Chapter 2: Introduction to Bayesian Data Analysis

강의 목표

- 통계 추론의 개념
  - Parameter
  - Prediction
  - 예 1) 회귀분석
  - 예 2) Z-테스트
- Bayesian vs. Frequentist
- 베이지 법칙의 이해

Bayesian Statistical Inference

- Bayesian statistical inference uses Bayes' Law to combine prior information and sample data to make conclusions about a parameter of interest.
- Questions:
  1. What is **Statistical Inference**?
  2. What is a **Parameter**?
  3. What is **Bayes'Law**?

Statistical Inference

- From numerical data, drawing conclusions about quantities that are not observed.
- 통계의 표본정보(sample)에 근거하여 전체 모집단의 특성, 파라미터 등을 추정하는 과학적 과정이다.
- Two Types:
  1. 가설 검정: Quantities that are not directly observable (parameter)
  2. 통계적 추정: Potentially observable quantities (prediction)

Example I: Regression Analysis

Notations

- y: response, output
- $x = (x_1, x_2, \ldots, x_p)$: predictors, input

Goal: model the relationship between y and $x_1, \ldots, x_p$

Example I: Regression Analysis

- General form: $y = f(x) + \epsilon$
  - $f(\cdot)$: underlying truth. Unknown
    - Linear: $f(x) = \beta_0 + \beta_1 x$
    - Polynomial: $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$
    - More complicated: $f(x) = \beta_0(cos(\beta_1 x) + sin(\beta_2 x))$
    - Non-parametric: No β's
  - ϵ: error

Example II: Estimation of Population Mean or Probability

Notations

- x: target variable, input

Goal: Find the statistics for the target variable.

Example II: Estimation of Population Mean or Probability

Let x has normal distribution.

$$x \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

Then we can estimate μ and σ^2. And these are called parameters.

Example: Baby Birth Weight

- From previous experience we know that the birth weights of babies in England are Normally distributed with a mean of 3000g and a standard deviation of 500g.
- We think that maybe babies in Australia have a mean birth weight greater than 3000g and we would like to test this hypothesis.

Motivation for Bayesian Modeling

- Bayesians treat unobserved data and unknown parameters in similar ways (random).
- Frequentists treat unknown parameters as constant.
- For $P(y \mid \mu, \sigma^2)$, Bayesians treat μ and σ^2 as random, while Frequentists treat them as constants.
- Therefore, Bayesians describe each parameter with a probability distribution.

Motivation for Bayesian Inference

- Bayesian inference differs from classical inference in that it specifies a probability distribution for the parameter of interest.
- 예를 들어 신뢰구간을 어떻게 해석하는가?

Motivation for Bayesian Inference

- As their model, Bayesians specify:
    1. A joint density function, which describes the form of the distribution of the **full sample** of data (given the parameter values)
    2. A prior distribution, which describes the behavior of the parameter(s) **unconditiona**l on the data.

Motivation for Bayesian Inference

- The prior could reflect:
  1. Uncertainty about a parameter that is actually fixed.
  2. The variety of values that a truly stochastic parameter could take.

Different Interpretations of Probability

Frequentists definition of the probability of an event: If we **repeat** an experiment a very large number of times, what is the proportion of times the event occurs?

- **Problem**: For some situations, it is impossible to repeat (or even conceive of repeating) the experiment many times.
- **Example**: The probability that Governor Haley is re-elected in 2014.

Different Interpretations of Probability

- Subjective probability: Based on an individual's degree of belief that an event will occur.
  - **Example**: A bettor is willing to risk up to \$200 betting that Haley will be re-elected, in order to win \$100. The bettor's subjective P[Haley wins] is 2/3.
  - The Bayesian approach can naturally incorporate subjective probabilities about the parameter, where appropriate.

