2026/10/16, 2026 확률 기초 세미나

# 이공학도를 위한 확률 및 통계학

## - 5장 이산형 확률분포 -

김 민 식 (minsik@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송 분포(Poisson Distribution)

# 이산형 확률분포

- 정의
  - 확률변수가 가질 수 있는 값이 셀 수 있거나 나열 가능한 경우, 각 값에 확률을 대응시킨 분포
    - e.g., 동전을 여러 번 던졌을 때 앞면이 나온 횟수
- 특징
  - 서로 다른 실험도 동일한 형태의 확률분포를 따를 수 있음
  - 가능한 모든 값의 확률을 더하면 1임

# 이산형 확률분포

- 종류
  - 이산형 균일분포 (Discrete Uniform Distribution)
  - 이항분포 (Binomial Distribution)
  - 다항분포 (Multinomial Distribution)
  - 초기하 분포 (Hypergeomertic Distribution)
  - 음이항 분포 (Negative Binomial Distribution)
  - 기하 분포 (Geometric Distribution)
  - 포아송 분포 (Poisson Distribution)

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 이산형 균일분포

- 정의

*모수(Parameter):
확률분포의 형태를 결정하는 고정된 값

  - 확률변수가 취할 수 있는 각 값들이 모두 같은 확률을 가지는 이산 확률 분포
    - e.g., 주사위를 한 번 던졌을 때 나오는 눈 1, 2, 3, 4, 5, 6의 각 확률은 $\frac{1}{6}$임

- 수식

*$k$: 확률변수 $X$가 가질 수 있는 가능한 값의 개수

확률변수 $X$가 $x_1, x_2, \dots, x_k$의 각 값을 취할 확률이 동일하다면, 이산형 균일분포는 다음과 같음

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

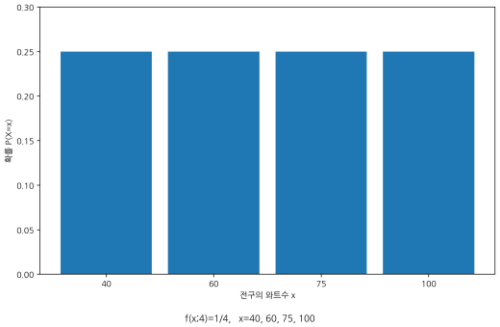
  - 균일분포는 모수 $k$에 종속되어 있음을 나타내기 위해 $f(x)$대신에 $f(x;k)$를 사용함

# 이산형 균일분포

- 예제 5.1

40와트, 60와트, 75와트, 100와트의 전구가 들어 있는 상자에서 임의로 하나의 전구를 꺼낼 때의 확률 분포를 구하시오.

- 표본공간 $S = \{40, 60, 75, 100\}$의 각 원소는 $\frac{1}{4}$의 발생확률을 가짐
- $X$의 확률분포는 $f(x; 4) = \frac{1}{4}, x = 40, 60, 75, 100$인 균일분포임

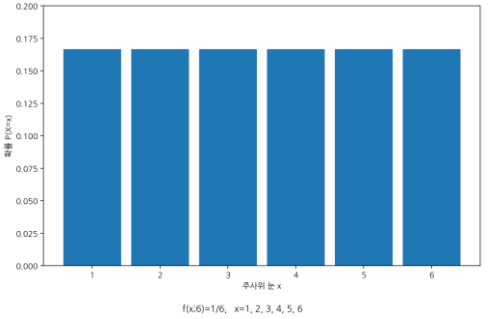



<그림 1> 예제 5.1의 확률분포 그래프

- 예제 5.2

하나의 주사위를 던지는 실험의 확률 분포를 구하시오.

- 표본공간 $S = \{1, 2, 3, 4, 5, 6\}$의 각 원소는 $\frac{1}{6}$의 발생확률을 가짐
- $X$의 확률분포는 $f(x; 6) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6$인 균일분포임




<그림 2> 예제 5.2의 확률분포 그래프

# 이산형 균일분포

- 평균과 분산 (1/2)
  - 정리

이산형 균일분포의 평균과 분산은

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

  - 증명

*$E(X)$: 확률변수 $X$의 기댓값

평균의 정의에 이산형 균일분포의 정의를 대입하면

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

이 됨. 또한 분산 정의에 의하여

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

# 이산형 균일분포

- 평균과 분산 (2/2)
  - 예제 5.3

예제 5.2에서 $X$의 평균과 분산을 구하시오.

- $X$의 확률분포는 $f(x;6)=\frac{1}{6},\ x=1,2,3,4,5,6$임
- $\mu=\frac{1+2+3+4+5+6}{6}=3.5$
- $\sigma^2=\frac{1}{6}[(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2]=\frac{17.5}{6}=\frac{35}{12}=2.92$

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 이항분포

- 베르누이 시행(Bernoulli Trial)
  - 정의
    - 한 번의 시행에서 결과가 성공 또는 실패 중 하나로만 나타나는 무작위 실험
- 베르누이 과정(Bernoulli Process) (1/2)
  - 개요
    - 이항분포는 독립적인 베르누이 시행의 반복에서 성공 횟수를 다루므로, 분포의 성립 과정과 조건을 이해하기 위해 베르누이 과정의 개념이 필요
  - 정의
    - 성공확률 $p$가 일정한 베르누이 시행을 서로 독립적으로 반복하는 확률과정

# 이항분포

- 베르누이 과정 (Bernoulli Process) (2/2)
  - 특징
    - 실험은 $n$번의 반복시행으로 구성됨
    - 각 시행의 결과는 성공이나 실패의 두 가지 중 하나가 됨
    - $p$로 표시되는 성공확률은 매 시행마다 일정함
    - 각 시행은 서로 독립임
  - 예제 5.4

제조과정에서 3개의 제품을 임의로 추출하여 검사하고, 양품과 불량으로 구분하는 베르누이 시행을 표현하시오.

- 불량을 발견하는 경우를 성공이라고 하면, 성공횟수는 0에서 3사이의 값을 취하는 확률변수 $X$가 됨
- 8가지 가능한 결과와 각 경우에 대한 $X$의 값은 다음과 같음

| 결과 ($N$:양품) | $NNN$ | $NDN$ | $NND$ | $DNN$ | $NDD$ | $DND$ | $DDN$ | $DDD$ |
|---|---|---|---|---|---|---|---|---|
| $X$ | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |

- 제품들 중 25%가 불량이라고 한다면, 공정에서 제품들이 독립적으로 추출되므로
- $P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{9}{64}$
- 나머지 결과들에 대해서도 확률을 구할 수 있으므로 $X$의 확률분포는 다음과 같음

# 이항분포

- 정의
  - 성공확률 $p$가 일정하고 서로 독립인 베르누이 시행을 $n$번 반복할 때, 성공 횟수가 따르는 이산 확률분포

*$x$: 성공 횟수
*$n$: 전체 시행 횟수
*$p$: 한 번의 시행에서 성공할 확률

- 수식

성공확률이 $p$, 실패확률이 $q = 1 - p$인 베르누이 시행의 $n$회 독립시행에서 성공의 횟수를 나타내는 이항확률변수 $X$의 확률분포는

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

- 예제 5.5

어떤 종류의 부품이 충격실험에서 충격을 견딜 확률이 $\frac{3}{4}$이다. 4개의 부품에 대해 실험했을 때 정확하게 2개가 충격을 견딜 확률을 구하라.

- 각 실험이 독립이고, 각 실험에 대해 $p = \frac{3}{4}$이므로
- $b\left(2; 4, \frac{3}{4}\right) = \binom{4}{2}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = (6)\left(\frac{3^2}{4^2}\right)\left(\frac{1}{4^2}\right) = \frac{27}{128}$

# 이항분포

- 예제 5.6

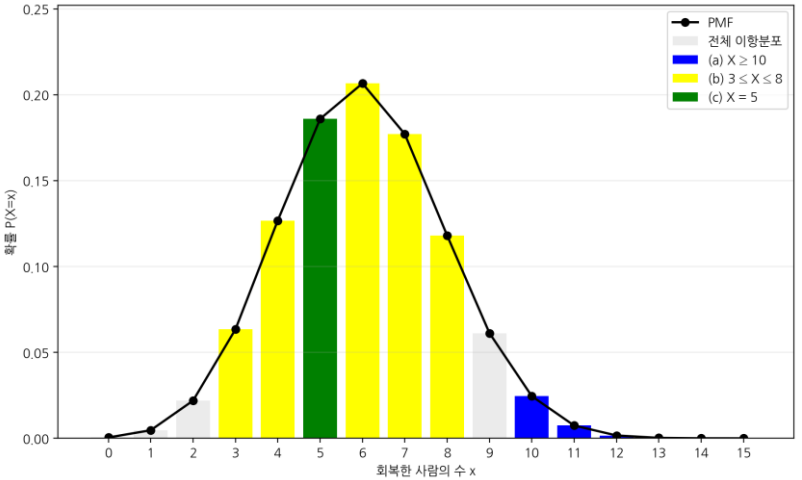
회복 확률이 0.4인 빈혈에 15명이 걸렸을 때 다음 확률을 구하시오.

a. 적어도 10명이 회복될 확률
b. 3명에서 8명 사이의 사람이 회복될 확률
c. 정확히 5명이 회복될 확률

a. 상위 꼬리 확률을 구하는 문제로, 전체 확률 1에서 $X \leq 9$의 누적확률을 빼서 계산함
$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$

b. 구간확률을 구하는 문제로, $X \leq 8$의 누적확률에서 $X \leq 2$의 누적확률을 빼서 계산함
$P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15,0.4) = \sum_{x=0}^{8} b(x; 15,0.4) - \sum_{x=0}^{2} b(x; 15,0.4) = 0.9050 - 0.0271 = 0.8779$

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




<그림 3> 예제 5.6의 이항분포 그래프

# 이항분포

- 예제 5.7

불량품일 확률이 3%인 제조업체가 있다고 했을 때, 다음 확률을 구하시오.
a. 제품 중 20개를 임의로 선택하였을 때, 20개 중 적어도 한 개의 불량품이 있을 확률
b. 한 달에 10번의 납품을 받고, 납품시마다 20개의 장비를 검사한다고 할 때, 적어도 한 대의 불량품이 포함된 납품이 3번 있을 확률

a. 20개의 장비 중 불량품의 수를 $X$라고 하면 $X$는 $b(x;20,0.03)$분포를 따름
따라서, $P(X \geq 1) = 1 - P(X = 0) = 1 - b(0;20,0.03) = 1 - (0.03)^0(1-0.03)^{20-0} = 0.4562$
b. 납품시마다 최소한 한 대의 불량품이 있거나 없기 때문에 납품시마다 하는 검사는 $p = 0.4562$인 베르누이 과정이라고 생각 가능함
   - 납품들이 서로 독립적이라고 가정하고 최소한 한 대의 불량품을 포함하는 납품의 수를 $Y$라고 하면, $Y$는 이항분포 $B(y;10,0.4562)$를 따르게 됨
   - $P(Y=3) = b(3;10,0.4562) = \binom{10}{3}(0.4562)^3(1-0.4562)^7 = 0.1602$

