5. 이항분포에 대한 베이지안 추론

이경재

인하대학교 통계학과

March 27, 2019

강의 목표

- Bayesian Inference for Binomial Distribution
- Parameter Estimation
  - Point Estimation
  - Credible Interval
- Prediction

Beta Posterior Distribution

- Consider 40 flips of a coin having $\Pr(Heads) = \theta$.
- We model the count of heads as binomial:

$$p(X = x \mid \theta) = \binom{40}{x} \theta^x (1 - \theta)^{40-x}, \quad x = 0, 1, \ldots, 40.$$

- Let's use a uniform prior for θ:

$$p(\theta) = 1, \quad 0 \leq \theta \leq 1.$$

Beta Posterior Distribution

- Then the posterior is:

$$\begin{aligned}\pi(\theta \mid x) &\propto p(\theta)L(\theta \mid x) \\ &\propto \theta^{x}(1-\theta)^{40-x} \\ &\propto \theta^{x+1-1}(1-\theta)^{40-x+1-1}, \quad 0 \leq \theta \leq 1.\end{aligned}$$

- Thus, $\theta \mid x \sim Beta(x+1, 40-x+1)$. (Why?)
- Suppose we observe 15 “heads”, i.e., $x = 15$ here.

  Then $\theta \mid x \sim Beta(16, 26)$.
- Then the point estimation for θ is:

$$\begin{aligned}\mathrm{Mode}(\theta \mid x) &= 15/(15+25) = 0.375 \\ \mathrm{E}(\theta \mid x) &= 16/(16+26) = 0.381 \\ \mathrm{Var}(\theta \mid x) &= 0.00548.\end{aligned}$$

Beta Posterior Distribution

- Then the posterior is:

$$
\begin{aligned}
\pi(\theta \mid x) &\propto p(\theta) L(\theta \mid x) \\
&\propto \theta^{x}(1-\theta)^{40-x} \\
&\propto \theta^{x+1-1}(1-\theta)^{40-x+1-1}, \quad 0 \leq \theta \leq 1.
\end{aligned}
$$

- Thus, $\theta \mid x \sim Beta(x+1, 40-x+1)$. (Why?)
- Suppose we observe 15 "heads", i.e., $x = 15$ here.

  Then $\theta \mid x \sim Beta(16, 26)$.
- Then the point estimation for θ is:

$$
\begin{aligned}
\mathrm{Mode}(\theta \mid x) &= 15/(15+25) = 0.375 \\
\mathrm{E}(\theta \mid x) &= 16/(16+26) = 0.381 \\
\mathrm{Var}(\theta \mid x) &= 0.00548.
\end{aligned}
$$

Beta Posterior Distribution

- Posterior distribution is a combination of prior information of θ and data.
- In this example,
  - Prior: 특정한 θ에 차별을 두지 않는다.
  - Data: θ가 0.375에 가까울 수록 확률이 높다.

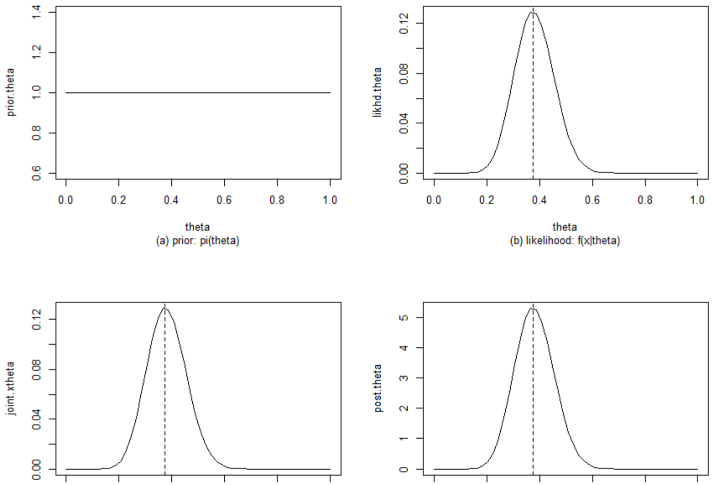
Prior vs Likelihood vs Posterior (R 실습)

```
> # theta ~ Beta(a, b)
> a=1 ; b=1
> # x|theta - B(n, theta)
> n=40 ; x=15
> # a discretization of the possible theta values
> theta = seq(0, 1, length=50)
> prior.theta = dbeta(theta, a, b)
> # prob of data\theta(likelihood)
> likhd.theta = dbinom ( x, n, theta)
> # joint prob of data & theta
> joint.xtheta = prior.theta*likhd.theta
> # posterior of theta
> post.theta = dbeta(theta, a+x, b+n-x)
```

