5-1 확률변수와 확률분포

5-1-1.확률변수

확률변수

확률변수란 실험의 표본공간으로부터 실수 값으로의 변환함수이다.

따라서 확률변수가 특정 실수 값을 가질 확률은 표본공간의 원소에 대한 확률로부터 유도된다.

확률변수란 임의실험에 의해 그 값이 결정되는 변수를 말함

예 동전 2개를 던지는 실험에서 표본공간은 다음과 같다.

Ω={HH, HT, TH, TT}

▶ 이 실험에서 각각의 원소들이 나타날 확률은 $\frac{1}{4}$ 이다.

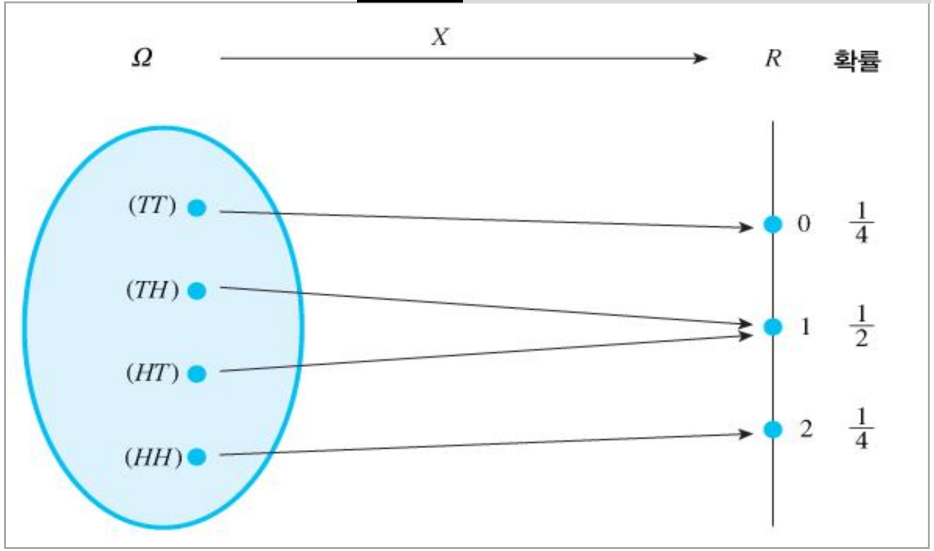
▶ 확률변수 X를 'X≡앞면의 수'로 정의할 때 가능한 값은,

X{HH}=2, X{HT, TH}=1, X{TT}=0 이고,

X의 각 값에 대한 확률은

$$P_r(X=2)=\frac{1}{4}, P(X=1)=\frac{1}{2}, P(X=0)=\frac{1}{4}$$ 과 같이 구할 수 있다.

그림 표본공간의 원소와 확률변수



'확률변수란 정의역(domain)이 표본공간이고 치역(range)이 실수값인 함수이다'

라고 할 수 있다. 따라서 확률변수 X는 $X : \Omega \rightarrow R$으로 표현할 수 있다.

정의역:함수가 정의하는 모든 수의 집합 / 치역:함수가 취하는 값 전체의 집합

이산형 확률변수

0이 아닌 확률 값을 갖는 실수 값이 셀 수 있는 경우임
즉, 확률변수가 이산점에서 0이 아닌 확률을 갖는 확률변수임

항상 정수의 값을 취하는 확률변수임

이산형 확률변수의 확률조건

1. $0 \le P_i \le 1, \ \ i = 1, 2, \cdots, n$

즉, 각 x_i 가 나타날 확률은 0과 1 사이의 값을 갖는다.

2. $\sum_{i=1}^{n} P_i = 1$

즉, 모든 가능한 경우의 확률의 합은 1이다.

확률질량함수

이산형 확률변수는 이산점(discrete points)에서 0이 아닌 확률 값을 가지며 확률은,

$$P_r(X = x_i) = P_i,\ \ i = 1,\ 2,\ \cdots,\ n$$

으로 표현한다.

이와 같이 각 이산점에 있어서 확률의 크기를 표현하는 함수를

확률질량함수(probability mass function)라고 한다.

