4.1절 ~ 4.4절

3주1강. 확률변수와 확률분포1

복습

복습 : 순열과 중복 순열

- **순열** : n개의 서로 다른 개체가 있을 때 그 중에서 r개를 선택하여 일렬로 세우는 방법의 수.
- 기호 : $${}_nP_r = n(n-1)(n-2)\cdots(n-r+1)$$ $$= \frac{n!}{(n-r)!}$$
- 약속 : $0! = 1,$ $\quad {}_nP_0 = 1$

- **중복순열** : 서로 다른 n개에서 중복을 허락해 r개를 선택하여 일렬로 나열하는 방법의 수

$${}_n\Pi_r = n^r$$

복습 : 조합과 중복조합

- **조합** : n개의 서로 다른 개체 중에서 임의로 r개를 선택하는 방법의 수

$$ {}_nC_r = \binom{n}{r} $$

$$ \binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!} $$

- $ {}_nC_r = {}_nC_{n-r},$ $\qquad {}_nC_n = 1,$ $\qquad {}_nC_0 = 1$

- **중복조합** : 서로 다른 n개에서 중복을 허락하여 r개를 선택하는 방법의 수

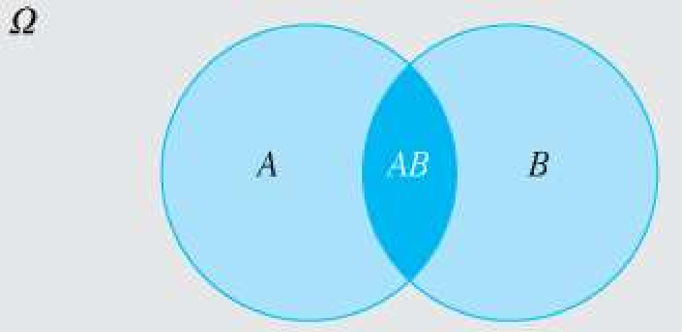
$$ {}_nH_r = {}_{n+r-1}C_r $$

복습 : 조건부 확률

- 사건 A가 주어졌을 때 사건 B의 조건부 확률은 $P(A) \neq 0$이라는 전제하에서 다음과 같이 정의한다.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- 사건 A와 B를 벤 다이어그램으로 그리면 다음과 같다.



- 즉, 조건부 확률은사건 A의 크기에 대한 곱사건 AB의 크기의 비율이라고 할 수 있다. $P(B|A)$는 실험에서 나타나는 결과를 A에 한정할 때 사건 B가 나타날 가능성을 측정하는 것이다.

복습 : 확률적 독립성--두 사건에 발생이 서로 영향을 미치지 않는다.

- 두 사건 A, B가 다음 조건 중 하나를 만족하면 서로 확률적으로 독립이라고 한다.

1. $P(A \cap B) = P(A) \cdot P(B)$
2. $P(A|B) = P(A)$
3. $P(B|A) = P(B)$

서로 배반과 서로 독립

▶ 서로 배반

두 사건 A, B가 서로 배반이라는 조건은 $AB = \emptyset$을 의미하므로 항상 $P(AB) = 0$ 이다.

▶ 서로 독립

두 사건 A, B가 독립이라는 조건은 $P(AB) = P(A) \cdot P(B)$이다.

