popReconstruct Demonstration

IUSSP International Population Conference, 2017

Mark Wheldon

United Nations, Population Division
and
Auckland University of Technology, New Zealand

Disclaimer: The views and opinions expressed in this presentation are those of the authors and do not necessarily represent those of the United Nations. This presentation has not been formally edited and cleared by the United Nations.

Getting the Files

- Create a directory on your hard drive to store the files.
- Go to
  https://github.com/markalava/Bayesian-Reconstruction
- Click "Clone or download".
- You can either download the zip file or 'clone' the *GitHub* repository.
- The repository contains the package as well as data files and *R* scripts to reconstruct the population of Thailand.

Getting the Files

Inputs

Initial Estimates

- The main inputs are initial estimates of fertility, migration, survival, population counts in the baseline year, population counts in subsequent years, and sex ratios at birth:

```
> (load("../data/thai_initial_ests.RData"))
[1] "asFertTHAI.mat"  "asMigTHAI.mat"  "asSurvTHAI.mat"  "baselineTHAI.mat"
[5] "censusTHAI.mat"  "srbTHAI.mat"
```

Separation Factors

- Life table separation factors aid conversion among demographic parameters.

```
> (load("../data/thai_sep_factors.RData"))
[1] "thaiFemale.sf"  "thaiMale.sf"
```

Inputs

Proposal Variances

- Proposal variances are additional inputs that ensure efficient running of the MCMC algorithm.
- Tuning functions in `2_tuning.R` can be used to help select good values for these parameters.

```
> (load("../data/thai_propvars.RData"))
[1] "thai.propvar"
```

Initial Estimates

Fertility Rates and SRBs

- Fertility rates and SRBs are matrices with five-year periods as columns, age-groups as rows.

```
> asFertTHAI.mat[1:12, 1:5]
          1960        1965        1970        1975        1980
0   0.00000000  0.00000000  0.00000000  0.00000000  0.000000000
5   0.00000000  0.00000000  0.00000000  0.00000000  0.000000000
10  0.00000000  0.00000000  0.00000000  0.00000000  0.000000000
15  0.05175459  0.06061538  0.06224863  0.06697348  0.056633004
20  0.24021806  0.23413036  0.22148542  0.21094695  0.163755256
25  0.28880806  0.27729894  0.24963584  0.19398416  0.154346419
30  0.27214926  0.25390511  0.21493142  0.15023581  0.103045913
35  0.21223853  0.19860950  0.16017444  0.10735185  0.060685573
40  0.11637142  0.12300785  0.09998354  0.05626398  0.041692712
45  0.02948599  0.03232336  0.02472358  0.02067992  0.008513132
50  0.00000000  0.00000000  0.00000000  0.00000000  0.000000000
55  0.00000000  0.00000000  0.00000000  0.00000000  0.000000000
```

Initial Estimates

Survival and Migration Proportions, Population Counts

- The remaining inputs are lists of two components, "female" and "male".

- Each component is a matrix with five-year periods as columns, age-groups as rows.

```
> lapply(asSurvTHAI.mat, "[", i = 1:3, j = 1:5)
$female
        1960      1965      1970      1975      1980
0  0.9142663 0.9237345 0.9377636 0.9546893 0.9692136
5  0.9612171 0.9621474 0.9724710 0.9805215 0.9857410
10 0.9899652 0.9911082 0.9923885 0.9936778 0.9954615

$male
        1960      1965      1970      1975      1980
0  0.9108313 0.9224748 0.9340966 0.9461544 0.9605498
5  0.9611747 0.9639159 0.9731235 0.9801511 0.9830052
10 0.9881956 0.9895983 0.9911646 0.9926470 0.9943731
```

Running the Reconstruction: `1_estimation.R`

- Set-up

```
> set.seed(1)
> library(coda)
> data.path <- "data"
> prog.path <- "R_Code"
> results.path <- "results"
> if (!file.exists(results.path)) dir.create(results.path)
> example(source, echo = FALSE)
> sourceDir(prog.path)

Calculate_Counts.R :
ccmp_femDom.R :
ChooseHyperparameters.R :
CondVariances.R :
EstModel_TwoSex.R :
Misc_Functions.R :
sampleFromPrior.R :
```

Running the Reconstruction: `1_estimation.R`

- Set the number of iterations.
  ```
  > n.iter <- 17
  > burn.in <- 3
  ```

- Convert elicited relative errors to hyperparameters.
  ```
  > invGam.params <- make.hyper.params(absDev = list(fert = 0.1,
  +     surv = 0.1, mig = 0.2, pop = 0.1, srb = 0.1), prob = list(fert = 0.9,
  +     surv = 0.9, mig = 0.9, pop = 0.9, srb = 0.9), alpha = list(fert = 0.5,
  +     surv = 0.5, mig = 0.5, pop = 0.5, srb = 0.5), s.star = unlist(asSurvTHAI.ma
  > invGam.params[1:2]
  $al.f
  [1] 0.5

  $be.f
  [1] 0.0001392358
  ```

Running the Reconstruction: `1_estimation.R`

- Load proposal variances.

```
> withVisible(load(file = file.path("data", "thai_propvars.RData")))$value
[1] "thai.propvar"
```

