Bayesian Methods in Syndromic Surveillance

Andrew B. Lawson
University of South Carolina

NYAM Syndromic Surveillance Meeting
October 2003

Structure of Talk

1)Issues in Syndromic Surveillance
2) Bayesian background
3) Bayesian alarm functions

Issues in Syndromic Surveillance

- - ***Surveillance*** methodologies are under-developed within statistical approaches to temporal, and in particular, spatial epidemiology.

  - Surveillance is based on the idea that *limits* are to be set on the incidence of disease within a temporal or spatial domain and actions taken if these limits are passed. These limits may be based on a single observed disease pattern or could consist of rules for intervention based on many ancillary variables.

- ***Example:*** outbreaks of short-latency intestinal disease could be related to pharmaceutical sales, and monitoring of the disease and sales may be needed

- ***Process monitoring***: statistical process control (SPC) has relevance to the issue of how to design disease surveillance systems.

Some Issues in Data Mining

- Health surveillance is now concerned with wider issues than single disease modelling
- Syndromic surveillance: early detection of adverse health events
  - inhalation insult reports; infectious disease spread; ER visits
  - Often ancillary information is available which can help make early detection possible
  - Pharmaceutical sales; job/school absenteeism; ER visits
  - Syndromes can consist of symptoms but need to be measurable/reported

- multiple times series of events or maps can be monitored together to allow early detection

Bayesian Background

- Methods are needed which can allow us to surveillance large quantities of data
- Need for innovative monitoring and modelling systems
- Bayesian methods are ideally suited to this task

- Bayesian methods allow the definition of a model for the data (a likelihood)
- They allow the inclusion of prior assumptions (however vague) about the parameters of the model
- They allow the testing of these assumptions via sensitivity analysis
- They have a natural sequential updating approach

- Define some data at time t as y_t. This could be a vector of disease incidences or rates. Also denote all data up to and including t as T.
- We assume that a likelihood for the data is available: $f(y_t|\theta)$ where θ are a set of parameters
  - This is just the distribution of the current data given parameters θ
  - a simple example: single disease, Poisson distribution, expected rate e_t, relative risk θ_t.
  - Then $f(y_t|\theta)$ is a Poisson distribution with mean $e_t\theta_t$. That is $y_t \sim Pois(e_t\theta_t)$
  - We could also assume a more complicated model for θ_t, where it is dependent on t itself or other covariables; we wont pursue this here.

Posterior and Prior Distributions

- In a Bayesian approach the parameters can also have distributions
- These are called *prior* distributions
  - In the Poisson example θ_t could have a gamma distribution with some parameters (α, β) say.
- The product of the likelihood and the prior distribution is called the *posterior* distribution of θ. This distribution tells us about the behaviour of θ given the observed data and our prior beliefs about θ.
  - In the Poisson example the posterior distribution was $f(y_t|\theta_t)$. The prior distribution was $g(\theta_t|\alpha, \beta)$where α, β are assumed known for simplicity. Hence the posterior distribution of θ_t is

$$f(y_t|\theta_t)g(\theta_t|\alpha, \beta).$$

  - This is the product of a Poisson likelihood and a gamma prior distribution

Sequential nature of the Posterior definition

- Bayesian methods lend themselves naturally to sequential problems
- This can be seen in the posterior definition: likelihood (current data)×prior
  - The prior here can be represented as the posterior distribution from *previous data.*This represents our beliefs about the data before we see the current data.
- Hence: denoting the posterior for a single parameter θ as $P(\theta|y)$:

$$P(\theta|y_t) = f(y_t|\theta)P(\theta|y_{t-1})$$

- We can also ask the question: what does the current model predict for the new time period? A predictive distribution can be constructed as $P(y_{t+1}|y_t)$

Bayesian alarm function

- y_t is the current data (counts usually) for a monitored site (could be a small area or address).
- y_T is the cumulative data on the disease up to and including time t.
- A parameter vector θ is defined.
- Syndromic variables are also available: x_t is one such variable and $\mathbf{x}_t$ is the vector of syndromic variables.
- Define the complete data and ancillary (syndromic) vector as

$$D_t = \left\{ \begin{array}{c} y_t \\ x_{1t} \\ x_{2t} \\ x_{3t} \\ . \end{array} \right. = \left\{ \begin{array}{c} y_t \\ \mathbf{x}_t \end{array} \right.$$

Posterior definition
Conditioning on $\mathbf{x}_T$

- A sequential posterior can be identified as

$$P(\theta|y_T, \mathbf{x}_T) \propto f(y_t|\theta, x_T)P(\theta|y_{T-1,}\mathbf{x}_{T-1})$$

where $P(\theta|y_{T-1,}\mathbf{x}_{T-1})$ is the posterior up to and including time $T-1$.

