4장 ~ 5장

7주1강 복습 및 문제풀이(4장,5장)

4장 복습

이산형 확률변수

- 이산형 확률변수란 0이 아닌 확률 값을 갖는 실수 값이 셀 수 있는 경우를 말한다. 즉, 확률변수가 이산점에서 0이 아닌 확률을 갖는 확률변수를 말한다.
- 이산형 확률변수 X의 가능한 값을 $x_1, x_2, \cdots, x_n$이라 하고 $X = x_i$일 때의 확률이 P_i라면 다음과 같이 표현한다.

$$P_r(X = x_i) = P_i, \qquad i = 1, 2, 3, \cdots, n$$

- **이산형 확률변수의 확률조건**

1. $0 \le P_i \le 1,\ i = 1, 2, 3, \cdots, n$. 즉, 각 x_i가 나타날 확률은 0과 1사이의 값을 갖는다.
2. $\sum_{i=1}^{n} P_i = 1$. 즉, 모든 가능한 경우의 확률의 합은 1이다.

확률 질량 함수

- 이산형 확률변수는 이산점(discrete points)에서 0이 아닌 확률 값을 가지며 확률은, $P_r(X = x_i) = P_i, \quad i = 1, 2, 3, \cdots, n$으로 표현한다. 이와 같이 각 이산점에 있어서 확률의 크기를 표현하는 함수를 **확률질량함수(probability mass function)**라고 한다.

연속형 확률변수

- **연속형 확률변수**란?
  가능한 값이 실수의 어느 특정구간 전체에 해당하는 확률변수를 말한다.
  즉, 특정 실수 구간에서 0이 아닌 확률을 갖는 확률변수이다.

- 연속형 확률변수는 특정한 실수구간 내에서 0이 아닌 확률을 가지므로 이 구간에 대한 확률은 함수의 형태로 표현한다.

확률밀도함수

- 연속형 확률변수 $\boldsymbol{X}$ 의 확률함수를 $\boldsymbol{f}(\boldsymbol{x})$라고 할 때, $\boldsymbol{f}(\boldsymbol{x})$는 **확률밀도함수 (probability density function; p.d.f)**라고 부르며 다음 조건을 만족한다.

1. 모든 $\boldsymbol{X}$ 값에 대하여 $\boldsymbol{f}(\boldsymbol{x}) \geq 0$이다. 즉, $\boldsymbol{X}$의 모든 실수 값에 대하여 확률밀도함수는 0 이상이다.
2. $\boldsymbol{X}$의 모든 가능한 값의 확률은 적분 $\int_{-\infty}^{\infty} f(x)\,dx$로 구하며 이 값은 항상 1이다.
3. 구간 $(\boldsymbol{a}, \boldsymbol{b})$의 확률은 $\int_{a}^{b} f(x)\,dx$이다. 즉 구간 $(\boldsymbol{a}, \boldsymbol{b})$에 대한 $\boldsymbol{X}$의 확률은 그 구간에 있어서 확률밀도함수 $\boldsymbol{f}(\boldsymbol{x})$로 만들어지는 면적의 크기이다.

누적분포함수

- **누적분포함수(cumulative distribution function ; c.d.f)**란?

특정값 a에 대하여 확률변수 X가 $X \le a$인 모든 경우의 확률의 합으로 다음과 같이 표현한다.

$$P_r(X \le a) = F_X(a)$$

- 이산형 확률변수에서는 $F_X(a) = \sum_{x_i \le a} P(X = x_i)$이고,

- 연속형 확률변수에서는 $F_X(a) = \int_{-\infty}^{a} f(x)\, dx$이다.

