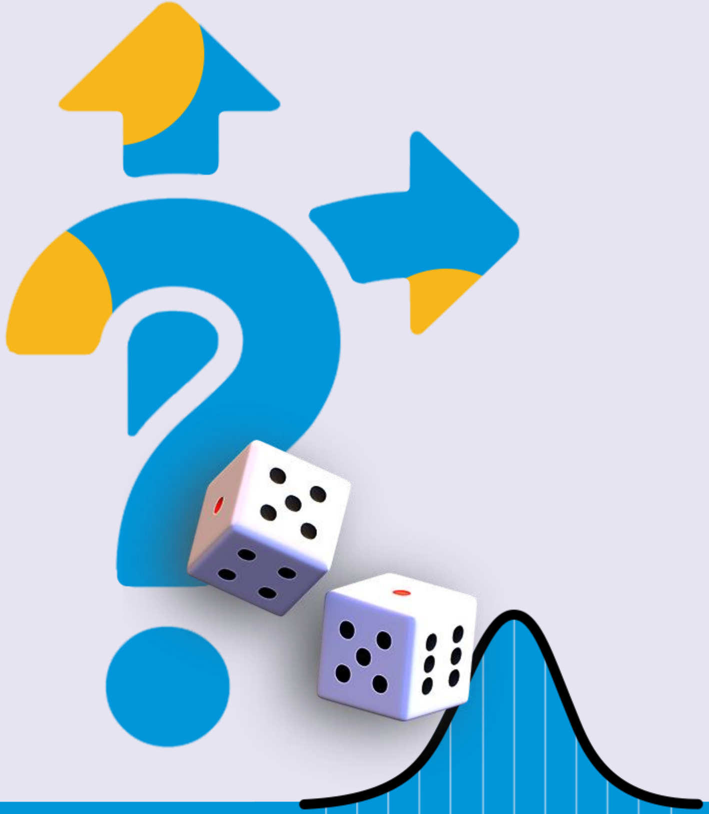
베이즈데이터분석 / 이재용 교수

12강

모형 선택과 진단



한국방송통신대학교

목차

- 예. 세 개의 모형
- 모형확률을 통한 모형 선택
- 예측값을 이용한 모형 선택과 진단

목차

- 예. 세 개의 모형
- 모형확률을 통한 모형 선택
- 예측값을 이용한 모형 선택과 진단

예. 세 개의 모형

데이터

$$
\begin{aligned}
x &= (x_1, \ldots, x_n) \\
&= (8.559, 8.343, 8.095, 8.783, 9.748, \\
&\quad 9.671, 10.910, 9.779, 11.121, 16.768)
\end{aligned}
$$

데이터 탐색

```
The decimal point is at the |

 8 |1368778
10 |91
12 |
14 |
16 |8
```

	x
Min.	: 8.095
1st Qu.	: 8.615
Median	: 9.710
Mean	: 10.178
3rd Qu.	: 10.627
Max.	: 16.768

예. 세 개의 모형

세 개의 모형

$$M_0: x_1, \ldots x_n | \theta \overset{i.i.d.}{\sim} N(\theta, 1)$$

$$M_1: x_1, \ldots x_n | \theta \overset{i.i.d.}{\sim} Ca(\theta, 1)$$

$$M_2: x_1, \ldots, x_n | \lambda \overset{i.i.d.}{\sim} Exp(\lambda).$$

목차

- 예. 세 개의 모형
- 모형확률을 통한 모형 선택
- 예측값을 이용한 모형 선택과 진단

베이즈 모형 선택의 틀

문제

J개의 모형 중 선택.

m : 모형 $(m = 1, \cdots, J)$

모형

모형이 $m = j$일 때,

$$x|\theta_j, m = j \sim f(x|\theta_j, m = j)$$

이라 하자.

사전분포

$$m \sim \pi(m), m = 1, 2, \ldots, J$$

$$\theta_j | m = j \sim \pi(\theta_j | m = j).$$

베이즈 모형 선택의 틀

사후확률

$$\pi(m = j|x) \propto \pi(m = j)f(x|m = j)$$

여기서

$$f(x|m = j) = \int_{\theta_j} f(x|\theta_j, m = j)\pi(\theta_j|m = j)d\theta_j$$

으로 모형 $m = j$ 하에서 관측치 x의 주변 밀도 함수

사후확률

- 하나의 모형 $m = 0$를 기준으로 베이즈 인자를 이용한 사후확률의 계산

$$\pi(m = j|x) = \frac{\pi(m = j)f(x|m = j)}{\sum_{m=h\in\mathcal{M}} \pi(m = h) f(x|m = h)}$$

$$= \frac{\pi(m = j)B_{j0}}{\pi(m = 0)\mathrm{X}1 + \sum_{h=1}^{M} \pi(m = h)B_{h0}}.$$

여기서

$$B_{h0} = \frac{f(x|m = h)}{f(x|m = 0)}.$$

예. 세 개의 모형

모형 M_0

$$x_1, \ldots x_n | \theta \overset{i.i.d.}{\sim} N(\theta, 1)$$

$$\theta \sim N(\mu = 9.7, \sigma^2 = 7.1^2)$$

θ의 사전분포는 데이터에 비해서 충분히 퍼져있도록 잡았다.

