확 률

(Probability)

3.1 기본개념

◉ 확률이 발생하는 상황에서의 공통적인 특징

① 주사위 던지기

② 앞면이 나올 때까지 동전 던지기

③ 판매한 스마트폰의 수명

④ 랜덤하게 선택된 학생의 혈액형

- 실험을 시행하기 전에 발생할 수 있는 모든 결과는 알 수 있음

  ① $\{1, 2, 3, 4, 5, 6\}$

  ② 앞면을 H, 뒷면을 T이라고 하면, $\{H, TH, TTH, \cdots\}$

  head (그림) tail (숫자)

  ③ x를 수명(단위 일)이라고 하면, $\{x \mid 0 \leq x\}$

- 실험을 하기 전까지 이들 결과 중 어떤 것이 발생할 것인지에 대해 확실하게 예측할 수 없음

- **확률실험**(random experiment), 통계적실험

  : 위의 두 성질을 만족시키는 실험

  ☞ 독립적으로 반복 수행 가정

- **표본공간**(sample space; Ω)
  : 확률실험에서 발생 가능한 모든 결과들의 집합

- **사건**(event; A, B): 표본공간 내에서 우리가 관심을 가지는 부분집합
  ① 홀수가 나오는 경우
  ② 3번 이상 던지는 경우
  ③ 365일 이전에 수명을 다 하는 경우

- 어떤 사건 A가 발생한다는 것은 실험 결과가 A에 속하는 원소 중 하나이거나 A에 포함되어 있다는 것을 의미

- **확률(probability)**: 이러한 사건이 발생할 가능성이 얼마나 되는지를 나타내는 수치적 측도 [0, 1] 사이의 실수값
  - 확률을 언급하기 위해서는 해당하는 확률실험이 전제되어야 하고 이에 따른 표본공간과 사건이 정해져야 함 → P(A)

※ 표본공간과 사건은 수학적 관점에서 보면 일종의 집합
  - 확률을 정의하고 계산하기 위해서는 집합에 대한 기본 정의와 연산을 알아야 함

【표 3.1】 집합의 정의와 연산

정의 및 법칙	표시 및 내용
• A와 B의 합사건(union)	$A \cup B = \{\omega \mid \omega \in A$ 또는 $\omega \in B\}$
• A와 B의 곱사건(intersection)	$A \cap B = \{\omega \mid \omega \in A$ 그리고 $\omega \in B\}$
• A의 여사건(complement)	$A^c = \{\omega \mid \omega \not\in A$ 그리고 $\omega \in \Omega\}$
• 교환법칙(commutative law)	$A \cup B = B \cup A,\ A \cap B = B \cap A$
• 결합법칙(associative law)	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
• 분배법칙(distributive law)	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
• 드모르간(De Morgan) 법칙	$(A \cup B)^c = A^c \cap B^c,\ (A \cap B)^c = A^c \cup B^c$
• 무한개의 사건이 존재하는 경우 $(A_1, A_2, \ldots)$	$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$

○ 서로 배반사건(disjoint, mutually exclusive)

: 두 사건 A와 B가 공통부분이 없는 경우, 즉 $A \cap B = \varnothing$

○ 벤다이어그램(Venn diagram) 이용

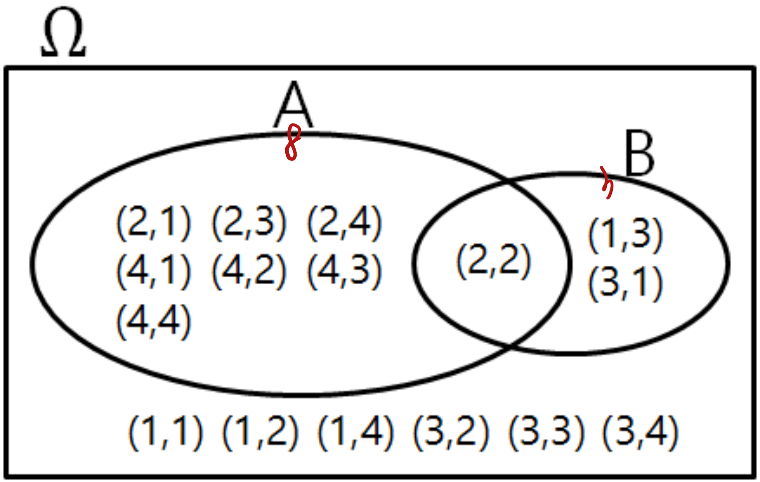
- 사건(집합)들 간의 관계를 쉽게 이해할 수 있음

[예제 3.1] 정사면체 주사위 두 개를 던지기

○ A: 첫 번째 주사위가 짝수인 사건

○ B: 두 주사위의 합이 4인 사건

$$
\begin{aligned}
\Omega &= \{(i,j) \mid i=1,2,3,4,\ \ j=1,2,3,4\} \\
A &= \{(i,j) \mid i=2,4,\quad j=1,2,3,4\} \\
B &= \{(i,j) \mid i+j=4\} = \{(1,3),\ (2,2),\ (3,1)\}
\end{aligned}
$$



○ $A^c \cap B = \{(1,3),\ (3,1)\}$

○ $A^c \cap B^c = (A \cup B)^c = \{(1,1),\ (1,2),\ (1,4),\ (3,2),\ (3,3),\ (3,4)\}$

