One-parameter models

Patrick Breheny

January 22

Introduction

- Binomial data is not the only example in which Bayesian solutions can be worked out in closed form
- Today's topic is other one-parameter models in which conjugacy can be employed
- These models are interesting and useful on their own, as well as serving as building blocks for more complicated models

Conjugate prior

- First, let's consider the Poisson distribution: $Y \sim \text{Pois}(\theta)$, with likelihood

$$L(\theta|y) \propto \theta^y e^{-\theta}$$

- We may recognize this as the kernel of a Gamma distribution:

$$p(\theta|\alpha, \beta) \propto \theta^{\alpha-1} e^{-\theta\beta} \quad \text{for } \theta > 0$$

- Thus, if we let θ have a Gamma prior, the posterior distribution will also be in the Gamma family

Posterior

- Specifically, if we let $\theta \sim \text{Gamma}(\alpha, \beta)$,

$$\theta | y \sim \text{Gamma}(\alpha + y, \beta + 1)$$

- Note that if $Y_i \overset{\text{iid}}{\sim} \text{Pois}(\theta)$, then $\sum_i Y_i \sim \text{Pois}(n\theta)$
- Thus, we also have that if we observe n iid Poisson random variables $\{Y_i\}$ and let $\theta \sim \text{Gamma}(\alpha, \beta)$,

$$\theta | y \sim \text{Gamma}\left(\alpha + \sum_i y_i, \beta + n\right)$$

Jeffreys prior

- For the Poisson likelihood, the Jeffreys prior is $p(\theta) \propto \theta^{-1/2}$ (homework)
- Unfortunately, $\theta^{-1/2}$ is not integrable over $[0, \infty)$
- This doesn't necessarily cause a problem, though – note that the Jeffreys prior can be thought of as a $\text{Gamma}(\frac{1}{2}, 0)$ distribution, leading to the posterior $\theta | y \sim \text{Gamma}(\frac{1}{2} + y, 1)$
- Note, however, that a $\text{Gamma}(\frac{1}{2}, 0)$ distribution is not really a distribution, in that it cannot integrate to 1

Improper priors

- Such a distribution is said to be *improper*; the use of them in Bayesian statistics is perhaps somewhat controversial
- Some statisticians have argued that such distributions cannot legitimately represent a prior belief and thus cannot be a rational part of Bayesian statistics
- However, most Bayesian statisticians consider them reasonable in the sense of representing a limit of proper posteriors:

$$\text{Gamma}\left(\frac{1}{2}, 0\right) = \lim_{\beta \to 0} \text{Gamma}\left(\frac{1}{2}, \beta\right),$$

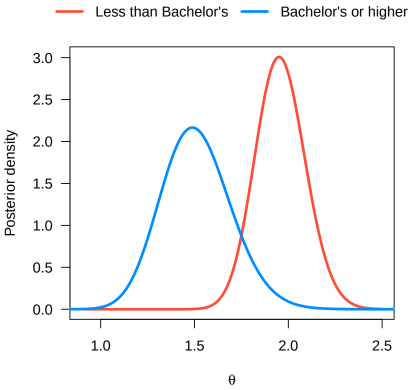
  with the posterior representing a similarly limiting case
- Nevertheless, it is important to be careful when dealing with improper priors; the limiting distribution idea produces reasonable conclusions only if the posterior is guaranteed to be proper

Example

- The General Social Survey (GSS) is a sociological survey of United States residents carried out every other year by the National Opinion Research Center at the University of Chicago
- The survey collected data on 155 women who were 40 years of age or older in the 1990s
- Among the 155 women were 111 whose highest educational level was less than a bachelor's degree (these women had a total of 217 children) and 44 women with at least a bachelor's degree (these women had a total of 66 children)

Example (cont'd)

Using the (improper) Jeffreys priors:



	Less than Bachelor's	Bachelor's or Higher
$\hat{\lambda}$	1.95	1.51
HDI_{95}	$(1.70, 2.22)$	$(1.16, 1.88)$

