Numerical approach: Discrete prior

Computing exercise

Suppose we have 15 independent trials and each trial results in one of two possible outcomes, success or failure. The probability of success remains constant for each trial. In that case, \( Y \mid \pi \) is binomial \((n = 15, \pi)\). Suppose that we observed \( y = 6 \) successes.

1. Calculate Bayesian estimator and 95\% credible interval with no information.

2. Calculate Bayesian estimator with Beta \((2, 4)\) prior.

Suppose the prior has the shape given by

\[
g(\pi) = \begin{cases} 
\pi & \text{for } \pi \leq .2, \\
.2 & \text{for } .2 < \pi < .3, \\
.5 - \pi & \text{for } .3 < \pi \leq .5, \\
0 & \text{for } .5 < \pi.
\end{cases}
\]

3. Plot three Bayesian estimators and \( \pi \) in the same graph.
Review

Bayesian view
+ our belief on the parameter is updated by data
+ posterior distribution of the parameter updated by data gives complete inference

Frequentist
+ likelihood of random sample = sampling distribution of data

Frequentist interpretation

Random sample
+ data is drawn from a population distribution with unknown and fixed \( p \)
+ sampling distribution ? conceptually, all possible random samples

Sampling Distribution

random sample : \((y_1, y_2, \ldots, y_n)\)
statistic : \(s = g(y_1, y_2, \ldots, y_n)\)
sampling distribution : \(f(s \mid \theta)\), \( \theta \) is fixed and unknown
+ measure how the statistic varies over all possible samples with the parameter

Likelihood
+ Joint PDF of random sample, data
+ function of the parameter, \( \theta \)
+ Frequentist : estimate the \( \theta \)
+ Bayesian : update the prior and get posterior \( \pi(\theta \mid data) \)

Point estimation

Frequentist
+ MLE : The value of parameter which maximizes the likelihood
+ MVUE : Factorization theorem, Rao Cramer Lower Bound, Rao-Blackwell Theorem
Bayesian
+ Posterior mean or median of posterior distribution of the $\theta$

Evaluating by Frequentist criteria
+ pre-data analysis: sampling dist. is treated as a random variable.
+ "what if the parameter has the value?"
  - estimation concept: max.
  - Bayesian: prior

Unbiased
+ Unbiased estimator: $E(\hat{\theta}) = \theta$, bias = 0
+ Bias: $B(\hat{\theta}) = \hat{\theta} - \theta$
+ Bayesian don't care of unbiased

MSE mean square of error
+ $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$: measure the precision of estimators
+ $MSE(\hat{\theta}) = V(\hat{\theta}) + B(\hat{\theta})^2$
+ MVUE: no bias, $MSE(\hat{\theta}) = V(\hat{\theta})$ estimator variance

Comparing estimators for proportion

Frequentist estimator $\hat{p}_i = \frac{y}{n}$, $y \sim B(n, p)$
+ $E(\hat{p}_i) = E(\frac{y}{n}) = p$, $V(\hat{p}_i) = V(\frac{y}{n}) = \frac{p(1-p)}{n}$
+ $MSE(\hat{p}_i) = \frac{p(1-p)}{n}$

Bayesian estimator depending prior
+ Uniform Prior $\pi(p) \sim Beta(1,1)$
+ Conjugate Prior $\pi(p) \sim Beta(a,b)$
  - posterior $\pi(p | data) \sim Beta(a^* = y + a, b^* = n - y + b)$
  - Bayesian estimator: $\hat{p}_B = \frac{y + a}{n + a + b} = \frac{y}{n + a + b} + \frac{a}{n + a + b}$
\[
\text{MSE}(\hat{p}_b) = V(\hat{p}_b) + B(\hat{p}_b)^2 = \left(\frac{1}{n+a+b}\right)^2 np(1-p) + \left(\frac{a-ap-bp}{n+a+b}\right)^2
\]

In R: (n=10), true p=0.3

```r
#uniform prior ~ Beta(1,1)
#conjugate prior ~ Beta(2,3)
n=10; y=seq(0,1,by=0.01); y=seq(0,10,by=1)

#Estimation
mle=y/n;
bayes_u=(y+1)/(n+1+1)
bayes_c=(y+2)/(n+2+3)

#Estimator 그래프
x11()
split.screen(c(1,1))
screen(1)
plot(y, mle, type='p', ylab='phat', main='Estimator', ylim=c(0,1), col="red")
screen(1)
plot(y, bayes_u, type='p', ylim=c(0,1), col="blue", ylab='phat')
screen(1)
plot(y, bayes_c, type='p', ylim=c(0,1), col="black", ylab='phat')
abline(h=c(0.3), lty=3)
```
When p is between 0.15~0.7, Conjugate prior is the best.

