결합분포와 조건부분포

4장

4.1 결합확률분포

동일한 표본공간에서 정의되는 두 개 이상의 확률변수의 분포를 결합확률분포(joint probability distribution) 또는 결합분포(joint distribution)라 한다. X와 Y의 결합확률밀도함수(joint probability density function 또는 $joint\ \ p.d.f.$)는 $f(x,\ y)$로 나타내고, 기호로는 $(X,\ Y)' \sim f(x,\ y)$로 표기한다.

결합확률밀도함수 $f(x,\ y)$는 다음의 조건을 만족한다. X와 Y가 모두 이산형인 경우에는

(a) $f(x,\ y) \geq 0 \quad for\ \ all\ \ x\ \ and\ \ y$

(b) $\sum_{all}\sum_{(x,\ y)} f(x,\ y) = 1$

(c) $P\{(X,\ Y) \in A\} = \sum\sum_{(x,\ y) \in A} f(x,\ y)$

을 만족하며, X와 Y가 모두 연속형인 경우에는

(a) $f(x,\ y) \geq 0 \quad for\ \ all\ \ x\ \ and\ \ y$

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

(c) $P\{(X,\ Y) \in A\} = \int_A\int f(x,\ y)dxdy$

을 만족한다. 위에서 A는 xy평면상에서의 임의의 집합이다.

정의 1 결합누적분포함수

X와 Y의 결합누적분포함수(joint cumulative distribution function 또는 *joint c.d.f.*)는

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

으로 정의된다. 특히 X와 Y가 연속형인 경우에는

$$f(x,\ y) = \frac{d^2}{dxdy}F(x,\ y)$$

이 성립된다.

정의 2 주변확률밀도함수

$(X,\ Y)' \sim f(x,\ y)$일 때, X의 주변확률밀도함수(marginal probability density function 또는 *marginal p.d.f.*)는

$$f_X(x) = \begin{cases} \sum_y f(x,\ y), & (X,\ Y):\ discrete \\ \int_{-\infty}^{\infty} f(x,\ y)\,dy, & (X,\ Y):\ continuous \end{cases}$$

으로 주어지며, Y의 *marginal p.d.f.*도 동일한 방법으로 정의된다.

정의 3 기댓값

$(X,\ Y)' \sim f(x,\ y)$일 때, $u(X, Y)$의 기댓값은

$$Eu(X,\ Y) = \begin{cases} \sum_x \sum_y u(x, y) f(x,\ y), & (X,\ Y):\ discrete \\ \int_{-\infty}^{-\infty} \int_{-\infty}^{\infty} u(x,\ y) f(x,\ y)\,dx\,dy, & (X,\ Y):\ continuous \end{cases}$$

으로 정의된다.

정의 4 결합적률생성함수

$(X,\ Y)'$의 결합적률생성함수(joint moment generating function 또는 *joint m.g.f.*)는

$$M(t_1,\ t_2) = E(e^{t_1X + t_2Y})$$

$$= \begin{cases} \sum_x \sum_y e^{t_1x + t_2y} f(x,\ y), & (X,\ Y): discrete \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x + t_2y} f(x,\ y)dxdy, & (X,\ Y): continuous \end{cases}$$

으로 정의된다.

4.2 두 확률변수의 독립

두 확률변수 간의 독립(independence)에 대한 정의는 다음과 같다.

정의 5 두 확률변수의 독립

$(X,\ Y)'$가 결합누적분포함수 $F(x,\ y)$를 가질 때

$$F(x,\ y) = F_X(x)F_Y(y) \quad for\ all\ \ x\ \ and\ \ y$$

$$\Leftrightarrow X\text{와 } Y\text{는 서로 독립}$$

참고 $(X,\ Y)' \sim F(x,\ y)$일 때, X의 주변누적분포함수(*marginal c.d.f.*)는

$$F_X(x) = P(X \leqq x) = P(X \leqq x,\ Y < \infty)$$

$$= \lim_{y \to \infty} F(x,\ y) = F(x,\ \infty)$$

을 통해 구해질 수 있다. 마찬가지로 $F_Y(y) = F(\infty,\ y)$이다.

정의 6 두 확률변수의 독립

$(X, Y)' \sim f(x, y)$일 때

$$f(x, y) = f_X(x)\, f_Y(y) \;\; for \;\; all \;\; x \;\; and \;\; y$$

$$\Leftrightarrow X\text{와 } Y\text{는 서로 독립}$$

위의 정의를 이용하기 위해서는 X와 Y의 *marginal p.d.f.*를 직접 구해야 하는 번거로움이 있으나, 아래의 정리를 이용하면 보다 쉬운 방법으로 독립성을 보일 수 있다.

