

# Bayesian Classification of Partially Observed Outbreaks Using Time-Series Data

COSMIN SAFTA, JAIDEEP RAY (Sandia National Laboratories, Livermore, CA, USA)

KAREN CHENG, DAVID CRARY (Applied Research Associates, Inc, Arlington, VA, USA)

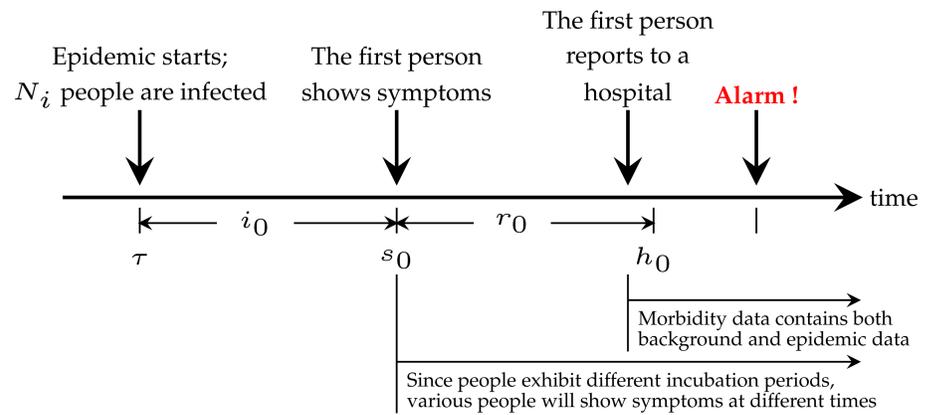
## Problem Statement

- Create Bayesian techniques to classify and characterize epidemics from a time-series of ICD-9 codes (will call this time-series a "morbidity stream")
- It is assumed the morbidity stream has already set off an alarm (through a Kalman filter anomaly detector) Starting with a set of putative diseases:
  - Identify which disease or set of diseases "fit the data best" and,
  - Infer associated information about it, i.e. number of index cases, start time of the epidemic, spread rate, etc

## Technical Approach

- Create probabilistic epidemic models for each disease
  - Sub-models: Probability density functions (PDFs) for incubation period, hospital visit delay & normalized infection rate
  - Identify unknown parameters and their epidemiological significance (e.g. number of index cases, decay rates of epidemics, parameters of visit delay distribution)
  - Identify prior beliefs for various epidemic parameters - this is often the place where there are differences between various epidemic models
- Pose and solve the statistical inverse problem for each epidemic model
  - Bayesian solution using an adaptive Markov Chain Monte Carlo technique - provides PDFs for the parameters being estimated
  - Model selection - use likelihood ratio tests, AIC & BIC to determine the most probable epidemic model

## Timeline of an Epidemic



- Note that the first person, "0", that shows symptoms at day  $s_0$  after  $i_0$  days since the start of the epidemic, might not be the first person to report to the hospital. Other person "j", with  $s_j > s_0$  may seek care earlier,  $h_j < h_0$ .
- $N_i$ ,  $\tau$ , and the parameters that control the incubation, spread rate, and hospital visit delays are inferred from morbidity data.

## Plague Model

The sub-models that comprise the plague epidemic are described below:

- Incubation period is modeled as a log-normal distribution (Gani and Leach [1])

$$f_{inc}(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{\log^2(t/4.3)}{2\sigma^2}\right)$$

where  $\sigma = 0.3762$  and  $t$  is the number of days since the start of the epidemic

- Visit delay and Infection rates are modeled using Gamma distribution with cumulative density functions (cdfs) given by [2]

$$F_{vd}(t, r_{vd}) = \frac{\gamma(1.992, r_{vd} \cdot t)}{\Gamma(1.992)}, \quad F_{ir}(t, r_{ir}) = \frac{\gamma(2, r_{ir} \cdot t)}{\Gamma(2)}$$

- $\gamma()$  is the lower incomplete Gamma function,  $r_{vd}$ ,  $r_{ir}$  are the rate parameters for the visit delay and infection rate distributions, respectively
- $t$  is the number of days after the person started showing symptoms

## Anthrax Model

The PDF of the incubation period is given by

$$f_{inc}(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{\log^2(t/\mu_D)}{2\sigma^2}\right)$$

where  $\mu_D$  and  $\sigma$  are computed based on Wilkening [3] A2 model.

