DEVISING AN EXPERIMENTAL PROTOCOL FOR CONTROLING HIV INFECTION THROUGH STRUCTURED TREATMENT INTERRUPTIONS

Shannon Kubiak[1], Heather Lehr[2], Rachel Levy[3], Todd Moeller[4], Albert Parker[5], Edward Swim[6]

Problem Presenter: Sarah Holte, Ph.D.
Fred Hutchinson Cancer Research Center

NCSU Faculty Consultant: H.T. Banks, Ph.D.

Student Consultants: Brian Adams and David Bortz

[1]Towson University
[2]University of Texas at Austin
[3]North Carolina State University
[4]Georgia Institute of Technology
[5]Montana State University - Bozeman
[6]Texas Tech University

Outline

- Introduction

- Sensitivity Analysis
  - -Time schedule
  - -Parameters
  - -Compartments

- Inverse Problem
  - -Parameter Estimates

Introduction

Highly Active Anti-Retroviral Therapy (HAART) of HIV infection has significantly reduced morbidity and mortality in developed countries. However, since these treatments can cause side effects and require strict adherence to treatment protocol, questions about whether or not treatment can be interrupted or discontinued with control of infection maintained by the host immune system remain to be answered. We examine a model incorporating structured treatment interruptions (STI) in order to suggest a treatment protocol for experiments to be performed at Merck. We suggest a timing schedule for treatment, how data from the compartments should be collected, as well as estimates of the paramaters of interest.

First documented case from patient who went off the drug without telling doctors

Goals

We want to guide the treatment protocol for experiments at Merck by answering the following questions:

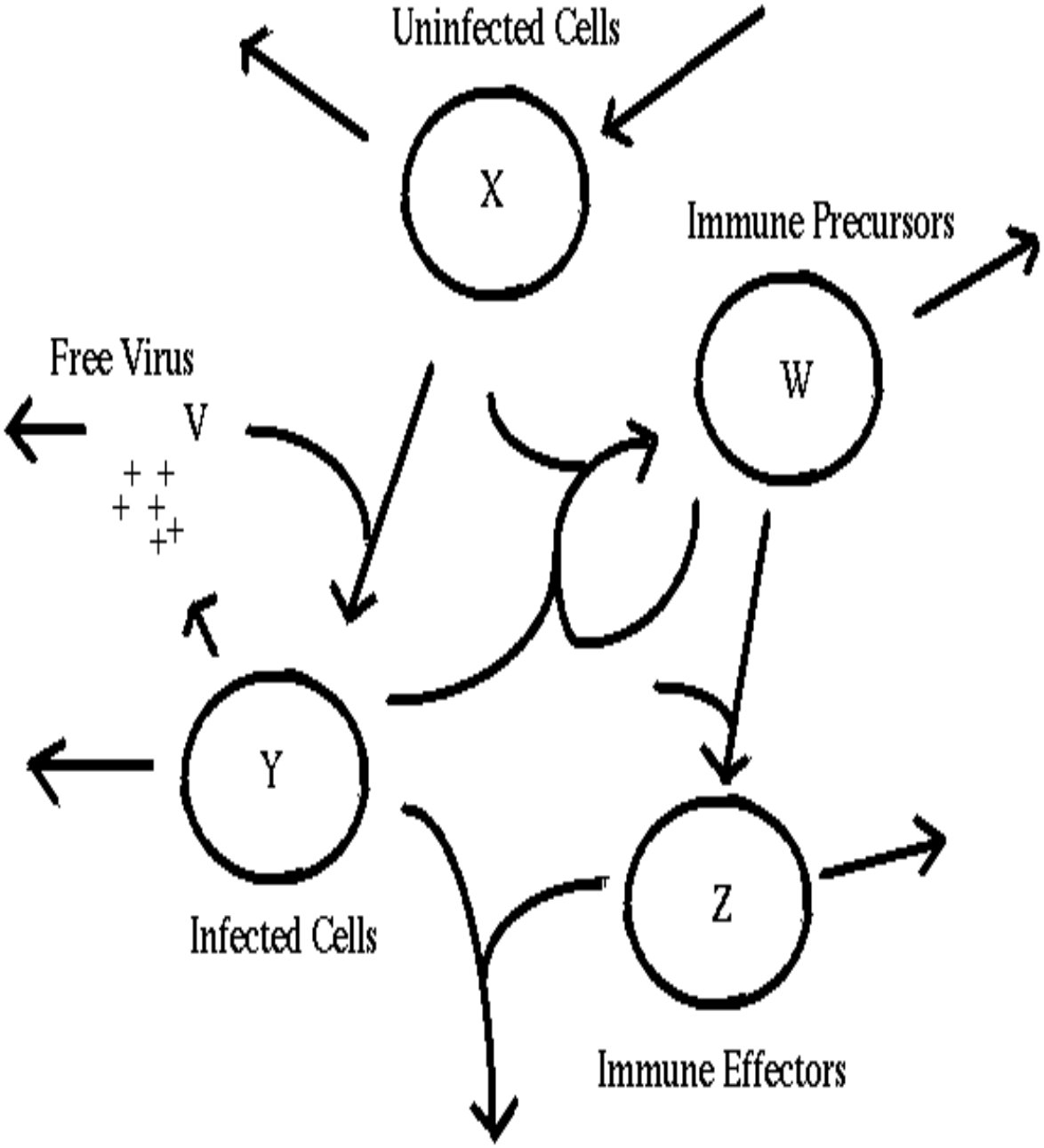
- When should data from compartments be collected?
- How should the compartments be collected?
- Which parameters most affect the model?
- What are good estimates for the parameters?