Overall Best

Uniform prior is the best. Therefore, when no information about p I available, the best choice is uniform prior.
Interval Estimation

Frequentist
+ pivot method: get an sampling distribution which has no parameter in the PDF
+ sampling distribution: \( Y \sim B(n, p) \) the parameter \( p \) is in the PDF
+ \( \frac{Y / n - p}{p(1 - p) / n} \sim z \) by CLT
+ 100(1-\( \alpha \))% Confidence interval: \( \frac{Y}{n} \pm \frac{z_{\alpha/2}}{n} \frac{y / n(1 - y / n)}{n} \)

Bayesian
+ We have exact posterior PDF and then, exact CI (called Credible Interval) exists
+ 100(1-\( \alpha \))% Credible interval: \( (B_{\alpha/2}^{-1}(a^*, b^*), B_{1-\alpha/2}^{-1}(a^*, b^*)) \)

Example 14 (continued from Chapter 8) Out of a random sample of \( n = 100 \) Hamilton residents, \( y = 26 \) said they support building a casino in Hamilton.

Bart’s uniform prior => posterior= \( Beta(27,75) \)

Hypothesis testing

One-sided test
+ null hypothesis: \( H_0: p = p_0 \)
+ alternative hypothesis: \( H_a: p > p_0 \) (or \( p < p_0 \))
+ \( y^* \) is observed out of \( n \) Bernoulli trials
+ (Frequentist) \( p \)-value = \( P(Y \geq y^* \mid Y \sim B(n, p_0)) \)
+ (Frequentist) \( \alpha \approx P(Y \geq y^{CR} \mid Y \sim B(n, p_0)) \) where \( y^{CR} \) is the critical value for the alternative hypothesis, \( p > p_0 \)
+ (Bayesian) \( P(H_0 : \pi \leq \pi_0 | y^*) = \int_{0}^{\pi_0} \pi(\pi | y^*)d\pi \) is compared to the significant level.

**Example 15** Suppose we wish to determine if a new treatment is better than the standard treatment. If so, \( \pi \), the proportion of patients who benefit from the new treatment, should be better than \( \pi_0 \), the proportion who benefit from the standard treatment. It is known from historical records that \( \pi_0 = 0.6 \). A random group of 10 patients are given the new treatment. \( Y \), the number who benefit from the treatment will be binomial \((n, \pi)\). We observe \( y = 8 \) patients benefit. This is better than we would expect if \( \pi = 0.6 \). But, is it enough better for us to conclude that \( \pi > 0.6 \) at the 10% level of significance?

+ Null hypothesis : \( H_0 : \pi = 0.6 \)
+ Alternative hypothesis : \( H_0 : \pi > 0.6 \)

```r
# hypothesis testing : alpha=5%
n=10; y=8
1-pbinom(7,10,0.6) # p-value > 1-pbinom(7,10,0.6) # p-value

[1] 0.1672898
```

There is no strong evidence that \( \pi > 0.6 \).

**Example 15 (continued)** Suppose we use a beta \((1, 1)\) prior for \( \pi \). Then given \( y = 8 \), the posterior density is beta \((9, 3)\).
```r
> pbeta(0.6,9,3) # p-value for Bayesian

[1] 0.1189168
```

**Two-sided test**

+ null hypothesis : \( H_0 : \pi = \pi_0 \) vs. alternative hypothesis : \( H_a : \pi \neq \pi_0 \)
+ \( y^* \) is observed out of \( n \) Bernoulli trials
+ (Frequentist) \( p\text{-value} = P(Y \geq y^* | Y \sim B(n, \pi_0)) * 2 \) (when \( y^* > n\pi_0 \))
+ (Frequentist) \( \alpha / 2 \geq P(Y \geq y^C_{\text{CR}} | Y \sim B(n, \pi_0)) \) where \( y^C_{\text{CR}} \) is the critical value
+ (Bayesian) using 100(1-\( \alpha \))% credible interval.

**Example 16** A coin is tossed 15 times, and we observe 10 heads. Are 10 heads out of 15 tosses enough to determine that the coin is not fair? In other words, is \( \pi \) the probability of getting a head different than \( \frac{1}{2} \)?

