3장 : 이산형 확률분포

3 이산형 확률분포

3.1. 베르누이 (Bernoulli) 확률변수와 이항분포 (Binomial Distribution)

- 결과값이 성공과 실패로 나누어지는 실험을 수행했다고 가정하자.

- 만약 결과값이 성공일 경우 X는 1이고 실패일 경우는 X는 0이라고 할때 pmf가 다음과 같은 확률변수를 베르누이 (Bernoulli) 확률변수라 한다.

$$p(0) = P\{X = 0\} = 1 - p, \qquad \text{and} \qquad p(1) = P\{X = 1\} = p,$$

이때 $p, 0 \le p \le 1$는 성공일 확률

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

- $Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1 - p).$

3 이산형 확률분포

3.1. 베르누이 (Bernoulli) 확률변수와 이항분포 (Binomial Distribution)

확률변수 : 시행 (n) 중 성공의 수 → 이항 확률변수(Binomial random variable)

- **만약 n 개의 독립적인 베르누이 시행을 $X_1, \ldots, X_n$ 이라 하고 각각의 X_i 가 1의 값을 가질 확률을 p 라 하면, $X = X_1 + \cdots + X_n$ 은 모수 n 과 p 를 갖는 이항분포 (Binomial distribution)을 가지며 $X \sim B(n, p)$ 로 나타내기로 한다.**

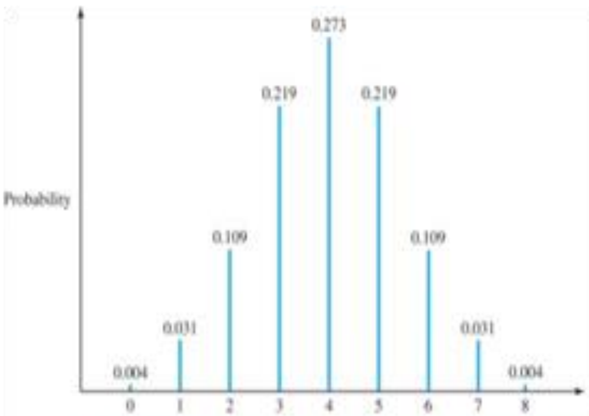
$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

- $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1.$

3 이산형 확률분포

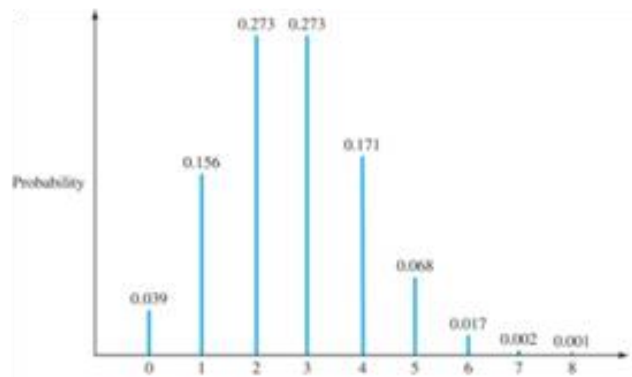
3.1. 베르누이 (Bernoulli) 확률변수와 이항분포 (Binomial Distribution)

Outcome of four Bernoulli trials	Probability	X = number of successes
0000	$(1-p)(1-p)(1-p)(1-p)$	0
0001	$(1-p)(1-p)(1-p)p$	1
0010	$(1-p)(1-p)p(1-p)$	1
0011	$(1-p)(1-p)pp$	2
0100	$(1-p)p(1-p)(1-p)$	1
0101	$(1-p)p(1-p)p$	2
0110	$(1-p)pp(1-p)$	2
0111	$(1-p)ppp$	3
1000	$p(1-p)(1-p)(1-p)$	1
1001	$p(1-p)(1-p)p$	2
1010	$p(1-p)p(1-p)$	2
1011	$p(1-p)pp$	3
1100	$pp(1-p)(1-p)$	2
1101	$pp(1-p)p$	3
1110	$ppp(1-p)$	3
1111	$pppp$	4



x	0	1	2	3	4	5	6	7	8
$P(X \le x)$	0.004	0.035	0.144	0.363	0.636	0.855	0.965	0.996	1.000

