제 4 장
확률이론

목차

제 1절 확률의 정의

- 상대빈도정의
  - 확률을 과학적으로 정의하기 위해서 상대빈도정의와 동등발생정의의 두 방법을 사용
  - 동전을 던졌을 때 앞면이 나올 확률이 1/2라고 어떻게 단정할 수 있는가?
    - 이는 동전을 수업이 많이 던졌을 대 전체시행횟수 중에서 앞면이 나타나는 빈도수가 전체시행횟수의 1/2에 접근한다는 의미
    - 이러한 개념에 입각하여 확률을 정의하는 방법이 상대빈도정의(relative frequency definition)

$$P(A) = \lim_{N \to \infty} \frac{n}{N} \qquad (4 \cdot 1)$$

$P(A)$: A사건이 발생할 확률

N : 총시행횟수

n : A사건이 발생한 횟수

제 1절 확률의 정의

- 상대빈도정의에 의하면
  - “어떤 사건이 나타날 확률은 실험을 무한에 가깝게 계속적으로 시행했을 때, 전체시행횟수에서 그 사건이 나타나는 빈도수를 상대적으로 나타낸 것”
  - 상대빈도정의에 입각한 확률의 정의를 식으로 나타낸 것이 식 4・1
  - 그러나 상대빈도정의에 의해 확률을 정의하면 어떤 사건의 확률을 알기 위해 실험을 무한에 가깝게 반복적으로 시행해야 한다는 문제 발생

제 1절 확률의 정의

- 동등발생정의

  - 동전의 앞면이 나올 확률이 1/2이라는 것은 실험에 근거를 둔 것이라기보다 앞면과 뒷면이 나올 확률은 동일할 것이라는 논리적 추론에 근거를 둔 것

    - 이와 같은 확률의 정의를 동등발생정의(equally likely definition)라고 함

  - 아래의 동등발생정의는 어떤 실험이나 관찰의 결과로 나타날 수 있는 모든 경우들이 각각 동일한 가능성을 가지고 발생할 것이라는 가정하에서 특정사건 A가 일어날 확률을 정의한 것

$$P(A) = \frac{\text{사건 } A\text{에 속하는 경우의 수}}{\text{발생할 가능성이 동일한 전체 경우의 수}} \qquad (4 \cdot 2)$$

제 1절 확률의 정의

- 동등발생정의에 입각한 확률은 나타날 수 있는 전체의 경우 중에서 어떤 특정사건이 차지하는 경우의 구성비율(proportion)과 같은 의미

- 그러므로 “확률=구성비율”이라는 생각을 염두에 두면 됨

  - 예를 들어 어떤 반에서 여자의 비율이 40%라는 것은 한 사람을 무작위로 뽑을 때 여자가 뽑힐 확률이 40%라는 것을 의미

- 이상의 두 가지 확률 정의는 서로 보완적으로 사용될 수 있으며, 어떠한 방법으로 확률을 정의하든 간에 그 결과는 동일

제 2절 집합이론과 확률의 개념

- 집합이론
  - 확률이론을 쉽게 설명하기 위해서는 집합이론의 용어와 부호를 사용하는 것이 편리
    - 집합(set)이란 개체 또는 원소(element)의 모임
    - 원소는 집합을 형성하는 개별사물을 가리키는 것으로서, 그 개별사물은 명확히 정의되어야 함
    - 특정한 문제에서 가능한 모든 원소의 집합을 전체집합(universal set), 이 모든 원소 중 일부를 나타내는 집합을 부분집합이라고 함
  - 확률이론 설명에 필요한 집합연산(set operation)
    - 여집합(complementary sets), 합집합(union of sets), 교집합(intersection of sets) 등

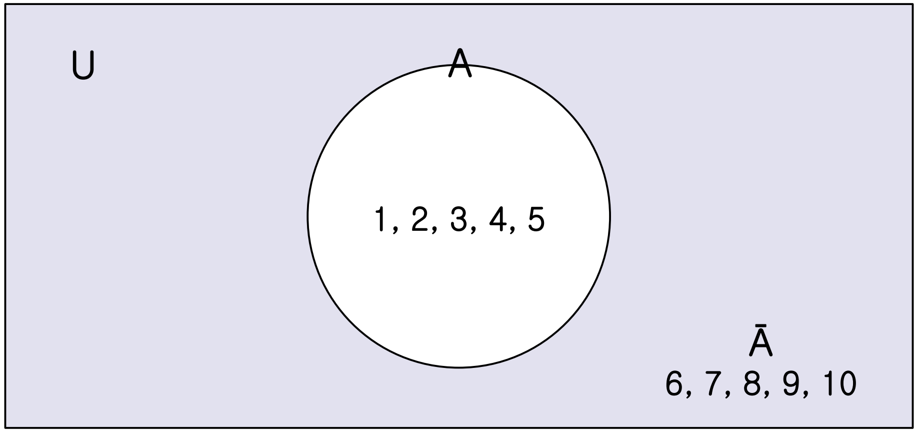
제 2절 집합이론과 확률의 개념

여집합

- 전체집합과 관심이 되는 부분집합 A가 정의되면 전체집합 중에서 집합 A에 포함되지 않는 부분이 생기는데, 이를 집합 A의 여집합이라 하여 $\bar{A}$로 나타냄

