

포아송 분포

* 단위시간이나 단위공간에서 어떤 사건의 출현횟수가 갖는 이산형 확률분포를 포아송분포라고 하며, 이항분포에서 $np=m$ 을 작은 값으로 일정하게 고정시키고 시료 n 을 증가시킨 극한 분포이다.

• 이러한 포아송 분포는 모분포가 단위 시간 또는 공간에서 작은 경우 적용되어 이항분포를 근사분포라고 하면, 포아송 분포는 결점분포라고 한다.

• 포아송 분포를 따르는 확률변수 X 의 확률질량함수(PMF) 및 기대값과 분산을 다음과 같다.

$$P(X=x) = \frac{e^{-m} m^x}{x!} \quad (\text{단, } m=np, m>0, x=0,1,2,\dots,n)$$

$$E(X) = np = m$$

$$\text{Var}(X) = np = m$$

포아송 분포의 특징

1. 기대값과 분산이 같은 이산형 확률분포이다.

2. $m \geq 5$ 일 때는 정규분포에 근사하는 분포이다.

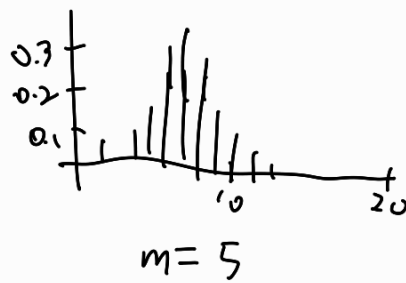
3. m 이 작을 때는 왼쪽으로 기울어진 비대칭 분포가 되나, m 이 커짐에 따라 좌-우대칭의 분포에 근사한다.

4. 모분포 적합도를 P 가 대단히 작아지는 경우 적용되는 이산형 확률분포이다.

(즉, $P < 0.1$ 이면 포아송 분포를 따른다.)

5. 부적합수(결점수), 단위당 부적합수, 결점률과 같은 계수지는 포아송 분포를 따른다.

이동아수 분포의 확률은 m 의 값이 변함에 따라 다른 값이 되기 때문에 분포의 형태도 m 의 변함에 따라 형태가 다르게 된다. $m=1, 5, 10$ 일 때 분포의 형태는 다음의 그림과 같다. 그림에서 $m=5$ 일 때 좌우대칭인 형태의 분포를 보여준다.



(예제) 어느 공장에서 종업원 1000명을 선정할 때 다음과 같이 특정일의 생일이 발생할 확률을 구정하라.

(1) 특정일이 아무도 생일이 되지 않을 확률

i) 종업원 1000명중 특정일이 생일일 기대값

$$P = \frac{1}{365} = 2.74 \times 10^{-3}, n=1000 \rightarrow E(X) = m = nP = \boxed{2.74}$$

ii) 특정일날 생일이 한 사람이 존재하지 않을 확률

$$P(X=0) = \frac{e^{-m} m^x}{x!} = \frac{e^{-2.74} m^0}{0!} = e^{-2.74} = \boxed{0.06459}$$

(2) 특정일이 2인 이상이 생일이 될 확률

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) \\ &= 1 - e^{-m}(1+m) = 1 - e^{-2.74}(1+2.74) \\ &= \boxed{0.76} \end{aligned}$$

이항분포

• 모집단 (N)의 크기가 유한 모집단일 때 사용하는 이항분포의 발생분포로서, 이항분포가 N 이 무한대이므로 무복귀합통계량 (무복귀량) p 가 거의 변하지 않는 복귀합통계량 방식이라면 이항분포는 비복귀합통계량 방식을 따른다.

• 크기 N 의 유한모집단 (총량) 속성 A 를 갖는 것이 M 개에서 비복귀합통계량으로 n 개의 시료 (속성 A 를 갖는 것이 x 개)를 취할 때, 그 중에 A 의 속성을 갖는 것의 개수를 나타내는 확률변수 X 라고 하자. 그러면 X 가 x 의 값을 가질 확률 $P(X=x)$

• 이항분포를 따르는 확률변수 X 의 확률질량함수 (pmf) 및 기댓값과 분산은 다음과 같다.

$$P(X=x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \frac{M^x \cdot (N-M)^{n-x}}{N^n}, \quad E(X) = np$$

$$Var(X) = \left(\frac{N-n}{N-1}\right) np(1-p) \quad (\text{여기서 } M=M^0, x=0, 1, 2, 3, \dots, n)$$

이항분포의 특징

1. 분포가 이산적 특징을 띈다.
2. $N \rightarrow \infty$ 에 근사하면 이항분포에 근사하는 분포이다.
3. $p=0.5$ 일 때는 도형균치에 대하여 좌우대칭의 분포이다.
4. 무한승정계수 $\left(\frac{N-n}{N-1}\right)$ 을 갖으며, 정수가 가장 높은 이산형 분포이다.

(예제) 흰 바둑돌 3개와 검은 바둑돌 5개가 들어있는 주머니에서 4개를 꺼낼 때 2개씩 흰 바둑돌과 2개의 검은 바둑돌이 나올 확률을 구하시오.

$$N=8, n=4, M=3, x=2, P(x=2) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \frac{3C_2 \times 5C_2}{8C_4} = \frac{3}{7} = \boxed{0.428572}$$

(예제) 40명으로 구성된 학급에는 3명의 소년소녀가장이 포함되어 있다. 임의로 5명의 학생을 선발하였을 때, 여기에 소년소녀 3명의 수를 확률변수 X 로 나타내고 가정하고 다음에 문제를 구하시오.

$$(1) P(X=1) \quad P(X=1) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \frac{3C_1 \times 37C_4}{40C_5} = \frac{595}{1976} = \boxed{0.30}$$

$$(2) E(X), Var(X) \quad E(X) = np = n \times \frac{M}{N} = 5 \times \frac{3}{40} = 0.375$$

$$(M=NP, P=\frac{M}{N})$$

$$Var(X) = \left(\frac{N-n}{N-1} \right) np(1-p) = \left(\frac{40-5}{40-1} \right) \times 5 \times \frac{3}{40} \left(1 - \frac{3}{40} \right)$$

$$= \frac{259}{832} = \boxed{0.31}$$