Homework 06

Ihsan Kahveci

2022-05-23

Contents

1 Questions

1.1 *Q1*

1.1.1 *Q1.a*

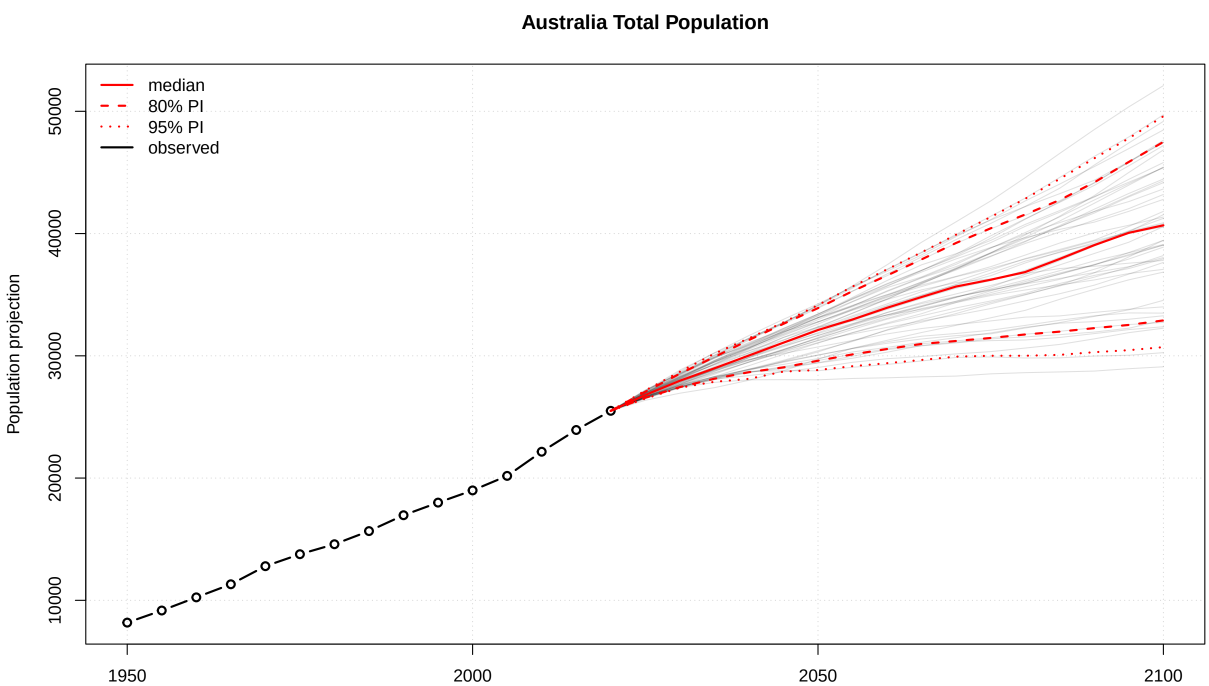
In `bayes.pop`, One can specify the quantities derived from population projections using `pop.expressions`.

> As another example, the potential support ratio can be defined as "PFR[5:13] / PFR[14:27]". (Ševčíková, Raftery 2016)

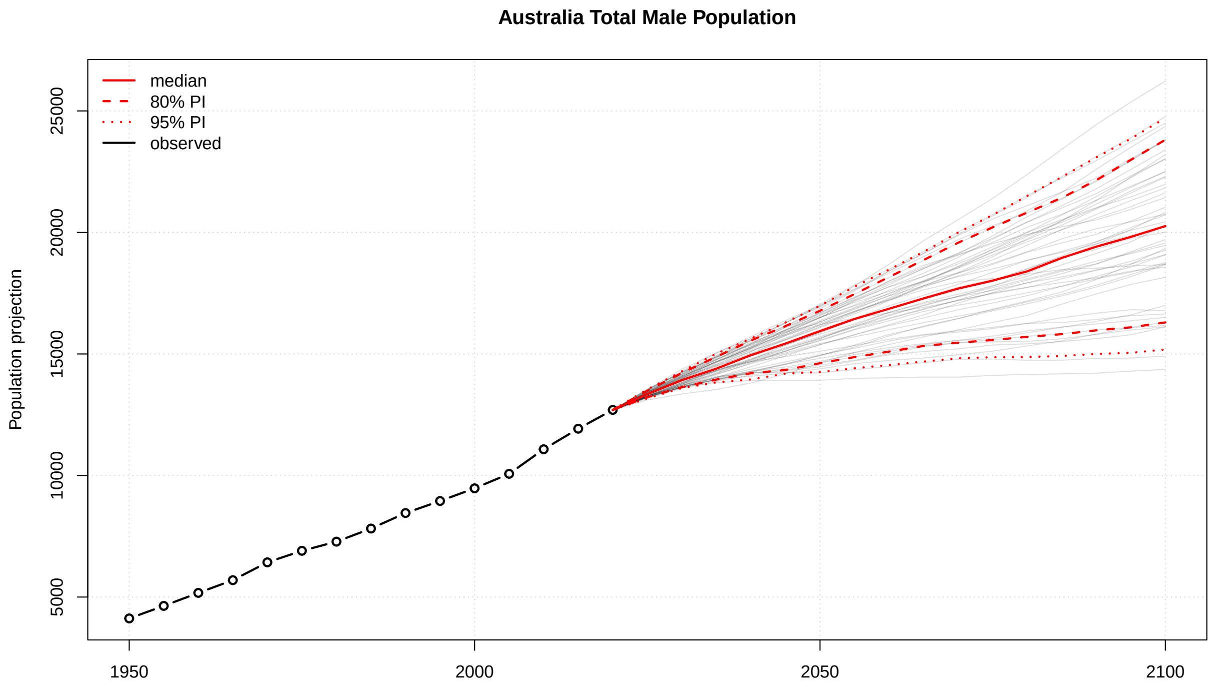
Ševčíková, H., & Raftery, A. E. (2016). bayesPop: Probabilistic Population Projections. Journal of Statistical Software, 75(5), 1–29. https://doi.org/10.18637/jss.v075.i05

