

Computational Thinking for Actuaries and Financial Professionals

With Applications in Julia

Alec Loudenback and Yun-Tien Lee

2024-12-16

Table of contents

Preface	1
I. Introduction	3
The approach	6
What you will learn	6
The Journey Ahead	6
Prerequisites	7
The Contents of This Book	7
Notes on formatting	8
Colophon	9
1. Why Program?	11
1.1. In this Chapter	11
1.2. Introduction	11
1.3. What's coding got to do with this?	14
1.4. The 10x Modeler	14
1.5. Avoiding Quick Fixes	15
1.6. Risk Governance	16
1.7. Managing and Leading the Transformation	16
1.8. Outlook	17
2. Why use Julia?	19
2.1. Expressiveness and Syntax	19
2.1.1. Example: Retention Analysis	20
2.1.2. Example: Random Sampling	21
2.2. The Speed	22
2.3. More of Julia's benefits	23
2.4. The Tradeoff	24
2.5. Package Ecosystem	24
2.6. Conclusion	25
3. Elements of Financial Modeling	27
3.1. In this Chapter	27
3.2. What is a model?	27
3.2.1. "Small world" vs "Large world"	28

Table of contents

3.3.	What is a <i>Financial Model</i> ?	28
3.4.	The Process of Building a Financial Model	29
3.5.	Predictive versus Explanatory Models	29
3.5.1.	A Historical Example	29
3.5.2.	Examples in the Financial Context	31
3.6.	What makes a good model?	32
3.6.1.	Achieving original purpose	32
3.6.2.	Usability	33
3.6.3.	Performance	34
3.6.4.	Separation of Model Logic and Data	34
3.6.5.	Abstraction of Modeled Systems	34
3.7.	What makes a good modeler?	35
3.7.1.	Domain Expertise	35
3.7.2.	Model Theory	36
3.7.3.	Curiosity	38
3.7.4.	Rigor	38
3.7.5.	Clarity	38
3.7.6.	Humble	38
3.7.7.	Architecture	38
3.7.8.	Planning	39
3.7.9.	Toolset	39
II. Conceptual Foundations: Programming and Abstractions		41
4.	Elements of Programming	43
4.1.	In this section	43
4.2.	Computer Science, Programming, and Coding	43
4.3.	Expressions and Control Flow	45
4.3.1.	Assignment and Variables	45
4.4.	Expressions	46
4.4.1.	Compound Expression	46
4.4.2.	Conditional Expressions	47
4.4.3.	Assignment and Variables	50
4.4.4.	Loops	51
4.4.5.	Performance of loops	52
4.5.	Data Types	52
4.5.1.	Numbers	53
4.5.2.	Type Hierarchy	55
4.5.3.	Collections	56
4.5.4.	Parametric Types	65
4.5.5.	Types for things not there	65
4.5.6.	Union Types	66
4.5.7.	Creating User Defined Types	66

Table of contents

4.5.8. Mutable structs	69
4.5.9. Constructors	70
4.6. Functions	71
4.6.1. Special Operators	71
4.6.2. Defining Functions	72
4.6.3. Defining Methods on Types	73
4.6.4. Keyword Arguments	75
4.6.5. Default Arguments	75
4.6.6. Anonymous Functions	75
4.6.7. First Class Nature	76
4.6.8. Broadcasting	77
4.6.9. Passing by Sharing	80
4.7. Scope	81
4.7.1. Modules and Namespaces	83
5. Functional Abstractions	85
5.1. In this section	85
5.2. Introduction	85
5.3. Imperative Style	86
5.3.1. Iterators	87
5.4. Functional Techniques and Terminology	89
5.4.1. map	91
5.4.2. accumulate	93
5.4.3. reduce	94
5.4.4. mapreduce	95
5.4.5. filter	96
5.4.6. More Tips on Functional Styles	97
5.5. Array-Oriented Styles	99
5.6. Recursion	100
6. Data and Types	103
6.1. In this section	103
6.2. Using Types to Value a Portfolio	103
6.3. Benefits of Using Types	103
6.4. Defining Types for Portfolio Valuation	104
6.4.1. (Multiple) Dispatch	106
6.5. Object Oriented Design	110
6.6. Assigning Behavior	110
6.7. Inheritance	110
6.7.1. Composition over Inheritance	111
7. Higher Levels of Abstraction	113
7.1. In this section	113
7.2. Introduction	113

Table of contents

7.3.	Principles for Abstraction	114
7.3.1.	Pragmatic Considerations for Model Design	115
7.4.	Interfaces	116
7.4.1.	Defining Good Interfaces	116
7.4.2.	Interfaces: A Financial Modeling Case Study	117
7.5.	Macros & Homoiconicity	121
7.5.1.	Metaprogramming in Financial Modeling	122
7.5.2.	Commonly Encountered Macros	123
III.	Conceptual Foundations: Building Performant Models	125
8.	Hardware and Its Implications	127
8.1.	In this section	127
8.2.	Computer Architecture	127
8.2.1.	Memory and Moving Data Around	127
8.2.2.	Stack vs Heap	128
8.2.3.	Processor	129
8.2.4.	Logistics Warehouse Analogy	132
8.2.6.	Speed of Computer Actions	133
9.	Writing Performant Single-Threaded Code	135
9.1.	In this section	135
9.2.	Introduction	135
9.3.	Patterns of Performant Sequential Code	136
9.3.1.	Minimize Memory Allocations	136
9.3.2.	Optimize Memory Access Patterns	137
9.3.3.	Use Efficient Data Types	141
9.3.4.	Avoid Type Instabilities	141
9.3.5.	Optimize for Branch Prediction	144
9.3.6.	Further Reading	145
10.	Debugging and Performance Measurement	147
10.1.	Setup	147
10.2.	Benchmarking	147
10.3.	Debugging	148
11.	Parallelization	151
11.1.	In this section	151
11.2.	Amdahl's Law and the Limits of Parallel Computing	151
11.3.	Types of Parallelism	153
11.4.	Vectorization	154
11.4.1.	Hardware	156
11.4.2.	Additional Packages	156

Table of contents

11.5. Multi-Threading	156
11.5.1. Tasks	156
11.5.2. Multi-Threading Overview	160
11.6. GPU and TPUs	165
11.6.1. Hardware	165
11.6.2. Utilizing a GPU	167
11.7. Multi-Processing / Multi-Device	174
11.8. Parallel Programming Models	177
11.8.1. Map-Reduce	177
11.8.2. Array-Based	181
11.8.3. Loop-Based	182
11.8.4. Task-Based	183
11.9. References	184
12. Julia Development Tips	185
12.1. In this section	185
IV. Conceptual Foundations: Learning from Related Disciplines	187
13. Elements of Computer Science	189
13.1. In this section	189
13.2. Computer Science for Financial Professionals	189
13.3. Algorithms & Complexity	190
13.3.1. Computational Complexity	190
13.3.2. Expected versus worst-case complexity	195
13.3.3. Complexity: Takeaways	195
13.4. Data Structures	195
13.4.1. Arrays	196
13.4.2. Linked Lists	196
13.4.3. Records/Structs	197
13.4.4. Dictionaries (Hash Tables)	198
13.4.5. Graphs	199
13.4.6. Trees	200
13.4.7. Data Structures Conclusion	201
13.5. Formal Verification	201
13.5.1. Basic Concept	202
13.5.2. Formal Verification in Practice	202
13.5.3. Related Topics	202
14. Applying Software Engineering Principles	205
14.1. In this section	205
14.2. Introduction	205
14.3. Testing	206

Table of contents

14.4. Documentation	206
14.5. Version Control	206
14.5.1. Collaborative Workflows	206
14.6. Distributing the Package	206
15. Statistical Inference and Information Theory	207
15.1. In This Chapter	207
15.2. Information Theory	207
15.2.1. Example: Classificaiton	209
15.2.2. Maximum Entropy Distributions	212
15.3. Bayes' Rule	216
15.3.1. Example: Model Selection via Likelihoods	217
15.4. Modern Bayesian Statistics	223
15.4.1. Background	223
15.4.2. Implications for Financial Modeling	226
15.4.3. Basics of Bayesian Modeling	227
15.4.4. Markov-Chain Monte Carlo	227
15.4.5. MCMC Algorithms	233
15.4.6. Rainfall Example (Continued)	234
15.4.7. Conclusion	244
15.4.8. Further Reading	244
V. Computational Thinking in an Actuarial and Financial Context	247
16. Modeling	249
16.1. In This Chapter	249
16.2. Key attributes	249
16.3. Possible approaches	250
16.4. How to work with data that feeds the models	252
16.5. Parsimony	253
17. Automatic Differentiation	257
17.1. In This Chapter	257
17.2. Motivation for (Automatic) Derivatives	257
17.3. Finite Differentiation	257
17.4. Automatic Differentiation	261
17.4.1. Dual Numbers	261
17.5. Performance of Automatic Differentiation	264
17.6. Automatic Differentiation in Practice	265
17.6.1. Performance	268
17.7. Forward Mode and Reverse Mode	269
17.8. Practical tips for Automatic Differentiation	269
17.8.1. Choosing between Reverse Mode and Forward Mode	269

Table of contents

17.8.2. Mutation	269
17.8.3. Custom Rules	269
17.8.4. Available Libraries	270
17.9. References	270
18. Optimization	271
18.1. In This Chapter	271
18.2. Setup	271
18.3. Differentiable programming	271
18.4. Gradient-Free Optimization	273
18.4.1. Linear optimization	273
18.4.2. Evolutionary algorithm	274
18.4.3. Simulated annealing	276
18.4.4. Bayesian optimization	277
18.4.5. BFGS	279
18.5. Model fitting	280
18.5.1. Root finding	280
18.5.2. Bracketed search algorithm	281
18.5.3. Best fitting curve	282
19. Sensitivity Analysis	285
19.1. In This Chapter	285
19.2. Setup	285
19.3. The Data	286
19.4. Common Sensitivity Analysis Methodologies	288
19.4.1. Finite Differences	288
19.4.2. Regression Analyses	288
19.4.3. Sobol Indices	289
19.4.4. Morris Method	290
19.4.5. Fourier Amplitude Sensitivity Tests	290
19.4.6. Automatic Differentiation	291
19.4.7. Scenario Analyses	291
19.5. Benchmarking	292
20. Stochastic Modeling	293
21. Visualizations	295
21.1. In This Chapter	295
22. Matrices and Their Uses	297
22.1. In This Chapter	297
22.2. Setup	297
22.3. Matrix manipulation	297
22.3.1. Multiplication	297

Table of contents

22.3.2. Inversion	298
22.4. Matrix decomposition	298
22.4.1. Eigenvalues	298
22.4.2. Singular values	299
22.4.3. Matrix factorization and fatorization machines	300
22.4.4. Principal component analysis	301
23. Learning from Data	303
23.1. In this chapter	303
23.2. Setup	303
23.3. Applications	303
23.3.1. Parameter fitting	303
23.3.2. Forecasting	303
23.3.3. Static and dynamic validation	303
23.3.4. Implied rate analysis	305
VI. Applications in Practice	307
24. Stochastic Mortality Projections	309
24.1. In This Chapter	309
24.2. Setup	309
24.3. The Data	310
24.4. Running the projection	313
24.4.1. Stochastic Projection	313
24.5. Benchmarking	315
24.6. Further Optimization	316
25. Scenario Generation	317
25.1. In This Chapter	317
25.2. Setup	317
25.3. The Data	317
25.4. Pseudo Random Number Generators	317
25.4.1. Common PRNGs	318
25.4.2. Consistent Interface	319
25.5. Common Economic Scenario Generation Approaches	319
25.5.1. Interest Rate Models	319
25.5.2. Stock Models	324
25.5.3. Copulas	328
25.6. Benchmarking	329
26. Similarity Analysis	331
26.1. In This Chapter	331
26.2. Setup	331

Table of contents

26.3. The Data	331
26.4. Common Similarity Measures	333
26.4.1. Euclidean Distance (L2 norm)	333
26.4.2. Manhattan Distance (L1 Norm)	333
26.4.3. Cosine Similarity	334
26.4.4. Jaccard Similarity	334
26.4.5. Hamming Distance	335
26.5. k-Nearest Neighbor (kNN) Clustering	335
26.6. Benchmarking	336
27. Portfolio Optimization	337
27.1. In This Chapter	337
27.2. Setup	337
27.3. The Data	337
27.4. Theory	337
27.5. Mathematical tools	338
27.5.1. Mean-variance optimization model	338
27.5.2. Efficient frontier analysis	340
27.5.3. Black-Litterman	341
27.5.4. Risk Parity	342
27.5.5. Sharpe Ratio Maximization	343
27.5.6. Robust Optimization	344
27.5.7. Asset weights from different methodologies	345
27.6. Practical considerations	346
27.6.1. Fractional purchases of assets	346
27.6.2. Large number of assets	347
28. Bayesian Mortality Modeling	349
28.1. Generating fake data	349
28.2. 1: A single binomial parameter model	353
28.2.1. Sampling from the posterior	354
28.3. 2. Parametric model	358
28.3.1. Plotting samples from the posterior	360
28.4. 3. Parametric model	362
28.5. Handling non-unit exposures	365
28.6. Predictions	368
29. Other Useful Techniques	371
29.1. In this chapter	371
29.2. Conceptual Techniques	371
29.2.1. Taking things to the Extreme	371
29.2.2. Range Bounding	371
29.3. Modeling Techniques	372
29.3.1. Serialization	372

Table of contents

VII. Appendices	373
30. Set up Julia and the Computing Environment	375
30.1. Installation	375
30.2. Package Management	375
30.3. Editors	376
30.3.1. Visual Studio Code	376
30.3.2. Notebooks	376
30.4. REPL	376
30.4.1. Help Mode	376
31. Environment and Package Management	377
31.1. In This Section	377
31.2. Projects, Manifests, and Dependencies	377
31.2.1. Project.toml	377
31.2.2. Manifest.toml	379
31.2.3. Reproducibility	380
31.3. Environments	380
31.4. Packages	381
31.4.1. Packages versus Projects	381
31.4.2. Basic Package Structure	381
31.4.3. Extension Packages	381
31.5. Registries	381
31.5.1. Local Registries	381
32. The Julia Ecosystem Today	383
32.0.1. Data	383
32.0.2. Plotting	384
32.0.3. Statistics	384
32.0.4. Machine Learning	385
32.0.5. Differentiable Programming	385
32.0.6. Utilities	386
32.0.7. Other packages	386
32.0.8. Actuarial packages	387
References	389

Preface

This book is intended to enable practitioners and advanced students of financial disciplines to utilize the tools, language, and ideas of computational and related sciences in their own work.

Part I.

Introduction

"I think one of the things that really separates us from the high primates is that we're tool builders. I read a study that measured the efficiency of locomotion for various species on the planet. The condor used the least energy to move a kilometer. And, humans came in with a rather unimpressive showing, about a third of the way down the list. It was not too proud a showing for the crown of creation. So, that didn't look so good. But, then somebody at Scientific American had the insight to test the efficiency of locomotion for a man on a bicycle. And, a man on a bicycle, a human on a bicycle, blew the condor away, completely off the top of the charts.

And that's what a computer is to me. What a computer is to me is it's the most remarkable tool that we've ever come up with, and it's the equivalent of a bicycle for our minds." - Steve Jobs (1990)

The world of financial modeling is incredibly complex and variegated. It, along with many of the sciences, is a place where practical goals harness computational tools to arrive at answers that (we hope) are meaningful in a way that tells us more about the world we live in. What this usually means specifically is that practitioners utilize computers to do the heavy work of processing data or running simulations which reveal the something about the complex systems we seek to represent. In this way, then, financial modelers must also be a craftsman who seeks not only to design new products, but must also think carefully about the tools and the process used therein.

This book seeks to aid the practitioner in developing that workmanship: we will develop new ways to look at the *process*, think about how to most clearly represent ideas, dive into details about computer hardware and bring it back up to the most abstract levels, and develop a vocabulary to more clearly express and communicate these concepts. The book contains a large number of practical examples to demonstrate that the end result is better for the journey we will take.

This book looks at programming for the applied financial professional and we will start by answering a very basic question: "*why is this relevant for financial modeling?*". The answer is simple: financial modeling is complex, data intensive, and often very abstract. Programming is the best tool humans have so far developed for rigorously transforming ideas and data into results. A builder may be the most skilled person in the world with a hammer but another with some basic training in a richer set of tools will build a better house. This book will enhance your toolkit with experience with multiple tools: a specific programming language, yes, but much more than that: a language to talk about solving problems, a deeper understanding of specific problem solving techniques, how to make decisions about what the architecture of a solution looks like, and practical advice from experienced practitioners.

The approach

The authors of the book are practicing actuaries, but we intend for the content to be applicable to nearly all practitioners in the financial industry. The discussion and examples may have an orientation towards insurance topics, but the concepts and patterns are applicable to a wide variety of related disciplines.

We will pull from examples on both sides of the balance sheet: the left (assets) and right (liabilities). We may also take the liberty to, at times, abuse traditional accounting notions: a liability is just an asset with the obligor and obligee switched. When the accounting conventions are important (such as modeling a total balance sheet) we will be mindful in explaining the accounting perspective. In practice, this means that we'll take examples that use examples of assets (fixed income, equity, derivatives) or liabilities (life insurance, annuities, long term care) and show that similar modeling techniques can be used for both.

What you will learn

It is our hope that with the help of this book, you will find it more efficient to discuss aspects of modeling with colleagues, borrow problem solving language from computer science, spot recurring structural patterns in problems that arise, and understand how best to make use of the “bicycle for your mind” in the context of financial modeling.

It is the experience of the authors that many professionals that do complex modeling as a part of their work have gotten to be very proficient *in spite of* not having substantive formal training on problem solving, algorithms, or model architecture. This book serves to fill that gap and provide the “missing semester” (or “years of practical learning”!). After reading this book, we hope that you will *appreciate* the attributes of Microsoft Excel that made it so ubiquitous, but that you *prefer* to use a programming language for the ability to more naturally express the relevant abstractions which make your models simpler, faster, or more usable by others.

Even if your direct responsibility does not entail hands-on-coding, be it management or “low-code”, the ideas and language should prove useful in guiding the work to a cleaner, more efficient solution.

The Journey Ahead

Learning a new topic, especially one that's not well trodden in a given field, can be intimidating. There are many resources available online, this book will recommend some others, and there are community support resources available - check the chat and

forums and look for the users talking about the topics that interest you. One of the wonderful things about the technology community is the degree to which content is available online for learning and reference.

Further, moving substantial parts of the financial services industry towards a digital-first, modern workflow is a monumental effort and you should seek partners on both the finance and information technology side. In general, good ideas and processes will prevail and the trick to encouraging adoption is finding the right place to plug a new idea or suggestion. One of the secondary, but still important, things this book should provide is the language and technical knowledge to partner with others (such as peers and IT) to make pragmatic decisions about the tradeoffs that will need to be made.

Prerequisites

Basic experience with financial modeling is not strictly required, but it will benefit the reader to be familiar so that the examples will not be attempting to teach both financial maths and computer science simultaneously.

Advanced financial maths (e.g. stochastic calculus) is *not* required. Indeed, this book is not oriented to the advanced technicalities of Wall Street “quants” and is instead directed at the multitudes of financial practitioners focused on producing results that are not measured in the microseconds of high-frequency trading.

Prior programming experience is *not* required either: Chapter 4 introduces the basic syntax and concepts while Chapter 30 covers setting up your environment to follow along. For readers with background in programming, we recommend skimming Chapter 4 and reading in full the sections which have a \square symbol in the margin, which is our way of highlighting Julia-specific content to be aware of.

i TODO

Create a venn diagram showing financial modeling at the intersection of statistics, financial math, computer science.

The Contents of This Book

Part 1 of the book addresses the theoretical and technical foundations of programming, as well as the conceptual basis for financial modelling. It familiarizes the readers with key functional programming principles, alongside introducing important aspects of software engineering relevant to financial modelling.

Parts 2 and 3 bridge the gap between theory and practical applications, underlining the features of Julia that make it a robust tool for real-world financial and actuarial contexts. Through a careful exploration of topics like sensitivity analysis, optimization, stochastic modeling, visualization, and practical financial applications, the book demonstrates how Julia's high-level, high-performance programming capabilities can enhance accuracy and efficiency in financial modelling. As an up-and-coming language loved for its speed and simplicity, Julia is ripe for wide adoption in the financial sector. The time for this book is ripe, as it will satiate the growing demand for professionals who want to blend programming skills with financial modelling acumen.

While we have chosen to use Julia for the examples in this book, the vast majority of the concepts presented are not Julia-specific. We will attempt to motivate why Julia works so well as a language for financial modeling but like mathematics and applied mathematics, the concepts are portable even if the numbers (language) changes. Readers are encouraged to follow along the examples on their own computer (see instructions for Julia in Chapter 30) and the entire book is available on GitHub at [#TODO: determine book URL].

Notes on formatting

When a concept is defined for the first time, the term will be **bold**. Code, or references to pieces of code will be formatted in inline code style like `1+1` or in separate code blocks:

"This is a code block that doesn't show any results"

"This is a code block that does show output"

"This is a code block that does show output"

When we show inline commands are to be sent to Pkg mode in the REPL (see Section 31.1), such as such as `add DataFrames`, we will try to make it clear in the context. If using Pkg mode in standalone codeblocks, it will be presented showing the full prompt, such as:

```
(@v1.10) pkg> add DataFrames
```

There will be various callout blocks which indicate tips or warnings. These should be self-evident but we wanted to point to a particular callout which is intended to convey advice that stems from practical modeling experience of the authors:

Financial Modeling Pro-tip

This box indicates a side note that's particularly applicable to improving your financial modeling.

Colophon

The HTML and PDF book were rendered using Quarto and Quarto's open source dependencies like Pandoc and LaTeX.

The HTML version of this book uses Lato for the body font and JuliaMono for the monospace font.

The PDF version of this book uses TeX Gyre Pagella for the body font and JuliaMono for the monospace font.

The cover was designed by Alec Loudeback using Affinity Designer with the graphic used under permission by user cormullion on Github.

This book was rendered on August 13, 2024. The system used to generate the code and benchmarks was:

```
versioninfo()
```

```
Julia Version 1.10.4
Commit 48d4fd48430 (2024-06-04 10:41 UTC)
Build Info:
  Official https://julialang.org/ release
Platform Info:
  OS: macOS (arm64-apple-darwin22.4.0)
  CPU: 8 × Apple M3
  WORD_SIZE: 64
  LIBM: libopenlibm
  LLVM: libLLVM-15.0.7 (ORCJIT, apple-m1)
Threads: 4 default, 0 interactive, 2 GC (on 4 virtual cores)
Environment:
  JULIA_NUM_THREADS = auto
```


1. Why Program?

"Humans are allergic to change. They love to say, 'We've always done it this way.' I try to fight that. That's why I have a clock on my wall that runs counterclockwise." - Grace Hopper (1987)

[Drafting Note: This chapter is pulled from the article published in 2020 and needs to be adapted for the book's audience. Also to include: why not low-code solutions?]

1.1. In this Chapter

We motivate why a financial professional should adopt programming skills which will improve their own capabilities and enjoyment of the discipline, whilst allowing themselves to better themselves and the industry we work in.

1.2. Introduction

The financial sector is undergoing a profound transformation. In an era defined by big data, (pseudo) artificial intelligence, and rapid technological advancement, the traditional boundaries of finance are expanding and blurring. From Wall Street to Main Street, from global investment banks to local credit unions, technology is reshaping how financial services are delivered, how risks are managed, and how decisions are made.

This digital revolution is not just changing the tools we use; it's fundamentally altering the skills required to succeed in finance. In the past, a strong foundation in mathematics, economics, and financial theory was sufficient for most roles in the industry. Today, these skills, while still crucial, are increasingly being augmented—and in some cases, superseded—by technological proficiency.

At the forefront of this shift is the growing importance of programming skills. In the beginning of the computer era in finance, the differentiating skill was being able to utilize digital computing, data processing, and calculation engines to automate, analyze, and report on the business. These skills required low level programming and the success of many of those early programs is evident in their legacy: many of them are still around in the 2020s!

1. Why Program?

At some point, due to regulatory pressures, attempts at organization efficiencies, or management decision making, the skill of programming became highly specialized and most financial professionals (investment analysts, actuaries, accountants, etc.) became relegated to being “business users”, utilizing either Microsoft Excel or a proprietary third-party software to accomplish their responsibilities. The reasons for this were not totally wrong, even in retrospect. At some point, between internet access, an increasingly complex stack of software sitting between the developer and the hardware, and the proliferation of computer security risks it makes some sense that many financial developers were pushed out of the programming trade and instead specialized, separate business and IT units were developed. Of course, this led to many inefficiencies and is now swinging back the other way.

What’s changed that’s enabling financial professionals to re-engage with the powerful tools that programming provides? Some reasons include:

1. **Code management tools.** Github and other version control systems provide best-in-class ways of managing codebase changes and collaboration. Tools exist to scan repositories for leaking secrets, security vulnerabilities, and dependency management.
2. **Increasingly accessible development.** Originally, very few layers of complexity existed between the written code and running it on the mainframe. Over time, drivers, operating systems, networking, dependencies, and compilers made development more complex. Today, languages, libraries, code editors, and deployment tools have smoothed many of these frictions.
3. **Competitive Pressures.** An increasingly commoditized financial product with evermore competition has led to a need to improve efficiency of manufacturing and selling financial products. Having a business developer is a lot more efficient than a business user who needs to get an IT developer to implement something. Further, pressures from outside the financial sector abound: It’s easier to teach a tech developer enough to be successful in a finance role than it is to teach a finance professional development skills.
4. **Regulatory and Risk Demands.** Pressures that previously motivated the move to proprietary software for modeling included regulatory reporting, internal risk metrics, and management performance evaluation. However, companies are realizing that their unique products, risk frameworks, preferred management measurements, and employee potential means that having a bespoke internal model is seen as a key capability. Many regulatory frameworks also encourage the use of a bespoke model, which is a particularly attractive option especially for those who view the given regulatory framework as inappropriately reflecting their own business and risk profile.

One subset of business analysts that *did not* start to migrate away from development as a strategic part of their value were “quants” or quantitative analysts who heavily utilized programming skills to develop unique products, trading strategies, modeling frame-

1.2. Introduction

works, and risk engines. This book is not really for that class of people and is instead geared towards the mass of financial professionals who want to get some of the benefits of the tools that the quants have been using for years. Whether you're an investment banker modeling complex derivatives, an actuary calculating insurance risks, a financial planner optimizing client portfolios, or a risk manager stress-testing scenarios, the ability to code is becoming as fundamental as the ability to use a spreadsheet was a generation ago. To remain competitive, adaptable, and effective in the evolving landscape of finance, professionals must embrace programming as a core skill.

As we delve into this topic, keep in mind that learning to code is not about replacing traditional financial acumen—it's about augmenting and enhancing it. It's about equipping yourself with the tools to tackle the complex, data-driven challenges of modern finance. In short, it's about future-proofing your career in an industry that is increasingly defined by its ability to innovate and adapt to technological change.

Today, there is a trend towards technological value-creation and is evident across many traditional sectors. Tesla claims that it's a technology company; Amazon is the #1 product retailer because of its vehement focus on internal information sharing¹; Airlines are so dependent on their systems that the skies become quieter on the rare occasion that their computers give way.

Why is it, that companies that are so involved in *things* (cars, shopping) and *physical services* (flights) are so much more focused on improving their technological operations than insurance companies *whose very focus is 'information-based'*? **The market has rewarded those who have prioritized their internal technological solutions.**

Commoditized investing services and challenging yield environments have reduced companies' comparative advantage to "manage money". Spread compression and the explosion of consumer-oriented investment services means a more competitive focus on the ability to manage the entire asset or policy's lifecycle efficiently (digitally), perform more real-time analysis of experience and risk management, and handle the growing product and regulatory complexity.

These are problems that have technological solutions and are waiting for insurance company adoption.

Companies that treat data like coordinates on a grid (spreadsheets) *will get left behind*. Two main hurdles have prevented technology companies from breaking into insurance:

1. High regulatory barriers to entry, and
2. Difficulty in selling complex insurance products without traditional distribution.

¹Have you had your Bezos moment? What you can learn from Amazon

1. Why Program?

Once those two walls are breached, traditional insurance companies without a strong technology core will struggle to keep up. The key to thriving is not just adding “developers” to an organization; it’s going to be **getting domain experts like actuaries to be an integral part of the technology transformation.**

1.3. What's coding got to do with this?

Everything. Programming is the optimal way to interact between the computer and actuary — and importantly between computer and computer. Programming is the actionable expression of ideas, math, analysis, and information. Think of programming as the 21st-century leap in the actuary’s toolkit, just as spreadsheets were in the preceding 40 years. Versus a spreadsheet-oriented workflow:

- More natural automation of, and between processes
- Better reproducibility
- Scaling to fit any size dataset and workload
- Statistics and machine learning capabilities
- Advanced visualizations to garner new views into your data

This list isn’t comprehensive and some benefits are subtle — when you are code-oriented instead of spreadsheet-oriented, you tend to want to structure your data in a portable and shareable way. For example, relying more on data warehouses instead of email attachments. This, in turn, enables data discovery and insights that otherwise wouldn’t be there. Investing in a code-oriented workflow is playing the long-game.

The financial modeler of the future needs to have coding as one of their core skills. Already today, the advances of business processes, insurance products, and financial ingenuity are written with lines of code — *not* spreadsheets. Not being able to code *necessarily* means that you are *following* what others are doing today.

It’s commonly accepted now that to gather insights from your data, you need to know how to code. Similar to your data, your business architecture, modeling needs, and product peculiarities are often better suited to customized solutions. Why stop at data science when learning how to solve problems with a computer?

1.4. The 10x Modeler

As we swing back to a technological focus, we do not leave the finance-driven complexity behind. The increasingly complex business needs will highlight a large productivity difference between a financial modeler who can code and one who can’t — simply because the former can react, create, synthesize, and model faster than the latter. From the

1.5. Avoiding Quick Fixes

efficiency of transforming administration extracts, summarizing and aggregating valuation output, to analyzing available data in ways that spreadsheets simply can't handle, you can become a "10x Modeler"².

Flipping switches in a graphical user interface versus being able to *build models* is the difference between having a surface-level familiarity and having full command over the analysis and the concepts involved — with the flexibility to do what your software can't.

Your current software might be able to perform the first layer of analysis but be at a loss when you want to visualize, perform sensitivity analysis, statistics, stochastic analysis, or process automation. Things that, when done programmatically, are often just a few lines of additional code.

Should you drop the license for your software vendor? No, not yet anyway. But the ability to supplement and break out of the modeling box has been an increasingly important part of most professionals' work and this trend appears to be accelerating.

Additionally, code-based solutions can leverage the entire-technology sector's progress to solve problems that are *hard* otherwise: scalability, data workflows, integration across functional areas, version control and versioning, model change governance, reproducibility, and more.

30-40 years ago, there were no vendor-supplied modeling solutions and so you had no choice but to build models internally. This shifted with the advent of vendor-supplied modeling solutions. Today, it's never been better for companies to leverage open source to support their custom modeling, risk analysis/monitoring, and reporting workflows.

1.5. Avoiding Quick Fixes

One tantalizing path to contemplate is to avoid *really* learning how to code. Between AI solutions being developed and low-code offerings, is there really a need to learn the fundamentals of coding? We argue that there is, for the basic reason that coding is not about mechanically typing out lines in an editor, but both a tool and a craft that is designed to enhance and apply your own creative and logical thinking.

The current generation of AI is fundamentally limited. Yann LeCun, Meta's (Facebook's) Chief AI Scientist describes the large-language model (LLM) approach as not even at the level of intelligence of a cat, and that we are decades away from true artificial general intelligence (AGI). These models have a "very limited understanding of logic... do not understand the physical world, do not have persistent memory, cannot reason in any reasonable definition of the term and cannot plan... hierarchically"

²The 10x [Rockstar] developer is NOT a myth

1. Why Program?

(Murphy and Criddle 2024). An important role for AI to play will be to *support* modelers in boilerplate, syntactical hurdles (“in VBA I would do it like this, but in Julia how do I do X, Y, or Z”), and basic algorithmic support. What is not likely to change in the short term is most of the value that a modeler brings to the table: creative thinking, understanding of company and market dynamics, and capability to understand broader architecture and conceptual aspects of modeling.

A similarly fraught path is low-code solutions. Low-code solutions are inherently limiting in their capabilities and lock you into a particular vendored solution. If you know enough about what you are trying to do to be able to state it in clearly in plain English, then you are most of the way to being able to program in a full coding solution (AI can actually help bridge this gap here). As soon as you hit a limitation of the system (“I’d like to use XYZ optimization algorithm at each timestep”), you are reliant on the vendor to implement that option in the “low code” solution. Further, you are out-sourcing a lot of the important inner-workings of the model to someone else and not building that expertise yourself of in-house somewhere.

1.6. Risk Governance

Code-based workflows are highly conducive to risk governance frameworks as well. If a modern software project has all of the following benefits, then why not a modern insurance product and associated processes?

- Access control and approval processes
- Version control, version management, and reproducibility
- Continuous testing and validation of results
- Open and transparent design
- Minimization of manual overrides, intervention, and opportunity for user error
- Automated trending analysis, system metrics, and summary statistics
- Continuously updated, integrated, and self-generating documentation
- Integration with other business processes through a formal boundary (e.g. via an API)
- Tools to manage collaboration in parallel and in sequence

1.7. Managing and Leading the Transformation

The ability to understand the concepts, capabilities, challenges, and lingo is not a dichotomy, it’s a spectrum. Most actuaries, even at fairly high levels, are still often involved in analytical work. Still above that, it’s difficult to lead something that you don’t understand.

1.8. Outlook

Conversely, the skill and practice of coding enhances managerial capabilities. When you are really skilled at pulling apart a problem or process into its constituent parts and designing optimal solutions; that's a core attribute of leadership: having the vision of where the organization *should be* instead of thinking about where it is now.

Nor is the skillset described here limiting in any other aspect of career development any more than mathematical ability, project collaboration, or financial acumen — just to name a few.

1.8. Outlook

It will increasingly be essential for companies to modernize to remain competitive. That modernization isn't built with big black-box software packages; it will be with domain experts who can translate the expertise into new forms of analysis - doing it faster and more robustly than the competition.

SpaceX doesn't just hire rocket scientists - they hire rocket scientists who code.

Be a modeler who codes.

2. Why use Julia?

[Drafting Note: This chapter is pulled from the article published in 2021 and needs to be adapted for the book's audience.]

Julia is relatively new¹, and *it shows*. It is evident in its pragmatic, productivity-focused design choices, pleasant syntax, rich ecosystem, thriving communities, and its ability to be both very general purpose and power cutting edge computing.

With Julia: math-heavy code looks like math; it's easy to pick up, and quick-to-prototype. Packages are well-integrated, with excellent visualization libraries and pragmatic design choices.

Julia's popularity continues to grow across many fields and there's a growing body of online references and tutorials, videos, and print media to learn from.

Large financial services organizations have already started realizing gains: BlackRock's Aladdin portfolio modeling, the Federal Reserve's economic simulations, and Aviva's Solvency II-compliant modeling². The last of these has a great talk on YouTube by Aviva's Tim Thornham, which showcases an on-the-ground view of what difference the right choice of technology and programming language can make. Moving from their vendor-supplied modeling solution was **1000x faster, took 1/10 the amount of code, and was implemented 10x faster**.

The language is not just great for data science — but also modeling, ETL, visualizations, package control/version management, machine learning, string manipulation, web-backends, and many other use cases. Julia is well suited for financial modeling work: easy to read and write and very performant.

2.1. Expressiveness and Syntax

Expressiveness is the *manner in which and scope of* ideas and concepts that can be represented in a programming language. **Syntax** refers to how the code *looks* on the screen and its readability.

¹Python first appeared in 1990. R is an implementation of S, which was created in 1976, though depending on when you want to place the start of an independent R project varies (1993, 1995, and 2000 are alternate dates). The history of these languages is long and substantial changes have occurred since these dates.

²Aviva Case Study

2. Why use Julia?

In a language with high expressiveness and pleasant syntax, you:

- Go from idea in your head to final product faster.
- Encapsulate concepts naturally and write concise functions.
- Compose functions and data naturally.
- Focus on the end-goal instead of fighting the tools.

Expressiveness can be hard to explain, but perhaps two short examples will illustrate.

2.1.1. Example: Retention Analysis

This is a really simple example relating `Cessions`, `Policies`, and `Lives` to do simple retention analysis.

First, let's define our data:

```
# Define our data structures
struct Life
    policies
end

struct Policy
    face
    cessions
end

struct Cession
    ceded
end
```

Now to calculate amounts retained. First, let's say what retention means for a `Policy`:

```
# define retention
function retained(pol::Policy)
    pol.face - sum(cession.ceded for cession in pol.cessions)
end
```

And then what retention means for a `Life`:

```
function retained(l::Life)
    sum(retained(policy) for policy in life.policies)
end
```

2.1. Expressiveness and Syntax

It's almost exactly how you'd specify it English. No joins, no boilerplate, no fiddling with complicated syntax. You can express ideas and concepts the way that you think of them, not, for example, as a series of dataframe joins or as row/column coordinates on a spreadsheet.

We defined `retained` and adapted it to mean related, but different things depending on the specific context. That is, we didn't have to define `retained_life(...)` and `retained_pol(...)` because Julia can be *dispatch* based on what you give it. This is, as some would call it, unreasonably effective.

Let's use the above code in practice then.

The julia> syntax indicates that we've moved into Julia's interactive mode (REPL mode):

```
# create two policies with two and one cessions respectively
julia> pol_1 = Policy( 1000, [ Cession(100), Cession(500)] )
julia> pol_2 = Policy( 2500, [ Cession(1000) ] )

# create a life, which has the two policies
julia> life = Life([pol_1, pol_2])

julia> retained(pol_1)
400

julia> retained(life)
1900
```

And for the last trick, something called “broadcasting”, which automatically vectorizes any function you write, no need to write loops or create `if` statements to handle a single vs repeated case:

```
julia> retained.(life.policies) # retained amount for each policy
[400 , 1500]
```

2.1.2. Example: Random Sampling

As another motivating example showcasing multiple dispatch, here's random sampling in Julia, R, and Python.

We generate 100:

- Uniform random numbers
- Standard normal random numbers
- Bernoulli random number
- Random samples with a given set

2. Why use Julia?

Table 2.1.: A comparison of random outcome generation in Julia, R, and Python.

Julia	R	Python
<code>using Distributions</code>	<code>runif(100)</code>	<code>import scipy.stats as sps</code>
	<code>rnorm(100)</code>	<code>import numpy as np</code>
<code>rand(100)</code>	<code>rbern(100, 0.5)</code>	
<code>rand(Normal(), 100)</code>	<code>sample(c("Preferred", "Standard"),</code>	<code>sps.uniform.rvs(size=100)</code>
<code>rand(Bernoulli(0.5), 100)</code>	<code>100, replace=TRUE)</code>	<code>sps.norm.rvs(size=100)</code>
<code>rand(["Preferred", "Standard"], 100)</code>		<code>sps.bernoulli.rvs(p=0.5, size=100)</code>
		<code>np.random.choice(["Preferred", "Standard"],</code>
		<code>size=100)</code>

By understanding the different types of things passed to `rand()`, it maintains the same syntax across a variety of different scenarios. We could define `rand(Cession)` and have it generate a random `Cession` like we used above.

2.2. The Speed

As the journal Nature said, “Come for the Syntax, Stay for the Speed”.

Recall the Solvency II compliance which ran 1000x faster than the prior vendor solution mentioned earlier: what does it mean to be 1000x faster at something? It’s the difference between something taking 10 seconds instead of 3 hours — or 1 hour instead of 42 days.

What analysis would you like to do if it took less time? A stochastic analysis of life-level claims? Machine learning with your experience data? Daily valuation instead of quarterly?

Speaking from experience, speed is not just great for production time improvements. During development, it’s really helpful too. When building something, I can see that I messed something up in a couple of seconds instead of 20 minutes. The build, test, fix, iteration cycle goes faster this way.

Admittedly, most workflows don’t see a 1000x speedup, but 10x to 1000x is a very common range of speed differences vs R or Python or MATLAB.

Sometimes you will see less of a speed difference; R and Python have already circumvented this and written much core code in low-level languages. This is an example of what’s called the “two-language” problem where the language productive to write in isn’t very fast. For example, more than half of R packages use C/C++/Fortran and core packages in Python like Pandas, PyTorch, NumPy, SciPy, etc. do this too.

2.3. More of Julia's benefits

Within the bounds of the optimized R/Python libraries, you can leverage this work. Extending it can be difficult: what if you have a custom retention management system running on millions of policies every night?

Julia packages you are using are almost always written in pure Julia: you can see what's going on, learn from them, or even contribute a package of your own!

2.3. More of Julia's benefits

Julia is easy to write, learn, and be productive in:

- It's free and open-source
 - Very permissive licenses, facilitating the use in commercial environments (same with most packages)
- Large and growing set of available packages
- Write how you like because it's multi-paradigm: vectorizable (R), object-oriented (Python), functional (Lisp), or detail-oriented (C)
- Built-in package manager, documentation, and testing-library
- Jupyter Notebook support (it's in the name! **Julia-Python-R**)
- Many small, nice things that add up:
 - Unicode characters like α or β
 - Nice display of arrays
 - Simple anonymous function syntax
 - Wide range of text editor support
 - First-class support for missing values across the entire language
 - Literate programming support (like R-Markdown)
- Built-in Dates package that makes working with dates pleasant
- Ability to directly call and use R and Python code/packages with the PyCall and RCall packages
- Error messages are helpful and tell you *what line* the error came from, not just the type of error
- Debugger functionality so you can step through your code line by line

For power-users, advanced features are easily accessible: parallel programming, broadcasting, types, interfaces, metaprogramming, and more.

These are some of the things that make Julia one of the world's most loved languages on the StackOverflow Developer Survey.

2. Why use Julia?

For those who are enterprise-minded: in addition to the liberal licensing mentioned above, there are professional products from organizations like Julia Computing that provide hands-on support, training, IT governance solutions, behind-the-firewall package management, and deployment/scaling assistance.

2.4. The Tradeoff

Julia is fast because it's compiled, unlike R and Python where (loosely speaking) the computer just reads one line at a time. Julia compiles code "just-in-time": right before you use a function for the first time, it will take a moment to pre-process the code section for the machine. Subsequent calls don't need to be re-compiled and are very fast.

A hypothetical example: running 10,000 stochastic projections where Julia needs to pre-compile but then runs each 10x faster:

- Julia runs in 2 minutes: the first projection takes 1 second to compile and run, but each 9,999 remaining projections only take 10ms.
- Python runs in 17 minutes: 100ms of a second for each computation.

Typically, the compilation is very fast (milliseconds), but in the most complicated cases it can be several seconds. One of these is the "time-to-first-plot" issue because it's the most common one users encounter: super-flexible plotting libraries have a lot of things to pre-compile. So in the case of plotting, it can take several seconds to display the first plot after starting Julia, but then it's remarkably quick and easy to create an animation of your model results. The time-to-first plot is a solvable problem that's receiving a lot of attention from the core developers and will get better with future Julia releases.

For users working with a lot of data or complex calculations (like actuaries!), the run-time speedup is worth a few seconds at the start.

2.5. Package Ecosystem

Using packages as dependencies in your project is assisted by Julia' bundled package manager.

For each project, you can track the exact set of dependencies and replicate the code/process on another machine or another time. In R or Python, dependency management is notoriously difficult and it's one of the things that the Julia creators wanted to fix from the start.

Packages can be one of the thousands of publicly available, or private packages hosted internally behind a firewall.

2.6. Conclusion

Another powerful aspect of the package ecosystem is that due to the language design, packages can be combined/extended in ways that are difficult for other common languages. This means that Julia packages often interop without any additional coordination.

For example, packages that operate on data tables work without issue together in Julia. In R/Python, many features tend to come bundled in a giant singular package like Python's Pandas which has Input/Output, Date manipulation, plotting, resampling, and more. There's a new Consortium for Python Data API Standards which seeks to harmonize the different packages in Python to make them more consistent (R's Tidyverse plays a similar role in coordinating their subset of the package ecosystem).

In Julia, packages tend to be more plug-and-play. For example, every time you want to load a CSV you might not want to transform the data into a dataframe (maybe you want a matrix or a plot instead). To load data into a dataframe, in Julia the practice is to use both the CSV and DataFrames packages, which help separate concerns. Some users may prefer the Python/R approach of less modular but more all-inclusive packages.

2.6. Conclusion

Looking at other great tools like R and Python, it can be difficult to summarize a single reason to motivate a switch to Julia, but hopefully this article piqued an interest to try it for your next project.

That said, Julia shouldn't be the only tool in your tool-kit. SQL will remain an important way to interact with databases. R and Python aren't going anywhere in the short term and will always offer a different perspective on things!

In an earlier article, I talked about becoming a **10x Actuary** which meant being proficient in the language of computers so that you could build and implement great things. In a large way, the choice of tools and paradigms shape your focus. Productivity is one aspect, expressiveness is another, speed one more. There are many reasons to think about what tools you use and trying out different ones is probably the best way to find what works best for you.

It is said that you cannot fully conceptualize something unless your language has a word for it. Similar to spoken language, you may find that breaking out of spreadsheet coordinates (and even a dataframe-centric view of the world) reveals different questions to ask and enables innovated ways to solve problems. In this way, you reward your intellect while building more meaningful and relevant models and analysis.

3. Elements of Financial Modeling

"Truth ... is much too complicated to allow anything but approximations" -
John von Neumann

3.1. In this Chapter

We explain what constitutes a financial model and what are common uses of a model.
We explain what makes an adept practitioner.

3.2. What is a model?

A **model** represents aspects of the world around us distilled down into simpler, more tractable components. It is impossible to fully capture the everything that may affect the objects of our interest. We may build models for a variety of reasons, as listed in Table 3.1.

Table 3.1.: The REDCAPE model use framework, from "The Model Thinker" by Scott Page.

Use	Description
Reason	To identify conditions and deduce logical implications.
Explain	To provide (testable) explanations for empirical phenomena.
Design	To choose features of institutions, policies, and rules.
Communicate	To relate knowledge and understandings.
Act	To guide policy choices and strategic actions.
Predict	To make numerical and categorical predictions of future and unknown phenomena.
Explore	To investigate possibilities and hypotheticals.

For example, say we want to simulate the returns for the stocks in our retirement portfolio. It would be impossible to try to build a model which would capture all of the individual people working jobs and making decisions, weather events that damage property,

3. Elements of Financial Modeling

political machinations, etc. Instead, we try to capture certain fundamental characteristics. For example, it is common to model equity returns as cumulative pluses and minuses from random movements where those movements have certain theoretical or historical characteristics.

Whether we are using this model of equity returns to estimate available retirement income or replicate an exotic option price, a key aspect of the model is the **assumptions** used therein. For the retirement income scenario we might *assume* a healthy eight percent return on stocks and conclude that such a return will be sufficient to retire at age 53. Alternatively, we may assume that future returns will follow a stochastic path with a certain distribution of volatility and drift. These two assumption sets will produce **output** - results from our model that must be inspected, questioned, and understood in the context of the “small world” of the model’s mechanistic workings. Lastly, to be effective practitioners we must be able to contextualize the “small world” results within the “large world” that exists around us.

3.2.1. “Small world” vs “Large world”

What do we mean by “small world” vs “large world”? Say that our model is one that discounts a fixed set of future cashflows using the US Treasury rate curve. If I run my model using current rates today, and then re-run my model tomorrow with the same future cashflows and the present value of those cashflows has increased by 5% I may ask why the result has changed so much in such a short period of time! In the “small”, mechanistic world of the model I may be able to see that the rates I used to discount the cashflows have fallen substantially. The “small world” answer is that the inputs have changed which produced a mechanical change in the output. The “large world” answer may be that the Federal Reserve lowered the Federal Funds Rate to prevent the economy from entering a deflationary recession. Of course, we can’t completely explain the relation between our model and the real world (otherwise we could capture that relationship in our model!). An effective practitioner will always try to look up from the immediate work and take stock of how the world at large *is* or *is not* reflected in the model.

3.3. What is a *Financial Model*?

Financial models are those used extensively to ascertain better understanding of complex contracts, perform scenario analysis, and inform market participants’ decisions related to perceived value (and therefore price). It can’t be quantified directly, but it is likely not an exaggeration that many billions of dollars is transacted each day as a result of decisions made from the output of financial models.

Most financial models can be characterized with a focus on the first or both of:

3.4. The Process of Building a Financial Model

1. Attempting to project pattern of cashflows or obligations at future timepoints
2. Reducing the projected obligations into a current value

Examples of this:

- Projecting a retiree's savings through time (1), and determining how much they should be saving today for their retirement goal (2)
- Projecting the obligation of an exotic option across different potential paths (1), and determining the premium for that option (2)

Models are sometimes taken a step further, such as transforming the underlying **economic view** into an accounting or regulatory view (such as representing associated debits and credits, capital requirements, or associated intangible, capitalized balances).

We should also distinguish a financial model from a purely statistical model, where often the inputs and output data are known and the intention is to estimate relationships between variables (example: linear regressions). That said, a financial model may have statistical components and many aspects of modeling is shared between the two kinds.

3.4. The Process of Building a Financial Model

i TODO: Describe model building process and make associated diagram

3.5. Predictive versus Explanatory Models

Given a set of inputs, our model will generate an output and we are generally interested in its accuracy. *The model need not have a realistic mechanism for how the world works.* That is, we may primarily be interested in accurately calculating an output value without the model having any scientific, explanatory power of how different parts of the real-world system interact.

3.5.1. A Historical Example

Consider the classic underdog story where Copernicus overthrew the status quo when he proposed (correctly) that the earth orbited the sun instead of the other way around¹.

¹Prof. Richard Fitzpatrick has excellent coverage of the associated mathematics and implications in "A Modern Almagest": <https://farside.ph.utexas.edu/books/Syntaxis/Almagest/Almagest.html>

3. Elements of Financial Modeling

The existing Ptolemy model used a geocentric view of the solar system in which the planets and sun orbited the Earth in perfect circles with an epicycle used to explain retrograde motion (as seen in Figure 3.1). Retrograde motion is the term used to describe the apparent, temporarily reversed motion of a planet as viewed from Earth when the Earth is overtaking the other planet in orbit around the sun. This was accurate enough to match the observational data that described the position of the planets in the sky.

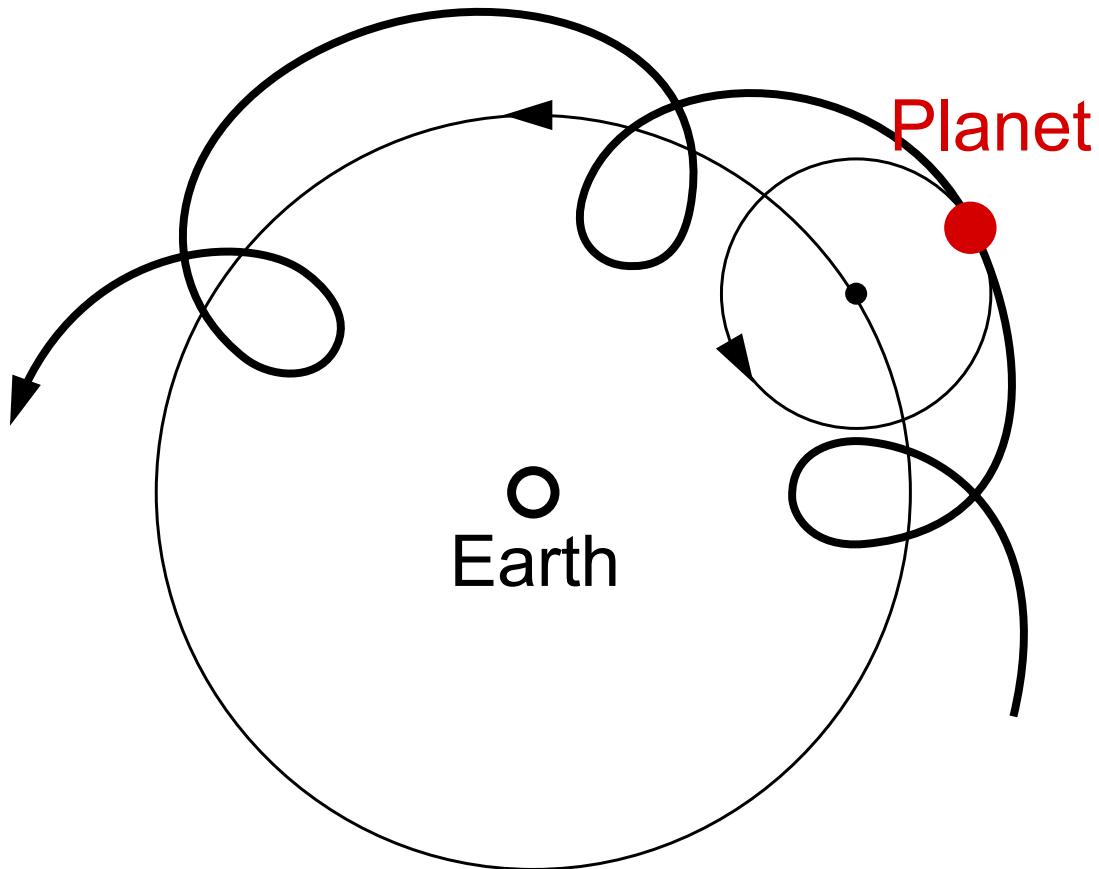


Figure 3.1.: In the Ptolemy solar model, the retrograde motion of the planets was explained by adding an epicycle to the circular orbit around the earth.

Famously, Copernicus came along and said that the sun, not the Earth, should be at the center (a heliocentric model). Earth revolves around the sun! Today, we know this to be a much better description of reality than one in which the Earth arrogantly sits at the center of the universe. However the model was actually slightly *less* accurate in predicting the apparent position of the planets (to the limits of observational precision at the time)! Why would this be?

3.5. Predictive versus Explanatory Models

First, the Copernican proposal still used perfectly circular orbits with an epicycle adjustment, which we know today to be inaccurate (in favor of an elliptical orbit consistent with the theory of gravity). *Despite being more scientifically correct, it was still not the complete picture.*

Second, the geocentric model was already very accurate because it was essentially a Taylor-series approximation which described to sufficient observational accuracy the apparent position of the planet relative to the Earth. *The heliocentric model was effectively a re-parameterization of the orbital approximation.*

Third, we have considered a limited criteria for which we are evaluating the model for accuracy, namely apparent position of the planets. *It's not until we contemplate other observational data that the Copernican model would demonstrate greater modeling accuracy:* apparent brightness of the planets as they undergo retrograde motion and angular relationship of the planets to the sun.

For modelers today, this demonstrates a few things to keep in mind:

1. Predictive models need not have a scientific, causal structure to make accurate predictions.
2. It is difficult to capture the complete scientific inter-relationships of a system and much care and thought needs to be given in what aspects are included in our model.
3. We should look at, or seek out, additional data that is related to our model because we may accurately fit (or overfit) to one outcome while achieving an increasingly poor fit to other related variables.

Striving to better understand the world is a *good thing to do* but trying to include more components into the model is not always going to help achieve our goals.

3.5.2. Examples in the Financial Context

3.5.2.1. Home Prices

American home prices which have a strong degree of seasonality and have the strongest prices around April of each year. We may find that including a simple oscillating term in our model captures the variability in prices *better* than if we tried to imperfectly capture the true market dynamics of home sales: supply and demand curves varying by personal (job bonus payment timing, school calendars), local (new homes built, company relocation), and national (monetary policy, tax incentives for home-ownership). In other words, one could likely predict a stable pattern like this with a model that contains a simple sinusoidal periodic component. One could likely spend months trying to build a more scientific model and not achieve as good of fit, *even though the latter tries to be more conceptually accurate.*

3. Elements of Financial Modeling



3.5.2.2. Replicating Portfolio

Another example in the financial modeling realm: in attempting to value a portfolio of insurance contracts a **replicating portfolio** of hypothetical assets will sometimes be constructed². The point of this is to create a basket of assets that can be more quickly (minutes to hours) valued in response to changing market conditions than it would take to run the actuarial model (hours to days). This is an example where the basket of assets has no ability to explain why the projected cashflows are what they are - but retains strong predictive accuracy.

3.6. What makes a good model?

The answer is: *it depends.*

3.6.1. Achieving original purpose

A model is built for a specific set of reasons and therefore we must evaluate a model in terms of achieving that goal. We should not critique a model if we want to use it outside of what it was intended to do. This includes: contents of output and required level of accuracy.

A model may have been created to for scenario analysis to value all assets in a portfolio to within half a percent of a more accurate, but much more computationally expensive model. If we try to add a never-before-seen asset class or use the model to order trades we may be extending the design scope of the original model.

²See, e.g., SOA Investment Symposium March 2010. *Replicating Portfolios in the Insurance Industry* (Curt Burmeister Mike Dorsel Patricia Matson)

3.6. What makes a good model?

3.6.2. Usability

How easy is it for someone to use? Does it require pages and pages of documentation, weeks of specialized training and an on-call help desk? *All else equal*, it is an indicator of how usable the model is by the amount of support and training. However, one may sometimes wish to create a highly capable, complex model which is known to require a high amount of experience and expertise. An analogy here might be the cockpit of a small Cessna aircraft versus a fighter jet: the former is a lot simpler and takes less training to master but is also more limited.

Figure 3.2 illustrates this concept and shows that if your goal is very high capability that you may need to expect to develop training materials and support the more complex model. On this view, a better model is one that is able to have a shorter amount of time and experience to achieve the same level of capability.

```
[Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```

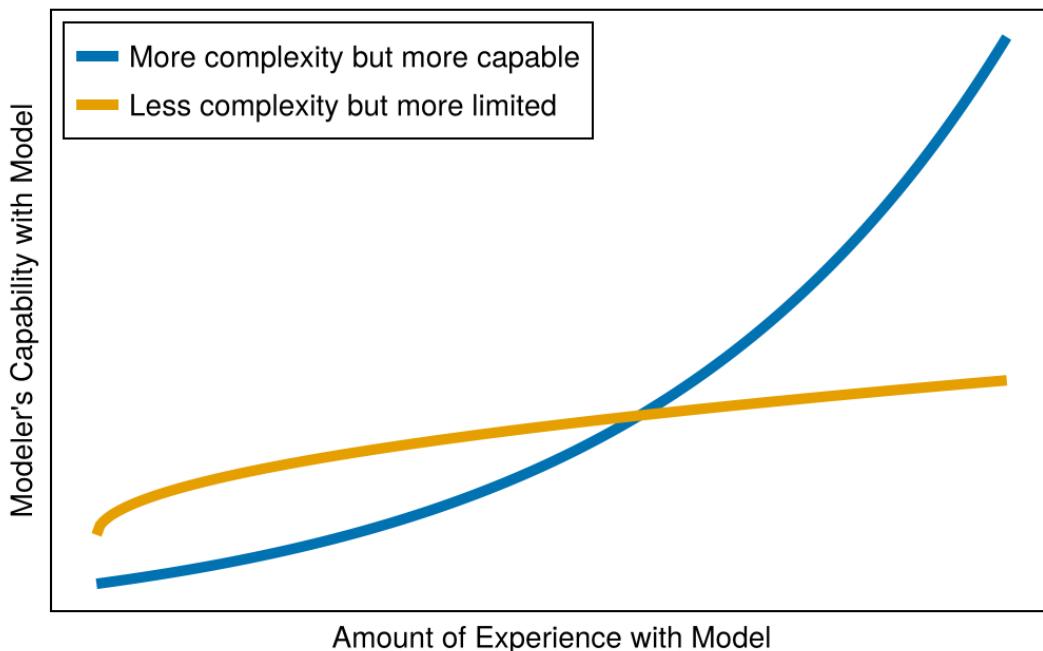


Figure 3.2.: Tradeoff between complexity and capability

3. Elements of Financial Modeling

3.6.3. Performance

Financial models are generally not used for their awe-inspiring beauty - users are results oriented and the faster a model returns the requested results, the better. Aside from direct computational costs such as server runtime, a shorter model runtime means that one can iterate faster, test new ideas on the fly, and stay focused on the problem at hand.

Many readers may be familiar with the cadence of (1) try running model overnight, (2) see results failed in the morning, (3) spend day developing, (4) repeat step 1. It is preferred if this cycle can be measured in minutes instead of hours or days.

Of course, requirements must be considered here too: needs for high frequency trading, daily portfolio rebalancing, and quarterly valuations are different when it comes to performance.

3.6.4. Separation of Model Logic and Data

When business logic is embedded within data, or data inputs are spread out across multiple locations it's tough to keep track of things. Using a spreadsheet as an example, often times it's incredibly difficult to ascertain a model's operation if inputs are spread out across locations on many tabs. Or if related calculations are performed in multiple locations, or if it's not clear where the line is drawn between calculations performed in the worksheets or in macros.

3.6.5. Abstraction of Modeled Systems

At different times we are interested in different **ladder of abstraction**: sometimes we are interested in the small details, but other times we are interested in understanding the behavior of systems at a higher level.

Say we are an insurance company with a portfolio of fixed income assets supporting long term insurance liabilities. We might delineate different levels of abstraction like so:

Table 3.2.: An example of the different levels of abstraction when thinking about modeling an insurance company's assets and liabilities.

Item	
More Abstract	Sensitivity of an entire company's solvency position
	Sensitivity of a portfolio of assets
More granular	Behavior over time of an individual contract
	Mechanics of an individual bond or insurance policy

3.7. What makes a good modeler?

At different times, we are often interested in different aspects of a problem. In general, you start to be able to obtain more insights and a greater understanding of the system when you move up the ladder of abstraction.

In fact, a lot of designing a model is essentially trying to figure out where to put the right abstractions. What is the right level of detail to model this in and what is the right level of detail to expose to other systems?

Let us also distinguish between **vertical abstraction**, as described above, and **horizontal abstraction** which will refer to encapsulating different properties, or mechanics of components of model that effectively exist on the same level of vertical abstraction. For example, both asset and liability mechanics sit at the most granular level in Table 3.2, But it may make sense in our model to separate the data and behavior from each other. If we were to do that, that would be an example of creating horizontal abstraction in service of our overall modeling goals.

3.7. What makes a good modeler?

A model is nothing without its operator, and a skilled practitioner is worth their weight in gold. What elements separate a good modeler from a mediocre modeler?

3.7.1. Domain Expertise

An expert who knows enough about all of the domains that are applicable is crucial. Imagine if someone said let's emulate an architect by having a construction worker and an artist work together. It's all too common for business to attempt to pair a business expert with an information technologist in the same way.

Unfortunately, this means that there's generally no easy way out of learning enough about finance, actuarial science, computers, and/or programming in order to be an effective modeler.

Also, a word of warning for the financial analysts out there: the computer scientists may find it easier to learn applied financial modeling than the other way around since the tools, techniques, and language of problem solving is already more a more general and flexible skill-set. There's more technologists starting banks than there are financiers starting technology companies.

3. Elements of Financial Modeling

3.7.2. Model Theory

If it is granted that financial modeling must involve, as the essential part, a building up of modeler's knowledge, the next issue is to characterize that knowledge more explicitly. The modeler's knowledge should be regarded as a theory, in the sense of Ryle's³ "Concept of the Mind." Very briefly: a person who has or possesses a theory in this sense knows how to do certain things and in addition can support the actual doing with explanations, justifications, and answers to queries, about the model and its results⁴.

A financial model is rarely left in a final state. Regulatory changes, additional mechanics, sensitivity testing, market dynamics, new products, and new systems to interact with force a model to undergo change and development through its entire life. And like a living thing, it must have nurturing caregivers. This metaphor sounds extended, but Naur's point is that unless the model also lives in the heads of its developers then it cannot successfully be maintained through time:

The conclusion seems inescapable that at least with certain kinds of large programs, the continued adaption, modification, and correction of errors in them, is essentially dependent on a certain kind of knowledge possessed by a group of programmers who are closely and continuously connected with them.

Assume that we need to adapt the model to fit a new product. One possessing a high degree of model theory includes:

- the ability to describe the trade-offs between alternate approaches that would accomplish the desired change
- relate the proposed change to the design of the current system and any challenges that will arise as a result of prior design decisions
- provide a quantitative estimation for the impact the change will have: runtime, risk metrics, valuation changes, etc.
- Analogize how the system works to themselves or to others
- Describe key limitations that the model has and where it is most divorced from the reality it seeks to represent.

Abstractions and analogies of the system are a critical aspect of model theory, as the human mind cannot retain perfectly precise detail about how the system works in each sub-component. The ability to, at some times, collapse and compartmentalize parts of

³Ryle, G. *The Concept of Mind*. Harmondsworth, England, Penguin, 1963, first published 1949. Applying "Theory Building"

⁴The idea of "model theory" is adapted from Peter Naur's 1985 essay, "Programming as Theory Building". Indeed, this whole paragraph is only a slightly modified version of Naur's description of theory in the programming context.

3.7. What makes a good modeler?

the model to limit the mental overload while at others recall important implementation details requires training - and is enhanced by learning concepts like those which will be covered in this book.

An example of how the right abstractions (and language describing those abstractions) can be helpful in simplifying the mental load:

Instead of:

The valuation process starts by reading an extract into three tabs of the spreadsheet. A macro loops through the list of policies on the first tab and in column C it gives the name of the applicable statutory valuation ruleset. The ruleset is defined as the combination of (1) the logic in the macro in the "Valuation" VBA module with, (2) the underlying rate tables from the tabs named XXX to ZZZ, along with (3) the additional policy level detail on the second tab. The valuation projection is then run with the current policy values taken from the third tab of the spreadsheet and the resulting reserve (equal to the actuarial present value of claims) is saved and recorded in column J of the first tab. Finally, a pivot table is used to sum up the reserves by different groups.

We could instead design the process so that the following could be said instead:

Policy extracts are parsed into a Policy datatype which contains a subtype ValuationKind indicating the applicable statutory ruleset to apply. From there, we map the valuation function over the set of Policies and perform an additive reduce to determine the total reserve.

There are terminologies and concepts in the second example which we will develop over the course of this section of the book - we don't want to dwell on the details bright now. However, we do want to emphasize that the process itself being able to condensed down to descriptions that are much more meaningful to the understanding of the model is a key differentiator for a code-based model instead of spreadsheets. It is not exaggerating that we could develop a handful of compartmentalized logics such that our primary valuation process described above could look like this in real code:

```
policies = parse(Policy, CSV.File("extract.csv"))
reserve = mapreduce(+, value, policies)
```

We've abstracted the mechanistic workings of the model into concise and meaningful symbols that not only perform the desired calculations but also make it obvious to an informed but unfamiliar reader what it's doing.

`parse`, `mapreduce`, `+`, `value`, `Policy` are all imbued with meaning - the first three would be understood by any computer scientist by the nature of their training (and is training that this book covers). The last two are unique to our model and have "real world" meaning that our domain expert modeler would understand which analogizes very directly to the way we would suggest implementing the details of `value` or `Policy`. The benefit of this, again, is to provide tools and concepts which let us more easily develop model theory.

3. Elements of Financial Modeling

3.7.3. Curiosity

A model never answers all of the questions and many times find itself overdrawn: sometimes more questions arise than answers provided. It is our experience that you modeler who continues to pursue questions that arise as a result of the analysis and in particular possesses an Insatiable itch for resolving apparent contradictions in model conclusions. That is, if an incomplete understanding or an incorrect model allows one to arrive at contradictory conclusions it's suggest that a deeper understanding or model revision is required.

3.7.4. Rigor

When developing a model it's important to ensure that assumptions and parameters are very clear, the methodology is in line with establish theory, inappropriate thought has been given to how the model will be used. Additionally one should be mindful of standards of practice. For example, professional actuarial societies have a long list of Actuarial Standards of Practice ("ASOPs"), some of which apply to modeling and the use of data that models ultimately rely on.

3.7.5. Clarity

A rigorous understanding of the fundamentals is important as it is all too easy to let imprecise communication and terminology interfere with the task at hand. Many terms in finance are overloaded with multiple meanings depending on the context such as the speakers background or company norms. When there is a term that is prone to misunderstanding because of its multiple overloaded meanings, a practitioner should take care to use that term And convey which definition is intended either explicitly or through the appropriate context clues.

3.7.6. Humble

Irreducible & epistemic/reducible uncertainty...

3.7.7. Architecture

Any sufficiently complex project benefits from architectural thinking

3.7. What makes a good modeler?

3.7.8. Planning

When tackling a large problem, it helps

3.7.9. Toolset

An experience professional is aware of a number of approaches that can be used in solving a problem. From heuristics that are able to be calculated on a napkin to complex economic models, the ability to draw on a wide tool set allows a practitioner to find the right solution for a given problem. Further, it is the intention of this book to enumerate a number of additional approaches that may prove useful in practice.

Part II.

Conceptual Foundations:
Programming and Abstractions

4. Elements of Programming

“Programming is not about typing, it’s about thinking.” — Rich Hickey
(2011)

4.1. In this section

Start building up computer science concepts by introducing tangible programming essentials. Data types, variables, control flow, functions, and scope are introduced.

On Your First Readthrough

This chapter is intended to be an introductory reference for most of the basic building blocks which we will build abstractions on top of in chapters that follow. We want this chapter to essentially be an easy and mildly opinionated stepping-stone on your journey.

At some point, you will likely find yourself seeking more precise or thorough documentation and will begin directly searching or reading the documentation of a language or library itself. However, it may be intimidating or frustrating reading reference documentation due to the density and terminology - let this chapter (and book writ large) be a bridge for you.

If reading this book in a linear fashion and new to programming, we suggest skipping the following sections and returning when encountering the concept or term later in the book:

- Section 4.5.4 through Section 4.5.9 which covers advanced and custom data types
- After Section 4.6.3 which deals with advanced function usage and program organization via scope

4.2. Computer Science, Programming, and Coding

Computer Science is the study of computing and information. As a science, it is distinct from programming languages which are merely coarse implementations of specific com-

4. Elements of Programming

puter science concepts¹. Programming (or “coding”) is the art and science of writing code in programming languages to have the computer perform desired tasks. While this may sound mechanistic, programming truly is one of the highest forms of abstract thinking and the design space of potential solutions is so large and potentially complex that much art and experience is needed to create a well-made program.

The language of computer science also provides a lexicon so that financial practitioners can discuss model architecture and problem characteristics. Having the language to describe a concept will also help see aspects of the problem in new ways, opening one up to more innovative solutions.

In the context of this financial modeling that we do, we can consider a financial model to be a type of computer program. It takes as input abstract information (data), performs calculations (an algorithm), and returns new data as an output. In this context, we generally do not need to consider many things that a software engineer may contemplate such as a graphical user interface, networking, or access restrictions. But there are many similarities: a good financial modeler must understand data types, algorithms, and some hardware details.

We will build up the concepts over this and the following chapter:

- This chapter will provide a survey of important concepts in computer science that will prove useful for our financial modeling. First, we will talk about data types, boolean logic, and basic expressions. We’ll build on those to discuss algorithms (functions) which perform useful work and use control flow and recursion.
- In the following chapters about abstraction, we will step back and discuss higher level concepts: the “schools of thought” around organizing the relationship between data and functions (functional versus object-oriented programming), design patterns, computational complexity, and compilation.

Tip

There will be brief references to hardware considerations for completeness, but hardware knowledge is not necessary to understand most programming languages (including Julia). It’s impossible to completely avoid talking about hardware when you care about the performance of your code, so feel free to gloss over the reference to hardware details on the first read and come back later after Chapter 8.

It’s highly recommended that you follow along and have a Julia session open (e.g. a REPL or a notebook) when first going through this chapter. See Chapter 30 if you haven’t gotten that set up yet. Follow along with the examples as we go.

¹Said differently, computer science may contemplate ideas and abstractions more generally than a specific implementation, as in mathematics where a theorem may be proved ($a^2 + b^2 = c^2$) without resorting to specific numeric examples ($3^2 + 4^2 = 5^2$).

 Tip

You can get some help in the REPL by typing a ? followed by the symbol you want help with, for example:

```
help?> sum
search: sum sum! summary cumsum cumsum! ...
sum(f, itr; [init])
```

Sum the results of calling function f on each element of itr.

... More text truncated...

 Caution

This introductory chapter is intended to provide a survey of the important concepts and building blocks, not to be a complete reference. For full details on available functions, more complete definitions, and a more complete tour of all language features, see the Manual at docs.julialang.org.

4.3. Expressions and Control Flow

4.3.1. Assignment and Variables

One of the first things it will be convenient to understand is the concept of variables. In virtually every programming language, we can assign values to make our program more organized and meaningful to the human reader. In the following example, we assign values to intermediate symbols to benefit us humans as we convert (silly!) American distance units:

```
feet_per_yard = 3
yards_per_mile = 1760

feet = 3000
miles = feet / feet_per_yard / yards_per_mile
```

0.56818181818182

4. Elements of Programming

Beyond readability, variables are a form of **abstraction** which allows us to think beyond specific instances of data and numbers to a more general representation. For example, the last line in the prior code example is a very generic computation of a unit conversion relationship and feet could be any number and the expression remains a valid calculation.

We will dive a little bit deeper into variables and assignment in Section 4.4.3.

4.4. Expressions

Having already seen some more illustrative examples above, we can zoom in onto smaller pieces called **expressions** which are effectively the basic block of code that gets evaluated. Here is an expression that adds two integers together that evaluate to a new integer (3 in this case):

```
1 + 2
```

```
3
```

A bigger program is built up of many of these smaller bits of code.

4.4.1. Compound Expression

There's two kinds of blocks where we can ensure that sub-expressions get evaluated in order and return the last expression as the overall return value: `begin` and `let` blocks.

