제 5 장 이산확률분포

1. 확률변수와 확률분포

1.1 확률변수(random variable)

확률변수란 실험 또는 관찰에서 일정한 확률을 가지고 발생하는 사건에 수치를 부여한 것을 말하며, 보통 X로 표시한다.

1.1.1 확률변수의 예

동전을 던졌을 때 앞면이 나오는 사건 - 앞면 아니면 뒷면이다.

앞면이 나오면 값 1을, 뒷면이 나오면 값 0을 부여하자. 이러한 값들을 X라 하자.
동전 앞면 또는 뒷면이 발생하는 사건은 일정한 확률을 가지고 발생하므로 X는 확률변수이다.
X가 1 또는 0의 값을 가질 확률은 $P(X = 0) = 0.5$, $P(X = 1) = 0.5$이다.

동전을 두 개 던졌을 때 앞면의 수 – 하나도 없거나, 한 개 아니면 두 개이다.
확률변수 X를 사용하여 없으면 0, 한 개면 1, 두 개면 2의 값을 부여하자.
관련 확률은 $P(X = 0) = 0.25, P(X = 1) = 0.50$, $P(X = 2) = 0.25$이다.

이산확률변수(discrete random variable), 연속확률변수(continuous random variable)

1.1.2 확률분포(probability distribution)

확률분포란 어떤 확률변수가 취할 수 있는 모든 값들과 이 값들이 나타날 확률을 나타낸 것이다.

확률분포의 예 - X: 동전을 두 번 던졌을 때 앞면이 나온 횟수

X	$P\,(X = x)$
0	0.25
1	0.50
2	0.25

앞면의 수 X	앞면의 수가 특정한 값 x일 확률 $P(X = x)$
{앞면의 수 =} 0	앞면의 수가 0일 확률 $P(X = 0)$
{앞면의 수 =} 1	앞면의 수가 1일 확률 $P(X = 1)$
{앞면의 수 =} 2	앞면의 수가 2일 확률 $P(X = 2)$

X	$P(X)$
0	0.25
1	0.50
2	0.25

앞면의 수 X	앞면의 수에 관한 확률 $P(X)$
{앞면의 수 =} 0	{앞면의 수가} 0일 확률 $P(0)$
{앞면의 수 =} 1	{앞면의 수가} 1일 확률 $P(1)$
{앞면의 수 =} 2	{앞면의 수가} 2일 확률 $P(2)$

1.1.2.1 상대도수분포와 확률분포

동전 2개씩 1,000번 던진 결과, 2개 중 앞면이 한 개도 없는 경우는 250, 앞면이 한 개인 경우는 500, 모두 앞면이 경우는 250번이었다. 이를 상대도수분포로 정리하고 이를 확률분포와 비교하면 다음과 같다.

계급	빈도수	상대도수
0	250	0.250
1	500	0.500
2	250	0.250
합계	1,000	1.000

확률분포와 비교하면 →

변수(X)	확률 P(X)
0	0.250
1	0.500
2	0.250
합계	1.000

2. 이산확률변수

2.1 이산확률분포

이산확률변수에서의 확률함수의 성질

(1) 모든 x값에 대하여 $p(x) \geq 0$→ $p(x)$, $P(X)$는 표현의 차이일 뿐이다.

(2) $\sum_x p(x) = 1$ → $\sum_x ()$ 는 모든 x에 대해 ()의 덧셈을 수행한다.

ex) 앞의 동전 던지기의 확률함수

2.2 누적분포함수(cumulative distribution function)

일반적으로 $F(x)$로 표기

$F(x) = P(X \leq x)$ → 여기서 X는 확률변수, x는 특정한 값

→ $F(x)$는 $P(X \leq x)$로 정의된다.

→ $P(X \leq x)$는 -∞에서 특정한 값 x까지에 해당하는 확률

2.2.1 앞의 동전 던지기의 예

$F(0) = P(X \leq 0) = P(X = 0)$ → X가 0이하인 경우를 모두 나열하면 X=0

$= P(0) = 0.25$ → $P(X = 0)$을 간략히 $P(0)$으로 표현

$F(1) = P(X \le 1)$ → X가 1 이하인 경우를 모두 나열하면 X=0,1
$= P(X = 0) + P(X = 1) = P(0) + P(1) = 0.75$

$F(2) = P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$
$= P(0) + P(1) + P(2) = 1.00$

2.2.2 누적상대도수와 누적분포

계급	빈도수	상대도수	누적상대도수
0	250	0.250	0.250
1	500	0.500	0.750
2	250	0.250	1.000
합계	1,000	1.000	1.000

변수(X)	확률 P(X)	누적 F(X)
0	0.250	0.250
1	0.500	0.750
2	0.250	1.000
합계	1.000	1.000

연습문제 1. 확률변수 X는 4, 8, 12의 **값을 갖는다**. $P(4) = 0.2$**이다.**

문제 1. $F(4) =$

문제 2. $P(8) = 0.3$ 이라면, $F(8) =$

문제 3. $P(8) = 0.3$ 이라면, $P(12) =$

문제 4. $F(X)$는 항상 증가한다. True, False

3. 이산확률분포의 측정

3.1 X와 X_i에 대한 설명

① X가 가질 수 있는 값의 수를 n로 표기하자. 동전을 두 개 던졌을 때 앞면이 나온 횟수를 X라 하면, X는 0, 1, 2의 3가지 값만 갖는다. 동전 두 개 던지는 실험을 1,000번 실시해도 X값의 유형은 총 3가지이다. 여기서 n = 3이다.

② X값 유형의 수를 n이라 하자. X값의 유형 중 첫 번째 값은 X_1, 두 번째 값은 X_2, ..., n번째 값은 X_n으로 표기하자. ①의 예에서 $X_1 = 0$, $X_2 = 1$, $X_3 = 2$가 된다.

$X = 2$ 앞면이 나온 횟수(X)는 2이다. → 앞면이 두 번 나왔다.

$X_2 = 1$ 앞면이 나오는 횟수 중 두 번째 경우의 수는 1이다.

$X = X_3$ 앞면이 나오는 횟수 중 세 번째 경우의 수가 발생했다.

3.2 평균(기댓값) Expected Value

이산확률변수 X가 X_1, $X_2, \ldots$, X_n의 값을 취하며 또한 X가 X_i $(i = 1, 2, \ldots, n)$의 값을 취할 확률이 $P(X_i)$일 때 X의 기댓값은 다음과 같다.

$$E(X) = \sum_{i=1}^{n} X_i \cdot P(X_i)$$

3.2.1 평균(기댓값)의 설명

① X의 평균은 어떻게 구할까? 예에서 X=0인 경우는 전체 중 0.25, X=1인 경우는 전체 중 0.50, X=2인 경우는 전체 중 0.25이다. 이에 적절한 방법은 가중평균이다. X의 가중평균은 $0.25\times0 + 0.50\times1 + 0.25\times2 = 1.0$이 된다.
여기서 0.25, 0.50, 0.25는 가중치인데, 이 값은 확률값과 동일하다.
그러므로, $\sum_{i=1}^{n} X_i \cdot P(X_i)$는 가중치가 확률값인 가중평균이다.
→ $\sum(\ )P(X_i)$는 ()의 평균을 의미한다.

② 일반적인 가중평균은 $\sum_{i=1}^{n} w_i \cdot X_i$ 형식으로 기입하지만, 기댓값의 경우는 $\sum_{i=1}^{n} X_i \cdot P(X_i)$로 기입한다.

Example - Expected Value

X_i	$P(X_i)$	$X_i \cdot P(X_i)$
-600	0.5	-300
600	0.2	120
1,000	0.3	300
합계	1.0	120 (기댓값)

연습문제 2.

