4장 이산확률분포 : 확률모형

§4.1. 이산균등분포

Example

(1) 동전을 한 번 던지는 확률실험에서 표본공간은

$$\Omega=\{앞면(H),뒷면(T)\}$$

이다. 앞면의 수를 확률변수 X라 하면, X의 치역공간은 $R_X=\{0,1\}$이고 확률질량함수는 다음과 같다.

$$f_X(x)=\begin{cases}\frac{1}{2} & , x=0,1\\ 0 & , \text{otherwise}\end{cases}$$

(2) 주사위를 한 번 던지는 확률실험에서 표본공간은

$$\Omega=\{⚀,⚁,⚂,⚃,⚄,⚅\}$$

이다. 주사위 윗면에 나온 눈의 수를 확률변수 X라 하면, X의 치역공간은 $R_X=\{1,2,3,4,5,6\}$이고 확률질량함수는 다음과 같다.

$$f_X(x)=\begin{cases}\frac{1}{6} & , x=1,2,3,4,5,6\\ 0 & , \text{otherwise}\end{cases}$$

Definition

이산확률변수 X의 치역공간이 $R_X=\{x_1,x_2,\cdots,x_n\}$이고 확률질량함수가

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

이면, X는 **이산균등분포**(discrete uniform distribution)를 따른다고 하고, 기호로 $X\sim DU(n)$으로 나타낸다.

Theorem

확률변수 X의 치역공간이 $R_X=\{x_1,x_2,\cdots,x_n\}$이고, $X\sim DU(n)$이면, 다음이 성립한다.

(1) $E[X]=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i=\mu_X$

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

(3) $m_X(t)=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}e^{tx_i}$

(proof)

Example

(1) 동전을 한 번 던지는 확률실험에서 앞면의 수를 X라 할 때, X의 평균과 분산을 구하여라.

(2) 주사위를 한 번 던지는 확률실험에서 주사위 윗면에 나온 눈의 수를 X라 할 때, X의 평균과 분산을 구하여라.

Corollary

확률변수 X의 치역공간이 $R_X=\{1,2,\cdots,n\}$이고, X가 이산균등분포를 따르면,

$$E[X]=\frac{n+1}{2},\quad Var[X]=\frac{n^2-1}{12}$$

이다.

(proof)

Example 1

1에서 10까지 숫자를 적은 동일한 모양의 카드가 들어 있는 주머니에서 임의로 하나 꺼낸 카드의 번호를 X라 할 때, 다음을 구하여라.

(1) X의 확률질량함수

(2) X의 평균과 분산

(3) $Y=2X-1$의 평균과 분산을 구하여라.

§4.3. 이항분포

▣ 베르누이 분포

Definition

확률실험의 결과가 '성공'과 '실패'와 같이 상반되는 두 가지로 나오는 경우, 즉 표본공간이

$$\Omega=\{성공(s),실패(f)\}$$

이면, 이와 같은 실험을 **베르누이 시행**(Bernoulli trial)이라 한다. 베르누이 시행에서 성공률이 $p(0\le p\le 1)$이면, 실패율은 $q=1-p$이다.
만약 확률변수 X를 성공이면 1, 실패이면 0으로 정의하면, X의 치역공간은 $R_X=\{0,1\}$이고 확률질량함수는

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

이다.
이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수 p인 **베르누이 분포**(Bernoulli distribution)라 하고, 기호로 $X\sim B(1,p)$로 나타낸다.

Theorem

$X\sim B(1,p)$이면, 다음이 성립한다.

(1) $E[X]=p$

(2) $Var[X]=p(1-p)=pq \quad$ (단, $q=1-p$)

(3) $m_X(t)=pe^t+q$

(proof)

Example 1

앞면이 나올 가능성이 $\frac{1}{3}$인 찌그러진 동전을 던지는 게임에서, 앞면이 나오면 성공한다고 하자. 이때 앞면이 나오는 사건에 대한 확률질량함수와 평균 그리고 분산을 구하여라.

▣ 이항 분포

Definition

성공할 확률이 p인 베르누이 시행을 독립적으로 n번 반복하는 확률실험에서 성공 횟수를 X라 하면, X의 치역공간은

$$R_X=\{0,1,2,\cdots,n\}$$

이고, X의 확률질량함수는

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

이다.
이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수가 n과 p인 **이항분포**(binormal distribution)라 하고, 기호로 $X\sim B(n,p)$로 나타낸다.

