제5장. 확률과 확률분포

Jinseog Kim
Dongguk University
jinseog.kim@gmail.com

2017-05-10

확률변수 (random variable)

- 확률 (Probability) : 표본공간의 부분집합(사건)들에 대하여 0과 1사이의 수를 대응시키는 함수, 즉,

$$P : S \to [0, 1].$$

- 동전을 한번 던지는 실험에서의 확률
  - 표본공간: $S = \{H, T\}$
  - 사건 : $\emptyset, \{H\}, \{T\}, \{H, T\}$
  - 각 사건들의 확률

사건 (event)	$\emptyset$	$\{H\}$	$\{T\}$	$\{H, T\}$
확률 (probability)	$0/2 = 0$	$1/2$	$1/2$	$2/2 = 1$

확률변수 (random variable)

- 실험의 결과를 실수값으로 대응시키는 함수, 통상적으로 영문 대문자로 표시

$$X : S \to R.$$

- 확률변수의 예
  1. 동전을 한번 던지는 실험에서의 확률변수의 정의 : 앞면을 1, 뒷면을 0 으로 대응시키는 경우

$$X(H) = 1, X(T) = 0$$

  2. 동전을 두번 던지는 실험에서의 앞면의 개수

$$X(\{H, H\}) = 2, X(\{H, T\}) = 1, , X(\{T, H\}) = 1, , X(\{T, T\}) = 0$$

  3. 중간고사의 점수
  4. 1학년 학생들의 키

확률변수의 종류와 확률분포

- 확률변수의 종류
  - 이산형 확률변수 : 대응되는 결과값(확률변수가 취하는 값, 치역의 원소의 개수)의 개수가 유한개, 또는 셀 수 있는 경우
  - 연속형 확률변수 : 대응되는 결과값의 개수가 셀 수 없는 무한 개
- 확률분포 (probability distribution): 확률변수가 취하는 값과 확률을 대응시키는 관계
  1. 동전을 한번 던지는 실험

$$P(X = 1) = P(s : X(s) = 1) = P(\{H\}) = 1/2$$

$$P(X = 0) = P(s : X(s) = 0) = P(\{T\}) = 1/2$$

  2. 동전을 두번 던지는 실험에서 앞면의 수
     - 확률변수

표본공간	$\{H, H\}$	$\{T, H\}$	$\{H, T\}$	$\{T, T\}$
확률변수(X)의 값	2	1	1	0

     - 확률분포

x	0	1	2
$P(X = x)$	1/4	1/2	1/4

확률분포의 성질

1. $P(X = x) \geq 0$
2. $\sum_{\forall x} P(X = x) = 0$
3. $P(X \leq x) = \sum_{a \leq x} P(X = a)$

확률변수의 평균과 분산

- 확률변수의 평균 = 확률변수의 기대값(expected value)

$$E(X) = \sum_{\forall x} xP(X = x)$$

1. 동전을 한번 던지는 실험

$$E(X) = 1 \times P(X = 1) + 0 \times P(X = 0) = 1/2$$

2. 동전을 두번 던지는 실험에서 앞면의 수

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) \\ &= 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 \\ &= 0 + 1/2 + 1/2 = 1 \end{aligned}$$

확률변수의 평균과 분산

- 확률변수의 분산

$$V(X) = E[X - E(X)]^2 = \sum_{\forall x} [x - E(X)]^2 P(X = x)$$

1. 동전을 한번 던지는 실험

   $V(X) = (1 - 1/2)^2 P(X = 1) + (0 - 1/2)^2 P(X = 0) = 1/4 \cdot 1/2 + 1/4 \cdot 1/2 = 1/4$

2. 동전을 두번 던지는 실험에서 앞면의 수

$$\begin{aligned} V(X) &= (0 - 1)^2 P(X = 0) + (1 - 1)^2 P(X = 1) + (2 - 1)^2 P(X = 2) \\ &= 1/4 + 0 + 1/4 = 1/2 \end{aligned}$$