어떤 수익 상품에 투자하는 데, 이익은 –600, 200, 800의 세 경우가 있다. 이익이 –600일 확률은 0.40, 이익이 200일 확률은 0.25, 800일 확률은 0.35이다. 이익금을 X라 하자.

문제 1. 발생가능한 이익금의 경우의 수(n)는?

문제 2. 이익금을 올림차순으로 정렬했다고 하자. X_1 = ______

문제 3. 이익금과 그 이익금이 발생할 확률을 표로 정리하고자 한다. 각 칸의 의미를 설명하시오.

i	X_i	$P(X_i)$
1		
2		
3		

문제 4. 이익금과 그 이익금이 발생할 확률을 표로 정리하고자 한다. 기호로 쓰시오.

i	X_i	$P(X_i)$
1		
2		
3		

문제 5. 이익금과 그 이익금이 발생할 확률을 표로 정리하고자 한다. 숫자로 쓰시오.

i	X_i	$P(X_i)$
1		
2		
3		

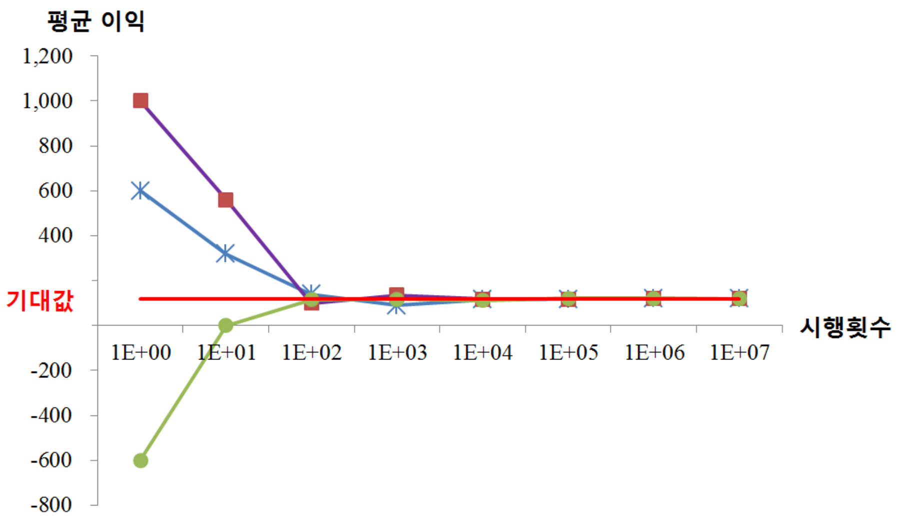
문제 6. 이 상품에 1,000 번 투자한다고 하자. 1,000 번 중 –600 의 이익이 실현되는 경우는 대략 몇 번쯤일까?

문제 7. 이 상품에 1,000,000 번 투자한다고 하자. 1,000,000 번 중 800 의 이익이 실현되는 경우는 대략 몇 번쯤일까?

문제 8. 이 상품의 기대이익을 구하시오.

3.2.2 기댓값의 컴퓨터 모의실험

시행횟수	해당 시행횟수까지의 이익 평균		
	실험 1	실험 2	실험 3
1	600.00	1,000.00	-600.00
10	320.00	560.00	0.00
100	140.00	100.00	116.00
1,000	90.00	135.60	116.80
10,000	115.40	117.40	112.36
100,000	118.55	117.34	120.95
1,000,000	118.86	120.11	120.74
10,000,000	119.59	119.93	119.72



지수형식으로 숫자 표기 방법

3.27E+06 = 3.27×10^6 = 3.27×1,000,000

3.27E-0.6 = $3.27\times\frac{1}{10^6}$ = $3.27\times\frac{1}{1{,}000{,}000}$

3.2.2.1 기댓값의 컴퓨터 모의실험 해석

① 실험 1, 2, 3 모두 시행횟수가 작은 경우에는 시행횟수까지의 평균값과 기댓값이 일치하지 않는다. 예) 시행횟수가 10인 경우, 평균은 각각 320, 560, 0이다.

② 시행횟수가 증가할수록 시행횟수까지의 평균값은 기댓값에 근접한다. 예) 시행횟수가 천만인 경우, 평균은 각각 119.59, 119.93, 119.72이다.

③ 기댓값 공식은 일종의 가중평균이다. 여기서 가중치는 확률이다. 상대빈도정의에 따르면 확률은 실험의 횟수가 무한대에 이를 때의 상대빈도이다.

④ 종합하면 시행횟수가 작을 때의 평균값은 기댓값과 다를 수 있지만, 횟수가 증가할수록 평균값은 기댓값에 근접하며, 횟수가 무한대이면 평균값은 기댓값과 일치한다.

3.2.3 산술평균과 기댓값

3.2.3.1 산술평균의 예

X_i는 i번째 관찰값을 의미한다. 예: X_3= -600

동일한 값(예: -600)이 반복될 수 있다.

번호	데이터값
1	1,000
2	600

번호	X
1	X_1
2	X_2

3	-600
4	1,000
5	-600
…	…
1,000	600

3	X_3
4	X_4
5	X_5
…	…
1,000	$X_{1,000}$

$$\mu = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_{1,000}}{N}$$

$$= \frac{1,000 + 600 - 600 + 1,000 + \cdots + 600}{1,000}$$

$$| \leftarrow \text{약 } 500\text{개} \rightarrow | \; | \leftarrow \text{약 } 200\text{개} \rightarrow | \quad | \leftarrow \text{약 } 300\text{개} \rightarrow |$$

$$= \frac{(-600 + \cdots - 600) + (600 + \cdots + 600) + (1,000 + \cdots + 1,000)}{1,000}$$ (분자를 정렬시키면)

$$= -600 \times \frac{500}{1,000} + 600 \times \frac{200}{1,000} + 1,000 \times \frac{300}{1,000}$$

= -600×0.5 + 600×0.2 + 1,000×0.3

3.2.3.2 기대값의 예

X_i는 X가 가질 수 있는 값 중 i번째 값을 의미한다. 예: X_3=1,000
여기서 동일한 값을 갖는 X_i는 없다.

앞 테이블에 있는 1,000개의 데이터를 도수분포표로 정리하면 아래와 같다.

번호	데이타값 (X)	빈도(f)	상대도수
1	-600	500	0.50 (500/1,000)
2	600	200	0.20 (200/1,000)
3	1,000	300	0.30 (300/1,000)
합계		1,000	1.00

번호	X	Prob.
1	X_1	$P(X = X_1)$
2	X_2	$P(X = X_2)$
3	X_3	$P(X = X_3)$
합계		1.00

위 테이블에서 총 관찰수는 1,000이지만, 관찰수가 ∞라면 (상대빈도정의에 의해) 상대도수는 확률이 된다.

3.2.4 기댓값의 특성

a, b: 상수, X: 확률변수

(1) $E(aX) = a \cdot E(X)$

(2) $E(X + b) = E(X) + b$

(3) $E(aX+b) = a \cdot E(X) + b$

$$E(aX+b) = \sum_{i=1}^{n} (aX_i + b) \cdot P(X_i)$$

$$= \sum_{i=1}^{n} aX_i \cdot P(X_i) + \sum_{i=1}^{n} b \cdot P(X_i)$$

$$= a\sum_{i=1}^{n} X_i \cdot P(X_i) + b\sum_{i=1}^{n} \cdot P(X_i)$$

$$= a \cdot E(X) + b \quad since \sum_{i=1}^{n} P(X_i) = 1$$

연습문제 3. 어떤 유형의 투자 사례 10개를 조사한 결과는 다음과 같다.
{-400, 100, 100, 600, -400, -400, 600, 600, 100, -400}
이번 조사 결과는 이 유형 투자의 전형적인 사례라 한다.

