확률 및 통계학

- 2장 확률 (Probability) -

김 지 혜 (jihye@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

- 보충
- 베이즈 정리(Bayes' Rule)

목 차

- 보충
- 베이즈 정리(Bayes' Rule)

보충

- 조건부 확률(Conditional Probability)
  - 독립사상(Independent Event)
    - 정의

정의 2.11

$P(B|A) = P(B)$이거나 $P(A|B) = P(A)$이면 두 사상 A와 B는 **독립**이다.

- 두 사상 A와 B에서 한 사상의 결과가 다른 사상에 영향을 주지 않음

    - 예제

카드 52장 중 2장의 카드를 연속적으로 복원추출하는 실험에서 첫 번째 카드가 A일 사상(A)과 두 번째 카드가 스페이드일 사상(B)에 대해 A와 B는 독립인가 종속인가?

$$P(B) = \frac{13}{52}, P(B|A) = \frac{13}{52}$$

$\therefore P(B|A) = P(B)$이므로 사상 A와 B는 독립

보충

- 조건부 확률(Conditional Probability)
  - 종속사상(Dependent Event)
    - 정의

정의 2.12

$P(B|A) \neq P(B)$이거나 $P(A|B) \neq P(A)$이면 두 사상 A와 B는 **종속**이다.

- 두 사상 A와 B에서 한 사상의 결과가 다른 사상에 영향을 줌

    - 예제

짝수가 홀수보다 2배만큼 더 많이 발생하는 주사위를 던져서 3보다 큰 수가 나올 사상(A)과 완전제곱수를 얻을 사상(B)에 대해 A와 B는 독립인가 종속인가?

- $A = \{4, 5, 6\}, B = \{1, 4\}, B|A = \{4\}$

$$P(B) = \frac{3}{9} = \frac{1}{3}, P(B|A) = \frac{2}{5}$$

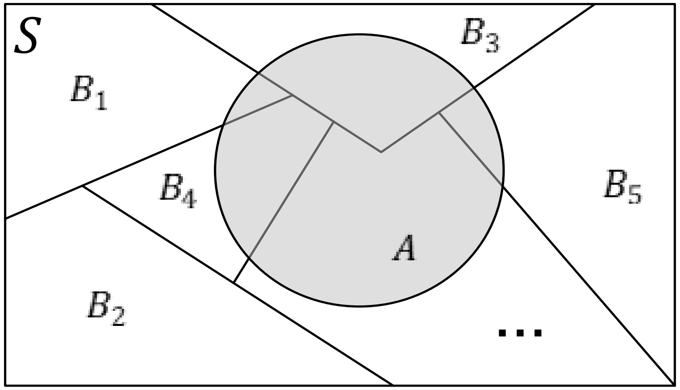
$\therefore P(B|A) \neq P(B)$이므로 사상 A와 B는 종속

목 차

- 보충
- 베이즈 정리(Bayes' Rule)

베이즈 정리(Bayes' Rule)

- 전확률의 정리(Theorem of Total Probability)
  - 정의
    - 서로 배반인 사상 $B_1, B_2, \cdots, B_n$이 표본공간 S의 분할일 때, 사상 A를 구하는 공식



<그림 22> 표본공간 S의 분할

정리 2.13

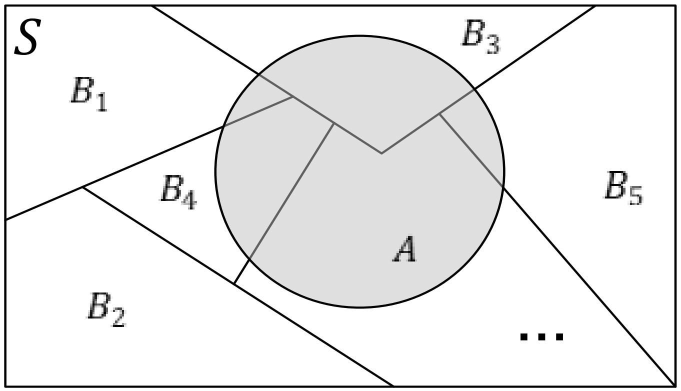
사상 $B_1, B_2, \cdots, B_k$를 표본공간 S의 분할이라 하고, $P(B_i) \neq 0,\ i = 1, 2, \cdots, k$ 라 하면, S의 임의의 사상 A에 대하여

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

가 성립한다.