5-1 -3. 연속형 확률변수

연속형 확률변수

가능한 값이 실수의 어느 특정구간 전체에 해당하는 확률변수임
즉, 특정 실수 구간에서 0이 아닌 확률을 갖는 확률변수임

일정한 범위내의 모든 실수 값을 연속적으로 취하는 확률변수

확률밀도함수

연속형 확률변수 X 의 확률함수를 $f(x)$ 라고 할 때,

$f(x)$ 는 **확률밀도함수(probability density function;** ***p.d.f*** **)**라고 부르며

다음 조건을 만족한다.

확률밀도함수의 조건

1. 모든 X 값에 대하여 $f(x) \geq 0$ 이다.

   즉, X 의 모든 실수 값에 대하여 확률밀도함수는 0 이상이다.

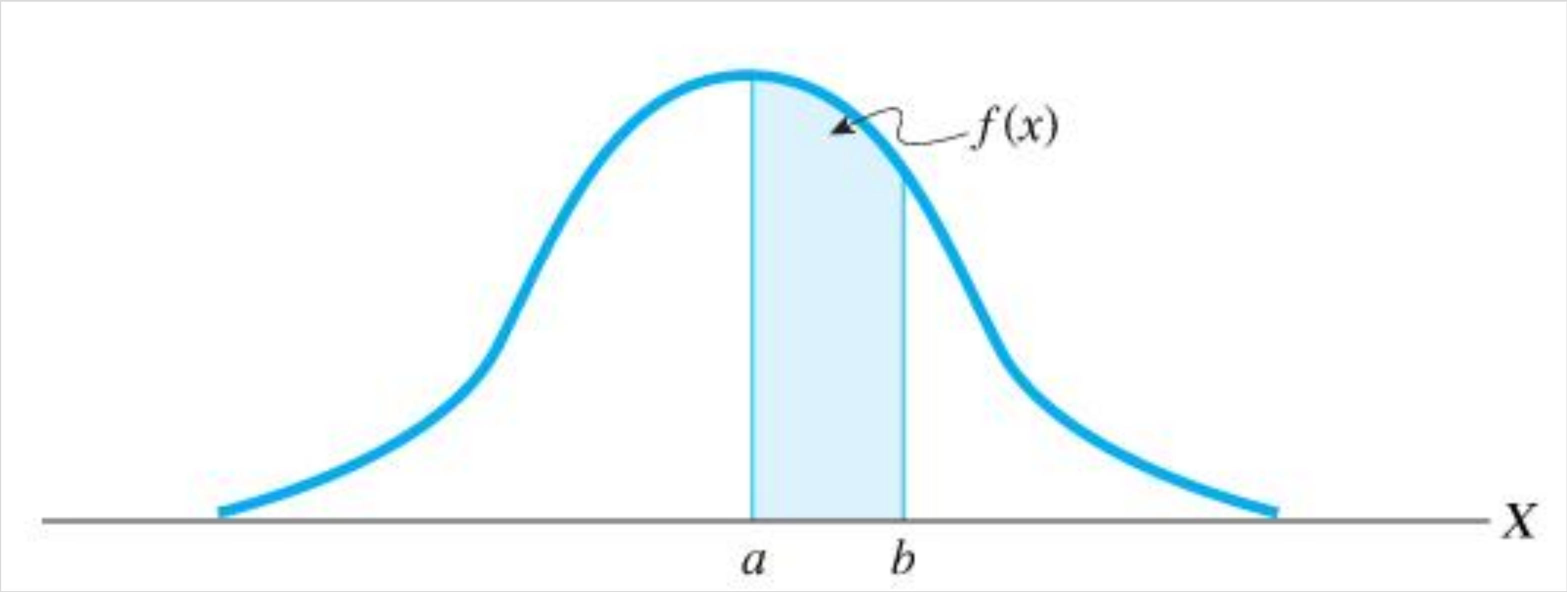
2. X 의 모든 가능한 값의 확률은 적분

   $\int_{-\infty}^{\infty} f(x)dx$로 구하며 이 값은 항상 1이다.

3. 구간 (a, b)의 확률은 $\int_{a}^{b} f(x)dx$ 이다. 즉, 구간 (a, b) 에 대한 X 의 확률은

   그 구간에 있어서 확률밀도함수 $f(x)$ 로 만들어지는 면적의 크기이다.

