STA 365: Applied Bayesian Statistics

Boris Babic
Assistant Professor, University of Toronto

Week 2B: Beta Binomial Model

Review: Bayesian Framework

- Prior distribution for θ:
$$\theta \sim \pi(\theta)$$
- Sample distribution (or likelihood) of $\boldsymbol{X}$ given θ:
$$\boldsymbol{X}|\theta \sim f(\boldsymbol{x}|\theta) = \pi(\boldsymbol{x}|\theta)$$
- Joint distribution of $\boldsymbol{X}$ and θ (this is our full model):
$$f(x, \theta) = f(\boldsymbol{x}|\theta)\pi(\theta)$$
- Recall, chain rule of probability: $p(E_1 \cap E_2) = p(E_2|E_1)p(E_1)$
- Marginal distribution of $\boldsymbol{X}$:
$$m(\boldsymbol{x}) = \int_{\theta \in \Omega} f(\boldsymbol{x}, \theta) d\theta = \int_{\theta \in \Omega} f(\boldsymbol{x}|\theta)\pi(\theta) d\theta$$
- Posterior distribution of θ (conditional distribution of θ given $\boldsymbol{X}$):
$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta)}{m(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{m(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta)\pi(\theta) \qquad \text{(Bayes' Rule)}$$

Bayesian Approach to the Coin Problem

- Let $X = 1$ denote the coin landing on heads.
- Since $X \sim \text{Bernoulli}(\theta)$, we know that $\sum_{i=1}^n X_i$ is Binomial in (n, θ):

$$f(x|\theta, n) = \binom{n}{\sum x_i} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$$

- Our prior for $\theta \sim \text{Beta}(\alpha, \beta)$:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- The posterior for $\theta \sim \text{Beta}(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i)$:

$$\pi(\theta) = \frac{\Gamma((\alpha + \sum_{i=1}^n x_i) + (\beta + n - \sum_{i=1}^n x_i))}{\Gamma(\alpha + \sum_{i=1}^n x_i)\Gamma(\beta + n - \sum_{i=1}^n x_i)} \theta^{\alpha+\sum x_i}(1-\theta)^{\beta+n-\sum x_i}$$

  Let $\alpha^* = \sum_{i=1}^n x_i + \alpha$. Let $\beta^* = n - \sum_{i=1}^n x_i + \beta$. Then our posterior distribution for θ is $\text{Beta}(\alpha^*, \beta^*)$:

$$\pi(\theta|\boldsymbol{x}) = \frac{\Gamma(\alpha^* + \beta^*)}{\Gamma(\alpha^*)\Gamma(\beta^*)} \theta^{\alpha^*-1}(1-\theta)^{\beta^*-1}$$

- This is because the beta distribution and the bernoulli/binomial likelihood belong to the same conjugate family.
- Bayesian updating is simplified by adding the pseudo heads to the observed heads and pseudo tails to observed tails.

Some properties of the Beta distribution

- Let $X \sim \text{Binomial}(n, \theta)$
- Mean: $\alpha/(\alpha+\beta)$
- Variance:

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

- Mode: $(\alpha-1)/(\alpha+\beta-2)$
- Note that the posterior mean is a compromise between prior information and data:

$$\mathrm{E}[\theta|x] = \frac{\alpha+x}{\alpha+\beta+n} = \frac{n}{\alpha+\beta+n}\left(\frac{x}{n}\right) + \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta}\right)$$

- This is:

$$p(\mathrm{MLE}) + 1 - p(\text{prior mean})$$

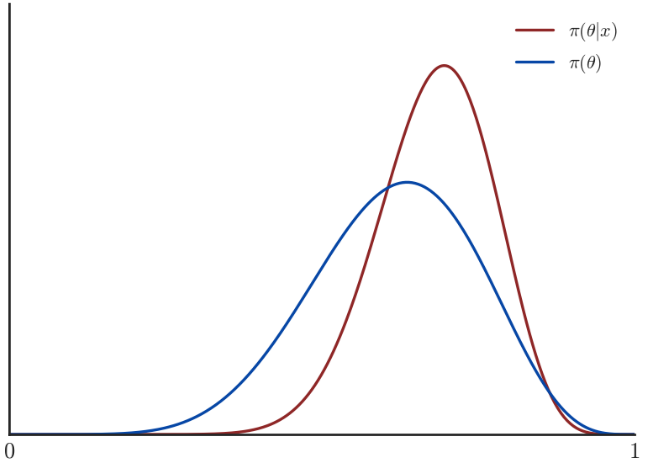
  where p corresponds to the weight of the data, i.e., to n, and $1-p$ corresponds to the weight of the prior (or "pseudo" data), i.e. to $\alpha+\beta$.
- What happens when the sample size goes to infinity?

$$\lim_{n\to\infty}(\mathrm{E}(\theta|x) - \frac{x}{n}) = \ ?$$

Bayesian Approach to the Coin Problem

- Our problem: a coin that was bent in a way that seemed to favor heads, (i.e., $\theta > 0.5$) landed on 9H and 3T.
- Suppose $\alpha = 8$ and $\beta = 5$.
- The posterior distribution is beta with $\alpha = 9 + 8 = 17$ and $\beta = 3 + 5 = 8$.

