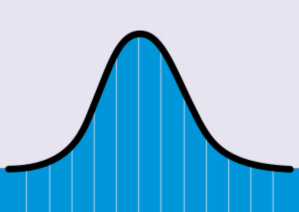
베이즈데이터분석 / 이재용 교수

02 강

베이즈 추론

목차



목차

통계적 추론의 문제

- 모수 (θ) 자연의 법칙을 나타내는 미지의 값
- 관측치 (x) 확률분포 $f(x|\theta)$를 따르는 확률변수
- 목표 x를 기초로 모수 θ에 대한 추론을 하고자 한다.

통계적 추론의 문제

비와 우산의 예

- A는 B와 지하의 사무실을 같이 쓰고 있다. 사무실이 지하에 있기 때문에 출근한 후에 날씨의 변화를 알 수 없다. 오늘 따라 A는 사무실에 일찍 나왔다. 그런데 B가 출근할 때 비에 젖은 우산을 들고 들어왔다고 하자. A는 이를 보고 지금 비가 오고 있을 것이라고 추론할 수 있다.
- 관측 값 x는 θ에 따라 분포가 달라진다. 비가 오면 x는 비에 젖은 우산을 든 B일 가능성이 높고, 비가 오지 않으면 x는 우산을 들지 않은 B일 가능성이 높다.
- x의 분포가 θ에 따라 달라지기 때문에 x는 θ에 대한 정보를 갖고 있고 통계적 추론은 이 정보를 이용하여 θ를 추론하는 것이다.

예 : 압정

문제

↘의 확률을 알고자 한다.

예 : 압정

- 압정 10개를 던져서 7개의 ⟍와 3개의 ⊥를 얻었다.
- ⟍의 확률에 대해 어떤 결론을 내릴 수 있나?

- **모수** θ $=$ ⋋의 확률
- **관측치** x = 압정을 10번 던졌을 때 ⋋가 나온 횟수

- **문제** $x = 7$를 관측했을 때 θ에 관한 추론은?

베이즈 추론

- 베이즈 추론은 다음의 세 가지 요소로 이루어져 있다.

  - **사전분포**: θ의 분포로 자료를 보기 전의 분석자의 θ에 관한 정보 (혹은 불확실성)를 나타낸다.
    $\pi(\theta)$로 나타낸다.

  - **확률모형**: $x|\theta \sim f(x|\theta)$.

  - **사후분포**: 자료가 주어졌을 때 θ의 확률분포로 자료를 본 후의 분석자의 θ에 관한 정보를 나타낸다.
    $\pi(\theta|x)$로 나타낸다.

베이즈 법칙: 사후분포의 계산방법

베이즈 법칙 혹은 베이즈 정리는 사후분포를 계산하는 수학적 방법이다.

$$\pi(\theta|x) \propto \pi(\theta)f(x|\theta)$$

혹은

$$\text{사후분포} \propto \text{사전분포} \times \text{가능도.}$$

압정의 예: 사후분포의 계산

- 사전분포: $\theta \sim U(0,1)$.

  자료 x를 보기 전에, 모든 θ값이 동일한 가능성을 갖고 있다면, 즉
  모든 $\theta_1 \neq \theta_2$에 대하여 $\pi(\theta_1) = \pi(\theta_2)$라면 $\theta \sim U(0,1)$.

- 가능도: $x|\theta \sim Bin(n,\theta)$.

- 사후분포

$$\begin{aligned}\pi(\theta|x) &\propto \pi(\theta) \times f(x|\theta) \\ &= 1 \times \binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &\propto \theta^x(1-\theta)^{n-x}, 0 < \theta < 1.\end{aligned}$$

$$\theta|x \sim Beta(x+1, n-x+1).$$

- 압정의 예에서는,

$$\theta|x = 7 \sim Beta(8,4).$$

참고. 분포의 예들

- **균등분포(uniform distribution)**: 연속형 확률변수 X가 밀도함수

$$f(x) = \frac{1}{b-a} I(a \leq x \leq b)$$

를 따르면, $X \sim U(a,b),\ a < b$라 쓴다.