Prior vs Likelihood vs Posterior (R 실습)

```
par (mfrow=c(2, 2)) # set up a 2x2 plotting window plot
plot (theta, prior.theta, type="l",
sub="(a) prior: pi(theta)")
plot(theta, likhd.theta, type="l",
sub="(b) likelihood: f(x|theta)")
abline(v=x/n, lty=2)
plot(theta, joint.xtheta, type="l",
sub="(c) prior x likelihood: pi (theta)x f(x|theta)")
abline(v=(a+x-1)/(a+b+n-2), lty=2)
plot (theta, post.theta, type="l",
sub="(d) posterior: pi(theta|x)")
abline(v=(a+x-1)/(a+b+n-2), lty=2)
```

Prior vs Likelihood vs Posterior (R 실습)



(a) prior: pi(theta)

(b) likelihood: f(x|theta)

(c) prior x likelihood: pi (theta)x f(x|theta)

(d) posterior: pi(theta|x)

Prior vs Likelihood vs Posterior (R 실습)

- (b) Likelihood와 (c) Uniform $\times$ Likelihood는 동일

  $\Longrightarrow f(x \mid \theta) = f(x \mid \theta)\pi(\theta)$ since $\pi(\theta) = 1$

- (d) Posterior와 (c) Uniform $\times$ Likelihood은 세로축만 다름

  $\Longrightarrow \pi(\theta \mid x) = f(x \mid \theta)\pi(\theta)/p(x) \propto f(x \mid \theta)\pi(\theta)$

- (d) Posterior와 (b) Likelihood의 평균은 다름

  $\Longrightarrow \pi(\theta \mid x) \neq f(x \mid \theta)$ (and not even proportional)

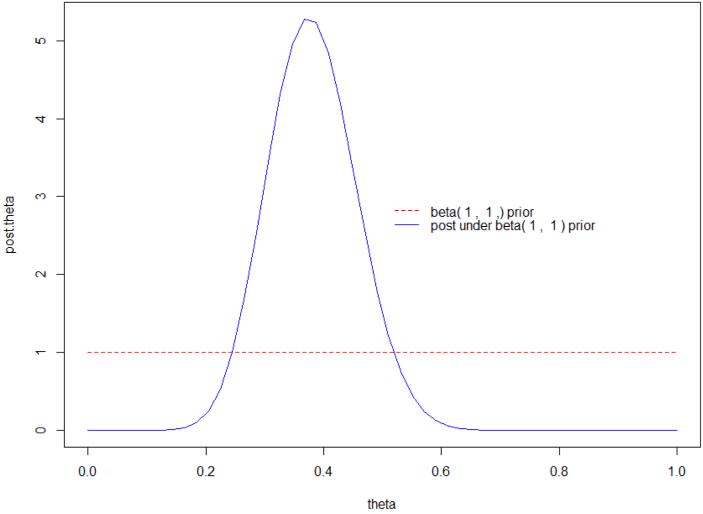
Prior vs Likelihood vs Posterior (R 실습)

우리의 사전적인 믿음(Prior)이 자료를 관측한 뒤 어떻게 바뀌었는가(Posterior)?

```
> par(mfrow=c(1, 1))
> plot(theta, post.theta, type="l", col="blue")
> lines(theta, prior.theta, col="red", lty=2)

> legend(.5, 3, legend=c(paste("beta(",a,", ",b,",) prior"),
+        paste("post under beta(",a, ", ",b,") prior")),
+        lty=c(2, 1), col=c("red", "blue"), bty="n")
```

Prior vs Likelihood vs Posterior (R 실습)



Prior vs Likelihood vs Posterior (R 실습)

- 사전정보: 어떤 특정한 θ 에 대하여 차별을 두지 않음
- 사후정보: θ가 0.375에 가까운 값일 확률이 매우 높음

Monte Carlo

- It is often very difficult to find the posterior distribution.
- Solution: Monte Carlo Method
- We can find information of the posterior based on the posterior samples.

Monte Carlo

$$I = \int_a^b h(x)dx$$

- 주어진 함수 $h(x)$에 대한 적분을 계산하고 싶음. 하지만 적분을 직접 하는 것이 불가능하다고 가정

Monte Carlo

$$I = \int_a^b h(x)dx = \int_a^b g(x)f(x)dx$$

- $f(x)$라는 분포로부터 샘플을 얻는 것이 가능하다고 가정
- $g(x) = h(x)/f(x)$라고 가정하면 원래 구하려던 적분 I를 다음과 같이 생각 가능

Monte Carlo

$$
\begin{aligned}
I = \int_a^b h(x)dx &= \int_a^b g(x)f(x)dx \\
&\approx \frac{1}{N}\sum_{i=1}^{N} g(X_i), \quad X_i \sim f
\end{aligned}
$$

