MEC 20005 Summer 2024
경영경제통계

오주희 교수

한동대학교 경영경제학부

E-mail: Jooheeoh@handong.edu

June 27th 2024

가중평균과 그룹화 자료 (The weighted mean and working with grouped data)

- 가중평균
- 그룹화 자료의 평균
- 그룹화 자료의 분산
- 그룹화 자료의 표준편차

가중평균(weighted mean)

- 관찰값의 $\boxed{\text{중요도}}$를 반영한 가중치를 각각의 자료값에 부여하여 평균을 계산할 때, 이러한 평균을 '가중 평균' 이라고 한다.

- $$\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

  여기서:

  $x_i = i$ 번째 관찰값

  $w_i = i$ 번째 관찰값의 가중치

- 관찰값이 중요도에 있어서 서로 다를 때, 분석자는 각 관찰값이 가지는 중요도를 가장 잘 반영할 수 있는 가중치를 선택하여야 한다.

예) 어느 금융자산 포트폴리오가 50%는 주식, 40%는 채권, 10%는 현금으로 구성되어 있다. 주식 수익률은 12%이며 채권 수익률은 7%f고 하자. 포트폴리오의 수익률은?

$R_w = 0.5 \times 0.12 + 0.4 \times 0.07 + 0.1 \times 0 = 0.088$ (8.8%)

그룹화된 자료(grouped data)

- 가중평균 계산법이 그룹화된 자료의 평균, 분산, 표준편차의 대략적인 값을 구하는데 사용된다.
- 가중평균을 계산하기 위해, 각 계급의 중간점을 그 계급의 평균처럼 가정하여 사용한다.
- 계급의 도수를 가중치로 사용하여 계급 중간점들의 가중평균을 계산한다.
- 분산과 표준편차를 계산할 때도 유사한 방법으로 계급의 도수를 가중치로 사용한다.

그룹화 자료의 평균

표본평균

$$\bar{x} = \frac{\sum f_i M_i}{n}$$

모집단 평균

$$\mu = \frac{\sum f_i M_i}{N}$$

여기서:

f_i = i 계급의 (빈)도수

M_i = i 계급의 중간점

그룹화 자료의 표본평균

앞선 예에서 본 70채의 아파트 표본 월세자료가 아래와 같이 도수분포 형식으로 그룹화되어 있다.

Rent ($)	Frequency
420-439	8
440-459	17
460-479	12
480-499	8
500-519	7
520-539	4
540-559	2
560-579	4
580-599	2
600-619	6

그룹화 자료의 표본 평균

Rent ($)	f_i	M_i	f_iM_i
420-439	8	429.5	3436.0
440-459	17	449.5	7641.5
460-479	12	469.5	5634.0
480-499	8	489.5	3916.0
500-519	7	509.5	3566.5
520-539	4	529.5	2118.0
540-559	2	549.5	1099.0
560-579	4	569.5	2278.0
580-599	2	589.5	1179.0
600-619	6	609.5	3657.0
Total	70		34525.0

$$\bar{x} = \frac{34,525}{70} = 493.21$$

이런 근사값은 실제 평균인 $490.80과는 $2.41 정도 차이가 있다.

그룹화된 자료의 분산

- 표본의 경우

$$s^2 = \frac{\sum f_i (M_i - \bar{x})^2}{n-1}$$

- 모집단의 경우

$$\sigma^2 = \frac{\sum f_i (M_i - \mu)^2}{N}$$

그룹화 자료에서 표본분산

Rent ($)	f_i	M_i	$M_i - \bar{x}$	$(M_i - \bar{x})^2$	$f_i(M_i - \bar{x})^2$
420-439	8	429.5	-63.7	4058.96	32471.71
440-459	17	449.5	-43.7	1910.56	32479.59
460-479	12	469.5	-23.7	562.16	6745.97
480-499	8	489.5	-3.7	13.76	110.11
500-519	7	509.5	16.3	265.36	1857.55
520-539	4	529.5	36.3	1316.96	5267.86
540-559	2	549.5	56.3	3168.56	6337.13
560-579	4	569.5	76.3	5820.16	23280.66
580-599	2	589.5	96.3	9271.76	18543.53
600-619	6	609.5	116.3	13523.36	81140.18
Total	70				208234.29

계속 →

그룹화 자료에서 표본 분산

- 표본 분산

$$s^2 = 208{,}234.29/(70-1) = 3{,}017.89$$

- 표본 표준편차

$$s = \sqrt{3{,}017.89} = 54.94$$

이러한 근사값은 실제 표준편차인 $54.74와는 겨우 $.20 정도 차이가 난다.

Sample Mean : HW

In Chapter 3, we learn about how we can calculate weighted averages in grouped data.

Please submit your answers to the following simple question in the LMS assignment section.