Using our converged life expectancy at birth and total fertility rate simulations, we can create probabilistic projections of the following population quantities:

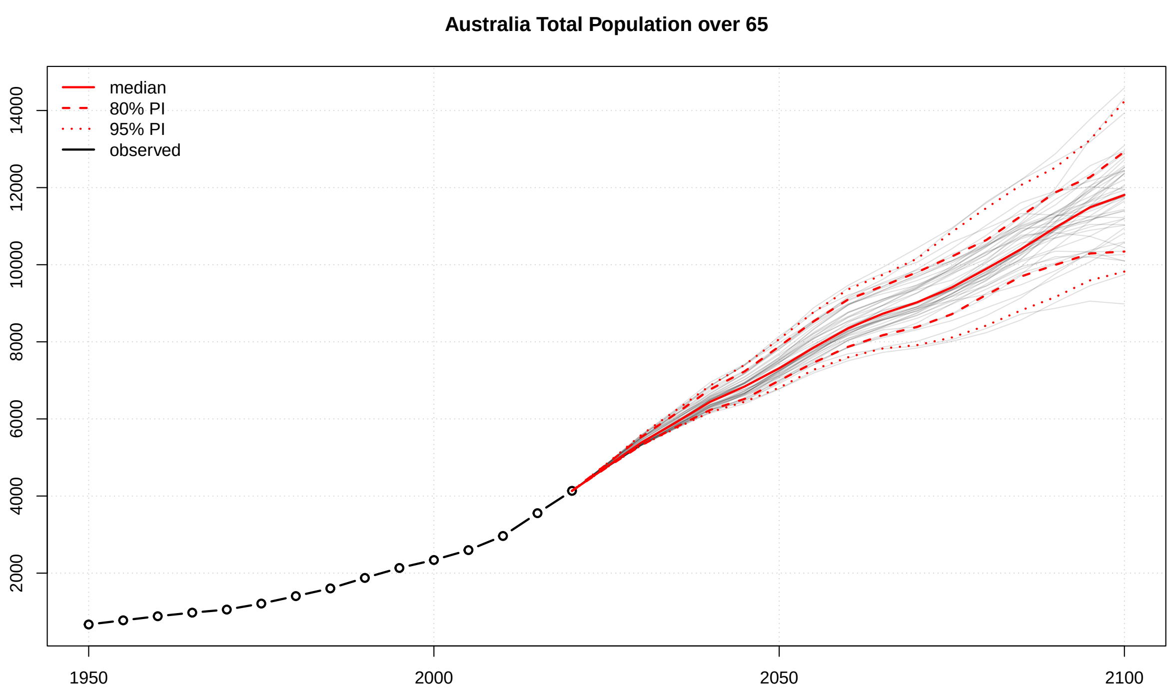
Note that the "Potential Support Ratio" is defined as: $\frac{\textit{people aged 20-64}}{\textit{people aged 65 and over}}$



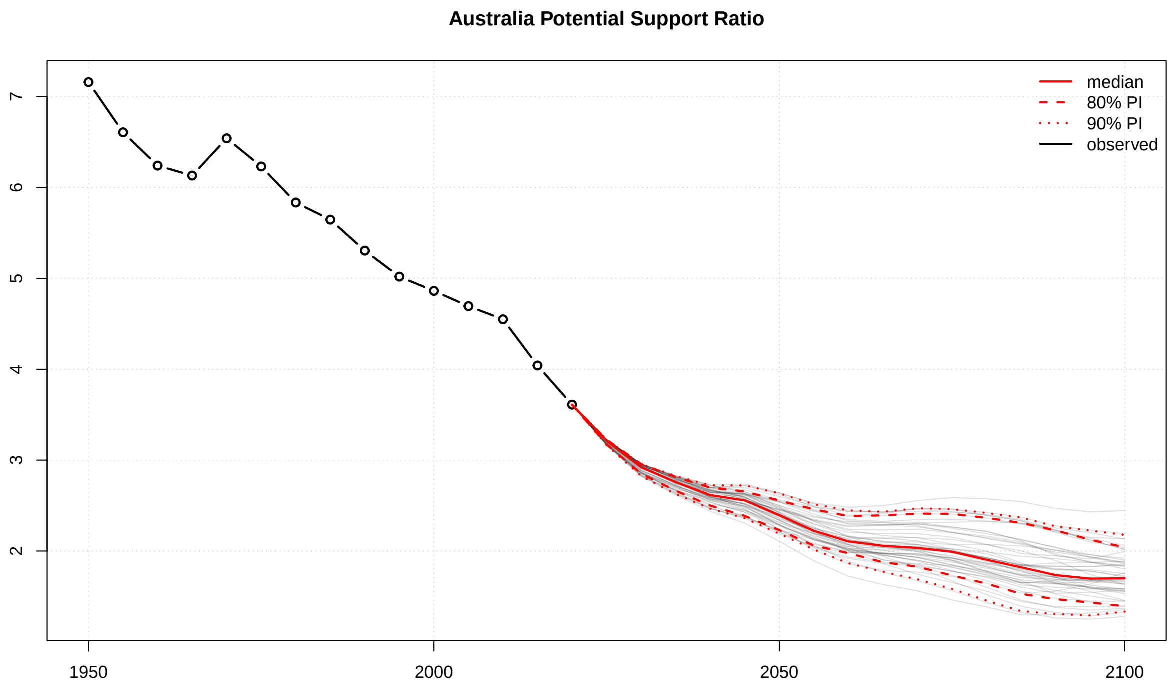
1.1.2 *Q1.b*



1.1.3 *Q1.c*



1.1.4 *Q1.d*



Looking at the projections of potential support ratio, we see continued declines into the future, though eventually at a slower rate. This indicates Australia's age structure is likely to continue shifting to older ages. From a social security perspective, this means that Australia's number of workers per retiree is going to decrease from 2020-2100.

