Bayesian Hierarchical Approaches to Spatial Analysis of Injury and Disaster Data

Charles DiMaggio, PhD*
Columbia University
Departments of Anesthesiology and Epidemiology

August 10, 2012

1 Introduction

The motivation for Bayesian approaches to spatial modeling lies in the difficulties of spatial data that we've discussed. Data points near each other will be very similar in terms of the kinds of variables, like demographics, SES and geographic features, in which we are likely to be interested as epidemiologists, making familiar approaches like linear or logistic regression inappropriate. Poisson models of the kinds of count data we find in spatial epidemiology while an attractive option, are subject to their own difficulties. The data tend to be over-dispersed, meaning that the variance is greater than the mean [1] While a number of effective approaches to spatial data analysis exist, the spatial data we work with as epidemiologists are most often not the kind of highly ordered, 'lattice' or point-process data for which many spatial analytic techniques have been developed.

In this chapter, we'll try to tackle Bayesian Hierarchical Modeling of spatial data. Bayesian analysis is a vast and rapidly expanding field. Space constraints here preclude a more general and thorough treatment of the topic of Bayesian epidemiological analysis. [2] We will for now limit ourselves to focused introduction and then (in the next chapter) return to applications in the New York City TBI vignette.

Most of this section of the notes are based on Andrew Lawson's texts and workshops which are well-worth pursuing if you would like a more thorough treatment of the subject. I particularly recommend Bayesian Disease Mapping. And if you have the opportunity, by all means attend one of Dr. Lawson's workshops.

*Email: cjd11@columbia.edu

[1] A Poisson distributed variable has a single parameter, λ for both its mean and variance

[2] There are some excellent such texts. I highly recommend Andrew Gelman's introductory text and David Speigelhalter's excellent early book on the subject

We'll first consider the types of data and questions for which Bayesian approaches are suited. Then we'll introduce the basic theory of Bayesian statistics, and proceed to describe the gamma-Poisson model as a flexible and useful approach in the multilevel setting frequently encountered in spatially distributed count data. I'll present an example of the methods, using data from Hurricane Katrina in Orleans Parish, Louisiana, and consider some of the advantages and limitations of Bayesian spatial analysis for the practicing epidemiologist.

2 The problem with place

As we have seen, aggregating data to the group level based on geography is not simple. Bringing data together spatially frequently results in heterogeneous and arbitrary groupings that may be too large and undifferentiated to capture risk appropriately. Analyses that rely on variable specification based on irregular geographic units, such as ZIP Codes, may be affected by extreme values based on a few cases in small populations. Rare events contribute to more heterogeneity than is assumed by commonly used epidemiological methods like the Poisson models. Finally, epidemiologically influential covariates of an outcome, which may be unmeasured, are likely to be similar in adjacent areas resulting in spatial autocorrelation.

A a basic measure of increased occurrence might be a ratio that compares observed to expected counts of an outcome in a geographic area like a census block. We could then calculate some risk that explains any change from the expected number to the observed number. So, for example, if there were no risk in a particular area, this risk factor would be equal to 1, and the observed number would be equal to the expected number. If, on the other hand, there was some increased or decreased risk of in a particular area this number would be greater than or less than 1. This is a standardized mortality (or morbidity) ratio (SMR), where for region i:

- y_i is the observed count for some outcome
- e_i is *expected* count
- θ_i is the (unknown) parameter for the relative risk

A crude estimate of the risk θ_i would be $smr_i = \frac{y_i}{e_i}$

As noted, this kind of data is subject to the problems of non-independence. Census blocks near each other are likely to be similar in important ways based on geography and population demographics. The areal units are also often defined in irregular and arbitrary ways unrelated to their potential use for health outcomes analyses. In the United States ZIP Codes are intended as a convenient means of delivering mail. They are far from regularly arranged yet they have been treated as lattice-like for point-process data. Finally, count data of the kind often used in spatial injury epidemiology are subject to over dispersion and instability. Small expected numbers in the denominator e.g. say only two household in a census

block, can lead to large inflated risk estimates if only one household is affected. These characteristics call into question the suitability of such approaches as simple Poisson models.

