5.1절 ~ 5.4절

5주1강. 이산형 확률분포1

5.1절

5.1. 확률분포와 모수

확률분포와 모수

1. **확률분포**

- 통계분석에서 자료를 수집하고 그 수집된 자료로부터 어떤 정보를 얻고자 하는 경우에는 항상 수집된 자료가 특정한 확률분포를 따른다고 가정한다.

확률분포

- **이산형 확률분포**
- **연속형 확률분포**

2. **모수(parameter)**

- 모든 확률분포는 그 분포의 모양을 결정하는 값이 있는데 그 값을 모수 (parameter)라고 한다. 모수는 확률분포의 특성을 나타내는 값으로 모든 확률분포는 모수에 의하여 구체적인 모양이 결정되어진다.

5.2절

5.2. 베르누이 확률분포

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.

- 따라서 베르누이 시행은 실험의 결과가 s 또는 f 인 확률실험이라고 할 수 있으며, 표본공간은 $\Omega = \{s, f\}$이 된다.

- 이 실험에서 결과가 s일 확률이 p라면 확률의 기본원리에 의하여 실험결과가 f일 확률은 $1 - p$이다.

예 5-1 베르누이 시행의 예

(a) 동전 하나를 던지는 실험
(sol) 결과 : 앞면(Head)/뒷면(Tail)

(b) 대학에 지원한 한 학생의 시험결과
(sol) 결과 : 합격/불합격

(c) 한 공장의 생산제품의 불량여부
(sol) 결과 : 합격품/불량품

(d) 활을 쏘아서 과녁 맞추기
(sol) 결과 : 성공/실패

베르누이 확률변수와 베르누이 확률분포

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, $\boldsymbol{X \sim \mathrm{Bernoulli}(p)}$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 $\boldsymbol{p}$**이다.

베르누이 확률분포의 평균과 분산

4. 베르누이 확률분포의 평균과 분산

① 평균

$$E(X) = \sum x_i p_i = 0 \times (1-p) + 1 \times p = p$$

② 분산

$$E(X^2) = \sum x_i^2 p_i = 0 \times (1-p) + 1 \times p = p$$

$$\sigma^2 x = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$$

5.3절

5.3. 이항분포

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. n개 독립인 확률변수의 합의 평균과 분산

- $X_1, X_2, \cdots, X_n$ 이 서로 독립이며 각각의 평균이 μ이고, 분산이 σ^2이라면 이 확률변수들의 합 $X = \sum_{i=1}^{n} X_i$의 평균과 분산은 다음과 같다.

$$E(X) = n\mu, \qquad Var(X) = n\sigma^2$$

$$\begin{aligned} E(X) &= E\left(\sum X_i\right) = E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) = \mu + \mu + \cdots + \mu = n\mu \end{aligned}$$

$$\begin{aligned} Var(X) &= E\big(X - E(X)\big)^2 = E\left(\sum X_i - n\mu\right)^2 = E\left(\sum (X_i - \mu)\right)^2 \\ &= E\Big(\sum (X_i - \mu)^2 + \sum\sum (X_i - \mu)(X_j - \mu)\Big) = E\left(\sum (X_i - \mu)^2\right) + E\Big(\sum\sum (X_i - \mu)(X_j - \mu)\Big) \\ &= E\left(\sum (X_i - \mu)^2\right) = \sum E(X_i - \mu)^2 = \sum Var(X_i) = n\sigma^2 \end{aligned}$$

- $X_1, X_2, \cdots, X_n$ 이 서로 독립이므로 $i \neq j$이면 $E\big[(X_i - \mu)(X_j - \mu)\big] = 0$

3. 이항분포의 평균과 분산

- 위의 설명 2번을 이용하여 이항분포의 평균과 분산을 다음과 같이 유도할 수 있다.
- $X_1, X_2, \cdots, X_n$ 가 각각 서로 독립인 베르누이 확률변수이고 각각의 확률은 $P_r(X_i = 1) = p$, $P_r(X_i = 0) = q$, $i = 1, 2, \cdots, n$이라고 할 때, 이항확률변수 X는 $X = \sum_{i=1}^{n} X_i$라고 정의할 수 있다.
- 따라서 이항확률변수 X의 평균과 분산은 다음과 같다.

