확률통계 및 공통 알고리즘

1. 확률의 개념

임채영

서울대학교

이번강의에서 다룰 내용

- 집합 - 확률을 소개하기 위한 집합의 정의, 벤다이어그램, 집합의 법칙
- 확률 - 확률의 공리, 조건부 확률, 확률의 법칙, 독립, 베이즈정리

집합의 개념

- 집합(Set)이란, 주어진 성질을 만족시키는 개체들의 모임이라고 할 수 있다. 그리고 이러한 개체들을 원소(element)라고 한다.
- 일반적으로 특정 개체가 집합의 원소인지의 여부는 항상 명확해야 한다.(속하거나, 속하지 않거나)
- 집합의 원소 개수가 1개보다 많을 때, 원소들은 서로 달라야 한다. 그리고 같은 원소가 여러 개 있을 수는 없다.

집합의 표현

집합을 표현하는 방법은 크게 두 가지로 나눌 수 있다.

- 원소 나열법 : 집합의 원소를 나열하여 집합을 표현하는 방법
  - 예시: $A = \{1, 2, 3, 4, 6, 12\}$
- 조건 제시법: 조건을 제시하여 집합을 표현하는 방법
  - 예시: $A = \{x | x\text{는 12의 약수이다}\}$

부분집합, 공집합

- 부분집합 (subset)
  - $A \subset B$
  - A는 B의 부분집합
  - $x \in A \Rightarrow x \in B \text{ for all } x$
  - $A \subset B \text{ and } B \subset A, \text{ then } A = B$
- 공집합 (empty set)
  - $\varnothing$
  - 공집합은 모든 집합의 부분집합이다. $\varnothing \subset A$

교집합, 합집합

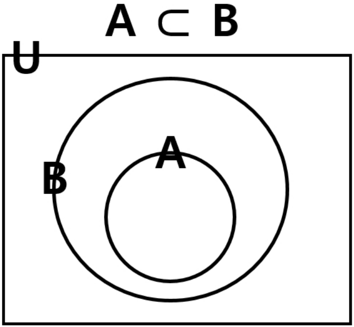
- 교집합 (intersection)
  - $A \cap B$
  - $A \cap B = \{x \mid (x \in A)\ and\ (x \in B)\}$
  - $A \cap B = \varnothing$, A 와 B 는 서로소 (disjoint)
- 합집합 (union)
  - $A \cup B$
  - $A \cup B = \{x \mid (x \in A)\ or\ (x \in B)\}$

차집합, 여집합

- 차집합 (subtraction)
  - $A - B$
  - $A - B = \{x \mid (x \in A)\ \mathit{and}\ (x \notin B)\}$
- 여집합 (complement)
  - A^c
  - $A^c = \{x \in U \mid x \notin A\} = \{x \in U \mid\ \sim (x \in A)\}$, U 는 전체집합
- $A - B \ = \ A \cap B^c$

벤 다이어그램 (Venn Diagram)

- 집합의 포함관계들을 시각화 하는 데에 사용하는 도표
- $A \subset B$의 벤 다이어그램



예제

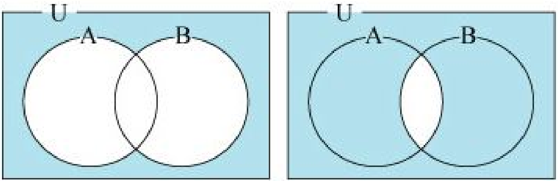
- 두 집합 $A = \{1, 2, 3, 4, 5, 6, 7\}$와 B$\{4, 5, 6, 7, 8, 9, 10\}$가 있을 때, A와 B의 교집합, 합집합, 차집합을 구하여라.

드 모르간의 법칙 (De Morgan's laws)

- 집합들의 다른 두 연산의 결과가 같은 집합을 의미할때
  - 예시
    $(A \cup B)^c = \{x| \sim (x \in A \; or \; x \in B)\}$
    $A^c \cap B^c = \{x|(\sim (x \in A)) \; and \; (\sim (x \in B))\}$

    두 집합의 조건은 같은 조건이므로 (벤다이어그램으로 확인 가능), $(A \cup B)^c = A^c \cap B^c$이 성립. 이를 드 모르간의 법칙이라 한다.

