2.1 The Distribution of two random variables.

〈Ex. 2.1.1〉

Imagine a box containing 2 white balls, 3 black balls, 4 red balls. Let X and Y denote the number of white and black balls respectively among the 3 balls chosen blindedly.

$$P(X=x, Y=y) = \frac{\binom{2}{x}\binom{3}{y}\binom{4}{3-x-y}}{\binom{9}{3}}, \quad \begin{matrix} x=0,1,2, & y=0,1,2,3, \\ x+y \le 3 \end{matrix}$$

The table below denotes the values which X and Y can take and their respective probabilities.

		Value of Y				Sum
		Y=0	Y=1	Y=2	Y=3	
Value of X	X=0	4/84	18/84	12/84	1/84	35/84
	X=1	12/84	24/84	6/84	0	42/84
	X=2	4/84	3/84	0	0	7/84
Sum		20/84	45/84	18/84	1/84	1

표 2.1.1

Let $\{(x_j, y_k) \mid j=1,2,\cdots, k=1,2,\cdots\}$ be the set of all possible combinations of two ordered pair random variables X and Y.

The function that maps the pair (x_j, y_k) with its respective probability is called the pdf of "bivariate random vector".

$$f(x_j, y_k) = P(X=x_j, Y=y_k)$$

[이산형 이변량 확률벡터의 확률밀도함수]:

(a) $f(x,y) \ge 0 \quad \forall x,y : -\infty < x < +\infty, \ -\infty < y < +\infty$
$f(x,y) = 0 \quad \forall (x,y) \ne (x_j, y_k) \quad (j=1,2,\cdots, k=1,2,\cdots)$

(b) $\sum_x \sum_y f(x,y) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} f(x_j, y_k) = 1$

(c) $\sum_{x:\, a \le x \le b} \ \sum_{y:\, c \le y \le d} f(x,y) = P(a \le X \le b, c \le Y \le d)$

[연속형 이차원 확률변수의 확률밀도함수]:

(a) $f(x,y) \ge 0 \quad \forall x,y : -\infty \le x \le +\infty, \ -\infty \le y \le +\infty$

(b) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) = 1$

(c) $\int_a^b \int_c^d f(x,y)\,dy\,dx = P(a \le X \le b, c \le Y \le d) \quad (a<b, c<d)$

* (a) 이중적분은 미분순인 $dy\,dx$가 나타내는 순서대로 적분해나가는 반복적분(iterated integral)을 뜻한다. 즉 $\int_a^b \int_c^d f(x,y)\,dx\,dy = \int_a^b \left\{ \int_c^d f(x,y)\,dy \right\} dx$

〈Ex. 2.1.2〉 Find the value of C for the following function to be a probability distribution function. Find $P(X \ge 2, Y \ge 3)$ when the function below is the probability distribution of (X,Y).

$$f(x,y) = c e^{-x-y} I_{(0 \le x \le y)}$$

$$\int_0^{\infty} \int_0^{y} c e^{-x-y}\,dx\,dy = c \int_0^{\infty} e^{-y} \int_0^{y} e^{-x}\,dx\,dy$$

$$= c \int_0^{\infty} e^{-y} \left[-e^{-x} \right]_0^{y} dy$$

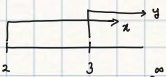
$$= c \int_0^{\infty} e^{-y}(1-e^{-y})\,dy$$

$$= c \int_0^{\infty} e^{-y} - e^{-2y}\,dy$$

$$= c \left[-e^{-y} + \frac{1}{2} e^{-2y} \right]_0^{\infty}$$

$$= \frac{1}{2} c = 1 \qquad \therefore c = 2$$

$$P(X \ge 2, Y \ge 3) = P(3 > X \ge 2, Y \ge 3) + P(X \ge 3, Y \ge 3)$$



$$= \underbrace{2 \int_3^{\infty} \int_2^{3} e^{-x} e^{-y}\,dx\,dy}_{①} + \underbrace{2 \int_3^{\infty} \int_x^{\infty} e^{-x} \cdot e^{-y}\,dy\,dx}_{②}$$

$$① \quad 2 \int_3^{\infty} \left[-e^{-x} \right]_2^3 e^{-y}\,dy = 2 \int_3^{\infty} (e^{-2} - e^{-3}) e^{-y}\,dy$$

$$= 2(e^{-2} - e^{-3}) \left[-e^{-y} \right]_3^{\infty}$$

$$= 2(e^{-2} - e^{-3})(e^{-3})$$

$$= 2(e^{-5} - e^{-6})$$

$$② \quad 2 \int_3^{\infty} \int_x^{\infty} e^{-x} \cdot e^{-y}\,dy\,dx = 2 \int_3^{\infty} e^{-x} \left[-e^{-y} \right]_x^{\infty} dx = 2 \int_3^{\infty} e^{-x}(e^{-x})\,dx$$

$$= 2 \int_3^{\infty} e^{-2x}\,dx = 2 \left[-\frac{1}{2} e^{-2x} \right]_3^{\infty}$$

$$= \left[-e^{-2x} \right]_3^{\infty} = e^{-6}$$

$$① + ② = 2(e^{-5} - e^{-6}) + e^{-6} = 2e^{-5} - e^{-6}$$ ✓✓

[Textbook Solution]

$$P(X \ge 2, X \ge 3) = \int_2^{\infty} \int_3^{\infty} 2e^{-x-y} I_{(0 \le x \le y)}\,dy\,dx$$

그런데 $y \ge 3, y \ge x$을 간략히 $y \ge 3 \vee x$[20]로 나타내면

$$\int_3^{+\infty} 2e^{-x-y} I_{(0 \le x \le y)}\,dy \le \int_{3 \vee x}^{\infty} 2e^{-x-y}\,dy = 2e^{-x} \left[-e^{-y} \right]_{3 \vee x}^{+\infty} = 2e^{-x-3 \vee x}$$

$$\therefore P(X \ge 2, Y \ge 3) = \int_2^{\infty} 2e^{-x-3 \vee x}\,dx = \int_2^3 2e^{-x-3}\,dx + \int_3^{+\infty} 2e^{-x-x}\,dx$$

$$\therefore P(X \ge 2, Y \ge 3) = 2e^{-3} \left[-e^{-x} \right]_2^3 + \left[-e^{-2x} \right]_3^{+\infty} = 2e^{-5} - e^{-6} \fallingdotseq 0.011$$

* Footnote: $a \vee b$ 는 a와 b 중 큰 수를 나타내는 기호이고, $a \wedge b$ 는 a와 b 중 작은 수를 나타내는 기호이다.

⟨Definition: Joint · Marginal Probability Distribution⟩

Joint pdf : (X,Y) distribution

Marginal pdf : (X) / (Y) distribution

\# How to get marginal pdf from joint pdf

$$P(a \le X \le b) = P(a \le X \le b, -\infty \le y \le +\infty)$$

$$= \begin{cases} \sum_{x:\, a \le x \le b} \sum_{y} f(x,y) & (X,Y) \text{ are discrete} \\ \int_a^b \int_{-\infty}^{+\infty} f(x,y)\, dx\, dy & (X,Y) \text{ are continuous} \end{cases}$$

Therefore, for $f_1(x)$:

$$f_1(x) = \begin{cases} \sum_{y} f(x,y) & (X,Y) \text{ are discrete} \\ \int_{-\infty}^{+\infty} f(x,y)\, dy & (X,Y) \text{ are continuous} \end{cases}$$

$$P(a \le X \le b) = \begin{cases} \sum_{x:\, a \le x \le b} f_1(x) & (X,Y) \text{ are discrete} \\ \int_a^b f_1(x)\, dx & (X,Y) \text{ are continuous} \end{cases}$$

We can observe that the probability of X is given as the integral of $f_1(x)$. Meaning, $f_1(x)$ is the pdf of X.

이차원 확률변수 (bivariate r.v.) 의 결합확률밀도함수 (joint pdf) 가 $f(x,y)$ 일 때, X 와 Y 의 주변확률밀도함수 $f_1(x)$ 와 $f_2(y)$ 는 각각 다음과 같이 주어진다.

$$f_1(x) = \begin{cases} \sum_{y} f(x,y) \\ \int_{-\infty}^{\infty} f(x,y)\, dy \end{cases} \qquad f_2(y) = \begin{cases} \sum_{x} f(x,y) & (X,Y) \text{ are discrete} \\ \int_{-\infty}^{\infty} f(x,y)\, dx & (X,Y) \text{ are continuous} \end{cases}$$

⟨Ex. 2.1.3⟩

⟨예 2.1.1⟩ 에서 (X,Y) 의 결합확률밀도함수는

$$f(x,y) = \frac{\binom{2}{x}\binom{3}{y}\binom{4}{3-x-y}}{\binom{9}{3}}, \quad x = 0,1,2,\ y = 0,1,2,3,\ x+y \le 3$$

이고, X 와 Y 의 주변확률밀도함수는 각각 다음과 같이 주어진다.[21] 이 결과는 ⟨표 2.1.1⟩ 의 세로와 가로 각각의 주변에 주어진 X 와 Y 의 주변확률분포와 일치하는 것을 알 수 있다.

$$f_1(x) = \sum_{y} f(x,y) = \binom{2}{x} \sum_{y=0}^{3-x} \binom{3}{y}\binom{4}{3-x-y} \Big/ \binom{9}{3} = \binom{2}{x}\binom{7}{3-x} \Big/ \binom{9}{3}, \quad x = 0,1,2$$

$$f_2(x) = \sum_{x} f(x,y) = \binom{3}{y} \sum_{x=0}^{2 \wedge (3-y)} \binom{2}{x}\binom{4}{3-y-x} \Big/ \binom{9}{3} = \binom{3}{y}\binom{6}{3-y}, \quad y = 0,1,2,3$$

★ 21) $m \ge l$, $n \ge l$ 일 때 $\sum_{y=0}^{l} \binom{m}{l}\binom{n}{l-y} = \binom{m+n}{l}$ 은

$(1+t)^m (1+t)^n = (1+t)^{m+n}$ 에서 t^l 의 계수이다.

$$(1+t)^{m+n} = \sum_{y=0}^{m+n} \binom{m+n}{y} t^y \cdot t^{(m+n)-y} = 1 + \cdots + \underbrace{\binom{m+n}{l} t^l}_{*} + \cdots + t^{m+n}$$

$$(1+t)^m = \sum \binom{m}{y} t^{m-y}$$

$$(1+t)^n = \sum \binom{n}{x} t^{n-x}$$

$$(1+t)^m (1+t)^n = \Big(\sum_{y} t^{m-y}\Big)\Big(\sum_{x} t^{n-x}\Big) = \Big(1 + mt + \binom{m}{2}t^2 + \cdots\Big)\Big(1 + nt + \binom{n}{2}t^2 \cdots\Big)$$

$$= \Big(\cdots + \underbrace{\sum_{z=0}^{l} \binom{m}{z}\binom{n}{l-z} t^l}_{**} + \cdots\Big)$$

$* = **$ $\therefore \sum_{z=0}^{l} \binom{m}{z}\binom{n}{l-z} = \binom{m+n}{l}$

★★ $\sum_{y=0}^{l} \binom{m}{l}\binom{n}{l-y} = \binom{m+n}{l}$

⟨Ex. 2.1.4⟩

From ⟨Ex. 2.1.2⟩, the joint pdf of (X,Y) is

$$f(x,y) = 2e^{-x-y} I_{(0 \le x \le y)}$$

The marginal pdfs are given as below

$$f(x) = \int_x^{\infty} 2e^{-x-y}\, dy = \Big[-2e^{-x-y}\Big]_x^{\infty} = 2e^{-2x} I_{(0 \le x)}$$

$$f(y) = \int_0^{y} 2e^{-x-y}\, dx = \Big[-2e^{-x-y}\Big]_0^{y} = 2e^{-y} - 2e^{-2y} = 2e^{-y}(1 - e^{-y}) I_{(0 \le y)}$$

⟨Definition: Joint cdf / pdf⟩

(a) (discrete)

If X and Y can have the values $x_1 < x_2 < \cdots < x_m < \cdots$, and $y_1 < y_2 < \cdots < y_n < \cdots$ respectively,

$$F(x_m, y_n) = P(X \le x_m, Y \le y_n) = \sum_{j=1}^{m} \sum_{k=1}^{n} f(x_j, y_k)$$

$$f(x_j, y_k) = P(X = x_j, Y = y_k)$$

$$= \{F(x_j, y_k) - F(x_{j-1}, y_k)\} - \{F(x_j, y_{k-1}) - F(x_{j-1}, y_{k-1})\}$$

(b) (continuous)

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(t,u)\,dt\,du$$

f 가 연속인 점 (x,y) 에서 $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$

〈Ex 2.1.5〉

In 〈Ex. 2.1.2〉 the joint cdf of (X,Y) is

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} 2e^{-t-u} I_{(0\le t\le u)}\,du\,dt$$

However,

$$\int_{-\infty}^{y} 2e^{-t-u} I_{(0\le t\le u)}\,du = 2e^{-t}\int_{t}^{y} e^{-u}\,du\, I_{(y\ge t\ge 0)} = 2e^{-t}(e^{-t}-e^{-y}) I_{(y\ge t\ge 0)}$$

$$\therefore F(x,y) = \int_{-\infty}^{x} 2e^{-t}(e^{-t}-e^{-y}) I_{(0\le t\le y)}\,dt$$

$$= \int_{0}^{x\wedge y} 2\left(e^{-2t} - e^{-t-y}\right) dt\, I_{(x\wedge y \ge 0)}$$

$$= \left\{1 - e^{-2x\wedge y} + 2e^{-y}\left(e^{-x\wedge y} - 1\right)\right\} I_{(x\ge 0,\, y\ge 0)}$$

(joint cdf)

(pdf)

확률밀도함수의 경우와 마찬가지로, X 와 Y의 결합누적분포함수와 각 확률변수의 누적분포함수를 구분하기 위하여 각 확률변수의 누적분포함수를 주변누적분포함수라고 한다.

r.v.s cdf → marginal cdf

즉, (cdf)

$$F_1(x) = P(X \le x),\quad F_2(x) = P(Y \le y)$$

를 각각 X의 주변누적분포함수, Y 의 주변누적분포함수라고 한다.

한편, 확률측도의 연속성으로부터

$$P(X \le x) = P(X \le x, Y \le +\infty) = \lim_{y\to+\infty} P(X \le x, Y \le y)$$

이므로, 다음과 같이 주변누적분포함수는 결합누적분포함수로부터 구할 수 있다.

$$F_1(x) = \lim_{y\to+\infty} F(x,y) \qquad F_2(y) = \lim_{x\to+\infty} F(x,y)$$

〈Ex. 2.1.6〉

〈예 2.1.5〉 에서 (X,Y)의 결합누적분포함수는

$$F(x,y) = \left\{1 - e^{-2x\wedge y} + 2e^{-y}\left(e^{-x\wedge y} - 1\right)\right\} I_{(x\ge 0, y\ge 0)}$$

따라서 X,Y의 주변누적분포함수는 각각 다음과 같이 구할 수 있다.

$$F_1(x) = \lim_{y\to+\infty}\left\{1 - e^{-2x\wedge y} + 2e^{-y}\left(e^{-x\wedge y} - 1\right)\right\} I_{(x\ge 0, y\ge 0)} = (1 - e^{-2x}) I_{(x\ge 0)}$$

$$F_2(x) = \lim_{x\to+\infty}\left\{1 - e^{-2x\wedge y} + 2e^{-y}\left(e^{-x\wedge y} - 1\right)\right\} I_{(x\ge 0, y\ge 0)} = (1 + e^{-2y} - 2e^{-y}) I_{(y\ge 0)}$$

($-2(x\wedge y)$)

이들을 미분하면 그 도함수가 〈예 2.1.4〉 에서 구한 주변확률밀도함수

$$f_1(x) = 2e^{-2x} I_{(x\ge 0)},\quad f_2(y) = 2e^{-y}(1 - e^{-y}) I_{(y\ge 0)}$$

와 연속인 점에서 일치하는 것을 알 수 있다.

