4. 결합확률분포(Joint probability distribution)

① 하나의 확률실험에서 두 개 이상의 확률변수를 정의하는 경우

한 개의 확률변수를 정의하는 경우	두 개 이상의 확률변수를 정의하는 경우
임의의 30세 성인, 키가 160cm에서 165cm 사이일 확률은?	임의의 30세 성인, 키가 160cm~165cm 사이이면서 동시에 몸무게가 60~70kg이 될 확률은?
$P(160 \le H \le 165)$	$P(160 \le H \le 165,\ 60 \le W \le 70)$
$P(160 \le H \le 165)$ $= \int_{160}^{165} p(h)dh$	$P(160 \le H \le 165,\ 60 \le W \le 70)$ $= \int_{60}^{70}\int_{160}^{165} p(h,\ w)\,dh\,dw$

② 이산확률변수와 연속확률변수의 결합확률분포함수

이산확률변수 X와 Y의 결합확률질량함수 $p_{XY}(x,\ y)$	연속확률변수 X와 Y의 결합확률밀도함수 $f_{XY}(x,\ y)$
㉠ $p_{XY}(x,\ y) \ge 0$ ㉡ $\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty} p_{XY}(x,\ y) = 1$ ㉢ $P\{(x,\ y) \in A\} = \sum\sum_{A} p_{XY}(x,\ y)$	㉠ $f_{XY}(x,\ y) \ge 0$ ㉡ $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,\ y)dx\,dy = 1$ ㉢ $P\{[X,\ Y] \in R\} = \iint_R f_{XY}(x,\ y)\,dxdy$
$X = \{x \mid x = 0,\ 1,\ 2\}$, $Y = \{y \mid y = 0,\ 1,\ 2,\ 3\}$ $A = \{(x,\ y) \mid x + y \le 2\}$에 대한 $P[(x,\ y) \in A]$	$f_{XY}(x,\ y) = \frac{1}{3}(x+y)$ $(0 < x < 1,\quad 0 < y < 2)$

③ 두 확률변수 $X,\ Y$에 대한 결합확률분포함수가 주어진 상태에서 각각의 확률변수의 확률분포를 주변분포(marginal distribution)라 한다. 주변분포는 결합확률분포함수로부터 한 가지 확률변수에 대한 정보만 알 수 있다.

이산확률변수 $X,\ Y$의 $p_{XY}(x,\ y)$	연속확률변수 $X,\ Y$의 $f_{XY}(x,\ y)$
각 확률변수의 주변확률질량함수 ㉠ $p_X(x) = P(X = x) = \sum_{y=-\infty}^{\infty} p_{XY}(x,\ y)$ ㉡ $p_Y(y) = P(Y = y) = \sum_{x=-\infty}^{\infty} p_{XY}(x,\ y)$	각 확률변수의 주변확률밀도함수 ㉠ $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,\ y)dy$ ㉡ $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,\ y)dx$

4-1. 이산확률변수의 결합분포

4-1-1. 결합확률질량함수

① 이산확률변수 X와 Y의 결합확률질량함수

$$p_{XY}(x,\ y) = P(X = x,\ Y = y)$$

② 성질

㉠ $p_{XY}(x,\ y) \geq 0$

㉡ $\displaystyle\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty} p_{XY}(x,\ y) = 1$

㉢ $P\{(x,\ y) \in A\} = \displaystyle\sum\sum_{A} p_{XY}(x,\ y)$

〚예〛 검은 구슬 3개, 붉은 구슬 2개, 흰 구슬 3개에서 임의로 2개를 꺼낼 때, 검은 구슬 개수를 X, 붉은 구슬 개수를 Y라고 하면 X와 Y의 결합확률분포는?
$A = \{(x,\ y) | X + Y \leq 1\}$일 때 $P[(x,\ y) \in A]$?

