STAT 517: Stochastic Modeling

Lecture 1:
Introduction to the Stochastic Theory of Epidemics

January 9, 2008
Syllabus

STATISTICS 517: Stochastic Modelling (4 credits) Winter Quarter - 2008

Instructor: I. Longini

Meeting time and place: Tuesdays and Thursdays 1:30 P.M. - 3:20 P.M., Room A114, Physics & Astronomy Auditorium

Office hours: By appointment

Office location: Department of Biostatistics, Room F-653, Health Sciences Building (In the F – Wing http://www.washington.edu/admin/ada/MagMap.FH7.jpg )

Objectives: The student will learn the stochastic theory of epidemics and infectious disease transmission. This will include the theory of random phenomena that is concerned with the flow of events in time and space, especially those exhibiting highly variable behavior that can be described by probability distributions. Specifically, the student will learn to deal with the branching process, random walks, martingales, Markov processes, Poisson process, birth and death processes as applied to epidemic theory. There will be an emphasis on learning methods of strong scientific importance as opposed to purely mathematical theory.

Course Website: http://www.stat.washington.edu/courses/stat517/winter08/

Lecture Notes:

Required text:

Reference texts:
Bailey, N.T.J. (1964). The Elements of Stochastic Processes with Applications to the Natural Sciences, Wiley, N.Y.

Prerequisites: Statistics 516 or suitable course in stochastic processes that includes inference
Additional helpful background: Real Analysis, Differential Equations (including nonlinear theory), Survival analysis
Evaluation: Homework  10%
               Midterm  30%
               Final    60%
<table>
<thead>
<tr>
<th>#§</th>
<th>Date</th>
<th>Topic</th>
<th>Reference in Chiang *&lt;br&gt;(Lecture Notes **)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 8</td>
<td>Introduction to the Stochastic Theory of Epidemics</td>
<td>Readings</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>Branching Processes</td>
<td>3.1-3.4, 4.1</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>Epidemics as Branching Processes</td>
<td>(3.2-3.3)</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>Gambler's Ruin, Gambling Systems, Discrete-time Martingales</td>
<td>4.3 (4.3)</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>Discrete State Space and Time Markov Processes, Deterministic Dynamic Processes</td>
<td>5.1-5.8</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>Chain Binomial Models, Reed-Frost Model of Epidemics</td>
<td>(5.5-5.5, 5.8)</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>Inference on Markov chains and chain binomial models</td>
<td>(5.3)</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
<td>Algebraic Treatment of Markov chains</td>
<td>6.1-6.3, 14.3</td>
</tr>
<tr>
<td>9</td>
<td>Feb 5</td>
<td>Continuous-Time Markov Processes</td>
<td>10.1-10.3</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>Stages of Disease Process, HIV progression</td>
<td>11.7 (5.7, 6.7)</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>Embedded Process and Semi-Markov Process, Inference on Continuous-Time Markov Processes</td>
<td>(6.5)</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>Mid-term exam (Jan 7 – 31 material)</td>
<td>In class, open book</td>
</tr>
<tr>
<td>13</td>
<td>21</td>
<td>Stochastic Models for Graphs</td>
<td>Pavel Krivitsky ***</td>
</tr>
<tr>
<td>14</td>
<td>26</td>
<td>Inference for Stochastic Epidemic Models – Efficacy Studies</td>
<td>Yang Yang ***</td>
</tr>
<tr>
<td>15</td>
<td>28</td>
<td>Inference for Stochastic Epidemic Models – Real-time Estimation for Outbreaks</td>
<td>Yang Yang ***</td>
</tr>
<tr>
<td>16</td>
<td>March 4</td>
<td>Hidden Markov Processes</td>
<td>(8)</td>
</tr>
<tr>
<td>17</td>
<td>6</td>
<td>Counting Processes and Continuous-Time Martingales</td>
<td>(7.1)</td>
</tr>
<tr>
<td>18</td>
<td>11</td>
<td>Inference Using Continuous-Time Martingales</td>
<td>(7.1-7.3)</td>
</tr>
<tr>
<td>19</td>
<td>13</td>
<td>Individual-level stochastic simulation models, agent based</td>
<td>Readings</td>
</tr>
</tbody>
</table>