(Chapter 3 Q.52) Question1: Calculate the weighted average of x using below data.

x	weight(w)
3.2	6
2.0	3
2.5	2
5.0	8

Answer: 3.69

Sample Mean : HW

(Optional Question: Chapter 3 Q.53)

Class	Mid-point	frequency
3-7	5	4
8-12	10	7
13-17	15	9
18-22	20	5

Question 2: Find the sample mean.
Question 3: Find the sample variance and sample standard deviation.

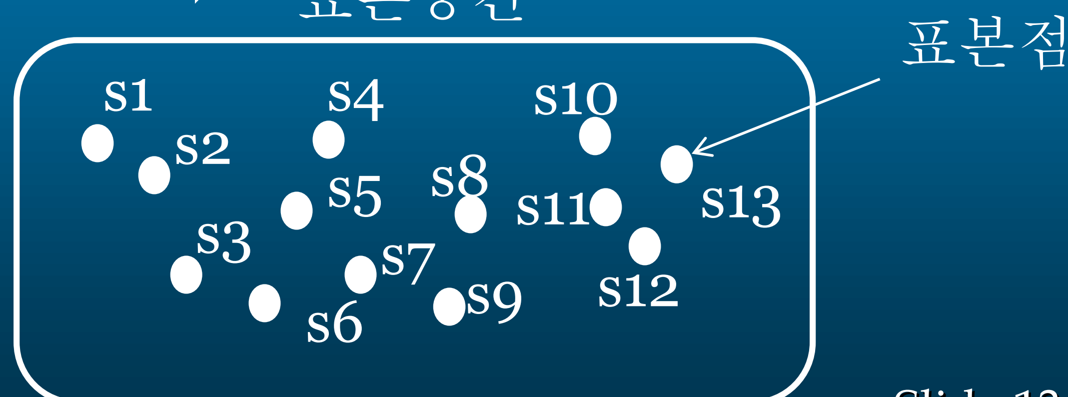
Q2: 13 (sample mean)
Q3: 25 (sample variance), 5 (standard deviation)

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



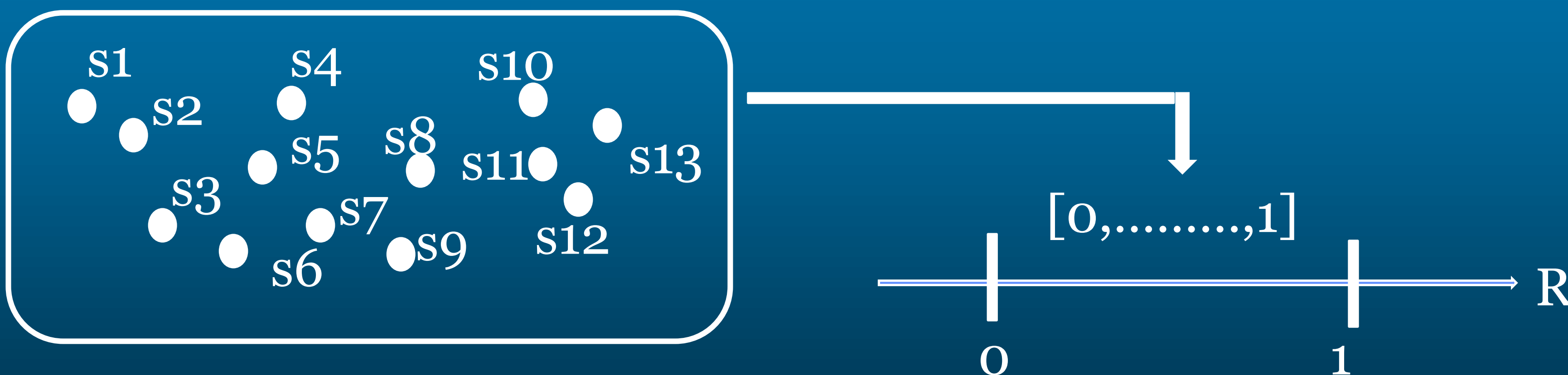
확률(probability)

