5. Multiple Random Variables

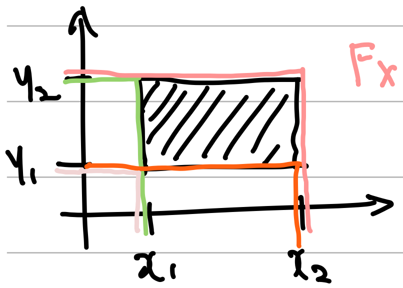
Joint CDF

$$F_{X,Y}(x,y) = P[X \le x,\ Y \le y]$$

- $0 \le F_{X,Y}(x,y) \le 1$
- $F_{X,Y}(\infty,\infty) = 1$
- $F_X(x) = F_{X,Y}(x,\infty)$
- $F_Y(y) = F_{X,Y}(\infty,y)$
- $F_{X,Y}(x,-\infty) = 0$
- $F_{X,Y}(-\infty,y) = 0$
- $x \le x_1$ 이고 $y \le y_1$ 이라면 $F_{XY}(x,y) \le F_{X,Y}(x_1,y_1)$

※ $P[x_1 < X \le x_2,\ y_1 < Y \le y_2]$

$$= F_{XY}(x_2,y_2) - F_{XY}(x_2,y_1) - F_{XY}(x_1,y_2) + F_{XY}(x_1,y_1)$$



Joint PMF

$$P_{XY}(x,y) = P[X=x,\ Y=y]$$

$$P[B] = \sum_{(x,y)\in B} P_{XY}(x,y)$$

Marginal PMF

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x,y)$$

$$P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x,y)$$

Joint PDF

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du$$

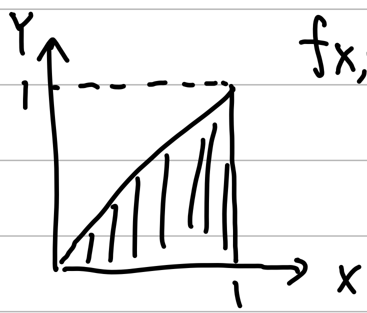
$$\downarrow$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x\, \partial y}$$

- $f_{X,Y}(x,y) \geq 0$

- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx = 1$

ex) Find CDF $F_{X,Y}(x,y)$



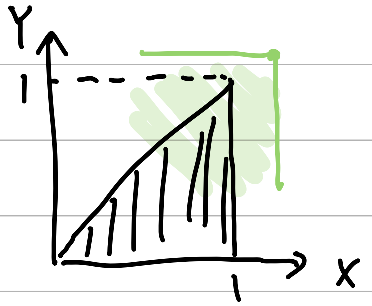
$f_{X,Y}(x,y) = 2$

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \leq y \leq x \leq 1 \\ 0 & \text{o.w.} \end{cases}$$

① $x<0$ 또는 $y<0$ 일 때

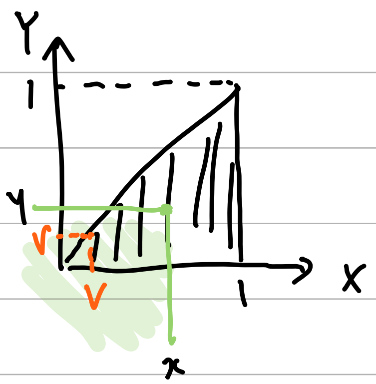
$F_{X,Y}(x,y) = 0$

② $x>1,\ y>1$ 일 때



$F_{X,Y}(x,y) = 1$

③ $0 \le y \le x \le 1$ 일 때



$$F_{X,Y}(x,y) = \int_0^y \int_v^x 2\,du\,dv$$

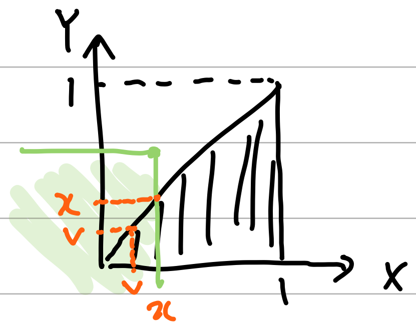
$$= \int_0^y \left([2u]_v^x \right) dv$$

