Section 3

PSTAT 115, Fall 2020

October 20, 2020

This lab will focus on the followng topics:

- Derive posterior distribution from prior and likelihood function

- Common Conjugate priors

- Posterior mean and MAP estimates

- Computing probability intervals with quantile functions

The following example will be adopted throughout to illustrate the concepts.

Estimating the probability of a female birth

The proportion of births that are female has long been a topic of interest both scientifically and to the lay public. Two hundred years ago it was established that the proportion of female births in European populations was less than 0.5. The currently accepted value of the proportion of female births in large European-race populations is 0.485.

- Assuming Bernoulli distribution for each child birth, and let p denote the probability of being a female. What is the probability of observing y female babies among n newborn babies (y is between 0 and n, inclusive)?

- Now assume we do not have any prior information about the probability of female birth, so a uniform prior on $[0, 1]$ is used. What is the posterior distribution under this prior?

$$p(\theta) = \text{dbeta}(\theta, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \quad \text{for } 0 \le \theta \le 1.$$

- We know the uniform distribution is a special case of the Beta distribution. Now what if we apply a $Beta(2, 2)$ distribution as our prior? What is the posterior distribution now?

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

$$= \frac{1}{p(y)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \times \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

$$= c(n,y,a,b) \times \theta^{a+y-1}(1-\theta)^{b+n-y-1}$$

$$= \text{dbeta}(\theta, a+y, b+n-y).$$

It is important to understand the last two lines above: The second to last line says that $p(\theta|y)$ is, as a function of θ, proportional to $\theta^{a+y-1} \times (1-\theta)^{b+n-y-1}$. This means that it has the same *shape* as the beta density dbeta$(\theta, a+y, b+n-y)$. But we also know that $p(\theta|y)$ and the beta density must both integrate to 1, and therefore they also share the same *scale*. These two things together mean that $p(\theta|y)$ and the beta density are in fact the same function. Throughout the book we will use this trick to identify posterior distributions: We will recognize that the posterior distribution is proportional to a known probability density, and therefore must equal that density.

Common Conjugate priors

Conjugacy is formally defined as follows. A collection of pdfs (or pmfs) is called a conjugate prior family for a model $X \sim f(x|\theta), \theta \in \Theta$, if whenever a prior $\xi(\theta)$ is chosen from the collection, it leads to a posterior $\xi(\theta|x)$ that is also a member of the collection, for every observation $X = x$. For instance, in our above binomial model, the Beta distribution is a conjugate prior to the binomial distribution.

Conjugate priors permit fast posterior computations, which can be valuable in high dimension problems. Meanwhile, there are handy formulas for mean and mode of the posterior distributions. The following table summarizes the common distribution we might encounter throughout the course.

Model	Parameter	Prior	Posterior
$X \sim \text{Binomial}(n, p)$	$0 \le p \le 1$	$\text{Beta}(a, b)$ $a > 0, b > 0$	$\text{Beta}(a', b')$ $a' = a + x$ $b' = b + n - x$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \text{Poisson}(\lambda)$	$\lambda > 0$	$\text{Gamma}(a, b)$ $a > 0, b > 0$	$\text{Gamma}(a', b')$ $a' = a + n\bar{x}$ $b' = b + n$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \text{Exponential}(\lambda)$	$\lambda > 0$	$\text{Gamma}(a, b)$ $a > 0, b > 0$	$\text{Gamma}(a', b')$ $a' = a + n$ $b' = b + n\bar{x}$
$X = (X_1, \cdots, X_n)$ $X_i \overset{\text{IID}}{\sim} \text{Normal}(\mu, \sigma^2)$ σ^2 known	$-\infty < \mu < \infty$	$\text{Normal}(a, b^2)$ $-\infty < a < \infty$ $b > 0$	$\text{Normal}(a', b'^2)$ $a' = \frac{nb^2\bar{x} + \sigma^2 a}{nb^2 + \sigma^2}$ $b'^2 = \frac{\sigma^2 b^2}{nb^2 + \sigma^2}$

Posterior mean and MAP estimates

An early study concerning the sex of newborn Germany babies found that of a total of 98 births, 43 were female. Assume we are using the uniform prior.

