DATA 5690 Class Notes

Computational Methods for Fintech

Dr. Tyler J. Brough

2024-08-26

Table of contents

5

Preface

The Course Syllabus

DATA 5690/6690: Computational Methods for FinTech

Fall 2024

Course Information

- Dates: Aug 26 – Dec 13
- Time: M/W 12:00 – 1:15 pm
- Room: Huntsman Hall 132
- Course Canvas Page
- Course GitHub Page

Instructor Information

- Tyler J. Brough
- Office: BUS 719
- Office Hours: By appointment
- Email: tyler.brough@usu.edu

Course Description

Fintech is a portmanteau of the words *finance* and *technology*. One of the main themes of fintech is that financial markets institutions and information systems and networks are converging. This course will provide a philosophical, historical, and mathematical introduction to the foundations of fintech.

This course will highlight catallactic theory, operational subjective statistical methods, and real options theory as the foundations of modern fintech. Additional topics will be Monte Carlo methods from financial engineering and reinforcement learning from machine learning.

Course Mechanics

- The mode of the course will be the Socratic method. See the first assessment item below.
- There will be assigned readings announced on each Thursday for the coming week. You should read each assigned source thoroughly and completely. I advise you to take notes, especially in the form of zettelkasten reading notes. More will be said about the zettelkasten method later in the course.
- At the beginning of each course period I will run a Python program that will implement a PRNG to select one of you to begin our discussion. It is expected that each member of the course will take part in each discussion. Your thorough readings should provide you ample material for ideas, questions, and discussion topics.
- More will be said about this during class.

Assessment, Deliverables, and Grading

Your grade in this course will be determined by the following:

- **The Socratic Method** (25%)
  - As stated above, the mode of this course will be the Socratic method.
  - The main part of this will be to read the book *The Socratic Method: A Practitioner's Handbook* by Ward Farnsworth.
  - The first deliverable will be a complete set of annotated bibliographies per chapter.
  - The second deliverable will be an essay on how the Socratic method possibly serves as a foundation for fintech entrepreneurship. Additional details will be given.
  - The main component of this item of assessment will be the use of the Socratic method during our course discussions. Your reading of the book will provide you information regarding its proper usage and implementation.
- **Computational Exercises** (15%)
  - There will be multiple assigned computational exercises throughout the course.
  - In each case, the deliverable will be a jupyter notebook file.
- **Midterm Exam** (30%)
  - This exam will cover the first two sections of the course.
  - The deliverable will be a jupyter notebook file.
- **Final Exam** (30%)
  - This exam will be comprehensive, but will mainly focus on options and computational methods in light of the main themes of the course.

  - The deliverable will be a jupyter notebook file.

I will provide a grading rubric for each item by which you will be assessed. Students may resubmit their deliverables as many times as they wish for re-grading conditional on timely original submission.

Important Dates

- 09/02 - Labor Day - no class
- 10/18 - Fall Break - Friday
- 11/27 - 11/29 - Thanksgiving Holiday - Wed-Fri
- 12/02 - 12/06 - No-Test Week
- 12/06 - Last Day of Classes
- 12/09 - 12/13 - Final Examinations

Grading Rubric for the Socratic Method

- At the beginning of the semester, I will allocate each student 60 points.
- Points will be deducted as follows:
  - Missing a class without a valid excuse provided in advance: -10 points
  - If Emmy or I witness you using your cell phone during class for reasons other than class discussion: -2 points
  - Coming to class late: every portion of 5-minute intervals: -2 points (10 points max)
  - If Emmy or I witness you using your laptop for non-course discussion related purposes: -4 points
  - Talking with a neighbor – outside of group work times: -3 points
  - Starting in Part II of the class, points will be deducted for unpreparedness in class discussions. This deduction will increase each week. This is to ensure that students who have done well through the first part of the course remain fully motivated.
- You will earn points by correctly responding to in-class questions, and performing in-class exercises. We will use a randomization device to select participants.
  - Points awarded will depend on the type of question and the quality, comprehensiveness, and thoughtfulness of your answer.
  - In-class exercises, in which you are asked to live code in front of class, are worth more than general discussion points. The former may earn up to 10 points while the latter may earn up to 2 points.
  - Asking an insightful question that reveals a depth of understanding: 1 point (to a maximum of 6)
  - Bringing in an outside source from scholarly literature or popular media that supports our discussion: 1 or 2 points (to a maximum of 10 points)
  - Demonstrating an example of synoptic reading and searching: 1 or 2 points (to a maximum of 10 points)
- Note that early on you can really only be deducted points for unprofessional behavior. As the semester progresses you can only earn points by demonstrating mastery and understanding of the course material.

Part I

The Political Economy of Fintech

Catallactics

This chapter will cover theory of catallactics. See **Mises2007a?**.

Mises on Economic Calculation

Economic Calculation The process by which fallible men acting in a changing world choose, on the basis of monetary prices, among an infinite variety of imaginable and possible methods of production. Economic calculation in the absence of a generally accepted medium of exchange (money) is inconceivable. In a market economy, money prices stem from the bids and asks of producers and consumers. Thus, prices reflect the rel- ative urgency of their various wants. The prices at which goods and ser- vices are exchanged influence the choices consumers and producers make when bidding for natural resources and final products, as well as for produced and semi-produced factors of production. Higher (lower) prices reflect greater (lesser) demand and/or greater (lesser) scarcity of a good or service and induce users to conserve (splurge), and/or expand (contract) production. Thus market prices enable individuals—consumers and entrepreneurs — to calculate and to guide production so that the means available tend to be devoted to the most urgent wants, leaving no more urgently felt want unsatisfied. The two requisites for economic calculation are (1) private ownership, not only of consumers' goods but also of factors of production, and (2) a common denominator, money, in which relative values may be expressed. Mises pointed out in 1920 that such calculations would be impossible in a socialist economy — because the two conditions for economic calculation would be lacking. Socialist planners would have to rely on outside prices to determine relative market values. This thesis evoked a lively, and ongoing, debate, as defenders of socialism attempted to refute Mises.