B(8, 0.5)



x	0	1	2	3	4	5	6	7	8
$P(X \le x)$	0.039	0.195	0.468	0.741	0.912	0.980	0.997	0.999	1.000

B(8, 1/3)

3 이산형 확률분포

3.1. 베르누이 (Bernoulli) 확률변수와 이항분포 (Binomial Distribution)

- 독립적인 베르누이 시행의 합으로 이항 확률변수를 고려해보자.
  $X = X_1 + \cdots + X_n$ 라면

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = np.$$

- $Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p).$

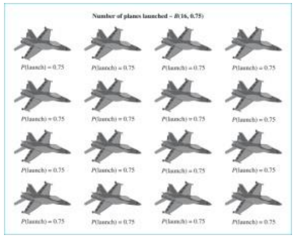
- **예제** : 네개의 동전을 던졌을 경우 각각의 동전의 결과는 서로 독립이라고 가정한다면 2개는 앞면 2개는 뒷면이 나올 확률은 얼마인가?

풀이) X 를 앞면이 나오는 회수라고 한다면, $X \sim Bin\left(4, \frac{1}{2}\right).$

$$Pr[X=2] = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{6}{16} = \frac{3}{8}.$$

3.1. 베르누이 (Bernoulli) 확률변수와 이항분포 (Binomial Distribution)

공군항공전투대대



- 한 공군 요격전대는 항상 비상출격 준비를 갖춘 16대의 항공기로 구성되어있다.
- 하지만 비행기의 엔진에 문제가 많아서, 고장이 발생하여 출격해야 하지만 출격을 못할 확률은 0.25이다.
- 전투대대 사령관은 얼마나 많은 비행기가 출격명령을 즉시 이륙할 수 있을지 관심을 가지고 있다.
- 오른쪽 그림에서처럼 성공적으로 출격할 수 있는 비행기의 수는 $n=16$, $p=0.75$ 인 이항분포를 따른다.
- 각각의 비행기는 성공확률 $p=0.75$ 인 베르누이 시행을 따른다고 볼 수 있다.
- 성공적으로 출격할 수 있는 비행기의 기대 대수 및 분산은

$$E(X)=np=16\times 0.75=12$$
$$Var(X)=np(1-p)=16\times 0.75\times 0.25=3$$

이 되며 표준편차 $\sigma=\sqrt{3}=1.73$ 대이다.

- 아래 그림에서 보듯이 정확히 12대의 비행기가 성공적으로 출격할 확률은

$$P(X=12)=\binom{16}{12}\times 0.75^{12}\times 0.25^{4}=\frac{16!}{12!4!}\times 0.75^{12}\times 0.25^{4}=0.225$$

이며,

최소 14대가 성공적으로 출격할 확률은 다음과 같다:

$$\begin{aligned} P(X\geq 14) &= P(X=14)+P(X=15)+P(X=16) \\ &= \binom{16}{14}\times 0.75^{14}\times 0.25^{2}+\binom{16}{15}\times 0.75^{15}\times 0.25^{1}+\binom{16}{16}\times 0.75^{16}\times 0.25^{0} \\ &= 0.134+0.054+0.010=0.198 \end{aligned}$$

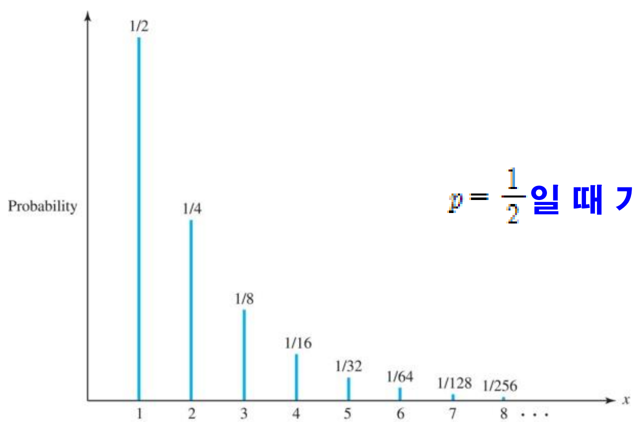
3 이산형 확률분포

3.2. 기하분포 (Geometric Distribution)

- 성공확률이 p일 때 성공이 발생할 때 까지 반복하였고 이때의 수행 회수를 X라 하자. 이 때 X는 모수 p 를 갖는 기하 확률변수라 한다.