3.2 확률의 이해

3.2.1 고전적 의미의 확률

- 17세기 중반 파스칼이나 페르마 등이 도박 문제에 대해 의견 교환
  - 카드 게임에서 어떤 패가 더 좋은 것인지 결정하기 위해 각 패의 발생할 수 있는 빈도를 계산
- 표본공간에서 사건에 해당하는 원소가 차지하는 비율
- 표본공간이 n개 원소로 이루어져 있고 **각 근원사건의 발생가능성이 동일(equally likely)**한 경우, k개의 원소를 가지는 사건 A의 확률

$$P(A) = \frac{k}{n}$$

[예제 3.2] 정사면체 주사위 2개 → Ω 16개

A 8개

○ 각 눈이 나올 가능성은 동일

$$P(A)=\frac{8}{16}=\frac{1}{2},\quad P(B)=\frac{3}{16}$$

$$P(A^c\cap B)=\frac{2}{16}=\frac{1}{8},\quad P(A^c\cap B^c)=\frac{6}{16}=\frac{3}{8}$$

○ 경우의 수(number of cases)

○ 확률을 계산하기 위해서는 표본공간과 사건에 있는 원소의 개수를 효율적으로 계산하는 것이 중요

○ 경우의 수의 기본 법칙: **곱의 법칙(multiplication rule)**

어떤 실험이 m 개의 연속된 단계로 이루어져 있고 i -번째 단계에서 발생 가능한 결과의 수가 n_i 개이면 전체 실험에서 발생 가능한 경우의 수는

$$n = n_1 \times n_2 \times \cdots \times n_m$$

[예제 3.3] 세트메뉴의 경우의 수

- 세트메뉴에는 4가지 음료수, 2가지 샐러드, 5가지 메인 요리, 4가지의 디저트 중에서 각각 하나씩을 선택
- 선택할 수 있는 세트의 종류는 $4 \times 2 \times 5 \times 4 = 160$

○ 경우의 수의 일반적인 문제

1부터 n까지 적힌 공이 들어 있는 주머니에서 k개를 무작위로 선택

◦ k개를 어떻게 추출하고 나열할 것인지에 따라 달라짐

◦ 추출방법:

- 복원(with replacement)추출
- 비복원(without replacement)추출

Q: 만약 한꺼번에 k개의 공을 뽑으면?

◦ 뽑힌 순서를 고려 여부

- 순서 고려하는 경우: (1, 2)와 (2, 1)을 다른 것으로 봄
- 순서 고려하지 않는 경우: (1, 2)와 (2, 1)을 같은 것으로 봄

Q: 뽑은 것을 크기 순서대로 정렬(sorting)한다고 하면?

배열 \ 추출	복원	비복원
순서고려	Ⓐ	Ⓑ
순서무시	~~Ⓒ~~	Ⓓ

Ⓐ **중복순열**

Ⓑ **순열(permutation)**

Ⓓ **조합(combination)**

~~Ⓒ 중복조합~~

거의사용X

Ⓐ **중복순열**의 수는 매번 n개 선택 가능

$$n \times \cdots \times n = n^k$$

Ⓑ **순열**은 각 단계에서 선택할 수 있는 수가 하나씩 줄어듦

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$$

Ⓓ 순서를 고려하지 않는다면 같은 번호로 이루어진 순열이 같은 것으로 취급 ⇨ 선택된 k개의 공의 순서열

$$k \times (k-1) \times \cdots \times 1 = k!$$

순서를 고려하지 않는 **조합의 수**는

$$\frac{n \times (n-1) \times \cdots \times (n-k+1)}{1 \times 2 \times \cdots \times k} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$

【표 3.2】 경우의 수

배열 \ 추출	복원	비복원
순서고려	n^k	$\dfrac{n!}{(n-k)!}$
순서무시		$\dbinom{n}{k}$

[예제 3.4] 나눔Lotto 6/45

- 1에서 45까지의 숫자에서 6개의 번호를 비복원 추출
- 1등은 선택한 6개의 번호가 당첨 번호와 모두 일치
- 5개만 일치하면 2등 또는 3등
- 전체 가능한 경우의 수

$$\#(\Omega) = \binom{45}{6} = \frac{45 \times 44 \times 43 \times 42 \times 41 \times 40}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 8{,}145{,}060$$

⇨ $P(1\text{등}) = 1/8145060$

- 2등과 3등: 6개 당첨번호 중 5개 선택, 39개 중 하나 선택

$$\#(2\text{등 또는 }3\text{등}) = \binom{6}{5} \times 39 = 6 \times 39 = 234$$

⇨ $P(2\text{등 또는 }3\text{등}) = 234/8145060 = 1/34807.95$

[예제 3.5] Birthday problem

- 1년은 365일이고, 365일 동안 태어날 가능성이 동일
- A: k명의 사람이 모두 다른 생일을 가지는 사건
- k명이 각각 365일 중 한 날을 선택할 수 있음: $\#(\Omega) = 365^{\mathrm{k}}$
- k명이 모두 생일이 다르다는 것은 각기 다른 날짜를 선택하는 것으로 365일을 k번 비복원 추출하는 방법

$$\#(A) = 365 \times 364 \times \cdots \times (365 - k + 1) = \frac{365!}{(365-k)!}$$

$$\Rightarrow P(A) = \frac{365!/(365-k)!}{365^k} = \frac{365!}{365^k (365-k)!} = \prod_{j=0}^{k-1} \left(1 - \frac{j}{365}\right)$$

k 명의 생일이 모두 다른 확률

k	5	10	15	**20**	30	40	50
$P(A)$	0.9729	0.8831	0.7471	0.5886	0.2937	0.1088	0.0296

