STAT 770 Dec. 2 Lecture 26

Bayesian R Calculations with GLMM

Reading and Topics for this lecture: Sections 13.6–13.7, `arm` and `bayesglm` documentation; and `BayesGLMM.RLog` script.

(1) Bayesian GLM analysis within `R` packages

(2) Is there a Bayesian GLMM ?

(3) Posterior-density Calculations in Logit-Normal Settings

(4) Simulated and "Small Area" Data examples

Bayesian Approach to GLM (cluster-level Logistic)

Consider logit-normal model, restricted to case where all Y_{ij} are binary, cluster i has n_i unit observations, and $\underline{X}_{ij} \equiv \underline{X}_i$.

Then as before, `logLik` $\log L_i(\beta)$ for i'th cluster given ϵ_i is

$$\log L_i(\beta, \underline{\epsilon}, \underline{Y}) = Y_{i+} \, (\underline{X}_i^{tr} \, \beta + \epsilon_i) - n_i \log \Big(1 + \exp(\underline{X}_i^{tr} \, \beta + \epsilon_i)\Big)$$

For the Bayesian, the ϵ_i and β are all unknowns with priors.

$$\pi(\beta, \, \underline{\epsilon} \, | \, \underline{Y}) \propto \pi_0(\beta, \, \underline{\epsilon}) \prod_{i=1}^{m} L_i(\beta, \underline{\epsilon}, \underline{Y})$$

All further special features of parameters and models must be expressed through the choice of prior.

General Comments about Bayesian Monte Carlo

Setting is always: unknown $\theta \sim \pi(t)$, data $\underline{Y}$, model $f_{\underline{Y}|\theta}(\underline{y}\,|\,t)$

Calculate posterior $\pi(\theta\,|\,\underline{Y}) = \pi(\theta)\, f_{\underline{Y}|\theta}(\underline{y}\,|\,t) \Big/ f_{\underline{Y}}(\underline{Y})$ if possible

Otherwise, simulate a long sequence $\{\theta_k\}_{k=0}^{B+T}$ from a stochastic (Markovian) process, with limiting stationary distribution (after B **burn-in** steps), $\theta_k \xrightarrow{\mathcal{D}} \pi(\cdot\,|\,\underline{Y})$ as $k \to \infty$

Perform statistical calculations by estimating:

$$E(h(\theta)\,|\,\underline{Y}) \approx \frac{1}{T} \sum_{k=B+1}^{B+T} h(\theta_k)$$

Bayesian GLMM ?

So a first Bayesian analysis might be to take $\beta \in \mathbb{R}^p$ from a prior (often with coordinates independent with large scale parameter) and ϵ_i *iid* from a specified $\mathcal{N}(0, \sigma_0^2)$. The `R` package `arm` with function `sim` allows us to generate as many approximate 'samples' as desired from the posterior $\pi(\beta, \underline{\epsilon} \,|\, \underline{Y})$.

With $\epsilon_i \equiv 0$ $(\sigma_0 = 0)$, have Bayes approach to fixed-effect GLM.

The 'samples' are approximately from the posterior of the indicated model, which is fine in fixed-effect GLM whebn $\sigma_0 = 0$, but this is a different model from the random-intercept model parameterized by (β, σ). But the function `sim` can also do that simulation, as indicated in `BayesGLMM.RLog` script.

Logit-Normal Posterior Calculations

Items of interest concerning posterior:

• posterior means and variances of cluster effects

• distributional characteristics (quantiles, for CIs) of parameter estimates and predictors (such as random-effect predictors) – for this, also need posterior-simulated σ output sequences

• maximum absolute random-intercepts could also be used to search for outlier-clusters

• various correlations between parameters and predictors might also be useful

Small-Area Example

The example using Table 13.2 will also be analyzed this way at the end of the `BayesGLMM.RLog` script.

This is the way you could do HW problems. If I find a simple way to get richer Bayesian MCMC output from an R package, without going deeply into WinBugs syntax, I will cover that in next Monday's lecture.

Otherwise we go on to a different topic next week: tree-based and "Machine Learning" classifiers to compete with logistic regression.