2.1 The Rev. Bayes meets Dr. Snow

A century before John Snow mapped cholera deaths in London Thomas Bayes, an English minister and mathematician, sparked an approach to conditional probability that bears his name[3] and that addresses, in many respects, the problems inherent in spatial epidemiological analysis. At it's most basic level, the Bayesian approach to knowledge asks: How do we combine what we expect with what we see? Or put somewhat differently, how do we learn from the data to update our knowledge?

Clinicians, who I believe tend to be natural Bayesians, are taught 'hoof beats usually mean a horse is approaching', and that much more information is needed before concluding it is a zebra. We turn to this mode of thinking in our approach to spatial analysis. How does what we expect to see in a region, based on the surrounding regions, impact our conclusions about what we actually see? Bayes Theorem formalizes and quantifies this common-sense approach to evidence and expectations.

2.2 From common sense, to numbers.

Statistically, in a Bayesian approach, we base our conclusions about the probability of a risk estimate given our data, $Pr[\theta|y]$, on a combination of our prior expectation, expressed as the probability of observing some risk estimate $Pr[\theta]$, and the likelihood of observing the data with which we are presented, $Pr[y|\theta]$:

$$Pr[\theta|y] \propto Pr[\theta] \cdot Pr[y|\theta] \tag{1}$$

When we have a lot of data based on our observations, the likelihood of that data tends to overwhelm any prior expectations we might have had to the contrary. The less data we have, the more influence our prior expectation will have.

Our *prior distribution* essentially dictates how we believe the parameter θ would behave if we had no data from which to make our decision. What, for example, might we expect is the probability that someone living within 3 miles of a certain location would die from a gun shot wound? Our best guess might be, for example, 1 in 20 or about 5%, and that this probability varies around this point estimate in a normal fashion with a variance of say 0.01 or 1%. This estimate may be based on previous studies, law enforcement data, clinical experience or a combination of sources. What then, if we conduct a study that indicates the risk of firearm related

[3]Stigler's Law of Eponomy states that 'No scientific discovery is named after its original discoverer'. In this case Thomas Bayes clearly got the ball rolling (if you know his original thought experiment, then pun intended) but folks like Richard Price and especially Pierre Simon Laplace should rightfully have their names attached to the theory. Still, 'Bayesian' has a certain ring to it.

fatality within 3 miles of the location is 45%? How do we revise what we think about the risk of firearm-related deaths in this area? Any revision will depend in large measure on how *likely* these data or observations are given our expectations (and model). This, then, is the second bit of information on the right side of the equation, referred to as the *likelihood* or the *likelihood function* which represents the probability we assign our observed data given the postulated parameters represented by our prior. Our *posterior distribution*, or revised probability of the parameter θ, is a combination of these two probabilities. [4] In a very common sense way, it tells us, for example, that if the results of our study differ markedly from our best existing information we should perhaps be somewhat skeptical.

2.3 From numbers to BUGS

For many years, approaches to combining the prior and the likelihood were restricted to a set of situations where the prior distribution was conjugate to, or in the same family, as the likelihood and could be derived by some fairly simple updating of the likelihood function. So, for example, in the setting of an standardized mortality ratio (SMR), a $\Gamma(\alpha, \beta)$ prior is conjugate to a $Pois(\lambda)$ likelihood and can be updated with the information in the likelihood function through the simple formula $\Gamma(\alpha+y, \beta+e)$ [5]

Many reasonably realistic problems, though, are not amenable to conjugate analyses. There may not be an appropriately conjugate prior. They may require higher order differential equations that do not have closed or simple solutions. [6] In these cases, simulation approaches can be use to solve for the parameter estimates, but for may years the required computing power for the kinds of simulations necessary, and more importantly, an approach or algorithm to define a valid sample space from which to simulate were not widely available or widely appreciated.

