Computational Simulation: Numerical Methods
Assignment 2

David Lawton
22337087

13th Oct. 2024.

Contents

1 Introduction

In this report we detail our analysis of the growth of human population since 1750, using an exponential growth model.

The data we used was obtained from both 'Gapminder' (1750 - 1940) and the 'United States Census Bureau' (1950 - 2016).

2 Exponential Growth Model

The first step in our analysis was to view the data on a logarithmic scale with respect to population, as this would allow us to determine whether the growth of the human population was approximately exponential.

Instead of a 'passive' logarithmic transformation, where we change the scale on the plot, we ran an 'active' logarithmic transformation on the data, where we took the natural logarithm of the population before plotting was done. This is done so that a linear regression can be applied to the transformed data, rather than an exponential one on the original data.

2.1 Theory

In an exponential model the data is assumed to follow the form

$$n(t) = n_0 \exp\left[\lambda(t - t_0)\right] \tag{1}$$

which after our logarithmic transformation, becomes

$$\ln n(t) = \ln n_0 + \lambda(t - t_0) \tag{2}$$

thus on our logarithmic plot, the transformed data follows a linear form.

$$f(t) = \ln n(t) = a + bt \tag{3}$$

where $a = \ln n_0 - \lambda t_0$, $b = \lambda$.

We now run a linear regression on subsets of the transformed data, which appear linear, to determine approximate values of a_i and b_i, in the ith subset.
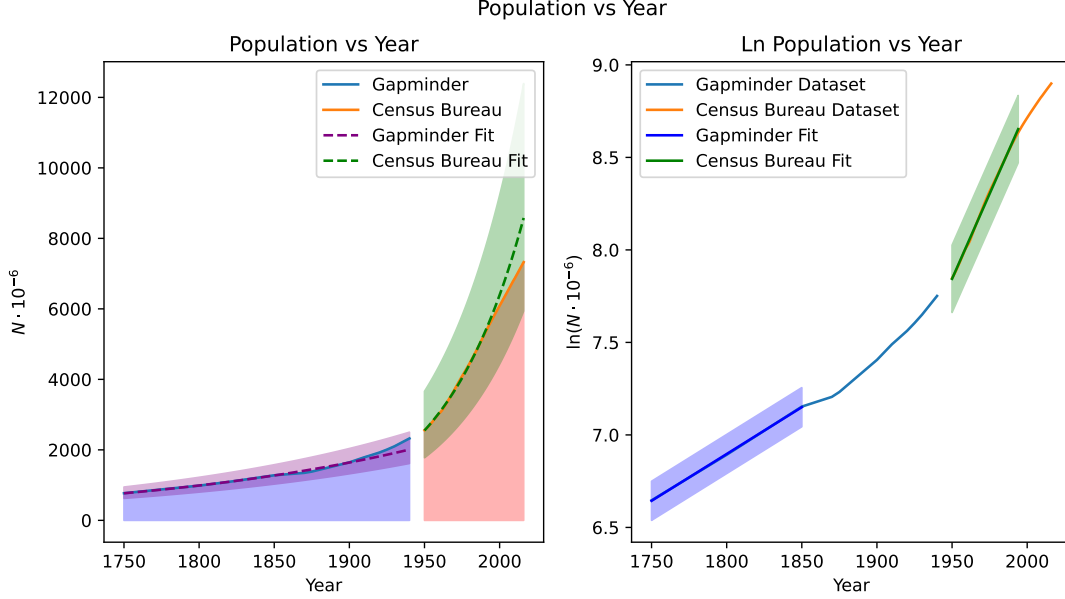

Figure 1:   Plots of Population and natural log of population against time, with the exponential growth fit plotted with approx. 65% confidence intervals in both. Note that the data is contained inside the confidence intervals, but does not follow the fit exactly.

