4장. 이산확률분포

이산균등분포(discrete uniform distribution)

확률변수 X의 확률질량함수가

$$f_X(x) = \begin{cases} \frac{1}{n}, & x \in R_X = \{x_1, x_2, \cdots, x_n\} \\ 0, & \text{otherwise} \end{cases}$$

일 때, X는 이산균등분포를 따른다고 하고 $X \sim U_{\{x_1, x_2, \cdots, x_n\}}$로 나타낸다.

x	x_1	x_2	$\cdots$	x_n
$P[X=x]$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$

Theorem

$X \sim U_{\{x_1, x_2, \cdots, x_n\}}$ 이면,

$$(1) \quad E[X] = \frac{1}{n}\sum_{i=1}^{n} x_i := \mu_X$$

$$(2) \quad \mathrm{Var}[X] = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_X)^2$$

$$(3) \quad m_X(t) = \frac{1}{n}\sum_{i=1}^{n} e^{x_i t}$$

베르누이분포(Bernoulli distribution)

성공 확률이 $p(0 \leqslant p \leqslant 1)$인 시행에서 X를 성공 횟수라 하면, $R_X = \{0, 1\}$이고 X의 확률질량함수가

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x}, & x \in R_X = \{0, 1\} \\ 0, & \text{otherwise} \end{cases}$$

이다. 이 확률변수 X는 베르누이분포를 따른다고 하고, $X \sim \mathrm{Ber}(p)$로 나타낸다.

x	0	1
$P[X=x]$	$1-p$	p

베르누이 시행(Bernoulli trial)

실험의 결과가 2 가지('성공'과 '실패', 또는 '찬성'과 '반대')로 주어지는 확률실험을 베르누이 시행이라 한다. 또 베르누이 시행에서, 확률변수 X를 성공이면 1, 실패이면 0으로 정의하고 또 성공할 확률이 p이면, X는 모수 p인 베르누이 확률변수가 된다. 즉, $X \sim \text{Ber}(p)$

Theorem

$X \sim \text{Ber}(p)$이면,

(1) $E[X] = p$

(2) $\text{Var}[X] = p(1-p)$

(3) $m_X(t) = pe^t + (1-p), \quad t \in \mathbb{R}$

이항분포(Binomial distribution)

성공 확률이 p인 베르누이 시행을 독립적으로 n 번 반복할 때 X를 성공 횟수라 하면, $R_X = \{0, 1, 2, \cdots, n\}$이고 X의 확률질량함수가

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in R_X = \{0, 1, 2, \cdots, n\} \\ 0, & \text{otherwise} \end{cases}$$

이다. 이 확률변수 X는 이항분포를 따른다고 하고, $X \sim B(n,p)$로 나타낸다.

Remark

(1)

$$\sum_{x \in \mathbb{R}} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \Big(p + (1-p)\Big)^n = 1^n = 1$$

(2) $n = 1$인 이항분포는 베르누이분포이다. 즉, $B(1,p) = \text{Ber}(p)$

Theorem

$X \sim B(n,p)$이면,

(1) $E[X] = np$

(2) $\text{Var}[X] = np(1-p)$

(3) $m_X(t) = \Big(pe^t + (1-p)\Big)^n, \quad t \in \mathbb{R}$

$$\begin{aligned} E[X] &= \sum_{x \in \mathbb{R}} x \cdot f_X(x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} x \cdot \frac{n!}{x! \cdot (n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} n \cdot \frac{(n-1)!}{(x-1)! \cdot (n-x)!} p^x (1-p)^{n-x} \\ &= \sum_{x=1}^{n} np \cdot \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\ &= np \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \\ &= np \cdot \Big(p + (1-p)\Big)^{n-1} = np \end{aligned}$$

Example

$X \sim \mathrm{Ber}(p)$, $Y \sim \mathrm{Ber}(p)$이고 X와 Y가 독립일 때, 확률변수 $Z = X + Y$의 확률질량함수 $f_Z(z)$를 구하시오.

