Capstone Milestone Report

Introduction

Population projections are routinely used by private, public and non-profit organizations for policy development ("Uses and Limitations", 2016). For example, projections are used in:

- Estimating central and local finance allocation
- Planning housing and land use
- Health care planning: for modelling and projecting health care indicators
- Weighting national and regional surveys
- Creating teacher workforce models at a national and local level
- Assessing ageing population and understanding its implications

Therefore, accurate and high-quality population projections are imperative for good policy-making. **For this assignment, we will be estimating aggregate, age-specific, and sex- specific population projections until 2100 for 200+ countries.**

Fertility, life expectancy and migration statistics are the three main independent variables for estimating population projections (Keneda & Bremner, 2014). Fertility is expressed as total fertility rate (TFR). Total fertility rate is the average number of children that a woman will have in her lifetime if the rate of childbearing in a given year remains unchanged over her entire life (Keneda & Bremner, 2014). Life expectancy at birth is the average number of years a person is expected to live if (s)he were to pass through life subject to age-specific mortality rates of a given period ("Life Expectancy at", 2000). For this study, we only take international migration into account. An international migrant is

defined as a person who is living in a country other than his or her country of birth (Menozzi, 2015).

For this assignment, we obtain data from United Nations Population Division Database (data points from 1950 to 2012). This data is readily available in R software in 'wpp2012' package. This package contains estimates of total fertility rates for each country, sex-specific rates of life expectancy for each country, age-specific and sex-specific estimates for population for each country, age-specific and sex-specific mortality rates for each country, age-specific and sex-specific net migration for each country, sex ratios at birth (as a ratio of female to male) for each country, distribution of age-specific fertility rates for each country, and the location dataset.

In order to estimate population projections for our assignment the following data input is required for each country:

- Sex-specific and age-specific population estimates at the initial time $t = 0$
- Projections of total fertility rate (TFR)
- Projections of fertility distribution over ages
- Projections of sex ratio at birth
- Projections of male and female life expectancy at birth (e0)
- Historical data on sex- and age-specific death rates (for $t \leq 0$)
- Projections of sex- and age-specific net migration

The input projections for fertility, life expectancy and migration mentioned above are generated using data from 'wpp2012' package in R software. After generating the required data input we estimate population projections for each country. The next section will describe this methodology in more detail.

Methodology

As per Sevcikova et al. (2013), population projections are estimated using the following demographic balancing equation:

$$P_{c,t} = P_{c,t-1} + B_{c,t} - D_{c,t} + M_{c,t} \quad \text{——————————————————— (i)}$$

where $P_{c,t}$ is the population prediction in country c at time t,

$P_{c,t-1}$ is the population prediction in country c at time t−1,

$B_{c,t}$ is the number of births in country c at time t,

$D_{c,t}$ the number of deaths in country c at time t,

$M_{c,t}$ net migration in country c at time t.

Theoretically, this equation implies that for any given time period the population of a country is dependant upon population prediction of that country in the previous time period, number of births in the country at that time, number of deaths in the country at that time and the net migration in country at that time.

Our analysis is divided into three main steps. The first two steps are the feature engineering part of the analysis — we estimate fertility and life expectancy parameters for each country. The final step is machine learning part of the analysis — we estimate population projections for each country using the parameters we engineered in the previous two steps.