Bayesian Statistics

Generalized Linear Mixed Effects Model

Nan Lin

Department of Mathematics
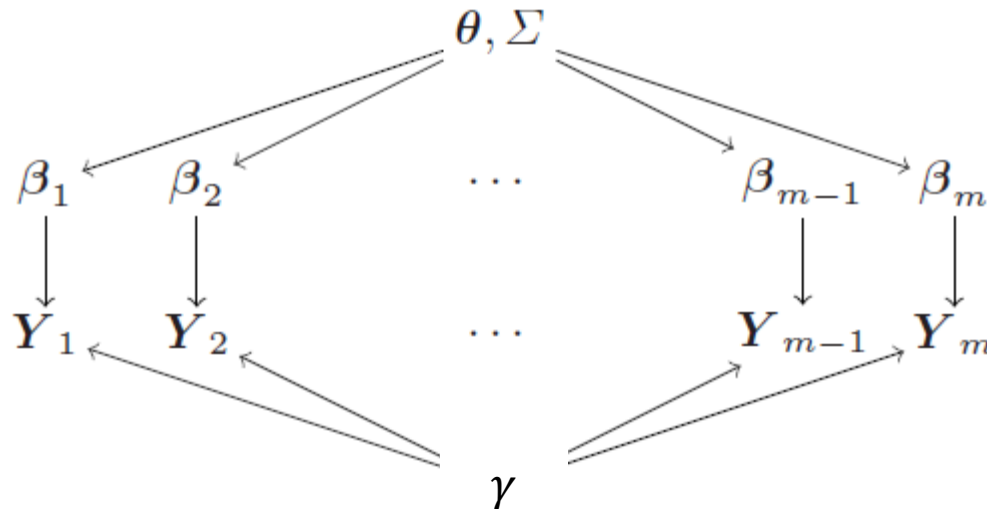
Washington University in St. Louis

Generalized Linear Mixed Effects Model (GLMM)

▸ **Extension of Linear mixed effects model to non-normal response**

  ▸ e.g. binary or count

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \sim \text{i.i.d. multivariate normal}(\boldsymbol{\theta}, \Sigma)$$

$$p(\boldsymbol{y}_j | \mathbf{X}_j, \boldsymbol{\beta}_j, \gamma) = \prod_{i=1}^{n_j} p(y_{i,j} | \boldsymbol{\beta}_j^T \boldsymbol{x}_{i,j}, \gamma),$$

Metropolis-within-Gibbs

▸ For simplicity, we assume no γ parameter (such as in logistic and Poisson regression)

▸ Full conditional distributions of (θ, Σ) can be standard under proper prior specifications

  ▸ Prior

$$\boldsymbol{\theta} \sim \text{multivariate normal}(\boldsymbol{\mu}_0, \Lambda_0)$$
$$\Sigma \sim \text{inverse-Wishart}(\eta_0, \mathbf{S}_0^{-1})$$

  ▸ Full conditional distributions are of the same type

▸ $\{\beta_j\}$ requires Metropolis updating

  ▸ We consider updating $\{\beta_j\}$ sequentially one at a time

Metropolis updating of $\{\beta_j\}$

▸ Consider a multivariate normal proposal distribution with mean as the current value $\beta_j^{(s)}$ with proposal variance $V_j^{(s)}$

  ▸ A common choice is to set $V_j^{(s)} = \text{scale} \times \Sigma^{(s)}$

1. Sample $\boldsymbol{\beta}_j^* \sim \text{multivariate normal}(\boldsymbol{\beta}_j^{(s)}, V_j^{(s)})$.
2. Compute the acceptance ratio

$$r = \frac{p(\boldsymbol{y}_j|\mathbf{X}_j, \boldsymbol{\beta}_j^*)p(\boldsymbol{\beta}_j^*|\boldsymbol{\theta}^{(s)}, \Sigma^{(s)})}{p(\boldsymbol{y}_j|\mathbf{X}_j, \boldsymbol{\beta}_j^{(s)})p(\boldsymbol{\beta}_j^{(s)}|\boldsymbol{\theta}^{(s)}, \Sigma^{(s)})}.$$

3. Sample $u \sim \text{uniform}(0,1)$. Set $\boldsymbol{\beta}_j^{(s+1)}$ to $\boldsymbol{\beta}_j^*$ if $u < r$ and to $\boldsymbol{\beta}_j^{(s)}$ if $u > r$.

