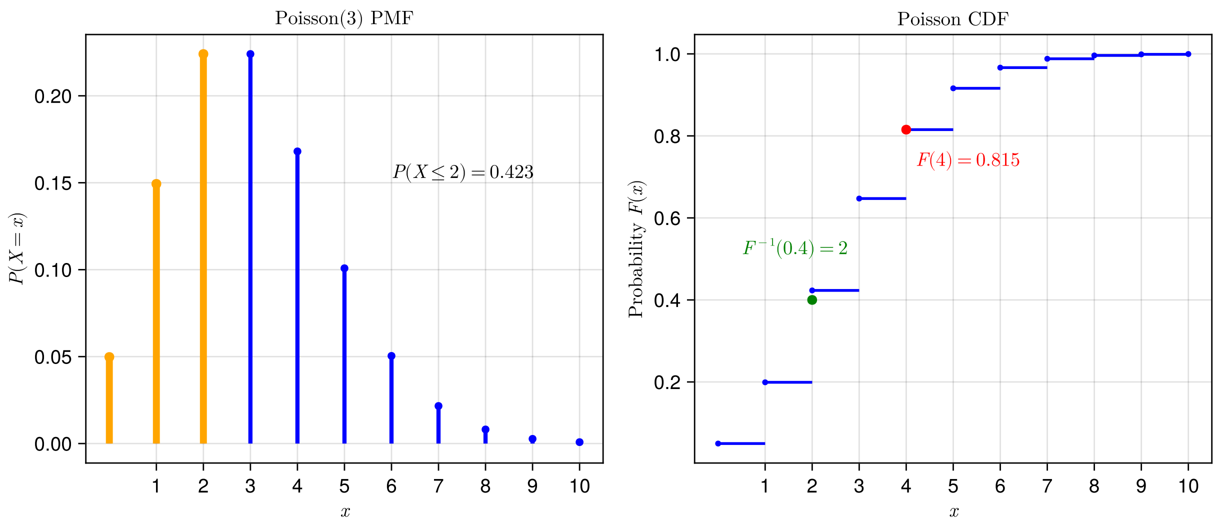
The normal distribution example shows how probability density relates to cumulative probability. The left panel demonstrates that probabilities correspond to areas under the density curve. The right panel shows the CDF's S-shaped curve and illustrates both forward operations (finding probabilities from values) and inverse operations (finding values from probabilities).

These operations are fundamental to risk assessment: forward operations answer "what's the probability of exceeding this threshold?" while inverse operations answer "what value corresponds to this probability?"

Discrete distributions: Poisson example

Discrete distributions illustrate the same concepts but with point masses rather than continuous densities. The Poisson distribution commonly models count data like the number of extreme events per year.

These helper functions handle the specific visualization needs of discrete distributions. Unlike continuous distributions, discrete probabilities are point masses, and CDFs are step functions.

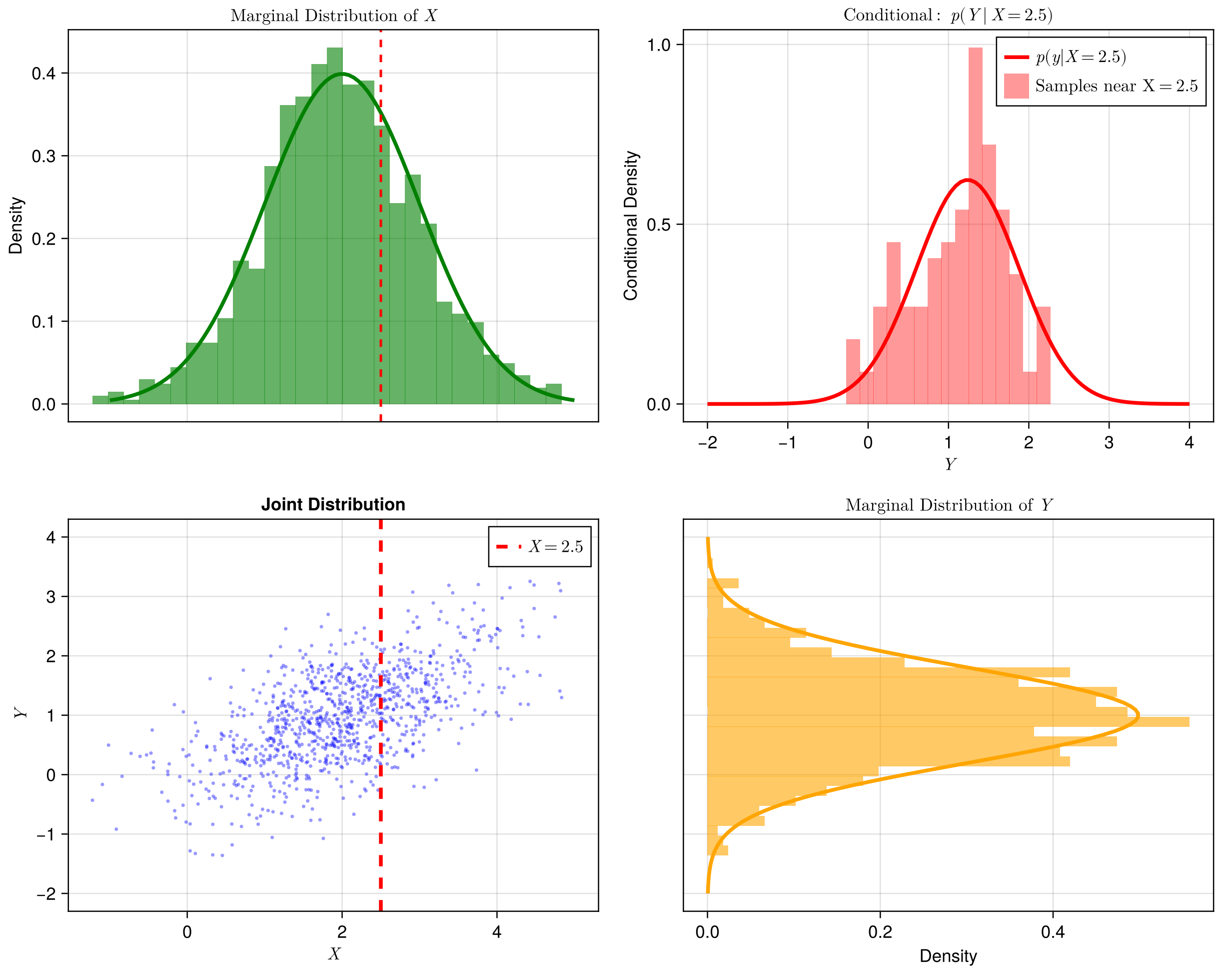


The Poisson distribution demonstrates these same fundamental concepts for discrete random variables. Individual probabilities are represented as point masses rather than areas under curves. The CDF becomes a step function that jumps at each possible value.

This distribution often appears in climate applications when modeling rare events like the annual number of hurricanes making landfall or the count of days exceeding extreme temperature thresholds.

Multiple variables and dependence

Real systems involve multiple interconnected variables. Understanding joint, marginal, and conditional distributions enables modeling of complex dependencies.



This multivariate example demonstrates how joint distributions decompose into marginal and conditional components. The joint distribution (bottom left) shows the full relationship between variables. Marginal distributions (top and right panels) show each variable's behavior independently. The conditional distribution (top right) shows how one variable behaves given specific values of another.

These concepts are essential for climate modeling where variables like temperature and precipitation are correlated. Understanding their joint behavior enables more accurate risk assessment than treating them independently.

Key insights and climate applications

The examples in this notebook illustrate fundamental principles that apply across all probability distributions:

Distribution functions work together: PDFs/PMFs, CDFs, and quantile functions provide complementary views of the same underlying uncertainty.

Discrete and continuous cases follow similar logic: The mathematical relationships remain consistent whether dealing with counts or continuous measurements.

Multiple variables require joint modeling: Real climate systems involve correlated variables that must be modeled together for accurate risk assessment.

In climate applications, these concepts appear when: - Modeling temperature distributions to assess heat wave probabilities - Analyzing extreme precipitation using heavy-tailed distributions - Understanding joint temperature-humidity relationships for heat stress assessment - Characterizing the frequency of compound events like concurrent drought and heat

The computational tools demonstrated here provide the foundation for more complex statistical inference methods covered in subsequent notebooks.

Inference Example: Flipping a Coin

This notebook demonstrates the fundamental principles of statistical inference using the pedagogically clean example of coin flipping. While simple, this example illuminates the core concepts that underlie all statistical analysis: likelihood functions, maximum likelihood estimation, and Bayesian inference.

The same principles apply whether we're estimating the probability of a fair coin or the frequency of extreme climate events. Understanding these methods enables principled learning from data under uncertainty.

Problem

Consider flipping a coin of unknown success rate θ (the probability of heads) multiple times. After observing the outcomes, we want to estimate θ. This mirrors many climate problems where we observe limited data and seek to understand underlying probabilities.