베이즈 정리(Bayes' Rule)

- 전확률의 정리(Theorem of Total Probability)
  - 증명



<그림 22> 표본공간 S의 분할

정리 2.10

두 사상 A와 B가 동시에 발생할 수 있다면

$$P(A \cap B) = P(A)P(B|A)$$

가 된다.

따름정리 2.2

사상 $A_1, A_2, \ldots, A_n$이 서로 배반이면

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

이다.

- 그림 22에서 사상 A는 서로 배반인 사상 $B_1 \cap A, B_2 \cap A, \cdots, B_k \cap A$의 합집합, 즉 $A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A)$임을 알 수 있다. 따름정리 2.2와 정리 2.10에 의해

$$P(A) = P[(B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A)]$$

$$= P(B_1 \cap A) + P(B_2 \cap A) + \cdots + P(B_k \cap A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

가 성립한다.

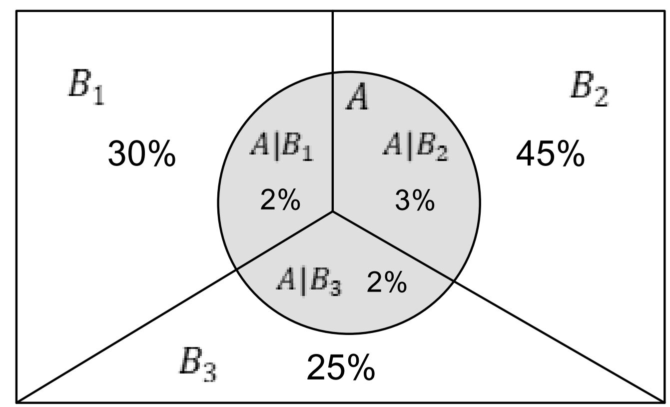
베이즈 정리(Bayes' Rule)

- 전확률의 정리(Theorem of Total Probability)
  - 예제 2.41

3대의 기계 B_1, B_2, B_3가 각각 전체생산량의 30%, 45%, 25%를 생산하는 어느 조립공장에서 과거의 경험으로부터 각 기계의 불량품 제조율이 2%, 3%, 2%임이 알려져 있다. 이때 완제품 중에서 임의로 하나를 선택했을 때, 그것이 불량품일 확률은 얼마인가?

- 그 제품이 불량품일 사상: A
- 그 제품이 기계 B_1, B_2, B_3에서 제조되었을 사상 : B_1, B_2, B_3
- 전확률의 정리를 이용하면 다음과 같이 쓸 수 있다.

$$\therefore P(A) = P(B_1 \cap A) + P(B_2 \cap A) + P(B_3 \cap A)$$
$$= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$
$$= 0.3 \times 0.02 + 0.45 \times 0.03 + 0.25 \times 0.02$$
$$= 0.0245$$



<그림 23> 예제 2.41

베이즈 정리(Bayes' Rule)

- 정의
  - 현재 관측된 결과가 이전 경험에 의해 발생되었는지 추론하는 사후 확률을 구하는 공식
    - 사전 확률: 사전에 주어진 확률
      - $P(A)$: 현재 관측된 결과 (전확률의 정리 활용)
      - $P(B_i)$: 이전 경험들
      - $P(A|B_i)$: 이전 경험들이 현재 관측된 결과에 미치는 가능성
    - 사후 확률: 추가 정보를 통해 추론해야 할 확률
      - $P(B_i|A)$: 현재 관측된 결과가 특정한 이전 경험(추가 정보)에 의해 발생되었다고 추론하는 확률
        - e.g., 예제 2.41에서 생산된 불량품이 B_1에서 제조되었을 확률