결합 확률 분포

1. 이산형 확률변수인 경우

- 확률변수 X의 가능한 값이 $x_1, \cdots, x_n$이고, 확률변수 Y의 가능한 값이 $y_1, \cdots, y_m$이라면, (X, Y)의 결합확률분포는 다음과 같이 정의한다.

$$P_r(X = x_i, Y = y_j) = P_{ij}, \qquad i = 1, 2, \cdots, n, \qquad j = 1, 2, \cdots, m$$

- 여기에서 함수 $P_r(X = x_i, Y = y_j) = P_{ij}$를 **결합확률질량함수(joint probability mass function)**라고 정의하며, 모든 (i, j)에 대해 다음이 성립한다.

$$P_{ij} \geq 0, \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} = 1$$

2. 연속형 확률변수인 경우

- 확률변수 X, Y의 **결합확률밀도함수(joint probability density function)**를 $f(x,y)$라고 정의할 때, 모든 (x,y)에 대해 다음이 성립한다.

$$f(x,y) \geq 0$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

주변확률분포의 정의

1. 이산형 확률변수

- 확률변수 X의 가능한 값이 $x_1, \cdots, x_n$이고, 확률변수 Y의 가능한 값이 $y_1, \cdots, y_m$이며, X와 Y의 결합확률분포가 $P_r\left(X = x_i, Y = y_j\right) = P_{ij},\ (i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m)$이라고 할 때, X와 Y의 주변확률분포는 각각 다음과 같이 구한다.
- X의 **주변확률분포** :

$$P(X = x_i) = \sum_{j=1}^{m} P_{ij} = P_{i1} + P_{i2} + \cdots + P_{im}$$

- Y의 **주변확률분포** :

$$P\left(Y = y_j\right) = \sum_{i=1}^{n} P_{ij} = P_{1j} + P_{2j} + \cdots + P_{nj}$$

2. 연속형 확률변수인 경우

- 두 확률변수 X, Y의 결합확률밀도함수를 $f(x, y)$라고 할 때, X, Y의 주변확률분포는 각각 다음과 같이 정의한다.
- X의 **주변확률분포** :

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\ dy$$

- Y의 **주변확률분포** :

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\ dx$$

- 연속형 확률분포에서 주변확률분포함수를 **주변확률밀도함수(marginal probability density function)**라고 부른다

두 확률변수의 독립성

1. 이산형 확률변수

- 두 이산형 확률변수 X, Y의 결합확률분포함수를 $P(X = x, Y = y)$ 라 하고, 각각의 주변분포함수를 $P(X = x)$, $P(Y = y)$라고 할 때, 모든 가능한 (x, y) 값에 대하여,

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

  일 때 한하여 두 변수 X 와 Y 는 **독립**이다.

2. 연속형 확률변수

- 두 연속형 확률변수 X, Y의 결합확률밀도함수를 $f(x, y)$라 하고, $f(x)$와 $f(y)$를 각각 X, Y의 주변확률밀도함수라고 하면,

$$f(x, y) = f(x) \cdot f(y)$$

  일 때 한하여 두 변수 X 와 Y 는 **독립**이다.

기댓값(평균)

- **기대값**(expected value) 또는 **평균**(mean, average)이라는 개념은 확률분포에서 **분포의 무게중심**을 말하며, 확률 값을 가중치로 하는 확률변수의 가능한 값에 대한 가중평균(weighted average)이라고 할 수 있다.

1. **이산형 확률변수**

- 이산형 확률변수 X의 가능한 값이 $(x_1, x_2, \cdots, x_n)$이며, $P(X = x_i) = P_i,\ i = 1, 2, \cdots, n$일 때, X 의 기대값 $E(X)$는 다음과 같다.

$$E(X) = \sum_{i=1}^{n} x_i P_i$$

2. **연속형 확률변수**

- 연속형 확률변수 X의 확률밀도함수를 $f(x)$라 하면, X의 기대값 $E(X)$는 다음과 같다.

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

분산

- 평균이 확률분포의 무게중심인 데 비하여 분산은 **확률분포의 흩어진 정도를 측정**하는 것으로 평균이 같은 경우에도 분산의 크기에 따라서 분포의 모양이 달라진다.

