Parameter estimation in stochastic kinetic models

James B. Rawlings

Department of Chemical and Biological Engineering
University of Wisconsin–Madison

WIDDOW Seminar Series
Madison, Wisconsin
April 25, 2011

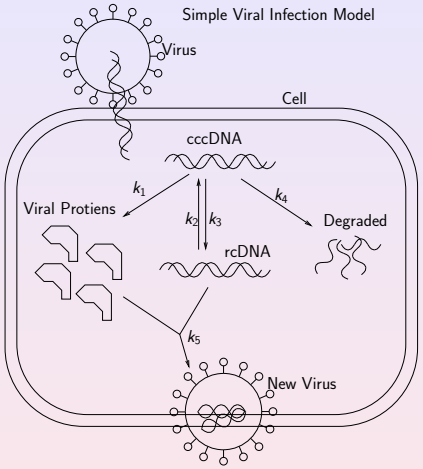
Outline

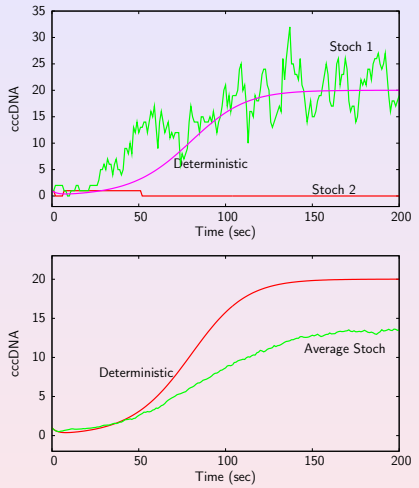
Introduction to stochastic kinetics

Stochastic kinetics

- Small species populations
- Species numbers are integers, reactions cause integer jumps
- Large fluctuations in species numbers and reaction rates
- Biological networks and catalyst particles

Reactions on small length scales: Virus infection

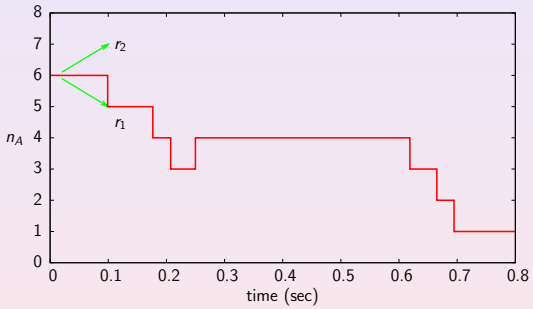




Average concentrations of small systems are not necessarily the same as the deterministic evolution.

Stochastic simulation (SSA) — Gillespie algorithm

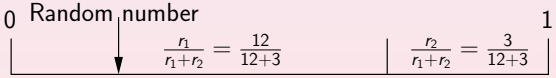
$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B \qquad k_1 = 2 \quad k_2 = 1 \qquad n_{A0} = 6 \quad n_{B0} = 3$$



SSA Algorithm

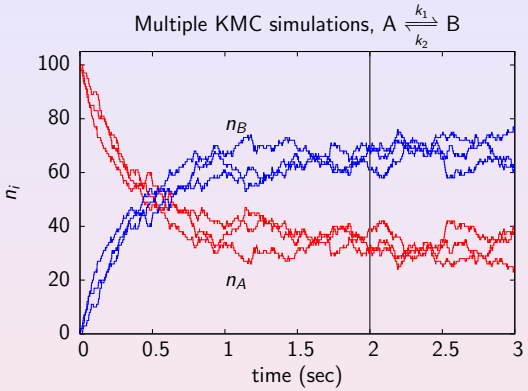
1. Choose which reaction
2. Choose time step
3. Repeat

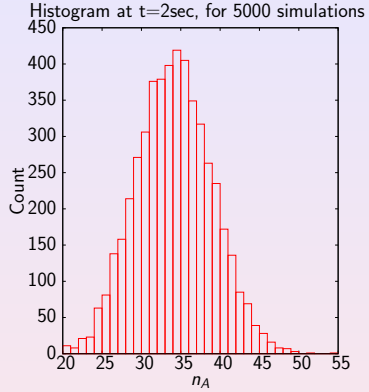
- Which reaction:



- Time step: Sample from an exponential distribution where the distribution mean is the sum of reaction rates.

SSA simulations and probability





- SSA simulations are samples of a probability distribution that evolves in time.
- We can write the evolution equation for the probability density (chemical master equation).

