5.5절 ~ 5.6절

5주2강. 이산형 확률분포2

5.5절

5.5. 다항분포

다항분포

- **1. 다항분포**
- 가능한 결과가 세 개 이상인 실험을 n번 반복하였을 때 **각 결과가 나타날 확률을 구하는 분포를 다항분포(multinomial distribution)**라고 한다.
- 각 실험의 실현치가 k개 범주 중의 하나이며, 각 범주가 나타날 확률이 $p_1, p_2, \cdots, p_k$인 확률 실험에서 실험을 n번 반복했을 때 각 범주의 관측치를 확률변수 $X_1, X_2, \dots, X_k$로 표현하면, $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$가 관측될 확률을 측정하는 분포를 **다항분포**라고 한다.
- 다항분포의 확률은 다음과 같이 구한다.

$$P_r(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!} \cdot {p_1}^{x_1} \cdot {p_2}^{x_2} \cdot \cdots \cdot {p_k}^{x_k}$$

- 여기에서 $n = x_1 + x_2 + \cdots + x_k$이며, $p_1 + p_2 + \cdots + p_k = 1$이다.

- 관측도수와 확률, 범주의 관계를 표로 정리하면 다음과 같다.

관측도수	x_1	x_2	x_3	$\cdots$	x_k
확률	p_1	p_2	p_3	$\cdots$	p_k
범주	1	2	3	$\cdots$	k

예 5-8 한 실험의 결과는 1, 2, 3, 4로 표현되는 4가지 속성 중 하나로 나타나며 각각의 속성이 나타날 확률이 순서대로 1/10 , 2/10 , 3/10 , 4/10 라고 한다. 이 실험에서 관측한 20개의 값에 있어서 각 속성의 관측치가 3, 4, 5, 8일 확률을 구하라.

(sol) X_1, X_2, X_3, X_4를 각각의 속성이 관측된 수를 의미하는 확률변수라고 할 때, 각각의 속성이 나타날 확률이 1/10 , 2/10 , 3/10 , 4/10 로 다음과 같이 나타낼 수 있다.

$$P_r(X_1 = 3, X_2 = 4, X_3 = 5, X_4 = 8) = \frac{20!}{3!\ 4!\ 5!\ 8!} \cdot \left(\frac{1}{10}\right)^3 \cdot \left(\frac{2}{10}\right)^4 \cdot \left(\frac{3}{10}\right)^5 \cdot \left(\frac{4}{10}\right)^8$$

5.6절

5.6. 포아송분포

시메옹 포아송

- 1781~1840년 프랑스 출생
- 공학자, 수학자, 물리학자

포아송 분포

- 이항분포의 특수한 케이스
- 시행횟수가 많다.
- 발생확률이 낮다.

포아송 분포의 확률질량함수 유도

- 하루(24시간) 동안 거리를 돌아 다니다가 길냥이를 만나는 횟수(X)
- 마주칠 확률 p : 매우 작음
- 시행횟수 n : $n \to \infty$
- $np =$ 하루동안 돌아다니며 길냥이를 만나는 평균 횟수 $= \mu \quad \Longrightarrow \quad p = \frac{\mu}{n}$

$$\begin{aligned}
\binom{n}{x} p^x (1-p)^{n-x} &= \binom{n}{x} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} \\
&= \frac{n!}{(n-x)!\ x!} \frac{\mu^x}{n^x} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\mu^x}{n^x} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x}
\end{aligned}$$

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!}\frac{\mu^x}{n^x}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$
$$= \frac{n(n-1)\cdots(n-x+1)}{n^x}\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$
$$= \left(\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right)\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$

$$P(X=x) = \lim_{n\to\infty}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-x+1}{n}\cdot\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}\right)$$
$$= \lim_{n\to\infty}\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n = \lim_{n\to\infty}\frac{\mu^x}{x!}\left(\left(1-\frac{\mu}{n}\right)^{-\frac{n}{\mu}}\right)^{-\mu} = \frac{\mu^x}{x!}e^{-\mu}$$

1. 포아송분포

- 포아송분포는 주어진 단위시간, 거리, 영역 등에서 어떤 사건이 발생되는 횟수를 측정하는 확률변수로 다음과 같은 …를 생각할 수 있다.
  - ① 특정지역에서 제한된 … 내에 발생되는 교통사고의 수의 분포
  - ② 타이프를 치는 데 있어서 페이… 오타의 수의 분포
  - ③ 특정시간에 고속도로 톨게이트를 지… …제차량 수의 분포
  - ④ 자동생산라인에서 특정시간에 발생되는 불… …의 분포

- 단위구간 내에서 어떤 사건이 평균 μ회 발생… 라고 할 때, $\boldsymbol{X} \sim$ **Poisson** $(\boldsymbol{\mu})$로 표현하며 사건…

$$P_r(X = k) = \frac{\mu^k}{k!} e^{-\mu}, \ldots$$

- 여기에서 e는 자연대수(log)의 밑수로 $e = 2.7\ldots$

1. 구간이 겹치지 않는 경우 사건의 수는 서로 독립니다.
2. 사건이 일어날 횟수의 평균은 구간의 길이에 비례한다.
3. 임의의 시간에 일어날 확률은 0이다.