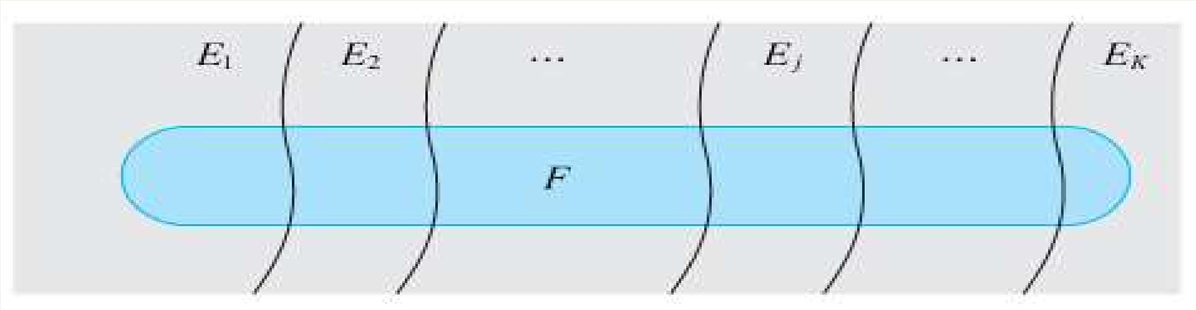
▶ 따라서 $P(A)$ 또는 $P(B)$ 가 0이 아닌 한, 두 사건이 상호배반이라는 조건과 두 사건이 상호독립이라는 조건은 동시에 만족할 수 없다.

복습 : 베이즈 정리 (Bayes' Theorem)

- 표본공간 Ω가 k개의 사건 열 $E_1, E_2, \dots, E_k$에 의하여 분할(partition)된다고 한다. 다른 사건 F가 일어났을 때 이 사건이 E_j에서 일어날 확률은 다음과 같이 계산한다.

$$P(E_j|F) = \frac{P(FE_j)}{P(F)} = \frac{P(FE_j)}{\sum_{i=1}^{k} P(FE_i)} = \frac{P(F|E_j)P(E_j)}{\sum_{i=1}^{k} P(F|E_i)P(E_i)}$$

- 여기에서 분할(partition)이란 $E_1, E_2, \dots, E_k$가 상호배반이며, $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$ 임을 의미한다.



복습 : 예제

- 4지 선다형 문제에서 한 학생이 정답을 알고 있을 확률이 0.8이고 정답을 임의로 선택할 확률이 0.2이며, 정답을 임의로 선택할 때, 정답을 맞출 확률이 0.25라고 한다. 그 학생이 정답을 맞추었다고 할 때, 그 학생이 정답을 임의로 선택했을 확률은 얼마인가?

4.1절

4.1. 확률변수

확률 변수

- **확률 변수** : 실험의 표본공간으로부터 실수 값으로의 변환함수이다. 따라서 확률변수가 특정 실수 값을 가질 확률은 표본공간의 원소에 대한 확률로부터 유도된다.

예4-1. 동전 2개를 던지는 실험에서 표본공간은 다음과 같다.

$$\Omega = \{HH, HT, TH, TT\}$$

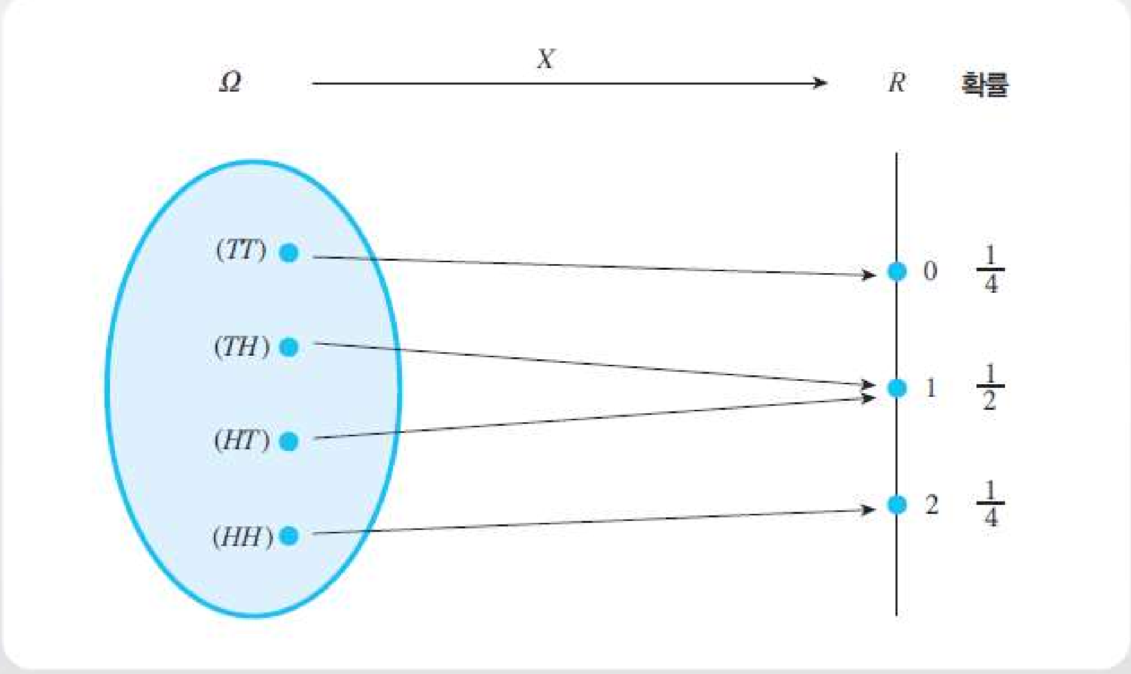
이 실험에서 각각의 원소들이 나타날 확률은 $\frac{1}{4}$이다. 확률변수 X를 'X =앞면의 수'로 정의할 때 가능한 값은?

