제13주 분산과 공분산

hylee@silla.ac.kr

확률및통계(2)

제 4절 분산과 공분산

- 자료가 평균에서 얼마나 떨어져있는가를 말해주는 지표로서 분산을 공부한다.

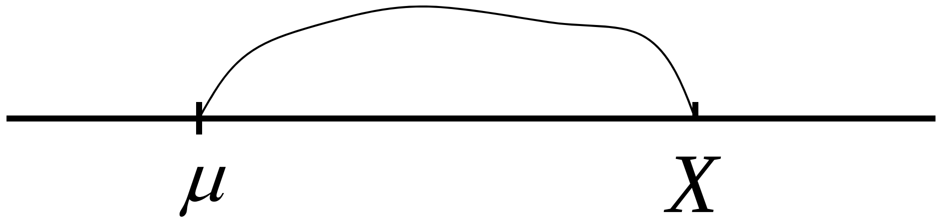
$$Var(X) = E\left(\text{X 와 평균의 떨어진 거리의 제곱에 대한 기대값} \right)$$

X 와 μ 의 떨어진 거리

$$X - \mu$$

$$(X - \mu)^2$$

$$E[(X - \mu)^2]$$



$$E(X) = \mu$$

[정의] 분산

X 는 기대값이 $E(X) = \mu$ 인 확률변수

일 때, X 의 **분산**(variance)을 $Var(X)$ 로

나타내며, 다음과 같이 정의한다.

$$Var(X) = E[(X - \mu)^2]$$

*** 분산의 간편식

$$
\begin{aligned}
Var(X) = E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - E(2\mu X) + E(\mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

예제 21

주사위를 던져서 나오는 눈의 수를 X 라 할 때, X 의 분산을 구하시오.

[풀이]

$$E(X) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$$

$$E(X^2) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + \cdots + 6^2\left(\frac{1}{6}\right) = \frac{91}{6}$$

이므로 분산은

$$Var(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

[정리]

임의의 상수 a, b 에 대하여 다음이 성립한다.

$$Var(aX + b) = a^2 Var(X)$$

[증명]