25명부터
0.5보다 작아짐

○ 연속표본공간

- 발생 가능성이 동일한 상황을 선이나 평면 등을 이용해 표현
- 사건 A가 발생한다는 것은, 영역 Ω내에서 임의의 한 점을 무작위로 택할 때 이 점이 영역 A에 있다는 의미
- 사건 A의 확률은 전체 영역에서 A가 차지하는 비율

$$P(A) = \frac{\|A\|}{\|\Omega\|}$$

여기서 $\|\cdot\|$는 길이, 면적, 부피 등을 의미

[예제 3.6] 원형 돌림판

- 교재 p.83

$\Omega \quad 0 \le x \le 12$
$A \quad 1 \le x \le 4$
$\rightarrow \frac{3}{12}$

3.2.2 상대도수의 극한개념

- (정상적인) 동전의 앞면이 나올 확률은 1/2?
- 고전적 확률: "앞면과 뒷면의 발생가능성이 동일하고

  $\Omega = \{H, T\}$, $A = \{H\}$ 이므로 $P(A) = 1/2$ 인 것으로 해석
- 동전던지기 실험

실험자	던진 횟수	앞면	상대도수
Buffon	4040	2048	0.5080
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

▶ 실험을 계속 수행하면 **상대도수가 0.5로 수렴**

Q: 동전이 찌그러진 경우는 ? 1/2이라고 하기 어렵다?

- 윷을 하나 던졌을 때 평면이 위로 나타날 확률은?
  - 윷을 n번 던졌을 때 평면이 나온 횟수를 $n(A)$라면

$$P(A) \simeq \frac{n(A)}{n}$$

  - 만약 실험을 무한히 반복하면 $n(A)/n$은 어떤 값으로 수렴하고 이 극한값을 사건 A가 일어날 확률로 정의

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

▶ "**그냥 도전 동전 돌리기**"에서 각각 실험에서 발생하는 결과는 표본

※ 실험을 무한히 반복 ⇒ "모집단 확률(분포)"로 해석할 수 있음

- 확률은 모집단이 어떤 형태로 이루어져(분포되어) 있는지를 보여줌
- 많은 표본을 통해 모집단의 특성을 파악하는 때문에 이런 방식으로 확률을 정의하는 경우 **통계적 확률**(statistical probability)이라고 함

※ 일기예보나 전쟁게임(war game)

- ○ 모의실험(simulation)을 통해 결과를 도출
- ○ 고속 컴퓨터의 보급으로 인해 더욱더 활용도가 높아짐

[예제 3.7] 생일문제

○ 365일 각각의 날에 태어날 가능성이 동일?X

- 2013년 월별 하루 평균 출생아수 ("월간인구동향" 통계청. p.37))

상황	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
Ⓐ	1427	1309	1243	1225	1149	1105	1167	1174	1235	1163	1128	1031
Ⓑ	2000	2000	2000	2000	2000	2000	1000	1000	1000	1000	1000	1000

Ⓐ 2013년에 태어난 436,455명의 신생아 월별 하루 평균출생아수

Ⓑ 1월에서 6월까지 일별 출생아가 7월에서 12월까지 일별 출생아의 2배라고 가정

표본의크기 (관측개수) 증가 → 확률값수렴

▶ Ⓐ와 Ⓑ를 만족한다는 가정으로 **모의실험**(simulation) → 극한개념

- 각각의 인원 k에 대해 일억 번 실시한 모의실험에서 비율

☞ **"통계적 확률"**

【표 3.3】 상황에 따른 생일문제 (생일 모두 다를 확률)

k	5	10	15	20	30	40	50
$P(A)$	**0.9729**	**0.8831**	**0.7471**	**0.5886**	**0.2937**	**0.1088**	**0.0296**
Ⓐ	0.9727	0.8823	0.7456	0.5865	0.2913	0.1072	0.0290
Ⓑ	0.9699	0.8709	0.7230	0.5551	0.2569	0.0853	0.0202

$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$ 으로

$\frac{\text{빈도}}{\text{전체개수}}$ = 상대빈도 $\xrightarrow{n \to \infty}$ 확률값수렴

3.2.3 공리적 확률

확률의 상식적 법칙을 수학적으로 정리

- 콜모고로프(A. N. Kolmogorov, 1903-1987)
- 확률의 공리는 확률이론의 기반
- $P(\cdot)$: 확률측도(probability measure)

그냥 받아들이기.