- 여집합 $\bar{A}$ = {전체집합 중에서 집합 A에 포함되지 않는 원소들}

[그림 4-1] A와 $\bar{A}$의 벤 다이어그램



$U = \{1, 2, \cdots, 10\}$
$A = \{1, 2, 3, 4, 5\}$
$\bar{A} = \{6, 7, 8, 9, 10\}$

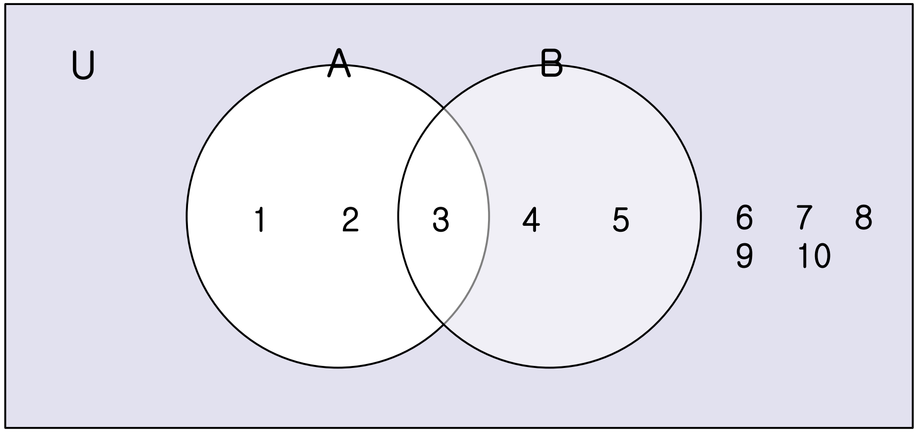
제 2절 집합이론과 확률의 개념

합집합

- 두 개의 집합 A와 B의 합집합은 집합 A나 B 중에서 하나의 집합에 속하거나, 두 집합 모두에 속하는 원소들로 구성된 집합으로 $A \cup B$로 나타냄

- 합집합 $A \cup B$ = {집합 A 또는 집합 B에 속하는 원소}

[그림 4-2] $A \cup B$의 벤 다이어그램



$U = \{1, 2, \cdots, 10\}$
$A = \{1, 2, 3\}$
$B = \{3, 4, 5\}$
$A \cup B = \{1, 2, 3, 4, 5\}$