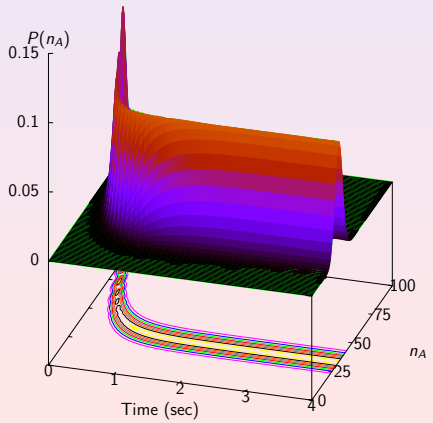
Chemical master equation

$$\frac{dP(x)}{dt} = \sum_{j=1}^{N_{rxn}} \underbrace{r_j(x-\nu_j)P(x-\nu_j)}_{\text{rate into state } x} - \underbrace{r_j(x)P(x)}_{\text{rate out of state } x}$$

$$\frac{dP}{dt} = AP$$

Master equation example

- A $\underset{k_2}{\overset{k_1}{\rightleftharpoons}}$ B
- $n_{A0} = 100, n_{B0} = 0$
- $k_1 = 2, k_2 = 1$
- 101 possible states
- 101 Coupled ODEs



Master equation — Important points

Chemical master equation

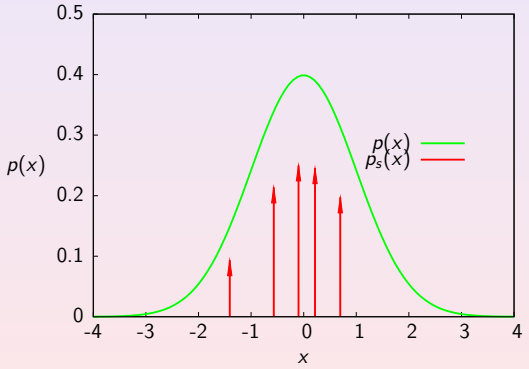
$$\frac{dP(x)}{dt} = \sum_{j=1}^{N_{rxn}} \underbrace{r_j(x-\nu_j)P(x-\nu_j)}_{\text{rate into state } x} - \underbrace{r_j(x)P(x)}_{\text{rate out of state } x}$$

$$\frac{dP}{dt} = AP$$

- Often the dimensionality of the master equation makes direct solution infeasible
- The master equation shows what probability distribution is sampled in an SSA simulation
- A reduced master equation can lead to a new/faster simulation schemes

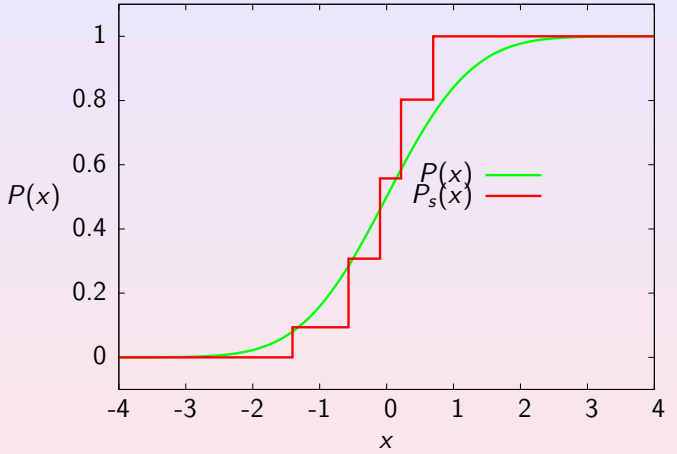
The sampled density

$$p_s(x) = \sum_{i=1}^{s} w_i \delta(x - x_i) \qquad x_i \text{ samples} \qquad w_i \text{ weights}$$



Exact density $p(x)$ and a sampled density $p_s(x)$ with five samples for $\xi \sim N(0,1)$

The cumulative sampled density



Corresponding exact $P(x)$ and sampled $P_s(x)$ cumulative distributions

Convergence of the sampled density with sample number

Define a measure of sampling error

$$D_s = \sup_x |P_s(x) - P(x)|$$

Theorem (Kolmogoroff (1933))

