

5 이산확률분포

2017년 10월 15일 일요일 오전 11:35

주제: 이산 확률분포

어떤 사건이 발생할 확률뿐만 아니라 영향에 대해서도 다룬다.

이산이란?

- 다른 데이터가 1, 2, 3, 4 ...처럼 구운할 수 있는 경우 이산 데이터라고 부른다

예제: 카지노의 슬롯머신에서 돈을 얼마나 벌 수 있는지를 판단할 때 사용

슬롯머신



작업



예제 1 (9. 238)

제일 양방

제일양 1달러 지출

달러, 달러, 달러 = 20 달러

달러 * 2, 체리 = 15 달러

체리, 체리, 체리 = 10 달러

레몬, 레몬, 레몬 = 5 달러

=====

특정한 그림이 특정한 장에 나타날 확률:

\$	체리	레몬	기타
0.1	0.2	0.2	0.5

전체: 세 개의 장에서 어떤 그림이 나타나는가는
서로 독립이다.

즉, 어느 하나의 장에 어떤 그림이 나타나는가는
다른 장에 어떠한 영향도 주지 않는다.

(P. 239)

① 달려, 달려, 달려의 확률:

$$\begin{aligned} P(\text{달려}, \text{달려}, \text{달려}) &= P(\text{달려}) * P(\text{달려}) * P(\text{달려}) \\ &= 0.1 * 0.1 * 0.1 \\ &= 0.001 \end{aligned}$$

② 달려, 달려, 체리의 확률 (일의 순서):

$$\begin{aligned} P(\text{달려}, \text{달려}, \text{체리}) + P(\text{달려}, \text{체리}, \text{달려}) \\ + P(\text{체리}, \text{달려}, \text{달려}) \\ = (0.1 * 0.1 * 0.2) * 3 \\ = 0.006 \end{aligned}$$

③ 레몬, 레몬, 레몬의 확률:

$$\begin{aligned} P(\text{레몬}, \text{레몬}, \text{레몬}) &= 0.2 * 0.2 * 0.2 \\ &= 0.008 \end{aligned}$$

④ 체리, 체리, 체리의 확률:

$$\begin{aligned} P(\text{체리}, \text{체리}, \text{체리}) &= 0.2 * 0.2 * 0.2 \\ &= 0.008 \end{aligned}$$

P. 241

슬롯 머신에 대한 확률 흐름

① 사건들의 확률

조합	그랑	레몬	체리	달려/체리	달려
확률	0.977	0.008	0.008	0.006	0.001

② 슬롯머신에서 빨 수 있는 금액에 대한 확률 분포

조합	꽝	레몬	체리	달러/체리	달러
파는 금액	-1	4	9	14	19
확률	0.977	0.008	0.008	0.006	0.001

==

확률 변수와 확률 분포 (P. 242)

* 확률 변수 : 특정한 확률과 연관되어 있는
값들에 대한 변수

예) 슬롯머신에서 확률 변수는
파는 금액 중 하나의 값을 가질 수
있다.

* 일반적으로 X, Y 등을 이용하여
확률 변수를 나타낸다.

그리고, 확률 변수 X가 x의 값을
가질 확률을 $P(X=x)$ 로 표시함.

==

이제 확률 변수를 이용하여 슬롯머신의
확률 분포를 아래와 같이 나타낼 수 있다.

조합	꽝	터운	체리	달러/체리	달러
α	-1	4	9	14	19
$P(X=x)$	0.977	0.008	0.008	0.006	0.001

기대치 계산 (P. 244)

기대치 : 기대되는 값.

예제) 슬롯머신의 경우, 1회 시행할 때마다
얼마를 벌게 될지 기대하는 값이

기대치이다.

기대치 공식 :

$$E(X) = \sum x \cdot P(X=x)$$

학률변수 X 에
대한 기대치
학률변수가 가질 수 있는 값과
그 값이 발생할 확률을 곱한다.