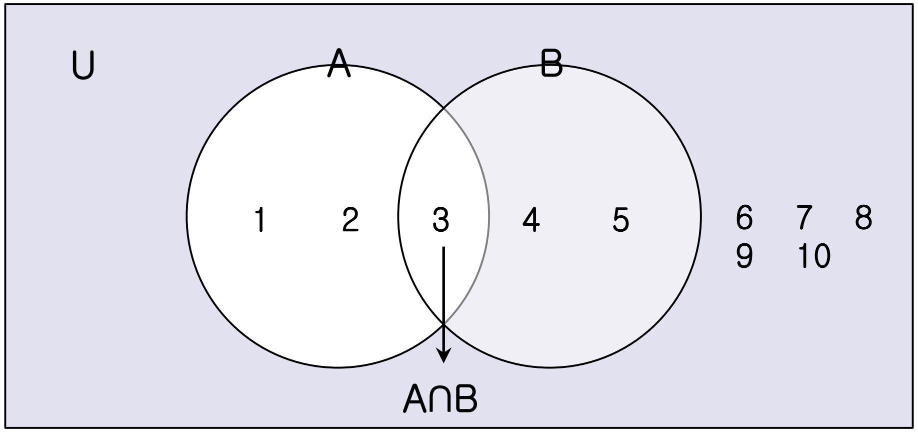
제 2절 집합이론과 확률의 개념

교집합

- 두 개의 집합 A와 B의 교집합은 A∩B로 표시되는데, 집합 A와 B에 공통적으로 속해 있는 원소만을 포함하는 집합

- 교집합 A∩B = {집합 A와 B의 공통원소}

[그림 4-3] A∩B의 벤 다이어그램



U = {1, 2, ⋯, 10}
A = {1, 2, 3}
B = {3, 4, 5}
A∩B = {3}

제 2절 집합이론과 확률의 개념

합집합의 계산

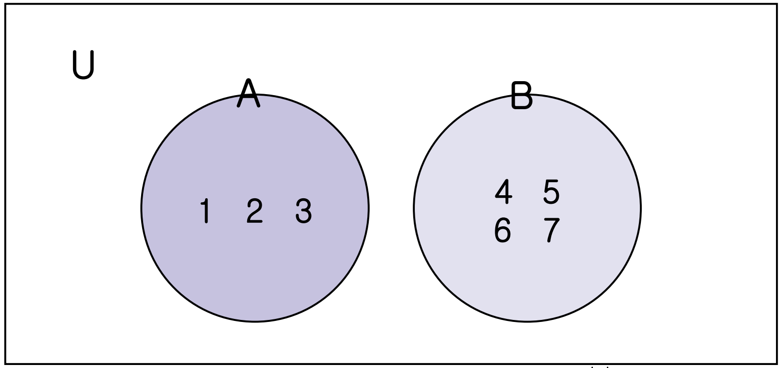
- 합집합 A∪B의 계산을 그림 4-2에서 살펴보면, 다음과 같은 식을 얻을 수 있음

$$A \cup B = A + B - A \cap B \qquad (4 \cdot 3)$$

- 집합 A와 B에서 두 집합의 공통원소가 없는 경우는 두 집합이 서로 배타적(mutually exclusive)이라고 하며, 서로 배타적인 집합의 합집합은 다음과 같이 표현

$$A \cup B = A + B \qquad (4 \cdot 4)$$

[그림 4-4] 배타적 집합의 벤 다이어그램



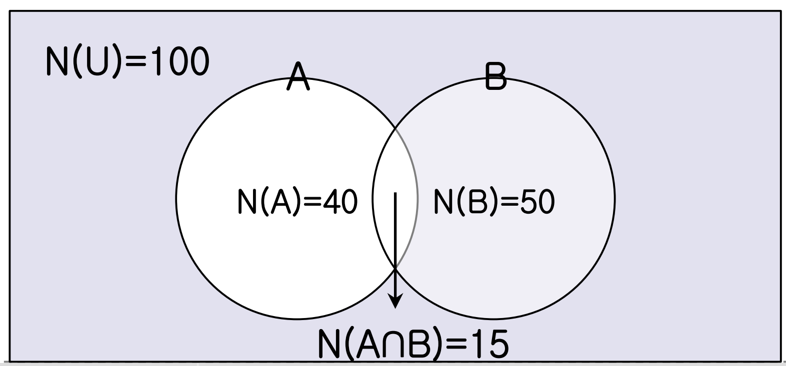
제 2절 집합이론과 확률의 개념

- 예제1
  - 100명의 학생이 시험을 본 결과, 영어시험에서 60점 이상 받은 학생은 40명이었으며, 국어시험에서는 60점 이상 받은 학생이 50명이었다. 영어시험과 국어시험 모두 60점 이상 받은 학생이 15명이라면, 최소한 한 과목에서 60점 이상을 받은 학생은 몇 명인가?

  – A를 영어시험 60점 이상, B를 국어시험 60점 이상 받은 학생의 집합이라 하면, $N(A) = 40$, $N(B) = 50$, 그리고 $N(A \cap B) = 15$

  – $A \cup B = A + B - A \cap B$이므로, $N(A \cup B) = N(A) + N(B) - N(A \cap B)$, 따라서 $N(A \cup B) = 40 + 50 - 15 = 75$

[그림 4-5] 국어시험과 영어시험에 대한 벤 다이어그램



제 2절 집합이론과 확률의 개념

- 집합이론과 확률이론
  - 집합이론과 확률이론에서 이용되는 용어들 중에는 동일한 개념이 다르게 표현되고 있는 경우가 있음
    - 실험이나 관찰에서 얻은 결과를 가리켜 확률이론에서는 사건 또는 사상이라고 하는데, 이는 집합이론에서 집합(set)의 개념과 같음
    - 확률이론에서는 여러 개의 사건 가운데 한 개의 사건을 가리킬 때 단일사건(simple event) 또는 경우라 하는데, 이는 집합이론에서의 원소(element)에 해당되는 개념
    - 또한 확률이론에서는 실험이나 관찰에서 생길 수 있는 모든 사건의 모임을 가리켜 표본공간(sample space)이라고 하는데, 이는 집합이론에서의 전체집합(universal set)과 대응되는 개념