드 모르간의 법칙에 대한 벤 다이어그램



$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$

집합의 법칙

- 교환법칙 (Commutative laws):
  $A \cup B = B \cup A$, $A \cap B = B \cap A$
- 결합법칙 (Associative laws):
  $(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
  $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$
- 분배법칙 (Distributive laws):
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

확률(Probability)이란?

어떤 사건(event)이 일어날 가능성을 나타내는 개념

확률을 정의하기 위해 필요한 개념소개

- 표본공간(Sample Space, $\mathcal{S}$) : 어떤 시행 (Experiement)에서 얻을 수 있는 가능한 모든 결과(outcome)들의 집합
  예시 : 하나의 주사위를 던지고, 나오는 눈의 수를 관찰할 때 표본공간 $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$
- 사건(event, 사상) : 표본공간의 부분집합으로 보통 집합 $A, B, C, \cdots$ 등으로 표현

표본공간과 사건 : Example

- 두 개의 동전을 동시에 던져서 나오는 면의 순서쌍 (앞면 H, 뒷면 T)
  - 표본공간, $\mathcal{S} = \{(H,H),(H,T),(T,H),(T,T)\}$
  - 앞면이 적어도 한번 나오는 사건을 A 라 하자.

$$A = \{(H,H),(H,T),(T,H)\}, \; A \subset \mathcal{S}$$

- 고객센터에 전화를 했을때 기다려야 하는 시간을 조사하기 위해 한 명의 고객이 기다린 시간(분)을 관측할때,
  - 표본공간 $\mathcal{S} = \{t | t \geq 0\}$
  - 기다린 시간이 3분 이상인 사건을 B 라하자.
  - $B = \{t \,|\, t \geq 3\}$

사건의 연산

집합 연산의 기호를 사용

- 합사건 : $A \cup B$
- 곱사건 : $A \cap B$
- 여사건 : A^C
- 배반사건 : $A \cap B = \emptyset$ 이면 A와 B는 서로 배반

확률의 정의 - 등확률모형의 경우

사건 A가 일어날 확률 $P(A)$는 다음과 같이 정의

$$P(A) = \frac{\text{사건 A에 속하는 원소의 개수}}{\text{표본공간 전체의 원소의 개수}}$$

- 예) 두 개의 동전을 동시에 던졌을 때, 앞면이 적어도 한 번 나올 확률

확률측도를 통한 확률의 정의

- 다음과 같은 성질을 만족하는 $P(\cdot)$를 확률측도(Probability Measure) 라고 한다 (Axiom of Probability)

(1) 표본공간 $\mathcal{S}$에서 임의의 사건 A에 대하여 $0 \leq P(A)$

(2) $P(S) = 1$

(3) 서로 배반인 사건 $A_1, A_2, \ldots$에 대하여

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

위의 공리로 부터 나오는 성질

- $P(\emptyset) = 0$
- $A \subset B$ 이면 $P(A) \leq P(B)$
- $0 \leq P(A) \leq 1$
- $P(A^C) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

예제

다섯개의 주사위를 던져 숫자의 합이 7 이상인 확률은?

조건부 확률(Conditional probability)

- 사건 A가 주어졌을 때 사건 B의 조건부확률은 $P(B|A)$로 나타내고 $P(A) > 0$이라는 가정하에 다음과 같이 정의

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

- 사건 A를 새로운 축소된 표본공간으로 간주했을 때, 사건 B가 일어날 확률
- 예제 : 세개의 동전을 차례로 던지는 경우, 앞면이 나온 수가 2 (A)일때, 첫번째 던지기에서 앞면이 나올 (B) 확률은?

