Tuning joint priors

Stat 340: Bayesian Statistics

Example: Comparing proportions

- Yen et al. (2009) studied breast cancer survivors to consider possible risk factors for lymphedema (LE)
- Condition caused by a blockage of lymph vessels that drain fluids from body tissues
- $n = 1211$ women were followed for 4 years
- Tracked LE status and number of lymph nodes examined during surgery

Research question:
Does the probability of LE vary according to the of nodes examined?

Example adapted from *Bayesian Thinking in Biostatistics*

Observed data

Nodes	Lymphedema		
examined	Yes	No	
High (>5)	126	351	477
Low (≤5)	53	681	734

- We're conditioning on the number of nodes
- Assume the number of LE cases in the high and low categories are independent binomial random variables

LE model

θ_1 = probability of LE for high node counts

θ_2 = probability of LE for low node counts

Likelihood: $Y_1|\theta_1 \sim \text{Binom}(n_1, \theta_1), \quad Y_2|\theta_2 \sim \text{Binom}(n_2, \theta_2)$

$Y_1|\theta_1, Y_2|\theta_2$ independent

Priors: $\theta_1 \sim \text{Beta}(a_1, b_1), \quad \theta_2 \sim \text{Beta}(a_2, b_2)$

θ_1, θ_2 independent

Tuning beta priors

We're assuming θ_1, θ_2 are independent, so we can tune the two beta priors separately

Prior information from literature:

- Indicates reasonable ranges for LE probability after surgery
- Risk is about 2% per year, continues for many years
- For 4 year period
  - set prior mode to 0.08
  - $P(\theta_i < 0.3) = 0.95$

Other priors

- In the absence of prior information, using independent reference priors is justified
- If we prior studies had been conducted, then we could use data augmentation priors
- We're not constrained to the beta family of priors, but it is convenient!

Example: Comparing rates

- Colditz et al. (1990) studied rates of breast cancer for 50-59-year-old postmenopausal women
- Two cohorts studied
    - estrogen replacement therapy
    - no estrogen replacement therapy

Research question:
Does estrogen replacement therapy impact the rate of breast cancer?

Example adapted from *Bayesian Thinking in Biostatistics*

Observed data

Group	Cancer cases	Person-years
Hormone therapy	123	46,524
None	288	145,159

Person-years: "The product of the number of years times the number of members of a population"

We'll *assume* the number of cancer cases in the two cohorts are independent

Cancer model

θ_1 = rate of occurrence of cancer per person-year in hormone therapy cohort

θ_2 = rate of occurrence of cancer per person-year in no-therapy cohort

Likelihood: $Y_1|\theta_1 \sim \text{Poisson}(\theta_1 M_1), \quad Y_2|\theta_2 \sim \text{Poisson}(\theta_2 M_2)$

$Y_1|\theta_1, Y_2|\theta_2$ independent

Priors: $\theta_1 \sim \text{Gamma}(a_1, b_1), \quad \theta_2 \sim \text{Gamma}(a_2, b_2)$

θ_1, θ_2 independent

Tuning gamma priors

1. Informative priors

   Ask experts to think about θ_i independently, use this information to tune the prior parameters

2. Non-informative/vague priors

   Use reference or diffuse priors

Reference priors

- A reference prior for a rate parameter of a Poisson distribution is

$$\pi(\theta) \propto 1/\sqrt{\theta}$$

- Equivalent to Gamma(0.5, 0), which is **improper**
- You can approximate it with Gamma(0.5, 0.001) in JAGS