5장 이산 확률 분포 (두번째)

p.252

- **확률 분포와 선형변환**

예) 슬롯머신의 비용과 상금이 아래와 같이 변경됨.

- 게임당 ~~1000원~~ 2000원 지불
- 아래 네 사건이 발생하는 경우 상금 받음.
  - ① ($, $, $) : ~~20,000원~~ 100,000원
  - ② $ * 2회, 🍒 * 1회 : ~~15,000원~~ 75,000원
  - ③ (🍒, 🍒, 🍒) : ~~10,000원~~ 50,000원
  - ④ (O, O, O) : ~~5,000원~~ 25,000원
- 나머지 경우 : 꽝
- 각각의 창에서 특정 그림에서 멈출 확률
  - $: 0.1
  - 🍒 : 0.2
  - O : 0.2
  - 기타 : 0.5

← 변하지 않음.

- 게임 비용과 상금이 변하였지만 사건의 종류와 확률은 변하지 않았다. 따라서 기존의 확률분포와 비슷하지만 새로운 확률 변수 Y가 가리키는 값들의 집합이 변했다.

→ 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

- 확률변수 Y에 대한 기대치와 분산을 공식으로 직접 계산하면 다음과 같다.

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

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

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

결론: 즉, 게임을 1회 할 때마다 확률적으로 십중팔구는

$$[-850 - 8{,}211.4,\ -850 + 8{,}211.4]$$

사이의 값을 받게 된다.

하지만 현실적으로는 2,000원을 잃거나 최소 23,000원을 따야 한다. 따라서 게임을 1회 할 때마다 십중팔구 2,000원을 잃는다는 결론이 나온다.

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

p.255

• 두 확률변수 사이의 선형변환 관계 확인하기

- 확률변수 X와 Y의 관계 확인하기
  * X : 상금액 변경 이전에 버는 액수를 가리키는 확률변수
  * Y : " " 이후에 " " " "

  * X와 Y 사이의 관계를 확인해야 한다.
  이를 위해 당첨금과 X, Y 사이의 관계를 확인해 보면 된다.

- 어떤 사건이 발생했을 때 원래 당첨금이 p 이면 아래 관계식이 성립한다.

$X = p - 1{,}000$ (비용이 1,000원)

$Y = 5*p - 2{,}000$ (상금 5배, 비용 2,000원)

따라서 $Y = 5(X + 1{,}000) - 2{,}000 = 5 \cdot X + 3{,}000$

이제 아래 식이 성립한다.

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

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

p.259

- 질문: a, b가 실수일 때, $E(a*X + b)$와 $Var(a*X + b)$를 $E(X)$와 $Var(Y)$로부터 유도하여 계산할 수 있는가?

- 대답: 그렇다. 아래 식이 성립한다.

$$\boxed{\begin{aligned} E(a*X + b) &= a*E(X) + b \\ Var(a*X + b) &= a^2 * Var(X) \end{aligned}}$$

p.258

- 따라서 상금이 변경된 슬롯머신에서 1회당 벌 수 있는 금액의 기대치와 분산은 다음과 같이 쉽게 계산할 수 있다.

$$E(Y) = 5 * E(X) + 3{,}000$$
$$= 5 * (-770) + 3{,}000$$
$$= -3{,}850 + 3{,}000$$
$$= -850$$

$$Var(Y) = 5^2 * Var(X)$$
$$= 25 * 2{,}697{,}100$$
$$= 67{,}427{,}500$$

이전에 계산한 값과 동일함.

• 지금까지 선형변환을 이용하여 하나의 확률분포에서 새로운 확률분포를 구하였다.

이제부터는 두 개 이상의 확률분포를 이용하여 새로운 확률분포를 구하고 다루는 방법을 살펴보고자 한다.
즉, 새로 구한 확률분포의 기대치와 분산을 쉽게 계산하는 방법을 배운다.

p. 261

• 선형변환과 독립관측의 차이

- 선형변환 : 슬롯머신 게임 예제에서 보았듯이 사건의 종류와 각 사건의 발생확률이 변하지 않았지만 각 사건이 발생했을 때의 영향(여기서는 버는액수)만이 선형관계를 따라 달라졌을 때를 가리킴.

- 독립관측 : 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 확률이 선형변환과는 다른 방식으로 얻어진다.

서로 영향을 주지 않는다는 의미

예) 1~10 사이의 숫자가 표기된 10개의 구슬이 들어있는 상자에서 한 개의 구슬을 뽑았을 때 10이 나오면 5000원을 벌고, 아니면 1000원을 잃는 게임을 한다고 가정하자.

"구슬뽑기 게임" 이라 부르자.

그러면 게임을 한 번 할 때 버는 금액을 대상으로 하는 확률분포는 다음과 같으며 상금을 두 배로 하는 선형변환의 경우와 동일한 모양과 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

선형변환

두 번의 독립관측

숫자만 두 배로!

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

확률 그대로!

