Modeling & Simulation
(Computational Immunology)

Steven H. Kleinstein

Departments of Pathology and Immunobiology
Yale University School of Medicine

steven.kleinstein@yale.edu

March 30, 2014
Simulated Experiment

Demonstrate full cycle of fitting model to data to estimate parameters

How can we estimate flow/proliferation/death rates?

Parameters used to create synthetic data

s = 0.003 per hour
p = 0.01 per hour
d = p + s (to achieve steady state)

Random noise added to each data point

BrdU withdrawn

0 20 40 60 80

Time (hours)

0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4

Fraction Labeled

A) Before BrdU administration
B) During BrdU administration
C) After BrdU administration

proliferation

source

d

d

d

T

L

U

L

U

p

p

p
How much confidence to put in estimate?

Construct confidence intervals for model parameters

Parameter estimates

s = 0.002 per hour
p = 0.01 per hour
Accuracy of Estimated Model Parameters

Underlying true set of model parameters ($\mathbf{a}_{\text{true}}$) known to Mother Nature but hidden from the experimenter

- True parameters are statistically realized as measured data set $\mathcal{D}_{(o)}$
- Fitting $\mathcal{D}_{(o)}$ yields estimated model parameters $\mathbf{a}_{(o)}$
- Other experiments could have resulted in data sets $\mathcal{D}_{(1)}$, $\mathcal{D}_{(2)}$, etc. which would have yielded model parameters $\mathbf{a}_{(1)}$, $\mathbf{a}_{(2)}$, etc.

Estimate probability distribution of $\mathbf{a}_{(i)} - \mathbf{a}_{\text{true}}$ without knowing $\mathbf{a}_{\text{true}}$
Monte Carlo Simulation of Synthetic Data Sets

- Assume that if $a^{(0)}$ is a reasonable estimate of $a_{\text{true}}$, then the distribution of $a^{(j)}-a^{(0)}$ should be similar to that of $a^{(j)}-a_{\text{true}}$.

- With the assumed $a^{(0)}$, and some understanding of the characteristics of the measurement noise, we can generate “synthetic data sets” $D^{S}_{(1)}, D^{S}_{(2)}, \ldots$ at the same $x_i$ values as the actual data set, $D_{(0)}$, that have the same relationship to $a^{(0)}$ as $D_{(0)}$ has to $a_{\text{true}}$.

- For each $D^{S}_{(j)}$, perform a model fit to obtain corresponding $a^{S}_{(j)}$, yielding one point $a^{S}_{(j)}-a^{(0)}$ for simulating the desired M-dimensional probability distribution. This is a very powerful technique!!

(Costa, Kleinstein and Hershberg, Sci Signal. 2011)
The Bootstrap Method

Estimating generalization error based on “resampling”: Randomly draw datasets with replacement from training data

- If don’t know enough about the measurement errors (i.e. cannot even say they are normally distributed) so Monte Carlo simulation cannot be used.
- Bootstrap Method uses actual data set $D_{(o)}$, with its N data points, to generate synthetic data sets $D^S_{(1)}$, $D^S_{(2)}$, … also with N data points.
- Randomly select N data points from $D_{(o)}$ with replacement, which makes $D^S_{(j)}$ differ from $D_{(o)}$ with a fraction of the original points replaced by duplicated original points.
- Fitting the $D^S_{(j)}$ data yields model parameter sets $a^S_{(j)}$ using actual measurement noise.

If sample is good approximation of population, bootstrap method will provide good approximation of sampling distribution of original statistic.
Bootstrap Methods

Randomly draw datasets with replacement from training data

- D = [3.0, 2.8, 3.7, 3.4, 3.5] → average = 3.28
- Bootstrap samples $D_N$ could be:
  - [2.8, 3.4, 3.7, 3.4, 3.5] → 3.36
  - [3.5, 3.0, 3.4, 2.8, 3.7] → 3.28
  - [3.5, 3.5, 3.4, 3.0, 2.8] → 3.24
  - ...