Generating example data

We start with a known scenario to validate our inference methods.

This setup allows us to evaluate how well different inference methods recover the true parameter value. In real applications, we never know the true value and must rely on the methods demonstrated here.

Likelihood-based inference

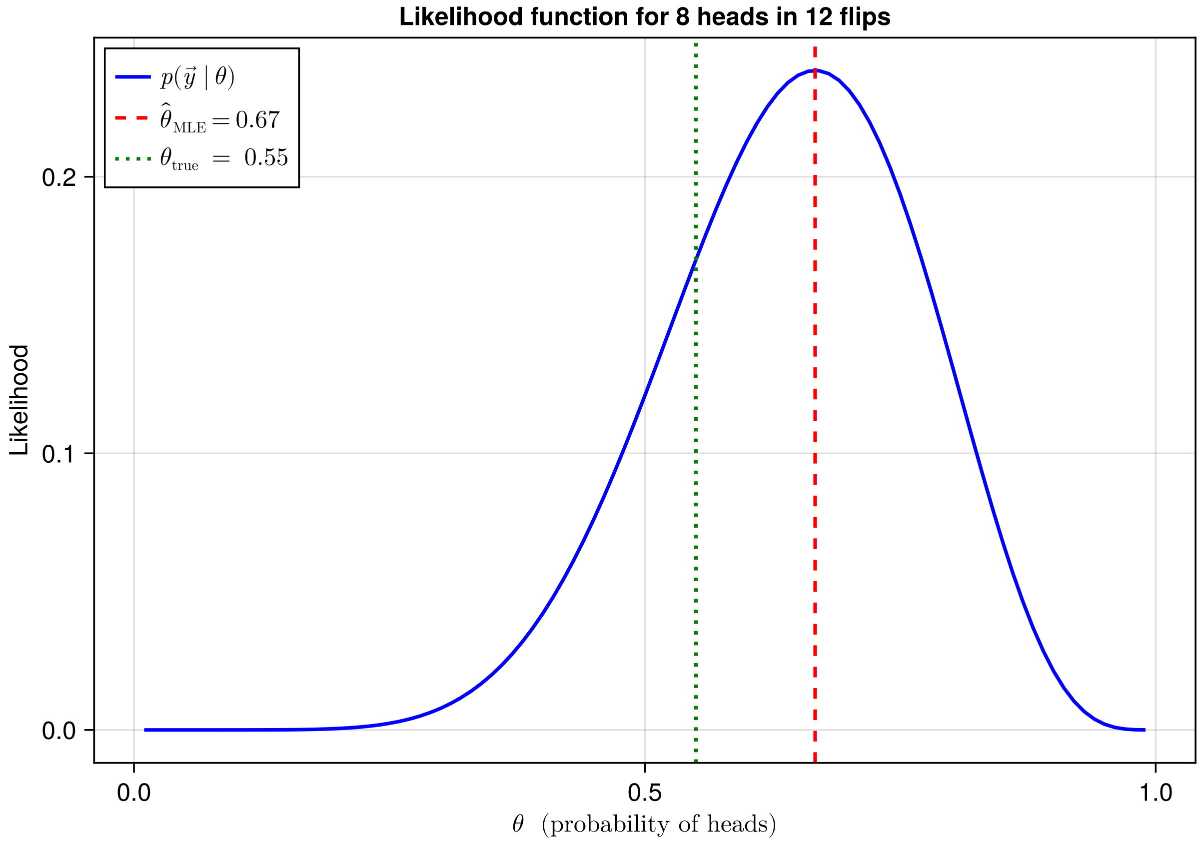
The likelihood function bridges observed data and unknown parameters. It quantifies how probable our observed data would be under different parameter values.

Understanding the likelihood function

For coin flips, the likelihood follows the binomial distribution:

$$p(\text{data} \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

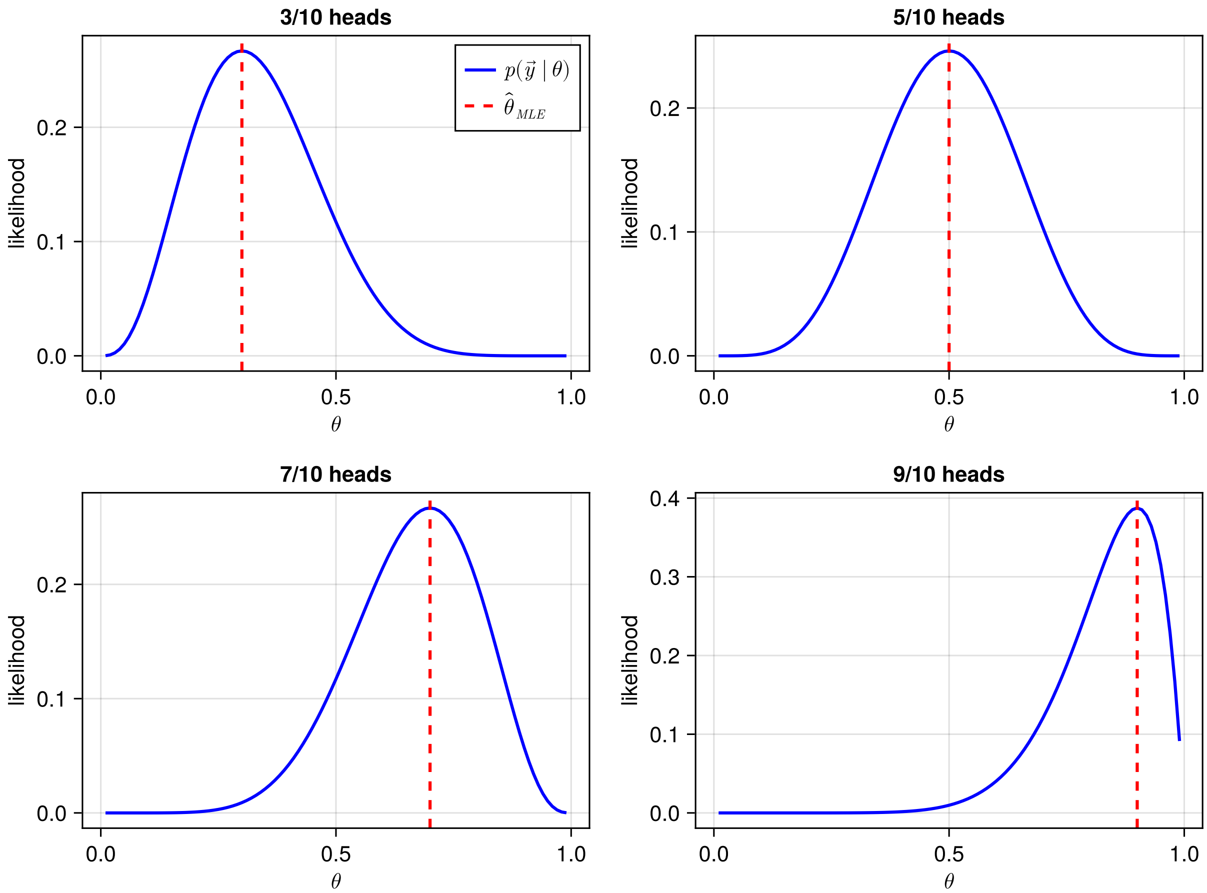
where k is the number of observed heads and n is the total number of flips.



The likelihood function shows how plausible different parameter values are given our observed data. It peaks at the observed proportion, which becomes our maximum likelihood estimate. Notice how the curve's width reflects uncertainty—more data would create a sharper peak.

Maximum likelihood estimation

Maximum likelihood estimation (MLE) finds the parameter value that makes the observed data most probable. For the binomial model, this has a simple analytical solution.



Each panel shows how the likelihood function changes with different observed outcomes. The maximum always occurs at the observed proportion, confirming that MLE equals the sample proportion for binomial data.

This relationship holds broadly: MLE provides intuitive estimates that often match simple summary statistics for well-behaved problems.