(sol) X의 가능한 값은 $X\{HH\} = 2,\ X\{HT, YH\} = 1,\ X\{TT\} = 0$이고, X의 각 값에 대한 확률은 다음과 같다.

$$P_r(X = 2) = \frac{1}{4},\ P_r(X = 1) = \frac{1}{2},\ P_r(X = 0) = \frac{1}{4}$$

‘확률변수란 정의역(domain)이 표본공간이고 치역(range)이 실수값인 함수이다’ 라고 할 수 있다. 따라서 확률변수 X는 $X: \Omega \rightarrow R$으로 표현할 수 있다.

그림 4-1 표본공간의 원소와 확률변수



예4-2. 동전 3개를 던지는 실험에서 확률변수 X를 '$X \equiv$앞면의 수'라고 정의할 때 X의 확률분포표를 구하라.

(sol) 동전 3개를 던지는 실험에서 표본공간은 다음과 같다.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

→ X의 가능한 값은 $0, 1, 2, 3$임을 알 수 있다.

- Ω은 8개의 원소로 되어 있으므로 각 원소의 확률은 $\frac{1}{8}$이고 따라서 X의 각 값에 대한 확률은 다음과 같다.

$(X = 0) = \{TTT\} = 1/8$

$(X = 1) = \{TTH, THT, HTT\} = 3/8$

$(X = 2) = \{HHT, HTH, THH\} = 3/8$

$(X = 3) = \{HHH\} = 1/8$

X의 확률분포표

X	확률
0	1/8
1	3/8
2	3/8
3	1/8
합	1

예 4-3. 두 개의 주사위를 던지는 실험에서 확률변수 X를 '$X \equiv$ 두 주사위 눈금의 합'이라고 정의할 때, X의 확률분포를 구하라.

(sol) 두 개의 주사위를 던지는 실험에서 표본공간은 다음과 같이 36개의 원소로 이루어지며, 따라서 각 원소가 발생될 확률은 1/36이다. X의 가능한 값은 $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$이며, 위의 표본공간에 의하여 X의 각 값에 대한 확률이 다음과 같음을 알 수 있다.

X	2	3	4	5	6	7	8	9	10	11	12	합
확률	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

(sol) 확률변수 Y를 '$Y \equiv$ 첫 주사위 눈금-두 번째 주사위 눈금'이라고 정의할 때 Y의 가능한 값은 $-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$이며 Y의 각 값에 대한 확률은 다음과 같음을 알 수 있다.

