제 4 장 확률이론

집합 용어 및 표기법

집합(set)은 서로 구별되는 대상을 순서와 무관하게 모은 것이고, 원소(element)는 집합에 속하는 각각의 대상을 말한다. 집합의 원소는 어떤 것이라도 무방하다. 예) 숫자, 글자, 국가 등

집합 표현

① 언어로 표현
A is the set whose members are the first ten positive integers.

② 원소나열법
A = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
A = {1, 2, 3, ..., 10} The ellipsis indicates that the list continues in the obvious way.

③ 조건제시법
A = { x : x is an integer, 1≤x≤10} ← {원소: 원소의 조건}
A = { x | x is an integer, 1≤x≤10} ← :(colon) 또는 |(vertical bar) 사용

원소가 두 개 이상의 변수로 이루어진 경우는 ()를 사용한다. 예: 통계학 시험 점수 → (중간고사점수, 기말고사점수, 퀴즈점수, 과제물점수) → (95, 98, 100, 85)

함수(function)의 이해

f(x) = y
Input x에 대응하는 Output y를 찾아준다.
일반적인 형식: 함수_이름(Input) = Output

예: 개당 가격이 천원이 어떤 제품을 x개 구입할 때의 구매 총액
구매_총액(x) = 1,000x

함수 이름은 “구매_총액”처럼 길어도 무관하며, 흔히 f, g, P, p, F 등을 많이 사용한다.

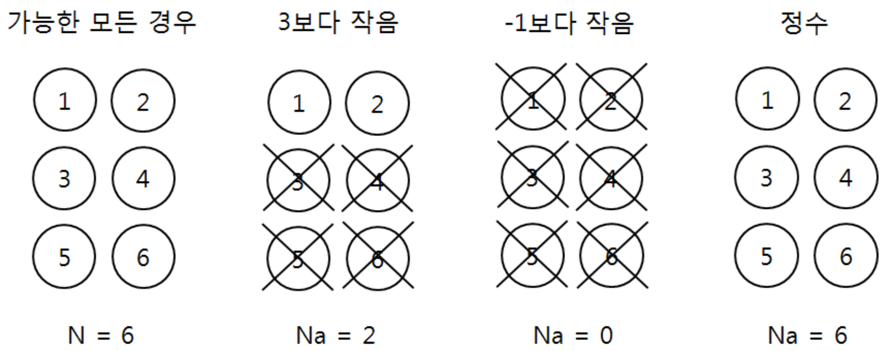
1. 확률의 이론

1.1 동등발생정의-equally likely definition

$$P(A) = \frac{\text{A가 일어날 경우의 수}}{\text{가능한 모든 경우의 수}} = \frac{N_A}{N}$$

여기서, N: 상호배타적이고 동등발생적인 결과의 수
(mutually exclusive and equally likely outcomes)

N_A: N개의 결과 중 사건 A에 해당(조건 A에 부합)하는 결과의 수



예) 주사위를 던졌을 때 (주사위 윗면에) 나오는 수 (1)

P(3보다 작은 수가 발생) ← 3보다 작은 수가 발생할 확률

= P(A) ← 3보다 작은 수가 발생하는 사건을 A라 정의하면

= $\frac{N_a}{N}$ ← 주사위 경우의 수를 N이라 하고, 3보다 작은 수들의 경우의 수를 N_a라 하면

= 2/6

예) 주사위를 던졌을 때 (주사위 윗면에) 나오는 수 (2)

가능한 모든 경우를 나열하면 1, 2, 3, 4, 5, 6이고, 모든 경우의 수는 6이다.
→ 주사위 던지는 실험을 아무리 많이 반복하여도 결과의 경우는 총 6개이다.

1.1.1 mutually exclusive:

사건 A와 B가 상호배타적이란 'A와 B가 동시에 발생하지 않는다.'를 의미한다. 1이 나오면서 2 또는 3이 나오는 경우는 없다. 한 번 시행에 발생하는 결과는 1부터 6까지의 숫자 중 단 하나일 뿐, 2개 이상이 겹치는 경우는 없다.

① 상호배타적 사건들: 1, 2
주사위를 던졌을 때 1이면서 2인 경우는 없다.

② 상호배타적이지 않은 사건들: '짝수', '5보다 작은 수'
주사위를 던졌을 때 2 또는 4가 나오면, '짝수'이면서 '5보다 작은 수' 모두에 해당된다.

1.1.2 equally likely

(주사위를 던지는) 시행의 결과로 1이 2보다 더 많이 발생하거나, 2가 3보다 더 많이 발생한다고 간주할 어떤 근거도 없다. 6개의 결과 중 어떤 것도 다른 것보다 발생 확률이 더 크다고 간주할 근거가 없으므로, 각 결과의 발생 확률을 동일한 값으로 간주한다.

① equally likely에 해당하는 event들:
주사위 던졌을 때 나오는 숫자들: 1, 2, ..., 6
동전 던졌을 때의 결과: 앞면, 뒷면

② equally likely에 해당되지 않는 event들:
성적 결과: 0, 1, 2, ..., 100

주사위를 던졌을 때 2가 나오는 경우의 수 .... 1 (2 하나만 해당)
짝수인 경우의 수 .... 3 (2, 4, 6 해당)
4보다 큰 경우의 수 .... 2 (5, 6 해당)
7보다 큰 경우의 수 .... 0 (해당되는 경우 없음)
0보다 작은 경우의 수 .... 0 (해당되는 경우 없음)

주사위를 던졌을 때 2가 나올 확률 = $\frac{\text{2가 나오는 경우의 수}}{\text{주사위를 던졌을 때 나오는 모든 경우의 수}}$

기호로 표현하면 → P(2) = $1/6$

주사위를 던졌을 때 짝수가 나올 확률
= P(짝수) = $3/6$ = 0.5

1.1.3 주사위 예를 통한 확률의 특성 이해

① $P(A)$의 값이 음수일 수 있는가?
동등발생정의에 의하면 $P(A) = \frac{\text{A가 일어날 경우의 수}}{\text{가능한 모든 경우의 수}}$이다.
'A가 일어날 경우의 수'는 0이므로(음수가 아니므로), $P(A)$는 음수가 될 수 없다.

② $P(A)$의 최소값은?
위에서 언급한 것처럼 'A가 일어날 경우의 수'는 최소 0이므로, $P(A)$의 최소값은 0이다.

