경영통계

원광대학교 경영학부

담당교수: 정호일

第6장 확률분포

이산확률분포

- Bernoulli 확률분포
- 이항분포
- 포아송분포
- 초기하분포

I. Bernoulli 시행

1. 베르누이 시행이란?
   실험이나 관찰의 결과 발생 가능한 사상이 단 두 가지인 경우의 시행을 말함.
2. 베르누이 시행 조건
   1) 베르누이 시행결과를 확률변수 X라 할 때 확률변수 X는 0 또는 1이다. 흔히 x=1의 사상을 성공, x=0의 사상을 실패라 한다.
   2) 각 실험에서 성공할 확률은 p , 실패할 확률은 (1-p)로 일정하고 성공확률과 실패확률의 합은 1이다.
   3) 여러 번의 베르누이 시행은 서로 독립이다.

I. Bernoulli 시행

3. 베르누이 확률분포
   확률변수를 X로 하는 베르누이 확률분포는 다음과 같다.
   실패:0, 성공:1 , p:성공확률 (1-p):실패확률

X	P(X)
0	1-p
1	p

4. 확률분포의 기댓값, 분산,표준편차

$$E(X) = 0 \times (1-p) + 1 \times p = p$$

$$Var(X) = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p)$$

$$\sigma(X) = \sqrt{p(1-p)}$$

I. 베르누이 확률분포 예제

예제1)

회덕분기점에서 광주방면과 부산방면으로 고속도로가 분기된다. 자동차 1,000대를 조사한 결과 광주방면은 400대, 부산방면은 600대이었다. 서울톨게이트를 출발한 자동차 한대를 임의로 뽑았을 때 그 차가 광주방면으로 갈 확률(p)은 0.4이다. 광주로 가는 차를 성공=1이라 하고 나머지는 실패=0이라 했을 경우 이의 기댓값 및 분산을 계산하라.

(풀이)

$E(X) = 0\times0.6 + 1\times0.4 = 0.4$

$Var(X)=0.4(1-0.6)=0.16$

예제2) 1 또는 3의 눈이 나오는 것을 '성공'이라 간주하고 기타는 실패로 하는 주사위던지기 실험의 결과를 확률변수 X로 하는 확률분포와 기댓값 및 분산은?

II. 이항분포

1. 이항실험(Binomial experiment)
   베르누이 시행조건 외에 동일한 시행을 n번 반복 할 때 나타나는 결과의 수에 관심을 가질 때 반복적인 베르누이 시행을 이항실험이라고 함.

2. 이항분포(Binomial probability distribution)
   베르누이 시행을 한번 이상 반복했을 경우 **시행결과의 합**을 **변수값**으로 하는 확률변수의 분포를 이항분포라 한다. 즉, 여러 번의 시행결과 각 시행의 결과가 성공과 실패 2가지의 경우로만 나타나고, 이 때 여러 번의 시행결과의 성공 또는 실패횟수의 합을 이항확률변수라고 하며 이러한 이항확률변수의 확률분포를 이항확률분포(이항분포)라 한다.

Ⅱ. 이항분포

- ‘성공’할 확률이 p이고 반복횟수가 n인 베르누이 시행에서, X=성공의 횟수라 할 때, 확률변수 X는 모수가 (n,p)인 **이항분포**(Binomial Distribution)를 따른다고 한다. 그리고 X～B(n,p)로 쓴다.
- 이항확률함수
  한 시행의 성공확률이 p이고 시행을 독립적으로 n회 반복할 때 성공횟수 X의 확률함수를 이항확률함수라고 함.

$$P(X = x) = {}_nC_x p^x (1-p)^{n-x} \quad (x = 0,1,2,\cdots,n)$$

여기서 ${}_nC_x$는 n번의 베르누이시행에서

성공 x개를 순서를 고려하지 않고

나열하는방법의수

Ⅱ. 이항분포

- 이항분포의 특징
  - 이항분포의 모양은 개별적인 베르누이 시행에서의 성공확률과 시행횟수에 따라 결정된다.
  - 베르누이의 각 시행은 상호독립적임.
  - 각각의 베르누이 시행에서 성공확률은 모두 동일함.
  - 이항분포의 확률변수 X는 n개의 독립적인 베르누이 시행의 합이므로 $X = X_1 + X_2 + X_3 + \ldots + X_n$ 이다.
- 이항분포의 기댓값과 분산

