Conjugate Priors

Hierarchical Models

Comparing Ranks through Simulation

Conjugate Priors

- General principle: choose a prior which represents your beliefs
- In practice, use a convenient prior which approximately represents your beliefs
- Choose a family which is computationally convenient, but is also flexible
- A conjugate prior is one that leads to a posterior distribution of the same form (it is likelihood-dependent)
- Examples: Prior- Likelihood $\rightarrow$ Posterior
  - Beta–binomial $\rightarrow$ Beta
  - Normal–normal $\rightarrow$ Normal
  - Gamma–gamma $\rightarrow$ Gamma
  - Gamma–poisson $\rightarrow$ Gamma

Exponential Families

If the density of Y can be written in the following form

$$p(y_i|\theta) = f(y_i)g(\theta)\exp(\phi(\theta)^T u(y_i))$$

then we say that Y is a member of an exponential family with natural parameter $\phi(\theta)$ (note θ and y_i may be vectors).

For a sequence of *iid* observations, the likelihood is

$$\begin{aligned} p(y|\theta) &= \left[\prod_{i=1}^{n} f(y_i)\right] g(\theta)^n \exp(\phi(\theta)^T \sum_i u(y_i)) \\ &\propto g(\theta)^n \exp \phi(\theta)^T t(y) \end{aligned}$$

where the quantity $t(y) = \sum u(y_i)$ is a sufficnent statistic for θ because the likelihood for θ depends on y only through the value $t(y)$.

Conjugate Analysis

If the prior distribution is of the form

$$p(\theta) \propto g(\theta)^{\eta} \exp(\phi(\theta)^T \nu)$$

then the posterior density for θ is of the form

$$p(\theta|y) \propto g(\theta)^{(n+\eta)} \exp(\phi(\theta)^T (\nu + t(y)))$$

which is in the same family of distributions as the prior distribution.

Note: the exponential families are the only classes of distributions that have conjugate prior distributions.

Why Hierarchical Models?

- Typical prior for binomial proportion parameter is a Beta
  - Beta is conjugate
  - Uniform is a special case
- What if the beta family is not flexible enough?
- What if our data are overdispersed?
  - Recall VA data are overdispersed

A Simple Hierarchical Model

- Typical prior is $Beta(a, b)$
- How do we choose a and b?
- Instead of choosing values for hyperparameters, treat them as unknown and also put priors on them
- Example:

$$
\begin{aligned}
y_i|\theta &\sim Bin(n_i, \theta) \\
\theta|\alpha, \beta &\sim Beta(\alpha, \beta) \\
\alpha|a, b &\sim \Gamma(a, b) \\
\beta|a, b &\sim \Gamma(a, b)
\end{aligned}
$$

- Advantages:
  - Hierarchical prior gives more flexibility to model
  - Posterior is less sensitive to choice of hyperparameters
- Disadvantages:
  - Model is harder to fit, as parameters are not “jointly conjugate”
    * Can marginalize
    * Can fit with Markov Chain Monte Carlo (MCMC)
  - Parameters can be more difficult to interpret
  - Does not account for overdispersion

A Different Hierarchical Model

- Account for overdispersion by fitting a separate θ_i for each VA hospital:

$$\begin{aligned} y_i|\theta_i &\sim Bin(n_i, \theta_i) \\ \theta_i|a,b &\sim Beta(a,b) \end{aligned}$$

- This is a random effects model
- By borrowing strength, can fit many parameters
- θ_i's are shrunk towards prior mean $a/(a+b)$
- Likelihood analysis based on Beta-Binomial model (estimate a and b)
- Could also put prior on hyperparameters; use of numerical integration or MC

Comparing Ranks through Simulation

Want to make posterior inference on the relative ordering of the quality of care of the VA hospitals in 1992.

Observed data give one ranking, but no indication of uncertainty

- see the 6th ranked hospital
- many are very similar

Simulation with Transformations

- (Jointly) Ranks are a function of the proportions
  - Observed ranks are a function of observed proportions
  - Posterior for ranks is a function of the posterior for proportions
- A simulation from the posterior for proportions produces a simulation from the posterior for ranks
- Evaluate posterior for ranks via Monte Carlo

- $y_i \sim \text{Bin}(n_i, \theta_i)$; $\theta_i \sim$Beta$(1, 1)$
- Simulate a vector of proportions from the posterior

$$\theta_i | Y \sim Beta\,(y_i + 1, n_i - y_i + 1)$$

  $\theta_i|Y$'s conditionally independent
- Generate k samples from the posterior for proportions
- Transform each sample into a sample of posterior ranks