Theorem

$X\sim B(n,p)$이면, 다음이 성립한다.

(1) $E[X]=np$

(2) $Var[X]=npq \quad$ (단, $q=1-p$)

(3) $m_X(t)=(pe^t+q)^n$

(proof)

Example 2

앞면이 나올 가능성이 $\frac{1}{3}$인 찌그러진 동전을 세 번 던져서 앞면이 나온 횟수를 X라 할 때, X의 확률질량함수와 $E[X]$와 $Var[X]$를 구하여라.

Example 3

1부터 4의 숫자가 적힌 사면체를 5번 던질 때, 숫자 1이 나온 횟수를 X라 한다.

(1) X의 확률질량함수를 구하여라.

(2) 앞면이 한 번 나올 확률을 구하여라.

(3) 앞면이 많아야 한 번 나올 확률을 구하여라.

(4) 앞면이 적어도 두 번 이상 나올 확률을 구하여라.

(5) X의 평균과 분산을 구하여라.

Example 4

5지 선다형으로 주어진 10문제에서 임의로 답안을 선정한다. 다음 이항누적분포표를 이용하여 다음을 구하여라.

(1) 정답을 선택한 문항수가 2개일 확률

(2) 적어도 5문제 이상 정답을 선택할 확률

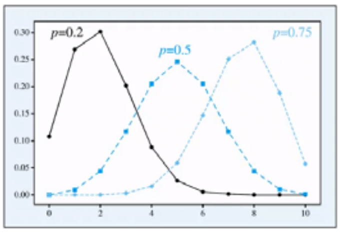
		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0	0.5987	0.3487	0.1969	0.1074	0.0562	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230

Remark

(1) $p < 0.5$이면, 이항분포는 왼쪽으로 치우치고 오른쪽으로 긴꼬리 모양을 가진다.

(2) $p > 0.5$이면, 이항분포는 오른쪽으로 치우치고 왼쪽으로 긴꼬리 모양을 가진다.

(3) $p = 0.5$이면, 이항분포는 n에 관계없이 $m_X = \frac{n}{2}$을 중심으로 좌우대칭이고, 이 경우 **대칭이항분포**(symmetric binomial distribution)라 한다.



p에 따른 이항분포의 비교

Theorem

(1) $X_1, X_2, \cdots, X_n$이 독립이고 모두 같은 베르누이 분포 $B(1,p)$를 따르는 확률변수이면, $X = X_1 + X_2 + \cdots + X_n$는 $B(n,p)$를 따른다. 이때,

$$\mu_X = E[X_1 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i] = np$$

$$\sigma_X{}^2 = Var[X_1 + \cdots + X_n] = \sum_{i=1}^{n} Var[X_i] = np(1-p)$$

이다.

(2) 두 확률변수 X와 Y가 독립이고 각각

$$X \sim B(n,p), \quad Y \sim B(m,p)$$

인 이항분포를 이룬다면, $Z = X + Y$일 때

$$P(Z = k) = \binom{m+n}{k} p^k (1-p)^{(m+n)-k}$$

이다. 즉, $Z \sim B(m+n, p)$이다.

$(proof)$

Example 5

임의로 선정된 남학생과 여학생이 5지 선다형으로 주어진 두 종류의 문제에서 임의로 답안을 선정한다. 남학생은 A형 문제지로 10문제를 풀고 여학생은 B형 문제지로 5문제를 푼다. 남학생이 정답을 선정한 문제 수를 X, 여학생이 정답을 선정한 문제 수를 Y라 할 때, 다음을 구하여라.

(1) 15문제에서 두 학생이 정답을 선택한 평균 문제 수

(2) 15문제에서 두 학생이 적어도 5문제 이상 정답을 선택할 확률

Theorem

$X \sim B(n,p)$일 때, 확률변수 $Y = \frac{X}{n}$는 성공의 비율을 나타내며, Y를 **표본비율**(sample proportion)이라 하고,

$$E[Y] = p, \quad Var[Y] = \frac{pq}{n}$$

이다.

$(proof)$

§4.4. 기하분포와 음이항분포

▣ 기하 분포

Theorem

성공할 확률이 p인 베르누이 시행을 독립적으로 계속 반복하는 확률실험에서 확률변수 X를

X='첫 번째 성공이 일어날 때까지의 시행 횟수'

로 정의하면, X의 치역공간은 $R_X=\{1,2,\cdots,n,\cdots\}$이고 X의 확률질량함수는

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

이다.

