

Eric Kernfeld

Summary of Wilkinson’s “Parameter inference for stochastic kinetic models of bacterial gene regulation”, a book chapter in [1].

1 Purpose

This paper develops tools to study bacterial behavior, which is sometimes nondeterministic. It investigates a possible mechanism for this randomness: fluctuations in biochemical systems that regulate cell metabolism. In other papers, these systems are modeled using continuous-time Markov jump processes. This gives a rigorous treatment of stochastic dynamics. One remaining issue is parameter inference from data that are noisy, discrete-time, and partial. The paper addresses this problem via Bayesian inference and MCMC.

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 it can be disorienting, I’ll relay some of the biological relationships in short sentences in the next paragraph.

The protein *flagellin* helps bacteria move. The protein σ^D promotes *flagellin*. The *fla / che* operon¹ contains many motility-related genes, including the one for σ^D . The protein σ^A and the protein *CodY* both suppress the *fla / che* operon. Thus, they suppress σ^D , and they indirectly suppress *flagellin*. In fact, *CodY* also downregulates *flagellin* directly. This is easiest to digest as a figure.

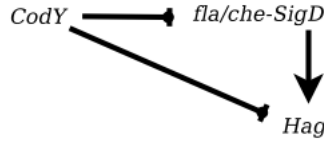


Figure 1: Regulatory relationships. *Hag* is the gene for *flagellin*, while *SigD* encodes σ^D .

EMK: Editing note: for consistency, fix every expression where θ is to the left of x .

3 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 producing q_{ij} particles of type j . I’ll refer to these en masse as $X(t)$, P , Q and once $R_i(t)$ is defined, $R(t)$. Defining the matrix S to be $Q^T - P^T$, and letting $R_i(t)$ denote the number of reactions of type i in $[0, t]$, then $X(t) - X(0) = SR(t)$. Under some assumptions, the different reaction channels evolve independently, and $R_i(t)$ is a Poisson process with intensity $c_i \int_0^t \prod_{j=1}^u \binom{X_j(t)}{p_{ij}}$. 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 ν_i . Define $t_0 \equiv 0$ and $t_{n+1} \equiv T$. When it comes to inference, you may recognize the likelihood as a competing hazards model from continuous Markov chain theory². EMK: This likelihood is a guess, because Wilkinson writes it using a function (h_0) that he never defines. Working for now on the fact that

¹Operons 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.

²At times, I choose to use equations that stand free of sentences. These I do not punctuate.

they're competing (homogeneous) Poisson processes. I'm fairly sure this is correct as I have it, or in other words, that h_0 is $\sum h_i$.

$$P(\nu, t|c) = \prod_{i=1}^n c_{\nu_i} \prod_{j=1}^u \binom{X_j(t_{i-1})}{p_{\nu_i j}} \exp(-c_{\nu_i} \int_0^T \binom{X_j(t)}{p_{\nu_i j}} dt)$$

Wilkinson points out that $\int_0^T \binom{X_j(t_{i-1})}{p_{\nu_i j}} dt$ is tractable; in fact, it is $\sum_{i=0}^n (t_{i+1} - t_i) \binom{X_j(t_i)}{p_{\nu_i j}}$. If the number of reactions of type k is r_k , then the likelihood becomes

$$P(\nu, t|c) = \prod_{k=1}^u \left\{ c_k^{r_k} \left\{ \prod_{j=1}^u \binom{X_j(t_{i-1})}{p_{kj}} \right\}^{r_k} \exp(-c_k \int_0^T \sum_j \binom{X_j(t)}{p_{kj}} dt) \right\}.$$

Still calling it the likelihood, Wilkinson omits the bracketed term $\prod_{j=1}^u \binom{X_j(t_{i-1})}{p_{kj}}^{r_k}$, probably because it does not involve c . Either way, the gamma distribution is conjugate for c_k , and setting independent priors $c_k \sim \Gamma(a_k, b_k)$, the posteriors are independent with $c_k \sim \Gamma(a_k + r_k, b_k + \int_0^T \sum_j \binom{X_j(t)}{p_{kj}} dt)$. This is an ideal scenario; missing data make inference more challenging.

4 Inference

In Wilkinson's data, not every reaction is recorded. Measurements are intermittent, and only some of the particle counts are measured. There is error in the measurements, to boot. The tidy scheme above scheme falls to pieces. Wilkinson says it better than I can: "Consider first the best-case scenario—perfect observation of the system at discrete times. Conditional on discrete-time observations, the Markov process breaks up into a collection of independent bridge processes that appear not to be analytically tractable." He goes on to mention alternative strategies, which include the MCMC scheme of [2] and the approximation [3] paired with inference via [EMK: cite more Wilkinson papers here](#).