1. is important since patients can't come in every day for blood extractions

2. is important since technicians can measure X+Y cheaply, but measuring them individually is not. So how much do we loose if we can only measure X+Y combined rather than X and Y seperately?

3. We don't want to waste our time estimating parameters that don't affect the model very much.

The Model


Uninfected Cells
X
Immune Precursors
W
Free Virus
V
Y
Infected Cells
Z
Immune Effectors

- There exists an X (surveillance) for each disease.

- Early models used X, Y and Z. Wodarz-Nowak uses W. We have added the V.

- This model is similar to the normal process incurredby theimmune system for any attacking virus.

- After initial interaction between X and its specified V, a population of W stays in the system, which is why vaccines work.

ODE model

$$
\begin{aligned}
\dot{X} &= \lambda - d\,X - \beta\,[1 - f\,u(t)]\,XV \\
\dot{Y} &= \beta\,[1 - f\,u(t)]\,XV - a\,Y - p\,YZ \\
\dot{W} &= c\,XYW - cq\,YW - bW \\
\dot{Z} &= cq\,YW - h\,Z \\
\dot{V} &= k\,Y - \mu\,V
\end{aligned}
$$

Compartments

$$
\begin{aligned}
X &= \text{Uninfected T helper cells} \\
Y &= \text{Infected T helper cells} \\
W &= \text{Immune Precursor Cytotoxic T Lymphocyte} \\
Z &= \text{Immune Effector Cytotoxic T Lymphocyte} \\
V &= \text{Free Virus}
\end{aligned}
$$

Parameters

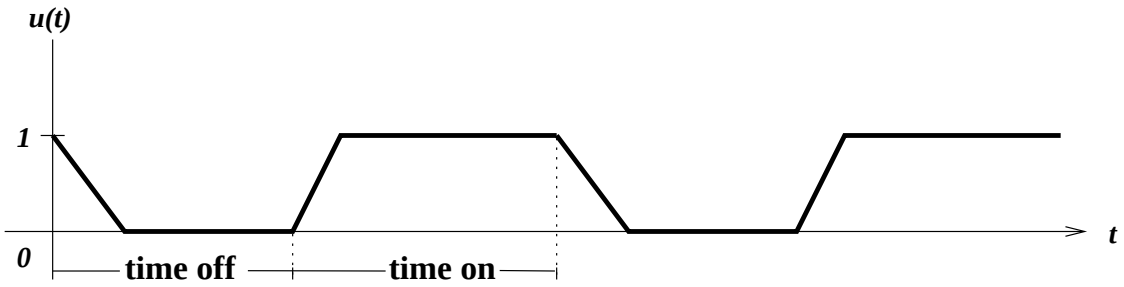
parameter	definition
λ	T helper target cell source from thymus
d	net natural growth/death of T helper target cells
β	infectivity of cells, due to XV
f	treatment efficacy factor ($0 \leq f \leq 1$)
a	net natural growth/death of infected cells
p	death rate of infected cells due to immune response (YZ)
c	growth of CTL precurors out of XYW
b	natural death of CTL precursors
$c\,q$	CTLe growth due to infected cells/CTLp (YW)
h	natural death of CTL effectors
k	growth rate of virions due to infected cells
μ	net natural growth/death of virions