제 3절 기본적인 확률법칙

확률의 덧셈법칙

- 확률의 덧셈법칙(addition law)은 집합이론에서 합집합의 계산에 대응되는 개념으로서, 다음과 같음

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (4 \cdot 5)$$

  - 식 4・5는 확률의 덧셈법칙의 일반식이나, A사건과 B사건이 서로 배타적 사건(mutually exclusive events)일 때는 $P(A \cap B)=0$이 되므로 다음과 같이 표현

$$P(A \cup B) = P(A) + P(B) \qquad (4 \cdot 6)$$

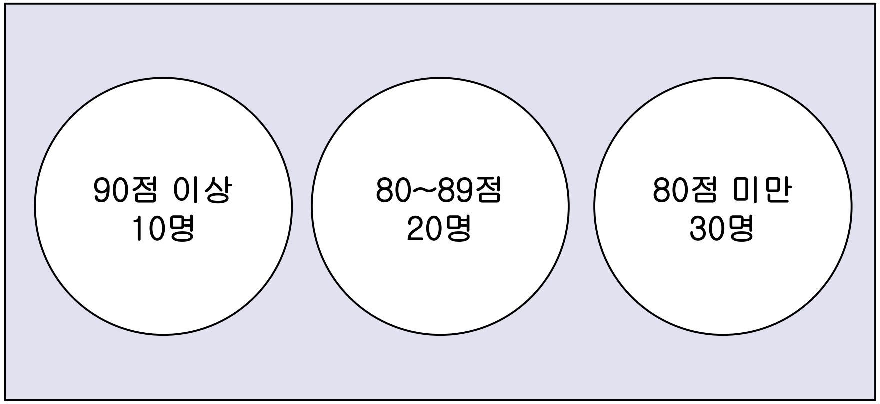
- 배타적 사건의 예
  - 어느 초등학교 6학년 1반의 국어점수를 알아보았더니 60명 중에서 10명이 90점 이상, 20명이 80점 이상 90점 미만, 30명이 80점 미만이었다. 한 학생을 무작위로 뽑을 때 90점 이상이거나 80점 미만의 학생일 확률은 얼마인가?

제 3절 기본적인 확률법칙

– 한 학생의 성적이 90점 이상이면서 동시에 80점 미만일 수는 없기 때문에 이 사건은 배타적 사건 임, 따라서 구하고자 하는 확률은 다음과 같음

P(90점 이상) = 10/60
P(80점 미만) = 30/60이므로,
P(90점 이상∪80점 미만) = P(90점 이상) + P(80점 미만)
10/60 + 30/60 = 40/60 = 0.67

[그림 4-6] 국어시험의 분포

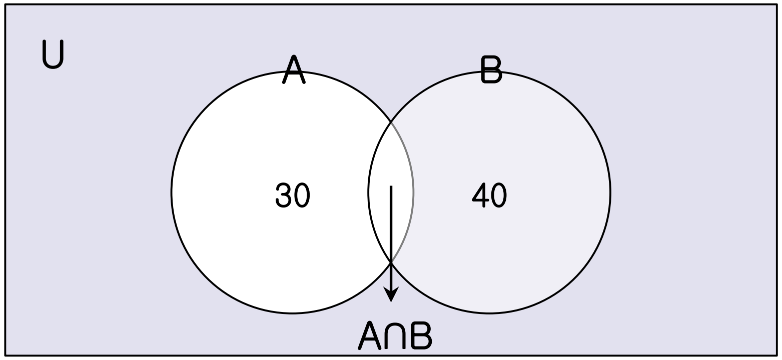


제 3절 기본적인 확률법칙

- 서로 배타적이 아닌 예
  - 통계학 강좌 수강생 60명의 필기도구를 조사하여 본 겨로가 연필을 가지고 있는 학생은 30명이었고, 볼펜을 가지고 있는 학생은 40명이었다. 임의로 한 학생을 뽑아서 조사했을 때, 볼펜 또는 연필을 가지고 있을 확률은 얼마나 될까?

  – 여기에서는 볼펜을 가진 학생이 뽑히는 사건과 연필을 가진 학생이 뽑히는 사건이 서로 배타적이 아님. 어떤 학생이 볼펜을 가지고 있다는 사실이 연필을 가지고 있지 않다는 것을 뜻하지는 않기 때문

[그림 4-7] 연필과 볼펜을 가진 학생



제 3절 기본적인 확률법칙

– 그림 4-7에서와 같이 연필을 가진 학생이 뽑힐 확률을 P(A), 볼펜을 가진 학생이 뽑힐 확률을 P(B)라 할 때, P(A∪B)를 알기 위하여는 볼펜과 연필을 모두 가진 학생이 뽑힐 확률 P(A∩B)를 알아야 함

– 연필과 볼펜을 둘 다 가진 학생의 수를 25명이라 한다면, P(A∪B)는 다음과 같이 계산

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= 30/60 + 40/60 - 25/60 \\
&= 45/60
\end{aligned}
$$

제 3절 기본적인 확률법칙

- 조건부확률

  - 앞에서는 한 번의 실험에서 어떤 사건이 발생할 확률에 대하여 살펴보았으나, 실제로는 두 단계 또는 그 이상의 실험단계를 고려하여 확률을 구해야 하는 경우가 많음