① 평균

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = p + p + \cdots + p = np$$

② 분산

$$Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = pq + pq + \cdots + pq = npq$$

4. 이항분포의 확률계산

- 이항확률분포에서 확률의 계산은 부록V의 [표1]에 의하여 구할 수 있다.
- 표본의 수 $n = 5, 10, 15, 20, 25$인 경우에 있어서 성공의 확률이 p인 이항확률변수 X의 값 a까지의 누적확률을 제시하여 준다.

$$P_r(X \leq a) = \sum_{x=0}^{a} P(x)$$

예 5-2 성공(s)이 나타날 확률이 p이고, 실패(f)가 나타날 확률이 $1-p$인 베르누이 시행을 독립적으로 3번 반복 실시하는 실험에서 확률변수 X를 '성공의 횟수'라고 정의할 때 X의 확률분포를 구하라.

- (sol) 실험에서 나타날 수 있는 모든 가능한 경우는 다음과 같이 8가지이며 각 경우의 확률은 각 시행이 서로 독립적이므로 각 시행 결과들의 확률의 곱과 같다. X의 확률분포는 다음과 같음을 알 수 있다.

실험결과			X	확률
첫째,	둘째,	셋째		
s	s	s	3	p^3
s	s	f	2	$p^2(1-p)$
s	f	s	2	$p^2(1-p)$
f	s	s	2	$p^2(1-p)$
s	f	f	1	$p(1-p)^2$
f	s	f	1	$p(1-p)^2$
f	f	s	1	$p(1-p)^2$
f	f	f	0	$(1-p)^3$
합				1

$$P_r(X=3)=p^3$$

$$P_r(X=2)=3p^2(1-p)$$

$$P_r(X=1)=3p(1-p)^2$$

$$P_r(X=0)=(1-p)^3$$

예 5-3 평균타율이 3할인 야구선수가 어느 경기에서 10회 타석에 섰을 때 매 타석마다 안타 또는 아웃이라고 한다. 이 선수가 때린 안타의 수의 분포를 구하고, 안타를 정확하게 3개 때릴 확률과 안타를 4개의 이하 때릴 확률, 그리고 이 선수가 때린 안타의 수에 대한 평균과 분산을 구하라.

(sol) 확률변수 X를 '이 선수가 때린 안타의 수'로 정의할 때, $X \sim B(10, 0.3)$이므로 X의 확률분포함수는 다음과 같다.

$$P_r(X = k) = \binom{10}{k}(0.3)^k(0.7)^{10-}\ , \qquad k - 0, 1, 2, \cdots, 10$$

정확하게 3개의 안타를 때릴 확률과 안타를 네 개 이하로 때릴 확률은 부록의 표를 이용하여 다음과 같이 계산된다.

$$P_r(X = 3) = \binom{10}{3}(0.3)^3(0.7)^7,$$
$$P_r(X = 3) = P_r(X \le 3) - P_r(X \le 2) = 0.650 - 0.383 = 0.267$$
$$P_r(X \le 4) = 0.850$$

안타의 수의 평균과 분산은 각각 다음과 같다.

$$E(X) = np = 10 \times 0.3 = 3, \quad Var(X) = npq = 10 \times 0.3 \times 0.7 = 2.1$$

예 5-4 제품 한 상자에 정확하게 5%의 불량품이 들어 있다고 한다. 이 상자에서 10개의 제품을 임의로 관찰하여 불량품이 2개 이상인 경우에는 이 상자를 폐기처분한다고 할 때 이 상자가 폐기처분될 확률을 구하라.

