통계학

– 제 3 장 확률과 확률분포 (2)

0011 0010 1010 1101 0001 0100 1011

확률표본

- 유한 모집단으로부터의 확률표본
  - 예 – [대학신문사의 편집위원 선발]
    - 대학신문사에서 두 명의 편집위원 선발
    - 5명의 지원자 : 여자 3명 + 남자 2명
    - 5명의 지원자 중에서 2명을 임의로 선발한다면,
      남자가 한 명도 선택되지 않을 확률?

      - 통계적 실험 : 5명 중 2명 선발
      - 표본공간 :

$$\Omega = \begin{Bmatrix} (\mathrm{F1},\ \mathrm{F2}),\ (\mathrm{F1},\ \mathrm{F3}),\ (\mathrm{F1},\ \mathrm{M1}),\ (\mathrm{F1},\ \mathrm{M2}),\ (\mathrm{F2},\ \mathrm{F3}), \\ (\mathrm{F2},\ \mathrm{M1}),\ (\mathrm{F2},\ \mathrm{M2}),\ (\mathrm{F3},\ \mathrm{M1}),\ (\mathrm{F3},\ \mathrm{M2}),\ (\mathrm{M1},\ \mathrm{M2}) \end{Bmatrix}$$

      - 5명의 지원자 중에서 남자가 한 명도 선택되지 않는 사건 :

$$A = \{(\mathrm{F1},\ \mathrm{F2}),\ (\mathrm{F1},\ \mathrm{F3}),\ (\mathrm{F2},\ \mathrm{F3})\}$$

      - 5명의 지원자 중에서 남자가 한 명도 선택되지 않을 확률 :

$$P(A) = \frac{3}{10}$$

- 조합법
  - N 개의 다른 객체를 가지고 있는 한 집단으로부터 r 개의 객체를 선택할 때, 가능한 선택방법의 수 :

$$\binom{N}{r} = \frac{N!}{r!(N-r)!} = \frac{N\times(N-1)\times\cdots\times(N-r+1)}{r\times(r-1)\times\cdots\times 2\times 1}$$

- 예 – [대학신문사의 편집위원 선발]
  - 조합법에 의한 계산
    - 표본공간 내의 모든 근원사상 수 :

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5\times 4}{2\times 1} = 10$$

    - 5명의 지원자 중에서 남자가 선택되지 않는 사건의 근원사상 수 :

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3\times 2}{2\times 1} = 3$$

- 확률표본

> 크기가 n 인 모든 묶음들이 모두 똑같은 확률을 갖는다면
>
> N 개의 서로 다른 객체를 가지고 있는 한 모집단으로부터 뽑혀진
>
> 크기 n 의 표본을 확률표본이라 함

  - 표본으로부터 모집단을 일반화하기 위해서는
    모든 표본들이 똑같은 선택의 기회를 가질 수 있도록
    공정한 추출과정이 필요함

확률변수

- 확률변수의 정의
  - 확률변수 (random variable)
    - 각각의 근원사상들에 대하여 하나의 실수값을 대응시켜주는 함수
      - 확률변수 : $X,\ Y,\ Z$
      - 확률변수가 가지는 값 : $x_1,\ x_2,\ \cdots,\ y_1,\ y_2,\ \cdots,\ z_1,\ z_2,\ \cdots$
  - 예 – [동전 던지기]
    - 두 개의 동전을 던지는 실험
    - 확률변수 X : 앞면(H)이 나오는 횟수

근원사상	(H, H)	(H, T)	(T, H)	(T, T)
X 의 값	2	1	1	0

X 의 값	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$
근원사상	(T, T)	(H, T), (T, H)	(H, H)
확률	1/4	1/2	1/4

- 확률변수의 종류
  - 이산확률변수
    - 확률변수가 취할 수 있는 가능한 값들이 셀 수 있는 값
    - 확률변수가 취하는 특정 값에 대한 확률을 계산 가능
    - 예 – 한 자동차 정비공장에서 하루 동안 정비되는 자동차 수
      – 20문제의 퀴즈 중 학생들이 올바르게 답한 문제의 수
  - 연속확률변수
    - 확률변수가 취할 수 있는 가능한 값들이 구간 내의 연속적인 값
    - 확률변수가 취하는 특정 구간에 대한 확률만 계산 가능
    - 예 – 과일 음료 안에 포함된 과일의 함유율
      – 다이어트 프로그램에 따른 몸무게의 감소량