문제 1. 이익금의 확률분포표를 작성하시오. 단, 이익금을 나타내는 확률변수는 X라 하자.

i	X_i	$P(X_i)$
1		
2		
3		
합계		

문제 2. 이익금의 기댓값을 구하시오.

문제 3. 사례별로 이익금이 10 배씩 커지고 다시 30 이 추가되었다고 하자. 이때의 이익금의 기댓값을 구하시오.

3.3 분산(Variance)과 표준편차(Standard Deviation)

3.3.1 분산의 정의-확률변수가 이산변수인 경우

$$Var(X) = V(X) = \sum_{i=1}^{n} \left(X_i - E(X)\right)^2 \cdot P(X_i)$$

표준편차 $\sigma = \sqrt{V(X)}$ (표준편차는 비음)

3.3.1.1 (확률변수가 아닌) 자료의 분산과 표준편차

① 모집단 - 분산 $= \sigma^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}$, 표준편차 $= \sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N}}$

σ^2는 $(X_i-\mu)^2$의 평균이다. → $\frac{\Sigma(\ )}{N}$은 ()의 평균을 의미한다.

② 표본 - 분산 = $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$, 표준편차 = S = $\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$

3.3.1.2 V(X)와 σ^2

① $V(X)$는 함수형식이다. 여기서 함수명은 V인데, 분산을 나타내는 함수이므로 Variance의 첫 글자를 함수명으로 사용했다. Input은 X이다. 정리하면, $V(X)$란 X의 분산을 의미한다.

② $V(X)$와 σ^2은 동일하다. $V(X)$는 분산을 함수형식으로 표기한 것이고 σ^2은 기호로 표기한 것뿐이다. 또한 $E(X)$와 μ도 동일하다. $E(X)$는 평균을 함수형식으로 표기한 것이고 μ는 기호로 표기한 것뿐이다.

3.3.1.3 V(X)는 $(X_i - E(X))^2$의 가중평균이다

① 편차제곱의 (가중)평균이다.
→ $\sum w_i()$ 또는 $\sum(\ )P(X_i)$는 ()의 가중평균을 의미한다.

3.3.2 확률분포표가 주어졌을 때 분산과 표준편차 구하는 방법

공식: $V(X) = \sum_{i=1}^{n}\left(X_i - E(X)\right)^2 \cdot P(X_i)$

단계 1. $E(X)$를 구한다.

단계 2. 각 X_i별로 $\left(X_i - E(X)\right)^2 \cdot P(X_i)$를 구한 후 그 값들을 더한다. → 분산

단계 3. 분산의 (양의) 제곱근을 구한다. → 표준편차

X	$P(X)$
-200	0.4
300	0.3
500	0.3
합계	1.0

단계 1. $E(X)$ = -200(0.4) + 300(0.3) + 500(0.3) = 160

단계 2.

X	$P(X)$	$\{X - E(X)\}^2$	$\{X - E(X)\}^2 \cdot P(X)$
-200	0.4	$(-200 - 160)^2$ = 129,600	51,840

300	0.3	$(300-160)^2$ = 19,600	5,880
500	0.3	$(500-160)^2$ = 115,600	34,680
합계	1.0		92,400 = $V(X)$

단계 3. (표준편차) $\sigma = \sqrt{V(X)} = \sqrt{92{,}400} = 303.97$

예제. 확률분포표에 주어진 확률에 따라 X를 무작위로 발생시킨다고 하자.

① 편차제곱이 129,600일 확률은? Ans. 0.4

② 편차제곱이 19,600일 확률은? Ans. 0.3

③ 편차제곱이 115,600일 확률은? Ans. 0.3

④ 편차제곱의 평균은 ___ 보다는 크다. Ans. 19,600

⑤ 편차제곱의 평균은 ___ 보다는 작다. Ans. 129,600

⑥ 편차제곱의 (가중)평균은? Ans. 129,600(0.4) + 19,600(0.3) + 115,600(0.3) = 92,400

연습문제 4.

X	$P(X)$
-200	0.2
100	0.5
300	0.3
합계	1.0

문제 1. X 의 기댓값

문제 2. X 의 분산과 표준편차

문제 3. 총 백만개의 데이터가 있으며, 그 값들은 위의 분포를 정확히 따른다고 하자. -200, 100, 300 은 몇 개씩 존재하는가?

문제 4. (문제 3 의) 데이터들의 평균을 구하시오.

문제 5. (문제 3 의) 데이터들의 분산을 구하시오.

3.3.3 분산의 기댓값 특성

(1) $V(aX) = a^2 \cdot V(X)$

(2) $V(X+b) \;=\; V(X)$

(3) $V(aX\,+\,b) \;=\; a^2 \cdot V(X)$

$V(aX\,+\,b) \;=\; \sum[aX_i + b - E(aX+b)]^2 P(X_i)$

$= \sum[aX_i +\; b - \;aE(X) - \;b\;]^2 P(X_i)$ since $E(aX+b) \;=\; aE(X) \;+\mathrm{b}$

$= \sum[aX_i - aE(X)]^2 P(X_i)$

$= \sum a^2[X_i - E(X)]^2 P(X_i)$

$= a^2 \sum[X_i - E(X)]^2 P(X_i)$

$= a^2 \cdot V(X)$

(4) $V(X) \;=\; E(X^2) \;-\; E(X)^2$

$V(X) \;=\; \sum_{i=1}^{n} \{X_i - E(X)\}^2 \cdot P(X_\mathrm{i})$ 분산의 정의

$= \sum_{i=1}^{n} \left\{X_i^2 - 2X_iE(X) \,+\, E(X)^2\right\} \cdot P(X_i)$ $(a+b)^2 \;=\; a^2 \;+2ab \;+\; b^2$

$= \sum_{i=1}^{n} X_i^2 \cdot P(X_i) \;+\; \sum_{i=1}^{n} -2\,X_iE(X) \cdot P(X_i) \;+\; \sum_{i=1}^{n} E(X)^2 \cdot P(X_i)$

$(a+b+c) \cdot d \;=\; ad \;+\; bd \;+\; cd$

여기서, $\sum_{i=1}^{n} X_i^2 \cdot P(X_i) \;=\; E(X^2)$ .... (1)

$E(X)$와 $E(X)^2$은 상수이므로 $\sum$ 앞으로 뺄 수 있다.

$\sum_{i=1}^{n} -2\,X_iE(X) \cdot P(X_i)$

$= -2E(X) \sum_{i=1}^{n} X_i \cdot P(X_i)$ $E(X)$는 상수

$= -2\;\; E(X) \;\cdot E(X) \;= -2E(X)^2$ .... (2) $\sum_{i=1}^{n} X_i \cdot P(X_i) \;=\; E(X)$

$\sum_{i=1}^{n} E(X)^2 \cdot P(X_i)$

$= \; E(X)^2 \cdot \sum_{i=1}^{n} P(X_i)$ $E(X)^2$는 상수

$= \; E(X)^2$ .... (3) $\sum_{i=1}^{n} P(X_i) \;= 1$

(1), (2)와 (3)으로부터 $V(X) \;=\; E(X^2) \;-2E(X)^2 \;+\; E(X)^2 \;=\; E(X^2) \;-\; E(X)^2$

연습문제 5.