```
c = begin
  a = 3
  b = 4
  a + b
end
```

```
a, b, c
```

```
(3, 4, 7)
```

The variables inside the `begin` block are evaluated in the same scope as `c` and therefore have the assigned values when we call `a` and `b` in the last line. Contrast that with the `let` block below, where `d` and `e` are not available when we try to get the value of `f`. This is because `let` creates a new inner scope that's not available in `f`'s scope. More on scope later in the chapter.

```
f = let
    d = 1
    e = 2
    d + e
end
f

3

d

LoadError: UndefVarError: `d` not defined
UndefVarError: `d` not defined
```

4.4.2. Conditional Expressions

Conditionals are expressions that evaluate to a **boolean** true or false. This is the beginning of really being able to assemble complex logic to perform useful work. Here are a handful expressions that would evaluate to true:

```
1 > 0
1 == 1 # check for equality
Float64 isa Rational
(5 > 0) & (-1 < 2) # "and" expression
(5 > 0) | (-1 > 2) # "or" expression
1 != 2
```

i Note

In Julia, the booleans have an integer equality: `true` is equal to 1 (`true == 1`) and `false` is equal to 0 (`false == 0`). However:

- `true != 5`. Only 1 is equal to true (in some languages, any non-zero number is “truthy”).
- `true` is not equal to 1 (`egal` is defined later in this chapter).

Conditionals can be used to assemble different logical paths for the program to follow and the general pattern is an `if` block:

4. Elements of Programming

```
if condition
    # do one thing
elseif condition
    # do something else
else
    # do something if none of the
    # other conditions are met
end
```

A complete example:

```
function buy_or_sell(my_value, market_price)
    if my_value > market_price
        "buy more"
    elseif my_value < market_price
        "sell"
    else
        "hold"
    end
end

buy_or_sell(10, 15), buy_or_sell(15, 10), buy_or_sell(10, 10)

("sell", "buy more", "hold")
```

4.4.2.1. Equality

The “Ship of Theseus²” problem is an example of how equality can be philosophically complex concept. In computer science we have the advantage that while we may not be able to resolve what’s the “right” type of equality, we can be more precise about it.

Here is an example for which we can see the difference between two types of equality:

- **Egal** equality is when a program could not distinguish between two objects at all
- **Equal** equality is when the values of two objects are the same

If two things are egal, then they are also equal.

In the following example, `s` and `t` are equal but not egal:

²The Ship of Theseus problem specifically refers to a legendary ancient Greek ship, owned by the hero Theseus. The paradox arises from the scenario where, over time, each wooden part of the ship is replaced with identical materials, leading to the question of whether the fully restored ship is still the same ship as the original. The Ship of Theseus problem is a thought experiment in philosophy that explores the nature of identity and change. It questions whether an object that has had all of its components replaced remains fundamentally the same object.

4.4. Expressions

```
s = [1, 2, 3]
t = [1, 2, 3]
s == t, s === t
```

```
(true, false)
```

One way to think about this is that while the values are equal, there is a way that one of the arrays could be made not equal to the other:

```
t[2] = 5
t
```

```
3-element Vector{Int64}:
1
5
3
```

Now t is no longer equal to s:

```
s == t
false
```

The reason this happens is that arrays are containers that can have their contents modified. Even though they originally had the same values, s and t are different containers, and *it just so happened* that the values they contained started out the same.

Some data can't be modified, including some kinds of collections. Immutable types like the following tuple, with the same stored values, are equal because there is no way for us to make them different:

```
(2, 4) === (2, 4)
true
```

Using this terminology, we could now interpret the "Ship of Theseus" as that his ship is "equal" but not "equal".

4. Elements of Programming

4.4.3. Assignment and Variables

When we say `x = 2` we are **assigning** the integer value of 2 to the variable `x`. This is an expression that lets us bind a something to a variable so that it can be referenced more concisely or in different parts of our code. When we re-assign the variable we are not mutating the value: `x = 3` does not change the 2.

When we have a mutable object (e.g. an Array or a `mutable struct`), we can mutate the value inside the referenced container. For example:

```
x = [1, 2, 3]                                ①  
x[1] = 5                                     ②  
x
```

- ① `x` refers to the array which currently contains the elements 1, 2, and 3.
- ② We re-assign the first element of the array to be the value 5 instead of 1

```
3-element Vector{Int64}:  
5  
2  
3
```

In the above example, `x` has not been reassigned. It is possible for two variables to refer to the same object:

```
x = [1, 2, 3]  
y = x  
x[1] = 6  
y
```

- ① `y` refers to the *same* underlying array as `x`

```
3-element Vector{Int64}:  
6  
2  
3
```

Note that variables can be re-assigned unless they are marked as `const`:

```
const PHI = π * 2 # <1>
```

- ① Capitalizing constant variables is a convention in Julia.

If we tried to re-assign `PHI`, we would get an error.

 Warning

Note that if we declare a `const` variable that refers to a mutable container like an array, the container can still be mutated. It's the reference to the container that remains constant, not necessarily the elements within the container.

4.4.4. Loops

Loops are ways for the program to move through a program and repeat expressions while we want it to. There are two primary loops: `for` and `while`.

for loops are loops that iterate over a defined range or set of values. Let's assume that we have the array `v = [6,7,8]`. Here are multiple examples of using a `for` loop in order to print each value to output (`println`):

```
# use fixed indices
for i in 1:3
    println(v[i])
end

# use indices the of the array
for i in eachindex(v)
    println(v[i])
end

# use the elements of the array
for x in v
    println(x)
end

# use the elements of the array
for x ∈ v           # ∈ is typed \in<tab>
    println(x)
end
```

while loops will run repeatedly until an expression is false. Here's some examples of printing each value of `v` again:

```
# index the array
i = 1
while i <= length(v)
    println(v[i])
    global i += 1
end
```

(1)

4. Elements of Programming

- ① `global` is used to increment `i` by 1. `i` is defined outside the scope of the `while` loop (see Section 4.7).

```
# index the array
i = 1
while true
    println(v[i])
    if i >= length(v)
        break
    end
    global i += 1
end
```

(1)

- ① `break` is used to terminate the loop manually, since the condition that follows the `while` will never be false.

4.4.5. Performance of loops

Loops are highly performant in Julia and often the fastest way to accomplish things. Those coming from Python or R may have developed a habit to avoid writing loops. *Fear the for loop not!*

4.5. Data Types

Data types are a way of categorizing information by intrinsic characteristics. We instinctively know that `13.24` is different than "this set of words" and types are how we will formalize this distinction. This is a key conceptual point, and mathematically it's like we have different sets of objects to perform specialized operations on. Beyond this set-like abstraction is implementation details related to computer hardware. You probably know that computers only natively "speak" in binary zeros and ones. Data types are a primary way that a computer can understand if it should interpret `01000010` as `B` or as `66`³.

Each 0 or 1 within a computer is called a **bit** and eight bits in a row form a **byte** (such as `01000010`). This is where we get terms like "gigabytes" or "kilobits per second" as a measure of the quantity or rate of bits something can handle⁴.

³This binary representations correspond to `B` and `66` with the *ASCII character set* and 8-bit integer encodings respectively, discussed later in this chapter.

⁴Some distinctions you may encounter: in short-form, "kb" means kilobits while the upper-case "B" in "kB" means kilobytes. Also confusingly, sometimes the "k" can be binary or decimal - because computers speak in binary, a binary "k" means 1024 (equal to 2^{10}) instead of the usual decimal 1000. In most computer contexts, the binary (multiples of 1024) is more common.

4.5.1. Numbers

Numbers are usually grouped into two categories: **integers** and **floating-point**⁵ numbers. Integers are like the mathematical set of integers while floating-point is a way of representing decimal numbers. Both have some limitations since computers can only natively represent a finite set of numbers due to the hardware (more on this in Chapter 8). Here are three integers that are input into the **REPL** (Read-Eval-Print-Loop)⁶ and the result is **printed** below the input:

2

2

423

423

1929234

1929234

And three floating-point numbers:

0.2

0.2

-23.3421

-23.3421
14e3 # the same as 14,000.0

14000.0

⁵The term floating point refers to the fact that the number's radix (decimal) point can “float” between the significant digits of the number.

⁶That is, it *reads* the code input from the user, *evaluates* what code was given to it, *prints* the result of the input to the screen, and *loops* through the process again.

4. Elements of Programming

On most systems, `0.2` will be interpreted as a 64-bit floating point type called `Float64` in Julia since most architectures these days are 64-bit⁷, while on a 32-bit system `0.2` would be interpreted as a `Float32`. Given that there are a finite amount of bits attempting to represent a continuous, infinite set of numbers means that some numbers are not able to be represented with perfect precision. For example, if we ask for `0.2`, the closest representations in 64 and 32 bit are:

- `0.20000000298023223876953125` in 32-bit
- `0.200000000000000011102230246251565404236316680908203125` in 64-bit

This leads to special considerations that computers take when performing calculations on floating point maths, some of which will be covered in more detail in Chapter 8. For now, just note that floating point numbers have limited precision and even if we input `0.2`, your computations will use the above decimal representations even if it will print out a number with fewer digits shown:

`x = 0.2` (1)

`big(x)` (2)

- ① Here, we **assign** the value `0.2` to a **variable** `x`. More on variables/assignments in Section 4.4.3.
- ② `big(x)` is a arbitrary precision floating point number and by default prints the full precision that was embedded in our variable `x`, which was originally `Float64`.

`0.200000000000000011102230246251565404236316680908203125`

i Note

Note the difference in what printed between the last example and when we input `0.2` earlier in the chapter. The former had the same (not-exactly equal to `0.2`) *value*, but it printed an abbreviated set of digits as a nicety for the user, who usually doesn't want to look at floating point numbers with their full machine precision. The system has the full precision (`0.20...3125`) but is truncating the output. In the last example, we've converted the normal `Float64` to a `BigFloat` which will not truncate the output when printing.

Integers are similarly represented as 32 or 64 bits (with `Int32` and `Int64`) and are limited to exact precision:

- -32,767 to 32,767 for `Int32`
- -2,147,483,647 to 2,147,483,647 for `Int64`

⁷This means that their central processing units (CPUs) use instructions that are 64 bits long.

Additional range in the positive direction if one chooses to use “unsigned”, non-negative numbers (`UInt32` and `UInt64`). Unlike floating point numbers, the integers have a type `Int` which will use the system bit architecture by default (that is, `Int(30)` will create a 64 bit integer on 64-bit systems and 32-bit on 32-bit systems).

Financial Modeling Pro-tip

Excel’s numeric storage and routine is complex and not quite the same as most programming languages, which follow the Institute of Electrical and Electronics Engineer’s standards (such as the IEEE 754 standard for double precision floating point numbers). Excel uses IEEE for the *computations* but results (and therefore the cells that comprise many calculations interim values) are stored with 15 significant digits of information. In some ways this is the worst of both worlds: having the sometimes unusual (but well-defined) behavior of floating point arithmetic *and* having additional modifications to various steps of a calculation. In general, you can assume that the programming language result (following the IEEE 754 standard) is a better result because there are aspects to the IEEE 754 defines techniques to minimize issues that arise in floating point math. Some of the issues (round-off or truncation) can be amplified instead of minimized with Excel.

In practice, this means that it can be difficult to exactly replicate a calculation in Excel in a programming language and vice-versa. It’s best to try to validate a programming model versus Excel model using very small unit calculations (e.g. a single step or iteration of a routine) instead of an all-in result. You may need to define some tolerance threshold for comparison of a value that is the result of a long chain of calculation.

4.5.2. Type Hierarchy

We can describe a *hierarchy* of types. Both `Float64` and `Int64` are examples of `Real` numbers (here, `Real` is an **abstract** Julia type which corresponds to the mathematical set of real numbers commonly denoted with \mathbb{R}). Both `Float64` and `Int32` are `Real` numbers, so why not just define all numbers as a `Real` type? Because for performant calculations, the computer must know in advance how many bits each number is represented with.

Figure 4.1 shows the type hierarchy for most built-in Julia number types.

The integer and floating point types described in the prior section are known as **concrete** types because there are no possible sub types (child types). Further, a concrete type can be a **bit type** if the data type will always have the same number of bits in memory: a `Float32` will always be 32 bits in memory, for example. Contrast this with strings (described below) which can contain an arbitrary number of characters.

4. Elements of Programming

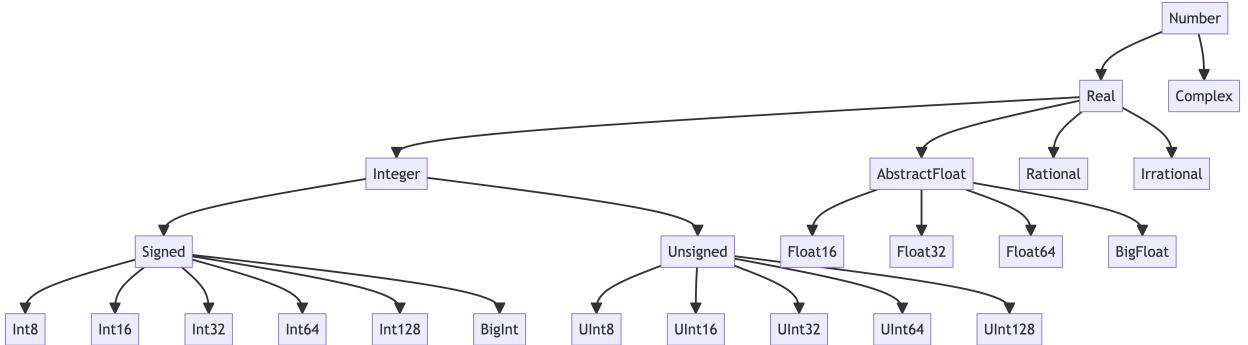


Figure 4.1.: Numeric Type Hierarchy in Julia. Leafs of the tree are concrete types.

4.5.3. Collections

Collections are types that are really useful for storing data which contains many elements. This section describes some of the most common and useful types of containers.

4.5.3.1. Arrays

Arrays are the most common way to represent a collection of similar data. For example, we can represent a set of integers as follows:

`[1, 10, 300]`

```
3-element Vector{Int64}:
 1
 10
 300
```

And a floating point array:

`[0.2, 1.3, 300.0]`

```
3-element Vector{Float64}:
 0.2
 1.3
 300.0
```

4.5. Data Types

Note the above two arrays are different types of arrays. The first is `Vector{Int64}` and the second is `Vector{Float64}`. These are arrays of concrete types and so Julia will know that each element of an array is the same amount of bits which will enable more efficient computations. With the following set of mixed numbers, Julia will **promote** the integers to floating point since the integers can be accurately represented⁸ in floating point.

```
[1, 1.3, 300.0, 21]
```

4-element `Vector{Float64}`:

```
1.0  
1.3  
300.0  
21.0
```

However, if we explicitly ask Julia to use a `Real`-typed array, the type is now `Vector{Real}`. Recall that `Real` is an abstract type. Having heterogeneous types within the array is conceptually fine, but in practice limits performance. Again, this will be covered in more detail in Chapter 8.

In Julia, arrays can be multi-dimensional. Here are two three-dimensional arrays with length three in each dimension:

```
rand(3, 3, 3)
```

3×3×3 `Array{Float64, 3}`:

```
[ :, :, 1 ] =  
0.518085 0.293179 0.252319  
0.865949 0.898116 0.897366  
0.982497 0.910202 0.39895  
  
[ :, :, 2 ] =  
0.48049 0.0231775 0.696707  
0.785279 0.965449 0.448145  
0.160946 0.307039 0.384671  
  
[ :, :, 3 ] =  
0.855621 0.155781 0.27221  
0.18449 0.371296 0.777745  
0.778503 0.735568 0.203806
```

```
[x + y + z for x in 1:3, y in 11:13, z in 21:23]
```

⁸Accurate only to a limited precision, as described in Section 4.5.1.

4. Elements of Programming

```
3x3x3 Array{Int64, 3}:
```

```
[:, :, 1] =  
33 34 35  
34 35 36  
35 36 37
```

```
[:, :, 2] =  
34 35 36  
35 36 37  
36 37 38
```

```
[:, :, 3] =  
35 36 37  
36 37 38  
37 38 39
```

The above example demonstrates **array comprehension** syntax which is a convenient way to create arrays in Julia.

A two-dimensional array has the rows by semi-colons (;):

```
x = [1 2 3; 4 5 6]
```

```
2x3 Matrix{Int64}:
```

```
1 2 3  
4 5 6
```

i Note

In Julia, a `Vector{Float64}` is simply a one-dimensional array of floating points and a `Matrix{Float64}` is a two-dimensional array. More precisely, they are **type aliases** of the more generic `Array{Float64,1}` and `Array{Float64,2}` names. Arrays with three or more dimensions don't have a type alias pre-defined.

4.5.3.2. Array indexing

Array elements are accessed with the integer position, starting at 1 for the first element⁹
¹⁰:

⁹Whether an index starts at 1 or 0 is sometimes debated. Zero-based indexing is natural in the context of low-level programming which deal with bits and positional *offsets* in computer memory. For higher level programming one-based indexing is more natural: in a set of data stored in an array, it is much more natural to reference the *first* (through n^{th}) datum instead of the *zeroth* (through $(n-1)^{th}$) datum.

¹⁰Arrays in Julia can actually be indexed with an arbitrary starting point: see the package `OffsetArrays.jl`

4.5. Data Types

```
v = [10, 20, 30, 40, 50]  
v[2]
```

20

We can also access a subset of the vector's contents by passing a range:

```
v[2:4]
```

```
3-element Vector{Int64}:  
20  
30  
40
```

And we can generically reference the array's contents, such as:

```
v[begin+1:end-1]
```

```
3-element Vector{Int64}:  
20  
30  
40
```

We can assign values into the array as well, as well as combine arrays and push new elements to the end:

```
v[2] = -1  
push!(v, 5)  
vcat(v, [1, 2, 3])
```

```
9-element Vector{Int64}:  
10  
-1  
30  
40  
50  
5  
1  
2  
3
```

4. Elements of Programming

4.5.3.3. Array Alignment

When you have an MxN matrix (M rows, N columns), a choice must be made as to which elements are next to each other in memory. Typical math convention and fundamental computer linear algebra libraries (dating back decades!) are column major and Julia follows that legacy. **Column major** means that elements going down the rows of a column are stored next to each other in memory. This is important to know so that (1) you remember that vectors are treated like a column vector when working with arrays (that is: a N element 1D vector is like a Nx1 matrix), and (2) when iterating through an array, it will be faster for the computer to access elements next to each other column-wise. A 10x10 matrix is actually stored in memory as 100 elements coming in order, one after another in single file.

This 3x4 matrix is stored with the elements of columns next to each other, which we can see with `vec`:

```
mat = [1 2 3; 4 5 6; 7 8 9]
```

```
3x3 Matrix{Int64}:
```

```
1 2 3  
4 5 6  
7 8 9
```

```
vec(mat)
```

```
9-element Vector{Int64}:
```

```
1  
4  
7  
2  
5  
8  
3  
6  
9
```

4.5.3.4. Ranges

A **range** is a representation of a range of numbers. We actually used them above to index into arrays. They are expressed as `start:stop`

We don't have to actually store all of these numbers on the computer somewhere as in an `Array`. Instead, this is an object that *represents* the ordered set of numbers. So for example, we can sum up 1 through the number of atoms on the earth instantaneously:

4.5. Data Types

This is possible due to two things:

1. not needing to actually store that many numbers in memory, and
 2. Julia being smart enough to apply the triangular number formula¹¹ when `sum` is given a consecutive range.

There are more general ways to construct ranges:

Step by another number instead of the default 1:

1:2:7

1:2:7

Specify the number of values within the range, inclusive of the first number¹²:

```
range(0, 10, 21)
```

0.0:0.5:10.0

4.5.3.5. Characters, Strings, and Symbols

Characters are represented in most programming languages as letters within quotation marks. In Julia, individual characters are represented using single quotes:

'a'

'a': ASCII/Unicode U+0061 (category Ll: Letter, lowercase)

¹¹The triangular numbers (sum of integers from 1 to n) are:

$$T_n = \sum_{k=1}^n k = 1 + 2 + \dots + n = \frac{n^2 + n}{2} = \frac{n(n+1)}{2} = \binom{n+1}{2}$$

¹²Whether the last number is in the resulting range depends on if the step evenly divides the end of the range.

4. Elements of Programming

Letters and other characters present more difficulties than numbers to represent within a computer (think of how many languages and alphabets exist!), and it essentially only works because the world at large has agreed to a given representation. Originally **ASCII** (American Standard Code for Information Interchange) was used to represent just 95 of the most common English characters ("a" through "z", zero through nine, etc.). Now, **UTF** (Unicode Transformation Format) can encode more than a million characters and symbols from many human languages.

Strings are a collection¹³ of characters, and can be created in Julia with double quotes:

```
"hello world"
```

```
"hello world"
```

It's easy to ascertain how 'normal' characters can be inserted into a string, but what about things like new lines or tabs? They are represented by their own characters but are normally not printed in computer output. However, those otherwise invisible characters do exist. For example, here we will use a **string literal** (indicated by the """") to tell Julia to interpret the string as given, including the invisible new line created by hitting return on the keyboard between the two words:

```
"""
hello
world
"""
```

```
"hello\nworld\n"
```

The output above shows the \n character contained within the string.

Symbols are a way of representing an identifier which cannot be seen as a collection of individual characters. :helloworld is distinct from "helloworld" - you can kind of think of the former as an un-executed bit of code - if we were to execute it (with eval(:helloworld)), we would get an error UndefVarError: 'a' not defined . Symbols can *look* like strings but do not behave like them. For now, it is best to not worry about symbols but it is an important aspect of Julia which allows the language to represent aspects of itself as data. This allows for powerful self-reference and self-modification of code but this is a more advanced topic generally out of scope of this book.

¹³Under the hood, strings are essentially a vector of characters but there are complexities with character encoding that don't allow a lossless conversion to individual characters of uniform bit length. This is for historical compatibility reasons and to avoid making most documents' file sizes larger than it needs to be.

4.5.3.6. Tuples

Tuples are a set of values that belong together and are denoted by values inside parenthesis and separated by a comma. An example might be x-y coordinates in 2 dimensional space:

```
x = 3
y = 4
p1 = (x, y)
```

(3, 4)

Tuple's values can be accessed like arrays:

```
p1[1]
```

3

Tuples fill a middle ground between scalar types and arrays in more ways than one:

- Tuples have no problem having heterogeneous types in the different slots.
- Tuples are **immutable**, meaning that you cannot overwrite the value in memory (an error will be thrown if we try to do `p[1] = 5`).
- It's generally expected that within an array, you would be able to apply the same operation to all the elements (e.g. square each element) or do something like sum all of the elements together which isn't generally case for a tuple.
- Tuples are generally stack allocated instead of being heap allocated like arrays¹⁴, meaning that a lot of times they can be faster than arrays.

4.5.3.6.1. Named Tuples

Named tuples provide a way to give each field within the tuple a specific name. For example, our x-y coordinate example above could become:

```
p2 = (x=3, y=4)
```

(x = 3, y = 4)

¹⁴What this means will be explained in Chapter 8 .

4. Elements of Programming

The benefit is that we can give more meaning to each field and access the values in a nicer way. Previously, we used `location[1]` to access the `x`-value, but with the new definition we can access it by name:

`p2.x`

3

4.5.3.7. Dictionaries

Dictionaries are a container which relates a **key** to an associated **value**. Kind of like how arrays relate an index to a value, but the difference is that a dictionary is (1) un-ordered and (2) the key doesn't have to be an integer.

Here's an example which relates a name to an age:

```
d = Dict(  
    "Joelle" => 10,  
    "Monica" => 84,  
    "Zaylee" => 39,  
)
```

```
Dict{String, Int64} with 3 entries:  
  "Monica" => 84  
  "Zaylee" => 39  
  "Joelle" => 10
```

Then we can look up an age given a name:

```
d["Zaylee"]
```

39

Dictionaries are super flexible data structures and can be used in many situations.

64

4.5.4. Parametric Types

We just saw how tuples can contain heterogeneous types of data inside a common container. Let's look at this a little bit closer by looking at the full type:

```
typeof(p1)
```

```
Tuple{Int64, Int64}
```

`location` is a `Tuple{Int64, Int64}` type, which means that its first and second elements are both `Int64`. Contrast this with:

```
typeof(("hello", 1.0))
```

```
Tuple{String, Float64}
```

These tuples are both of the form `Tuple{T,U}` where `T` and `U` are both types. Why does this matter? We and the compiler can distinguish between a `Tuple{Int64, Int64}` and a `Tuple{String, Float64}` which allows us to reason about things ("I can add the first element of tuple together only if both are numbers") and the compiler to optimize (sometimes it can know exactly how many bits in memory a tuple of a certain kind will need and be more efficient about memory use). Further, we will see how this can become a powerful force in writing appropriately abstracted code and more logically organize our entire program when we encounter "multiple dispatch" later on.

4.5.5. Types for things not there

`nothing` represents that there's nothing to be returned - for example if there's no solution to an optimization problem or if a function just doesn't have any value to return (such as in the case with input/output like `println`).

`missing` is to represent something *should* be there but it's not, as is all too common in real-world data. Julia natively supports `missing` and three-value logic, which is an extension of the two-value boolean (true/false) logic, to handle missing logical values:



Tip

`Missing` and `Nothing` are the *types* while `missing` and `nothing` are the *values* here¹⁵. This is analogous to `Float64` being a type and `2.0` being a value.

¹⁵ `Missing` and `Nothing` are instances of **singleton type**, which means that there is only a single value that either type can take on.

4. Elements of Programming

Table 4.1.: Three value logic with true, missing, and false.

		(a) Not logic	
NOT (!)	Value		
true	false		
missing	missing		
false	true		

(b) And logic			
AND (&)	true	missing	false
true	true	missing	false
missing	missing	missing	false
false	false	false	false

(c) Or Logic			
OR ()	true	missing	false
true	true	true	true
missing	true	missing	missing
false	true	missing	false

4.5.6. Union Types

When two types may arise in a context, **union types** are a way to represent that. For example, if we have a data feed and we know that it will produce *either* a `Float64` or a `Missing` type then we can say that the value for this is `Union{Float64, Missing}`. This is much better for the compiler (and our performance!) than saying that the type of this is Any.

4.5.7. Creating User Defined Types

We've talked about some built-in types but so much additional capabilities come from being able to define our own types. For example, taking the x-y-coordinate example from above, we could do the following instead of defining a tuple:

```
struct BasicPoint
    x::Int64
    y::Int64
end

p3 = BasicPoint(3, 4)
```

4.5. Data Types

```
BasicPoint(3, 4)
```

`BasicPoint` is a **composite type** because it is composed of elements of other types. Fields are accessed the same way as named tuples:

```
p3.x, p3.y
```

(1)

- ① Note that here, Julia will return a tuple instead of a single value due to the comma separated expressions.

```
(3, 4)
```

`structs` in Julia are immutable like tuples above.

But wait, didn't tuples let us mix types too via parametric types? Yes, and we can do the same with our type!

```
struct Point{T}
    x::T
    y::T
end
```

Line 1 The `{T}` after the type's name allows for different Points to be created depending on what the type of the underlying `x` and `y` is.

Here's two new points which now have different types:

```
p4 = Point(1, 4)
p5 = Point(2.0, 3.0)
```

```
p4, p5
```

```
(Point{Int64}(1, 4), Point{Float64}(2.0, 3.0))
```

Note that the types are not equal because they have different type parameters!

```
typeof(p4), typeof(p5), typeof(p4) == typeof(p5)
```

```
(Point{Int64}, Point{Float64}, false)
```

But both are now subtypes of `PPoint2D`. The expression `X isa Y` is true when `X` is a (sub)type of `Y`:

```
p4 isa Point, p5 isa Point
```

4. Elements of Programming

```
(true, true)
```

Note though, that the `x` and `y` are both of the same type in each `PPoint2D` that we created. If instead we wanted to allow the coordinates to be of different types, then we could have defined `PPoint2D` as follows:

```
struct Point{T,U}
    x :: T
    y :: U
end
```

i Note

Can we define the `structs` above without indicating a (parametric) type? Yes!

```
struct Point
    x # no type here!
    y # no type declared here either!
end
```

But! `x` and `y` will both be allowed to be `Any`, which is the fallback type where Julia says that it doesn't know any more about the type until runtime (the time at which our program encounters the data when running). This means that the compiler (and us!) can't reason about or optimize the code as effectively as when the types are explicit or parametric. This is an example of how Julia can provide a nice learning curve - don't worry about the types until you start to get more sophisticated about the program design or need to extract more performance from the code.

The above `structs` that we have defined are examples of **concrete types** types which hold data. **Abstract types** don't directly hold data themselves but are used to define a hierarchy of types which we will later exploit (Chapter 7) to implement custom behavior depending on what type our data is.

Here's an example of (1) defining a set of related types that sits above our `Point2D`:

```
abstract type Coordinate end
abstract type CartesianCoordinate <: Coordinate end
abstract type PolarCoordinate <: Coordinate end

struct Point2D{T} <: CartesianCoordinate
    x :: T
    y :: T
end
```

```

struct Point3D{T} <: CartesianCoordinate
    x :: T
    y :: T
    z :: T
end

struct Polar2D{T} <: PolarCoordinate
    r :: T
    θ :: T
end

```

💡 Unicode Characters

Julia has wonderful Unicode support, meaning that it's not a problem to include characters like θ . The character can be typed in Julia editors by entering \theta and then pressing the TAB key on the keyboard.

Unicode is helpful for following conventions that you may be used to in math. For example, the math formula $\text{circumference}(r) = 2 \times r \times \pi$ can be written in Julia with `circumference(r) = 2 * r * π`.

The name for the characters follows the same for LaTeX, so you can search the internet for, e.g. "theta LaTeX" to find the appropriate name. Furhter, you can use the REPL help mode to find out how to enter a character if you can copy and paste it from somewhere:

```
help?> θ
"θ" can be typed by \theta<tab>
```

4.5.8. Mutable structs

It is possible to define structs where the data can be modified - such a data field is said to be **mutable** because it can be changed or mutated. Here's an example of what it would look like if we made Point2D mutable:

```

mutable struct Point2D{T}
    x :: T
    y :: T
end

```

You may find that this more naturally represents what you are trying to do. However, recall that an advantage of an immutable datatype is that costly memory doesn't necessarily have to be allocated for it. So you may think that you're being more efficient by re-using the same object... but it may not actually be faster. Again, more will be revealed in Chapter 8.

4. Elements of Programming

💡 Financial Modeling Pro-tip

Generally you should default to using immutable types and consider only moving to mutable types in specific circumstances. You'll see some examples in the applications later in the book.

4.5.9. Constructors

Constructors are functions that return a data type (functions will be covered more generally later in the chapter). When we declare a `struct`, an implicit function is defined that takes a tuple of arguments and returns the data type that was declared. In the following example, after we define `MyType` the `struct`, Julia creates a function (also called `MyType`) which takes two arguments and will return the datatype `MyType`:

```
struct MyDate
    year::Int
    month::Int
    day ::Int
end

methods(MyDate)

# 2 methods for type constructor:
[1] MyDate(year::Int64, month::Int64, day::Int64)
    @ In[54]:2
[2] MyDate(year, month, day)
    @ In[54]:2
```

Implicit constructors are nice in that you don't have to define a default method and the language does it for you. Sometimes there's reasons to want to control how an object is created, either for convenience or to enforce certain restrictions.

We can use an inner constructor (i.e. inside the `struct` block) to enforce restrictions:

```
struct MyDate
    year::Int
    month::Int
    day ::Int

    function MyDate(y,m,d)
        if ~(m in 1:12)
```

```

        error("month is not between 1 and 12")
else if ~ (d in 1:31)
    error("day is not between 1 and 31")
else
    return new(y,m,d)
end

end

```

And outer constructors are simply functions defined that have the same name as the data type , but are not defined inside the struct block. Extending the MyDate example, say we want to provide a default constructor for if no day is given such that the date returns the 1st of the month:

```

function MyDate(y,m)
    return MyDate(y,m,1)
end

```

4.6. Functions

Functions are a set of expressions that take inputs and return specified outputs.

4.6.1. Special Operators

Operators are the glue of expressions which combine values. We've already seen quite a few, but let's develop a little bit of terminology for these functions.

Unary operators are operators which only take a single argument. Examples include the ! which negates a boolean value or - which negates a number:

```
!true, -5
```

```
(false, -5)
```

Binary operators take two arguments and are some of the most common functions we encounter, such as + or - or >:

```
1 + 2, 1 - 2, 1 > 2
```

```
(3, -1, false)
```

4. Elements of Programming

The above unary and binary operators are special kinds of functions which don't require the use of parenthesis. However, they can be written with parathesis for greater clarity:

```
!(true), -(5), +(1, 2), -(1, 2)  
(false, -5, 3, -1)
```

In Julia, we distinguish between **functions** which define behavior that maps a set of inputs to outputs. But a single function can adapt its behavior to the arguments themselves. We have just seen the function `-` be used in two different ways: negation and subtraction depending on whether it had one or two arguments given to it. In this way there is a conceptual hierarchy of functions that complements the hierarchy we have discussed in relation to types:

- `-` is the overall function
- `-(x)` is a unary function which negates its values, `-(x,y)` subtracts `y` from `x`
- Specific methods are then created for each combination of concrete types: `-(x::Float64)` is a different method than `-(x::Int)`

Methods are specific compiled versions of the function for specific types. This is important because at a hardware level, operations for different types (e.g. integers versus floating point) differ considerably. By optimizing for the specific types Julia is able to achieve nearly ideal performance without the same sacrifices of other dynamic languages. We will develop more with respect to methods when we talk about dispatch in Chapter 7.

4.6.2. Defining Functions

Funcitons more generally are defined like so:

```
function functionname(arguments)  
    # ... code that does things  
end
```

Here's an example which returns the difference between the highest and lowest values in a collection:

```
function value_range(collection)  
  
    hi = maximum(collection)  
    lo = minimum(collection)  
    return hi - lo  
end
```

- ① `return` is optional but recommended to convey to readers of the program where you expect your function to terminate and return a value.

4.6.3. Defining Methods on Types

Here's another example of a function which calculates the distance between a point and the origin:

```
function distance(point)
    return sqrt(point.x^2 + point.y^2) (1)
(2)
end
```

- ① A function block is declared with the name `distance` which takes a single argument called `point`
- ② We compute the distance formula for a point with `x` and `y` coordinates. The `return` value make explicit what value the function will output.

`distance` (generic function with 1 method)

i Note

An alternate, simpler function syntax for `distance` would be:

```
distance(point) = sqrt(point.x^2 + point.y^2)
```

However, we might at this point note a flaw in our function's definition if we think about the various Coordinates we defined earlier: our definition would currently only work for `Point2D`. For example, if we try a `Point3D` we will get the wrong answer:

`distance(Point3D(1, 1, 1))`

1.4142135623730951

The above value should be $\sqrt{3}$, or approximately 1.73205.

What we need to do is define a refined distance for each type, which we'll call `dist` to distinguish from the earlier definition.

```
"""
dist(point)

The euclidean distance of a point from the origin.
"""
```

4. Elements of Programming

```
dist(p::Point2D) = sqrt(p.x^2 + p.y^2)
dist(p::Point3D) = sqrt(p.x^2 + p.y^2 + p.z^2)
dist(p::Polar2D) = p.r
```

dist (generic function with 3 methods)

Now our result will be correct:

```
dist(Point3D(1, 1, 1))
```

1.7320508075688772

This is referred to **dispatching** on the argument types. Julia will look up to find the most specific method of a function for the given argument types, and falling back to a generic implementation if one is defined.

In Chapter 7 we will see how dispatch (single and multiple) can provide very nice abstractions to simplify the design of a model.

Docstrings (Documentation Strings)

Notice the strings preceding the definition of `dist`. In Julia, putting a string ("...") or string literal (""""...""") right above the definition will allow Julia to recognize the string as documentation and provide it to the user in help mode (Section 30.4.1) and/or have a documentation tool create a webpage or PDF documentation resource.

Defining Methods for Parametric Types

We learned that `Float64 <: Real` in the type hierarchy. However, note that `Tuple{Float64}` is not a sub-type of `Tuple{Real}`. This is called being **invariant** in type theory... but for our purposes this just practically means that when we define a method we need to specify that we want it to apply to all subtypes.

For example, `myfunction(x::Tuple{Real})` would *not* be called if `x` was a `Tuple{Float64}` because it's not a sub-type of `Tuple{Real}`. To act the way we want, would define the method with the signature of `myfunction(Tuple{<:Real})` or `myfunction{T}(Tuple{T})` where `{T<:Real}`.

4.6.4. Keyword Arguments

Keyword arguments are arguments that are passed to a function but do not use *position* to pass data to functions but instead used named arguments. In the following example, `filepath` is a **positional argument** while the two arguments after the semicolon (`;`) are keyword arguments.

```
function read_data(filepath; normalize_names, has_header_row)
    # ... function would be defined here
end
```

The function would need to be called and have the two keyword arguments specified:

```
read_data("results.csv"; normalize_names=true, has_header_row=false)
```

4.6.5. Default Arguments

We are able to define default arguments for both positional and keyword arguments via an assignment expression in the function signature. For example, we can make it so that the user need not specify all the options for each call. Modifying the prior example so that typical CSVs work with less customization from the user:

```
function read_data(filepath;
    normalize_names = true,
    has_header = false
)
```

This is a simplified example, but if you look at the documentation for most data import packages you'll see a lot of functionality defined via keyword arguments which have sensible defaults so that most of the time you need not worry about modifying them.

4.6.6. Anonymous Functions

Anonymous functions are functions that have no name and are used in contexts where the name does not matter. The syntax is `x → ...expression with x....`. As an example, say that we want to create a vector from another where each element is squared. `map` applies a function to each member of a given collection:

```
v = [4, 1, 5]
map(x → x^2, v)
```

(1)

(1) The `x → x^2` is the anonymous function in this example.

4. Elements of Programming

```
3-element Vector{Int64}:
16
1
25
```

They are often used when constructing something from another value, or defining a function within optimization or solving routines.

4.6.7. First Class Nature

Functions in many languages including Julia are **first class** which means that functions can be assigned and moved around like data variables.

In this example, we have a general approach to calculate the error of a modeled result compared to a known truth. In this context, there are different ways to measure error of the modeled result and we can simplify the implementation of loss by keeping the different kinds of error defined separately. Then, we can assign a function to a variable and use it as an argument to another function:

```
function square_error(guess, correct)
    (correct - guess)^2
end

function abs_error(guess, correct)
    abs(correct - guess)
end

# obs meaning "observations"
function loss(modeled_obs,
            actual_obs,
            loss_function
        )
    sum(
        loss_function.(modeled_obs, actual_obs)
    )
end

let
    a = loss([1, 5, 11], [1, 4, 9], square_error)          (2)
    b = loss([1, 5, 11], [1, 4, 9], abs_error)             (3)
    a, b
end
```

- ① `loss_function` is a variable that will refer to a function instead of data.
- ② Using a `let` block here is good practice to not have temporary variables `a` and `b` scattered around our workspace.
- ③ Using a function as an argument to another function is an example of functions being treated as “first class”.

(5, 3)

4.6.8. Broadcasting

Looking at the prior definition of `dist`, what if we wanted to compute the squared distance from the origin for a set of points? If those points are stored in an array, we can **broadcast** functions to all members of a collection at the same time. This is accomplished using the **dot-syntax** as follows:

```
points = [Point2D(1, 2), Point2D(3, 4), Point2D(6, 7)]
dist.(points) .^ 2
```

```
3-element Vector{Float64}:
 5.000000000000001
 25.0
 85.0
```

Let’s unpack that a bit more:

1. The `.` in `dist.(points)` tells Julia to apply the function `dist` to each element in `points`.
2. The `.` in `.^` tells Julia to square each values as well

Why broadcasting is useful:

1. Without needing any redefinition of functions we were able to transform the function `dist` and exponentiation (`^`) to work on a collection of data. This means that we can keep our code simpler and easier to reason about (operating on individual things is easier than adding logic to handle collections of things).
2. When multiple broadcasted operations are joined together, Julia can **fuse** the operations so that each operation is performed at the same time instead of each step sequentially. That is, if the operation were not fused, the computer would first calculate `dist` for each point, and then apply the square on the collection of distances. When it’s fused, the operations can happen at the same time without creating an interim set of values.

4. Elements of Programming

i Note

For readers coming from numpy-flavored Python or R, broadcasting is a way that can feel familiar to the array-oriented behavior of those two languages. Once you feel comfortable with Julia in general, you may find yourself relaxing and relying less on array-oriented design and instead picking whichever iteration paradigm feels most natural for the problem at hand: loops or broadcasting over arrays.

4.6.8.1. Broadcasting Rules

What happens if one of the collections is not the same size as the others? When broadcasting, singleton dimensions (i.e. the 1 in 1xN, “1-by-N”, dimensions) will be expanded automatically when it makes sense. For example, if you have a single element and a one dimensional array, the single element will be expanded in the function call without using any additional memory (if that dimension matches one of the dimensions of the other array).

The rules with an MxN and a PxQ array:

- either (M and P) or (N and Q) need to be the same, *and*
- one of the non-matching dimensions needs to be 1

Some examples might clarify. This 1x1 element is being combined with a 4x1, so there is a compatible dimension (N and Q match, M is 1):

```
2 .^ [0, 1, 2, 3]
```

```
4-element Vector{Int64}:
1
2
4
8
```

Here, this 1x3 works with the 2x3 (N and Q match, M is 1)

```
[1 2 3] .+ [1 2 3; 4 5 6]
```

```
2×3 Matrix{Int64}:
2 4 6
5 7 9
```

This 3x1 isn't compatible with this 2x3 array (neither M and P nor N and Q match)

```
#| error: true
[1, 2, 3] .+ [1 2 3; 4 5 6]
```

This 2x4 isn't compatible with the 2x3 (M and P match, but N nor Q is 1):

```
#| error: true
[1 2; 3 4] .+ [1 2 3; 4 5 6]
```

4.6.8.2. Not Broadcasting

What if you do not want the array to be used element-wise when broadcasting? Then you can wrap the array in a `Ref`, which is used in broadcasting to make the array be treated like a scalar. In the example below, `in(needle,haystack)` searches a collection (`haystack`) for an item (`needle`) and returns `true` or `false` if the item is in the collection:

```
in(4, [1 2 3; 4 5 6])
```

```
true
```

What if we had an array of things ("needles") that we wanted to search for? By default, broadcasting would effectively split the array up into collections of individual elements to search:

```
in.([1, 9], [1 2 3; 4 5 6])
```

2x3 BitMatrix:

```
1 0 0
0 0 0
```

Effectively, the result above is the result of this broadcasted result:

```
in(1, [1,2,3]) # the first row of the above result
in(9, [4,5,6])
```

If we were expecting Julia to return `[1,0]` (that the first needle is in the haystack but the second needle is not), then we need to tell Julia not to broadcast along the second array with `Ref`:

```
in.([1, 9], Ref([1 2 3; 4 5 6]))
```

2-element BitVector:

```
1
0
```

4. Elements of Programming

4.6.9. Passing by Sharing

We often want to share data between scopes, such as between modules or by passing something into a function's scope. Arguments to a function in Julia are **passed-by-sharing** which means that an outside variable can be mutated from within a function. We can modify the array in the outer scope (scope discussed later in this chapter) from within the function. In this example, we modify the array that is assigned to `v` by doubling each element:

```
v = [1, 2, 3]

function double!(v)
    for i in eachindex(v)
        v[1] = 2 * v[i]
    end
end

double!(v)

v
```

3-element Vector{Int64}:

6
2
3

💡 Tip

Convention in Julia is that a function that modifies its arguments has a `!` in its name and we follow this convention in `double!` above. Another example would be the built-in function `sort!` which will sort an array in-place without allocating a new array to store the sorted values.

We won't discuss all potential ways that programming languages can behave in this regard, but an alternative that one may have seen before (e.g. in Matlab) is pass-by-value where a modification to an argument only modifies the value within the scope. Here's how to replicate that in Julia by copying the value before handing it to a function. This time, `v` is not modified because we only passed a copy of the array and not the array itself:

```
v = [1, 2, 3]
double!(copy(v))
v
```

```
3-element Vector{Int64}:
1
2
3
```

4.7. Scope

In projects of even modest complexity, it can be challenging to come up with unique identifiers for different functions or variables. **Scope** refers to the bounds for which an identifier is available. We will often talk about the **local scope** that's inside some expression that creates a narrowly defined scope (such as a function or let or module block) or the **global scope** which is the top level scope that contains everything else inside of it. Here are a few examples to demonstrate scope.

```
i = 1
let
    j = 3
    i + j
end
```

- ① i is defined in the global scope and would be available to other inner scopes.
- ② The let ... end block creates a local scope which inherits the defined global scope definitions.
- ③ j is only defined in the local scope created by the let block.

4

In fact, if we try to use j outside of the scope defined above we will get an error:

j

```
LoadError: UndefVarError: `j` not defined
UndefVarError: `j` not defined
```

Tip

let blocks are a great way to organize your code in bite-sized chunks or to be able to re-use common variable names without worrying about conflict. Here's an example of using let blocks to:

1. Perform intermediate calculations without fear of returning a partially modified variable

4. Elements of Programming

2. Re-use common variable names

```
bonds = let
    df = CSV.read("bonds.csv", DataFrame)
    df.issuer = lookup_issuer(df.CUSIP)
    df
end

mortgages = let
    df = CSV.read("bonds.csv", DataFrame)
    df.issuer = lookup_issuer(df.CUSIP)
    df
end
```

If we were running this interactively (e.g. step-by step in VS Code, the REPL, or notebooks) then these two code blocks will run completely and will run separately. The short, descriptive name `df` is reused, but there's no chance of conflict. We also can't easily run the block of code (`let ... end`) and get a partially evaluated result (e.g. getting the dataframe before it has been appropriately modified to add the `issuer` column).

Here is an example with functions:

```
x = 2
base = 10
foo() = base^x
foo(x) = base^x
foo(x, base) = base^x
①
②
③

foo(), foo(4), foo(4, 4)
```

- ① Both `base` and `x` are inherited from the global scope.
- ② `x` is based on the local scope from the function's arguments and `base` is inherited from the global scope.
- ③ Both `base` and `x` are defined in the local scope via the function's arguments.

`(100, 10000, 256)`

In Julia, it's always best to explicitly pass arguments to functions rather than relying on them coming from an inherited scope. This is more straight-forward and easier to reason about and it also allows Julia to optimize the function to run faster because all relevant variables coming from outside the function are defined at the function's entry point (the arguments).

4.7.1. Modules and Namespaces

Modules are ways to encapsulate related functionality together. Another benefit is that the variables inside the module don't "pollute" the **namespace** of your current scope. Here's an example:

```
module Shape (1)

    struct Triangle{T}
        base::T
        height::T
    end

    function area(t::Triangle) (2)
        return 1 / 2 * t.base * t.height
    end
end

t = Shape.Triangle(4, 2) (3)
area = Shape.area(t) (4)
```

- ① module defines an encapsulated block of code which is anchored to the namespace Shape
- ② Here, area a *function* defined within the Shape module.
- ③ Outside of Shape module, we can access the definitions inside via the `Module.identifier` syntax.
- ④ Here, area is a *variable* in our global scope that *does not* conflict with the area defined within the Shape module. If `Shape.area` were not within a module then when we said `area = ...` we would have reassigned area to no longer refer to the function and instead would refer to the area of our triangle.

4.0

Note

Summarizing related terminology:

- A **module** is a block of code such as `module MySimulation ... end`
- A **package** is a module that has a specific set of files and associated metadata. Essentially, it's a module with a `Project.toml` file that has a name and unique identifier listed, and a file in a `src/` directory called `MySimulation.jl`
 - **Library** is just another name for a package, and the most common context this comes up is when talking about the packages that are bundled

4. Elements of Programming

with Julia itself called the **standard library** (`stdlib`).

5. Functional Abstractions

The purpose of abstraction is not to be vague, but to create a new semantic level in which one can be absolutely precise. - Edsger Dijkstra (1972)

5.1. In this section

Demonstrate different approaches to a problem which gradually introduce more reusable or general techniques. These techniques will allow for constructing sophisticated models while maintaining consistency and simplicity. Imperative programming, functional programming, and recursion.

5.2. Introduction

This chapter will center around a simple task: calculate the present value of a portfolio of a single fixed, risk-free, coupon-paying bonds under two different interest rate environments. The focus will be on describing different approaches to this problem, not be adding complexity to the problem (e.g. no getting into credit spreads, settlement timing, etc.).

```
cf_bond = [10, 10, 10, 10, 110];
rate = [0.05, 0.06, 0.05, 0.04, 0.05];
```

① The rates given are the one year rate for time zero, time one, etc.

The other bond and set of rates is described later in the chapter.

Mathematically, the problem is to determine the Present Value, where:

$$\text{Present Value} = \sum \text{Cashflow}_t \times \text{Discount Factor}_t$$

Where

$$\text{Discount Factor}_t = \prod^t \frac{1}{1 + \text{Discount Rate}_i}$$

5. Functional Abstractions

We will repeatedly solve the same problem before extending it to more examples. It may feel repetitive but the focus here is not the problem, but rather the variations between the different approaches.

5.3. Imperative Style

One of the most familiar styles of programming is called **imperative** (or **procedural**), where we provide step-by-step commands are provided to the computer. The programmer defines the data involved and how that data moves through the program one step at a time. It commonly uses loops to perform tasks repeatedly or across a set of data. The program's **state** (assignment and logic of the program's variables) is defined and managed by the programmer explicitly.

Here's an imperative style of calculating the present value of the bond.

```
let (1)
    pv = 0.0
    discount = 1.0

    for i in 1:length(cf_bond) (2)
        discount = discount / (1 + rate[i])
        pv = pv + discount * cf_bond[i] (3)
    end
    pv
end
```

- ① Declare variables to keep track of the discount rate and running total for value
- ② Loop the length of the cashflow vector.
- ③ At each step of the loop, look up (via index *i*) update the discount factor to account for the prevailing rate and add the discounted cashflow to the running total present value.

121.48888490821489

This style is simple, digestible, and clear. If we were performing the calculation by hand, it would likely follow a pattern very similar to this. Look up the first cashflow and discount rate, compute a discount factor, and subtotal the value. Repeat for the next set of values.

5.3.1. Iterators

Note that in the prior code example we defined an index variable `i` and had to manually define the range over which it would operate (1 through the length of the bond's cashflow vector). A couple of reasons this could be sub-optimal:

1. We are making the *assumption* that the indices of the vectors start with one, when in reality Julia arrays *can* be defined to start at 0 or another arbitrary index.
2. We manually perform the lookup of the values within each iteration.

We can solve the first one (partially) by letting Julia return an iterable set of values corresponding to the indices of the `cf_bond` vector. This is an example of an **iterator** which is an object upon which we can repeatedly ask for the next value until it tells us to stop.

By using `eachindex` we can get the indices of the vector since Julia already knows what they are:

```
eachindex(cf_bond)
```

```
Base.OneTo(5)
```

Lazy Programming

The result, `Base.OneTo(5)` is a **lazy** object which represents a collection that does not get fully instantiated until asked to (which may not actually be necessary). Many (most?) iterators are lazy but we can interact with them without fully instantiating the data that they represent. For example, we could find the largest index:

```
maximum(eachindex(cf_bond))
```

```
5
```

The point is if we have an object that *represents* a set, we need not actually enumerate each element of the set to interact with it.

We can fully instantiate an iterator with `collect`

```
collect(eachindex(cf_bond))
```

```
5-element Vector{Int64}:
```

```
1  
2  
3  
4  
5
```

5. Functional Abstractions

Laziness is generally a good thing in programming because sometimes it can be computationally or memory expensive to fully instantiate the collection of interest.

And when used in context:

```
let
    pv = 0.0
    discount = 1.0

    for i in eachindex(cf_bond)
        discount = discount / (1 + rate[i])
        pv = pv + discount * cf_bond[i]
    end
    pv
end
```

121.48888490821489

Here Julia gave us the index associated with the bond cashflows, but we are still looking up the values (why not just ask for the values instead of their index?) as well as assuming that the indices are the same for the discount rates.

We can get the value and the associated index with enumerate:

```
collect(enumerate(cf_bond))

5-element Vector{Tuple{Int64, Int64}}:
(1, 10)
(2, 10)
(3, 10)
(4, 10)
(5, 110)
```

This would allow us to skip the step of needing to look up the bond's cashflows. However, we can go even further by just asking for value associated with both collections. With zip (named because it's sort of like zipping up two collections together), we get an iterator that provides the values of the underlying collections:

```
collect(zip(cf_bond, rate))
```

5.4. Functional Techniques and Terminology

```
5-element Vector{Tuple{Int64, Float64}}:  
(10, 0.05)  
(10, 0.06)  
(10, 0.05)  
(10, 0.04)  
(110, 0.05)
```

This provides the simplest implementation of the imperative approaches:

```
let  
    pv = 0.0  
    discount = 1.0  
  
    for (cf, r) in zip(cf_bond, rate)  
        discount = discount / (1 + r)  
        pv = pv + discount * cf  
    end  
    pv  
end
```

```
121.48888490821489
```

The primary downsides to this approach are:

1. Needing to keep track of state is fine in simple cases, but can quickly become difficult to reason about and error prone as the number and complexity of variables grows.
2. Program flow is explicitly stated, leaving fewer places that the compiler can automatically optimize or parallelize.

Note that it's when state is `mutable` that a program tends to be more complex.

5.4. Functional Techniques and Terminology

Functional programming is a paradigm which attempts to minimize state via composing functions together.

Table 5.1 introduces a set of core functional methods to familiarize yourself with. Note that anonymous functions (#sec-anonymous-functions) are used frequently to define intermediary steps.

5. Functional Abstractions

Table 5.1.: Important Functional Methods.

Function	Description	Example
<code>map(f, v)</code>	Apply function <code>f</code> to each element of the collection <code>v</code> .	<code>map(</code> <code>x → x^2,</code> <code>[1, 3, 5]</code> <code>) # [1, 9, 25]</code>
<code>reduce(op, v)</code>	Apply binary <code>op</code> to pairs of values, reducing the dimension of the collection <code>v</code> .	<code>reduce(</code> <code>*</code> , <code>[1, 3, 5]</code> <code>) # 15</code>
	Has a couple of important, optional keyword arguments to note (which also apply to other variants of <code>reduce</code> below):	<ul style="list-style-type: none"> • <code>init</code> defines the identity element (e.g. the initial value of <code>+</code> and <code>*</code> is <code>0</code> and <code>1</code> respectively) • <code>dims</code> defines which dimension to reduce across (if the dimension of <code>v</code> is more than one).
<code>mapreduce(op, f, v)</code>	Maps <code>f</code> over collection <code>v</code> and returns a reduced result using <code>op</code> .	<code>mapreduce(</code> <code>*</code> , <code>x → x^2,</code> <code>[1, 3, 5]</code> <code>) # 35</code>
<code>foldl(op, v)</code>	Like <code>reduce</code> , but applies <code>op</code> from left to right (<code>foldl</code>) or right to left (<code>foldr</code>). Also has <code>mapfoldl</code> and <code>mapfoldr</code> versions.	<code>foldl(</code> <code>*</code> , <code>[1, 3, 5]</code> <code>) # 15</code>
<code>accumulate(op, v)</code>	Apply <code>op</code> along <code>v</code> , creating a vector with the cumulative result.	<code>accumulate(</code> <code>+</code> , <code>[1, 3, 5]</code> <code>) # [1, 4, 9]</code>
<code>filter(f, v)</code>	Apply <code>f</code> along <code>v</code> and return a copy of <code>v</code> with elements where <code>f</code> is true	<code>filter(</code> <code>>=(3),</code> <code>[1, 3, 5]</code> <code>) # [3, 5]</code>

5.4. Functional Techniques and Terminology

This paradigm is very powerful in a few ways:

1. It provides a language for talking about what a computation is doing. Instead of “looping over a collection called `portfolio` and calling a `value` function” we can more concisely refer to this as `mapreduce(value, portfolio)`.
2. Often times you are forced to think about the design of the program more deeply, recognizing the core calculations and data used within the model.
3. The compiler is free to apply more optimizations. For example, with `reduce`, the compiler could drive the calculation in any order since the operation is associative.
4. The lack of mutable state

Let's build a version of the present value calculation using the functional building blocks described above.

5.4.1. map

`map` is so named for the mathematical concept of mapping an input to an output. Here, it's effectively the same thing. We take a collection and use the given function to calculate an output. The size of the output equals the size of the input.

First, to show how we could calculate the discount factor we will use `map` to compute the one-period discount factors:

```
map(x → 1 / (1 + x), rate)
```

```
5-element Vector{Float64}:
0.9523809523809523
0.9433962264150942
0.9523809523809523
0.9615384615384615
0.9523809523809523
```

`map` transforms the `rate` collection by applying the anonymous function `x → 1 / (1 + x)`, which is the single period discount factor. This operation is conveyed visually in Figure 5.1.

Tip

`map` is an absolute workhorse of a function and the authors recommend using it liberally within your code. We find ourselves using `map` frequently, usually avoiding defining an explicit loop (unless we are modifying some existing collection).

An anti-pattern where `map` would likely be a better tool than a loop often looks like this:

5. Functional Abstractions

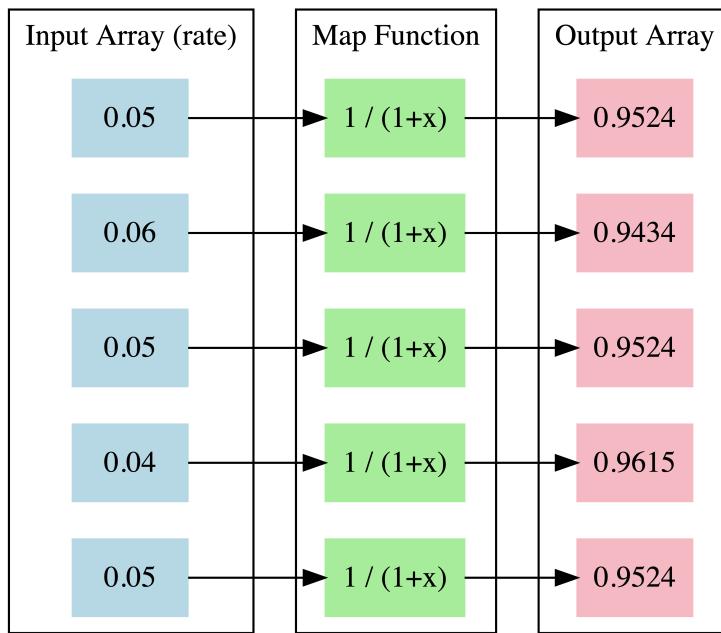


Figure 5.1.: A diagram showing that `map` creates a new collection mirroring the old one, output =after applying the given function to each element in the original collection.

```
for x in collection
    result = # ... do stuff ...
    push!(output,result)
end
output
```

Instead, `map` simplifies this to:

```
map(collection) do x
    # ... do stuff
end
```

Not only does this have the advantage of being clearer, more concise, and less work... it also let's Julia manage the output type of your computation so you don't have to worry about the type of output.

5.4.2. accumulate

`accumulate` takes an operation and a collection and returns a collection where each element is the cumulative result of applying the operation from the first element to the current one. For example, to calculate the cumulative product of the one-period discount factors:

```
accumulate(*, map(x → 1 / (1 + x), rate))
```

5-element Vector{Float64}:

```
0.9523809523809523
0.898472596585804
0.8556881872245752
0.822777103100553
0.7835972410481457
```

This results in a vector of the cumulative discount factors for each point in time corresponding to the given cashflows.

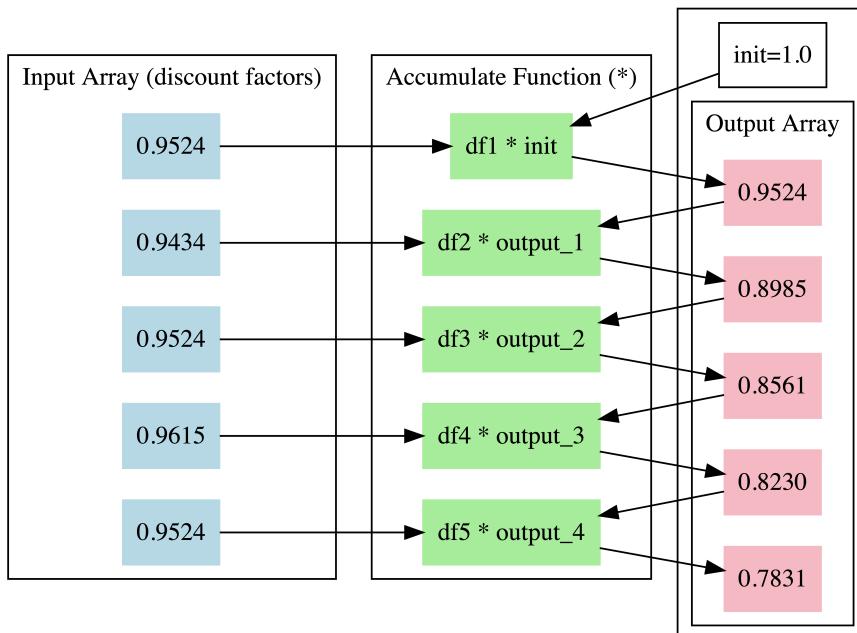


Figure 5.2.: A diagram showing that `accumulate` creates a new collection where each element is the cumulative result of applying the given operation to all previous elements.

5. Functional Abstractions

i Note

For `accumulate` and `reduce`, an important value is the `init` (an optional keyword argument), which is the initial value to start the accumulation or reduction. By default, for common operations this **identity element** is predefined. For example, for `+` the identity is `0` while for `*` it is `1`. The identity element e is the one where for a given binary operation \odot , that $x \odot e = x$.

Another example beyond addition and subtraction is string concatenation. In Julia, two strings are concatenated with `*` (like in mathematics, $a * b$ is also written as ab). The identity element for strings where the binary operation $\odot = *$ is `" "`. For example:

```
accumulate(*, ["a", "b", "c"], init="")  
  
3-element Vector{String}:  
"a"  
"ab"  
"abc"
```

This is a taste of a branch of mathematics known as Category Theory, a very rich subject but largely beyond the immediate scope of this book. The category theoretical term for sets of things that work with the binary operator and identity elements as described above is a monoid. You may ignore this fact.

5.4.3. `reduce`

`reduce` takes an operation and a collection and applies the operation repeatedly to pairs of elements until there is only a single value left.

For example, we start with the calculation of the vector of discounted cashflows

```
dfs = accumulate(*, map(x → 1 / (1 + x), rate))  
discounted_cfs = map(*, cf_bond, dfs)
```

```
5-element Vector{Float64}:  
9.523809523809524  
8.98472596585804  
8.556881872245752  
8.22777103100553  
86.19569651529602
```

Then we can sum them with `reduce`:

5.4. Functional Techniques and Terminology

```
reduce(+, discounted_cfs)
```

121.48888490821487

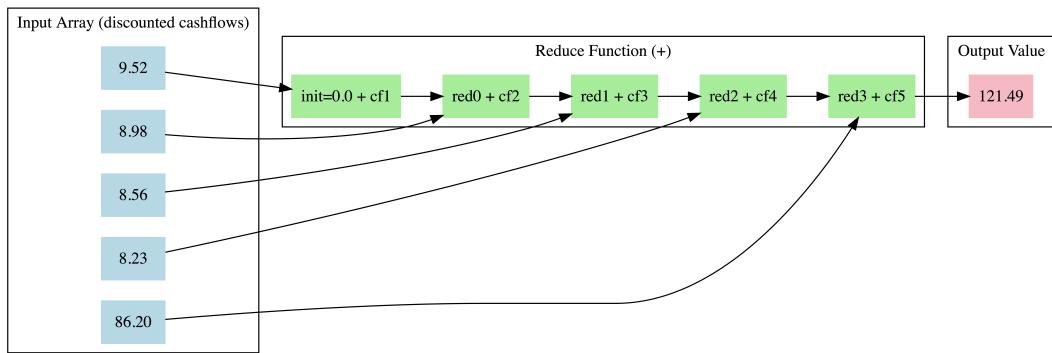


Figure 5.3.: A diagram showing how `reduce` applies the given operation to pairs of elements, ultimately reducing the collection to a single value.

5.4.4. mapreduce

We can combine `map`, `accumulate` and `reduce` to concisely calculate the present value in a functional style:

```
dfs = accumulate(*, map(x → 1 / (1 + x), rate))
mapreduce(*, +, cf_bond, dfs)
```

121.48888490821487

This calculates the discount factors, applies them to the cashflows with `map`, and sums the result with a reduction.

Tip

At the risk of sounding obvious, an easy way to make the program more “functional” is to simply use more functions. Do this one thing and it will improve the model’s organization, maintainability, and reduce bugs!

Take the example from earlier:

```
pv = 0.0
discount = 1.0
```

5. Functional Abstractions

```
for (cf, r) in zip(cf_bond, rate)
    discount = discount / (1 + r)
    pv = pv + discount * cf
end
pv
```

We can easily turn this code into a function so that it can operate on data beyond the single pair of `cf_bond` and `rate` previously defined:

```
function pv(rates,cashflows)
    pv = 0.0
    discount = 1.0

    for (cf, r) in zip(rates, cashflows)
        discount = discount / (1 + r)
        pv = pv + discount * cf
    end
    pv
end
```

①

- ① Here, `cf_bond` and `rate` would refer to whatever was passed as arguments to the function instead of any globally defined values.

Now we could use this definition of `pv` on other instances of `rates` and `cashflows`.

5.4.5. filter

`filter` does what you might think - filter a collection based on some criterion that can be determined as true or false.

For example filtering out even numbers using the `isodd` function:

```
filter(isodd, 1:6)
```

```
3-element Vector{Int64}:
1
3
5
```

Or filtering out things that don't match a criteria:

```
filter(x → ~(x == 5), 1:6)
```

5.4. Functional Techniques and Terminology

```
5-element Vector{Int64}:
1
2
3
4
6
```

While we didn't need `filter` to calculate a bond's present value in the example above, one can imagine how you may want to filter dates that a bond might pay a cashflow, say last day of a quarter:

```
using Dates
let d = Date(2024, 01, 01)
    filter(d → lastdayofquarter(d) == d, d:Day(1):lastdayofyear(d))
end
```

```
4-element Vector{Date}:
2024-03-31
2024-06-30
2024-09-30
2024-12-31
```

5.4.6. More Tips on Functional Styles

5.4.6.1. do Syntax for Function Arguments

In more complex situations such as with multiple collections or multi-line logic, there is a clearer syntax that is often used. `do` is a reserved keyword in Julia that creates an anonymous function and passes its arguments to a function like `map`. For example, this (terrible) code which decides if a number is prime. The anonymous function requires a `begin` block since the logic of the function is extended into multiple lines.