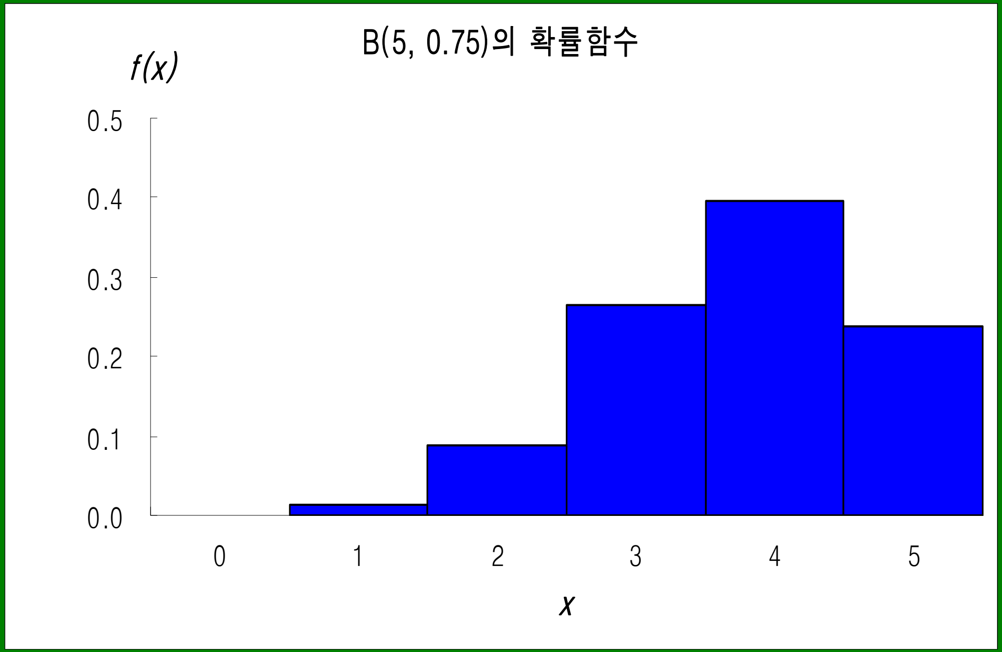
$$
\begin{aligned}
E(X) &= E(x_1 + x_2 + \cdots + x_n) \\
&= E(x_1) + E(x_2) + \cdots + E(x_n) \\
&= np \quad \because E(x_1) = E(x_2) = \cdots E(x_n) = p
\end{aligned}
$$

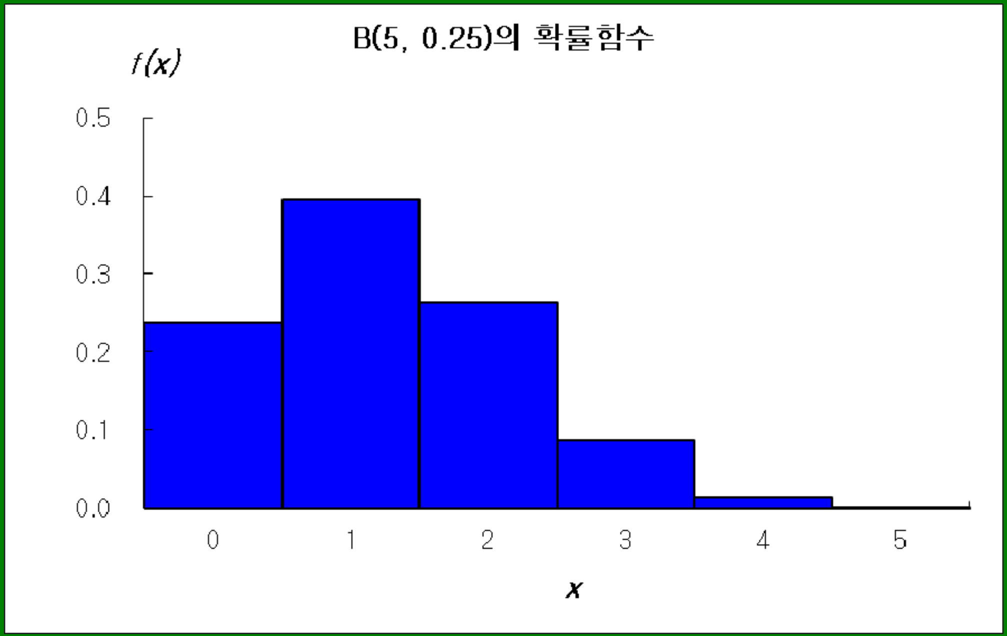
$$
\begin{aligned}
Var(X) &= Var(x_1 + x_2 + \cdots + x_n) \\
&= Var(x_1) + Var(x_2) + \cdots + Var(x_n) \\
&= np(1 - p) \\
&\because Var(x_i) = p(1 - p)
\end{aligned}
$$

