Chap 4. 확률(probability)

Ho Sun Shon

목 차

생활속에서 만나는 확률이야기

- ➢ 로또에 당첨될 확률
- ➢ 야구 경기에서 특정팀이 승리 할 확률
- ➢ 흡연자가 폐암에 걸릴 확률
- ➢ 집값이 오를 확률
- ➢ 일기 예보(비 올 확률이 60%), 우산을 가지고 갔을 때 불편 때문에 생긴 손실 m, 우산을 가지고 가지 않았을 때 손실을 r 이라 할 때

표 4.1 의사결정과 예상손실

의사결정	예상손실
우산을 가져간다	m
우산을 가져가지 않는다	$0.6 \times r + 0.4 \times 0 = 0.6r$

- ❖ $m < 0.6r$ 이면 우산을 가져간다.
- ❖ $m > 0.6r$ 이면 내일 우산을 가져가지 않는다
- ❖ $m = 0.6r$ 이면 내일 우산을 가져가도 되고 가져가지 않아도 된다

표본 공간 (sample space)

- **통계적 시행(trial) 또는 통계적 실험(experiment)**
  - 어떠한 시행이 수행될 때 일어날 수 있는 가능한 결과들 중 한가지 결과가 나타나며 그 시행을 수행하기 전에는 어떤 결과가 나올 지 모름
- **표본공간(sample space)**
  - 통계적 실험을 통해 일어날 수 있는 모든 가능한 결과들의 집합
  - 예제1: 동전을 두 번 던질 때 표본공간?
  - 예제2: 동전의 앞면이 나올 때까지 던질 때 표본공간?
  - 예제3: 암환자가 치료를 받은 후 생존시간을 t라 할 때, 생존시간에 대한 표본 공간? $S = \{t|t > 0\}$

표본 공간 (sample space)

- 이산 표본공간(Discreate sample space)
  - 표본공간을 구성하는 원소들을 셀 수 있는 경우
- 연속 표본공간(Continuous sample space)
  - 표본공간을 구성하는 원소들을 셀 수 없는 경우
- 사건(event)
  - 표본공간의 부분집합
  - 단순사건: 표본공간 안에서 단하나만의 결과를 갖는 사건
- 배반(mutually exclusive)
  - $A \cap B = \emptyset$

문제 1

- 동전을 3번을 던질 때 표본공간은?
  - S={HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}
- 주사위를 2번 던지는 실험에서
  - 주사위를 2번 던질 때 표본공간을 구하면?
  - 두 주사위의 합이 5가 될 확률은?
  - 두 주사위의 합이 2이하가 될 확률은?

확률 (Probability)

- 여러 가지 가능한 결과 중 하나가 일어나는 실험에서, 그 중 일부가 일어날 가능성을 0과 1 사이의 값으로 나타낸 것
- 고전적 확률(수학적 확률)
  - 표본공간 안의 n개의 원소들이 일어날 가능성은 모두 같고 그 가운데 반드시 한 개는 일어 날 때

$$P(A) = \frac{\text{사건}A\text{에 속하는 모든 원소의 개수}}{\text{표본공간 안의 모든 원소의 개수}}$$

  - 예제4.4
    - 사건 A를 주사위를 던졌을 때 짝수의 눈이 나올 확률
    - 예제: 사건 B를 45개 숫자 가운데 임의의 6개의 숫자를 골라내는 것이라 할 때 사건 B의 확률은?

확률 (Probability)

- 상대도수의 극한적 개념의 확률(통계적 확률)
  - ❖ 예: 임의로 선택한 학부생이 승용차를 소유하고 있을 확률
  - ❖ 예: 한 공장에서 생산되는 컴퓨터의 특정 부품의 불량률
  - ❖ 상대도수의 극한적 개념을 이용하여 확률을 정의
  - ❖ **통계적 실험을 무한히 반복했을 때 사건 A가 n번 일어난 상대도수를 의미**

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

  - ❖ n= 시행횟수, #(A)=n번 시행 중 사건 A가 일어나는 횟수

확률 (Probability)

➢ 예제 4.5: 앞면과 뒷면이 있는 정상적인 동전 한 개를 던질 때, 동전의 앞면이 나타날 확률을 상대 도수의 극한 개념을 사용하여 구하여라.

