Bayes Homework4

Chen Xupeng

Textbook: # Chapter6:6

Variety of predictive reference sets: in the example of binary outcomes on page 147, it is assumed that the number of measurements, n, is fixed in advance, and so the hypothetical replications under the binomial model are performed with n = 20. Suppose instead that the protocol for measurement is to stop once 13 zeros have appeared.

    a. Explain why the posterior distribution of the parameter θ under the assumed model does not change.

using new protocol we have:

$$p(y, n|\theta) = \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}1_{(y_n=0)}1$$

For the given data it is $\theta^7(1-\theta)^{13}$, which is the likelihood. If the prior for θ is unchanged, the posterior distribution will not be changed

    b. Perform a posterior predictive check, using the same test quantity, T = number of switches, but simulating the replications y^{rep} under the new measurement protocol. Display the predictive simulations, $T(y^{rep})$, and discuss how they differ from Figure 6.5.

```
test <- NULL
for (i in 1:1000){
  theta <- rbeta (1,8,14)
  y.rep <- rbinom (1,1,theta)
  while (sum(y.rep==0) < 13)
    y.rep <- c(y.rep, rbinom(1,1,theta))
  n.rep <- length(y.rep)
  test <- c(test,
            sum (y.rep[2:n.rep] != y.rep[1:(n.rep-1)]))}
hist (test, xlab="T (y-rep)", yaxt="n", breaks=seq(-.5, max(test)+.5), cex=2)
```

Histogram of test
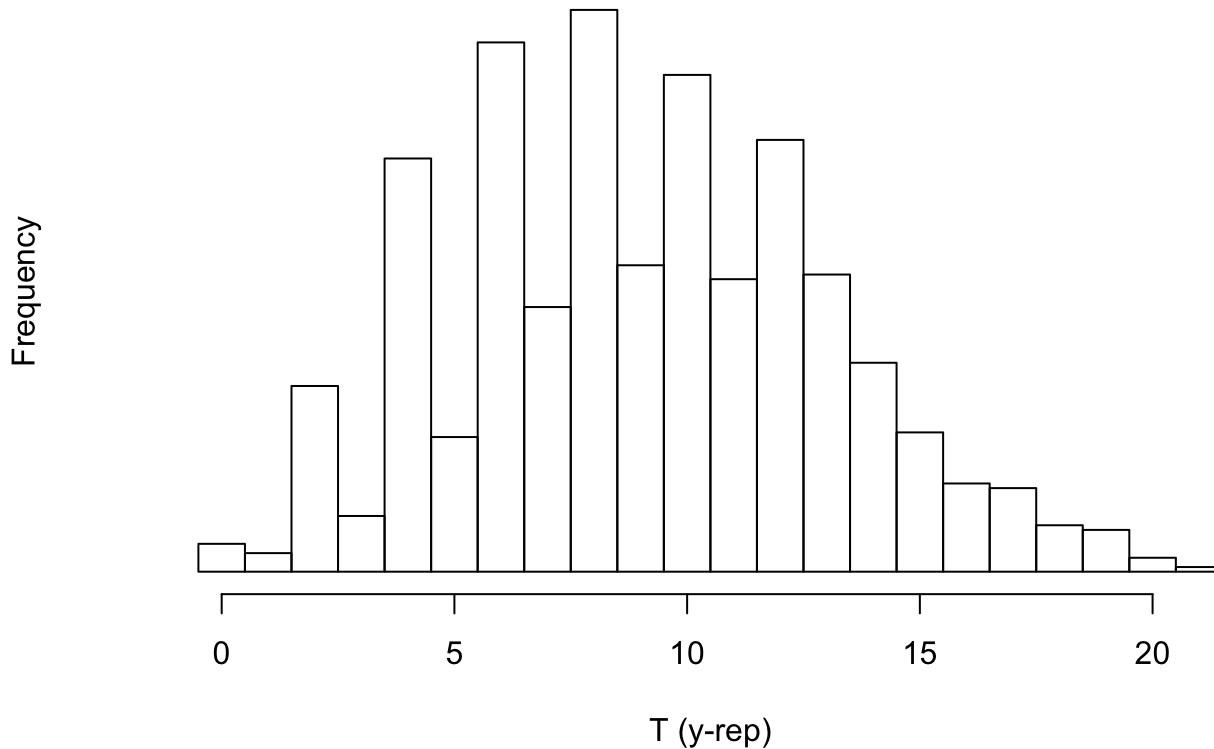
there are two differences: the data distribution is more dispersed because now n is also a random variable. and there are more even values because of particular data collection procedure.

Hypothesis Testing in Bayesian Framework (the homework mentioned in class):

Suppose we flip a coin, and get 9 heads and 3 tails. Denote θ as the probability of head. Consider

$$H_0 : \theta = 1/2 \textbf{ vs } H_1 : \theta > 1/2$$

Suppose we have no preference between null hypothesis and alternative hypothesis in prior. Please use Bayes Factor to compare these two hypothesis, under Binomial model and Negative binomial model, separately. Is there difference between these two results?

$$\text{Bayes factor } (H_2; H_1) = \frac{p(y|H_2)}{p(y|H_1)} = \frac{\int p(\theta_2|H_2) p(y|\theta_2, H_2) d\theta_2}{\int p(\theta_1|H_1) p(y|\theta_1, H_1) d\theta_1}$$

using bayes factor we can have posterior odds from prior odds

$$\frac{p(H_1|y)}{p(H_0|y)} = \frac{p(H_1)}{p(H_0)} \times \text{Bayes factor } (H_1; H_0)$$

For binomial model

$$p(y|H_0) = C_{12}^9 \cdot (\frac{1}{2})^{12}$$

$$p(y|H_1) = \int_{\frac{1}{2}}^1 p(y|\theta)p(\theta)d\theta$$

$$= \int_{\frac{1}{2}}^1 C_{12}^9 \theta^9 (1-\theta)^3 \cdot 2d\theta$$

```
p_h0 <- choose(12,9)*0.5**9*0.5**3
p_h1_function <- function(x) {choose(12,9)*x**9*(1-x)**3*2}
p_h1 <- integrate(p_h1_function, lower = 0.5, upper = 1)
p_h0
```

```
## [1] 0.05371094
```

```
p_h1
```

```
## 0.1467473 with absolute error < 1.6e-15
```

```
p_h1$value/p_h0
```

```
## [1] 2.732168
```

Bayes factor: $p(y|H_1)/p(y|H_0) = 2.732168$

from the bayes factor criteria, there is weak proof to support the alternative hypothesis that $\theta > \frac{1}{2}$.

for negative binomial distribution

$$p(y|H_0) = C_{11}^8 \cdot (\frac{1}{2})^{12}$$

$$p(y|H_1) = \int_{\frac{1}{2}}^1 p(y|\theta)p(\theta)d\theta$$

$$= \int_{\frac{1}{2}}^1 C_{11}^8 \theta^9 (1-\theta)^3 \cdot 2d\theta$$

```
p_h0 <- choose(11,8)*0.5**9*0.5**3
p_h1_function <- function(x) {choose(11,8)*x**9*(1-x)**3*2}
p_h1 <- integrate(p_h1_function, lower = 0.5, upper = 1)
p_h0
```

```
## [1] 0.0402832
```

```
p_h1
```

```
## 0.1100605 with absolute error < 1.2e-15
```

```
p_h1$value/p_h0
```

```
## [1] 2.732168
```

Bayes factor: $p\left(y|H_1\right)/p\left(y|H_0\right) = 2.732168$

from the bayes factor criteria, there is weak proof to support the alternative hypothesis that $\theta > \frac{1}{2}$.