Suppose that $P(x)$ is continuous. Then for every fixed $z \geq 0$ as $s \to \infty$

$$\Pr\left(D_s \leq z s^{-1/2}\right) \to L(z)$$

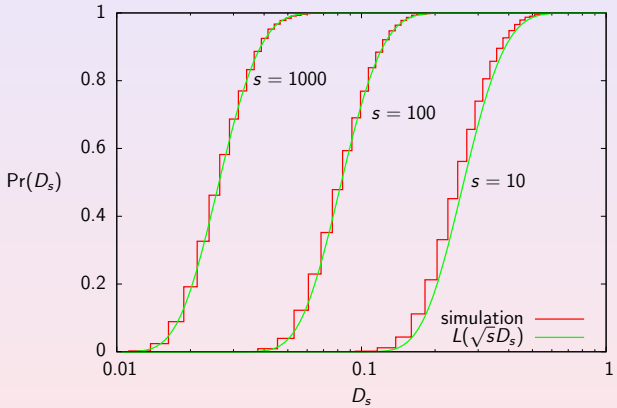
in which $L(z)$ is the cumulative distribution function given for $z > 0$ by

$$L(z) = \sqrt{2\pi} z^{-1} \sum_{\nu=1}^{\infty} e^{-(2\nu-1)^2 \pi^2 / 8z^2}$$

and $L(z) = 0$ for $z \leq 0$.

The sampling error distribution in pictures

Cumulative distribution for the sampling error $\Pr(D_s)$ of a unit variance normal for three different sample sizes, $s = 10, 100, 1000$



Distribution from simulation using 5000 samples (red) and Kolmogorov limiting distribution (green)

Simpler model . . . call me the tumbling dice. . .

We suspect the die may be unfair on the values of 1 and 6

$$p_1 = \frac{1-\theta}{6} \qquad p_2 = p_3 = p_4 = p_5 = \frac{1}{6} \qquad p_6 = \frac{1+\theta}{6}$$

We watch $n = 100$ rolls and want to estimate the unfairness θ.

The experimental measurement; $n = 100$ rolls

```
y =
     2     4     2     2     2     6     6     4     6     4     2     4     4     4     6     4     6     3     4     3
     5     5     1     2     5     2     6     6     2     1     6     5     1     4     6     6     5     2     4     2
     4     5     6     3     5     6     3     4     6     5     3     4     5     3     5     4     5     6     6     5
     6     4     4     6     4     6     6     2     6     3     2     3     5     3     1     1     6     4     1     1
     3     1     6     4     4     5     4     3     5     1     6     3     2     1     5     6     6     2     6     3
```

What are the odds of obtaining this outcome again by sampling?

$$\begin{aligned} p(y;\theta) &= \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\cdots\left(\frac{1+\theta}{6}\right)\cdots\left(\frac{1-\theta}{6}\right)\cdots\left(\frac{1}{6}\right) \\ &= (1/6)^{100} \quad (1-\theta)^{10}(1+\theta)^{26} \\ &= 1.5\times 10^{-78} \quad (1-\theta)^{10}(1+\theta)^{26} \end{aligned}$$

We should live so long

The odds of obtaining this outcome are of the order

$$10^{-78}$$

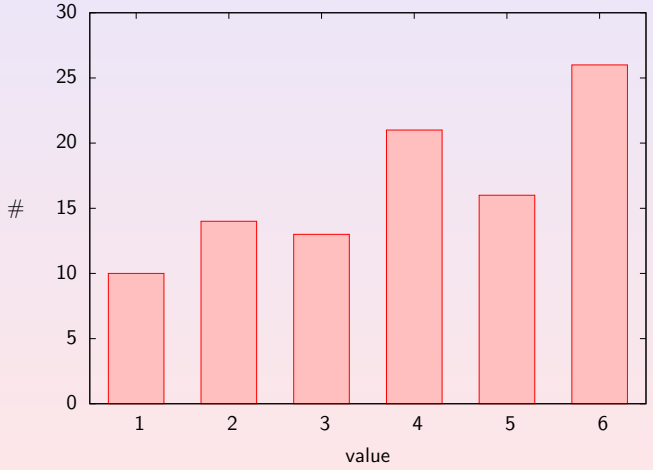
and until we sample this exact outcome of 100 die rolls again, the conclusion is

$$p_s(y;\theta) = 0$$

and that's a highly inefficient way to analyze the experiment!

The frequency count of the measurement; $n = 100$ rolls

Because the rolls are assumed *independent*, the ordering of the outcomes is not important. So then we look at frequency count



Our friend the multinomial distribution

Reinterpret the meaning of "measurement" to frequency count

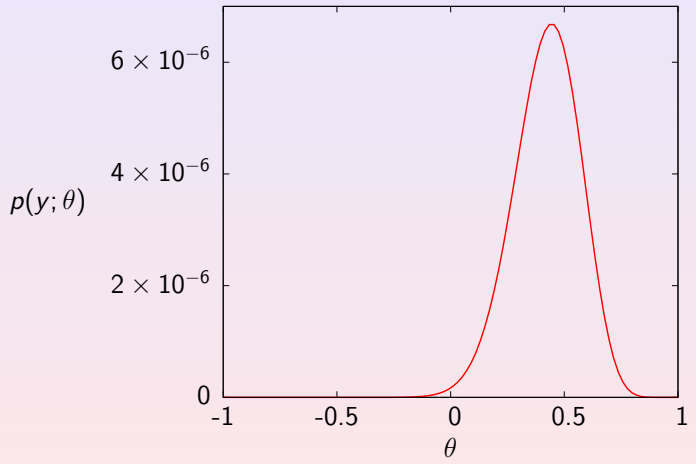
$$y = \begin{bmatrix} 10 & 14 & 13 & 21 & 16 & 26 \end{bmatrix}$$

