

22S:138
Bayesian Statistics

Inference for Proportions, continued

Lecture 5
 Sept. 10, 2008

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

Revisiting the uniform distribution (a noninformative prior for a proportion)

- The uniform distribution is a special case of the Beta distribution.
- What are its parameters?

$$U(0, 1) = \text{Beta}(?, ?)$$

- What is the equivalent prior sample size for a $U(0, 1)$ prior?

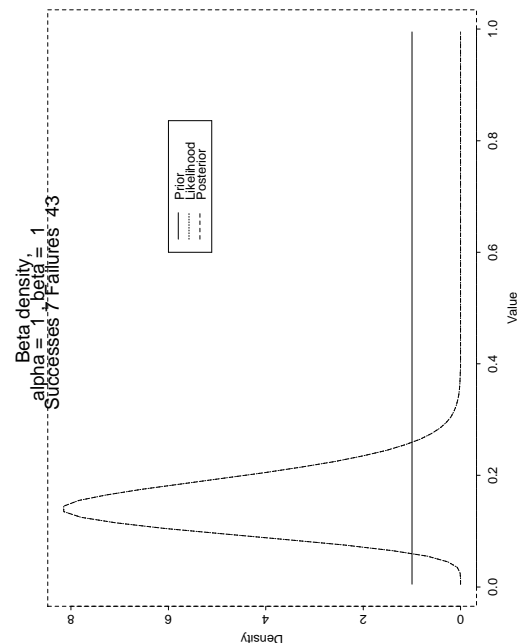
– There is disagreement as to whether the equivalent prior sample size should be defined as

- * $\alpha + \beta$
- * $\alpha + \beta - 1$
- * $\alpha + \beta - 2$

- What is the posterior distribution produced with a $U(0, 1)$ prior and a binomial likelihood?

$$p(p|y) = \text{Beta}(1 + y, 1 + n - y) \\ \propto p^y (1 - p)^{n-y}$$

proportional to the likelihood, as we said before



- Is the posterior mean equal to the MLE \hat{p} ?
- Note that the *mode* of a $Beta(\alpha, \beta)$ distribution is $\frac{\alpha-1}{\alpha+\beta-2}$. So the mode of the posterior distribution given above is $\frac{y}{n} = \hat{p}$.

Estimation

- point estimates
- measures of spread
- Bayesian intervals

The posterior distribution contains all the current information about the unknown parameter

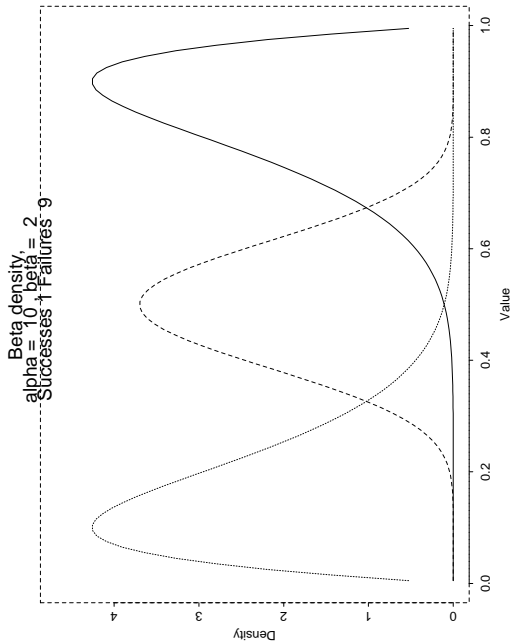
All Bayesian inference is based on the posterior distribution.

- estimation
 - estimating values of unknown parameters that can never be observed or known.
- testing
- prediction
 - estimating the values of potentially observable but currently unobserved quantities.
 - e.g., we might want to predict the number of “yesses” in a future survey of 50 UI students

The posterior variance

- The posterior variance is one summary of the spread of the posterior distribution.
- The larger the posterior variance, the more uncertainty we still have about the parameter.
- See the table of distributions from GCSR for the formula for the variance of a random variable with a beta distribution.
- For a uniform prior and a binomial likelihood, the posterior variance is (almost) always smaller than the prior variance.

When posterior variance is *not* smaller than prior variance



Bayesian intervals

- called “posterior intervals” or “credible sets”
- two kinds
 - equal tail credible sets
 - highest posterior density regions

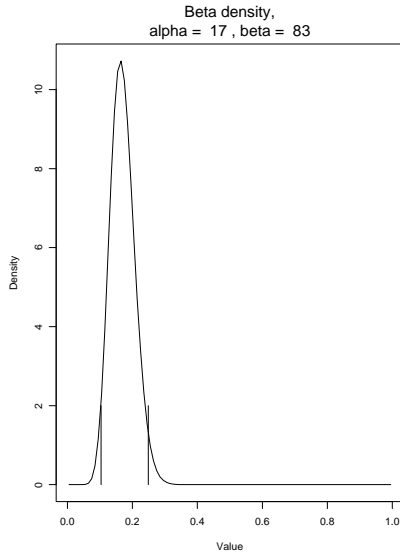
- In our school-quitting example
 - uniform prior
 - * prior variance = $\frac{1}{12} = 0.083$
 - * posterior variance = .00246
 - Beta(10, 40) prior
 - * prior variance = 0.00314
 - * posterior variance = .00140

Equal tail credible sets

- A $100(1 - \alpha)\%$ equal tail credible set is the interval from the $\frac{\alpha}{2}$ quantile to the $1 - \frac{\alpha}{2}$ quantile of the posterior distribution.
- e.g. if we want a 95% equal tail credible set, α is .05 and we need the .025 and the .975 quantiles.
- We can use built-in Splus functions to get quantiles of standard distributions.