예) $P(-1) = \frac{-1\text{이 발생할 경우의 수}}{\text{가능한 모든 경우의 수}}$

$= \frac{0\,\{\text{주사위에 }-1\text{이 없으므로 발생 경우의 수는 }0\text{이다.}\}}{6\,\{\text{가능한 모든 경우의 수는 }1,2,\ldots,6\text{ 총 }6\text{이다.}\}}$

③ $P(A)$의 최대값은?
A가 {가능한 모든 경우의 수}와 같다면, 분자와 분모가 동일하므로 $P(A)$는 1이 된다.
예) A는 정수라 하자. 주사위 결과 1, 2, ..., 6 중 정수에 해당하는 결과의 수는 모두 6이다. 그러므로 $\mathrm{P}(\text{정수}) = \frac{6}{6} = 1$이 된다.

정리하면, $0 \leq P(A) \leq 1$이 성립한다.

이 개념에 의한 확률계산은 순전히 이론적인 추론에 근거한 것(실제 실험자료를 수집할 필요 없다.)

연습문제 1.

문제 1. 동등발생정의를 수식으로 표현하시오. 단, N: 상호배타적이고 동등발생적인 결과의 수, n: N개의 결과 중 사건 A 에 해당(조건 A 에 부합)하는 결과의 수를 나타냅니다.

문제 2. 동등발생정의를 수식으로 표현하시오. 단, N: 상호배타적이고 동등발생적인 결과의 수, N_b: N개의 결과 중 사건 B 에 해당(조건 B 에 부합)하는 결과의 수를 나타냅니다.

문제 3. 주사위를 던졌을 때 3 이 나올 확률은?

문제 4. 주사위를 던졌을 때 2 이하의 값이 나올 확률은?

문제 5. ___ $\leq P(A) \leq$ ___

1.2 상대빈도정의-relative frequency definition

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

$P(A)$: A 사건이 발생할 확률
n: 총시행횟수
r: A 사건이 발생한 횟수

Example: 동전을 던졌을 때 앞면이 나올 확률에 관한 컴퓨터 모의실험

컴퓨터가 생성한 Random Number 값이 0.5 미만이면 앞면이 나온 것으로 간주하고, 그렇지 않으면 뒷면이 나온 것으로 간주한다. 여기서, Random Number는 0과 1 사이의 값을 갖는 무작위값이다.

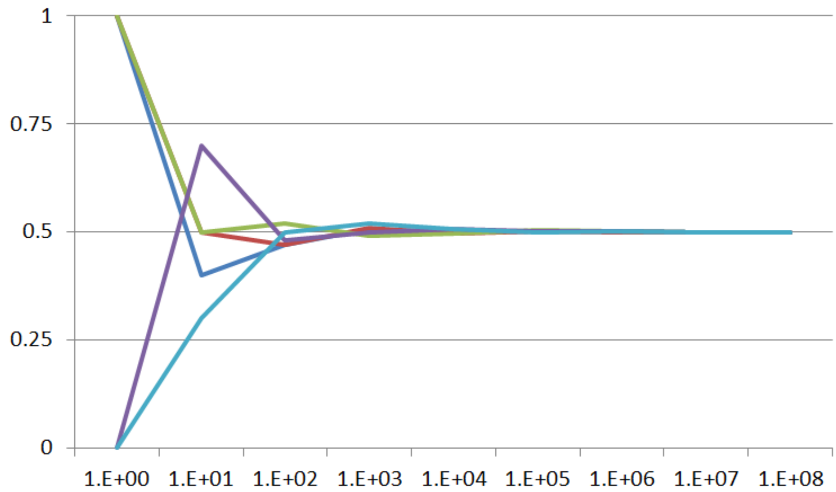
총 실험횟수: 10

시행횟수 (A)	Random 값	앞면 발생 여부 *	누적발생횟수(B)	확률 (B/A)
1	0.5021	0	0	0.0000
2	0.5637	0	0	0.0000
3	0.6343	0	0	0.0000
4	0.4624	1	1	0.2500
5	0.1630	1	2	0.4000
6	0.4065	1	3	0.5000
7	0.5009	0	3	0.4286
8	0.5408	0	3	0.3750

9	0.8204	0	3	0.3333
10	0.9753	0	3	0.3000

앞면이면 (0.5 미만이면) 1, 아니면 (0.5 이상이면) 0의 값을 갖는다.

i번째 시행	i 번째까지의 발생횟수 평균				
	실험 1	실험 2	실험 3	실험 4	실험 5
1	1.000000	1.000000	1.000000	0.000000	0.000000
10	0.400000	0.500000	0.500000	0.700000	0.300000
100	0.470000	0.470000	0.520000	0.480000	0.500000
1,000	0.507000	0.510000	0.492000	0.499000	0.520000
10,000	0.499400	0.496400	0.495400	0.505600	0.505600
100,000	0.501620	0.500230	0.503050	0.501330	0.499200
1,000,000	0.500734	0.499970	0.500968	0.500317	0.500380
10,000,000	0.500007	0.499865	0.499889	0.499806	0.499781
100,000,000	0.499997	0.499991	0.500003	0.499993	0.499998



연습문제 2.

문제 1. 상대빈도정의를 수식으로 표현하시오. 단, $P(A)$: A 사건이 발생할 확률, n: 총시행횟수, r: A 사건이 발생한 횟수를 의미합니다.

문제 2. 상대빈도정의에 의하면, 주사위를 12 번 던졌을 때 6 이 나온 상대빈도는 정확히 1/6 이다. True, False

1.3 주관적 확률개념-subjective probability concept

이용 가능한 정보에 근거하여 어떤 사건이 일어날 것이라고 주관적으로 믿고 있는 정도

2. 실험과 표본 공간

실험: experiment, 어떤 결과에 대한 관찰이거나 측정치를 구하는 절차

표본공간: sample space, 실험의 결과로 얻어진 가능한 모든 사건의 집합

Example 두 개의 동전을 던지는 실험

앞면이면 0, 뒷면이면 1; 표기 (첫째 동전의 실험결과, 둘째 동전의 실험결과)

표본공간 S = {(0,0), (0,1), (1,0), (1,1)}

3. 확률법칙

3.1 확률의 특성 (확률의 공리)

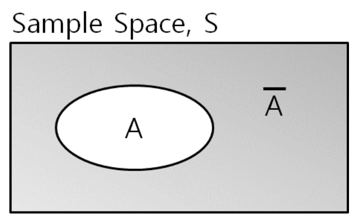
공리 1. $0 \le P(A) \le 1$

공리 2. $P(S) = 1$

공리 3. $P(\overline{A}) = 1 - P(A)$, 여기서 $\overline{A}$는 사건 A가 발생하지 않는 사건이다.