정리 1 $(X, Y)' \sim f(x, y)$일 때

$$f(x, y) = g(x)h(y)$$

$$\Leftrightarrow X\text{와 } Y\text{는 서로 독립}$$

위 정리의 의미는 *joint p.d.f.*가

$$(x\text{만의 함수}) \times (y\text{만의 함수})$$

의 형태로 표현되기만 하면 X와 Y는 서로 독립이며, 그렇지 않으면 독립이 아님을 의미한다. 이때, (x, y)의 범위(영역)도 (x 범위)×(y 범위)의 형태 즉, 사각형의 형태($(-\infty, \infty) \times (-\infty, \infty)$꼴 포함)로 표현되지 않으면 절대 두 변수가 독립이 될 수가 없음에 유의할 필요가 있다.

정리 2 X와 Y가 독립이면

(a) $u(X)$와 $v(Y)$도 독립이며, 따라서

(b) $E[u(X)v(Y)] = [E\,u(X)][E\,v(Y)]$

가 성립한다. 여기서 $u(X)$와 $v(Y)$는 각각 X와 Y만의 함수이다.

증명 (b) 기댓값의 정의를 이용하여 쉽게 보일 수 있다.

예제 1 $(X,\ Y)' \sim f(x,\ y) = \begin{cases} x+y, & 0<x<1,\ 0<y<1 \\ 0, & otherwise \end{cases}$

일 때, X와 Y가 독립인가?

풀이 [방법 1] X와 Y의 $marginal\ p.d.f.$는

$$f_X(x) = \int_0^1 (x+y)dy = \begin{cases} x+\frac{1}{2}, & 0<x<1 \\ 0, & otherwise \end{cases}$$

$$f_Y(y) = \begin{cases} y+\frac{1}{2}, & 0<y<1 \\ 0, & otherwise \end{cases}$$

이다. 따라서

$$f(x,\ y) \neq f_X(x) \cdot f_Y(y)$$

이므로 X와 Y는 독립이 아니다.

[방법 2] X와 Y의 영역은 $(0,\ 1)\times(0,\ 1)$로 곱의 형태로 표현되나, $f(x,\ y)$가 $g(x)h(y)$ 꼴로 분해되지 않으므로 X와 Y는 독립이 아니다. ■

예제 2 $Z \sim N(0,\ 1)$, $V \sim \chi^2(r)$ 이고, Z와 V가 독립일 때

$$T = \frac{Z}{\sqrt{V/r}}$$

의 기댓값은

$$E(T) = E\left(\frac{Z}{\sqrt{V/r}}\right) = \sqrt{r}\,E(Z)E\left(\frac{1}{\sqrt{V}}\right)$$

을 통해 $E(T) = 0$이 됨을 알 수 있다.

참고 위의 예제에서 T는 자유도가 r인 t분포의 정의로 사용된다(6장 [정의 1] 참고).

예제 3 두 확률변수의 비 $\frac{X}{Y}$와 그의 분모 Y가 독립이면

$$E\left[\left(\frac{X}{Y}\right)^k\right] = \frac{E(X^k)}{E(Y^k)}$$

이 성립함을 보여라. 단 $E(Y^k) \neq 0$ $(k = 1, 2, \cdots)$이다.

풀이 $\frac{X}{Y}$와 Y가 독립이면, $\left(\frac{X}{Y}\right)^k$와 Y^k도 독립이므로

$$E\left[\left(\frac{X}{Y}\right)^k Y^k\right] = E\left[\left(\frac{X}{Y}\right)^k\right]E[Y^k]$$

이고, 위 식의 좌변이 $E(X^k)$이므로 위 결과가 성립된다. ■

다음의 정리는 $m.g.f.$를 이용하여 독립성을 보이는 데 유용하다.

정리 3 $(X, Y)' \sim M(t_1, t_2)$일 때,

$$M(t_1, t_2) = M(t_1, 0)\,M(0, t_2)$$

$$\Leftrightarrow X\text{와 } Y\text{는 서로 독립}$$

위의 정리에서

$$M(t_1,\, 0) = E\, e^{t_1 X} = M_X(t_1)$$

$$M(0,\, t_2) = E\, e^{t_2 Y} = M_Y(t_2)$$

이며, 이는 각각 X와 Y의 $marginal\ \ m.g.f.$가 되며, 이에 대응하는 분포가 X와 Y의 $marginal\ \ p.d.f.$가 된다.