[공리 1] $P(\Omega)=1$ (Ω는 표본공간)

[공리 2] $0 \le P(A) \le 1$, $A \subset \Omega$

[공리 3] 서로배반인 사건 $A_1, A_2, \ldots, A_n$에 대해,

$A_i \cap A_j = \phi$ $(i \neq j)$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

■ 확률의 기본정리

- 공리로부터 얻을 수 있는 확률의 기본적 성질

① $P(A^c) = 1 - P(A)$

▶ $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$

○ $P(\varnothing) = 0$ $\Omega^c = \phi$

○ 생일문제: k 명 중 적어도 두 사람 이상이 같은 생일을 가지는 사건은 A^c

$$P(A^c) = 1 - P(A) = 1 - \frac{365!}{365^k (365-k)!} = 1 - \prod_{j=0}^{k-1}\left(1 - \frac{j}{365}\right)$$

[예제 3.8] 1,000장 중 4장이 당첨복권

- 4장의 복권 구입한 경우, 적어도 한 장 이상 당첨복권 구입 확률은?
- A: 한 장 이상의 당첨복권을 구입할 사건

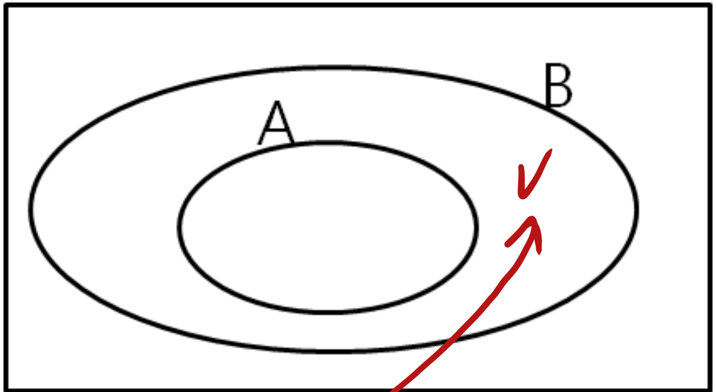
  = 당첨복권이 한 장, 두 장, 세 장, 네 장인 경우

⇨ A^c: 구입한 4장 모두 당첨되지 않을 사건

$$P(A^c) = \frac{996 \times 995 \times 994 \times 993}{1000 \times 999 \times 998 \times 997} = 0.9841$$

⇨ $P(A) = 1 - P(A^c) = 0.0159$

② $A \subset B$이면 $P(A) \leq P(B)$



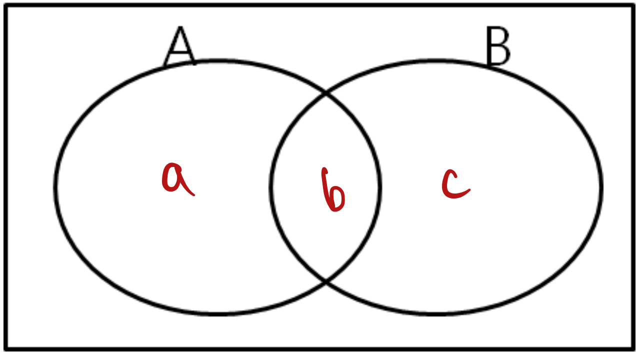
○ $B = A \cup (\boxed{B \cap A^c})$

$P(B) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0}$ $(\because$ 공리2$)$

따라서 $P(B) \geq P(A)$

③ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

모두배반사건으로 쪼개서 생각하기



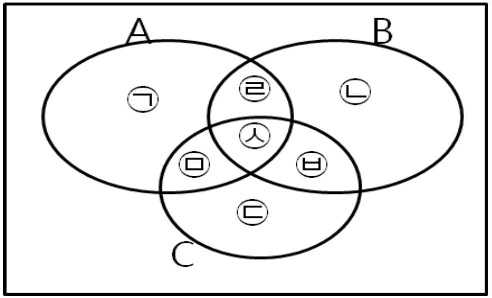
$$A = (A \cap B) \cup (A \cap B^c)$$

$$\Rightarrow \; P(A) = P(A \cap B) + P(A \cap B^c)$$

$$\& \; P(B) = P(A \cap B) + P(A^c \cap B)$$

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

○ 사건 A, B, C



$$P(A\cup B\cup C) = P(㉠) + P(㉡) + P(㉢) + P(㉣) + P(㉤) + P(㉥) + P(㉦)$$

- $P(㉠) = P(A) - P(A\cap B) - P(A\cap C) + P(A\cap B\cap C)$
- $P(㉣) = P(A\cap B) - P(A\cap B\cap C)$
- $P(㉦) = P(A\cap B\cap C)$

⇨ $$P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C)$$

○ n개의 사건 $A_1, A_2, \ldots, A_n$에 대해 일반화

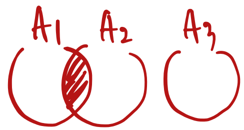
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$

$$+ \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

④ $P(A \cup B) \leq P(A) + P(B)$

○ 부울의 부등식(Boole's inequality)

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$



⇒ 겹치는부분이있으므로

○ 본페로니 부등식(Bonferroni's inequality) 드모르간법칙이용

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

$P(A_1 \cap A_2) = 1 - P((A_1 \cap A_2)^c)$

부울의부등식

$\hookrightarrow A_1^c \cup A_2^c \leq P(A_1)^c + P(A_2)^c$

$1 - P(A)$ $1 - P(B)$

[예제 3.10] 1000장 중 4장이 당첨복권 ([예제 3.8] 참조)

○ A_i : i 번째 복권이 당첨될 사건 ⇨ $P(A_i) = 0.004$

비복원일때도 i에 상관없이 확률이 같은가?

