확률의 기본개념

- 확률(Probability)이란?
  경험 혹은 실험의 결과로 특정한 사건(event)이나 결과가 발생할 가능성

주사위를 던졌을 때, 각 각의 눈이 나올 수 있는 확률은?

확률의 종류

1. 주관적 확률 : 의사결정자의 지식, 정보 및 경험을 바탕으로 한 확률

2. 객관적 확률 : 특정사건이 발생할 가능성이 객관적으로 명확하여 **누구나 쉽게 알 수 있는 경우**나 **동일한 실험을 무수히 반복적으로 수행할 경우** 특정한 사건이 발생할 확률

객관적 확률

고전적 확률

1. 이론에 근거한 사전 확률

2. 모든 결과가 발생할 가능성이 명확하고, 각 결과는 서로 배타적, 즉 서로 다른 2개 이상의 결과가 동시에 발생할 수 없는 경우 활용

장기적 상대도수 확률

1. 실제 실험에 근거한 사후 확률

2. 같은 실험을 반복적으로 무수히 수행 할 경우 특정 사건이 발생할 수 있는 상대적 빈도로 정의

실험과 표본공간

· **실험(experiment)이란** 어떤 행위의 결과를 관찰하고 측정하여 그 결과에 대해 구체적인 값을 부여하는 것을 의미한다.

→ **어떤 일을 수행하고 그 결과를 나온 값을 기록하여 자료로 남기는 행위**

실험결과

- **표본점(sample point)**: 한번의 실험으로부터 얻은 결과
- **표본공간(sample space)**: 실험을 통해서 얻을 수 있는 가능한 표본점들의 집합

한계확률(marginal probability) - 1

· **한계확률이란?** 2개의 변수가 교차되어 만들어진 빈도교차상에서 하나의 관측치가 행이나 열 어느 한 변수에 의해서만 구분되는 특정 집단에 속할 확률을 의미




☞ **전체 20명의 남녀학생들을 대상으로 가장 좋아 하는 과일이 무엇인지를 조사하여 정리한 빈도교차표**

	사과	딸기	포도	수박	행의 합
남	2	2	3	1	8
여	3	4	1	4	12
열의 합	5	6	4	5	20

한계확률(marginal probability) - 2

· **한계확률이란?** 아무런 조건이 없는 상태에서 A라는 사건이 발생할 확률

→ 비조건확률(unconditional probability)

→ P(A)

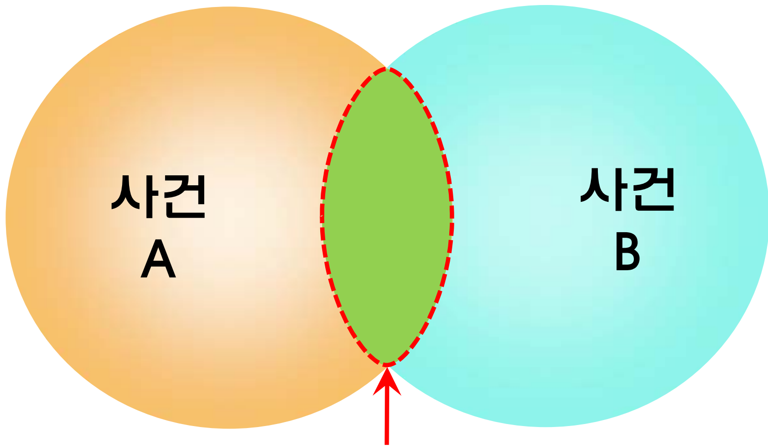
	사과	딸기	포도	수박	행의 합
남	2	2	3	1	8
여	3	4	1	4	12
열의 합	5	6	4	5	20

총 20명의 학생 중에서 1명을 선정할 경우,
이 학생이 사과를 좋아하는 학생일 확률은?

$$\frac{5}{20}$$

결합확률(joint probability)

· **결합확률이란?** 두 개 이상의 사건이 동시에 발생할 가능성을 나타내는 확률



n(A∩B) = 사건 A와 사건 B가 동시에 발생하는 경우의 수

$$P(A\cap B) = \frac{n(A\cap B)}{n(S)} \Rightarrow \frac{\text{사건 A와 사건 B가 동시에 발생하는 경우의 수}}{\text{전체 표본공간의 경우의 수}}$$

조건부 확률(conditional probability)