- **이항분포(binomial distribution)**:

이산형 확률변수 X가 밀도함수

$$P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

를 따르면,$X \sim Bin(n,p)$라 쓴다.

참고. 분포의 예들

- **베타분포(beta distribution)**: 연속형 확률변수 X가 밀도함수

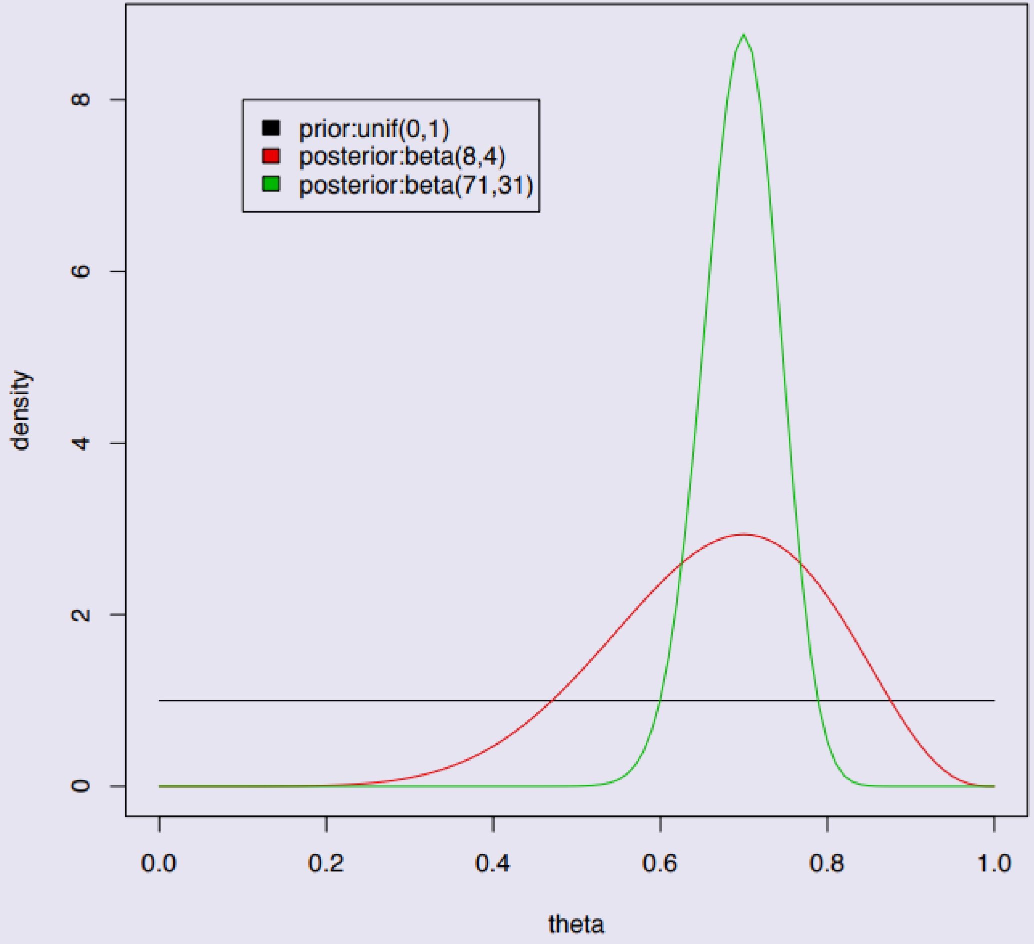
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x, 1$$

를 따르면 $X \sim Beta(\alpha, \beta), \alpha, \beta > 0$라 쓴다.