1.2 *Q2*

1.2.1 *Q2.a*

Table 1: Crude net migration rate (CNMR) for Australia, 1950-2100

Period start	CNMR
1950	9.134
1955	8.235
1960	7.132
1965	13.377
1970	3.908
1975	3.325
1980	5.997
1985	8.021
1990	4.018
1995	4.201
2000	6.018
2005	11.441

Period start	CNMR
2010	8.587
2015	6.403

1.2.2 *Q2.b*

I fit an order 1 autoregressive model to the subset of Australia Crude Net Migration Rates obtained from UN, and extract some model parameters below. *Note that the AR(1) model was fit using the "mle" method.*

Table 2: AR(1) Model for Australia CNMR, 1950-2020

	Model 1
ar1	0.07 [−0.45, 0.58]
intercept	7.13 [5.56, 8.71]
Num.Obs.	14
AIC	74.8
BIC	76.7
Log.Lik.	−34.385

1.2.3 *Q2.c*

Using these model parameters, we can find an analytic solution for the predictive probability distribution of net migration rates in Australia for 2020-2025, which takes the form:

$$(F_{t+1} - \mu) = \rho * (F_t - \mu) + \epsilon$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (\hat{y_i} - y_i)^2}{N}$$

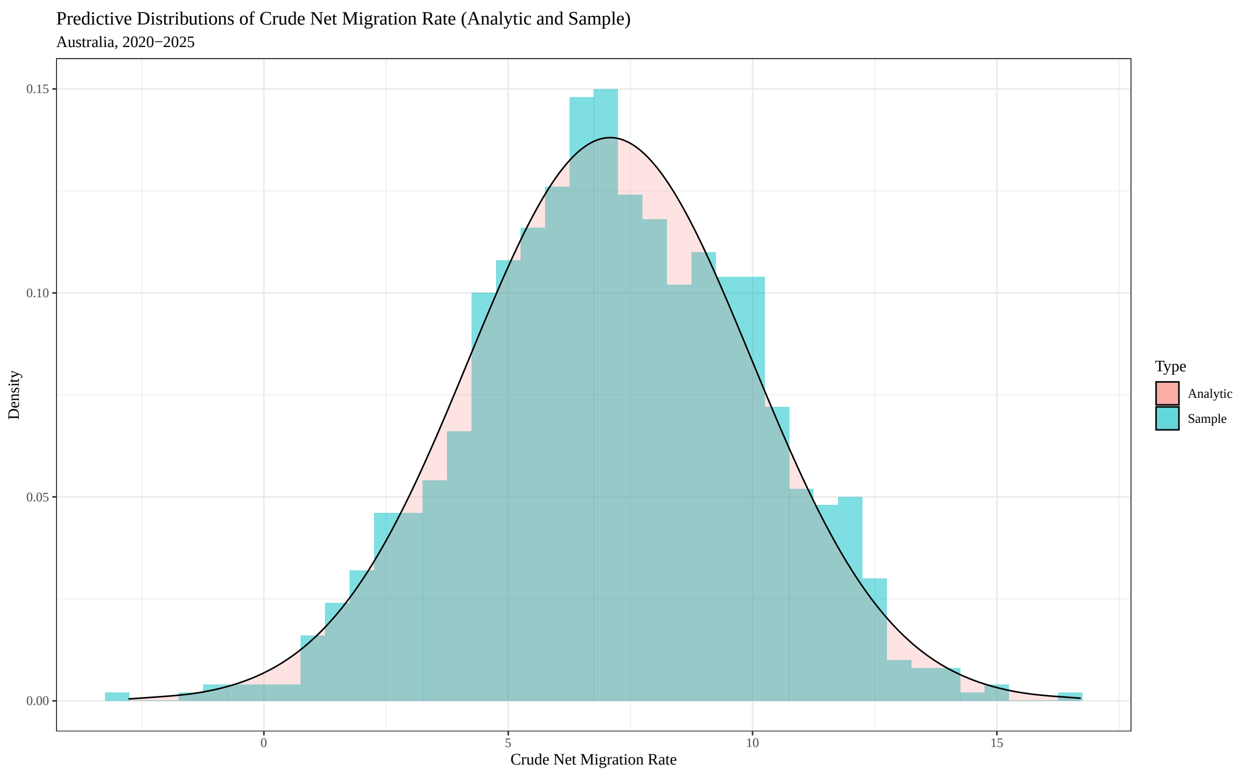
- $\sigma^2 = 7.9562007$
- $F_{2020} = 7.0860154$

Table 3: Predictive Distribution of Australia Crude Net Migration Rate, 2020-2025

Mean	Median	2.5% CI	97.5% CI
7.086	7.086	1.557	12.615

1.2.4 *Q2.d*

We can also sample from this distribution to show it does indeed follow what we predict analytically:



1.2.5 *Q2.e*

Assuming that age and sex distribution of migrants has not changed from 2012, which is the last time the UN reported age/sex-specific net migration rates.

