22S:138
Bayesian Statistics

Introduction to Multi-Parameter Models

Lecture 9
Sept. 20, 2006

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Multiparameter models

- Real problems in statistics nearly always involve more than one unknown quantity.
- However, usually only one, or a few, parameters or predictions are of substantive interest.
- Example: newt healing rates
  - We may be primarily interested in the population mean healing rate μ, but of course we don't really know the value of the population variance σ^2.
  - So in a realistic model, we must also treat σ^2 as an unknown parameter.
- "nuisance parameters"

- In cases of this kind, the aim of Bayesian analysis is to obtain the posterior *marginal* distribution of the parameter(s) of interest, e.g.

$$p(\mu|\mathbf{y})$$

- The general approach is to estimate the *joint* posterior distribution of all unknown quantities in the model, and then integrate out the one(s) we aren't interested in.
- Example: in normal means example, we will find

$$p(\mu,\ \sigma^2\ |\mathbf{y})$$

then

$$p(\mu \mid \mathbf{y}) \ = \ \int p(\mu,\ \sigma^2\ |\mathbf{y})\ d\sigma^2$$

Example: normal data with both μ and σ^2 unknown

- Need joint prior on both unknown parameters.
- Consider first the conventional noninformative prior for this problem

$$p(\mu, \sigma^2) \ \propto \ \frac{1}{\sigma^2}$$

- This arises by considering μ and σ^2 *a priori* independent and taking the product of the standard noninformative priors for each.
  - *A priori* independence may be a reasonable assumption here; it says that if we knew something about one of the unknown parameters, that wouldn't give us information about the distribution of the other one.

- Recall standard noninformative priors for μ when σ^2 is assumed known, and for σ^2 when μ is assumed known.

- This is not quite a conjugate prior; we will see that the posterior distribution does not factor like this into an inverse gamma times an independent normal.
- Note that this prior is improper, and the joint posterior is improper if there are fewer than two observations in the current data.

Joint posterior distribution with conventional noninformative prior

- The joint posterior is

$$\begin{aligned} p(\mu,\ \sigma^2) &\propto \frac{1}{\sigma^2} \ \times \ \frac{1}{(\sigma^2)^{\frac{n}{2}}}\ exp(\left(-\frac{1}{2\ \sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right) \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} exp\left(-\frac{1}{2\ \sigma^2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 \ + \ n(\bar{y}-\mu)^2\right]\right) \\ &= \frac{1}{(\sigma^2)^{\frac{n}{2}+1}} exp\left(-\frac{1}{2\ \sigma^2}\left[(n-1)\ s^2 \ + \ n(\bar{y}-\mu)^2\right]\right) \end{aligned}$$

where s^2 is the sample variance of the y_is:

$$s^2 \ = \ \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

- $\bar{y}$ and s^2 are the sufficient statistics for μ and σ^2.

Steps to the marginal posterior distribution of μ

- We will use these identities from conditional probability

$$\begin{aligned} p(\mu \mid \mathbf{y}) &= \int p(\mu,\ \sigma^2\ |\mathbf{y})\ d\sigma^2 \\ &= \int p(\mu \mid \sigma^2,\ \mathbf{y}) p(\sigma^2|\mathbf{y}) d\sigma^2 \end{aligned}$$

- It can be shown by direct integration (GCSR p. 75) that the marginal posterior distribution of σ^2 is

$$p(\sigma^2|\mathbf{y}) \ \propto \ \frac{1}{(\sigma^2)^{\frac{n+1}{2}}} exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)$$

- What parametric density is this?

The conditional posterior distribution of μ given σ^2

- Use what we already know about the posterior mean of μ with *known* variance and a uniform prior on μ:

$$p(\mu \mid \sigma^2, \mathbf{y}) \ = \ N(\bar{y}, \frac{\sigma^2}{n})$$

- Again, it can be shown by direct integration (GCSR p. 68-69) that the *marginal* posterior distribution of μ

$$p(\mu \mid \mathbf{y}) \ = \ \int p(\mu \mid \sigma^2,\ \mathbf{y}) p(\sigma^2|\mathbf{y}) d\sigma^2$$

is a Student's t distribution with

- mean $\bar{y}$
- scale parameter $\frac{s^2}{n}$
- degrees of freedom $n-1$

Return of the newts

An informative semi-conjugate joint prior on μ and σ^2 in for the normal distribution

- An intuitive procedure for specifying a joint prior distribution $p(\mu,\ \sigma^2)$ if we had prior information on both is:
  - Assume *a priori* independence
  - Place an inverse gamma prior on σ^2
  - Place a normal prior on μ
  - Then the joint prior is the product of these two priors
- This is called a "semi-conjugate" prior. Why?

- However, it is *not* a conjugate prior!
- In fact, the marginal posterior distributions $p(\sigma^2|\mathbf{y})$ and $p(\mu|\mathbf{y})$ have no simple conjugate forms.

- We can find posterior means, variances, quantiles, etc. by numerical integration or simulation.
- We will use the WinBUGS program in lab to do this.

What are Markov chain Monte Carlo methods used for?

- to fit models that are too complex, high-dimensional, or otherwise wierd to fit by other methods
- especially frequently used for fitting Bayesian models

BUGS and WinBUGS are general-purpose packages that use Gibbs sampling to fit Bayesian models.

- constructs a Markov chain whose stationary distribution is the joint posterior of the unknowns (model parameters and missing data) of the specified model, conditional on the observed data
- exploits the fact that, under certain regularity conditions, this joint posterior distribution is the product of the "full conditional

distributions" of each unknown given all the other model quantities

- generates a sample path from the Markov chain
  - at each iteration, generates a realization of each unknown

What does the WinBUGS user have to input?

- model specification in terms of the distributional relationships between observables and parameters
  - distributions of observables as functions of parameters (likelihood)
  - prior distributions of parameters
- auxiliary files containing
  - data
  - initial values for unknowns

WinBUGS output is *samples*.

- correlated
- of quantities user has requested WinBUGS to "monitor"
  - parameters
  - missing data
  - functions of either of these