Stat 136 (Bayesian Statistics)

Lesson 1.3: Bayesian Inference for the Binomial Proportion: Continuous Prior

Motivation

- A limitation of specifying a discrete prior for p is when a plausible value is not specified in the prior distribution (e.g. $p = 0.2$), it will be assigned a 0 probability in the posterior distribution
- Ideally, we want a distribution that allows p to be any value in [0, 1]
- Two possible distributions come into mind:
  - Continuous uniform distribution in the interval [0,1]
  - Beta distribution

The continuous uniform distribution as prior

The probability density function of the continuous Uniform distribution on the interval [a, b] is

$$f(p) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq p \leq b \\ 0, & \text{elsewhere} \end{cases}$$

If we use $U(0,1)$ as a prior distribution, then

$$f(p) = \begin{cases} 1, & \text{if } 0 \leq p \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

Recall that the likelihood function for p is

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

Thus, based on the Bayes' rule, the posterior distribution is

$$f(p|y) = \binom{n}{y} p^y (1-p)^{n-y}$$

The Beta distribution as prior

The density function of the Beta distribution with parameters a and b is given by

$$f(p) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, \text{ if } 0 \leq p \leq 1 \\ 0, \text{elsewhere} \end{cases}$$

Hence, this distribution is suitable for modeling a random variable whose domain is between 0 and 1, say the proportion p.

Note that the $U(0,1)$ distribution is a special case of the Beta distribution with $a = b = 1$.

Recall that if $p \sim Beta(a,b)$, then

$$E(p) = \frac{a}{a+b}$$
$$Mode(p) = \frac{a-1}{a+b-2}, \text{if } a,b > 1$$
$$Var(p) = \frac{ab}{(a+b)^2(a+b+1)}$$

Multiplying the likelihood function for p given by

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

with the Beta prior

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

will give us the posterior density

$$f(p|y) \propto f(p) \times f(y|p) = p^{y+a-1}(1-p)^{n-y+b-1}$$

If we multiply the above product by

$$\frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)}$$

then the resulting pdf is that of a beta distribution with $a' = y + a$ and $b' = n - y + b$.

We have seen that if we use a Beta prior for the Binomial proportion, we will obtain a Beta posterior.

That is,

$$p \sim Beta(a,b)$$

and

$$Y \sim Binomial(n, p)$$

then

$$p|Y \sim Beta(y + a, n - y + b)$$

This result shows an important concept in Bayesian statistics: **Conjugate** distributions

Conjugate distribution or **conjugate** pair means a pair of a sampling distribution and a prior distribution for which the resulting posterior distribution belongs to same parametric family of distributions as the prior distribution.

How to determine the values of *a* and *b* of the Beta prior?

The parameters of the $Beta(a, b)$ prior are called as **hyperparameters**. These are chosen to reflect the researcher's prior beliefs about p. But it ios difficult to guess values of a and b in Beta(a, b). The solution is to specify quantile(s) of the beta distribution and solve for a and b. Recall that the q^{th} quantile is the value y of the random variable Y such that

$$P(Y \leq y_q) = q$$

In R, we use the *beta.select()* function in the **ProbBayes** package to find the parameters a and b of the Beta density curve.

An Example: Minahal Kita Resto

Suppose the restaurant owner uses his knowledge to specify the 0.5 and 0.9 quantiles of the distribution of the proportion p as follows:

- First, the restaurant owner thinks of a value q_{50} such that the proportion p is equally likely to be smaller or larger than q_{50} . After some thought, he thinks that $q_{50} = 0.55$.
- Next, the the restaurant owner thinks of a value q_{90} that he is prety sure (with probability 0.90) that the proportion p is smaller than q_{90}. After some more thought, he thinks that $q_{90} = 0.80$

```
library(ProbBayes)
beta.select(list(x=0.55,p=0.5),
            list(x=0.80,p=0.90))
```

```
## [1] 3.06 2.56
```

