

확률 **(Probability)**

■ 기본개념

◎ 확률이 발생하는 상황에서의 공통적인 특징

- ① 주사위 던지기
- ② 앞면이 나올 때까지 동전 던지기
- ③ 구매 한 스마트폰의 수명

- 실험을 시행하기 전에 발생할 수 있는 모든 결과는 알 수 있음
 - ① $\{1, 2, 3, 4, 5, 6\}$
 - ② 앞면¹⁾을 H , 뒷면을 T 이라고 하면, $\{H, TH, TTH, \dots\}$
 - ③ x 를 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)**: 위의 두 성질을 가지는 실험

1) 일반적으로 그림이 있는 부분을 앞면 숫자가 있는 부분을 뒷면이라고 함

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

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

- 어떤 사건 A 가 발생한다는 것은 실험결과가 A 에 속하는 원소 중 하나이거나 A 에 포함되어 있다는 것을 의미

- **확률(probability)**: 이러한 사건이 발생할 가능성이 얼마나 되는지를 나타내는 수치적 측도
 - 확률을 언급하기 위해서는 해당하는 확률실험이 전제되어야 하고 이에 따른 표본공간과 사건이 정해져야 함

- 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
• A 와 B 의 합사건(union)	$A \cup B = \{\omega \mid \omega \in A \text{ 또는 } \omega \in B\}$
• A 와 B 의 곱사건(intersection)	$A \cap B = \{\omega \mid \omega \in A \text{ 그리고 } \omega \in B\}$
• A 의 여사건(complement)	$A^c = \{\omega \mid \omega \notin A \text{ 그리고 } \omega \in \Omega\}$
• 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)의 법칙	$(A \cup B)^c = A^c \cap B^c, (A \cap B)^c = A^c \cup B^c$
• 무한개의 사건이 존재하는 경우 ($A_1, A_2, \dots$)	$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$ $\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$

- 서로배반사건(disjoint, mutually exclusive): 임의의 두 사건 A 와 B 가 공통부분이 없는 경우, 즉 $A \cap B = \emptyset$
- 벤 다이어그램(Venn diagram) 이용