X	$P(X)$
-200	0.2
100	0.5
300	0.3
합계	1.0

문제 1. $E(X)$와 $V(X)$

문제 2. $V(10X)$

문제 3. $V(X+50)$

문제 4. $V(10X+50)$

4. 이항분포(Binomial Distribution)

4.0 (기초 정리) 계승(factorial), 조합(combination)

계승(factorial): n이 양의 정수이면 그 수보다 작거나 같은 모든 양의 정수의 곱이다.

$$n! = \prod_{k=1}^{n} k = 1 \times 2 \times \cdots \times n$$

0! = 1,　1! = 1,　2! = 1×2 = 2,　4! = 1× 2 × 3 × 4=24

조합(combination): n개 중 r개를 고르는 경우(순서는 무시된다.)

$${}_nC_r = \frac{n!}{r!\,(n-r)!}$$

$${}_7C_2 = \frac{7!}{2!\,(7-2)!} = \frac{7!}{2!\,5!} = \frac{7\times 6\times 5!}{2!\,5!} = 21$$

4.1 베르누이 시행 조건

1. 각 시행의 결과는 상호배타적인 두 사건으로 구성.
   두 사건을 편의상 성공(S)과 실패(F)로 분류
   → 확률변수 x, 흔히 성공이면 x에 1의 값을, 실패이면 0의 값을 부여한다.
2. 성공의 확률 = p, 실패의 확률 = 1-p. → 성공 확률 + 실패 확률 = 1
3. 각 시행은 독립적이다.

예) 야구에서 타자가 진루하면 (진루) 성공, 진루에 실패하면 (진루) 실패라 하자. 어떤 타자가 진루할 확률은 0.4이라 하자. 확률변수 x는 성공이면 1, 아니면 0의 값을 갖도록 하자.
진루할 확률 = P(진루 성공) = $P(x=1)$ = $P(1)$ = 0.4
진루에 실패할 확률 = P(진루 실패) = $P(x=0)$ = $P(0)$ = 1-0.4 = 0.6
성공 확률을 기호 p로 표기하면, 이 경우 p = 0.4이다.
실패 확률을 기호 p를 사용하여 표기하면, 이 경우 실패확률은 1-p이다.
여기서 P는 $P(x)$라는 함수의 이름이며, p는 성공확률의 구체적인 값을 나타내는 기

호이다. p를 v 또는 τ 등 다른 기호로 표현해도 무방하지만, 주로 p를 사용한다.

4.2 이항분포의 의의

4.2.1 이항분포 사례

10명 중 3명이 지하철과 버스 중 지하철을 이용할 확률

5명 중 2명이 두 개의 메뉴 중 첫 번째 메뉴를 선택할 확률

부품 20개 중 불량품이 19개일 확률

4.2.2 이항분포의 예-동전을 n번 던졌을 때 앞면의 수 분포

S: 앞면, F: 뒷면, n: 시행횟수, X: 앞면의 수

동전을 n번 던졌을 때 앞면이 나오는 횟수의 분포를 알고자 한다.

① 동전을 1회 던졌을 때 결과는 앞면 또는 뒷면이다. → 상호배타적인 두 사건이다.
앞면이 나오면 성공 S, 뒷면이 나오면 F로 표기하자.

② 동전을 n번 던졌을 때 앞면(성공) 횟수를 X라 하자.
X는 $0, 1, 2, \ldots, n$의 값을 갖는다.

③ 여기서 관심 사항은 n번 던졌을 때 성공 횟수 X의 확률분포이다.

4.2.2.1 $n = 3$일 때 발생 case

case	시도			성공 횟수	실패 횟수	발생확률	발생확률 (일반화)
	1	2	3				
1	S	S	S	3	0	$p\ p\ p$	$p^3(1-p)^0$
2	S	S	F	2	1	$p\ p(1\text{-}p)$	$p^2(1-p)^1$
3	S	F	S	2	1	$p(1\text{-}p)p$	$p^2(1-p)^1$
4	S	F	F	1	2	$p(1\text{-}p)(1\text{-}p)$	$p^1(1-p)^2$
5	F	S	S	2	1	$(1\text{-}p)p\ p$	$p^2(1-p)^1$
6	F	S	F	1	2	$(1\text{-}p)p(1\text{-}p)$	$p^1(1-p)^2$
7	F	F	S	1	2	$(1\text{-}p)(1\text{-}p)p$	$p^1(1-p)^2$
8	F	F	F	0	3	$(1\text{-}p)(1\text{-}p)(1\text{-}p)$	$p^0(1-p)^3$

주의: $p^0 = 1$이다. (예: $0.7^0 = 1$)

성공횟수	해당 case	해당 case 확률 합	확률의 일반화
3	1	$1 \times p^3(1-p)^0$	${}_3C_3\, p^3(1-p)^0$
2	2, 3, 5	$3 \times p^2(1-p)^1$	${}_3C_2\, p^2(1-p)^1$

1	4, 6, 7	$3\times p^1(1-p)^2$	${}_3C_1\,p^1(1-p)^2$
0	8	$1\times p^0(1-p)^3$	${}_3C_0\,p^0(1-p)^3$

3번 시도하여 성공횟수가 1일 확률

$= P(X=1|n=3)$

= 위 표에서 4번, 6번 또는 7번이 발생할 확률의 합

$= P(S,F,F) + P(F,S,F) + P(F,F,S)$

$= p(1-p)(1-p) + (1-p)p(1-p) + (1-p)(1-p)p$

$= 3p(1-p)(1-p)$

즉, ${}_3C_1 \cdot p(1-p)(1-p)$이다.

연습문제 6.

문제 1. 4 번 시도하여 2 번 성공할 성공/실패 경우를 모두 나열하시오.

문제 2. (문제 1 의) 각 경우별 발생확률을 기입하시오. 단, 1 회 성공확률은 p로 한다.

문제 3. 4 번 시도하여 2 번 성공할 확률은? 단, 1 회 성공확률은 p로 한다.

문제 4. n번 시도하여 x번 성공할 확률은? 단, 1 회 성공확률은 p로 한다.

4.3 이항확률함수

$P(X = x\,|n,\ p) = {}_nC_x\ p^x(1-p)^{n-x}$

X: (성공횟수를 나타내는) 확률변수, x: (특정한) 성공횟수, n: 시행횟수,

p: (단일 시행에서의) 성공확률, 1-p: 실패확률, ${}_nC_x = \frac{n!}{x!(n-x)!}$

Note: ${}_nC_x\ p^x(1-p)^{n-x} = \frac{(\text{시도횟수})!}{(\text{성공횟수})!\times(\text{실패횟수})!} \times \text{성공확률}^{\text{성공횟수}}\ \text{실패확률}^{\text{실패횟수}}$

Note: $\sum_{x=0}^{n}\ {}_nC_x\,p^x q^{n-x} = 1.0$이다. → 확률의 합은 1.0이다.
(여기서, q= 1-p = 실패확률)

$p + q = 1$이므로, n값이 무엇이든 $(p+q)^n = 1$이 성립한다.
$(p+q)^n$를 전개하면

$(p+q)^n = {}_nC_0p^0q^n + {}_nC_1p^1q^{n-1} + {}_nC_2p^2q^{n-2} + \cdots + {}_nC_np^nq^0 = 1$

$_nC_0p^0q^n$ + $_nC_1p^1q^{n-1}$ + $_nC_2p^2q^{n-2}$ + ⋯ + $_nC_np^nq^0$ 부분을
$\sum$으로 표기하면, $\sum_{x=0}^{n}$ $_nC_xp^xq^{n-x}$이다.