Software developed over the past decade or so, of which the WinBUGS package developed in the UK is a notable example, takes advantage of both advances in computing and in the understanding of constructing Monte Carol Markov Chains to make such simulations approach. WinBUGS (which stands for Windows Bayes Using Gibbs Sampling) samples from a proposed posterior distribution using either Gibb's (for which it's named) or Metropolis Hasting algorithms.

The details of the sampling schemes used in packages like WinBUGS is beyond

[4] And leads to the essential mantra of Bayesian analysis: 'The posterior is proportional to the prior times the likelihood.' Which should be invoked like an incantation when doubt and confusion arise.

[5] Don't worry too much if this isn't entirely clear at this point.

[6] For a meaningful appreciation of Bayesian analysis, you may have to review some (though not necessarily a lot) of your college calculus. In this case, you may think of "higher order" as referring to equations that require multiple factors to solve, and "simple or closed" as equations that have only one solution or that can be integrated to 1. This last point is critical, since we're dealing with probabilities which, by definition must sum or integrate to 1.

the scope of this short description, [7] but it is important to note that this approach carries with it additional responsibilities for the analyst.

First, because these are Markov Chains, each value is dependent on the prior value in the chain; we must assess correlation of sample values. We do this by reviewing autocorrelation graphs and statistics. Ideally, we would like any correlation among values to drop off after the first lag.

Second, since we are trying to sample the (posterior) target distribution as fully and as efficiently as possible, our proposal distribution should sample an appropriately wide area of the proposal distribution. This part of practical Bayesian analysis can get messy, particularly if our proposal distribution is too narrow or our starting value is off somewhere in the hinterlands of a large multi-dimensional target posterior distribution. One approach to this task is to calculate acceptance rates. For random walk chains with normal proposal densities, ideal acceptance rates would be about 50% for one parameter model and about 25% for multi-parameter models. Gibb's sampling, which is based on defining a multi-parameter distribution conditional on one parameter given all the others, can help obviate some of this fine tuning necessary for more traditional Metropolis-Hastings algorithms, which require that the proposal distribution be specified *a priori*.

Third, to help ensure that we are, indeed, sampling from the stable underlying target distribution, we also do things like evaluate whether the chain of sample values actually *converges* to a stable distribution that is consistent with the posterior distribution in which we are interested. Practically, we do this by running 2 or 3 *chains* of samples each starting with an *initial value* chosen from widely dispersed areas of the target distribution, [8] and then evaluating the chains to make sure they are all sampling form the same distribution. This evaluation can be as simple as looking at the kernal density of the distribution graphically to make sure they are reasonably overlapping, to calculating statistics such as the Brooks-Gellman-Rubin (BGR) that compares within chain variation to across chain variation.

Finally, and in some respects most critically, we are obligated to evaluate our choice of prior distributions, some of which may be more influential or appropriate than others. This can be accomplished with sensitivity analyses substituting and evaluating the effects of different prior distributions.

[7] In addition to Lawson, a couple of excellent references and introductory texts for all this stuff are: *(1)* Albert, J. (2009) Bayesian Computation with R. New York: Springer. *(2)* Banerjee, S., Carlin, B. P., and Gelfand, A. E. (Eds.). (2004). Hierarchical modeling and analysis for spatial data. Boca Raton, Chapman and Hall/CRC. *(3)* Gelman A, Carlin JB. (2009) Bayesian Data Analysis. Boca Raton: Chapman and Hall/CRC. *(4)* Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol, 35(3), 765-775. and *(5)* Spiegelhalter D, Abrams K, and Myles J. (2004) Bayesian Approaches to Clinical Trials and Healthcare Evaluation. West Sussex, John Wiley and Sons.

[8] If this sounds a lot easier said than done, it is.

