□ 조건부확률

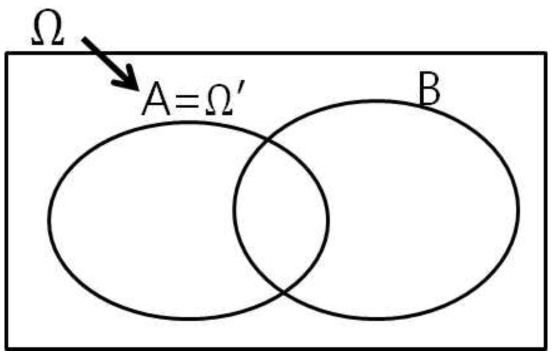
◉ 동전 두 개를 던질 때 두 동전 모두 앞면일 사건의 확률은?

- $\Omega = \{HH, \ TH, \ HT, \ TT\}$
- $P(\{HH\}) = 1/4$

● 추가정보: 어떤 한 동전이 앞면이라는 것을 알았을 때, 두 동전 모두 앞면일 사건의 확률은?

- 표본공간 ⇒ $\{HH, TH, HT\}$로 축소
- $P(\{HH\}) = 1/3$

- **조건부 확률(conditional probability)**: 확률실험에서 새로운 정보 또는 조건(A)이 추가되었을 때, 사건 B의 확률



○ 사건 A가 발생했다면 A 이외의 것은 일어날 수 없음

⇨ A가 새로운 표본공간 Ω'이 되고, B가 발생한다는 것은 $A \cap B$에 있는 원소가 발생하는 것을 의미

- 사건 A가 주어졌을 때 사건 B의 조건부 확률

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0$$

○ 사망률(mortality rate)

- 어느 해의 40대 사망률: 그해 **40대 이상인 사람들 중**에서 40대에 사망한 사람의 비율
  - 표본공간이 전체 연령대에서 40대 이상으로 축소

- 생존율(survival rate): 40대 이상인 사람 중 그 해 생존한 사람의 비율
  - 생존율 = 1-사망률

◉ 완전생명표

- 통계청에서 발표한 2012년 자료의 일부분
- 인구10만 명에서 시작하여 각 연령까지 생존한 사람의 수

연령	생존자		
	전체	남자	여자
0세	100,000	100,000	100,000
1세	99,709	99,686	99,733
40세	98,158	97,727	98,619
41세	98,048	97,581	98,546
80세	64,812	53,265	75,732
81세	61,712	49,691	72,910

○ 0세 사망자는 10만 중 100000-99709=291명

⇨ 영아 사망률 = $\frac{291}{100000} = 0.00291 = 0.29\%$

○ 40세의 남성사망률은 40세 이상 남자 생존자 97727명 중 40세에 사망한 97727-97581=146명

⇨ 40세 남성사망률 = $\frac{146}{97727} = 0.00149 = 0.15\%$

○ 80세 여성 생존율 = $\frac{72910}{75732} = 0.967 = 96.7\%$

○ 조건부확률의 활용

① $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

- ○ 곱사건이 순차적인 사건들의 조건부확률의 곱으로 표시

◉ 정상제품(정) 90개와 불량품(불) 10개가 들어있는 상자에서 무작위로 2개를 비복원으로 추출

- $\Omega = \{(정_1, 정_2), (정_1, 불_2), (불_1, 정_2), (불_1, 불_2)\}$
- $(정_1, 정_2)$의 확률은?
  - 첫 번째가 정상일 확률은 90/100
  - 두 번째**도** 정상일 확률은 89/99
  - $P(정_1, 정_2) = \underbrace{\frac{90}{100}}_{P(정_1)} \times \underbrace{\frac{89}{99}}_{P(정_2 | 정_1)} = \frac{89}{110}$

SOOKMYUNG WOMEN'S UNIVERSITY K-MOOC

○ A_1, A_2, A_3에 대해 $P(A_1 \cap A_2) > 0$이면

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\frac{P(A_1 \cap A_2)}{P(A_1)}\frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$$

$$\Rightarrow \; P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

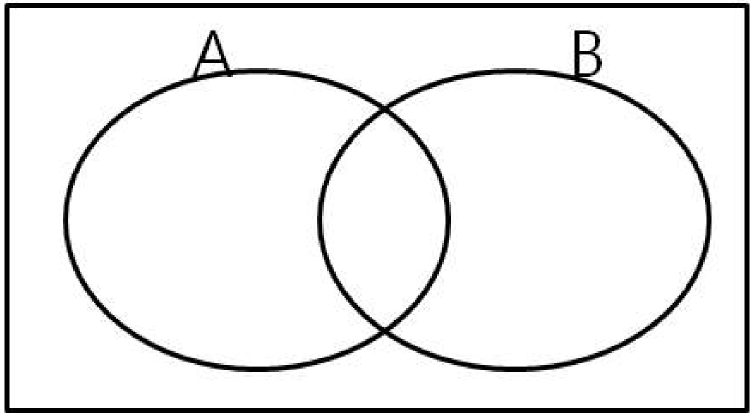
○ 수학적 귀납법을 이용하면, $P(A_1 \cap \cdots \cap A_{n-1}) > 0$

