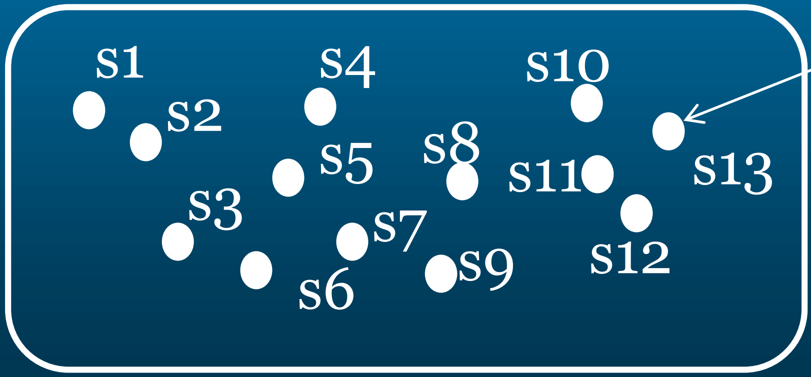
실험(experiment)과 표본 공간(sample space)

- 실험이란 잘 정의된 결과들을 산출하는 과정을 말한다.
- 실험에서 표본공간이란 발생 가능한 모든 실험 결과들의 집합을 말한다.
- 실험결과는 표본점이라고도 한다.

실험 ⟶ 표본공간

동전던지기: 앞면, 뒷면
불량품검사: 불량품 개수
주사위 굴리기: 1,2,3,4,5,6
축구경기: 승리, 패배, 무승부



표본점

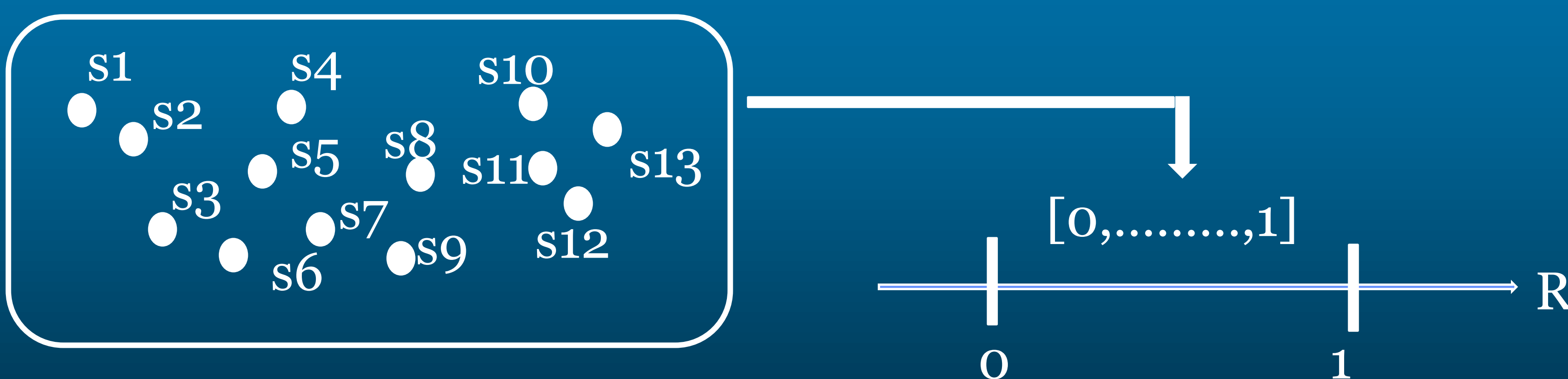
확률(probability)

- 확률은 특정 사건이 발생할 가능성에 대한 수치적 척도이다.
- 확률값은 항상 0부터 1사이의 값을 가진다.

