04 CHAPTER

확률변수와 확률분포

1

제4장
확률변수와 확률분포

1. 확률변수와 확률분포의 개념

2/13

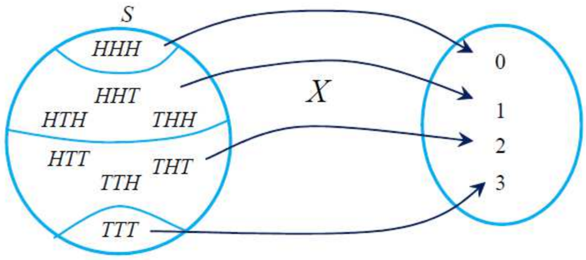
2

1. 확률변수와 확률분포의 개념

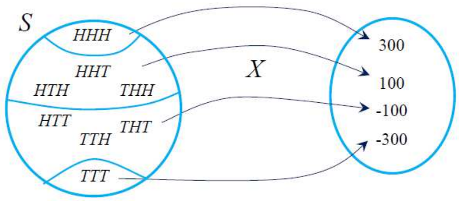
[정의 4-1] 확률변수(random variable)
- 표본공간의 각 원소를 실수 값으로 바꾸는 함수
- 확률분포를 가짐

- 동전을 세 번 던지는 실험에서의 표본공간

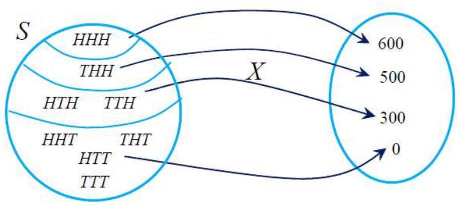
  S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
- 동전의 뒷면이 나오는 횟수 → 확률변수



[예 4-1] 동전을 세 번 던져 나온 (앞면의 개수 - 뒷면의 개수)만큼 100원 씩 주고받는 게임에서 수익 → 확률변수 X



- 첫 번째 앞면 → 100, 두 번째 앞면 → 200, 세 번째 앞면 → 300
- 뒷면이 나오면 → 0으로 초기화



- 이산표본공간(discrete sample space)
  - 유한개 또는 셀 수 있는 무한개의 원소로 구성된 표본공간
    - 이산확률변수(discrete random variable)
      - ✓ 동전의 앞면이 나올 때까지의 시행 횟수
      - ✓ 100개의 제품 중 불량품의 수
- 연속표본공간(continuous sample space)
  - 실직선 상의 임의의 구간으로 나타낼 수 있는 표본공간
    - 연속확률변수(continuous random variable)
      - ✓ 사람의 키와 몸무게
      - ✓ 제품의 수명

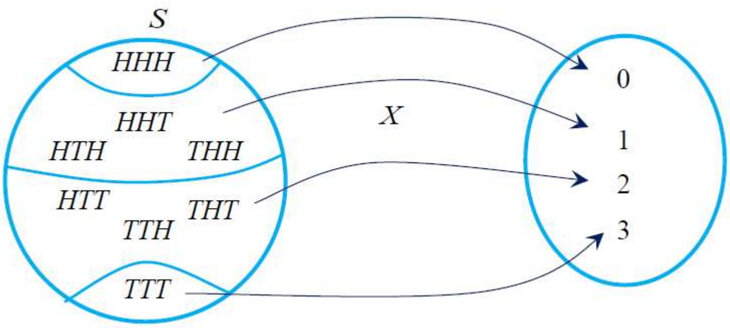
1.2 확률분포

[정의 4-2] 이산확률분포(discrete probability distribution)
- 이산표본공간의 확률변수로부터 생성된 확률분포
  - 확률질량함수(probability mass function)

$$P(X = x) = f(x)$$

$$\sum_x f(x) = 1,\ 0 \le f(x) \le 1$$

[예 4-2] 동전을 세 번 던지는 시행에서의 뒷면의 개수 X의 확률분포

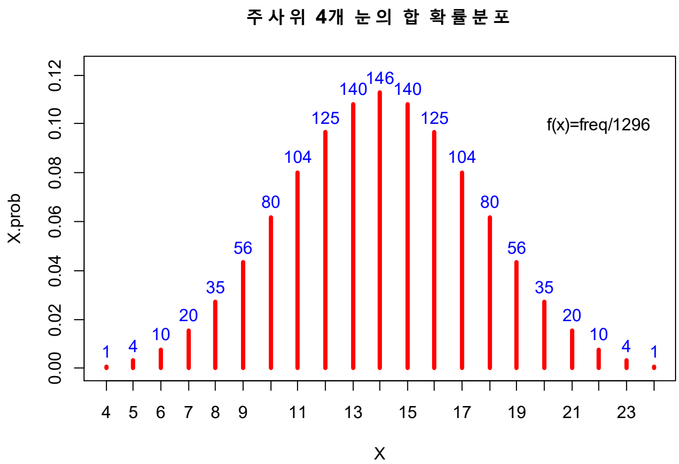


$$\Rightarrow f(0)=\frac{1}{8},\ f(1)=\frac{3}{8},\ f(2)=\frac{3}{8},\ f(3)=\frac{1}{8}$$