§ Lecture number

* Material covered in Chiang and in Lecture Notes in many cases

** Material covered only in the Lecture Notes

*** Guest lecture
Stochastic Epidemic Theory

- 1700’s: French probabilists and gambling
  - de Moivre
  - James Bernoulli
  - Many others
  - *Games, Gods and Gambling* (1961) by Florence Nightingale David

- 1766, Daniel Bernoulli argues for small pox vaccination using statistical arguments
Abraham de Moivre  Daniel Bernoulli
Florence Nightingale David
Reed-Frost Model History

- P. D. En’ko (1889)
- L. Reed & W.H. Frost (1930)
- M. Greenwood (1931)
- H. Abbey (1952)
- L. Elveback, J.P. Fox, E. Ackerman (1960)
Reed-Frost Model

Lowell Reed  
1886 - 1966

Wade Hampton Frost  
1880–1938

Both Former Deans: Johns Hopkins School of Public Health
Helen Abbey
1915 - 2001

Eugene Ackerman
1920 -
This Could Happen!
Vaccination post-alert; 10M doses/week for 25 weeks of low-efficacy vaccine, $R_0 = 1.9$
Real Time Detection, Estimation and Control
Publications


Reproductive numbers

- Basic reproductive number, $R_0$:
  Average number of new infections that a typical infected person produces in a completely susceptible population
  - $R_0 > 1$ (or R) for sustained transmission

- $R_f$:
  Average reproductive number with fraction $f$ of the population vaccinated
The Problem

- Observed illness onset times
  - Exposure information
- Is the disease infectious?
- If so, what are estimates of transmission parameters?
  - Secondary attack rates
  - Reproductive number
- How effective are interventions?
- Calibration of intervention models.
Information Needed

• Data
  – Illness onset times
  – Crude exposure information
  – Other information
    • Treatment and prophylaxis
    • Hospitalization and death
    • Infection information
    • Covariates
    • Illness serial interval distribution

• Natural History
  – Incubation and infectious periods
• Consider a community composed of households, and three channels through which infection may occur.
  
  – **Close contact within households:** the probability that a susceptible is infected by an infective in the same household in one day is $p_1$.
  
  – **Casual contact within community:** the probability that a susceptible is infected by an infective in the same community but different household in one day is $p_2$.
  
  – **Common source of infection** (e.g., zoonotic source or visiting infectives from outside of the community): the probability that a susceptible is infected by the common source in one day is $b$. 
Statistical Model
Hypothesis to be tested.

\[ H_0 : p_1 = p_2 = 0 \ vs. \]
\[ H_1 : p_1 > 0 \ or \ p_2 > 0, \]

Test statistic

\[ \lambda = -2 \log \left( \frac{\sup_b L_0(b|\tilde{t}_i, i = 1, \ldots, N)}{\sup_{b,p_1,p_2} L(b, p_1, p_2|\tilde{t}_i, i = 1, \ldots, N)} \right), \]

where \( \tilde{t}_i \) is the symptom onset time for person \( i \), and \( N \) is the population size.
- More assumptions:
  - Random mixing in households and in the community.
  - The latent period coincides with the incubation period.
  - Distributions of the latent period ($\delta$) and the infectious period ($\eta$) are known:
    * $\delta \sim g(l) = \Pr(\delta = l), l = \delta_{\text{min}}, \delta_{\text{min}} + 1, \ldots, \delta_{\text{max}}$.
    * $\eta \sim f(l) = \Pr(\eta \geq l), l = \eta_{\text{min}}, \eta_{\text{min}} + 1, \ldots, \eta_{\text{max}}$.
  - Observation starts from day 1 to day $T$, and exposure to the common source starts from day 1 to day $S \leq T$.
    * When $S < T - \delta_{\text{min}}$, asymptotic method does not work.
• The probability that an infective $j$ infects an susceptible $i$ on day $t$ is

$$p_{ji}(t) = p_1^{I(j \in H_i)} p_2^{I(j \notin H_i)} f(t - \tilde{t}_j),$$

where $H_i$ is the set of household members of person $i$.