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(A \cap B_i)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

* 분자: 조건부 확률의 승법 정리 / 분모: 조건부 확률의 전확률의 정리

베이즈 정리(Bayes' Rule)

정리 2.14

사상 $B_1, B_2, \cdots, B_k$가 표본공간 S의 분할이고, 모든 $i(i = 1, 2, \cdots, k)$에 대하여 $P(B_i) \neq 0$이라 하자. 그리고 임의의 사상 A에 대하여 $P(A) \neq 0$이라 하자. 그러면 각 정수값 $r(r = 1, 2, \cdots, k)$에 대하여

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

이 성립한다.

- 증명

  - 조건부 확률의 정의에 의하여

$$P(B_r|A) = \frac{P(B_r \cap A)}{P(A)}$$

이고, 분모에 정리 2.13을 적용하면 다음 결과를 얻을 수 있다.

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

베이즈 정리(Bayes' Rule)

- 예제 2.42 (1/2)

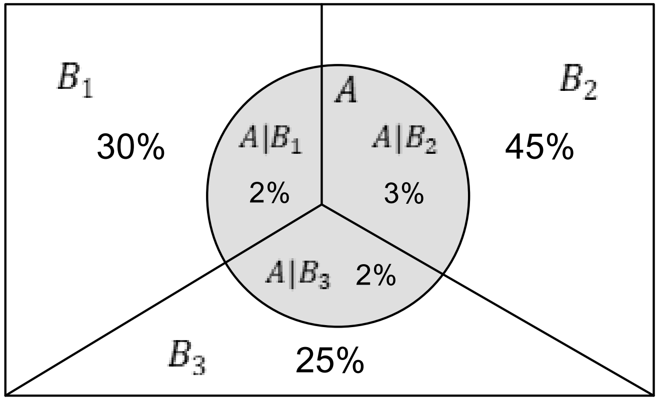
3대의 기계 B_1, B_2, B_3가 각각 전체생산량의 30%, 45%, 25%를 생산하는 어느 조립 공장에서 과거의 경험으로부터 각 기계의 불량품 제조율이 2%, 3%, 2%임이 알려져 있다. 이때 완제품 중에서 불량품을 선택했을 때, 그 제품이 기계 B_3에서 제조되었을 확률은 얼마인가?

① 관측된 결과와 이전 경험 구분

- 관측된 결과: $P(A)$ (완제품 생산 결과, 불량품)
- 이전 경험: 3대의 기계에 따른 세 가지 확률

② 표본공간에서 분할인 사건들(B_i) 정의

- $P(B_1) = 0.3$: B_1에서 생산된 제품의 비율
- $P(B_2) = 0.45$: B_2에서 생산된 제품의 비율
- $P(B_3) = 0.25$: B_3에서 생산된 제품의 비율



<그림 23> 예제 2.41

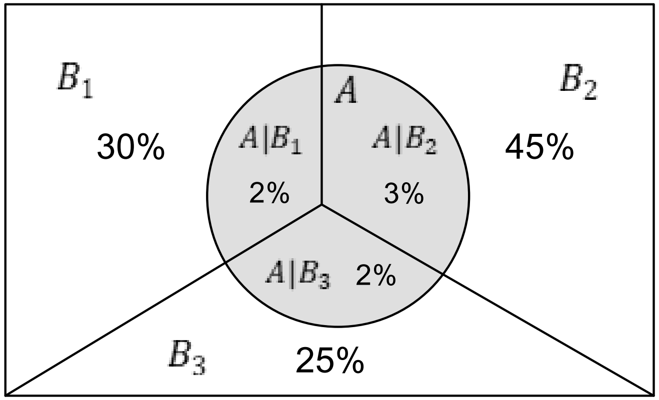
베이즈 정리(Bayes' Rule)

- 예제 2.42 (2/2)

3대의 기계 B_1, B_2, B_3가 각각 전체생산량의 30%, 45%, 25%를 생산하는 어느 조립 공장에서 과거의 경험으로부터 각 기계의 불량품 제조율이 2%, 3%, 2%임이 알려져 있다. 이때 완제품 중에서 불량품을 선택했을 때, 그 제품이 기계 B_3에서 제조되었을 확률은 얼마인가?

