Stat 136 (Bayesian Statistics)

Lesson 2.2: Bayesian Inference for a Proportion Using a 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

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

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

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

The Beta distribution as prior

- Multiplying the likelihood function for p given by

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

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

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

- If we multiply the above product by

$$\frac{\Gamma(n+a+b)}{\Gamma(a+s)\Gamma(b+f)}$$

then the resulting pdf is that of a beta distribution with $a' = a + s$ and $b' = b + f$

From Beta prior to Beta posterior

- 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

  $$p|Y \sim Beta(a + s, b + f)$$

- 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 *hyperparameters*
- These are chosen to reflect the researcher's prior beliefs about p
- Difficult to guess values of a and b in Beta(a, b)
- The solution is to specify quantile(s) of the beta distribution
- 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

Revisiting the sleep patterns of graduating college students

- Suppose the researcher believes that the median and 90^{th} percentiles are given by 0.3 and 0.5, respectively.

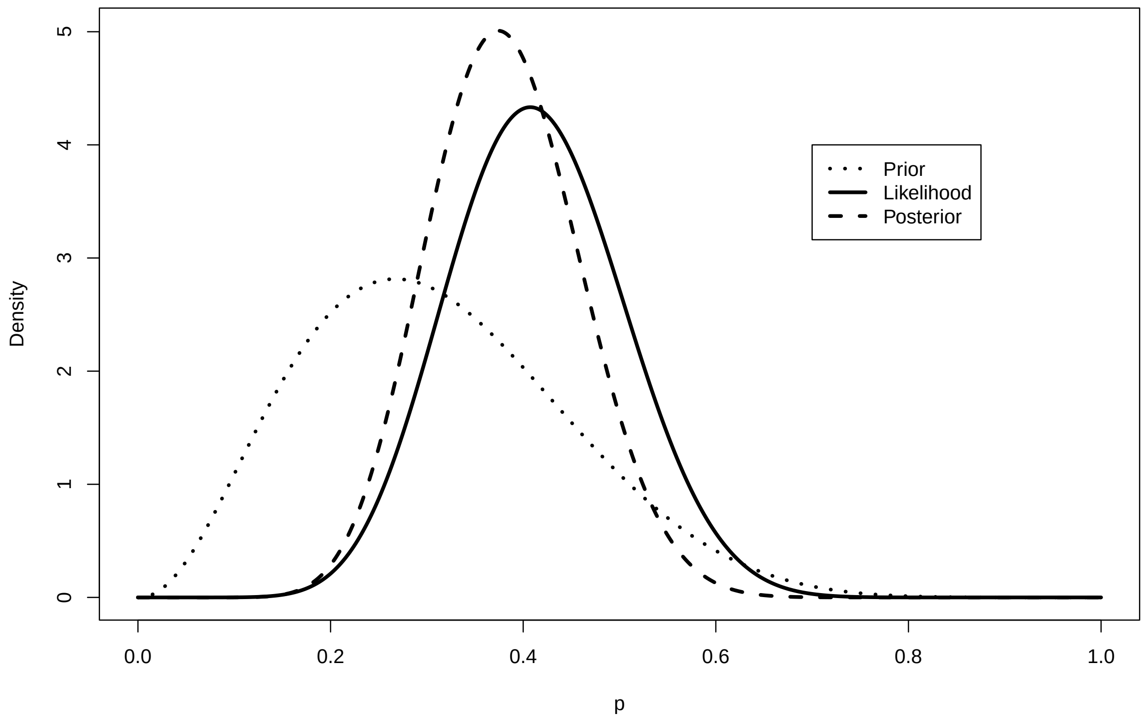
```
library(ProbBayes)
beta.select(list(x=0.3,p=0.50),
            list(x=0.5,p=0.90))
```

```
[1] 3.26 7.19
```

- The prior distribution of p is Beta(3.26, 7.19), the likelihood is Binomial (27, p)
- Thus, the posterior is Beta(14.26, 23.19)

```
p <- seq(0, 1, length = 1000)
a <- 3.26
b <- 7.19
s <- 11
f <- 16
prior <- dbeta(p, a, b)
like <- dbeta(p, s+1, f+1)
post <- dbeta(p, a+s, b+f)
plot(p, post, type = "l", ylab = "Density", lty = 2, lwd = 3)
lines(p, like, lty = 1, lwd = 3)
```

```r
lines(p, prior, lty = 3, lwd = 3)
legend(.7, 4, c("Prior", "Likelihood", "Posterior"),
       lty=c(3, 1, 2), lwd = c( 3, 3, 3))
```



