Pomona College

Senior Thesis in Mathematics

Mathematical Growth

Author:
Michael He

Advisor:
Professor Gizem Karaali

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

May 5, 2022

Abstract

Growth is a fundamental process in the natural world. However, there are many different forms of growth when quantified, analyzed, and modeled in mathematics. This dissertation seeks to parse several common models of mathematical growth, including linear, exponential, logarithmic, combinatorial, sigmoid (S-Shaped), and fractal growth processes. We will look at real world examples, as well as examine the outcomes of growth processes.

Contents

1 Introduction **1**

1.1 Contexts 1

1.2 Structure 3

1.3 Progress 5

2 Linear Growth **8**

2.1 Linear Growth 8

2.1.1 Definitions 8

2.1.2 Applications 9

3 Exponential Growth **11**

3.1 Contexts 11

3.2 Exponentiation 12

3.2.1 Introduction 12

3.2.2 History 13

3.2.3 Definition 16

3.2.4 Properties 18

3.2.5 Bases 19

3.2.6 Distribution of Growth Rates 20

3.3 Compounding and Doubling Time 21

3.3.1 History 21

3.3.2 Doubling Time and Rule of 72 22

3.3.3 Connection to X-fold growth 25

3.3.4 Exponential Decay and Half-Life 25

3.4 Applications 26

3.4.1 Inflation 26

3.4.2 Berkshire Hathaway 30

3.4.3 S&P 500 500 Index 33

3.4.4 Relentless Arithmetic 37

3.5 Inflation Data 43

3.6 Berkshire Hathaway Data 44

3.7 Humble Arithmetic Data 46

4 Logarithmic Growth **47**

4.1 Logarithms 47

4.1.1 Definition 47

4.1.2 Natural Logarithm and Properties 49

4.1.3 Graphing 50
4.2 Logarithmic Model 51
4.2.1 Function 51
4.2.2 Applications 53

5 S-Shaped, Sigmoid Growth **54**
5.1 Introduction 54
5.2 Logistics Model 55
5.3 Sigmoid Growth 56
5.4 Notes on Prediction 56
5.4.1 Notable Examples 58
5.4.2 Overuse 58
5.4.3 Reasons for Failure 58

6 Combinatorial Growth **60**
6.1 Premise 60
6.2 Combinatorics 61
6.3 Factorial Growth 63
6.4 Combinatorial Explosion 65
6.5 Economics - Weitzman 65

7 Fractal Growth **72**
7.1 Premise 72

7.2 History 72
7.3 Definition and Contexts 75
7.4 Sierpinski Triangles 77

8 Collective Outcomes **82**
8.1 Premise 82
8.2 Normal and Lognormal 83
8.2.1 Normal Distribution 83
8.2.2 Central Limit Theorem 84
8.2.3 Application 84
8.2.4 Log-Normal 85
8.3 Power Law 86
8.3.1 Introduction 86
8.3.2 Pareto Distribution 87
8.3.3 Zipf's Law 89
8.3.4 Issues 90

A Proof of Central Limit Theorem **94**
A.0.1 What is CLT? 94
A.0.2 Version 95
A.0.3 Relevant Concepts 95
A.0.4 Technical Definition 97

A.1 Background Concepts 97
A.1.1 Mean 97
A.1.2 Variance 98
A.1.3 Moment Generating Functions 98
A.1.4 Standard Normal Distribution 99
A.2 Before The Proof 99
A.2.1 $E[Y]$ and $Var[Y]$ 99
A.2.2 Y^* 99
A.2.3 Change of Variables 99
A.2.4 Moment Generating Function Properties 100
A.3 Proof 100

Chapter 1

Introduction

> A scoffer seeks wisdom and does not find it, But knowledge is easy to him who understands. - Proverbs 14:6

1.1 Contexts

This thesis is the culmination of my intellectual journey so far.

I started college as a humanities major. While I enjoyed those courses, I realized the cycle of reading books and writing essays was not the best use of my time.

While I excelled at math in grade school and high school, calculus was the stopping point. I thought I was done with calculus given how many math courses I already did.

But to balance my curriculum, I took math courses with incredible professors. Going through MATH31H opened my eyes to what math could do. Ideas that seemed unrelated suddenly came together. I realized that math is incredibly diverse. Connections between fields are much more fluid than one assumes. For example, probability theory is directly related to calculus, which doesn't make sense on surface.

Right before the pandemic sent everyone home, I declared to be a math major. Twelve math courses and five semesters later, I am writing this thesis on mathematical growth. Switching major in the middle of my college life and turbo-charging through the electives was a lot of work. It has not been an easy journey by any means, but everything appears easy in hindsight.

This is partly why I'm fascinated by the idea of growth. In hindsight things change so much, but you don't feel that way when they are happening.

My pivot came with a lot of work and frustration, but it actually prompted me to learn more about interesting things on my own. My interests in history and human nature expanded naturally to economics and sociology, as I felt more confident in parsing statistics and technical writing (though often they are too dry to bother). Meanwhile, my background in math led me to appreciate computer science and natural sciences. In addition, spending many hours during lockdown pushed me to dig deeper into music, art, and writing.

And mathematics connected the dots behind the scenes for all these interests.

They converge on the theory of growth in different contexts. On the macro side, growth includes the universe, the biosphere, the economy, society and culture. On the micro side growth includes the human mind, faith, performance and of course, personal growth.

The interesting thing is how we perceive growth in a dualist sense. We are taught and trained to measure growth quantitatively, yet the most striking observations are mostly qualitative. It speaks volumes about the narrative animals us humans are.

A big part of my learning experience is understanding the difference between what we do know and what we don't but feel like we do. Writing this dissertation feels exactly like that.

Understanding growth is important. Growth is present in all walks of life. Growth happens in the world whether we measure it or not, because time exists and growth is simply the result of time interacting with everything else. In a philosophical sense, growth indicates progress, something akin to our civilization's undertaking.

Not to belabor the point, but mathematics carries an inherent beauty and growth is one way to demonstrate that. Of course, understanding growth is also highly practical, as this exposition seeks to address. For example, my journey is also one of growth.

As Charlie Munger says, learn the big ideas from each discipline and find the connections between them to build a "lattice of mental models". He cannot be more correct. There is an intellectual beauty that invigorates the mind when connections are made between seemingly unrelated ideas. The cosmic forces seem to converge. At the same time, the opportunity cost of not doing that is simply too high. There is no reason to leave sound and valuable insights and wisdom on the table. When you consider the value of basic arithmetic in our daily life, it seems too much a price to pay for people who don't understand basic calculations well.

My goal is to uncover the mathematics behind different types of growth and see how they work mathematically, in real life, and with one another to an extent.

1.2 Structure

What is growth? How does it work? Why does it matter? These are the main questions to explore in this dissertation.

While things are so different, the theory of growth somehow connects the rise in GDP per capita to the massive improvement in computing speed. Such tangible growth is relevant to us as much as the intangible growth we often examine with a philosophical lens.

Vaclav Smil defines growth as, "a function of time ... [with] trajectories in countless graphs with time usually plotted on the abscissa (horizontal or x axis) and the growing variable measured on the ordinate (vertical or y axis)." It captures the quantitative aspects succinctly, mainly that of change of variable(s) in comparison to another variable (usually time and usually singular). [22]

I define growth in a simple manner: growth is change over time in quantitative and qualitative manners.

It is possible to trace growth in more than two dimensions at a time, but it is not for the faint of hearts. Human cognition is mostly limited to two-dimensional spaces.

Growth is generally understood with two (time and measurable output). Time is the near-universal component of any growth phenomena and growth function.

Of course, time is not always the measuring stick. One can also measure growth by comparing changes in other such as a two-dimensional plot of height and weight. However, time exists in those variables implicitly as well and deciphering these relationships without considering time are much harder.

We will cover the following six types of growth at different lengths:

1. Linear growth
2. Exponential growth
3. Logarithmic growth
4. Combinatorial growth
5. Sigmoid growth
6. Fractal growth

The one sentence description (growth over time) for each goes like this. Linear growth changes by the same amount per time unit. Exponential growth changes by the same proportion per time unit. Logarithmic growth is the inverse of exponential growth, or one over said proportion. Combinatorial growth changes by an exponentially increasing base and an accelerating growth rate per time unit. Sigmoid growth involves exponential growth, but with an upper bound. Fractal growth changes dimensionality.

We are most familiar with linear growth and exponential growth, but we already don't understand a lot much about exponential growth to do with calculus. The other growth types are also important, though they are much less explained or taught.

In each chapter I will define one growth type, explain how it works and explore its applications. Each chapter includes definitions, properties, and visualizations.

While each chapter is standalone, I also hope readers can gain a cumulative (hence growing) understanding on mathematical growth. If one reads this, finds it interesting, and feels confident to apply a few ideas in their daily life or has an epiphany on how X or Y works, then I will consider my job well done. I wrote this for myself, but I hope others may find this interesting in some way.

After exploring the six types of growth, I will explore some common patterns of growth and their outcomes, which involve statistics and probability theory. This chapter covers things relating to normal distribution and asymmetrical distribution, including important properties, and real world examples. The appendix will include a longer proof.

1.3 Progress

By the nature of the scope, this thesis requires a breadth of literature, ranging from books to articles to even Youtube videos.

Vaclav Smil's book Growth is the foundation of this inquiry. [22] Geoffrey West's book Scale gives a good primer on exponential and logarithmic growth. [26]

For each chapter, there are also primary articles that offer technical details and insights, in addition to the books mentioned above. I will also make use of secondary resources to present real world applications, nuance behind the mathematical models, and everything in-between. To honor the liberal arts fashion, I aim for this thesis to be as multi-disciplinary and generalist and general public facing as possible, while maintaining its firm roots in mathematics and logical writing.

The topic warrants a breadth of literature and sources ranging from books to articles to even YouTube videos and interactive online lessons. As I come across things over time, I will add them organically (yet another demonstration of growth, this time specifically for this project).

Initially I wanted to write about growth due to Vaclav Smil's Growth and Geoffrey West's Scale. The former gives a broad overview of growth in various parts of the real world, from biology to machines to even cities and economies. The latter is a great primer on exponential and logarithmic growth and the math that connects seemingly unrelated phenomena.

Each chapter includes articles that offer technical details and insights in addition to the books mentioned above. I will also make use of secondary resources to present real world applications, nuance behind the mathematical models, and everything in-between. To honor the liberal arts fashion, I aim for this thesis to be as multi-disciplinary and public-facing as possible, while maintaining its firm roots in mathematics and logical writing. Anyone who knows high school math concepts well should have no problem reading this if they skip the more technical portions.

While previous version of the introduction include a roadmap with due dates, it's way past that stage now. Instead, I want to expand on how I proceed to write this dissertation. I read interesting and relevant materials, write out the main ideas, let my mind process the information on the backburner, and fill out the details later. Then it's followed with rewrites, heavy edits, and a repeat of the cycle.

Writing this is another prime example of growth. From the initial stump and not knowing where to start to now cranking words and insights out at a high speed, the last year has been a blessing in my writing career. While academic writing early in college stunted my growth, my natural interests are slowly taking over my writing chops. Both quantity and quality are rapidly improving, surely the best type of growth trajectory!

After finishing this dissertation, I plan to write a series of explainer posts on the mathematics and findings. It won't just stop there, since there is an entire life ahead of graduation. I intend this project to play a role in lifelong learning and a life well lived. Alas, that is the remanents of a former humanities student…

Chapter 2

Linear Growth

> Remember this: Whoever sows sparingly will also reap sparingly, and whoever sows generously will also reap generously. - 2 Corinthians 9:6

2.1 Linear Growth

2.1.1 Definitions

Linear growth is the most intuitive to understand, thus we will only cover it briefly. Usage of linear equations goes back thousands of years, even though proper development in linear algebra didn't happen until the seventeenth century.

The most common form of a linear function is written as:

$$y = mx + b \tag{2.1}$$

where m is the slope, or a constant value per unit of change in x, and b is the intercept (or y-intercept), which is the initial value..

The two parameters m and b are fixed and the two variables x and y are well, variable. They depend on the two parameters. When $x = 0$, $y = mx + b$ becomes $y = b$.

x is known as an input variable and y is known as an output variable. Generally, we change the values of x to get different values of y. Of course, we can revert the process and manipulate values of y to get values of x we want. When y is written as $f(x)$ or something equivalent, y is also known as a function.

We can rewrite this equation as

$$N_t = N_0 + kt \tag{2.2}$$

, where k is the slope, N_0 is the intercept, t is the input variable x and N_t is the output variable y at time t.

This form is less abstract then the usual algebraic form learned in high school algebra courses and connects the four "things" together.

Linear growth is an increase in quantity over time when the rate of change (slope) stays the same and the output grows by the same amount for each unit of time (or any input variable). We can work backwards with linear functions: as long as we know three of the four components, we can easily calculate the missing one.

2.1.2 Applications

There are numerous examples of linear growth. I will only mention two lesser-known but equally intriguing examples.

Linear Regression Linear regression is an important topic in statistics. It examines the relationship between explanatory variable(s) and a response variable. For example, we want to find the relationship between age and height. Simple linear regression has one explanatory variable and one response variable. Multiple linear regression has multiple explanatory variables. Linear regression can show us the connection (correlation) between variables and also give us some prediction power.

Rocks Stalagmites are rocks formed in caves due to drippings from the cave ceilings. They grow in a linear fashion over a long time period, often in units of millennia. [22]

One interesting observation is even if the absolute growth of a stalagmite is the same over time, its growth rate slows down. For example, assume a growth rate of 1 mm/year and an initial height of 1000 mm. This stalagmite will grow 1000 mm in a thousand years to $2000 = 1000 + 1000 \cdot 1$ mm. In the first year, the growth rate is $\frac{1}{1000} = 0.1\%$, but one thousand years later, the growth rate is now $\frac{1}{2000} = 0.05\%$, half of the starting point.

Chapter 3

Exponential Growth

> So Jesus said to them, "Because of your unbelief; for assuredly, I say to you, if you have faith as a mustard seed, you will say to this mountain, 'Move from here to there,' and it will move; and nothing will be impossible for you." - Matthew 17:20

3.1 Contexts

I've known exponential growth for a long time. I heard the quote "compound interest is the eighth wonder of the universe" as a child, though I had o idea what compound interest or x-th wonder meant. When I learned about exponential function in algebra, this quote made a bit more sense.

I didn't grasp the power of exponential growth until high school, when I started to read extensively about investing. Being a Warren Buffett fan, I naturally came across the idea of compounding. I learned about the rule of 72, doubling time, and all the arithmetic tricks for doing these calculations in my head. It is still fun to do.

And that led to my exploration of exponential growth in economics.

China's GDP post-1979 grew eighty-two-fold in 41 years, which comes down to a 11.4% compounded annual growth rate (CAGR). Four decades of sustained growth produced such drastic gains for China, while the US grew at 5.2% during the same time for an eight-fold growth. A little more than double the US growth rate allowed China's GDP to grow 10 times more than the US GDP in terms of proportional changes!

This explains the catching-up of China's GDP to the United States, as noted in the graph below, courtesy of the World Bank. [2]

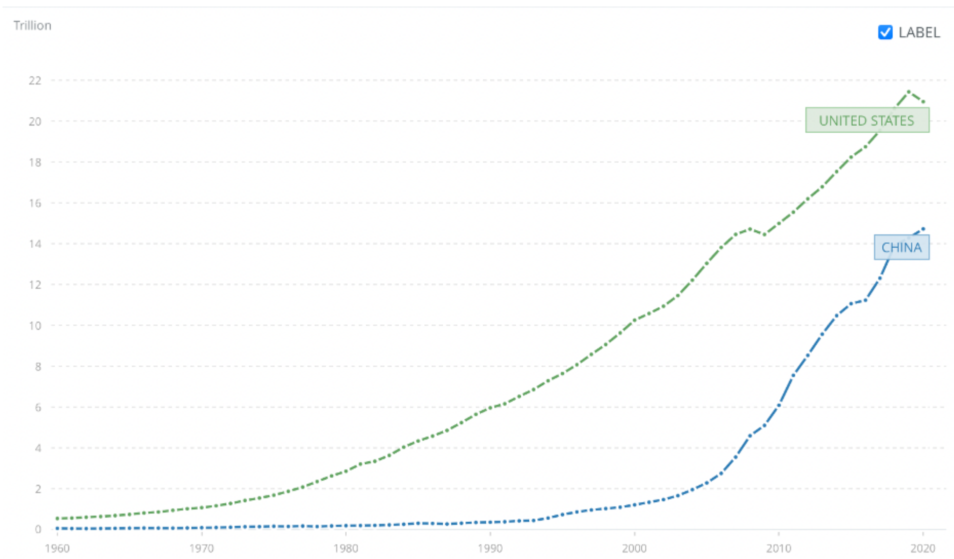


Figure 3.1: GDP growth of China and Unites States, 1960 - 2020

3.2 Exponentiation

3.2.1 Introduction

Exponentiation is an extension of multiplication.

Let's start with the number 2. $2 \cdot 1 = 2$. $2 \cdot 2 = 4$. $2 \cdot 2 \cdot 2 = 8$. $2 \cdot 2 \cdot 2 \cdot 2 = 16$. These results all share a common 2, which we call the base. We can rewrite this with an exponent, or how many times the base multiplies by itself.

For now, let's keep the exponent as a positive integer because negative exponents or non-integer exponents are less straightforward to understand from a multiplication point-of-view.

We write an exponent as a superscript and call the expression "base-b raised to the nth power", n being the exponent's value. Two specific exponential powers are squares for 2^2 and cubes for 2^3. People have used them in geometry for millennia.

Let's rewrite the four numbers. $2 = 2^1 . 4 = 2 \cdot 2 = 2^2 . 8 = 2 \cdot 2 \cdot 2 = 2^3 . 16 = 2 \cdot 2 \cdot 2 \cdot 2 = 2^4$. The pattern is each successive number is twice the earlier number, but its exponent only increases by 1. This is hardly surprising, since this is the definition of an exponent.

Exponentiation is important to calculating and measuring growth for this exact reason. Assume a steady growth rate, when each round yields the same percentage growth, we can shorten the expression with exponentiation.

There are two common forms of writing exponentiation. The first is like the example above, b^n, where b is the base and n is the exponent. One can easily adapt the algebraic notation to other forms such as a^x, b^x, x^n, and so forth.

The other common way is called exponential function, based on Euler's number e. In fact, this is the default form when working with exponential functions. This is commonly written as

$$f(x) = e^x \tag{3.1}$$

.

It can also be written as $f(x) = exp(x)$. Euler's number is $e = exp(1)$.

3.2.2 History

To understand Euler's number, we must look at the history of exponentiation.

Ancient Greeks already made use of exponentiation called amplification, from

Euclid to Archimedes. For the longest time, mathematicians used different notations and names for the concept of exponentiation. German mathematician Michael Stifel coined the word "exponent" in Arithmetica Integra in 1544.

Several mathematicians drove significant progress in the field of exponential functions. [22] In 1683 Jacob Bernoulli found the expression for continuous compound interest to be

$$\lim_{n \to \infty} (1 + \frac{1}{n})^n \tag{3.2}$$

.

This means the value continues to grow as n increases. A plot of such increasing values is below. One can see as n increases, the value gets increasing close to 2.7. Indeed, the approximate value of e will eventually converge to around 2.718.

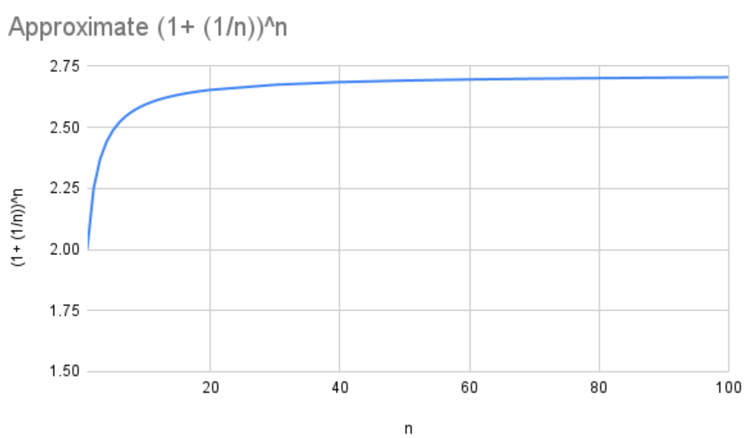


Figure 3.2: An approximation of Bernoulli's expression

Calculus is closely related to exponential functions and the discovery of the number e. Leonhard Euler was instrumental in the development of the exponential function. He first defined the exponential function $e^x = \exp x$ as

$$\lim_{n\to\infty}(1+\frac{x}{n})^n \tag{3.3}$$

.

He also defined Euler's number as

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots \tag{3.4}$$

when $x = 1$ in e^x.

Euler established the precedence for non-integer exponents to exist. He wrote in 1748 that

> consider exponentials or powers in which the exponent itself is a variable. It is clear that quantities of this kind are not algebraic functions, since in those the exponents must be constant."

Euler's number is special for many reasons. It is the base of natural logarithms (subject in the next chapter). It is the limit for $(1+\frac{1}{n})^n$ as n approaches infinity. It is the sum of the infinite series Euler defined above

$$e = \sum_{n=0}^{\infty}\frac{1}{n!} \tag{3.5}$$

.

Euler's number and exponential function is also unique in that its derivative is itself. There are more properties to Euler's number and the exponential function, Eli Maor wrote an entire book about e.

