



2026/03/01, 2026 확률 기초 세미나



이공학도를 위한 확률 및 통계학

-5장 이산형 확률분포-

민 윤 홍(yunhong@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

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

목 차

- 보충

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

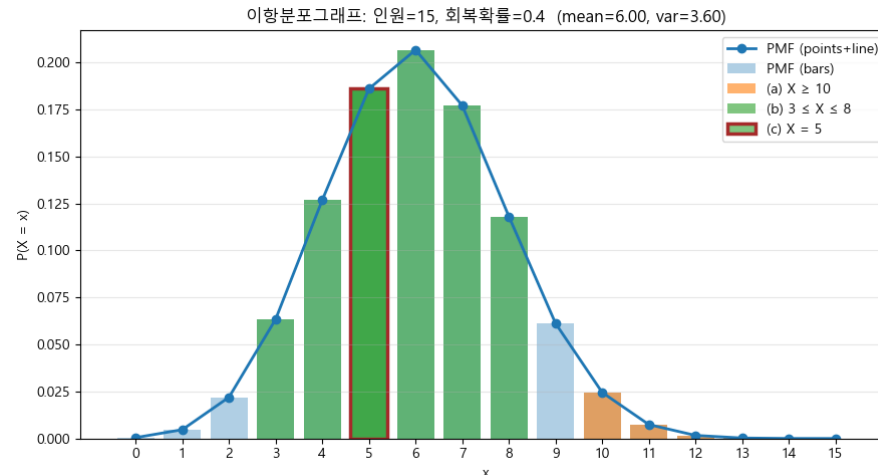
보충

• 예제 5.6

회복 확률이 0.4인 binomial에 15명이 걸렸을 때 다음 확률을 구하시오

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

- (a) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^9 b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$
- (b) $P(3 \leq X \leq 8) = \sum_{x=3}^8 b(x; 15, 0.4) = \sum_{x=0}^8 b(x; 15, 0.4) - \sum_{x=0}^2 b(x; 15, 0.4) = 0.9050 - 0.0271 = 0.8779$
- (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

• 예제 5.9

예제 5.6의 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 ~ 9.794가 됨
- 체비셰프 정리는 빈혈에 걸린 15명의 환자 중 회복될 사람의 수가 2.206과 9.794 사이, 혹은 자료가 이산형이므로 3명에서 9명까지 될 확률이 적어도 $\frac{3}{4}$ 이 됨을 표현함

보충

• 예제 5.10

3개의 활주로가 있는 공항에서 각 활주로가 사용될 확률은 다음과 같다.

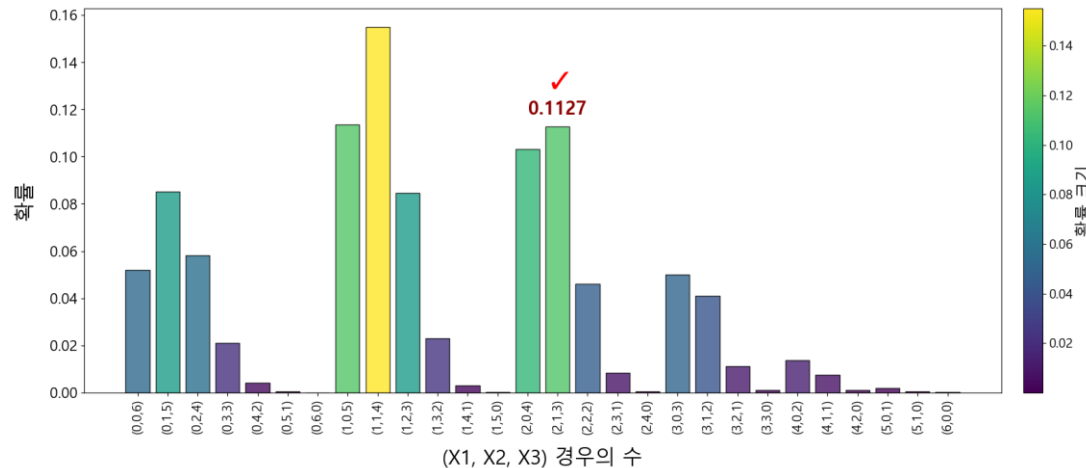
활주로 1 : $p_1 = \frac{2}{9}$, 활주로 2 : $p_2 = \frac{1}{6}$, 활주로 3 : $p_3 = \frac{11}{18}$

임의로 도착하는 6대의 비행기가 다음과 같이 활주로에 도착할 확률을 구하시오
활주로 1 : 2대, 활주로 2 : 1대, 활주로 3 : 3대

- 다항분포 수식을 활용하면

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

예제 5.10: 다항분포 $P(X_1, X_2, X_3)$ 시각화



<그림 4> 예제 5.10의 다항분포 그래프

보충

• 평균과 분산

음이항분포 $b^*(x; k, p)$ 의 평균과 분산은

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

• 증명 (1/2)

- 평균의 정의에 음이항분포 형식을 적용하면 $E(X) = \sum_{x=k}^{\infty} x \binom{x-1}{k-1} p^k (q)^{x-k}$
- $x \binom{x-1}{k-1} = \frac{x!}{(k-1)!(x-k)!} = k \frac{x!}{k!(x-k)!} = k \binom{x}{k}$ 로 변형
- 정리하면 $E(X) = \sum_{x=k}^{\infty} k \binom{x}{k} p^k (q)^{x-k}$
- 음이항분포 형식이 나오도록 k 를 분리하고 p 를 곱했다가 나누면
$$E(X) = k \sum_{x=k}^{\infty} \binom{x}{k} p^k q^{x-k} = \frac{k}{p} \sum_{x=k}^{\infty} \binom{x}{k} p^{k+1} q^{x-k}$$
- $x+1 = y, k+1 = m$ 으로 치환하면 $\frac{k}{p} \sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m}$
- $\sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m} = 1$ 이므로 음이항분포의 평균은 $\frac{k}{p}$

보충

• 평균과 분산

• 증명 (2/2)

- $E(X^2) - (E(X))^2$ 에서 $E(X^2)$ 대신 $E[X(X-1)] + E(X)$ 으로 계산
- $E[X(X-1)] = \sum_{x=k}^{\infty} x(x-1) \binom{x-1}{k-1} p^k q^{x-k}$
- 평균 증명과정과 마찬가지로 $x(x-1) \binom{x-1}{k-1} = k(k+1) \binom{x}{k+1}$ 로 변형
- 정리하면 $E[X(X-1)] = \sum_{x=k}^{\infty} k(k+1) \binom{x}{k+1} p^k q^{x-k}$
- 우변을 음이항분포 꼴로 만들기 위해 p^2 을 분리, $\frac{k(k+1)}{p^2} \sum_{x=k}^{\infty} \binom{x}{k+1} p^{k+2} q^{x-k}$
- $x+1=y, k+2=m$ 으로 치환, $\sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m+1}$
- $E[X(X-1)] = \frac{k(k+1)q}{p^2}, E(X) = \frac{k}{p}$ 이므로 $E(X^2) = \frac{k(k+1)q}{p^2} + \frac{k}{p}$
- $E(X^2) - \left(\frac{k}{p}\right)^2 = \frac{k(k+1)}{p^2} + \frac{k}{p} - \frac{k^2}{p^2} = \frac{k(k+1)q + kp - k^2}{p^2}$
- $q = 1-p$ 를 대입, 정리하면 $\frac{k(1-p)}{p^2}$

보충

- 정의

- 일정한 시간 간격 또는 일정한 영역 안에서 발생하는 결과의 개수를 세는 실험의 확률 분포
 - e.g., PEL에 시간 당 오는 메일의 수

- 특징

1. 단위시간간격이나 일정영역에서 발생하는 결과의 수는 서로 겹치지 않는 다른 시간간격이나 영역에서 발생하는 수와 독립임을 가정
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

보충

• 특징

• 증명

- t : 시간, $N(t)$: $t = 0$ 부터 t 까지 발생한 사건의 확률변수일 때,
- 시간 t 까지 정확히 n 번 발생할 확률을 $P_n(t) = P(N(t) = n)$ 라고 가정
- 시간 $t + h$ 까지 0번 발생하기 위해선 시간 t 까지 0번, 그 다음 h 동안 0번 발생
- 특징 1에 의해 $P_0(t + h) = P_0(t)(1 - \lambda h + o(h))$ 가 성립
 - λ 는 단위시간 또는 단위면적에서 발생하는 결과의 평균 수
- 정리하면 $\frac{P_0(t+h)-P_0(t)}{h} = -\lambda P_0(t) + \frac{P_0(t)o(h)}{h}$
- 양변에 $\lim_{h \rightarrow 0}$ 을 취하면 $P'_0(t) = -\lambda P_0(t)$
- 양변을 $P_0(t)$ 로 나누고 적분한 후 지수를 취하면 $P_0(t) = e^C e^{-\lambda t}$, (C 는 적분상수)
- $N(0) = 0$ 이므로 $P_0(0) = P(N(0) = 0) = 1$
- 따라서 $1 = P_0(0) = e^C$, $P_0(t) = e^{-\lambda t}$
- $n = 1$ 일 때 $P_1(t) = \lambda t e^{-\lambda t}$, $n = 2$ 일 때 $P_2(t) = \frac{(\lambda t)^2}{2!} e^{-\lambda t}$
- 일반화하면 $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$

보충

• 정리 (1/2)

분포	확률질량함수	평균	분산	사용 상황
이산형 균일분포	$f(x; k) = \frac{1}{k}$	$\frac{1}{k} \sum_{i=1}^k x_i$	$\frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2$	모든 결과가 동일한 확률 (주사위 등)
이항분포	$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$	np	npq	독립 베르누이 시행 n 번을 진행하여 성공 횟수를 파악
다항분포	$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n)$ $= \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$	np_i	$np_i(1 - p_i)$	이항분포의 확장인 상황 (범주 3개 이상)

보충

• 정리 (2/2)

분포	확률질량함수	평균	분산	사용 상황
초기하분포	$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$	$\frac{nk}{N}$	$\frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \left(1 - \frac{k}{N}\right)$	비복원추출을 진행하는 상황
음이항분포	$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$	r 번째 성공까지 시행 횟수를 파악
기하분포	$g(x; p) = pq^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	첫 성공까지 시행 횟수를 파악
포아송분포	$\frac{e^{-\lambda t} (\lambda t)^x}{x!}$	λ	λ	단위 시간/공간 내 사건 횟수 파악

목 차

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

이산형 확률분포

- 정의

- 이산형 확률변수를 나타내는 분포

- 확률변수가 가질 수 있는 값의 개수를 셀 수 있거나 나열이 가능한 확률변수

- e.g., 주사위를 던져서 나올 수 있는 경우의 수는 1, 2, 3, 4, 5, 6

- 특징

- 서로 다른 실험이라도 결과값의 분포 형태가 동일하게 나타날 수 있음

- 동전의 앞, 뒤를 구분하는 것과 공정품의 정상, 불량을 판별하는 경우 모두 이항분포를 따름

이산형 확률분포

- 종류 (1/2)
 - 이산형 균일분포
 - 확률변수가 취할 수 있는 각 값들이 모두 같은 확률을 가지는 분포
 - 이항분포
 - 성공 확률이 일정한 독립적인 베르누이 시행을 반복했을 때 성공 횟수의 확률 분포를 나타내는 이산 확률 분포
 - 베르누이 시행(Bernoulli Trial)
 - 결과가 성공과 실패 둘 중 하나로만 나타나는 실험
 - 다항분포
 - 각 시행에서 가능한 결과가 두 개 이상인 확률 분포

이산형 확률분포

- 종류 (2/2)

- 초기하 분포

- 각 시행에서 독립이 전제되지 않은 이산 확률 분포

- 음이항 분포

- 독립적인 베르누이 시행에서 목표 성공 횟수를 달성하기까지 필요한 총 시행 횟수를 나타내는 이산 확률 분포

- 기하 분포

- 독립적인 베르누이 시행에서 처음으로 성공할 때까지 필요한 시행 횟수에 대한 이산 확률 분포

- 포아송 분포

- 단위 시간 안에 어떤 사건이 몇 번 발생할 것인지를 표현하는 이산 확률 분포

목 차

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

이산형 균일분포

- 정의

*모수(Parameter)

모수는 확률분포의 형태를 결정하는 고정된 값이다.