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = \; & P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots \\ & \times P(A_n|A_1 \cap \cdots \cap A_{n-1}) \end{aligned}$$

◉ 당첨복권이 4장인 복권 1000장 발매

- A_i: 구입한 4장의 복권 중에서 i번째 복권이 당첨될 사건
  - $P(A_1) = 0.004$
- $P(A_2)$?
  - $P(A_2) = P(A_1 \cap A_2) + P(A_1^c \cap A_2)$

② $$P(B) = P(A \cap B) + P(A^c \cap B)$$
$$= P(A)P(B|A) + P(A^c)P(B|A^c)$$



○ $$P(A_2) = P(A_1)P(A_2|A_1) + P(A_1^c)P(A_2|A_1^c)$$
$$= \frac{4}{1000}\frac{3}{999} + \frac{996}{1000}\frac{4}{999} = \frac{4}{1000} = 0.004$$

- 어떤 일련의 사건들이 순차적으로 결합된 경우

  ⇨ 특정 시점에서의 사건 확률은 앞에서 발생할 수 있는 상황이나 연결된 상황들의 확률을 모두 더하여 구할 수 있음

◉ 스팸메일 필터

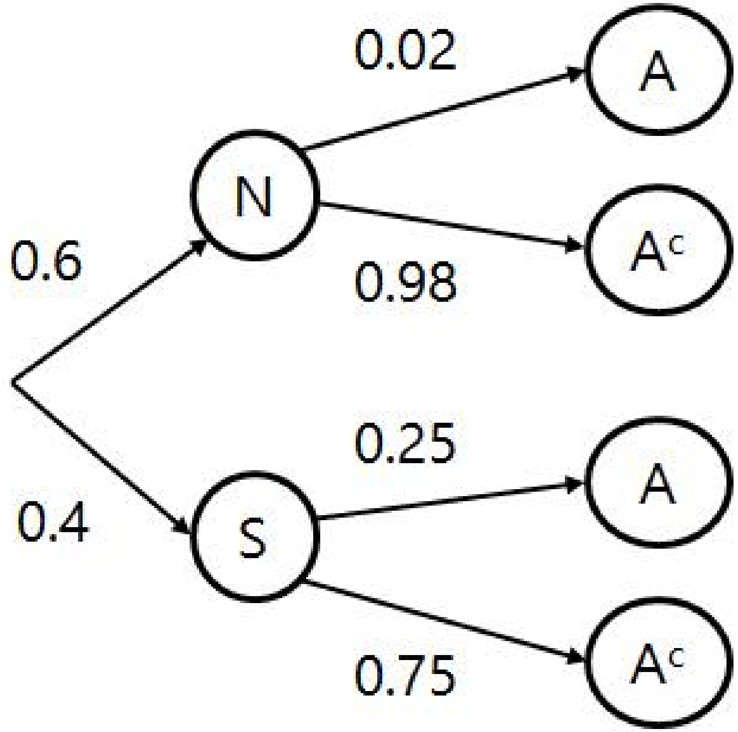
- ○ 어떤 메일시스템의 수신메일 중 40%가 스팸메일(S)이고 나머지는 정상메일(N)
  - · $P(S)=0.4$, $P(N)=0.6$
- ○ 스팸메일 중 25%는 "A"라는 단어를 포함하고 정상메일 중 2%가 이 단어를 포함
  - · $P(A|S)=0.25$, $P(A|N)=0.02$

Q. 전체 메일 중 "A" 단어를 포함한 메일의 비율은?

$$
\begin{aligned}
P(A) &= P(S \cap A) + P(N \cap A) \\
&= P(S)P(A|S) + P(N)P(A|N)
\end{aligned}
$$

$$\Rightarrow \ P(A) = 0.4 \times 0.25 + 0.6 \times 0.02 = 0.1 + 0.012 = 0.112$$

○ 확률수형도(probability tree)



- **표본공간의 분할(partition):** 사건 $A_1,\ \ldots,A_n$ 가

  Ⓐ 서로배반사건, 즉 모든 $i \neq j$에 대해 $A_i \cap A_j = \varnothing$

  Ⓑ 전체를 이루는 사건(exhaustive), 즉 $A_1 \cup \cdots \cup A_n = \Omega$

- 사건 $A_1,\ \ldots,A_n$ 이 표본공간 Ω의 분할이면

$$P(B) = P(B \cap \Omega) = P((B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n))$$

$$= P(B \cap A_1) + \cdots + P(B \cap A_n)$$

$$= P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)$$

- **정리**

  - $P(B|A) = \dfrac{P(A \cap B)}{P(A)}, \quad P(A) > 0$

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

  - $$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A)P(B|A) + P(A^c)P(B|A) \end{aligned}$$