```
map(x → begin
    if x == 1
        "prime"
    elseif x == 2
        "not prime"
    elseif x == 3
        "prime"
    elseif x > 4
        "probably not prime"
    end
```

5. Functional Abstractions

```
end,  
[1, 2, 3, 10]  
)
```

This can be written more cleanly with the do syntax:

```
map([1, 2, 3, 10]) do x  
if x == 1  
"prime"  
elseif x == 2  
"not prime"  
elseif x == 3  
"prime"  
elseif x > 4  
"probably not prime"  
end)
```

5.4.6.2. Multiple Collections

map and the other functional operators discussed in this section can take multiple arguments. This is convenient if you have multiple arguments to a function:

```
discounts = [0.9, 0.81, 0.73]  
cashflows = [10, 10, 10]  
  
map((d, c) → d * c, discounts, cashflows)  
  
3-element Vector{Float64}:  
9.0  
8.100000000000001  
7.3
```

Or an example with the do syntax:

```
map(discounts, cashflows) do d, c  
d * c  
end  
  
3-element Vector{Float64}:  
9.0  
8.100000000000001  
7.3
```

5.4.6.3. Mixing Functional And Imperative Styles

One of the best things about Julia is how natural it can be to mix the different styles. Sometimes the best is the mix of both styles and that's one of the benefits of Julia: use the style that's most natural to the problem.

Note

Lisp ("list processing") is another, much older language than Julia (created in the 1950s!). One of its claims to fame is how flexible and powerful the tools are within the language to build upon. There's a couple aspects of this curse that we wish to describe because we can learn from it while Julia is still a relatively young language. Part of the "curse" is that: because there's so much freedom in what can be expressed in the language, there's not an obvious "best" way of doing things. This can lead to decision paralysis where you are trying to over-analyze what's the best way to write part of your code. Our advice: *don't worry about it!* A working implementation of something is better than an over-optimized idea.

The other part of the "curse" is that because is that it's relatively easy to implement so many things from the building blocks that Julia provides and compose them together to do what you want. This has a downside because the general approach to packages is smaller, standalone pieces that you call as needed. For example, consider Python's Pandas library, upon which Python's data science community was built. It came bundled with a CSV reader, Excel reader, Database reader, DataFrame type, visualization library, and statistical functions. In Julia, each of those are separate packages that specialize for the respective topics. This is advantageous in that they can progress independently from one another, you don't have to include functionality that you don't need, and you can mix and match libraries depending on your preference.

5.5. Array-Oriented Styles

Array-oriented programming is one that is practiced in two main contexts:

1. GPU programming
2. Python numerical computing

The former because GPUs want large blocks of similar data to operate in parallel. The latter is because native Python is too slow for many modeling problems so libraries like NumPy, SciPy, and tensor libraries define C++ (or similar) libraries for users to call out to.

5. Functional Abstractions

Array-oriented programming is not always natural for financial and actuarial applications. Differences in behavior or timing of underlying cashflows can make a set of otherwise similar products difficult to capture in nicely gridded arrays. Nonetheless, certain applications (scenario generation, some valuation routines) fit very naturally into this paradigm. Furthermore, for those that work well it's often a great way to extract additional performance due to the parallelization offered via CPU or GPU array programming.

Table 5.2 shows the bond present value example in this style.

Table 5.2.: Julia's broadcasting makes for an array-oriented style, similar to the approach that would be used with Python's NumPy.

Julia	Python (NumPy)
<pre>cf_bond = [10, 10, 10, 10, 110] rate = [0.05, 0.06, 0.05, 0.04, 0.05] discount_factors = cumprod(1 ./ (1 .+ rate)) sum(cf_bond .* discount_factors)</pre>	<pre>import numpy as np cf_bond = np.array([10, 10, 10, 10, 110]) discount_factors = np.array([0.05, 0.06, 0.05, 0.04, 0.05]) result = np.sum(cf_bond * discount_factors)</pre>

The downsides to this style are:

1. Sometimes it is unnatural because of non-uniformity of the data we are working with. For example if the length of the cashflows were shorter than the discount rates, we would have to perform intermediate steps to shorten or lengthen arrays in order to get them to be the same size.
2. A good bit of runtime performance is lost because the computer needs to allocate and fill many intermediate arrays (note how in Table 5.2, the `discount_factors` needs to instantiate an entirely new vector even though it's only temporarily used). See more on allocations in Chapter 8.

5.6. Recursion

The Fibonacci sequence is a classic example of a recursive algorithm:

$$F(n) = \begin{cases} 0, & \text{if } n = 0 \\ 1, & \text{if } n = 1 \\ F(n - 1) + F(n - 2), & \text{if } n > 1 \end{cases}$$

In code, this translates into a function definition that refers to itself:

```
function fibonacci(n)
    if n == 0
        return 0
    elseif n == 1
        return 1
    else
        return fibonacci(n-1) + fibonacci(n-2)
    end
end
```

One could imagine a possible pattern where the value of a stream of cashflows is defined as the sum of the value of the discounted next period plus the cashflows that occur this period.

```
function pv_recursive(rate,cashflows,accumulated_value=0.0)
    if isempty(cashflows)
        return accumulated_value
    else
        v = (accumulated_value + cashflows[end]) / (1 + rate)
        return pv_recursive(rate,cashflows[begin:end-1], v)
    end
end

pv_recursive (generic function with 2 methods)
```

And an example of its use:

```
pv_recursive(0.05,[10,10,10])
```

```
27.232480293704782
```

Generally, the recursive pattern includes defining a ‘base case’ where you stop the recursive behavior and return the result that has been accumulated up to that point.

6. Data and Types

I am only one, but I am one. I can't do everything, but I can do something.
The something I ought to do, I can do. And by the grace of God, I will -
Edward Everett Hale (1902)

6.1. In this section

The powerful benefits that using assigning types to data has within the model's system, some examples and relating some aspects of object oriented design.

6.2. Using Types to Value a Portfolio

We will assemble a set of interfaces that let's us value a portfolio of assets. Using the constructs introduced in the prior chapter, we can describe this as additively reducing over the value-mapped collection of assets in the portfolio. Or more concisely:

```
mapreduce(value,+,portfolio)
```

In `portfolio`, the assets may be heterogeneous so we will need to define what the valuation semantics are for the different kinds of assets. To get to our end goal, we will need to:

1. Define the different kinds of assets within our portfolio
2. How the assets are to be valued.

6.3. Benefits of Using Types

Defining types allows us to do several things:

1. **Separate concerns.** For example, deciding how to value an option need not know how we value a bond. The code and associated logic is kept distinct which is easier to reason about and to test.

6. Data and Types

2. **Re-use code.** When a set of types within a hierarchy all share the same logic, then we can define the method at the highest relevant level and avoid writing the method for each possible type. In our simple example we won't get as much benefit here since the hierarchy is simple and the set of types small.
3. **Dispatch on type.** By defining types for our assets, we can use multiple dispatch to define specialized behavior for each type. This allows us to write generic code that works with any asset type, and the Julia compiler will automatically select the appropriate method based on the type of the asset at runtime. This is a powerful feature that enables extensibility and modularity in our code.
4. **Improve readability and clarity.** By defining types for our assets, we make our code more expressive and self-documenting. The types provide a clear indication of what kind of data we are working with, making it easier for other developers (or ourselves in the future) to understand and maintain the codebase.^w
5. **Enable type safety.** By specifying the expected types for function arguments and return values, we can catch type-related errors at compile time rather than at runtime. This helps prevent bugs and makes our code more robust.

With these benefits in mind, let's start by defining the types for our assets. We'll create an abstract type called `Asset` that will serve as the parent type for all our asset types. If you haven't read it already, Section 4.5.7 is a good reference for details on types at the language level (this section is focused on organization and building up the abstracted valuation process).

6.4. Defining Types for Portfolio Valuation

We will define five types of assets in this simplified universe:

- Cash
- Risk Free Bonds (coupon and zero-coupon varieties)
- European Puts and Calls on Equities

To do the valuation of these, we need some economic parameters as well: risk free rates and option implied volatilities, which we will pass via named tuples. In a more robust model it would be wise to use types to differentiate between different kinds of economic assumption sets but we will limit the scope here such that

Here's the outline of what follows to get an understanding of types, type hierarchy, and multiple dispatch.

1. Define the Cash and Bond types.
2. Define the most basic economic parameter set.
3. Define the value functions for Cash and Bonds.

6.4. Defining Types for Portfolio Valuation

```

## Data type definitions
abstract type AbstractAsset end (1)

struct Cash <: AbstractAsset
    balance::Float64
end

abstract type AbstractBond <: AbstractAsset end (2)

struct CouponBond <: AbstractBond
    par::Float64
    coupon::Float64
    tenor::Int
end

struct ZeroCouponBond <: AbstractBond
    par::Float64
    tenor::Int
end

```

- ① General convention is to name abstract types beginning with Abstract...

Now to define the economic parameters:

```

struct EconomicAssumptions{T}
    riskfree::T
end

```

This is a parametric type because later on we will vary what objects we use for riskfree. For now, we will use simple scalar values, like in this potential scenario:

```

::: {.cell_execution_count=3} {.julia .cell-code} econ_baseline= EconomicAssumptions(0.05)
::: {.cell-output .cell-output-display execution_count=4} EconomicAssumptions{Float64}(0.05)
::: :::

```

Now on to defining the valuation for Cash and AbstractBonds. Cash is always equal to it's balance:

```

value(asset::Cash, ea::EconomicAssumptions) = asset.balance

value (generic function with 1 method)

```

Risk free bonds are the discounted present value of the riskless cashflows. We first define a method that generically operates on any fixed bond, all that's left to do is for

6. Data and Types

different types of bonds to define how much cashflow occurs at the given point in time by defining `cashflow` for the associated type.

```
function value(asset::AbstractBond, r::Float64)          (2)
    discount_factor = 1.0
    value = 0.0
    for t in 1:asset.tenor
        discount_factor /= (1 + r)
        value += discount_factor * cashflow(asset, t)      (1)
    end
    return value
end

function cashflow(bond::CouponBond, time)
    if time == bond.tenor
        (1 + bond.coupon) * bond.par
    else
        bond.coupon * bond.par
    end
end

function value(bond::ZeroCouponBond, r::Float64)         (3)
    return bond.par / (1 + r)^bond.tenor
end
```

(1) `x /= y`, `x += y`, etc. are shorthand ways to write `x = x / y` or `x = x + y`

(2) `value` is defined for `AbstractBonds` in general...

(3) ... and then more specifically for `ZeroCouponBonds`. This will be explained when discussing “dispatch” below.

`value` (generic function with 3 methods)

6.4.1. (Multiple) Dispatch

When a function is called, the computer has to decide which method to use. In the example above, when we want to value a `ZeroCouponBond`, does the `value(asset::AbstractBond, r)` or `value(bond::ZeroCouponBond, r)` version get used? **Dispatch** is the process of determining the right method to use and the rule is that *the most specific defined method gets used*. In this case, that means that even though our `ZeroCouponBond` is an `AbstractBond`, the routine that will be used is the most specific `value(bond::ZeroCouponBond, r)`.

Already, this is a powerful tool to simplify our code. Imagine the alternative of a long chain of conditional statements trying to find the right logic to use:

6.4. Defining Types for Portfolio Valuation

```
# don't do this!
function value(asset,r)
    if asset.type == "ZeroCouponBond"
        # special code for Zero coupon bonds
        #
    elseif asset.type == "ParBond"
        # special code for Par bonds
        #
    elseif asset.type == "AmortizingBond"
        # special code for Amortizing Bonds
        #
    else
        # here define the generic AbstractBond logic
    end
end
```

A more general concept is that of **multiple dispatch**, where the types of *all arguments* are used to determine which method to use. This is a very general paradigm, and in many ways is more extensible than traditional object oriented approaches, (more on that in the next section).

In our definition of `value` above, we used a simple scalar interest rate to determine the rate to discount the cash flows. What if instead of a scalar interest rate value we wanted to instead pass an object that represented a term structure of interest rates? Note how in the definition of `value` for `ZeroCouponBond`, we have defined a *more specific* signature: both the first and second arguments are specific, concrete types. When we call `value(ZeroCouponBond(100.0,3),0.05)`, we avoid the loop that's defined in the generic case and jump immediate to a more efficient definition of its `value`. This is **dispatching** on the combination of types and picking the most relevant (specific) version for what has been passed to it.

Despite the definitions above, the following will error because we haven't defined a method for `value` which takes as its second argument a type of `EconomicAssumptions`:

```
#| error: true
value(ZeroCouponBond(100.0,5),econ_baseline)
```

Let's fix that. Here we define a method which takes the economic assumption type and just relays the relevant risk free rate to the `value` methods already defined (which take an `AbstractBond` and a scalar `r`).

```
value(bond::AbstractBond,econ::EconomicAssumptions) = value(bond,econ.riskfree)

value (generic function with 4 methods)
```

6. Data and Types

Now this following works:

```
value(ZeroCouponBond(100.0, 5), econ_baseline)
```

78.35261664684589

Here's an example of how this would be used:

```
portfolio = [
    Cash(50.0),
    CouponBond(100.0, 0.05, 5),
    ZeroCouponBond(100.0, 5),
]
map(asset→ value(asset,econ_baseline), portfolio)
```

3-element Vector{Float64}:

50.0
99.9999999999999
78.35261664684589

This is very close to the goal that we set out at the end of the section. We can complete it by reducing over the collection to sum up the value:

```
mapreduce(asset → value(asset,econ_baseline), +, portfolio)
```

228.3526166468459

i Note

This code:

```
mapreduce(asset→ value(asset,econ_baseline), +, portfolio)
```

is more verbose than what we set out do at the start (`mapreduce(value,+,portfolio)`) due to the two-argument `value` function requiring a second arugment for the economic variables. This works well! However, there is a way to define it which avoids the anonymous function, which in some cases will end up needing to be compiled more frequently than you want it to. Sometime we want a lightweight, okay-to-compile-on-the-fly function. Other times, we know it's something that will be passed around in compute-intensive parts of the code. A technique in this situation is to define an object which "locks in" one of the arguments but behaves like the anonymous version. There is a pair

6.4. Defining Types for Portfolio Valuation

of types in the Base module, Fix1 and Fix2, which represent partially-applied versions of the two-argument function f, with the first or second argument fixed to the value "x".

This is, Base.Fix1(f, x) behaves like $y \rightarrow f(x, y)$ and Base.Fix2(f, x) behaves like $y \rightarrow f(y, x)$.

In the context of our valuation model, this would look like:

```
val = Base.Fix2(value,econ_baseline)
mapreduce(val,+,portfolio)
```

228.3526166468459

Extending the example, we can use a time-varying risk free rate instead of a constraint. For fun, let's say that the risk free rate has a sinusoidal pattern:

```
econ_sin = EconomicAssumptions(t → 0.05 + sin(t) / 100)
```

```
EconomicAssumptions{var "#15#16"}(var "#15#16"())
```

Now value will not work, because we've only defined how value works on bonds if the given rate is a Float64 type:

```
#| error: true
value(ZeroCouponBond(100.0, 5), econ_sin)
```

We can extend our methods to account for this:

```
function value(bond::ZeroCouponBond, r::T) where {T<:Function}           ①
    return bond.par / (1 + r(bond.tenor))bond.tenor                      ②
end
```

- ① The `r :: T ... where {T<:Function}` says use this method if r is any concrete subtype of the (abstract) Function type.
- ② r is a function, where we call the time to get the zero coupon bond (a.k.a. spot) rate for the given timepoint.

```
value (generic function with 5 methods)
```

Now it works:

```
value(ZeroCouponBond(100.0, 5), econ_sin)
```

82.03058910862806

6. Data and Types

6.5. Object Oriented Design

There's enough general familiarity with object oriented ("OO") design that it's worth describing for understanding how it compares and contrasts to other design patterns. Object oriented systems attempt to form the analogy that various parts of the system are their own objects which encapsulate both data and behavior. Object oriented design is often one the first computer programming abstractions introduced because it very relatable¹, however this comparative discussion will point out a number of its flaws as well. That said, much of OO design can be emulated in Julia except for data inheritance.

We bring up object oriented design not because of the authors (admittedly subjective) opinion that the object-oriented paradigm can be less suitable for financial modeling, but because by having a (potentially more relatable) contrasting approach we can better illuminate certain ideas and concepts.

6.6. Assigning Behavior

The `value` function is a good example of where the OO requirement to ascribe behavior to a single type (class) can lead to confusing design. If we had to assign `value` to one of the objects involved, should it be the economic parameters of the asset contracts? The choice is not obvious at all. Isn't it the market (economic parameters) that determines the value? But then if `value` were to be a method wholly owned by the economic parameters, how could it possibly define in advance the valuation semantics of all types of assets? What if one wanted to extend the valuation to a new asset class? These are problems presented in traditional OO designs and that resolve so elegantly with multiple dispatch.

6.7. Inheritance

We discussed the type hierarchy in Chapter 4 and in most OO implementations this hierarchy comes with inheriting both data *and* behavior. This is different from Julia where subtypes inherit behavior but not data from the parent type.

Inheriting the data tends to introduce a tight coupling between the parent and the child classes in OO systems. This tight coupling can lead to several issues, particularly as systems grow in complexity. For example, changes in the parent class can inadvertently affect the behavior of all its child classes, which can be problematic if these changes are not carefully managed. This is often referred to as the "fragile base class problem,"

¹And projections, which is handled by defining a `ProjectionKind`, such as a cashflow or accounting basis. This topic is covered in more detail in the `FinanceModels.jl` documentation.

where base classes are delicate and changes to them can have widespread, unintended consequences.

Another issue with inheritance in OO design is the temptation to use it for code reuse, which can lead to inappropriate hierarchies. Developers might create deep inheritance structures just to reuse code, leading to a scenario where classes are not logically related but are forced into a hierarchy. This can make the system harder to understand and maintain.

Moreover, inheritance can sometimes lead to the duplication of code across the hierarchy, especially if the inherited behavior needs to be slightly modified in different child classes. This goes against the DRY (Don't Repeat Yourself) principle, which is a fundamental concept in software engineering advocating for the reduction of repetition in code.

6.7.1. Composition over Inheritance

To mitigate some of the problems associated with inheritance, there's a growing preference for *composition*. Composition involves creating objects that contain instances of other objects to achieve complex behaviors. This approach is more flexible than inheritance as it allows for the creation of more modular and reusable code. There is a general preference for "composition over inheritance" among professional developers these days.

In composition, objects are constructed from other objects, and behaviors are delegated to these contained objects. This approach allows for greater flexibility, as it's easier to change the behavior of a system by replacing parts of it without affecting the entire hierarchy, as is often the case with inheritance.

Composition looks like this:

```
struct CUSIP
    code::string
end

struct FixedIncome
    coupon::Float64
    tenor::Float64
end

struct MunicipalBond
    cusip::CUSIP
    fi::FixedIncome
end

struct ListedOption
```

6. Data and Types

```
cusip::CUSIP
#... other data fields
end

struct UnlistedBond
    fi::FixedIncome
end

# define behavior which relies on defining
last_transaction(c::CUSIP) = # ... perform lookup of data
last_transaction(asset) = last_transaction(asset.cusip)

duration(f::FixedIncome) = # ... calculate duration
duration(asset) = duration(asset.fi)
```

In the above example, there are number of asset classes that have CUSIP related attributes (i.e. the 9 character code) and behavior (e.g. being able to look up transaction data). Other assets have fixed income attributes (e.g. calculating a duration). But not all of these assets have a CUSIP! Composition lets us bundle the data and behavior together without needing complex chains of inheritance.

i Note

A CUSIP (Committee on Uniform Security Identification Procedures) number, is a unique nine-character alphanumeric code assigned to securities, such as stocks and bonds, in the United States and Canada. This code is used to facilitate the clearing and settlement process of securities and to uniquely identify them in transactions and records.

7. Higher Levels of Abstraction

“Simple things should be simple, complex things should be possible.” —
Alan Kay (1970s)

7.1. In this section

Why we talk about abstraction as a technique in and of itself, discussion of abstraction at the level of code organization and interfaces.

7.2. Introduction

In programming and modeling, as in mathematics, abstraction permits the definition of interchangeable components and patterns that can be reused. Abstraction is a selective ignorance—focusing on the aspects of the problem that are relevant, and ignoring the others. The last two chapters described what we might call “micro” level abstractions: specific functions or types. In this chapter, we zoom out and examine some principles that guide good model development and how that manifests itself in architectural concerns such as how different parts of the code are organized, what parts of the program are considered ‘public’ versus ‘private’, and patterns themselves.

Chapter 4 Described a number of tools that we can utilize as interfaces within our model. We use these tools that are provided by our programming language *in service of* the conceptual abstraction described above.

- Functions let us implement behavior, where we need trouble ourselves with the low level details.
- Data types provide a hierarchical structure to provide meaning to things, and to group those things together into more meaningful structures.
- Modules allow us to combine data, and or function, into a related group of concepts which can be shared in different parts of our model

7. Higher Levels of Abstraction

7.3. Principles for Abstraction

Here is a list of some principles that arise when developing a particular abstraction. Not all abstractions serve all of these purposes but generally fit one or more of them.

Table 7.1.: Finding abstractions generally means finding patterns that fit into one of these principles.

Principle	What	Why	Example
Separation of Concerns	Divide the system into distinct parts, each addressing a separate concern	Promote modularity and reduce high degree of dependence (coupling) between components	Separating data retrieval, data processing, and output generation steps in a process
Encapsulation	Hide the internal details of a component and expose only a clean, well-defined set of functionality (interface)	Don't let other parts of the program modify internal data and make the system easier to understand and maintain	Defining a type or module with well defined behavior and responsibility
Composability	Design simple components that can be combined to create more complex behaviors, as opposed to a single component that attempts to handle all behavior.	Promote reuse and allow for the components to be combined creatively	Separate details about economic conditions into different types than contracts/instruments
Generalization	Identify common patterns and create generic components that can be specialized as needed. Often this means identifying the common behavior that arises repeatedly in a model	Avoid duplication and make the system more expressive and extensible	Defining a generic Instrument type that can be specialized for different asset classes

These principles provide guidance for creating abstractions that are modular, reusable, and maintainable. By following these principles, developers can create financial models

7.3. Principles for Abstraction

that are easier to understand, extend, and adapt to changing requirements.

7.3.1. Pragmatic Considerations for Model Design

7.3.1.1. Behavior-Oriented

This strategies is to effectively group together components with a model that behaves similarly. So, in our example of bonds and interest-rate swaps fundamentally, they share many characteristics and are used in very similar ways within a model. Therefore, it might make sense to group them together when developing a model.

7.3.1.2. Domain Expertise

It may be that components of the model require sufficient expertise that different persons or groups are involved in the development. This may warrant separating a models design, So that different groups contributing to the model can focus on any more narrow aspect, Regardless of inherent similarity of components. For example, at a higher vertical level of obstruction, financial derivatives may fall under similar grouping, but sufficient differences exist for equity credit or foreign exchange derivatives that the model should separate those three asset classes for development purposes.

7.3.1.3. Composability versus All-in-One

For some model design goals, it may be warranted to attempt to bundle together more functionality instead of allowing users to compose a functionality that comes from different packages. For example, perhaps a certain visualization of a model result is particularly useful, It is not easy to create from scratch, And virtually everyone using the model, will desire to see the model output visualized that way. Instead of relying on the user to install a separate visualization package and develop the visualization themselves, it could make sense to bundle visualization functionality with a model that is otherwise unconcerned with graphical capabilities.

In general, though it is preferred to try to loosely couple systems, you can pick and choose which components you use and that those components work well together.

7. Higher Levels of Abstraction

7.4. Interfaces

Interfaces are the boundary between different encapsulated abstractions. The user-facing interface is the set of functionality and details that the user of the package or model must consider, which is separate from the intermediate variables, logic, and complexity that may be contained within.

Example of an interface

When looking up a ticker for a market quote, one need not be mindful of the underlying realtime databases, networking, rendering text to the screen, memory management, etc. The interface is “put in symbol, get out number”. By design there are multiple layers of interfaces and abstractions that are used but the financial modeler need only be actively concerned about the points that he or she comes in contact with, not the entire chain of complexity.

For a financial model this might mean that there is an interface for bonds, or there is an interface for interest-rate swaps. There may be a different interface for calculating risk metrics or visualizing the results.

Financial model this might mean that there is an interface for bonds, or there is an interface for interest-rate swaps. There may be a different interface for calculating risk metrics or visualizing the results. A better system design will separate the concern of visualizing output from the mechanics of a fixed income contract. This is what it means to put boundaries on different parts of a models logic. One of the easiest places to see this is with the available open source packages. There are packages available for visualizations, data frames, file, storage, statistical analysis, etc. for many of these it's easy to see where the natural boundary lies.

However, it's often difficult to find where to draw lines within financial models. For example, should bonds and interest-rate swaps be in separate packages? Or both part of a broader fixed income package? This is where much of the art and domain expertise of the financial professional comes to bear in modeling. There would be no way for a pure software engineer to think about the right design for the system without understanding how underlying components share, similarities or differences and how those components interact.

7.4.1. Defining Good Interfaces

A well-designed interface should follow these principles:

1. **Be minimal and focused.** The interface should provide only the essential functionality needed, without unnecessary clutter or features. This makes the interface

easier to understand and facilitates building the necessary complexity through digestible, composable components.

2. **Be consistent and intuitive.** The interface should use consistent naming conventions, parameter orders, and behaviors. It should match the user's mental model and expectations.
3. **Hide implementation details.** The interface should abstract away the internal complexity and expose only what the user needs to know. This of details allows the implementation to change without affecting users of the interface.
4. **Be documented and contractual.** The interface should clearly specify what inputs it expects and what outputs or behaviors it provides. It forms a contract between the implementation and the users.
5. **Be testable.** A good interface allows the functionality to be easily tested through the public interface, without needing to access internal details.

7.4.2. Interfaces: A Financial Modeling Case Study

As a case study, we'll look at the `FinanceModels.jl` and related packages to discuss some of the background and design choices that went into the functionality. This suite was written by one of the authors and is publically available as set of installable Julia packages.

7.4.2.1. Background

In actuarial work, it is common to need to work with interest rate and bond yield curves to determine current forward rates, estimates of the shape of future yield curves, or discount a series of cashflows to determine a present value. Determining things like "given a par yield curve, what's the implied discount factor for a cashflow at time 10" or "what is the 10 year BBB public corporate rate implied by the current curve in five years' time" is cumbersome at best in a spreadsheet.

For example, to determine the answer to the first one ("a discount factor for time 10") actually requires quite a bit of detail and assumption to derive:

- Reference market data and a specification for how that market data should be interpreted. For example, if given the rate `0.05` for time 10, quoted as a continuous rate or annual effective? Is that a par rate, a zero-coupon bond (spot) rate, or a one-year-forward rate from time 10?
- Smoothing, interpolation, or extrapolation for noisy or sparse data. Should the rates be bootstrapped or fit to a parametrically specified curve?

This is the type of complexity that we wish to save the user from needing to keep front of mind when the primary goal is, e.g., valuation of a stream of riskless life insurance payments, which might look like this:

7. Higher Levels of Abstraction

```
risk_free_rates = [0.05,0.06,...0.06]
tenors = [1/12,3/12,...30]
yield_curve = Yields.Par(risk_free_rates,tenors)

cashflow_vector = [1e6,3e6,...,1e3]
present_value(yield_curve,cashflow_vector)
```

This is very clear from the variable and function names what the purpose and steps in the analysis are. Imagine starting with rates and cashflows in a spreadsheet, needing to perform the bootstrapping, interpolation, and discounting before getting to the simple present value sought in the analysis. What can be, with the right abstractions, distilled into five lines of code would take hundreds of cells in a spreadsheet. Providing abstractions like this at the hand of financial modelers is a productivity multiplier.

7.4.2.2. Initial Versions

There were two main abstractions to talk about from early versions of the packages.

7.4.2.2.1. Rates

Utilizing the benefit of the type system, it was decided that it would be most useful to represent rates not as simple floating point numbers (e.g. 0.05) but instead with dedicated types to distinguish between rate conventions. The abstract type `CompoundingFrequency` had two subtypes: `Continuous` and `Periodic` so that a 5% rate compounded continuously versus an effective per period rate would be distinguished via `Continuous(0.05)` versus `Periodic(0.05,1)`. The two could be converted between by extending the built-in `Base.convert` function.

This was useful because once rates were converted into `Rates` within the ecosystem, that data contained within itself characteristics that could distinguish how downstream functionality should treat the rates.

7.4.2.2.2. Yield Curves

At first, only bootstrapping was supported as a method to construct curve objects. This required that there was only one rate given per time period (no noisy data) and only supported linear, quadratic, and cubic splines.

Further, there was a specific constructor for different common types of instruments. From the old documentation:

- `Yields.Zero(rates,maturities)` using a vector of zero, or spot, rates
- `Yields.Forward(rates,maturities)` using a vector of one-period
- `Yields.Constant(rate)` takes a single constant rate for all times

- `Yields.Par(rates,maturities)` takes a series of yields for securities priced at par. Assumes that maturities ≤ 1 year do not pay coupons and that after one year, pays coupons with frequency equal to the CompoundingFrequency of the corresponding rate.
- `Yields.CMT(rates,maturities)` takes the most commonly presented rate data (e.g. Treasury.gov) and bootstraps the curve given the combination of bills and bonds.
- `Yields.OIS(rates,maturities)` takes the most commonly presented rate data for overnight swaps and bootstraps the curve.

This covered a lot of lightweight use-cases, but made a lot of implicit assumptions about how the given rates should be interpreted.

7.4.2.3. The Birth of FinanceModels

There were a multiple of insights that led to a more flexible interface in more recent versions.

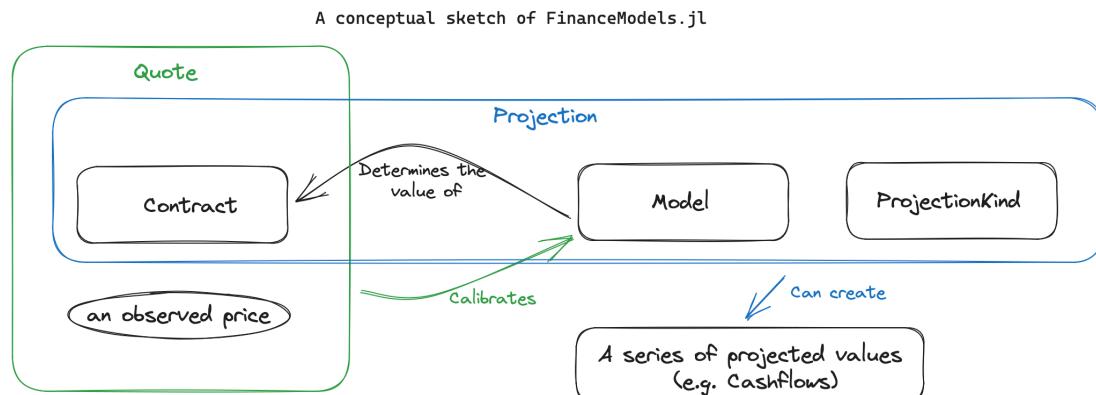


Figure 7.1.: A conceptual sketch of `FinanceModels.jl` components.

First, realizing that yield curves were just a particular kind of model - one that used interest rates to discount cashflows. But you can have different kinds of models - such as Black-Scholes option valuation or a Monte Carlo valuation approach. Likewise, the cashflows need not simply be a vector of floating point values, and instead it could be the representation of a generic financial contract. As long as the model knew how to value it, an appropriate present value could be derived.

Where previously it was:

```
present_value(yield_curve,cashflow_vector)
```

7. Higher Levels of Abstraction

Now, it was

```
present_value(model,contract)
```

Second, that a model was simple some generic box that had been “fit” to previously observed prices for similar types of contracts we would be trying to value in the model. The combination of a contract and a price constituted a “quote” and with multiple quotes a model could be fit using various algorithms.

With these changes, the package that was originally called Yields.jl was renamed to FinanceModels.jl. The updated code from the earlier example now would be implemented like this:

```
risk_free_rates = [0.05,0.06,...0.06]
tenors = [1/12,3/12,...30]
quotes = ParYield.(risk_free_rates,tenors)
model = fit(Spline.Cubic(),quotes,Fit.Bootstrap())

cashflow_vector = [1e6,3e6,...,1e3]
present_value(model,cashflow_vector)
```

It's slightly more verbose, but notice how much more powerful and extensible `fit(Spline.Cubic(), quotes, Fit.Bootstrap())` is than `Yields.Par(risk_free_rates, tenors)`. The end result is the same, but now the same package and interface can clearly interchange other options, such as a NelsonSiegelSvensson curve instead of a spline. And the quotes could be a combination of observed bonds of different technical parameters (though still sharing characteristics which make it relevant for the model being constructed).

The same pattern also applies for option valuation, such as this example of vanilla euro options with an assumed constant volatility assumption:

```
a = Option.EuroCall(CommonEquity(), 1.0, 1.0)                                ①
b = Option.EuroCall(CommonEquity(), 1.0, 2.0)

qs = [
    Quote(0.0541, a),
    Quote(0.072636, b),
]

model = Equity.BlackScholesMerton(0.01, 0.02, Volatility.Constant())           ③
m = fit(model, qs)                                                               ④
present_value(m,qs[1].instrument)                                               ⑤
```

7.5. Macros & Homoiconicity

- ① The arguments to EuroCall are the underlying asset type, strike, and maturity time.
- ② A vector of observed option prices.
- ③ A BSM model with a given risk free rate, dividend yield, and a to-be-fit constant volatility component.
- ④ Fits the model and derives an approximate volatility of 0.15 .
- ⑤ Values the contract and in such a simple, noiseless model we recover the original price of 0.0541

With a consistent interface able to handle a wide variety of situations, the modeler is free to expand the model in new directions of analysis with the built in functionality allowing him or her to compose pieces together that was not possible with the less abstracted design. For example, the equity option example had no parallel when all of the available constructors were `Yields.Zero` or `Yields.Par` and would have required a completely from-scratch implementation with newly defined functions.

Further, and critically, the new design allows modelers to create their own models or contracts¹ and extend the existing methods rather than needing to create their own: the function signature `fit(model, quotes)` handles a very wide variety of cases, as does `present_value(model, contract)`.

7.5. Macros & Homoiconicity

We've talked about transforming data and restructuring logic in order to make the model more effective. We can go still deeper!(Or is it higher level?) We can actually abstract the process of writing code itself! This subject is a bit advanced, so we are simply going to introduce it because you will likely find many convenient instances of it as a *user* even if you never find a need to implement this yourself.

Homoiconicity refers to the property of a programming language where the language's code can be represented and manipulated as a data structure in the language itself. In other words, the code is data and can be treated as such. This enables powerful metaprogramming (i.e. code that writes other code) capabilities, where code can be generated or transformed during the compilation process.

Macros are a metaprogramming feature that leverage homoiconicity in Julia. They allow the programmer to write code that generates or manipulates other code at compile-time. Macros take code as input, transform it based on certain rules or patterns, and return the modified code which then gets compiled.

For example, a built-in macro is `@time` which will measure the elapsed runtime for a piece of code².

¹And projections, which is handled by defining a `ProjectionKind`, such as a cashflow or accounting basis.
This topic is covered in more detail in the `FinanceModels.jl` documentation.

²(`time?`) is a simple, built-in function. For true benchmarking purposes, see Section 10.2.

7. Higher Levels of Abstraction

```
@time exp(rand())
```

Will effectively expand to:

```
t0 = time_ns()
value = exp(rand())
t1 = time_ns()
println("elapsed time: ", (t1-t0)/1e9, " seconds")
value
```

Here it is when we run it:

```
@time exp(rand())
```

```
0.000003 seconds
```

```
1.2421471640746253
```

7.5.1. Metaprogramming in Financial Modeling

In the context of financial modeling, macros can be used to simplify repetitive or complex code patterns, enforce certain conventions or constraints, or generate code based on data or configuration.

Here are a few potential use cases of macros in financial modeling. Again, these are more advanced use-cases but knowing that these paths exist may benefit your work in the future.

1. Defining custom DSLs (Domain-Specific Languages): Macros can be used to create expressive and concise DSLs tailored to financial modeling. For example, a macro could allow defining financial contracts using a syntax closer to the domain language, which then gets expanded into the underlying implementation code.
2. Automating boilerplate code: Macros can help reduce code duplication by generating common patterns or boilerplate code. This can include generating accessor functions³, constructors, or serialization logic based on type definitions.
3. Enforcing conventions and constraints: Macros can be used to enforce coding conventions, such as naming rules or type checks, by automatically transforming code that doesn't adhere to the conventions. They can also be used to add runtime assertions or checks based on certain conditions.

³Accessor functions are useful when working with nested data structures. For example, if you have a `struct` within a `struct` and want to conveniently access an inner `struct`'s field.

4. Optimizing performance: Macros can be used to perform code optimizations at compile-time. For example, a macro could unroll loops, inline functions, or specialize generic code based on specific types or parameters, resulting in more efficient runtime code.
5. Generating code from data: Macros can be used to generate code based on external data or configuration files. For example, a macro could read a specification file and generate the corresponding financial contract types and functions.

7.5.2. Commonly Encountered Macros

Table 7.2.: Useful macros for modeling work. There are others related to parallelism which will be covered in Chapter 11.

Macro	Description
BenchmarkTools.@benchmark	Runs the given expression multiple times, collecting timing and memory allocation statistics. Useful for benchmarking and performance analysis.
BenchmarkTools.@btime	Similar to @benchmark, but focuses on the minimum execution time and provides a more concise output.
@edit	Opens the source code of a function or module in an editor for inspection or modification.
@which	Displays the method that would be called for a given function call, helping to understand method dispatch.
@code_warntype	Shows the type inference results for a given function call, highlighting any type instabilities or performance issues.
@info, @warn, @error	Used for logging messages at different severity levels (info, warning, error) during program execution.
@assert	Asserts that a given condition is true, throwing an error if the condition is false. Useful for runtime checks and debugging.
@view, @views	Access a subset of an array without copying the data in that slice. @views applies to all array slicing operations within the expressions that follow it.
Test.@test, Test.@testset	Used for defining unit tests. @test checks that a condition is true, while @testset groups related tests together.
@raw	Encloses a string literal, disabling string interpolation and escape sequences. Useful for writing raw string data. This is especially helpful when working with filepaths where the \ in Windows paths otherwise needs to be escaped with a leading slash (e.g. \\).

7. Higher Levels of Abstraction

Macro	Description
@fastmath	Enables aggressive floating-point optimizations within a block, potentially sacrificing strict IEEE compliance for performance.
@inbounds	Disables bounds checking for array accesses within a block, improving performance but removing safety checks.
@inline	Suggests to the compiler that a function should be inlined at its call sites, potentially improving performance by reducing function call overhead.

Part III.

Conceptual Foundations: Building Performant Models

8. Hardware and Its Implications

a CPU is literally a rock that we tricked into thinking.

not to oversimplify: first you have to flatten the rock and put lightning inside it.

- Twitter user daisyowl, 2017

8.1. In this section

A discussion of why a cursory understanding of modern computing hardware and architecture is important for making the right design decisions within a modeling context. Stack vs heap allocations, pointers, and bit types.

8.2. Computer Architecture

We can think of a computer as handling data at rest (in memory) or being acted upon (processed).

8.2.1. Memory and Moving Data Around

The core of modeling on computers is to perform computations on data, but unfortunately the speed at which data can be *accessed* has grown much slower than the rate the actual computations can be performed. Further, the size of the available persistent data storage (HDDs/SSDs) has ballooned, exacerbating the problem, making the speed and volume of memory the typical constraint in most workflows. Solutions have been developed to create a pipeline intended to most efficiently shuttle data to and from the processor and the persistent storage. This extends further to data between different data stores and computers.

We will focus primarily on the architecture of a single computer, as even laptop computers today contain enough power for most modeling tasks, *if the computer is used effectively*. Further, learning how to optimize a program for a single computer/processes or is almost always a precursor step to effective parallelization as well.

8. Hardware and Its Implications

8.2.1.1. Memory Types and Location

Memory has an inverse relationship between size and proximity to the processor units. The closer the data is to the processor units, the smaller the storage and the less likely the data will persist at that location for very long.

Kind	Rough Size	Lifecycle
Solid State Disk (SSD) or Hard Disk Drive (HDD)	TBs	Persistent/Permanent
Random Access Memory (RAM)	Dozens to Hundreds of GBs	Seconds to Hours (while computer is powered on)
CPU Cache - L3	8 MB to 128 MB	Microseconds to Milliseconds
CPU Cache - L2	2 MB to 16 MB	Nanoseconds to Microseconds
CPU Cache - L1	~16 kB	Nanoseconds

After requesting data from a persistent location like a Solid State Drive (SSD), the memory is read into Random Access Memory (RAM). The advantage of RAM over a persistent location is speed, typically that memory can be accessed and modified many times faster than the persistent data location. The tradeoff is that RAM is not persistent: when the computer is powered down, the RAM loses the information stored within. When data is needed by the CPU for data is read from RAM into a small hierarchy of caches. The **CPU Caches** are small (physically and in capacity), but very fast. The caches are also physically colocated with the CPU for efficiency. Data is organized and funneled through the caches as an intermediary between the CPU and RAM and is fed from Level 3 (L3) cache in steps down to L1 cache as the data gets closer to the processor.

8.2.2. Stack vs Heap

Sitting within the RAM region of memory are two sections called Stack and Heap. These are places where variables created from our program's code can be stored. In both cases, the program will request memory space but they have some differences to be aware of.

The **stack** stores small, fixed-size (known bit length), data and program components. The stack is a last-in-first-out queue of data that is able to be written to and read from very quickly. The **heap** is a region which can be dynamically sized and has random read/write (you need not access the data in a particular order). The heap is much slower but more flexible.

8.2.2.1. Garbage Collector

The **garbage collector** is a program that gets run to free up previously requested/allocated memory. It accomplishes this by keeping track of references to data in memory by section of your code. If a section of code is no longer reachable (e.g. inside a function that will never get called again, or a loop that ran earlier in the program but is now complete) then periodically the garbage collector will pause execution of the primary program in order to “sweep” the memory and mark it as now able to be used by your program or the operating system.

8.2.3. Processor

The processor reads lines from the cache into registers and then executes instructions (e.g. take the bytes from register 10 and add the bytes from register 11 to them). This is really all a processor does at the lowest level: combine the electrical status of bits using logical circuits. Logical circuits (**transistors**) are an arrangement of wires that output a new electrical signal that varies depending on the input. From a collection of smaller building block gates (e.g. AND, OR, NOR, XOR) more complex operations can be built up¹, into operations like addition, multiplication, division, etc. These electrical signals move the state of the program forward once time per CPU cycle. CPU cycles are what’s quoted as cycles per second, e.g. when a chip is advertised as as 3.0 GHz (or 3 billion cycles per second).

The programmer (or compiler, if we are working in a higher level language like Julia) tells the CPU which instruction to run. The set of instructions that are valid for a given processor are called the Instruction Set Architecture, or ISA. In computer Assembly language (roughly one-level above directly manipulating the bits), the instructions are given names like ADD, SUB, MUL, DIV, MOV, etc. These are not all created equal, however, as some instructions take many CPU cycles to complete (floating point DIV takes 10-20 CPU cycles while ADD only takes a single CPU cycle).

Some architecture examples that may be familiar:

- Intel x86-64 (a.k.a. AMD64) are common computer processors that use registers that are 64 bits wide (the prior generation was 32 bits wide) and use the **x86** instruction set.
- ARM chips, including the Apple M-Series processors are characterized by the use of the **ARM** instruction set and recent processors of this kind are also 64 bit.

¹In fact, only a single logical gate is needed to reproduce all boolean logical gates: NAND (Not AND) and NOR gates can be composed to create AND, OR, NOR, etc. gates.

8. Hardware and Its Implications

The ARM is known as a **reduced instruction set chip** (RISC), which means that it has fewer available instructions compared to, e.g. the x86 architecture. The benefit of the reduced instruction set is that it is generally much more power efficient, but comes at the cost of sacrificing specialized instructions such as string manipulation, or in lower-end chips, even the division operation (which have to be implemented via software routines instead of CPU operations). However, for specialized workloads, the availability of a key instruction can make a program run on the CPU 10-100x faster at times. An example of this is that at the time of writing, AVX512 processors are becoming available (see Chapter 11 for a discussion of vectorization) which can benefit some workloads greatly.

Tip

Trying to optimize your program via selecting specialized chips should be one of the *last* ways that you seek to optimize runtime, as generally a similar order of magnitude speedup can be achieved through more efficient algorithm design or general parallelization techniques - in this way the performance is *portable* and can be used on other systems and not just special architectures.

When writing in Julia, you need not be concerned with the low-level instructions as the compiler will optimize the execution for you. However, should it be useful, it is easy to inspect the compiled code. For example, if we create a function to add three numbers, we can see that the ADD instruction is called twice: first adding the first and second arguments, and then adding the third argument to that intermediate sum.

```
myadd(x, y, z) = x + y + z
@code_native myadd(1, 2, 3)

.section    __TEXT,__text,regular,pure_instructions
.build_version macos, 14, 0
.globl _julia_myadd_1092           ; -- Begin function julia_myadd_1092
.p2align   2
_julia_myadd_1092:                 ; @julia_myadd_1092
; | @ In[2]:1 within `myadd'
; %bb.0:                           ; %top
; | @ operators.jl:587 within `+' @ int.jl:87
    add x8, x1, x0
    add x0, x8, x2
    ret
; LL                                ; -- End function
.subsections_via_symbols
```

 Tip

Compilers are complex, hyper-optimized programs which turn your source code into the raw bits executed by the computer. Key milestones in the Julia to binary compilation pipeline include the following items. Note the `@code_...` macros which Julia has and allow the programmer to inspect the intermediate representations.

Step	Description	Example
Julia Source Code	The level written by the programmer in a high level language.	<code>myadd(x,y,z) = x + y + z</code>
Lowered Abstract Syntax Tree (AST)	An intermediate representation of the code after the first stage of compilation, where the high-level syntax is simplified into a more structured form that's easier for the compiler to work with.	<code>julia> @code_lowered myadd(1,2,3)</code> CodeInfo(1 - %1 = x + y + z └ return %1)
LLVM	Low-Level Virtual Machine language, which is a massively popular compiler used by languages like Julia and Rust.	<code>julia> @code_llvm myadd(1,2,3)</code> ; @ REPL[7]:1 within `myadd` define i64 @julia_myadd_2022(i64 signext top: ; ↳ @ operators.jl:587 within `+` @ int. %3 = add i64 %1, %0 %4 = add i64 %3, %2 ret i64 %4 ; ↳ } julia> @code_native myadd(1,2,3)
Native	The final machine code output, specific to the target CPU architecture. This is at the same level as Assembly language.	<code>.section __TEXT,__text,re</code> .build_version macos, 14, 0 .globl _julia_myadd_1851 .p2align 2 _julia_myadd_1851: ; ↳ @ REPL[7]:1 within `myadd` ; %bb.0: ; ↳ @ operators.jl:587 within `+` @ int. add x8, x1, x0 add x0, x8, x2 ret ; LL

8. Hardware and Its Implications

8.2.3.1. Increasing Complexity in Search of Performance

Transistors are the building-block that creates the CPU and enables the physical process which governs the computations. For a very long time, the major source of improved computer performance was simply to make smaller transistors, allowing more of them to be packed together to create computer chips. This worked for many years and the propensity for the transistor count to double about every two years. In this way, software performance improvements came as side effect of the phenomenal scaling in hardware capability. However, raw single core performance and clock frequency (CPU cycle speed) dramatically flattened out starting a bit before the year 2010. This was due to the fact that transistor density has been starting to be limited by:

1. Pure physical constraints (transistors can be measured in width of atoms) where we have limited ability to manufacture something so small.
2. Thermodynamics, where heat can't be removed from the CPU core fast enough to avoid damaging the core and therefore operations per second are capped.

To obtain increasing performance, two main strategies have been employed in lieu of throwing more transistors into a single core:

1. Utilize multiple, separate cores and operate in an increasingly parallel way.
2. Use clever tricks to predict, schedule, and optimize the computations to make better use of the memory pipeline and otherwise idle CPU cycles.

We will cover techniques to utilize concurrent/parallel processing in Chapter 11. As for the second technique, it is capable of very impressive accelerations (on the order 2x to 100x faster than a naive implementation). However, it has sometimes caused issues. There have been some famous security vulnerabilities such as Spectre and Meltdown, which exploited speculative execution – a technique used to optimize CPU performance which will execute code *before being explicitly asked to* because the scheduler *anticipates* the next steps (with very good, but imperfect accuracy).

8.2.4. Logistics Warehouse Analogy

The problem is analogous to a logistics warehouse (persistent data) which needs to package up orders (processor instructions). There's a conveyor belt of items being constantly routed to the packaging station. In order to keep the packing station working at full capacity, the intermediate systems (RAM & CPU caches) are funneling items they *think* will be needed to the packager (data that's *expected* to be used in the processor). Most of the time, the necessary item (data) is optimally brought to the packaging station(process), or a nearby holding spot (CPU cache).

This system has grown very efficient, but sometimes the predictions miss or a never-before-ordered item needs to be picked from the far side of the warehouse and this

causes significant delays to the system. Sometimes a package will start to be assembled before the packager has even gotten to that order (branch prediction) which can make the system faster most of the time, but if the predicted package isn't actually what the customer ordered, then the work is lost and has to be redone (branch mispredict).

There are a lot more optimizations along the way:

- Since the items are already mostly arranged so that related items are next to each other, the conveyor belt will bring nearby items at the same time it brings the requested item (memory blocks).
- If an item usually ordered after another one is, the conveyor system will start to bring that second item as soon as the first one is ordered (prefetching).
- Different types of packaging stations might be used for specialized items (e.g. vector processing or cryptography instructions in the CPU).

8.2.5.

8.2.6. Speed of Computer Actions

Operation	Time (ns)	Distance Light Traveled
Single CPU Cycle (e.g. one ADD or OR operation on a register)	0.3	9 centimeters
L1 cache reference	1	30 centimeters
Branch mispredict	5	150 centimeters
L2 cache reference	5	150 centimeters
Main memory reference	100	30 meters
Read 1MB sequentially from RAM	250,000	75 km (~2 marathons)
Round trip within a datacenter	500,000	150km (the thickness of Earth's atmosphere)
Read 1MB sequentially from SSD	1,000,000	300km (distance Washington D.C. to New York City)
Hard disk seek	10,000,000	3,000km (width of continental United States)
Send packet CA->Netherlands->CA	150,000,000	45,000km (circumference of earth)

Source: <https://cs.brown.edu/courses/csci0300/2022/assign/labs/lab4.html>

9. Writing Performant Single-Threaded Code

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away. - Antoine de Saint-Exupéry, Airman's Odyssey

9.1. In this section

...

9.2. Introduction

With today's hardware, the highest throughput computations utilize GPUs for massive parallelization. However, writing parallel code, let alone performant parallel code, typically relies heavily on understanding patterns of non-parallel code. Secondly, many problems are not massively parallelizable and a primarily sequential model architecture is required. For these reasons, it's critical to understand sequential patterns before moving onto parallel code.

Further, with many high level languages the *only* way to achieve reasonable runtime is to utilize libraries facilitating parallelism. With fast languages like Julia, it may be that a simpler, easier to maintain sequential model is preferable to a more complex parallel version as the speed requirements may be met without parallelization. Like the quote that opened the chapter, you may prefer a simpler sequential version of a model to a more complex parallel one.

Developer time (your time) is often more expensive than runtime, so be prepared to accept a good-enough but sub-optimal code instead of spending a lot of time optimizing every last nanosecond out of it!

9. Writing Performant Single-Threaded Code

9.3. Patterns of Performant Sequential Code

9.3.1. Minimize Memory Allocations

Allocating memory onto the Heap takes a lot more time than (1) not using intermediate memory storage at all, or (2) utilizing the Stack. Each allocation requires time for memory management and requires the garbage collector, which can significantly impact performance, especially in tight loops or frequently called functions.

i Note

Tight loops or **hot loops** are the performance critical section of the code that are performed many times during a computation. They are often the “inner-most” loop of a nested loop algorithm.

In Julia, a general rule of thumb is that dynamically sizable or mutable objects (arrays, mutable structs) will be heap allocated while small fixed size objects can be stack allocated. For mutable objects, a common technique is to pre-allocate an array and then re-use that array for subsequent calculations. In the following example, note how we pre-allocate the output vector instead of creating vectors for each bond *and then* summing the vectors together at the end:

```
end_time = 10
cashflow_output = zeros(end_time)

par_bonds = map(1:1000) do i
    (tenor=rand((3, 5, 10)), rate=rand() / 10)
end

for asset in par_bonds
    for t in 1:end_time
        if t == asset.tenor
            cashflow_output[t] += 1 + asset.rate
        else
            cashflow_output[t] += asset.rate
        end
    end
end
cashflow_output

10-element Vector{Float64}:
 49.841737407128015
 49.841737407128015
```

9.3. Patterns of Performant Sequential Code

```
385.8417374071286
49.841737407128015
358.8417374071278
49.841737407128015
49.841737407128015
49.841737407128015
49.841737407128015
49.841737407128015
404.8417374071284
```

Julia's `@allocated` macro will display the number of bytes allocated by an expression, helping you identify and eliminate unnecessary allocations.

```
random_sum() = sum([rand() for _ in 1:10])
@allocated random_sum()
```

144

9.3.2. Optimize Memory Access Patterns

Optimizing memory access patterns is essential for leveraging the CPU's cache hierarchy effectively. Modern CPUs have multiple levels of cache (L1, L2, L3), each with different sizes and access speeds. By structuring your code to access memory in a cache-friendly manner, you can significantly reduce memory latency and improve overall performance.

What is cache-friendly memory access? Essentially it boils down to spatial and temporal locality.

9.3.2.1. Spatial Locality

Spatial locality refers to accessing data that is physically near each other in memory (e.g contiguous blocks of data in an array).

For example, it's better to access data in a linear order rather than random order. For example, if we sum up the elements of an array in order it will be significantly faster than if we do it randomly:

```
using BenchmarkTools, Random

# Create a large array of structs to emphasize memory access patterns
struct DataPoint
    value::Float64
    # Add padding to make each element 64 bytes (a typical cache line size)
```

9. Writing Performant Single-Threaded Code

```
padding::NTuple{7,Float64}
end

function create_large_array(n)
    [DataPoint(rand(), tuple(rand(7)...)) for _ in 1:n]
end

# Create a large array
const N = 1_000_000
large_array = create_large_array(N)

# Function for sequential access
function sequential_sum(arr)
    sum = 0.0
    for i in eachindex(arr)
        sum += arr[i].value
    end
    sum
end

# Function for random access
function random_sum(arr, indices)
    sum = 0.0
    for i in indices
        sum += arr[i].value
    end
    sum
end

# Create shuffled indices
shuffled_indices = shuffle(1:N)

# Benchmark
println("Sequential access:")
@btime sequential_sum($large_array)

println("\nRandom access:")
@btime random_sum($large_array, $shuffled_indices)
```

```
Sequential access:
859.208 μs (0 allocations: 0 bytes)
```

```
Random access:
```

9.3. Patterns of Performant Sequential Code

```
3.070 ms (0 allocations: 0 bytes)
```

```
500660.6135955673
```

When the data is accessed in a linear order, it means that the computer can load chunks of data into the cache and it can operate on that cached data for several cycles before new data needs to be loaded into the cache. In contrast, when accessing the data randomly, then the cache frequently needs to be populated with a different set of bits from a completely different part of our array.

9.3.2.1.1. Column vs Row Major Order

All multi-dimensional arrays in computer memory are actually stored linearly. When storing the multi-dimensional array, an architectural decision needs to be made at the language-level and Julia is column-major, similar to many performance-oriented languages and libraries (e.g. LAPACK, Fortran, Matlab). Values are stored going down the columns instead of across the rows.

For example, this 2D array would be stored as [1,2,3,...] in memory, which is made clear via `vec` (which turns a multi-dimensional array into a 1D vector):

```
let
    array = [
        1 4 7
        2 5 8
        3 6 9
    ]
    vec(array)
end

9-element Vector{Int64}:
 1
 2
 3
 4
 5
 6
 7
 8
 9
```

9. Writing Performant Single-Threaded Code

When working with arrays, prefer accessing elements in column-major order (the default in Julia) to maximize spatial locality. This allows the CPU to prefetch data more effectively.

You can see how summing up values across the first (column) dimension is much faster than summing across rows:

```
let
```

```
    @btime sum(arr, dims=1) setup = arr = rand(1000, 1000)
    @btime sum(arr, dims=2) setup = arr = rand(1000, 1000)
end
```

```
140.750 μs (1 allocation: 8.00 KiB)
141.542 μs (5 allocations: 8.08 KiB)
```

```
1000×1 Matrix{Float64}:
510.9133573317914
483.13346342458965
499.80418511254595
501.81252618126103
487.28972442309106
496.73115779696684
503.9279080555015
500.16512357407214
493.3758161779059
504.88080095885636
497.0208816155849
480.6332551313387
502.88904781813926
:
494.7730650191918
500.9987544481897
493.2872673955195
497.85748932138483
480.7698068269132
500.009279177882
497.2378695446579
483.5437585558339
498.2471106657064
492.7435818227162
498.84224615275076
511.8006559647578
```

9.3.2.2. Temporal Locality

The scheduler and branch prediction will recognize data that is accessed together closely in time and prefetch relevant blocks of data. This is an example of keeping “hot” data more readily accessible to the CPU than “cold” data.

Sometimes this happens at the operating system level - if you load a dataset that exceeds the available RAM, some of the “active” memory will actually be kept in a space on the persistent disk (e.g. your SSD) to avoid the computer crashing from being out of memory. Segments of the data that have been accessed recently will be in RAM while sections of the data not recently accessed are likely to be in the portion stored on the persistent disk.

9.3.3. Use Efficient Data Types

The right data type can lead to more compact memory representations, better cache utilization, and more efficient CPU instructions. This is another case of where having a smaller memory footprint allows for higher utilization of the CPU since computers tend to be memory-constrained in speed.

On some CPUs, you may find performance if use the smallest data type that can accurately represent your data. For example, prefer Int32 over Int64 if your values will never exceed 32-bit integer range. For floating-point numbers, use Float32 instead of Float64 if the reduced precision is acceptable for your calculations. These smaller types not only save memory but also allow for more efficient vectorized operations (see `?@sec-parallelism`) on modern CPUs. Sometimes using the smaller datatype isn’t beneficial: if you have a 64-bit architecture machine then the overhead of converting to/from 64-bit numbers may outweigh any speedup from higher memory throughput.

For collections, choose appropriate container types based on your use case. Arrays are efficient for calculations that loop through all or most elements, while Dictionaries are better for sparse look-ups or outside of the “hot loop” portion of a computation.

Consider using StaticArrays for small, fixed-size arrays, as they can be allocated on the stack and lead to better performance in certain scenarios than dynamically sizeable arrays. The trade-off is that the static arrays require more up-front compile time and after a certain point (length in the 50-100 element range) it usually isn’t worth trying to use StaticArrays.

9.3.4. Avoid Type Instabilities

Type instabilities occur when the compiler cannot infer a single concrete type for a variable or function return value. These instabilities can significantly hinder Julia’s ability to generate optimized machine code, leading to performance degradation. When the

9. Writing Performant Single-Threaded Code

compiler is not able to infer the types at compile-time (**compile time dispatch**), then while the program is running a lookup needs to be performed to find the most appropriate functions for the given type (**runtime dispatch**). Preferably, when the types are known at compile-time, Julia is able to create machine code that will point directly to the desired function instead of needing to perform that lookup.

To avoid type instabilities, ensure that functions have consistent return types across all code paths. For example:

```
function multiple_return_types()
    if rand() > 0.5
        return 1.0 # Float
    else
        return 2    # Int
    end
end

function single_return_type()
    if rand() > 0.5
        return 1.0 # Float
    else
        return 2.0 # Float
    end
end

@btime sum(multiple_return_types() for _ in 1:1000)
@btime sum(single_return_type() for _ in 1:1000)

5.201 μs (0 allocations: 0 bytes)
814.976 ns (0 allocations: 0 bytes)
```

1478.0

Note that having heterogeneous types as above is not the same thing as **type instability**, which is when Julia cannot determine in advance what the data types will be. In the example above, the return type is not unstable: the compiler recognizes that the single parametric type `Union{Float64, Int64}` will be returned., even though two different types can be returned. When the types cannot be determined by the compiler, it leads to runtime dispatch.

The following function illustrates a common anti-pattern wherein an `Any` typed array is created and then elements are added to it. Because any type can be added to an `Any` array (we happen to just add floats to it) then Julia's not sure what types to expect inside the container and therefore has to determine it at runtime.

9.3. Patterns of Performant Sequential Code

```
function unstable_function()
    values = [] # Implicitly Any[]
    push!(values, 1.0, 2.0)
    maximum(values)
end

unstable_function (generic function with 1 method)
```

Employ Julia's `@code_warntype` macro to identify type instabilities in your code. Code that cannot be inferred will be annotated with red text showing the point of the code that will return unstable types. In this Quarto document, the informative colors will not render with the way it does in a coding environment, but look for the `:: Any` annotation which indicates the compiler could not identify the type to return.

```
@code_warntype unstable_function()

MethodInstance for unstable_function()
  from unstable_function() @ Main In[9]:1
Arguments
  #self#::Core.Const(unstable_function)
Locals
  values::Vector{Any}
Body ::Any
1 -      (values = Base.vect())
|      Main.push!(values, 1.0, 2.0)
|      %3 = Main.maximum(values)::Any
└      return %3
```

More advanced exploration of type inference can be had via `Cthulhu.jl`¹, where you can interactively (via the REPL) navigate through different lines of code to explore inferred code.

When working with parametric types, look to avoid usage of generic type parameters (e.g. `Array{Any}`) whenever possible. For custom types, make use of parametric types to create type-stable abstractions. For example, the latter `struct Bond2` or `struct Bond3` will allow Julia to create distinct concrete types and methods as opposed to needing generic runtime dispatch due to the unpredictable potential types of the `struct` fields. The difference between `Bond2` and `Bond3` is that the fields in `Bond3` could be any type (such as `Float16` or `String`) as long as both of them equaled the type variable `T`.

¹So-named for the “slow descent into madness” when descending into functions to follow the Julia compiler’s type inference across many layers of function calls.

9. Writing Performant Single-Threaded Code

```
struct Bond1
    par
    coupon
end

struct Bond2
    par::Float64
    coupon::Float64
end

struct Bond3{T}
    par::T
    coupon::T
end
```

9.3.5. Optimize for Branch Prediction

Modern CPUs use branch prediction to speculatively execute instructions before knowing the outcome of conditional statements. Optimizing your code for branch prediction can significantly improve performance, especially in tight loops or frequently executed code paths.

To optimize for branch prediction:

1. Structure your code to make branching patterns more predictable. For instance, in if-else statements, put the more likely condition first. This allows the CPU to more accurately predict the branch outcome.
2. Use loop unrolling to reduce the number of branches. This technique involves manually repeating loop body code to reduce the number of loop iterations and associated branch instructions. See Section 11.4 for more on what this means.
3. Consider using Julia's `@inbounds` macro to eliminate bounds checking in array operations when you're certain the accesses are safe. This reduces the number of conditional checks the CPU needs to perform.
4. For performance-critical sections with unpredictable branches, consider using branch-free algorithms or bitwise operations instead of conditional statements. This can help avoid the penalties associated with branch mispredictions.
5. In some cases, it may be beneficial to replace branches with arithmetic operations (e.g., using the ternary operator or multiplication by boolean values) to allow for better vectorization and reduce the impact of branch mispredictions.

Here's an example demonstrating the impact of branch prediction:

```

function sum_if_positive(arr)
    sum = 0.0
    for x in arr
        if isodd(x)
            sum += x
        else
            sum -= x
        end
    end
    sum
end

# Benchmark
arr = rand(Int, 1_000_000)
arr_mostly_odds = fill(3, 999_999)
push!(arr_mostly_odds, 2) # add one even to get to 1M elements
@btime sum_if_positive($arr)
@btime sum_if_positive($arr_mostly_odds);

735.291 μs (0 allocations: 0 bytes)
685.291 μs (0 allocations: 0 bytes)

```

In this example, having consistently seen odd numbers means that the CPU will predict that the branch that will be used is the `sum += x` branch of the `if` statement.

Remember that optimizing for branch prediction often involves trade-offs. The benefits can vary depending on the specific hardware and the nature of your data. If performance critical, profile your code to ensure that your optimizations are actually improving performance in your specific use case. Overoptimizing on one set of hardware (e.g. local computer) may not translate the same on another set of hardware (e.g. server deployment).

9.3.6. Further Reading

- What scientists must know about hardware to write fast code
- Optimizing Serial Code, ScIML Book

10. Debugging and Performance Measurement

“Debugging is twice as hard as writing the code in the first place. Therefore, if you write the code as cleverly as possible, you are, by definition, not smart enough to debug it.” Brian Kernighan

10.1. Setup

```
using BenchmarkTools  
using Logging  
using Infiltrator
```

10.2. Benchmarking

Benchmarking is the process of evaluating performance by comparing it to a standard or a set of best practices. Julia provides handy benchmarking tools for performance measurement.

```
function example_function(n)  
    sum = 0  
    for i in 1:n  
        sum += i^2  
    end  
    return sum  
end  
  
# Benchmark the function  
@btime example_function(1000)  
@btime example_function(2000)  
  
1.250 ns (0 allocations: 0 bytes)  
1.250 ns (0 allocations: 0 bytes)
```

10. Debugging and Performance Measurement

2668667000

10.3. Debugging

Debugging in Julia involves identifying and fixing errors or bugs in your code. Julia provides several tools and techniques to aid in debugging.

- Using print and println Statements

The simplest form of debugging involves adding print or println statements to output variable values or checkpoints in your code. This helps track the flow and state of execution.

```
function add(a, b)
    println("a: ", a)
    println("b: ", b)
    result = a + b
    println("result: ", result)
    return result
end

add(3, 5)
```

```
a: 3
b: 5
result: 8
```

8

- Using the Debugger Package

Julia has a powerful debugger package called Debugger.jl, which allows us to step through code, set breakpoints, and inspect variables. Note that this needs to be run in an interactive session (e.g. the REPL or Jupyter).

```
function example_function(x)
    y = x * 2
    @enter begin
        z = y + 1
        return z
    end
end

example_function(5)
```

10.3. Debugging

When the debugger is started, we can interactively inspect variables, step through the code, and continue execution.

- Commands Some common commands in the debugger are:

- n (Next): Step to the next line of code.
- s (Step): Step into the function call.
- c (Continue): Continue execution until the next breakpoint or the end of the program.
- p (Print): Print the value of a variable.

- Using (`debug?`) Logging

The Logging module in Julia allows us to log messages at different levels (e.g., (`debug?`), (`info?`), (`warn?`), (`error?`)). This is useful for more controlled logging compared to print statements.

```
function compute(x)
    @debug "Starting computation with x = $x"
    y = x^2
    @debug "Computed y = $y"
    return y
end

compute(10)
```

100

To see debug logs, it is necessary to configure the logging level:

```
global_logger(ConsoleLogger(stderr, Logging.Debug))
```

- Using Infiltrator for Interactive Debugging

The Infiltrator.jl package provides another approach to interactive debugging by letting you insert debugging sessions into your code.