# 이항분포

- 예제 5.8

어느 시골에 있는 우물 중 30% 는 불순물이 있다고 알려져 있다. 10개의 우물을 임의로 선정하여 조사하기로 했을 때, 각 확률을 구하시오.

a. 3개의 우물에 불순물이 있을 확률 (이항분포를 이용)
b. 3개를 초과한 우물에 불순물이 있을 확률
c. 10개의 우물을 검사한 결과 6개 이상의 우물에서 불순물이 발견될 확률

a. 이항분포를 이용하면

$$b(3;10,0.3) = P(X=3) = \sum_{x=0}^{3} b(x;10,0.3) - \sum_{x=0}^{2} b(x;10,0.3) = \binom{10}{3}(0.3)^3(0.7)^7 \approx 0.2668$$

b. $P(X>3) = 1 - \sum_{x=0}^{3} b(x;10,0.3) = 1 - 0.6496 = 0.3504$

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

# 이항분포

- 평균과 분산 (1/3)
  - 정리

이항분포 $b(x;n,p)$의 평균과 분산은

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

  - 증명 (1/2)

$j$번째 시행에서의 결과를 베르누이 확률변수 $I_j$로 표시하면 $I_j$는 $q$와 $p$의 확률로 각각 0과 1의 값을 가짐. 따라서, 이항실험에서 성공횟수는 $n$개의 독립된 베르누이 변수의 합으로 쓰여질 수 있음. 그러므로,

$$X = I_1 + I_2 + \cdots + I_n$$

$I_j$의 평균은 $E(I_j) = (0)(q) + (1)(p) = p$임

$g(A,B) = A$이고 $h(A,B) = B$라 했을 때 $E[A \pm B] = E[A] \pm E(B)$ 를 이용하면,

이항분포의 평균은

$$\mu = E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = p + p + \cdots + p = np$$

# 이항분포

- 평균과 분산 (2/3)
  - 증명 (2/2)

$I_j$의 분산은

$$\sigma_{I_j}^2 = E[(I_j - p)^2] = E(I_j^2) - p^2 = (0)^2(q) + (1)^2(p) - p^2 = p(1-p) = pq$$

서로 독립인 확률변수의 합의 분산은 각 분산의 합이므로 이항분포의 분산은

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = pq + pq + \cdots + pq = npq$$

# 이항분포

- 평균과 분산 (3/3)
  - 예제 5.9

*체비셰프 정리:
평균에서 일정 거리 이내에 포함되는 값의 최소 비율을 보장하는 정리

회복 확률이 $0.4$인 빈혈에 $15$명이 걸렸을 때의 이항확률변수에 대한 평균과 분산을 구하고, 체비셰프 정리를 사용하여 구간 $\mu \pm 2\sigma$의 의미를 설명하라.

- $n = 15, p = 0.4$인 이항실험이므로 분산 공식을 적용하여
$$\sigma^2 = (15)(0.4)(0.6) = 3.6, \sigma = 1.897$$
- 문제에서 요구하는 구간은 $2.206 \sim 9.794$가 됨
- 체비셰프 정리는 빈혈에 걸린 $15$명의 환자 중 회복될 사람의 수가 $2.206$과 $9.794$ 사이, 혹은 자료가 이산형이므로 $3$명에서 $9$명까지 될 확률이 적어도 $\frac{3}{4}$이 됨을 표현함

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 다항분포

- 정의
  - 각 시행에서 가능한 결과가 둘 이상인 다항실험을 $n$번 독립적으로 반복할 때, 각 결과가 발생한 횟수의 결합확률분포

- 수식

*$n$: 전체 시행 횟수
*$k$: 가능한 결과의 종류 수
*$x_i$: $i$번째 결과가 나온 횟수
*$p_i$: 한 번의 시행에서 $i$번째 결과가 나올 확률

각 시행에서 $p_1, p_2, \ldots, p_k$의 확률로 $k$개의 결과 $E_1, E_2, \ldots, E_k$ 중 어느 하나가 발생하면, $n$번의 독립시행에서 각각 $E_1, E_2, \ldots, E_k$ 의 발생횟수를 나타내는 확률변수 $X_1, X_2, \ldots, X_k$의 확률분포는 다음과 같음

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

여기서 $\sum_{i=1}^{k} x_i = n$ 이고, $\sum_{i=1}^{k} p_i = 1$임

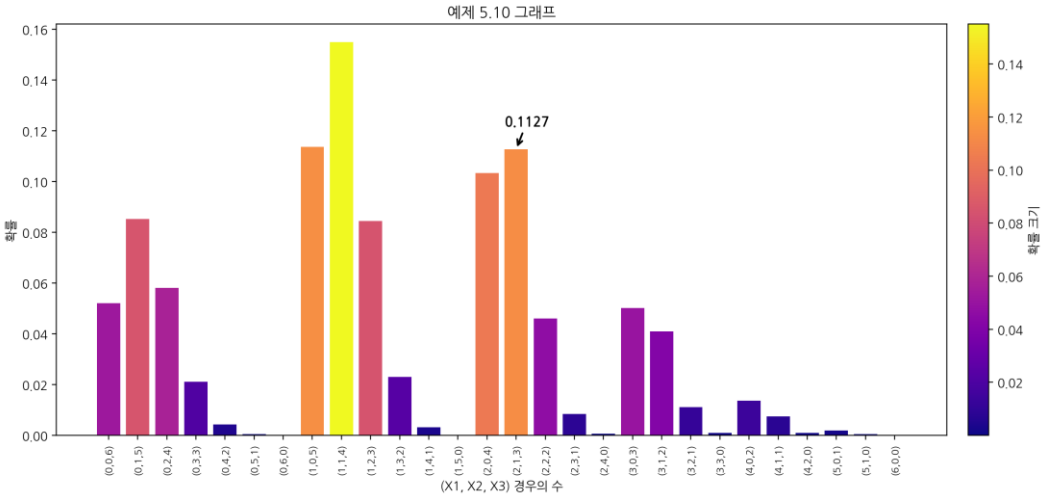
# 다항분포

- 예제 5.10

3개의 활주로가 있는 이 공항에서 각 활주로가 사용될 확률은 다음과 같다.
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 확률은 얼마인가?
활주로 1 : 2대, 활주로 2 : 1대, 활주로 3 : 3대

- 다항분포를 이용하면

$$f(2,1,3;\frac{2}{9},\frac{1}{6},\frac{11}{18},6) = \binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{11}{18}\right)^3 = \frac{6!}{2!\,1!\,3!}\cdot\frac{4}{81}\cdot\frac{1}{6}\cdot\frac{11^3}{18^3} = 0.1127$$




<그림 4> 예제 5.10의 다항분포 그래프

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 초기하분포

- 정의
  - 유한한 모집단에서 비복원추출을 할 때, 표본에 포함된 성공 항목의 개수가 따르는 이산 확률분포
- 특징
  - 크기 $N$의 유한모집단으로부터 크기 $n$의 확률표본을 비복원으로 추출
  - $N$개 중 $k$개는 성공으로, 나머지 $(N-k)$개는 실패로 분류
- 수식

*$N$: 전체 개수
*$k$: 가능한 결과의 종류 수
*$n$: 실제로 뽑는 개수
*$x$: 뽑힌 것 중 성공 개수

$k$개의 성공과 $N-k$개의 실패로 구성된 크기 $N$의 유한모집단에서 크기 $n$인 확률표본을 취할 때, 성공의 개수를 나타내는 초기하확률변수 $X$의 확률분포는 다음과 같음

$$h(x;N,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},\ \max\{0,n-(N-k)\}\le x\le \min\{n,k\}$$

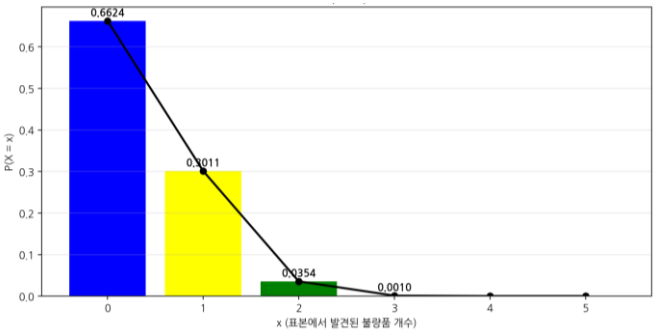
# 초기하분포

## • 예제 5.11

40개의 부품들로 구성된 한 로트에 불량품이 3개 이상 들어 있으면 거부한다고 한다.한 로트에서 임의로 5개의 부품을 취하여 하나의 불량품이라도 발견되면 그 로트를 거부한다고 할 때, 3개의 불량품이 들어 있는 로트에서 5개를 취했을 때 정확히 1개의 불량품이 발견될 확률은 얼마인가?

- $n = 5, N = 40, k = 3$이고 $x = 1$인 초기하분포를 이용하면, 구하고자 하는 확률은 다음과 같음

$$h(1; 40,5,3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011$$




<그림 5> 예제 5.11의 초기하분포 그래프

## • 예제 5.12

100개의 제품 중 12개가 불량품이었다. 100개 중에서 10개를 임의로 뽑았을 때 이 중 3개가 불량일 확률은 얼마인가?