예제 4 $(X,\, Y)$의 $p.d.f.$가

$$f(x,\, y) = \begin{cases} \dfrac{n!}{x!y!(n-x-y)!}\, p_1^x\, p_2^y\, p_3^{n-x-y}, & x, y = 0, 1, \cdots, n,\ \ x+y \le n \\ 0, & otherwise \end{cases}$$

일 때, $(X,\, Y)$는 삼항분포(trinomial distribution)를 따른다고 하고, 기호로는 $(X,\, Y)' \sim Trinomial(n,\, p_1,\, p_2)$로 나타낸다. 여기서 $0 < p_1,\, p_2 < 1$, $p_3 = 1 - p_1 - p_2$이다.

이때, $(X,\, Y)$의 $joint\ \ m.g.f.$는

$$\begin{aligned} M(t_1,\, t_2) &= \sum_{x=0}^{n} \sum_{y=0}^{n-x} e^{t_1 x + t_2 y} \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y} \\ &= \sum_{x=0}^{n} \sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!} (p_1 e^{t_1})^x (p_2 e^{t_2})^y p_3^{n-x-y} \\ &= (p_1 e^{t_1} + p_2 e^{t_2} + p_3)^n \end{aligned}$$

이 된다. X와 Y의 $marginal\ \ m.g.f.$는

$$\begin{aligned} M_X(t_1) &= M(t_1,\, 0) \\ &= (p_1 e^{t_1} + p_2 + p_3)^n = [p_1 e^{t_1} + (1 - p_1)]^n \end{aligned}$$

$$M_Y(t_2) = M(0,\ t_2)$$
$$= (p_1 + p_2 e^{t_2} + p_3)^n = [p_2 e^{t_2} + (1 - p_2)]^n$$

이므로

$$X \sim B(n,\ p_1), \quad Y \sim B(n,\ p_2)$$

이다. 또한

$$M(t_1,\ t_2) \neq M(t_1,\ 0)M(0,\ t_2)$$

이므로 X와 Y는 서로 독립이 아니라고 말할 수 있다.

4.3 조건부 분포와 조건부 기댓값

정의 7 조건부확률밀도함수

$(X,\ Y)$가 $f(x,\ y)$를 따른다고 하자. $Y = y$로 주어질 때, X의 조건부확률밀도함수(conditional probability density function 또는 *conditional p.d.f.*)는

$$f(x|y) = \frac{f(x,\ y)}{f(y)}, \quad f(y) > 0$$

으로 정의되며, 기호로는 $X|y$ 또는 $X|Y = y \sim f(x|y)$로 나타낸다.

$(X,\ Y)$가 이산형일 때, 조건부확률밀도함수 $f(x|y)$는 다음의 조건

(a) $f(x|y) \geq 0 \quad for\ all\ x$

(b) $\sum_x f(x|y) = 1$

(c) $P(a \leq x \leq b|\ Y = y) = \sum_{x \in [a,b]} f(x|y)$

을 만족한다.

(X, Y)가 연속형일 때, 조건부확률밀도함수 $f(x|y)$는 다음의 조건

(a) $f(x|y) \geq 0 \quad for\ \ all\ \ x$

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

(c) $P(a \leq X \leq b \mid Y = y) = \int_a^b f(x \mid y)dx$

을 만족한다.

Remarks 조건부확률밀도함수 $f(x \mid y)$의 의미는 다음과 같다.

1. (X, Y)가 이산형인 경우에는

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

으로 $Y = y$상에 있는 확률이 0 이상인 점들만으로, 전체 확률의 합이 1이 되도록 만든 분포라 말할 수 있다.

2. (X, Y)가 연속형인 경우에는 $Y = y$상의 확률은 0이므로

$$f(x \mid y) \approx \frac{f(x,\ y)\triangle x \triangle y}{f(y)\triangle y} = \frac{f(x,\ y)}{f(y)}\triangle x$$

으로 정의하면 된다.