Y	-5	-4	-3	-2	-1	0	1	2	3	4	5	합
확률	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

4.2절

4.2. 이산형 확률변수

이산형 확률변수

- 이산형 확률변수란 0이 아닌 확률 값을 갖는 실수 값이 셀 수 있는 경우를 말한다. 즉, 확률변수가 이산점에서 0이 아닌 확률을 갖는 확률변수를 말한다.
- 이산형 확률변수 X의 가능한 값을 $x_1, x_2, \cdots, x_n$이라 하고 $X = x_i$일 때의 확률이 P_i라면 다음과 같이 표현한다.

$$P_r(X = x_i) = P_i, \qquad i = 1, 2, 3, \cdots, n$$

- **이산형 확률변수의 확률조건**

1. $0 \le P_i \le 1,\ i = 1, 2, 3, \cdots, n$. 즉, 각 x_i가 나타날 확률은 0과 1사이의 값을 갖는다.
2. $\sum_{i=1}^{n} P_i = 1$. 즉, 모든 가능한 경우의 확률의 합은 1이다.

확률 질량 함수

- 이산형 확률변수는 이산점(discrete points)에서 0이 아닌 확률 값을 가지며 확률은, $P_r(X = x_i) = P_i, \quad i = 1, 2, 3, \cdots, n$으로 표현한다. 이와 같이 각 이산점에 있어서 확률의 크기를 표현하는 함수를 **확률질량함수(probability mass function)**라고 한다.

4.3절

4.3. 연속형 확률변수

연속형 확률변수

- **연속형 확률변수**란?
  가능한 값이 실수의 어느 특정구간 전체에 해당하는 확률변수를 말한다.
  즉, 특정 실수 구간에서 0이 아닌 확률을 갖는 확률변수이다.

- 연속형 확률변수는 특정한 실수구간 내에서 0이 아닌 확률을 가지므로 이 구간에 대한 확률은 함수의 형태로 표현한다.

확률밀도함수

- 연속형 확률변수 $\boldsymbol{X}$ 의 확률함수를 $\boldsymbol{f}(\boldsymbol{x})$라고 할 때, $\boldsymbol{f}(\boldsymbol{x})$는 **확률밀도함수 (probability density function; p.d.f)**라고 부르며 다음 조건을 만족한다.

1. 모든 $\boldsymbol{X}$ 값에 대하여 $\boldsymbol{f}(\boldsymbol{x}) \geq 0$이다. 즉, $\boldsymbol{X}$의 모든 실수 값에 대하여 확률밀도함수는 0 이상이다.
2. $\boldsymbol{X}$의 모든 가능한 값의 확률은 적분 $\int_{-\infty}^{\infty} f(x)\, dx$로 구하며 이 값은 항상 1이다.
3. 구간 $(\boldsymbol{a}, \boldsymbol{b})$의 확률은 $\int_a^b f(x)\, dx$이다. 즉 구간 $(\boldsymbol{a}, \boldsymbol{b})$에 대한 $\boldsymbol{X}$의 확률은 그 구간에 있어서 확률밀도함수 $\boldsymbol{f}(\boldsymbol{x})$로 만들어지는 면적의 크기이다.

예4-4. 확률변수 X가 0과 1 사이에서 균등한 분포를 가질 때 X의 확률밀도함수는 다음과 같이 표현한다.

$$\boldsymbol{f(x)} = \begin{cases} \mathbf{1}, & \mathbf{0 \le x \le 1} \\ \mathbf{0}, & \text{다른 곳에서} \end{cases}$$

(sol) 모든 실수값 X에 대하여 $f(x) = 0$ 또는 1이므로 $f(x) \ge 0$의 조건을 만족하며 다음과 같은 식을 만족하므로 $f(x)$는 확률밀도함수의 조건을 만족한다.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{1} f(x)\,dx = \int_{0}^{1} 1\,dx = 1$$

이 확률분포에서 X가 0.2에서 0.5 사이의 값을 가질 확률은 다음과 같다.

$$P_r(0.2 \le x \le 0.5) = \int_{0.2}^{0.5} f(x)\,dx = \int_{0.2}^{0.5} 1\,dx = 0.3$$

예 4-5. 확률변수 X의 확률밀도함수가 다음과 같을 때, $f(x)$를 그림으로 표현하면 아래 그림과 같다.

