Mathematical Biostatistics Boot Camp: Lecture 13, Binomial Proportions

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

October 18, 2012

Table of contents

Intervals for binomial parameters

- When $X \sim \text{Binomial}(n, p)$ we know that
  a. $\hat{p} = X/n$ is the MLE for p
  b. $E[\hat{p}] = p$
  c. $\mathrm{Var}(\hat{p}) = p(1-p)/n$
  d. $\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}}$ follows a normal distribution for large n
- The latter fact leads to the Wald interval for p

$$\hat{p} \pm Z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

Some discussion

- The Wald interval performs terribly
- Coverage probability varies wildly, sometimes being quite low for certain values of n even when p is not near the boundaries
    - Example, when $p = .5$ and $n = 40$ the actual coverage of a 95% interval is only 92%
- When p is small or large, coverage can be quite poor even for extremely large values of n
    - Example, when $p = .005$ and $n = 1,876$ the actual coverage rate of a 95% interval is only 90%

Simple fix

- A simple fix for the problem is to add two successes and two failures
- That is let $\tilde{p} = (X+2)/(n+4)$
- The (Agresti- Coull) interval is

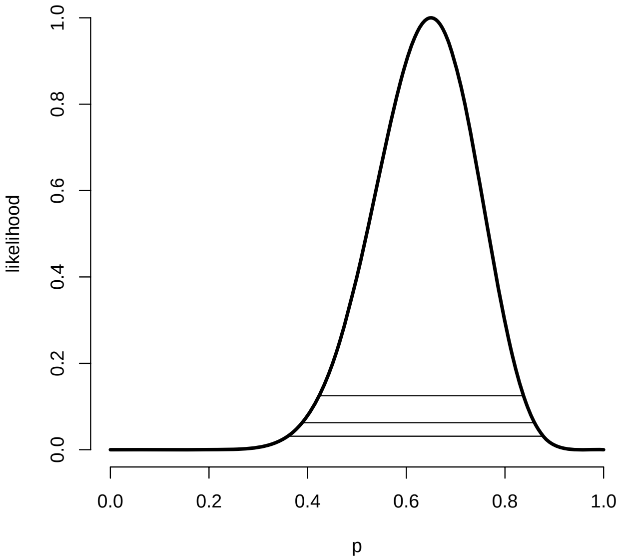
$$\tilde{p} \pm Z_{1-\alpha/2}\sqrt{\tilde{p}(1-\tilde{p})/\tilde{n}}$$

- Motivation: when p is large or small, the distribution of $\hat{p}$ is skewed and it does not make sense to center the interval at the MLE; adding the pseudo observations pulls the center of the interval toward .5
- Later we will show that this interval is the inversion of a hypothesis testing technique

Example

Suppose that in a random sample of an at-risk population 13 of 20 subjects had hypertension. Estimate the prevalence of hypertension in this population.

- $\hat p = .65$, $n = 20$
- $\tilde p = .63$, $\tilde n = 24$
- $Z_{.975} = 1.96$
- Wald interval $[.44, .86]$
- Agresti-Coull interval $[.44, .82]$
- $1/8$ likelihood interval $[.42, .84]$


likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
p

Bayesian analysis

- Bayesian statistics posits a **prior** on the parameter of interest
- All inferences are then performed on the distribution of the parameter given the data, called the **posterior**
- In general,
$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$
- Therefore (as we saw in diagnostic testing) the likelihood is the factor by which our prior beliefs are updated to produce conclusions in the light of the data

Beta priors

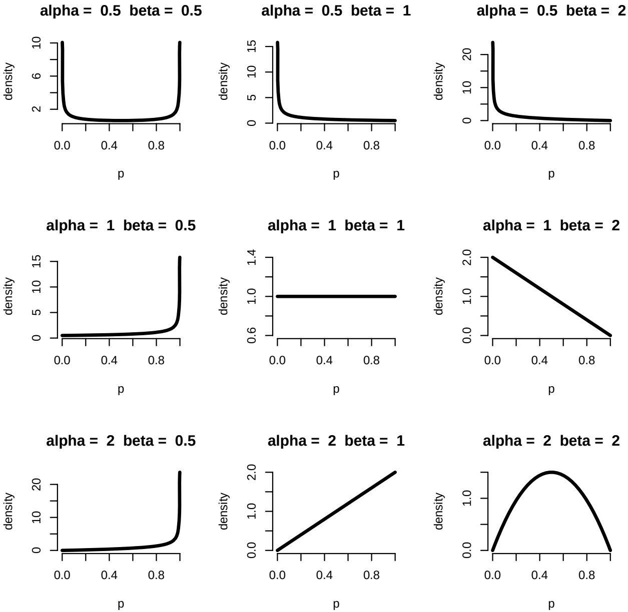
- The beta distribution is the default prior for parameters between 0 and 1.
- The beta density depends on two parameters α and β

