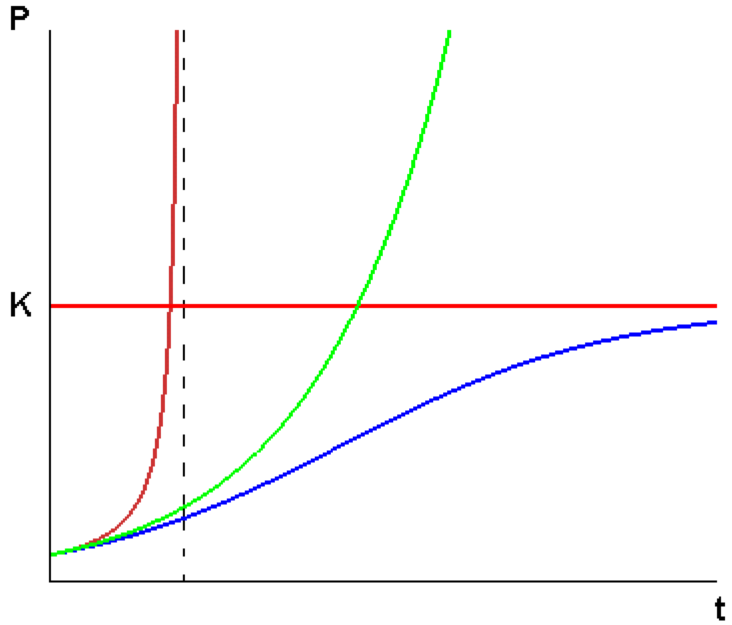
Identifying The Logistic and Limited Exponential Functions Models in Common Data



Project by Vladimir Filatov, Abdullah Alrubian, Kristian Don

For Mathematical Modelling II

December 2024

Table of Contents

Introduction

The limited exponential dates to the 18th century when it was first formulated by Thomas Malthus through a comparison of population and its growth. The limited exponential takes the carrying capacity [L] into account unlike the original exponential, [L] also represents limitations that can occur in the real world, such as environmental factors, and competition.

$$y = L + (y_0 - L)e^{-kx}$$

The logistic function was introduced by Pierre Verhulst who developed it during the 18[th] century with the intention of studying how resources limit the size and age of the population. Verhulst improved the Malthusian curve by implementing the density dependent competition, which can lead to a more balanced population in a carrying capacity.

$$y = \frac{L}{1 + (\frac{L}{y_0} - 1)e^{-kx}}$$

There are two equations used in modelling and plotting the data, each is used for specific reasons,

$ln(\frac{L - y}{L - y_0}) =- kx$, in this equation the data is well-balanced, hence we can guess a value for L,

$\frac{L-y_{n+1}}{L-y_n} = e^{-kd}$, and in this equation the time difference is constant (d).

From the limited exponential and logistic equations, the linear equation is produced:

$$Y = aX + b$$

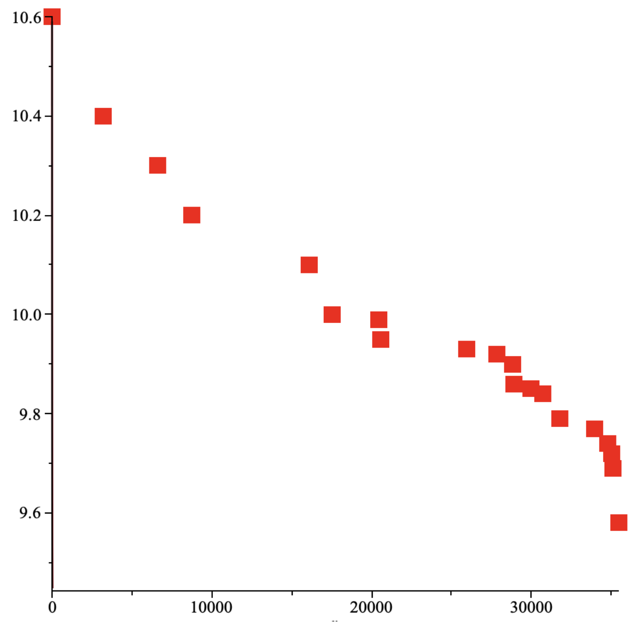
Where $a = e^{-kd}$ and $b = L(1 - e^{-kd})$.

The following report will illustrate the usage of both limited exponential and logistic equations in calculating 2 different problems each having 3 various tasks. The first problem will be the speed limit of the men's 100m sprint speed limit, in which the first task will calculate the error and see if the prediction is accurate, the second task is going to further improve the prediction, and the last task is going to illustrate the marathon record. Next will be life expectancy, where the first task will determine which of the models has the best function, the second task will compare the different countries, and the last task will be a future prediction based on the calculated data.

Men's 100m sprint speed limit

Task 1

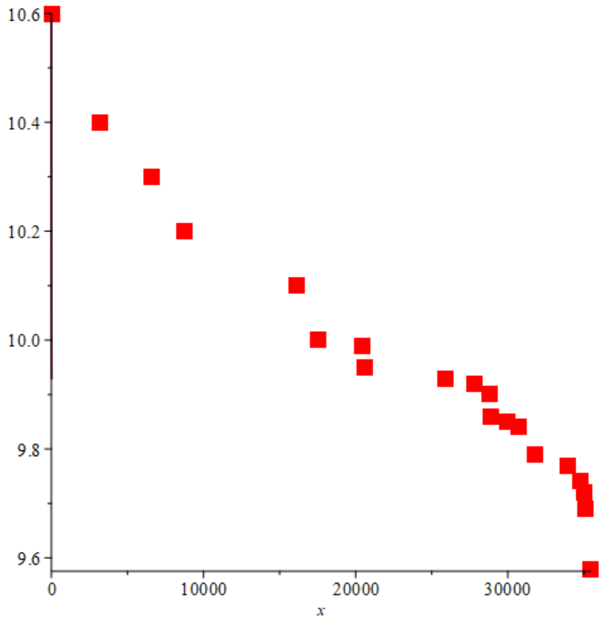
Firstly, we were tasked with fitting a limited exponential model to a set of data which included the men's 100m world records since 1912 and the dates they were broken. We had to input all of the values and calculate an error using our model. Usain Bolt estimates that the world records will stop at 9.40 seconds, therefore we used this value as our limit.



