생생한 사례로 배우는 확률과 통계

[강의교안 이용 안내]

한빛아카데미
Hanbit Academy, Inc.

Chapter 06

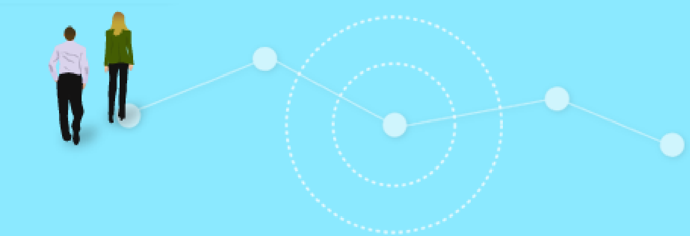
생생한 사례로 배우는
확률과 통계
PROBABILITY & STATISTICS

확률분포

Probability Distribution

목 차

6.1 이산확률분포



이산균등분포

이산균등분포[discrete uniform distribution] :
이산확률변수 X의 상태공간 $S_X = \{1, 2, \ldots, n\}$에서 확률질량함수가 균등한 확률분포로 $X \sim DU(n)$으로 나타낸다.

$$f(x) = \begin{cases} \dfrac{1}{n} & , x = 1, 2, \cdots, n \\ 0 & ,\text{다른 곳에서} \end{cases}$$

이산균등분포

❖ 이산균등분포 $X \sim DU(n)$의 평균과 분산

X의 평균 : $$\mu = E(X) = \sum_{x=1}^{n} x \cdot \frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n} x = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

X의 2차 적률 : $$E(X^2) = \sum_{x=1}^{n} x^2 \cdot \frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

X의 분산 : $$\sigma^2 = E(X^2) - E(X)^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{n^2-1}{12}$$

① 평균 : $\mu = E(X) = \dfrac{n+1}{2}$

② 분산 : $\sigma^2 = \dfrac{n^2-1}{12}$

이산균등분포

예제 6-1

각 면에 1, 2, 3, 4의 번호가 적힌 공정한 사면체를 던져 바닥에 놓인 수를 확률변수 X라 할 때, 확률질량함수와 평균 및 분산을 구하라.

풀이

공정한 사면체를 던져서 $1 \sim 4$의 숫자가 바닥에 놓일 확률은 동등하게 이므로
확률변수 X의 확률질량함수는 다음과 같다.

$$f(x) = \frac{1}{4}, \quad x = 1, 2, 3, 4$$

$X \sim DU(4)$이므로 평균과 분산은 각각 다음과 같다.

$$\mu = E(X) = \frac{4+1}{2} = \frac{5}{2}, \quad \sigma^2 = \frac{4^2 - 1}{12} = \frac{5}{4}$$

베르누이 분포와 이항분포

베르누이분포^Bernoulli distribution^ :
$P(X = 0) = 1 - p$, $P(X = 1) = p$인 분포이고, 이때 $X \sim B(1, p)$로 나타낸다.

$$f(x) = \begin{cases} 1-p & ,x=0 \\ p & ,x=1 \\ 0 & ,\text{다른 곳에서} \end{cases}$$

베르누이 분포와 이항분포

❖ 1차 적률과 2차 적률 :

$$E(X) = 0 \times f(0) + 1 \times f(1) = p$$

$$E(X^2) = 0^2 \times f(0) + 1^2 \times f(1) = p$$

① 평균 : $\mu = p$

② 분산 : $\sigma^2 = p(1-p)$

분포함수 :

$$\boxed{F(x) = \begin{cases} 0 & , x < 0 \\ p & , 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}}$$

베르누이 분포와 이항분포

❖ 성공률 인 베르누이분포의 평균과 분산은 성공률 p에 의하여 결정되는 것을 알 수 있다. 이와 같이 확률분포를 결정짓는 상수를 확률분포의 **모수**parameter라 한다.

❖ 베르누이분포는 <u>서로 상반되는 두 가지 결과로 구성된 통계실험</u>, 즉

- 동전 던지기의 앞면과 뒷면,
- 전기회로 스위치의 ON과 OFF,
- 설문조사의 YES와 NO,
- 상품의 불량품과 양품

등과 같은 경우의 확률모형에 사용된다.

베르누이 분포와 이항분포

❖ 관심의 대상이 되는 실험결과를 성공 그렇지 않은 결과를 실패라 하고, 확률변수 X를 성공이면 $X = 1$ 그리고 실패이면 $X = 0$으로 정의한다.

베르누이 실험을 독립적으로 반복하여 시행하는 과정을 **베르누이 시행**Bernoulli trial이라 한다.

베르누이 시행은 다음 세 가지 가정에 기초한 반복시행을 나타낸다.

① 매 시행에서 실험결과는 성공(S)과 실패(F) 두 가지뿐이다.
② 매 시행에서 성공률은 동일하다.
③ 각 시행의 실험결과는 독립적이다.

베르누이 분포와 이항분포

예제 6-2

성공률이 $p = \dfrac{1}{4}$ 인 베르누이 실험을 독립적으로 3번 반복하여 시행한다. 매 시행에서 실험 결과가 성공이면 $X = 1$, 실패이면 $X = 0$으로 정의할 때, 다음을 구하라.

(a) 확률변수 X의 확률질량함수와 평균 및 분산

(b) 반복 시행한 3번 중에서 꼭 한 번만 성공할 확률

풀이

(a) 확률변수 X는 성공률이 $p = 1/4$인 베르누이분포를 이루므로 확률질량함수는 다음과 같다.

$$f(x) = \begin{cases} 3/4 & , x = 0 \\ 1/4 & , x = 1 \\ 0 & ,\text{다른 곳에서} \end{cases}$$

베르누이 분포와 이항분포

확률변수 X의 평균과 분산은 각각 다음과 같다.

$$\mu = \frac{1}{4}, \quad \sigma^2 = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$$

(b) 3번의 시행에서 꼭 한 번만 성공하는 경우 : (1, 0, 0), (0, 1, 0), (0, 0, 1)
매 시행에서 성공을 S, 실패를 F라 하면, $P(S) = 1/4, P(F) = 3/4$이다. 꼭 한 번만 성공하는 경우를 S와 F로 나타내면, SFF, FSF, FFS이고 각 사행은 독립이므로 다음 확률을 얻는다.

$$P(SFF) = P(S)P(F)P(F) = \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{9}{64}$$

$$P(FSF) = P(F)P(S)P(F) = \frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{9}{64}$$

$$P(FFS) = P(F)P(F)P(S) = \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{9}{64}$$

따라서 구하고자 하는 확률은 $\frac{9}{64} + \frac{9}{64} + \frac{9}{64} = \frac{27}{64}$ 이다.

베르누이 분포와 이항분포

예 매회 성공률이 1/4인 베르누이 실험을 독립적으로 세 번 반복 시행하여, 성공의 횟수를 확률변수 X라 하자.

표본공간 : $S = \{SSS,\ SSF,\ SFS,\ FSS, SFF, FSF,\ FFS,\ FFF\}$

$$\text{성공률}: P(S) = \frac{1}{4}, \quad \text{실패율}: P(F) = \frac{3}{4}$$

각 표본점에 대한 확률변수 X와 확률

표본점	SSS	SSF	SFS	FSS	SFF	FSF	FSS	FFF
성공 횟수	3	2	2	2	1	1	1	0
확률	$\left(\frac{1}{4}\right)^3$	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$	$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$	$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$	$\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2$	$\left(\frac{3}{4}\right)^3$

베르누이 분포와 이항분포

베르누이 실험을 독립적으로 세 번 반복하여 성공한 횟수 X의 확률분포

X	0	1	2	3
$P(X=x)$	$1\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3$	$3\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2$	$3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1$	$1\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0$

조합의 수 : $\binom{3}{0}=1,\ \binom{3}{1}=3,\ \binom{3}{2}=3,\ \binom{3}{3}=1$

X	0	1	2	3
$P(X=x)$	$\binom{3}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3$	$\binom{3}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2$	$\binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1$	$\binom{3}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0$

확률변수 X의 확률함수 : $p(x)=\binom{3}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x},\quad x=0,1,2,3$

베르누이 분포와 이항분포

❖ 확률함수의 구조 :

시행 횟수

$$P(X=1)=\binom{3}{1}\left(\frac{1}{4}\right)^{1}\left(\frac{3}{4}\right)^{2=3-1}\quad,\quad x=0,1,2,3$$

성공 횟수 성공률

❖ 매회 성공률이 p인 베르누이 실험을 독립적으로 n번 반복 시행하여, 성공의 횟수 X의 확률질량함수

$$f(x)=\begin{cases}\binom{n}{x}p^{x}(1-p)^{n-x} & ,x=0,1,2,\cdots,n\\ 0 & ,\text{다른 곳에서}\end{cases}$$

베르누이 분포와 이항분포

이항분포binomial distribution :
매회 성공률이 p인 베르누이 실험을 독립적으로 n번 반복 시행하여, 성공의 횟수 X의 확률분포이며, $X \sim B(n, p)$로 나타낸다.

❖ $X_i \sim B(1, p), i = 1, 2, \ldots, n$이라 할 때, $X = X_1 + X_2 + X_n$이라 하면, X_i가 취하는 값은 0 또는 1이므로 X는 n번 베르누이 시행을 독립적으로 반복할 때 성공의 횟수이다.
즉, 독립인 $X_i \sim B(1, p), i = 1, 2, \ldots, n$에 대하여 $X \sim B(n, p)$이다.

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np$$

$$Var(X) = Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + E(X_2) + \cdots + E(X_n) = np(1-p)$$

① 평균 : $\mu = np$

② 분산 : $\sigma^2 = np(1 - p)$

베르누이 분포와 이항분포

예제 6-3

집적회로 10개에 포함된 불량품의 수를 조사하고자 한다. 각 집적회로가 불량품일 가능성은 독립적으로 0.05이다. 불량품의 수를 확률변수 X라 할 때, 다음을 구하라.

(a) 확률질량함수

(b) 평균과 표준편차

(c) 집적회로 10개 중에 불량품이 2개일 확률

풀이

(a) 확률변수 X는 불량률이 0.05인 집적회로 10개를 독립적으로 조사하여 불량품의 수를 나타낸다. 그러므로 X는 모수 $n = 10, p = 0.05$인 이항분포에 따르고, 따라서 확률질량함수는 다음과 같다.

$$f(x) = \binom{10}{x} 0.05^x 0.95^{10-x}, \quad x = 0, 1, 2, \cdots, 10$$

베르누이 분포와 이항분포

(b) 모수가 $n = 10, p = 0.05$인 이항분포이므로 X의 평균과 분산 그리고 표준편차는 각각 다음과 같다.

$$\mu = 10 \times 0.05 = 0.5, \quad \sigma^2 = 10 \times 0.05 \times 0.95 = 0.475, \quad \sigma = \sqrt{0.475} = 0.6892$$

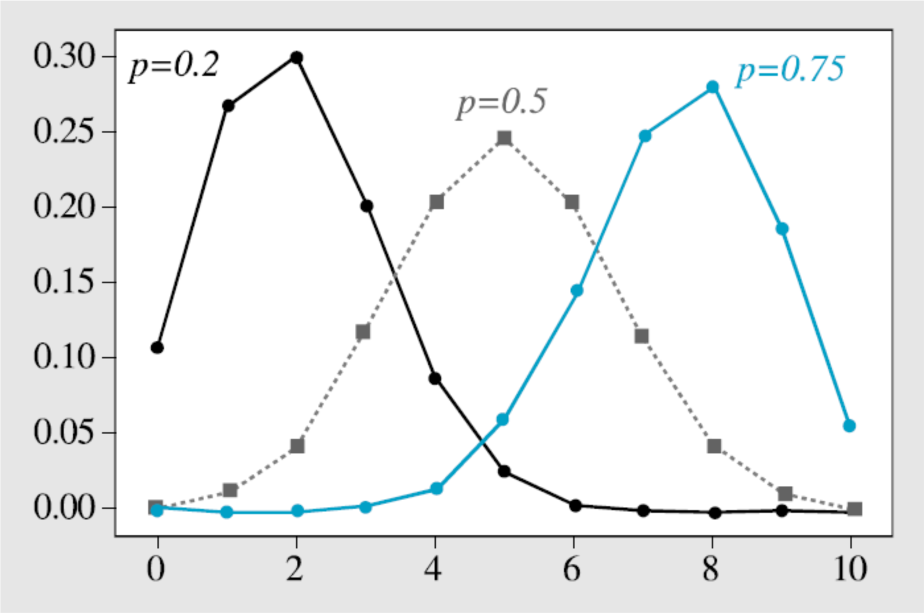
(c) 불량품이 두 개일 확률 : $P(X = 2) = f(2) = \binom{10}{2} 0.05^2 0.95^8 = 0.0746$

이항분포의 모양

① $p < 0.5$이면 이항분포는 왼쪽으로 치우치고 오른쪽으로 긴 꼬리 모양을 가지는 양의 비대칭이며,

② $p > 0.5$이면 오른쪽으로 치우치고 왼쪽으로 긴 꼬리 모양을 가지는 음의 비대칭인 분포를 이룬다.

③ $p = 0.5$ 이면 n에 관계없이 $\mu = n/2$을 중심으로 좌우대칭이고, 이 경우에 **대칭이항분포**symmetric binomial distribution라 한다.



확률 계산

❖ $X \sim B(n, p), a = 0, 1, 2, \ldots, n$일 때, 다음과 같이 확률을 계산한다.

① $P(X = a) = P(X \le a) - P(X \le a - 1)$
② $P(a \le X \le b) = P(X \le b) - P(X \le a - 1)$
③ $P(X \ge a) = 1 - P(X \le a - 1)$

❖ $X \sim B(5, 0.25)$에 대하여 $P(X \le 1)$을 구하는 방법

시행 횟수 성공 횟수 성공률 $P(X \le 1)$

n	x	p = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815

확률 계산

예제 6-4

[예제 6-3]의 집적회로에 대해 이항누적확률표를 이용하여 다음 확률을 구하라.