이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수가 p인 **기하분포**(geometric distribution)라 하고, 기호로 $X\sim G(p)$로 나타낸다.

참고

시행횟수	사상	확률
$X=1$	p	p
$X=2$	q p	$(1-p)p$
$X=3$	q q p	$(1-p)^2p$
$X=4$	q q q p	$(1-p)^3p$
⋮	⋮	⋮
$X=x$	q q q q … q p	$(1-p)^{x-1}p$
⋮	⋮	⋮

Theorem

$X\sim G(p)$이면, 다음이 성립한다.

(1) $E[X]=\dfrac{1}{p}$

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

(3) $m_X(t)=\dfrac{pe^t}{1-qe^t}$

(*proof*)

Theorem 무기억성 성질(memorylessness property)

$X\sim G(p)$이면, 임의의 양의 정수 a,b에 대하여

$$P(X>a+b\,|\,X>a)=P(X>b)$$

이다.

(*proof*)

Example 1

1 또는 2의 눈이 처음 나올 때까지 주사위를 던지는 게임을 할 때, 주사위를 던진 횟수를 확률변수 X라 하자.

(1) X의 확률질량함수와 평균 $E[X]$, 분산 $Var[X]$을 구하여라.

(2) 주사위를 3 번째 던졌을 때, 처음으로 1 또는 2의 눈이 나올 확률을 구하여라.

(3) 5번 이내에 1 또는 2의 눈이 나올 확률을 구하여라.

(4) 주사위를 5번 던졌을 때까지 1 또는 2의 눈이 나오지 않았을 때, 처음으로 1 또는 2의 눈이 나올 때까지 적어도 3번 더 주사위를 던질 확률을 구하여라.

(5) 5번째에 처음으로 1 또는 2의 눈이 나왔다는 조건 아래서, 그 이후로 다시 3번째에서 처음으로 1 또는 2의 눈이 나올 확률을 구하여라.

Example

불량률이 5%인 제품의 더미에서 일일이 제품을 검사하여 불량품을 선별해낸다. 20개의 검사가 끝날 때까지 불량품은 발견되지 않았을 때, 첫 불량품이 발견될 때까지 앞으로 적어도 5개의 상품을 더 관찰해야할 확률은?

▣ 음이항 분포

Definition

성공할 확률이 p인 베르누이 시행을 독립적으로 계속 반복하는 확률실험에서 확률변수 X를

X=‘r번째 성공이 일어날 때까지의 시행 횟수’

로 정의하면, X의 치역공간은 $R_X=\{r, r+1, \cdots\}$이고 X의 확률질량함수는

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

이다.

이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수가 r, p인 **음이항분포**(negative binomial distribution)라 하고, 기호로 $X\sim NB(r,p)$로 나타낸다.

만약 $r=1$이면, $NB(1,p)=G(p)$이다.

참고 $x-1$번 중에 $r-1$번 성공 $\quad$ x번째 성공 $\quad$ 확률

ⓠⓠ…ⓠⓟⓠⓠ…ⓠⓟⓠⓠ…ⓠ | ⓟ $\quad \binom{x-1}{r-1}q^{x-r}p^r$

참고 **이항급수**(binomial series)

모든 실수 $r\in\mathbb{R}$에 대하여, $|x|<1$일 때 $(1+x)^r=\sum_{n=0}^{\infty}\binom{r}{n}x^n$이다.

① $(1+x)^{-r}=1+(-r)x+\frac{(-r)(-r-1)}{2!}x^2+\frac{(-r)(-r-1)(-r-2)}{3!}x^3+\cdots$

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

② $(1-x)^{-r}=1-r(-x)+\frac{r(r+1)}{2!}(-x)^2+\frac{r(r+1)(r+2)}{3!}(-x)^3+\cdots$

$=1+rx+\frac{r(r+1)}{2!}x^2+\frac{r(r+1)(r+2)}{3!}x^3+\cdots$

$=\binom{r-1}{0}+\binom{r}{1}x+\binom{r+1}{2}x^2+\binom{r+2}{3}x^3+\cdots$

$=\binom{r-1}{r-1}+\binom{r}{r-1}x+\binom{r+1}{r-1}x^2+\binom{r+2}{r-1}x^3+\cdots=\sum_{n=r}^{\infty}\binom{n-1}{r-1}x^{n-r}$

Theorem

$X\sim NB(r,p)$이면, 다음이 성립한다.

(1) $E[X]=\frac{r}{p}$

(2) $Var[X]=\frac{rq}{p^2}$ (단, $q=1-p$)

(3) $m_X(t)=\left(\frac{pe^t}{1-qe^t}\right)^r$

(proof)

$|\leftarrow X_1 \rightarrow||\leftarrow X_2 \rightarrow|\ |\leftarrow \cdots \ \rightarrow||\leftarrow X_{r-1} \rightarrow|\ |\leftarrow X_r \rightarrow|$

ⓠⓠ…ⓠⓟ ⓠⓠ…ⓠⓟ ⓠⓠ…ⓠⓟ ⓠⓠ…ⓠⓠⓟ ⓠⓠ…ⓠⓟ

Example 2

주사위를 던지는 게임에서 1 또는 2의 눈이 나오면 성공이라 한다. 세 번째 성공이 있기까지 주사위를 반복하여 던진 횟수를 확률변수 X라 할 때, 다음을 구하여라.

(1) X의 확률질량함수

(2) X의 평균과 분산

(3) 5번째 시행에서 세 번째 성공이 일어날 확률

Example

A, B 두 사람이 주사위 한 개를 던져서 주사위 눈의 개수가 짝수이면 A가 이기고, 홀수이면 B가 이긴다고 한다.

이 게임을 독립적으로 반복하여 둘 중 한사람이 먼저 5게임을 이길 때까지 계속한다.

(1) 게임이 7회에 끝날 확률

(2) 게임이 7회에 끝났을 때, A가 승리할 확률

§4.5. 푸아송분포

Definition

임의의 양수 m에 대하여 확률변수 X의 확률질량함수가

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

이면, X는 모수가 m인 **푸아송분포**(Poisson distribution)라 하고, 기호로 $X\sim Poisson(m)$으로 나타낸다.

참고 (1) 이항분포에서 p가 매우 작고 n이 크면 $(p\to 0, n\to\infty)$, 푸아송분포에 근사시켜 계산한다.
실제 많은 문제에서 $np\le 5$, $n\ge 20$인 경우에 적용된다.
$(\because)$

(2) $\sum_{x\in R_X} f_X(x)$

Theorem

$X\sim Poi(\mu)$이면, 다음이 성립한다.

(1) $E[X]=\mu$

(2) $Var[X]=\mu$

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

(proof)

Example 1

평균 $m=1.5$인 푸아송 확률변수 X에 대하여 다음을 구하여라.

(1) $P(X=1)$

(2) $P(X\le 2)$

(3) $P(X\ge 3)$

Example

세 쌍둥이가 탄생할 확률일 0.0001이라 할 때, 10000명의 산모 중에서 적어도 4명 이상이 세 쌍둥이를 낳았을 확률을 구하여라.

Example 2

어떤 공정라인에서 생산된 제품의 불량률은 0.05이다. 이 공정라인에서 생산된 제품 20개를 임의로 선정했을 때, 다음을 구하여라.

(1) 불량품이 하나일 확률

(2) 푸아송분포에 의한 (1)의 근사확률

Remark

어떤 단위 시간이나 길이나 또는 공간을 나타내는 단위 구간 안에서 특정한 사건이 일어나는 비율이 λ라면, 길이가 t인 구간 동안에 그 사건이 일어나는 평균적인 횟수는 $m=\lambda t$라고 할 수 있다.
길이가 t인 구간 동안 일어나는 특정한 사건의 횟수 $X(t)$ (또는 $X_{(a,a+t)}$)는 모수 $m=\lambda t$인 푸아송분포에 따른다.

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

참고 예 일정 시간 동안 식당에 찾아오는 손님의 수
백과사전 한 페이지에 나타나는 오타의 수
단위 길이의 전선줄에 나타나는 흠의 수
단위 시간당 어떤 물질에 의하여 방출된 방사능 입자의 수

푸아송가정

① **비집락성** : $P(X_{(t,t+\Delta t)}\geq 2)=0$
충분히 작은 시간 Δt에서 사건은 최대 1건만 발생한다.

② **비례성** : $P(X_{(a,a+\Delta t)}=1)=\lambda\Delta t$
충분히 작은 시간에서 한 사건이 발생할 가능성은 존재하고, 그 크기는 시간 Δt에 비례한다.