Based on the graph we can see that Bolt's prediction was quite close. The error we calculated using his prediction of 9.4 seconds was just 5%.

Task 2

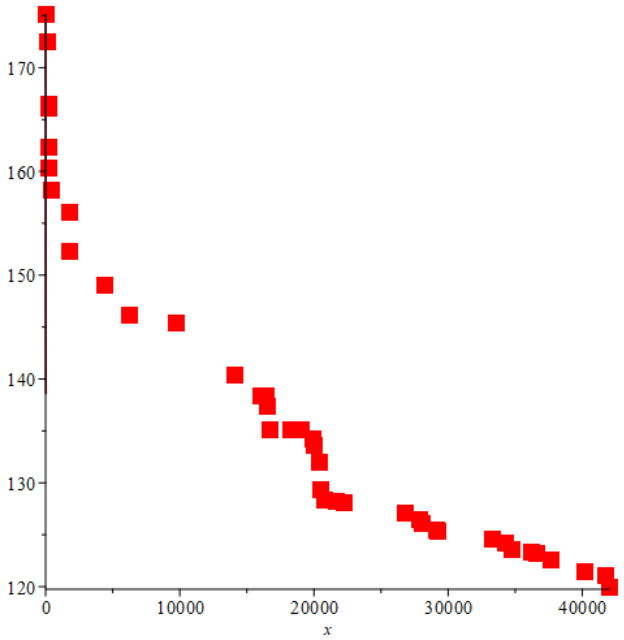
For task 2 we had to improve on Usain Bolt's prediction of 9.4s. To do this, we experimented with different limit values using trial and error, then using the value with the smallest error. This is what we came up with:



As seen in the graph above we found a value even closer and more accurate than Bolt's prediction. With our limit equal to 9.9 seconds the error fell to 2%, therefore successfully improving on Usain Bolt's prediction.

Task 3

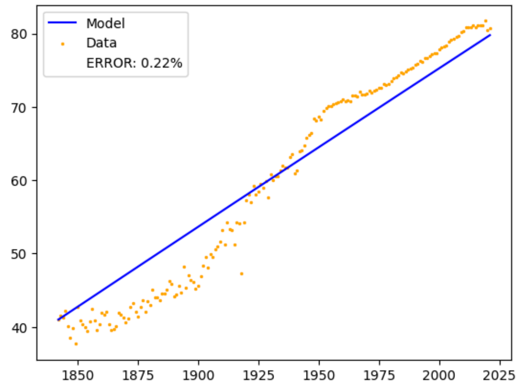
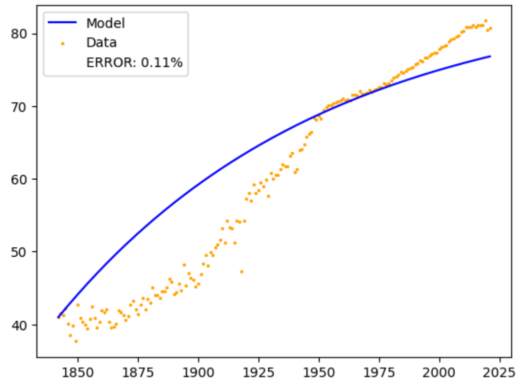
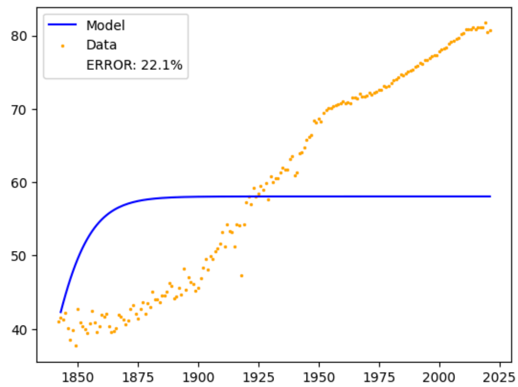
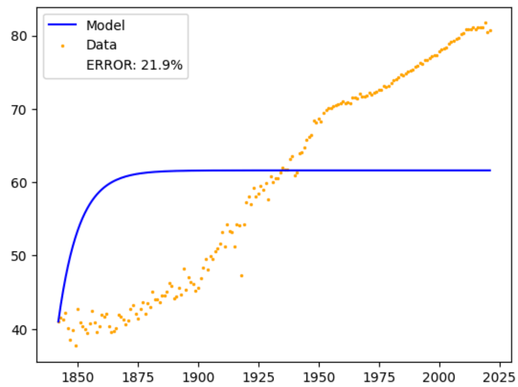
Task 3 requirs us to solve a similar problem using marathon record times since 1908. We used the same procedure and steps, again using experimentation through trial and error to give a rough estimate for the limit.



Based on the graph above and the limit value we used (137) we calculated 10% error using the marathon time and dates.