$$p(y) = \frac{n!}{y_1! y_2! \cdots y_6!} p_1^{y_1} p_2^{y_2} \cdots p_6^{y_6}$$

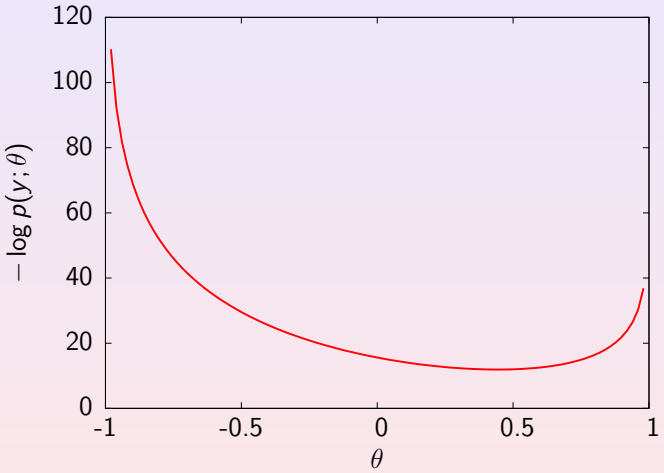
For our problem,

$$p(y;\theta) = \frac{100!}{10!14! \cdots 26!} \left(\frac{1-\theta}{6}\right)^{10} (1/6)^{14} \cdots (1/6)^{16} \left(\frac{1+\theta}{6}\right)^{26}$$
$$p(y;\theta) = 1.1 \times 10^{71} \quad (1/6)^{100} \quad (1-\theta)^{10}(1+\theta)^{26}$$
$$p(y;\theta) = 1.7 \times 10^{-7} \quad (1-\theta)^{10}(1+\theta)^{26}$$

The likelihood for our measurement



The (negative) log likelihood for our measurement



How many samples?

Still, the probability of obtaining the measured outcome

$$y = \begin{bmatrix} 10 & 14 & 13 & 21 & 16 & 26 \end{bmatrix}$$

is of the order

$$10^{-7}$$

and we require on the order 10^7 samples (simulations) before we have a reasonable chance of concluding

$$p_s(y;\theta) \neq 0$$

Although considerably better than 10^{78} simulations, 10^7 simulations is still inefficient.

We would like to obtain a good estimate of $p(y;\theta)$ for a given θ from a *single* simulation.

Measurement error? What measurement error?

Now let's introduce the concept of error in the measurement process. Why?

1. It's often required for a good model of the sensor
2. Even discrete measurements may have error
   Bush v. Gore, 2000, the hanging chad
   2011 WI Supreme Court election, the Waukesha county votes
3. It introduces nonzero probability without exact match of experiment

New random variable v, which distinguishes the state of the die from our measurement of the state

$$y = x + v \qquad v \sim N(0, R)$$

We often model v as a zero mean normal with covariance R. Both discrete and continuous v are useful.

Likelihood of data with measurement error

The explanation of a nonmatching measurement y to the sampled density p_s is the measurement error

$$p_s(y \mid x) = p_v(y - x)$$

$$p_s(y \mid x) = \frac{1}{(2\pi)^{n/2} |R|^{1/2}} \quad e^{-\frac{1}{2}(y-x)' R^{-1}(y-x)}$$