③ 각 사건들(B_i)이 현재 관측된 결과에 미치는 가능성

- $P(A|B_1) = 0.02$: B_1에서 불량품이 나올 확률
- $P(A|B_2) = 0.03$: B_2에서 불량품이 나올 확률
- $P(A|B_3) = 0.02$: B_3에서 불량품이 나올 확률



<그림 23> 예제 2.41

④ 완제품 생산 결과가 불량품일 때, B_3에서 제조되었을 확률

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} = \frac{0.25 \times 0.02}{0.3 \times 0.02 + 0.45 \times 0.03 + 0.25 \times 0.02}$$

$$= \frac{10}{49} \fallingdotseq 0.21$$

- 같은 방식으로 B_1의 경우 약 0.24, B_2의 경우 약 0.55이다.
- 따라서, 불량품이 선택되었을 때 그 제품이 B_3에서 제조되지는 않았을 것이다.

베이즈 정리(Bayes' Rule)

- 예제 2.43 (1/3)

어느 제조업체에서 제품 개발 시 세 가지 분석법을 적용하고 있는데, 비용의 문제로 분석법 1, 2, 3을 제품의 30%, 20%, 50%에 각각 적용하고 있다. 세 가지 방법에 따라 불량률은 다음과 같이 각각 다른 것으로 알려졌다.

$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02$$

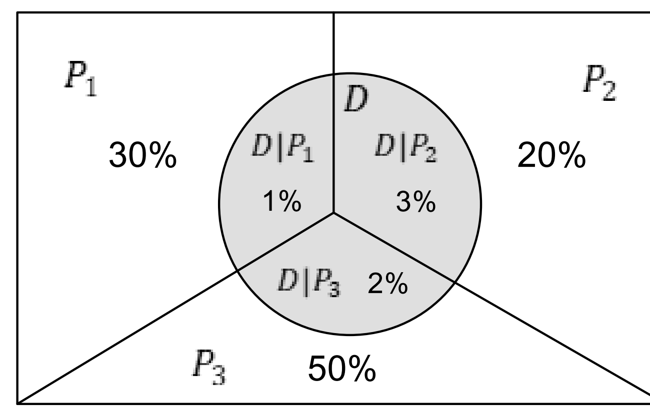
여기에서 $P(D|P_j)$는 분석법 j였을 때 불량품일 확률을 나타낸다. 한 제품을 임의로 선택한 후 검사한 결과 불량이었다면, 어느 분석법을 사용했을 가능성이 가장 많은가?

① 관측된 결과와 이전 경험 구분

- 관측된 결과: $P(D)$ (제품 검사 결과, 불량품)
- 이전 경험: 분석법 $1, 2, 3$에 따른 세 가지 확률

② 표본공간에서 분할인 사건들(P_i) 정의

- $P(P_1) = 0.3$: 분석법 1으로 검사된 비율
- $P(P_2) = 0.2$: 분석법 2로 검사된 비율
- $P(P_3) = 0.5$: 분석법 3으로 검사된 비율



<그림 24> 예제 2.43

베이즈 정리(Bayes' Rule)

- 예제 2.43 (2/3)