- **감마함수(gamma function)**:

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx, \alpha > 0.$$

사후확률밀도함수



prior:unif(0,1)
posterior:beta(8,4)
posterior:beta(71,31)
density
theta
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0

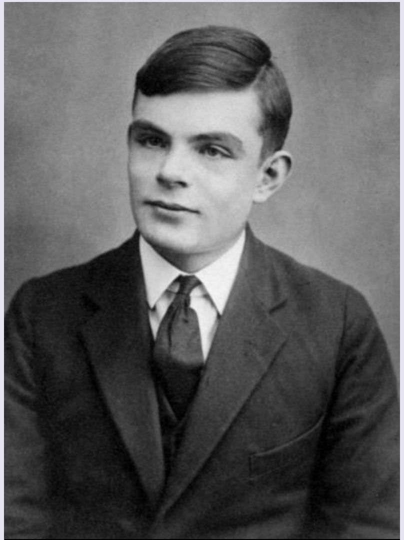
베이즈 통계의 장점

- 새로운 자료가 관측되었을 때, 정보를 업데이트 하는 것이 자연스럽다. 과거의 사후분포가, 현재의 사전분포가 된다.

$$\boxed{\pi(\theta)} \rightarrow X_1 \rightarrow \boxed{\pi(\theta|X_1)} \rightarrow X_2 \rightarrow \boxed{\pi(\theta|X_1, X_2)} \cdots$$

- (전문가 의견의 이용) 과거의 경험으로부터 강한 사전 정보 혹은 의견이 있을 때, 이를 추론에 이용할 수 있다.
- (계층모형) 비슷한 값들을 동시에 추정해야 할 때 정보를 종합해서 추론하는데 장점이 있다.
- 구간 추정과 가설 검정의 결과의 해석이 자연스럽다.
- 추론을 할 때, 대표본이론을 이용한 근사를 이용하지 않아도 된다. 베이즈추론은 유한 표본에서도 정확한 추론분포를 이용한다.
- 빈도론 추론 방법이 베이즈 방법의 일종이 되는 경우가 많다.

베이즈 통계의 적용 예들



Alan Turing이
2차 세계대전에서
독일군의
암호를 풀 때
베이즈 통계를
사용하였다.

미국의 전 연방준비제도
이사회 회장이었던
Alan Greenspan은
금융정책 결정과정에
베이즈 결정론을
이용한다고 말했다.

MS 오피스의
길잡이도
베이즈 통계를
이용하였다.
빌 게이트는 베이즈
주의자라고 알려져 있다.

이항모형

- 사전분포: $\theta \sim Beta(\alpha, \beta), \alpha, \beta > 0$
- 모형: $x|\theta \sim Bin(n, \theta)$
- 사후분포:

$$\theta|x \sim Beta(\alpha + x,\ \beta + n - x), x = 0, 1, \cdots, n.$$

이항모형

- 베이즈 추정량은 사전분포의 평균과 최대가능도 추정량의 가중평균이다.

$$\hat{\theta}^B = \mathbb{E}(\theta|x) = \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n}\frac{x}{n}.$$

$\mathbb{E}(\theta) = \frac{\alpha}{\alpha+\beta}$이고 $\hat{\theta}^{MLE} = \frac{x}{n}$이다.

- 표본의 크기가 커지면서 사전분포는 사라진다. 즉, 자료가 사전분포를 압도한다.

$$\lim_{n\to\infty} \hat{\theta}^B = \hat{\theta}^{MLE}.$$

- 사전분포 파라미터의 결정: $\alpha + \beta$를 사전표본의 크기로 해석한다.

균등사전분포와 불충분근거의 원칙

θ에 대한 사전분포로

$$\theta \sim U(0,1)$$

를 많이 사용한다.

