확률 및 통계학

- 2장 확률(Probability) -

임연주 (yeonjoo@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

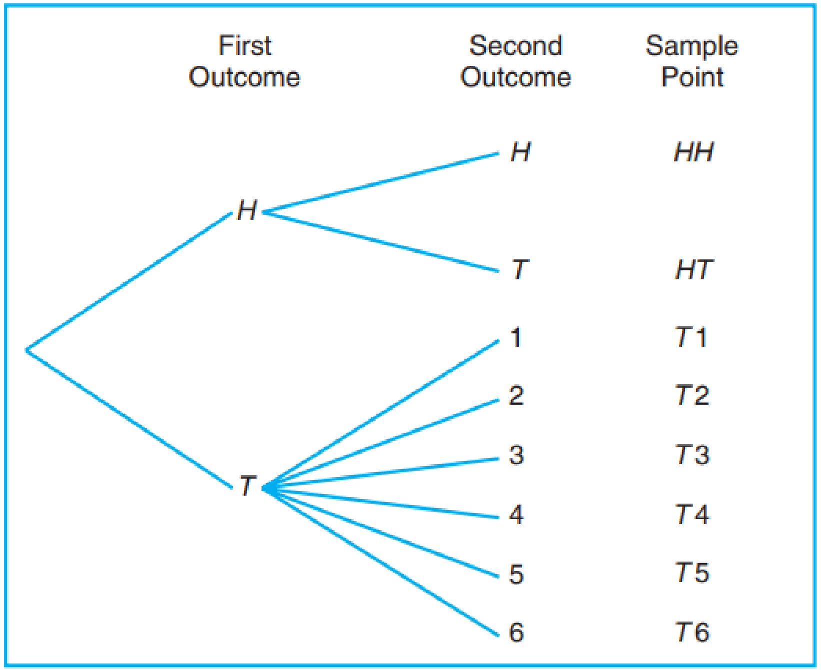
목 차

표본 공간

- 표본공간(Sample space)
  - 정의
    - 실험에서 발생 가능한 모든 결과들의 집합을 의미
    - 표본 개개의 원소(Element 또는 Member)를 표본점(Sample point)라고 함
    - 기호 S로 표시
      - $S = \{표본점1, 표본점2\}$
  - 특징
    - 실험마다 하나 이상의 표본공간이 사용될 수 있음
      - 표본공간들 중에서도 실험의 결과로써 가장 많은 정보가 부각될 수 있는 표본공간을 사용해야 함

표본 공간

- 표본공간(Sample space)
  - 수형도 (Tree Diagram)
    - 표본공간을 트리구조로 표현하여 원소들을 나열할 수 있는 다이아그램
    - 수형도 표현
      - 실험 예제
        - 동전 앞면 → 동전
        - 동전 뒷면 → 주사위
      - $S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$



표본 공간

- 표본공간(Sample space)
  - 식 표현
    - $S = \{H, T\}$
      - 동전 한 개를 던지는 실험
    - $S = \{x \mid x\text{는 인구 백만이 넘는 도시}\}$
      - 전세계 인구 백만이 넘는 도시들의 집합
    - $S = \{(x, y) \mid x^2 + y^2 \leq 4\}$
      - 반지름이 2인 원의 원주상과 내부에 있는 모든 점(x, y)의 집합
    - $S = \{D, ND, NND, NNND, \cdots\}$
      - 불량품 D가 발견될 때까지 표본을 추출