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad \text{for} \;\; 0 \leq p \leq 1$$

- The mean of the beta density is $\alpha/(\alpha+\beta)$
- The variance of the beta density is

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- The uniform density is the special case where $\alpha = \beta = 1$


alpha = 0.5 beta = 0.5
alpha = 0.5 beta = 1
alpha = 0.5 beta = 2
alpha = 1 beta = 0.5
alpha = 1 beta = 1
alpha = 1 beta = 2
alpha = 2 beta = 0.5
alpha = 2 beta = 1
alpha = 2 beta = 2
density
p
0.0
0.4
0.8

Posterior

- Suppose that we chose values of α and β so that the beta prior is indicative of our degree of belief regarding p in the absence of data
- Then using the rule that

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

and throwing out anything that doesn't depend on p, we have that

$$\begin{aligned} \text{Posterior} &\propto p^x(1-p)^{n-x} \times p^{\alpha-1}(1-p)^{\beta-1} \\ &= p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \end{aligned}$$

- This density is just another beta density with parameters $\tilde{\alpha} = x + \alpha$ and $\tilde{\beta} = n - x + \beta$

Posterior mean

- Posterior mean

$$
\begin{aligned}
E[p \mid X] &= \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}} \\
&= \frac{x + \alpha}{x + \alpha + n - x + \beta} \\
&= \frac{x + \alpha}{n + \alpha + \beta} \\
&= \frac{x}{n} \times \frac{n}{n + \alpha + \beta} + \frac{\alpha}{\alpha + \beta} \times \frac{\alpha + \beta}{n + \alpha + \beta} \\
&= \text{MLE} \times \pi + \text{Prior Mean} \times (1 - \pi)
\end{aligned}
$$

- The posterior mean is a mixture of the MLE ($\hat{p}$) and the prior mean
- π goes to 1 as n gets large; for large n the data swamps the prior
- For small n, the prior mean dominates
- Generalizes how science should ideally work; as data becomes increasingly available, prior beliefs should matter less and less
- With a prior that is degenerate at a value, no amount of data can overcome the prior

Posterior variance

- The posterior variance is

$$\begin{aligned}\mathrm{Var}(p \mid x) &= \frac{\tilde\alpha\tilde\beta}{(\tilde\alpha+\tilde\beta)^2(\tilde\alpha+\tilde\beta+1)} \\ &= \frac{(x+\alpha)(n-x+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}\end{aligned}$$

- Let $\tilde p = (x+\alpha)/(n+\alpha+\beta)$ and $\tilde n = n+\alpha+\beta$ then we have

$$\mathrm{Var}(p \mid x) = \frac{\tilde p(1-\tilde p)}{\tilde n + 1}$$

Discussion

- If $\alpha = \beta = 2$ then the posterior mean is

$$\tilde{p} = (x + 2)/(n + 4)$$

and the posterior variance is

$$\tilde{p}(1 - \tilde{p})/(\tilde{n} + 1)$$

- This is almost exactly the mean and variance we used for the Agresti-Coull interval

Example

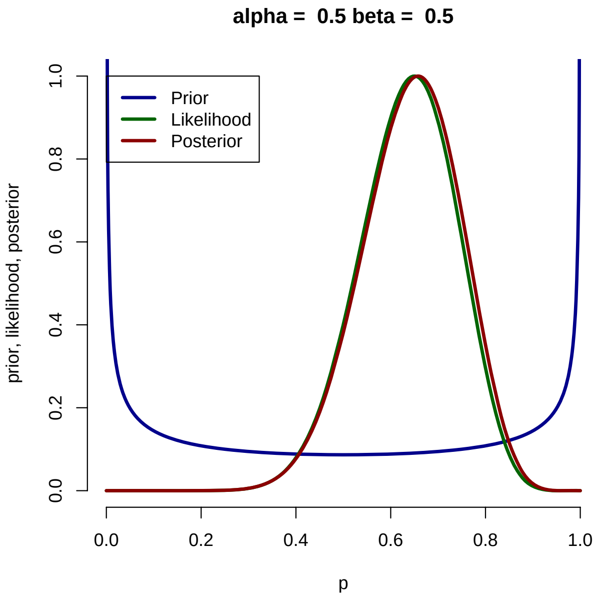
- Consider the previous example where $x = 13$ and $n = 20$
- Consider a uniform prior, $\alpha = \beta = 1$
- The posterior is proportional to (see formula above)