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

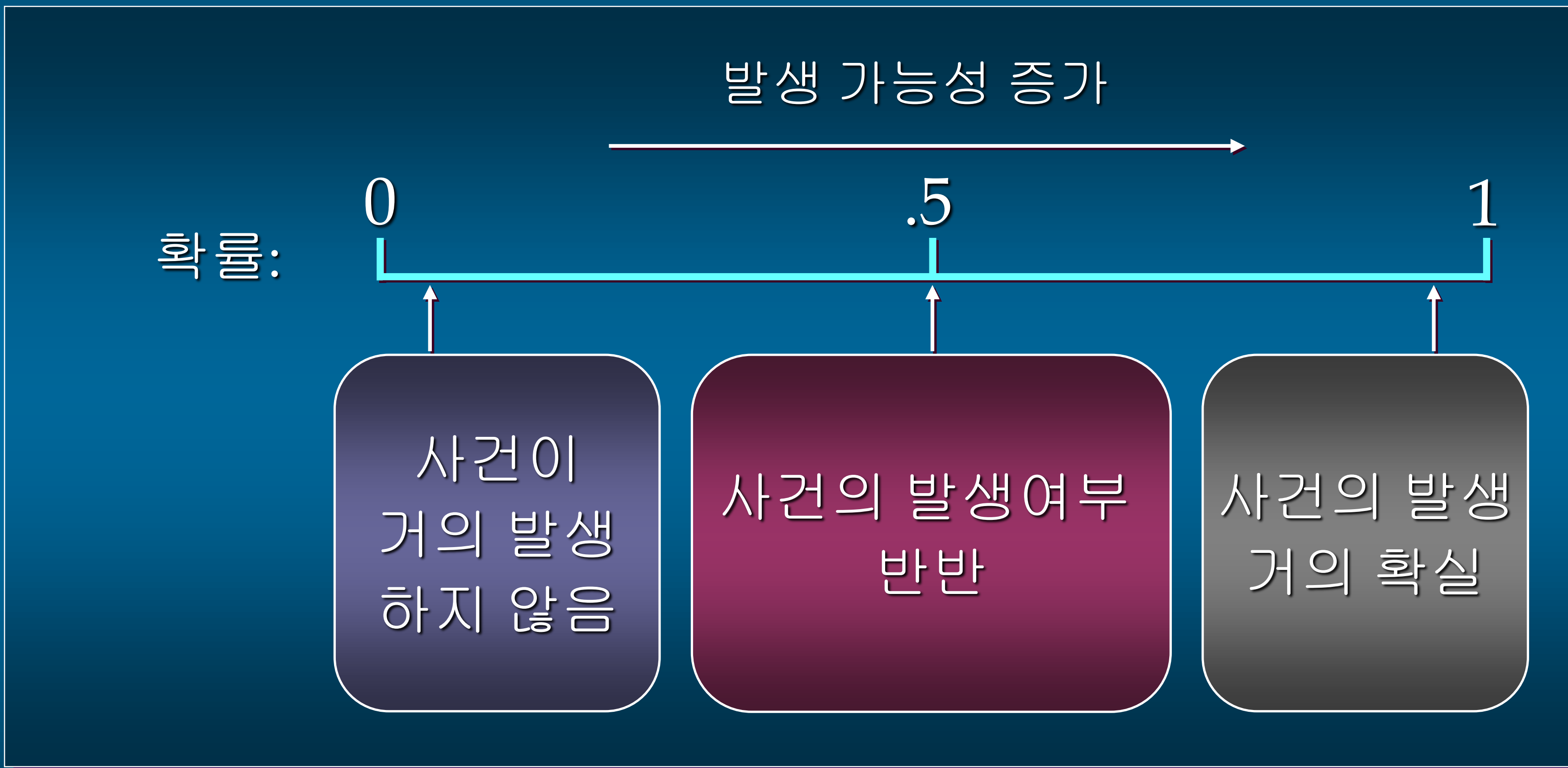
표본점
사건

표본공간

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



사건의 발생 가능성에 대한 척도로서의 확률



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 **N**개의 개체 중에서 순서를 고려하여 **n**개의 개체를 뽑는 실험결과의 개수를 알 수 있게 해준다.

N개의 개체 중에서 *n*개의 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.

$$P(E_1) + P(E_2) + . . . + P(E_n) = 1$$

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

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

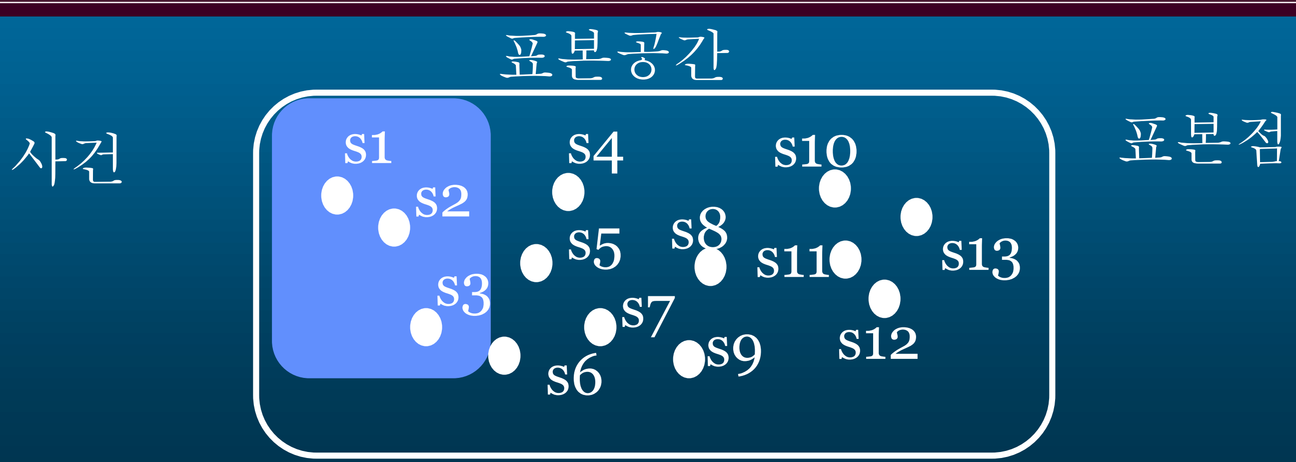
각 부여된 확률은 총도수(총 일수)로 해당 도수 (해당 일수)를 나눈 값이다.