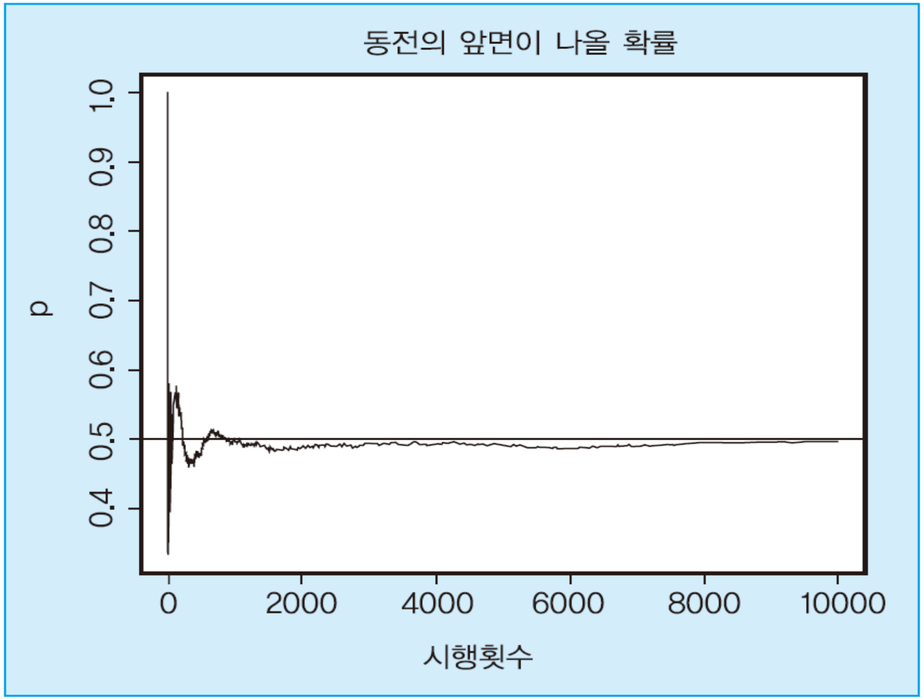


그림 4.1 동전을 10,000번 던졌을 때 앞면이 나타날 확률

확률 (Probability)

➢ 주관적 확률

- ❖ 개인의 경험이나 지식에 근거한 추측 또는 간접경험에 의하여 의사결정자가 주관적으로 정의
- ❖ 예:
  - ▪ 어떤 유명 가수가 전국 투어 공연에서 일정액의 흥행수익을 올릴 확률
  - ▪ 주식에서 일정기간 투자하여 원하는 수익을 달성 할 확률

확률의 법칙

➢ 확률의 공리

❖ 모든 사건(표본공간)의 임의 사건 A에 대하여 $0 \leq P(A) \leq 1$

❖ $P(S) = 1$

❖ 사건 $A_1, A_2, \ldots$ 가 서로 배반일 때 (즉, 서로 다른 i, j에 대하여 $A_i \cap A_i = \emptyset$)

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

❖ 여사건의 확률

- $P(A^c) = 1 - P(A)$

여사건 (Complement Event)

➢ 예제 4.6

❖ 타이타닉호의 생존 여부

표 4.2 타이타닉호의 참사현황

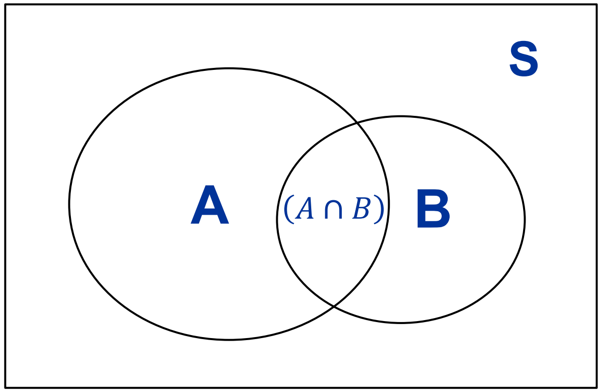
생존 여부	계층				합계
	1등석	2등석	3등석	승무원	
생존	202	118	178	212	710
사망	123	167	528	673	1,491
합계	325	285	706	885	2,201