If sample is good approximation of population, bootstrap method will provide good approximation of sampling distribution of original statistic.
Bootstrapping Parameter Confidence Intervals

1) Fit model to data to obtain parameter estimates
2) Draw a bootstrap sample of the residuals (Fixed-X Bootstrapping)
3) Create bootstrap sample of observations by adding randomly sampled residual to predicted value of each observation

Repeat 1000x

 Estimate parameters for bootstrap samples

Bootstrapping observations also possible – asymptotically equivalent
Bootstrapping Parameter Confidence Intervals


Percentile Intervals

Calculate the parameter for each bootstrap sample and select \( \alpha \) (e.g., 0.05)

\[
\text{LCL} = \alpha / 2^{th} \text{ percentile.}
\]

\[
\text{UCL} = (1-\alpha/2)^{th} \text{ percentile.}
\]

Use MATLAB’s prctile function:

\[
\text{UCL} = \text{prctile}(\text{bootstrap estimates}, 0.025)
\]

Parameter estimates for synthetic data

Estimate of \( s = 0.0017 \) [0.0009, 0.0030]

Estimate of \( p = 0.0099 \) [0.0095, 0.0100]

May not have correct coverage when sampling distribution skewed
Practical reference for these kinds of methods

Numerical Recipes:
Includes source code for integration, optimization, etc.

Free NR versions online at http://www.nr.com/oldverswitcher.html
Hepatitis C Viral Dynamics and Interferon-α Therapy

Modeling 23 patients during 14 days of therapy (daily doses)

Viral loads exhibit short delay followed by biphasic decline in viral load

How does interferon therapy work?
Model of Hepatitis C Viral Dynamics

Includes virus along with target (T) and infected (I) cells

\[ \frac{dT}{dt} = s - dT \]  \hspace{1cm} (1)

\[ \frac{dI}{dt} = \ ? - \delta I \]  \hspace{1cm} (2)

\[ \frac{dV}{dt} = pI - cV \]  \hspace{1cm} (3)

Before therapy, virus load is approximately constant
Model of Interferon-α Therapy

Includes virus along with target (T) and infected (I) cells

Target Cells
\[ \frac{dT}{dt} = s - dT - \beta VT \]  (1)

Infected Cells
\[ \frac{dI}{dt} = \beta VT - \delta I \]  (2)

Virus (HCV RNA)
\[ \frac{dV}{dt} = pI - cV \]  (3)

Therapy can reduce the rate of infection, or production of virions
Hepatitis C Viral Dynamics and Interferon-α Therapy

Modeling 23 patients during 14 days of therapy (daily doses)

