Bayesian Disease Mapping

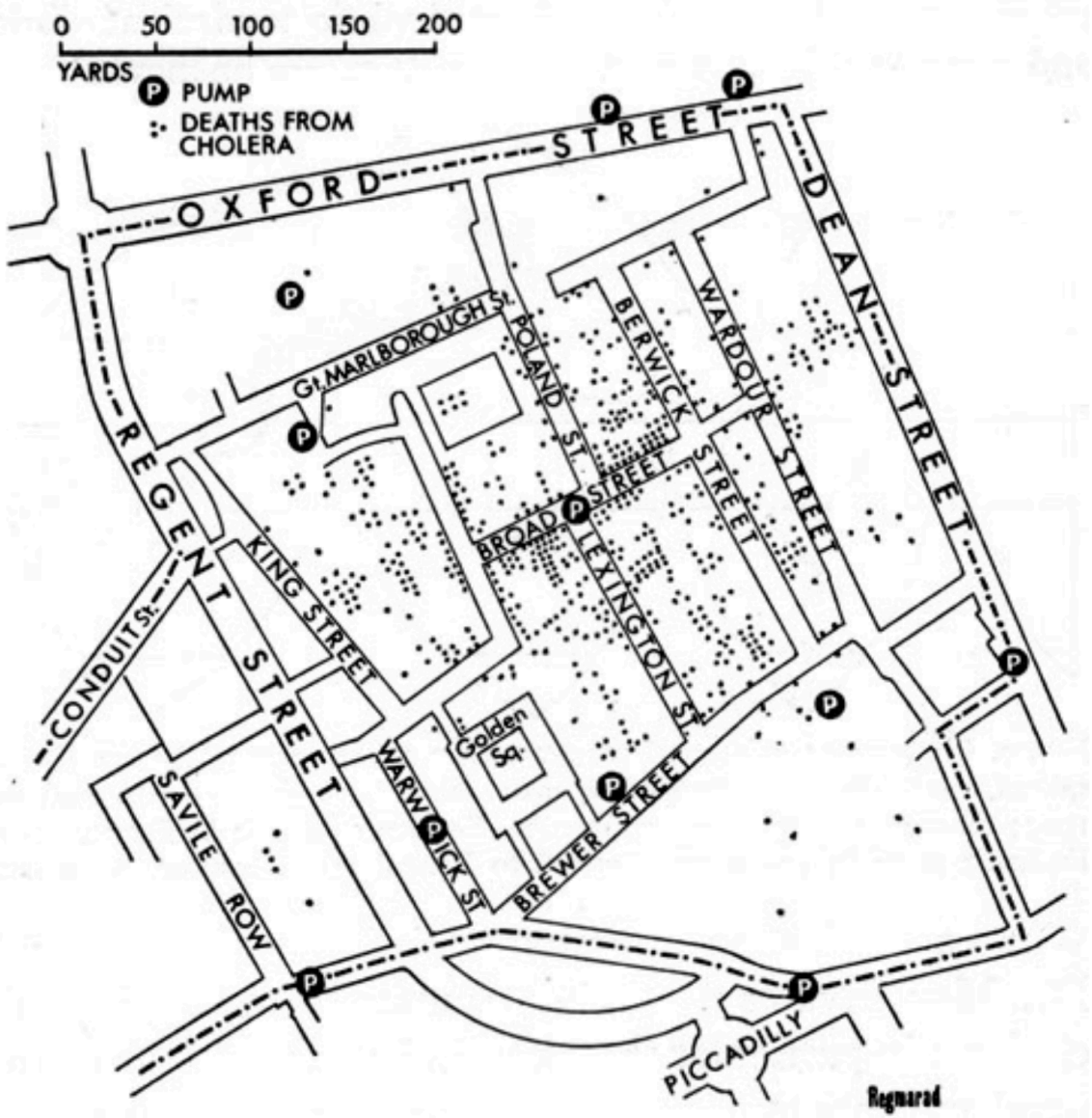
Yinuo Zeng

Background

- Apply Bayesian hierarchical models in geographical analysis of disease

- The goal of disease mapping is to provide visual summary of spatial information and identify patterns (spatial variation of the disease, areas of usually high risk) from the map

- Cholera outbreak in London's Board Street Region in 1854, by studying the spatial distribution of cholera victims around that area, John Snow find Cholera was spread through contaminated water.



Data

- HealthData.gov
- Amebiasis (a disease) in California
- 812 observations (58 Counties, 14 Years)
- 4 variables (County, Year, Count, Population)

County	Year	Count	Population
San Francisco	2001	162	782223
San Francisco	2002	141	783255
San Francisco	2003	98	781870
San Francisco	2004	92	780699
San Francisco	2005	119	779655
San Francisco	2006	116	782928
San Francisco	2007	93	791334
San Francisco	2008	120	798673
San Francisco	2009	94	801799
San Francisco	2010	106	806314
San Francisco	2011	80	813595
San Francisco	2012	58	822403
San Francisco	2013	60	830956
San Francisco	2014	49	837831