❖ 임의의 한 사람을 선택 했을 때 생존 할 확률: P(A)

❖ 임의의 한 사람을 선택 했을 때 사망 할 확률: 1-P(A)

확률의 덧셈 정리

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

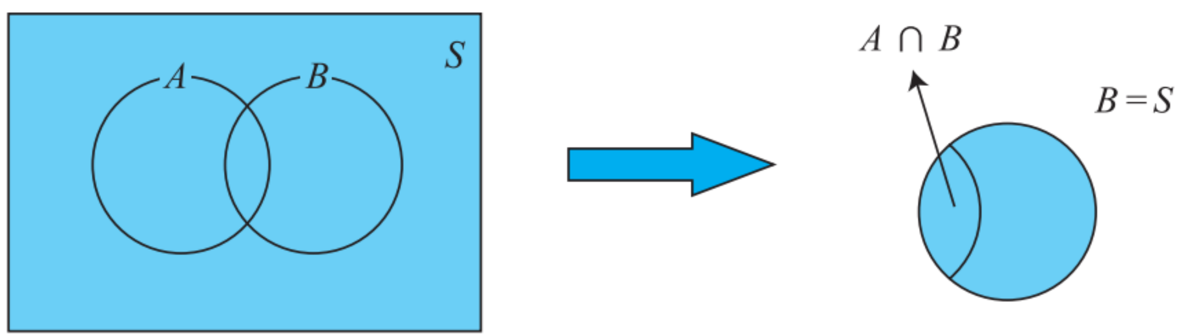


- 예제 4.7
  - 총 2,201명의 사람들 중 임의로 선택한 사람이 생존하거나 승무원일 확률은?
  - $P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{710}{2201} + \frac{885}{2201} - \frac{212}{2201} = 0.6284$

조건부 확률 (Conditional Probability)

- 사건 B가 발생했다는 정보가 주어졌을 때, 사건 A가 발생할 확률을 조건부확률이라고 하며 $P(A|B)$라고 나타냄
- 일반적으로 아무런 정보가 없는 상태에서 사건 A가 발생할 확률 $P(A)$와 사건 B가 발생했다는 조건 하에서 사건 A가 발생할 조건부확률인 $P(A|B)$은 같지 않음

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



조건부 확률 (Conditional Probability)

➢ 예제 4.8:

표 4.2 타이타닉호의 참사현황

생존 여부	계층				합계
	1등석	2등석	3등석	승무원	
생존	202	118	178	212	710
사망	123	167	528	673	1,491
합계	325	285	706	885	2,201

(1) 총 2,201명 가운데서 임의로 선택한 사람이 1등석 승객이면서 생존할 확률

❖ $P(A \cap B) = \frac{202}{2201} \approx 0.0918$

(2) 1등석 승객 중 임의로 선택한 사람이 생존할 확률(1등석 승객: A_1, 생존사건: B)

❖ $P(B|A_1) = \frac{P(B \cap A_1)}{P(A_1)} = \frac{202/2201}{325/2201} = 0.6215$

조건부 확률 (Conditional Probability)

- 예제 4.8:

(3) 생존 한 사람 중 임의로 선택한 사람이 1등석 승객일 확률

$$P(A_1|B) = \frac{\frac{202}{2201}}{\frac{710}{2201}} = 0.2845$$

(4) 총 2201명 가운데서 임의로 선택한 사람이 3등석 승객이면서 생존할 확률

$$P(A_3 \cap B) = \frac{178}{2201} = 0.0809$$

(5) 3등석 승객 중 임의로 선택한 사람이 생존 할 확률

$$P(B|A_3) = \frac{P(B \cap A_3)}{P(A_3)} = \frac{\frac{178}{2201}}{\frac{706}{2201}} = 0.2521$$