숫자가 완전히 달라짐.

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

둘 다 잃은 경우 $(0.9*0.9)$

한 번 따고 한 번 잃은 경우 $(0.9*0.1*2)$

둘 다 딴 경우 $(0.1*0.1)$

P.262

- 독립관측

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

경우 ①: 두 대의 동일한 슬롯머신을 대상으로 게임을 진행한 결과에 대한 확률분포

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

- 사건과 관측

- 예를 들어, 하나의 슬롯머신에서 여러 번 게임을 할 때, 각 게임을 사건이라고 하며, 당첨금액처럼 각 게임의 결과값을 관측이라고 한다.

- 그런데 슬롯머신의 경우, 각 게임의 관측이 동일한 기대치와 분산을 갖는 반면에 게임을 할 때마다의 결과값은 독립적으로 달라진다.

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

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

- 예제

- 구슬뽑기 게임을 두 번 진행할 때의 결과에 대한 확률분포를 얻는 과정은 아래 그림과 같다.

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

복제할 때마다 다른 확률변수 사용

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_1과 X_2는 X의 독립관측이라 부른다.

독립관측 2회

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

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

$$Z = X_1 + X_2$$

- 또한 아래 공식이 성립한다.

$$E(Z) = E(X_1+X_2) = E(X_1) + E(X_2) = E(X) + E(X) = 2E(X)$$

복제 결과

$$Var(Z) = Var(X_1+X_2) = Var(X_1) + Var(X_2) = Var(X) + Var(X) = 2Var(X)$$

* 주의 : 독립관측의 확률분포는 기존의 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 얻어진 새로운 확률변수

p.263

- **독립관측의 덧셈**
  - 앞서 배운 내용을 일반화 하면 다음과 같다.
  - 먼저 $X_1, X_2, \cdots, X_n$ 이 X의 독립관측이라 하자.
    예를들어, 구슬뽑기게임을 n번 실행 경우를 생각하면 된다.
    그러면 다음 공식이 성립한다.

$$\boxed{\begin{aligned} Z &= X_1 + X_2 + \cdots + X_n \\ E(Z) &= E(X_1 + \cdots + X_n) = n \cdot E(X) \\ Var(Z) &= Var(X_1 + \cdots + X_n) = n \cdot Var(X) \end{aligned}}$$

p. 263 ~ 270

- **확률변수의 덧셈**
  - 독립관측과는 달리 이제부터는 서로 상관이 전혀 없으며 서로 독립인 두 개의 확률변수 X와 Y의 합을 다루고자 한다.
  - 전제 : ① $E(X)$와 $E(Y)$ 가 서로 상관 없음
    ② $Var(X)$와 $Var(Y)$가 서로 상관 없음
  - 하지만 이 경우에도 아래 공식이 성립한다.

$$\boxed{\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ Var(X+Y) &= Var(X) + Var(Y) \end{aligned}}$$

p.269

예제

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

- 가정 : 각 슬롯머신의 확률분포는 다음과 같다.

x	-5,000	395,000
$P(X=x)$	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	⋯	0.0006	0.0001

10개의 경우 존재

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

- 하지만 앞서 다룬 공식을 이용하면 매우 쉽게 기대치와 분산을 계산할 수 있다.

p.271

- 확률변수의 뺄셈 서로 독립!

  - 두 확률변수 X와 Y의 차이를 다룰 때 X−Y라는 확률분포를 사용함.
  - X와 Y의 확률분포가 주어졌을 때, X−Y의 확률분포는 X+Y의 경우와 유사하게 구해진다. 다만 두 확률변수가 갖는 값의 차이를 다룬다는 점만 다를 뿐이다.

$x-y$			∘∘∘		
$P(X-Y=x-y)$			∘∘∘		

p.273

예제) A와 B 두 개의 레스토랑에서 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 메뉴를 선택할 때 지불하는 비용의 차이를 대상으로 하는 확률분포는 X−Y로 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−Y에 대한 확률분포표를 새로 작성하지 않아도 X−Y의 기대치와 분산은 쉽게 구할 수 있다. 아래 공식이 성립하기 때문이다.

$$\boxed{\begin{aligned} E(X-Y) &= E(X)-E(Y) \\ Var(X-Y) &= Var(X)+Var(Y) \end{aligned}}$$

⇐ 주의!

• 주의: 두 확률변수의 차이의 분산을 구할 때 빼지 않고 더한다.

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

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

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

반면에 X-Y의 값은 2,000원에서 35,000원 사이.
즉, 최대 33,000원 차이.
그리고 25,000 + 8,000 = 33,000 이다!

P.272

• 최종정리: 선형변환과 확률분포의 덧셈과 뺄셈.

$$E(aX+b) = aE(X)+b$$
$$Var(aX+b) = a^2 Var(X) + b^2 Var(Y)$$
$$E(aX-b) = aE(X)-b$$
$$Var(aX-b) = a^2 Var(X) + b^2 Var(Y)$$