6.4절

6주2강 연속형 확률분포2

복습

복습 : 균등분포(uniform distribution)

1. **균등분포(uniform distribution)**는 연속형 분포에서 가장 단순한 분포형태로 특정구간 내의 값들이 가능성이 균등한 분포를 말한다.

- 연속형 확률변수 X가 실수구간 $[a, b]$에서 나타날 가능성이 균등할 때, X는 균등분포를 따른다고 하며 $\boldsymbol{X \sim U(a, b)}$로 표현한다. X의 확률밀도함수는 다음과 같다.

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le X \le b \\ 0, & \text{다른 곳에서} \end{cases}$$

복습

2. 균등분포의 평균과 분산

- 확률변수 X가 $X \sim U(a, b)$라고 할 때, X의 평균과 분산은 다음과 같다.

  ① 평균 : $E(X) = \frac{b+a}{2}$

  ② 분산 : $Var(X) = \frac{1}{12}(b-a)^2$

3. 균등분포의 확률계산

- 구간 $\alpha \le X \le \beta$에서의 확률은 다음과 같다.

$$P_r(\alpha \le X \le \beta) = \frac{\beta - \alpha}{b - a}, \qquad a \le \alpha, \qquad \beta \le b$$

복습 : 정규분포

- **정규분포**란 가능한 값이 $(-\infty, \infty)$사이의 모든 실수값이고 분포의 형태가 종모양(bell shape)인 분포를 말한다. 정규분포(normal distribution)는 통계이론에 있어서 가장 중요한 확률분포로 **대부분의 통계분석은 수집된 자료가 정규분포를 따른다고 전제**한다.

- 확률변수 X가 평균 μ와 분산 σ^2을 갖는 정규분포를 따른다고 할 때, X 확률밀도함수는 다음과 같다.

$$X \sim N(\mu, \sigma^2) \qquad f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty$$

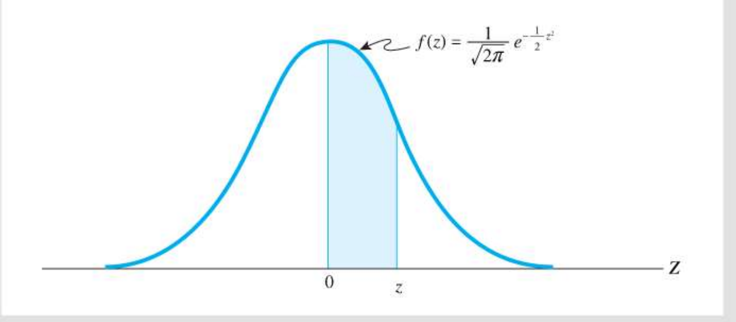
복습 : 표준정규분포

- 확률변수 Z가 평균이 0이고 분산이 1인 정규분포를 따를 때 Z는 **표준정규분포**를 따른다고 하며, $Z \sim N(0,1)$으로 표현한다. Z의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty < z < \infty$$

- 표준정규분포의 형태는 다음과 같으며, 중심 0에서부터 양의 값 Z까지의 확률은 색칠한 부분의 넓이와 같다. 수식을 통해 확률값을 구하는 것이 쉽지 않으므로 확률 $P(0 \le Z \le z)$은 주로 표준정규분포표를 이용하여 확률값을 구한다.

그림 8－5 표준정규분포



복습 : 정규분포의 확률계산

표준화

- 정규분포를 따르는 확률변수 $X \sim N(\mu, \sigma^2)$에서 $Z = \frac{X-\mu}{\sigma}$로 정의하면, $Z \sim N(0,1)$이다.
- 위의 표준화 공식을 이용하여 정규분포를 따르는 확률변수를 표준정규분포로 표준화시켜 특정구간의 확률을 구할 수 있다. 표준정규분포의 특정 실수영역 (a, b)에 있어서 확률은 다음과 같다.

$$P_r(a \leq Z \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

- 위의 적분계산은 매우 복잡하여 쉽지 않음을 알 수 있다. 위의 표준정규분포에서 설명한 바와 같이 부록의 표3에 제시된 확률값을 이용해 확률을 계산해 낼 수 있다.