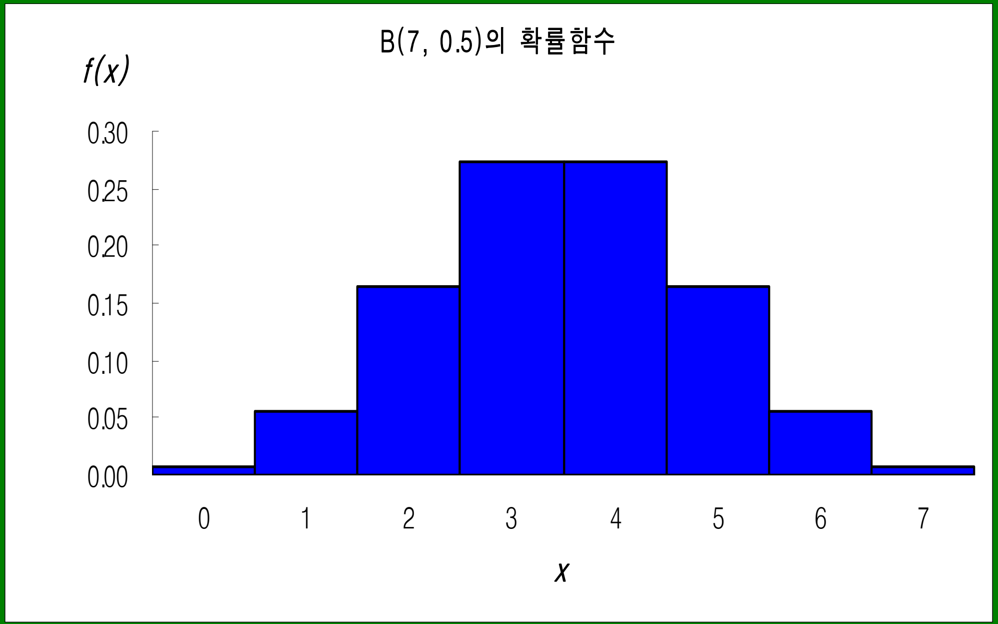
II. 이항분포

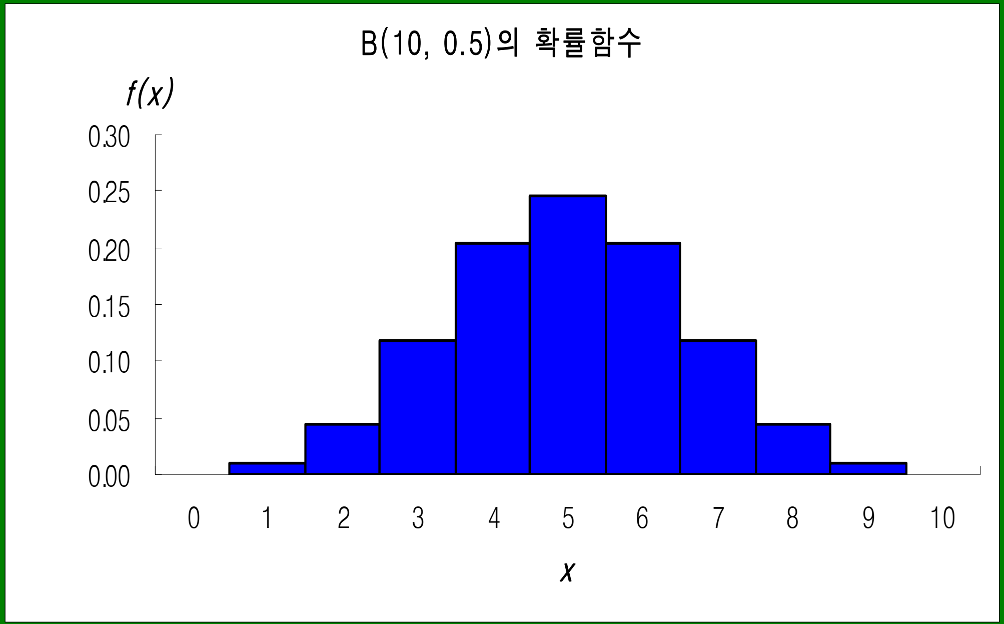
- 이항분포의 형태
  - X~B(n,p)　　n:시행횟수(반복횟수) p:성공확률
  - n(시행횟수): n이 커지면 연속성이 높아진다.
  - p=0.5이면 히스토그램이 좌우대칭이다.
  - p>0.5이면 히스토그램이 왼쪽 꼬리분포이다.
  - P<0.5이면 히스토그램이 오른쪽 꼬리분포이다.

