ReadMe

The following scripts are a demonstration of utilizing the methods proposed in a "Bayesian Framework for Generation of Fully Synthetic Mixed Datasets" to create a fully synthetic version of a simulated dataset. The simulated dataset is comprised of an unordered categorical, binary, continuous and count variable. The count variable in this case is simulated as a complex function of the other variables, and as a result, is left out of the initial copula model for targeted univariate synthesis.

The scripts are meant to be run sequentially, with the ordering specified by the numeric prefix in the naming convention. Each script will create variables stored in the global environment that may be accessed at subsequent points in the framework. In addition, the script 'sampleMGP.R' should be run or sourced prior to running the sampler, as it contains the functions necessary for sampling the parameters of the multiplicative gamma process prior utilized in our sampling algorithm.

Version 2.2 or later of the TruncatedNormal package is required to run '2_Synthesizer_demo.R'. At the bottom of the '3_NonlinearSynthesizer_demo.R', the user can find simple univariate and bivariate summaries that confirm the utility of the synthetic data.