04 CHAPTER

이산형 확률분포

1

제4-3장 이산형 확률분포

1. 이산균일분포
2. 이항분포
3. 초기하분포
4. 포아송분포
5. 기하분포
6. 음이항분포*
7. 다항분포*

2

1. 이산균일분포

[정의 6-1] 이산균일분포(discrete uniform distribution)
n 개의 결과값이 균일한 확률로 발생하는 확률분포

- pdf $f(x)=\frac{1}{n},\ x=1,2,\ldots,n$
- 기댓값 $E(X)=\sum_{x=1}^{n} x\left(\frac{1}{n}\right)=\frac{1}{n}\times\frac{n(n+1)}{2}=\frac{n+1}{2}$
- 분산 $E(X^2)=\sum_{x=1}^{n} x^2\left(\frac{1}{n}\right)=\frac{1}{n}\times\frac{n(n+1)(2n+1)}{6}=\frac{(n+1)(2n+1)}{6}$

$$\Rightarrow Var(X)=E(X^2)-E(X)^2$$
$$=\frac{(n+1)(2n+1)}{6}-\left(\frac{n+1}{2}\right)^2=\frac{(n+1)(n-1)}{12}$$

3/23

3

1. 이산균일분포

[예 6-1] 1에서 20까지 번호가 적혀 있는 동일한 20개의 공이 들어 있는 상자에서 임의로 하나의 공을 꺼냈을 때 나온 번호 X

(1) X의 확률분포함수 $f(x)=\frac{1}{20},\ x=1,2,\ldots,20$

(2) X의 기댓값과 분산 $E(X)=\frac{20+1}{2}=10.5$

$$Var(X)=\frac{(20+1)(20-1)}{12}=33.25$$

(3) 15 이상의 번호가 나올 확률

$$P(X\geq 15)=\sum_{x=15}^{20}\frac{1}{20}=\frac{6}{20}=0.3$$

4/23

4

2. 이항분포

[정의 6-2] 베르누이분포(Bernoulli distribution) $X \sim B(1, p)$
: 성공 확률이 일정한 1회의 시행에서 나오는 성공 횟수의 확률분포

- pdf $f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases} \Leftrightarrow f(x) = p^x(1-p)^{1-x},\ x = 0, 1$
- 기댓값 $E(X) = \sum_{x=0}^{1} xf(x) = 0 \times (1-p) + 1 \times p = p$
- 분산 $E(X^2) = \sum_{x=0}^{1} x^2 f(x) = 0^2 \times (1-p) + 1^2 \times p = p$
  $\Rightarrow Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$
- MGF $m(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx} f(x) = e^0 \times (1-p) + e^t \times p = 1 - p + pe^t$
  $\Rightarrow E(X) = m'(0) = pe^0 = p \quad E(X^2) = m''(0) = pe^0 = p$
  $\Rightarrow Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$

5/23

5

2. 이항분포

[정의 6-3] 이항분포(binomial distribution) $X \sim B(n, p)$
: 성공 확률이 일정한 n 회의 시행에서 나오는 성공 횟수의 확률분포

- Pdf $f(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, 2, \ldots, n$
  $X_1, X_2, \cdots, X_n \overset{iid}{\sim} B(1, p) \Rightarrow X \equiv \sum_{i=1}^{n} X_i \sim B(n, p)$
- 기댓값 $E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$
- 분산 $Var(X) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$
- MGF $m(t) = m_{X_1}(t) \times m_{X_2}(t) \times \cdots \times m_{X_n}(t) = (1 - p + pe^t)^n$

6/23

6

$X \sim B(5, p)$ $\quad f(x) = \binom{5}{x} p^x (1-p)^{5-x}, \ x = 0, 1, 2, \ldots, 5$