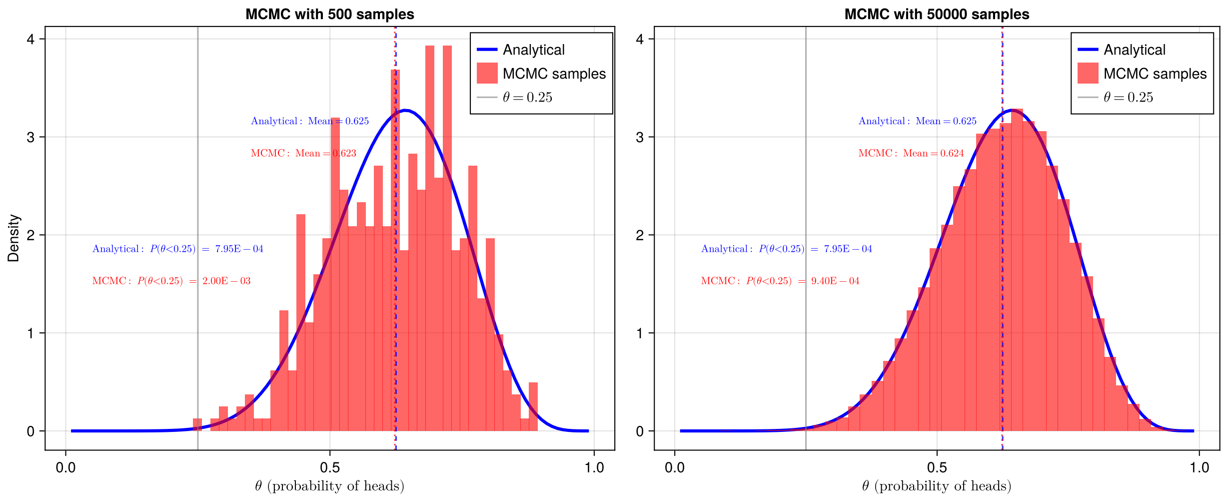
Bayesian inference

First we define our probabilistic model in Turing.

```
coin_flip_model (generic function with 2 methods)
```

Then, we compare the analytical posterior (which is available due to conjugacy) to the posterior obtained via MCMC sampling.

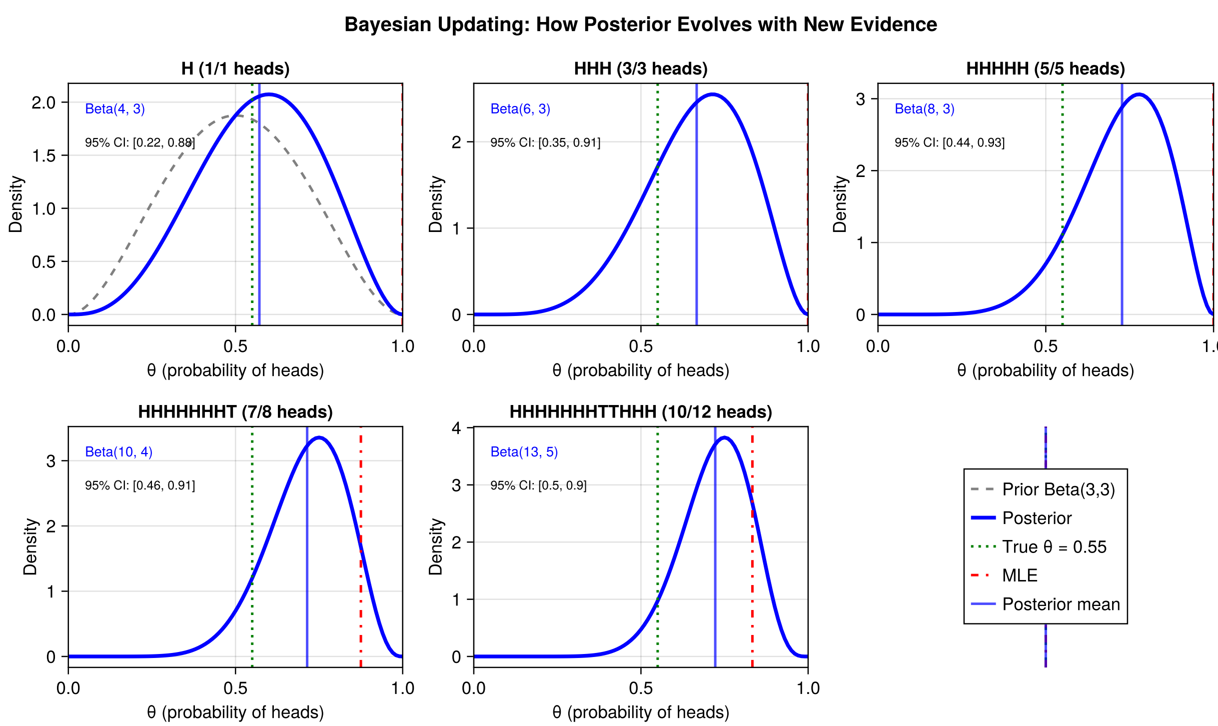
We can then compare how well we can sample the analytical distribution with different numbers of samples.



We can see that with a small number of samples, we do a reasonable job of approximating the mean, but the tails are poorly estimated. This is a general principle in Monte Carlo methods: estimating tail probabilities requires many more samples than estimating central tendencies like the mean or median. Sometimes, specialized techniques like importance sampling are needed to accurately capture tail behavior.

Sequential Bayesian updating

The following example demonstrates how posterior distributions evolve as we observe more coin flips:



Linear regression: three perspectives on the same problem

This notebook demonstrates how the same statistical problem can be approached from three different perspectives: curve fitting, maximum likelihood estimation, and Bayesian inference. Linear regression provides an ideal example because it bridges simple parameter estimation with multivariate modeling while maintaining analytical tractability.

Understanding these different approaches illuminates the philosophical foundations underlying statistical methods and their practical implications for uncertainty quantification.

The regression problem

Linear regression models the relationship between a predictor variable x and a response variable y as:

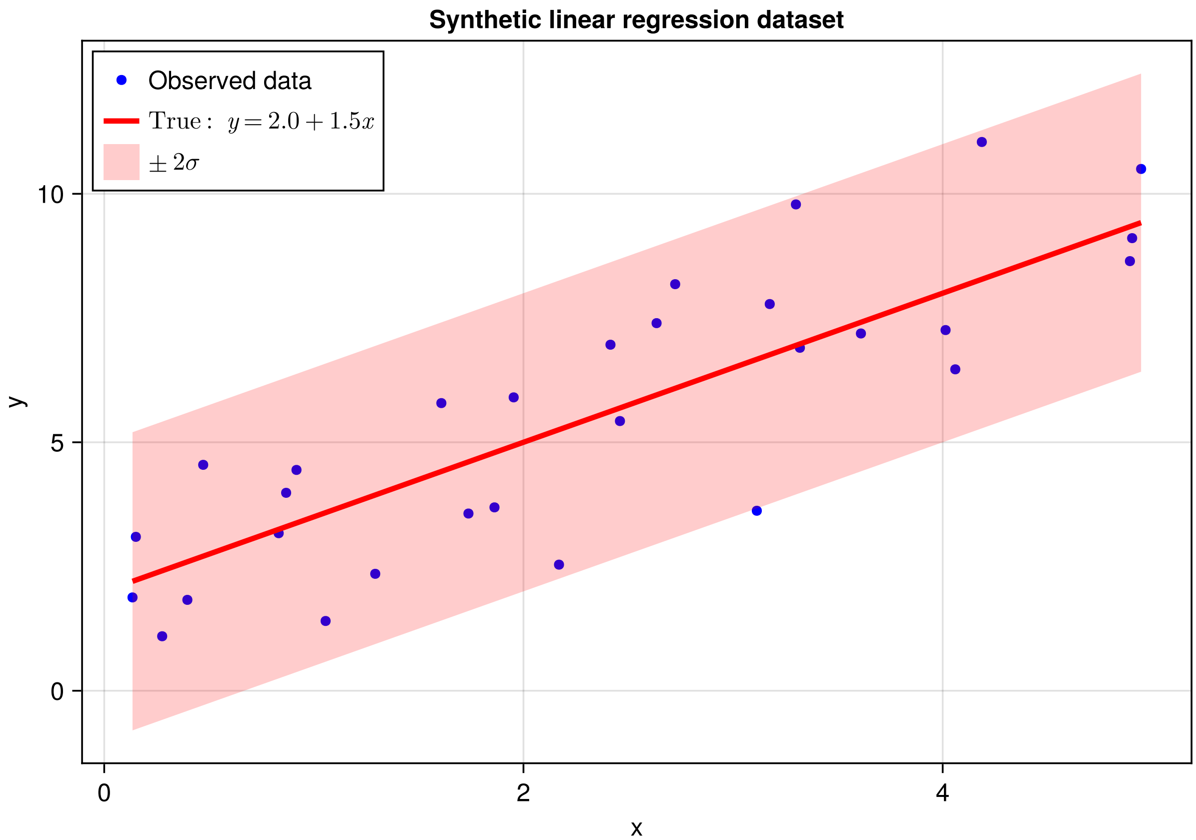
$$y_i = \alpha + \beta x_i + \epsilon_i$$

where α is the intercept, β is the slope, and ϵ_i represents random noise.

This framework applies broadly in climate science: relating temperature to elevation, precipitation to atmospheric pressure, or sea level rise to global temperature.

Generating synthetic data

We create synthetic data from a known linear relationship to evaluate how well different methods recover the true parameters. First, we generate and plot our synthetic data.



The plot shows our synthetic data with the true linear relationship and noise level. The scattered points represent our "observations," while the red line and shaded band show the true underlying process we're trying to recover.

Approach 1: Curve fitting (least squares)

The curve fitting perspective treats regression as an optimization problem: find the line that minimizes prediction errors. This approach focuses on algorithmic solutions without probabilistic interpretation.