(a) 적어도 집적회로 1개가 불량일 확률

(b) 많아야 2개가 불량일 확률

(c) 3개 이상 5개 이하가 불량일 확률

풀이

(1) $X \sim B(10, 0.05)$이므로 구하고자 하는 확률은 다음과 같다.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.5987 = 0.4013$$

(2) 많아야 2개가 불량일 확률은 다음과 같다.

$$P(X \leq 2) = 0.9885$$

(3) 구하고자 하는 확률은 다음과 같다.

$$P(3 \leq X \leq 5) = P(X \leq 5) - P(X \leq 2) = 1 - 0.9885 = 0.0115$$

		p		
n	x	…	0.05	…
10	0	…	0.5987	…
	1	…	0.9139	…
	2	…	0.9885	…
	3	…	0.9990	…
	4	…	0.9999	…
	5	…	1.0000	…
	6	…	1.0000	…

독립인 두 이항확률변수의 합

❖ $X \sim B(n,p), Y \sim B(m,p)$이고 X와 Y가 독립이면, $X+Y \sim B(n+m,p)$이다. 전확률 공식에 의해 다음을 얻는다.

$$
\begin{aligned}
P(X+Y=k) &= \sum_{x=0}^{n} P(X=x, x+Y=k) = \sum_{x=0}^{n} P(X=x)P(Y=k-x) \\
&= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{k-x} p^y (1-p)^{m-(k-x)} \\
&= \sum_{x=0}^{n} \binom{n}{x} \binom{m}{k-x} p^k (1-p)^{(n+m)-k} \\
&= \binom{n+m}{k} p^k (1-p)^{(n+m)-k}, \quad k = 0,1,2,\cdots,n+m
\end{aligned}
$$

표본비율

❖ $X \sim B(n,p)$에 대하여 $Y = X/n$은 n번의 베르누이 실험에서 성공한 비율을 나타내며, 이 비율을 **표본비율**sample proportion이라 한다.

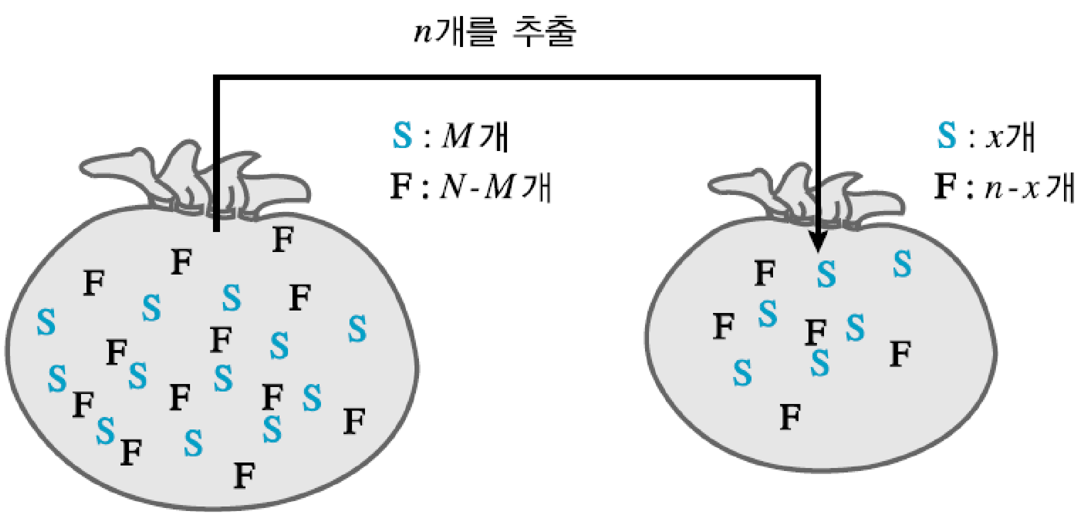
① 표본비율 Y의 평균 : $\mu_Y = p$

② 표본비율 Y의 분산 : $\sigma_Y^2 = \dfrac{pq}{n}, \quad q = 1 - p$

초기하분포

1. 모집단은 N개의 $item$으로 구성된다.
2. 각 item은 성공(S)과 실패(F)로 구성되며, 모집단 안에 성공의 개수는 M이고 실패의 개수는 $N - M$이다.
3. 모집단으로부터 n개의 $item$을 선정한다.

표본으로 선정된 n개의 $item$ 안에 포함된 성공의 수 X에 관한 확률분포를 모수 N, M, n인 **초기하분포**hypergeometric distribution라 하고, $X \sim H(N, M, n)$으로 나타낸다.



초기하분포

❖ N개 중에서 n개의 $item$을 꺼내는 방법의 수 :

N개 안에 포함된 성공의 개수 M개 중에서 x개를 꺼내는 방법의 수 : $\binom{M}{x}$

성공의 수가 x인 각각에 대해 실패의 개수 $N - M$개 중에서

$n - x$개를 꺼내는 방법의 수 : $\binom{N-M}{n-x}$

초기하분포

❖ N개 중에서 n개의 $item$을 꺼낼 때, 성공인 $item$ 이 x개일 확률 :

$$P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

이때 n개 안에 포함된 성공인 $item$의 수는 다음과 같다.

$$\max(0, n + M - N) \le x \le \min(n, M)$$

보편적으로 $x = 0, 1, 2, \dots, n$을 많이 사용한다.

초기하분포의 평균과 분산

$$E(X)=\sum_{x=0}^{n} xf(x)=\sum_{x=0}^{n} x\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$=\sum_{x=0}^{n}\frac{\dfrac{x\cdot M!}{x!(M-x)!}\cdot\dfrac{(N-M)!}{(n-x)![(N-M)-(n-x)]!}}{\dfrac{N!}{n!(N-n)!}}$$

$$=\sum_{x=0}^{n}\frac{\dfrac{M(M-1)!}{(x-1)![(M-1)-(x-1)]!}\cdot\dfrac{[(N-1)-(M-1)]!}{[(n-1)-(x-1)]![(N-M)-(n-x)]!}}{\dfrac{N(N-1)!}{n(n-1)![(N-1)-(n-1)]!}}$$

$$=n\frac{M}{N}\sum_{x=0}^{n}\frac{\dfrac{(M-1)!}{(x-1)![(M-1-(x-1)]!}\cdot\dfrac{[(N-1)-(M-1)]!}{[(n-1)-(x-1)]![(N-M)-(n-x)]!}}{\dfrac{(N-1)!}{(n-1)![(N-1)-(n-1)]!}}$$

$t = x - 1$이라 하면, 다음을 얻는다.

초기하분포의 평균과 분산

$$E(X) = n\frac{M}{N}\sum_{t=0}^{n-1}\frac{\dfrac{(M-1)!}{t![(M-1)-t]!}\cdot\dfrac{[(N-1)-(M-1)]!}{[(n-1)-t]![(N-M)-((n-1)-t)]!}}{\dfrac{(N-1)!}{(n-1)![(N-1)-(n-1)]!}}$$

$$= n\frac{M}{N}\sum_{t=0}^{n-1}\boxed{\frac{\binom{M-1}{t}\binom{(N-1)-(M-1)}{(n-1)-t}}{\binom{N-1}{n-1}}}$$

$X \sim H(N-1, M-1, n-1)$의 확률질량함수

$$\boxed{\sum_{t=0}^{n-1}\frac{\binom{M-1}{t}\binom{(N-1)-(M-1)}{(n-1)-t}}{\binom{N-1}{n-1}} = 1}$$

$$E(X) = n\frac{M}{N}$$

같은 방법으로 $E[X(X-1)]$의 기댓값은 다음과 같다.

$$E[X(X-1)] = n(n-1)\frac{M(M-1)}{N(N-1)}$$

초기하분포의 평균과 분산

$$E(X^2) = E[X(X-1)] + E(X) = n(n-1)\frac{M(M-1)}{N(N-1)} + n\frac{M}{N}$$

$$= \frac{nMN - n^2M - nM^2 + (nM)^2}{N(N-1)}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{nMN - n^2M - nM^2 + (nM)^2}{N(N-1)} - \left(n\frac{M}{N}\right)^2$$

$$= n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$$

① 평균 : $\mu = n\frac{M}{N}$

② 분산 : $\sigma^2 = n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}$

초기하분포의 평균과 분산

예제 6-5

집적회로 20개가 들어 있는 상자 안에 불량품이 5개 포함되어 있다. 이 상자에서 집적회로 4개를 선정할 때, 다음을 구하라.

(a) 선정된 집적회로 안에 포함된 불량품 수에 대한 확률질량함수

(b) 선정된 집적회로 안에 포함된 불량품 수의 평균과 표준편차

(c) 불량품 2개가 포함될 확률

풀이

(a) 선정된 집적회로 안에 포함된 불량품 수를 X라 하면, $X \sim H(20, 5, 4)$이므로 확률질량함수는 다음과 같다.

$$f(x) = \frac{\binom{5}{x}\binom{15}{4-x}}{\binom{20}{4}} = \frac{1}{4845}\binom{5}{x}\binom{15}{4-x}, \quad x = 0, 1, 2, 3, 4$$

초기하분포의 평균과 분산

(b) $N = 20, M = 5, n = 4$이므로 평균과 분산은 각각 다음과 같다.

$$\mu = 4 \times \frac{5}{20} = 1, \quad \sigma^2 = 4 \times \frac{5}{20} \times \left(1 - \frac{5}{20}\right) \times \frac{20-4}{20-1} = \frac{12}{19} = 0.6316$$

따라서 표준편차는 $\sigma = \sqrt{0.6316} = 0.7947$ 이다.

(c) 구하고자 하는 확률은 다음과 같다.

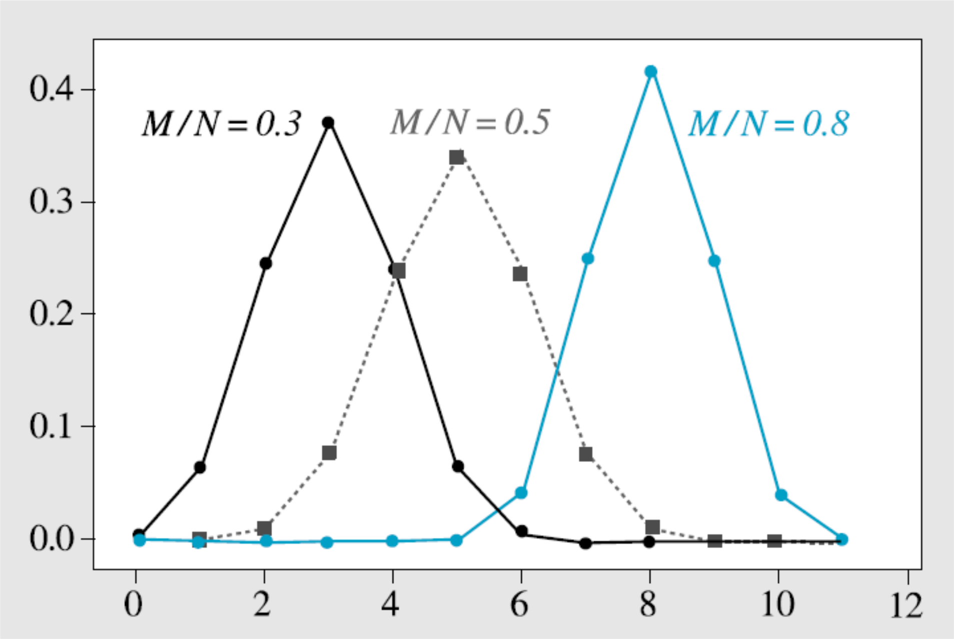
$$P(X = 2) = f(2) = \frac{\binom{5}{2}\binom{15}{2}}{\binom{20}{4}} = \frac{10 \times 105}{4845} = \frac{70}{323}$$

초기하분포의 모양

① $M/N < 0.5$이면 초기하분포는 왼쪽으로 치우치고 오른쪽으로 긴 꼬리 모양을 가지는 양의 비대칭이며,

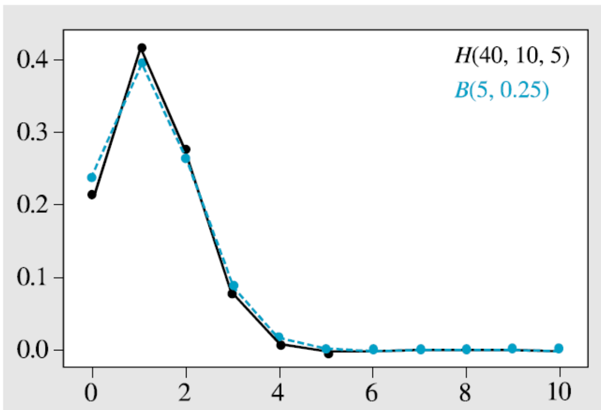
② $M/N > 0.5$ 이면 오른쪽으로 치우치고 왼쪽으로 긴 꼬리 모양을 가지는 음의 비대칭인 분포를 이룬다.

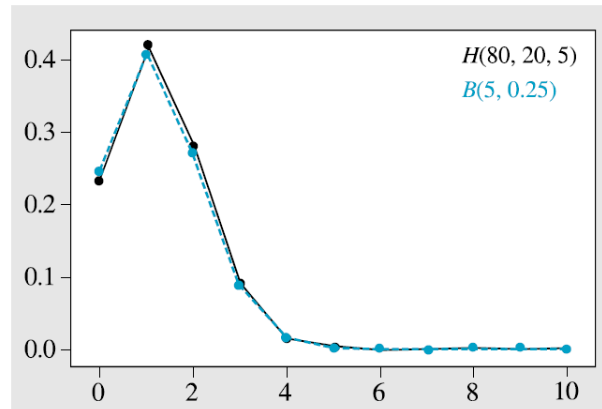
③ $N/M = 0.5$ 이면 좌우대칭이다.