〈Theorem 2.1.1 Properties of joint cumulative distribution function〉

(a) (increasing)

(i) $F(x,y) \le F(x+h, y)$, $F(x,y) \le F(x, y+k)$ $\forall h>0,\ k>0$

(ii) $F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \ge 0$

(b) (Total Volatility)

$$\lim_{x\to-\infty} F(x,b) = 0,\quad \lim_{y\to-\infty} F(a,y) = 0,\quad \lim_{\substack{x\to+\infty\\ y\to+\infty}} F(x,y) = 1.\quad \forall a,b$$

(c) (Right Continuity)

$$\lim_{\substack{h\to 0\\ h>0}} F(x+h,y) = F(x,y),\quad \lim_{\substack{k\to 0\\ k>0}} F(x,y+k) = F(x,y) \qquad \forall\, x,y$$

(proof)

(a)-(i), (b), (c) can be proved in a similar manner as the Single random variable case and thus the details will be skipped.

(a)-(ii) is clear from the following

$$P(x_1 < X < x_2,\ y_1 < Y < y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$$

As in the Single r.v. case, joint cdf's left limit is:

$$F(x^-, y) = \lim_{\substack{h\to 0\\ h>0}} F(x-h, y) = P(X < x, Y \le y)$$

Left limit may not be continuous.

Thus, the "jump" at $X = x$ from the left is:

$$P(X = x, Y \le y) = P(X \le x, Y \le y) - P(X < x, Y \le y)$$
$$= F(x,y) - F(x^-, y)$$

2.2. CHARACTERISTIC'S OF JOINT PROBABILITY DISTRIBUTIONS

Expectation of joint probability distribution

If the bivariate random variable (X,Y)'s pdf is $f(x,y)$, for a real value function $g(x,y)$, if the following are defined in $\mathbb{R}$,

$$\begin{cases} \sum_x \sum_y g(x,y) f(x,y) & ((X,Y) \text{ is discrete}) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f(x,y)\, dx\, dy & ((X,Y) \text{ is continuous}) \end{cases}$$

we call those values the expectation of $g(X,Y)$ and denote it as $E[g(X,Y)]$

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f(x,y) & ((X,Y) \text{ is discrete}) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f(x,y)\, dx\, dy & ((X,Y) \text{ is continuous}) \end{cases}$$

Ex. 2.2.1

The pdf of the bivariate r.v. is $f(x,y) = 120xy(1-x-y)\mathbb{I}_{(x\geq 0, y\geq 0, x+y\leq 1)}$
Find $E[XY]$.

$$\begin{aligned} E[XY] &= \int_0^1 \int_0^{1-y} xy \cdot 120xy(1-x-y)\, dx\, dy \\ &= \int_0^1 \int_0^{1-y} 120x^2y^2 - 120x^3y^2 - 120x^2y^3\, dx\, dy \\ &= \int_0^1 \left[40x^3y^2 - 30x^4y^2 - 40x^3y^3 \right]_0^{1-y} dy \\ &= \int_0^1 40(1-y)^3y^2 - 30(1-y)^4y^2 - 40(1-y)^3y^3\, dy \\ &= \int_0^1 40(1-3y+3y^2-y^3)y^2 - 30(1-4y+6y^2-4y^3+y^4)y^2 \\ &\quad - 40(1-3y+3y^2-y^3)y^3\, dy \\ &= \int_0^1 \cancel{40y^2} - \cancel{120y^3} + \cancel{120y^4} - 40y^5 - \cancel{30y^2} + \cancel{120y^3} - \cancel{180y^4} + \cancel{120y^5} - \cancel{30y^6} \\ &\quad - 40y^3 + \cancel{120y^4} - \cancel{120y^5} + \cancel{40y^6}\, dy \\ &= \int_0^1 10y^2 - 40y^5 - 40y^3 + 60y^4 + 10y^6\, dy \\ &= \left[\frac{10}{3}y^3 - \frac{40}{6}y^6 - 10y^4 + 12y^5 + \frac{10}{7}y^7 \right]_0^1 \\ &= \frac{10}{3} - \frac{40}{6} - 10 + 12 + \frac{10}{7} \\ &= -\frac{10}{3} + 2 + \frac{10}{7} = \frac{-70+42+30}{21} = \frac{2}{21} \end{aligned}$$

〈Theorem 2.2.1 Properties of Expectation of joint distribution〉

(a) (linearity) $E[C_1 g_1(X,Y) + C_2 g_2(X,Y)] = C_1 E[g_1(X,Y)] + C_2 E[g_2(X,Y)]$
* C_1, C_2 are constants

(b) (monotonicity) $g_1(X,Y) \leq g_2(X,Y) \Rightarrow E[g_1(X,Y)] \leq E[g_2(X,Y)]$

From the definition of Expectation, we know that an Expectation that depends on a single r.v. is an expectation of a marginal pdf

For instance

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x) f(x,y)\, dy\, dx = \int_{-\infty}^{+\infty} g(x) \left\{ \int_{-\infty}^{+\infty} f(x,y)\, dy \right\} dx \\ &= \int_{-\infty}^{+\infty} g(x) \underbrace{f_1(x)}_{\text{marginal pdf}}\, dx \end{aligned}$$

Ex. 2.2.2

From Ex. 2.2.1 the pdf is:

$$f(x,y) = 120xy(1-x-y)\mathbb{I}_{(x\geq 0, y\geq 0, x+y\leq 1)}$$

$$\begin{aligned} E[X] &= \iint_{\substack{0\leq x \\ 0\leq y \\ x+y\leq 1}} x \cdot 120xy(1-x-y)\, dx\, dy \\ &= \int_0^1 x \left\{ 120x \int_0^{1-x} y(1-x-y)\, dy \right\} dx \\ &= \int_0^1 x \left\{ 120x \int_0^{1-x} (1-x)y - y^2\, dy \right\} dx \\ &= \int_0^1 120x^2 \left[(1-x)\cdot\frac{1}{2}y^2 - \frac{1}{3}y^3 \right]_0^{1-x} dx \\ &= \int_0^1 120x^2 \left[\frac{1}{6}(1-x)^3 \right] dx \\ &= \int_0^1 20x^2(1-x)^3\, dx \\ &= \int_0^1 20(1-t)^2 t^3\, dt = 1/3 \end{aligned}$$

22) 이러한 선형성과 단조성은 각각의 기대값이 실수로 정의되는 것을 전제로 하는 것이다.

Definition: Covariance & Correlation Coefficient

(a) (Covariance) $\mathrm{Cov}(X,Y) = E[(X-\mu_1)(Y-\mu_2)]$
$(= \sigma_{1,2}$ or $\sigma_{X,Y})$ ↳ $E[X]$ ↳ $E[Y]$

(b) (Correlation Coefficient) $\mathrm{Corr}(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$ $(\sigma_1, \sigma_2 \neq 0)$

- meaning: If X moves together with Y (i.e. when x is below the mean, Y is also below its mean and vice versa), Covariance is positive and otherwise, negative. If the change in magnitude is similar (i.e. when $|X-\mu_1|$ is large, so is $|Y-\mu_2|$ and vice versa) the value of covariance ($|\mathrm{Cov}(X,Y)|$) is large.

- limitation: Covariance depends on the r.v.s measure unit (e.g. cm, m, g, Kg). Thus it is unclear whether the direction(방향) / magnitude(크기)* is due to the difference in measure unit(측정단위) or the magnitude of change(변화량의 크기) * of covariance

Thus, we divide the covariance with the standard deviation to eliminate the effects of measure unit.

〈Theorem 2.2.2 Properties of Covariance and it Calculation〉

(a) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

(b) $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$, a,b,c,d are Constants

(c) $\mathrm{Cov}(X,Y) = E[X,Y] - E[X]E[Y]$

(proof)

(a) $\mathrm{Cov}(X,Y) = E[(X-\mu_1)(Y-\mu_2)] = E[(Y-\mu_2)(X-\mu_1)] = \mathrm{Cov}(Y,X)$
$\mathrm{Cov}(X,X) = E[(X-\mu_1)(X-\mu_1)] = E[(X-\mu_1)^2] = \mathrm{Var}(X)$

(b) $\mathrm{Cov}(aX+b, cY+d) = E[(aX+b-E[aX+b])(cY+d-E[cY+d])]$
$= E[(aX-aE[X])(cY-cE[Y])]$
$= E[a(X-E[X])\cdot c(Y-E[Y])]$
$= ac\,E[(X-\mu_1)(Y-\mu_2)]$
$= ac\,\mathrm{Cov}(X,Y)$

(c) $\mathrm{Cov}(X,Y) = E[(X-\mu_1)(Y-\mu_2)] = E[XY - X\cdot\mu_2 - Y\cdot\mu_1 + \mu_1\mu_2]$
$= E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1\mu_2$
$= E[XY] - \mu_2\mu_1 - \mu_1\mu_2 + \mu_1\mu_2$
$= E[XY] - \mu_1\mu_2$
$= E[XY] - E[X]E[Y]$

Ex 2.2.3

If the pdf is $f(x,y) = 120xy(1-x-y)\,I_{(x\geq 0, y\geq 0, x+y\leq 1)}$

$E[XY] = 2/21$ $E[X] = 1/3$ and $E[Y] = 1/3$ from Ex 2.2.1 & 2.2.2

$\therefore \mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{2}{21} - \left(\frac{1}{3}\right)^2 = \frac{2}{21} - \frac{1}{9} = \frac{18-21}{189} = -\frac{1}{63}$

Ex. 2.2.4

If the pdf of $f(x,y) = 120xy(1-x-y)\,I_{(x\geq 0, y\geq 0, x+y\leq 1)}$

we know from previous exercises that the marginal pdf is

$$f_1(x) = 20x(1-x)^3 I_{(0\leq x\leq 1)}$$

and $E[X] = 1/3$

$\therefore \mathrm{Var}(X) = E[X^2] - E[X]^2 = \int_0^1 x^2 \cdot 20x(1-x)^3 dx - \left(\frac{1}{3}\right)^2$

$= \int_0^1 20x^3(1-x)^3 dx - 1/9$

$= \int_0^1 20x^3(1-3x+3x^2-x^3)dx - 1/9$

$= \int_0^1 20x^3 - 60x^4 + 60x^5 - 20x^6 dx - 1/9$

$= \left[5x^4 - 12x^5 + 10x^6 - \frac{20}{7}x^7\right]_0^1 - 1/9$

$= 5 - 12 + 10 - \frac{20}{7} - 1/9 = 3 - \frac{20}{7} - 1/9 = \frac{1}{7} - \frac{1}{9} = \frac{2}{63}$

Similarly $\mathrm{Var}(Y) = 2/63$

$\therefore \mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)} = -1/63 / 2/63 = -1/2$

〈Theorem 2.2.3 properties of Correlation Coefficient〉

$E[X] = \mu_1$, $\mathrm{Var}(X) = \sigma_1^2$, $E[Y] = \mu_2$, $\mathrm{Var}(Y) = \sigma_2^2$, $\mathrm{Cov}(X,Y) = \rho$

(a) $\mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}\right) = 1-\rho^2$

(b) $-1 \leq \rho \leq 1$

(c) $\rho = 1 \Leftrightarrow P\left(\frac{Y-\mu_2}{\sigma_2} = \frac{X-\mu_1}{\sigma_1}\right) = 1$

$\rho = -1 \Leftrightarrow P\left(\frac{Y-\mu_2}{\sigma_2} = -\frac{X-\mu_1}{\sigma_1}\right) = 1$

(proof)

(a) $E\left[\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}\right] = 0$

$\therefore \mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}\right) = E\left[\left(\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}\right)^2\right]$

$= E\left[\left(\frac{Y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{Y-\mu_2}{\sigma_2}\right)\left(\frac{X-\mu_1}{\sigma_1}\right) + \rho^2\left(\frac{X-\mu_1}{\sigma_1}\right)^2\right]$

$= \mathrm{Var}(Y') - 2\rho\,\mathrm{Corr}(Y,X) + \rho^2\mathrm{Var}(X')$; Y' and X'

$= 1 - 2\rho^2 + \rho^2 = 1-\rho^2$: Standardized Y & X

(a)-(ii) (내가 증명한 것)

$$\mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}\right)=E\left[\left(\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}\right)^2\right]$$

$$=E\left[\left(\frac{Y-\mu_2}{\sigma_2}\right)^2-2\left(\frac{Y-\mu_2}{\sigma_2}\right)\left(\frac{X-\mu_1}{\sigma_1}\right)+\left(\frac{X-\mu_1}{\sigma_1}\right)^2\right]$$

$$=\mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2}\right)-2\,\mathrm{Corr}(Y,X)+\mathrm{Var}\left(\frac{X-\mu_1}{\sigma_1}\right)$$

$$=1-2\rho+1=2(1-\rho)$$

(b) from (a), we know that $1-\rho^2\geq 0 \quad \because \mathrm{Var}(X)\geq 0$

$1\geq\rho^2 \Leftrightarrow -1\leq\rho\leq 1$

(c) $\rho=\pm 1$ 인 경우 (a)에서의 분산이 0 인 경우로서 <정리 1.6.4>로부터 (c)가 성립하는 것을 알 수 있다.

↳ Suppose $\rho=1 \Leftrightarrow \mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2}-\rho\frac{X-\mu_1}{\sigma_1}\right)=0$

$\Leftrightarrow \mathrm{Var}\left(\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}\right)=0$ Let $Z=\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}$

$\Leftrightarrow P(Z=E[Z])=1$, $E[Z]=E\left[\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}\right]=0$

* 정리 1.6.4

$\Leftrightarrow P(Z=0)=1$

$\Leftrightarrow P\left(\frac{Y-\mu_2}{\sigma_2}-\frac{X-\mu_1}{\sigma_1}=0\right)=1$

$\Leftrightarrow P\left(\frac{Y-\mu_2}{\sigma_2}=\frac{X-\mu_1}{\sigma_1}\right)=1$ Q.E.D.

<정리 2.2.3>에서와 같이, X, Y의 상관계수를 $\rho_{X,Y}$ 또는 $\rho_{1,2}$로 나타내거나 X, Y를 생략하여도 혼동이 없을 때에는 ρ로 나타낸다.

한편, <정리 2.2.3>에서 주어진 상관계수의 범위를 공분산과 표준편차의 관계로 나타내면 다음의 부등식[24]이 성립하는 것을 알 수 있다.

$$|\mathrm{Cov}(X,Y)|\leq\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$$

$|\rho|\leq 1 \Rightarrow \frac{|\mathrm{Cov}(X,Y)|}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}\leq 1$

$\Rightarrow |\mathrm{Cov}(X,Y)|\leq\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$

24) 일반적으로 다음 부등식이 성립하고, 이를 코쉬-슈바르츠 부등식(Cauchy-Schwartz) 부등식이라고 한다.

$$|E(VW)|\leq\sqrt{E[V^2]E[W^2]}$$

<정리 2.2.3>에서 알 수 있듯이 상관계수의 절대값이 커질수록 (X,Y)의 분포는 $(y-\mu_2)/\sigma_2=\rho(x-\mu_1)/\sigma_1$ 주위에 가깝게 분포되어 나타날 것이다. 이러한 뜻에서 상관계수는 두 변수 사이의 직선 관계를 나타내는 특성치라고 할 수 있다

상관계수나 공분산이 두 확률변수의 직선 관계의 특성을 설명하는 것과 같이, 결합확률분포의 특성을 나타내는 것으로는 다음과 같은 결합적률이 있다.

〈Definition: Joint Moment〉

If $E[|X^rY^s|]<+\infty$

$$E[X^rY^s]=\begin{cases}\sum_x\sum_y x^r y^s f(x,y) & (X,Y)\text{ is discrete}\\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^r y^s f(x,y)\,dx\,dy & (X,Y)\text{ is Continuous}\end{cases}$$

is called the mn-th joint moment and is denoted as $m_{r,s}(X,Y)$ or $m_{r,s}$

Let's consider a joint moment generating function. From the single random variable case, and from the exponential function's power series expansion we know the following:

$$e^{t_1X}e^{t_2Y}=\sum_{r=0}^{\infty}\frac{(t_1X)^r}{r!}\sum_{s=0}^{\infty}\frac{(t_2Y)^s}{s!}=\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{X^rY^s}{r!\,s!}t_1^r t_2^s$$

$$\therefore E[e^{t_1X}e^{t_2Y}]=\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{E[X^rY^s]}{r!\,s!}t_1^r t_2^s$$ 일 것으로 예상할 수 있다.

