Construct a population model - 1 Bases

Thomas Laure

November, 2018

Resume :

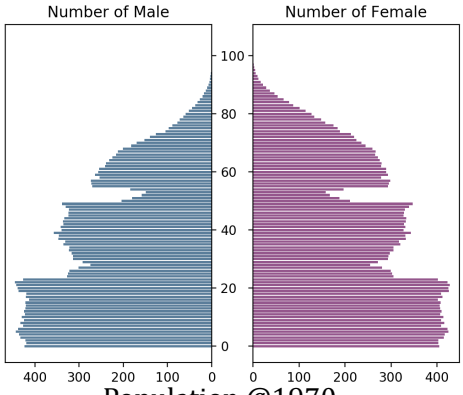
This report describe a first approach employed to project population across time. If this first part only consider male & female by age more characteristics can be added and the technics will be fine tuned. Applications can fit to an important amount of problematics including economic, commercial, statistics…

1. Data

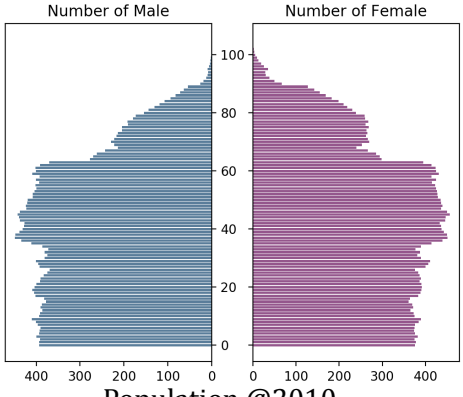
As explained in the resume, this first approach focus on the projection of population according to the sex and age. This data is widely available on internet and for most countries it can be found on the UN website (http://data.un.org/). The exemple used later will be based on the french demographic data from 1901 to 2017. To project the data it is necessary to establish a mortality rate by age as well as a way to project births, it is important to note that for this work migrations will not be taken into account. To perform the projection of the births we will use an additional data which is the evolution of the purchasing power across time, this information is available on the french statistical institute INSEE (https://www.insee.fr).

2. First analysis

The pyramid shape evolve across time, here are two examples on the selected dataset :



Population @1970

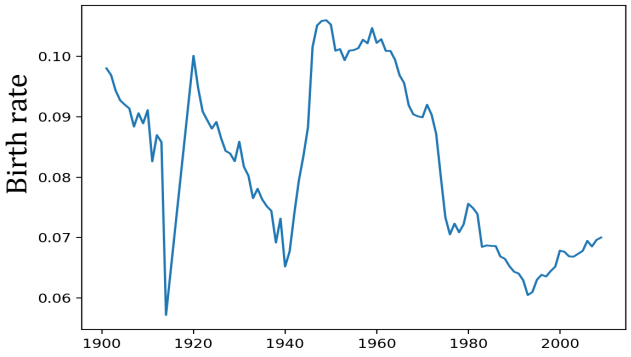


Population @2010

3. Project variables

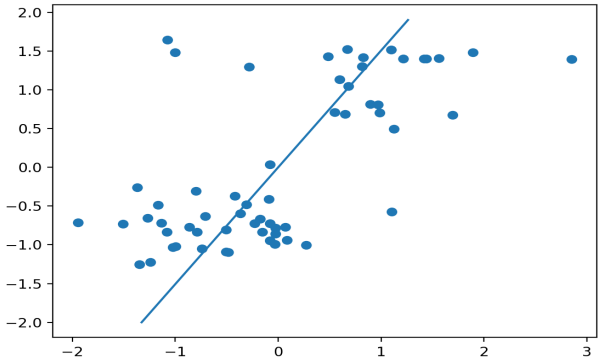
Project the population in this exemple require to project for each year the number of birth as well as the deaths for each age. To calculate the number of births we will use a regression of a rate calculated as the births for a year divided by the average number of man & women in age to give birth, the formula is :

$$BirthRate_n = \frac{Pop_{age=0,year=n+1,sex=(M,F)}}{Mean(Pop_{age\in[18,45],year=n,sex=M}, Pop_{age\in[18,45],year=n,sex=F}}$$



The birth rate is chaotic against time

Predicting the birth rate at a point in time do not seems easy. The first two lows followed by highs are explained by wars but the evolution after seems subject to more external factors. In the first approach, only one external factor will be used : the evolution of the purchasing power. The scatter plot below show the relationship between the two variables.