- Bayesian inference (point and credible interval estimates, and test of hypothesis) will be based on Beta(14.26, 23.19)
- In fact, a Bayesian point estimate of p is the mean of Beta(14.26, 23.19) which is

$$\hat{p} = \frac{a'}{a' + b'} = \frac{14.26}{14.26 + 23.19} \approx 0.38$$

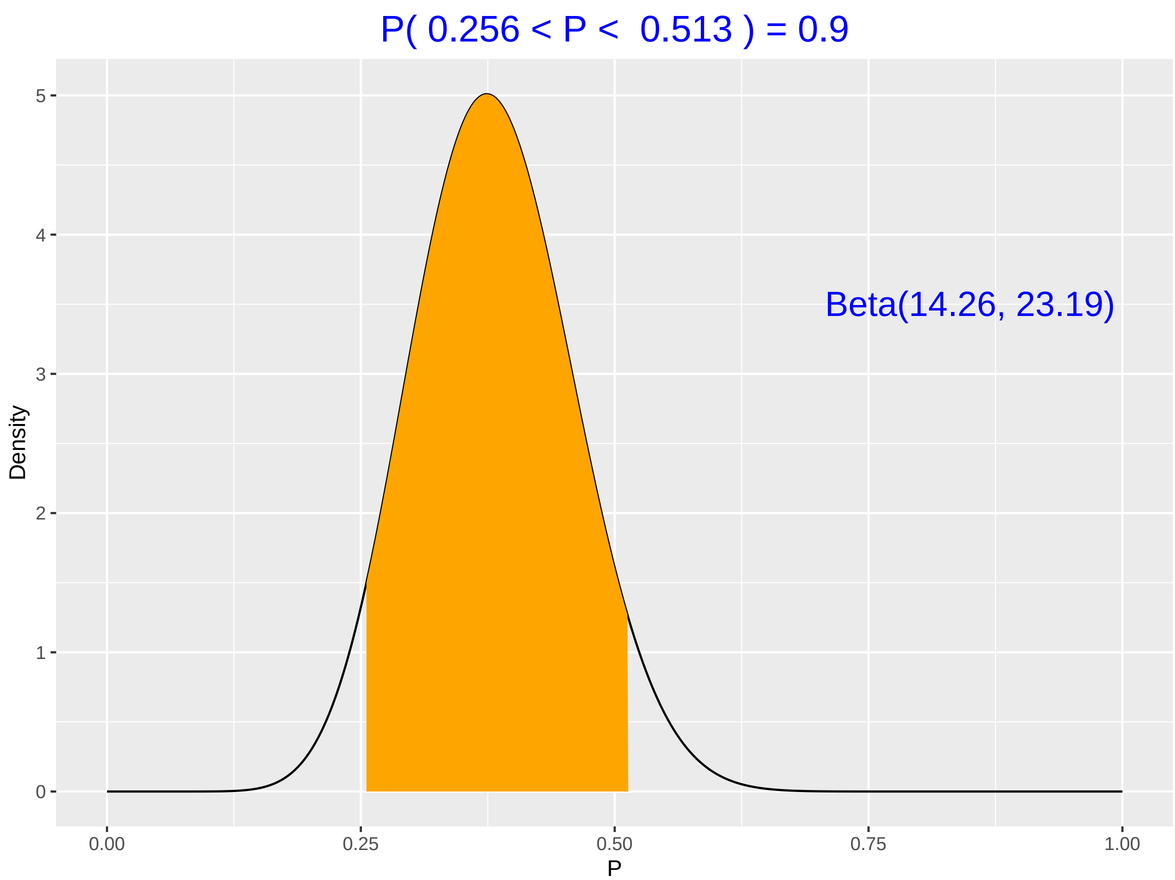
- Alternatively, we can simulate observations from Beta(14.26, 23.19) and compute the mean of these observations

```r
round(mean(rbeta(1000,14.26, 23.19)),2)
```

```
[1] 0.39
```

- 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 *qbeta()* function in the **stats** package and the **ProbBayes** package.

```
ProbBayes::beta_interval(0.9,c(14.26, 23.19))
```

