2016/02/03

[2016-동계세미나-확률 기초]

# Chapter 5 이산형 확률분포

# 5.1~5.3

이 부 형(boohyung@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

# Contents

- 이산형 균일분포
- 이항분포
- 다항분포

# 이산형 균일분포

- 이산형 균일분포(discrete uniform distribution)
  - 확률변수 X가 $x_1, x_2, x_3, \ldots x_k$의 각 값을 취할 확률이 동일하다면, 이산형 확률분포는

$$f(x;k) = \frac{1}{k}, \qquad x = x_1, x_2, x_3, \ldots x_k$$

  - 예) 하나의 주사위를 던지는 실험에서 각각 하나의 면이 나올 확률

$$f(x;6) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6$$

# 이산형 균일분포

- 이산형 균일분포(discrete uniform distribution)
  - 이산형 균일분포의 평균과 분산은

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i \text{ 와} \qquad \sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

  - 증명

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

$$\sigma^2 = \sum_{i=1}^{k} f(x)\,(x_i - \mu)^2, \qquad f(x) = \frac{1}{k}$$

# 이항분포

- 베르누이 과정(Bernoulli process)
  - 실험은 n번의 반복시행으로 구성됨
  - 각 시행의 결과는 성공이나 실패의 두 가지 중 하나가 됨
  - $p$로 표시되는 성공확률은 매 시행마다 일정
  - 각 시행은 서로 독립

  - 예제1) 한 벌의 카드에서 연속적으로 $n$번 카드를 뽑아서 그 카드가 하트라면 '성공', 아니라면 '실패' 라고 할 때 매 시행의 '성공' 일 확률
    (조건: 복원 추출, 다음 번 추출 전에 충분히 섞여짐)

    '성공' 확률 $p = \frac{1}{4}$ 로 일정하고 각 시행은 서로 독립

# 이항분포

- 베르누이 과정(Bernoulli process)
  - 실험은 n번의 반복시행으로 구성됨
  - 각 시행의 결과는 성공이나 실패의 두 가지 중 하나가 됨
  - $p$로 표시되는 성공확률은 매 시행마다 일정
  - 각 시행은 서로 독립

  - 예제2) 한 공격자가 타겟 시스템에 n번의 공격을 한다고 할 때, 성공할 확률은 $\frac{1}{n}$ (단, 각각의 공격은 독립)

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 이항확률변수(binomial random variable): $n$번의 베르누이 시행에서의 성공횟수 $X$
  - 성공확률이 $p$, 실패확률이 $q = 1 - p$인 베르누이 시행의 $n$회 독립시행에서 성공의 횟수를 나타내는 이항확률변수 $X$의 확률분포는

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, 3, \dots, n$$

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 예제1) 어떤 핸드폰이 충격실험에서 충격을 견딜 확률이 ¾ 이다. 4개의 핸드폰에 대해 실험했을 때 정확하게 2개가 충격을 견딜 확률을 구하여라.

    풀이) 각 시행은 독립이고 일정한 확률을 가지기 때문에 베르누이 시행이고, 이항분포로 나타낼 수 있음

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

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 예제2) 어떤 공격자가 시스템을 공격한다고 할 때, 공격을 견딜 확률이 $\frac{3}{4}$ 이다. 4개 부분으로 이루어진 한 시스템에서 시스템이 작동하지 않을 확률은?
    (단, 각 공격은 독립이고, 4개 중 2개의 부분이 파괴되면 시스템이 작동하지 않음, 시스템이 작동하지 않으면 공격자는 공격을 멈춤)