Mathematical foundation

Least squares minimizes the sum of squared residuals:

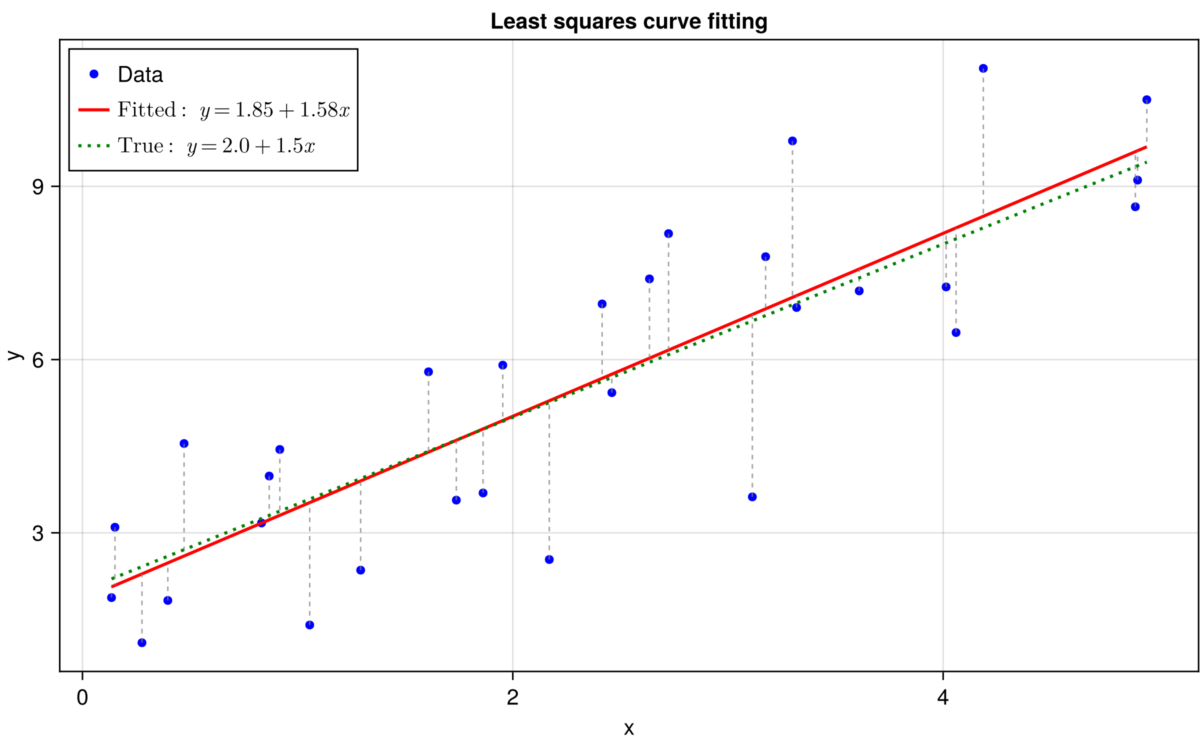
$$\text{minimize} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

This optimization problem has a closed-form analytical solution.

	parameter	estimate	true_value
	String	Float64	Float64?
1	Intercept	1.85082	2.0
2	Slope	1.58365	1.5
3	MSE	2.22587	*missing*
4	R²	0.711843	*missing*

The least squares estimates are close to the true values, demonstrating that the method works well for this problem. However, this approach provides no direct way to quantify uncertainty in the parameter estimates.

Visualization with residuals



The residual lines (gray dashes) show prediction errors for each data point. Least squares minimizes the sum of squared lengths of these lines, providing the “best fit” in this specific sense.

Approach 2: Maximum likelihood estimation

The probabilistic perspective treats regression as a statistical model with explicitly specified error distributions. This enables likelihood-based inference and uncertainty quantification.

Statistical model specification

We assume:

$$y_i \sim \text{Normal}(\alpha + \beta x_i, \sigma^2)$$

This specifies that responses are normally distributed around the linear predictor with constant variance.

	Parameter	MLE	OLS	True_Value
	String	Float64	Float64	Float64
1	Intercept	1.85082	1.85082	2.0
2	Slope	1.58365	1.58365	1.5
3	Sigma	1.49193	1.49193	1.5

Turing's `optimize` function with `MLE()` provides a clean, robust approach to maximum likelihood estimation without requiring custom optimization code. The MLE estimates for intercept and slope match the OLS results exactly, confirming the theoretical equivalence between these approaches for linear regression with normal errors. Additionally, MLE provides an estimate of the noise parameter σ.

Approach 3: Bayesian inference

Bayesian inference treats all parameters as random variables with probability distributions. It combines prior beliefs with observed data to produce posterior distributions that fully characterize parameter uncertainty.

Bayesian inference implementation

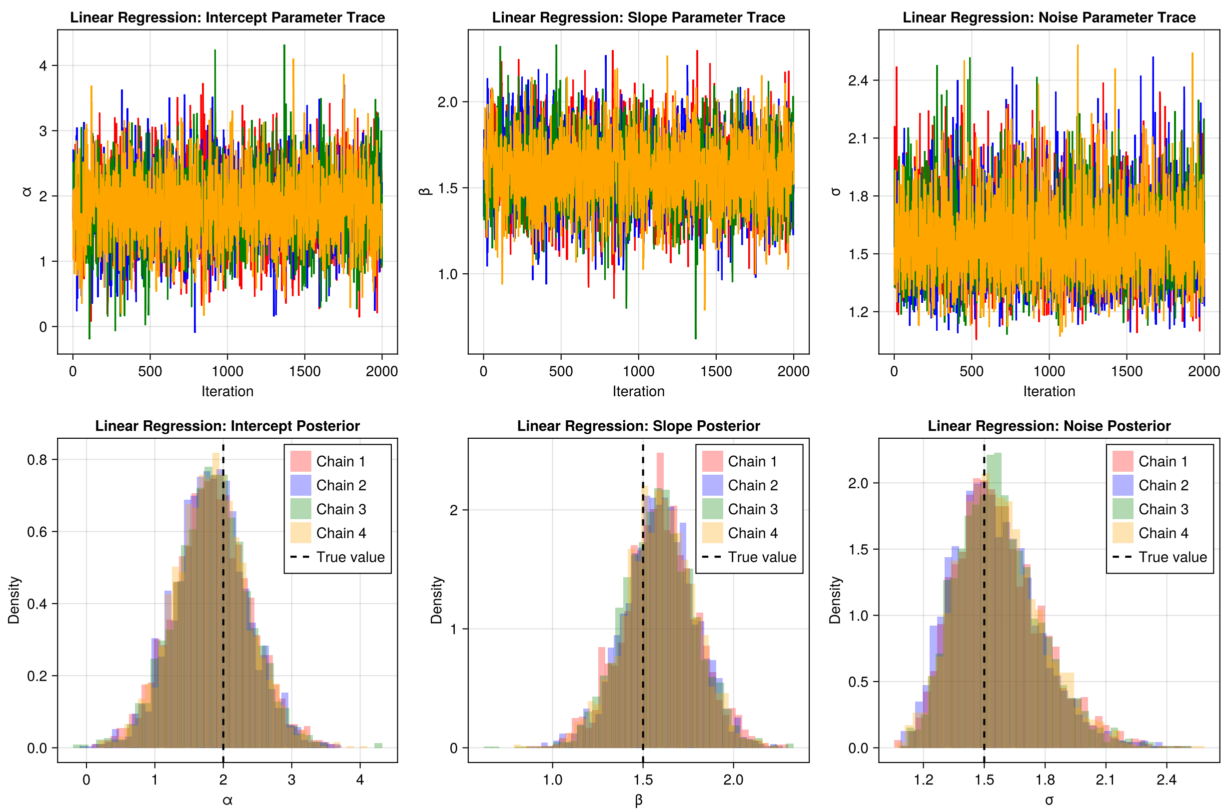
```
Warning: Only a single thread available: MCMC chains are not sampled in parallel
@ AbstractMCMC ~/.julia/packages/AbstractMCMC/FSyVk/src/sample.jl:382
Info: Found initial step size
    = 0.05
Info: Found initial step size
    = 0.2
Info: Found initial step size
    = 0.2
Info: Found initial step size
    = 0.025
```

```
"MCMCChains.Chains{Float64, AxisArrays.AxisArray{Float64, 3, Array{Float64, 3}, Tuple{AxisArray
```

The diagnostic table is consistent with, *though does not guarantee*, MCMC convergence and accurate sampling. Fundamentally, we can check for comon signs of non-convergence and rule them out, but we cannot actually prove that we are sampling from the true posterior distribution. While there are no perfect rules, generally

- Low values of $\hat{R}$ (`rhat`) close to 1 (e.g., <1.1) are consistent with "good mixing", meaning that the chains have converged to the target distribution.
- High effective sample sizes (`ess_bulk`, `ess_tail`) indicate that the samples are more approximately independent, which is desirable for accurate posterior estimation.