- $n = 10, N = 100, k = 12$이고 $x = 3$인 초기하분포를 이용하면, 구하고자 하는 확률은 다음과 같음

$$h(3; 100,10,12) = \frac{\binom{12}{3}\binom{88}{7}}{\binom{100}{10}} = 0.0807$$

# 초기하분포

- 평균과 분산
  - 정리

초기하분포 $h(x;\ N,n,k)$의 평균과 분산은

$$\mu = \frac{nk}{N}, \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\ (1 - \frac{k}{N})$$

  - 평균 증명

- $E(X) = \sum_{x=0}^{n} x\ \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = k\sum_{x=1}^{n} \frac{(k-1)!}{(x-1)!(k-x)!} \cdot \frac{\binom{N-k}{n-x}}{\binom{N}{n}} = k\sum_{x=1}^{n} \frac{\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}}$
- $y = x - 1$이라 하면, $E(X) = k\sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{N-k}{n-y-1}}{\binom{N}{n}}$
- $\binom{N-k}{n-y-1} = \binom{(N-1)-(k-1)}{n-y-1}$이고 $\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1}$이므로

$$E(X) = \frac{nk}{N}\sum_{y=0}^{n-1} \frac{\binom{k-1}{y}\binom{(N-1)-(k-1)}{n-y-1}}{\binom{N-1}{n-1}} = \frac{nk}{N}$$

# 초기하분포

- 이항분포와의 관계
  - 두 분포 모두 $n$번의 시행(추출)에서 성공의 개수 $X$를 다룸
  - 모집단 크기 $N$에 비해 표본크기 $n$이 매우 작아 $n/N \leq 0.05$ 이면, 비복원추출의 영향이 작아져 초기하분포를 이항분포로 근사할 수 있음
  - $p = \frac{k}{N}$ 이고 $h(x; N, n, k) \approx b(x; n, p)$로 표현 가능
  - 예제 5.13

자동차 타이어 제조업자에 의하면 판매대리점으로 보내기 위해서 선적된 5,000개의 타이어 중에서 1,000개가 약간의 결함을 가지고 있다고 한다.어떤 사람이 이 타이어 중에서 임의로 10개를 구입했을 때 그 중 3개가 결함을 가지고 있을 확률은 얼마인가?

- 표본크기 $n = 10$ 에 대해서 $N = 5{,}000$이 $\frac{n}{N} \leq 0.05$이므로 구하고자 하는 확률은 이항분포를 이용하여 근사적으로 계산이 가능함
- 결함이 있는 타이어를 고를 확률은 $0.2$, 따라서 3개의 결함이 있는 타이어를 구입할 확률은

$$h(3; 5000, 10, 1000) \approx b(3; 10, 0.2) = \sum_{x=0}^{3} b(x; 10, 0.2) - \sum_{x=0}^{2} b(x; 10, 0.2) = 0.8791 - 0.6778 = 0.2013$$

# 초기하분포

- 다변량 초기하분포 (1/2)
  - 정의
    - 여러 종류로 나뉜 유한모집단에서 $n$개를 비복원추출할 때, 각 종류가 몇 개씩 뽑히는지를 나타내는 결합 확률분포
      - e.g., 빨간 공 3개, 파란 공 4개, 초록 공 3개가 있는 상자에서 공 5개를 비복원 추출할 때, 각 색의 공이 몇 개씩 뽑히는지 보는 경우
  - 수식

$N$개의 유한모집단이 $k$개의 집합 $A_1, A_2, \cdots, A_k$로 분할될 수 있고, 각각의 집합이 $a_1, a_2, \cdots, a_k$개의 원소를 가지고 있다면, $n$개의 확률표본 중 $A_1, A_2, \cdots, A_k$ 원소를 나타내는 확률변수 $X_1, X_2, \cdots, X_k$의 확률분포는 다음과 같음

$$f(x_1, x_2, \cdots, x_k; a_1, a_2, \cdots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$$

여기서 $\sum_{i=1}^{k} x_i = n$이고, $\sum_{i=1}^{k} a_i = N$ 임

# 초기하분포

- 다변량 초기하분포 (2/2)
  - 예제 5.14

어느 생물학 실험의 대상이 되는 $10$명의 그룹이 있다. 이 중 $3$명의 혈액형이 O형, $4$명은 A형이고, 나머지 $3$명은 B형이다. $5$명을 임의로 선택할 때, O형인 사람이 $1$명, A형인 사람이 $2$명, B형인 사람이 $2$명일 확률은 얼마인가?

- 임의로 선택한 O형의 수를 $x_1 = 1, 10$명 중 전체 O형의 수를 $a_1 = 3$
- 임의로 선택한 A형의 수를 $x_2 = 2, 10$명 중 전체 A형의 수를 $a_2 = 4$
- 임의로 선택한 B형의 수를 $x_3 = 2, 10$명 중 전체 B형의 수를 $a_3 = 3$으로 표기
- $N = 10$이고 $n = 5$인 초기하분포의 확장을 이용하면

$$f(1,2,2;3,4,3,10,5) = \frac{\binom{3}{1}\binom{4}{2}\binom{3}{2}}{\binom{10}{5}} = \frac{3}{14} = 0.2143$$

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 음이항분포

- 음이항실험
  - 정의
    - 성공확률 $p$가 일정한 독립적인 베르누이 시행을 $k$번째 성공이 발생할 때까지 반복하는 실험
- 음이항분포
  - 정의
    - 음이항 실험에서 $k$번째 성공이 발생할 때까지 필요한 총 시행 횟수 $X$가 따르는 이산 확률분포
      - e.g., 성공확률이 일정한 자유투를 반복할 때, 3번째 성공이 나올 때까지 필요한 시도 횟수
  - 정보보안에서의 활용 예시
    - 모의 피싱 등에서 일정 횟수의 보안 이벤트가 발생하기까지 필요한 시도 횟수 분석

# 음이항분포

- 수식

*$k$: 목표 성공 횟수
*$x$: $k$번째 성공까지의 총 시행 횟수

독립적인 반복시행에서 성공확률이 $p$, 실패확률이 $q$라 하면, $k$번째 성공이 일어날 때까지의 시행횟수인 확률변수 $X$의 확률분포는 다음과 같음(가능한 배열 개수 × 배열 하나의 확률)

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, x = k, k+1, k+2, \ldots$$

- 평균과 분산
  - 정리

음이항분포 $b^*(x;k,p)$의 평균과 분산은

$$\mu = \frac{k}{p}, \qquad \sigma^2 = \frac{k(1-p)}{p^2}$$

# 음이항분포

- 평균과 분산
  - 증명 - 평균

*기대값의 선형성(Linearity of Expectation)
여러 확률변수의 합의 기댓값은 각 확률변수의 기댓값의 합과 같음
$E(X_1 + X_2 + \cdots + X_k) = E(X_1) + E(X_2) + \cdots + E(X_k)$

- $X_i$를 $i-1$번째 성공 이후 $i$번째 성공이 나올 때까지 필요한 시행 횟수라고 정의함
- $k$번째 성공까지 필요한 전체 시행 횟수는
$$X = X_1 + X_2 + \cdots + X_k$$
- 성공 한 번이 나올 때까지 필요한 평균 시행 횟수를 $E(X_i)$라 하면, 한 번 시행 후 실패할 확률 $q$로 다시 같은 상황이 되므로
$$E(X_i) = 1 + qE(X_i)$$
- 정리하면
$$E(X_i)(1-q) = 1, \qquad 1-q = p, \qquad E(X_i) = \frac{1}{p}$$
- 기댓값의 선형성을 적용하면
$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_k) = \frac{1}{p} + \frac{1}{p} + \cdots + \frac{1}{p}$$
- 따라서 음이항분포의 평균은 $\mu = \frac{k}{p}$

# 음이항분포

- 평균과 분산
  - 증명 - 분산

- 앞에서 구한 각 $X_i$의 기댓값은 $E(X_i) = \frac{1}{p}$
- 첫 시행에서 성공하면 $X_i = 1$, 실패하면 1회를 시행한 뒤 다시 처음과 같은 상황이 반복되므로

$$E\left(X_i^2\right) = p \cdot 1^2 + qE[(1 + X_i)^2]$$

- 전개하면

$$E\left(X_i^2\right) = p + q\left(1 + 2E(X_i) + E\left(X_i^2\right)\right)$$

- 여기에 $p + q = 1, E(X_i) = \frac{1}{p}$를 넣어서 정리하면

$$E\left(X_i^2\right) = \frac{2 - p}{p^2}$$

- 이제 분산 공식 $\sigma_i^2 = E\left(X_i^2\right) - [E(X_i)]^2$에 대입

$$\sigma_i^2 = \frac{2 - p}{p^2} - \frac{1}{p^2} = \frac{1 - p}{p^2}$$

- 각 $X_i$가 서로 독립이므로 분산을 더해서

$$\sigma^2 = \sigma_1^2 + \cdots + \sigma_k^2 = k\frac{1 - p}{p^2}$$

# 음이항분포

- 예제 5.15

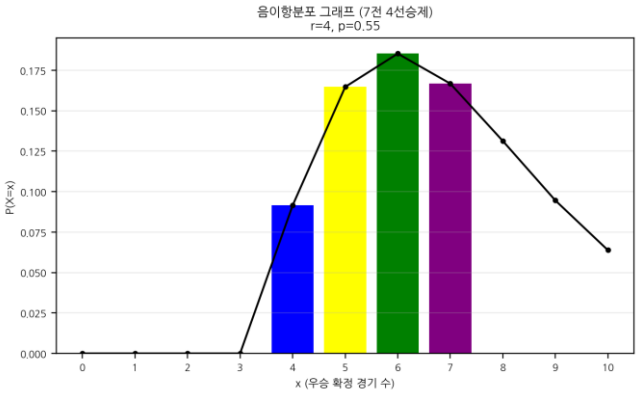
NBA 결승전은 7전 4선승제로 벌어진다. 결승전에서 만난 두 팀 $A$와 $B$의 한 경기에서 $A$팀이 $B$팀을 이길 확률이 0.55일 때, 다음 확률을 구하시오.

a. $A$팀이 6게임째에서 우승할 확률
b. $A$팀이 우승할 확률
c. 이 두 팀이 5전 3선승제인 플레이오프에서 만났다면 $A$팀이 플레이오프에서 우승할 확률

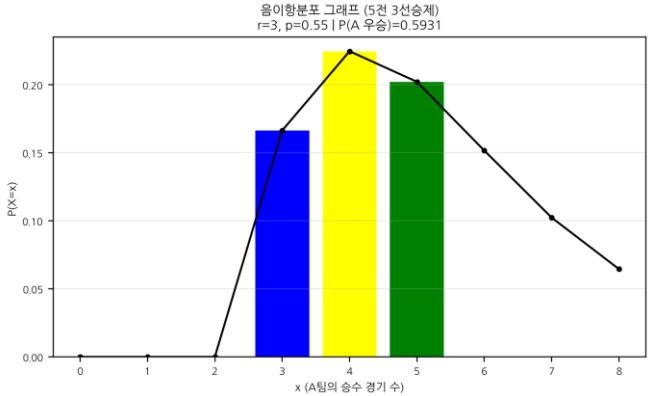
a. $b^*(6;4,0.55) = \binom{5}{3}(0.55)^4(1-0.55)^{6-4} = 0.1853$

b. $P(A\text{팀이 } NBA \text{ 우승}) = b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55)$
$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083$

c. $P(A\text{팀이 플레이오프 우승}) = b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55)$
$= 0.1664 + 0.2246 + 0.2021 = 0.5931$