The crux of the paper is a new competitor to all of these. It builds upon the fact that exact simulations from this model are possible, and also on the fact that measurement error helps soften the requirements on where the bridge should begin and end. Let θ include c , controlling the reaction rates, and τ , controlling the degree of measurement error. Let x denote the true state of the chain, but measured only at discrete times. That means the likelihood $P(x|\theta)$ is thought to be intractable. 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 a new value (x^*, θ^*) from $P(x, \theta|\mathcal{D}) = P(\theta)P(x|\theta)P(\mathcal{D}|x, \theta)$ using a proposal $f(\theta^*, x^*|\theta, x)$, it works out that the acceptance ratio must be the min of 1 and $\frac{f(\theta^*, x^*|\theta, x)}{f(\theta, x|\theta^*, x^*)} \times \frac{P(\theta)}{P(\theta^*)} \times \frac{P(x|\theta)}{P(x^*|\theta^*)} \times \frac{P(\mathcal{D}|x, \theta)}{P(\mathcal{D}|x^*, \theta^*)}$. In the model Wilkinson considers, $P(\mathcal{D}|x, \theta)$ is simple, but $P(x|\theta)$ is not feasible to compute. The insight is that one can cancel the 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 spills straight out if instead of drawing both x and θ from out-of-the-box proposals, we draw $\theta^* \sim f(\theta^*|\theta)$ and compute x^* via simulation with parameters θ^* . The end result is that the ratio of interest simplifies:

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

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

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} over time and use an M-H

step that alters only one coordinate of x . Suppose you draw inputs from $P(\theta, x | \mathcal{D}_{1:t})$, propose (θ^*, x_t^*) by perturbing (θ, x_t) slightly, and propose x_{t+1}^* with a new forward simulation. Suppose you then accept with probability $\min\{1, \frac{P(\mathcal{D}_{t+1} | x_{t+1}^*, \theta^*)}{P(\mathcal{D}_{t+1} | x_{t+1}, \theta)}\}$. The claim is that you are left with an exact draw from $P(\theta, x | \mathcal{D}_{1:t+1})$. **EMK: Why is this true? This doesn't quite seem to match any of the schemes discussed in 516. It seems to be sequential monte carlo, which I'm now trying to teach myself.** This leads to higher acceptance rates, and by starting with the prior and moving up through time, you arrive at samples from the posterior conditioned on all the data.

At this point, Wilkinson describes a connection to Approximate Bayesian Computation (ABC), and he describes software and computing tools that can be used to run the model.

5 An experiment

Wilkinson next describes a twelve-reaction regulatory model for these chemicals, spelling it out in table 1, and he lists three known reaction rates that can be used to test the model's performance. The prior distributions are each log-uniform over 4 orders of magnitude. The experiment assumes D_t is the number of σ^D molecules observed with Gaussian error s.d. 10 molecules. The initial state of the cell is assumed known, and observations occur every 5 minutes for 2 hours. Wilkinson says "1,000,000 particles were used, together with a burn-in of 1,000 iterations and a thin of 5[.]" **EMK: He never defines particles, so pending further reading, my best guess is that this means he starts with 1e6 samples from the prior, carrying each one forward. Not sure how the burn-in fits in.**

He also repeats this experiment, altering the measurement process so that it relies on a proxy for σ^D , a fluorescent protein (GFP). He performs inference by a) plugging GFP levels into the algorithm as if they were exactly proportional to σ^D and b) including the production and decay of GFP in the biochemical model. He concludes that plugging in GFP for σ^D leads to strong, incorrect posteriors, but that the full model is still useful.

6 Extensions

Wilkinson mentions many areas for future work.

- His likelihood-free MCMC scheme applies to any hidden, discretely observed continuous Markov process. It also can be used for multiple time-series $\mathcal{D}^1, \dots, \mathcal{D}^p$ by using posteriors from one as priors for the next.
- This method can fuse data from two bacterial strains, one missing a certain gene, by using posteriors from one experiment as priors for the next. It can also accomodate multiple fluorescent "reporters" for different molecules.
- Unsolved problems include scaling the method up to process data on batches of cells observed en masse and integrating data from microarrays or RNA-seq.

References

- [1] Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M.: Bayesian statistics 9 ninth Valencia international meeting on Bayesian statistics, Benidorm, Spain, 03-08.06.2010. Oxford U.P., Oxford (2012)
- [2] Boys, R.J., Wilkinson, D.J., Kirkwood, T.B.: Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing* **18**(2) (June 2008) 125–135
- [3] Gillespie, D.T.: The chemical langevin equation. *The Journal of Chemical Physics* **113**(1) (2000) 297–306