- f라는 분포를 따르는 확률변수 X의 평균 $\mathrm{E}X$를 구하려 할 때, 다음과 같이 표본평균으로 근사할 수 있음: $\mathrm{E}X \approx \frac{1}{N}\sum_{i=1}^{N} x_i$
- 표본의 크기 N이 커질 수록 더 정확한 근사가 됨

위와 같이, 계산하고자 하는 적분을 해당 분포로부터 얻은 샘플을 이용해 근사하는 방법을 몬테 카를로(Monte Carlo)라 함

Monte Carlo: 예제

$$I = \int_{-1}^{1} x^2 dx$$

- $h(x) = x^2$인 경우

Monte Carlo: 예제

$$
\begin{aligned}
I = \int_{-1}^{1} x^2 dx \quad &= \quad \int_{-1}^{1} 2x^2 \cdot f(x) dx \\
&\approx \quad \frac{1}{N} \sum_{i=1}^{N} 2X_i^2, \quad X_i \sim Unif(-1, 1)
\end{aligned}
$$

- $f(x) = \frac{1}{2}, \quad -1 < x < 1$: f는 -1에서 1 사이의 균일분포. 이는 아래와 같이 표현할 수 있음

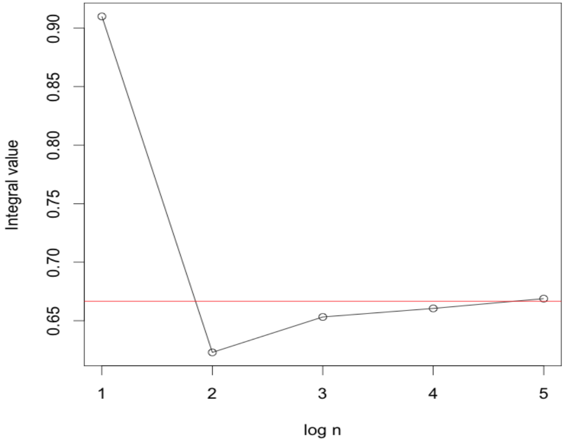
$$
X \sim Unif(-1, 1)
$$

Monte Carlo: 예제 (R code)

```
n = c(10, 100, 1000, 10000, 100000) # N values
int = rep(0, 5) # integral

for(i in 1:5){
xval = runif(n[i], min=-1, max=1)
int[i] = mean(2*xval^2)
}
```

Monte Carlo: 예제



Monte Carlo Method (R 실습)

사후분포 통계량들을 구할 때 Monte Carlo 방법을 이용해보자.

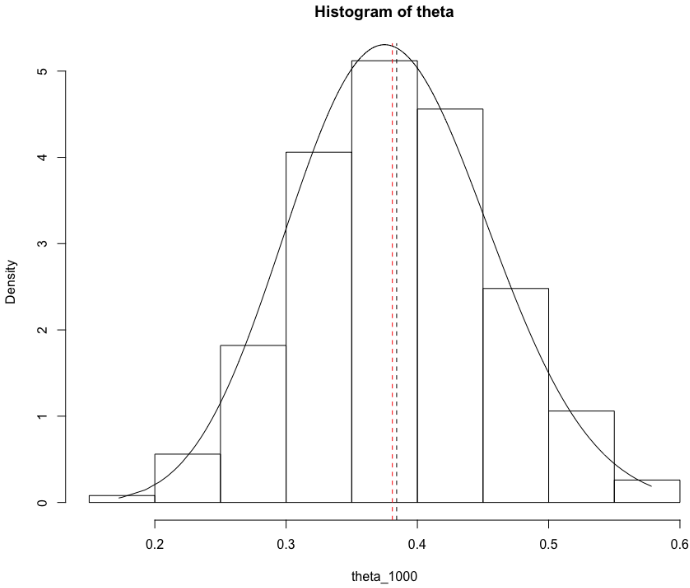
```
> theta_1000 = rbeta(1000, a+x, b+n-x) # generate posterior samples
> quantile(theta_1000, c(.025, .975)) # simulation-based quantiles
2.5%      97.5%
0.2412677 0.5268378
> qbeta(c(.025, .975), a+x, b+n-x) # theoretical quantiles
[1] 0.2420110 0.5306375
> mean(theta); var(theta)  # simulation-based estimates
[1] 0.379344
[1] 0.005324879
> # theoretical estimates
> (a+x)/(a+b+n); (a+x)*(b+n-x)/((a+b+n+1)*(a+b+n)^2)
[1] 0.3809524
[1] 0.005484364
```

Monte Carlo Method (R 실습)