Solution. $X \sim \mathrm{Ber}(p)$, $Y \sim \mathrm{Ber}(p)$ 일 때, X와 Y의 확률질량함수:

$$f_X(x) = \begin{cases} p^x(1-p)^{1-x}, & x \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} p^y(1-p)^{1-y}, & y \in \{0,1\} \\ 0, & \text{otherwise} \end{cases}$$

확률벡터 (X, Y)의 결합확률질량함수를 $f_{X,Y}(x,y)$:(X, Y 독립)

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$
$$= \begin{cases} p^{x+y}(1-p)^{2-x-y}, & (x,y) \in \{(0,0),(0,1),(1,0),(1,1)\} \\ 0, & \text{otherwise} \end{cases}$$

$Z = X + Y$의 확률질량함수 $f_Z(z)$ (치공간 $R_Z = \{0,1,2\}$):

$$f_Z(z) = P[Z = z] = P[X + Y = z] = \sum_{x+y=z} f_{X,Y}(x,y)$$

$$= \begin{cases} \sum_{x+y=0} p^{x+y}(1-p)^{2-x-y}, & z = 0 \quad ; \quad (x,y) = (0,0) \\ \sum_{x+y=1} p^{x+y}(1-p)^{2-x-y}, & z = 1 \quad ; \quad (x,y) = (1,0),(0,1) \\ \sum_{x+y=2} p^{x+y}(1-p)^{2-x-y}, & z = 2 \quad ; \quad (x,y) = (1,1) \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} (1-p)^2, & z = 0 \\ p(1-p) + p(1-p) = 2p(1-p), & z = 1 \\ p^2, & z = 2 \\ 0, & \text{otherwise} \end{cases}$$

$$= \binom{2}{z} p^z (1-p)^{2-z} \cdot I_{\{0,1,2\}}$$

따라서 $Z \sim B(2, p)$이다.

기하분포(Geometric distribution)

성공 확률이 $p(0 < p < 1)$인 베르누이 시행을 독립적으로 계속 반복할 때, X를 첫 번째 성공이 일어날 때까지의 시행 횟수라 하면, $R_X = \{1, 2, 3, \cdots\}$이고 X의 확률질량함수가

$$f_X(x) = \begin{cases} (1-p)^{x-1} \cdot p, & x \in R_X = \{1, 2, 3, \cdots\} \\ 0, & \text{otherwise} \end{cases}$$

이다. 이 확률변수 X는 기하분포를 따른다고 하고, $X \sim \mathrm{Ge}(p)$로 나타낸다.

Remark

$$\sum_{x \in \mathbb{R}} f_X(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} \cdot p = \frac{p}{1-(1-p)} = 1$$

Theorem

$0 < p < 1$이고 $X \sim \mathrm{Ge}(p)$이면,

(1) $E[X] = \dfrac{1}{p}$, $\mathrm{Var}[X] = \dfrac{1-p}{p^2}$

(2) $m_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$, $\quad t < -\ln(1-p)$

Proof.

$$m_X(t) = E\left[e^{tX}\right] = \sum_{x \in \mathbb{R}} e^{tx} f_X(x) = \sum_{x=1}^{\infty} e^{tx} \cdot (1-p)^{x-1} \cdot p$$
$$= \frac{p}{1-p} \sum_{x=1}^{\infty} \left((1-p) \cdot e^t\right)^x = \frac{p}{1-p} \cdot \frac{(1-p)e^t}{1-(1-p)e^t}$$
$$= \frac{pe^t}{1-(1-p)e^t}, \qquad (\text{단}, \ (1-p)\, e^t < 1)$$

Theorem(무기억성)

$0 < p < 1$이고 $X \sim \text{Ge}(p)$이면, 모든 $a,\ b \in \mathbb{N}$에 대해

$$P\left[X > a + b \,\middle|\, X > a\right] = P[X > b]$$

Proof.

$$P[X > b] = \sum_{x=b+1}^{\infty} p \cdot (1-p)^{x-1} = \frac{p(1-p)^b}{1-(1-p)} = (1-p)^b$$

$$\begin{aligned} P\left[X > a + b \,\middle|\, X > a\right] &= \frac{P[X > a+b]}{P[X > a]} = \frac{(1-p)^{a+b}}{(1-p)^a} \\ &= (1-p)^b \end{aligned}$$

Example

1. 어떤 사람이 운전면허시험에 합격할 확률은 0.7이라고 한다. 이 사람이 네 번 이전에 시험에 합격할 확률을 구하시오.

2. 불량률이 5%인 제품의 더미에서 일일이 검사하여 불량품을 선별해 낸다. 20 개의 검사가 끝났을 때까지 불량품은 발견되지 않았다고 한다. 첫 불량품이 발견될 때까지 앞으로 적어도 5 개의 상품을 더 관찰해야 할 확률을 구하시오.