Based on age distribution of net migration stocks in 2012, I decomposed the projected migration rate for 2020 by age groups using the formula:

$$\text{ASMR}_{20} = \text{PASMR}_{12} * \text{CNMR}_{20}$$

Table 4: Predicted Age-spesific Migration Rates (ASMR) for Australia, 2020-2025

Age Grop	PASMR	ASMR
0-4	0.11	0.76
5-9	0.10	0.69
10-14	0.10	0.74
15-19	0.12	0.87
20-24	0.14	0.97
25-29	0.13	0.93
30-34	0.12	0.82
35-39	0.09	0.65
40-44	0.06	0.42
45-49	0.03	0.22
50-54	0.01	0.07
55-59	0.00	0.01
60-64	0.00	-0.02
65-69	0.00	-0.02
70-74	0.00	-0.01
75-79	0.00	0.00
80-84	0.00	0.00
85-89	0.00	0.00
90-94	0.00	0.00
95-99	0.00	0.00
100+	0.00	0.00

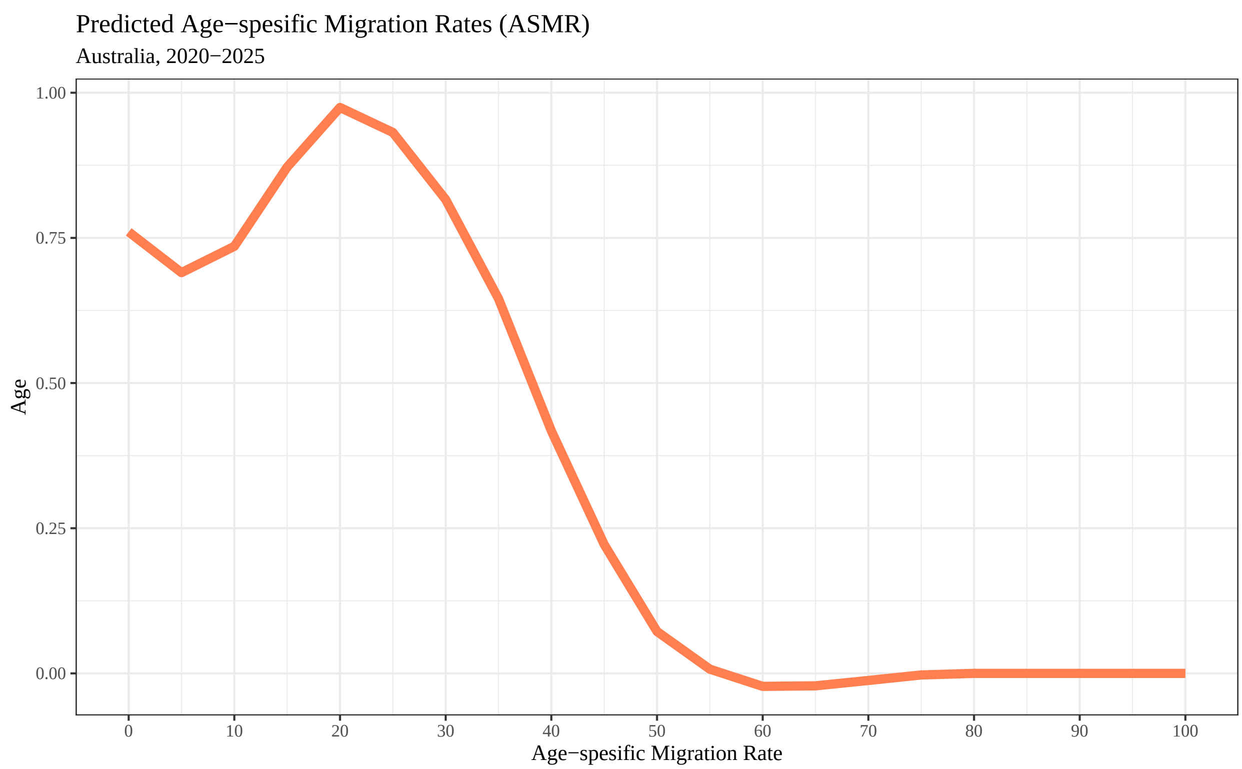


Table 5: Rogers-Castro Model Age-spesific Migration Rates (ASMR) for Australia, 2020-2025

Age Grop	ASMR
0-4	0.13
5-9	0.09
10-14	0.06
15-19	0.08
20-24	0.23
25-29	0.21
30-34	0.15
35-39	0.10
40-44	0.07
45-49	0.05
50-54	0.04
55-59	0.03
60-64	0.03
65-69	0.02
70-74	0.02
75-79	0.02
80-84	0.02
85-89	0.02
90-94	0.02
95-99	0.02
100+	0.02