Ⅱ. 이항분포


B(5, 0.75)의 확률함수
f(x)
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
x


B(5, 0.25)의 확률함수
f(x)
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
x


B(7, 0.5)의 확률함수
f(x)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6
7
x


B(10, 0.5)의 확률함수
f(x)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6
7
8
9
10
x

Ⅱ. 이항분포

예1) : 주사위를 3번 던져서 1의 눈이 나온 횟수라 할 때 확률분포를 나타내라.

(풀이) ‘성공’의 확률이 1/6이고 반복횟수가 3인 베르누이 시행, X~B(3, 1/6)
1의 눈이면 성공 S 아니면 실패 F라 할 때

X	경우의 수		확률
0	FFF	${}_3C_0$	${}_3C_0(1/6)^0\ (5/6)^3=0.579$
1	SFF FSF FFS	${}_3C_1$	${}_3C_1(1/6)^1\ (5/6)^2=0.347$
2	SSF SFS FSS	${}_3C_2$	${}_3C_2(1/6)^2\ (5/6)^1=0.069$
3	SSS	${}_3C_3$	${}_3C_3(1/6)^3\ (5/6)^0=0.005$

II. 이항분포

예2) :회덕분기점에서 광주방면으로 가는 것을 성공이라 하고 기타 방면으로 가는 것을 실패라 하자. 성공확률을 0.4라 할 때 임의로 5대를 뽑았을 때 2대 이하가 광주로 갈 확률을 구하라.

(풀이) '성공'의 확률이 0.4이고 반복횟수가 5인 베르누이 시행, X~B(5, 0.4)

$$P(X\leq 2)=P(X=0)+P(X=1)+P(X=2)$$

$$={}_5C_0(0.4)^0(0.6)^5+{}_5C_1(0.4)^1(0.6)^4+{}_5C_2(0.4)^2(0.6)^2$$

$$=0.0778+0.2592+0.3456=0.6826$$

만약 시행횟수가 커진다면 계산이 복잡할 것이다. 이러한 경우 이항분포표를 활용하면 편리하다.

방법: 이항확률함수의 모수(n,p)를 이용한다. 세로축: 시행횟수와 성공횟수 가로축:성공확률로 하여 n번의 베르누이 시행에 대한 x번의 성공확률을 계산해놓은 표.

예3) 어느 대학교에 다니는 학생들 중 50%의 학생이 버스를 이용하여 통학을 하고 있다고 한다. 임의로 20명의 학생을 추출할 때, 이들 중 12명 이상이 버스를 이용할 확률을 구하라.(이항분포표활용)

(풀이) $P(X\geq 12)=0.1201+0.0739+0.0370+0.0148+0.0046+0.0011+0.0002+0.0000+0.0000$

II. 이항분포

예4) 원광자동차(주) 부품공장에서 생산하는 특정부품의 불량률은 15%라고 하자. 검사원이 부품 100개를 무작위 추출했을 때 정상부품이 나오는 횟수를 확률변수 X라 할 때 그 가운데서 어느 정도가 정상품이라고 기대할 수 있는가? 분산과 표준편차는?