2. 포아송 분포의 평균과 분산

- 확률변수 X가 평균 μ인 포아송분포를 따를 때, X의 평균과 분산은 다음과 같이 동일하다.
  ① 평균 : $E(X) = \mu$
  ② 분산 : $Var(X) = \mu$

$$E(X) = \sum_{k=0}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} = \sum_{k=1}^{\infty} k\frac{\mu^k}{k!}e^{-\mu} = \mu e^{-\mu}\sum_{k=1}^{\infty}\frac{\mu^{k-1}}{(k-1)!}$$

$$= \mu e^{-\mu}\sum_{n=0}^{\infty}\frac{\mu^n}{n!} = \mu e^{-\mu}\cdot e^{\mu} = \mu$$

$\boldsymbol{Var(X) = \mu}$

$$Var(X) = E(X^2) - [E(X)]^2 = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} - \mu^2 = \sum_{k=1}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} - \mu^2$$

$$= \sum_{k=1}^{\infty} k \frac{\mu \cdot \mu^{k-1}}{(k-1)!} e^{-\mu} - \mu^2 = \mu e^{-\mu} \sum_{k=1}^{\infty} k \frac{\mu^{k-1}}{(k-1)!} - \mu^2 = \mu e^{-\mu} \sum_{n=0}^{\infty} (n+1) \frac{\mu^n}{n!} - \mu^2$$

$$= \mu e^{-\mu} \left[\sum_{n=0}^{\infty} n \frac{\mu^n}{n!} + \sum_{n=0}^{\infty} \frac{\mu^n}{n!} \right] - \mu^2 = \mu e^{-\mu} \left[\sum_{n=1}^{\infty} \frac{\mu \cdot \mu^{n-1}}{(n-1)!} + e^{\mu} \right] - \mu^2$$

$$= \mu e^{-\mu} \left[\mu \sum_{s=0}^{\infty} \frac{\mu^s}{s!} + e^{\mu} \right] - \mu^2 = \mu e^{-\mu} [\mu e^{\mu} + e^{\mu}] - \mu^2 = \mu^2 + \mu - \mu^2 = \mu$$

- **3. 포아송 분포의 확률계산**
- 아래는 포아송분포에서 각각의 평균값 μ에 대한 확률변수 X의 c까지의 누적확률값 $P_r(X=k)=\sum_{k=0}^{c}\frac{\mu^k}{k!}e^{-\mu}$을 계산해 나타낸 표이다.

표 7-1 **포아송분포표**(〈부록 IV〉의 [표 2]의 일부)

c \ μ	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0	6.2	6.4
0	.012	.010	.008	.007	.006	.005	.004	.003	.002	.002	.002
1	.066	.056	.048	.040	.034	.029	.024	.021	.017	.015	.012
2	.185	.163	.143	.125	.109	.095	.082	.072	.062	.054	.046
3	.359	.326	.294	.265	.238	.213	.191	.170	.151	.134	.119
4	.551	.513	.476	.440	.406	.373	.342	.313	.285	.259	.235
5	.720	.686	.651	.616	.581	.546	.512	.478	.446	.414	.384
6	.844	.818	.791	.762	.732	.702	.670	.638	.606	.574	.542
7	.921	.905	.887	.867	.845	.822	.797	.771	.744	.716	.687
8	.964	.955	.944	.932	.918	.903	.886	.867	.847	.826	.803
9	.985	.980	.975	.968	.960	.951	.941	.929	.916	.902	.886
10	.994	.992	.990	.986	.982	.977	.972	.965	.957	.949	.939
11	.998	.997	.996	.995	.993	.990	.988	.984	.980	.975	.969
12	.999	.999	.999	.998	.997	.996	.995	.993	.991	.989	.986
13	1.000	1.000	1.000	.999	.999	.999	.998	.997	.996	.995	.994
14				1.000	1.000	1.000	.999	.999	.999	.998	.997
15							1.000	1.000	.999	.999	.999
16									1.000	1.000	1.000

예 5-9 최근 올림픽도로에서는 하루 평균 5건의 교통사고가 발생한다. 교통사고의 발생횟수가 포아송분포를 따른다고 할 때, 어느 날 교통사고가 전혀 일어나지 않을 확률은 얼마인가? 또 교통사고가 3번 이상 일어날 확률은 얼마인가?

- (sol) 확률변수 X가 하루에 발생하는 교통사고의 횟수를 나타낸다고 하면, 하루에 발생되는 교통사고의 횟수가 평균 5회이므로 $X \sim \text{Poisson}(5)$이고 다음과 같이 나타낼 수 있다.

$$P_r(X = k) = \frac{5^k}{k!} e^{-5}$$

- 교통사고가 전혀 발생하지 않을 확률과 3회 이상 발생할 확률은 다음과 같이 계산된다.

$$P_r(X = 0) = \frac{5^0}{0!} e^{-5} = e^{-5} \approx 0.00674$$

$$P_r(X \geq 3) = 1 - P_r(X < 3) = 1 - \big(P_r(X = 0) + P_r(X = 1) + P_r(X = 2)\big)$$

$$= 1 - 0.1247 = 0.8753$$

예 5-10 한 공장의 자동화기계가 제품을 생산하는 데 생산된 제품 중에서 불량품의 수가 시간당 평균 3개씩 발생된다고 한다. 특정시간에 생산된 제품 중에서 불량품이 2개 이상 발생될 확률은 얼마인가?

(sol) 확률변수 X가 불량품의 수를 나타낼 때, $X \sim \text{Poisson}(3)$ 이며, 부록 Ⅲ의 [표 2]에 의하여 $\mu = 3.0$ 일 때 다음과 같이 계산된다.

$$P_r(X \geq 2) = 1 - P_r(X \leq 1) = 1 - 0.199 = 0.801$$

끝~~ ♥♥