1000개의 샘플에 기반하여 그린 사후분포와 실제 사후분포의 비교

```
> hist(theta_1000, prob=T, main="Histogram of theta")
   # simulation-based density
> theta_1000 = theta_1000[order(theta_1000)]
> lines(theta_1000, dbeta(theta_1000, a+x, b+n-x))
   # theoretical density
> mean.theta = mean(theta_1000)
> abline(v=mean.theta, lty=2)
> abline(v=(a+x)/(a+b+n), lty=2, col = "red")
```

Monte Carlo Method

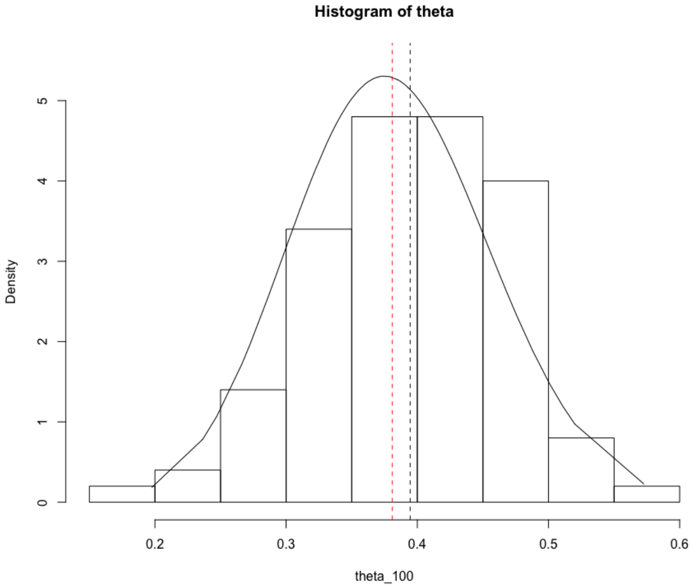


Monte Carlo Method (R 실습)

100개의 샘플에 기반하여 그린 사후분포와 실제 사후분포의 비교

```
> theta_100 = rbeta(100, a+x, b+n-x) # generate posterior samples
> hist(theta_100, prob=T, main="Histogram of theta", ylim=c(0,5.5))
   # simulation-based density
> theta_100 = theta_100[order(theta_100)]
> lines(theta_100, dbeta(theta_100, a+x, b+n-x))
   # theoretical density
> mean.theta = mean(theta_100)
> abline(v=mean.theta, lty=2)
> abline(v=(a+x)/(a+b+n), lty=2, col = "red")
```

Monte Carlo Method



Beauty of Monte Carlo

- Suppose we are interested in the log odds ratio,

$$\eta = \log\left(\frac{\theta}{1-\theta}\right).$$

- In this case, it might be difficult to calculate the posterior of η directly.
- Instead, we can simply
  1. obtain posterior samples $\theta^{(i)}$,
  2. use the transformation $\eta^{(i)} = \log(\theta^{(i)}/(1-\theta^{(i)}))$ and
  3. apply the Monte Carlo method ($\mathrm{E}(\eta) = \mathrm{E}\left[\log(\frac{\theta}{1-\theta})\right]$).

Monte Carlo Method (R 실습)

```
> ## log odds ratio
> a=b=1
> X=15;n=40
> theta=rbeta(10000,a+x,b+n-x)
> eta=log(theta/(1-theta))
> hist(eta, prob=T, main="Histogram of eta")
> lines(density(eta), lty=2)
> mean(eta); var(eta)
[1] -0.4947897
[1] 0.1035466
```

Beta/Binomial Bayesian Model

- Suppose we observe

$$X_1, \dots, X_n \mid p \overset{iid}{\sim} Ber(p).$$

  We wish to estimate the "success probability" p via the Bayesian approach.

- We will use a prior $p \sim Beta(a, b)$.
- Note that $Y = \sum_{i=1}^{n} X_i \sim B(n, p)$.
- We first write the joint density of Y and p.

Complete Derivation of Beta/Binomial Model

$$
\begin{aligned}
f(y,p) &= f(y \mid p)\pi(p) \\
&= \left[\binom{n}{y} p^{y}(1-p)^{n-y}\right]\left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}\right] \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{y+a-1}(1-p)^{n-y+b-1}
\end{aligned}
$$

Derivation of Beta/Binomial Model

The marginal density of Y.

$$
\begin{aligned}
f(y) &= \int_0^1 f(y,p)dp \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{y+a-1}(1-p)^{n-y+b-1}dp \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)} \\
&\times \int_0^1 \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} p^{y+a-1}(1-p)^{n-y+b-1}dp \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}.
\end{aligned}
$$

Derivation of Beta/Binomial Model

- Then the posterior $\pi(p \mid y) = f(y, p)/f(y)$ is

$$\pi(p \mid y) = \frac{\Gamma(n + a + b)}{\Gamma(y + a)\Gamma(n - y + b)} p^{y+a-1}(1 - p)^{n-y+b-1}$$

  for some $0 \leq p \leq 1$.

