Bayesian Inference for One Parameter Models

The Binomial Model

Yingbo Li

Clemson University

MATH 9810

Teenagers and Televisions

In 1998, the New York Times and CBS News polled 1048 randomly selected 13 - 17 year olds to ask them if they had a television in their room.

- $n = 1048$ sampled teenagers
- $y = 692$ had a television in their room
- Inference about θ, the proportion of 13 - 17 year olds who have a television in their room (in 1998)
- What is a reasonable probability model for data?

Binomial Model

- Independent Bernoulli trials $X_i, (i = 1, \ldots, n)$
- "Success" probability $\theta : p(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$
- Y = number of successes = $\sum_{i=1}^{n} X_i$
- $Y|\theta \sim \text{Bin}(n, \theta)$

$$p(Y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \text{ for } y = 0, 1, \ldots, n$$

- $E(Y|\theta) = n\theta$, $Var(Y|\theta) = n\theta(1-\theta)$
- R functions: `dbinom`, `pbinom`, `qbinom`, `rbinom`

Classical Inference for θ

Assumptions $\rightarrow Y \mid \theta \sim \text{Bin}(n, \theta)$

- Usual point estimator: observed proportion

$$\hat{\theta} = y/n$$

- $E(\hat{\theta}|\theta) = \theta$, so that $\hat{\theta}$ is an unbiased estimator
- $Var(\hat{\theta}|\theta) = \theta(1-\theta)/n$, so that precision (1/variance) increases for large n and θ near zero or one

Likelihood Function

- Likelihood function: $L(\theta) \propto p(y|\theta)$

$$L(\theta) = \binom{1048}{692} \theta^{692}(1-\theta)^{356} \propto \theta^{692}(1-\theta)^{356}$$

- For FIXED y, look at how probability of the data changes as θ varies over parameter space
- For each value of θ, the likelihood indicates how well that value of θ explains the observed data
- Usually drop terms that do not depend on parameters
- Calculations easier with log likelihood function

$$\log L(\theta) \propto y \log(\theta) + (n-y)\log(1-\theta)$$

Maximum Likelihood Estimate

- What is most likely value of θ for these data?
- Maximum likelihood estimate: value of θ that maximizes the likelihood

Using calculus, we take the derivative of $\log L(\theta)$ and set equal to zero:

$$\begin{aligned} 0 &= \frac{\partial \log L(\theta)}{\partial \theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta} \\ (n-y)\theta &= y(1-\theta) \\ \theta &= y/n \end{aligned}$$

The MLE is $\hat{\theta} = y/n = 692/1048 = .660$.

Functions of Parameters: Odds

- odds: $o(\theta) = \theta/(1-\theta)$
- Likelihood is same under one-to-one transformation
  $p(y \mid o) = p(y \mid \theta(o))$
- MLE of $g(\theta)$ is $g(\hat{\theta})$
- estimated probability that a 13-17 year old will have a TV in their room is 0.660
- estimated odds that a 13-17 year old will have a TV in their room is 1.94 to 1
- estimated odds that a 13-17 year old will not have a TV in their room is $.515$ to 1

Classical Interval Estimates for θ

For large n, central limit theorem:

$$\hat{\theta} = y/n \mathrel{\dot\sim} N(\theta, \theta(1-\theta)/n)$$

This leads to 95% confidence interval for θ,

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

$$95\% \text{ CI for } \theta : (0.632, 0.689)$$

Probability that interval covers θ (prior to seeing the data) equals 0.95.

Not same as $p(0.632 < \theta < 0.689) = 0.95$.

Bayesian Inference about θ

Conditional on observed outcome y (and n) the posterior distribution of θ is

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \quad 0 < \theta < 1$$

where

$$p(y) = \int_0^1 p(y|\theta)p(\theta)d\theta$$

is the marginal density of data.
Alternatively, we have

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

subject to normalization to unit integral.