- 라플라스는 균등분포가 완전한 무지를 표현한다고 생각했다.
- 불충분근거의 원칙이란 근거가 없는 경우 두 개의 다른 사건은 동일한 사전 확률을 갖는다는 원칙이다.
- 베이즈는 균등분포를 사전분포로 쓸 때, x의 주변분포가 이산균등분포가 되는 것을 보고 $U(0,1)$을 사전분포로 써야 하는 이유라고 생각했다.
- $\theta \sim U(0,1) = Beta(1,1),\ x|\theta \sim Bin(n,\theta)$이면, $\theta|x \sim Beta(x+1,\ n-x+1)$이다.

정규모형 : 평균을 모르고 분산은 아는 경우

- (모형) σ^2은 아는 값이고, θ에 대해 추론하고자 한다.

$$X_1, \cdots, X_n | \theta \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$$

- (사전분포) $\theta \sim N(\mu, \tau^2)$

- (사후분포)

$$\theta | x \sim N\left(\frac{\frac{1}{\tau^2}\mu + \frac{n \cdot}{\sigma^2}\bar{X}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right).$$

여기서 $x = (x_1, \ldots, x_n)$이다.

정규모형 : 평균을 모르고 분산은 아는 경우

- 베이즈 추정량은 사전분포의 평균과 최대 가능도 추정량의 가중 평균이다.

$$\hat{\theta}^B = \frac{\frac{1}{\tau^2}\mu + \frac{n\cdot}{\sigma^2}\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} = \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\mu + \frac{\frac{n\cdot}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\bar{x}.$$

  여기서 μ는 사전분포의 평균이고 $\bar{x}$는$\hat{\theta}^{MLE}$이다.

- 관측치의 개수가 커지면서, 자료가 사전분포를 압도한다. 즉

$$\lim_{n\to\infty} \hat{\theta}^B = \hat{\theta}^{MLE}.$$

- 사전분포 모수의 결정: $\tau^2 = \sigma^2/k$라 놓을 때, k를 사전표본의 개수라 한다.

- 점근적으로 사후분포와 최대 가능도 추정량의 분포가 동일하다.

$\theta \in \mathbb{R}$에 대한 균등분포

- (균등사전분포일 때의 사후분포)

$$\theta \sim \pi(\theta) = 1, \forall\theta \in \mathbb{R}$$

균등분포 $\pi(\theta)$는 엄밀한 의미에서 확률분포가 아니다.
그럼에도 불구하고, 사후분포를 계산해보면

$$\theta|x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

으로, 자료가 사전분포를 압도한 사후분포의 값으로 합리적이다.

- 균등사전분포일 때의 사후분포는 어떤 사전분포의 사후분포들의 극한이다.

$$\lim_{\tau^2 \to \infty} [\theta|x]_\pi = N(\mu, \tau^2) = [\theta|x]_{\pi(\theta)} = 1.$$

감마분포로 정의되는 분포들

- (감마분포) $X \sim Ga(\alpha, \beta), \alpha, \beta > 0.$
  - 밀도함수: $f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0.$
  - $\mathbb{E}X = \frac{\alpha}{\beta}, \mathbb{V}ar(X) = \frac{\alpha}{\beta^2}.$
- (카이제곱분포) $X \sim \mathcal{X}^2(\nu) = Ga\left(\frac{\nu}{2}, \frac{1}{2}\right).$
  - 밀도함수:

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x}, x > 0 .$$

  - $\mathbb{E}X = \nu, \mathbb{V}ar(X) = 2\nu.$

감마분포로 정의되는 분포들

- (척도 카이제곱분포) $X \sim \mathcal{X}^2(\nu, \lambda) = Ga(\frac{\nu}{2}, \frac{\nu}{2}\lambda)$.
  - 밀도함수:
    $$f(x) = \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{\nu\lambda}{2}x}, x > 0.$$
  - $\mathbb{E}X = \frac{1}{\lambda}$, $\mathbb{V}ar(X) = \frac{2}{\nu\lambda^2}$.