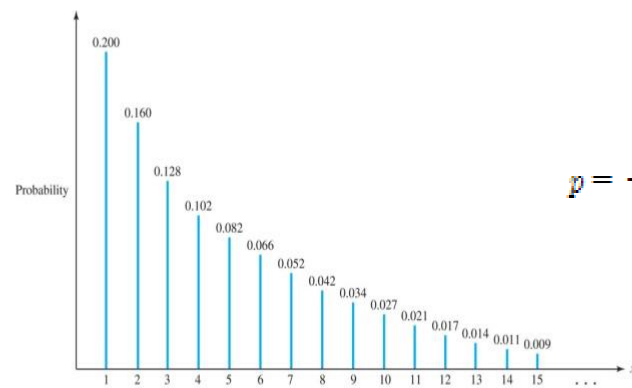
$$P(X=x)=(1-p)^{x-1}p\ ,\ x=0,1,\cdots$$

x번째 시행에서 p의 확률을 가지고 성공 !!
$(x-1)$ 까지는 실패 !!



$p=\dfrac{1}{2}$일 때 기하분포

x	1	2	3	4	5	6	7	8	···
$P(X \le x)$	1/2	3/4	7/8	15/16	31/32	63/64	127/128	255/256	···



$p=\dfrac{1}{5}$일 때 기하분포

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	···
$P(X \le x)$	0.200	0.360	0.488	0.590	0.672	0.738	0.790	0.832	0.866	0.893	0.914	0.931	0.945	0.956	0.965	···

3.2. 기하분포 (Geometric Distribution)

- $E(X) = 1/p, \quad Var(X) = (1-p)/p^2.$

$$\begin{aligned} P(X > x) &= (1-p)^x p + (1-p)^{x+1} p + \cdots \\ &= p(1-p)^x (1 + (1-p) + (1-p)^2 + \cdots) \\ &= p(1-p)^x \frac{1}{p} = (1-p)^x. \end{aligned}$$

$$P(X > n + x | X > n) = \frac{P(X>n, X>n+x)}{P(X>n)} = \frac{P(X>n+x)}{P(X>n)}$$

$$= (1-p)^x = P(X > x).$$

- **기하 분포는 무기억 특성을 갖는 유일한 이산형 분포**

3.2. 기하분포 (Geometric Distribution)

> **예제** : 어느 회사는 바쁜 시간에 전화량이 폭주한다. 만약 이 시간대에서 전화가 연결될 확률이 $p = 0.05$ 라 한다면, 전화가 연결되기 위해서 5번의 시도가 필요할 확률은 얼마인가?

풀이) 만약 X를 첫번째 성공을 위한 시행 횟수라 한다면, $X \sim \text{Geo}(p)$ 이고 $x = 5, p = 0.05,\ Pr(X = 5) = (0.95)^4(0.05) = 0.041.$ 이다.
X 의 기대값과 분산은 다음과 같다.

$$E[X] = 1/0.05 = 20, \qquad Var(X) = \frac{1-0.05}{0.05^2} = 380$$

3 이산형 확률분포

3.3. 음이항분포 (Negative Binomial Distribution)

- 성공확률이 p 인 베르누이 실험을 r 번째 성공이 발생할 때까지의 수행한 회수를 X 라고 하자. 이 때 X 는 모수가 p 와 r 인 음이항 분포를 가진다.