- 확률변수 X의 **분산(variance)**은 $E(X) = \mu$라고 할 때, X와 μ의 편차의 제곱, 즉 $(X-\mu)^2$의 기대값으로 $Var(X)$ 또는 σ_X^2으로 표현되며,

$$\sigma_X^2 = E(X-\mu)^2$$

- 으로 구한다. 여기에서 확률변수 X를 나타낼 필요가 없는 경우에는 σ_X^2에서 X를 없애고 σ^2으로 표현한다. 분산의 양의 제곱근을 **표준편차(standard deviation)**라고 하며 σ로 표현한다. 즉, $\sigma = \sqrt{\sigma^2}$이다.

분산의 계산공식

$$\boldsymbol{\sigma^2 = E\left(X^2\right) - (E(X))^2 = E\left(X^2\right) - \mu^2}$$

- $E(X^2)$의 계산

$$E(X^2) = \sum_{i=1}^{n} x_i^2 P_i \qquad (X : \text{이산형 확률변수})$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx \qquad (X : \text{연속형 확률변수})$$

공분산

- 공분산은 두 확률변수의 결합분포를 알고 있는 경우에 구할 수 있는 모수이다. 주로 두 변수 사이의 관계의 밀접도를 측정하는 상관계수(correlation coefficient)를 구하는 과정에서 계산되는 경우가 많다.

- 두 확률변수 X와 Y의 **공분산(covariance)**은 $Cov(X,Y)$ 또는 σ_{XY}로 표현하며 다음과 같이 계산한다.

$$\boldsymbol{\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X \cdot \mu_Y}$$

4장 연습문제 풀이

연습문제 11. 확률변수 X의 확률밀도함수가 $f(x) = \begin{cases} \frac{1}{2}x, & 0 \le x \le 2 \\ 0, & \text{다른 곳에서} \end{cases}$ 일 때,

(a) X의 확률분포를 그림으로 표현하라.

풀이)

(b) X의 평균 μ_X와 분산 σ_X^2을 구하여라.

풀이) $\mu_X = E(X) = \int_0^2 \frac{1}{2}x \cdot x\, dx = \int_0^2 \frac{1}{2}x^2\, dx = \left[\frac{1}{6}x^3\right]_0^2 = \frac{4}{3}$

$E(X^2) = \int_0^2 \frac{1}{2}x \cdot x^2\, dx = \int_0^2 \frac{1}{2}x^3\, dx = \left[\frac{1}{8}x^4\right]_0^2 = 2$

$\sigma_X^2 = E(X^2) - \mu_X^2 = 2 - \frac{16}{9} = \frac{2}{9}$

(c) X의 누적분포함수를 구하고, 이를 그림으로 표현하여라.

풀이) $F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \le x < 2 \\ 1, & 2 \le x \end{cases}$

연습문제 11. 확률변수 X의 확률밀도함수가 $f(x) = \begin{cases} \frac{1}{2}x, & 0 \le x \le 2 \\ 0, & \text{다른 곳에서} \end{cases}$ 일 때,

(d) $P_r(X > 1)$을 계산하여라.

풀이) $P_r(X > 1) = \int_1^2 \frac{1}{2}x\,dx = \left[\frac{1}{4}x^2\right]_1^2 = \frac{1}{4}(4-1) = \frac{3}{4}$

(e) $X > 1$이 주어졌을 때, $X > \frac{3}{2}$에 대한 조건부 확률을 구하여라.

풀이) $P_r\left(X > \frac{3}{2} \mid X > 1\right) = \frac{P_r\left(X>\frac{3}{2}\right)}{P_r(X>1)} = \frac{7/16}{3/4} = \frac{7}{12}$

(f) 확률변수 Y를 $Y \equiv X^2$으로 정의할 때, Y의 확률밀도함수를 구하여라.