<풀이> 이항분포의 기댓값을 의미한다. 그러므로

$E(X)=0\times{}_{100}C_{0}(0.85)^{0}(0.15)^{100}+1\times{}_{100}C_{100}(0.85)^{1}(0.15)^{99}+\cdots+100\times$

${}_{100}C_{100}(0.85)^{100}(0.15)^{0}$ $=100\times0.85=85$개

$Var(X)=100\times0.85\times0.15=3.57$(개)

III. 포아송분포

일정한 단위시간, 단위공간, 단위거리,단위면적 등에서 발생하는 관심 있는 사건의 발생 가능한 횟수를 변수 값으로 하는 확률변수의 분포

포아송분포의 조건

- 어떤 단위시간에서 발생하는 평균수는 이 단위시간의 크기에 비례한다.
- 단위시간에서 발생하는 발생횟수는 단위시간간에 독립적이다.
- 충분히 작은 단위시간에서 둘 이상의 사상이 동시에 발생할 확률은 0이다.
- 동일한 간격에서 발생하는 사상의 확률은 동일하다.

III. 포아송분포

포아송분포 확률함수

$$P(X) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-\mu}(\mu)^x}{x!}$$

$$e = 2.71828$$

$$x = \text{발생횟수}$$

$$\lambda = \text{단위당 평균발생횟수}$$

$$t = \text{단위수} \quad \lambda t = \mu$$

III. 포아송분포

포아송확률함수 유도

$$P(X) = \lim_{n\to\infty} {}_nC_x P^x (1-P)^{n-x} \qquad \mu = np \text{라고 하면 } p = \frac{\mu}{n}$$

$$= \lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots.(n-x+1)}{X!} \left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{\mu^x}{x!} \frac{n(n-1)\ldots.(n-x+1)}{n^x} \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-x}$$

$$= \frac{\mu^x}{x!} \lim_{n\to\infty} \left(\frac{n}{n}\frac{n-1}{n}\frac{n-2}{n}\ldots\ldots\frac{n-x+1}{n}\right) \left(1-\frac{\mu}{n}\right)^n \left(1-\frac{\mu}{n}\right)^{-x}$$

$$= \frac{\mu^x}{x!} \lim_{n\to\infty} \left(1-\frac{\mu}{n}\right)^n$$

$$= \frac{\mu^x}{x!} e^{-\mu} \quad \because \lim_{n\to\infty} \left(1-\frac{\mu}{n}\right)^n = e^{-\mu}$$

Ⅲ. 포아송분포

포아송분포의 기댓값과 분산

$X \sim \text{Poisson}(\mu)$ 일 때,

(1) $E(X) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$

$= \sum_{x=1}^{\infty} x \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$

$= e^{-\mu} \cdot \mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!}$

$= e^{-\mu} \cdot \mu \cdot \sum_{y=0}^{\infty} \frac{\mu^y}{y!}$

$= e^{-\mu} \cdot \mu \cdot e^{\mu} = \mu.$

(2) $E(X^2) = \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$

$= \sum_{x=0}^{\infty} (x^2 - x) \frac{e^{-\mu} \cdot \mu^x}{x!} + \sum_{x=0}^{\infty} x \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$

$= \sum_{x=2}^{\infty} x(x-1) \frac{e^{-\mu} \cdot \mu^x}{x!} + \sum_{x=1}^{\infty} x \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$

$= e^{-\mu} \cdot \mu^2 \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} + \mu$

$= e^{-\mu} \cdot \mu^2 \sum_{y=0}^{\infty} \frac{\mu^y}{y!} + \mu = \mu^2 + \mu$

$\Longrightarrow \text{Var}(X) = E(X^2) - \{E(X)\}^2 = \mu^2 + \mu - (\mu)^2 = \mu$