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

4/40

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.



확률의 기본 관계

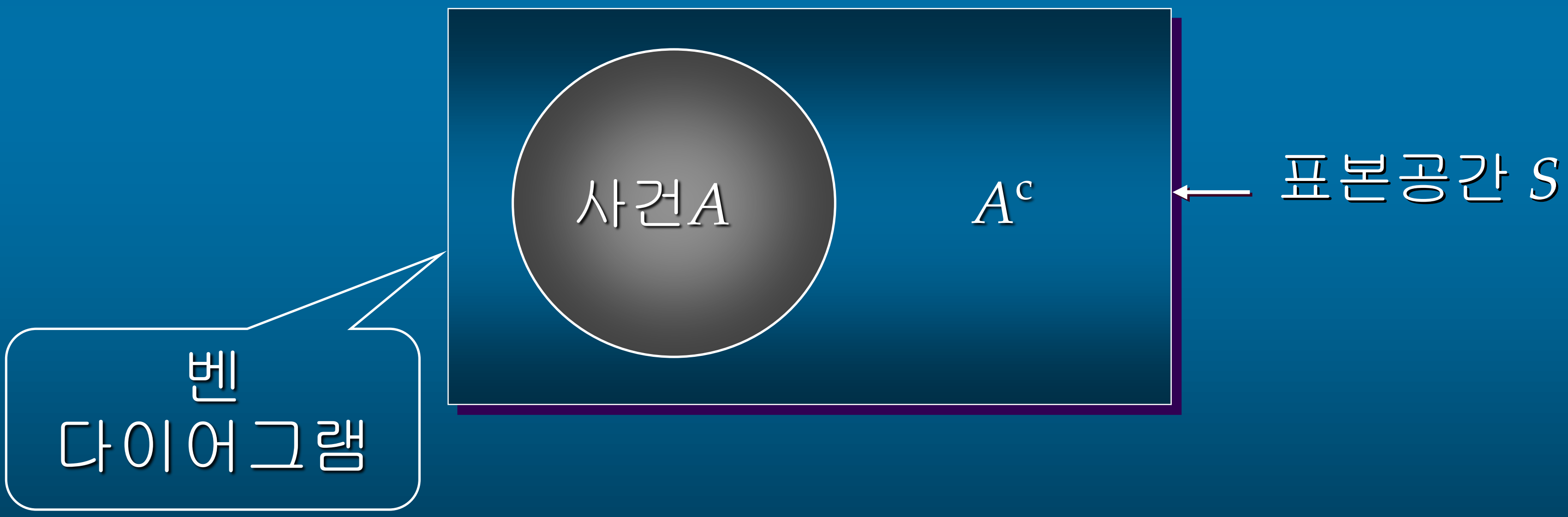
모든 표본점들의 확률을 모르더라도 사건의 확률을 계산하기 위해 사용될 수 있는 확률의 기본적인 관계가 존재 한다.

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

사건의 여집합(complement)

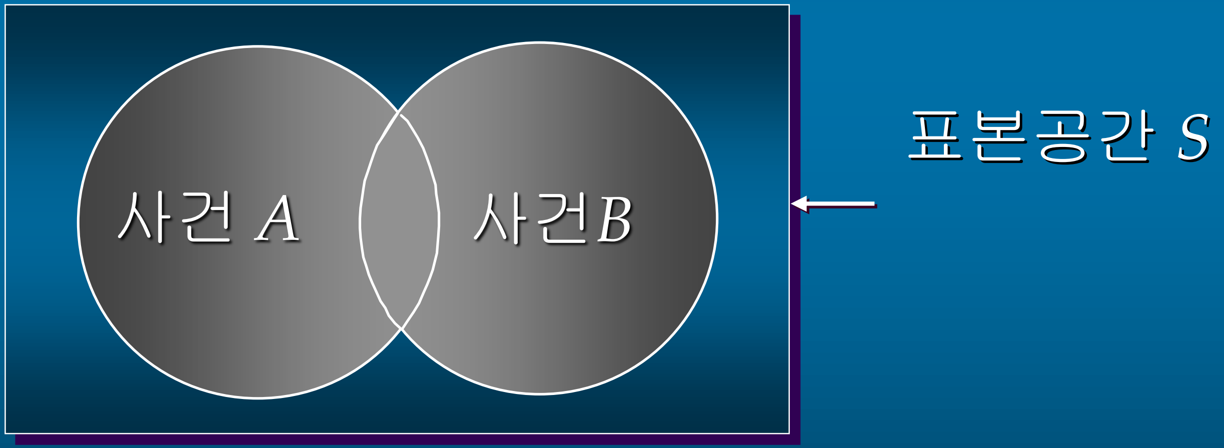
A의 여집합이란 사건 A에 속하지 않은 표본점들로 구성된 사건을 말한다.

