22S:138 Bayesian Statistics

Inference for Proportions, concluded Introduction to Other One-Parameter Models

Lecture 7
Sept. 12, 2005

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

Nonconjugate prior distributions

- conjugate prior distributions are a convenience but may not reflect true prior knowledge
  - simplify computations
  - easy to understand results
  - may be decent approximations to true prior knowledge
- Bayes' rule for updating from prior to posterior applies with nonconjugate priors as well.

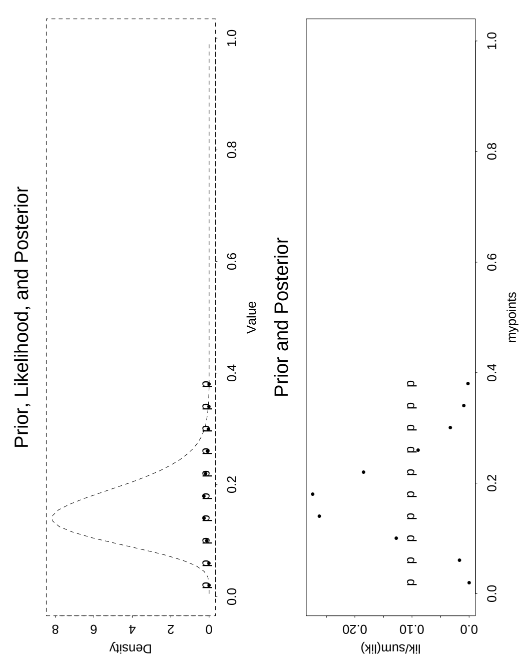
Example: discrete prior for binomial proportion

- In the survey problem, suppose we chose to put all prior probability on a discrete set of values such as

$$p(p) = .1, \quad p = 0.02, 0.06, 0.10, \ldots 0.38$$

- Then, regardless of the likelihood, the posterior probability will be 0 for all values of p except these.
- $posterior \propto prior \times likelihood$
- $p(p|y) \propto .1 \times p^y (1-p)^{n-y}$

Plot of prior, likelihood, and posterior with discrete prior



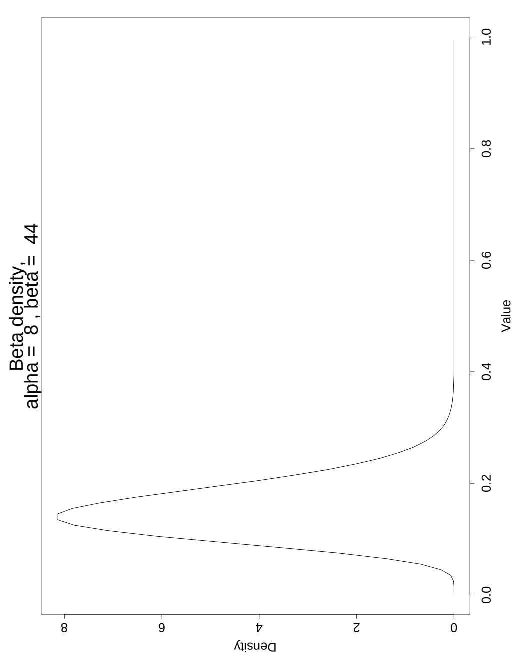
Example: a histogram prior for a binomial proportion

- advantages of histogram priors
  - easy to specify
  - requires no parametric assumptions
  - provides great flexibility in specifying prior beliefs
- how to construct for binomial proportion
  - divide interval (0,1) into predefined subintervals
  - assign probabilities to each interval in accordance with prior belief that population proportion lay in that interval

- e.g.

interval	prior probability
(0, .2]	0.9
(.2, .4]	0.075
(.4, .6]	0.02
(.6, .8]	0.004
(.8, 1.0)	0.001

Plot of histogram prior and likelihood



Plot of resulting posterior

Other one-parameter models: Inference for the mean of a normal population with known variance

- For what kind of problem might a normal likelihood be appropriate?
  - random variable is continuous-valued
  - expect roughly symmetric distribution of values in the population
  - not too "heavy-tailed" of a distribution
- form of normal p.d.f.

$$y \sim N(\mu, \sigma^2)$$
$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

Example

- We wish to estimate the population mean μ of the rate at which skin wounds heal in a particular species of newt.
- Biologists measured the rate at which new cells closed the skin of 18 anesthetized newts. They measured in micrometers (millionths of a meter) per hour.
- Scientists usually assume that animal subjects are simple random samples from their species or genetic type. We will do this with these observations.
- We will pretend that we know that the population standard deviation σ for healing rate is 8 micrometers per hour.

Exchangeability: Part I
Exchangeable experiments

- First consider *experiments* with a discrete set of outcomes.
- Two such experiments are exchangeable (for a particular person) if:
  1. The possible outcomes are the same in both experiments.
  2. The probability of each outcome in one experiment is the same as in the other experiment.
  3. The conditional probabilities for the second experiment, given the results of the first experiment, are the same as the conditional probabilities for the first, given the results of the second.

Example: randomly drawing two cards, one at a time, from a set of four cards

- All cards are aces. One is from each suit.
- We will consider each of the two draws a separate experiment.
- Are the two draws independent?

- Are they exchangeable? Consider each criterion.
  1. Possible outcomes of first draw are diamond, heart, club, spade (D,H,C,S). Possible outcomes of second draw are D,H,C,S.