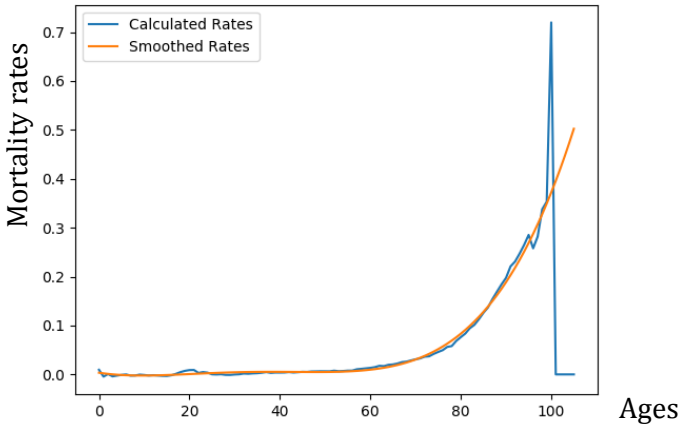


correlation exists however the model can be improved

To project the mortality rates the first approach is also the simplest. It starts by calculating the historical mortality rates with the formula :

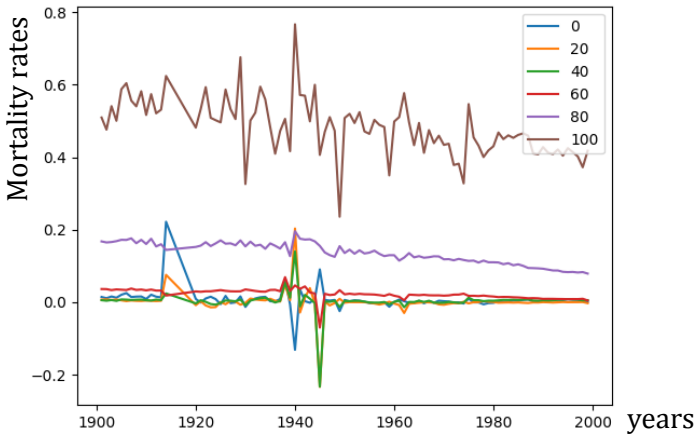
$$MortalityRate_{age=a,year=n} = \frac{Pop_{age=a+1,year=n+1}}{Pop_{age=a,year=n}}$$

The last container of the pyramid data contain the population with an age $\geq age_{limit}$ to correct this effect and have a smoother curve we will use a regression of the mortality rates across ages :



Mortality rates calc from the pyramid @year=2000

To perform the projection of mortality rates we observe the evolution of mortality rates for each age across time.



Mortality rates declines almost linearly in time

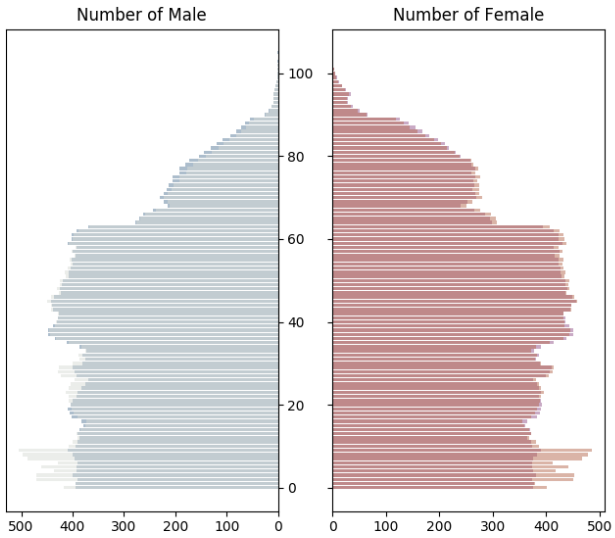
In this first approach mortality rates for the future will be calculated using a linear estimation.

4. Algorithm & results

The code can be summed up as follow :

a. input enter the starting year and the number of year of projection
b. According to the historical data perform a regression of the mortality rate by age
c. Use the regression to project the mortality rates for the future years
d. Calculate the birth rates and the proportion of male vs female births on the historical data. Perform the regression on the birth rate against the evolution of the purchasing power.
e. Using real or estimation of the purchasing power evolution for the years of projection calculate the birth rate.
f. Using the projected data perform a loop on the population starting with the initial year.

Here is the result for a 10 years projection starting from 2000 (light) compared to the actual 2010 population pyramid :



Projected births are , the rest of the structure fit far better

5. Conclusion

As explained this work is a first approach of a more global project. It set up algorithms that will be re-used later for specific improvement of this model or used in other projects. The code is available on GitHub and I stay available for any clarification or to discuss the subject.