Since the prior distribution of p is Beta(3.06, 2.56), and the likelihood is Binomial (20, p), thus, the posterior is Beta(15.06, 10.56). Bayesian inference (point and credible interval estimates, and test of hypothesis) will be based on Beta(15.06, 10.56).In fact, a Bayesian point estimate of p is the mean of Beta(15.06, 10.56) which is

$$\hat{p} = \frac{15.06}{15.06 + 10.56} = 0.587822 \approx 0.59$$

Alternatively, we can simulate observations from Beta(15.06, 10.56) and compute the mean of these observations

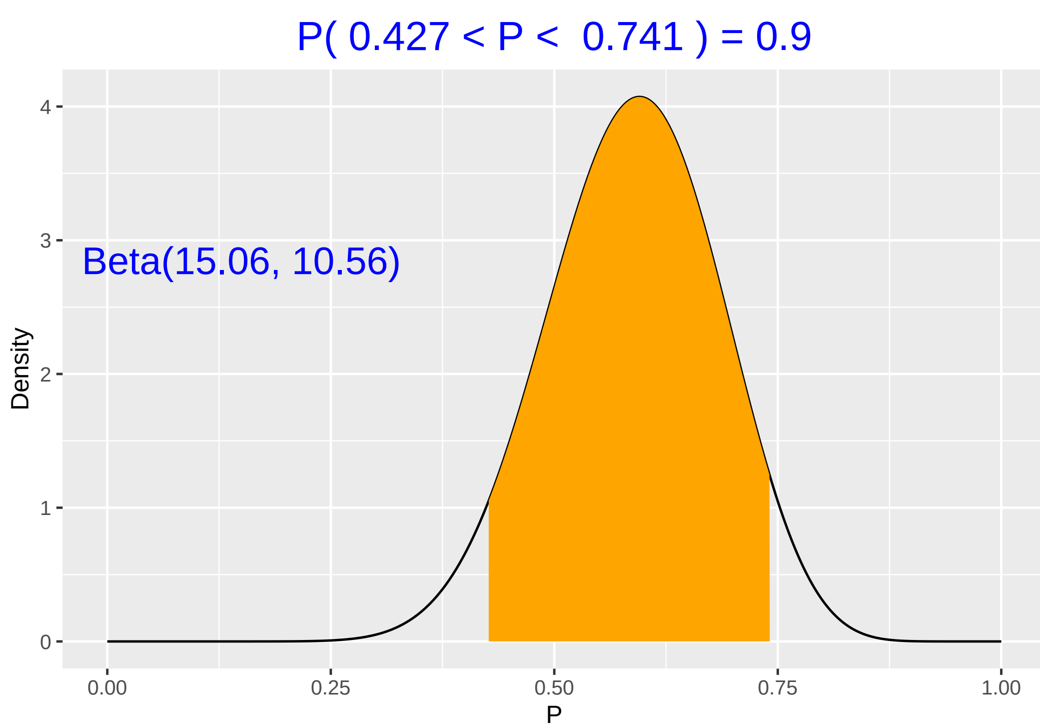
```
mean(rbeta(1000,15.06,10.56))
```

```
## [1] 0.589832
```

Another type of inference is a Bayesian credible interval, an interval that one is confident contains p. Such an interval provides an uncertainty estimate for the parameter p. A 90% Bayesian credible interval is an interval that contains 90% of the posterior probability. A (1- α)% credible interval for p can be obtained using the *beta_interval()* function in the **ProbBayes** package and the *qbeta()* function in the **stats** package.

```
#90% credible interval for p:

beta_interval(0.9,c(15.06, 10.56))
```



```r
qbeta(c(0.05, 0.95), 15.06, 10.56)
```