3.2.3 Definition

Exponential growth is an increase in quantity over time when the rate of change is proportional to the quantity itself. Exponential growth in the opposite direction is called exponential decay.

The difference between exponential growth and linear growth is the linear rate of change stays the same in absolute terms and the exponential rate of change stays the same in proportional terms.

There are two exponential growth functions covered in this chapter. The first type is based on discrete variables, or variables that can be counted. An example is a six-face dice with six numbers.

$$N_t = N_0(1 + r)^t \tag{3.6}$$

refers to exponential growth over discrete intervals, such as $t = 1, 2, 3, \ldots$.

For example, a 1000-dollar bank account with a 1% annual interest rate will become 1010 dollars at end of the year. The interest next year is based on the 1010 dollars, instead of the the 1000 dollars from the start. It is also possible to add more to the base during the process, such as additional deposits into a retirement account when the original investment is already growing. The new added amount will now join the growth.

The second type of exponential growth function is based on continuous variables, or those that cannot be easily counted. An example is all real numbers, since you can continue to write $3, 3.01, 3.001, \ldots$

$$N_t = N_0 e^{rt} \tag{3.7}$$

refers to continuous exponential growth.

The difference between the two types of exponential growth is illustrated below by two functions $N_t = 1.1^t$ and $N_t = e^{0.1t}$, both growing at a 10 % rate. As expected, the continuous growth function $N_t = e^{0.1t}$ grows slightly faster than the discrete growth function $N_t = 1.1^t$.

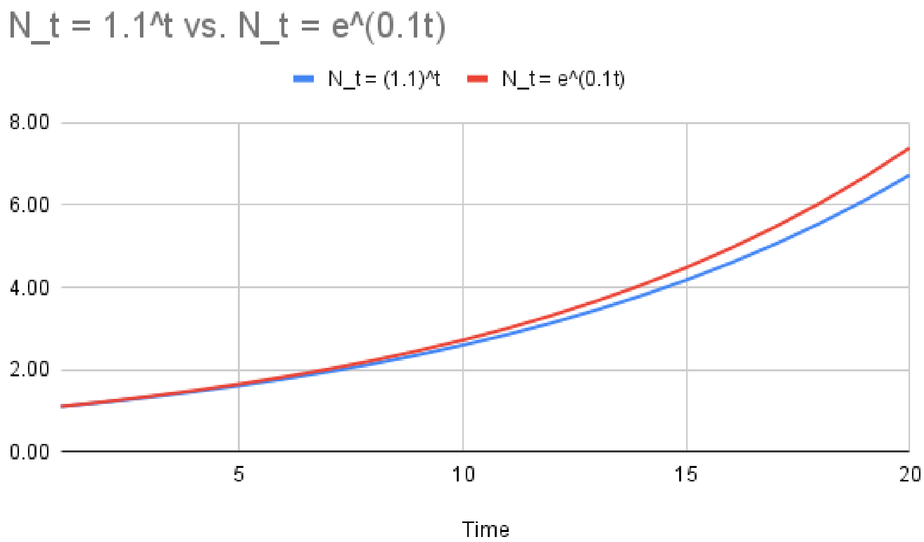


Figure 3.3: Discrete versus continuous exponential growth

There are four parameters for exponential growth functions three input variables and one output variable. Changing any of the inputs can drastically change the output.

The first parameter is the initial quantity N_0, or the input.

The second parameter is time t. Time can be any unit interval, usually ranging from seconds to years.

The third parameter is growth rate r, usually written in decimal form.

We can also calculate $(1+r)^t$ directly to derive "how many times or percent will the initial amount grow to at end of time t?".

The last parameter is end quantity at time t N_t, or the output.

Similar to linear growth, N_t in exponential functions can be calculated both forwards and backwards. To calculate the compounded growth rate, we need to know the other three parameters and vice versa. The most common calculation is for compounded annually growth rate (CAGR), commonly used in business and finance:

$$CAGR = (\frac{N_t}{N_0})^{(1/t)} - 1 \tag{3.8}$$

.

We first derive the amount of growth from beginning period 0 to the ending period t, then raise it to its inverted power, which squares it how ever many times. This leaves us with a multiplier, so to get the growth percentage, we subtract 1 or 100%.

So the calculations for the GDP examples in the beginning of the chapter are as follows: China GDP growth = $(147200/1780)^{(1/41)} - 1 = 11.36\%$ and US GDP growth = $(209400/26270)^{(1/41)} - 1 = 5.19\%$.

3.2.4 Properties

There are several key properties of exponents with base b and exponent power x.

Zero $b^0 = 1$ for any b.

Negative $b^{-x} = \frac{1}{b^x}$ and $b^x = \frac{1}{b^{-x}}$ for for any integer x and non-zero base b.

Product of powers $b^x \cdot b^y = b^{x+y}$. For example $2^3 \cdot 2^5 = 2^{3+5} = 2^8$.

Quotient of powers $\frac{b^x}{b^y} = b^{x-y}$. For example $\frac{3^4}{3^2} = 3^{4-2} = 3^2$.

Power of power $(b^x)^y = b^{x \cdot y} = b^{xy}$. For example $(5^2)^4 = 5^{2 \cdot 4} = 5^8$.

Power of product $(b \cdot a)^x = b^x \cdot a^x$. For example $(7 \cdot 4)^5 = 7^5 \cdot 4^5$.

Power of quotient $(\frac{b}{a})^x = \frac{b^x}{a^x}$. For example $(\frac{7}{2})^3 = \frac{7^3}{2^3}$.

3.2.5 Bases

There are a few widely-used bases in exponentiation.

Powers of 10. In the base-ten or decimal number system used today, powers of 10 are extremely common and written as 1 followed by number of zeroes indicated by the exponent (negative exponents move in the other direction, where zeroes precede the 1). $10^0 = 1.10^1 = 10.10^2 = 100.10^3 = 1000$ and so forth. Inversely, $10^(- 1) = 0.1.10^(- 2) = 0.01$ and so on.

Scientific notation. This is written as 10^x and is used to denote numbers large and small. It's a standard way to standardize numbers from very different sizes. For example, the speed of light is 299792458 m/s, or roughly $3 * 10^8 m/s$.

Orders of magnitude. An order of magnitude is a difference of 1 in the base-10 exponent, or a difference of 10 times. For example, 1000 is an order of magnitude larger than 100.

Powers of 2. The first few negative powers of 2 is common used. $2^{-1} = \frac{1}{2}$ and $2^{-2} = \frac{1}{4}$. and so on. They are also called half, quarter, etc. Base 2 is the foundation to computer science, since 2^n denotes the number of possible values of a n-bit binary number (0 or 1).

Power of 1. Any powers of 1 is one, so $1^n = 1$. The first power of any number is the number itself, so $n^1 = n$.

Power of 0. If exponent n is positive ($n > 0$), then $0^n = 0$. If exponent n is negative ($n < 0$), then 0^n is undefined, since it's the same as $\frac{1}{0^{-n}}$ or $\frac{1}{0}$. The value of 0^0 is under debate.

Power of -1. If n is an even integer, then $(-1)^n = 1$, since $(-1)^2 = 1$ and all even integers are multiples of 2. If n is an odd integer, then $(-1)^n = -1$. This property is often used in induction proofs.

3.2.6 Distribution of Growth Rates

Below is a table of exponential powers for the first ten integers. Each successive cell in each row is a multiple of the previous cell.

n	n^2	n^3	n^4	n^5	n^6
1	1	1	1	1	1
2	4	8	64	512	262144
3	9	27	729	19683	387420489
4	16	64	4096	262144	68719476736
5	25	125	15625	1953125	3814697265625
6	36	216	46656	10077696	101559956668416
7	49	343	117649	40353607	1628413597910450
8	64	512	262144	134217728	18014398509482000
9	81	729	531441	387420489	150094635296999000
10	100	1000	1000000	1000000000	1000000000000000000

Interestingly, each column displays an exponential decay in growth rate. For example, in the column of n^3, the initial growth is from 1 to 8, then to 27. The growth rate decreases from 700% to 237.5% percent. From 27 to 64, the growth rate slows down to 137%.

The obvious reason in this exponential decay is because $1^n = 1$ for all n, while 2 raised to a large power can become very big, so the growth rate from 1 to 2^n is 2^n itself. If we don't account for the guaranteed initial spike and start with base 2, going from 2^n to 2^{n+1} will still demonstrate decelerating growth rate like the plot below.

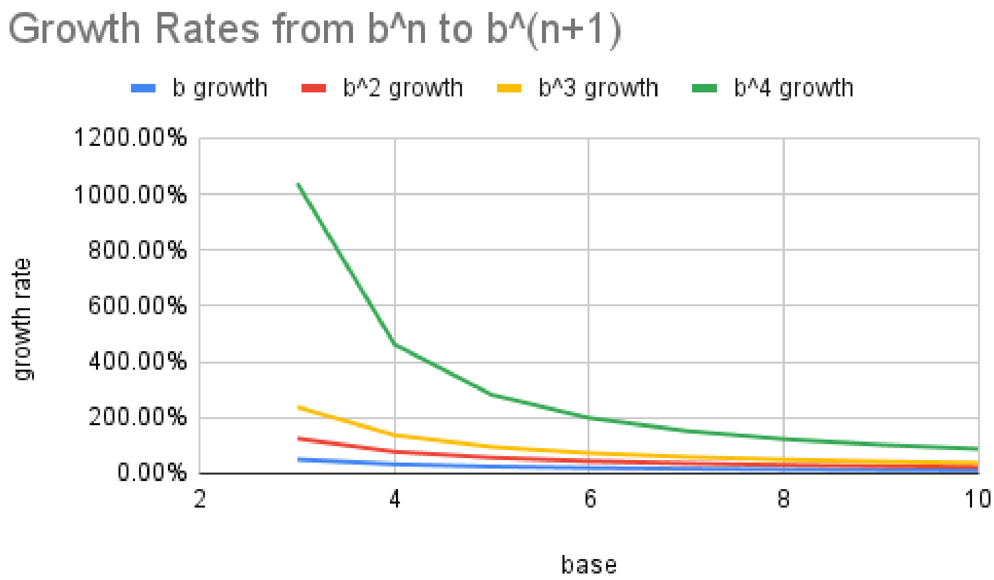


Figure 3.4: Decreasing growth in exponential functions with regards to bases

3.3 Compounding and Doubling Time

3.3.1 History

Compound interest is adding new interest to the combination of existing principal (amount originally borrowed) and existing interest. It is different from a simple interest, which pays the same amount of interest for every time period. The act of adding more than the same amount is called compounding. Compound interest grows exponentially.

There are two types of compound interests, namely discrete $N_t = N_0(1+r)^t$ and the continuous $N_t = N_0 e^{rt}$.

The Florentine merchant Francesco Balducci Pegolotti included a compound interest table in his book Pratica della mercatura in the 14th century. It details the interests on a 100 lire loan from 1 percent to 8 percent in half-percent increments for 20 years. [17]

Mathematician Richard Witt wrote the first book entirely about compound interest called Arithmeticall Questions in 1610. It included tables and clear explanations on how to calculate with different compounding rates and various use cases. [13]

And as mentioned before, mathematician Jacob Bernoulli discovered the number e in 1683 by studying compound interest, which led to the formalization of exponential growth. [22]

3.3.2 Doubling Time and Rule of 72

Doubling time is the amount of time it takes to double in size or quantity. When the growth rate is constant, the growth itself is exponential.

Note variable growth rates can still lead to doubling, but the time needed will be unpredictable, unlike exponential growth.

The Rule of 72 is an arithmetic rule used to estimate doubling time in conjunction with compound interest mechanism. TSuch rules are used out of convenience.

Italian mathematician Luca Pacioli (whose major achievement includes double-entry accounting and introducing Arabic numerals to Europe) mentioned the Rule of 72 in his book Summa de Arithmetica in 1494. He suggested that to estimate the doubling time of an investment, simply divide the interest rate in as percentages by 72 to derive the appropriate time period. Given his lack of detailed explanation, the rule was presumably widely used at the time. [3]

> In wanting to know of any capital, at a given yearly percentage, in how many years it will double adding the interest to the capital, keep as a rule [the number] 72 in mind, which you will always divide by the interest, and what results, in that many years it will be doubled. Example: When the interest is 6 percent per year, I say that one divides 72 by 6; 12 results, and in 12 years the capital will be doubled.

The doubling time t_d is defined as follows:

$$t_d = \frac{ln(2)}{ln(1+r)} \tag{3.9}$$

,

where r is the growth rate in decimal form.

It can be easily calculated from $N_t = N_0(1+r)^t$ and $N_t = N_0 e^{rt}$ using natural logarithms. Since our desired quantity is twice the original amount, N_t will be 2 and N_0 will be 1 in this case.

Set $2 = 1 \cdot (1+r)^{t_d}$. Now $2 = (1+r)^{t_d}$. $ln(2) = t_d \cdot ln(1+r)$. $t_d = \frac{ln(2)}{ln(1+r)}$.

Set $2 = 1 \cdot e^{rt_d}$. Now $ln(2) = r \cdot t_d$. $t_d = \frac{ln(2)}{r}$.

This leads to the rule of 72, because $ln(2) \approx 0.69$. In addition, there is the Rule of 70, Rule of 69.3. The Rule of 72 is most widely used because it shares many divisors (1,2,3,4,6,8,9,12), making it the easiest to do mental arithmetic with, especially when compounding is discrete. Note the answer is a rough estimate, but it is close enough.

The table below compares three simple doubling time rules and an adjustment rule. The most accurate simple rule is in bold.

The Rule of 70 and Rule of 69.3 are more accurate at lower growth rates. The Rule of 72 is the most accurate from 5 percent and onward, but becomes less accurate from the real doubling time at higher growth rates.

Note if the compounding is continuous instead of the discrete, then the Rule of 69.3 and 70 are more accurate than 72 because $ln(2) \approx 0.693$.

The 72-adjusted rule improves the doubling time estimation, which adjusts the value 72 beyond 8% growth rate. For every 3% increase, the value (denominator) increases by 1%.

$$t \approx \frac{72 + (r-8)/3}{r} \tag{3.10}$$

.

This simplifies to $t \approx \frac{208}{3r} + \frac{1}{3}$ where $208/3 \approx 69.3$.

Rate	Actual Years	Rule of 72	Rule of 70	Rule of 69.3	72 adjusted
0.25	277.61	288.00	280.00	**277.20**	277.67
0.50	138.98	144.00	140.00	**138.60**	139.00
0.75	92.77	96.00	93.33	**92.40**	92.78
1.00	69.66	72.00	70.00	**69.30**	69.67
2.00	35.00	36.00	**35.00**	34.65	35.00
3.00	23.45	24.00	**23.33**	23.10	23.44
4.00	17.67	18.00	**17.50**	17.33	17.67
5.00	14.21	**14.40**	14.00	13.86	14.20
6.00	11.90	**12.00**	11.67	11.55	11.89
7.00	10.24	**10.29**	10.00	9.90	10.24
8.00	9.01	**9.00**	8.75	8.66	9.00
9.00	8.04	**8.00**	7.78	7.70	8.04
10.00	7.27	**7.20**	7.00	6.93	7.27
11.00	6.64	**6.55**	6.36	6.30	6.64
12.00	6.12	**6.00**	5.83	5.78	6.11
13.00	5.67	**5.54**	5.38	5.33	5.67
14.00	5.29	**5.14**	5.00	4.95	5.29
15.00	4.96	**4.80**	4.67	4.62	4.96
16.00	4.67	**4.50**	4.38	4.33	4.67
17.00	4.41	**4.24**	4.12	4.08	4.41
18.00	4.19	**4.00**	3.89	3.85	4.19
19.00	3.98	**3.79**	3.68	3.65	3.98
20.00	3.80	**3.60**	3.50	3.47	3.80
25.00	3.11	**2.88**	2.80	2.77	3.11
30.00	2.64	**2.40**	2.33	2.31	2.64
40.00	2.06	**1.80**	1.75	1.73	2.07
50.00	1.71	**1.44**	1.40	1.39	1.72
60.00	1.47	**1.20**	1.17	1.16	1.49
70.00	1.31	**1.03**	1.00	0.99	1.32

3.3.3 Connection to X-fold growth

We can easily derive tripling time, quadrupling time, and so forth. Simply calculate $t = \frac{ln(X)}{ln(1+r)}$ where X is the desired amount (positive non-integers allowed) and r is the growth rate in decimal form.

The table below shows the time to grow X-fold under 1% to 10% growth rate. As the growth rate increases per column, the growth time required goes down naturally. As the end target increases in columns, the growth time required at any specific rate goes up, but that growth rate is decreasing.

rate	2-fold	3-fold	4-fold	5-fold	6-fold	7-fold	8-fold	9-fold	10-fold
1.00	69.66	110.41	139.32	161.75	180.07	195.56	208.98	220.82	231.41
2.00	35.00	55.48	70.01	81.27	90.48	98.27	105.01	110.96	116.28
3.00	23.45	37.17	46.90	54.45	60.62	65.83	70.35	74.33	77.90
4.00	17.67	28.01	35.35	41.04	45.68	49.61	53.02	56.02	58.71
5.00	14.21	22.52	28.41	32.99	36.72	39.88	42.62	45.03	47.19
6.00	11.90	18.85	23.79	27.62	30.75	33.40	35.69	37.71	39.52
7.00	10.24	16.24	20.49	23.79	26.48	28.76	30.73	32.48	34.03
8.00	9.01	14.27	18.01	20.91	23.28	25.28	27.02	28.55	29.92
9.00	8.04	12.75	16.09	18.68	20.79	22.58	24.13	25.50	26.72
10.00	7.27	11.53	14.55	16.89	18.80	20.42	21.82	23.05	24.16
20.00	3.80	6.03	7.60	8.83	9.83	10.67	11.41	12.05	12.63
30.00	2.64	4.19	5.28	6.13	6.83	7.42	7.93	8.37	8.78
40.00	2.06	3.27	4.12	4.78	5.33	5.78	6.18	6.53	6.84
50.00	1.71	2.71	3.42	3.97	4.42	4.80	5.13	5.42	5.68
60.00	1.47	2.34	2.95	3.42	3.81	4.14	4.42	4.67	4.90
70.00	1.31	2.07	2.61	3.03	3.38	3.67	3.92	4.14	4.34

3.3.4 Exponential Decay and Half-Life

Exponential decay is the opposite of exponential growth. Simply put, the amount decreases by the same proportion during every time period.

The formula is

$$N_t = N_0 e^{-rt} \tag{3.11}$$

where $-r$ is the decay rate and assume r is positive.

Exponential decay is widely used in sciences, especially physics and chemistry. A specific type of exponential decay is half-life, or $t_{1/2}$. It is the inverse of doubling time, or the time it takes for the amount to decrease by half.

If we set $0.5 = e^{-rt}$, then $t = \frac{ln(0.5)}{-r}$ is the half-life for decay constant r. r is often written as γ in textbooks, but I write it as r for clarity.

3.4 Applications

Exponential growth can be applied to many areas, including economics, personal finance and more.

3.4.1 Inflation

Inflation is one of the most straightforward applications of exponential growth, as well as one of the most relevant to us.

Inflation is the general increase in prices of goods and services. It is not a stranger, yet we don't often understand the mathematics behind inflation.

Simply put, when the price of something goes up, the same amount of money buys less of that something. Suppose ten dollars can buy ten bananas at one dollar each. In an economy with 10% inflation, each banana now costs 1.1 dollar, so you can only buy nine bananas with the ten dollars. We can say the price has gone up, but we can also say our purchasing power has decreased.

The main metric to measure inflation in the U.S. is the Consumer Price Index (CPI), calculated by the U.S. Bureau of Labor Statistics since 1913. It is a broad metric that explains how fast are prices increasing for various goods and services American consumers usually need.

Each successive year's index divided by the previous year's gives us the inflation in percentage change. The price level of 1981 grows in proportion to that of 1980 and so on.

Below is the inflation rate since 1913, with the detailed table at the end of the chapter. [14]

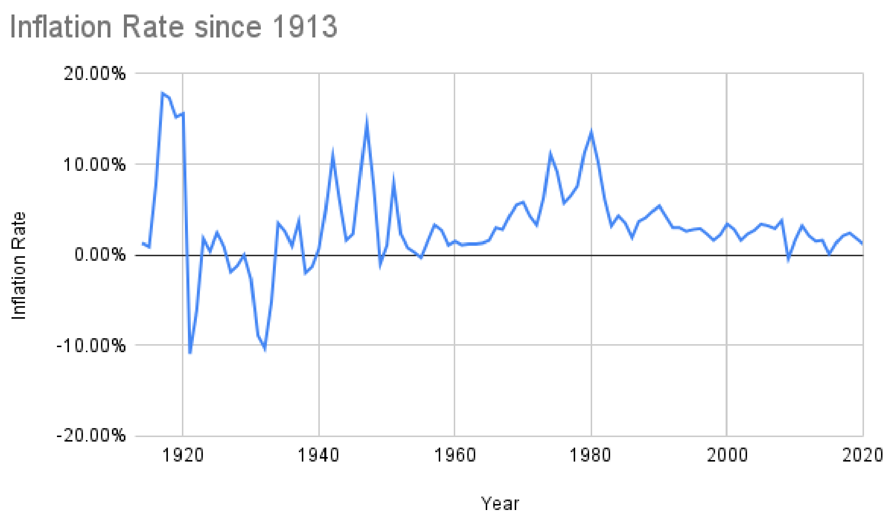


Figure 3.5: Inflation rates from 1913 - 2021

The graph reveals four things. One, inflation rates vary a lot. The spikes indicate huge price increases, which generally mean the economy is not doing so well. Prices change faster than income, so any quick increase spells various issues for the economy.