사상

- 사상(Event)
  - 표본공간의 부분집합을 의미
    - 공집합, 여집합이 될 수 있음
  - 사건이라고도 지칭함

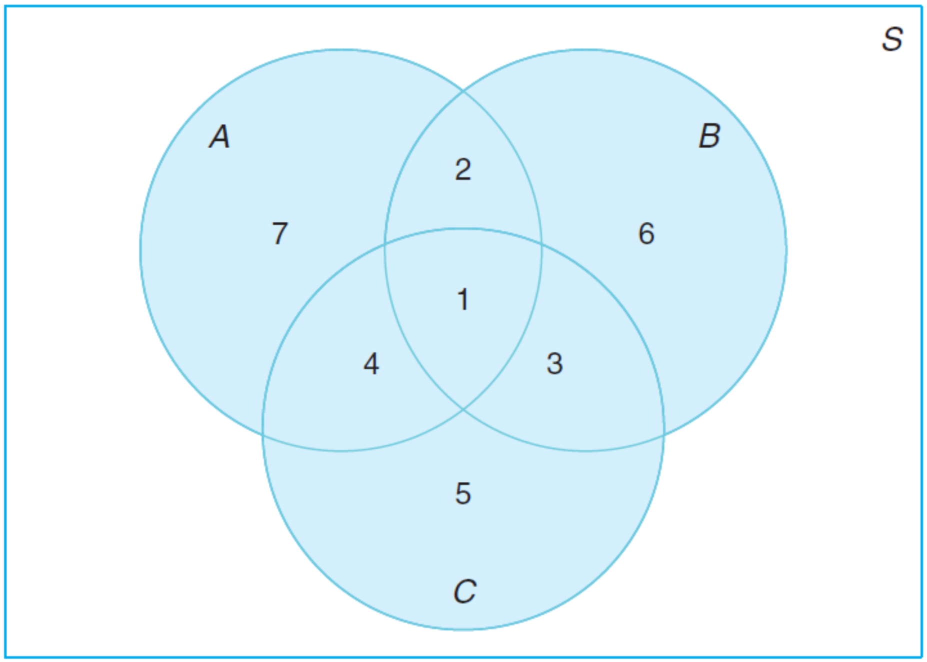
  - 사상 예제
    - $S = \{t \mid t \geq 0\}$인 표본공간에서 5년 내에 고장 날 사상 A는 $A = \{t \mid 0 \leq t \leq 5\}$이며, S의 부분집합
      - t: 전자부품 수명(단위: 년)

사상

- 사상(Event)
  - 사상과 표본공간과의 관계를 벤 다이어그램(Venn diagram)을 통해 표현
    1. 사상 A의 여집합
       - A 사상 집합의 원소가 아닌 S의 모든 원소들을 의미
       - A'로 표시
    2. 사상 A와 B의 교집합
       - A와 B에 공통으로 속하는 모든 원소들을 포함하는 사상
       - $A \cap B$로 표시
    3. 사상 A와 B의 합집합
       - A 혹은 B에 속하는 모든 원소들을 포함하는 사상
       - $A \cup B$로 표시
    4. 사상 A와 B의 상호 배반
       - $A \cap B = \varnothing$인 경우, 사상 A와 B는 상호배반(Mutually exclusive)함

사상

- 사상 벤 다이어그램



$$
\begin{aligned}
A \cup C &= \text{regions } 1,\ 2,\ 3,\ 4,\ 5, \text{ and } 7,\\
B' \cap A &= \text{regions } 4 \text{ and } 7,\\
A \cap B \cap C &= \text{region } 1,\\
(A \cup B) \cap C' &= \text{regions } 2,\ 6, \text{ and } 7,
\end{aligned}
$$

사상

- 사상 Question
  - 다음 중 동일한 사상은 어느 것인가?
    - (a) $A = \{1,3\}$
    - (b) $B = \{x \mid x$는 주사위 눈금 수$\}$
    - (c) $C = \{x \mid x^2 - 4x + 3 = 0\}$
    - (d) $D = \{x \mid x$는 6개의 동전을 던졌을 때 나오는 앞면의 수$\}$

경우의 수

- 경우의 수(Number of cases)
  - 실험에서 발생할 수 있는 모든 사상(Event)의 가짓수
  - 표본공간의 표본점의 수를 계산하는 것

- 경우의 수 계산 방법
  - 합의 법칙($A \cup B$)
    - 사상 A **또는** B가 독립적으로 일어나는 경우의 수
  - 곱의 법칙($A \cap B$)
    - 사상 A, B가 **동시에** 일어나는 경우의 수