Midterm I: The Calculation Debate 2.0

In his famous article, the data scientist Michael I. Jordan has argued against the popular view that what is going on in artificial intelligence has to do with human-imitative systems - that is, building *in silico* computational learning algorithms that mimic human intelligence. There are others who strongly disagree. In this midterm, you will be asked to consider this modern debate about the nature of artificial intelligence in the context of the socialist calculation debate between Austrian economists (Mises and Hayek) on the one hand and the market socialists (Lange and Lerner) on the other. You will be evaluated on your understanding and application of the principles that we have discussed in our various readings up to this point in the course.

1. In his famous article [1], Michael Jordan has claimed that AI is not fundamentally about developing human-imitative intelligence systems. Jordan believes that instead, what is going on is the emergence of a new engineering discipline that we might call *data science* that is about building human *augmentative intelligence* systems and *intelligent infrastructure* systems. On the other hand, the computer scientists Parkes and Wellman (see [2]) state explicitly:

> *The field of artificial intelligence (AI) strives to build rational agents capable of perceiving the world around them and taking actions to advance specified goals. Put another way, AI researchers aim to construct a synthetic* ***homo economicus****, the mythical perfectly rational agent of neoclassical economics.*

- *(a)* Relate Jordan's position in the AI debate to Mises's and Hayek's position in the calculation debate. How are the arguments similar? How do they differ? Emphasize especially the concepts of knowledge problems, dynamic processes, and institutions as coordinating devices.
- *(b)* In his article, Jordan states that historically other disciplines such as chemistry or biology did not strive to advance their respective fields by designing an artificial agent such as an *artificial chemist* or an *artificial biologist*. Is this true for the field of economics, or could one interpret what Lange and Lerner were trying to do as exactly that: construct an artificial economic planner?

- *(c)* Relate Parkes' and Wellman's project to that of Lange and Lerner. How are the two projects similar? Different? Do any of Mises's or Hayek's arguments to Lange and Lerner apply to Parkes and Wellman? Explain.
- *(d)* Who do you think won the original calcuation debate and why?
- *(e)* Who do you think is ahead in the current AI debate and why?

2. Phelan and Wenzel, two modern Austrians, have entered the AI debate. See [3]. Not surprisingly, they argue that AI hasn't changed anything fundamental that would change the Austrian position since the time of the original calculation debate.

- *(a)* Sketch the argument that Phelan and Wenzel make in some detail. How does it relate to the original debate? What things have changed since the early 20th century that makes the debate worth rehashing? Do you find their argument compelling? How does your view of what AI is influence your position?
- *(b)* Is there anything that you think Phelan and Wenzel get wrong or simply miss or don't understand about modern AI technology? Does this fundamentally change what you think they should conclude? Why or why not?

3. In a somewhat recent special volume of the *Econometrics Journal*, there was a scholarly discussion of the role of machine learning in econometrics. In that special volume, Igami seeks to draw parallels between *machine learning* and *structural econometrics*. See [4]. In particular, he mentions the DeepMind AI *AlphaGo* that was trained to play the game of Go and had remarkable success against human opponents. You are not expected to know all of the technical details in these complex papers, but you are expected to understand the basic issues at hand at a high level.

- *(a)* Please watch the documentary *AlphaGo*. Take a position on whether or not this is a human imitative or human augmentative system. Defend your answer with logic and evidence and using concepts we have learned together in this class.
- *(b)* The Editor of the special volume, John Rust, and his colleagues in their introductory article discuss the essential role of human creativity in econometric model building and claim that AI/ML systems will not replace human econometricians any time soon. See [5]. Likewise Bayesian econometrician John Geweke in his fascinating monograph claims that econometric model building takes place in a setting of unstructured uncertainty and requires human creativity. See [6], chapter 1. What does this mean? What is the

role of human creativity in machine learning and econometrics? How does this relate to Hayekian knowledge problems and the concept of the Austrian entrepreneur? Consider especially James Buchanan's concepts of *natural* and *artifactual* man and *reactive* and *creative* choice. See especially *Natural and Artifacutal Man* in [7] and [8].

- *(c)* One of the lessons of the socialist calculation debate was an understanding, not just of the feasibility of economic planning in the real world, but of the epistemic problem of Walrasian equilibrium theory. Namely, that there is a parallel in the epistemic problems of the economic planner and the economic theorist. Sketch out this argument. Now, consider an extension of this symmetry in the modern AI debate. Sketch out a similar epistemic symmetry between the Austrian entrepreneur engaged in economic calculation and the Geweke-style creative econometrician. How is their knowledge problem similar? How deep does this analogy go and what does this mean for econometric modeling building and practice?

- *(d)* How do you think AI/ML will effect econometric theory? Include in your answer a working definition of what AI/ML is for you.

Part II

Foundations for Fintech

The Basic Theory of Interest

This section will cover basic time value of money calculations. The main references will **Luenberger2013?**.

Principal and Interest

Simple Interest

$$V = (1 + rn)A$$

If the proportional rule holds for fraction years, then after any time t (always measured in years!), the account value is

$$V = (1 + rt)A$$

The account grows *linearly* with time.

Compound Interest

The Seven-Ten Rule *Money invested at 7% per year doubles in approximately 10 years. Also, money invested at 10% per year doubles in approximately 7 years.*

Compounding at Various Intervals

The effective rate is the number r' that satisfies: $1 + r' = [1 + (r/m)]^m$.

Continuous Compounding

$$\lim_{m\to\infty}[1+(r/m)]^m = e^r$$

where $e = 2.7818\ldots$ is the base of the natural logarithm.

The effective annual rate r' is the value satisfying $1 + r' = e^r$

Example: If the nominal interest rate is 8% per year, then with continuous compounding the growth would $e^{0.08} = 1.0833$, and hence the effective interest rate is 8.33%

We can calculate how much an account will have grown after any arbitrary length of time (including fractional periods). We denote time by the variable t, measured in years. Thus $t = 1$ is one year, and $t = 0.25$ is 3 months.

General formula for compouding:

$$[1+(r/m)]^k = [1+(r/m)]^{mt} = [1+(r/m)]^{m^t} \to e^{rt}$$

Debt

If you *borrow* money then the compounding works in exactly the same way but in this case you pay the interest rather than receiving it in the investment case.

