

Eric Kernfeld  
Summary of Wilkinson's "Parameter inference for stochastic kinetic models of bacterial gene regulation",  
a book chapter in [1].

report; only 1 author?

## Abstract

uninformative priors?; why, what does this achieve?

nice abstract —  
this is the spirit;  
see also a few  
coments—>>

In this paper, Wilkinson attempts to infer reaction rates for biochemical networks in a setting with discrete observations, missing data, and measurement error. He uses vague priors and likelihood-free MCMC methods within a Bayesian model. He runs four main simulations. The first three iterate through successively more difficult and realistic measurement models, and they show the approach can accurately infer three key reaction rates with a useful precision. The fourth studies a naive model, showing it leads to overconfident, incorrect inferences. All tests are conducted using synthetic data on only one vector of true parameters.

text

experiments - not th

..not the  
experiments, but  
what is  
demonstrated by  
them (if anything)

I plan to implement the method in Julia and reproduce the experiments. To honestly test the method, I need to do more simulations, too: what happens when the log-space mean of the prior is not near the true values, or when the true values are somewhere other than Wilkinson's choice? What happens when all the parameters are unknown, rather than just the three we are interested in?

# 1 Introduction

## 1.1 Summary

very nice  
summary! you  
could make it  
like abstract.

This paper develops tools to study bacterial behavior. There is a line of research modeling biochemical systems as continuous-time Markov jump processes, and this approach has physical motivation. The inherent stochasticity can also explain why three bacteria in similar environments may act differently. Much is known about biochemical network structure, but less about parameter values [2], so within this area, Wilkinson's research fill a gap in parameter inference for Markov jump processes. The problem is difficult because in practice the data are partial, discrete-time, and noisy; for mathematical flexibility, his paper uses Bayesian inference and MCMC. The method is computationally intensive and excruciatingly serial, but it is accurate and shows potential for flexibility. Wilkinson's method or its variants may ultimately help reverse-engineer the regulatory networks that determine bacterial behavior and survival, though because of problems with scaling, it will probably be superseded by other research <sup>1</sup>.

## 1.2 Bare-Bones Biology

As a motivating case, Wilkinson uses the "decision" of *Bacillus subtilis* whether to become mobile. The paper centers around a gene encoding *flagellin*, which is a protein component of organelles that allow motility. Because biological networks can be disorienting, I'll outline some of the biological relationships in short sentences in the next paragraph.

move this at the beginning; here it interrupts the flow

The protein *flagellin* helps bacteria move. The protein  $\sigma^D$  promotes *flagellin*. The *fla / che* operon <sup>2</sup> contains many motility-related genes, including the one for  $\sigma^D$ . The protein  $\sigma^A$  and the protein *CodY* both suppress the *fla / che* operon. Thus, they suppress  $\sigma^D$ , and they indirectly suppress *flagellin*. In fact, *CodY* also downregulates *flagellin* directly. This is easiest to digest as a figure.

for this case, better put this in the caption, say the interactions form a bayes net (if they do)

The paper focuses on measurements of  $\sigma^D$ . It attempts to infer the rates of binding of *fla/che* repressors, unbinding of *fla/che* repressors, and production of  $\sigma^D$ .

Sig^D better than \sigma^D

# 2 The Model

Since we're simulating a biochemical system, suppose there are  $X_j(t)$  particles of type  $j$  at time  $t$ ,  $j \in \{1 \dots u\}$ . These particles interact via a set of reactions  $\mathcal{R}_i$ ,  $i \in \{1 \dots v\}$ , with  $\mathcal{R}_i$  consuming  $p_{ij}$  particles of type  $j$  and

<sup>1</sup>Much of the alternative research Wilkinson cites is from his own research group, which seems to be trying out every available strategy on this problem—see the brief section on other work.

<sup>2</sup>Proteins are complex molecules that take myriad forms and roles within a cell. Operons are the basic transistor-like elements of the genome. In response to an outside stimulus, such as high levels of the sugar lactose, a normally active operon may become inactive or vice versa. By "active", I mean that the DNA encoded by the operon can be transcribed into RNA; this is the first step in the production of proteins.

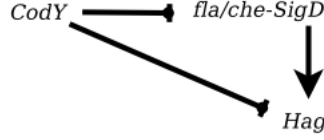


Figure 1: Regulatory relationships. *Hag* is the gene for *flagellin*, while *SigD* encodes  $\sigma^D$ .

producing  $q_{ij}$  particles of type  $j$ . Let  $R_i(t)$  denote the number of reactions of type  $i$  in  $[0, t]$ . I'll refer to these using the arrays  $X(t)$ ,  $P$ ,  $Q$  and  $R(t)$ . Defining the matrix  $S$  to be  $Q^T - P^T$ , the model says that  $X(t) - X(0) = SR(t)$ . Under some assumptions, the different reaction channels evolve independently, and  $R_i(t)$  is a Poisson process with intensity  $c_i \int_0^t \prod_{j=1}^u (X_j(t))_{p_{ij}}$ . The  $c_i$ 's are unknown.