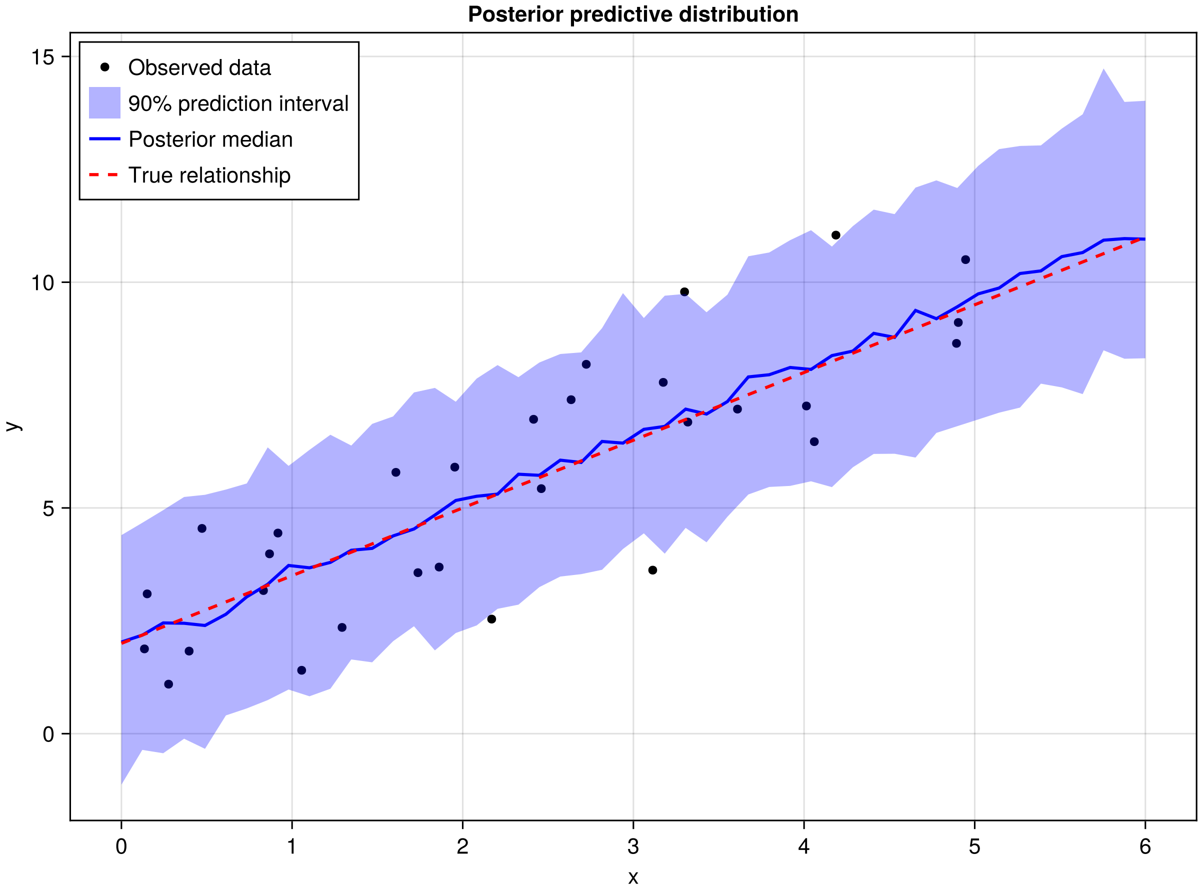
Trace plot



We can see from this plot that the chains visually appear to be **well-mixed** and **stationary**, which again is consistent with, though not definitive proof of, convergence.

Posterior predictive distribution

The posterior predictive distribution represents our beliefs about new data given what we've observed. It's computed as $p(y_{\text{new}}|y_{\text{obs}}) = \int p(y_{\text{new}}|\theta)p(\theta|y_{\text{obs}})d\theta$.



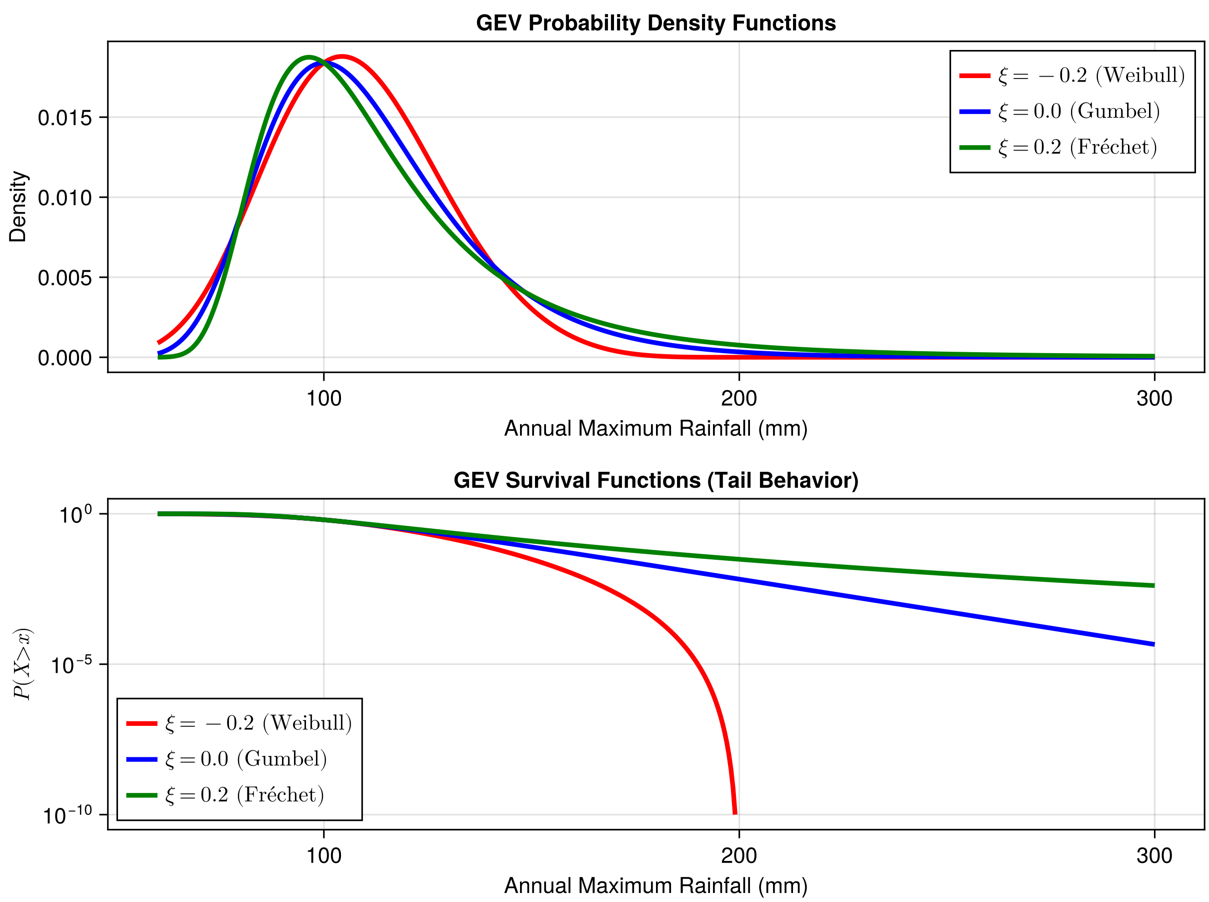
The posterior predictive distribution incorporates both parameter uncertainty and observation noise. The 90% prediction interval shows where we expect new observations to fall, while the median line shows our best prediction. This is the fundamental quantity for forecasting and decision-making under uncertainty.

Extreme Value Theory Examples

This notebook demonstrates extreme value analysis concepts using real Houston precipitation data, illustrating the fundamental extrapolation problem in climate risk assessment.

Shape parameter implications

The shape parameter in extreme value distributions fundamentally determines how heavy the tail is, which directly affects our extrapolation to rare events. Let's visualize how different shape parameter values lead to dramatically different tail behavior:



The key insight: **the shape parameter ξ controls how rapidly probabilities decrease in the tail**.

Houston precipitation data

Let's look at this using a realistic example

```
Loaded 70366 raw GHCN daily observations
First few rows:
```

	USC00414313	19430101	PRCP	0	Column5	Column6	6	Column8
	String15	Int64	String7	Int64	String1?	Missing	String1	Int64?
1	USC00414313	19430102	PRCP	0	*missing*	*missing*	6	*missing*
2	USC00414313	19430103	PRCP	0	*missing*	*missing*	6	*missing*
3	USC00414313	19430104	PRCP	0	*missing*	*missing*	6	*missing*

This loads the raw GHCN (Global Historical Climatology Network) data file. Each row contains weather measurements with a specific format that we need to decode.

