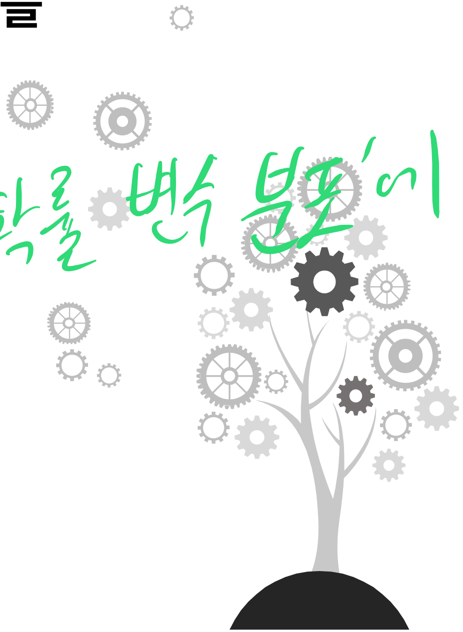
Contents

→ 전부 이산확률 변수 분포에 해당한다.



이항분포와 그에 관련된 분포들

이산 확률 변수 분포 중 하나!!!

01 베르누이 시행

- 베르누이 시행 (Bernoulli trial) : 성공과 실패의 두 가지 중에서 하나가 나타나는 실험

① 베르누이 시행의 각 시행은 성공과 실패의 두 결과만을 갖는다.
② 각 시행에서 성공할 확률은 p, 실패할 확률은 $1-p$로 일정하다.
③ 각 시행은 서로 독립으로 서로 영향을 미치지 않는다.

독립시행

해당 실험의 근원사건으로만 표본공간은 두 개의 이루어져 있음

$P(S)=p$, $P(F)=1-p$

성공 실패

두 확률 합하면 '1'

- 예 : 동전을 1번 던져서 앞면이 나오는 실험
앞면이 나오면 성공(1), 뒷면이 나오면 실패(0)

성공을 확률변수로 나타낼 때, '1'로 나타냄

확률이 p라면

$\Omega = \{1, 0\}$ 이것의 확률은 1-p.

- 예제 1, 예제 2

확률 변수 X : 앞면이 나오는 횟수

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

이항분포와 그에 관련된 분포들

- 베르누이 분포 (Bernoulli distribution) :
  X : 성공의 확률이 p인 베르누이 시행의 결과

  $X \sim \text{Bernoulli}(p)$:
  확률변수 X는 모수가 p인 베르누이 분포를 따른다

베르누이 시행을 한 번할.

- X의 확률분포

X	$P(X = x)$
0 실패	$1 - p$
1 성공	p
합계	1

X는 오직 0과 1.

$P(X = x) = p^x \cdot (1 - p)^{1-x}$, $x = 0, 1$

$E(X) = p$, $E(X^2) = \cancel{x}\ p$

$0^2 \times P(X=0) + 1^2 \times P(X=1)$
$= 1^2 \times P(X=1) = 1^2 \times p$
$= p$

$Var(X) = p(1 - p)$

$Var(X) = E(X^2) - \{E(X)\}^2$
$= p - p^2$
$= p(1-p)$

- 모수 p(베르누이 시행의 성공 확률)에 대해서 표본으로부터 추론

이항분포와 그에 관련된 분포들

베르누이 시행이 이항분포를 따른다.

02 이항분포

베르누이 시행을 n번 시행했을 때, 'n번 중에서 성공의 수'를 확률변수로 갖는 분포.

여러번 시행

- 이항 분포 (Binomial distribution) :
  X : m번 베르누이 시행 중 성공의 수
  (각 베르누이 시행은 서로 독립이며 동일한 성공 확률 p를 가짐)

'이항분포'

$X \sim \text{Binomial}(m, p)$ or $X \sim \text{Bin}(m, p)$:
확률변수 X는 모수가 (m, p)인 이항분포를 따른다

$m=1$ 인 이항분포 <=> 베르누이 분포.

'~분포를 따른다.'

- X의 확률분포

$$P(X=x) = \binom{m}{x} p^x \cdot (1-p)^{m-x}, \quad x = 0, 1, \cdots, m$$

$$E(X) = mp, \quad Var(X) = mp(1-p)$$

q

- 모수 p(베르누이 시행의 성공 확률)에 대해서 표본을 이용하여 추론
- 모수 m(베르누이 시행의 시행 회수)는 처음부터 주어지는 정보

표본 집단에게 실험을 시켜 모수 p를 추론함

Chapter 7 이항분포와 그에 관련된 분포들

예) $X \sim \text{Bin}(m = 5, p = 0.7)$

$$P(X = 1) = \binom{5}{1} \times 0.7^{1} \times 0.3^{5-1}$$

- 이항분포는 서로 독립이고 같은 분포를 갖는 베르누이 분포들의 합 ⇒ 즉, 여러번의 시행들이 서로 '독립'이다.

$$X = \sum_{i=1}^{m} X_i \sim \text{Bin}(m, p)$$

m: 총 시행 횟수

$$X_i \sim \text{Bernoulli}(p), \quad i = 1, \cdots, m, \quad E(X_i) = p, \quad Var(X_i) = p(1-p)$$

← 베르누이 분포

$$P(X_i = x_i) = \binom{1}{x_i} p^{x_i} \cdot (1-p)^{1-x_i} = p^{x_i} \cdot (1-p)^{1-x_i}, \quad x_i = 0, 1$$

X: m번 중 성공 횟수

⇒ 성공 횟수 : x번

('c'라고 쓰임)

↑ '애는 무조건 1'

$$P(X = x) = \binom{m}{x} p^{x} \cdot (1-p)^{m-x}, \quad x = 0, 1, \cdots, m$$

← 이항분포.

ex) $(1-p) \cdot p \cdot (1-p) \cdot p \cdot p \cdots$

$$E(X) = mp, \quad Var(X) = mp(1-p)$$

↑ 이 경우같은 다른 경우도 존재한다.!