- 확률변수가 취할 수 있는 각 값들이 모두 같은 확률을 가지는 이산 확률 분포

- 수식

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

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

로 주어진다

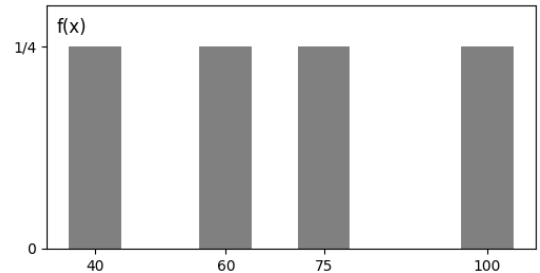
- 균일분포는 모수* k 에 종속되어 있음을 나타내기 위해 $f(x)$ 대신에 $f(x; k)$ 를 사용
 - x 에 대한 함수지만, 형태가 k 에 의해 결정

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
- $\therefore X$ 의 확률분포는 $f(x; 4) = \frac{1}{4}, x = 40, 60, 75, 100$ 인 균일분포임

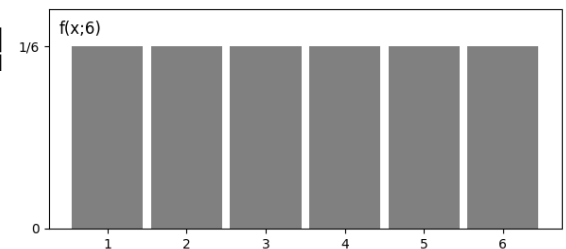


<그림 1> 예제 5.1의 확률분포 그래프

• 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
- $\therefore X$ 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

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

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 된다. 또한 분산의 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

- 예제 5.3

예제 5.2에서 X 의 분산과 평균을 구하시오

- 예제 5.2에서 X 의 확률분포는 $f(x; 6) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6$ 인 균일분포임
- $\mu = \frac{1+2+3+4+5+6}{6} = 3.5$
- $\sigma^2 = \frac{1}{6}[(1 - 3.5)^2 + (2 - 3.5)^2 + \dots + (6 - 3.5)^2] = \frac{17.5}{6} = \frac{35}{12} = 2.92$

목 차

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

이항분포

- 베르누이 시행(Bernoulli Trial)
 - 결과가 성공 혹은 실패 둘 중 하나로만 나타나는 독립적인 무작위 실험
- 특징
 - 실험은 n 번의 반복실험으로 구성
 - 각 시행의 결과는 성공이나 실패, 두 가지 중 하나
 - 성공확률은 매 시행마다 일정
 - 각 시행은 서로 독립

이항분포

• 베르누이 시행(Bernoulli Trial)

• 예제 5.4

제조과정에서 3개의 제품을 임의로 추출하여 양품과 불량을 구분하는 베르누이 시행을 표현하시오

- 불량을 발견하는 경우를 성공이라고 하면, 성공횟수는 0에서 3사이의 값을 가지는 확률변수가 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
- $P(X = x) = \binom{3}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}$, $P(NDN) = P(N)P(D)P(N) = \left(\frac{3}{4}\right) \left(\frac{1}{4}\right) \left(\frac{3}{4}\right) = \frac{9}{64}$
- 나머지 결과들도 동일하게 구할 수 있으므로 X 의 확률분포는 다음과 같음

x	0	1	2	3
$f(x)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

이항분포

- 정의

- 성공 확률이 일정한 독립적인 베르누이 시행을 반복했을 때, 성공 횟수의 확률 분포를 나타내는 이산 확률 분포

- 수식

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

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

- $(q + p)^n$ 의 이항전개에서 각 항의 $x = 0, 1, 2, \dots, n$ 일 때 $b(x; n, p)$ 의 각 값들과 일치한다는 사실에서 파생

이항 정리는 두 개의 항으로 이루어진 $(a + b)^n$ 을 조합을 이용하여 전개하는 정리

$$(a + b)^n = \sum_{r=0}^n \binom{n}{r} a^{n-r} b^r = \binom{n}{r} a^n + \binom{n}{r} a^{n-1} b + \dots \binom{n}{r} a^{n-r} b^r + \dots + \binom{n}{r} b^n$$

이항분포

• 예제 5.5

어떤 종류의 부품이 충격실험에서 충격을 견딜 확률이 $\frac{3}{4}$ 이라고 할 때, 4개의 부품에 대해 실험했을 때 2개가 충격에 견딜 확률을 구하시오

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

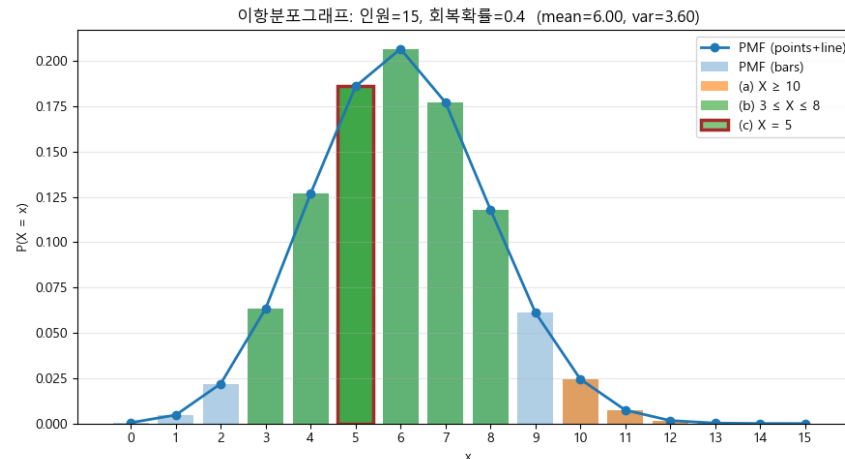
이항분포

• 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) $P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{x=0}^9 b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$
- (b) $P(3 \leq X \leq 8) = \sum_{x=3}^8 b(x; 15, 0.4) = \sum_{x=0}^8 b(x; 15, 0.4) - \sum_{x=0}^2 b(x; 15, 0.4) = 0.9050 - 0.0271 = 0.8779$
- (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 $X = b(x; 20, 0.03)$ 이고 확률을 이항분포를 사용해 나타내면

$$P(X \geq 1) = 1 - P(X = 0) = 1 - b(0; 20, 0.03) = 1 - 0.03^0(1 - 0.03)^{20-0} = 0.4562$$

- (b) 납품시마다 최소한 한 대의 불량품이 있거나 없기 때문에 납품시마다 하는 검사는 $p = 0.4562$ 인 베르누이 과정으로 생각 가능함
 - 납품들이 서로 독립적이라고 가정하고 최소한 한 대의 불량품을 포함하는 납품수를 Y 라고 한다면 $Y = b(y; 10, 0.4562)$

$$\therefore P(Y = 3) = \binom{10}{3} 0.4562^3 = (1 - 0.4562)^7 = 0.1602$$

이항분포

• 평균과 분산 (1/4)

이항분포 $b(x; n, p)$ 의 평균과 분산은

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

• 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 $E(I_j) = (0)(q) + (1)(p) = p$ 이다.

$g(X, Y) = X$ 이고 $h(X, Y) = Y$ 라 했을 때 $E[X \pm Y] = E[X] \pm E[Y]$ 를 이용하면, 이항분포의 평균은

$$\mu = E(X) = E(I_1) + E(I_2) + \cdots + E(I_n) = \underbrace{p + p + \cdots + p}_{n\text{개}} = np$$

이항분포

- 평균과 분산 (2/4)
- 증명 (2/2)

I_j 의 분산은

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

서로 독립인 확률변수의 합의 분산은 각 분산의 합이므로 이항분포의 분산은

$$\sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2 = \underbrace{pq + pq + \cdots + pq}_{n\text{개}} = npq$$

이항분포

- 평균과 분산 (3/4)

- 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 (4/4)

- 예제 5.9

회복 확률이 0.4인 빈혈에 15명이 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 ~ 9.794가 됨
- 체비셰프 정리는 빈혈에 걸린 15명의 환자 중 회복될 사람의 수가 2.206과 9.794 사이, 혹은 자료가 이산형이므로 3명에서 9명까지 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 (Hypergeometric Distribution)
- 음이항분포 (Negative Binomial Distribution)
- 기하분포 (Geometric Distribution)
- 포아송 분포 (Poisson Distribution)

다항분포

- 정의

- 다항실험을 n 번 반복했을 때, 각 결과가 발생한 횟수의 결합 확률분포
- 다항실험
 - 각 시행에서 가능한 결과가 두 개 이상인 경우의 실험

- 수식

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

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

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

다항분포

• 예제 5.10

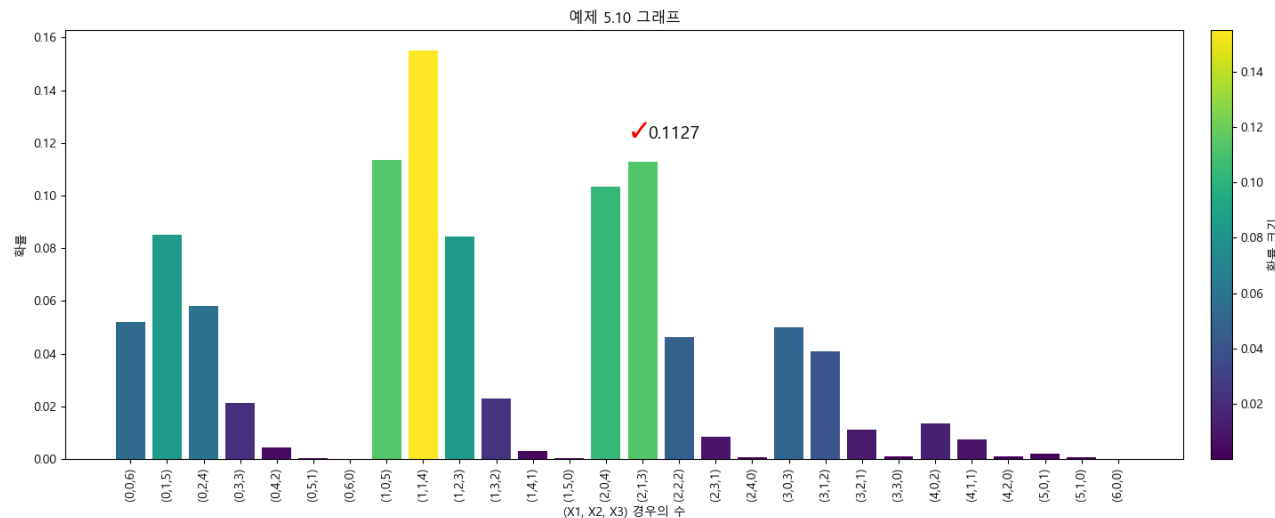
3개의 활주로가 있는 공항에서 각 활주로가 사용될 확률은 다음과 같다.

