Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study
Authors: Lekone and Finkenstädt

Jon Azose

April 30, 2013
Ebola outbreak in DRC in 1995 infected 316, had an 81% mortality rate.

There’s no cure, but measures to prevent the spread of the disease are very effective.

**Question**: How does preventative intervention influence the course of the epidemic?

Model the progression of the disease with a susceptible-exposed-infectious-recovered (SEIR) model.
The real data
What would complete data look like?

<table>
<thead>
<tr>
<th>Person</th>
<th>Time of exposure</th>
<th>Time of contagion</th>
<th>Time of removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>1.32</td>
<td>6.45</td>
<td>14.90</td>
</tr>
<tr>
<td>Person 2</td>
<td>Not exposed</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Person 3</td>
<td>Not exposed</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Person N</td>
<td>145.01</td>
<td>152.77</td>
<td>155.81</td>
</tr>
</tbody>
</table>

Complete data would be:

- linked
- continuous
- complete (i.e. have $B(t)$, $C(t)$, and $D(t)$ with no missing records.)
The plan for making inferences

- Define a model that gives a likelihood function for unlinked, discrete-time data given a set of parameters.
- Choose prior distributions for parameters.
- Alternate sampling from distributions for (missing data|parameters) and (parameters|data).
- Make inferences based on posterior distribution of parameters.
A model for the complete data

Assume a population of size $N$ with one initial infection. At the individual level:

- Time until $S \rightarrow E$ transitions is exponentially distributed with rate $\beta(t)I(t)/N$, where
  
  $$\beta(t) = \begin{cases} 
  \beta & \text{if } t < t_* \\
  \beta e^{-q(t-t_*)} & \text{if } t \geq t_*.
  \end{cases}$$

- Time until $E \rightarrow I$ transitions is exponentially distributed with rate $\rho$.
- Time until $I \rightarrow R$ transitions is exponentially distributed with rate $\gamma$. 
\( \beta \): base transmission rate per infected individual (before intervention)

\( q \): decay in rate of transmission after intervention

1/\( \varrho \): mean incubation period

1/\( \gamma \): mean infectious period

Put Gamma priors on \( \beta, q, \varrho, \) and \( \gamma \).
We don’t need linked or continuous data

We don’t need linked data:

- Exponential distribution is memoryless.
- Sufficient statistics look like dwell times summed over all individuals.

We don’t need continuous data. On a discrete time scale, we can use

\[ B(t) \sim \text{Bin} \left( S(t), \left( 1 - \exp \left( -\frac{\beta(t)}{N} I(t) \right) \right) \right) \]

\[ C(t) \sim \text{Bin} \left( E(t), (1 - e^{-\varrho}) \right) \]

\[ D(t) \sim \text{Bin} \left( I(t), (1 - e^{-\gamma}) \right). \]
We can impute $B(t)$, $C(t)$, and $D(t)$

How to sample from $B|C, D, \Theta$:

1. Start with some $B$ that is not impossible. (e.g. $B$ must be such that no new infections can occur if there are no infectious.)

2. Propose a configuration $B'$ by randomly picking days $t_+$ and $t_-$. Take $B'(t_+) = B(t_+) + 1$ and $B'(t_-) = B(t_-) - 1$.

3. Accept the move to the proposed $B'$ with probability

$$\min \left( \frac{\pi(B'|C, D, \Theta) \ p(B' \rightarrow B)}{\pi(B|C, D, \Theta) \ p(B \rightarrow B')} , 1 \right)$$

4. Repeat steps 2 and 3.

The same procedure will work on the missing observations from $C(t)$ and $D(t)$. 
Example imputation
Imputation of $B$ induces a negative correlation between posterior draws of $\varrho$ (rate of $E \rightarrow I$ transition) and $q$ (rate of transmission decay).
Scientific contribution: How does intervention impact the spread of Ebola?

Statistical contribution: How can we make inference in an SEIR model when the $S \rightarrow E$ transitions are entirely unobserved?

Problem (for me): Markov chain convergence.