<그림 6> 예제 5.15-b의 음이항분포 그래프




<그림 7> 예제 5.15-c의 음이항분포 그래프

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 기하분포

- 정의
  - 독립적인 베르누이 시행에서 첫 번째 성공이 일어날 때까지 필요한 시행 횟수 $X$가 따르는 이산 확률분포
- 수식

독립적인 반복시행에서 성공확률이 $p$, 실패확률이 $q$라 하면, 첫 번째 성공이 일어날 때까지의 시행횟수인 확률변수 $X$의 확률분포는 다음과 같음

$$g(x;p) = pq^{x-1}, x = 1, 2, 3, \ldots$$

- 정보보안에서의 활용 예시
  - 비밀번호·인증코드 추측 공격의 최초 성공까지 필요한 시도 횟수 분석

# 기하분포

- 평균과 분산
  - 정리

기하분포를 따르는 확률변수의 평균과 분산은 각각 다음과 같음

$$\mu = \frac{1}{p}, \sigma^2 = \frac{1-p}{p^2}$$

  - 증명 - 평균

- 첫 번째 성공이 나올 때까지 필요한 시행 횟수를 $X$라 함
- 첫 시행은 반드시 1번 수행하며, 실패할 확률 $q$로 다시 동일한 상황이 반복되므로

$$E(X) = 1 + qE(X)$$

- 정리하면

$$E(X) - qE(X) = 1, \qquad E(X)(1-q) = 1$$

- $1 - q = p$이므로 $\mu = \frac{1}{p}$

# 기하분포

- 평균과 분산
  - 증명 - 분산

- 분산은 제곱의 기댓값에서 평균의 제곱을 빼서 구함

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

- 첫 시행에서 성공하면 $X = 1$, 실패하면 1회를 시행한 뒤 다음 시행부터 동일한 조건으로 다시 시작하므로

$$E(X^2) = p \cdot 1^2 + qE[(1+X)^2]$$

- 전개하면

$$E(X^2) = p + q\big(1 + 2E(X) + E(X^2)\big)$$

- $p + q = 1$, $E(X) = \frac{1}{p}$를 대입하면

$$E(X^2) = 1 + \frac{2q}{p} + qE(X^2)$$

- 정리하면

$$pE(X^2) = 1 + \frac{2q}{p}, \qquad E(X^2) = \frac{2-p}{p^2}$$

- 분산 공식에 대입하면

$$\sigma^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

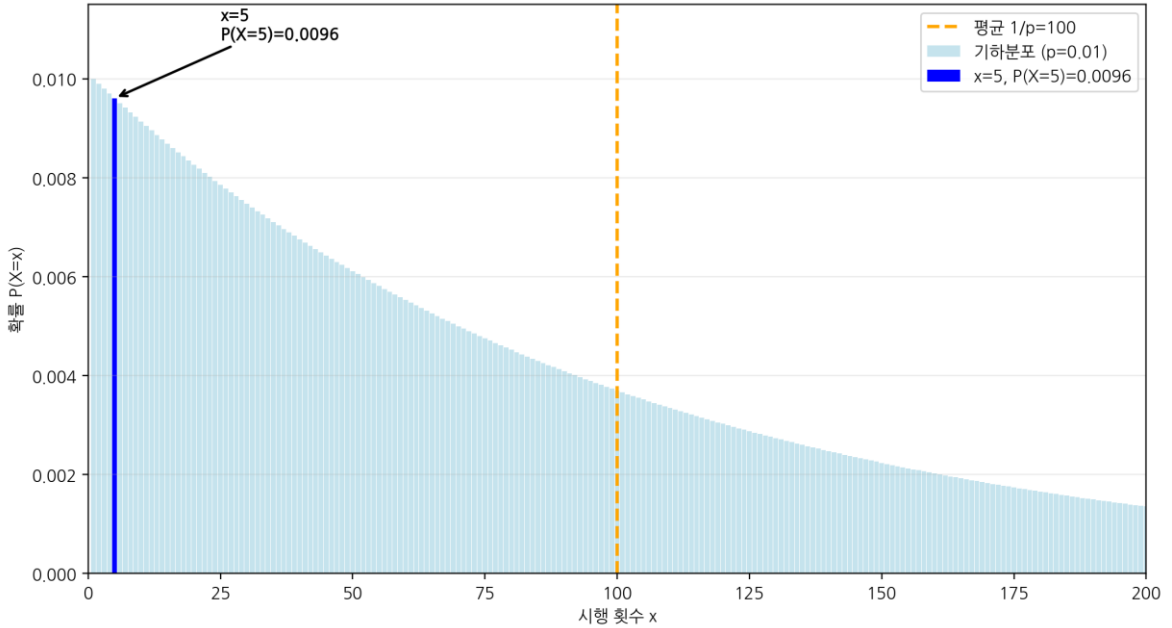
# 기하분포

- 예제 5.16

어떤 제조공정에서 100개의 제품마다 평균적으로 한 개의 불량품이 들어 있다고 알려져 있다. 제품을 하나씩 검사할 때 첫 번째 불량품이 다섯 번째 검사에서 발견될 확률은 얼마인가?

- $x = 5, p = 0.01$인 기하분포를 따르므로,

$$g(5; 0.01) = 0.01(0.99)^4 = 0.0096$$




<그림 8> 예제 5.16의 기하분포 그래프

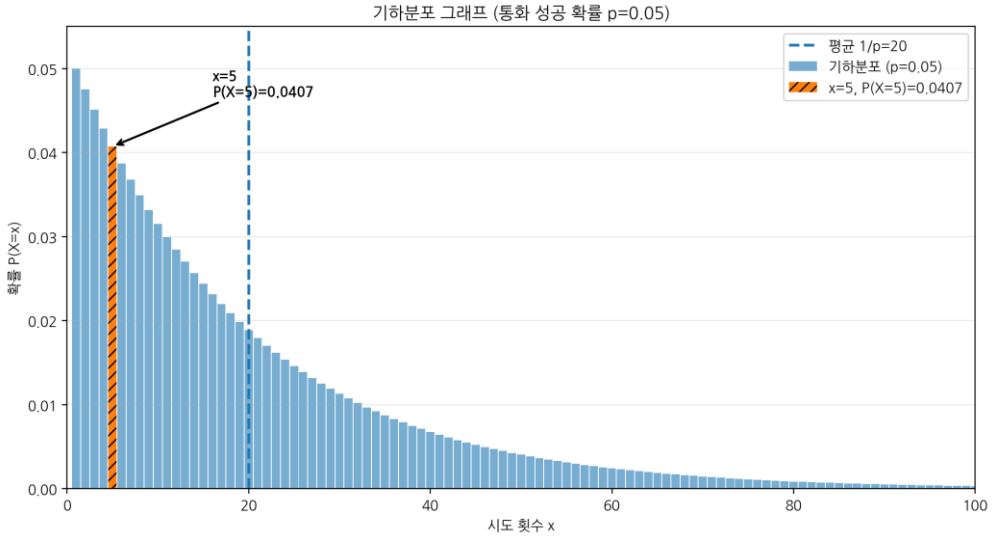
# 기하분포

- 예제 5.17

전화통화가 폭주하는 시간대에 한 번의 시도로 상대방과 통화할 수 있는 가능성은 $p = 0.05$라고 할 때, 다섯 번째 시도에서 상대방과 통화할 수 있는 확률을 구하라.

- $x = 5, p = 0.05$인 기하분포를 따르므로,

$$P(X = x) = g(5; 0.05) = 0.05(0.95)^4 = 0.0407$$




<그림 9> 예제 5.17의 기하분포 그래프

# 목 차


- 이산형 확률분포(Discrete Probability Distribution)
- 이산형 균일분포(Discrete Uniform Distribution)
- 이항분포(Binomial Distribution)
- 다항분포(Multinomial Distribution)
- 초기하분포(Hypergeomertic Distribution)
- 음이항분포(Negative Binomial Distribution)
- 기하분포(Geometric Distribution)
- 포아송분포(Poisson Distribution)

# 포아송분포

- 정의
  - 일정한 시간·영역 내 사건 발생 횟수 $X$가 따르는 이산확률분포
- 성질
  1. 서로 겹치지 않는 구간에서 발생한 사건 수는 서로 독립
     - e.g ., 오전 9~10시 로그인 실패 횟수와 10~11시 로그인 실패 횟수는 서로 영향을 주지 않음
  2. 매우 짧은 시간간격이나 작은 영역에서 단 한 번의 결과가 일어날 확률은 시간간격의 길이나 영역의 크기에 비례함
  3. 매우 짧은 시간간격이나 작은 영역에서 둘 이상의 결과가 일어날 확률은 무시할 수 있음
- 정보보안에서의 활용 예시
  - 단위 시간당 로그인 실패나 공격 요청 횟수를 이용한 이상 징후 탐지

# 포아송분포

- 수식

*$\lambda$: 단위 시간 또는 영역당 평균 발생 횟수
*$x$: 실제 발생 횟수
*$t$: 관찰 구간의 크기
*$\lambda t$: 관찰 구간에서의 평균 발생 횟수

일정한 시간간격 $t$ 또는 일정영역 $t$에서 발생하는 결과의 수를 나타내는 포아송 확률변수 $X$의 확률분포는 다음과 같음

$$p(x;\lambda t) = \frac{e^{(-\lambda t)}(\lambda t)^x}{x!}, x = 0, 1, 2, \ldots$$

여기서 λ는 단위시간 또는 단위면적에서 발생하는 결과의 평균 수이고, $e = 2.71828 \cdots$

- 평균과 분산 (1/3)
  - 정리

포아송 분포 $p(x;\lambda t)$의 평균과 분산은 다음과 같음

$$\mu = \sigma^2 = \lambda t$$

# 포아송분포

- 평균과 분산 (2/3)
  - 증명 – 평균

- 포아송분포의 확률질량함수는 $p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$ 이므로 기댓값의 정의에 따라

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

- $x = 0$인 항은 0이므로

$$E(X) = \sum_{x=1}^{\infty} \frac{e^{-\lambda t}(\lambda t)^x}{(x-1)!}$$

- $\lambda t$를 하나 밖으로 빼면

$$E(X) = \lambda t\, e^{-\lambda t} \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1}}{(x-1)!}$$

- $y = x - 1$로 치환하면

$$E(X) = \lambda t\, e^{-\lambda t} \sum_{y=0}^{\infty} \frac{(\lambda t)^{y}}{y!}$$

- 여기서 $\sum_{y=0}^{\infty} \frac{(\lambda t)^y}{y!} = e^{\lambda t}$ 이므로 $E(X) = \lambda t\, e^{-\lambda t} e^{\lambda t}$
- 따라서 $\mu = E(X) = \lambda t$