```
function compute(x)
    @infiltrate
    y = x^2
    return y
end

compute(10)
```

10. Debugging and Performance Measurement

Infiltrator.jl needs a fully-functional Julia REPL.

100

When execution reaches the (**infiltrate?**) line, it will drop us into an interactive prompt where we can inspect variables and execute commands.

- Error Messages and Stack Traces

Julia's error messages and stack traces can be quite informative. When an error occurs, Julia provides a traceback that shows the function call stack leading to the error, which helps in identifying where things went wrong.

```
function faulty_function(x)
    return 1 / x
end

faulty_function(0) # This will raise a `DivisionError`
```

Inf

The traceback will show us the sequence of function calls that led to the error.

Debugging in Julia involves a mix of strategies, including using print statements, the Debugger package for step-by-step inspection, logging with the Logging module, and interactive debugging with Infiltrator. These tools and techniques can help you identify and fix issues in our code efficiently.

150

11. Parallelization

Quote TBD

11.1. In this section

Fundamentals of parallel workloads, different mechanisms to distribute work: vectorization, multi-threading, GPU, and multi-device workflows. Different programming models: map-reduce, arrays, and tasks.

11.2. Amdahl's Law and the Limits of Parallel Computing

An important ground-truth in computing is that there is an upper limit to how fast a workload can be sped up through distributing the workload among multiple processor units. For example, if there is a modeling workload wherein 90% of the work is independent (say policy or asset level calculations) and the remaining 10% of the workload is an aggregate (say company or portfolio level), then the theoretical maximum speedup of the process is 10x faster (1 / 90% parallelizable load). This is captured in a law known as **Amdahl's Law** and it reflects the *theoretical* maximum speedup a workload could see. In practice, the speedup is worse than this due to overhead of moving data around, scheduling the tasks, and aggregating results. This is why in many cases a good effort in sequential workloads (see @#sec-performance-single) is often a more fruitful effort than trying to parallelize some workloads.

That said, there are still many modeling use-cases for parallelization. Modern investment and insurance portfolios can easily contain 100's of thousands or millions of seriatim holdings. In many cases, these can be evaluated independently, though on the often times there is interaction with the total portfolio (contract dividends, non-guaranteed elements, profit sharing, etc.). Further, even if the holdings are not parallelizable across the holdings dimension, we are often interested in independent evaluations across economic scenarios which is amendable to parallelization.

$$S(n) = \frac{1}{(1-p) + \frac{p}{n}}$$

Where:

11. Parallelization

- $S(n)$ is the theoretical speedup of the execution of the whole task
- n is the number of processors
- p is the proportion of the execution time that benefits from improved resources

We can visualize this for different combinations of p and n in Figure 11.1.

```
using CairoMakie

function amdahl_speedup(p, n)
    return 1 / ((1 - p) + p / n)
end

function main()
    fig = Figure(size=(800, 600))
    ax = Axis(fig[1, 1],
              title="Amdahl's Law",
              xlabel=L"Number of processors ($n$)",
              ylabel="Speedup",
              xscale=log2,
              xticks=2 .^ (0:16),
              xtickformat=x → "2^" .* string.(Int.(log.(2, x))),
              yticks=0:2:20
    )
    n = 2 .^ (0:16)
    parallel_portions = [0.5, 0.75, 0.9, 0.95]
    linestyles = [:solid, :dash, :dashdot, :solid]
    for (i, p) in enumerate(parallel_portions)
        speedup = [amdahl_speedup(p, ni) for ni in n]
        lines!(ax, n, speedup, label="$(Int(p*100))%", linestyle=linestyles[i])
    end
    xlims!(ax, 1, 2^16)
    ylims!(ax, 0, 20)
    axislegend(ax, L"Parallel portion ($p$)", position=:lt)
    fig
end

main()
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `S
└ @ Makie ~/julia/packages/Makie/GtFuI/src/scenes.jl:227
```

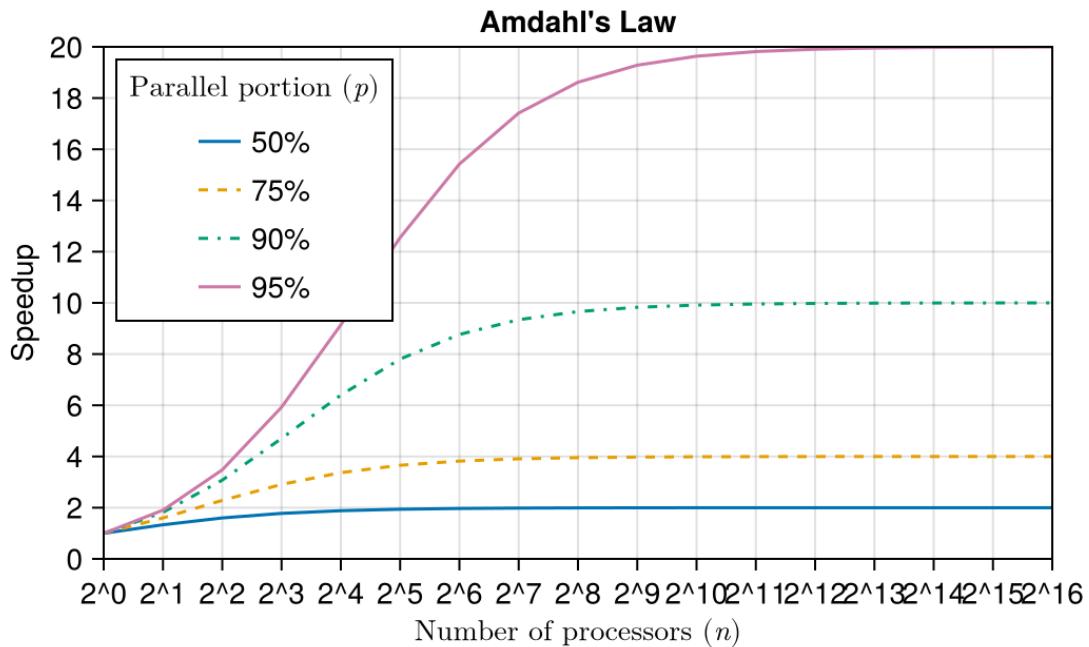


Figure 11.1.: Theoretical upper bound for speedup of a workload given the parallelizable portion p and number of processors n .

With this understanding, we will be able to set expectations and analyze the benefit of parallelization.

11.3. Types of Parallelism

Parallel processing comes in different flavors and is related to the details of hardware as discussed in Chapter 8. We will necessarily extend the discussion of hardware here, as parallelization is (mostly) inextricably tied to hardware details (we will revisit this in Section 11.8).

11. Parallelization

Table 11.1.: Major types of computational parallelism highlighting their key characteristics, advantages, and potential drawbacks.

Type	Description	Strengths	Weaknesses
Vectorization (SIMD)	Performs same operation on multiple data points simultaneously	Efficient for data-parallel tasks, uses specialized CPU instructions	Limited to certain types of operations, data must be contiguous
Multi-Threading	Executes multiple threads concurrently on a single CPU	Good for task parallelism, utilizes multi-core processors effectively	Overhead from thread management, potential race conditions
GPU	Uses graphics processing units (GPUs) for parallel computations	Excellent for massively parallel tasks, high throughput	Specialized programming required, data transfer overhead
Multi-Device / Distributed	Spreads computation across multiple machines or devices	Scales to very large problems, can use heterogeneous hardware	Complex to implement and manage, network latency issues

11.4. Vectorization

Vectorization in the context of parallel processing refers to special circuits within the CPU wherein the CPU will load multiple data units (e.g. 4 or 8 floating point numbers) in a contiguous block and perform the same instruction on them at the same time. This is also known as **SIMD, or Single-Instruction Multiple Data**.

The requirements for SIMD-able code are that:

- The intended section for SIMD is inside the inner-most loop.
- There are no branches (if-statements) inside the loop body.
 - Indexing an array is actually a possible branch, as two cases could arise: the index is inbounds or out-of-bounds. To avoid this, either use `for x in collection`, `for i in eachindex(collection)` or `for i in 1:n; @inbounds collection[i]` though the last of these is discouraged in favor of the prior, safer options.
 -

```

using BenchmarkTools

function prevent_simd(arr)
    sum = 0
    for x in arr
        if x > 0
            sum += x
        end
    end
    return sum
end

function allow_simd(arr)
    sum = 0
    for x in arr
        sum += max(x, 0)
    end
    return sum
end

let
    x = rand(10000)

    @btime prevent_simd($x)
    @btime allow_simd($x)
end

```

32.000 µs (0 allocations: 0 bytes)
44.375 µs (0 allocations: 0 bytes)

4978.1373721250375

In testing the above code, the `allow_simd` version should be several times faster than the `prevent_simd` example. The reason is that `prevent_simd` has a branch (`if x > 0`) where the behavior of the code may change depending on the value in `arr`. Conversely, the behavior of `allow_simd` is always the same in each iteration, no matter the value of `x`. This allows the compiler to generate vectorized code automatically.

Note that the compiler is able to identify vectorizable code in many cases, though through some cases may benefit from a more manual hint to the compiler through macro annotations (see `?@simd` for details).

11. Parallelization

Other types of parallelism that we will discuss in this chapter have some risk of errors or data corruption if not used correctly. SIMD isn't prone to issues like this because if the code is not SIMD-able then the compiler will not auto-vectorize the code block.

11.4.1. Hardware

Vectorization is hardware dependent. If the CPU does not support vectorization you will not see speedups from it. Many consumer and professional chips have AVX2 (Advanced Vector Extensions, with the 2 signifying second-generation 256 bit width, allowing four simultaneous 64-bit operations). The next generation is AVX512, having twice the SIMD capacity as AVX2. However, as of 2025 most consumer chips do not yet have that and commercial chips may not actually be faster than the AVX2 due to thermal restrictions (SIMD uses more power and generates more heat).

11.4.2. Additional Packages

Some additional packages to be aware of include:

- LoopVectorization.jl which can enhance the vectorized loops even further, such as handling the “tail” of a vectorized loops more efficiently than the base compiler. The “tail” refers to situations like where you have a vector width of 8, but don’t have a collection that’s a nice multiple of 8 (say 1001 elements).
- Octavian.jl implements a linear algebra-like library, utilizing parallelism via vectorization to generate efficient code for the system it’s running on.
- Tulio.jl is an einsum library, a domain-specific language for tensor (a specific subset of vectors) operations, common in machine learning and linear algebra.

11.5. Multi-Threading

11.5.1. Tasks

To understand multithreading examples, we first need to discuss **Tasks**, which are chunks of computation that get performed together, but after which the computer is free to switch to a new task. Technically, there are some instructions within a task that will let the computer pause and come back to that task later (such as `sleep`). Tasks are not, by themselves, allow for multiple computations to be performed in parallel. For example, one task might be loading a data file from persistent storage into RAM. After that task is complete, the computer continues on with another task in the queue (rendering a web page, playing a song, etc.). In this way even with a single processor and core, a computer could be “doing multiple things at once” (or “multi-tasking”)

even though nothing is running in parallel. The scheduling of the tasks is handled automatically by Julia and the operating system.

Here's an example of a couple of tasks where we write to an array. The second task actually writes to the array first, since we asked the first task to sleep (which allows the computer to yield to other tasks in the queue)¹.

```
let
    shared_array = zeros(5)

    task1 = @task begin
        sleep(1)
        shared_array[1] = 1

        println("Task 1: ", shared_array)
    end

    task2 = @task begin
        shared_array[2] = 2
        println("Task 2: ", shared_array)
    end

    schedule(task1); schedule(task2)
    wait(task1)
    wait(task2)

    println("Main: ", shared_array)
end
```

```
Task 2: [0.0, 2.0, 0.0, 0.0, 0.0]
Task 1: [1.0, 2.0, 0.0, 0.0, 0.0]
Main: [1.0, 2.0, 0.0, 0.0, 0.0]
```

11.5.1.1. Channels

Channels are a way to communicate data in an ordered way between tasks. You specify a type of data that the buffer will contain and how many elements it can hold. It then stores items (via `put!`) in a first-in-first-out (FIFO) queue, which can be popped off the queue (via `'take!'`) by other tasks.

¹Technically, it's possible that the second task doesn't write to the array first. This could happen if there's enough tasks (from our program or others on the computer) that saturate the CPU during the first task's `sleep` period such that the first task gets picked up again before the second one does.

11. Parallelization

Here's an example of a system which generates trades in the financial markets at random time intervals, and a monitoring tasks takes the results and tabulates running statistics:

```
let

    function trade_producer(channel,i)
        sleep(rand())
        profit = randn()
        put!(channel, profit)
        println("Producer: Trade Result #$i $(round(profit, digits=3))")
    end

    function portfolio_monitor(channel,n)
        sum = 0.0
        for _ in 1:n
            profit = take!(channel)
            sum += profit
            println("Monitor: Received $(round(profit, digits=3)), Cumulative profit: $(round(sum, digits=3))")
        end
    end

    channel = Channel{Float64}(32)

    # Start producer and consumer tasks
    @sync begin
        for i in 1:5; @async trade_producer(channel,i); end
        @async portfolio_monitor(channel,5)
    end

    # Close the channel and wait for tasks to finish
    close(channel)
end
```

- ① Random sleep between 0 and 1 seconds
- ② Generate a random number from standard normal distribution to simulate profit or loss from a trade
- ③ In this teaching example, we've limited the system to produce just five "trades". In practice, this could be kept running indefinitely via, e.g., `while true`.
- ④ Create a channel with a buffer size of 32 floats (in this limited example, we could have gotten away with just 5 since that's how many the demonstration produces). In practice, you want this to be long enough that the consumer of the channel never

11.5. Multi-Threading

gets so far behind that the channel fills up. The channel is created outside of the `@sync` block so that channel is in scope when we close it.

- ⑤ `@sync` waits (like `wait(task)`) for all of the scheduled tasks within the block to complete before proceeding with the rest of the program.
- ⑥ `@async` does the combination of creating a task via `@task` and scheduling in one, simpler call.

```
Producer: Trade Result #1 0.685
Monitor: Received 0.685, Cumulative profit: 0.685
Producer: Trade Result #3 -1.578
Monitor: Received -1.578, Cumulative profit: -0.893
Producer: Trade Result #4 -0.276
Monitor: Received -0.276, Cumulative profit: -1.169
Producer: Trade Result #2 0.011
Monitor: Received 0.011, Cumulative profit: -1.158
Producer: Trade Result #5 0.374
Monitor: Received 0.374, Cumulative profit: -0.784
```

This is really useful for handling events that are “external” to our program. If we were just doing a modelign excersise using static data, then we could control the order of processing and not need to worry about monitoring a volatile source of data. Nonetheless, tasks can still be useful in some cases even if a model is not using “live” data: for example if one of the steps in a model is to load a very large dataset, it may be possible to perform some computations while chunked task requests are queued to load more data from the disk.

while Julia’s garbage collector will eventually clean up unclosed channels, it’s a good practice to explicitly close them to ensure proper resource management, clear signaling of completion, and to avoid potential blocking or termination issues in your programs.

An additional thing to be aware of: if the task never finishes properly inside the `@sync`, then your program may get stuck in an infite loop and hang. Such as if one of the tasks never has a termination conidtion such as an upper bound on a loop, or a clear way to break out of a `while true` loop. While not different than a normal loop, such issues become less obvious underneath the layer of task abstractions.

The key takeaway for tasks is that it’s a way to chunk work into bundles that can be run in a concurrent fashion, even if nothing is technically being processed in parallel. The multi-threading and parallel programming paradigms sections build off of tasks so an understanding of tasks is helpful. However, some of the higher level libraries hide the task-based building blocks from you as the user/developer and so an intricate understanding of tasks is not required to be successful in parallelizing your Julia code.

11. Parallelization

11.5.2. Multi-Threading Overview

When a program starts on your computer, a **process** is created which is where the operating system allocates some overhead items (keeping track of the code and memory allocations and layout) and block of memory in RAM that can be utilized by that process. Different processes do not have access to each other's allocated memory.

i Note

Readers may be familiar with starting Excel in different processes. When Microsoft Excel is opened multiple times, in different processes, then the workbooks in each respective process do not share memory and cannot create links or use full copy/paste functionality between them. It's only when workbooks are opened within the same process that the workbooks may seamlessly talk to each other.

Within each process, a main thread is created. That thread is where the running of the code occurs. For the level of the discussion here, you can mainly think of a process as a container with shared memory for threads, which do the real work (as illustrated in Figure 11.2). Besides the main thread, other threads can be created within the process and access the same shared memory.

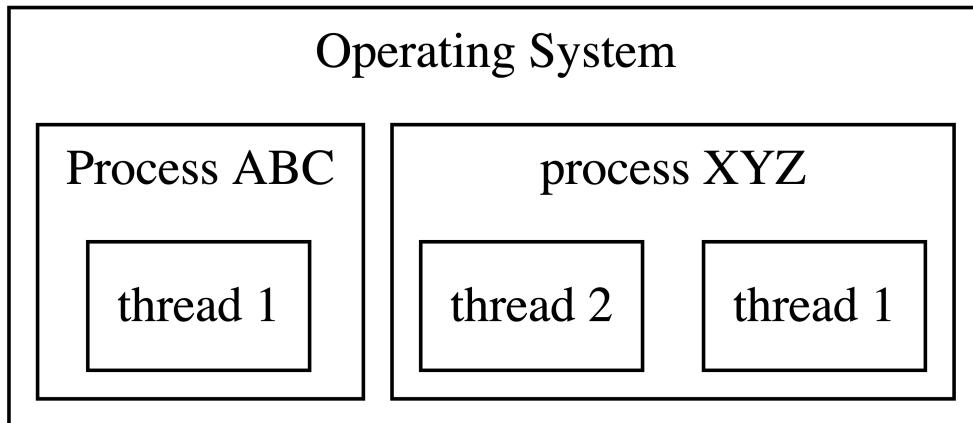


Figure 11.2.: When a program starts, the operating system creates a process for which multiple threads (a *main* thread plus optional additional threads) share memory.

The advantage of threads is that within a single physical processor, there may be multiple cores. Those cores can access the shared process memory and run tasks from different threads simultaneously. This is a technique that takes advantage of modern processor architecture wherein several (sometimes as many as 32 or more) cores exist on the

 Note
same chip.

Technically, there are different flavors of threading. While not critical for the understanding and modeling-focused discussion here, for completeness here is a bit more detail on different thread types.

- Recall that **Tasks** are chunks of computation that get performed together, but after which the computer is free to switch to a new task. For example, one task might be loading a data file from persistent storage into RAM. After that task is complete, the computer continues on with another task in the queue (rendering a web page, playing a song, etc.). In this way even with a single processor and core, a computer could be “doing multiple things at once” (or “multi-tasking”) even though nothing is running in parallel.
- **Operating System Threads** or just **Threads** are managed (as the name implies) at the operating system level. The benefit to this is that operating system level threads have more power: the operating system can pause or limit throughput on running programs if the operating system needs the resources for something it deems higher priority. It’s technically possible to use this power to force a higher priority for your own code, but Julia and many other languages do not offer creating of these types of threads in favor of the next type of threads. Operating system threads have a higher amount overhead (time and memory) involved in creating and destroying the threads.
- **Green threads, cooperative threads, fibers, or user-threads** are the type of threads that Julia provides. They are managed at the process (Julia) level and don’t have as much overhead in their creation as operating system threads. Also in Julia, a thread is implemented via Tasks

Parallelism in modern computing comes in many flavors, occurs at many different levels (hardware, OS, software, network), and has many different implementations of similar concepts. The terminology of threading in practice and online documentation is confusing and prone to confusion. If you are having a discussion or asking a question, feel free to take the time to ask for clarification on the terminology being used at a given point in time.

 Important

To use multi-threading, Julia needs to be started with more than one thread. This can be done by either setting the environment variable `JULIA_NUM_THREADS` to either auto or specify a number like 4. You can also specify how many threads to start julia with if given the `-t` command line argument (such as running `julia -t`

11. Parallelization

4 to start Julia with four threads from the command line). For the examples in this book and how many threads are used, see `?@sec-colophon` for the system settings. Why doesn't it automatically start with more than one thread? Between "hyper-threading" (synthetic additional thread capacity), multi-core architectures, and the different types of threads it's actually difficult to predict how many threads will be optimal for a given system. Julia's current default is to take the more conservative approach and start single-threaded unless otherwise specified. The "auto" option is a best-guess but can, on certain systems and configurations, be very bad for performance. The authors recommend for most common systems to just use "auto".

11.5.2.1. Multi-Threading Pitfalls

Different threads being able to access the same memory is a double-edged sword. It is useful because we do not need to create multiple copies of the data in RAM or in the cache² and can improve the overall throughput of our usually memory-bandwidth-limited machines. The downside is that if we are mutating the shared data for which our program relies upon, then our program may produce unintended results if the modification occurs carelessly. There are a couple of related issues to be aware of:

11.5.2.1.1. Race Conditions

The first issue is known as a **race condition**, which occurs when a block of memory has been read from or written to in an unintended order. For example, if we have two threads which are accumulating a sub-total, each process may read the running sub-total before the other thread has finished its update.

In the following example, we use the `Threads.@threads` to tell Julia to automatically distribute the work across threads.

```
function sum_bad(n)
    subtotal = 0
    Threads.@threads for i in 1:n
        subtotal += i
    end
    subtotal
end

sum_bad(100_000)
```

²There are some chips which do not have access to the same memory in a multi-threading context, and are known as non-uniform memory access (NUMA). These architectures work more like those in Section 11.7.

```
937554695
```

The result of this should be that `subtotal` doesn't equal `5000050000` because the different threads pick up the subtotal before the other thread finishes adding it's current value. As a result, when one of the threads finishes it's work another may be starting from a place that ignores the updated value.

11.5.2.2. Avoiding Multi-threading Pitfalls

We will cover several ways to manage multi-threading race conditions, but it is the recommendation of the authors to primarily utilize higher level library code, which will be demonstrated after covering some of the more basic, manual techniques.

11.5.2.2.1. Chunking up work into single-threaded work

In this example, we can correct `sum_bad` by splitting the work into different threads, each of which is independent. Then, we can aggregate the results of each of the chunks.

```
function sum_single(a)
    s = 0
    for i in a
        s += i
    end
    s
end
@btime sum_single(1:100_000)

1.250 ns (0 allocations: 0 bytes)
```

```
5000050000
```

Note that in the single-threaded case, Julia is able to identify this common pattern and use a shortcut, calculating the sum of the integers 1 through n as $\frac{n(n+1)}{2}$.

```
function sum_chunker(a)

chunks = Iterators.partition(1:a, a ÷ Threads.nthreads())
tasks = map(chunks) do chunk
    Threads.@spawn sum_single(chunk)
end
chunk_sums = fetch.(tasks)
```

11. Parallelization

```
    return sum_single(chunk_sums)

end

@btime sum_chunker(100_000)
```

1.154 μs (34 allocations: 2.52 KiB)

5000050000

11.5.2.2. Using Locks

Locks prevent memory from being accessed from more than one thread at a time.

```
function sum_with_lock(n)
    subtotal = 0
    lock = ReentrantLock()
    Threads.@threads for i in 1:n
        Base.@lock lock begin
            subtotal += i
        end
    end
    subtotal
end
@btime sum_with_lock(100_000)
```

6.975 ms (199483 allocations: 3.05 MiB)

5000050000

11.5.2.3. Using Atomics

Atomics are certain primitive values with a reduced set of operations for which Julia and the compiler can automatically create thread-safe code. This is often significantly faster than the context-switching overhead needed with locking and unlocking memory for threaded tasks. Compared with locks, atomics are simpler to implement and easier to reason about. The downside is that atomics are limited to the available primitive atomics types and methods.

```

function sum_with_atomic(n)
    subtotal = Threads.Atomic{Int}(0)
    Threads.@threads for i in 1:n
        Threads.atomic_add!(subtotal, i)
    end
    subtotal[]
end

@btime sum_with_atomic(100_000)

```

481.125 μ s (22 allocations: 2.16 KiB)

5000050000

11.6. GPU and TPUs

11.6.1. Hardware

Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs) are hardware accelerators for massively parallel computations. A TPU is very similar to a GPU but have special ability to handle data types and instructions that are more specialized for linear algebra operations; going forward we will simply refer to these types of accelerators as GPUs.

GPUs have similar components as the CPU as discussed in Chapter 8. They have RAM, caches for the cores, and cores that run the coded instructions on the data. The differences from a CPU are primarily:

- A GPU typically has thousands of cores while a CPU generally has single or double digit cores.
 - The cores typically operate at a *slower* clock speed than CPUs, relying on the sheer number of cores to perform computations faster.
- The GPU cores essentially have to be running the same set of instructions on all of the data, not unlike vectorization (Section 11.4).
 - GPU code is not suited for code with branching conditions (e.g. if statements) and so is more limited in the kinds of computations it can handle compared to the CPU.
- The RAM is typically much more constrained, typically less than a quarter of what primary RAM might be.

11. Parallelization

- As a result, GPUs may need strategies to move chunks of data to and from the GPU memory for moderately large datasets. Further, it's actually fairly common to use a lower-precision datatype (e.g. `Float16` or `Float32`) to improve overall program throughput at the cost of some precision.
- The caches are similar in concept to CPU, but unlike most CPU caches, there is locality to data wherein core #1 will have much quicker access to a different subset of data than, say, core #1024.
- A GPU is usually a secondary device of sorts: its physically and in device architecture is separate from the CPU which remains in charge of overall computer execution. To some extent, this is changing with some of the latest computer hardware. For example, the M-series of Apple processors have the CPU, GPU, and RAM in a single tightly integrated package for efficiency and computational power.
 - The implication of this (as with any movement of memory) is that there is overhead to moving data to and from the GPU. Your calculations will need to be in the single milliseconds range of time in order to start to see benefit from utilizing a GPU.

11.6.1.1. Notable Vendors and Libraries

Like the difference between x86 and ARM architectures, GPU also have specific architectures which vary by the vendor. To make full use of the hardware, the vendors need to (1) provide device drivers which allow the CPU to talk to the GPU, and (2) provide the libraries (lower level application programming interfaces, or APIs) which allow developers to utilize different hardware features without needing to write machine code. As of the mid 2020s, here are the most important GPU vendors and the associated programming library for utilizing their specific hardware:

Table 11.2.: Important GPU and TPU vendors and the associated library/interface.

Vendor	Hardware	API Library/Package
NVIDIA	Geforce, GTX/RTX, various Data Center focused hardware	CUDA
AMD	Radeon, various Data Center focused hardware	ROCm
Intel	Core, Xeon, Arc processors	OneAPI
Apple	M Series processors	Metal
Google	Tensor processors	TensorFlow

11.6.2. Utilizing a GPU

With some of the key conceptual differences between CPUs and GPUs explained, let's explore how to incorporate these powerful hardware accelerators.

11.6.2.1. Julia GPU Libraries

There's essentially two types of GPU programming we will discuss here:

1. Array-based programming can, in Julia, be fairly easily translated into code that will run on GPUs.
2. Generated kernels wherein higher level Julia code is written but is compiled into GPU-compatible code.

i Note

Kernels is a term used to describe explicit instructions for operating on data, as opposed to general code where the compiler can translate higher level functions into explicit instructions.

It is possible to explicitly write a GPU kernel in Julia or vendor API library (CUDA, Metal, etc.) but this is more advanced usage which will not be covered here.

Julia has wonderful support for several of the primary vendors (at the time of writing, CUDA, Metal, OneAPI, and ROCm) via the JuliaGPU organization. Installation of the required dependencies is also very straightforward and the interfaces at the array and generated kernel levels are very similar. The differences are obvious at the lower level vendor-API wrappers (which is the lower-level technique that will not be covered here).

The benefit of the consistency of the higher level libraries we will use here is that examples written for one of the types of accelerators will be largely directly translate-able to another. This is especially true for array programming, a little less so for the kernel style, and the least true for for the low level vendor-API wrapper functionality.

This book will be rendered on a Mac and therefore the examples will use Metal in order to run computational cells, however we'll show a CUDA translation for some of the examples in order to show the straight-forward nature of translating higher level GPU code in Julia is.

GPU API	GPU Array Type	Kernel Macro
CUDA	CuArray	@cuda
Metal	MtlArray	@metal
oneAPI	oneArray	@oneAPI

11. Parallelization

GPU API	GPU Array Type	Kernel Macro
ROCm	ROCArray	@roc

11.6.2.2. Array Programming on the GPU

First described in Section 5.5, array programming eschews writing loops and instead favors initializing blocks of heap-allocated memory and filling it with data to be operated on at a single point in time. While this is often not the most efficient way to utilize CPUs, it's essentially the required style of code to utilize GPUs.

For the example below, we will calculate the present value of a series of cashflows across a number of different scenarios. An explanation of the code is given below the example.

```
using Metal

function calculate_present_values!(present_values,cashflows, discount_matrices) ①
    # Perform element-wise multiplication and sum along the time dimension
    present_values .= sum(cashflows .* discount_matrices, dims=1) ②
end

# Example usage using 100 time periods, 100k scenarios
num_scenarios = 10^5
pvs = zeros(Float32,1,num_scenarios)
cashflows = rand(Float32, 100) ③
discount_matrices = rand(Float32, 100, num_scenarios) ④

# copy the data to the GPU
pvs_GPU = MtlArray(pvs)
cashflows_GPU = MtlArray(cashflows)
discount_matrices_GPU = MtlArray(discount_matrices) ⑤

@btime calculate_present_values!($pvs,$cashflows, $discount_matrices)
@btime calculate_present_values!($pvs_GPU,$cashflows_GPU, $discount_matrices_GPU) ⑥
```

- ① The function `calculate_present_values!` is written the same way as if we were just writing CPU code. Note that we are also passing a pre-allocated vector, `present_values` to store the result. This will allow us to isolate the performance of the computation, rather than including any overhead of allocating the array for the result.
- ② The code is broadcasted across the first dimension so that the single set of cashflows is discounted for each scenario's discount vector.

- ③ Metal only supports 32 bit floating point (some CUDA hardware will support 64 bit floating point)
- ④ Using 100 thousand scenarios for this example.
- ⑤ `MtlArray(array)` will copy the array values to the GPU.
- ⑥ Note that the data still lives on the GPU and is of the `MtlMatrix` (a type alias for a 2-D `MtlArray`).

```
[ Info: Precompiling Metal [dde4c033-4e86-420c-a63e-0dd931031962]
Precompiling SpecialFunctionsExt
  ✓ Metal → SpecialFunctionsExt
  1 dependency successfully precompiled in 5 seconds. 66 already precompiled.
[ Info: Precompiling SpecialFunctionsExt [05d8ebbe-653a-54ed-ba56-24759129d732]
  Warning: Module Metal with build ID fabfcfd-8fd8-00ae-0000-4f3b57ab4261 is missing from the cache
  This may mean Metal [dde4c033-4e86-420c-a63e-0dd931031962] does not support precompilation but it
  @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing SpecialFunctionsExt [05d8ebbe-653a-54ed-ba56-24759129d732]

2.960 ms (4 allocations: 38.53 MiB)
169.042 µs (587 allocations: 16.70 KiB)

1×100000 MtlMatrix{Float32, Private}:
 26.9773  28.4593  25.7268  27.7698 ... 25.4993  25.8076  28.0196  29.0483
```

The testing suggests approximately 200 times faster computation when performed on the GPU. Note however, that does not include the overhead of (1) moving the data to the GPU (in the initial `MtlArray(cashflows)` call), or (2) returning the data to the CPU (since the return type for the GPU version is `MtlArray`). We can measure this overhead by wrapping the data transfer inside another function and benchmarking it:

```
function GPU_overhead_test(present_values, cashflows, discount_matrices)
    pvs_GPU = MtlArray(present_values)
    cashflows_GPU = MtlArray(cashflows)                                     ⑤
    discount_matrices_GPU = MtlArray(discount_matrices)

    calculate_present_values!(pvs_GPU, cashflows_GPU, discount_matrices_GPU)

    Array(pvs_GPU) # convert to CPU array
end

@btime GPU_overhead_test($pvs,$cashflows,$discount_matrices)

9.112 ms (1024 allocations: 421.60 KiB)
```

11. Parallelization

```
1x100000 Matrix{Float32}:
 26.9773 28.4593 25.7268 27.7698 ... 25.4993 25.8076 28.0196 29.0483
```

With the additional overhead, the computation on the GPU takes more total time than if the work were done just on the CPU. This is a very simple example, and the balance tips heavily in favor of the GPU when:

1. The computational demands are significantly higher (e.g. we were to do more calculations than just a simple multiply/divide/sum).
2. The data size grows bigger.

Note

The previous example can be translated to CUDA by simply exchanging `MtLArray` for `CuArray`.

Warning

This example again underscores that hardware parallelization is not an automatic “win” for performance. A lot of uninformed discussion around modeling performance is to simply try to get things to run on the GPU and it is often *not* the case that the models will run faster. Further, as the modeling logic gets more complex, it does require greater care to keep in mind GPU constraints (acceptable data types, memory limitations, avoiding scalar operations, data transfer between CPU and GPU, etc.). A best practice is to contemplate sequential performance and memory usage before leveraging GPU accelerators.

11.6.2.3. Kernel Programming on the GPU

Another approach to GPU programming is often referred to as kernel programming, or being much more explicit about *how* a computation is performed. This is as opposed to the declarative approach in the array-oriented style (Section 11.6.2.2) wherein we specified *what* we wanted the computation to be.

The key ideas here are that we need to manually specify several aspects which came ‘free’ in the array-oriented style. The tradeoff is that we can be more fine-tuned about how the computation leverages our hardware, potentially increasing performance.

The GPU libraries in Julia abstract much of the low level programming typically necessary for this style of programming, but we still need to explicitly look at:

1. How the GPU will iterate across different cores/threads threads.

2. How many threads to utilize, the optimal number depends on the shape of the computation (long vectors, multi-dimensional arrays), memory constraints, and hardware specifics.

 - GPU threads: Individual units of execution within a kernel. Each thread runs the same kernel code but operates on a different portion of the data.

3. How to chunk (group) the data to distribute the data to the different GPU threads

Our strategy for the present values example will be to distribute the work such that different GPU threads are working on different scenarios. Within a scenario, the loop is a very familiar approach: initialize a subtotal to zero and then accumulate the calculated present values.

```
function calculate_present_values_kernel!(present_values,cashflows, discount_matrices)
    idx = thread_position_in_grid_1d()                                ①
    pv = 0.0f0
    for t in 1:size(cashflows, 1)
        pv += cashflows[t] * discount_matrices[t, idx]               ③
    end
    present_values[idx] = pv                                         ④
    return nothing
end                                                               ⑤
```

- ① As the work is distributed across threads, `thread_position_in_grid_1d()` will give the index of the current thread so that we can index data appropriately for the work as we decide to splitit up (we've split up thw work by scenario in this example).
- ② Recall that we are working with `Float32` on the GPU here, so the zero value is set via the `f0` notation indicating a 32-bit floating point number.
- ③ The loop is across timesteps within each thread, while the thread index is tracked with `idx`.
- ④ The result is written to the pre-allocated array of present values, and we avoid race conditions because the different threads are working on difference scenarios.
- ⑤ We don't explicitly have to `return nothing` here, but it makes it extra clear that the intention of the function is to mutate the `present_values` array given to it. This mutation intention is also signaled by the `!` convention in the function name.

```
calculate_present_values_kernel! (generic function with 1 method)
```

The kernel above was fairly similar to how we might write code for CPU-threaded approaches, but we now need to specify the technicals of launching this on the GPU. The

11. Parallelization

threads defines how many independent calculations to run at a given time, and the maximum will be dependent on the hardware used. The groups argument defines the number of threads that share memory and synchronize results together (meaning that group will wait for all threads to finish before moving onto the next chunk of data). The push-pull here is that threads that can share data avoid needing to create duplicate copies of that data in memory, but if there is variability in how long each calculation will take, then the waiting time for synchronizing results may slow the overall computation down.

Our task utilizes shared memory of the cashflows for each thread, so through some experimentation in advance, we find that a relatively large group size of ~512 is optimal.

We bring this all together through the use of the kernel macro @metal:

```
threads = 1024
groups = cld(num_scenarios, 512)

@btime @metal threads=$threads groups=$groups calculate_present_values_kernel!(
    $pvs_GPU,
    $cashflows_GPU,
    $discount_matrices_GPU
)

17.167 μs (171 allocations: 4.08 KiB)
```

```
Metal.HostKernel{typeof(calculate_present_values_kernel!), Tuple{MtlDeviceMatrix{Float32, 1}, M
```

This is approximately seven times faster than the array-oriented style above, meaning that the GPU kernel version's computation is over 1000 times faster than the CPU version. However, we saw previously that the cost of moving the data to the GPU memory and then back to the CPU memory was the biggest time sink of all - again we'd need to have more scale in the problem to make offloading to the GPU beneficial overall.

Note

The Metal GPUs are able to iterate threads across three different dimensions. In the prior example, we only used one dimension and thus used `thread_position_in_grid_1d()`. If we were distributing the threads across, say, three dimensions then we would use `thread_position_in_grid_2d()`.

How do you determine how many dimensions to use? A good approach is to mimic the data you are trying to parallelize. In the example of calculating a vector of present values across 100k scenarios, that was the primary 'axis of parallelization'. If instead of a one-dimensional set of cashflows (e.g. a single asset with fixed cashflows), we had a two-dimensional set of cashflows (e.g. a portfolio of many

assets), then we may find the best balance of code simplicity and performance to iterate across two dimensions of threads (but we are still limited by the same number of total available threads).

Note

The above example would be translated to CUDA by changing just a few things:

- The thread indexing would be `idx = threadIdx().x` instead of `i = threadIdx().x`
- The GPU arrays should be created with `CuArray` instead of `MtlArray`.
- The kernel macro would be `@cuda threads=1024 calculate_present_values_kernel!(...)` instead of `@metal threads=threads groups=groups calculate_present_values_kernel!(...)`. The memory sharing and synchronizing between threads is more manual than Metal.jl, but this is not strictly necessary for our example.

THIS CODE NEEDS TO BE TESTED ON A CUDA MACHINE:

```
function calculate_present_values_kernel!(present_values,cashflows, discount_matrices)
    idx = threadIdx().x

    pv = 0.0f0
    for t in 1:size(cashflows, 1)
        pv += cashflows[t] * discount_matrices[t, idx]
    end

    present_values[idx] = pv
    return nothing
end

groups = cld(num_scenarios, 512)

@cuda threads=threads calculate_present_values_kernel!(
    pvs_GPU,
    cashflows_GPU,
    discount_matrices_GPU
)
```

11.7. Multi-Processing / Multi-Device

Mutiple device, or **multi-device** computer refers to using separate groups of memory/processor combinations to accomplish tasks in parallel. This can be as simple as multiple distinct cores on within a single desktop computer, or many separate computers networked across the internet, or many processesors within a high performance cluster or a computing-as-a-service provider like Amazon Web Services or JuliaHub.

Everything discussed previously related to hardware (Chapter 8, Section 11.5, Section 11.6) continues to apply. The additional complexity is attempting to syncronize the computation across multiple sets of the same (homogenous) or different (heterogeneous) hardware.

As you might imagine, approaches to multi-device computing can vary widely. Julia's approach tries to strike the balance between capability and user-friendliness and uses a primary/worker model wherein one of the processors is the main coordinator while other processors are "workers". If only one processor is started, then the main processor is also a worker processor. This main/worker approach uses a "one-sided" approach to coordination. The main worker utilizes high level calls and the workers respond, with some of the communication and handoff handled by Julia transparently from the user's perspective.

A useful mental model is the asynchronous task-based concepts discussed in Section 11.5.1, as the main worker will effectively queue up work with the worker processors. Because there may be a delay associated with the computation or the communication between the processors, the worker runs asynchronously.

Description	Task API	Distributed Analogue
Create a new task	Task()	@spawnat
Run task asynchronously	@async	@spawnat
Retrieve task result	fetch	fetch
Wait for task completion	@sync	sync
Communication between tasks	channel	RemoteChannel

Adapting the trade producer and monitor example from above to run on multiple processors (see #sec-channels to review the base model and algorithm), we make a few key changes:

- using `Distributed` loads the `Distributed` standard Julia library, providing the interface for multi-processing across different hardware.
- `addprocs(n)` will add n worker processors (the main processor is already counted as one worker). When adding local machine processors, the processors are part of

11.7. Multi-Processing / Multi-Device

the local machine. This starts new Julia processes (you can see this in the task manager of the machine) which inherit the package environment (i.e. `Project.toml` and environment variables) from the main process; this does not occur automatically if not part of the same local machine.

- To add processors from other machines, see the Distributed Computing section of the Julia docs.
- `myid()` is the identification number of the given processor that's been spun up.
- We use a `RemoteChannel` instead of a `Channel` to facilitate communication across processors.
- Instead of `@async`, we use `@spawnat n` to create a task for processor number `n` (or `:any` will automatically assign a processor).

See

```
using Distributed
let

# Add worker processes if not already added
if nworkers() == 1
    addprocs(4) # Add 4 worker processes
end

@everywhere function trade_producer(channel, i)
    sleep(rand())
    profit = randn()
    put!(channel, profit)
    println("Producer $(myid()): Trade Result #$i $(round(profit, digits=3))")
end

@everywhere function portfolio_monitor(channel, n)
    sum = 0.0
    for _ in 1:n
        profit = take!(channel)
        sum += profit
        println("Monitor $(myid()): Received $(round(profit, digits=3)), Cumulative profit: $(sum)")(3)
    end
end

function run_distributed_simulation()
    channel = RemoteChannel(() → Channel{Float64}(32))(4)
    # Start producer and consumer tasks
    @sync begin(5)
```

11. Parallelization

```
for i in 1:5
    @spawnat :any trade_producer(channel, i)
end
@spawnat :any portfolio_monitor(channel, 5)
end

# Close the channel and wait for tasks to finish
close(channel)
end

# Run the simulation
run_distributed_simulation()
end
```

```
From worker 5: Producer 5: Trade Result #4 0.722
From worker 3: Monitor 3: Received 0.722, Cumulative profit: 0.722
From worker 3: Producer 3: Trade Result #2 1.428
From worker 3: Monitor 3: Received 1.428, Cumulative profit: 2.149
From worker 3: Monitor 3: Received -0.599, Cumulative profit: 1.551
From worker 4: Producer 4: Trade Result #3 -0.599
From worker 3: Monitor 3: Received 0.743, Cumulative profit: 2.293
From worker 3: Monitor 3: Received -0.117, Cumulative profit: 2.176
From worker 2: Producer 2: Trade Result #1 0.743
From worker 2: Producer 2: Trade Result #5 -0.117
```

Given the similarity to the single-process version, what's the motivation for this approach? A few differences:

- In this simplified example, we are simply starting additional Julia processes on the same machine. Like a threaded approach, the work will be split across the same multi-core processor. In this context, the main difference is that the processes do not share memory.
 - Communicating across processes generally has a little bit more overhead than communicating across threads.
- If distributing across machines, not sharing memory is advantageous if using multi-processing across machines that have their own memory stores which need not compete with the main process (such as distributing chunks of large datasets).
- The worker processors don't need to be the same architecture as the main processor, allowing usage of different machines or cloud computing that is communicating with a local main process.

11.8. Parallel Programming Models

The previous sections have explained the different parallel programming models and how to directly utilize them to harness additional computing power. Each approach (multi-threading, GPU, distributed processing, etc.) has unique considerations and tradeoffs. These approaches in Julia are generally much more accessible to beginning and intermediate users than other languages, but admittedly still requires a decent amount of thought and care.

It is possible, if you are willing to give up some fine-grained control, to utilize some higher level approaches which look to abstract away some of the particularities of the implementation.

11.8.1. Map-Reduce

Map-Reduce (Section 5.4.4) operations are inherently parallelizable and various libraries provide parallelized versions of the base `mapreduce`. This is the workhorse function of many ‘big data’ workloads and many statistical operations are versions of `mapreduce`.

11.8.1.1. Multi-Threading

11.8.1.1.1. OhMyThreads

`ThreadsX.jl` provides the threaded versions of essential functions such as `tmap`, `tmapreduce`, `tcollect`, and `tforeach` (see Table 5.1). In most cases, the chunking and data sharing is handled automatically for you.

```
import OhMyThreads
@btime OhMyThreads.tmapreduce(x → x, +, 1:100_000)

[ Info: Precompiling OhMyThreads [67456a42-1dca-4109-a031-0a68de7e3ad5]
[ Info: Precompiling InverseFunctionsUnitfulExt [f5f6e0dd-5310-5802-bcb2-1cb72ad693d4]
[ Info: Precompiling StatsFunsInverseFunctionsExt [da3fed98-1718-55bb-8128-3e4a2e454b06]
Precompiling AccessorsIntervalSetsExt
  ✓ Accessors → AccessorsIntervalSetsExt
  ✓ Accessors → AccessorsStaticArraysExt
2 dependencies successfully precompiled in 1 seconds. 17 already precompiled.
[ Info: Precompiling AccessorsIntervalSetsExt [727f68c9-d1d4-5b40-b284-36502e629768]
└ Warning: Module Accessors with build ID fabfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the
  This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation
└ @ Base loading.jl:1948
```

11. Parallelization

```
[ Info: Skipping precompilation since __precompile__(false). Importing AccessorsIntervalSetsExt [Precompiling AccessorsStructArraysExt
    ✓ Accessors → AccessorsStructArraysExt
    1 dependency successfully precompiled in 0 seconds. 27 already precompiled.
[ Info: Precompiling AccessorsStructArraysExt [deedf894-762e-575a-ad8b-1df4bba63293]
[ Warning: Module Accessors with build ID fafbfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the
| This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation
| @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing AccessorsStructArraysExt [Info: Precompiling AccessorsStaticArraysExt [91548973-bbcf-5128-ac3c-c8b871e934a5]
[ Warning: Module Accessors with build ID fafbfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the
| This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation
| @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing AccessorsStaticArraysExt [Precompiling AccessorsUnitfulExt
    ✓ Accessors → AccessorsUnitfulExt
    1 dependency successfully precompiled in 1 seconds. 22 already precompiled.
[ Info: Precompiling AccessorsUnitfulExt [0f33c9ce-b40b-5f58-839e-64dee873ac84]
[ Warning: Module Accessors with build ID fafbfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the
| This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation
| @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing AccessorsUnitfulExt [0f33c9ce-b40b-5f58-839e-64dee873ac84]
[ Precompiling BangBangChainRulesCoreExt
    ✓ BangBang → BangBangChainRulesCoreExt
    ✓ BangBang → BangBangTablesExt
    ✓ BangBang → BangBangStaticArraysExt
    3 dependencies successfully precompiled in 1 seconds. 39 already precompiled.
[ Info: Precompiling BangBangChainRulesCoreExt [47e8a63d-7df8-5da4-81a4-8f5796ea640c]
[ Warning: Module BangBang with build ID fafbfcfd-3113-422e-0000-4f4d4b2ede85 is missing from the
| This may mean BangBang [198e06fe-97b7-11e9-32a5-e1d131e6ad66] does not support precompilation
| @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing BangBangChainRulesCoreExt [Info: Precompiling BangBangStaticArraysExt [a9f1882a-14fa-573e-a12d-824431257a23]
[ Warning: Module BangBang with build ID fafbfcfd-3113-422e-0000-4f4d4b2ede85 is missing from the
| This may mean BangBang [198e06fe-97b7-11e9-32a5-e1d131e6ad66] does not support precompilation
| @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing BangBangStaticArraysExt [Precompiling BangBangStructArraysExt
    ✓ BangBang → BangBangStructArraysExt
    1 dependency successfully precompiled in 1 seconds. 37 already precompiled.
[ Info: Precompiling BangBangStructArraysExt [d139770a-8b79-56c4-91f8-7273c836fd96]
[ Warning: Module BangBang with build ID fafbfcfd-3113-422e-0000-4f4d4b2ede85 is missing from the
```

```

| This may mean BangBang [198e06fe-97b7-11e9-32a5-e1d131e6ad66] does not support precompilation b
└ @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing BangBangStructArraysExt [c
[ Info: Precompiling BangBangTablesExt [476361b5-ac10-5c09-8bec-30d098a22a5b]
┌ Warning: Module BangBang with build ID fafbfcfd-3113-422e-0000-4f4d4b2ede85 is missing from the
| This may mean BangBang [198e06fe-97b7-11e9-32a5-e1d131e6ad66] does not support precompilation b
└ @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing BangBangTablesExt [476361b

```

7.954 μs (31 allocations: 2.42 KiB)

5000050000

11.8.1.1.2. ThreadsX

ThreadsX.jl is built off of the wonderful Transducers.jl package, though the latter is a bit more advanced (more abstract, but as a result more composable and powerful). ThreadsX provides threaded versions of many popular base functions. It offers a wider set of readymade threaded functions, but has a much more complex codebase. For the vast majority of threading needs, OhMyThreads.jl should be sufficient and performant. See the documentation for all of the implemented functions, but for our illustrative example:

```
import ThreadsX
@btime ThreadsX.mapreduce(x → x, +, 1:100_000)
```

```

Precompiling ThreadsX
  ✓ MicroCollections
  ✓ Transducers
  ✓ Transducers → TransducersReferenceablesExt
  ✓ Transducers → TransducersOnlineStatsBaseExt
  ✓ ThreadsX
5 dependencies successfully precompiled in 3 seconds. 72 already precompiled.
[ Info: Precompiling ThreadsX [ac1d9e8a-700a-412c-b207-f0111f4b6c0d]
┌ Warning: Module BangBang with build ID fafbfcfd-3113-422e-0000-4f4d4b2ede85 is missing from the
| This may mean BangBang [198e06fe-97b7-11e9-32a5-e1d131e6ad66] does not support precompilation b
└ @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing ThreadsX [ac1d9e8a-700a-41
[ Info: Precompiling MicroCollections [128add7d-3638-4c79-886c-908ea0c25c34]
┌ Warning: Module Accessors with build ID fafbfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the
| This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation
└ @ Base loading.jl:1948

```

11. Parallelization

```
[ Info: Skipping precompilation since __precompile__(false). Importing MicroCollections [128add7c...]  
[ Info: Precompiling Transducers [28d57a85-8fef-5791-bfe6-a80928e7c999]  
[ Warning: Module Accessors with build ID fabfcfd-8e60-dfff-0000-4f4d5c05c3e0 is missing from the...  
| This may mean Accessors [7d9f7c33-5ae7-4f3b-8dc6-eff91059b697] does not support precompilation |  
| @ Base loading.jl:1948  
[ Info: Skipping precompilation since __precompile__(false). Importing Transducers [28d57a85-8fef...]  
[ Info: Precompiling TransducersReferenceablesExt [befac7fd-b390-5150-b72a-6269c65d7e1f]  
[ Warning: Module Transducers with build ID ffffffff-ffff-ffff-0000-4f534f3cbc85 is missing from the...  
| This may mean Transducers [28d57a85-8fef-5791-bfe6-a80928e7c999] does not support precompilation |  
| @ Base loading.jl:1948  
[ Info: Skipping precompilation since __precompile__(false). Importing TransducersReferenceablesExt [...]
```

13.208 µs (218 allocations: 16.48 KiB)

5000050000

11.8.1.2. Multi-Processing

`reduce(op, pmap(f, collection))` will use a distributed map and reduce the resulting map on the main thread. This pattern works well if each application of `f` to elements of `collection` is costly.

`@distributed (op) for x in collection; f(x); end` is a way to write the loop with the reduction `op` for which the `f` need not be costly.

The difference between the two approaches is that with `pmap`, `collection` is made available to all workers. In the `@distributed` approach, the collection is partitioned and only a subset is sent to the designated workers.

Here's an example of both of these, calculating a simple example of counting coin flips:

```
# this is a example of poor utilization of pmap, since the operation is  
# fast and the overhead of moving the whole collection dominates  
@btime reduce(+, pmap(x → rand((0,1)), 1:10^3))
```

93.937 ms (48569 allocations: 2.03 MiB)

483

180

```

function dist_demo()
    @distributed (+) for _ in 1:10^5
        rand((0,1))
    end
end

@btime dist_demo()

```

187.625 μs (300 allocations: 12.86 KiB)

49904

11.8.2. Array-Based

Array based approaches will often utilize the parallelism of SIMD on the CPU or many cores on the GPU. It's as simple as using generic library calls which will often be optimized at the compiler level. Examples:

```

let
    x = rand(Float32, 10^8)
    x_GPU = MtlArray(x)
    @btime sum($x)
    @btime sum($x_GPU)
end

6.367 ms (0 allocations: 0 bytes)
4.392 ms (878 allocations: 23.88 KiB)

```

5.0002308f7

`sum(x)` compiles to SIMD instructions on the CPU, while using the GPU array type in `sum(x_GPU)` is enough to let the compiler dispatch on the GPU type and emit efficient, parallelized code for the GPU.

Distributed array types allow for large datasets to effectively be partitioned across multiple processors, and have implementations in the `DistributedArrays.jl` and `Dagger.jl` libraries.

11. Parallelization

11.8.3. Loop-Based

Loops which don't have race conditions can easily become multi-threaded. Here, we have three versions of updating a collection to square the contained values:

```
let v = collect(1:10000)

for i in eachindex(v)
    v[i] = v[i]^2
end
v[1:3]
end
```

3-element Vector{Int64}:

```
1
4
9
```

Using multi-threading

```
let v = collect(1:10000)
Threads.@threads for i in eachindex(v)
    v[i] = v[i]^2
end
v[1:3]
end
```

3-element Vector{Int64}:

```
1
4
9
```

Using multi-processing:

```
let v = collect(1:10000)
Distributed.@distributed for i in eachindex(v)
    v[i] = v[i]^2
end
v[1:3]
end
```

```
3-element Vector{Int64}:
1
2
3
```

For more advanced usage, including handling shared memory see Section 11.7 and Section 11.5.

11.8.4. Task-Based

Task based approaches attempt to abstract the scheduling and distribution of work from the user. Instead of saying how the computation should be done, the user specifies the intended operations and allows the library to handle the workflow. The main library for this in Julia is Dagger.jl.

Effectively, the library establishes a network topology (a map of how different processor nodes can communicate) and models the work as a directed, acyclic graph (a DAG, which is like a map of how the work is related). The library is then able to assign the work in the appropriate order to the available computation devices. The benefit of this is most apparent with complex workflows or network topologies where it would be difficult to manually assign, communicate, and schedule the workflow.

Here's a very simple example which demonstrates Dagger waiting for the two tasks which work in parallel (we already added multiple processors to this environment in Section 11.7):

```
import Dagger

# This runs first:
a = Dagger.@spawn rand(100, 100)

# These run in parallel:
b = Dagger.@spawn sum(a)
c = Dagger.@spawn prod(a)

# Finally, this runs:
wait(Dagger.@spawn println("b: ", b, ", c: ", c))

[ Info: Precompiling Dagger [d58978e5-989f-55fb-8d15-ea34adc7bf54]
Precompiling TransducersOnlineStatsBaseExt
  ✓ Transducers → TransducersOnlineStatsBaseExt
  1 dependency successfully precompiled in 1 seconds. 73 already precompiled.
[ Info: Precompiling TransducersOnlineStatsBaseExt [6f45943c-d98e-5e8a-8912-adbe5bfabdb6]
```

11. Parallelization

```
[ Warning: Module Transducers with build ID ffffffff-ffff-ffff-0000-4f534f3cbc85 is missing from the cache  
[ This may mean Transducers [28d57a85-8fef-5791-bfe6-a80928e7c999] does not support precompilation  
[ @ Base loading.jl:1948  
[ Info: Skipping precompilation since __precompile__(false). Importing TransducersOnlineStatsBase  
Precompiling GraphVizSimpleExt  
    ✓ Dagger → GraphVizSimpleExt  
    1 dependency successfully precompiled in 16 seconds. 78 already precompiled.  
[ Info: Precompiling GraphVizSimpleExt [8f367522-86d3-5221-bdf5-df974a7b0ff8]  
[ Warning: Module Dagger with build ID fafbfcfd-fb66-bf3a-0000-4f5f8a7aae53 is missing from the cache  
[ This may mean Dagger [d58978e5-989f-55fb-8d15-ea34adc7bf54] does not support precompilation but  
[ @ Base loading.jl:1948  
[ Info: Skipping precompilation since __precompile__(false). Importing GraphVizSimpleExt [8f367522-86d3-5221-bdf5-df974a7b0ff8]  
  
b: 5021.100551617885, c: 0.0
```

11.9. References

- https://book.sciml.ai/notes/06-The_Different_Flavors_of_Parallelism/
- <https://docs.julialang.org/en/v1/manual/parallel-computing/>
- <https://enccs.github.io/julia-for-hpc/>

12. Julia Development Tips

12.1. In this section

Julia's trade-off between runtime speed and compile time, Revise and code re-loading, precompilation of packages, creating sysimages, and other techniques to improve developer and user experience.

Part IV.

Conceptual Foundations: Learning from Related Disciplines

13. Elements of Computer Science

“Fundamentally, computer science is a science of abstraction—creating the right model for a problem and devising the appropriate mechanizable techniques to solve it. Confronted with a problem, we must create an abstraction of that problem that can be represented and manipulated inside a computer. Through these manipulations, we try to find a solution to the original problem.” - Al Aho and Jeff Ullman (1992)

13.1. In this section

Adapting computer science concepts to work for financial professionals. Concepts like computability, computational complexity, the language of algorithms and problem solving.

13.2. Computer Science for Financial Professionals

Computer science as a term can be a bit misleading because of the overwhelming association with the physical desktop or laptop machines that we call “computers”. The discipline of computer science is much richer than consumer electronics: at its core, computer science concerns itself with areas of research and answering tough questions:

- **Algorithms and Optimization.** How can a problem be solved efficiently? How can that problem be solved *at all?* Given constraints, how can one find an optimal solution?
- **Information Theory.** Given limited data, what *can* be known or inferred from it?
- **Theory of Computation.** What sorts of questions are even answerable? Is an answer easy to compute or will resolving it require more resources than the entire known universe? Will a computation ever stop calculating?
- **Data Structures.** How to encode, store, and use data? How does that data relate to each other and what are the trade-offs between different representations of that data?

13. Elements of Computer Science

For a reader in the twenty-first century we hope that's it's patently obvious how impactful the *applied* computer science has been as an end-user of the internet, artificial intelligence, computational photography, safety control systems, etc., etc. have been to our lives. It is a testament to the utility of being able to harness some of the ideas of this science is. Many of the most impactful advances occur at the boundary between two disciplines. It's here in this chapter that we desire to bring together the financial discipline together with computer science and to provide the financial practitioner with the language and concepts to leverage some of computer science's most relevant ideas.

In this section, we will refer back to a problem called the travelling salesperson problem (TSP).

13.3. Algorithms & Complexity

Algorithms is a general term for a process that transforms an input to an output. It's the dirty, down-to-earth implementation of a mathematical function or process. Further, we should indicate that a process needs to be specified in sufficient detail to be able to call itself an algorithm versus a heuristic which does not indicate with enough detail how the process would unfold.

13.3.1. Computational Complexity

We can characterize the computational complexity of a problem by looking at how long an algorithm takes to complete a task when given an input of size n . We can then compare two approaches to see which is computationally less complex for a given n .

Note that computational complexity isn't quite the same as how fast an algorithm will run on your computer, but it's a very good guide. Modern computer architectures can sometimes execute multiple instructions in a single cycle of the CPU making an algorithm that is, on paper, slower than another actually run faster in practice. Additionally, sometimes algorithms are able to substantially limit the number of *computations* to be performed, at the expense of using a lot more *memory* and thereby trading CPU usage with RAM usage.

You can think of computational complexity as a measure of how much work is to be performed. Sometimes the computer is able to perform certain kinds of work more efficiently.

Further, when we analyze an algorithm recall that ultimately our code gets translated into instructions for the computer hardware. Some instructions are implemented in a way that for any type of number (e.g. floating point), it doesn't matter if the number is 1.0 or 0.41582574300044717, the operation will take the exact same time and number of instructions to execute (e.g. for the addition operation).

13.3. Algorithms & Complexity

Sometimes a higher level operation is implemented in a way that takes many machine instructions. For example, division instructions may require many CPU cycles when compared to multiplication or division. Sometimes this is an important distinction and sometimes not, but for this book we will ignore this level of analysis.

13.3.1.1. Example: Sum of Consecutive Integers

Take for example the problem of determining the sum of integers from 1 to n . We will explore three different algorithms and the associated computational complexity for them.

13.3.1.2. Constant Time

A mathematical proof can show a simple formula for the result. This allows us to compute the answer in **constant time**, which means that for any n , our algorithm is essentially the same amount of work.

```
nsum_constant(n) = n * (n + 1) / 2
```

```
nsum_constant (generic function with 1 method)
```

In this we see that we perform three operations: a multiplication, a sum, and a division, no matter what n is. If n is 10_000_000 we'd expect this to complete in about a single unit of time.

13.3.1.3. Linear Time

This algorithm performs a number of operations which grows in proportion with n by individually summing up each element in 1 through n :

```
function nsum_linear(n)
    result = 0
    for i in 1:n
        result += i
    end
    result
end

nsum_linear (generic function with 1 method)
```

13. Elements of Computer Science

If n were 10_000_000, we'd expect it to run with roughly 10 million operations, or about 3 million times as many operations as the constant time version. We can say that this version of the algorithm will take approximately n steps to complete.

13.3.1.4. Quadratic Time

What if we were less efficient, and instead said that the operation $n + 42$ was to be implemented not as the basic addition of two numbers, but that we should *add one to n forty-two times*? That is, we'll see that we add a second loop which increments our result by a unit instead of simply adding the current i to the running total result :

```
function nsum_quadratic(n)
    result = 0
    for i in 1:n
        for j in 1:i
            result += 1
        end
    end
    result
end
```

(1) (2)

- (1) The outer loop with iterator i .
- (2) The inner loop with iterator j .

`nsum_quadratic (generic function with 1 method)`

Breaking down the steps:

- When i is 1 there is 1 addition in the inner loop
- When i is 2 there are 2 additions in the inner loop
- ...
- When i is n there are n additions in the inner loop

Therefore, this computation takes $1 + \dots + (n - 2) + (n - 1) + n$ steps to complete. We actually know that this simplifies down to our constant time formula $n * (n + 1) \div 2$ or $n^2 + n \div 2$ steps to complete.

13.3.1.5. Comparison

13.3.1.5.1. Big-O Notation

We can categorize the above implementations using a convention called **Big-O Notation**¹ which is a way of distilling and classifying computational complexity. We characterize the algorithms by the most significant term in the total number of operations. Table 13.1 shows for the examples constructed above what the description, order, and order of magnitude complexity is.

Table 13.1.: Complexity comparison for the three sample cases of summing integers from 1 to n .

Function	Computational Cost	Complexity Description	Big-O Order	Steps ($n = 10,000$)
nsum_constant	fixed	Constant	$O(1)$	~1
nsum_linear	n	Linear	$O(n)$	~10,000
nsum_quadratic	$n^2 + n \div 2$	Quadratic	$O(n^2)$	~100,000,000

Table 13.2 shows a comparison of a more extended set of complexity levels. For the most complex categories of problems, the cost to compute grows so fast that it boggles the mind. What sorts of problems fall into the most complex categories? $O(2^n)$, or exponential complexity, examples include the traveling salesman problem if solved with dynamic programming or the recursive approach to calculating the n th Fibonacci number. The beastly $O(n!)$ algorithms include brute force solving the traveling salesman problem or enumerating all partitions of a set. In financial modeling, we may encounter these sorts of problems in portfolio optimization (using the brute-force approach of testing every potential combination assets to optimize a portfolio).

Table 13.2.: Different Big-O Orders of Complexity

Big-O Order	Description	$n = 10$	$n = 1,000$	$n = 1,000,000$
$O(1)$	Constant Time	1	1	1
$O(n)$	Linear Time	10	1,000	1,000,000
$O(n^2)$	Quadratic Time	100	1,000,000	10^{12}
$O(\log(n))$	Logarithmic Time	3	7	14

¹“Big-O”, so named because of the “O” in used in $O(1)$. $O(n)$, etc. Not one of the sciences’ more creative names.

13. Elements of Computer Science

Big-O Order	Description	$n = 10$	$n = 1,000$	$n = 1,000,000$
$O(n \times \log(n))$	Linearithmic Time	30	7,000	14,000,000
$O(2^n)$	Exponential Time	1,024	$\sim 10^{300}$	$\sim 10^{301029}$
$O(n!)$	Factorial Time	3,628,800	$\sim 10^{2567}$	$\sim 10^{5565708}$

i Note

We care only about the most significant term because when n is large, the most significant term tends to dominate. For example, in our quadratic time example which has $n^2 + n \div 2$ steps, if n is a large number like 10 million, then we see that it will result in:

$$n^2 + n \div 2(10^6)^2 + 10^6 \div 2(10^{12}) + 5^6$$

10^{12} is significantly more important than 5^6 (sixty-four million times as important, to be precise).

Conversely, if n is small then we don't really care about computational complexity in general. This is why Big-O notation reduces the problem down to only the most significant complexity cost term.

13.3.1.5.2. Empirical Results

```
using BenchmarkTools
@btime nsum_constant(10_000)

0.750 ns (0 allocations: 0 bytes)

50005000

@btime nsum_linear(10_000)

1.250 ns (0 allocations: 0 bytes)

50005000

@btime nsum_quadratic(10_000)

2.398 μs (0 allocations: 0 bytes)
```

```
50005000
```

The preceding examples of constant, linear, and exponential times are *conceptually* correct but if we try to run them in practice we see that the description doesn't seem to hold at all for the linear time version, as it runs as quickly as the constant time version.

What happened was that the compiler was able to understand and optimize the linear version such that it effectively transformed it into the constant time version and avoid the iterative summation that we had written. For examples that are simple enough to use as a teaching problem, the compiler can often optimize different written code down to the same efficient machine code (this is the same Triangular Number optimization we saw in Section 4.5.3.4).

13.3.2. Expected versus worst-case complexity

Another consideration is that there may be one approach which performs better in the majority of cases, at the expense of having very poor performance in specific cases. Sometimes we may risk those high cost cases if we expect the benefit to be worthwhile on the rest of the problem set.

13.3.3. Complexity: Takeaways

The idea of algorithmic complexity is important because it grounds us in the harsh truth that some problems are *very* difficult to compute. It's in these cases that a lot of the creativity and domain specific heuristics can become the foremost consideration. We must remember to be thoughtful about the design of our models and when searching for additional performance to look for the loops-within-loops or combinatorical explosions. It's often at this level, rather than micro-optimizations, that you can transform the performance of the overall model (unless the fundamental complexity of the problem at hand forbids it).

13.4. Data Structures

Data structures is the art and science of how to represent data in discrete objects. There are many common kinds and many specialized sub-kinds, and we will describe some of the most common ones here. Julia has many data structures available in the Base library, but an extensive collection of other data structures can be found in the `DataStructures.jl` package.

13. Elements of Computer Science

13.4.1. Arrays

An **array** is a contiguous block of memory containing elements of the same type, accessed via integer indices. Arrays have fast random access and are the fastest data structure for linear/iterated access of data.

In Julia, an array is a very common data structure and is implemented with a simple declaration, such as:

x = [1, 2, 3]

In memory, the integers are stored as consecutive bits representing the integer values of 1, 2, and 3, and would look like this (with the different integers shown on new lines for clarity):

This is great for accessing the values one-by-one or in consecutive groups, but it's not efficient if values need to be inserted in between. For example, if we wanted to insert 0 between the 1 and 2 in `x`, then we'd need to overwrite the second position in the array, ask the operating system to allocate more memory², and re-write the bytes that come after our new value. Inserting values at the end (`push!(array, value)`) is usually fast unless more memory needs to be allocated.

13.4.2. Linked Lists

A **linked list** is a chain of nodes where each node contains a value and a pointer to the next node. Linked lists allow for efficient insertion and deletion but slower random access compared to arrays.

In Julia, a simple linked list node could be implemented as:

```
mutable struct Node
    value::Any
    next::Union{Node, Nothing}
end
```

```
z = Node(3, Nothing)
```

²In practice, the operating system may have already allocated space for an array that's larger than what the program is actually using so far, so this step may be 'quick' at times, while other times the operating system may actually need to extend the block of memory allocated to the array.

