

확률통계 및 공통 알고리즘

1. 확률의 개념

임채영

서울대학교

이번강의에서 다룰 내용

- ▶ 확률의 공리, 조건부 확률, 확률의 법칙, 독립, 베이즈정리

확률(Probability)이란?

- ▶ 어떤 사건(event)이 일어날 가능성을 나타내는 개념
확률을 정의하기 위해 필요한 개념소개
 - 표본공간(Sample Space, S) : 어떤 시행 (Experiment)에서 얻을 수 있는 가능한 모든 결과(outcome)들의 집합
예 : 하나의 주사위를 던지고, 나오는 눈의 수를 관찰할 때
표본공간 $S = \{1, 2, 3, 4, 5, 6\}$
 - 사건(event, 사상) : 표본공간의 부분집합으로 보통 집합 A, B, C, \dots 등으로 표현

표본공간과 사건 : Example

- ▶ 두 개의 동전을 동시에 던져서 나오는 면의 순서쌍 (앞면 H , 뒷면 T)

- 표본공간, $S = \{(H, H), (H, T), (T, H), (T, T)\}$
- 앞면이 적어도 한번 나오는 사건을 A 라 하면,

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

- ▶ 고객센터에 전화를 했을때 기다려야 하는 시간을 조사하기 위해 한 명의 고객이 기다린 시간(분)을 관측할때,

- 표본공간 $S = \{t | t \geq 0\}$
- 기다린 시간이 3분 이상인 사건을 B 라고 하면,

$$B = \{t | t \geq 3\}$$

사건의 연산

집합 연산의 기호를 사용

- ▶ 합사건 : $A \cup B$
- ▶ 곱사건 : $A \cap B$
- ▶ 여사건 : A^C
- ▶ 배반사건 : $A \cap B = \emptyset$ 이면 A 와 B 는 서로 배반

확률의 정의 - 등확률모형의 경우

사건 A 가 일어날 확률 $P(A)$ 는 다음과 같이 정의

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

- ▶ 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 S 에서 임의의 사건 A 에 대하여 $0 \leq P(A)$

(2) $P(S) = 1$

(3) 서로 배반인 사건 A_1, A_2, \dots 에 대하여

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

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

곱셈법칙

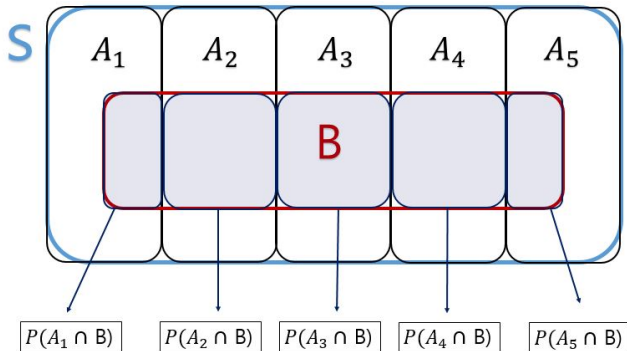
$P(A) > 0, P(B) > 0$ 이면

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

예제 : 불량품 10개와 정상품 90개가 들어있는 상자에서 2개를 단순 랜덤추출(무작위추출, 비복원)할 때, 2개 모두 불량품일 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 $\{A_1, \dots, A_n\}$ 을 생각하자. 표본공간의 분할은 다음을 만족한다.

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

- ▶ 이때, 전확률공식은

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

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%, C공장의 10%, D공장의 2%가 김치냉장고 생산 비율이라고 하자. 하나의 제품을 단순랜덤추출했을 때 그 제품이 김치냉장고일 확률을 구하여라.

독립(Independence)

서로 독립(mutually independence)

- ▶ 사건 A 가 일어났다고 하더라도 사건 B 가 일어날 확률에 아무런 영향을 미치지 않는 것

- ▶ 사건 A 와 B 가 서로 독립

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

- ▶ 두 사건 A 와 B 가 독립이 아니면 종속이라고 한다.
- ▶ 참고 : $A \cap B = \emptyset$ 인 두 사건 A 와 B 는 서로 배반(mutually disjoint), 즉 두 사건이 동시에 일어날 수 없음을 의미하고 A 와 B 는 종속 사건이다.

- ▶ 사건 A_1, \dots, A_n 이 서로 독립(mutually independent)이다
 \Leftrightarrow

$$\forall 1 \leq i_1 < \dots < i_k \leq n \ (2 \leq k \leq n)$$

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

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

예제

불량품 20개와 정상품 80개로 이루어진 100개들이 박스에서 2개의 제품을 단순랜덤추출할 때, 첫번째 제품이 불량품일 사건을 A , 두번째 제품이 불량품일 사건을 B 라 하면 A 와 B 는 독립인가?

베이즈 정리(Bayes Theorem)

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

베イズ 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?