Putting together

1. Sample $\boldsymbol{\theta}^{(s+1)}$ from its full conditional distribution.
2. Sample $\Sigma^{(s+1)}$ from its full conditional distribution.
3. For each $j \in \{1, \ldots, m\}$,
    a) propose a new value $\boldsymbol{\beta}_j^*$;
    b) set $\boldsymbol{\beta}_j^{(s+1)}$ equal to $\boldsymbol{\beta}_j^*$ or $\boldsymbol{\beta}_j^{(s)}$ with the appropriate probability.

Example: tumor location data*

▶ A certain population of laboratory mice experiences a high rate of intestinal tumor growth. One item of interest to researchers is how the rate of tumor growth varies along the length of the intestine.

▶ Study: the intestine of each of 21 sample mice was divided into 20 sections and the number of tumors occurring in each section was recorded.

▶ Code the 20 locations as $x \in (0.05, 0.10, \ldots, 0.95)$
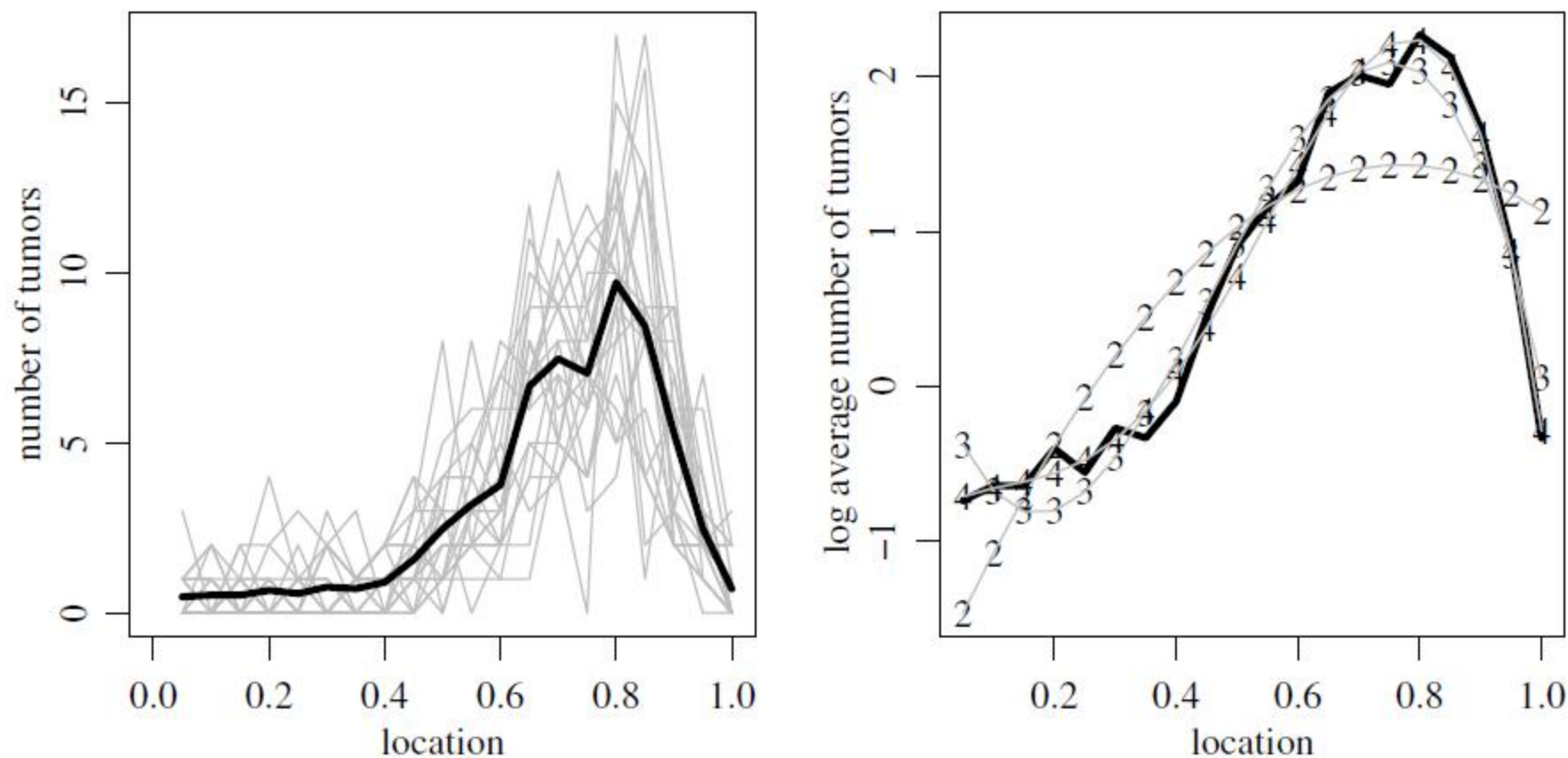
▶ Response: # of tumors

