베이지안 통계학 과제1

202485-010007 김동현

1. 서로 독립인 관측값들 $(x_1, x_2, \cdots, x_n)$가 평균이 $\frac{1}{\theta}$인 $Exp(\theta), \theta > 0$일 때, 적절한 사전분포를 제시(이유포함)하고, θ의 사후분포를 구하시오.

관측값이 지수분포를 따를 때, 적절한 사전분포는 감마분포이다.

지수분포의 확률밀도함수는 다음과 같다.

$$f(x_i|\theta) = \theta e^{-\theta x_i}, \quad x_i \geq 0$$

지수분포의 가능도함수는 다음과 같다.

$$L(\theta|x_i) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum x_i}$$

θ의 사전분포가 감마분포라 하면, 확률밀도함수는 다음과 같다.

$$f(\theta|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \alpha, \beta > 0$$

베이즈정리를 이용하여 사후분포를 다음과 같이 구할 수 있다.

$$\pi(\theta|x_i) \propto \theta^{\alpha-1} \cdot e^{-\beta\theta} \cdot \theta^n e^{-\theta \sum x_i}$$

$$= \theta^{\alpha+n-1} \cdot e^{-(\beta + \sum x_i)\theta} = Gamma(\alpha + n, \beta + \sum x_i)$$

관측값이 지수분포를 따를 때, 사전분포를 감마분포를 선택하면 사전분포와 사후분포가 같은 켤레사전분포가 된다. 사전분포와 사후분포가 다르다면 사전분포와 사후분포의 성질을 이해하기 위해 2개 분포를 이해해야 한다. 켤레사전분포는 사전분포와 사후분포가 같기 때문에 하나의 분포에 대해서 이해하면 사전분포와 사후분포를 모두 이해할 수 있는 장점이 있다.

2. 지난 10일 동안 7번 비가 왔다. 비가 오는 사건이 과거에 비가 오는 사건들과 서로 독립이라면, 내일과 모레 이틀간 연속적으로 비가 올 확률의 분포를 구하시오. 비가 올 확률의 사전분포는 0과 1사이에서 정의된 균등분포를 사용하시오.

사전분포는 0과 1사이에서 정의된 균등분포이다.

$$\pi(\theta) = 1, \quad 0 \leq \theta \leq 1$$

비가 오는 사건이 독립적이라면 이항분포라 할 수 있다.

이항분포의 확률질량함수는 다음과 같다.

$$P(X = k|\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

지난 10일 동안 7번 비가 온 경우 이항분포의 가능도 함수는 다음과 같다.

$$f(x|\theta) = \binom{10}{7}\theta^7 \cdot (1-\theta)^3$$

베이즈정리를 이용하면 $\pi(\theta|x) \propto \pi(\theta) \cdot f(x|\theta)$ 이므로

$\pi(\theta|x) \propto \theta^7(1-\theta)^3$ 와 같이 단순화 할 수 있고, 이는 θ에 대한 Beta(8,4) 분포이다.

내일과 모레 이틀간 연속적으로 비가 올 확률은 θ^2이며,
베타분포의 성질에 따라

$$E(\theta^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} = \frac{8 \cdot 9}{12 \cdot 13} = \frac{6}{13} = 0.4615$$ 이다.

따라서, 내일과 모레 연속으로 비가 올 확률은 46.15%이다.

3. 연속인 관측치들을 얻어졌다고 하자. 관측치들이 N(0,1)를 따르는지 Ca(0,1)을 따르는지 다음의 가설 $H_0 : x_1, \cdots, x_n \sim N(0,1) \;\; vs \;\; H_1 : x_1, \cdots, x_n \sim Ca(0,1)$ 을 검정하고자 할 때, 각 가설의 사후확률을 구하시오. 이 때, 각 가설의 사전 확률이 1/2로 가정하시오. 참고로, Ca(0,1)는 코시 분포(Cauchy distribution)로 다음과 같은 pdf를 갖는다.

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

1. 사전확률

$$\pi_0 = \pi(\theta_0) = \frac{1}{2}, \; \pi_1 = \pi(\theta_1) = \frac{1}{2}$$

2. 각 가설의 확률밀도함수, 가능도 함수

(1) 가설 H_0 가 성립할 경우(관측치들이 표준정규분포를 따르는 경우)

확률밀도함수는 다음과 같다.

$$f(x|\theta_0) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

가설 H_0 하에서 n개의 관측치가 있을 때 가능도함수는 다음과 같다.

$$L(\theta_0|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}e^{-\frac{x_i^2}{2}}$$

(2) 가설 H_1 가 성립할 경우(관측치들이 코시분포를 따르는 경우)

확률밀도함수는 다음과 같다.

$$f(x|\theta_1) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

가설 H_1 하에서 n개의 관측치가 있을 때 가능도함수는 다음과 같다.

$$L(\theta_1|X) = \prod_{i=1}^{n} \frac{1}{\pi(1+x^2)}$$

3. 사후확률

(1) 가설 H_0 성립할 경우의 사후확률은 다음과 같다.

$$\begin{aligned}\alpha_0 &= \frac{\pi_0 \cdot Ł(\theta_0|X)}{\pi_0 \cdot Ł(\theta_0|X) + \pi_1 \cdot Ł(\theta_1|X)} \\ &= \frac{1/2 \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}}{1/2 \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} + 1/2 \cdot \prod_{i=1}^{n} \frac{1}{\pi(1+x_i^2)}}\end{aligned}$$

(2) 가설 H_1 성립할 경우의 사후확률은 다음과 같다.

$$\begin{aligned}\alpha_1 &= 1 - \alpha_0 \\ &= \frac{1/2 \cdot \prod_{i=1}^{n} \frac{1}{\pi(1+x_i^2)}}{1/2 \cdot \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} + 1/2 \cdot \prod_{i=1}^{n} \frac{1}{\pi(1+x_i^2)}}\end{aligned}$$