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 확률을 구하시오
활주로 1 : 2대, 활주로 2 : 1대, 활주로 3 : 3대

- 다항분포 수식을 활용하면

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



<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 (Hypergeometric Distribution)
- 음이항분포 (Negative Binomial Distribution)
- 기하분포 (Geometric Distribution)
- 포아송 분포 (Poisson Distribution)

초기하분포

- 정의

- 유한한 집단에서 비복원추출을 할 때, 특정 항목이 뽑힐 확률을 계산하는 이산 확률 분포

- 특징

- 크기 N 의 유한모집단으로부터 크기 n 의 확률표본을 비복원으로 추출
- N 개 중 k 개는 성공으로, 나머지 $(N - k)$ 개는 실패로 분류

- 수식

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

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

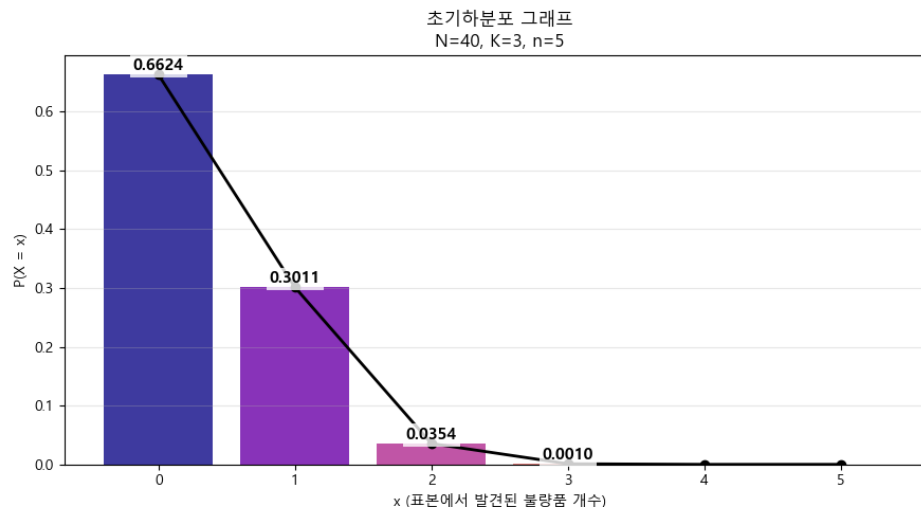
초기하분포

• 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

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



<그림 5> 예제 5.11의 초기하분포 그래프

초기하분포

• 평균과 분산

초기하분포 $h(x; N, n, k)$ 의 평균과 분산은

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

• 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

• 예제 5.12

100개의 제품 중 12개가 불량품이라고 하자. 100개중 10개를 임의로 뽑았을 때 이 중 3개가 불량일 확률을 구하시오

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

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

• 예제 5.13

예제 5.11에서 확률변수의 평균과 분산을 구하고, 체비셰프 정리를 이용하여 구간 $\mu \pm 2\sigma$ 의 의미를 설명하라

- $n = 5$ $N = 40$, $k = 3$ 인 초기하실험이므로, 분산 공식에 의하여

$$\mu = \frac{(5)(3)}{40} = \frac{3}{8} = 0.3750, \sigma^2 = \left(\frac{40-5}{39}\right) (5) \left(\frac{3}{40}\right) \left(1 - \frac{3}{40}\right) = 0.3113, \sigma \approx 0.5579$$

- $\mu \pm 2\sigma$ 는 $-0.7410 \sim 1.4910$ 로 체비셰프 정리는 3개의 불량품이 포함된 크기 40인 로트에서 임의로 5개를 취할 때 얻어지는 불량품의 개수가 $-0.7410 \sim 1.4910$ 사이에 있을 확률이 적어도 $\frac{3}{4}$ 이 됨

초기하분포

- 이항분포와의 관계 (1/2)

- 모집단의 크기 N 과 표본의 개수 n 이 $n/N \leq 0.05$ 일 때, 초기하분포와 이항분포는 거의 차이가 없음
- 증명

- $P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ 인 초기하분포 확률식에서 조합(C)를 곱으로 풀어 쓰면
- $\binom{K}{x} = \frac{K(K-1)\cdots(K-x+1)}{x!}$, $\binom{N-K}{n-x} = \frac{(N-K)(N-K-1)\cdots(N-K-n+x+1)}{(n-x)!}$, $\binom{N}{n} = \frac{N(N-1)\cdots(N-n+1)}{n!}$
- 대입하여 정리하면 $P(X = x) = \frac{\binom{n}{x} \frac{(K(K-1)\cdots(K-x+1))((N-K)(N-K-1)\cdots(N-K-n+x+1))}{N(N-1)\cdots(N-n+1)}}$
($\frac{K}{N} \frac{K-1}{N-1} \cdots =$ 성공이 x 번 나올 부분, $\frac{N-K}{N-x} \frac{N-K-1}{N-x-1} \cdots =$ 실패가 $n-x$ 번 나올 부분)
- n 이 N 보다 작을 경우에 성공은 $\frac{K}{N}$ 에, 실패는 $\frac{N-K}{N}$ 에 근사
- 근사식을 쓰면 $P(X = x) \approx \binom{n}{x} \left(\frac{K}{N}\right)^x \left(\frac{N-K}{N}\right)^{n-x}$
- $p = \frac{K}{N}$ 로 치환하면 $\binom{n}{x} p^x (1-p)^{n-x}$

초기하분포

• 이항분포와의 관계 (2/2)

• 예제 5.14

5000개의 타이어 중에서 1000개가 결함을 가지고 있다고 가정했을 때, 임의로 10개를 뽑았을 때 3개가 결함을 가지고 있을 확률을 구하시오

- 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
- $\therefore$ 3개의 결함이 있는 타이어를 구입할 확률은

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

초기하분포

- 다변량 초기하분포 (1/2)

- 모집단에 여러 종류(그룹)가 있을 때, 비복원으로 표본을 뽑아 각 종류가 몇 개씩 나오는지 확인하는 분포
- 수식

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

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

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

초기하분포

• 다변량 초기하분포 (2/2)

• 예제 5.15

- 생체실험 대상자 10명 중에서 3명은 혈액형이 O형, 4명은 A형, 나머지 3명은 B형이라고 가정하자. 5명을 임의로 선택할 때, O형인 사람이 1명, A형인 사람이 2명, B형인 사람이 2명일 확률을 구하시오
- 임의로 선택한 O형의 수를 $x_1 = 1$, 10명 중 전체 O형의 수를 $a_1 = 3$
- 임의로 선택한 A형의 수를 $x_2 = 2$, 10명 중 전체 A형의 수를 $a_2 = 4$
- 임의로 선택한 B형의 수를 $x_3 = 2$, 10명 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 (Hypergeometric Distribution)
- 음이항분포 (Negative Binomial Distribution)
- 기하분포 (Geometric Distribution)
- 포아송 분포 (Poisson Distribution)

음이항분포

- 정의

- 음이항실험에서 k 번째 성공이 일어날 때까지의 시행횟수 X 의 확률분포
 - 음이항실험
 - 고정된 시행에서 k 번째 성공이 x 번째 시행에서 일어날 확률을 구하는 실험

- 수식

- 독립적인 반복시행에서 성공확률이 p , 실패확률이 q 라 하면, k 번째 성공이 일어날 때까지의 시행횟수인 확률변수 X 의 확률분포는 다음과 같다

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

음이항분포

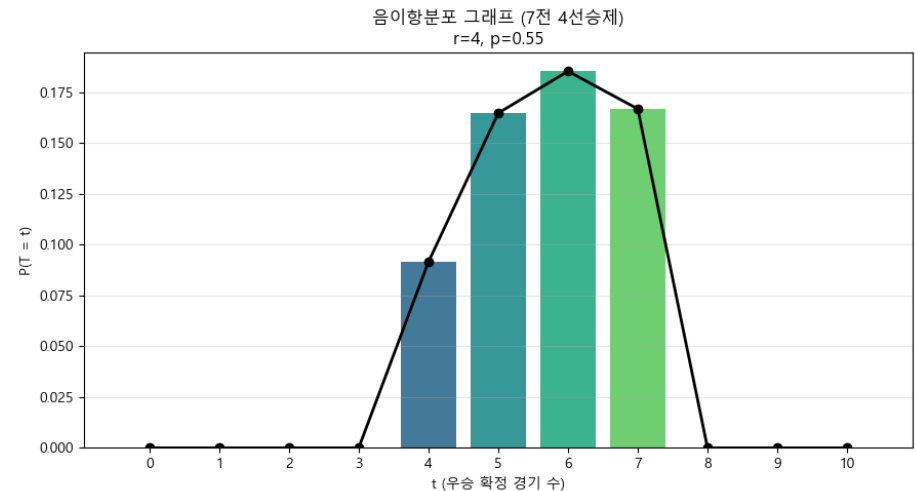
• 예제 5.16 (1/2)

- NBA 결승전은 7전 4선승제로 벌어진다. 결승전에서 만난 두 팀 A와 B의 한 경기에서 A팀이 B팀을 이길 확률이 0.55 일 때, 다음 확률을 구하시오
- (a) A팀이 6게임 짜에서 우승할 확률
- (b) A팀이 우승할 확률
- (c) 이 두 팀이 5전 3선승제인 플레이오프에서 만났다면 A팀이 플레이오프에서 우승할 확률

- (a) 성공할 확률이 0.55, 실패할 확률이 $1 - 0.55$ 이므로 공식을 사용하면

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

- (b) $P(\text{A팀이 NBA 우승})$
 $= b^*(4; 4, 0.55) + b^*(5; 4, 0.55)$
 $+ b^*(6; 4, 0.55) + b^*(7; 4, 0.55)$
 $= 0.0915 + 0.1647 +$
 $0.1853 + 0.1665 = 0.6083$

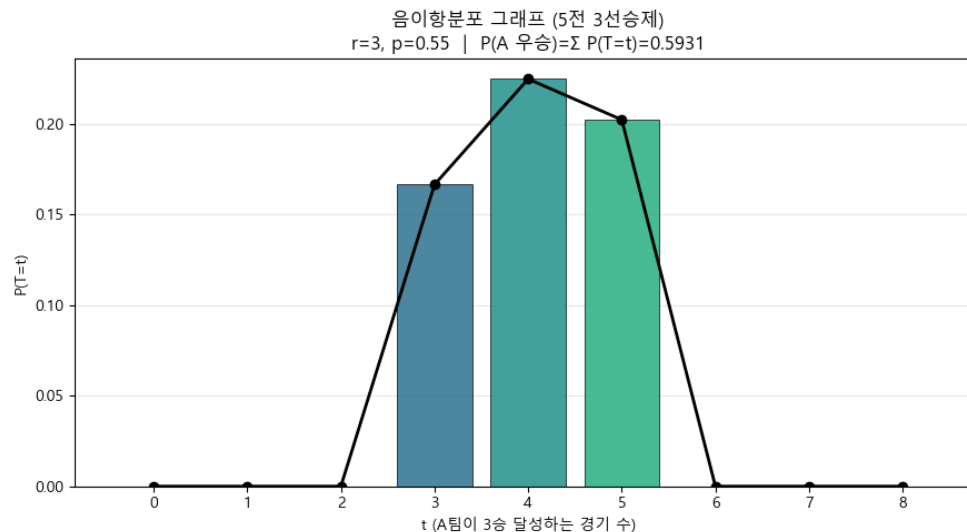


<그림 6> 예제 5.16-(b)의 음이항분포 그래프

음이항분포

• 예제 5.16 (2/2)

- NBA 결승전은 7전 4선승제로 벌어진다. 결승전에서 만난 두 팀 A와 B의 한 경기에서 A팀이 B팀을 이길 확률이 0.55 일 때, 다음 확률을 구하시오
 - (a) A팀이 6게임 짜에서 우승할 확률
 - (b) A팀이 우승할 확률
 - (c) 이 두 팀이 5전 3선승제인 플레이오프에서 만났다면 A팀이 플레이오프에서 우승할 확률
- (c) $P(\text{A팀이 플레이오프 우승}) = b^*(3; 3, 0.55) + b^*(4; 3, 0.55) + b^*(5; 3, 0.55)$
 $= 0.1664 + 0.2246 + 0.2021 = 0.5931$