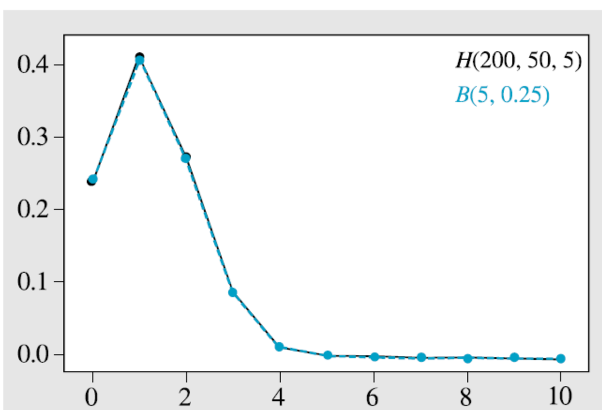


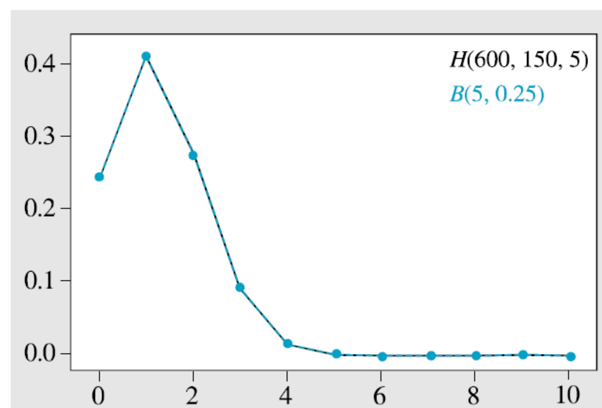
초기하분포와 이항분포

❖ $M/N = p$로서 일정하고, $N \to \infty$이면 초기하분포는 평균 $\mu = np$, 분산 $\sigma^2 = np(1-p)$인 이항분포에 근사한다.









초기하분포

예제 6-6

집적회로 100개가 들어 있는 상자에서 집적회로 4개를 선정한다. 상자 안에는 불량품 5개가 포함되어 있다.

(a) 선정된 불량품의 수에 대한 확률질량함수를 구하라.

(b) 불량품 2개가 포함될 확률을 구하라.

(c) 이항분포를 이용하여 불량품 2개가 포함될 근사확률을 구하라.

풀이

(a) 선정된 집적회로 안에 포함된 불량품 수를 X라 하면, $X \sim H(100,5,4)$이므로 확률질량함수는 다음과 같다.

$$f(x) = \frac{\binom{5}{x}\binom{95}{4-x}}{\binom{100}{4}}, \quad x = 0,1,2,3,4$$

초기하분포

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X=2)=f(2)=\frac{\binom{5}{2}\binom{95}{2}}{\binom{100}{4}}=\frac{5!/(2!3!)\times 95!/(2!93!)}{100!/(4!96!)}=\frac{1786}{156849}=0.01139$$

(c) $N = 100, M = 5$이므로 X는 이항분포 $B(4, 0.6)$에 근사한다. 그러므로 구하고자 하는 확률은 $P(X = 2) = P(X \leq 2) - P(X \leq 1) = 0.9995 - 0.9860 = 0.0135$이다.

기하분포

예 **매회 성공률이 p인 베르누이 실험을 독립적으로 세 번 반복 시행하여, 처음 성공할 때까지 반복적으로 시행한 횟수를 확률변수 X라 하자.**

표본공간 : $S = \{ S, FS, FFS, FFFS, FFFFS, FFFFFS, \ldots \}$

성공률 : $P(S) = p$, 실패율 : $P(F) = 1 - p$

상태공간 : $S_X = \{ 1, 2, 3, 4, 5, 6, \ldots \}$

$$P(X=1) = P(S) = p$$
$$P(X=2) = P(F \cap S) = P(F)P(S) = (1-p)p$$
$$P(X=3) = P(F \cap F \cap S) = P(F)P(F)P(S) = (1-p)^2 p$$
$$P(X=4) = P(F \cap F \cap F \cap S) = P(F)P(F)P(F)P(S) = (1-p)^3 p$$
$$\vdots$$
$$P(X=x) = P(\underbrace{\boxed{F \cap \cdots \cap F}}_{x-1\text{번}} \cap S) = \underbrace{\boxed{P(F)\cdots P(F)}}_{x-1\text{번}} P(S) = (1-p)^{x-1} p$$

기하분포

기하분포geometric distribution :
매 시행에서 성공률이 p인 베르누이 실험을 처음 성공할 때까지 독립적으로 반복 시행한 횟수 X의 확률분포. 이때 $X \sim G(p)$로 나타낸다.

$$f(x) = \begin{cases} pq^{x-1} & ,x = 1,2,3,\cdots,\ q = 1-p \\ 0 & ,\text{다른 곳에서} \end{cases}$$

기하분포의 평균과 분산

$$E(X)=\sum_{x=0}^{\infty} xf(x)=\sum_{x=0}^{\infty} xpq^{x-1}=p\sum_{x=0}^{\infty} xq^{x-1} \qquad \left(xq^{x-1}=\frac{d}{dq}q^x \text{라하면} \right)$$

$$=p\sum_{x=1}^{n}\frac{d}{dq}q^x=p\frac{d}{dq}\left(\sum_{x=1}^{n}q^x\right)=p\frac{d}{dq}\left(\frac{q}{1-q}\right)$$

$$=\frac{p}{(1-q)^2}=\frac{1}{p}$$

같은 방법으로, $E[X(X-1)] = 2q/p^2$이고, 따라서 분산은 다음과 같다.

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

$$=\frac{2q}{p^2}+\frac{1}{p}-\left(\frac{1}{p}\right)^2=\frac{q}{p^2}$$

① 평균 : $\mu=\dfrac{1}{p}$

② 분산 : $\sigma^2=\dfrac{q}{p^2}$

기하분포의 평균과 분산

예제 6-7

해안지역에 건설되는 건물은 내구 설계에 대한 건축법규를 따른다. 이 건축법규에 의하면 설계 풍속을 평균 25년 주기로 찾아오는 강풍을 명시하도록 되어 있다.

(a) 건물을 새로 건설한 첫해 1년 동안 이 풍속이 처음 찾아올 확률을 구하라.

(b) 건물을 새로 건설한 이후로 x 번째 해에 이 풍속이 처음 찾아올 확률을 구하라.

(c) 건물을 새로 건설한 이후로 3년 이내에 이 풍속이 찾아올 확률을 구하라.

풀이

(a) 이 풍속이 평균 25년 주기로 찾아오므로 건물을 새로 건설한 첫해 1년 동안 이 풍속이 처음 찾아올 확률은 다음과 같다.

$$p = \frac{1}{\mu} = \frac{1}{25} = 0.04$$

(b) x번째 해에 이 풍속이 처음 찾아올 확률은 다음과 같다.

$$P(X = x) = 0.04 \times 0.96^{x-1}, \quad x = 1, 2, 3, \cdots$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X \le 3) = 0.04 \times (1 + 0.96 + 0.96^2) = 0.04 \times 2.8816 = 0.1153$$

기하분포의 비기억성 성질

❖ $X \sim G(p)$이면 양의 정수 m,n에 대하여 다음이 성립한다.

$$P(X > n+m \mid X > n) = P(X > m)$$

$$P(X > n+m \mid X > n) = \frac{P(X > n, X > n+m)}{P(X > n)} = \frac{P(X > n+m)}{P(X > n)} \quad \text{이고}$$

$$P(X > m) = P(X \geq m+1) = \sum_{x=m+1}^{\infty} pq^{x-1} = p \cdot \frac{q^m}{1-q} = q^m \quad \text{이므로}$$

$$P(X > n) = q^n, \quad P(X > n+m) = q^{n+m}$$

$$P(X > n+m \mid X > n) = \frac{P(X > n+m)}{P(X > n)} = \frac{q^{n+m}}{q^n} = q^m = P(X > m)$$

기하분포의 비기억성 성질에 대한 의미

① 처음 성공할 때까지 반복 시행한 횟수는 그 이후로 다시 처음 성공할 때까지 반복 시행한 횟수에 독립이고 항등분포이다.

② 기하분포는 적당한 시간 단위로 구분되고 베르누이 시행으로 모형화 되는 확률 문제에 많이 적용된다.

③ 매 단위 시간마다 특정한 사건의 발생 유무는 독립이고, 연속적으로 동일한 두 사건이 발생할 때까지 반복 시행한 횟수는 동일한 분포를 이루는 것을 나타낸다

기하분포

예제 6-8

해양 구조물은 평균 해면으로부터 파고가 10m를 넘는 파도에 견디도록 고안되어 있다. 10m가 넘는 파도가 찾아올 확률이 연간 4%일 때, 다음을 구하라.

(a) 10m를 초과하는 파도가 찾아올 평균 년 수

(b) 10m를 초과하는 파도가 처음으로 찾아올 년 수에 대한 확률질량함수

(c) 이 구조물이 세워진 이후로 2년 이내에 10m를 초과하는 파도가 찾아올 확률

(d) 이 구조물이 세워진 이후로 2년 이후에 첫 번째로 이 파도가 찾아온다고 할 때, 정확히 5차 년도에 이 파도가 찾아올 확률

(e) (d)의 가정 아래서 2차 년도 이후로 5년을 넘겨야 이 파도가 찾아올 확률

풀이

(a) 파고 10m가 넘는 파도가 처음으로 찾아올 확률이 매년 $p = 0.04$이므로 이 파도가 찾아올 평균 년 수는 $m = 1/0.04 = 25$(년)이다.

기하분포

(b) 구조물이 세워진 이후로 10m를 초과하는 파도가 처음으로 찾아올 년 수를 X라 하면, 구하고자 하는 확률질량함수는 다음과 같다.

$$f(x) = 0.04 \times 0.96^{x-1}, \quad x = 1, 2, 3, \cdots$$

(c) 구하고자 하는 확률은 $P(X \le 2) = f(2) = (0.04)(0.96) = 0.0384$이다.

(d) 구하고자 하는 확률은 다음과 같다.

$$P(X = 5 | X > 2) = \frac{P([X = 5] \cap [X > 2])}{P(X > 2)} = \frac{P(X = 5)}{1 - P(X \le 2)}$$

$$= \frac{0.04 \times 0.96^4}{1 - 0.0384} = \frac{0.03397}{0.9616} = 0.0353$$

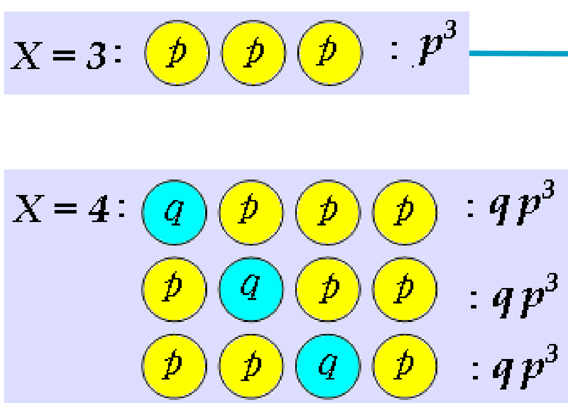
(e) 구하고자 하는 확률은 다음과 같다.

$$P(X > 5 + 2 | X > 2) = P(X = 5) = 0.04 \times 0.96^4 = 0.03397$$

음이항분포

음이항분포negative binomial distribution :
매 시행에서 성공률이 p인 베르누이 실험을 처음 r번째 성공할 때까지 독립적으로 반복시행한 횟수 X의 확률분포.
$X \sim NB(r, p)$로 나타낸다.

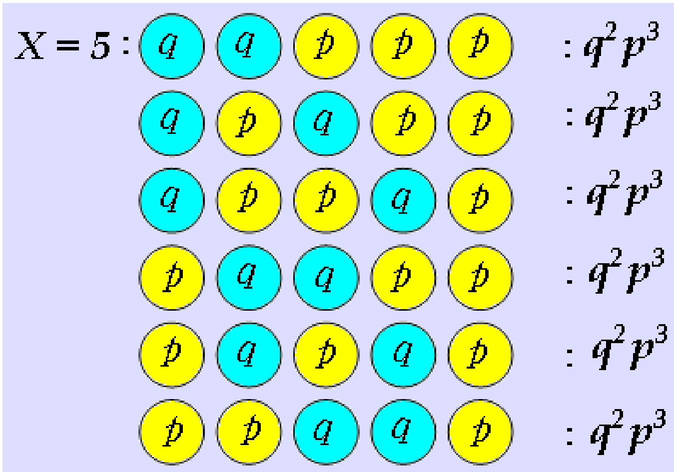
예 **매 회 성공률 p인 베르누이 실험을 세 번째 성공할 때까지 독립적으로 반복 시행한 횟수를 확률변수 X라 하자.**



경우의 수 : 1가지
각 경우의 확률 : p^3
$P(X=3) = p^3$

경우의 수 : 3가지
각 경우의 확률 : qp^3
$P(X=4) = 3qp^3 = \binom{3}{2} qp^3$

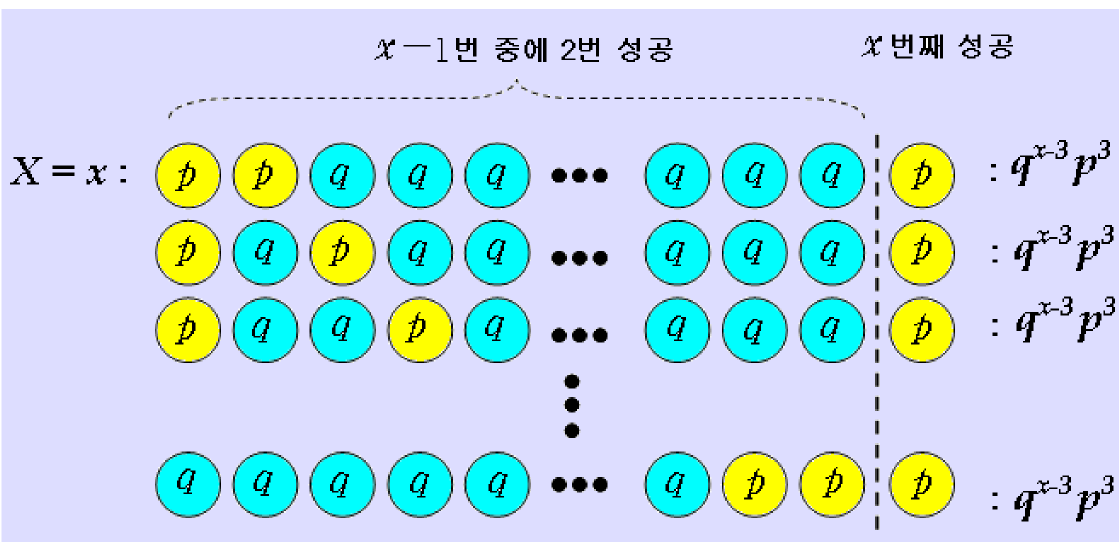
음이항분포



경우의 수 : 6가지

각 경우의 확률 : q^2p^3

$P(X=4)=6qp^3=\binom{5}{2}q^2p^3$



경우의 수 : $\binom{x-1}{2}$가지

각 경우의 확률 : $q^{x-3}p^3$

$P(X=x)=\binom{x-1}{2}q^{x-3}p^3$

음이항분포

❖ 매 시행에서 성공률이 p인 베르누이 실험을 처음 r번째 성공할 때까지 독립적으로 반복시행한 횟수 X의 확률분포:

확률변수 X가 취할 수 있는 가장 작은 값은 처음부터 r번 연속하여 성공하는 경우이므로 X의 상태공간은 $S_X = \{r, r+1, r+2, r+3, \ldots\}$이다.

x번째 시행에서 r번째 성공이 이루어졌다면,
$x-1$번의 시행에서 꼭 $r-1$번 성공하고 x번째 시행에서 마지막 r번째 성공이 이루어져야 한다.
이때 $x-1$번의 시행에서 성공의 횟수는 이항분포 $B(x-1, p)$에 따르고, 따라서 꼭 $r-1$번 성공할 확률은 다음과 같다.

$$P(x-1\text{번 시행에서 } r-1\text{번 성공}) = \binom{x-1}{r-1} p^{r-1} q^{(x-1)-(r-1)} = \binom{x-1}{r-1} p^{r-1} q^{x-r}$$

음이항분포

매 시행이 독립이므로 x번째 시행에서 r번째 성공이 이루어질 확률은 다음과 같다.