Processing GHCN format

GHCN daily files use a fixed format where each row represents one measurement at one station on one date:

```
After filtering for precipitation (PRCP):
- Total PRCP observations: 25225
- Date range (raw): 19430102 to 20131130
```

GHCN includes many weather variables (temperature, snow, etc.), but we only need precipitation (PRCP). The date is stored as YYYYMMDD format and values are in tenths of millimeters.

Converting dates and units

```
After format conversion:
- Daily observations: 25225
- Date range: 1943-01-02 to 2013-11-30
- Years of data: 71
- Max daily rainfall: 272.3 mm
```

Now we have properly formatted dates and rainfall measurements in standard millimeter units. This represents over 70 years of daily precipitation measurements - our "high-frequency observations."

Filling missing dates for complete time series

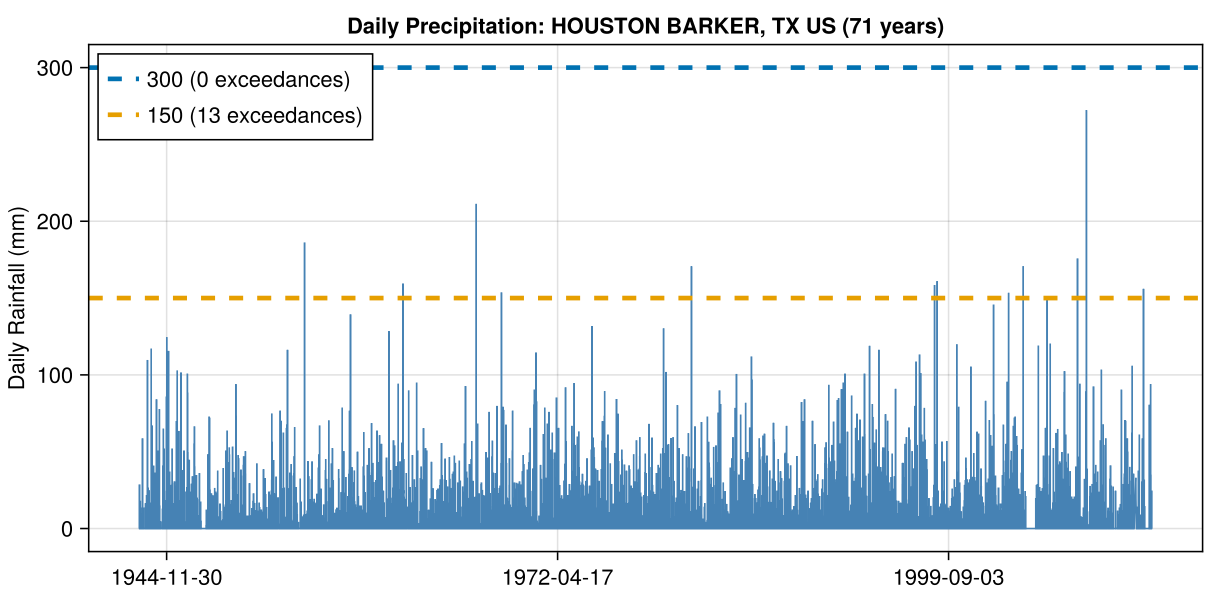
GHCN data only includes dates where measurements were recorded. For proper time series analysis and plotting, we need to fill in missing dates:

```
Date completeness:
- GHCN observations: 25225 days
- Complete date range: 25901 days
- Missing days: 676
- Complete time series: 25901 days
- Missing rainfall values: 676
```

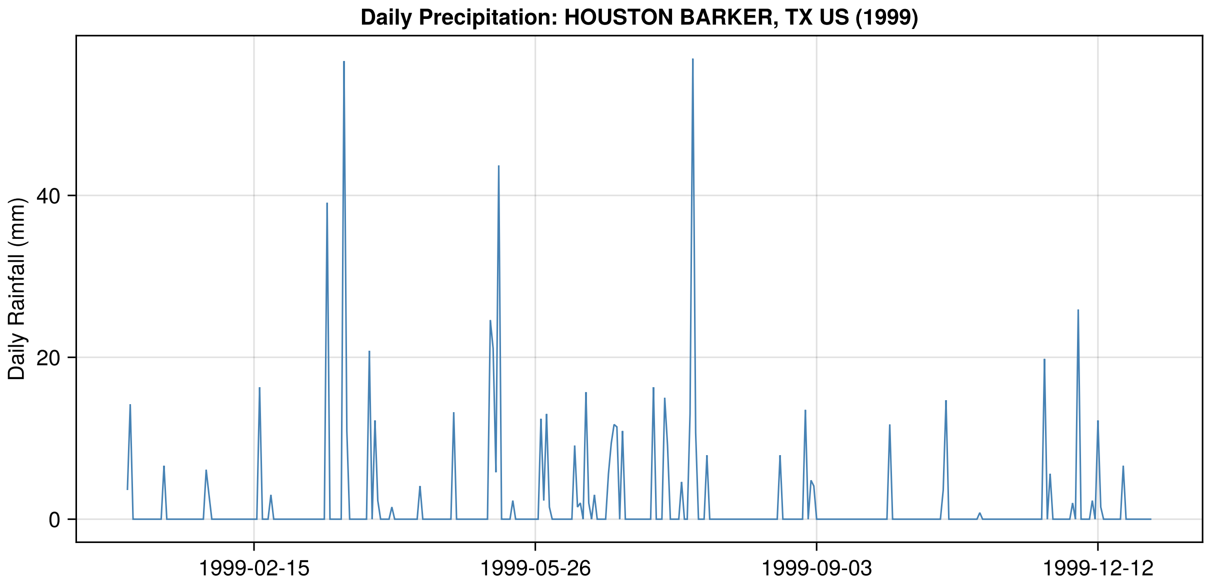
This creates a complete daily time series from $(minimum(precip_data.date)) to $(maximum(precip_data.date)) with `missing` values for days without measurements. Missing values are common in historical weather data due to equipment issues, observer absence, or data quality problems.

Daily precipitation time series

Now let's visualize the daily precipitation data to see the "high-frequency observations" mentioned in our extrapolation problem:



Or a subset


Daily Precipitation: HOUSTON BARKER, TX US (1999)
Daily Rainfall (mm)
0
20
40
1999-02-15
1999-05-26
1999-09-03
1999-12-12

Mosquito bites and beer consumption: simulation-based inference

This notebook demonstrates computational approaches to hypothesis testing using a practical example: whether drinking beer affects mosquito bite frequency. The analysis illustrates how simulation can provide intuitive understanding of statistical concepts without relying on complex mathematical assumptions.

This example showcases the power of computational thinking in statistics, where complex theoretical concepts become accessible through direct simulation. The same principles apply to climate data analysis where traditional parametric assumptions may not hold.

The research question

We investigate a practical question with broader implications for experimental design and causal inference:

> Does drinking beer reduce the likelihood of being bitten by mosquitos?

This question exemplifies common challenges in environmental and health research where controlled experiments must account for multiple confounding factors. The statistical methods demonstrated here apply broadly to climate impact studies and policy evaluation.

Learning objectives

After working through this analysis, you should understand: - The logic underlying permutation testing and simulation-based inference - How computational methods can replace complex mathematical derivations - The interpretation of p-values through direct simulation - When simulation-based tests offer advantages over parametric approaches

Experimental data

The dataset comes from a controlled experiment where participants were randomly assigned to consume either beer or water, then exposed to mosquitos in a controlled environment. Researchers counted the number of mosquito bites received by each participant.

	Group	N_participants	Mean_bites	Std_bites	Min_bites	Max_bites
	String	Int64	Float64	Float64	Int64	Int64
1	Beer drinkers	25	23.6	4.1332	17	31
2	Water drinkers	18	19.2222	3.67112	12	24

The descriptive statistics reveal that beer drinkers received more mosquito bites on average. However, we need statistical analysis to determine whether this difference could reasonably be attributed to random variation.

Initial analysis: observed difference

The most direct approach compares the average number of bites between groups. This test statistic captures the core research question in a single number.

	Statistic	Value
	String	Float64
1	Beer group mean	23.6
2	Water group mean	19.2222
3	Difference (Beer - Water)	4.37778

Beer drinkers received approximately 5.1 more bites on average than water drinkers. But is this difference statistically meaningful, or could it arise from random variation alone?

The null hypothesis and permutation testing

The skeptic's position forms our null hypothesis: drinking beer has no effect on mosquito bite frequency. Under this assumption, the group assignment (beer vs. water) is irrelevant, and the observed difference arose by chance.

We can test this hypothesis through permutation testing, which simulates what would happen if the null hypothesis were true.

Two approaches to hypothesis testing

Parametric approach (t-test): - Assumes data follow specific distributions (normal) - Uses mathematical theory for p-value calculation - Fast but relies on assumptions that may not hold

Simulation approach (permutation test): - Makes minimal distributional assumptions - Uses computational power instead of mathematical theory - More intuitive and flexible

Implementing permutation testing

The logic is straightforward: if treatment assignment doesn't matter, we can randomly shuffle group labels and recalculate the difference.

```
Example permuted difference: 0.46
```

This function simulates one possible outcome under the null hypothesis. By repeating this process thousands of times, we can characterize the full distribution of differences we'd expect by chance alone.