A와 $\overline{A}$: 상호배타적(mutually exclusive)이며 동시에 완전포괄적(collectively exhaustive)인 관계이다.

$\overline{A}$는 A' 또는 A^c로 표기하기도 한다.

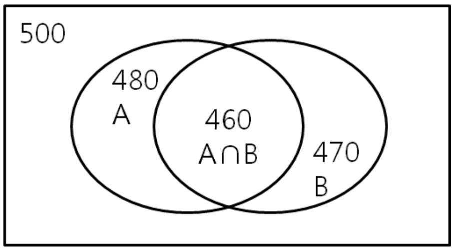


3.2 덧셈법칙

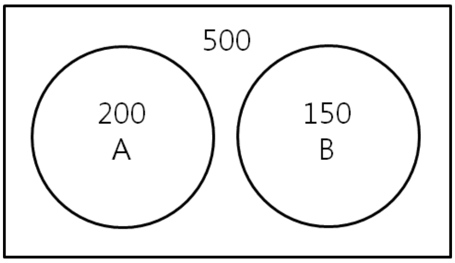
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

사건 A와 B가 배타적인 경우: $P(A \cup B) = P(A) + P(B) \because P(A \cap B) = 0$

3.2.1 Example) $P(A \cup B) = ?$



$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{480}{500} + \frac{470}{500} - \frac{460}{500} = \frac{490}{500}$$



$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= P(A) + P(B) \; since \; P(A \cap B) = 0 \\ &= \frac{200}{500} + \frac{150}{500} = 0.70 \end{aligned}$$

연습문제 3. 사건 A는 주사위를 던졌을 때 2 이하의 값이 나오는 사건을, 사건 B는 짝수가 나오는 사건을 의미한다고 하자.

문제 1. 사건 A 는 주사위를 던졌을 때 2 이하의 값이 나오는 사건을 의미하다면, $\overline{A}$란 ______ 사건을 의미한다.

문제 2. 주사위 6 개의 값 중 사건 A 에 속하는 값을 모두 나열하시오.

문제 3. $P(A) =$

문제 4. $P(\overline{A}) =$

문제 5. $P(A) + P(\overline{A}) =$

문제 6. 주사위 6 개의 값 중 사건 B 에 속하는 값을 모두 나열하시오.

문제 7. $P(B) =$

문제 8. 주사위 6 개의 값 중 사건 A 에도 속하고 사건 B 에도 속하는 값을 모두 나열하시오.

문제 9. $P(A \cap B)$ =

문제 10. 주사위 6 개의 값 중 사건 A 또는 B 에 속하는 값을 모두 나열하시오.

문제 11. $P(A \cup B)$ =

3.3 결합확률과 한계확률

특정 사안에 대한 성별 찬성, 반대 인원 (총인원 1,000명)

→ 횡단적 자료(cross-sectional data)

	찬성(B1)	반대(B2)	합계
남(A1)	250	450	700
여(A2)	200	100	300
합계	450	550	1,000

	찬성(B1)	반대(B2)	합계
남(A1)	①	②	①+②
여(A2)	③	④	③+④
합계	①+③	②+④	①+②+③+④

연습문제 4.

문제 1. 위의 표에서 250 의 의미는?

문제 2. 위의 표에서 300 의 의미는?

문제 3. 위의 표에서 1,000 의 의미는?

문제 4. 어떤 횡단적 자료의 부분 정보이다. 전체 표를 완성하시오.

	국산	외제	합계
소형	10		25
중형		10	50
대형			
합계	70		100

특정 사안에 대한 성별 찬성, 반대 확률

	찬성(B1)	반대(B2)	합계
남(A1)	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	$P(A_1 \cap B_1)+P(A_1 \cap B_2)$ $= P(A_1)$
여(A2)	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	$P(A_2 \cap B_1)+P(A_2 \cap B_2)$ $= P(A_2)$
합계	$P(A_1 \cap B_1)+P(A_2 \cap B_1)=$ $P(B_1)$	$P(A_1 \cap B_2)+P(A_2 \cap B_2)$ $= P(B_2)$	$P(A_1)+P(A_2) = P(B_1)+P(B_2)$ $= P(S)$

	찬성(B1)	반대(B2)	합계
남(A1)	$\frac{250}{1,000}$	$\frac{450}{1,000}$	$\frac{700}{1,000}$
여(A2)	$\frac{200}{1,000}$	$\frac{100}{1,000}$	$\frac{300}{1,000}$
합계	$\frac{450}{1,000}$	$\frac{550}{1,000}$	$\frac{1,000}{1,000}$

3.3.2 결합확률 – joint probability, 두 사건이 동시에 일어날 확률

$$P(A_1 \cap B_1) = \frac{250}{1,000} = 0.25$$

3.3.3 한계확률 – marginal probability, 관련된 결합확률을 더한 확률

$$P(B_1) = P(A_1 \cap B_1) + P(A_2 \cap B_1) = 0.25 + 0.20 = 0.45$$

위의 예에서 개체(item)는 사람

각 사람마다 두 속성을 관찰 – 속성 1. 성별; 속성 2. 찬반

결합확률이란 특정한 두 속성값을 모두 충족하는 비율(예: 남자-반대, 또는 찬성-여자 등)을 의미하고

한계확률이란 특정한 한 속성값을 만족시키는 비율(예: 반대 – 남녀 불문, 여자 – 찬반 불문 등)을 의미한다.

연습문제 5.

문제 1. 위의 표에서 0.250 의 의미는?

문제 2. 위의 표에서 0.300 의 의미는?