③ **독립성** : $a\leq b\leq c\leq d$일 때,
$$P(X_{(a,b)}=k\,|\,X_{(c,d)}=h)=P(X_{(a,b)}=k)$$
중복되지 않는 시간에서 일어나는 사건은 서로 독립이다
동일한 크기의 시간에서 관찰된 횟수는 동일한 분포를 따른다.

$(proof)$

Example

오전 10시부터 11시까지 평균 3명의 손님이 찾아오는 상점에서 오늘 오전 10시부터 10시30분 사이에 적어도 2명의 손님이 찾아올 확률을 구하여라.

Example

실험실에서 0.001초 동안 카운터를 통과하는 방사능 입자의 평균수는 2이다.
0.002초 동안 6개의 입자가 카운터를 통과할 확률은?

Example 3

시간에 따라 컴퓨터 시스템에 수신되는 어떤 신호의 관찰 횟수 $X(t)$가 $\lambda=8$인 푸아송과정을 이룬다고 할 때, 다음 확률을 구하여라.

$$P(X(2.5)=17,\ X(3.7)=22,\ X(4.3)=36)$$

§4.2. 초기하분포

Example

N개의 공이 들어 있는 주머니 속에 흰색 공이 r개, 검정색 공이 $N-r$개 들어있다.
이 주머니에서 임의로 비복원추출법에 의해 공을 n개를 꺼낼 때, n개 공 안에 포함된 흰색 공이 정확히 x개 포함될 확률과 x의 범위를 구하여라.

Definition

크기가 N인 유한 모집단이 크기가 r 과 $r_0=N-r$ 인 두 개의 부분 모집단 A, B로 나누어져 있다. 모집단에서 n개의 표본을 임의로 비복원추출 할 때, 추출된 n개의 표본에서 부분 모집단 A안의 원소의 개수를 X라 하면, X의 확률질량함수는

$$f_X(x)=\begin{cases} \dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & , \ x\in R_X \\ 0 & , \ \text{otherwise} \end{cases}$$

이고, X의 치역공간은

$$R_X=\{x \mid \max\{0, n-(N-r)\}\le x\le \min\{n,r\}\}$$

이다.
이와 같은 확률질량함수를 갖는 확률변수 X의 확률분포를 모수 N, r, n 인 **초기하분포**(hypergeometry distribution)라 하고, 기호로 $X\sim HG(N;r,n)$로 나타낸다.

Theorem

$X\sim HG(N;r,n)$이면, 다음이 성립한다.

(1) $E[X]=n\dfrac{r}{N}=np$ (단, $p=\dfrac{r}{N}$)

(2) $Var[X]=n\dfrac{r}{N}\left(1-\dfrac{r}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

(3) N이 충분히 크고 $\dfrac{n}{N}\le 0.1$이면, $X\approx B(n,\dfrac{r}{N})$이다.

즉, $P(X=x)\approx\binom{n}{x}p^x(1-p)^{n-x}$이다.

참고 분산식에서 $\left(\dfrac{N-n}{N-1}\right)$을 **유한 모집단 수정계수**라 하며, N이 커질수록 이 값은 1로 수렴한다.

(proof)

Example

30개의 부품들로 구성된 한 ※로트에서 불량품이 3개 들어있다. 그 로트에서 임의로 3개의 부품을 취할 때, 정확히 1개의 불량품이 발견될 확률은 얼마인가?

※ 로트 : 1회에 생산되는 특정수의 제품 단위

Example 1

10개의 동전이 들어 있는 주머니 안에 100원짜리 동전이 4개 들어 있다. 이 주머니에서 임의로 동전 5개를 선정할 때, 선정된 동전 안에 들어 있는 100원짜리 동전의 개수를 X라 한다. 이때 다음을 구하여라.

(1) X에 대한 확률질량함수와 확률분포

(2) 100원짜리 동전이 한 개 또는 두 개 나올 확률

(3) 선정된 100원짜리 동전의 평균과 분산

Example 6 (§4.3)

빨간 공 75개와 파란 공 1425개가 들어 있는 상자에서 10개의 공을 임의로 꺼낸다. 이때 꺼낸 공 안에 빨간 공이 2개 들어 있을 근사확률을 구하여라.