· **조건부확률이란?** 이미 하나의 사건이 발생한 상태에서 또 다른 사건이 발생할 가능성을 나타내는 확률

$$\frac{\text{사건 A와 사건 B가 동시에 발생할 가능성을 나타내는 결합확률}}{\text{사건 A가 발생할 가능성을 나타내는 한계확률}}$$

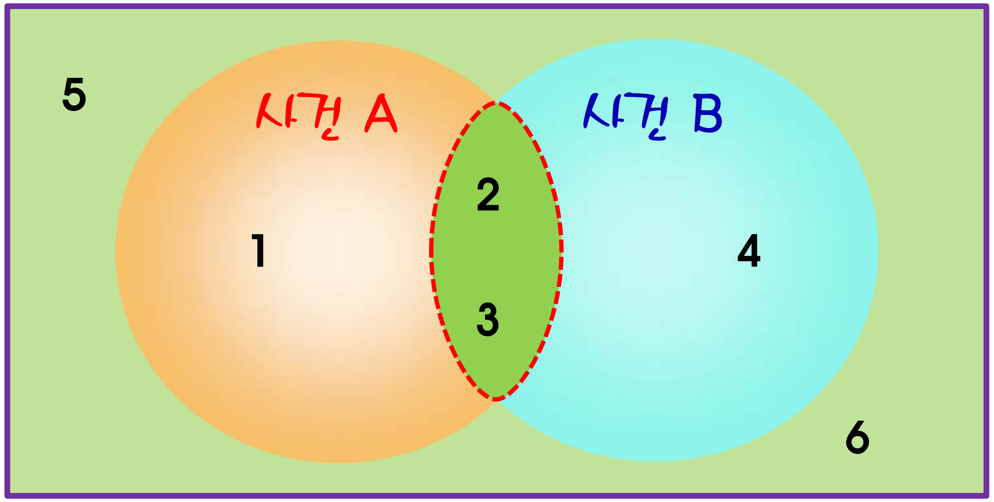
사건 A가 발생한 조건하에서 또 다른 사건 B가 발생할 확률 ⟶ $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

확률의 연산법칙 – 덧셈법칙(집합)

· 표본공간 S = {1, 2, 3, 4, 5, 6}

· 사건 A = {1, 2, 3}

· 사건 B = {2, 3, 4}



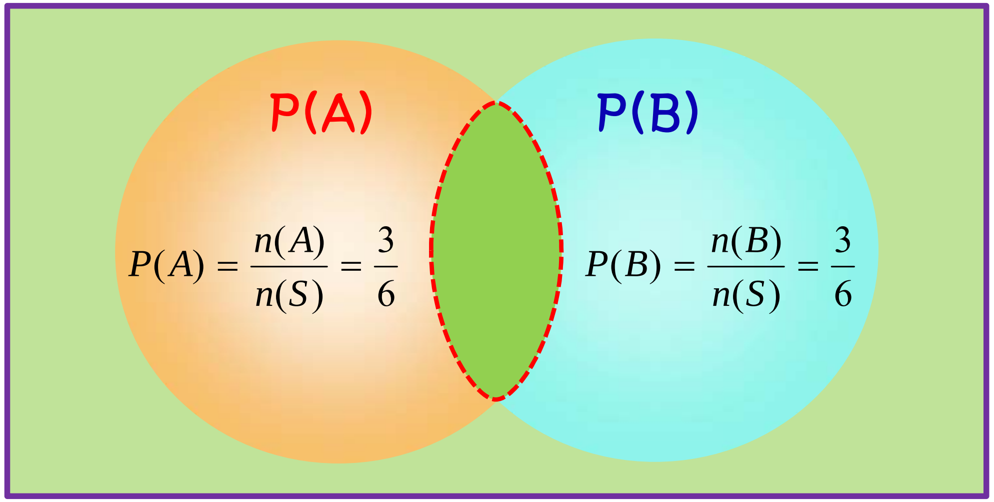
$$(A \cup B) = ?$$

$$(A \cap B) = ?$$

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

확률의 연산법칙 – 덧셈법칙(확률)

· 표본점 수 n(S) = 6, n(A) = 3, n(B) = 3



$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = ?$$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = ?$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

확률의 연산법칙 – 곱셈법칙

· 확률의 곱셈법칙은 앞에서 살펴본 조건부 확률을 이용한다.