$$= \int_0^y (2(x-v))\,dv$$

$$= \left[2xv - 2\tfrac{1}{2}v^2 \right]_0^y$$

$$= 2xy - y^2$$

④ $0 \le x < y,\ 0 \le x \le 1$ 일 때



$$F_{X,Y}(x,y) = \int_0^x \int_v^x 2\,du\,dv$$

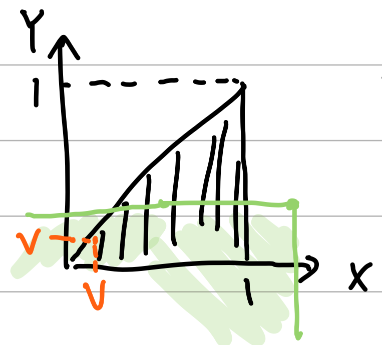
$$= \int_0^x [2u]_v^x \, dv$$

$$= \int_0^x 2(x-v)\, dv$$

$$= [2xv - v^2]_0^x$$

$$= 2x^2 - x^2 = x^2$$

⑤ $0 \le y \le 1 \quad x > 1$ 일 때



$$F_{X,Y}(x,y) = \int_0^y \int_v^1 2\,du\,dv$$

$$= \int_0^y [2u]_v^1 \, dv$$

$$= \int_0^y 2-2v \, dv$$

$$= [2v - v^2]_0^y$$

$$= 2y - y^2$$

$$F_{X,Y}(x,y) = \begin{cases} 0 & x<0 \text{ or } y<0 \\ 2xy - y^2 & 0 \le y \le x \le 1 \\ x^2 & 0 \le x < y,\ 0 \le x \le 1 \\ 2y - y^2 & 0 \le y \le 1,\ x > 1 \\ 1 & x > 1,\ y > 1 \end{cases}$$

Marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$

Independent Random Variable

RV X and Y are independent iff

$$P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)$$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

for all x, y

Expected Value

$W = g(x,y)$

$$E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x,y)\, P_{X,Y}(x,y)$$

$$E[W] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)$$

$E[X+Y] = E[X] + E[Y]$

$Var[X+Y] = Var[X] + Var[Y] + 2E[(X-\mu_X)(Y-\mu_Y)]$

Covariance

$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$

Correlation Coefficient

$\rho_{X,Y} = \dfrac{Cov[X,Y]}{\sqrt{Var[X]\,Var[Y]}} = \dfrac{Cov[X,Y]}{\sigma_X \sigma_Y}$, $-1 \le \rho_{X,Y} \le 1$

- if $\hat{X} = aX + b$, $\hat{Y} = cY + d$

$\rho_{\hat{X},\hat{Y}} = \rho_{X,Y}$, $Cov[\hat{X},\hat{Y}] = a \cdot c \cdot Cov[X,Y]$

Correlation

$r_{X,Y} = E[X \cdot Y]$

$Cov[X,Y] = E[XY] - E[X]E[Y] = r_{X,Y} - \mu_X \mu_Y$

$Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$

- $X = Y$ 라면?

$Cov[X,Y] = E[X^2] - E[X]^2 = Var[X]$

Orthogonal RV

$r_{X,Y} = E[XY] = 0$ 이면 Orthogonal 하다고 한다.

Uncorrelated RV

$Cov[X,Y] = \rho_{X,Y} = 0$ 이면 Uncorrelated 하다.

- X, Y가 independent 하다면,

$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)]$

$r_{X,Y} = E[XY] = E[X]E[Y]$

$Cov[X,Y] = \rho_{X,Y} = 0$

$Var[X+Y] = Var[X] + Var[Y] + \underbrace{2Cov[X,Y]}_{=0}$

independent 이면 → uncorrelated 하다. (역은 성립 X)

- Multivariate Joint CDF

$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P[X_1 \le x_1, \cdots, X_n \le x_n]$$

- Multivariate Joint PDF

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \frac{\partial^n F_{X_1,\cdots,X_n}(x_1,\cdots,x_n)}{\partial x_1 \cdots \partial x_n}$$

- $f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) \geq 0$
- $F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)\, du_n \cdots du_1$
- $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\cdots,X_n}(x_1,\cdots,x_n)\, dx_n \cdots dx_1 = 1$

- Multivariate Joint PMF

$$P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P[X_1 = x_1, \cdots, X_n = x_n]$$

- $P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) \geq 0$
- $\sum_{x_1 \in S_{X_1}} \cdots \sum_{x_n \in S_{X_n}} P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = 1$

N independent

$X_1 \cdots X_n$ 이 모두 independent 하면,

$$P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P_{X_1}(x_1)\, P_{X_2}(x_2) \cdots P_{X_n}(x_n)$$

N independent and identically distributed (iid)

$$P_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = P_X(x_1)\, P_X(x_2) \cdots P_X(x_n)$$