

5장 이산 확률 분포 (두 번째)

P.252

• 확률 분포와 선형변환

예) 슬롯머신의 배팅과 상금이 아래와 같이 변경된다.

2007년

- 게임당 ~~1000원~~ 지불
- 아래의 사건이 발생하는 경우 상금 받음.

$$\begin{array}{l} \textcircled{1} (\$,\$,\$) : \cancel{=0,000원} \quad 100,000원 \\ \textcircled{2} \$ * 2회, \$ * 1회 : \cancel{15,000원} \quad 75,000원 \\ \textcircled{3} (\$, \$, \$) : \cancel{10,000원} \quad 50,000원 \\ \textcircled{4} (0, 0, 0) : \cancel{5,000원} \quad 25,000원 \end{array}$$

- 나머지 경우 : 광
- 각각의 경우에서 특정 그림에서 벌금 확률

- \\$: 0.1
- \\$: 0.2
- \\$: 0.2
- 기타 : 0.5

← 연습지 단답문.

- 게임비용과 상금이 변하였지만 사건의 종류와 확률은 변하지 않았다. 따라서 기존의 확률분포와 비슷하지만 새로운 확률변수 Y 가 가리키는 그룹들의 진동성이�하였다.

→ 새로운 확률분포인
(새로운 확률변수)
사용

y	98,000	73,000	48,000	23,000	-2,000
$P(Y=y)$	0.001	0.006	0.008	0.008	0.977

- 학률 변수 Y 에 대한 기대치와 표본은 공식으로 적용
계산하면 다음과 같다.

$$\begin{aligned}
 E(Y) &= 98,000 * 0.001 + 73,000 * 0.006 \\
 &\quad + 48,000 * 0.008 + 23,000 * 0.009 \\
 &\quad + (-2,000) * 0.977 \\
 &= -850
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(Y) &= E(Y - \mu)^2 \quad \leftarrow \mu = E(Y) \\
 &= (98,000 + 850)^2 * 0.001 + (73,000 + 850)^2 * 0.006 \\
 &\quad + (48,000 + 850)^2 * 0.008 + (23,000 + 850)^2 * 0.009 \\
 &\quad + (-2,000 + 850)^2 * 0.977 \\
 &= 67,427,500
 \end{aligned}$$

$$\sigma = \sqrt{\text{Var}(Y)} = 8,211.4$$

질문: 즉, 기액을 1회 할 때마다 학률적으로 성장률구는

$$[-850 - 8,211.4, -850 + 8,211.4]$$

사이의 확률을 알게 된다.

하지만 현실적으로는 2,000 원을 뺏거나 차소
23,000 원을 따야 한다. 따라서 기액을 1회
할 때마다 성장률구 2,000 원을 뺏는다는
질문이 나온다.

- 질문: 하지만 이렇게 매번 새로 계산해야 하나?
질의에 따라 계산이 매우 비싼 비용을 요구한다면
어떻게 하나?
- 대답: 선행변환인 경우 간단한 해결책 존재!

p. 255

두 확률변수 사이의 선형변환 관계 확인하기

- 확률변수 X 와 Y 의 관계 확인하기

* X : 상급액 변동에 이전에 버는 액수를 가지는 확률변수

* Y : " " " 이전에 " " " "

* X 와 Y 사이의 관계를 확인해야 한다.

이를 위해 당첨금과 X , Y 사이의 관계를 확인해 보면 된다.

- 어떤 사건이 빨리 생겼을 때 원래 당첨금이 P 이면 아래 식이 성립한다.

$$X = P - 1,000 \quad (\text{비율 } 1,000\%)$$

$$Y = 5P - 2,000 \quad (\text{상금 5배, 비용 } 2,000\%)$$

따라서 $Y = 5(X + 1,000) - 2,000 = 5X + 3,000$

아래 식이 성립한다.

$$E(Y) = E(5X + 3,000)$$

$$\text{Var}(Y) = \text{Var}(5X + 3,000)$$

- 질문: a, b 가 실수일 때, $E(a*X + b)$ 와 $\text{Var}(a*X + b)$ 를 $E(X)$ 와 $\text{Var}(Y)$ 를 통해 유도하여 계산할 수 있는가?

- 대답: 그렇다. 아래 식이 성립한다.

$$E(a*X + b) = a * E(X) + b$$

$$\text{Var}(a*X + b) = a^2 * \text{Var}(X)$$

p. 259

9. 25일

- 따라서 상점이 번영한 속도마진에서 1회당 벌 수 있는 금액의 기대치와 총판은 판매과 같이 쉽게 계산할 수 있다.

$$\begin{aligned}E(Y) &= 5 * E(X) + 3,000 \\&= 5 * (-770) + 3,000 \\&= -3,850 + 3,000 \\&= -850\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= 5 * \text{Var}(X) \\&= 25 * 2,697,100 \\&= 67,427,500\end{aligned}$$

이전에 계산한
값과 동일함.

- 지금까지 선형변환을 이용하여 하나의 확률분포에서 새로운 확률분포를 구하였다.

