5장 이산 확률 분포

- **주제**: 어떤 사건의 발생할 확률에 대한 정보를 활용하여 해당 사건이 미칠 수 있는 영향을 알아본다.

- **알아야 할 개념**
  - 이산 데이터: 1, 2, 3, 4, ... 처럼 각각의 값이 딱딱 끊어져서 구분될 수 있는 데이터
  - 이산 데이터 예제
    ① 연, 월, 일 관련 데이터
    ② 학년, 학점, 월급 관련 데이터
  - 연속 데이터: 질량, 길이, 시간 등 엄밀한 측정을 요하는 값들로 이루어진 데이터

- **5장 주요 예제: 슬롯머신**
  - 창: 3개
  - 그림: \$, 🍒, ○, 기타
    ↓ \$ → 달러, 🍒 → 체리, ○ → 레몬
  - 잭팟: 정해진 순서대로 그림이 나오는 사건
    책에서는 (\$, \$, \$)가 나오는 사건을 가리킴.

p.238

- 게임 방법
  - 게임당 1000원 지불 (주의: 책에서는 달러 사용)
  - 아래 4 사건이 발생하는 경우 상금 받음.

① (달러 사건) : ($, $, $)	상금 20,000원
② ($/체리 사건): $ * 2회, 🍒 * 1회	상금 15,000원
③ (체리 사건) : (🍒, 🍒, 🍒)	상금 10,000원
④ (레몬 사건) : (🍋, 🍋, 🍋)	상금 5,000원
⑤ (꽝 사건): 나머지 경우	상금 0원

- 각각의 창에서 특정 그림이 나올 확률

$	🍒	🍋	기타
0.1	0.2	0.2	0.5

- 전제: 각각의 창은 서로 독립이다.
  즉, 어느 하나의 창에 어떤 그림이 나올지는
  다른 창에 전혀 영향을 주지 않는다.

p.239

- 각 사건이 발생할 확률은 다음과 같다.

P(달러 사건) = P($, $, $) = 0.1 * 0.1 * 0.1 = (0.1)^3 = **0.001**

P($/체리 사건) = P($, $, 🍒) + P($, 🍒, $) + P(🍒, $, $)
= (0.1 * 0.1 * 0.2) * 3 = **0.006**

P(체리 사건) = P(🍒, 🍒, 🍒) = (0.2)^3 = **0.008**

P(레몬 사건) = P(🍋, 🍋, 🍋) = (0.2)^3 = **0.008**

P(꽝) = 1 − (0.001 + 0.006 + 0.008 * 2) = **0.977**

• 확률분포표와 기대치

- 확률분포표: 사건들의 확률 분포를 담은 표

- 각각의 사건은 서로 배반이어야 하며, 각 사건의 확률의 합은 1이어야 한다.

확률의 합이 1임을 반영

즉, 모든 발생 가능한 경우가 어떤 사건에는 포함되어야 하며, 단 하나의 사건에만 포함되어야 한다.

← 사건들이 서로 배반임을 반영.

예) 슬롯머신의 경우

사건	달러 사건	$/체리 사건	체리 사건	레몬 사건	꽝
확률	0.001	0.006	0.008	0.008	0.977

← 서로 배반

← 합이 1임

⇐ 슬롯머신에서 발생하는 모든 경우를 5가지 사건이 포함함. 또한 각 사건은 서로 배반임.

- 따라서 아래처럼 슬롯머신 게임에서 1회당 벌 수 있는 금액에 대한 확률분포표를 구할 수 있다.

사건	달러 사건	$/체리 사건	체리 사건	레몬 사건	꽝
버는 금액	19,000	14,000	9,000	4,000	-1,000
확률	0.001	0.006	0.008	0.008	0.977

- 기대치
  - 확률분포표에 포함된 사건이 발생할 때 얻을 수 있는 값에 대한 기대치
  - 예) 슬롯머신의 경우, 게임을 1회 실행할 때 벌 수 있는 돈의 기대치

(20,000 - 1,000) * P(달러 사건)
+ (15,000 - 1,000) * P($/체리 사건)
+ (10,000 - 1,000) * P(체리 사건)
+ (5,000 - 1,000) * P(레몬 사건)
+ (0 - 1,000) * P(꽝)

= -774

- **확률변수와 확률분포**
  - 확률변수: 특정한 확률과 연관되어 있는 값들을 대상으로 하는 변수

    예) 슬롯머신에서 확률변수는 다는 액수 중 하나를 가리킴. 즉, 19,000, 14,000, 9,000, 4,000, -1,000 중에 하나를 가리키는 변수.

