SOOKMYUNG WOMEN'S UNIVERSITY K-MOOC

■ 기댓값

- $E(X) = \sum_x x f_X(x)$, $E(Y) = \sum_y y f_Y(y)$

- 확률변수 X와 Y에 대해, $X+Y$의 기댓값? XY의 기댓값?
  ⇨ 두 변수를 고려한다는 것은 일단 두 변수에 대한 결합분포가 있다는 것을 전제
  ⇒ 결합확률질량함수나 결합확률밀도함수를 이용

SOOKMYUNG WOMEN'S UNIVERSITY K-MOOC

● 이산확률변수

○ $$\begin{aligned} E(X+Y) &= \sum_x \sum_y (x+y) f(x,y) \\ &= \sum_x \sum_y x f(x,y) + \sum_x \sum_y y f(x,y) \\ &= \sum_x x f_X(x) + \sum_y y f_Y(y) \\ &= E(X) + E(Y) \end{aligned}$$

○ $$E(XY) = \sum_x \sum_y xy f(x,y)$$

- **기댓값 정리**

① $E(X+Y) = E(X) + E(Y)$

② X와 Y가 독립이면 $E(XY) = E(X)E(Y)$

SOOKMYUNG WOMEN'S UNIVERSITY K-MOOC

■ 공분산(Covariance)

- 표본공분산 $c_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
  - 표본이 가질 수 있는 값 $\{x_1, x_2, \ldots, x_{k1}\}$, $\{y_1, y_2, \ldots, y_{k2}\}$
  - n_{ij}: 표본 중 (x_i, y_j) 값을 가지는 표본의 수

$$\begin{aligned} c_{x,y} &= \frac{1}{n-1}\sum_{i=1}^{k1}\sum_{j=1}^{k2} n_{ij}(x_i - \overline{x})(y_j - \overline{y}) \\ &= \frac{n}{n-1}\sum_{i=1}^{k1}\sum_{j=1}^{k2} p_{ij}(x_i - \overline{x})(y_j - \overline{y}) \end{aligned}$$

- 두 확률변수 X와 Y의 공분산

$$
\begin{aligned}
Cov(X, Y) &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y) = E((X-\mu_X)(Y-\mu_Y)) \\
&= \sum_x \sum_y xyf(x,y) - \mu_X \mu_Y = E(XY) - E(X)E(Y)
\end{aligned}
$$

- X와 Y가 독립이면, $E(XY) = E(X)E(Y)$

$$\Rightarrow \; Cov(X, Y) = 0$$

  - 역은 일반적으로 성립하지 않음

⦿ 결합확률분포표

$x \backslash y$	-1	0	1	$f_X(x)$
0	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{2}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
$f_Y(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

○ $E(X) = 1/3$, $E(Y) = 0$, $E(XY) = 0$ ⇨ $Cov(X, Y) = 0$

○ $f(0,-1) = \frac{1}{3} \neq \frac{2}{3} \times \frac{1}{3} = f_X(0) f_Y(-1)$ ⇨ 독립 아님

● 기댓값 정리

③ $Cov(aX+b, cY+d) = abCov(X, Y)$

④ $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$

⑤ X와 Y가 독립이면, $Var(X \pm Y) = Var(X) + Var(Y)$

■ 상관계수(coefficient of correlation)

- 표준화 변수들의 공분산
  - $U = \dfrac{X - \mu_X}{\sigma_X}, \quad V = \dfrac{Y - \mu_Y}{\sigma_Y} \quad \Rightarrow \; E(U) = E(V) = 0$
  - $Cov(U, V) = E(UV) = \dfrac{Cov(X, Y)}{\sigma_X \sigma_Y}$
- 두 확률변수 X와 Y의 상관계수

$$\rho_{XY} = Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

- 상관계수의 성질
  - $-1 \le \rho \le 1$
  - 어떤 직선을 중심으로 확률(밀도)이 모여 있을수록 $|\rho|$는 1에 근접
  - 상수 $a \neq 0$에 대해, $Y = aX + b$이면 $|\rho_{XY}| = 1$
  - $Cor(aX+b, cY+d) = sign(a)sign(b)Cor(X, Y)$

■ 요약

- 공분산: $Cov(X, Y) = \begin{cases} \sum\sum (x-\mu)(y-\mu) f(x, y) \\ \int\int (x-\mu)(y-\mu) f(x, y)\, dy\, dx \end{cases}$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- X, Y가 독립이면, $E(XY) = E(X)E(Y)$
- 상관계수: $\rho_{XY} = Cor(X, Y) = \dfrac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$