시행 결과 ●=성공, ○=실패	확률 $P(X=x)$	시행 결과 ●=성공, ○=실패	확률 $P(X=x)$
○○○○○	$f(0) = \binom{5}{0}(1-p)^5$	●●●●●	$f(5) = \binom{5}{5}p^5$
●○○○○		●●●●○	
○●○○○		●●●○●	
○○●○○	$f(1) = \binom{5}{1}p(1-p)^4 = 5p(1-p)^4$	●●○●●	$f(4) = \binom{5}{4}p^4(1-p) = 5p^4(1-p)$
○○○●○		●○●●●	
○○○○●		○●●●●	
●●○○○		●●●○○	
●○●○○		●●○●○	
●○○●○		●●○○●	
●○○○●		●○●●○	
○●●○○	$f(2) = \binom{5}{2}p^2(1-p)^3 = 10p^2(1-p)^3$	●○●○●	$f(3) = \binom{5}{3}p^3(1-p)^2 = 10p^3(1-p)^2$
○●○●○		●○○●●	
○●○○●		○●●●○	
○○●●○		○●●○●	
○○●○●		○●○●●	
○○○●●		○○●●●	

2. 이항분포

\# 이항분포 R 함수

\# 확률분포함수 (size=n=표본크기, prob=p=성공확률)

dbinom(x, size, prob)

\# Excel 함수 = BINOM.DIST(x, size, prob, FALSE)

\# 누적분포함수 (q=분위수, lower.tail=TRUE=아래로부터 누적)

pbinom(q, size, prob, lower.tail = TRUE)

\# Excel 함수 = BINOM.DIST(x, size, prob, TRUE)

\# 분위수 (p=누적확률)

qbinom(p, size, prob, lower.tail = TRUE)

\# Excel 함수 = BINOM.INV(size, prob, p)

\# 이항 확률변수(n=난수의 개수)

rbinom(n, size, prob)

\# Excel 함수는 없으나, 아래와 같이 한 개의 난수 생성

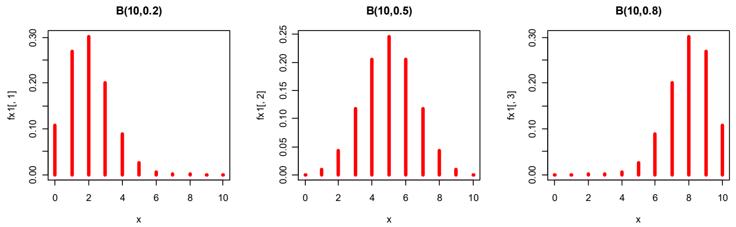
= BINOM.INV(size, prob, RAND( ))

[예 6-2] 성공확률이 각각 0.2, 0.5, 0.8인 무한모집단에서 10개씩 표본을 취하였을 때 나타나는 성공회수의 확률분포

(1) p=0.2 $X \sim B(10, 0.2)$ $\quad f(x) = \binom{10}{x}(0.2)^x(0.8)^{10-x},\ x = 0, 1, 2, \dots, 10$

(2) p=0.5 $X \sim B(10, 0.5)$ $\quad f(x) = \binom{10}{x}(0.5)^x(0.5)^{10-x},\ x = 0, 1, 2, \dots, 10$

(3) p=0.8 $X \sim B(10, 0.8)$ $\quad f(x) = \binom{10}{x}(0.8)^x(0.2)^{10-x},\ x = 0, 1, 2, \dots, 10$



2. 이항분포

[예 6-3] 불량률이 0.03인 공정에서 20개의 표본을 추출하여 검사하여 발견한 불량개수 X $X \sim B(20, 0.03)$

(1) 확률분포함수 $\quad f(x) = \binom{20}{x}(0.03)^x(0.97)^{20-x},\ x = 0, 1, 2, \dots, 20$

(2) 평균과 분산 $\quad E(X) = 20 \times 0.03 = 0.6 \qquad Var(X) = 20 \times 0.03 \times 0.97 = 0.582$

(3) P(X = 2) $\quad f(2) = \binom{20}{2}(0.03)^2(0.97)^{20-2} = 190 \times (0.03)^2(0.97)^{18} \doteq 0.099$

(4) P($X \geq$ 3)