```
y = Node(2,z)  
x = Node(1,y)
```

- ① Here, 'Nothing' would represent the end of the linked list.

Inserting a new node between existing nodes is efficient - if we wanted to inser a new node between the ones with value 2 and 3, we could do this:

```
a = Node(0,z)                                ①  
y.next = a
```

However, accessing the nth element requires traversing the list from the beginning, making it $O(n)$ time complexity for random access. Also, if you have an intermediate node such as y, y itself does not know about x so there's no way to move 'up' the list to get to previous values.

13.4.3. Records/Structs

An aggregate of named fields, typically of fixed size and sequence. Records group related data together. We've encountered structs in Section 4.5.7, but here we'll add that simple structs with primitive fields can themselves be represented without creating pointers to the data stored:

```
struct SimpleBond
    id::Int
    par::Float64
end

struct LessSimpleBond
    id::String
    par::Float64
end

a = SimpleBond(1, 100.0)
b = LessSimpleBond("1", 100.0)
isbits(a), isbits(b)

(true, false)
```

Because `a` is comprised of simple elements, it can be represented as a contiguous set of bits in memory. It would look something like this in memory:

13. Elements of Computer Science

- ① The bits of 1
- ② The bits of 100.0

In contrast, the `LessSimpleBond` uses a `String` to represent the ID of the bond. Strings are essentially arrays of containers, and the arrays themselves are mutable containers which is by definition not a constant set of bits. In memory, `b` would look like:

```
.... a pointer ... # <1>
010000000101100100000000000000000000000000000000000000000000000000000000 # <2>
```

- ① a pointer/reference to the array of characters that comprise the string ID
- ② The bits of 100.0

In performance critical code, having data that is represented with simple bits instead of references/pointers can be much faster (see Chapter 24 for an example).

i Note

For many mutable types, there are immutable, bits-types alternatives. For example:

- Arrays have a `StaticArray` counterpart (from the `StaticArrays.jl` package).
- Strings have `InlineStrings` (from the `InlineStrings.jl` package) which use fixed-width representations of strings.

The downsides to the immutable alternatives (other than the loss of potentially desired flexibility that mutability provides) are that they can be harder on the compiler (more upfront compilation cost) to handle the specialized cases involved.

13.4.4. Dictionaries (Hash Tables)

13.4.4.1. Hashes and Hash Functions.

Hashes are the result of a **hash function** that maps arbitrary data to a fixed size value. It's sort of a "one way" mapping to a simpler value which has the benefits of:

1. One way so that if someone knows the hashed value, it's *very* difficult to guess what the original value was. This is most useful in cryptographic and security applications.
2. Creating (probabilistically) unique IDs for a given set of data.

For example, we can calculate a type of hash called an SHA hash on any data:

```

import SHA
let
    a = SHA.sha256("hello world") ▷ bytes2hex
    b = SHA.sha256(rand(UInt8, 10^6)) ▷ bytes2hex
    println(a)
    println(b)
end

```

b94d27b9934d3e08a52e52d7da7dabfac484efe37a5380ee9088f7ace2efcde9
713d14843b5639de07596c0043d0dfe843ad5130a490ece05990a84fa44caf2c

We can easily verify that the sha256 hash of "hello world" is the same each time, but it's virtually impossible to guess "hello world" if we are just given the resulting hash. This is the premise of trying to "crack" a password when the stored password hash is stolen.

One way to check if two set of data are the same is to compute the hash and see if the resulting hashes are equal. For example, maybe you want to see if two data files with different names contain the same data - comparing the hashes is a sure way to determine if they contain the same data.

13.4.4.2. Dictionaries

Dictionaries map a *key* to a *value*. More specifically, they use the *hash of a key* to store a reference to the *value*.

Dictionaries offer constant-time average case access but must handle potential collisions of keys (generally, the more robust the collision handling means higher fixed cost for access).

13.4.5. Graphs

A **graph** is a collection of nodes (also called vertices) connected by edges to represent relationships or connections between entities. Graphs are versatile data structures that can model various real-world scenarios such as social networks, transportation systems, or computer networks.

In Julia, a simple graph could be implemented using a dictionary where keys are nodes and values are lists of connected nodes:

13. Elements of Computer Science

```
struct Graph
    nodes::Dict{Any, Vector{Any}}
end

function add_edge!(graph::Graph, node1, node2)
    push!(get!(graph.nodes, node1, []), node2)
    push!(get!(graph.nodes, node2, []), node1)
end

g = Graph(Dict())
add_edge!(g, 1, 2)
add_edge!(g, 2, 3)
add_edge!(g, 1, 3)
```

This implementation represents an undirected graph. For a directed graph, you would only add the edge in one direction.

Graphs can be traversed using various algorithms such as depth-first search (DFS) or breadth-first search (BFS). These traversals are useful for finding paths, detecting cycles, or exploring connected components.

For more advanced graph operations, the `Graphs.jl` package provides a comprehensive set of tools for working with graphs in Julia.

13.4.6. Trees

A tree is a hierarchical data structure with a root node and child subtrees. Each node in a tree can have zero or more child nodes, and every node (except the root) has exactly one parent node. Trees are widely used for representing hierarchical relationships, organizing data for efficient searching and sorting, and in various algorithms.

A simple binary tree node in Julia could be implemented as:

```
mutable struct TreeNode
    value::Any
    left::Union{TreeNode, Nothing}
    right::Union{TreeNode, Nothing}
end

# Creating a simple binary tree
root = TreeNode(1,
    TreeNode(2,
        TreeNode(4, nothing, nothing),
        TreeNode(5, nothing, nothing))
```

```
    ),
    TreeNode(3,
              nothing,
              TreeNode(6, nothing, nothing)
    )
)
```

Trees have various specialized forms, each with its own properties and use cases:

- Binary Search Trees (BST): Each node has at most two children, with all left descendants less than the current node, and all right descendants greater.
 - AVL Trees: Self-balancing binary search trees, ensuring that the heights of the two child subtrees of any node differ by at most one.
 - B-trees: Generalization of binary search trees, allowing nodes to have more than two children. Commonly used in databases and file systems.
 - Trie (Prefix Tree): Used for efficient retrieval of keys in a dataset of strings. Each node represents a common prefix of some keys.

Trees support efficient operations like insertion, deletion, and searching, often with $O(\log n)$ time complexity for balanced trees. They are fundamental in many algorithms and data structures, including heaps, syntax trees in compilers, and decision trees in machine learning.

13.4.7. Data Structures Conclusion

Data structures have strengths and weakness depending on whether you want to prioritize computational efficiency, memory (space) efficiency, code simplicity, and/or mutability. Due to the complexity of real world modeling needs, it can be the case that different representations of the data are more natural or more efficient for the use case at hand.

13.5. Formal Verification

Formal verification is a technique used to prove or disprove the correctness of algorithms with respect to a certain formal specification or property. In essence, it's a mathematical approach to ensuring that a system behaves exactly as intended under all possible conditions.

13. Elements of Computer Science

13.5.1. Basic Concept

In formal verification, we use mathematical methods to:

1. Create a formal model of the system
2. Specify the desired properties or behaviors
3. Prove that the model satisfies these properties

This process can be automated using specialized software tools called theorem provers or model checkers.

13.5.2. Formal Verification in Practice

It sounds like the perfect risk management and regulatory technique: prove that the system works exactly as intended. However, there has been very limited deployment of formal verification in industry. This is for several reasons:

1. Incomplete Coverage: It's often impractical to formally verify entire large-scale financial systems. Verification, if at all, is typically limited to critical components.
2. Incomplete Specification: Actually reasoning through how the system should behave in all scenarios requires actually contemplating mathematically complete and rigorous possibilities that could occur.
3. Model-Reality Gap: The formal model may not perfectly represent the real-world system, especially in finance where market behavior can be unpredictable.
4. Changing Requirements: Financial regulations and market conditions change rapidly, potentially outdated formal verifications.
5. Performance Trade-offs: Systems designed for easy formal verification might sacrifice performance or flexibility.
6. Cost: The process can be expensive in terms of time and specialized labor.

13.5.3. Related Topics

13.5.3.1. Property Based Testing

Testing will be discussed in more detail in Chapter 14, but an intermediate concept between Formal Verification and typical software testing is **property-based** testing, which tests for general rules instead of specific examples.

For example, a function which is associative ($(a + b) + c = a + (b + c)$) or commutative ($a + b = b + a$) can be tested with simple examples like:

```
using Test

myadd(a,b) = a + b

@test myadd(1,2) == myadd(2,1)
@test myadd(myadd(1,2),3) == myadd(1,myadd(2,3))
```

However, we really haven't proven the associative and commutative properties in general. There are techniques to do this, which is a more comprehensive alternative to testing specific examples above. Packages like Supposition.jl provide functionality for this. Note that like Formal Verification, property-based testing is a more advanced topic.

13.5.3.2. Fuzzing

Fuzzing is kind of like property based testing, but instead of testing general rules, we generalize the simple examples using randomness. For example, we could test the commutative property using random numbers instead, therefore statistically checking that the property holds:

```
@testset for i in 1:10000
    a = rand()
    b = rand()

    @test myadd(a,b) == myadd(b,a)
end
```

This is a good advancement over the simple `@test myadd(1,2) == myadd(2,1)`, in terms of checking the correctness of `myadd`, but it comes at the cost of more computational time and non-deterministic tests.

14. Applying Software Engineering Principles

“Programs must be written for people to read, and only incidentally for machines to execute.” — Harold Abelson and Gerald Jay Sussman (1984)

14.1. In this section

We describe modern software engineering practices such as version control, testing, documentation, and pipelines which can be utilized by the financial professional to make their own work more robust and automated. Data practices and workflow advice.

14.2. Introduction

In addition to the core concepts from computer science described so far, there’s also a similar set of ideas about the *practice* and *experience* of working with a code-based workflow that makes the approach more powerful than what the code itself can do.

That is, the majority of a professional financial modeler’s time is often spent *doing things other than building models*, such as testing the model’s results, writing documentation, collaborating with others on the design, and figuring out how to share the model with others inside the company. This chapter covers how a code-first workflow makes each one of those responsibilities easier or more effective.

14. Applying Software Engineering Principles

14.3. Testing

14.4. Documentation

14.5. Version Control

14.5.1. Collaborative Workflows

14.6. Distributing the Package

Note some more advanced testing topics in Section 13.5.3.

15. Statistical Inference and Information Theory

"My greatest concern was what to call [the amount of unpredictability in a random outcome]. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.'

When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place, your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.' "- Claude Shannon (1971)

15.1. In This Chapter

A brief introduction to information theory and its foundational role in statistics. Entropy and probability distributions. Bayes' rule and model selection comparison via likelihoods. A brief tour of modern Bayesian statistics.

15.2. Information Theory

Probability, statistics, machine learning, signal processing, and even physics have a foundational link in **information theory** which is the description and analysis of how much useful data is contained within something.

Let's consider the following number that we encounter while reading a report which contains estimates of total amount of assets held. Unfortunately, for one reason or another one of the digits is not visible to you. Here's what you can read, with the _ indicating that the digit is not visible:

32,000,_00

15. Statistical Inference and Information Theory

Now you probably already formed an opinion on what the missing number is, but let's look at how we can quantify the analysis.

Given that we know the number was an estimate and the tendency of humans to like nice round numbers, our **prior assumption** for what the probability of the missing digit is may be something like the $p(x_i)$ row of Table 15.1. We shall call the individual outcomes x_i and the overall set of probabilities $\{x_0, x_1, \dots, x_9\}$ is called X .

The information content of an outcome, $h(x)$ is measured in bits and defined as¹:

$$h(x_i) = \log_2 \frac{1}{p(x_i)} \quad (15.1)$$

So if we were to find out that the missing digit were indeed 0, we have gained less information relative to our expectation than if the missing digit were anything other than 0.

We can characterize the entire distribution X via the **entropy**, $H(X)$, of a probability set is the ensemble's average information content:

$$H(X) = \sum p(x_i) \log_2 \frac{1}{p(x_i)} \quad (15.2)$$

The entropy $H(X)$ of the presumed outcomes in Table 15.1 distribution of outcomes is 0.722bits.

Table 15.1.: Probability distribution of missing digit, knowing the human inclination to prefer round numbers when estimating.

x_i	0	1	2	3	4	5	6	7	8	9
$p(x_i)$.91	.01	.01	.01	.01	.01	.01	.01	.01	.01
$h(x_i)$	0.136	6.644	6.644	6.644	6.644	6.644	6.644	6.644	6.644	6.644

Note that we have take a view on the probability distribution for the missing digit, and we'll refer to this as the **prior assumption** (or just **prior**). This is an opinionated assumption, so what if we had another colleague who believed humans are completely rational and without bias for certain numbers. They would then be arguing for a prior assumed distribution consistent with Table 15.2.

With the uniform prior assumption, $H(X) = 3.322$ bits and $h(x_i)$ is also uniform. We will not prove it here, but a uniform probability over a set of outcomes is the highest entropy distribution that can be assumed.

¹Log base two turns out to be the most natural representation of information content as it mimics the fundamental 0 or 1 value bit. A more complete introduction is available in "Information Theory, Inference, and Learning Algorithms" by David MacKay.

Table 15.2.: Probability distribution of missing digit with uniform, maximal entropy for the assumed probability distribution.

x_i	0	1	2	3	4	5	6	7	8	9
$p(x_i)$.10	.10	.10	.10	.10	.10	.10	.10	.10	.10
$h(x_i)$	3.322	3.322	3.322	3.322	3.322	3.322	3.322	3.322	3.322	3.322

The choice of prior assumption can significantly impact the interpretation and analysis of the missing information. If we have strong reasons to believe that the human bias prior is more appropriate given the context (e.g., knowing that the number is an estimate), then we would expect the missing digit to be '0' with high probability. However, if we have no specific knowledge about the nature of the number and prefer to make a more conservative assumption, the uniform prior may be more suitable.

In real-world scenarios, the choice of prior assumptions often depends on domain knowledge, available data, and the specific problem at hand. It is important to carefully consider and justify the prior assumptions used in information-theoretic and statistical analyses.

15.2.1. Example: Classification

In this example, we will determine the optimal splits for a decision tree² based on the information gained at each node in the tree.

```
using DataFrames
```

```
employed = [true, false, true, true, true, false, true]
good_credit = [true, true, false, true, false, false, true]
default = [true, false, true, true, true, false, true]
default_data = DataFrame(; employed, good_credit, default)
```

The entropy of the default rate data is, per Equation 15.2:

```
H0 = let
    p1 = sum(default_data.default) / nrow(default_data)
    p2 = 1 - p1
    p1 * log2(1 / p1) + p2 * log(1 / p2)
end
```

0.6578517147391054

²A decision tree is a classification algorithm which attempts to optimally classify an output based on if/else type branches on the input variables.

15. Statistical Inference and Information Theory

Table 15.3.: Fictional data regarding loan attributes and whether or not a loan defaulted before its maturity.

	employed	good_credit	default
	Bool	Bool	Bool
1	1	1	1
2	0	1	0
3	1	0	1
4	1	1	1
5	1	0	1
6	0	0	1
7	0	0	0
8	1	1	1

Our goal is to determine which attribute (`employed` or `good_credit`) to use as the first split in the decision tree. We will decide this by calculating the information gain, which is the difference in entropy between the prior node and the candidate node. In our case we start with H_0 as calculated above for the output variable `default` and calculate the difference in entropy between it and the average entropy of the data if we split on that node. **Information gain**, $IG(inputs, attributes)$, is:

...

Let's first consider splitting the tree based on the `employed` status. We will calculate the entropy of each subset: with employment and without employment.

If we split the data based on being employed, we'd get two sub-datasets:

```
df_employed = filter(:employed => ==(true), default_data)
```

	employed	good_credit	default
	Bool	Bool	Bool
1	1	1	1
2	1	0	1
3	1	1	1
4	1	0	1
5	1	1	1

and

```
df_unemployed = filter(:employed => ==(false), default_data)
```

	employed	good_credit	default
	Bool	Bool	Bool
1	0	1	0
2	0	0	1
3	0	0	0

let's call it's entropy H_{employed} , which should be zero because there is no variability in the default outcome for this subset.

```
H_employed = let
    p1 = sum(df_employed.default) / nrow(df_employed)
    p2 = 1 - p1
    # p1 * log2(1 / p1) + p2 * log(1 / p2)
    p1 * log2(1 / p1) + 0
end
```

①

- ① In the case of $p_i = 0$ the value of h (the second term in the sum above) is taken to be 0, which is consistent with the $\lim_{p \rightarrow 0^+} p \log(p) = 0$.

0.0

And the corresponding candidate leaf is $H_{\text{unemployed}}$:

```
H_unemployed = let
    p1 = sum(df_unemployed.default) / nrow(df_unemployed)
    p2 = 1 - p1
    p1 * log2(1 / p1) + p2 * log(1 / p2)
end
```

0.7986309056458281

The average of the two is weighted by the size of the data that would fall into each leaf:

```
H1_employment = let
    p_emp = nrow(df_employed) / nrow(default_data)
    p_unemp = 1 - p_emp
    p_emp * H_employed + p_unemp * H_unemployed
end
```

0.29948658961718555

The information gain for splitting the tree using employment status is the difference between the root entropy and the entropy of the employment split:

15. Statistical Inference and Information Theory

IG_employment = H0 - H1_employment

0.35836512512191987

We could repeat the analysis to determine the information gain if we were to split the tree based on having good credit. However, given that there are only two attributes we can already conclude that employed is a better attribute to split the data on. This is because the information gain of IG_employment (0.358) is the majority of the overall entropy H0 (0.658). Entropy is always additive and you cannot have negative entropy, therefore no other other attribute could have greater information gain. This also matches our intuition when looking at Table 15.3 as the eye can spot a higher correlation between employed and default than good_credit and default.

The above example demonstrates how we can use information theory to create more optimal inferences on data.

15.2.2. Maximum Entropy Distributions

Why is information theory a useful concept? Many financial models are statistical in nature and concepts of randomness and entropy are foundational. For example, when trying to estimate parameter distributions or assume a distribution for a random process you can lean on information theory to use the most conservative choice: the distribution with the highest entropy given known constraints. These distributions are referred to as **maximum entropy distributions**. Some discussion of maximum entropy distributions in the context of risk assessment is available in an article by Duracz³. probability distributions and risk asses

Table 15.4.: Maximum Entropy Distributions and the conditions under which they are applicable.

Constraint	Discrete Distribution	Continuous Distribution
Bounded range	Uniform (discrete)	Uniform (continuous)
Bounded range (0 to 1) with information about the mean or variance		Beta
Mean is finite, two possible values	Binomial	

³https://www.researchgate.net/publication/239752412_Derivation_of_Probability_Distributions_for_Risk_Assessment

Constraint	Discrete Distribution	Continuous Distribution
Mean is finite and positive	Geometric	Exponential
Mean is finite and range is > zero		Gamma
Mean and Variance is finite		Guassian (Normal)
Positive and equal mean and variance	Poisson	

The distributions in Table 15.4 arise again and again in nature because of the second law of thermodynamics - nature likes to have constantly increasing entropy and therefore it should be no surprise (random) processes that maximize entropy pop up all over the place. As an example, let's look at processes that behave like the Gaussian (Normal) distribution.

15.2.2.1. Processes that give rise to certain distributions

A random walk can be viewed as the cumulative impact of nudges pushing in opposite directions. This behavior culminates in the random, terminal position being able to be described by a Gaussian distribution. The center of a Gaussian distribution is "thick" because there are many more ways for the cumulative total nudges to mostly cancel out, while its increasingly rare to end up further and further from the starting point (mean). The distribution then spreads out as flat (randomly) as it can while still maintaining the constraint of having a given, finite variance. Any other continuous distribution that has the same mean and variance has lower entropy than the Guassian.

Table 15.5.: Underlying processes create typical probability distributions. That there is significant overlap with the distributions in Section 15.2.2 is not a coincidence.

Process	Distribution of Data	Examples
Many <i>additive</i> pluses and minus that move an outcome in one dimension	Normal	Sum of many dice rolls, errors in measurements, sample means (Central Limit Theorem)
Many <i>multiplicative</i> pluses and minus that move an outcome in one dimension	Log-normal	Incomes, sizes of cities, stock prices

15. Statistical Inference and Information Theory

Process	Distribution of Data	Examples
Waiting times between independent events occurring at a constant average rate	Exponential	Time between radioactive decay events, customer arrivals
Discrete trials each with the same probability of success, counting the number of successes	Binomial	Coin flips, defective items in a batch
Discrete trials each with the same probability of success, counting the number of trials until the first success	Geometric	Number of job applications until getting hired
Continuous trials each with the same probability of success, measuring the time until the first success	Exponential	Time until a component fails, time until a sales call results in a sale
Waiting time until the r-th event occurs in a Poisson process	Gamma	Time until the 3rd customer arrives, time until the 5th defect occurs

💡 Probability Distributions

There are a *lot* of specialized distributions. There are lists of distributions you can find online or in references such as Leemis and McQueston (2008) which has a full-page network diagram of the relationships.

The information-theoretic and Bayesian perspective on it is to eschew memorization of a bunch of special cases and statistical tests. If you pull up the aforementioned diagram in Leemis and McQueston (2008), you can see just a handful of distributions that have the most central roles in the universe of distributions. Many distributions are simply transformations, limiting instances, or otherwise special cases of a more fundamental distribution. Instead of trying to memorize a bunch of probability distributions, it's better to think critically about:

1. The fundamental processes that give rise to the randomness.
2. Transformations of the data to make it nicer to work with, such as translations, scaling, or other non-destructive changes.

Then when you encounter a wacky dataset you don't need to comb the depths of Wikipedia to find the perfect probability distributions.

15.2.2.2. Additive and Multiplicative Processes

Table 15.5 describes some examples, let us discuss further what it means to have a process that arises via an additive vs multiplicative effect⁴.

An outcome is additive if it's the sum or difference of multiple independent processes. One of the simplest examples of this is rolling multiple dice and taking their sum. Or a random walk along the natural numbers wherein with equal probability you take a step left or right. The distribution of the position after n steps converges rapidly to a normal distribution. Another common one is when you are looking at the mean of a sample - since you are summing up the individual measurements you end up with a normal distribution (the Central Limit Theorem).

However, many processes are multiplicative in nature. For example the population density of cities is distributed in a log-normal fashion. If we think about the factors that contribute to choice of place to live, we can see how these factors multiply: an attractive city might make someone 10% more likely to move, a city with water features 15% more likely, high crime 30% less likely, etc. These forces combine in a multiplicative way in the generative process of deciding where to move.

💡 Tip 1: Logarithms

The logarithm of a geometric process transforms the outcomes into "log-space". The information is the same, but is often a more convenient form for the analysis. That is, if:

$$Y = x_1 \times x_2 \times \dots \times x_i$$

Then,

$$\log(Y) = \log(x_1) + \log(x_2) + \dots + \log(x_i)$$

This is effectively the transformation that gives rise to the Normal versus Log-Normal distribution.

Bringing this back to the context of computational thinking:

First, we should think about how to transform data or modeling outcomes into a more convenient format. The log transform doesn't eliminate any information but may map the information into a shape that is easier for an optimizer or Monte Carlo simulation to explore.

⁴Multiplicative processes are often referred to as "geometric", as in "geometric Brownian motion" or "geometric mean". Additive processes are sometimes referred to as "arithmetic". This root of this confusing terminology appears to be due to the fact that series involving repeated multiplication were solved via geometric (triangles, angles, etc.) methods while those using sums and differences were solved via arithmetic.

15. Statistical Inference and Information Theory

Second, per Chapter 4, floating point math is a *lossy* transformation of real numbers into a digital computer representation. Some information (in the literal Shannon information sense) is lost when computing and this tends to be worst with very small real numbers, such as those we encounter frequently in probabilities and likelihoods. Logarithms map very small numbers into negative numbers that don't encounter the same degree of truncation error that tiny numbers do

Third, modern CPUs are generally much faster at adding or subtracting numbers than multiplying or dividing. Therefore working with the logarithm of processes may be computationally faster than the direct process itself.

15.3. Bayes' Rule

The minister and statistician Thomas Bayes derived a relationship of conditional probabilities that we today know as **Bayes' Rule**, commonly written as:

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D)}$$

The components of this are:

- $P(H | D)$ is the conditional probability of event H occurring given that D is true.
- $P(D | H)$ is the conditional probability of event D occurring given that H is true.
- $P(H)$ is the prior probability of event H .
- $P(D)$ is the prior probability of event D .

If we take the following:

- D is the available data
- H is our hypothesis

Then we can draw conclusions about the probability of a hypothesis being true given the observed data. When thought about this way, Bayes' rule is often described as:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

This is a very useful framework, which we'll return to more completely in Section 15.4. First, let's look at combining information theory and Bayes' rule in an applied example.

15.3.1. Example: Model Selection via Likelihoods

Let's say that we have competing hypothesis about a data generating process, such as: "given a set of data representing risk outcomes, what distribution best fits the data"?

We can compare these models using Bayes' rules by observing the following: Suppose we have two models, H_1 and H_2 , and we want to compare their likelihoods given the observed data, D . We can use Bayes' rule to calculate the posterior probability of each model: $\text{P}(H_1|D) = (P(D|H_1) * P(H_1)) / P(D)$

$$\text{P}(H_2|D) = (P(D|H_2) * P(H_2)) / P(D)$$

Where:

- $P(H_1|D)$ and $P(H_2|D)$ are the posterior probabilities of models H_1 and H_2 , respectively, given the data D .
- $P(D|H_1)$ and $P(D|H_2)$ are the likelihoods of the data D under models H_1 and H_2 , respectively.
- $P(H_1)$ and $P(H_2)$ are the prior probabilities of models H_1 and H_2 , respectively.
- $P(D)$ is the marginal likelihood of the data, which serves as a normalizing constant.

To compare the likelihoods of the two models, we can calculate the ratio of their posterior probabilities, known as the Bayes factor, BF :

$$BF = \frac{P(H_1|D)}{P(H_2|D)}$$

Substituting the expressions for the posterior probabilities from Bayes' rule, we get:

$$BF = \frac{P(D|H_1) \times P(H_1)}{P(D|H_2) \times P(H_2)}$$

The marginal likelihood $P(D)$ cancels out since it appears in both the numerator and denominator. If we assume equal prior probabilities for the models, i.e., $P(H_1) = P(H_2)$, then the Bayes factor simplifies to the likelihood ratio:

$$BF = \frac{P(D|H_1)}{P(D|H_2)}$$

The interpretation of the Bayes factor is as follows:

- If $BF > 1$, the data favor H_1 over H_2 .
- If $BF < 1$, the data favor H_2 over H_1 .
- If $BF = 1$, the data do not provide evidence in favor of either model.

15. Statistical Inference and Information Theory

In practice, the likelihoods $P(D|H_1)$ and $P(D|H_2)$ are often calculated using the probability density or mass functions of the models, evaluated at the observed data points. The prior probabilities $P(H_1)$ and $P(H_2)$ can be assigned based on prior knowledge or assumptions about the models. By comparing the likelihoods of the models using the Bayes factor, we can quantify the relative support for each model given the observed data, while taking into account the prior probabilities of the models.

Another way of interpreting this is the more simplistic evaluation of which model has the higher likelihood given the data: this is simply a matter of comparing the magnitude of the likelihoods.

⚠ Null Hypothesis Statistical Test

Null Hypothesis Statistical Tests (NHST) is the idea of trying to statistically support an alternative hypothesis over a null hypothesis. The support in favor of alternative versus the null is reported via some statistical power, such as the **p-value** (the probability that the test result is as, or more extreme, than the value computed). The idea is that there's some objective way to push science towards greater truths and NHST was seen as a methodology that avoided the subjectivity of the Bayesian approach. However, while pure in concept, the NHST choices of both null hypothesis and model contain significant amounts of subjectivity! We might as well call the null hypothesis a prior and stop trying to disprove it absolutely. Instead: focus on model comparison, model structure, and posterior probabilities of the competing theories.

Over 100 statistical tests have been developed in service of NHST Lewis (2013), but it's widely viewed now that a focus on NHST has led to *worse* science due to a multitude of factors, such as:

- “P-hacking” or trying to find subsets of data which can (often only by chance) support rejecting some null
- Cognitive anchoring to the importance of a p-value of 0.05 or less - why choose that number versus 0.01 or 0.001 or 0.49?
- Bias in research processes where one may stop data collection or experimentation after achieving a favorable test result
- Inappropriate application of the myriad of statistical tests
- Focus on p-values rather than effects that simply matter more or have greater effect
 - For example, which is of more interest to doctors? A study indicating a 1 in a billion chance of serious side effect (p-value 0.0001) or a study indicating a 1 in 3 chance (p-value 0.06)? Many journals would only publish the former study.
- Difficulty to determine *causal* relationships.

There is subjectivity in the null hypothesis, data collection methodologies, study design, handling of missing data, choice of data *not* to include, which statistical tests to perform, and interpretation of relationships.

The authors of this book recommend against basic NHST and memorization of statistical tests in favor of principled Bayesian approaches. For the actuarial readers, NHST is analogous to traditional credibility methods (of which the authors also prefer more modern statistical approaches).

The example we'll look at relates to the annual rainfall totals for a specific location in California⁵, which could be useful for insuring flood risk or determining the value of a catastrophe bond. Acknowledging that we are attempting to create a geocentric model⁶ instead of a scientifically accurate weather model, we narrow the problem to finding a probability distribution that matches the historical rainfall totals. Our goal is to recommend a model that best fits the data and justify that recommendation quantitatively. Before even looking at the data, Table 15.6 shows three competing models based on thinking about the real-world outcome we are trying to model. These three are chosen for the increasingly sophisticated thought process that might lead the modeler to recommend them - but which is supportable by the statistics?

Table 15.6.: Three alternative hypothesis about the distribution of annual rainfall totals.

Hypothesis	Process	Possible Rationale
H_1	A Normal (Gaussian) distribution	The sum of independent rainstorms creates annual rainfall totals that are normally distributed
H_2	A LogNormal distribution	Since it's normal-ish, but skewed and can't be negative
H_3	A Gamma Distribution	Since rainfall totals would be the sum of exponentially-distributed independent rainfall events

i Note

In the literature, H_3 (the Gamma distribution) is known as the "Log-Pearson Type III distribution". It's actually recommended by the US Corps of Army Engineers as the recommended way to model rainfall totals.

```
rain = [
    39.51, 42.65, 44.09, 41.92, 28.42, 58.65, 30.18, 64.4, 29.02,
```

⁵<https://data.ca.gov/dataset/annual-precipitation-data-for-northern-california-1944-current>

⁶See @sec-predictive-vs-explanatory.

15. Statistical Inference and Information Theory

```
37.00, 32.17, 36.37, 47.55, 27.71, 58.26, 36.55, 49.57, 39.84,  
82.22, 47.58, 51.18, 32.28, 52.48, 65.24, 51.12, 25.03, 23.27,  
26.11, 47.3, 31.8, 61.45, 94.95, 34.8, 49.53, 28.65, 35.3, 34.8,  
27.45, 20.7, 36.99, 60.54, 22.5, 64.85, 43.1, 37.55, 82.05, 27.9,  
36.55, 28.7, 29.25, 42.32, 31.93, 41.8, 55.9, 20.65, 29.28, 18.4,  
39.31, 20.36, 22.73, 12.75, 23.35, 29.59, 44.47, 20.06, 46.48,  
13.46, 9.34, 16.51, 48.24  
];
```

When we do plot it, we can see some of the characteristics that align with our prior assumptions and knowledge about the system itself, such as: the data being constrained to positive values and a skew towards having some extreme weather years with lots of rainfall.

```
using CairoMakie  
hist(rain)
```

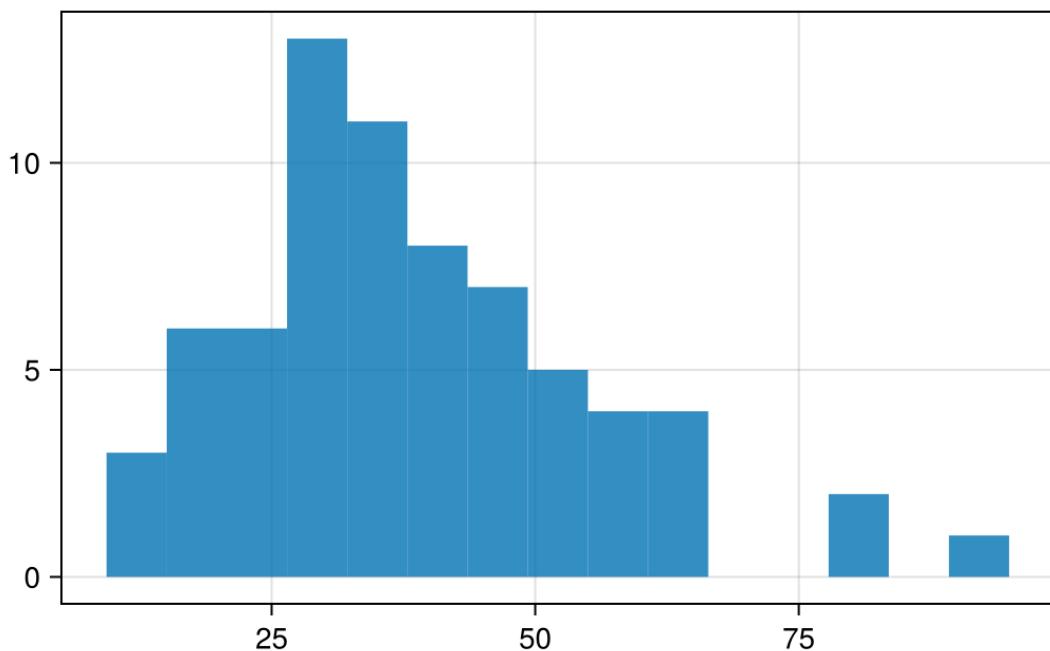


Figure 15.1.: Annual rainfall totals for a specific location in California.

We will show the likelihood of the three models after deriving the **maximum likelihood (MLE)**, which is simply finding the parameters that maximize the calculated likelihood. In general, this can be accomplished by an optimization routine, but here we will just use the functions built into Distributions.jl:

```

using StatsBase
using Distributions

n = fit_mle(Normal, rain)
ln = fit_mle(Normal, log.(rain))
lg = fit_mle(Gamma, log.(rain))
@show n
@show ln
@show lg;

n = Normal{Float64}(\mu=38.91442857142857, σ=16.643603630714306)
ln = Normal{Float64}(\mu=3.5690550009062663, σ=0.44148379736539156)
lg = Gamma{Float64}(α=61.58531301458412, θ=0.05795302201453571)

let x = rain

range = 1:0.1:100
fig, ax, _ = lines(range, cdf.(n, range), label="Normal", axis=(xgridvisible=false, ygridvisible=false))
lines!(ax, range, cdf.(ln, log.(range)), label="LogNormal")
lines!(range, cdf.(lg, log.(range)), label="LogGamma")
lines!(quantile.(Ref(x), 0.01:0.01:0.99), 0.01:0.01:0.99, label="Data", color(:black, 0.6))
fig[1, 2] = Legend(fig, ax, "Model", framevisible=false)
fig
end

```

15. Statistical Inference and Information Theory

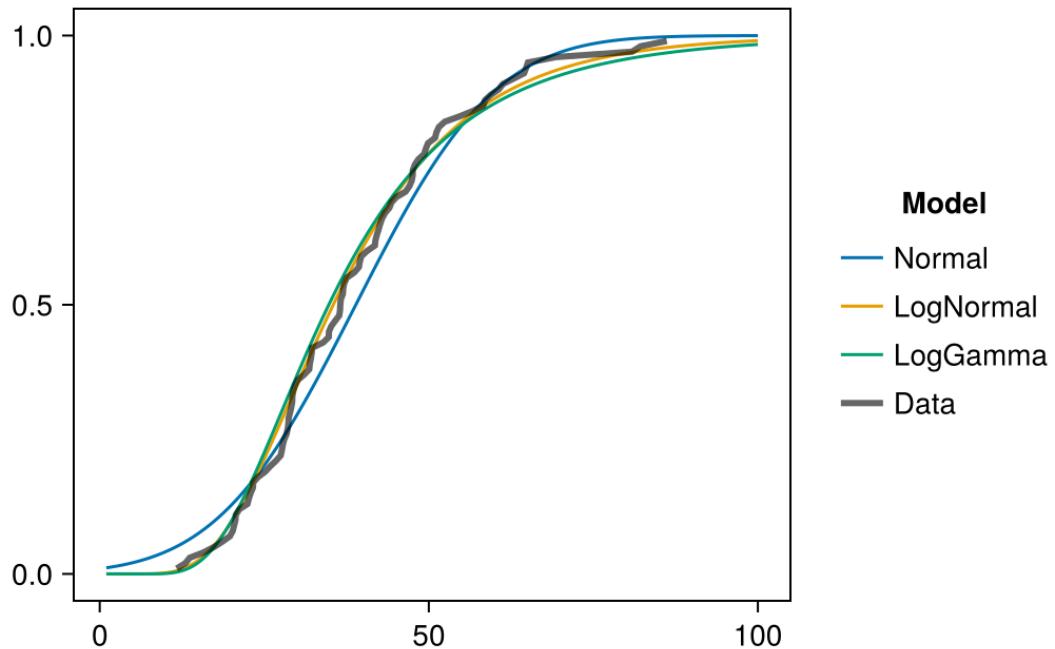


Figure 15.2.

Let's look at the likelihoods. For the practical reasons described in Tip 1, we will compare the log-likelihoods to maintain convention with what you'd likely see or deal with in practice. Taking the log of the likelihood does not change the ranking of the likelihoods.

```
let
    n_lik = sum(log.(pdf.(n, rain)))
    ln_lik = sum(log.(pdf.(ln, log.(rain))))
    lg_lik = sum(log.(pdf.(lg, log.(rain))))
    @show n_lik
    @show ln_lik
    @show lg_lik
end;

n_lik = -296.1675156647812
ln_lik = -42.09272021737914
lg_lik = -43.79151806348801
```

The results indicate that the LogNormal and the Gamma model for rainfall distribution are very superior to the Normal model, consistent with the visual inspection of the quan-

tiles in Figure 15.2. We reach that conclusion by noting how much more likely the latter two are, as the likelihoods of -42 and -44 is much greater than -296^7 .

We evaluated the likelihood at a single point estimate of the parameters, but a true posterior probability of the parameters of the distributions will be represented by a *distribution* rather than a point. Expanding the analysis to account for that point will be the focus of the remainder of this chapter.

15.4. Modern Bayesian Statistics

15.4.1. Background

Bayesian statistics is generally *not* taught in undergraduate statistics - Bayes' rule is introduced, basic probability exercises are assigned, and then statistics moves on to a curriculum of regression and NHSTs. Why is the applied practice of statistics then gravitating towards Bayesian approaches? There are both philosophical and practical reasons why.

Philosophically, one of the main reasons why Bayesian thinking is appealing is its ability to provide a straightforward interpretation of statistical conclusions.

For example, when estimating an unknown quantity, a Bayesian probability interval can be directly understood as having a high probability of containing that quantity. In contrast, a Frequentist confidence interval is typically interpreted only in the context of a series of similar inferences that could be made in repeated practice. In recent years, there has been a growing emphasis on interval estimation rather than hypothesis testing in applied statistics. This shift has strengthened the Bayesian perspective since it is likely that many users of standard confidence intervals intuitively interpret them in a manner consistent with Bayesian thinking.

Another meaningful way to understand the contrast between Bayesian and Frequentist approaches is through the lens of decision theory, specifically how each view treats the concept of randomness. This perspective pertains to whether you regard the data being random or the parameters being random.

Frequentist statistics treats parameters as fixed and unknown, and the data as random — this is reflective of the view that data you collect is but one realization of an infinitely repeatable random process. Consequently, Frequentist procedures, like hypothesis testing or confidence intervals, are generally based on the idea of long-run frequency or repeatable sampling.

⁷The values are negative because we are taking the logarithm of a number less than 1. The likelihoods are less than 1 because the likelihood is the joint (multiplicative) probability of observing each of the individual outcomes.

15. Statistical Inference and Information Theory

Conversely, Bayesian statistics turns this on its head by treating the data as fixed — after all, once you've collected your data, it's no longer random but a fixed observed quantity. Parameters, which are unknown, are treated as random variables. The Bayesian approach then allows us to use probability to quantify our uncertainty about these parameters.

The Bayesian approach tends to align more closely with our intuitive way of reasoning about problems. Often, you are given specific data and you want to understand what that particular set of data tells you about the world. You're likely less interested in what might happen if you had infinite data, but rather in drawing the best conclusions you can from the data you do have.

Practically, recent advances in computational power, algorithm development, and open-source libraries have enabled practitioners to adapt the Bayesian workflow.

Deriving the posterior distribution is analytically intractable so computational methods must be used. Advances in raw computing power only in the 1990's made non-trivial Bayesian analysis possible, and recent advances in algorithms have made the computations more efficient. For example, one of the most popular algorithms, NUTS, was only published in the 2010's.

Many problems require the use of compute clusters to manage runtime, but if there is any place to invest in understanding posterior probability distributions, it's insurance companies trying to manage risk!

The availability of open-source libraries, such as Turing.jl, PyMC3, and Stan provide access to the core routines in an accessible interface. The exercise remains undoubtedly one that benefits from the computational thinking described in this book - understanding model complexity, model transformations and structure, data types and program organization, etc.

15.4.1.1. Advantages of the Bayesian Approach

The main advantages of this approach over traditional actuarial techniques are:

1. **Focus on distributions rather than point estimates of the posterior's mean or mode.** We are often interested in the distribution of the parameters and a focus on a single parameter estimate will underestimate the risk distribution.
2. **Model flexibility.** A Bayesian model can be as simple as an ordinary linear regression, but as complex as modeling a full insurance mechanics.
3. **Simpler mental model.** Fundamentally, Bayes' theorem could be distilled down to an approach where you count the ways that things could occur and update the probabilities accordingly.

4. **Explicit Assumptions.**: Enumerating the random variables in your model and explicitly parameterizing prior assumptions avoids ambiguity of the assumptions inside the statistical model.

15.4.1.2. Challenges with the Bayesian Approach

With the Bayesian approach, there are a handful of things that are challenging. Many of the listed items are not unique to the Bayesian approach, but there are different facets of the issues that arise.

1. **Model Construction.** One must be thoughtful about the model and how variables interact. However, with the flexibility of modeling, you can apply (actuarial) science to make better models!
2. **Model Diagnostics.** Instead of R^2 values, there are unique diagnostics that one must monitor to ensure that the posterior sampling worked as intended.
3. **Model Complexity and Size of Data.** The sampling algorithms are computationally intensive - as the amount of data grows and model complexity grows, the runtime demands cluster computing.
4. **Model Representation.** The statistical derivation of the posterior can only reflect the complexity of the world as defined by your model. A Bayesian model won't automatically infer all possible real-world relationships and constraints.

Subjectivity of the Priors?

There are two ways one might react to subjectivity in a Bayesian context: It's a feature that should be embraced or it's a flaw that should be avoided.

15.4.1.3. Subjectivity as a Feature

A Bayesian approach to defining a statistical model is an approach that allows for explicitly incorporating professional judgment. Encoding assumptions into a Bayesian model forces the actuary to be explicit about otherwise fuzzy predilections. The explicit assumption is also more amenable to productive debate about its merits and biases than an implicit judgmental override.

15.4.1.4. Subjectivity as a Flaw

Subjectivity is inherent in all useful statistical methods. Subjectivity in traditional approaches include how the data was collected, which hypothesis to test, what significant levels to use, and assumptions about the data-generating processes. In fact, the “objective” approach to null hypothesis testing is so prone to abuse and misinterpretation that in 2016, the American Statistical Association issued a statement intended to steer statistical analysis into a “post $p < 0.05$ era.” That “ $p < 0.05$ ”

15. Statistical Inference and Information Theory

approach is embedded in most traditional approaches to actuarial credibility⁸ and therefore should be similarly reconsidered.

15.4.2. Implications for Financial Modeling

Like Bayes' Formula itself, another aspect of actuarial literature that is taught but often glossed over in practice is the difference between process risk (volatility), parameter risk, and model formulation risk. Often when performing analysis that relies on stochastic result, in practice only process/volatility risk is assessed.

Bayesian statistics provides the tools to help actuaries address parameter risk and model formulation. The posterior distribution of parameters derived is consistent with the observed data and modeled relationships. This posterior distribution of parameters can then be run as an additional dimension to the risk analysis.

Additionally, best practices include skepticism of the model construction itself, and testing different formulation of the modeled relationships and variable combinations to identify models which are best fit for purpose. Tools such as Information Criterion, posterior predictive checks, Bayes factors, and other statistical diagnostics can inform the actuary about trade-offs between different choices of model.

i Bayesian Versus Machine Learning

Machine learning (ML) is *fully compatible* with Bayesian analysis - one can derive posterior distributions for the ML parameters like any other statistical model and the combination of approaches may be fruitful in practice.

However, to the extent that actuaries have leaned on ML approaches due to the shortcomings of traditional actuarial approaches, Bayesian modeling may provide an attractive alternative without resorting to notoriously finicky and difficult-to-explain ML models. The Bayesian framework provides an explainable model and offers several analytic extensions beyond the scope of this introductory chapter:

- Causal Modeling: Identifying not just correlated relationships, but causal ones, in contexts where a traditional experiment is unavailable.
- Bayes Action: Optimizing a parameter for, e.g., a CTE95 level instead of a parameter mean.
- Information Criterion: Principled techniques to compare model fit and complexity.
- Missing data: Mechanisms to handle the different kinds of missing data.
- Model averaging: Posteriors can be combined from different models to syn-

⁸Note that the approach discussed here is much more encompassing than the Bühlmann-Straub Bayesian approach described in the actuarial literature.

thesize different approaches.

15.4.3. Basics of Bayesian Modeling

A Bayesian statistical model has four main components to focus on:

1. **Prior** encoding assumptions about the random variables related to the problem at hand, before conditioning on the data.
2. A **Model** that defines how the random variables give rise to the observed outcome.
3. **Data** that we use to update our prior assumptions.
4. **Posterior** distributions of our random variables, conditioned on the observed data and our model

Having defined the first two components and collected our data, the workflow involves computationally sampling the posterior distribution, often using a technique called **Markov Chain Monte-Carlo** (MCMC). The result is a series of values that are sampled statistically from the posterior distribution.

15.4.4. Markov-Chain Monte Carlo

While computing the posterior distribution for most model parameters is analytically intractable, we can probabilistically sample from the posterior distribution and achieve an approximation of the posterior distribution. MCMC samplers, as they are called, do this by moving through the parameter space and travel to different points in proportion to the posterior probability. It is a Markov-Chain because the probability of the next point's location is influenced by the prior sampling point's location.

Here is a simple example demonstrated with one of the oldest MCMC algorithm, called Metropolis-Hastings. The general idea is this:

1. Start at an arbitrary point and make that the `current_state`.
2. Propose a new point which is the `current_state` plus some movement that comes from a random distribution, `proposal_dist`.
3. Calculate the likelihood ratio of the proposed versus current point (`acceptance_ratio` below).
4. Draw a random number - if that random number is less than the `acceptance_ratio`, then move to that new point. Otherwise do not move.
5. Repeat steps 2-4 until the distribution of points converges to a stable posterior distribution.

15. Statistical Inference and Information Theory

Here is what a complete example looks like. We will try to find the posterior of an arbitrary 2D density function. Picking the product of an Exponential and Beta distribution looks like this:

```
σ = 0.15
μ = 0.1

n_observations = 100
return_dist = Normal(μ,σ)
returns = rand(return_dist,n_observations)

# plot the target_distirubution
μ_range = LinRange(-0.5, 0.5, 400)
σ_range = LinRange(0.0, 3.0, 400)

f = Figure()
ax1 = Axis(f[1,1],title="True Distribution of Returns")
ax2 = Axis(f[2,1],title="Simulated Outcomes", xlabel="Return")
plot!(ax1,return_dist)
vlines!(ax1,[μ],color=(:black,0.7))
text!(ax1,μ,0;text="mean ($μ)",rotation=pi/2)
hist!(ax2,returns)
vlines!(ax2,[mean(returns)],color=(:black,0.7))
text!(ax2,mean(returns),0;text="mean ($(round(mean(returns);digits=3)))",rotation=pi/2)

linkxaxes!(ax1,ax2)

f
```

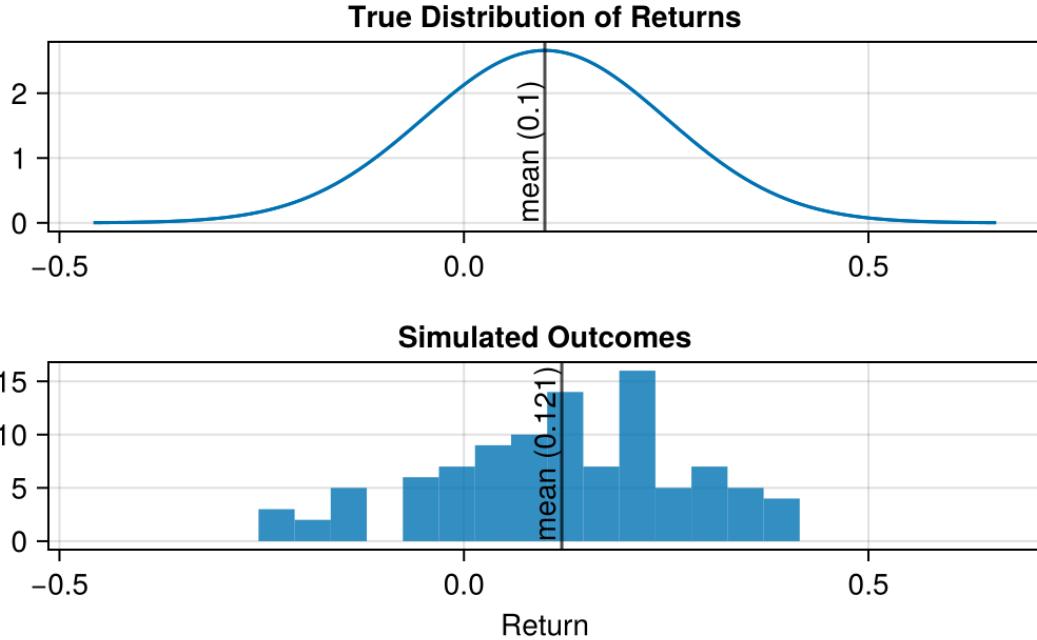


Figure 15.3.: The target probability densities which we will attempt to infer via MCMC.

In reality, we don't observe the parameters we are interested in determining the values for.

We will next define a probability distribution for the random step that we take from the `current_point`. We choose a 2D Gaussian for this. The `proposal_std` controls how big of a movement is taken at each step.

```
# Define the proposal distribution
proposal_std = 0.05
proposal_dist = Normal(0,proposal_std)
```

`Normal{Float64}($\mu=0.0$, $\sigma=0.05$)`

We next define how many steps we want the chain to sample for, and implement the algorithm's main loop containing the logic described above. The resulting chain contains a list of points that the algorithm has moved along during the sampling process. Note that there is a `burn_in` parameter. This is because we want the chain to be effectively independent of both (1) the starting point for the sample, and (2) so that different chains are effectively independent.

15. Statistical Inference and Information Theory

```
# MCMC parameters
num_samples = 5000
burn_in = 500

# Define priors
μ_prior = Normal(0, 0.25)
σ_prior = Gamma(0.5)

# Initialize the Markov chain
μ_current, σ_current = 0.0, 0.25
current_prob = sum(logpdf(Normal(μ_current, σ_current), r) for r in returns) + logpdf(μ_prior,
chain = zeros(num_samples, 2)

count = 0
# MCMC sampling loop
while count < num_samples
    # Generate a new proposal
    ḡ, ḡ = μ_current + rand(proposal_dist), σ_current + rand(proposal_dist)
    if ḡ > 0

        # Calculate the acceptance ratio

        proposal_prob = sum(logpdf(Normal(ḡ, ḡ), r) for r in returns) +
                        logpdf(μ_prior, ḡ) + logpdf(σ_prior, ḡ)
        log_acceptance_ratio = proposal_prob - current_prob

        # Accept or reject the proposal
        if log(rand()) < log_acceptance_ratio
            μ_current, σ_current = ḡ, ḡ
            current_prob = proposal_prob
        end

        # Store the current state as a sample
        count += 1
        chain[count, :] .= μ_current, σ_current
    else
        # skip because σ can't be negative
    end
end

chain
```

```
5000×2 Matrix{Float64}:
0.0          0.25
0.0159363  0.17485
0.0293783  0.230514
0.0293783  0.230514
0.0293783  0.230514
0.0293783  0.230514
0.0293783  0.230514
0.0293783  0.230514
0.0319267  0.14624
0.0319267  0.14624
0.086885   0.16042
0.086885   0.16042
0.086885   0.16042
0.086885   0.16042
⋮
0.143334   0.176833
0.143334   0.176833
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
0.130407   0.160402
```

After having performed the sampling, we can now visualize the chain versus the target_distribution. A few things to note:

1. The red line indicates the “warm up” or “burn-in” phase and we do not consider that as part of the sampled chain because those values are too correlated with the arbitrary starting point.
2. The blue line indicates the path traveled by the Metropolis-Hastings algorithm. Long asides into low-probability regions are possible, but in general the path will traverse areas in proportion to the probability of interest.

```
# Plot the chain
let
    f = Figure()
    ax1 = Axis(f[1, 1], yreversed=true)

    scatterlines!(ax1, chain[1:burn_in, 1], 1:burn_in, color=(:red, 0.5), markercolor=:red)
```

15. Statistical Inference and Information Theory

```
scatterlines!(ax1,chain[burn_in+1:end,1],burn_in+1:num_samples,color=(:blue,0.5),markercol

ax2 = Axis(f[2,1],xlabel="μ")

hist!(ax2,chain[burn_in+1:end,1],color=:blue)

linkxaxes!(ax1, ax2)

ax3 = Axis(f[1, 2],yreversed=true)

scatterlines!(ax3,chain[1:burn_in,2],1:burn_in,color=(:red,0.5),markercolor=:red)
scatterlines!(ax3,chain[burn_in+1:end,2],burn_in+1:num_samples,color=(:blue,0.5),markercol

ax4 = Axis(f[2,2],xlabel="σ")
hist!(ax4,chain[burn_in+1:end,2],color=:blue)

linkxaxes!(ax3, ax4)

f
end
```

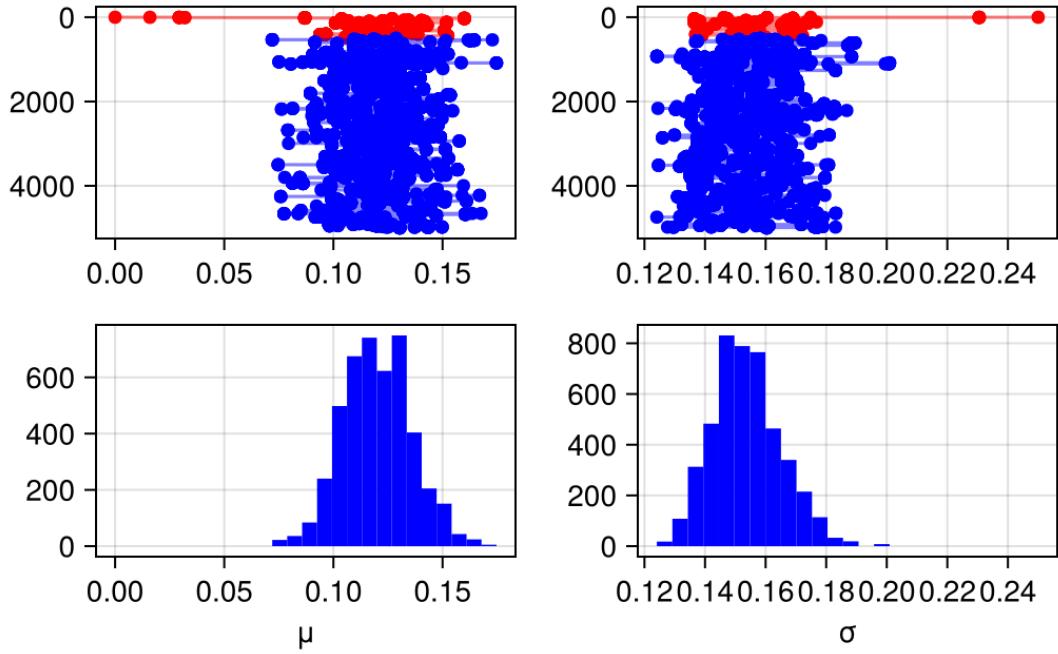


Figure 15.4.: The blue lines of the MCMC chain explore the posterior density of interest (after discarding the burn-in samples in red).

1. Iterate through pairs of points to plot as separate line segments so that the transparent blue line visually stack and the higher density near the center is apparent.

This was a very quick introduction to MCMC sampling, with the point being that the technique is not magic. More advanced libraries will perform the sampling for you with more advanced algorithms than Metropolis-Hastings, but by demonstrating that we could do-it-from-scratch in a small amount of code is intended to make the approach less intimidating.

15.4.5. MCMC Algorithms

The Metropolis-Hastings algorithm is simple, but somewhat inefficient. Some challenges with MCMC sampling are both mathematical and computational:

1. Often times the algorithm will back-track (take a “U-Turn”), wasting steps in regions already explored.
2. The algorithm can have a very high rate of rejecting proposals if the proposal mechanism generates steps that would move the current state into a low-probability regions.

15. Statistical Inference and Information Theory

3. The choice of proposal distribution and parameters can greatly influence the speed of convergence. Too large of movement and key regions can be entirely skipped over, while small movements can take much longer than necessary to explore the space.
4. As the number of parameters grows, the dimensionality of the parameter space to explore also grows making posterior exploration much harder.
5. The shape of the posterior space can be more or less difficult to explore. Complex models may have regions of density that are not nicely “round” - regions may be curved, donut shaped, or disjointed.

The problems above mean that MCMC sampling is very computationally expensive for more complex examples. Compared with Metropolis-Hastings, modern algorithms (such as the No-U-Turn (NUTS)) algorithm explore the posterior distribution more efficiently by avoiding back-tracking to already explored regions and dynamically adjusting the proposals to adaptively fit the posterior. Many of them take direct influence from particle physics, with the algorithm keeping track of the energy of the current state as it explores the posterior space.

Algorithms have only brought so much relief: much falls back onto the modeler to design models that are computationally efficient, transformed to eliminate oddly-shaped density regions, or find the right simplifications to the analysis in order to make the problem tractable.

i Note

What does it mean to transform the parameter space?

An example will be shown in Chapter 28 where we want to ensure that a binomial variable is constrained to the region $[0, 1]$ but the underlying factors are allowed to vary across the entire real numbers. We use a logit (or inverse logit, a.k.a. logistic) to transform the parameters to the required probability range for the binomial outcome.

Another common transform is “Normalizing” the data to center the data around zero and to scale the outcomes such that the sample standard deviation is equal to one.

15.4.6. Rainfall Example (Continued)

We will construct a Bayesian model using the Turing.jl library which allows us to step back from the intricacies of defining our own sampler and routine. The goal is to fit the parameters of one of the competing models from above in order to demonstrate an MCMC analysis workflow and essential concepts.

The first thing that we will do is use Turing’s `@model` macro to define a model. This has a few components:

1. The “model” is really just a Julia function that takes in data and relates the data to the statistical outcomes modeled.
2. The `~` is the syntax to either relate a parameter to a prior assumptions.
3. A loop (or broadcasted `.~`) that ties specific data observations to the random process.

Think of the `@model` block really as a model *constructor*. When you pass data to the model, then you get an instantiated `Model` type⁹.

```
using Turing
```

```
@model function rainLogNormal(logdata) (1)

    # Prior Assumptions for the (Log) Normal Parameters
    μ ~ Normal(4,1) (2)
    σ ~ Exponential(0.5) (3)

    # Link observations to the random process
    for i in 1:length(logdata)
        logdata[i] ~ Normal(μ, σ)
    end
end

m = rainLogNormal(log.(rain));
```

- ① Defining the model uses the `@model` macro from Turing.
- ② We know that there will be positive rainfall and 96% of mean annual rainfall will be between $\exp(2)$ and $\exp(6)$, or 7 and 403 inches.
- ③ In a LogNormal model, 0.5 deviations covers a lot of variation in outcomes.

Precompiling Turing

```
✓ BangBang → BangBangDataFramesExt
✓ Transducers → TransducersDataFramesExt
2 dependencies successfully precompiled in 17 seconds. 386 already precompiled.
Warning: Module DataFrames with build ID fabfcfd-62f2-8f0e-0003-a8c289d47b70 is missing from the
| This may mean DataFrames [a93c6f00-e57d-5684-b7b6-d8193f3e46c0] does not support precompilation
└ @ Base loading.jl:1948
Warning: Module DataFrames with build ID fabfcfd-62f2-8f0e-0003-a8c289d47b70 is missing from the
| This may mean DataFrames [a93c6f00-e57d-5684-b7b6-d8193f3e46c0] does not support precompilation
└ @ Base loading.jl:1948
Precompiling SciMLBaseMakieExt
✓ SciMLBase → SciMLBaseMakieExt
```

⁹Specifically: a `DynamicPPL.Model` type (PPL = Probabilistic Programming Language).

15. Statistical Inference and Information Theory

```
1 dependency successfully precompiled in 9 seconds. 310 already precompiled.
```

```
Warning: Module SciMLBase with build ID fafbfcfd-4eea-96f2-0003-a8d2cf761c55 is missing from the
```

```
This may mean SciMLBase [0bca4576-84f4-4d90-8ffe-ffa030f20462] does not support precompilation
```

```
@ Base loading.jl:1948
```

15.4.6.1. Setting Priors

In the example above, we used “weakly informative” priors. We constrained the prior probability to plausible ranges, knowing enough about the system of study (rainfall) that it would be completely implausible for there to be a $\text{Uniform}(0, \text{Inf})$ distribution of mean log-rainfall total. We admit a level of subjectivity when we encode into the model some limitations on what we would believe to be true. Admittedly, we haven’t confirmed with a meteorologist that $\exp(20)$ (485 million) inches of rain per year is impossible. But such is the beauty of the transparency of Bayesian analysis that the prior assumption is right there! Front and center!

“Strongly informative” priors would be something where we want to encode a stronger assumption about the plausible range of outcomes, such as if we knew enough about the problem domain that we could tell given the location of the rainfall, we’d expect 95% of the rainfall to be between, say, 10 and 30 inches per year.

“Uninformative” priors use only maximum entropy or uniform priors to avoid encoding bias into the model.

15.4.6.2. Sampling

Analysis should begin by evaluating the prior assumptions for reasonability and coverage over possible outcomes of the process we are trying to model. The top plot in Figure 15.8 shows the modeled rainfall outcomes taking on a wide range of possible outcomes. If we had more knowledge of the system we could enforce a stronger (narrower) prior assumption to constrain the model to a smaller set of values.

The object returned is an MCMCChains structure containing the samples as well as diagnostic information. Summary information gets printed below.

```
chain_prior = sample(m, Prior(), 1000)
```

We now sample the posterior by using the No-U-Turns (NUTS) algorithm and drawing 1000 samples (not including the warm-up phase). This is the primary result we will analyze further.

```
chain_posterior = sample(m, NUTS(), 1000)
```

Chains MCMC chain (1000×3×1 Array{Float64, 3}):

```
Iterations      = 1:1:1000
Number of chains = 1
Samples per chain = 1000
Wall duration    = 0.18 seconds
Compute duration = 0.18 seconds
parameters       = μ, σ
internals        = lp
```

Summary Statistics

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
μ	4.0442	0.9953	0.0316	988.1617	1025.2091	0.9998	...
σ	0.4776	0.4650	0.0148	967.9518	908.2284	1.0001	...

1 column omitted

Quantiles

parameters	2.5%	25.0%	50.0%	75.0%	97.5%
Symbol	Float64	Float64	Float64	Float64	Float64
μ	2.1332	3.3771	4.0275	4.7251	5.9615
σ	0.0093	0.1526	0.3456	0.6495	1.6957

Figure 15.5.: Model output for the sampled prior. This isn't running an MCMC algorithm, it's simply taking draws from the defined prior assumptions.

15. Statistical Inference and Information Theory

Chains MCMC chain (1000×14×1 Array{Float64, 3}):

```
Iterations      = 501:1:1500
Number of chains = 1
Samples per chain = 1000
Wall duration    = 2.34 seconds
Compute duration = 2.34 seconds
parameters       = μ, σ
internals        = lp, n_steps, is_accept, acceptance_rate, log_density, hamiltonian_energy, hamil
```

Summary Statistics

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	e ...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
μ	3.5698	0.0520	0.0017	985.5310	738.0395	1.0047	...
σ	0.4506	0.0389	0.0013	986.3862	674.5299	1.0045	...

1 column omitted

Quantiles

parameters	2.5%	25.0%	50.0%	75.0%	97.5%
Symbol	Float64	Float64	Float64	Float64	Float64
μ	3.4691	3.5343	3.5698	3.6066	3.6713
σ	0.3814	0.4230	0.4468	0.4747	0.5341

Figure 15.6.: Model output for the sampled posterior.

15.4.6.3. Diagnostics

Before analyzing the result itself, we should check a few things to ensure the model and sampler were well behaved. MCMC techniques are fundamentally stochastic and randomness can cause an errant sampling path. Or a model may be mis-specified such that the parameter space to explore is incompatible with the current algorithm (or any known so far).

A few things we can check:

First, the `ess` or **effective sample size** which adjusts the number of samples for the degree of autocorrelation in the chain. Ideally, we would be able to draw independent samples from the posterior but due to the Markov-Chain approach the samples can have autocorrelation between neighboring samples. Therefore we collect less information about the posterior in the presence of positive autocorrelation. An `ess` greater than our sample indicates that there was less (negative) autocorrelation than we would have expected for the chain. An `ess` much less than the number of samples indicates that the chain isn't sampling very efficiently but, aside from needing to run more samples, isn't necessarily a problem.

```
ess(chain_posterior)
```

ESS		
parameters	ess	ess_per_sec
Symbol	Float64	Float64
μ	985.5310	420.9872
σ	986.3862	421.3525

Second, the `rhat` (\hat{R}) is the Gelman-Rubin convergence diagnostic and it's value should be very close to 1.0 for a chain that has converged properly. Even a value of 1.01 may indicate an issue and quickly gets worse for higher values.

```
rhat(chain_posterior)
```

R-hat	
parameters	rhat
Symbol	Float64
μ	1.0047
σ	1.0045

15. Statistical Inference and Information Theory

Next, we can look at the “trace” plots for the parameters being sampled (Figure 15.7). These are sometimes called “hairy caterpillar” plots because in a healthy chain sample, we should see a series without autocorrelation and that the values bounce around randomly between individual samples.

```
let
    f = Figure()
    ax1 = Axis(f[1,1],ylabel="μ")
    lines!(ax1,vec(get(chain_posterior,:μ).μ.data))
    ax2 = Axis(f[2,1],ylabel="σ")
    lines!(ax2,vec(get(chain_posterior,:σ).σ.data))
    f
end
```

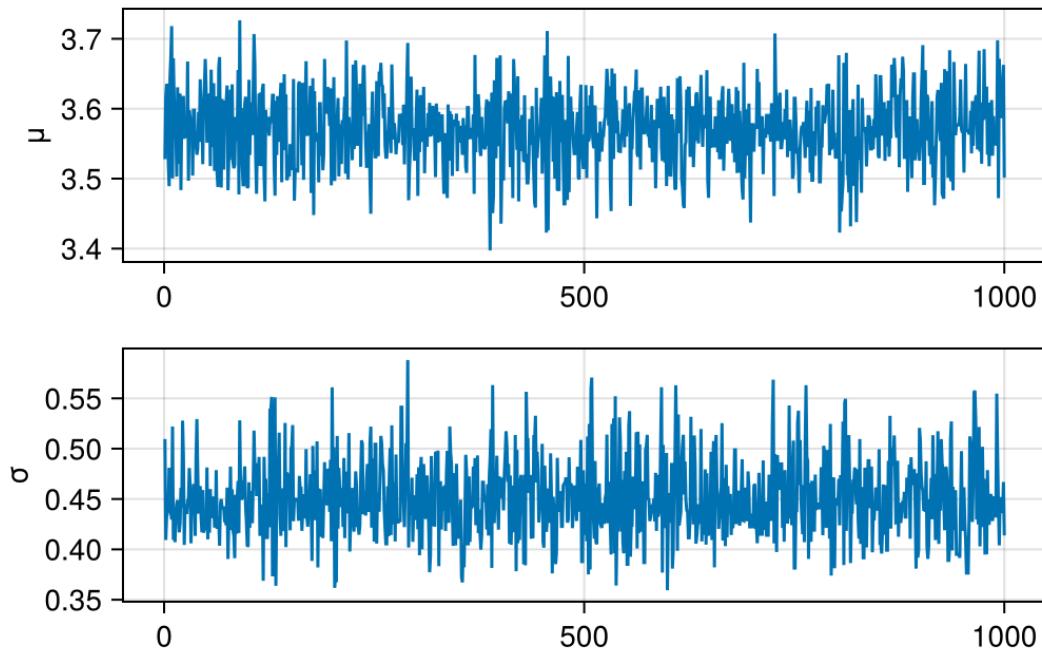


Figure 15.7.: The trace plots indicate low autocorrelation which is desirable for an MCMC sample.

The ess, rhat, and trace plots all look good for our sampled chain so we can next we will analyze the results proper.

15.4.6.4. Analysis

Let's see how it looks compared to the data first. Figure 15.8 shows 200 samples from the prior and posterior. The prior (top) shows how wide the range of possible rainfall outcomes could be using our weakly informative prior assumptions. The bottom shows that after having learned from the data, the posterior probability of rainfall has narrowed considerably.