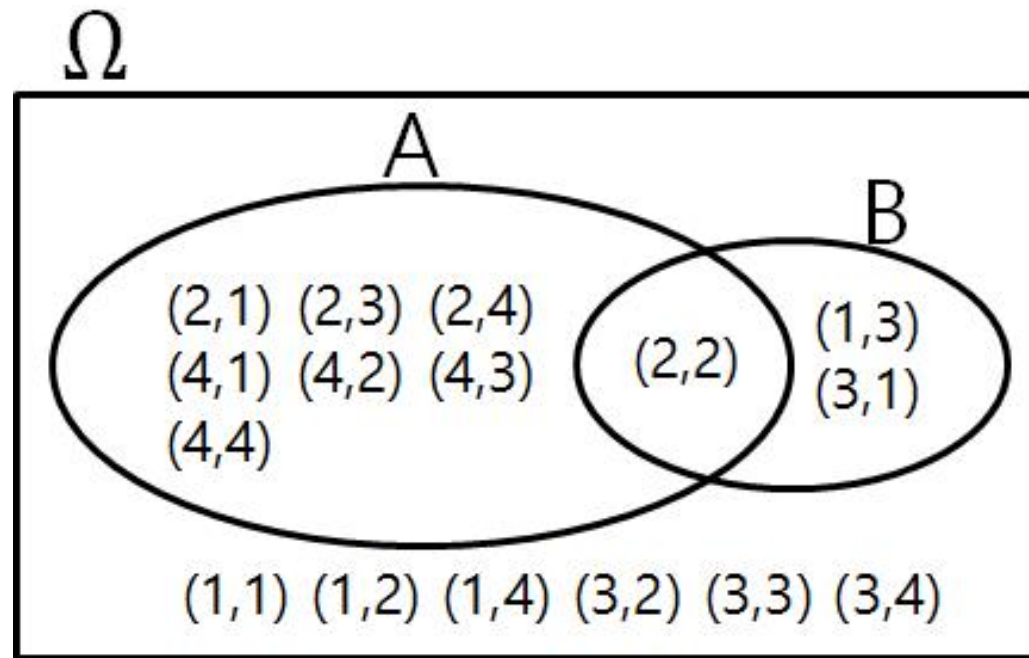
● 정사면체 주사위 두 개를 던지기

- A : 첫 번째 주사위가 짝수인 사건
- B : 두 주사위의 합이 4인 사건

$$\Omega = \{(i, j) \mid i = 1, 2, 3, 4, j = 1, 2, 3, 4\}$$

$$A = \{(i, j) \mid i = 2, 4, j = 1, 2, 3, 4\}$$

$$B = \{(i, j) \mid i + j = 4\} = \{(1, 3), (2, 2), (3, 1)\}$$



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

■ 확률의 이해

□ 고전적 확률

- 17세기 중반 파스칼이나 페르마 등이 도박문제에 대해 의견을 교환
 - 카드게임에서 어떤 패가 더 높은 패인지를 결정하기 위해 각 패의 발생할 수 있는 빈도를 계산
- 표본공간에서 사건이 해당되는 원소가 차지하는 비율

- 표본공간이 n 개의 원소로 이루어져 있고 각 근원사건의 발생가능성이 동일(equally likely)한 경우, k 개의 원소를 가지는 사건 A 의 확률

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

● 정사면체 주사위

- 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(the number of cases)

- 확률을 계산하기 위해서는 표본공간과 사건에 있는 원소의 개수를 효율적으로 계산하는 것이 중요
- 경우의 수의 기본 법칙은 **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

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

◎ 세트메뉴의 경우의 수

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

- 경우의 수의 일반적인 문제 : 1번부터 n 번까지 적혀있는 공이 들어 있는 주머니에서 k 개를 무작위로 선택
- k 개를 어떻게 추출하고 나열할 것인지에 따라 달라짐
- 추출방법:
 - 복원(with replacement)추출
 - 비복원(without replacement)추출
 - Q? 만약 한꺼번에 k 개의 공을 뽑으면?
- 뽑힌 순서를 고려 여부
 - 순서 고려 O: (1, 2)와 (2, 1)을 다른 것
 - 순서 고려 X: (1, 2)와 (2, 1)을 같은 것
 - Q? 뽑은 것을 크기 순서대로 정렬(sorting)한다고 하면?

배열 \ 추출	복원	비복원
	순서고려	순서무시
순서고려	㉠	㉡
순서무시	㉢	㉣

㉡ 순열(permutation)

㉠ 중복순열

㉣ 조합(combination)

㉢ 중복조합

- 순열은 각 단계에서 선택할 수 있는 수가 하나씩 줄어듦

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

- 중복순열의 수는 매번 n 개 선택 가능

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

- 순서를 고려하지 않는다면 같은 번호로 이루어진 순열이 같은 것으로 취급 $\Rightarrow$ 선택된 k 개의 공의 순서열

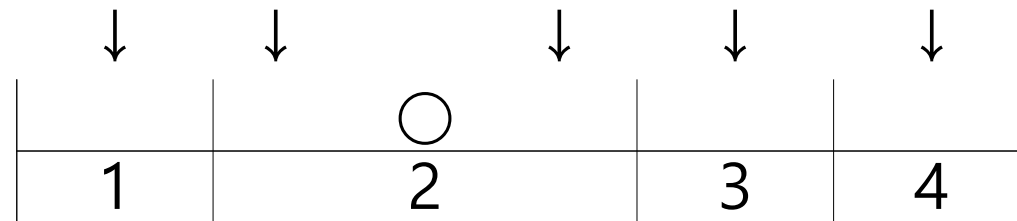
$$k \times (k-1) \times \cdots \times 1 = k!$$

순서를 고려하지 않는 조합의 수는

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

○ 중복조합의 경우, $n^k/k!$?