Money Markets

- US Treasury bills: 3, 13, 26, or 52 weeks
- US Treasury notes: 2, 3, 5, 7, and 10 years
- US Treasury bonds: 30 years

Present Value

Discount factor:

$$d_k = \frac{1}{[1+(r/m)]^k}$$

The present value of a payment of A to be received k periods in the future is $d_k A$

$$d_k A = \frac{1}{[1+(r/m)]^k} \times A = \frac{A}{[1+(r/m)]^k}$$

Present and Future Values of Streams

Ideal Bank

- Same rate of interest to both deposits and loans
- Has no service charges or fees
- Its interest applies to any size of principal: from 1 cent to $1 million (or more!)

Future Value

Future value of a stream *Given a cash flow stream* $(x_0, x_1, \dots, x_n)$ *and interest rate* r *each period, the future value of the stream is*

$$FV = x_0(1+r)^n + x_1(1+r)^{n-1} + \cdots + x_n.$$

Present Value

Present value of a stream *Given a cash flow stream* $(x_0, x_1, \dots, x_n)$ *and an interest rate* r *per period, the present value of the cash flow stream is*

$$PV = x_0 + \frac{x_1}{1+r} + \frac{x_2}{(1+r)^2} + \cdots + \frac{x_n}{(1+r)^n}.$$

Frequent and Continuous Compounding

Suppose r is the nominal annual interest rate and interest is compounded at m equally spaced periods per year. Cash flows occur initially and at the end of each of n periods, forming a stream $(x_0, x_1, \dots, x_n)$. Then according to the preceding we can state that

$$PV = \sum_{k=0}^{n} x(t_k)e^{-rt_k}$$

If the nominal annual interest rate is compounded continuously and cash flows occur at times $t_0, t_1, \dots, t_k$. Denote the cash flow at time t_k by $x(t_k)$. Then we have

$$PV = \sum_{k=0}^{n} x(t_k)e^{-rt_k} \quad ,$$

which is the continously compounding formula for present value.

Present Value and an Ideal Bank

Main theorem on present value *The cash flow streams* $\boldsymbol{x} = (x_0, x_1, \dots, x_n)$ *and* $\boldsymbol{y} = (y_0, y_1, \dots, y_n)$ *are equivalent for a constant ideal bank with interest rate* r *if and only if the present values of the two streams, evaluated at the bank's interest rate, are equal.*

See the textbook for the proof.

- This is important because it says that present value is the only number needed to fully characterize a cash flow stream (given an ideal bank)
- The stream can be transformed in various ways (even engineered!) but as long as the present value remains the same then the streams are equivalent
- So if you encounter a cash flow stream, it says that all you have to do is to calculate the present value and use the ideal bank to tailor the stream to any shape you desire.
- This enables the very powerful force of risk transfer.

Internal Rate of Return

Internal rate of return *Let* $(x_0, x_1, \dots, x_n)$ *be a cash flow stream. Then the internal rate of return of this stream is a number* r *satisfying the equation*

$$0 = x_0 + \frac{x_1}{(1+r)} + \frac{x_2}{(1+r)^2} + \cdots + \frac{x_n}{(1+r)^n}$$

Equivalently, it is a number r *satisfying* $\frac{1}{(1+r)} = c$ *[that is,* $r = (1/c) - 1$*], where* c *satisfies the polynomial equation*

$$0 = x_0 + x_1 c + x_2 c^2 + \cdots + x_n c^n$$

Main theorem of internal rate of return *Suppose the cash flow stream* $(x_0, x_1, \dots, x_n)$ *has* $x_0 < 0$ *and* $x_k \geq 0$ *for all* k, $k = 1, 2, \dots, n$, *with at least one term being strictly positive. Then there is a unique positive root to the equation* $0 = x_0 + x_1 c + x_2 c^2 + \cdots + x_n c^n$.

Example This example comes from the documentation for the IRR function in Microsoft Excel on its website.

The data for that example are reproduced here for convenience:

Data	Description
-$70,000	Initial cost of a business
$12,000	Net income for the first year
$15,000	Net income for the second year
$18,000	Net income for the third year
$21,000	Net income for the fourth year
$26,000	Net income for the fifth year

The internal rate of return (irr) for this cash flow stream is the rate r which solves

$$0 = -\$70,000 + \frac{\$12,000}{(1+r)} + \frac{\$15,000}{(1+r)^2} + \frac{\$18,000}{(1+r)^3} + \frac{\$21,000}{(1+r)^4} + \frac{\$26,000}{(1+r)^5}$$

We can us the Python module `numpy-financial`'s `irr` function to solve this as follows:

```python
import numpy as np
import numpy_financial as npf

cash_flows = np.array(
            [-70_000.0, 12_000.0, 15_000.0, 18_000.0, 21_000.0, 26_000.0])
the_irr = npf.irr(cash_flows)
print(f"The IRR is: {the_irr : .5f}")
```

```
The IRR is:  0.08663
```

Here is the same result using Microsoft Excel:

	A	B
1	**Data**	**Description**
2	($70,000)	Initial cost of a business
3	$12,000	Net income for the first year
4	$15,000	Net income for the second year
5	$18,000	Net income for the third year
6	$21,000	Net income for the fourth year
7	$26,000	Net income for the fifth year
8		
9	**IRR**	8.663%

Figure 1: Microsoft Excel IRR function

Evaluation Criteria

Net Present Value

Yada

The Probability Calculus

This section will cover probability. The main references will be Chapter 2 of [9].

Introduction

This presentation is chosen to be representative of the typical approach.

Probability

Background

Some definitions:

- *random experiment* – a process or action whose outcome cannot be predicted with certainty
- *sample space* – the set of all outcomes from an experiment
- *random variable* – denoted by uppercase letters, such as X. Technically, a real-valued function defined on the sample space
- *discrete random variable* – can take on values from a finite or countably infinite set of numbers
  - Ex: number of defective parts, number of typographical errors on a page
- *continuous random variable* – can take on values from an interval or real numbers
  - Ex: inter-arrival times of planes at runway, weight of tablets in pharmaceutical production line, voltage of a power plant at different times

Q: which type of random variable are prices?