Use as our estimator

$$\lim_{R \to 0} \max_{\theta} p_s(y \mid x; \theta)$$

The limit $R \to 0$ accounts for the zero measurement error case

Equivalent estimator

$$\lim_{R \to 0} \max_{\theta} \quad p_s(y \mid x; \theta)$$

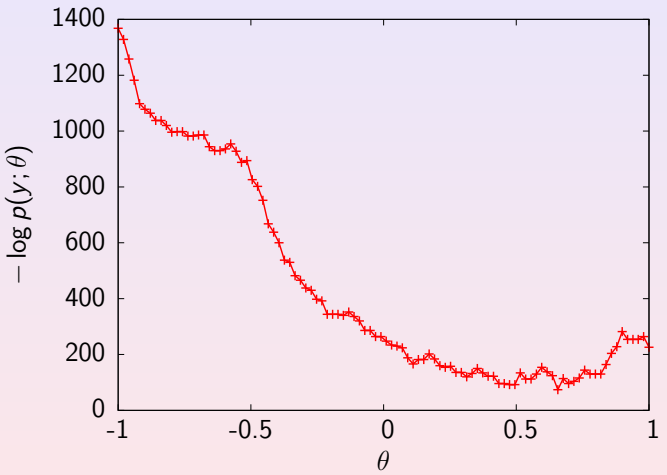
$$\lim_{R \to 0} \min_{\theta} \quad -\log p_s(y \mid x; \theta)$$

$$\lim_{R \to 0} \min_{\theta} \quad (n/2)\log(2\pi) + (1/2)\log(|R|) + (1/2)(y - x)' R^{-1}(y - x)$$

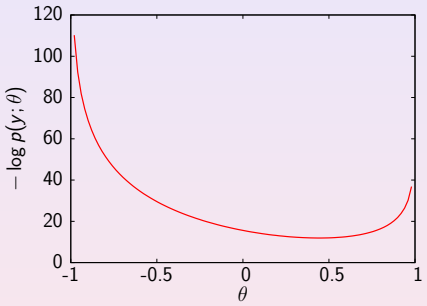
$$\min_{\theta} \quad (y - x)' R^{-1}(y - x)$$

- Our good friend, least squares.
- If we keep sampling until some $x = y$, then all the other x's drop out of the calculation as $R \to 0$. This limit captures using the frequency count of the matches as the likelihood.

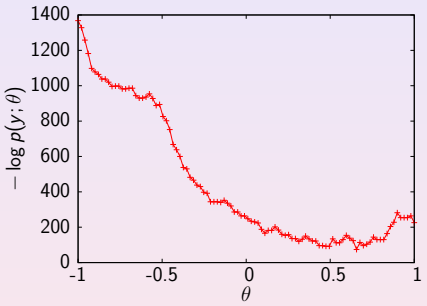
Likelihood with one simulation at each θ



Recall the true likelihood



true likelihood



one simulation

Likelihood with one simulation at each θ; the *same* versus *different* randomness

Same random numbers

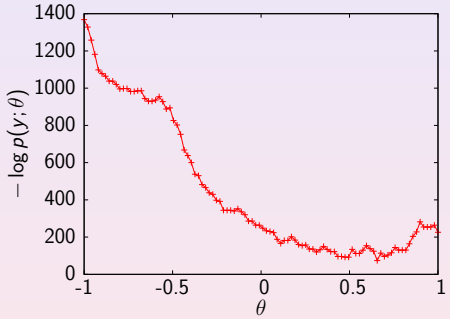
```
u = rand(nrolls,1);
thetavec = linspace(-1, 1, 100);
for j = 1:length(thetavec)
  theta = thetavec(j);
  p(1) = 1/6*(1 - theta);
  p(6) = 1/6*(1 + theta);
  P = cumsum(p);
  % do one simulation for each theta
  for i = 1:nrolls
    rollx(i) = sum(u(i) >= P) + 1;
  endfor
```

Different random numbers

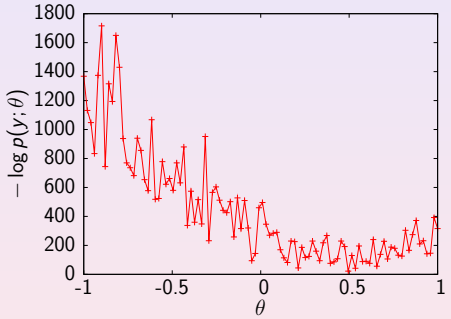
```
for j = 1:length(thetavec)
  theta = thetavec(j);
  p(1) = 1/6*(1 - theta);
  p(6) = 1/6*(1 + theta);
  P = cumsum(p);
  % do one simulation for each theta
  u = rand(nrolls,1);
  for i = 1:nrolls
    rollx(i) = sum(u(i) >= P) + 1;
  endfor
```


Likelihood with one simulation at each θ; the *same* versus *different* randomness

Same random numbers



Different random numbers



In the limit of infinite samples, it won't matter, but we live in the world of finite (often small!) samples.

Sources of randomness in the likelihood function

- The measurement is random
- The samples, and hence the sampled density are random
- We have to live with the effects of the first one (or ask for more measurements).
- We can reduce the effects of the second one with increased simulation.