곱하기가 끝나면
모든 값을 더해준다.

주의: 평균값 구하는 공식과 유사하다.

$$\mu = \frac{\sum x}{n}$$

따라서, 경우에 따라 $E(X)$ 를 μ 로
표시하기도 함.

예제) 슬롯머신에 대한 기대치

$$\begin{aligned}E(X) &= (-1 * 0.977) + (4 * 0.008) \\&\quad + (9 * 0.008) + (14 * 0.001) \\&\quad + (19 * 0.001) \\&= -0.77\end{aligned}$$

즉, 슬롯머신 레버를 한번 당길 때마다
0.77 달러를 잃을 것으로 기대됨.

만약 100번 하면 77달러 (8만원 정도)
잃는 것이 기대됨.

=

확률분포의 운산 (P. 245)

질문: 슬롯머신을 할 때마다 무조건 돈을
잃을까?

대답: 그렇지 않다. 기대치는 말 그대로
기대치일 뿐이며, 항상 그렇다는 의미는
아니다.

따라서 경우에 따라 돈을 딸 수도 있고
훨씬 더 많이 잃을 수도 있다.
그렇다면 딱거나 잃을 수 있는 범위를

측정할 수 있는 방식이 필요하다.

즉, 값들의 분포의 양식이 필요하며,

분산을 이용하면 된다.

학점예상 분산 공식 (P. 246)

$$\text{Var}(X) = E(X - \mu)^2, \text{ 단 } \mu = E(X)$$

= $E(X - \mu)^2$ 계산법 :

$$E(X - \mu)^2 = \sum (x - \mu)^2 * P(X=x)$$

각 x 에 대해 $(x - \mu)$ 의 제곱을 계산.

예제 : 슬롯머신의 분산 (P. 247)

$$\begin{aligned} \text{Var}(X) &= E(X - \mu)^2 \\ &= (-1 + 0.77)^2 * 0.977 \\ &\quad + (4 + 0.77)^2 * 0.008 \\ &\quad + (9 + 0.77)^2 * 0.008 \\ &\quad + (14 + 0.77)^2 * 0.006 \\ &\quad + (19 + 0.77)^2 * 0.001 \\ &= 2.6971 \end{aligned}$$

확률분포의 표준편차 :

$$\sigma = \sqrt{\text{Var}(X)}$$

예제) 슬롯머신의 표준편차

$$\sigma = \sqrt{2.6971} = 1.642$$

따라서, 슬롯머신 레이를 한번 누를 때마다

$$[-0.77 - 1.642, -0.77 + 1.642] \text{ 구간}$$

$$= [-2.412, 0.872]$$

사이의 돈을 땔 수 있다고 기대할 수 있다.

그런데 매회 1달러 이상 잃을 수

없으므로, 땔 수 있는 금액에 대한

기대값은 $[-1, 0.872]$ 구간에서

움직인다고 할 수 있다.

확률분포와 선형변환

예제 (P. 252)

슬롯머신의 비용과 상금이 아래와 같이 설정됨.

① 게임당 2를 지출

② 상금이 5씩 떨어짐. 즉,

↗ 달러, 달러, 달러 : 100 달러
 달러, 달러, 체리(임의는지) : 75 달러
 체리, 체리, 체리 : 50 달러
 레몬, 레몬, 레몬 : 25 달러

=> 새로운 확률분포

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

=> 기대치 :

$$\begin{aligned}
 E(Y) &= (-2 \times 0.977) + (23 \times 0.008) \\
 &\quad + (48 \times 0.008) + (73 \times 0.006) \\
 &\quad + (98 \times 0.001) \\
 &= -0.85 \quad (\leftarrow \text{기대치가 더 낮아짐})
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(Y) &= E(Y-\mu)^2 \\
 &= (-2 + 0.85)^2 \times 0.977 \\
 &\quad + (23 + 0.85)^2 \times 0.008
 \end{aligned}$$

$$\begin{aligned}
 & + (48 + 0.85)^2 * 0.008 \\
 & + (73 + 0.85) * 0.006 \\
 & + (98 + 0.85) * 0.001 \\
 & = 67.4275
 \end{aligned}$$

$\Rightarrow \boxed{\sigma = 8.2114}$

\Leftrightarrow (돈을 딸 수 있는 금액의 변화가)
이 거칠

질문: 하지만 이렇게 매번 새로 계산해야 하나?