여기부터는 두 개 이상의 확률분포를 이용하여 새로운 확률분포를 구하고 각각의 발생률을 살펴보자 한다.

즉, 서로 다른 확률분포의 기대치와 총판을 쉽게 계산하는 방법을 배운다.

P. 261

선형변환과 확률변수의 차이

- **선형변환** : 슬롯머신 게임 예제에서 사건의 종류와 각 사건의 발생 확률이 변하지 않았지만 각 사건이 발생했을 때의 영향(예: 베는 액수) 만이 선형변환에 따라 달라졌을 때를 가리킨다. 서로 영향을 주지 않겠다는 의미
- **확률변수** : 예를 들어, 두 대의 **독립적인** 슬롯머신을 동시에 실행한 결과를 대상으로 확률분포를 확인하면 한 대로 게임을 진행할 때와 완전히 다른 결과를 얻는다. 즉, 사건의 종류와 각 사건의 영향이 선형변환과는 다른 방식으로 달라진다.

예) 1~10 사이의 숫자가 표기된 10개의 구슬이 들어 있는 상자에서 한 개의 구슬을 뽑았을 때 10이 나오면 5000 원을 벌고, 아니면 1000 원을 얻는 게임을 한다고assumption 하자.
구슬 뽑기 게임
이라assumption 으로자.
그러면 게임을 한 번 할 때 베는 금액을 대상으로 하는 확률분포는 다음과 같은 식으로 선형변환의 경우와 동일한 조망과 내용물을 가진 두 개의 상자에서 동시에 구슬을 하나씩 뽑았을 때의 경우 확률분포의 변화는 아래와 같다.

X	-1,000	5,000
$P(X=x)$	0.9	0.1

$2X$	2,000	10,000
$P(2X=z)$	0.9	0.1

Z	-2,000	4,000	10,000
$P(Z=z)$	0.81	0.18	0.01

↑
한 번만 뽑았을 때

↓
숫자만 두 번째!

↑
한 번만 뽑았을 때
숫자만 두 번째!

↓
숫자가 완전히 달라짐!

↓
한 번만 뽑았을 때

↑
한 번만 뽑았을 때
한 번만 뽑았을 때
한 번만 뽑았을 때
(0.9 * 0.9) (0.9 * 0.1 + 0.1 * 0.9) (0.1 * 0.1)

• 독립관측

- 먼저 아래 두 경우는 동일한 확률분포를 갖는다는 사실을 이해해야 한다.

경우 ① : 두 대의 슬롯머신을 대상으로 개인을 진행한 결과에 대한 확률분포

경우 ② : 한 대의 슬롯머신에서 개인을 두 번 진행한 결과에 대한 확률분포

• 사건과 관측

- 예를 들어, 하나의 슬롯머신에서 여러 번 개인을 할 때, 각 개인을 **사건**이라고 하며, 당첨금액처럼 각 개인의 결과값은 **관측**이라고 한다.

- 그런데 슬롯머신의 경우, 각 개인의 관측이 동일한 기대치와 분산을 갖는 한편에 개인을 할 때마다의 결과값은 독립적으로 달라진다.

이는 경우 ②가 경우 ①과 동일한 확률분포를 갖는다는 설명과 동일하게 이해할 수 있다.

- 따라서, 동일한 기대치와 분산을 갖는 사건, 예를 들어 슬롯게임을 한 번씩으로 실행한다 해도 각각의 사건을 구별해야 한다.

• 예제

- 구슬 뽑기 게임을 두 번 진행할 때의 결과에 대한 확률분포를 만드는 과정은 아래 그림과 같다.

복제할 때마다
이론 확률변수 사용

X	-1,000	5,000
$P(X=x)$	0.9	0.1

복제

복제

X_1	-1,000	5,000
$P(X_1=x_1)$	0.9	0.1

X_2	-1,000	5,000
$P(X_2=x_2)$	0.9	0.1

X_1 과 X_2 는
 X 의 독립변수
이라 부른다.

Z	-2,000	4,000	10,000
$P(Z=z)$	0.81	0.18	0.01

독립변수
2회

- 이제 세 확률변수 X_1, X_2, X_3 사이의 관계를 유도하면 다음과 같다.

$$Z = X_1 + X_2$$

- 또한 아래 공식이 성립한다.

$$\begin{aligned} E(Z) &= E(X_1 + X_2) = E(X_1) + E(X_2) \quad (= E(X) + E(X)) \\ &= 2E(X) \end{aligned}$$

복제 결과

$$\begin{aligned} \text{Var}(Z) &= \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) \quad (= \text{Var}(X) + \text{Var}(X)) \\ &= 2\text{Var}(X) \end{aligned}$$

*주의: 독립변수의 확률분포는 기준의 확률분포와 동일하다.
따라서 기대치와 분산도 동일하다.

- 주의: $E(X_1) = E(X_2) = E(X)$ 가 성립한다 하더라도 $X_1 + X_2$ 와 $2X$ 가 서로 다르다.