2.4 From BUGS to a model

Bayesian thinking lends itself naturally to the kind of hierarchical models suited to areal spatial analysis. We can specify not only a distribution for how we believe *individual* risk (θ_i) is distributed, but also, by specifying an additional set of parameters, how we believe θ varies across higher levels of organization, such as geographic units. One could, for example, say that y_i is the empirical (observed) rate of some event in a geographic area i, θ is the true underlying rate, and some additional parameter(s) how that true rate varies across all such areas in which we are interested.

To begin building a model, we must define our prior and our likelihood. Let's start with the likelihood, or data component of the model. For count data of the kind with which we frequently work in epidemiology, we assume an underlying Poisson distribution. Often described as the distribution of rare events, the Poisson distribution is characterized by a single parameter, λ (i.e., both $\mu = \lambda$, and $var = \lambda$). λ is the rate per unit time at which some event k occurs, and the Poisson distribution is defined as:

$$Pois(\lambda) = e^{\lambda^k}/k! \tag{2}$$

The y_i counts in area i, are independently identically Poisson distributed and have an expectation in area i of e_i , the expected count, times θ_i , the risk for area i :

$$y_i, iid \sim Pois(e_i\theta_i) \tag{3}$$

Having defined our likelihood, we next define a prior distribution for this likelihood. A useful and commonly used prior distribution for θ in the setting of spatial analyses is the gamma (Γ) distribution:

$$\theta \sim \Gamma(\alpha, \beta) \tag{4}$$

where, $\mu = \alpha/\beta$, and $var = \alpha/\beta^2$

The Gamma distribution consists of a very flexible class of probability distributions. For example, $\Gamma(1, b)$ is exponential with $\mu = 1/b$, $\Gamma(\frac{v}{2}, \frac{1}{2})$ is chi square distributed with v degrees of freedom. Gamma distributions are constrained to be positive, which is necessary when dealing with count data, and setting the two parameters equal to each other $\alpha = \beta$ results in a null value of 1, which is useful for modeling risk estimates. Finally, the Gamma distribution is conjugate to the Poisson distribution, making our prior and our likelihood of the same family which not only allows for simplified statistics in one parameter problems, but carries with it additional advantages in terms of a valid choice of prior in more complicated analyses.

2.4.1 The Poisson-gamma model

Since a basic Bayesian assumption is that any parameter in a problem has a prior distribution of its own, the α and β parameters in the gamma also have prior distributions. The usual approach is to put exponential distributions on α and β. [9]:

$$y_i \sim Pois(e_i\theta_i) \tag{5}$$
$$\theta \sim \Gamma(\alpha, \beta) \tag{6}$$
$$\alpha \sim exp(v), \tag{7}$$
$$\beta \sim exp(\rho) \tag{8}$$

Below is an example of the code for this hierarchical model in BUGS language. The code looks a lot like R. But not really.

```
model
{
For (i in 1:m)                # loop to repeat for every spatial level
{
       y[i]~dpois(mu[i])      # Poisson likelihood for observed counts
      mu[i]<-e[i]*theta[i]
      theta[i]~dgamma(a,b)    # relative risk
}
        a~dexp(0.1)           # hyperprior distributions
        b~dexp(0.1)
}
```

Following that is an illustration of the basic hierarchy in a directed acyclic graph (called a 'Doodle' in the BUGS world) for the model (on the left) and a description of the components of the graph (on the right).

[9] At this point we begin to see the inherently hierarchical nature of the Bayesian approach