이항급수

(1) $x, r \in \mathbb{R}, |x| < 1$ 일 때, $\quad (1+x)^r = \displaystyle\sum_{n=0}^{\infty} \binom{r}{n} x^n$

(2) $|x| < 1$인 실수 x와 자연수 r에 대해,

$$\begin{aligned} (1+x)^{-r} &= 1 + (-r)x + \frac{(-r)(-r-1)}{2!}x^2 + \frac{(-r)(-r-1)(-r-2)}{3!}x^3 + \cdots \\ &= 1 - rx + \frac{r(r+1)}{2!}x^2 - \frac{r(r+1)(r+2)}{3!}x^3 + \cdots \end{aligned}$$

$$\begin{aligned} (1-x)^{-r} &= 1 + rx + \frac{r(r+1)}{2!}x^2 + \frac{r(r+1)(r+2)}{3!}x^3 + \cdots \\ &= \binom{r-1}{0} + \binom{r}{1}x + \binom{r+1}{2}x^2 + \binom{r+2}{3}x^3 + \cdots \\ &= \binom{r-1}{r-1} + \binom{r}{r-1}x + \binom{r+1}{r-1}x^2 + \binom{r+2}{r-1}x^3 + \cdots \\ &= \sum_{n=r}^{\infty} \binom{n-1}{r-1} x^{n-r} \end{aligned}$$

음이항분포(Negative Binomial Distribution)

성공 확률이 p인 베르누이 시행을 독립적으로 계속 반복할 때, X를 r 번째 성공이 일어날 때까지의 시행 횟수라 하면, $R_X = \{r,\ r+1\ \cdots\}$이고 X의 확률질량함수가

$$f_X(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r}, & x \in R_X = \{r,\ r+1\ \cdots\} \\ 0, & \text{otherwise} \end{cases}$$

이다. 이 확률변수 X는 음이항분포를 따른다고 하고, $X \sim \text{NB}(r, p)$로 나타낸다.

Remark

X를 $r = 3$ 번째 성공이 일어날 때까지의 시행 횟수:

$x = 3:\quad SSS$

$x = 4:\quad SSFS,\ SFSS,\ FSSS$

$x = 5:\quad SSFFS,\ SFSFS,\ SFFSS,\ FSSFS,\ FSFSS,\ FFSSS$

Theorem

$0 < p < 1$이고 $X \sim \mathrm{NB}(r,p)$이면,

(1) $E[X] = \frac{r}{p}$, $\mathrm{Var}[X] = \frac{r(1-p)}{p^2}$

(2) $m_X(t) = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r$, $\quad t < \ln(\frac{1}{1-p})$

Proof.

$$m_X(t) = E\left[e^{tX}\right] = \sum_{x\in\mathbb{R}} e^{tx} f_X(x) = \sum_{x=r}^{\infty} e^{tx} \cdot \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

$$= \sum_{x=r}^{\infty} \binom{x-1}{r-1} \left(pe^t\right)^r \left((1-p)e^t\right)^{x-r}$$

$$= \frac{(pe^t)^r}{\left(1-(1-p)e^t\right)^r} \sum_{x=r}^{\infty} \binom{x-1}{r-1} \left(1-(1-p)e^t\right)^r \left(1-(1-(1-p)e^t)\right)^{x-r}$$

$$= \left(\frac{pe^t}{1-(1-p)e^t}\right)^r \qquad \left(\text{단, } (1-p)e^t < 1, \ \text{ 즉, } \ t < \ln(\frac{1}{1-p})\right)$$

Example

$p = 0.4$

눈 수술을 받은 환자의 40%가 후유증이 없다고 한다. 이 눈수술을 7 명의 환자까지 시행하여 비로소 4 명의 환자가 후유증이 없을 확률을 구하시오.

$r=4 \qquad X \sim NB(4, 0.4)$

$P[X=7] \qquad {}_6C_3 (0.4)^4 (0.6)^3$

Theorem

$i = 1, 2, \cdots, r$에 대하여 $X_i \sim \mathrm{Ge}(p)$이고 $X_1, X_2, \cdots, X_r$이 서로 독립이면, 확률변수 $Z = X_1 + X_2 + \cdots + X_r$은 음이항분포 $\mathrm{NB}(r,p)$를 따른다.

여러가지 이산분포의 비교

- 표본공간 $\Omega = \{S(\text{성공}), F(\text{실패})\}$
- $P(\{S\}) = p$
- X : 확률변수

	X : 성공 횟수	X : 성공할 때 까지의 시행 횟수	
시행 횟수가 1번	베르누이분포 $\mathrm{Ber}(p)$	기하분포 $\mathrm{Ge}(p)$	마지막 시행에서 1 번째 성공
시행 횟수가 n번	이항분포 $B(n,p)$	음이항분포 $\mathrm{NB}(r,p)$	마지막 시행에서 r 번째 성공

푸아송분포의 도입

어떤 현상에서 시간 $(0,t)$ 내에 발생하는 사건들의 수를 확률변수 $X_{(0,t)}$라 하자. 그러면 $X_{(0,t)}$는 치공간이 $R_{X_{(0,t)}} = \{0,1,2,3,\cdots\}$인 이산확률변수이다.