이산확률변수와 확률분포

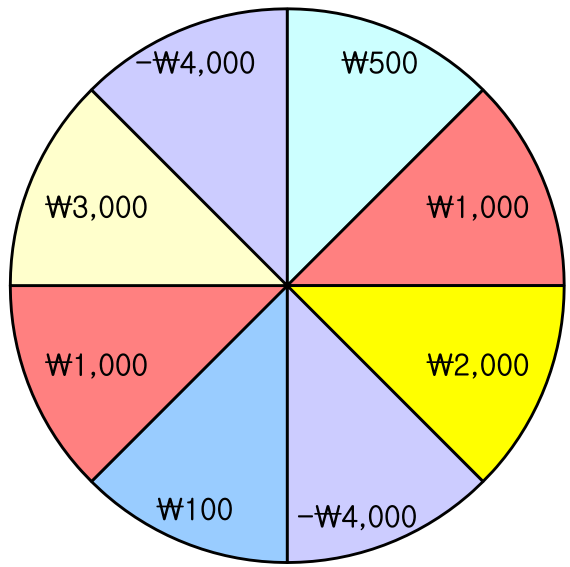
- 확률분포의 정의
  - 확률분포 (probability distribution)
    - 확률변수가 갖는 값들과 그에 대응하는 확률값을 나타내는 것
  - 확률질량함수 (probability mass function, p.m.f.)
    - 이산확률변수 X가 취할 수 있는 값 $x_1,\ x_2,\ x_3,\ \cdots$ 에 대하여 확률 $P(X=x_1),\ P(X=x_2),\ P(X=x_3),\ \cdots$ 를 대응시켜주는 함수

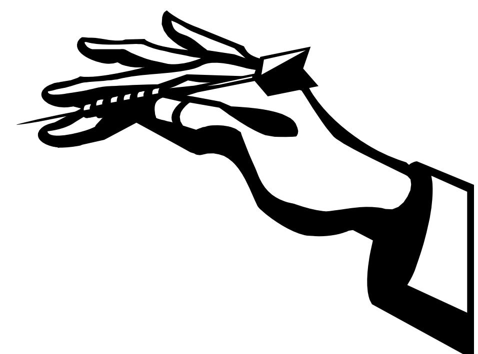
$$\boxed{f(x_i) = P(X = x_i)}$$

    - 확률질량함수의 성질
      - 모든 x_i 값에 대해 $0 \le f(x_i) \le 1$
      - $\sum_{x_i} f(x_i) = 1$

확률변수의 평균과 분산

- 기대값과 분산
  - 확률변수의 기대값 (expected value)
    - 확률변수의 평균
    - 확률변수가 가질 수 있는 값들 중에서 확률분포의 중심 위치
  - 예 – [다트 게임]
    - 다트를 한 번 던졌을 때 생기는 이익 : X





- 이산확률변수의 기대값
  - 확률변수에 대해 확률을 가중치로 하여 계산된 가중평균

$$E[X] = \sum_{x_i} x_i \times f(x_i) = \mu$$

- 예 – [다트 게임]
  - X의 확률분포

X의 값	x_1	x_2	x_3	x_4	x_5	x_6
	100	500	1000	2000	3000	−4000
$f(x_i)$	1/8	1/8	2/8	1/8	1/8	2/8

  - 기대값 : $E[X] = \mu = 100 \times \frac{1}{8} + 500 \times \frac{1}{8} + \cdots - 4000 \times \frac{2}{8} = -50$

- 이산확률변수 함수의 기대값
  - 확률변수 X의 함수 $h(X)$의 기대값

$$E[h(X)] = \sum_{x_i} h(x_i) \times f(x_i)$$

- 예 – [다트 게임]
  - X^2의 확률분포

X^2 의 값	x_1^2	x_2^2	x_3^2	x_4^2	x_5^2	x_6^2
	10,000	250,000	1,000,000	4,000,000	9,000,000	16,000,000
$f(x_i)$	1/8	1/8	2/8	1/8	1/8	2/8

  - 기대값 : $E\left[X^2\right] = 10,000 \times \frac{1}{8} + \cdots + 16,000,000 \times \frac{2}{8} = 5,907,500$

- 확률변수의 분산과 표준편차
  - 확률변수 X의 분산

$$
\begin{aligned}
Var[X] &= E[X - E[X]]^2 \\
&= E\left[X^2 - 2XE[X] + (E[X])^2\right] \\
&= E\left[X^2\right] - 2E[X]E[X] + (E[X])^2 \\
&= E\left[X^2\right] - (E[X])^2 \\
&= E\left[X^2\right] - \mu^2 \\
&= \sigma^2
\end{aligned}
$$

  - 확률변수 X의 표준편차

$$\sigma = \sqrt{Var[X]} = \sqrt{\sigma^2}$$

- 예 – [다트 게임]
  - 분산 :

$$Var[X] = \sigma^2 = E\left[X^2\right] - \left(E[X]\right)^2 = 5,907,500 - 2,500 = 5,905,000$$

  - 표준편차 :

$$\sigma = \sqrt{\sigma^2} = \sqrt{5,905,000} \doteq 2430.02$$