NB: we cannot list all outcomes from an experiment when we observe a continuous random variable because there are an infinite number of possibilities. But we can specify the interval of values that X can take on.

- Ex: if the r.v. is tensile strength of cement, then the sample space might be $(0, \infty) \quad kg/cm^2$
- *event* – is a subset of outcomes in the sample space.
  - Ex: a piston ring is defective or that the tensile strength of cement is in the range 40 to 50 $\quad kg/cm^2$

Notation

Discrete random variables: letting 1 represent a defective piston ring and letting 0 represent a good piston ring, then the probability of the event that the piston ring is defective is

$$P(X = 1)$$

Continuous random variables: Let X denote the tensile strength of cement. The probability that an observed tensile stength is in the range 40 to 50 $\quad kg/cm^2$ is

$$P(40 \quad kg/cm^2 \leq X \leq 50 \quad kg/cm^2)$$

- *mutually exclusive events* – events that cannot occur simultaneously or jointly
  - Ex: heads and tails of a coin flip

Probability

- *probability* – a measure of the likelihood that some event will occur (or that an observed outcome will take on certain values)
- *probability distribution* – describes the probabilities associated with each possible value for the random variable
- *equally likelihood model* (classical) – for an experiment where each of n possible outcomes is equally likely, then we assign probability mass of $1/n$ to each outcome
  - Ex: flipping a fair coin, tossing a fair die, or randomly selecting a card from a standard deck of cards
- *relative frequency method* (frequentist) – we conduct an experiment n times and record the outcome

$$P(E) = \frac{f}{n}$$

where f denotes the number of experimental outcomes that satisfy event E

- *probability density function* – $f(x)$ for a continuous random variable
- *probability mass function* – $f(x)$ for a discrete random variable

To find the probability that a continuous random variable falls in a particular interval of real numbers, we have to calculate the appropriate area under the curve of $f(x)$.

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

From the text:

> The area under the curve of $f(x)$ between a and b represents the probability that an observed value of the random variable X will assume a value between a and b. This concept is illustrated in Figure 2.1 where the shaded area represents the desired probability.

NB: a valid probability density function should be non-negative, and the total area under the curve must equal 1.

- *cumulative distribution function* – $F(X)$ is defined as the probability that the random variable X assumes a value less than or equal to a given x

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

For a discrete random variable X, that can take on values $x_1, x_2, ...$, the probability mass function is given by

$$f(x_i) = P(X = x_i); \quad i = 1, 2, ...$$

and the cumulative distribution function is

$$F(a) = \sum_{x_i \leq a} f(x_i) \quad i = 1, 2, ...$$

The Axioms of Probability

We let S represent the sample space and E be an event that is a subset of S.

AXIOM 1

The probability of event E must be between 0 and 1:

$$0 \leq P(E) \leq 1$$

AXIOM 2

$$P(S) = 1$$

AXIOM 3

For mutually exclusive events $E_1, E_2, \ldots, E_k$,

$$P(E_1 \cup E_2 \cup \ldots \cup E_k) = \sum_{i=1}^{k} P(E_i)$$

Notes:

- Axiom 1 states that a probability must be between 0 and 1
- Axiom 2 says that an outcome from the experiment must occur
- Axiom 3 enables us to calculate the probability that at least one of the mutually exclusive events $E_1, E_2, \ldots, E_k$ occurs by summing the individual probabilities

Conditional Probability and Independence

Conditional Probability

Conditional probability arises in situations where we need to calculate a probability based on sume partial information concerning the experiment, and we will see that it plays a vital role in supervised learning applications.

The *conditional probability* of event E given event F is defined as follows:

CONDITIONAL PROBABILITY

$$P(E|F) = \frac{P(E \cap F)}{P(F)}; \quad P(F) > 0.$$

Here $P(E \cap F)$ represents the *joint probability* that both E and F occur together, and $P(F)$ is the probability that event F is the probability that event F occurs.

MULTIPLICATION RULE

$$P(E \cap F) = P(F)P(E|F)$$

Independence

If *events* are *independent*, then knowing that one event has occurred does not change our degree of belief or the likelihood that the other event occurs. If *random variables* are independent, then the observed value of one random variable does not affect the observed value of another.

In general, the conditional probability $P(E|F)$ is not equal to $P(E)$. In these cases, the events are called *dependent*.

INDEPENDENT EVENTS

Two events E and F are said to be independent if an only if any of the following are true:

$$P(E \cap F) = P(E)P(F)$$
$$P(E) = P(E|F)$$

NB: if E and F are independent then the multiplication rule becomes

$$P(E \cap F) = P(E)P(F)$$

This can be extended to k events

$$P(E_1 \cap E_2 \cap ... \cap E_k) = \prod_{i=1}^{k} P(E_i)$$

where events E_i and E_j (for all i and j, $i \neq j$) are independent.

Bayes' Theorem

BAYES' THEOREM

$$P(E_i|F) = \frac{P(E_i)P(F|E_i)}{P(E_1)P(F|E_1) + \cdots + P(E_k)P(F|E_k)}$$

Appendix: Notes on Set Theory

Will use capital letters such as A, B, C to denote sets of points. If the elemonts in a set are a_1, a_2, a_3 we will write

$$A = \{a_1, a_2, a_3\}$$

Let S be the set of all elements under consideration. in other words S is the universal set.

Subset

Take two sets A and B. A is a subset of B, or A is contained in B, denoted

$$A \subset B$$

if every element of A is also in B.

The null set or empty set, denoted by $\emptyset$ is the set containing no points. Thus $\emptyset$ is a subset of every set.

We can visualize set's with Venn diagrams.

Venn Diagram for $A \subset B$

- TODO: insert graph here!

Union

Consider now two arbitrary sets A and B. The union of A and B, denoted $A \cup B$ is the set of points in A or B or both. That is, the union of A and B contains all points in at least one of the sets.

We can express this using *set builder notation*:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

Venn Diagram for $A \cup B$

- TODO: insert graph here.

The key word is "or" (meaning A or B or both).

Intersection

The intersection of A and B, denoted by $A \cap B$ is the set of all points in both A and B

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

The key word here is "and"

Venn Diagram for $A \cap B$

- TODO: insert graph here!