[풀이]

- 검은 구슬 X : ${}_3C_x\ (0 \leq x \leq 3)$
- 붉은 구슬 Y : ${}_2C_y\ (0 \leq y \leq 2)$
- 흰 구슬 : ${}_3C_{2-x-y}$

따라서 X와 Y의 결합확률분포함수는

$$\mathrm{P}(X = x,\ Y = y) = \frac{{}_3C_x \times {}_2C_y \times {}_3C_{2-x-y}}{{}_8C_2}$$

$Y \backslash X$	0	1	2	p_Y
0	3/28	9/28	3/28	15/28
1	6/28	6/28	0	12/28
2	2/28	0	0	1/28
p_X	10/28	15/28	3/28	1

이다. 오른쪽 표는 X와 Y의 결합확률분포표이다.
또한, $A = \{(x,\ y) | X + Y \leq 1\}$이므로 $A = \{(0,\ 0),\ (0,\ 1),\ (1,\ 0)\}$이므로
$P[(x,\ y) \in A] = (3+9+6)/28 = 9/14$이다.

〚예〛 위 예에서 X와 Y의 주변질량함수와 X의 주변평균 및 X의 주변분산은?

[풀이]

$$p_X(x) = P(X = x) = \frac{{}_3C_x \times {}_5C_{2-x}}{{}_8C_2}\ (x = 0,\ 1,\ 2)$$

$$p_Y(y) = P(Y = y) = \frac{{}_2C_y \times {}_5C_{2-y}}{{}_8C_2}\ (y = 0,\ 1,\ 2)$$

X의 주변평균 $\mu_X = E(X) = \displaystyle\sum_{x=-\infty}^{\infty} x p_x(x) = 0 \times \frac{10}{28} + 1 \times \frac{15}{28} + 2 \times \frac{3}{28}$

X의 주변분산 $\sigma_X^2 = V(X) = \displaystyle\sum_{x=-\infty}^{\infty} (x - \mu_x)^2 p_x(x) = \frac{45}{112}$

4-1-2. 조건부확률질량함수

① 두 이산확률변수 $X,\ Y$의 결합확률질량함수 $p_{XY}(x,\ y)$에 대하여

㉠ $X=x$로 주어진 확률 변수 Y의 조건부 확률질량함수 $p_{Y|x}(y)$

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

㉡ $Y=y$로 주어진 확률 변수 X의 조건부 확률질량함수 $p_{X|y}(x)$

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

② $p_{Y|x}(y)\geq 0,\ p_{X|y}(x)\geq 0$

③ $\sum_{R_x} p_{Y|x}(y)=1,\ R_x=\{(X,\ Y)$의 치역에서 $X=x$를 만족하는 부분집합$\}$

〖예〗 다음 결합확률분포표를 참고하여 $X=3$일 때 Y의 조건부 분포를 구하시오.

$p_{XY}(x,\ y)$		Y					$p_X(x)$
		0	1	2	3	4	
X	0	0.00000	0.00000	0.00002	0.00004	0.00004	0.0001
	1	0.00003	0.00035	0.00138	0.00184	*	0.0036
	2	0.00194	0.0156	0.0311	*	*	0.0486
	3	0.0583	0.2333	*	*	*	0.2916
	4	0.6561	*	*	*	*	0.6561
$p_Y(y)$		0.7164	0.24925	0.0325	0.00184	0.00004	1.0000

[풀이]

$x=3$일 때, Y의 가능한 값은 0, 1 뿐이다.

즉, $R_x=\{(x,\ y):x=3\}=\{(3,\ 0),\ (3,\ 1)\}$이다.

$$p_{Y|3}(0)=P(Y=0|X=3)=\frac{P(X=3,\ Y=0)}{P(X=3)}=\frac{0.0583}{0.2916}=0.1999$$

$$p_{Y|3}(1)=P(Y=1|X=3)=\frac{P(X=3,\ Y=1)}{P(X=3)}=\frac{0.2333}{0.2916}=0.8001$$

이다. 따라서 $\sum_{y=0}^{1} p_{Y|3}(y)=0.1999+0.8001=1.0$이다.