1.2.6 *Q2e* Alternative

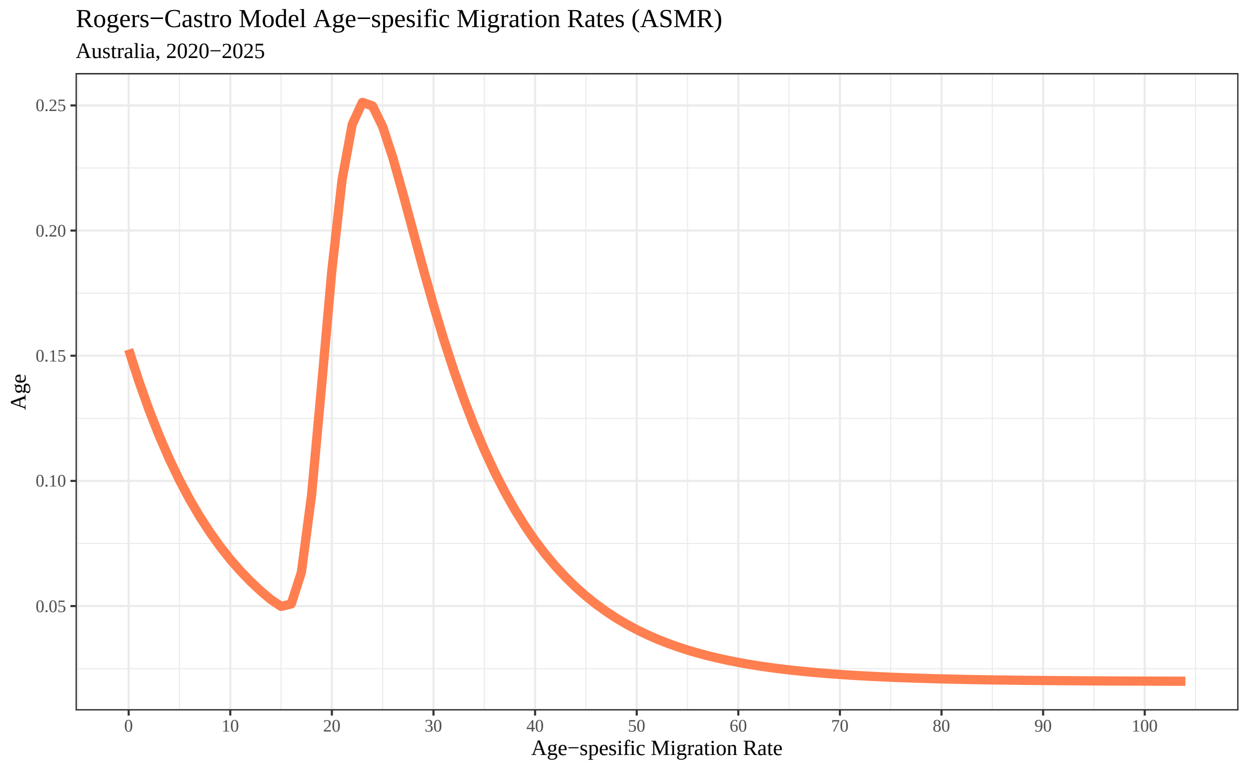
Migration schedules without a retirement peak may be represented by a "reduced" model with seven parameters, because in such instances the *post-labor force* component of the model is omitted (Rogers & Castro 1981). The simplified 7 parameter model:

$$M(x) = a_1 e^{-\alpha_1 x} + a_2 e^{-\alpha_2 (x-\mu) - e^{-\lambda(x-\mu)}} + c$$

`migest` package provides fundamental parameters for Rogers-Castro migration schedule. I plug them into the equation above:

$$M(x) = 0.02 e^{-0.1x} + 0.06 e^{-0.1(x-20) - e^{-0.4x-20)}} + 0.003$$

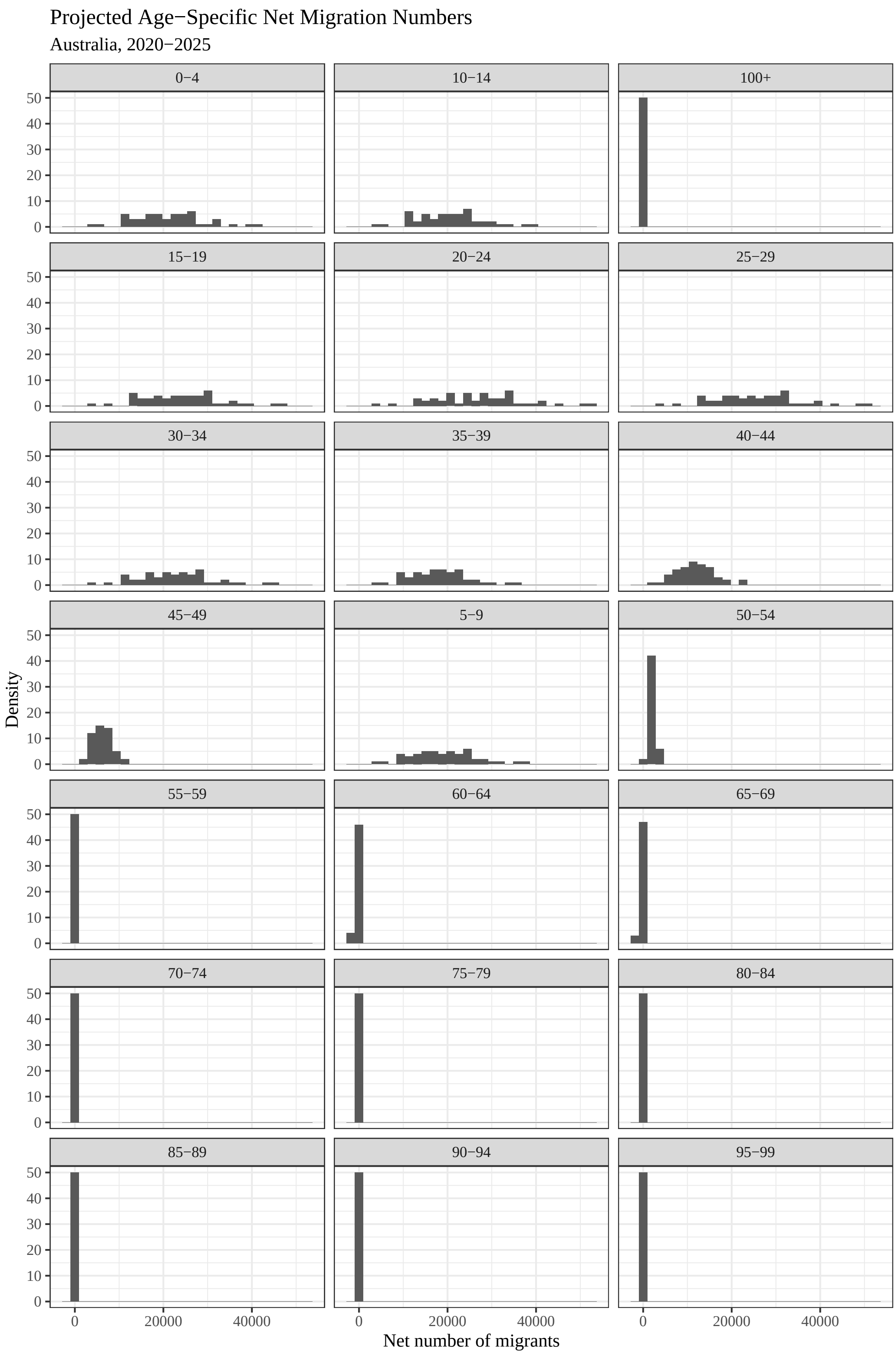
The result is modeled age schedule for migration rates. Hence, if we multiply the resulting vector with the predicted migration rate for 2020-2025, we get age-spesific migration rates for 2020-2025. We need to standardize the output so that it sums to 1.