그림 연속형 확률변수의 확률분포



누적분포함수

특정값 ***a*** 에 대하여 확률변수 ***X*** 가 ***X* ≤ *a*** 인 모든 경우의 확률의 합

$$P_r = (X \le a) = F_X(a)$$

이산형 확률변수 : $F_X(a) = \sum_{\substack{a\text{이하인} \\ \text{모든}x_i}} P(X = x_i)$

연속형 확률변수 : $F_X(a) = \int_{-\infty}^{a} f(x)dx$

누적확률함수는 다음의 3조건을 만족시킨다.

- ✓ 모든 X의 값에 대해서 $0 \le F_X(x) \le 1$ 이다.
- ✓ $x_1 \langle x_2$ 이면, $F_X(x_1) \le F_X(x_2)$ 이다.
- ✓ $F_X(-\infty) = 0$ 이며 $F_X(\infty) = 1$ 이다.

이산형 확률변수의 누적분포함수

이산형 확률변수 ***X***가 이산점 $x_1 < x_2 < \ldots < x_n$에서 0이 아닌 확률을 가지며 그 확률을 $P_i, i=1,\ldots,n$ 이라고 할 때, ***X***의 누적분포함수는 다음과 같다.

1. $a < x_1$ 인 경우, $F(a)=0$ 이다.
2. $a \ge x_1$ 인 경우, $F(a)=\sum_{x_i \le a} P_i$ 이다.
3. $a \ge x_n$ 인 경우, $F(a)=\sum_{x_i \le a} P_i = 1$ 이다.
4. $F(a)$는 ***X***의 각 이산점 $x_1 < x_2 < \ldots < x_n$에서 $P_i, i=1,\ldots,n$ 만큼씩 도약하는 계단함수이다.

예 동전 3개를 던지는 실험에서 확률변수 X를 'X≡앞면의 수'라고 정의할 때 X의 확률분포는 다음과 같다.

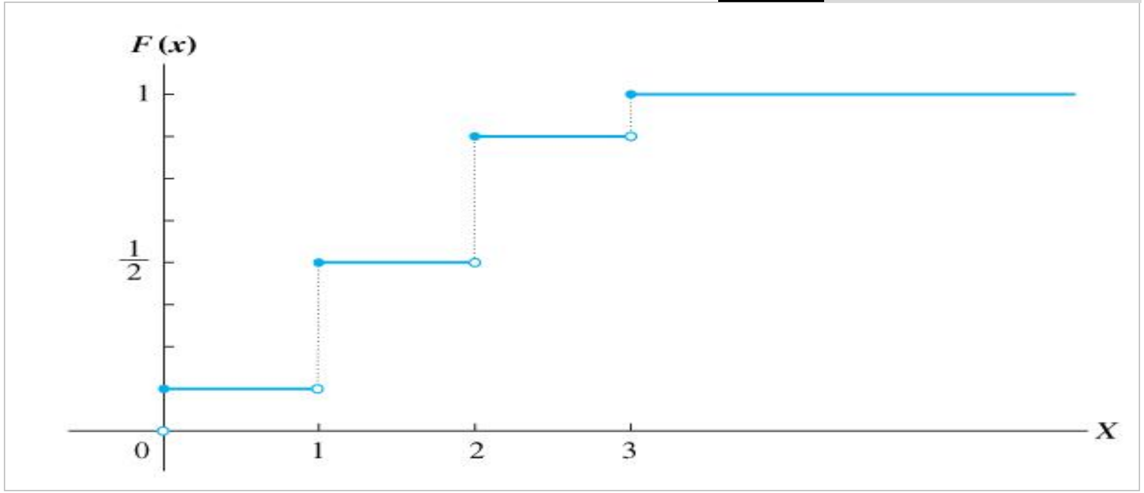
X	0	1	2	3	합
확률	1/8	3/8	3/8	1/8	1

이 확률분포의 누적분포함수는 다음과 같다.

x	$F(x) = Pr(X \le x)$
$x<0$	0
$0 \le x < 1$	1/8
$1 \le x < 2$	1/2
$2 \le x < 3$	7/8
$x \ge 3$	1

▶ 위에서 구한 누적분포함수값을 그래프로 표현하면 다음과 같다.

그림 **X의 누적분포함수**



▶ 이산형 확률변수는 이산점에서 0이 아닌 확률을 가지므로 누적분포함수는 각 이산점에서 도약(jump)하는 **계단함수(step function)**형태이다.