(6) 생존한 사람이 임의로 선택한 사람이 3등석 승객일 확률

$$P(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = \frac{\frac{178}{2201}}{\frac{710}{2201}} = 0.2507$$

조건부 확률 (Conditional Probability)

- (7) (1)번부터 (6)번 까지를 통하여 알 수 있는 사실은 무엇인가?
  - 모든 계층의 생존 비율(사망 비율)이 비슷한가?
  - 상류층의 생존비율(62.15%)과 하류층의 생존비율(25.07%)이 다른 것을 확인 할 수 있음

문제 1

- 문제) 어느 회사의 직원 100명 중에서 성별과 흡연 여부에 따른 분할표

	흡연	비흡연	합계
남	30	30	60
녀	15	25	40
합계	45	55	100

- 임의로 한 명을 선택할 때, 선택된 사람이 흡연자일 확률은? 선택된 사람이 남성이라는 정보가 주어졌을 때, 이 사람이 흡연자일 확률은?
- A: 선택된 사람이 흡연자인 사건
  B: 선택된 사람이 남성인 사건

문제 2

➢ 예) 어느 회사 직원을 체중과 혈압에 따른 비율

	비만	정상	정상 이하	합계
고혈압	0.10	0.08	0.02	0.20
정상 혈압	0.15	0.45	0.20	0.80
합계	0.25	0.53	0.22	1.00

❖ 한 명을 임의 추출할 때, 이 사람이 고혈압일 확률은?

❖ 임의로 선택된 사람이 비만일 때, 이 사람이 고혈압일 확률은?

❖ A: 선택된 사람이 고혈압인 사건

B: 선택된 사람이 비만인 사건

확률의 곱셈정리

- 조건부확률을 이용하면

$$P(A \cap B) = P(B)P(A|B):$$ **곱사건의 확률법칙**

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

- 표본공간 S에서 정의되는 사건을 $A_1,\ A_2,\ A_3, \cdots, A_n$

$$P(A_1A_2A_3 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2A_3 \cdots A_{n-1})$$

where, $A_1A_2A_3 \cdots A_k = A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_k$

- **확률의 곱셈정리**: 표본공간 S 안의 두 사건 A와 B가 일어난다면 교사건의 확률 $P(A \cap B)$ 에 대해 다음이 성립

$$P(A \cap B) = \begin{cases} P(A)P(B|A): & P(A) > 0 \\ P(B)P(A|B): & P(B) > 0 \end{cases}$$

확률의 곱셈정리

➢ 예제 4.9

❖ 한 방에 모여 있는 사람의 수를 N이라 할 때 서로 같은 생일을 갖게 될 확률을 구하라.(1년을 365일로 가정)

❖ $P(N$명의 생일이 서로 다른 사건)=$\frac{365\times364\times363\cdots\times(365-N+1)}{365^N}$

❖ $P(N$명 가운데 최소한 두 명이 같은 생일을 갖는다)

❖ =1-P(N명의 생일이 서로 다르다) $= 1 - \frac{365\times364\times363\cdots\times(365-N+1)}{365^N}$

표 4.3 *N*명이 있을 때 서로 같은 생일을 갖게 될 확률

N	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
P	0.027	0.117	0.253	0.411	0.569	0.706	0.814	0.891	0.941	0.970	0.986	0.994	0.998	0.999	0.9997