문제 3. 위의 표에서 여자일 확률은 어떤 확률들의 합과 동일한가? (한계확률)

3.4 조건확률(조건부확률, conditional probability)

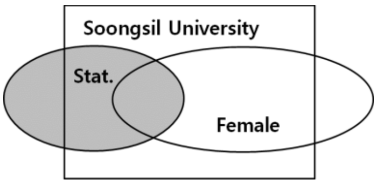
관련된 사건이 일어난 조건 하에서 또 다른 사건이 발생할 확률

어떤 사건이 일어난 또는 일어날 조건 하에서, 즉 변화된 표본공간에서 어떤 사건이 일어날 확률

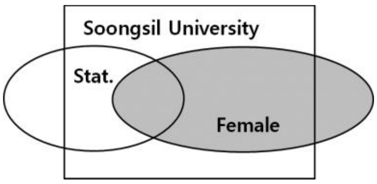
$P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B) > 0$ ← the probability of A given B

example) $P(A_2|B_2) = \frac{P(A_2 \cap B_2)}{P(B_2)} = \frac{0.10}{0.55} = 0.18$

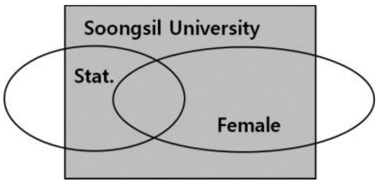
Example) Stat.-현재 통계학 수강생, Female-여학생, Soongsil Univ.-숭실대학교



Stat.

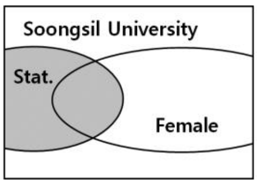


Female

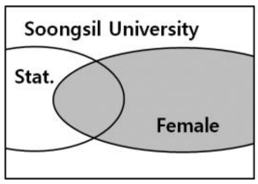


Soongsil University

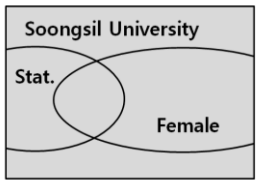
숭실대학교가 표본공간 S이라면, Stat.이나 Female 모두 Soongsil University 이내에 있는 원소들만 의미가 있다.



숭실대 내의 Stat.



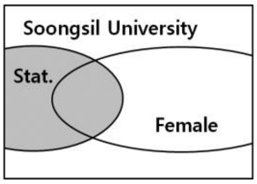
숭실대 내의 Female



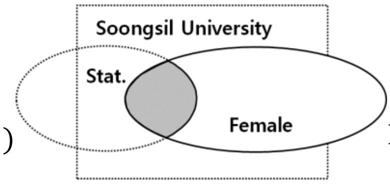
숭실대 내의 Stat.은 P(Stat.|Soongsil University)로 표기해야 하지만, 모든 원소는 표본공간인 Soongsil University 내에서만 의미가 있으므로 P(Stat.|Soongsil University)에서 | 뒷부분은 생략하고 P(Stat.)으로 표기한다.

조건부 확률이란 동일한 사건을 다루지만, 표본공간이 변경되는 경우이다. 예를 들어 P(Stat.)과 P(Stat.|Female)은 모두 Stat.이란 Event를 다루지만, P(Stat.)의 경우는 표본공간이 Soongsil이고 P(Stat.|Female)의 경우는 표본공간이 Female이다. 단, 여기서 Stat.이

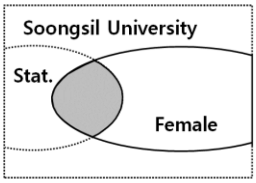
나 Female 모두 원래 표본공간 내에서의 Stat.과 Female을 의미한다.



P(Stat.)



P(St|Fe)의 잘못된 표현



P(St|Fe)의 정확한 표현

연습문제 6.

	찬성(B_1)	반대(B_2)	합계
남(A_1)	0.25	0.45	0.70
여(A_2)	0.20	0.10	0.30
합계	0.45	0.55	1.00

문제 1. 임의로 한 명을 선택했을 때, 그 사람이 찬성일 확률은?

문제 2. 한 명을 선택했더니 남자였다. 그 사람이 찬성일 확률은? (조건부 확률)

문제 3. 한 명을 선택했더니 찬성이었다. 그 사람이 남자일 확률은? (조건부 확률)

연습문제 7.

No	거주형태	거주지	연령	성별	No	거주형태	거주지	연령	성별
1	단독세대	비수도권	35세 미만	여자	11	아파트	수도권	35세 미만	남자
2	단독세대	비수도권	35세 이상	남자	12	아파트	수도권	35세 미만	남자
3	아파트	비수도권	35세 미만	남자	13	아파트	수도권	35세 이상	남자
4	아파트	수도권	35세 미만	여자	14	아파트	수도권	35세 미만	남자
5	아파트	수도권	35세 이상	남자	15	단독세대	수도권	35세 미만	남자
6	아파트	비수도권	35세 미만	남자	16	아파트	비수도권	35세 이상	남자
7	단독세대	수도권	35세 이상	여자	17	단독세대	수도권	35세 미만	남자
8	단독세대	수도권	35세 이상	여자	18	단독세대	수도권	35세 이상	남자
9	아파트	비수도권	35세 이상	남자	19	아파트	비수도권	35세 미만	남자
10	단독세대	비수도권	35세 미만	남자	20	아파트	비수도권	35세 미만	남자

문제 1. P(단독세대, 수도권)

문제 2. P(단독세대)

문제 3. P(수도권 | 단독세대)

문제 4. P(단독세대 | 수도권)

문제 5. P(단독세대 | 수도권, 남자)

문제 6. P(단독세대 | 수도권, 남자, 35 세 미만)

3.5 확률의 곱셈법칙

$P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$ ← 조건부확률 공식으로부터 유도

$P(A \cap B)$는 $P(A, B)$로도 표기함.

연습문제 8.

문제 1. $P(A \cap B) = 0.7$, $P(B) = 0.9$ 일 때, $P(A|B)$

문제 2. $P(A \cap B) = 0.3$, $P(B) = 0.9$ 일 때, $P(A|B)$

문제 3. $P(A \cap B) = 0.5$, $P(B) = 0.9$ 일 때, $P(A|B)$

문제 4. $P(A \cap B) = 0.6$, $P(B) = 0.9$ 일 때, $P(A|B)$

문제 5. $P(A|B) = 0.3$, $P(B) = 0.5$ 일 때, $P(A \cap B)$

문제 6. $P(A|B) = 0.3$, $P(B) = 0.4$ 일 때, $P(A \cap B)$

3.6 복원추출과 비복원추출

Example 복원추출과 비복원추출

주머니에 붉은 공(R) 4개와 파란 공(B) 6개가 있다. 처음에 파란 공이 나오고 다음에 빨간 공이 나올 확률은?

3.6.1 복원추출

복원추출의 경우(공을 꺼낸 뒤 색을 확인하고 다시 넣는다. 그러므로 주머니 안의 공의 수는 항상 일정하다.)

P(첫 시행에서 파란 공을 꺼내고 그리고 다음 시행에서 빨간 공을 꺼낸다.)