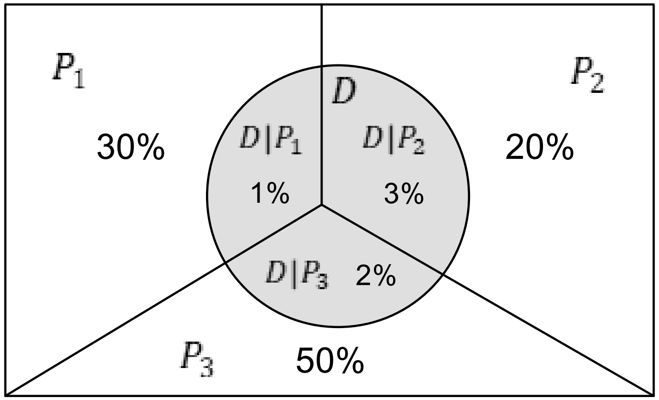
어느 제조업체에서 제품 개발 시 세 가지 분석법을 적용하고 있는데, 비용의 문제로 분석법 1, 2, 3을 제품의 30%, 20%, 50%에 각각 적용하고 있다. 세 가지 방법에 따라 불량률은 다음과 같이 각각 다른 것으로 알려졌다.

$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02$$

여기에서 $P(D|P_j)$는 분석법 j였을 때 불량품일 확률을 나타낸다. 한 제품을 임의로 선택한 후 검사한 결과 불량이었다면, 어느 분석법을 사용했을 가능성이 가장 많은가?

③ 각 사건들(P_i)이 현재 관측된 결과에 미치는 가능성

- $P(D|P_1) = 0.01$: 분석법 1 검사로 불량품이 나올 확률
- $P(D|P_2) = 0.03$: 분석법 2 검사로 불량품이 나올 확률
- $P(D|P_3) = 0.02$: 분석법 3 검사로 불량품이 나올 확률



<그림 24> 예제 2.43

베이즈 정리(Bayes' Rule)

• 예제 2.43 (3/3)

어느 제조업체에서 제품 개발 시 세 가지 분석법을 적용하고 있는데, 비용의 문제로 분석법 1, 2, 3을 제품의 30%, 20%, 50%에 각각 적용하고 있다. 세 가지 방법에 따라 불량률은 다음과 같이 각각 다른 것으로 알려졌다.

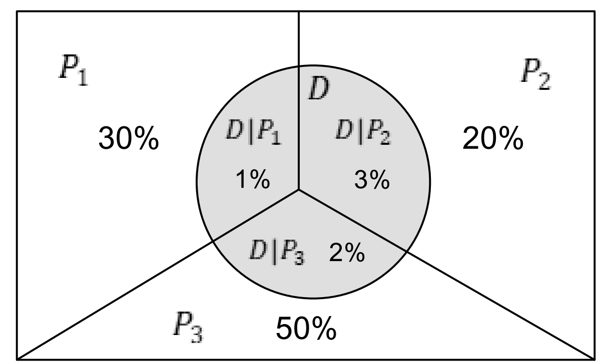
$$P(D|P_1) = 0.01, P(D|P_2) = 0.03, P(D|P_3) = 0.02$$

여기에서 $P(D|P_j)$는 분석법 j였을 때 불량품일 확률을 나타낸다. 한 제품을 임의로 선택한 후 검사한 결과 불량이었다면, 어느 분석법을 사용했을 가능성이 가장 많은가?

④ 제품 검사 결과 불량품일 때, 분석법 1으로 검사했을 확률

$$P(P_1|D) = \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)}$$

$$= \frac{0.3 \times 0.01}{0.3 \times 0.01 + 0.2 \times 0.03 + 0.5 \times 0.02} = 0.158$$



<그림 24> 예제 2.43

- 같은 방법으로 $P(P_2|D) = \frac{0.03 \times 0.2}{0.019} = 0.316,\ P(P_3|D) = \frac{0.02 \times 0.5}{0.019} = 0.526$이다.
- 따라서, 제품이 불량이었을 때 분석법 3일 조건부 확률이 세 가지 중 가장 크므로, 분석법 3을 사용했을 가능성이 가장 크다.