Bayes' Law

- Suppose that there are two events A and B.
- Bayes' Law relates the conditional probabilities.
- Recall that

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

  and

$$P(B \mid A) = \frac{P(A, B)}{P(A)}$$

- Hence,

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

Bayes' Law

- Then,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

  and

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

Bayes' Law

- Marginal distribution of Y: $P(Y)$.
- Joint distribution of X, Y: $P(X, Y)$.
- Recall that
$$P(X) = \sum_Y P(X, Y).$$
- Then,

$$P(X) = \sum_Y P(X, Y) = \sum_Y P(X \mid Y) P(Y).$$

Bayes' Law

- Then,
$$P(x) = \sum_y P(x, y) = \sum_y P(x \mid y)P(y).$$

Bayes' Law Example

- **Example**: 1975 British national referendum on whether the UK should remain part of the European Economic Community
- Suppose 52% of voters supported the Labour Party and 48% the Conservative Party. Suppose 55% of Labour voters wanted the UK to remain part of the EEC and 85% of Conservative voters wanted this.
- Let **L** denote "Labour" and **Y** denote "Yes".
- What is the probability that a person voting "Yes" to remaining in EEC is a Labour voter?

$$P(L \mid Y) = \frac{P(Y \mid L)P(L)}{P(Y)}.$$

Bayes' Law Example

A simple probability rules implies that

$$P(Y) = P(Y, L) + P(Y, L^c) = P(Y \mid L)P(L) + P(Y \mid L^c)P(L^c).$$

Hence,

$$\begin{aligned} P(L \mid Y) &= \frac{P(Y \mid L)P(L)}{P(Y \mid L)P(L) + P(Y \mid L^c)P(L^c)} \\ &= \frac{(.55)(.52)}{(.55)(.52) + (.85)(.48)} \\ &= 0.41. \end{aligned}$$

Bayes' Law with Multiple Events

- Let **D** represent some observed data and let A, B, and C be mutually exclusive (and exhaustive) events conditional on **D**.
- Note that

$$
\begin{aligned}
P(D) &= P(A \cap D) + P(B \cap D) + P(C \cap D) \\
&= P(D \mid A)P(A) + P(D \mid B)P(B) + P(D \mid C)P(C).
\end{aligned}
$$

- Using Bayes' Law,

$$
\begin{aligned}
P(A \mid \mathbf{D}) &= \frac{P(\mathbf{D} \mid A)P(A)}{P(\mathbf{D})} \\
&= \frac{P(\mathbf{D} \mid A)P(A)}{(\mathbf{D} \mid A)P(A) + P(\mathbf{D} \mid B)P(B) + P(\mathbf{D} \mid C)P(C)}.
\end{aligned}
$$

Bayes' Law with Multiple Events

- Denoting A, B, C by $\theta_1, \theta_2, \theta_3$, we can write this more generally as

$$P(\theta_i \mid D) = \frac{P(\theta_i)P(D \mid \theta_i)}{\sum_{j=1}^{3} P(\theta_j)P(D \mid \theta_j)}.$$

- If there are k distinct discrete outcomes $\theta_1, \theta_2, ..., \theta_k$, we obtain for any $i \in \{\{1, 2, ..., k\}$:

$$P(\theta_i \mid D) = \frac{P(\theta_i)P(D \mid \theta_i)}{\sum_{j=1}^{k} P(\theta_j)P(D \mid \theta_j)}.$$

- The denominator equals $P(D)$, the marginal distribution of the data.
- Note if the values of θ are portions of the continuous real line, the sum may be replaced by an integral.

Bayes' Law Example (4 Classes)

- In the 1996 General Social Survey, for males (age 30+):
  - 11% of those in the lowest income quartile were college graduates.
  - 19% of those in the second-lowest income quartile were college graduates.
  - 31% of those in the third-lowest income quartile were college graduates.
  - 53% of those in the highest income quartile were college graduates.
- What is the probability that a college graduate falls in the lowest income quartile?

Bayes' Law Example (4 Classes)

- G: college graduate, Q_j: j^{th} quartile.

$$
\begin{aligned}
P(Q_j \mid G) &= \frac{P(Q_1)P(G \mid Q_1)}{\sum_{j=1}^{4} P(Q_j)P(G \mid Q_j)} \\
&= \frac{.11 \cdot .25}{.11 \cdot .25 + .19 \cdot .25 + .31 \cdot .25 + .53 \cdot .25} \\
&= 0.09.
\end{aligned}
$$

- Find $P(Q_2 \mid G)$.
- How does this **conditional** distribution differ from the **unconditional** distribution $P(Q1), ..., P(Q4)$?

