이항분포

(p. 333, 336, 337)

동일한 시도를 n번 시도할 때, 특정 사건이 r번 나올 확률들의 분포를 이항분포라 한다.

조건: 각각의 시도는 상호 독립적이며, 각각의 시도에서 특정 사건이 발생할 확률은 동일하다.

예제1: 4지선다 문제

하나의 질문을 임의로 찍었을 때 맞출 확률: $p = \frac{1}{4}$

질문:

n개의 질문을 모두 임의로 찍어서 맞출 때 r개의 정답을 얻을 확률?

맞히는 개수 →	r	0	1	⋯	r	⋯	⋯	$n-1$	n
r개 맞힐 확률 →	$P(X=r)$								

$P(X=r)$

어떻게 계산하나?

$P(X=r)$ 계산 예제 : $n=3$ 인 경우

즉, 3 개의 문제 중에 r 개를 맞힐 확률.

먼저 $r \in \{0, 1, 2, 3\}$ 임.

따라서 4 가지 경우를 살펴보아야 함.

① 모두 틀린 경우

$$P(X=0) = \left(\frac{3}{4}\right)^3 = {}^3C_0 \cdot \left(\frac{1}{4}\right)^0 \cdot \left(\frac{3}{4}\right)^3$$

② 한 문제 맞힐 경우

$$P(X=1) = \underbrace{3}_{={}^3C_1} \cdot \underbrace{\frac{1}{4} \cdot \left(\frac{3}{4}\right)^2}$$

3 경우 각각 발생할 확률

${}^3C_1$ ← 3 문제 중에 1개 선택.

③ 두 문제 맞힐 확률

$$P(X=2) = \underbrace{{}^3C_2} \cdot \underbrace{\left(\frac{1}{4}\right)^2 \cdot \frac{3}{4}}$$

3 경우 각각이 발생할 확률

→ 3 문제 중에 2개 선택

④ 세 문제 모두 맞힐 확률

$$P(X=3) = \left(\frac{1}{4}\right)^3 = {}^3C_3 \cdot \left(\frac{1}{4}\right)^3 \cdot \left(\frac{3}{4}\right)^0$$

$$\Rightarrow P(X=r) = {}^3C_r \cdot \left(\frac{1}{4}\right)^r \cdot \left(\frac{3}{4}\right)^{3-r}$$

일반화

한 번의 시도에서 성공할 확률 $= p$

원하는 결과가 발생한다는 의미

n번 시도에서 r번 성공할 확률

$_nC_r$ 경우 각각이 발생할 확률

$$P(X=r) = {}_nC_r \cdot p^r \cdot (1-p)^{n-r}$$

$$= {}_nC_r \cdot p^r \cdot q^{n-r} \quad (q = 1-p)$$

n개 중에서 r개를 선택하는 방법의 수

선택한 r개가 성공할 확률

r개를 제외한 나머지가 실패할 확률

이런 경우 X는 이항분포를 따른다 말하고

아래와 같이 표기한다.

$$X \sim B(n, p)$$

이항분포 반복 횟수 한 번의 시도에서 성공할 확률

이항분포 적용 연습 (교재에 없음)

하나의 정상적인 동전을 30번 던져서 24번

앞면이 나올 확률?

결론답안:

정상적인 동전을 30번 던져서 앞면이 나오는 횟수에 대한 확률분포 X는 이항분포를 따르며, 아래 관계를 만족시킨다.

$$X \sim B\left(30, \frac{1}{2}\right)$$

따라서 앞면이 24번 나올 확률은 다음과 같다.

$$P(X=24) = {}_{30}C_{24} \cdot \left(\frac{1}{2}\right)^{24} \cdot \left(\frac{1}{2}\right)^{6} = {}_{30}C_{24} \cdot \left(\frac{1}{2}\right)^{30}$$

$$\fallingdotseq 0.0006$$

즉, 10,000분의 6의 확률이다.

또한 앞면이 24회 이상 나올 확률은 다음과 같다.

$$P(X \geq 24) = P(X=24) + \cdots + P(X=30)$$

$$= 0.0007$$

즉, 10,000분의 7 수준이다.

일반적으로 사건 발생확률이 5% 이하면 그런 사건이 실제로 발생하기 어렵다고 본다. 따라서 동전을 30번 던져서 앞면이 24번 나온다면 그 동전은 정상적이지 않다고 판단한다.

결론

이후에 가설검정을 배울때 자세히 다룬다.

이항분포의 기대치와 분산 (p.338 ~ 341)

$X \sim B(n, p)$ 라 하자.

그러면:

$E(X) = n \cdot p$

$Var(X) = n \cdot p \cdot q \quad (q = 1-p)$

설명:

① X_1을 한 번 시도한 경우의 확률분포라 하자.

즉,

r	0	1
$P(X=r)$	q	p

← 한 번 시도하면 성공 또는 실패만 발생함.

따라서, $E(X_1) = 0 \cdot q + 1 \cdot p = p$

$$
\begin{aligned}
Var(X_1) &= (0-p)^2 \cdot q + (1-p)^2 \cdot p \\
&= p^2 \cdot (1-p) + (1-2p+p^2) \cdot p \\
&= p^2 - \cancel{p^3} + p - 2p^2 + \cancel{p^3} \\
&= p - p^2 = p(1-p) \\
&= p \cdot q
\end{aligned}
$$

② X 와 X_1 사이에 아래 관계가 성립한다 :

X는 X_1을 반복적이며 독립적으로 n번 실행하는 확률분포,
즉, X_1을 n번 독립관측 하는 것이다.