3   Results

From our fit we obtained the following values for a and b, in each subset

Subset i	a_i	b_i	σ_a	σ_b
Subset 1	-2.213	0.005061	0.1056	5.85e-05
Subset 2	-28.01	0.01839	0.1818	9.22e-05

This gives us the corresponding values for n_0 and λ, where n_0 is the population at t_0 for each subset, and λ is the exponential growth rate.

Subset i	n_0	λ	σ_{n_0}	σ_λ
Subset 1	768.0	0.005061	113.0	5.85e-05
Subset 2	2.55e+03	0.01839	651.9	9.22e-05

Clearly, from figure 1, while the total dataset is within the margin of error of the fit of the subset, the data does seem to lag below the perfect exponential. This is because, effectively, the other variables determining

population growth which are assumed to have, on aggregate, no effect in this model (λ constant), are not constant.

In terms of assessing a 'current' population evolution, the fit is likely not very helpful. A more accurate method would perhaps be a machine learning model, using interpolation to increase the dataset size, and modelling many smaller systems (countries, continents) to gain a greater insight into the correlation of population growth with various other variables (GDP, war, education, quality of life, access to medicine, ...).

A more rudimentary method would be to use this model, but to fit some functions (exponential, power law, sigmoidal, ...) to the λ values, to examine how the growth rate may change in the future. Of course in this case, n_0 is the current population, and t_0 is the current year.

In this fashion we made a basic example, setting the current year to 2016, and the current population to the last value in the data. We then fitted a linear function to the λ values, and projected values until 2030.

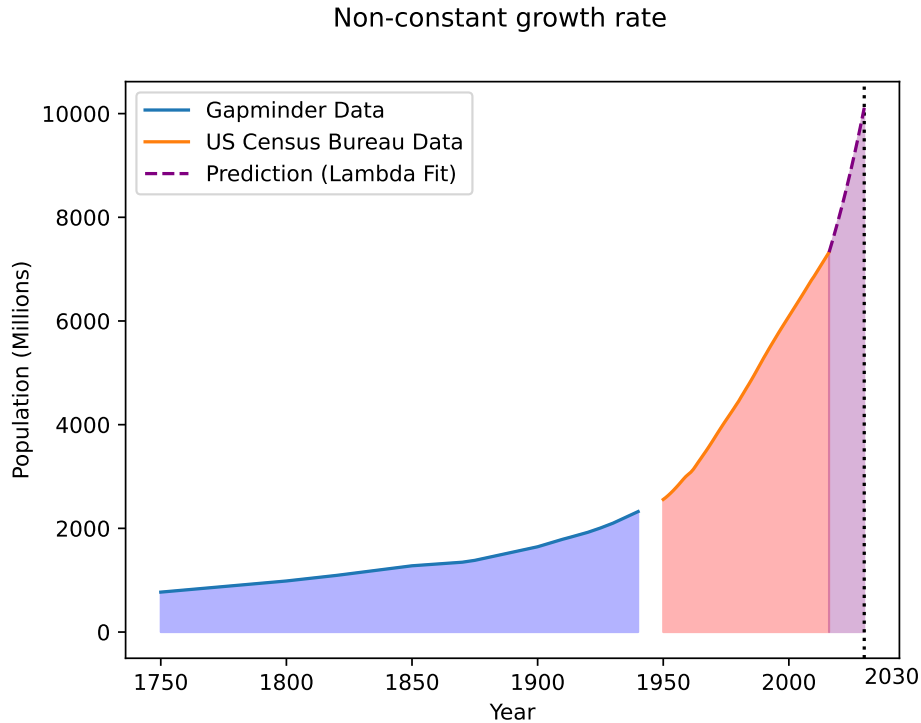## Non-constant growth rate

Figure 2: Plot of the predicted population growth until 2030, as marked, using the linear fit of the λ values. While this is a slight overshoot as of 2022 world population data, it is a reasonable estimate, which would only be made more accurate by a larger dataset.

I would like to mention that while GitHub Copilot was used to assist in the coding involved, it was used exclusively to autofill parts already written for other subsets, and answer questions regarding errors, and not to generate 'new' code itself, nor text for this report.