확률의 곱셈정리

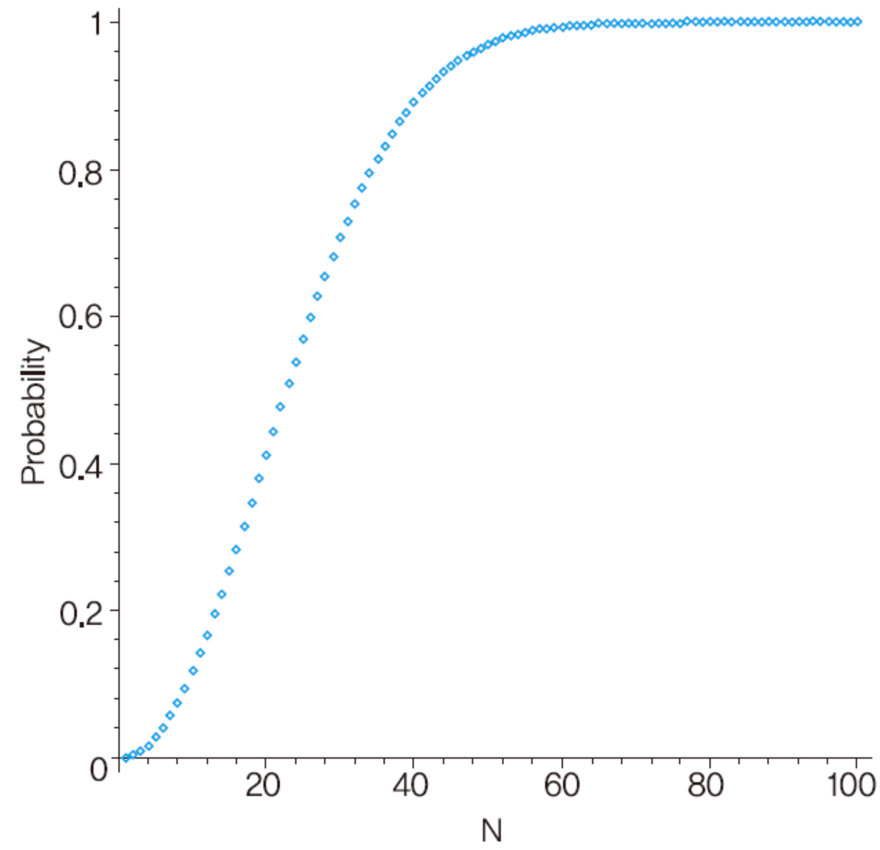


그림 4.3 N명이 있을 때 서로 같은 생일을 갖게 될 확률

확률의 곱셈정리

- 예제: 남학생 7명과 여학생 5명으로 이루어진 모임에서 임의로 2명을 선택할 때, 두 사람 모두 남학생일 확률은?

1. $\dfrac{\binom{7}{2}}{\binom{12}{2}} = \dfrac{7\times6/2}{12\times11/2} = \dfrac{7\times6}{12\times11}$

2. A: 두 번째로 선택된 학생이 남학생인 사건

   B: 첫 번째로 선택된 학생이 남학생인 사건

   $$P(A \cap B) = P(B)P(A|B) = \frac{7}{12} \times \frac{6}{11} = \frac{7\times6}{12\times11}$$

문제 3

- 한 학생이 동전을 던져 공부를 하러 학교에 갈 것인가(S) 아니면 영화를 보러 갈 것인가(M)를 결정하고 공부하게 되면 A, B, C 과목 중 하나를 임의로 선택하여 공부하고, 영화는 D, E 프로 중 하나를 임의로 선택하여 보기로 하였다. 이 학생이 C 과목을 공부할 확률은?

독립사건 (Independent Event)

- 두 사건 A와 B에 대하여 $P(A|B) = P(A)$이 성립한다면, 이는 어떤 의미인가?
  - 사건 A가 발생할 확률과 사건 B가 발생했다는 조건 하에서 사건 A가 발생할 조건부확률이 같다.
  - $P(A|B) = P(A)$이 성립하면 $P(A|B^c) = P(A)$도 성립한다.
  - 사건 A가 발생할 확률은 사건 B가 발생하든 발생하지 않든 상관없이 같은 값을 가진다.
  - 사건 B가 발생했다는 정보가 사건 A의 확률을 구하는데 영향을 주지 않는다.
  - 사건 A와 사건 B는 서로 영향을 주지 않는다.

독립사건 (Independent Event)

- 두 사건 A와 B에 대하여

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

  가 성립할 때, 사건 A와 사건 B는 서로 독립(independent)

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

독립사건

표본공간 S 안의 두 사건 A와 B가 $P(A)>0$, $P(B)>0$이며, 다음 식이 성립하면 두 사건 A와 B는 서로 독립이라고 한다.

