Bayesian Statistics

Basic examples of Bayesian inference

Nan Lin
Department of Mathematics
Washington University in St. Louis

Dirichlet distribution

- How to simulate from a dirichlet distribution?

$$Z_i \sim \text{ Gamma}(\alpha_i, \beta) \text{ independently,}$$

$$\downarrow$$

$$S = \sum_{i=1}^{K} Z_i \sim \text{ Gamma}\left(\sum_{i=1}^{K} \alpha_i, \beta\right)$$

$$\downarrow$$

$$V = (V_1, \cdots, V_K) = (Z_1/S, \cdots, Z_K/S) \sim \text{ Dir}(\alpha_1, \cdots, \alpha_K)$$

```
n=20
a=c(1,1,1)
d1 <- length(a)
y <- matrix(rgamma(d1 * n, a), ncol = d1, byrow = TRUE)
x <- sweep(y, 1, rowSums(y), FUN = '/')
```

A multinomial model (Section 4.3)

- CBS News survey before 1988 presidential election
  - $n = 1447$
  - George Bush, $y_1 = 727$
  - Michael Dukakis, $y_2 = 583$
  - No opinion, $y_3 = 137$
- Model: $y = (y_1, y_2, y_3) \sim Multinomial(n, \theta_1, \theta_2, \theta_3)$
- Uniform prior: $Dirichlet(1,1,1)$
- Posterior:

$$(\theta_1, \theta_2, \theta_3)|y \sim Dirichlet(y_1 + 1, y_2 + 1, y_3 + 1)$$

- By simulation, one can study the posterior distribution of $\theta_1 - \theta_2$

Electoral votes

- 2008 Election: Obama vs MaCain
  - In State j:
    - θ_{Oj} vs θ_{Mj}
    - EV_j electoral votes in total
  - The number of electoral votes for Obama

$$EV_O = \sum_{j=1}^{51} EV_j \, I(\theta_{Oj} > \theta_{Mj})$$

- Data:
  - (q_{Oj}, q_{Mj}), sample proportion of voters for Obama and MaCain in state j
  - Assume each poll is based on a sample of 500 voters
- Uniform prior and independence between states
  - $(\theta_{Oj}, \theta_{Mj}, 1-\theta_{Oj}-\theta_{Mj}) \sim Dirichlet(1,1,1)$
- Posterior
  - $(\theta_{Oj}, \theta_{Mj}, 1-\theta_{Oj}-\theta_{Mj}) | data \sim Dirichlet(500q_{Oj}+1{,}500q_{Mj}+1{,}500(1-q_{Oj}-q_{Mj})+1)$
  - also independent between states
- Simulate the posterior distribution of EV_O

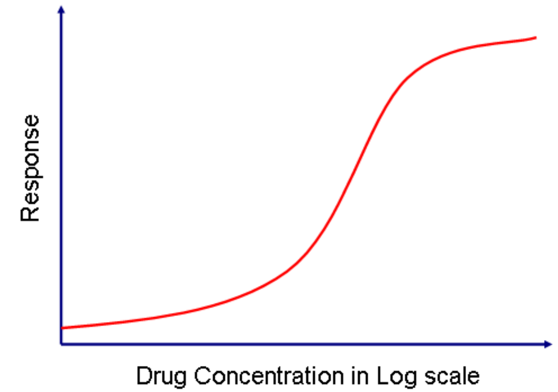
Electoral votes (cont)

- Obtain $P(\theta_{Oj} > \theta_{Mj})$ in the jth state
  - `prob.Obama()`
- Then simulate Obama's electoral votes
  - `sim.selection()`

A Bioassay experiment (section 4.4)

Table 4.1. Data from the bioassay experiment

Dose	Deaths	Sample size
−0.86	0	5
−0.30	1	5
−0.05	3	5
0.73	5	5



- data: $y_i | p_i \sim Bin(n_i, p_i)$
- Logistic regression: $\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_i$
- Parameters: $\theta = (\beta_0, \beta_1)$
- Frequentist: maximum likelihood estimation [`glm()`]

A Bioassay experiment (cont)

- Bayesian
  - Prior specification
    - If dose is at $x_L = -0.7$, the median and 90th percentile of p_L are 0.2 and 0.5. → `beta.select()` ➔ $beta(1.12{,}3.56)$
    - If dose is at $x_H = 0.6$, the median and 90th percentile of p_H are 0.8 and 0.98. ➔ $beta(2.1{,}0.74)$
    - Assume prior independence
    - Since (p_L, p_H) is a one-to-one transformation of (β_0, β_1), we can then obtain the corresponding prior of (β_0, β_1)

$$p_L = \frac{\exp(\beta_0 + \beta_1 x_L)}{1 + \exp(\beta_0 + \beta_1 x_L)}, p_H = \frac{\exp(\beta_0 + \beta_1 x_H)}{1 + \exp(\beta_0 + \beta_1 x_H)}$$

    - The induced prior is then

$$p(\beta_0, \beta_1) \propto p_L^{1.12}(1 - p_L)^{3.56} p_H^{2.1}(1 - p_H)^{0.74}$$

The likelihood function of the unknown regression parameters β_0 and β_1 is given by

$$L(\beta_0, \beta_1) = \prod_{i=1}^{4} p_i^{y_i}(1 - p_i)^{n_i - y_i},$$

where $p_i = \exp(\beta_0 + \beta_1 x_i)/(1 + \exp(\beta_0 + \beta_1 x_i))$. If the standard flat noninformative prior is placed on (β_0, β_1), then the posterior density is proportional to the likelihood function.

If using the informative prior in the previous slide, it is as if we have two more observations in the following table

Dose	Deaths	Sample size
-0.7	1.12	4.68
0.6	2.10	2.84

Non-conjugate prior

- How to simulate from the posterior distribution?
  - Brute-force method by discrete approximation

1. Choose a grid of values of (β_0, β_1) over a lattice that covers the posterior density
2. Compute the posterior density value $p(\beta_0, \beta_1) \propto L(\beta_0, \beta_1)\pi(\beta_0, \beta_1)$ on the grid
3. Normalize by dividing by the sum of over all grid points. In this step, the posterior density is approximated by a discrete probability distribution on the lattice
4. Using the R function sample(), take a random sample with replacement from the discrete distribution.
5. The resulting simulated draws are an approximate sample from the posterior distribution.

Posterior inference

- Plot the contour of the posterior distribution
  - `Mycontour()`
- Simulate from the posterior distribution
  - `Simcontour()`
- For display, one may resort to the frequentist 95% confidence interval to decide the range

Effective/lethal dosage (from wiki)

- The **median lethal dose**, **LD_{50}** (abbreviation for "lethal dose, 50%"), of a toxin, radiation, or pathogen is the dose required to kill half the members of a tested population after a specified test duration. LD50 figures are frequently used as a general indicator of a substance's acute toxicity.
- In logistic regression,

$$LD_{50} = -\frac{\beta_0}{\beta_1}$$