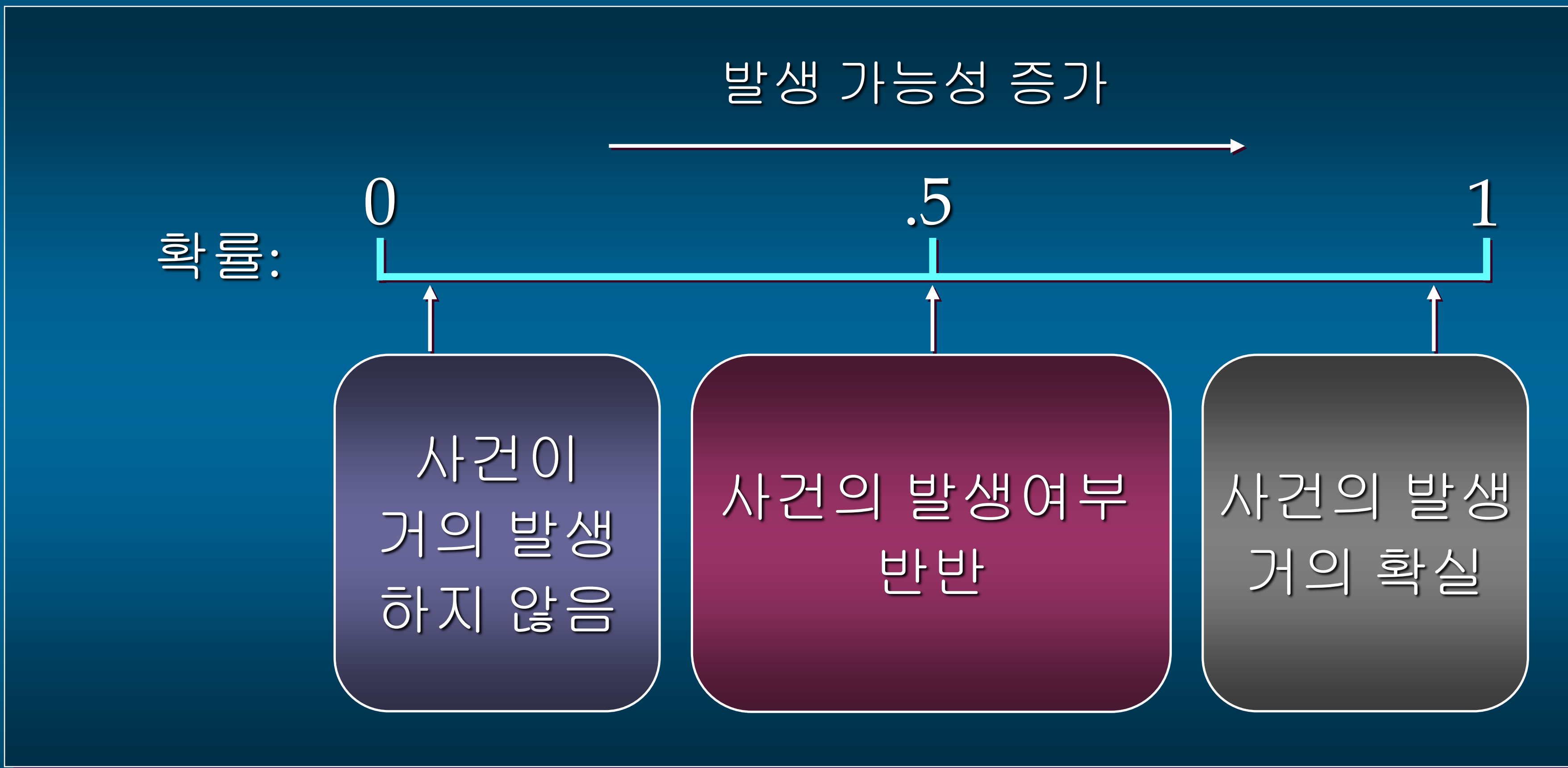
표본점
사건

표본공간

확률P: 표본공간의 all elements 집합→[0,1]사이의 실수로 mapping해주는 함수



사건의 발생 가능성에 대한 척도로서의 확률



예: Bradley Investments

Bradley 는 Markley Oil 과 Collins Mining의 주식에 투자를 하였다. Bradley 는 이들 투자로부터 향후 3개월 동안 가능한 결과들(outcomes)이 아래와 같을 것이라고 결정하였다.

3개월간 투자수익 ($000)

Markley Oil	Collins Mining
10	8
5	–2
0	
–20	

다단계 실험(multi-step experiments)에 대한 계산법칙

- 만약 실험이 k 단계로 구성되어 있으며, 첫 번째 단계에서 n_1 개 가능한 결과가 있고, 두 번째 단계에서 n_2 개 가능한 결과가 있고, 등등 이면,
  이 실험 결과의 총 개수는 $(n_1)(n_2) \ldots (n_k)$가 된다.

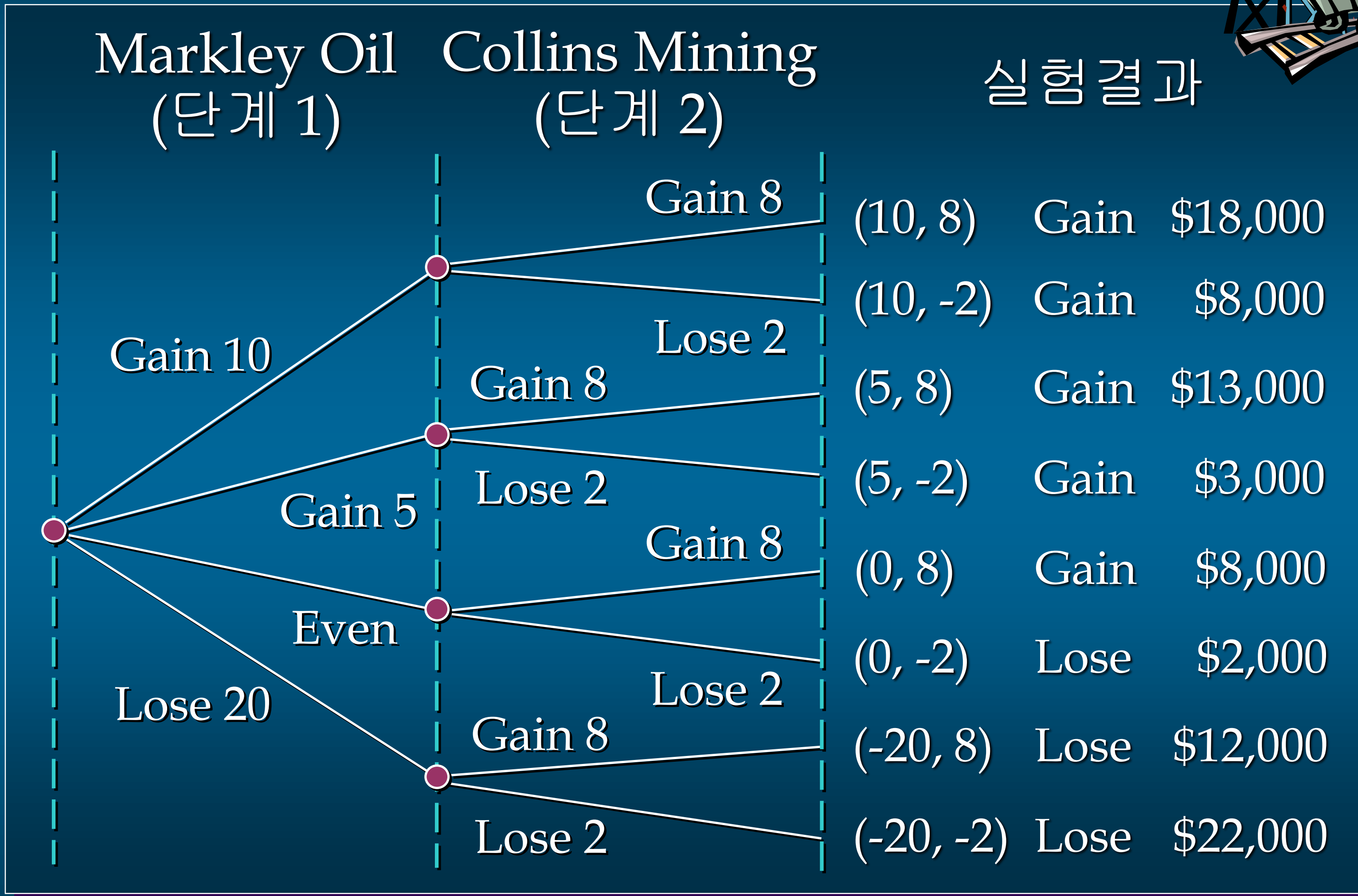
- 다단계 실험 결과를 파악하는데 도움이 되는 그래프가 계통도(tree diagram)이다.