예제 1 $(X,\ Y)' \sim f(x,\ y) = \begin{cases} \dfrac{x+y}{21}, & x = 1, 2, 3,\ \ y = 1, 2 \\ 0, & otherwise \end{cases}$ 일 때, X와 Y의

$marginal\ \ p.d.f.$는

$$f_1(x) = \sum_y f(x,\ y) = \begin{cases} \dfrac{2x+3}{21}, & x = 1, 2, 3 \\ 0, & otherwise \end{cases}$$

$$f_2(y) = \sum_x f(x,\, y) = \begin{cases} \dfrac{3y+6}{21}, & y = 1, 2 \\ 0, & otherwise \end{cases}$$

이고, $X|y$와 $Y|x$ 의 *conditional p.d.f.*는

$$f(x|y) = \frac{f(x,\, y)}{f_2(y)} = \begin{cases} \dfrac{x+y}{3y+6}, & x = 1, 2, 3\ (y = 1, 2) \\ 0, & otherwise \end{cases}$$

$$g(y|x) = \frac{f(x,\, y)}{f_1(x)} = \begin{cases} \dfrac{x+y}{2x+3}, & y = 1, 2\ (x = 1, 2, 3) \\ 0, & otherwise \end{cases}$$

이다. 직관적인 이해를 위해 위의 *joint p.d.f.*를 표로 나타내면

Y \ X	1	2	3	합
1	2/21	3/21	4/21	9/21
2	3/21	4/21	5/21	12/21
합	5/21	7/21	9/21	1

이다. 이로부터 $X\,|\,Y=1$과 $X\,|\,Y=2$의 *conditional p.d.f.*는

$$f(x|1) = \begin{cases} \dfrac{2}{9}, & x = 1 \\ \dfrac{3}{9}, & x = 2 \\ \dfrac{4}{9}, & x = 3 \\ 0, & otherwise \end{cases} \qquad \& \qquad f(x|2) = \begin{cases} \dfrac{3}{12}, & x = 1 \\ \dfrac{4}{12}, & x = 2 \\ \dfrac{5}{12}, & x = 3 \\ 0, & otherwise \end{cases}$$

이 됨을 쉽게 확인할 수 있다.

정의 8 조건부 기댓값

$(X, Y)' \sim f(x, y)$라 하자. $Y=y$일 때 $u(X)$의 조건부 기댓값(conditional expectation)은

$$E[u(X)|y)] = \begin{cases} \sum_x u(x) f(x|y), & (X, Y): discrete \\ \int_{-\infty}^{\infty} u(x) f(x|y)\, dx, & (X, Y): continuous \end{cases}$$

으로 정의된다.

예를 들어, $Y=y$일 때 X의 조건부 기댓값은

$$\mu_{X|y} \equiv E(X|y) = \begin{cases} \sum_x x f(x|y), & (X, Y): discrete \\ \int_{-\infty}^{\infty} x f(x|y)\, dx, & (X, Y): continuous \end{cases}$$

이고, $Y=y$일 때 X의 조건부 분산은

$$\begin{aligned} \sigma^2_{X|y} \equiv Var(X|y) &= E\{[X - E(X|y)]^2 | y\} \\ &= E(X^2|y) - E(X|y)^2 \end{aligned}$$

으로 주어진다.

Remarks 1. 조건부 기댓값 $E[u(X)|y]$는 y만의 함수임에 주의할 필요가 있다. 따라서 $E(X|y)$나 $Var(X|y)$도 모두 y만의 함수이다.

2. 조건부 기댓값 $E[u(X)|y]$에서 소문자 y대신 대문자 Y를 사용한 표현은

$$E[u(X)|Y] = E[u(X)|y]\,|_{y=Y}$$

을 의미한다.

예제 2 $(X,\ Y)' \sim f(x,\ y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & otherwise \end{cases}$

일 때, X의 *marginal p.d.f.*는

$$f_1(x) = \int_x^1 2\,dy = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

이고, $Y|x$의 *conditional p.d.f.*는

$$f(y|x) = \frac{f(x,\ y)}{f_1(x)} = \begin{cases} \dfrac{1}{1-x}, & x < y < 1\ \ (0 < x < 1) \\ 0, & otherwise \end{cases}$$

이다. 이로부터 $E(Y|x)$와 $Var(Y|x)$는

$$E(Y|x) = \int_x^1 y\, f(y|x)\, dy = \int_x^1 y \frac{1}{1-x} dy = \frac{1+x}{2}, \quad 0 < x < 1$$