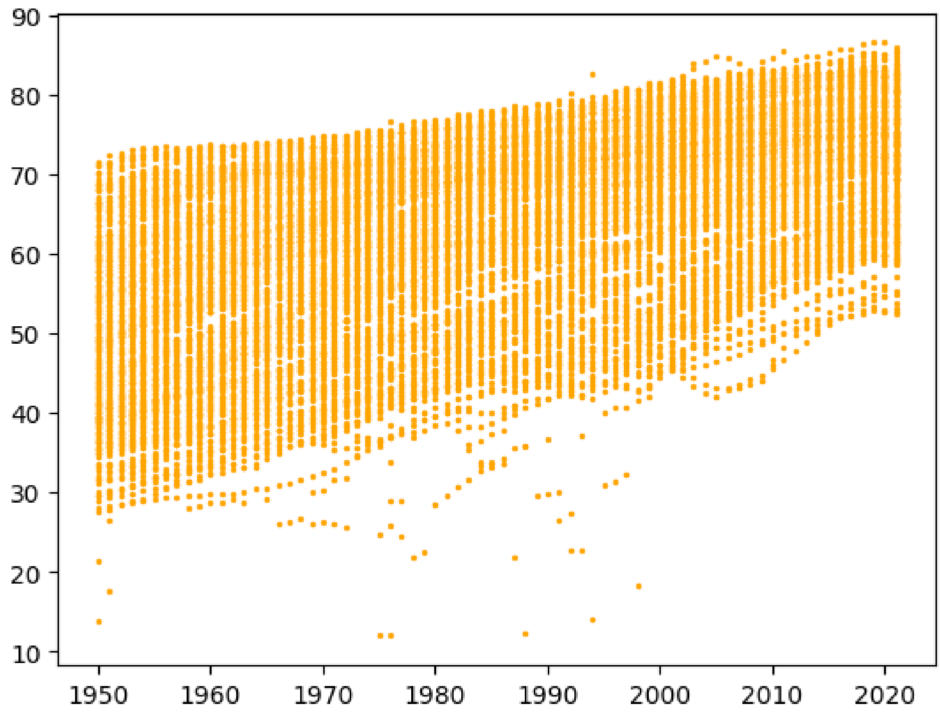
Life Expectancy Modelling

Having introduced the two methods for each model our team began experimenting to see which of the methods is more accurate. Once again, the first method is to estimate the value L, the function value at which the graph is supposed to converge, while the second method is to find L from the a and b values found after applying linear regression.
Here are the results having the United Kingdom data as an example:

Method №	Logistic	Limited Exponential
1		
2		

A few things to take note of: increasing L in the Logistic Method 1 variant does not affect the graph anyhow, while increasing L in the Limited Exponential Method 1 “straightens” the curve, decreasing the error(since the data looks closest to a linear regression). The sampling period starts from 1842, since this is when the United Kingdom has data for every single year after that, so we can correctly assume a constant step $d = 1$ (most countries have data starting from the year 1950). Additionally, changing the period or observed country does not fix Method 2’s incorrect modelling for both functions.

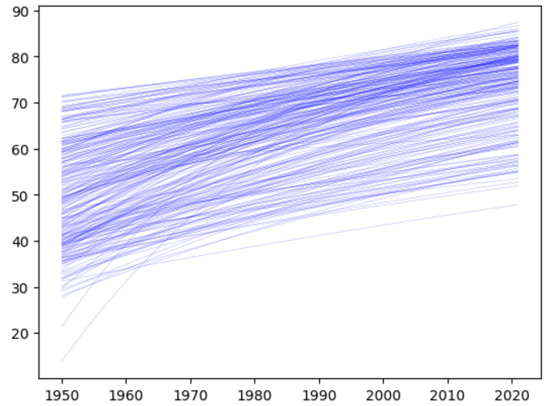
Task 1: Which Function Is Better?

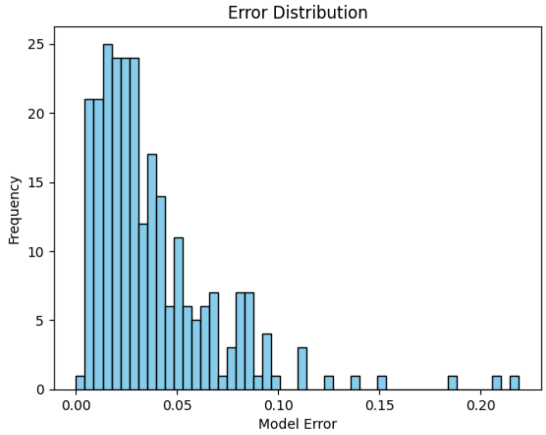


This chart shows every country’s progression in life expectancy(excluding Northern Ireland, Scotland and the USSR, countries that do not have data for the 1950-2021 period each for their own reason). Some interesting patterns to notice: the top “limit” is more horizontal, than the bottom “limit”, showing that there is room to grow for the countries with low life expectancy while it is harder for the top countries to increase the metric at a fast pace. Additionally, the reliability of this graph is questionable, since 1950 is right after the second world war, which is when the age expectancy was low, so the graph might show a more drastic improvement in the metric than it would have been in a different time period. As for the other aspects of the cleanness of this data, doubts arise for some of the points below 20 years.
On a positive note, it shows that there wasn't a single country that generally decreased in life expectancy over these 70 years. Moreover, most countries tremendously increased with an average growth of 22.6 years with only 8 countries having a growth of less than 10 years.

Having said that, here are the models using the limited exponential function:

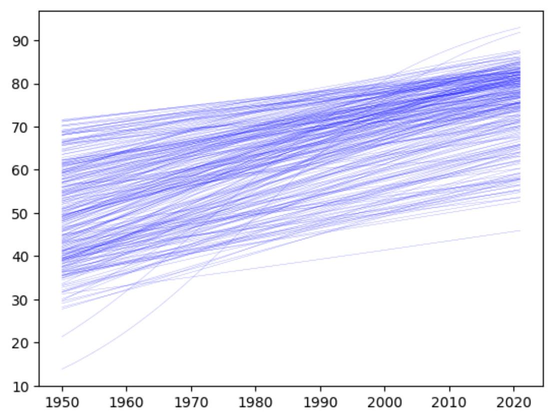
Limited Exponential Function

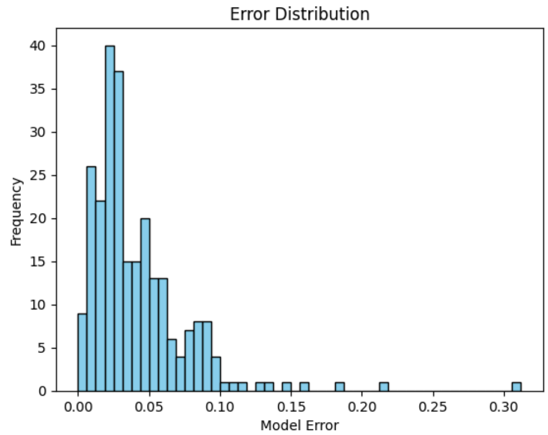




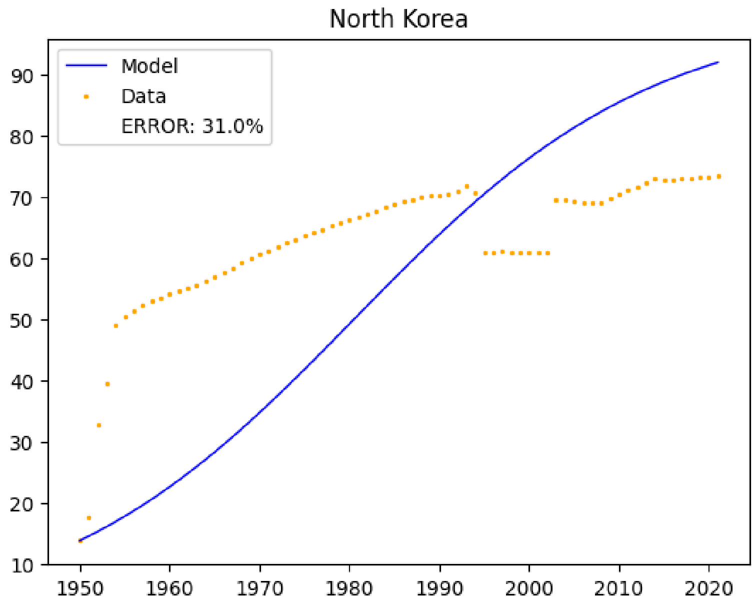
The graph sums up the models of each country in one of the most important metrics. Our team picked an L value of 100, which means that each line should converge at that value after a certain time period. From the model error distribution it is visible that the limited exponential is generally a good model, as even the highest error countries do not exceed 25%. The average error was 3.8%.

Logistic Function





Apart from a couple of hairs that actually depict a logistic function, the vast majority of lines are close to straight, which is a small difference from the limited exponential. Nevertheless, the error distribution was very similar to the previous model, except for a lower variance. The average error was 4.2% and there was only one extreme case of bad fitting data, and here is it’s plot:



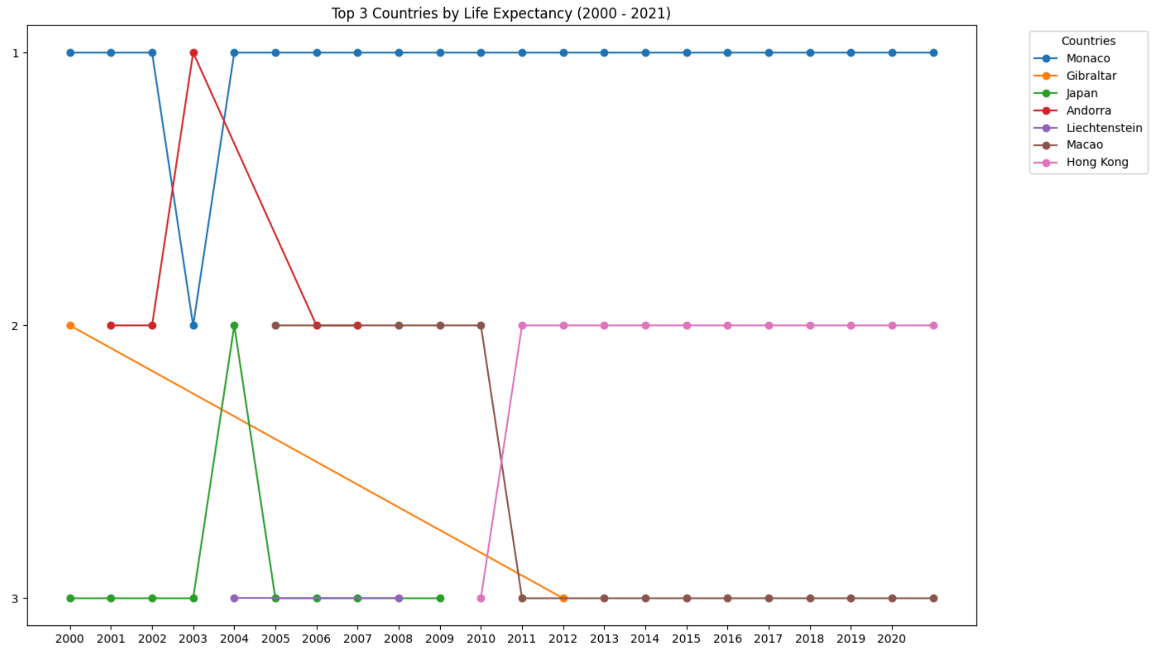
Of course, with such a country it is hard to have open and reliable data, which is the reason it had 13 years of life expectancy in 1950 and a sudden drop in the 2000s.