기대값 $E(X) = \mu = \lambda t$

분산 $Var(X) = \sigma^2 = \mu$

III. 포아송분포

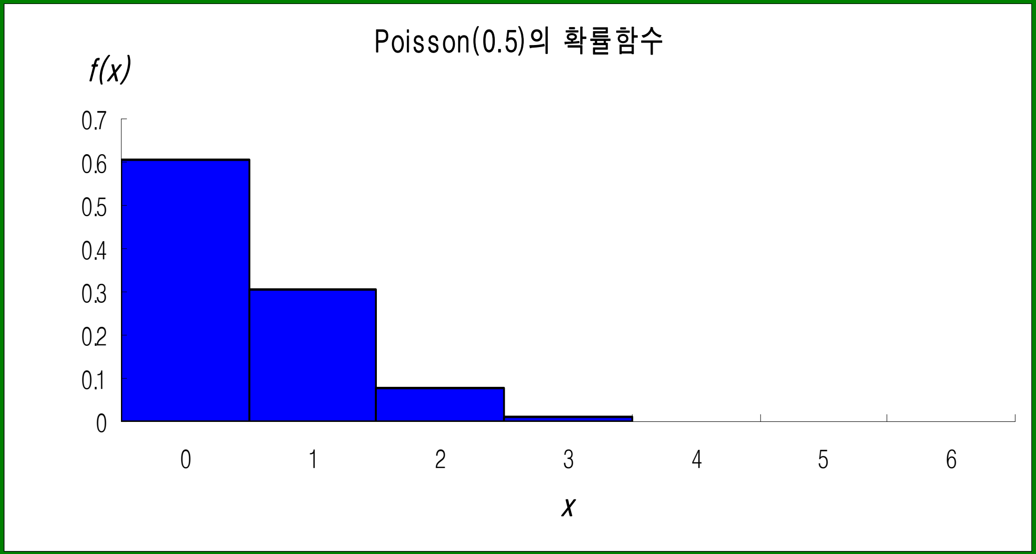
포아송분포와 이항분포의 관계

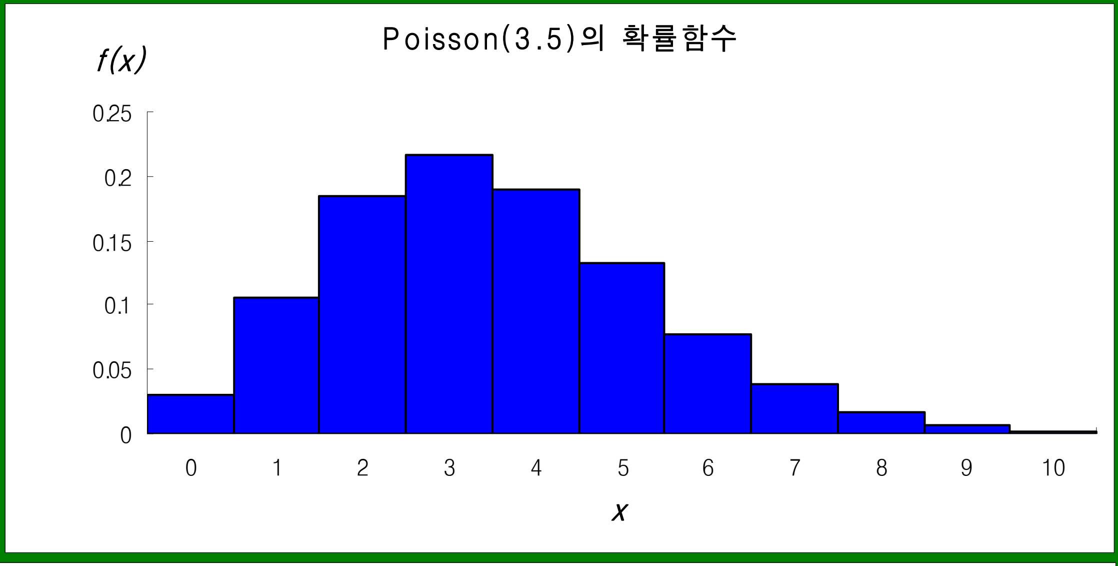
- 이항분포의 시행횟수(n)가 크고 성공확률(p)이 작으면 이항분포 대신 포아송 분포를 근사치로 이용할 수 있다. n이 무한 히 커질 수록 모수 n과 p를 갖는 이항확률 변수 X는 모수가 $\mu=np$인 포아송 확률변수가 된다.
- 일반적으로 $n>20$이고 $\mu<5$이면 이항분포는 포아송분포와 차이가 미미하다.

