확률 및 통계학

- 2장 확률(Probability) -

명 세인(sein@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

- 표본 공간
- 사상
- 경우의 수
- 사상의 확률
- 가법 정리
- 조건부확률, 독립사상, 승법정리
- 베이즈 정리

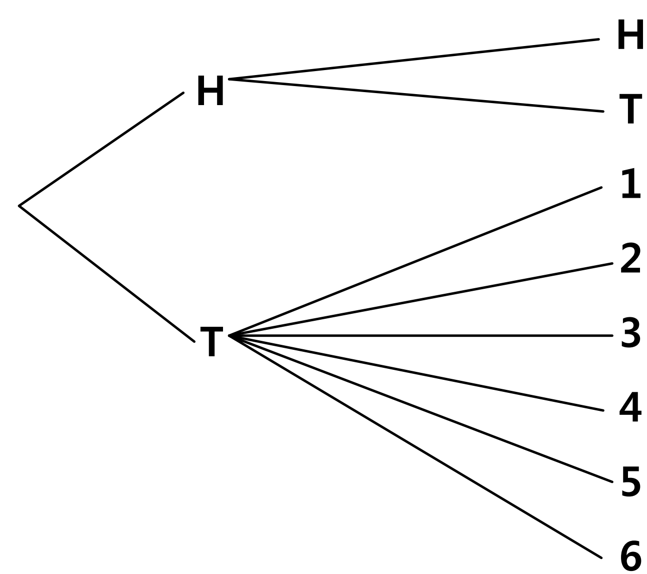
표본 공간

- 표본공간(Sample Space)
  - 통계적 실험에서 발생가능한 모든 결과들의 집합
  - 기호 S로 표기
  - 표본공간의 각 결과는 원소(Element or Member)또는 표본점 (Sample Point)

  - 수형도(Tree Diagram)
    - 표본공간을 트리구조로 표현하여 가짓수를 헤아릴 수 있는 그래프
    - 표본점으로 요약하여 표현가능

표본 공간

- 표본공간(Sample Space)
  - 식 표현
    - S = {앞, 뒤}
    - S = {x | x는 인구 백만이 넘는 도시}
    - S = {(x,y) | $x^2 + y^2 \le 4$ }
  - 수형도 표현
    - 동전을던져 앞면이면 동전을
    - 뒷면이면 주사위를