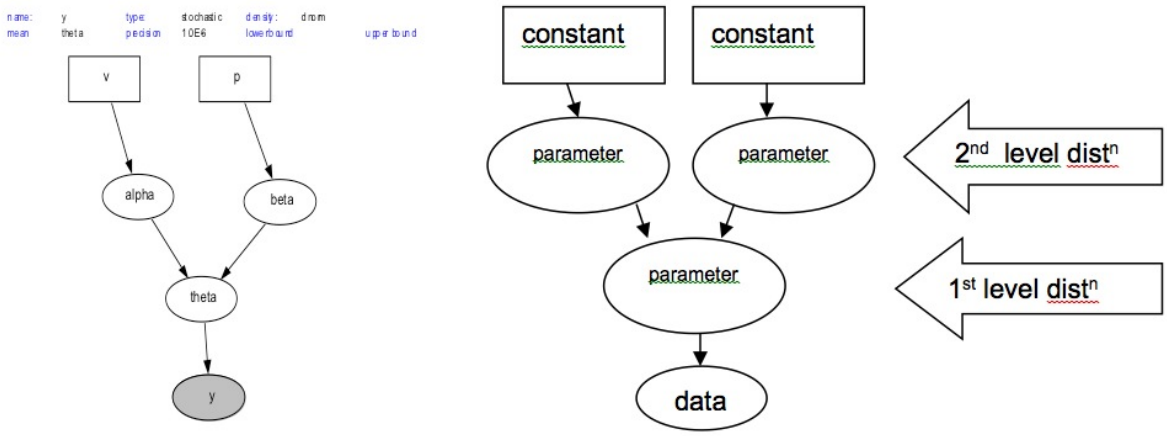

Figure 1: Poisson-gamma Spatial Model Hierarchy

2.4.2 Random and spatial effects

Taking the natural logarithm of our hierarchical model allows the inclusion of linear regression terms, a random effects term and a spatial effects term:

$$y_i \sim Pois(\mu_i) \tag{9}$$

$$log(\mu_i) = \beta_n + T_1 + T_2 \tag{10}$$

$$\tag{11}$$

Where β_n represents a vector a log-linear regression terms for variables that might capture potential confounders like age, gender, or socio-economic status, T_1 represents a random effect term, and T_2 represents a spatial effects term.

Random effects terms have been proposed [10] as a useful way to account for group-level heterogeneity . Basically, when there is more variation, or 'noise', in data than is accounted for by the individual-level model and error, we separate out part of the error or residual variance ($r \sim nl(0, \sigma)$) that we believe is due to the groups that give rise to or within which the individuals are nested ($v \sim nl(0, \sigma)$). This variance or heterogeneity though, is not explicitly spatially structured.

The explicit spatial effects component in the model can be represented by a conditional autoregressive (CAR) term, which we first encountered in our discussion

[10] Though not always accepted...

of spatial linear models. As you may recall, a CAR model is based on a set of spatial neighborhoods. In the usual formulation, each neighborhood consists of adjacent spatial shapes that share a common border. The mean θ_j in neighborhood j is normally distributed with its parameters defined as μ_j the average of the μ_{ij}'s in the neighborhood and σ_j equal to the σ's of the neighborhood μ_{ij}'s divided by the number (δ_j) of spatial shapes in the neighborhood. [11]

$$\mu_j \sim nl(\bar{\mu_\delta}, \tau_\mu / n_\delta) \tag{12}$$

In much the same way the random effect term captures unstructured heterogeneity, the CAR term captures spatially structured heterogeneity or variance in the data that is not captured by your risk model.

Our updated model, then looks like this:

$$y_i \sim Pois(e_i \theta_i = \mu_i) \tag{13}$$
$$log(\mu_i) = \beta_n + \nu_i + \upsilon_i \tag{14}$$
$$\nu \sim nl(0, \tau_\nu) \tag{15}$$
$$\upsilon \sim nl(\bar{\upsilon_\delta}, \tau_\upsilon / n_\delta) \tag{16}$$

The WinBUGS package makes it relatively easy to specify a CAR model. You first define an adjacency matrix, which is basically a list of all block groups that share an adjacency. You then define a set of weights for those adjacencies. The most straightforward and most commonly used approach is a weight of 1 when two spatial shapes share an adjacency and a weight of zero when they do not. The following code demonstrates what a model might look like in BUGS:

```
model
{

for( i in 1 : m ) {
      y[i] ~ dpois(mu[i])
      mu[i] <- e[i] * rr[i]
      log(rr[i]) <- b0 + b1*variable1 + b2*variable2 + b3*variable3 +v[i] + h[i]
                                        # h is CAR term
                                        # v is random effects term
r[i]<-(y[i]-mu[i])                      # r is residual

v[i]~dnorm(0,tau.v)        # prior on random effects needs to be inside loop
```

