STAT 578: Advanced Bayesian Modeling

Week 4 – Lesson 3

Hierarchical Models: More Details

Fall 2019

Partial Conjugacy

Review

Recall conjugacy: Posterior is from same family of distributions as prior.

Many standard sampling distributions have natural conjugate priors in non-hierarchical models (will discuss later).

For hierarchical models, exact natural conjugacy is usually impossible (even in simple cases, like hierarchical normal model).

Partial Conjugacy

Even if natural conjugacy is impossible, something similar may be possible:

Terms like **partial conjugacy**, **conditional conjugacy**, and **semi-conjugacy** refer vaguely to the case that the distribution specified for a parameter in the prior is in the same family as its conditional posterior.

That is, if

$$\theta = (\theta_1, \theta_2)$$

and θ_1 has a (possibly conditional) prior in the DAG model from a given family, then the conditional posterior $p(\theta_1 \mid \theta_2, y)$ is also from that same family.

Example

For the hierarchical normal model (with known variances), recall

$$\theta_1, \ldots, \theta_J \mid \mu, \tau \quad \sim \quad \text{iid } \mathrm{N}(\mu, \tau^2)$$

led to

$$\theta_1, \ldots, \theta_J \mid \mu, \tau, y \quad \sim \quad \text{indep. } \mathrm{N}(\hat{\theta}_j, V_j)$$

reflecting the fact that the normal family is a partially conjugate for each θ_j.

(In fact, independent – *not* iid – normals form a partially conjugate family for the θ_js jointly.)

Can often discover partial conjugacy by examining form of joint density of model.

By conditioning and using Bayes' rule:

$$p(\theta_1 \mid \theta_2, y)\, p(\theta_2 \mid y) \;=\; p(\theta_1, \theta_2 \mid y) \;\propto\; p(\theta_1, \theta_2)\, p(y \mid \theta_1, \theta_2)$$

and so

$$p(\theta_1 \mid \theta_2, y) \;\propto\; p(\theta_1, \theta_2)\, p(y \mid \theta_1, \theta_2)$$

where the proportionality is in θ_1 only.

So if the joint model density has a recognizable form in θ_1 that is the same as the (possibly conditional) prior in θ_1, the family of that form is partially conjugate.

Example

Hierarchical binomial sampling model:

$$y_j \mid \theta_j \;\sim\; \text{indep. Bin}(n_j, \theta_j) \qquad\qquad \theta_j \mid \alpha, \beta \;\sim\; \text{iid Beta}(\alpha, \beta)$$

Joint model density proportional to

$$p(\alpha, \beta) \cdot \prod_{j=1}^{J} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \cdot \theta_j^{y_j}(1-\theta_j)^{n_j-y_j} \;\propto\; \prod_{j=1}^{J} \theta_j^{\alpha+y_j-1}(1-\theta_j)^{\beta+n_j-y_j-1}$$

(proportional in θ_js only, $0 < \theta_j < 1$).

So independent betas form conditional posterior of θ_js, like the conditional prior.

Some Natural Conjugate Families

Sampling Distribution		Parameter	Natural Conjugate
Binomial:	$\mathrm{Bin}(n, \theta)$	θ	beta distribution
Normal:	$\mathrm{N}(\mu, \sigma^2)$	μ	normal distribution
Normal:	$\mathrm{N}(\mu, \sigma^2)$	σ^2	inverse gamma distribution (scaled inverse chi-square)
Poisson:	$\mathrm{Poisson}(\lambda)$	λ	gamma distribution

Warning: Choosing priors based on (partial) conjugacy does not always lead to a good model (e.g., 2016 polls example).