    풀이) 예제1과 같음

# 이항분포

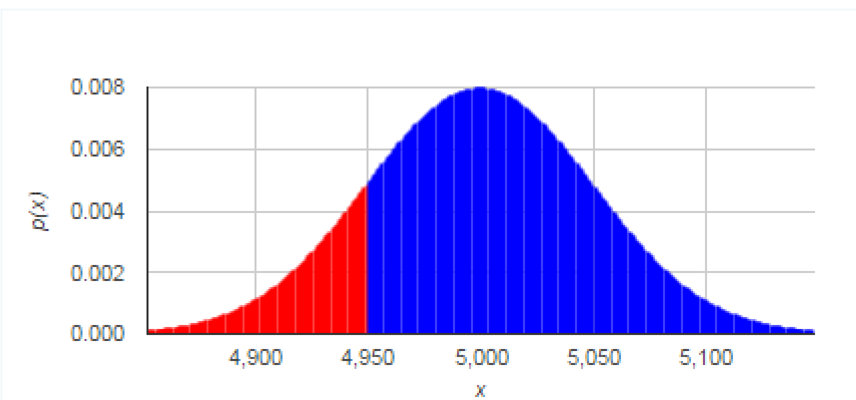
- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 이항분포의 누적합 $B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$
    - 누적이항분포표 이용(부록 A.1)
      - $p = 0.1 \sim 0.9$
      - $n = 1 \sim 20$
  - 예제) 빈혈환자가 회복될 확률이 0.4라 하자. 15명이 빈혈에 걸렸을 때 다음 확률을 누적이항분포표를 이용하여 구하여라.
    (a) 적어도 10명이 회복될 확률

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

$$= 1 - 0.9662 = 0.0338$$

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 이항분포의 누적합 $B(r;n,p) = \sum_{x=0}^{r} b(x;n,p)$
    - 누적이항분포표 이용(부록 A.1)
      - $p = 0.1 \sim 0.9$
      - $n = 1 \sim 20$
    - 이항분포 그래프$(\mathrm{e.g.},\ r = 4.950)$



# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - 이항분포의 누적합 $B(r;n,p)=\sum_{x=0}^{r} b(x;n,p)$
    - 누적이항분포표 이용(부록 A.1)
      - $p=0.1\sim0.9$
      - $n=1\sim20$
  - 예제) 빈혈환자가 회복될 확률이 0.4라 하자. 15명이 빈혈에 걸렸을 때 다음 확률을 누적이항분포표를 이용하여 구하여라.
    (b) 정확히 5명이 회복될 확률
    $P(X=5)=b(5;15,0.4)$
    $$=\sum_{x=0}^{5} b(x;15,0.4)-\sum_{x=0}^{4} b(x;15,0.4)=0.4032-0.2173$$

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - $P(X=r) = \binom{n}{r} p^r q^{n-r}$ (단, $p+q=1, r=0,1,,2,\ldots,n$)를 따를 때, 이항분포 $b(x;n,p)$의 평균과 분산은
  $$\mu = np, \qquad \sigma^2 = npq$$
  - 예) 하나의 주사위를 120번 던져서 1의 눈이 나오는 횟수는 직관상 몇 번 정도일까? $20$번; $120 \text{ X} \frac{1}{6} = 20$
    - 여기서 “직관상 몇 번 정도” 는 평균 $\mu$를 의미함
    - $n = 120, p = \frac{1}{6}$

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - $P(X=r)= \binom{n}{r} p^r q^{n-r}$ (단, $p+q=1, r=0,1,,2,\ldots,n$)를 따를 때, 이항분포 $b(x;n,p)$의 평균과 분산은

$$\mu = np, \qquad \sigma^2 = npq$$

  - 예제1) 빈혈환자가 회복될 확률이 0.4라 하자. 15명이 빈혈에 걸렸을 때 평균과 분산을 구하고, 체비셰프 정리를 이용해 $\mathrm{\mu} \pm 2\sigma$의 의미를 말하라.