- anthrax is not contagious so there is no need to model infection rate
- the incubation period depends on the average dose of anthrax spores  $D$ .
- the visit delay uses the same model as plague

## Fitting Models to Observations

- Predict  $\nu((t_i, t_{i+1}]) = M(t_i; \Theta)$ , the number of symptomatic people reporting for care in the time interval  $(t_i, t_{i+1}]$  (usually a day);  $\Theta$  is the set of model parameters
- Plague time series consists of patients who were index cases,  $\nu_i$ , and patients that caught from someone else,  $\nu_s$ .

$$\nu_i((t_i, t_{i+1}]; \Theta_P) = (1 - \alpha) N_{tot} \int_{\tau}^{t_{i+1}} f_{inc}(s - \tau) (F_{vd}(t_{i+1} - s, r_{vd}) - F_{vd}(t_i - s, r_{vd})) ds$$

$$\nu_s((t_i, t_{i+1}]; \Theta_P) = \alpha N_{tot} \int_{s_1=\tau}^{t_{i+1}} \int_{s_2=\tau}^{t_{i+1}} f_{ir}(s_1 - \tau) f_{inc}(s_2 - s_1) (F_{vd}(t_{i+1} - s_2, r_{vd}) - F_{vd}(t_i - s_2, r_{vd})) ds_2 ds_1$$

The set of unknown parameters is  $\Theta_P = \{N_{tot}, \alpha, \tau, r_{vd}, r_{ir}\}$ , where  $N_{tot}$  is the total number of people in the epidemic,  $\alpha$  is the fraction of people who are not index cases,  $\tau$  is the start of the epidemic, and  $r_{vd}$ ,  $r_{ir}$  are parameter which control the visit delay and spread rate, respectively.

- Anthrax time series consists of patients who were index cases only,  $\nu_i$

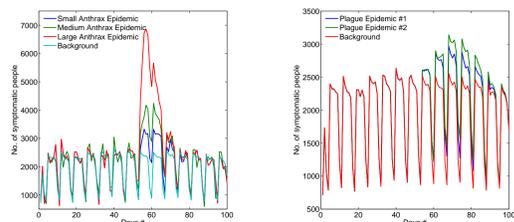
$$\nu_i((t_i, t_{i+1}]; \Theta_P) = N_i \int_{\tau}^{t_{i+1}} f_{inc}(s - \tau; \log_{10} D) (F_{vd}(t_{i+1} - s, r_{vd}) - F_{vd}(t_i - s, r_{vd})) ds$$

The set of unknown parameters is  $\Theta_A = \{N_i, \tau, \log_{10} D, r_{vd}\}$ , where  $N_i$  is the number of index cases,  $\tau$  is the start of the epidemic, and  $r_{vd}$  controls the visit delay.

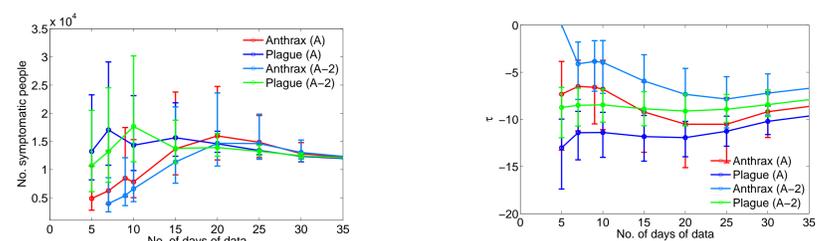
- Use Markov Chain Monte Carlo (with Gaussian errors) to find the best fits  $\Theta_P$  &  $\Theta_A$  for the time-series of observed symptomatic patients