- (역감마분포) $X \sim Inv - Ga(\alpha, \beta) \Longleftrightarrow 1/X \sim Ga(\alpha, \beta), \alpha, \beta > 0.$
  - 밀도함수: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\beta/x}, x > 0$
  - $\mathbb{E}X = \frac{\beta}{\alpha-1}$, $\alpha > 1$

    $\mathbb{V}ar(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$.

감마분포로 정의되는 분포들

- (역카이제곱분포) $X \sim Inv\text{-}\chi^2(\nu) = Inv\text{-}Ga\left(\frac{\nu}{2}, \frac{1}{2}\right)$.
  - 밀도함수: $f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{x}\right)^{\frac{\nu}{2}+1} e^{-\frac{1}{2x}}, x > 0$.
  - $\mathbb{E}X = \frac{1}{\nu-2}, \nu > 2$

    $\mathbb{V}ar(X) = \frac{2}{(\nu-2)^2(\nu-4)}, \nu > 4$.
- (역척도 카이제곱분포) $X \sim Inv\text{-}\chi^2(\nu, \lambda) = Inv\text{-}Ga\left(\frac{\nu}{2}, \frac{\nu}{2}\lambda\right)$.
  - 밀도함수 : $f(x) = \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{1}{x}\right)^{\frac{\nu}{2}+1} e^{-\frac{\nu\lambda}{2x}}, x > 0$.
  - $\mathbb{E}X = \frac{1}{\nu-2}\lambda, \nu > 2$

    $\mathbb{V}ar(X) = \frac{2\nu^2\lambda^2}{(\nu-2)^2(\nu-4)}, \nu > 4$.

정규모형 : 평균을 알고 분산에 대해 추론하는 경우

모형 θ는 아는 값이고, $\sigma^2 > 0$에 대해 추론하고자 한다.

$$x_1, \cdots, X_n \mid \sigma^2 \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$$

사전분포

$$\sigma^2 \sim Inv - \chi^2(\nu_{0,} \sigma_0^2)$$

$$\Leftrightarrow \tau^2 = \frac{1}{\sigma^2} \sim \chi^2(\nu_{0,} \sigma_0^2) = Ga\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

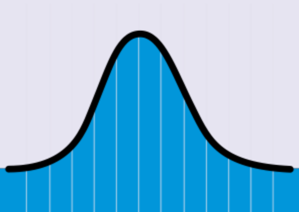
사후분포

$$\tau^2 | x \sim \chi^2\left(\nu_n = \nu_0 + n,\ \sigma_n^2 = \frac{\nu_0 \sigma_0^2 + ns^2}{\nu + n}\right)$$

혹은 $\sigma^2 | x \sim \mathrm{Inv} - \chi^2(\nu_{n,}\ \sigma_n^2)$

여기서 $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \theta)^2$.

목차



추정: 세 가지 사후분포의 한 점 요약

사후분포의 평균

$$\hat{\theta}^{M} \;:=\; \mathbb{E}\left(\theta|x\right) = \int \theta\pi(\theta|x)d\theta\,.$$

최대사후밀도추정량

$$\hat{\theta}^{MAP} \;:=\underset{\theta}{\operatorname{argmax}}\; \pi(\theta|x)$$

- 사후밀도함수의 모드 혹은 최대값
- MAP (maximum a posteriori) 추정량이라고도 한다.
- 사전분포가 균등분포일 때 최대가능도 추정량

사후분포의 중앙값

$$\hat{\theta}^{med} \;:= \mathrm{med}^{\pi(\theta|x)}(\theta)\,.$$

참고: 최대가능도추정량

- (정의)

$$\hat{\theta}^{MLE} := \underset{\theta}{\operatorname{argmax}} \ L(\theta).$$

여기서 $L(\theta)$는 모형의 가능도로 iid 모형의 경우

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

이다.

- 최대가능도추정량(MLE, maximum likelihood estimator)는 가능도를 최대로 하는 θ 값이다.