곱셈법칙

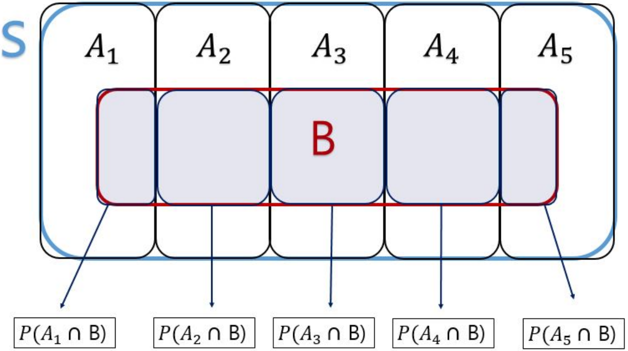
$P(A) > 0, P(B) > 0$이면

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

예제 : 빨간 구슬 10개와 파란구슬 90개가 들어있는 상자에서 2개를 단순 랜덤추출할 때, 2개 모두 빨간구슬일 확률을 구하여라.

전확률공식 (law of total probability)

어떤 사건 B 의 확률 $P(B)$을 구할때, 표본공간의 분할정보를 이용하는 공식



$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$$
$$+P(B|A_4)P(A_4) + P(B|A_5)P(A_5)$$

전확률공식

- 표본공간 $\mathcal{S}$의 분할 $\{A_1, \cdots, A_n\}$을 생각하자. 표본공간의 분할은 다음을 만족한다.

  $$A_i \cap A_j = \emptyset \ (i \neq j),\ A_1 \cup A_2 \cup \cdots \cup A_n = \mathcal{S}$$

- 이때, 전확률공식은

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$$

전확률공식 : 예제

냉장고/김치냉장고 제조 회사는 총 4개의 공장 (A,B,C,D) 을 가지고 있다. 각 공장의 생산량(%)은 다음과 같다. 30% (A), 25% (B), 25% (C) 그리고 20% (D). 그런데 A공장의 50%, B공장의 30%, C 공장의 10%, D공장의 2%가 김치냉장고 생산 비율이라고 하자. 하나의 제품을 단순랜덤추출했을 때 그 제품이 김치냉장고일 확률을 구하여라.

독립(Independence)

서로 독립(mutually independence)

- 사건 A가 일어났다고 하더라도 사건 B가 일어날 확률에 아무런 영향을 미치지 않는 것
- 사건 A와 B가 서로 독립

  $P(B|A) = P(B)$ 또는 $P(A \cap B) = P(A)P(B)$
- 두 사건 A와 B가 독립이 아니면 종속이라고 한다.
- 참고 : $A \cap B = \varnothing$인 두 사건 A와 B는 서로 배반(mutually disjoint), 즉 두 사건이 동시에 일어날 수 없음을 의미하고 A와 B는 종속 사건이다.

- 사건 $A_1, \ldots A_n$이 서로 독립(mutually independent)이다 $\Leftrightarrow$

$$\forall 1 \le i_1 < \cdots < i_k \le n \; (2 \le k \le n)$$

$$P(\cap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$$

- A, B가 독립사건이면 A^C, B (또는 A, B^C 또는 A^C, B^C)가 독립

예제

불량품 20개와 정상품 80개로 구성된 로트(lot)에서 2개의 제품을 단순랜덤추출할 때, 첫번째 제품이 불량품일 사건을 A, 두번째 제품이 불량품일 사건을 B라 하면 A와 B는 독립인가?

베이즈 정리(Bayes Theorem)

- $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$
- 두 사건 A, B의 확률 P(A),P(B), 조건부 확률 $P(B|A) = \frac{P(A \cap B)}{P(A)}$ 를 알고 있을때, $P(A \mid B)$를 구함.
- $P(B)$ 는 전확률공식을 이용하여 $P(B|A)P(A) + P(B|A^c)P(A^c)$로 바꿀수 있다.

베이즈 정리 : 예제

- 전체 국민의 0.1% 가 걸리는 특정 질병에 대해 질병에 걸렸는지 판정하는 검사법이 있다고 하자. 질병이 없는데 있다고 틀리게 판정할 확률이 5%, 질병이 있는데 질병이 있다고 맞게 판정할 확률은 100% 라고 할때, 검사법에 의해 질병이 있다고 판정받은 사람이 실제로 질병이 걸린 사람의 확률은?