• The probability that subject $i$ escapes infection from all infective sources on day $t$ is

$$e_i(t) = (1 - b)^{I(t \leq S)} \prod_{j=1}^{N} (1 - p_{ji}(t)).$$
• A likelihood for $b$, $p_1$ and $p_2$ contributed by person $i$ is

$$L_i(b, p_1, p_2 | \tilde{t}_j, j = 1, \ldots, N)$$

$$= \begin{cases} 
\prod_{t=1}^{T} e_i(t), & \text{not infected,} \\
\sum_t g(\tilde{t}_i - t)(1 - e_i(t)) \prod_{\tau=1}^{t} e_i(\tau), & \text{otherwise,}
\end{cases}$$

(3)

• When $p_1 = p_2 = 0$, (2) reduces to

$$e_i(t) = (1 - b)^{I(t \leq S)}.$$
Important parameters

- Household SAR\(_1\) = \(\sum_i f(l)(1 - (1 - p_1)^i)\)

- Community SAR\(_2\) = \(\sum_i f(l)(1 - (1 - p_2)^i)\)

- Local \(R_0 = (M - 1) \times SAR_1 + (N - M) \times SAR_2\)
H5N1 Influenza in Family Cluster in North Sumatra, May 2006
### Family Member Information

<table>
<thead>
<tr>
<th>Case</th>
<th>Relation</th>
<th>Age</th>
<th>PCR Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Case</td>
<td>Mother</td>
<td>35 yr</td>
<td>PCR-</td>
</tr>
<tr>
<td>Case 2</td>
<td>Son</td>
<td>15 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td>Case 3</td>
<td>Son</td>
<td>17 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Son</td>
<td>10 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G’mother</td>
<td>55 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Daughter</td>
<td>21 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fiancé</td>
<td>?? yr</td>
<td></td>
</tr>
<tr>
<td>Case 4</td>
<td>Mother</td>
<td>29 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Father</td>
<td>32 yr</td>
<td></td>
</tr>
<tr>
<td>Case 5</td>
<td>Daughter</td>
<td>18 mo</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Son</td>
<td>10 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Daughter</td>
<td>6 yr</td>
<td></td>
</tr>
<tr>
<td>Case 6</td>
<td>Father</td>
<td>32 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Mother</td>
<td>29 yr</td>
<td></td>
</tr>
<tr>
<td>Case 7</td>
<td>Son</td>
<td>10 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Son</td>
<td>6 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Index's Brother</td>
<td>?? yr</td>
<td></td>
</tr>
<tr>
<td>Case 8</td>
<td>Father</td>
<td>25 yr</td>
<td>PCR+</td>
</tr>
<tr>
<td></td>
<td>Mother</td>
<td>?? yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Son</td>
<td>3 yr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Son</td>
<td>5 mo</td>
<td></td>
</tr>
</tbody>
</table>

### Event Timeline
- **April 23**: Family Gathering – April 29
- **April 30**: Cared for Sick Son
- **May 7**: Cared for Index Case
- **May 14**: Exposed to Husband
- **May 21**: Frequently Visited House 1
Tests and estimates

• Statistical evidence of person-to-person transmission ($p = 0.009$)

• Household SAR: 29% (95% CI, 15-51%)

• Lower bound on the local $R_0$: 1.14 (95% CI, 0.61-2.14)
  – 88% chance of no further spread
TranStat

To analyze data from outbreaks of acute infectious disease, MIDAS scientists at the University of Washington and the Fred Hutchinson Cancer Research Center have developed TranStat, a tool for data entry, storage, and rapid analysis. It is being used to test for the presence of human-to-human transmission (or animal-to-animal transmission in veterinary settings) and to estimate the epidemiological characteristics of the disease, such as secondary attack rates and the local reproductive number.

**Purpose.** The key to controlling a pandemic is early detection, containment, and mitigation. The TranStat tool was developed to enable field personnel and researchers to enter and revise data from local outbreaks. From these data, TranStat provides a means of testing for the presence of human-to-human (or animal-to-animal) transmission. If this transmission is detected, estimates of the household-specific and neighborhood-specific secondary attack rates and local reproductive number are provided.

