확률과 통계 : 07 연속확률분포

아이비에듀

July 13, 2022

목차

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{}, \boxed{}, \boxed{}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{0}, \boxed{1}, \boxed{2}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{0}, \boxed{1}, \boxed{2}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) 가능한 각각의 X값에 대하여, 확률을 계산해볼 수 있다 ;

$$P(X = 0) = {}_2C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$P(X = 1) = {}_2C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{1}{2}$$

$$P(X = 2) = {}_2C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$

이렇게, 0을 넣으면 $\frac{1}{4}$가 나오고, 1을 넣으면 $\frac{1}{2}$가 나오고, 2를 넣으면 $\frac{1}{4}$가 나오는 함수 $P(X = x)$를 생각할 수 있다. 이 함수를 $\boxed{}$라고 부른다.

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{0}, \boxed{1}, \boxed{2}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) 가능한 각각의 X값에 대하여, 확률을 계산해볼 수 있다 ;

$$P(X=0) = {}_2C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$P(X=1) = {}_2C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{1}{2}$$

$$P(X=2) = {}_2C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$

이렇게, 0을 넣으면 $\frac{1}{4}$가 나오고, 1을 넣으면 $\frac{1}{2}$가 나오고, 2를 넣으면 $\frac{1}{4}$가 나오는 함수 $P(X=x)$를 생각할 수 있다. 이 함수를 확률질량함수라고 부른다.

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{0}, \ \boxed{1}, \ \boxed{2}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) 가능한 각각의 X값에 대하여, 확률을 계산해볼 수 있다 ;

$$P(X=0) = {}_2C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

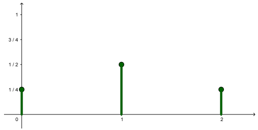
$$P(X=1) = {}_2C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{1}{2}$$

$$P(X=2) = {}_2C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$

이렇게, 0을 넣으면 $\frac{1}{4}$가 나오고, 1을 넣으면 $\frac{1}{2}$가 나오고, 2를 넣으면 $\frac{1}{4}$가 나오는 함수 $P(X=x)$를 생각할 수 있다. 이 함수를 $\boxed{\text{확률질량함수}}$라고 부른다.

(3) 확률질량함수는 다음과 같이 표로 나타낼 수도 있고, 그래프로 나타낼 수도 있다.

X	0	1	2
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



이 표를 $\boxed{}$라고 하고, 이 그래프를 $\boxed{}$이라고 부른다.

이산확률분포(복습)

동전을 두 번 던졌을 때, 앞면이 나온 횟수를 X라고 하자.

(1) 가능한 X의 값은

$$X = \boxed{0},\ \boxed{1},\ \boxed{2}$$

이다. 따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) 가능한 각각의 X값에 대하여, 확률을 계산해볼 수 있다 ;

$$P(X=0) = {}_2C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

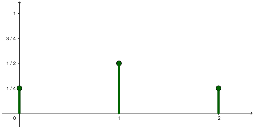
$$P(X=1) = {}_2C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{1}{2}$$

$$P(X=2) = {}_2C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$

이렇게, 0을 넣으면 $\frac{1}{4}$가 나오고, 1을 넣으면 $\frac{1}{2}$가 나오고, 2를 넣으면 $\frac{1}{4}$가 나오는 함수 $P(X=x)$를 생각할 수 있다. 이 함수를 $\boxed{\text{확률질량함수}}$라고 부른다.

(3) 확률질량함수는 다음과 같이 표로 나타낼 수도 있고, 그래프로 나타낼 수도 있다.

X	0	1	2
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



이 표를 $\boxed{\text{확률분포표}}$라고 하고, 이 그래프를 $\boxed{\text{히스토그램}}$이라고 부른다.

이산확률분포(복습)

(4) 이 확률분포에 대한 기댓값과 분산, 표준편차는 다음과 같이 계산한다.

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

$$V(X) = E((X - m)^2) = (0 - 1)^2 \times \frac{1}{4} + (1 - 1)^2 \times \frac{1}{2} + (2 - 0)^2 \times \frac{1}{4} = \frac{1}{2}$$

$$\sigma(X) = \sqrt{V(X)} = \frac{\sqrt{2}}{2}.$$

이산확률분포(복습)

(4) 이 확률분포에 대한 기댓값과 분산, 표준편차는 다음과 같이 계산한다.

