

Lesson 19 (Actually Lesson 21...)

Clark

As we left off with last class, often we want to fit models that are more complicated than a shared effect across clusters. Perhaps we want to consider AR(1) correlated errors within our Poisson GLMM. That is, we might want to model:

In cases like this, it will often be beneficial to use Bayesian methods to conduct inference. I'm going to start today with some big picture ideas of Bayesian inference and then after spring break we will talk specifically correlated error structure inside of a GLMM.

To start, though, let's consider the goal of (parametric) statistics. Often, we use statistical methods in order to make inference about a parameter. To accomplish this we must assume that our data comes from a known density. For instance, we might say that $y \sim \text{Binom}(n, \pi)$ and use our data to estimate π . To formally conduct inference we would:

After we do this, we then might create confidence intervals and, if we remember from MA206/256, we always are careful to say something like "We are 95% confident..." or "If we repeat our experiment over and over again, approximately 95% of the time our CI will cover our true parameter". Or something similar to this.

In either case, perhaps it was beat into our heads to be very careful NOT to say "The probability that our parameter is in our CI is 95%." Why?

An alternative approach is Bayesian inference and, as it turns out, Bayesians have no issue saying “The probability that our parameter is in XX interval is 95%” If we remember Bayes rule, it stated that:

The same concept can be applied to statistical inference. If we think of our probability distribution as a conditional probability, we can write:

Now, if we have multiple observations, this becomes:

While this might look complicated, in many instances it’s actually very straight forward after we pick an appropriate *prior distribution*. For instance we might have:

However, in many instances we might want to pick a prior distribution that does not lead to a convenient posterior distribution. In those cases, we would need to simulate or estimate the posterior (We will discuss this more next lesson).

I want to go back to linear regression now, let's write out a model and consider what prior distributions we have to pick:

As it turns out, we don't have to stop there. We could add parameters to our priors and estimate what are called hyper-parameters... Just as a quick aside, if we have a regression model $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$ and we place a prior on β of $\beta \sim N(0, \tau\mathbf{I})$ the resulting posterior mean is the same as conducting ridge regression that you have discussed likely in MA477.

The addition of another level of parameters often can allow us to create structured dependence in another way. Let's consider the following example (from Bayesian Data Analysis):

In the evaluation of drugs for possible clinical application, studies are routinely performed on rodents. For a particular study drawn from the statistical literature, suppose the immediate aim is to estimate θ , the probability of tumor in a population of female laboratory rats that receive a zero dose of the drug. The data show that 4 out of 14 rats developed polyps. However, this isn't the first experiment that has been conducted, perhaps all the data look like:

```
y <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,
      2,1,5,2,5,3,2,7,7,3,3,2,9,10,4,4,4,4,4,4,4,10,4,4,4,5,11,12,
      5,5,6,5,6,6,6,6,16,15,15,9,4)
n <- c(20,20,20,20,20,20,20,19,19,19,19,18,18,17,20,20,20,20,19,19,18,18,25,24,
      23,20,20,20,20,20,20,10,49,19,46,27,17,49,47,20,20,13,48,50,20,20,20,
      20,20,20,48,19,19,19,22,46,49,20,20,23,19,22,20,20,20,52,46,47,24,14)
```

Let's think through different ways to model this data.

If we take the hierarchical approach, our entire posterior is found through:

Here the primary interest is likely on our hyper-parameters, but we can also find $\theta^*|\alpha, \beta, y$, that is, likely values for our next experiment. To find the distribution of $\alpha, \beta|y$ we could do:

Now, usually this isn't helpful, but for a large class of GLMMs it turns out that this can be approximated to make this possible to do (more to follow when we talk about INLA next lesson).

Finally, I want to talk about how we build up regression models, then we can discuss how we create dependence inside of these models. Let's consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Recall, this is a GLM with $y_i \sim N(\mu_i, \sigma^2)$ What parameters do we need to put prior distributions on? We can draw this diagram as:

Here, there's some possible choices of priors that are considered standard, uninformative priors. We could perhaps, extend this to now consider an AR(1) structure in the model as maybe our data are observed over time?

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + z_t + \epsilon_t \\ z_t | z_{t-1} &\sim N(\phi z_{t-1}, \sigma_z^2) \\ \epsilon_t &\sim N(0, \sigma_e^2) \end{aligned}$$

Let's tie this concept together with hierarchical models that we previously considered and also put this into a GLM framework.

Let's consider the number of lip cancer cases in Scotland in the years 1975-1980 at the county level.

```
library(SpatialEpi)
data(scotland)
map <- scotland$spatial.polygon
plot(map)
```



Let's think through how we could go about modeling this.

Note that this model implies a hierarchical model for β_0 . We could also write:

Note that this is what's considered an epidemiological model. One way to account for population is to factor in the expected number of cases. That is:

This, then, gets handled similarly to an offset. Again, our full model is:

Perhaps we want to account for the spatial structure that is NOT accounted for in $X\beta$? One model could be: