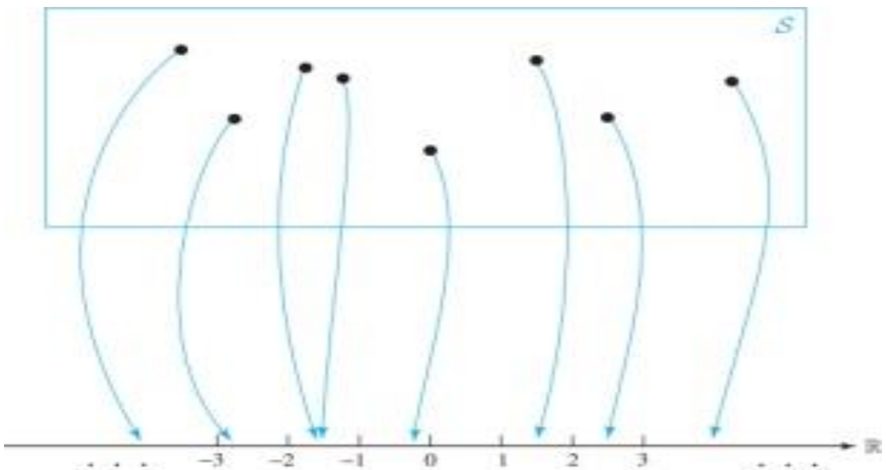
2장 : 확률변수

2 확률변수

2.1. 이산형 확률변수 (Discrete random variables)

- 확률변수는 특정실험의 표본공간에 속하는 각각의 결과에 실수치를 부여함에 의해 생성할 수 있다.
- 확률변수는 함수이다.

$$X : S \Rightarrow \mathbb{R}$$



2 확률변수

2.1. 이산형 확률변수 (Discrete random variables)

- **확률변수가 셀 수 있는** (countable) **집합** (**기계고장의 수, 인터뷰 지원자의 수, 자동차 사고수**, etc.)**으로부터 값을 취한다고 할 때 이산형 확률변수라고 한다.**

- **함수** $p(x)$ **는 이산형 확률변수** X**에 대한 확률질량함수** **(probability mass function: pmf)라고 하며 다음을 만족한다.**

  1. $Pr(X = x) = p(x) \quad x = 0, 1, 2, \ldots$
  2. $p(x) \geq 0$
  3. $\sum_x p(x) = 1$

- **누적분포함수** **(cumulative distribution function: cdf)**

$$P(x) = P(X \leq x).$$

2 확률변수

2.1. 이산형 확률변수 (Discrete random variables)

- **예제 2.1** : 두개의 주사위를 던져 확률변수 X 를 나오는 주사위 수의 합으로 정의할때

$$X = 2, 3, \ldots, 11, 12.$$

$$p(2) = Pr(X=2) = \tfrac{1}{36}, \qquad p(3) = Pr(X=3) = \tfrac{2}{36},$$
$$p(4) = Pr(X=4) = \tfrac{3}{36}, \cdots \qquad p(7) = Pr(X=7) = \tfrac{6}{36},$$
$$p(8) = Pr(X=8) = \tfrac{5}{36}, \cdots \qquad p(12) = Pr(X=12) = \tfrac{1}{36}.$$

- $\sum_{x=2}^{12} p(x) = 1 \qquad \text{cdf: } P(6) = Pr(2 \le X \le 6) = \frac{15}{36}$

2.2. 연속형 확률변수 (Continuous Random Variables)

- **연속형 확률변수 (Continuous random variables)**는 연속적인 영역에서 값을 취한다. 예를 들면 전지고장시간은 $[0, \infty)$ 에서 값을 취하며, 콘크리트 판 파괴강도는 $[120, 150]$ 에서 값을 취한다.

- 함수 $f(x)$ 는 다음을 만족할 때 연속형 확률변수 X 의 확률밀도함수 **(probability density function: pdf)** 라 하며 다음을 만족한다.
  1. $f(x) \geq 0, \text{ for all } x \in \mathbb{R}$
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$
  3. $Pr(a < X < b) = \int_a^b f(x)dx = F(x)\Big|_a^b = F(b) - F(a)$
  4. $Pr(X = a) = \int_a^a f(x)\,dx = 0.$

- pdf $f(x)$ 를 가지는 확률변수 X 에 대한 **누적분포함수 (Cumulative distribution function: cdf)**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \qquad \text{for } -\infty < x < \infty.$$

- $f(x) = \frac{dF(x)}{dx}$

2.2. 연속형 확률변수 (Continuous Random Variables)

예제 2.2 : 전지고장시간

$$f(x) = \frac{2}{(x+1)^3}, \qquad x \geq 0$$

1. $\int_0^\infty f(x)\,dx = \int_0^\infty \frac{2}{(x+1)^3}\,dx = -\frac{1}{(x+1)^2}\Big|_0^\infty = 0 - (-1) = 1.$
2. $F(x) = 1 - \frac{1}{(x+1)^2}, \qquad x \geq 0.$
3. $F(5) = Pr(0 \leq X \leq 5) = 1 - \frac{1}{36} = \frac{35}{36}.$

2.3. 확률변수의 기대값 (Expectation)

- **기대값 (expectation)은 확률변수의 평균을 나타냄.**

- **pmf $P(X = x_i) = p_i$ 를 가지는 이산형 확률변수의 기대값 (expectation or mean)**

$$\mu = E(X) = \sum_i p_i x_i .$$

- **pdf $f(x)$ 를 가지는 연속형 확률변수의 기대값 (expectation or mean)**

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx .$$

- **함수 $g(x)$ 의 기대값**

이산형 : $$E(g(X)) = \sum_i p_i g(x_i)$$

연속형 : $$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- $e.g.) E[g(X)] \neq g(E[x])$

2.3. 확률변수의 기대값 (Expectation)

예제 2.3 : 두개의 주사위를 던져 나오는 눈의 합을 확률변수 X 라고 할때

$$E[X] = \sum_{x=2}^{12} Pr(X = x) \cdot x$$

$$= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7$$

예제 2.4 : 전지고장시간

$$E[X] = \int_0^\infty x \cdot \frac{2}{(x+1)^3}\, dx = \int_0^\infty \left(\frac{2}{(x+1)^2} - \frac{2}{(x+1)^3} \right)$$

$$= \left(-\frac{2}{x+1} + \frac{1}{(x+1)^2} \right) \Big|_0^\infty = 0 - (-1) = 1.$$

2.3. 확률변수의 기대값 (Expectation)

- Cdf $F(x)$ 를 가지는 연속형 확률변수 X 의 중앙값 (Medians)은 다음을 만족하는 x 의 값에 해당.

$$F(x) = P(X \leq x) = 0.5.$$

- 좌우대칭 확률변수 (Symmetric random variables)

  - 연속형 확률변수 X 가 pdf $f(x)$를 가진다면 특정 값 α에 대하여

$$f(\alpha + x) = f(\alpha - x) \qquad x \in \mathbb{R}$$

즉, X 는 α에 대하여 좌우대칭이다.

$\mu = E(X) = \alpha$는 중앙값

2.3. 확률변수의 기대값 (Expectation)

- **예제 2.5** : 전지고장시간

$$F(x) = 1 - \frac{1}{(x+1)^2} = 0.5$$

$$\text{Median } x_{.5} = \sqrt{2} - 1 = 0.414 \approx 25 \text{ min.}$$

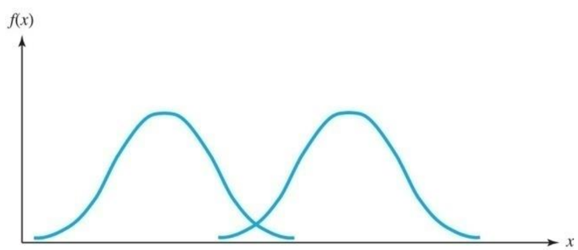
- 기대값 및 중앙값은 우리에게 유용한 information을 제공.

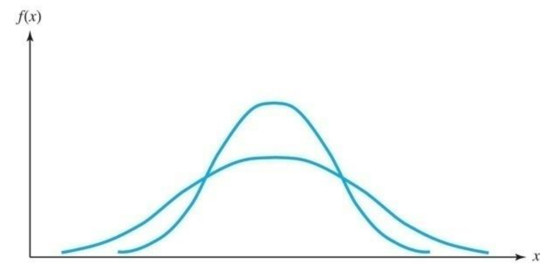
즉 MP3에 위의 battery를 사용할 때 24시간 사용하기 위해서는 평균 24개의 battery가 필요.

또한, 우리가 위의 battery 6개를 6개의 MP3에 사용할 때 그 6개의 MP3중 3개도 30분 이상 사용할 수 없다는 의미.

2.4. 확률변수의 분산 (Variance)

> **분산 (Variance)은 확률변수에 대한 흩어짐 혹은 변동성을 측정한다.**





평균값 다름

but! 평균값에 대하여 밀도함수의 형태 혹은 산포 정도가 같음

따라서 분산은 같음

평균값 같음

but! 분산이 다름

평평하고 퍼져있는 밀도함수의 분산이 큼

평균 또는 기대값 : μ σ

분산 : σ^2

분산의 제곱근(분포의 표준편차) : σ

※ 확률변수 x가 초단위라면 표준편차는 '초단위'이며, 분산은 '초단위의 제곱'

2.4. 확률변수의 분산 (Variance)

- **확률변수 X 에 대한 분산**

$$\sigma^2 = Var(X) = E[(X - E(X))^2]$$
$$= E[X^2 - 2XE[X] + (E[x])^2] = E(X^2) - (E(X))^2.$$

$$Var(X) = \sum_i p_i(x_i - E[X])^2 = \sum_i p_i x_i^2 - \left(\sum_i p_i x_i\right)^2$$ **이산형인 경우**

$$= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)\, dx\right)^2$$ **연속형인 경우**

- **확률변수 X 의 표준편차 (Standard deviation) :** $\sigma = \sqrt{Var(X)}.$

2.4. 확률변수의 분산 (Variance)

예제 2.6 : 주사위를 2개 던져 나오는 눈의 수의 합을 확률변수 X 라고 할때

$$\sigma^2 = Var(X) = \sum_{x=2}^{12} p(x)(x - E[X])^2$$

$$= \frac{1}{36}(2-7)^2 + \frac{2}{36}(3-7)^2 + \cdots + \frac{2}{36}(11-7)^2 + \frac{1}{36}(12-7)^2 = \frac{210}{36}.$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\frac{210}{36}}.$$

2.4. 확률변수의 분산 (Variance)

예제 2.7 : 하나의 로트(Lot)에서 품질검사자는 7개의 표본을 추출한다. 그 로트가 4개의 양품과 3개의 불량품을 포함하고 있다고 한다. 만약 3개의 표본을 추출하여 표본내의 양품의 개수를 확률변수 X 라고 할때, X 의 pmf는

$$p(x) = Pr(X = x) = \frac{\begin{pmatrix} 4 \\ x \end{pmatrix} \times \begin{pmatrix} 3 \\ 3-x \end{pmatrix}}{\begin{pmatrix} 7 \\ 3 \end{pmatrix}}, \qquad x = 0, 1, 2, 3.$$

$$p(0) = Pr(X = 0) = \tfrac{1}{35}, \qquad p(1) = Pr(X = 1) = \tfrac{12}{35}$$

$$p(2) = Pr(X = 2) = \tfrac{18}{35}, \qquad p(3) = Pr(X = 3) = \tfrac{4}{35}$$

$$P(x) = \begin{cases} 0 & x < 0 \\ 1/35 & 0 \le x < 1 \\ 13/35 & 1 \le x < 2 \\ 31/35 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

2 확률변수

2.4. 확률변수의 분산 (Variance)

$$\mu = E[X] = 0 \cdot \frac{1}{35} + 1 \cdot \frac{12}{35} + 2 \cdot \frac{18}{35} + 3 \cdot \frac{4}{35} = \frac{12}{7}.$$

$$E[X^2] = 0^2 \cdot \frac{1}{35} + 1^2 \cdot \frac{12}{35} + 2^2 \cdot \frac{18}{35} + 3^2 \cdot \frac{4}{35} = \frac{24}{7}.$$

$$Var(X) = E[X^2] - (E[X])^2 = \frac{24}{7} - \frac{144}{49} = \frac{24}{49}.$$

다른 해법 : 위의 pmf**를 가지는 분포는 모수** $N = 7, n = 3, r = 4$ **인 초기하 분포** (hypergeometric distribution)**이므로**

$$\mu = E[X] = \frac{n \times r}{N} = \frac{3 \times 4}{7} = \frac{12}{7}.$$

2.4. 확률변수의 분산 (Variance)

예제 2.8 : 한 지역의 편의점에서 코카콜라에 대한 주당 수요(단위: 1000리터)에 대한 확률변수 X는 다음의 pdf를 가진다.

$$f(x) = \begin{cases} 2(x-1) & 1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

1. $f(x) \geq 0$
2. $\int_1^2 2(x-1)\,dx = 2 \cdot \left(\frac{x^2}{2} - x \right) \Big|_1^2 = 1.$
3. $\mu = E[X] = 2\int_1^2 x(x-1)\,dx = 2 \cdot \left(\frac{x^3}{3} - \frac{x^2}{2} \right) \Big|_1^2 = \frac{5}{3}.$
4. $E[X^2] = 2\int_1^2 x^2(x-1)\,dx = \frac{17}{6}.$
5. $Var(X) = E[X^2] - \mu^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$

2.4. 확률변수의 분산 (Variance)

체비셰프 부등식 (Chebyshev's Inequality)

- Theorem 2.1 확률변수 X 가 평균 μ 분산 σ^2을 가진다고 할 때

$$P(\mu - c\sigma \le X \le \mu + c\sigma) \ge 1 - \frac{1}{c^2} \qquad \text{for } c \ge 1.$$

- PROPOSTION 1. (Markov 부등식) $u(X)$를 확률변수 X 의 비음수 함수라고 하자. 만약 $E[u(X)]$가 존재한다면 양의 상수 c에 대하여

$$Pr(u(X) \ge c) \le \frac{E[u(X)]}{c}.$$

- 증명. $A = \{x : u(x) \ge c\}.$

$$\begin{aligned} E[u(X)] &= \int_{-\infty}^{\infty} u(x)f(x)\,dx = \int_A u(x)f(x)\,dx + \int_{\bar{A}} u(x)f(x)\,dx \\ &\ge \int_A u(x)f(x)\,dx \\ &\ge c\int_A f(x)\,dx = Pr(X \in A) = Pr(u(X) \ge c). \end{aligned}$$

2.4. 확률변수의 분산 (Variance)

- **Chebyshev's Inequality** (Cont'd)

  - Chevyshev's Inequality의 증명

$$Pr\left((X-\mu)^2 \le c^2\sigma^2\right) \le \frac{E[(X-\mu)^2]}{c^2\sigma^2} = \frac{1}{c^2}$$

$$Pr[|X-\mu| \ge c\sigma] \le \frac{1}{c^2}$$

$$Pr[|X-\mu| \le c\sigma] \ge 1-\frac{1}{c^2}.$$

- $Pr[\mu-2\sigma \le X \le \mu+2\sigma] \ge 1-\frac{1}{2^2} = 0.75$

- $Pr[\mu-3\sigma \le X \le \mu+3\sigma] \ge 1-\frac{1}{3^2} = 0.89.$

- **정확한 분포를 모르더라도 평균과 분산을 구할 수 있으면 평균값에 대해 몇 sigma 안에 들어있을 지에 대한 확률값을 구할 수 있다.**

2.4. 확률변수의 분산 (Variance)

- cdf $F(x)$ 를 가지는 확률변수 X 의 p **분위수 (Quantile)** 혹은 $p \times 100th$ **백분위수 (Percentile)**는

$$F(x) = p$$

를 만족하는 x 의 값으로서 0.5 Quantile (or 50 percentile)은 **중앙값(median)**.

- **상한 사분위수 (Upper quartile)** : 분포의 75 백분위수

  하한 사분위수 (Lower quartile) : 분포의 25 백분위수

  사분위범위 (Interquartile range)는 상한 사분위수와 하한 사분위수의 거리

2.4. 확률변수의 분산 (Variance)

예제 2.9 : **예제** 2.8 (**계속**)

$$F(x) = \int_1^x 2(t-1)\,dt = (t^2 - 2t)\Big|_1^x = (x-1)^2, \qquad 1 < x < 2$$

$$\text{upper quartile: } F(x) = (x-1)^2 = 0.75, \qquad x_{.75} = 1.87.$$

$$\text{lower quartile: } F(x) = (x-1)^2 = 0.25, \qquad x_{.25} = 1.5.$$

$$\text{interquartile range: } x_{.75} - x_{.25} = 1.87 - 1.5 = 0.37.$$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

- 두 확률변수 X 와 Y 의 **결합확률분포 (Joint probability distribution)**

$$p(x,y) = P(X=x, Y=y)$$

이며 연속형의 경우 $f(x,y) = P(X=x, Y=y)$

- 두 확률변수 X 와 Y 의 **결합누적분포함수(Joint cumulative distribution function)**

$$P(x,y) = \sum_{i:x_i \le x} \sum_{j:y_j \le y} p(x,y)$$ 이산형인 경우

$$F(x,y) = \int_{w=-\infty}^{x} \int_{z=-\infty}^{y} f(w,z)\, dz\, dw$$ 연속형인 경우

- **주변확률분포 (Marginal Probability Distribution)**

$$p_X(x) = \sum_y p(x,y), \qquad p_Y(y) = \sum_x p(x,y)$$ 이산형인 경우

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$ 연속형인 경우

2.5. 결합확률변수 (Jointly Distributed Random Variables)

예제 2.10 : 항아리에 1부터 6까지 숫자가 적인 공이 들어있다고 하자. 항아리에서 비복원추출(without replacement)로 3개의 공을 임의로 선택한다고 할때

X=maximum

Sol) 총 경우의 수는 $P_3^6 = 120$. X=maximum of the three numbers selected,
Y= minimum of the three numbers selected.

x	3	4	5	6
$p(x)$	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{10}{20}$

y	1	2	3	4
$p(y)$	$\frac{10}{20}$	$\frac{6}{20}$	$\frac{3}{20}$	$\frac{1}{20}$

	1	2	3	4	$p_X(x)$
3	$\frac{1}{20}$	0	0	0	$\frac{1}{20}$
4	$\frac{2}{20}$	$\frac{1}{20}$	0	0	$\frac{3}{20}$
5	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	0	$\frac{6}{20}$
6	$\frac{4}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{10}{20}$
$f_Y(y)$	$\frac{10}{20}$	$\frac{6}{20}$	$\frac{3}{20}$	$\frac{1}{20}$	$\sum = 1$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

예제 2.11 : 연속형인 경우

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

1. $\int_0^1 \int_0^1 f(x,y)\,dx\,dy = \int_0^1 \frac{2}{5}\left(x^2+3xy\right)\Big|_0^1 dy$

$$= \int_0^1 \left(\frac{2}{5}+\frac{6}{5}y\right)\,dy = \frac{2}{5}y+\frac{3}{5}y^2\Big|_0^1 = \frac{2}{5}+\frac{3}{5} = 1.$$

2. $f_X(x) = \int_0^1 \frac{2}{5}(2x+3y)\,dy = \left(\frac{4}{5}xy+\frac{3}{5}y^2\right)\Big|_0^1 = \frac{4}{5}x+\frac{3}{5}.$

3. $f_Y(y) = \int_0^1 \frac{2}{5}(2x+3y)\,dx = \left(\frac{2}{5}x^2+\frac{6}{5}xy\right)\Big|_0^1 = \frac{6}{5}y+\frac{2}{5}.$

4. $Pr\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) = \int_{1/4}^{1/2} \int_0^{1/2} \frac{2}{5}(2x+3y)\,dx\,dy$

$$= \int_{1/4}^{1/2} \frac{2}{5}\left(x^2+3xy\right)\Big|_0^{1/2} dy = \int_{1/4}^{1/2} \left(\frac{1}{10}+\frac{3}{5}y\right)\,dy$$

$$= \left(\frac{1}{10}y+\frac{3}{10}y^2\right)\Big|_{1/4}^{1/2} = \frac{13}{160}.$$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

- 주어진 특정함수 $u(x,y)$에 대하여

$$E[u(x,y)] = \begin{cases} \sum_{i:x_i \le x} \sum_{j:y_j \le y} u(x,y)p(x,y) & \text{이산형인 경우} \\ \int_{w=-\infty}^{x} \int_{z=-\infty}^{y} u(w,z) f(w,z)\, dz\, dw & \text{연속형인 경우} \end{cases}$$

- 확률변수 X 가 주어졌다는 조건하에 Y 의 **조건부분포 (Conditional Distribution)**

$$p_{X|Y=y} = P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p(x,y)}{p_Y(y)}$$

$$p_{Y|X=x} = \frac{p(x,y)}{p_X(x)}$$ 이산형인 경우

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)}$$ 연속형인 경우

2.5. 결합확률변수 (Jointly Distributed Random Variables)

예제 2.12 : **예제** 2.10 (**계속**)

y	1	2	3
$p_Y(y\|X=5)$	$\frac{3}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

– $E[Y|X=5] = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{5}{3}.$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

- 이산형 확률변수 X 와 Y에 대하여

$$Pr(a \le X \le b|Y = y) = \sum_{x=a}^{b} p(x|y).$$

- 두 확률변수 X 와 Y의 **독립 (Independence)**

한 확률변수가 또다른 하나의 확률변수가 취하는 값에 의존하지 않을 때 확률변수 X와 Y는 **독립**이라고 한다.

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \Rightarrow \quad f(x,y) = f_Y(y) \cdot f_X(x|y) = f_Y(y) \cdot f_X(x)$$

$p_{x,y} = p_X(x)p_Y(y)$ **이산형인 경우**

$f_{x,y} = f_X(x)f_Y(y)$ **연속형인 경우**

2.5. 결합확률변수 (Jointly Distributed Random Variables)

예제 2.13 :

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4} & 0 \le x \le 2, 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

1. $f_X(x) = \int_0^1 \frac{x}{4}\left(1+3y^2\right)\, dy = \frac{x}{4}\left(y+y^3\right)\Big|_0^1 = \frac{x}{2}, \qquad 0 < x < 2.$

2. $f_Y(y) = \int_0^2 \frac{x}{4}\left(1+3y^2\right)\, dx = \frac{x^2}{8}\left(1+3y^2\right)\Big|_0^2 = \frac{1}{2}\left(1+3y^2\right) \qquad 0 < y < 1.$

3. $f_X(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2} = f_X(x) \Longleftrightarrow f_X(x|y) = f_X(x) \cdot f_Y(y).$

4. $Pr\left(\frac{1}{4} < X < \frac{1}{2} | Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2}\, dx = \frac{3}{64}.$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

- 두 확률변수 X 와 Y 의 **공분산 (Covariance)**

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

- 만약 X 와 Y 독립이면 $E[XY] = E[X] \cdot E[Y]$

즉, $Cov(X,Y) = 0.$

- **공분산의 특성**

1. $Cov(X,X) = var(X)$
2. $Cov(X,Y) = Cov(Y,X)$
3. $Cov(cX,Y) = cCov(X,Y)$
4. $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$

2.5. 결합확률변수 (Jointly Distributed Random Variables)

5. $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} cov(X_i, Y_j).$

6. $Var\left(\sum_{i=1}^{n} X_i\right) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} X_j\right)$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Cov(X_i, X_i) + \sum_{i=1}^{n} \sum_{j \neq i} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i=1}^{n} \sum_{j<i} Cov(X_i, X_j).$$

- **두 확률변수 X 와 Y 의 상관 (Correlation)**

$$-1 \leq \rho_{XY} = Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \leq 1$$

2.6. 확률변수의 조합 및 함수 (Combinations and Functions of Random Variables)

- $Y = aX + b$ 이며 $a, b \in \mathbb{R}$ 일때 $E[Y] = aE[X] + b$ 이며
  $Var(Y) = a^2 Var(X).$

- $E[X_1 + X_2] = E[X_1] + E[X_2],$
  $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2).$

- 만약 X_1 과 X_2 가 독립이라면
  $Var(X_1 + X2) = Var(X_1) + Var(X_2).$

- $E[a_1 X_1 + \cdots + a_n X_n + b] = a_1 E[X_1] + \cdots + a_n E[X_n] + b.$

- 확률변수 $X_1, X_2, \ldots, X_n$ 가 독립일때
  $Var(a_1 X_1 + \cdots + a_n X_n + b) = a_1^2 Var(X_1) + \cdots + a_n^2 Var(X_n).$

2.6. 확률변수의 조합 및 함수 (Combinations and Functions of Random Variables)

- **PROPOSTION 1**. **확률변수** $X_1, \ldots, X_n$ **가 기대값** μ **이며 분산** σ^2**을 가지는 독립이며 동일한 분포**(independent and identically distributed (iid))**를 따른다면 산술평균** $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ **에 대하여**

$$(a)\ \ E[\bar{X}] = \mu.$$

$$(b)\ \ Var(\bar{X}) = \sigma^2/n.$$

$$(c)\ \ Cov(\bar{X}, X_i - \bar{X}) = 0.$$

- **(c)의 증명** :

$$\begin{aligned} Cov(\bar{X}, X_i - \bar{X}) &= Cov(\bar{X}, X_i) - Cov(\bar{X}, \bar{X}) \\ &= \tfrac{1}{n} Cov\left(X_i + \textstyle\sum_{j \neq i} X_j, X_i\right) - Var(\bar{X}) \\ &= \tfrac{1}{n} Cov(X_i, X_i) + \tfrac{1}{n} Cov\left(\textstyle\sum_{j \neq i} X_j, X_i\right) - \tfrac{\sigma^2}{n} = 0 \end{aligned}$$

2.6. 확률변수의 조합 및 함수 (Combinations and Functions of Random Variables)

확률변수의 비선형 함수 (Nonlinear function of a random variable)

- 확률변수 X 의 비선형 함수 $Y = g(X)$는 또다른 확률변수이며 비선형 함수는 다음과 같은 형태를 포함한다 : $Y = X^2,\ Y = \sqrt{X},\ Y = e^X.$

- 확률변수 X 가 0과 1사이에 균일분포하게 분포한다고 할때 pdf는

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

- 확률변수 Y 의 비선형함수가 $Y = e^X$ 일때

$$\begin{aligned} F_Y(y) &= Pr(Y \le y) \\ &= Pr(e^X \le y) = Pr(X \le \log(y)) = F_X(\log(y)) = \log(y) \end{aligned}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{y}. \quad for \quad 1 \le y \le 2.718.$$

$$E[Y] = \int_1^{2.718} y f_Y(y)\, dy = \int_1^{2.718} y \cdot \frac{1}{y}\, dy = 1.718.$$

- $E[Y] = E[g(X)] = E[e^X] \ne g(E[X]) = e^{E[X]} = e^{0.5} = 1.649.$