- Plotting the prior distribution (in blue lines), binomial likelihood (in green lines) and the posterior distribution (in red lines).

```r
# prior
curve(p/p, from = 0, to = 1, xname = "p", xlab = "p", ylab = "density",
      ylim = c(0, 8), col = 'blue')

# likelihood
curve(choose(98, 43) * p^43 * (1-p)^(98-43),
      from = 0, to = 1,  xname = "p",
```
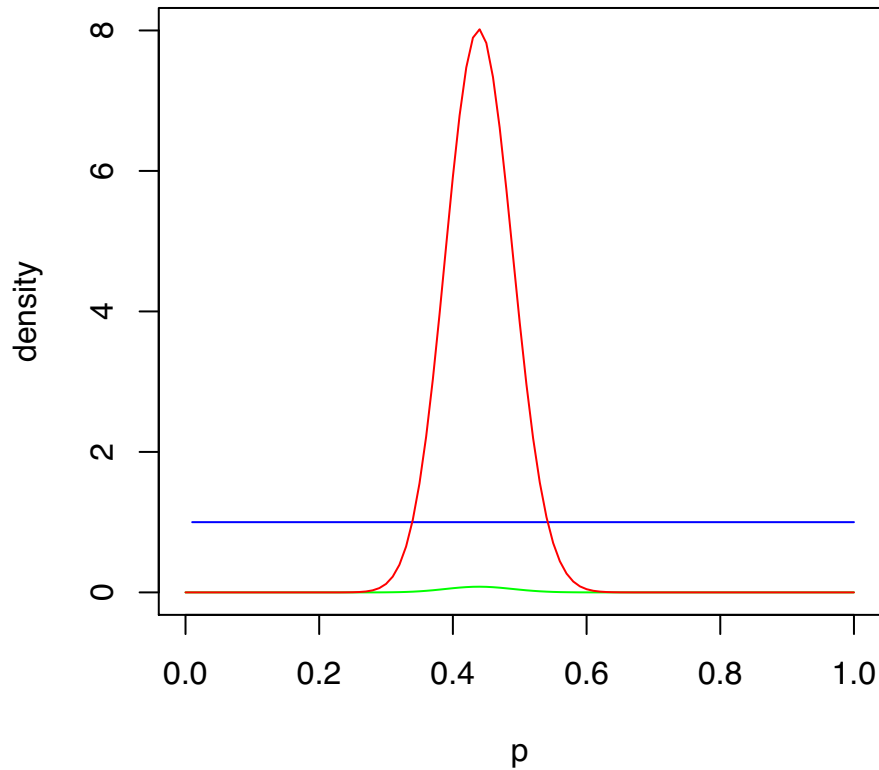
```
        xlab = "p", ylab = "likelihood", add = TRUE,  ylim = c(0, 8), col = 'green')

# posterior
a_post = 1 + 43
b_post = 1 + 98 - 43
curve(gamma(a_post + b_post)/gamma(a_post)/gamma(b_post) *
      p^(a_post - 1) * (1-p)^(b_post - 1), from = 0, to = 1, xname = "p",
      xlab = "p", ylab = "density", add = TRUE, ylim = c(0, 8), col = 'red')
```
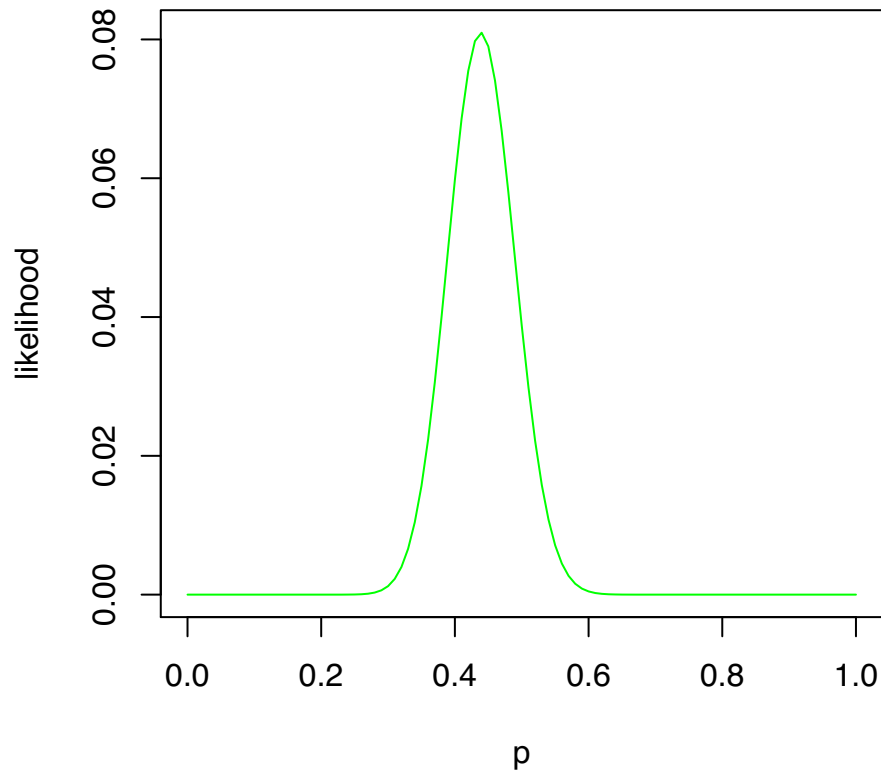