베이즈 정리(Bayes' Rule)

- 예제 2.44 (1/2)

제약사에서 환자가 감기에 걸린지 확인하는 시약을 만들었다. 감기는 전체 인구 중 걸린 사람이 0.2%인 병이다. 감기에 걸린 환자에게 시약을 테스트한 결과 99%의 확률로 양성 반응을 보였다. 하지만, 감기에 걸리지 않은 환자에게 시약 검사를 했을 때, 양성 반응, 즉 잘못된 결과가 나타난 확률이 5%이다. 감기에 걸린지 확인이 되지 않은 어떤 환자가 이 시약을 테스트한 결과, 양성 반응을 보였다면 이 환자가 감기에 걸려 있을 확률은 얼마인가?

① 관측된 결과와 이전 경험 구분

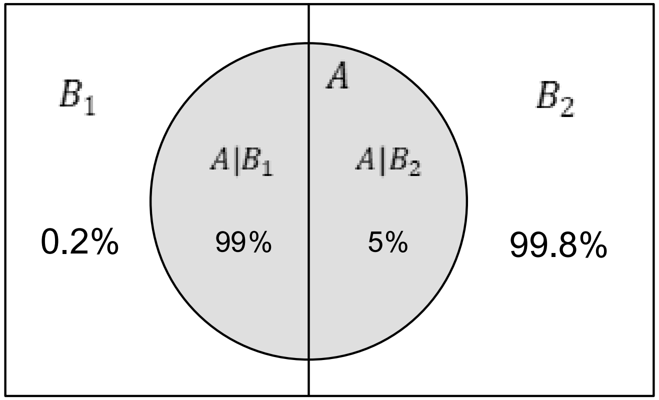
- 관측된 결과: $P(A)$ (시약 테스트 결과, 양성 반응)
- 이전 경험: 감기에 걸린 여부에 따른 두 가지 확률

② 표본공간에서 분할인 사건들(B_i) 정의

- $P(B_1) = 0.002$: 감기에 걸린 사람의 비율
- $P(B_2) = 0.998$: 감기에 걸리지 않은 사람의 비율

③ 각 사건들(B_i)이 현재 관측된 결과에 미치는 가능성

- $P(A|B_1) = 0.99$: 감기에 걸린 사람에게 양성이 나온 확률
- $P(A|B_2) = 0.05$: 감기에 걸리지 않은 사람에게 양성이 나온 확률



<그림 25> 예제 2.44

베이즈 정리(Bayes' Rule)

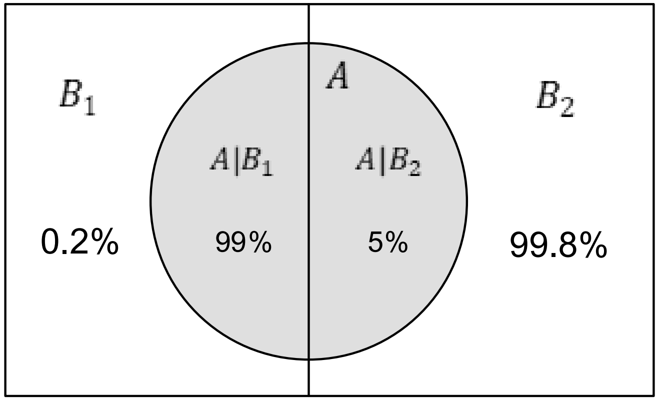
- 예제 2.44 (2/2)

제약사에서 환자가 감기에 걸린지 확인하는 시약을 만들었다. 감기는 전체 인구 중 걸린 사람이 0.2%인 병이다. 감기에 걸린 환자에게 시약을 테스트한 결과 99%의 확률로 양성 반응을 보였다. 하지만, 감기에 걸리지 않은 환자에게 시약 검사를 했을 때, 양성 반응, 즉 잘못된 결과가 나타난 확률이 5%이다. 감기에 걸린지 확인이 되지 않은 어떤 환자가 이 시약을 테스트한 결과, 양성 반응을 보였다면 이 환자가 감기에 걸려 있을 확률은 얼마인가?