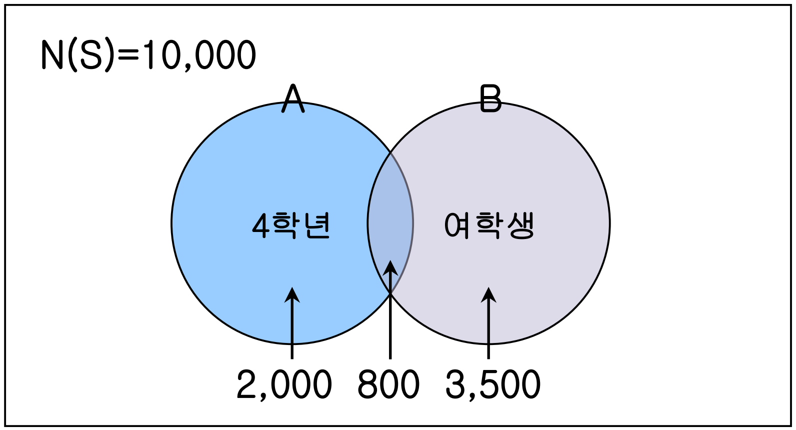
  - 조건부확률(conditional probability)이란 어떤 사건이 일어난 또는 일어날 조건하에서, 즉 변화된 표본공간에서 어떤 사건이 일어날 확률

  - 예제2

    - 잠실초등학교의 전체학생 10,000명 중에서 여학생은 3,500명이다. 2,000명이 4학년이며, 이 중 여학생은 800명이다. 전체학생을 S, 여학생을 B, 그리고 4학년 학생을 A라 하자. 그러면 이 때 $A \cap B$는 여학생이며 4학년인 학생을 나타낸다. 이를 그림 4-8을 이용해서 설명해 보자.

제 3절 기본적인 확률법칙

[그림 4-8] 조건부확률의 예



- 사건 B가 발생했다는 조건하에서 사건 A가 발생할 확률은 P(A|B)로 쓰며, "사건 B가 발생할 조건하에서 사건 A가 발생할 확률"이라고 읽음

- 뽑힌 학생이 4학년이란 조건하에서 그 학생이 여학생일 확률은 P(여학생|4학년)으로 표시하며, 다음과 같이 계산

$$P(\text{여학생} \mid 4\text{학년}) = \frac{\text{여학생} \cap 4\text{학년}}{4\text{학년}} = \frac{800}{2{,}000} = 0.40$$

제 3절 기본적인 확률법칙

- 또는 표본공간 S로 나누어서

$$P(여학생|4학년) = \frac{여학생 \cap 4학년}{4학년} = \frac{(여학생 \cap 4학년)/S}{(4학년)/S}$$

$$= \frac{P(여학생 \cap 4학년)}{P(4학년)} = \frac{800/10{,}000}{2{,}000/10{,}000} = 0.40$$

- 이처럼 조건부확률이란 그림 4-8의 벤 다이어그램에서 보듯이 S를 표본공간으로 보는 것이 아니라, 뽑힌 학생이 4학년일 사건 A를 표본공간으로 간주하고, 이 중에서 A∩B의 구성비율을 구하는 개념으로서 식 4・7과 같이 간단히 표시할 수 있음

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad (4 \cdot 7)$$

- 따라서 조건부확률은 표본공간이 줄어들었기 때문에 독립확률보다 큼
- 앞의 예에서 P(A)는 전체에서 4학년이 차지하는 비율로서 P(A) = 0.2며, P(B)는 전체에서 여학생이 차지하는 비율이므로 P(B) = 0.35고, P(A∩B) = 0.08
- 그러므로 $P(B|A) = \frac{0.08}{0.2} = 0.4$ 즉, 4학년을 한 사람 선택했을 때 그 학생이 여학생일 가능성은 0.4

제 3절 기본적인 확률법칙

확률의 곱셈법칙

- 확률의 곱셈법칙(multiplication law)은 집합이론에서 교집합의 개념에 대응되는 확률을 말함
- 교집합의 확률은 조건부확률의 개념을 이용하면 쉽게 이해할 수 있음

$$P(A \cap B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A) \quad (4 \cdot 8)$$

  - 식 4・8은 사건 A와 B가 동시에 일어날 확률은 사건 A가 일어날 확률과 사건 A가 일어난 다음 사건 B가 일어날 확률을 곱한 것과 같다는 의미

- 예제3
  - 잠원초등학교의 총 학생수는 10,000명이다. 그 중 4학년은 2,000명이며, 여학생은 3,500명이다. 여학생 중에서 4학년인 학생은 8/35이다. 그러면 어느 학생을 무작위로 선택했을 때, 그 학생이 4학년 여학생일 확률은 얼마인가?