- 독립적인 기하분포 확률변수의 합으로 음이항 분포를 고려할 경우

$$P(X=x)=\binom{x-1}{r-1}(1-p)^{x-r}p^r, \qquad x=r,r+1,\ldots$$

- $E[X_i]=1/p$ 이고 $Var(X)=(1-p)/p^2$ 인 독립적인 기하분포의 합 $X=X_1+\cdots+X_r$, where X_i, $i=1,\ldots,r$ 일 경우

$$E(X)=E(X_1)+E(X_2)+\cdots+E(X_r)=np$$

$$Var(X)=Var(X_1)+\cdots+Var(X_r)=\frac{r(1-p)}{p^2}.$$

3.3. 음이항분포 (Negative Binomial Distribution)

예제 : NBA 결승전에서 총 7전중 4게임을 먼저 이기는 팀이 승리한다. 만약 A팀이 B팀을 이길 확률이 0.55이고 이 두 팀이 결승전에서 대결한다고 가정하자.

a) A팀이 6전을 치루고 우승할 확률은 얼마인가?

b) A팀이 우승할 확률은 얼마인가?

풀이-a) X **를 A팀이 우승하기 위해 필요한 총 게임수라고한다면**

$$X \sim \mathrm{NB}(r = 4, p = 0.55).$$

$$\mathrm{NB}(4, 0.55) = \begin{pmatrix} 6-1 \\ 4-1 \end{pmatrix} (1-0.55)^{6-4} 0.55^4 = 0.1853.$$

풀이-b) Pr[A팀이 우승]

$$= 0.55^4 + \begin{pmatrix} 5-1 \\ 4-1 \end{pmatrix} (0.45)^{5-4} 0.55^4 + \begin{pmatrix} 6-1 \\ 4-1 \end{pmatrix} (0.45)^{6-4} 0.55^4$$

$$+ \begin{pmatrix} 7-1 \\ 4-1 \end{pmatrix} (0.45)^{7-4} 0.55^4$$

$$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083.$$

3 이산형 확률분포

3.4. 초기하분포 (Hypergeometric Distribution)

- N 아이템 중 r 개가 관심대상인 결과를 가지는 경우를 고려해보자.
  예로서 생산된 N 제품 중 r 개가 불량품에 해당하는 경우를 가정할 수 있다

- 만약 그 중 한 개를 임의로 선택하여 관심대상의 결과를 가질 확률은
  $p = r / N$ 이 된다.

- 만약 n 개의 아이템을 복원(with replacement)방법으로 임의 추출하였을 때 그 중 불량품의 개수를 X 라고 하면

  $X \sim Bin\left(n, \frac{r}{N}\right)$ 이다.

- 하지만 n 개의 아이템을 비복원 (without replacement)으로 임의 추출할 경우, 적합한 분포는 초기하분포 (hypergeometric distribution)이다.

3 이산형 확률분포

3.4. 초기하분포 (Hypergeometric Distribution)

초기하분포의 확률질량함수는 다음과 같다.

$$P(X = x) = \frac{\begin{pmatrix} r \\ x \end{pmatrix} \times \begin{pmatrix} N-r \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}},$$

$$max\{0, n+r-N\} \leq x \leq \min\{n, r\}$$

$N - r - n + x \geq 0.$ **이므로**

$$E[X] = \frac{nr}{N}$$

$$Var(X) = \left(\frac{N-n}{N-1}\right) \times n \times \frac{r}{N} \times \left(1 - \frac{r}{N}\right)$$

3 이산형 확률분포

3.4. 초기하분포 (Hypergeometric Distribution)

> **예제** : 어느 제품은 7개 부품을 가지고 있고 이를 검사하였다. 이 중 4개는 정상이고 3개는 불량이었다. 총 7개 부품 중 3개의 표본을 추출하였다. X 를 이 표본에서의 불량 부품의 수라고 한다면 X 의 pmf는 다음과 같다.

$$P(X=x)=\frac{\begin{pmatrix}3\\x\end{pmatrix}\times\begin{pmatrix}4\\3-x\end{pmatrix}}{\begin{pmatrix}7\\3\end{pmatrix}},\qquad x=0,1,2,3.$$

- $p(0)=Pr(X=0)=\frac{4}{35},\qquad p(1)=Pr(X=1)=\frac{18}{35}$
- $p(2)=Pr(X=2)=\frac{12}{35},\qquad p(3)=Pr(X=3)=\frac{1}{35}$

Here, $N=7, n=3, r=3$, thus $\mu=E[X]=\frac{3\times 3}{7}$.

