베이지안 통계의 기초

이경재

인하대학교 통계학과

March 3, 2019

베이지안의 기초

베이지안의 개념

- 관측되지 않은 모든 모수를 확률변수로 생각
- 모수의 분포에 대한 믿음대로 사전분포를 부여
- 자료의 가능도와 모수의 사전분포를 이용해, 사후분포로부터 미지의 모수를 추론
- 베이지안은 기존의 통계방법(이후로 빈도론이라 부름)과 대응되는 하나의 패러다임

가능도(모형) (Likelihood)

$$f(x \mid \theta) \geq 0 \qquad \text{where} \int f(x \mid \theta)dx = 1 \tag{1}$$

- 자료의 분포를 말하며, 가정하고 있는 모형을 대표
- x: 관측된 자료
- θ: 미지의 모수. 추론하고 싶은 대상
- (1)을 x의 함수로 보면 밀도함수, θ의 함수로 보면 가능도라 함

가능도의 예시

- (이항분포)
  $X \sim B(n,p) \quad \Leftrightarrow \quad f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \ldots, n$
- (포아송분포) $X \sim Poi(\theta) \quad \Leftrightarrow \quad f(x \mid \theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, \ldots.$
- (정규분포)
  $X \sim N(\mu, \sigma^2) \quad \Leftrightarrow \quad f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}$

사전분포(Prior)

$$\pi(\theta) \geq 0 \qquad \text{where} \int \pi(\theta) d\theta = 1 \tag{2}$$

- 미지의 모수가 따를 것이라 기대되는 확률분포
- 알려진 사실이나 개인적인 믿음을 반영

사전분포의 예시

- (베타분포)
  $\theta \sim Beta(a,b) \quad \Leftrightarrow \quad \pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, \quad \theta \in (0,1)$
- (감마분포)
  $\theta \sim Gamma(a,b) \quad \Leftrightarrow \quad \pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}, \quad \theta \in (0,\infty)$
- (균일분포)
  $\theta \sim Unif(a,b) \quad \Leftrightarrow \quad \pi(\theta) \propto 1$ with $\int \pi(\theta)d\theta = 1, \quad \theta \in [a,b]$

사후분포(Posterior)

$$\pi(\theta \mid x) \quad \propto \quad \pi(\theta) f(x \mid \theta)$$

$$\pi(\theta \mid x) \geq 0 \qquad \text{where} \int \pi(\theta \mid x) d\theta = 1 \tag{3}$$

- 관측된 자료 x에 대한 모수 θ의 조건부 분포
- 자료의 가능도와 모수의 사전분포로 이루어짐
- (일반적으로) 자료가 더 많이 관측될 수록 추론의 성능이 향상됨

베이지안을 이용한 통계분석

추론(Inference)

- 베이지안 추론은 $\pi(\theta \mid x)$로부터 θ에 대한 추론을 하는 것이다.
- 추론이라 함은 사후분포 (3)의 분포를 이용하여 샘플을 뽑거나 평균 등의 대표값을 이용해서 추정값을 구하는 것이다.

추론(Inference)

- 사후분포를 구하는 데에는 다음의 베이즈 정리가 필요하다:

$$P(A,B) = P(A)P(B \mid A).$$

- 베이즈 정리를 이용하여 아래와 같이 사후분포를 구할 수 있다:

$$\begin{aligned}\pi(\theta \mid x) &= \frac{f(\theta, x)}{f(x)}, \quad f(x) \equiv \int \pi(\theta) f(x \mid \theta) d\theta \\ &= \frac{\pi(\theta) f(x \mid \theta)}{f(x)}.\end{aligned}$$

- 위의 식에서 θ에 대한 부분만 뽑아오면

$$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta). \tag{4}$$

예측(Prediction)

$$p(x_{new} \mid x)$$

- 관측된 자료(x)를 이용하여 미래에 관측될 자료(x_{new})를 예측
- 베이지안 방식의 예측은 아래와 같은 식을 이용

$$p(x_{new} \mid x) \quad = \quad \int f(x_{new} \mid \theta)\pi(\theta \mid x)d\theta$$

가설검정(Hypothesis test)

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1$$

- 주어진 자료를 바탕으로 귀무가설(H_0)과 대립가설(H_1) 중 하나를 채택
- $\pi(H_0 \mid x)$와 $\pi(H_1 \mid x)$로 판단. 베이즈인자(Bayes factor)등을 이용

베이지안 추론의 장점

- 관측된 자료에 의거한 추론이 가능
- 알려진 사실, (필요한 경우) 개인적인 믿음 등을 추론에 반영
- 계층모형 등의 경우 CV (Cross validation)등을 쓰지 않고 자연스러운 추론이 가능
- 신뢰구간 등의 해석이 더 자연스러움

베이지안 추론의 단점

- 사후분포의 계산이 어려운 경우가 많음
- 고정된 갯수의 자료를 이용할 때, 사전분포가 모수의 실제 분포와 많이 다른 경우 추론의 정확도가 떨어짐

베이지안과 빈도론의 비교

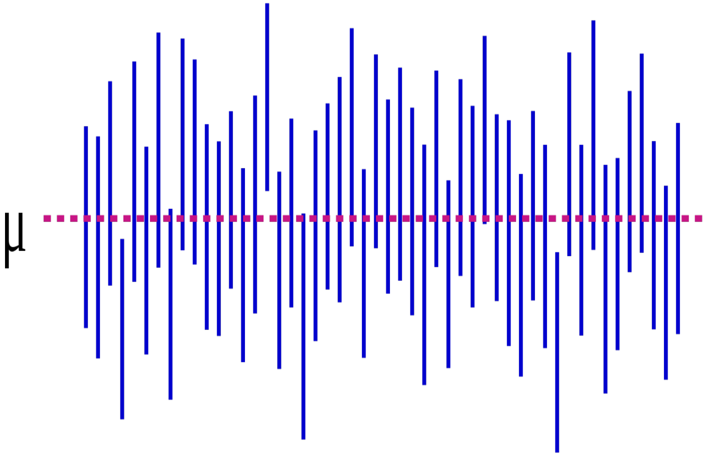
θ 신뢰구간: 빈도론

- 다음 식을 만족하는 구간 $[L(X), U(X)]$를 θ에 대한 $(1-\alpha)100\%$ 신뢰구간(Confidence interval)이라 한다:

$$P_\theta(L(X) \le \theta \le U(X)) \ge 1-\alpha, \quad \forall\theta.$$

- 신뢰구간이 자료의 분포에 의존

신뢰구간의 의미



θ 신용구간: 베이지안

- 다음 식을 만족하는 구간 $[L(X), U(X)]$를 θ에 대한 $(1-\alpha)100\%$ 신용구간(Credible interval)이라 한다:

$$\pi(L(x) \le \theta \le U(x) \mid x) \ge 1-\alpha.$$

- 신뢰구간이 모수의 사후분포에 의존