Bayes Inference about θ

- Initial prior uncertainty about θ described by a prior distribution $p(\theta)$.
- Uniform density is a common default choice

$$p(\theta) = 1, \qquad 0 < \theta < 1$$

- Flat density, each point equally weighted
- "uninformative" about true value

Results with Uniform Prior

Posterior $p(\theta|y) \propto \theta^y(1-\theta)^{n-y}(1) = \theta^{y+1-1}(1-\theta)^{n-y+1-1}$

- Recognize that kernel of density is a Beta$(y+1, n-y+1)$

$$p(\theta \mid y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\,\theta^{a-1}(1-\theta)^{b-1} \qquad \text{for } 0<\theta<1$$

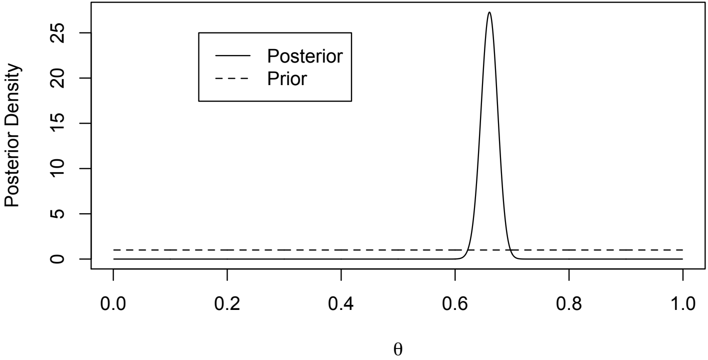
- $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. When x is integer, $\Gamma(x) = (x-1)!$
- Under uniform prior, posterior distribution for θ is Beta$(693, 357)$
- In R: `dbeta`, `pbeta`, `qbeta`, `rbeta`
    - quantiles (percentiles),

```
> qbeta(c(0.025, 0.5, 0.975), y + 1, n - y + 1);
[1] 0.6310792 0.6601016 0.6883435
```

Posterior Distribution for Television Data

Under uniform prior, posterior distribution for θ is Beta$(693, 357)$

```
theta = seq(0.001, 0.999, by = 0.001)
plot(theta, dbeta(theta,693,357),t="l",lty=1)
lines(theta, dbeta(theta, 1,1), lty=2)
```



Beta Distributions $\theta \sim \text{Beta}(a, b)$

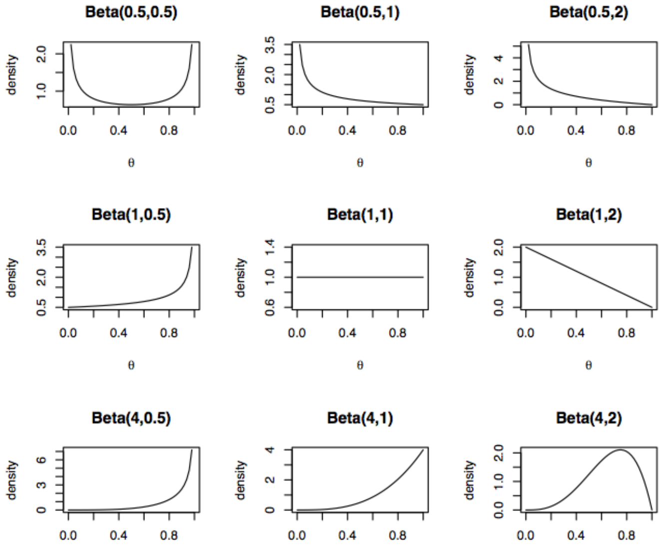
$$p(\theta) = \frac{1}{B(a,b)} \theta^{a-1}(1-\theta)^{b-1} \text{ for } 0 < \theta < 1$$

- $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$E(\theta) = \frac{a}{a+b}, \quad Var(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$$

- More concentrated, or "precise", for larger $a+b$
- unique mode at $(a-1)/(a+b-2)$ if $a, b > 1$
- mode at 0 if $a \leq 1$, and/or one at 1 if $b \leq 1$