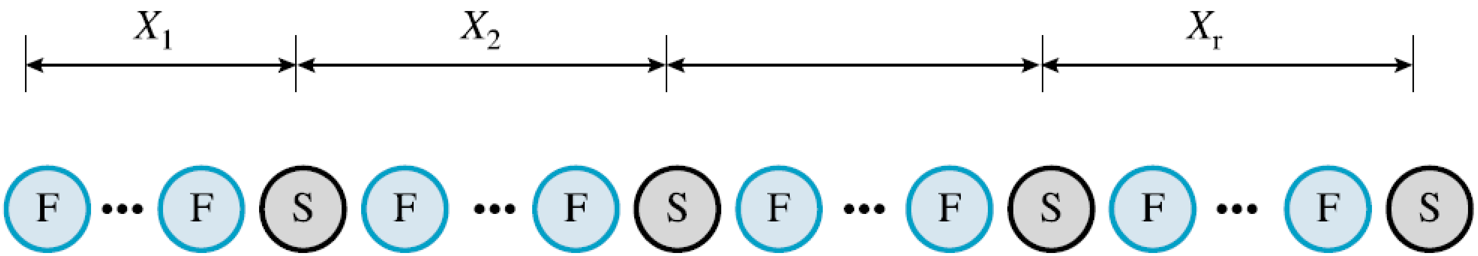
$$\begin{aligned}P(X=x) &= P(x-1\text{번 시행에서 } r-1\text{번 성공}) \times P(x\text{번째 시행에서 } r\text{번째 성공}) \\ &= \binom{x-1}{r-1} p^{r-1} q^{x-r} \times p \\ &= \binom{x-1}{r-1} p^{r} q^{x-r}\end{aligned}$$

$X \sim NB(r,p)$의 확률질량함수

$$f(x) = \begin{cases} \binom{x-1}{r-1} p^r q^{x-r} & ,x = r, r+1, r+2, \cdots, \quad q = 1-p \\ 0 & ,\text{다른 곳에서} \end{cases}$$

음이항분포와 기하분포

❖ $r = 1$인 경우에 음이항분포는 모수 p인 기하분포이다.
$X_i \sim G(p)$라 하면 $X = X_1 + X_2 + \ldots + X_r$은 음이항분포에 따른다.



음이항분포의 평균과 분산

❖ $X_i \sim G(p)$이면 $E(X_i)=\dfrac{1}{p}$, $Var(X_i)=\dfrac{q}{p^2}$ 이므로

$$E(X)=E(X_1)+E(X_2)+\cdots+E(X_r)=\frac{r}{p}$$

$$Var(X)=Var(X_1)+Var(X_2)+\cdots+Var(X_r)=\frac{rq}{p^2}$$

① 평균 : $\mu=\dfrac{r}{p}$

② 분산 : $\sigma^2=\dfrac{rq}{p^2}$

음이항분포

예제 6-9

[예제 6−8]에서 파고 10m가 넘는 파도가 두 번 찾아오는 경우에 대해 다음을 구하라.

(a) 10m를 넘는 파도가 두 번 찾아올 평균 년 수

(b) 10m를 초과하는 파도가 두 번 찾아올 년 수에 대한 확률질량함수

(c) 이 구조물이 세워진 이후로 5년째 되는 해에 두 번째 파도가 찾아올 확률

풀이

(a) 파고 10m가 넘는 파도가 처음으로 찾아올 확률이 매년 $p = 0.04$이므로 이 파도가 두 번 찾아올 평균 년 수는 $m = 2/0.04 = 50$(년)이다.

(b) 구조물이 세워진 이후로 10m를 초과하는 파도가 두 번째로 찾아올 년 수를 X라 하면, 구하고자 하는 확률질량함수는 다음과 같다.

$$f(x) = \binom{x-1}{1}(0.04)^2(0.96)^{x-2}, \quad x = 2,3,4,\cdots$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X=5) = f(5) = 4 \times 0.04^2 \times 0.96^3 = 0.0057$$

푸아송분포

❖ 푸아송분포는 다음과 같이 주어진 영역 안에서 특별한 사건이 발생한 횟수에 관한 확률분포이다.

- 1초 동안 방출된 방사능 입자의 수
- 1시간 동안 걸려온 스팸문자 횟수
- 특정 시간 동안 톨게이트를 통과한 자동차 수
- 음료수 1mm^3 당 박테리아 수
- 소설책의 한 면 당 오자의 개수

푸아송분포의 구조

❖ $X \sim B(n,p)$에 대해 $m = np$로 일정하고 $n \to \infty$라고 하자. 이항확률변수 X의 확률질량함수 $f(x)$는 다음 극한을 갖는다.

$$
\begin{aligned}
f(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!}\left(\frac{m}{n}\right)^x \left(1-\frac{m}{n}\right)^{n-x} \\
&= \frac{m^x}{x!}\frac{n(n-1)\cdots(n-x+1)}{n^{n-x}} \cdot \left[\left(1-\frac{m}{n}\right)^{-\frac{n}{m}}\right]^{-m} \cdot \left(1-\frac{m}{n}\right)^{-x} \\
&= \frac{m^x}{x!} 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-x+1}{n} \cdot \left[\left(1-\frac{m}{n}\right)^{-\frac{n}{m}}\right]^{-m} \cdot \left(1-\frac{m}{n}\right)^{-x} \\
&= \frac{m^x}{x!} e^{-m}, \quad x = 0,1,2,\cdots
\end{aligned}
$$

$*\quad \dfrac{n-1}{n} \cdot \dfrac{n-2}{n} \cdots \dfrac{n-x+1}{n} \to 1, \quad \left(1-\dfrac{m}{n}\right)^{-\frac{n}{m}} \to e, \quad \left(1-\dfrac{m}{n}\right)^{-x} \to 1$

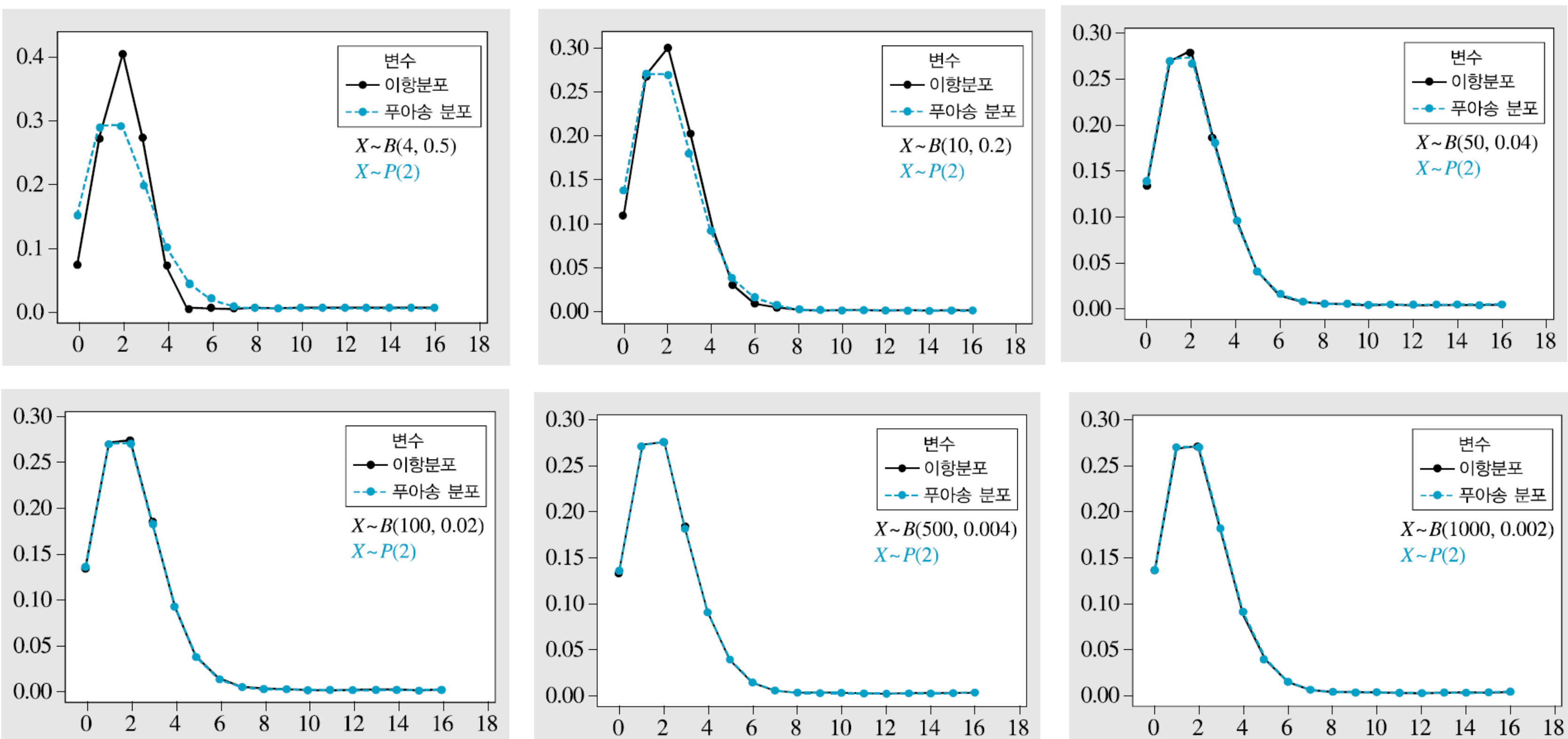
푸아송분포의 구조

포아송분포Poisson distribution **:**
양의 실수 m에 대하여 확률질량함수가 다음과 같은 확률변수 X의 확률분포. 이때 $X \sim P(m)$으로 나타낸다.

$$f(x) = \begin{cases} \dfrac{m^x}{x!} e^{-m} & , x = 0, 1, 2, \cdots \\ 0 & ,\text{다른 곳에서} \end{cases}$$

푸아송분포의 구조

❖ 모수 m인 푸아송분포는 $m = np$로 일정한 이항분포 $X \sim B(n,p)$의 극한분포이다.



푸아송분포의 구조

예제 6-10

$X \sim B(1000,\ 0.002)$일 때, 푸아송 분포를 이용하여 $P(X=4)$를 근사적으로 구하라.

풀이

X는 모수 $n = 1000, p = 0.002$인 이항분포에 따르므로 X는 근사적으로 모수 2인 푸아송분포에 근사한다. 따라서 X의 확률질량함수는 근사적으로 다음과 같다.

$$f(x) = \frac{2^x}{x!} e^{-2}, \quad x = 0, 1, 2, \cdots$$

따라서 구하고자 하는 확률은 다음과 같다.

$$P(X=4) = f(4) = \frac{2^4}{4!} e^{-2} = \frac{2}{3} e^{-2} = 0.0902$$

푸아송분포의 구조

평균 :$E(X)=\sum_{x=0}^{\infty} xf(x)=\sum_{x=0}^{\infty} x\cdot\frac{m^x}{x!}e^{-m}=\sum_{x=1}^{\infty}\frac{m\cdot m^{x-1}}{(x-1)!}e^{-m}$

$=m\sum_{t=0}^{\infty}\boxed{\frac{m^t}{t!}e^{-m}}=m$

$P(m)$인 확률질량함수

같은 방법으로

2차 기댓값 : $E[X(X-1)]=m^2$

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

분산 :$Var(X)=E(X^2)-E(X)^2=(m^2+m)-m^2=m$

① 평균 : $\mu=m$

② 분산 : $\sigma^2=m$

푸아송분포의 구조

예제 6-11

1000분의 1초 동안 평균적으로 방사능 입자 3개가 계수관을 통과한다고 할 때, 다음을 구하라.

(a) 계수관을 통과하는 방사능 입자 수에 대한 확률질량함수

(b) 단위시간 동안 계수관을 통과한 입자가 적어도 하나 이상일 확률

(c) 이 시간 동안 정확히 한 입자가 통과할 확률

풀이

(a) 계수관을 통과하는 방사능 입자 수를 X라 하면, 평균 3인 푸아송분포에 따른다. 따라서 구하고자 하는 확률질량함수는 다음과 같다.

$$f(x) = \frac{3^x}{x!} e^{-3}, \quad x = 0, 1, 2, \cdots$$

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - f(0)$$
$$= 1 - \frac{3^0}{0!} e^{-3} = 1 - e^{-3} = 0.9502$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X = 1) = f(1) = \frac{3^1}{1!} e^{-3} = 3e^{-3} = 0.1494$$

확률 계산

❖ $X \sim P(\mu)$, $a = 0, 1, 2, \ldots$일 때, 다음과 같이 확률을 계산한다.

