

# 확률 및 통계학

- 5장 이산형 확률분포(1) -

명 세인([sein@pel.smuc.ac.kr](mailto:sein@pel.smuc.ac.kr))

상명대학교 프로토콜공학연구실

# 목 차

---

- 이산형 확률분포
- 이산형 균일분포
- 이항 분포
- 다항 분포

# 이산형 확률분포

---

- 의미

- 이산형 자료의 확률 분포

- 모든 확률이 동일할 때: 이산형 균일 분포
- 베르누이 시행의 확률분포: 이항 분포
- 베르누이 시행이 아니며, 독립 시행 확률분포: 다항 분포
- 베르누이 시행이 아니며, 비 복원 시행 확률분포: 초기하분포
- 특정 범위 안에 사건이 몇 번 발생하는지를 표현: 포아송분포

# 이산형 균일분포

---

- 의미

- 이산형 확률변수가 취하는 각 확률이 모두 같은 확률

- 정의

- 이산형 확률변수  $X$ 의 확률 값  $x_1, x_2, \dots, x_k$ 의 각 값의 확률이 동일 하다면 이산형 균일분포(Discrete Uniform Distribution)
  - $f(x; k) = \frac{1}{k}, x = x_1, x_2, \dots, x_k$
  - 균일 분포는 모수  $k$ 에 종속됨을 나타냄

# 이산형 균일분포

---

- 예제 5.1

- 40와트, 60와트, 75와트, 100와트의 전구가 들어있는 상자에서 임의로 하나의 전구를 꺼낼 때, 표본공간  $S = \{40, 60, 75, 100\}$ 의 각 원소는  $\frac{1}{4}$ 의 발생 확률
  - $X$ 의 확률분포  $f(x; 4) = 1/4, x = 40, 60, 75, 100$ 인 균일분포

- 예제 5.2

- 하나의 주사위를 던지는 실험 표본공간  $S = \{1, 2, 3, 4, 5, 6\}$ 의 각 원소의 발생 확률은  $1/6$ 
  - $X$ 의 확률분포  $f(x; 6) = 1/6, x = 1, 2, 3, 4, 5, 6$

# 이산형 균일분포

---

- 정리: 이산형 균일 분포의 평균과 분산

- 평균:  $\mu = \frac{1}{k} \sum_{i=1}^k x_i$

- 분산:  $\sigma^2 = \frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2$

- 증명

- $\mu = E(X) = \sum_{i=1}^k x_i f(x; k) = \sum_{i=1}^k \frac{x_i}{k} = \frac{1}{k} \sum_{i=1}^k x_i$

- $\sigma^2 = E[(X - \mu)^2] = \sum_{i=1}^k (x_i - \mu)^2 f(x; k) = \frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2$

# 이산형 균일분포

---

- 예제 5.3

- 예제 5.2에서 확률변수  $X$ 의 평균과 분산

- 평균:  $\mu = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{1+2+3+4+5+6}{6} = 3.5$

- 분산:  $\sigma^2 = \frac{1}{6} \sum_{i=1}^6 (x_i - 3.5)^2$   
 $= \frac{1}{6} [(1 - 3.5)^2 + (2 - 3.5)^2 + \dots + (6 - 3.5)^2] = 35/12$

# 이항 분포

---

- 의미

- $n$ 번의 베르누이 시행에서의 성공 횟수  $X$ 를 이항 확률 변수
- 이항 확률변수(Binomial Random Variable)의 확률분포는 이항 분포(Binomial Distribution)
- 베르누이 과정(Bernoulli Process)
  - 실험은  $n$ 번의 반복 시행(Trial)
  - 각 시행의 결과는 성공 또는 실패 두 가지 중 하나
  - $p$ 로 표시되는 성공확률은 매 시행마다 일정
  - 각 시행은 서로 독립