Convergence with number of samples

Data

$$y = \begin{bmatrix} 10 & 14 & 13 & 21 & 16 & 26 \end{bmatrix}$$

Measurement error

$$y = x + v \qquad v \sim N(0, R)$$

Samples

$$x_1 = \begin{bmatrix} 7 & 15 & 22 & 21 & 10 & 25 \end{bmatrix}$$
$$x_2 = \begin{bmatrix} 7 & 16 & 26 & 12 & 14 & 25 \end{bmatrix}$$
$$\vdots$$
$$x_s = \begin{bmatrix} 21 & 13 & 8 & 20 & 20 & 18 \end{bmatrix}$$

Computing the likelihood

From sampled density of x to measurement y

$$p(y) = \int p(y \mid x) p_s(x) dx$$

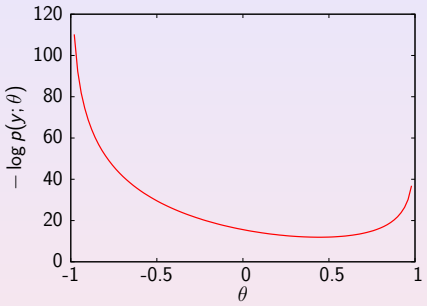
Measurement equation and sampled density

$$p(y \mid x) = p_v(y - x) \qquad p_s(x) = \frac{1}{s} \sum_{i=1}^{s} \delta(x - x_i)$$

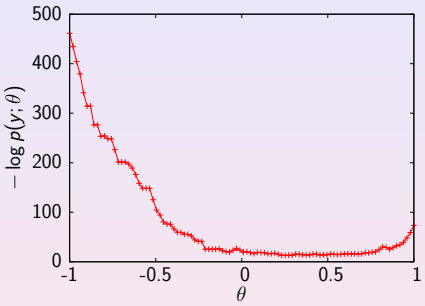
Combining

$$p(y) = \int p_v(y - x) \frac{1}{s} \sum_{i=1}^{s} \delta(x - x_i) dx$$
$$= \frac{1}{s} \sum_{i=1}^{s} p_v(y - x_i)$$
$$p(y) = \frac{1}{s(2\pi)^3 |R|^{1/2}} \sum_{i=1}^{s} e^{-\frac{1}{2}(y - x_i)' R^{-1} (y - x_i)}$$

Convergence with number of samples



true likelihood



$s = 1000$ samples

It's nice to establish convergence as $s \to \infty$ (and $R \to 0$), but convergence won't be achieved in the class of applications of interest.

Confidence intervals for the parameter

- Now consider θ to be a *random* variable with prior $p(\theta)$

$$p(\theta \mid y) = \frac{p(y \mid \theta) p(\theta)}{\int p(y, \theta) d\theta}$$
$$\propto p(y \mid \theta) p(\theta)$$
$$\propto L(y; \theta) p(\theta)$$

- For a uniform (noninformative) prior on a chosen compact set

$$p(\theta \mid y) \propto L(y; \theta)$$

- May use a sampling strategy to obtain mean and variance of the posterior $p(\theta \mid y)$.
- May instead use a quadratic approximation of $L(y; \theta)$ near θ^0.

Literature overview for parameter estimation of stochastic kinetic models

- Boys et al. (2008) propose generating many samples of the full master equation consistent with the given measurement. They then use Markov chain Monte Carlo to obtain the posterior distribution of the parameter. The first step is computationally intractable for the models of interest here.
- Golightly and Wilkinson (2008) use the Fokker-Planck approximation of the master equation. This diffusion approximation is not generally applicable in stochastic chemical kinetics.
- For ODE models, Toni et al. (2009) approximate the likelihood by measuring the distance between experimental data and a simulation. They use sequential Monte Carlo to obtain the posterior.

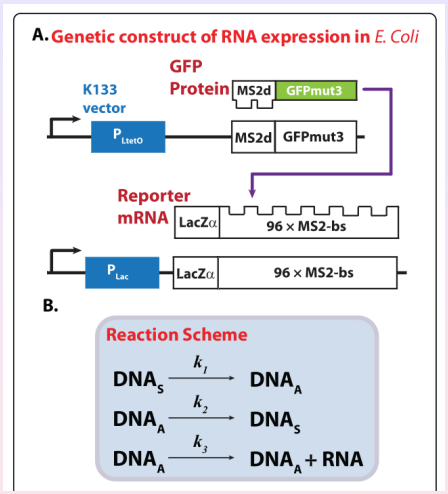
Literature overview for parameter estimation of stochastic kinetic models

- Tian et al. (2006) express the likelihood $p(y|\theta)$ as a product of transition densities $p(y|\theta) = \prod_{i=1}^{n} p(y_{i+1}|y_i, \theta)$. Each $p(y_{i+1}|y_i, \theta)$ is evaluated using 5000 SSA simulations. A genetic algorithm is used to maximize $p(y|\theta)$. This procedure is computationally inefficient because 5000 SSA simulations are used for each transition.
- Reinker et al. (2006) calculate the likelihood analytically using an artificial maximum number of reactions that can occur within a given time interval. They use a quasi-newton method to maximize the likelihood. The assumption about the maximum number of reactions is unrealistic.

