3장. 결합확률분포

결합분포함수(joint distribution function)

(1) 두 확률변수 $X, Y : \Omega \to \mathbb{R}$에 대하여, 함수

$$(X, Y) : \Omega \to \mathbb{R} \times \mathbb{R}, \quad (X, Y)(\omega) = (X(\omega), Y(\omega)),$$

를 이차원 확률벡터 또는 이차원 확률변수라고 한다.

(2) (X, Y)를 이차원 확률벡터라 할 때,

- 2변수 함수 $F_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$F_{X,Y}(x, y) = P[X \leqslant x, Y \leqslant y]$$
$$= P\Big(\{\omega \in \Omega \mid X(\omega) \leqslant x, Y(\omega) \leqslant y\}\Big),$$

  를 (X, Y)의 결합분포함수라 한다.

- $R_{X \times Y} = \{(x, y) \in \mathbb{R}^2 \mid X(\omega) = x, Y(\omega) = y, \omega \in \Omega\}$를 (X, Y)의 치공간이라 한다.

결합확률질량함수(joint probability mass function)

$(X, Y) : \Omega \to \mathbb{R} \times \mathbb{R}$를 이차원 확률벡터라 하자.
(X, Y)의 치공간 $R_{X \times Y}$가 가산집합인 경우 확률벡터 (X, Y)를 이산확률벡터라 하고, 함수 $f_{X,Y} : \mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$f_{X,Y}(x, y) = P[X = x, Y = y]$$
$$= P[\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}],$$

를 (X, Y)의 결합확률질량함수라 한다.

결합확률질량함수의 성질

(1) $\forall (x, y) \in \mathbb{R}^2, \ f_{X,Y}(x, y) \geqslant 0$

(2) $\displaystyle\sum_{(x,y) \in \mathbb{R}^2} f_{X,Y}(x, y) = \sum_{x \in \mathbb{R}} \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) = 1$

(3) $\displaystyle F_{X,Y}(x, y) = P[X \leqslant x, Y \leqslant y] = \sum_{x' \leqslant x} \sum_{y' \leqslant y} f_{X,Y}(x', y')$

Example

한 주머니에 흰 공이 3개, 검은 공이 2개, 붉은 공이 3개가 들어 있다. 이 주머니에서 2개의 공을 꺼낼 때 X를 흰 공의 개수, Y를 검은 공의 개수라 한다.

(1) 확률변수 X와 Y의 치공간
(2) 확률변수 X와 Y의 확률질량함수
(3) 확률벡터 (X, Y)의 치공간
(4) 확률벡터 (X, Y)의 결합확률질량함수
(5) X, Y의 주변확률질량함수
(6) (2)와 (5)의 결과 비교

결합확률밀도함수(joint probability density function)

이차원 확률벡터 $(X,Y):\Omega\to\mathbb{R}\times\mathbb{R}$의 결합분포함수를 $F_{X,Y}$라 하자. 적당한 음이아닌 함수 $f_{X,Y}:\mathbb{R}^2\longrightarrow\mathbb{R}$가 존재하여 모든 $x,y\in\mathbb{R}$에 대하여

$$F_{X,Y}(x,y)=P[X\leqslant x,Y\leqslant y]=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{X,Y}(s,t)dt\,ds$$

를 만족할 때, 확률벡터 (X,Y)를 연속확률벡터라 하고 함수 $f_{X,Y}$를 (X,Y) 의 결합확률밀도함수라 한다.

결합확률밀도함수의 성질

(1) $\forall(x,y)\in\mathbb{R}^2,\ f_{X,Y}(x,y)\geqslant 0$

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

(3) $\displaystyle P[a_1\leqslant X\leqslant a_2,b_1\leqslant Y\leqslant b_2]=\int_{a_1}^{a_2}\int_{b_1}^{b_2}f_{X,Y}(x,y)dy\,dx$

Example

연속확률벡터 (X,Y)의 결합확률밀도함수가 다음과 같다.