4.3.2.1 **n=3, p=0.3인 이항분포의 (컴퓨터) 모의실험**

실험 횟수	시행 차수별 성공여부			성공 횟수	빈도				상대빈도				합계
	1	2	3		0	1	2	3	-	1	2	3	
1	0	0	0	0	1	-	-	-	1.0000	-	-	-	1.0000
2	0	0	1	1	1	1	-	-	0.5000	0.5000	-	-	1.0000
3	0	1	0	1	1	2	-	-	0.3333	0.6667	-	-	1.0000
4	0	1	0	1	1	3	-	-	0.2500	0.7500	-	-	1.0000
5	1	1	1	3	1	3	-	1	0.2000	0.6000	-	0.2000	1.0000
6	0	0	0	0	2	3	-	1	0.3333	0.5000	-	0.1667	1.0000
7	0	0	0	0	3	3	-	1	0.4286	0.4286	-	0.1429	1.0000
8	1	1	0	2	3	3	1	1	0.3750	0.3750	0.1250	0.1250	1.0000
9	0	0	1	1	3	4	1	1	0.3333	0.4444	0.1111	0.1111	1.0000
10	0	0	0	0	4	4	1	1	0.4000	0.4000	0.1000	0.1000	1.0000
100	1	0	1	2	36	41	20	3	0.3600	0.4100	0.2000	0.0300	1.0000
1,000	0	1	0	1	344	436	197	23	0.3440	0.4360	0.1970	0.0230	1.0000
10,000	0	1	1	2	3,476	4,384	1,873	267	0.3476	0.4384	0.1873	0.0267	1.0000

실험횟수 1: 1, 2, 3차 시도에서 모두 실패, 이번 3차례 시도에서 총 성공횟수 x는 0, 현재까지 성공횟수 x가 0인 빈도는 1, 성공횟수가 1, 2, 3인 빈도는 0, 상대빈도는 x가 0인 경우는 1/1 = (해당 빈도/빈도의 합) = 1, 나머지는 모두 0/1 = 0. 모든 상대빈도의 합은 1.0이다.

실험횟수 2: 1, 2차 시도는 실패, 3차 시도는 성공, 이번 3차례 시도에서 총 성공횟수 x는 1, (실험 1과 2의) 현재까지 성공횟수 x가 0과 1인 빈도는 1, x가 2, 3인 빈도는 0, 상대빈도는 x가 0과 1인 경우는 1/2, 나머지는 모두 0/1 = 0. 모든 상대빈도의 합은 1.0이다.

실험횟수 10: 1, 2, 3차 모두 실패, 이번 3차례 시도에서 x = 0, 현재까지 (실험 1에서 10까지) x = 0, 1인 빈도는 4, x = 2, 3인 빈도는 1, 상대빈도는 x =0과 1인 경우는 각각 0.4, 2와 3인 경우는 가각 0.1이며, 상대빈도의 합은 1.0이다.

연습문제 7.

문제 1. 위의 표에서 확률변수(우리가 관심을 갖고 있는 대상)는 무엇인가?

문제 2. 위의 표를 도수분포표로 변환하시오.

문제 3. (상대빈도가 아닌) 빈도만을 사용하여 x의 평균, 분산, 표준편차를 구하시오. 단, 10,000 개를 대상으로 하며, 모집단으로 간주한다.

문제 4. 빈도가 아닌 상대빈도를 사용하여 평균, 분산, 표준편차를 구하시오. 단, 10,000 개를 대상으로 하며, 모집단으로 간주한다.

문제 5. 문제 4 에서 0, 1, 2, 3 을 $x_1, ..., x_4$로, 상대빈도를 $P(x_i)$로 변환하여 정리하시오. 단, 평균은 $E(X)$로, 분산은 $V(X)$로, 표준편차는 σ로 표기한다.

연습문제 8.

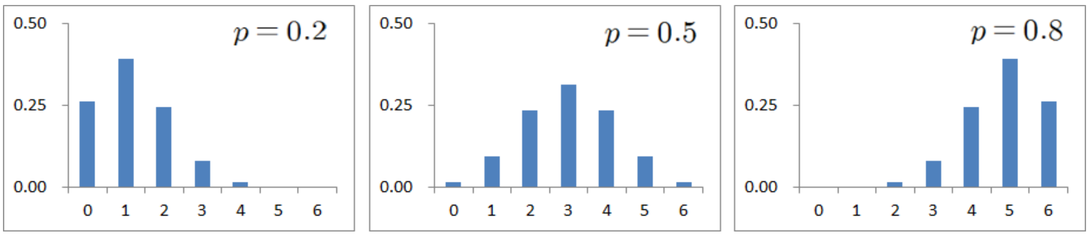
문제 1. 확률변수 X는 이항분포를 따른다. 모든 x에 대해 $P(X = x \mid n=3,\ p=0.3)$를 구하시오.

문제 2. 문제 1 에서 구한 확률값과 연습문제 7 에서 구한 상대빈도값이 다른 이유를 설명하시오.

문제 3. 확률변수 X는 이항분포를 따른다. 모든 x에 대해 $P(X = x \mid n=5,\ p=0.7)$를 구하시오.

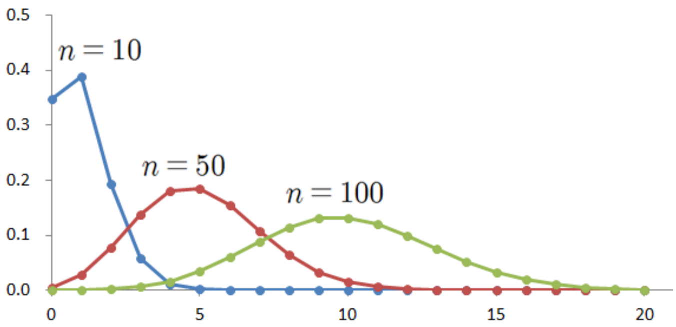
4.4 n과 p에 따른 이항확률분포그래프의 특징

n=6일 때 서로 다른 p의 이항분포



→ p=0.5일 때 대칭

p=0.1일 때 서로 다른 n의 이항분포



$\rightarrow$ p가 0.5가 아니라도 n이 충분히 커지면 대칭에 가까워진다.

4.5 이항분포의 기댓값과 분산

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

$V(X) = \sigma^2 = np(1-p) = npq$ 여기서 q는 실패확률이며 당연히 1-성공확률(p)이다.

$\sigma = \sqrt{npq}$

---

1 $E(x) = \sum_{i=1}^{n} X_i \cdot P(X_i)$ 이산확률변수의 기대값 공식

$= \sum_{x=0}^{n} x \cdot P(x)$ 이항분포의 x는 $0, 1, 2, \ldots, n$

$= \sum_{x=0}^{n} x \cdot {}_nC_x p^x q^{n-x}$ 이항분포에서 $P(x) = {}_nC_x p^x q^{n-x}$

$= \sum_{x=1}^{n} x \cdot {}_nC_x p^x q^{n-x}$ since $x = 0$이면 $x \cdot {}_nC_x p^x q^{n-x} = 0$

$$= \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)! \cdot x!} \cdot p^x q^{n-x}$$

$$= \sum_{x=1}^{n} \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)! \cdot (x-1)!} \cdot p^{(x-1)} q^{n-x}$$

$z = x - 1$로 설정하면

$$= np \sum_{z=0}^{n-1} \frac{(n-1)!}{(n-1-z)! \cdot z!} \cdot p^z q^{n-1-z}$$

$m = n - 1$로 설정하면

$$= np \sum_{z=0}^{m} \frac{m!}{(m-z)! \cdot z!} \cdot p^z q^{m-z}$$

$= np$ since $\sum_{z=0}^{m} \frac{m!}{(m-z)! \cdot z!} \cdot p^z q^{m-z} = 1$

Note: $E(X) = \sum_{i=1}^{n} X_i \cdot P(X_i)$와 $E(X) = np$ 의 차이

① 모든 이산확률변수의 $E(X)$는 $\sum_{i=1}^{n} X_i \cdot P(X_i)$으로 정의된다.

