베이지안통계학 과제6 solution

Jieun Shin

Spring 2022

문제 7-2. (25점)

- 사후분포: $\theta|\bar{X} \sim N(150, 1)$, 95% 최대사후구간: [148.04, 151.96] (5점, 사후구간이 없으면 -2)
- 사후분포: $\theta|\bar{X} \sim N(150, \frac{21}{20})$, 95% 최대사후구간: [147.99, 152.01] (5점, 사후구간이 없으면 -2)
- 그래프로 비교 (5점, 그래프가 없으면 -5)
- 예측분포: $(Z - \theta + \theta)|x_1, \ldots, x_{20} \sim N(150, \frac{21+1}{10})$ (5점, 과정에서 틀린경우 -2)
- $P(150 < Z < 153) = 0.4784$ (5점)

(1) 데이터의 분포가 $\bar{X}|\theta \sim N(\theta, \frac{\sigma^2}{n})$ 이고, 여기서 $n = 20, \sigma^2 = 21$ 이다. 그리고 θ의 사전분포는 $\pi(\theta) \sim N(\mu_0, \sigma_0^2)$ 이고 여기서 $\mu_0 = 150, \sigma_0^2 = 21$ 이다.

θ의 사후분포는

$$\begin{aligned}\pi(\theta|\bar{x}) &\propto f(\bar{x}|\theta)\pi(\theta)\\ &\propto \exp\left(-\frac{n}{2\sigma^2}(\theta-\bar{x})^2 - \frac{1}{2\sigma_0^2}(\theta-\mu_0)^2\right)\\ &\propto \exp\left(-\frac{1}{2\sigma_\pi^2}(\theta-\mu_\pi)^2\right)\end{aligned}$$

에 비례하며 여기서 $\sigma_\pi^2 = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} = \frac{1}{\frac{20}{21}+\frac{1}{21}} = 1$ 이고, $\mu_\pi = \frac{\frac{n}{\sigma^2}\bar{x}+\frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}} = \frac{\frac{20}{21}150+\frac{1}{21}150}{\frac{20}{21}+\frac{1}{21}} = 150$ 이다.

따라서 $\theta|\bar{X} \sim N(150, 1)$ 이고, 95% 최대사후구간은 [148.04, 151.96]이다.

```
qnorm(c(0.025, 0.975), mean = 150, sd = 1)
```

```
## [1] 148.04 151.96
```

(2) θ의 사전분포를 무정보 사전분포 $\pi(\theta) = 1$ 이라 하자.

그러면 θ의 사후분포는

$$\begin{aligned}\pi(\theta|\bar{x}) &\propto f(\bar{x}|\theta)\pi(\theta)\\ &\propto \exp\left(-\frac{n}{2\sigma^2}(\theta-\bar{x})^2\right)\\ &\propto \exp\left(-\frac{1}{2\sigma_\pi^2}(\theta-\mu_\pi)^2\right)\end{aligned}$$

에 비례하며 여기서 $\sigma_\pi^2 = \frac{\sigma^2}{n} = \frac{21}{20}$ 이고, $\mu_\pi = \bar{x} = 150$ 이다.

따라서 무정보 사전분포를 준 경우 사후분포는 $\theta|\bar{X} \sim N(150, \frac{21}{20})$ 이고, 95% 최대사후구간은 [147.99, 152.01] 이다.

```r
qnorm(c(0.025, 0.975), mean = 150, sd = sqrt(21/20))
```

```
## [1] 147.9916 152.0084
```

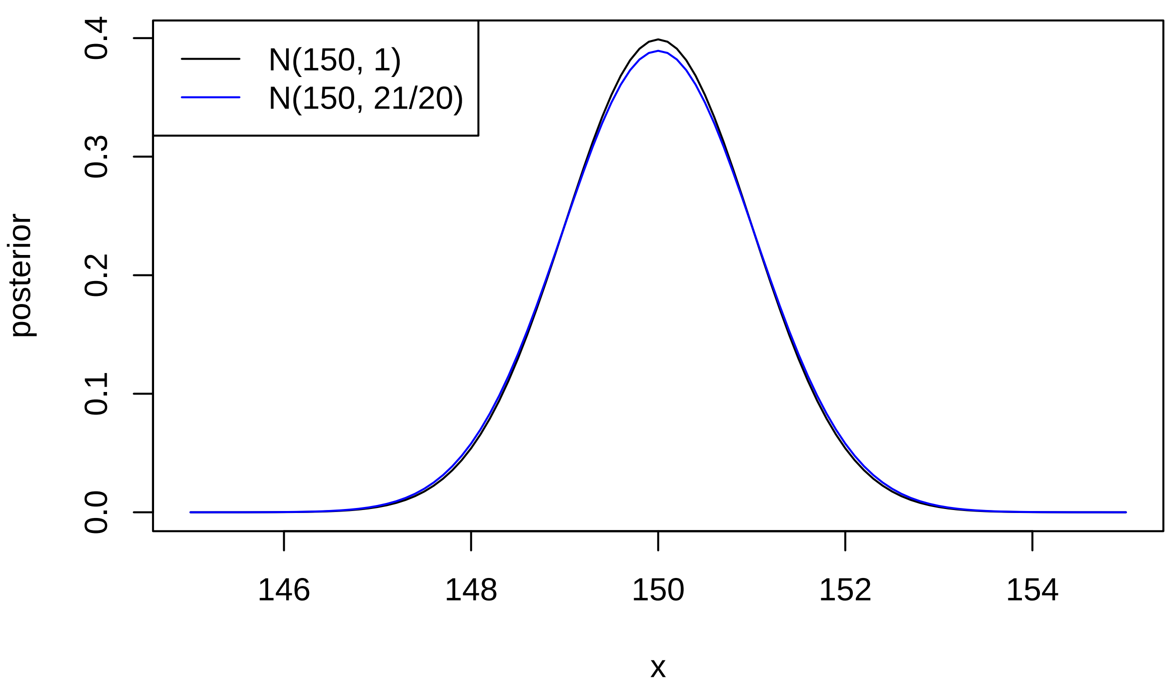
(3) 그래프와 비교

사전분포가 정규분포를 따를 때, 무정보 사전분포보다 분산이 작으므로 사후분포의 분산이 가능도함수의 분산보다 더 작아진다. 그리고 사전분포를 무정보 사전분포로 가정했을 때, 사후분포는 가능도함수와 분포가 같아짐을 볼 수 있다.

```r
# (1)번 분포
curve(dnorm(x, mean = 150, sd = 1), 150 - 5, 150 + 5, ylab = 'posterior')

# (2)번 분포
curve(dnorm(x, mean = 150, sd = sqrt(21/20)), 150 - 5, 150 + 5, ylab = 'posterior', col = 'blue', add =

legend('topleft', lty = c(1, 1), col = c("black", "blue"),
       legend = c("N(150, 1)", "N(150, 21/20)"))
```



(4) 새로운 관측치 10개 $X_{21}, \ldots, X_{30}$를 추가한 경우, $Z = \frac{1}{10}(X_{21} + \cdots + X_{30})$의 예측분포는 $(Z - \theta + \theta)|x_1, \ldots, x_{20} \sim N(150, \frac{21}{10} + 1)$이 된다 ($Z - \theta \sim N(0, \frac{21}{10}$이고, $\theta \sim N(150, 1)$이므로).

(5) $P(150 < Z < 153)$를 R 코드를 통해 구하여 계산하면 0.4558이다.

```r
ci = pnorm(c(150, 153), mean = 150, sd = sqrt(21/10 + 1))
ci[2] - ci[1]
```

```
## [1] 0.4557988
```