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

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

순열(permutation)

· 서로 다른 n개의 개체에서 k개를 선택한 다음, 순서를 고려하여 배열하는 경우의 수를 순열(permutation)이라 하며, n개에서 k개를 선택하여 배열하는 순열을 기호로는 ${}_nP_k$로 표시한다.

$$
{}_nP_k = \frac{n!}{(n-k)!}
$$

테이블 위에 있는 4개의 숫자 중에서 2개를 뽑아서 만들 수 있는 2자리 정수의 개수는?

조합(combination)

· 서로 다른 n개의 개체에서 순서를 고려하지 않고 k개를 선택하고자 하는 경우, k개로 이루어진 집합의 수를 구하는 방법을 n개에서 k개를 택하는 조합이라 하며, 기호로는 ${}_nC_k$ 로 표시한다.

$$ {}_nC_k = \frac{n!}{k!(n-k)!} $$

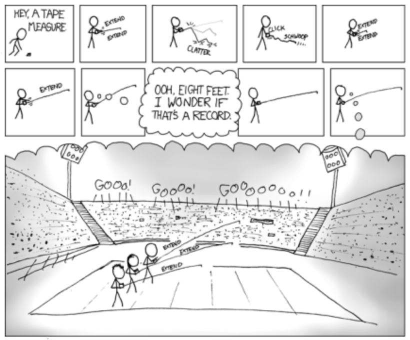




확률과 확률변수

- 확률이란?
  경험 혹은 실험의 결과로 특정한 사건(event)이나 결과가 발생할 가능성

- 변수란?
  관심대상의 속성을 척도를 이용하여 측정한 값들을 대표하는 것



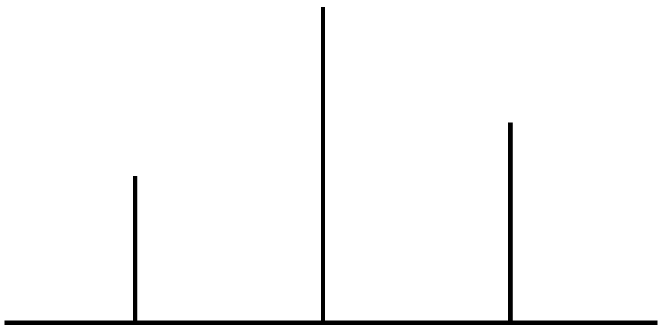
확률변수는 특정한 값을 가질 수 있는 확률이 주어진 변수

확률변수의 종류

이산확률변수

· **정수와 같이 명확한 값을 변수 값**으로 함.

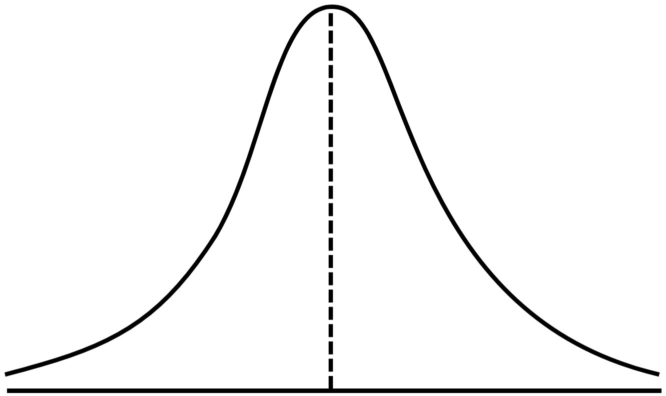
· 확률변수가 가질 수 있는 값의 수가 한정되어 그 수를 셀 수 있는 변수



연속확률변수

· **변수 값이 정수처럼 명확하지 못함**.

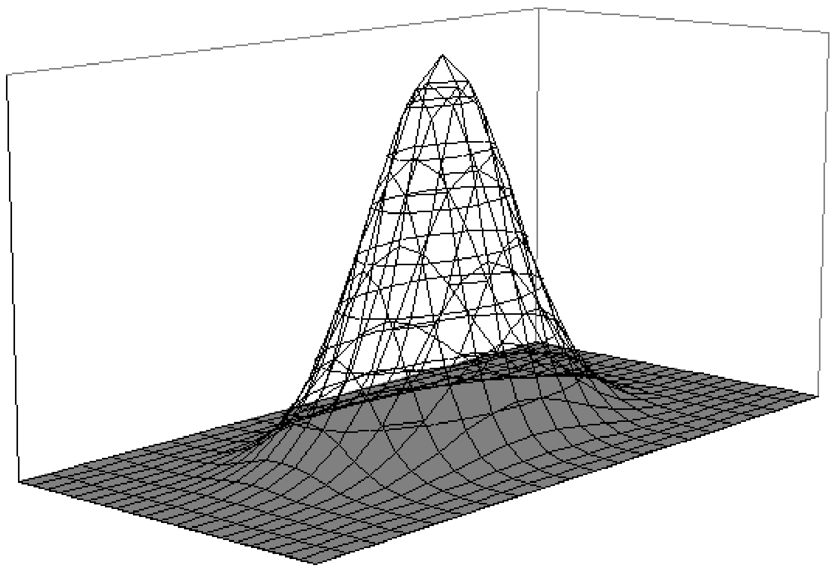
· 확률변수가 연속량으로 표기되어 가능한 변수 값의 개수를 셀 수 없는 변수