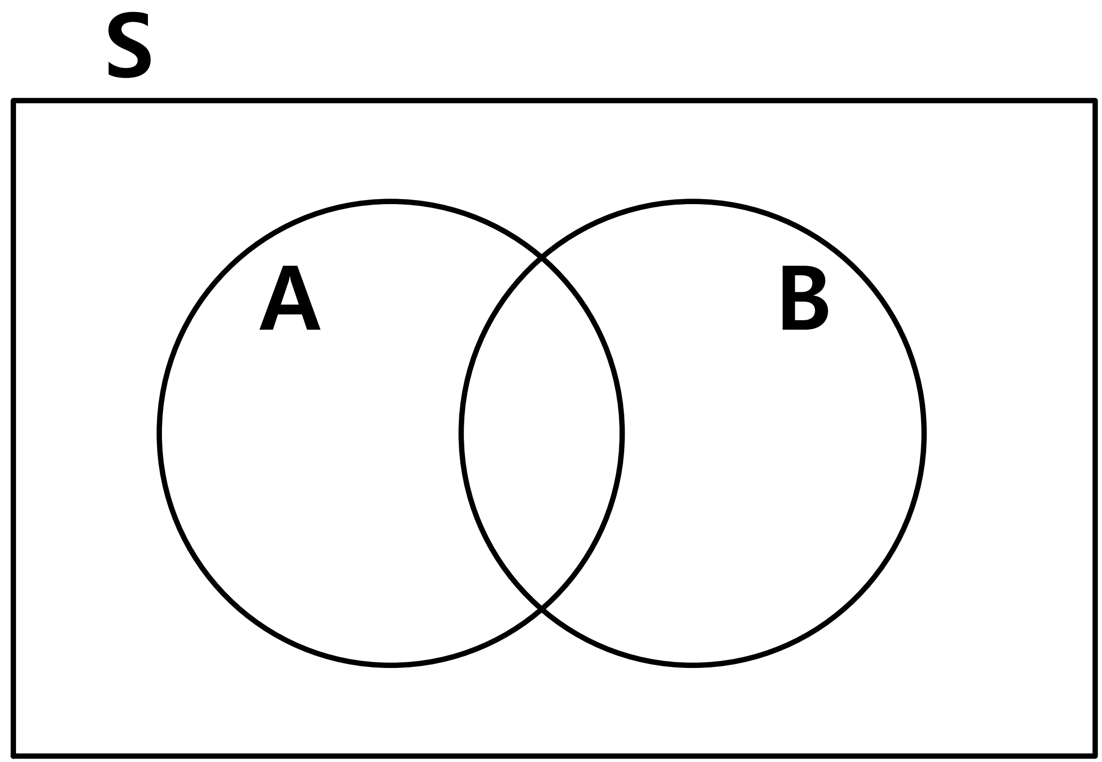
사상

- 사상
  - 표본 공간의 부분집합
  - 사건(Event)과 같은 의미

  - 표본공간과 사상의 관계를, 벤 다이어그램을 이용한 집합개념의 그래프로 표현
    - 공집합(Empty Set)
    - 여집합(Complement)
    - 교집합(Intersection)
    - 합집합(Union)
    - 상호배반(Mutually Exclusive or Disjoint)

사상

- 사상
  - 벤다이어그램(Venn Diagram)
    - $A \cup B$
    - $A \cap B$
    - $\emptyset$
    - A^c
    - $A \cap B = \emptyset$



사상

- 사상
  - 사건의 해석 문제
    - 원점에 중심을 두고 반경이 3인 원의 내부에서 제 1사분면에 있는 점들로 구성된 표본 공간S를 수식을 사용하여 표현하라

경우의 수

- 경우의 수
  - 실험이 수행되며 특정 사상에 관련된 결과의 수를 계산
  - 곱의 법칙
    - 첫번째 시행의 경우의 수와 두번째 시행의 **경우의수를 곱**하여 결정할 수 있는 **총 가지수를 계산**
    - 주사위를 두 번 던져 나올 수 있는 사건의 수
      - 첫번째 n1 = 6
      - 두번째 n2 = 6
      - 사건의 수: $n1 \times n2 = 36$

경우의 수

- 경우의 수
  - 순열(Permutation)
    - 대상물 집합의 **원소가 중복이 없을 때**, **선택하여(n) 나열(r)하는 가짓수(nPr)**
      - 카드가 섞여서 나올 수 있는 가짓수
      - $nPr = \frac{n!}{(n-r)!}$
    - 원순열(Circular Permutation)
      - **원형으로 나열**하는 가짓수, 한 원소를 기준점으로 선정하고 나열
      - $(n-1)!$
    - 분할(Partition)
      - 대상물 집합에 중복이 존재하는경우, **부분집합을 나열**하는 경우
      - $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\ldots n_r!}$

경우의 수

- 경우의 수
  - 순열의 해석
    - 단어 infinity로 구성할 수 있는 단어의 재배열 순열은?