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

독립사건 (Independent Event)

- 예) 성별과 길거리 금연에 대한 찬반 의견 조사

	금연 찬성	금연 반대	합계
남	24	16	40
녀	14	6	20
합계	38	22	60

- 이 중에서 임의로 한 명을 선택할 때,

  M = 선택된 사람이 남성인 사건

  A = 선택된 사람이 금연에 찬성할 사건

- $P(M) = \frac{40}{60} = \frac{2}{3}, P(A) = \frac{38}{60} = \frac{19}{30}$ => $P(M \cap A) = \frac{24}{60} = \frac{2}{5}$

  $\rightarrow P(M \cap A) \neq P(M)P(A)$ 이므로 A와 M은 독립이 아니다.

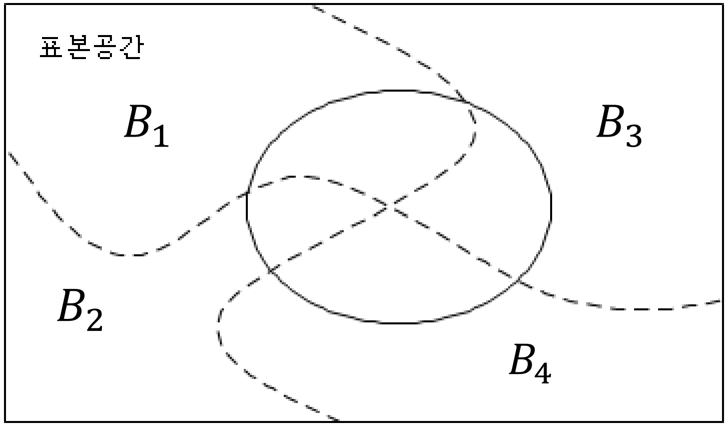
  혹은 $P(A) = \frac{38}{60}, P(A|M) = \frac{24}{40}$로 서로 다르므로 A와 M은 독립이 아니다

베이즈 정리 (Bayes' theorem)

➢ 분할 (partition)

❖ 표본공간 S 안에서 n개의 사건 $B_1, B_2, \dots, B_n$ 가 서로 배반이고,

❖ $B_1 \cup B_2 \cup \cdots \cup B_n = S$일 때, 즉 $B_1, B_2, \dots, B_n$이 표본공간을 분할



(1) 임의의 i와 j에 대하여 $P(B_i \cap B_j) = 0$

(2) $\bigcup_{i=1}^{n} B_i = S$

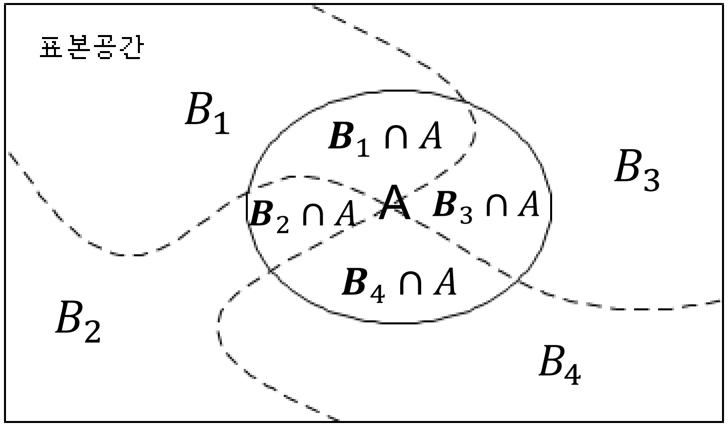
베이즈 정리 [Bayes' theorem]