Literature overview for parameter estimation of stochastic kinetic models

- Poovathingal and Gunawan (2010) propose to evaluate likelihood using the solution to the master equation. Their proposed function is not the likelihood, but some other merit function. They estimate the solution of the master equation by SSA simulations. This is computationally intensive and requires a binning strategy. They use directed evolution to optimize.

Stochastic kinetics example — RNA dynamics in E. coli[1]

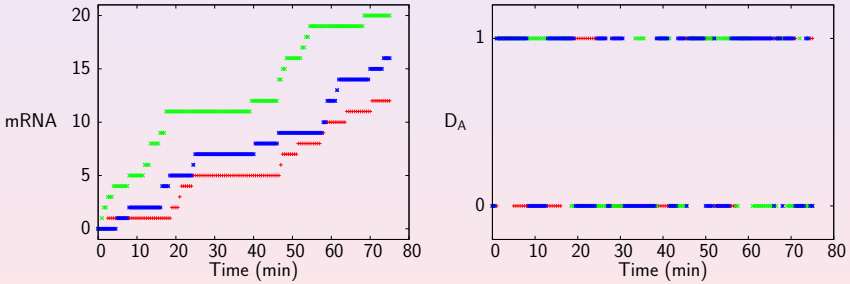


- Model to explain mRNA dynamics in E. Coli
- Three unknown parameters k_1, k_2, k_3

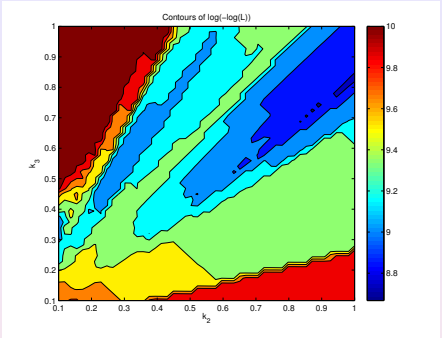
[1] Poovathingal and Gunawan (2010); Golding et al. (2005)

Experimental data generation

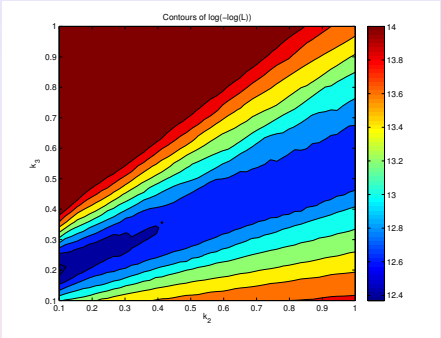
Simulate the model using SSA to generate data:
$k_1 = 0.28, k_2 = 0.17, k_3 = 0.4$



Parameter estimate with one experiment



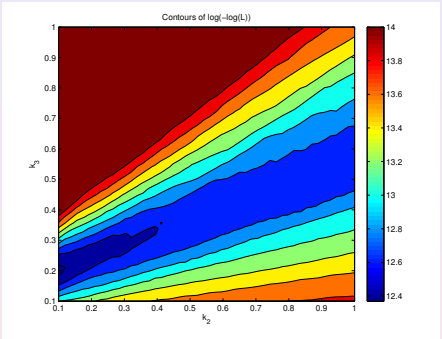
1 simulation



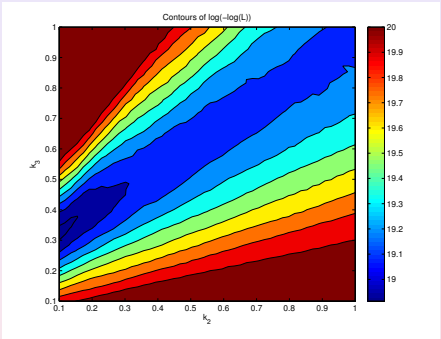
200 simulations

- True parameters are $k_1 = 0.277, k_2 = 0.1667, k_3 = 0.4$
- Assume that the true value of k_1 is known
- Likelihood is nonsmooth with only one experiment and one simulation