추정의 두 가지 예

이항모형

- (사전분포) $\theta \sim Beta(\alpha, \beta), \alpha, \beta > 0$
- (모형) $x|\theta \sim Bin(n, \theta)$
- (사후분포)

$$\theta|x \sim Beta(\alpha + x, \beta + n - x), x = 0, 1, \cdots, n.$$

- (문제) 세 가지 추정량을 구해보자.

추정의 두 가지 예

정규모형

- (모형) σ^2 은 아는 값이고, θ는 모르는 값이다.

$$x_1, \cdots, X_n \; | \theta \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$$

- (사전분포) $\theta \; \sim \; N(\mu, \tau^2)$

- (사후분포)

$$\theta | x \sim N\left(\frac{\frac{1}{\tau^2}\mu + \frac{n \cdot}{\sigma^2}\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right).$$

여기서 $\mathrm{x} = (x_1, \ldots, x_n)$이다.

- (문제) 세 가지 추정량을 구해보자.

추정오차(estimation error)

- 추정치 δ의 사후분포의 표준편차와 분산을 δ의 추정오차로 쓴다.
- 사후분포의 산포에 관한 측도로 사후분포의 분산과 표준편차를 쓴다.

예

사후분포가

$$\theta|x \sim N\left(\frac{\frac{1}{\tau^2}\mu + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}, \left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)^{-1}\right)$$

와 같이 주어질 때, 사후분포의 분산과 표준편차를 구해보자.
또한 추정량 δ(x) =$\bar{x}$의 사후분산을 구해보자

목차

신용집합(credible set)

$100(1-\alpha)\%$ 신용집합의 정의

집합 $C \subseteq \Theta$이 $\pi(C|x) \geq 1-\alpha$를 만족하면, C를 θ의 $100(1-\alpha)\%, 0 \leq \alpha \leq 1$ 신용집합이라고 한다.

최고사후밀도(Highest posterior density, HPD) 신용집합

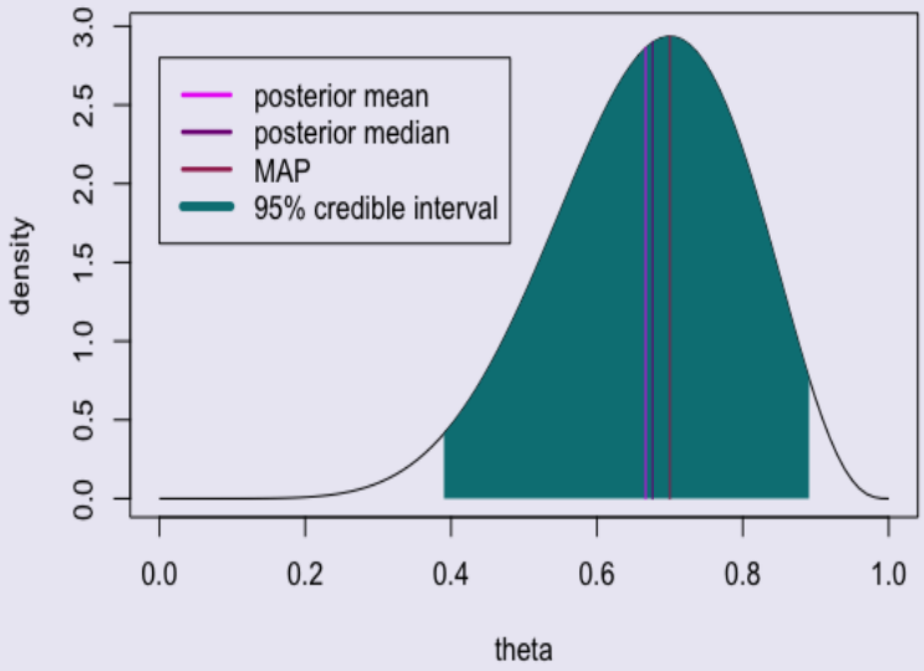
θ의 $100(1-\alpha)\%$ 최고사후밀도 신용집합 $C \subseteq \Theta$는 $C = \{\theta \in \Theta : \pi(\theta|x) \geq k\}$와 같이 정의된다. 여기서 k는 $P(C|x) \geq 1-\alpha$가 되도록 정해진다.