$$\boldsymbol{f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}}$$

(sol) 모든 x에 대하여 $f(x) \geq 0$이고, 다음과 같은 식을 만족하므로 $f(x)$는 확률밀도함수의 조건을 만족한다.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} f(x)\,dx = \int_{0}^{\infty} e^{-x}\,dx = \lim_{t \to \infty} [-e^{-x}]_0^t = \lim_{t \to \infty} (1 - e^{-t}) = 1$$

이 확률분포에서 X가 0에서 1 사이의 값을 가질 확률은 다음과 같다.

$$P_r(0 \leq x \leq 1) = \int_{0}^{1} f(x)\,dx = \int_{0}^{1} e^{-x}\,dx = [-e^{-x}]_0^1 = 1 - \frac{1}{e}$$

예 4-6. 확률변수 X가 다음과 같은 확률밀도함수를 가질 때 상수 c를 구하라.

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{cx}, & \boldsymbol{0 \le x \le 2} \\ \boldsymbol{0}, & \text{다른 곳에서} \end{cases}$$

- (sol)

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{2} f(x)\,dx = \int_{0}^{2} cx\,dx = \left[\frac{1}{2}cx^2\right]_0^2 = 2c = 1$$

- 따라서 $C = \frac{1}{2}$이다.

4.4절

4.4. 누적분포함수

누적분포함수

- **누적분포함수(cumulative distribution function ; c.d.f)**란?

  특정값 a에 대하여 확률변수 X가 $X \le a$인 모든 경우의 확률의 합으로 다음과 같이 표현한다.

$$P_r(X \le a) = F_X(a)$$

- 이산형 확률변수에서는 $F_X(a) = \sum_{x_i \le a} P(X = x_i)$이고,

- 연속형 확률변수에서는 $F_X(a) = \int_{-\infty}^{a} f(x)\, dx$이다.

이산형 확률변수의 누적분포함수

- 이산형 확률변수 X가 이산점 $x_1 < x_2 < \cdots < x_n$에서 0이 아닌 확률을 가지며 $P_r(X = x_i) = P_i$, $i = 1, 2, 3, \cdots, n$이라고 할 때, X의 누적분포함수는 다음과 같다.

1. $a < x_1$인 경우, $F(a) = 0$이다.
2. $a \geq x_1$인 경우, $F(a) = \sum_{x_i \leq a} P_i$이다.
3. $a \geq x_n$인 경우, $F(a) = \sum_{x_i \leq a} P_i = 1$이다.
4. $F(a)$는 X의 각 이산점 $x_1 < x_2 < \cdots < x_n$에서 P_i, $i = 1, 2, 3, \cdots, n$만큼씩 도약하는 계단함수이다.

예 4-7. 동전 3개를 던지는 실험에서 확률변수 X를 '$X \equiv$ 앞면의 수'라고 정의할 때 X의 확률분포는 다음과 같다. 이 확률분포의 누적분포함수를 구하라.

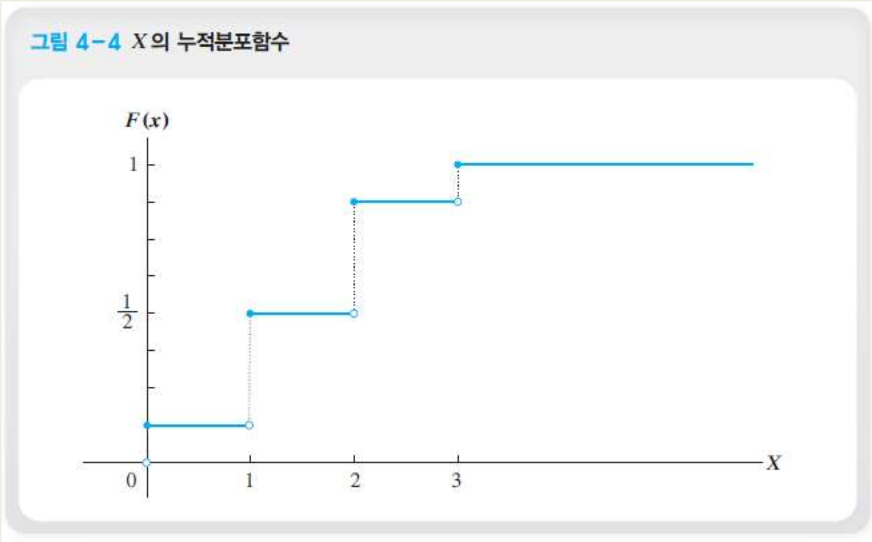
X	0	1	2	3	합
확률	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