Compliment

If A is a subset of S, then the compliment of A, denoted by $\bar{A}$, is the set of all points that are in S, but not in A.

$$\bar{A} = \{x : x \in S \quad | \quad x \notin A\}$$

Venn Diagram for $\bar{A}$

- TODO: insert graph here!

NB: $A \cup \bar{A} = S$

Disjoint

Two sets A and B are said to be mutually exclusive or disjoint if $A \cap B = \emptyset$. That is, mutually exclusive sets have no points in common.

Venn Diagram for $A \cup B = \emptyset$

- TODO: insert graph here!

An Example

Consider, the example of tossing a die.

$$S = \{1, 2, 3, 4, 5, 6\}$$

Let

$$\begin{aligned} A &= \{1, 2\} \\ B &= \{1, 3\} \\ C &= \{2, 4, 6\} \end{aligned}$$

- $A \cup B = \{1, 2, 3\}$ they are NOT disjoint
- $A \cap B = \{1\}$
- $\bar{A} = \{3, 4, 5, 6\}$
- $B \cup C = \emptyset$ are disjoint
- $A \cup \bar{A} = \{1, 2, 3, 4, 5, 6\}$ the Universal set

Conditional Probability in Set Notation

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

Note:

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

Thus

$$P(A \cap B) = P(B \mid A)P(A)$$

and

$$P(A \cap B) = P(A \mid B)P(B)$$

and thus

$$P(A \mid B)P(B) = P(B \mid A)P(A)$$
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

This is ***Bayes' Rule***!

Sampling Theory

This chapter will cover the frequentist approach to probability and statistics. The main reference will be Chapter 3 of **MartinezMartinez2016?**.

Lectures on Statistical Inference

Here are two lectures covering core concepts of statistical inference, complete with examples, a study guide, and a glossary:

Lecture 1: Introduction to Statistical Inference

Statistical inference focuses on drawing conclusions about a population based on information obtained from a sample. It encompasses methods like estimating population parameters, testing hypotheses, and building probability density estimates. The reliability of our inferences hinges on the sample being representative of the population, often achieved through random sampling.

Sample Mean and Sample Variance

Sample Mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sample Variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Sample Standard Deviation

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

Sampling Distributions

The ***sampling distribution*** is the probability distribution for a statistic.

NB: *a statistic is a random variable.*

The Law of Large Numbers

Law of Large Numbers *(LLN) is a theorem that describes the result of performing the same experiment a large number of times. According to the LLN, the average of the results obtained from a large number of trials should be arbitrarily close to the expected value, and will tend to become closer as more trials are performed.*

See the Wikipedia article on the Law of Large Numbers for more details.

Definition

The law which states that the larger a sample, the nearer its mean is to that of the parent population from which the sample is drawn. More formally: for every $\varepsilon > 0$, the probability

$$\{|\bar{Y} - Y| > \varepsilon\} \to 0 \quad \text{as} \quad n \to \infty$$

where n is the sample size, $\bar{Y}$ is the sample mean, and μ is the parent mean.

More rigorous definitions are the following:

For i.i.d sequences of one-dimensional random variables $Y_1, Y_2, ...$, let $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$.

The *weak law of large numbers* states that $\bar{Y}_n$ converges in probability to $\mu = E\{Y_i\}$ if $E\{|Y_i|\} < \infty$.

The *strong law of large numbers* states that $\bar{Y}_n$ converges almost surely to μ if $E\{|Y_i|\} < \infty$.

Both results hold under the more stringent but easily checked condition that $var\{Y_i\} = \sigma^2 < \infty$.

Using Simulation to Check the Law of Large Numbers

We can use simulation to check the Law of Large Numbers. Consider a fair die with six sides and outcomes $Y = \{1, 2, 3, 4, 5, 6\}$, each with $P[Y_i = y] = \frac{1}{6}$. The true mean is

$$\mu = E\{Y\} = \frac{1}{6}[1 + 2 + 3 + 4 + 5 + 6] = 3.5$$

We can verify this in `Python`:

```
import numpy as np
import matplotlib.pyplot as plt

m = 10000
sizes = np.arange(1,m + 1)
means = np.zeros((m,))

for i in range(len(sizes)):
    y = np.random.randint(1,7, size=sizes[i])
    means[i] = y.mean()

plt.plot(means, 'g', lw = 2.5)
plt.grid(True)
plt.title("Simulating the Toss of a Fair Die to Demonstrate the Law of Large Numbers")
plt.xlabel("Sample Size")
plt.ylabel("Estimated Mean")
```

Central Limit Theorem

Let $f(x)$ represent a probability density with finite variance σ^2 and mean μ. Also, let $\bar{X}$ be the sample mean for a random sample of size n drawn from this distribution. For large n, the distribution of $\bar{X}$ is approximately normally distributed with mean μ and variance given by $\frac{\sigma^2}{n}$.

The theorem that states that, if samples of size n are taken from a parent population with mean μ and standard deviation σ, then the distribution of their means will be approximately normal with:

$$\text{Mean} = \mu$$

and

$$\text{Standard deviation} = \frac{\sigma}{\sqrt{n}}$$

As the sample size n increases, this distribution approaches the normal distribution with increasing accuracy. Thus in the limit, as $n \to \infty$, the distribution of the sample means $\to$ Normal, mean μ, standard deviation $\sigma/\sqrt{n}$.

If the parent population is itself normal, the distribution of the sample means will be normal, whatever the sample size. If the parent population is of finite size N, two possibilties arise:

1. If the sampling is carried out with replacement, the theorem stands as stated;
2. If there is no replacement, the standard deviation of the sample mean is:

$$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

The factor $\sqrt{\frac{N-n}{N-1}}$ is called the **finite population correction.**

The central limite theorem provides the basis for much of sampling theory; it can be summed up, as follows. If n is not small, the sampling distribution of the means is approximately normal, has mean $= \mu$ (the parent mean), and has standard deviation $\sigma/\sqrt{n}$ (where σ is the parent standard deviation).

Using Simulation to Check the Central Limit Theorem

We can use simulation to build intuition for the central limit theorem as well. Consider the mean of a sample from an exponential distribution. Recall that the density of the exponential distribution is the following:

$$f(x) = \frac{1}{\theta}e^{-x/\theta}$$

for $\theta > 0$ and $x > 0$.