동일꼬리 신용집합(equal tail credible set)

θ의 $100(1-\alpha)\%$ 동일꼬리 신용집합은 $[A, B]$와 같이 정의되고,A, B는 $\pi(\theta \leq A|x) = \frac{\alpha}{2}$와$\pi(\theta \geq B|x) = \frac{\alpha}{2}$ 을 만족한다.

정규모형의 예

압정의 예



- θ의 95% 신용구간은

$$(0.390, 0.891)$$

이다.

구간추정의 해석

- 빈도론자의 95% 신뢰구간 해석

  동일한 실험을 무한히 반복해서 동일한 방법으로 신뢰구간들을 계산하면 이 중 95%의 신뢰구간이 모수 θ를 포함한다.

  이 해석은 당장 우리 손 안에 있는 신뢰구간에 대해 아무런 말도 하지 못한다.

- 베이즈주의자의 95% 신용구간에 대한 해석

  주어진 자료에 대해, 계산된 신용구간이 모수를 포함할 확률이 95%이다.

확률의 의미

- 빈도론적 해석.

  빈도론적 해석은 무한히 반복 가능한 실험에 대해서만 적용 가능하다. 사건 A의 확률은

$$\frac{\text{사건 A의 빈도}}{\text{실험의 반복수}}.$$

- 주관적 확률의 해석.

  개인적 믿음의 정도

- 베이즈주의자들은 주관적 확률의 해석을 믿고 모든 불확실성은 확률로 표현 가능하다고 믿는다.

베이즈주의자들이 믿는 것들

- 미지의 모수에 대한 추론은 자료(와 사전분포)에 대해서만 의존해야 하고 관측할 수도 있었던 값들에 의존하면 안된다.

  놀랍게도 빈도론적 추론에서는 관측될 수도 있으리라 믿어지는 값들이지만 정작 관측되지는 않은 값들에 의해 추론이 의존한다.

- 모든 불확실성은 확률로 표현 가능하다.

  따라서 모수에 대한 불확실성은 자료가 관측되었건 안되었건 확률분포로 표현 가능하다.

신용집합(credible set): 정규모형의 예

집합 $C \subseteq \Theta$ 이 $\pi(C|x) \geq 1-\alpha$를 만족하면, C를 θ의 $100(1-\alpha)\%, 0 \leq \alpha \leq 1$ **신용집합**이라고 한다.

θ의 $100(1-\alpha)\%$ **최고사후밀도 신용집합** $C \subseteq \Theta$ 는 $C = \{\theta \in \Theta : \pi(\theta|x) \geq k\}$와 같이 정의된다. 여기서 k는 $P(C|x) \geq 1-\alpha$가 되도록 정해진다.

θ의 $100(1-\alpha)\%$ **동일꼬리 신용집합**은 $[A, B]$와 같이 정의되고, A, B는 $\pi(\theta \leq A|x) = \frac{\alpha}{2}$와 $\pi(\theta \geq B|x) = \frac{\alpha}{2}$을 만족한다

신용집합(credible set): 정규모형의 예

- (모형) σ^2은 아는 값이고, θ는 모르는 값이다.

$$x_1, \cdots, X_n \mid \theta \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$$

- (사전분포) $\theta \sim N(\mu, \tau^2)$

- (사후분포)

$$\theta | x \sim N\left(\frac{\frac{1}{\tau^2}\mu + \frac{n\cdot}{\sigma^2}\bar{x}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \right).$$

여기서 $x = (x_1, \ldots, x_n)$이다.

- (문제) θ의 95% 신용구간을 구하자.

다음시간

03 강

R의 소개와 실습