⇒ 뒤에서 다시확인

⇨ 확률계산? (복원 & 비복원)

○ $A = \bigcup_{i=1}^{4} A_i$ 로 표시 (적어도 한 장 이상 당첨복권)

$$P(A) = P\left(\bigcup_{i=1}^{4} A_i\right) = 0.0159 \le \sum_{i=1}^{4} P(A_i) = 0.016$$

20장일때 → 0.016×5 = 0.08 임을 쉽게계산가능

※ 부울의 부등식(Boole's inequality)

항상 두값사이에 큰차이가 없는것은 아니지만 어떠한 확률의 상한값을 계산할때 유용하다

3.4 조건부확률 Conditional probablity

◉ 동전 두 개를 던지는 실험에서 어떤 한 동전이 앞면이라는 것을 알았을 때, 두 동전 모두 앞면일 사건의 확률은?

- $\Omega = \{HH, \ TH, \ HT, \ TT\}$ → $\frac{1}{4}$
- 하나가 앞면이라는 추가적인 정보가 주어지면 표본공간에서 $\{TT\}$가 발생할 수 없음

⇨ 표본공간은 $\{HH, TH, HT\}$로 축소 → $\frac{1}{3}$

◉ 주사위 빨간색 $\{1, 2, 3, 4\}$, ~~검정색 $\{5, 6\}$~~

↱ 5,6은 더이상 표본공간의 element가 아님.

- 빨간색 면이 나온 것을 알았을 때 ☜ "추가적인 정보"

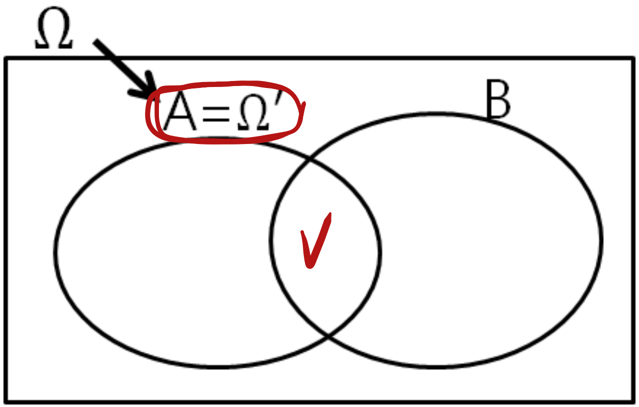
$\{1\}$일 확률? ⇨ $P(\{1\} \mid \underbrace{\{1,2,3,4\}}_{\Omega'}) = P(B \mid \underbrace{A}_{\Omega'}) = \dfrac{P(B \cap A)}{P(A)}$

(새로운 표본공간)

(+) 전체 표본공간에서 B의 확률은 $P(B \mid \Omega)$지만 생략한것

- **조건부 확률**(conditional probability):

  확률실험에서 새로운 정보 또는 조건이 추가되었을 때 사건의 확률



- 사건 A가 발생했다면 A 이외의 것은 일어날 수 없음

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

- A하에서 B의 조건부확률은 A에서 $A \cap B$가 차지하는 비율
- 사건 A가 주어졌을 때 사건 B의 조건부 확률은 $P(B|A)$로 표시

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

분모≠0

○ 사망률(mortality rate) ☞ 조건부 확률로 정의되는 사회지표

- 어느 해의 40대 사망률

  : 그 해 40대 이상인 사람들 중에서 40대에 사망한 사람의 비율

  ☞ 표본공간이 전체 연령대에서 40대 이상으로 축소

- 생존율(survival rate)

  : 40대 이상인 사람 중 그 해 생존한 사람의 비율로 (1-사망률)로 계산

cf) 전연령에 관한표를 10만명기준으로 재조정한것

[예제 3.12] 완전생명표

- 통계청에서 발표한 2012년 자료의 일부분
- 인구 10만 명에서 시작하여 각 연령까지 생존한 사람의 수

【표 3.4】 완전생명표

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
60세	92,679	89,823	95,657
61세	92,146	89,046	95,379
80세	64,812	53,265	75,732
81세	61,712	49,691	72,910

−291

0세 사망률 $\frac{291}{100,000}$

−543.

60세 사망률 $\frac{543}{92,679}$

○ 0세 사망자는 10만 중 100000-99709=291명

⇨ 영아 사망률= 291/100000=0.00291(0.29%)

○ 40세의 남성사망률은 40세 이상 남자 생존자 97,727명 중 40세에 사망한 97727-97581=146명

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

전연령 사망율 남성 > 여성

【표 3.5】 연령별 사망자와 사망률

상대도수기반 data table → 확률로의 해석 (∵극한값 개념)

연령	구분	전체	남자	여자
0세	사망자	291	314	267
	사망률	0.29%	0.31%	0.27%
20세	사망자	33	44	21
	사망률	0.03%	0.04%	0.02%
40세	사망자	110	146	73
	사망률	0.11%	0.15%	0.07%
60세	사망자	533	777	278
	사망률	0.58%	0.87%	0.29%
80세	사망자	3,100	3,574	2,822
	사망률	4.78%	6.71%	3.73%

영아사망율↑ 사고위험성, …

2만8천번

○ 80세 여성 생존율: 1-2822/75732=1-0.0373=0.967로 96.7%

□ 조건부확률의 활용

$P(B|A) = \frac{P(B\cap A)}{P(A)} \rightarrow P(B|A) \times P(A) = P(B\cap A)$

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

: 곱사건은 순차적인 사건들의 조건부확률의 곱으로 표시

[예제 3.13]

