5강

확률 및 확률분포함수2

통계·데이터과학과 이기재 교수

한국방송통신대학교

목차

01

확률변수

확률변수(random variable)

› 표본공간의 각 원소에 실수값을 대응시켜 주는 함수

- 이산형 확률변수
  - 불량품의 수, 고속도로에서의 사고건수, 방문자수 등
- 연속형 확률변수
  - 전구의 수명, 몸무게, 체온, 통근시간 등

확률변수 예제1

- 동전을 두 번 던지는 실험

X = "두 번 던질 때 나온 앞면의 수"

{앞 앞}	→	2
{앞 뒤}	→	1
{뒤 앞}	→	1
{뒤 뒤}	→	0

$$P(X = 2) = P(\{\text{앞앞}\}) = \frac{1}{4}$$

$$P(X = 1) = P(\{\text{앞뒤, 뒤앞}\}) = \frac{2}{4}$$

$$P(X = 0) = P(\{\text{뒤뒤}\}) = \frac{1}{4}$$

확률변수 예제2

- 동전을 2개 던지는 실험

  X = "동전 2개 던질 때 나온 앞면의 수"

  - 확률변수 X의 확률분포함수

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

  - 확률변수 X의 누적확률분포함수

X	0	1	2
$P(X\leq x)$	1/4	3/4	1

확률변수 예제3

- 200가구를 조사대상으로 지난 일년 동안 각 가구에서 병원 방문 회수 조사

병원방문 횟수	0	1	2	3	4
가구수	74	80	30	10	6

확률변수 X = "병원 방문 회수 " 의 확률분포함수와 누적 확률분포함수는?

확률분포함수		누적확률분포함수	
$X=x$	$P(X=x)$	$X=x$	$P(X\le x)$
0	0.37	0	0.37
1	0.40	1	0.77
2	0.15	2	0.92
3	0.05	3	0.97
4	0.03	4	1.00
계	1.00	–	–

이산형 확률분포의 성질

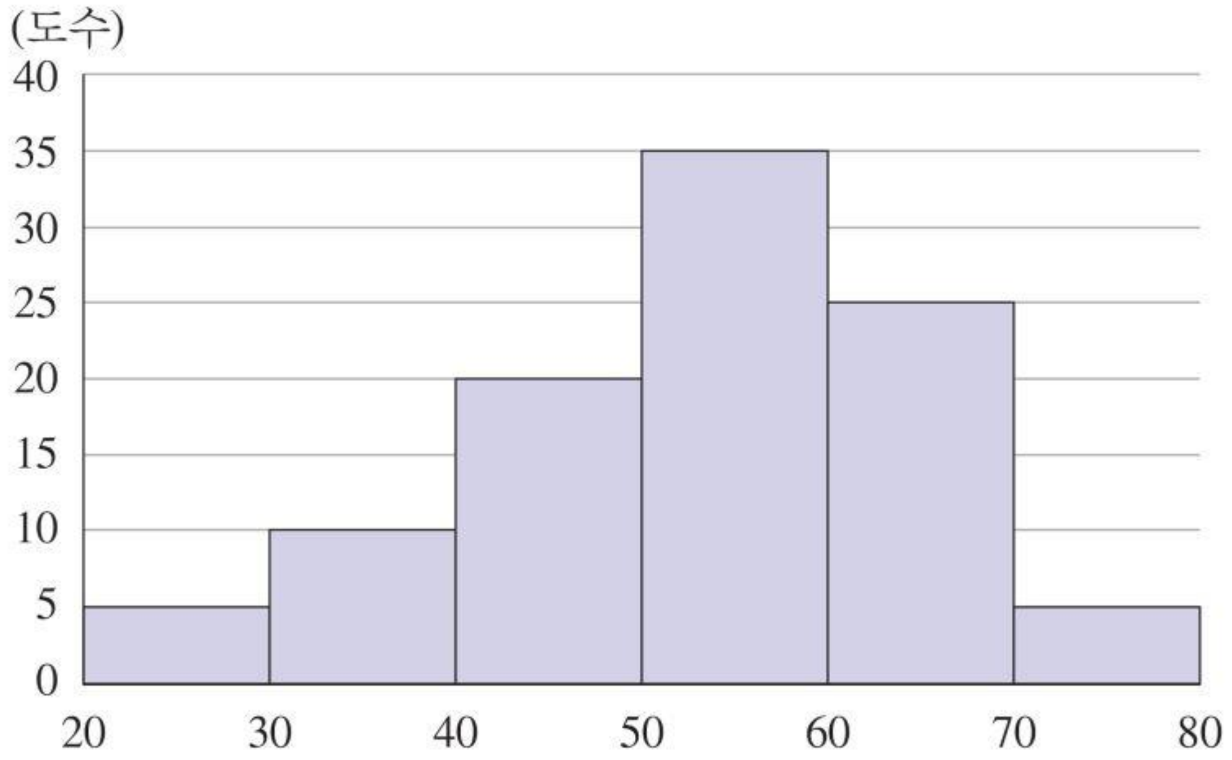
- 확률분포함수 $p(x) = P(X = x)$에 대하여
  - $0 \le p(x) \le 1,$
  - $\displaystyle\sum_{\text{모든 } x} p(x) = 1$
  - $\displaystyle P(a < X \le b) = \sum_{a < x \le b} p(x)$