# 포아송분포

- 평균과 분산 (3/3)
  - 증명 – 분산

- 분산은 제곱의 기댓값에서 평균의 제곱을 빼서 구함

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

- $X^2 = X(X-1) + X$이므로

$$E(X^2) = E[X(X-1)] + E(X)$$

- 포아송분포의 확률식을 이용하면

$$E[X(X-1)] = \sum_{x=2}^{\infty} x\,(x-1)\frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

- 정리하면

$$E[X(X-1)] = (\lambda t)^2 e^{-\lambda t} \sum_{x=2}^{\infty} \frac{(\lambda t)^{x-2}}{(x-2)!}$$

- $y = x - 2$로 치환하면

$$\sum_{y=0}^{\infty} \frac{(\lambda t)^y}{y!} = e^{\lambda t}$$

- 따라서 $E[X(X-1)] = (\lambda t)^2 e^{-\lambda t} e^{\lambda t} = (\lambda t)^2$
- 앞에서 구한 $E(X) = \lambda t$를 이용하면 $E(X^2) = (\lambda t)^2 + \lambda t$
- 분산 공식에 대입하면 $\sigma^2 = (\lambda t)^2 + \lambda t - (\lambda t)^2$
- 따라서 $\sigma^2 = \lambda t$

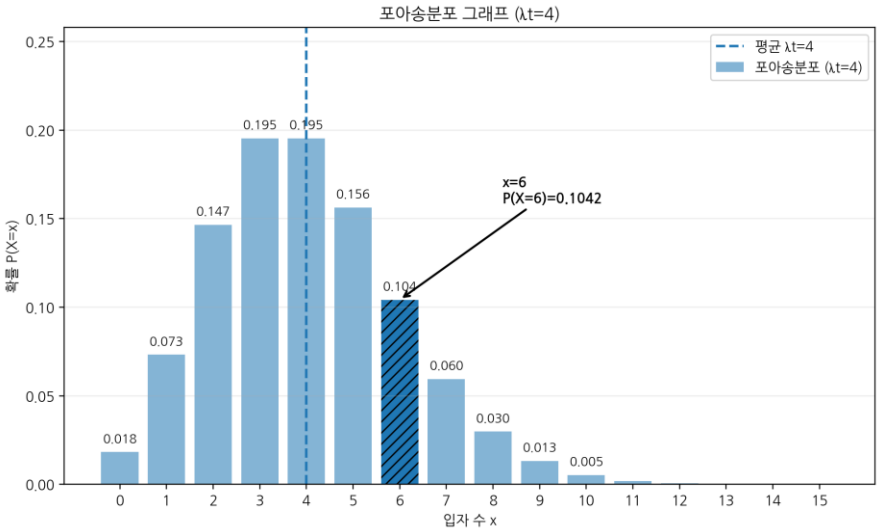
# 포아송분포

- 예제 5.18

실험실에서 $\frac{1}{1000}$ 초 동안 카운터를 통과하는 방사능 입자의 평균수는 4이다. $\frac{1}{1000}$ 초 동안 6개의 입자가 카운터를 통과할 확률을 구하시오.

- $x = 6, \lambda t = 4$ 인 포아송분포를 사용하면,

$$p(6;4) = \frac{e^{(-4)4^6}}{6!} = \sum_{x=0}^{6} p(x;4) - \sum_{x=0}^{5} p(x;4) = 0.8893 - 0.7851 = 0.1042$$




<그림 10> 예제 5.18의 포아송 분포 그래프

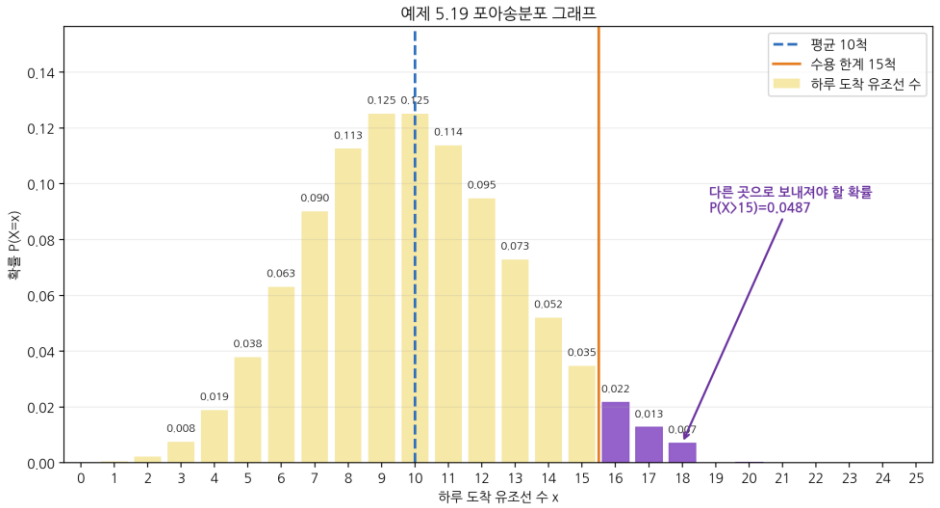
# 포아송분포

- 예제 5.19

어느 항구도시에 하루에 도착하는 유조선은 평균 10척으로 알려져 있다. 그 항구에 있는 시설로는 하루에 기껏해야 15척의 유조선을 취급할 수 있다고 할 때, 어느 날 유조선이 다른 곳으로 보내져야 할 확률은 얼마인가?

- $\lambda t = 10,\ p(x;10) = \frac{e^{(-10)10^x}}{x!}$ 인 포아송분포를 사용하면

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x;10) = 1 - 0.9513 = 0.0487$$




<그림 11> 예제 5.19의 포아송 분포 그래프

# 포아송분포

- 이항분포와의 관계 (1/4)
  - 이항분포에서 $n$이 매우 크고 $p$가 충분히 작을 때 포아송분포로 근사할 수 있으며, 일반적으로 $p \leq 0.05, n \geq 20$ 정도를 근사 적용의 경험적 기준으로 사용하기도 함
  - 이항분포에서 베르누이 시행 간의 독립성은 포아송 과정의 독립성 조건에 대응함
  - 평균은 $\mu = np$로 설정함
    - 수식

$X$를 $b(x;\ n, p)$를 따르는 이항확률변수라 하자. $n \to \infty, p \to 0$이고,
$np \xrightarrow{n\to\infty} \mu$가 상수이면 $b(x; n, p) \xrightarrow{n\to\infty} p(x; \mu)$

# 포아송분포

- 이항분포와의 관계 (2/4)
  - 증명

- $b(x;n,p)=\binom{n}{x}p^x\,q^{n-x}=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}=\frac{n(n-1)\cdots(n-x+1)}{x!}p^x(1-p)^{n-x}$
- $p=\frac{\mu}{n}$로 치환하면 $b(x;n,p)=\frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}$

$$=1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$

- $x$와 $\mu$가 상수일 때 $n\to\infty$임에 따라

$$\lim_{n\to\infty}1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)=1,\ \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-x}=1$$ 이고

- $e$의 정의로부터 $\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n=e^{-\mu}$
- 그러므로 주어진 극한조건 안에서

$$b(x;n,p)\to\frac{e^{-\mu}\mu^x}{x!},x=0,1,2,\ldots$$

# 포아송분포

- 이항분포와의 관계 (3/4)
  - 예제 5.20

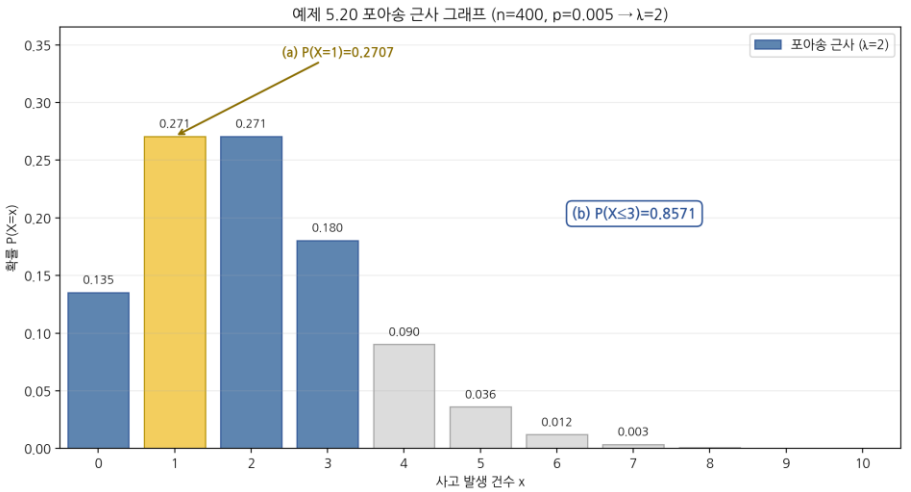
어느 산업현장에서는 사고가 종종 발생하는데, 어느 날 사고가 일어날 확률은 0.005이고, 사고들은 서로 독립적이라고 알려져 있다. 이때 다음 확률을 구하시오.

a. 400일 동안 사고가 한 건 발생할 확률
b. 사고일이 많아야 3일일 인 이항확률변

a. $X$를 $n = 400$이고 $p = 0.005$인 이항확률변수라고 하면, $np = 2$이고 포아송근사를 이용

b. $p(X = 1) = \frac{e^{-2}2^1}{1!} = 0.2707$

c. $P(X \leq 3) = \sum_{x=0}^{3} \frac{e^{-2}2^x}{x!} = 0.8571$




<그림 12> 예제 5.20의 포아송 분포 그래프

# 포아송분포

- 이항분포와의 관계 (4/4)
  - 예제 5.21

유리제품의 제조공정에서 1000개의 제품마다 평균적으로 1개에 하나 이상의 기포가 생긴다고 알려져 있다. 8000개의 확률표본 중에서 기포를 가지고 있는 유리제품의 개수가 7보다 작을 확률은 얼마인가?

