Designing experiments to understand the variability in biochemical reaction networks.

Jakob Ruess

ETH Zurich

June 5, 2013
Contents

1 Stochastic Modeling - Why and how?

2 Experimental Design

3 Applications
1 Stochastic Modeling - Why and how?

2 Experimental Design

3 Applications
1 Stochastic Modeling - Why and how?

2 Experimental Design

3 Applications
Stochastic Models

\[
\begin{align*}
& a \quad \text{mRNA} \\
& b \quad \text{mRNA} \\
& c \quad \text{mRNA} + \text{protein} \\
& d \quad \text{protein}
\end{align*}
\]

ODEs

Protein Mean

Mean Number of Molecules

Time

Jakob Ruess (ETH Zurich)  
Automatic Control Laboratory  
June 5, 2013  
4 / 19
Stochastic Models

Chemical Master Equation

Jakob Ruess (ETH Zurich)
Stochastic Models

Parameter Inference:
Search for parameters that lead to agreement of model predictions and measurements.
Stochastic Models

Chemical Master Equation

Experimental Design:
Search for the experiment which maximizes information about the parameters.
Stochastic Models

Parameter Inference:
Search for parameters that lead to agreement of model predictions and measurements.

Experimental Design:
Search for experiment which maximizes information about the unknown parameters.
Stochastic Models

Parameter Inference: Search for parameters that lead to agreement of model predictions and measurements.

Experimental Design: Search for experiment which maximizes information about the unknown parameters.

Chemical Master Equation
Parameter Inference: Search for parameters that lead to agreement of model predictions and measurements.

Experimental Design: Search for experiment which maximizes information about the unknown parameters.

Moment Equations

- Protein Mean
- Protein Variance
Outline

1. Stochastic Modeling - Why and how?
2. Experimental Design
3. Applications
A measure of the information that an observed random variable $Y$ carries about a vector $\theta$ of unknown parameters which parametrizes its distribution $f(Y; \theta)$:

\[
(I(\theta))_{i,j} = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(Y; \theta) \right) \left( \frac{\partial}{\partial \theta_j} \log f(Y; \theta) \right) \right]
\]

In our case: $\theta$ are the model parameters and $f(Y; \theta)$ is the distribution of the measured species.
Experimental Design

The Fisher Information

\[(I(\theta))_{i,j} = \mathbb{E} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(Y; \theta) \right) \left( \frac{\partial}{\partial \theta_j} \log f(Y; \theta) \right) \right] \]

Optimal Experiment

\[u^* = \arg \max_{u \in \mathcal{E}} \{ \det (I(\theta, u)) \}, \]

where \(I(\theta, u)\) is the Fisher information for experiment \(u\) and \(\mathcal{E}\) is the set of possible experiments.
The Fisher information for Gaussian measurements.

Assume for a second that $f(Y; \theta)$ is a Gaussian distribution with mean $m$ and variance $\sigma^2$. Then it holds that

Fisher information for Gaussian measurements

$$(I(\theta))_{i,j} = \frac{\partial m}{\partial \theta_i} \frac{\partial m}{\partial \theta_j} \sigma^2 + \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j} \sigma^4.$$ 

This can be used to compute the information under the linear noise approximation.

However, in most applications the process cannot be well approximated by a Gaussian...
The Fisher information for Gaussian measurements.

Assume for a second that $f(Y; \theta)$ is a Gaussian distribution with mean $m$ and variance $\sigma^2$. Then it holds that

\begin{equation}
(I(\theta))_{i,j} = \frac{\partial m}{\partial \theta_i} \frac{\partial m}{\partial \theta_j}{\sigma^2} + \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j}{\sigma^4}.
\end{equation}

This can be used to compute the information under the linear noise approximation.

However, in most applications the process cannot be well approximated by a Gaussian...
The Fisher information for Gaussian measurements.

Assume for a second that $f(Y; \theta)$ is a Gaussian distribution with mean $m$ and variance $\sigma^2$. Then it holds that

\[
(I(\theta))_{i,j} = \frac{\partial m}{\partial \theta_i} \frac{\partial m}{\partial \theta_j} \frac{1}{\sigma^2} + \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j} \frac{1}{\sigma^4}.
\]

This can be used to compute the information under the linear noise approximation. However, in most applications the process cannot be well approximated by a Gaussian...
The Fisher information of sample mean and variance.

**Approach:** Instead of trying to compute the total information of the sample we focus on the information of sample mean and variance.

The central limit theorem implies for large enough sample size $n$:

The joint distribution of sample mean and variance:

$$Y := [\hat{\mu}_1, \hat{\mu}_2]^T \sim \mathcal{N}([\mu_1, \mu_2]^T, \Sigma),$$

where

$$\Sigma = \frac{1}{n} \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \frac{n-3}{n-1} \mu_2^2 \end{pmatrix}.$$  

Hence, the distribution $f(Y, \theta)$ is now really a Gaussian (even though we did not use a Gaussian approximation of the underlying process).
The Fisher information of sample mean and variance.

