Modelling and surveillance of infectious diseases - or why there is an $R$ in SARS

Michael Höhle$^{1,2}$

$^1$Department of Statistics, Ludwig-Maximilians-Universität München, Germany

$^2$Centre for Mathematical Sciences, Technische Universität München München, Germany

useR!2008
Kaleidoscope II Session
Dortmund, 12 August 2008
Motivation

- How can R assist in understanding and controlling infectious diseases – be it in human, plant or veterinary epidemiology.
- Two R packages exist:
  1. **RLadyBug** contains a set of functions for the simulation and parameter estimation in spatially heterogeneous SIR models.
  2. **surveillance** contains algorithms for the detection of aberrations in time series of counts arising from routine public health surveillance.
- This talk intends to give an overview of using R for especially (1) – deeper mathematical details are suspended to the lunch break.

```
> library("RLadyBug")
```
Applications (1): SARS in Hong Kong 2003

- Daily number of new cases of the severe acute respiratory syndrome (SARS) in Hong Kong (Anonymous, 2003)
- Epidemic curve created with package epitools (Aragon, 2007).
Applications (2): CSF Transmission Experiment

- Experiment by Laevens et al. (1999) with classical swine fever (CSF) using $S(0) = (5, 5, 6)$ and $E = (0, 1, 0)$.
- Event history of each pig with inoculation as origin

> data("laevens")
> plot(laevens, type = individual ~ time / position)
Applications (3) – CSF surveillance

- Classical swine fever (CSF) in Brandenburg (BB) and Mecklenburg-Western Pomerania (MP), Germany
- Total of 81 infected farms out of 3290 during 1993-2004

Interest in investigating the connection between the CSF incidence among domestic pigs and wild boars
Applications (4) - Spatial incidence of CSF in MPBB

Domestic pigs in MPBB
- No case
- 1 case
- 2–5 cases
- More than 5 cases

Wild boars in MPBB
- No case
- 1 case
- 2–5 cases
- 6–20 cases
- More than 20 cases
Stochastic epidemic models (1)

- SEIR model: A closed population of $n + m$ individuals divided into susceptible, exposed, infectious, and recovered
- $S(0) = n$, $E(0) = m$, $I(0) = 0$ and $R(0) = 0$
- At time $t$, an individual $j$ meets infectious at rate

$$\lambda_j(t|\mathcal{H}_t) = \sum_{i=1}^{n+m} \mathbb{1}_{i \in \text{Infectious}(t)} \cdot f(i, j),$$

where $f(\cdot) \geq 0$ is a function of the distance between $i$ and $j$
- If a susceptible meets an infected, it becomes exposed
Stochastic epidemic models (2) – Distance kernels

1. Homogeneous model: \( \forall i, j : f(i, j) = \beta > 0 \) and hence

\[
\lambda_j(t|H_t) = \beta I(t)
\]

2. Heterogeneous model: The population is made up of \( k \) units arranged on a grid in space. For \( j \) in unit \( u_j \):

\[
\lambda_j(t|H_t) = \beta I_{u_j}(t) + \beta_\eta \sum_{u \in N(u_j)} I_u(t)
\]

3. Heterogeneous model, where individuals have locations in \( \mathbb{R}^2 \) and \( f(i, j) \) is a function on the Euclidean distance dist\((i, j)\)
SARS in Hong Kong 2003

- Assuming a constant incubation time of 6.4 days and a constant recovery time of 34 days as suggested by the mean of the distributions in Donelly et al. (2003) we obtain

\[
\text{Basic reproduction number } R_0 = R_0(m1, \text{hksars}) = 1.0012.
\]
Motivation

Methods

Results

Surveillance

Discussion

References

CSF Transmission Experiment (1)

- Exposure times are not observed, instead of imposing we assume $T_E \sim Ga(\delta_E, \gamma_E)$ and $T_I \sim Ga(\delta_I, \gamma_I)$
- A Bayesian setting with MCMC is used for parameter inference

```r
> print(m2 <- seir(laevens, laevens.opts.mcmc))
```

An object of class LBInferenceMCMC

Parameter Estimations (posterior mean from 2500 samples):

<table>
<thead>
<tr>
<th>Parameter:</th>
<th>beta</th>
<th>betaN</th>
<th>gammaE</th>
<th>deltaE</th>
<th>gammaI</th>
<th>deltaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.03706</td>
<td>0.02837</td>
<td>56.82000</td>
<td>9.37400</td>
<td>2.16200</td>
<td>0.25640</td>
</tr>
</tbody>
</table>

StandardErrors (posterior std.dev. from 2500 samples):

<table>
<thead>
<tr>
<th>Parameter:</th>
<th>beta</th>
<th>betaN</th>
<th>gammaE</th>
<th>deltaE</th>
<th>gammaI</th>
<th>deltaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.018500</td>
<td>0.009481</td>
<td>45.510000</td>
<td>7.761000</td>
<td>0.738100</td>
<td>0.097760</td>
</tr>
</tbody>
</table>
CSF Transmission Experiment (2)

- MCMC output can be further analysed by e.g. coda package
- Posterior density of $\beta/\beta_n$ and $R_0$

> `plot(m2, which = "betabetaN")`
> `quantile(R0(m2, laevens), c(0.025, 0.5, 0.975))`

\[
\begin{align*}
\beta_n & \sim N(0.01, 0.1) \\
\eta & \sim N(0, 0.1)
\end{align*}
\]
CSF surveillance (1)

- CSF surveillance data consists of multiple outbreaks
- SIR Extension: Risk of infection consists of two components
  - *endemic component*: Time to infection from external sources modelled by a Cox model
  - *epidemic component*: Similar to heterogeneous SIR model with distance weighting of infectives
- Rate of infection has the following form

\[
\lambda_j(t|H_t) = \exp\left( h_0(t) + z_j(t)'\alpha \right) + \sum_{i=1}^{n+m} \mathbb{1}_{i\in\text{Infectious}(t)} \cdot f(i,j)
\]

- When using a linear basis expansion of \( f(i,j) \) this rate is similar to the conditional intensity of an additive-multiplicative hazard model from survival analysis
CSF surveillance (2)

- **Endemic component**: piecewise exponential baseline and time varying covariates boars and vaccination area
- **Epidemic component**: $f(i,j) = \beta > 0$
- **Inference using penalized loglikelihood with a model syntax similar to the timereg package** (Scheike, 2006)

```r
> m3 <- spatialSIR(Surv(start, stop, event) ~ fconst +
+ cox(boar) + cox(vacc), data = mpbb.evHist, ...)
> coef(m3)[c("fconst", "cox(boar)", "cox(vacc)")]
> diag(vcov(m3))[c("fconst", "cox(boar)", "cox(vacc)")]
```

```
fconst  cox(boar)  cox(vacc)
3.814e-06  2.108e+00  1.261e+00
7.371e-12  9.263e-02  1.729e-01
```
Plot of the total intensity $\sum_{j=1}^{n+m} \lambda_j(t|H_t)$, the log baseline hazard $h_0(t)$ (with a 95% CI) and the epidemic proportion.
The surveillance package (1)

```r
> library("surveillance")
> data("shadar")
> control = list(range = 105:295, ret = "cases", alpha = 0,
+     c.ARL = 5)
> plot(algo.glrnb(shadar, control = control))
```

Analysis of shadar using glrpois: intercept
The surveillance package (2)

- Surveillance algorithms:
  - cdc – Stroup et al. (1989)
  - farrington – Farrington et al. (1996)
  - cusum – Rossi et al. (1999)

- Comparison of surveillance algorithms using sensitivity, specificity and its variants in simulations

- Surveillance time series models:
  - hhh - Held et al. (2005); Paul et al. (2008)
  - twins - Held et al. (2006) (Experimental)
First steps towards R functionality for infectious disease modelling. More complex and realistic models imaginable.

Packages contain many additional visualization and simulation procedures (Sellke construction, Ogata’s modified thinning)

Combining database, R, Sweave/odfWeave and LaTeX/OpenOffice allows for easy generation of daily bulletins or reports

> motd

Message of the day

Packages are on CRAN. Starting points are H. (2007); H. and Feldmann (2007). Maybe they are of help. If adaptation is needed for your problem let me know.
Acknowledgements

Persons:

- Ulrike Feldmann, Sebastian Meyer and Valentin Wimmer, Ludwig-Maximilians-Universität München, Germany
- Michaela Paul and Andrea Riebler, Institute of Social and Preventive Medicine, University of Zurich, Switzerland
- Christoph Staubach, Federal Research Institute for Animal Health (FLI), Germany

Financial Support:

- Munich Center of Health Sciences
Literature I