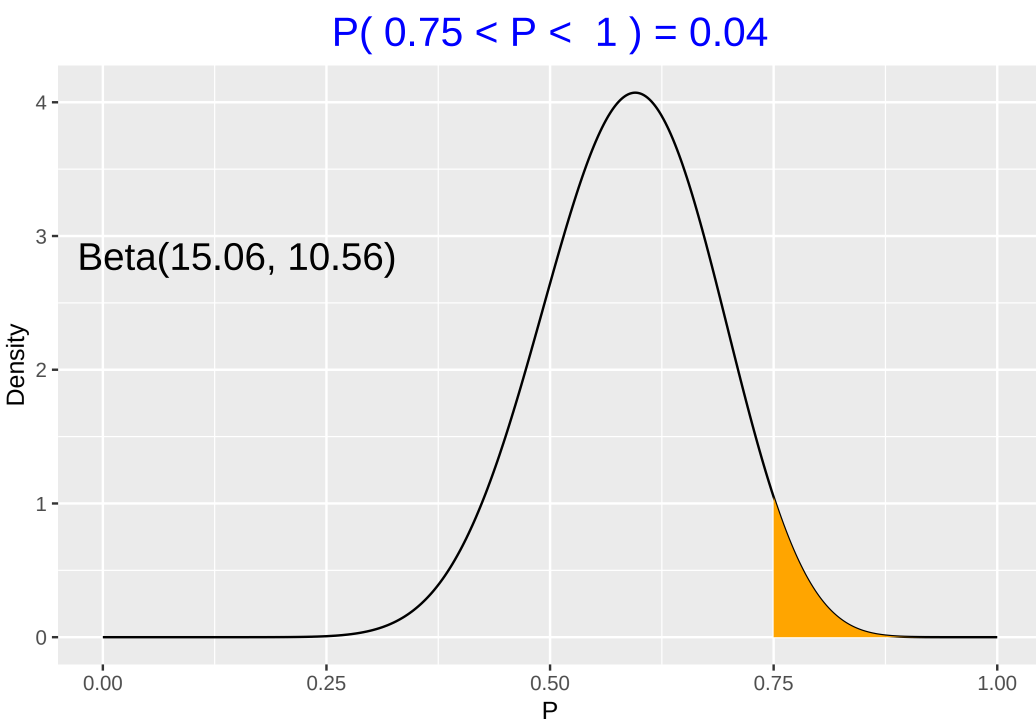
```
## [1] 0.4266788 0.7410141
```

Suppose one of the workers in Minahal Kita resto claims that at least 75% of the customers prefer Friday. Is this a reasonable claim?

From a Bayesian viewpoint, we find the posterior probability that $p \geq 0.75$ and make a decision based on the value of the posterior probability. If the probability is small, we can reject this claim.

In R, we can use the *beta_area()* function in the **ProbBayes** package. Use the *pbeta()* function in the **stats** package and simulate observations and calculate the proportion of the simulated values which are ≥ 0.75

```r
beta_area(lo = 0.75, hi = 1.0,
          shape_par = c(15.06, 10.56))
```




```r
1- pbeta(0.75,15.06, 10.56)
```

```
## [1] 0.03973022
```

Using the above code we have $P(p \geq 0.75)$=0.0397302 $\approx$ 0.04 We can also simulate random values from Beta(15.06,10.56).

```r
#simulates n=1000 observations
rval <- rbeta(1000, 15.06, 10.56)
prop <- sum(rval >= 0.75) / 1000
print(prop)
```

```
## [1] 0.041
```

In all cases, the $P(Y \geq 0.75) \approx 0.04$, thus, the claim that at least 75% of customers prefer Friday for a dinner out can be rejected at $\alpha = 0.05$.

Bayesian Prediction

Prediction is a typical task of Bayesian inference and statistical inference, in general. Once we are able to make inference about the parameter in our statistical model, one may be interested in predicting future observations.

For example, if a second survey is given by the owner of Minahal Kita Resto to m customers, what is expected number of customers who prefers Friday?

