SDS 394 Project Proposal

Carlos Gillett (ctg576) & Kayla Schaefer (khs426)

Spring 2015

1 Project Topic

We are going to do our project on Approximate Bayesian Computation, specifically looking at what has been deemed "Tiny Data". To have a concrete example for our project, we will look at the everyday example of estimating the number of socks a person has given a random sample of size < 15.

The ABC algorithm works as follows [1]:

1. Construct a generative model that produces the same type of data as you are trying to model. Assume prior probability distributions over all the parameters that you want to estimate.
2. Sample tentative parameters values from the prior distributions, plug these into the generative model and simulate a dataset.
3. Check if the simulated dataset matches the actual data you are trying to model. If yes, add the tentative parameter values to a list of retained probable parameter values, if no, throw them away.
4. Repeat step 2 and 3 to build up the list of probable parameter values.
5. Finally, the distribution of the probable parameter values represents the posterior information regarding the parameters.

Therefore, we intend to let the user can input how many individual/paired socks you find, what prior parameters you want to use, and how many processors over which to parallel. Then we will run the algorithm and return how many socks you have as singles, as pairs, and overall.

2 Meeting Requirements

Because we are running an acceptance-rejection algorithm, this should extend easily to parallelization with MPI. We also will use the scientific library GSL to aid with the probability distribution functions. We also intend to supply the user with a final output plot of the posterior distributions for the overall number of socks, number of pairs, and number of single socks. Lastly, before reaching our final version, we will try to optimize the code using a profiler.

References

[1] Baath, R. (2014). Tiny Data, Approximate Bayesian Computation and the Socks of Karl Broman. Retrieved from http://www.sumsar.net/blog/2014/10/tiny-data-and-the-socks-of-karl-broman/.