**Approach:** Instead of trying to compute the total information of the sample we focus on the information of sample mean and variance. The central limit theorem implies for large enough sample size $n$:

The joint distribution of sample mean and variance

\[ Y := [\hat{\mu}_1, \hat{\mu}_2]^T \sim \mathcal{N}\left( [\mu_1, \mu_2]^T, \Sigma \right), \text{ where} \]

\[ \Sigma = \frac{1}{n} \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \frac{n-3}{n-1} \mu_2 \end{pmatrix}. \]

Hence, the distribution $f(Y, \theta)$ is now really a Gaussian (even though we did not use a Gaussian approximation of the underlying process).
The Fisher information of sample mean and variance.

**Approach:** Instead of trying to compute the total information of the sample we focus on the information of sample mean and variance. The central limit theorem implies for large enough sample size $n$:

The joint distribution of sample mean and variance

\[ Y := [\hat{\mu}_1, \hat{\mu}_2]^T \sim \mathcal{N} \left( [\mu_1, \mu_2]^T, \Sigma \right), \text{ where} \]

\[ \Sigma = \frac{1}{n} \begin{pmatrix} \mu_2 & \frac{\mu_3}{n-1} \\ \mu_3 & \mu_4 - \frac{n-3}{n-1} \mu_2 \end{pmatrix}. \]

Hence, the distribution $f(Y, \theta)$ is now really a Gaussian (even though we did not use a Gaussian approximation of the underlying process).
The Fisher information of sample mean and variance.

Information of sample mean and variance

\[(I_S(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \mu_2 + n \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right) \mu_2^2 (\mu_4 - \mu_2^2) - \mu_2 \mu_3^2 \]
Some special cases.

So what happens when the underlying distribution is actually Gaussian?

Information (general)

\[
(l_S(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \mu_2 + \frac{n}{\mu_2} \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right) \mu_2 \left( \mu_4 - \mu_2^2 \right) - \mu_2 \mu_3^2.
\]
Some special cases.

Information (general)

\[
(I_S(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \mu_2 + n \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right) \frac{\mu_2^2}{\mu_2^2 (\mu_4 - \mu_2^2) - \mu_2 \mu_3^2}.
\]

For a Gaussian it holds that $\mu_3 = 0$ and $\mu_4 = 3 \mu_2^2 = 3 \sigma^4$. 
Some special cases.

Information (general)

\[
(Is(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \mu_2 + n \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right).
\]

For a Gaussian it holds that \( \mu_3 = 0 \) and \( \mu_4 = 3\mu_2^2 = 3\sigma^4 \).

Information for a Gaussian

\[
(Is(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \frac{1}{\sigma^2} + n \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j} \frac{1}{\sigma^4}.
\]
Some special cases.

Information (general)

\[(I_S(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j}{\mu_2} + n \frac{\left(\mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3\right) \left(\mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3\right)}{\mu_2^2 (\mu_4 - \mu_2^2) - \mu_2 \mu_3^2}.\]

Information for a Gaussian

\[(I_S(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j}{\sigma^2} + n \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j}{\sigma^4}.\]

What about a Poisson distribution?
Some special cases.

**Information (general)**

\[
(l_s(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \mu_2 + n \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right).
\]

**Information for a Gaussian**

\[
(l_s(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \sigma^2 + n \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j} \sigma^4.
\]

What about a Poisson distribution?
For a Poisson distribution it holds that \( \mu_1 = \mu_2 = \mu_3. \)
Some special cases.

### Information (general)

\[
(I_{S}(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \frac{\partial \mu_1}{\partial \theta_i} + n \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_i} - \frac{\partial \mu_1}{\partial \theta_i} \mu_3 \right) \left( \mu_2 \frac{\partial \mu_2}{\partial \theta_j} - \frac{\partial \mu_1}{\partial \theta_j} \mu_3 \right) \frac{\mu_2^2}{\sigma^2} \left( \mu_4 - \mu_2^2 \right) - \mu_2 \mu_3^2.
\]

### Information for a Gaussian

\[
(I_{S}(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \frac{\partial \sigma^2}{\partial \theta_i} + \frac{1}{2} \frac{\partial \sigma^2}{\partial \theta_i} \frac{\partial \sigma^2}{\partial \theta_j}.
\]

### Information for a Poisson

\[
(I_{S}(\theta))_{i,j} \approx n \frac{\partial \mu_1}{\partial \theta_i} \frac{\partial \mu_1}{\partial \theta_j} \frac{\partial \mu_1}{\partial \theta_j} \frac{\partial \mu_1}{\partial \theta_i}.
\]
An experimental design framework for chemical reaction systems.

\[ u^* = \arg \max_{u \in \mathcal{E}} \{ \det (I_S(\theta, u)) \}, \]

where \( I_S(\theta, u) \) is computed from the moments.