A의 여집합은 A^c로 표시된다.



두 사건의 합집합(union)

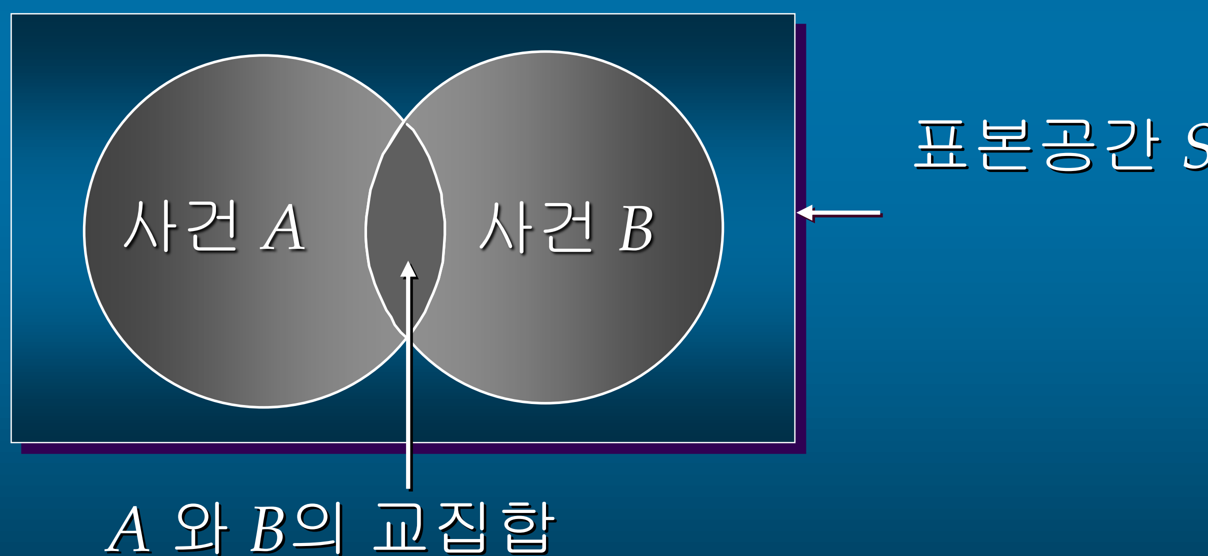
- 사건 A 와 B의 합집합은 A 나 B 또는 양쪽 모두에 속하는 모든 표본점들을 포함하고 있는 사건이다.
- A와 B 의 합집합은 $A \cup B$로 표시된다.



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.



덧셈(addition) 법칙

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

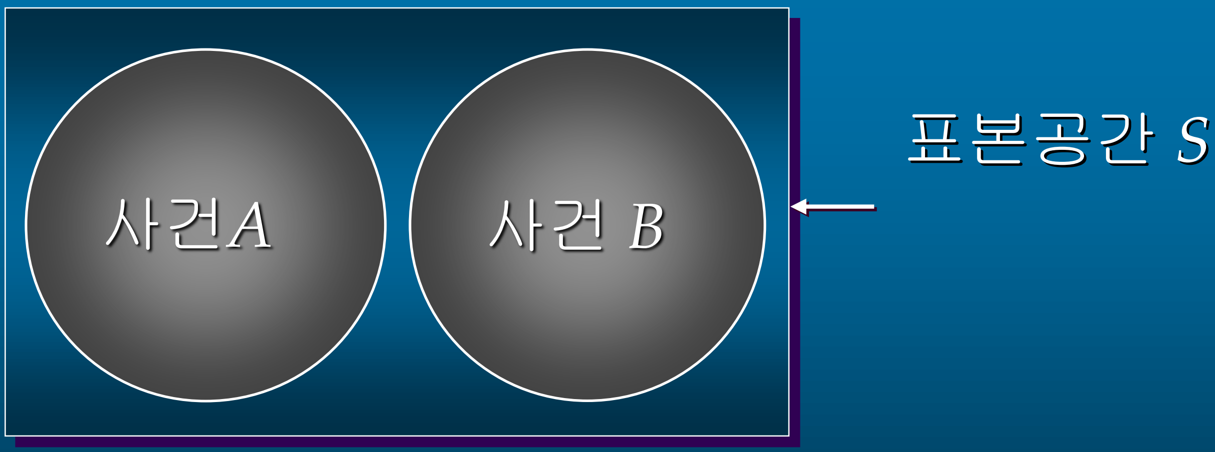
덧셈 법칙은 아래와 같다:

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

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

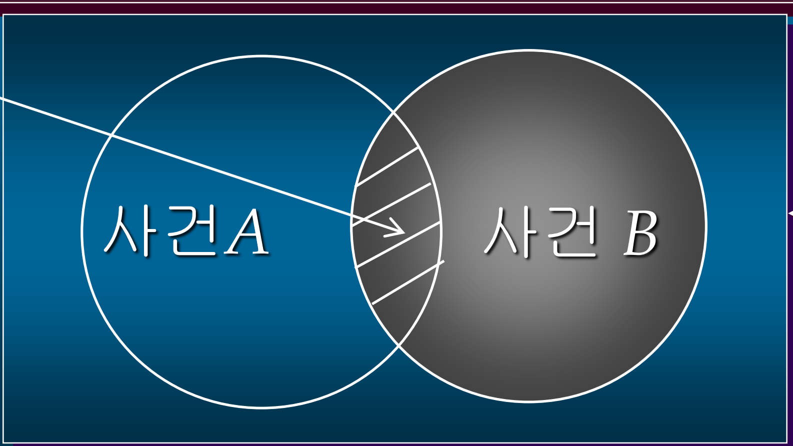
한 사건이 발생했을 때 다른 사건이 발생할 확률을 조건부 확률이라고 한다.

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

조건부 확률은 아래와 같이 계산된다

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

사건 A교집합 B



표본공간

곱셈(multiplication)법칙

곱셈법칙은 두 사건의 곱의 확률을 계산하는데 쓰인다.

곱셈법칙은 아래와 같다:

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

독립(independent) 사건

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

사건A 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:

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

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

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

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

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

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

Review: Conditional Probability

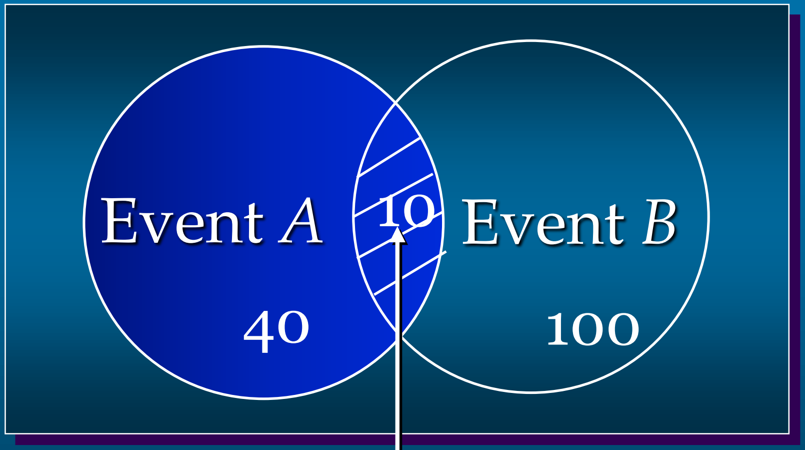
- Conditional probability: Probability of an event given that another event has occurred. Probability that B will occur given the condition of event A
- The conditional probability of B given A is denoted by P(B|A).