Suppose  $n$  reactions occur between time 0 and time  $T$ . Suppose the  $i$ th one occurs at time  $t_i$  and suppose it has type  $\nu_i$ . Define  $t_0 \equiv 0$  and  $t_{n+1} \equiv T$ . When it comes to inference, the likelihood follows a competing-hazards model from continuous Markov chain theory. If the number of reactions of type  $k$  is  $r_k$ , then the likelihood has a conjugate Gamma prior  $c_k \sim \Gamma(a_k, b_k)$ , and the corresponding posteriors are independent with  $c_k \sim \Gamma(a_k + r_k, b_k + \sum_{i=0}^n (t_{i+1} - t_i) \binom{X_j(t_i)}{p_{\nu_i j}})$ . This is an ideal scenario; missing data make inference more challenging.

### 3 Inference

data collected in this type of experiment

In Wilkinson's data, not every reaction is recorded. Measurements are intermittent, with error, and only  $\sigma^D$  or a fluorescent reporter gets measured. Conjugate priors cannot be used because, conditioned on discrete observations, the likelihood has no closed form—in fact, even evaluating the likelihood is not feasible. To adapt, the paper modifies likelihood-free methods from [3]. To make the mathematics easier to digest, I'll introduce first the predecessor and only then Wilkinson's method.

Let  $\theta$  include  $c$ , controlling the reaction rates, and  $\tau$ , controlling the scale of measurement error. Let  $x$  denote the true state of the chain, but measured only at discrete times. The likelihood  $P(x|\theta)$  cannot be evaluated. Let  $\mathcal{D}$  be  $x$  measured with error and possibly with missing data for some particle types. If we want to construct a Metropolis-Hastings scheme to sample from  $P(x, \theta|\mathcal{D}) \propto P(\theta)P(x|\theta)P(\mathcal{D}|x, \theta)$  using a proposal  $f(\theta^*, x^*|\theta, x)$ , it works out that the acceptance ratio must be the min of 1 and

$x$  = set of all counts  $X$ ?

$$\frac{f(\theta^*, x^*|\theta, x)}{f(\theta, x|\theta^*, x^*)} \times \frac{P(\theta)}{P(\theta^*)} \times \frac{P(x|\theta)}{P(x^*|\theta^*)} \times \frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)}. \quad P(..)$$

This genre of algorithms builds upon the fact that exact simulations from this model are possible, and also that measurement error helps soften the requirements on where the bridge process should begin and end. The insight from from [3] is that one can cancel the intractable term  $\frac{P(x|\theta)}{P(x^*|\theta^*)}$  by constructing a proposal that contains  $\frac{P(x^*|\theta^*)}{P(x|\theta)}$  as a factor. This factor arises if instead of drawing both  $x^*$  and  $\theta^*$  from simple out-of-the-box proposals, we draw only  $\theta^* \sim f(\theta^*|\theta)$  and compute  $x^*$  via simulation with parameters  $\theta^*$ . In the end result, the ratio of interest simplifies:

$$\begin{aligned} & \frac{f(\theta^*, x^*|\theta, x)}{f(\theta, x|\theta^*, x^*)} \times \frac{P(x|\theta)}{P(x^*|\theta^*)} \times \frac{P(\theta)}{P(\theta^*)} \times \frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)} \\ &= \frac{f(\theta^*|\theta)}{f(\theta|\theta^*)} \times \frac{P(x^*|\theta^*)}{P(x|\theta)} \times \frac{P(x|\theta)}{P(x^*|\theta^*)} \times \frac{P(\theta)}{P(\theta^*)} \times \frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)} \\ &= \frac{f(\theta^*|\theta)}{f(\theta|\theta^*)} \times \frac{P(\theta)}{P(\theta^*)} \times \frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)}. \end{aligned}$$

specifiy

To simplify further, use an independence sampler with the prior as a proposal, and the calculation reduces to  $\frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)}$ .

one summary sentence about how nice/useful this idea

If the measurement error is small, or  $\mathcal{D}$  is high-dimensional, or both, this scheme leads to very high rejection rates, so it is not usable. Instead, you can break down  $X$  and  $\mathcal{D}$ , adding only one data point at a time. The exact approach is best understood by referring to Algorithm 1. In broad strokes, the procedure alternates between generating large samples from  $P(\theta, x_{t_{1:i}} | \mathcal{D}_{t_{1:i}})$ , done via likelihood-free MCMC, and folding in new data (incrementing  $i$ ). To be clear, instead of running MCMC just once, in this method runs another five million steps through the sampler *for every time point in the dataset*. It isn't as bad as it sounds: for a sampler that runs only once, the chain will not mix without a sophisticated proposal, and constructing a usable proposal is often linear in the size of the data anyway.

can you say this clearer?

aha!