- Instead of solving the complete CME we only need to solve the moment equations for the moments of order up to 4.
- The only required assumption is that the sample is of sufficient size for applicability of the central limit theorem.
- However, the moment equations may not be solvable...
An experimental design framework for chemical reaction systems.

\[ u^* = \arg \max_{u \in \mathcal{E}} \{ \det (I_S(\theta, u)) \}, \]

where \( I_S(\theta, u) \) is computed from the moments.

- Instead of solving the complete CME we only need to solve the moment equations for the moments of order up to 4.
- The only required assumption is that the sample is of sufficient size for applicability of the central limit theorem.
- However, the moment equations may not be solvable...
An experimental design framework for chemical reaction systems.

\[ u^* = \arg \max_{u \in E} \{ \det (I_S(\theta, u)) \}, \]

where \( I_S(\theta, u) \) is computed from the moments.

- Instead of solving the complete CME we only need to solve the moment equations for the moments of order up to 4.
- The only required assumption is that the sample is of sufficient size for applicability of the central limit theorem.
- However, the moment equations may not be solvable...
An experimental design framework for chemical reaction systems.

\[ u^* = \arg \max_{u \in \mathcal{E}} \{ \det (I_S(\theta, u)) \}, \]

where \( I_S(\theta, u) \) is computed from the moments.

- Instead of solving the complete CME we only need to solve the moment equations for the moments of order up to 4.
- The only required assumption is that the sample is of sufficient size for applicability of the central limit theorem.
- However, the moment equations may not be solvable...
Outline

1. Stochastic Modeling - Why and how?
2. Experimental Design
3. Applications
A simple model of gene expression.
A simple model of gene expression.

To make this a bit more interesting

Varying mRNA production rate

\[ da_t = r(m - a_t)dt + s\sqrt{a_t}dW_t, \]
A simple model of gene expression.

To make this a bit more interesting

Varying mRNA production rate

\[ da_t = r(m - a_t)dt + s\sqrt{a_t}dW_t, \]

Parameters: \( b, c, d, \mu_a, V_a, r \)
A simple model of gene expression.

To make this a bit more interesting:

**Varying mRNA production rate**

\[ da_t = r(m - a_t)dt + s\sqrt{a_t}dW_t, \]

Parameters: \( c, \mu_a, V_a, r \)
A comparison of unplanned and optimal experiments.

**Table:** Comparison of different experimental approaches.

<table>
<thead>
<tr>
<th></th>
<th>Unplanned experiment</th>
<th>Optimally designed experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_a$</td>
<td>0.0009</td>
<td>1.1125</td>
</tr>
<tr>
<td>$V_a$</td>
<td>0.0002</td>
<td>0.1286</td>
</tr>
<tr>
<td>$c$</td>
<td>0.0009</td>
<td>1.1817</td>
</tr>
<tr>
<td>$r$</td>
<td>0.0012</td>
<td>0.0129</td>
</tr>
</tbody>
</table>
Confidence Ellipses

A

B

C

D

E

F
Light induced gene expression

Experiment

[a] Red light (650 nm)
Gal4 AD
PIF3
Promoter
PhyB
Pr
Gal1 UAS
Venus

[b] Far-red light (730 nm)
Gal4 AD
PIF3
Promoter
PhyB Pfr
Gal1 UAS Venus

YFP signal

[c] Average fluorescence fold change
Time (min)
R
R
R
R
R
0 50 100 150 200 250 300

[d] +PCB –light
Light induced gene expression

Experiment

Overall worst, but best to identify the protein production rate

Overall best, but worst to identify the protein production rate

Average fluorescence fold change

Time (min)

2 6 10 14 18

0 50 100 150 200 250 300

R R R R

0 50 100 150 200 250 300 350

1 2 3 4 5 6 7

FR FR FR FR

Red light (650 nm)

Far-red light (730 nm)

Gal4 AD
PIF3

PhyB

Gal4 BD

Red light

Far-red light

Glt1 UAS

Venus

Overall worst, but best to identify the protein production rate

Overall best, but worst to identify the protein production rate

Jakob Ruess (ETH Zurich)
Automatic Control Laboratory
June 5, 2013
Light induced gene expression

Experiment

Average fluorescence fold change

Time (min)

0 50 100 150 200 250 300
R
R
R
R
R

0 50 100 150 200 250 300
FR
FR
FR
FR
FR

Best for r

Red light (650 nm)

PhyB

Pr

Gal1 UAS

Far-red light (730 nm)

PhyB Pfr

Gal1 UAS Venus

Jakob Ruess (ETH Zurich)
Automatic Control Laboratory
June 5, 2013
Thank you for your attention