[11] The spatially structured component CAR is sometimes described as a Gaussian process $\lambda \sim Nl(W, \tau_\lambda)$ where W represents the matrix of neighbors that defines the neighborhood structure, and the conditional distribution of each λ_i , given all the other λ_i 's, is normal with $\mu =$ the average λ of its neighbors and a precision (τ_λ).

```
   }
# priors
b0~dflat()
b1~dnorm(0, 0.001)
b2~dnorm(0, 0.001)
b3~dnorm(0, 0.001)

h[1:m]~car.normal(adj[], weights[], num[], tau.h)    #prior on CAR
 tau.v~dgamma(0.001,0.001)               # priors on v and h
             tau.h~dgamma(0.001,0.001)
}
}
```

3 From theory to practice: Post-Katrina Repatriation Model

Before applying these methods to our ongoing example about traumatic brain injury in New York City, I'll illustrate the basic approach with a model I developed to try to document and help explain why some New Orleans neighborhoods seemed to rebound relatively quickly, while others languished (and some continue to languish) for years following Hurricane Katrina. The data is US Postal Service counts of the number of households in a census block actively receiving mail. We compare pre-hurricane counts in June 2005 (our expected number) with June 2009 counts (observed number). θ_i is an estimate for the increase or decrease in the observed vs. the expected number of households receiving mail in a census block group using a Bayesian smoothed estimate of y_i/e_i. (Note that we are interested primarily in the decreased 'risk' of mail delivery in a census block). Based on some preliminary, non-spatial, analyses, we are interested in the role of income and geographic elevation above sea-level as predictors of repatriation.

We first plot the unadjusted observed vs. the expected proportion of households receiving mail in block groups pre and post hurricane. The results are fairly heterogeneous, with some areas actually experiencing a net increase in population following the disaster.

We then map our social and environmental variables to see how they vary geographically and if that variation may be consistent with changes in repatriation.

There appears to be, even on this subjective level, an inverse relationship between these two variables. Our next step, is to examine their relationship to repatriation more rigorously. For this, we turn to a Bayesian hierarchical model.

In addition to assessing whether poverty and/or elevation above sea level explains whether folks returned to their homes in Orleans Parish after Hurricane Katrina we also include a conditional auto-regression (CAR) term to 'smooth' the outcome estimate and a a random effects term to capture additional non-modeled variability

Figure 2: Percent Observed vs. Expected Households Receiving Mail, Orleans County, LA, June 2005 vs. May 2009, by Census Block Group.

across the ecologic units of census block groups. We compared models using the Deviance Information Criterion (DIC) which is a Bayesian analogue of the Akaike Information Criterion (AIC), where a relatively smaller value is considered a 'better' model. We found that by this criterion, a model consisting of the CAR term, poverty, elevation and an interaction term for poverty and elevation is the 'winner'.

```
  model
{
for( i in 1 : m ) {
```

Figure 3: Proportion of Population Living Below Federal Poverty Level, Orleans County, LA, 2000, by Census Block Group.

```
      y[i] ~ dpois(mu[i])
      mu[i] <- e[i] * rr[i]
log(rr[i]) <- b0 + b1*((elevation[i]-3.7)/3.6) + b2*((poverty[i]-28.4)/18.2) + b3*((elev
r[i]<-(y[i]-mu[i])                                  # r is residual
prob50[i]<-step(0.5-rr[i])                # step function to calculate and plot exceedenc
   }
h[1:m]~car.normal(adj[], weights[], num[], tau.h) #no need weights in model get from mat
                         b0~dflat()
                         b1~dnorm(0, 0.001)
```

Figure 4: Elevation (in Meters) Above Sea Level. Orleans County, LA, by Census Block Group

```
b2~dnorm(0, 0.001)
b3~dnorm(0, 0.001)
tau.h~dgamma(0.001,0.001)    }
```

Notice that we transformed or re-parameterized our poverty and elevation variables by subtracting the mean dividing by the standard error. The explanation for this lies in assessing convergence to the assumed posterior distribution. Our first plot tracings of the Markov chains appear to be 'wandering' and not efficiently

sampling the the sample space for the posterior distribution.