$$med_i x_i = 9.7$$

$$\max_j |med_i x_i - x_j| = 7.1.$$

M_0하의 x의 주변밀도함수

$$f\,(x|m=0) = \int \prod_{i=1}^{n} N\,(x_i|\theta, 1) N(\theta|\mu, \sigma^2) d\theta$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{n\sigma^2+1}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2 + n\bar{x}^2 + \frac{\mu^2}{\sigma^2} - \frac{(n\bar{x} + \frac{\mu}{\sigma})^2}{n + \frac{1}{\sigma}} \right\}.$$

예. 세 개의 모형

모형 M_1

$$x_1, \ldots x_n | \theta \overset{i.i.d.}{\sim} Ca(\theta, 1)$$

$$\theta \sim N(\mu = 9.7, \sigma^2 = 7.1^2)$$

M_1 하의 x 의 주변밀도함수

$$f(x|m = 1) = \int \prod_{i=1}^{n} Ca\,(x_i|\theta, 1) N(\theta|\mu, \sigma^2) d\theta$$

$$= \mathbb{E}\left(\frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1 + (x_i - \theta)^2}\right), \theta \sim N(\mu, \sigma^2).$$

예. 세 개의 모형

모형 M_2

$$x_1, \dots x_n | \lambda \overset{i.i.d.}{\sim} Exp(\lambda)$$

$$\lambda \sim Ga(\alpha, \beta)$$

M_2하의 x의 주변밀도함수

$$f(x|m = 2) = \int \prod_{i=1}^{n} Exp(x_i|\lambda)\, Ga(\lambda|\alpha, \beta) d\theta$$

$$= \frac{\beta^{\alpha}}{(\beta + \sum_i x_i)^{\alpha+n}} \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}.$$

사전분포의 설정

$$\frac{\alpha}{\beta} = \mu = 9.7$$

$$\frac{\alpha}{\beta^2} = \sigma^2 = 7.1^2$$

이라 놓고 α와 β를 구하면

$$\alpha = \frac{\mu^2}{\sigma^2}$$

$$\beta = \frac{\mu}{\sigma^2}.$$

예. 세 개의 모형 사후확률

모형의 사전확률

$$(\pi(m_0), \pi(m_1), \pi(m_2)) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

모형의 사후확률

$$(\pi(m_0|x), \pi(m_1|x), \pi(m_2|x)) \propto (f(x|m_0)\pi(m_0), f(x|m_1)\pi(m_1), f(x|m_2)\pi(m_2))$$

$$\propto (7.48 \times 10^{-6}, 9.99 \times 10^{-1}, 3.97 \times 10^{-8})$$

$$(\pi(m_0|x), \pi(m_1|x), \pi(m_2|x)) = (7.48 \times 10^{-6} 0.999, 3.97 \times 10^{-8})$$

모형의 불확실성이 있을 때의 추정

목표

모형의 불확실성이 있을 때

$$\delta = \delta(\theta_j, m = j)$$

를 추정하고자 한다. δ는 모든 모형에서 정의될 수 있어야 한다.

1개의 모형 추정량

여러 개의 모형 중 가장 사후확률이 큰 $\widehat{m}$을 선택하여
$\widehat{m}$하에서 δ를 추정한다.

$$\hat{\delta}^{one,plug} = \delta(\hat{\theta}_{\widehat{m}}, m = \widehat{m})$$

베이즈 모형평균추정량

$$\widehat{\delta}^{BMA} = \sum_{m} \mathbb{E}\,(\delta(\theta_m, m)|x, m)\pi(m|x)$$

$$= \sum_{m} \hat{\delta}^{B}\,(m)\pi(m|x).$$

예. 세 개의 모형 베이즈 모형평균추정량

x_i분포의 중앙값 추정

$$\delta \begin{cases} \theta, & m = 0 \\ \theta, & m = 1 \\ \dfrac{0.693}{\lambda}, & m = 2 \end{cases}$$

예. 세 개의 모형 베이즈 모형평균추정량

δ의 모형평균추정량

각 모형하에서 스탠을 이용해 모수들의 베이즈 추정량을 구하면

$$\hat{\theta}_{m=0} = 9.24$$

$$\hat{\theta}_{m=1} = 3.33$$

$$\hat{\lambda}_{m=2} = 0.11$$

이를 이용해서 δ의 베이즈 모형 평균 추정량을 구하면

$$\hat{\delta}^{BMA} = 7.48 \times 10^{-6} \times 9.24 + 0.999 \times 3.33 + 3.97 \times 10^{-8} \times \frac{0.693}{0.11} = 3.32$$

을 얻는다.

베이즈 인자의 라플라스 근사

- 베이즈 인자는 주변분포의 비로 표현된다.