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{3}{11}y(x+y), & 0<x<1,\ 0<y<2\\ 0, & \text{otherwise}\end{cases}$$

(1) $P[0<X<1,1<Y<2]$
(2) 확률벡터 (X,Y)의 결합분포함수
(3) X,Y의 주변확률질량함수

주변확률질량함수(marginal pmf)

(X,Y)가 이산확률벡터일 때, X,Y의 확률질량함수 f_X,f_Y를 각각 X,Y의 주변확률질량함수라 한다.

Theorem

$f_{X,Y}$를 이산확률벡터 (X,Y)의 결합확률질량함수라 하면,

(1) $\displaystyle f_X(x)=\sum_{y\in\mathbb{R}}f_{X,Y}(x,y)$

(2) $\displaystyle f_Y(y)=\sum_{x\in\mathbb{R}}f_{X,Y}(x,y)$

(3) $\displaystyle F_X(x)=\lim_{y\to\infty}F_{X,Y}(x,y)$

(4) $\displaystyle F_Y(y)=\lim_{x\to\infty}F_{X,Y}(x,y)$

주변확률밀도함수(marginal pdf)

(X,Y)가 연속확률벡터일 때, X,Y의 확률밀도함수 f_X,f_Y를 각각 X,Y의 주변확률밀도함수라 한다.

Theorem

$f_{X,Y}$를 연속확률벡터 (X,Y)의 결합확률밀도함수라 하면,

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

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

(3) $\displaystyle F_X(x)=\lim_{y\to\infty}F_{X,Y}(x,y)$

(4) $\displaystyle F_Y(y)=\lim_{x\to\infty}F_{X,Y}(x,y)$

확률변수의 독립성

(X, Y)는 $f_{X,Y}(x, y)$를 결합확률밀도(질량)함수로 갖는 2차원 연속 (이산)확률벡터이고, $f_X(x)$와 $f_Y(y)$는 각각 X와 Y의 주변확률밀도(질량)함수일 때,

$$\forall (x, y) \in \mathbb{R}^2, \quad f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

이면, 확률변수 X와 Y는 확률적으로 서로 독립(stochastically independent), 간단히 독립이라고 한다.

Remark

- 결합확률밀도(질량)함수 $f_{X,Y}(x, y)$가 주어지면 주변확률밀도(질량)함수 $f_X(x)$와 $f_Y(y)$를 구할 수 있다.
- X와 Y가 독립이고 주변확률밀도(질량)함수 $f_X(x)$와 $f_Y(y)$가 주어지면 결합확률밀도(질량)함수 $f_{X,Y}(x, y)$를 구할 수 있다.

Example

2차원 연속 확률벡터 (X, Y)의 결합확률밀도함수가 다음과 같다.

$$f_{X,Y}(x, y) = \begin{cases} 18x^2y^5, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

X와 Y가 (확률적으로) 독립인지 확인하시오.

Example

2차원 연속 확률벡터 (X, Y)의 결합확률밀도함수가 다음과 같다.

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{5}(x + 4y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

X와 Y가 (확률적으로) 독립인지 확인하시오.

Definition

(X, Y)는 $f_{X,Y}(x, y)$를 결합확률밀도(질량)함수로 갖는 2차원 연속 (이산)확률벡터이고, $g(X, Y)$를 X와 Y의 함수라 할 때, $g(X, Y)$의 평균을 다음과 같이 정의한다.

$$E\Big[g(X, Y)\Big] = \begin{cases} \displaystyle\sum_{x\in\mathbb{R}} \sum_{y\in\mathbb{R}} g(x, y) f_{X,Y}(x, y), & \text{이산형} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy, & \text{연속형} \end{cases}$$

Theorem

- $g(X, Y) = X$이면, $E\Big[g(X, Y)\Big] = E[X]$
- $E[X + Y] = E[X] + E[Y]$

Example

확률벡터 (X, Y)의 결합확률질량함수와 X와 Y의 주변확률질량함수가 다음과 같다.

$x \diagdown y$	2	4	6	$f_X(x)$
1	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
$f_Y(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

확률변수 $g(X, Y) = X + Y$의 기댓값을 구하시오.