〈Definition: Joint Moment Generoting Function〉

If $E[e^{t_1X+t_2Y}]\in\mathbb{R}$ $\forall t_1,t_2$ in an open interval containing 0,

$M(t_1,t_2)=E[e^{t_1X+t_2Y}]$, $-h_1<t_1<h_1$, $-h_2<t_2<h_2$ $(\exists h_1>0, \exists h_2>0)$

is called the joint moment generating function (결합적률생성함수)

〈Theorem 2.2.4. Properties of Joint Moment Generating Function〉

(a) (Joint Moment Generation)

If the moment generating function of (X,Y) exists, meaning $M(t_1,t_2)=E[e^{t_1X+t_2Y}]<+\infty$, $-h_1<t_1<h$, $-h_2<t_2<h_2$ $(\exists h_1>0, \exists h_2>0)$ all joint moments of (X,Y) exists

$$E[X^rY^s]=\left[\frac{\partial^{r+s}}{\partial t_1^r\partial t_2^s}M(t_1,t_2)\right]_{\substack{t_1=0\\ t_2=0}}$$

$$M(t_1,t_2)=\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{E[X^rY^s]}{r!\,s!}t_1^r t_2^s,\quad -h_1<t_1<h,\ -h_2<t_2<h_2$$

(b) (Determines Distribution)

Suppose that the moment generating functions of two bivariate random variables (X_1,X_2) and (Y_1,Y_2) exists and are $M_{X_1,X_2}(t_1,t_2)$ and $M_{Y_1,Y_2}(t_1,t_2)$ respectively. Also suppose that

$M_{X_1,X_2}(t_1,t_2)$ and $M_{Y_1,Y_2}(t_1,t_2)$ are identical on an open interval containing 0:

$$M_{X_1,X_2}(t_1,t_2)=M_{Y_1,Y_2}(t_1,t_2)\quad \forall t_i:\ -h_i<t_i<h_i\ (\exists h_i>0),\ i=1,2$$

Then the joint pdf and cdfs of (X_1,X_2) and (Y_1,Y_2) are identical.

As the case of a single r.v.

$C(t_1,t_2) = \log M(t_1,t_2) = \log E[e^{t_1X+t_2Y}]$, $-h_i < t_i < h_i$ $(\exists h_i > 0)$ $|i=1,2)$ is called the joint cumulant distribution function and the power series expansion is:

$$C(t_1,t_2) = \sum_{\substack{r=0 \\ r+s\geq 1}}^{\infty}\sum_{s=0}^{\infty} \frac{C^{(r+s)}(0,0)}{r!\,s!}\, t_1^r t_2^s \,, \quad C^{(r+s)} = \left[\frac{\partial^{r+s}}{\partial^r \partial^s} C(t_1,t_2)\right]_{\substack{t_1=0 \\ t_2=0}}$$

The coefficient of $t_1^r t_2^s / r!\,s!$, $C^{(r,s)}(0,0)$, is the rs-th joint cumulant of (X,Y) and is denoted as $C_{r,s}(X,Y)$ or $C_{r,s}$.

The relationship between $C_{r,s}$ and $m_{j,k}$ is

$$\log\left(1 + \sum_{\substack{j=0 \\ j+k\geq 1}}^{\infty}\sum_{k=0}^{\infty} \frac{m_{j,k}}{j!\,k!}\, t_1^j t_2^k\right) = \sum_{\substack{r=0 \\ r+s\geq 1}}^{\infty}\sum_{s=0}^{\infty} \frac{C_{r,s}}{r!\,s!}\, t_1^r t_2^s ,$$

$j=0, k=0$; $r+s=0 \to C_{0,0}=0$?

$-h_i < t_i < h_i$ $(\exists h_i > 0)$ $|i=1,2)$

If we use the power series expansion of the log function:

$\log(1+A) = A - \frac{1}{2}A^2 + \frac{1}{3}A^3 - \cdots$, $|-1 < A < 1)$

$A = (m_{1,0}t_1 + m_{0,1}t_2) + (m_{2,0}t_1^2/2! + m_{1,1}t_1t_2 + m_{0,2}t_2^2/2!) + \cdots$

If we organize it in an ascending order of t_1, t_2

$C_{1,0} = m_{1,0} = E[X]$ $\quad$ $C_{0,1} = m_{0,1} = E[Y]$

$C_{2,0} = m_{2,0} - m_{1,0}^2 = E[X^2] - E[X]^2 = \mathrm{Var}(X)$

$C_{1,1} = m_{1,1} - m_{1,0}\,m_{0,1} = E[XY] - E[X]E[Y] = \mathrm{Cov}(X,Y)$

$C_{0,2} = m_{0,2} - (m_{0,1})^2 = E[Y^2] - E[Y]^2 = \mathrm{Var}(Y)$

Ex. 3.3.5

From Ex. 2.1.a, the joint pdf of (X,Y) is

$$f(x,y) = 2e^{-x-y} I_{(0\leq x\leq y)}$$

The joint moment generating function of (X,Y) is

$$E[e^{t_1X+t_2Y}] = \int_0^{\infty}\int_0^{y} e^{t_1x+t_2y}\, 2\cdot e^{-x-y}\, dx\,dy$$

$$= \int_0^{\infty} 2e^{(t_2-1)y} \int_0^{y} e^{(t_1-1)x}\, dx\, dy$$

$$= \int_0^{\infty} 2e^{-(1-t_2)y} \int_0^{y} e^{-(1-t_1)x}\, dx\, dy$$

$$= \int_0^{\infty} 2e^{-(1-t_2)y} \left[-\frac{1}{1-t_1} e^{-(1-t_1)x}\right]_0^{y} dy$$

$$= \int_0^{\infty} 2e^{-(1-t_2)y} \left\{\frac{1}{1-t_1}\left(1 - e^{-(1-t_1)y}\right)\right\} dy$$

$$= \int_0^{\infty} 2(1-t_1)^{-1}\left(e^{-(1-t_2)y} - e^{-(2-t_1-t_2)y}\right) dy$$

$$= 2(1-t_1)^{-1}\left\{(1-t_2)^{-1} - (2-t_1-t_2)^{-1}\right\}$$

$$= 2(1-t_1)^{-1}\cdot\frac{2-t_1-t_2-1+t_2}{(1-t_2)(2-t_1-t_2)} = 2(1-t_1)^{-1}\frac{1-t_1}{(1-t_2)(2-t_1-t_2)}$$

$$= 2(1-t_2)^{-1}(2-t_1-t_2)^{-1}$$

$$C_{X_1,X_2}(t_1,t_2) = \log\left(M_{X_1,X_2}(t_1,t_2)\right)$$

$$= \log(2) - \log(1-t_1) - \log(2-t_1-t_2)$$

* $\log(1-t_2) = -t_2 - \frac{1}{2}t_2^2 - \frac{1}{3}t_2^3 - \cdots$

$\therefore -\log(1-t_2) = t_2 + \frac{1}{2}t_2^2 + \frac{1}{3}t_2^3 + \cdots$

* $\log(2-t_1-t_2) = \log(1-(t_1+t_2-1))$

Let $A = t_1+t_2-1$

$\log(2-t_1-t_2) = -(t_1+t_2-1) - \frac{1}{2}(t_1+t_2-1)^2 - \frac{1}{3}(t_1+t_2-1)^3$

$\therefore -\log(2-t_1-t_2) = t_1+t_2-1 + \frac{1}{2}(t_1^2+t_2^2+1-2t_1-2t_2+2) + \cdots$

Problem: $\log(1-(t_1+t_2-1))$'s power series expansion does not allow us to isolate t_1, t_2 and t_1t_2. This is due to the presence of (-1) inside the bracket. For example $(t_1+t_2-1)^n = t_1^n + \cdots + (-1)^{n-1}t_1^1 + \cdots$

instead, denote $M_{X_1,X_2}(t_1,t_2) = (1-t_2)^{-1}\left(\frac{2}{2-t_1-t_2}\right)$

$$\log M_{X_1,X_2}(t_1,t_2) = -\log(1-t_2) - \log\left(1 - \frac{t_1+t_2}{2}\right)$$

$$= \left(t_2 + \frac{1}{2}t_2^2 + \cdots\right) + \left(\frac{t_1+t_2}{2}\right) + \frac{1}{2}\left(\frac{t_1+t_2}{2}\right)^2 + \cdots$$

$$= \frac{1}{2}t_1 + \frac{3}{2}t_2 + \frac{1}{8}t_1^2 + \frac{5}{8}t_2^2 + \frac{1}{4}t_1t_2 + \cdots$$

$$= \frac{1}{2}t_1 + \frac{3}{2}t_2 + \frac{1}{4}\frac{1}{2!}t_1^2 + \frac{5}{4}\frac{1}{2!}t_2^2 + \frac{1}{4}\frac{1}{1!}\frac{1}{1!}t_1t_2$$

$C_{1,0} = 1/2 = E[X]$
$C_{0,1} = 3/2 = E[Y]$
$C_{1,1} = 1/4 = Cov(X,Y)$
$C_{2,0} = 1/4 = Var(X)$
$C_{0,2} = 5/4 = Var(Y)$

$$Corr(X,Y) = Cov(X,Y) / \sqrt{Var(X)\,Var(Y)} = 1/4 \,/\, (1/4 \cdot 5/4)^{1/2} = 1/\sqrt{5}$$

In order to distinguish the joint mgf of (X,Y) and the mgf of X and Y we call the mgf of a single random variable a "marginal mgf"

$M_X(s) = E[e^{sX}]$, $-h_1 < s < h_1$ $(\exists\, h_1 > 0)$ (X's marginal mgf)
$M_Y(t) = E[e^{tY}]$, $-h_2 < t < h_2$ $(\exists\, h_2 > 0)$ (Y's marginal mgf)

And the log of these functions are called marginal cgf.
We can find marginal mgf / cgf as below:

$$M_X(s) = M_{X,Y}(s,0), \quad C_X(s) = C_{X,Y}(s,0)$$

$$M_Y(t) = M_{X,Y}(0,t), \quad C_Y(t) = C_{X,Y}(0,t)$$

Ex. 2.2.6

We know from Ex 2.1.4 that the marginal pdf of X's pdf given in Ex 2.1.2 is:

$$f_1(x) = 2e^{-x} I_{(x \geq 0)}$$

Thus, the marginal mgf is: $M_X(s) = \int_0^{+\infty} e^{sx} \cdot 2 \cdot e^{-2x}\, dx = (1 - s/2)^{-1}$, $s < 2$

This is equivalent to Ex. 2.2.5 $M_{X,Y}(s,0)$

2.3 CONDITIONAL DISTRIBUTION & CONDITIONAL EXPECTATION.

Given that X, Y are discrete, under $X = x$, the conditional probability

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

is defined only if $P(X = x) > 0$.

Ex 2.3.1 Distribution Table of conditional Probability

Two white balls, three black balls, and four red balls are in a box. Suppose that you choose 3 balls from the box randomly.
Let X denote the number of white balls and Y the number of black balls. From Ex. 2.1.1 we know that the probability distribution is given as below.

		Value of Y				Sum
		$Y=0$	$Y=1$	$Y=2$	$Y=3$	
Value of X	$X=0$	4/84	18/84	12/84	1/84	35/84
	$X=1$	12/84	24/84	6/84	0	42/84
	$X=2$	4/84	3/84	0	0	7/84
Sum		20/84	45/84	18/84	1/84	1

표 2.1.1

Let's find the probability distribution of Y given $X = 0$.
The conditional probability we want to find is:

$$P(Y=y \mid X=0) = P(X=0, Y=y) / P(X=0) \quad (y = 0,1,2,3)$$

$P(Y=y \mid X=0)$ value				Total
$Y=0 \mid X=0$	$Y=1 \mid X=0$	$Y=2 \mid X=0$	$Y=3 \mid X=0$	
4/35	18/35	12/35	1/35	1

표 2.3.1

Definition: Conditional pdf (discrete)

: 이차원 확률변수 (X,Y)가 이산형으로서 (X,Y)의 결합확률밀도함수가 $f_{1,2}(x,y)$이고 X의 주변확률밀도함수가 $f_1(x)$일 때, $X = x$인 조건에서 Y의 가능한 값에 조건부확률을 대응시킨 함수

$$f_{2|1}(y|x) = \frac{f_{1,2}(x,y)}{f_1(x)} = \frac{P(X=x, Y=y)}{P(X=x)} = P(Y=y \mid X=x)$$

을 $X = x$인 조건에서 Y의 조건부 확률밀도함수라고 한다.

$$f_{2|1}(y|x) = \frac{f_{1,2}(x,y)}{f_1(x)} = P(X=x \mid Y=y), \quad (x : f_1(x) > 0)$$

(a) $f_{2|1}(y|x) \geq 0 \quad \forall y : -\infty < y < +\infty$
$f_{2|1}(y|x) = 0 \quad \forall y : y \neq y_k \ (k = 1, 2, \cdots)$

(b) $\sum_y f_{2|1}(y|x) = 1$

Definition: Conditional pdf (continuous)

If the bivariate r.v. is continuous, $P(c \le Y \le d, X = x) = 0$, $P(X = x) = 0$. Thus, we cannot use the previous definition. But we can consider the following for $h > 0$

$$P(c \le Y \le d \mid x \le X \le x+h) = \frac{P(c \le Y \le d, x \le X \le x+h)}{P(x \le X \le x+h)}$$

If the joint pdf of the bivariate r.v. $f_{1,2}$ and its marginal pdf is continuous, the following limit equation is true as $h \to 0$

$$P(c \le Y \le d, x \le X \le x+h)/h = \frac{1}{h}\int_x^{x+h}\left\{\int_c^d f_{1,2}(s,y)\,dy\right\}ds \to \int_c^d f_{1,2}(x,y)\,dy$$

$$P(x \le X \le x+h)/h = \frac{1}{h}\int_x^{x+h} f_1(s)\,ds \to f_1(x)$$

Thus, similar to the discrete case, if $f_1(x) > 0$, for $f_{2|1}(y|x)$ defined as:

$$f_{2|1}(y|x) = \frac{f_{1,2}(x,y)}{f_1(x)}$$

the following equality holds.

$$\lim_{\substack{h\to 0\\ h>0}} P(c \le Y \le d \mid x \le X \le x+h) = \int_c^d f_{2|1}(y|x)\,dy \quad (*)$$

We define the LHS limit (*) as the conditional probability of Y given $X = x$ and denote it as:

$$P(c \le Y \le d \mid X = x) = \int_c^d f_{2|1}(y|x)\,dy$$

and we call $f_{2|1}(y|x)$ the "conditional pdf of Y given $X = x$"

연속형인 경우의 조건부확률밀도함수

$$f_{2|1}(y|x) = \frac{f_{1,2}(x,y)}{f_1(x)} \quad (x : f_1(x) > 0)$$

(a) $f_{2|1}(y|x) \ge 0 \quad \forall y : -\infty < y < +\infty$

(b) $\int_{-\infty}^{+\infty} f_{2|1}(y|x)\,dy = 1$

(c) $P(c \le Y \le d \mid X = x) = \int_c^d f_{2|1}(y|x)\,dy$

Ex. 2.3.2

From Ex. 2.1.2, 2.1.4 the joint pdf and marginal pdf are:

$$f_{1,2}(x,y) = 2e^{-x-y} I_{(0 \le x \le y)}, \quad f_1(x) = 2e^{-2x} I_{(x \ge 0)}$$

$$f_{2|1}(y|x) = 2e^{-(y-x)} I_{(y \ge x)}$$

$$P(Y \ge c \mid X = x) = \int_c^{+\infty} e^{-(y-x)}\,dy = \int_{c\vee x}^{+\infty} e^{-(y-x)}\,dy = e^{-(c\vee x - x)}$$

⟨Theorem 2.3.1 Properties of conditional Probability⟩

$$P(a \le X \le b, c \le Y \le d) = \begin{cases} \sum_{a \le x \le b} P(c \le Y \le d \mid X = x) f_1(x), & \text{discrete} \\ \int_a^b P(c \le Y \le d \mid X = x) f_1(x)\,dx, & \text{continuous} \end{cases}$$

(proof)

$$\int_a^b P(c \le Y \le d \mid X = x) f_1(x)\,dx = \int_a^b \left\{\int_c^d f_{2|1}(y|x)\,dy\right\} f_1(x)\,dx$$

$$= \int_a^b \int_c^d f_{1,2}(x,y)\,dy\,dx$$

$$= P(a \le X \le b, c \le Y \le d)$$

이산형의 경우 같은 방법으로 증명이 가능하다. 노트 p.1 Ex. 2.1.1 밑의 정의 참고

Ex. 2.3.3

From Ex. 2.1.2 and Ex. 2.3.2, the joint pdf of (X,Y) is

$$f_{1,2}(x,y) = 2e^{-x-y} I_{(0 \le x \le y)}$$

the marginal pdf and conditional probabilities are:

$$f_1(x) = 2e^{-2x} I_{(x \ge 0)}, \quad P(Y \ge c \mid X = x) = e^{-(c\vee x - x)}$$

Thus, $P(X \ge a, Y \ge 3) = \int_2^{\infty} e^{-(3\vee x - x)} 2e^{-2x}\,dx$

$$= \int_2^3 2e^{-3-x}\,dx + \int_3^{+\infty} 2e^{-2x}\,dx$$

$$= 2e^{-5} - e^{-6}$$

the results are equivalent to Ex. 2.1.2

Definition: Conditional Mean (조건부 평균)

If the conditional pdf of Y given $X = x$ if $f_{2|1}(y|x)$, and

$$\begin{cases} \sum_y y f_{2|1}(y|x) & \text{(discrete)} \\ \int_{-\infty}^{+\infty} y f_{2|1}(y|x)\,dy & \text{(continuous)} \end{cases}$$

are defined in $\mathbb{R}$, we call it the conditional mean of Y given $X = x$ and we denote it as $E[Y|X=x]$ or $\mu_{2|1}(x)$

$$\mu_{2|1}(x) = \begin{cases} \sum_y y f_{2|1}(y|x) & \text{(discrete)} \\ \int_{-\infty}^{+\infty} y f_{2|1}(y|x)\,dy & \text{(continuous)} \end{cases}$$

Ex. 2.3.4

$$f_{1,2}(x,y) = 120xy(1-x-y)I_{(x\ge 0, y\ge 0, x+y\le 1)}$$

$$f_1(x) = 20x(1-x)^3 I_{(0\le x\le 1)}$$

$$\therefore f_{2|1}(y|x) = 6y(1-x)^{-3}(1-x-y)I_{(1-x\ge y\ge 0)}$$

Thus, the conditional mean of Y given $X=x$ is

$$\int_0^{1-x} y\cdot 6y(1-x)^{-3}(1-x-y)\,dy$$

$$= (1-x)^{-3}\int_0^{1-x} 6y^2 - 6xy^2 - 6y^3\,dy$$

$$= (1-x)^{-3}\left[2y^3 - 2xy^3 - \frac{3}{2}y^4\right]_0^{1-x}$$

$$= 2(1-x)^{-3}(1-x)^3 - 2(1-x)^{-3}(1-x)^3\cdot x - \frac{3}{2}(1-x)^{-3}(1-x)^4$$

$$= 2 - 2x - \frac{3}{2}(1-x) = \frac{1}{2}(1-x) \checkmark\checkmark$$

Definition: Conditional Expectation and Conditional Variance

$$E[g(X,Y)|X=x] = \begin{cases} \sum_y g(x,y) f_{2|1}(y|x) & \text{(discrete)} \\ \int_{-\infty}^{+\infty} g(x,y) f_{2|1}(y|x)\,dy & \text{(continuous)} \end{cases}$$

($\mathbb{R}$) 실수로 정의되어야 함

"conditional expected value of Y, given $X=x$"

Definition: Conditional Variance

$E[(Y-\mu_{2|1}(x))^2 | X=x]$ is called the "Conditional variance of Y given $X=x$". Here, $\mu_{2|1}(x) = E[Y|X=x]$.

$$Var(Y|X=x) = E[(Y-\mu_{2|1}(x))^2 | X=x]$$

$$= \begin{cases} \sum_y (y-\mu_{2|1}(x)) f_{2|1}^2(y) & \text{(discrete)} \\ \int_{-\infty}^{+\infty} (y-\mu_{2|1}(x))^2 f_{2|1}(y)\,dy & \text{(continuous)} \end{cases}$$

〈Theorem 2.3.2. Properties of Conditional Expectation〉

(a) (i) $E[c_1 g_1(Y) + c_2 g_2(Y) | X=x]$
$= c_1 E[g_1(Y)|X=x] + c_2 E[g_2(Y)|X=x]$ (c_1, c_2 are constants)

(ii) $E[c(X)g(Y)|X=x] = c(x)E[g(Y)|X=x]$

(b) $g_1(Y) \le g_2(Y) \Rightarrow E[g_1(Y)|X=x] \le E[g_2(Y)|X=x]$

〈Theorem 2.3.3. Conditional Variance's Calculation〉

$$Var(Y|X=x) = E[Y^2|X=x] - \{E(Y|X=x)\}^2$$

(proof)

$$\begin{aligned} Var(Y|X=x) &= E[(Y-\mu_{2|1}(x))^2 | X=x] \\ &= E[Y^2 - 2Y\cdot\mu_{2|1} + \mu_{2|1}^2 | X=x] \\ &= E[Y^2|X=x] - 2E[Y\cdot\mu_{2|1}|X=x] + E[\mu_{2|1}^2|X=x] \\ &= E[Y^2|X=x] - 2\cdot\mu_{2|1}\cdot E[Y|X=x] + \mu_{2|1}^2 \\ &= E[Y^2|X=x] - 2\mu_{2|1}^2 + \mu_{2|1}^2 \\ &= E[Y^2|X=x] - \mu_{2|1}^2 \\ &= E[Y^2|X=x] - E[Y|X=x]^2 \end{aligned}$$

Ex. 2.3.5

In Ex. 2.3.4, the conditional pdf and conditional mean of Y given $X=x$ are

$$f_{2|1}(y|x) = \frac{f_{1,2}(x,y)}{f_1(x)} = \frac{6y(1-x-y)}{(1-x)^3} I_{(0\le y\le 1-x)}$$

$$\mu_{2|1}(x) = E[Y|X=x] = (1-x)/2$$

Thus, the conditional variance of Y given $X=x$ is

$$\int_0^{1-x} y^2\cdot\frac{6y(1-x-y)}{(1-x)^3}\,dy - \left[\frac{1-x}{2}\right]^2$$

$$= (1-x)^{-3}\int_0^{1-x} 6y^3 - 6xy^3 - 6y^4\,dy - (1-x)^2/4$$

$$= (1-x)^{-3}\left[\frac{6}{4}y^4 - \frac{6}{4}xy^4 - \frac{6}{5}y^5\right]_0^{1-x} - (1-x)^2/4$$

$$= \frac{3}{2}(1-x) - \frac{3}{2}x(1-x) - \frac{6}{5}(1-x)^2 - (1-x)^2/4$$

$$= (1-x)^2\left\{\frac{3}{2} - \frac{6}{5} - \frac{1}{4}\right\} = (1-x)^2\frac{30-24-5}{20}$$

$$= \frac{(1-x)^2}{20}$$

Definition: Conditional Expectation ($E[g(Y)|X]$) as a random variable.

If we say that the conditional expectation of $g(Y)$ given $X=x$ is:

$$h(x) = E[g(Y)|X=x]$$

we call the random variable having the values $h(x)$ based on the distribution of X as: "Conditional Expectation of $g(Y)$ given X". It is denoted as $E[g(Y)|X] = h(X)$

Ex 2.3.6.

From Ex. 2.3.4 and Ex. 2.3.5

$\mu_{2|1}(x) = E[Y|X=x] = (1-x)/2$, $\sigma^2_{2|1}(x) = Var[Y|X=x] = (1-x)^2/20$

Thus, $E(Y|X) = \mu_{2|1}(x) = (1-x)/2$, $Var(Y|X) = \sigma^2_{2|1}(x) = (1-x)^2/20$

〈Theorem 2.3.4. Properties of Conditional Expectation as a random variable〉

*If you can, memorize the proof

(a) $E[E[Y|X]] = E[Y]$

(b) $Cov(Y - E[Y|X], v(X)) = 0 \quad \forall v(x)$

(proof)

(a) $E[Y|X] = \int_{-\infty}^{+\infty} y \cdot f_{2|1}(y|x)\,dy$

$$E[E[Y|X]] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y \cdot f_{2|1}(y|x)\, f(x)\,dy\,dx$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y\, f(y,x)\,dy\,dx$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{+\infty} y\, f(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{+\infty} y \cdot f(y)\,dy = E[Y]$$

(b) $Cov(Y - E[Y|X], v(X))$

$= E[(Y - E[Y|X])\,v(X)] - E[Y - E[Y|X]]\,E[v(X)]$

$= E[Y \cdot v(X)] - E[E[Y|X]\,v(X)] - E[Y]E[v(X)] + E[Y]E[v(X)]$

$= E[E[Y \cdot v(X) | X = x]] - E[E[Y|X]\,v(X)]$

$= E[E[Y|X]\,v(X)] - E[E[Y|X]\,v(X)]$

$= 0$

[Textbook Solution]

Let $z = Y - E[Y|X]$. From (a) we know that $E[z] = 0$ and $E[z|X] = 0$

$$E[z|X] = E[Y|X] - E\{E[Y|X]\,|\,X\} = E[Y|X] - E[Y|X] = 0$$

$$E[z] = E[E[z|X]] = 0$$

$$\therefore Cov(z, v(x)) = E[z \cdot v(x)] - E[z]E[v(x)]$$
$$= E\{E(z \cdot v(x)|x)\} = E\{v(x)\,E(z|x)\}$$
$$= 0$$

*To understand the formal proof of $E[E[Y|X]|x] = E[Y|x]$, you need to learn measure theory.

Intuitive intepretation: $E[Y|X]$ is the "best guess" of Y based on the information in X. Because this guess is constructed entirely using X, it is effectively a constant once X is known.
↳ But we are treating it as a r.v. here.

If you are asked to guess the value of a variable that is already known to you (because it is a function of X and you are given X), your best guess is simply the value itself - Gemini 3.0

Ex. 2.3.7

Ex. 2.2.1 and Ex. 2.2.2, (X,Y) joint pdf and X's marginal pdf are respectively

$$f_{1|2}(x,y) = 120\,xy\,(1-x-y)\,I_{(x\geq 0,\, y\geq 0,\, x+y\leq 1)}$$

$$f_1(x) = 20\,x(1-x)^3\,I_{(0\leq x\leq 1)}$$

From Ex. 2.3.4 and Ex. 2.3.6, $E[Y|X] = (1-X)/2$

thus, $E[E(Y|X)] = E[(1-X)/2] = (1 - E[X])/2$

However, from Ex. 2.2.2, $E[X] = \frac{1}{3}$ and similarly for Y, $E[Y] = 1/3$

Thus, $E[E[Y|X]] = (1 - E(X))/2 = 1/3 = E[Y]$

Meanwhile,

$$Cov(Y - E[Y|X]) = E[\{Y - E[Y|X]\}X]$$
$$= E[\{Y - (1-X)/2\}X]$$
$$= E[XY] - (E(X) - E(X^2))/2$$

From Ex. 2.2.2, Ex. 2.2.3 and Ex. 2.2.4

$E[X] = 1/3$, $E[XY] = 2/21$, $E[X^2] = 1/7$

$\therefore Cov(Y - E[Y|X], X) = 2/21 - (1/3 - 1/7)/2 = 0$

* $E[Y|X]$는 Y를 예측하기 "가장 좋은" X의 함수라는 뜻을 갖고 있고, 흔히 회귀함수(regression function)라고 한다.

〈Theorem 2.3.5 Least Squares Predictor〉

The function $u^*(x)$ that minimizes $E[(Y-u(x))^2]$ is $E[Y|X]$

$$E[(Y - E[Y|X])^2] \leq E[(Y - u(x))^2], \quad \forall u(x) \quad \cdots (1)$$

The mean squared prediction error $E[(Y-u(x))^2]$'s minimum value is

$$E[(Y - E(Y|X))^2] = E[Var(Y|X)] \quad \cdots (2)$$

(proof)

(1) $E[(Y-u(x))^2] = E[(Y - E[Y|X] + E[Y|X] - u(x))^2]$

$= E[(Y-E[Y|X])^2 + 2(Y-E[Y|X])(E[Y|X]-u(x)) + (E[Y|X]-u(x))^2]$

($2(Y-E[Y|X])(E[Y|X]-u(x)) = 0$)

$= E[(Y-E[Y|X])^2] + E[(E[Y|X]-u(x))^2]$

$E[(E[Y|X]-u(x))^2] \geq 0$

$\therefore E[(Y-u(x))^2] \geq E[(Y-E[Y|X])^2]$

(2) $E[(Y-E[Y|X])^2] = E[Y^2 - 2YE[Y|X] + E[Y|X]^2]$

$= E[E[Y^2|X] - 2E[Y|X]E[Y|X] + E[Y|X]^2]$

$= E[E[Y^2|X] - E[Y|X]^2]$

$= E[Var(Y|X)]$

[Textbook Proof][26]

The key idea of this proof is to reformulate the expression into the sum of two r.v.s whose covariance equals 0

$Y-u(X) = (Y-E[Y|X]) \oplus (E[Y|X]-u(X))$

Let $v(X) = E[Y|X]-u(X)$

$[Y-u(X)]^2 = [(Y-E[Y|X]) + v(X)]^2$

$= (Y-E[Y|X])^2 + 2v(X)(Y-E[Y|X]) + v(X)^2$

$E[(Y-u(X))^2] = E[(Y-E[Y|X])^2] + 2\,Cov(v(X), Y-E[Y|X]) + E[v(X)^2]$

$= E[(Y-E[Y|X])^2] + 0 + E[v(X)^2]$

$\geq E[(Y-E[Y|X])^2]$

Meanwhile, let $W=(Y-E(Y|X))^2$. Using the property of conditional expectation

$E[W|X] = E[(Y-E[Y|X])^2|X] = Var(Y|X)$

$\therefore E[(Y-E[Y|X])^2] = E[E[W]] = E[Var(Y|X)]$

*26) 이제부터 확률변수의 합을 $\oplus$의 기호로 나타낼 때에는 확률변수 사이의 상관계수가 0인 것을 전제로 하는 것을 약속한다.

〈Theorem 2.3.6 : Decomposition of Variance〉

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

(proof)

$Var(Y) = E[Y^2] - E[Y]^2$

$= E[E[Y^2|X]] - E[E[Y|X]]^2$

$= E[E[Y^2|X]] - E[E[Y|X]^2] + E[E[Y|X]^2] - E[E[Y|X]]^2$

$= E[E[Y^2|X] - E[Y|X]^2] + E[E[Y|X]^2] - E[E[Y|X]]^2$

$= E[Var(Y|X)] + Var(E[Y|X])$ □

[Textbook proof]

Let $E[Y] = \mu$

$Y-\mu = (Y-E[Y|X]) \oplus (E[Y|X]-\mu)$

However, $E[(Y-E[Y|X])(E[Y|X]-\mu)] = Cov(Y-E[Y|X], E[Y|X]) = 0$

$E[(Y-\mu)^2] = E[(Y-E(Y|X))^2] + E[(E[Y|X]-\mu)^2]$

Here, $E[E[Y|X]] = E[Y] = \mu$

$\therefore Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$

Ex. 2.3.8.

In Ex. 2.2.1, the joint pdf of (X,Y) is

$f_{1,2}(x,y) = 120xy(1-x-y)I_{(x\geq 0, y\geq 0, x+y\leq 1)}$

In Ex. 2.3.6 $E[Y|X] = (1-X)/2$, $Var(Y|X) = (1-x)^2/20$

In Ex. 2.2.2 and Ex. 2.2.4 $E[X] = 1/3$, $E[X^2] = 1/7$, $Var(X) = 2/63$

Hence, $Var[E(Y|X)] + E[Var(Y|X)] = Var[(1-X)/2] + E[(1-X)^2/20]$

$= Var(X)/4 + [E(X^2) - 2E(X) + 1]/20$

$= 2/63$

We can see that the result is equivalent to Ex. 2.2.4 $Var(Y) = 2/63$

2.4. INDEPENDENCE OF RANDOM VARIABLES

Definition: Mutually Independent

For two random variables X and Y if

$$P(a \le X \le b,\ c \le Y \le d) = P(a \le X \le b)P(c \le Y \le d),$$

X and Y are mutually independent

〈Theorem 2.4.1 : Independence of r.v.s〉

The following conditions are the necessary and sufficient conditions for r.v.s X and Y to be independent.

(a) (cdf) $cdf_{1,2}(x,y) = cdf_1(x) \cdot cdf_2(y) \quad \forall x,y : -\infty < x < +\infty, -\infty < y < +\infty$

(b) (pdf) $pdf_{1,2}(x,y) = pdf_{1,2}(x,y) = pdf_1(x)\,pdf_2(y)$ $\forall x,y : -\infty < x < +\infty, -\infty < y < +\infty$

(c) (mgf)[28] $mgf_{1,2}(t_1,t_2) = mgf_1(t_1)\,mgf_2(t_2) \quad \forall t_i : -h_i < t_i < h_i \ (\exists h_i > 0)$ $(i = 1,2)$

(d) (Probability Measure)[29] $P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \quad \forall A, B$

(proof)

(a) $\Leftrightarrow$ (b) due to the relationship between cdf and pdf

(b) $\Leftrightarrow$ (c) True because of 1-1 mapping between pdf and mgf

(b) $\Leftrightarrow$ (d) because of 1-1 mapping between pdf and probability measure

2 Meanwhile, if we find the integral/summation of pdf over an interval, we get the probability over the interval. Thus if (b) is true, Thm. 2.4.1 holds. On the contrary, if thm 2.4.1 holds for $a = c = -n$, $b = x$, $d = y$ $(n = 1, 2, \cdots)$, by taking the limit for n, we know that (a) holds from the continuity of probability measure.

* 28) (c)의 조건은 적률생성함수가 존재할 때에 나머지 함수들과 동치이다.

* 29) $X \in A$는 $\{S : X(S) \in A\}$, 즉 역상상 (inverse image)을 뜻한다.

Ex. 2.4.2

From Ex. 2.1.2 and Ex. 2.1.4, the joint pdf of (X,Y) is

$$f_{1,2}(x,y) = 2e^{-x-y} I_{(0 \le x \le y)}$$

the marginal pdfs are :

$$f_1(x) = 2e^{-2x} I_{(x \ge 0)}, \quad f_2(x) = 2e^{-y}(1 - e^{-y}) I_{(y \ge 0)}$$

$$\therefore f_{1,2}(x,y) \ne f_1(x) f_2(y) \quad \forall x,y$$

Hence X and Y are not independent.

As Ex. 2.4.2, random variables that are not independent are called mutually dependent.

To show mutual dependence, find a counter example to (a)-(d) in thm. 2.4.1

It is not always necessary to find the marginal pdfs to show independence. Instead of thm. 2.4.1 (b) you can apply (b').

(b') there exists functions* $g_1(x)$ and $g_2(y)$ which consists of only x and y respectively.

$$pdf_{1,2}(x,y) = g_1(x)\,g_2(y) \quad \forall x, y$$

If (b') is true and continuous, integrate over all x and y

$$\int_{-\infty}^{+\infty} g_1(x)\,dx \int_{-\infty}^{+\infty} g_2(y)\,dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} pdf_{1,2}(x,y)\,dx\,dy = 1$$

From the equality above, we know that (b) holds.

$$g_1(x)\,g_2(y) = \underbrace{\left\{ g_1(x) \Big/ \int_{-\infty}^{+\infty} g_1(x)\,dx \right\}}_{f_1(x)} \underbrace{\left\{ g_2(y) \Big/ \int_{-\infty}^{+\infty} g_2(y)\,dy \right\}}_{f_2(y)} = pdf_1(x)\,pdf_2(y)$$

* 누락: $g_1(x)$, $g_2(y)$ must be non-negative functions.

For an example: $f_{1,2}(x,y) = 2e^{-x-2y} I_{(x \ge 0, y \ge 0)} = \underbrace{2e^{-x} I_{(x \ge 0)}}_{g_1(x)} \underbrace{e^{-2y} I_{(y \ge 0)}}_{g_2(x)}$

$\therefore$ X and Y are independent

Similarly for (a) and (c), we can verify the independence of X and Y without finding the marginal cdf or mgf.

★ **〈Theorem 2.4.2 Independent Variable's Function〉**

If X and Y are independent, g(X) and g(Y) are independent.

(proof) Let the inverse image of g be $g^{-1}(A) = \{ x : g(x) = A \}$

$(g_1(X) \in A, g_2(Y) \in B) = (X \in g_1^{-1}(A), Y \in g_2^{-1}(B))$

$(g_1(X) \in A) = (X \in g_1^{-1}(A))$, $(g_2(Y) \in B) = (Y \in g_2^{-1}(B))$

Meanwhile, if X and Y are independent, from thm 2.4.1 (d)

$P(X \in g_1^{-1}(A), Y \in g_2^{-1}(B)) = P(X \in g_1^{-1}(A))P(Y \in g_2^{-1}(B))$

$\therefore P(g_1(X) \in A, g_2(Y) \in B) = P(g_1(X) \in A)P(g_2(Y) \in B)$

Thus $g_1(X)$ and $g_2(Y)$ are independent.

[Gemini 3.0. proof]

1. Setup and definitions

Definition of Independence of random variables:
X and Y are independent if and only if for all Borel sets $A, B \in \mathcal{B}(\mathbb{R})$

$$P(X \in A, Y \in B) = P(X \in A)\,P(Y \in B)$$

This is equivalent to saying that the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$ are independent.

2. Understanding the Generated σ-algebra

The σ-algebra $\sigma(X)$ is the collection of all events that can be determined by the value of X.

Mathematically, it is defined as: $\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\}$

Now consider the function $Z = g_1(X)$. The σ-algebra generated by Z is:

$$\sigma(g_1(X)) = \{(g_1 \circ X)^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$$

Using the properties of inverse images, we know that

$$(g_1 \circ X)^{-1}(B) = X^{-1}(g_1^{-1}(B))$$

Since g_1 is a measureable function, the set $g_1^{-1}(B)$ is a Borel set. Let's call this set A. Therefore, $X^{-1}(g_1^{-1}(B)) = X^{-1}(A)$, which is an element of $\sigma(X)$

3. Proof By Containment

Because every set in $\sigma(g_1(X))$ is also a set in $\sigma(X)$, we have the following relationship:

$$\sigma(g_1(X)) \subseteq \sigma(X) \quad \text{and} \quad \sigma(g_2(Y)) \subseteq \sigma(Y)$$

In plain English: "The information contained in $g_1(X)$ is a subset of the information contained in X".

4. Conclusion

If two σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent, then any of their Sub σ-algebras are automatically independent.

Since $\sigma(g_1(X))$ is a sub-σ-algebra of $\sigma(X)$ and $\sigma(g_2(Y))$ is a sub-σ-algebra of $\sigma(Y)$, it follows that:

$$P(g_1(X) \in B_1, g_2(Y) \in B_2) = P(g_1(X) \in B_1)\,P(g_2(Y) \in B_2)$$

for all Borel sets B_1, B_2. Thus, $g_1(X)$ and $g_2(Y)$ are independent.

Q) What is a Borel Set? - Gemini 3.0

One line Summary: any set that can be constructed from open intervals (or open sets) using a countable number of operations: unions, intersections, and complements.

1. Ingredients: To build a Borel set, you start with the basic building blocks of topology.

- Open sets : e.g. $(0,1)$
- Closed sets : e.g. $[0,1]$
- Single points : e.g. $\{5\}$

2. The rules (The σ-Algebra)
You are allowed to take these blocks and combine them using:

2-1. Countable unions: $A_1 \cup A_2 \cup A_3 \cup \cdots$
2-2. Countable intersections: $A_1 \cap A_2 \cap A_3 \cap \cdots$
2-3. Complements: Everything "outside" the set

The collection of all sets that can be created this way is called the Borel σ-algebra. Any individual member of that collection is a Borel set.

Key takeaways: If a set is a Borel set, it means we can mathematically assign it a "size" or "probability" without breaking the laws of logic.

〈Theorem 2.4.3: The expectation of the product of independent random variable (독립인 확률변수들의 곱의 기대값)〉

If the random variables X and Y are independent,

$$E[g_1(X)\,g_2(Y)] = E[g_1(X)]\,E[g_2(X)]$$

(proof)

$$E[g_1(X)\,g_2(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(x)\,g_2(y)\,pdf_{1,2}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{+\infty} g_1(x)\,pdf_1(x)\,dx \int_{-\infty}^{+\infty} g_2(y)\,pdf_2(y)\,dy$$

$$= E[g_1(x)]\,E[g_2(y)]$$

〈Theorem 2.4.4 Independence and correlation〉

If r.v.s X and Y are independent, $Cov(X,Y)=0$

(proof)

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$
$$= E[X]E[Y] - E[X]E[Y]$$
$$= 0$$

Ex. 2.4.3. Covariance = 0 but not independent r.v.s

$$f_{1,2}(x,y) = \frac{9}{40}(1+x^2y^2)\,I_{(-1\le x\le 1)}\,I_{(-1\le y\le 1)}$$

marginal pdfs are:

$$f_1(x) = \int_{-1}^{1}\frac{9}{40}(1+x^2y^2)\,dy = \left[\frac{9}{40}y + \frac{3}{40}x^2y^3\right]_{-1}^{1} = \frac{9}{20} + \frac{3}{20}x^2 = \frac{3}{20}(3+x^2)$$
$$-1\le x\le 1$$

$$f_2(y) = \cdots = \frac{3}{20}(3+y^2) \qquad -1\le y\le 1$$

$E[X]=0$, $E[Y]=0$, $E[XY] = \int_{-1}^{1}\int_{-1}^{1}\frac{9}{40}(xy + x^3y^3)\,dx\,dy$

$$= \int_{-1}^{1}\frac{9}{40}\left[\frac{1}{2}x^2y + \frac{1}{4}x^4y^3\right]_{-1}^{1}dy$$

$$= \int_{-1}^{1} 0\,dy = 0$$

$\therefore Cov(X,Y) = E[XY] - E[X]E[Y] = 0$

Also, $pdf_{1,2}(x,y) \neq pdf_1(x)\cdot pdf_2(y)$

$\therefore$ X and Y are not independent.

〈Theorem 2.4.5 Variance of Sum of random variables〉

(a) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

(b) X and Y are independent $\Rightarrow Var(X+Y) = Var(X) + Var(Y)$

(proof)

(a) $Var(X+Y) = E[(X+Y-\mu_1-\mu_2)^2]$, $\mu_1 = E[X]$, $\mu_2 = E[Y]$

$$= E[(X-\mu_1)^2 + 2(X-\mu_1)(Y-\mu_2) + (Y-\mu_2)^2]$$
$$= Var(X) + 2Cov(X,Y) + Var(Y)$$

(b) X and Y are independent $\Rightarrow Cov(X,Y)=0$

$\therefore Var(X,Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

$$= Var(X) + Var(Y)$$

2.5 MULTIVARIATE PROBABILITY DISTRIBUTION

Definition: K variate random vector

vector[**] $(X_1, X_2, \cdots X_K)^t$, whose components X_i is a random variable, is called a K variate random vector.

*) vector is always a column vector unless said otherwise and t denotes a vector or the transpose of a matrix.

Definition: Discrete Multivariate Joint pdf:

(a) $f(x_1, \cdots, x_K) \geq 0 \quad \forall x_i : -\infty \leq x_i \leq +\infty \quad (i = 1, 2, \cdots, K)$
$f(x_1, \cdots, x_K) = 0 \quad \forall x_i \notin \{x_{i1}, x_{i2}, \cdots\} \quad (i = 1, 2, \cdots, K)$

(b) $\sum_{x_1} \cdots \sum_{x_K} f(x_1, \cdots, x_K) = 1$

(c) $\sum_{x_1 : a_1 \leq x_1 \leq b_1} \cdots \sum_{x_K : a_K \leq x_K \leq b_K} f(x_1, \cdots, x_K) = P(a_1 \leq X_1 \leq b_1, \cdots, a_K \leq X_K \leq b_K)$

Definition: Continuous Multivariate Joint pdf:

(a) $f(x_1, \cdots, x_K) \geq 0 \quad \forall x_i : -\infty < x_i < +\infty \quad (i = 1, 2, \cdots K)$

(b) $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \cdots, x_K)\, dx_K \cdots dx_1 = 1$

(c) $\int_{a_1}^{b_1} \cdots \int_{a_K}^{b_K} f(x_1, \cdots, x_K)\, dx_K \cdots dx_1 = P(a_1 \leq X_1 \leq b_1, \cdots, a_K \leq X_K \leq b_K)$

Ex. 2.5.1

Choose four balls randomly from a box containing 2 white balls, 3 black balls, 4 red balls, and 5 yellow balls. Let X_1, X_2 and X_3 be the number of white balls, black balls and red balls respectively. The joint pdf of $(X_1, X_2, X_3)^t$ is

$$f(x_1, x_2, x_3) = \frac{\binom{2}{x_1}\binom{3}{x_2}\binom{4}{x_3}\binom{5}{4 - x_1 - x_2 - x_3}}{\binom{14}{4}}$$

$x_1 = 0,1,2 \quad x_2 = 0,1,2,3 \quad x_3 = 0,1,2,3,4 \quad x_1 + x_2 + x_3 \leq 4$

Ex. 2.5.2

Find the value of c for the following function to be a joint pdf

$$f(x_1, x_2, x_3) = c e^{-x_1 - x_2 - x_3} I_{(0 \leq x_1 \leq x_2 \leq x_3)}$$

$$\int_0^\infty \int_0^{x_3} \int_0^{x_2} c e^{-x_1 - x_2 - x_3}\, dx_1 dx_2 dx_3$$

$$= \int_0^\infty c e^{-x_3} \int_0^{x_3} e^{-x_2} \int_0^{x_2} e^{-x_1}\, dx_1 dx_2 dx_3$$

$$= \int_0^\infty c e^{-x_3} \int_0^{x_3} e^{-x_2} \left[-e^{-x_1}\right]_0^{x_2} dx_2 dx_3$$

$$= \int_0^\infty c e^{-x_3} \int_0^{x_3} (e^{-x_2} - e^{-2x_2})\, dx_2 dx_3$$

$$= \int_0^\infty c e^{-x_3} \left[-e^{-x_2} + \frac{1}{2} e^{-2x_2}\right]_0^{x_3} dx_3$$

$$= \int_0^\infty c e^{-x_3} \left[-e^{-x_3} + \frac{1}{2} e^{-2x_3} + 1 - \frac{1}{2}\right] dx_3$$

$$= c \int_0^\infty -e^{-2x_3} + \frac{1}{2} e^{-3x_3} + \frac{1}{2} e^{-x_3}\, dx_3$$

$$= c \left[\frac{1}{2} e^{-2x_3} - \frac{1}{6} e^{-3x_3} - \frac{1}{2} e^{-x_3}\right]_0^\infty$$

$$= c \left(\frac{1}{6}\right) \quad \therefore c = 6 \quad \checkmark\checkmark$$

Definition: Marginal pdf

If the pdf of $(X_1, X_2, \cdots, X_K)^t$ is $f(x_1, \cdots, x_K)$,

(a) $f_1(x_1) = \begin{cases} \sum_{x_2} \cdots \sum_{x_K} f(x_1, x_2, \cdots, x_K) & \text{(discrete)} \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \cdots, x_K)\, dx_2 \cdots dx_K & \text{(continuous)} \end{cases}$

(b) $f_{1,2}(x_1, x_2) = \begin{cases} \sum_{x_3} \cdots \sum_{x_K} f(x_1, \cdots, x_K) \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, x_3, \cdots, x_K)\, dx_3 \cdots dx_K \end{cases}$

Ex. 2.5.3

The marginal pdf of $(X_1, X_2)^t$ for the pdf given in Ex. 2.5.1 is:

$$f_{1,2}(x_1, x_2) = \sum_{x_3} \binom{2}{x_1}\binom{3}{x_2}\binom{4}{x_3}\binom{5}{4 - x_1 - x_2 - x_3} \Big/ \binom{14}{4}$$

$$= \binom{2}{x_1}\binom{3}{x_2} \sum_{x_3} \binom{4}{x_3}\binom{5}{4 - x_1 - x_2 - x_3} \Big/ \binom{14}{4}$$

$$= \binom{2}{x_1}\binom{3}{x_2}\binom{9}{4 - x_1 - x_2} \Big/ \binom{14}{4} \quad * \sum_{y=0}^{l} \binom{m}{l}\binom{n}{l-y} = \binom{m+n}{l}$$

$x = 0,1,2 \quad y = 0,1,2,3 \quad x + y \leq 4$

Ex 2.5.4.