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

$$P(A) = \frac{n_A}{n_S} = \frac{50}{150}$$

$$P(B) = \frac{n_B}{n_S} = \frac{110}{150}$$

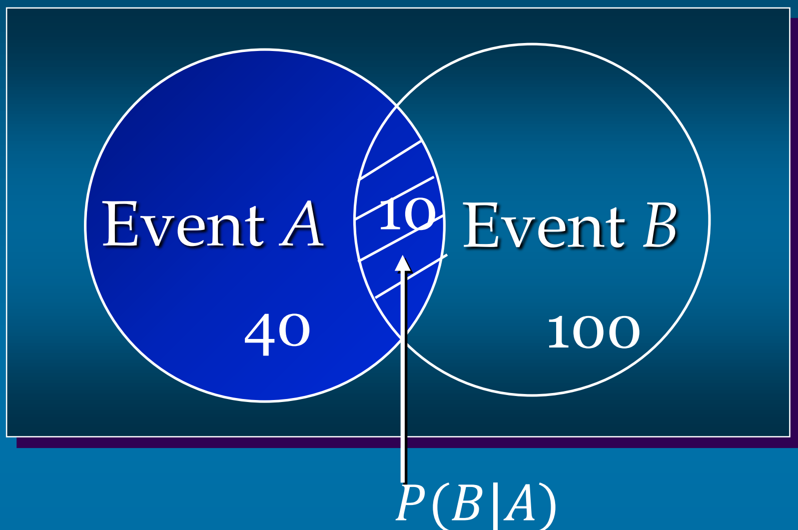
$$P(A \cap B) = \frac{10}{150}$$



$P(A \cap B)$

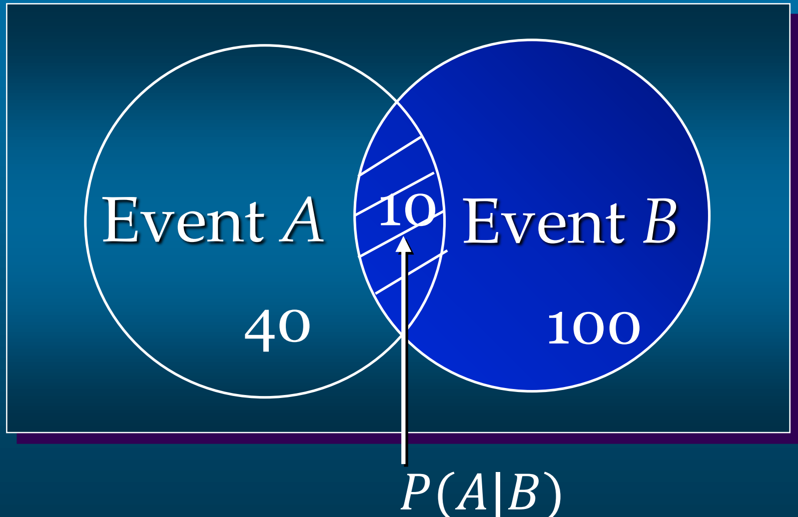
$$P(B|A) = \frac{P_{A \cap B}}{P_A} = \frac{10/150}{50/150}$$

Review: Conditional Probability



$$P(A \cap B) = \frac{10}{150}$$

$$P(B|A) = \frac{P_{A\cap B}}{P_A} = \frac{10/150}{50/150}$$



$$P(A \cap B) = \frac{10}{150}$$

$$P(A|B) = \frac{P_{A\cap B}}{P_B} = \frac{10/150}{110/150}$$

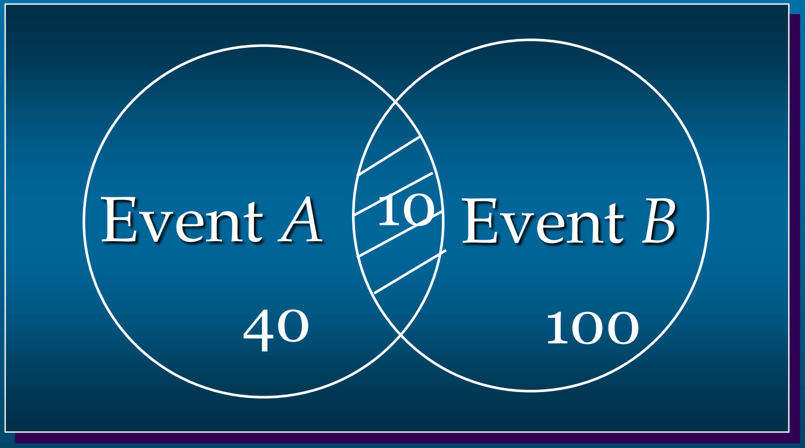
Review: Conditional Probability

- Application of Conditional Probability
  - $P(A \cap B)$ can be described as a <u>multiplication</u> of conditional probability

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

$$P(A \cap B) = P(A) \times P(B|A) = \frac{50}{150} \times \frac{10}{50} = \frac{10}{150}$$

$$P(A \cap B) = P(B) \times P(A|B) = \frac{110}{150} \times \frac{10}{110} = \frac{10}{150}$$



Review: Bayes' Theorem

- Bayes' Theorem
  - Probability theory that calculates the relationship between prior and posterior probabilities using conditional probabilities

사전확률 조건부확률

  - Prior probability P(A) : Event of Interest. (no new information)
  - Conditional probability P(B | A): Given A, we have new information B

(+customer report + complain no.+..)

  - Posterior probability P(A | B): Given new information B → Prob. A

사후확률

Review: Bayes' Theorem

- Prior Probability 사전확률 $P(A)$

Event of interest A

- Conditional Probability 조건부확률

New information B arrived given A

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

- Posterior Probability 사후확률

Given new information B,
what is the probability of event A?

