

Datasets and Code

C code for implementing the MCMC sampling algorithms, R code for running all the sampling algorithms, post-processing the samples and creating plots and entries for tables described in the article are available here.

The user would first need to install the R package **BAS**, then use the R script **sample.r** to run all the sampling algorithms and then use **plots.r** or **tables.r** to create plots or entries for tables described in the article. Please note that the C code was compiled and run, using the “system” command from within R under a Linux OS. Following is a list of all files.

1. **BAS_0.90.tar.gz** : R package **BAS** for implementing the Bayesian adaptive sampling (BAS) algorithm
2. **crimecen.txt**: Centered U.S. crime data used as input to **crime-rs-thin.c**
3. **proteincen.txt**: Centered protein data used as input to **protein-rs-thin.c**
4. **simcen-x.txt**, **simcen-y.txt**: Centered simulated design matrix and vector of response variables respectively, to be used as input to **sim-mc3.c**, **sim-rs.c** and **sim-rs-thin.c**
5. **crime-rs-thin.c**: C code for implementing RS-Thin for U.S. crime data
6. **protein-rs-thin.c**: C code for implementing RS-Thin for protein activity data
7. **protein-rsthin-pred.c**: C code for out of sample prediction using RS-Thin for protein activity data
8. **sim-mc3.c**: C code for implementing MC³ for simulated data
9. **sim-rs.c**: C code for implementing RS for simulated data
10. **sim-rs-thin.c**: C code for implementing RS-Thin for simulated data
11. **sample.r**: R code for running all sampling algorithms for inference and prediction
12. **plots.r**: R code for post-processing output from algorithms and creating plots in the article

13. `tables.r`: R code for post-processing output from algorithms and creating entries for tables in the article
14. `arrays.hpp`: Header file required for C code for memory allocation
15. `allmodels.r`, `bindec.r`, `repeatsamp.r`: Auxiliary R functions