```

function chn_cdf!(axis,chain,rain)
    n = 200
    s = sample(chain, n)
    vals = get(s, [:μ, :σ])
    ds = Normal.(vals.μ, vals.σ) # ,get(s,:σ)[i])
    rg = 1:200
    for (i, d) in enumerate(ds)
        lines!(axis, rg,cdf.(d,log.(rg)),color=(:gray,0.3))
    end

    # plot the actual data
    percentiles= 0.01:0.01:0.99
    lines!(axis,quantile.(Ref(rain),percentiles),percentiles,linewidth=3)
end

let
    f = Figure()
    ax1 = Axis(f[1,1],title="Prior", xgridvisible=false, ygridvisible=false,ylabel="Quantile")
    chn_cdf!(ax1,chain_prior,rain)

    ax2 = Axis(f[2,1],title="Posterior", xgridvisible=false, ygridvisible=false,xlabel="Annual
    chn_cdf!(ax2,chain_posterior,rain)

    linkxaxes!(ax1, ax2)

    f
end

```

15. Statistical Inference and Information Theory

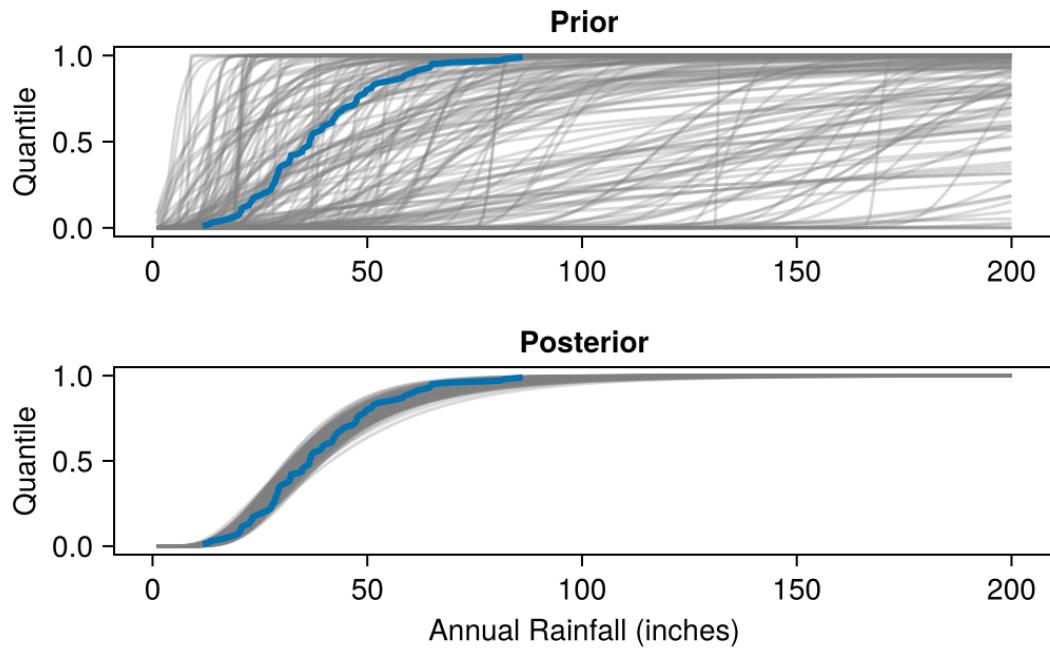


Figure 15.8.: The fitted posterior model (bottom) has good coverage of the observed data (shown in blue).

Comparing to the maximum likelihood analysis from before by plotting the MLE point estimate onto the marginal densities in Figure 15.9. The peak of the the posterior is referred to as the **maximum a posteriori** (MAP) and would be the point estimate proposed by this Bayesian analysis. However, the Bayesian way of thinking about distributions of outcomes rather than point estimates is one of the main aspects we encourage for financial modelers. Using the posterior distribution of the parameters, we can assess parameter uncertainty directly instead of ignoring it as we tend to do with point estimates.

```
let
    # get the parameters from the earlier MLE approach
    p = params(ln)

    f = Figure()

    # plot μ posterior
    ax1 = Axis(f[1,1],title="μ posterior",xgridvisible=false)
    hideydecorations!(ax1)
    d = density!(ax1,vec(get(chain_posterior,:μ).μ.data))
    l = vlines!(ax1,[p[1]],color=:red)
```

```

# plot σ posterior
ax2 = Axis(f[2,1],title="σ posterior", xgridvisible=false)
hideydecorations!(ax2)
density!(ax2,vec(get(chain_posterior,:σ).σ.data))
vlines!(ax2,[p[2]],color=:red)

Legend(f[1,2],[d,l],["Posterior Density", "MLE Estimate"])

f
end

```

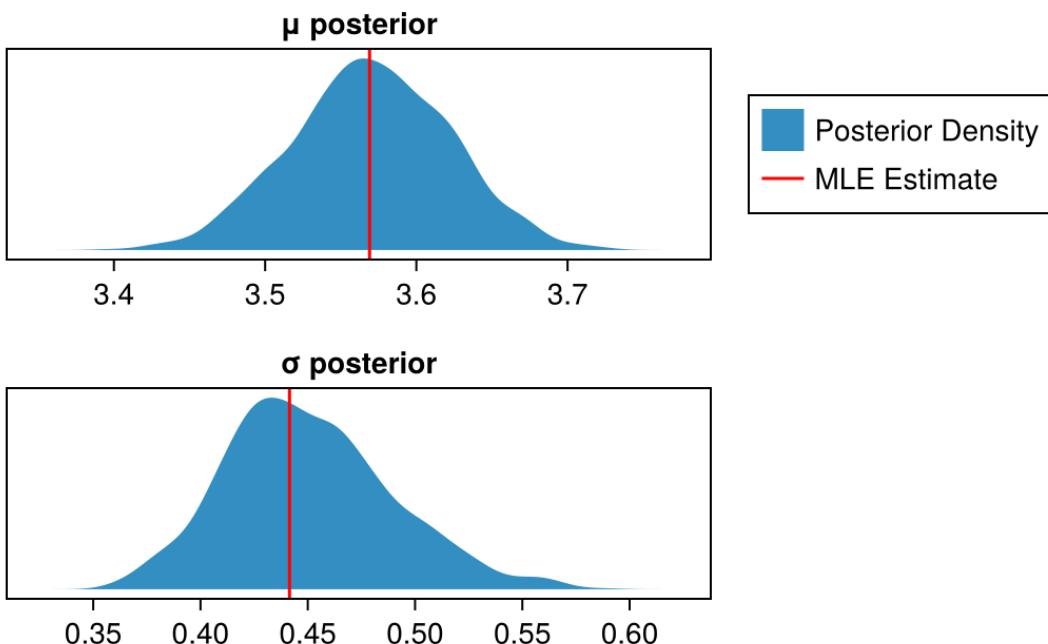


Figure 15.9.: The MLE point estimate need not necessarily align with the peak or center of posterior densities (e.g. in the case of a bimodal distribution).

15.4.6.5. Model Limitations

We have built and assessed a simple statistical model that could be used in the estimation of risk for a particular location. Nowhere in our model did we define a mechanism to capture a more sophisticated view of the world. There is no parameter for changes over time due to climate change, or inter-annual seasonality for El Niño or La Niña cycles,

15. Statistical Inference and Information Theory

or any of a multitude of other real-world factors that can influence the forecasting. All we've defined is that there is a LogNormal process generating rainfall in a particular location. This may or may not be sufficient to capture the dynamics of the problem at hand.

Part of the benefits of the Bayesian approach is that it allows us to extend the statistical model to be arbitrarily complex in order to capture our intended dynamics. We are limited by the availability of data, computational power and time, and our own expertise in the modeling. Regardless of the complexity of the model, the same fundamental techniques and idea apply in the Bayesian approach.

15.4.6.6. Continuing the Analysis

Like any good model, you can often continue the analysis in any number of directions, such as: collecting more data, evaluating different models, creating different visualizations, making predictions about future events, creating a multi-level model that predicts rainfall for multiple related locations simultaneously, among many other

Earlier we discussed model comparison. To compute a real Bayes Factor in comparing the different models, we would take the average likelihood across the posterior samples instead of just comparing the maximum likelihood points as we did earlier. There are more sophisticated tools for estimating out of sample performance of the model, or measures that evaluate a model for over-fitting by penalizing the diagnostic statistic for the model having too many free parameters. See LOO (leave-one-out) cross-validation and various "information criteria" in the resources listed in Section 15.4.8.

15.4.7. Conclusion

This chapter has attempted to make accessible the foundations of statistical inference and the modern tools and approaches available. Underlying this approach to thinking about statistical problems are informational theoretic and mathematical concepts that can be challenging to learn when traditional finance and actuarial curricula is not centered on the necessary computational foundations that are associated with modern statistical analysis. Further, the approach of treating estimation not as an exercise in determining a "best estimate" but instead as a range of outcomes will enhance financial analysis and quantification of risks.

15.4.8. Further Reading

Bayesian approaches to statistical problems are rapidly changing the professional statistical field. To the extent that the actuarial profession incorporates statistical procedures, financial professionals should consider adopting the same practices. The benefits of this

15.4. Modern Bayesian Statistics

are a better understanding of the distribution of risk and return, results that are more interpretable and explainable, and techniques that can be applied to a wider range of problems. The combination of these things would serve to enhance best practices related to understanding and communicating about financial quantities.

Textbooks recommended by the author are:

- Statistical Rethinking (McElreath)
- Bayes Rules! (Johnson, Ott, Dogucu)
- Bayesian Data Analysis (Gelman, et. al.)

Chi Feng has an interactive demonstration of different MCMC samplers available at:
<https://chi-feng.github.io/mcmc-demo/>.

Part V.

Computational Thinking in an Actuarial and Financial Context

16. Modeling

16.1. In This Chapter

We discuss how to approach a problem and identify the key attributes to include in the model, what are the inherent trade-offs with different approaches, and how to work with data that feeds your model.

16.2. Key attributes

When creating a model, whether a data model, a conceptual model, or any other type, certain key attributes are generally important to include to ensure it is effective and useful. Some essential attributes include:

- Objective: Clearly define what the model aims to achieve.
- Boundaries: Specify the limits and constraints of the model to avoid scope creep.
- Variables: Identify and define all variables involved in the model.
- Parameters: Include constants or coefficients that influence the variables.
- Dependencies: Describe how variables interact with each other.
- Relationships: Detail the connections between different components of the model.
- Inputs: Specify the data or resources required for the model to function.
- Outputs: Define what results or predictions the model produces.
- Underlying Assumptions: Document any assumptions made during the model's development to clarify its limitations and validity.
- Validation Criteria: Outline how the model's accuracy and reliability are tested.
- Performance Metrics: Define the metrics used to evaluate the model's performance.
- Scalability: Ensure the model can handle increased data or complexity if needed.
- Adaptability: Allow for adjustments or updates as new information or requirements arise.
- Documentation: Provide comprehensive documentation explaining how the model works, including algorithms, data sources, and methods.
- Transparency: Make the model's workings understandable to stakeholders or users.
- User Interface: Design an intuitive interface if the model is interactive.

16. Modeling

- Ease of Use: Ensure that the model is user-friendly and accessible to its intended audience.
- Ethics: Address any ethical concerns related to the model's application or impact.
- Regulations: Ensure compliance with relevant laws and regulations.

Including these attributes helps create a robust, reliable, and practical model that effectively serves its intended purpose.

16.3. Possible approaches

Different modeling approaches come with their own sets of trade-offs. Common modeling approaches, and the inherent trade-offs, may include:

1. Statistical Models

Examples: Linear regression, logistic regression

Trade-offs: - Simplicity vs. Accuracy: Statistical models are often simpler and more interpretable but may not capture complex relationships as well as more sophisticated models. - Assumptions: These models typically rely on assumptions (e.g., linearity, normality) that may not always hold true, potentially affecting their accuracy.

2. Machine Learning Models

Examples: Decision trees, random forests, neural networks

Trade-offs: - Complexity vs. Interpretability: Machine learning models, especially deep learning models, can capture complex patterns but are often less interpretable. - Overfitting: More complex models risk overfitting the training data, requiring careful validation and tuning to ensure generalizability. - Data Requirements: These models often require large amounts of data to perform well, and their performance can degrade with limited or noisy data.

3. Simulation Models

Examples: Monte Carlo simulations, agent-based models

Trade-offs: - Accuracy vs. Computational Expense: Simulations can model complex systems and scenarios but can be computationally expensive and time-consuming. - Detail vs. Generalization: High-fidelity simulations can be very detailed but might be overkill for problems where a simpler model would suffice.

4. Theoretical Models

16.3. Possible approaches

Examples: Economic models, physical models

Trade-offs: - Precision vs. Practicality: Theoretical models provide a foundational understanding but may rely on idealizations or simplifications that don't fully capture real-world complexities. - Applicability: They may be highly accurate in specific contexts but less applicable to broader or more variable situations.

5. Hybrid Models

Examples: Combining statistical and machine learning approaches, or combining theoretical and simulation models

Trade-offs: - Complexity vs. Versatility: Hybrid models aim to leverage the strengths of different approaches but can be complex to design and manage. - Integration Challenges: Combining different types of models may present challenges in integrating them effectively and ensuring consistency in their outputs.

6. Empirical Models

Examples: Time series forecasting, econometric models

Trade-offs: - Data Dependence vs. Predictive Power: Empirical models rely heavily on historical data and may not perform well in scenarios where patterns change or data is sparse. - Context Sensitivity: These models can be very accurate for the specific data they are trained on but might not generalize well to different contexts or conditions.

7. Probabilistic Models

Examples: Bayesian networks, probabilistic graphical models

Trade-offs: - Flexibility vs. Computational Complexity: Probabilistic models can handle uncertainty and complex relationships but often require more sophisticated computations and can be harder to implement and interpret.

8. Summary of Common Trade-offs:

- Complexity vs. Simplicity: More complex models can capture more nuanced details but are harder to understand and manage.
- Accuracy vs. Interpretability: High-accuracy models may be less interpretable, making it harder to understand their decision-making process.
- Data Requirements: Some models require large amounts of data or very specific types of data, which can be a limitation in practice.
- Computational Resources: More sophisticated models or simulations can require significant computational power, which may not always be feasible.

Understanding these trade-offs helps in selecting the most appropriate modeling approach based on the specific needs of the problem at hand and the resources available.

16. Modeling

16.4. How to work with data that feeds the models

Working effectively with data that feeds the models involves several key steps to ensure the data is suitable for modeling and that the model performs well. The steps may include:

1. Data Collection

- Source Identification: Identify and gather data from relevant and reliable sources.
 - Data Acquisition: Use appropriate methods for collecting data, such as web scraping, surveys, sensors, or databases.

2. Data Exploration and Understanding

- Descriptive Statistics: Generate summary statistics (mean, median, standard deviation) to understand the data's central tendencies and variability.
- Visualization: Use plots (histograms, scatter plots, box plots) to visually inspect distributions and relationships between variables.
- Data Profiling: Assess data quality, completeness, and consistency.

3. Data Cleaning

- Handling Missing Values: Decide how to address missing data—options include imputation, interpolation, or removing incomplete records.
- Outlier Detection: Identify and handle outliers that may affect model performance. Outliers can be treated or removed based on their cause and impact.
- Data Transformation: Normalize or standardize data if needed, especially if the model requires data in a specific format or scale.

4. Feature Engineering

- Feature Selection: Choose relevant features that contribute to the model's predictive power. This may involve techniques like correlation analysis or feature importance scores.
- Feature Creation: Create new features from existing data that might provide additional insights or improve model performance. This could include polynomial features, interaction terms, or domain-specific transformations.

5. Data Splitting

- Training and Testing Sets: Split the data into training and testing sets (and sometimes a validation set) to evaluate model performance and avoid overfitting.
- Cross-Validation: Use cross-validation techniques (e.g., k-fold cross-validation) to assess model performance on different subsets of the data.

6. Data Preprocessing

- Scaling and Normalization: Apply techniques such as min-max scaling or z-score normalization to ensure features are on a similar scale.
- Encoding Categorical Variables: Convert categorical variables into numerical formats using methods like one-hot encoding or label encoding.
- Data Augmentation: For certain applications (e.g., image processing), augment the data to increase the size and variability of the dataset.

7. Data Integration

- Combining Datasets: If using multiple data sources, merge datasets carefully, ensuring consistent formats and handling discrepancies.
- Data Alignment: Ensure that the data from different sources are aligned in terms of timing, units, and granularity.

8. Data Storage and Management

- Data Warehousing: Store data in a structured format that facilitates easy access and management, such as databases or data lakes.
- Version Control: Track changes to datasets over time to maintain reproducibility and manage updates.

9. Ethical Considerations

- Bias and Fairness: Evaluate data for biases and ensure that the model does not perpetuate or amplify them.
- Privacy: Protect sensitive information and comply with data privacy regulations such as GDPR or CCPA.

10. Continuous Monitoring and Updating

- Performance Monitoring: Regularly assess the model's performance using new data and update the model as needed.
- Data Drift: Monitor for changes in data distribution over time (data drift) and retrain the model if necessary.

By following these steps, one can effectively manage data for your model, ensuring that it is clean, relevant, and capable of delivering accurate and reliable results.

16.5. Parsimony

Parsimony in modeling refers to the principle of simplicity: choosing the simplest model that sufficiently explains the data or solves a problem. A parsimonious model aims to achieve a balance between complexity and performance, avoiding overfitting while capturing essential patterns.

16. Modeling

- Key Aspects of Parsimony

Simplicity vs. Complexity - Simplicity: A model with fewer parameters and a simpler structure is easier to interpret and less likely to overfit. - Complexity: More complex models may fit the training data better but can be prone to overfitting and may lack generalizability.

- Occam's Razor

This principle suggests that among competing hypotheses, the one with the fewest assumptions should be selected. In modeling, this translates to preferring simpler models when they perform comparably to more complex ones.

- Model Selection Criteria
 - AIC (Akaike Information Criterion): Penalizes models for the number of parameters, balancing goodness-of-fit with model complexity.
 - BIC (Bayesian Information Criterion): Similar to AIC but includes a stronger penalty for complexity, especially for larger datasets.
 - Cross-Validation: Helps assess model performance on unseen data, which aids in choosing a parsimonious model that generalizes well.
- Trade-offs
 - Bias-Variance Trade-off: Simpler models have higher bias but lower variance, while more complex models have lower bias but higher variance. Parsimonious models aim to balance this trade-off.
 - Interpretability vs. Performance: Simpler models are often more interpretable but might not capture all nuances of the data as well as more complex models.
- Techniques for Achieving Parsimony
 - Feature Selection: Choose only the most relevant features to reduce complexity.
 - Regularization: Techniques like Lasso (L1 regularization) and Ridge (L2 regularization) add penalties to the size of coefficients to encourage simpler models.
 - Pruning: In decision trees and neural networks, pruning can simplify the model by removing nodes or connections that contribute little to performance.
- Model Evaluation
 - Generalization: A parsimonious model should generalize well to new, unseen data, demonstrating good performance across different datasets.
 - Complexity Management: Evaluate the model's performance in terms of both accuracy and complexity to ensure it's neither too simple nor unnecessarily complex.

16.5. Parsimony

- Practical Considerations
- Context and Domain Knowledge

The appropriate level of parsimony can depend on the specific problem domain and the nature of the data. Domain expertise can guide the selection of relevant features and model structure.

- Model Evolution

Start with a simple model and incrementally increase complexity only if necessary. This iterative approach helps in understanding the impact of added complexity.

- Computational Efficiency

Parsimonious models are generally more computationally efficient, which is beneficial for large datasets or real-time applications.

- Model Interpretation

Simpler models are easier to explain to stakeholders and can be crucial when the model's decisions need to be transparent and justifiable.

- Summary

Parsimony involves favoring simpler models that are still capable of capturing essential patterns in the data. By achieving a balance between complexity and performance, a parsimonious model avoids overfitting, improves interpretability, and often enhances generalizability. Techniques such as feature selection, regularization, and cross-validation are essential for developing parsimonious models. Ultimately, the goal is to find the simplest model that performs well on unseen data while maintaining clarity and efficiency.

17. Automatic Differentiation

17.1. In This Chapter

Harnessing the chain rule to compute derivatives not just of simple functions, but of complex programs.

17.2. Motivation for (Automatic) Derivatives

Derivatives are one of the most useful analytical tools we have. Determining the rate of change with respect to an input is effectively sensitivity testing. Knowing the derivative lets you optimize things faster (see Chapter 18). You can test properties and implications (monotonicity, maxima/minima).

17.3. Finite Differentiation

Finite differentiation is evaluating a function $f(x)$ at a value x and then at a nearby value $x+\epsilon$. The line drawn through these two points effectively estimates the line that is tangent to the function f at x : effectively the derivative has been found by approximation. That is, we are looking to approximate the derivative using the property:

$$f'(x) = \lim_{\epsilon \rightarrow 0} \frac{f(x_0 + \epsilon) - f(x_0)}{\epsilon}$$

We can approximate the result by simply choosing a small ϵ .

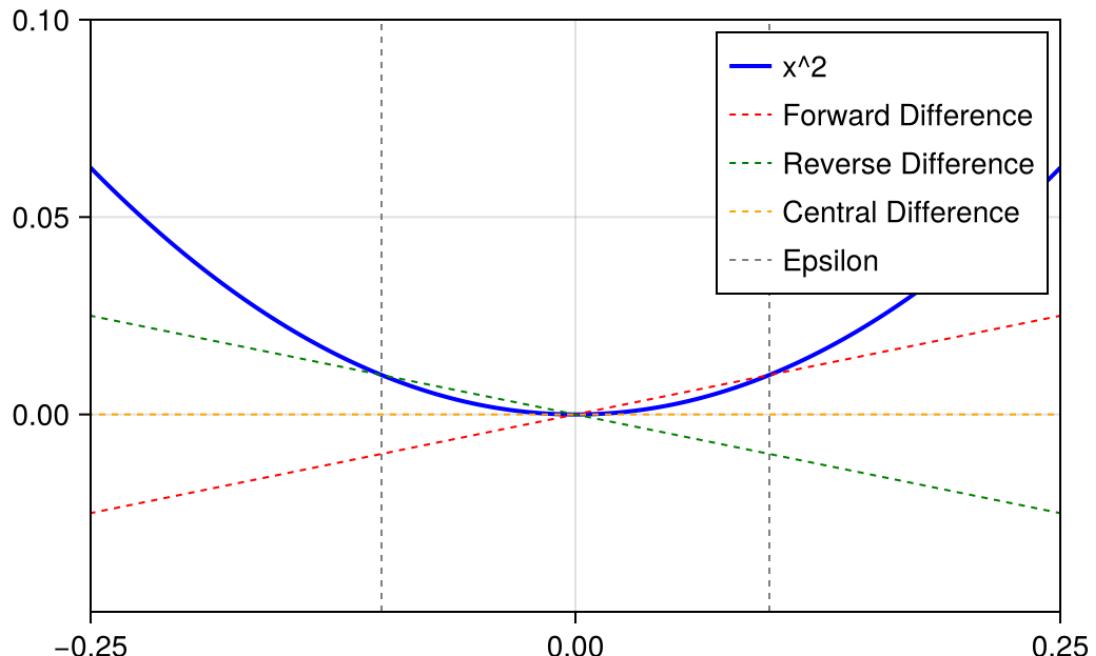
There's also flavors of finite differentiation to approximate derivatives to be aware of:

- forward difference is as defined in the above equation, where ϵ is added to x_0
- reverse difference is as defined in the above equation, where ϵ is subtracted from x_0
- central difference is where we evaluate at $x_0 \pm \epsilon$ and then divide by 2ϵ

17. Automatic Differentiation

The benefit of the central difference is that it limits issues around minima and maxima where the trough or peak respectively would seem much steeper if using forward or reverse. Here's a picture of this:

```
[ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
[ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```



One benefit of the central difference method is that it is often more accurate than forward or reverse. However it comes at the cost of needing to evaluate the function an additional time in many circumstances. Take, for example, the process of optimizing a function to find a maxima or minima. The process usually involves guessing an initial point, evaluating the function at that point, and determining what the derivative of the function is at that point. Both items are used to update the guess to one that's closer to the solution. This approach is used in many optimization algorithms such as Newton's Method.

At each step you need to evaluate the function three times: for x , $x + \epsilon$, and $x - \epsilon$. With forward or reverse finite differences, you can reuse the prior function evaluation of the prior guess x . As one of the components in the estimation of the derivative, thereby saving an evaluation of the function for each iteration.

17.3. Finite Differentiation

There are additional challenges with the finite differences method. In practice, we are often interested in much more complex functions than x^2 . For example, we may actually be interested in the sum of a series that is many elements long or contains more complex operations than basic algebra. In the prior example, the ϵ is set unusually wide for demonstration purposes. As ϵ grow smaller generally, the accuracy of all three finite different methods increases. However, that's not always the case due to both the complexity of the function that you may be trying to differentiate or due to numerical inaccuracies of floating point math.

To demonstrate, here is a more complex example using an arbitrary function

for this example we'll show the results of the three methods calculated at different values of ϵ :

```
using DataFrames
```

```
f(x) = exp(x)
ε = 10 .^ (range(-16, stop=0, length=100))
x₀ = 1
estimate = @. (f(x₀ + ε) - f(x₀ - ε)) / 2ε
actual = f(x₀)                                ①

fig = Figure()
ax = Axis(fig[1, 1], xscale=log10,yscale=log10, xlabel="ε", ylabel="absolute error")
scatter!(ax, ε, abs.(estimate .- actual))
fig
```

- ① The derivative of $f(x) = \exp(x)$ is itself. That is $f'(x) = f(x)$ in this special case.

```
┌ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/julia/packages/Makie/GtFuI/src/scenes.jl:227
```

17. Automatic Differentiation

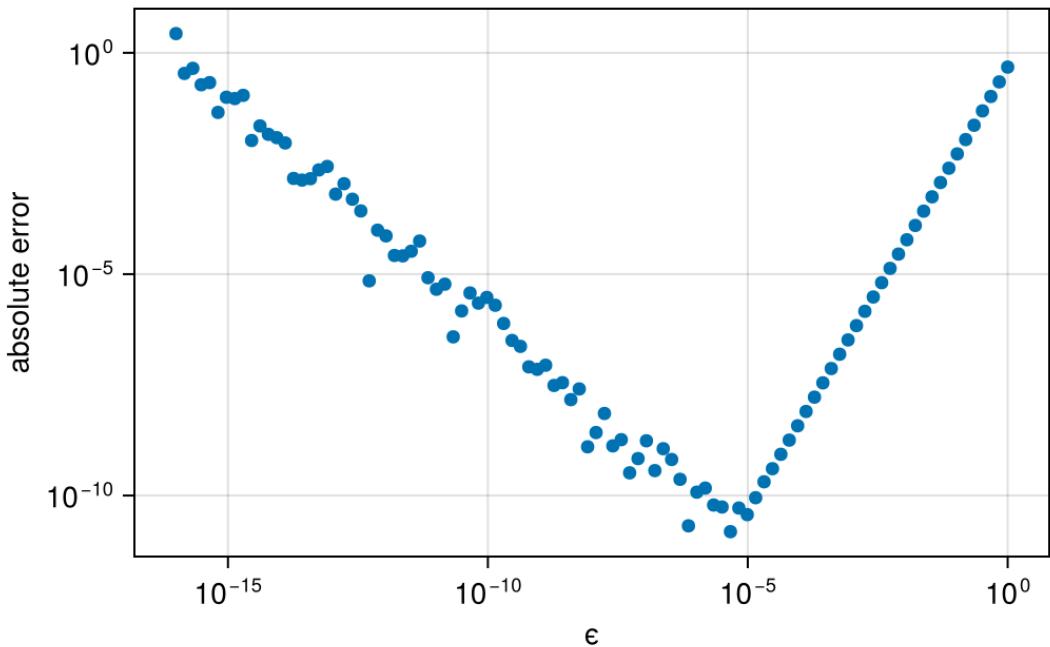


Figure 17.1.: A log-log plot showing the absolute error of the finite differences. Further to the left, roundoff error dominates while further to the right, truncation error dominates.

Note

The `@.` in the code example above is a macro that applies broadcasting each function to its right. `@. (f(x0 + ε) - f(x0 - ε)) / 2ε` is the same as `(f.(x0 .+ ε) .- f.(x0 .- ε)) ./ (2 .* ε)`

A few observations:

1. At virtually every value of ϵ we observe some error from the true derivative.
2. That error is the sum of two parts: **truncation error** is inherent in that we are using a given value for ϵ and not determining the limiting analytic value as $\epsilon \rightarrow 0$. The other component is **roundoff error** which arises due to the limited precision of floating point math.

The implications of this are that we need to often be careful about the choice of ϵ , as the optimal choice will vary depending on the function and the point we are attempting to evaluate. This presents a number of practical difficulties in various algorithms.

Additionally, when computing the finite difference we must evaluate the function multiple times to determine a single estimate of the derivative. When performing something

like optimization the process typically involves iteratively making many guesses — plus the number of guesses required to find the right answer can depends on the ability to accurate determine the derivative at a point!

Admittedly, despite the accuracy and computational overhead, finite differences can be very useful in many circumstances. However, a more appealing alternative approach will be covered next.

17.4. Automatic Differentiation

Automatic differentiation (“autodiff” or “AD” for short) is essentially the practice of defining algorithmically what the derivatives of function should be. We are able to do this through a creative application of the chain rule. Recall that the **chain rule** allows us to compute the derivative of a composite function using the derivatives of the component functions:

$$\begin{aligned} h(x) &= f(g(x)) \\ h'(x) &= f'(g(x))g'(x) \end{aligned}$$

Using this rule, we can define how elementary operations act when differentiated. Combined with the fact that most computer code is building up from a bunch of elementary operations, we can get a very long way in differentiating complex functions.

17.4.1. Dual Numbers

To understand where we are going, let's remind ourselves about complex numbers. Complex numbers are of the form which has an real part (r) and an imaginary part (iq):

$$r + iq$$

By definition we say that $i^2 = -1$. This is useful because it allows us to perform certain types of operations (e.g. finding a square root of a negative number) that is otherwise unsolvable with just the real numbers¹. After defining how the normal algebraic operations (addition, multiplication, etc.) work for the imaginary number, we are able to utilize the imaginary numbers for a variety of practical mathematical tasks.

¹Richard Feynman has a wonderful, short lecture on algebra here:
https://www.feynmanlectures.caltech.edu/I_22.html

17. Automatic Differentiation

What is meant by extending the algebraic operations for imaginary numbers? For example, stating how addition should work for imaginary numbers:

$$(r + iq) + (s + iu) = (r + s) + i(q + u)$$

In a similar fashion as extending the Real (\mathbb{R}) numbers with an *imaginary* part, for automatic differentiation we will extend them with a *dual* part. A **dual number** is one of the form:

$$a + \epsilon b$$

Where $\epsilon^2 = 0$ and $\epsilon \neq 0$ by definition. For our purposes here, one can think of b as the derivative of the function evaluated at the same point as a . An initial example should make this clearer. First let's define a DualNumber:

```
struct DualNumber{T,U}
    a :: T
    b :: U
    function DualNumber(a :: T, b :: U=zero(a)) where {T,U}
        return new{T,U}(a, b)
    end
end
```

- ① We define this type parametrically to handle all sorts of `<:Real` types and allow `a` and `b` to vary types in case a mathematical operation causes a type change (e.g. as in the case of integers becoming a floating point number like `10/4 == 2.5`)
- ② `zero(a)` is a generic way to create a value equal to zero with the same type of the argument `a`. `zero(12.0) == 0.0` and `zero(12) == 0`.

Now let's define how dual numbers work under addition. The mathematical rule is:

$$(a + \epsilon b) + (c + \epsilon d) = (a + c) + (b + d)\epsilon$$

We then need to define how it works for the combinations of numbers that we might receive as arguments to our function (this is an example where multiple dispatch greatly simplifies the code compare to object oriented single dispatch!):

```
Base.:+(d::DualNumber, e::DualNumber) = DualNumber(d.a + e.a, d.b + e.b)
Base.:+(d::DualNumber, x) = DualNumber(d.a + x, d.b)
Base.:+(x, d::DualNumber) = d + x
```

And here's how we would get the derivative of a very simple function:

```
f1(x) = 5 + x
f1(DualNumber(10, 1))

DualNumber{Int64, Int64}(15, 1)
```

That's not super interesting though - the derivative of f_1 is just 1 and we supplied that in the construction of `DualNumber`. We did at least prove that we can add the 10 and 5!

Let's make this more interesting by also defining the multiplication operation on dual numbers. We'll follow the product rule:

$$(u \times v)' = u' \times v + u \times v'$$

```
Base.*(d::DualNumber, e::DualNumber) = DualNumber(d.a * g.a, d.b * g.a + d.a * g.b)
Base.(x, d::DualNumber) = DualNumber(d.a * x, d.b * x)
Base.(d::DualNumber, x) = x * d
```

Now what if we evaluate this function:

```
f2(x) = 5 + 3x
f2(DualNumber(10, 1))
```

```
DualNumber{Int64, Int64}(35, 3)
```

We have found that the second component is 3, which is indeed the derivative of $5 + 3x$ with respect to x . And in the first part we have the value of f_2 evaluated at 10.

i Note

When calculating the derivative, why do we start with 1 in the dual part of the number? Because the derivative of a variable with respect to itself is 1. From this unitary starting point, the various operations applied accumulate the derivative of the various operations in the b part of $a + \epsilon b$.

We can also define this for things like transcendental functions:

```
Base.exp(d::DualNumber) = DualNumber(exp(d.a), exp(d.a) * d.b)
Base.sin(d::DualNumber) = DualNumber(sin(d.a), cos(d.a) * d.b)
Base.cos(d::DualNumber) = DualNumber(cos(d.a), -sin(d.a) * d.b)
exp(DualNumber(1, 1))
```

17. Automatic Differentiation

```
DualNumber{Float64, Float64}(2.718281828459045, 2.718281828459045)
```

```
sin(DualNumber(0, 1))
```

```
DualNumber{Float64, Float64}(0.0, 1.0)
```

```
cos(DualNumber(0, 1))
```

```
DualNumber{Float64, Float64}(1.0, -0.0)
```

And finally, to put it all together in a more usable wrapper, we can define a function which will calculate the derivative of another function at a certain point:

```
derivative(f, x) = f(DualNumber(x, one(x))).b
```

```
derivative (generic function with 1 method)
```

And then evaluating it on a more complex function like $f(x) = 5e^{\sin(x)} + 3x$ at $x = 0$, we would analytically derive 8, which matches what we calculate next:

```
let
    f(x) = 5 * exp(sin(x)) + 3x
    derivative(f, 0)
end
```

```
8.0
```

We have demonstrated that through the clever use of dual numbers and the chain rule that complex expressions can be automatically differentiated by the computer to an exact level, limited only by the same machine precision that applies to our primary function of interest as well.

17.5. Performance of Automatic Differentiation

Recall that in the finite difference method, we generally had to evaluate the function two or three times to *approximate* the derivative. Here we have a single function call that provides both the value and the derivative at that value. How does this compare performance-wise to simply evaluating the function a single time?

```
using BenchmarkTools
@btime f2(rand())
```

17.6. Automatic Differentiation in Practice

```
2.541 ns (0 allocations: 0 bytes)
```

```
5.700791782272078
```

```
@btime f2(DualNumber(rand(), 1))
```

```
2.583 ns (0 allocations: 0 bytes)
```

```
DualNumber{Float64, Int64}(6.693653587892538, 3)
```

In performing this computation, the compiler has been able to optimize it such that we effectively are able to compute the function and its derivative at effectively the same speed as just evaluating the function itself! As the function gets more complex, the overhead does increase but is still a *much* preferred option versus finite differentiation. This advantage becomes more pronounced as we contemplate derivatives with respect to many variables at once or for higher-order derivatives.

i Note

In fact, it's largely due to the advances in applications of automatic differentiation that has led to the explosion of machine learning and artificial intelligence techniques in the 2010s/2020s. The "learning" process relies on solving parameter weights and would be too computationally expensive if using finite differences. These applications of autodifferentiation in specialized C++ libraries underpin the libraries like PyTorch, Tensorflow, and Keras. These libraries specialize in allowing for autodiff on a limited subset of operations. Julia's available automatic differentiation is more general and can be applied to many more scenarios.

17.6. Automatic Differentiation in Practice

We have, of course, not defined an exhaustive list of operations, covering only `+`, `*`, `exp`, `sin`, and `cos`. There are only a few more arithmetic (`-`, `/`) and transcendental (`log`, more trigonometric functions, etc.) before we would have a very robust set of algebraic operations defined for our `DualNumber`. In fact, it's possible to go even further and to define the behavior through conditional expressions and iterations to differentiate fairly complex functions or to extend the mechanism to partial derivatives and higher-order derivatives as well.

17. Automatic Differentiation

```
import Distributions
import ForwardDiff

N(x) = Distributions.cdf(Distributions.Normal(), x)

function d1(S, K, τ, r, σ, q)
    return (log(S / K) + (r - q + σ^2 / 2) * τ) / (σ * √(τ))
end

function d2(S, K, τ, r, σ, q)
    return d1(S, K, τ, r, σ, q) - σ * √(τ)
end

"""

eurocall(parameters)
```

Calculate the Black-Scholes implied option price for a european call where 'parameters' is a vector.

- 'S' is the current asset price
- 'K' is the strike or exercise price
- 'τ' is the time remaining to maturity (can be typed with \\tau[tab])
- 'r' is the continuously compounded risk free rate
- 'σ' is the (implied) volatility (can be typed with \\sigma[tab])
- 'q' is the continuously paid dividend rate

```
"""

function eurocall(parameters)
    S, K, τ, r, σ, q = parameters
    iszero(τ) && return max(zero(S), S - K)
    d₁ = d1(S, K, τ, r, σ, q)
    d₂ = d2(S, K, τ, r, σ, q)
    return (N(d₁) * S * exp(τ * (r - q)) - N(d₂) * K) * exp(-r * τ)
end
```

- ① We put the various variables inside a single parameters vector to allow calling a single gradient call instead of multiple derivative calls for each parameter.

eurocall

```
S = 1.0
K = 1.0
τ = 30 / 365
r = 0.05
σ = 0.2
q = 0.0
```

```
params = [S, K, τ, r, σ, q]
eurocall(params)
```

```
0.02493376819403728
```

 Tip

Some terminology in differentiation:

- **Derivative** is generally the scalar rate of change in output relative to a scalar input and can be used in the context of partial derivatives for a multi-variate function (e.g. $\frac{d}{dx} f(x, y, z)$).
- **Gradient** is the first derivative with respect to all dimensions of a function that outputs a scalar. For a function $f(x, y, z)$ the gradient would be a vector of partial derivatives such that you would get $[\frac{d}{dx}, \frac{d}{dy}, \frac{d}{dz}]$
- **Jacobian** is the first derivative with respect to all dimensions of a function that outputs a vector.
- **Hessian** is the second derivative with respect to all dimensions of a function that outputs a scalar.

With the above code, now we can get the partial derivatives with respect to each parameter. The first, third, fourth, fifth, and sixth correspond to the common “greeks” *delta*, *theta*, *rho*, *vega*, and *epsilon* respectively. The second term is the parital derivative with respect to the strike price:

```
ForwardDiff.gradient(eurocall, params)
```

6-element Vector{Float64}:

```
0.5399635456230838
-0.5150297774290467
0.16420676980838977
0.042331214583209334
0.11379886104405816
-0.04438056539367815
```

We can also get the second order greeks with a simple call. In addtion to many uncommon second order parital derivatives. *Gamma* is in the [1,1] position for example:

```
ForwardDiff.hessian(eurocall, params)
```

6x6 Matrix{Float64}:

6.92276	-6.92276	0.242297	0.568994	-0.0853491	-0.613375
---------	----------	----------	----------	------------	-----------

17. Automatic Differentiation

```
-6.92276    6.92276   -0.07809   -0.526663   0.199148   0.568994  
 0.242297   -0.07809   -0.846846   0.521448   0.685306   -0.559878  
 0.568994   -0.526663   0.521448   0.0432874  -0.0163683  -0.0467667  
-0.0853491   0.199148   0.685306   -0.0163683  0.00245525  0.007015  
-0.613375   0.568994   -0.559878   -0.0467667  0.007015   0.0504144
```

17.6.1. Performance

Earlier we examined the impact on performance for the derivatives using the `DualNumber` developed in this chapter on a very basic function. What about if we take a more realistic example like `eurocall`? We can observe approximately a 9x slowdown when computing all of the first order derivatives which isn't bad considering we are computing 6x of the outputs!

```
@btime eurocall($params)
```

```
35.834 ns (0 allocations: 0 bytes)
```

```
0.02493376819403728
```

```
let  
  g = similar(params)                                     ①  
  @btime ForwardDiff.gradient!($g, eurocall, $params)  
end
```

- ① To avoid benchmarking allocating a new array we are able to pre-allocate the memory to store the result and then call `gradient!` to fill in `g` for each result.

```
312.762 ns (2 allocations: 704 bytes)
```

```
6-element Vector{Float64}:  
 0.5399635456230838  
-0.5150297774290467  
 0.16420676980838977  
 0.042331214583209334  
 0.11379886104405816  
-0.04438056539367815
```

17.7. Forward Mode and Reverse Mode

The approach of autodiff outlined about is called **forward mode** auto-differentiation where the derivative is brought forward through the computation and accumulated through each step. The alternative to this is to first evaluate the function and then work backwards by accumulating the partial derivatives in what's called **reverse mode** automatic differentiation.

Reverse mode requires more book-keeping because unlike the forward mode the derivative needs to be carried backwards, unlike the `DualNumber` approach of forward mode.

17.8. Practical tips for Automatic Differentiation

Here are a few practical tips to keep in mind.

17.8.1. Choosing between Reverse Mode and Forward Mode

Forward mode is more efficient when the number of outputs is much larger than the number inputs. When the number of inputs is much larger than the number of outputs, then reverse mode will generally be more efficient. Examples of the number of inputs being larger than the outputs might be in a statistical analysis where many features are used to predict a limited number of outcome variables or a complex model with a lot of parameters.

17.8.2. Mutation

Auto-differentiation works through most code, but a particularly tricky part to get right is when values within arrays are mutated (changed). It's possible to do so but may require a little bit more boilerplate to setup. As of 2024, Enzyme.jl has the best support for functions with mutation inside of them.

17.8.3. Custom Rules

Custom rules for new or unusual functions can be defined, but this is an area that should be explored equipped with a bit of calculus and a deeper understanding of both forward-mode and reverse-mode. ChainRules.jl provides an interface for defining additional rules that hook into the AD infrastructure in Julia as well as provide a good set of documentation on how to extend the rules for your custom function.

17. Automatic Differentiation

17.8.4. Available Libraries

- **ForwardDiff.jl** provides robust forward-mode AD.
- **Zygote.jl** is a reverse-mode package with the innovations of being able to differentiate structs in addition to arrays and scalars.
- **Enzyme.jl** is a newer package which allows for both forward and reverse mode, but has the advantage of supporting array mutation. Additionally, Enzyme works at the level of LLVM code (an intermediate level between high level Julia code and machine code) which allows for different, sometimes better, optimizations.

In the authors experience, they would probably recommend ForwardDiff.jl first and then Enzyme.jl if reaching for more advanced functionality or looking for reverse mode.

17.9. References

- [https://book.sciml.ai/notes/08-Forward-Mode_Automatic_Differentiation_\(AD\)_via_High_Dimension](https://book.sciml.ai/notes/08-Forward-Mode_Automatic_Differentiation_(AD)_via_High_Dimension)
- <https://blog.esciencecenter.nl/automatic-differentiation-from-scratch-23d50c699555>

18. Optimization

18.1. In This Chapter

Optimization as root finding or minimization/maximization of defined objectives. Differentiable programming and the benefits to optimization problems. Other non gradient based optimization approaches. Model fitting as an optimization problem.

18.2. Setup

```
using Flux
using LsqFit
using CairoMakie
using JuMP
using GLPK
using Optim
using Distributions
using Random
using LinearAlgebra

[ Info: Precompiling Flux [587475ba-b771-5e3f-ad9e-33799f191a9c]
Precompiling ZygoteColorsExt
  ✓ Zygote → ZygoteColorsExt
    1 dependency successfully precompiled in 12 seconds. 77 already precompiled.
[ Info: Precompiling ZygoteColorsExt [e68c091a-8ea5-5ca7-be4f-380657d4ad79]
  Warning: Module Zygote with build ID fafbfccfd-752b-f4d6-0003-a8fd3fa4d2ab is missing from the cache
  This may mean Zygote [e88e6eb3-aa80-5325-afca-941959d7151f] does not support precompilation but
  @ Base loading.jl:1948
[ Info: Skipping precompilation since __precompile__(false). Importing ZygoteColorsExt [e68c091a-8ea5-5ca7-be4f-380657d4ad79]
```

18.3. Differentiable programming

Differentiable programming is an approach to programming where functions are defined using differentiable operations, allowing automatic differentiation to be applied

18. Optimization

to them. Automatic differentiation is a technique used to efficiently compute derivatives of functions, and it is crucial in many machine learning algorithms, optimization techniques, and scientific computing applications.

Elements in differentiable programming

- Differentiable functions: Functions are defined using operations that are differentiable. These operations include basic arithmetic operations (addition, subtraction, multiplication, division), as well as more complex operations like exponentials, logarithms, trigonometric functions, etc.
- Automatic differentiation (AD): Automatic differentiation is used to compute derivatives of functions with respect to their inputs or parameters. AD exploits the fact that every computer program, no matter how complex, executes a sequence of elementary arithmetic operations (addition, subtraction, multiplication, division), and elementary functions (exponentials, logarithms, trigonometric functions). By applying the chain rule repeatedly to these operations, derivatives of arbitrary order can be computed automatically, accurately to working precision, and using at most a small constant factor more arithmetic operations than the original program. Refer to chapter on automatic differentiation.
- Optimization and machine learning: Differentiable programming is particularly useful in optimization problems, where gradients or higher-order derivatives are required to find the minimum or maximum of a function. It's also widely used in machine learning, where optimization algorithms like gradient descent are used to train models by adjusting their parameters to minimize a loss function.
- Gradient calculation: The gradient plays a crucial role in optimization problems primarily because it provides the direction of the steepest ascent of a function. Optimization algorithms often iteratively update parameters in the direction opposite to the gradient (for minimization problems), which tends to converge towards a local minimum (or maximum for maximization problems). Besides, computing the gradient is often computationally feasible and relatively inexpensive compared to other methods for determining function behavior, such as higher-order derivatives or function evaluations at different points. Beyond just the direction, the magnitude (or norm) of the gradient also indicates how steep the function change is in that direction. This information is used to adjust step sizes in optimization algorithms, balancing between convergence speed and stability.
- Local and global optimization: A local optimal value refers to a solution where the objective function (or cost function) has the best possible value in a neighborhood surrounding that solution. A global optimal value, on the other hand, is the best possible value of the objective function across the entire feasible domain. For smooth and convex functions, the gradient points towards the global minimum (or maximum), making it extremely efficient for finding the optimal

18.4. Gradient-Free Optimization

solution. Even for non-convex functions, the gradient provides valuable information about the direction to move towards improving the objective function value locally.

```
# Define a differentiable function
f(x) = 3x^2 + 2x + 1
# Define an input value
x = 2.0

@show "Value of f(x) at x=$x: ", f(x)
@show "Gradient of f(x) at x=$x: ", gradient(x → f(x), x)

("Value of f(x) at x=$(x): ", f(x)) = ("Value of f(x) at x=2.0: ", 17.0)
("Gradient of f(x) at x=$(x): ", gradient((x->begin
    #= In[4]:7 =
    f(x)
end), x)) = ("Gradient of f(x) at x=2.0: ", (14.0,))

("Gradient of f(x) at x=2.0: ", (14.0,))
```

18.4. Gradient-Free Optimization

This category includes algorithms that do not rely on gradients or derivative information. They often explore the objective function using heuristics or other types of probes to guide the search.

18.4.1. Linear optimization

Linear optimization, also known as linear programming (LP), is a mathematical method for finding the best outcome in a mathematical model with linear relationships. It involves optimizing a linear objective function subject to a set of linear equality and inequality constraints. Linear programming has a wide range of applications across various fields, including operations research, economics, engineering, and logistics.

```
# Define the objective coefficients
c = [1.0, 2.0, 3.0]
# Define the constraint matrix (A) and right-hand side (b)
A = [1.0 1.0 0.0;
      0.0 1.0 1.0]
b = [10.0, 20.0]
# Create a JuMP model
```

18. Optimization

```
linear_model = Model(GLPK.Optimizer)
# Define decision variables
@variable(linear_model, x[1:3] >= 0)
# Define objective function
@objective(linear_model, Max, dot(c, x))
# Add constraints
@constraint(linear_model, constr[i=1:2], dot(A[i, :], x) <= b[i])
# Solve the optimization problem
optimize!(linear_model)

# Print results
println("Objective value: ", objective_value(linear_model))
println("Optimal solution:")
for i in 1:3
    println("\tx[$i] = ", value(x[i]))
end

Objective value: 70.0
Optimal solution:
x[1] = 10.0
x[2] = 0.0
x[3] = 20.0
```

18.4.2. Evolutionary algorithm

An evolutionary algorithm (EA) is a family of optimization algorithms inspired by the principles of biological evolution. They are particularly useful for solving complex optimization problems where traditional gradient-based methods may struggle due to non-linearity, multimodality, or high dimensionality of the search space.

```
# Define the objective function to maximize
function objective_function(x)
    return sum(x .^ 2) # Example function: sum of squares
end
# Define the evolutionary algorithm function
function evolutionary_algorithm(dim::Int, pop_size::Int, max_gen::Int)
    # Initialization
    population = rand(Bool, dim, pop_size) # Random binary population
    best_solution = population[:, argmax([objective_function(sol) for sol in eachcol(population)])]
    best_fitness = objective_function(best_solution)
    # Evolution loop
    for gen in 1:max_gen
```

18.4. Gradient-Free Optimization

```

# Selection: Binary tournament selection
parents = zeros(Bool, dim, pop_size)
for i in 1:pop_size
    idx1, idx2 = rand(1:pop_size, 2)
    parents[:, i] = if objective_function(population[:, idx1]) > objective_function(population[:, idx2])
        population[:, idx1]
    else
        population[:, idx2]
    end
end
# Crossover: Single-point crossover
offspring = zeros(Bool, dim, pop_size)
for i in 1:2:pop_size
    point = rand(1:dim-1)
    offspring[1:point, i] = parents[1:point, i]
    offspring[1:point, i+1] = parents[1:point, i+1]
    offspring[point+1:end, i] = parents[point+1:end, i+1]
    offspring[point+1:end, i+1] = parents[point+1:end, i]
end
# Mutation: Uniform mutation
mutation_rate = 0.05
for i in 1:pop_size
    for j in 1:dim
        if rand() < mutation_rate
            offspring[j, i] = !offspring[j, i] # Flip the bit
        end
    end
end
# Replace population with offspring
population = offspring
# Update best solution found so far
current_best = population[:, argmax([objective_function(sol) for sol in eachcol(population)])
current_best_fitness = objective_function(current_best)
if current_best_fitness > best_fitness
    best_solution = current_best
    best_fitness = current_best_fitness
end
# Print best fitness in each generation (optional)
println("Generation $gen: Best fitness = $best_fitness")
end
return best_solution, best_fitness
end

```

18. Optimization

```
dim = 10          # Dimension of the problem (e.g., number of variables)
pop_size = 50    # Population size
max_gen = 10     # Maximum number of generations
# Run the evolutionary algorithm
best_solution, best_fitness = evolutionary_algorithm(dim, pop_size, max_gen)

# Print the best solution found
println("\nBest solution found:")
println("Solution = ", best_solution)
println("Fitness = ", best_fitness)

Generation 1: Best fitness = 9
Generation 2: Best fitness = 9
Generation 3: Best fitness = 10
Generation 4: Best fitness = 10
Generation 5: Best fitness = 10
Generation 6: Best fitness = 10
Generation 7: Best fitness = 10
Generation 8: Best fitness = 10
Generation 9: Best fitness = 10
Generation 10: Best fitness = 10

Best solution found:
Solution = Bool[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Fitness = 10
```

18.4.3. Simulated annealing

Simulated Annealing (SA) is a probabilistic optimization technique inspired by the annealing process in metallurgy. It is used to find near-optimal solutions to optimization problems, particularly in cases where traditional gradient-based methods may get stuck in local minima/maxima. SA accepts worse solutions with a certain probability, allowing it to explore the search space more broadly initially and then gradually narrow down towards better solutions as it progresses.

```
# Objective function to minimize
function objective_function(x)
    return x^2
end
# Simulated Annealing function
function simulated_annealing(initial_solution, initial_temperature, cooling_rate, num_iterations)
    current_solution = initial_solution
```

```

best_solution = initial_solution
current_temperature = initial_temperature
for iteration in 1:num_iterations
    # Generate a new solution near the current solution
    new_solution = current_solution + randn() # Example: small random change
    # Calculate objective function values
    current_value = objective_function(current_solution)
    new_value = objective_function(new_solution)
    # Decide whether to accept the new solution
    if new_value < current_value || rand() < exp(-(new_value - current_value) / current_te
        current_solution = new_solution
    end
    # Update the best solution found so far
    if objective_function(current_solution) < objective_function(best_solution)
        best_solution = current_solution
    end
    # Cool down the temperature
    current_temperature *= cooling_rate
end
return best_solution, objective_function(best_solution)
end

initial_solution = 5.0      # Initial solution guess
initial_temperature = 10.0   # Initial temperature
cooling_rate = 0.95         # Cooling rate
num_iterations = 1000       # Number of iterations
# Run simulated annealing
best_solution, best_value = simulated_annealing(initial_solution, initial_temperature, cooling

# Print results
println("Best solution found: x = ", best_solution)
println("Objective function value at best solution: ", best_value)

Best solution found: x = 0.0010264028654319377
Objective function value at best solution: 1.0535028421668925e-6

```

18.4.4. Bayesian optimization

Bayesian Optimization (BO) is a powerful technique for global optimization of expensive-to-evaluate black-box functions. It leverages probabilistic models to predict the objective function's behavior across the search space and uses these models to make informed decisions about where to evaluate the function next. This approach efficiently

18. Optimization

balances exploration (searching for promising regions) and exploitation (exploiting regions likely to yield optimal values), making it particularly suitable for optimization problems where function evaluations are costly, such as tuning hyperparameters of machine learning models or optimizing parameters of complex simulations.

```
# Define your objective function to be optimized
function objective(x::Float64)
    return -(x^2 + 0.1 * sin(5 * x)) # Example objective function (negative because we seek minima)
end
# Bayesian optimization function
function bayesian_optimization(objective, bounds::Tuple{Float64,Float64}, num_iterations::Int)
    Random.seed!(1234) # Setting a seed for reproducibility
    X = Float64[] # List to store evaluated points
    Y = Float64[] # List to store objective values
    # Initial random point (you can choose other initial points as well)
    x_init = rand() * (bounds[2] - bounds[1]) + bounds[1]
    push!(X, x_init)
    push!(Y, objective(x_init))
    # Main loop
    for i in 1:num_iterations
        # Fit a model to the observed data (Gaussian Process in this case)
        # For simplicity, let's just use the current best observed value
        x_next = rand() * (bounds[2] - bounds[1]) + bounds[1] # Random sampling
        # Evaluate the objective function at the chosen point
        y_next = objective(x_next)
        # Update the data with the new observation
        push!(X, x_next)
        push!(Y, y_next)
        # Here, we will just print the current best observed value
        println("Iteration $i: Best value = $(maximum(Y))")
    end
    # Return the best observed value and corresponding parameter
    best_idx = argmax(Y)
    return X[best_idx], Y[best_idx]
end

best_x, best_value = bayesian_optimization(objective, (-5.0, 5.0), 10)
println("Best x found: $best_x, Best value: $best_value")

Iteration 1: Best value = -0.3041807254074535
Iteration 2: Best value = -0.3041807254074535
Iteration 3: Best value = -0.3041807254074535
Iteration 4: Best value = -0.3041807254074535
```

```

Iteration 5: Best value = -0.3041807254074535
Iteration 6: Best value = -0.3041807254074535
Iteration 7: Best value = -0.3041807254074535
Iteration 8: Best value = 0.02504980758369785
Iteration 9: Best value = 0.02504980758369785
Iteration 10: Best value = 0.02504980758369785
Best x found: -0.057501331095793695, Best value: 0.02504980758369785

```

18.4.5. BFGS

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is a popular iterative optimization algorithm used for unconstrained optimization problems. It belongs to the family of quasi-Newton methods, which are designed to find the local minimum of a differentiable objective function without needing its Hessian matrix directly. Instead, BFGS iteratively constructs an approximation to the inverse Hessian matrix using gradients of the objective function.

```

# Define the objective function to minimize
function objective_function(x)
    return sum(x .* x)
end

# Initial guess for the minimization
initial_x = [1.0]
# Perform optimization using BFGS method
result = optimize(objective_function, initial_x, BFGS())
# Extract the optimized solution
solution = result.minimizer
minimum_value = result.minimum

# Print the result
println("Optimized solution: x = ", solution)
println("Minimum value found: ", minimum_value)

```

```

Optimized solution: x = [-6.359357485052897e-13]
Minimum value found: 4.0441427622698303e-25

```

18.5. Model fitting

18.5.1. Root finding

Root finding, also known as root approximation or root isolation, is the process of finding the values of the independent variable (usually denoted as x) for which a given function equals zero. In mathematical terms, if we have a function $f(x)$, root finding involves finding values of x such that $f(x) = 0$.

There are various algorithms for root finding, each with its own advantages and disadvantages depending on the characteristics of the function and the requirements of the problem. One notable approach is Newton's method, an iterative method that uses the derivative of the function to approximate the root with increasing accuracy in each iteration.

```
# Define a differentiable function
f(x) = 3x^2 + 2x + 1
# Define an initial value
x = 1.0
# tolerance of difference in value
tol = 1e-6
# maximum number of iteration of the algorithm
max_iter = 100
iter = 0
while abs(f(x)) > tol && iter < max_iter
    x -= f(x) / gradient(x → f(x), x)[1]
    iter += 1
end
if iter == max_iter
    @show "Warning: Maximum number of iterations reached."
else
    @show "Root found after", iter, " iterations."
end
@show "Approximate root: ", x

"Warning: Maximum number of iterations reached." = "Warning: Maximum number of iterations reached."
("Approximate root: ", x) = ("Approximate root: ", -1.391591884376212)

("Approximate root: ", -1.391591884376212)
```

18.5.2. Bracketed search algorithm

A bracketed search algorithm is a technique used in optimization and numerical methods to confine or “bracket” a minimum or maximum of a function within a specified interval. The primary goal is to reduce the search space systematically until a satisfactory solution or range containing the optimal value is found.

```
function bisection_method(f, a, b; tol=1e-6, max_iter=100)
    """
    Bisection method to find a root of the function f(x) within the interval [a, b].
    Parameters:
    - f: Function to find the root of.
    - a, b: Initial interval [a, b] where the root is expected to be.
    - tol: Tolerance for the root (default is 1e-6).
    - max_iter: Maximum number of iterations allowed (default is 100).

    Returns:
    - root: Approximate root found within the tolerance.
    - iterations: Number of iterations taken to converge.
    """
    fa = f(a)
    fb = f(b)
    if fa * fb > 0
        error("The function values at the endpoints must have opposite signs.")
    end
    iterations = 0
    while (b - a) / 2 > tol && iterations < max_iter
        c = (a + b) / 2
        fc = f(c)
        if fc == 0
            return c, iterations
        end
        if fa * fc < 0
            b = c
            fb = fc
        else
            a = c
            fa = fc
        end
        iterations += 1
    end
    root = (a + b) / 2
```

18. Optimization

```
    return root, iterations
end

# Define the function we want to find the root of
function f(x)
    return x^3 - 6x^2 + 11x - 6.1
end

# Initial interval [a, b] and tolerance
a = 0.5
b = 10
tolerance = 1e-6
# Apply the bisection method
root, iterations = bisection_method(f, a, b, tol=tolerance)

# Print results
println("Approximate root: ", root)
println("Iterations taken: ", iterations)
println("Function value at root: ", f(root))
```

```
Approximate root: 3.046680122613907
Iterations taken: 23
Function value at root: -9.356632642010254e-7
```

18.5.3. Best fitting curve

In model fitting, the “best fitting curve” refers to the curve or function that best describes the relationship between the independent and dependent variables in the data. The goal of model fitting is to find the parameters of the chosen curve or function that minimize the difference between the observed data points and the values predicted by the model.

The process of finding the best fitting curve typically involves:

- Choosing a model: Based on the nature of the data and the underlying relationship between the variables, a suitable model or family of models are selected.
- Estimating parameters: Using the chosen model, one estimates the parameters that best describe the relationship between the variables. This is often done using optimization techniques such as least squares regression, maximum likelihood estimation, or Bayesian inference.

- Evaluating the fit: Once the parameters are estimated, one evaluates the goodness of fit of the model by comparing the predicted values to the observed data. Common metrics for evaluating fit, or error functions, include the residual sum of squares, the coefficient of determination (R-squared), and visual inspection of the residuals.
- Iterating if necessary: If the fit is not satisfactory, one may need to iterate on the model or consider alternative models until you find a satisfactory fit to the data.

```

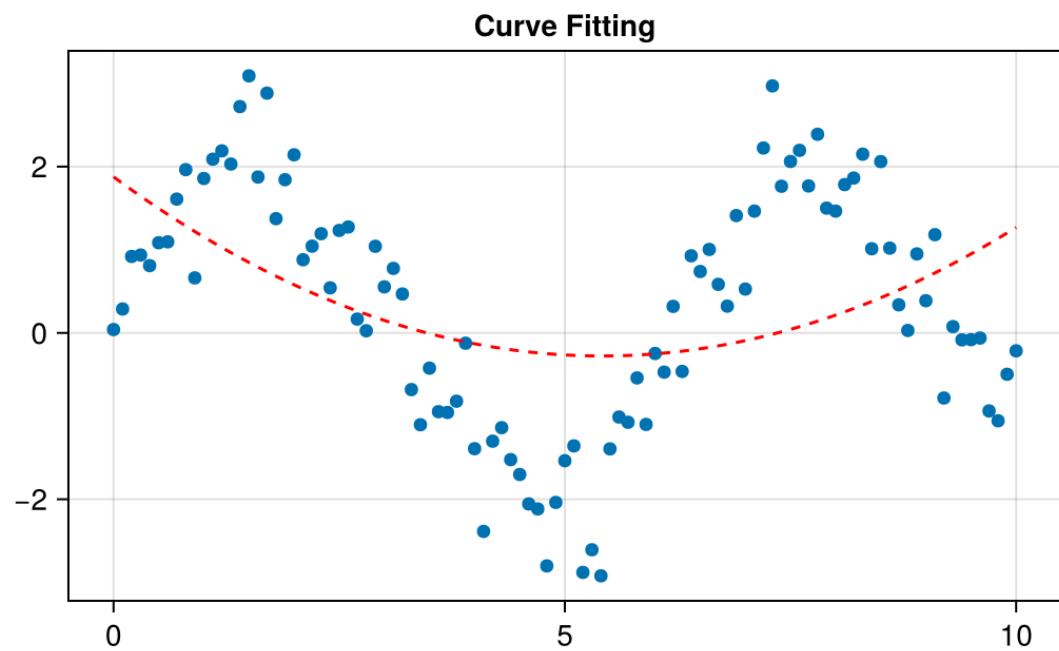
x_data = 0:0.1:10
y_data = 2 .* sin.(x_data) .+ 0.5 .* randn(length(x_data))
# Define the model function
curve_model(x, p) = p[1] * x .^ 2 + p[2] * x .+ p[3]
# Initial parameter guess
p₀ = [1.0, 1.0, 1.0]
# Fit the model to the data
fit_result = curve_fit(curve_model, x_data, y_data, p₀)
# Extract the fitted parameters
params = coef(fit_result)
# Evaluate the model with the fitted parameters
y_fit = curve_model(x_data, params)
# Plot the data and the fitted curve
fig = Figure()
Axis(fig[1, 1], title="Curve Fitting")
scatter!(x_data, y_data, label="Data")
lines!(x_data, y_fit, label="Fitted Curve", linestyle=:dash, color=:red)
fig

```

Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S'

└ @ Makie ~/julia/packages/Makie/GtFuI/src/scenes.jl:227

18. Optimization



19. Sensitivity Analysis

[Drafting note: based on some examples. Needs to be revised with more exposition.]

19.1. In This Chapter

Different approaches to understanding the sensitivity of a model to changes in its inputs: derivatives, finite differences, global sensitivity analysis approaches, and statistical approaches.

19.2. Setup

```
using CSV, DataFrames
using MortalityTables, Dates
using GlobalSensitivity
using QuasiMonteCarlo
using CairoMakie

@enum Sex Female = 1 Male = 2
@enum Risk Standard = 1 Preferred = 2

mutable struct Policy
    id::Int
    sex::Sex
    benefit_base::Float64
    COLA::Float64
    mode::Int
    prem::Float64
    pp::Int
    issue_date::Date
    issue_age::Int
    risk::Risk
end
```

19. Sensitivity Analysis

19.3. The Data

```
sample_csv_data =
    IOBuffer(
        raw"id,sex,benefit_base,COLA,mode,prem,pp,issue_date,issue_age,risk
        1,M,100000.0,0.03,1,1000.0,3,1999-12-05,30,Std"
    )

mort = Dict(
    Male => MortalityTables.table(988).ultimate,
    Female => MortalityTables.table(992).ultimate,
)

Dict{Sex, OffsetArrays.OffsetVector{Float64, Vector{Float64}}} with 2 entries:
Male => [0.022571, 0.022571, 0.022571, 0.022571, 0.022571, 0.022571, 0.0225...
Female => [0.00745, 0.00745, 0.00745, 0.00745, 0.00745, 0.00745, 0.00745, 0.0...

policies = let

    # read CSV directly into a dataframe
    # df = CSV.read("sample_inforce.csv",DataFrame) # use local string for notebook
    df = CSV.read(sample_csv_data, DataFrame)

    # map over each row and construct an array of Policy objects
    map(eachrow(df)) do row
        Policy(
            row.id,
            row.sex == "M" ? Male : Female,
            row.benefit_base,
            row.COLA,
            row.mode,
            row.prem,
            row.pp,
            row.issue_date,
            row.issue_age,
            row.risk == "Std" ? Standard : Preferred,
        )
    end
end

1-element Vector{Policy}:
Policy(1, Male, 100000.0, 0.03, 1, 1000.0, 3, Date("1999-12-05"), 30, Standard)
```

19.3. The Data

Given a basic insurance product, a pure whole of life (WOL) policy with level benefits and level premiums payable within the first 10 years, the reserve at the end of the y^{th} policy year is defined by

$$res(y) = \sum_{t=age+y}^{120} (sur_{t-age-y} * mort_t * B_y * \sqrt{1+r}) - (P_y * sur_{t-age-y})$$

where

- $mort_t$ is the mortality at age t
- p_y is the survival probability adjusted with COLA, with values of
 - $p_{y-1} = 1$,
 - $p_x = p_{x-1} * (1 - mort_{age+y}) / (1 + COLA)$ for $x \geq y$, and
 - 0 for $x < y - 1$ or $age + x \geq 120$, or ultimate age of the current mortality table
- B_y is the level benefit throughout the policy
- P_y is the level premium within the first 10 policy years which is 0 for policy years after 10
- r is the level interest rate throughout the policy

```
function sur(y::Int, pol::Policy)
    if y == 0
        1
    elseif y < 0 || 120 - y <= pol.issue_age
        0
    else
        sur(y - 1, pol) * (1 - mort[pol.sex][pol.issue_age+y]) / (1 + pol.COLA)
    end
end

function res(y::Int, pol::Policy)
    s = 0.0
    if y >= 1 && y <= 120 - pol.issue_age
        for t in (pol.issue_age+y):120
            prem = 0.0
            if y <= pol.pp
                prem = pol.prem
            end
            s += sur(t - pol.issue_age - y, pol) * mort[pol.sex][t] * pol.benefit_base - prem
        end
    end
end
```

19. Sensitivity Analysis

```
s  
end  
  
res (generic function with 1 method)
```

19.4. Common Sensitivity Analysis Methodologies

19.4.1. Finite Differences

Define a customized finite difference function with respect to the COLA, rippled by a small difference.

```
function res_wrt_r_fd(y::Int, pol::Policy, r::Float64, h=1e-3)  
    p_+, p_- = deepcopy(pol), deepcopy(pol)  
    p_+.COLA, p_-.COLA = r + h, r - h  
    (res(y, p_-) - res(y, p_+)) / (2res(y, pol))  
end  
  
res_wrt_r_fd(2, policies[1], 0.03) # changes in reserve at year 2 when the interest rate at 3%  
0.021366520936389077
```

19.4.2. Regression Analyses

```
function r1_wrt_r(r)  
    p = deepcopy(policies[1])  
    p.COLA = r[2]  
    p.prem = r[3]  
    res(Int(floor(r[1])), p)  
end  
  
# reserve @ year 1/2, interest rate @ 0.03 ± 0.01, prem @ 1000.0 ± 0.1  
reg_anal = gsa(r1_wrt_r, RegressionGSA(), [[1, 2], [0.029, 0.031], [999.9, 1000.1]], samples=1000)  
@show reg_anal.pearson  
  
reg_anal.pearson = [0.0022668260330361793 -0.999960323782643 -0.0032559466968495794]  
  
1x3 Matrix{Float64}:  
0.00226683 -0.99996 -0.00325595
```

The Pearson Spearman coefficients show the correlation coefficient matrix between inputs and outputs.