① $P(X = a) = P(X \le a) - P(X \le a - 1)$

② $P(a \le X \le b) = P(X \le b) - P(X \le a - 1)$

③ $P(X \ge a) = 1 - P(X \le a - 1)$

❖ $X \sim P(1.4)$에 대하여 $P(X \le 4)$를 구하는 방법

발생 횟수 $P(X \le 4)$ 평균

	μ								
x	1.10	1.20	0.30	1.40	1.50	1.60	1.70	1.80	1.90
0	0.333	0.301	0.273	0.247	0.223	0.202	0.183	0.165	0.150
1	0.699	0.663	0.627	0.592	0.553	0.525	0.493	0.463	0.434
2	0.900	0.879	0.857	0.833	0.809	0.783	0.757	0.731	0.704
3	0.974	0.966	0.957	0.946	0.934	0.921	0.907	0.891	0.875
4	0.995	0.992	0.989	0.986	0.981	0.976	0.970	0.964	0.954
5	0.999	0.998	0.998	0.997	0.996	0.994	0.992	0.990	0.987
6	1.000	1.000	1.000	0.999	0.999	0.999	0.998	0.997	0.997

푸아송분포

예제 6-12

〈부록 A.2〉의 푸아송누적확률표를 이용하여 $X \sim P(2.5)$에 대해 다음 확률을 구하라.

(a) $P(X \leq 3)$ (b) $P(X = 4)$ (c) $P(X \geq 4)$

풀이

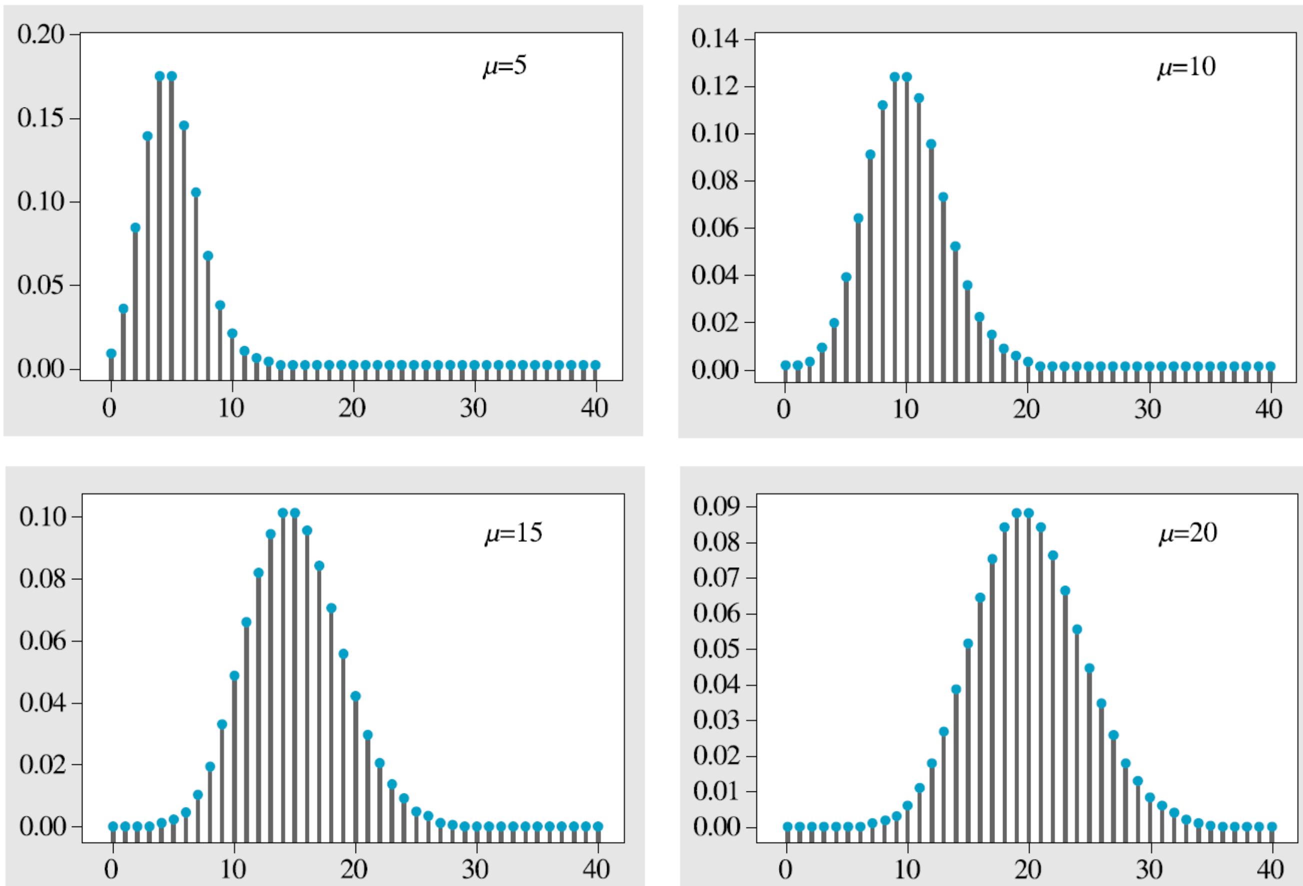
(a) 푸아송 누적확률표에 의해 $P(X \leq 3) = 0.758$이다.

(b) 구하고자 하는 확률은 $P(X = 4) = P(X \leq 4) - P(X \leq 3) = 0.891 - 0.758 = 0.133$이다.

(c) 구하고자 하는 확률은 $P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.758 = 0.242$이다.

푸아송분포

❖ 평균μ인 푸아송분포는μ가 커질수록 평균을 중심으로 갖는 종모양에 근접한다.



독립인 두 푸아송확률변수의 합

❖ $X \sim P(\mu),\ Y \sim P(\lambda)$이고 X와 Y가 독립이면, $X \sim P(\mu),\ Y \sim P(\lambda)$이다.
전확률 공식에 의해 다음을 얻는다.

$$P(X+Y=v) = \sum_{y=0}^{v} P(X+y=v, Y=y) = \sum_{y=0}^{v} P(X=v-y)P(Y=y)$$

$$= \sum_{y=0}^{v} \frac{\mu^{v-y}}{(v-y)!} e^{-\mu} \cdot \frac{\lambda^{y}}{y!} e^{-\lambda} = \sum_{y=0}^{v} \frac{\mu^{v-y}\lambda^{y}}{y!(v-y)!} e^{-(\mu+\lambda)}$$

$$= \sum_{y=0}^{v} \frac{v!}{y!(v-y)!} \frac{\mu^{v-y}\lambda^{y}}{(\mu+\lambda)^{v}} \frac{(\mu+\lambda)^{v}}{v!} e^{-(\mu+\lambda)}$$

$$= \frac{(\mu+\lambda)^{v}}{v!} e^{-(\mu+\lambda)} \sum_{y=0}^{v} \boxed{\binom{v}{y}\left(\frac{\lambda}{\mu+\lambda}\right)^{y}\left(1-\frac{\lambda}{\mu+\lambda}\right)^{v-y}}$$

이항분포

$$= \frac{(\mu+\lambda)^{v}}{v!} e^{-(\mu+\lambda)}, \quad v = 0, 1, 2, \cdots$$

다항분포

다항분포multivariate distribution :
매회 실험결과가 k개의 서로 배반인 사건 $A_1, A_2, \dots, A_k$로 구성되고, 각각의 사건이 발생할 가능성이 $p_i = P(Ai), i = 1,2, \dots, k$인 통계실험을 n번 독립적으로 반복 실행할 때, 각 사건이 나타난 횟수인 $X_1, X_2, \dots, X_k$에 대한 확률분포.
이때 $(X_1, X_2, \dots, X_k) \sim Mult(n, p_1, \dots, p_k)$으로 나타낸다.

$$f(x_1, x_2, \cdots, x_k) = \frac{n!}{(x_1)!(x_2)!\cdots(x_k)!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k} \qquad \begin{array}{l} 0 \le x_i \le n, i = 1, 2, \cdots k \\ x_1 + x_2 + \cdots + x_k = n, \\ p_1 + p_2 + \cdots + p_k = 1 \end{array}$$

다항분포의 성질

❖ X_i 는 이항분포 $B(n, pi)$에 따른다.

X_i와 X_j 의 공분산은 다음과 같다.

$$Cov(X_i, X_j) = -np_1p_2$$

X_i와 X_j 의 상관계수는 다음과 같다.

$$\rho = -\left[\frac{p_ip_j}{(1-p_i)(1-p_j)}\right]^{\frac{1}{2}}$$

예제 6-13

어느 회사는 판매한 제품에 대한 민원이 청구되면 민원의 내용에 따라 A4 용지 1쪽, 2쪽, 3쪽으로 작성된 답변을 민원인에게 제공한다. 예년의 답변 결과에 따르면 1쪽, 2쪽, 3쪽으로 답변한 비율이 각각 $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{4}$이다. 10건의 민원 중에서 답변이 1쪽, 2쪽, 3쪽인 횟수를 각각 X, Y, Z라 할 때, 다음을 구하라.

(a) X, Y, Z의 결합확률질량함수

(b) $P(X=6,\ Y=3,\ Z=1)$

(c) 확률변수 Y의 평균 μ_Y와 분산 σ_Y^2

(d) X와 Y의 공분산과 상관계수

다항분포

풀이

(a) X,Y,Z의 결합확률질량함수는 다음과 같다.

$$f(x,y,z)=\frac{10!}{x!y!z!}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{4}\right)^{y}\left(\frac{1}{4}\right)^{z}=\frac{10!}{x!y!z!}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{4}\right)^{y+z}$$

여기서 $x+y+z=10, x,y,z=0,1,2,\ldots,10$이다.

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X=6,Y=3,Z=1)=f(6,3,1)=\frac{10!}{6!3!1!}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{4}\right)^{3+1}=\frac{840}{2^{14}}=\frac{105}{2048}$$

풀이

(c) $Y \sim B(10, ¼)$이므로 구하고자하는 평균과 분산은 각각 다음과 같다.

$$\mu_Y = 10 \times \frac{1}{4} = 2.5, \quad \sigma_Y^2 = 10 \times \frac{1}{4} \times \frac{3}{4} = \frac{15}{8}$$

(d) $X \sim B(10, ½)$이고 $Y \sim B(10, ¼)$이므로 X와 Y의 공분산과 상관계수는 각각 다음과 같다.

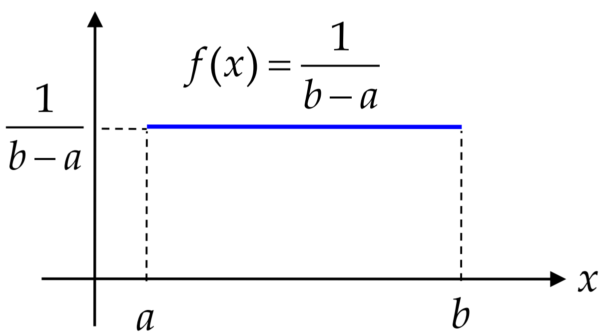
$$Cov(X,Y) = -10 \times \frac{1}{2} \times \frac{1}{4} = -\frac{5}{4}, \quad \rho = -\sqrt{\frac{\frac{1}{2} \times \frac{1}{4}}{\frac{1}{2} \times \frac{3}{4}}} = -\sqrt{\frac{1}{3}} = -0.577$$

6.2 연속확률분포

균등분포

균등분포uniform distribution :
확률변수 X의 상태공간 $S_X = [a,b]$에 대하여 다음 확률밀도함수를 갖는 확률분포. 이때 $X \sim U(a, b)$로 나타낸다.

$$f(x)=\begin{cases} \dfrac{1}{b-a} & ,a \le x \le b \\ 0 & ,\text{다른 곳에서} \end{cases}$$



균등분포

평균 : $$E(X)=\int_a^b xf(x)dx=\frac{1}{b-a}\int_a^b xdx=\frac{1}{b-a}\left(\frac{1}{2}x^2\right)\Bigg|_a^b$$

$$=\frac{1}{b-a}\cdot\frac{b^2-a^2}{2}=\frac{a+b}{2}$$

2차 기댓값 : $$E(X^2)=\int_a^b x^2 f(x)dx=\frac{1}{b-a}\int_a^b x^2dx=\frac{1}{b-a}\left(\frac{1}{3}x^3\right)\Bigg|_a^b$$

$$=\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}=\frac{a^2+ab+b^3}{3}$$

분산 : $$Var(X)=E(X^2)-E(X)^2$$

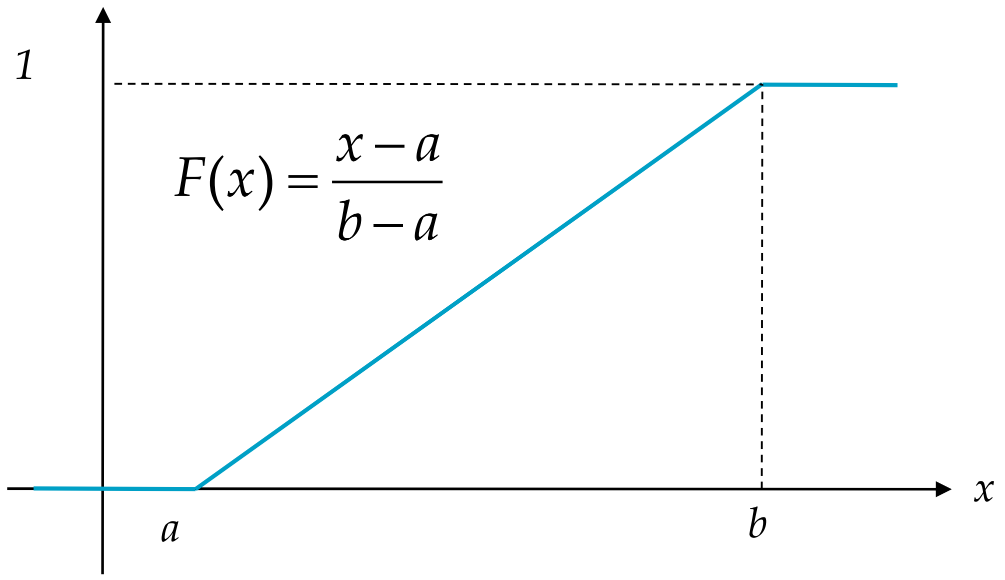
$$=\frac{a^2+ab+b^3}{3}-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}$$