Two, inflation rates are mostly positive, except the dips between 1920 and 1940 and the tiny -0.4% in 2009, corresponding to the post World War I economic depression, the Great Depression, and the Great Recession of 2008.

Three, given the definition of average inflation rate, some prices have increased a lot more than the average and some have increased slower. For example, healthcare costs has increased much faster by food prices.

Four, even though the inflation rates vary a lot, the underlying index has steadily marched upward with very occasional trending downwards. Volatility in the relative sense does not stop increase in the absolute sense, thanks to arithmetic.

While inflation rate varies from two-digits in the 1970s and early 1980s (an era known as stagflation) to low single-digits today (until 2020 at least), the law of exponential growth means the price level keeps rising. The annual average graph below captures the general rise in price level over the last century, courtesy of the Minneapolis Federal Reserve. [14]

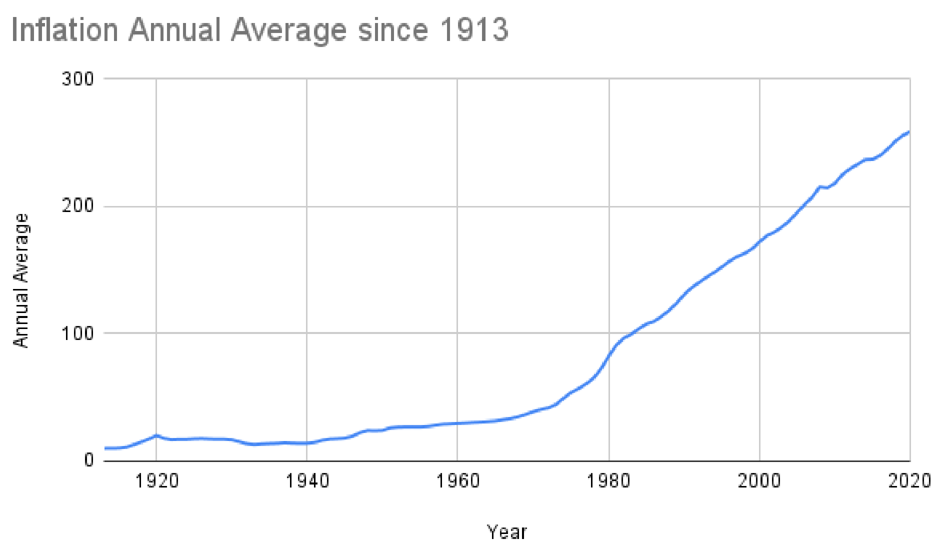


Figure 3.6: Inflation index change 1913 - 2021

From 1913 to 2021 (108 years) the annual average has grown $\frac{271.0}{9.9} = 2737\%$, which is an incredible amount. Yet the compounding annual inflation rate is a mere $(27.37)^{\frac{1}{108}} - 1 = 3.11\%$.

Take a more recent look at the last forty years, the CPI has grown $\frac{271.0}{90.9} = 298\%$. Prices have generally increased three times in four decades, which means some things are more than three times as expensive in 2021 than in 1981. This is all due to a mere $(2.98)^{\frac{1}{40}} - 1 = 2.77\%$ compounded annual inflation rate.

During my life time (circa 1999), the CPI has grown $\frac{271.0}{166.6} = 163\%$ with a $(1.63)^{\frac{1}{22}} - 1 = 2.24\%$ compounded annual inflation rate.

Perhaps a graph is more illustrative of the erosive power of inflation than any words. Below is the declining purchasing power since 1913, calculated using the inverse of inflation rate. It is a classic illustration of exponential decay.

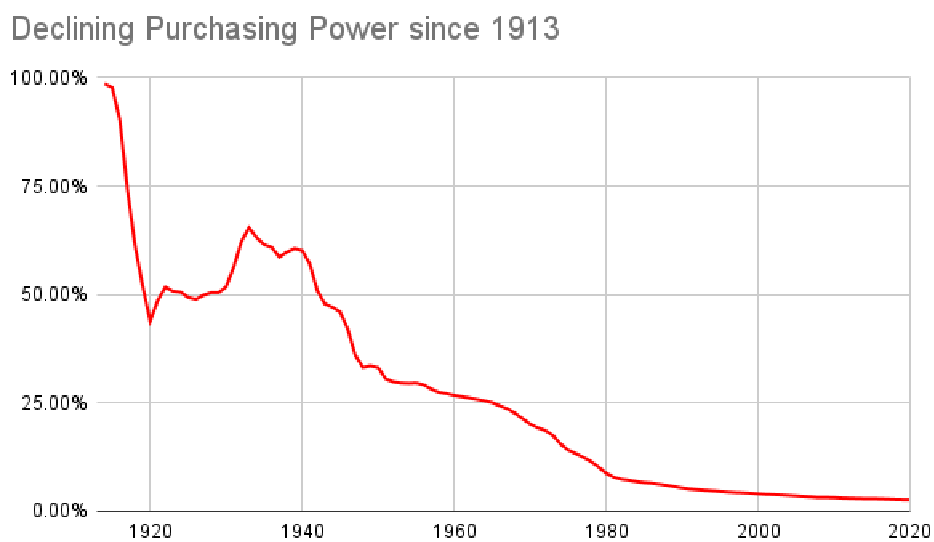


Figure 3.7: Declining purchasing power 1913 - 2021

One dollar from 1913 is only worth 2.56% of its original amount in 2021. One dollar from 1981 is worth 32.32% of its original amount. And during my lifetime, one dollar from 1999 is worth 60.58% of its original amount. That means around 40% of my purchasing power since birth has been robbed by inflation!

Exponential growth is ruthless and we are surely losing money every year due to price increases. Hence, keeping all of our money as cash is actually not a smart financial decision. Having enough savings is important, but we must make good use of extra savings to grow (at least preserve) its purchasing power. This leads to another application of exponential growth in the stock market.

3.4.2 Berkshire Hathaway

Berkshire Hathaway is one of the largest public corporations in the world. It owns a diverse set of businesses such as insurance companies (e.g. GEICO), famous brands (e.g. Fruit of the Loom and Dairy Queen), utilities, large industries (e.g. BNSF railroad), and my favorite See's Candies.

But it is most famous for its CEO and chairman Warren Buffett. Since he took control of the company in the 1960s, Buffett and his business partner vice chairman Charlie Munger have completely transformed the company from a failing textile manufacturer into one of the most successful companies in the world.

Each year Warren Buffett writes a letter to the company shareholders with the first page detailing the performance of the company stock from 1965 to 2021. I updated the stock data up to April 1, 2022. [6]

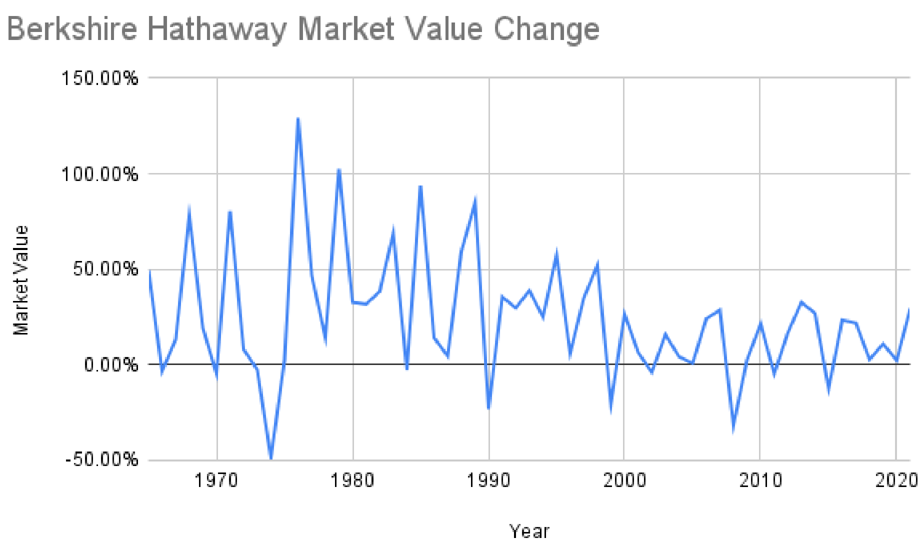


Figure 3.8: Market value change of Berkshire Hathaway 1965 - 2022

Once again, it looks "messy" similar to the inflation annual change graph. However,the y-axis is very different. In the inflation chart the maximum rate is 20%, but here 150% is the max, which means 20% is highly probable. In fact, 29 of the 58 recorded years have a gain more than 20%, 18 years have a mild positive gain, and only 11 years have a negative gain. When you visualize the actual performance of the stock price over the last six decades, the results look staggering.

Note, this performance is in comparison to the S&P 500 index with its annual dividends reinvested, which is a form of exponential growth as well.

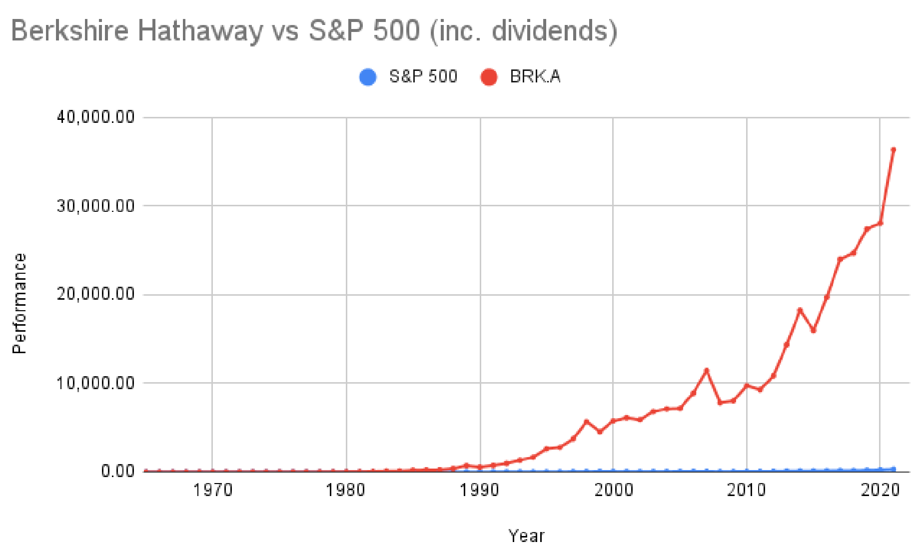


Figure 3.9: Berkshire Hathaway vs S&P 500 index with reinvested dividends

One, the scale of the y-axis once again. To show any sort of growth in this graph, one has to grow more than 1000 times or 1000000%, which is one-tenth of the first major tick at 10000 times or 10000000%.

Two, the blue line, which is the performance of the S&P 500 500 index with dividends invested (meaning it grows faster than the index itself), seems to be flat. In reality, it is not and grew 284-fold! This means one dollar from 1965 will turn into a nice 284 dollars in April 2022, not bad at all. But in comparison to Berkshire's monstrous 41448-fold gain, the S&P 500 500 performance is only $\frac{284}{41448} = 0.68\%$ of that potential growth!

To truly compare the growth, we have to log-transform the y-axis to see that the S&P 500 500 index has also grown quite well (more on logarithmic growth in the next chapter). Each tick is 10 times of the previous one, so the red endpoint is more than 100 times that of the blue endpoints. Once again, this is difficult to intuitively understand due to our perception bias when it comes to exponential and logarithmic growth.

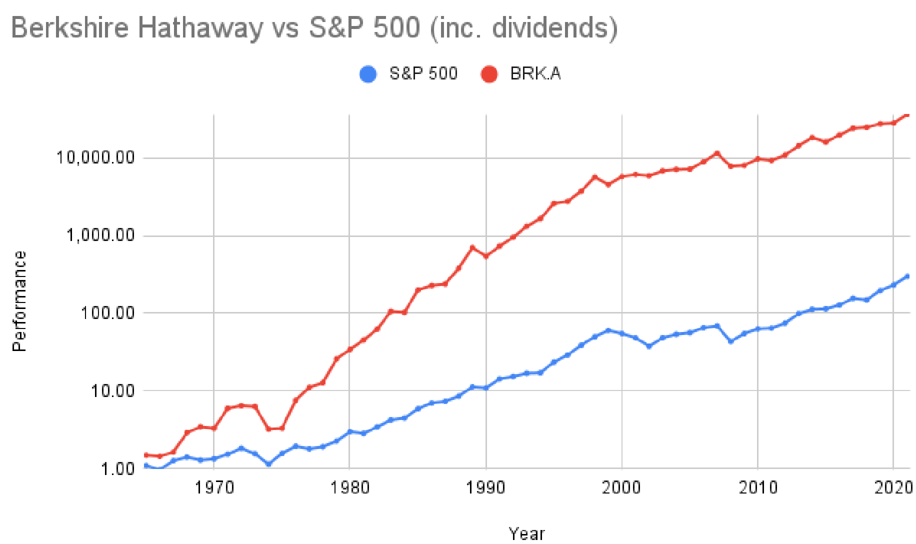


Figure 3.10: Berkshire Hathaway vs S&P 500 index in log

Three, the annual change graph shows rather low performance over the last decade compared to previous decades (often near or above 100% growth), yet after 2010 the stock has still performed well, growing over four-fold. This is all due to high single-digit and low double-digit growth.

Four, following the previous observation, the last ten years have added three times of the gains of the previous four decades. This is a key property of exponential growth. A small relative increase in the later stage is often larger than all of the previous gains combined due to the now incredibly large base.

What is behind this crazy growth from a multi-million dollar company to one worth more than 700 billion today? Mathematically, the power of compounding and exponential growth. The business side, however, is a lot more complicated and deserves much more analysis. Buffett, Munger, and Berkshire employees have made major decisions correctly, not waste money, and diligently re-invest profits into growing the business.

For the Berkshire shareholders who have held company stock for many years, they are probably smiling left to right. However, let's not forget the difficulty to find such a maverick business, which is little to impossible.

3.4.3 S&P 500 500 Index

After discussing the almost miraculous growth of Berkshire Hathaway, now let's look at the growth of the S&P 500 500 index that Berkshire trounced. Turns out, it's not doing too bad either. In fact, consensus says this is the best way for most Americans to invest their savings.

The S&P 500 500 index tracks the performance of America's largest 500 companies. It essentially represents Corporate America as a whole at its most profitable and strongest level.

Indeed, The S&P 500 index is a benchmark metric for the entire U.S. and global stock market. Once again, these ups and downs still produce incredible return over the last five decades.

The following three charts are in similar format to Berkshire Hathaway: annual change, performance over time, and performance log-transformed. We see an EKG-like graph with high volatility.

One can see many more instances of decline in comparison to Berkshire Hathaway. This is expected given the S&P 500 index is comprised of five hundred companies, whose stock prices rise and fall all the time. It is the more frequent declines that stop the compounding from happening. Similarly, the log-transformed graph doesn't say as much as the previous one.

And it is this principle of reversion to the mean that makes the S&P 500 performance pale in comparison to Berkshire Hathaway.

As Matthew 20:16 says:

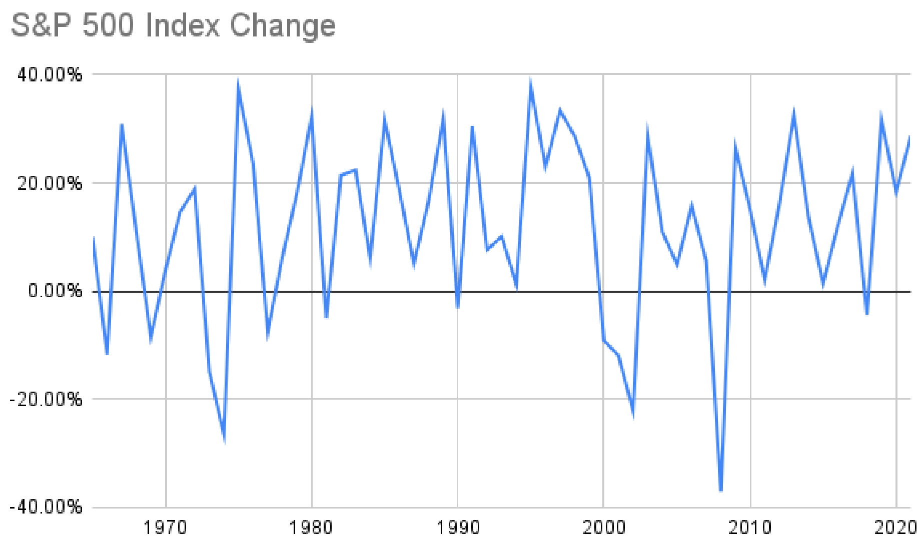


Figure 3.11: Annual change in S&P 500 index 1965 - 2021

Figure 3.12: S&P 500 index 1965 - 2021

Figure 3.13: S&P 500 index 1965 - 2021 in log

"the last shall be first and the first shall be last."

When a good year (positive gain) is more or less offset by negative gains, then the exponential growth cannot really take off.

The following two graphs are the S&P 500 index performances over nine decades, since its existence. It is more dramatic in terms of the log-transformed graph, but in the grand scheme of things such drama fades out.

What appears lackluster may still be surprising. We know that if one invests and re-invests dividends every year, then one dollar one dollar from 1965 will turn into 284 dollars in April 2022. Why is that?

This is once again the power of compounding. Since the dividend rate is positive every year, it actually enhances the index performance. A few percentage points of improvement actually leads to dramatic improvement over a long time period, as the next example will illustrate.

Figure 3.14: S&P 500 index 1920 - 2021 in log

Figure 3.15: S&P 500 index 1920 - 2021 in log

3.4.4 Relentless Arithmetic

Yet it's this exact S&P 500 Index that can help millions of people achieve financial security. Even if its performance isn't that great in comparison to blockbuster companies, it is actually the most manageable and achievable goal. After all, index funds are designed to be held for a long time (e.g. for retirement) and carry little to no hassles and fees.

John Bogle, the founder of Vanguard and a pioneer father of index funds, wrote extensively on the benefits and logic behind index-fund investing. He wrote that the biggest reason behind why indexing works it the ”relentlessness of the humble arithmetic”. In essence, exponential growth!

He makes clear one very important lesson to exponential growth: a few percentage points make a huge difference given enough time. This is the exact answer to the S&P 500 index versus its re-invested version question.

Bogle lists a few major sources of the percentage difference: inflation, lagging performance of active fund managers, fees, and taxes. Let's see how these friction-like forces affect long-term growth. [4]

Assume an initial 1000 dollar investment growing at a steady rate (only achievable in the long-run), how would the results fare each decade?

In the first decade, the results are already different. At a 10% rate, the end result is 2358 dollars, when it is only 1551 dollars for a 5% growth rate. The difference is $\frac{2358-1551}{1551} = 52\%$ for the 5% option! But that is only the start.

At the end of the second decade, the results are truly different. Whereas the 10% rate yields 6116 dollars, the 5% rate only produces 2527 dollars. This is a difference of $\frac{6116-2527}{2527} = 142\%$ for the 5% option!

Notice also how in ten years the amount for 10% more than doubled? This happens to every growth rate greater or equal to 7% due to the rule of 72. In fact, it will keep going for every decade. You can see for the 7% growth rate, every decade it roughly doubles.

Now at the end of the fourth decade (when people enter retirement if they start investing in their twenties), the results couldn't be more different.