---

**Algorithm 1:** Wilkinson's sequence of MCMC Samplers

---

Given a hidden continuous-time Markov process  $\{x_t\}_{t=0}^T$  with:

Unknown parameters  $\theta$

Known initial state  $x_0$

Data points  $\mathcal{D}_{t_i}$  at times  $t_i, i \in \{1, \dots, I\}$

A simple, tractable error model  $P(\mathcal{D}_{t_i} | x_{t_i}, \theta)$

notation for simulator to use later

A simulator for paths of  $x$  given  $\theta$  the process

A big array  $B_0$  (Wilkinson uses length 1,000,000) of samples from a prior on  $\theta, x_0$

Empty arrays  $B_i$  of the same length

For each time point (for  $i \in \{1, \dots, I\}$ ), fill  $B_i$  with samples from this Metropolis Hastings scheme:

Initialize  $(\theta, x_{t_i})$

general comments:

for 1=1  
until...

Until  $B_i$  is full:

Draw  $(\theta^*, x_{t_{i-1}}^*)$  from  $B_{i-1}$  or a KDE of its contents (note  $t_0 \equiv 0$ )

Using  $(\theta^*, x_{t_{i-1}}^*)$ , simulate up to  $x_{t_i}^*$ , the state at time  $t_i$

Set  $A = \min(1, \frac{P(\mathcal{D}_{t_i} | x_{t_i}^*, \theta^*)}{P(\mathcal{D}_{t_i} | x_{t_i}, \theta)})$

With probability  $A$ , overwrite  $(\theta, x_{t_i})$  with  $(\theta^*, x_{t_i}^*)$

If the number of times through this loop exceeds 1000 ( for the burn-in) and equals one modulo five (for the thinning), add  $(\theta, x_{t_i})$  to  $B_i$

- yes, very organized, readable  
- 1000, 5 replace with n\_burnin

1000, n\_thin = 5

- seeing numbers in front of st

helps readers/people who have

questions refer to individual steps

In any M-H scheme, the acceptance probability is either 1 or the proposal ratio times the posterior ratio. For step  $i$ , the proposal ratio  $\times$  posterior ratio can be expanded as follows:

$$\frac{P(x_{t_i}^* | x_{t_{i-1}}^*, \theta^*, \mathcal{D}_{t_{i-1}})}{P(x_{t_i} | x_{t_{i-1}}, \theta, \mathcal{D}_{t_{i-1}})} \frac{P(x_{t_{i-1}}^*, \theta^* | \mathcal{D}_{t_{i-1}})}{P(x_{t_{i-1}}, \theta | \mathcal{D}_{t_{i-1}})} \times \frac{P(x_{t_i}^* | x_{t_{i-1}}^*, \theta^*, \mathcal{D}_{t_{i-1}})}{P(x_{t_i} | x_{t_{i-1}}, \theta, \mathcal{D}_{t_{i-1}})} \frac{P(x_{t_{i-1}}, \theta | \mathcal{D}_{t_{i-1}})}{P(x_{t_{i-1}}^*, \theta^* | \mathcal{D}_{t_{i-1}})} \frac{P(\mathcal{D}_{t_i} | x_{t_i}, \theta)}{P(\mathcal{D}_{t_i} | x_{t_i}^*, \theta^*)}.$$

This shows how the compact form in Algorithm 1 arises. The Markov model would allow us to include or omit red terms at will, which is key: otherwise, the proposal, by design not conditioned on the data, would not cancel the intractable likelihood term. Note that using a KDE does not affect the ratio, because usually the kernel  $K(\cdot | \cdot)$  is symmetric, so the extra term  $\frac{K(\theta^* | \theta)}{K(\theta | \theta^*)}$  is just 1.

very nice; a summary sentence -  
with this we achieved..

## 4 What could Wilkinson have done differently? Experiments and other research

### 4.1 Other work

Wilkinson mentions many areas for future work in terms of biological applications for this method. In terms of new methodology, further work is needed to integrate data from microarrays or RNA-seq and to produce scalable methods that can process data on batches of cells.

which end?