Convergence with replication of experiments

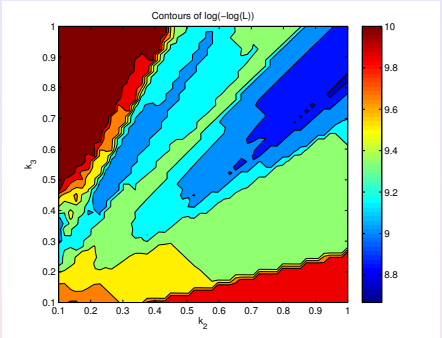


1 experiment and 200 simulations

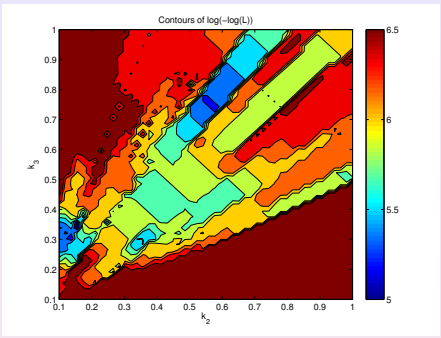


200 experiments and 200 simulations

Parameter estimate with one experiment



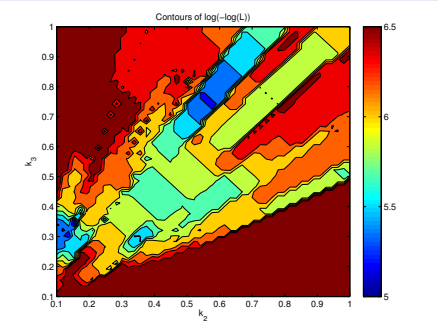
1 experiment and 1 simulation



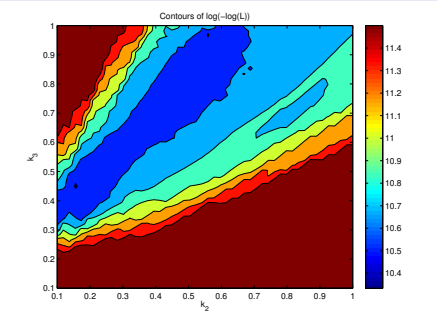
1 experiment and 100 simulations

- True parameters are $k_1 = 0.277, k_2 = 0.1667, k_3 = 0.4$
- Assume that the true value of k_1 is known
- Likelihood is nonsmooth with only one experiment and one simulation

Convergence with replication of experiments



1 experiment and 100 simulations



100 experiments and 100 simulations

What's left to do? Lots!

Optimizers

- The optimization desirables: efficient methods to deal with high dimensional parameter vector, ill-conditioned estimation problem, noisy likelihood, and constraints.
- Adaptively decide when and by how much to increase sampling to reduce effects of noise
- Criteria for termination

What's left to do?

Analysts/Probabilists/Statisticians

- How to efficiently calculate parametric derivatives (gradients)
- How to efficiently calculate (approximate) confidence intervals
- Issues of convergence, variability, bias; effects of sample size

What's left to do?

Experimentalists

- Do the models handle the experiments of interest
- Are the data demands realistic
- Feedback on: (i) application with relevant datasets; (ii) conceptual framework; (iii) software implementation
- Efficiency, accuracy, robustness, convenience

What's left to do?

Software designers

- To have wide impact, software tools have to be developed!
- Who are the end users? How much experience is required?
- What is an appropriate interface for the expected users? The expert users? Do they use the same interface? How do we support software for multiple user groups?
- How do we maintain and extend the first generation software? How do we obtain necessary user feedback?
- What is the expected shelf-life of the developed methods?

Acknowledgment

JBR would like to thank Rishi Srivastava and Ankur Gupta of UW-Madison for helpful feedback on the presentation.

Further Reading I

R. Boys, D. Wilkinson, and T. Kirkwood. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*, 18(2):125–135, 2008. ISSN 0960-3174.

I. Golding, J. Paulsson, S. Zawilski, and E. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, 2005. ISSN 0092-8674.

A. Golightly and D. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674–1693, 2008. ISSN 0167-9473.

A. Kolmogoroff. Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari*, 4:1–11, 1933.

S. Poovathingal and R. Gunawan. Global parameter estimation methods for stochastic biochemical systems. *BMC bioinformatics*, 11(1):414, 2010. ISSN 1471-2105.

Further Reading II

S. Reinker, R. Altman, and J. Timmer. Parameter estimation in stochastic biochemical reactions. *IEE Proc.-Syst. Biol*, 153(4):168, 2006.

T. Tian, S. Xu, J. Gao, and K. Burrage. Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics*, 23(1):84, 2006. ISSN 1367-4803.

T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6 (31):187, 2009. ISSN 1742-5689.