① 평균 : $\mu=\dfrac{a+b}{2}$

② 분산 : $\sigma^2=\dfrac{(b-a)^2}{12}$

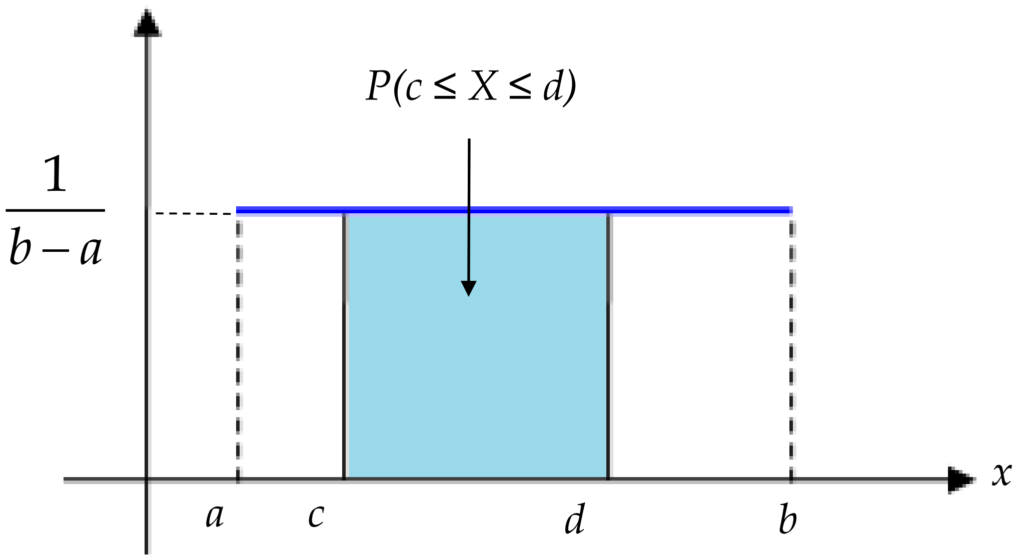
분포함수

$$F(x)=\int_a^x f(t)dt=\frac{1}{b-a}\int_a^x dt=\frac{x-a}{b-a},\quad a\le x\le b$$



확률 계산

$$P(c \le X \le d) = F(d) - F(c)$$
$$= \frac{d-a}{b-a} - \frac{c-a}{b-a} = \frac{d-c}{b-a}$$



균등분포

예제 6-14

모뎀에 들어오는 신호의 위상각이 0과 2π에서 균등하게 분포한다고 할 때, 다음을 구하라.

(a) 신호의 위상각 X의 확률밀도함수와 분포함수

(b) 확률변수 X의 평균과 분산

(c) $P\left(\frac{\pi}{4} < X \le \frac{\pi}{2}\right)$

(d) $P\left(\frac{\pi}{3} < X \le \frac{3\pi}{4} \mid X > \frac{\pi}{2}\right)$

풀이

(a) X가 $[0,\ 2\pi]$에서 균등분포를 이루므로 확률밀도함수와 분포함수는 다음과 같다.

$$f(x) = \begin{cases} \dfrac{1}{2\pi} & ,0 \le x \le 2\pi \\ 0 & ,\text{다른 곳에서} \end{cases} \qquad F(x) = \begin{cases} 0 & ,x < 0 \\ \dfrac{x}{2\pi} & ,0 \le x < 2\pi \\ 1 & ,x \ge 2\pi \end{cases}$$

균등분포

(b) 평균과 분산은 각각 다음과 같다.

$$\mu = \frac{2\pi - 0}{2} = \pi \qquad \sigma^2 = \frac{(2\pi - 0)^2}{12} = \frac{\pi^2}{3}$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P\left(\frac{\pi}{4} < X \le \frac{\pi}{2}\right) = \frac{1}{2\pi}\left(\frac{\pi}{2} - \frac{\pi}{4}\right) = \frac{1}{8}$$

(d) 구하고자 하는 확률은 다음과 같다.

$$P\left(\frac{\pi}{3} < X \le \frac{3\pi}{4} \middle| X > \frac{\pi}{2}\right) = \frac{P\left(\frac{\pi}{3} < X \le \frac{3\pi}{4}, X > \frac{\pi}{2}\right)}{P\left(X > \frac{\pi}{2}\right)} = \frac{P\left(\frac{\pi}{2} < X \le \frac{3\pi}{4}\right)}{P\left(X > \frac{\pi}{2}\right)}$$

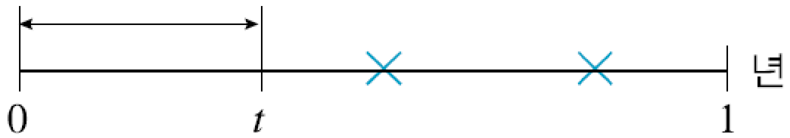
$$= \frac{\frac{1}{2\pi}\left(\frac{3\pi}{4} - \frac{\pi}{2}\right)}{1 - \frac{\pi}{2} \times \frac{1}{2\pi}} = \frac{1/8}{3/4} = \frac{1}{6}$$

지수분포

예 **1년 동안 작업장에서 발생하는 안전사고가 평균 2회인 푸아송분포에 따라 발생한다고 하자.**

그리고 1월 1일부터 관측한 이후로 첫 번째 안전사고가 발생할 때까지 걸리는 시간에 대한 확률모형을 구한다고 하자. 그러면 1년 동안 2회의 비율로 안전사고가 발생하므로 t일 까지 발생한 사고 건수는 평균 $2t$이다. 그리고 이때까지 발생한 사고 건수를 확률변수 X라 하면, X의 확률질량함수는 다음과 같다.

$$f(x) = \frac{(2t)^x}{x!} e^{-2t}, \quad x = 0, 1, 2, \cdots$$



지수분포

처음 사고가 발생할 때까지 걸리는 시간을 T라 하면 $[T > t]$인 사건은 관측을 시작한 이후로 t일을 초과한 이후에 교통사고가 발생함을 의미하고, 이것은 $[0, t]$에서 사고가 전혀 발생하지 않음을 의미한다. 즉, 두 사건 $[T > t]$와 $[X(t) = 0]$은 동치이다.
따라서 다음 확률을 얻는다.

$$P(T > t) = P[X(t) = 0] = e^{-2t}$$

➡

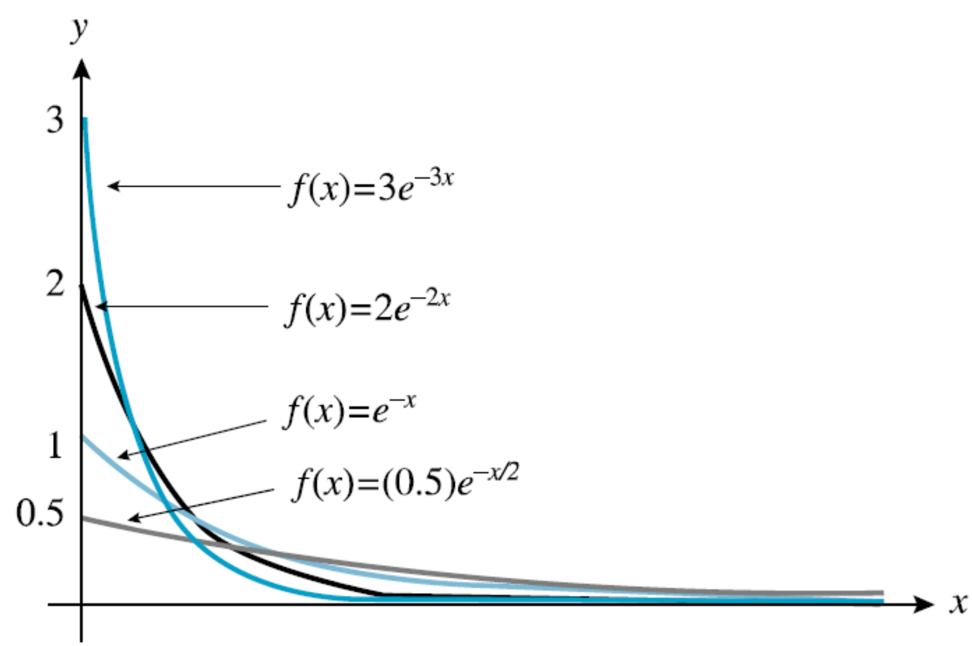
$$F(t) = P(T \le t) = 1 - P(T > t) = 1 - e^{-2t}$$

$$f(t) = \frac{d}{dt} F(t) = 2e^{-2t}$$

지수분포

지수분포exponential distribution :
확률변수 X의 상태공간 $S_X = \{x : x > 0\}$에서 다음 확률밀도함수를 갖는 확률분포. 이때 l는 양수이고 $X \sim Exp(\lambda)$로 나타낸다.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x > 0 \\ 0 & ,\text{다른 곳에서} \end{cases}$$



지수분포

$$\mu = E(X) = \int_0^\infty x(\lambda e^{-\lambda x})dx = \lim_{u\to\infty}\left(-\frac{\lambda x+1}{\lambda} e^{-\lambda x}\Big|_0^u \right) = \frac{1}{\lambda}$$

$$E(X^2) = \int_0^\infty x^2(\lambda e^{-\lambda x})dx = \lim_{u\to\infty}\left(-\frac{\lambda^2 x^2 + 2\lambda x + 2}{\lambda^2} e^{-\lambda x}\Big|_0^u \right) = \frac{2}{\lambda^2}$$

$$\sigma^2 = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

① 평균 : $\mu = \frac{1}{\lambda}$

② 분산 : $\sigma^2 = \frac{1}{\lambda^2}$

분포함수

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = \lim_{u \to \infty} \left(-e^{-\lambda t} \Big|_0^x \right) = 1 - e^{-\lambda x}, \quad x > 0$$

분포함수

예제 6-15

방사성 선원이 푸아송 분포를 따라 1분당 평균 10개의 입자를 발사한다고 할 때, 다음을 구하라.

(a) 첫 번째 입자가 발사될 때까지 걸리는 시간 X의 확률밀도함수

(b) 첫 번째 입자가 발사될 때까지 걸리는 평균 시간과 분산

(c) 분포함수 $F(x)$

풀이

(a) 단위시간 당 평균 10개의 입자가 발사되므로 첫 번째 입자가 발사될 때까지 걸리는 시간 X의 확률밀도함수는$f(x) = 10e^{-10x}, x > 0$이다.

(b) $X \sim Exp(10)$이므로 의 평균은 $\mu = \dfrac{1}{10}$ (분)이고 분산은 $\sigma^2 = \dfrac{1}{100}$ 이다.

(c) $X \sim Exp(10)$이므로 에 대하여 분포함수와 생존함수는 각각 다음과 같다.

$$F(x) = 1 - e^{-10x},\ S(x) = e^{-10x}$$

비기억성 성질

❖ $X \sim Exp(\lambda)$이면 양의 정수 a, b에 대하여 다음이 성립한다.

$$\boxed{P(X > a+b \mid X > a) = P(X > b)}$$

$$P(X > a+b \mid X > a) = \frac{P(X > a, X > a+b)}{P(X > a)} = \frac{P(X > a+b)}{P(X > a)}$$ 이고

$$P(X > b) = \int_b^\infty \lambda e^{-\lambda x} dx = \lim_{u \to \infty} -e^{-\lambda x}\Big|_b^u = e^{-\lambda b}$$ 이므로

$$P(X > a) = e^{-\lambda a}, \quad P(X > a+b) = e^{-\lambda(a+b)}$$

$$P(X > a+b \mid X > a) = \frac{P(X > a+b)}{P(X > a)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = P(X > b)$$

비기억성 성질

예제 6-16

[예제 6−15]에서 12초 전에 첫 번째 입자가 발사되지 않았다고 할 때, 앞으로 12초를 더 기다려야 할 확률을 구하라.