(sol) 확률변수 X가 10개의 관찰치 중에서 불량품의 수를 나타낸다고 할 때, 각 관찰치가 불량품일 확률은 0.05이므로 X의 확률분포는 다음과 같다.

$$P_r(X = k) = \binom{10}{k}(0.05)^k(0.95)^{10-k}, \qquad k - 0, 1, 2, \cdots, 10$$

이 상자가 폐기될 확률은 부록의 표를 이용하여 다음과 같이 계산된다.

$$P_r(X \geq 2) = 1 - P_r(X < 2) = 1 - P_r(X \leq 1) = 1 - 0.914 = 0.086$$

예 5-5 어느 가게에 오후 5시부터 7시 사이에 들어오는 손님들 중 30%가 소품목 계산대를 이용한다. 5명의 손님을 랜덤하게 선택했을 때, 다음 확률을 구하여라.

(sol) 확률변수 X를 5명의 손님 중 소품목 계산대를 이용하는 손님의 수라고 할 때, 들어오는 손님이 소품목 계산대를 이용할 확률은 0.3이므로, X의 확률분포는 다음과 같다.

$$P_r(X = k) = \binom{5}{k}(0.3)^k(0.7)^{5-k}, \qquad k - 0, 1, 2, \cdots, 5$$

a. 소품목 계산대를 이용하는 손님이 2명일 확률

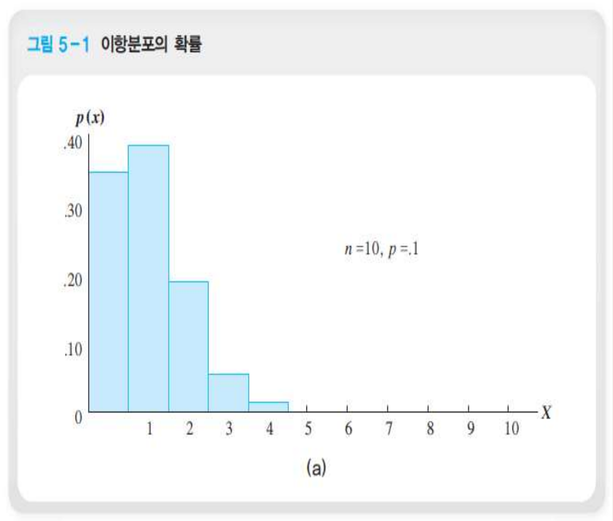
(sol) $P_r(X = 2) = P_r(X \leq 2) - P_r(X \leq 1) = 0.837 - 0.528 = 0.309$