$= P(B, R) = P(B) \cdot P(R|B) = \frac{6}{10} \cdot \frac{4}{10} = 0.24$

→ 복원추출의 경우, $P(R|B) = P(R)$이 성립하며, $P(R|R) = P(R)$도 성립한다.

→ 먼저 시행 결과가 다음 시행 결과에 영향을 끼치지 않는다.

→ 두 사건은 서로 독립적이다.

3.6.2 비복원추출

비복원추출의 경우(한 번 꺼낸 공을 다시 주머니에 넣지 않는다. 그러므로 주머니 안의 공의 수는 시행할 때마다 하나씩 줄어든다.)

P(첫 시행에서 파란 공을 꺼내고 그리고 다음 시행에서 빨간 공을 꺼낸다.)

$= P(B,R) = P(B) \cdot P(R|B) = \dfrac{6}{10} \cdot \dfrac{4}{9} = 0.27$

→ 비복원추출의 경우, $P(R) = \frac{4}{10}$, $P(R|B) = \frac{4}{9}$, $P(R|R) = \frac{3}{9}$이다.

→ 먼저 시행 결과가 다음 시행 결과에 영향을 끼친다.

→ 두 사건은 서로 종속적이다.

3.7 독립사건

$P(A \mid B) = P(A), P(B \mid A) = P(B)$ → 표본공간이 변화해도 확률 불변

$P(A,B) = P(A) \cdot P(B|A) = P(A) \cdot P(B) \quad since\ P(B \mid A) = P(B)$

연습문제 9. 단지에 붉은 공(R) 4개와 파란 공(B) 2개가 있다고 하자.

문제 1. 복원 추출을 가정하자. $P(B) = ?$

문제 2. 복원 추출을 가정하자. $P(B \mid R) = ?$
(첫 번째 추출에서 R 이 나온 조건 하에서 두 번째 추출에서 B 가 나올 확률)

문제 3. 비복원 추출을 가정하자. $P(B \mid R) = ?$

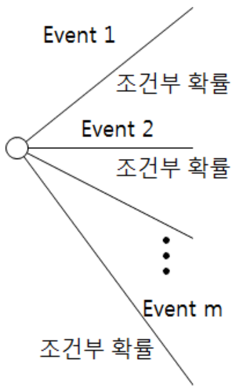
문제 4. 복원 추출에서 $P(B \cap R)$ 값은?
(첫 번째 추출에서 B 가 그리고 두 번째 추출에서 R 이 선택될 확률은?)

문제 5. 비복원 추출에서 $P(B \cap R)$ 값은?
(첫 번째 추출에서 B 가 그리고 두 번째 추출에서 R 이 선택될 확률은?)

문제 6. 비복원 추출의 경우, 먼저 추출된 공과 이후에 추출된 공은 서로 독립적이다.

3.8 확률나무

3.8.1 확률나무 표기법



○은 마디(node), —은 가지(branch)라 한다.

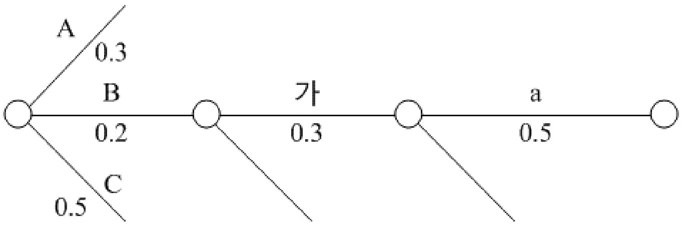
node에서 오른쪽으로 뻗은 branch들은 node에서 발생하는 모든 경우들을 표기한다.

branch 위에는 해당 사건을, 아래에는 해당 사건이 발생할 확률을 기입한다.

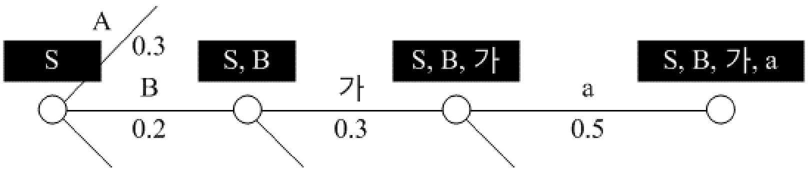
각각의 사건들은 상호배타적이고 완전포괄적이다 (mutually exclusive and collectively exhaustive): 어떤 event도 다른 event와 함께 발생할 수 없고, 노드에서 발생하는 어떤 event도 반드시 event 1부터 event m 중 하나에 속한다.

node에 속한 조건부 확률의 합은 항상 1이다.

확률나무의 예

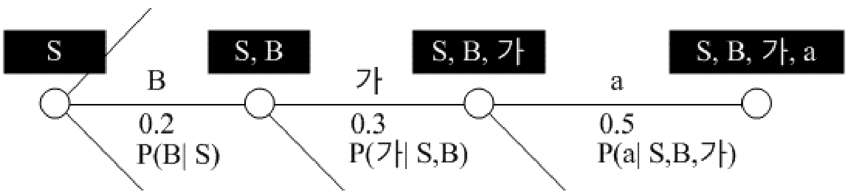


① 결합사건



첫 번째 마디는 전체를 의미하고, 두 번째 마디는 전체에서 B를 거친 단계를, 마지막 마디는 전체에서 B, 가, a를 거친 단계를 의미한다.

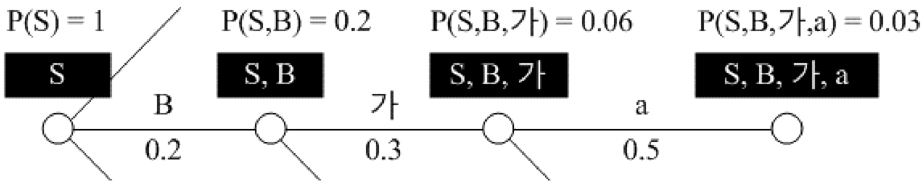
② 조건부확률



첫 번째 조건부 확률은 전체에서 B로 갈 확률(S에서 B로 갈 확률)을, 두 번째 조건부 확률은 전체에서 B를 거친 단계에서 가로 갈 확률(S,B에서 가로 갈 확률)을, 세 번째

조건부 확률은 전체에서 B, 가를 거친 단계에서 a로 갈 확률(S,B,가에서 a로 갈 확률)을 의미한다.