뽑은걸 다시 넣지 않음

정상 90개, 불량 10개 있는 상자에서 무작위로 2개 **비복원추출**

$$\Omega = \{(정상_1, 정상_2), (정상_1, 불량_2), (불량_1, 정상_2), (불량_1, 불량_2)\}$$ 순서를따지는경우

○ (정상1, 정상2)의 확률은? $\frac{90}{100} \times \frac{89}{99}$

- 첫 번째가 정상일 확률은 90/100
- 두 번째도 정상일 확률은 89/99

- $P(\text{정상}_1, \text{정상}_2) = \dfrac{90}{100} \times \dfrac{89}{99} = \dfrac{89}{110}$

$P(\text{정상}_1)$ ↲ ↳ $P(\text{정상}_2 | \text{정상}_1)$

즉 정상1 이 나왔을때 정상2가 나올확률

$$P(\text{정상}_1, \text{불량}_2) = P(\text{정상}_1)P(\text{불량}_2 | \text{정상}_1) = \frac{90}{100} \times \frac{10}{99} = \frac{10}{110}$$

$$P(\text{불량}_1, \text{정상}_2) = P(\text{불량}_1)P(\text{정상}_2 | \text{불량}_1) = \frac{10}{100} \times \frac{90}{99} = \frac{10}{110}$$

$$P(\text{불량}_1, \text{불량}_2) = P(\text{불량}_1)P(\text{불량}_2 | \text{불량}_1) = \frac{10}{100} \times \frac{9}{99} = \frac{1}{110}$$

▶ 사건이 n개인 경우로 일반화

○ $A_1,\ A_2,\ A_3$에 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, $P(A_1 \cap \cdots \cap A_{n-1}) > 0$

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

▶ 점심값 내기

20개 쪽지 중 당첨 쪽지를 뽑은 4명이 냄 (비복원추출!)

Q : 제비뽑기에서 몇 번째로 뽑는 것이 유리한가?

⇒ 순서에 상관없이 모두 같다

○ A_i : i 번째 당첨될 사건

○ $P(A_i) = ?$

A→당첨X, B→당첨일때

$P(B_1)\ P(B_2|B_1) \rightarrow P(B_1 \cap B_2)$

$$P(A_1) = \frac{4}{20}$$

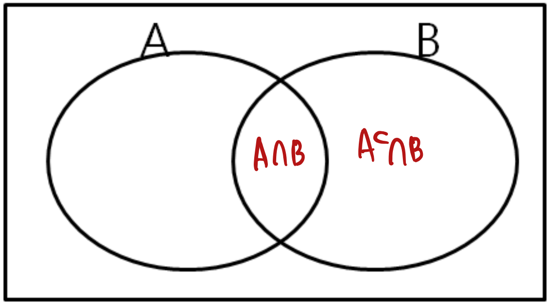
$$P(A_2) = \frac{4}{20}\frac{3}{19} + \frac{16}{20}\frac{4}{19} = \frac{4}{20}$$

$P(A_1)\ P(B_2|A_1) \rightarrow P(A_1 \cap B_2)$

$$\& \ \ P(A_3) = \frac{4}{20} \quad \ldots$$

▶ 전확률의 법칙 (rule of total probability)

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



제비뽑기 : $P(A_2) = \boxed{P(A_1)}P(A_2|A_1) + \boxed{P(A_1^c)}P(A_2|A_1^c)$
$$= \frac{4}{20}\frac{3}{19} + \frac{16}{20}\frac{4}{19} = \frac{4}{20}$$

$$
\begin{aligned}
P(A_3) = \; & P(A_1 \cap A_2 \cap A_3) + \boxed{P(A_1^c \cap A_2 \cap A_3)} \\
& + P(A_1 \cap A_2^c \cap A_3) + P(A_1^c \cap A_2^c \cap A_3) \\
= \; & P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \\
\frac{16}{20} \cdot \frac{4}{19} \cdot \frac{3}{18} \quad & + \boxed{P(A_1^c)P(A_2 | A_1^c)P(A_3 | A_1^c \cap A_2)} \\
& + P(A_1)P(A_2^c | A_1)P(A_3 | A_1 \cap A_2^c) \\
& + P(A_1^c)P(A_2^c | A_1^c)P(A_3 | A_1^c \cap A_2^c) = \frac{4}{20}
\end{aligned}
$$

- 어떤 일련의 사건들이 순차적으로 결합된 경우

  ⇨ 특정 시점에서의 사건 확률은 앞에서 발생할 수 있는 모든 상황에 대한 조건부 확률을 통해 구할 수 있음

[예제 3.15] 스팸메일 필터

$P(S)=0.4$, $P(N)=0.6$

- 메일시스템 수신메일 중 **40%가 스팸메일**(S), 나머지는 정상메일(N)
- **스팸메일** 중 내용에 "X"라는 단어가 있는 메일은 **25%**이고 $\rightarrow P(X|S)=0.25$
  정상메일 중 이 단어가 있는 경우는 **2%** $\rightarrow P(X|N)=0.02$
- 전체 메일 중 "X" 단어를 포함한 메일의 비율은?