④ 양성 반응일 때, 감기에 걸린 사람일 확률

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)}$$

$$= \frac{0.002 \times 0.99}{(0.002 \times 0.99) + (0.998 \times 0.05)} = 0.038$$



<그림 25> 예제 2.44

베이즈 정리(Bayes' Rule)

• 예제 2.45 (1/2)

어느 도시에서 남성의 6%, 여성의 1%의 키가 180cm 이상이다. 그리고 여성은 도시 인구의 62%를 차지한다. 임의로 한 사람을 선택했더니 키가 180cm 이상이라면, 이 사람이 여성일 확률은?

① 관측된 결과와 이전 경험 구분

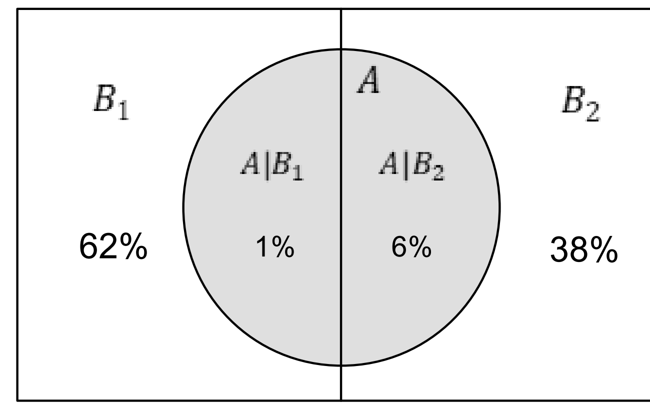
- 관측된 결과: $P(A)$ $(180\text{cm}$인 확률$)$
- 이전 경험: 여성과 남성인 경우에 대한 각각의 확률

② 표본공간에서 분할인 사건들(B_i) 정의

- $P(B_1) = 0.62$: 여성 인구 비율
- $P(B_2) = 0.38$: 남성 인구 비율

③ 각 사건들(B_i)이 현재 관측된 결과에 미치는 가능성

- $P(A|B_1) = 0.01$: 여성인 경우, 키가 180cm 이상인 확률
- $P(A|B_2) = 0.06$: 남성인 경우, 키가 180cm 이상인 확률



<그림 26> 예제 2.45

베이즈 정리(Bayes' Rule)

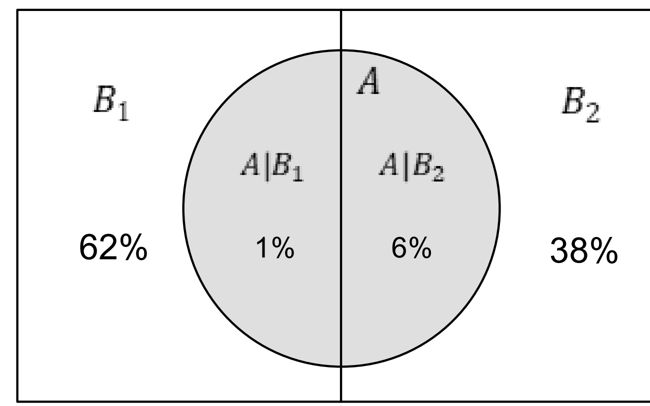
- 예제 2.45 (2/2)

어느 도시에서 남성의 6%, 여성의 1%의 키가 180cm 이상이다. 그리고 여성은 도시 인구의 62%를 차지한다. 임의로 한 사람을 선택했더니 키가 180cm 이상이라면, 이 사람이 여성일 확률은?

④ 180cm 이상인 사람을 선택했을 때, 여성인 확률

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)}$$

$$= \frac{0.62 \times 0.01}{(0.62 \times 0.01) + (0.38 \times 0.06)} = 0.2138$$



<그림 26> 예제 2.45

Thanks!

김 지 혜 (jihye@pel.sejong.ac.kr)