- $\{1, 2, 3, 4\}$ 에서 2개를 선택하는 중복조합은 10개: (1,1), (1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,3), (3,4), (4,4)
- $4^2/2! = 8 \neq 10$
- n 개의 선택 가능한 공간에 k 개의 공을 넣는 실험



- 첫 번째는 4가지 두 번째는 5가지 $4 \times 5 = 20$
- 순서열의 수 $2! \Rightarrow \frac{5 \times 4}{2!} = \frac{5!}{2!3!} = \binom{5}{2} = 10$

- 만약 3개를 선택한다면 세 번째 공의 화살표는 6개가 되어 $6 \times 5 \times 4$ 가 되고 선택된 3개의 순서열의 수는 3!

$$\frac{6 \times 5 \times 4}{3!} = \frac{6!}{3!3!} = \binom{4+2}{3} = 20$$

- 중복조합의 경우의 수

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

【표 3.2】 경우의 수

배열 \ 추출	복원	비복원
순서고려	n^k	$\frac{n!}{(n-k)!}$
순서무시	$\binom{n+k-1}{k}$	$\binom{n}{k}$

● 나눔Lotto 6/45

- 1에서 45까지의 숫자에서 6개의 번호를 비복원 추출하고 크기순서대로 정렬한 당첨번호를 제공
- 1등: 선택한 6개의 번호가 당첨번호와 모두 일치
- 5개만 일치하면 2등 또는 3등
- 전체 가능한 경우의 수

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

$$\Rightarrow P(1\text{등}) = 1/8145060$$

- 2등과 3등: 순서와 관계없이 6개 당첨번호 중 5개를 선택하고 나머지 하나는 다른 39개 중 하나를 선택

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

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

◎ Birthday problem

- 1년을 365일이고 **365일 동안 태어날 가능성**이 동일
- A : k 명의 사람이 모두 다른 생일을 가지는 사건
- k 명의 사람이 각각 365일 중 한 날을 선택할 수 있으므로

$$\#(\Omega) = 365^k$$

- 모든 사람이 생일이 다르다는 것은 각기 다른 날짜를 선택하는 것으로 365일을 k 번 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	5	10	15	20	30	40	50
$P(A)$	0.9729	0.8831	0.7471	0.5886	0.2937	0.1088	0.0296

○ 연속표본공간

- 발생가능성이 동일한 상황을 선이나 평면 등을 이용
- 사건 A 가 발생한다는 것은, 영역 Ω 내에서 임의의 한 점을 무작위로 택할 때 이 점이 영역 A 에 있다는 의미
- 사건 A 의 확률은 전체 영역에서 A 가 차지하는 비율

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

- $\| \cdot \|$ 는 길이, 면적, 부피 등을 의미

□ 상대도수의 극한개념

- 동전의 앞면이 나올 확률은 $1/2$?
- 고전적 확률: "앞면과 뒷면의 발생가능성이 동일하고 $\Omega = \{H, T\}$, $A = \{H\}$ 이므로 $P(A) = 1/2$ 이다"라고 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

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

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

- 만약 실험을 무한히 반복한다면 $n(A)/n$ 은 어떤 값으로 수렴하는데 이 극한값을 사건 A 가 일어날 확률을 정의

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

- “그냥도전 동전 돌리기”에서 각각의 실험에서 발생하는 결과는 표본이고 **실험을 무한히 반복**한다는 것은 **표본이 결국에는 모집단**이 됨

- **확률은 모집단**이 어떻게 형태로 이루어져 있는지를 보여줌
- 상대도수의 극한으로써의 확률은 많은 표본을 통해 모집단의 특성을 파악하는 방식을 따른다고 해서 **통계적 확률(statistical probability)**이라고 함
- 일기예보나 전쟁게임과 같이 모의실험(simulation)을 통해 결과를 도출하는 분야에서 많이 사용되고 있으며 고속 컴퓨터의 보급으로 인해 더욱더 활용도가 높아지고 있음

● 생일문제

- 365일 각각의 날에 태어날 가능성이 동일?