경우의 수

- 순열(Permutation)
  - 서로 다른 n개의 원소 중 r개를 중복없이 골라 순서에 상관있게 나열하는 것
    - n개에서 r개를 선택하는 순열
    - $nPr = \frac{n!}{(n-r)!}$
  - 순열 Question
    - 30명의 학생으로 구성된 동아리의 회장과 총무를 뽑으려고 할 때, 다음 각 경우에 가능한 선출방법은 얼마나 되는가?
      - (a) 선출 방법에 제한이 없을 때
      - (b) A는 회장이 아니면 하지 않으려고 할 때
      - (c) B와 C는 같이 하는 경우가 아니면 하지 않으려고 할 때
      - (d) D와 E는 서로 같이는 하지 않으려고 할 때

경우의 수

- 순열(Permutation)
  - 원순열(Circular permutation)
    - n개를 나열하는데, 원형으로 나열하는 경우의 수
    - 한 원소를 기준점으로 선정하고 그 다음 원소부터 나열됨
    - $(n-1)!$
- 조합(Combination)
  - 서로 다른 n개의 원소에서 r개를 중복되지 않으면서 순서에 상관없이 뽑아 나열하는 것
  - $nCr = \binom{n}{r,n-r} = \binom{n}{r} = \frac{n!}{(n-r)!\,r!} = \frac{nPr}{r!}$

경우의 수

- 분할(Partition)
  - 서로소인 부분집합들의 합집합이 원래의 집합이 되는 가짓수
  - 분할의 수 계산
    - $\binom{n}{n_1, n_2, \ldots, n_r} = \left(\frac{n!}{n_1! n_2! \ldots n_r!}\right)$
      - n개의 집합에서 r개의 부분으로 나누는 경우
        - 단, $n_1 + n_2 + \cdots + n_r = n$
  - 분할 Question
    - 9명의 PEL 연구생이 5인용 객실 1개와 2인용 객실 2개에 투숙하는 방법은 몇 가지가 있는가? (단, 2인용 객실 1개에는 여학생 2명 투숙해야 함)

사상의 확률

- 사상의 확률
  - 실험에서 사상에 대한 가중치(Weight) 또는 확률(Probability)을 의미
  - 특징
    - 사상 A가 발생할 가능성을 나타내는 0~1 사이의 값
      - 어떤 사건이 일어날 확률(P)의 범위
        - $0 \leq p \leq 1$
    - 표본공간의 모든 표본점 확률들의 합은 1
      - $P(S) = 1$

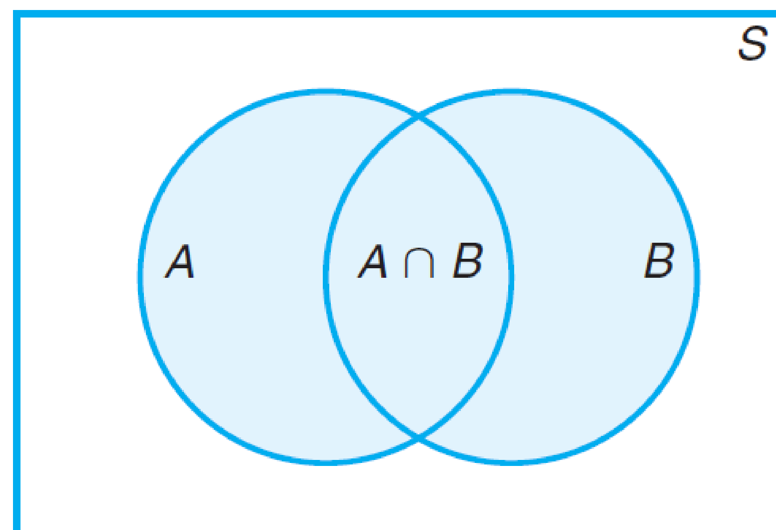
사상의 확률

- 사상의 확률
  - Question
    - 짝수가 홀수보다 2배만큼 더 많이 발생하는 주사위를 한번 던져서 4보다 작은 수가 나올 사상 E의 $P(E)$를 구하라