와

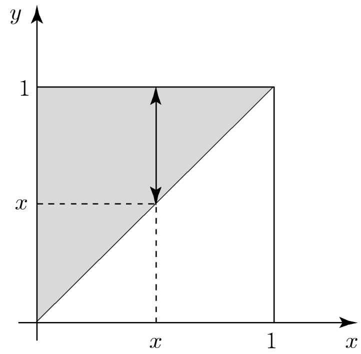
$$\begin{aligned} Var(Y|x) &= E(Y^2\,|\,x) - E(Y|\,x)^2 \\ &= \int_x^1 y^2 \frac{1}{1-x} dy - \left(\frac{1+x}{2}\right)^2 \\ &= \frac{1}{12}(x-1)^2, \quad 0 < x < 1 \end{aligned}$$

이다. 구체적인 확률계산의 예를 들면

$$\begin{aligned} P\left(\frac{5}{8} < Y < \frac{7}{8} \,\middle|\, X = \frac{1}{4}\right) &= \int_{\frac{5}{8}}^{\frac{7}{8}} f(y\,|\,\tfrac{1}{4})\ dy \\ &= \int_{\frac{5}{8}}^{\frac{7}{8}} \frac{4}{3}\ dy = \frac{1}{3} \end{aligned}$$

이다.

Remark 위의 예제에서 (X, Y)의 분포는 아래의 음영 부분의 영역에서, $p.d.f.$의 높이가 일정한, 균일분포를 따른다고 할 수 있다.



위 그림에서 $X=x$로 주어질 때 확률이 존재하는 Y의 공간은 $[x, 1]$구간이고, 이 영역에서 확률이 균일하게 분포하므로 $Y\,|\,x \sim Uniform(x, 1)$이 되고, 균일분포의 성질에 의해

$$E(Y|x)=\frac{x+1}{2}, \quad Var(Y|x)=\frac{(1-x)^2}{12}$$

이 되는 것은 직관적으로 당연한 결과이다.

또한, 위 그림으로부터 $X=1/4$로 주어질 때 $Y|X=1/4$은 $U(1/4, 1)$이므로

$$P\left(\frac{5}{8}< Y< \frac{7}{8}\,\middle|\,X=\frac{1}{4}\right)=\left(\frac{7}{8}-\frac{5}{8}\right)/\left(1-\frac{1}{4}\right)=\frac{1}{3}$$

이 됨은 자명하다.

예제 3 $(X, Y)' \sim Trinomial(n, p_1, p_2)$일 때, $Y|x$의 조건부 $p.d.f.$는

$$X \sim B(n, p_1)$$

으로부터(4.2절 예제4 참고)

$$f(y|x) = \frac{f(x,\, y)}{f(x)}$$

$$= \frac{\dfrac{n!}{x!n!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y}}{\dfrac{n!}{x!(n-x)!} p_1^x (1-p_1)^{n-x}}$$

$$= \begin{cases} \dfrac{(n-x)!}{y!(n-x-y)!} \left(\dfrac{p_2}{1-p_1}\right)^y \left(\dfrac{p_3}{1-p_1}\right)^{n-x-y}, & y = 0, 1, \cdots, n-x \\ 0, \quad otherwise \end{cases}$$

이므로

$$Y|X = x \sim B(n-x,\, \frac{p_2}{1-p_1})$$

이다. 이로부터 조건부 기댓값은

$$E(Y|x) = (n-x)\frac{p_2}{1-p_1}$$

이 된다.

마찬가지로 $X|Y = y \sim B(n-y,\, \frac{p_1}{1-p_2})$이고, $E(X|y) = (n-y)\frac{p_1}{1-p_2}$이 된다.

정리 4 이중기댓값정리 : Double Expectation Theorem

두 확률변수 X와 Y에 대하여

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

이 성립한다.

증명 기댓값의 정의로부터

$$
\begin{aligned}
E[E(X \mid Y)] &= E\varphi(Y), \quad \varphi(Y) = E(X \mid Y) \\
&= \int_{-\infty}^{\infty} \varphi(y) f(y)\, dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f(x \mid y) dx \right] f(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,\ y)\, dx\, dy \\
&= E(X)
\end{aligned}
$$

이 성립된다.

이중기댓값정리는 다음과 같이

(비조건부) 기댓값 = 조건부 기댓값의 평균

으로 해석될 수 있다.

위의 정리를 일반화하면

$$E[u(X)] = E\{E[u(X) \mid Y]\}$$

이다.

정리 5 두 확률변수 X와 Y에 대하여

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

이 성립한다.