<그림 7> 예제 5.16-(c)의 음이항분포 그래프

음이항분포

• 평균과 분산

음이항분포 $b^*(x; k, p)$ 의 평균과 분산은

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

• 증명 (1/2)

- 평균의 정의에 음이항분포 형식을 적용하면 $E(X) = \sum_{x=k}^{\infty} x \binom{x-1}{k-1} p^k (q)^{x-k}$
- $x \binom{x-1}{k-1} = \frac{x!}{(k-1)!(x-k)!} = k \frac{x!}{k!(x-k)!} = k \binom{x}{k}$ 로 변형
- 정리하면 $E(X) = \sum_{x=k}^{\infty} k \binom{x}{k} p^k (q)^{x-k}$
- 음이항분포 형식이 나오도록 k 를 분리하고 p 를 곱했다가 나누면
$$E(X) = k \sum_{x=k}^{\infty} \binom{x}{k} p^k q^{x-k} = \frac{k}{p} \sum_{x=k}^{\infty} \binom{x}{k} p^{k+1} q^{x-k}$$
- $x+1 = y, k+1 = m$ 으로 치환하면 $\frac{k}{p} \sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m}$
- $\sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m} = 1$ 이므로 음이항분포의 평균은 $\frac{k}{p}$

음이항분포

• 평균과 분산

• 증명 (2/2)

- $E(X^2) - (E(X))^2$ 에서 $E(X^2)$ 대신 $E[X(X-1)] + E(X)$ 으로 계산
- $E[X(X-1)] = \sum_{x=k}^{\infty} x(x-1) \binom{x-1}{k-1} p^k q^{x-k}$
- 평균 증명과정과 마찬가지로 $x(x-1) \binom{x-1}{k-1} = k(k+1) \binom{x}{k+1}$ 로 변형
- 정리하면 $E[X(X-1)] = \sum_{x=k}^{\infty} k(k+1) \binom{x}{k+1} p^k q^{x-k}$
- 우변을 음이항분포 꼴로 만들기 위해 p^2 을 분리, $\frac{k(k+1)}{p^2} \sum_{x=k}^{\infty} \binom{x}{k+1} p^{k+2} q^{x-k}$
- $x+1=y, k+2=m$ 으로 치환, $\sum_{y=m}^{\infty} \binom{y-1}{m-1} p^m q^{y-m+1}$
- $E[X(X-1)] = \frac{k(k+1)q}{p^2}, E(X) = \frac{k}{p}$ 이므로 $E(X^2) = \frac{k(k+1)q}{p^2} + \frac{k}{p}$
- $E(X^2) - \left(\frac{k}{p}\right)^2 = \frac{k(k+1)}{p^2} + \frac{k}{p} - \frac{k^2}{p^2} = \frac{k(k+1)q + kp - k^2}{p^2}$
- $q = 1-p$ 를 대입, 정리하면 $\frac{k(1-p)}{p^2}$

목 차

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

기하분포

- 정의

- 음이항분포에서 첫 번째 성공이 나올 때까지의 시행 횟수에 대한 확률 분포

- 수식

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

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

- 평균과 분산

- 기하분포를 따르는 확률변수의 평균과 분산은 각각 다음과 같다

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

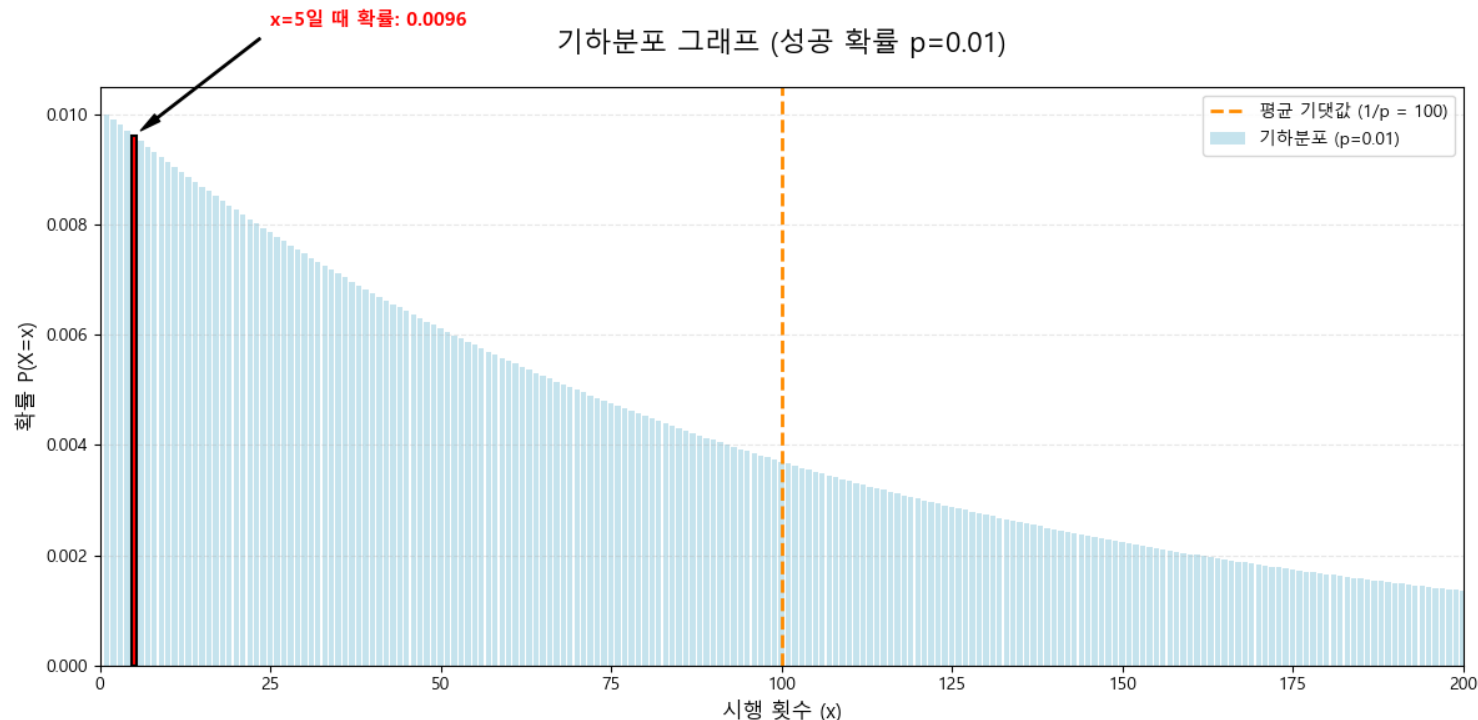
기하분포

• 예제 5.17

- 제조공정에서 100개의 제품마다 한 개의 불량품이 들어 있다고 한다. 제품을 하나씩 검사할 때 첫 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 5.17의 기하분포 그래프

기하분포

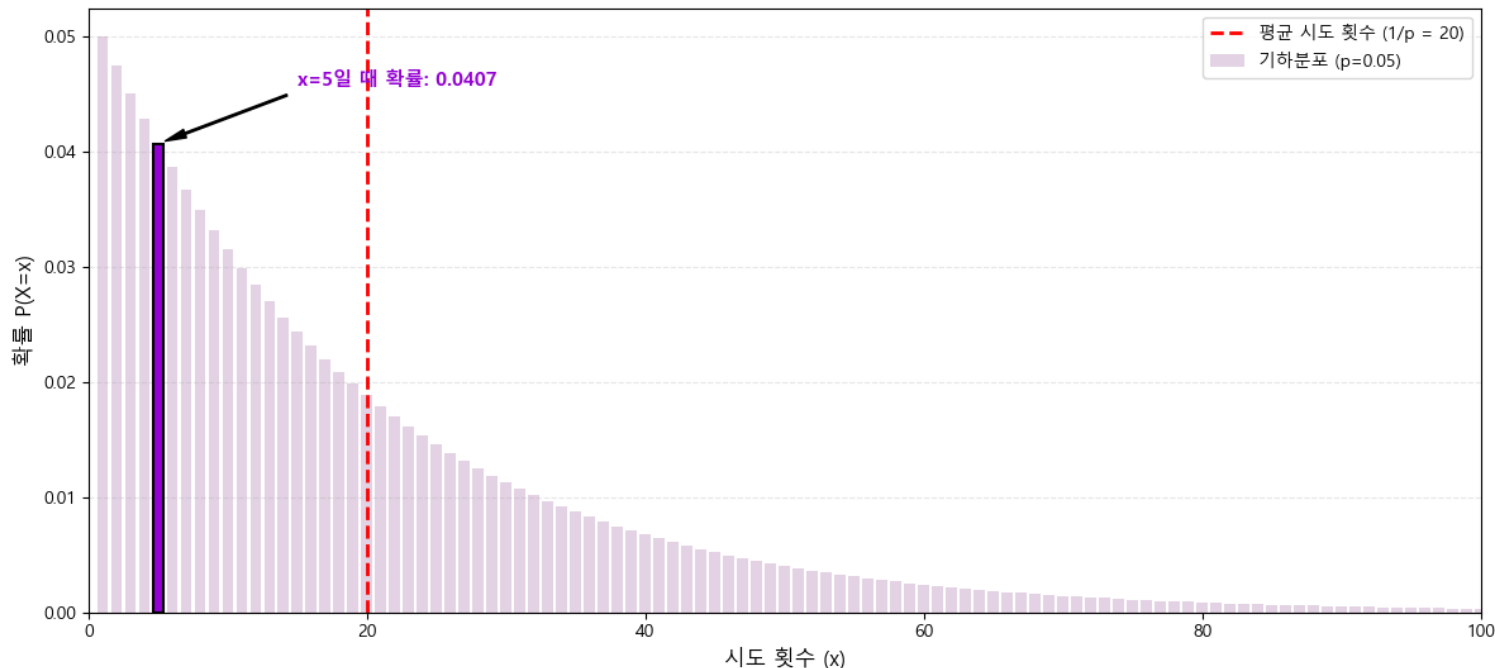
• 예제 5.18

- 전화통화가 물리는 시간대에 한 번의 시도로 상대방과 통화할 수 있는 확률이 0.05라고 할 때, 다섯 번째 시도에서 상대방과 통화할 수 있는 확률은 얼마인가?

- $x = 5, p = 0.05$ 인 기하분포를 따르므로

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

기하분포 그래프 (통화 성공 확률 $p=0.05$)



<그림 9> 예제 5.18의 기하분포 그래프

목 차

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

포아송분포

- 정의

- 일정한 시간 간격 또는 일정한 영역 안에서 발생하는 결과의 개수를 세는 실험의 확률 분포
 - e.g., PEL에 시간 당 오는 메일의 수

- 특징

1. 단위시간간격이나 일정영역에서 발생하는 결과의 수는 서로 겹치지 않는 다른 시간간격이나 영역에서 발생하는 수와 독립임을 가정
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

포아송분포

• 특징

• 증명

- t :시간, $N(t)$: $t = 0$ 부터 t 까지 발생한 사건의 확률변수일 때,
- 시간 t 까지 정확히 n 번 발생할 확률을 $P_n(t) = P(N(t) = n)$ 라고 가정
- 시간 $t + h$ 까지 0번 발생하기 위해선 시간 t 까지 0번, 그 다음 h 동안 0번 발생
- 특징 1에 의해 $P_0(t + h) = P_0(t)(1 - \lambda h + o(h))$ 가 성립
 - λ 는 단위시간 또는 단위면적에서 발생하는 결과의 평균 수
- 정리하면 $\frac{P_0(t+h)-P_0(t)}{h} = -\lambda P_0(t) + \frac{P_0(t)o(h)}{h}$
- 양변에 $\lim_{h \rightarrow 0}$ 을 취하면 $P'_0(t) = -\lambda P_0(t)$
- 양변을 $P_0(t)$ 로 나누고 적분한 후 지수를 취하면 $P_0(t) = e^C e^{-\lambda t}$, (C 는 적분상수)
- $N(0) = 0$ 이므로 $P_0(0) = P(N(0) = 0) = 1$
- 따라서 $1 = P_0(0) = e^C$, $P_0(t) = e^{-\lambda t}$
- $n = 1$ 일 때 $P_1(t) = \lambda t e^{-\lambda t}$, $n = 2$ 일 때 $P_2(t) = \frac{(\lambda t)^2}{2!} e^{-\lambda t}$
- 일반화하면 $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$

포아송분포

• 수식

- 일정한 시간간격 t 또는 일정영역 t 에서 발생하는 결과의 수를 나타내는 포아송 확률변수 X 의 확률분포는 다음과 같다.