- **전확률 (total probability)**
  - N개의 배반사건 $B_1, B_2, \dots, B_n$이 표본공간 S를 분할 할 때 임의의 사건 A의 전확률 (total probability)을 다음과 같이 표현
  - $P(A) = P((B_1 \cap A) \cup \cdots \cup (B_n \cap A))$

$$= P(B_1 \cap A) + \cdots + P(B_n \cap A)$$

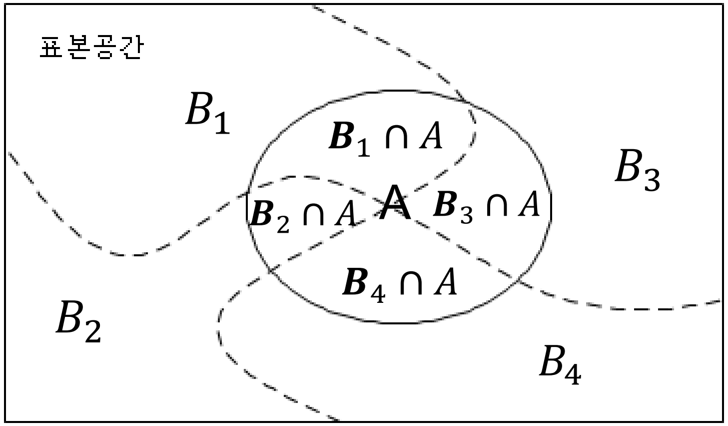
$$= P(B_1)P(A|B_1) + \cdots + P(B_n)P(A|B_n)$$

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



베이즈 정리 (Bayes' theorem)

- n 개의 배반사건 $B_1, B_2, \dots, B_n$이 표본공간 S를 분할 할 때, 임의의 사건 A가 주어진다면, 특정한 사건 B_j로 부터 나왔을 확률 $P(B_j|A)$?
- $P(B_j|A)$ =?



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

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

$$= \frac{P(B_j)P(A|B_j)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n)}$$

- 특정한 사건 $P(B_j)$를 사전확률(prior probability)
- A가 주어졌을 때 B_j로 부터 나왔을 확률 $P(B_j|A)$를 사후확률(posterior probability)

베이즈 정리 (Bayes' theorem)

➢ 예제 4.11: 세 개의 상자에 반도체들이 들어 있다. 첫 번째 상자에 있는 20개의 반도체 가운데 5개가 불량품이고, 두 번째 상자에 있는 30개의 반도체 가운데 3개가 불량품이며, 세 번째 상자에 있는 50개의 반도체 가운데에 12개가 불량품이다.

1) 상자 중 임의로 하나의 반도체를 선택 했을 때, 반도체 하나를 골라내어 검사한 결과 불량품일 확률을 계산하여라.
2) 1)에서 검사한 반도체가 불량품 일 때 첫번째 상자에서 나왔을 확률을 계산하여라.

베이즈 정리 (Bayes' theorem)

- 각 상자를 B라 할 때
  - 첫 번째 상자: $P(B_1) = \frac{20}{100}$
  - 두 번째 상자: $P(B_2) = \frac{30}{100}$,
  - 세 번째 상자: $P(B_3) = \frac{50}{100}$
- 불량품인 사건을 A라 하면
  - 첫 번째 상자의 불량품 확률: $P(A|B_1) = \frac{5}{20}$
  - 두 번째 상자의 불량품 확률: $P(A|B_2) = \frac{3}{30}$
  - 세 번째 상자의 불량품 확률: $P(A|B_3) = \frac{12}{50}$
- $P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$

$$= \left(\frac{20}{100} \times \frac{5}{20}\right) + \left(\frac{30}{100} \times \frac{3}{30}\right) + \left(\frac{50}{100} \times \frac{12}{50}\right) = 0.2$$

- $P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)} = 0.25$

문제 4

- 문제 4
  - 어느 회사의 공장 A_1, A_2, A_3에서 주기판(motherboard)의 30%, 50%, 20%를 생산한다. 공장 A_1, A_2, A_3의 불량률이 각각 2%, 1%, 5%이다.

1. 이 회사제품 중 임의로 하나를 선택하였을 때, 이 제품이 불량일 확률은?
2. 선택된 제품이 불량일 때, 이 제품이 공장 A_2에서 생산되었을 확률은?

요약

- 표본공간
- 확률, 확률의 법칙
- 조건부 확률
- 독립사건
- 베이즈 정리