Definition

크기가 N인 유한 모집단이 크기가 $r_1, r_2, \cdots, r_k$ 인 k개의 부분 모집단 $A_1, A_2, \cdots, A_k$로 나누어져 있다. 모집단에서 n개의 표본을 임의로 비복원추출 할 때, 추출된 n개의 표본에서 부분 모집단 $A_1, A_2, \cdots, A_k$안의 원소의 개수를 $X_1, X_2, \cdots, X_k$라 하면, n의 표본 중에

$$X_1 = x_1,\ X_2 = x_2,\ \cdots,\ X_k = x_k$$

일 확률은 다음과 같다.

$$P(X_1 = x_1, \cdots, X_k = x_k) = \frac{\binom{r_1}{x_1}\binom{r_2}{x_2}\cdots\binom{r_k}{x_k}}{\binom{N}{n}}$$

$$(x_1 + x_2 + \cdots + x_k = n,\ 0 \le x_j \le r_j,\ 1 \le j \le k)$$

이러한 확률분포를 **다변량 초기하 분포**(multivariable hypergeometry distribution)라 한다.

Example 2

주머니 안에 빨간 공과 파란 공 그리고 노란 공이 각각 8개, 10개, 10개씩 들어 있는 주머니에서 임의로 공 10개를 꺼낸다. 이때 선정된 10개 중에서 빨간 공과 파란 공 그리고 노란 공의 수를 각각 X, Y, Z라 한다.

(1) 확률변수 X, Y, Z의 결합확률질량함수를 구하여라.

(2) 선정된 10개 중에서 빨간 공과 파란공, 그리고 노란 공이 각각 3, 3 그리고 4개씩 포함될 확률을 구하여라.

(3) 선정된 10개 중에 포함될 노란 공의 평균을 구하여라.

§4.6. 다항분포

Definition

매회 실험결과가 k개의 서로 배반인 사건 $A_1, A_2, \cdots, A_k$로 구성되고, 각각의 사건이 매회 발생할 가능성이

$$P(A_i) = p_i, \ \ i = 1, 2, \cdots, k$$

이고 $p_1 + p_2 + \cdots + p_k = 1$인 실험을 n번 독립적으로 반복 시행한다고 하자.

n번의 독립인 시행에서 사건 A_i의 발생횟수를 X_i라 하면, k차 확률변수 $(X_1, X_2, \cdots, X_k)$의 치역공간은

$$R_{X_1 \times \cdots \times X_k} = \{(x_1, \cdots, x_k) \mid 0 \le x_i \le n, \ x_1 + \cdots + x_k = n\}$$

이고 결합확률질량함수는

$$f_{X_1, \cdots, X_k}(x_1, \cdots, x_k) = \begin{cases} \dfrac{n!}{x_1! \cdots x_k!} p_1^{\ x_1} \cdots p_k^{\ x_k}, & (x_1, \cdots, x_k) \in R_{X_1 \times \cdots \times X_k} \\ 0 & , \text{otherwise} \end{cases}$$

이다.

이와 같은 확률질량함수를 갖는 확률변수 $X_1, \cdots, X_k$는 모수 n과 $p_1, \cdots, p_k$를 갖는 **다항분포**(mulinomial distribution)을 이룬다고 하고, 기호로 $X \sim M(n, p_1, \cdots, p_k)$로 나타낸다.

참고 각 확률변수 X_i의 주변확률분포 $X_i \sim B(n, p_i)$이고, 평균과 분산은

$$E[X_i] = np_i, \ \ Var[X_i] = np_i(1 - p_i)$$

이다.

Example

$(X_1, X_2, X_3, X_4) \sim Mult\left(20, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\right)$일 때, 다음을 구하여라.

(1) (X_1, X_2, X_3, X_4)의 결합확률질량함수

(2) 각 확률변수 X_i의 평균과 분산

(3) X_2가 2이하일 확률

Example

주머니 안에 빨간 공 3개, 파란 공 4개, 노란 공 3개가 들어 있고 복원추출에 의하여 차례대로 공 5개를 꺼낸다. 이때 5개의 공 안에 포함된 빨간 공의 수를 X, 파란 공의 수를 Y, 노란공의 수를 Z라 한다.

(1) X, Y, Z의 결합확률질량함수를 구하여라.

(2) 빨간 공이 1개, 파란 공이 2개, 노란 공이 2객 나오는 확률을 구하여라.