복습 : 지수분포

- 지수분포란 어떤 사건이 포아송분포에 의하여 발생될 때, 지정된 시점으로부터 이 사건이 처음 발생될 때까지 걸린 시간을 측정하는 확률분포이다.
- 확률변수 T가 다음과 같은 확률밀도함수를 가질 때 T는 모수 λ를 갖는 지수분포를 따른다고 하며 다음과 같이 표현한다.

$$f(t) = \lambda e^{-\lambda t}, \qquad t > 0$$

- 지수분포의 평균과 분산

① $E(T) = \frac{1}{\lambda}$,

② $Var(X) = \frac{1}{\lambda^2}$

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.

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

- 이항확률변수 X의 평균과 분산은 다음과 같다.

  ① 평균 : $E(X) = np$

  ② 분산 : $Var(X) = npq$

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, $\boldsymbol{X \sim \textbf{Poisson}\ (\mu)}$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.

- 확률변수 X가 평균 μ인 포아송분포를 따를 때, X의 평균과 분산은 다음과 같이 동일하다.

  ① 평균 : $E(X) = \mu$

  ② 분산 : $Var(X) = \mu$

6.4절

6.4 이항분포의 근사확률

1. n이 큰 경우 이항분포의 확률계산

- 실제 문제에 있어서 n이 큰 경우 이항분포의 확률계산은 쉽지 않다. 이항분포에서 n이 작은 경우에는 부록 Ⅲ의 [표 1]을 이용하여 확률을 구할 수 있으나, 위의 경우와 같이 n이 큰 경우에는 p의 값의 크기에 따라 표준정규분포 또는 포아송분포를 이용하여 근사확률을 구한다.

① 이항분포의 정규근사

- 확률변수 X가 $X \sim B(n, p)$인 이항분포를 따를 때, X의 평균과 분산은 각각 다음과 같음을 알고 있다.

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

- 이러한 이항분포에서 n이 크고 p가 $\frac{1}{2}$에 가까운 경우에 X 의 분포는 평균 np 와 분산 npq를 갖는 정규분포와 근사적으로 동일하다고 할 수 있으며, X에 대한 확률은 $X \sim N(np, npq)$를 이용하여 근사적으로 구할 수 있다.
- 이항분포는 이산형 확률변수이고 정규분포는 연속형 확률변수 이므로 확률계산에서 약간의 조정을 하는 것이 필요하다. 정규분포를 이용하여 이항분포의 근사확률을 구하는 경우에 있어서 0.5를 조정하는 것을 이산형 확률분포의 확률을 연속형 확률분포를 이용하여 구하는 데 따르는 수정이라고 하여 **연속수정(continuity correction)**이라고 한다.

② 이항분포의 포아송근사

- 이항분포 $X \sim B(n, p)$에서 n이 크고 p가 작은 경우에 있어서의 확률은 포아송분포를 이용하여 근사적으로 구할 수 있다.
- 포아송분포에서는 평균 μ가 모수이고, 이항분포 $X \sim B(n, p)$ 에서는 평균이 np이므로 $X \sim B(n, p)$ 는 $X \sim \text{Poisson}\ (np)$를 이용하여 근사확률을 구한다.

예 6-12. 주사위를 100번 던지는 실험에서 짝수의 눈금이 60회 이상 나타날 확률을 구하라.

(sol) 매번 실행에서 짝수가 나타날 확률은 이므로 확률변수 X를 100번 던지는 실험에서 나타난 짝수 눈금의 수라고 할 때 $X \sim B(100, 1/2)$이고, X의 평균과 분산은 각각 다음과 같다.

$$E(X) = np = 100 \times \frac{1}{2} = 50$$

$$Var(X) = npq = 100 \times \frac{1}{2} \times \frac{1}{2} = 25$$

$P(X \geq 60)$을 정규분포를 이용하여 근사확률을 구하면 다음과 같다.

$$\begin{aligned} P_r(X \geq 60) &= P_r\left(\frac{X - np}{\sqrt{npq}} \geq \frac{59.5 - 50}{\sqrt{25}}\right) = P_r\left(Z \geq \frac{9.5}{5}\right) \\ &= P_r(Z \geq 1.9) = 0.5 - P_r(0 \leq Z \leq 1.9) = 0.5 - 0.4713 = 0.0287 \end{aligned}$$

예 6-13. 서울시민 중에서 30%의 사람들이 A정당을 지지한다고 한다. 한 모임에 모인 50명의 서울시민 중에서 다음의 확률을 구하여라.