풀이) $F_Y(y) = P_r(Y \le y) = P_r(X^2 \le y)$

$= P_r\left(-\sqrt{y} \le X \le \sqrt{y}\right) = P_r\left(0 \le X \le \sqrt{y}\right)$

$= \int_0^{\sqrt{y}} \frac{1}{2}x\,dx = \left[\frac{1}{4}x^2\right]_0^{\sqrt{y}} = \frac{1}{4}y,\ 0 \le y \le 4$

연습문제 12. 확률변수 X의 확률밀도함수가 $f(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0, & \text{다른 곳에서} \end{cases}$ 일 때,

(a) X의 확률분포를 그림으로 표현하라.

풀이)

(b) X의 누적분포함수를 구하고, 이를 그림으로 표현하여라.

풀이) $F(x) = P_r(X \le x) = \int_0^x 6t(1-t)\,dt = [3t^2 - 2t^3]_0^x = 3x^2 - 2x^3, \quad 0 \le x < 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ 3x^2 - 2x^3, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$$

(c) X의 평균 μ_X와 분산 σ_X^2을 구하여라.

풀이) $\mu_X = E(X) = \int_0^1 6x(1-x) \cdot x\,dx = \int_0^1 (6x^2 - 6x^3)\,dx = \left[2x^3 - \frac{6}{4}x^4\right]_0^1 = 2 - \frac{3}{2} = \frac{1}{2}$

$$E(X^2) = \int_0^1 6x(1-x) \cdot x^2\,dx = \int_0^1 (6x^3 - 6x^4)\,dx = \left[\frac{6}{4}x^4 - \frac{6}{5}x^5\right]_0^1 = \frac{6}{4} - \frac{6}{5} = \frac{6}{20} = \frac{3}{10}$$

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$$

연습문제14. 두 확률변수 X와 Y의 결합확률분포함수가 다음과 같을 때,

Y \ X	2	3	4
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{1}{16}$	$\frac{1}{4}$	0
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(a) X와 Y의 주변확률분포를 구하여라.

X	0	1	2	계
확률	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{6}{16}$	1

Y	2	3	4	계
확률	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

(b) $E(X+Y)$를 구하여라.

풀이) $E(X) = 0 \times \frac{5}{16} + 1 \times \frac{5}{16} + 2 \times \frac{6}{16} = \frac{17}{16}$, $E(Y) = 2 \times \frac{1}{4} + 3 \times \frac{1}{2} + 4 \times \frac{1}{4} = \frac{12}{4} = 3$

$E(X+Y) = E(X) + E(Y) = \frac{17}{16} + 3 = \frac{65}{16}$

연습문제14. 두 확률변수 X와 Y의 결합확률분포함수가 다음과 같을 때,

$X \backslash Y$	2	3	4
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{1}{16}$	$\frac{1}{4}$	0
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

(c) $\mathrm{Cov}(X, Y)$를 구하여라.

풀이)

$X \cdot Y$	0	2	3	4	6	8
확률	$\frac{5}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$E(XY) = 0 \times \frac{5}{16} + 2 \times \frac{1}{16} + 3 \times \frac{1}{4} + 4 \times \frac{1}{8} + 6 \times \frac{1}{8} + 8 \times \frac{1}{8} = \frac{1+6+4+6+8}{8} = \frac{25}{8}$$

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{25}{8} - \frac{17}{16} \times 3 = \frac{50-5}{16} = -\frac{1}{16}$$

(b) X와 Y가 독립인가를 보여라.

풀이) $P_r(X = 0) = \frac{5}{16}$, $P_r(Y = 2) = \frac{1}{4}$, $P_r(X = 0, Y = 2) = \frac{1}{16}$이다.

$P_r(X = 0, Y = 2) = \frac{1}{16} \neq P_r(X = 0) \times P_r(Y = 2) = \frac{5}{64}$이므로 X와 Y는 서로 독립이 아니다.