Fig. 11.4. Tumor count data. The first panel gives mouse-specific tumor counts as a function of location in gray, with a population average in black. The second panel gives quadratic, cubic and quartic polynomial fits to the log sample average tumor count.

Hierarchical Poisson regression

▸ Let $Y_{x,j}$ be mouse j's tumor count at location x of their intestine

▸ Level 1:

  ▸ Assume $Y_{x,j} \sim Poisson(e^{f_j(x)})$

  ▸ $f_j(x) = \beta_{1,j} + \beta_{2,j}x + \cdots + \beta_{p,j}x^{p-1}$, use $p = 5$

▸ Level 2:

$$\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \sim \text{i.i.d. multivariate normal}(\boldsymbol{\theta}, \Sigma)$$

▸ Level 3:

$$\boldsymbol{\theta} \sim \text{multivariate normal}(\boldsymbol{\mu}_0, \Lambda_0)$$
$$\Sigma \sim \text{inverse-Wishart}(\eta_0, \mathbf{S}_0^{-1})$$

Choosing the hyper-prior

▸ Unit information prior

  ▸ the prior distributions for the parameters are weakly centered around estimates derived from the observed data

▸ Get a rough estimate $\tilde{\beta}_j$ by fitting a polynomial regression using data from mouse j only

  ▸ Regress $\left\{ \log\left(y_{1,j} + \frac{1}{n}\right), \dots, \log\left(y_{n,j} + \frac{1}{n}\right) \right\}$ on $\{x_1, \dots, x_{20}\}$

▸ Set $\mu_0 = \frac{1}{m} \sum_j \tilde{\beta}_j$ and Λ_0 as the sample covariance matrix from $\tilde{\beta}_j$'s

▸ Set S_0 also as the sample covariance matrix from $\tilde{\beta}_j$'s

▸ $\eta_0 = p + 2 = 7$

MCMC

▸ Proposal distribution for β_j: multivariate normal $(\beta_j^{(s)}, \Sigma^{(s)}/2)$