## Epidemics' Time Series

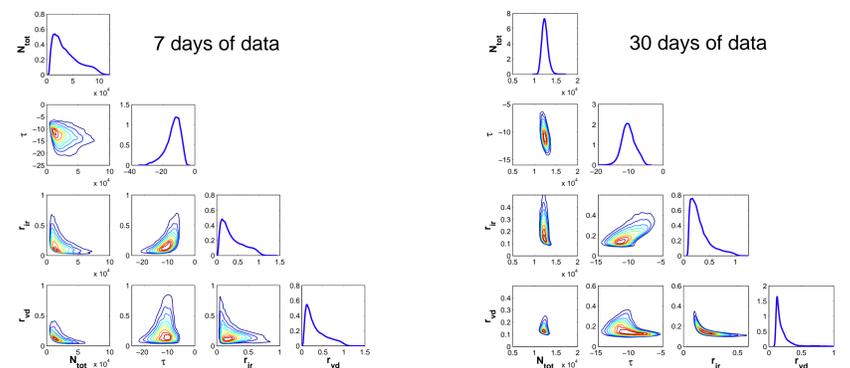
- Epidemics start at day 50.
- The alarm system is triggered at day 56.
- The background morbidity time-series is extracted from the data before model fitting.



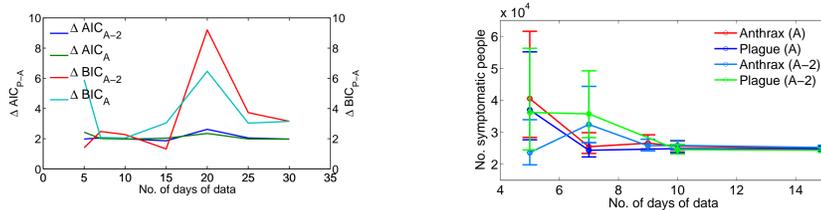
## Plague Epidemics - Parameter Estimation



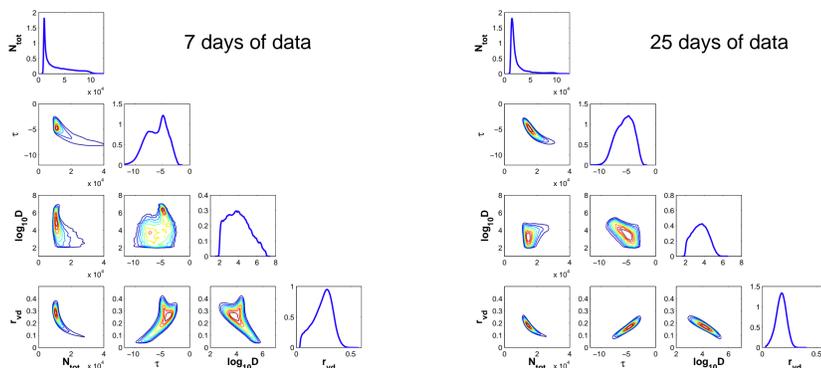
## Parameters' Joint & Marginal Distributions



## Anthrax Epidemics Likelihood Test & Parameter Estimation



## Parameters' Joint & Marginal Distributions



## Summary

- Results show that a time-series based classification may be possible
- For the test cases considered, the correct model can be selected and the number of index case can be captured within  $\pm\sigma$  with 5-10 days of data
- The low signal-to-noise ratio makes the classification difficult for small epidemics.

## References

- [1] R. Gani and S. Leach. Transmission potential of smallpox in contemporary populations. *Nature*, 414, 2001.
- [2] J. Ray, P. T. Boggs, D. M. Gay, M. N. Lemaster, and M. E. Ehlen. Risk-based decision making for staggered bioterrorist attacks: Resource allocation and risk reduction in. Technical report.
- [3] D. Wilkening. Sverdlovsk revisited: Modeling human inhalational anthrax. *Proceedings of the National Academy of Science*, 103:7589-7594, 2006.