가법의 정리

- 가법정리(Additive rules)
  - 확률의 계산을 간단하게 해주는 법칙

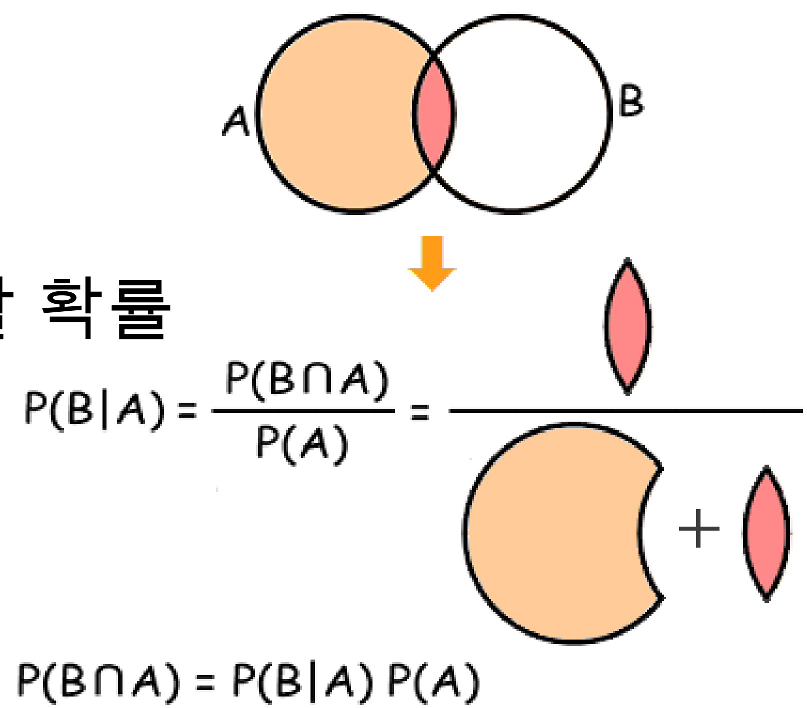
1. 확률의 덧셈정리
   - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   - 두 사상 A, B가 서로 상호배반사상이면 $A \cap B = \emptyset$ 이면 $P(A \cup B) = P(A) + P(B)$



2. 여사건의 확률
   - 사상 A가 일어나지 않을 확률
   - $P(A) = 1 - P(A')$

조건부 확률, 독립사상, 승법정리

- 조건부 확률(Conditional Probability)
  - 사상 A가 일어났다는 전제 조건에서 사상 B가 일어날 확률
    - 즉, 어떤 조건하에서 추가적으로 일어날 사상의 확률을 의미
      - 어떤 사상이 발생함을 바탕으로 또 다른 사상의 발생 확률을 다시 계산할 수 있음
        - 확률 P(A)의 갱신(Update)
          - 전체적인 결과가 변함
  - $P(B|A)$로 표기
    - 의미: A가 일어났을 때 B가 일어날 확률
    - $P(B|A) = \frac{P(B \cap A)}{P(A)},\ P(A)\ > 0$



조건부 확률, 독립사상, 승법정리

- 독립, 종속과 배반
  - 독립 사상(Independence event)
    - 서로 다른 사건이 일어날 확률에 영향을 주지 않는 사건들
      - 두 사상 A, B에서 B의 확률에 아무런 영향을 미치지 않음
    - 3 조건과 동치(하나만 성립하면 독립 사상)
      - $P(A \cap B) = P(A) * P(B)$
      - $P(B|A) = P(B)$
      - $P(A|B) = P(A)$
  - 종속 사상(Dependent event)
    - 사건 A가 일어났을 경우와 일어나지 않았을 경우에 따라 사건 B의 확률이 다를 경우
    - $P(B|A) \neq P(B|A')$
    - $P(B \cap A) = P(A \cap B) = P(A) * P(B|A)$

