European Doctoral School of Demography 2019-20

Population Projections

Lecture 1

Ugofilippo Basellini

basellini@demogr.mpg.de

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

ined
INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

Online course
1st June 2020

Presentations

- A little bit about myself ...
- ... and a little bit about you

Miscellaneous

- My first one-week course, virtual teaching, hope that it will be fun for everyone
- Course material available at `https://github.com/ubasellini/EDSD2020-population-projections`
- Feel free to ask questions during the course (please type "question" in the chat) or email me at any time
- Slides inspired by Chapter 6 of Preston et al. (2001) & Guy Abel's EDSD course in 2015-16
- Lectures consist of theoretical slides and small exercises in `R` to make the class more entertaining
- On top of projections, we will learn how to use animations to make nice dynamic visualizations, and how to use `shiny` apps

Assignment

- Group assignment, with three students for each groups
- Six questions (out of seven), which will be presented after each class
- The exercises follow up on what we have touched during the class
- Please email me a single `.pdf` file containing your answers, as well as the `R` codes and data to generate the results. It can also be an Rmarkdown document
- Deadline: Sunday 7^{th} June
- And the groups are: let's draw them in `R`

Brief course summary

Lecture 1:

- introduction to population projections
- first (simple) model of population projections

Lecture 2: cohort component method

Lecture 3: matrix projections

Lecture 4: more matrix projections

Motivations for population projections I

"*Population projection is probably the demographic technique that is most frequently requested by demography's clients*" (Preston et al. 2001)

- Governments: to anticipate demands for schools, roads, medical personnel, ...
- Private businesses: potential size of their future market
- Public and private sector: future pension expenditures

$\Rightarrow$ more generally, they are the base for many other projections that are essential for social and economic planning, e.g. labour, education, health, housing, social security, energy, climate, transport, ...

Motivations for population projections II

Demographers use population projections for a variety of needs:

- to analyse the implications of a certain set of demographic parameters on population size, composition and growth
- to help understand the determinants of population change and inform policy discussions
- to illustrate the implications of certain demographic characteristics (the model's *inputs*) on population parameters over time (the model's *outputs*)
- typical inputs: fertility, mortality and migration - although methodology can be extended to different demographic processes (marriage patterns, contraceptive use regimes, ...)
- to identify realistic goals and targets for future development trends

Projections vs forecasts I

"*Population projections are calculations which show the future development of a population when certain assumptions are made about the future course of fertility, mortality and migration. They are in general purely formal calculations, developing the implications of the assumptions that are made.*" (United Nations 1958)

- quality of projections is determined by their *internal* validity: do projections obey fundamental demographic accounting relations?
- can be used to answer "*what if*" questions, i.e. purely hypothetical scenarios corresponding to different assumptions. For example, what would happen to the population:
  - *if* mortality had not changed in the last ten years
  - *if* fertility had been 10% lower
  - *if* ...

Projections vs forecasts II

"*A population forecast is a projection in which the assumptions are considered to yield a realistic picture of the probable future development of a population*" (United Nations 1958)

- quality of forecasts is determined by their *external* validity: are the predictions of subsequent events accurate?
- important feature of forecasts is its **uncertainty**
- stochastic forecasts are characterized by probabilistic distributions from which one can compute:
  - central (point) forecast
  - uncertainty via prediction interval

Projections vs forecasts III

In Nathan Keyfitz (1972) words: "*Population projections are correct beyond any testing against the subsequent population performance; in fact they can be incorrect only in the trivial sense that the author made an arithmetic error that prevents his final numbers from being consistent with his initial assumptions.*"

"*Yet in fact projections and forecasts are not easily distinguished. ... A demographer makes a projection, and his reader uses it as a forecast*"

- this course: population projections **only**
- next week: forecasting with Prof. Heather Booth

Population projection methodology

A desirable methodology should:

- obey the demographic accounting relationships, i.e. balancing equation of population growth
- be implementable with the available inputs
- provide the desired level of detail in the output
    - only total population
    - disaggregated by sex
    - by age
    - by age and sex
    - 1y or 5y intervals
    - including additional information (education, ethnicity, urban/rural, ...)

The constant exponential growth model

Given the total population at time 0, $N(0)$, and the mean annualized growth rate between 0 and T, $\bar{r}[0, T]$, what is the total population at time T, $N(T)$?

$$N(T) = N(0)e^{\int_0^T r(t)dt} = N(0)e^{\bar{r}[0,T]T} \tag{1}$$

Let's simplify things: assume that population growth rate is constant over time, i.e. $\bar{r}[0, T] = r$ for all T. Then, using the first two terms of the exponential Taylor series:

$$N(T) = N(0)e^{rT} = N(0)(1 + r)^T \tag{2}$$

The constant exponential growth model - exercise

Open your `R` session.

Exercise

Assume two populations with $N_a(0) = 100$, $N_b(0) = 90$, $r_a = 0.05$ and $r_b = 0.10$. After how many years will population b surpass a? Plot your results.