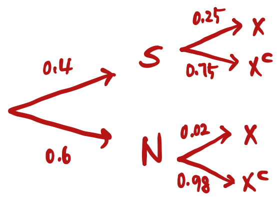
$$P(X)=P(S\cap X)+P(N\cap X)=P(S)\,P(X|S)+P(N)P(X|N)$$

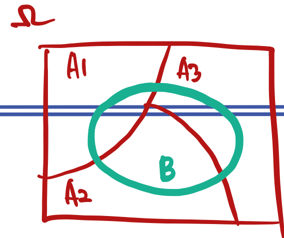
$$\Rightarrow P(S)=0.4,\ P(N)=0.6,\ P(X|S)=0.25,\ P(X|N)=0.02$$

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

○ 확률수형도(probability tree)

교재 98쪽 <그림 3.10> 참조





○ 표본공간의 분할(partition)

- 사건 $A_1, \dots, A_n$ 가 표본공간 Ω의 **분할(partition)**
  - Ⓐ **서로 배반사건**, 즉 모든 $i \neq j$에 대해 $A_i \cap A_j = \varnothing$
  - Ⓑ **전체를 이루는 사건**, 즉 $A_1 \cup \cdots \cup A_n = \Omega$

- 사건 $A_1, \dots, A_n$ 이 표본공간 Ω의 **분할(partition)**이면

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

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

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

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

○ 베이즈정리(Bayes' theorem)

- $P(B|A)$은 순서적으로 볼 때, 대부분 사건 A가 먼저 발생하고 B가 이어서 발생하는 상황에 대한 확률
  - ○ A는 원인, B는 결과의 형태를 가짐
  - ○ 원인 가능성 $P(A)$ 또는 $P(A^c)$는 사건 B가 관측되기 이전 확률
    - ⇨ **사전확률(prior probability)**

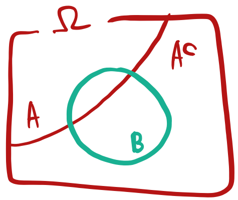
  - ○ 결과 B를 관측했을 때 그 원인이 사건 A일 확률은? $P(A|B)$**?**
  - ○ 이 확률을 사건 B가 관측된 후의 A의 확률이라고 해서
    - ⇨ $P(A|B)$를 **사후확률(posterior probability)**이라고 함

 신부님.

▶ 베이즈(Thomas Bayes, 1701-1973)

○ 만약 $P(B)>0$, $P(A|B)=P(A\cap B)/P(B)$

- $P(A)>0$, $P(A^c)>0$이면, ①과 ②에 의해



베이즈 정리(Bayes' theorem)

$$\underbrace{P(A|B)}_{\text{인 과}}=\frac{P(A\cap B)}{P(B)}=\frac{\overset{P(A\cap B)}{P(A)P(B|A)}}{\underbrace{P(A)P(B|A)}_{P(A\cap B)}+\underbrace{P(A^c)P(B|A^c)}_{P(A^c\cap B)}}$$

※ 어떤 문제에서는 결과를 얻은 상태에서 그 결과가 발생하게 된 원인을 역으로 추정

⇨ **후향적연구(retrospective study), 사례·대조연구(case-control study)**

[예제 3.17] 신종 코로나 감염여부 진단

인 과
$P(+|A)=0.96$

○ 감염 여부 간이진단 검사를 실시, 감염되었을 때 양성 반응이 나올 확률은 0.96, 감염되지 않았을 때 양성 반응이 나올 확률이 0.05

$P(+|A^c)=0.05$

- ○ 만약 이 검사에서 양성 반응이 나왔다면?
- ○ 확률적 표현: A를 코로나에 감염된 사건
  - $P(+|A)=0.96, \quad P(+|A^c)=0.05$
  - 양성 반응이 나왔을 때 코로나에 걸렸을 확률은 $P(A|+)$

  ※ 참고: $P(+|A) \neq P(A|+)$

걸렸을때 양성인경우　　양성반응일때 실제로도 걸린 경우

(전체국민) 코로나감염확률

○ $P(A|+)$를 계산하기 위해서는 $P(A)$를 알아야 함

만약 $P(A)=0.01$이라고 하면

사전확률

새로운정보+로 인해 $P(A)$가 $0.01 \rightarrow 0.1624$로 바뀌었음

$$P(A|+) = \frac{P(A)P(+|A)}{P(A)P(+|A)+P(A^c)P(+|A^c)}$$

$$= \frac{0.01\times 0.96}{0.01\times 0.96 + 0.99 \times 0.05} = \frac{0.0096}{0.0591} = 0.1624$$

$P(A^c|+) = 0.8376$
$= 1 - 0.1624$
(높다)

양성반응일때실제감염도 됐을 경우 (낮다)

⇒ ① 안걸렸는데 양성인경우 $P(+\cap A^c)\downarrow$

② 전체감염자비율 $P(A)\uparrow$

면 이 확률이 커질 수 있다

$+\ P(+|A)\uparrow$

걸렸을때 양성인경우

▶ $P(+|A)=0.96$ 과는 완전히 다름

$$\frac{P(A\cap +)}{P(+)} = \frac{P(A\cap +)}{P(A\cap +)+P(A^c\cap +)}$$

◎ 베이즈 정리의 일반식

- 사건 $A_1, \ldots, A_n$ 은 표본공간 Ω의 분할
- 모든 i에 대해 $P(A_i) > 0$ 이면

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

$P(A_k \cap B)$

$\sum P(A_i \cap B)$

※ $A_1, \ldots, A_n$: 유전형질(원인) & B : 혈우병(결과)

⇨ $P(A_k|B)$ 역으로 결과를 보고 원인을 찾음!!