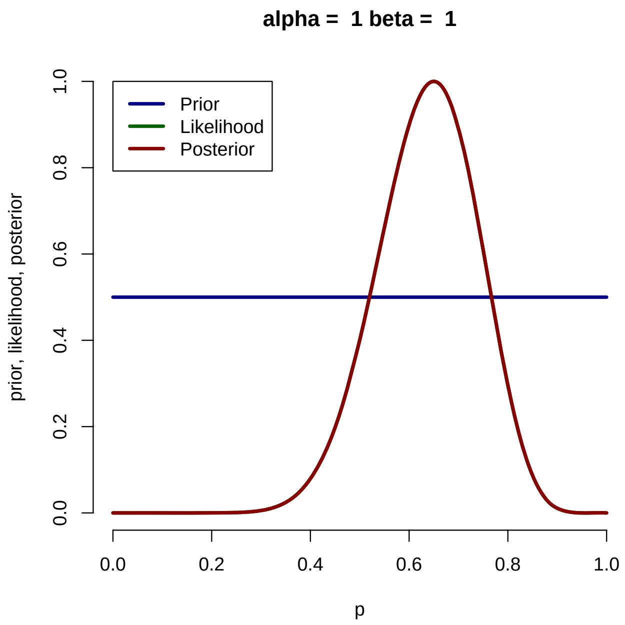
$$p^{x+\alpha-1}(1-p)^{n-x+\beta-1} = p^{x}(1-p)^{n-x}$$

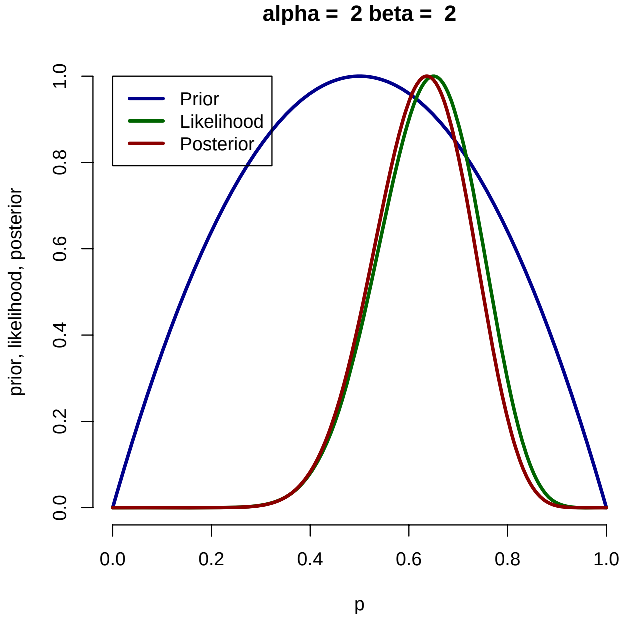
  that is, for the uniform prior, the posterior is the likelihood
- Consider the instance where $\alpha = \beta = 2$ (recall this prior is humped around the point .5) the posterior is

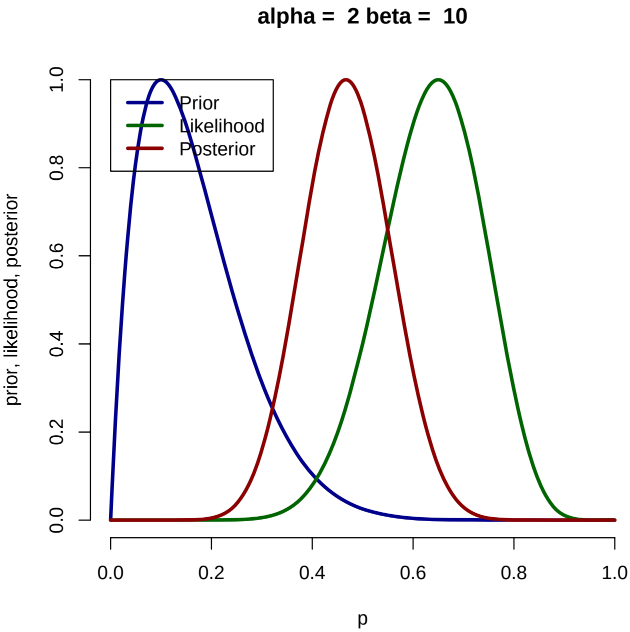
$$p^{x+\alpha-1}(1-p)^{n-x+\beta-1} = p^{x+1}(1-p)^{n-x+1}$$