다단계 실험에 대한 계산법칙

Bradley의 투자는 2단계 실험으로 볼 수 있다. 두 종목의 주식으로 되어 있으며, 각 종목은 각각 실험 결과의 집합을 가지고 있다.

Markley Oil:	$n_1 = 4$
Collins Mining:	$n_2 = 2$
총 실험결과 수:	$n_1 n_2 = (4)(2) = 8$

계통도(tree diagram)



조합(combinations)에 관한 계산법칙

두 번째 유용한 계산법칙은 N개의 개체 중에서 n개의 개체를 뽑는 실험결과의 개수를 알 수 있게 해준다.

N개의 개체 중에서 n개의 개체를 한 번에 뽑을 경우 조합의 수:

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

여기서 : $N! = N(N-1)(N-2)\ldots(2)(1)$

$n! = n(n-1)(n-2)\ldots(2)(1)$

$0! = 1$

예) 품질관리부서에서 불량품 여부를 판정하기 위해 제품 5개 중에서 2개를 무작위로 뽑아 검사한다고 할 때, 5개 중에서 2개를 취하는 경우의 수: 5!/(2!3!)=120/12=10

순열(permutations)에 관한 계산법칙

- 세 번째 유용하게 사용되는 계산법칙은 N개의 개체 중에서 순서를 고려하여 n개의 개체를 뽑는 실험결과의 개수를 알 수 있게 해준다.

N개의 개체 중에서 n개의 개체를 한 번에 뽑을 경우 순열의 수:

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!}$$

여기서 : $N! = N(N-1)(N-2)\ldots(2)(1)$

$n! = n(n-1)(n-2)\ldots(2)(1)$

$0! = 1$

예) 앞서 품질관리부서의 예에서 부품 5개 중에서 2개를 순서를 고려해서 검사한다고 할 때, 5개 중에서 2개를 검사하는 순열의 수: 5!/3!=120/6=20

확률의 부여

■ 확률 부여를 위한 기본 요건

1. 각 실험결과에 부여되는 확률은 0과 1사이의 있어야 한다.

모든 i에 대해 $0 \le P(E_i) \le 1$

여기서,
E_i는 i번째 실험결과,
$P(E_i)$는 그 확률을 말한다.

확률의 부여

2. 모든 실험에 대한 확률의 합은 1이 되어야 한다.

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1$$

여기서, n 은 실험결과의 수

확률 부여 방법

고전적 방법

결과의 발생확률이 같다는 가정하에 확률을 부여

상대도수 방법

실험이나 역사적 자료에 기초하여 확률을 부여

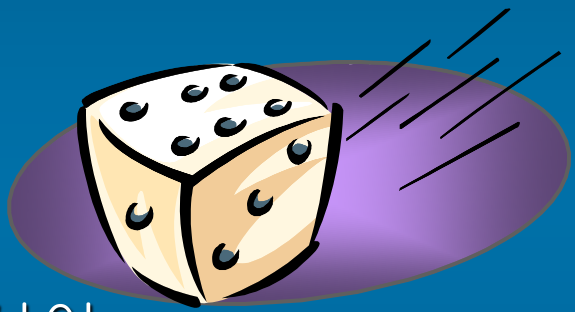
주관적 방법

주관적 판단에 기초하여 확률을 부여

고전적(classic) 방법

만약 n 개의 가능한 실험결과가 있을 때, 각 결과에 $1/n$의 확률을 부여한다.

예:



실험: 주사위 던지기

표본 공간: $S = \{1, 2, 3, 4, 5, 6\}$

확률: 각 표본점은 1/6의 발생가능성이 있다.

상대도수(relative frequency) 방법

- 예 : Lucas Tool Rental(공구 대여)

  Lucas Tool Rental 은 하루에 대여되는 자동차 광택기의 대수에 대하여 확률을 부여하려고 한다. 지난 40일간 일 대여 대수의 기록이 아래와 같다.

대여 대수	대여 일수
0	4
1	6
2	18
3	10
4	2

상대도수 방법

각 부여된 확률은 총도수(총 일수)로
해당 도수 (해당 일수)를 나눈 값이다.

대여 대수	일수	확률
0	4	.10
1	6	.15
2	18	.45
3	10	.25
4	2	.05
	40	1.00

4/40

주관적(subjective) 방법

- 경제적 상황이나 기업환경이 빠르게 변할 때 확률 부여를 역사적 자료에만 의존하는 것은 적절하지 못하다.

- 경험이나 직관과 같은 다른 이용 가능한 자료도 사용할 수 있다. 그러나 확률값은 궁극적으로 실험의 결과가 일어날 것이라는 믿음의 정도를 표시해야 한다.

- 가장 좋은 확률 추정값은 고전적 방법이나 상대도수 방법에 의한 추정값에 주관적 방법의 추정값을 결합하여 구해질 수 있다.