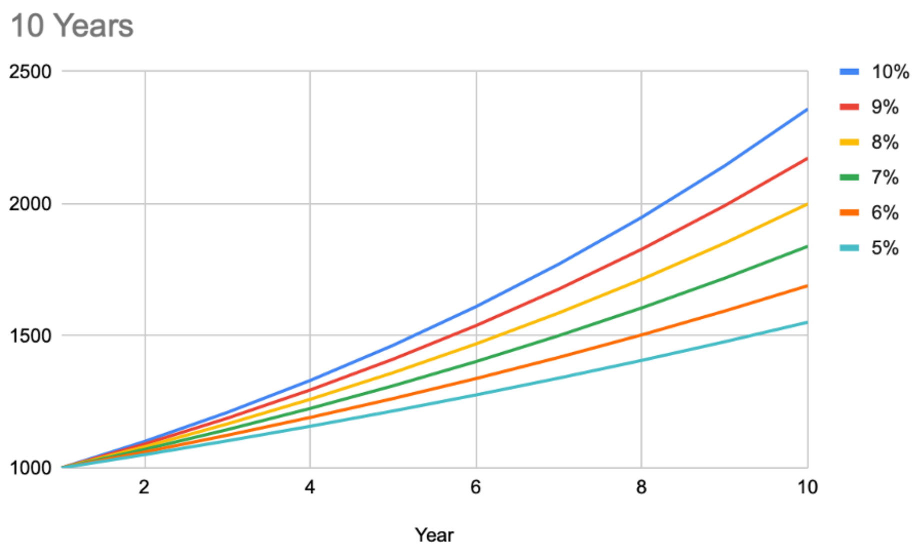


Figure 3.16: 10 year growth results from 5% to 10%

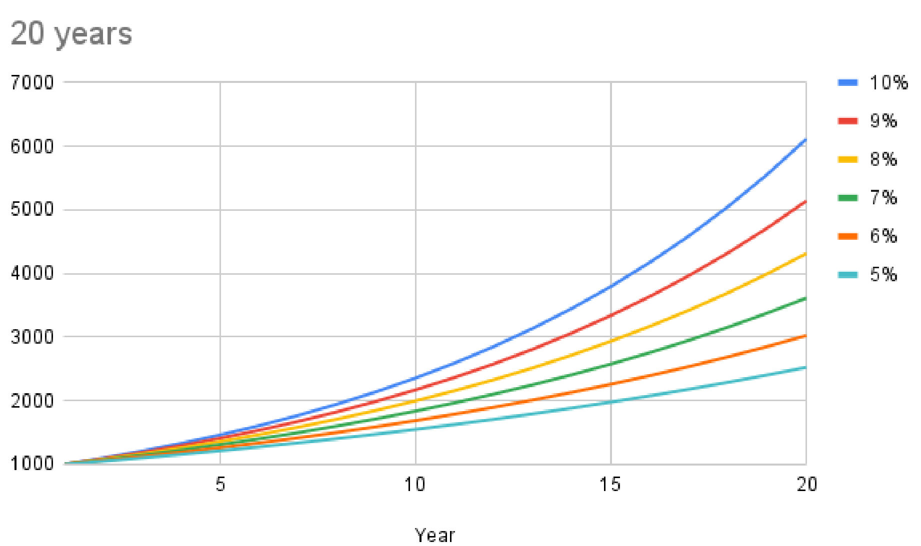


Figure 3.17: 20 year growth results from 5% to 10%

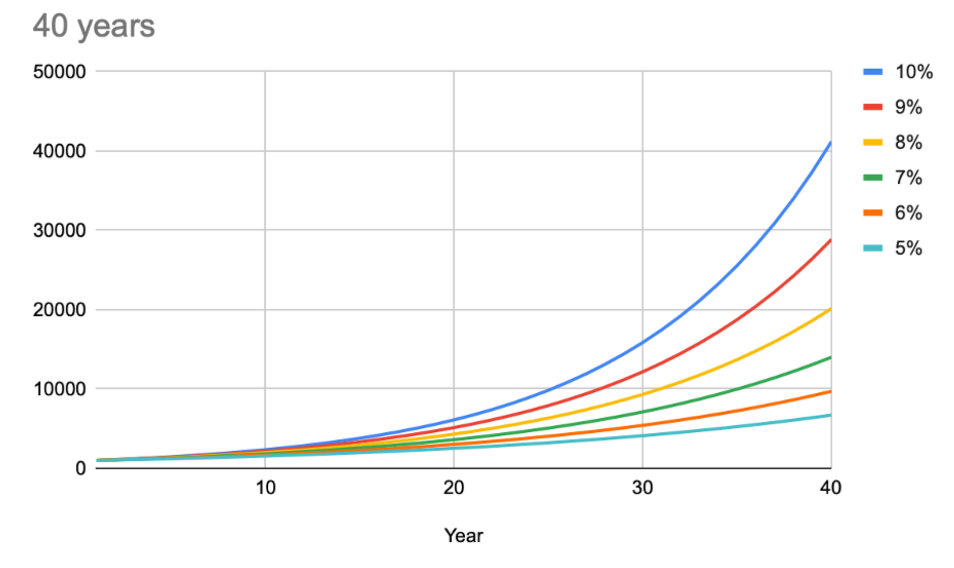


Figure 3.18: 40 year growth results from 5% to 10%

Whereas the 10% rate yields 41144 dollars, the 5% rate only produces 6705 dollars. This is a difference of $\frac{41144-6705}{6705} = 514\%$ for the 5% option!

The 10% option has taken off in comparison to everyone else. It is double that of 8% at the end of forty years. Yes, a mere 2% difference in growth rate can mean twice as much difference in the result. Since the long-term inflation rate since 1999 is slightly above 2%? That's right, this inflation will eat away half of your deserved wealth in forty years, without you doing anything. And that is if inflation can stay low for this entire time, which hasn't always been the case.

There are many more things to unpack here, but given the charts and data, readers can infer.

Now let's look at a more realistic scenario. It is unlikely one will invest once and then wait for decades. Periodic addition is much more likely, for example through one's retirement accounts. Let's calculate the power of performance drag here as well.

Assume an addition of 6000 dollars each year for 40 years, which is the maximum contribution to one's Roth IRA account. This is already a tax-free account, so it's at least one less friction to worry about.

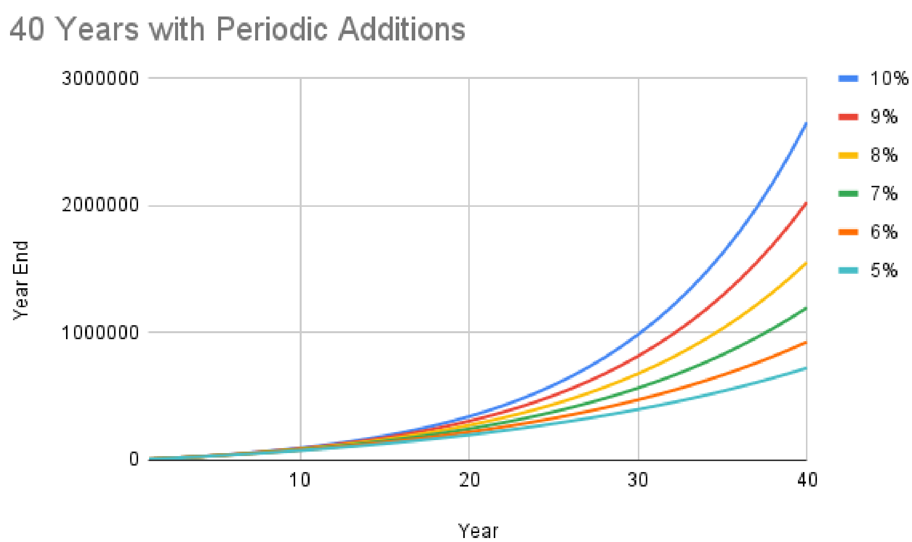


Figure 3.19: 40 year growth results from 5% to 10% with periodic additions

The difference is huge. The 10% scenario yields close to 3 million dollars, whereas the 5% option produces only above 700000 dollars. But it is a four-time difference versus the five-time previously. The gap is smaller this time. The periodic addition gave the 5% option some room to catch up on, but not by much.

However, note the absolute difference between the last two scenarios. Even the losers here are still way larger than the winner before, which implies the importance of consistency. No matter the growth rate, keep adding will increase the chance of greater return, which is only natural.

Another way of examining the effect of compounding drag is to look at the loss ratio, or how much you have "lost" in comparison to the original 10% benchmark.

We can see that even with a 1% friction, after forty years one-quarter of your deserved wealth is gone. For 3%, more than half. For the unfortunate 5%, three-quarters!

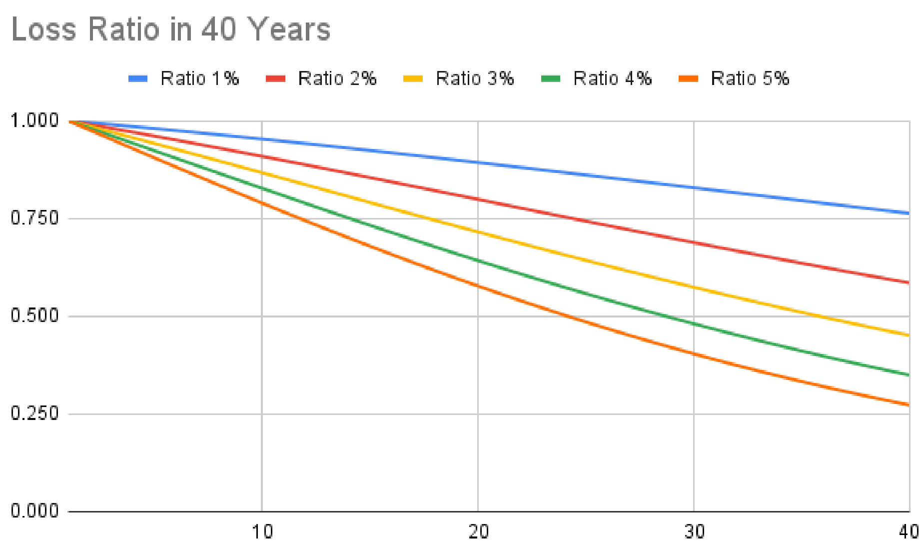


Figure 3.20: Loss ratios in 40 years

If this relatively simple arithmetic is still not enough motivation for young people to start saving and investing as early as possible, then I'm not sure what will.

But not all hope is lost. As shown by the period addition example, even at a lower investment return, consistent effort pays off. The 5% option still returned more than 700000 dollars after forty years, which isn't a trivial amount. And this is only based on a six-thousand-dollar annual addition, which is surely going to increase as one grows older and earns more.

At the same time, the 5% return isn't entirely crazy either, since the long-term stock market index return is around 7% according to Bogle and 10.24% including re-invested dividends. So if we aim for the low bar and do our fair share, the rewards are still bountiful. But if you can improve your annual return even by a percent or two, the end results will be very different. A 19.66% growth rate over 56 years produced a 4144833% return for Berkshire Hathaway.

3.5 Inflation Data

Year	Annual	Inflation
1914	10	1.30
1915	10.1	0.90
1916	10.9	7.70
1917	12.8	17.80
1918	15	17.30
1919	17.3	15.20
1920	20	15.60
1921	17.9	-10.90
1922	16.8	-6.20
1923	17.1	1.80
1924	17.1	0.40
1925	17.5	2.40
1926	17.7	0.90
1927	17.4	-1.90
1928	17.2	-1.20
1929	17.2	0.00
1930	16.7	-2.70
1931	15.2	-8.90
1932	13.6	-10.30
1933	12.9	-5.20
1934	13.4	3.50
1935	13.7	2.60
1936	13.9	1.00
1937	14.4	3.70
1938	14.1	-2.00
1939	13.9	-1.30
1940	14	0.70
1941	14.7	5.10
1942	16.3	10.90
1943	17.3	6.00
1944	17.6	1.60
1945	18	2.30
1946	19.5	8.50
1947	22.3	14.40
1948	24	7.70
1949	23.8	-1.00

Year	Annual	Inflation
1950	24.1	1.10
1951	26	7.90
1952	26.6	2.30
1953	26.8	0.80
1954	26.9	0.30
1955	26.8	-0.30
1956	27.2	1.50
1957	28.1	3.30
1958	28.9	2.70
1959	29.2	1.08
1960	29.6	1.50
1961	29.9	1.10
1962	30.3	1.20
1963	30.6	1.20
1964	31	1.30
1965	31.5	1.60
1966	32.5	3.00
1967	33.4	2.80
1968	34.8	4.30
1969	36.7	5.50
1970	38.8	5.80
1971	40.5	4.30
1972	41.8	3.30
1973	44.4	6.20
1974	49.3	11.10
1975	53.8	9.10
1976	56.9	5.70
1977	60.6	6.50
1978	65.2	7.60
1979	72.6	11.30
1980	82.4	13.50
1981	90.9	10.30
1982	96.5	6.10
1983	99.6	3.20
1984	103.9	4.30
1985	107.6	3.50

Year	Annual	Inflation
1986	109.6	1.90
1987	113.6	3.70
1988	118.3	4.10
1989	124	4.80
1990	130.7	5.40
1991	136.2	4.20
1992	140.3	3.00
1993	144.5	3.00
1994	148.2	2.60
1995	152.4	2.80
1996	156.9	2.90
1997	160.5	2.30
1998	163	1.60
1999	166.6	2.20
2000	172.2	3.40
2001	177.1	2.80
2002	179.9	1.60
2003	184	2.30
2004	188.9	2.70
2005	195.3	3.40
2006	201.6	3.20
2007	207.3	2.90
2008	215.3	3.80
2009	214.5	-0.40
2010	218.1	1.60
2011	224.9	3.20
2012	229.6	2.10
2013	233	1.50
2014	236.7	1.60
2015	237	0.10
2016	240	1.30
2017	245.1	2.10
2018	251.1	2.40
2019	255.7	1.80
2020	258.8	1.20
2021	271.4	4.80

3.6 Berkshire Hathaway Data

Year	Per-Share Value	S&P 500
1965	49.5	10
1966	-3.4	-11.7
1967	13.3	30.9
1968	77.8	11
1969	19.4	-8.4
1970	-4.6	3.9
1971	80.5	14.6
1972	8.1	18.9
1973	-2.5	-14.8
1974	48.7	-26.4
1975	2.5	37.2
1976	129.3	23.6
1977	46.8	-7.4
1978	14.5	6.4
1979	102.5	18.2
1980	32.8	32.3
1981	31.8	-5
1982	38.4	21.4
1983	69	22.4
1984	-2.7	6.1
1985	93.7	31.6
1986	14.2	18.6
1987	4.6	5.1
1988	59.3	16.6
1989	84.6	31.7
1990	-23.1	-3.1
1991	35.6	30.5
1992	29.8	7.6
1993	38.9	10.1
1994	25	1.3

Year	Per-Share Value	S&P 500
1995	57.4	37.6
1996	6.2	23
1997	34.9	33.4
1998	52.2	28.6
1999	-19.9	21
2000	26.6	-9.1
2001	6.5	-11.9
2002	-3.8	-22.1
2003	15.8	28.7
2004	4.3	10.9
2005	0.8	4.9
2006	24.1	15.8
2007	28.7	5.5
2008	-31.8	-37
2009	2.7	26.5
2010	21.4	15.1
2011	-4.7	2.1
2012	16.8	16
2013	32.7	32.4
2014	27	13.7
2015	-12.5	1.4
2016	23.4	12
2017	21.9	21.8
2018	2.8	-4.4
2019	11	31.5
2020	2.4	18.4
2021	29.6	28.7
2022*	14.1	-4.5
CAGR	19.66	10.2
Overall	4,144,800	28,485

3.7 Humble Arithmetic Data

Year	10%	9%	8%	7%	6%	5%
1	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0
2	1100.0	1090.0	1080.0	1070.0	1060.0	1050.0
3	1210.0	1188.1	1166.4	1144.9	1123.6	1102.5
4	1331.0	1295.0	1259.7	1225.0	1191.0	1157.6
5	1464.1	1411.6	1360.5	1310.8	1262.5	1215.5
6	1610.5	1538.6	1469.3	1402.6	1338.2	1276.3
7	1771.6	1677.1	1586.9	1500.7	1418.5	1340.1
8	1948.7	1828.0	1713.8	1605.8	1503.6	1407.1
9	2143.6	1992.6	1850.9	1718.2	1593.8	1477.5
10	2357.9	2171.9	1999.0	1838.5	1689.5	1551.3
11	2593.7	2367.4	2158.9	1967.2	1790.8	1628.9
12	2853.1	2580.4	2331.6	2104.9	1898.3	1710.3
13	3138.4	2812.7	2518.2	2252.2	2012.2	1795.9
14	3452.3	3065.8	2719.6	2409.8	2132.9	1885.6
15	3797.5	3341.7	2937.2	2578.5	2260.9	1979.9
16	4177.2	3642.5	3172.2	2759.0	2396.6	2078.9
17	4595.0	3970.3	3425.9	2952.2	2540.4	2182.9
18	5054.5	4327.6	3700.0	3158.8	2692.8	2292.0
19	5559.9	4717.1	3996.0	3379.9	2854.3	2406.6
20	6115.9	5141.7	4315.7	3616.5	3025.6	2527.0
21	6727.5	5604.4	4661.0	3869.7	3207.1	2653.3
22	7400.2	6108.8	5033.8	4140.6	3399.6	2786.0
23	8140.3	6658.6	5436.5	4430.4	3603.5	2925.3
24	8954.3	7257.9	5871.5	4740.5	3819.7	3071.5
25	9849.7	7911.1	6341.2	5072.4	4048.9	3225.1
26	10834.7	8623.1	6848.5	5427.4	4291.9	3386.4
27	11918.2	9399.2	7396.4	5807.4	4549.4	3555.7
28	13110.0	10245.1	7988.1	6213.9	4822.3	3733.5
29	14421.0	11167.1	8627.1	6648.8	5111.7	3920.1
30	15863.1	12172.2	9317.3	7114.3	5418.4	4116.1
31	17449.4	13267.7	10062.7	7612.3	5743.5	4321.9
32	19194.3	14461.8	10867.7	8145.1	6088.1	4538.0
33	21113.8	15763.3	11737.1	8715.3	6453.4	4764.9
34	23225.2	17182.0	12676.0	9325.3	6840.6	5003.2
35	25547.7	18728.4	13690.1	9978.1	7251.0	5253.3
36	28102.4	20414.0	14785.3	10676.6	7686.1	5516.0
37	30912.7	22251.2	15968.2	11423.9	8147.3	5791.8
38	34003.9	24253.8	17245.6	12223.6	8636.1	6081.4
39	37404.3	26436.7	18625.3	13079.3	9154.3	6385.5
40	41144.8	28816.0	20115.3	13994.8	9703.5	6704.8

Chapter 4

Logarithmic Growth

> Again he said, "What shall we say the kingdom of God is like, or what parable shall we use to describe it? It is like a mustard seed, which is the smallest of all seeds on earth. Yet when planted, it grows and becomes the largest of all garden plants, with such big branches that the birds can perch in its shade. - Mark 4:30 - 32

4.1 Logarithms

4.1.1 Definition

The logarithm is the inverse to exponentiation. It involves two parts, a base b and a given number x. It is written as $\log_b(x)$ with the base specified. We make extensive use of our exponentiation knowledge to understand.

The logarithm of the given number x is how many times base b must be exponentiated to produce x. For example, the logarithm of 10000 with a base 10 is 4 or $\log_{10}(10000) = 4$, since $10^4 = 10000$.

As noted with the previous chapter, there are several common bases to logarithms. The example above is in base-10. In base-2, $\log_2(32) = 5$ since

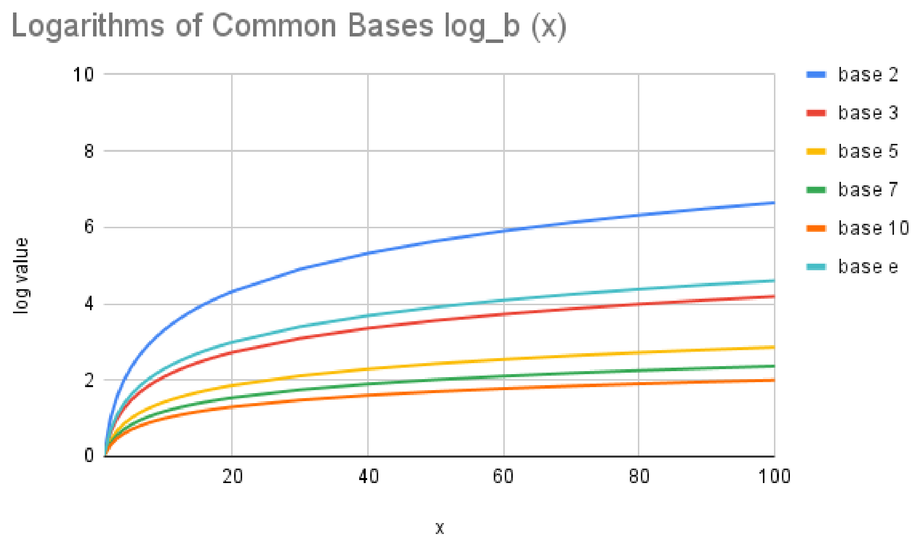


Figure 4.1: Logarithmic growth of common bases

$2^5 = 32$.

Of course we can calculate logarithms of any base b against any number x. Most of the times the logarithm will not be in integer form. For example $\log_2(31) = 4.954$. Non-integer logs reveal that the number x is in between its nearest two integers. For example, we know 31 is in between $\log_2(16) = 4$ and $\log_2(32) = 5$, especially close to 32, hence the 4.954 for log value.

Logs can also be negative. For example, $\log_2(0.5) = -1$ since $2^{-1} = \frac{1}{2}$.

The plot below shows the logarithms of different bases. As one can see, the increase in x is much faster than the increase in the logarithm values. This means logarithm graphs are much better at handling massive increase in values, which exponential graphs are not particularly good at.

John Napier first wrote about logarithms in 1614.The product rule in logarithm is hugely important to the development of logarithm since it reduces computation efforts of multiplications and divisions into additions, subtractions and looking up logarithm tables (a historical artifact that is no longer in wide use).

The concept of logarithm, especially that of the natural logarithm is important to understanding exponential functions in a different perspective.

The natural logarithm was invented before Euler, but he defined it as such:

$$ln(x) = \lim_{n\to\infty} (x^{(1/n)} - 1) \tag{4.1}$$

as n approaches infinity.

He then wrote the inverse of $y = a^x asx = log_a y$.

4.1.2 Natural Logarithm and Properties

Following with the number e, we have the concept of natural logarithm. It is the direct inverse to e^x. The most basic definition of natural log is $\ln e^1 = 1$.

Note that $e^{(\ln x)} = x$ and $\ln e^x = x$. This confirms the inverse relationship between exponential and logarithmic functions.

As with exponentiation, logarithms come with a few properties. Natural logarithm properties work the same as the properties below.

Note for these properties to work they must share a common base b.

Identities For any base b, $\log_b(b) = 1$ and $\log_b(1) = 0$ since $b^1 = b$ and $b^0 = 1$.

Product $\log_b(xy) = log_b(x) + log_b(y)$.

Quotient $\log_b \frac{x}{y} = log_b(x) - log_b(y)$.

Power Rule $\log_b(x^y) = y \log_b(x)$.