<table>
<thead>
<tr>
<th>Regimen</th>
<th>Patient</th>
<th>Initial VL (10^6 copies per milliliter)</th>
<th>Delay (hours)</th>
<th>Virion clearance (c) (1/day) ± error</th>
<th>Efficacy (e) Percent ± error</th>
<th>Infected cell death (δ) (1/day) ± error</th>
<th>Production (10^6 copies per day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>5.6</td>
<td>8</td>
<td>5.9 ± 1.1</td>
<td>79 ± 4.0%</td>
<td>0 ± 0.01</td>
<td>495 ± 146</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>1.9</td>
<td>8</td>
<td>6.4 ± 1.8</td>
<td>75 ± 7.0%</td>
<td>0.12 ± 0.02</td>
<td>290 ± 72</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>14.2</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>1</td>
<td>D</td>
<td>7.1</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td>1.1</td>
<td>11</td>
<td>7.0 ± 0.6</td>
<td>86 ± 0.1%</td>
<td>0.32 ± 0.04</td>
<td>125 ± 46</td>
</tr>
<tr>
<td>1</td>
<td>F</td>
<td>6.5</td>
<td>7</td>
<td>5.0 ± 0.8</td>
<td>89 ± 8.0%</td>
<td>0 ± 0.01</td>
<td>601 ± 15</td>
</tr>
<tr>
<td>1</td>
<td>G</td>
<td>3.3</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>1</td>
<td>H</td>
<td>4.1</td>
<td>10</td>
<td>6.9 ± 0.2</td>
<td>75 ± 1.0%</td>
<td>0 ± 0.01</td>
<td>498 ± 14</td>
</tr>
<tr>
<td>1: Mean</td>
<td>± SD</td>
<td>5.5 ± 4.1</td>
<td>9 ± 1.5</td>
<td>6.2 ± 0.8</td>
<td>81 ± 8%</td>
<td>0.09 ± 0.14</td>
<td>402 ± 191</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>6.1</td>
<td>7</td>
<td>3.6 ± 0.2</td>
<td>86 ± 0.5%</td>
<td>0.12 ± 0.01</td>
<td>410 ± 144</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>16.7</td>
<td>9</td>
<td>6.0 ± 0.3</td>
<td>98 ± 0.4%</td>
<td>RB</td>
<td>1409 ± 44</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>8.6</td>
<td>8</td>
<td>6.8 ± 0.8</td>
<td>96 ± 1.0%</td>
<td>0.11 ± 0.03</td>
<td>1089 ± 33</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>1.0</td>
<td>7</td>
<td>5.6 ± 0.5</td>
<td>95 ± 1.0%</td>
<td>0.16 ± 0.04</td>
<td>92 ± 18</td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td>59.0</td>
<td>10</td>
<td>11.2 ± 0.6</td>
<td>99.7 ± 0.01%</td>
<td>0.07 ± 0.02</td>
<td>1219 ± 41</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>10.9</td>
<td>7</td>
<td>4.4 ± 0.1</td>
<td>96 ± 0.9%</td>
<td>0.04 ± 0.01</td>
<td>965 ± 29</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>23.8</td>
<td>7</td>
<td>4.8 ± 0.1</td>
<td>92 ± 0.8%</td>
<td>RB</td>
<td>1780 ± 52</td>
</tr>
<tr>
<td>2: Mean</td>
<td>± SD</td>
<td>16.1 ± 18.9</td>
<td>8 ± 1</td>
<td>6.3 ± 2.4</td>
<td>95 ± 4.6%</td>
<td>0.1 ± 0.05</td>
<td>2282 ± 4045</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>6.7</td>
<td>8</td>
<td>3.7 ± 0.3</td>
<td>99.7 ± 0.4%</td>
<td>0.12 ± 0.04</td>
<td>405 ± 144</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>4.1</td>
<td>11</td>
<td>9.5 ± 3.7</td>
<td>91 ± 2.0%</td>
<td>0.11 ± 0.03</td>
<td>761 ± 29</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>5.8</td>
<td>13</td>
<td>5.7 ± 0.7</td>
<td>98 ± 0.5%</td>
<td>ND</td>
<td>523 ± 18</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>0.4</td>
<td>5</td>
<td>6.0 ± 0.8</td>
<td>99.0 ± 0.2%</td>
<td>0.4 ± 0.05</td>
<td>42 ± 13</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>18.3</td>
<td>7</td>
<td>6.0 ± 0.9</td>
<td>97.5 ± 1.6%</td>
<td>RB</td>
<td>2136 ± 52</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>1.1</td>
<td>14</td>
<td>5.8 ± 0.6</td>
<td>90 ± 0.3%</td>
<td>0.33 ± 0.03</td>
<td>112 ± 29</td>
</tr>
<tr>
<td>3: Mean</td>
<td>± SD</td>
<td>6.0 ± 5.9</td>
<td>9.5 ± 3.5</td>
<td>6.1 ± 1.9</td>
<td>96 ± 4.9%</td>
<td>0.24 ± 0.15</td>
<td>663 ± 769</td>
</tr>
<tr>
<td>All: Mean</td>
<td>± SD</td>
<td>9.4 ± 12.4</td>
<td>8.7 ± 2.3</td>
<td>6.2 ± 1.8</td>
<td>—</td>
<td>0.14 ± 0.13</td>
<td>1276 ± 498</td>
</tr>
</tbody>
</table>

Average virion production rate of $1.3 \times 10^{12}$ virions per day
Hepatitis C Viral Dynamics and Interferon-α Therapy

Modeling 23 patients during 14 days of therapy (daily doses)

Suggests immune control has important role in lowering viral load

Patients with undetectable HCV after 3 months of therapy (filled symbols) had significantly faster cell death rates
Major impact on understanding HIV/AIDS

HIV-I protease inhibitor given to twenty infected patients in order to perturb the balance between virus production and clearance.

Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection


AIDS Diamond Research Center, NYU School of Medicine, 480 First Avenue, New York, New York 10016, USA
* Santa Fe Institute, Santa Fe, New Mexico 87501, USA
† Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
‡ Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

Treatment of infected patients with ABT-538, an inhibitor of the protease of human immunodeficiency virus type 1 (HIV-1), causes plasma HIV-1 levels to decrease exponentially (mean half-life, 2.1 ± 0.4 days) and CD4 lymphocyte counts to rise substantially. Minimum estimates of HIV-1 production and clearance and of CD4 lymphocyte turnover indicate that replication of HIV-1 in vivo is continuous and highly productive, driving the rapid turnover of CD4 lymphocytes.

Discussion

We believe our new kinetic data have important implications for HIV-1 therapy and pathogenesis. It is self evident that, with rapid turnover of HIV-1, generation of viral diversity and the attendant increased opportunities for viral escape from therapeutic agents are unavoidable sequelae. Treatment strategies, if they are to have a dramatic clinical impact, must therefore be initiated as early in the infection course as possible, perhaps even during seroconversion. The rapid turnover of HIV-1 in plasma also suggests that current protocols for monitoring the acute antiviral activity of novel compounds must be modified to focus on the first few days following drug initiation. Our interventional

Viral dynamics applied to a wide variety of systems
The SIR Model of Epidemics

Model for many infectious diseases including measles

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \mu I \\
\frac{dR}{dt} &= \mu I
\end{align*}
\]

Other versions allow recovered individual to be re-infected
The basic reproductive ratio: $R_0$

average number of secondary cases caused by an infectious individual in a totally susceptible population

$$R_0 = \frac{\beta}{\mu} \times S(0)$$

$R_0 < 1$: disease dies out
$R_0 > 1$: disease can invade

<table>
<thead>
<tr>
<th>Disease</th>
<th>$R_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>2 to 5</td>
</tr>
<tr>
<td>Smallpox</td>
<td>3 to 5</td>
</tr>
<tr>
<td>Measles</td>
<td>16 to 18</td>
</tr>
<tr>
<td>Malaria</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

$R_0$ indicates whether population at risk from disease
ODEs are deterministic

Predicts epidemic even with non-zero chance that disease dies out

Stochasticity → risk of disease extinction when number of cases is small, even if $R_0 > 1$.

Simulate using stochastic approach – Gillepsie Method
Simulating Stochastic Models

How can we generate random number stream(s)?

Be careful on computer clusters (streams can be correlated)
Pseudo-Random Number Generators (PRNGs)

Starting with the same seed will give you equivalent stream

**Uniform deviates: [0,1)**
Linear congruential generator

$$I_{j+1} = aI_j + c \pmod{m}$$

$I_0$ is the seed (common to use system clock)

$$I_{j+1} = 3I_j + 7 \pmod{10}$$

Produces: 6,5,2,3

**Period**: time before stream repeats itself
(maximum m)

Fast, but sequential calls can be correlated, so not used much

**Better approach**
Mersenne Twister
(period $2^{19937} - 1$)

Be careful on computer clusters (streams can be correlated)
Simulating from other distributions

**Transformation Method**: indefinite integral of $p(y)$ must be known and invertible

Transformation to generate exponential distribution (Poisson process)

$$\text{Exponential}(\alpha) = -\frac{1}{\alpha} \ln \left[ \text{Uniform}(0,1) \right]$$

Methods based on underlying ability to generate uniform distributions
Boolean Network Models

Qualitative approach
Can be useful where kinetic parameters are not sufficiently known

- A directed graph (network)
  - Nodes represent the elements of a system
  - Edges represent regulatory relationships between elements
- Nodes characterized by True/False state
  - Network with N nodes will have $2^N$ possible states
- As time passes, node state determined by the states of neighbors, through a rule called a transfer function
  - Eg, logical function using the operators NOT, AND, OR
  - Output of transfer function determines state of the node

Often matches biological intuition: eg, genes are on/off.
Boolean Network Models

Qualitative approach
Can be useful where kinetic parameters are not sufficiently known

Time is discrete, specifying instances in which the state of the nodes may change