19.4.3. Sobol Indices

Sobol is a variance-based method, and it decomposes the variance of the output of the model or system into fractions which can be attributed to inputs or sets of inputs. This helps to get not just the individual parameter's sensitivities, but also gives a way to quantify the affect and sensitivity from the interaction between the parameters.

$$Y = f_0 + \sum_{i=1}^d f_i(X_i) + \sum_{i < j}^d f_{ij}(X_i, X_j) + \dots + f_{1,2,\dots,d}(X_1, X_2, \dots, X_d)$$

$$Var(Y) = \sum_{i=1}^d V_i + \sum_{i < j}^d V_{ij} + \dots + V_{1,2,\dots,d}$$

The Sobol Indices are “ordered”, the first order indices given by $S_i = \frac{V_i}{Var(Y)}$, the contribution to the output variance of the main effect of X_i . Therefore, it measures the effect of varying X_i alone, but averaged over variations in other input parameters. It is standardized by the total variance to provide a fractional contribution. Higher-order interaction indices S_{ij} , S_{ijk} and so on can be formed by dividing other terms in the variance decomposition by $Var(Y)$.

```
# reserve @ year 1/2, interest rate @ 0.03 ± 0.01, prem @ 1000.0 ± 0.1
L, U = QuasiMonteCarlo.generate_design_matrices(1000, [1, 0.029, 999.9], [2, 0.031, 1000.1], S)
s = gsa(r1_wrt_r, Sobol(), L, U)
@show s.S1
@show s.ST
```

```
[ Warning: The `generate_design_matrices(n, d, sampler, R = NoRand(), num_mats)` method does not p
  Prefer using randomization methods such as `R = Shift()`, `R = MatousekScrambling()`, etc., se
  @ QuasiMonteCarlo ~/.julia/packages/QuasiMonteCarlo/KvLfb/src/RandomizedQuasiMonteCarlo/iterat
```

```
s.S1 = [0.0, 1.2308143537488831, -0.00010730838653271867]
s.ST = [0.0, 1.0014202549551279, 5.831510556439337e-6]
```

3-element Vector{Float64}:

```
0.0
1.0014202549551279
5.831510556439337e-6
```

The output shows the first order and total order of variations in different input parameters.

19. Sensitivity Analysis

19.4.4. Morris Method

The Morris method also known as Morris's OAT method where OAT stands for One At a Time can be described in the following steps:

$$EE_i = \frac{f(x_1, x_2, \dots x_i + \Delta, \dots x_k) - y}{\Delta}$$

We calculate local sensitivity measures known as “elementary effects”, which are calculated by measuring the perturbation in the output of the model on changing one parameter.

These are evaluated at various points in the input chosen such that a wide “spread” of the parameter space is explored and considered in the analysis, to provide an approximate global importance measure. The mean and variance of these elementary effects is computed. A high value of the mean implies that a parameter is important, a high variance implies that its effects are non-linear or the result of interactions with other inputs. This method does not evaluate separately the contribution from the interaction and the contribution of the parameters individually and gives the effects for each parameter which takes into consideration all the interactions and its individual contribution.

```
# reserve @ year 1/2, interest rate @ 0.03 ± 0.01, prem @ 1000.0 ± 0.1
m = gsa(r1_wrt_r, Morris(), [[1, 2], [0.029, 0.031], [999.9, 1000.1]])
@show m.means
@show m.variances

m.means = [0.0 -714326.4358633403 -17.23601505988768]
m.variances = [0.0 1.9057971936038154e8 0.012468630640914313]

1x3 Matrix{Float64}:
 0.0  1.9058e8  0.0124686
```

From the means it can be observed which variables are more important, and the variances imply higher degree of nonlinearity or interactions with other variables.

19.4.5. Fourier Amplitude Sensitivity Tests

FAST offers a robust, especially at low sample size, and computationally efficient procedure to get the first and total order indices as discussed in Sobol. It utilizes monodimensional Fourier decomposition along a curve, exploring the parameter space. The curve is defined by a set of parametric equations,

19.4. Common Sensitivity Analysis Methodologies

$$EE_i = \frac{f(x_1, x_2, \dots x_i + \Delta, \dots x_k) - y}{\Delta}$$

where s is a scalar variable varying over the range $-\infty < s < +\infty$, G_i are transformation functions and $w_i, \forall i = 1, 2, \dots, N$ is a set of different (angular) frequencies, to be properly selected, associated with each factor for all N (samples) number of parameter sets.

```
# reserve @ year 1/2, interest rate @ 0.03 ± 0.01, prem @ 1000.0 ± 0.1
fast = gsa(r1_wrt_r, eFAST(), [[1, 2], [0.029, 0.031], [999.9, 1000.1]], samples=1000)
@show fast.S1
@show fast.ST

fast.S1 = [5.045003000409528e-12 0.9976898154045825 5.7959545686720465e-6]
fast.ST = [6.75062992394615e-7 0.9999937118855862 0.002300936414847965]

1x3 Matrix{Float64}:
 6.75063e-7  0.999994  0.00230094
```

The output shows the first order and total order of variations in different input parameters.

19.4.6. Automatic Differentiation

By applying the chain rule repeatedly on elementary operations of computer calculations, automatic differentiation can be applied to measure impacts of small differences. More details in the chapter on automatic differentiation.

19.4.7. Scenario Analyses

Scenarios can be generated following scenario generation methodologies to evaluate impacts. More details in the chapter on scenario generation.

When scenarios are generated to evaluate sensitivities, one may need to take the following into consideration.

- Reverse stress testing. Reverse stress testing in scenario analysis involves identifying extreme scenarios that could potentially lead to catastrophic outcomes for a financial institution or a system. Unlike traditional sensitivity testing to simulate the impact of adverse events on the system, reverse stress testing starts with a catastrophic outcome and works backwards to determine the combination of events or circumstances that could lead to such an outcome.

19. Sensitivity Analysis

- Stylistic scenarios. Developing stylistic scenarios in scenario analysis involves creating narratives or storylines that describe plausible future states or situations. These scenarios are crafted to capture key uncertainties, trends, and factors that could significantly impact the organization, industry, or environment under study.
- Backtesting against historical data. Backtesting in scenario analysis involves an iterative process of using past data to validate the effectiveness and accuracy of scenarios developed for forecasting future outcomes. Scenarios are first defined and applied on selective historical data, and refined after any discrepancies of scenario outcomes versus historical results are identified.

19.5. Benchmarking

20. Stochastic Modeling

The Monte Carlo Method: (i) A last resort when doing numerical integration, and (ii) a way of wastefully using computer time. - Malvin H. Kalos¹ (c. 1960)

¹Kalos was a pioneer in Monte Carlo techniques, quoted via https://doi.org/10.1007/978-3-540-74686-7_3

21. Visualizations

21.1. In This Chapter

The evolved brain and pattern recognition, recommended principles for looking at data, and avoiding common mistakes. Exploratory visualization versus visualizations intended for an audience.

22. Matrices and Their Uses

22.1. In This Chapter

Matrices and their myriad uses: reframing problems through the eyes of linear algebra, an intuitive refreshing on applicable maths, and recurring patterns of matrix operations in financial modeling.

22.2. Setup

```
using LinearAlgebra
using Recommendation
using SparseArrays
using MLDataUtils
using Statistics
using MultivariateStats
```

WARNING: using `Recommendation.isdefined` in module `Main` conflicts with an existing identifier.

22.3. Matrix manipulation

22.3.1. Multiplication

```
# Define two matrices
A = [1 2 3;
      4 5 6;
      7 8 9]
B = [9 8 7;
      6 5 4;
      3 2 1]
# Perform matrix multiplication
C = A * B
# Display the result
```

22. Matrices and Their Uses

```
println("Result of matrix multiplication:")
println(C)
```

```
Result of matrix multiplication:
[30 24 18; 84 69 54; 138 114 90]
```

22.3.2. Inversion

```
# Define a matrix
A = [1 2; 3 4]
# Compute the inverse of the matrix
A_inv = inv(A)
# Display the result
println("Inverse of matrix A:")
println(A_inv)
```

```
Inverse of matrix A:
[-1.999999999999996 0.999999999999998; 1.499999999999998 -0.499999999999999]
```

22.4. Matrix decomposition

22.4.1. Eigenvalues

Eigenvalue decomposition, also known as eigendecomposition, is a matrix factorization that decomposes a matrix into its eigenvectors and eigenvalues.

```
# Create a square matrix
A = [1 2 3;
      4 5 6;
      7 8 9]
# Perform eigenvalue decomposition
eigen_A = eigen(A)
# Extract eigenvalues and eigenvectors
λ = eigen_A.values
V = eigen_A.vectors
# Display the results
println("Original Matrix:")
println(A)
println("\nEigenvalues:")
```

```

println(λ)
println("\nEigenvectors:")
println(V)

Original Matrix:
[1 2 3; 4 5 6; 7 8 9]

Eigenvalues:
[-1.1168439698070434, -8.582743335036247e-16, 16.11684396980703]

Eigenvectors:
[-0.7858302387420671 0.4082482904638635 -0.2319706872462857; -0.0867513392566285 -0.81649658092

```

22.4.2. Singular values

Singular value decomposition breaks a matrix into three matrices U , Σ , and V , representing the left singular vectors, the singular values (diagonal matrix), and the right singular vectors, respectively.

```

# Create a random matrix
A = rand(4, 3)
# Perform Singular Value Decomposition (SVD)
U, Σ, V = svd(A)
# U: Left singular vectors
# Σ: Singular values (diagonal matrix)
# V: Right singular vectors (transpose)
# Reconstruct original matrix
A_reconstructed = U * Diagonal(Σ) * V'

# Display the results
println("Original Matrix:")
println(A)
println("\nLeft Singular Vectors:")
println(U)
println("\nSingular Values:")
println(Σ)
println("\nRight Singular Vectors:")
println(V)
println("\nReconstructed Matrix:")
println(A_reconstructed)

```

Original Matrix:

22. Matrices and Their Uses

```
[0.9915848655668824 0.3345269182924432 0.16344567066449178; 0.5457492343389718 0.56209763388839  
Left Singular Vectors:  
[-0.41279523033041177 0.8751936757562201 -0.11184544257476875; -0.5635674444882998 -0.159823949  
Singular Values:  
[2.0109769173759795, 0.750778492629152, 0.2843779035524414]  
Right Singular Vectors:  
[-0.5423417277329626 0.813809991012953 -0.20875523678969615; -0.5742781886754388 -0.17772084859  
Reconstructed Matrix:  
[0.9915848655668821 0.3345269182924434 0.163445670664492; 0.5457492343389713 0.5620976338883944
```

22.4.3. Matrix factorization and fatorization machines

Matrix factorization is a popular technique in recommendation systems for modeling user-item interactions and making personalized recommendations. The core idea behind matrix factorization is to decompose the user-item interaction matrix into two lower-dimensional matrices, capturing latent factors that represent user preferences and item characteristics. By learning these latent factors, the recommendation system can make predictions for unseen user-item pairs.

Factorization Machines (FM) are a type of supervised machine learning model designed for tasks such as regression and classification, especially in the context of recommendation systems and predictive modeling with sparse data. FM models extend traditional linear models by incorporating interactions between features, allowing them to capture complex relationships within the data.

```
# Generate synthetic user-item interaction data
num_users = 100
num_items = 50
num_ratings = 500
user_ids = rand(1:num_users, num_ratings)
item_ids = rand(1:num_items, num_ratings)
ratings = rand(1:5, num_ratings)
# Create a sparse user-item matrix
user_item_matrix = sparse(user_ids, item_ids, ratings)
# Split data into training and testing sets
train_data, test_data = splitobs(user_item_matrix, 0.8)
# Set parameters for matrix factorization
num_factors = 10
num_iterations = 10
```

```

# Train matrix factorization model
data = DataAccessor(user_item_matrix)
recommender = MF(data) # FactorizationMachines(data) alternatively
fit!(recommender)
# Predict ratings for the test set
rec = Dict()
for user in 1:num_users
    rec[user] = recommend(recommender, user, num_items, collect(1:num_items))
end
# Evaluate model performance
predictions = []
for (i, j, v) in zip(findnz(test_data.data)[1], findnz(test_data.data)[2], findnz(test_data.da
    for p in rec[i]
        if p[1] == j
            push!(predictions, p[2])
            break
        end
    end
end
rmse = measure(RMSE(), predictions, nonzeros(test_data.data))
println("Root Mean Squared Error (RMSE): ", rmse)

```

Root Mean Squared Error (RMSE): 1.2871054077515227

22.4.4. Principal component analysis

Principal Component Analysis (PCA) is a widely used technique in various fields for dimensionality reduction, data visualization, feature extraction, and noise reduction. PCA can also be applied to detect anomalies or outliers in the data by identifying data points that deviate significantly from the normal patterns captured by the principal components. Anomalies may appear as data points with large reconstruction errors or as outliers in the low-dimensional space spanned by the principal components.

```

# Generate some synthetic data
data = randn(100, 5) # 100 samples, 5 features
# Perform PCA
pca_model = fit(PCA, data; maxoutdim=2) # Project to 2 principal components
# Transform the data
transformed_data = transform(pca_model, data)
# Access principal components and explained variance ratio
principal_components = pca_model.prinvars
explained_variance_ratio = pca_model.prinvars / sum(pca_model.prinvars)

```

22. Matrices and Their Uses

```
# Print results
println("Principal Components:")
println(principal_components)
println("Explained Variance Ratio:")
println(explained_variance_ratio)
```

```
Principal Components:
[34.00886849934543, 30.42344782187357]
Explained Variance Ratio:
[0.527823155228482, 0.47217684477151795]
```

23. Learning from Data

23.1. In this chapter

Using data to inform a model: fitting parameters, forecasting, and fundamental limitations on prediction.

23.2. Setup

```
using MLJ
using StatsBase
using Flux
```

23.3. Applications

23.3.1. Parameter fitting

Refer to the chapter on Optimization for more details.

23.3.2. Forecasting

23.3.3. Static and dynamic validation

Static validation typically involves splitting the dataset into training and testing sets, where the testing set is held out and not used during model training. The model is trained on the training set and then evaluated on the held-out testing set to assess its performance. This approach helps to measure how well the model generalizes to unseen data.

Dynamic validation, on the other hand, involves using a rolling or expanding window to train and test the model iteratively over time. In each iteration, the model is trained on past data and tested on future data, simulating how the model would perform in a

23. Learning from Data

real-world scenario where new data becomes available over time. This approach helps to assess the model's ability to adapt to changing patterns and trends in the data.

```
# Generate synthetic time series data
num_samples = 100
data = rand(num_samples)
X = [ones(num_samples) data]
y = 2data .+ 1 .+ 0.1 * randn(num_samples, 1) # dependent variable with noise
# Train the model on the training set
θ = X \ y
# Predictions
y_pred = θ[2] .* data .+ θ[1]
# Compute evaluation metrics
mse = mean((y_pred .- y) .^ 2)
mae = mean(abs.(y_pred .- y))

println("Static validation results:")
println("Mean Squared Error (MSE): ", mse)
println("Mean Absolute Error (MAE): ", mae)

# Dynamic validation to update model over time and evaluate
num_updates = 5
mse_dyn = Float64[]
mae_dyn = Float64[]
for i in 1:num_updates
    data = rand(num_samples)
    X = [ones(num_samples) data]
    y = 2data .+ 1 .+ 0.1 * randn(num_samples, 1) # dependent variable with noise
    # Train the model on the training set
    θ = X \ y
    # Predictions
    y_pred = θ[2] .* data .+ θ[1]
    # Compute evaluation metrics
    mse = mean((y_pred .- y) .^ 2)
    mae = mean(abs.(y_pred .- y))
    push!(mse_dyn, mse)
    push!(mae_dyn, mae)
end

println("Dynamic validation results:")
println("Mean Squared Error (MSE): ", mean(mse_dyn))
println("Mean Absolute Error (MAE): ", mean(mae_dyn))
```

Static validation results:

```

Mean Squared Error (MSE): 0.009277783481528551
Mean Absolute Error (MAE): 0.07406233393494578
Dynamic validation results:
Mean Squared Error (MSE): 0.00970644911256275
Mean Absolute Error (MAE): 0.07768355131631507

```

23.3.4. Implied rate analysis

Implied rates are rates that are derived from the prices of financial instruments, such as bonds or options. For example, in the context of bonds, the implied rate is the interest rate that equates the present value of future cash flows from the bond (coupons and principal) to its current market price.

```

# Define the bond cash flows and prices
cash_flows = [100, 100, 100, 100, 1000] # Coupons and principal
prices = [95, 96, 97, 98, 1050]           # Market prices
# Define a function to calculate the present value of cash flows given a rate
function present_value(rate, cash_flows)
    pv = 0
    for (i, cf) in enumerate(cash_flows)
        pv += cf / (1 + rate)^i
    end
    return pv
end
# Define a function to calculate the implied rate using bisection method
function implied_rate(cash_flows, price)
    f(rate) = present_value(rate, cash_flows) - price
    return rootassign(f, 0.0, 1.0)
end
function rootassign(f, l, u)
    # Define an initial value
    x = 1.0
    # tolerance of difference in value
    tol = 1e-6
    # maximum number of iteration of the algorithm
    max_iter = 100
    iter = 0
    while abs(f(x)) > tol && iter < max_iter
        x -= f(x) / gradient(x → f(x), x)[1]
        iter += 1
    end
    if iter < max_iter && l < x < u
        return x
    end
end

```

23. Learning from Data

```
    else
        return -1.0
    end
end
# Calculate implied rates for each bond
implied_rates = [implied_rate(cash_flows, price) for price in prices]
# Print the results
for (i, rate) in enumerate(implied_rates)
    println("Implied rate for bond $i: $rate")
end

Implied rate for bond 1: -1.0
Implied rate for bond 2: -1.0
Implied rate for bond 3: -1.0
Implied rate for bond 4: -1.0
Implied rate for bond 5: -1.0
```

Part VI.

Applications in Practice

24. Stochastic Mortality Projections

[Drafting note: taken from a tutorial on JuliaActuary.org. Needs to be revised with more exposition.]

24.1. In This Chapter

A term life insurance policy is used to illustrate: selecting key model features, design tradeoffs between a few different approaches, and a discussion of the performance impacts of the different approaches to parallelism.

24.2. Setup

```
using CSV, DataFrames
using MortalityTables, ActuaryUtilities
using Dates
using ThreadsX
using BenchmarkTools
using Random
using CairoMakie
```

Define a datatype. Not strictly necessary, but will make extending the program with more functions easier.

Type annotations are optional, but providing them is able to coerce the values to be all plain bits (i.e. simple, non-referenced values like arrays are) when the type is constructed. This makes the whole data be stored in the stack and is an example of data-oriented design. It's much slower without the type annotations (~0.5 million policies per second, ~50x slower).

```
@enum Sex Female = 1 Male = 2
@enum Risk Standard = 1 Preferred = 2
```

24. Stochastic Mortality Projections

```
struct Policy
    id::Int
    sex::Sex
    benefit_base::Float64
    COLA::Float64
    mode::Int
    issue_date::Date
    issue_age::Int
    risk::Risk
end
```

24.3. The Data

```
sample_csv_data =
    IOBuffer(
        raw"id,sex,benefit_base,COLA,mode,issue_date,issue_age,risk
        1,M,100000.0,0.03,12,1999-12-05,30,Std
        2,F,200000.0,0.03,12,1999-12-05,30,Pref"
    )

IOBuffer(data=UInt8[...], readable=true, writable=false, seekable=true, append=false, size=152,
policies = let

    # read CSV directly into a dataframe
    # df = CSV.read("sample_inforce.csv",DataFrame) # use local string for notebook
    df = CSV.read(sample_csv_data, DataFrame)

    # map over each row and construct an array of Policy objects
    map(eachrow(df)) do row
        Policy(
            row.id,
            row.sex == "M" ? Male : Female,
            row.benefit_base,
            row.COLA,
            row.mode,
            row.issue_date,
            row.issue_age,
            row.risk == "Std" ? Standard : Preferred,
        )
    end
end
```

```

end

2-element Vector{Policy}:
Policy(1, Male, 100000.0, 0.03, 12, Date("1999-12-05"), 30, Standard)
Policy(2, Female, 200000.0, 0.03, 12, Date("1999-12-05"), 30, Preferred)

```

Define what mortality gets used:

```

mort = Dict(
    Male => MortalityTables.table(988).ultimate,
    Female => MortalityTables.table(992).ultimate,
)

function mortality(pol::Policy, params)
    return params.mortality[pol.sex]
end

mortality (generic function with 1 method)

```

This defines the core logic of the policy projection and will write the results to the given out container (here, a named tuple of arrays).

This is using a threaded approach where it could be operating on any of the computer's available threads, thus achieving thread-based parallelism - as opposed to multi-processor (multi-machine) or GPU-based computation, which requires formulating the problem a bit differently (array/matrix based). For the scale of computation here, I think I'd apply this model of parallelism.

```

function pol_project!(out, policy, params)
    # some starting values for the given policy
    dur = duration(policy.issue_date, params.val_date)
    start_age = policy.issue_age + dur - 1
    COLA_factor = (1 + policy.COLA)
    cur_benefit = policy.benefit_base * COLA_factor^(dur - 1)

    # get the right mortality vector
    qs = mortality(policy, params)

    # grab the current thread's id to write to results container without conflicting with other
    tid = Threads.threadid()

    ω = lastindex(qs)

```

24. Stochastic Mortality Projections

```
# inbounds turns off bounds-checking, which makes hot loops faster but first write loop wi
@inbounds for t in 1:min(params.proj_length, w - start_age)

    q = qs[start_age+t] # get current mortality

    if (rand() < q)
        return # if dead then just return and don't increment the results anymore
    else
        # pay benefit, add a life to the output count, and increment the benefit for next
        out.benefits[t, tid] += cur_benefit
        out.lives[t, tid] += 1
        cur_benefit *= COLA_factor
    end
end
end

pol_project! (generic function with 1 method)
```

Parameters for our projection:

```
params = (
    val_date=Date(2021, 12, 31),
    proj_length=100,
    mortality=mort,
)
```

```
(val_date = Date("2021-12-31"), proj_length = 100, mortality = Dict{Sex, OffsetArrays.OffsetVector})
```

Check the number of threads we're using:

```
Threads.nthreads()
```

4

```
function project(policies, params)
    threads = Threads.nthreads()
    benefits = zeros(params.proj_length, threads)
    lives = zeros(Int, params.proj_length, threads)
    out = (; benefits, lives)
    ThreadsX.foreach(policies) do pol
        pol_project!(out, pol, params)
```

```

end
map(x → vec(reduce(+, x, dims=2)), out)
end

project (generic function with 1 method)
```

24.4. Running the projection

Example of a single projection:

```
project(repeat(policies, 100_000), params)
```

```
(benefits = [5.629243560753166e10, 5.67483091822927e10, 5.7125986538619194e10, 5.743023483251459e10])
```

24.4.1. Stochastic Projection

Loop through and calculate the results n times (this is only running the two policies in the sample data "n times").

```

function stochastic_proj(policies, params, n)

    ThreadsX.map(1:n) do i
        project(policies, params)
    end
end

stochastic_proj (generic function with 1 method)

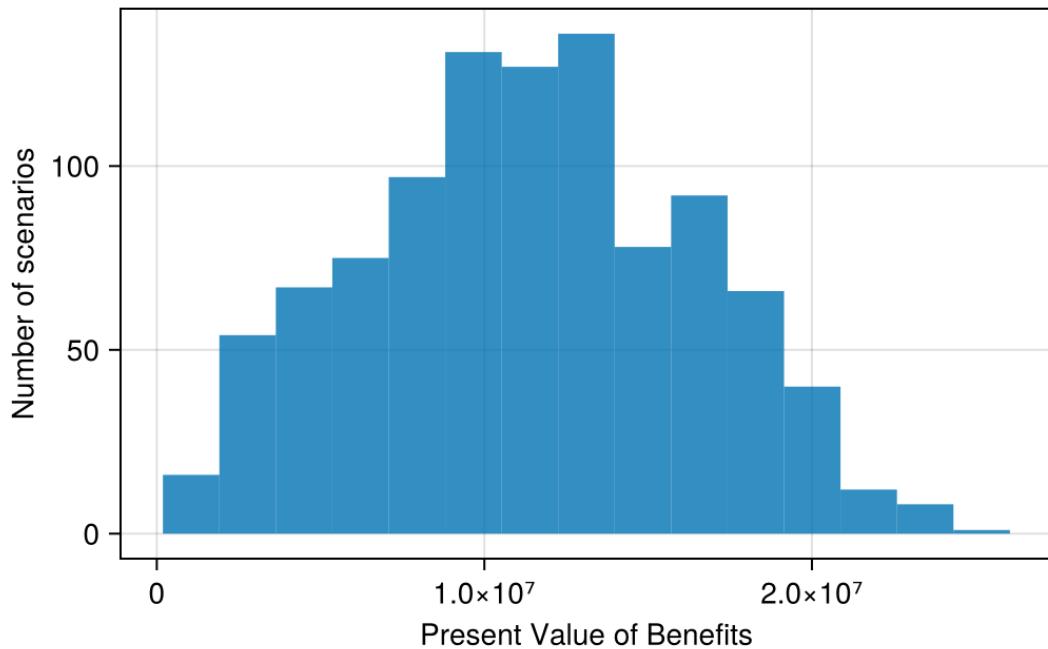
stoch = stochastic_proj(policies, params, 1000)

1000-element Vector{@NamedTuple{benefits::Vector{Float64}, lives::Vector{Int64}}}:
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977])
(benefits = [574831.0226582347, 394717.3022253211, 406558.82129208074, 418755.5859308432, 431313])
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977])
(benefits = [383220.68177215644, 394717.3022253211, 406558.82129208074, 418755.5859308432, 431313])
(benefits = [574831.0226582347, 394717.3022253211, 406558.82129208074, 418755.5859308432, 431313])
```

24. Stochastic Mortality Projections

```
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
:
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 394717.3022253211, 406558.82129208074, 418755.5859308432, 43131
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 203279.41064604037, 209377.7929654216, 21565
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 209377.7929654216, 215659
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 394717.3022253211, 406558.82129208074, 418755.5859308432, 43131
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
(benefits = [574831.0226582347, 592075.9533379817, 609838.2319381211, 628133.3788962648, 646977
let
    v = [pv(0.03, s.benefits) for s in stoch]
    hist(v,
        bins=15,
        axis=(
            xlabel="Present Value of Benefits",
            ylabel="Number of scenarios"
        )
    )
end
```

```
┌ Warning: Found `resolution` in the theme when creating a `Scene`. The `resolution` keyword for `S
└ @ Makie ~/julia/packages/Makie/GtFuI/src/scenes.jl:227
```



24.5. Benchmarking

Using a 2022 Macbook Air M2 laptop, about 30 million policies able to be stochastically projected per second:

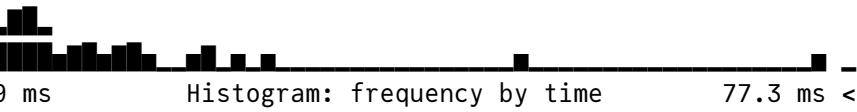
```

policies_to_benchmark = 3_000_000
# adjust the `repeat` depending on how many policies are already in the array
# to match the target number for the benchmark
n = policies_to_benchmark ÷ length(policies)

@benchmark project(p, r) setup = (p = repeat($policies, $n); r = $params)
  
```

24. Stochastic Mortality Projections

```
markTools.Trial: 59 samples with 1 evaluation.  
  e (min ... max): 65.950 ms ... 89.459 ms | GC (min ... max): 0.00% ... 0.00%  
  (median):       66.981 ms                  | GC (median):      0.00%  
  (mean ± σ):    67.841 ms ± 3.386 ms | GC (mean ± σ):  0.00% ± 0.00%
```



```
...y estimate: 29.20 KiB, allocs estimate: 222.
```

24.6. Further Optimization

In no particular order:

- the RNG could be made faster: <https://bkamins.github.io/julialang/2020/11/20/rand.html>
- Could make the stochastic set distributed, but at the current speed the overhead of distributed computing is probably more time than it would save. Same thing with GPU projections
- ...

25. Scenario Generation

[Drafting note: based on some examples. Needs to be revised with more exposition.]

25.1. In This Chapter

How to generate synthetic data for your model using sub-models, with applications to economic scenario generation and portfolio composition.

25.2. Setup

```
using CSV, DataFrames
using Random
using StatsBase, Distributions
using CairoMakie
```

25.3. The Data

25.4. Pseudo Random Number Generators

Modern computers utilize Pseudo random number generators (PRNGs) to generate random-like numbers. PRNGs are algorithms used to generate sequences of numbers that appear to be random but are actually determined by an initial value, known as the seed. These generators are called “pseudo-random” because the sequences they produce are deterministic; if you provide the same seed, you’ll get the same sequence of numbers. In addition, they have a finite period, which means that after a certain number of generated values, the sequence will repeat. It’s important to choose or design PRNGs with a long enough period for practical applications.

25. Scenario Generation

25.4.1. Common PRNGs

25.4.1.1. Mersenne Twister

One of the strengths of the Mersenne Twister is its exceptionally long period. The period is $2^{19937} - 1$, which means it can generate $2^{19937} - 1$ pseudo random numbers before repeating. This long period is crucial for applications requiring a large number of independent random numbers. It is also known for its good statistical properties. It passes many standard tests for randomness and provides a relatively uniform distribution of random numbers. Moreover, it is designed to allow multiple independent instances to be used concurrently without interfering with each other. This makes it suitable for parallel computing. Although there are faster generators for specific use cases, the Mersenne Twister is still often favored for its balance between speed and quality.

25.4.1.2. Xorshift

Xorshift is a family of PRNGs known for their simplicity and relatively fast operation. The name “xorshift” comes from the bitwise XOR (exclusive or) and bit-shifting operations that are the core of the algorithm. Xorshift generators are often used in applications where speed is a priority and cryptographic-strength randomness is not a strict requirement. Xorshift PRNGs use bitwise XOR, left shifts, and right shifts to update the internal state and generate pseudo-random numbers. The basic idea is to repeatedly apply these operations to the state to produce a sequence of numbers. The period of a typical xorshift generator is relatively short compared to some other PRNGs like the Mersenne Twister. However, there are variations of xorshift algorithms that can have longer periods. One of the main advantages of xorshift is its simplicity and speed. The bitwise XOR and bit-shifting operations can be efficiently implemented in hardware, making xorshift generators suitable for applications where fast random number generation is crucial.

25.4.1.3. Xoshiro

Xoshiro is a family of PRNGs known for their high performance and good statistical properties. The name “Xoshiro” is derived from the Japanese word “xoroshiro,” meaning “random.” Xoshiro algorithms, including Xoshiro128 and others, use a combination of bitwise XOR, bit-shifting, and addition operations. They often have more complex update rules than basic Xorshift algorithms. In addition, they typically have longer periods, making them suitable for applications that require more pseudo-random numbers before repetition.

25.5. Common Economic Scenario Generation Approaches

25.4.2. Consistent Interface

Julia offers a consistent interface for random numbers due to its design and multiple dispatch principles. Consider the following random numbers in different data types.

```
rng = MersenneTwister(1234)
rand(Int, (2, 3))
```

```
2×3 Matrix{Int64}:
 4031799989282868969   6026628416658068938   -418203448389577244
 -4230338824039096837  -7781844405943780478  -3913909816083977254
```

```
rng = MersenneTwister(1234)
rand(Float64, (2, 3))
```

```
2×3 Matrix{Float64}:
 0.965998  0.401312  0.946917
 0.364053  0.87962   0.663841
```

```
rng = Xoshiro(1234)
rand(Bool, (2, 3))
```

```
2×3 Matrix{Bool}:
 1  0  0
 0  0  0
```

25.5. Common Economic Scenario Generation Approaches

Economic scenario generation involves the development of plausible future economic scenarios to assess the potential impact on financial portfolios, investments, or decision-making processes. Various approaches are used to generate economic scenarios, including stochastic differential equations (SDEs) and Monte Carlo simulations.

25.5.1. Interest Rate Models

25.5.1.1. Vasicek and Cox Ingersoll Ross (CIR)

The Vasicek model is a one-factor model commonly used for simulating interest rate scenarios. It describes the dynamics of short-term interest rates using a stochastic differential equation (SDE). In a Monte Carlo simulation, we can use the Vasicek model to generate multiple interest rate paths. The CIR model is an extension of the Vasicek

25. Scenario Generation

model with non-constant volatility. It addresses the issue of negative interest rates by ensuring that interest rates remain positive. Vasicek is defined as

$$dr(t) = \kappa(\theta - r(t)) dt + \sigma dW(t)$$

where

- $r(t)$ is the short-term interest rate at time t .
- κ is the speed of mean reversion, representing how quickly the interest rate reverts to its long-term mean.
- θ is the long-term mean or equilibrium level of the interest rate.
- σ is the volatility of the interest rate.
- $dW(t)$ is a Wiener process or Brownian motion, representing a random shock.

And CIR is defined as

$$dr(t) = \kappa(\theta - r(t)) dt + \sigma\sqrt{r(t)} dW(t)$$

where

- $r(t)$ is the short-term interest rate at time t .
- κ is the speed of mean reversion, representing how quickly the interest rate reverts to its long-term mean.
- θ is the long-term mean or equilibrium level of the interest rate.
- σ is the volatility of the interest rate.
- $dW(t)$ is a Wiener process or Brownian motion, representing a random shock.

The following code shows a simplified implementation of a CIR model. The specification of dr can be changed to become a Vasicek model.

```
# Set seed for reproducibility
Random.seed!(1234)

# CIR model parameters
κ = 0.2      # Speed of mean reversion
θ = 0.05     # Long-term mean
σ = 0.1      # Volatility

# Initial short-term interest rate
r₀ = 0.03

# Number of time steps and simulations
num_steps = 252
num_simulations = 1_000
```

25.5. Common Economic Scenario Generation Approaches

```
# Time increment
Δt = 1 / 252

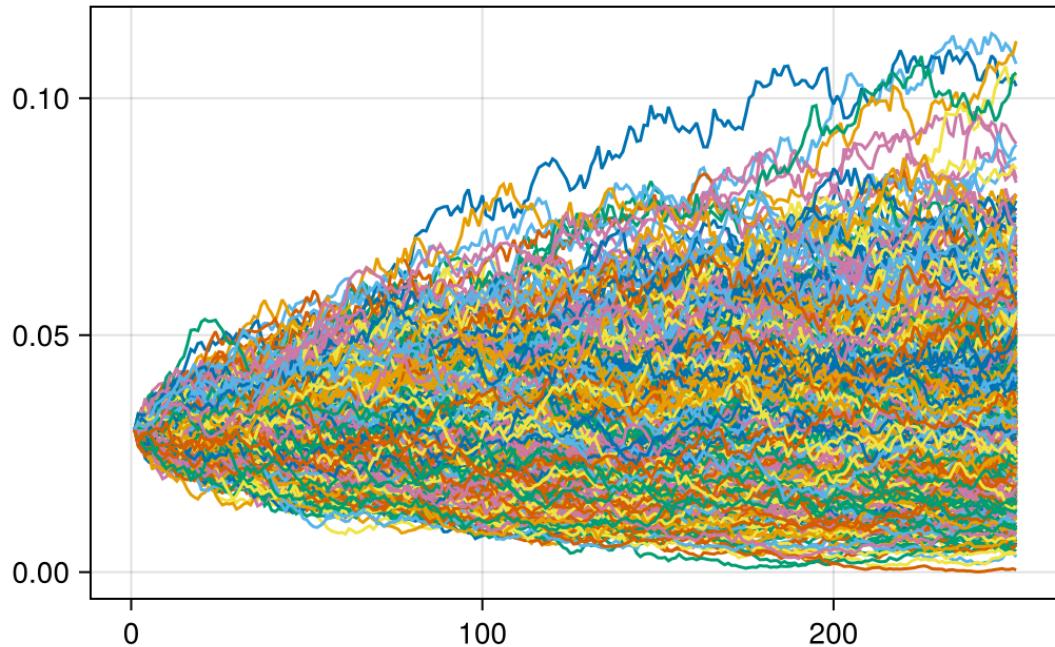
# Function to simulate CIR process
function cir_simulation(κ, θ, σ, r₀, Δt, num_steps, num_simulations)
    interest_rate_paths = zeros(num_steps, num_simulations)
    for j in 1:num_simulations
        interest_rate_paths[1, j] = r₀
        for i in 2:num_steps
            dW = randn() * sqrt(Δt)
            # for Vasicek
            # dr = κ * (θ - interest_rate_paths[i-1, j]) * Δt + σ * dW
            dr = κ * (θ - interest_rate_paths[i-1, j]) * Δt + σ * sqrt(interest_rate_paths[i-1, j])
            interest_rate_paths[i, j] = max(interest_rate_paths[i-1, j] + dr, 0) # Ensure non-negative
        end
    end
    return interest_rate_paths
end

# Run CIR simulation
cir_paths = cir_simulation(κ, θ, σ, r₀, Δt, num_steps, num_simulations)

# Plot the simulated interest rate paths
f = Figure()
Axis(f[1, 1])
for i in 1:num_simulations
    lines!(1:num_steps, cir_paths[:, i])
end
f
```

```
┌ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/julia/packages/Makie/GtFuI/src/scenes.jl:227
```

25. Scenario Generation



25.5.1.2. Hull White

The Hull-White model is a one-factor model that extends the Vasicek model by allowing the mean reversion and volatility parameters to be time-dependent. It is commonly used for pricing interest rate derivatives. Brace-Gatarek-Musiela (BGM) Model extends the Hull-White model to incorporate more factors. It is one of the Libor Market Model (LMM) that describes the evolution of forward rates. It allows for the modeling of both the short-rate and the entire yield curve. It is defined as

$$dr(t) = (\theta(t) - ar(t)) dt + \sigma(t) dW(t)$$

where

- $r(t)$ is the short-term interest rate at time t .
- θ is the long-term mean or equilibrium level of the interest rate.
- a is the speed of mean reversion.
- $\sigma(t)$ is the time-dependent volatility of the interest rate.
- $dW(t)$ is a Wiener process or Brownian motion, representing a random shock.

```
# Set seed for reproducibility
Random.seed!(1234)
```

25.5. Common Economic Scenario Generation Approaches

```
# Hull-White model parameters
α = 0.1          # Mean reversion speed
σ = 0.02         # Volatility
r₀ = 0.03        # Initial short-term interest rate

# Number of time steps and simulations
num_steps = 252
num_simulations = 1_000

# Time increment
Δt = 1 / 252

# Function to simulate Hull-White process
function hull_white_simulation(α, σ, r₀, Δt, num_steps, num_simulations)
    interest_rate_paths = zeros(num_steps, num_simulations)
    for j in 1:num_simulations
        interest_rate_paths[1, j] = r₀
        for i in 2:num_steps
            dW = randn() * sqrt(Δt)
            dr = α * (σ - interest_rate_paths[i-1, j]) * Δt + σ * dW
            interest_rate_paths[i, j] = interest_rate_paths[i-1, j] + dr
        end
    end
    return interest_rate_paths
end

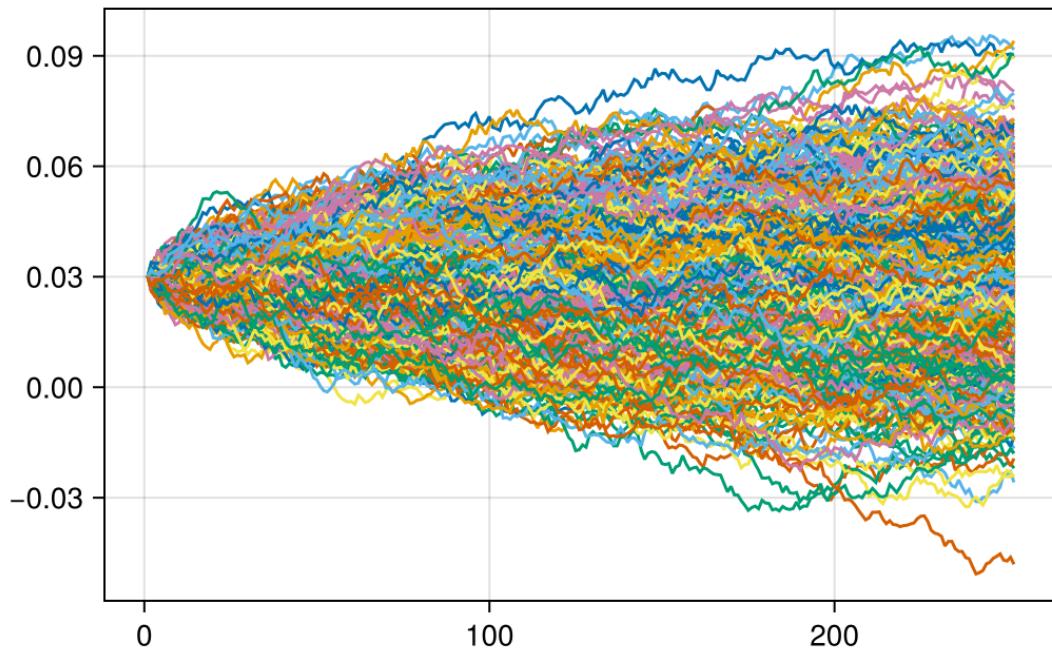
# Run Hull-White simulation
hull_white_paths = hull_white_simulation(α, σ, r₀, Δt, num_steps, num_simulations)

# Plot the simulated interest rate paths
f = Figure()
Axis(f[1, 1])
for i in 1:num_simulations
    lines!(1:num_steps, hull_white_paths[:, i])
end
f
```

Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'Scene'

└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227

25. Scenario Generation



25.5.2. Stock Models

25.5.2.1. Geometric Brownian Motion (GBM)

GBM is a stochastic process commonly used to model the price movement of financial instruments, including stocks. It assumes constant volatility and is characterized by a log-normal distribution. It is defined as

$$dS(t) = \mu S(t) dt + \sigma S(t) dW(t)$$

where

- $S(t)$ is the stock price at time t .
- μ is the drift coefficient (expected return).
- σ is the volatility coefficient.
- $dW(t)$ is a Wiener process or Brownian motion, representing a random shock.

```
# Set seed for reproducibility
Random.seed!(1234)

# GBM parameters
μ = 0.05      # Drift (expected return)
```

25.5. Common Economic Scenario Generation Approaches

```
σ = 0.2          # Volatility

# Initial stock price
S₀ = 100

# Number of time steps and simulations
num_steps = 252
num_simulations = 1_000

# Time increment
Δt = 1 / 252

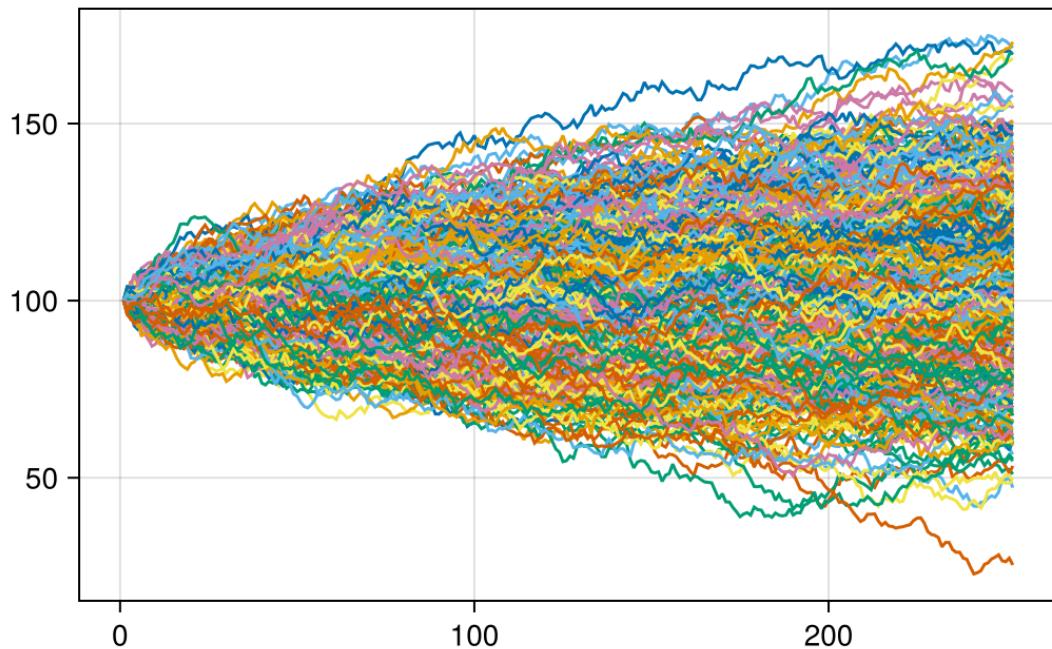
# Function to simulate GBM
function gbm_simulation(μ, σ, S₀, Δt, num_steps, num_simulations)
    stock_price_paths = zeros(num_steps, num_simulations)
    for j in 1:num_simulations
        stock_price_paths[1, j] = S₀
        for i in 2:num_steps
            dW = randn() * sqrt(Δt)
            dS = μ * S₀ * Δt + σ * S₀ * dW
            stock_price_paths[i, j] = stock_price_paths[i-1, j] + dS
        end
    end
    return stock_price_paths
end

# Run GBM simulation
gbm_paths = gbm_simulation(μ, σ, S₀, Δt, num_steps, num_simulations)

# Plot the simulated stock price paths
f = Figure()
Axis(f[1, 1])
for i in 1:num_simulations
    lines!(1:num_steps, gbm_paths[:, i])
end
f
```

Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'Scene' is deprecated and will be removed in a future version of Makie.

25. Scenario Generation



25.5.2.2. Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

GARCH models capture time-varying volatility. They are often used in conjunction with other models to forecast volatility. It is defined as

$$\sigma_t^2 = \omega + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$r_t = \varepsilon_t \sqrt{\sigma_t^2}$$

- σ_t^2 is the conditional variance at time t
- r_t is the return at time t
- ε_t is a white noise or innovation process
- $\omega, \alpha_1, \beta_1$ are model parameters

```
# Set seed for reproducibility
Random.seed!(1234)

# GARCH(1,1) parameters
α₀ = 0.01      # Constant term
α₁ = 0.1       # Coefficient for lagged squared returns
```

25.5. Common Economic Scenario Generation Approaches

```
β₁ = 0.8          # Coefficient for lagged conditional volatility

# Number of time steps and simulations
num_steps = 252
num_simulations = 1_000

# Time increment
Δt = 1 / 252

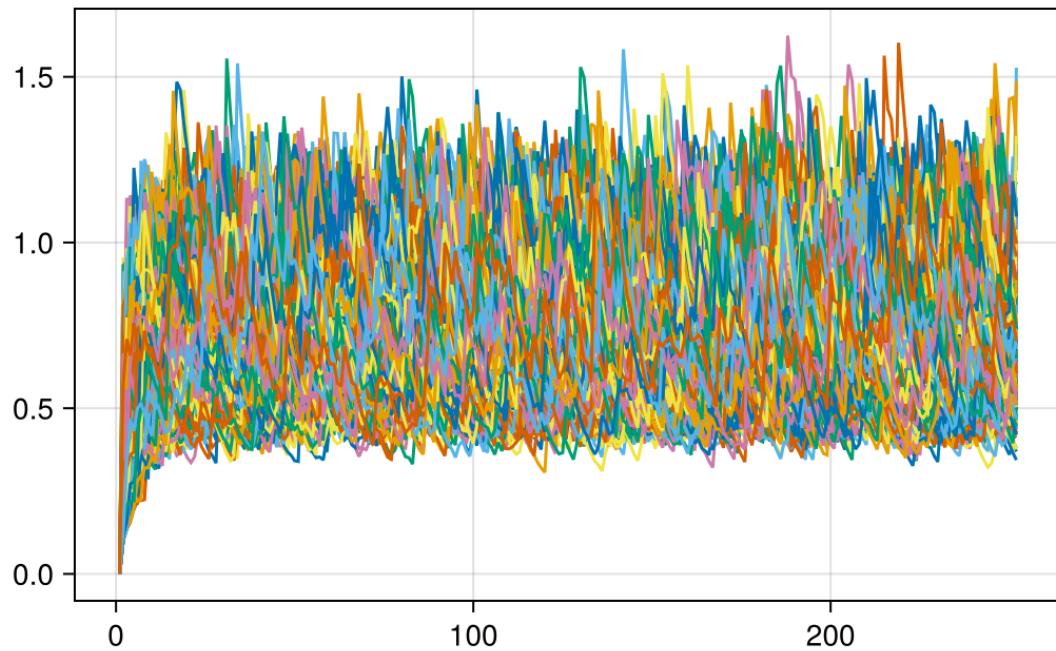
# Function to simulate GARCH(1,1) volatility
function garch_simulation(α₀, α₁, β₁, num_steps, num_simulations)
    volatility_paths = zeros(num_steps, num_simulations)
    for j in 1:num_simulations
        ε = randn(num_steps)
        squared_returns = zeros(num_steps)
        for i in 2:num_steps
            squared_returns[i] = α₀ + α₁ * ε[i-1]^2 + β₁ * squared_returns[i-1]
            volatility_paths[i, j] = sqrt(squared_returns[i])
        end
    end
    return volatility_paths
end

# Run GARCH simulation
garch_paths = garch_simulation(α₀, α₁, β₁, num_steps, num_simulations)

# Plot the simulated volatility paths
f = Figure()
Axis(f[1, 1])
for i in 1:num_simulations
    lines!(1:num_steps, garch_paths[:, i])
end
f
```

```
┌ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```

25. Scenario Generation



25.5.3. Copulas

Simulating data using copulas involves generating multivariate samples with specified marginal distributions and a copula structure.

```
# Set seed for reproducibility
Random.seed!(1234)

# Marginal distributions (e.g., normal)
marginal1 = Normal(0, 1)
marginal2 = Normal(0, 1)

# Clayton copula parameters
theta = 0.5

# Number of data points
num_points = 1000

# Generate independent samples from marginals
u1 = rand(marginal1, num_points)
u2 = rand(marginal2, num_points)
```

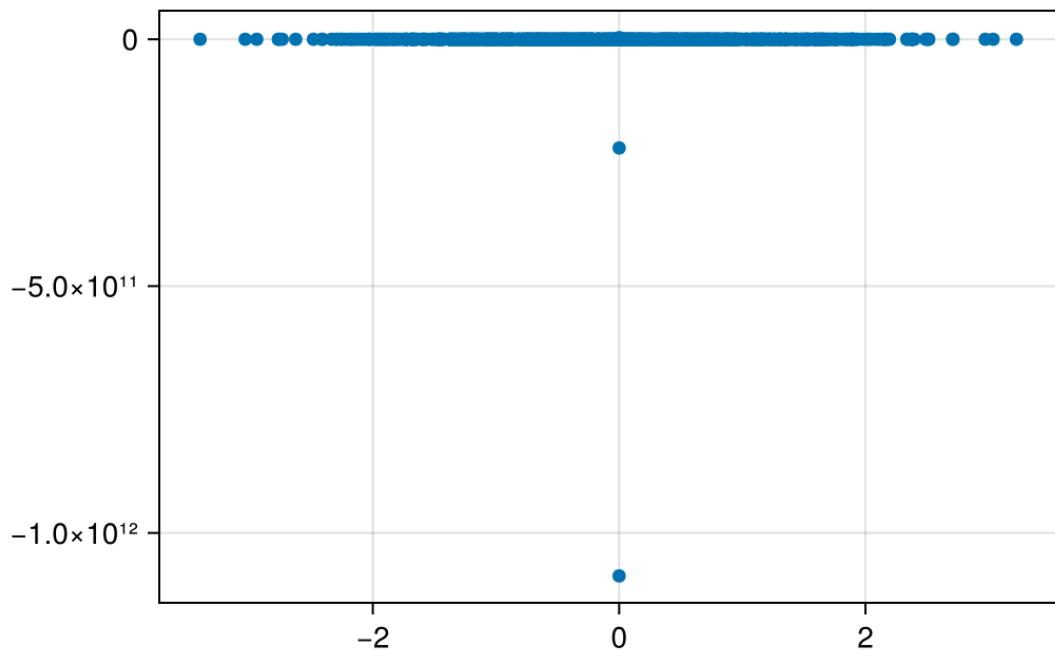
25.6. Benchmarking

```
# Clayton copula simulation
function clayton_copula_simulation(u1, u2, theta)
    v1 = u1
    v2 = u2 .* ((theta .* u1) .^ (-1 / theta - 1))
    return v1, v2
end

# Simulate Clayton copula
v1, v2 = clayton_copula_simulation(u1, u2, theta)

# Plot the simulated bivariate data
f = Figure()
Axis(f[1, 1])
scatter!(v1, v2)
f
```

```
[ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```



25.6. Benchmarking

26. Similarity Analysis

[Drafting note: based on some examples. Needs to be revised with more exposition.]

26.1. In This Chapter

Given a set of interest, understanding the relative similarity (or not) of features of interest is useful in classification and data compression techniques.

26.2. Setup

```
using CSV, DataFrames
using LinearAlgebra
using StatsBase, TableTransforms
using CairoMakie
using NearestNeighbors
```

26.3. The Data

Stored data can generally be categorized into two formats: tabular (structured) and non-tabular (unstructured). Structured data format is a structured way of organizing and presenting data in rows and columns, resembling a table. This format is widely used for storing and representing structured datasets, making it easy to read, analyze, and manipulate data. The most common example of structured data is a spreadsheet, where data is organized into rows and columns. Structured data can also be stored in relational databases for easier lookups and matching. On the other hand, unstructured data refers to data that lacks a predefined data model or structure. Unlike structured data, which fits neatly into tables or databases, unstructured data does not have a predefined schema. It can include text documents, images, audio files, video files, social media posts, and more.

Structured data can be further categorized into numerical and categorical data based on the types of values they represent. The following data tables will be referenced throughout the chapter. Real numerical data can easily be converted or normalized to a series

26. Similarity Analysis

of floating points, and real categorical data to a series of binary literals through one-hot encoding procedures.

```
sample_csv_data =
    IOBuffer(
        raw"id,sex,benefit_base,education,occupation,issue_age
1,M,100000.0,college,1,30.0
2,F,200000.0,master,3,20.0
3,M,150000.0,high_school,4,40.0
4,F,50000.0,college,2,60.0
5,M,250000.0,college,1,40.0
6,F,200000.0,high_school,2,30.0"
    )

IOBuffer(data=UInt8[...], readable=true, writable=false, seekable=true, append=false, size=278,
df = CSV.read(sample_csv_data, DataFrame)
df_num = apply(MinMax(), df[:, [:benefit_base, :issue_age]])[1]

| benefit_base | issue_age | |
|---|---|---|
| Float64 | Float64 |
| 1 | 0.25 | 0.25 |
| 2 | 0.75 | 0.0 |
| 3 | 0.5 | 0.5 |
| 4 | 0.0 | 1.0 |
| 5 | 1.0 | 0.5 |
| 6 | 0.75 | 0.25 |

arr_cat = hcat(indicatormat(df.sex)', indicatormat(df.education)', indicatormat(df.occupation)'

6x9 Matrix{Bool}:
0 1 1 0 0 1 0 0 0
1 0 0 0 1 0 0 1 0
0 1 0 1 0 0 0 0 1
1 0 1 0 0 0 1 0 0
0 1 1 0 0 1 0 0 0
1 0 0 1 0 0 1 0 0
```

For unstructured data, due to the nature of their variety, the choice of representation depends on the type of data and the specific task at hand. For text data, a Word2Vec embedding is commonly used, while Convolutional Neural Networks (CNNs) are for image data and wave transforms are for audio data. No matter which transformation is applied, unstructured data can generally be converted to a series of floating points, just like numerical structured data.

26.4. Common Similarity Measures

The following measures are commonly used to calculate similarities.

26.4.1. Euclidean Distance (L2 norm)

Euclidean distance, also known as the L2 norm, is defined as

$$d = \sqrt{\sum_{i=1}^n (w_i - v_i)^2}$$

The distance is usually meaningful when applied to numerical data. The following Julia code shows the Euclidean distance for the first two rows in df_num.

```
#d12 = √(Σ((Array(df_num[1, :]) .- Array(df_num[2, :])) .* (Array(df_num[1, :]) .- Array(df_num[2, :])))
d12 = LinearAlgebra.norm(Array(df_num[1, :]) .- Array(df_num[2, :]))
```

0.5590169943749475

26.4.2. Manhattan Distance (L1 Norm)

Manhattan distance, also known as the L1 norm, is defined as

$$d = \sum_{i=1}^n |w_i - v_i|$$

The distance is also usually meaningful when applied to numerical data. The following Julia code shows the Euclidean distance for the first two rows in df_num.

```
#d12 = Σ(abs.(Array(df_num[1, :]) .- Array(df_num[2, :])))
d12 = LinearAlgebra.norm1(Array(df_num[1, :]) .- Array(df_num[2, :]))
```

0.75

26. Similarity Analysis

26.4.3. Cosine Similarity

Cosine similarity is defined as

$$d = \frac{\sum_{i=1}^n w_i \cdot v_i}{\sqrt{\sum_{i=1}^n w_i^2} \cdot \sqrt{\sum_{i=1}^n v_i^2}}$$

The distance would be meaningful when applied to both numerical and categorical data.

The following Julia code shows the cosine similarity for the first two rows in df_num.

```
d12 = (Array(df_num[1, :]) · Array(df_num[2, :])) / norm(df_num[1, :]) / norm(df_num[2, :])
```

```
0.7071067811865475
```

The following Julia code shows the cosine similarity for the first and the third rows in arr_cat.

```
d13 = (arr_cat[1, :] · arr_cat[3, :]) / norm(arr_cat[1, :]) / norm(arr_cat[3, :])
```

```
0.3333333333333337
```

Note how similar the syntax of processing for numerical or categorical data is. Multiple dispatch allows Julia to identify most efficient underlying procedure for different types of data. For categorical data, the *dot* operation on binary vectors is essentially count of 1's, while for numerical data it is the *dot* operation for most numerical processing libraries.

26.4.4. Jaccard Similarity

Jaccard similarity is defined as

$$d = \frac{|W \cap V|}{|W \cup V|}$$

The distance is usually meaningful when applied to categorical data. The following Julia code shows the Jaccard similarity for the first and the third rows in arr_cat.

```
d13 = (arr_cat[1, :] · arr_cat[3, :]) / sum(arr_cat[1, :] .| arr_cat[3, :])
```

```
0.2
```

26.5. *k*-Nearest Neighbor (*kNN*) Clustering

26.4.5. Hamming Distance

Hamming distance is defined as $d = \text{Number of positions at which } w \text{ and } v \text{ differ}$. The distance is usually meaningful when applied to categorical data. The following Julia code shows the Hamming distance for the first and the third rows in arr_cat.

```
d13 = sum(arr_cat[1, :] .⊤ arr_cat[3, :])
```

4

26.5. *k*-Nearest Neighbor (*kNN*) Clustering

kNN is primarily known as a classification algorithm, but it can also be used for clustering, particularly in the context of density-based clustering. Density-based clustering identifies regions in the data space where the density of data points is higher, and it groups points in these high-density regions. The core idea of *kNN* clustering is to assign each data point to a cluster based on the density of its neighbors. A data point becomes a core point if it has at least a specified number of neighbors within a certain distance.

```
# Create a kNN model
k = 1
knn_model = KDTree(Array(df_num))

# Query point for prediction
query_point = rand(2)

# Find k nearest neighbors
indices, distances = knn(knn_model, query_point, k)

# Display results
println("Query Point: $query_point")
println("Nearest Neighbors Indices: $indices")
println("Distances to Neighbors: $distances")

f = Figure()
Axis(f[1, 1])
scatter!(df_num[:, 1], df_num[:, 2])
scatter!(query_point[1], query_point[2])
f
```

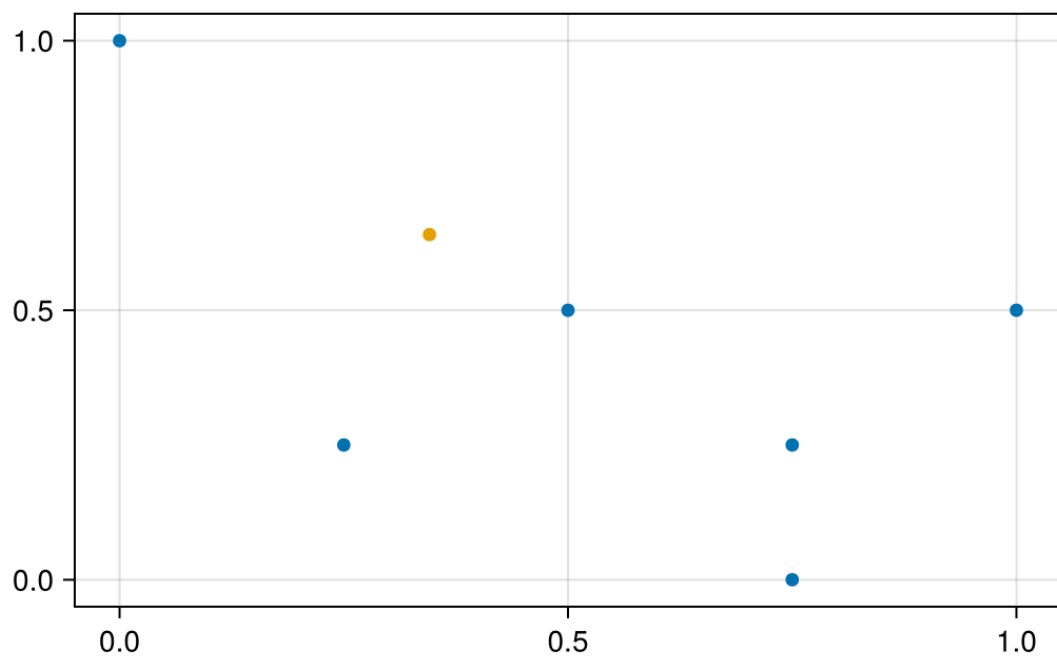
26. Similarity Analysis

Query Point: [0.3455600517323645, 0.6402966372892922]

Nearest Neighbors Indices: [1]

Distances to Neighbors: [0.14548728905691138]

```
[ Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
└ @ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```



26.6. Benchmarking

27. Portfolio Optimization

27.1. In This Chapter

Optimization in a portfolio context with examples of asset selection under different constraints and objectives.

27.2. Setup

```
using CSV, DataFrames
using JuMP, Ipopt, LinearAlgebra
using CairoMakie
```

27.3. The Data

```
μ = [0.1, 0.15, 0.12] # returns
ρ = [0.1 0.05 0.03;
      0.05 0.12 0.04;
      0.03 0.04 0.08] # covariances
n_a = length(μ) # number of assets
```

3

27.4. Theory

Harry Markowitz introduced the modern portfolio theory in 1952. The main idea is that investors are pursuing to maximize their expected return of a portfolio given a certain amount of risk. By definition any portfolio yielding a higher return must have higher amount of risk, so there is a trade-off between desired expected returns and allowable risks. The risk versus maximized expected return relationship can be plotted out as a curve, a.k.a. the efficient frontier.

27.5. Mathematical tools

27.5.1. Mean-variance optimization model

Mean-variance optimization is a mathematical framework that seeks to maximize expected returns while minimizing portfolio variance (or standard deviation). It involves calculating the expected return and risk of individual assets and finding the optimal combination of assets to achieve the desired risk-return tradeoff.

$$\begin{aligned} \text{minimize} \quad & w^T \Sigma w \\ \text{subject to} \quad & R^T \geq \mu_{\text{target}} \\ & 1^T w = 1 \\ & w \geq 0 \end{aligned}$$

```
# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Objective: minimize portfolio variance
@objective(model, Min, sum(w[i] * ρ[i, j] * w[j] for i in 1:n_a, j in 1:n_a))
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or positive
@constraint(model, sum(w) == 1)
# May also add additional constraints
# target_return = 0.1
# @constraint(model, dot(μ, w) >= target_return)
# Solve the optimization problem
optimize!(model)
# Print results
@show "Optimal Portfolio Weights:"
for i = 1:n_a
    @show ("Asset ", i, ": ", value.(w)[i])
end

*****
This program contains Ipopt, a library for large-scale nonlinear optimization.
Ipopt is released as open source code under the Eclipse Public License (EPL).
For more information visit https://github.com/coin-or/Ipopt
*****
```

"Optimal Portfolio Weights:" = "Optimal Portfolio Weights:"

```
("Asset ", i, ":", value.(w)[i]) = ("Asset ", 1, ":", 0.3333333012309821)
("Asset ", i, ":", value.(w)[i]) = ("Asset ", 2, ":", 0.16666675086886984)
("Asset ", i, ":", value.(w)[i]) = ("Asset ", 3, ":", 0.4999999479001481)
```

In mean-variance portfolio optimization, incorporating a cost of risk-based capital on assets is a practical consideration that reflects the additional capital required to support riskier assets in a portfolio. This approach ensures that the optimization process not only maximizes returns relative to risk but also considers the regulatory or internal cost implications associated with holding riskier assets.

$$\begin{aligned} & \text{maximize} && w^T R_{adj} \\ & \text{subject to} && w^T \Sigma w \leq \sigma_{max}^2 \\ & && 1^T w = 1 \\ & && w \geq 0 \end{aligned}$$

where $R_{adj} = [(\mu_1 - \lambda_1), (\mu_2 - \lambda_2), \dots, (\mu_N - \lambda_N)]$ is the adjusted expected returns.

```
# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
r = μ .- [0.01, 0.02, 0.05] # risk adjusted returns
σ²_max = 0.1 # maximum portfolio variance
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Objective: minimize portfolio variance
@objective(model, Max, sum(w[i] * r[i] for i in 1:n_a))
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or positive
@constraint(model, sum(w) == 1)
# Constraints: Sum of allowable portfolio variance is limited
@constraint(model, sum(w[i] * ρ[i, j] * w[j] for i in 1:n_a, j in 1:n_a) <= σ²_max)
# May also add additional constraints
# target_return = 0.1
# @constraint(model, dot(μ, w) >= target_return)
# Solve the optimization problem
optimize!(model)
# Print results
@show "Optimal Portfolio Weights:"
for i = 1:n_a
    @show ("Asset ", i, ":", value.(w)[i])
end
```

27. Portfolio Optimization

```
"Optimal Portfolio Weights:" = "Optimal Portfolio Weights:"
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 1, ": ", 0.16666454953827514)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 2, ": ", 0.8333339906581219)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 3, ": ", 1.4598036030472386e-6)
```

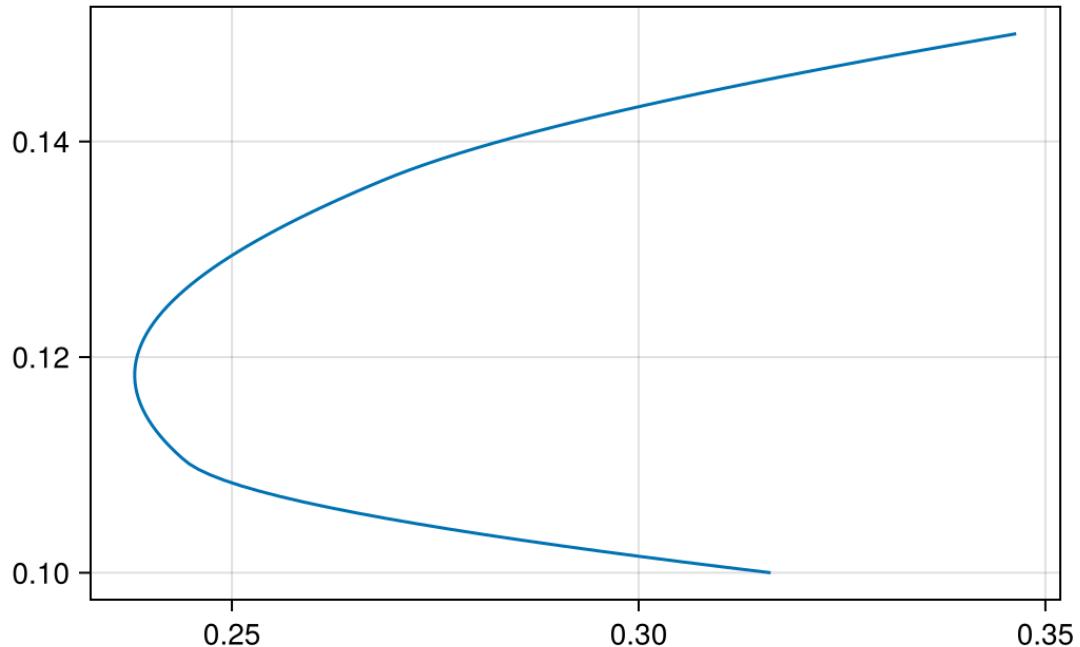
27.5.2. Efficient frontier analysis

The efficient frontier represents the set of portfolios that offer the highest expected return for a given level of risk or the lowest risk for a given level of return. Efficient frontier analysis involves plotting risk-return combinations for different portfolios and identifying the optimal portfolio on the frontier.