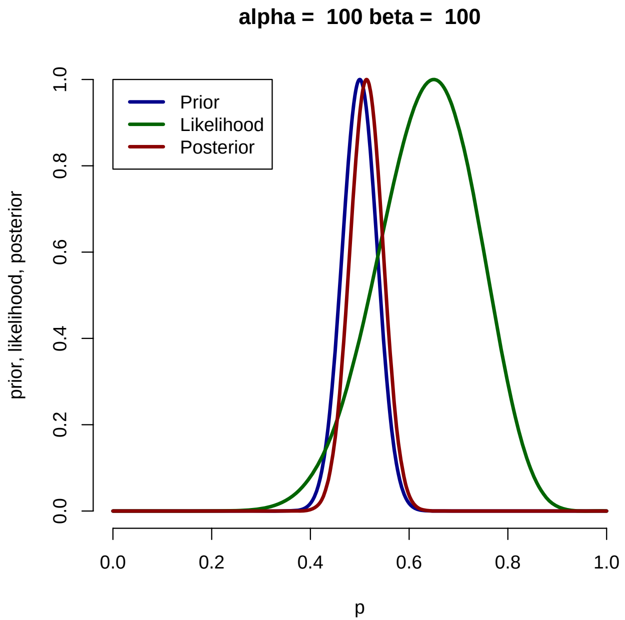
- The "Jeffrey's prior" which has some theoretical benefits puts $\alpha = \beta = .5$


alpha = 0.5 beta = 0.5
Prior
Likelihood
Posterior
prior, likelihood, posterior
0.0
0.2
0.4
0.6
0.8
1.0
p


alpha = 1 beta = 1
Prior
Likelihood
Posterior
prior, likelihood, posterior
0.0
0.2
0.4
0.6
0.8
1.0
p


alpha = 2 beta = 2
Prior
Likelihood
Posterior
prior, likelihood, posterior
0.0
0.2
0.4
0.6
0.8
1.0
p
0.0
0.2
0.4
0.6
0.8
1.0


alpha = 2 beta = 10
Prior
Likelihood
Posterior
prior, likelihood, posterior
0.0
0.2
0.4
0.6
0.8
1.0
p

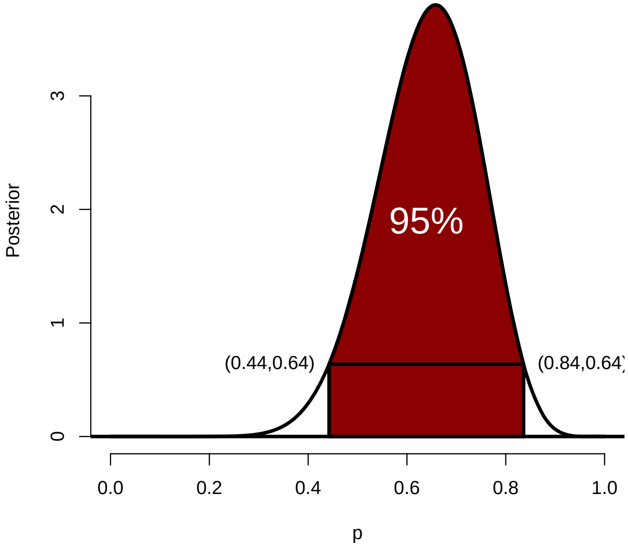

alpha = 100 beta = 100
Prior
Likelihood
Posterior
prior, likelihood, posterior
0.0
0.2
0.4
0.6
0.8
1.0
p

Bayesian credible intervals

- A *Bayesian credible interval* is the Bayesian analog of a confidence interval
- A 95% credible interval, $[a, b]$ would satisfy

$$P(p \in [a, b] \mid x) = .95$$

- The best credible intervals chop off the posterior with a horizontal line in the same way we did for likelihoods
- These are called highest posterior density (HPD) intervals


95%
(0.44,0.64)
(0.84,0.64)
Posterior
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
p

R code

Install the `binom` package, then the command

```
library(binom)
binom.bayes(13, 20, type = "highest")
```

gives the HPD interval. The default credible level is 95% and the default prior is the Jeffrey's prior.

Interpretation of confidence intervals

- Confidence interval: (Wald) $[.44, .86]$
- Fuzzy interpretation:

  We are 95% confident that p lies between .44 to .86

- Actual interpretation:

  The interval .44 to .86 was constructed such that in repeated independent experiments, 95% of the intervals obtained would contain p.

- Yikes!

Likelihood intervals

- Recall the $1/8$ likelihood interval was $[.42, .84]$
- Fuzzy interpretation:

  The interval $[.42, .84]$ *represents plausible values for* p*.*

- Actual interpretation

  The interval $[.42, .84]$ *represents plausible values for* p *in the sense that for each point in this interval, there is no other point that is more than* 8 *times better supported given the data.*

- Yikes!

Credible intervals

- Recall the Jeffrey's prior 95% credible interval was $[.44, .84]$
- Actual interpretation

The probability that p is between .44 and .84 is 95%.