엑셀.

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

n	x	p = 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

(sol)

(1) $P(X=2) = P(X \le 2) - P(X \le 1) = 0.6778 - 0.3758 = 0.3020$

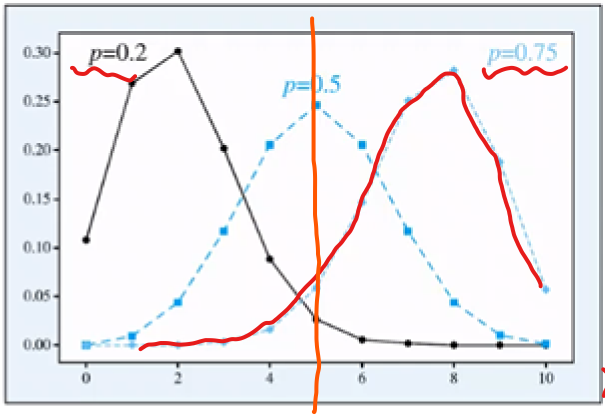
(2) $P(X \ge 5) = 1 - P(X \le 4) = 1 - 0.9672 = 0.0328$

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 $\mu_X = \frac{n}{2}$을 중심으로 좌우대칭이고, 이 경우 **대칭이항분포**(symmetric binomial distribution)라 한다.



p에 따른 이항분포의 비교

$X_i \sim B(1,p)$

$X = \sum_{i=1}^{n} X_i \sim B(n,p)$

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, 선형성

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

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

이다.

$X_1, X_2, \cdots, X_n$: 독립

(2) 두 확률변수 X와 Y가 독립이고 각각

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

인 이항분포를 이룬다면, $Z = X + Y$일 때

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

이다. 즉, $Z \sim B(m+n, p)$이다.

전체성공횟수 $= k$

$Z = X + Y = k$

$X = x$, $Y = k - x$

$(proof)$

(2) $P(Z=k) = \sum_{x \in R_X} P(X=x, Y=k-x) = \sum_{x \in R_X} P(X=x)P(Y=k-x)$ ← X, Y : 독립

$$= \sum_{x \in R_X} \binom{m}{x} p^{\cancel{x}} (1-p)^{m-\cancel{x}} \binom{n}{k-x} p^{k-\cancel{x}} (1-p)^{n-(k-\cancel{x})}$$

$$= \sum_{x \in R_X} \binom{m}{x}\binom{n}{k-x} p^k (1-p)^{(m+n)-k} = \binom{m+n}{k} p^k (1-p)^{(m+n)-k}$$

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

상수

$$\sum_{x=0}^{m} \binom{m}{x}\binom{n}{k-x} = \binom{m+n}{k}$$

순서상관없이 k개를 선택

Example 5

1명 1명.

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.

10+5=

(1) 15문제에서 두 학생이 정답을 선택한 평균 문제 수

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

(*sol*)

(1) $X \sim B(\overset{=m}{10}, 0.2)$, $Y \sim B(\overset{=n}{5}, 0.2)$이고, X와 Y는 독립이다.

$Z = X + Y \sim B(\overset{=m+n}{15}, 0.2)$이므로, 평균은 $\mu_Z = 15\left(\frac{1}{5}\right) = 3$

(2) $P(Z \geq 5) = 1 - P(Z \leq 4) = 1 - 0.8358 = 0.1642$

교재-이항 누적분포표

누적성공횟수.

n	x	p = 0.20
15	0	0.0352
	1	0.1671
	2	0.3980
	3	0.6482
	4	$\boxed{0.8358}$
	5	0.9386

Theorem 시행횟수 성공률.

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

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

성공률

이다.

$n=10.$

2번 성공.

$\dfrac{2}{10}$ = 성공비율

(*proof*)

$$E[Y] = E\left[\frac{X}{n}\right] = \frac{1}{n} E[X] = \frac{1}{\cancel{n}} \cancel{n} p = p$$

$$Var[Y] = Var\left[\frac{X}{n}\right] = \frac{1}{n^2} Var[X] = \frac{1}{n^{\cancel{2}}} \cancel{n} p q = \frac{pq}{n}$$

§4.4. 기하분포와 음이항분포

이항분포 : X = n번의 시행 중 성공횟수

기하분포 : X = 첫번째 성공까지의 시행횟수

1. 이산균등분포
2. 이항분포
3. 기하분포 ✓
4. 음이항분포
5. 푸아송분포
6. 초기하분포
7. 다항분포

■ 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

자연수 전체 집합.

X의 확률질량함수는

실패확률 성공확률.

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

이다.

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

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

$x-1$번. x번째

$$\sum_{n=1}^{\infty} r^n$$

공비가 r인 등비수열의 합 = 기하급수

Geometric series.

$p=\frac{1}{2} \Rightarrow$ 2번의 시행.

$p=\frac{1}{3} \Rightarrow$ 3번의 시행.

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)

공비 $r=qe^t$ 인 등비수열의 합.

