집합

↳ 확률을 수학적으로 정의하기 위해서는 집합론을 알아야 하기 때문에, 집합론의 기초를 복습.

집합과 원소

- 집합: 구별 가능한 객체의 모임
- 원소: 집합에 포함된 객체, 숫자일 필요는 없다.
- 파이썬에서는 집합을 "set"과 "frozenset"으로 나타낸다.
  - set: 내용을 변경할 수 있는 mutable 자료형.
  - frozenset: 내용을 변경할 수 없는 immutable 자료형 => Dictionary 자료형의 키나 set 자료형의 원소가 됨.

집합의 크기

- 집합의 크기(Cardinality, 카디널리티)는 집합이 가지는 원소의 갯수를 말한다. || 기호나 Card 기호를 사용해 나타낸다. 만약 $A=\{1,2,3\}$ 이면, $|A| = card(A) = 3$
- 집합의 크기 in python: len(A)
- 두 실수 간에 항상 다른 실수가 존재하므로 다음과 같은 실수구간 집합은 무한개의 원소를 가짐.

합집합: 각 집합의 원소를 모두 포함하는 집합을 말하고 $\cup$로 표기 => $A \cup B$ (파이썬: union or |)

교집합: 두 사건의 모두에 속하는 원소로만 이루어진 집합, $\cap$로 표기 => $A \cap B$ (파이썬: intersection or &)

부분집합: 어떤 집합의 원소중 일부만을 포함하는 집합 (ex. $A \subset \Omega$)

전체집합: 원래의 집합(Ω), 모든 집합은 자기자신의 부분집합.

차집합: 어떤 집합 A에 속하면서 다른 집합 B에는 속하지 않는 원소로 이루어진 A의 부분집합

$A-B$ => 차집합(difference)

여집합: 전체집합 Ω 중에서 부분집합 A에 속하지 않은 원소로만 이루어진 부분집합을 여집합(Complement)

A^c => 여집합(complement)

$A^c = \Omega - A$

공집합: 아무런 원소도 포함하지 않는 집합(공집합), $\emptyset$

↳ 공집합은 모든 부분의 부분집합이 된다.

$\emptyset \subset A$, for all A.

$A \cap \emptyset = \emptyset$

$A \cup \emptyset = A$

$A \cap A^c = \emptyset$

부분집합의 수

↳ 집합 $A=\{1,2\}$는 다음과 같이 4개의 부분집합을 가진다.

$$A_1=\{\},\ \emptyset$$
$$A_2=\{1\}$$
$$A_3=\{2\}$$
$$A_4=\{1,2\}$$

↳ 일반적으로, 원소의 개수가 N개인 집합은 2^N개의 부분집합을 가진다.

연습문제 1

$$\Omega=\{HH, HT, TH, TT\}$$

1. 이 집합의 부분집합의 갯수는? $2^4=$ 16개
2. 이 집합의 모든 부분집합의 frozenset 자료형 객체를 만들고, 이를 원소로 가지는 set 객체를 만든다.

합집합과 교집합의 분배 법칙

곱셈의 분배법칙과 덧셈 법칙은 아래와 같다.

$$a\times(b+c)=a\times b+b\times c$$

↳ $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$

↳ $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$

연습문제 2

다음 세 집합 A, B, C에 대해 두 가지 분배 법칙이 성립하는지 증명하라.

$$A=\{1,3,5\}$$
$$B=\{1,2,3\}$$
$$C=\{2,4,6\}$$

확률의 수학적 정의

↳ 확률의 수학적 정의를 아는 것은 확률과 관련된 복잡한 문제를 푸는 것 뿐 아니라
지금까지 생각해오던 확률에 대한 관점을 새롭게 환기시키는데 도움이 된다.

확률 표본 (Sample)

↳ 풀고자 하는 확률적 문제에서 발생할 수 있는 하나의 현상, 혹은 선택될 수 있는 하나의 경우.

표본공간 (Sample space)

↳ 값이 될 수 있는 혹은 선택될 수 있는 모든 표본의 집합. 보통 Ω (대문자 오메가)로 표기.

- 표본공간을 정의한다는 것은 어떤 현상이 가능하고 어떤 현상이 가능하지 않은가를 정의하는 작업

연습문제 1

1. 주사위를 던져 나오는 숫자를 구하는 문제를 확률론적으로 접근할 때 표본공간 Ω_d를 구하라.

⇒ If dice is a fair dice, then $2^6 = 64$

2. 약속 날짜를 정하기로 했다. 결정된 날짜가 며칠인가 아닌가 구하는 문제를 확률론적으로 접근할 때, 표본공간 Ω를 구하라.

⇒ 날짜라는 개념이 있으면, $\frac{n}{31}$ or, $\frac{7}{365}$ or 윤달이면 그것도 포함.

연습문제 2

1. 잠을 자다가 깨어나 시계를 보았을 때, 분침이 가리키는 각도를 표본이라고 하면, 표본공간 Ω_1은? 표본공간의 표본개수는?