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - [f(0) + f(1) + f(2)]$$
$$= 1 - [{}_{20}C_0(0.03)^0(0.97)^{20} + {}_{20}C_1(0.03)^1(0.97)^{19} + {}_{20}C_2(0.03)^2(0.97)^{18}]$$
$$= 1 - [(0.97)^{20} + 20 \times 0.03 \times (0.97)^{19} + 190 \times (0.03)^2(0.97)^{18}]$$
$$\doteq 1 - (0.544 + 0.336 + 0.099) = 1 - 0.979 = 0.021$$

```
dbinom(0:2, 20, 0.03)
[1] 0.54379434 0.33636763 0.09882967
1-sum(dbinom(0:2, 20, 0.03)); pbinom(2, 20, 0.03, lower=F)
[1] 0.02100836        [1] 0.02100836
```

4. 포아송분포

[정의 6-5] 포아송분포(Poisson distribution) $Poi(\lambda)$

: 일정한 단위에서 발생한 희소한 사건수의 확률분포

$$f(x) = \lambda^x \frac{e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

$$\lim_{n\to\infty} f(x) = \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \lim_{n\to\infty} \frac{n!}{x!(n-x)!} p^x (1-p)^x$$

$$\lambda = np \Rightarrow p = \frac{\lambda}{n} \quad = \frac{1}{x!} \lim_{n\to\infty} \frac{n!}{(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n / \left(1 - \frac{\lambda}{n}\right)^x$$

$$= \lambda^x \frac{e^{-\lambda}}{x!}$$

11/23

4. 포아송분포

- 확률분포함수의 조건

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \lambda^x \frac{e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

- MGF, 평균 및 분산

$$m(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \lambda^x \frac{e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$$m'(t) = \lambda e^t e^{\lambda(e^t-1)} \Rightarrow E(X) = m'(0) = \lambda e^0 e^0 = \lambda$$

$$m''(t) = \lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)} \Rightarrow E(X^2) = m''(0) = \lambda + \lambda^2$$

$$\Rightarrow Var(X) = E(X^2) - E(X)^2 = \lambda$$

12/23

4. 포아송분포

포아송분포 R 함수

# 확률분포함수 (lambda=기댓값)
dpois(x, lambda)
# Excel 함수 = POISSON.DIST(x, lambda, FALSE)

# 누적분포함수 (q=분위수, lower.tail=TRUE=아래로부터 누적)
ppois(q, lambda, lower.tail = TRUE)
# Excel 함수 = POISSON.DIST(x, lambda, TRUE)

# 분위수 (p=누적확률)
qpois(p, lambda, lower.tail = TRUE)

# 포아송 확률변수(n=난수의 개수)
rpois(n, lambda)

13/23

13

4. 포아송분포

[예 6-6] 일정 단위당 평균 발생회수가 각각 2개, 5개, 8개인 세 종류의 무한모집단에서 일정 단위의 표본을 취하였을 때, 포아송 확률분포

(1) 단위당 평균 발생회수가 2인 경우 $Poi(2)$

$$f(x) = 2^x \frac{e^{-2}}{x!}, \quad x = 0, 1, 2, \ldots$$

(2) 단위당 평균 발생회수가 5인 경우 $Poi(5)$

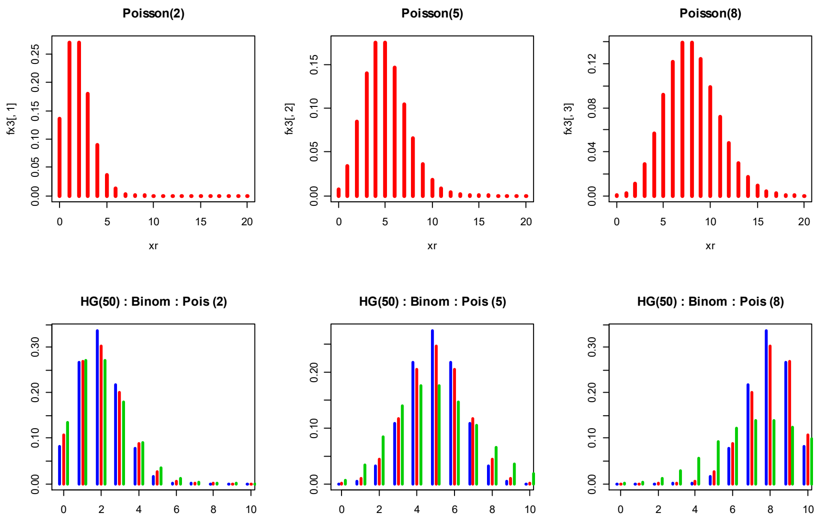
$$f(x) = 5^x \frac{e^{-5}}{x!}, \quad x = 0, 1, 2, \ldots$$

