Hierarchical models for areal data

- Areal data: often aggregate outcomes over a well-defined area. Example: number of cancer cases per county or proportion of people living in poverty in a set of census tracks.

- What are the inferential issues?
  (a) Identifying a spatial pattern and its strength. If data are spatially correlated, measurements from areas that are ‘close’ will be more alike.
  (b) Smoothing and to what degree. Observed measurements often present extreme values due to small samples in small areas. Maximal smoothing: substitute observed measurements by the overall mean in the region. Something less extreme is what we discuss later.
  (c) Prediction: for a new area, how would we predict $Y$ given measurements in other areas?

Defining neighbors

- A proximity matrix $W$ with entries $w_{ij}$ spatially connects areas $i$ and $j$ in some fashion.
- Typically, $w_{ii} = 0$.
- There are many choices for the $w_{ij}$:
  - Binary: $w_{ij} = 1$ if areas $i, j$ share a common boundary, and is 0 otherwise.
  - Continuous: decreasing function of intercentroidal distance.
  - Combo: $w_{ij} = 1$ if areas are within a certain distance.
- $W$ need not be symmetric.
- Entries are often standardized by dividing into $\sum_j w_{ij} = w_{ii+}$. If entries are standardized, the $W$ will often be asymmetric.
Defining neighbors (cont’d)

• The \( w_{ij} \) can be thought of as weights and provide a means to introduce spatial structure into the model.

• Areas that are 'closer by' in some sense are more alike.

• For any problem, we can define first, second, third, etc order neighbors. For distance bins \((0, d_1], (d_1, d_2], (d_2, d_3], \ldots\) we can define
  
  (a) \( W^{(1)} \), the first-order proximity matrix with \( w_{ij}^{(1)} = 1 \) if distance between \( i \) and \( j \) is less than \( d_1 \).

  (b) \( W^{(2)} \), the second-order proximity matrix with \( w_{ij}^{(2)} = 1 \) if distance between \( i \) and \( j \) is more than \( d_1 \) but less than \( d_2 \).

Areal data models

• Because these models are used mostly in epidemiology, we begin with an application in disease mapping to introduce concepts.

• Typical data:
  
  \( Y_i \) observed number of cases in area \( i, i = 1, \ldots, I \)

  \( E_i \) expected number of cases in area \( i \).

• The \( Y \)'s are assumed to be random and the \( E \)'s are assumed to be known and to depend on the number of persons \( n_i \) at risk.
Areal data models (cont’d)

- An *internal* standardized estimate of $E_i$ is
  \[
  E_i = n_i \bar{r} = n_i \left( \frac{\sum_j y_i}{\sum_j n_i} \right),
  \]
corresponding to a constant disease rate across areas.

- An *external* standardized estimate is
  \[
  E_i = \sum_j n_{ij} r_j,
  \]
  where $r_j$ is the risk for persons of age group $j$ (from some existing table of risks by age) and $n_{ij}$ is the number of persons of age $j$ in area $i$.

Standard frequentist approach

- For small $E_i$,
  \[
  Y_i | \eta_i \sim \text{Poisson}(E_i \eta_i),
  \]
  with $\eta_i$ the true relative risk in area $i$.

- The MLE is the *standard mortality ratio*
  \[
  \hat{\eta}_i = SMR_i = \frac{Y_i}{E_i},
  \]

- The variance of the $SMR_i$ is
  \[
  \text{var}(SMR_i) = \frac{\text{var}(Y_i)}{E_i^2} = \frac{\eta_i}{E_i},
  \]
  estimated by plugging $\hat{\eta}_i$ to obtain
  \[
  \text{var}(SMR_i) = \frac{Y_i}{E_i^2}.
  \]
**Frequentist approach (cont’d)**