Distribution of $\theta|y$ vs. $Y|y$

- It is worth distinguishing between the distribution of $\theta|y$, which represents our uncertainty about the *average* number of children in these two groups, and the distribution of the actual number of children in the two groups
- We can derive this distribution as well, however; letting Y denote the random number of children a woman in a particular group might have and y the observed data,

$$p(Y|y) = \int p(Y|\theta)p(\theta|y)d\theta$$

$$\ldots$$

$$= \text{NegBin}\left(a_n, \frac{b_n}{b_n + 1}\right)$$

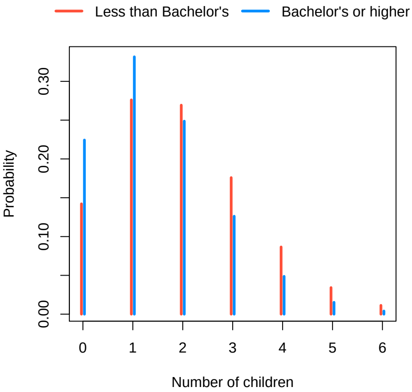
Posterior predictive distributions

- This idea of integrating over θ (or "marginalizing over" θ) applies to any random quantity of interest ω that depends on the posterior distribution:

$$p(\omega) = \int p(\omega|\theta, y)p(\theta|y)d\theta,$$

  and is an essential idea in Bayesian inference and decision theory
- This is sometimes called the *posterior predictive distribution*; we will see many more examples throughout the course

Distribution of $Y|y$



	Less than Bachelor's	Bachelor's or Higher
$\mathrm{E}(\theta \vert y)$	1.96	1.51
$\mathrm{E}(Y \vert y)$	1.96	1.51
$\mathrm{Var}(\theta \vert y)$	0.02	0.03
$\mathrm{Var}(Y \vert y)$	1.98	1.55

Remarks

Note that

- The means, $\mathrm{E}(\theta|y)$ and $\mathrm{E}(Y|y)$, are exactly the same
- The variance of $Y|y$ is much larger than the variance of $\theta|y$; correspondingly, the distributions of $Y|y$ overlap heavily, unlike those of $\theta|y$
- This is a crucial distinction in both Bayesian and frequentist statistics: strong evidence of a difference between two populations does not mean that the difference itself is large

Remarks (cont'd)

- $\mathrm{Var}(Y|y)$ is larger than the variance we would get if we simply plugged in the mean as a point estimate
- This is an appealing point of Bayesian statistics; with a few exceptions, in frequentist statistics we must simply plug a point estimate (usually the MLE) into a predictive distribution
- This fails to account for the multiple sources of variability (random variability in the quantity of interest and uncertainty about θ), and leads to prediction intervals that are too narrow

Conjugate prior, posterior

- Next, let us consider the exponential distribution:

$$L(\theta|y) = \theta e^{-\theta y}$$

- Note that the exponential distribution also has the gamma as a conjugate prior, leading to the posterior distribution

$$\theta|\mathbf{y} \sim \text{Gamma}\left(\alpha + n, \beta + \sum_i y_i\right)$$

Combining conjugate analyses

- Even when all parameters can be analyzed using conjugate methods, Monte Carlo methods may still be required to study the posterior distribution of a quantity of interest
- For example, consider the (hypothetical) heart transplant study in section 3.5 of our book:
    - 10 patients receive a heart transplant, of which 8 survive
    - The surviving patients are monitored, and survive for 2, 3, 4, 4, 6, 7, 10, and 12 years following the transplant
- If we assume that, given transplant success, survival time follows an exponential distribution, we can carry out separate conjugate analyses for π, the probability of success, and λ, the survival rate

Transplant example

- Our quantity of interest, however, is $\omega = \pi/\lambda$, the average total survival
- We can calculate its posterior through Monte Carlo means, drawing π and λ from their posterior distributions and then taking the Monte Carlo integral of π/λ, which has posterior mean 5.1 years
- Suppose that the average life span of a patient who does not receive a transplant is 2 years; we would be interested in $Pr(\omega > 2|\mathbf{y})$, which is 99%
- Note, again, that this is very different from the posterior probability that an individual patient will survive at least 2 years, which is only 55%

Summary

- We have looked at binomial, exponential, and Poisson likelihoods, but all distributions in the exponential family have natural conjugate priors; Table 3.1 in our text provides a list along with posteriors and predictive distributions
- Up next: distributions with two (or more) parameters, in particular the normal distribution