$$E(aX+b)=aE(X)+b$$

이므로

$$Var(aX+b)=E[(aX+b)^2]-[E(aX+b)]^2$$

$$=E[a^2X^2+2abX+b^2]$$

$$-\{a^2(E(X))^2+2abE(X)+b^2\}$$

$$=a^2E(X^2)+2abE(X)+b^2$$

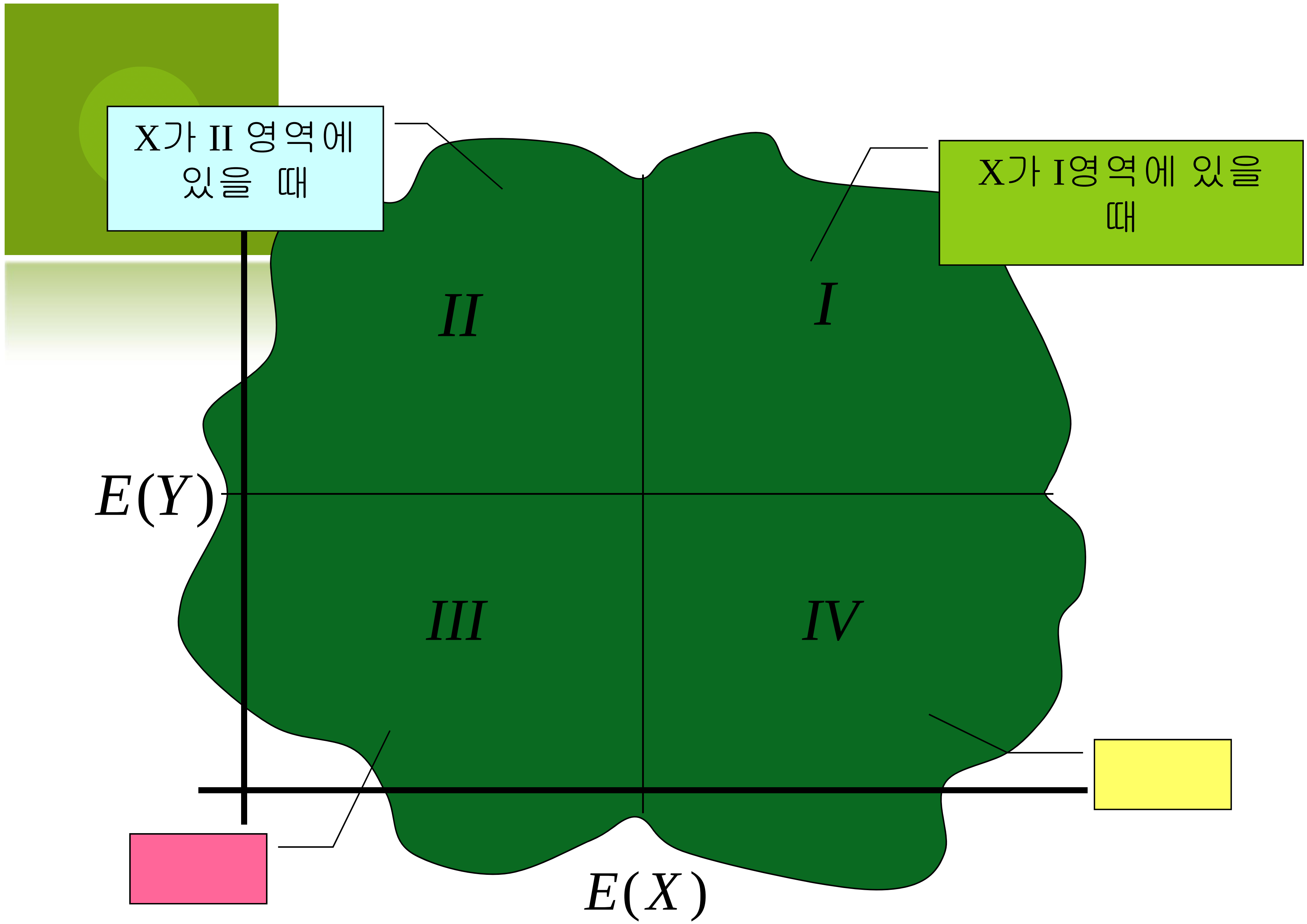
$$-a^2(E(X))^2-2abE(X)-b^2$$

$$=a^2\{E(X^2)-[E(X)]^2\}=a^2Var(X)$$

[정의] 공분산

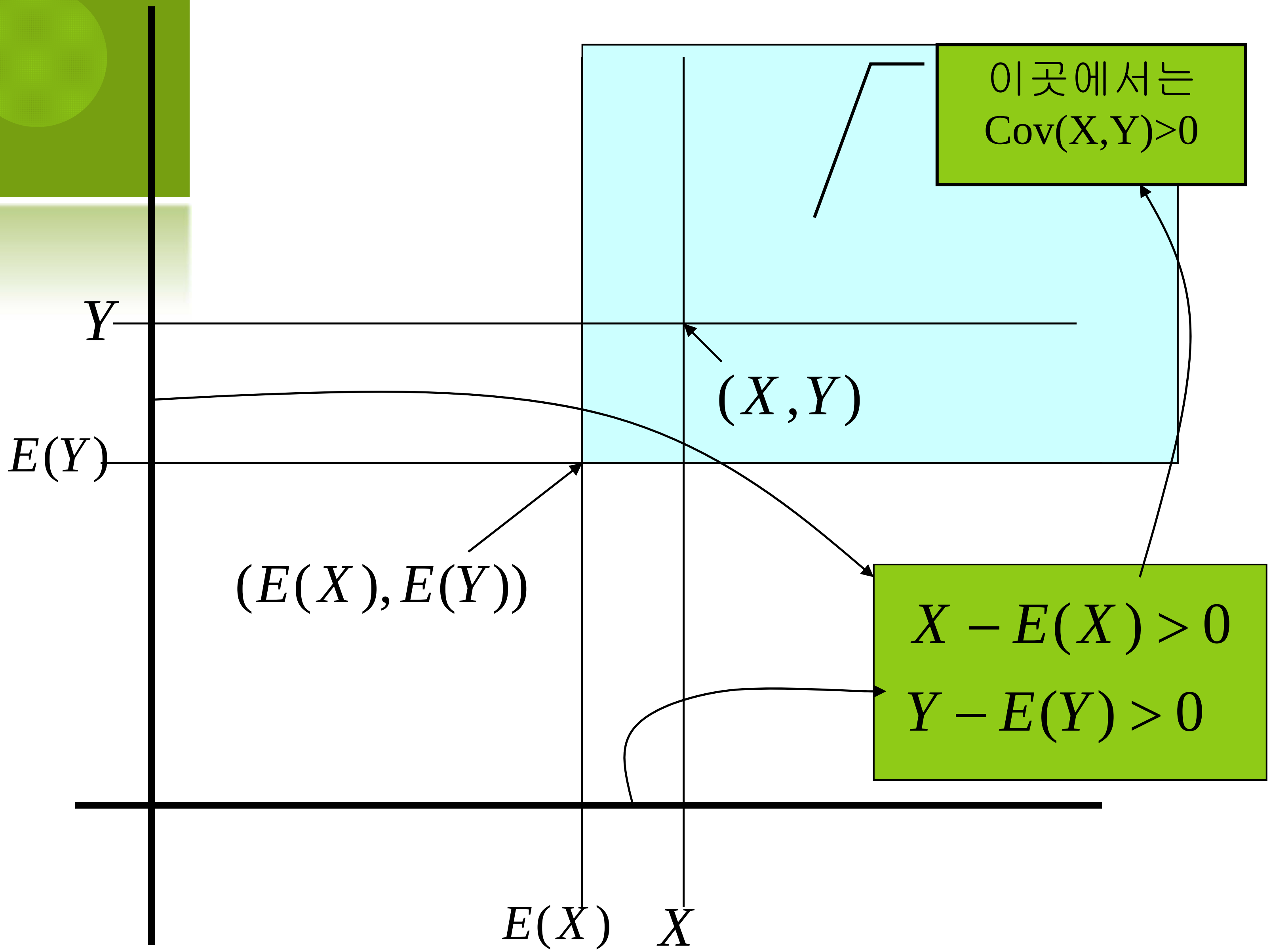
확률변수 X 와 Y 의 공분산(covariance)을 $Cov(X,Y)$ 로 나타내며, 다음과같이 정의한다.

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


X가 II 영역에 있을 때
X가 I영역에 있을 때
II
I
III
IV
$E(Y)$
$E(X)$

각 영역에 있을 때 $X - E(X)$ 의 값과 $Y - E(Y)$ 의 값을 비교해보자.

	X-E(X)	Y-E(Y)	((X-E(X))(Y-E(Y)))
I 영역	양수	양수	양수
II 영역	음수	양수	음수
III 영역	음수	음수	양수
IV 영역	양수	음수	음수



이곳에서는
Cov(X,Y)>0
Y
$E(Y)$
(X, Y)
$(E(X), E(Y))$
$X - E(X) > 0$
$Y - E(Y) > 0$
$E(X)$
X

위 정의를 전개해보자

$$Cov(X,Y)$$

$$= E[XY - E(X)Y - XE(Y) + E(X)E(Y)]$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y)$$

이다. 특히, X 와 Y 가 서로독립이면

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$= 0$$

두 확률변수의 합의 분산

$$
\begin{aligned}
Var(X+Y) &= E[(X+Y-E(X+Y))^2] \\
&= E[(X+Y-E(X)-E(Y))^2] \\
&= E[((X-E(X))+(Y-E(Y))^2] \\
&= E[(X-E(X))^2+(Y-E(Y))^2 \\
&\qquad\qquad +2(X-E(X))(Y-E(Y))] \\
&= E(X-E(X))^2+E(Y-E(Y))^2 \\
&\qquad\qquad +2E[(X-E(X))(Y-E(Y))]
\end{aligned}
$$

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

가 된다. $X - Y$ 에 대한 분산은

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