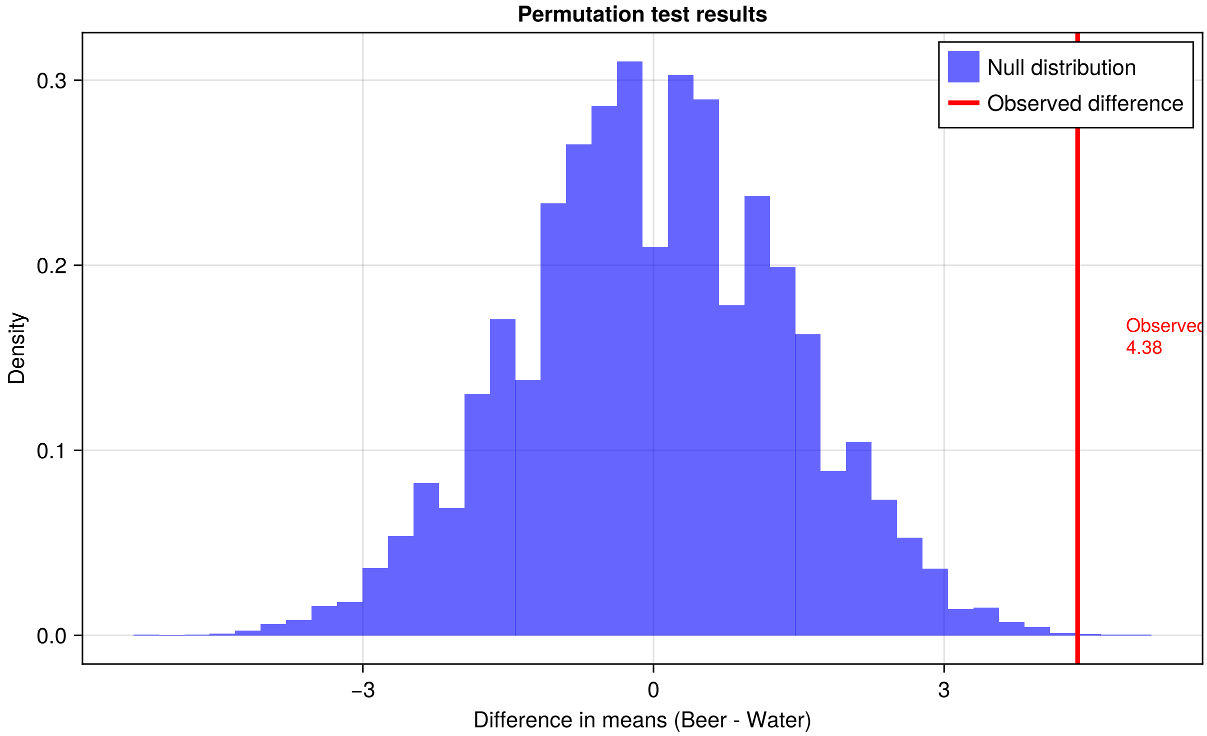
Generating the null distribution

	Statistic	Value
	String	Float64
1	Mean	0.00119
2	Std Dev	1.38007
3	Min	-5.36889
4	Max	5.14222
5	2.5th percentile	-2.69333
6	97.5th percentile	2.65778

The null distribution shows what group differences we'd expect if treatment assignment were random. Most permuted differences cluster around zero, as expected when there's no true effect.

Visualizing the results

The key insight comes from comparing our observed difference to the null distribution. How extreme is our observed value relative to what we'd expect by chance?



The visualization clearly shows that our observed difference lies in the extreme tail of the null distribution. This suggests that the observed difference is unlikely to have occurred by chance alone.

Statistical significance assessment

	Measure	Value
	String	String
1	Observed difference	4.38 bites
2	One-sided p-value	0.0003
3	Two-sided p-value	0.0008
4	Effect size (Cohen's d)	1.11
5	Interpretation	Statistically significant

Comparison with parametric testing

For completeness, we can compare our simulation-based results with traditional parametric tests.

	Method	P_value	Test_statistic
	String	Float64	String
1	Permutation test	0.00082	Difference in means
2	Equal variance t-test	0.000883127	t = 3.587
3	Unequal variance t-test	0.000747402	t = 3.658

All three methods reach similar conclusions, but the permutation test requires fewer assumptions about the underlying data distribution.

Key insights and broader applications

This analysis demonstrates several important statistical concepts:

Simulation-based inference: Complex statistical concepts become intuitive when approached through simulation rather than mathematical theory.

Assumption-free testing: Permutation tests work without requiring specific distributional assumptions, making them robust for real-world data.

P-value interpretation: The p-value directly represents the probability of observing such an extreme difference under the null hypothesis, clarified through simulation.

Effect size matters: Statistical significance alone doesn't indicate practical importance—effect size measures like Cohen's d provide additional context.

Climate science applications

These simulation-based methods prove particularly valuable in climate research where:

- **Non-normal data**: Climate variables often have skewed or heavy-tailed distributions
- **Small sample sizes**: Paleoclimate records or extreme event counts may have limited observations
- **Complex dependencies**: Traditional parametric assumptions may not hold for climate time series
- **Policy decisions**: Robust statistical inference supports high-stakes climate adaptation decisions

The computational approach demonstrated here scales naturally to more complex problems while maintaining the same intuitive logic: simulate what would happen under different assumptions and compare with observed data.

References

Abernathey, Ryan. 2024. *An Introduction to Earth and Environmental Data Science*. https://earth-env-data-science.github.io/intro.html.

Applegate, Patrick, and Klaus Keller. 2015. *Risk Analysis in the Earth Sciences*. Leanpub. https://leanpub.next/raes.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. "Generative AI Without Guardrails Can Harm Learning: Evidence from High School Mathematics." *Proceedings of the National Academy of Sciences* 122 (26): e2422633122. https://doi.org/10.1073/pnas.2422633122.

Bishop, Christopher M., and Hugh Bishop. 2024. *Deep Learning: Foundations and Concepts*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-45468-4.

Blitzstein, Joseph K., and Jessica Hwang. 2019. *Introduction to Probability, Second Edition*. 2nd Edition. Boca Raton: Chapman and Hall/CRC. http://probabilitybook.net.

Coles, Stuart. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer.

Cressie, Noel A. C., and Christopher K. Wikle. 2011. *Statistics for Spatio-Temporal Data*. Hoboken, N.J.: Wiley.

Downey, Allen B. 2021. *Think Bayes*. "O'Reilly Media, Inc.". https://allendowney.github.io/ThinkBayes2/.

Farnham, David J, James Doss-Gollin, and Upmanu Lall. 2018. "Regional Extreme Precipitation Events: Robust Inference from Credibly Simulated GCM Variables." *Water Resources Research* 54 (6). https://doi.org/10.1002/2017wr021318.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics Springer, Berlin.

Gelman, Andrew. 2021. *Regression and Other Stories*. Analytical Methods for Social Research. Cambridge, United Kingdom ; Cambridge University Press.

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. Chapman & Hall/CRC Boca Raton, FL, USA.

Ghil, M, P Yiou, S Hallegatte, B D Malamud, P Naveau, A Soloviev, P Friederichs, et al. 2011. "Extreme Events: Dynamics, Statistics and Prediction." *Nonlinear Processes in Geophysics* 18 (3): 295–350. https://doi.org/10/fvzxvv.

Helsel, Dennis R., Robert M. Hirsch, Karen R. Ryberg, Stacey A. Archfield, and Edward J. Gilroy. 2020. *Statistical Methods in Water Resources. Techniques and Methods*. U.S. Geological Survey. https://doi.org/10.3133/tm4A3.

Herman, Jon, and Will Usher. 2017. "SALib: An Open-Source Python Library for Sensitivity Analysis." *Journal of Open Source Software* 2 (9): 97. https://doi.org/10.21105/joss.00097.

Jaynes, Edwin T. 2003. *Probability Theory: The Logic of Science*. New York, NY: Cambridge University Press.

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. 2025. "Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task." June 10, 2025. https://doi.org/10.48550/arXiv.2506.08872.

Lafferty, David C., and Ryan L. Sriver. 2023. "Downscaling and Bias-Correction Contribute Considerable Uncertainty to Local Climate Projections in CMIP6." *Npj Climate and Atmospheric Science* 6 (1, 1): 1–13. https://doi.org/10.1038/s41612-023-00486-0.

Lanzante, John R, Keith W Dixon, Mary Jo Nath, Carolyn E Whitlock, and Dennis Adams-Smith. 2018. "Some Pitfalls in Statistical Downscaling of Future Climate." *Bulletin of the American Meteorological Society* 99 (4): 791–803. https://doi.org/10.1175/bams-d-17-0046.1.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Second edition. Texts in Statistical Science Series. Boca Raton ; CRC Press, Taylor & Francis Group.