Conclusion

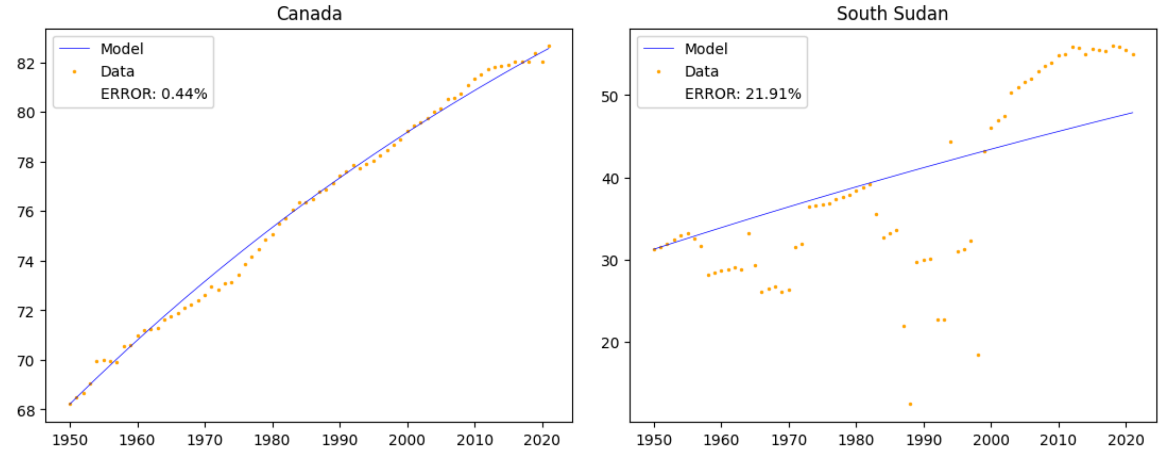
The limited exponential function wins by 0.4% of average error difference which is why our team will be pursuing it in the following tasks. To the naked eye, after going through each individual country's data, it was visible that <5 countries had a graph similar to a logistic one, while there were plenty of examples where the data resembled a limited exponential. Of course, had the logistic function had a smaller error, our team would reject this hypothesis.

Task 2: Compare Countries

Most of the countries started recording the data in 1950, which leaves us with quite a small period to test our models. During the 1950-2021 period only 31 unique countries appeared in the top 10 world rankings for each year: Norway, Netherlands, Iceland, Sweden, Denmark, Guernsey, Jersey, New Zealand, Australia, Switzerland, San Marino, Latvia, United Kingdom, Monaco, Faroe Islands, Andorra, Canada, Bonaire Sint Eustatius and Saba, Japan, Spain, Hong Kong, Greece, Gibraltar, Macao, Liechtenstein, Bermuda, Italy, Singapore, Malta, South Korea, Martinique. This information highlights that countries with high life expectancy tend to stay at the top while it is a rare occasion for a new country to appear in the top ranks. It was found using this chart which looks fine with small sample size, however gets messier with more data depicted:

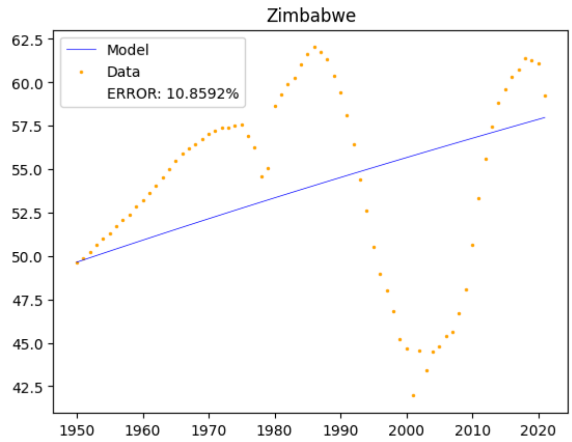


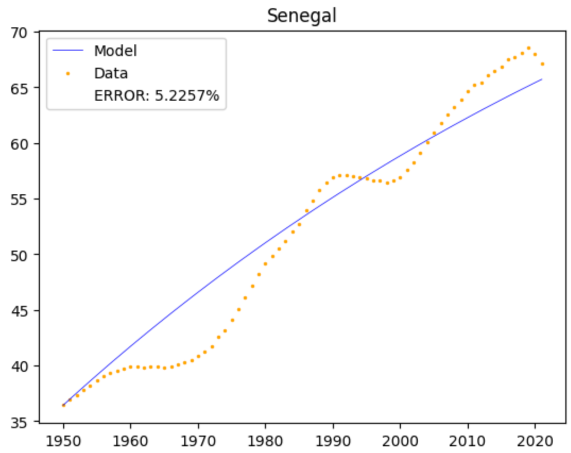
Here are the best fitting and worst fitting countries from the 236 total:

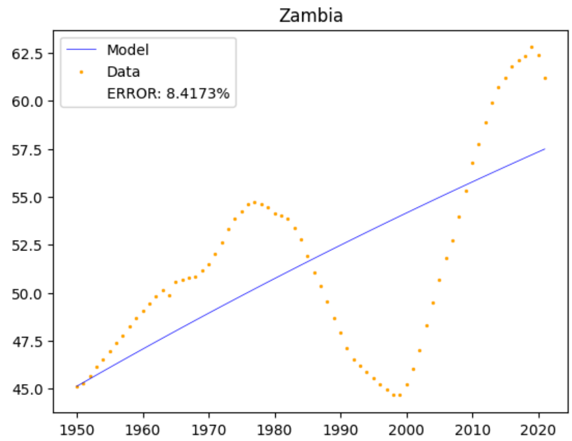


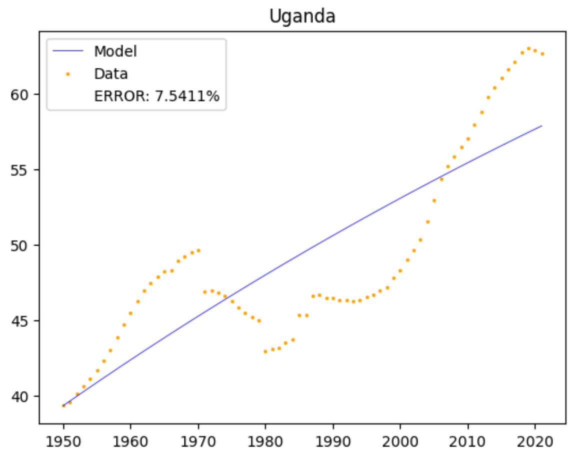
Other highlights include:

Africa

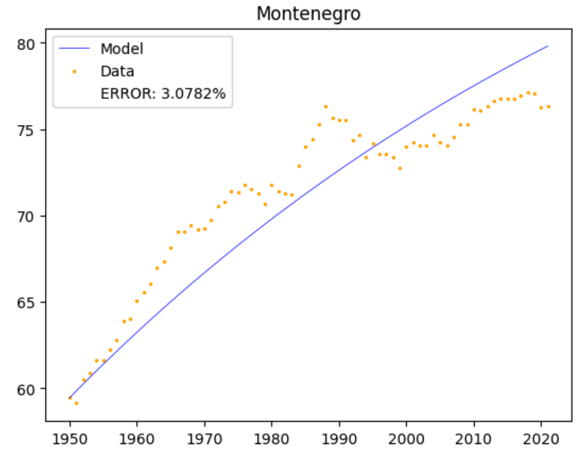

Zimbabwe
Model
Data
ERROR: 10.8592%
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
42.5
1950
1960
1970
1980
1990
2000
2010
2020

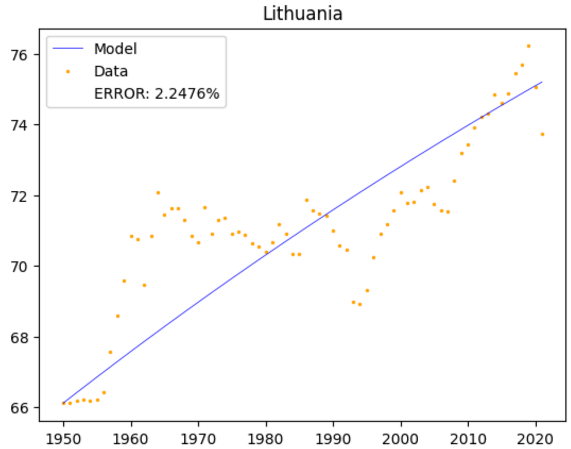

Senegal
Model
Data
ERROR: 5.2257%
70
65
60
55
50
45
40
35
1950
1960
1970
1980
1990
2000
2010
2020

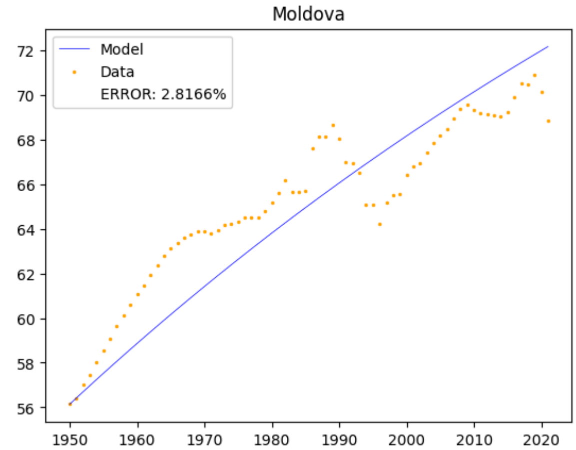

Zambia
Model
Data
ERROR: 8.4173%
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
1950
1960
1970
1980
1990
2000
2010
2020

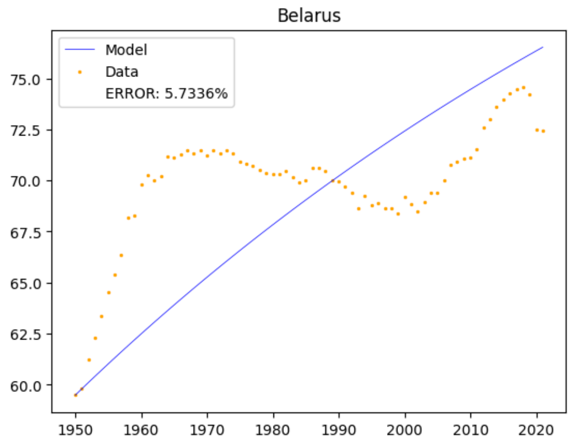

Uganda
Model
Data
ERROR: 7.5411%
60
55
50
45
40
1950
1960
1970
1980
1990
2000
2010
2020