제 3절 기본적인 확률법칙

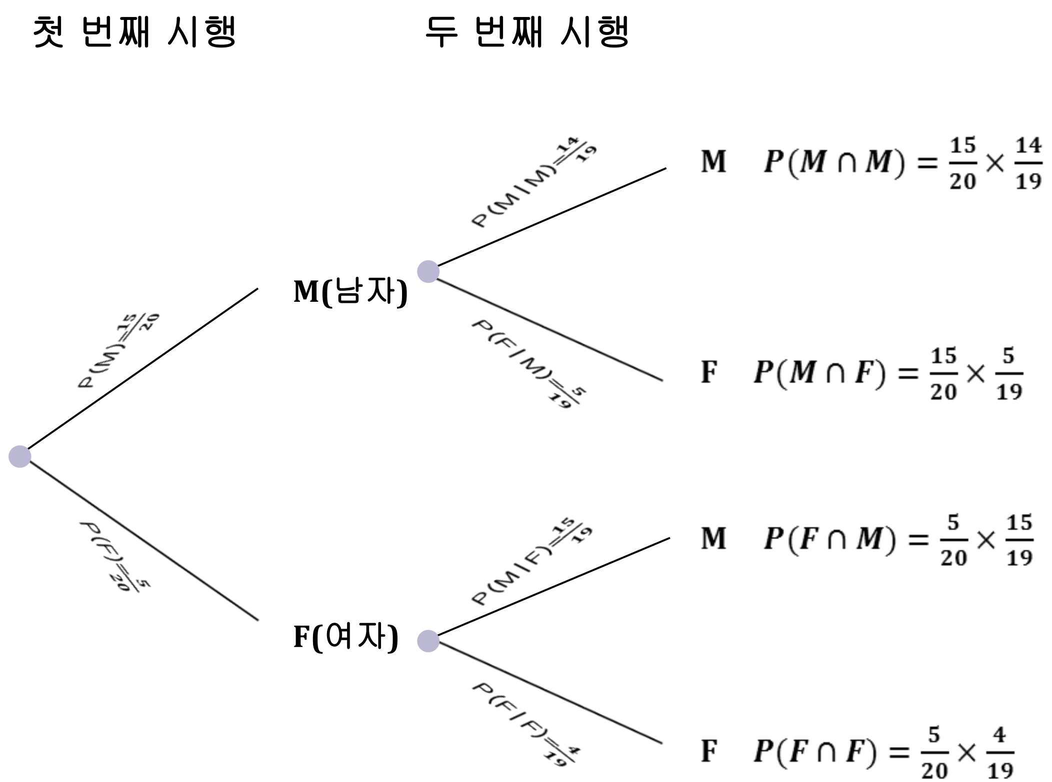
$$P(4학년 \cap 여학생) = P(여학생) \cdot P(4학년 | 여학생)$$
$$= 3{,}500/10{,}000 \times 8/35 = 0.08$$

- 곱셈법칙과 의사결정수
  - 의사결정수(decision tree)는 그림 4-9에서 보는 바와 같이 여러 단계를 거치는 확률실험에서 나타날 수 있는 모든 가능한 결과를 그린 것
  - 이를 이용하면 복잡한 확률계산문제를 쉽게 계산할 수 있음
  - 예를 들어, 남자 15명과 여자 5명으로 구성되어 있는 반에서 두 사람을 차례로 뽑으려고 할 때, 처음에 남자가 뽑히고 다음에 여자가 뽑힐 확률은 얼마인가?

$$P(M \cap F) = P(M) \cdot P(F | M)$$
$$= 15/20 \times 5/19 = 15/76$$

제 3절 기본적인 확률법칙

[그림 4-9] 의사결정수 이용의 예



제 3절 기본적인 확률법칙

- 독립사건과 종속사건
  - 처음에 어떤 결과가 나왔느냐 하는 것이 다음에 어떤 사건이 발생할 확률에 아무 영향을 주지 않을 때 이 두 사건을 독립사건(independent event)이라고 함
  - 반대로 조건부확률처럼 한 사건의 발생이 다음에 발생할 사건에 영향을 주는 경우를 종속사건(dependent event)이라고 함
  - 독립사건의 정의