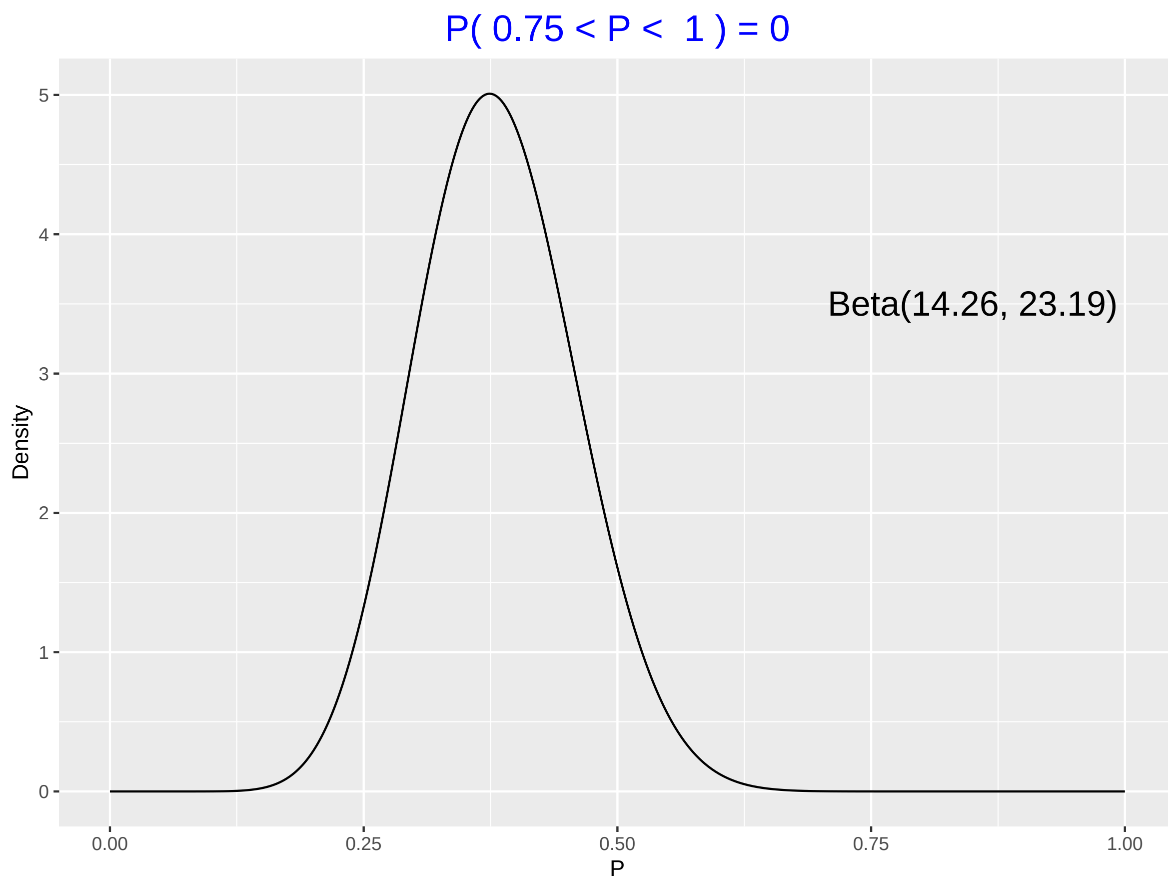


- Suppose it is claimed that at least 75% of graduating college students slept at least 8 hours every night. Is this a reasonable claim?
- From a Bayesian viewpoint,
  - Find the posterior probability that $p \geq 0.75$.

  - 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

```
ProbBayes::beta_area(lo = 0.75, hi = 1.0,
          shape_par = c(14.26, 23.19))
```



- Alternatively, we can use the *pbeta()* function in the **stats** package

```
1- pbeta(0.75,14.26, 23.19)
```

```
[1] 9.529512e-07
```

- Using the above code we have $P(p \geq 0.75) \approx 0$

- Or simulate observations and calculate the proportion of the simulated values which are ≥ 0.75

```
#simulates n=1000 observations
rval <- rbeta(1000, 14.26, 23.19)
prop <- sum(rval >= 0.75) / 1000
print(prop)
```

```
[1] 0
```

- In all cases, the $P(Y \geq 0.75) \approx 0$, thus, the claim that at least 75% of graduating college students slept at least 8 hours every night is very unlikely, almost impossible to happen.

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 another survey is conducted using a sample of m graduating students, what is expected number of students who would say that they have at least 8 hours of sleep per night?
- Let $\widetilde{Y}$ be a new observation. We want the probability function

$$f(\widetilde{Y} = \tilde{s} | Y = s)$$

  where $\tilde{s}$ 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

$$f(\widetilde{Y} = \tilde{s}, p | Y = s) = f(\widetilde{Y} = \tilde{s} | p) f(p | Y = s)$$

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

$$f(\widetilde{Y} = \tilde{s} | Y = s) = \int f(\widetilde{Y} = \tilde{s} | p) f(p | Y = s) dp$$

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

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

- This is the Beta-Binomial distribution with parameters $m, a + s, b + f$
- 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 s and n are quantities from the likelihood
- Let us use the Beta-Binomial distribution to compute the predictive probability that there are $\tilde{s}$ successes in a new survey of size m.
- We shall use the *pbetap()* function in the **ProbBayes** package for the calculation of predictive probabilities

Bayesian Prediction: an example

- Suppose we survey an additional $m = 20$ graduating students. What is the probability that $\tilde{s} = 12$ of them will say that they have at least 8 hours of sleep every night?

```
pred.prob <- pbetap(c(14.26, 23.19),20,12)
print(pred.prob)
```

```
[1] 0.03976734
```