**Data Input.** TranStat uses information on the

- outbreak,
- population at risk,
- exposure events, and
- estimated incubation and infectious periods.

Outbreak details include the number of neighborhoods, households, cases, and noncases and the timing of the outbreak. If a community-wide intervention has occurred, the timing of the intervention is also specified.

Information on the population at risk includes sex, age, neighborhood and household of residence, dates of illness onset for cases; dates of hospitalization, if any; and dates of receiving treatments for cases or prophylaxis for noncases, if any.

Exposure details for each person include the neighborhoods and households visited and the dates of the visits.

The estimated distributions of the infectious or incubation periods are specified in terms of minimum and maximum number of days and probabilities for each particular duration (e.g., day, hour, week).

**Method.** A discrete-time maximum likelihood model is used to estimate the time-specific probabilities of transmission within and between households, from which the secondary attack rates are derived. In addition, the time-invariant infection probability from a common source is calculated.

To test for human-to-human transmission, the likelihood ratio is calculated, comparing the likelihood

https://www.epimodels.org/midas/about.do
assuming that there is no transmission (i.e., the within- and between-household transmission probabilities are zero) with the likelihood when no assumption is made. In TranStat, the assumption of no human-to-human transmission reduces to a model with transmission from common sources only. Because this transmission does not vary between individuals, a simple permutation method was developed to estimate the distribution of likelihood ratio statistics under the assumption of no human-to-human transmission. The infection and symptom status of the observed cases are repeatedly reassigned to different groups of subjects. The observed likelihood ratio statistic can then be compared with the generated distribution of likelihood ratio statistics to obtain the p-value.1

Results. TranStat results include estimates (and 95% confidence intervals, as appropriate) for

- the case fatality rate,
- the daily probability of infection by common source (such as a zoontic source),
- the within- and between-household daily transmission probabilities and secondary attack rates,
- local reproductive number, and
- the p-value for testing human-to-human transmission.

Availability. The TranStat tool can be downloaded from http://www.midasmodels.org. The version 0.1 Java code can be run on Windows platforms, and later versions can be run on any platform, including Linux.

For information about TranStat, send e-mail to yang@scharp.org for questions about the tool methods and dianis@jdrf.org for questions about operating TranStat.

For more information about MIDAS

MIDAS Web site: http://www.midasmodels.org

National Institute of General Medical Sciences (NIGMS): http://www.nigms.nih.gov/Initiatives/MIDAS

MIDAS Scientific Director at NIGMS: Irene Eckstrand, ieckstr@nigms.nih.gov


Acknowledgments and product names used are trademarks, registered trademarks, or trade names of their respective holders. Microsoft Windows is a registered trademark of Microsoft Corporation, Linux is a registered trademark of Linux Torvalds.

About MIDAS

Funded by the National Institute of General Medical Sciences (NIGMS), one of the National Institutes of Health, the MIDAS mission is to investigate computational and mathematical models of existing and emerging infectious diseases in order to prepare the nation to respond to disease outbreaks.

http://www.midasmodels.org
Control of transmission at the source
Publications


Dealing with pandemic flu

• Containment of pandemic influenza at the source is the best strategy
  – Mobile stockpiles antiviral agents
    • 5 million courses, WHO donated by Roche
  – Mobile stockpile vaccines
    • 50 million doses, WHO donated by GSK
Pre-pandemic Vaccines and Antiviral Agents
Vaccines and antivirals will slow the spread of a pandemic.
Measures of Vaccine Efficacy*

- \{\text{VE}(t)_S, \text{VE}(t)_P, \text{VE}(t)_I\}

- \text{VE}_{SP}(t) = 1 - (1 - \text{VE}_S(t)) \cdot (1 - \text{VE}_P(t))

- Phase III, usually only \text{VE}_{SP}
- Phase I and II vaccine trials
  - Correlates of immunity and transmission
- Virus challenge studies in humans or animals

What we conclude at this point about VE parameters for H5N1 pre-pandemic influenza vaccines for heterologous virus?
Vaccine efficacy kinetics model