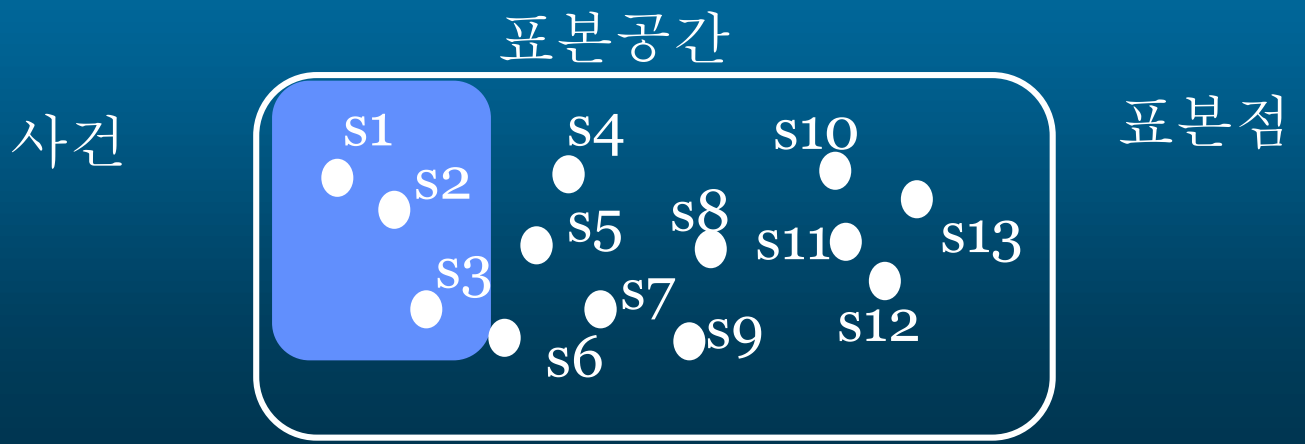
주관적 방법

한 분석가가 주관적 방법을 적용하여 아래와 같이 확률을 부여하였다.

실험결과	순 손익	확률
(10, 8)	$18,000 Gain	.20
(10, –2)	$8,000 Gain	.08
(5, 8)	$13,000 Gain	.16
(5, –2)	$3,000 Gain	.26
(0, 8)	$8,000 Gain	.10
(0, –2)	$2,000 Loss	.12
(–20, 8)	$12,000 Loss	.02
(–20, –2)	$22,000 Loss	.06

사건(events)과 확률

- 사건은 표본점들의 집합이다.
- 사건의 확률은 그 사건에 속하는 표본점들의 확률의 합과 같다.
- 실험의 모든 표본점들과 그들의 확률을 알 수 있다면 사건의 확률을 계산할 수 있다.



사건과 확률

사건 M = Markley Oil 양(+)의 수익

$$M = \{(10, 8), (10, -2), (5, 8), (5, -2)\}$$

$$P(M) = P(10, 8) + P(10, -2) + P(5, 8) + P(5, -2)$$

$$= .20 + .08 + .16 + .26$$

$$= .70$$

확률의 기본 관계

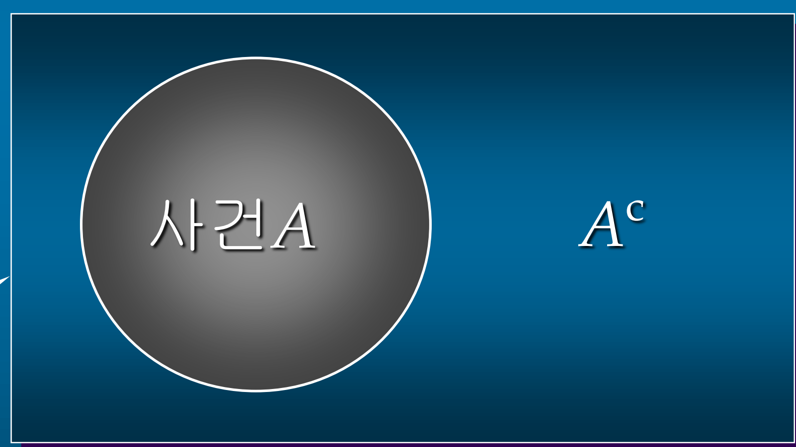
모든 표본점들의 확률을 모르더라도 사건의 확률을 계산하기 위해 사용될 수 있는 확률의 기본적인 관계가 존재 한다.

- 사건의 여집합
- 두 사건의 합집합
- 두 사건의 교집합
- 상호 배타(반)적 사건

사건의 여집합(complement)

A의 여집합이란 사건 A에 속하지 않은 표본점들로 구성된 사건을 말한다.

A의 여집합은 A^c로 표시된다.

