베이지안통계학 과제3 solution

Spring 2022

문제 4-1. (10점)

- θ의 우도함수 (5점)
- θ의 최우추정치 (5점)

$X_1, \ldots, X_7 \sim Ber(\theta)$이므로 $f_{X_i}(x) = \theta^x(1-\theta)^{1-x}, i = 1, 2, \ldots, 7$이다. 관측치가 $(x_1, \ldots, x_7) = (1, 1, 0, 1, 0, 0, 0)$이므로 결합 확률분포는 $f(x_1, \ldots, x_7; \theta) = \theta^3(1-\theta)^4$이다. 우도함수와 로그 우도함수로부터 다음과 같이 최우추정치를 구할 수 있다.

$$\begin{aligned} L(\theta; x_1, \ldots, x_7) &= f(x_1, \ldots, x_7; \theta) = \theta^3(1-\theta)^4 \\ \log L(\theta; x_1, \ldots, x_7) &= 3\log\theta + 4\log(1-\theta) \\ \frac{d\log L}{d\theta} &= \frac{3}{\theta} - \frac{4}{1-\theta} \end{aligned}$$

$\frac{d\log L}{d\theta} = 0$을 만족하게 하는 θ는

$$\begin{aligned} \frac{3}{\theta} - \frac{4}{1-\theta} &= 0 \\ 7\theta &= 3 \\ \therefore \hat{\theta} &= \frac{3}{7} \end{aligned}$$

이다.

문제 4-2. (10점)

- $f(x; \theta) = g(T(x), \theta)h(x)$ 꼴로 $T(x) = \sum_{i=1}^n X_i^2$가 충분통계량임을 보이면 10점

$X_1, \ldots, X_n \sim N(0, \theta)$이므로 결합확률밀도함수는

$$f(X_1, \ldots, X_n; \theta) = (2n\theta)^{-n/2} \exp\left\{ -\frac{1}{2\theta}\sum_{i=1}^n X_i^2 \right\}$$

이다. $T(x) = \sum_{i=1}^n X_i^2$일 때 $h(x) = (2n\theta)^{-n/2}, g(T(x); \theta) = \exp\left\{ -\frac{1}{2\theta}\sum_{i=1}^n X_i^2 \right\}$가 존재하고, $f(X_1, \ldots, X_n; \theta) = g(T(x); \theta)h(x)$로 표현 가능하므로 $T(x) = \sum_{i=1}^n X_i^2$가 충분통계량임을 알 수 있다.

문제 4-6. (10점)

- θ의 사후분포 (식 or 분포값 or 그래프) (5점)
- θ의 95% 최대 사후구간 $= \{0.1, 0.2, 0.3, 0.4\}$(5점)

$Y = \sum_{i=1}^{20} X_i \sim B(20, \theta)$ 이므로 Y 의 pdf는 $f(y) = \binom{20}{y}\theta^y(1-\theta)^{20-y}$ 이다. 그리고 $i = 1, \dots, 9$ 에 대해 θ 는 각각 $\theta_i = 0.1, \dots, 0.9$ 의 값을 가지며, 사전분포 $\pi(\theta_i) = 1/9$ 을 가정한다. 그러면 θ 의 사후분포 $f(\theta; Y)$ 는

$$f(\theta; Y) = \frac{f(Y;\theta)}{f(Y)} = \frac{f(Y|\theta)\pi(\theta)}{\sum_{i=1}^{9} f(Y|\theta_i)\pi(\theta_i)} = \frac{\binom{20}{5}\theta^5(1-\theta)^{15}\frac{1}{9}}{\sum_{i=1}^{9}\binom{20}{5}\theta_i^5(1-\theta_i)^{15}\frac{1}{9}}$$

이다.

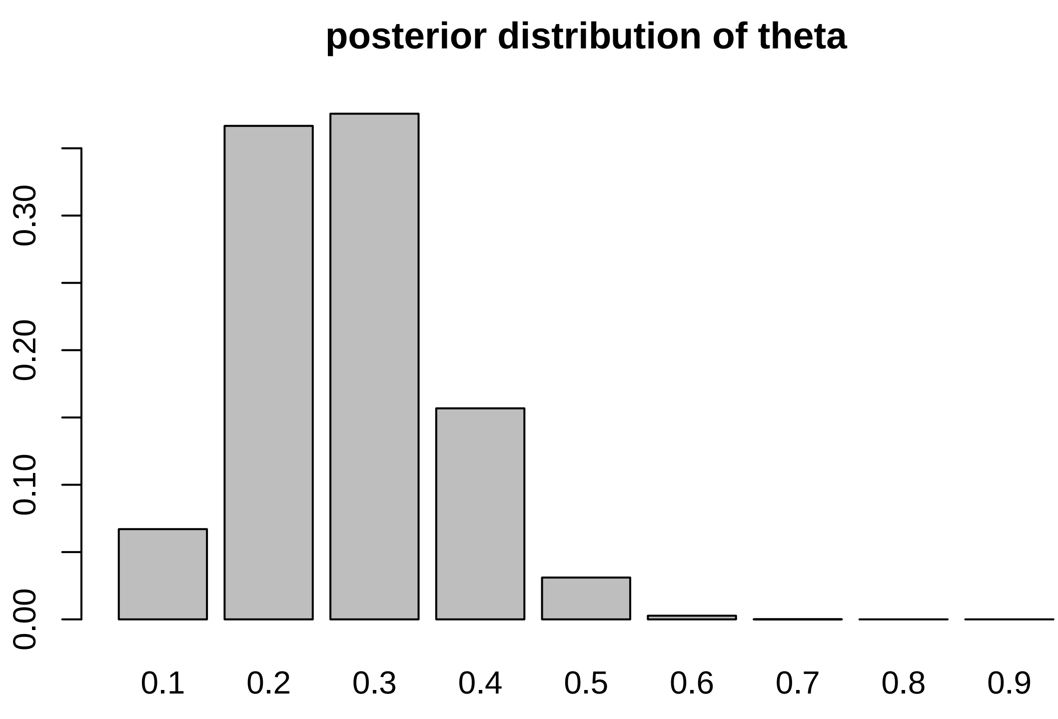
사후분포는 다음의 그래프와 같고, 95% 최대사후구간은 $P(\theta \in C|x) \geq 0.95$ 를 만족하는 집합 C 중에서 집합 C의 사후확률이 최대한 0.95에 가까운 것을 선택하면 된다. 이 문제에서는 $C = \{0.1, 0.2, 0.3, 0.4\}$ 로 설정 가능하다.

```r
theta = seq(0.1, 0.9, length.out = 9)
px = dbinom(x = 5, size = 20, prob = theta) / sum( dbinom(x = 5, size = 20, prob = theta) )

px # pdf
```

```
## [1] 6.704636e-02 3.666379e-01 3.756769e-01 1.567856e-01 3.105544e-02
## [6] 2.718903e-03 7.853199e-05 3.496532e-07 1.922871e-11
```

```r
names(px) = theta # 그래프
barplot(px, main = "posterior distribution of theta")
```



```r
# 95% 최대사후구간
sum(px[1:4])
```

```
## [1] 0.9661468
```