- Set arguments

```
> estModArgs <- list(al.f = invGam.params$al.f, be.f = invGam.params$be.f,
+     al.s = invGam.params$al.s, be.s = invGam.params$be.s,
+     be.g = invGam.params$be.g, al.n = invGam.params$al.n,
+     al.srb = invGam.params$al.srb, be.srb = invGam.params$be.srb,
+     n.iter = n.iter, burn.in = burn.in,
+     start.f = asFertTHAI.mat, start.s = asSurvTHAI.mat,
+     start.b = baselineTHAI.mat, start.srb = srbTHAI.mat,
[..etc..]
```

- Run, then save the reconstruction.

```
> ThaiMcmc <- do.call(popRecon.pop.est.sampler.aug07, args = estModArgs)
> save(ThaiMcmc, file = file.path(results.path, "thai_mcmc.RData"))
```

Tuning `2_tuning.R`

- MCMC acceptance ratios can be examined to help find good proposal variances.
- Raftery-Lewis diagnostic can be run to help choose chain length.

Proposal Variances

- MCMC acceptance ratios between about 0.2 and 0.5 are acceptable.
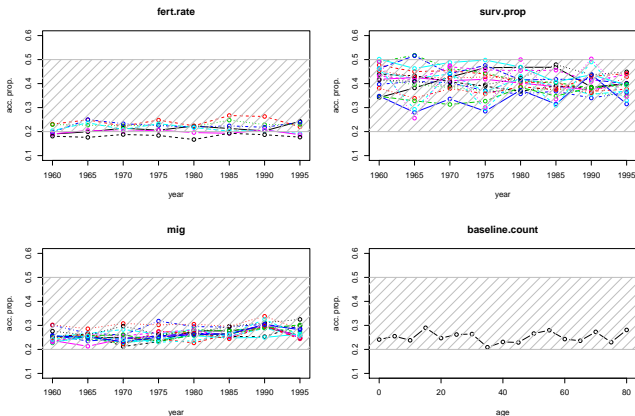- These can be plotted for each input parameter.

Figure: Metropolis acceptance ratios for fertility rate parameters. Example run with 1000 iterations and 1000 burn-in.

Proposal Variances

- Automatic suggestions for proposal variances are generated if they need to be tuned.

- Finding good proposal variances is a matter of trial and error with short runs of the model.

```
> vitalCondVars <- list()
> vitalCondVars$fert.rate <- chain.cond.vars.feb08(log(ThaiMcmc$fert.rate.mcmc))
> vitalCondVars$surv.prop <-
+     (chain.cond.vars.feb08(logit(ThaiMcmc$surv.prop.mcmc[["female"]])) +
+     chain.cond.vars.feb08(logit(ThaiMcmc$surv.prop.mcmc[["male"]])))/2
```

Sample from the Prior Distribution
2_sample_from_prior.R

- To be able to create plots that compare the prior and posterior distributions, run 2_sample_from_prior.R.
- Note: this can take a long time.

Post Process `3_post_process.R`

- Output from `1_estimation.R` needs to be post-processed (e.g., to convert rates to counts).

```
### Calculate Counts from Posterior

> nF <- ncol(ThaiMcmc$fixed.params$mean.mig.prop$female) *
+     nrow(ThaiMcmc$fixed.params$mean.mig.prop$female)

> proj.to.counts.aug22(fert.rate.mcmc = ThaiMcmc$fert.rate.mcmc
+                      ,surv.prop.mcmc = list(female = ThaiMcmc$surv.prop.mcmc$female
+                       ,male = ThaiMcmc$surv.prop.mcmc$male)
+                      ,mig.prop.mcmc = list(female = ThaiMcmc$mig.prop.mcmc$female
+                       ,male = ThaiMcmc$mig.prop.mcmc$male)
+                      ,baseline.count.mcmc = list(female =
+                       ,male = ThaiMcmc$baseline.count.mcmc$male)
+                      ,pop.count.mcmc = list(female = ThaiMcmc$lx.mcmc$female
+                       ,male = ThaiMcmc$lx.mcmc$male)
+                      ,sep.factors = list(female = thaiFemale.sf[1:nF]
+                       ,male = thaiMale.sf[1:nF])
+                      ,srb.mcmc = ThaiMcmc$srb.mcmc
+                      ,name.pref = "Thai."
+                      ,name.suf = ""
+                      ,file.backed = TRUE
+                      ,ccmp.f = "ccmp.femDom.jun12"
+                      )
[..etc..]
```

Plot Results `4_plots_results.R`

- To plot some results, use the functions and calls in
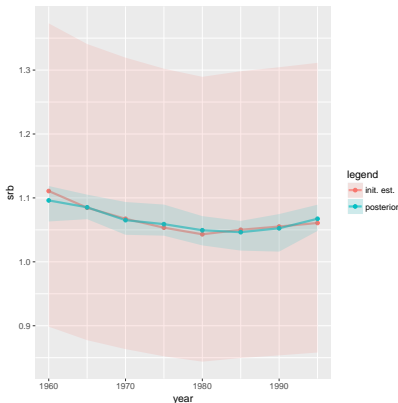 `4_plots_results.R`.

Figure: Sex ratio at birth. Prior (red) and posterior (blue) 95% credible intervals and prior and posterior medians.