```
# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Define objective function: minimize portfolio variance
portfolio_variance = w'ρ * w
@objective(model, Min, portfolio_variance)
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or p
@constraint(model, sum(w) == 1)
# Generate a range of target returns
points = 100
target_returns = range(minimum(μ), maximum(μ), length=points)

efficient_frontier = []
for target_return in target_returns
    # Add additional constraint for target return
    @constraint(model, c, dot(μ, w) == target_return)
    # Solve the problem
    optimize!(model)
    # Show solution
    if termination_status(model) == MOI.LOCALLY_SOLVED
        push!(efficient_frontier, (sqrt(objective_value(model)), target_return))
    end
    unregister(model, :c)
    delete(model, c)
end
# Plot Efficient Frontier
fig = Figure()
Axis(fig[1, 1])
lines!(map(x → x[1], efficient_frontier), map(x → x[2], efficient_frontier))
fig
```

```
[Warning: Found 'resolution' in the theme when creating a 'Scene'. The 'resolution' keyword for 'S
@ Makie ~/.julia/packages/Makie/GtFuI/src/scenes.jl:227
```



27.5.3. Black-Litterman

The Black-Litterman model combines the views of investors with market equilibrium assumptions to generate optimal portfolios. It starts with a market equilibrium portfolio and adjusts it based on investor views and confidence levels. The model incorporates subjective opinions while maintaining diversification and risk management principles.

$$\begin{aligned} & \text{maximize} \quad \mu^T w - \lambda \cdot \frac{1}{2} w^T \Sigma w \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1 \\ & \quad w_i \geq 0, \quad \forall i \end{aligned}$$

```
λ = 2.5 # risk aversion
rfr = 0.02 # risk free rate
```

27. Portfolio Optimization

```
# Market equilibrium parameters (prior)
μ_market = [0.08, 0.08, 0.08] # Market equilibrium return
Σ_market = ρ # Market equilibrium covariance matrix
# Investor views
Q = μ # Expected returns on assets according to investor views
P = [1 0 0; 0 1 0; 0 0 1]      # Pick matrix specifying which assets views are on
Ω = [0.001^2 0.0 0.0; 0.0 0.002^2 0.0; 0.0 0.0 0.003^2] # Views uncertainty (covariance matrix)

# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Black-Litterman expected return adjustment
Σ_prior_inv = inv(Σ_market)
τ = 0.05 # Scaling factor
# Calculate the posterior expected returns
μ_posterior = Σ_prior_inv * (τ * Σ_market * (Σ_prior_inv + P' * inv(Ω) * P)) \
              + (τ * Σ_market * (Σ_prior_inv * μ_market + P' * inv(Ω) * Q)) + Σ_prior_inv * μ_market
# Objective: maximize sharpe ratio
sr = (w' * μ_posterior - rfr) / (λ / 2 * w' * Σ_market * w)
@objective(model, Max, sr)
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or positive
@constraint(model, sum(w) == 1)
# Solve the optimization problem
optimize!(model)
# Print results
v = sqrt(value.(w)' * Σ_market * value.(w))
@show "Optimal Portfolio Weights, Expected Portfolio Return, Portfolio Volatility:", v
for i = 1:n_a
    @show ("Asset ", i, ": ", value.(w)[i], value.(w)[i] * μ_posterior[i])
end
```

```
("Optimal Portfolio Weights, Expected Portfolio Return, Portfolio Volatility:", v) = ("Optimal Portfolio Weights, Expected Portfolio Return, Portfolio Volatility:", v)
("Asset ", 1, ": ", value.(w)[1], value.(w)[1] * μ_posterior[1]) = ("Asset ", 1, ": ", 2.0824146208)
("Asset ", 2, ": ", value.(w)[2], value.(w)[2] * μ_posterior[2]) = ("Asset ", 2, ": ", 0.2311617288)
("Asset ", 3, ": ", value.(w)[3], value.(w)[3] * μ_posterior[3]) = ("Asset ", 3, ": ", 0.7688382502)
```

27.5.4. Risk Parity

Risk parity is an asset allocation strategy that allocates capital based on risk rather than traditional measures such as market capitalization or asset prices. It aims to balance risk contributions across different assets or asset classes to achieve a more stable portfolio.

Risk parity portfolios often include assets with different risk profiles, such as stocks, bonds, and commodities.

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^N (w_i \cdot \sqrt{\sigma_i})^2 \\ \text{subject to} \quad & \sum_{i=1}^N w_i = 1 \\ & w_i \geq 0, \quad \forall i \end{aligned}$$

```
# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Objective: minimize portfolio variance
portfolio_variance = w'ρ * w
margin = (ρ * w ./ sqrt(portfolio_variance)) .* w
risk_contributions = margin ./ sum(margin)
target = repeat([1.0 / n_a], n_a)
@objective(model, Max, sum((risk_contributions .- target) .^ 2))
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or p
@constraint(model, sum(w) == 1)
# Solve the optimization problem
optimize!(model)
# Print results
@show "Optimal Portfolio Weights:"
for i = 1:n_a
    @show ("Asset ", i, ": ", value.(w)[i])
end

"Optimal Portfolio Weights:" = "Optimal Portfolio Weights:"
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 1, ": ", -6.957484531612737e-9)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 2, ": ", 1.0000000131375544)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 3, ": ", -6.180069741122123e-9)
```

27.5.5. Sharpe Ratio Maximization

The Sharpe ratio measures the risk-adjusted return of a portfolio and is calculated as the ratio of excess return to volatility. Maximizing the Sharpe ratio involves finding the portfolio allocation that offers the highest risk-adjusted return. This approach focuses on achieving the best tradeoff between risk and return.

27. Portfolio Optimization

$$\begin{aligned} & \text{maximize} \quad \frac{E[R_p] - R_f}{\sigma_p} \\ & \text{subject to} \quad \sum_{i=1}^N w_i = 1 \\ & \quad w_i \geq 0, \quad \forall i \end{aligned}$$

```
# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Objective: minimize portfolio variance
rfr = 0.05 # risk free rate
@objective(model, Max, (dot(mu, w) - rfr) / sqrt(sum(w[i] * rho[i, j] * w[j] for i in 1:n_a, j in 1:n_a)))
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or positive
@constraint(model, sum(w) == 1)
# Solve the optimization problem
optimize!(model)
# Print results
@show "Optimal Portfolio Weights:"
for i = 1:n_a
    @show ("Asset ", i, ": ", value.(w)[i])
end

"Optimal Portfolio Weights:" = "Optimal Portfolio Weights:"
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 1, ": ", 0.010841995514843134)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 2, ": ", 0.5352292318109132)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 3, ": ", 0.45392877267424375)
```

27.5.6. Robust Optimization

Robust optimization techniques aim to create portfolios that are resilient to uncertainties and fluctuations in market conditions. These techniques consider a range of possible scenarios and optimize portfolios to perform well across different market environments. A robust parameter in robust portfolio optimization is typically chosen to ensure the portfolio's performance remains stable and satisfactory under different market conditions or variations in input data. Robust optimization may involve incorporating stress tests, scenario analysis, or robust risk measures into the portfolio construction process.

$$\begin{aligned}
 & \text{minimize} && w^T \Sigma w + \gamma \|w - w_0\|_2^2 \\
 & \text{subject to} && \sum_{i=1}^N w_i = 1 \\
 & && w_i \geq 0, \quad \forall i \\
 & && \|(\Sigma^{1/2}(w - w_0))\|_2 \leq \epsilon
 \end{aligned}$$

```

# Create an optimization model
model = Model(optimizer_with_attributes(Ipopt.Optimizer, "print_level" => 0))
# Set up weights as variables to optimize
@variable(model, w[1:n_a] >= zero(0.0))
# Objective: minimize portfolio variance
ε = 0.05 # Uncertainty level
γ = 0.1 # Robustness parameter
w₀ = [0.3, 0.4, 0.3] # expected weights
@objective(model, Min, dot(w, ρ * w) + γ * sum((w[i] - w₀[i])^2 for i in 1:n_a))
# Constraints: Sum of portfolio weights should equal to 1, and all weights should be zero or p
@constraint(model, sum(w) == 1)
@constraint(model, sum((ρ[i, j] * (w[i] - w₀[i]) * (w[j] - w₀[j]))) for i in 1:n_a, j in 1:n_a) <
# Solve the optimization problem
optimize!(model)
# Print results
@show "Optimal Portfolio Weights:"
for i = 1:n_a
    @show ("Asset ", i, ": ", value.(w)[i])
end

"Optimal Portfolio Weights:" = "Optimal Portfolio Weights:"
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 1, ": ", 0.31250000098314346)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 2, ": ", 0.31250000376951037)
("Asset ", i, ": ", value.(w)[i]) = ("Asset ", 3, ": ", 0.37499999524734623)

```

27.5.7. Asset weights from different methodologies

Table 27.1.: Optimized asset weights from different methodologies

Methodology	Asset weights
Standard mean variance (with RBC costs)	[0.33, 0.17, 0.50] [0.17, 0.83, 0.00]

27. Portfolio Optimization

Methodology	Asset weights
Black-Litterman	[0.00, 0.23, 0.77]
Risk parity	[0.00, 1.00, 0.00]
Sharpe ratio	[0.01, 0.54, 0.45]
Robust	[0.31, 0.31, 0.38]

Seeing from the asset weights from a standard mean variance approach, due to RBC costs asset weights shifted to ones with higher yields. Asset weights from Sharpe ratio approach aligns with the Sharpe ratio for each asset. Those from Robust approach seek results not far away from expected weights under different conditions.

27.6. Practical considerations

27.6.1. Fractional purchases of assets

In traditional portfolio optimization, fractional purchases of assets refer to the ability to allocate fractions or percentages of capital to individual assets. However, in certain contexts or practical implementations, fractional purchases may not be allowed or considered.

- Practical constraints. Some investment vehicles or platforms may restrict investors from purchasing fractions of shares or assets. For instance, certain mutual funds, exchange-traded funds (ETFs), or other investment products may require whole units of shares to be purchased.
- Simplicity and cost-effectiveness. Handling fractional shares can add complexity and operational costs to portfolio management, especially in terms of transaction fees, administrative overhead, and reconciliation processes.
- Market liquidity. Some assets may have limited liquidity or trading volumes, making it impractical or difficult to execute fractional purchases without significantly impacting market prices or transaction costs.
- Regulatory considerations. Regulations in certain jurisdictions may impose restrictions on fractional share trading or ownership, potentially limiting the ability to include fractional purchases in portfolio optimization strategies.

27.6.2. Large number of assets

In portfolio optimization, a penalty factor for a large volume of assets typically refers to a mechanism or adjustment applied to the optimization process to mitigate the potential biases or challenges that arise when dealing with a large number of assets. This concept is particularly relevant in the context of mean-variance optimization and other optimization frameworks where computational efficiency and practical portfolio management considerations come into play. Too many assets may have the following issues.

- Dimensionality. As the number of assets (or dimensions) increases in a portfolio, traditional optimization methods may become computationally intensive or prone to overfitting. This is because the complexity of the optimization problem grows exponentially with the number of assets.
- Sparsity and concentration. In practice, not all assets may contribute equally to portfolio performance. Some assets may have negligible impact on the overall portfolio characteristics (such as risk or return) due to low weights or correlations with other assets.
- Penalizing excessive complexity. A penalty factor can be introduced to penalize portfolios that overly diversify or allocate small weights to a large number of assets. This encourages the optimization process to focus on more significant assets or reduce the complexity of the portfolio structure.

There are various ways to implement a penalty factor for a large volume of assets:

- Regularization techniques. Techniques like Lasso (L1 regularization) or Ridge (L2 regularization) regression can penalize small weights or excessive diversification by adding a penalty term to the objective function.
- Subset selection. Methods that explicitly select a subset of assets based on their contribution to portfolio performance, rather than including all assets indiscriminately.
- Heuristic adjustments. Introducing heuristic rules or adjustments based on practical portfolio management principles or empirical observations.

28. Bayesian Mortality Modeling

! Drafting Notes

Ideas: - First plot graph of outcomes and discuss some key features: - More variance when the subset are smaller

- Just plot the data first, with the group colors, to explain what we are looking at and for consistency in subsequent plots
- Generate sample data using parameters sampled from chain and show the bands on the associated outcomes.

28.1. Generating fake data

The problem of interest is to look at mortality rates, which are given in terms of exposures (whether or not a life experienced a death in a given year).

We'll grab some example rates from an insurance table, which has a "selection" component: When someone enters observation, say at age 50, their mortality is path dependent (so for someone who started being observed at 50 will have a different risk/mortality rate at age 55 than someone who started being observed at 45).

Additionally, there may be additional groups of interest, such as:

- high/medium/low risk classification
- sex
- group (e.g. company, data source, etc.)
- type of insurance product offered

The example data will start with only the risk classification above. “””

```
using MortalityTables
using Turing
using UUIDs
using DataFramesMeta
using MCMCChains
```

28. Bayesian Mortality Modeling

```
using LinearAlgebra
using CairoMakie
using StatsBase
using OffsetArrays

n = 10_000
inforce = map(1:n) do i
    (
        id=uuid1(),
        issue_age=rand(30:70),
        risk_level=rand(1:3),
    )
)

end

10000-element Vector{@NamedTuple{id::UUID, issue_age::Int64, risk_level::Int64}}:
(id = UUID("77729530-5802-11ef-30e9-01057ec6aee4"), issue_age = 46, risk_level = 1)
(id = UUID("7772954e-5802-11ef-2968-95b0968224ee"), issue_age = 60, risk_level = 1)
(id = UUID("77729558-5802-11ef-160b-87ce1f11c836"), issue_age = 57, risk_level = 1)
(id = UUID("77729558-5802-11ef-3eb3-9d68f1d278fe"), issue_age = 54, risk_level = 1)
(id = UUID("77729558-5802-11ef-0dc9-416481c9987d"), issue_age = 30, risk_level = 1)
(id = UUID("77729562-5802-11ef-21f5-b7cb29f4ca5a"), issue_age = 43, risk_level = 1)
(id = UUID("77729562-5802-11ef-04cf-9f778b81c8aa"), issue_age = 51, risk_level = 2)
(id = UUID("77729562-5802-11ef-09dd-2deb8e8b51b8"), issue_age = 66, risk_level = 3)
(id = UUID("7772956c-5802-11ef-358a-cd8a222e2953"), issue_age = 58, risk_level = 3)
(id = UUID("7772956c-5802-11ef-0a63-b19c91f92bc0"), issue_age = 54, risk_level = 3)
(id = UUID("7772956c-5802-11ef-39e6-6d21337da798"), issue_age = 51, risk_level = 3)
(id = UUID("7772956c-5802-11ef-0235-d3001f1dba9f"), issue_age = 55, risk_level = 2)
(id = UUID("77729578-5802-11ef-11b4-bf112c53279a"), issue_age = 39, risk_level = 3)
⋮
(id = UUID("77731050-5802-11ef-2961-e35e65944449"), issue_age = 30, risk_level = 1)
(id = UUID("77731050-5802-11ef-253c-1f94c7de77af"), issue_age = 58, risk_level = 3)
(id = UUID("7773105a-5802-11ef-1b1a-2d592191e13e"), issue_age = 37, risk_level = 3)
(id = UUID("7773105a-5802-11ef-2a5f-e1799a46bef9"), issue_age = 39, risk_level = 1)
(id = UUID("7773105a-5802-11ef-0c5a-054859e34e4e"), issue_age = 38, risk_level = 3)
(id = UUID("77731064-5802-11ef-1d85-23d77cfdae01"), issue_age = 31, risk_level = 3)
(id = UUID("77731064-5802-11ef-2ed3-afba1ad470f3"), issue_age = 48, risk_level = 2)
(id = UUID("77731064-5802-11ef-11cc-7b01383252f6"), issue_age = 64, risk_level = 3)
(id = UUID("7773106e-5802-11ef-036e-c5a839efdfa8"), issue_age = 45, risk_level = 3)
(id = UUID("7773106e-5802-11ef-1d52-a9011a03adf8"), issue_age = 49, risk_level = 1)
(id = UUID("7773106e-5802-11ef-271a-b53d64405fb7"), issue_age = 41, risk_level = 1)
(id = UUID("77731078-5802-11ef-1049-4f4e279d79dc"), issue_age = 67, risk_level = 3)
```

28.1. Generating fake data

```
base_table = MortalityTables.table("2001 VBT Residual Standard Select and Ultimate - Male Nonsmokers")

function tabular_mortality(params, issue_age, att_age, risk_level)
    q = params.ultimate[att_age]
    if risk_level == 1
        q *= 0.7
    elseif risk_level == 2
        q = q
    else
        q *= 1.5
    end
end

tabular_mortality (generic function with 1 method)

function model_outcomes(inforce, assumption, assumption_params; n_years=5)

    outcomes = map(inforce) do pol
        alive = 1
        sim = map(1:n_years) do t
            att_age = pol.issue_age + t - 1
            q = assumption(
                assumption_params,
                pol.issue_age,
                att_age,
                pol.risk_level
            )
            if rand() < q
                out = (att_age=att_age, exposures=alive, death=1)
                alive = 0
                out
            else
                (att_age=att_age, exposures=alive, death=0)
            end
        end
        filter!(x → x.exposures == 1, sim)
    end

    df = DataFrame(inforce)

    df.outcomes = outcomes
```

28. Bayesian Mortality Modeling

```
df = flatten(df, :outcomes)

df.att_age = [x.att_age for x in df.outcomes]
df.death = [x.death for x in df.outcomes]
df.exposures = [x.exposures for x in df.outcomes]
select!(df, Not(:outcomes))

end

exposures = model_outcomes(inforce, tabular_mortality, base_table)
data = combine(groupby(exposures, [:issue_age, :att_age])) do subdf
    (exposures=nrow(subdf),
     deaths=sum(subdf.death),
     fraction=sum(subdf.death) / nrow(subdf))
end

data2 = combine(groupby(exposures, [:issue_age, :att_age, :risk_level])) do subdf
    (exposures=nrow(subdf),
     deaths=sum(subdf.death),
     fraction=sum(subdf.death) / nrow(subdf))
end
```

28.2. 1: A single binomial parameter model

	issue_age	att_age	risk_level	exposures	deaths	fraction
	Int64	Int64	Int64	Int64	Int64	Float64
1	30	30	1	85	0	0.0
2	30	30	2	80	0	0.0
3	30	30	3	78	0	0.0
4	30	31	1	85	0	0.0
5	30	31	2	80	0	0.0
6	30	31	3	78	0	0.0
7	30	32	1	85	0	0.0
8	30	32	2	80	1	0.0125
9	30	32	3	78	0	0.0
10	30	33	1	85	0	0.0
11	30	33	2	79	0	0.0
12	30	33	3	78	0	0.0
13	30	34	1	85	0	0.0
14	30	34	2	79	0	0.0
15	30	34	3	78	1	0.0128205
16	31	31	1	70	0	0.0
17	31	31	2	73	0	0.0
18	31	31	3	69	0	0.0
19	31	32	1	70	0	0.0
20	31	32	2	73	0	0.0
21	31	32	3	69	0	0.0
22	31	33	1	70	0	0.0
23	31	33	2	73	0	0.0
24	31	33	3	69	0	0.0
25	31	34	1	70	0	0.0
26	31	34	2	73	0	0.0
27	31	34	3	69	0	0.0
28	31	35	1	70	0	0.0
29	31	35	2	73	0	0.0
30	31	35	3	69	0	0.0
...

28.2. 1: A single binomial parameter model

Estiamte q , the average mortality rate, not accounting for any variation within the population/sample. Our model is defines as:

$$q \sim Beta(1, 1) \\ p(death) \sim Binomial(q)$$

28. Bayesian Mortality Modeling

```

@model function mortality(data, deaths)
    q ~ Beta(1, 1)
    for i = 1:nrow(data)
        deaths[i] ~ Binomial(data.exposures[i], q)
    end
end

m1 = mortality(data, data.deaths)

DynamicPPL.Model{typeof(mortality), (:data, :deaths), (), (), Tuple{DataFrame, Vector{Int64}}, T}

```

Row	issue_age	att_age	exposures	deaths	fraction
	Int64	Int64	Int64	Int64	Float64
1	30	30	243	0	0.0
2	30	31	243	0	0.0
3	30	32	243	1	0.00411523
4	30	33	242	0	0.0
5	30	34	242	1	0.00413223
6	31	31	212	0	0.0
7	31	32	212	0	0.0
8	31	33	212	0	0.0
9	31	34	212	0	0.0
10	31	35	212	0	0.0
11	32	32	214	1	0.0046729
:	:	:	:	:	:
196	69	69	254	9	0.0354331
197	69	70	245	8	0.0326531
198	69	71	237	7	0.0295359
199	69	72	230	9	0.0391304
200	69	73	221	11	0.0497738
201	70	70	238	3	0.012605
202	70	71	235	9	0.0382979
203	70	72	226	5	0.0221239
204	70	73	221	6	0.0271493
205	70	74	215	5	0.0232558

184 rows omitted, deaths = [0, 0, 1, 0, 1, 0, 0, 0, 0, 0 ... 9, 8, 7, 9, 1]

28.2.1. Sampling from the posterior

We use a No-U-Turn-Sampler (NUTS) technique to sample multiple chains at once:

```

num_chains = 4
chain = sample(m1, NUTS(), MCMCThreads(), 400, num_chains)

```

28.2. 1: A single binomial parameter model

Chains MCMC chain (400×13×4 Array{Float64, 3}):

```

Iterations      = 201:1:600
Number of chains = 4
Samples per chain = 400
Wall duration    = 1.89 seconds
Compute duration = 7.46 seconds
parameters       = q
internals        = lp, n_steps, is_accept, acceptance_rate, log_density, hamiltonian_energy, hamil

Summary Statistics
parameters      mean      std      mcse    ess_bulk    ess_tail     rhat   e ...
Symbol          Float64   Float64   Float64   Float64   Float64   Float64   ...
q              0.0082   0.0004   0.0000   548.3059   998.5838   1.0136   ...
                                         ...                                          
                                         1 column omitted

Quantiles
parameters      2.5%      25.0%     50.0%     75.0%     97.5%
Symbol          Float64   Float64   Float64   Float64   Float64
q              0.0074   0.0079   0.0082   0.0085   0.0090

```

Here, we have asked for the outcomes to be modeled via a single parameter for the population. We see that the posterior distribution of q is very close to the overall population mortality rate:

```
sum(data.deaths) / sum(data.exposures)
```

```
0.008174442190669371
```

However, We can see that the sampling of possible posterior parameters doesn't really fit the data very well since our model was so simplified. The lines represent the posterior binomial probability.

This is saying that for the observed data, if there really is just a single probability p that governs the true process that came up with the data, there's a pretty narrow range of values it could possibly be:

```

let
  data_weight = log.(data.exposures)
  #data_weight = ./(data_weight ./ maximum(data_weight) .* 20)
  f = Figure(title="Parametric Bayesian Mortality")

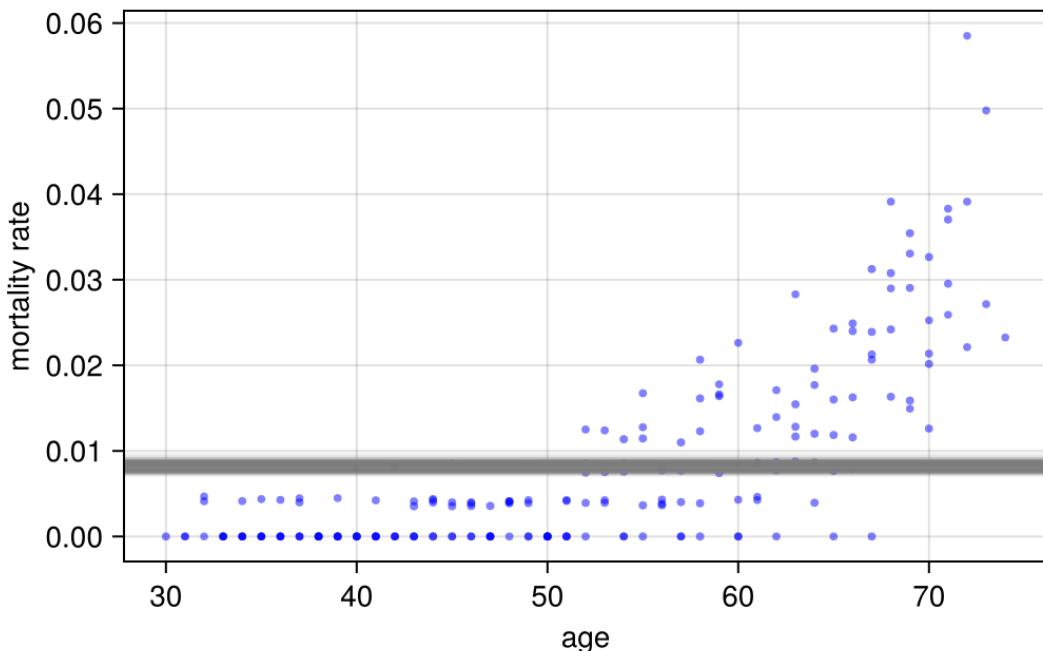
```

28. Bayesian Mortality Modeling

```
)  
ax = Axis(f[1, 1],  
          xlabel="age",  
          ylabel="mortality rate",  
          # ylims=(0.0, 0.25),  
)  
scatter!(ax,  
        data.att_age,  
        data.fraction,  
        markersize=data_weight,  
        color=(:blue, 0.5),  
        label="Experience data point (size indicates relative exposure quantity),")  
  
# show n samples from the posterior plotted on the graph  
n = 300  
ages = sort!(unique(data.att_age))  
  
q_posterior = sample(chain, n)[:q]  
  
for i in 1:n  
    hlines!(ax, [q_posterior[i]], color=(:grey, 0.1))  
end  
  
# Need to simulate at individual level and then aggregate?  
  
sim05 = Float64[]  
sim95 = Float64[]  
for r in eachrow(data)  
    outcomes = map(1:n) do i  
        rand(Binomial(r.exposures, q_posterior[i]), 500)  
    end  
    push!(sim05, quantile(Iterators.flatten(outcomes), 0.05) / r.exposures)  
    push!(sim95, quantile(Iterators.flatten(outcomes), 0.95) / r.exposures)  
end  
  
f
```

28.2. 1: A single binomial parameter model

```
end
```



```
let  
n = 300  
q_posterior = sample(chain, n)[:q]
```

```
end
```

```
2-dimensional AxisArray{Float64,2,...} with axes:  
:iter, 1:300  
:chain, 1:1
```

```
And data, a 300×1 Matrix{Float64}:  
0.007502951724144315  
0.008562087781377417  
0.007169458392007883  
0.008096427340539426  
0.008595755248581636  
0.008280669478329361  
0.008805900762599236  
0.00802399748233711  
0.008158228044380975  
0.008412240110158597
```

28. Bayesian Mortality Modeling

```
0.008324785351523478
0.00792766736608433
0.008670803376835047
::
0.008395089882455966
0.007820396289529371
0.008338916954301065
0.008053989130394741
0.008458822915592867
0.008100686018561612
0.008085550530297754
0.008333910547751955
0.008381973726059307
0.008240749386746137
0.008618202581263902
0.00805475325843088
```

28.3. 2. Parametric model

In this example, we utilize a MakehamBeard parameterization because it's already very similar in form to a logistic function. This is important because our desired output is a probability (ie the probability of a death at a given age), so the value must be constrained to be in the interval between zero and one.

The **prior** values for a,b,c, and k are chosen to constrain the hazard (mortality) rate to be between zero and one.

This isn't an ideal parameterization (e.g. we aren't including information about the select underwriting period), but is an example of utilizing Bayesian techniques on life experience data. "

```
@model function mortality2(data, deaths)
    a ~ Exponential(0.1)
    b ~ Exponential(0.1)
    c = 0.0
    k ~ truncated(Exponential(1), 1, Inf)

    # use the variables to create a parametric mortality model
    m = MortalityTables.MakehamBeard(; a, b, c, k)

    # loop through the rows of the dataframe to let Turing observe the data
    # and how consistent the parameters are with the data
    for i = 1:nrow(data)
```

28.3. 2. Parametric model

```

age = data.att_age[i]
q = MortalityTables.hazard(m, age)
deaths[i] ~ Binomial(data.exposures[i], q)
end
end

mortality2 (generic function with 2 methods)
```

We combine the model with the data and sample from the posterior using a similar call as before:

```
m2 = mortality2(data, data.deaths)

chain2 = sample(m2, NUTS(), MCMCThreads(), 400, num_chains)
```

Chains MCMC chain (400×15×4 Array{Float64, 3}):

```

Iterations      = 201:1:600
Number of chains = 4
Samples per chain = 400
Wall duration    = 7.5 seconds
Compute duration = 26.32 seconds
parameters       = a, b, k
internals        = lp, n_steps, is_accept, acceptance_rate, log_density, hamiltonian_energy, hamil
```

Summary Statistics

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	e ...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
a	0.0000	0.0000	0.0000	442.2301	403.1110	1.0106	...
b	0.0997	0.0064	0.0003	440.8886	456.6988	1.0108	...
k	1.9351	0.8926	0.0310	564.8268	455.9055	1.0056	...

1 column omitted

Quantiles

parameters	2.5%	25.0%	50.0%	75.0%	97.5%
Symbol	Float64	Float64	Float64	Float64	Float64
a	0.0000	0.0000	0.0000	0.0000	0.0001
b	0.0866	0.0956	0.0998	0.1040	0.1117
k	1.0179	1.2554	1.6750	2.3023	4.2121

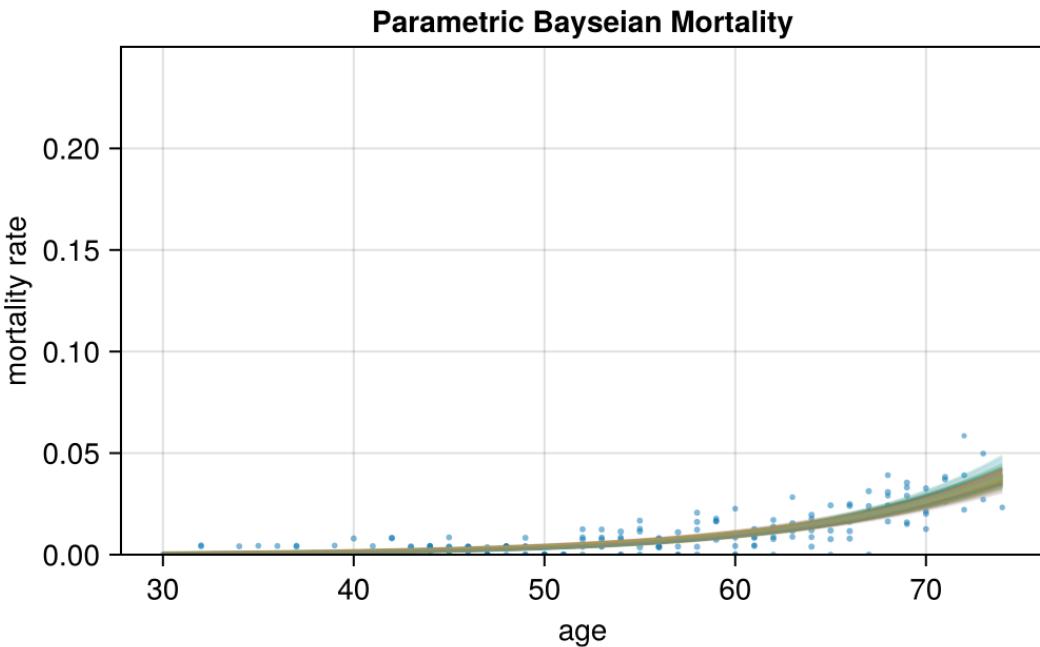
28. Bayesian Mortality Modeling

```
summarize(chain2)  
plot(chain2)
```

28.3.1. Plotting samples from the posterior

We can see that the sampling of possible posterior parameters fits the data well:

```
let  
    data_weight = data.exposures ./ sum(data.exposures)  
    data_weight = .√(data_weight ./ maximum(data_weight) .* 20)  
  
    p = scatter(  
        data.att_age,  
        data.fraction,  
        markersize=data_weight,  
        alpha=0.5,  
        label="Experience data point (size indicates relative exposure quantity)",  
        axis=  
            xlabel="age",  
            limits=(nothing, nothing, 0.0, 0.25),  
            ylabel="mortality rate",  
            title="Parametric Bayesian Mortality"  
    )  
)  
  
# show n samples from the posterior plotted on the graph  
n = 300  
ages = sort!(unique(data.att_age))  
  
for i in 1:n  
    s = sample(chain2, 1)  
    a = only(s[:a])  
    b = only(s[:b])  
    k = only(s[:k])  
    c = 0  
    m = MortalityTables.MakehamBeard(; a, b, c, k)  
    lines!(ages, age → MortalityTables.hazard(m, age), alpha=0.1, label="")  
end  
p  
end
```



```

let
    data_weight = log.(data.exposures)
    #data_weight = .√(data_weight ./ maximum(data_weight) .* 20)
    f = Figure(title="Parametric Bayesian Mortality")
)
ax = Axis(f[1, 1],
    xlabel="age",
    ylabel="mortality rate",
    # ylims=(0.0, 0.25),
)
scatter!(ax,
    data.att_age,
    data.fraction,
    markersize=data_weight,
    color=(:blue, 0.5),
    label="Experience data point (size indicates relative exposure quantity),")

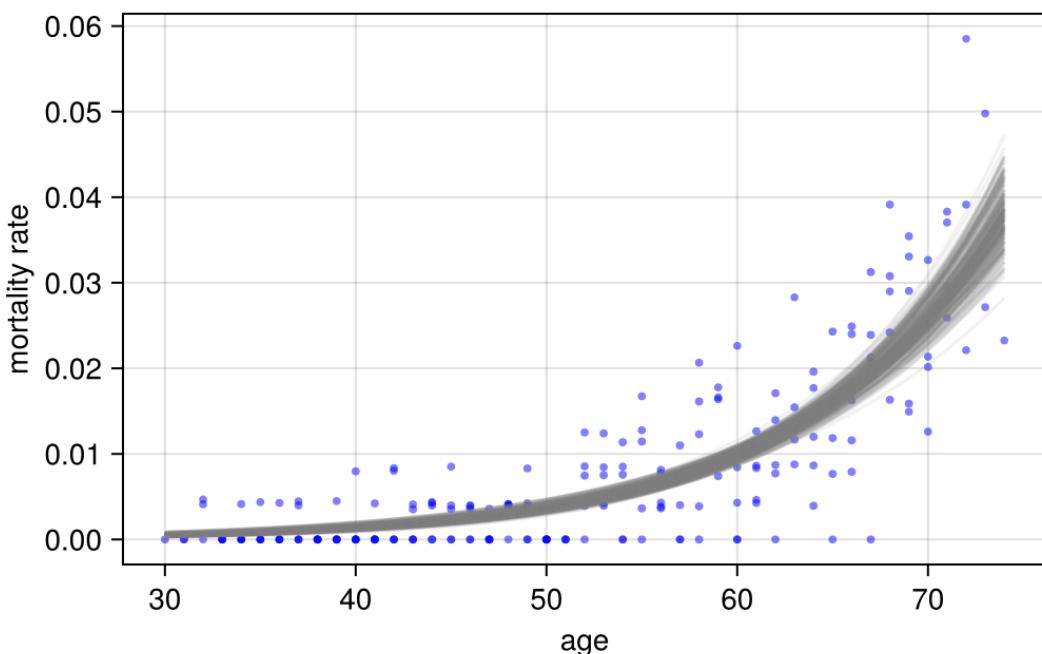
# show n samples from the posterior plotted on the graph
n = 300
ages = sort!(unique(data.att_age))

for i in 1:n
    s = sample(chain2, 1)

```

28. Bayesian Mortality Modeling

```
a = only(s[:a])
b = only(s[:b])
k = only(s[:k])
c = 0
m = MortalityTables.MakehamBeard(; a, b, c, k)
qs = MortalityTables.hazard.(m, ages)
lines!(ax, ages, qs, color=(:grey, 0.1))
end
f
end
```



28.4. 3. Parametric model

This model extends the prior to create a multi-level model. Each risk class (`risk_level`) gets its own a parameter in the `MakehamBeard` model. The prior for a_i is determined by the hyperparameter \bar{a} .

```
@model function mortality3(data, deaths)
    risk_levels = length(levels(data.risk_level))
    b ~ Exponential(0.1)
    ā ~ Exponential(0.1)
    a ~ filldist(Exponential(ā), risk_levels)
```

```

c = 0
k ~ truncated(Exponential(1), 1, Inf)

# use the variables to create a parametric mortality model

# loop through the rows of the dataframe to let Turing observe the data
# and how consistent the parameters are with the data
for i = 1:nrow(data)
    risk = data.risk_level[i]

    m = MortalityTables.MakehamBeard(; a=a[risk], b, c, k)
    age = data.att_age[i]
    q = MortalityTables.hazard(m, age)
    deaths[i] ~ Binomial(data.exposures[i], q)
end
end

m3 = mortality3(data2, data2.deaths)

chain3 = sample(m3, NUTS(), 1000)

summarize(chain3)

```

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	e ...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
b	0.1011	0.0059	0.0004	260.9032	312.0404	1.0040	...
ā	0.0001	0.0001	0.0000	414.7803	515.2298	1.0032	...
a[1]	0.0000	0.0000	0.0000	264.9345	354.1170	1.0031	...
a[2]	0.0000	0.0000	0.0000	252.2647	315.9927	1.0054	...
a[3]	0.0000	0.0000	0.0000	261.9619	321.2269	1.0025	...
k	1.8983	0.8670	0.0347	402.5420	253.8603	1.0017	...

1 column omitted

```

let data = data2

data_weight = data.exposures ./ sum(data.exposures)
data_weight = .√(data_weight ./ maximum(data_weight) .* 20)
color_i = data.risk_level

p = scatter(
    data.att_age,

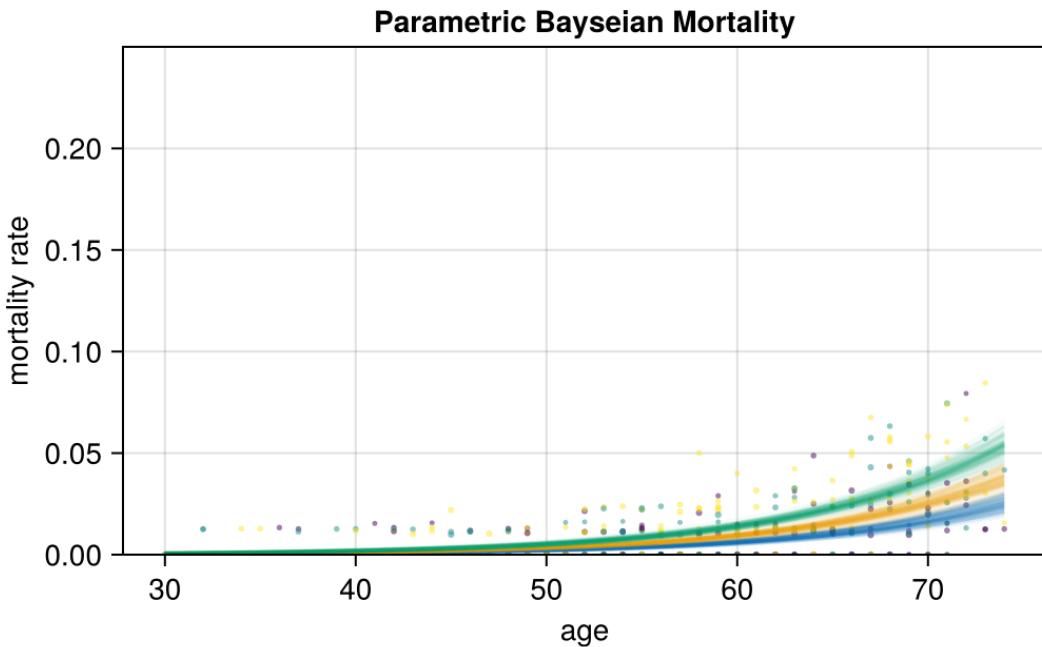
```

28. Bayesian Mortality Modeling

```
data.fraction,
marksize=data_weight,
alpha=0.5,
color=color_i,
label="Experience data point (size indicates relative exposure quantity)",
axis=(
    xlabel="age",
    limits=(nothing, nothing, 0.0, 0.25),
    ylabel="mortality rate",
    title="Parametric Bayesian Mortality"
)
)

# show n samples from the posterior plotted on the graph
n = 100

ages = sort!(unique(data.att_age))
for r in 1:3
    for i in 1:n
        s = sample(chain3, 1)
        a = only(s[Symbol("a[$r]")])
        b = only(s[:b])
        k = only(s[:k])
        c = 0
        m = MortalityTables.MakehamBeard(; a, b, c, k)
        if i == 1
            lines!(ages, age → MortalityTables.hazard(m, age), label="risk level $r", alpha=0.5)
        else
            lines!(ages, age → MortalityTables.hazard(m, age), label="", alpha=0.2, color=color_i)
        end
    end
end
p
end
```



28.5. Handling non-unit exposures

The key is to use the Poisson distribution, which is a continuous approximation to the Binomial distribution:

```
@model function mortality4(data, deaths)
    risk_levels = length(levels(data.risk_level))
    b ~ Exponential(0.1)
    ā ~ Exponential(0.1)
    a ~ filldist(Exponential(ā), risk_levels)
    c ~ Beta(4, 18)
    k ~ truncated(Exponential(1), 1, Inf)

    # use the variables to create a parametric mortality model

    # loop through the rows of the dataframe to let Turing observe the data
    # and how consistent the parameters are with the data
    for i = 1:nrow(data)
        risk = data.risk_level[i]

        m = MortalityTables.MakehamBeard(; a=a[risk], b, c, k)
```

28. Bayesian Mortality Modeling

```

age = data.att_age[i]
q = MortalityTables.hazard(m, age)
deaths[i] ~ Poisson(data.exposures[i] * q)
end
end

m4 = mortality4(data2, data2.deaths)

chain4 = sample(m4, NUTS(), 1000)

```

Chains MCMC chain (1000×19×1 Array{Float64, 3}):

```

Iterations          = 501:1:1500
Number of chains   = 1
Samples per chain  = 1000
Wall duration      = 35.05 seconds
Compute duration   = 35.05 seconds
parameters         = b, ā, a[1], a[2], a[3], c, k
internals          = lp, n_steps, is_accept, acceptance_rate, log_density, hamiltonian_energy, hamil

```

Summary Statistics

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	e ...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
b	0.1190	0.0095	0.0006	226.2427	277.0332	0.9996	...
ā	0.0000	0.0002	0.0000	336.0061	300.7459	1.0031	...
a[1]	0.0000	0.0000	0.0000	235.4927	246.7399	0.9999	...
a[2]	0.0000	0.0000	0.0000	237.6709	224.6614	0.9995	...
a[3]	0.0000	0.0000	0.0000	221.5668	231.8254	0.9996	...
c	0.0010	0.0003	0.0000	334.7884	345.1912	0.9998	...
k	2.0597	0.9574	0.0418	364.1287	291.2789	1.0007	...

1 column omitted

Quantiles

parameters	2.5%	25.0%	50.0%	75.0%	97.5%
Symbol	Float64	Float64	Float64	Float64	Float64
b	0.1002	0.1127	0.1184	0.1251	0.1380
ā	0.0000	0.0000	0.0000	0.0000	0.0001
a[1]	0.0000	0.0000	0.0000	0.0000	0.0000
a[2]	0.0000	0.0000	0.0000	0.0000	0.0000
a[3]	0.0000	0.0000	0.0000	0.0000	0.0000
c	0.0004	0.0007	0.0010	0.0012	0.0017

28.5. Handling non-unit exposures

```

k      1.0433    1.3603    1.7836    2.5436    4.6717

risk_factors4 = [mean(chain4[Symbol("a[$f]")]) for f in 1:3]

risk_factors4 ./ risk_factors4[2]

let data = data2

data_weight = data.exposures ./ sum(data.exposures)
data_weight = ./maximum(data_weight) .* 20
color_i = data.risk_level

p = scatter(
    data.att_age,
    data.fraction,
    markersize=data_weight,
    alpha=0.5,
    color=color_i,
    label="Experience data point (size indicates relative exposure quantity)",
    axis=(xlabel="age",
          limits=(nothing, nothing, 0.0, 0.25),
          ylabel="mortality rate",
          title="Parametric Bayesian Mortality"
    )
)

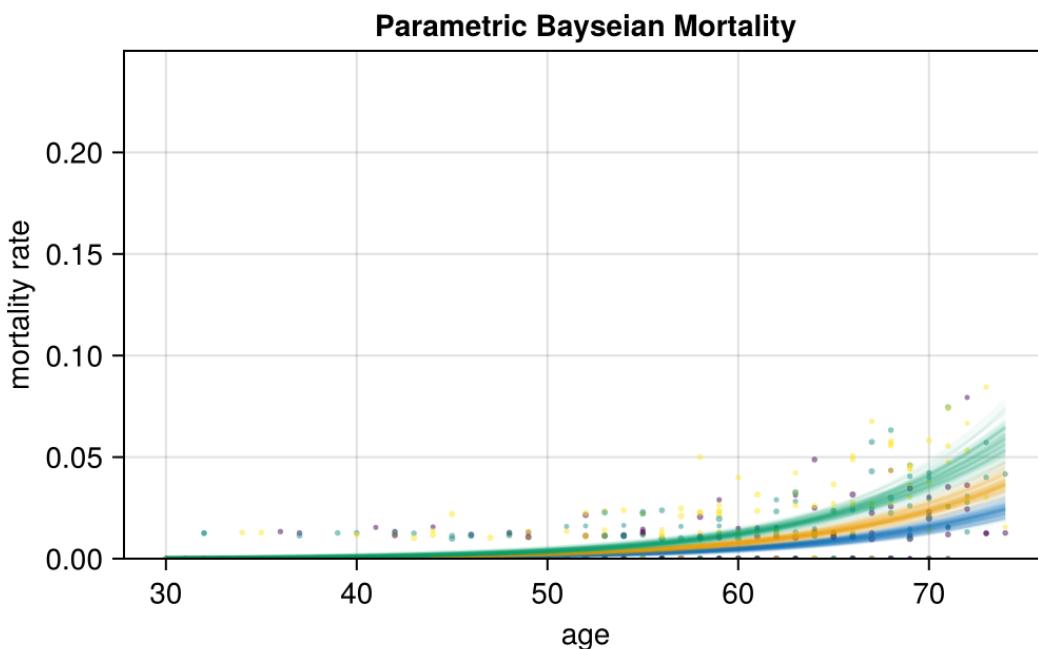
# show n samples from the posterior plotted on the graph
n = 100

ages = sort!(unique(data.att_age))
for r in 1:3
    for i in 1:n
        s = sample(chain4, 1)
        a = only(s[Symbol("a[$r]")])
        b = only(s[:b])
        k = only(s[:k])
        c = 0
        m = MortalityTables.MakehamBeard(; a, b, c, k)
        if i == 1
            lines!(ages, age → MortalityTables.hazard(m, age), label="risk level $r", alpha=0.2, color=color_i)
        else
            lines!(ages, age → MortalityTables.hazard(m, age), label="", alpha=0.2, color=color_i)
        end
    end
end

```

28. Bayesian Mortality Modeling

```
    end  
end  
p  
end
```



28.6. Predictions

We can generate predictive estimates by passing a vector of `missing` in place of the outcome variables and then calling `predict`.

We get a table of values where each row is the prediction implied by the corresponding chain sample, and the columns are the predicted value for each of the outcomes in our original dataset.

```
preds = predict(mortality4(data2, fill(missing, length(data2.deaths))), chain4)
```

Chains MCMC chain (1000×615×1 Array{Float64, 3}):

```
Iterations      = 1:1:1000
Number of chains = 1
Samples per chain = 1000
parameters      = deaths[1], deaths[2], deaths[3], deaths[4], deaths[5], deaths[6], deaths[7]
```

```
internals      =
```

Summary Statistics

parameters	mean	std	mcse	ess_bulk	ess_tail	rhat	...
Symbol	Float64	Float64	Float64	Float64	Float64	Float64	...
deaths[1]	0.0920	0.2994	0.0096	969.8982	975.4719	0.9994	...
deaths[2]	0.0990	0.3276	0.0103	990.7121	964.7356	1.0000	...
deaths[3]	0.1080	0.3201	0.0107	892.8681	901.2020	0.9993	...
deaths[4]	0.0850	0.3031	0.0098	957.0014	949.6680	1.0003	...
deaths[5]	0.0960	0.3080	0.0092	1118.7967	1004.0241	0.9995	...
deaths[6]	0.1160	0.3326	0.0104	1022.2281	1012.2033	0.9990	...
deaths[7]	0.1040	0.3245	0.0107	942.3261	951.6927	0.9991	...
deaths[8]	0.1040	0.3214	0.0109	884.1969	921.1422	0.9991	...
deaths[9]	0.1000	0.3164	0.0101	989.8808	1003.8824	0.9994	...
deaths[10]	0.1200	0.3489	0.0110	999.8412	995.2423	0.9991	...
deaths[11]	0.0810	0.2873	0.0101	814.5509	817.3605	1.0019	...
deaths[12]	0.0980	0.3138	0.0107	871.8743	885.6581	0.9990	...
deaths[13]	0.1190	0.3507	0.0118	884.5493	859.5603	0.9996	...
deaths[14]	0.0960	0.3080	0.0096	1034.2194	1012.2033	0.9990	...
deaths[15]	0.1150	0.3285	0.0105	972.4146	941.7170	0.9990	...
deaths[16]	0.0870	0.2958	0.0090	1091.7732	1010.1399	0.9994	...
deaths[17]	0.0780	0.2757	0.0088	977.3639	970.2756	0.9991	...
deaths[18]	0.0980	0.3263	0.0104	968.9211	937.7982	1.0004	...
deaths[19]	0.0830	0.2902	0.0093	990.6260	1010.0085	1.0016	...
deaths[20]	0.0910	0.3047	0.0100	916.1305	913.4316	0.9995	...
deaths[21]	0.1080	0.3232	0.0100	1031.7879	1012.2033	1.0000	...
deaths[22]	0.0840	0.2984	0.0097	937.7273	931.7600	0.9991	...
deaths[23]	0.0810	0.2942	0.0090	1062.9747	774.7252	0.9990	...
:	:	:	:	:	:	:	~

1 column and 592 rows omitted

Quantiles

parameters	2.5%	25.0%	50.0%	75.0%	97.5%
Symbol	Float64	Float64	Float64	Float64	Float64
deaths[1]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[2]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[3]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[4]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[5]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[6]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[7]	0.0000	0.0000	0.0000	0.0000	1.0000

28. Bayesian Mortality Modeling

deaths[8]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[9]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[10]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[11]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[12]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[13]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[14]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[15]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[16]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[17]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[18]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[19]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[20]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[21]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[22]	0.0000	0.0000	0.0000	0.0000	1.0000
deaths[23]	0.0000	0.0000	0.0000	0.0000	1.0000
:	:	:	:	:	:

592 rows omitted

29. Other Useful Techniques

29.1. In this chapter

Other useful techniques are surveyed, such as: memoization to avoid repeated computations, psuedo-monte carlo, creating a model office, and tips on modeling a complete balance sheet. Also covered are elements of practical review such as static and dynamic validations, and implied rate analysis.

29.2. Conceptual Techniques

29.2.1. Taking things to the Extreme

Consider what happens if something is taken to an extreme. For example, what happens in the model if we input negative rates? Where should negative rates be allowed and can the model handle them?

29.2.2. Range Bounding

Sometimes you just need to know that an outcome is within a certain range - if you can develop a "high" and "low" estimate by making assumptions that you know are outside of feasible ranges, then you can determine whether something is reasonable or within tolerances.

To take an example from the pages of interview questions: say you need to determine if a mortgaged property's value is greater than the amount of the outstanding loan (say \$100,000). You don't have an appraisal, but know that it's in reasonable condition and that (1) a comparable house with many more issues sold for \$100 per square foot. You also don't know the square footage of the house, but know from the number of rooms and layout that it must be at least 1000 square feet. Therefore you know that the value should at least be greater than:

$$\frac{\$100}{\text{sq. ft}} \times 1000 \text{sq. ft} = \$100,000$$

29. Other Useful Techniques

We'd then conclude that the value of the house very likely exceeds the outstanding balance of the loan and resolves our query without complex modeling or expensive appraisals.

29.3. Modeling Techniques

29.3.1. Serialization

Part VII.

Appendices

30. Set up Julia and the Computing Environment

30.1. Installation

Julia is open source and can be downloaded from JuliaLang.org and is available for all major operating systems. After you download and install, then you have Julia installed and can access the **REPL**, or Read-Eval-Print-Loop, which can run complete programs or function as powerful day-to-day calculator. However, many people find it more comfortable to work in a text editor or **IDE** (Integrated Development Environment).

If you are looking for managed installations with a curated set of packages for use within an organization, there are ways to self-host package repositories and otherwise administratively manage packages. Julia Computing offers managed support with enterprise solutions, including push-button cloud compute capabilities.

30.2. Package Management

Julia comes with **Pkg**, a built-in package manager. With it, you can install packages, pin certain versions, recreate environments with the same set of dependencies, and upgrade/remove/develop packages easily. It's one of the things that *just works* and makes Julia stand out versus alternative languages that don't have a de-facto way of managing or installing packages.

Package installation is accomplished interactively in the REPL or executing commands.

- In the REPL, you can change to the Package Management Mode by hitting `]` and, e.g., add `DataFrames CSV` to install the two packages. Hit `[backspace]` to exit that mode in the REPL.
- The same operation without changing REPL modes would be: `using Pkg; Pkg.add(["DataFrames", "CSV"])`

Related to packages, are **environments** which are a self-contained workspaces for your code. This lets you install only packages that are relevant to the current work. It also lets you 'remember' the exact set of packages and versions that you used. In fact, you can

30. Set up Julia and the Computing Environment

share the environment with others, and it will be able to recreate the same environment as when you ran the code. This is accomplished via a `Project.toml` file, which tracks the direct dependencies you've added, along with details about your project like its version number. The `Manifest.toml` tracks the entire dependency tree.

Reproducibility via the environment tools above is a really key aspect that will ensure Julia code is consistent across time and users, which is important for financial controls.

30.3. Editors

Because Julia is very extensible and amenable to analysis of its own code, you can typically find plugins for whatever tool you prefer to write code in. A few examples:

30.3.1. Visual Studio Code

Visual Studio Code is a free editor from Microsoft. There's a full-featured Julia plugin available, which will help with auto-completion, warnings, and other code hints that you might find in a dedicated editor (e.g. PyCharm or RStudio). Like those tools, you can view plots, search documentation, show datasets, debug, and manage version control.

30.3.2. Notebooks

Notebooks are typically more interactive environments than text editors - you can write code in cells and see the results side-by-side.

The most popular notebook tool is Jupyter ("Julia, Python, R"). It is widely used and fits in well with exploratory data analysis or other interactive workflows. It can be installed by adding the `IJulia.jl` package.

`Pluto.jl` is a newer tool, which adds reactivity and interactivity. It is also more amenable to version control than Jupyter notebooks because notebooks are saved as plain Julia scripts. Pluto is unique to Julia because of the language's ability to introspect and analyze dependencies in its own code. Pluto also has built-in package/environment management, meaning that Pluto notebooks contains all the code needed to reproduce results (as long as Julia and Pluto are installed).

30.4. REPL

30.4.1. Help Mode

31. Environment and Package Management

31.1. In This Section

How to effectively utilize environments to ensure consistent and reproducible results.
How to use and manage packages. How to create a package and share with others.
How to use local registries.

31.2. Projects, Manifests, and Dependencies

Julia comes bundled with Pkg.jl, an environment and package manager. It enables installation of packages from registries, pinning versions for compatibility, and analyzing your dependencies. It uses a couple of files to record this to your project: `Project.toml` and `Manifest.toml`.

31.2.1. Project.toml

A `Project.toml` file defines attributes about the current project and its dependencies. Julia uses this to understand how to reference your current project and what dependencies it should look for from registries when instantiating the project.

 Note

TOML (Tom's Obvious Markup Language) is a modern configuration file format used to store settings and data in a human-readable, plaintext format.

This is a bit abstract, so here is a quick, annotated tour of an example `Project.toml` file:

```
name = "FinanceCore"                                     ①
uuid = "b9b1ffdd-6612-4b69-8227-7663be06e089"          ②
authors = ["alecloudenback <alecloudenback@users.noreply.github.com> and contributors"] ③
version = "2.1.0"

[deps]
Dates = "ade2ca70-3891-5945-98fb-dc099432e06a"           ④
```

31. Environment and Package Management

```
LoopVectorization = "bdcacae8-1622-11e9-2a5c-532679323890"  
Roots = "f2b01f46-fcfa-551c-844a-d8ac1e96c665"
```

```
[compat]  
Dates = "1"  
LoopVectorization = "^0.12"  
Roots = "^1.0, 2"  
julia = "1.6"
```

(5)

- ① The name is the name of your current project which only matters if you turn your project into a package.
- ② A **UUID** is a unique identifier and can be created with Julia's UUIDs standard library.
- ③ The version follows Semantic Versioning ("SemVer") to convey to Pkg (and users!) information that ties a specific version to a specific code commit¹.
- ④ The deps section records the name of direct dependencies and their UUIDs so that Julia can know which packages to grab in order to make your project run.
- ⑤ The compat section defines compatibility with packages can be enforced (via SemVer) to clarify which versions are allowed to be installed in case incompatibilities arise.

When you instantiate a project (see Section 31.3 for more), Julia will essentially add the packages listed under deps, and will **resolve** the compatible versions, generally picking the highest version number for the packages so long as the compat section rule are not broken.

When adding the dependencies, those packages themselves likely specify their own set of dependencies and Julia must resolve the entire **dependency graph** or **dependency tree** to allow your current project to work.

Semantic Versioning

Semantic Versioning ("SemVer") is a scheme which uses the three-component version code to convey meaning about different versions of a package to both users and computer systems. With the version scheme vMAJOR.MINOR.PATCH, the meaning is roughly as follows:

1. MAJOR increments denote changes to the code which make it incompatible with prior versions.
2. MINOR increments denote changes which add features that are compatible with the prior versions.
3. PATCH increments denote changes which fix issues in prior versions and code written against the prior version is still compatible.

As an example, say we are currently using v2.10.4 of a package, and the following

¹When registering a package to a repository, the repository will record the version indicated in the Project.toml file to the git commit id of the package when it is registered.

theoretical options are available for us to upgrade to:

- v2.10.5 - The 4 to 5 indicates that something may have been broken in the prior release and so we should upgrade without fear that we need to make changes to our code (unless we relied on the previously broken code!).
- v2.11.0 - The 10 to 11 bump suggests that the new release contains some features which should not require us to change any of our previously written code.
- v3.0.0 - The 2 to 3 indicates that we will potentially have to modify code that we have written that interfaces with this dependency.

SemVer cannot distill all possible compatibility and upgrade information about a set of packages (e.g. an author may release an update with a MINOR version which also includes fixes).

31.2.2. `Manifest.toml`

The `Manifest.toml` file includes a record of all external dependencies used by the project at hand. Unlike `Project.toml`, this file gets machine generated when Julia instantiates or updates the environment. The contents are basically a long list of your direct dependencies and the dependencies of those direct dependencies and looks something like this:

```
julia_version = "1.10.0"
manifest_format = "2.0"
project_hash = "5fea00df4808d89f9c977d15b8ee992bd408081b"

[[deps.AbstractFFTs]]
deps = ["LinearAlgebra"]
git-tree-sha1 = "d92ad398961a3ed262d8bf04a1a2b8340f915fef"
uuid = "621f4979-c628-5d54-868e-fcf4e3e8185c"
version = "1.5.0"
weakdeps = ["ChainRulesCore", "Test"]

[deps.AbstractFFTs.extensions]
AbstractFFTsChainRulesCoreExt = "ChainRulesCore"
AbstractFFTsTestExt = "Test"

... many more lines
```

31. Environment and Package Management

Note

Starting in Julia 1.11, Manifest files will include a version indication, making it nicer to work with multiple Julia versions at one time on a single system.

31.2.3. Reproducibility

Reproducibility fulfills both practical and principled goals. *Practical* in that we can record the complex chain of dependencies that is used in modern computing in order to potentially re-create a result or demonstrate an audit trail of the tools used. *Principled* in that there are circumstances (like science research) in which we want to be able to replicate results. The combination of `Project.toml` and `Manifest.toml` go a long way towards accomplishing this, as you can share both and with the same hardware and Julia version should be able to get the exact same set of dependencies and therefore run the same code. In practice, this level of reproducibility isn't *usually* needed, as most time a set of code can be run accurately without requiring the exact same set of dependencies.

Since dependencies can have variation between systems (Windows/Mac) and architectures (x86 vs x64), you may not be able to recreate the Manifest exactly. Nevertheless, it's a fairly low bar if you are trying to maintain the utmost level of rigor around the toolchain and Julia is one of the most robust languages regarding tools to support open replication of results.

Artifacts

Julia has a system called **artifacts** which allows specification of a location and hash (a cryptographic key) for data and binaries. The artifact system used to download and verify the contents of a file match the hash. This is designed for more permanent data and less end-user workflows, but we call it out here as another example where Julia takes steps to promote consistency and reproducibility.

For more on data workflows for the end-user, see Chapter 14.

31.3. Environments

Environment is meant to mean, in general, the computer you use and software installed in it. When we speak about **environments** in the Julia context, this means the Julia version and packages available to the current Julia code. For example, from the current code is a given package installed and usable?

31.4. Packages

If you open a Julia REPL, by default you will be in the *global* environment. If you hit] to enter Pkg mode, you should see:

```
(@v1.10) pkg>
```

The (@1.10) indicates that you are using the global environment for the current Julia version (there is no global environment which applies across all Julia versions installed). You can activate a new environment with `activate [environment name]`.

```
(@v1.10) pkg> activate MyNewEnv  
Activating new project at `~/MyNewEnv`
```

This will... not do anything. Yet! When we add a package to this environment, *then* it will create a `Project.toml` and `Manifest.toml` file in that directory. Now that directory is a full fledged Julia project!

Tip

Activate a temporary environment with `activate --temp`. This will give you a temporary environment with a random name, which is very useful for testing out things in a clean, simplified environment (the global environment, like @1.10 still applies.)

31.4. Packages

31.4.1. Packages versus Projects

31.4.2. Basic Package Structure

31.4.3. Extension Packages

31.5. Regisries

31.5.1. Local Registries

32. The Julia Ecosystem Today

A tour of relevant available packages as of 2023.

The Julia ecosystem favors composability and interoperability, enabled by multiple dispatch. In other words, because it's easy to automatically specialize functionality based on the type of data being used, there's much less need to bundle a lot of features within a single package.

As you'll see, Julia packages tend to be less vertically integrated because it's easier to pass data around. Counterexamples of this in Python and R:

- Numpy-compatible packages that are designed to work with a subset of numerically fast libraries in Python
- special functions in Pandas to read CSV, JSON, database connections, etc.
- The Tidyverse in R has a tightly coupled set of packages that works well together but has limitations with some other R packages

Julia is not perfect in this regard, but it's neat to see how frequently things *just work*. It's not magic, but because of Julia features outside the scope of this article it's easy for package developers (and you!) to do this.

Julia also has language-level support for documentation, so packages can follow a consistent style of help-text and have the docs be auto-generated into web pages available locally or online.

The following highlighted packages were chosen for their relevance to typical actuarial work, with a bias towards those used regularly by the authors. This is a small sampling of the over 6000 registered Julia Packages¹

32.0.1. Data

Julia offers a rich data ecosystem with a multitude of available packages. Perhaps at the center of the data ecosystem are `CSV.jl` and `DataFrames.jl`. `CSV.jl` is for reading and writing files text files (namely CSVs) and offers top-class read and write performance. `DataFrames.jl` is a mature package for working with dataframes, comparable to Pandas or `dplyr`.

¹(`time?`) is a simple, built-in function. For true benchmarking purposes, see Section 10.2.

32. The Julia Ecosystem Today

Other notable packages include `ODBC.jl`, which lets you connect to any database (given you have the right drivers installed), and `Arrow.jl` which implements the Apache Arrow standard in Julia.

Worth mentioning also is `Dates`, a built-in package making date manipulation straightforward and robust.

Check out [JuliaData.org](https://juliadata.org) for more packages and information.

32.0.2. Plotting

`Plots.jl` is a meta-package providing an interface to consistently work with several plotting backends, depending if you are trying to emphasize interactivity on the web or print-quality output. You can very easily add animations or change almost any feature of a plot.

`StatsPlots.jl` extends `Plots.jl` with a focus on data visualization and compatibility with dataframes.

`Makie.jl` supports GPU-accelerated plotting and can create very rich, beautiful visualizations, but its main downside is that it has not yet been optimized to minimize the time-to-first-plot.

32.0.3. Statistics

Julia has first-class support for missing values, which follows the rules of three-valued logic so other packages don't need to do anything special to incorporate missing values.

`StatsBase.jl` and `Distributions.jl` are essentials for a range of statistics functions and probability distributions respectively.

Others include:

- `Turing.jl`, a probabilistic programming (Bayesian statistics) library, which is outstanding in its combination of clear model syntax with performance.
- `GLM.jl` for any type of linear modeling (mimicking R's `glm` functionality).
- `LsqFit.jl` for fitting data to non-linear models.
- `MultivariateStats.jl` for multivariate statistics, such as PCA.

You can find more packages and learn about them [here](#).

32.0.4. Machine Learning

Flux, Gen, Knet, and MLJ are all very popular machine learning libraries. There are also packages for PyTorch, Tensorflow, and SciKitML available. One advantage for users is that the Julia packages are written in Julia, so it can be easier to adapt or see what's going on in the entire stack. In contrast to this design, PyTorch and Tensorflow are built primarily with C++.

Another advantage is that the Julia libraries can use automatic differentiation to optimize on a wider range of data and functions than those built into libraries in other languages.

32.0.5. Differentiable Programming

Sensitivity testing is very common in actuarial workflows: essentially, it's understanding the change in one variable in relation to another. In other words, the derivative!

Julia has unique capabilities where almost across the entire language and ecosystem, you can take the derivative of entire functions or scripts. For example, the following is real Julia code to automatically calculate the sensitivity of the ending account value with respect to the inputs:

```
julia> using Zygote

julia> function policy_av(pol)
    COIs = [0.00319, 0.00345, 0.0038, 0.00419, 0.0047, 0.00532]
    av = 0.0
    for (i,coi) in enumerate(COIs)
        av += av * pol.credit_rate
        av += pol.annual_premium
        av -= pol.face * coi
    end
    return av           # return the final account value
end

julia> pol = (annual_premium = 1000, face = 100_000, credit_rate = 0.05);

julia> policy_av(pol)      # the ending account value
4048.08

julia> policy_av'(pol)     # the derivative of the account value with respect to the inputs
(annual_premium = 6.802, face = -0.0275, credit_rate = 10972.52)
```

32. The Julia Ecosystem Today

When executing the code above, Julia isn't just adding a small amount and calculating the finite difference. Differentiation is applied to entire programs through extensive use of basic derivatives and the chain rule. **Automatic differentiation**, has uses in optimization, machine learning, sensitivity testing, and risk analysis. You can read more about Julia's autodiff ecosystem [here](#).

32.0.6. Utilities

There are also a lot of quality-of-life packages, like `Revise.jl` which lets you edit code on the fly without needing to re-run entire scripts.

`BenchmarkTools.jl` makes it incredibly easy to benchmark your code - simply add `@benchmark` in front of what you want to test, and you will be presented with detailed statistics. For example:

```
julia> using ActuaryUtilities, BenchmarkTools

julia> @benchmark present_value(0.05,[10,10,10])

BenchmarkTools.Trial: 10000 samples with 994 evaluations.

Range (min ... max): 33.492 ns ... 829.015 ns | GC (min ... max): 0.00% ... 95.40%
Time   (median):      34.708 ns                  | GC (median):      0.00%
Time   (mean ± σ):   36.599 ns ± 33.686 ns    | GC (mean ± σ):  4.40% ± 4.55%
                                                 ┌─────────────────────────────────────────────────────────────────────────┐
                                                 └────────────────────────────────────────────────────────────────────────┘
                                                 33.5 ns          Histogram: log(frequency) by time          45.6 ns <
```

`Test` is a built-in package for performing testsets, while `Documenter.jl` will build high-quality documentation based on your inline documentation.

`ClipData.jl` lets you copy and paste from spreadsheets to Julia sessions.

32.0.7. Other packages

Julia is a general-purpose language, so you will find packages for web development, graphics, game development, audio production, and much more. You can explore packages (and their dependencies) at <https://juliahub.com/>.

32.0.8. Actuarial packages

Saving the best for last, the next article in the series will dive deeper into actuarial packages, such as those published by JuliaActuary for easy mortality table manipulation, common actuarial functions, financial math, and experience analysis.

References

- Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.
- Leemis, Lawrence M, and Jacquelyn T McQueston. 2008. "Univariate Distribution Relationships." *The American Statistician* 62 (1): 45–53. <https://doi.org/10.1198/000313008x270448>.
- Lewis, N D. 2013. *100 Statistical Tests*. Createspace.
- Murphy, Hannah, and Cristina Criddle. 2024. "Meta AI Chief Says Large Language Models Will Not Reach Human Intelligence." <https://www.ft.com/content/23fab126-f1d3-4add-a457-207a25730ad9>.