상황	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
Ⓐ	1452	1450	1394	1337	1271	1273	1294	1335	1390	1348	1283	1106
Ⓑ	2000	2000	2000	2000	2000	2000	1000	1000	1000	1000	1000	1000

- Ⓐ: 2012년에 태어난 484,300명의 아이들의 월별 하루 평균출생아수
- Ⓑ: 1월에서 6월까지의 일별 출생아가 7월에서 12월까지 일별 출생아의 두 배 가정
- 각각의 k 에 대해 일억 번 실시한 비율

【표 3.3】 상황에 따른 생일문제의 확률 차이

k	5	10	15	20	30	40	50
$P(A)$	0.9729	0.8831	0.7471	0.5886	0.2937	0.1088	0.0296
Ⓐ	0.9728	0.8825	0.7464	0.5870	0.2922	0.1076	0.0292
Ⓑ	0.9699	0.8709	0.7230	0.5551	0.2569	0.0853	0.0202

□ 공리적 확률

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

[공리 1] $P(\Omega) = 1$

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

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

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

■ 확률의 기본정리

① $P(A^c) = 1 - P(A)$

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

○ $P(\emptyset) = 0$

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

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

◎ 1000장 중 4장이 당첨복권

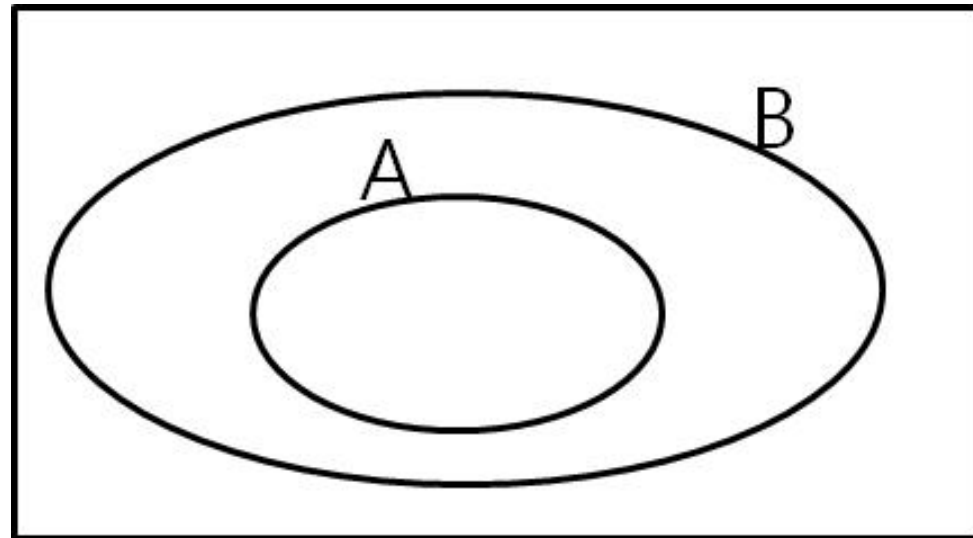
- 4장의 복권을 구입한다면 적어도 한 장 이상의 당첨복권을 구입하게 될 확률은?
- A : 한 장 이상의 당첨복권을 구입할 사건 = 당첨복권이 한 장, 두 장, 세 장, 네 장인 경우