4-2. 연속확률변수의 결합분포

4-2-1. 결합확률밀도함수

① 두 연속확률변수 X와 Y의 결합확률밀도함수 $f_{XY}(x,\ y)$; 2차원 평면상에서 정의

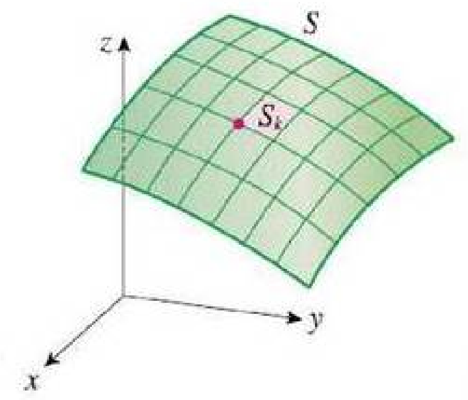
② 성질

㉠ $f_{XY}(x,\ y) \geq 0$

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

㉢ 2차원 평면상의 임의의 영역 R에 대해

$P\{[X,\ Y] \in R\} = \iint_R f_{XY}(x,\ y)\,dxdy$: 체적



〖예〗 두 연속확률변수 $X,\ Y$의 결합확률밀도함수

$$f_{XY}(x,\ y) = axy\ (1 < x < 3,\ 2 < y < 4,\ a\text{는 상수})$$

에서 상수 a, $P(1 \leq X \leq 2,\ 2 \leq Y \leq 3)$은?

[풀이]

$\int_2^4\int_1^3 axy\,dxdy = a\int_2^4 4y\,dy = 24a = 1$이므로 상수 a의 값은 $\frac{1}{24}$이다.

$P(1 \leq X \leq 2,\ 2 \leq Y \leq 3) = \int_2^3\int_1^2 \frac{1}{24}xy\,dxdy = \frac{15}{96}$

Q? X의 주변밀도함수 $f_X(x)$는?

[풀이]

$f_X(x) = \int_2^4 \frac{1}{24}xy\,dy = \frac{1}{4}x\ (1 < x < 3)$

X의 주변평균 $\mu_X = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_1^3 x\frac{1}{4}x\,dx = \frac{13}{6}$

X의 주변분산 $\sigma_X^2 = V(x) = \int_{-\infty}^{\infty}(x-\mu_X)^2 f_X(x)\,dx = \int_1^3\left(x-\frac{13}{6}\right)^2\frac{1}{4}x\,dx = \frac{11}{36}$

4-2-2. 조건부확률밀도함수

① 두 연속확률변수 X, Y의 결합확률밀도함수 $f_{XY}(x,\ y)$에서 $X=x$를 가정한 Y의 조건부 확률밀도함수 $f_{Y|x}(y)$는

$$f_{Y|x}(y)=\frac{f_{XY}(x,\ y)}{f_X(x)}\ (\text{단},\ f_X(x)>0)$$

② $f_{Y|x}(y) \geq 0$

③ $\int_{R_x} f_{Y|x}(y)\,dy=1$

$(R_x=\{(X,\ Y)$의 치역에서 $X=x$를 만족하는 부분집합$\}$

④ $Y=y$를 가정한 X의 조건부 확률밀도함수 $f_{X|y}(x)$

〚예〛 두 연속확률변수 X, Y의 결합확률질량함수

$$f_{XY}(x,\ y)=\frac{1}{3}(x+y)\ (0<x<1,\quad 0<y<2)$$

일 때, $Y=y$를 가정한 X의 조건부 확률밀도함수?

[풀이]

$$f_{X|y}(x)=\frac{f_{XY}(x,\ y)}{f_Y(y)}=\frac{\frac{1}{3}(x+y)}{\frac{1}{3}\int_0^1 (x+y)\,dx}=\frac{2(x+y)}{2y+1}\quad (0<x<1,\ 0<y<2)$$

【참고】 두 변수 X와 Y의 독립 확인

결합확률질량함수	결합확률밀도함수
① 실숫값 x, y에 대해 $p_{XY}(x,\ y)=p_X(x)p_Y(y)$ ② $p_X(x)>0$인 모든 x, y에 대해 $p_{Y\|x}(y)=p_Y(y)$ ③ $p_Y(y)>0$인 모든 x, y에 대해 $p_{X\|y}(x)=p_X(x)$	① 실숫값 x, y에 대해 $f_{XY}(x,\ y)=f_X(x)f_Y(y)$ ② $f_X(x)>0$인 모든 x, y에 대해 $f_{Y\|x}(y)=f_Y(y)$ ③ $f_Y(y)>0$인 모든 x, y에 대해 $f_{X\|y}(x)=f_X(x)$