- 제품 1개에 기포가 있을 확률은

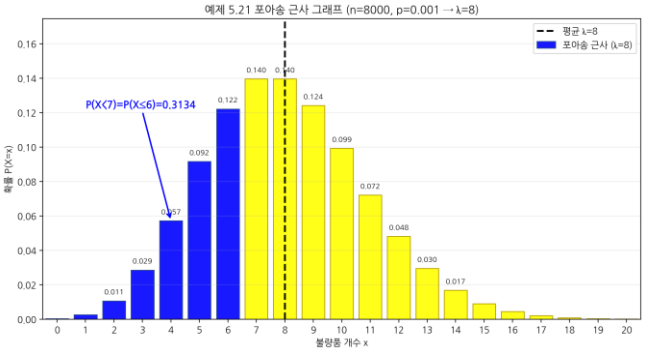
$$p = \frac{1}{1000} = 0.001, \qquad n = 8000$$

- 따라서

$$X \sim B(8000, 0.001), \qquad \mu = np = 8000(0.001) = 8$$

- $n$이 크고 $p$가 작으므로 평균이 8인 포아송분포로 근사하면,

$$P(X < 7) = P(X \leq 6) = \sum_{x=0}^{6} b(x; 8000, 0.001) \approx \sum_{x=0}^{6} p(x; 8) = \sum_{x=0}^{6} \frac{e^{-8} 8^x}{x!} = 0.3134$$




<그림 13> 예제 5.21의 포아송 분포 그래프

# 포아송분포

- 정리 (1/2)

| 분포 | 고정 조건 | 확률변수 의미 | 평균 | 분산 |
|---|---|---|---|---|
| 이산형 균일분포 | 가능한 값 $k$개, 동일한 확률 | 선택된 값 $X$ | $\frac{1}{k}\sum x_i$ | $\frac{1}{k}\sum(x_i-\mu)^2$ |
| 이항분포 | 시행횟수 $n$, 성공확률 $p$ | $n$번 중 성공 횟수 | $np$ | $npq$ |
| 다항분포 | 시행횟수 $n$, 각 범주 확률 $p_i$ | 각 범주의 발생 횟수 $X_i$ | $np_i$ | $np_i(1-p_i)$ |
| 초기하분포 | 모집단 $N$, 성공 $k$, 표본 $n$ | 비복원추출한 표본의 성공 개수 | $\mu=\frac{nk}{N}$ | $\frac{N-n}{N-1}n\frac{k}{N}\left(1-\frac{k}{N}\right)$ |

# 포아송분포

- 정리 (2/2)

| 분포 | 고정 조건 | 확률변수 의미 | 평균 | 분산 |
|---|---|---|---|---|
| 음이항분포 | 목표 성공횟수 $k$, 성공확률 $p$ | $k$번째 성공까지의 총 시행 횟수 | $\frac{k}{p}$ | $\frac{k(1-p)}{p^2}$ |
| 기하분포 | 성공확률 $p$ | 첫 성공까지의 시행 횟수 | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ |
| 포아송분포 | 평균 발생 횟수 $\lambda t$ | 일정 시간·영역 내 사건 발생 횟수 | $\lambda t$ | $\lambda t$ |

# Thanks!

김 민 식(minsik@pel.sejong.ac.kr)

# 부록#1 그래프 Python 코드

- 예제 5.1

```
1.  import matplotlib.pyplot as plt
2.  import numpy as np
3.  from matplotlib import font_manager, rcParams

4.  preferred_fonts = ["NanumGothic", "Malgun Gothic", "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
5.  available_fonts = {f.name for f in font_manager.fontManager.ttflist}
6.  for f in preferred_fonts:
7.              if f in available_fonts:
8.                  rcParams["font.family"] = f
9.                  break
10. rcParams["axes.unicode_minus"] = False
11. labels1 = ["40", "60", "75", "100"]
12. xpos1 = np.arange(len(labels1))
13. p1 = np.repeat(1/4, len(labels1))

14. plt.figure(figsize=(8.5, 5.5))
15. plt.bar(xpos1, p1, width=0.85)
16. plt.ylim(0, 0.30)
17. plt.xticks(xpos1, labels1)
18. plt.xlabel("전구의 와트수 x")
19. plt.ylabel("확률 P(X=x)")
20. plt.figtext(0.5, 0.01, "f(x;4)=1/4,   x=40, 60, 75, 100", ha="center", fontsize=11)
21. plt.tight_layout(rect=[0, 0.05, 1, 1])
22. plt.show()
```

# 부록#2 그래프 Python 코드

- 예제 5.2

```
1.  import matplotlib.pyplot as plt
2.  import numpy as np
3.  from matplotlib import font_manager, rcParams

4.  preferred_fonts = ["NanumGothic", "Malgun Gothic", "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
5.  available_fonts = {f.name for f in font_manager.fontManager.ttflist}

6.  for f in preferred_fonts:
7.      if f in available_fonts:
8.          rcParams["font.family"] = f
9.          break

10. rcParams["axes.unicode_minus"] = False

11. labels1 = ["40", "60", "75", "100"]
12. xpos1 = np.arange(len(labels1))
13. p1 = np.repeat(1/4, len(labels1))
14. plt.figure(figsize=(8.5, 5.5))
15. plt.bar(xpos1, p1, width=0.85)
16. plt.ylim(0, 0.30)
17. plt.xticks(xpos1, labels1)
18. plt.xlabel("전구의 와트수 x")
19. plt.ylabel("확률 P(X=x)")
20. plt.figtext(0.5, 0.01, "f(x;4)=1/4,   x=40, 60, 75, 100", ha="center", fontsize=11)
21. plt.tight_layout(rect=[0, 0.05, 1, 1])
22. plt.show()
```

# 부록#3 그래프 Python 코드

- 예제 5.6

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams

5.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
6.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

7.  for f in preferred_fonts:
8.      if f in available_fonts:
9.          rcParams["font.family"] = f
10.         break

11. rcParams["axes.unicode_minus"] = False

12. # 예제 5.6
13. n = 15
14. p = 0.4
15. x = np.arange(0, n + 1)
16. pmf = np.array([math.comb(n, xi) * (p ** xi) *
    ((1 - p) ** (n - xi)) for xi in x])

17. plt.figure(figsize=(9, 5.5))
```

```
18.plt.bar(x, pmf, width=0.8, color="lightgray",
   alpha=0.45, label="전체 이항분포")
19. plt.plot(x, pmf, marker="o", linewidth=1.8,
    color="black", label="PMF")
20. # (a) X ≥ 10 : 파랑
21. mask_a = x >= 10
22. plt.bar(x[mask_a], pmf[mask_a], width=0.8,
    color="blue", label="(a) X ≥ 10")
23. # (b) 3 ≤ X ≤ 8 : 노랑
24. mask_b = (x >= 3) & (x <= 8)
25. plt.bar(x[mask_b], pmf[mask_b], width=0.8,
    color="yellow", label="(b) 3 ≤ X ≤ 8")

26. # (c) X = 5 : 초록
27. mask_c = x == 5
28. plt.bar(x[mask_c], pmf[mask_c], width=0.8,
    color="green", label="(c) X = 5")

29. plt.xlabel("회복한 사람의 수 x")
30. plt.ylabel("확률 P(X=x)")
31. plt.xticks(x)
32. plt.ylim(0, pmf.max() * 1.22)
33. plt.grid(axis="y", alpha=0.25)
34. plt.legend()
35. plt.tight_layout()
36. plt.show()
```

# 부록#4 그래프 Python 코드

- 예제 5.10

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  import matplotlib.cm as cm
5.  import matplotlib.colors as mcolors
6.  from matplotlib import font_manager, rcParams

7.  # 한글 폰트 설정
8.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
9.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

10. for f in preferred_fonts:
11.     if f in available_fonts:
12.         rcParams["font.family"] = f
13.         break

14. rcParams["axes.unicode_minus"] = False

15. # 예제 5.10 다항분포
16. n = 6
17. p1, p2, p3 = 2/9, 1/6, 11/18

18. cases = []
19. probs = []
```

```
20. for x1 in range(n + 1):
21.     for x2 in range(n + 1 - x1):
22.         x3 = n - x1 - x2

23.         coef = math.factorial(n) / (
24.             math.factorial(x1) *
25.             math.factorial(x2) *
26.             math.factorial(x3)
27.         )

28.         prob = coef * (p1**x1) * (p2**x2) *
    (p3**x3)

29.         cases.append((x1, x2, x3))
30.         probs.append(prob)

31. probs = np.array(probs)
32. x = np.arange(len(cases))
33. labels = [f"({a},{b},{c})" for a, b, c in cases]

34. # 확률 크기에 따라 색상 적용
35. norm = mcolors.Normalize(vmin=probs.min(),
    vmax=probs.max())
36. cmap = cm.plasma
```

# 부록#4 그래프 Python 코드

- 예제 5.10

```
37. fig, ax = plt.subplots(figsize=(13, 5.8))

38. ax.bar(
39.     x,
40.     probs,
41.     width=0.8,
42.     color=cmap(norm(probs))
43. )
44. ax.set_title("예제 5.10 그래프", fontsize=12)
45. ax.set_ylabel("확률")
46. ax.set_xlabel("(X1, X2, X3) 경우의 수")
47. ax.set_xticks(x)
48. ax.set_xticklabels(
49.     labels,
50.     rotation=90,
51.     fontsize=8
52. )
53. ax.set_ylim(0, 0.162)
54. # 컬러바
55. sm = cm.ScalarMappable(
56.     norm=norm,
57.  cmap=cmap
58. )
59. sm.set_array([])

60. cbar = fig.colorbar(
```

```
61.     sm,
62.     ax=ax,
63.     pad=0.02
64. )
65. cbar.set_label("확률 크기")
66. # (2,1,3) 강조
67. target = (2, 1, 3)
68. target_idx = cases.index(target)
69. target_prob = probs[target_idx]
70. ax.annotate(
71.     f"{target_prob:.4f}",
72.     xy=(target_idx, target_prob),
73.     xytext=(target_idx - 0.4, 0.122),
74.     color="black",
75.     fontsize=11,
76.     fontweight="bold",
77.     arrowprops=dict(
78.         arrowstyle="->",
79.         color="black",
80.         lw=1.5
81.     )
82. )
83. plt.tight_layout()
84. plt.show()
```

# 부록#5 그래프 Python 코드

- 예제 5.11

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams

5.  # 한글 폰트 설정
6.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
7.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

8.  for f in preferred_fonts:
9.      if f in available_fonts:
10.         rcParams["font.family"] = f
11.         break

12. rcParams["axes.unicode_minus"] = False

13. # 예제 5.11
14. N = 40
15. K = 3
16. n = 5

17. def hypergeom_pmf(x, N, K, n):
18.     if x < 0 or x > K or x > n:
19.         return 0.0
20.     if n - x > N - K:
21.  return 0.0
22.     return (math.comb(K, x) * math.comb(N - K,
    n - x)) / math.comb(N, n)

23. x = np.arange(0, n + 1)
24. pmf = np.array([hypergeom_pmf(xi, N, K, n) for
    xi in x])

25. # 파랑 → 노랑 → 초록 순서
26. bar_colors = ["blue", "yellow", "green", "blue",
    "yellow", "green"]

27. plt.figure(figsize=(9, 5))