Rogers−Castro Model Age−spesific Migration Rates (ASMR)
Australia, 2020−2025
0.25
0.20
0.15
0.10
0.05
Age
0
10
20
30
40
50
60
70
80
90
100
Age−spesific Migration Rate

1.2.7 *Q2.f*

For this question, I created 50 different migration rate estimates for 2025 based on the AR(1) model *Q2.c*. Then, using the same age schedule from *Q2.e*, I created age-spesific migration rates for each sample. Here is the pseudo-equation:

$$\text{ASMR}_{2025} = \text{TotalPop}_{2020} \times \text{NMR}_{2025} \times \text{PASMR}_{2012}$$
$$\text{NMR}_{20205} = N(\mu = 7.0860154, \sigma = 2.8206738)$$



Projected Age−Specific Net Migration Numbers
Australia, 2020−2025
0−4
10−14
100+
15−19
20−24
25−29
30−34
35−39
40−44
45−49
5−9
50−54
55−59
60−64
65−69
70−74
75−79
80−84
85−89
90−94
95−99
Density
0
10
20
30
40
50
0
20000
40000
Net number of migrants

1.2.8 *Q2.g*

We will use the same code we used to generate population projections in *Q1*, except we will plug our migration estimates into the `pop.predict` function. `pop.predict` function has a built-in function to disagregated total net migration into age-spesific migration by applying a Rogers-Castro schedule. Hence, we only need to plug the total migration estimates for 2025.

The challenge in this question is to transform our predicted trajectories into a format that `pop.predict` function understands. Here is the relevant section from the vignette:

> migMtraj, migFtraj, migtraj Comma-delimited CSV file with male/female age-specific migration trajectories, or total migration trajectories (migtraj). If present, it replaces deterministic projections given by the mig* items. It has a similar format as e.g. e0M.file with columns "LocID", "Year", "Trajectory", "Age" (except for migtraj) and "Migration". For a five-year simulation, the "Age" column must have values "0-4", "5-9", "10-14", ..., "95-99", "100+". In an annual simulation, age is given by a single number between 0 and 100.

Australia Country Code: 36

Table 6: Predicted MalePopulation for Australia, 2020-2025

Age	Median	2.5% PI	97.5 PI
0-4	817.36	648.97	992.38
5-9	889.81	889.81	889.97
10-14	874.63	874.63	874.73
15-19	863.87	863.87	864.07
20-24	817.81	817.81	818.20
25-29	848.48	848.48	848.97
30-34	941.95	941.95	942.54
35-39	964.47	964.46	965.22
40-44	935.42	935.42	936.44
45-49	813.85	813.85	815.19
50-54	824.41	824.41	826.54
55-59	756.04	756.03	759.10
60-64	746.22	746.21	750.96
65-69	663.26	663.25	670.00
70-74	559.24	559.22	568.36
75-79	465.68	465.66	478.11
80-84	291.47	291.44	303.82
85-89	158.28	158.26	168.23
90-94	63.69	63.68	68.86
95-99	15.33	15.33	16.66
100-104	1.54	1.54	1.65
105-109	0.05	0.05	0.05
110-114	0.00	0.00	0.00
115-119	0.00	0.00	0.00
120-124	0.00	0.00	0.00
125-129	0.00	0.00	0.00
130+	0.00	0.00	0.00

The following projections include estimated migration counts for 2025. This is for the male population by age:

This is for the female population by age:

Table 7: Predicted Female Population for Australia, 2020-2025

Age	Median	2.5% PI	97.5 PI
0-4	777.14	617.62	943.11
5-9	844.68	844.64	844.78
10-14	831.95	831.93	832.01
15-19	823.16	823.12	823.25
20-24	783.85	783.79	783.99
25-29	822.12	822.06	822.27
30-34	905.34	905.27	905.54
35-39	962.99	962.88	963.29
40-44	937.72	937.55	938.18
45-49	824.73	824.48	825.37
50-54	842.92	842.51	843.98
55-59	782.47	781.88	783.98
60-64	770.09	769.18	772.39
65-69	707.12	705.80	710.43
70-74	605.20	603.35	609.83
75-79	516.66	514.03	523.31
80-84	345.36	342.52	352.72
85-89	211.23	208.49	218.57
90-94	107.72	105.80	113.04
95-99	34.06	33.37	36.02
100-104	5.03	4.94	5.27
105-109	0.25	0.25	0.26
110-114	0.00	0.00	0.00
115-119	0.00	0.00	0.00
120-124	0.00	0.00	0.00
125-129	0.00	0.00	0.00
130+	0.00	0.00	0.00