**Example 16 (continued)** If we use a uniform prior distribution, the posterior is the beta \((10 + 1, 5 + 1)\) distribution.
```r
# hypothesis testing : alpha=5%
n=15; y=10
2*(1-pbinom(y-1,n,0.5)) # p-value > 2*(1-pbinom(y-1,n,0.5)) # p-value

[1] 0.3017578

> qbeta(0.025,11,6) # 95% lower credible interval

[1] 0.4133794

qbeta(0.025,11,6) # 95% lower credible interval > qbeta(0.025,11,6) # 95%

[1] 0.8480163
```

Professor Kwon, Sehyung | Dept. of Statistics, HANNAM University

http://wolfpack.hnu.ac.kr
Exercise #1

The standard method of screening for a disease fails to detect the presence of the disease in 15% of the patients who actually do have the disease. A new method of screening for the presence of the disease has been developed. A random sample of \( n = 75 \) patients who are known to have the disease is screened using the new method. Let \( \pi \) be the probability the new screening method fails to detect the disease.

(a) What is the distribution of \( y \), the number of times the new screening method fails to detect the disease?

(b) Of these \( n = 75 \) patients, the new method failed to detect the disease in \( y = 6 \) cases. What is the frequentist estimator of \( \pi \)?

(c) Use a \textit{beta} (1, 6) prior for \( \pi \). Find \( g(\pi|y) \), the posterior distribution of \( \pi \).

(d) Find the posterior mean and variance.

(e) If \( \pi \geq .15 \), then the new screening method is no better than the standard method. Test

\[
H_0 : \pi \geq .15 \quad \text{versus} \quad H_1 : \pi < .15
\]

at the 5% level of significance in a Bayesian manner.

Exercise #2

In the same study of water quality, \( n = 145 \) samples were taken from streams having a high environmental impact from dairying. Out of these \( y = 9 \) had a high \textit{Campylobacter} level. Let \( \pi \) be the true probability that a sample of water from this type of stream has a high \textit{Campylobacter} level.

(a) Find the frequentist estimator for \( \pi \).

(b) Use a \textit{beta} (1, 10) prior for \( \pi \). Calculate the posterior distribution \( g(\pi|y) \).

(c) Find the posterior mean and variance. What is the Bayesian estimator for \( \pi \)?

(d) Find a 95% credible interval for \( \pi \).

(e) Test the hypothesis

\[
H_0 : \pi = .10 \quad \text{versus} \quad H_1 : \pi \neq .10
\]

at the 5% level of significance.
Monte Carlo

We will perform a Monte Carlo study approximating the sampling distribution of two estimators of $\pi$. The frequentist estimator we will use is $\hat{\pi}_f = \frac{\hat{y}}{n}$, the sample proportion. The Bayesian estimator we will use is $\hat{\pi}_B = \frac{\hat{y} + 1}{n + 1}$, which equals the posterior mean when we used a uniform prior for $\pi$. We will compare the sampling distributions (in terms of bias, variance, and mean squared error) of the two estimators over a range of $\pi$ values from 0 to 1.

(a) For $\pi = 0.1, 0.2, \ldots, 0.9$

i. Draw 5000 random samples from binomial ($n = 10, \pi$).
ii. Calculate the frequentist estimator $\hat{\pi}_f = \frac{\hat{y}}{n}$ for each of the 5000 samples.
iii. Calculate the Bayesian estimator $\hat{\pi}_B = \frac{\hat{y} + 1}{n + 2}$ for each of the 5000 samples.
iv. Calculate the means of these estimators over the 5000 samples, and subtract $\pi$ to give the biases of the two estimators. Note that this is a function of $\pi$.
v. Calculate the variances of these estimators over the 5000 samples. Note that this is also a function of $\pi$.
vi. Calculate the mean squared error of these estimators over the 5000 samples. The first way is

$$MS(\hat{\pi}) = (bias(\hat{\pi}))^2 + Var(\hat{\pi}).$$

The second way is to take the sample mean of the squared distance the estimator is away from the true value over all 5000 samples. Do it both ways, and see that they give the same result.

(b) Plot the biases of the two estimators versus $\pi$ at those values and connect the adjacent points. (Put both estimators on the same graph.)

i. Does the frequentist estimator appear to be unbiased over the range of $\pi$ values?
ii. Does the Bayesian estimator appear to be unbiased over the range of the $\pi$ values?

(c) Plot the mean squared errors of the two estimators versus $\pi$ over the range of $\pi$ values, connecting adjacent points. (Put both estimators on the same graph.)

i. Does your graph resemble Figure 9.2?
ii. Over what range of $\pi$ values does the Bayesian estimator have smaller mean squared error than that of the frequentist estimator?