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

- λ 는 단위시간 또는 단위면적에서 발생하는 결과의 평균 수
- $e = 2.71828 \dots$

• 평균과 분산

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

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

- 증명 과정은 부록#1 참조

포아송분포

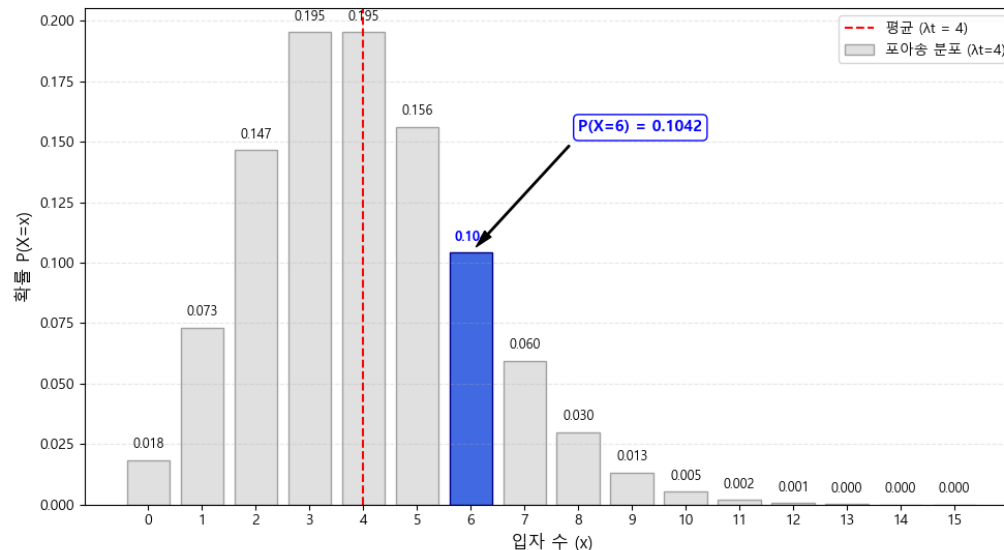
• 예제 5.19

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

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

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

포아송 분포 그래프 (예제 5.19, $\lambda t=4$)



<그림 10> 예제 5.19 포아송분포 그래프

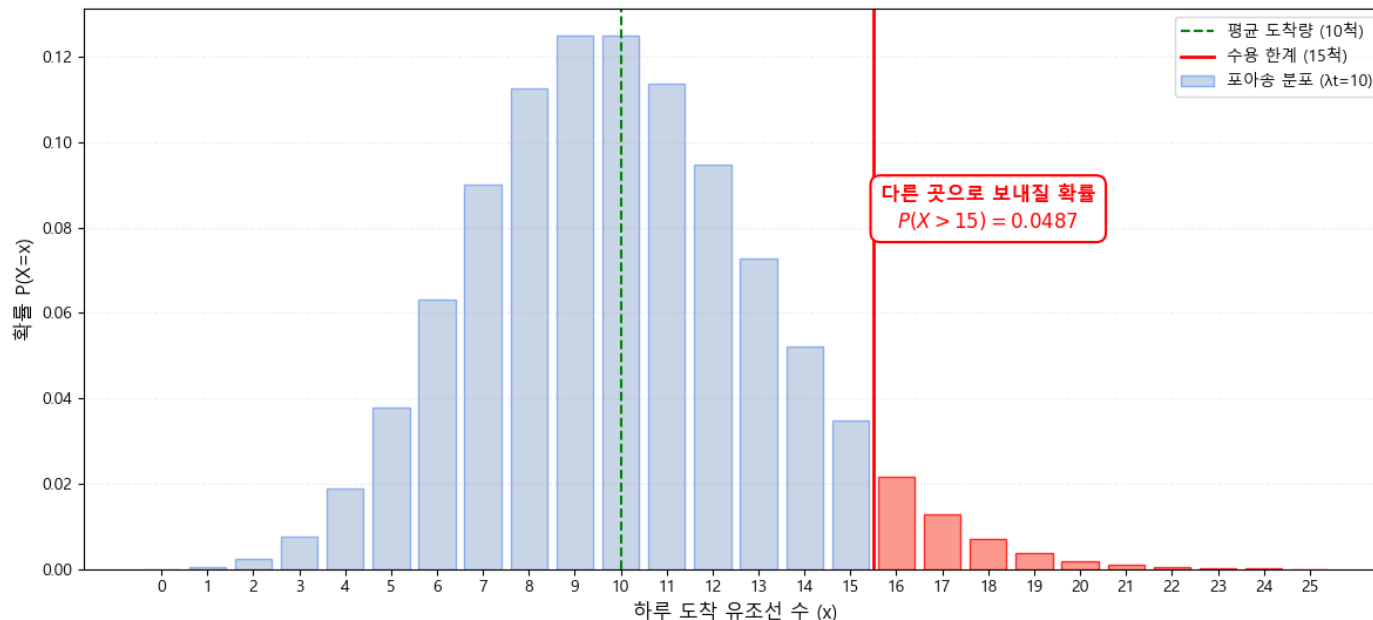
포아송분포

• 예제 5.20

- 어느 항구에 하루에 도착하는 유조선이 평균 10척이라고 하자. 그 항구는 하루에 최대 15척의 유조선을 취급할 수 있다고 할 때, 어느 날 유조선이 다른 곳으로 보내져야 할 확률을 구하시오

- $\lambda t = 10, f(x; 10) = \frac{e^{-10} 10^x}{x!}$ 인 포아송분포를 사용하면
- $P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513 = 0.0487$

포아송 분포 그래프 (예제 5.20, $\lambda t=10$)



<그림 11> 예제 5.20의 포아송분포 그래프

포아송분포

- 이항분포와의 관계

- 이항분포에서 시행횟수 n 이 매우 크고 성공확률 p 가 매우 작다면, 극한으로 보냈을 때 포아송분포와 동일한 특성을 보임
 - $n \geq 20, p \leq 0.05$ 기준을 대부분 채용

- 수식

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

포아송분포

• 증명

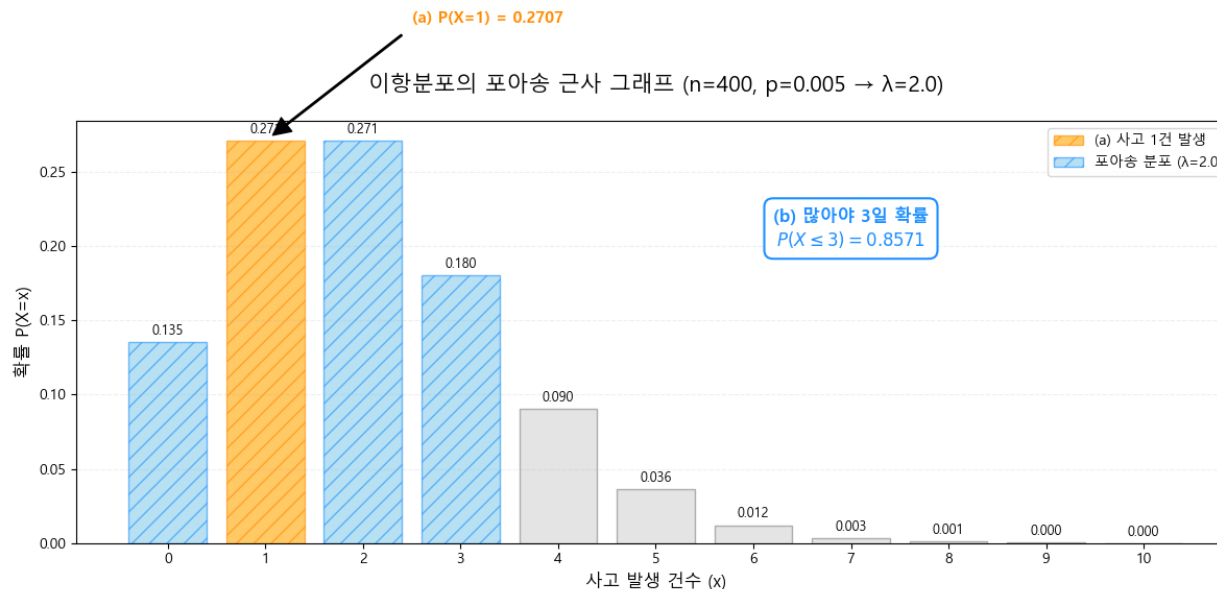
- $b(x; n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x}$
- $p = \frac{\mu}{n}$ 로 치환하면
- $b(x; n, p) = \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x}$
 $= 1 \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{x-1}{n}\right) \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x}$
- x 와 μ 가 상수일 때 $n \rightarrow \infty$ 임에 따라
- $\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n \cdots \left(1 - \frac{x-1}{n}\right) = 1, \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^{-x} = 1$ 이고 e 의 정의로부터
- $\lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n = \lim_{n \rightarrow \infty} \left\{ \left[1 + \frac{1}{\left(-\frac{n}{\mu}\right)} \right]^{-\frac{n}{\mu}} \right\}^{-\mu} = e^{-\mu}$
- 그러므로 주어진 극한조건 안에서

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

포아송분포

• 예제 5.21

- 사고가 일어날 확률이 0.005이고 사고들은 서로 독립적인 산업현장에서 일어난다고 할 때, 다음확률을 구하시오
 - (a) 400일 동안 사고가 한 건 발생할 확률
 - (b) 사고일이 많아야 3일일 확률
- X 를 $n = 400$ 이고 $p = 0.005$ 인 이항확률변수라고 하면 $np = 2$ 이고 포아송근사를 이용,
 - (a) $p(X = 1) = e^{-2}2^1 = 0.2707$
 - (b) $P(X \leq 3) = \sum_{x=0}^3 e^{-2}2^x/x! = 0.8571$



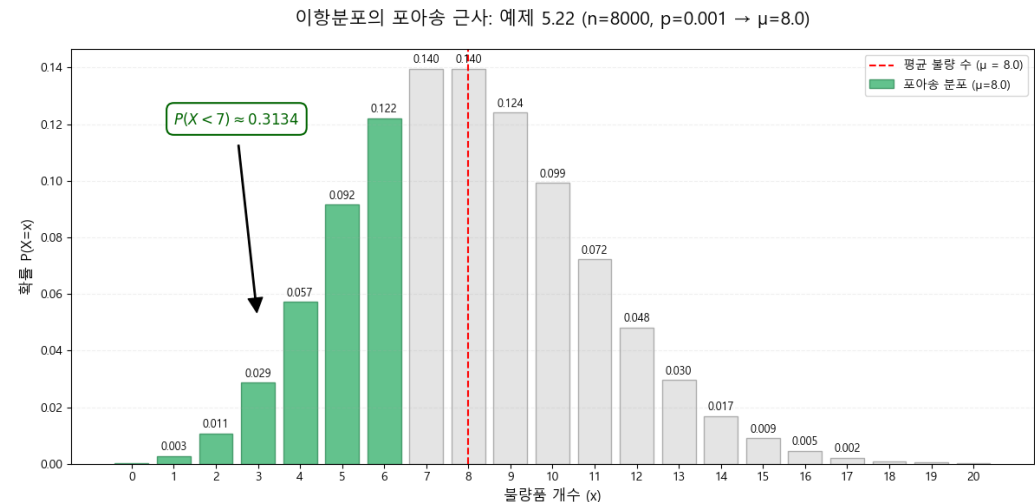
<그림 12> 예제 5.21의 포아송분포 그래프

포아송분포

• 예제 5.22

- 1000개의 제품마다 평균적으로 1개에 불량품이 나오는 공정이 있다고 할 때, 8000개의 확률표본 중에서 불량을 가지고 있는 제품이 7보다 작을 확률을 구하시오
- $n = 8000$ 이고 $p = 0.001$ 인 이항실험이지만, p 가 0에 근사한 값이고 n 이 매우 크므로 $\mu = (8000)(0.001) = 8$ 인 포아송 분포로 근사시킬 수 있음
- 구하고자 하는 확률은