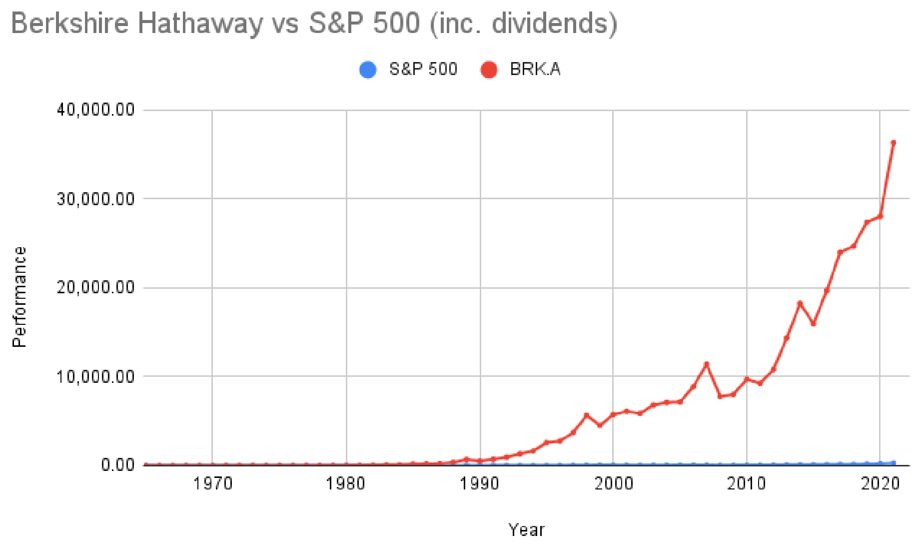


Figure 4.2: Berkshire Hathaway vs S&P 500 index with reinvested dividends

4.1.3 Graphing

As shown in the previous chapter, sometimes we want to use logarithm scales to better visualize and understand massive growth, such as exponential growth.

For example, since Berkshire Hathaway's stock price has increased over forty-thousand-fold since 1964, the more than two-hundred-fold growth of the S & P 500 index fund (with dividends re-invested) appears non-existent but is actually significant in its own way.

Once we adjust the y-axis by logarithm (semi-log), we can compare both returns in a much clearer fashion. Even though the blue line seems closer to the red line, we know in fact there is a 100-time gap between the blue line and the red line after year 2000. This is a caveat to understanding logarithm graph: each tick grows in exponential fashion, not in linear fashion! It looks linear but it is not.

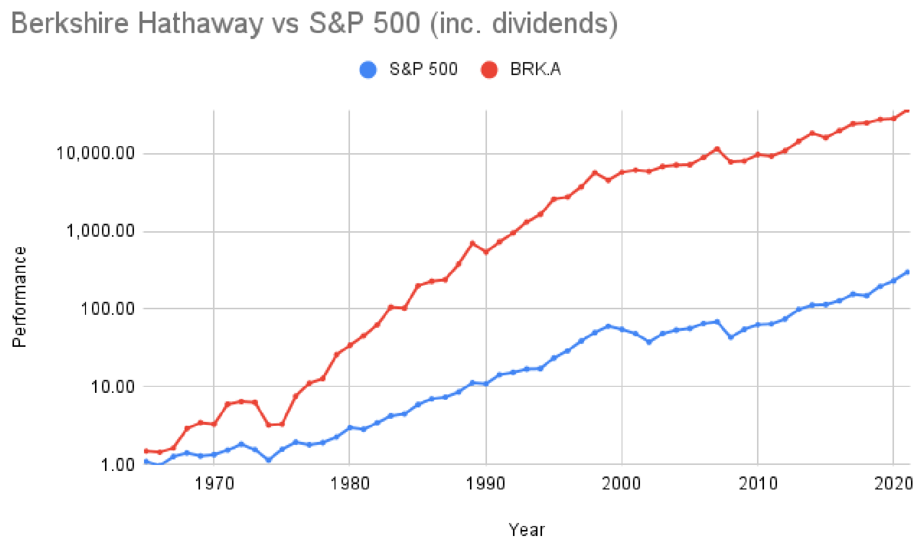


Figure 4.3: Berkshire Hathaway vs S&P 500 index in log

Log-Log versus Semi-log The example above and many data visualizations ask for semi-log plots, which usually condenses exponential growth in y into a more "readable" format while keeping the x intact. However, sometimes both variables are graphed in logarithmic form, known as log-log. The most prominent use case of log-log is in power law, especially Zipf's Law.

4.2 Logarithmic Model

4.2.1 Function

We can write a basic form of logarithmic function as

$$y = a \ln x + b \tag{4.2}$$

, where a is the rate of growth and b is the initial amount.

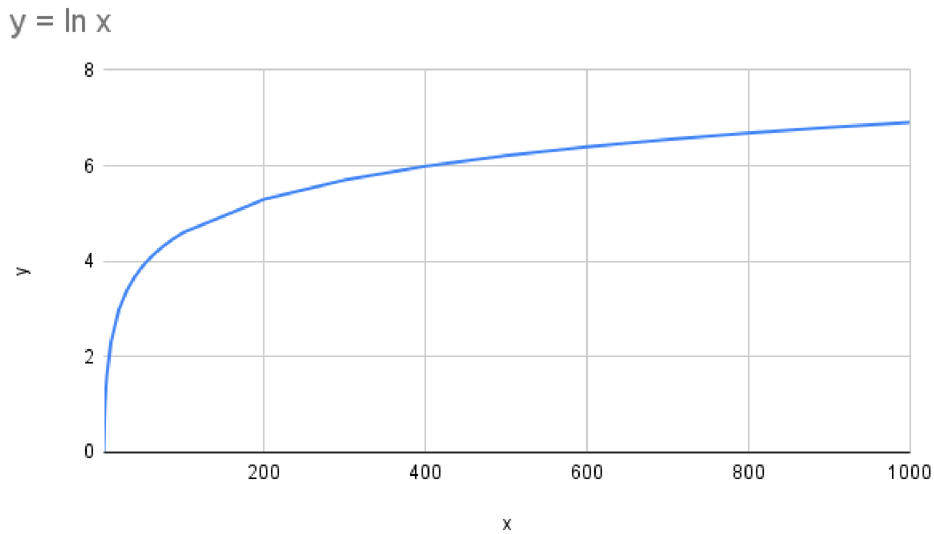


Figure 4.4: Approximation of ln x

Another way of looking at this model is to exponentiate both sides such that $e^y = e^{a \ln x + b} = e^{a \ln x} \cdot e^b = e^{\ln x^a} \cdot e^b = x^a \cdot e^b$. Both a and e^b are constants, so x is the only variable affecting the outcome e^y. Note that if $b = 0$ then $x^a \cdot e$ is essentially a multiplication of two exponential functions.

This model has a few distinct features. First, it has no upper bound, as an increase in x in theory can increase infinitely. Second, it grows fastest in the initial period and gradually slows down.

The graph below is an illustration of $y = \ln x$. As we can see there is a fast initial increase in y as x grows. Indeed, $\ln 1 = 0$, $\ln 2 \approx 0.693$, $\ln 3 \approx 1.099$, $\ln 8 \approx 2.079$, $\ln 20 \approx 2.995$ and $\ln 60 \approx 4.094$. The difference in x that leads to one unit increase in y is rapidly increasing. This leads to a slower growth as x increases, the exact opposite growth pattern to an exponential growth function.

4.2.2 Applications

Log scales are often used to assess various real-world phenomena. [11]

Earthquake The Richter scale is used to measure earthquake intensity. The simplified format is $R = \log intensity$, where R is known as the magnitude and the intensity is a reference number. Every increase of 1 in R means the intensity of the earthquake is 10 times stronger. This is why an earthquake of $R = 7$ is enough to cause huge casualties and $R = 8$ is often disastrous.

pH level The pH level is used to measure acidity of a chemical ranging from 0 to 14. A pH level of 7 means neutral, 0 to 7 fall under acidic range and 7 to 14 fall under basic range. It is written as $pH = -\log[H^+]$, where $[H^+]$ is the concentration of hydrogen ions measured in moles per liter. Due to the negative coefficient, more hydrogen ions result in smaller pH and higher acidity.

Chapter 5

S-Shaped, Sigmoid Growth

> And He has made from one blood every nation of men to dwell on all the face of the earth, and has determined their preappointed times and the boundaries of their dwellings – Acts 17:26

5.1 Introduction

Sigmoid growth is a S-shaped growth curve. Logistic growth is the most well-known form of sigmoid growth. All logistic curves are sigmoid curves, but not all sigmoid curves are logistic curves.

There are three general patterns of all sigmoid curves. One, they have an upper limit. This leads to the second pattern, which is fast initial growth at different rates that eventually slows down.

Three, due to the slowing down and tapering near the system's maximum, we can find an inflection point. This is a point where the growth of growth rate (not growth itself) changes from positive to negative, indicating a decelerating growth rate. Imagine a dashboard of a car. The speed will always be zero or positive, but it can go from 0 mph to 40 mph or slow down from 70 mph to 50 mph, which corresponds to acceleration and deceleration.

In calculus terms, the growth rate is the first derivative and the inflection point measures the derivative of the first derivative, or the second derivative.

Different sigmoid curves grow differently, most notably the locations of inflection points and the accelerating to decelerating phase. Due to that complexity, we will only examine the growth function in a broad perspective. In addition, advanced calculus is necessary to understand the mathematics behind sigmoid growth curves, which is not assumed for all readers.

5.2 Logistics Model

Mathematician Pierre Francois Verhulst came up with the logistic model as a solution to his teacher Adolphe Quetelet's question on population growth. The name of logistic model is based on its French name logistique, not from the word logistics used for military. [22]

The logistic model is written as a differential equation. It must be solved by calculus integration. [20]

The simplest form of a differential equation is written as

$$\frac{dN}{dt} = rN \tag{5.1}$$

where r is the growth rate and N is the amount. This is our old friend exponential equation.

After integration we get the solution

$$N(t) = N_0 e^{rt} \tag{5.2}$$

.

However Verhulst considered a maximal level for a population, since the environment must have a carrying capacity. Thus he adds a carrying capacity K as proportion to the current size N and the logistic model becomes

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}) \quad (5.3)$$

with a solution

$$N(t) = \frac{KN_0}{(K - N_0)e^{-rt} + N_0} \quad (5.4)$$

where N_0 is the initial population size.

There are three parameters to the logistic growth model. We have the initial growth rate r, the carrying capacity K, and the changing rate $\frac{N}{K}$.

A key aspect to the logistic model is its inflection is exactly in the middle of the growth curve.

5.3 Sigmoid Growth

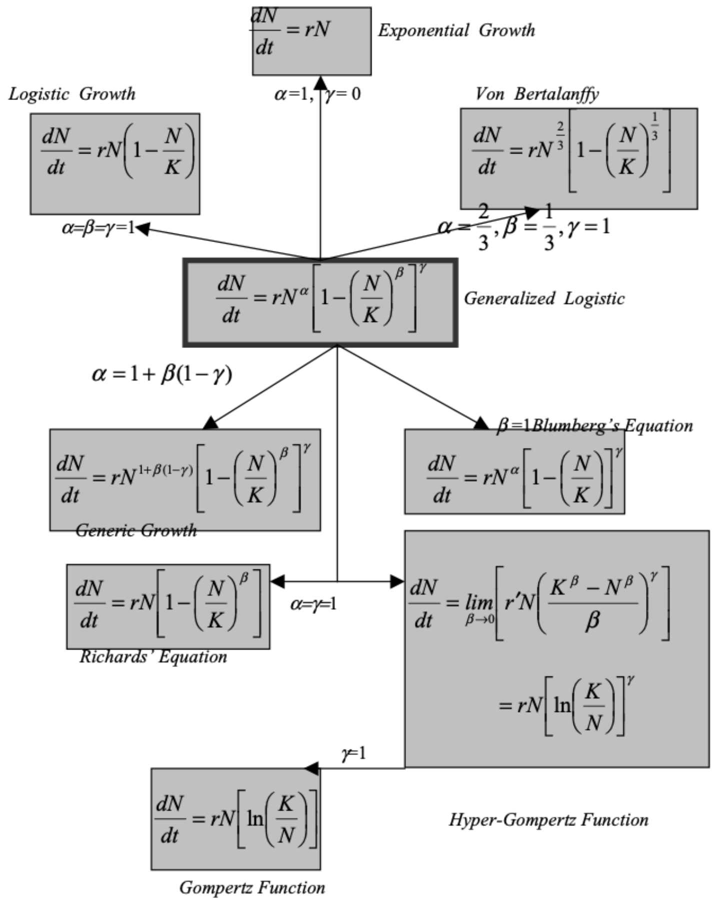
There are many types of sigmoid growth models depending on the specific example. Myhrvold has collected over seventy logistic and sigmoid growth models. Most of them are specific to applied fields, given the specific parameter tuning. Since most models are specific to circumstances and very complex in nature, we will skip over them collectively. [15]

Tsoularis in his 2002 paper analyzed major types of logistic models to present a more "generalizable" equation. Key parameters can be fine tuned to fit the growth and the general graph. [23]

5.4 Notes on Prediction

As much as I enjoy writing about the accuracy and fascinating nature of logistic models, I find the implications of over-using and misusing the growth patterns much more important.

Figure 5.1: A compilation of common sigmoid models



Even though logistic growth and sigmoid curves can quite accurately describe many growth patterns, we must be cautious with applying this type of growth model for prediction and estimation.

There are two ways to consistently produce biased and misleading predictions. It is due to either predicting the inflection point too soon, leading to predicting slowing down while the end result can be higher, or by underestimating the eventual upper limit. Combined together, these two forces often dramatically underestimate the actual results. [20]

5.4.1 Notable Examples

There are several notable examples of failures in using sigmoid growth curves, especially that of the logistic growth model.

For decades people have tried to predict peak oil, or when crude oil production will reach its maximum before its inevitable decline, first due to oil shortage and now due to the transition from fossil fuel to renewable energy sources. A famous 1956 prediction said US oil production would peak around 1970, but that has simply not been true even as predictions kept refining. [20]

Another spectacular failure in prediction is in global car ownership. The 1990 prediction was that global car ownership would saturate around 475 million cars, but more than one billion cars were registered by 2017, more than twice the original prediction. [22]

5.4.2 Overuse

A really big problem is in being over confident in this model's predictability. As Sandberg writes, "We cannot forecast a system with the same confidence with which we can backcast the system". [20]

Early on scientists like Pearl was overly confident and overused logistic models on nearly everything he could find. It was a new concept at the time and took a few decades to prove things wrong. Hindsight is always clear but foresight is much harder. However, by then the habits are already entrenched and an entire generation of scientists were educated by this method and now understanding the model's limitation becomes harder to teach. [22]

5.4.3 Reasons for Failure

Why doesn't prediction work?

Constance Crozier notes it is incredibly hard to predict the inflection point with earlier data. However even with more data points, the correct curve is still very difficult to find until the curve starts tapering near the end. [7]

The real-world limit or carrying capacity is shaped by physical constraints, such as physics, biology, chemistry, etc. However, technology changes the fundamental nature of things by improving efficiencies on various fronts.

When predicting future population, none of the forecasters can account for the dramatic increase of crop yields due to fertilizers, better seeds, more efficiency farming techniques, mechanization of agriculture, and various technologies that constitute the Green Revolution.

In terms of oil production, more oilfields are being uncovered all the time. The obvious ones in Saudi Arabia are being supplanted by deep ocean sources and also hidden ones in oil states such as Texas and Saudi Arabia itself. Oilfield discovery technology makes that possible. At the same time, machines are becoming more fuel efficient, so despite a massive increase in energy consumption, the rate of depletion (as a percentage) based on efficiency is actually slowing down.

Another interesting example on Mozart's output. Some claim that by using logistic growth curve, Mozart actually produced 90 percent of his best works when he died at a young age of thirty-five. However, Smil found multiple growth curves that equally fit historical data without depending on logistic curve parameters. This seems to imply that multiple forms of models can fit the same dataset with vastly different results and interpretations. [22]

Hence this leads to another key insight. The more important thing is not interpreting the results but knowing what to choose for the analysis and why. It's understanding how these growth curves work at its most foundational level. Extending the logic, perhaps the biggest insight about studying sigmoid growth curves is understanding the limitations and tradeoffs involved in choosing each type of growth model.

Chapter 6

Combinatorial Growth

> He told them another parable: "The kingdom of heaven is like a mustard seed, which a man took and planted in his field. Though it is the smallest of all seeds, yet when it grows, it is the largest of garden plants and becomes a tree, so that the birds come and perch in its branches. – Matthew 13:31-32

6.1 Premise

My first encounter with combinatorial growth is an article Professor Karaali shared with me called "Exponential Growth Isn't Cool. Combinatorial Growth Is." It is an application of combinatorial growth (also known as combinatorial explosion) to the digital technology industry. It also coincided with a question I had when learning combinatorics: if things grow like factorials, how does the underlying dynamics work? [1]

An interesting application of combinatorics is in economics. Economists are using combinatorics retrospectively to explain the mathematics behind the Industrial Revolution and its explosive growth, when situated in a long time-horizon (e.g. two thousand years).

In this chapter we will explore each concept starting with an introduction to basic combinatorics.

6.2 Combinatorics

Combinatorics is difficult to define, so let's get to the point. Its main purpose to count is demonstrated throughout history.

Combinatorics comes with several basic rules and definitions. [24]

Rule of sum If there are $n(A)$ ways to do something in set A and $n(B)$ ways to do something in set B and they are distinct from each other, then there are $n(A) + n(B)$ ways to do A or B.

Example You want to order one food item. A taco truck gives you 15 different options and a Chinese takeout restaurant gives you 25 completely different options. There are no overlaps, so there are 15 + 25 = 40 total items to choose your one item.

Rule of product If there are $n(A)$ ways to do something in set A and $n(B)$ ways to do something in set B and doing A and B is independent (whatever you choose for A won't affect the number of choices for B), then there are $n(A) \cdot n(B)$ ways to do A and B.

Example There are 4 types of burgers (regular, cheeseburger, double cheeseburger, BBQ burger) and 3 types of drinks (water, cola, lime soda) to choose at the local burger shop. You need to create a combo with the 4 burger options and 3 drink options. There are $4 \cdot 3 = 12$ total ways to create a unique combo.

Inclusion-Exclusion Principle To avoid double-counting (or multi-counting), we need to subtract overlaps. Suppose there are two sets A_1 and A_2, each

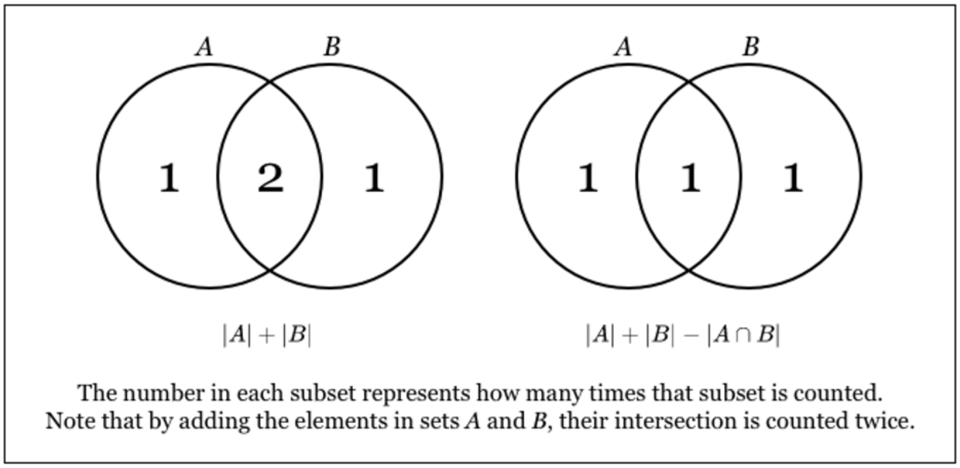


Figure 6.1: Principle of Inclusion and Exclusion

with $|A_1|$ and $|A_2|$ ways of counting (with some ways overlapping). If we want to count every combination between the two sets, we must subtract the overlapping ways. Hence $|A_1 \cap A_2| = |A_1| + |A_2| - |A_1 \cup A_2|$. Note that as the number of sets increase, the principle's complexity increases.

Example How many numbers between 1 and 100 are divisible by 3 or 4? There are $\frac{100}{3} \approx 33$ numbers divisible by 3 and $\frac{100}{4} = 25$ numbers divisible by 4. But some of these numbers are both divisible by 3 and 4 (e.g. 12, 24, etc.), so we must consider numbers divisible by $3 \cdot 4 = 12$, or $\frac{100}{12} \approx 8$. There are $33 + 25 - 8 = 50$ numbers divisible by 3 or 4.

Also consider this Venn Diagram, which illustrates the application of the principle. [5]

Factorials A factorial is defined as $n!$ (where n is a non-negative integer). It is the product of all positive integers less or equal to n or the product of n and its next smallest factorial $(n-1)!$. The math notation is $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 = n \cdot (n-1)!$.

Example $5! = 54 \cdot 3 \cdot 2 \cdot 1 = 120$.

Permutation Permutation is each arrangement of things with a specific order. The things come from a specific set and the order matters. The most common permutation is n permute k, or $nP_k = \frac{n!}{(n-k)!}$ ways to order k things out of n unique things.

Example A set includes 4 letters A, B, C, and D and we need to make a word of 2 letters, including any gibberish. Since the spelling matters, we use permutation and there are $\frac{4!}{(4-2)!} = \frac{24}{2} = 12$ ways to make a 2-letter word. There are 4 options to assign the first letter and 3 leftover options for the second letter, hence $4 \cdot 3 = 12$.