N → X
S → X
인 과

◉ 스팸메일 필터 ☞ [예제 3.15] 참조

수신메일 내용에 "X"라는 단어가 있을 때 (☜ 추가정보)

이 메일이 스팸메일일 확률은? → 높다면 X가 들어간 것을 모두 스팸처리

- 확률식: $P(S|X) = \dfrac{P(S \cap X)}{P(X)}$

- $P(X) = 0.112$

먼저 구해야함 - $P(S \cap X) = P(S)P(X|S) = 0.4 \times 0.25 = 0.1$

⇨ $P(S|X) = \dfrac{P(S \cap X)}{P(X)} = \dfrac{0.1}{0.112} = 0.8929$

※ **사전확률** $P(S) = 0.4$ → **사후확률** $P(S|A) = 0.8929$

스팸일 확률 update X라는 글자가 있을때 스팸일 확률

3.5 독립사건(independent events)

- $P(A) > 0$이고 $P(B) > 0$이면

  $P(A \cap B) = P(A)\underbrace{P(B|A)}_{P(B)} = P(B)\underbrace{P(A|B)}_{P(A)}$

- 만약 A가 B에 영향을 주지 않고, B가 A에 영향을 주지 않는다면,

  $P(B|A) = P(B)$, $P(A|B) = P(A)$

- 사건 A와 B가 서로 영향을 주고받지 않는 경우,

  "**사건 A와 B는 독립사건(independent events)이다**"라고 함

  $$\rightleftarrows \boxed{P(A \cap B) = P(A)\,P(B)}$$

  필충

[참고]

○ 표본공간과 공집합은 임의의 사건 A와 독립

$$P(\Omega \cap A) = P(A) = P(\underset{1}{\Omega})P(A)$$

$$P(\varnothing \cap A) = P(\varnothing) = P(\underset{0}{\varnothing})P(A)$$

○ 사건 A와 B가 서로 독립적이면 ☞ [예제 3.22] 참조

$\rightleftarrows$ A와 B^c도 독립적임 ex. A는 남성, B는 선호 일경우

$\rightleftarrows$ A^c와 B도 독립적임

$\rightleftarrows$ A^c와 B^c도 독립적임

[예제 3.21] 두 개의 정육면체 주사위 → Ω는 6x6 =36개

- A 두 주사위의 합이 6인 사건
- B 두 주사위의 합이 7인 사건
- C 첫 번째 주사위의 눈이 3인 사건

$$A = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

$$B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

$$C = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$$

Q. A와 C는 독립인가? B와 C는 독립인가?

$\underbrace{P(A \cap C)}_{\frac{1}{36}} \neq \underbrace{P(A)}_{\frac{5}{36}} \times \underbrace{P(C)}_{\frac{6}{36}}$ → 독립X, $\underbrace{P(B \cap C)}_{\frac{1}{36}} = \underbrace{P(B)}_{\frac{6}{36}} \times \underbrace{P(C)}_{\frac{6}{36}}$ → 독립O

○ 3개의 사건 A, B, C에 대해 다음의 네 등식이 모두 성립하면 3개의 사건 A, B, C는 독립사건 또는 서로 독립적(mutually independent)이라고 함

모든조건을 다확인해야함

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

$$P(A \cap C) = P(A)\,P(C)$$

$$P(B \cap C) = P(B)\,P(C)$$

$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

[예제 3.24] 주사위 3개 던지기

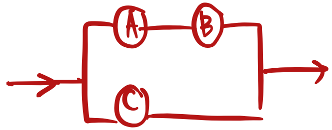
여사건사용

- "6"이 최소한 한 번 이상 나올 확률은?
- A: 주사위 눈이 "6"인 경우가 최소한 한 번 이상 나올 사건
- A_i: i 번째 주사위 눈이 "6"인 사건

$$P(A) = 1 - P(A^c) = 1 - P(A_1^c \cap A_2^c \cap A_3^c) = 0.4213$$

[예제 3.25] 전기전달시스템 <그림 3.12> p. 108

- 세 개의 ON/OFF 스위치로 구성
- 스위치 A, B, C가 ON일 확률은 각각 0.7, 0.8, 0.6
- 각각의 스위치는 **독립적으로** 작동
- A와 B는 직렬, C와 A는 병렬로 구성



▶ 시스템이 전기를 전달할 사건은 $C \cup (A \cap B)$

$$\begin{aligned} P(\text{전기전달}) &= P(C \cup (A \cap B)) \\ &= P(C) + P(A \cap B) - P(A \cap B \cap C) \end{aligned}$$

- 각각의 스위치가 **독립적으로** 작동하므로

$$\begin{aligned} P(\text{전기전달}) &= P(C) + P(A)\,P(B) - P(A)\,P(B)\,P(C) \\ &= 0.6 + 0.7 \times 0.8 - 0.7 \times 0.8 \times 0.6 = 0.824 \end{aligned}$$