② 이항분포에서 $E(X)$을 ①에 따라 구하면, 그 값은 np 이 된다(정리하면 np 이 된다).

연습문제 9. 확률변수 X는 n=3이고 p=0.3인 이항분포를 따른다.

문제 1. n과 p를 사용하여 $E(X)$를 구하고, 그 결과를 연습문제 7에서 구한 평균과 비교하시오.

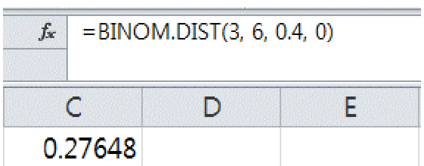
문제 2. n과 p를 사용하여 $V(X)$를 구하고, 그 결과를 연습문제 7에서 구한 분산과 비교하시오.

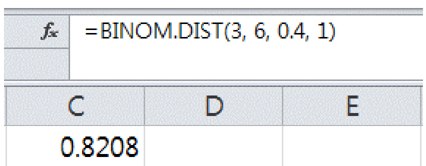
4.6 이항확률분포표 읽기

		p		
n	x	0.05	0.10	0.15
1	0	0.9500	0.9000	0.8500
	1	0.0500	0.1000	0.1500
2	0	0.9025	0.8100	0.7225
	1	0.0950	0.1800	**0.2550**
	2	0.0025	0.0100	0.0225
3	0	0.8574	0.7290	0.6141
	1	0.1354	0.2430	0.3251

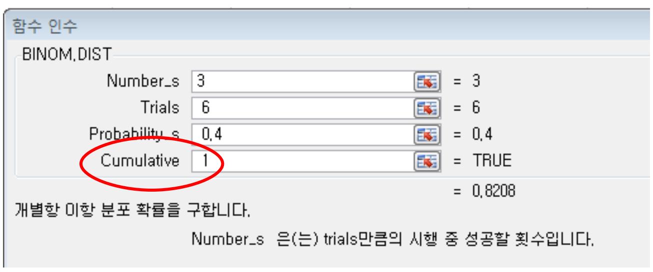
← $P(x = 1 | n = 2,\ p = 0.15) = 0.2550$

Excel 2010에서 이항분포 확률값 찾기

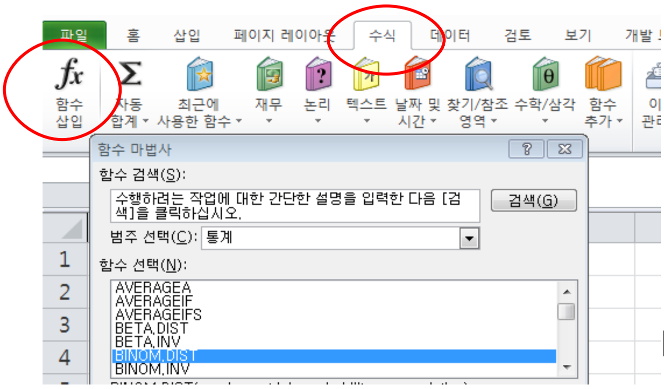




$P(x = 3 | n = 6, p = 0.4)$ $P(x \leq 3 | n = 6, p = 0.4)$



cumulative: 1이면 누적확률값을, 0이면 해당 확률값을 제시한다.



[수식 리본텝] - 함수 삽입 선택

5. 포아송분포(Poisson Distribution)

5.0 (기초 정리) e(자연로그 밑수)

오일러의 수, 네이피어 상수, 자연상수로 불린다.

(1) e의 정의

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

$$e = \sum_{n=0}^{\infty}\frac{1^n}{n!} = \frac{1}{0!}+\frac{1^1}{1!}+\frac{1^2}{2!}+\frac{1^3}{3!}+\cdots$$

(2) e의 주요 공식

$$e^z = \lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n$$

$$e^z = \frac{z^0}{0!}+\frac{z^1}{1!}+\frac{z^2}{2!}+\frac{z^3}{3!}+\cdots = \sum_{n=0}^{\infty}\frac{z^n}{n!}$$

5.1 포아송분포 사례

매장 방문하는 시간당 고객의 수

단위 면적당 결점의 수

일별 불만 건수

페이지당 오타의 수

5.2 포아송분포의 조건

(1) 관심대상이 되는 시간(또는 공간)은 아주 작은 단위구간(subinterval)으로 나누어질 수 있으며, 그 단위구간에서 어떤 사건이 발생할 가능성은 적다.

(2) 일정한 단위시간(또는 공간)에서 어떤 사건이 발생할 횟수와 다른 구간에서 그 사건이 발생할 횟수는 서로 독립적이다.

(3) 아주 작은 단위구간에서 둘 또는 그 이상의 사건이 발생할 확률은 극히 작다(0으로 간주)

(4) 관심의 대상이 되는 시간 내의 일정시간에서 발생하는 사건수의 확률분포는 다른 시간에서 발생하는 사건수의 확률분포와 같다.

5.3 포아송분포의 확률함수

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

여기서, e: 2.71828..... (자연로그의 밑수)

λ : (단위 시간 또는 공간당) 평균발생횟수, (lambda→λ, Lambda→Λ)

x : 사건발생횟수, 0, 1, 2, 3, …

5.3.1 Example. 포아송분포

10분당 3.5명 손님 내방 → $\lambda = 3.5$

(1) 10분간 방문 손님의 수 = 0일 확률

$$P(X = 0) = \frac{e^{-3.5} \cdot 3.5^{0}}{0!} = 0.0302$$

(2) 10분간 최대 방문자의 수가 3일 확률

$$P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= \frac{e^{-3.5} \cdot 3.5^{0}}{0!} + \frac{e^{-3.5} \cdot 3.5^{1}}{1!} + \cdots + \frac{e^{-3.5} \cdot 3.5^{3}}{3!} = 0.5367$$

5.3.2 Excel로 확률 구하기

fx =POISSON.DIST(0, 3.5, 0)

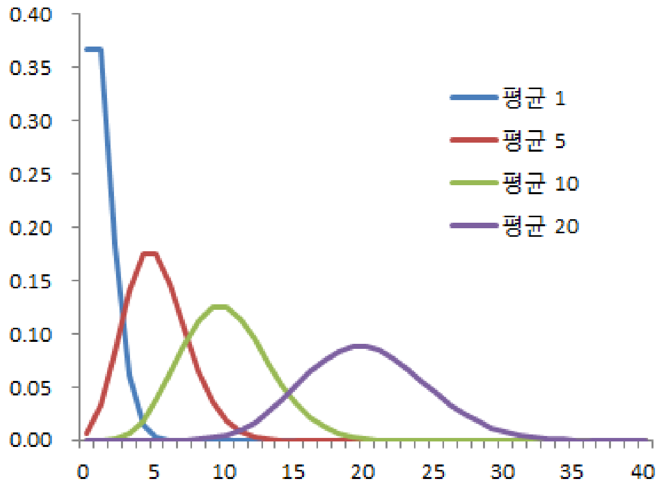
C	D	E
0.030197		

fx =POISSON.DIST(3, 3.5, 1)

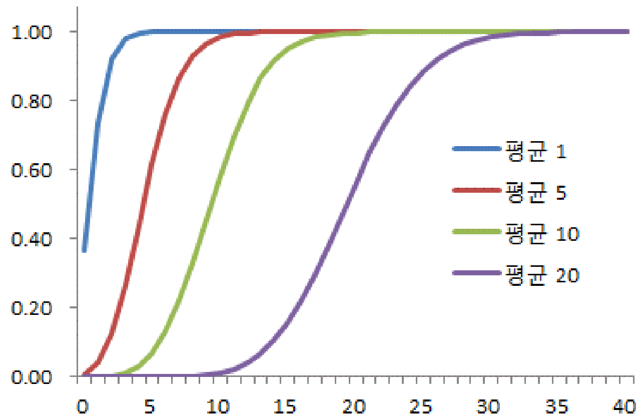
C	D	E
0.536633		

5.4 λ에 따른 포아송 확률분포 그래프

$\lambda = 1, 5, 10, 20$



$\lambda = 1, 5, 10, 20$인 포아송 누적확률분포 그래프



연습문제 10. 확률변수 X는 $\lambda = 3.5$인 포아송분포를 따른다.