Reminder:

$$N(T) = N(0)(1 + r)^T$$

The constant exponential growth model - one possible solution

Example

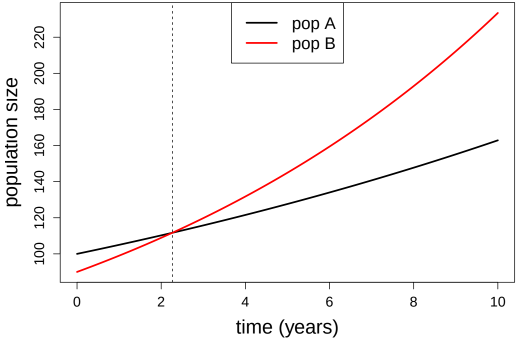
```
rm(list=ls(all=TRUE))
N0_a <- 100
N0_b <- 90
t <- seq(0,10,0.01)
r_a <- 0.05
r_b <- 0.10
PopProj <- function(N0,r,t){
NT <- N0 * (1+r)^t
return(NT)
}
Na <- PopProj(N0_a,r_a,t)
Nb <- PopProj(N0_b,r_b,t)
t_hat <- t[which(Nb > Na)[1]]
```

$$\Rightarrow \hat{t} = 2.27 \text{ years}$$

One possible solution (contd)

Example

```
plot(t,Na,t="l",lwd=2,ylim=range(Na,Nb),
xlab="time (years)",ylab="population size")
lines(t,Nb,col=2,lwd=2)
abline(v=t_hat,lty=2)
legend("top",c("pop A","pop B"),col=1:2,lwd=2)
```



Adding demographic components

Let $B[0,T], D[0,T]$ and $M[0,T]$ denote then number of births, deaths and net migrations between 0 and T. From the balancing equation of population growth:

$$N(T) = N(0) + B[0,T] - D[0,T] + M[0,T], \tag{3}$$

and dividing by the number of person-years lived, $PY[0,T]$, we get:

$$\begin{array}{ccccccc} \dfrac{N(T)-N(0)}{PY[0,T]} & = & \dfrac{B[0,T]}{PY[0,T]} & - & \dfrac{D[0,T]}{PY[0,T]} & + & \dfrac{M[0,T]}{PY[0,T]} \\ r & = & b & - & d & + & m \end{array} \tag{4}$$

i.e. we can decompose the crude growth rate r in terms of the three demographic components, i.e. birth (b), death (d) and net migration (m) crude rates.

Adding demographic components - exercise

Exercise

Assume again two populations with $N_a(0) = 100$, $N_b(0) = 90$, $b_a = 0.05$, $d_a = 0.07$ and $m_a = 0.01$, and $b_b = 0.10$, $d_b = 0.05$ and $m_b = 0.03$. After how many years will population b surpass a?

Hint: substitute $r = b - d + m$ into

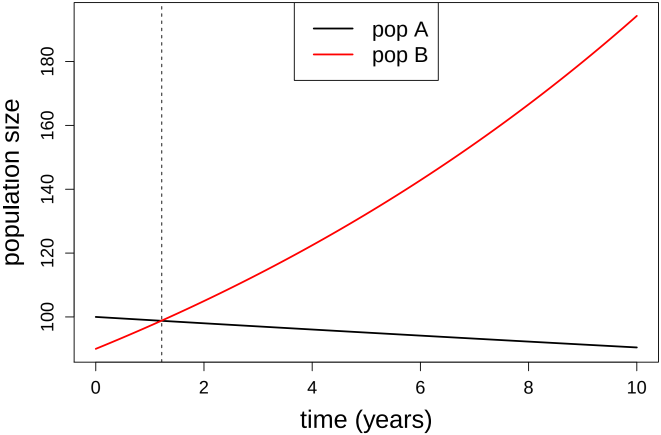
$$N(T) = N(0)(1 + r)^T$$

Adding demographic components - one possible solution

Example

```
N0_a <- 100
N0_b <- 90
t <- seq(0,10,0.01)
b_a <- d_b <- 0.05
d_a <- 0.07
m_a <- 0.01
b_b <- 0.10
m_b <- 0.03
PopProj <- function(N0,b,d,m,t){
r <- b - d + m
NT <- N0 * (1+r)^t
return(NT)
}
Na <- PopProj(N0_a,b_a,d_a,m_a,t)
Nb <- PopProj(N0_b,b_b,d_b,m_b,t)
t_hat <- t[which(Nb > Na)[1]]
```

Adding demographic components - one possible solution



$$\Rightarrow \hat{t} = 1.22 \text{ years}$$

From net migration rates to counts

Rogers, A. (1990). *Requiem for the net migrant*. Geographical Analysis, 22(4), 283-300:

- the net migrant is "a nonexistent category of individuals"
- net migration rate is a flawed measure: it is the difference between a true rate (out-migration) and a prevalence (in-migration). Denominators are different!
- they obscure regularities in age profiles of migration

Use net migration counts instead:

$$N(t+1) = N(t)(1 + b - d) + M \tag{5}$$

or immigration counts (I) and emigration rates (e):

$$N(t+1) = N(t)(1 + b - d - e) + I \tag{6}$$

From net migration rates to counts - exercise

Exercise

Take a population with $N(0) = 100$, $b = 0.03$, $d = 0.02$ and $e = 0.005$ and $I = 0.75$. What is the projected population after 20 years?