- We ignore diffusion (hence we have a system of ODE's)

- There are two stable equilibria: one where the virus dominates and one where the immune system dominates. Set $\dot{z} = 0$ and solve for z giving $z = z_{equil}$. Next, find the e-vals of the Jacobian evaluated at z_{equil}. $re(e - vals) < 0 \implies z_{equil}$ is stable.

- The drugs we are trying to model are *Reverse Transcriptase inhibitors* which prevent viral RNA from translating to the DNA of the cell. They are not *protein inhibitors* which prevent protein envelopes (bulgees which house the attacking virus) from forming.

STI function $u(t)$



$u(t) = 0 \implies$ no treatment

$u(t) = 1 \implies$ full treatment

Sensitivity Analysis

Let $\mathbf{z} = [X, Y, W, Z, V]$ and
$\mathbf{q} = [\lambda, d, \beta, a, p, c, q, b, h, k, \mu, f]$. To get the *sensitivity matrix* $\frac{\partial z}{\partial q}$, we rewrite the model as

$$\begin{aligned} \dot{z} &= f(z(t); q) \\ z(0) &= z_0 \end{aligned}$$

Differentiating with respect to q and formally passing the time derivative through yields

$$\left(\frac{\partial \dot{z}}{\partial q}\right)(t) = \frac{\partial f}{\partial z}(z(t,q); q) \cdot \frac{\partial z}{\partial q} + \frac{\partial f}{\partial q}(z(t,q); q).$$

For $q = q_0$, this can be written as a 5×12 matrix system of ODEs for the sensitivity matrix $r(t) = \frac{\partial z}{\partial q}$

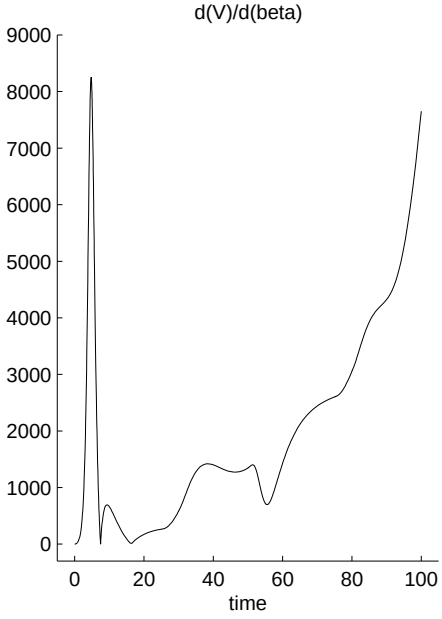
$$\begin{aligned} \dot{r}(t) &= A_0(t)\, r(t) + g_0(t) \\ r(0) &= 0. \end{aligned}$$

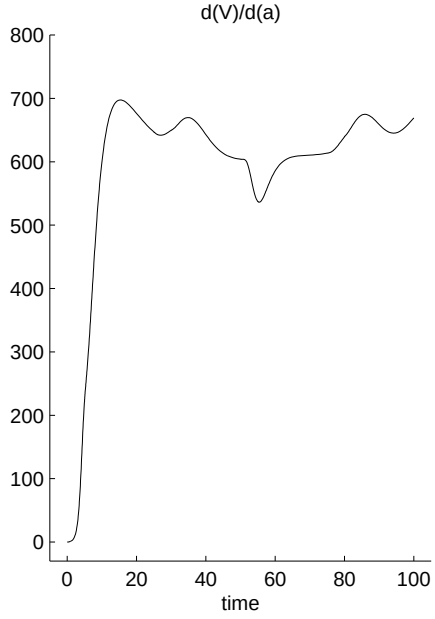
This yields the local system sensitivity about the point $q_0 \in \Re^m$.