28. plt.bar(
29.     x,
30.     pmf,
31.     width=0.8,
32.     color=bar_colors
33. )

34. plt.plot(
35.     x,
36.     pmf,
37.     marker="o",
38.     linewidth=2,
```

# 부록#5 그래프 Python 코드

- 예제 5.11

```
39. color="black"
40. )

41. # 확률값 표시
42. for xi, pi in zip(x, pmf):
43.     if pi > 0:
44.         plt.text(
45.             xi,
46.             pi + 0.008,
47.             f"{pi:.4f}",
48.             ha="center",
49.             va="bottom",
50.             fontsize=10,
51.             fontweight="bold"
52.         )

53. plt.title("초기하분포 그래프\nN=40, K=3, n=5")
54. plt.xlabel("x (표본에서 발견된 불량품 개수)")
55. plt.ylabel("P(X = x)")
56. plt.xticks(x)
57. plt.ylim(0, pmf.max() * 1.05)
58. plt.grid(True, axis="y", alpha=0.3)
59. plt.tight_layout()

60. plt.show()
```

# 부록#6 그래프 Python 코드

- 예제 5.15

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams
5.  # 한글 폰트 설정
6.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
7.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

8.  for f in preferred_fonts:
9.      if f in available_fonts:
10.         rcParams["font.family"] = f
11.         break
12. rcParams["axes.unicode_minus"] = False
13. p = 0.55
14. r = 4
15. def negbin_pmf(x, r, p):
16.     if x < r:
17.         return 0.0
18.     return math.comb(x - 1, r - 1) * (p ** r) *
    ((1 - p) ** (x - r))
19. # x축 0~10
20. x = np.arange(0, 11)
21. pmf = np.array([negbin_pmf(xi, r, p) for xi in x])
22. # 7전 4선승제이므로 7경기 이후는 0
23. pmf[x > 7] = 0
```

```
24. # 막대는 실제 우승 가능한 4~7경기만
25. bar_x = np.arange(4, 8)
26. bar_y = pmf[4:8]
27. bar_colors = ["blue", "yellow", "green", "purple"]
28. plt.figure(figsize=(6.6, 4.4))
29. plt.bar(
30.     bar_x,
31.     bar_y,
32.     width=0.8,
33.     color=bar_colors
34. )
35. plt.plot(
36.     x,
37.     pmf,
38.     color="black",
39.     marker="o",
40.     linewidth=1.4,
41.     markersize=3
42. )
43. plt.title("음이항분포 그래프 (7전 4선승제)\nr=4, p=0.55")
44. plt.xlabel("x (우승 확정 경기 수)")
45. plt.ylabel("P(X=x)")
46. plt.xticks(x)
47. plt.grid(axis="y", alpha=0.25)
48. plt.tight_layout()
49. plt.show()
```

# 부록#7 그래프 Python 코드

- 예제 5.15

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams
5.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
6.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}
7.  for f in preferred_fonts:
8.      if f in available_fonts:
9.          rcParams["font.family"] = f
10.         break
11. rcParams["axes.unicode_minus"] = False
12. p = 0.55
13. r = 3

14. def negbin_pmf(x, r, p):
15.     if x < r:
16.         return 0.0
17.     return math.comb(x - 1, r - 1) * (p ** r) *
    ((1 - p) ** (x - r))
18. # x축 0~8
19. x = np.arange(0, 9)
20. # 음이항분포 계산
21. pmf = np.array([negbin_pmf(xi, r, p) for xi in x])
22. # 5전 3선승제이므로 5경기 이후는 0
23. pmf[x > 5] = 0
```

```
24. # 막대는 실제 우승 가능한 3~5경기만
25. bar_x = np.arange(3, 6)
26. bar_y = pmf[3:6]
27. bar_colors = ["blue", "yellow", "green"]
28. plt.figure(figsize=(6.6, 4.4))
29. plt.bar(
30.     bar_x,
31.     bar_y,
32.     width=0.8,
33.     color=bar_colors
34. )
35. plt.plot(
36.     x,
37.     pmf,
38.     color="black",
39.     marker="o",
40.     linewidth=1.4,
41.     markersize=3
42. )
43. plt.title("음이항분포 그래프 (5전 3선승제)\nr=3, p=0.55")
44. plt.xlabel("x (우승 확정 경기 수)")
45. plt.ylabel("P(X=x)")
46. plt.xticks(x)
47. plt.grid(axis="y", alpha=0.25)
48. plt.tight_layout()
49. plt.show()
```

# 부록#8 그래프 Python 코드

- 예제 5.16

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt
3.  from matplotlib import font_manager, rcParams

4.  # 한글 폰트 설정
5.  preferred_fonts = ["NanumGothic", "Malgun Gothic",
    "AppleGothic", "Noto Sans CJK KR", "Noto Sans KR"]
6.  available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

7.  for f in preferred_fonts:
8.      if f in available_fonts:
9.          rcParams["font.family"] = f
10.         break
11. rcParams["axes.unicode_minus"] = False
12. # 기하분포
13. p = 0.01
14. target_x = 5

15. x = np.arange(1, 201)

16. # P(X=x)=p(1-p)^(x-1)
17. pmf = p * (1 - p) ** (x - 1)

18. target_prob = p * (1 - p) ** (target_x - 1)
19. mean_x = 1 / p
```

```
20. plt.figure(figsize=(10, 5.5))
21. # 전체 분포
22. plt.bar(
23.     x,
24.     pmf,
25.     width=0.9,
26.     color="lightblue",
27.     alpha=0.7,
28.     label=f"기하분포 (p={p})"
29. )
30. # x=5 강조
31. plt.bar(
32.     target_x,
33.     target_prob,
34.     width=0.9,
35.     color="blue",
36.     label=f"x=5, P(X=5)={target_prob:.4f}"
37. )
38. # 평균선
39. plt.axvline(
40.     mean_x,
41.     color="orange",
42.     linestyle="--",
43.     linewidth=2,
44.     label=f"평균 1/p={int(mean_x)}"
45. )
```

# 부록#8 그래프 Python 코드

- 예제 5.16

```
46. # x=5 주석
47. plt.annotate(
48.     f"x=5\nP(X=5)={target_prob:.4f}",
49.     xy=(target_x, target_prob),
50.     xytext=(25, target_prob + 0.0012),
51.     fontsize=10,
52.     fontweight="bold",
53.     arrowprops=dict(
54.         arrowstyle="->",
55.         lw=1.5
56.     )
57. )

58. plt.xlabel("시행 횟수 x")
59. plt.ylabel("확률 P(X=x)")
60. plt.xlim(0, 200)
61. plt.ylim(0, 0.0115)
62. plt.grid(axis="y", alpha=0.25)
63. plt.legend()
64. plt.tight_layout()

65. plt.show()
```

# 부록#9 그래프 Python 코드

- 예제 5.17

```
1.  import numpy as np
2.  import matplotlib.pyplot as plt
3.  from matplotlib import font_manager, rcParams
4.  # 한글 폰트 설정
5.  preferred_fonts = [
6.      "NanumGothic",
7.      "Malgun Gothic",
8.      "AppleGothic",
9.      "Noto Sans CJK KR",
10.     "Noto Sans KR"
11. ]
12. available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

13. for f in preferred_fonts:
14.     if f in available_fonts:
15.         rcParams["font.family"] = f
16.         break

17. rcParams["axes.unicode_minus"] = False

18. # 예제 5.17: 기하분포
19. p = 0.05
20. target_x = 5

21. x = np.arange(1, 101)
```

```
22. # P(X=x)=p(1-p)^(x-1)
23. pmf = p * (1 - p) ** (x - 1)

24. target_prob = p * (1 - p) ** (target_x - 1)
25. mean_x = 1 / p

26. plt.figure(figsize=(10, 5.5))

27. # 전체 기하분포
28. plt.bar(
29.     x,
30.     pmf,
31.     width=0.9,
32.     alpha=0.6,
33.     label=f"기하분포 (p={p})"
34. )

35. # x=5 강조
36. plt.bar(
37.     target_x,
38.     target_prob,
39.     width=0.9,
40.     hatch="///",
41.     label=f"x=5, P(X=5)={target_prob:.4f}"
42. )
```

# 부록#9 그래프 Python 코드

- 예제 5.17

```
43. plt.axvline(
44.     mean_x,
45.     linestyle="--",
46.     linewidth=2,
47.     label=f"평균 1/p={int(mean_x)}"
48. )
49. # x=5 주석
50. plt.annotate(
51.     f"x=5\nP(X=5)={target_prob:.4f}",
52.     xy=(target_x, target_prob),
53.     xytext=(16, target_prob + 0.006),
54.     fontsize=10,
55.     fontweight="bold",
56.     arrowprops=dict(
57.         arrowstyle="->",
58.         lw=1.5
59.     )
60. )
61. plt.title("기하분포 그래프 (통화 성공 확률 p=0.05)")
62. plt.xlabel("시도 횟수 x")
63. plt.ylabel("확률 P(X=x)")
64. plt.xlim(0, 100)
65. plt.ylim(0, 0.055)
66. plt.grid(axis="y", alpha=0.25)
67. plt.legend()
68. plt.tight_layout()
69. plt.show()
```

# 부록#10 그래프 Python 코드

- 예제 5.18

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams
5.  preferred_fonts = [
6.      "NanumGothic",
7.      "Malgun Gothic",
8.      "AppleGothic",
9.      "Noto Sans CJK KR",
10.     "Noto Sans KR"
11. ]
12. available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

13. for f in preferred_fonts:
14.     if f in available_fonts:
15.         rcParams["font.family"] = f
16.         break
17. rcParams["axes.unicode_minus"] = False
18. lam = 4
19. target_x = 6
20. x = np.arange(0, 16)
21. pmf = np.array([
22.     math.exp(-lam) * (lam ** xi) /
    math.factorial(xi)
23.     for xi in x
24. ])
25. target_prob = pmf[target_x]
26. plt.figure(figsize=(9, 5.5))
27. # 전체 분포
28. bars = plt.bar(
29.     x,
30.     pmf,
31.     width=0.8,
32.     alpha=0.55,
33.     label=f"포아송분포 (λt={lam})"
34. )
35. # x=6 강조
36. bars[target_x].set_alpha(1.0)
37. bars[target_x].set_hatch("///")
38. plt.axvline(
39.     lam,
40.     linestyle="--",
41.     linewidth=1.8,
42.     label=f"평균 λt={lam}"
43. )
44. for xi, pi in zip(x, pmf):
45.     if pi > 0.002:
46.         plt.text(
47.             xi,
48.             pi + 0.004,
49.             f"{pi:.3f}",
50.             ha="center",
51.             va="bottom",
52.             fontsize=9
53.         )
```

# 부록#10 그래프 Python 코드

- 예제 5.18