# 이항 분포

---

- 정의: 이항 분포

- 성공확률이  $p$ , 실패확률이  $q = 1 - p$ 인 베르누이 시행의  $n$ 회 독립시행에서 성공의 횟수  $x$ 를 나타내는 이항확률변수  $X$ 의 확률분포

- $b(x; n, p) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n$

- $\binom{n}{x} = nCx = \frac{n!}{x!(n-x)!}$

- 이항 분포의 누적

- 확률분포의 조건:  $p + q = \sum_{x=0}^n b(x; n, p) = 1$
- 이항분포 누적 합  $B(r; n, p) = \sum_{x=0}^r b(x; n, p) \rightarrow$  부록 A.1

# 이항 분포

---

- 예제 5.4

- 어떤 종류의 부품이 충격실험에서 충격을 견딜 확률은  $3/4$   
4개의 부품에 대해 실험할 때, 2개가 충격을 견딜 확률

- $p = \frac{3}{4}$

- $b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2} \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 = \left(\frac{4!}{2! 2!}\right) \left(\frac{3^2}{4^4}\right) = \frac{27}{128}$

# 이항 분포

---

- 예제 5.5

- 빈혈환자가 회복될 확률  $p = 0.4$ , 15명이 빈혈에 걸렸을 경우 다음 확률

- 적어도 10명이 회복될 확률

- $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^9 b(x; 15, 0.4) = 1 - 0.9662$

- 3명에서 8명 사이의 사람이 회복될 확률

- $P(3 \leq X \leq 8) = \sum_{x=3}^8 b(x; 15, 0.4) =$

- $\sum_{x=0}^8 b(x; 15, 0.4) - \sum_{x=0}^2 b(x; 15, 0.4) = 0.9050 - 0.0271 = 0.8779$

- 정확히 5명이 회복될 확률

- $P(X = 5) = b(5; 15, 0.4) = \sum_{x=0}^5 b(x; 15, 0.4) - \sum_{x=0}^4 b(x; 15, 0.4)$   
 $= 0.4032 - 0.2173 = 0.1859$

# 이항 분포

---

- 정리: 이항분포의 평균과 분산

- $P(X = r) = \binom{n}{r} p^r q^{n-r}$ 이며  $p + q = 1, r = 0, 1, 2, \dots, n$ 를 따르는 이항확률 변수  $X$ 의 평균과 분산

- 평균:  $\mu = np$

- $$E(X) = \sum_{r=1}^n r \binom{n}{r} p^r q^{n-r} = \sum_{r=1}^n n \binom{n-1}{r-1} p^r q^{n-r} = np \sum_{r=1}^n \binom{n-1}{r-1} p^{r-1} q^{n-r} = np(p + q)^{n-1} = np$$

# 이항 분포

- 정리: 이항분포의 평균과 분산

- $P(X = r) = \binom{n}{r} p^r q^{n-r}$ 이며  $p + q = 1, r = 0, 1, 2, \dots, n$ 를 따르는 이항확률 변수  $X$ 의 평균과 분산

- 분산:  $\sigma^2 = npq$

- $$\begin{aligned}\sigma^2 &= E(X^2) - [E(X)]^2 = \sum_{r=0}^n r^2 \binom{n}{r} p^r q^{n-r} - (np)^2 \\&= \sum_{r=0}^n (r^2 - r + r) \binom{n}{r} p^r q^{n-r} - (np)^2 \\&= \sum_{r=0}^n r(r-1) \binom{n}{r} p^r q^{n-r} + \sum_{r=0}^n r \binom{n}{r} p^r q^{n-r} - (np)^2 \\&= \sum_{r=2}^n n(n-1) \binom{n-2}{r-2} p^r q^{n-r} + np - (np)^2 \\&= n(n-1)p^2 \sum_{r=2}^n \binom{n-2}{r-2} p^{r-2} q^{n-r} + np - (np)^2 \\&= n(n-1)p^2 (p+q)^{n-2} + np - (np)^2 = np(1-p) = npq\end{aligned}$$