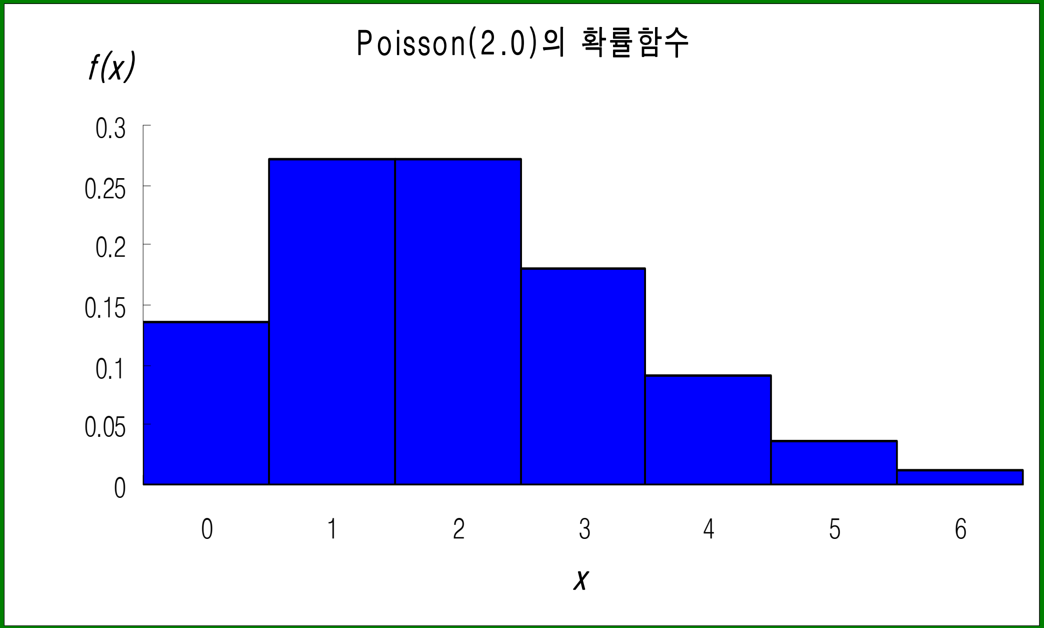
III. 포아송분포

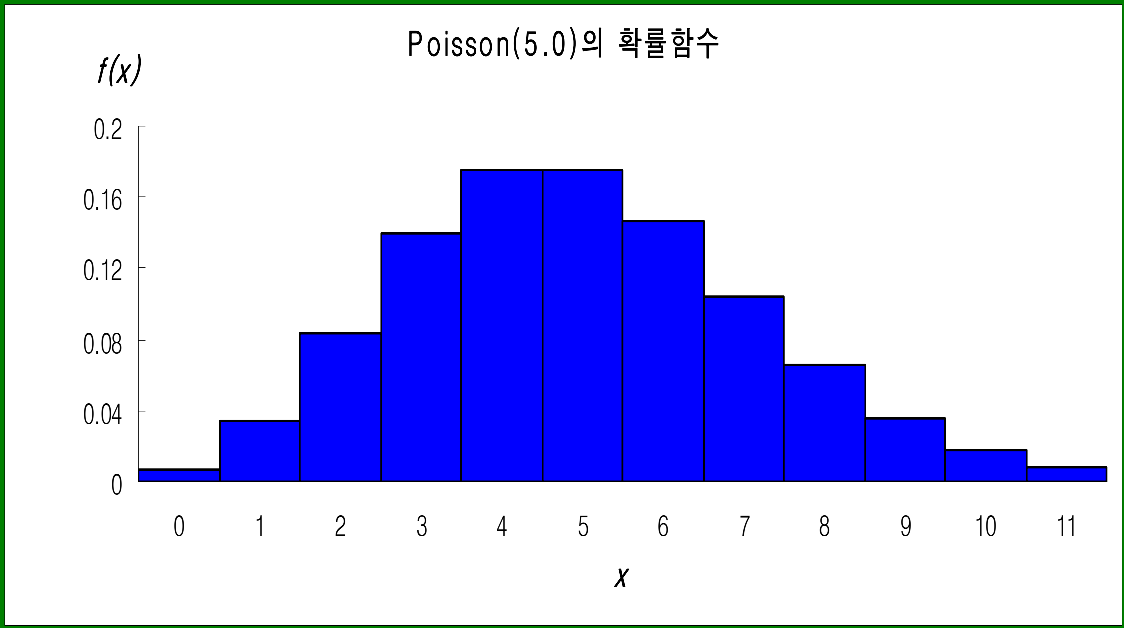
포아송분포 모양

$$X \sim Poisson(\mu)$$









III. 포아송분포

포아송분포의 사례

- 새 자동차에서 발견되는 흠집의 수
- 보험회사에서 접수한 매월의 사망자 의수
- 콜센터에 걸려오는 전화 건수
- 은행창구에 도착하는 시간당 고객의 수
- 증권회사의 분당 매도 건수
- 비단 한 필에서 발견되는 결점의 수
- A4 한 페이지당 오타 건수

Ⅲ. 포아송분포

포아송분포의 예제

예제1) 배가 항구에 시간당 두 척씩 도착한다고 한다. 이 과정이 포아송분포를 따른다고 할 때 30분 동안 배 한 척도 도착하지 않을 확률을 구하라.