- The equivalent (posterior) predictive distribution is given by:

$$P(y_t|y_{T-1}) = \int f(y_t|\theta, \mathbf{x}_T)P(\theta|y_{T-1,}\mathbf{x}_{T-1})d\theta.$$

- Within an MCMC sampler this can be approximated via:

$$\approx \frac{1}{G}\sum_{g=1}^{G} f(y_t|\theta_{T-1}^{g}, \mathbf{x}_T)$$

where θ_{T-1}^{g} is the sampled parameter vector for the g th iteration from the posterior at $T-1$.

- This is called recursive Bayesian learning

Unconditional Version

- D_t is the vector of count data and syndromic variables at time *t*. The posterior given the evolution up to and including *t* is

$$P(\theta|D_T) \propto f(D_t|\theta)P(\theta|D_{T-1})$$

where $f(D_t|\theta)$ is the new data likelihood which could include correlations between elements (which could be *maps* or *time series*).

- The associated predictive distribution is given by :

$$P(D_t|D_{T-1}) = \int f(D_t|\theta)P(\theta|D_{T-1})d\theta$$

where $P(\theta|D_{T-1}) = f(D_{t-1}|\theta)P(\theta|D_{T-2})$.

Bayesian version of the optimal surveillance alarm function

- Define a frequentist alarm function for the current time (s) as:

$$P(x_s) = \sum_{k=1}^{s} \pi_k \prod_{u=k}^{s} \frac{f(x(u)|\mu')}{f(x(u)|\mu^0)} / \sum_{k=1}^{s} \pi_k.$$

  Here the function is designed to detect any change (of μ^0 to μ') on the range $k = 1, \ldots, s$. π_k is the probability of a jump at k given there hasn't been one before.
- Often for discrete times the geometric distribution is used for π_k.
- A Bayesian version of this would have

$$P(x_s) = \sum_{k=1}^{s} h(k) \frac{\prod_{u=k}^{s} f(x(u)|\mu')g(\mu'|u)}{\prod_{u=k}^{s} f(x(u)|\mu^0)g(\mu^0|u)} / \sum_{l=1}^{s} h(l)$$

Here $h(k)$ is the probability of a jump at k, and $g(\mu'|u)$ is the conditional prior distribution of the new μ value given the time u.

Note that for an alarm which is simply concerned with the jump at the present time (s) (and only then) the alarm function simplifies down to the Bayes Factor:

$$BF = \frac{f(x(s)|\mu')g(\mu'|s)}{f(x(s)|\mu^0)g(\mu^0|s)}.$$

Otherwise the alarm functions is a weighted product of posteriors for the $s-k+1$ time points with weights $w_k = h(k)/\sum_{l=1}^{s} h(l)$.

Definition of $\boldsymbol{f}(\boldsymbol{x}(\boldsymbol{u})|\boldsymbol{\mu})$

The density $f(x(u)|\mu)$ can be defined in different ways depending on the surveillance task.

Syndromic Vector Monitoring

Adopting the notation of section ref: uncond, the vector density for D_t yields $f(x(u)|\mu) \equiv f(D_u|\theta)$, and we generalize the jump to a vector form. In this case,

$$P(D_s) = \sum_{k=1}^{s} w_k \frac{\prod_{u=k}^{s} f(D_u|\theta')g(\theta'|u)}{\prod_{u=k}^{s} f(D_u|\theta^0)g(\theta^0|u)}.$$

This alarm could be extended to include dependence on previous observed data.

Conclusions

- Syndromic Surveillance is well suited to Bayesian methods
- Computational problems with evaluation of posteriors can be overcome using particle filters, windows, likelihood approximations and special spatial computational algorithms.
- Sequential nature of method an advantage

- Finally: What if model fails to hold ?