Computation Time Verification & Fitting Chains for Data Analysis

Overview

This notebook was used to

1. get an approximate runtime required to fit the models on North Carolina data
2. fit models and also record the runtime for the S.Atlantic Census Division. These chains were saved and used for the data analysis (Section 6 of the manuscript)

This was run on a M1 Macbook Pro (2021).

Run on: 2024-12-25.

The samplers are loaded here

```
source("samplers/fh_fit.R")
source("samplers/dm_fit.R")
source("samplers/car_fit.R")
source("samplers/bym_fit.R")
source("samplers/ssd_fit.R")
```

North Carolina

Load Data

```
#north carolina
load("data/data_NC/data.RDA")

#data
covs = c("degree", "assistance", "no_car", "povPerc",
         "white", "black", "native", "asian", "hispanic")
X <- model.matrix(~., all_data[, covs, drop=F])
n <- nrow(X); j <- ncol(X) # number of covariates INCLUDES intercept

#response (with transformation)
y <- all_data$rentBurden
d.var <- all_data$rentBurdenSE^2
y.star <- log(y)
d.star <- d.var / y^2
```

Run samplers

```r
set.seed(7)
# ---- Run Chains ----
start <- proc.time()
fh_res <- fh_fit(X, y.star, d.star, ndesired=2500, nburn=10000, nthin=1, verbose=FALSE)
fh_time <- proc.time() - start

start <- proc.time()
dm_res <- dm_fit(X, y.star, d.star, ndesired=2500, nburn=10000, nthin=1, verbose=FALSE)
dm_time <- proc.time() - start

start <- proc.time()
car_res <- car_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
car_time <- proc.time() - start

start <- proc.time()
bym_res <- bym_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
```

```
## The legacy packages maptools, rgdal, and rgeos, underpinning the sp package,
## which was just loaded, will retire in October 2023.
## Please refer to R-spatial evolution reports for details, especially
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
## The sp package is now running under evolution status 2
##      (status 2 uses the sf package in place of rgdal)
```

```r
bym_time <- proc.time() - start

start <- proc.time()
new_res <- ssd_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
ssd_time <- proc.time() - start
```

Computation time **in seconds**:

```
## [1] "--- FH Computation time ---"

##    user  system elapsed
##   3.839   0.082   3.938

## [1] "--- DM Computation time ---"

##    user  system elapsed
##   6.202   0.051   6.266

## [1] "--- CAR Computation time ---"

##    user  system elapsed
##  41.395   0.214  41.676

## [1] "--- BYM Computation time ---"
```

```
##    user  system elapsed
##  29.851   0.686  31.959
```

```
## [1] "--- SSD Computation time ---"
```

```
##    user  system elapsed
##  51.668   0.772  53.833
```

South Atlantic Census Division

Load Data

```
library(tidycensus)
#south atlantic census division
load("data/data_multi-state/data.RDA")
all_data = all_data %>% mutate(state_code=strtrim(all_data$fips, 2)) %>%
  left_join(fips_codes %>%
              select(state, state_code, state_name) %>%
              unique.data.frame(), by="state_code")

#data
covs = c("degree", "povPerc", "black", "state")
X <- model.matrix(~., all_data[, covs, drop=F])
n <- nrow(X); j <- ncol(X) # number of covariates INCLUDES intercept

#response (with transformation)
y <- all_data$rentBurden
d.var <- all_data$rentBurdenSE^2
y.star <- log(y)
d.star <- d.var / y^2
```

Run samplers

```
set.seed(7)
# ---- Run Chains ----
start <- proc.time()
fh_res <- fh_fit(X, y.star, d.star, ndesired=2500, nburn=10000, nthin=1, verbose=FALSE)
fh_time <- proc.time() - start

start <- proc.time()
dm_res <- dm_fit(X, y.star, d.star, ndesired=2500, nburn=10000, nthin=1, verbose=FALSE)
dm_time <- proc.time() - start

start <- proc.time()
car_res <- car_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
car_time <- proc.time() - start

start <- proc.time()
```

```r
bym_res <- bym_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
bym_time <- proc.time() - start

start <- proc.time()
new_res <- ssd_fit(X, y.star, d.star, A, ndesired=2500, nburn=2000, nthin=1, verbose=FALSE)
ssd_time <- proc.time() - start
```

Computation time **in seconds**:

```
## [1] "--- FH Computation time ---"

##    user  system elapsed
##  15.648   0.108  15.802

## [1] "--- DM Computation time ---"

##    user  system elapsed
##  28.775   0.377  29.438

## [1] "--- CAR Computation time ---"

##     user   system  elapsed
## 2882.464   24.887 2926.238

## [1] "--- BYM Computation time ---"

##     user   system  elapsed
## 2220.692   22.473 2260.645

## [1] "--- SSD Computation time ---"

##     user   system  elapsed
## 3925.040   37.130 3988.454
```

Computation time in **minutes (spatial models)**:

```
## [1] "--- CAR Computation time (minutes)---"

##       user     system    elapsed
## 48.0410667  0.4147833 48.7706333

## [1] "--- BYM Computation time (minutes)---"

##     user   system  elapsed
## 37.01153  0.37455 37.67742

## [1] "--- SSD Computation time (minutes)---"

##       user     system    elapsed
## 65.4173333  0.6188333 66.4742333
```

Save the chains for S.Atlantic Division

To be used for the data analysis.

```
save(fh_res, file="data_analysis_chains/fh_res.RDA")
save(dm_res, file="data_analysis_chains/dm_res.RDA")
save(car_res, file="data_analysis_chains/car_res.RDA")
save(bym_res, file="data_analysis_chains/bym_res.RDA")
save(new_res, file="data_analysis_chains/new_res.RDA")
```