조건부 확률, 독립사상, 승법정리

- 독립, 종속과 배반
  - 배반 사상(Exclusive event)
    - 동시에 일어날 수 없는 두 사건
    - $P(A \cup B) = P(A) + P(B)$ 이 성립해야 함
      - $A \cap B = \emptyset \rightarrow P(A \cap B)$ = 0

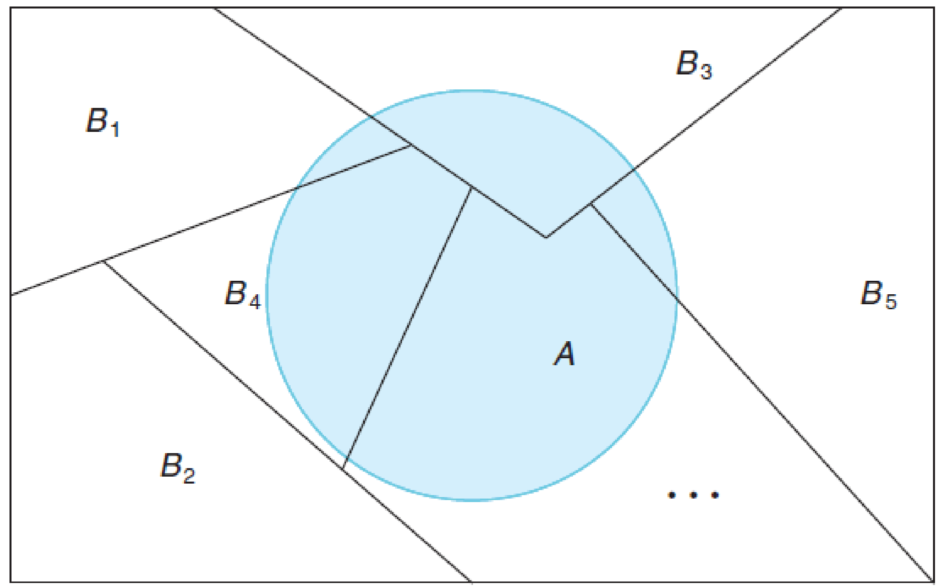
- 승법공식(Multiplicative rule)
  - 기존 공식에 계산특성을 이용해 두 사상이 함께 일어날 확률을 계산할 수 있음
    - 조건부 확률 공식($\frac{P(B \cap A)}{P(A)}$)에 양변에 $P(A)$를 곱함
  - 종속 관계일 때 성립

베이즈 정리

- 전확률(Total Probability)의 법칙
  - 사상 A를 서로 배반사건 B와의 합집합으로 나타낼 수 있음
    - $B \cap B` = 0,\ B \cup B`$이 전체 표본공간을 의미
    - $A = (B \cap A) \cup (B` \cap A)$
    - $P(A) = P[(B \cap A) \cup (B` \cap A)] = P(B \cap A) + P(B` \cap A)$
      $= P(B)P(A|B) + P(B`)P(A|B')$
  - 전확률의 정리(Theorem of total probability)
    - 표본공간이 k개의 부분집합으로 분할되는 경우, 식을 일반화하여 공식을 도출
    - $P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$
      - 사건 $B_1, B_2, \ldots, B_k$를 표본공간 S의 분할이고, $P(B) \neq 0$, $\mathrm{i} = 1, 2, \ldots, \mathrm{k}$이면 S의 임의 사건 A는 위 식을 성립함