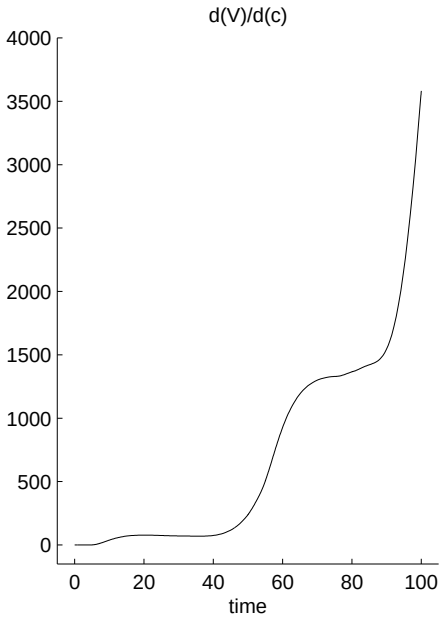
Given z_0 and q_0, we do an ODE solve (using MATLAB) to find a solution $z(t, q_0)$ over some time grid t. Since we have analytical expressions for $\frac{\partial f}{\partial z}(z(t,q);q)$ and $\frac{\partial f}{\partial q}$ then we know $A_0(t)$ and $g_0(t)$ (which depend on t through $z(t)$). Thus, we can do another (linear) ODE solve to get $r(t) = \frac{\partial z}{\partial q}$.

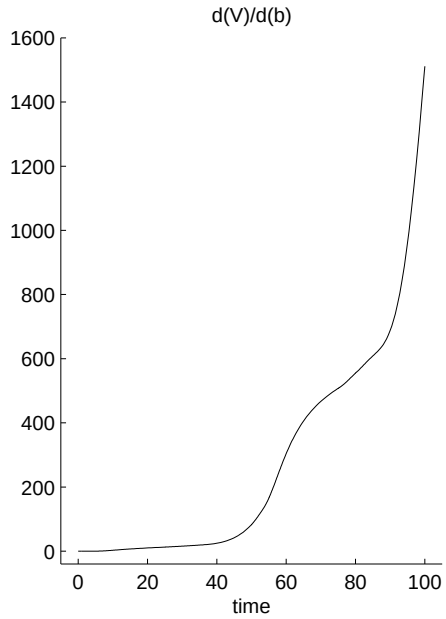
When should data from compartments be collected?

Use $\frac{\partial z}{\partial q}|_{q=q_0}$ to determine the sensitivity of the viral load over 100 days of a periodic STI:









- $\frac{\partial z}{\partial q}$ is a 3-D matrix (for each point in time t, we have a 5 x 12 matrix).

- After viewing similar graphs for each of the partials of X, Y, W, Z, V with respect to $\lambda, d, \beta, a, p, c, q, b, h, k, \mu, f$ we recommended that compartments be measured once at 5 days, again at 30 days, then once a week thereafter

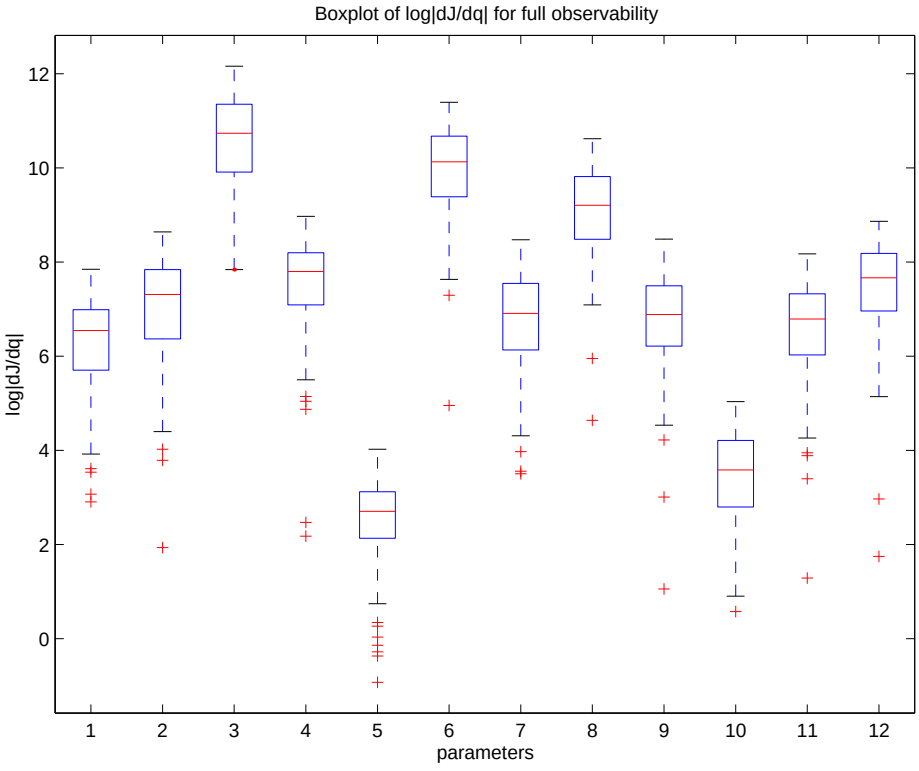
Which parameters most affect the model?