문제 1. 아래 확률분포표를 완성하시오.

x	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	x	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
0	0.03020	8	
1	0.10569	9	0.00656
2	0.18496	10	0.00230
3	0.21579	11	0.00073
4		12	0.00021
5	0.13217	13	0.00006
6	0.07710	14	0.00001

7	0.03855	15	0.00000

문제 2. 기대값(평균)과 분산을 구하기 위한 아래 표를 완성하시오.

x	$P(x)$	$x \cdot P(x)$	$\left(x - E(X)\right)^2$	$\left(x - E(X)\right)^2 \cdot P(x)$
0	0.0302	-	12.2499	0.3699
1	0.1057	0.1057	6.2499	0.6606
2	0.1850	0.3699	2.2500	0.4161
3				
4	0.1888	0.7552	0.2500	0.0472
5	0.1322	0.6608	2.2500	0.2974
6				
7	0.0385	0.2698	12.2501	0.4722
8	0.0169	0.1349	20.2501	0.3415
9				
10	0.0023	0.0230	42.2502	0.0970
11	0.0007	0.0080	56.2502	0.0411
12				
13	0.0001	0.0007	90.2503	0.0052
14	0.0000	0.0002	110.2503	0.0016
15	0.0000	0.0001	132.2503	0.0004
	합계	3.5000		3.4999

5.5 포아송분포의 기댓값과 분산

$E(X) = \lambda$ [2]

---

[2] $E(x) \;=\; \sum_x x \cdot p(x)$ 모든 x에 대하여

$= \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!}$ 포아송확률변수는 0, 1, 2, ..., ∞의 값을 지닌다.

$= \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!}$ $x = 0$인 경우는 $x \cdot p(x) \;=\; 0$

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

$V(X) = \lambda$

연습문제 11. 확률변수 X는 평균이 λ인 포아송분포를 따른다.

문제 1. $E(X)$를 정의대로 기술하시오.

문제 2. $V(X)$를 정의대로 기술하시오.

연습문제 12. 확률변수 X는 $\lambda = 3.5$ 인 포아송분포를 따른다.

문제 1. $E(X)$

문제 2. $V(X)$

5.6 이항분포와 포아송분포의 연관성

n이 충분히 크고 ($n \geq 20$) 성공확률 p가 매우 작은 ($p \leq 0.05$) 경우 이항분포 확률값은 포아송분포 확률값에 접근한다.

${}_nC_x \; p^x(1-p)^{n-x} \simeq \frac{e^{-\lambda}\lambda^x}{x!}$ where $\lambda = np$ [3]

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

$$= \lambda \cdot e^{-\lambda} \cdot e^{\lambda} \qquad since\ \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}$$

$$= \lambda$$

3 $p(x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

$$= \frac{n!}{x!\,(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \quad Let\ \lambda = np \rightarrow p = \frac{\lambda}{n}$$

$$= \frac{n!}{x!\,(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)\cdot(n-x)!}{x!\,(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^x}$$

Example $p = 0.03$, n=20, x=2

이항분포: $P(X=2) = {}_{20}C_2(0.03)^2 \cdot (0.97)^{18}$ = 0.09883

포아송분포: $P(X=2) = \frac{e^{-(20\times0.03)}\times(20\times0.03)^2}{2!}$ = 0.09879 since $\lambda = np$

6. 제6절 초기하분포(Hypergeometric Distribution)

6.1 초기하분포의 조건

베르누이 시행과 흡사하나 비복원의 경우이다.

6.2 초기하분포의 확률함수

$$P(N_1 \text{ 중 } x_1,\ N_2 \text{ 중 } x_2) = \frac{{}_{N_1}C_{x_1} \times {}_{N_2}C_{x_2}}{{}_{N_1+N_2}C_{x_1+x_2}}$$

Example 총 20개의 제품 중 불량품은 5. 4개를 선택했을 때 2개가 불량일 확률은?

비복원 경우: $N_1 = 15,\ N_2 = 5,\ x_1 = 2,\ x_2 = 2.$

$P(X=2) = \frac{{}_{15}C_2 \times {}_5C_2}{{}_{20}C_4} = 0.2167$

복원의 경우: n=4, x=2, p=0.25 (since $p = \frac{5}{20}$)

$P(X=2) = {}_4C_2 \times 0.25^2 \times (1-0.25)^2$ = 0.2109

연습문제 정답

For n large and p small

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

$$p(x) = \frac{n(n-1)\cdots(n-x+1)}{n^x}\ \frac{\lambda^x}{x!}\ \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^x}\left(1-\frac{\lambda}{n}\right)^x = \frac{e^{-\lambda}\cdot\lambda^x}{x!}$$

1. (1) 0.2 (2) $F(4) + P(8) = 0.5$ (3) $F(12) - F(8) = 0.5$ (4) F, 감소하지 않는다.

2. (1) 3 (2) -600

(3)

i	X_i	$P(X_i)$
1	이익금의 첫 번째 경우의 값	이익금의 첫 번째 경우의 값이 발생할 확률
2	이익금의 두 번째 경우의 값	이익금의 두 번째 경우의 값이 발생할 확률
3	이익금의 세 번째 경우의 값	이익금의 세 번째 경우의 값이 발생할 확률

(4)

i	X_i	$P(X_i)$
1	X_1	P(X_1)
2	X_2	P(X_2)
3	X_3	P(X_3)

(5)

i	X_i	$P(X_i)$
1	-600	0.40
2	200	0.25
3	800	0.35

(6) 400번 (1,000×0.4) (7) 350,000번 (1,000,000×0.35)

(8) -600(0.4)+200(0.25)+800(0.35)

3. (1)

i	X_i	$P(X_i)$
1	-400	0.4
2	100	0.3
3	600	0.3
합계		1.0

(2) E(X) = -400(0.4) + 100(0.3) + 600(0.3) = 50

(3) $\mathrm{E}(10\mathrm{X} + 30) = 10E(X) + 30 = 10(50) + 30 = 530$

4. (1)

	X_i	$P(X_i)$	$X_i \cdot P(X_i)$
1	-200	0.2	-40

2	100	0.5	50
3	300	0.3	90
합계		1.0	100 = $E(X)$

(2)

X	$P(X)$	$\{X-E(X)\}^2$	$\{X-E(X)\}^2 \cdot P(X)$
-200	0.2	$(-200-100)^2$ = 90,000	18,000
100	0.5	$(100-100)^2$ = 0	0
300	0.3	$(300-100)^2$ = 40,000	12,000
합계	1.0		30,000 = $V(X)$

표준편차 = $\sqrt{V(X)}$ = $\sqrt{30{,}000}$ = 173.21

(3) 각각 2십만개, 5십만개, 3십만개,

(4) $\mu = \frac{-200 \cdot 200{,}000 + 100 \cdot 500{,}000 + 300 \cdot 300{,}000}{1{,}000{,}000}$ = 100

(5) 가능한 편차제곱은 3가지이다. $(-200-100)^2$, $(100-100)^2$, $(300-100)^2$

이들은 각각 각각 2십만개, 5십만개, 3십만개씩 존재한다.