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



<그림 13> 예제 5.22의 포아송분포 그래프

포아송분포

• 정리 (1/2)

분포	확률질량함수	평균	분산	사용 상황
이산형 균일분포	$f(x; k) = \frac{1}{k}$	$\frac{1}{k} \sum_{i=1}^k x_i$	$\frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2$	모든 결과가 동일한 확률 (주사위 등)
이항분포	$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$	np	npq	독립 베르누이 시행 n 번을 진행하여 성공 횟수를 파악
다항분포	$f(x_1, x_2, \dots, x_k; p_1, p_2, \dots, p_k, n)$ $= \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$	np_i	$np_i(1 - p_i)$	이항분포의 확장인 상황 (범주 3개 이상)

포아송분포

• 정리 (2/2)

분포	확률질량함수	평균	분산	사용 상황
초기하분포	$h(x; N, n, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$	$\frac{nk}{N}$	$\frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \left(1 - \frac{k}{N}\right)$	비복원추출을 진행하는 상황
음이항분포	$b^*(x; k, p) = \binom{x-1}{k-1} p^k q^{x-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$	r 번째 성공까지 시행 횟수를 파악
기하분포	$g(x; p) = pq^{x-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	첫 성공까지 시행 횟수를 파악
포아송분포	$\frac{e^{-\lambda t} (\lambda t)^x}{x!}$	λ	λ	단위 시간/공간 내 사건 횟수 파악

Thanks!

민 윤 홍(yunhong@pel.sejong.ac.kr)

부록#1 포아송분포 평균, 분산 증명

• 평균

- 포아송 분포 $p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, x = 0, 1, 2, \dots$
- 평균 $\mu = \sum_{x=0}^{\infty} xp(x; \lambda t) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda t}(\lambda t)^x}{x!}$
- $x \frac{(\lambda t)^x}{x!} = (\lambda t) \frac{(\lambda t)^{x-1}}{(x-1)!}$ 로 변형하고, $x = 0$ 일 때 값은 0 이므로
$$\mu = e^{-\lambda t} (\lambda t) \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1}}{(x-1)!}$$
- $x - 1 = k$ 라고 하면 $(\lambda t) e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!}$
- 지수급수 성질(테일러 전개)로 $\sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{\lambda t}$

부록#1 포아송분포 평균, 분산 증명

• 분산

- $E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)p(x; \lambda t) = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda t} (\lambda t)^x}{x!}$
- $x(x-1) \frac{(\lambda t)^x}{x!} = (\lambda t)^2 \frac{(\lambda t)^{x-2}}{(x-2)!}$
- $E([X(X-1)]) = e^{-\lambda t} (\lambda t)^2 \sum_{x=2}^{\infty} \frac{(\lambda t)^{x-2}}{(x-2)!} = (\lambda t)^2$
- $\sigma^2 = Var(X) = E(X^2) - \mu^2 = E[X(X-1)] + E(X) - \mu^2$
- 대입하면 $\sigma^2 = (\lambda t)^2 + (\lambda t) - (\lambda t)^2 = \lambda t$

부록#2 누적 포아송분포표

• $\sum_{x=0}^r p(x; \lambda)$ 의 누적 포아송분포표

r	$\lambda = 3$	$\lambda = 3.5$	$\lambda = 4$	$\lambda = 4.5$	$\lambda = 5.0$
0	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.4232	0.3208	0.2381	0.1736	0.1247
3	0.6472	0.5366	0.4335	0.3423	0.265
4	0.8153	0.7254	0.6288	0.5321	0.4405
5	0.9161	0.8576	0.7851	0.7029	0.616
6	0.9665	0.9347	0.8893	0.8311	0.7622
7	0.9881	0.9733	0.9489	0.9134	0.8666
8	0.9962	0.9901	0.9786	0.9597	0.9319
9	0.9989	0.9967	0.9919	0.9829	0.9682
10	0.9997	0.999	0.9972	0.9933	0.9863
11	0.9999	0.9997	0.9991	0.9976	0.9945
12	1	0.9999	0.9997	0.9992	0.998
13	1	1	0.9999	0.9997	0.9993
14	1	1	1	0.9999	0.9998

부록#2 누적 포아송분포표

• $\sum_{x=0}^r p(x; \lambda)$ 의 누적 포아송분포표

r	$\lambda = 6$	$\lambda = 7$	$\lambda = 8$	$\lambda = 9$	$\lambda = 10$
0	0.0025	0.0009	0.0003	0.0001	0.0000
1	0.0174	0.0073	0.0030	0.0012	0.0005
2	0.0620	0.0296	0.0138	0.0062	0.0028
3	0.1512	0.0818	0.0424	0.0212	0.0103
4	0.2851	0.1730	0.0996	0.0550	0.0293
5	0.4457	0.3007	0.1912	0.1157	0.0671
6	0.6063	0.4497	0.3134	0.2068	0.1301
7	0.7440	0.5987	0.4530	0.3239	0.2202
8	0.8472	0.7291	0.5925	0.4557	0.3328
9	0.9161	0.8305	0.7166	0.5874	0.4579
10	0.9574	0.9015	0.8159	0.7060	0.5830
11	0.9799	0.9467	0.8881	0.8030	0.6968
12	0.9912	0.9730	0.9362	0.8758	0.7916

부록#3 그래프 Python 코드

• 예제 5.6

```
1. import math
2. import numpy as np
3. import matplotlib.pyplot as plt
4. import matplotlib as mpl

5. mpl.rcParams['font.family'] = 'Malgun Gothic'
6. mpl.rcParams['axes.unicode_minus'] = False

7. n = 15
8. p = 0.4

9. x = np.arange(0, n + 1)

10. pmf = np.array([math.comb(n, xi) * (p ** xi) * ((1 - p) ** (n - xi)) for xi in x])

11. show_line = True
12. shade_a = True
13. shade_b = True
14. shade_c = True

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

16. plt.bar(x, pmf, width=0.8, alpha=0.35, label="PMF
    (bars)")
```

```
17. if show_line:
18.     plt.plot(x, pmf, marker="o", linewidth=2,
        label="PMF (points+line)")

19. def shade_range(x_lo, x_hi, label):
20.     mask = (x >= x_lo) & (x <= x_hi)
21.     plt.bar(x[mask], pmf[mask], width=0.8,
        alpha=0.6, label=label)

22. if shade_a:
23.     shade_range(10, n, "(a)  $X \geq 10$ ")

24. if shade_b:
25.     shade_range(3, 8, "(b)  $3 \leq X \leq 8$ ")

26. if shade_c:
27.     shade_range(5, 5, "(c)  $X = 5$ ")

28. mu = n * p
29. var = n * p * (1 - p)

30. plt.title(f"이항분포그래프: 인원={n}, 회복확률={p}
    (mean={mu:.2f}, var={var:.2f})")
31. plt.xlabel("x")
32. plt.ylabel("P(X = x)")
33. plt.xticks(x)
```

부록#3 그래프 Python 코드

• 예제 5.6

```
34. plt.grid(True, axis="y", alpha=0.3)
35. plt.legend()
36. plt.tight_layout()
37. plt.show()

38. pa = pmf[x >= 10].sum()
39. pb = pmf[(x >= 3) & (x <= 8)].sum()
40. pc = pmf[x == 5].sum()

41. print(f"P(X >= 10) = {pa:.6f}")
42. print(f"P(3 <= X <= 8) = {pb:.6f}")
43. print(f"P(X = 5) = {pc:.6f}")
```

부록#3 그래프 Python 코드

• 예제 5.10

```
1. import math
2. import numpy as np
3. import matplotlib.pyplot as plt
4. import matplotlib as mpl
5. import matplotlib.cm as cm
6. import matplotlib.colors as colors
7. from mpl_toolkits.mplot3d import Axes3D # noqa:
   F401

8. mpl.rcParams['font.family'] = 'Malgun Gothic'
9. mpl.rcParams['axes.unicode_minus'] = False

10. # -----
11. # Multinomial example 5.10
12. # -----
13. n = 6
14. p1 = 2/9
15. p2 = 1/6
16. p3 = 11/18

17. def multinom_pmf(x1, x2, x3):
18.     coef = math.factorial(n) / (
19.         math.factorial(x1) *
20.         math.factorial(x2) *
21.         math.factorial(x3)
22.     )
23.     return coef * (p1**x1) * (p2**x2) * (p3**x3)
```

```
24. x_vals = []
25. y_vals = []
26. z_vals = []

27. for x1 in range(n + 1):
28.     for x2 in range(n + 1 - x1):
29.         x3 = n - x1 - x2
30.         prob = multinom_pmf(x1, x2, x3)

31.         x_vals.append(x1)
32.         y_vals.append(x2)
33.         z_vals.append(prob)

34. x_vals = np.array(x_vals)
35. y_vals = np.array(y_vals)
36. z_vals = np.array(z_vals)

37. dx = 0.5
38. dy = 0.5
39. z_base = np.zeros_like(z_vals)

40. # 색상 그라데이션 설정
41. norm = colors.Normalize(vmin=min(z_vals),
42.                          vmax=max(z_vals))
42. cmap = cm.viridis # viridis / plasma / inferno
   / coolwarm 등
43. bar_colors = cmap(norm(z_vals))
```

부록#3 그래프 Python 코드

• 예제 5.10

```
44. #불투명도(alpha)
45. alpha_value = 0.8
46. bar_colors[:, -1] = alpha_value # 마지막 채널이 alpha

47. # 그래프 생성
48. fig = plt.figure(figsize=(10, 7))
49. ax = fig.add_subplot(111, projection='3d')

50. ax.bar3d(
51.     x_vals,
52.     y_vals,
53.     z_base,
54.     dx,
55.     dy,
56.     z_vals,
57.     color=bar_colors,
58.     shade=True
59. )

60. ax.set_title(
61.     f"다항분포 3D 막대그래프 (n={n})\n"
62.     f"p1={p1:.3f}, p2={p2:.3f}, p3={p3:.3f}"
63. )
64. ax.set_xlabel("x1 (활주로 1)")
65. ax.set_ylabel("x2 (활주로 2)")
66. ax.set_zlabel("확률")
```

```
67. # x3 텍스트 공간 표시
68. ax.text(
69.     n/2,
70.     -1.2,
71.     max(z_vals)*1.1,
72.     "x3 (활주로 3 도착 대수)",
73.     fontsize=12
74. )

75. # 컬러바 추가
76. mappable = cm.ScalarMappable(norm=norm, cmap=cmap)
77. mappable.set_array([])
78. cbar = fig.colorbar(mappable, ax=ax, pad=0.1)
79. cbar.set_label("확률 크기")

80. plt.tight_layout()
81. plt.show()
```