- Vaccine efficacy build up exponentially
- Two doses of vaccines are given at week 0 and 3.
- For the first dose, it takes two weeks to build up its full protection ($k=0.05$); for the second one, it takes one week ($k=0.4$)
- Two weeks after the first dose, the vaccine efficacy is at 50% of its full strength of two doses

$$r_{ve}(t) = a \cdot e^{bt} + c$$

- $r_{ve}(0) = 0$
- $r_{ve}(1) = k$
- $r_{ve}(2) = 1$

$$a = \frac{1}{\ln \frac{1}{k} - 1}$$

$$b = \ln(\frac{1}{k} - 1)$$

$$c = -a$$

$0 < k < 1$
Vaccine efficacy buildup assumption

1st dose, prime

2nd dose, boost

Week
Large-Scale, Individual-based, Stochastic Simulation Models
Transmission Models

• The four key elements of our models
  – Disease natural history model and parameters
  – Community-level transmission between people, through various contact groups (household, work group, school)
  – Census demographics (where people live) and workerflow data (where they work), at tract-level resolution
  – Transportation statistics on long-distance travel
The higher $R_0$, the earlier and the higher the peak incidence of the pandemic
Pre-pandemic Vaccination Strategies

• Mass pre-vaccination
  – Two doses at least five weeks before initial case

• Reactive mass vaccination
  – Begin vaccinating x days after first case in a geographic region

• Ring vaccination
  – Begin vaccinating x days after first case in ring, then in a ring after each subsequent case
Antiviral efficacies used in the model: Oseltamivir

- Antiviral efficacy of reducing susceptibility to infection: \( \text{AVE}_S = 0.48, [0.17, 0.67] \) 95% CI*

- Antiviral efficacy of reducing illness given infection: \( \text{AVE}_P = 0.56, [0.10, 0.73] \) 95% CI*

- Antiviral efficacy of reducing illness with infection: \( \text{AVE}_{SD} = 0.80, [0.35, 0.94] \) 95% CI*
  - Mult.: \( \text{AVE}_{SP} = 1 - (1 - \text{AVE}_S)(1 - \text{AVE}_P) = 0.77 \)

- Antiviral efficacy of reducing infectiousness to others: \( \text{AVE}_I = 0.80, [0.45, 0.93] \) 95% CI*

TAP: Targeted antiviral prophylaxis using neuraminidase inhibitors (oseltamivir/zanamivir)
Containment at the source

- Stochastic, individual based simulations of Southeast Asian population of 500,000 individuals

- Transmission occurs in households, schools, workplaces, clusters of households, social places, and community

Longini et al. Science 2005; 309: 1083-1087
Containing pandemic influenza at the source: Goal of modeling

- Contain a reassorted or mutated strain of influenza at the source
- Avian A(H5N1) is a likely virus
- Source could be in SE Asia

Longini et al. Science 2005; 309: 1083-1087
Containing pandemic influenza at the source:
Practical example of modeling

Rural population of 500,000 in Thailand

Population matched to non-municipal area household-size and age distributions.*

*Population and Housing Census 2000 data used where available (www.nso.go.th); other National Statistical Office reports and tables used as necessary.
Population Characteristics

- 36 localities each of size ~14,000
- Total area: 75 km X 75 km = 5,625 km²
- Population density ~89/km²
Simulated pandemic influenza outbreak

$R_0 = 1.4$

Without intervention

80% TAP

Longini et al. Science 2005; 309: 1083-1087
Simulated pandemic influenza outbreak

$R_0 = 1.7$

80% TAP

80% TAP + 50% Pre-vacc

Longini et al. Science 2005; 309: 1083-1087
Simulated pandemic influenza outbreak

$R_0 = 1.7$

70% reactive vaccination

No intervention
Simulated pandemic influenza outbreak
\[ R_0 = 1.7 \]

70\% reactive vaccination

70\% reactive vaccination + 80\% TAP
70% reactive vaccination R0=1.2 intervention at day 18
70% reactive vaccination R0=1.7 intervention at day 18
Comparison of epidemic size for Reactive vaccination and baseline at $R_0=1.7$
Sensitive, unpublished material has been deleted after this point.
THE END