따라서, 아래의 식을 얻는다.

$$X = \underbrace{X_1 + \cdots + X_1}_{n\text{번}}$$

③ 위 식을 이용하여 $E(X)$와 $Var(X)$를 아래와 같이 계산할 수 있다.

$$E(X) = E(\underbrace{X_1 + \cdots + X_1}_{n\text{번}})$$

$$= n \cdot E(X_1)$$

$$= n \cdot p$$

$$Var(X) = Var(X_1 + \cdots + X_n)$$

$$= n \cdot Var(X_1)$$

$$= n \cdot p \cdot q$$

이항분포 예제

① 동전을 n번 던져 앞면이 r번 나올 확률들의 분포:

$$X_{coin} \sim B\left(n, \frac{1}{2}\right)$$

② 주사위를 n번 던져 2보다 큰 소수가 r번 나올 확률들의 분포:

$$X_{dice} \sim B\left(n, \frac{1}{3}\right)$$

이유: 2보다 큰 소수는 3과 5이다.

푸아송 분포 (p. 346~348)

특정 기간 동안 어떤 사건이 발생하는 횟수의 평균값 또는 비율이 주어졌을 때, 미래의 동일한 기간 동안 해당 사건이 r번 발생할 확률들의 분포

예제) 영화관의 낡은 팝콘 기계

- 주중에 평균 3.4회 고장 발생

특정 기간

사건 발생 빈도수의 평균값

질문: 다음 주중에 한 번도 고장이 발생하지 않을 확률은?

해법: 이런 경우 푸아송 분포를 활용한다.
팝콘 기계의 경우 아래처럼 표기한다.

$$X \sim Po(3.4)$$

일반화 :

어떤 주어진 구간에 사건이 발생하는 수를 X라 할 때, 만약 X가 구간마다 λ번 발생하는 평균값을 갖는다면,

람다라 읽는다

X를 푸아송 분포를 따른다고 말하고 아래와 같이 표기 한다.

그리스어 알파벳 (영어의 L에 해당)

$$X \sim P_o(\lambda)$$

특정 구간에 언급된 사건이 r번 발생할 확률

X가 푸아송 분포를 따를 때 $P(X=r)$은 아래와 같다.

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

($e \simeq 2.71$, 오일러 상수)

예제 : 영화관 팝콘기계 (p.349~350)

영화관 팝콘기계의 고장율은 푸아송 분포를 따른다 : $X \sim Po(3.4)$

따라서, 예를 들어 다음 주에 팝콘기계가 한 번도 고장나지 않을 확률은 다음과 같다

$$P(X=0) = \frac{e^{-3.4} \cdot (3.4)^0}{0!}$$

$$= \frac{1}{e^{3.4}}$$

$$\approx 0.033$$

즉, 3% 정도이다.

결론 : 아무래도 새 팝콘기계를 준비하는 게 좋을 듯 싶다.

푸아송 분포의 기대치와 분산 (p.348)

푸아송 분포의 기대치와 분산은 매우 간단하다.

$X \sim Po(\lambda)$ 인 경우

$$E(X) = \lambda$$
$$Var(X) = \lambda$$

상식적으로 생각할 수 있는 값과 일치한다.

이다.

경험적으로 특정기간에
평균 λ번 어떤 사건이
발생하였다면, 미래에도
동일한 기간동안 λ번
동일한 사건이 발생할 것이라
기대할 수 있다.

포아송 분포의 결합 (p.352~353)

X와 Y가 독립확률변수일 때 아래가 성립한다.
(5장 참조)

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

따라서 두 개의 독립확률변수 X, Y에 대해서도
아래 관계가 성립한다.

$$X \sim Po(\lambda_x)$$
$$Y \sim Po(\lambda_y)$$

두 사건이 서로 독립이며
동일한 기간을
대상으로 해야 함.

$$\Rightarrow \begin{cases} X+Y \sim Po(\lambda_x + \lambda_y) \\ E(X+Y) = Var(X+Y) = \lambda_x + \lambda_y \end{cases}$$

예제 : 팝콘가게 | 음료수가게

$X \sim Po(3.4)$ | $Y \sim Po(2.3)$

팝콘가게
일주일 동안
고장횟수

음료수가게
일주일 동안
고장횟수

서로 독립임.

두 가게중 어느 하나의 일주일 동안의 고장횟수는
$X+Y \sim Po(3.4+2.3)$의 포아송 분포를 따른다.
따라서, 예를 들어, 다음 주 일주일 동안 어느 가게도
고장나지 않을 확률은 아래와 같다.

$$P(X+Y=0) = \frac{e^{-5.7} \cdot 5.7^0}{0!} = 0.003 \quad (0.3\%)$$

이항분포와 푸아송 분포의 관계 (p.356)

$X \sim B(n, p)$ 이고 $n \geq 50$, $p \leq 0.1$ 인 경우 $X \sim P_o(np)$ 라고 해도 무방하다.

즉, 이항분포인 경우 n이 적당히 크고, p가 꽤 작으면 푸아송 분포를 이용하여 확률을 계산할 수 있다.

예제 (p.357)

50개의 문제를 무작정 찍어서 5개의 답을 맞출 확률은 얼마인가?
단, 한 문제를 찍어 정답을 맞출 확률은 0.05 이다.

결론답안: X가 정답 개수라 할 때 $X \sim B(50, 0.05)$ 이다. 따라서, 푸아송 분포를 이용할 수 있다.
즉, $X \sim P_o(2.5)$ 라고 보자.
따라서, $P(X=5) = \dfrac{e^{-2.5} \cdot (2.5)^5}{5!} = 0.067$
즉, 6.7% 이다. $\boxed{\text{ㅁ}}$