⇒ A^c : 구입한 4장 모두 당첨되지 않을 사건

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

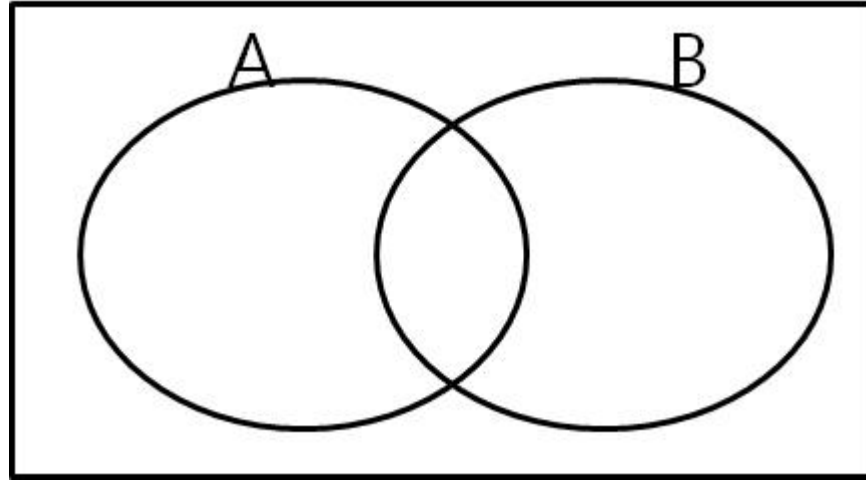
$$\Rightarrow P(A) = 1 - P(A^c) = 0.0159$$

② $A \subset B$ 이면 $P(A) \leq P(B)$



○ $B = A \cup (B \cap A^c)$

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



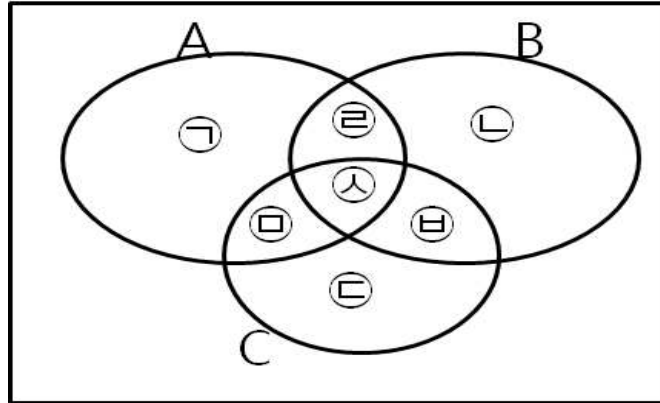
◦ $A = (A \cap B) \cup (A \cap B^c)$

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

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

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

- 사건 A, B, C



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

$$- P(\text{㉠}) = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$$

$$- P(\text{㉣}) = P(A \cap B) - P(A \cap B \cap C)$$

$$- P(\text{㉦}) = P(A \cap B \cap C)$$

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

- n 개의 사건 $A_1, A_2, \dots, A_n$ 에 대해

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &\quad + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n) \end{aligned}$$

④ $P(A \cup B) \leq P(A) + P(B)$

○ 부울의 부등식(Boole's inequality): $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$

○ $P(A^c \cup B^c) \leq P(A^c) + P(B^c)$

$\Rightarrow P(A^c \cup B^c) = 1 - P(A \cap B) \leq 1 - \{P(A) + P(B)\}$

○ 본페로니 부등식(Bonferroni's inequality)