To measure how well values predicted by the model, $z(t_i)$, fit real data, $\hat{z}_i$, we use the least squares cost function

$$J(q) = \frac{\sum_i |\log(C \cdot \mathbf{z}(t_i, \mathbf{q})) - \log(C \cdot \hat{\mathbf{z}}_\mathbf{i})|^2}{\sigma_i^2}$$

$$\Longrightarrow \frac{\partial J}{\partial \mathbf{q}}(\mathbf{q}) = \sum_i 2 \frac{\log(C \cdot \mathbf{z}(t_i, \mathbf{q})) - \log(C \cdot \hat{\mathbf{z}}_\mathbf{i})}{\sigma_i^2} \cdot \frac{C}{C \cdot \mathbf{z}(t_i, \mathbf{q})} \left(\frac{\partial \mathbf{z}}{\partial \mathbf{q}}(t_i, \mathbf{q}) \right)$$

$\frac{\partial \mathbf{z}}{\partial \mathbf{q}}$ is the Sensitivity Matrix, C is the observation matrix, and σ^2 is the measurement error. Therefore, given data $\{\hat{z}_i\}$, $\frac{\partial J}{\partial \mathbf{q}}$ quantifies how sensitive our model is to small changes in the parameters.



- The boxplot was generated over 100 synthetic data sets.

- No real data was available. We had to generate data synthetically by

$$\log \hat{\mathbf{z}}_\mathbf{i} = \log \mathbf{z}_\mathbf{i} + \sigma\epsilon_i,$$

where $\epsilon_i = \epsilon(t_i) \sim N(0,1)$ and we assumed the vector of measurement errors were $\sigma^2 = [.01.01.01.01.25] \cdot \lambda$, for $\lambda \geq 1$. That is, σ^2 is the error incurred when a clinician actually measures each compartment.

- J can be used to answer the question: "*Given a data set, does our model describe the data set?*" The answer is **yes** if there is a value of q that makes J "small".

- J is a noisy function. Each function evaluation requires two ODE solves.

- dJ/dq can be used to answer the question "*Given a data set, how do small changes in the parameters affect the model?*" If $dJ/dq(i)$ is "larger" than $dJ/dq(j)$ for some i not equal to j, then the answer is **yes**, $q(i)$ affects the model a lot more than $q(j)$, and hence $q(i)$ is a parameter that we want to make sure that we estimate correctly.

