2강 확률분포(1)

학습내용
1. 확률분포함수를 이해한다.
2. 기댓값과 분산을 이해한다.
3. 마코프 부등식을 이해한다.
4. 적률생성함수를 이해한다.

01 확률분포함수

1. 확률변수

- 이산형 확률변수
  - 확률변수의 가능한 값이 유한 개이거나 셀 수 있는 경우
  - 확률질량함수: f(x) = P(X=x)
  - 확률질량함수의 성질
    - $0 \le f(x) \le 1$
    - $\sum_{x=1}^{n} f(x) = 1$
    - $P(a \le X \le b) = \sum_{a}^{b} f(x)$
- 예) 불량품 3개, 양품 2개인 상자에서 2개의 제품을 꺼낼 때 불량품 수의 확률질량함수는?
  - $f(x) = \frac{{}_3C_x \cdot {}_2C_{2-x}}{{}_5C_2}, \quad x = 0, 1, 2$
- 연속형 확률변수
  - 확률변수의 가능한 값이 무한 개이며 셀 수 없는 경우
  - 연속형 확률변수의 분포를 결정하는 함수: 누적분포함수
    - $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt, \quad f(x) = \frac{d}{dx}F(x)$
  - 누적분포함수를 미분하면 확률밀도함수가 된다.
- 예) 10분 간격으로 출발하는 버스를 기다리는 시간의 확률밀도함수는?
  - $F(x) = P(X \le x) = \frac{x-0}{10} = \frac{x}{10} = \int_{0}^{x} \frac{1}{10}dt$
  - 그런데,

    $f(x) = \frac{d}{dx}F(x)$

    이므로

$$f(x) = \begin{cases} \frac{1}{10} & \text{if} \quad 0 \le x \le 10 \\ 0 & \text{if} \quad x < 0 \;\; or \;\; x > 10 \end{cases}$$
이다.

- 확률밀도함수의 성질
  - $0 \le f(x)$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $P(a \le X \le b) = \int_a^b f(x)dx$
- 예) 다음 확률밀도함수에서 C와 P(1<x<2)를 구하시오.
  - $f(x) = \begin{cases} Ce^{-x} & \text{if} \;\; x \ge 0 \\ 0 & \text{if} \;\; x < 0 \end{cases}$
    $\int_{-\infty}^{\infty} f(x)dx = 1$
  - 인데, x<0 인 경우 f(x)=0 이므로
    $\int_0^{\infty} f(x)dx = \int_0^{\infty} Ce^{-x}dx = 1$
    이다.
  - $\int_0^{\infty} Ce^{-x}dx = C\int_0^{\infty} e^{-x}dx = C\left[-e^{-x}\right]_0^{\infty} = C\left[0-(-1)\right] = C = 1$
  - C가 1이므로
  - $P(1 < x < 2) = \int_1^2 e^{-x}dx = \left[-e^{-x}\right]_1^2 = -e^{-2} - (-e)^{-1} = e^{-1} - e^{-2}$

2. 기댓값

- 확률분포
  - 확률변수 모든 값에 대응하는 확률을 나타내는 분포
    - 이산형 확률변수: 확률질량함수
    - 연속형 확률변수: 확률밀도함수
- 기댓값
  - 분포의 무게중심(균형점)
    - 이산형 확률변수:
      $E(X) = \sum_{x=0}^{\infty} xf(x)$
    - 연속형 확률변수:
      $E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- 예) 동전을 2번 던져서 나온 앞면의 수의 기댓값은?

X	0	1	2
E(X)	1/4	1/2	1/4

$$E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

- 예) 다음 확률밀도함수를 가지는 X의 기댓값은?

$$f(x) = \begin{cases} 2(1-x) & \text{if } \ 0 \le x \le 1 \\ 0 & \text{if } \ x < 0, x > 1 \end{cases}$$

$$E(X) = \int_0^1 x \cdot 2(1-x)dx = \int_0^1 (2x - 2x^2)dx$$
$$= \left[2 \cdot \frac{1}{2}x^2 - 2 \cdot \frac{1}{3}x^3\right]_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$$

3. 기댓값의 성질

- 기댓값의 성질
  - E(aX+b) = aE(X) + b
  - E(g(X)+h(X)) = E(g(X)) + E(h(X))

03 분산과 표준편차

1. 분산

- 분산(variance): 확률변수의 흩어져 있는 정도