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

◎ 1000장 중 4장이 당첨복권

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

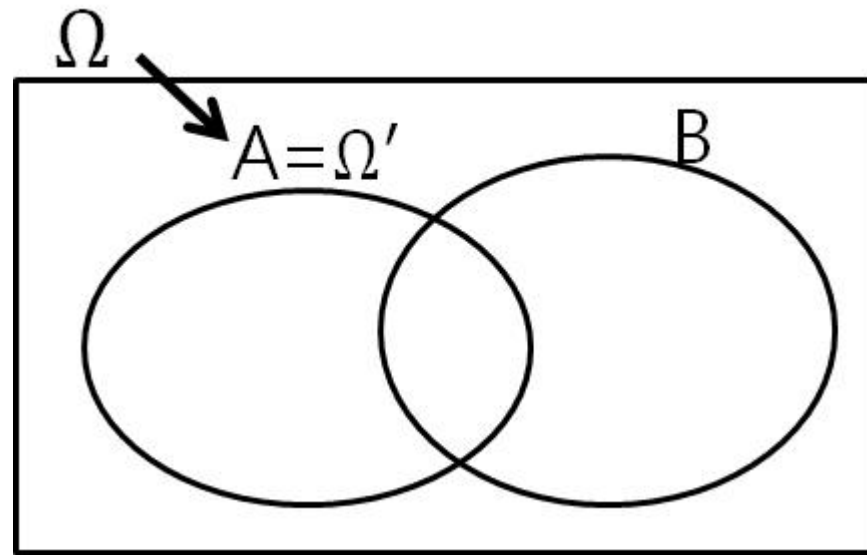
○ $A = \bigcup_{i=1}^4 A_i$ 로 표시

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

■ 조건부 확률

- ◎ 동전 두 개를 던지는 실험에서 어떤 한 동전이 앞면이라는 것을 알았을 때, 두 동전 모두 앞면일 사건의 확률은?
 - $\Omega = \{HH, TH, HT, TT\}$
 - 어떤 한 동전이 앞면이라는 정보가 추가로 주어지면 표본공간에서 $\{TT\}$ 가 발생할 수 없음
 - ⇒ 표본공간은 $\{HH, TH, HT\}$ 로 축소

- **조건부 확률(conditional probability)**: 확률실험에서 새로운 정보 또는 조건이 추가되었을 때 사건의 확률



- 사건 A 가 발생했다면 A 이외의 것은 일어날 수 없기 때문에 A 가 새로운 표본공간 Ω' 이 되고, B 가 발생한다는 것은 A 안에서 $A \cap B$ 에 있는 원소가 발생하는 것을 의미

- A 하에서 B 의 조건부확률은 A 에서 $A \cap B$ 가 차지하는 비율
- 사건 A 가 주어졌을 때 사건 B 의 조건부 확률은 $P(B|A)$ 로 표시

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

○ 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 비율로 (1-사망률)로 계산

● 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

- 0세 사망자는 10만 중 $100000-99709=291$ 명
 $\Rightarrow$ 영아 사망률 = $291/100000=0.00291(0.29\%)$
- 40세의 남성사망률은 40세 이상 남자 생존자 97727명 중
 40세에 사망한 $97727-97581=146$ 명
 $\Rightarrow$ 40세 남성사망률 = $\frac{146}{97727} = 0.00149 = 0.15\%$
- 80세 여성 생존율은 $1-2822/75732=1-0.0373=0.967$ 로
 96.7%

【표 3.5】 연령별 사망자와 사망률

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%

□ 조건부확률의 활용

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

- 곱사건이 순차적인 사건들의 조건부확률의 곱으로 표시

● 정상 제품 90개와 불량품 10개가 들어있는 상자에서 무작위로 2개를 비복원으로 추출

- $\Omega = \{(\text{정상}_1, \text{정상}_2), (\text{정상}_1, \text{불량}_2), (\text{불량}_1, \text{정상}_2), (\text{불량}_1, \text{불량}_2)\}$

- (정상1, 정상2)의 확률은?

- 첫 번째가 정상일 확률은 90/100
- 두 번째 **도** 정상일 확률은 89/99

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

$$P(\text{정상}_1) \quad \leftarrow \quad \hookrightarrow \quad P(\text{정상}_2 | \text{정상}_1)$$

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

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

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

- A_1, A_2, A_3 에 대해 $P(A_1 \cap A_2) > 0$ 이면

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

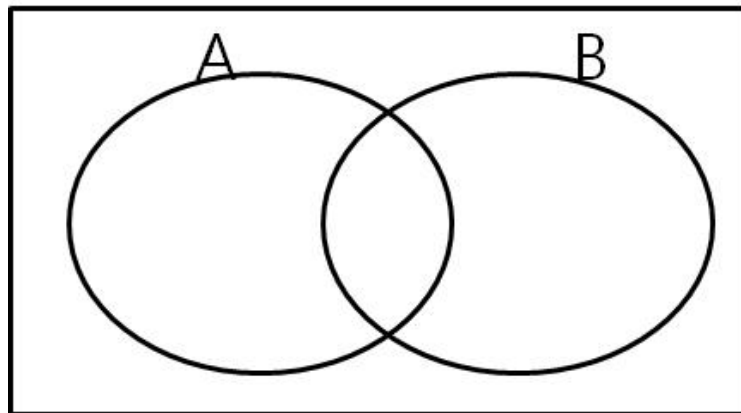
$$\Rightarrow P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