풀이

12초는 $\frac{1}{5}$ 분, 24초는 $\frac{2}{5}$ 분이고 지수분포는 비기억성 성질을 가지므로 구하고자 하는 확률은 다음과 같다.

$$P\left(X \ge \frac{2}{5} \middle| X > \frac{1}{5}\right) = P\left(X \ge \frac{1}{5}\right) = 1 - \int_0^{1/5} 10e^{-10x}dx$$

$$= 1 - \left(1 - e^{-2}\right) = e^{-2} = 0.1353$$

감마분포의 생성

감마함수 : $\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt, \quad \Rightarrow \quad \Gamma(\alpha)>0$

이때 $t=\dfrac{x}{\beta}, \quad \beta>0$ 이라 하면, $\quad dt=\dfrac{1}{\beta}dx$ 이므로

$$\Gamma(\alpha)=\int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}\frac{1}{\beta}dx=\int_0^\infty \frac{1}{\beta^\alpha}x^{\alpha-1}e^{-x/\beta}dx$$

$$\Rightarrow \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}dx=1$$

❖ 다음 함수는 확률밀도함수의 조건을 만족한다.

$$f(x)=\begin{cases}\dfrac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta} & ,x>0\\ 0 & ,\text{다른 곳에서}\end{cases}$$

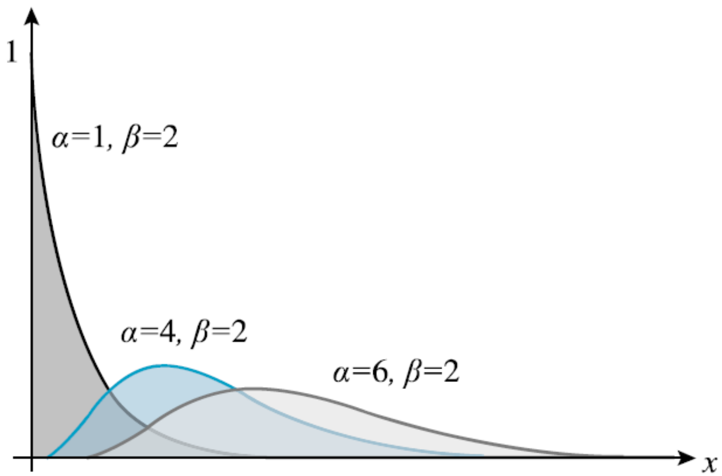
감마분포

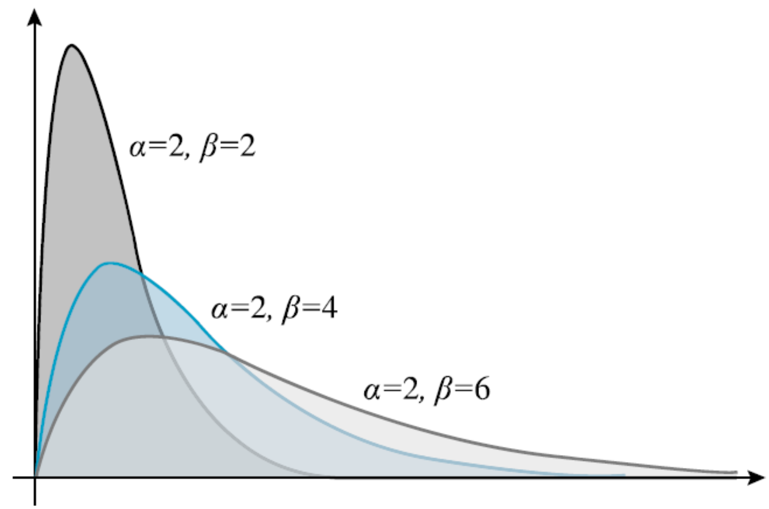
감마분포Gamma distribution :
확률변수 X의 상태공간 $S_X = \{x : x > 0\}$에서 다음 확률밀도함수를 갖는 확률분포.
이때 $X \sim \Gamma(\alpha, \beta)$로 나타낸다. 여기서 $\alpha > 0,\ \beta > 0$ 이다.

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} & ,x > 0 \\ 0 & ,\text{다른 곳에서} \end{cases}$$

감마분포

❖ 모수 α를 **형상모수**shape parameter , β를 **척도모수**scale parameter라 한다.





감마분포

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} \frac{x}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx$$

$$= \int_0^{\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha} e^{-x/\beta} dx$$

$$= \frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)} \int_0^{\infty} \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{(\alpha+1)-1} e^{-x/\beta} dx$$

$$= \frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)} = \frac{\Gamma(\alpha)\alpha\beta}{\Gamma(\alpha)} = \alpha\beta$$

$$E(X^2) = \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)} = \frac{\Gamma(\alpha)\alpha(\alpha+1)\beta^2}{\Gamma(\alpha)} = \alpha(\alpha+1)\beta^2$$

$$Var(X) = E(X^2) - E(X)^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2$$

① 평균 : $\mu = \alpha\beta$

② 분산 : $\sigma^2 = \alpha\beta^2$

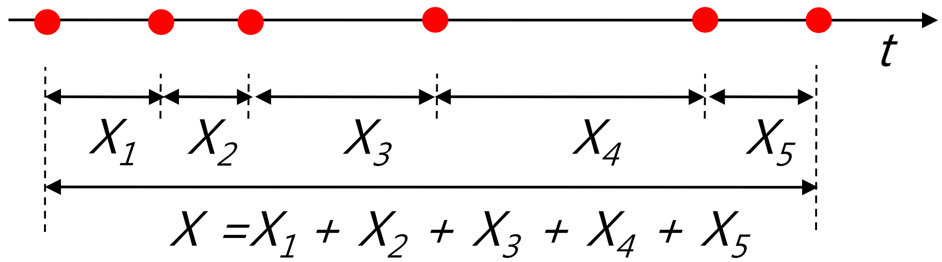
감마분포

❖ 모수 $\alpha = 1$인 감마분포는 모수 $\lambda = 1/\beta$ 인 지수분포이다.

$$\alpha = 1 \quad \Rightarrow \quad \Gamma(1) = 1 \quad \Rightarrow \quad f(x) = \frac{1}{\beta} e^{-x/\beta}$$

예 **비율 λ인 포아송분포에 따라 첫 번째 사건이 발생할 때까지 걸리는 시간을 X_1 , 첫 번째 사건 이후로 다시 첫 번째 사건이 발생할 때까지 걸리는 시간을 X_2 , 같은 방법으로 n – 1번째 사건과 n번째 사건 사이의 대기시간을 X_n이라 하면 n번째 사건이 발생할 때까지 걸리는 시간 $X = X_1 + X_2 + \ldots + X_n$에 대한 확률분포는 감마분포이다.**

감마분포



다섯 번째 사건이 발생할 때까지 걸리는 총시간 : X

X 의 평균과 분산 : $\mu_X = E(X_1 + X_2 + \cdots + X_n) = \dfrac{n}{\lambda}$

$$\sigma_X^2 = Var(X_1 + X_2 + \cdots + X_n) = \frac{n}{\lambda^2}$$

감마분포

❖ 독립인 확률변수 $X_i \sim Exp(\lambda)$, $i = 1, 2, \ldots, n$에 대하여 $X = X_1 + X_2 + \ldots + X_n$은 $\alpha = n$, $\beta = 1/\lambda$인 감마분포이다. 즉, 다음을 얻는다.

$$X_i : i.i.d \ \ Exp(\lambda), i = 1, 2, \cdots, n \quad \Rightarrow \quad X = X_1 + X_2 + \cdots + X_n \sim \Gamma\left(n, \frac{1}{\lambda}\right)$$

❖ 감마분포의 형상모수 a가 자연수인 경우, 즉 자연수 $a = n$인 감마분포를 **얼랑분포**Erlang distribution라 하며, 얼랑분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, \quad x > 0$$

감마분포

예제 6-17

전화 교환대에 1분당 평균 2회인 푸아송 분포를 따라 전화가 걸려온다. 교환원을 교체한 이후로 3건의 전화가 걸려올 때까지 걸리는 시간을 X라 할 때, 다음을 구하라.

(a) X 의 확률밀도함수

(b) 3건의 전화가 걸려올 때까지 걸리는 평균 시간

(c) 3건의 전화가 걸려올 때까지 걸리는 시간이 1분 이내에 있을 확률

풀이

(a) 1분 당 평균 2회의 전화가 걸려오므로 $\beta = \dfrac{1}{2}$ 이고 3건의 전화가 걸려올 때까지 걸리는 시간이 X이므로 $a = 3$이다. 그러므로 X의 확률밀도함수는 다음과 같다.

$$X \sim \Gamma\left(3, \frac{1}{2}\right) \quad \Rightarrow \quad f(x) = \frac{1}{\Gamma(3)(1/2)^3} x^2 e^{-2x} = 4x^2 e^{-2x}, \quad x > 0$$

(b) $\alpha = 3,\ \beta = ½$이므로 $\mu = 3/2 = 1.5$(분)이다.

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X \le 1) = \int_0^1 4x^2 e^{-2x} dx = \left[-(2x^2 + 2x + 1)e^{-2x}\right]_0^1 = 1 - 5e^{-2}$$

베타분포의 생성

❖ 베타함수 : $\mathrm{Beta}(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx, \quad \alpha>0,\beta>0, \quad \Rightarrow \quad \mathrm{Beta}(\alpha,\beta)>0$

$$\int_0^1 \frac{1}{\mathrm{Beta}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx=1$$

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx=1$$

$\mathrm{Beta}(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ 이므로

❖ 다음 함수는 확률밀도함수의 조건을 만족한다.

$$f(x)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & ,0<x<1 \\ 0 & ,\text{다른 곳에서}\end{cases}$$

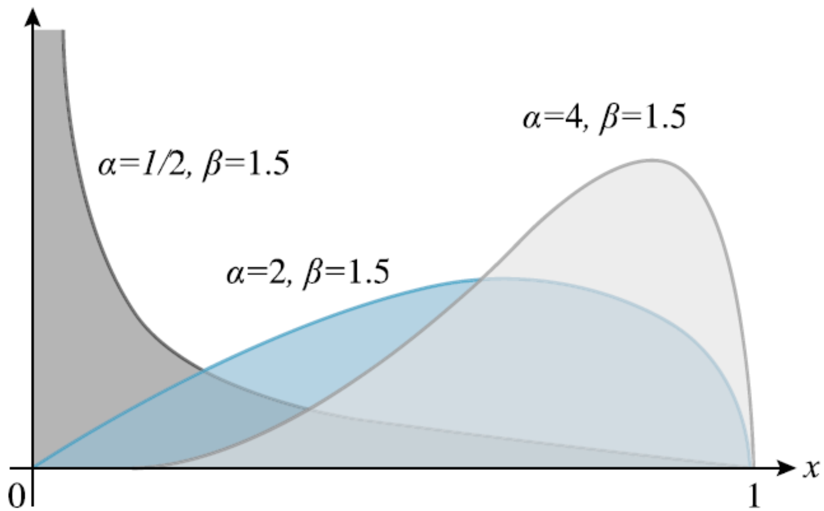
베타분포의 생성

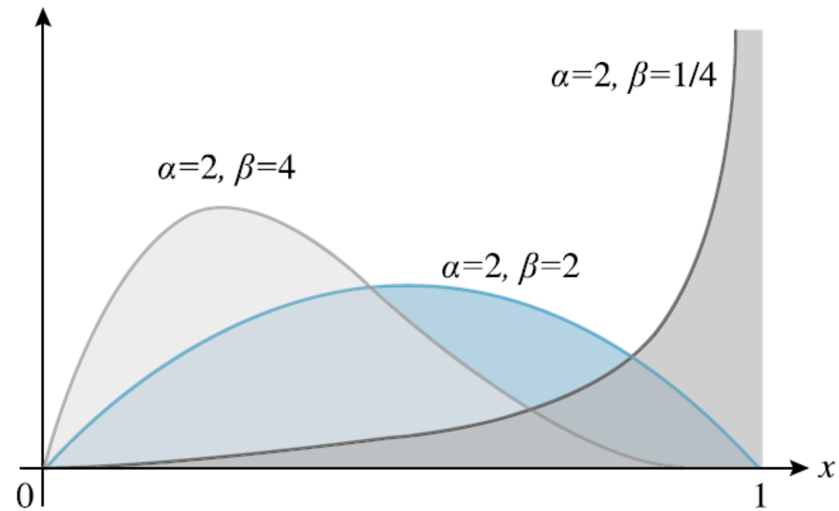
베타분포 Beta distribution :
확률변수 X의 상태공간 $S_X=\{x : 0 < x < 1\}$에서 다음 확률밀도함수를 갖는 확률분포. 이때 $X \sim \text{Beta}(\alpha, \beta)$로 나타낸다. 여기서 $\alpha > 0, \ \beta > 0$ 이다.

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & , 0 < x < 1 \\ 0 & ,\text{다른 곳에서} \end{cases}$$

베타분포의 생성

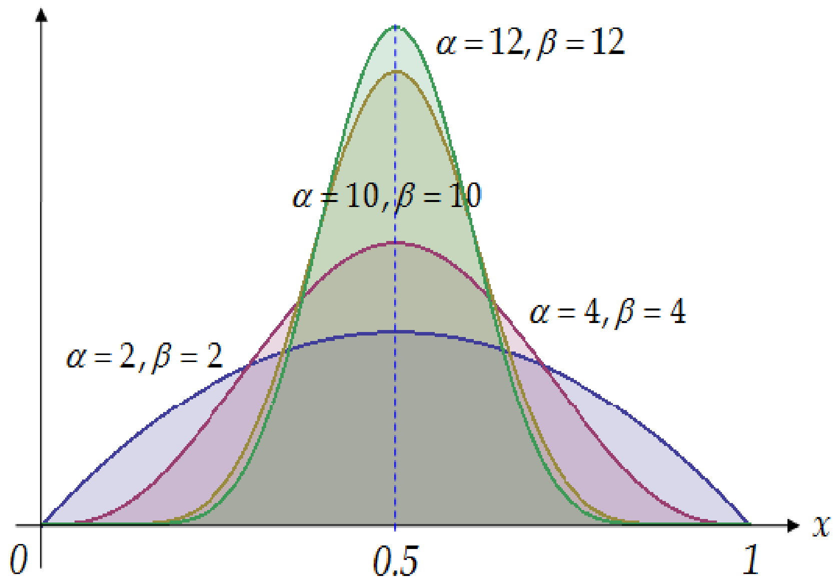
❖ $\alpha \neq \beta$ 인 경우 베타분포의 그래프





베타분포의 생성

❖ $\alpha = \beta$ 인 경우 베타분포의 그래프



❖ $\alpha = \beta$ 가 커질수록 베타분포는 종모양을 이룬다.

❖ $\alpha = \beta = 1$ 이면 $0 < x < 1$에서 균등분포를 이룬다. 즉, $Beta(1,1) = U(0,1)$이다.

베타분포의 생성

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha}(1-x)^{\beta-1}dx$$

$$=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)}\int_0^1\frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)}x^{(\alpha+1)-1}(1-x)^{\beta-1}dx$$

$$=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+1)}{\Gamma(\alpha+1+\beta)}=\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+1+\beta)}$$

$$=\frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)}=\frac{\alpha}{\alpha+\beta}$$

$$E(X^2)=\frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}$$

$$Var(X)=E(X^2)-E(X)^2=\frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}-\left(\frac{\alpha}{\alpha+\beta}\right)^2$$

$$=\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

① 평균 : $\mu=\dfrac{\alpha}{\alpha+\beta}$

② 분산 : $\sigma^2=\dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

베타분포의 생성

예제 6-18

금속합금에 포함된 주석의 비율 X가 모수 $\alpha = 4$, $\beta = 3$인 베타분포를 따른다고 할 때, 다음을 구하라.