Hint: write a `for` loop in your PopProj function for the formula

$$N(t+1) = N(t)(1 + b - d - e) + I$$

From net migration rates to counts - one possible solution

Example

```
N0 <- 100
t <- 1:21   ## one more than projection period
b <- 0.03
d <- 0.02
e <- 0.005
I <- 0.75
PopProj <- function(N0,b,d,e,I,t){
r <- b - d - e
NT <- rep(NA,t)
NT[1] <- N0
for (i in 2:t){
NT[i] <- NT[i-1] * (1+r) + I
}
return(NT)
}
NT <- PopProj(N0,b,d,e,I,max(t))
NT[max(t)]
```

Incorporating future assumptions

We can extend formula

$$N(t+1) = N(t)(1 + b - d - e) + I \tag{7}$$

to incorporate future assumptions on demographic components

$$N(t+1) = N(t)(1 + b_{t,t+1} - d_{t,t+1} - e_{t,t+1}) + I_{t,t+1} \tag{8}$$

This allows us to experiment different assumptions in order to answer "*what if*" questions (the aim of projections).

Incorporating future assumptions - exercise

Exercise

Take the same population with values $N(0) = 100$, $b = 0.03$, $d = 0.02$ and $e = 0.005$ and $I = 0.75$, and project again for 20 years ahead. However, make two assumptions about migration. After ten projected years: (i) the emigration rate e doubles, and (ii) the immigration counts I converge to zero at the last projected year. Compare this scenario with the previous projection.

Hint: within your `for` loop, feed vectors rather than constant numbers

Incorporating future assumptions - one possible solution

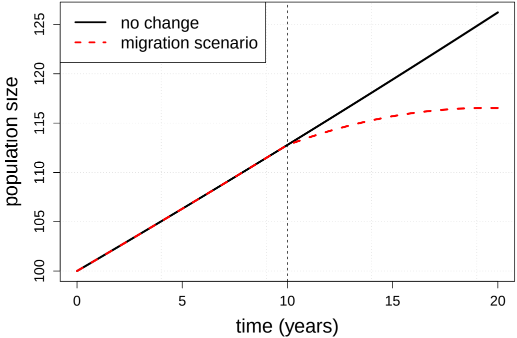
Example

```
e <- c(rep(0.005,11),rep(0.01,10))
I <- c(rep(0.75,11),seq(0.75,0,length.out = 10))
PopProj <- function(N0,b,d,e,I,t){
NT <- rep(NA,t)
NT[1] <- N0
for (i in 2:t){
r <- b - d - e[i]
NT[i] <- NT[i-1] * (1+r) + I[i]
}
return(NT)
}
NT_scenario <- PopProj(N0,b,d,e,I,max(t))
```

Incorporating future assumptions - one possible solution

Example

```
plot(t,NT,t="n",xlab="time (years)",ylab="population size",axes="F")
axis(1,at=c(1,6,11,16,21),labels = c(0,5,10,15,20))
axis(2,cex.axis=1.25);grid();box()
lines(t,NT,lwd=2)
lines(t,NT_scenario,col=2,lwd=2,lty=2)
abline(v=11,lty=2)
legend("topleft",c("no change","migration scenario"),col=1:2,lwd=2,lty=c(1,2))
```



Some final remarks

- Reminder: crude rates are very sensitive to age-compositional effects
- The constant exponential growth model shown today is sensible when we can assume:
  - demographic components do not vary much over projection period, *and*
  - the age distribution remains constant
- however, the age distribution itself is determined by population change. An improved projection methodology should take into account age effects by modeling the age distribution as by-product of fertility, mortality and migration conditions

$\Rightarrow$ cohort component method (tomorrow's lecture)

Assignment

Exercise #1

Take the population time-series for any country (data from HMD or UN), and answer to the following three questions:

- Project the population 20-year ahead using the constant exponential model (baseline projection)
- Make an assumption about one of the demographic components, and compare this projection with the baseline one
- Assume that your observed data ends 5 years before the last available data point, and project the population for 5 years. Are the projected values close to the observed ones?

References

- Keyfitz, N. (1972). On Future Population. *Journal of the American Statistical Association*, **67**(338), 347–363
- Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.
- Rogers, A. (1990). Requiem for the Net Migrant. *Geographical Analysis*, **22**(4), 283–300.
- United Nations (1958). *Multilingual Demographic Dictionary*. United Nations Population Studies, No. 29