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

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

- ◎ 당첨복권이 4장인 복권 1000장 발매
 - A_i : 구입한 4장의 복권 중에서 i 번째 복권이 당첨될 사건
 - $P(A_i) = ?$
 - $P(A_1) = 0.004$

$$\begin{aligned}
 \textcircled{2} \quad P(B) &= P(A \cap B) + P(A^c \cap B) \\
 &= P(A)P(B|A) + P(A^c)P(B|A)
 \end{aligned}$$



$$\begin{aligned}
 \circ \quad P(A_2) &= P(A_1)P(A_2|A_1) + P(A_1^c)P(A_2|A_1^c) \\
 &= \frac{4}{1000} \frac{3}{999} + \frac{996}{1000} \frac{4}{999} = \frac{4}{1000} = 0.004
 \end{aligned}$$

- $$\begin{aligned}
 P(A_3) &= P(A_1 \cap A_2 \cap A_3) + P(A_1^c \cap A_2 \cap A_3) \\
 &\quad + P(A_1 \cap A_2^c \cap A_3) + P(A_1^c \cap A_2^c \cap A_3) \\
 &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\
 &\quad + P(A_1^c)P(A_2|A_1^c)P(A_3|A_1^c \cap A_2) \\
 &\quad + P(A_1)P(A_2^c|A_1)P(A_3|A_1 \cap A_2^c) \\
 &\quad + P(A_1^c)P(A_2^c|A_1^c)P(A_3|A_1^c \cap A_2^c) = 0.004
 \end{aligned}$$
- 어떤 일련의 사건들이 순차적 또는 결합된 형태로 되어 있는 상황에서 특정 시점이나 위치에 해당되는 사건의 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)
- 스팸메일 중 내용에 "A"라는 단어가 있는 메일은 25%이고 정상메일 중 이 단어가 있는 경우는 2%
- 전체 메일 중 "A" 단어를 포함한 메일의 비율은?

$$P(A) = P(S \cap A) + P(N \cap A) = P(S)P(A|S) + P(N)P(A|N)$$

- $P(S) = 0.4, P(N) = 0.6, P(A|S) = 0.25, P(A|N) = 0.02$

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

○ 확률수형도(probability tree)

○ 표본공간의 분할(partition)

○ 사건 $A_1, \dots, A_n$ 가

① 서로배반사건, 즉 모든 $i \neq j$ 에 대해 $A_i \cap A_j = \emptyset$

② 전체를 이루는 사건(exhaustive), 즉 $A_1 \cup \dots \cup A_n = \Omega$

이면, 사건 $A_1, \dots, A_n$ 을 표본공간 Ω 의 분할이라고 함

○ 사건 $A_1, \dots, A_n$ 이 표본공간 Ω 의 분할이면

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

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

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

○ 베이즈정리(Bayes' theorem)

- $P(B|A)$ 은 순서적으로 볼 때, 대부분 사건 A 가 먼저 발생하고 B 가 이어 발생하는 상황에 대한 확률
 - A 는 원인, B 는 결과의 형태를 가짐
 - 원인의 가능성인 $P(A)$ 또는 $P(A^c)$ 는 사건 B 가 관측되기 이전의 확률
 - ⇒ 사전확률(prior probability)

- 어떤 문제에서는 결과를 얻은 상태에서 그 결과가 발생하게 된 원인을 역으로 추정 $\Rightarrow$

사례-대조연구(case-control study)

- 결과 B 의 관측했을 때 그 원인이 A 일 사건의 확률은?