연습문제18. 동전 세 개를 던지는 실험에서 확률변수 X는 '앞면의 수 ' 를 나타내고, 확률변수

Y는 $Y = \begin{cases} 1, & \text{첫째 동전이 앞면인 경우} \\ 0, & \text{첫째 동전이 앞면이 아닌 경우} \end{cases}$ 와 같다고 할 때

(a) X와 Y의 결합확률분포함수를 구하여라.

풀이)

X \ Y	0	1	계
0	$\frac{1}{8}$	0	$\frac{1}{8}$
1	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
2	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$
3	0	$\frac{1}{8}$	$\frac{1}{8}$
계	$\frac{1}{2}$	$\frac{1}{2}$	1

(b) X와 Y의 주변확률분포를 구하여라.

풀이)

X	0	1	2	3
확률	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Y	0	1	계
확률	$\frac{1}{2}$	$\frac{1}{2}$	1

(c) $E(X+Y)$와 $E(X-Y)$를 구하여라.

풀이) $E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{2}$, $E(Y) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$

$E(X+Y) = E(X) + E(Y) = \frac{3}{2} + \frac{1}{2} = 2$, $E(X-Y) = E(X) - E(Y) = \frac{3}{2} - \frac{1}{2} = 1$

연습문제18. 동전 세 개를 던지는 실험에서 확률변수 X는 '앞면의 수 ' 를 나타내고, 확률변수 Y는 $Y = \begin{cases} 1, & \text{첫째 동전이 앞면인 경우} \\ 0, & \text{첫째 동전이 앞면이 아닌 경우} \end{cases}$ 와 같다고 할 때

(d) X와 Y의 분산을 구하여라.

풀이) $E(X^2) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} = 3,\quad E(Y^2) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$

$\sigma_X^2 = E(X^2) - [E(X)]^2 = 3 - \frac{9}{4} = \frac{3}{4},\qquad \sigma_Y^2 = E(Y^2) - [E(Y)]^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

(e) X와 Y의 공분산을 구하여라.

풀이)

$X \cdot Y$	0	1	2	3
확률	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

$E(XY) = 0 \times \frac{1}{2} + 1 \times \frac{1}{8} + 2 \times \frac{2}{8} + 3 \times \frac{1}{8} = 1,\qquad \mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = 1 - \frac{3}{2} \times \frac{1}{2} = \frac{1}{4}$

(f) X와 Y가 독립인가를 보여라.

풀이) $P_r(X = 0) = \frac{1}{8},\qquad P_r(Y = 0) = \frac{1}{2},\qquad P_r(X = 0, Y = 0) = \frac{1}{8}$이다.

$P_r(X = 0, Y = 0) = \frac{1}{8} \neq P_r(X = 0) \times P_r(Y = 0) = \frac{1}{16}$이므로 X와 Y는 서로 독립이 아니다.

5장 복습

베르누이 시행(Bernoulli trial)

1. 베르누이 시행(Bernoulli trial)

- 실험에서 결과가 둘 중의 하나로 나타나는 실험을 말한다. 그 중 하나를 **성공(success : s)**이라 하고 다른 하나를 **실패(failure : f)**라고 정의한다.

2. 베르누이 확률변수

- 베르누이 시행에서 결과가 s이면 '1'이고, 결과가 f이면 '0'이라고 정의된 확률변수를 말한다.

3. 베르누이 확률분포

- 확률변수 X의 분포가 다음과 같을 때, X를 **베르누이 확률변수**라 하고, $X \sim \mathrm{Bernoulli}(p)$로 정의한다.

$$P_r(X = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1$$

- 베르누이 분포의 모양은 확률 p의 값에 의하여 결정되므로 **이 분포의 모수는 p**이다.