Let $\widetilde{Y}$ be a new observation. We want the probability function

$$f(\widetilde{Y} = \widetilde{y} | Y = y)$$

where $\widetilde{y}$ is a value of $\widetilde{Y}$. But the conditional distribution of $\widetilde{Y}$ given a value of the proportion p is Binomial(m, p) and the current beliefs about p are described by the posterior density. So the joint density of $\widetilde{Y}$ and p is

$$f(\widetilde{Y} = \widetilde{y}, p | Y = y) = f(\widetilde{Y} = \widetilde{y} | p) \pi(p | Y = y)$$

Thus, to get $f(\widetilde{Y} = \widetilde{y} | Y = y)$, we integrate out p. That is,

$$f(\widetilde{Y} = \widetilde{y} | Y = y) = \int f(\widetilde{Y} = \widetilde{y} | p) \pi(p | Y = y) dp$$

After the substitution of densities and an integration step, we get

$$f(\widetilde{Y} = \widetilde{y} | Y = y) = \binom{m}{\widetilde{y}} \frac{B(a + y + \widetilde{y}, b + n - y + m - \widetilde{y})}{B(a + y, b + n - y)}$$

This is the Beta-Binomial distribution with parameters $m, a + y, b + n - y$.

Recall that $B(a, b)$ is the beta function defined as

$$B(a, b) = \int_0^1 p^{a-1} (1 - p)^{b-1} dp = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}$$

In summary, Bayesian prediction distribution of a new observation is a Beta-Binomial distribution where m is the number of observations in the new sample, a and b are the parameters of the Beta prior, and y and n are quantities from the likelihood.

Let us use the Beta-Binomial distribution to compute the predictive probability that there are $\tilde{y}$ successes in a new survey of size m. We shall use the *pbetap()* function in the **ProbBayes** package for the calculation of predictive probabilities.

Suppose we survey an additional 20 customers. What is the probability that 12 of them will prefer Friday for a dinner out?

```
pred.prob <- pbetap(c(15.06,10.56),20,12)
print(pred.prob)
```

```
## [1] 0.1338812
```

It is desirable to construct an interval that will contain $\widetilde{Y}$ with a high probability. Suppose the desired probability content is 0.90. One constructs this prediction interval by putting in the most likely values of $\widetilde{Y}$ until the probability content of the set exceeds 0.90.

```
pred.prob.dist <- pbetap(c(15.06,10.56),20,0:20)
discint(cbind(0:20,pred.prob.dist), 0.90)
```

```
## $prob
## [1] 0.9185699
##
## $set
##  [1]  7  8  9 10 11 12 13 14 15 16
```

Thus, $P(7 \leq \widetilde{Y} \leq 16) \approx 0.92$

In some situations it is difficult to derive the exact predictive distribution. To solve this issue, we simulate values from the predictive distribution.

- Step 1: Simulate draws of the parameter (in this case the proportion (p) from its posterior distribution
- Step 2: Then simulating values of the future observation $\tilde{y}$ from the sampling density (say the Binomial distribution using the simulated p's in Step 1)

```
pred_p_sim <- rbeta(1, a + y, b + n - y)
pred_y_sim <- rbinom(1, n, pred_p_sim)
```

Let us generate 1000 proportions (*pred_p_sim*) from the posterior distribution. Input these simulated proportions into the Binomial pmf with m=20. Then get the average of the simulated random y-values to get the expected value of $\widetilde{Y}$.

```r
pred_p_sim <- rbeta(1000, 15.06, 10.56)
pred_y_sim <- rbinom(1000, 20, pred_p_sim)
print(mean(pred_y_sim))
```

```
## [1] 11.672
```

Predictive Checking: *Comparing Bayesian Models*

Suppose another restaurant worker is more pessimistic about the likelihood of people dining on Friday. The worker's prior median of the proportion p is 0.2 and her 90^{th} percentile is 0.4.

```r
beta.select(list(x=0.20,p=0.5),
            list(x=0.40,p=0.90))
```

```
## [1] 2.07 7.32
```

Based on the above code chunk we have $a = 2.07$ and $b = 7.32$. That is the prior is $Beta(2.07, 7.32)$. With the same likelihood for $Y = 12$ which is $Binomial(20, p)$ then the posterior is $Beta(14.07, 15.32)$. The mean of this posterior distribution is 0.48.