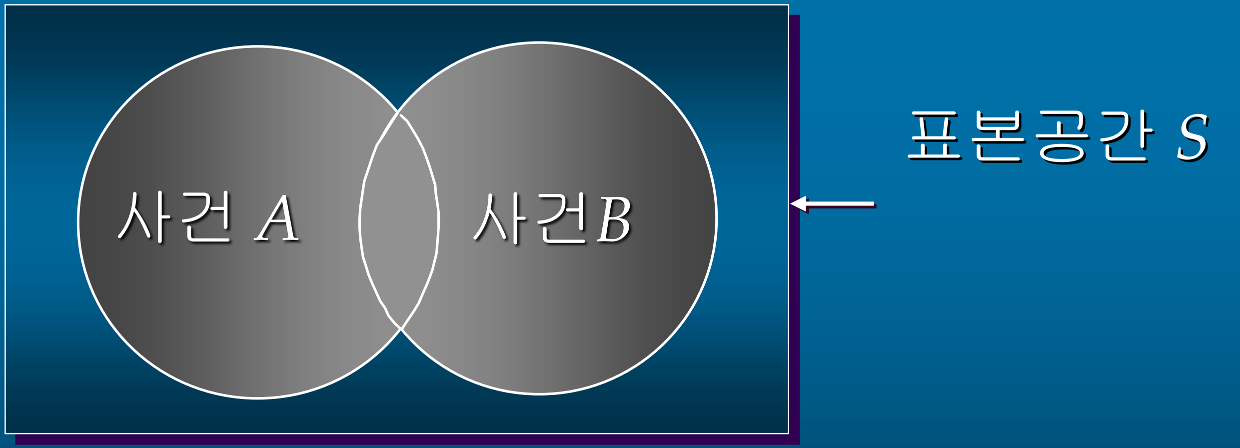


표본공간 S

벤 다이어그램

두 사건의 합집합(union)

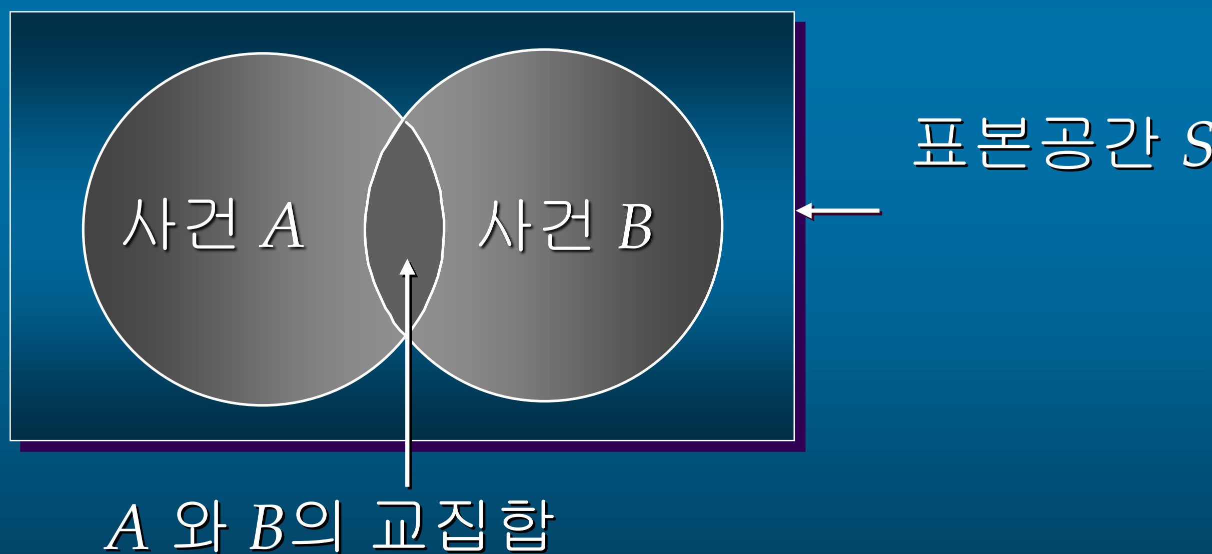
- 사건 A 와 B의 합집합은 A 나 B 또는 양쪽 모두에 속하는 모든 표본점들을 포함하고 있는 사건이다.
- A와 B 의 합집합은 $A \cup B$로 표시된다.



두 사건의 교집합(intersection)

A와 **B**의 교집합은 **A**와 **B** 양쪽에 다 속하는 표본점들의 집합이다.

사건 A 와 B의 교집합은 $A \cap B$로 표시된다.



두 사건의 교집합

사건 M = Markley Oil 양(+)의 수익

사건 C = Collins Mining 양(+)의 수익

$M \cap C$ = Markley Oil 그리고 Collins Mining 양(+)의 수익

$$M \cap C = \{(10, 8), (5, 8)\}$$

$$P(M \cap C) = P(10, 8) + P(5, 8)$$

$$= .20 + .16$$

$$= .36$$

덧셈(addition) 법칙

덧셈 법칙은 사건 A 나 B 또는 양쪽 모두가 발생할 확률을 계산할 때 사용하는 방법이다.

덧셈 법칙은 아래와 같다:

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

덧셈 법칙

사건 M = Markley Oil 양(+)의 수익

사건 C = Collins Mining 양(+)의 수익

$M \cup C$ = Markley Oil 또는
Collins Mining 양(+)의 수익

앞에서: $P(M) = .70$, $P(C) = .48$, $P(M \cap C) = .36$

그래서: $P(M \cup C) = P(M) + P(C) - P(M \cap C)$

$= .70 + .48 - .36$

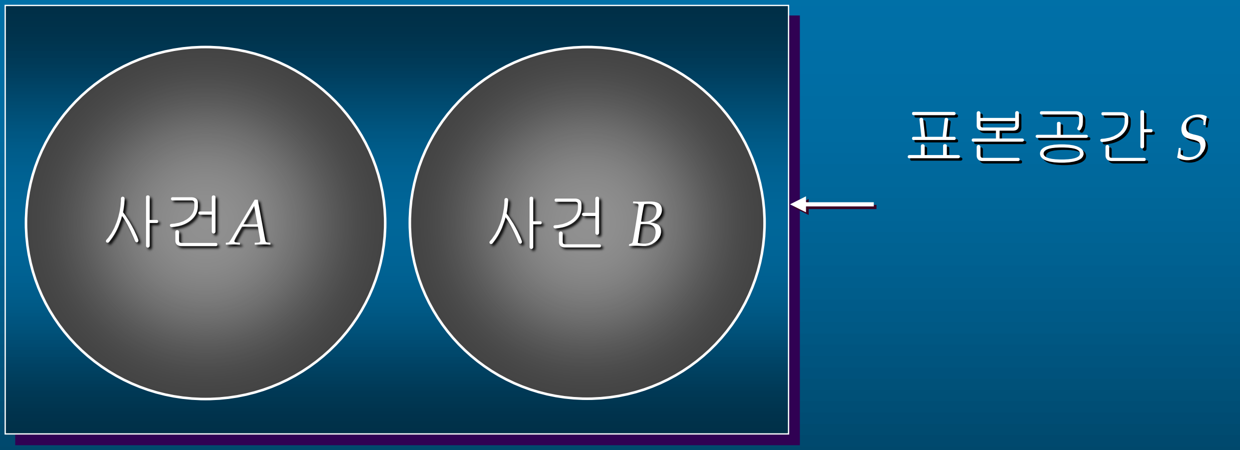
$= .82$

(이 결과는 앞서 한 사건의 확률정의를 이용하여 구한 결과와 같다)