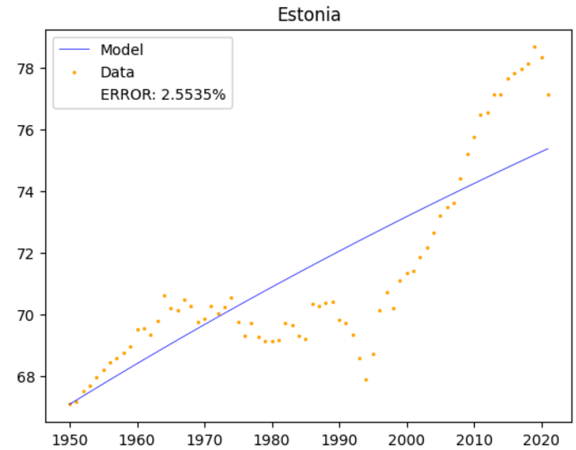
Eastern Europe & Post-Soviet

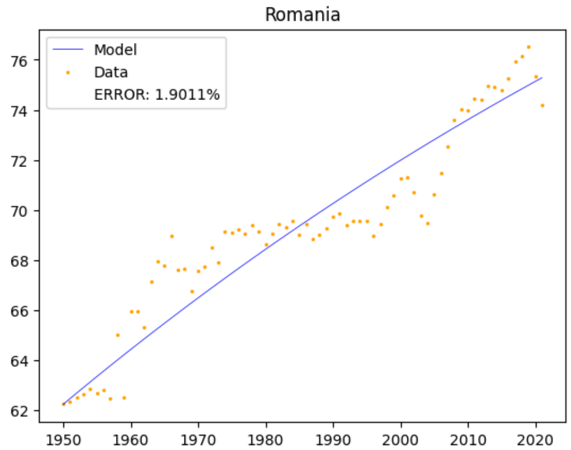




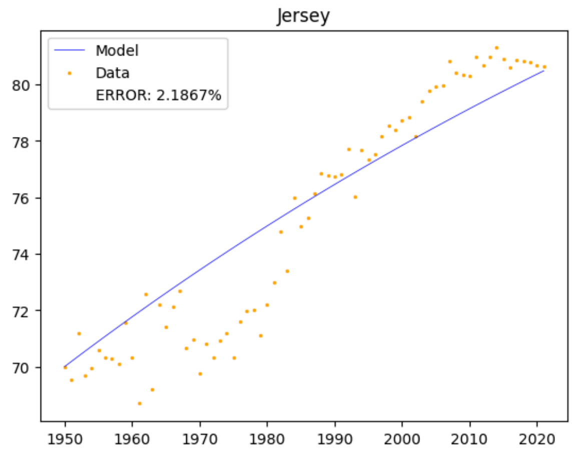


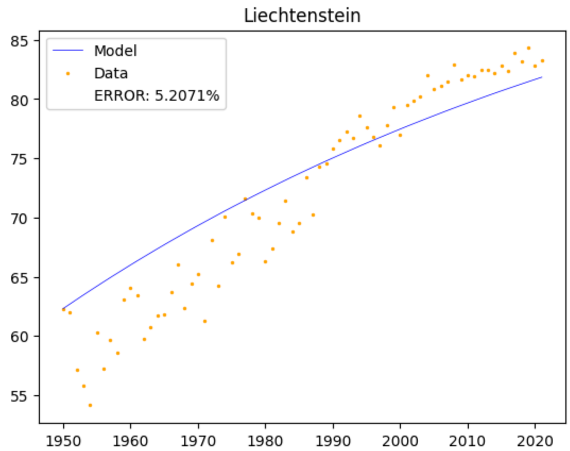


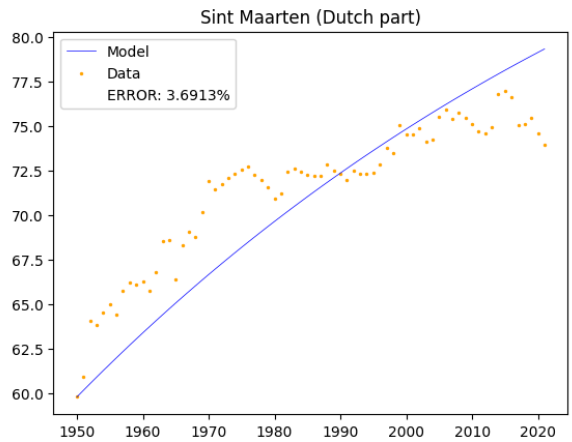


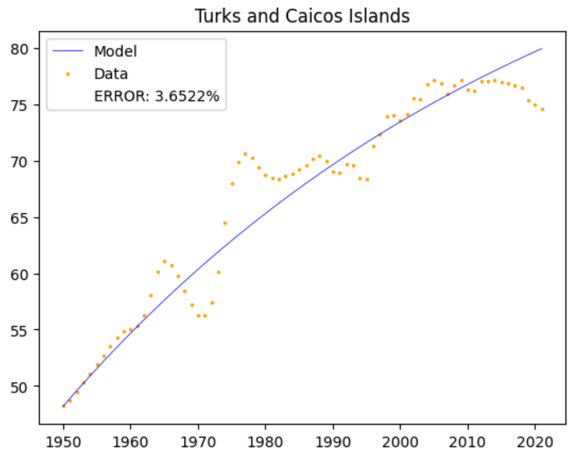


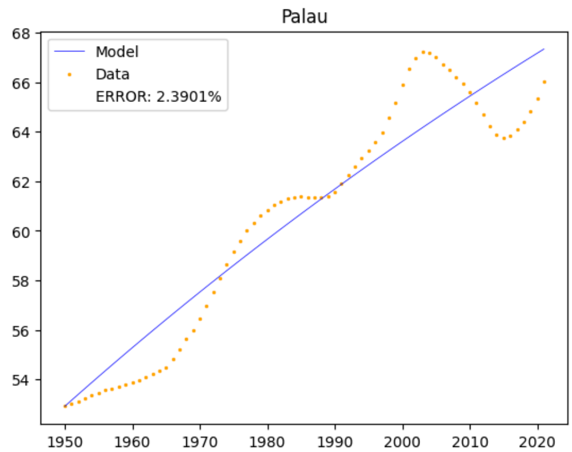
Miscellaneous

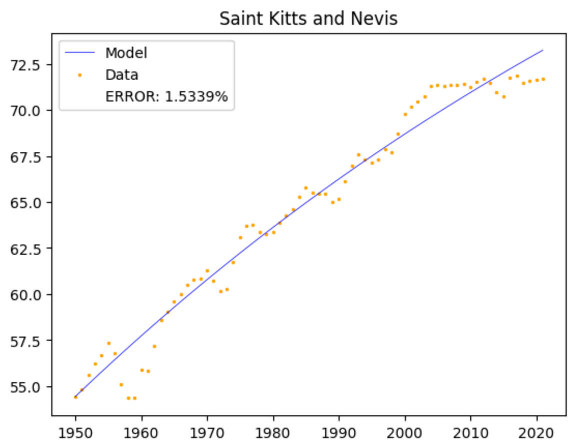






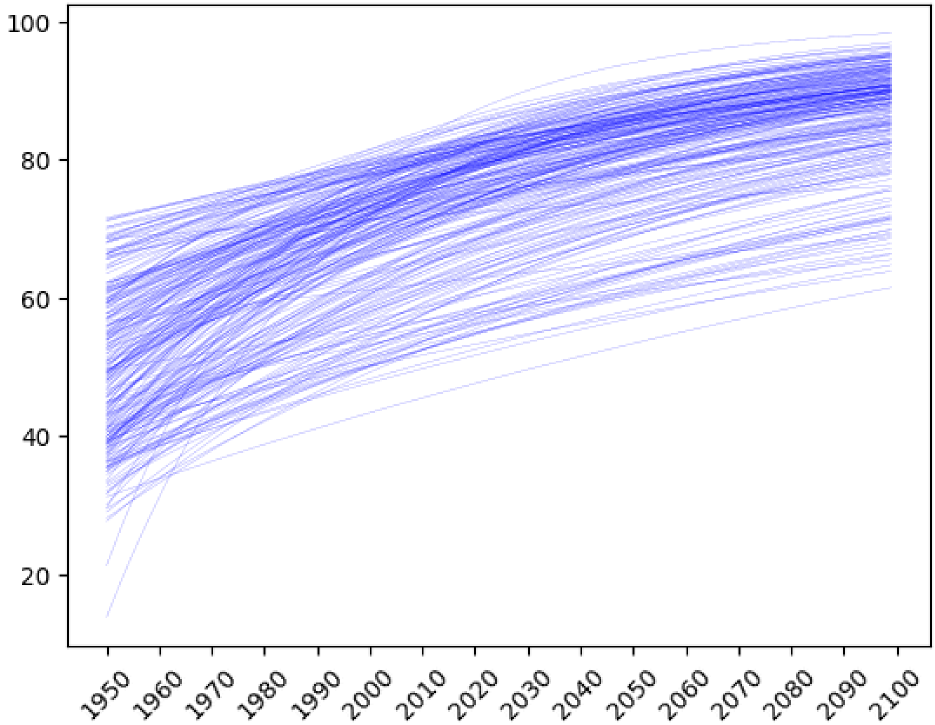




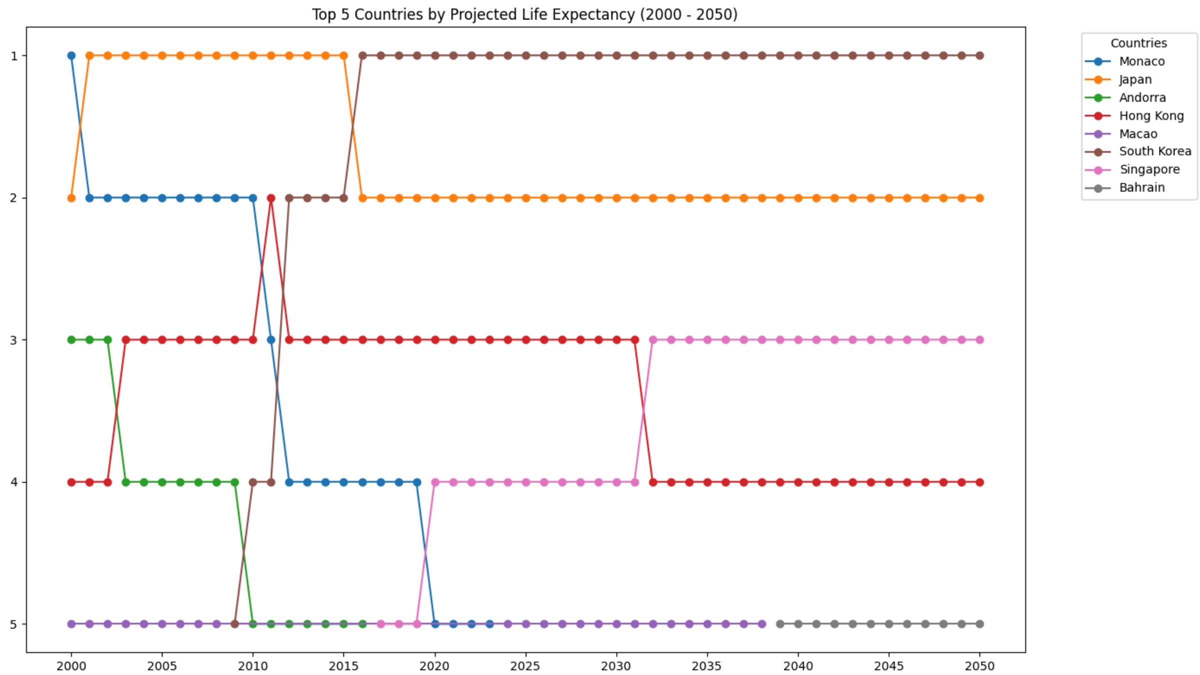


Task 3: Predictions for the Future

We decided to extend the models until the year 2100 to find insights.



Using the ranking chart from before our team calculated the top 5 countries by life expectancy predicted by our models until the year 2050:



Answering the task question, it seems Spain will not be overtaking Japan as the country with the longest life expectancy by 2040. Moreover, the models predict that South Korea will be the unanimous leader up until 2100. Unlike real data, our models do not take into account conflicts, migration and other factors affecting life expectancy. For that reason the chart shows very continuous and curvy changes in the rankings: some countries move slowly up, some slowly move down the top charts and in the end it all converges to a more and more defined order, in our case that is

1. South Korea
2. North Korea
3. Bahrain
4. Oman
5. Singapore

By the year 2600 (and the order holds until 3000). So what is common between these 5 countries? The answer our team found is that all 5 countries had a tremendous increase between 1950 and 2021 and therefore, the model predicted that the same rate would apply further on as well.

Conclusion

The limited exponential and logistic functions are two of the most important and widely used functions in certain areas of science, they are unique for their convergence and gradual change. In the case of our examples of using them, it was found that the functions are good at modelling life expectancy growth and modelling world records times based on an estimate. This is clearly shown in task 2 of the 100m sprint times where based on our estimate, the error was just 2%. These functions are versatile and extremely accurate when used to model certain cases.

Resources

- Life Expectancy csv file - Our World In Data
- Nature Education: Exponential and Logistic Equations in Population Growth
- World record progression of 100m
- Marathon world record progression

Contributions

Vladimir Filatov	Life Expectancy
Abdullah Alrubian	Introduction, Presentation
Kristian Don	100m Sprint