(3) 단위당 평균 발생회수가 8인 경우 $Poi(8)$

$$f(x) = 8^x \frac{e^{-8}}{x!}, \quad x = 0, 1, 2, \ldots$$

14/23

14



4. 포아송분포

[예 6-7] 단위당 평균 결점수가 1.5 개인 제품을 생산하는 프로세스에서 샘플링검사 실시 $X \sim Poi(1.5)$

(1) 확률분포함수 $f(x) = (1.5)^x \dfrac{e^{-1.5}}{x!},\ x = 0, 1, 2, \ldots$

(2) 기댓값 $E(X) = \lambda = 1.5$

(3) 분산 $Var(X) = \lambda = 1.5$

(4) P(2개) $f(2) = (1.5)^2 \dfrac{e^{-1.5}}{2!} \doteq 0.251$

(5) P(3개 이상) $P(X \geq 3) = 1 - P(X \leq 2) = 1 - e^{-1.5} \times (1 + 1.5 + 1.5^2 / 2) \doteq 0.191$

(6) P(10단위에서 20개) $Y \sim Poi(15) \Rightarrow P(Y = 20) = 15^{20} \times \dfrac{e^{-15}}{20!} \doteq 0.042$

```
dpois(0:2, 1.5)
[1] 0.2231302 0.3346952 0.2510214
1-sum(dpois(0:2, 1.5)); ppois(2, 1.5, lower=F)
[1] 0.1911532        [1] 0.1911532
```

5. 기하분포

[정의 6-6] 기하분포(geometric distribution) $X \sim G(p)$

: 성공 확률이 일정한 시행에서 첫 번째 성공이 발생할 때까지 시행한 횟수의 확률분포

$$f(x) = (1-p)^{x-1} p,\ x = 1, 2, \ldots$$

첫 번째 성공이 나오기까지 시행 시나리오 (● : 성공 ○ : 실패)	총 시행횟수 (확률변수 X)	확률 $P(X=x)$
●	$X=1$	p
○ ●	$X=2$	$(1-p)p$
○ ○ ●	$X=3$	$(1-p)^2p$
○ ○ ○ ●	$X=4$	$(1-p)^3p$
○ ○ ○ ○ ●	$X=5$	$(1-p)^4p$
○ ○ ○ ○ ○ ●	$X=6$	$(1-p)^5p$
⋮	⋮	⋮
○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ⋯ ○ ●	$X=x$	$(1-p)^{x-1}p$

17/23

17

5. 기하분포

- 확률분포함수의 조건

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} p = \frac{p}{1-(1-p)} = 1$$

- 누적분포함수

$$F(x) \equiv P(X \le x) = \sum_{y=1}^{x} (1-p)^{y-1} p = \frac{p[1-(1-p)^x]}{1-(1-p)} = 1-(1-p)^x$$

- 모멘트생성함수

$$m(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p$$

$x-1 \rightarrow y$

$$= pe^t \sum_{y=0}^{\infty} (e^t q)^y = \frac{pe^t}{1-qe^t},\ qe^t < 1$$

18/23

18

5. 기하분포

- 기댓값과 분산 $E(X)=\dfrac{1}{p}$ $Var(X)=\dfrac{1-p}{p^2}$ $m(t)=\dfrac{pe^t}{1-qe^t}$

$$m'(t)=\frac{pe^t(1-qe^t)+pe^tqe^t}{(1-qe^t)^2}=\frac{pe^t}{1-qe^t}+\frac{pqe^{2t}}{(1-qe^t)^2}$$

$$\Rightarrow E(X)=m'(0)=\frac{p}{p}+\frac{p(1-p)}{p^2}=\frac{1}{p}$$

$$m''(t)=\frac{pe^t(1-qe^t)+pe^tqe^t}{(1-qe^t)^2}+\frac{2pqe^{2t}(1-qe^t)^2+2pqe^{2t}q(1-qe^t)}{(1-qe^t)^4}$$