확률분포(probability distribution)

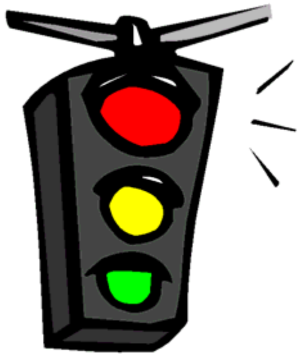
- **확률분포는** 확률변수가 특정한 값을 가질 확률, 즉 상대적 가능성을 나타낸 것으로 모든 가능한 **확률변수값**과 그 값이 발생할 가능성인 **확률값을 도수분포표나 그래프로 나타낸 것이다.**

- **확률함수는** 그래프로 확률분포를 나타내는 경우 확률분포를 나타내는 함수이다.



확률분포의 종류

- 이산확률분포
  - 이항분포
  - 초기하분포
  - 포아송분포



- 연속확률분포
  - 정규(Z)분포
  - t 분포
  - Chi-square 분포
  - F 분포

위의 7가지 분포는 무수히 많은 확률분포들 중에서 분포의 특성을 파악하여 함수로 나타낼 수 있는 대표적인 분포들이다.

확률함수(probability function)

- 확률분포를 함수로 나타낸 것을 확률함수라고 한다.

확률질량함수(p.m.f : probability mass function)

- 확률변수 X가 이산확률변수일 경우, **확률분포함수를 P(x)로 표시**하며, 이를 **확률질량함수**라고 한다.

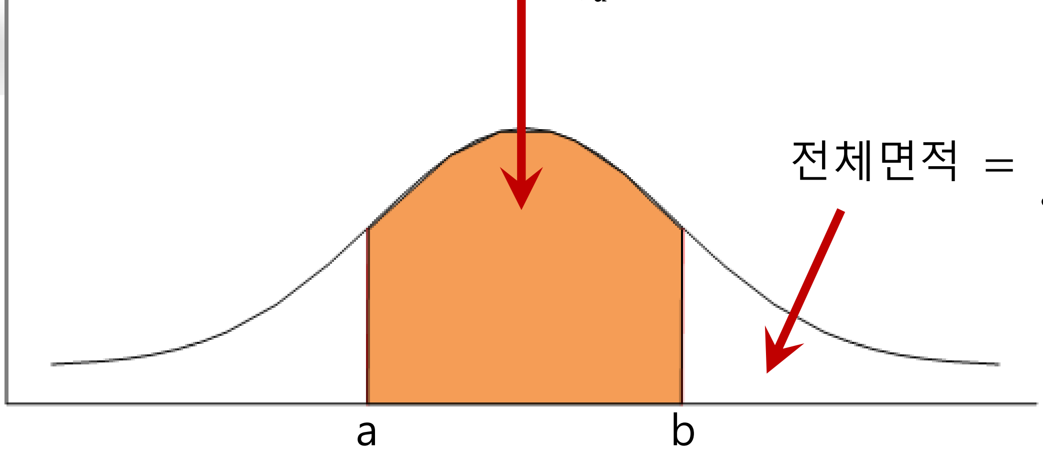
확률밀도함수(p.d.f : probability density function)

- 확률변수 X가 연속확률변수일 경우, **확률분포함수를 f(x)로 표시**하며, 이를 **확률밀도함수**라고 한다.

확률밀도함수의 특성

- 확률은 확률밀도함수를 X축상의 일정구간(a, b)에서 적분한 값이다.
- **전체 확률밀도함수의 면적**, 즉 확률밀도함수를 전구간(-∞, +∞)에서 적분한 값은 **1이다.**

$$P(a \le x \le b) = \int_a^b f(x)\,dx$$



$$\text{전체면적} = \int_{-\infty}^{\infty} f(x)\,dx = 1$$