b. 소품목 계산대를 이용하는 손님이 1명 이하일 확률

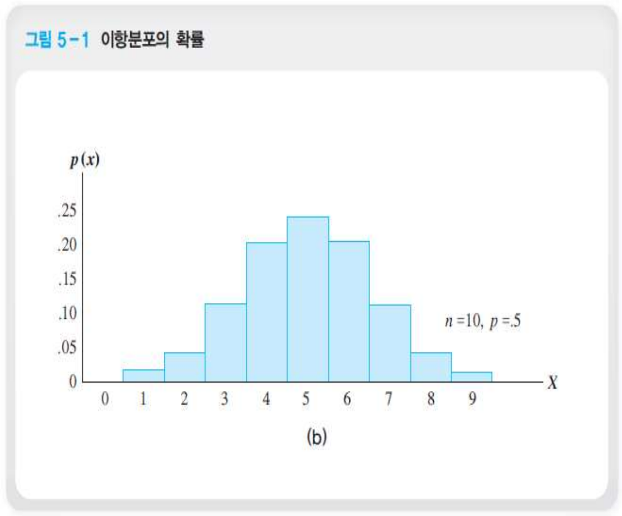
(sol) $P_r(X \leq 1) = 0.528$

5. 이항분포의 그래프

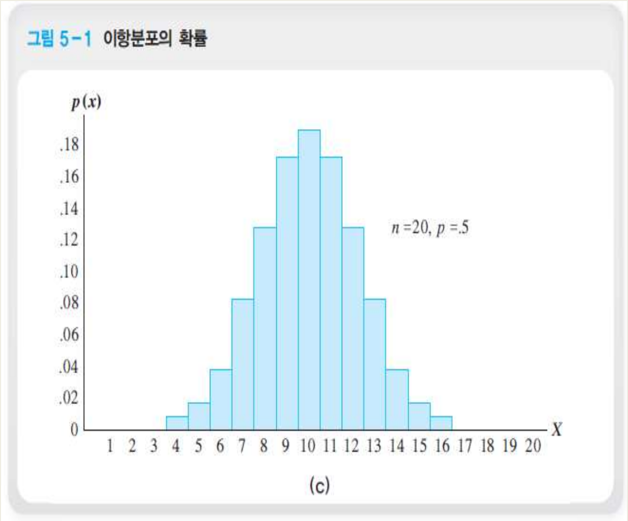
- 이항분포는 성공확률 p가 0.5이면 평균 $\mu = np = \frac{n}{2}$을 중심으로 좌우대칭인 분포를 가지며, p와 n의 크기에 따라 모양이 결정된다. 세 가지 경우에 대한 이항분포의 그림은 다음과 같다.



(a)



(b)



(c)

5.4절

5.4. 기하분포

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 p**이다.

❑ 2. 기하분포의 평균과 분산

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

$$E(X) = \sum kq^{k-1}p = p + 2qp + 3q^2p + \cdots$$

$$qE(X) = qp + 2q^2p + 3q^3p + \cdots$$

$$\Longrightarrow \quad (1-q)E(X) = p + qp + q^2p + q^3p + \cdots = \frac{p}{1-q} = 1$$

$$\therefore \quad E(X) = \frac{1}{p}$$

2. 기하분포의 평균과 분산

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

$$E(X^2) = \sum k^2 q^{k-1} p = p + 4qp + 9q^2p + 16q^3p + \cdots$$
$$qE(X^2) = qp + 4q^2p + 9q^3p + 16q^4p + \cdots$$
$$(1-q)E(X^2) = p + 3qp + 5q^2p + 7q^3p + 9q^4p + \cdots$$
$$E(X^2) = 1 + 3q + 5q^2 + 7q^3 + 9q^4 + \cdots$$
$$qE(X^2) = q + 3q^2 + 5q^3 + 7q^4 + 9q^5 + \cdots$$
$$(1-q)E(X^2) = 1 + 2q + 2q^2 + 2q^3 + 2q^4 + \cdots = 1 + \frac{2q}{1-q} \quad \Longrightarrow \quad E(X^2) = \frac{1}{p} + \frac{2q}{p^2}$$

$$\therefore \quad Var(X) = \frac{1}{p} + \frac{2q}{p^2} - \frac{1}{p^2} = \frac{p + 2q - 1}{p^2} = \frac{q}{p^2}$$

예 5-6 매주 발매하여 주말에 추첨하는 복권에서 당첨될 확률이 20%라고 한다. 한 사람이 복권을 당첨될 때까지 매주 하나씩 계속하여 산다고 할 때, 이 사람이 5번째 주에 가서 처음 당첨될 확률은 얼마인가?

(sol) 확률변수 X를 이 사람이 복권에 처음 당첨될 때까지 복권을 산 횟수로 정의하면 X의 분포는 다음과 같다.

$$P_r(X = k) = (0.8)^{k-1} \cdot 0.2, \qquad k = 1, 2, 3, \cdots$$

따라서 5번째에 처음 당첨될 확률은 다음과 같이 계산된다.

$$P_r(X = 5) = (0.8)^4 \cdot 0.2 = 0.08192$$

예 5-7 주사위 1개를 6이 나올 때까지 던진다고 할 때 다음을 계산하라.

- (sol) 확률변수 X를 주사위를 던져서 6이 처음 나오기 전까지의 던진 횟수라고 할 때, 매번 시행에서 6이 나올 확률이 1/6 이므로 $X \sim \text{Geometric}(1/6)$이다.

- a. 5번째에 6이 나올 확률

- (sol) $P_r(X = 5) = q^4 p = \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6}$

- b. 6이 나올 때까지의 평균실행횟수

- (sol) $E(X) = \frac{1}{6} = \frac{1}{\frac{1}{6}} = 6$

끝~~ ♥♥