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

$$V(X) = E((X-m)^2) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-0)^2 \times \frac{1}{4} = \frac{1}{2}$$

$$\sigma(X) = \sqrt{V(X)} = \frac{\sqrt{2}}{2}.$$

분산을 구할 때에는 $V(X) = E(X^2) - E(X)^2$을 활용하여 구해도 된다.

$$E(X^2) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}$$

$$V(X) = E(X^2) - E(X)^2 = \frac{3}{2} - 1 = \frac{1}{2}$$

이산확률분포(복습)

(4) 이 확률분포에 대한 기댓값과 분산, 표준편차는 다음과 같이 계산한다.

$$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$
$$V(X) = E((X-m)^2) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-0)^2 \times \frac{1}{4} = \frac{1}{2}$$
$$\sigma(X) = \sqrt{V(X)} = \frac{\sqrt{2}}{2}.$$

분산을 구할 때에는 $V(X) = E(X^2) - E(X)^2$을 활용하여 구해도 된다.

$$E(X^2) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}$$
$$V(X) = E(X^2) - E(X)^2 = \frac{3}{2} - 1 = \frac{1}{2}$$

(5) 한편, X는 사실 $\boxed{}$를 따른다고 볼 수 있다. 즉 $X \sim B(2, \frac{1}{2})$이다.

이산확률분포(복습)

(4) 이 확률분포에 대한 기댓값과 분산, 표준편차는 다음과 같이 계산한다.

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

$$V(X) = E((X-m)^2) = (0-1)^2 \times \frac{1}{4} + (1-1)^2 \times \frac{1}{2} + (2-0)^2 \times \frac{1}{4} = \frac{1}{2}$$

$$\sigma(X) = \sqrt{V(X)} = \frac{\sqrt{2}}{2}.$$

분산을 구할 때에는 $V(X) = E(X^2) - E(X)^2$을 활용하여 구해도 된다.

$$E(X^2) = 0^2 \times \frac{1}{4} + 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{3}{2}$$

$$V(X) = E(X^2) - E(X)^2 = \frac{3}{2} - 1 = \frac{1}{2}$$

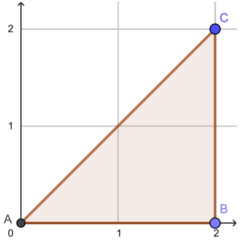
(5) 한편, X는 사실 $\boxed{\text{이항분포}}$를 따른다고 볼 수 있다. 즉 $X \sim B(2, \frac{1}{2})$이다. 따라서

$$E(X) = 2 \times \frac{1}{2} = 1$$

$$V(X) = 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}.$$

연속확률분포

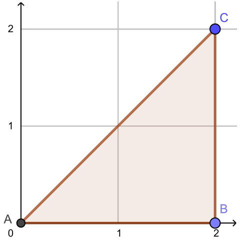
오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{}$이다.

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



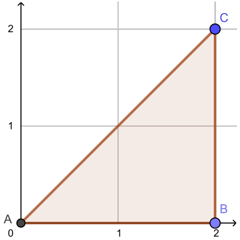
(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

(2) 가능한 X의 값의 범위는 $\boxed{\phantom{0 < X < 2}}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



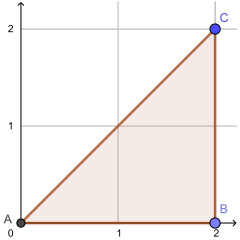
(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

(2) 가능한 X의 값의 범위는 $\boxed{0 \leq X \leq 2}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

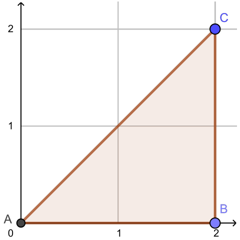
(2) 가능한 X의 값의 범위는 $\boxed{0 \leq X \leq 2}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(3) $P(0 \leq X \leq 1) = \boxed{}$, $P(1 \leq X \leq 2) = \boxed{}$

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

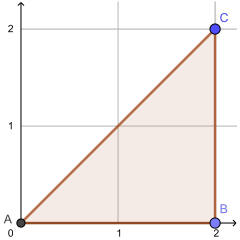
(2) 가능한 X의 값의 범위는 $\boxed{0 \leq X \leq 2}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(3) $P(0 \leq X \leq 1) = \boxed{\frac{1}{4}}$, $\quad P(1 \leq X \leq 2) = \boxed{\frac{3}{4}}$

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

(2) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 2}$이다.

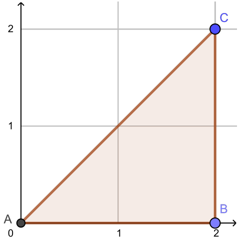
따라서 X는 (이산확률변수, 연속확률변수)이다.