Posterior Distribution



Inferences on θ

- Any statements that we wish to make about θ can be easily computed from the posterior distribution.
- The posterior distribution describes all our beliefs about θ after viewing the data.
- For example, we way want to make a point estimate using the posterior mean.

  This is given by $\alpha/(\alpha+\beta)$.

  Before seeing the data, this was $8/(8+5) \approx 0.61$.

  After seeing the data, this is $17/(17+8) \approx 0.68$.

  Note that the sample mean is 0.75. The data has nudged our prior toward a stronger belief in the coin's bias toward heads.

  But is is not as strong as the MLE because (remember) the posterior mean is a weighted average of the MLE and the prior mean.
- We may also want the mode, which is the value we think most likely. This is $(\alpha-1)/(\alpha+\beta-2) = (17-1)/(17+8-2) \approx 0.69$.

Bayesian Hypothesis Tests

- Recall that what we really wanted to know was a simple question: is the coin biased toward heads?
- Now we can answer it directly:

$$\begin{aligned}\Pr(\theta > 0.5) &= \int_{0.5}^{1} \pi(\theta|\boldsymbol{x}) d\theta \\ &= 1 - CDF(\theta|\boldsymbol{x})|_{\theta=0.5} \\ &= 1 - 0.03 \\ &= 0.97\end{aligned}$$

- R code: `1 - pbeta(0.5, 17, 8)`
- We are 97% confident that the coin is biased toward heads.
- We now have an answer to a one-sided hypothesis test:

$$H_0 : \theta \leq 0.5 \qquad\qquad H_1 : \theta > 0.5$$

- But instead of accepting/rejecting the null hypothesis, we make probabilistic statements about the research hypothesis from the posterior distribution.

Bayesian Two-Sided Hypothesis Tests

- But what if we want to know whether the coin is fair or not? That is,

$$H_0 : \theta = 0.5 \qquad\qquad H_1 : \theta \neq 0.5$$

- On the picture developed so far, we cannot do this.
- The probability that θ takes on any specific value is 0. Thus the posterior probability for any such H_0 will be 0.
- We will see how to make binary decisions in the Bayesian framework once we introduce the notions of loss and Bayes risk.

Credible Intervals

- However, we can calculate a $(1-\alpha)100\%$ credible interval for θ. For example, a 95% credible interval for θ is,

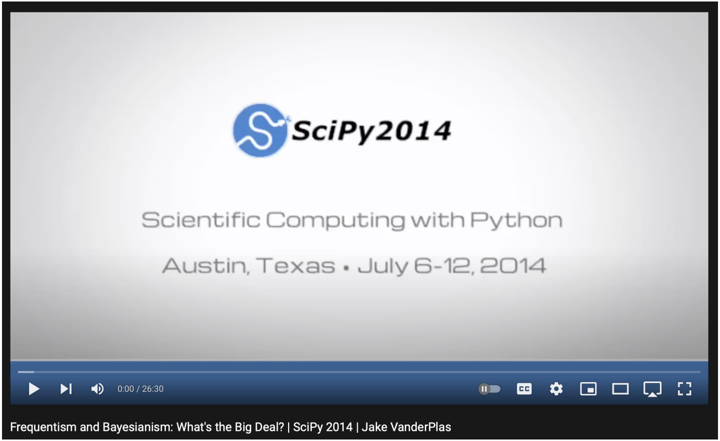
$$\Pr(a < \theta < b) = \int_a^b \pi(\theta|\boldsymbol{x}) d\theta = 0.95$$

- In our case, $a = 0.49$ and $b = 0.84$.
- R code: `qbeta(c(0.025,0.975),17,8)`
- We can also compute the probability that θ is in any desired region of the posterior distribution. This gives us a probabilistic statement about a small region around a point null hypothesis. For example:

$$\begin{aligned}\Pr(0.4 < \theta < 0.6) &= \int_{0.4}^{0.6} \pi(\theta|\boldsymbol{x}) d\theta \\ &= CDF(\theta|\boldsymbol{x})|_{\theta=0.6} - CDF(\theta|\boldsymbol{x})|_{\theta=0.4} \\ &= 0.19\end{aligned}$$

- R code: `pbeta(0.6, 17, 8) - pbeta(0.4, 17, 8)`.
- We are about 20% confident that θ is between 0.4 and 0.6.
- Compare this to the confidence interval from Class 1A. This is now a probabilistic statement about θ, treated as a random quantity. And not a probabilistic statement about X and the proportion of cases in which it will cover θ if sampled repeatedly!

Video



https://youtu.be/KhAUfqhLakw?t=1231