In Python we can simulate from the exponential distribution as follows:

```
np.random.exponential(size=100)
```

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import gaussian_kde

m = 10000
n = 1000 # start at 10 and move up to 10000

means = np.zeros((m,))
```

```python
for i in range(m):
    x = np.random.exponential(size=n)
    means[i] = x.mean()

density = gaussian_kde(means)
xs = np.linspace(0.5,1.5,200)
density.covariance_factor = lambda : .25
density._compute_covariance()
plt.plot(xs,density(xs), lw = 2)
plt.title("Kernel Densit Plot")
plt.grid(True)
plt.show()
```

Key Terms:

- **Population:** The entire group of objects or individuals we seek information about.
- **Sample:** A subset of the population selected for observation.
- **Parameter:** A numerical characteristic of a population (e.g., mean (μ), standard deviation (σ)).
- **Statistic:** A function of the observed data in a sample, used to estimate a population parameter (e.g., sample mean ($\bar{X}$), sample standard deviation (S)).
- **Sampling Distribution:** The probability distribution of a statistic.

Types of Sampling:

- **Simple Random Sampling:** Each sample of size n has an equal chance of being selected. Can be done with or without replacement.
- **Stratified Random Sampling:** The population is divided into levels (strata), and a random sample is taken from each level.

Estimators and Their Properties:

- **Estimator:** A function that maps the data space to the parameter space, used to estimate population parameters from sample data. It's a statistic and, if the sample is random, a random variable. Denoted as $T(X_1, X_2, \dots, X_n)$.
- **Bias:** The difference between the expected value of an estimator and the true parameter value. An unbiased estimator has zero bias.
- **Mean Squared Error (MSE):** Measures the average squared difference between an estimator and the true parameter value. MSE is the sum of the variance and the squared bias of the estimator.
- **Efficiency:** Compares estimators based on their MSE. A more efficient estimator has a lower MSE.

- **Standard Error:** The standard deviation of the sampling distribution of an estimator. It measures the precision of the estimator.

Study Guide:

1. **Define statistical inference and its primary goals.**
2. **Differentiate between a population and a sample, a parameter and a statistic.**
3. **Describe the importance of sampling distributions in statistical inference.**
4. **Explain the concept of an estimator and discuss its desirable properties: unbiasedness, low MSE, high efficiency, and low standard error.**

Lecture 2: Methods of Parameter Estimation

This lecture introduces common methods for deriving estimators and explores criteria for evaluating their performance:

Maximum Likelihood Estimation (MLE):

- **Likelihood Function:** Represents the probability of observing the given sample data as a function of the unknown parameter(s).
- **Maximum Likelihood Estimator:** The value of the parameter that maximizes the likelihood function, making the observed data most probable. Found by solving the likelihood equation: dL( )/d = 0, where L( ) is the likelihood function.

Method of Moments:

- **Population Moments:** Expected values of powers of the random variable (e.g., mean (= E[X]), variance (^{2} = E[X^2] - ^{2})).
- **Sample Moments:** Analogous to population moments, but calculated from the sample data.
- **Method of Moments Estimation:** Expresses population parameters in terms of population moments and then substitutes them with corresponding sample moments to obtain parameter estimates.

Example: Estimating Normal Distribution Parameters:

For a random sample from a normal distribution with unknown mean ( ) and variance (^{2}):

- **MLE:** The MLE for is the sample mean (X), and the MLE for ^{2} is the sample variance (S^2) multiplied by (n-1)/n to correct for bias.
- **Method of Moments:** Would yield the same estimators as MLE in this case.

Empirical Distribution Function and Quantiles:

- **Empirical Distribution Function (EDF):** A nonparametric estimate of the cumulative distribution function, based on the order statistics of the sample.

- **Quantile:** A value that divides the distribution into segments with specific probabilities. For example, the median (q_{0.5}) divides the distribution into two equal halves.
- **Estimating Quantiles:** Can be done using the EDF, interpolation methods, or other specialized techniques.

Study Guide:

1. **Describe the principles of maximum likelihood estimation and explain how to find MLEs.**
2. **Explain the method of moments estimation and discuss its steps.**
3. **Define the empirical distribution function and its uses.**
4. **Explain the concept of quantiles and describe methods for estimating them from data.**

Glossary:

- **Population:** The complete set of elements under study.
- **Sample:** A subset of the population selected for analysis.
- **Parameter:** A descriptive measure of a population.
- **Statistic:** A descriptive measure of a sample.
- **Estimator:** A function used to estimate a population parameter from a sample.
- **Sampling Distribution:** Probability distribution of a statistic.
- **Bias:** Average difference between an estimator and the true parameter.
- **Mean Squared Error (MSE):** Average squared difference between an estimator and the true parameter.
- **Efficiency:** Compares estimators based on their MSE.
- **Standard Error:** Standard deviation of an estimator's sampling distribution.
- **Likelihood Function:** Probability of observing the data given parameter values.
- **Maximum Likelihood Estimator (MLE):** Parameter value maximizing the likelihood function.
- **Population Moments:** Expected values of powers of a random variable.
- **Sample Moments:** Counterparts to population moments, computed from sample data.
- **Empirical Distribution Function (EDF):** A nonparametric estimate of the CDF.
- **Quantile:** Value dividing a distribution into segments with specific probabilities.

Operational Subjective Statistical Methods

The thesis we propound … is that scientific reasoning is reasoning in accordance with the calculus of probabilities. **HowsonUrbach2006?**

The logic of probability merely formalizes the coherent implications of specified beliefs about events when the beliefs are expressed in a graded scale from 0 (complete denial) to 1 (complete affirmation). This logic is relevant to ***anyone's*** *uncertain knowledge about any unknown observable quantities. The "scientific method" empowers the scientist with precisely the same inferential principles that it empowers the person on the street in the conduct of daily affairs, no more and no less.* [10, p. 13]

Computational Exercises

Software design skills are essential for the practicing data scientist. These exercises are designed to help you develop some basic design skills within the broad context of numerical computing.

You should seek for a proper mix of elegance and efficiency in your programs.

The Chicken Nuggets Problem

This problem is known as the *chicken nuggets* problem (or sometimes the *coins* problem). It goes like this: you walk into Chick Fil-A with an unlimited budget (and appetite!). You can purchase nuggets in boxes of 6, 9, and 20 pieces.

Write a program to tell you the ***highest*** number of nuggets that you ***cannot*** purchase. Re-read that just in case it went past you the first time. The highest number that you cannot get. For example, you can get 15 nuggets by purchasing a box of 6 and a box of 9 nuggets. You can get 18 by purchasing 2 boxes of 9 or 3 boxes of 6. But with no combination of 6, 9, or 20 can you purchase 17 nuggets. What is the highest number that you cannot get?

This simple game will give you experience assembling different bits of `Python` programming to find the solution. The most direct approach and simplest approach will also employ a very simple numerical method called *brute force* search.

Please write your solution starting with the code snippet below.