- 확률변수의 분산의 간편식

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

확률함수와 누적분포함수

- 이산형 확률변수 X에 대한 확률을 다음과 같이 표현 : 확률질량함수(probability mass function)

$$p(x) = P(X = x)$$

- 확률변수 X가 주어진 값 x보다 작거나 같게 될 확률 : 누적분포함수(cumulative distribution function)

$$F_X(x) = P(X \leq x) = \sum_{y=-\infty}^{x} P(X = y)$$

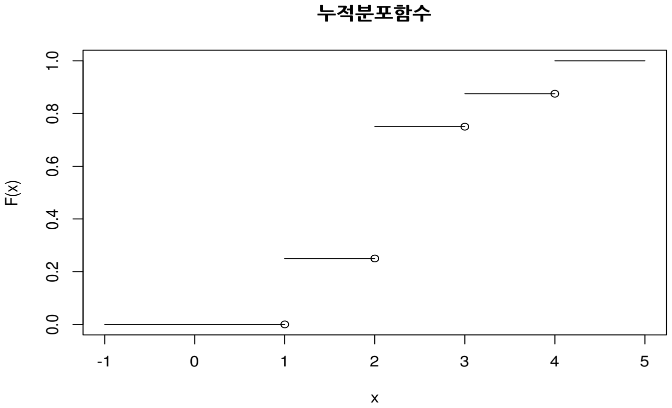
예제 13

- 확률(질량)함수가 아래인 경우 누적분포함수는?

$$p(1) = 1/4, p(2) = 1/2, p(3) = 1/8, p(4) = 1/8$$

$$F(x) = \begin{cases} 0, x < 1 \\ 1/4, x < 2 \\ 3/4, x < 3 \\ 7/8, x < 4 \\ 1, x \geq 4 \end{cases}$$

예제 13



누적분포함수의 성질

1. $a < b$에 대하여 $F(a) < F(b)$
2. $\lim_{x \to -\infty} F(x) = 0$
3. $\lim_{x \to \infty} F(x) = 1$
4. $\lim_{x \to a+} F(x) = F(a)$

- 확률질량함수와 누적분포함수의 관계

$$F_X(x) - F(x-) = P(X \leq x) - P(X < x) = P(X = x)$$

단, 여기서 $F(x-) = \lim_{y \to x-} F(y)$

연속형 확률변수와 확률밀도함수

- 연속형확률변수인 경우의 확률질량함수는 항상 0 : 의미가 없음

$$P(X = x) = P(X \leq x) - P(X < x) = 0$$

- 연속형확률변수의 확률밀도함수(probability density function): 누적분포함수의 미분으로 정의함

$$f(x) = \frac{dF(x)}{dx}$$

연속형 확률변수와 확률밀도함수

1. 모든 x에 대하여, $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\int_{-\infty}^{x} f(x)dx = F(x)$

연속형 확률변수의 평균과 분산

- 확률변수의 기대값(expected value)

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- 확률변수의 분산

$$V(X) = E[X - E(X)]^2 = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx$$

예제 14

아래의 확률밀도함수에 대하여

$$f(x) = c(4x - 2x^2), 0 < x < 2$$

1. 상수 c?

$$\int f(x)dx = \int_0^2 c(4x - 2x^2)dx = \frac{8c}{3} = 1, c = \frac{3}{8}$$

2. $F(1)$?

$$F(1) = \int_0^1 \frac{3}{8}(4x - 2x^2)dx = 1/2$$

3. $E(X)$?

$$E(X) = \int_0^2 x\frac{3}{8}(4x - 2x^2)dx = 1$$

4. $E(X^2)$?

$$E(X^2) = \int_0^2 x^2\frac{3}{8}(4x - 2x^2)dx = 6/5$$

5. $V(X)$?

$$E(X^2) - [E(X)]^2 = 6/5 - 1^2 = 1/5$$