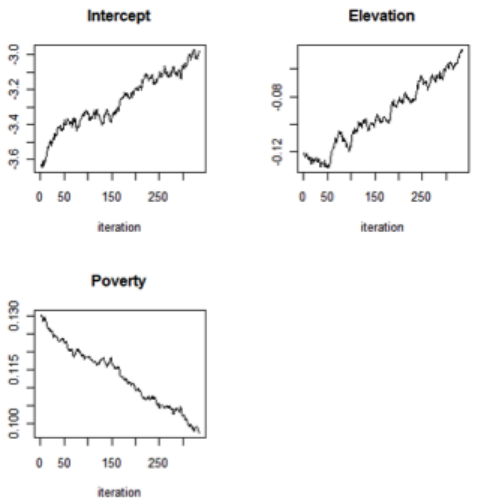

Figure 5: Sample tracings, before re-parameterization.

This issue of potential non-convergence prevents us from making valid inferences on the parameters. There are a few ways to address it. We can re-parameterize the problematic variables by centering them (subtracting the mean value), standardize them (dividing by the standard error) or transform them (taking the log of the value). We find that centering and standardizing the values returns plot tracings much more consistent with convergence. We can test this assessment through use of a statistic, such as the Brooks-Gellman-Rubin statistic, which in this case indicated adequate convergence.

Having fit the model, we now proceed to make inferences and apply the results to smoothing the mapped risk estimates and identifying areas of unusually low repatriation rates. We begin by plotting the probability distributions for our parameter estimates. These are probability density plots of the beta coefficients. That last statement should give you some pause, and actually represents perhaps the most profound difference between a Bayesian approach and the more commonly encountered frequentist or Pearson-Neyman approach to biostatistics. From these data, we can directly calculate the probability of a parameter value. [12]

[12]Take a moment to think about that, and ponder wither the p-value.

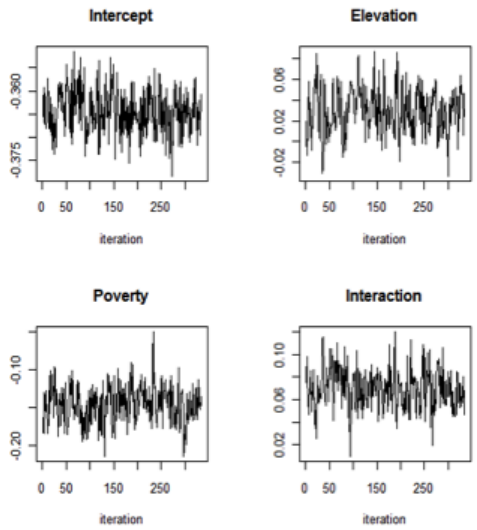

Figure 6: Sample tracings, after re-parameterization.

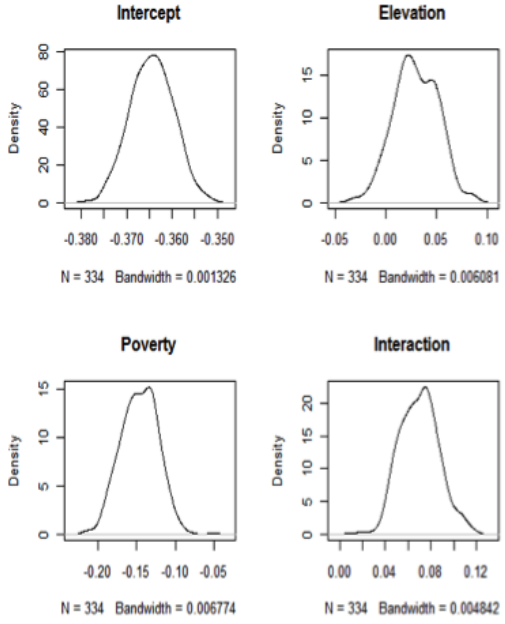

Figure 7: Kernel density plots of model parameters.