- Example, for our quitting school problem with the Beta(10,40) prior, the posterior was Beta(17, 83).

```
> qbeta( c(0.025, 0.975), 17, 83 )
[1] 0.1033333 0.2491463
```



Interpretation of Bayesian intervals

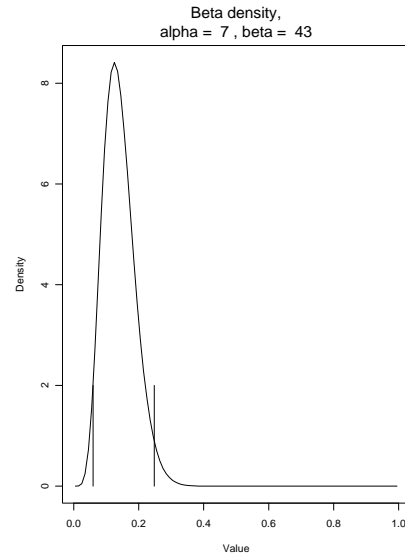
- Recall that the posterior distribution represents our updated subjective probability distribution for the unknown parameter.
- Thus, for us, the interpretation of the 95% credible set is that the probability is .95 that the true p is in that interval.
- If the Beta(10,40) had been a true representation of our prior beliefs or knowledge about the parameter p , then after seeing our survey data, we would believe that

$$P(0.103 < p < 0.249) = .95$$

- Contrast this with the interpretation of a frequentist confidence interval.

- If we had instead used a uniform prior, the posterior was Beta(8,44).

```
> qbeta( c(0.025, 0.975), 8,44)
[1] 0.07024083 0.26255154
```



Highest posterior density regions

- the density at any point *inside* the interval is greater than the density at any point outside it
- shortest possible interval trapping the desired probability
- preferable to equal-tail credible sets when posterior is highly skewed or multimodal
- generally difficult to compute; tables of HDRs for certain densities are available

What would go wrong if the new data were used to formulate the prior?

- Worst case:
 - Suppose we know nothing about the problem; our true prior is uniform (ignorance!)
 - suppose we looked at our own survey data and used its normalized likelihood as our prior

$$p(p) \propto p^7 (1-p)^{43}, \quad 0 < p < 1$$

- prior would be Beta(8, 44)
- posterior would be Beta(15, 87)
- posterior mean would be $\frac{15}{15+87} = .147$
- posterior variance would be 0.0012
- 95% credible set

```
> qbeta( c(0.025, 0.975), 15, 87 )
[1] 0.08557233 0.22162269
```

Using posterior probabilities to test hypotheses

- Suppose we wanted to test the following hypotheses regarding p :

$$H_0 : p \leq .1$$

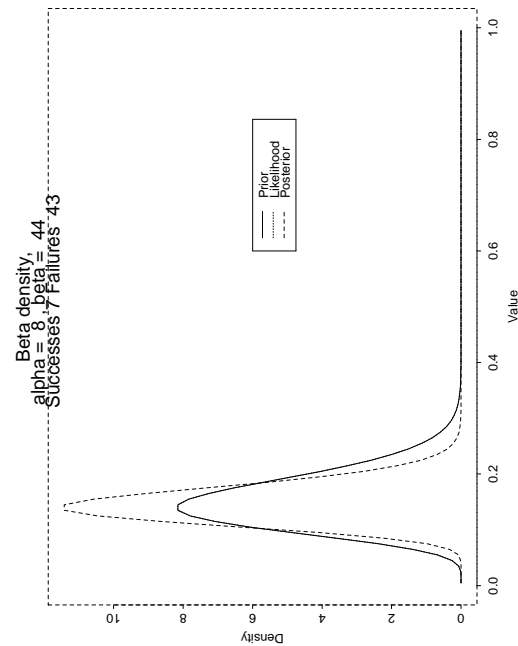
$$H_1 : p > .1$$

- We simply need the posterior probabilities of these two ranges of values for p .
- Suppose that the Beta(10, 40) had been our true prior, so our posterior distribution is Beta(17, 83). We can use a built-in Splus function to obtain $P(p < .1 | y)$.

```
> pbeta(.1, 17, 83)
[1] 0.01879825
```

With this prior, we would conclude that $P(p \leq .1 | y) = .019$.

- If we instead had used the uniform prior, so our posterior was Beta(8, 44),



```
> pbeta(.1, 8, 44)
[1] 0.1329079
```

With this prior, we would conclude that $P(p \leq .1 | y) = .133$.

- The interpretation here is totally different from that of a frequentist p-value.

Robustness

- An inference is *robust* if it is not seriously affected by changes in the assumptions on which it is based.
- Assumptions include
 - form of the likelihood
 - parametric family for prior
 - parameters of prior
 - etc.
- Whether an inference is “seriously affected” depends on the purpose of the analysis.
- In this case, if the primary purpose of the analysis was to get a point estimate for p , we might decide that estimation was quite robust to changes in the prior parameters.
- If our primary purpose was the hypothesis test, we might decide otherwise.