③ 결합확률



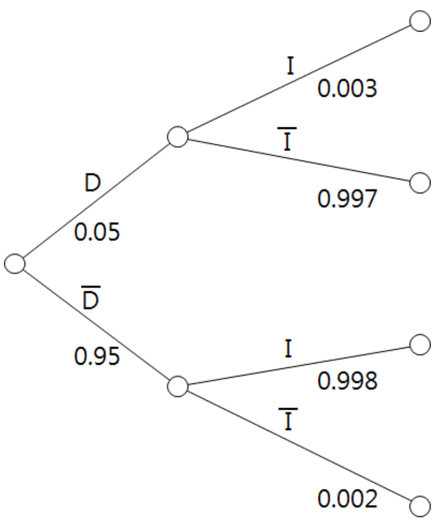
확률의 곱셈법칙에 따라 결합확률은 다음과 같이 계산된다.

$P(S, B) = P(B \mid S) \times P(S) = 0.2$

$P(S, B, 가) = P(가 \mid S, B) \times P(S, B) = 0.06$

$P(S, B, 가, a) = P(a \mid S, B, 가) \times P(S, B, 가) = 0.03$

3.8.2 확률나무 계산의 예



$P(D,I) = P(D) \times P(I \mid D) = 0.05 \times 0.003$

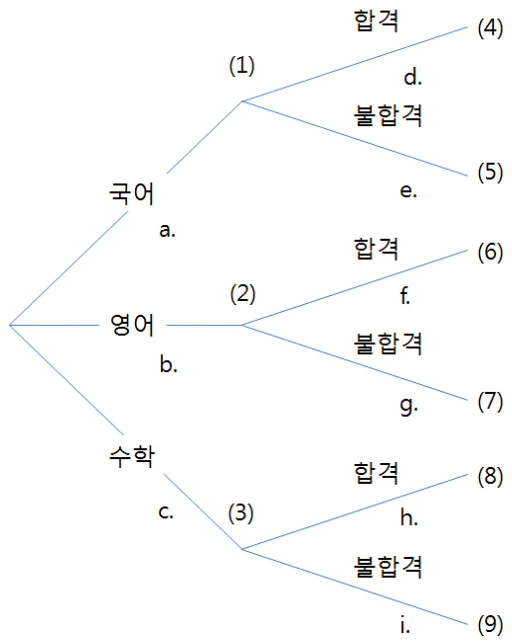
$P(D,\bar{I}) = P(D) \times P(\bar{I} \mid D) = 0.05 \times 0.997$

$P(\bar{D},I) = P(\bar{D}) \times P(I \mid \bar{D}) = 0.95 \times 0.998$

$P(\bar{D},\bar{I}) = P(\bar{D}) \times P(\bar{I} \mid \bar{D}) = 0.95 \times 0.002$

연습문제 10. 어떤 테스트 결과는 다음과 같다. 확률나무에서 알파벳은 조건부확률, (숫자)는 결합확률을 의미한다.

	합격	불합격
국어	15	25
영어	10	15
수학	20	15



문제 1. a.의 값은?

문제 2. (1)의 값은?

문제 3. d.의 값은?

문제 4. (4)의 값은?

문제 5. 다음 중 합이 1.0 이 아닌 것은 모두 표기하시오.
① a + b + c ② (1) + (2) + (3) ③ f + g ④ (4) + (5)

문제 6. 전체를 (1), (2), (3)으로 표현하시오. (결합확률을 결합확률로 표기)

문제 7. (1), (2), (3)을 (4), (5), ..., (9)로 표기 (결합확률을 결합확률로 표기)

문제 8. a, b, c 를 (1), (2), (3)으로 표현하시오. (결합확률로 조건부확률 표기)

문제 9. d, e, f, g, h, i 를 (4), (5), ..., (9)로 표현하시오. (결합확률로 조건부확률 표기)

문제 10. (1), (2), (3)을 a, b, c 로 표현하시오. (조건부확률로 결합확률 표기)

문제 11. (4), (5), ..., (9)를 a, b, ..., i 로 표현하시오. (조건부확률로 결합확률 표기)

4. 베이즈정리

영국의 Bayes(1702~1761)-관찰된 표본자료로부터 그 자료가 속한 모집단을 어떻게 추론할 것인가를 연구

4.1 베이즈 정리의 Example

어떤 회사 제품의 불량률은 1%. 어떤 검사 기법을 사용하여 불량 여부 판정. 그 검사 기법으로 불량인 제품을 불합격 판정할 확률은 0.98이고, 정상 제품을 불합격 판정할 확률은 0.04이다. 판정 결과가 합격일 때, 실제 그 제품이 불량일 확률은?

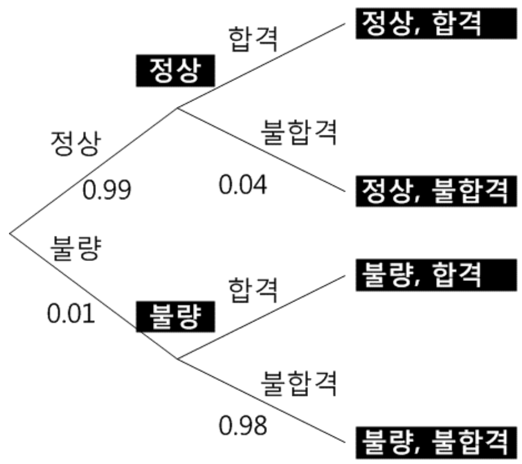
$P(\text{불량}) = 0.01, P(\text{정상}) = 0.99$ → prior probability

$P(\text{불합격} \mid \text{불량}) = 0.98$ ... 불량 제품을 불합격 판정하는 조건부 확률 → likelihood

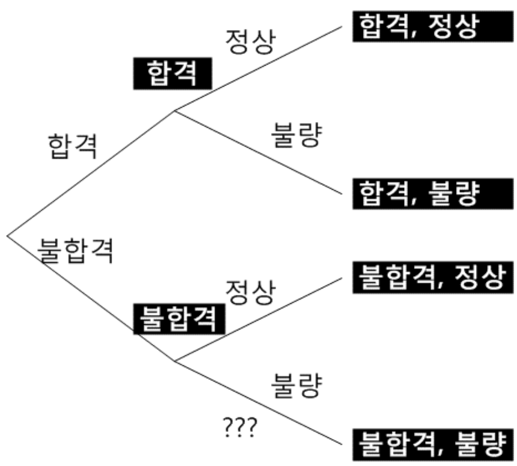
$P(\text{불합격} \mid \text{정상}) = 0.04$ ... 정상 제품을 불합격 판정하는 조건부 확률