베이즈 정리

- 전확률(Total Probability)의 법칙
  - 분할 개념 그림



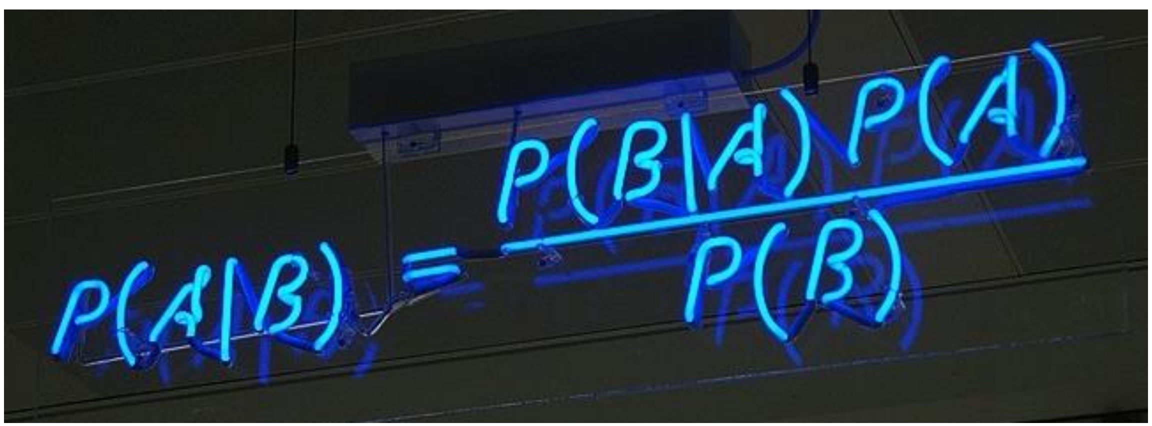
  - Question
    - 3대의 기계 B_1, B_2, B_3가 각각 전체 생산량의 30%, 45%, 25%를 생산하는 조립공장에서 불량품 제조율이 각각 2%, 3%, 2%이면, 불량품을 선택할 확률은 얼마인가?

베이즈 정리

- 개요
  - 사전확률(Prior probability)과 사후확률(Posterior probability)과의 관계를 설명하는 확률 이론
    - 주어진 조건부 확률 $\mathrm{P}(B|A)$을 이용하여 다른 조건부 확률 $\mathrm{P}(A|B)$을 구하는 것
      - 아래 두 가지 확률을 알아야 함
        - 사건 A, B가 각각 일어날 확률
        - 사건 A가 일어난 상태에서의 사건 B가 일어날 확률

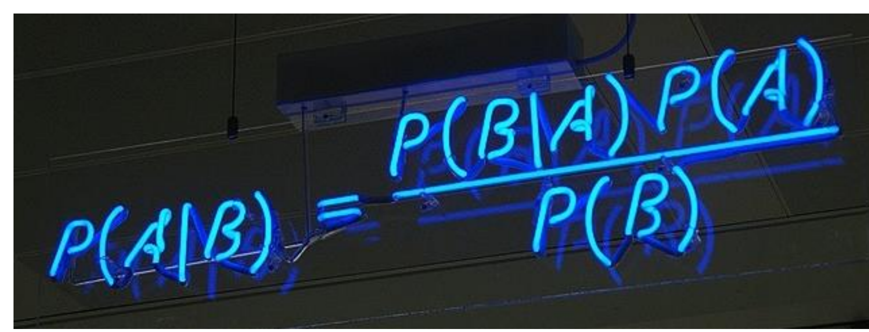
베이즈 정리

- 정의
  - 사건 B가 $B_1, B_2, \dots, B_n$ 중 하나라고 했을 때 다음이 성립
  - $P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$



베이즈 정리

- 스팸 메일 필터 문제
  - 수신 메일의 내용에 Free라는 단어가 있을 때 이 메일이 스팸 메일일 확률은?
    - $P(Spam) = \frac{3}{10}$
    - $P(Free) = \frac{4}{10}$
    - $P(Spam|Free) = \frac{2}{3}$