(sol) 확률변수 X를 모임에 모인 사람 중에서 A정당을 지지하는 사람의 수라 할 때,
$X \sim B(50, 0.3)$임을 알 수 있으며, 평균과 분산은 각각 $E(X) = 50 \times 0.3 = 15, Var(X) = 50 \times 0.3 \times 0.7 = 10.5$이다.
다음의 각 확률을 정규분포 $N(15, 10.5)$를 이용하여 근사확률을 구하면 다음과 같다.

(a) A정당을 지지하는 사람이 20명 이상일 확률
(sol) 이항분포의 정규근사를 이용해 다음과 같이 구할 수 있다.

$$P_r(X \geq 20) = P_r(X \geq 19.5) = P_r\left(\frac{X - np}{\sqrt{npq}} \geq \frac{19.5 - 15}{\sqrt{10.5}}\right)$$
$$= P_r(Z \geq 1.39) = 0.5 - P_r(0 \leq Z \leq 1.39) = 0.5 - 0.4177 = 0.0823$$

(b) A정당을 지지하는 사람이 5명 이내일 확률

(sol) 이항분포의 정규근사를 이용해 다음과 같이 구할 수 있다.

$$P_r(X \leq 5) = P_r(X \leq 5.5) = P_r\left(\frac{X - np}{\sqrt{npq}} \leq \frac{5.5 - 15}{\sqrt{10.5}}\right) = P_r(Z \leq -2.93)$$
$$= P_r(Z \geq 2.93) = 0.5 - P_r(0 \leq Z \leq 2.93) = 0.5 - 0.4983 = 0.00173$$

(c) A정당을 지지하는 사람이 10명에서 20명 사이에 있을 확률

(sol) 이항분포의 정규근사를 이용해 다음과 같이 구할 수 있다.

$$P_r(10 \leq X \leq 20) = P_r(9.5 \leq X \leq 20.5) = P_r\left(\frac{9.5 - 15}{\sqrt{10.5}} \leq \frac{X - np}{\sqrt{npq}} \leq \frac{20.5 - 15}{\sqrt{10.5}}\right)$$
$$= P_r(-1.70 \leq Z \leq 1.70) = 2 \times P_r(0 \leq Z \leq 1.70) = 2 \times 0.4554 = 0.9108$$

예 6-14. 이동표적을 맞추는 사격에서 매번 시행에서 표적을 맞추는 확률이 0.05라고 한다. 이를 100번 시행할 때 다음 확률을 구하라.

(sol) 확률변수 X를 100번 시행에서 표적을 맞춘 횟수라고 할 때, $X \sim B(100, 0.05)$이다.
$E(X) = 100 \times 0.05 = 5$이므로, X에 대한 확률은 $\text{Poisson}(5)$(포아송 근사)를 이용하여 구할 수 있으며 부록 Ⅲ의 [표 2]에 의하여 확률을 다음과 같이 구할 수 있다.

(a) 표적을 5번 이상 맞출 확률

(sol) 이항분포의 포아송 근사를 이용해 다음과 같이 구할 수 있다.

$$P_r(X \geq 5) = 1 - P_r(X \leq 4) = 1 - 0.440 = 0.560$$

(b) 표적을 3번 이내로 맞출 확률

(sol) 이항분포의 포아송 근사를 이용해 다음과 같이 구할 수 있다.

$$P_r(X \leq 3) = P_r(X = 0) + P_r(X = 1) + P_r(X = 2) + P_r(X = 3) = 0.265$$

끝~~ ♥♥