The joint pdf of $(X_1, X_2, X_3)^t$ in Ex. 2.5.2 is

$$f(x_1, x_2, x_3) = 6 e^{-x_1 - x_2 - x_3} I_{(0 \leq x_1 \leq x_2 \leq x_3)}$$

The marginal pdf of $(X_1, X_3)^t$ is

$$\int_{x_1}^{x_3} 6 e^{-x_1 - x_2 - x_3}\, dx_2 = \left[-6 e^{-x_1 - x_2 - x_3}\right]_{x_1}^{x_3}$$
$$= 6 e^{-x_1 - x_3} (e^{-x_1} - e^{-x_3}) I_{(0 \leq x_1 \leq x_3)} \quad \checkmark\checkmark$$

The marginal pdf for (X_1) is

$$\int_{x_1}^{\infty} 6 e^{-2x_1 - x_3} - 6 e^{-x_1 - 2x_3}\, dx_3$$

$$= \left[-6 e^{-2x_1 - x_3} + 3 e^{-x_1 - 2x_3}\right]_{x_1}^{\infty}$$
$$= 6 e^{-3x_1} - 3 e^{-3x_1} = 3 e^{-3x_1} I_{(0 \leq x_1)} \quad \checkmark\checkmark$$

Similar to the single variable case, in the multivariate case there is one-to-one mapping between pdfs and cdfs. For example, the following holds:

$$F(x_1,\cdots,x_K) = P(X_1 \le x_1,\cdots,X_K \le x_K) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_K} f(t_1,\cdots,t_K)\,dt_K\cdots dt_1$$

$$f(t_1,\cdots,t_K) = \frac{\partial^K}{\partial x_1\cdots\partial x_K} F(x_1,\cdots,x_K) \text{ on point } (x_1,\cdots,x_K) \text{ that are Continuous on } f$$

Definition: Expectation of a function for the multivariate case.

If the joint pdf of $(X_1,\cdots,X_K)^t$ is $f(x_1,\cdots,x_K)$, for a real value function $g(x_1,\cdots,x_K)$

$$E[g(X_1,\cdots,X_K)] = \begin{cases} \sum_{x_1}\cdots\sum_{x_K} g(x_1,\cdots,x_K) f(x_1\cdots x_K) & \text{(discrete)} \\ \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x_1,\cdots,x_K) f(x_1,\cdots,x_K)\,dx_K\cdots dx_1 & \text{(continuous)} \end{cases}$$

$E[g(X_1,\cdots,X_K)]$ is called the expectation of $g(X_1,\cdots,X_K)$

Ex. 2.5.5

If the pdf of $(X_1,X_2,X_3)^t$ is

$f(x_1,x_2,x_3) = 120\,x_1(1-x_1-x_2-x_3)\,I_{\{x_1\ge 0,\, x_2\ge 0,\, x_3\ge 0,\, x_1+x_2+x_3\le 1\}}$

Find $E[X_1X_2]$.

$$E[X_1X_2] = \int_0^1\int_0^{1-x_1}\int_0^{1-x_1-x_2} x_1x_2\cdot 120\,x_1(1-x_1-x_2-x_3)\,dx_3\,dx_2\,dx_1$$

$$= 120\int_0^1 x_1^2\int_0^{1-x_1} x_2\int_0^{1-x_1-x_2} 1-x_1-x_2-x_3\,dx_3\,dx_2\,dx_1$$

$$= 120\int_0^1 x_1^2\int_0^{1-x_1} x_2\left[(1-x_1-x_2)x_3-\tfrac{1}{2}x_3^2\right]_0^{1-x_1-x_2} dx_2\,dx_1$$

$$= 120\int_0^1 x_1^2\int_0^{1-x_1} x_2\,(1-x_1-x_2)^2/2\,dx_2\,dx_1$$

$$= 60\int_0^1 x_1^2\int_0^{1-x_1} x_2(1-x_1-x_2)^2\,dx_2\,dx_1 \quad \text{* integration by parts}$$

$$= 60\int_0^1 \left\{\left[-\tfrac{1}{3}x_2(1-x_1-x_2)^3\right]_0^{1-x_1} - \int_0^{1-x_1} -\tfrac{1}{3}(1-x_1-x_2)^3\,dx_2\right\} dx_1$$

$$= 60\int_0^1 x_1^2\left\{0-\left[\tfrac{1}{12}(1-x_1-x_2)^4\right]_0^{1-x_1}\right\} dx_1$$

$$= \int_0^1 5x_1^2(1-x_1)^4\,dx_1$$

$$= \left[5x_1^2\cdot-\tfrac{1}{5}(1-x_1)^5\right]_0^1 - \int_0^1 10x_1\cdot-\tfrac{1}{5}\cdot(1-x_1)^5\,dx_1$$

$$= 2\int_0^1 x_1(1-x_1)^5\,dx_1 = 2\left\{\left[x_1\cdot-\tfrac{1}{6}(1-x_1)^6\right]_0^1 - \int_0^1 -\tfrac{1}{6}(1-x_1)^6\,dx_1\right\}$$

$$= \tfrac{1}{3}\int_0^1 (1-x_1)^6\,dx_1 = \tfrac{1}{3}\left[-\tfrac{1}{7}(1-x_1)^7\right]_0^1 = \tfrac{1}{21}$$ ✓✓

⟨Theorem 2.5.1 Properties of Expectation (Multivariate Case)⟩

(a) (linearity) for arbitrary constants c_1, c_2

$$E[c_1g_1(X_1,\cdots,X_K) + c_2g_2(X_1,\cdots,X_K)] = c_1E[g_1(X_1,\cdots,X_K)] + c_2E[g_2(X_1,\cdots,X_K)]$$

(b) (monotonicity) If $g_1(X_1,\cdots,X_K) \le g_2(X_1,\cdots,X_K)$,

$$E[g_1(X_1,\cdots,X_K)] \le E[g_2(X_1,\cdots,X_K)]$$

Definition: Random Vector[35] Mean and Covariance Matrix.

Mean vector / variance-covariance matrix's elements are the mean / variance·covariance of $X_1,\cdots,X_K$

$$E[X] = (\mu_1,\cdots,\mu_K)^t = (E[X_1],\cdots,E[X_K])^t$$

$$Var(X) = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,K} \\ \vdots & \ddots & \vdots \\ \sigma_{K,1} & \cdots & \sigma_{K,K} \end{pmatrix} = (Cov(X_i,X_j))_{\substack{1\le i\le K \\ 1\le j\le K}}$$

*35) 벡터의 기호로는 $\underline{X}$, $\underline{a}$, $\underline{\mu}$ 와 같이 나타내어 실수 값의 X, a, μ 와 구분하기도 하지만 이 책에서는 같은 기호를 사용하기로 한다. 따라서 문장이나 문단의 맥락에서 X, a, μ 가 벡터 또는 실수인가를 판단하여야 한다.

Ex. 2.5.6

From Ex 2.2.1 the random vector $(X_1,X_2)^t$'s joint pdf is

$$f(x,y) = 120\,xy(1-x-y)\,I_{\{x\ge 0,\, y\ge 0,\, x+y\le 1\}}$$

From Ex 2.2.2, Ex. 2.2.3, Ex. 2.2.4 we know that

$$E[X] = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix},\quad Var(X) = \begin{pmatrix} 2/63 & -1/63 \\ -1/63 & 2/63 \end{pmatrix}$$

Ex. 2.5.7

In Ex. 2.1.2, the joint pdf of the random vector is:

$$f(x,y) = 2e^{-x-y}\,I_{\{0\le x\le y\}}$$

From Ex. 2.2.5,

$$E[X] = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix},\quad Var(X) = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 5/4 \end{pmatrix}$$

⟨Theorem 2.5.2. Properties of the Expectation of Random Matrix⟩

The following holds for random matrices $V=(V_{ij})$ and $W=(W_{ij})$

(a) $E[CWD] = CE[W]D$ (C and D are constant matrices)

(b) $E[V+W] = E[V] + E[W]$

Let the mean of the random vector $X=(X_1,\cdots,X_K)^t$ be $\mu=(\mu_1,\cdots,\mu_K)^t$

Then, $(X-\mu)(X-\mu)^t = ((X_i-\mu_i)(X_j-\mu_j))$

Hence we can express the covariance matrix of X as

$$Var(X) = (Cov(X_i,X_j)) = (E[(X_i-\mu_i)(X_j-\mu_j)])$$
$$= E[(X-\mu)(X-\mu)^t]$$

Let $Y=(Y_1,\cdots,Y_L)^t$ and its mean vector be $\eta=(\eta_1,\cdots,\eta_L)^t$

The covariance matrix $Cov(X,Y)=(Cov(X_i,Y_j))_{1\le i\le K,\ 1\le j\le L}$ $(K\times L)$ is:

$$Cov(X,Y)=(Cov(X_i,Y_j))=(E[(X_i-\mu_i)(Y_j-\eta_j)])=E[(X-\mu)(Y-\eta)^t]$$

〈Theorem 2.5.3. Properties of mean vector and Covariance Matrix〉

(a) $E[AX+b]=AE[X]+b$ (A: constant matrix, b: constant vector)

(b) $Var(AX+b)=AVar(X)A^t$ (〃)

(c) $Cov(AX+b,CY+d)=ACov(X,Y)C^t$ (A,C: matrix; b,d: vector)

(d) $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$

$Cov(X,Z+W)=Cov(X,Z)+Cov(X,W)$

(e) $Cov(Y,X)=(Cov(X,Y))^t$, $Var(X)=Cov(X,X)$

(f) $Var(X+Y)=Var(X)+Var(Y)+Cov(X,Y)+Cov(Y,X)$

(proof) 기대값의 성질과 공분산행렬의 정의로부터 명백하다.

(b) $Var(AX+b)=E[(AX+b-A\mu-b)(AX+b-A\mu-b)^t]$

$=E[A(X-\mu)(X-\mu)^tA^t]$

$=AE[(X-\mu)(X-\mu)^t]A^t$

$=AVar(X)A^t$

Ex. 2.5.8

The pdf of $X=(X_1,X_2)^t$ from Ex.2.1.a is $f(x,y)=2e^{-x-y}I_{(0\le x\le y)}$

From Ex. 2.5.7, we know that

$$E[X]=\begin{pmatrix}1/2\\3/2\end{pmatrix},\quad Var(X)=\begin{pmatrix}1/4 & 1/4\\1/4 & 5/4\end{pmatrix}$$

If we let $Y=(Y_1,Y_2)^t=(2X_1,X_2-X_1)^t$

$$Y=AX,\quad A=\begin{pmatrix}2 & 0\\-1 & 1\end{pmatrix}$$

Thus, the expectation and variance of Y are

$$E[Y]=AE[X]=\begin{pmatrix}2 & 0\\-1 & 1\end{pmatrix}\begin{pmatrix}1/2\\3/2\end{pmatrix}=\begin{pmatrix}1\\1\end{pmatrix}\ \checkmark\checkmark$$

$$Var(Y)=AVar(X)A^t=\begin{pmatrix}2 & 0\\-1 & 1\end{pmatrix}\begin{pmatrix}1/4 & 1/4\\1/4 & 5/4\end{pmatrix}\begin{pmatrix}2 & -1\\0 & 1\end{pmatrix}$$

$$=\begin{pmatrix}1/2 & 1/2\\0 & 1\end{pmatrix}\begin{pmatrix}2 & -1\\0 & 1\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\ \checkmark\checkmark$$

〈Theorem 2.5.4. Properties of the Covariance Matrix〉

Variance-Covariance matrix ($Var(X)$) of the random vector $X=(X_1,\cdots,X_K)^t$ is a nonnegative definite symmetric matrix, meaning

$$Var(X)=Var(X)^t,\quad a^tVar(X)a\ge 0\quad \forall a\in\mathbb{R}^K$$

(proof) It is trivial that $Var(X)$ is symmetrical.

a^tX is a real value random variable (1 dim) $\quad a^tX=\sum_{i=1}^{K}a_iX_i$

$$a^tVar(X)a=Var(a^tX)\ge 0\quad \forall a\in\mathbb{R}^K$$

Variance Matrix is a nonnegative definite symmetric matrix (음이 아닌 정부호의 대칭행렬). From p.500 diagonalization Theorem and p.504's condition of nonnegative definite symmetric matrix,

we know that an orthogonal matrix

$$P=(p_1,\cdots,p_K)^t,\quad p_i^tp_i=1,\ p_i^tp_j=0\ (i\ne j)$$

and the eigenvalues $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_K\ge 0$ exists s.t.

$$Var(X)=P\,diag(\lambda_i)P^t,\quad PP^t=P^tP=I,\quad \lambda_1\ge\lambda_2\ge\cdots\ge\lambda_K\ge 0$$

Thus, $Var(P^tX)=diag(\lambda_i)$, $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_K\ge 0$

In P^tX*, the correlation between the components are 0 and their variances are $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_K\ge 0$.

The component in P^tX that corresponds to λ_i is called the ith principal component

* P^tX is a coordinate variable (좌표계의 변수) of X expressed using the basis (기저) of $P=(p_1,p_2,\cdots,p_K)^t$

From the preceding explanation, the eigenvectors and eigenvalues of the variance matrix shows the distribution's direction and magnitude of dispersion (분포의 주된 방향과 산포의 크기)

Definition: Joint Moment of Multivariate Random Variable

If $E[|X_1^{r_1}\cdots X_K^{r_K}|]<+\infty$

$$E[X_1^{r_1}\cdots X_K^{r_K}]=\begin{cases}\sum_{x_1}\cdots\sum_{x_K}x_1^{r_1}\cdots x_K^{r_K}f(x_1,\cdots,x_K) & \text{(discrete)}\\ \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}x_1^{r_1}\cdots x_K^{r_K}f(x_1\cdots x_K)\,dx_K\cdots dx_1 & \text{(continuous)}\end{cases}$$

is called the $(r_1,\cdots,r_K)^{th}$ joint moment of $X=(X_1,\cdots,X_K)^t$ and denote it as $m_{r_1\cdots r_K}(X_1,\cdots,X_K)$ or $m_{r_1\cdots r_K}$

Definition: Joint MGF of Multivariate Random Variable

For values of $t_1,\cdots,t_K$ in an open interval containing 0, in $E[e^{t_1X_1+\cdots+t_KX_K}]\in\mathbb{R}$, the function

$$M(t_1,\cdots,t_K)=E[e^{t_1X_1+\cdots+t_KX_K}],\quad -h_i<t_i<h_i\ (\exists h_i>0)\ [i=1,2,\cdots,K]$$

is called the joint MGF of $X=(X_1,\cdots,X_K)^t$

〈Theorem 2.5.5. Properties of joint MGF〉

(a) [Generation of joint moments]

If the joint MGF for $X=(X_1,\cdots,X_k)^t$ exists, meaning,

$$M(t_1,\cdots,t_k)=E\left(e^{t_1X_1+\cdots+t_kX_k}\right)<+\infty,\quad -h_i<t_i<h_i,\ (\exists h_i>0)\ (i=1,2,\cdots,k)$$

the all joint moments of $X=(X_1,\cdots,X_k)^t$ exists

$$E[X_1^{r_1}\cdots X_k^{r_k}]=\left[\frac{\partial^{r_1+\cdots+r_k}}{\partial t_1^{r_1}\cdots\partial t_k^{r_k}}M(t_1,\cdots,t_k)\right]_{t_1=\cdots=t_k=0}$$

$$M(t_1,\cdots,t_k)=\sum_{r_1=0}^{\infty}\cdots\sum_{r_k=0}^{\infty}\frac{E[X_1^{r_1}\cdots X_k^{r_k}]}{r_1!\cdots r_k!}t_1^{r_1}\cdots t_k^{r_k}$$

$$-h_i<t_i<h_i\ (\exists h_i>0)\quad (i=1,\cdots,k)$$

(b) [Distribution Deterministic : 분포 결정성]

If the joint MGFs of $X=(X_1,\cdots,X_k)^t$ and $Y=(Y_1,\cdots,Y_k)^t$ exists and are equal over an open interval containing 0, meaning:

$$M_{X_1,\cdots,X_k}(t_1,\cdots,t_k)=M_{Y_1,\cdots,Y_k}(t_1,\cdots,t_k)$$

their joint pdfs and cdfs are equal.