x	$F(x) = P_r(X \leq x)$
$x < 0$	0
$0 \leq x < 1$	$\frac{1}{8}$
$1 \leq x < 2$	$\frac{1}{2}$
$2 \leq x < 3$	$\frac{7}{8}$
$x \geq 3$	1

(sol)

- $P_r(X \leq b) = 0,\ b < 0$
- $P_r(X \leq 0) = P_r(X = 0) = \frac{1}{8}$
- $P_r(X \leq 1) = P_r(X = 0) + P_r(X = 1) = \frac{1}{2}$
- $P_r(X \leq 2) = P_r(X = 0) + P_r(X = 1) + P_r(X = 2) = \frac{7}{8}$
- $P_r(X \leq a) = P_r(X = 0) + P_r(X = 1) + P_r(X = 2) + P_r(X = 3) = 1, \quad a \geq 3.$

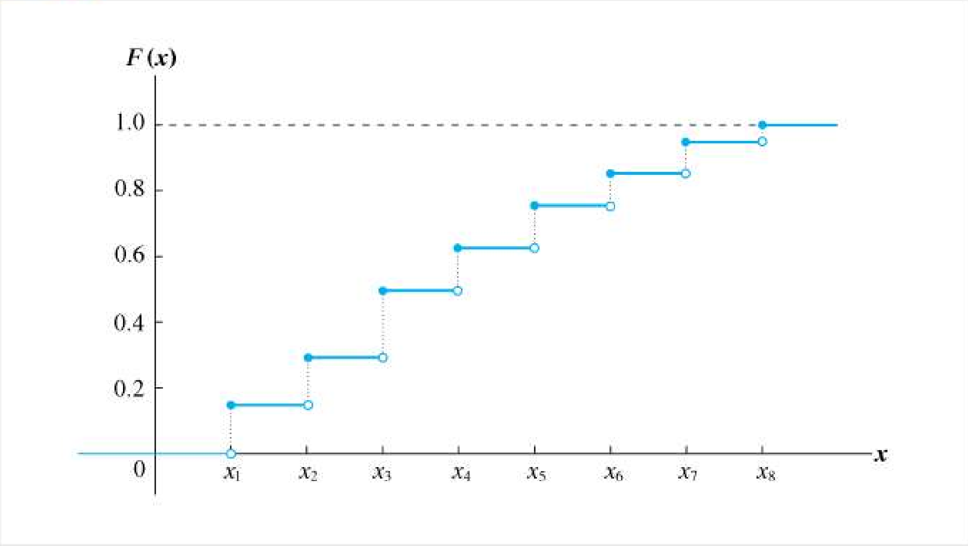
- 앞에서 구한 누적분포함수값을 그래프로 표현하면 다음과 같다.



- 이산형 확률변수는 이산점에서 0이 아닌 확률을 가지므로 누적분포함수는 각 이산점에서 도약(jump)하는 계단함수(step function) 형태이다.

- 이산형 확률변수 X가 이산점 $x_1 < x_2 < \cdots < x_n$에서 확률 $P_r(X = x_i) = P_i, \ \ i = 1, 2, \cdots, 8$을 가질 때, 누적분포함수는 다음과 같이 표현할 수 있다.