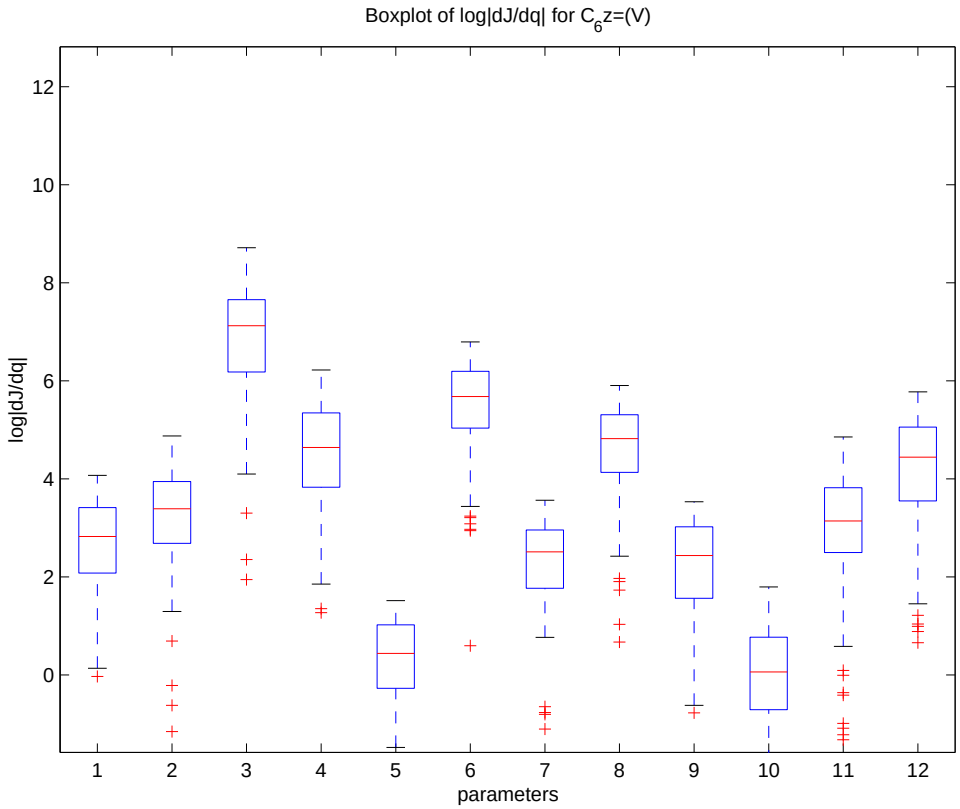
- The boxplot shows that the five parameters that affect the model the most are:

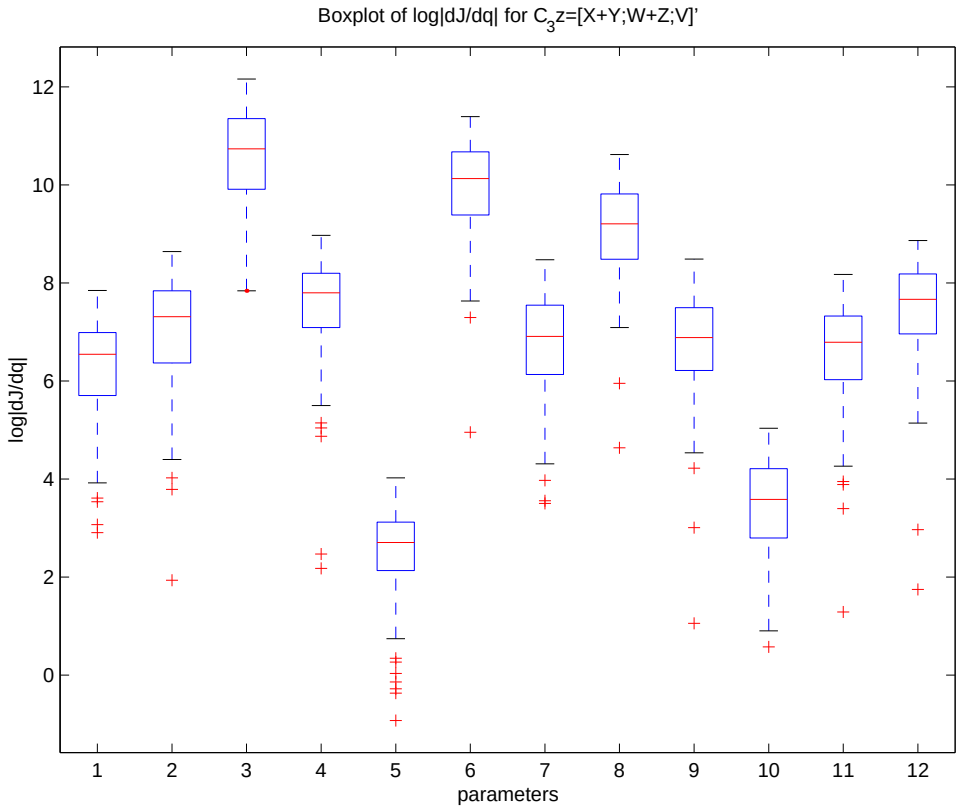
  β = proliferation rate of Infected T helper cells,
  c = proliferation rate of Immune Precursors CTL,
  b = natural death rate of Immune Precursors CTL,
  a = natural death rate of Infected T helper cells, and
  f = drug efficacy.