Similar to the bivariate case, if the joint MGF exists,

$$C(t_1,\cdots,t_k)=\log M(t_1,\cdots,t_k)=\log E[e^{t_1X_1+\cdots+t_kX_k}]$$
$$-h_i<t_i<t_i\ (\exists h_i>0)\ (i=1,2,\cdots,k)$$

is called the joint cumulant generating function (cgf)

We can find the expectation and the variance-covariance matrix by finding the derivative or the power series expansion.

$$\left[\frac{\partial}{\partial t_i}C(t_1,\cdots,t_k)\right]_{t_1=\cdots=t_k=0}=E[X_i]$$

$$\left[\frac{\partial^2}{\partial t_i\partial t_j}C(t_1,\cdots,t_k)\right]_{t_1=\cdots=t_k=0}=\mathrm{Cov}(X_i,X_j)$$

Definition: Mean and Variance of joint cgf:

$$C(t)=\log(M(t))=\log E[e^{t_1X_1+\cdots+t_kX_k}],\quad -h_i<t_i<h_i\ (\exists h_i>0)\ i=1,\cdots,k)$$

Let $\dot{C}(t)$ and $\ddot{C}(t)$ be the first and second derivative of $C(t)$ respectively. (* 38, 39)

$$C(t)=C(0)^t+\frac{1}{2}t^t\ddot{C}(0)t+\cdots,\quad E[X]=\dot{C}(0),\quad \mathrm{Var}(X)=\ddot{C}(0)$$

* 38) $\dot{C}(t)=\left(\frac{\partial}{\partial t_1}C(t),\cdots,\frac{\partial}{\partial t_k}C(t)\right)^t$ is called the "gradient vector".

* 39) $\ddot{C}(t)=\left(\frac{\partial^2}{\partial t_i\partial t_j}C(t)\right)$ is called the Hessian matrix.

Ex. 2.5.9.

In Ex. 2.5.2 the pdf of the random vector $X=(X_1,X_2,X_3)^t$ is

$$f(x_1,x_2,x_3)=6e^{-x_1-x_2-x_3}I_{(0\le x_1\le x_2\le x_3)}$$

the joint mgf is:

$$M_X(t)=\iiint_{0\le x_1\le x_2\le x_3}e^{t_1x_1+t_2x_2+t_3x_3}\cdot 6e^{-x_1-x_2-x_3}dx_3dx_2dx_1$$

$$=6\int e^{-(1-t_1)x_1}\int e^{-(1-t_2)x_2}\int e^{-(1-t_3)x_3}dx_3dx_2dx_1$$

$$=6\int e^{-(1-t_1)x_1}\int e^{-(1-t_2)x_2}\left[-\frac{1}{(1-t_3)}e^{-(1-t_3)x_3}\right]_{x_2}^{\infty}$$

$$=6\int e^{-(1-t_1)x_1}\int e^{-(1-t_2)x_2}(1-t_3)^{-1}e^{-(1-t_3)x_2}dx_2dx_1$$

$$=6\int e^{-(1-t_1)x_1}\left[-\frac{1}{(1-t_3)(2-t_2-t_3)}e^{-(2-t_2-t_3)x_2}\right]_{x_1}^{\infty}dx_1$$

$$=6\int e^{-(1-t_1)x_1}(1-t_3)^{-1}(2-t_2-t_3)^{-1}e^{-(2-t_2-t_3)x_1}dx_1$$

$$=6(1-t_3)^{-1}(2-t_2-t_3)^{-1}\int_0^{\infty}e^{-(3-t_1-t_2-t_3)x_1}dx_1$$

$$=6(1-t_3)^{-1}(2-t_2-t_3)^{-1}\left[-\frac{1}{3-t_1-t_2-t_3}e^{-(3-t_1-t_2-t_3)x_1}\right]_0^{\infty}$$

$$=6(1-t_3)^{-1}(2-t_2-t_3)^{-1}(3-t_1-t_2-t_3)^{-1}$$

$t_1+t_2+t_3<3$, $t_2+t_3<2$, $t_3<1$

$$C(t)=\log M_X(t)$$
$$=\log 6(1-t_3)^{-1}(2-t_2-t_3)^{-1}(3-t_1-t_2-t_3)^{-1}$$
$$=\log(1-t_3)^{-1}\left(1-\frac{t_2+t_3}{2}\right)^{-1}\left(1-\frac{t_1+t_2+t_3}{3}\right)^{-1}$$
$$=-\ln(1-t_3)-\ln\left(1-\frac{t_2+t_3}{2}\right)-\ln\left(1-\frac{t_1+t_2+t_3}{3}\right)$$

* cgf 구하는 연습 필요. 중요!

* $-\log(1-A)=A+A^2/2+A^3/3+\cdots$

$$\left(t_3+t_3^2/2+\cdots\right)+\left\{\frac{t_2+t_3}{2}+\frac{(t_2+t_3)^2}{4\cdot 2}+\cdots\right\}+\left\{\frac{t_1+t_2+t_3}{3}+\cdots\right\}$$

$$=t_1/3+5t_2/6+11/6\,t_3+\cdots$$

$$(\cdots)=t_3^2/2+\frac{t_2^2+2t_2t_3+t_3^2}{8}+\frac{t_1^2+t_2^2+t_3^2+2t_1t_2+2t_2t_3+2t_1t_3}{18}+\cdots$$

$$=\left(\frac{1}{2}+\frac{1}{8}+\frac{1}{18}\right)t_3^2+\left(\frac{1}{8}+\frac{1}{18}\right)t_2^2+\left(\frac{1}{18}\right)t_1^2+\frac{13}{36}t_2t_3+\frac{1}{9}t_1t_2+\frac{1}{9}t_1t_3$$

$$=\frac{49}{72}t_3^2+\frac{13}{72}t_2^2+\frac{1}{18}t_1^2+\frac{13}{36}t_2t_3+\frac{1}{9}t_1t_2+\frac{1}{9}t_1t_3\ \checkmark\checkmark$$

$$=\frac{1}{2}\left(\frac{1}{9}t_1^2+\frac{13}{36}t_2^2+\frac{49}{36}t_3^2+\frac{1}{9}2t_1t_2+\frac{1}{9}2t_1t_3+\frac{13}{36}2t_2t_3\right)$$

↳ $2\frac{2}{1}\frac{4}{2}$ 이후 $\frac{9}{11}$

$$\therefore E[X]=\begin{pmatrix}C_{1,0,0}\\C_{0,1,0}\\C_{0,0,1}\end{pmatrix}=\begin{pmatrix}1/3\\5/6\\11/6\end{pmatrix}\checkmark\checkmark$$

$\sum\frac{C_{s,r}}{s!\,r!}t_1^s t_2^r$ look at Ex. 2.2.25

$$\mathrm{Var}(X)=\begin{bmatrix}C_{2,0,0}&C_{1,1,0}&C_{1,0,1}\\C_{1,1,0}&C_{0,2,0}&C_{0,1,1}\\C_{1,0,1}&C_{0,1,1}&C_{0,0,2}\end{bmatrix}=\begin{bmatrix}1/9&1/9&1/9\\1/9&13/36&13/36\\1/9&13/36&49/36\end{bmatrix}\checkmark\checkmark$$

Definition: Conditional pdf of Multivariate Random Variable.

For multivariate random variables X and Y, if the joint pdf is $f_{X,Y}(x,y)$ and the marginal pdf of X is $f_X(x)$, the conditional pdf of Y given $X=x$ is:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \quad (x : f_X(x) > 0)$$

Ex. 2.5.10

In Ex. 2.5.5, the joint pdf of the random vector $(X_1, X_2, X_3)^t$ is

$$f(x_1,x_2,x_3) = 120\, x_1 (1-x_1-x_2-x_3) I_{(x_1\ge 0,\ x_2\ge 0,\ x_3\ge 0,\ x_1+x_2+x_3\le 1)}$$

The marginal pdfs of X_1 and $(X_2, X_3)^t$ are:

$$f_1(x_1) = \int_0^{1-x_1}\int_0^{1-x_1-x_2} 120\, x_1(1-x_1-x_2-x_3)\, dx_3\, dx_2$$

$$= 120\, x_1 \int_0^{1-x_1} \left[(1-x_1-x_2)x_3 - \frac{1}{2}x_3^2 \right]_0^{1-x_1-x_2} dx_2$$

$$= 120\, x_1 \int_0^{1-x_1} (1-x_1-x_2)^2 \cdot 1/2 \; dx_2$$

$$= 120\, x_1 \left[-\frac{1}{6}(1-x_1-x_2)^3 \right]_0^{1-x_1}$$

$$= 20\, x_1 (1-x_1)^3 I_{(0\le x_1\le 1)}$$ ✓✓

$$f_{2,3}(x_2,x_3) = \int_0^{1-x_2-x_3} 120\, x_1(1-x_1-x_2-x_3)\, dx_1$$

$$= \left[-40x_1^3 + 60x_1^2(1-x_2-x_3) \right]_0^{1-x_2-x_3}$$

$$= 20(1-x_2-x_3)^3 I_{(0\le x_2,\ 0\le x_3,\ x_2+x_3\le 1)}$$ ✓✓

Thus, $f_{2,3|1}(x_2,x_3|x_1) = 6(1-x_1-x_2-x_3)(1-x_3)^{-3} I_{(0\le x_2,\ 0\le x_3,\ x_2+x_3\le 1-x_1)}$

$f_{1|2,3}(x_1|x_2,x_3) = 6x_1(1-x_1-x_2-x_3)(1-x_2-x_3)^{-3} I_{(0\le x_1\le 1-x_2-x_3)}$

Definition: Conditional Expectation of Random Vector [40]

$$E[g(X,Y)|X=x] = \begin{cases} \sum_{y_1}\cdots\sum_{y_L} g(x,y_1,\cdots,y_L) f_{Y|X}(y_1,\cdots,y_L|x) & \text{(discrete)} \\ \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(x,y_1,\cdots,y_L) f_{Y|X}(y_1,\cdots,y_L|x)\, dy_L\cdots dy_1 & \text{(continuous)} \end{cases}$$

For multivariate random variable $X=(X_1,\cdots,X_K)^t$, $Y=(Y_1,\cdots,Y_L)^t$,
If the conditional pdf of Y given $X=x$ is $f_{Y|X}(y|x)$, the expectation of real value function $g(X,Y)$ is $E[g(X,Y)|X=x]$

40) 조건부 기대값의 정의에서 오른쪽 변의 합이나 적분은 절대수렴하는 것을 전제로 하는 것이다.

Meanwhile, for multivariate random variable $X=(X_1,\cdots,X_K)^t$ $Y=(Y_1,\cdots,Y_L)^t$, the conditional mean and Conditional variance-Covariance matrix is expressed as:

$$E[Y|X] = (E(Y_1|X),\cdots,E(Y_L|X))^t$$

$$Var(Y|X) = (Cov(Y_i,Y_j|X))_{\substack{1\le i\le L\\ 1\le j\le L}}$$

Thus, the properties of conditional expectation holds if X and Y are multivariate random variables(다차원 확률변수)

⟨Theorem 2.5.6. Properties of Conditional Expectation⟩

(a) $E[E[Y|X]] = E[Y]$

(b) $Cov[Y-E[Y|X], v(X)] = 0 \quad \forall v(X)$

Ex. 2.5.11

In Ex. 2.5.10 the Conditional probability of $(X_2,X_3)^t$ given $X_1=x_1$ is

$$f_{2,3|1}(x_2,x_3|x_1) = \frac{f_{1,2,3}(x_1,x_2,x_3)}{f_1(x_1)}$$

$$= \frac{6(1-x_1-x_2-x_3)}{(1-x_1)^3} I_{(x_2\ge 0, x_3\ge 0,\ x_2+x_3\le 1-x_1)}$$

Thus, $E[X_2|X_1=x_2] = \int_0^{1-x_1}\int_0^{1-x_1-x_2} x_2 \cdot 6(1-x_1-x_2-x_3)\, dx_3\, dx_2 (1-x_1)^{-3}$

$$= \int_0^{1-x_1} \left[6x_2(1-x_1-x_2)x_3 - 3x_2x_3^2 \right]_0^{1-x_1-x_2} dx_2 (1-x_1)^{-3}$$

$$= \int_0^{1-x_1} 3x_2(1-x_1-x_2)^2\, dx_2 (1-x_1)^{-3}$$

$$= \left[-x_2(1-x_1-x_2)^3 \right]_0^{1-x_1} - \int_0^{1-x_1} -(1-x_1-x_2)^3 dx_2 (1-x_1)^{-3}$$

$$= \left[-\frac{1}{4}(1-x_1-x_2)^4 \right]_0^{1-x_1} (1-x_1)^{-3}$$

$$= \frac{1}{4}(1-x_1)^4 (1-x_1)^{-3} = (1-x_1)/4$$ ✓✓

$E[X_2^2|X_1=x_1] = \cdots = \int_0^{1-x_1} 3x_2^2(1-x_1-x_2)^3 dx_2 (1-x_1)^{-3}$

$$= \left\{ \left[3x_2^2 \cdot -\frac{1}{3}(1-x_1-x_2)^3 \right]_0^{1-x_1} - \int_0^{1-x_1} 6x_2(1-x_1-x_2)^3 \cdot \left(-\frac{1}{3}\right) dx_2 \right\} (1-x_1)^{-3}$$

$$= \int_0^{1-x_1} 2x_2(1-x_1-x_2)^3 \cdot dx_2 (1-x_1)^{-3}$$

$$= \left\{ \left[2x_2\left(-\frac{1}{4}\right)\cdot(1-x_1-x_2)^4 \right]_0^{1-x_1} - \int_0^{1-x_1} -\frac{1}{2}(1-x_1-x_2)^4 dx_2 \right\} (1-x_1)^{-3}$$

$$= \left[\frac{1}{2}\cdot -\frac{1}{5}(1-x_1-x_2)^5 \right]_0^{1-x_1} (1-x_1)^{-3}$$

$$= \frac{1}{10}(1-x_1)^2$$ ✓✓

$$E[X_2X_3|X_1=x_1] = \int_0^{1-x_1}\int_0^{1-x_1-x_2} x_2x_3 \cdot 6(1-x_1-x_2-x_3)\, dx_3\, dx_2 (1-x_1)^{-3}$$

$$= \int_0^{1-x_1} x_2 \left[3(1-x_1-x_2)x_3^2 - 2x_3^3 \right]_0^{1-x_1-x_2} dx_2 (1-x_1)^{-3}$$

$$= \int_0^{1-x_1} x_2(1-x_1-x_2)^3\, dx_2 (1-x_1)^{-3}$$

$$= \int_0^{1-x_1} \frac{1}{4}(1-x_1-x_2)^4 dx_2 (1-x_1)^{-3}$$

$$= \left[-\frac{1}{20}(1-x_1-x_2)^5 \right]_0^{1-x_1} (1-x_1)^{-3} = (1-x_1)^2/20$$ ✓✓

$$\mathrm{Var}(X_2|X_1=x) = E[X_2^2|X_1=x] - E[X_2|X_1=x]^2$$
$$= (1-x_1)^2/10 - [(1-x_1)/4]^2$$
$$= \frac{(16-10)(1-x_1)^2}{160} = \frac{3}{80}(1-x_1)^2 \checkmark\checkmark$$

$$\mathrm{Cov}(X_2X_3|X_1=x_1) = E[X_2X_3|X_1=x_1] - E[X_2|X_1=x_1]E[X_3|X_1=x_1]$$
$$= (1-x_1)^2/20 - (1-x_1)/4 \cdot (1-x_1)/4$$
$$= \frac{(16-20)\cdot(1-x_1)^2}{16\cdot 20} = -\frac{(1-x_1)^2}{80} \checkmark\checkmark$$

$$\therefore E\left[\begin{pmatrix}X_2\\X_3\end{pmatrix}|X_1\right] = \begin{bmatrix}E[X_2|X_1]\\E[X_3|X_1]\end{bmatrix} = \begin{pmatrix}1/4\\1/4\end{pmatrix}(1-X_1) \checkmark\checkmark$$

$$\mathrm{Var}\left[\begin{pmatrix}X_2\\X_3\end{pmatrix}|X_1\right] = \begin{bmatrix}\mathrm{Cov}(X_2,X_2|X_1) & \mathrm{Cov}(X_2,X_3|X_1)\\ \mathrm{Cov}(X_3,X_2|X_1) & \mathrm{Cov}(X_3,X_3|X_1)\end{bmatrix}$$
$$= \frac{1}{80}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}(1-X_1)^2 \checkmark\checkmark$$