- Clearly, this posterior is $Beta(y + a, n - y + b)$.

Conjugate prior

- In this example, for the binomial model,
  1. we use the prior $Beta(a, b)$ and
  2. obtain the posterior $Beta(y + a, n - y + b)$.
- When prior and posterior belong to the same distribution family, we say that the prior is conjugate family of the likelihood (model).
- Thus, for the binomial model, beta distribution is the conjugate family, or for short, beta is the conjugate prior (for the binomial likelihood).

Conjugate prior (cont'd)

- Pros.
  - Easy to derive the posterior distribution
  - Easy to apply Monte Carlo method
  - Easy to add new data
- Cons.
  - Restricted form of the prior distribution

Inference with Beta/Binomial Model

- Consider letting $\hat{p} = \mathrm{E}(p \mid y)$ be the posterior mean.
- The mean of $Beta(y + a, n - y + b)$ (posterior) is

$$\hat{p} = \frac{y + a}{y + a + n - y + b} = \frac{y + a}{a + b + n},$$

where we can decompose it into

$$\hat{p} = \frac{n}{a + b + n} \underbrace{\left(\frac{y}{n}\right)}_{\text{sample mean}} + \frac{a + b}{a + b + n} \underbrace{\left(\frac{a}{a + b}\right)}_{\text{prior mean}}.$$

Inference with Beta/Binomial Model

$$\hat{p} = \frac{n}{a+b+n} \underbrace{\left(\frac{y}{n}\right)}_{\text{sample mean}} + \frac{a+b}{a+b+n} \underbrace{\left(\frac{a}{a+b}\right)}_{\text{prior mean}}.$$

- The "Bayes estimator" $\hat{p}$ is a weighted average of the usual frequentist estimator (sample mean) and the prior mean.
- As n increases $(n \to \infty)$, the sample data are weighted more heavily and the prior information less heavily. In fact, we have $\hat{p} \approx y/n$ for all large n.
- In general, as the sample size increases, the likelihood dominates the prior.

Inference with Beta/Binomial Model

$$\hat{p} = \frac{n}{a+b+n} \underbrace{\left(\frac{y}{n}\right)}_{\text{sample mean}} + \frac{a+b}{a+b+n} \underbrace{\left(\frac{a}{a+b}\right)}_{\text{prior mean}}.$$

- $\frac{n}{a+b+n}$ and $\frac{a+b}{a+b+n}$ are relative weights of n and $a+b$.
  We call $a+b$ the "prior sample size".
- Determination of (a, b)
  - $a+b$: prior sample size
  - $\frac{a}{a+b}$: prior guess of p

Prediction

- The prediction probability of X_{n+1} based on the data $x_1, \ldots, x_n$:

$$
\begin{aligned}
& P(X_{n+1} = 1 \mid x_1, \ldots, x_n) \\
= & \int_0^1 P(X_{n+1} = 1 \mid \theta, x_1, \ldots, x_n) \pi(\theta \mid x_1, \ldots, x_n) d\theta \\
= & \int_0^1 P(X_{n+1} = 1 \mid \theta) \pi(\theta \mid x_1, \ldots, x_n) d\theta \\
= & \int_0^1 \theta \pi(\theta \mid x_1, \ldots, x_n) d\theta \\
= & \mathrm{E}(\theta \mid x_1, \ldots, x_n) \\
= & \frac{a + \sum_i x_i}{a + b + n}.
\end{aligned}
$$

Prediction

- When we have the observed data $x_1, ..., x_n$, what is the prediction distribution of $Z = X_{n+1} + \cdots + X_{n+m}$?

$$
\begin{aligned}
&P(Z = z \mid x_1, \ldots, x_n) \\
&= \binom{m}{z} \frac{\Gamma(a+b+n)}{\Gamma(a+\sum x_i)\Gamma(b+n-\sum x_i)} \\
&\times \frac{\Gamma(a+\sum x_i+z)\Gamma(b+n-\sum x_i+m-z)}{\Gamma(a+b+n+m)}.
\end{aligned}
$$

- The above prediction distribution is called the Beta-Binomial distribution.

Prediction

- When we have the observed data $x_1, ..., x_n$, what is the prediction distribution of $Z = X_{n+1} + \cdots + X_{n+m}$?

$$
\begin{aligned}
& P(Z = z \mid x_1, \ldots, x_n) \\
& = \binom{m}{z} \frac{\Gamma(a+b+n)}{\Gamma(a+\sum x_i)\Gamma(b+n-\sum x_i)} \\
& \times \frac{\Gamma(a+\sum x_i + z)\Gamma(b+n-\sum x_i + m - z)}{\Gamma(a+b+n+m)}.
\end{aligned}
$$

- The above prediction distribution is called the Beta-Binomial distribution.