$\sigma^2 = \frac{(-200-100)^2 \cdot 200{,}000 + (100-100)^2 \cdot 500{,}000 + (300-100)^2 \cdot 300{,}000}{1{,}000{,}000}$

= 30,000

5. (1) $E(X)$ = 100, $V(X)$ = 30,000

(2) $E(10X)$ = $10E(X)$ = 1,000, $V(10X)$ = 10^2 $V(X)$ = 3,000,000

$10X$	$P(10X)$	$\{10X-E(10X)\}^2$	$\{10X-E(10X)\} \cdot P(X)$
-2,000	0.2	$(-2{,}000-1{,}000)^2$ = 9,000,000	1,800,000
1,000	0.5	$(1{,}000-1{,}000)^2$ = 0	0
3,000	0.3	$(3{,}000-1{,}000)^2$ = 4,000,000	1,200,000
합계	1.0		3,000,000

(3) $E(X+50)$ = $E(X)$ + 50 = 150, $V(X+50)$ = $V(X)$ = 30,000

X+50	$P(X+50)$	$\{X+50-E(X+50)\}^2$	$\{X-E(X)\}^2 \cdot P(X)$
-150	0.2	$(-150-150)^2$ = 90,000	18,000
150	0.5	$(150-150)^2$ = 0	0
350	0.3	$(350-150)^2$ = 40,000	12,000
합계	1.0		30,000

(4) $E(10X+50)$ = $E(10X)+50$ = 1,050, $V(10X+50)$ = $10^2V(X)$ = 3,000,000

$10X$+50	$P(10X+50)$	$\{10X+50-E(10X+50)\}^2$	$\{X-E(X)\}^2 \cdot P(X)$
-1,950	0.2	$(-1{,}950-1{,}050)^2$ = 9,000,000	1,800,000
1,050	0.5	$(1{,}050-1{,}050)^2$ = 0	0
3,050	0.3	$(3{,}050-1{,}050)^2$ = 4,000,000	1,200,000
합계	1.0		3,000,000

6. (1) SSFF, SFSF, SFFS, FSSF, FSFS, FFSS (총 ${}_4C_2 = \frac{4!}{2!(4-2)!} = 6$)

   (2) P(S,S,F,F) = $p\ p(1\text{-}p)(1\text{-}p) = p^2(1-p)^2$

   P(S,F,S,F) = $p(1\text{-}p)p(1\text{-}p) = p^2(1-p)^2$

   ....

   P(F,F,S,S) = $(1\text{-}p)(1\text{-}p)pp = p^2(1-p)^2$

   (3) ${}_4C_2 \times p^2(1-p)^2$

   (4) ${}_nC_x \times p^x(1-p)^{n-x}$

   Note. 여기서 $p^x(1-p)^{n-x}$는 성공확률$^{\text{성공횟수}}$ 실패확률$^{\text{실패횟수}}$이다.

7. (1) 성공횟수 x (시행 횟수 n은 조건이다.)

   (2)

	빈도					상대빈도				
성공횟수	1	10	100	1,000	10,000	1	10	100	1000	10000
x=0	1	4	36	344	3,476	1.0	0.4	0.36	0.344	0.3476
x=1	-	4	41	436	4,384	-	0.4	0.41	0.436	0.4384
x=2	-	1	20	197	1,873	-	0.1	0.2	0.197	0.1873
x=3	-	1	3	23	267	-	0.1	0.03	0.023	0.0267
합계	1	10	100	1,000	10,000	1.000	1.000	1.000	1.000	1.000

   (3) 평균 $= \frac{0(3{,}476)+1(4{,}384)+2(1{,}873)+3(267)}{3{,}476+4{,}384+1{,}873+267} = \frac{8{,}931}{10{,}000} = 0.8931$

   분산 $= \frac{(0-0.8931)^2\cdot 3{,}476+(1-0.8931)^2\cdot 4{,}384+(2-0.8931)^2\cdot 1{,}873+(3-0.8931)^2\cdot 267}{3{,}476+4{,}384+1{,}873+267}$

   $= \frac{6{,}302.724}{10{,}000} = 0.6303$

   표준편차 $= \sqrt{0.6303} = 0.7939$

   (4) 평균 = 0(0.3476) + 1(0.4384) + 2(0.1873) + 3(0.0267) = 0.8931

   분산 $= (0-0.8931)^2 \cdot 0.3476 + (1-0.8931)^2 \cdot 0.4384$

   $+(2-0.8931)^2 \cdot 0.1873 + (3-0.8931)^2 \cdot 0.0267$

$= 0.6303$

표준편차 $= \sqrt{0.6303} = 0.7939$

(5) $E(X) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \cdots + x_4 \cdot P(x_4) = \sum_{i=1}^{4} x_i \cdot P(x_i)$

$V(X) = \left(x_1 - E(X)\right)^2 \cdot P(x_1) + \left(x_2 - E(X)\right)^2 \cdot P(x_2) + \cdots + \left(x_4 - E(X)\right)^2 \cdot P(x_4)$

$= \sum_{i=1}^{4} \left(x_i - E(X)\right)^2 \cdot P(x_i)$

$\sigma = \sqrt{V(X)}$

8. (1)

X(성공횟수)	${}_nC_x$	$p^x(1-p)^{n-x}$	${}_nC_x\, p^x(1-p)^{n-x}$
0	1	0.3430	0.3430
1	3	0.1470	0.4410
2	3	0.0630	0.1890
3	1	0.0270	0.0270
합계			1.0000

(2) 확률의 정의 중 상대빈도정의에 의하면 실험의 횟수가 ∞에 이를 때 상대빈도값은 확률에 일치한다. 연습문제 7에서의 실험횟수는 ∞가 아닌 10,000이기 때문에 오차가 발생한다.

(3)

X(성공횟수)	${}_nC_x$	$p^x(1-p)^{n-x}$	${}_nC_x\, p^x(1-p)^{n-x}$
0	1	0.0024	0.0024
1	5	0.0057	0.0284
2	10	0.0132	0.1323
3	10	0.0309	0.3087
4	5	0.0720	0.3602
5	1	0.1681	0.1681
합계			1.0000

9. (1) $\mu = E(X) = np = 3 \times 0.3 = 0.9$ 연습문제 7에서는 0.8931

(2) $V(X) = \sigma^2 = npq = 3 \times 0.3 \times 0.7 = 0.63$ 연습문제 7에서는 0.6303

10. (1)

x	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
4	0.18881
8	0.01687

(2)

x	$P(x)$	$x \cdot P(x)$	$\left(x-E(X)\right)^2$	$\left(x-E(X)\right)^2 \cdot P(x)$
3	0.2158	0.6474	0.2500	0.0539
6	0.0771	0.4626	6.2501	0.4819
9	0.0066	0.0590	30.2502	0.1984
12	0.0002	0.0026	72.2503	0.0154

11. (1) $E(X) = \sum_{x=0}^{\infty} x \cdot P(x) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda}\lambda^x}{x!}$
(이 값을 수학적으로 정리하면 λ가 된다.)

(2) $V(X) = \sum_{x=0}^{\infty} \{x - E(x)\}^2 \cdot P(x) = \sum_{x=0}^{\infty} \{x - E(x)\}^2 \cdot \frac{e^{-\lambda}\lambda^x}{x!}$
(이 값을 수학적으로 정리하면 λ가 된다.)

12. (1) 3.5 (2) 3.5