Model

Model 1:

$$Y_{it} \sim \text{Binomial}(p_{it}, n_{it})$$

$$\text{logit}(p_{it}) = a_0 + v_i + u_i + g_t$$

$$a_0 \propto c$$

$$v_i \sim \text{Normal}(0, \sigma_v^2)$$

$$u_i \sim \text{CAR}(\sigma_u^2), \text{ conditional autorgressive prior}$$

$$g_t \sim \text{Normal}(0, \sigma_g^2)$$

$$\sigma_v^2 \sim \text{IG}(0.01, 0.01)$$

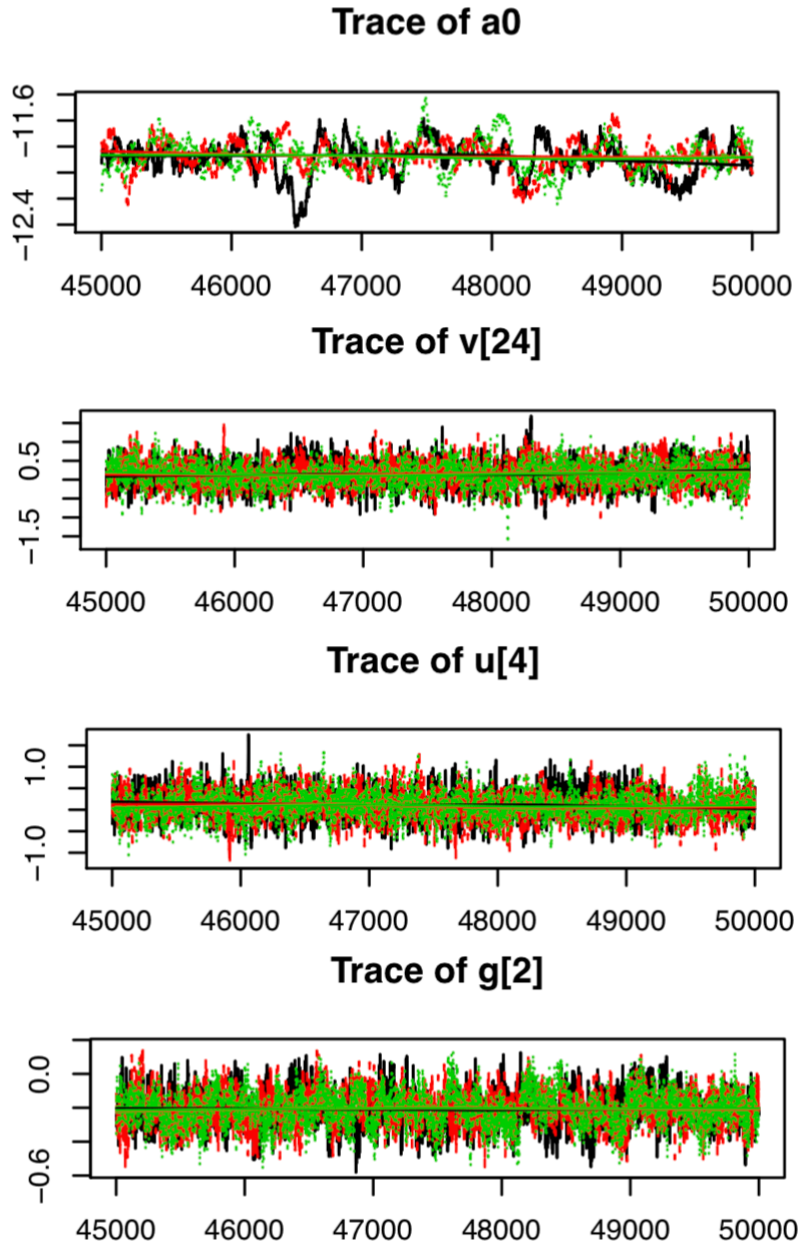
$$\sigma_u^2 \sim \text{IG}(0.01, 0.01)$$

$$\sigma_g^2 \sim \text{IG}(0.01, 0.01)$$

Scale Reduction Factor

min	mean	median	max
0.9999	1.0025	1.0002	1.1329

u[37]	u[41]	v[37]	v[41]
1.111583	1.114969	1.101102	1.132879



Model comparison

Model 2:

$$Y_{it} \sim \text{Binomial}(p_{it}, n_{it})$$

$$\text{logit}(p_{it}) = a_0 + v_i + u_i + \beta t$$

$$a_0 \propto c$$

$$\beta \sim \text{Normal}(0, 0.0001)$$

$$v_i \sim \text{Normal}(0, \sigma_v^2)$$

$$u_i \sim \text{CAR}(\sigma_u^2)$$

$$\sigma_v^2 \sim \text{IG}(0.01, 0.01)$$

$$\sigma_u^2 \sim \text{IG}(0.01, 0.01)$$

Model	DIC
1	1171
2	2518

Bayesian Disease Mapping

Likelihood: $Y_{it} \sim \text{Binomial}(p_{it}, n_{it})$

MLE: $p_{it} = \frac{y_{it}}{n_{it}}$

R Shiny:

Bayesian estimate vs Frequentist estimate

Bayesian Disease Mapping

Summary

- Unlike MLE, posterior mean is more smooth

- Improvements:
  - Interaction of spatial and temporal effects
  - covariates (temperature, age)

Thank You!

Identifiability Problem

$$Y_{it} \sim \text{Binomial}(p_{it}, n_{it})$$

$$a_0 \propto c$$

$$v_i \sim \text{Normal}(0, \sigma_v^2)$$

$$u_i \sim \text{CAR}(\sigma_u^2)$$

$$g_t \sim \text{Normal}(0, \sigma_g^2)$$

$$w_t \sim \text{Normal}(w_{t-1}, \sigma_w^2) \text{ for t=2...14}$$

$$w_1 \sim \text{Normal}(0, \sigma_w^2)$$

$$\sigma_v^2 \sim \text{IG}(0.01, 0.01)$$

$$\sigma_u^2 \sim \text{IG}(0.01, 0.01)$$

$$\sigma_g^2 \sim \text{IG}(0.01, 0.01)$$

$$\sigma_w^2 \sim \text{IG}(0.01, 0.01)$$