확률변수 $X_{(0,t)}$의 확률질량함수 $f_{X_{(0,t)}}(x)$를 생각하자.

- 푸아송 가정 : 확률현상에 관한 가정

(A1) $P[X_{(t,t+\Delta t)} = 1] = \lambda\Delta t + o(\Delta t)$

$o(\Delta t)$는 $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$을 만족하는 양

(A2) $P[X_{(t,t+\Delta t)} \geqslant 2] = o(\Delta t)$
충분히 작은 시간 Δt 내에 두 사건 이상이 일어날 확률은 0

(A3) $P[X_{(s,t)} = k \mid X_{(u,v)} = h] = P[X_{(s,t)} = k]$ $(u \leqslant v \leqslant s \leqslant t)$
즉, 중복되지 않는 관측시간 내에서 일어나는 사건들의 수는 서로 독립이다.

푸아송분포(Poisson distribution)

확률변수 X의 확률질량함수가

$$f_X(x) = \begin{cases} \dfrac{m^x e^{-m}}{x!}, & x \in R_X = \{0,\, 1,\, 2,\, \cdots\} \\ 0, & \text{otherwise} \end{cases}$$

이면, X는 푸아송분포를 따른다고 하고 $X \sim \mathrm{Pois}(m)$로 나타낸다.

Remark

$$\cancel{\sum_{x=0}^{\infty}} f_X(x) = \sum_{x=0}^{\infty} \frac{m^x e^{-m}}{x!} = e^{-m} \sum_{x=0}^{\infty} \frac{m^x}{x!} = e^{-m} \cdot e^{m} = 1$$

$$= \sum_{x\in\mathbb{R}} f_X(x)$$

Theorem

$X \sim \mathrm{Pois}(m)$이면,

(1) $E[X] = m$

(2) $\mathrm{Var}[X] = m$

(3) $m_X(t) = e^{m(e^t-1)}$

Proof.

$$m_X(t) = E\left[e^{tX}\right] = \sum_{x\in\mathbb{R}} e^{tx} \cdot f_X(x) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{m^x e^{-m}}{x!}$$

$$= e^{-m} \sum_{x=0}^{\infty} \frac{(me^t)^x}{x!} = e^{-m} \cdot e^{me^t} = e^{m(e^t-1)}$$

Example

$m=3$

$X \sim \mathrm{Pois}(3)$

$(0,4)$

0 4

1. 전화교환대에 전화벨 소리가 매 4 분마다 평균 3건의 비율로 울린다고 한다. 만약 전화벨 소리의 수가 푸아송분포를 따른다면 8분 동안 5건 이상의 전화벨 소리가 울릴 확률을 구하시오.

   0 8

   $m = 6$

   $X \sim \mathrm{pois}(6)$

   $(0,8)$

   $$P[X \geq 5] = \sum_{x=5}^{\infty} \frac{6^x e^{-6}}{x!} = 1 - \sum_{x=0}^{4} \frac{6^x e^{-6}}{x!}$$

2. 어떤 회사에서 근무시간에 발신되는 e-메일의 수는 평균적으로 시간당 30건이라 한다.
   (1) 이 회사에서 임의로 선택된 3 분 동안에 e-메일의 발신이 없을 확률을 구하시오.
   (2) 이 회사에서 임의로 선택된 6 분 동안에 e-메일의 발신이 5건 이상일 확률을 구하시오.

초기하분포(Hypergeometric Distribution)

크기 N인 유한 모집단에 속성 A를 가진 집단의 원소 개수를 D, 속성 A를 갖지 않는 집단의 원소 개수를 $N-D$라 하자. 이런 모집단에서 n개의 표본을 임의로 비복원추출할 때, X를 추출된 표본에서 속성 A를 가진 집단의 원소 개수라 하면, $R_X = \{0,\, 1,\, 2,\, \cdots,\, \min\{n, D\}\}$이고 X의 확률질량함수가

$$f_X(x) = \begin{cases} \dfrac{\binom{D}{x} \cdot \binom{N-D}{n-x}}{\binom{N}{n}}, & x \in R_X = \{0,\, 1,\, 2,\, \cdots,\, \min\{n, D\}\} \\ 0, & \text{otherwise} \end{cases}$$

이다. 이 확률변수 X는 초기하분포를 따른다고 하고, $X \sim \mathrm{HG}(N; n, D)$ 나타낸다.