(풀이)

$t=1/2$시간 $\mu=\lambda t=2(1/2)=1$ $P(X=0)=(e^{-1}1^{0})/0!=0.3679$

예제2) 어느 은행의 지점에 5분 동안 고객이 평균 네 명씩 도착한다고 한다. 포아송 분포의 조건이 만족된다고 할 때 포아송 분포를 구하라.

(풀이)

X	0	1	2	3	…………
P(X)	0.0183	0.0733	0.1456	0.1954	…………

IV. 초기하분포

초기하분포의 정의

N개(N_1+N_2)의 원소로 이루어진 모집단이 두 개의 그룹(성공집단과 실패집단)으로 분할되고 모집단으로부터 n개의 표본을 **비복원 추출하는 경우**, 성공집단에 속한 성공의 갯수를 확률변수 X라 할 때 (단, 성공집단 원소의 수를 N_1 실패집단 원소의 수는 N_2 일 때 $n \le N_1$, $n \le N_2$ 인 경우만 가정.) 확률변수 X가 다음과 같은 확률함수를 가질 때 이러한 분포를 **초기하분포**(Hypergeometric Distribution)이라 부른다.

$$\therefore \quad \Pr(X=x) = \frac{{}_{N_1}C_x \cdot {}_{N_2}C_{n-x}}{{}_{N}C_n}, \quad x = 0,1,2,\ldots,n.$$

Ⅳ. 초기하분포

초기하분포의 기댓값과 분산

$$E(X) = n \cdot \frac{N_1}{N} = np.$$

$$\begin{aligned} Var(X) &= n \cdot \frac{N_1}{N} \cdot \left(1 - \frac{N_1}{N}\right) \cdot \frac{N-n}{N-1} \\ &= n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \end{aligned}$$

Ⅳ. 초기하분포

초기하분포의 예제

예1) 어느 모임에 회원이 100명있다고 한다. 그 중에 남자가 60명, 여자가 40명이 있다고 하자. 전체회원에서 임의로 5명을 복원추출한다면 이 중 여자의 수를 확률변수 X라할 때 3명이 여자일 확률은?

<풀이>

X~B(5,0.4)의 분포이다.

$$\therefore\ {}_5C_3(0.4)^3(0.6)^2 = 0.2304$$

예2) 위 문제에서 전체회원에서 임의로 5명을 비복원추출한다면 이 중 여자의 수를 확률변수 X라 할 때 3명이 여자일 확률은?

<풀이>각 시행은 종속적 관계이므로 이항분포가 아닌 초기하분포가 된다. 따라서 모집단의 개수(N)=100명, 여자집단의 개수(N_1)=40명, 여자집단에 속하지 않는 개수(N_2)=60명, 반복시행횟수(n)=5명, 성공의 개수(x)=3인 초기하분포이다.

$$\therefore\quad \Pr(X=3) = \frac{{}_{40}C_3 \cdot {}_{100-40}C_{5-3}}{{}_{100}C_5} = 0.2323$$

Ⅳ. 초기하분포

초기하분포의 예제

예3) 어느 회사에서 생산하고 있는 제품 10개 중에 불량품이 2개가 있다고 한다. 이 중에서 3개를 비복원추출할 경우 1개가 불량품일 확률은?

<풀이>

$$\therefore \quad \Pr(X=1) = \frac{{}_{2}C_{1} \cdot {}_{10-2}C_{3-1}}{{}_{10}C_{3}} = 0.4667$$

예4) 주머니속에 노란공 6개와 파란공 4개가 들어 있다. 주머니에 들어 있는 10개의 공 가운데 비복원추출로 3개의 공을 꺼낼 경우, 이 중에서 2개의 공이 파란공일 확률은?

<풀이>

$$\therefore \quad \Pr(X=2) = \frac{{}_{4}C_{2} \cdot {}_{10-4}C_{3-2}}{{}_{10}C_{3}} = 0.3$$