Introduction to Bayesian Mapping Methods

Andrew B. Lawson©
Arnold School of Public Health
University of South Carolina*

* now at Public Health Sciences
MUSC, Charleston, SC, USA
• South Carolina congenital abnormality deaths 1990
Mapping issues

• Relative risk estimation
• Disease Clustering
• Ecological analysis
Relative risk estimation

• SMRs (standardized mortality /morbidity ratios
Congenital abnormality deaths SMR 1990 using 8 year rate

- 1.51 to 4.1 (9)
- 1.09 to 1.51 (9)
- 0.78 to 1.09 (9)
- 0.5 to 0.78 (9)
- 0 to 0.5 (10)
Some notation

- For each region on the map:
  - $y_i$ is the count of disease in the $i^{th}$ region
  - $e_i$ is the expected count in the $i^{th}$ region
  - $\theta_i$ is the relative risk in the $i^{th}$ region
- The SMR is just $smr_i = y_i / e_i$
- This is just an estimate of $\theta_i$
SMR problems

• Notoriously unstable
• Small expected count can lead to large SMRs
• Zero counts aren’t differentiated
• The SMR is *just the data!*
Smoothing for risk estimation

• Modern approaches to relative risk estimation rely on smoothing methods
• These methods often involve additional assumptions or model components
• Here we will examine only one approach: Bayesian modeling
Bayesian Modeling

Some statistical ideas:

• Likelihood…….we usually assume that counts of disease have a Poisson distribution so that $y_i$ has a Poisson distribution with expected value $e_i \theta_i$

• We usually write this as $y_i \sim \text{Pois}(e_i \theta_i)$ for short

• The counts have a Poisson likelihood
Likelihood

• The counts have a joint probability of arising based on the likelihood \( L(y, \theta) \):
• \( L(y, \theta) \) is the product of Poisson probabilities for each of the regions
• This tells us how likely the data are given the expected rates \( (e_i \theta_i) \)
• It also tells us what the most likely values of \( \theta \) are given the data observed.
Maximum Likelihood

• The SMR is the value of $\theta$ which gives the highest likelihood for the data (under a simple Poisson model)….this is called maximum likelihood (ML)

• This approach is often used in statistics to get good estimates of parameters

• *Here we go beyond ML*
Smoothing using Bayesian methods

• One way to produce smoother relative risk estimators is to assume that the risk has a distribution

• In Bayesian terms this is called a prior distribution

• In the Poisson count example the commonest prior distribution is to assume that $\theta_i$ has a Gamma distribution
A simple Hierarchy

- $y_i \sim \text{Poiss}(e_i \theta_i)$
- $\theta_i \sim \text{Gamma}(\alpha, \beta)$
- This a very simple example which allows the risk to vary according to a distribution
- $\alpha$ and $\beta$ are unknown here and we can either try to estimate them from the data OR give then a distribution also:
- E.g. $\alpha \sim \text{exp}(\upsilon), \beta \sim \text{exp}(\rho)$
Model hierarchy
Summary

- Bayesian models are useful for smoothing disease relative risk estimates
- They use prior distributions for parameters
- The priors can be multi-level
- The prior distributions can control the model results
- Sensitivity to prior distributions is important
A basic Hierarchy

- Data
  - Parameter
  - Parameter
  - Parameter
  - Parameter

- Data
  - 1st level
  - 2nd level
- distribution
distribution

CDC 2003 cancer conference
Modern Posterior inference

• Unlike the usual ML estimates of risk, a Bayesian model is described by a distribution and so a range of values of risk will arise (some more likely than others)

• Posterior distributions are sampled to give a range of these values (posterior sample)

• This contains a large amount of information about the parameter of interest
A Bayesian Model

• A Bayesian model consists of a likelihood and prior distributions

• The product of the likelihood and the prior distributions gives the most important distribution: the posterior distribution

• In Bayesian modeling all the inference about parameters is made from the posterior distribution.
Posterior Sampling

• The posterior distribution gives information about the distribution of parameters: not just about the most likely value
• It is now relatively simple to obtain samples of parameters from posterior distributions
• The commonest method for this is Gibbs Sampling
WinBUGS

- This package has been set up to provide relatively easy access to Gibbs Sampling for a range of hierarchical models
- The package is very flexible and implements Gibbs Sampling (and other Markov Chain Monte Carlo (MCMC) methods)
- It also includes a GIS module called GeoBUGS which allows the mapping of the resulting fitted parameters (e.g. relative risks)
Disease Mapping on WinBUGS

• WinBUGS is a very powerful tool which can be applied to:
  – Relative risk estimation
  – Putative health hazards (focused clustering)
  – Ecological analysis
A Simple Example

- South Carolina congenital abnormality deaths 1990
- Data: counts of deaths in counties of South Carolina
- Expected rates available as age x sex adjusted rates
- The SMR map is next:
SMR for congenital anomalies

CDC 2003 cancer conference
Gamma Poisson model: WinBUGS
Using WinBUGS

• WinBUGS is a windowed version of the BUGS package. BUGS stands for Bayesian inference using Gibbs Sampling
• The package must be programmed to sample form Bayesian models
• For simple models there is an interactive Doodle editor; more complex models must be written out fully.
WinBUGS Introduction

License Agreement

Introduction

This software and any associated documentation whether electronic or printed (hereinafter called "WinBUGS PACKAGE") is made available under a licence agreement and may be used only in accordance with the terms of that agreement.

This is a legal agreement between you (the Licencee), and MRC and Imperial College of Science, Technology and Medicine (the Licensor). The terms of the licence are provided in the following pages.

Users are required to register and to pay a fee for the use of the WinBUGS PACKAGE. Details of fees and the procedure for registration and acceptance of the licence terms is provided here.

There is no fee payable for the use of the demonstration (Internet) version of the WinBUGS package. Users of the demonstration version of the WinBUGS package can upgrade to the full version on payment of a fee.

The current fee is zero dollars ($0).

By completing and sending the registration you demonstrate your agreement to the terms of this licence and will become legally bound to the terms therein.

It should be emphasised that the statistical tools provided in the WinBUGS PACKAGE are by their very nature partly subjective. The Licensor cannot offer advice on interpretation of results obtained using the WinBUGS PACKAGE. Any assistance will be strictly limited to attempting to help if there are problems formulating the statistical problem with the WinBUGS PACKAGE.
Doodle Editor

- The doodle editor allows you to visually set up the ingredients of a model
- It then automatically writes the BUGS code for the model
BUGS code and Doodle stages

```r
model
{
  for (i in 1:m)
  {
    y[i]~dpois(mu[i])
    log(mu[i])<-log(e[i])+v[i]
    theta[i]=-exp(v[i])
    v[i]~dnorm(0,tau)
  }
}
```
Final doodle
Demonstration
Demonstration

- Doodle example with simple nodes
- SC congenital anomalies 1990
- Example 6.1.2 (burn-in 2000, final 6000 iterations)
- Example 6.1.3 Log-normal model (6000 iterations)
- Example 6.1.5 CAR –normal model (15000 iterations)
Extensions

- Space-time modeling (Section 6.1 6)
- Mixture modeling (section 6.1.7)
- Focused clustering (analysis of putative health hazards) (Chapter 7)
- Binomial models (Section 8.3.2)
- Ecological regression (chapter 8)
- Spatial survival analysis (Chapter 9)
Conclusions

• WinBUGS provides a free and relatively easy-to-use tool for disease mapping with small area count data
• Allows state-of-the-art approach to relative risk and ecological regression
• Available from: www.mrc-bsu.cam.ac.uk/bugs