

README for GitHub Repository

“BayesIV_Simulations”

Samrachana Adhikari, Sherri Rose, Sharon-Lise Normand

1 What is in this repository?

Scripts to run and summarize simulations for the paper “**Nonparametric Bayesian Instrumental Variable Analysis: Evaluating Heterogeneous Effects of Coronary Arterial Access Site Strategies**”, under strong instrument assumptions with Gamma distribution for the errors on outcome. The .R scripts included in this repository can be broadly divided into following three groups:

1. Scripts to run the simulations with 3 different sample sizes
 - “Simulation_strongIV_n_100_example.R”: for $n = 100$
 - “Simulation_strongIV_n_500_example.R”: for $n = 500$
 - “Simulation_strongIV_n_2000_example.R”: for $n = 2000$
2. Scripts to summarize the outputs for each n
 - “SummarizeSimulations_n_100.R”
 - “SummarizeSimulations_n_500.R”
 - “SummarizeSimulation_n_2000.R”
3. Script to run the simulation and summarize for all three n ’s in a sequence: “runSimulations.R”

1.1 Overall structure of the simulation files:

Structure of the .R files for different sample size is very similar and differs only in the number of observations n . We created different versions to automate the data generation, model fitting and summarization of the fitted models. As an example we discuss the structure of files for simulation with $n = 100$. The file ‘Simulation_strongIV_n_100_example.R’ is organized to repeat the following for 50 times:

- setup parameters to generate covariates, instrument and outcome data
- fit three different models that we compare in the paper

- save the generated data and the fitted models as .RData file.

Next, the file ‘SummarizeSimulations_n_100.R’ is organized to summarize the model fit as:

- load the simulated data and the fitted models
- compute posterior estimates of the treatment effects
- finally, summarize the estimates to obtain bias, coverage and width of the credible interval over 50 replications.

Run time: For $n = 100$, scales linearly with n .

2 Example Dataset:

Example dataset is included in the R package ‘BayesIV’. The following command in R console after loading the package displays the content of the example data:

```
Data(package='BayesIV')
```

The example data has following objects:

Y_{obs} = a vector of the observed outcome

D = a vector of the treatment assignment

X = a matrix of confounders

Z = a vector of the instrumental variable

The proposed model can be fitted in this dataset with following steps:

1. install the package in R from GitHub repository.

```
install.packages('devtools')
library(devtools)
install_github('SamAdhikari/BayesIV')
library(BayesIV)
```

2. fit the NP-Bayes IV model with default initialization and priors to draw 100 posterior samples. Refer to BayesIV-manual for more detail.

```
Model = mcmcRun_Normal_DPM(Yobs = Yobs, Tr = D, X = X, Z = Z, niter = 100,
priors = NULL, initialVals = NULL)
```

3. summarize the fitted model by specifying burnin and thinning values for the MCMC chain.

```
ATE_chain = getATE_posterior_NDPM(fittedModel = Model, X = X, niter = 100,
burnin = 0, thin = 1)
ATE = mean(ATE_chain)
```