답: 아니라. 신형변환은 경우 간단하게
계산할 수 있다.

(P. 255)

X를 금액변경 이전의 확률변수,
Y를 금액변경 이후의 확률변수라 하자.

그러면 아래 관계가 성립한다.

X는 금액변경 이전에 한 번 게임을 할 때마다
딸 수 있는 금액을 대표한다.

그런데 딸 수 있는 금액은 당첨금에서 게임비용을
제외한 액수이다. 따라서 아래 관계가
성립한다.

게임비용

$$X = \text{원래 양절률} - 1$$

동일한 이유로 해서, 금액변경 이전에
한 번 기입을 할 때마다 떨 수 있는
금액을 대표하는 확률변수 Y 에 대해
아래 관계가 성립한다.

이로운
기입비용

$$Y = \text{변경 후 양절률} - 2$$

이제 X 와 Y 사이의 관계를 아래와 같이
推導할 수 있다.

포인트 : ① 변경후 양절률 = $5 * (\text{원래 양절률})$

② 원래 양절률 = $X + 1$

$$\begin{aligned}\Rightarrow Y &= 5 * (\text{원래 양절률}) - 2 \\ &= 5 * (X + 1) - 2 \\ &= 5 * X + 3\end{aligned}$$

따라서 Y 의 기대값과 분산에 대해
다음 관계가 성립한다.

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

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

반면에 아래 공식이 성립한다.

(P. 259)

일반적으로 아래 공식이 성립한다.

$$E(aX+b) = aE(X) + b$$

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

(P. 258)

따라서 변형된 슬롯머신의 기대값과 표준은
자유롭고 같이 구할 수 있다.

$$\begin{aligned} E(Y) &= E(5X+3) = 5 \cdot E(X) + 3 \\ &= 5 \cdot (-0.77) + 3 \\ &= -3.85 + 3 \\ &= -0.85 \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(5Y+3) \\ &= 5^2 \cdot \text{Var}(Y) \end{aligned}$$

$$= 25 * 2.6971$$
$$= 69.4275$$

앞서 선형변환을 이용하여 하나의 확률분포에서 새로운 확률분포를 구하였다.

이제 우리는 두 개 이상의 확률분포를 이용하여 새로운 확률분포를 구한다.

또한 새롭게 구한 확률분포의 기대값과
분산을 찾기 계산하는 방법을 익힌다.

독립관측 (P. 262)

예를 들어, 하나의 슬롯머신에서 여러 번
게임을 할 때, 각 게임은 **사건**이라고
하되, **각 게임의 결과값은 관측**
이라고 한다.

당첨금액처럼

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

✓

그런데 슬롯머신의 경우 각 게임의 관측이 동일한 기대치와 흔성을 갖는 반면에 값은 다를 수 있다.

즉, 게임을 할 때마다 다른 결과가 나오라는 의미이다.

따라서, 동일한 기대치와 흔성을 갖는 사건, 즉 동일한 슬롯기임을 반복해서 한다 해도 각각의 사건을 구분해야 한다.

질서의
예를 들어, 슬롯머신에서 빨록 있는 굽백에 대한 확률분포 X 가 다음과 같이 주어졌다 하자.

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

그리면 두 번 연속 개입을 할 때
 X_1, X_2 처럼 다른 이름을 이용하여

첫 번째 관측의 확률분포를 X_1 ,
두 번째 관측의 확률분포를 X_2 로
불러야 한다.