- $2X$: 관측값만 두 배로 커진 새로운 확률변수

- $X_1 + X_2$: 두 개의 확률변수가 결합하여 일어진 서로운 확률변수

P.263

- 독립관측의 덧셈

- 앞서 배운 내용을 일반화 하면 다음과 같다.

- 먼저 X_1, X_2, \dots, X_n 이 X 의 독립관측이라 하자.

예를들어, 주는 봉기개인을 1번 실행 경우를 생각해보면 된다.

그러면 다음 공식이 성립한다.

$$Z = X_1 + X_2 + \dots + X_n$$

$$E(Z) = E(X_1 + \dots + X_n) = n \cdot E(X)$$

$$\text{Var}(Z) = \text{Var}(X_1 + \dots + X_n) = n \cdot \text{Var}(X)$$

7. 268 ~ 270

- 확률변수의 덧셈

- 독립관측라는 말은 딸리 이제부터는 서로 상관이 전혀 없으며 서로 독립인 두 개의 확률변수 X 와 Y 의 합을 다루고자 한다.

- 먼저: ① $E(X)$ 와 $E(Y)$ 가 서로 상관 없음을

② $\text{Var}(X)$ 와 $\text{Var}(Y)$ 가 서로 상관 없음을

- 하지만 이 경우에도 아래 공식이 성립한다.

$$E(X+Y) = E(X) + E(Y)$$

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

연습

서로 다른 규칙과 상관을 갖는 두 대의 슬롯머신을
각각 한 번씩 실행했을 때 알 수 있는 금액의
기대치와 분산을 구하고자 한다.

- 가정: 각 슬롯머신의 확률분포는 다음과 같다.

X	-5,000	39,3000
$P(X=x)$	0.99	0.01

Y	-2,000	23,000	48,000	73,000	98,000
$P(Y=y)$	0.977	0.008	0.008	0.006	0.001

- 원래는 $X+Y$ 의 확률분포를 아래 표 모양으로 복잡하게 계산해서 구해야 한다.

$X+Y$	-7,000	18,000	...	468,000	493,000
$P(X+Y=x+y)$	0.96723	0.00792	...	0.00006	0.00001

10개의 경우 존재

이후에 기대값 $E(X+Y)$ 와 표준분산 $Var(X+Y)$ 를 공식에 따라 직접 계산해야 한다.

- 하지만 앞서 가본 공식을 이용하면 매우 쉽게 기대치와 표준분산을 계산할 수 있다.

• 확률변수의 뺄셈

서로 독립!

- 두 확률변수 X 와 Y 의 차이를 다룰 때 $X - Y$ 라는 확률변수를 사용함.
- X 와 Y 의 확률분포가 주어졌을 때, $X - Y$ 의 확률분포는 $X + Y$ 의 경우와 유사하게 구해진다. 다만 두 확률변수가 갖는 합의 차이를 다루는 것만 다를 뿐이다.

$X - Y$...		
$P(X - Y = x - y)$...		

예제) A와 B 두 개의 리스토랑에서 제공하는 메뉴의 가격과 선택비율에 대한 확률분포가 아래와 같다.

A 레스토랑	X	20,000	30,000	40,000	45,000
	$P(X=x)$	0.3	0.4	0.2	0.1

B 레스토랑	y	10,000	15,000	18,000
	$P(Y=y)$	0.2	0.6	0.2

A,B 레스토랑에서 손님이 하나의 페부를 선택할 때 지불하는 비용의 차이를 대상으로 하는 확률분포는 $X - Y$ 로 나타내며, 아래 모양을 갖는다.

$X - Y$	2,000	5,000	...	30,000	35,000
$P(X - Y = x - y)$	0.06	0.18	...	0.06	0.02

하지만 $X - Y$ 에 대한 확률분포표를 새로 작성하지 않아도 $X - Y$ 의 기대치와 분산을 쉽게 구할 수 있다. 아래 공식이 성립하기 때문이다.

$$E(X - Y) = E(X) - E(Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

☞ 주의!

• 주의: 두 확률변수의 차이의 분산을 구할 때 빼지 않고 더한다.

이유: 하나의 변수에서 다른 변수를 뺀더라도 확률분포의 분산은 증가한다. 위 레스토랑 예제에서 이 현상을 확인할 수 있다.

A 레스토랑: 가격이 20,000 원에서 45,000 원 사이.
즉, 최대 25,000 원 차이.

B 레스토랑: 가격이 10,000 원에서 18,000 원 사이.
즉, 최대 8,000 원 차이.

반면에 $X - Y$ 의 값은 2,000 원에서 35,000 원 사이.
즉, 최대 33,000 원 차이.

그리고 $25,000 + 8,000 = 33,000$ 이다!

P. 272

• 최종정리: 선형변환과 확률분포의 덧셈과 뺄셈.

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

$$\text{Var}(aX + bY) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y)$$

$$E(aX - bY) = a \cdot E(X) - b \cdot E(Y)$$

$$\text{Var}(aX - bY) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y)$$