Combination Combination is the ways to combine k things out of n total possibilities without considering order. The most common combination is n choose k, or $nCk = \frac{n!}{(n-k)!k!}$ ways to combine k things out of n unique total things.

Example Suppose we need to mix and match 2 unique ice cream flavors out of 20 total possible flavors at the local ice cream shop (e.g. Bert and Rocky's). We can choose $\frac{20!}{(20-2)!2!} = \frac{20!}{18!2!} = \frac{20 \cdot 19}{2} = 190$ combinations. There are 20 options for the first scoop and 19 leftover options for the second scoop, but getting chocolate and then vanilla is no different than getting vanilla and then chocolate, hence we divide the repeats to get $20 \cdot 19/2 = 190$.

6.3 Factorial Growth

A simple exploration of factorials and its connection to exponential growth reveals some interesting insights.

By calculating the values of factorial $n!$ and the corresponding 2^n, we can now calculate the ratio $r(n) = \frac{n!}{2^n}$. This is a direct way of assessing the growth rate between factorials and exponential functions. Then we can also calculate a recursive ratio $\frac{r(n+1)}{r(n)}$. The results are as follows.

n	n!	2^n	$n!/2^n$	recursive ratio
1	1	2	0.50	
2	2	4	0.50	1.0
3	6	8	0.75	1.5
4	24	16	1.50	2.0
5	120	32	3.75	2.5
6	720	64	11.25	3.0
7	5040	128	39.38	3.5
8	40320	256	157.50	4.0
9	362880	512	708.75	4.5
10	3628800	1024	3543.75	5.0
11	39916800	2048	19490.63	5.5
12	479001600	4096	116943.75	6.0
13	6227020800	8192	760134.38	6.5
14	87178291200	16384	5320940.63	7.0
15	1307674368000	32768	39907054.69	7.5

As we can see, when $n < 4$, factorials grow slower than exponentials, but then $r(n)$ grows faster and faster, as the recursive ratio tells us. For example, $r(10)$ is 5 times that of r(9), which itself is 4.5 times of r(8). This indicates that $\frac{10!}{8!}$ is growing $5 \cdot 4.5 = 22.5$ times faster in proportion than $\frac{2^{10}}{2^8} = 4$. Indeed the math checks out, as $\frac{9 \cdot 10}{4} = 22.5$.

This seemingly rocket-ship like growth trajectory of the ratio is a key characteristic of factorial and combinatorial growth. In the later stage of this growth function, all previous gains pale in comparison, as if they don't exist. When plotting such growth, almost all previous growth seem to be hovering around 0, due to the latest round of growth expanding the scale by one or multiple orders of magnitude.

This means while different recursive ratio grows in a linear fashion, the ratio $r(n)$ grows much faster in a factorial manner. Even the rapid growth of exponential functions cannot tame the monstrous speed bigger factorials carry, as we need to take the log of $r(n)$ to display earlier values of $r(n)$ in details.

In fact, this pattern repeats even as we increase the base for the exponential function b^n, since the factorial values n! do not change. The ratios and recursive ratio will be different (growing slower), but the pattern continues.

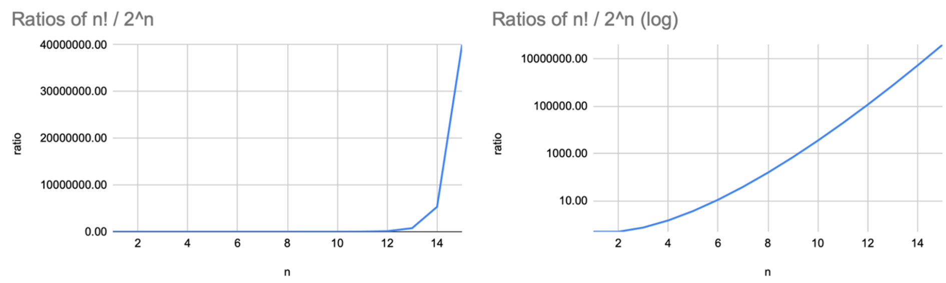


Figure 6.2: Growth of recursive ratios in linear and semi-log plot

6.4 Combinatorial Explosion

As observed above, factorials grow very quickly. While exponential growth is the initial amount growing by a constant proportion of its base (and linear growth is when the initial amount grows by a constant fixed amount), combinatorial growth is the initial amount growing by an increasing speed. It's not $2\cdot2\cdot2\cdot2\cdot2$ but $1\cdot2\cdot3\cdot4\cdot5$, which can be rewritten as $2^0\cdot2^1\cdot2^{1.58}\cdot2^2\cdot2^{2.32} = 2^{6.9}$. You don't even need to consider the "long run" to see an explosive growth. In fact, my corollary is it will explode sooner than you expect. Here it is 2^5 versus $2^{6.9}$, which will accelerate soon.

In short, the relationship between the first base unit and the next base unit is increasing – 2, then 3, then 4, and so on... An interesting side note is such change grows in an exponential decay fashion. Below is a plot of the growth rate from one base integer to the next. For example, 2 grows 50 percent to 3 and 3 grows 33 percent to 4 and so on.

6.5 Economics - Weitzman

The late Harvard economist Martin Weitzman wrote about a combinatorial model on innovation in 1998. The premise is that innovation occurs as a combination of different ideas (or technologies), which combinatorics can help explain. A classic example is how Thomas Edison experimented with thousands of combinations of materials before finding the bamboo filament

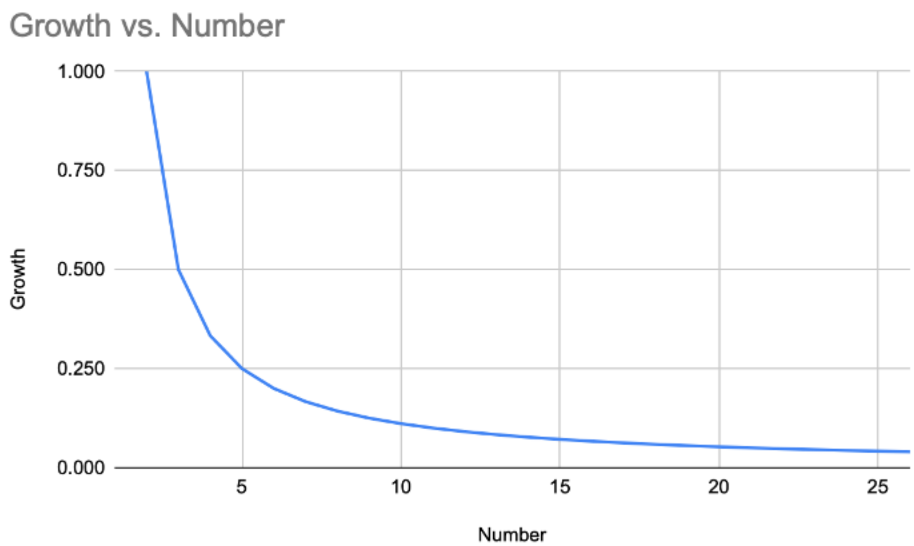


Figure 6.3: Growth from one integer base to the next

that produced the first commercially viable incandescent light bulb. [25]

This model carries a simplified assumptions. Only two ideas are combined at once because considering x ideas combinations at the same time will drastically complicate the calculations.

An implication of this model is once a new combination is considered "useful", it can be added back to the original stock and be combined freely with any other idea (and subsequent pairings). For example, the light bulb can now become a lamp.

So this growth model has five parameters. First, the initial starting ideas. Second, how many ideas we choose to combine each time (here we choose two). Third, the percentage of the new idea combination deemed useful. Fourth, the new total that becomes the initial starting ideas for the next round. Last, time itself, which indicates how many "rounds" (thorough exhaustion of ideas) instead of this constant moving unit on the horizontal axis. And it is this assumption that has a huge implication on how this model applies to economics.

Round	Initial	Unique pairs	New pairs	Useful ideas	Total
1	100	4950	4950	49	149
2	149	11026	6076	61	210
3	210	21945	10919	109	319
4	319	50721	28776	288	607
5	607	183921	133200	1332	1939
6	1939	1878891	1694970	16950	18889
7	18889	17887716	16008825	160088	178977
8	178977	16016293776	16016114799	160161148	160340125

We can devise a simple model of such process.

$$N_{r+1} = N_r + \binom{N_r}{2} \cdot useful \tag{6.1}$$

, where N_r denotes the number of ideas for round r.

Let's start with a thought experiment. Imagine we start with 100 ideas and find all the 2-idea combinations. That yields 100 choose 2 = 4950 unique combinations. Assume 1 percent of the unique pairs are useful, which we add back to the original mix. Now we have 100 + 49 = 149 ideas for the next round of mixing. 149 choose 2 = 11026, which contains the 4950 pairs already found. There are 11026 – 4950 = 6076 new combinations from round 0 and 1 percent of that is 61. Now we have 149 + 61 = 210 ideas in the mix. Repeat this a few times and the results look like the following table.

We see the first four rounds with relatively fast growth, but they are not nearly as crazy as the next three rounds. Then there is an astronomical leap from round 7 to round 8, a growth of $\frac{160340125}{178977} = 895$ times! That is a moment of combinatorial explosion. The following plot shows a hockey stick pattern. Each growth curve appears much flatter in the subsequent round accounting for the last round of fast growth (and even explosion). All the impressive growth beforehand now appears miniscule and barely noticeable (often looking like a flat line near zero). All the growth from 100 in round 1 to 179000 in round 7 pales in comparison to the 160 million new pairs round 8 contributes to the total. It's about three orders of magnitudes (1000 times) larger!

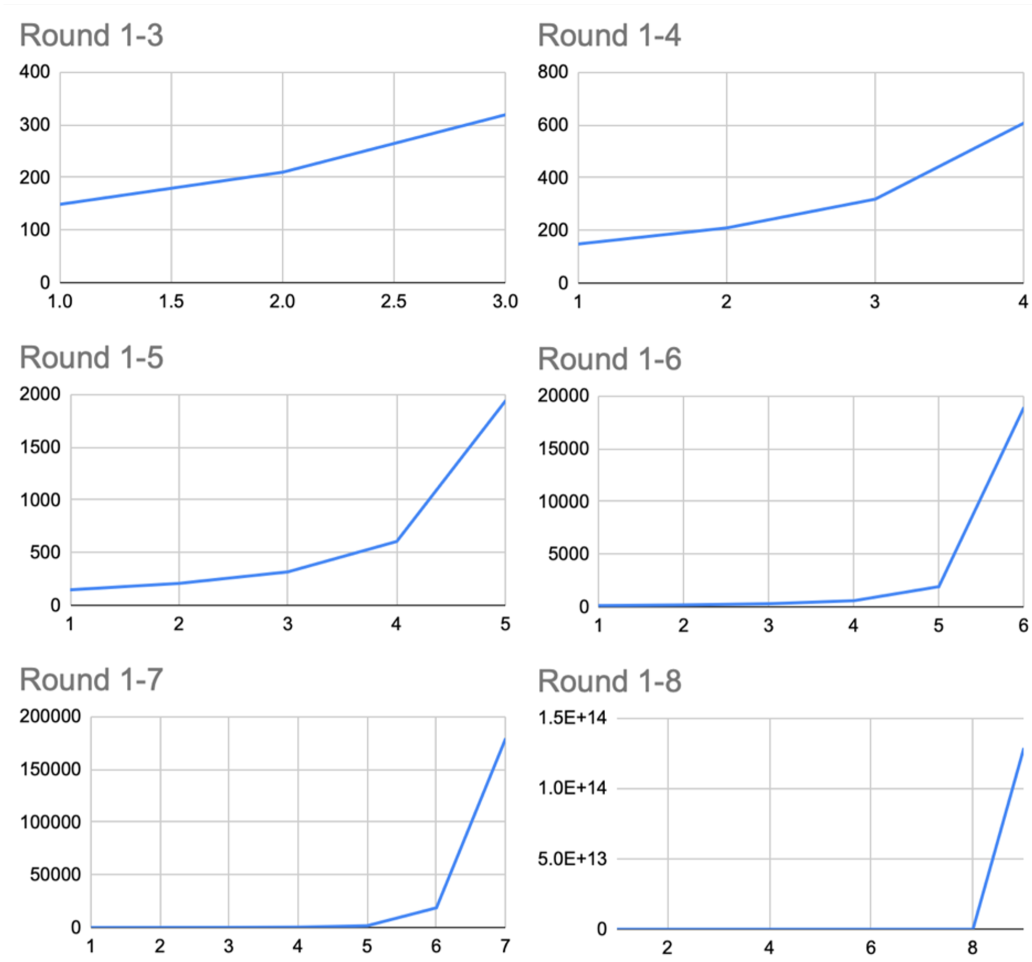


Figure 6.4: The number of ideas for each round

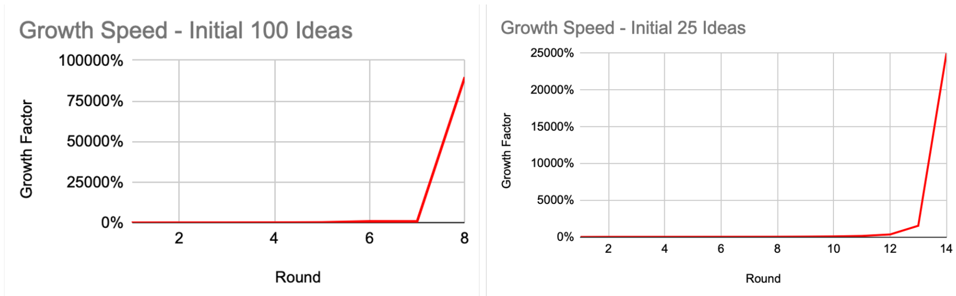


Figure 6.5: Speed of growth of two different initial idea count

The earlier rounds of combinatorial growth can still resemble exponential growth, but it is the sharp hockey-stick behavior that tells us we are in the combinatorial explosion territory. But that is not enough evidence – we also have to examine the growth rate of each round. Exponential growth rate stays fixed (e.g. 5 percent, 25 percent, etc.) and combinatorial growth rate accelerates exponentially.

Another observation is in the "inevitability" of such explosion. Even if we start with 25 ideas (assume we have caveman technologies), we will still get the explosion at round 14.

It may be easy to confuse exponential growth with combinatorial growth before such pre-explosion, so a helpful way is to examine the speed of such growth, namely the new quantity divided by the old quantity. Plotting them for both the initial 100 ideas and initial 25 cases, we can see a similar explosion in growth rate.

There are several caveats to Weitzman's model due to our assumptions. This thought experiment relies on each round of total combination to happen and exhausts simultaneously, but that is not true in real life. It may take many years for someone to produce one useful new idea. Putting stirrups on horses took centuries. Idea growth is most likely staggered left to right.

And if we were to apply the combinatorial growth model, then why is our economy (by means of GDP) growing at an exponential rate? Weitzman suggests that because researching and developing new ideas and technologies also cost a lot of resources and those costs go up much faster than how many

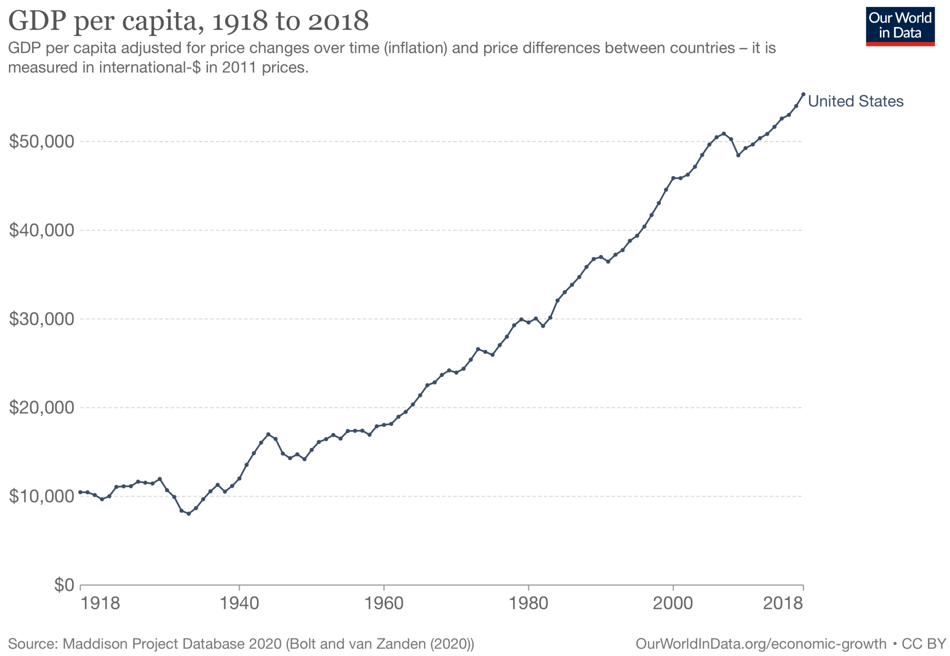


Figure 6.6: GDP per capita of the U.S. 1918 - 2018

new ideas we can generate, so we cannot pursue all the useful ideas. That is true, given not each new useful idea takes the same amount of resource to invent (e.g. a basic diesel engine and a semiconductor assembly line). And of course, a sequence exists in technological development. You cannot go from horses to cars without combustion engine, metallurgy, oil drilling, etc. Resource constraints drag the potential combinatorial growth back to the exponential arena. [8]

Maybe we should take comfort in the long-term implication of combinatorial growth. Even if at present we don't observe explosive growth, in the long run things are getting much much better. [9]

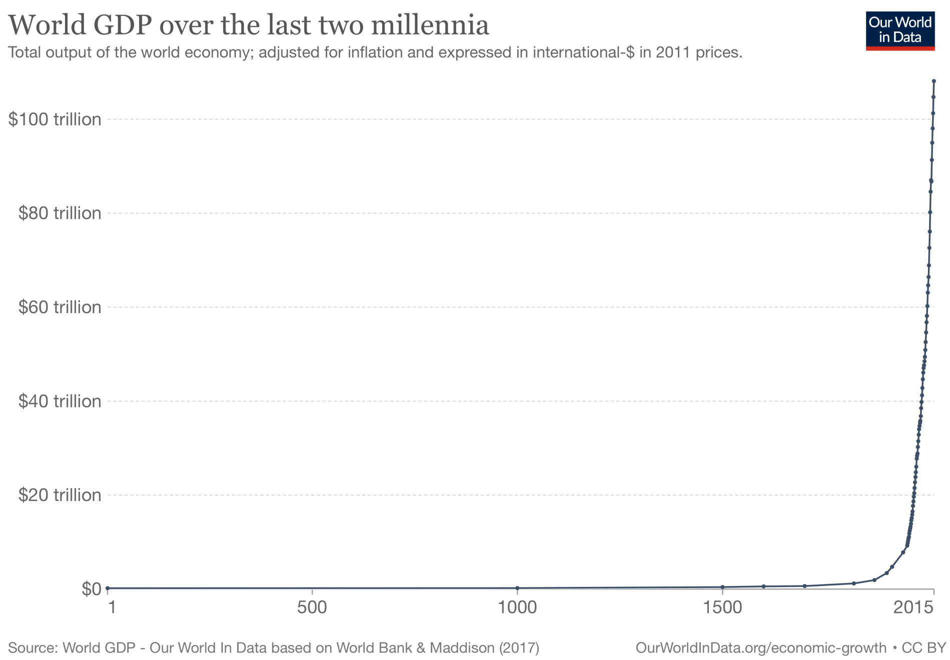


Figure 6.7: World GDP over two thousand years

Chapter 7

Fractal Growth

> But as He who called you is holy, you also be holy in all your conduct, because it is written, “Be holy, for I am holy.” - Peter 1:15 - 16

7.1 Premise

Fractal has become a buzzword in mathematics. Everyone has heard of it at some point with pictures of snowflakes or videos of a seemingly-endless pattern (known as a Mandelbrot set). But very few of us understand what a fractal is, what makes it special, and how it comes to be.

This chapter is an exploration on fractals.

7.2 History

According to Leonard Sander, “a fractal is an object with a sprawling, tenuous pattern. As the pattern is magnified it reveals repetitive levels of detail, so that similar structure exists on all scales.” [21]

In Gulliver's Travels Jonathan Swift referred to the basic idea of a fractal:

> So, naturalists observe, a flea hath smaller fleas that on him prey; and these have smaller still to bite 'em; and so proceed ad infinitum.

The definition of a fractal was not so clear for many years. Mathematicians have pondered the properties of fractals, most notably Gottfried Leibniz, Helge von Koch, Waclaw Sierpinski, Felix Hausdorff, Lewis Fry Richardson, and then Benoit Mandelbrot (who coined the term).

British mathematician Lewis Fry Richardson among other things pioneered mathematical methods to predict weather patterns. He was working on a theory of war and peace in the 1950s. His hypothesis was based on the length of two nations' borders, so measuring the borders became something he researched. [26]

The conventional view is when you use more precise instruments to measure the same length (e.g. a meter stick versus a tape measure based in centimeters), you should get a more accurate reading since the ultimate, objective length does not change but the instrument is more precise.

However, Richardson found out that zooming in changed the border length of Britain! The more he zoomed in (from 100 km to 10 km and so on), the longer the border seems to be. For example, using a measuring stick the length of 200 km, the border of Britain is at around 2400 km. Shortening the measuring stick to 100 km and 50 km, the border length extends to 2800 km and 3400 km respectively.