$$\Rightarrow \sum_x f(x)=1$$

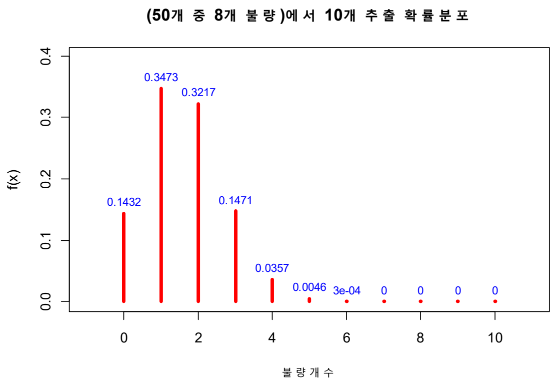
[예 4-3] 주사위를 네 번 던지는 실험에서 나오는 숫자 합 X의 확률분포

확률변수 X의 분포 그래프 [그림 4-3]



[예 4-4] 50개의 제품 중 8개의 불량품이 있는 상자로부터 10개의 제품을 랜덤 샘플링했을 때, 발견되는 불량품 개수 X에 대한 확률분포

$$f(x) = P(X = x) = \frac{{}_8C_x \times {}_{42}C_{10-x}}{{}_{50}C_{10}}, \; x = 0, 1, \ldots, 8$$



[정의 4-3] 연속확률분포(continuous probability distribution)

- 연속적인(셀 수 없는) 값을 갖는 확률변수의 확률분포
- 확률분포함수 f(x)는 확률 P(a<X<b)를 구하기 위한 확률밀도함수

$$P(a < X < b) = \int_a^b f(x)dx$$

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

[예 4-5] 밀도함수 : $f(x) = 2e^{-2x}, \; 0 < x < \infty$

(1) 확률분포? $f(x) \geq 0 \qquad \int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} 2e^{-2x}dx = [-e^{-2x}]_0^{\infty} = 1$

(2) P(0<X<1) $P(0 < X < 1) = \int_0^1 2e^{-2x}dx = [-e^{-2x}]_0^1 = 1 - e^{-2} \approx 0.8647$

[정의 4-4] 누적분포함수(cumulative distribution function)

– 확률변수 X가 특정한 값 x 이하일 확률

$$F(x) = P(X \le x)$$

[예 4-6] 동전을 세 번 던지는 확률실험에서 뒷면의 개수 X에 대한 누적 분포함수

$$f(0) = \frac{1}{8},\ f(1) = \frac{3}{8},\ f(2) = \frac{3}{8},\ f(3) = \frac{1}{8}$$

$$\Rightarrow F(x) = \begin{cases} 0, & x < 0 \\ 1/8, & 0 \le x < 1 \\ 1/2, & 1 \le x < 2 \\ 7/8, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

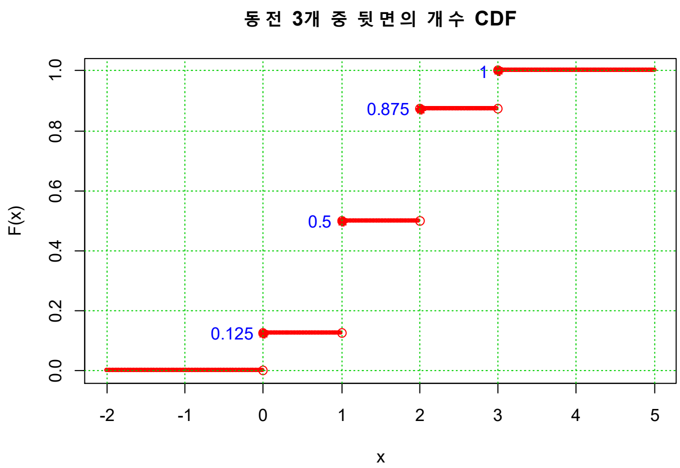
```
# 누적분포함수 F(x) 정의
Fx <- function(x) {
        if (x<0) {y <-0
        } else if (x<1) {y <- 1/8
        } else if (x<2) {y <- 1/2
        } else if (x<3) {y <- 7/8
        } else y <- 1
        return(y)  }
```

```
# 누적분포함수 벡터화 ⇒ Vectorize( ) 함수
VFx <- Vectorize(Fx, "x")
```

11/13

11

누적분포함수 F(x) 플롯



12/13

12

[예 4-7] 연속확률분포의 CDF $f(x) = 2e^{-2x},\ 0 < x < \infty$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} 2e^{-2y}dy = [-e^{-2y}]_0^x = 1 - e^{-2x},\ x > 0$$