Shapes of Beta Priors



Beta-Binomial Model

- Prior: $\theta \sim \text{Beta}(a, b) \Rightarrow p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$
- Likelihood: $L(\theta) = \theta^{y}(1-\theta)^{n-y}$
- Posterior: $p(\theta \mid Y) \propto L(\theta)p(\theta) \propto \theta^{y+a-1}(1-\theta)^{n-y+b-1}$
- $\theta \mid Y \sim \text{Beta}(y+a, n-y+b)$
- Posterior Mean:

$$\frac{y+a}{n+a+b} = \left(\frac{n}{n+a+b}\right)\left(\frac{y}{n}\right) + \left(\frac{a+b}{n+a+b}\right)\left(\frac{a}{a+b}\right)$$

  weighted average of MLE and prior mean
- a is prior number of "1"s and $a+b$ is prior sample size

Conjugate Prior Distributions

Consider a class of prior distributions, $p(\theta) \in \mathcal{P}$.

We say that the class is conjugate for a sampling model $p(y \mid \theta)$, if $p(\theta) \in \mathcal{P}$ implies that $p(\theta \mid Y) \in \mathcal{P}$ for all $p(\theta) \in \mathcal{P}$ and data y.

- If y has a binomial distribution, then the class of Beta prior distributions is conjugate.
- We will see that sampling models based on exponential families all have conjugate priors.

Posterior Intervals

Credible intervals, or confidence region

$$P(l(y) < \theta < u(y) \mid Y = y) = 0.95$$

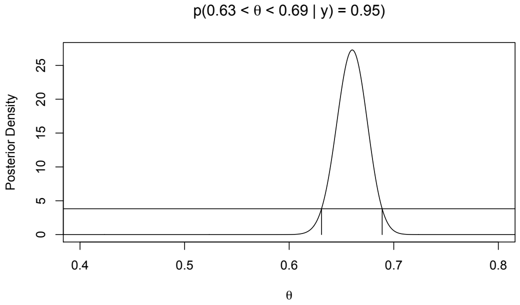
- Quantile-based intervals (equal tails): 95% interval
  $l(y) = 2.5\%$ quantile of $p(\theta \mid y)$
  $u(y) = 97.5\%$ quantile of $p(\theta \mid y)$

```
> qbeta(c(0.025, 0.975), y + 1, n - y + 1);
[1] 0.6310792 0.6883435
```

- We can also compute one-sided quantile-based intervals.

Highest Posterior Density (HPD) intervals

- $P(\theta \in s(y) \mid Y = y) = 0.95$
- For any $\theta_1 \in s(y)$ and $\theta_2 \notin s(y)$, $p(\theta_1 \mid Y = y) > p(\theta_2 \mid Y = y)$.




```
solve.HPD.beta(y, n, h = 0.14, xlim = c(0.4, 0.8))
```

Based on the samples, there is a 95% chance that between 63% and 69% of 13 - 17 year olds who have a television in their room.

Finding an HPD Interval

1. Find mode of posterior $\tilde{\theta}$
2. Construct relative density $(0 < r(\theta) \leq 1)$

$$r(\theta) = \frac{p(\theta \mid Y)}{p(\tilde{\theta} \mid Y)} = \frac{L(\theta \mid Y)p(\theta)}{L(\tilde{\theta} \mid Y)p(\tilde{\theta})}$$

3. Find points of equal density: start with initial height $h \in (0, 1)$ ($h = 0.1$ often is close) and solve for θ_{lh} such that $r(\theta_{lh}) = h$ where $0 < \theta_{lh} < \tilde{\theta}$. Solve for θ_{uh} such that $r(\theta_{uh}) = h$ with $\tilde{\theta} < \theta_{uh} < 1$
4. Calculate $P(\theta_{lh} < \theta < \theta_{uh} \mid Y)$
5. decrease/increase h and repeat Steps 3 – 4 until the probability is approximately 0.95