How should data from the compartments be collected?

The immunologist measures data in different linear combinations of X, Y, W, Z and V. We compute $\frac{\partial J}{\partial \mathbf{q}}$ with observation matrices that correspond to many of these realistic scenerios:

$$
\begin{aligned}
C_1\,\mathbf{z} &= \begin{pmatrix} X \\ Y \\ W \\ Z \\ V \end{pmatrix} \\
C_2\,\mathbf{z} &= \begin{pmatrix} X \\ Y \\ W+Z \\ V \end{pmatrix} \\
C_3\,\mathbf{z} &= \begin{pmatrix} X+Y \\ W+Z \\ V \end{pmatrix} \\
C_4\,\mathbf{z} &= \begin{pmatrix} X+Y \\ W+Z \end{pmatrix} \\
C_5\,\mathbf{z} &= \begin{pmatrix} X \\ Y \\ V \end{pmatrix} \\
C_6\,\mathbf{z} &= \begin{pmatrix} V \end{pmatrix}
\end{aligned}
$$


Boxplot of log|dJ/dq| for C6z=(V)
log|dJ/dq|
parameters
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12


Boxplot of log|dJ/dq| for C3z=[X+Y;W+Z;V]'
log|dJ/dq|
parameters
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12

- Boxplot from 100 synthetic data sets.

- We are assuming that the errors are independent

Inverse Problem

- *Is our model a good model?* That is, *Given a data set, does our model describe the data set?* The answer is **yes** if for

$$q* = \text{argmin}_{q \in Q_{ad}} J(q) = \frac{\sum_i |\log(C \cdot \mathbf{z}(t_i, \mathbf{q})) - \log(C \cdot \hat{\mathbf{z}}_\mathbf{i})|^2}{\sigma_i^2}$$

  $J(q^*)$ is "small".

- *What are good estimates for the parameters?* Simulate data by

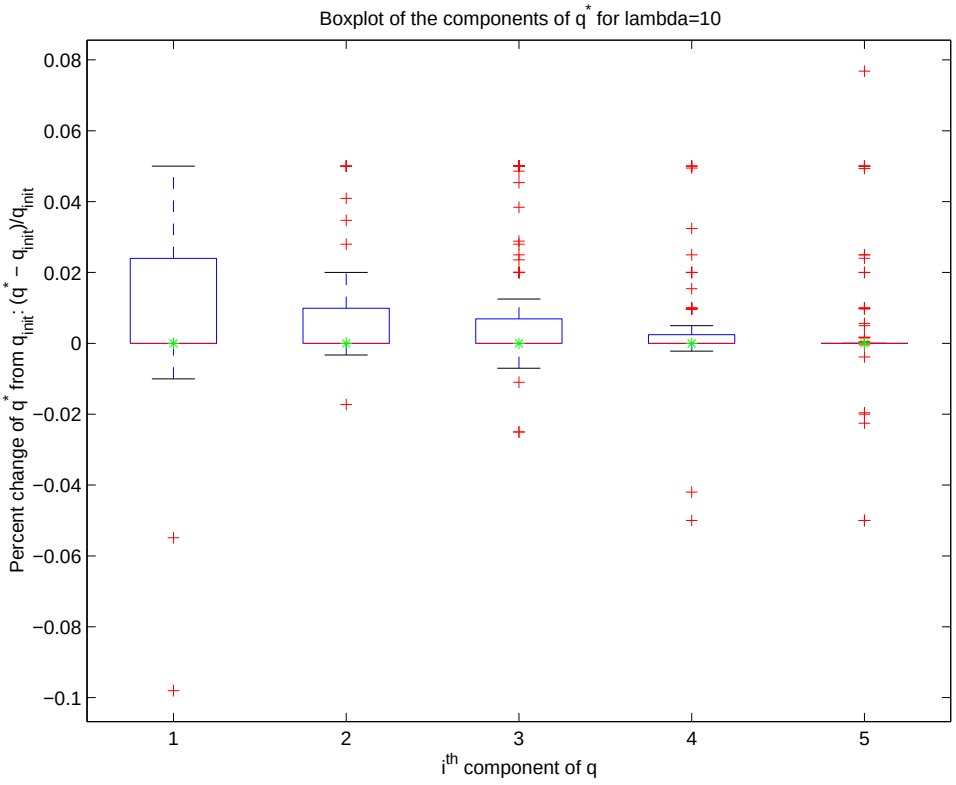
$$\log \hat{\mathbf{z}}_\mathbf{i} = \log \mathbf{z}_\mathbf{i} + \sigma \epsilon_i,$$