```
56.plt.annotate(
57.     f"x=6\nP(X=6)={target_prob:.4f}",
58.     xy=(target_x, target_prob),
59.     xytext=(8.2, target_prob + 0.055),
60.     fontsize=10,
61.     fontweight="bold",
62.     arrowprops=dict(
63.         arrowstyle="->",
64.         lw=1.5
65.     )
66. )
67. plt.title("포아송분포 그래프 (λt=4)")
68. plt.xlabel("입자 수 x")
69. plt.ylabel("확률 P(X=x)")
70. plt.xticks(x)
71. plt.ylim(0, pmf.max() * 1.32)
72. plt.grid(axis="y", alpha=0.25)
73. plt.legend()
74. plt.tight_layout()
75. plt.show()
```

# 부록#11 그래프 Python 코드

- 예제 5.19

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams
5.  # 한글 폰트 설정
6.  preferred_fonts = [
7.      "NanumGothic",
8.      "Malgun Gothic",
9.      "AppleGothic",
10.     "Noto Sans CJK KR",
11.     "Noto Sans KR"
12. ]
13. available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

14. for f in preferred_fonts:
15.     if f in available_fonts:
16.         rcParams["font.family"] = f
17.         break

18. rcParams["axes.unicode_minus"] = False

19. # 하루 평균 유조선 10척, 최대 15척까지 수용
20. lam = 10
21. capacity = 15
```

```
22. x = np.arange(0, 26)
23. pmf = np.array([
24.     math.exp(-lam) * (lam ** xi) / math.factorial(xi)
25.     for xi in x
26. ])
27. # P(X > 15)
28. prob_over_capacity = 1 - sum(
29.     math.exp(-lam) * (lam ** xi) / math.factorial(xi)
30.     for xi in range(capacity + 1)
31. )

32. plt.figure(figsize=(10, 5.5))

33. # 전체 분포
34. bars = plt.bar(
35.     x,
36.     pmf,
37.     width=0.8,
38.     color="#F6E7A1",
39.     alpha=0.95,
40.     label="하루 도착 유조선 수"
41. )

42. # 15척 초과 영역 강조
43. for i, xi in enumerate(x):
44.     if xi > capacity:
45.         bars[i].set_color("#8E5AC7")
46.         bars[i].set_alpha(0.95)
```

# 부록#11 그래프 Python 코드

- 예제 5.19

```
47.plt.axvline(
48.     lam,
49.     color="#2E6FBE",
50.     linestyle="--",
51.     linewidth=2,
52.     label="평균 10척"
53. )
54. # 수용 한계 15척
55. plt.axvline(
56.     capacity + 0.5,
57.     color="#E67E22",
58.     linestyle="-",
59.     linewidth=2,
60.     label="수용 한계 15척"
61. )
62. # 주요 확률값 표시
63. for xi, pi in zip(x, pmf):
64.     if pi > 0.006:
65.         plt.text(
66.             xi,
67.             pi + 0.003,
68.             f"{pi:.3f}",
69.             ha="center",
70.             va="bottom",
71.             fontsize=8
72.         )
```

```
73.# P(X>15) 강조
74. plt.annotate(
75.     f"다른 곳으로 보내져야 할
    확률\nP(X>15)={prob_over_capacity:.4f}",
76.     xy=(18, pmf[18]),
77.     xytext=(18.8, 0.09),
78.     fontsize=10,
79.     fontweight="bold",
80.     color="#6A329F",
81.     arrowprops=dict(
82.         arrowstyle="->",
83.         lw=1.6,
84.         color="#6A329F"
85.     )
86. )

87. plt.title("예제 5.19 포아송분포 그래프")
88. plt.xlabel("하루 도착 유조선 수 x")
89. plt.ylabel("확률 P(X=x)")
90. plt.xticks(x)
91. plt.xlim(-0.5, 25.5)
92. plt.ylim(0, pmf.max() * 1.25)
93. plt.grid(axis="y", alpha=0.22)
94. plt.legend()
95. plt.tight_layout()
96. plt.show()
```

# 부록#12 그래프 Python 코드

- 예제 5.20

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams

5.  # 한글 폰트 설정
6.  preferred_fonts = [
7.      "NanumGothic",
8.      "Malgun Gothic",
9.      "AppleGothic",
10.     "Noto Sans CJK KR",
11.     "Noto Sans KR"
12. ]

13. available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}
14. for f in preferred_fonts:
15.     if f in available_fonts:
16.         rcParams["font.family"] = f
17.         break
18. rcParams["axes.unicode_minus"] = False
19. # 예제 5.20: 이항분포의 포아송 근사
20. n = 400
21. p = 0.005
22. lam = n * p  # λ = 2
23. x = np.arange(0, 11)
24. # 포아송 PMF
25. pmf = np.array([
26.     math.exp(-lam) * (lam ** xi) / math.factorial(xi)
27.     for xi in x
28. ])
29. # (a) P(X=1)
30. prob_a = pmf[1]

31. # (b) P(X<=3)
32. prob_b = pmf[:4].sum()

33. plt.figure(figsize=(10, 5.5))

34. # 전체 분포
35. bars = plt.bar(
36.     x,
37.     pmf,
38.     width=0.8,
39.     color="#D9D9D9",
40.     edgecolor="#A0A0A0",
41.     alpha=0.9,
42.     label="포아송 근사 (λ=2)"
43. )

44. # (b) X <= 3 구간 : 파랑
45. for i in range(0, 4):
46.     bars[i].set_color("#4C78A8")
47.     bars[i].set_edgecolor("#2F5597")
```

# 부록#12 그래프 Python 코드

- 예제 5.20

```
# (a) X = 1 : 노랑
bars[1].set_color("#F2C94C")
bars[1].set_edgecolor("#B38F00")
for xi, pi in zip(x, pmf):
    if pi > 0.002:
        plt.text(
            xi,
            pi + 0.006,
            f"{pi:.3f}",
            ha="center",
            va="bottom",
            fontsize=9
        )
plt.annotate(
    f"(a) P(X=1)={prob_a:.4f}",
    xy=(1, prob_a),
    xytext=(2.4, 0.34),
    fontsize=10,
    fontweight="bold",
    color="#8A6D00",
    arrowprops=dict(
        arrowstyle="->",
        color="#8A6D00",
        lw=1.5
    )
)
```

```
# (b) 강조
plt.text(
    6.2,
    0.20,
    f"(b) P(X≤3)={prob_b:.4f}",
    fontsize=11,
    fontweight="bold",
    color="#2F5597",
    bbox=dict(
        boxstyle="round,pad=0.4",
        fc="white",
        ec="#2F5597",
        lw=1.2
    )
)

plt.title("예제 5.20 포아송 근사 그래프 (n=400, p=0.005
→ λ=2)")
plt.xlabel("사고 발생 건수 x")
plt.ylabel("확률 P(X=x)")
plt.xticks(x)
plt.xlim(-0.5, 10.5)
plt.ylim(0, pmf.max() * 1.35)
plt.grid(axis="y", alpha=0.22)
plt.legend()
plt.tight_layout()

plt.show()
```

# 부록#13 그래프 Python 코드

- 예제 5.21

```
1.  import math
2.  import numpy as np
3.  import matplotlib.pyplot as plt
4.  from matplotlib import font_manager, rcParams

5.  # 한글 폰트 설정
6.  preferred_fonts = [
7.      "NanumGothic",
8.      "Malgun Gothic",
9.      "AppleGothic",
10.     "Noto Sans CJK KR",
11.     "Noto Sans KR"
12. ]
13. available_fonts = {f.name for f in
    font_manager.fontManager.ttflist}

14. for f in preferred_fonts:
15.     if f in available_fonts:
16.         rcParams["font.family"] = f
17.         break
18. rcParams["axes.unicode_minus"] = False

19. # 예제 5.21: 이항분포의 포아송 근사
20. n = 8000
21. p = 0.001
22. lam = n * p   # λ = 8
23. x = np.arange(0, 21)

24. # 포아송 PMF
25. pmf = np.array([
26.     math.exp(-lam) * (lam ** xi) / math.factorial(xi)
27.     for xi in x
28. ])

29. # P(X < 7) = P(X <= 6)
30. prob_target = pmf[:7].sum()

31. plt.figure(figsize=(10, 5.5))

32. # 전체 분포: 노랑
33. bars = plt.bar(
34.     x,
35.     pmf,
36.     width=0.8,
37.     color="yellow",
38.     edgecolor="#B59B00",
39.     alpha=0.9,
40.     label="포아송 근사 (λ=8)"
41. )

42. # X < 7 구간 강조: 파랑
43. for i in range(0, 7):
44.     bars[i].set_color("blue")
45.     bars[i].set_edgecolor("#1D3F8F")
```

# 부록#13 그래프 Python 코드

- 예제 5.21

```
46.# 평균 λ=8 표시
47. plt.axvline(
48.     lam,
49.     color="black",
50.     linestyle="--",
51.     linewidth=2,
52.     label="평균 λ=8"
53. )

54. # 주요 확률값 표시
55. for xi, pi in zip(x, pmf):
56.     if pi > 0.01:
57.         plt.text(
58.             xi,
59.             pi + 0.003,
60.             f"{pi:.3f}",
61.             ha="center",
62.             va="bottom",
63.             fontsize=8
64.         )
65. # 구하고자 하는 확률 강조
66. plt.annotate(
67.     f"P(X<7)=P(X≤6)={prob_target:.4f}",
68.     xy=(4, pmf[4]),
69.     xytext=(1.0, pmf.max() * 0.88),
```

```
70. # 구하고자 하는 확률 강조
71. plt.annotate(
72.     f"P(X<7)=P(X≤6)={prob_target:.4f}",
73.     xy=(4, pmf[4]),
74.     xytext=(1.0, pmf.max() * 0.88),
75.     fontsize=11,
76.     fontweight="bold",
77.     color="blue",
78.     arrowprops=dict(
79.         arrowstyle="->",
80.         color="blue",
81.         lw=1.6
82.     )
83. )

84. plt.title("예제 5.21 포아송 근사 그래프 (n=8000, p=0.001
    → λ=8)")
85. plt.xlabel("불량품 개수 x")
86. plt.ylabel("확률 P(X=x)")
87. plt.xticks(x)
88. plt.xlim(-0.5, 20.5)
89. plt.ylim(0, pmf.max() * 1.25)
90. plt.grid(axis="y", alpha=0.22)
91. plt.legend()
92. plt.tight_layout()

93. plt.show()
```