상호배타적(mutually exclusive) 사건

두 사건이 공통으로 가지고 있는 표본점이 하나도 없는 경우 상호배타적 사건이라 한다.

한 사건이 발생하면 다른 사건이 일어날 수 없을 때 이 두 사건을 상호배타적이라고 한다.



상호배타적 사건에 대한 덧셈법칙

사건 A 와 B 가 상호배타적이라면 $P(A \cap B) = 0$

상호배타적 사건에 대한 덧셈 법칙:

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

"$-P(A \cap B)$"을 포함할 필요가 없다.

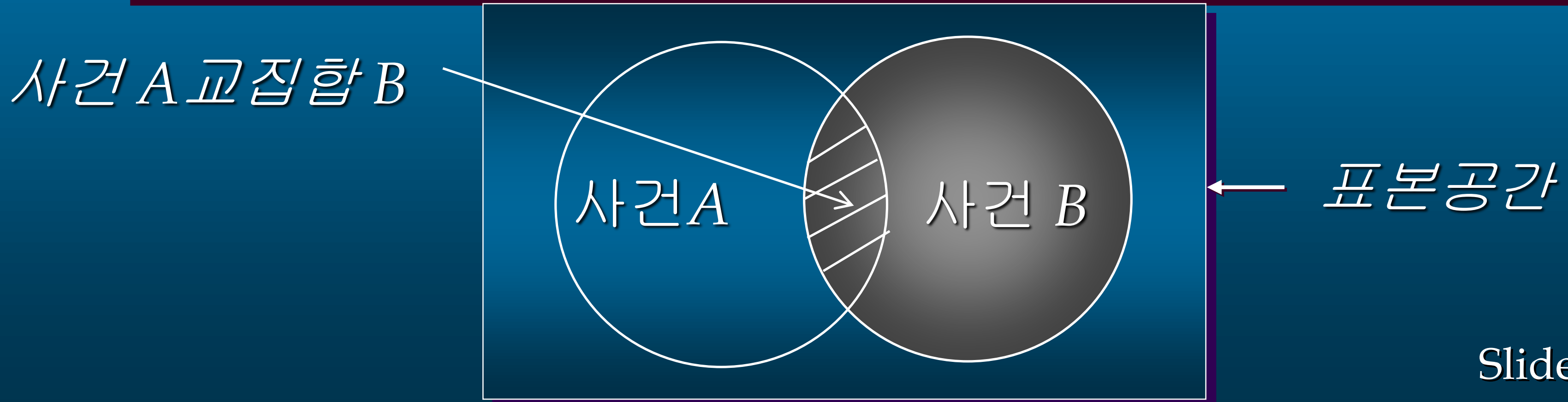
조건부(conditional) 확률

한 사건이 발생했을 때 다른 사건이 발생할 확률을 조건부 확률이라고 한다.

사건 B가 발생했을 때 사건 A의 조건부 확률은 $P(A|B)$로 표시된다.

조건부 확률은 아래와 같이 계산된다

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



조건부 확률

사건 M = Markley Oil 양(+)의 수익

사건 C = Collins Mining 양(+)의 수익

$P(C\,|\,M)$ = Markley Oil 양(+)의 수익 경우에 Collins Mining 양(+)의 수익 확률

앞에서: $P(M \cap C) = .36$, $P(M) = .70$

$$\text{그래서:} P(C\,|\,M) = \frac{P(C \cap M)}{P(M)} = \frac{.36}{.70} = .5143$$

곱셈(multiplication)법칙

곱셈법칙은 두 사건의 곱의 확률을 계산하는데 쓰인다.

곱셈법칙은 아래와 같다:

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

곱셈 법칙

사건 M = Markley Oil 양(+)의 수익

사건 C = Collins Mining 양(+)의 수익

$M \cap C$ = Markley Oil 양(+)의 수익
그리고 Collins Mining 양(+)의 수익

앞에서: $P(M) = .70,\ P(C|M) = .5143$

그래서: $P(M \cap C) = P(M)P(C|M)$

$= (.70)(.5143)$

$= .36$

이 결과는 앞서 한 사건의 확률정의를 이용하여 구한 결과와 같다)

독립(independent) 사건

만약 사건 A의 확률이 사건 B의 발생여부에 의해 변화되지 않을 때, 사건 A와 B는 독립적이다.

사건 A 와 B의 관계가 아래와 같다면 독립적이다:

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

또는

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

독립사건에 대한 곱셈법칙

곱셈법칙은 두 사건이 독립적인지를 판단하는데 사용될 수 있다.

곱셈법칙은 아래와 같다:

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

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

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

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

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

독립사건에 대한 곱셈법칙

사건 M = Markley Oil 양(+)의 수익

사건 C = Collins Mining 양(+)의 수익

사건 M 과 C 는 독립적인가?

$P(M \cap C) = P(M)P(C)$ 인가 ?