```python
def main():
    # your code goes here!
    print("Good luck!") # remove this line of code

if __name__ == "__main__":
    main()
```

```
Good luck!
```

Guess My Number

In the book Python Programming for the Absolute Beginner, 3rd Edition the author teaches `Python` through some simple game programming. One of the first games that he shows how to write is the so-called *Guess My Number* game, which is the children's game of guessing someone's secret number (a number between 1 and 100).

First, write a version of the game in which the computer chooses a secret number and the enduser must guess it.

The output of an implementation of this game might look like this:

```
        Welcome to 'Guess My Number'!
        I'm thinking of a number between 1 and 100.
        Try to guess it in as few attempts as possible.

    Take a guess: 50
    Lower ...
    Take a guess: 25
    Lower ...
    Take a guess: 12
    Higher...
    Take a guess: 18
    Lower ...
    Take a guess: 15
    Lower ...
    Take a guess: 13
    You guessed it! The number was 13
    And it only took you 6 tries!


    Press the enter key to exit.
```

Now write a version of the game where you and the computer switch roles! That is right: you think of a number and the computer must guess it in as few attempts as possible. You will need to encode your guessing logic to the program solution.

This might seem like silly game play, but in order to solve the problem you must use an algorithm called *binary search* or the *bisection method* to solve the problem correctly. This is our first attempt at programming a simple algorithm. This algorithm is used often in data analytics applications!

Please use the code cell below to write your solution:

```
def main():
    # your code goes here!
    print("Good luck!") # delete this line of code

if __name__ == "__main__":
    main()
```

```
Good luck!
```

Monte Carlo Simulation of Pi

"Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin." – John von Neumann

Statement of the Problem

The *Monte Carlo Method* is a computer simulation algorithm that works by using *pseudorandom number generators* to mimic real-world randomness. The purpose of this exercise is to give you an opportunity to practice the Monte Carlo method in a simplified setting.

Consider a square with sides of 1 unit and a unit circle contained within the square. It turns out that we can estimate the value of π by simulating many points within the square and counting the proportion that fall within the circle to the total number of points (i.e. those within the square).

Your task in this exercise is write a `Python` script that estimates the value of π by Monte Carlo simulation.

Recall that the area of the circle is given by the following equation:

$$A = \pi r^2$$

Where r is the radius of the circle, which in this case is $\frac{1}{2}$. Thus, the area of the unit cirlce is $\pi \left(\frac{1}{2}\right)^2 = \frac{\pi}{4}$.

Notice that in general the $\frac{\text{area of the cirlce}}{\text{area of the square}} \approx \frac{N_c}{N_T}$ where N_c is the number of simulated points falling inside the circle and N_T is the total number of simulated points.

Then for the unit circle it will be true that:

$$\pi \approx 4 \times \frac{N_c}{N_T}$$

Recall that for the equation for the unit circle is $x^2 + y^2 = 1$. Use this to classify a simulated xy-point as inside or outside the circle. ***Hint:*** simulate two standard uniform draws to form your xy-pairs.

Monte Carlo relies upon a *law of large numbers* argument where the approximation gets more accurate the more data points are simulated. Evaluate your program for repetitions of:

```python
n = [10**i for i in range(2, 8)]
n
```

```
[100, 1000, 10000, 100000, 1000000, 10000000]
```

Produce a table of the results.

Some Interesting Historical Background

The great mathematician Stanislaw Ulam invented the Monte Carlo method while his friend and colleague, John von Neumann was building the ENIAC machine with his team of engineers. For an interesting discussion of the history of the episode see here: The Monte Carlo Algorithm - George Dyson.

See the chapter *10. Monte Carlo* in **Dyson2012?** for a more in-depth accounting.

The Cherry Tootsie Rolls Problem

There are 3 candy machines in front of you that each dispense either chocolate or cherry flavored Tootsie Rolls. Each has a different probability of dispensing chocolate versus cherry flavors. Your absolute favorite flavor is cherry. You can draw one Tootsie Roll at a time from a single candy machine. The dispensing probabilities remain constant over time.

- Outline a strategy to obtain as many cherry flavored Tootsie Rolls as possible for a fixed set of draws.
- This is a tough problem. Use computational and statistical thinking to come up with a strategy. Use your imagination. Be adventurous.
- ***Hint:*** Use Bayes's Rule!

Monte Carlo Simulation of the St. Petersburg Paradox

The *St. Petersburg Paradox* is a very important paradox in economics and decision theory. It is named after the city where Daniel Bernoulli lived when he published the original paper on the subject. The paradoxical nature of the problem stems from the fact that a lottery is offered with infinite expected value but in practice few people are willing to pay more than \$20 for a chance to play the game. Thus the paradox. How can a lottery that has infinite value only be worth a relatively very small amount to most people?!

The rules of the game are as follows: A coin is flipped until a head (H) occurs. If a H appears on the nth flip, the player earns $\$2^n$. Of course, the game has an infinite number of outcomes (the coin might be flipped many many times and never come up H though that is extremely unlikely), but it is easy to write down the first few possibilities:

$$x_1 = \$2, x_2 = \$4, x_3 = \$8, \dots, x_n = \$2^n$$

This can be computed in `Python` as follows:

```python
for i in range(1,8):
    print(f"(i={i}, x=${2.0**i})")
```

```
(i=1, x=$2.0)
(i=2, x=$4.0)
(i=3, x=$8.0)
(i=4, x=$16.0)
(i=5, x=$32.0)
(i=6, x=$64.0)
(i=7, x=$128.0)
```

The probability of getting a H for the first time on the ith flip is $\left(\frac{1}{2}\right)^i$; it is the probability of getting $(i-1)$ tails (T) followed by a H. Hence the corresponding probabilities of the first few values are:

$$p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, \dots, p_n = \frac{1}{2^n}$$

So the expected value of the lottery is given by the following:

$$E(x) = \sum_{i=1}^{\infty} p_i x_i = \sum_{i=1}^{\infty} 2^i(1/2^i)$$

$$= 1 + 1 + 1 + \cdots + 1 + \cdots = \infty$$

Your task is to conduct a Monte Carlo simulation of the game as follows:

- You play a game against the dealer.
- The dealer puts \$2 into the pot.
- For each round of the game a fair coin is tossed.
- If the coin comes up T the dealer doubles the pot.
- If the coin comes up H the game ends and you win the pot.

What is your expected value for playing this game? How much would you be willing to pay to play this game?

Structure your solution to the game such that the player can parameterize the number of repetitions for the simulation as well the entry fee for playing the game. Calculate the average net earnings as well as median earnings across the number of repetitions. Print this output for each run to the console when the simulation is terminated.

See *Chapter 7: Uncertainty* in **Nicholson2012?** for more details.

Monte Carlo Comparison of MLE & MAP Estimators

In this exercise you will compare maximum likelihood estimation to maximum a posteriori estimation in three different scenarios.

1. The Beta-Binomial conjugate model
    - Select a value of θ (the "true" Bernoulli probability of success) that is not "fair"
    - Derive the maximum likelihood estimator
    - Derive the maximum a posteriori estimator
    - Choose an "informative" prior
    - Set $M = 100,000$ and simulate a dataset with $N = 50$
    - Estimate each model and store the estimate
    - Make tables and histograms to compare the sampling distributions of each estimator
    - Describe the results verbally
2. The Gamma-Poisson conjugate model
    - Repeat the above for this model
    - What is the MLE?
    - What is your prior?
3. The Normal-Normal conjugate model
    - Repeat the above for this model
    - Assume a fixed and known value for σ (*is this realistic?*)
    - What is the MLE?
    - What is your prior?

Part III

Computational Methods for Fintech

Option Pricing Theory

... when judged by its ability to explain the empirical data, option pricing theory is the most successful theory not only in finance, but in all of economics. – Stephen Ross

Computational Methods for Option Pricing

*The most important maxim for data analysis to heed, and one which many statisticians seem to have shunned, is this: "Far better an approximate answer to the **right** question, which is often vague, than an **exact** answer to the wrong question, which can always be made precise." Data analysis must progress by approximate answers, at best, since its knowledge of what the problem really is will at best be approximate. It would be a mistake not to face up to this fact, for by denying it, we would deny ourselves the use of a great body of approximate knowledge, as well as failing to maintain alertness to the possible importance in each particular instance of particular ways in which our knowledge is incomplete.* **TukeyAMS1962?**

Turning to the succor of modern computing machines, let us renounce all analytical tools. – Richard Bellman, *Dynamic Programming*

Some might view the reliance on Monte Carlo methods as defective relative to mathematical analysis…. On the contrary, I believe that Monte Carlo methods … can be superior to mathematical analysis…. Using Monte Carlo methods frees the applied statistician to explore a great variety of models with relative ease, and thus statisticians can pursue the scientific goals of matching models to data more effectively and with less algebraic digression than if mathematical analysis were the only tool… Bayesian statistics and Monte Carlo methods are ideally suited to the taks of passing many models over one data set. – **RubinAnnStat1984?**

*That we are thus **uncertain** about the answer to a numerical problem is one of the two central insights of **probabilistic numerics** (PN)…. The second of the central insights of PN is that a numerical algorithm can treated as an **agent**. For our purposes, an agent is an entity able to take actions so as to achieve its goals. These agents make predictions, and then use the predictions to decide how to interact with the environment. Machine learning often aims to build such agent, most explicitly within its subfield of reinforcement learning.* – **HennigOsborneKersting2022?**

Real Options Theory

- Search theory
- Reinforcement learning
- Real options (OHMC)

References

[1] Jordan, Michael I., "Artificial Intelligence—The Revolution Hasn't Happened Yet," *Harvard Data Science Review*, vol. 1, no. 1, 2019.

[2] Parkes, David C. and Wellman, Michael P., "Economic Reasoning and Artificial Intelligence," *Science*, vol. 349, no. 6245, pp. 267–272, 2015.

[3] Phelan, Steven E. and Wenzel, Nikolai G., "Big Data, Quantum Computing, and the Economic Calculation Debate: Will Roasted Cyberpigeons Fly into the Mouths of Comrades?" *Journal of Economic Behavior and Organization*, vol. 206, pp. 172–181, 2023.

[4] Igami, Mitsuru, "Artificial Intelligence as Structural Estimation: Deep Blue, Bonanza, and AlphaGo," *Econometrics Journal*, vol. 23, pp. S1–S24, 2020.

[5] Iskhakov, Fedor, Rust, John, and Schjerning, Bertel, "Machine Learning and Structural Econometrics: Contrasts and Synergies," *Econometrics Journal*, vol. 23, pp. S81–S124, 2020.

[6] J. Geweke, *Complete and Incomplete Econometric Models*. Princeton University Press, 2010.

[7] J. M. Buchanan, *What Should Economists Do?* Liberty Fund, 1979.

[8] Buchanan, James M., "The Domain of Subjective Economics: Between Predictive Science and Moral Philosophy," in *Method, Process, and Austrian Economics: Essays in Honor of Ludwig Von Mises*, Lexington Books, 1982.

[9] Martinez, Wendy L. and Martinez, Angel R., *Computational Statistics Handbook with MATLAB*, Third Edition. Chapman & Hall/CRC, 2015.

[10] F. Lad, *Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction*. John Wiley & Sons, Inc., 1996.