연속형 확률변수 사례

- 회사까지의 출근 소요 시간
  - 확률변수 X = "출근 소요 시간"

$a \leq X < b$	도수	상대도수
$20 \leq X < 30$분	5일	5/100
$30 \leq X < 40$분	10일	10/100
$40 \leq X < 50$분	20일	20/100
$50 \leq X < 60$분	40일	40/100
$60 \leq X < 70$분	20일	20/100
$70 \leq X < 80$분	5일	5/100
합계	100일	1

출근시간의 히스토그램

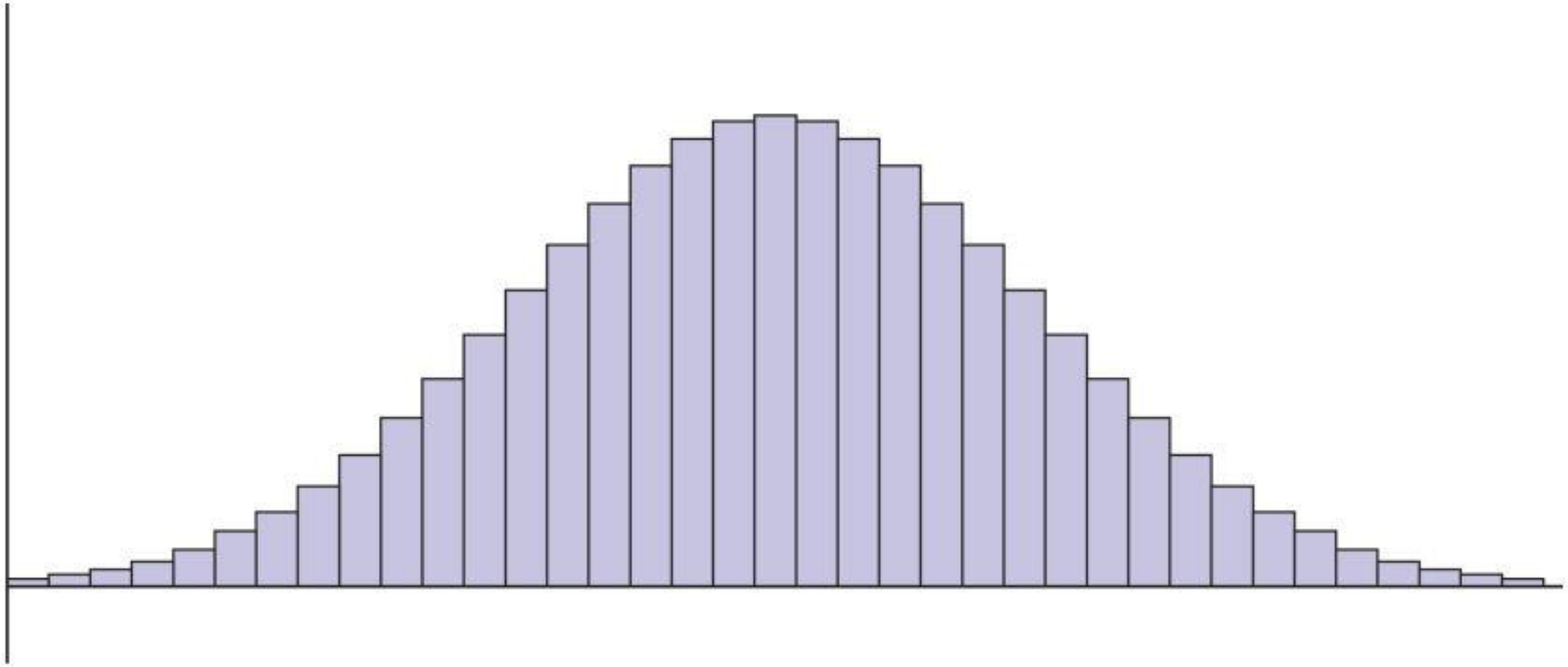


"출근시간이 30분에서 50분 사이일 확률

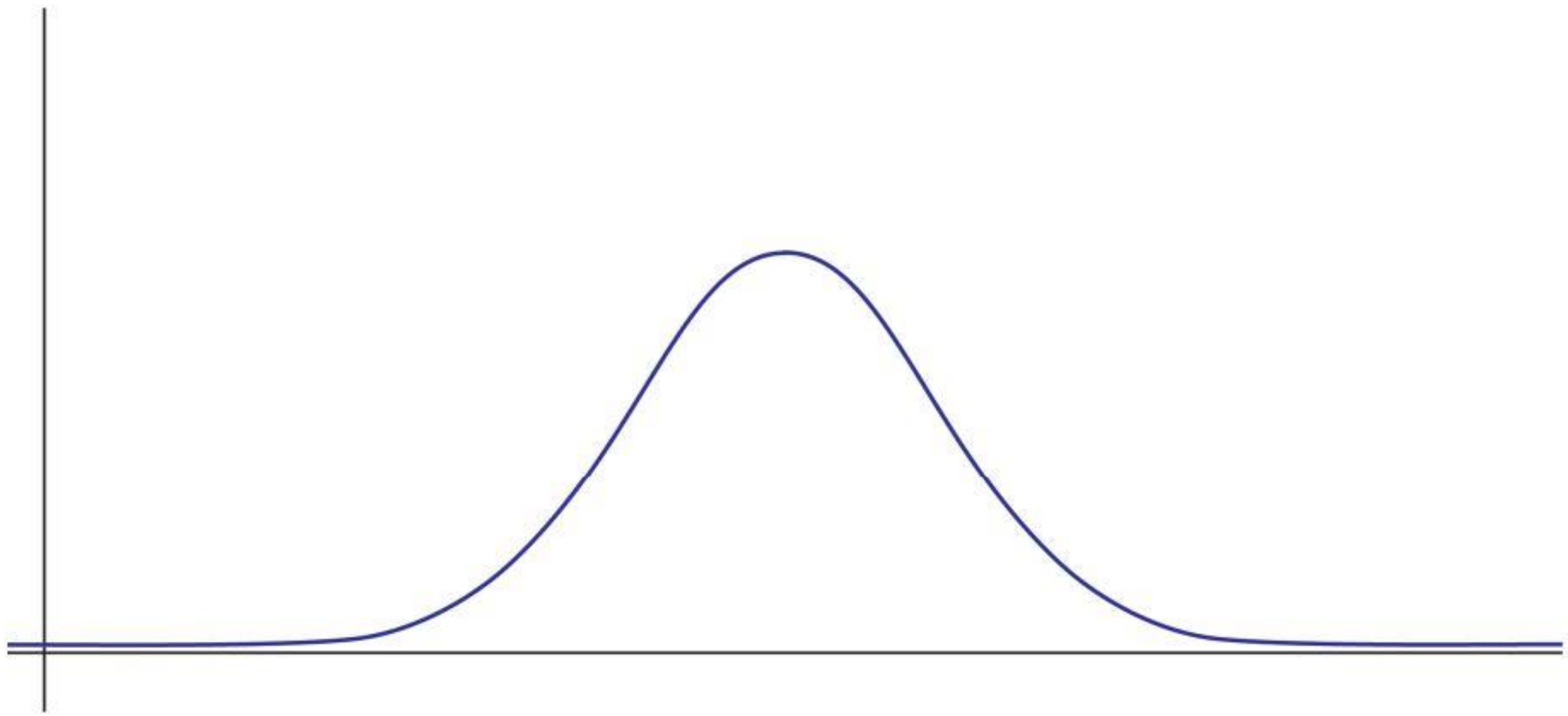
$$\Rightarrow P(30 \leq X < 50) = \frac{10}{100} + \frac{20}{100} = 0.3$$