경우의 수

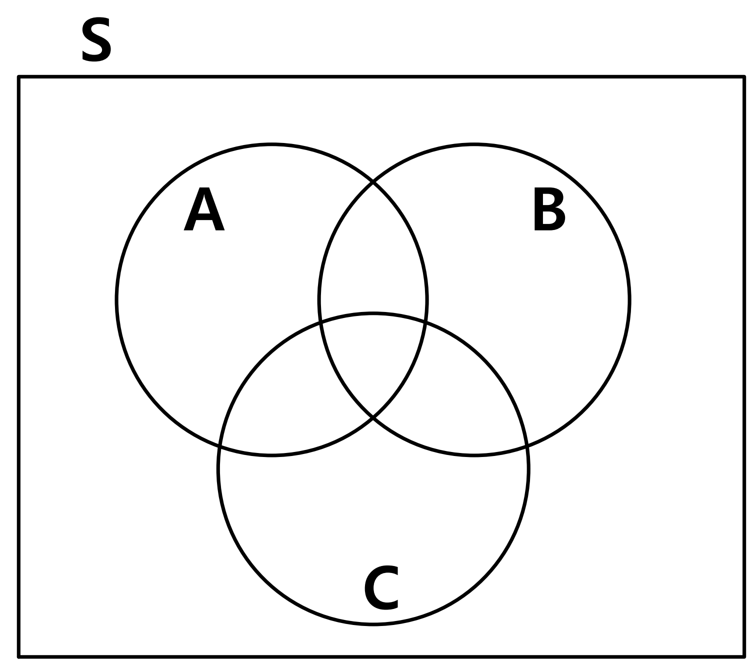
- 경우의 수
  - 조합(Combination)
    - 전체 집합의 전체 또는 일부의 **비순서적 배열**
    - $\boldsymbol{nCr = \frac{n!}{r!(n-r!)} = \frac{n(n-1)\ldots(n-r+1)}{r!} = \frac{nPr}{r!}}$
  - 조합의 해석
    - 10개의 아케이드 게임에서 3게임, 5개의 스포츠 게임에서 2게임 총5게임을 뽑는 가짓수는?

사상의 확률

- 사상의 확률
  - 확률(Probability) 또는 가중치(Weight)를 의미
  - 사건이 발생할 가능성을 나타내는 0~1사이의 값
    - P로 표기
    - 표본공간 S의 확률은 1, P(S) = 1
    - 배반 사건은 각사건의 확률이 사건의 합의 확률과 같다
      - $P(E_1 + E_2 + E_3) = P(E_1) + P(E_2) + P(E_3)$
  - 확률 해석
    - 두번의 동전을 던져 앞면을 볼 확률은?

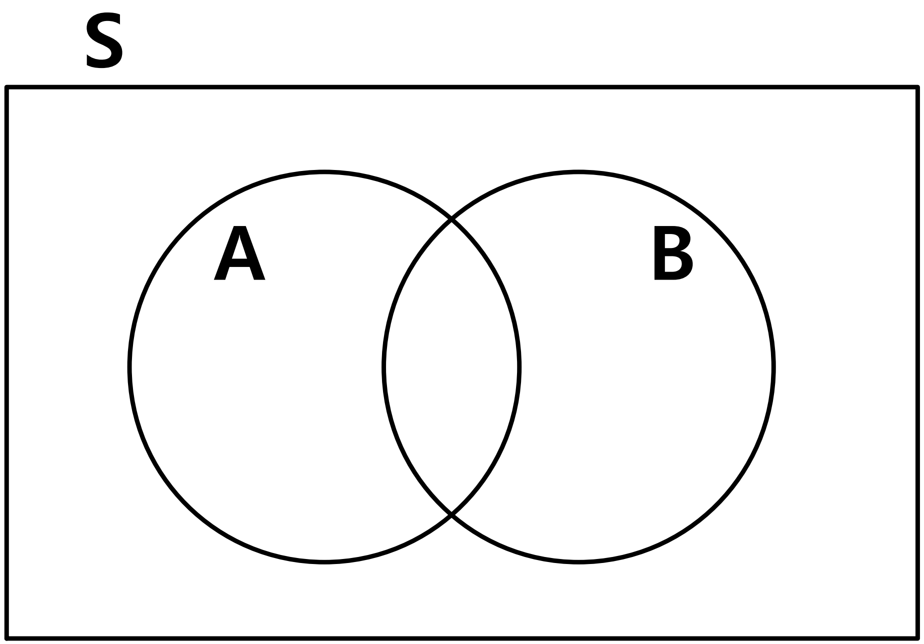
가법정리

- 가법정리
  - 확률의 계산 방법을 정리
  - 합집합 등의 계산규칙에 응용
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - $P(A \cap B) = \emptyset$이면 배반 사건
    - $P(A \cup B \cup C) = ?$
      - $P(A \cup B \cup C) = P(A \cup B) \cup P(C)$를 이용