It helps to plot the likelihood alone.

```
# likelihood
curve(choose(98, 43) * p^43 * (1-p)^(98-43),
      from = 0, to = 1,  xname = "p",
      xlab = "p", ylab = "likelihood", col = 'green')
```

The posterior probability distribution contains all the current information about the parameter θ. A graphical report on the entire posterior distribution is definitely meaningful and useful. For many practical cases, however, various numerical summaries of the distributions are desirable.

- As for all distributions, the mean of the distribution is an important location summary. What is the mean of the posterior distribution of our above example?

- Meanwhile, the MAP (maximum a posteriori), which is the mode of the posterior distribution, can be interpreted as the single "most likely" value of the parameter, given the data and the model. What is the MAP estimate in our posterior distribution?

- Is there always a unique MAP value in the posterior distribution?

Computing probability intervals with quantile functions

In addition to point summaries, it is nearly always important to report posterior uncentainty. Therefore, as in conventional statistics, an interval summary is desirable. A central interval of posterior probability, which corresponds, in the case of a $100(1-\alpha)\%$ interval, to the range of values above and below which lies exactly $100(\alpha/2)\%$ of the posterior probability.

- What is the 95% central interval of the above posterior distribution?

```
alpha = 1 - 0.95
low = qbeta(alpha/2, a_post, b_post)
high = qbeta(1 - alpha/2, a_post, b_post)
print(c(low, high))
```

```
## [1] 0.3445430 0.5377312
```

- Visualize the above central interval

```
curve(gamma(a_post + b_post)/gamma(a_post)/gamma(b_post) *
      p^(a_post - 1) * (1-p)^(b_post - 1), from = 0, to = 1, xname = "p",
      xlab = "p", ylab = "density")
abline(v = low, col = "red", lty = 2)
abline(v = high, col = "red", lty = 2)
```
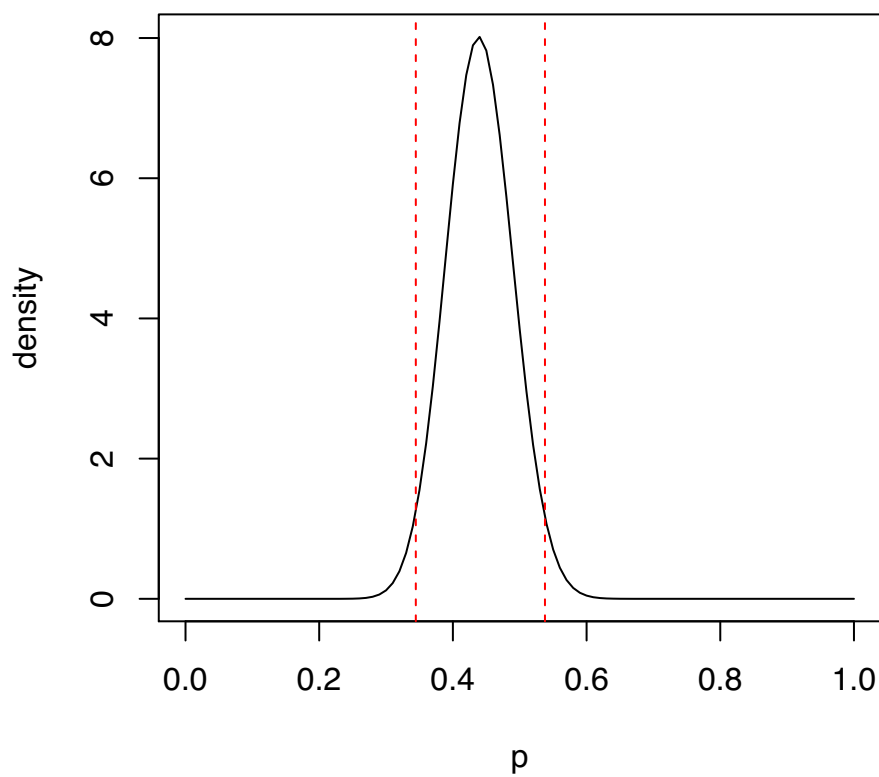

- How to interpret this central interval?