- To get a confidence interval for $\eta_i$, first assume that $\log(SMR_i)$ is approximately normal.
- From a Taylor expansion:
  \[
  \text{Var}[\log(SMR_i)] \approx \frac{1}{SMR_i^2} \text{Var}(SMR_i) = \frac{E_i^2}{Y_i^2} \times \frac{Y_i}{E_i} = \frac{1}{Y_i}.
  \]
- An approximate 95% CI for $\log(\eta_i)$ is
  \[SMR_i \pm 1.96/(Y_i)^{1/2}.
  \]
- Transforming back, an approximate 95% CI for $\eta_i$ is
  \[(SMR_i \exp(-1.96/(Y_i)^{1/2}), \ SMR_i \exp(1.96/(Y_i)^{1/2})).\]

**Frequentist approach (cont’d)**

- Suppose we wish to test whether risk in area $i$ is high relative to other areas. Then test
  \[H_0 : \eta_i = 1 \text{ versus } H_a : \eta_i > 1.
  \]
- This is a one-sided test.
- Under $H_0$, $Y_i \sim \text{Poisson}(E_i)$ so the $p-$value for the test is
  \[
p = \text{Prob}(X \geq Y_i | E_i) = 1 - \text{Prob}(X \geq Y_i | E_i) = 1 - \sum_{x=0}^{Y_i-1} \frac{(-E_i)^x E_i^x}{x!}.
  \]
- If $p < 0.05$ we reject $H_0$. 

\[7\]

\[8\]
Hierarchical models for areal data

- To estimate and map underlying relative risks, might wish to fit a random effects model.
- Assumption: true risks come from a common underlying distribution.
- Random effects models permit borrowing strength across areas to obtain better area-level estimates.
- Alas, models may be complex:
  - High-dimensional: one random effect for each area.
  - Non-normal if data are counts or binomial proportions.
- We have already discussed hierarchical Poisson models, so material in the next few transparencies is a review.

Poisson-Gamma model

- Consider
  \[ Y_i | \eta_i \sim \text{Poisson}(E_i \eta_i) \]
  \[ \eta_i | a, b \sim \text{Gamma}(a, b). \]
- Since \( E(\eta_i) = a/b \) and \( Var(\eta_i) = a/b^2 \), we can fix \( a, b \) as follows:
  - A priori, let \( E(\eta_i) = 1 \), the null value.
  - Let \( Var(\eta_i) = (0.5)^2 \), large on that scale.
- Resulting prior is \( \text{Gamma}(4, 4) \).
- Posterior is also Gamma:
  \[ p(\eta_i | y_i) = \text{Gamma}(y_i + a, E_i + b). \]
Poisson-Gamma model (cont’d)

• A point estimate of $\eta_i$ is
  
  \[ E(\eta_i|y_i) = E(\eta_i|y_i) = \frac{y_i + a}{E_i + b} = \frac{y_i + \frac{[E(\eta_i)]^2}{\text{Var}(\eta_i)}}{E_i + \frac{E(\eta_i)}{\text{Var}(\eta_i)}} = \frac{E_i(\frac{y_i}{E_i})}{E_i + \frac{E(\eta_i)}{\text{Var}(\eta_i)}} + \frac{E(\eta_i)}{E_i + \frac{E(\eta_i)}{\text{Var}(\eta_i)}} = w_iSMR_i + (1 - w_i)E(\eta_i), \]

  where $w_i = E_i/[E_i + (E(\eta_i)/\text{Var}(\eta_i))]$.

• Bayesian point estimate is a weighted average of the data-based $SMR_i$ and the prior mean $E(\eta_i)$.

Example

• In area $i$ we observe $y_i = 27$, and $E_i = 21$ were expected.

• Using Gamma(4, 4) as the prior, the posterior is Ga(31, 25).

• See figure: 1000 draws from prior and posterior.

• Posterior has mean $1.24 = 31/25$ so estimated risk is 24% higher than expected.

• Probability that risk is higher than 1 is $\Pr(\eta_i > 1|y_i) = 0.863$ so there is substantial but not definitive evidence that risk is area $i$ is elevated.

• The 95% credible interval for $\eta_i$ is $(0.842, 1.713)$ which covers 1, so we do not conclude that risk in area $i$ is significantly different from what we expected.