Based on these probability distributions, we can easily calculate the mean and upper and lower credible intervals at the 2.5% and 97.5 probability level for our

parameters.

We see these estimates in the following table.

Node (parameter)	mean	sd	MC error	2.5%	97.5%
Intercept	-0.3643	0.004737	1.777E-4	-0.3739	-0.3552
Elevation	0.02958	0.02042	9.978E-4	-0.00882	0.06738
Poverty	-0.1443	0.025	0.001272	-0.1913	-0.09626
PovertyElevation	0.07187	0.01735	8.302E-4	0.04037	0.1067

The MC error refers to the Monte Carlo error. It is the standard deviation divided by the number of sampling iterations, so is an indication of the increasing precision of our estimates as our sampling increases. As a rule of thumb, the Monte Carlo error should be 1% to 5% of the posterior standard deviation. We conclude from these results that the percentage of individuals below the poverty level was the most consistent predictor of repatriation following this disaster, even more so than elevation above sea level.

The CAR model also returns risk estimates for each census block in our data, conditioned on all the surrounding risk estimates. We can map these results as smoothed risk estimates.

Comparing this map to our previous un-smoothed map (Figure 2) we see there has been some 'shrinkage' toward local mean values.

With this admittedly healthy preamble, we turn our attention once again to traumatic brain injury in New York City to see if a Bayesian approach returns more valid or reliable results than those of non-Bayesian approaches.

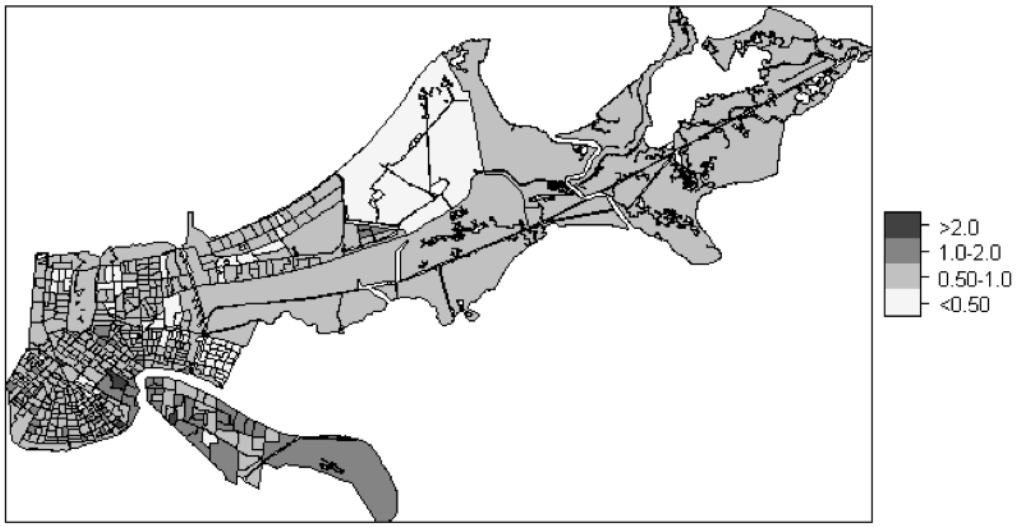

Figure 8: Bayesian Smoothed Risk Estimates Orleans Parish Census Block-Group Level Depopulation September 2008 vs. August 2005 Based on Number of Household Receiving Mail Delivery