부록#3 그래프 Python 코드

• 예제 5.11

```
1. import math
2. import numpy as np
3. import matplotlib.pyplot as plt
4. import matplotlib as mpl
5. import matplotlib.cm as cm
6. import matplotlib.colors as colors

7. mpl.rcParams['font.family'] = 'Malgun Gothic'
8. mpl.rcParams['axes.unicode_minus'] = False
9. N = 40
10. K = 3
11. n = 5
12. def hypergeom_pmf(x, N, K, n):
13.     if x < 0 or x > K or x > n:
14.         return 0.0
15.     if n - x > N - K:
16.         return 0.0
17.     return (math.comb(K, x) * math.comb(N - K, n -
18.         x)) / math.comb(N, n)
19. # x축 0~5
20. x = np.arange(0, n + 1)
21. pmf = np.array([hypergeom_pmf(xi, N, K, n) for xi in
22.     x])
23. plt.figure(figsize=(9, 5))
```

```
24. cmap = cm.plasma # viridis / plasma / inferno /
25.     coolwarm 가능
26. norm = colors.Normalize(vmin=min(x), vmax=max(x))
27. bar_colors = cmap(norm(x))
28. plt.bar(x, pmf, width=0.8, color=bar_colors,
29.     alpha=0.8)
30. plt.plot(x, pmf, marker="o", linewidth=2,
31.     color="black")
32. # 확률값 표시
33. for xi, pi in zip(x, pmf):
34.     if pi > 0:
35.         plt.text(
36.             xi,
37.             pi,
38.             f"{pi:.4f}",
39.             ha='center',
40.             va='bottom',
41.             fontsize=11,
42.             fontweight='bold',
43.             bbox=dict(
44.                 facecolor='white',
45.                 edgecolor='none',
46.                 alpha=0.8,
47.                 pad=2
48.             ),
49.             zorder=5
50.         )
```

부록#3 그래프 Python 코드

• 예제 5.11

```
46. plt.title(f"초기하분포 그래프\ nN={N}, K={K}, n={n}")
47. plt.xlabel("x (표본에서 발견된 불량품 개수)")
48. plt.ylabel("P(X = x)")
49. plt.xticks(x)
50. plt.grid(True, axis="y", alpha=0.3)
51. plt.tight_layout()
52. plt.show()

53. print("가능한 x:", list(x))
54. for xi, pi in zip(x, pmf):
55.     print(f"P(X={xi}) = {pi:.6f}")
```

부록#3 그래프 Python 코드

• 예제 5.16 – (b)

```
1. import math
2. import numpy as np
3. import matplotlib.pyplot as plt
4. import matplotlib as mpl
5. import matplotlib.cm as cm
6. import matplotlib.colors as colors

7. mpl.rcParams['font.family'] = 'Malgun Gothic'
8. mpl.rcParams['axes.unicode_minus'] = False

9. p = 0.55
10. r = 4

11. # 0~10까지 표시하되 7까지만 확률 존재
12. t = np.arange(0, 11)

13. def negbin_T_pmf(t, r, p):
14.     if t < r:
15.         return 0.0
16.     return math.comb(t - 1, r - 1) * (p ** r) * ((1
17. - p) ** (t - r))
17. # 원래 음이항 확률
18. pmf_raw = np.array([negbin_T_pmf(ti, r, p) for ti in
19. t])
19. pmf_truncated = pmf_raw.copy()
20. pmf_truncated[t > 7] = 0
```

```
21. # 정규화
22. normalizer = pmf_truncated.sum()
23. pmf_truncated = pmf_truncated / normalizer

24. # 색상
25. cmap = cm.viridis
26. norm = colors.Normalize(vmin=min(t), vmax=max(t))
27. bar_colors = cmap(norm(t))
28. bar_colors[:, -1] = 0.85

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

30. plt.bar(t, pmf_truncated,
31.         width=0.8,
32.         color=bar_colors,
33.         edgecolor="black",
34.         linewidth=0.5,
35.         zorder=2)

36. plt.plot(t, pmf_truncated,
37.         marker="o",
38.         linewidth=2,
39.         color="black",
40.         zorder=3)
```

부록#3 그래프 Python 코드

• 예제 5.16 – (b)

```
41. plt.title("음이항분포 그래프\nr=4, p=0.55")
42. plt.xlabel("t (우승 확정 경기 수)")
43. plt.ylabel("P*(T=t)")
44. plt.xticks(t)
45. plt.grid(True, axis="y", alpha=0.3)
46. plt.tight_layout()
47. plt.show()

48. print("절단 후 확률 합:", pmf_truncated.sum())
```

부록#3 그래프 Python 코드

• 예제 5.16 – (c)

```
1. import math
2. import numpy as np
3. import matplotlib.pyplot as plt
4. import matplotlib as mpl
5. import matplotlib.cm as cm
6. import matplotlib.colors as colors

7. mpl.rcParams['font.family'] = 'Malgun Gothic'
8. mpl.rcParams['axes.unicode_minus'] = False

9. p = 0.55
10. r = 3 # 3번째 승리

11. # x축 0~8까지 (보기 좋게 확장)
12. t = np.arange(0, 9)

13. def negbin_T_pmf(ti, r, p):
14.     if ti < r:
15.         return 0.0
16.     return math.comb(ti - 1, r - 1) * (p ** r) * ((1 - p) ** (ti - r))

17. # 이론적 음이항
18. pmf_raw = np.array([negbin_T_pmf(ti, r, p) for ti in t])
```

```
19. # 5전 3선승제 반영 (5 초과는 0)
20. pmf_series = pmf_raw.copy()
21. pmf_series[t > 5] = 0.0

22. # 우승 확률
23. p_win = pmf_series.sum()

24. # 색상 설정
25. cmap = cm.viridis
26. norm = colors.Normalize(vmin=min(t), vmax=max(t))
27. bar_colors = cmap(norm(t))
28. bar_colors[:, -1] = 0.85

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

30. plt.bar(
31.     t, pmf_series,
32.     width=0.8,
33.     color=bar_colors,
34.     edgecolor="black",
35.     linewidth=0.5,
36.     zorder=2
37. )
```

부록#3 그래프 Python 코드

• 예제 5.16 – (c)

```
38. plt.plot(t, pmf_series,  
39.         marker="o",  
40.         linewidth=2,  
41.         color="black",  
42.         zorder=3)  
  
43. plt.title(  
44.     f"음이항분포 그래프 (5전 3선승제)\n"  
45.     f"r=3, p={p} | P(A 우승)=Σ P(T=t)={p_win:.4f}"  
46. )  
  
47. plt.xlabel("t (A팀이 3승 달성하는 경기 수)")  
48. plt.ylabel("P(T=t)")  
49. plt.xticks(t)  
50. plt.grid(True, axis="y", alpha=0.3)  
51. plt.tight_layout()  
52. plt.show()  
  
53. # 검산 출력  
54. for ti, pi in zip(t, pmf_series):  
55.     print(f"P(T={ti}) = {pi:.6f}")  
56. print("P(A 우승) =", p_win)
```

부록#3 그래프 Python 코드

• 예제 5.17

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import geom

5. # 한글 폰트 설정
6. mpl.rcParams['font.family'] = 'Malgun Gothic'
7. mpl.rcParams['axes.unicode_minus'] = False

8. p = 0.01 # 성공 확률 (예: 100개 중 1개의 불량품)
9. target_x = 5 # 예제에서 질문한 5번째 시행

10. # x축 범위 설정 (성공 확률이 낮으므로 넉넉하게 200까지 설정)
11. x = np.arange(1, 201)

12. # 확률질량함수(PMF) 계산:  $g(x; p) = p * (1-p)^{(x-1)}$ 
13. pmf_values = geom.pmf(x, p)

14. plt.figure(figsize=(12, 6))

15. # 1. 전체 분포 그리기
16. plt.bar(x, pmf_values, color='lightblue', alpha=0.7,
          label=f'기하분포 (p={p})')
```

```
17. # 2. 예제 포인트(x=5) 강조하기
18. prob_at_5 = geom.pmf(target_x, p)
19. plt.bar(target_x, prob_at_5, color='red',
          edgecolor='black', linewidth=1.5, zorder=3)
20. plt.annotate(f'x=5일 때 확률: {prob_at_5:.4f}',
21.              xy=(target_x, prob_at_5),
22.              xytext=(target_x + 20, prob_at_5 +
23.                      0.002),
24.              arrowprops=dict(facecolor='black',
25.                              shrink=0.05, width=1, headwidth=5),
26.              fontsize=11, fontweight='bold',
27.              color='red')

28. # 3. 평균선 표시 ( $E(X) = 1/p$ )
29. mean_x = 1 / p
30. plt.axvline(mean_x, color='darkorange',
31.             linestyle='--', linewidth=2, label=f'평균 기댓값
32.             (1/p = {int(mean_x)})')

33. # 그래프 꾸미기
34. plt.title(f'기하분포 그래프 (성공 확률 p={p})',
35.          fontsize=16, pad=20)
36. plt.xlabel("시행 횟수 (x)", fontsize=12)
37. plt.ylabel("확률 P(X=x)", fontsize=12)
38. plt.legend()
39. plt.grid(True, axis='y', alpha=0.3, linestyle='--')
```

부록#3 그래프 Python 코드

• 예제 5.17

```
34. # x축 범위 제한
35. plt.xlim(0, 200)
36. plt.tight_layout()
37. plt.show()

38. print(f"첫 불량품이 5번째에서 발견될 확률:
    {prob_at_5:.4f}")
```

부록#3 그래프 Python 코드

• 예제 5.18

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import geom

5. # 한글 폰트 설정
6. mpl.rcParams['font.family'] = 'Malgun Gothic'
7. mpl.rcParams['axes.unicode_minus'] = False

8. # -----
9. # 기하분포 파라미터 설정 (통화 시도 예제)
10. # -----
11. p = 0.05      # 한 번의 시도로 통화할 수 있는 확률
12. target_x = 5  # 다섯 번째 시도에서 성공할 확률

13. # x축 범위 설정 (성공 확률이 0.05이므로 100회 정도까지
    시각화)
14. x = np.arange(1, 101)

15. # 확률질량함수(PMF) 계산:  $g(x; p) = p * (1-p)^{(x-1)}$ 
16. pmf_values = geom.pmf(x, p)

17. plt.figure(figsize=(12, 6))

18. # 1. 전체 분포 그리기
19. plt.bar(x, pmf_values, color='thistle', alpha=0.7,
    label=f'기하분포 (p={p})')
```

```
20. # 2. 예제 포인트(x=5) 강조하기
21. prob_at_5 = geom.pmf(target_x, p) # 계산 결과:
    0.0407
22. plt.bar(target_x, prob_at_5, color='darkviolet',
    edgecolor='black', linewidth=1.5, zorder=3)
23. plt.annotate(f'x=5일 때 확률: {prob_at_5:.4f}',
24.               xy=(target_x, prob_at_5),
25.               xytext=(target_x + 10, prob_at_5 +
    0.005),
26.               arrowprops=dict(facecolor='black',
    shrink=0.05, width=1, headwidth=5),
27.               fontsize=11, fontweight='bold',
    color='darkviolet')

28. # 3. 평균선 표시 ( $E(X) = 1/p$ )
29. mean_x = 1 / p
30. plt.axvline(mean_x, color='red', linestyle='--',
    linewidth=2, label=f'평균 시도 횟수 ( $1/p =$ 
    {int(mean_x)})')

31. # 그래프 꾸미기
32. plt.title(f'기하분포 그래프 (통화 성공 확률 p={p})',
    fontsize=16, pad=20)
33. plt.xlabel("시도 횟수 (x)", fontsize=12)
34. plt.ylabel("확률 P(X=x)", fontsize=12)
35. plt.legend()
36. plt.grid(True, axis='y', alpha=0.3, linestyle='--')
```

부록#3 그래프 Python 코드

• 예제 5.18

```
37. # x축 범위 제한
38. plt.xlim(0, 100)
39. plt.tight_layout()
40. plt.show()

41. print(f"다섯 번째 시도에서 통화할 확률: {prob_at_5:.4f}")
```