증명

$$
\begin{aligned}
Var(X \mid Y) &= E(X^2 \mid Y) - E(X \mid Y)^2 \\
&= E(X^2 \mid Y) - \varphi(Y)^2, \quad \varphi(Y) = E(X \mid Y)
\end{aligned}
$$

이므로

$$E[Var(X \mid Y)] = E[E(X^2 \mid Y)] - E[\varphi(Y)^2]$$
$$= E(X^2) - E[\varphi(Y)^2]$$

이고,

$$Var[E(X \mid Y)] = Var[\varphi(Y)], \quad \varphi(Y) = E(X \mid Y)$$
$$= E[\varphi(Y)^2] - [E\varphi(Y)]^2$$
$$= E[\varphi(Y)^2] - E(X)^2$$

이므로, 위 사실이 증명된다.

위의 정리는 다음과 같이

(비조건부) 분산 = 조건부 분산의 기댓값 + 조건부 기댓값의 분산

으로 해석될 수 있다.

예제 4 앞의 (예제 1)에서

$$E(X) = E[E(X \mid Y)]$$

임을 보이시오.

풀이 먼저

$$E(X \mid y) = \sum_{x=1}^{3} x f(x \mid y)$$
$$= \sum_{x=1}^{3} x \frac{x+y}{3y+6} = \frac{6y+14}{3y+6}, \quad y = 1, 2$$

으로부터

$$E[E(X\,|\,Y)] = \sum_{y=1}^{2} E(X|y)f_2(y) = E(X|1)f_2(1) + E(X|2)f_2(2)$$

$$= \frac{20}{9}\cdot\frac{9}{21} + \frac{13}{6}\cdot\frac{12}{21} = \frac{46}{21}$$

이다. 또한

$$E(X) = \sum_{x=1}^{3} x f_1(x) = \sum_{x=1}^{3} x\frac{2x+3}{21} = \frac{46}{21}$$

이다. ■

예제 5 $X_1, X_2, \cdots, X_n, \cdots$을 평균 μ_X, 분산 σ_X^2인 분포로부터의 확률표본의 수열이라 하자. N이 정수값을 가지는 확률변수일 때, 확률변수들의 랜덤합(random sum) $S_N = \sum_{i=1}^{N} X_i$의 기댓값과 분산을 구하여라.

풀이 다음의 관계

$$E(S_N) = E\,[E\,(S_N\,|\,N)]$$

에서

$$E(\,S_N\,|\,N = n) = E(\,\sum_{i=1}^{n} X_i) = n\,\mu_X$$

이므로

$$E(S_N) = E(N\,\mu_X) = \mu_X\,E(N) = \mu_N\,\mu_X$$

이다. 또한

$$Var(S_N\,|\,N = n) = Var(\sum_{i=1}^{n} X_i) = n\sigma_X^2 \quad (X_i \text{가 } i.i.d. \text{ 이므로})$$

이므로

$$\begin{aligned} Var(S_N) &= E[Var(S_N \mid N)] + Var[E(S_N \mid N)] \\ &= E(N\sigma_X^2) + Var(N\mu_X) \\ &= \sigma_X^2\, E(N) + \mu_X^2\, Var(N) \\ &= \mu_N\, \sigma_X^2 + \sigma_N^2\, \mu_X^2 \end{aligned}$$

이다. ■

4.4 상관계수

$(X,\ Y)' \sim f(x,\ y)$일 때, 두 변수 X와 Y간의 선형관계의 정도를 나타내는 양으로 공분산과 상관계수가 있다.

정의 9 공분산

확률변수 X와 Y의 공분산(covariance)은

$$\begin{aligned} \sigma_{XY} &= Cov(X, Y) = E(X-\mu_X)(Y-\mu_Y) \\ &= E(XY) - \mu_X \mu_Y \end{aligned}$$

으로 정의한다.

공분산은 X와 Y가 그들의 평균으로부터 동시에 커지거나 작아지는 경향이 강할수록 양의 큰 값을 가지며, X가 커질 때 Y가 작아지거나 그 반대의 현상이 강할수록 음의 큰 값을 가지게 된다. 따라서 (X, Y)의 결합분포의 확률이 양의 기울기를 가지는 직선에 집중될수록 공분산은 양의 큰 값을, 음의 기울기를 가지는 직선에 집중될수록 음의 큰 값을 가진다. 공분산이 0인 경우는 (X, Y)결합분포의 확률이 $(\mu_X,\ \mu_Y)$점으로부터 이차원 공간상에 균등하게(거리와 방향에 치우침이 없이) 퍼져있는 경우라 할 수 있다.

공분산의 크기는 X와 Y가 취하는 값의 단위에 의존하므로, 그 값만으로 두 변수