2 Appendix

```r
# Prep work ------------------------------------------------------------

# Load libraries
library(tidyverse)
library(bayesLife)
library(bayesPop)
library(migest)

# Data folders
e0.dir <- "../data/e0/sim03092016"
tfr.dir <- "../data/tfr/sim01192018"
pop.dir <- "../data/pop/sim05222022"
mig.file <- "../data/WPP2019_Period_Indicators_Medium.csv"


# Control randomness
set.seed(57)
options(scipen = 999)

# Question 1 -----------------------------------------------------------

# Run this once to get pop predictions
# pop.pred <- pop.predict(
#                 end.year = 2100, start.year = 1950, present.year = 2020,
#                 wpp.year = 2019, output.dir = pop.dir, nr.traj = 50,
#                 inputs = list(tfr.sim.dir = tfr.dir,
#                               e0F.sim.dir = e0.dir,
#                               e0M.sim.dir = "joint_"),
#                 keep.vital.events = FALSE, replace.output=TRUE)

pop_sim_pred <- get.pop.prediction(pop.dir)


# Question 1a ----------------------------------------------------------

# Defined using `?pop.expressions`
over65_exp <- "PAU[14:27]"
support_exp <- "PAU[5:13] / PAU[14:27]"

pop.trajectories.plot(
  pop_sim_pred, "Australia",
  sex = "both",
  sum.over.ages = TRUE,
  main = "Australia Total Population"
)
# Question 1b ----------------------------------------------------------

pop.trajectories.plot(
  pop_sim_pred, "Australia",
  sex = "male",
  sum.over.ages = TRUE,
```

```r
  main = "Australia Total Male Population"
)

# Question 1c -----------------------------------------------------------

pop.trajectories.plot(
  pop_sim_pred, "Australia",
  expression = over65_exp,
  sex = "both",
  sum.over.ages = TRUE,
  main = "Australia Total Population over 65"
)

# Question 1d -----------------------------------------------------------

pop.trajectories.plot(
  pop_sim_pred, "Australia",
  expression = support_exp,
  sex = "both",
  sum.over.ages = TRUE,
  main = "Australia Potential Support Ratio",
  show.legend = FALSE)

legend("topright", c("median",  "80% PI", "90% PI", "observed"),
       col = c("red", "red", "red", "black"),
       lty = c(1,2,3, 1), cex = 1, bty = "n", lwd = c(2,2,2,2))


# Question 2 ------------------------------------------------------------

au_mig <- read_csv(mig.file) %>%
  filter(Location == "Australia") %>%
  transmute(year = MidPeriod - 3, nmr = CNMR, nm = NetMigrations) %>%
  filter(year < 2020)


# Question 2a -----------------------------------------------------------
au_mig %>%
  select(-nm) %>%
  knitr::kable(
  booktabs = TRUE, digits = 3,
  col.names = c("Period start", "CNMR"),
  caption = "Crude net migration rate (CNMR) for Australia, 1950-2100"
)

# Question 2b -----------------------------------------------------------
au_ts <- au_mig %>%
  pull(nmr) %>%
  ts()

au_model = arima(au_ts, order=c(1,0,0))
modelsummary::msummary(au_model,
                       estimate = "{estimate} [{conf.low}, {conf.high}]",
```

```r
                        statistic = NULL, fmt = 2, output = "kableExtra",
                        caption = "AR(1) Model for Australia CNMR, 1950-2020")  %>%
  kableExtra::kable_styling(latex_options = "hold_position")

# Question 2c -------------------------------------------------------------

au_mig_last = au_mig$nmr[length(au_mig$nmr)]
au_sd = sqrt(au_model$sigma2)
au_mean = au_model$coef["intercept"]
au_ar1 = au_model$coef["ar1"]
au_mean_pred = unname(au_ar1) * (au_mig_last - au_mean) + au_mean
au_pred_dist <- qnorm(seq(.001, .999, .001), mean = au_mean_pred, sd = au_sd)

au_pred_tbl <- tibble(
  Mean = au_mean_pred,
  Median = au_mean_pred,
  `2.5% CI` = Mean - 1.96 * au_sd,
  `97.5% CI` = Mean + 1.96 * au_sd)
knitr::kable(
  au_pred_tbl,
  booktabs = TRUE,
  digits = 3,
  caption = "Predictive Distribution of Australia Crude Net Migration Rate, 2020-2025") %>%
  kableExtra::kable_styling(latex_options = "hold_position")


# Question 2d -------------------------------------------------------------

au_pred_dist <- qnorm(seq(.0005, .9995, .001), mean = au_mean_pred, sd = au_sd)

au_2020_sample <- rnorm(1000, mean = au_mean_pred, sd = au_sd)

au_2020_tbl <-
  tibble(
    Analytic = au_pred_dist,
    Sample = au_2020_sample
  ) %>%
  pivot_longer(everything(), names_to = "dist", values_to = "value")

ggplot(au_2020_tbl, aes(x = value, fill = dist)) +
  geom_histogram(
    data = filter(au_2020_tbl, dist == "Sample"),
    aes(y = ..density..),
    binwidth = .5,
    alpha = .5) +
  geom_density(
    data = filter(au_2020_tbl, dist == "Analytic"),
    alpha = .2) +
  theme_bw() +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Predictive Distributions of Crude Net Migration Rate (Analytic and Sample)",
    subtitle = "Australia, 2020-2025",
```

```
    x = "Crude Net Migration Rate",
    y = "Density",
    fill = "Type"
  )

# Question 2e -------------------------------------------------------------

library(wpp2012)
data("migrationF", package = "wpp2012")
data("migrationM", package =  "wpp2012")

au_mig12 <- bind_cols(
  migrationF %>%
  filter(country == "Australia") %>%