In fact, if we do prediction for $m = 20$ for this posterior distribution we also get a different result.

```r
pred_p_sim1 <- rbeta(1000, 14.07,15.32)
pred_y_sim1 <- rbinom(1000, 20, pred_p_sim1)
mean(pred_y_sim1)
```

```
## [1] 9.625
```

These results show that a the prior distribution has a very important contribution to the posterior distribution. So, which of the two priors is credible?

We compute the so-called **Bayes Factor** which is the ratio of the predictive densities of 2 models.

The function *binomial.beta.mix()* is used to find the Bayes factor. One inputs the prior probabilities of the two models (priors), and the vectors of Beta shape parameters that define the owner's prior and the worker's prior.

```r
probs <- c(0.5, 0.5)
# assumes that the 2 priors are equally plausible
beta_prior1 <- c(3.06, 2.56)
beta_prior2 <- c(2.07, 7.32)
beta_par <- rbind(beta_prior1, beta_prior2)
output <- binomial.beta.mix(probs, beta_par, c(12, 8))
posterior_odds <- output$probs[1] / output$probs[2]
print(posterior_odds)
```

```
## beta_prior1
##    6.777823
```

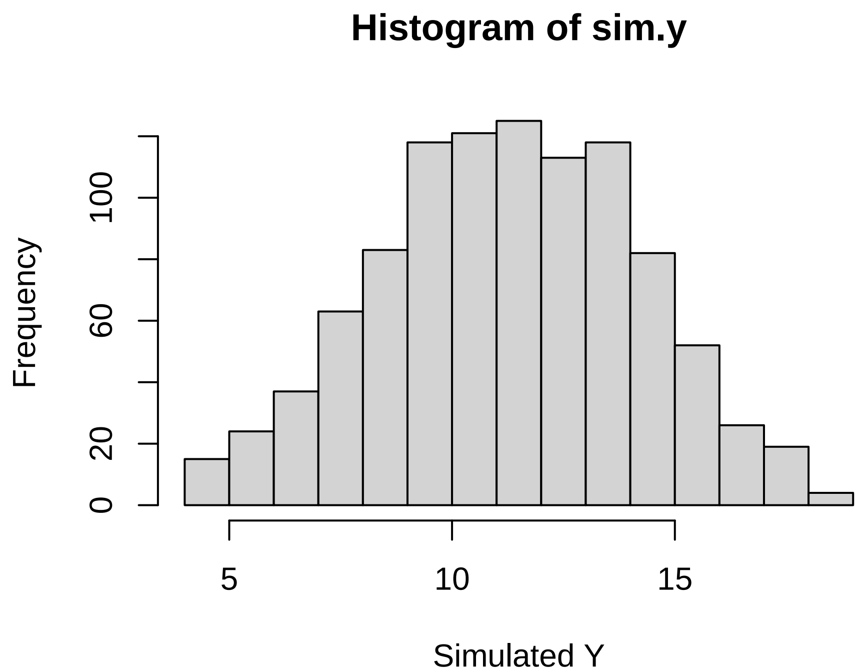
For the given observation (12 successes in 20 trials), there is 6.77 times more support for the owner's prior than for the worker's prior. Although the prior predictive distribution is useful in model checking, it has some disadvantages. For example, the prior predictive distribution may not exist in the situation where the prior is not a proper probability distribution. A related issue is that a prior may be assigned that may not accurately reflect one's prior beliefs about a parameter. An alternative method of checking the suitability of a Bayesian model is based on the posterior predictive distribution.

We compute the posterior predictive distribution of a replicated dataset, that is a dataset of the same sample size as our observed sample, then we check if the observed value of y is in the middle of this predictive distribution. If this is true, then this means that the observed sample is consistent with predictions of replicated data; else some model misspecification exists.

Posterior predictive checking

The process is similar to predictive checking wherein we simulate values of p from the posterior distribution and use these simulated proportions as input to simulate y values from the Binomial distribution.

```r
sim.p <- rbeta(1000, 15.06, 10.56)
sim.y <- rbinom(1000, 20, sim.p)
hist(sim.y,xlab="Simulated Y")
```


Histogram of sim.y
Frequency
0
20
60
100
5
10
15
Simulated Y