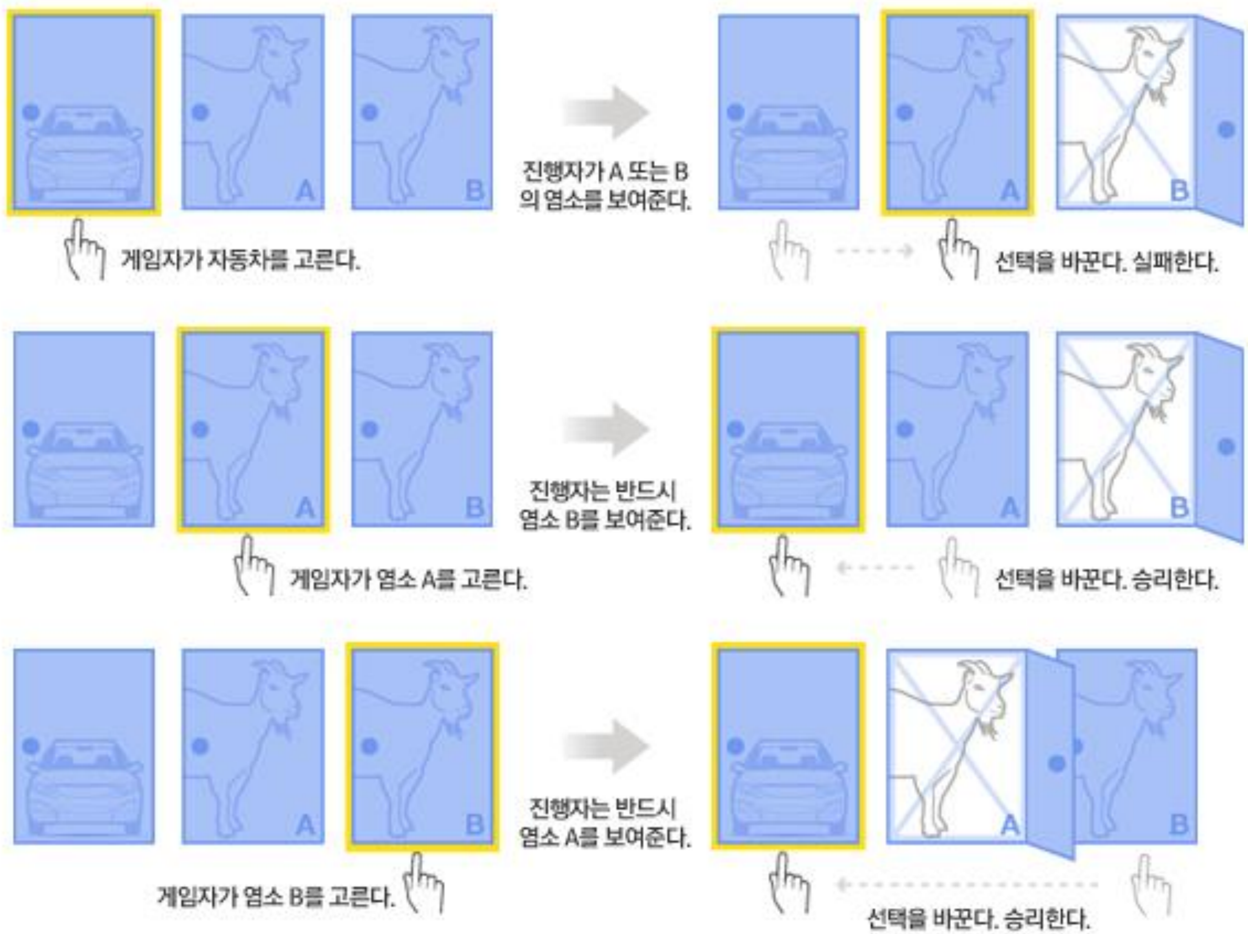


베이즈 정리

- 몬티홀 문제
  - 내용
    - 세 개의 문 중에 하나를 선택하여 문 뒤에 있는 선물을 가질 수 있는 게임 쇼
    - 한 문 뒤에는 자동차, 나머지 두 문 뒤에는 염소가 있음
    - 선택을 완료한 참가자에게 진행자는 염소가 있는 문을 하나 열어주며 선택을 바꾸겠냐고 묻는다.
    - 참가자가 자동차를 가지려고 할 때 원래 선택했던 번호를 바꾸는 것이 유리할까?

베이즈 정리

- 몬티홀 문제
  - 선택을 바꿀 것인가?



감사합니다!