  ▸ Acceptance rate is around 31%

▸ 50,000 scans

▸ Thinning: every 10th scan

```
## data
data(chapter11)
Y<-XY.tumor$Y ; X<-XY.tumor$X ; m<-dim(Y)[1] ; p<-dim(X)[2]

## priors
BETA<-NULL
for(j in 1:m)
{
  BETA<-rbind(BETA,lm(log(Y[j,]+1/20)~-1+X[,,j])$coef)
}

mu0<-apply(BETA,2,mean)
S0<-cov(BETA) ; eta0<-p+2
iL0<-iSigma<-solve(S0)

## MCMC
THETA.post<-NULL ; set.seed(1)
for(s in 1:50000)
{
  ##update theta
  Lm<-solve( iL0 + m*iSigma )
  mum<-Lm%*%( iL0%*%mu0 + iSigma%*%apply(BETA,2,sum) )
  theta<-t(rmvnorm(1,mum,Lm))
  ##

  ##update Sigma
  mtheta<-matrix(theta,m,p,byrow=TRUE)
  iSigma<-rwish(1,eta0+m,
          solve( S0+t(BETA-mtheta)%*%(BETA-mtheta)) )
  ##
```

```
##update beta
Sigma<-solve(iSigma)  ;  dSigma<-det(Sigma)
for(j in 1:m)
{
   beta.p<-t(rmvnorm(1,BETA[j,],.5*Sigma))

   lr<-sum( dpois(Y[j,],exp(X[,,j]%*%beta.p),log=TRUE ) -
        dpois(Y[j,],exp(X[,,j]%*%BETA[j,]),log=TRUE ) ) +
      ldmvnorm( t(beta.p),theta,Sigma,
           iSigma=iSigma,dSigma=dSigma ) -
      ldmvnorm( t(BETA[j,]),theta,Sigma,
           iSigma=iSigma,dSigma=dSigma )

   if( log(runif(1))<lr ) { BETA[j,]<-beta.p }
 }
##

##store some output
 if(s%%10==0){THETA.post<-rbind(THETA.post,t(theta))}
 ##

}
```
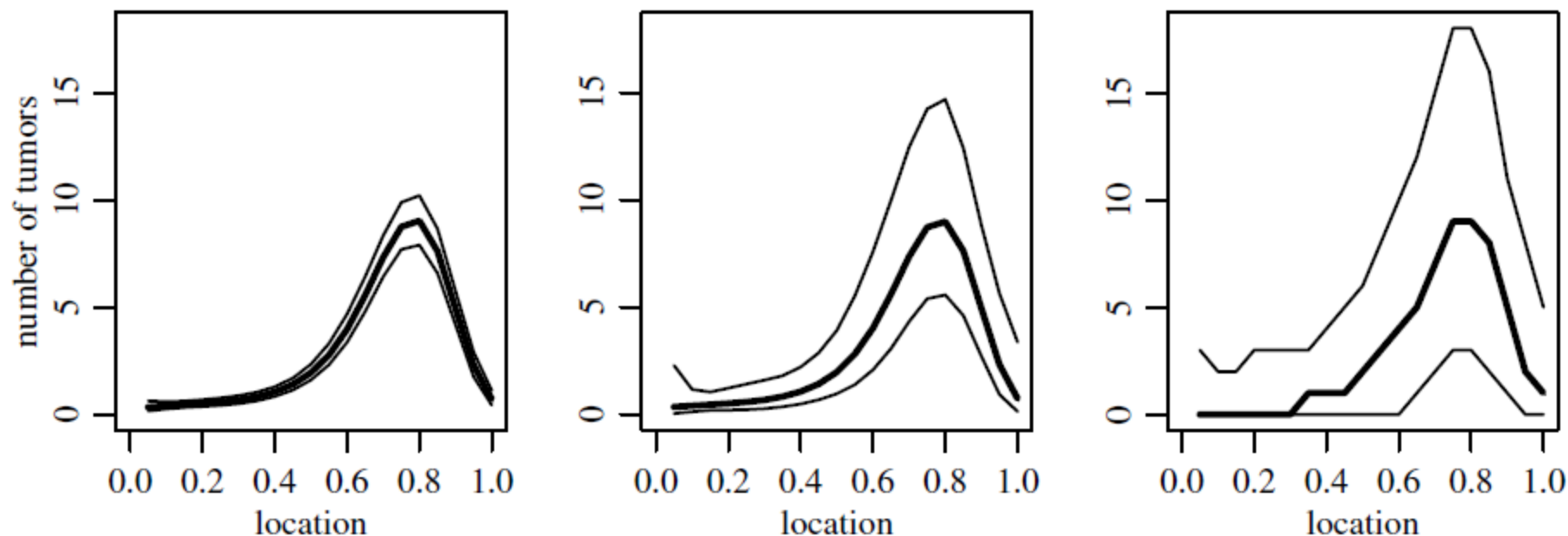
Results

Fig. 11.5. 2.5, 50 and 97.5% quantiles for $\exp(\boldsymbol{\theta}^T \boldsymbol{x})$, $\exp(\boldsymbol{\beta}^T \boldsymbol{x})$ and $\{Y|\boldsymbol{x}\}$.

The widening confidence bands of the three plots in this figure describe cumulative sources of uncertainty
- Uncertainty in theta + across-mouse heterogeneity + fluctuation in the response around its mean

Discussion

- Posterior approximation via MCMC for hierarchical models can suffer from poor mixing.

  - Many of the parameters in the model are highly correlated, and generating them one at a time in the Gibbs sampler can lead to a high degree of autocorrelation

- R package `MCMCglmm`

  - Tutorial of the package