(3) $P(0 \le X \le 1) = \boxed{\frac{1}{4}}$, $\quad P(1 \le X \le 2) = \boxed{\frac{3}{4}}$

(4) $P(X = 1) = \boxed{}$, $\quad P(X = 2) = \boxed{}$

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

(2) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 2}$이다.

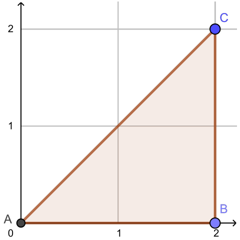
따라서 X는 (이산확률변수, 연속확률변수)이다.

(3) $P(0 \le X \le 1) = \boxed{\frac{1}{4}}$, $P(1 \le X \le 2) = \boxed{\frac{3}{4}}$

(4) $P(X = 1) = \boxed{0}$, $P(X = 2) = \boxed{0}$

연속확률분포

오른쪽 그림과 같이 세 점 $A(0,0)$, $B(2,0)$, $C(2,2)$으로 이루어진 삼각형 ABC가 있다. 이 삼각형의 내부에 임의로 한 점 Q를 잡을 때, Q의 x좌표를 X라고 하자.



(1) 이 삼각형의 넓이는 $\boxed{2}$이다.

(2) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 2}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

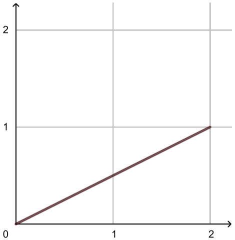
(3) $P(0 \le X \le 1) = \boxed{\frac{1}{4}}$, $P(1 \le X \le 2) = \boxed{\frac{3}{4}}$

(4) $P(X = 1) = \boxed{0}$, $P(X = 2) = \boxed{0}$

따라서, 이 경우에는 확률질량함수 $P(X = x)$를 생각하는 것이 아무런 의미가 없다. 대신, 새로운 함수 $f(x)$를 고려하는데, 이 함수는

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

를 만족시키는 함수로서, 확률밀도함수 라고 부른다. 이 경우에 확률밀도함수는 $f(x) = \frac{1}{2}x$이다.



연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) $P(0 \le X \le 20) = \boxed{}$, $P(30 \le X \le 45) = \boxed{}$

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) $P(0 \le X \le 20) = \boxed{\frac{1}{3}}$, $P(30 \le X \le 45) = \boxed{\frac{1}{4}}$

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) $P(0 \le X \le 20) = \boxed{\frac{1}{3}}$, $P(30 \le X \le 45) = \boxed{\frac{1}{4}}$

(3) $P(X = 20) = \boxed{}$, $P(X = 45) = \boxed{}$

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

(2) $P(0 \le X \le 20) = \boxed{\frac{1}{3}}$, $P(30 \le X \le 45) = \boxed{\frac{1}{4}}$

(3) $P(X = 20) = \boxed{0}$, $P(X = 45) = \boxed{0}$

연속확률분포

철수는 12시에서 1시 사이에 아무때나 은행에 방문하기로 하였다. 철수가 은행에 방문한 시각을 12시 X분이라고 하자.

(1) 가능한 X의 값의 범위는 $\boxed{0 \le X \le 60}$이다.

따라서 X는 (이산확률변수, 연속확률변수)이다.

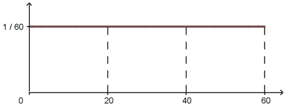
(2) $P(0 \le X \le 20) = \boxed{\frac{1}{3}}$, $\quad P(30 \le X \le 45) = \boxed{\frac{1}{4}}$

(3) $P(X = 20) = \boxed{0}$, $\quad P(X = 45) = \boxed{0}$

이 경우에도 확률질량함수 $P(X = x)$를 생각하는 것이 아무런 의미가 없다. 대신, 확률밀도함수 $f(x)$를 고려하는데, 이 함수는

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

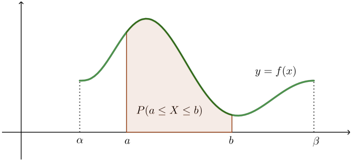
를 만족시킨다. 이 문제에서 확률밀도함수는 $f(x) = \frac{1}{60}$이다.



연속확률분포

연속확률변수 X가 값을 가질 수 있는 범위가 $\alpha \le X \le \beta$일 때, 다음 세 가지 성질을 만족시키는 함수 $f(x)$를 X의 확률밀도함수라고 한다.