Example

독립인 두 확률변수 X와 Y의 확률밀도함수가 각각 다음과 같다.

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}, \quad f_Y(y) = \begin{cases} 1, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

확률변수 $Z = X + Y$의 기댓값을 구하시오.

Theorem

2차원 확률벡터 (X, Y)에 대하여 X, Y가 독립이면,

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

특히, $E[XY] = E[X]\ E[Y]$이고 $m_{X+Y}(t) = m_X(t)m_Y(t)$이다.

Remark
$E[XY] = E[X]\ E[Y]$이지만 (X, Y)가 독립이 아닌 경우가 있다.

공분산과 상관계수

(X, Y)를 이차원 확률벡터라 하고

$$E[X] = \mu_X, \quad E[Y] = \mu_Y, \quad \mathrm{Var}[X] = \sigma_X^2, \quad \mathrm{Var}[Y] = \sigma_Y^2$$

이 각각 존재할 때, 다음을 정의한다.

(1) X와 Y의 공분산(covariance):

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

(2) X와 Y의 상관계수(correlation coefficient):

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}}$$

Remark
$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X), \quad \mathrm{Cov}(X, X) = \mathrm{Var}[X]$

Theorem

이차원 확률벡터 (X, Y)에 대하여 다음이 존재한다고 하자.

$$E[X] = \mu_X, \qquad E[Y] = \mu_Y,$$
$$\mathrm{Var}[X] = \sigma_X^2, \qquad \mathrm{Var}[Y] = \sigma_Y^2, \qquad E[XY] = \mu_{XY}$$

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

(2) X와 Y가 독립이면, $\mathrm{Cov}(X, Y) = 0$

(3) $\forall a, b, c, d \in \mathbb{R}, \quad \mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$

(4) $\forall a, b, c, d \in \mathbb{R}, \quad \rho(aX + b, cY + d) = \mathrm{sgn}(ac)\,\rho(X, Y) = \frac{ac}{|ac|}\rho(X, Y)$

(5) $\Big|\mathrm{Cov}(X, Y)\Big| \leqslant \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}$ (Cauchy-Schwarz 부등식)

(6) $-1 \leqslant \rho(X, Y) \leqslant 1$

(7) $\rho(X, Y) = \pm 1$이면, $X = \mu_X \pm \dfrac{\sigma_Y}{\sigma_X}(Y - \mu_Y)$

상관계수 $\rho(X, Y)$의 의미

$\rho = \rho(X, Y)$를 확률변수 X와 Y의 상관계수라 할 때,

$\lvert\rho\rvert \geqslant 0.8$:	강한 선형관계
$0.5 \leqslant \lvert\rho\rvert < 0.8$:	보통 선형관계
$\lvert\rho\rvert < 0.5$:	약한 선형관계
$\lvert\rho\rvert = 1$:	완전환 선형관계
$\rho = 0$:	상관이 없음=비상관=무상관

Remark
X와 Y가 독립이면 $\mathrm{Cov}(X, Y) = 0$이다. 하지만 일반적으로 그 역은 성립하지 않는다.

Example

확률벡터 (X, Y)의 결합확률질량함수가 다음과 같을 때, X와 Y의 공분산 $\mathrm{Cov}(X, Y)$와 상관계수 $\rho(X, Y)$를 구하시오.

$x \setminus y$	0	1	2	3	$f_X(x)$
0	0	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
1	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
2	$\frac{1}{8}$	$\frac{1}{8}$	0	0	$\frac{1}{4}$
$f_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Example

연속확률벡터 (X, Y)의 결합확률밀도함수가 다음과 같을 때, X와 Y의 공분산 $\mathrm{Cov}(X, Y)$와 상관계수 $\rho(X, Y)$를 구하시오.

$$f_{X,Y}(x, y) = \begin{cases} 8xy, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Theorem