히스토그램

- 지난 1년 동안의 통근시간에 대한 상대도수 히스토그램

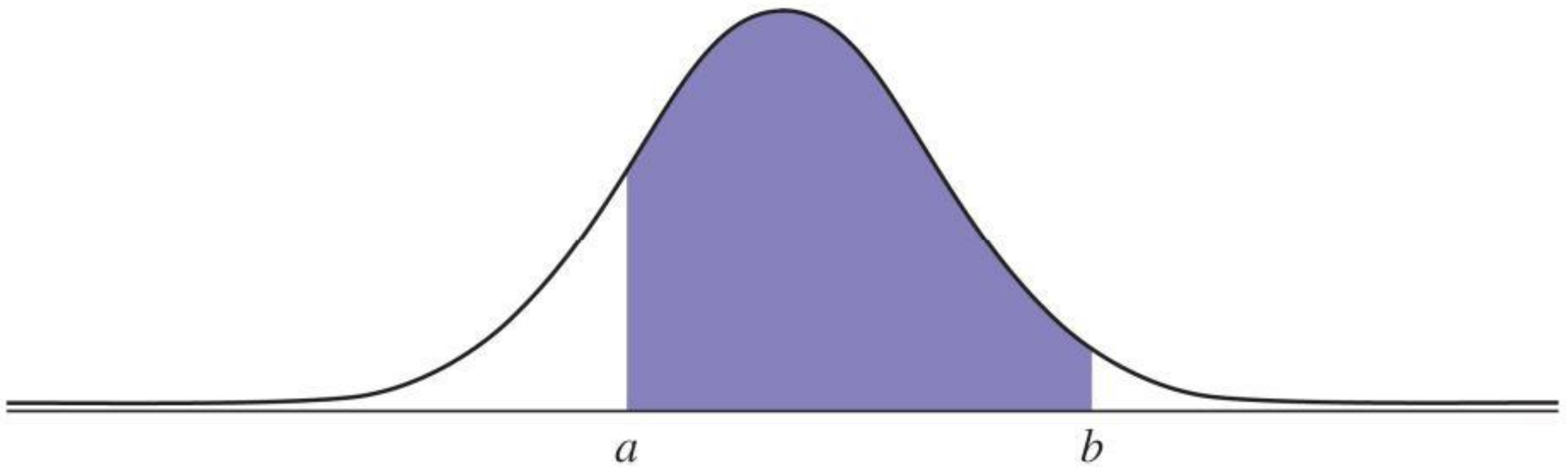


연속형 확률변수의 확률분포함수



연속형 확률변수의 확률분포함수

- 연속형 확률변수 X에 대한 $P(a \leq X < b)$



- 수학적 표현

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

확률밀도함수 $f(x)$의 성질

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) = 1$
- $P(a < X \leq b) = \int_{a}^{b} f(x)dx$

02

기댓값과 분산

확률변수 X의 기댓값(평균)

- 확률변수 X의 확률밀도함수를 $f(x)$라고 하면
  - 이산형

$$E(X) = \mu = \sum x_i f(x_i)$$

$$E(g(X)) = \sum g(x_i) f(x_i)$$

  - 연속형

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

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

확률변수 X의 기댓값

- 어느 자동차 판매영업소의 1주일간 자동차 판매대수 확률분포

  'X = 1주일간 판매대수'

x_i	0	1	2	3	4	5	계
$f(x_i)$	0.1	0.1	0.2	0.3	0.2	0.1	1.0

$$E(X) = \sum x_i f(x_i)$$

$$= 0 \times 0.1 + 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.2 + 5 \times 0.1$$

$$= 2.7$$

확률변수 X의 분산

- $Var(X) = \sigma^2 = E[(X-\mu)^2]$
- 확률변수 X의 확률밀도함수를 $f(x)$라고 하면
  - 이산형

$$\sum_i (x_i - \mu)^2 f(x_i)$$

  - 연속형

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

- $\sigma = \sqrt{Var(x)}$: 표준편차

예제1

기댓값과 분산의 계산

X = '동전을 2개 던질 때 앞면이 나온 횟수'

풀이

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

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

$$\begin{aligned} Var(X) &= E(X^2) - \mu^2 \\ &= 0^2 \times \frac{1}{4} + 1^2 \times \frac{2}{4} + 2^2 \times \frac{1}{4} - 1 = \frac{1}{2} \end{aligned}$$

새로운 확률변수 $aX + b$

- 새로운 확률변수 $aX + b$의 기댓값과 분산
  - $E(aX + b) = aE(X) + b$
  - $Var(aX + b) = a^2 Var(X)$

  여기서, a와 b는 임의의 상수

표준화된 확률변수
(standardized random variable)

- 평균이 μ, 표준편차가 σ인 확률변수 X에 대해서
  - 표준화된 확률변수 $Z = \frac{X-\mu}{\sigma}$
  - 확률변수 Z의 평균은 0, 분산은 1

다음시간 안내

6강

확률분포와 표본분포1