

베이지데이터분석 / 이재용 교수

05강

사전분포



Figures don't lie, but liars figure.

숫자는 거짓말을 하지 않는다. 하지만 거짓말쟁이는 공리를 한다.

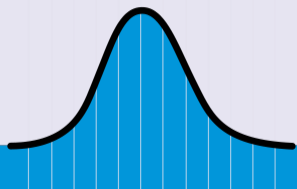
- Carroll D. Wright



목차

- 주관적 확률과 밀도함수를 이끌어 내는 방법들

- 무정보 사전분포

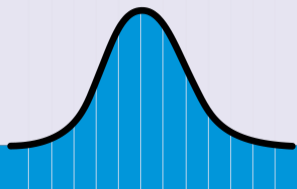




목차

- 주관적 확률과 밀도함수를 이끌어 내는 방법들
-

- 무정보 사전분포



확률의 주관적 결정 방법들

- ▶ (상대적 가능도(relative likelihoods)를 이용하는 방법)
사건 E 의 확률 $\mathbb{P}(E)$ 를 주관적으로 결정할 때,
 E 와 E^c 의 상대적 가능도를 고려한다.
 E 가 E^c 보다 두배 더 가능성이 높다고 생각하면

$$\mathbb{P}(E) = \frac{2}{3}, \mathbb{P}(E^c) = \frac{1}{3}$$

이다.

확률의 주관적 결정 방법들

➤ (내기(betting)를 통해 확률을 결정하는 방법)

- 사건 E 가 일어나면 z 원 만큼 따고,
사건 E^c 가 일어나면 1원 만큼 잃는 내기를 고려한다.
- $U(x)$ 는 금액 x 에 대한 분석자의 효용(utility)이다.
- 내기가 공정하다면 기대 효용은 0이 되어야 한다. 즉,

$$0 = U(z)\mathbb{P}(E) - U(1)(1 - \mathbb{P}(E))$$

이어야 한다. 즉,

$$P(E) = \frac{U(1)}{U(1) + U(z)}$$

이어야 한다.

확률의 주관적 결정 방법들

- $U(z) = cz, \quad c > 0$

$$\mathbb{P}(E) = \frac{1}{1+z}$$

가 된다.

- 실제 문제에 적용할 때는 여러 z 의 값에 대해 내기가 공정하다고 느끼는지 판단해서 적절한 z 값을 결정한다.

예: 대한민국의 월드컵 4강 진출 확률

- 대한민국이 4강에서 떨어지면 만원을 상대에게 주고, 대한민국이 4강에 들어가면 z 원을 받는 내기에서 공정한 z 값을 먼저 정하자. z 값에 따라서 이 내기를 나에게 유리하다고 생각하는지 안하는지 생각해 보자.
- 내기가 유리하다고 생각하면 o, 유리하지 않다고 생각하면 x로 표기를 했다.

z 값	5,000원	10,000원	20,000원	50,000원	100,000원
내기가 유리하다.	x	x	x	x	x
z 값	200,000원	500,000원	1,000,000원	2,000,000원	5,000,000원
내기가 유리하다.	x	o	o	o	o

예: 대한민국의 월드컵 4강 진출 확률

- 공정한 z 값을 40만원이라고 생각한다고 찾아냈다.
이제 선형 효용 함수를 가정하고 A가 생각하는
다음 월드컵에서 4강에 합류할 확률을 계산하면

$$\mathbb{P}(E) = \frac{1}{1 + 40} = 0.02439$$

를 얻는다.

주관적 확률 결정의 어려움

- 주관적 확률의 결정은 숙달된 데이터 분석가에게도 어려운 일이다.
- 효용 함수를 결정하기 어렵다.
일반적으로 작은 액수에 대한 효용 함수는 선형으로 움직이지만
액수가 커지면 효용의 상대적 크기가 점점 작아진다고 알려져 있다.
- 매우 작은 주관적 확률을 결정하는 것은 어려운 일이다.
예를 들어 내일 A가 교통사고를 당할 확률을 고려해보자.
이 확률은 0.0001일까, 0.00001일까?
이 두 값 중 어떤 값에 더 가까운 값일까?

주관적 확률 결정의 어려움

- 여러 개의 사건들의 주관적 확률을 결정할 때 확률들 간에 일관성(coherency)을 유지하는 것 역시 어려운 일로 알려져 있다.

두 사건 E_1 과 E_2 의 확률을 각각 $\mathbb{P}(E_1) = \frac{1}{3}$ 과 $\mathbb{P}(E_2) = \frac{1}{4}$ 이라 말한 후에

$E_1 \cup E_2$ 의 확률은 $\mathbb{P}(E_1 \cup E_2) = \frac{3}{4}$ 라 얘기하는 경우도 있을 수 있다. 하지만

$$\mathbb{P}(E_1 \cup E_2) \leq \frac{1}{3} + \frac{1}{4} < \frac{3}{4}$$

으로 가능하지 않다.

주관적 확률 밀도를 이끌어 내는 방법들

▶ 히스토그램 방법

: Θ 를 구간으로 나누고, 각 구간의 확률을 주관적으로 구한다.

▶ 상대적 가능도 방법

: Θ 의 여러 점들의 상대적 가능도를 결정하고 이 값들을 연결한다.

▶ 주어진 함수 형태에서 찾는 방법

- 적률 이용 방법
- 분위수 이용 방법
- 사전표본크기 이용 방법

주관적 확률 밀도를 이끌어 내는 방법들

▶ 누적분포함수를 결정하는 방법

: 몇 개의 $\alpha \in (0, 1)$ 에 대해 분위수를 주관적으로 이끌어내고 이를 부드럽게 연결한다.

▶ 다변량 사전 분포 이끌어 내기

: $\pi(\theta_1)$ 을 먼저 주관적으로 결정하고,
주어진 θ_1 에 대해 $\pi(\theta_2)$ 를 주관적으로 이끌어낸다.

예. 적률이용 방법. 베타사전분포 모수의 결정.

- ▶ θ 의 사전 분포의 모임이 $Beta(\alpha, \beta)$, $(\alpha, \beta > 0)$ 이라 하자. α, β 를 결정하자.
- ▶ θ 의 사전 평균과 사전 분산이 얼마가 될지 결정한다.
- ▶ 이 값들이 각각 m 과 v 라면

$$m = \frac{\alpha}{\alpha + \beta}$$
$$v = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

$$\alpha = m \left[\frac{1}{v} m(1 - m) - 1 \right]$$

$$\beta = (1 - m) \left[\frac{1}{v} m(1 - m) - 1 \right]$$

예. 적률이용 방법. 베타사전분포 모수의 결정.

- ▶ 적률은 분포의 꼬리의 두께에 크게 영향을 받고, 분석자는 대개의 경우 분포 꼬리에 대한 매우 막연한 정보를 갖고 있는 경우가 많다.
이와 같은 이유로 사전분포의 적률을
분석자로부터 이끌어내는 것은 쉽지 않다.

예. 분위수 이용 방법. 정규사전분포 모수의 결정.

▶ 사전 분포의 모임이 $N(\mu, \sigma^2)$ 이라 하자. μ, σ 를 결정하자.

▶ θ 의 50%와 75% 분위수 $x_{0.5}$ 와 $x_{0.75}$ 를 결정한다.

▶ $\mathbb{P} \mu, \sigma^2(\theta \leq x_{0.5}) = 0.5$

$$\mathbb{P} \mu, \sigma^2(\theta \leq x_{0.75}) = 0.75$$

를 만족하는 μ 와 σ^2 을 구하면 된다.

위의 식은

$$\frac{x_{0.5} - \mu}{\sigma} = 0$$

$$\frac{x_{0.75} - \mu}{\sigma} = 0.674$$

와 동치이고, 이를 풀면

$$\mu = x_{0.5}$$

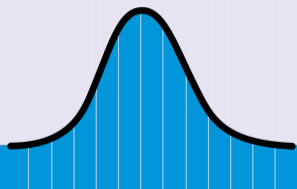
$$\sigma = \frac{x_{0.75} - x_{0.5}}{0.674}$$



목차

➤ 주관적 확률과 밀도함수를
이끌어 내는 방법들

➤ 무정보 사전분포



- ▶ $\Theta = \mathbb{R}$ 일 때, 상대적 가능도 방법을 이용하여 θ 에 대한 완전한 무지를 표현하면

$$\pi(\theta) = c, \forall \theta \in \mathbb{R}, c > 0$$

을 얻는다.

- ▶ (균등사전분포의 문제점)

- (임프로퍼 사전 분포(improper prior))

$$\int_{\mathbb{R}} \pi(\theta) d\theta = \infty$$

- 변환 불변성을 만족하지 않는다.

▶ (위치 모수의 변환 불변 사전분포)

- f 가 \mathbb{R} 에서 확률 밀도 함수.

$$x \sim f(x - \theta), \theta \in \mathbb{R} \quad (1)$$

을 위치 모형(location model)이라 하고, θ 를 위치 모수라 한다.

- θ 의 변환 불변 사전분포는

$$\pi(\theta) = C, \forall \theta$$

▶ (척도 모수의 변환 불변 사전분포)

- f 가 $\mathbb{R}_+ = (0, \infty)$ 에서 확률 밀도 함수.

$$x = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right), x > 0 \quad (2)$$

를 척도 모형(scale model)이라 하고, $\sigma > 0$ 를 척도 모수라 한다.