$$P(A|B)$$

- 이 확률을 사건 B 가 관측된 후의 A 의 확률이라고 해서
사후확률(posterior probability)이라고 함

● 암진단

- 암에 대한 간이진단 검사를 실시하는데 암에 걸렸을 때 양성반응이 나올 확률은 0.96이고 암에 걸리지 않았을 때 양성반응이 나올 확률이 0.05
- 만약 이 검사에서 양성반응이 나왔다면?
- 확률적 표현: A 를 암에 걸린 사건
 - $P(+|A) = 0.96, \quad P(+|A^c) = 0.05$
 - 양성반응이 나왔을 때 암에 걸렸을 확률은 $P(A|+)$
$$P(+|A) \neq P(A|+)$$

- 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

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

- $P(A|+)$ 를 계산하기 위해서는 $P(A)$ 를 알아야 함
 - 만약 $P(A) = 0.01$ 이라고 하면

$$\begin{aligned} P(A|+) &= \frac{P(A)P(+|A)}{P(A)P(+|A) + P(A^c)P(+|A^c)} \\ &= \frac{0.01 \times 0.96}{0.01 \times 0.96 + 0.99 \times 0.05} = \frac{0.0096}{0.0591} = 0.1624 \end{aligned}$$

- $P(+|A) = 0.96$ 과는 상당한 차이

- 베이즈 정리의 일반식
 - 사건 $A_1, \dots, A_n$ 은 표본공간 Ω 의 분할
 - 모든 i 에 대해 $P(A_i) > 0$ 이면

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

◎ 스팸메일 필터

- 수신메일 내용에 "A"라는 단어가 있을 때 이 메일이 스팸메일일 확률은?

- 확률식: $P(S|A)$

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

- $P(A) = 0.112$

- $P(S \cap A) = P(S)P(A|S) = 0.4 \times 0.25 = 0.1$

$$\Rightarrow P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{0.1}{0.112} = 0.8929$$

■ 독립사건(independent events)

- $P(A) > 0$ 이고 $P(B) > 0$ 이면

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

- 만약 사건 A 가 사건 B 의 발생에 영향을 주지 않고 B 가 사건 A 에 영향을 주지 않는다면,

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

- 사건 A 와 B 가 서로 영향을 주고받지 않는 경우, "사건 A 와 B 는 독립사건(independent events)이다."라고 함

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

- 표본공간과 공집합은 임의의 사건 A 와 독립

$$P(\Omega \cap A) = P(A) = P(\Omega)P(A)$$

$$P(\emptyset \cap A) = P(\emptyset) = P(\emptyset)P(A)$$

● 두 개의 정육면체 주사위

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

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

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

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

- A 와 C 는 독립인가? B 와 C 는 독립인가?

- 셋 사건 A, B, C 에 대해 다음의 네 등식이 모두 성립하면 셋 사건 A, B, C 는 독립사건 또는 서로 독립적(mutually independent)이라 함

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

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

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

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

◎ 주사위 3개를 던지기

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

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

● 전기전달시스템

- 세 개의 ON/OFF 스위치로 구성
- 스위치 A , B , C 가 ON일 확률은 각각 0.7, 0.8, 0.6
- 각각의 스위치는 독립적으로 세팅
- A 와 B 는 직렬, C 와 A 는 병렬로 구성

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

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

- 각각의 스위치가 독립적으로 세팅

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

○ 용어 및 정의

- 확률실험(random experiment)
- 표본공간(sample space)
- 사건(event)
- 서로배반사건(disjoint, mutually exclusive)
- 발생가능성 동일(equally likely)
- 복원(with replacement) & 비복원(without replacement)
- 통계적 확률(statistical probability)
- 조건부 확률(conditional probability)
- 분할(partition) - exhaustive
- 독립사건(independent events)

○ 정리(theorem)

- 경우의 수: 곱의 법칙
- 확률의 공리
- $P(A^c) = 1 - P(A)$
- $A \subset B$ 이면 $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
- $P(A \cup B) \leq P(A) + P(B)$
- $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

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

- 베이즈정리

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