$m(x) = \int f(x|\theta) \cdot \pi(\theta) d\theta$의라플라스 근사를 이용한 추정량은

$$\hat{m}(x) = \left(\frac{2\pi}{n}\right)^{p/2} |\Sigma(\theta^*)|^{\frac{1}{2}} e^{\log[f(x|\theta^*)\pi(\theta^*)]} \left(1 + O\left(\frac{1}{n}\right)\right)$$

으로 주어진다. 여기서$\theta^* = argmax_\theta f(x|\theta)\pi(\theta)$이고

$$\Sigma(\theta^*) = -\left(\frac{d^2}{\theta^2}\log[f(x|\theta^*)\pi(\theta^*)]\right)^{-1}$$

이다.

베이즈 인자의 라플라스 근사

- 따라서

$$\log \hat{m}(x) = \frac{p}{2}(\log(2\pi) - \log n) + \frac{1}{2}\log|\Sigma(\theta^*)| + \log f(x|\theta^*) + \log \pi(\theta^*)$$

가 된다.

- $-2\log \hat{m}(x)$에서 적당한 값들을 무시한 값이 베이즈정보기준(BIC, Bayesianinformation Criteria)으로

$$-2logf(y|\theta^*) + p\log n$$

와 같이 주어진다.

목차

- 예. 세 개의 모형
- 모형확률을 통한 모형 선택
- 예측값을 이용한 모형 선택과 진단

예측분포

모형

$$x|\theta \sim f(x|\theta)$$

$$\theta \sim \pi(\theta)$$

예측하고자 하는 값

예측하고자 하는 미래의 값 z의 분포는
x와는 독립이고 분포가 같은 카피라 하자.

$$z \sim f(z|\theta)$$

z의 예측분포

$$p(z|x) = \int p(z|x,\theta)\pi(\theta|x)d\theta$$

$$= \int f(z|\theta)\pi(\theta|x)d\theta.$$

사후예측 p 값

- $c(x,\theta)$을 이격도(discrepancy)라 하자.
  큰 값은 모형과 자료가 잘 맞지 않는다는 뜻이다. 사후예측 p 값은

$$p_{post} = \mathbb{P}[c(z,\theta) \; > \; c(x,\theta)|x]$$

  으로 주어진다. 0.01 혹은 0.05와 같이 작은 p_{post}값은 모형이 맞지 않는다는 뜻이다.

- 사후표본을 이용해 쉽게 추정할 수 있다.

$$\frac{1}{T}\sum_{t=1}^{T} I\left(c(z^{(t)},\theta^{(t)}) \; > \; c(x,\theta^{(t)})\right).$$

- 자료를 두 번 써서 보수적으로 판단한다는 비판이 있다.
  (Bayarri and Berger, 2000)

LPML (log pseudo marginal likelihood)

- 데이터 $\mathrm{x} = (x_1, \ldots, x_n)$
- (CPO) 모형이 관측치 x_i를 잘 설명한다면,
  x_i를 뺀 데이터 $\mathrm{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$가 주어졌을 때
  x_i의 조건부 밀도 함수의 확률이 큰 곳에 관측치 x_i가 있어야 한다. 즉,

$$CPO_i = p(x_i | \mathrm{x}_{-i}), i = 1, 2, \ldots, n$$

  값이 커야 한다.

  CPO는 conditional predictive ordinate의 약자이다.

LPML (log pseudo marginal likelihood)

- (LPML의 정의)

$$LPML := \sum_{i=1}^{n} \log CPO_i$$

- 여러 개의 모형을 비교할 때, 각각의 모형에서 LPML을 구해서 가장 큰 값을 갖는 모형을 LPML의 기준으로 가장 좋은 모형으로 판단한다.
- LPML 은 베이즈 버전의 교차 검증 (cross − validation) 혹은 하나 빼기 교차 검증 (leave − one − out − cross − validation, LOOCV) 이라 볼 수 있다.

DIC (Deviance Information Criterion)

- z는 x와 동일한 분포를 갖는 확률변수로 x와는 서로 독립.

  z는 관측치 x를 얻는 실험을 미래에한번 더 했을 때 얻는 값.

- 우리가 추정한 모형이 데이터를 잘 설명한다면

$$-2\log f(z|\tilde{\theta}(x))$$

  값은 작아야 한다. 여기서 $\tilde{\theta}(x)$는 x에 기반한 θ의 추정량.

- 하나의 z가 아니라 무한히 많은 z에 대해 평가를 하고자

$$\mathbb{E}_{z|\theta_0}(-2\log f(z|\tilde{\theta}(x)))$$

  을 비교한다.

DIC (Deviance Information Criterion)

- DIC는 $\mathbb{E}_{z|\theta_0}(-2\log f(z|\tilde{\theta}(x)))$의 추정량으로

$$DIC := D(\bar{\theta}) + 2p_D = 2\overline{D(\theta)} - D(\bar{\theta})$$

으로 주어진다. 여기서 $D(\theta) := -2\log f(x|\theta)$, $\bar{\theta}$는 θ의 사후 평균,

$p_D = \overline{D(\theta)} - D(\bar{\theta})$는 모형의 차원의 추정량, $\overline{D(\theta)}$는 $D(\theta)$의 사후 평균이다.

다음시간

13강

선형 회귀 모형