LGE - SNU DS 고급 과정

확률통계 및 공통 알고리즘

과제2 - 확률변수 및 확률분포

2022. 01

※필요하다면 다음의 값을 이용하시오 : $z_{0.977} = 2, z_{0.934} = 1.5,\ z_{0.691} = 0.5$

문제 1. 1부터 5까지의 자연수가 각각 하나씩 적혀있는 5개의 서랍이 있다. 5개의 서랍 중 임의로 2개가 배정된다고 한다. 이 때 배정되는 서랍에 적혀있는 자연수 중 큰 수를 확률변수 X 라 하고 작은 수를 확률변수 Y 라 하자.

(1) $E(X-Y)$를 구하시오.

(2) $P(1 \le Y \le 3)$를 구하시오.

[풀이].

(1) X 의 확률분포는 다음과 같다.

x	2	3	4	5
$p(x)$	0.1	0.2	0.3	0.4

$$E(X) = 1/10 \times 2 + 2/10 \times 3 + 3/10 \times 4 + 4/10 \times 5 = 4$$

Y 의 확률분포는 다음과 같다.

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1

$$E(Y) = 4/10 + 3/10 \times 2 + 2/10 \times 3 + 1/10 \times 4 = 2$$

$$\Rightarrow E(X-Y) = 2$$

(2) $P(1 \le Y \le 3) = P(Y=1) + P(Y=2) + P(Y=3) = 0.9$

문제 2. 확률변수 X 의 확률밀도 함수가 다음과 같을 때 물음에 답하여라.

$$f(x) = \begin{cases} x^2/3 & \text{if } -2 < x < 1 \\ 0 & \text{else} \end{cases}$$

(a) $f(x)$가 확률밀도함수임을 증명하라.

(b) $P(X \leq 0)$ 를 구하라.

(c) X 의 평균과 분산을 구하라.

[풀이].

(a) $\int_{-\infty}^{\infty} f(x)dx = \int_{-2}^{1} x^2/3\, dx = 1^3/9 - (-2)^3/9 = 1,\ f(x) \geq 0$

(b) $P(X \leq 0) = \int_{-2}^{0} x^2/3dx = 8/9$

(d) $E(X) = \int_{-2}^{1} x^3/3dx = 1^4/12 - (-2)^4/12 = -5/4$
$E(X^2) = \int_{-2}^{1} x^4/3dx = 1^4/15 - (-2)^5/15 = 11/5$
$Var(X) = E(X^2) - E(X)^2 = 51/80$

문제 3. X 와 Y 의 결합확률분포가 다음과 같다.

$y \setminus x$	1	2	3
1	0.05	0.10	0.15
2	0.20	0	0.15
3	0.10	0.20	0.05

(a) X 와 Y 의 주변확률분포를 구하여라.

(b) $E(X+Y)$, $E(X-2Y)$를 구하여라.

(c) X 와 Y 는 서로 독립인가? 그 이유는?

(d) $Y = 2$ 일때 X 의 조건부확률분포를 구하여라.

[풀이].

(a)

x	1	2	3
$p_X(x)$	0.35	0.30	0.35

y	1	2	3
$p_Y(y)$	0.30	0.35	0.35

(b) $E(X) = 2, E(Y) = 2.05 \quad \Rightarrow \quad E(X+Y) = 4.05, E(X-2Y) = -2.1$

(c) $P(X=2, Y=2) = 0 \neq P(X=2) \times P(Y=2) = 0.105$ 이므로

X 와 Y 는 독립이 아니다.

(d)

x	1	2	3
$p_{X\|Y=2}(x)$	4/7	0	3/7

문제 4. 각 제품의 불량일 확률은 서로 독립이며 불량률이 0.1 이라고 한다. 4개의 제품을 조사하였을 때 다음 물음에 답하여라.

(a) 불량품의 수가 3개 이상일 확률은?

(b) 4개의 제품 중 불량품 개수의 평균과 분산은?

[풀이].

(a) 4개의 제품 중 불량품의 개수를 X 라 하면, $X \sim B(4, 0.1)$ 이다.

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) \\ &= \binom{4}{3}(0.1)^3(0.9)^1 + \binom{4}{4}(0.1)^4 \\ &= 0.0037 \end{aligned}$$

(b) $E(X) = 4 \times 0.1 = 0.4$
$Var(X) = 4 \times 0.1 \times 0.9 = 0.36$

문제 5. 어느 설비를 통해 생산되는 시간당 전력량(kWh)는 근사적으로 평균이 100이고 분산이 16인 정규분포를 따른다고 한다. 시간당 생산 전력량이 98 이상 106 이하일 확률을 근사적으로 계산하여라.

[풀이]. $X=$ 시간당 생산 전력량이라 하면,
$P(98 \le X \le 106) = P(-0.5 \le \frac{X-100}{4} \le 1.5) = P(-0.5 \le Z \le 1.5) = 0.934 - (1-0.691) = 0.625$

문제 6. A, B 두집단의 시험성적은 각각 $N(80, 10^2), N(70, 5^2)$을 따른다고 한다. 전체 인원 중 A, B 집단에 속하는 비율은 50%, 50%라고 한다. 임의로 조사한 인원의 시험성적이 80점이하일 확률을 구하여라.

[풀이]. $X=$ 임의로 조사한 인원의 시험성적이라 하면

$$\begin{aligned} P(X \le 80) &= P(A)P(X \le 80|A) + P(B)P(X \le 80|B) \\ &= 0.5 \times P(Z \le 0) + 0.5 \times P(Z \le 2) \\ &= 0.7385 \end{aligned}$$

문제 7. 제품의 품질점수를 조사하여 상위 25%의 제품에는 A등급을 매기고자 한다. 품질점수가 0에서 50 사이의 균등분포를 따른다고 할 때, A등급을 받기위한 최소 품질점수를 구하여라.

[풀이]. $X=$ 제품의 품질점수라 하면 $X \sim U(0, 50)$이므로
$P(X \ge a) = \frac{50-a}{50} = 0.25$에서 $a = 37.5$이다.