- σ 의 변환 불변 사전분포는

$$\pi(\sigma) = \frac{C}{\sigma}, \forall \sigma.$$

- (척도 변환 불변 사전분포의 여러 형태)

$$\frac{1}{\sigma} d\sigma = \frac{1}{2\sigma^2} d\sigma^2 = d \log \sigma$$

예. 척도 정규모형

➤ (모형)

$$x_1, \dots, x_n | \sigma \stackrel{i.i.d}{\sim} N(\theta_0, \sigma^2), \theta_0 \in \mathbb{R}, \sigma > 0$$

➤ (사전분포)

$$\pi(\sigma) d\sigma = \frac{1}{\sigma} d\sigma, \sigma > 0$$

예. 척도 정규모형

▶ (사후분포)

$$\begin{aligned}\pi(\sigma|x)d\sigma &\propto \frac{1}{\sigma} d\sigma \prod_{i=1}^n \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(x_i-\theta_0)^2} \\ &= \frac{1}{\sigma^{n+1}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-\theta_0)^2} d\sigma.\end{aligned}$$

$\lambda = \frac{1}{\sigma^2}$ 로 변환하면

$$\begin{aligned}\frac{1}{\lambda} d\lambda &= \frac{1}{\sigma} d\sigma. \\ \pi(\sigma|x)d\sigma &\propto \lambda^{\frac{n}{2}-1} e^{-\frac{1}{2}\lambda \sum_{i=1}^n (x_i-\theta_0)^2} d\lambda \\ &\propto Ga(\lambda | \frac{n}{2}, \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2)\end{aligned}$$

➤ (모형)

$$x \sim f(x|\theta), \theta \in \Theta \subset \mathbb{R}^k$$

➤ (제프리스 사전분포)

$$\pi^J(\theta) \propto |\det I(\theta)|^{1/2}, \theta \in \Theta$$

➤ (변환불변성) $\eta = \eta(\theta)$ 일 때, θ 의 제프리스 사전분포를 구한 후 이를 η 로 변수 변환하면, η 의 제프리스 사전분포와 동일하다.

➤ (제프리스 사전분포의 문제점)

- 제프리스 사전분포는 가능도 원칙을 따르지 않는다.
- 제프리스 사전분포는 최적이지 아닌 추론을 줄 수 있다.

제프리스 사전분포의 예

▶ (이항모형)

$$x \sim \text{Bin}(n, p), p \in (0, 1), n \in \mathbb{N}$$

$$I(p) = np^{-1}(1-p)^{-1}$$

$$\pi^J(p) \propto p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} = \text{Beta}(p | \frac{1}{2}, \frac{1}{2})$$

▶ (정규모형)

$$x \sim N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0$$

$$I(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

$$\pi^J(\mu, \sigma^2) d\mu d\sigma^2 = \frac{1}{\sigma^3} d\mu d\sigma^2$$

- ▶ (모형) $x \sim f(x|\theta)$
- ▶ 모수 θ 를 관심 모수 θ_1 과 장애 모수 θ_2 로 $\theta = (\theta_1, \theta_2)$ 로 나눈다.
- ▶ (기준 사전분포)

$$\pi^R(\theta_1, \theta_2) = \pi^R(\theta_1)\pi^R(\theta_2|\theta_1)$$

- $\pi^R(\theta_2|\theta_1)$ 은 고정된 θ_1 에 대해 모형 $f(x|\theta_1, \theta_2)$ 에 대한 θ_2 의 제프리스 사전 분포

- $\pi^R(\theta_1)$ 는 모형

$$f(x|\theta_1) = \int f(x|\theta_1, \theta_2)\pi^R(\theta_2|\theta_1)d\theta_2$$

의 제프리스 사전 분포

다른 무정보 사전분포

포함 확률 일치 사전분포

: 사후 분포를 이용해 구한 θ 의 $100(1 - \alpha)\%$, $\alpha \in [0, 1]$ 신용 집합이 동일한 포함 확률의 빈도론적 신뢰 구간이 되도록 구한 θ 의 사전 분포

최대 엔트로피 사전분포: Θ 상의 엔트로피

$$H(\pi) = - \sum_{\theta \in \Theta} \pi(\theta) \log \pi(\theta), \quad \Theta : \text{이산집합}$$

를 최대로 하는 사전분포

제 2종 최대 가능도 사전분포

▶ 사전분포의 모임

$$\Gamma = \{\pi(\theta|\lambda) : \lambda \in \Lambda\}$$

중 하나의 사전분포를 고르려고 할 때, x 의 주변 분포를 최대화하는 λ 를 선택하는 것을 제 2종 최대 가능도 방법(type II maximum likelihood method)라 한다.

이 때 선택된 사전 분포 $\pi_{\hat{\lambda}}$ 를 제 2종 최대 가능도 사전분포라 한다.

여기서

$$\hat{\lambda} := \operatorname{argmax}_{\lambda} m(x|\lambda) = \operatorname{argmax}_{\lambda} \int f(x|\theta)\pi(\theta|\lambda)d\theta$$

이다.

- ▶ 제 2종 최대 가능도 사전 분포를 쓰는 베이지 방법은 경험적 베이지 방법(empirical Bayes inference)의 일종이다.

다음시간

06 강 _____

몬테카를로방법