Bayes' Law Example II

- 갓태어난 수지의 친아빠가 누구인지 밝혀내는 소송에서 다음과 같은 정보가 주어졌다. 엄마의 혈액형은 O형이며 아빠로 지목된 남자 Albert의 혈액형은 AB형이다. 소송을 진행하는 과정에서 수지의 혈액형을 조사하니 B형으로 나타났다. 수지의 혈액형이 B라는 사실을 사건 B로, Albert가 수지의 아빠일 사건을 F라고 하자.

Bayes' Law Example II

- 우리가 구하고자 하는 것은 $P(F \mid B)$이다. 그런데 혈액형 이외에, 그동안 수지의 엄마와 Albert와의 관계 또는 주변사람들의 증언 등에 의하여 Albert가 수지의 아빠일 가능성 $P(F)$를 추측할 수 있을 것이다. 이때 $P(F)$는 자료, 즉 혈액형을 측정하기 이전의 확률이므로 사전확률이라고 한다. 반면 $P(F \mid B)$는 자료측정 이후의 확률이므로 사후확률이라고 한다.
- 베이즈 정리에 의하여 $P(F \mid B)$를 구하면

$$P(F \mid B) = \frac{P(B \mid F)P(F)}{P(B \mid F)P(F) + P(B \mid F^c)P(F^c)}$$

Bayes' Law Example II

- 이제 $P(B \mid F)$와 $P(B \mid F^c)$를 구하자.
- Albert가 수지의 아빠일 경우 멘델의 법칙에 의하여 $P(B \mid F) = 0.5$이다. 또 $P(B \mid F^c)$를 생각해 보면 Albert가 수지의 아빠가 아닐 경우에도 수지의 혈액형이 B가 나올 수 있는데 이를 전체 코카시안 중에서 B형인 비율인 9%로 놓기로 하자. 이들을 위의 식에 대입하면

$$P(F \mid B) = \frac{0.5 \times P(F)}{0.5 \times P(F) + 0.09 \times P(F^c)} = \frac{50P(F)}{41P(F) + 9}.$$

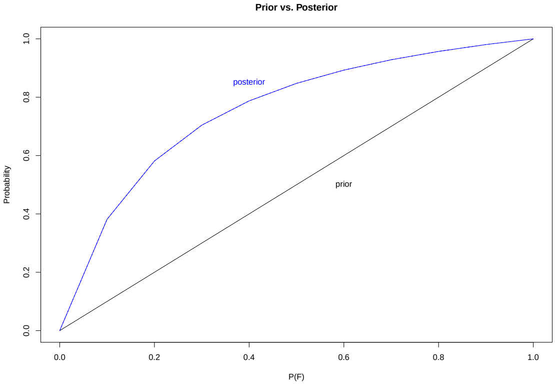
Bayes' Law Example II

- $P(F)$와 $P(F \mid B)$를 비교해보자.
- R을 이용한 simulation:

```
> aa <- function(pf){
+   posterior = 50*pf/ (41*pf +9)
+   return(posterior)
+ }
> x <- seq(0,1, by = 0.1)
> x
[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
> prob <- aa(x)
> prob
[1] 0.0000000 0.3816794 0.5813953 0.7042254 0.7874016
0.8474576 0.8928571 0.9283820 0.9569378 0.9803922 1.0000000
```

Bayes' Law Example II

```
> plot(x, x, type ="l", xlab ="P(F)", ylab="Probability")
> lines(x, prob, lty = 1, col=4)
> text(0.6,0.5,"prior"); text(.4,.85,"posterior",col=4)
> title("Prior vs. Posterior");
```



Bayes' Law Example II

- 사전확률 $P(F)$가 0 혹은 1에 가깝지 않은 경우 $P(F)$에 비하여 $P(F \mid B)$가 상당히 큼을 알 수 있는데, 이는 사전증거가 확실치 않은 경우 혈액형에 의한 증거가 상당한 영향을 미침을 의미한다.