Q: 불합격 판정받은 제품이 실제 불량일 확률, $P(\text{불량} \mid \text{합격})$? → posterior probability

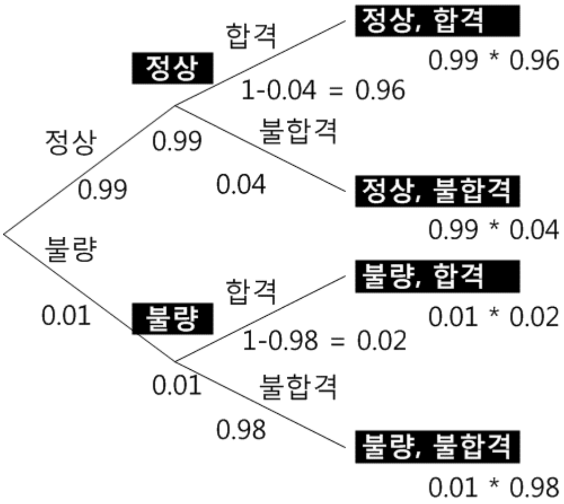
4.1.1 주어진 정보



4.1.2 도출해야 할 정보



4.1.3 도출과정



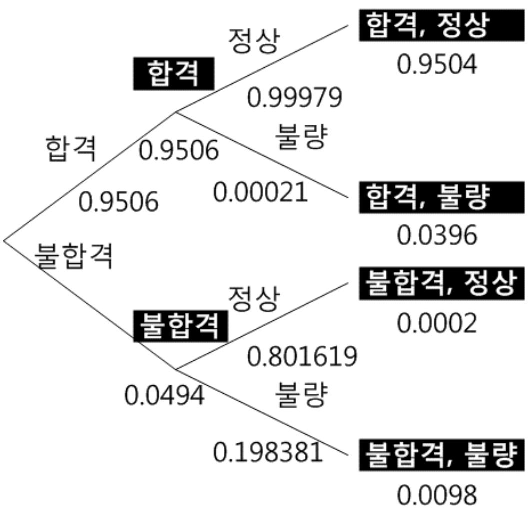
	합격	불합격	합계
정상	0.9504	0.0396	0.9900
불량	0.0002	0.0098	0.0100
합계	0.9506	0.0494	1.0000

$P(\text{정상}|\text{합격}) = 0.9504/0.9506 = 0.99979$

$P(\text{불량}|\text{합격}) = 0.0002/0.9506 = 0.00021$

$P(\text{정상}|\text{불합격}) = 0.0396/0.0494 = 0.801619$

$P(\text{불량}|\text{불합격}) = 0.0098/0.0494 = 0.198381$

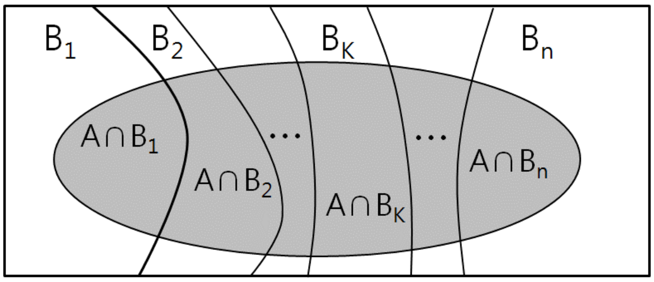


$$P(\text{불량}|\text{합격}) = \frac{P(\text{합격} \cap \text{불량})}{P(\text{합격})}$$

$$= \frac{P(\text{합격} \cap \text{불량})}{P(\text{합격} \cap \text{불량}) + P(\text{합격} \cap \text{정상})}$$

$$= \frac{P(\text{불량}) \cdot P(\text{합격}|\text{불량})}{P(\text{불량}) \cdot P(\text{합격}|\text{불량}) + P(\text{정상}) \cdot P(\text{합격}|\text{정상})}$$

4.2 베이즈 정리(Bayes' Theorem) 공식



주어진 정보 $P(B_1), P(B_2), \cdots, P(B_n)$

$P(A|B_1), P(A|B_2), \cdots, P(A|B_n)$

도출하고 싶은 정보 $P(B_k|A)$

$P(B_k|A) = \dfrac{P(B_k \cap A)}{P(A)}$ 이므로, $P(B_k \cap A)$와 $P(A)$ 값을 구하면 된다.

주어진 정보로부터 $P(B_k \cap A)$와 $P(A)$ 구하는 법

$$P(B_k \cap A) = P(B_k) \cdot P(A|B_k)$$

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) + \cdots + P(A \cap B_n) \\ &= P(B_1) \cdot P(A|B_1) + P(B_2) \cdot P(A|B_2) + \cdots + P(B_n) \cdot P(A|B_n) \\ &= \sum_{i=1}^{n} P(B_i) \cdot P(A|B_i) \end{aligned}$$

정리하면,

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

연습문제 11. 동일한 형태의 상자들이 3개 있는 데, 2개는 흰 색이고 1개는 검은 색이다. 모든 흰 상자에는 빨간 볼펜이 3자루씩, 파란 볼펜이 2자루씩 들

어 있고, 검은 상자에는 빨간 볼펜이 1자루, 파란 볼펜이 4자루 들어 있다.

문제를 풀기 전에 위의 정보를 사용하여 아래 표를 완성한다.

	빨간 볼펜	파란 볼펜	합계
흰 상자	2/3×3/5 =6/15	2/3×2/5 =4/15	2/3
검은 상자	1/3×1/5 =1/15	1/3×4/5 =4/15	1/3
합계	7/15	8/15	1.00

문제 1. 흰 상자를 택한다면 파란 볼펜을 꺼낼 확률은? $P(\text{파란 볼펜} \mid \text{흰 상자})$

문제 2. 상자를 하나 택하여 볼펜을 하나 꺼내 보는 실험을 반복한다. 이때 파란 색 볼펜이 선택될 확률은?

문제 3. 상자 하나를 택하여 볼펜을 하나 꺼내 보니 파란 색 볼펜이었다. 이때 볼펜을 꺼낸 상자가 흰 색 상자일 확률은?