This phenomenon is because borders are not straight lines, so measuring border lengths with straightedges will ignore plenty of the natural twists and turns of the actual border.

Similarly, Richardson discovered that the increase in length is not random when adjusted for scale. When you plot the length of borders versus the measuring unit (how long the measuring stick is), there is evidence for power law scaling.

Figure 7.1: Approximating Britain's borders with different scale

Richardson found that the crinkliness (or fractality) of west Britain is 0.25, 0.52 for Norway, 0.02 for South Africa, and 0.18 for the border between Spain and Portugal. This means if we double the resolution, say from a 100 km measuring stick to a 50 km measuring stick, the length of Britain's west coast grows by 25 percent, more than 50 percent for Norway, and so forth.

This means the value of a measured length is not helpful unless we know the scale of resolution to measure it in the first place. We must know the unit.

However, Richardson's findings were dormant until Benoit Mandelbrot published his paper "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension."

Benoit Mandelbrot at the time was a researcher at IBM. He began studying the data on coastlines based on Richardson's discovery and called the relationship between the zooming-in similarity fractality (or crinkliness). If the length changes depending on the resolution (or how "zoomed-in" the line is), then the fractality is non-zero. A fractality of zero means the shape is perfectly smooth, because no matter how much you zoom in, the actual length (or perimeter) does not change. A basic circle is one such example.

Mandelbrot defined the word "fractal" and went on to develop the entire field of fractal geometry, giving shape (pun intended) to how we understand fractals.

7.3 Definition and Contexts

There are several important concepts to understanding fractals: self-similarity, fractal dimension, scaling, and networks.

Self-similarity is the key characteristic of fractals. It means each smaller component of something has the same structure as the larger piece. An analogy to a self-similar structure is Russian nesting dolls. A Romanesco broccoli head contains many smaller broccoli heads that have the same structure for multiple "rounds".

Likewise, the mathematics behind self-similarity must be the same as well. As noted above, the 0.25 for west coast of Britain is self-similar, because as one doubles the resolution, the border length grows by 25 percent.

This means fractals are scale-invariant. Scaling up or down the dimension will not change the structure of things. The picture below illustrates this property well. Each small X groups together in fives to form a bigger X, which joins together to form yet another bigger X. There are four such levels below.

Mandelbrot famously wrote: "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line." This is talking about how shapes in the real world are not beautiful nor orderly. Because of this, fractal geometry plays a role in our understanding, especially in considering something's dimension. [26]

In classical (Euclidean) geometry, dimensions are whole numbers. A line (length) is one-dimensional. A square (area) is two-dimensional. A cube (volume) is three-dimensional.

Fractal dimensions are different. One can calculate it by adding 1 to the fractality, like the values above. Hence, the fractal dimension of Norway is 1.52, 1.02 for South Africa, and so on.

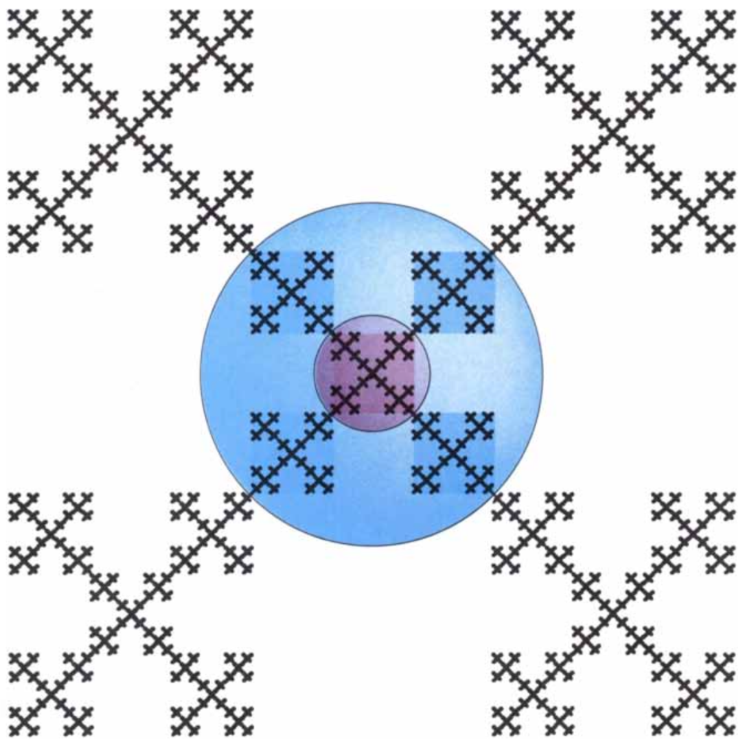


Figure 7.2: A five-repeated fractal

As Smil notes, Fractal dimension equals 1 for smooth Euclidian shapes, between 1 and 2 for two-dimensional shapes – seacoast length has D of 1.25 – and as much as 2.9 (of possible 3) for such complex three-dimensional networks as human lungs. [22]

7.4 Sierpinski Triangles

How else can we derive the fractal dimensions? We first look at classical geometry.

When we double the length of a one-dimensional line, the length increases by a factor of 2. For a square, the area increases by $2^2 = 4$. For a cube, the volume increases by $2^3 = 8$. This is essentially the scaling factor 2 raised to the n-th power, n being the dimension.

With fractals, things are different.

Let's take a look at Sierpinski Triangles, named after the mathematician Waclav Sierpinski. Start with an equilateral triangle, divide it into four even parts and remove the center triangle. Now one can repeat this process for as long as needed.

As one can see, if we scale a smaller Sierpinski triangle (say the top one of the three plus the empty center) by a factor of two, the area increases by 3, instead of 4. We know the increase is only 3 because the center small triangle is missing and the other two small triangles are the same as the top small triangle.

This means our dimension following classical geometry reads like $2^d = 3$, where d is the dimension. Using logarithm we derive $d = log_2(3) = 1.585$, which is a non-integer! 1.585 is the fractal dimension of Sierpinski triangle. As we zoom in on each smaller triangle, we realize this pattern keeps on repeating. The area increases or shrinks by a factor of 3, instead of 4. This is why it's not a fully 2-d shape.

This leads to an important property of fractals, that of scaling. As West notes, "Power law scaling is the mathematical expression of self-similiarty

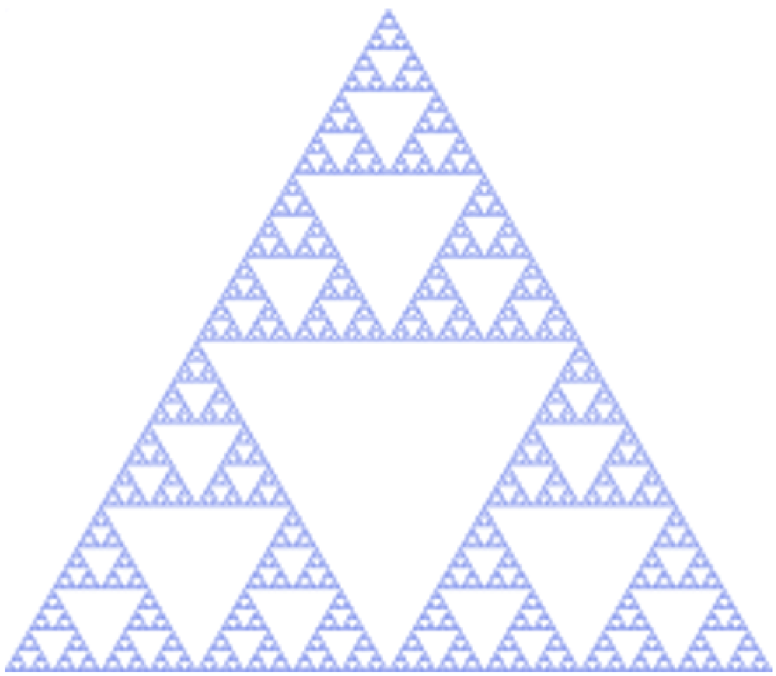


Figure 7.3: Sierpinski triangle

and fractality.” [26]

West has also noted that through natural selection, fractal networks (such as blood vessels) can optimize their distribution of energy so that “organisms operate as if they were in four dimensions”. That extra dimension is what makes fractal geometry so interesting and often counter-intuitive.

As noted from the Sierpinski triangle example, the density of all fractals decreases as its size increases. The area increases by a factor of 3 due to doubling its side length, but the ratio of the area over the “full” area will keep decreasing due to hollowing out the center triangle.

Another notable example of fractals is Koch Snowflakes, named after Helge von Koch.

It is constructed as following. Start with an equilateral triangle and divide one of its line segment into equal threes. Draw an equilateral triangle on top of the middle third line segment like a little hill. Then keep doing that to every line segment and repeat. You will get the picture below.

If you start with an equilateral triangle, the first iteration should result in the Star of David. Then it gradually looks like the following.

During each iteration, the side length increase by a factor of 4 though scaling is only by 3. This means the fractal dimension of Koch snowflakes must satisfy $3^d = 4$, or $d = log_3(4) = 1.262$.

And the boundary of the Mandelbrot set has fractal dimension of 2 — meaning it is as rough a coastline as it could possibly be. [22]

Fractals are used in many different areas, notably in video game and film graphic renderings. Computer software can easily generate background images for mountains and such due to the fractal nature of well, nature.

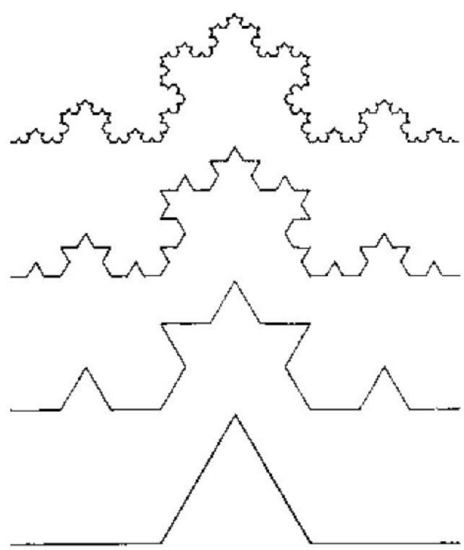


Figure 7.4: Constructing a Koch fractal from one triangle

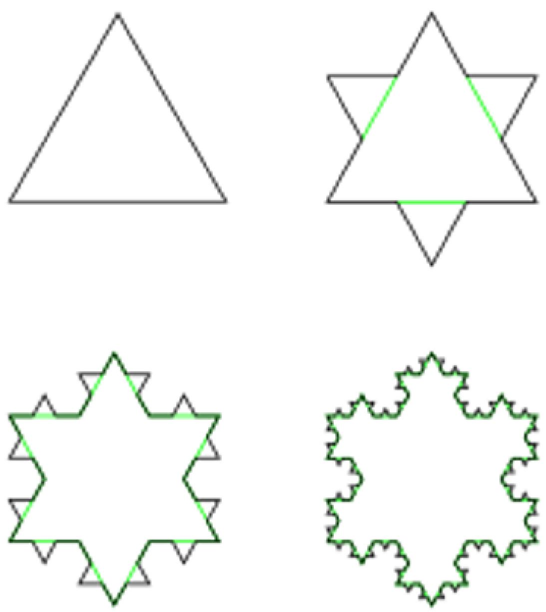


Figure 7.5: Constructing a Koch fractal from one triangle

Chapter 8

Collective Outcomes

> For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them. - Matthew 25:29

8.1 Premise

After examining six types of mathematical growth, the question becomes "what can we learn about things after the growth process?".

When you examine the results of different growth functions, they typically fall under two major buckets: normal and log-normal distribution, or asymmetrical distribution (also known as power law).

8.2 Normal and Lognormal

8.2.1 Normal Distribution

Normal distribution is commonly known as having a bell-shaped curve. It is also known as the Gaussian distribution, named after the famous mathematician Carl Friederich Gauss.

Gauss was not the first mathematician to discover the normal distribution. Pierre-Simon Laplace published Memoire sur la probabilite in 1774, thirty-five years before Gauss published his Theoria motus corporum coellestium in 1809. In fact Laplace wasn't the first mathematician either. Abraham de Moivre published Doctrine of Chances in 1738, which contained normal distribution curve. [22]

Normal distribution has several key characteristics. The distribution is a bell-shaped curve, which means the shape is symmetrical and continuous with no abrupt cut-offs. To obtain the distribution, it requires a certain sample size.

The mean and standard deviation define the shape and spread of the curve. The empirical rule states 68.3% of the results fall within one SD of the mean and 95.4% of the results fall within two standard deviations of the mean.

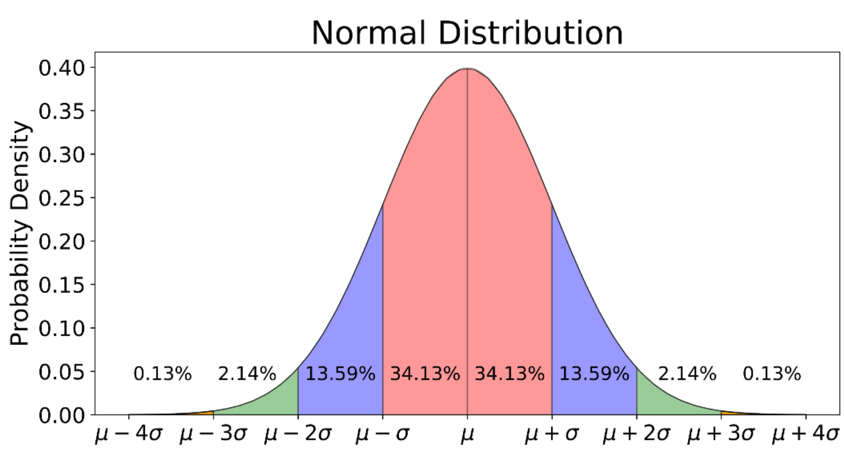


Figure 8.1: Normal distribution with empirical rule

Due to the nature of normal distribution, we often look at the ends of the curve, which are the outliers. Very often we look specifically at the tail on the right end in statistics.

8.2.2 Central Limit Theorem

Central limit theorem is the foundation to all normal distributions. It means the sum of a numerous independent random variable tends to be normally distributed regardless of its underlying distribution. In other words, a lot of observations will result in normal distribution.

The power of central limit theorem lies in the normal distribution of averages (or means). Even if the observed data varies drastically, if we compute the averages of such sample and repeat that process many times, the averages will become normally distributed.

The technical definition is as follows: $\sum X \sim N(\mu n, \sigma\sqrt{n})$, where μ is the mean and σ is the standard deviation.

For more details, please refer to the appendix for a complete proof of the central limit theorem, along with an overview of probability and statistical concepts.

8.2.3 Application

A key observation is that most outcomes of growth processes are normally distributed. For example, distribution of heights, which themselves are functions of linear growth with upper limit. The picture below describes the height distribution of Italian soldiers born in 1900. [10]

One potential problem is applying arithmetic mean from the normal distribution. If the results deviate widely, then the inference will have big issues. For example, if we have include very tall people in our sample and assume the population height to be normally distributed, it essentially assumes the existence of very short people who are actually nonexistent.

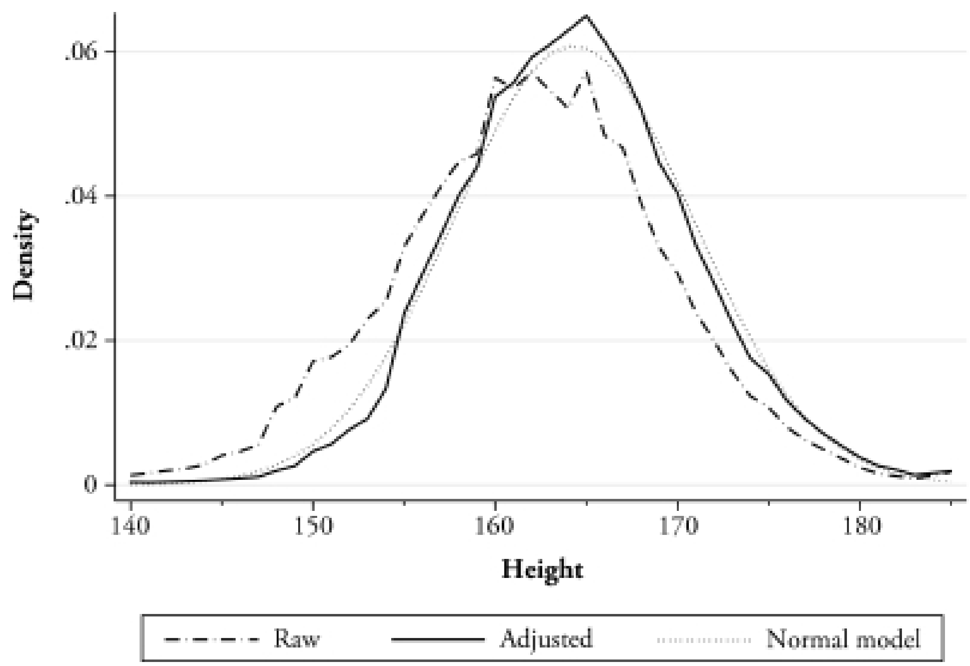


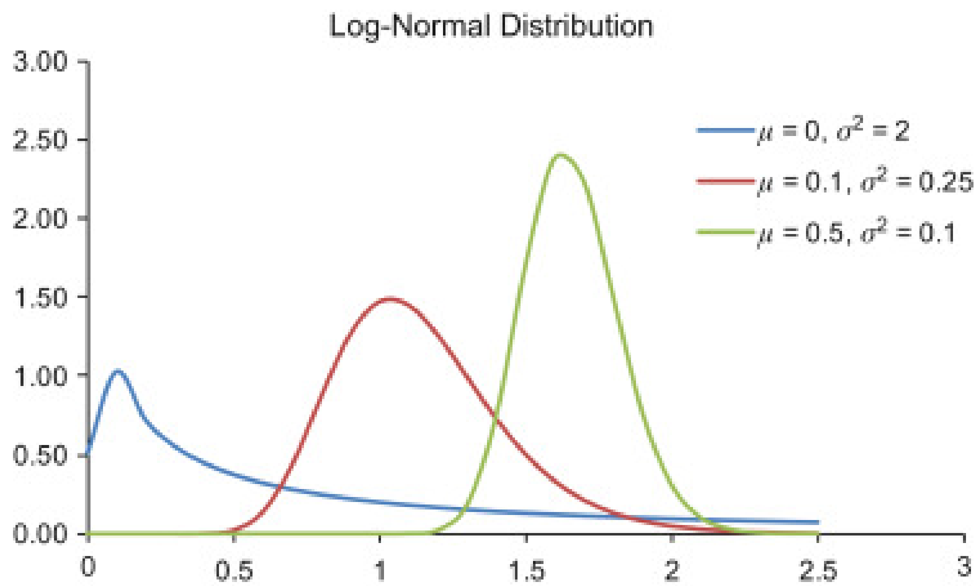
Figure 8.2: Normal distribution of Italian soldiers born in 1900

Hence we want to be careful about the degree of confidence we place in normal distribution.

8.2.4 Log-Normal

Sometimes the distribution is skewed but still has a bell-shaped curve. Such skewed (nonnormal) distributions are due to specific growth. They often skew to the left, also known as having a long tail to the right.

Once we transform the horizontal axis into log form, the distribution will look similar to a normal curve, hence its name log-normal.



8.3 Power Law

Not all growth outcomes are distributed normally. Sudden growth can produce asymmetrical distributions that span orders of magnitude. These distributions are often exponential functions or power-law functions.

8.3.1 Introduction

Power-law functions have the following attributes. They can approximate $f(x) = ax^{-k}$, where a, k are constants. In "perfect" condition, they produce L-shaped curve on linear plot. That means the trend is almost the direct opposite of exponential growth if seen across the mirror. This is because exponential distribution is in fact a form of power law distribution.

If you transform the axis or axes in log, the functions will show linear qualities again. For example, the global distribution of volcanic eruptions display such tendency. The linear plot is that of exponential decline, but when transformed in semi-log the relationship suddenly appears linear. [16]

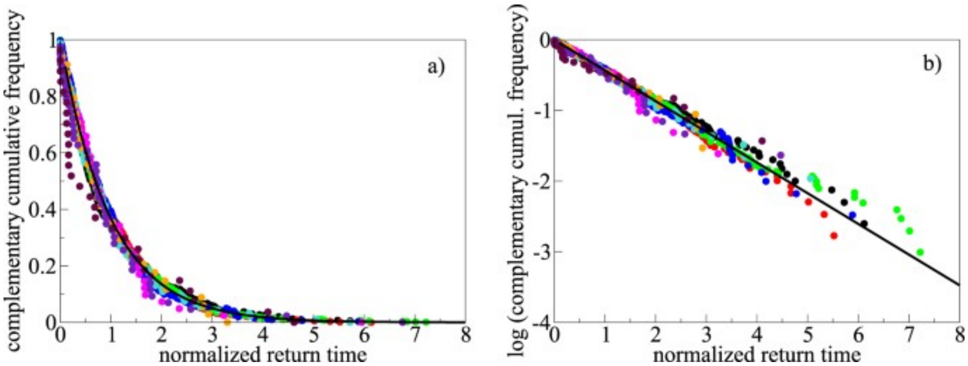


Figure 8.3: Global time-size distribution of volcanic eruptions on Earth

One can also see the connection between fractals and power law distribution, since the concept of self-similarity will quickly produce an increase or decrease in outcome growth.

8.3.2 Pareto Distribution

The most famous power-law distribution is the Pareto distribution, named after Italian economist Vilfredo Pareto. He noted that 20% of rich Italians owned 80% of land in Italy and this pattern could be applied to many phenomena.