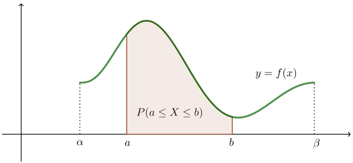
- $f(x) \ge 0$
- $\displaystyle\int_\alpha^\beta f(x)\,dx = 1$
- $\displaystyle\int_a^b f(x)\,dx = P(a \le X \le b)$



연속확률분포

연속확률변수 X가 값을 가질 수 있는 범위가 $\alpha \le X \le \beta$일 때, 다음 세 가지 성질을 만족시키는 함수 $f(x)$를 X의 확률밀도함수라고 한다.

- $f(x) \ge 0$
- $\displaystyle\int_{\alpha}^{\beta} f(x)\,dx = 1$
- $\displaystyle\int_{a}^{b} f(x)\,dx = P(a \le X \le b)$



- $0 \le p_i \le 1$
- $\displaystyle\sum_{i=1}^{n} p_i = 1$
- $\displaystyle\sum_{i=j}^{k} p_i = P(x_i \le X \le x_k)$

X	x_1	x_2	$\cdots$	x_n
$P(X = x_i)$	p_i	p_2	$\cdots$	p_n

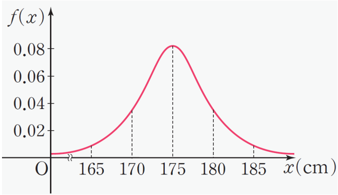
정규분포

예시 1

어느 고등학교에서 남학생의 키를 X라고 할 때, X의 분포가 다음과 같이 생겼다고 한다.

(1) 이 고등학교 남학생들의 키의 평균은 $\boxed{}$cm 이다.

(2) 키가 175cm에서 180cm인 학생들보다는 180cm에서 186cm인 학생들이 더 (많다, 적다).



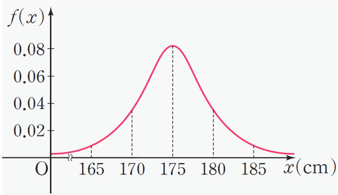
정규분포

예시 1

어느 고등학교에서 남학생의 키를 X라고 할 때, X의 분포가 다음과 같이 생겼다고 한다.

(1) 이 고등학교 남학생들의 키의 평균은 $\boxed{175}$cm 이다.

(2) 키가 175cm에서 180cm인 학생들보다는 180cm에서 186cm인 학생들이 더 (많다, 적다).



정의 2) 정규분포

사회나 자연에서 일어나는 많은 현상들에서 나타나는 자료들은 위의 그래프처럼 좌우 대칭인 종 모양의 분포를 가지는 경우가 많다. 이와 같은 분포를 정규분포라고 한다.

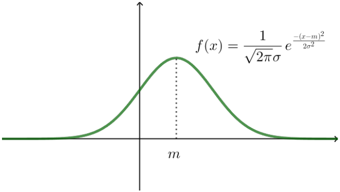
연속확률변수 X가 평균이 m이고 표준편차가 σ인 정규분포를 따를 때 기호로

$$X \sim N(m, \sigma)$$

라고 쓴다. 이때, 확률밀도함수 $f(x)$는

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma}}$$

이다.



정규분포

정리 3) 정규분포 곡선의 성질

(1) 정규분포의 확률밀도함수는 $x = m$일 때 최댓값을 가진다. 따라서, 평균 m이 커질수록 정규분포 곡선은 (왼쪽으로, 오른쪽으로) 이동한다.

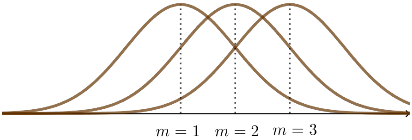
(2) 표준편차 σ는 이 분포가 얼마나 넓게 분포되어 있는지를 나타내는 값이다. 따라서, 표준편차 σ가 커질수록 정규분포 곡선은 (넓게 퍼진다, 뾰족하게 모인다).

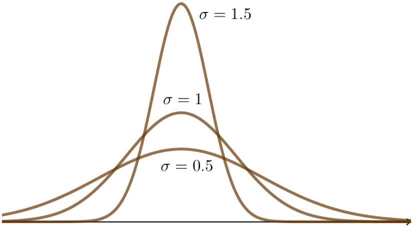
정규분포

정리 3) 정규분포 곡선의 성질

(1) 정규분포의 확률밀도함수는 $x = m$일 때 최댓값을 가진다. 따라서, 평균 m이 커질수록 정규분포 곡선은 (왼쪽으로, 오른쪽으로) 이동한다.

(2) 표준편차 σ는 이 분포가 얼마나 넓게 분포되어 있는지를 나타내는 값이다. 따라서, 표준편차 σ가 커질수록 정규분포 곡선은 (넓게 퍼진다, 뾰족하게 모인다).





표준정규분포