$$P(A|B) = P(A)$$
$$P(B|A) = P(B) \qquad (4 \cdot 9)$$

- 식 4・9의 의미는 사건 A가 나올 확률은 사건 B의 결과에 관계없이 언제나 같다는 것

제 3절 기본적인 확률법칙

- 종속사건과 독립사건의 곱셈법칙
  - 종속사건의 곱셈법칙

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \qquad (4 \cdot 10)$$

  - 독립사건의 곱셈법칙

$$P(A \cap B) = P(A) \cdot P(B) \qquad (4 \cdot 10)$$

  - 단순히 두 번의 시행에서 뿐만 아니라, 그 이상의 시행에서도 각 사건이 독립적이라면 다음의 곱셈법칙이 성립

$$P(A \cap B \cap C \cap \cdots) = P(A) \cdot P(B) \cdot P(C) \cdots$$

제 3절 기본적인 확률법칙

- 복원추출과 비복원추출
  - 같은 실험이라고 해도 표본선택 후 복원하느냐 그렇지 않느냐에 따라 매 번의 시행에서 나타나는 결과는 독립적인 사건이 될 수도 있으며, 종속적인 사건이 될 수도 있음
  - 표본을 뽑은 후 다시 복원시키면 표본공간은 언제나 변화가 없으므로 이런 경우에 나타나는 결과는 서로 독립적
  - 흰 공이 4개, 붉은 공이 6개가 들어 있는 주머니에서 공을 한 개씩 두 번 꺼낼 때

    ① 처음 꺼낸 공을 다시 집어넣지 않는 경우(비복원추출) – 종속사건

    – 처음 흰 공을 꺼내는 사건을 A, 두 번째 흰 공을 꺼내는 사건을 B라 하면, 두 번 모두 흰 공일 확률은?

$$P(B|A)=\frac{3}{9} \qquad \therefore P(A\cap B)=P(A)\cdot P(B|A)=\frac{4}{10}\times\frac{3}{9}=\frac{2}{15}$$

② 처음 꺼낸 공을 다시 집어넣는 경우(복원추출) – 독립사건

– 처음 흰 공을 꺼내는 사건을 A, 두 번째 흰 공을 꺼내는 사건을 B라 하면, 두 번 모두 흰 공일 확률은?

$$P(B) = P(B|A) = \frac{4}{10} \qquad \therefore P(A \cap B) = P(A) \cdot P(B) = \frac{4}{10} \times \frac{4}{10} = \frac{4}{25}$$