This in turn creates the the Pareto principle, or the 80/20 principle, that 80% of the outcomes are from 20% of the causes. It is essentially saying a small group accounts from a disproportionately large amount of the total outcomes.

Benoit Mandelbrot came up with the form of Pareto distribution: $Pr(X > x) = x^{-D}$.

Wealth disparity is one of the most famous and discussed example of power law. It is often dubbed the Matthew effect, named after the famous biblical verse (in the beginning of this chapter).

According to the Federal Reserve's consumer finance survey, the top 1%

group in the United States holds around 46 trillion dollars of wealth by the end of 2021, a staggering $\frac{46}{142} = 32.4\%$ of total national wealth. Add in the next 9% with their 53 trillion dollars of wealth, the top 10% hold $\frac{99}{142} = 69.7\%$ of total national wealth. [18]

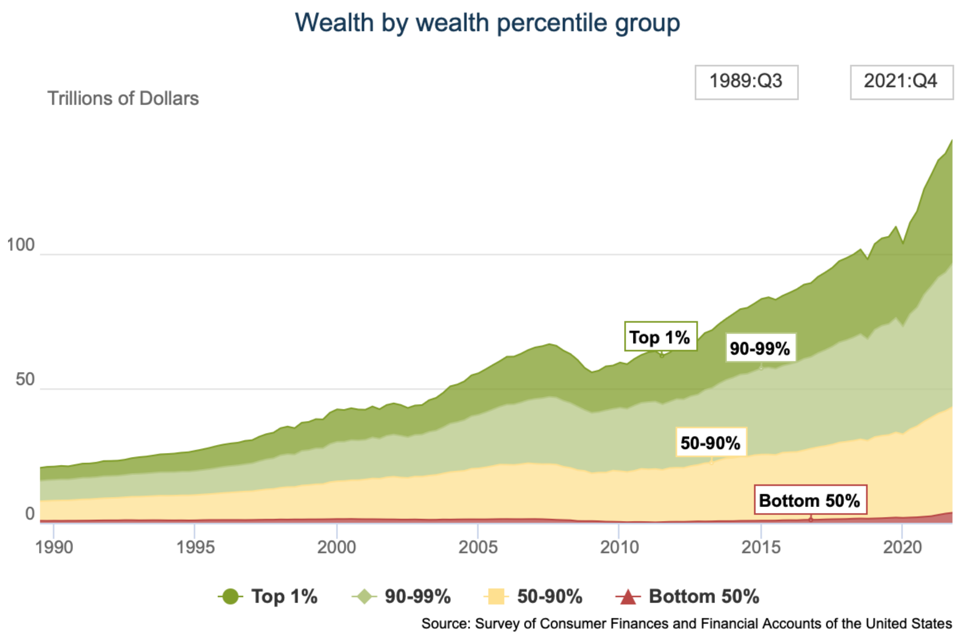


Figure 8.4: Distribution of Household Wealth in the U.S. since 1989

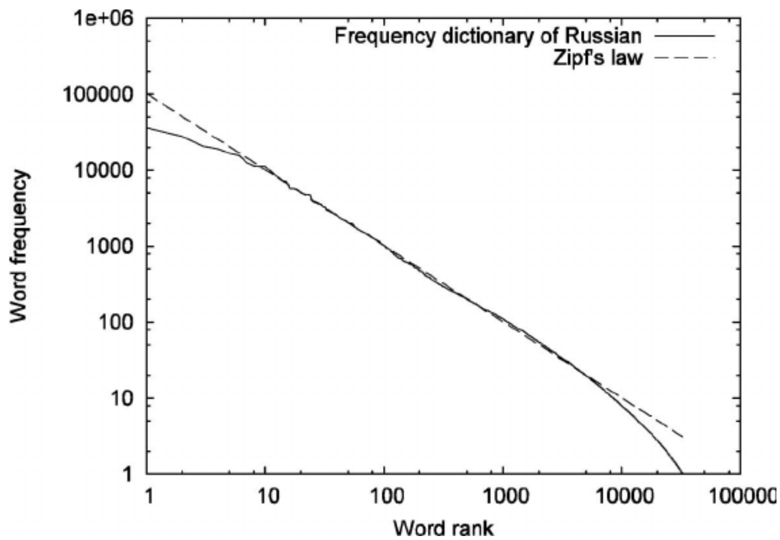


Figure 8.5: Word rank order plotted against frequency in Russian

8.3.3 Zipf's Law

Linguist George Zipf claimed that word frequency (P_n) is inversely proportional to its rank n in the frequency table: $P_n \propto \frac{1}{n^a}$, with the exponent a being close to 1. This means the most frequently used the word is about twice as frequently used as the second most common word, as illustrated by the plot below.

Zipf also studied the inverse power relationship between the ranking r of cities by their population size x: $x = r^{-1}$. It is actually the same as the Pareto distribution with the variables swapped. The figure from below measures the population of U.S. cities against U.S. population in 2009.

Notice both observations are based on log-log scale, which means if plotted linearly, the graph will look extremely skewed with a very long tail.

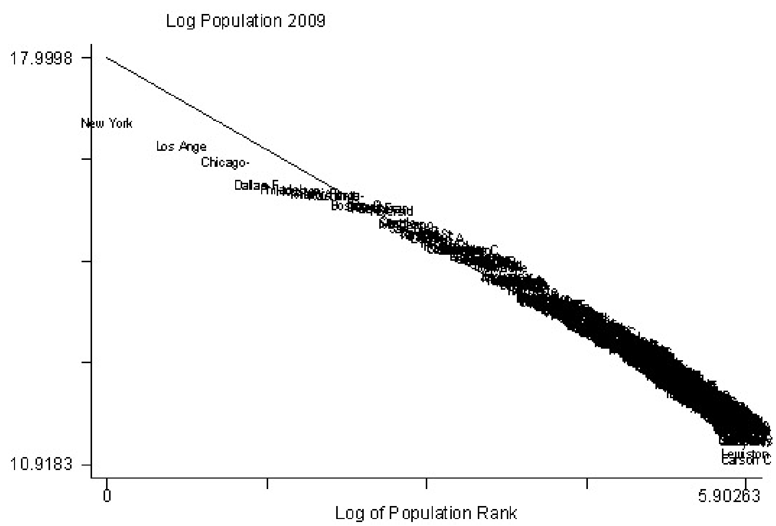


Figure 8.6: Population rank of U.S. cities plotted against 2009 population in log-log

8.3.4 Issues

When the sample size is relatively small and the data has large variance, different distributions with long tails are hard to separate. Power-law, log-normal, and Weibull distributions are such examples. In addition, log-normal and power-law functions often behave similarly. [22]

So the power law doesn't always hold true, partly because we are trying to explain complex reality with simple models (which we know are never correct but sometimes useful).

Bibliography

[1] Tor Bair. *Exponential Growth Isn't Cool. Combinatorial Growth Is.* URL: https://medium.com/@TorBair/exponential-growth-isn-t-cool-combinatorial-growth-is-85a0b1fdb6a5.

[2] World Bank. *GDP (current US dollar) - United States, China.* URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US-CN.

[3] P.L. Bernstein. *Against the Gods: The Remarkable Story of Risk.* Wiley, 1998.

[4] J.C. Bogle. *The Little Book of Common Sense Investing.* Wiley, 2017.

[5] Brilliant. *Principle of Inclusion and Exclusion (PIE).* URL: https://brilliant.org/wiki/principle-of-inclusion-and-exclusion-pie/.

[6] Warren Buffett. *Berkshire Hathaway Letter to Shareholders.* URL: https://berkshirehathaway.com/letters/2021ltr.pdf.

[7] Constance Crozier. *Forecasting s-curves is hard.* URL: https://constancecrozier.com/2020/04/16/forecasting-s-curves-is-hard/.

[8] Our World in Data. *GDP per capita, 1918 to 2018.* URL: https://ourworldindata.org/grapher/gdp-per-capita-maddison-2020?country=~USA.

[9] Our World in Data. *World GDP over the last two millennia.* URL: https://ourworldindata.org/grapher/world-gdp-over-the-last-two-millennia.

[10] A. Hearn. "Height and the normal distribution: evidence from Italian military data". In: *Demography* (2009). URL: https://doi.org/10.1353/dem.0.0049.

[11] James Jones. *4.5 - Exponential and Logarithmic Models*. URL: https://people.richland.edu/james/lecture/m116/logs/models.html.

[12] Vlad Krokhmal. *Introductory Probability and the Central Limit Theorem*. 2011. URL: https://www.math.uchicago.edu.

[13] C. G. Lewin. "Compound interest in the seventeenth century". In: *Journal of the Institute of Actuaries* 108.3 (1981), pp. 423–442. DOI: 10.1017/S0020268100040865.

[14] Federal Reserve Bank of Minneapolis. *Consumer Price Index, 1913-* . URL: https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-.

[15] Nathan P. Myhrvold. "Revisiting the Estimation of Dinosaur Growth Rates". In: *PLOS ONE* (2013). DOI: 10.1371/journal.pone.0081917. URL: https://doi.org/10.1371/journal.pone.0081917.

[16] P. G. Papale. "lobal time-size distribution of volcanic eruptions on Earth". In: *Nature* (2018). URL: https://doi.org/10.1038/s41598-018-25286-y.

[17] Francesco Balducci Pegolotti. *La Pratica della Mercatura*. Mediaeval Academy of America, 1936.

[18] The Federal Reserve. *Distribution of Household Wealth in the U.S. since 1989*. URL: https://www.federalreserve.gov/releases/z1/dataviz/dfa/distribute/chart/.

[19] Andrew Rothman. *Central Limit Theorem: Proofs Actually Working Through the Math*. URL: https://towardsdatascience.com/central-limit-theorem-proofs-actually-working-through-the-math-a994cd582b33.

[20] Anders Sandberg et al. "Sigmoids behaving badly: why they usually cannot predict the future as well as they seem to promise". In: (2021). URL: https://arxiv.org/abs/2109.08065.

[21] Leonard M. Sander. "Fractal Growth". In: *Scientific American* (1987). URL: http://www.jstor.org/stable/24979300.

[22] Vaclav Smil. *Growth*. MIT Press, 2019.

[23] A. Tsoularis. “Analysis of Logistic Growth Models”. In: (2001). URL: https://mro.massey.ac.nz/bitstream/handle/10179/4341/Analysis_of_Logistic_Growth_Models.pdf?sequence=1&isAllowed=y.

[24] unknown. *Combinatorics Sum and Product Rules Some Subtler Examples*. URL: http://www.cs.cornell.edu/courses/cs280/2004fa/280wk6_x4.pdf.

[25] Martin L. Weitzman. “Recombinant Growth”. In: *The Quarterly Journal of Economics* (1998). URL: https://doi.org/10.1162/003355398555595.

[26] Geoffrey West. *Scale: The Universal Laws of Life, Growth, and Death in Organisms, Cities, and COmpanies*. Penguin Books, 2017.

Appendix A

Proof of Central Limit Theorem

This proof is designed to be self-containing, hence it can be used as a brief primer on the necessary contexts for understanding the Central Limit Theorem. [19] [12]

A.0.1 What is CLT?

The Central Limit Theorem (CLT) is fundamental to probability and statistics. It can provide valuable insights from random-seeming things.

In plain English, CLT means given a large sample size, the distribution of sample means will roughly become normal, regardless of the population distribution. For example, cows at a local ranch have different weights, which may or may not form a bell-shaped curve. However, if you take the average of the cow weight and do that many times with randomly generated data-set or results from other comparable ranches, then the mean weight will form a bell curve.

A.0.2 Version

There are many versions of the Central Limit Theorem, depending on different assumptions. The most common version is the Lindberg-Levy CLT, which is what we will prove here.

A.0.3 Relevant Concepts

Several concepts are important to understanding the Central Limit Theorem.

Note The mathematical symbols and formulae used in this proof are simplified to my best ability, since increasing abstraction makes understanding more difficult, especially for those without background in probability theory or statistics.

Random Variables A random variable (RV) is a variable X that assigns outcome values randomly depending on the type and the function. It can be discrete, continuous, or a mix of both. A discrete example is flipping a fair coin, where the input is a flip and the outcome is one of two coin sides.

i.i.d An independently and identically distributed (i.i.d) random variable has the same probability distribution as the other random variables and they are mutually independent of one another. For example, flipping a fair coin means outcome from each flip does not depend on the outcome from previous flips.

Discrete and Continuous Variables Without loss of generality or precision, I define discrete and continuous variables as follows. Discrete variables can be counted, for example numbers on a six-faced dice, binary 0 and 1, or simply all non-negative integers. Continuous variables cannot be counted, because they can extend to infinity. An example is all real numbers, since you can write $3, 3.01, 3.001, \ldots$ infinitely.

Probability Mass Function A probability mass function $p(x) = P(X = x)$ provides the probability that a discrete random variable X is exactly equal to some value $-\infty < x < \infty$. The probabilities of a probability mass function must be non-negative and add up to 1, or $p(x) \geq 0$ and $\sum_x p(x) = 1$.

Probability Density Function A probability density function (PDF) $f(x)$ provides the probability that a continuous random variable X is equal to. We can integrate the PDF to derive the cumulative distribution function (CDF), in this case $F(x) = \int_{-\infty}^{\infty} f(x)dx$.

Mean The mean (colloquially known as the average) of X is formally known as the expected value $E[X]$.

If X is a discrete random variable with finite outcomes $x_1, x_2, \ldots, x_k$ and corresponding probability $p_1, p_2, \ldots, p_k$, the expected value is $E[X] = \sum_{i=1}^{k} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots x_k p_k$.

For example, a fair six-faced die with numbers $1, 2, 3, 4, 5, 6$ has an expected value of $E[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$.

If X is a continuous random variable with a probability density function of $f(x)$, then the expected value is $E[X] = \int_{-\infty}^{\infty} x f(x)dx$.

Variance If X is a random variable with a mean of μ (or $E[X] = \mu$), then the variance of X is $Var(X) = E[(X - \mu)^2]$.

For example, the variance of a fair six-faced die with $1, 2, 3, 4, 5, 6$ is $Var(X) = \frac{1}{6}((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2) = \frac{35}{12}$.

Note that $Var(X) = E[(x - E[X]^2)^2] = E[X^2] - E[X]^2$, which we will cover in the later section.

Standard Deviation The standard deviation (sd) of a random variable X is the square root of its variance. Therefore $SD = \sqrt{Var(X)}$.

Moments We will define moments in more details later. For now, remember that the expected value (mean) is the first moment of a probability distribution (PMF or PDF) and the variance is the second moment of such probability distribution.

A.0.4 Technical Definition

Given a sequence of n independently and identically distributed (i.i.d) random variables (RV) $X_1, X_2, \ldots, X_N$ with a mean $\mu = 0$ and variance of σ^2. Note that $\sigma^2 < \infty$.

Let a RV Y be the mean of this sequence of i.i.d. RVs $X_1, X_2, \ldots, X_N$, so $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$. The mean of Y is μ and the variance of Y is $\frac{\sigma^2}{n}$.

If Y^* is a RV with a center of μ and variance of $\frac{\sigma^2}{n}$, then Y^* distribution is approximately same as the standard normal distribution $Y^* \sim N(0, 1)$.

Alternative From "Applied Statistics and Probability for Engineers" by Douglas Montgomery and George Runger, the CLT is defined as "if $X_1, X_2, \ldots, X_N$ are n random samples drawn from a population with overall mean μ and finite variance σ^2, and if $\bar{X_n}$ is the sample mean, then the limiting form of the distribution, $Z = \lim_{n\to\infty} \sqrt{n}\frac{\bar{X_n}-\mu}{n}$ is a standard normal distribution" (241, Montgomery/Runger).

A.1 Background Concepts

There are several background concepts I need to cover, after finishing the proof.

A.1.1 Mean

There are several properties about expected values.

Linearity: $E[X+Y] = E[X] + E[Y]$

$E[aX] = aE[X]$

If X and Y are independent, then $E[XY] = E[X] \cdot E[Y]$

A.1.2 Variance

There are several properties about variance.

$Var(X) = E[(x - E[X]^2)^2] = E[X^2] - E[X]^2$

$Var(a) = 0$ for a constant a.

$Var(X + a) = Var(X)$.

$Var(aX) = a^2 Var(X)$.

The variance of a sum of two random variables

A.1.3 Moment Generating Functions

The moment generating function (MGF) of a random variable X is $M(t) = E[e^{tX}]$ if the expected value is defined. It may or may not exist for any specific value of t.

If the random variable X is discrete, then $M(t) = \sum_x e^{tx} p(x)$.

If the random variable X is continuous, then $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.

$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} x e^{tx} f(x) dx$.

$M'(0) = \int_{-\infty}^{\infty} x f(x) dx = E[X]$.

A.1.4 Standard Normal Distribution

Explain what is a standard normal distribution and how to write it.

A.2 Before The Proof

There are several preparatory steps before the final proof.

A.2.1 $E[Y]$ and $Var[Y]$

Remember that $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$E[Y] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$.

$Var(Y) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}vat(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

In conclusion, $E[Y] = \mu$ and $Var(Y) = \frac{\sigma^2}{n}$.

A.2.2 Y^*

Let's create a new random variable Y^* that is centered by its mean $E[Y]$ and has a standard deviation of $\sqrt{Var(Y)}$.

$Y^* = \frac{Y-E[Y]}{\sqrt{Var(Y)}} = \frac{Y-\mu}{\sqrt{\sigma^2/n}} = \frac{\sqrt{n}(Y-\mu)}{\sigma} = \frac{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i-\mu)}{\sigma}$.

A.2.3 Change of Variables

Define the random variable S as the sum of $X_1, X_2, \ldots, X_N$, so $S = \sum_{i=1}^{n} X_i$.

$E[S] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu = n\mu$.

$Var(S) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2.$

$S^* = \frac{S-E[S]}{\sqrt{Var(S)}} = \frac{S-n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sigma\sqrt{n}} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\sqrt{n}\mu)}{\sigma}.$

Remember $Y^* = \frac{Y-E[Y]}{\sqrt{Var(Y)}} = \frac{Y-\mu}{\sqrt{\sigma^2/n}} = \frac{\sqrt{n}(Y-\mu)}{\sigma} = \frac{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i-\mu)}{\sigma} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\sqrt{n}\mu)}{\sigma}.$

It turns out that $Y*$ and $S*$ are equivalent in terms of estimating the sample distribution! Since $S*$ is easier to use for our proof, we will work with $S*$ instead of $Y*$.

A.2.4 Moment Generating Function Properties

$M_Z(t) = M_X(t)M_Y(t)$ if $Z = X + Y$ and X and Y are independent.

Let's call this new standardized random variable Z and find its moment generating function. Several intermediate steps are skipped due to algebra, as one can use complete the square, exponential laws, and algebraic distributions to write exponents as powers.

$M_Z(t) = E[e^{xt}] = \int_{-\infty}^{\infty} e^{xt}\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{tx-\frac{x^2}{2}}\,dx = e^{\frac{1}{2}t^2}.$

A few properties to keep in mind.

If M_A and M_B are independent, then $M_c(t) = M_A(t)M_B(t)$.

$M_A^r(0) = E[A^r]$.

$M_X^t(0) = E[X^t]$.

A.3 Proof

Remember that $S^* = \frac{S-E[S]}{\sqrt{Var(S)}} = \frac{S-n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^{n} X_i-n\mu}{\sigma\sqrt{n}} = \frac{(X_1-\mu)+(X_2-\mu)+\cdots+(X_n-\mu)}{\sigma\sqrt{n}} = \frac{(X_1-\mu)}{\sigma\sqrt{n}} + \frac{(X_2-\mu)}{\sigma\sqrt{n}} + \cdots + \frac{(X_n-\mu)}{\sigma\sqrt{n}}.$

Now let's find the MGF of S^*.

$M_{s^*}(t) = E[e^{S^*}] = E[e^{t(\frac{(X_1-\mu)}{\sigma\sqrt{n}}+\cdots+\frac{(X_n-\mu)}{\sigma\sqrt{n}})}] = (E[e^{t(\frac{X-\mu}{\sigma\sqrt{n}})}])^n$
$= (M_{(\frac{X-\mu}{\sigma\sqrt{n}})}(t))^n = (M_{X-\mu}(\frac{t}{\sigma\sqrt{n}}))^n$.

Expand MGF to Taylor series.

$M_{X-\mu}(\frac{t}{\sigma\sqrt{n}}) = E[e^{t(\frac{X-\mu}{\sigma\sqrt{n}})}] = 1 + (\frac{t}{\sigma\sqrt{n}})(E[X-\mu]) + (\frac{t^2}{2n\sigma^2})(E[(X-\mu)^2]) = 1 + (\frac{t}{\sigma\sqrt{n}})(0) + (\frac{t^2}{2n\sigma^2})(\sigma^2) = 1 + \frac{t^2}{2n}$.

Since $\lim_{n\to\infty}(1+\frac{x}{n})^n = e^x$, $\lim_{n\to\infty} M_{S*}(t) = \lim_{n\to\infty}(1+\frac{t^2}{2n})^n = e^{1/2t^2}$

If $M_Z(t) = e^{\frac{1}{2}t^2}$, then $\lim_{n\to infty} M_{S*}(t) = M_Z(t)$, and the distribution is also similar that $N(0,1)$.

Acknowledgements

1. Professor Gizem Karaali
2. Professors at the Pomona College Math Department
3. My friends
4. My mother