2. Probabilities of each outcome on first draw are $\frac{1}{4}$. What about second draw?

   – we need *marginal* (unconditional) probabilities of each possible outcome on second draw.
   – Pr(draw 2 = D) = ?

3. What about *conditional* probabilities going in each direction?
   – What is $Pr(draw2 = H|draw1 = D)$?
   – What is $Pr(draw1 = H|draw2 = D)$?
   – What is $Pr(draw2 = H|draw1 = H)$?
   – What is $Pr(draw1 = H|draw2 = H)$?

More on criterion 3

- Criterion 3 may be restated to say:
  The experiments are *symmetric* in that the joint probabilities are the same regardless of the order in which the experiments are observed.

- In the example, consider the joint event of getting a D and a C in the two draws. We must check:

$$Pr(draw\ 1 = D\ and\ draw\ 2 = C) = Pr(draw\ 1 = C\ and\ draw\ 2 = D)$$

- This criterion is usually stated more formally as:
  The joint probabilities are invariant to permutations of the indices.

Experiments with continuous outcomes

- When experiments have continuous outcomes, we will be concerned with probability distributions in criteria 2 and 3.

Exchangeable populations, samples, observations

- Consider a population, e.g.
  - UI students
  - all newts
- We plan to select a member of the population and make a measurement, e.g.
  - response to survey question
  - rate of wound healing
- If our probability distributions on the possible responses are the same for all members of the population, then the corresponding experiments (making the measurements) are exchangeable.
  - and so we say the members of the populations themselves are exchangeable

- Observations in a sample are exchangeable if
  - possible outcomes of measurements on them are the same
  - probability distributions on possible outcomes are the same
  - learning is symmetric
- Often we consider sample observations exchangeable because we don't have enough information to tell them apart. If we later obtain more information, our subjective assessment of exchangeability my change.

Exchangeability and conditional independence

- If we consider observations in a sammple exchangeable, we will often specify their joint sampling distribution by treating the observations as conditionally independent given one or more shared parameters.
- This is what we did on Wed.

Joint probability distribution for newts' healing rates

$$\begin{aligned} p(y_1, y_2, \ldots y_n|\mu, \sigma^2) &= \prod_{i=1}^{18} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(-\frac{\Sigma_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(-\frac{\Sigma_{i=1}^n (y_i-\bar{y})^2}{2\sigma^2}\right) exp\left(-\frac{n\ (\bar{y}-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

We want to estimate the unknown population mean μ, assuming that σ^2 is a known constant. So viewing this expression as a likelihood for μ:

$$L(\mu|\mathbf{y}) \propto exp\left(-\frac{n\ (\bar{y}-\mu)^2}{2\sigma^2}\right)$$

Sufficient statistics

- Note that $\bar{y}$ appears in the likelihood instead of all the individual values y_i from each observation.
- Recall that a *statistic* is a number that can be calculated from sample data without our having to know the values of an unknown parameters.
- $\bar{y}$ is a statistic.
- When, as in this case, a statistic contains all the information in the data that is useful in estimating the unknown parameter of interest, the statistic is called a *sufficient statistic.*
- Using sufficient statistics when they exist makes Bayesian computation much easier.

Conjugate prior for a normal mean, variance assumed known

- The conjugate prior for a normal mean is also normal.
- If we express the likelihood in terms of the sufficient statistic $\bar{y}$

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

- and the prior

$$p(\mu) \;=\; N(\mu_0,\ \sigma_0^2)$$

- Then the posterior distribution will also be normal.

Easier to think about parameters of the posterior normal distribution using precisions rather than variances

- The precision is the inverse of the variance. The more spread out a distribution is (larger variance), the less precise it is (smaller precision).
- When reading Bayesian literature, you must always make sure whether normal distributions are parameterized in terms of the variance or the precision.
- Let's rewrite our likelihood and prior using precisions.

$$\begin{aligned} p(\bar{y} \mid \mu, \tau^2) &= N(\mu, n\tau^2) \\ p(\mu) &= N(\mu_0,\ \tau_0^2) \end{aligned}$$

where $\tau^2 = \frac{1}{\sigma^2}$ and $\tau_0^2 = \frac{1}{\sigma_0^2}$

- then the posterior distribution is also normal

$$p(\mu|\mathbf{y}) \;= N(\frac{n\tau^2\bar{y} \;+\; \tau_0^2\mu_0}{n\tau^2 + \tau_0^2},\; n\tau^2 + \tau_0^2)$$

- The posterior mean is a weighted average of $\bar{y}$ and the prior mean. The weights are proportional to the respective *precisions*.
- The posterior precision is the sum of the precisions from the prior and the likelihood.

Example: The newt data

- Suppose that we had information from a different species of salamanders that the mean healing rate was 20 micrometers per hour. We are very uncertain about whether this information is applicable to the species of newts that we are interested in.
- we will choose a fairly vague prior, parameterized in terms of its precision. (we'll choose a value together in class).

$$p(\mu) \;=\; N(20, ?)$$

- Here are the 18 actual data values.

29	27	34	40	22	28	14	35	26
35	12	30	23	17	11	22	23	33

- $\bar{y} = 25.6$.
- We are pretending that we know that the population standard deviation $\sigma = 8$.
- How do we proceed?