

Eric Kernfeld

Project proposal for stat 518

Paper is Wilkinson's "Parameter inference for stochastic kinetic models of bacterial gene regulation"

- The paper summary is a separate document.
- Parts of the paper I will implement:
 - the Gillespie algorithm for exact sampling from the underlying stochastic model
 - the particle MCMC scheme for estimating parameters
- Programming language—Vladimir Minin recommends Rcpp. I may write an early “toy” version in a higher-level language.
- Using outside code: the details of the model are specified via systems biology markup language. I don't know if it will be best to use tools that process it directly or if I should translate it into R or C. With your permission, I'd rather not replicate all of the details of the systems biology.
- For extensions:
 - On page 5, Wilkinson says “Conditional on discrete-time observations, the Markov process breaks up into a collection of independent bridge processes that appear not to be analytically tractable.” It sounds like he doesn't know this for certain, and it might be worth a look.
 - As Wilkinson states, the MCMC at the core of this paper is a general tool. I'd like to try it out for epidemiology, maybe replacing the Gillespie algorithm on the inside using chain-binomial infection models, or for ecology, which also uses stochastic process models. I wonder if Jon Wakefield or Vladimir Minin would find this interesting, or whether I ought to look in some other direction.
 - From a practical perspective, what happens when the forward model is close, but wrong, perhaps missing molecules that interact with the ones present? Also, can this method help infer regulatory network structure, or only reaction rates within a fixed structure?
 - I also have a pet project called Basis Expansion Monte Carlo that I'm itching to try out here or somewhere else. It's an approximation to Metropolis-Hastings that allows parallelization and might tolerate low acceptance rates. I won't go into the details here, but I have a small write-up that I work on when I get the chance. It's included as the file “BEMC.pdf”.
- Problems to fix—I haven't seen any yet, but I would like to flesh out more of the mathematical details, especially about using the model sequentially for different datasets.
- The datasets used in the paper are all synthetic.
- Experimental evaluation: As a scaffold, I want to do one set of checks to ensure mathematical correctness, then another to reproduce Wilkinson's first experiment. Beyond that, I need to discuss this with faculty. On the plus side, since the datasets in the paper are all synthetic, I don't need to hunt them down.
- a list of related papers (background, followup research, competing methods, applications):
 - **EMK: Will work on this**

Schedule

- Week 2:

By Thursday: draft the project plan **EMK: Check!**, prepare a 5-minute presentation for class **EMK: Check!**, and turn in my summary of the paper **EMK: Check!**.

- Week 3:
Catch up from week 2 [EMK: Check!](#); fill in more details on related work [EMK: To do!](#) and discuss with Jon/Vladimir/Marina what exactly I will implement [EMK: discussed first steps with Marina on Friday](#). In conjunction with stat 534, learn C [EMK: Working on it: I've now successfully called C from R](#) I'm trying to figure out how to use math libraries in Rcpp.
- Week 4: Hammer out pseudo-code for the implementation [EMK: Much of the planning is finished](#). Begin work on a larger presentation and write-up [EMK: Began working on two new graphics and a presentation on sequential MC.](#).
- Week 5: Work on full-scale code. Document experiments to be done and maybe test small versions of them. Continue updating presentation and write-up.
- Week 6: Work on code. Perform checks for mathematical correctness. Document these. Continue updating presentation and write-up.
- Week 7: Work on code. Perform other experiments, creating scripts to build figures. Continue updating presentation and write-up.
- Week 8: Half the same as week 7, with the other half spent searching the literature and writing my reactions to what I find.
- Week 9: Extra time for whatever gets behind schedule. Proofread.