그림 6-5 이산형 확률변수의 누적분포



연속형 확률변수의 누적분포함수

- 연속형 확률변수 $\boldsymbol{X}$의 확률밀도함수를 $\boldsymbol{f(x)}$라고 할 때, $\boldsymbol{X}$의 누적분포함수 $\boldsymbol{F}(x)$는 다음과 같다.

$$\boldsymbol{F}(x) = P_r(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

- 피적분함수 $f(x)$가 $f(x) \ge 0$, '모든 x에 대하여'의 조건을 만족하므로 $F(x)$는 감소하지 않는 연속함수(non-decreasing continuous function)이다.
- 누적분포함수가 확률밀도함수의 적분에 의하여 정의되므로 확률밀도함수는 누적분포함수의 미분으로 구할 수 있다.

연속형 확률변수의 누적분포함수

$$F(-\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 0$$

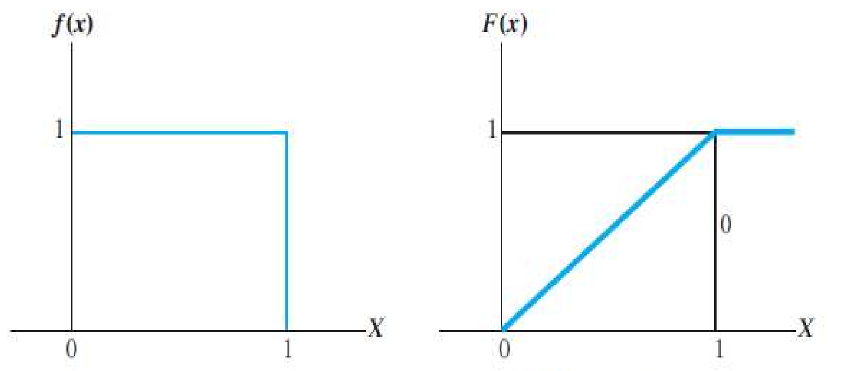
$$F(a) = \int_{-\infty}^{a} f(x)\,dx$$

$$F(\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

(예 4-4) 에 있는 확률변수 X의 분포함수와 누적분포함수

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

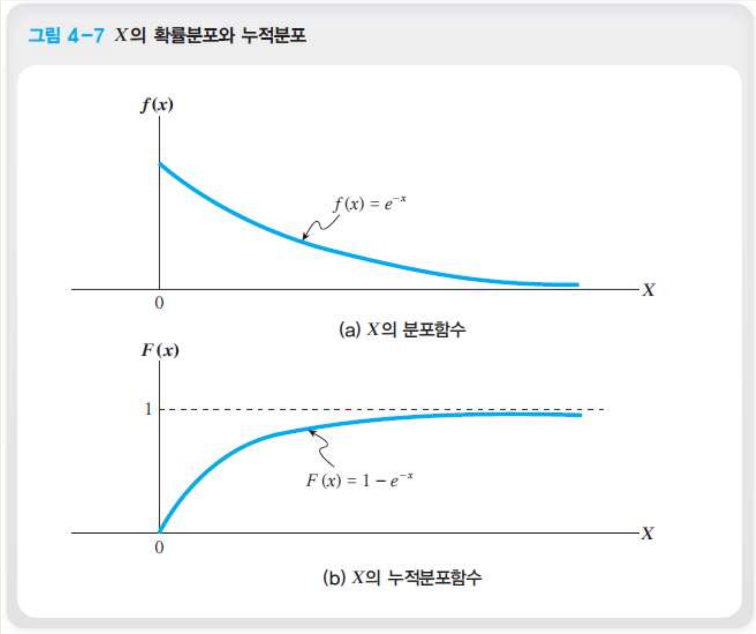
그림 4-6 X의 확률분포와 누적분포



(a) X의 분포함수　　(b) X의 누적분포함수

(예 4-5)에 있는 확률변수 X의 분포함수 누적분포함수

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



예 4-8 [예 4.4]에 있는 확률변수 X의 분포함수에 대한 누적분포함수를 구하라.

$$\boldsymbol{f(x)} = \begin{cases} \mathbf{1}, & \mathbf{0} \le \boldsymbol{x} \le \mathbf{1} \\ \mathbf{0}, & \text{다른 곳에서} \end{cases}$$

(sol) $f(x)$가 $0 \le x \le 1$ 구간에서만 0이 아닌 확률을 가지므로 $F(x)$는 x값에 따라 세 구간으로 나누어서 구하여야 한다.