4. 베르누이 확률분포의 평균과 분산

$$E(X) = p, \qquad \sigma^2(X) = p(1-p)$$

이항분포

1. 이항분포

- **이항분포(binomial distribution)**는 베르누이 시행을 독립적으로 n번 반복했을 때 나타나는 결과에 있어서 성공(s)의 횟수에 대한 분포를 구하는 것이다.
- 성공의 확률이 p이고 실패의 확률이 q ($q = 1 - p$)인 베르누이 독립적으로 n번 반복하였을 때 나타나는 성공의 횟수를 확률변수 X라 할 때, X를 이항확률변수(binomial random variable)라 하고, $\boldsymbol{X \sim B\,(n, p)}$로 정의하며 확률분포는 다음과 같다.

$$P_r(X = k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \cdots, n$$

- 이항분포에서 **모수는 각 시행에서 성공이 나타날 확률 $\boldsymbol{p}$와 시행횟수 $\boldsymbol{n}$**이다.

2. 이항분포의 평균과 분산

$$E(X) = np, \qquad Var(X) = npq$$

기하분포

1. 기하분포 (geometric distribution)

- 기하분포는 베르누이 시행으로부터 유도된 확률분포로써, 매번 시행에서 성공의 확률이 p 이고 실패일 확률이 $q = 1-p$인 베르누이 시행을 독립적으로 반복하여 실시할 때 **첫 성공이 나타날 때까지의 실행횟수에 대한 확률분포**이다.
- 확률변수 X가 첫 성공이 나올 때 까지의 시행 횟수라고 하면, X는 기하분포를 따르며 $\boldsymbol{X \sim \mathrm{Geometric}(p)}$로 표현하고,확률은 다음과 같이 계산한다.

$$P_r(X = k) = q^{k-1}p, \qquad k = 1, 2, 3, \cdots$$

- 여기에서 X의 가능한 값은 모든 양의 정수이며, 기하분포의 **모수는 성공의 확률 $\boldsymbol{p}$**이다.

2. 기하분포의 평균과 분산

$$E(X) = \frac{1}{p} \qquad Var(X) = \frac{q}{p^2}$$

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.

1. **포아송분포**

- 포아송분포는 주어진 단위시간, 거리, 영역 등에서 어떤 사건이 발생되는 횟수를 측정하는 확률변수이다.
- 단위구간 내에서 어떤 사건이 평균 μ회 발생한다고 한다. 확률변수 X를 사건의 발생횟수라고 할 때, $\boldsymbol{X \sim \mathrm{Poisson}\ (\mu)}$로 표현하며 사건이 k번 발생될 확률은 다음과 같다.

$$P_r(X = k) = \frac{\mu^k}{k!} e^{-\mu}, \qquad k = 0, 1, 2 \cdots$$

- 여기에서 e는 자연대수(log)의 밑수로 $e = 2.71828 \ldots$ 이다.

2. 포아송 분포의 평균과 분산

$$E(X) = \mu, \qquad Var(X) = \mu$$

5장 연습문제 풀이

연습문제1(f). 확률변수 X가 이항분포 $B(n, p)$를 따를 때, 다음 각 경우에 대한 확률을 구하여라.

(f) $n = 2,\ p = 0.3,\ X = 1$

풀이) $P_r(X = 1) = \binom{2}{1} \times 0.3 \times 0.7 = 2 \times 0.3 \times 0.7 = 0.42$

연습문제7. 한 농구선수가 자유투의 75%를 성공시킨다고 할 때

(a) 12번의 자유투에서 성공횟수의 기댓값은 얼마인가?

풀이) $E(X) = 12 \times \frac{3}{4} = 9$

(b) 12번의 자유투에서 10번 성공할 확률은 얼마인가?

풀이) $P_r(X = 10) = \binom{12}{10} \times \left(\frac{3}{4}\right)^{10} \times \left(\frac{1}{4}\right)^{2} = 0.232$

(c) 12번의 자유투에서 최소한 10번 성공할 확률은 얼마인가?

풀이) $P_r(X \geq 10) = \binom{12}{10} \times \left(\frac{3}{4}\right)^{10} \times \left(\frac{1}{4}\right)^{2} + \binom{12}{11} \times \left(\frac{3}{4}\right)^{11} \times \left(\frac{1}{4}\right) + \binom{12}{12} \times \left(\frac{3}{4}\right)^{12} = 0.391$

(d) 자유투의 세 번째 시도에서 처음 성공할 확률은 얼마인가?