$$\Rightarrow E(X^2)=m''(0)=\frac{p^2+p(1-p)}{p^2}+\frac{2pqp^2+2pq^2p}{p^4}$$

$$=\frac{p}{p^2}+\frac{2p(1-p)+2(1-p)^2}{p^2}=\frac{2-p}{p^2}$$

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

19/23

19

5. 기하분포

\# 기하분포 R 함수

\# Excel 함수 없음

\# 확률분포함수 (x=실패횟수, prob=p=성공확률)

```
dgeom(x, prob)
```

\# 누적분포함수 (q=분위수, lower.tail=TRUE=아래로부터 누적)

```
pgeom(q, prob, lower.tail = TRUE)
```

\# 분위수 (p=누적확률)

```
qgeom(p, prob, lower.tail = TRUE)
```

\# 기하 확률변수(n=난수의 개수)

```
rgeom(n, prob)
```

20/23

20

5. 기하분포

[예 6-8] 성공확률이 각각 0.1, 0.2, 0.3, 0.5인 네 유형의 무한모집단에서 첫 번째 성공을 얻을 때까지 시행

$$f(x) = p(1-p)^{x-1}$$

(1) 성공확률이 0.1인 경우 $f(x) = 0.1 \times 0.9^{x-1},\ x = 1, 2, 3, \ldots$

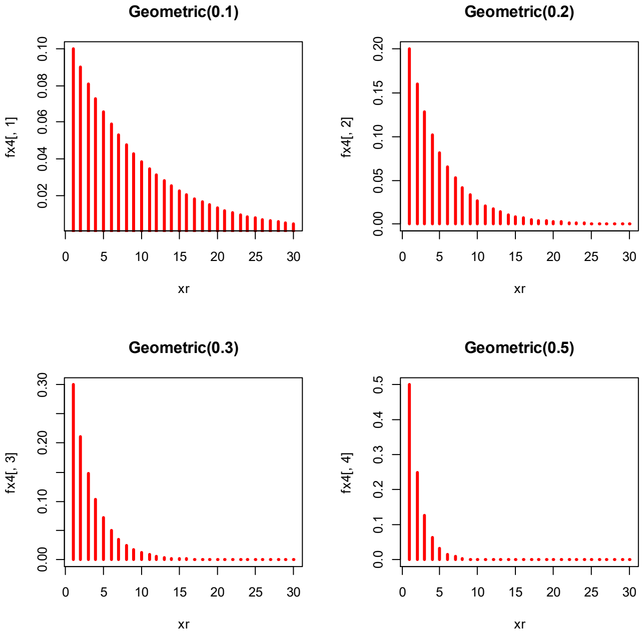
(2) 성공확률이 0.2인 경우 $f(x) = 0.2 \times 0.8^{x-1},\ x = 1, 2, 3, \ldots$

(3) 성공확률이 0.3인 경우 $f(x) = 0.3 \times 0.7^{x-1},\ x = 1, 2, 3, \ldots$

(4) 성공확률이 0.5인 경우 $f(x) = 0.5 \times 0.5^{x-1},\ x = 1, 2, 3, \ldots$

21/23

21



22/23

22

5. 기하분포

[예 6-9] 주사위 1개를 '6'이 나올 때까지 반복해서 굴리는 실험에서 총 시행회수 X

(1) X의 확률분포함수 $f(x)=\frac{1}{6}\left(\frac{5}{6}\right)^{x-1}, \ x=1,2,3,\ldots$

(2) X의 기댓값과 분산 $E(X)=\frac{1}{p}=6 \quad Var(X)=\frac{q}{p^2}=\frac{5/6}{(1/6)^2}=30$

(3) 3회의 시행 이내에 '6'이 나올 확률 $F(3)=1-q^3=1-\left(\frac{5}{6}\right)^3=\frac{91}{216}\doteq 0.421$

- 비기억 특성 (memoryless property)

$$P(X>x+y \mid X>x)=P(X>y)$$

$$P(X>x+y \mid X>x)=\frac{P(X>x+y)}{P(X>x)}=\frac{q^{x+y}}{q^x}=q^y=P(X>y)$$