이럴 때 X_1 과 X_2 는 X 의 독립관측
이라 쓰겠다.

X_1 과 X_2 는 이름은 다르지만 동일한
확률분포이다.

X_1	-1	5
$P(X_1 = x)$	0.9	0.1

X_2	-1	5
$P(X_2 = x)$	0.9	0.1

==

아제 슬롯머신에서 수행하는 두 개입에
대한 확률분포는 다음과 같다.

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

$P(Z=z)$	0.81	0.18	0.01
----------	------	------	------

=
이거 :

① $z = -2$ 인 경우 : 두 번째 틀을 때

$$P(Z=z) = 0.9 * 0.9 = 0.81$$

② $z = 4$ 인 경우 : 한번 틀었고
한번 떠는 경우

$$\begin{aligned} P(Z=z) &= 0.9 * 0.1 \\ &\quad + 0.1 * 0.9 \end{aligned}$$

③ $z = 10$ 인 경우 : 두 번째 떠는 경우

$$P(Z=z) = 0.1 * 0.1 = 0.01$$

=

따라서,

$$\begin{aligned} E(Z) &= -2 * 0.81 \\ &\quad + 4 * 0.18 \\ &\quad + 10 * 0.01 \\ &= -0.8 \end{aligned}$$

$$\text{Var}(Z) = (-2 + 0.8)^2 * 0.81$$

$$\begin{aligned}
 & + (4+0.8)^2 * 0.18 \\
 & + (10+0.8)^2 * 0.01 \\
 & = 6.48
 \end{aligned}$$

=

그런데, 확률분포 Z 의 기대값과 표산을
위와 같이 복잡한 과정을 가지지 않고
 X_1 과 X_2 의 기대값과 표산을 이용하여
구할 수 있다.

먼저, 아래 관계에 주의 한다.

Z 의 경우처럼 두 개의 독립간측인
 X_1 과 X_2 를 이용한 간측은
다음과 같이 표시한다.

$$Z = X_1 + X_2$$

그리고 아래 공식이 성립한다.

- -

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

실제로 $E(X) = -1 \times 0.9 + 5 \times 0.1 = \underline{\underline{-0.4}}$ 이며,

따라서 $E(Z) = -0.4 \times 2 = \underline{\underline{-0.8}}$ 이다.

앞서 계산한 기대값과 일치함을 알 수 있다.

동일한 이유로 해서 다음 공식도 성립한다.

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

따라서,

$$\begin{aligned}
 \text{Var}(Z) &= \text{Var}(X) \times 2 \\
 &= 3.24 \times 2 = \underline{\underline{6.48}}
 \end{aligned}$$

왜냐하면,

z

$$\begin{aligned} \text{Var}(X) &= (-1+0.4) * 0.9 \\ &\quad + (5+0.4)^2 * 0.1 \\ &= \underline{\underline{3.24}} \end{aligned}$$

독립관측들의 합에 대한
공식을 정리하면 다음과 같다.

X_1, X_2 가 X 의 독립관측인 경우,

$$E(X_1 + X_2) = 2 \cdot E(X)$$

$$\text{Var}(X_1 + X_2) = 2 \cdot \text{Var}(X)$$

(P. 263)

이 공식을 아래와 같이
 일반화 할 수 있다.

먼저, $X_1, X_2, X_3, \dots, X_m$ 이 X 의
특장관측이라 가정하자.

그러면 다음 공식이 성립한다.

$$\begin{aligned} E(X_1 + X_2 + X_3 + \dots + X_m) \\ = m \cdot E(X) \end{aligned}$$

$$\begin{aligned} \text{Var}(X_1 + X_2 + X_3 + \dots + X_m) \\ = m \cdot \text{Var}(X) \end{aligned}$$

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

이유: $E(2X)$ 는 관측값이 2배로 키진
새로운 관측값을 대상으로 하는.