부록#3 그래프 Python 코드

• 예제 5.19

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import poisson

5. mpl.rcParams['font.family'] = 'Malgun Gothic'
6. mpl.rcParams['axes.unicode_minus'] = False

7. lamb_t = 4 # 평균 발생 수 ( $\lambda t$ )
8. target_x = 6 # 우리가 구하고자 하는 사건 횟수
9. x = np.arange(0, 16)
10. pmf_values = poisson.pmf(x, lamb_t)

11. plt.figure(figsize=(10, 6))

12. # 1. 전체 분포 막대 그래프 그리기
13. bars = plt.bar(x, pmf_values, color='lightgray',
    edgcolor='gray', alpha=0.7, label=f'포아송 분포
    ( $\lambda t = \{lamb\_t\}$ )')

14. # 2. 예제 포인트(x=6) 강조하기
15. prob_at_6 = poisson.pmf(target_x, lamb_t) # 계산 결과:
    약 0.1042
16. bars[target_x].set_color('royalblue')
17. bars[target_x].set_edgcolor('navy')
18. bars[target_x].set_alpha(1.0)
```

```
19. # 3. 평균( $\lambda t$ ) 위치 표시
20. plt.axvline(lamb_t, color='red', linestyle='--',
    linewidth=1.5, label=f'평균 ( $\lambda t = \{lamb\_t\}$ )')

21. # 4. 막대 위에 확률값 표시
22. for i, prob in enumerate(pmf_values):
23.     color = 'blue' if i == target_x else 'black'
24.     fontweight = 'bold' if i == target_x else
        'normal'
25.     plt.text(i, prob + 0.005, f'{prob:.3f}',
        ha='center', fontsize=9, color=color,
        fontweight=fontweight)

26. # 그래프 꾸미기
27. plt.title(f"포아송 분포 그래프 (예제 5.19,
     $\lambda t = \{lamb\_t\}$ )", fontsize=15, pad=20)
28. plt.xlabel("입자 수 (x)", fontsize=12)
29. plt.ylabel("확률 P(X=x)", fontsize=12)
30. plt.xticks(x)
31. plt.legend()
32. plt.grid(True, axis='y', alpha=0.3, linestyle='--')
```

부록#3 그래프 Python 코드

• 예제 5.19

```
33. # 5. 예제 계산식 주석 추가
34. plt.annotate(f'P(X=6) = {prob_at_6:.4f}',
35.              xy=(target_x, prob_at_6),
36.              xytext=(target_x + 2, prob_at_6 + 0.05),
37.              arrowprops=dict(facecolor='black',
38.                              shrink=0.05, width=1, headwidth=6),
39.              fontsize=11, fontweight='bold',
40.              color='blue', bbox=dict(boxstyle="round,pad=0.3",
41.                                      fc="white", ec="blue", lw=1))

39. plt.tight_layout()
40. plt.show()

41. print(f"평균이 4일 때, 6개의 입자가 통과할 확률:
        {prob_at_6:.4f}")
```

부록#3 그래프 Python 코드

• 예제 5.20

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import poisson

5. mpl.rcParams['font.family'] = 'Malgun Gothic'
6. mpl.rcParams['axes.unicode_minus'] = False

7. lamb_t = 10 # 하루 평균 도착 유조선 수 ( $\lambda t$ )
8. capacity = 15 # 항구의 최대 수용 능력

9. # x축 범위 설정
10. x = np.arange(0, 26)

11. # 확률질량함수(PMF) 계산
12. pmf_values = poisson.pmf(x, lamb_t)

13. plt.figure(figsize=(12, 6))

14. # 1. 전체 분포 그리기
15. bars = plt.bar(x, pmf_values, color='lightsteelblue',
    edgcolor='cornflowerblue', alpha=0.7, label=f'포아송
    분포 ( $\lambda t = \{lamb\_t\}$ )')
```

```
16. # 2. 용량 초과 부분( $x > 15$ ) 강조하기
17. for i in range(capacity + 1, len(x)):
18.     bars[i].set_color('salmon')
19.     bars[i].set_edgcolor('red')
20.     bars[i].set_alpha(0.8)

21. # 3. 평균 및 수용 한계선 표시
22. plt.axvline(lamb_t, color='green', linestyle='--',
    linewidth=1.5, label=f'평균 도착량 ( $\{lamb\_t\}$ 척)')
23. plt.axvline(capacity + 0.5, color='red',
    linestyle='-', linewidth=2, label=f'수용 한계
    ( $\{capacity\}$ 척)')

24. # 4. 확률 합계 계산 및 텍스트 표시
25. #  $P(X > 15) = 1 - P(X \leq 15)$ 
26. prob_over_capacity = 1 - poisson.cdf(capacity,
    lamb_t) # 계산 결과: 0.0487

27. plt.text(18, 0.08, f'다른 곳으로 보내질 확률\n $P(X > 15) = \{prob\_over\_capacity:.4f\}\$',
    fontsize=12, color='red',
    fontweight='bold', ha='center',
    28.
    29.     bbox=dict(boxstyle="round,pad=0.5",
    fc="white", ec="red", lw=1.5))$ 
```

부록#3 그래프 Python 코드

• 예제 5.20

```
30. plt.title(f"포아송 분포 그래프 (예제 5.20,  $\lambda_t=\{\text{lamb\_t}\}$ )",  
    fontsize=16, pad=20)  
31. plt.xlabel("하루 도착 유조선 수 (x)", fontsize=12)  
32. plt.ylabel("확률  $P(X=x)$ ", fontsize=12)  
33. plt.xticks(x)  
34. plt.legend()  
35. plt.grid(True, axis='y', alpha=0.2, linestyle='--')  
  
36. plt.tight_layout()  
37. plt.show()  
  
38. print(f"유조선이 다른 곳으로 보내져야 할 확률:  
    {prob_over_capacity:.4f}")
```

부록#3 그래프 Python 코드

• 예제 5.21

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import poisson

5. mpl.rcParams['font.family'] = 'Malgun Gothic'
6. mpl.rcParams['axes.unicode_minus'] = False

7. n = 400
8. p = 0.005
9. lamb = n * p #  $\lambda = 2$ 

10. x = np.arange(0, 11)
11. pmf_values = poisson.pmf(x, lamb)
12. plt.figure(figsize=(12, 6))

13. # 1. 전체 분포 그리기
14. bars = plt.bar(x, pmf_values, color='lightgray',
    edgcolor='gray', alpha=0.6, label=f'포아송 분포
    ( $\lambda={lamb}$ )')

15. # 2. (a)번 케이스 강조: 사고가 정확히 1건 발생 (x=1)
16. prob_a = pmf_values[1] # 0.2707
17. bars[1].set_color('orange')
18. bars[1].set_edgcolor('darkorange')
19. bars[1].set_label('(a) 사고 1건 발생')
```

```
20. # 3. 사고가 많아야 3일 ( $x \leq 3$ )
21. for i in range(4):
22.     if i != 1: # 1은 (a)에서 이미 칠함
23.         bars[i].set_color('skyblue')
24.         bars[i].set_edgcolor('dodgerblue')
25.         bars[i].set_hatch('///')
26. # 4. 확률 수치 표시
27. for i, prob in enumerate(pmf_values):
28.     plt.text(i, prob + 0.005, f'{prob:.3f}',
        ha='center', fontsize=9)

29. # 5. 수식 및 결과 주석 추가
30. plt.annotate(f'(a)  $P(X=1) = {prob_a:.4f}$ ',
31.             xy=(1, prob_a), xytext=(2.5, 0.35),
32.             arrowprops=dict(facecolor='black',
        shrink=0.05, width=1),
33.             fontsize=11, fontweight='bold',
        color='darkorange')

34. prob_b = poisson.cdf(3, lamb) # 0.8571
35. plt.text(7, 0.20, f'(b) 많아야 3일 확률  $\sum_{x \leq 3} P(X \leq 3) = {prob_b:.4f}$ ',
36.         fontsize=12, color='dodgerblue',
        fontweight='bold', ha='center',
37.         bbox=dict(boxstyle="round,pad=0.5",
        fc="white", ec="dodgerblue", lw=1.5))
```

부록#3 그래프 Python 코드

• 예제 5.21

```
38. plt.title(f"이항분포의 포아송 근사 그래프 (n={n}, p={p} →  
    λ={lamb})", fontsize=15, pad=20)  
39. plt.xlabel("사고 발생 건수 (x)", fontsize=12)  
40. plt.ylabel("확률 P(X=x)", fontsize=12)  
41. plt.xticks(x)  
42. plt.legend()  
43. plt.grid(True, axis='y', alpha=0.2, linestyle='--')  
  
44. plt.tight_layout()  
45. plt.show()  
  
46. print(f"(a) 1건 발생 확률: {prob_a:.4f}")  
47. print(f"(b) 3건 이하 발생 확률: {prob_b:.4f}")
```

부록#3 그래프 Python 코드

• 예제 5.22

```
1. import numpy as np
2. import matplotlib.pyplot as plt
3. import matplotlib as mpl
4. from scipy.stats import poisson

5. mpl.rcParams['font.family'] = 'Malgun Gothic'
6. mpl.rcParams['axes.unicode_minus'] = False

7. n = 8000
8. p = 0.001
9. mu = n * p #  $\mu = 8$  (포아송 분포의 평균)

10. x = np.arange(0, 21)
11. pmf_values = poisson.pmf(x, mu)
12. plt.figure(figsize=(12, 6))

13. # 1. 전체 분포 그리기
14. bars = plt.bar(x, pmf_values, color='lightgray',
    edgcolor='gray', alpha=0.6, label=f'포아송 분포 ( $\mu={\mu}$ )')

15. # 2. 구하고자 하는 확률 영역 강조
16. for i in range(7):
17.     bars[i].set_color('mediumseagreen')
18.     bars[i].set_edgcolor('seagreen')
19.     bars[i].set_alpha(0.8)
```

```
20. # 3. 평균( $\mu$ ) 위치 표시
21. plt.axvline(mu, color='red', linestyle='--',
    linewidth=1.5, label=f'평균 불량 수 ( $\mu = {\mu}$ )')

22. # 4. 확률 수치 표시 (막대 위)
23. for i, prob in enumerate(pmf_values):
24.     if prob > 0.001: # 너무 낮은 확률은 표시 생략
25.         plt.text(i, prob + 0.002, f'{prob:.3f}',
            ha='center', fontsize=9)

26. # 5. 결과값 주석 추가
27. #  $P(X < 7) = P(X \leq 6)$ 
28. prob_res = poisson.cdf(6, mu) # 약 0.3134
29. plt.annotate(f'$P(X < 7) \approx {prob_res:.4f}$',
    xy=(3, 0.05), xytext=(1, 0.12),
    arrowprops=dict(facecolor='black',
    shrink=0.05, width=1),
    fontsize=12, fontweight='bold',
    color='darkgreen',
    bbox=dict(boxstyle="round,pad=0.5",
    fc="white", ec="darkgreen", lw=1.5))
```

부록#3 그래프 Python 코드

• 예제 5.22

```
34. # 그래프 꾸미기
35. plt.title(f"이항분포의 포아송 근사: 예제 5.22 (n={n},  
    p={p} →  $\mu={\mu}$ )", fontsize=15, pad=20)
36. plt.xlabel("불량품 개수 (x)", fontsize=12)
37. plt.ylabel("확률 P(X=x)", fontsize=12)
38. plt.xticks(x)
39. plt.legend()
40. plt.grid(True, axis='y', alpha=0.2, linestyle='--')

41. plt.tight_layout()
42. plt.show()

43. print(f"불량품이 7개보다 적을 확률 P(X < 7):  
    {prob_res:.4f}")
```