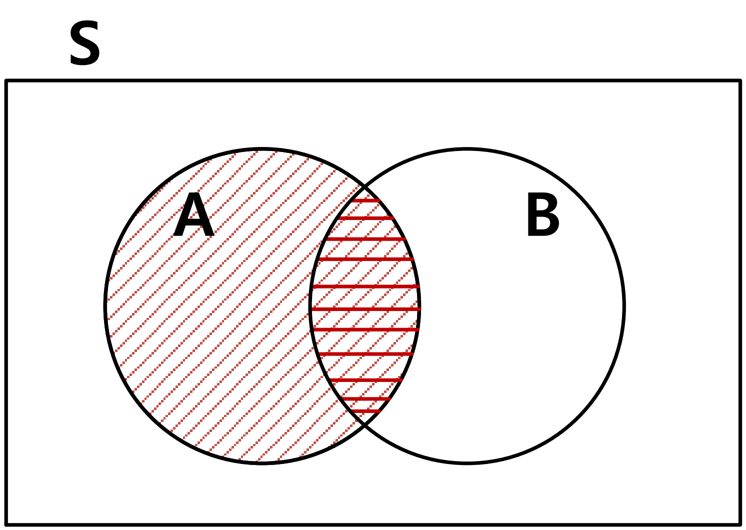
가법정리

- 가법정리
  - 여사건의 확률
    - $P(A) = 1 - P(A^C)$
  - 여사건 정리
    - $P(B) = P(A \cap B) + P(A^C \cap B)$
    - $P(B - A) = P(B) - P(A \cap B)$



조건부 확률, 독립사상, 승법정리

- 조건부 확률(Conditional Probability)
  - 사건 A가 일어날 때 사건 B가 나타날 확률
  - P(B | A)로 표기
    - 식: $P(B|A) = \frac{P(B \cap A)}{P(A)},\ P(A) > 0$
    - 벤다이어그램



조건부 확률, 독립사상, 승법정리

- 조건부 확률(Conditional Probability)
  - 조건부 확률의 해석
    - 180명 대상으로한 조사

	비흡연	적당한 흡연	과다흡연
고혈압	21	36	30
정상	48	26	19

    - 과다흡연자일때 고혈압 환자일 확률
    - 고혈압 환자가 아닐때 비흡연자일 확률

조건부 확률, 독립사상, 승법정리

- 독립사건(Independence)
  - 두 사건 A, B에 대하여 A와 B의 확률이 상호 관계가 없는경우
  - 다음 식이 성립
    - $P(B) = P(B|A) = P(B|A^C)$
    - $P(A) = P(A|B) = P(A|B^C)$
  - 관계가 있는경우 종속(Dependence)
    - 종속인지 독립인지를 판단하기위한 **독립의 필요충분조건식**
      - $P(A \cap B) = P(A)\,P(B)$
      - $P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$