  - 일반적으로 X, Y, Z, X_1, X_2, ... 등을 확률변수로 사용함.
  - 확률변수가 가리키는 값들의 집합이 다르면 다른 확률변수 기호를 사용한다.
  - 즉, 서로 다른 확률변수는 기본적으로 서로 다른 집합의 값들을 가리킨다.
  - 하지만, 서로 다른 기호가 동일한 집합의 값들을 가리킬 수는 있다. (교재 262쪽 독립관측 설명 참조)
  - 확률변수 X가 특정값 x를 가질 확률을 $P(X=x)$라고 표기함.
  - 주의①: 지금까지는 특정사건 A가 발생할 확률 $P(A)$를 다뤘지만 지금부터는 다루는 상황에 따라 다른 집합의 값을 다룬다는 점을 명시하기 위해 확률변수 기호를 함께 표기한다
  - 주의②: $P(X=x)$를 사용하기 위해서는 확률변수 X가 특정 값 x를 가질 확률을 계산할 수 있다는 것을 전제해야 한다.

    8장에서 연속데이터를 다룰 때에는 $P(X=x)$의 값은 항상 0이 된다. 대신에 $P(a < X < b)$처럼 X가 a와 b 구간 사이의 값을 취할 확률을 구한다.

- 슬롯머신 게임을 1회당 벌 수 있는 액수에 대한 확률분포를 확률변수를 이용하여 작성하면 다음과 같다.

x	19,000	14,000	9,000	4,000	-1,000
$P(X=x)$	0.001	0.006	0.008	0.008	0.977

p.244

- **확률변수 X에 대한 기대치**

- 확률변수 X에 대해 확률분포 $P(X=x)$가 주어졌을 때 확률변수 X에 대한 기대치 E(X)를 계산하는 일반적인 식은 다음과 같다:

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

* 주의: 평균값(μ) 계산과 유사한 모양을 가짐. 실제로 평균값은 기대치의 특별한 경우임.

$$\mu = \frac{\sum x}{n} = \sum x \cdot \frac{1}{n}$$

즉, 모든 x에 대해 $P(X=x) = \frac{1}{n}$ 이면 기대치는 평균값이 된다.

따라서 μ 라고 표기하기도 함

예) 슬롯머신의 경우 1회당 버는 액수를 가리키는 확률변수 X의 평균은 앞서 계산한 결과와 같다.

$$E(X) = \mu = -774$$

즉, 슬롯머신 게임을 1회 진행할 때 마다 평균적으로 774원을 잃을 것으로 기대할 수 있다.

p.245

- ## 확률변수 X에 대한 분산과 표준편차

- 질문: 앞서 계산한 기대치 결과에 의하면 슬롯머신 게임을 1회 진행할 때마다 평균적으로 77센트를 잃을 것이라고 기대한다.
그런데, 게임을 할 때마다 무조건 잃기만 할까?

- 대답: 그렇지 않다. 기대치는 기대치일 뿐이며 항상 그렇다는 의미는 아니다.
경우에 따라 돈을 딸 수도 있고 훨씬 더 많이 잃을 수도 있다. 그렇다면 얼마의 범위에서 돈을 따거나 잃을지 측정할 수 있는 방법이 필요하다.
즉, 값들의 어떻게 분포하는지 측정할 수단이 필요하며 이를 위해 값들의 분산을 이용한다.

p.246

- 확률변수 X에 대한 분산도 기존의 정의와 비슷하다.

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

* 주의: 기존에 정의한 분산의 정의와 유사함.

$$\sigma^2 = \frac{\sum (x-\mu)^2}{n} = \sum (x-\mu)^2 \cdot \frac{1}{n}$$

즉, $P(X=x) = \frac{1}{n}$.

- 확률변수 X의 표준편차는 분산의 루트값이다.

$$표준편차(\sigma) = \sqrt{Var(X)}$$

p. 247

예) 슬롯머신의 경우 1회당 버는 액수를 가리키는 확률변수 X의 평균, 분산, 표준편차는 다음과 같다.

← 앞서 계산하였음.

$$Var(X) = (19000+774)^2 * 0.001 + (14000+774)^2 * 0.006 + (9000+774)^2 * 0.008 + (4000+774)^2 * 0.008 + (-1000+774)^2 * 0.977$$

$$= 2,697,100$$

$$\sigma = \sqrt{Var(X)} = 1,642$$

설명: 슬롯머신 게임을 한 번 할 때마다 확률적으로 십중팔구 아래 구간의 값을 벌게 된다.

$$[-774-1,642 \ , -774+1,642]$$

즉, -2,416원에서 868원 사이의 돈을 십중팔구 번다.
하지만 현실에서는 게임 1회당 천원을 잃거나 최소 4,000원을 벌어야 한다. 따라서 게임 1회당 십중팔구 1,000원을 잃게 된다는 결론을 얻는다.