• Covariates could be incorporated into the Poisson-Gamma model following the approach of Christensen and Morris (1997) that we discussed earlier.
Poisson-lognormal models with spatial errors

- The Poisson-Gamma model does not allow (easily) for spatial correlation among the $\eta_i$.
- Instead, consider the Poisson-lognormal model, where in the second stage we model the log-relative risks $\log(\eta_i) = \psi_i$:

$$Y_i | \psi_i \sim \text{Poisson}(E_i \exp(\psi_i))$$

$$\psi_i = x_i' \beta + \theta_i + \phi_i,$$

where $x_i$ are area-level covariates.

- The $\theta_i$ are assumed to be exchangeable and model between-area variability:

$$\theta_i \sim N(0, 1/\tau_h).$$

- The $\theta_i$ incorporate global extra-Poisson variability in the log-relative risks (across the entire region).

Poisson-lognormal model (cont’d)

- The $\phi_i$ are the ’spatial’ parameters; they capture regional clustering.
- They model extra-Poisson variability in the log-relative risks at the local level so that ’neighboring’ areas have similar risks.
- One way to model the $\phi_i$ is to proceed as in the point-referenced data case. For $\phi = (\phi_1, ..., \phi_I)$, consider

$$\phi | \mu, \lambda \sim N_I(\mu, H(\lambda)),$$

and $H(\lambda)_{ii'} = \text{cov}(\phi_i, \phi_{i'})$ with hyperparameters $\lambda$.

- Possible models for $H(\lambda)$ include the exponential, the powered exponential, etc.
- While sensible, this model is difficult to fit because
  - Lots of matrix inversions required
  - Distance between $\phi_i$ and $\phi_{i'}$ may not be obvious.
CAR model

- More reasonable to think of a neighbor-based proximity measure and consider a conditionally autoregressive model for $\phi$:
  $$\phi_i \sim N(\bar{\phi}_i, 1/(\tau_c m_i)),$$
  where
  $$\bar{\phi}_i = \sum_{i \neq j} w_{ij}(\phi_i - \phi_j),$$
  and $m_i$ is the number of neighbors of area $i$. Earlier we called this $w_{ij}$.

- The weights $w_{ij}$ are (typically) 0 if areas $i$ and $j$ are not neighbors and 1 if they are.

- CAR models lend themselves to the Gibbs sampler. Each $\phi_i$ can be sampled from its conditional distribution so no matrix inversion is needed:
  $$p(\phi_i | \text{all}) \propto \text{Poi}(y_i | E_i e^{x_i \beta + \theta_i + \phi_i}) N(\phi_i | \bar{\phi}_i, \frac{1}{m_i \tau_c}).$$

Difficulties with CAR model

- The CAR prior is improper. Prior is a pairwise difference prior identified only up to a constant.

- The posterior will still be proper, but to identify an intercept $\beta_0$ for the log-relative risks, we need to impose a constraint: $\Sigma_i \phi_i = 0$.

- In simulation, constraint is imposed numerically by recentering each vector $\phi$ around its own mean.

- $\tau_h$ and $\tau_c$ cannot be too large because $\theta_i$ and $\phi_i$ become unidentifiable. We observe only one $Y_i$ in each area yet we try to fit two random effects. Very little data information.
Difficulties with CAR (cont’d)

• Hyperpriors for $\tau_h$, $\tau_c$ need to be chosen carefully.

• Consider

$$\tau_h \sim \text{Gamma}(a_h, b_h), \quad \tau_c \sim \text{Gamma}(a_c, b_c).$$

• To place equal emphasis on heterogeneity and spatial clustering, it is tempting to make $a_h = a_c$ and $b_h = b_c$. This is not correct because

(a) The $\tau_h$ prior is defined marginally, where the $\tau_c$ prior is conditional.

(b) The conditional prior precision is $\tau_cm_i$. Thus, a scale that satisfies

$$sd(\theta_i) = \frac{1}{\sqrt{\tau_h}} \approx \frac{1}{0.7\sqrt{\overline{m}\tau_c}} \approx sd(\phi_i)$$

with $\overline{m}$ the average number of neighbors is more 'fair' (Bernardinelli et al. 1995, *Statistics in Medicine*).