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

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

$$\boxed{P(A|B)} = \frac{P(A \cap B)}{P(B)} = \frac{\overbrace{P(A)P(B|A)}^{P(A \cap B)}}{\underbrace{\underbrace{P(A)P(B|A)}_{P(A \cap B)} + \underbrace{P(A')P(B|A')}_{P(A' \cap B)}}}$$

Review: Bayes' Theorem

Example of cancer detection

- Mr. Kim, 45 years old, was tested for lung cancer via chest CT scanning test, and the test result was "positive".
- When he has lung cancer, the test detects the lung cancer as "positive" result with a probability of <u>90%</u> correctness.
- Even if it is not a lung cancer, there is still <u>7% chance</u> of a positive test outcome.
- Men in their 40s and 50s have a <u>0.8% chance</u> of getting lung cancer.

- What is the probability that Mr. Kim has a lung cancer (when tested positive)?

Review: Bayes' Theorem

- Let's summarize the event given probability

A={cancer, normal}
B={positive, negative}

Event of interest A
New information B

- Probability of getting lung cancer

P(A): Prior probability

P(cancer)=0.008 (0.8%)

P(normal)=1-0.008=0.992 (99.2%)

- Conditional probability:

P(B|A): B new information given A

$$P(positive|cancer) = \frac{P(cancer \cap positive)}{P(cancer)} = 0.9\ (90\%)$$

$$P(positive|normal) = \frac{P(normal \cap positive)}{P(normal)} = 0.07\ (7\%)$$

Review: Bayes' Theorem

- Probability of having lung cancer when the test result is positive:

P(A|B): Posterior probability (given B)

$$P(cancer|positive)$$

$$= \frac{P(cancer)P(positive|cancer)}{P(cancer)P(positive|cancer) + P(normal)P(positive|normal)}$$

$$= \frac{(0.008)(0.9)}{(0.008)(0.9) + (0.992)(0.07)}$$

$$= \frac{0.0072}{0.0072 + 0.0694} \quad = \frac{0.0072}{0.0766}$$

$$= 0.0939\ (9.39\%)$$

Example Questions

Ch4. 45. In an article about investment alternatives, *Money* magazine reported that drug stocks provide a potential for long-term growth, with over 50% of the adult population of the United States taking prescription drugs on a regular basis.
For adults age 65 and older, 82% take prescription drugs regularly.
For adults age 18 to 64, 49% take prescription drugs regularly.
The 18-64 age group accounts for 83.5% of the adult population (Statistical Abstract of the United States, 2008).

a. *What is the probability that a randomly selected adult is 65 or older?*
Let A=age 65 or older.
P(A)=1-0.835=0.165   16.5%

a. *Given that an adult takes prescription drugs regularly, what is the probability that the adult is 65 or older?*
Let D=take drugs regularly.

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A)P(D|A)}{P(A)P(D|A) + P(A')P(D|A')}$$

$$= \frac{0.165\ (0.82)}{0.165(0.82)+0.835(0.49)} = 0.2485$$

기타

올해 250주년 맞은 베이즈 정리, 과학을 정복하다

2013.06.12 09:11

가 가

[강석기의 과학카페 130] 확률적 믿음을 수식화 해

"기하학에 피타고라스 정리가 있다면 확률론에는 베이즈 정리가 있다." - 해럴드 제프리스 경

에스라인 몸매를 자랑하던 S씨는 10여 년 전 어느 날 샤워를 하다 문득 가슴에서 작은 덩어리가 만져지는 것 같은 느낌이 들었다. 며칠간 불면의 밤을 보내다 용기를 내 병원을 찾았고 유방암 검사를 했다. 당시 의사는 검사 정확도가 90%라고 알려줬다. 그리고 검사 결과 양성으로 나왔다. 자신이 유방암일 확률이 90%라는데 충격을 받은 S씨는 그 자리에 털썩 주저앉았다.

"이 결과로는 유방암일 확률이 10%도 안 되니 너무 걱정하지 말고 추가 검사를 해봅시다."
"그게 무슨 말씀이세요?"

의사 말에 따르면 유방암에 걸린 여성은 성인 여성의 1% 수준이고 검사 정확도가 90%이므로 정상인데도 검사에서 유방암에 걸린 것으로 나올 확률은 10%다. 따라서 설사 검사에서 양성으로 나왔더라도 진짜 유방암에 걸렸을 확률은 8%에 불과하다는 것.

의사는 화이트보드에 수식까지 쓰며 설명해줬지만 S씨는 무슨 말인지 알아들을 수가 없었다. 아무튼 여러 검사를 한 결과 다행히 유방암이 아닐 걸로 판정됐다. 당시 의사가 S씨에게 설명하려고 했던 게 바로 베이즈 정리(Bayes' theorem)로 확률을 얻는 방법이다. 올해는 베이즈 정리가 발표된 지 250주년 되는 해다.