Approximate Bayesian Computation

Carlos Gillett & Kayla Schaefer

12/02/2014

Project Goals

Moving from Big Data to Tiny Data

- With the example of modeling socks in the laundry
  - How many socks? How many pairs / unique socks?
  - Very small sample size
  - Can't use usual bootstrapping or machine learning techniques

Solution

It's as easy as ABC

- Approximate Bayesian Computation

What is ABC?

- An acceptance-rejection algorithm

1. Need a generative model and prior distributions
2. Sample from the priors and plug into the model to get data
3. Does this data match our sample?
4. Repeat 2 and 3
5. This gives the posterior distribution of data

How did we implement this?

Our program lets the user give

- The sample of socks
- Best guesses about the distribution of socks
- Prior parameters

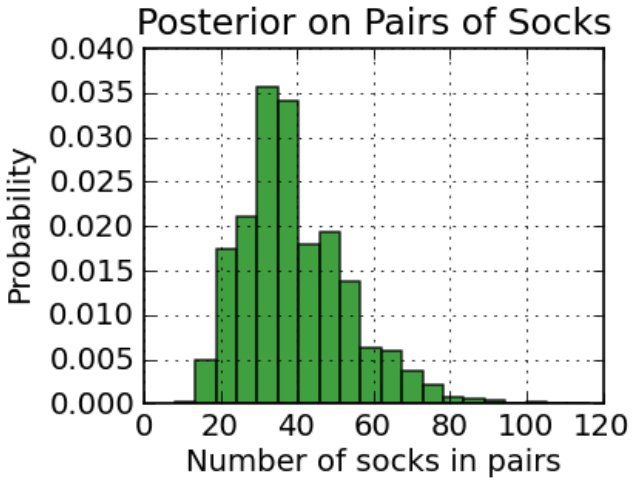
How did we implement this?

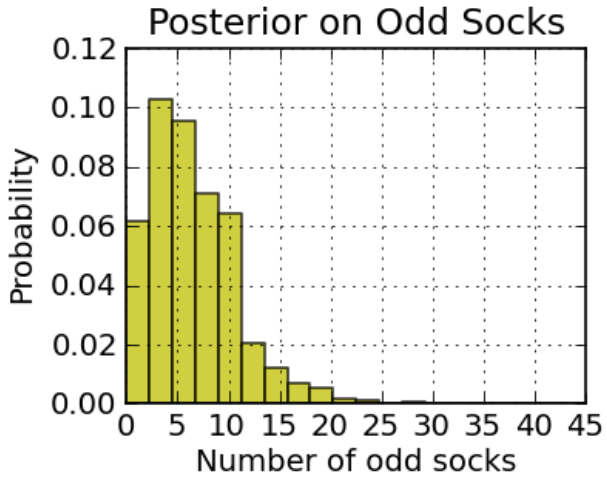
We build up the prior distributions with GSL, then sample from them. Along the way, we check if the samples are consistent with what the user gave.

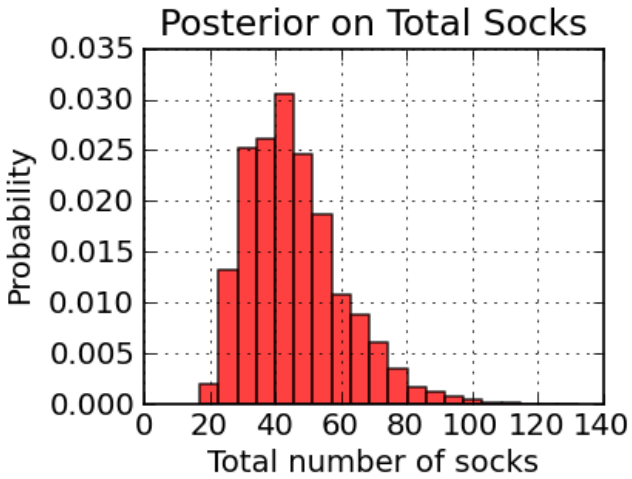
Four files are created that filter out just the 'good' samples. Median estimates of the sock counts are returned.

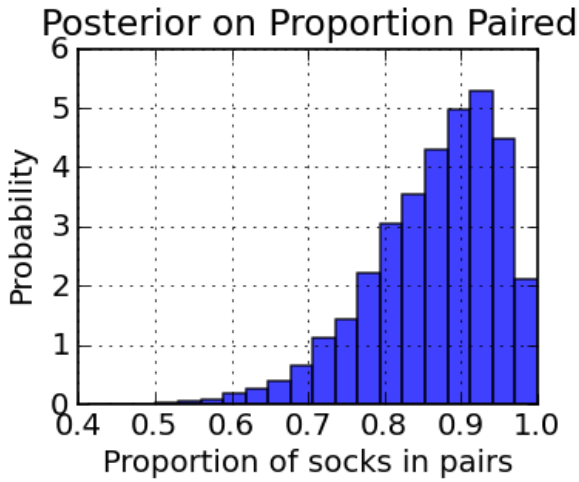
Plots

Generated using python's matplotlib.pyplot









Still to go

- Final tweaks to parallelizing the sampling process
- Code profiling and adjustment

Demo Time

References

Baath, R. (2014). *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/