제 4절 베이즈정리

베이즈정리

- 베이즈정리(Bayes' theorem)란 사전에 알고 있는 정보에 기준을 두고 어떤 사건이 일어나게 될 확률을 계산하는 이론

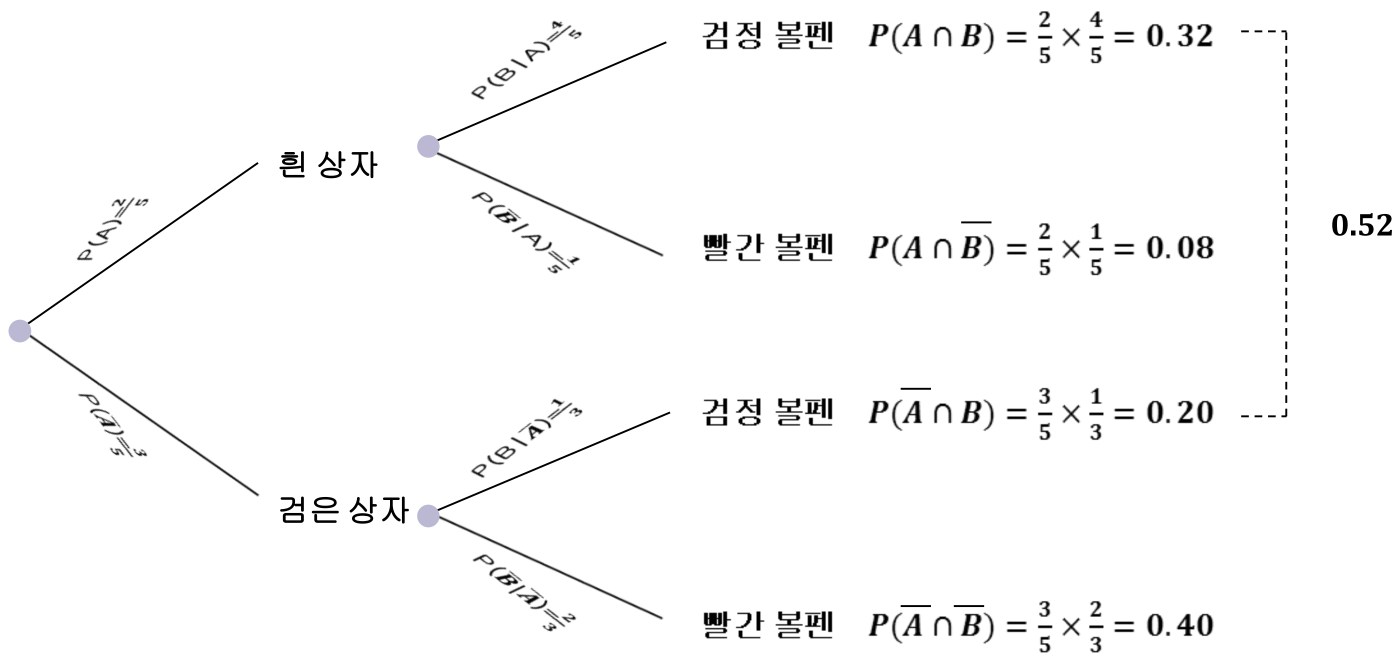
- 예를 들어,

  - 상자가 5개 있다. 두 개는 흰 상자, 세 개는 검은 상자다. 흰 상자에는 빨간 볼펜 1개, 검정 볼펜 4개, 검은 상자에는 빨간 볼펜 2개, 검정 볼펜 1개씩 들어있다. 어느 사람이 무작위로 1개의 볼펜을 뽑았을 때 이 볼펜이 검정 볼펜이었다면, 이 사람이 흰 상자를 택했을 확률은 얼마인가?

    – 의사결정수를 그리기 위해서 흰 상자를 선택하는 사건을 A, 검은 상자를 선택하는 사건을 $\overline{A}$, 검정 볼펜을 선택하는 사건을 B, 빨간 볼펜을 선택하는 사건을 $\overline{B}$라고 정의

제 4절 베이즈정리

[그림 4-10] 빨간 볼펜과 검정 볼펜을 선택할 확률



제 4절 베이즈정리

- 검정 볼펜이 나올 경우는 A∩B거나 $\bar{A}$∩B가 되므로, 검정 볼펜이 나올 확률은 P(A∩B)와 P($\bar{A}$∩B)를 더한 것

- 검정 볼펜이 흰 상자에서 나올 확률은?

  – 검정 볼펜이 흰 상자에서 나왔을 확률은 P(A|B)로 표시할 수 있는데, P(A|B)는 검정 볼펜이 나올 전체의 확률에서 흰 상자로부터 검정 볼펜이 뽑혔을 확률의 비율을 의미

$$P(A|B)=\frac{P(A\cap B)}{P(A\cap B)+P(\bar{A}\cap B)}$$

- 이상의 논리를 일반화 시켜보면

  - 사건 B가 일어났을 때, 사건 B를 일으킬 수 있는 사건들이 n가지가 있고, 이를 A_i(i=1,⋯,n)으로 나타내면, 사건 B가 발생할 수 있는 모든 확률 P(B)는 다음과 같음

$$P(B)=P(A_1\cap B)+P(A_2\cap B)+\cdots\cdots+P(A_k\cap B)$$

제 4절 베이즈정리

따라서 B가 발생하였을 때, 이것이 A_k에서 나왔을 가능성은 다음과 같음

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k \cap B)}{\sum P(A_i \cap B)}$$

- 베이즈정리(Bayes' theorem)

$$P(A_k|B) = \frac{P(A_k \cap B)}{\sum P(A_i \cap B)} \qquad (4 \cdot 11)$$

그런데 곱셈법칙으로부터 $P(A_i \cap B) = P(A_i) \cdot P(B|A_i)$ 이므로, 식 4・11은 아래와 같이 쓸 수 있음

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum P(A_i) \cdot P(B|A_i)} \qquad (4 \cdot 12)$$

제 4절 베이즈정리

- 예제 4

  - 비가 많이 올 확률 P(A) = 0.4, 비가 중간 정도 올 확률 P(B) = 0.3, 비가 아주 적게 올 확률 P(C) = 0.3이고, 비가 많이 올 때 풍년이 될 확률 P(K|A) = 0.6, 흉년이 될 확률 $P(\overline{K}|A)$ = 0.4, 비가 중간 정도 올 때 풍년이 될 확률 P(K|B) = 0.5, 흉년이 될 확률 $P(\overline{K}|B)$ = 0.5, 비가 아주 적게 왔을 때 풍년이 될 확률 P(K|C) = 0.2, 흉년이 될 확률 $P(\overline{K}|C)$ = 0.8이라 하자. 풍년이 됐는데 비가 아주 적게 왔을 가능성은 얼마인가?

  – 풍년이 됐는데, 비가 아주 적게 왔을 확률은 P(C|K)로 나타낼 수 있음. P(C|K)를 구하기 위하여 베이즈정리를 이용하면

$$P(C|K)=\frac{P(C\cap K)}{P(A\cap K)+P(B\cap K)+P(C\cap K)}$$

$$=\frac{P(C)\cdot P(K|C)}{P(A)\cdot P(K|A)+P(B)\cdot P(K|B)+P(C)\cdot P(K|C)}$$

$$=\frac{0.3\times 0.2}{0.4\times 0.6+0.3\times 0.5+0.3\times 0.2}=\frac{0.06}{0.45}=0.133$$