  select(age, migF = `2015-2020`),
  migrationM %>%
  filter(country == "Australia") %>%
  select(migM = `2015-2020`)) %>%
  mutate(migTotal = migF + migM,
         pasmr = migTotal/sum(migTotal))


au_mig20 = tibble(age = au_mig12$age, pasmr = au_mig12$pasmr) %>%
  mutate(asmr = pasmr * au_mean_pred)

au_mig20 %>%
  knitr::kable(
  booktabs = TRUE, digits = 2,
  col.names = c("Age Grop", "PASMR", "ASMR"),
  caption = "Predicted Age-spesific Migration Rates (ASMR) for Australia, 2020-2025",
)

au_mig20 %>%
  mutate(age = seq(0, 100, 5)) %>%
  ggplot(aes(age, asmr)) +
  geom_line(color="coral", size=3) +
  theme_bw(base_size = 15) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Predicted Age-spesific Migration Rates (ASMR)",
    subtitle = "Australia, 2020-2025",
    x = "Age-spesific Migration Rate",
    y = "Age" ) +
  scale_x_continuous(n.breaks = 10)
rc_decomposition = function(x, params){
  a1 = params[["a1"]]
  alpha1 = params[["alpha1"]]
  a2 = params[["a2"]]
  alpha2 = params[["alpha2"]]
  mu = params[["mu2"]]
  lambda = params[["lambda2"]]
  c = params[["c"]]
```

```
  mx = a1 * exp(-alpha1 * x) + #first component
    a2 * exp(-alpha2 * (x - mu) - exp(-lambda * (x - mu))) + #second component
    c #constant component

  return(mx)
}

rc_params = deframe(migest::rc_model_fund)
au_mig20_alt = tibble(age = 0:104) %>%
  mutate(pasmr = rc_decomposition(age, rc_params),
         pasmr = pasmr/sum(pasmr),
         asmr = pasmr * au_mean_pred)

  au_alt_grouped = au_mig20_alt %>%
  mutate(age_group = rep(1:21, each=5))%>%
  group_by(age_group) %>%
  summarise(asmr = mean(asmr)) %>%
  mutate(age_group = factor(age_group, labels = au_mig12$age))

au_alt_grouped %>%
  knitr::kable(
  booktabs = TRUE, digits = 2,
  col.names = c("Age Grop", "ASMR"),
  caption = "Rogers-Castro Model Age-spesific Migration Rates (ASMR) for Australia, 2020-2025",
)
au_mig20_alt %>%
  ggplot(aes(x=age, y=asmr )) +
  geom_line(color = "coral", size = 3) +
  theme_bw(base_size = 15) +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Rogers-Castro Model Age-spesific Migration Rates (ASMR)",
    subtitle = "Australia, 2020-2025",
    x = "Age-spesific Migration Rate",
    y = "Age" ) +
  scale_x_continuous(n.breaks = 10)

# Question 2f ------------------------------------------------------------
data("pop", package="wpp2019")
au_pop20 = pop %>% filter(name == "Australia") %>% pull(`2020`)
mig_est = rnorm(50,  mean = au_mean_pred, au_sd)
# storing age categories as a sepearate variable
age_cat = au_mig12$age
pasmr = au_mig12 %>% select(age, pasmr)

au_mig25 =
  tibble(sample = 1:50,
       pop20 = au_pop20,
       nmr25 = mig_est) %>%
  mutate(mig25 = au_pop20 * mig_est) %>%
  slice(rep(1:n(), each = 21)) %>%
  mutate(age = rep_len(age_cat, length.out = n()), .after=sample) %>%
  left_join(pasmr, by = "age") %>%
```

```r
  mutate(mig_counts25 = mig25*pasmr)


ggplot(au_mig25, aes(x = mig_counts25)) +
  geom_histogram() +
  facet_wrap(vars(age), ncol = 3) +
  theme_bw() +
  theme(text = element_text(family = "serif")) +
  labs(
    title = "Projected Age-Specific Net Migration Numbers",
    subtitle = "Australia, 2020-2025",
    x = "Net number of migrants",
    y = "Density"
  )
# Question 2g --------------------------------------------------------------
out.dir = "../data/au_2025"
mig.traj = "../data/mig/mig_trajectories.csv"

mig_out  =
  au_mig25 %>%
  mutate(LocID = "36", Period = "2020-2025", Year = "2023") %>%
  #mutate(mig25 = mig25*1000) %>%
  select(LocID, Year, Trajectory=sample, Migration=mig25) %>%
  distinct(Trajectory, .keep_all = TRUE)


#write_csv(mig_out,mig.traj)




# # Run this once to get pop predictions
# pop.pred <- pop.predict(
#                 end.year = 2100, start.year = 1950, present.year = 2020,
#                 wpp.year = 2019, output.dir = out.dir, nr.traj = 50,
#                 countries = "Australia",
#                 inputs = list(tfr.sim.dir = tfr.dir,
#                               e0F.sim.dir = e0.dir,
#                               e0M.sim.dir = "joint_",
#                               migtraj = mig.traj),
#                 keep.vital.events = FALSE, replace.output=FALSE)

au_pred_25 <- get.pop.prediction(out.dir)



outM = bayesPop::pop.byage.table(au_pred_25, country = "Australia",
                                 sex=c("male"), year = 2025)
outF =  bayesPop::pop.byage.table(au_pred_25, country = "Australia",
                                 sex=c("female"), year = 2025)



tblM = outM %>%
```

```r
  as.data.frame() %>%
  tibble::rownames_to_column(var = "age") %>%
  select(-`0.1`, -`0.9`)


tblF = outF %>%
  as.data.frame() %>%
  tibble::rownames_to_column(var = "age") %>%
  select(-`0.1`, -`0.9`)

tblM %>%
  knitr::kable(
  booktabs = TRUE, digits = 2,
  col.names = c("Age", "Median", "2.5% PI", "97.5 PI"),
  caption = "Predicted MalePopulation for Australia, 2020-2025")
tblF %>%
  knitr::kable(
  booktabs = TRUE, digits = 2,
  col.names = c("Age", "Median", "2.5% PI", "97.5 PI"),
  caption = "Predicted Female Population for Australia, 2020-2025")
```