풀이) $n=15, p=0.4$

$$\mu = 15 \ * 0.4, \qquad \sigma^2 = 15 \ * 0.4 \ * 0.6$$

$\sigma = 1.897$이므로 $6 \pm 1.897 = 2.206 \sim 9.794$
따라서, 15명의 환자 중 회복될 사람의 수가 2.206~9.974 사이(3명~9명)일 확률은 적어도 75% 이상

# 이항분포

- 이항분포(binomial distribution): 이항확률변수의 확률분포
  - $P(X = r) = \binom{n}{r} p^r q^{n-r}$ (단, $p + q = 1, r = 0, 1, , 2, \dots, n$)를 따를 때, 이항분포 $b(x; n, p)$의 평균과 분산은
    $$\mu = np, \qquad \sigma^2 = npq$$
  - 예제2) 한 침입탐지시스템(IDS)에서 오탐율이 $p$로 정해져 있을 때, $n$번의 침입(각각의 침입은 독립관계) 중 $\mu - k\sigma$번~$\mu + k\sigma$번 오탐지될 확률은 적어도 $1 - \frac{1}{k^2}$ 이다.

    (<-체비셰프 정리)

    풀이) $k$를 임의로 정하여 IDS의 오탐지율을 구할 수 있음
    → 그래프로 표현하여 IDS의 성능을 평가하는 지표로 사용 가능

# 다항분포

- 다항분포(multinomial distribution): 여러 개의 값을 가질 수 있는 독립 확률변수들에 대한 확률분포
  - 이항분포의 일반화
  - 각 시행에서 $p_1, p_2, p_3, \dots, p_k$의 확률로 k개의 결과 $E_1, E_2, E_3, \dots, E_k$ 중 어느 하나가 발생한다면, n번의 독립시행에서 각각 $E_1, E_2, E_3, \dots, E_k$의 발생횟수를 나타내는 확률변수 $X_1, X_2, X_3, \dots, X_k$의 확률분포는

$$f(x_1, x_2, x_3, \dots, x_k; p_1, p_2, p_3, \dots, p_k, n) = \binom{n}{x_1, x_2, x_3, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

(단, $\sum_{i=1}^{k} x_i = n, \sum_{i=1}^{k} p_i = 1$)

# 다항분포

- 다항분포(multinomial distribution): 여러 개의 값을 가질 수 있는 독립 확률변수들에 대한 확률분포
  - 예제) 한 번의 실험에서 $k$개의 사건이 발생하는 경우를 생각하여 보자. $k$가 3인 경우의 예를 들어보면, 주머니 안에 흰 공 3개, 검은 공 4개, 노란 공 5개가 들어있는 경우에는 결과가 세 가지이므로 삼항분포가 된다. 이 주머니에서 6개의 공을 복원 추출할 때, 흰 공의 수를 $X_1$, 검은 공의 수를 $X_2$, 그리고 노란 공의 수를 $X_3$라 할 때 $(X_1, X_2, X_3)$의 분포를 조사하라.

# 다항분포

- 다항분포(multinomial distribution): 여러 개의 값을 가질 수 있는 독립 확률변수들에 대한 확률분포
  - 풀이) $P(\text{흰 공}) = \frac{3}{12}, P(\text{검은 공}) = \frac{4}{12}, P(\text{노란 공}) = \frac{5}{12}$
    흰 공이 $X_1$ 개, 검은 공이 $X_2$ 개, 노란 공이 $X_3$ 개일 확률은 $\left(\frac{3}{12}\right)^{x_1}\left(\frac{4}{12}\right)^{x_2}\left(\frac{5}{12}\right)^{x_3}$ 이며, 이러한 경우의 수는 $\frac{6!}{x_1!x_2!x_3!}$
    - $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{6!}{x_1!x_2!x_3!}\left(\frac{3}{12}\right)^{x_1}\left(\frac{4}{12}\right)^{x_2}\left(\frac{5}{12}\right)^{x_3}$
      $(x_1 = 0, 1, 2, 3, x_2 = 0, 1, 2, 3, 4, x_3 = 0, 1, 2, 3, 4, 5$ 이며 $x_1 + x_2 + x_3 = 6)$