=) 그래서 조합이 곱해진다.

Chapter 7 이항분포와 그에 관련된 분포들

- 부록 표-1

이항 분포의 확률 변수는 '성공 횟수'를 의미함.

부록에 있는 표:

$$P(X \le c) = \sum_{x=0}^{c} {}_{m}C_{x} \cdot p^{x}(1-p)^{m-x}, \quad {}_{m}C_{x} = \binom{m}{x}$$

성공 횟수

성공확률(p)

n=시행수
(c=성공횟수)

P(X≤c)

$X \sim \mathrm{Bin}(m = 3, p = 0.3), \quad x = 0, 1, \cdots, m = 3$

ex) X~Bin(3, 0.3)

P(X=2)
=P(X≤2) − P(X≤1)

		$\cdots$	$p = 0.3$	$\cdots$
$m = 3$	$c = 0$		$P(X \le 0)$	
	$c = 1$		$P(X \le 1)$	
	$c = 2$		$P(X \le 2)$	
	$c = 3$		$P(X \le 3)$	

Chapter 7 이항분포와 그에 관련된 분포들

- 이항분포에서의 확률계산

부록 표-1의 확률표를 이용하여 아래의 확률들을 계산 가능

$P(1 \le X \le 4) = P(X \le 4) - P(X \le 0)$

$$P(a \le X \le b) = P(X \le b) - P(X \le a-1)$$
$$P(a < X < b) = P(X \le b-1) - P(X \le a)$$
$$P(a \le X < b) = P(X \le b-1) - P(X \le a-1)$$
$$P(a < X \le b) = P(X \le b) - P(X \le a)$$

$P(1 < X < 5) = P(X \le 4) - P(X \le 1)$

$$P(X = a) = P(X \le a) - P(X \le a-1)$$
$$P(X \ge a) = 1 - P(X \le a-1)$$

'이산 확률 변수'이기 때문에 가능.

확률 계산 시, 부등호와 등호에 따른 차이에 주의

$P(1 \le X < 4)$

$P(2 < X \le 6) = P(X \le 6) - P(X \le 2) = P(X \le 3) - P(X \le 0)$

N개

그룹1 D개 | 그룹2 N-D개

03 초기하분포

<유한한 모집단에서 비복원 추출하는 경우.>

- 초기하분포 (Hypergeometric distribution) :
  X : D개의 원소로 이루어진 그룹 1과 $N-D$개의 원소로 이루어진 그룹 2에서 비복원추출한 m개의 표본 내 그룹 1의 원소의 수

N : 총 원소의 개수

D : 그룹 1에 포함된 원소 개수

X : m개 중 1번 그룹의 원소수

유한 모집단의 개수

$X \sim \text{Hypergeometric}(N, m, D)$:
확률변수 X는 모수가 (N, m, D)인 초기하분포를 따른다

- X의 확률분포

$$P(X=x) = \frac{\binom{D}{x}\times\binom{N-D}{m-x}}{\binom{N}{m}}, \quad x = 0, 1, \cdots, m$$

$=N$, $=m$

$$\frac{{}_{D}C_{x} \times {}_{N-D}C_{m-x}}{{}_{N}C_{m}}$$

${}_{n}C_{m}$

= 모집단의 크기

N : 총 원소의 개수, D : 그룹 1의 원소의 수, $N-D$: 그룹 2의 원소의 수
m : 비복원추출한 표본의 개수, X : 표본 내 그룹 1의 원소의 수

(비복원 추출하는

이항분포와 그에 관련된 분포들

- 초기하분포의 평균과 분산, 초기하분포의 이항분포 근사

Let $p = \frac{D}{N}$,

그룹1의 원소의 개수

총 원소의 개수

$$E(X) = m \cdot \frac{D}{N} = m \cdot p, \quad Var(X) = mp(1-p) \cdot \frac{N-m}{N-1} \xrightarrow{m < 0.05\,N} mp(1-p)$$

베르누이 분포와 형태가 비슷하지만, 절대 베르누이 분포가 아니다.

비복원 추출이기 때문!!!

(추출한 표본의 개수가 모집단의 크기의 5%보다 작다면,)

(if $N \to \infty$, $\frac{N-m}{N-1} \to 1$)

"N이 월등히 크다면"

$$\therefore X \sim \text{Hypergeometric}(N, m, D) \xrightarrow{m < 0.05\,N} \text{Bin}\left(m, p = \frac{D}{N}\right)$$

N(총 원소의 개수)이 m(비복원추출한 표본의 개수) 보다 월등히 크다면, 초기하분포에서의 각각의 비복원추출을 복원추출(베르누이 시행)로 볼 수 있다. 그러면 X를 m번 베르누이 시행 중 성공의 수로 취급할 수 있다.

So, 해당분포는 '이항분포'가 되어진다.