Easy to model combinatorial regulatory relationships

(Thakar and Alberta, 2010)
Boolean Interaction Network of Immune Response

Can be useful where kinetic parameters are not sufficiently known

Future states of each node decided by transition rules using Boolean operators

Nodes encompass related functions

Cells with unique functions are incorporated as nodes

Bacteria expressing generic virulence factors

Cytokines with similar functions are grouped together

Thakar J., et.al. (2007) PloS CB
ODEs Neglect Spatial Structure

Several approaches to including spatial effects

- **Partial Differential Equations (PDEs)**
  - Allows quantities to vary over both space and time
  - Continuous and deterministic
- **Compartment Modeling**
  - Compartments assumed to be well-mixed
  - Elements present in each compartment tracked using ODEs.
  - ODEs incorporate coupling between compartments
- **Agent-Based Modeling (ABM)**
  - object-oriented, discrete-event, rule-based, stochastic
  - views system as an aggregation of components (agents) that follow intrinsic rules of behavior (agent-rules)

“Right” approach depends on question, and available data
Cellular Automata Models

A regular grid of cells, each in one of a finite number of states

A classic example is Conway's Game of Life based on the following rules of occupancy of 8 surrounding cells:

- **Birth:** A dead cell with exactly three live neighbors becomes a live cell (birth).
- **Survival:** A live cell with two or three live neighbors stays alive (survival).
- **Death:** In all other cases, a cell dies or remains dead (overcrowding or loneliness).

Gosper's Glider Gun

A new generation is created (advancing t by 1), according to some fixed rule (generally, a mathematical function) that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood.
Agent-based Models (ABMs): IMMSIM

Individual cells given unique properties: receptors and internal state

A computer model of cellular interactions in the immune system
Franco Celada and Philip E. Seiden
Immunology Today 56 Vol. 13 No. 2 1992

study immune receptor signal–based cellular behavior with a bit-string representation for receptor specificities

(Kohler et al, 2000)
Detailed spatial pattern formation
Realistic models of cell diffusion and response to chemokines

New intravital imaging techniques provide underlying data
## Range of Current Modeling Frameworks

Various types of computational models can be built.

<table>
<thead>
<tr>
<th>Modeling approach</th>
<th>Typical applications</th>
<th>Limitations</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual particle-based stochastic</td>
<td>Small subcellular signaling processes, aspects of bacterial biochemistry</td>
<td>Applies only to small systems (in terms of space and chemical complexity)</td>
<td>MCell (32), Smoldyn (314), ChemCell (315), GetBonNie (nonspatial) (49)</td>
</tr>
<tr>
<td>Particle number stochastic</td>
<td>Signaling processes with important stochastic aspects (due to small system size or high sensitivity)</td>
<td>Applies only to small systems (in terms of space and chemical complexity), has less detail than individual particle simulation</td>
<td>MesoRD (35), SmartCell (33), GetBonNie (nonspatial)</td>
</tr>
<tr>
<td>Concentration-based, spatial, nonstochastic</td>
<td>Cellular signaling processes with important spatial aspects</td>
<td>Provides either high spatial resolution or biochemical complexity, has no stochasticity</td>
<td>Virtual Cell (37), Simmune (36)</td>
</tr>
<tr>
<td>Concentration-based, nonspatial, nonstochastic</td>
<td>Cellular signaling processes without spatial aspects</td>
<td>Assumes global biochemical homogeneity in the simulated system</td>
<td>Copasi (46), E-cell (44), Cellware (45), Systems Biology Workbench (47), GetBonNie</td>
</tr>
</tbody>
</table>

(Germain et al, 2010)

Each method has advantages and limitations – no one right approach.
Interchange format for computer models

XML encoding: wide variety of models can be described

A software package can read in a model expressed in SBML and translate it into its own internal format for model analysis.

Still, most researchers develop models from scratch for every project.
For more information...

Getting Started in Computational Immunology

Steven H. Kleinstein*
Interdepartmental Program in Computational Biology and Bioinformatics, and Department of Pathology, Yale University School of Medicine, New Haven, Connecticut, United States of America

Biomedical Model Fitting and Error Analysis

Kevin D. Costa,1,* Steven H. Kleinstein,2,3 Uri Hershberg4

Feel free to email me with questions:
steven.kleinstein@yale.edu