$X_1, X_2, \cdots, X_n$과 $Y_1, Y_2, \cdots, Y_m$을 적분가능한 확률변수라 하고 모든 i, j에 대하여 $E[X_i]$, $E[Y_j]$가 존재하면,

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

$$(2) \quad \mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathrm{Var}[X_i] + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j)$$

(3) $X_1, X_2, \cdots, X_n$의 모든 짝이 무상관이면,

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathrm{Var}[X_i]$$

조건부 확률분포

(X, Y)는 $f_{X,Y}(x, y)$를 결합확률밀도(질량)함수로 갖는 2차원 연속(이산)확률벡터이고, $f_X(x)$와 $f_Y(y)$는 각각 X와 Y의 주변확률밀도(질량)함수라 하자.

(1) $Y = y$ 에 대한 X의 조건부 확률밀도(질량)함수

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad f_Y(y) \neq 0$$

(2) $X = x$ 에 대한 Y의 조건부 확률밀도(질량)함수

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

Theorem

$f_{X,Y}(x, y)$를 2차원 연속(이산)확률벡터 (X, Y)의 결합확률밀도(질량)함수라 하자.

(1) $f_X(x) \neq 0$인 $X = x$에서, $\forall y \in \mathbb{R}$, $f_{Y|X}(y|x) \geqslant 0$

(2) $f_X(x) \neq 0$인 $X = x$에서,

(i) (X, Y)가 이산확률벡터이면, $\displaystyle\sum_{y \in \mathbb{R}} f_{Y|X}(y|x) = 1.$

(ii) (X, Y)가 연속확률벡터이면, $\displaystyle\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dy = 1.$

Example

연속확률벡터 (X, Y)의 결합확률밀도함수가

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(x+4y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

일 때, 다음 물음에 답하시오.

(1) $Y = y$에 대한 X의 조건부 확률밀도함수를 구하시오.

(2) (1)을 이용하여 확률 $P\left[X \leqslant \frac{1}{2} \;\middle|\; Y = \frac{1}{2}\right]$를 구하시오.

조건부 기댓값(conditional expectation)

(X, Y)는 $f_{X,Y}(x,y)$를 결합확률밀도(질량)함수로 갖는 2차원 연속(이산)확률벡터라 하자.

(1) $Y = y$ 로 주어졌을 때 X의 조건부 기댓값(단, $f_Y(y) \neq 0$)

$$E[X \mid y] = E[X \mid Y = y] = \begin{cases} \displaystyle\sum_{x \in \mathbb{R}} x \cdot f_{X|Y}(x|y), & \text{이산형} \\ \displaystyle\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx, & \text{연속형} \end{cases}$$

(2) $X = x$ 로 주어졌을 때 Y의 조건부 기댓값(단, $f_X(x) \neq 0$)

$$E[Y \mid x] = E[Y \mid X = x] = \begin{cases} \displaystyle\sum_{y \in \mathbb{R}} y \cdot f_{Y|X}(y|x), & \text{이산형} \\ \displaystyle\int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\, dy, & \text{연속형} \end{cases}$$

Example

확률벡터 (X, Y)의 결합확률질량함수와 X와 Y의 주변확률질량함수가 다음과 같을 때, $Y = 1$로 주어졌을 때 X의 조건부 기댓값을 구하여라.

$x \setminus y$	0	1	2	3	$f_X(x)$
0	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{120}$	$\frac{7}{15}$
1	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{20}$	0	$\frac{7}{15}$
2	$\frac{1}{24}$	$\frac{1}{40}$	0	0	$\frac{1}{15}$
$f_Y(y)$	$\frac{7}{24}$	$\frac{21}{40}$	$\frac{7}{40}$	$\frac{1}{120}$	1

Example

2차원 연속 확률벡터 (X, Y)의 결합확률밀도함수가 다음과 같다.

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(x+4y), & 0 < x < 1,\ 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

(1) $Y = \frac{1}{2}$로 주어졌을 때 X의 조건부 기댓값을 구하시오.

(2) $X = \frac{1}{2}$로 주어졌을 때 Y의 조건부 기댓값을 구하시오.