Merz, Bruno, Jeroen C J H Aerts, Karsten Arnbjerg-Nielsen, M Baldi, A Becker, A Bichet, Günter Blöschl, et al. 2014. "Floods and Climate: Emerging Perspectives for Flood Risk Assessment and Management." *Natural Hazards and Earth System Science* 14 (7): 1921–42. https://doi.org/10/gb9nzm.

Mignan, Arnaud. 2024. *Introduction to Catastrophe Risk Modelling: A Physics-based Approach*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009437370.

Mudelsee, Manfred. 2020. "Statistical Analysis of Climate Extremes / Manfred Mudelsee." In *Statistical Analysis of Climate Extremes*. Cambridge, United Kingdom ; Cambridge University Press.

Naghettini, Mauro, ed. 2017. *Fundamentals of Statistical Hydrology*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-43561-9.

Piironen, Juho, and Aki Vehtari. 2017. "Comparison of Bayesian Predictive Methods for Model Selection." *Statistics and Computing* 27 (3): 711–35. https://doi.org/10.1007/s11222-016-9649-y.

Pyrcz, Michael J. 2024. *Applied Machine Learning in Python: A Hands-on Guide with Code*. https://geostatsguy.github.io/MachineLearningDemos_Book.

Rackauckas, Christopher, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. 2020. "Universal Differential Equations for Scientific Machine Learning." 2020. https://doi.org/10.48550/ARXIV.2001.04385.

Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470725184.ch1.

Thuerey, N., B. Holzschuh, P. Holl, G. Kohl, M. Lino, Q. Liu, P. Schnell, and F. Trost. 2024. *Physics-Based Deep Learning*. https://physicsbaseddeeplearning.org.

A Software Setup

If you want to run the computational notebooks in this book, or apply a similar workflow, then these instructions are for you.

A.1 Quick start

This section provides step-by-step instructions to get your development environment set up and running.

A.1.1 Installation steps

1. **Install Visual Studio Code** - your code editor
    - There are other good IDEs out there, and you can absolutely use one.
    - VS Code is a good and well-supported starting point
2. **Install Quarto** - for creating documents with code
    - For step 1, choose your operating system
    - For step 2, choose VS Code as your tool
3. **Install Julia using JuliaUp** - the programming language
    - Follow the directions on the GitHub page based on your operating system
    - Don't worry about the Continuous Integration (CI) section or anything below it
    - Install Julia 1.11 using `juliaup add 1.11`
    - Set this to be your default version using `juliaup default 1.11`
    - You should get a message that says something like `Configured the default Julia version to be '1.11'`
4. **In VS Code: Install extensions** from the Extensions marketplace
    - Install the **Julia** extension (provides syntax highlighting, code completion, and integrated REPL)
    - Install the **Quarto** extension (provides syntax highlighting and preview capabilities for `.qmd` files)
5. **Install GitHub Desktop** - for version control
    - This is optional if you prefer to use git through the command line or another app, but GitHub Desktop is a good default recommendation

A.1.2 Verification

After installation, you should be able to:

- Open VS Code and see the Julia and Quarto extensions listed
- Open a terminal and type `julia` to start the Julia REPL
- Create a new Quarto document (`.qmd` file) in VS Code with syntax highlighting

A.2 Dig deeper

A.2.1 Julia

Julia is a fast, modern programming language designed for scientific computing. Its syntax closely mirrors mathematical notation, making it intuitive for researchers while delivering performance comparable to C and Fortran.

JuliaUp is the official Julia version manager. It simplifies installation, allows you to maintain multiple Julia versions simultaneously, and keeps your installation current with the latest releases. This is especially useful as the Julia ecosystem evolves rapidly.

See the Julia page for more.

A.2.2 Quarto

Quarto is a scientific publishing system that enables you to combine code, results, and narrative text in reproducible documents. Think of it as the next generation of R Markdown, but with multi-language support (Julia, Python, R, and more).

This textbook is written in Quarto. Unlike traditional notebooks, Quarto documents are plain text files that render to multiple output formats (HTML, PDF, Word, presentations) while maintaining computational reproducibility.

You can learn more at:

- Official Tutorial: Hello, Quarto - basic document creation
- Official Tutorial: Computations - integrating code
- Comprehensive Quarto documentation

A.2.2.1 Writing with Markdown and math

Quarto uses Markdown syntax with LaTeX math notation. Essential references:

- Markdown Cheatsheet - basic text formatting
- LaTeX Cheatsheet - mathematical notation
- Mathpix Snip - convert equation images to LaTeX code (free tier available)
- Detexify - draw symbols to find LaTeX commands

A.2.3 Visual Studio Code

Visual Studio Code is a free, open-source code editor developed by Microsoft. Its strength lies in its extensibility—thousands of extensions add language support, debugging capabilities, and productivity tools.

For our workflow, the Julia extension transforms VS Code into a full Julia development environment with syntax highlighting, intelligent code completion, integrated debugging, and a built-in REPL. The Quarto extension provides similar capabilities for computational documents, including live preview and cell execution.

You can learn more at the official tutorial.

A.2.4 Git and GitHub

Git is a distributed version control system that tracks changes in your code over time. GitHub is a cloud-based platform that hosts Git repositories and adds collaboration features like issue tracking, pull requests, and project management.

Version control is essential for reproducible research—it allows you to track changes, collaborate with others, recover from mistakes, and share your work publicly. This textbook itself is maintained on GitHub.

You can learn more at:

- Git and GitHub for Poets - beginner-friendly video series
- GitHub Hello World – official docs
- Version Control - comprehensive guide from MIT's "Missing Semester"

B Julia Learning Resources

The computational examples in this textbook use the Julia programming language.

B.1 Why Julia?

Julia is a fast, modern, open-source programming language designed for scientific and numerical computing. The language is designed to be fast, dynamic, and easy to use and maintain.

Key advantages for this textbook include:

- High-Level Syntax: Julia has a clean and expressive syntax that closely parallels mathematical notation.
- Performance: Julia compiles to efficient machine code, achieving speeds comparable to low-level languages like C and Fortran. This solves the "two-language problem," where you might prototype in a high-level language but need to rewrite for performance.
- Simplified Dependencies: Eliminates or reduces the need for dependencies on C and Fortran libraries, which simplifies installation and maintenance.
- Open-Source and Shareable: Julia is completely open-source with excellent package management for reproducible research environments.
- Strong Ecosystem: Despite being newer, Julia has a rapidly growing ecosystem of high-quality libraries for scientific domains.

While Julia is powerful for computational thinking and research, many ecosystems remain stronger in other languages (like Python's deep learning and climate data analysis tools), so a well-rounded programmer benefits from learning multiple languages.

You can read more about Julia's design philosophy:

- Julia Data Science textbook is didactic and clear
- Why We Created Julia from the founders
- Why Julia Manifesto is more comprehensive

B.2 Learning resources

This textbook aims to never reinvent the wheel. There are lots of exceptional resources for learning Julia, or for learning computational concepts with Juila. Here are some favorites:

- MIT's Introduction to Computational Thinking: Julia-based course covering applied mathematics and computational thinking
- Julia for Nervous Beginners: free course for people hesitant but curious about learning Julia

- Julia Data Science: comprehensive introduction to data science with Julia
- FastTrack to Julia cheatsheet
- Comprehensive Julia Tutorials: YouTube playlist covering Julia topics
- Matlab-Python-Julia Cheatsheet: helpful if you're experienced in one of these languages

B.2.1 Specialized topics

Here are some additional resources for specific Julia tools and packages developed in this class

- Plotting: Makie Tutorials and MakieCon 2023 YouTube Channel
- Statistical Modeling: Turing.jl tutorials has detailed examples of using Turing for modeling

C Large Language Models ("AI")

Coding is an integral part of real-world climate-risk analysis, and large language models (LLMs; often referred to as "AI" models) are rapidly changing how some kinds of coding happen. Beyond web-based chatbots, you may have useed tools like GitHub Copilot (free for students and educators) or Claude Code (see free Deeplearning.AI Course). LLMs use powerful new technologies that can support learning and replace tedious tasks, but they can also threaten your intellectual growth and skill development (Kosmyna et al. 2025; Bastani et al. 2025).

It is clear that there are some tasks that should be delegated to these models and some tasks that must remain human-driven. However, there are tremendous differences of opinion about how most tasks in the middle can or should be allocated. As you wrestle with these questions for yourself, you should explore resources like:

- AI Snake Oil is a blog that seeks to dispel hype, remove misconceptions, and clarify the limits of AI. The authors are in the Princeton University Department of Computer Science.
- AI software assistants make the hardest kinds of bugs to spot from Pluralistic is a thoughtful and deep blog post about the perils of (mis)using LLMs for coding.
- One Useful Thing is a newsletter about AI focused on implications for work and education. The authors' prompt library is also a good resource for working with LLMs.
- Ed Zitron's Where's Your Ed At is a newsletter that takes a critical perspective on the business models and hype narratives around AI.