(a) X의 확률밀도함수

(b) 금속합금에 포함된 주석의 평균 비율

(c) 금속합금에 주석이 30% 이내로 포함될 확률

풀이

(a) X가 모수$\alpha = 4$, $\beta = 3$인 베타분포에 따르므로 우선 다음을 먼저 구한다.

$$\frac{\Gamma(7)}{\Gamma(4)\Gamma(3)} = \frac{6!}{(3!)(2!)} = 60$$

따라서 X의 확률밀도함수는 다음과 같다.

$$f(x) = 60x^3(1-x)^2, \quad 0 < x < 1$$

베타분포의 생성

(b) $\alpha = 4,\ \beta = 3$이므로 $\mu = 4/7$이다.

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X \le 0.3) = \int_0^{0.3} 60x^3(1-x)^2\,dx = \left[10x^6 - 24x^5 + 14x^4\right]_0^{0.3} = 0.07047$$

와이블분포

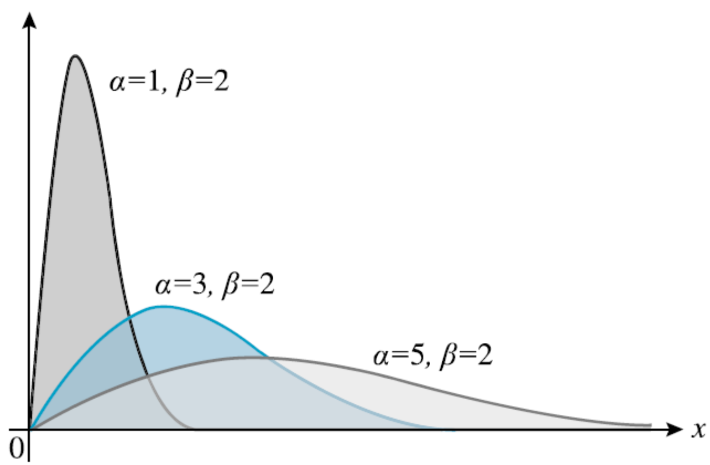
❖ 와이블이 재료의 파괴강도를 분석하는 과정에서 고안한 확률분포이며, 다음과 같은 경우에 확률모형을 설명하기 위하여 많이 사용한다.

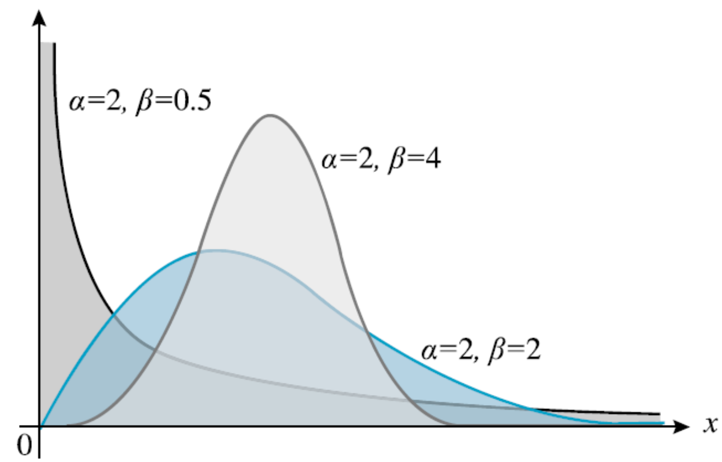
- 금속 및 복합 재료의 강도
- 전자 및 기계부품의 수명
- 의료사고와 같은 재해에 대비하기 위한 재해보험

와이블분포

와이블분포Weibull distribution :
확률변수 X의 상태공간 $S_X = \{x: x > 0\}$에서 다음 확률밀도함수를 갖는 확률분포. 이때 $X \sim \mathrm{Wei}(\alpha, \beta)$로 나타낸다. 여기서 $\alpha > 0$, $\beta > 0$ 이다.

❖ 일반적으로 와이블 분포는 다음과 같이 양의 비대칭인 분포를 이룬다.





와이블분포

$$E(X)=\int_0^\infty x\cdot\frac{\beta}{\alpha^\beta}x^{\beta-1}e^{-(x/\alpha)^\beta}dx$$

$$\left[(x/\alpha)^\beta=t \quad\Rightarrow\quad x=\alpha t^{1/\beta},\ dx=\frac{\alpha}{\beta}t^{(1/\beta)-1}dt\right]$$

$$=\int_0^\infty\frac{\beta}{\alpha^\beta}\cdot\frac{\alpha}{\beta}\left(\alpha t^{1/\beta}\right)^\beta t^{(1/\beta)-1}e^{-t}dt$$

$$=\alpha\int_0^\infty t^{1+(1/\beta)-1}e^{-t}dt=\alpha\Gamma\left(1+\frac{1}{\beta}\right)$$

$$E(X^2)=\alpha^2\Gamma\left(1+\frac{2}{\beta}\right)$$

$$Var(X)=E(X^2)-E(X)^2=\alpha^2\Gamma\left(1+\frac{2}{\beta}\right)-\alpha^2\Gamma\left(1+\frac{1}{\beta}\right)^2$$

① 평균 : $\mu=\alpha\Gamma\left(1+\frac{1}{\beta}\right)$

② 분산 : $\sigma^2=\alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma\left(1+\frac{1}{\beta}\right)^2\right]$

와이블분포

❖ 분포함수 : $F(x) = 1 - e^{-(x/\alpha)^{\beta}}, \quad x > 0$

❖ $\beta > 1$이면 시간이 흐름에 따라 시스템의 위험률이 시간의 멱승에 비례하면서 증가

$0 < \beta < 1$이면 $x = 0$부근에서 급격히 증가한다.

$\beta = 1$이면 고장률은 지수분포($\lambda = 1/\alpha$)와 같이 일정하다.

❖ 모수 $\alpha = \sqrt{2}\theta, \ \beta = 2$ 이면 와이블분포는 **레일리 분포** **Rayleigh distribution**가 된다.

$$\boxed{f(x) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)} \quad , x > 0}$$

와이블분포

예제 6-19

어떤 고온에서 박테리아의 생존시간(분) X가 모수 $\alpha = 5$, $\beta = 3$인 와이블 분포를 따른다고 할 때, 다음을 구하라.

(a) 분포함수 $F(x)$와 확률 $P(X \le 4)$

(b) X의 평균, 분산, 중위수

(c) $F(x_0) = 0.95$를 만족하는 x_0

풀이

(a) 박테리아의 생존시간을 X라 하면, 모수 $\alpha = 5, \beta = 3$ 인 와이블 분포에 따르므로 분포함수와 구하고자 하는 확률은 다음과 같다.

$$F(x) = 1 - e^{-(x/5)^3} = 1 - \exp\left(-\frac{x^3}{125}\right), \quad x > 0$$

$$P(X \le 4) = 1 - \exp\left(-\frac{4^3}{125}\right) = 0.4007$$

와이블분포

(b) 구하고자 하는 평균과 중위수 그리고 분산은 다음과 같다.

평균 : $\mu = 5\Gamma\left(1+\frac{1}{3}\right) = 5\times\frac{1}{3}\times\Gamma\left(\frac{1}{3}\right)(= 4.4649)$

분산 : $\sigma^2 = 25\left[\Gamma\left(1+\frac{2}{3}\right) - \Gamma\left(1+\frac{1}{3}\right)^2\right] = 25\left[\frac{2}{3}\Gamma\left(\frac{2}{3}\right) - \frac{1}{9}\Gamma\left(\frac{1}{3}\right)^2\right](= 2.6333)$

중위수 : $F(x_0) = 1-\exp\left(-\frac{x_0^3}{125}\right) = 0.5;\quad \exp\left(-\frac{x_0^3}{125}\right) = 0.5;\quad -\frac{x_0^3}{125} = \ln 0.5$

$$x_0^3 = -125\ln 0.5 = 86.6434;\quad x_0 = \sqrt[3]{86.6434} = 4.425$$

(c) 구하고자 하는 x_0은 다음과 같다.

$$F(x_0) = 1-\exp\left(-\frac{x_0^3}{125}\right) = 0.95;\quad \exp\left(-\frac{x_0^3}{125}\right) = 0.05;\quad -\frac{x_0^3}{125} = \ln 0.05$$

$$x_0^3 = -125\ln 0.05 = 374.467;\quad x_0 = \sqrt[3]{374.467} = 7.2073$$

약 7.2분 안에 박테리아의 *95%*가 죽는다.

파레토분포

❖ 파레토 분포는 경제적 효율성과 수입의 재분배에 대한 연구에서 사용된 확률모형이지만 게임이론과 공학에도 다음과 같이 자연 환경적 현상에 대한 확률모형에 적용된다.

- 최대 풍속
- 최대 강우량
- 지진 혹은 산불과 같은 자연 재해의 규모
- 유전 지역에서 오일 매장량

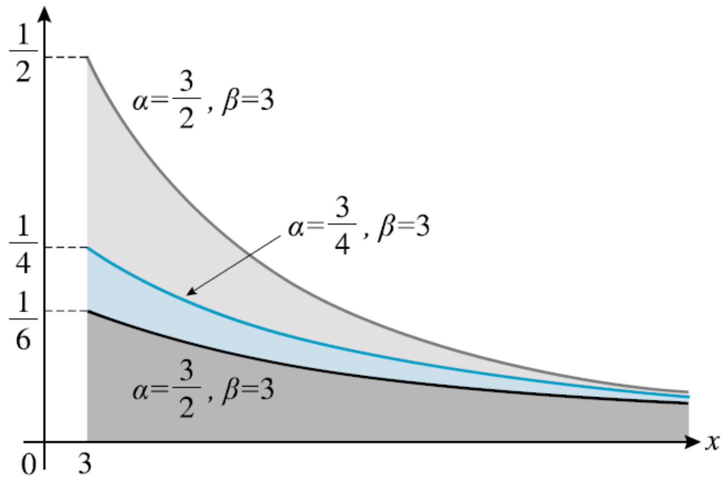
파레토분포

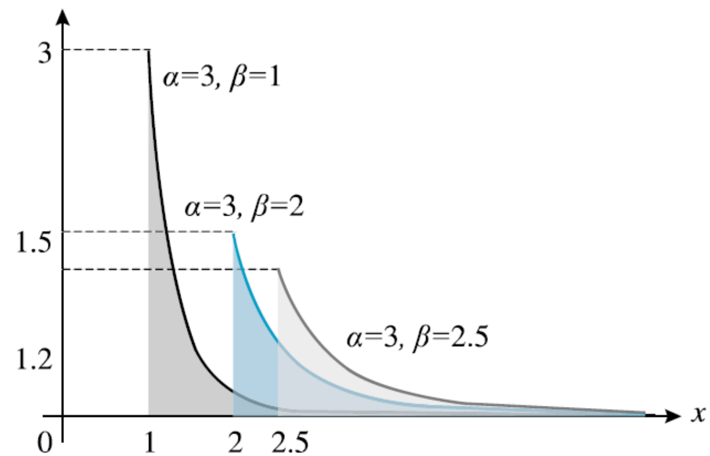
파레토분포[Pareto distribution] :
확률변수 X의 상태공간 $S_X=\{x : x > \beta\}$에서 다음 확률밀도함수를 갖는 확률분포. 이때 $X \sim \mathrm{Pareto}(\alpha, \beta)$로 나타낸다.
여기서 $\alpha > 0,\ \beta > 0$이다.

$$f(x) = \begin{cases} \dfrac{\alpha\beta^{\alpha}}{x^{\alpha+1}} & , x > \beta \\ 0 & ,\text{다른 곳에서} \end{cases}$$

파레토분포

❖ α는 모양을 결정하는 형상모수이고 β는 시작하는 위치를 나타내는 위치모수이다.





파레토분포

❖ α는 분포의 산포도를 나타내는 척도로서 α가 작을수록 분포의 꼬리가 두터워지며 산포가 커진다.

$$E(X)=\int_{\beta}^{\infty} x\cdot\frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}dx=\alpha\beta^{\alpha}\int_{\beta}^{\infty}x^{-\alpha}dx$$

$$=\alpha\beta^{\alpha}\left[-\frac{1}{\alpha-1}x^{-\alpha+1}\right]_{\beta}^{\infty}=\alpha\beta^{\alpha}\times\frac{\beta^{-\alpha+1}}{\alpha-1}=\frac{\alpha\beta}{\alpha-1},\quad \alpha>1$$

$$E(X^2)=\frac{\alpha\beta^2}{\alpha-2},\quad \alpha>2$$

$$Var(X)=E(X^2)-E(X)^2=\frac{\alpha\beta^2}{\alpha-2}-\left(\frac{\alpha\beta}{\alpha-1}\right)^2$$

$$=\frac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2},\quad \alpha>2$$

① 평균 : $\mu=\dfrac{\alpha\beta}{\alpha-1},\quad \alpha>1$

② 분산 : $\sigma^2=\dfrac{\alpha\beta^2}{(\alpha-2)(\alpha-1)^2},\quad \alpha>2$

예제 6-20

보험증권에 의한 보험 지급금이 $\alpha = 3$, $\beta = 4$인 파레토 분포를 따른다고 할 때, 다음을 구하라. 이때 단위는 100만원이다.

(a) 보험 지급금의 평균과 분산

(b) 보험 지급금이 6을 초과할 확률

(c) (b)의 조건 아래서 보험 지급금이 8을 초과할 확률

풀이

(a) 보험지급금은 모수 $\alpha = 3$, $\beta = 4$인 파레토 분포에 따르므로 평균과 분산은 각각 다음과 같다.

$$\mu = \frac{3 \times 4}{2} = 6, \quad \sigma^2 = \frac{3 \times 4^2}{2^2 \times 1} = 12$$

파레토분포

(b) 구하고자 하는 확률은 다음과 같다.

$$P(X>6)=S(6)=\left(\frac{4}{6}\right)^3=\frac{8}{27}$$

(c) 구하고자 하는 확률은 다음과 같다.

$$P(X>8|X>6)=\frac{P(X>8, X>6)}{P(X>6)}=\frac{P(X>8)}{P(X>6)}=\frac{S(8)}{S(6)}=\frac{1/8}{8/27}=\frac{27}{64}$$

Q&A