조건부 확률, 독립사상, 승법정리

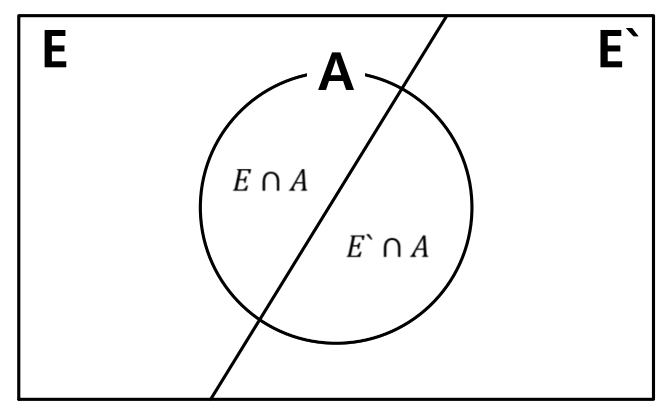
- 승법정리(Multiplicative Rule)
  - 조건부 확률 공식에 대하여 확률계산특성을 이용해 두 사건이 함께 발생할 확률을 구함
    - 두 사건 A,B가 동시에 발생할 수 있으면,

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

    - $P(A \cap B) = P(B \cap A)$이므로 동시에 발생하는 두 사건에 대한 식은 동일
  - 승법정리의 해석
    - 퓨즈 20개중 5개가 불량품이고, 비복원으로 2번 추출할 때, 둘 다 불량품일 확률은?

베이즈 정리

- 전확률(Total Probability)의 법칙
  - $P(A) = (E \cap A) \cup (E` \cap A)$
  - $P(A) = P[(E \cap A) \cup (E` \cap A)] = P(E \cap A) + P(E` \cap A)$
    $= P(E)\, P(A|E) + P(E`)P(A|E`)$
  - 위식을 일반화 하여 전확률 정리식을 도출
    - 사건 $B_1, B_2, \ldots, B_k$를 표본공간 S의 분할이고, $P(B) \neq 0, i = 1, 2, \ldots, k$이면, S의 임의의 사건 A는 다음식이 성립
      - $P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$



베이즈 정리

- 전확률(Total Probability)의 법칙
  - 전확률의 해석
    - 3대의 기계 B_1, B_2, B_3가 각각 전체 생산량의 30%, 45%, 25%를 생산하는 조립공장에서 불량품 제조율이 각 각 2%, 3%, 2%이면, 완제품을 선택하는경우 불량품일 확률은?

베이즈 정리

- 베이즈 정리
  - 베이즈 정리
    - 특정 사후확률$P(B_r|A)$구할 때, 조건부 확률 식($P(B|A) = \frac{P(B \cap A)}{P(A)}$) 으로 바로 구할 수 없을 때,
    - $P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)\, P(A|B_r)}{\sum_{i=1}^{k} P(B_i)\, P(A|B_i)}$

베이즈 정리

- 베이즈 정리
  - 베이즈 정리해석
    - 20P 불량품이 선택된 경우각 공장에서 제조되었을 확률은?

베이즈 정리

- 베이즈 정리
  - 베이즈 정리해석
    - 40세 이상인 성인이 암을 가질 확률: 0.05
      의사가 암환자를 암이라고 진단할 확률: 0.78
      암환자가 아닌 사람을 암이라고 진단할 확률; 0.06
      - 40세 이상인 사람이 암으로 진단될 확률은?
      - 진단받은 암 환자가 실제 암 환자일 확률은?

베이즈 정리

- 베이즈 정리
  - 베이즈 정리해석
    - 라텍스 페인트와 세미 페인트를 판매하는 상점의 통계
      - 라텍스를 구매할 확률은 0.75
      - 라텍스 구매고객의 60%는 롤러를 구입
      - 세미 페인트 구매고객은 30%만 롤러를 구입
      - 특정 고객이 롤러와 페인트를 구매했을 경우 라텍스 페인트일 확률은?

베이즈 정리

- 베이즈 정리
  - 베이즈 정리해석
    - 몬티홀 문제
      - 각각 자동차와 두 염소가있는 문 3개
      - 사회자는 정답을 앎
      - 선택자가 선택한 후 사회자는 염소를 보여줄 때 선택을 바꿀수 있다면 바꾸는것이 확률이 좋은가?

감사합니다!