↳ ~~표현~~ "$0° \leq x \leq 360$"

↳ "무제한"

2. 측정한 체온을 표본이라고 하면, 표본공간 Ω_a? 표본공간의 표본개수는?

↳ "무제한"

↳ "무제한"

사건 (event, 표본공간 Ω의 부분집합)

↳ 즉, 전체표본공간 중에서 우리가 관심을 가지고 있는 일부표본의 집합.

↳ $A = \{\} = \phi$

$B = \{H\}$

$C = \{T\}$

$D = \{H, T\} = \Omega$

— 예로든 B라는 사건은 {동전 앞면이 나오는 경우}

D라는 사건은 {동전 앞면이나 뒷면이 나오는 경우}

확률

↳ 사건을 입력하면 숫자가 출력되는 함수
(부분집합)

사건(부분집합) ⇒ 숫자.

↳ 세가지 규칙을 지켜야 한다.

i) 모든 사건에 대해 확률은 0 또는 양수이면서 실수이다. $P(A) \in \mathbb{R}, P(A) \geq 0$

ii) 표본공간이라는 사건에 대한 확률은 1이다. $P(\Omega) = 1$

iii) 공통원소가 없는 두사건의 합집합의 확률은 각각의 사건의 확률의 합이다.

↳ $A \cap B = \phi$, $P(A \cup B) = P(A) + P(B)$

이 세가지를 "콜모고로프의 공리(Kolmogorov's axioms)라고 한다.

확률은 표본이 아닌 사건을 입력으로 가지는 함수

— 확률은 표본이 아닌 사건(부분집합)에 대해 정의

$\cancel{P(6) = \frac{1}{6}}$, $P(\{6\}) = \frac{1}{6}$

— 표본의 개수가 유한하고, 각 사건에 대한 원소의 개수 이외에 다른 정보가 없으면, 각 사건의 확률은 원소의 개수에 비례한다고 보는 것이 타당하다.

$$P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$$

↳ 그러나, 만약 다른 곳에서 얻은 정보(Domain 지식...)이나 표본에 대한 Data가 존재하는 경우에는 보다 믿을 수 있는 확률값을 계산할 수 있다.

단순사건과 확률질량함수

- 표본 1개만 가지는 사건 : 단순사건(elementry event, atomic event)
- Atomic event(단순 사건)은 교집합을 가지지 않으므로 유한개의 사건이 있는 경우, 모든 단순사건의 확률값을 알면 콜모고로프의 세번째 공리에 의해 다른 모든 사건의 확률값계산가능

for example $p(\{H\}) = 0.1$, $p(\{T\}) = 0.2$, $p(\{Q\}) = 0.3$, $p(\{K\}) = 0.4$

$$p(\{T, Q\}) = 0.5$$

- 하지만 표본이 숫자인 경우, 사건 대신 표본을 입력으로 가지고 확률값을 출력으로 가지는 함수 p를 정의 할 수 있는데, 이를 확률 질량 함수(probability mass function)라고 한다. 확률과 확률질량함수 ...

확률 $\Rightarrow p(\{1\}) = 0.2$.

확률질량함수 $\Rightarrow p(1) = 0.2$

확률 $\Rightarrow p(\{1, 2\}) = 0.3$

확률질량함수 $\Rightarrow$ ~~p(~~ $p(1, 2)$: 정의되지 않는다.

표본의 수가 무한한 경우

- 왜 확률을 정의할 때, 입력은 표본이 아닌 사건으로 정의했을까? 다시 말해 왜 확률값을 표본에 할당하지 않고 사건에 대해 할당 했을까? 그 이유는 표본의 수가 무한한 경우를 다루기위해서
- 잠에서 깨어나 시계를 보았을 때, 분침이 이루는 각도를 설정하는 문제를 생각하자. 정각 12시를 가리키면 0도이고 정각 1시를 가리키면 30°이다. 그렇다면, 이 시계바늘 문제에서 분침이 정각 12시를 가리킬 확률, 즉 각도가 정확하게 0도가 될 확률은 0°이다.

$$p(\{\theta = 0°\}) = 0$$

- 각도가 아닌 어떤 경우도 마찬가지다. 예를 들어서 시계바늘이 1시를 가르킬 때,

$$p(\{\theta = 30°\}) = 0$$

- why? 모든 각도가 가능성이 똑같으므로 그 확률은 x라는 값으로 하자.

- 표본의 수가 무한하고, 모든 표본에 대해 표본 하나만을 가질 사건의 확률이 동등하다면, 표본에 하나에 대한 사건의 확률은 언제나 0 이다.
- 이번에는, 시계바늘이 12시와 1시 사이에 있는 경우 $\Rightarrow p(\{0 \leq x \leq 30°\}) = ? = \boxed{\frac{1}{12}}$
- 원소의 수가 무한개인 표본집합의 경우에는 사건의 수도 무한개이므로 확률을 할당하거나 어떻게 할당 했는지 설명하기가 어렵다. 이 경우는 확률밀도 함수를 사용한다.