수렴 $|r|=|qe^t|<1$

$$(3)\ m_X(t)=E[e^{tX}]=\sum_{x\in R_X=\{1,2,3,\cdots\}} e^{tx} f_X(x)=\sum_{x=1}^{\infty} e^{tx} q^{x-1} p=\frac{p}{q}\sum_{x=1}^{\infty}(qe^t)^x=\frac{p}{q}\cdot\frac{qe^t}{1-qe^t}=\frac{pe^t}{1-qe^t}$$

공비

t에 관해서 미분.

$$m_X{}'(t)=\frac{pe^t(1-qe^t)-pe^t(-qe^t)}{(1-qe^t)^2}=\frac{pe^t}{(1-qe^t)^2}$$

$pe^t-\cancel{pq(e^t)^2}+\cancel{pq(e^t)^2}$

$$m_X{}''(t)=\frac{pe^t(1-qe^t)^2-pe^t\{2(1-qe^t)(-pe^t)\}}{(1-qe^t)^4}=\frac{pe^t(1+qe^t)}{(1-qe^t)^3}$$

$$(1)\ E[X]=m_X{}'(0)=\frac{p}{(1-q)^2}=\frac{1}{p}$$

$$(2)\ E[X^2]=m_X{}''(0)=\frac{p(1+q)}{(1-q)^3}=\frac{1+q}{p^2}$$

$$Var[X]=E[X^2]-(E[X])^2=\frac{1+q}{p^2}-\left(\frac{1}{p}\right)^2=\frac{q}{p^2}$$

X : 기하분포.

Theorem 무기억성 성질(memorylessness property)

$X \sim G(p)$이면, 임의의 양의 정수 $a,\ b$에 대하여

$$P(X > a+b \mid X > a) = P(X > b) = (1-p)^b$$

조건부확률

이다.

$P(X \geq a+b \mid X > a) = P(X \geq b).$

(*proof*) 기하분포의 무기억성 성질은 처음 성공할 때까지 반복 시행한 횟수는 그 이후로 다시 성공할 때까지 반복 시행한 횟수에 독립이다.

① $$P(X > a) = \sum_{x=a+1}^{\infty} f_X(x) = \sum_{x=a+1}^{\infty} (1-p)^{x-1} p = \frac{(1-p)^a p}{\cancel{1} - (\cancel{1} - p)} = (1-p)^a$$

초항:

등비수열.

공비

a번째까지 실패

② $$P(X > a+b \mid X > a) = \frac{P(X > a+b)}{P(X > a)} = \frac{(1-p)^{a+b}}{(1-p)^a} = (1-p)^b = P(X > b).$$

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.

1 2 3 4 5 6 7

$P(X \le 5) = 1 - P(X > 6)$

(sol)

(1) 주사위를 던져서 1 또는 2의 눈이 나올 확률은 $\frac{1}{3}$이므로, $X \sim G(\frac{1}{3})$ 이다.

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

$= \frac{\frac{2}{3}}{(\frac{1}{3})^2}$

$(\therefore)\ p = \frac{1}{3},\ q = 1 - p = \frac{2}{3}$이므로, $\mu_X = \frac{1}{p} = 3,\ \sigma_X{}^2 = \frac{q}{p^2} = 6$ 이다.

(2) $P(X = 3) = f_X(3) = (\frac{1}{3})(\frac{2}{3})^2 = \boxed{\frac{4}{27}}$

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.

(*sol*)

(3) $P(X \le 5) = 1 - P(X > 5) = 1 - \left(\frac{2}{3}\right)^5 = \frac{211}{243}$

$\frac{32}{243}$

(4) $P(X \ge 5+3 \mid X > 5) = P(X \ge 3) = P(X > 2) = \left(\frac{2}{3}\right)^2 = \frac{4}{9}$

(5) X_1 = 처음으로 1 또는 2의 눈이 나올 때까지 던진 횟수

X_2 = 첫 번째 성공이후, 처음으로 1 또는 2의 눈이 나올 때까지 던진 횟수

$$(\therefore)\ P(X_2 = 3 \mid X_1 = 5) = \frac{P(X_2=3, X_1=5)}{P(X_1=5)} = \frac{P(X_2=3)P(X_1=5)}{P(X_1=5)} = P(X_2=3) = \boxed{\frac{4}{27}}$$

X_1, X_2 : 독립

무기억성.

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?

$\Rightarrow \frac{5}{100} = \frac{1}{20}$

(sol

$X \sim G(p = 0.05)$이므로, 기하분포의 무기억성에 의해

외워.

$$P(X \geq 25 \mid X > 20) = P(X \geq 5)$$

$$= P(X > 4) = (0.95)^4 = 0.81450625$$

반올림.

넷째자리

이항분포 : X = n번까지 시행에서 성공횟수
기하분포 : X = 첫번째 성공까지 시행횟수
음이항분포 : X = r번째 성공까지 시행횟수

■ 음이항 분포

Definition

성공할 확률이 p인 베르누이 시행을 독립적으로 계속 반복하는 확률실험에서 확률변수 X를

$$X=\text{'}r\text{번째 성공이 일어날 때까지의 시행 횟수'}$$

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

r이상의 자연수

X의 확률질량함수는

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

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.

시행에서 r번째 성공

참고 $x-1$번 중에 $r-1$번 성공 　 x번째 성공 　 확률

q q ⋯ q p q q ⋯ q p q q ⋯ q 　 p 　 $\binom{x-1}{r-1} q^{x-r} p^r$

$\binom{x-1}{r-1}$

순서상관없이!

참고 이항급수(binomial series)

이항정리. 이항계수.

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.

$$\binom{r}{n} = \frac{\overbrace{r\cdot(r-1)\cdots(r-n+1)}^{n\text{개}}}{n!}$$

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

$$\binom{-r}{1},\quad \binom{-r}{2},\quad \binom{-r}{3}$$

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

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

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

$1 = \binom{r}{0}$

$1 = \binom{r-1}{0}$

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

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

이항계수

$\binom{r}{n} = \binom{r}{r-n}$ 을 이용.