Example: Beta-Binomial distribution

- Suppose $X_1, \ldots, X_n \mid \theta \overset{iid}{\sim} Ber(\theta)$ and $x = \sum_{i=1}^n x_i = 15$ with $n = 40$
- Want to predict $Z = X_{41} + \cdots + X_{50}$, where $X_i \mid \theta \overset{iid}{\sim} Ber(\theta)$
- Note that $\hat{\theta} = n^{-1} \sum_{i=1}^n x_i = 0.375$.
- Frequentist:

$$P(Z = z \mid \hat{\theta} = 0.375) = \binom{10}{z} 0.375^z (1 - 0.375)^{10-z}, \quad z = 0, \ldots, 10.$$

- $\mathrm{E}(Z \mid \hat{\theta}) = 3.75$

Example: Beta-Binomial distribution

- Bayesian:

$$
\begin{aligned}
&P(Z = z \mid x_1, \ldots, x_n) \\
&= \binom{10}{z} \frac{\Gamma(1+1+40)}{\Gamma(1+15)\Gamma(1+40-15)} \\
&\times \frac{\Gamma(1+15+z)\Gamma(1+40-15+10-z)}{\Gamma(1+1+40+10)} \\
&= \binom{10}{z} \frac{\Gamma(42)}{\Gamma(16)\Gamma(26)} \frac{\Gamma(16+z)\Gamma(36-z)}{\Gamma(52)}.
\end{aligned}
$$

- Prediction for Z is $\mathrm{E}(Z \mid x_1, \ldots, x_n) = 3.8095$.
- Note that $\mathrm{E}(Z \mid x_1, \ldots, x_n) = 3.8095 > \mathrm{E}(Z \mid \hat{\theta}) = 3.75$.

Example: Beta-Binomial distribution

- Prediction variance (frequentist):

$$\mathrm{Var}(Z \mid \hat{\theta}) = 2.3438$$

- Prediction variance (Bayesian):

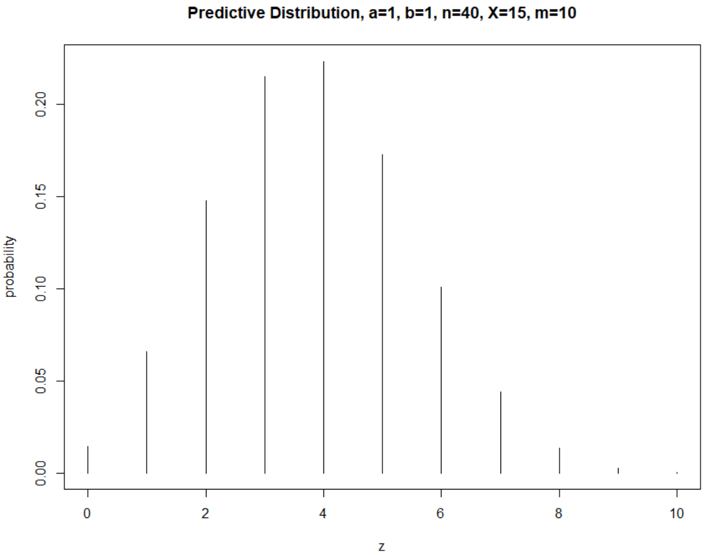
$$\mathrm{Var}(Z \mid x_1, \ldots, x_n) = 2.8558$$

- The Bayesian method provides a larger prediction variance because it considers the variability of θ.
- The frequentist method may suffer from the underestimate problem.

Example: Beta-Binomial distribution (R 실습)

```
> ## beta binomial distribution ####
> a=b=1
> n=40;x=15
> m=10;z=c(0:10)
> pred.z = gamma(m+1)/gamma(z+1)/gamma(m-z+1)*beta(a+z+x,
+          b+n-x+m-z)/beta(a+x, b+n-x)
> plot(z, pred.z, xlab="z", ylab="probability", type="h")
> title("Predictive Distribution, a=1, b=1, n=40, X=15, m=19")
```

Example: Beta-Binomial distribution (R 실습)



Monte Carlo Method Example

The prediction density is the posterior mean of $f(z \mid \theta)$:

$$\begin{aligned} f(Z = z \mid x_1, \dots, x_n) &= \int f(z \mid \theta)\pi(\theta \mid x_1, \dots, x_n)d\theta \\ &\equiv \mathrm{E}^{\pi}(f(z \mid \theta) \mid x_1, \dots, x_n) \end{aligned}$$

Thus, we can approximate $f(Z = z \mid x_1, \dots, x_n)$ using Monte Carlo method.