〈Theorem 2.5.7. Decomposition of the variance-covariance matrix〉

$$\mathrm{Var}(Y) = \mathrm{Var}(E[Y|X]) + E[\mathrm{Var}(Y|X)]$$

(proof)

$$\mathrm{Var}(Y) = \mathrm{Var}(Y - E[Y|X] + E[Y|X])$$
$$= \mathrm{Var}(Y-E[Y|X]) + \mathrm{Var}(E[Y|X]) + 2\mathrm{Cov}(Y-E[Y|X], E[Y|X])$$
$$= \mathrm{Var}(Y-E[Y|X]) + \mathrm{Var}(E[Y|X]) \quad \because \mathrm{Cov}(Y-E[Y|X], V(X)) = 0$$
$$= E[(Y-E[Y|X])(Y-E[Y|X])^t] + E[Y-E[Y|X]]E[Y-E[Y|X]]^t + \mathrm{Var}(E[Y|X])$$
$$= E[E[(Y-E[Y|X])(Y-E[Y|X])^t|X]] + \mathrm{Var}(E[Y|X])$$
$$= E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$$

[Textbook Proof]
(…)
Let $W = Y - E[Y|X]$
$E[WW^t|X] = E[(Y-E[Y|X])(Y-E[Y|X])^t|X] = \mathrm{Var}(Y|X)$

$\mathrm{Var}(W) = E[WW^t] = E[E[WW^t|X]] = E[\mathrm{Var}(Y|X)]$

$\therefore \mathrm{Var}(Y) = \mathrm{Var}(W) + \mathrm{Var}[E[Y|X]] = E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$

Ex. 2.5.12 Probability of $X \geq 60_K$ (16시그마 밖의 확률)

In Ex 2.5.10, the marginal pdf of $(X_2, X_3)^t$ is:

$$f_{2,3}(x_2,x_3) = 20(1-x_2-x_3)^3 I_{|x_2\geq 0, x_3\geq 0, 0\leq x_2+x_3\leq 1)}$$

Thus the covariance matrix of $Y=(X_2,X_3)^t$ is:

$$\mathrm{Var}(Y) = \begin{bmatrix}\mathrm{Var}(X_2) & \mathrm{Cov}(X_2,X_3)\\ \mathrm{Cov}(X_3,X_2) & \mathrm{Var}(X_3)\end{bmatrix} = \frac{1}{252}\begin{pmatrix}5 & -1\\-1 & 5\end{pmatrix}$$

Meanwhile in Ex. 2.5.11,

$$E[Y|X_1] = \begin{pmatrix}1/4\\1/4\end{pmatrix}(1-X_1),\ \mathrm{Var}(Y|X_1) = \frac{1}{80}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}(1-X_1)^2$$

In Ex 2.5.10, the marginal pdf of X_1 is:

$$f_1(x_1) = 20x_1(1-x_1)^3 I_{|0\leq x_1\leq 1)}$$

Hence, $E[(1-X_1)^2] = 10/21$ $\quad \mathrm{Var}(X_1) = 2/63$

$$E[\mathrm{Var}(Y|X_1)] = E\left[\frac{1}{80}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}(1-X_1)^2\right]$$
$$= \frac{1}{80}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}E[(1-X_1)^2]$$
$$= \frac{1}{80}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix}\cdot\frac{10}{21} = \frac{1}{168}\begin{pmatrix}3 & -1\\-1 & 3\end{pmatrix} \checkmark$$

$$\mathrm{Var}(E[Y|X_1]) = \mathrm{Var}\left(\begin{pmatrix}1/4\\1/4\end{pmatrix}(1-X_1)\right)$$
$$= \begin{pmatrix}1/4\\1/4\end{pmatrix}\mathrm{Var}(1-X_1)\begin{pmatrix}1/4 & 1/4\end{pmatrix}$$
$$= \begin{pmatrix}1/4\\1/4\end{pmatrix}\mathrm{Var}(X_1)\begin{pmatrix}1/4 & 1/4\end{pmatrix}$$
$$= \begin{pmatrix}1/4\\1/4\end{pmatrix}\cdot\frac{2}{63}\cdot\begin{pmatrix}1/4 & 1/4\end{pmatrix}$$
$$= \frac{2}{16\cdot 63}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} = \frac{1}{504}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$$

$$E[\mathrm{Var}(Y|X_1)] + \mathrm{Var}[E[Y|X_1]] = \frac{1}{504}\begin{pmatrix}10 & -2\\-2 & 10\end{pmatrix} = \frac{1}{252}\begin{pmatrix}5 & -1\\-1 & 5\end{pmatrix}$$

$$\therefore \mathrm{Var}(Y) = \frac{1}{252}\begin{pmatrix}5 & -1\\-1 & 5\end{pmatrix}$$

〈Theorem 2.5.8: Mean Squared Error of Multivariate Random Variable〉

Let $u(x) = (u_1(x), \cdots, u_K(x))^t$ be a vector value function (벡터 값 함수) of a multivariate r.v. X.

$$E[\|Y-u(x)\|^2] = E[(Y_1-u_1(x))^2 + \cdots + (Y_K-u_K(x))^2]$$

The function which minimizes the preceding expectation is $E[Y|X] = (E[Y_1|X], \cdots, E[Y_K|X])^t$. i.e.,

$$E[\|Y-E(Y|X)\|^2] \leq E[\|Y-u(x)\|^2],\ \forall u(x)$$

(proof)
$$E[\|Y-u(x)\|^2] = E[(Y-u(x))^t(Y-u(x))]$$
$$= E[(Y-E[Y|X]+E[Y|X]-u(x))^t(Y-E[Y|X]+E[Y|X]-u(x))]$$
$$= E[(Y-E[Y|X])^t(Y-E[Y|X])] + 2E[Y-E[Y|X]](E[Y|X]-u(x)) + E[(E[Y|X]-u(x))^t(E[Y|X]-u(x))]$$
$$= E[\|Y-E[Y|X]\|^2] + 2\{E[Y]-E[Y]\}(E[Y|X]-u(x)) + E[\|E[Y|X]-u(x)\|^2]$$
$$= E[\|Y-E[Y|X]\|^2] + E[\|E[Y|X]-u(x)\|^2]$$
$$\geq E[\|Y-E[Y|X]\|^2]$$

[Textbook] 조건부기대값의 성질 (b)를 이용하여 〈정리 2.3.5〉와 같은 방법으로 증명할 수 있다.

For multiple r.v.s $X_1, X_2, \cdots, X_n$ if
$P(X_1\in A_1, \cdots, X_n\in A_n) = P(X_1\in A_1)P(X_2\in A_2)\cdots P(X_n\in A_n)$ $\forall A_i\,(i=1,2,\cdots,n)$
we say that the random variables $X_1, \cdots, X_n$ are mutually independent.
This is also true for the multivariate case.

〈Theorem 2.5.9. Independence of multivariate random variables〉

The following conditions are necessary and sufficient conditions for $X_1, \cdots, X_n$ to be mutually independent.

(a) (pdf) $pdf_{1,\cdots,n}(x_1,\cdots,x_n) = pdf_1(x_1)\cdots pdf_n(x_n) \quad \forall x_i \ (i=1,\cdots,n)$

(b) (mgf) $mgf_{1,\cdots,n}(t_1,\cdots,t_n) = mgf_1(t_1)\cdots mgf_n(t_n) \quad \forall t_i : \|t_i\| < h_i \ (\exists h_i > 0)(i=1,2,\cdots,n)$

Ex. 2.5.13

In Ex. 2.2.2, the joint pdf of $(X_1, X_2, X_3)^t$ is:

$$f_{1,2,3}(x_1,x_2,x_3) = 6e^{-x_1-x_2-x_3} I_{(0 \le x_1 \le x_2 \le x_3)}$$

The marginal pdfs are

$$f_1(x_1) = \int_{x_1}^{\infty}\int_{x_2}^{\infty} 6e^{-x_1-x_2-x_3}\,dx_3\,dx_2$$
$$= 6e^{-x_1}\int_{x_1}^{\infty} e^{-x_2}\int_{x_2}^{\infty} e^{-x_3}\,dx_3\,dx_2$$
$$= 6e^{-x_1}\int_{x_1}^{\infty} e^{-x_2}\left[-e^{-x_3}\right]_{x_2}^{\infty} dx_2$$
$$= 6e^{-x_1}\int_{x_1}^{\infty} e^{-2x_2}\,dx_2$$
$$= 6e^{-x_1}\left[-\tfrac{1}{2}e^{-2x_2}\right]_{x_1}^{\infty}$$
$$= 3e^{-2x_1} I_{(0 \le x_1)} \quad \checkmark\checkmark$$

$$f_2(x_2) = \int_0^{x_1}\int_{x_2}^{\infty} 6e^{-x_1-x_2-x_3}\,dx_3\,dx_1$$
$$= 6e^{-x_2}\int_0^{x_2} e^{-x_1}\int_{x_2}^{\infty} e^{-x_3}\,dx_3\,dx_1$$
$$= 6e^{-x_2}\int_0^{x_2} e^{-x_1}\left[-e^{-x_3}\right]_{x_2}^{\infty} dx_1$$
$$= 6e^{-2x_2}\int_0^{x_2} e^{-x_1}\,dx_1$$
$$= 6e^{-2x_2}(1-e^{-x_2}) I_{(0 \le x_2)} \quad \checkmark\checkmark$$

$$f_3(x_3) = \int_0^{x_3}\int_0^{x_2} 6e^{-x_1-x_2-x_3}\,dx_1\,dx_2$$
$$= 6e^{-x_3}\int_0^{x_3} e^{-x_2}\int_0^{x_2} e^{-x_1}\,dx_1\,dx_2$$
$$= 6e^{-x_3}\int_0^{x_3} e^{-x_2}\left[-e^{-x_2}\right]_0^{x_2} dx_2$$
$$= 6e^{-x_3}\int_0^{x_3} e^{-x_2}-e^{-2x_2}\,dx_2$$
$$= 6e^{-x_3}\left[-e^{-x_2}+\tfrac{1}{2}e^{-2x_2}\right]_0^{x_3}$$
$$= 6e^{-x_3}\left(\tfrac{1}{2}-e^{-x_3}+\tfrac{1}{2}e^{-2x_3}\right)$$
$$= 3e^{-x_3}(1-2e^{-x_3}+e^{-2x_3})$$
$$= 3e^{-x_3}(1-e^{-x_3})^2 I_{(0 \le x_3)} \quad \checkmark\checkmark$$

$$f(x_1,x_2,x_3) \neq f_1(x_1)f_2(x_2)f(x_3)$$

$\therefore X_1, X_2$ and X_3 are NOT mutually independent.

To show the independence of multivariate random variables, it is not always necessary to find the marginal pdfs.

Instead of the preceding (a), (b), application of (a') and (b') is also possible.

(a') Functions $g_1 \cdots g_n$ that satisfies the following equality exists:

$$pdf_{1,2,\cdots,n}(x_1,x_2,\cdots,x_n) = g_1(x_1)\cdots g_n(x_n) \quad \forall x_i \ (i=1,\cdots,n)$$

(b') Functions $m_1,\cdots,m_n$ that satisfies the following equality exists:

$$mgf_{1,\cdots,n}(t_1\cdots t_n) = m_1(t_1)\cdots m_n(t_n)$$

$$\forall t_i : \|t_i\| < h_i \ (\exists h_i > 0) \ (i=1,2,\cdots,n)$$

Ex. 2.5.14

For the multivariate random vector $X = (X_1, X_2, X_3)^t$ in Ex. 2.5.9, let $Y_1 = X_1$, $Y_2 = X_2 - X_1$ and, $Y_3 = X_3 - X_2$

$$f(x_1,x_2,x_3) = 6e^{-x_1-x_2-x_3} I_{(0 \le x_1 \le x_2 \le x_3)}$$

$$X_1 = Y_1$$
$$X_2 = Y_1 + Y_2$$
$$X_3 = Y_1 + Y_2 + Y_3$$

$$|J| = \left\|\left(\partial x_i/\partial y_j\right)_{\substack{1\le i\le 3\\ 1\le j\le 3}}\right\| = 1$$

$$(0 \le x_1 \le x_2 \le x_3) = (0 \le y_1 \le y_1+y_2 \le y_1+y_2+y_3)$$

$$f(y_1,y_2,y_3)\ 6e^{-3y_1-2y_2-y_3} I_{(0<y_1,\,0<y_2,\,0<y_3)}$$

$$f(y_1) = \int_0^{\infty}\int_0^{\infty} 6e^{-3y_1-2y_2-y_3}\,dy_2\,dy_3$$
$$= \int_0^{\infty}\left[-3e^{-3y_1-2y_2-y_3}\right]_0^{\infty} dy_3$$
$$= \int_0^{\infty} 3e^{-3y_1-y_3}\,dy_3 = 3e^{-3y_1} I_{(0<y_1)}$$

$$f(y_2) = 1\cdots) = 2e^{-2y_2} I_{(0<y_2)}$$

$$f(y_3) = e^{-y_3} I_{(0<y_3)}$$

$\therefore f(y_1,y_2,y_3) = f(y_1)f(y_2)f(y_3)$

$\therefore Y_1, Y_2$ and Y_3 are independent.

[Textbook Solution]

From Ex. 2.5.9, the joint mgf of $(X_1,X_2,X_3)^t$ is:

$$M_X(t_1,t_2,t_3) = E[e^{t_1X_1+t_2X_2+t_3X_3}]$$
$$= 6(3-t_1-t_2-t_3)^{-1}(2-t_2-t_3)^{-1}(1-t_3)^{-1},$$
$$(t_1+t_2+t_3)<3,\ (t_2+t_3)<2,\ t_3<1$$

The mgf of $Y=(Y_1,Y_2,Y_3)^t$ is:

$$\begin{aligned}
M_Y(s_1,s_2,s_3) &= E[e^{s_1Y_1+s_2Y_2+s_3Y_3}] \\
&= E[e^{(s_1-s_2)X_1+(s_2-s_3)X_2+s_3X_3}] \\
&= 6(3-(s_1-s_2)-(s_2-s_3)-s_3)^{-1}(2-(s_2-s_3)-s_3)^{-1}(1-s_3)^{-1} \\
&= 6(3-s_1)^{-1}(2-s_2)^{-1}(1-s_3)^{-1} \\
&= (1-\tfrac{s_1}{3})^{-1}(1-\tfrac{s_2}{2})^{-1}(1-s_3)^{-1}, \quad s_1<3,\ s_2<2,\ s_3<1 \\
&= m_1(s_1)m_2(s_2)m_3(s_3)
\end{aligned}$$

Thus Y_1, Y_2 and Y_3 are mutually independent

You can show the independence of random variables using mgf!

〈Theorem 2.5.10. Properties of independent multivariate random variables〉

(a) (Function of independent r.v.s) If the multivariate r.v.s $X_1,\cdots X_n$ are mutually independent, their respective functions $g_1(X_1)\cdots g_n(X_n)$ are mutually independent also.

(b) (Expectation of the product of independent r.v.s) for independent multivariate r.v.s $X_1,\cdots,X_n$,

$$E[g_1(X_1)\cdots g_n(X_n)] = E[g_1(X_1)]\cdots E[g_n(X_n)]$$

for real value functions $g_1,\cdots,g_n$.

(c) (Independence and correlation) If the multivariate r.v.s are independent,

$$Cov(X_i,X_j)=0 \quad (i\neq j) \quad (i,j=1,\cdots,n)$$

〈Theorem 2.5.11 Sum of mutually independent random variables〉

(a) (Covariance matrix of sum) If the multivariate r.v.s $X_1,\cdots,X_n$ are mutually independent,

$$Var(X_1+\cdots+X_n) = Var(X_1)+\cdots+Var(X_n)$$

(b) (mgf of sum) If the multivariate r.v.s $X_1,\cdots,X_n$ are mutually independent,

$$mgf_{X_1+\cdots+X_n}(t) = mgf_{X_1}(t)\cdots mgf_{X_n}(t)$$

(proof)

(a) From 〈theorem 2.5.3〉-(d)

$$\begin{aligned}
Var(X_1+\cdots+X_n) &= Cov\Big(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\Big) \\
&= \sum_{i=1}^n\sum_{j=1}^n Cov(X_i,X_j)
\end{aligned}$$

Thus, from 〈Theorem 2.5.10〉-(c), we can show that (a) holds.

(b) From the definition of mgf and 〈Theorem 2.5.10〉-(b)

$$\begin{aligned}
mgf_{X_1+\cdots+X_n}(t) &= E[e^{t^t(X_1+\cdots+X_n)}] \\
&= E[e^{t^tX_1}]\cdots E[e^{t^tX_n}] \\
&= mgf_{X_1}(t)\cdots mgf_{X_n}(t)
\end{aligned}$$

* $t=(t_1\cdots t_n)^t$