풀이) $P_r(\text{첫 번째 실패, 두 번째 실패, 세 번째 성공}) = \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{64}$

연습문제9. 한 학교의 통계학개론 과목에서 10%의 학생들이 낙제를 한다고 할 때, 25명이 있는 한 통계학계론 과정에서

$$X \sim B\left(25, \frac{1}{10}\right)$$

(a) 낙제하는 학생들 수의 평균과 분산을 구하여라.

풀이) $E(X) = 25 \times \frac{1}{10} = \frac{25}{10},\quad Var(X) = 25 \times \frac{1}{10} \times \frac{9}{10} = \frac{225}{100}$

(b) 낙제생의 수가 2명 이내일 확률을 구하여라.

풀이) $P_r(X \le 2) = \binom{25}{0} \times \left(\frac{9}{10}\right)^{25} + \binom{25}{1} \times \left(\frac{9}{10}\right)^{24} \times \left(\frac{1}{10}\right) + \binom{25}{2} \times \left(\frac{9}{10}\right)^{23} \times \left(\frac{1}{10}\right)^{2} = 0.537$

연습문제14. 한 사격선수가 이동표적을 맞출 확률이 0.9라고 할 때

$p = 0.9$인 기하분포 모형

(a) 이 선수가 세 번째 시도에서 표적을 처음 맞출 확률은 얼마인가?

풀이) $P_r(X \leq 3) = (0.1)^2 \times 0.9 = 0.009$

(b) 이 선수가 표적을 처음 맞추는데 필요한 평균사격횟수는 얼마인가?

풀이) $E(X) = \frac{1}{p} = \frac{1}{0.9} = \frac{10}{9}$

(c) 이 선수가 두 번째 시도 이내에 표적을 맞출 확률은 얼마인가?

풀이) $P_r(X \leq 2) = P_r(X = 1) + P_r(X = 2) = 0.9 + 0.1 \times 0.9 = 0.99$

연습문제16(d). 확률변수 X가 세 개의 가능한 경우가 있는 다항분포 Multinomial$(n: p_1, p_2, p_3)$를 따를 때, 다음 각 경우의 확률을 구하여라.

(d) $P_r(X_1 = 4, X_2 = 4, X_3 = 2), n = 10,\ p_1 = 0.4,\ p_2 = 0.3,\ p_3 = 0.3$

풀이) $\frac{10!}{4!4!2!} \times (0.4)^4 \times (0.3)^4 \times (0.3)^2 = 0.0588$

연습문제18(d). 확률변수 X가 포아송분포 Poisson(μ)를 따를 때, 다음 각 경우의 확률을 구하여라.

- <부록V>의 [표 2]를 이용해 확률을 구한다.

(d) $P_r(X \geq 1),\ \mu = 5$

풀이) $P_r(X \geq 1) = 1 - P_r(X = 0) = 1 - 0.007 = 0.993$

연습문제21. 어느 한 책이 오자가 평균 10페이지에 20개씩 발생하는 포아송분포를 따른다고 할 때, 다음 각 경우의 확률을 구하여라.

평균 10페이지에 20개이므로 페이지당 평균 2개이며, 네 페이지당 평균 8개씩 오자가 발생된다.

(a) 어느 페이지에 오자가 하나도 없을 확률

풀이) $P_r(X = 0) = 0.135,\ \mu = 2$

(b) 어느 페이지에 오자가 두 개 이상 있을 확률

풀이) $P_r(X \geq 2) = 1 - P_r(X \leq 1) = 1 - 0.406 = 0.594, \mu = 2$

(c) 특정한 네 페이지에 오자가 하나도 없을 확률

풀이) $P_r(X = 0) = 0.000,\ \mu = 8$

끝~~ ♥♥