To that end, there exist faster alternatives. Some do not perform exact inference; a popular alternative is to approximate the process using a continuous-state diffusion process satisfying a certain SDE [4, 5] or a linear approximation to such a diffusion process [6]. Other non-exact methods include Approximate Bayesian Computation (ABC), which is compared with LF-MCMC in [7]. Some alternative methods are

a bit of organization here: from what you say, there exist inexact but faster competitors; should we use them or not? and when do you they would be useful? about the exact alternatives - what's their relationship with this method? are they later improvements? complete

exact, with a general theme of using approximations to generate good MCMC proposals. For examples, [8] uses a parallelized ABC algorithm, [9] uses a method of moments scheme, and [10] explores several proposal-generating options, giving a particularly nice set of references in the process.

## 4.2 Summary of Experiments

Wilkinson runs four main simulations. All tests are conducted using synthetic data on only one vector of true parameters. The first three iterate through successively more difficult and realistic measurement models:

- First experiment: observe  $\sigma^D$  directly
- Second experiment: observe  $Hag$  rather than  $\sigma^D$
- Third experiment: observe only a fluorescent reporter

why do they run these experiments? to demonstrate what? to explore what?

The fourth shows that a naive model, assuming the fluorescent reporter protein is proportional to the protein of interest, leads to strong and incorrect claims in the posterior probabilities.

### 4.2.1 Some details needed to reproduce experiments

Wilkinson describes a twelve-reaction regulatory model for these chemicals, spelling it out in table 1, and he lists three scientifically important reaction rates that, for tests of the inference method, will be treated as a “ground truth.” The prior distributions cover 4 orders of magnitude, and they are uniform on a log scale. The experiment assumes  $D_t$  is the number of molecules observed with Gaussian error of standard deviation 10 molecules. The initial state of the cell is assumed known EMK: but does he specify it?, and observations occur every 5 minutes (300 seconds) for 2 hours (7200 seconds).

## 4.3 Critique of experiments

Some authors [11] claim that reaction rates range over seven orders of magnitude, but Wilkinson’s prior covers only four orders of magnitude. Furthermore, it is centered around the ground truth. What happens when the log-space mean of the prior is not near the true values? Wilkinson recognizes prior selection as an issue, citing [12], but declines to pursue it, with an implicit claim that reasonable choices of prior will not jeopardize the method.

Wilkinson also treats parameters as known except for the three parameters of interest. How well does this method work when the rest of the reaction rates must also be inferred, or when the measurement error variance must be inferred?

yes, very nice — this is exactly what one expects of a critical reading

## References

- [1] Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M.: Ninth Valencia international meeting on Bayesian statistics, Benidorm, Spain, 03-08.06.2010. Oxford U.P., Oxford (2012)
- [2] Klipp, E., Liebermeister, W., Wierling, C.: Inferring dynamic properties of biochemical reaction networks from structural knowledge. *Genome Informatics* **15**(1) (2004) 125–137

- [3] Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.: Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences* **100**(26) (2003) 15324–15328
- [4] Golightly, A., Wilkinson, D.J.: Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **61**(3) (2005) 781–788
- [5] Golightly, A., Wilkinson, D.J.a.: Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo. *Interface Focus* **1**(6) (10 2011) 807–820
- [6] Fearnhead, P., Giagos, V., Sherlock, C.: Inference for reaction networks using the linear noise approximation. *Biometrics* **70**(2) (2014) 457–466
- [7] Owen, J., Wilkinson, D.J., Gillespie, C.S.: Likelihood free inference for markov processes: a comparison. *arXiv preprint arXiv:1410.0524* (2014)
- [8] Owen, J., Wilkinson, D.J., Gillespie, C.S.: Scalable inference for markov processes with intractable likelihoods. *Statistics and Computing* (2014) 1–12
- [9] Milner, P., Gillespie, C.S., Wilkinson, D.J.: Moment closure based parameter inference of stochastic kinetic models. *Statistics and Computing* **23**(2) (2013) 287–295
- [10] Golightly, A., Wilkinson, D.J.: Bayesian inference for markov jump processes with informative observations. *Statistical Applications in Genetics and Molecular Biology* (2014)
- [11] Schlosshauer, M., Baker, D.: Realistic protein–protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Science : A Publication of the Protein Society* **13**(6) (06 2004) 1660–1669
- [12] Liebermeister, W., Klipp, E.: Biochemical networks with uncertain parameters. *IEE Proceedings-Systems Biology* **152**(3) (2005) 97–107