- 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(14.26, 23.19),20,0:20)
discint(cbind(0:20,pred.prob.dist), 0.90)
```

```
$prob
[1] 0.9088212

$set
[1]  4  5  6  7  8  9 10 11 12
```

- Thus, $P(4 \leq \widetilde{Y} \leq 12) \approx 0.91$

Bayesian Prediction

- 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 $\tilde{s}$ from the sampling density (say the Binomial distribution using the simulated p's in Step 1)
- For example, let us generate 1000 proportions ($pred\_p\_sim$) from the posterior distribution
- Input these simulated proportions into the Binomial pmf with $m = 20$.
- Get the average of the simulated random y-values to get the expected value of $\widetilde{Y}$

```r
pred_p_sim <- rbeta(1000, 14.26, 23.19)
pred_y_sim <- rbinom(1000, 20, pred_p_sim)
print(mean(pred_y_sim))
```

```
[1] 7.736
```

Predictive Checking: *Comparing Bayesian Models*

- Suppose another researcher (B) is more pessimistic about the likelihood of students having at least 8 hours of sleep per night
- This researcher (B) believes that the prior median of the proportion p is 0.2 and the 90^{th} percentile is 0.4

```
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 distibution we also get a different result

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

```
[1] 9.488
```

- 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 two priors

```
probs <- c(0.5, 0.5)
# assumes that the 2 priors are equally plausible
beta_prior1 <- c(3.26, 7.19)
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)
```

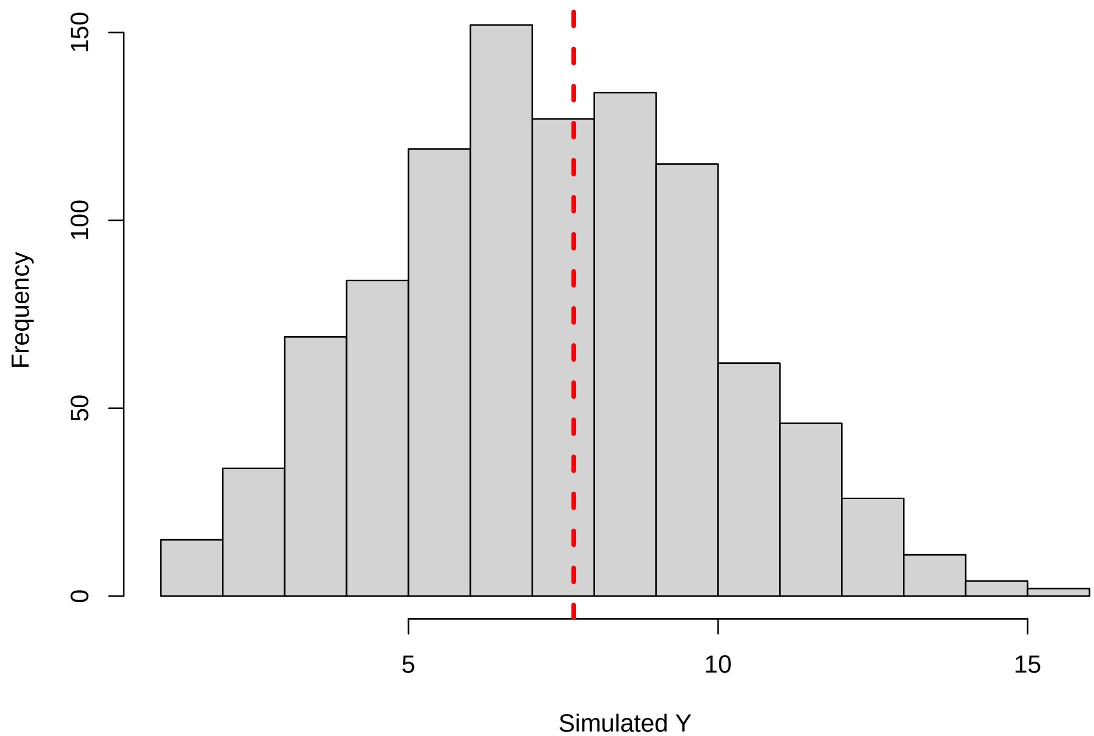
```
beta_prior1
   2.536484
```

- For the given observation (12 successes in 20 trials), there is about 2.54 times more support for the A's prior than for Bs prior.

Posterior predictive checking

- 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
- We then check if the observed value of s 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
- 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 s values from the Binomial distribution

```r
sim.p <- rbeta(1000, 14.26, 23.19)
sim.y <- rbinom(1000, 20, sim.p)
hist(sim.y,xlab="Simulated Y", main = " ")
abline(v = 7.668, col = "red", lwd = 3, lty = 2)
```


Frequency
0
50
100
150
5
10
15
Simulated Y