3 이산형 확률분포

3.5. 포아송분포 (Poisson Distribution)

- 특정한 기간 동안 발생하는 사건의 수를 확률변수로 정의하자. (예를 들어 어느 제품의 불량의 수, 일과 기간 중 받은 전화의 수)

- $0, 1, \ldots,$ 의 값을 갖는 이 확률 변수가 다음의 확률밀도 함수를 갖을 때 이를 모수 $\lambda > 0,$를 갖는 포아송 확률 변수라고 한다.

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

- 이것은 다음과 같이 표현된다. $X \sim Pois(\lambda).$

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

3.5. 포아송분포 (Poisson Distribution)

- 다음과 같은 이항분포를 생각해 보자. $n \to \infty, p \to 0, np \to \lambda > 0 (p = \lambda/n)$.
이 이항분포 $X \sim Bin(n,p)$, 는 다음과 같다.

$$
\begin{aligned}
Pr(X=x) &= \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n \\
&= \frac{n!}{(n-x)!x!} \cdot p^x (1-p)^{n-x} = \frac{n!}{(n-x)!x!} \cdot \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \cdot \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x} \\
&= \frac{e^{-\lambda}\lambda^x}{x!}
\end{aligned}
$$

3.5. 포아송분포 (Poisson Distribution)

Since, $n \to \infty, p \to 0,$

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}, \ \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x} \to 1, \left(1-\frac{\lambda}{n}\right)^x \to 1$$

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} \\ &= \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} \qquad \text{let } k = x-1 \\ &= \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda \end{aligned}$$

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) \cdot \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2 \end{aligned}$$

$$Var(X) = E[X(X-1)] + E[X] - (E[X])^2 = \lambda.$$

3.5. 포아송분포 (Poisson Distribution)

- **예제** : 어느 실험실 실험에서 1 millisecond 당 통과하는 방사성 물질의 평균 수는 4이다. 어느 주어진 1 millisecond에서 6개의 방사성 물질이 통과할 확률은 얼마인가?

$$Pr[X = 6] = \frac{e^{-4}4^6}{6!} = 0.1043.$$

3.6. 다항분포 (Multinomial Distribution)

- $p_1, \ldots, p_k$ 의 확률을 갖는 k 개의 결과값 $X_1, \ldots, X_k$ 을 갖는 n 개의 사건을 고려해 보자.
각각의 결과값의 빈도수를 나타내는 확률변수는 다항분포를 갖는다고 하고 이들의 결합밀도 함수는 다음과 같다.

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \times \cdots \times p_k^{x_k},$$

for nonnegative integer values of the x_i satisfying $x_1 + \cdots + x_k = n.$

3.6. 다항분포 (Multinomial Distribution)

- 예를 들어 고장을 전기적 고장, 기술적 고장, 오작동에 의한 고장으로 나눌수 있다고 가정하자. X_1, X_2, and X_3 를 각각의 원인에 의한 고장의 수이고 를 각각의 고장이 발생할 확률이 다음과 같다: p_1, p_2, and p_3,

  만약 10번의 고장이 발생했다면 $X_1 + X_2 + X_3 = 10,$

$$Pr[X_1 = x_1, X_2 = x_2, X_3 = x_3] = \frac{10!}{x_1!x_2!x_3!} p_1^{x_1} \times p_2^{x_2} \times p_3^{x_3}$$

  만약 $p_1 = 0.2, p_2 = 0.5, p_3 = 0.3,$ 이면

$$Pr[X_1 = 3, X_2 = 5, X_3 = 2] = \frac{10!}{3!5!2!}(0.2)^3 \times (0.5)^5 \times (0.3)^2 = 0.057$$

- $E(X_i) = np_i$, and $Var(X_i) = np_i(1 - p_i)$, 그러나 서로 독립이지 않음