연습문제 정답

1. (1) $P(A) = \frac{n}{N}$, (2) $P(B) = \frac{n_b}{N}$, (3) 1/6, (4) 1/3, (5) $0 \leq P(A) \leq 1$

2. (1) $P(A) = \lim_{n\to\infty} \frac{r}{n}$, (2) False

3. (1) 3 이상의 값이 나오는, (2) {1, 2}, (3) 1/3, (4) 2/3, (5) 1.0,
   (6) {2, 4, 6}, (7) 1/2, (8) {2}, (9) 1/6, (10) {1, 2, 4, 6}, (11) 2/3

4. (1) 남자이면서 찬성인 사람의 수 (2) Ans. 여자의 수(찬성, 반대 불문)
   (3) 전체 인원(남녀, 찬성, 반대 불문)
   (4)

	국산	외제	합계
소형	10	15	25
중형	40	10	50
대형	20	5	25
합계	70	30	100

5. (1) 전체 인원 중 임의로 한 명을 선택했을 때, 그 사람이 남자이면서 찬성일 확률 (결합확률)
   (2) 전체 인원 중 임의로 한 명을 선택했을 때, 그 사람이 여자일 확률 (한계확률)
   (3) 여자이면서 찬성일 확률 + 여자이면서 반대일 확률

6. (1) $\mathrm{P(B_1)} = 0.45 \leftarrow \frac{450}{1{,}000}$
   Note. 표본공간 S에서의 사건 $\mathrm{B_1}$에 대한 확률을 자세히 표기하면 $\mathrm{P(B_1|S)}$이다. 사건 $\mathrm{A_1, A_2, B_1, B_2}$ 모두 표본공간 S에서 정의된 사건들이므로 $\mathrm{P(B_1|S)}$는 간략히 $\mathrm{P(B_1)}$으로 표기한다.

   $\mathrm{P(B_1|S)}$를 조건부확률 공식으로 풀면 $\mathrm{P(B_1|S)} = \dfrac{\mathrm{P(B_1 \cap S)}}{\mathrm{P(S)}}$ 인데,

   $\dfrac{\mathrm{P(B_1 \cap S)}}{\mathrm{P(S)}} = \mathrm{P(B_1)}$이고 $\mathrm{P(S)} = 1$이므로 $\mathrm{P(B_1|S)} = \mathrm{P(B_1)}$가 된다.

	찬성(B1)	반대(B2)	합계
남(A1)	0.25	0.45	0.70
여(A2)	0.20	0.10	0.30
합계	0.45	0.55	1.00

   굵은 테두리는 표본공간을, 굵은 숫자는 사건의 확률을 의미한다.
   (2) $P(B_1|A_1) = \frac{P(B_1 \cap A_1)}{P(A_1)} = \frac{0.25}{0.70}$

   Note. $P(B_1|A_1)$은 사건 B_1에 대한 확률이지만 표본공간이 S에서 A_1으로 수정된다.
   표본공간이 변경되었으므로, 사건 B_1은 표본공간 A_1에 속한 B_1만 고려한다.
   → 찬성이면서 남자만 고려하지, 찬성이면서 여자는 고려치 않는다.
   → 찬성이면서 남자라는 사건은 $B_1 \cap A_1$이다.

	찬성(B_1)	반대(B_2)	합계
남(A_1)	**0.25**	0.45	0.70
여(A_2)	0.20	0.10	0.30
합계	0.45	0.55	1.00

   여기서 굵은 테두리는 표본공간, 굵은 숫자는 사건 확률을 의미한다.

   (3) $P(A_1|B_1) = \frac{P(A_1 \cap B_1)}{P(B_1)} = \frac{0.25}{0.45}$

   Note. $P(A_1|B_1)$은 사건 A_1에 대한 확률이지만 표본공간이 S에서 B_1으로 수정된다.

	찬성(B_1)	반대(B_2)	합계

남(A_1)	**0.25**	0.45	0.70
여(A_2)	0.20	0.10	0.30
합계	0.45	0.55	1.00

7. (1) $5/20$ = 0.25, (2) $8/20$ = 0.40, (3) $4/8$ = 0.50, (4) $4/11$ = 0.3636
   (5) $3/8$ = 0.375, (6) $2/5$ = 0.40

8. (1) 0.7/0.9 (2) 0.3/0.9 (3) 0.5/0.9 (4) 0.6/0.9 (5) 0.3×0.5 (6) 0.3×0.4

9. (1) 2/6, (2) 2/6, (3) 2/5, (4) $P(B) \cdot P(R|B) = P(B) \cdot P(R) = \frac{2}{6} \cdot \frac{4}{6} = \frac{2}{9}$,
   (5) $P(B) \cdot P(R|B) = \frac{2}{6} \cdot \frac{4}{5} = \frac{4}{15}$, (6) False

10. (1) 0.40 전체 중 국어의 비율 -조건부 (2) 0.40 전체 중 국어의 비율 – 결합
    (3) 15/40 = 3/8 국어 중 합격 비율 –조건부
    (4) 0 .15 전체 중 국어이면서 합격의 비율 –결합
    (5) 4 → 전체 중 국어의 비율과 같다.
    (6) 전체 = (1)+(2)+(3), P(전체) = P(국어)+P(영어)+P(수학)
    (7) (1) = (4)+(5), (2) = (6)+(7), (3) = (8)+(9), P(국어) = P(국어, 합격)+P(국어, 불합격),
    P(영어) = P(영어, 합격)+P(영어, 불합격),
    P(수학) = P(수학, 합격)+P(수학, 불합격)
    (8) $a = \frac{(1)}{(1)+(2)+(3)} = (1)$ $\because$ (1)+(2)+(3) = 1, $b = \frac{(2)}{(1)+(2)+(3)} = (2)$,
    $c = \frac{(3)}{(1)+(2)+(3)} = (3)$,
    (9) $d = \frac{(4)}{(4)+(5)}$, $e = \frac{(5)}{(4)+(5)}$, $f = \frac{(6)}{(6)+(7)}$, $g = \frac{(7)}{(6)+(7)}$, ..., $i = \frac{(9)}{(8)+(9)}$,
    (10) (1) = $a \cdot 1 = a$, (2) = $b \cdot 1 = b$, (3) = $c \cdot 1 = c$,
    (11) (4) = (1)$\cdot d = ad$, (5) = (1)$\cdot e = ae$, (6) = (2)$f = bf$, (7) = (2)$g = bg$, (8) = (3)$h = ch$,
    (9) = (3)$i = ci$

11. (1) P(파란 볼펜 | 흰 상자) = (4/15)/(2/3) = 2/5, (2) P(파란 볼펜) = 8/15,
    (3) P(흰 상자 | 파란 볼펜) = (4/15)/(8/15) = 1/2