앞에서: $P(M \cap C) = .36$, $P(M) = .70$, $P(C) = .48$

$P(M)P(C) = (.70)(.48) = .34$, .36이 아님

그래서: M 과 C 는 독립적이지 않다.

상호배타성 vs. 독립

- 상호배타적 사건과 독립사건의 개념을 혼돈하지 말라
- 발생확률이 0이 아닌 두 사건은 상호배타적이고 동시에 독립적이 사건이 될 수 없다.
- 상호배타적인 경우 한 사건이 일어나면 다른 사건은 일어날 수 없다. 즉, 다른 사건의 발생확률은 0이다 (그러므로 두 사건은 서로 의존적이다)
- 상호배타적이 아닌 경우 두 사건은 독립적이 될 수도 있고 안될 수 도 있다.

베이즈 정리(Bayes' Theorem)

- 종종 확률분석을 할 때 초기 또는 사전확률로 시작한다.
- 그 다음에는 표본이나 특별보고서, 제품 테스트로부터 사건에 대한 추가적인 정보를 얻는다.
- 이러한 새로운 정보에 의해 수정된 사후확률을 계산한다.
- 베이즈 정리는 사전확률을 갱신하는 방법을 제공한다.
- 사전확률 → 새로운 정보 → 베이즈 정리 적용 → 사후확률

베이즈 정리

- 예 : L. S. 원단

현재 계획된 쇼핑센터는
L. S. 원단과 같은 시내기업들에게
강한 경쟁을 제공할 것이다 .
그래서 쇼핑센터가 세워지면,
L. S. 원단 사장은 쇼핑센터로
이전 하기를 바랄 것이다.

만약 시의회에서 지목변경을 해주지 않으면 쇼핑센터를
세울 수 없다. 그래서 추진단이 먼저 의회에 지목변경에
대한 긍정적인 또는 부정적인 의견을 제안하여야 한다.

베이즃 정리

- 사전(prior) 확률

  사건:

  A_1 = 의회가 지목변경을 승인한 경우
  A_2 = 의회가 지목변경을 승인하지 않은 경우

  주관적 판단을 사용하여:

  $P(A_1) = .7, \quad P(A_2) = .3$

베이즈 정리

- 새로운 정보

  추진단이 지목변경에 대한 추천서를 제출하였다. 사건 B 는 추진단이 지목변경에 대한 부정적인 추천을 한 경우를 나타낸다고 하자.

  B 가 일어났을 때, L. S. 원단은 시의회가 지목 변경에 대한 승인 또는 승인하지 않을 확률에 대해 개정을 해야 하는가?

베이즈 정리

- 조건부 확률

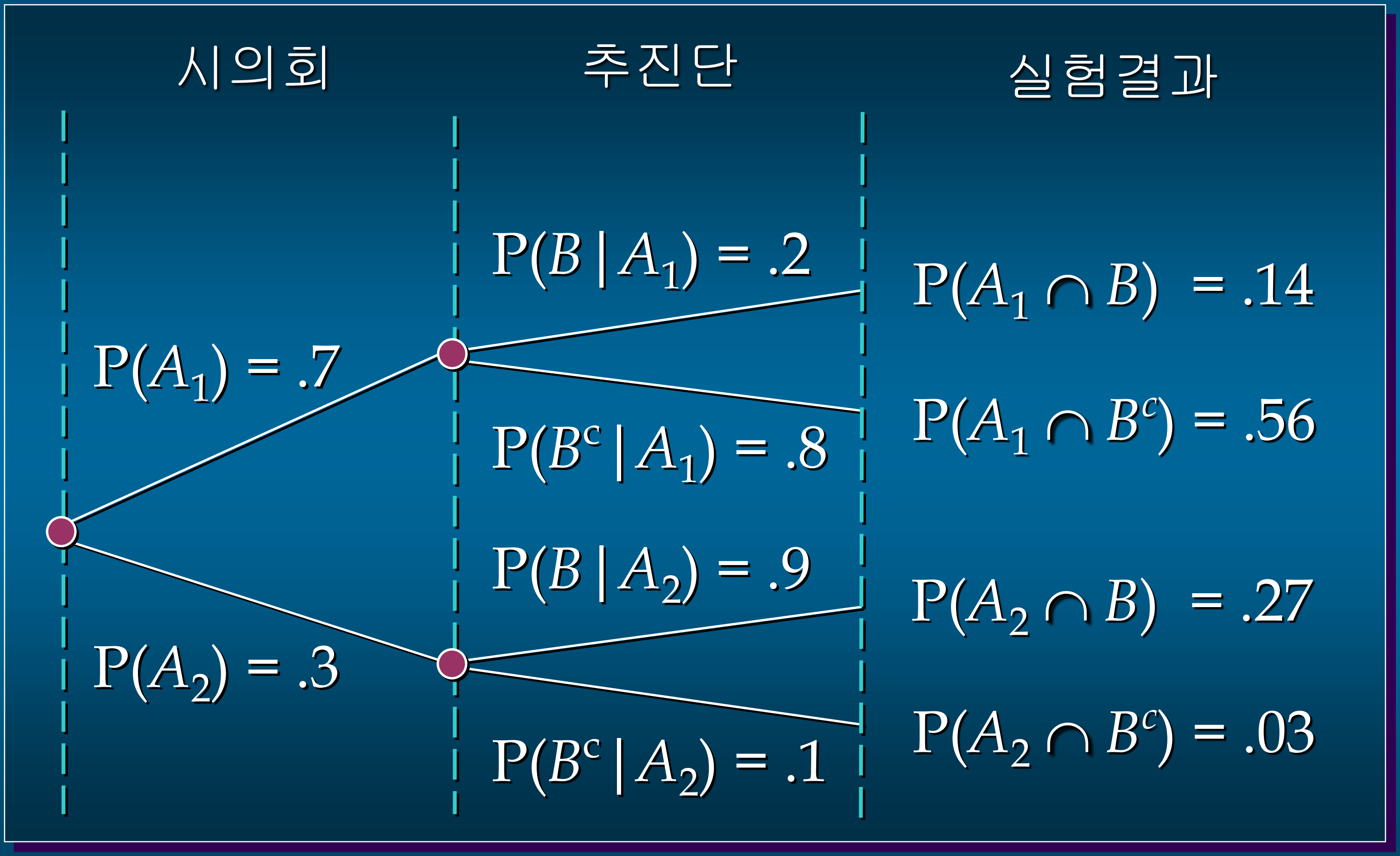
  시의회와 추진단의 과거자료로부터 아래 확률을 알 수 있었다고 하자:

$$P(B \mid A_1) = .2 \qquad P(B \mid A_2) = .9$$

그래서: $$P(B^C \mid A_1) = .8 \qquad P(B^C \mid A_2) = .1$$

베이즈 정리

계통도


시의회
추진단
실험결과
$P(A_1) = .7$
$P(A_2) = .3$
$P(B | A_1) = .2$
$P(B^c | A_1) = .8$
$P(B | A_2) = .9$
$P(B^c | A_2) = .1$
$P(A_1 \cap B) = .14$
$P(A_1 \cap B^c) = .56$
$P(A_2 \cap B) = .27$
$P(A_2 \cap B^c) = .03$

베이즈 정리

- 사건 B 가 일어났을 때 사건 A_i 가 일어날 사후(posterior) 확률을 찾기 위해 베이즈 정리를 적용한다.

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)}$$

- 베이즈 정리는 상호 배타적이며 모든 사건의 합집합이 전체 표본 공간이 되는 사건들의 사후확률을 구할 때 적용된다.

베이즈 정리

- 사후 확률

  추진단이 지목변경을 승인하지 않는 추천서를 제출하였을 때, 아래와 같이 사전확률을 개정할 수 있다:

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{(.7)(.2)}{(.7)(.2) + (.3)(.9)}$$

$$= .34$$

베이즈 정리

- 결론

추진단의 추천은 L. S.원단에게는 좋은 소식이다. 시의회가 지목변경을 할 사후확률은 .34 이다. 반면 사전확률은 .70 이었다.

표에 의한 방법(tabular approach)

- 단계 1

  다음 3개 열을 준비한다:

  제 1열: 사후확률이 필요한 상호배타적 사건

  제 2열: 이 사건들의 사전 확률

  제 3열: 각 사건이 주어진 경우 새로운 정보 B의 조건부 확률

표에 의한 방법

(1)	(2)	(3)	(4)	(5)
사건 A_i	사전확률 $P(A_i)$	조건부 확률 $P(B \mid A_i)$		
A_1	.7	.2		
A_2	.3	.9		
	1.0			

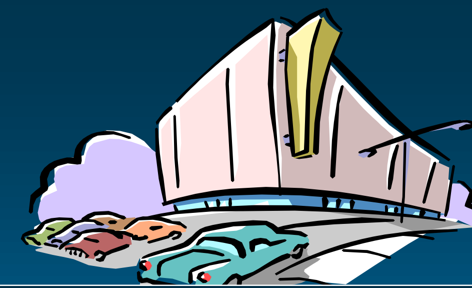
표에 의한 방법

- 단계 2

  제 4열: 곱셈 법칙을 사용하여 각 사건과 새로운 정보 B에 대한 결합확률을 계산한다. 이 결합확률은 제 2열의 사전확률과 제 3열의 조건부 확률을 곱하면 된다.

  그래서, $P(A_i \cap B) = P(A_i)\, P(B \mid A_i)$

표에 의한 방법



(1)	(2)	(3)	(4)	(5)
사건 A_i	사전확률 $P(A_i)$	조건부 확률 $P(B\,\vert\,A_i)$	결합확률 $P(A_i \cap B)$	
A_1	.7	.2	.14	
A_2	.3	.9	.27	
	1.0			

.7 x .2

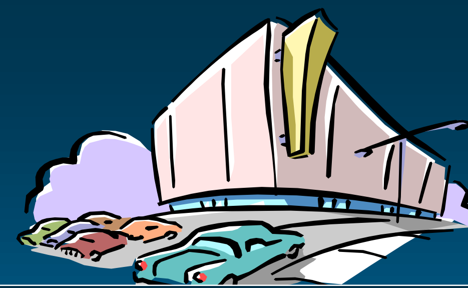
표에 의한 방법

- 단계 3

 제 4열

 결합확률을 합한다. 이 합은 새로운 정보에 대한 확률 $P(B)$이 된다. 즉, (.14 + .27)은 추진단이 부정적 추천을 할 전반적인 확률 .41 을 나타낸다.

표에 의한 방법



(1)	(2)	(3)	(4)	(5)
사건 A_i	사전확률 $P(A_i)$	조건부 확률 $P(B\mid A_i)$	결합확률 $P(A_i \cap B)$	
A_1	.7	.2	.14	
A_2	.3	.9	.27	
	1.0		$P(B)$ = .41	

표에 의한 방법

- 단계 4

  제 5열

  조건부 확률의 기본 관계식을 사용하여 사후 확률을 계산한다.

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)}$$

결합확률 $P(A_i \cap B)$는 제 4열에 있고, 확률 $P(B)$ 은 제 4열의 합이다.

표에 의한 방법

(1)	(2)	(3)	(4)	(5)
사건	사전확률	조건부 확률	결합확률	사후확률
A_i	$P(A_i)$	$P(B \mid A_i)$	$P(A_i \cap B)$	$P(A_i \mid B)$
A_1	.7	.2	.14	.3415
A_2	.3	.9	.27	.6585
	1.0		$P(B) =$ .41	1.0000

.14/ .41

4장 끝