- First Method: Suppose that we sample $\theta_1, ..., \theta_N \overset{iid}{\sim} \pi(\theta \mid x_1, \dots, x_n)$.

$$\widehat{f}(z \mid X_1, ..., X_n) = \frac{1}{N}\sum_{i=1}^{N} f(z \mid \theta_i) = \frac{1}{N}\sum_{i=1}^{N} \binom{m}{z} \theta_i^z (1 - \theta_i)^{m-z}.$$

Monte Carlo Method Example

- Second Method: Using the following property

$$\begin{aligned} f(z,\theta \mid x_1,\ldots,x_n) &= f(z \mid \theta, x_1,\ldots,x_n)\pi(\theta \mid x_1,\ldots,x_n) \\ &= f(z \mid \theta)\pi(\theta \mid x_1,\ldots,x_n), \end{aligned}$$

  we randomly choose N samples $\{z_i, \theta_i\}_{i=1}^N$ from

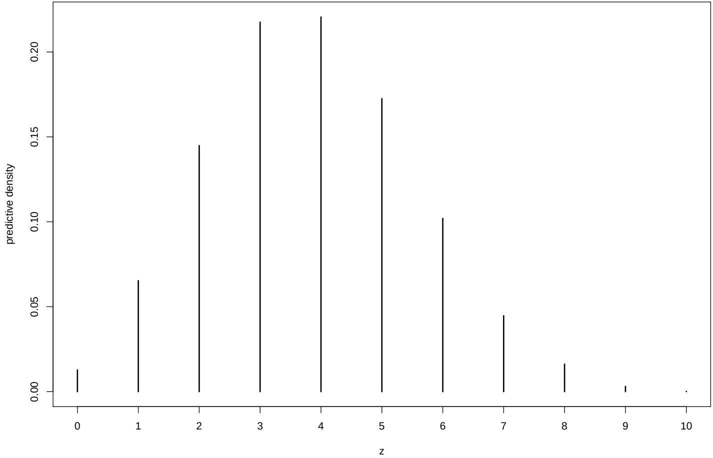
$$\begin{aligned} \theta_i &\sim \pi(\theta \mid x_1,\ldots,x_n) = Beta(a+x, b+n-x) \\ z_i \mid \theta_i &\sim f(z \mid \theta) = Bin(10, \theta_i). \end{aligned}$$

- If we choose only $\{z_i\}_{i=1}^N$, then it is a random sample from $f(z \mid X_1, ..., X_n)$. (Why?)

Monte Carlo Method Example (R 실습)

```
### Monte Carlo Method ####
a=b=1; X=15; n=40; m=10; N=10000
theta = rbeta(N,a+x,b+n-x)
pred.z=c(1:(m+1))*0
for(z in c(0:m)) pred.z[z+1]=mean(dbinom(z,m, theta))
zsample=rbinom(N, m, theta)
plot(table(zsample)/N, type="h", xlab="z", ylab="predictive density",
  main="")
mean(zsample)
[1] 3.8373
var(zsample)
[1] 2.891118
```

Monte Carlo Method Example (R 실습)



Bayesian Credible Interval

- Consider

$$\begin{aligned} X \mid \theta &\sim B(n, \theta) \\ \theta &\sim Unif(0,1) = Beta(1,1). \end{aligned}$$

- Assume that $n = 10$ and $x = 2$.
- Then we have

$$\theta \mid x \sim Beta(x+1, n-x+1) = Beta(3,9).$$

Bayesian Credible Interval

Let's calculate the Bayesian C.I using Grid Search Method.

```
a=1; b=1
x=2; n=10
theta = seq(0,1,length = 1001)
ftheta=dbeta(theta, a+x, n-x+b)
prob=ftheta/sum(ftheta)
HPD = HPDgrid(prob, 0.95)
HPD.grid=c( min(theta[HPD$index]), max(theta[HPD$index]))
HPD.grid
[1] 0.041 0.484
```

Frequentist Confidence Interval

Let's calculate the frequentist C.I.

```
install.packages("binom")
library(binom)
n=10; X=2
CI.exact=binom.confint(X, n, conf.level = 0.95, methods = c("exact"))
CI.exact=c(CI.exact$lower, CI.exact$upper)
CI.exact
[1] 0.02521073 0.55609546
```