(a) $x < 0$ 인 경우

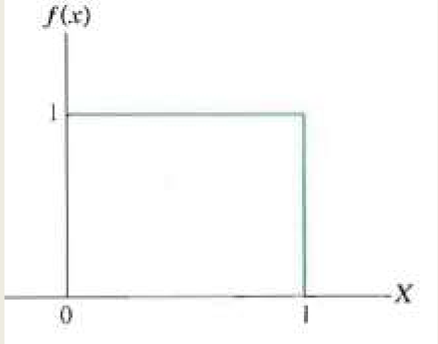
$$\boldsymbol{F}(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$$

(b) $0 \le x < 1$인 경우

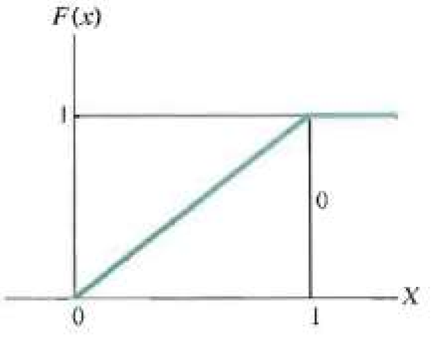
$$\boldsymbol{F}(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} 1\,dt = x$$

(c) $x \ge 1$인 경우

$$\boldsymbol{F}(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{1} f(t)\,dt = \int_{0}^{1} 1\,dt = 1$$



(a) X의 분포함수



(b) X의 누적분포함수

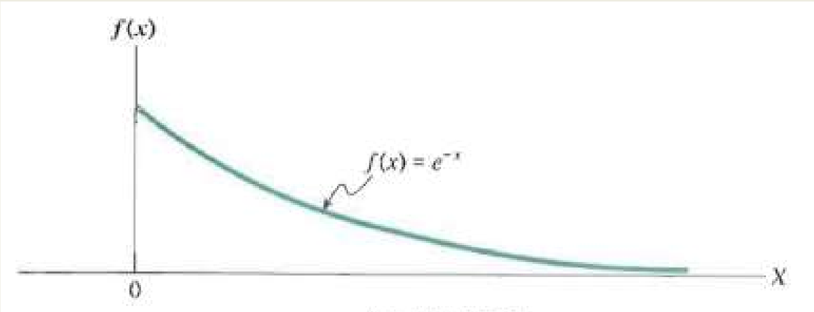
예 4-9 [예 4.5]에 있는 확률변수 X의 분포함수에 대한 누적분포함수를 구하라.

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

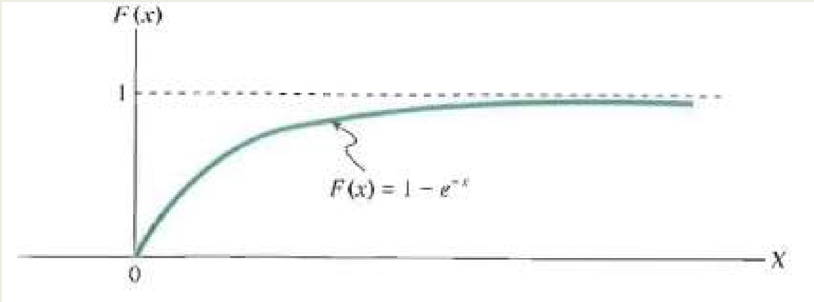
(sol) $f(x)$가 $x \geq 0$인 경우 0이 아닌 확률을 가지므로 $F(x)$는 다음과 같이 구할 수 있다.

(a) $x < 0$인 경우 : $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} 0\,dt = 0$

(b) $x \geq 0$인 경우 : $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} f(t)\,dt = \int_{0}^{x} e^{-t}\,dt = 1 - \frac{1}{e}$



(a) X의 분포함수



(b) X의 누적분포함수

끝~~ ♥♥