서로운 확률변수 (Y)의 기대값이.
즉, $Y = 2X$ 를 만족하는 $E(Y)$ 일.

반면에,

$E(X_1 + X_2)$ 는 두 개의 확률변수가
결합된 상황에서의 기대값일.

확률변수의 조성

(p. 269 ~ 270)

앞에서 다른 경위에는 살피 이제부터는
 X 와 Y 는 서로 상관이 없고, 서로 독립인
확률변수이다.

즉,

$E(X)$ 와 $E(Y)$ 가 서로 상관이 없고,
또한 $\text{Var}(X)$ 와 $\text{Var}(Y)$ 도 서로

상관없다고 가정한다.

하지만 이 경우에도 아래 공식이 성립한다.

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

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

예제 (P. 269)

서로 다른 규칙과 상관을 갖는 두 대의
슬롯머신을 각각 한 번씩 게임했을 때의
기대치와 확률을 구하라고 한다.

슬롯머신 각각의 확률변수는 다음과 같다고
가정한다.

X	-5	395
$P(X=x)$	0.99	0.01

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

원래는, $X+Y$ 의 확률분포를 아래 표 모양으로 복잡하게
계산해서 구해야 하고

$X+Y$	-7	18	43	68	93	393	418	443	468	493
$P(X+Y=x+y)$	0.967 23	0.000992

이후에 기대값 $E(X+Y)$ 와 흔한 $\text{Var}(X+Y)$ 를
공식에 맞추어 계산해야 한다.

하지만 이 공식을 이용하면

$E(X+Y)$ 와 $\text{Var}(X+Y)$ 를 쉽게
구할 수 있다.

확률변수의 학센

(P. 271)

두 개의 확률변수 X 와 Y 의 차이를
다룰 때 $X-Y$ 라는 확률변수를
사용한다.

$X-Y$ 의 확률분포는 아래 모양을 갖는다.

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

예제 (P. 279)

A, B 주 레스토랑에서 지불하는 비용의

학점을 끈포가 각각 X와 Y라 하고,
아래의 값을 갖는다고 가정하자.

리스트 A:

X	20	30	40	45
$P(X=x)$	0.3	0.4	0.2	0.1

리스트 B:

y	10	15	18
$P(Y=y)$	0.2	0.6	0.2

그리면 두 리스트에서 지불한 비용의 차이에
대한 학점을 끈포는 다음과 같다.

$X - Y$	2	5	10	12	15	20	22	25	27	30	35
$P(X-Y=x-y)$	0.06	0.18	0.06	0.08	0.24	0.08	0.04	0.12	0.02	0.1	0.02

그런데 위와 같이 비용의 차이에 대한
학점을 끈포를 구하지 않아도 비용의 차이에
대한 대체로 $E(X - Y)$ 와
문서 $\text{Var}(X - Y)$ 를 대해 궁금으로
구할 수 있다.

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

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

주의: 차이의 분산을 구할 때 빼기 순서를
더한다.

이유: 하나의 변수에서 다른 변수를 빼더라도
확률분포의 분산은 증가한다.

이 카드로랑 예제에서 이 현상을
확인할 수 있다.

A 카드로랑 : 28에서 45 사이
 $\frac{3}{7}, \frac{25}{\underline{8}}$ 정도 차이.

B 카드로랑 : 16에서 18 사이
즉, $\frac{8}{\underline{8}}$ 정도 차이

둘다에 $X-Y$ 의 값은 2에서 35 사이,
즉, $\frac{33}{\underline{33}}$ 정도 차이.

그런데 $\frac{33}{\underline{33}} = 25 + \frac{8}{\underline{8}}$ 이다.

=====
| 친증정리 | (P. 272)

선형변환과 확률분포의 학습과 예제 :

$$E(aX+bY) = aE(X) + bE(Y)$$

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

$$E(aX - bY) = aE(X) - bE(Y)$$

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