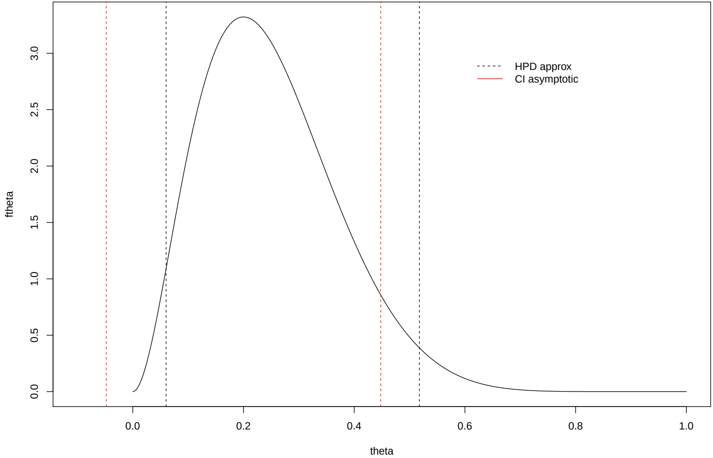
Bayesian C.I vs Frequentist C.I

- The Bayesian C.I is shorter than the frequentist C.I.
- Recall that $Unif(0,1) = Beta(1,1)$ prior has an effect corresponding to the prior sample size $1 + 1 = 2$.
- Thus, we can think that we have more data in this case.

Bayesian C.I vs Frequentist C.I

```
> HPD.approx=qbeta(c(0.025, 0.975),a+x, n-x+b)
> p=x/n
> CI.asympt=c(p-1.96*sqrt(p*(1-p)/n), p+1.96*sqrt(p*(1-p)/n))
> HPD.approx
[1] 0.06021773 0.51775585
> CI.asympt
[1] -0.04792257  0.44792257
```

Bayesian C.I vs Frequentist C.I



Comparison of two proportions

- Consider the model

$$
\begin{aligned}
Y_1, \ldots, Y_{n_1} \mid \theta_1 &\overset{iid}{\sim} Ber(\theta_1), \\
Z_1, \ldots, Z_{n_2} \mid \theta_2 &\overset{iid}{\sim} Ber(\theta_2),
\end{aligned}
$$

  where Y_i's and Z_i's are independent. Let $X_1 = \sum_{i=1}^{n_1} Y_i$ and $X_2 = \sum_{j=1}^{n_2} Z_j$.

- If we assume $\theta_i \sim Beta(a_i, b_i)$ for $i = 1, 2$, we have

$$
\begin{aligned}
\theta_1 \mid x_1 &\sim Beta(a_1 + x_1, b_1 + n_1 - x_1), \\
\theta_2 \mid x_2 &\sim Beta(a_2 + x_2, b_2 + n_2 - x_2).
\end{aligned}
$$

Comparison of two proportions

- We are interested in the comparison of two proportions (or success probabilities) θ_1 and θ_2.
- In this case, $\theta_1 - \theta_2$ may not be appropriate.
  (e.g.) $(\theta_1, \theta_2) = (0.001, 0.0001)$ and $(\theta_1, \theta_2) = (0.8, 0.809)$ satisfy $\theta_1 - \theta_2 = 0.009$, but they are quite different.
- Instead, it would be better to use the log odds ratio:

$$\xi = \log\left(\frac{\theta_1/(1-\theta_1)}{\theta_2/(1-\theta_2)}\right).$$

Comparison of two proportions: Example

- 어느 대학에서 통계학1을 수강하는 학생 18명과 통계학2를 수강하는 학생 10명을 랜덤 추출하여, 수업 수강이 통계학에 흥미를 가지는 데 도움이 되었는지 여부를 조사하였다.
- 통계학1의 18명 중 12명이, 통계학2의 10명 중 8명이 도움이 된다고 답하였다.
- 두 수업 수강생들 간에 비율에 차이가 있을까?

Comparison of two proportions: Example

- θ_1: 통계학1 수강생 중, 수업이 도움이 된다고 생각하는 학생의 비율
- θ_2: 통계학2 수강생 중, 수업이 도움이 된다고 생각하는 학생의 비율
- 다음과 같이 베이지안 모형을 세울 수 있다:

$$
\begin{aligned}
X_1 \mid \theta_1 &\sim B(18, \theta_1), \\
X_2 \mid \theta_2 &\sim B(10, \theta_2), \\
\theta_1, \theta_2 &\overset{iid}{\sim} Beta(a, b).
\end{aligned}
$$

- 그러면 다음의 사후분포를 얻는다:

$$
\begin{aligned}
\theta_1 \mid x_1 &\sim Beta(a + x_1, b + n_1 - x_1), \\
\theta_2 \mid x_2 &\sim Beta(a + x_2, b + n_2 - x_2)
\end{aligned}
$$

Comparison of two proportions: Example

```
a=b=1
n1=18; x1=12; n2=10; x2=8
theta1 = rbeta(10000, a+x1, b+n1-x1)
theta2 = rbeta(10000, a+x2, b+n2-x2)
xi = log( (theta1/(1-theta1)) /(theta2/(1-theta2)) )

HPD=HPDsample(xi)
plot(density(xi), type="l", xlab="log odds ratio",
  ylab="posterior density", main="")
abline(v=HPD, lty=2)
text(mean(xi), 0.06, "95% HPD interval")
```

Comparison of two proportions: Example