# 이항 분포

---

- 예제 5.7

- 어느 시골의 우물 중 30%는 불순물이 포함됨을 경험적으로 알고, 전체 우물 중 10개의 우물만 선정하여 검사할 때

- 정확히 3개의 우물에 불순물이 있을 확률

- $b(3; 10, 0.3) = P(X = 3) = \sum_{x=0}^3 b(x; 10, 0.3) - \sum_{x=0}^2 b(x; 10, 0.3)$   
 $0.6496 - 0.3828 = 0.2668$

- 3개를 초과한 우물에 불순물이 있을 확률

- $P(X > 3) = 1 - 0.6496 = 0.3504$

# 이항 분포

---

- 예제 5.8

- 예제 5.5의 이항확률변수에 대한 평균과 분산을 구하고, 체비셰프 정리를 사용하여 구간  $\mu \pm 2\sigma$ 의 의미는
  - $\sigma^2 = (15)(0.4)(0.6) = 3.6$ 
    - $\mu = 6, \sigma = 1.897$
  - $\mu \pm 2\sigma \rightarrow 6 \pm (2) \cdot (1.897)$  즉 2.206과 9.794사이
    - 자료가 이산형 이므로 15명의 환자 중, 3명에서 9명 사이의 사람이 회복될 확률은  $\frac{3}{4}$  이상

# 이항 분포

---

- 예제 5.9

- 예제 5.7에서 '30%가 오염' 은 추측이며, 10개의 우물을 검사했을 때 6개의 우물에서 불순물이 발견되면

- $$P(X \geq 6) = \sum_{x=6}^{10} b(x; 10, 0.3) = 1 - \sum_{x=0}^5 b(x; 10, 0.3)$$
$$= 1 - 0.9527 = 0.0473 \rightarrow 4.7\%$$

- 실제 검사 시 6개의 우물이 오염된 경우 30%오염 이라는 가정에 의문이 생길 수 있음



# 다항 분포

- 의미

- 다항 실험(Multinomial Experiment)의 확률분포, 각 시행의 결과가 두 가지 이상이며(베르누이 시행이 아닌) 각 시행은 독립

- 정의

- 각 시행에서  $p_1, p_2, \dots, p_k$ 의 확률로  $k$ 개의 결과  $E_1, E_2, \dots, E_k$  중 하나가 발생,  $n$ 번의 독립시행에서 각각  $E_1, E_2, \dots, E_k$ 의 발생 횟수를 나타내는 확률변수  $X_1, X_2, \dots, X_k$ 의 확률분포
  - $f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n) = \binom{n}{x_1, x_2, \dots, x_k} (p_1^{x_1} p_2^{x_2} \dots p_k^{x_k})$
  - $\sum_{i=1}^k x_i = n$ 이고,  $\sum_{i=1}^k p_i = 1$

# 이항 분포

## • 예제 5.10

- 항공기 이착륙 상황에 대한 이상적인 조건을 알아보기 위해 컴퓨터 시뮬레이션 수행, 3개의 활주로와 각 활주로가 사용될 확률이 아래와 같이 알려졌을 때

- 활주로 1:  $p_1 = \frac{2}{9}$ , 활주로 2:  $p_2 = \frac{1}{6}$ , 활주로 3:  $p_3 = \frac{11}{18}$

- 임의로 도착하는 6대의 비행기가 다음과 같이 활주로에 도착할 확률

- 활주로 1: 2대, 활주로 2: 1대, 활주로 3: 3대

- $$f\left(2,1,3; \frac{2}{9}, \frac{1}{6}, \frac{11}{18}, 6\right) = \binom{6}{2,1,3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3 = \frac{6!}{2!1!3!} \cdot \frac{2^2}{9^2} \cdot \frac{1}{6} \cdot \frac{11^3}{18^3}$$
$$= 0.1127$$

---

감사합니다!