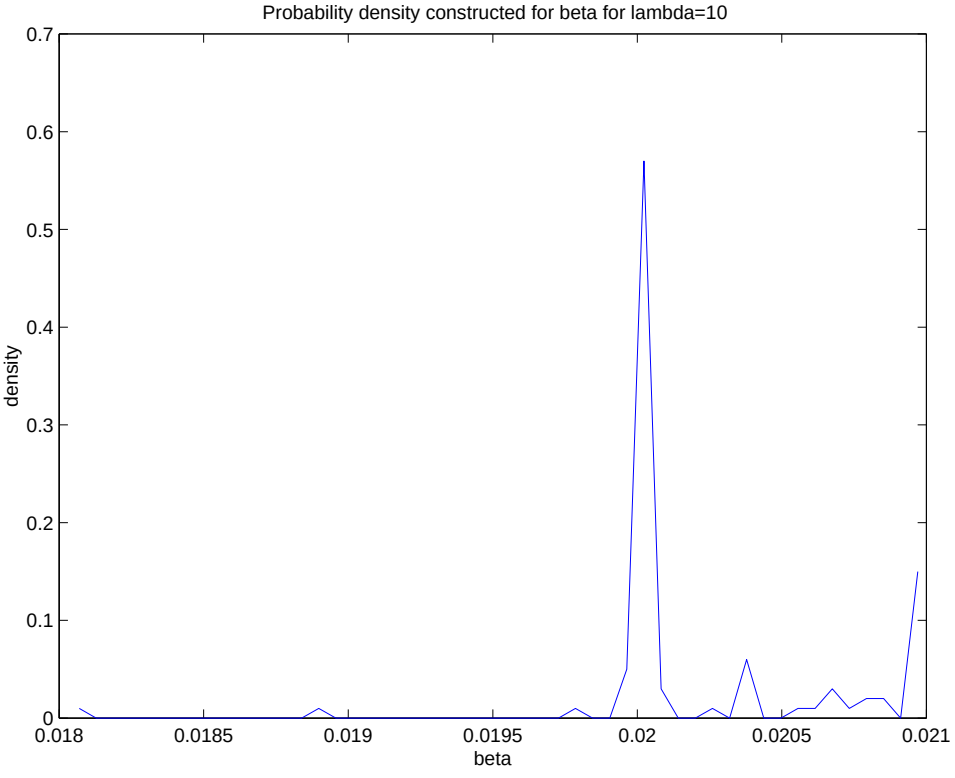
  where $\epsilon_i = \epsilon(t_i) \sim N(0,1)$, $\sigma^2 =$ [.01 .01 .01 .01 .25]$\cdot\lambda$ for $\lambda \geq 1$. Find q^* over many data sets to construct a probability density for each parameter.

- Found q^* using Nelder-Meade simplex method.

- Gradient type methods did not work very well. This is not surprising since J is a noisy function (each function evaluation requires two ODE solves). One could find the vectors corresponding to zeros (or values close to zero) in the SVD decomposition of $\frac{\partial z}{\partial q}$ which would indicate which linear combinations of the parameters are difficult to resolve.

- Generating pdf's in the manner described assumes independence of the parameters (the components of q).


Boxplot of the components of q* for lambda=10
Percent change of q* from q_init: (q* − q_init)/q_init
0.08
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
1
2
3
4
5
i^th component of q


Probability density constructed for beta for lambda=10
density
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.018
0.0185
0.019
0.0195
0.02
0.0205
0.021
beta

Conclusions

- When to collect data: on days 5, 30, and then every week thereafter
- How to collect data: measure $X+Y$, $W+Z$, and V
- Which parameters most affect the model: β, c, a, b, and f
- What are good estimates for the parameters: consult probability densities
- We solved the inverse problem for synthetic data
- Next: Solve inverse problem with real data