4.3 표준화(standardization)

$$Z = \frac{X - E(X)}{sd(X)}$$

E(Z)=0

Var(Z)=1

정리하면 나온다.

5.1 두 확률변수의 결합분포
: 두 개의 확률변수가 취할 수 있는 값들의 모든 순서 짝에 확률이 흩어진 정도를 합이 1인 양수로 나타낸 것

예) 서로 다른 세 개의 동전 A,B,C 를 던지는 실험
X=앞면의 수
Y=뒷면의 수

5.2 결합확률밀도함수 (joint pdf)

- 이산형 확률변수(X,Y) 의 확률밀도함수

$$p(x,y)=\begin{cases} P(X=x_i, Y=y_j) & (x,y)=(x_i,y_j), \quad i,j=1,2,\dots \\ 0 & o.w \end{cases}$$

이산형 확률밀도함수의 성질
(1) $0 \le p(x,y) \le 1,\ \sum_{모든(x,y)} p(x,y)=1$
(2) $P(a<X\le b, c<X\le d)=\sum_{a<x\le b}\sum_{c<y\le d} p(x,y)$

- 연속확률변수 (X,Y) 에 대하여 함수 $p(x,y)$가 다음을 만족할 때,

(1) $p(x,y) > 0, \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dx\,dy = 1$

(2) $P(a \le X \le b, c \le X \le d) = \int_c^d \int_a^b p(x,y)dxdy$ (단, $-\infty \le a < b \le \infty, -\infty \le c < d \le \infty$)

$p(x,y)$를 (X,Y) 의 확률밀도함수라고 한다.

5.4 주변확률밀도함수 (marginal pdf)

결합확률밀도함수 $p(x,y)$ 에서 유도된 각 확률변수 X,Y의 확률밀도함수

(1) 이산형 : $p_X(x) = \sum_y p(x,y)$, $p_Y(y) = \sum_x p(x,y)$

(2) 연속형 : $p_X(x) = \int_{-\infty}^{\infty} p(x,y)dy$, $p_X(x) = \int_{-\infty}^{\infty} p(x,y)dx$

예) X,Y의 결합분포가 다음과 같을 때, $p_X(x)$, $p_Y(y)$를 구하여라.

x \ y	0	1	2	3	행의 합
0	0.05	0.05	0.10	0.00	0.20
1	0.05	0.10	0.25	0.10	0.50
2	0.00	0.15	0.10	0.05	0.30
열의 합	0.10	0.30	0.45	0.15	1.00

5.5 두 확률변수의 함수의 기대값

X,Y의 결합확률밀도함수 $p(x,y)$ 및 함수 g 에 대하여 확률변수 g(X,Y)의 기대값

$$(1)\ E(g(X,Y)) = \begin{cases} \sum_{x\in s_X}\sum_{y\in s_Y} g(x,y)p(x,y) & (X,Y)\text{: 이산확률변수} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p(x,y)\,dx\,dy & (X,Y)\text{: 연속확률변수} \end{cases}$$

(2) $E(ag_1(X,Y) + bg_2(X,Y)) = aE(g_1(X,Y)) + b\mathrm{E}(g_2(X,Y))$, a,b는 상수

5.6 두 확률변수의 독립성

X,Y의 결합확률밀도함수 $p(x,y)$와 각각의 주변밀도함수 $p_X(x)$, $p_Y(y)$에 대하여 $p(x,y)=p_X(x)p_Y(y)$ 가 모든 (x,y)에 대하여 성립할 때, X와 Y는 서로 독립이다.

예) X,Y의 결합분포가 다음과 같을 때, E(XY)를 구하시오. X와 Y는 서로 독립인가?

	0	1	행의 합
0	1/8	1/8	1/4
1	1/4	1/4	1/2
2	1/8	1/8	1/4
열의 합	1/2	1/2	1.00

6 공분산과 상관계수

6.1 공분산 (covariance)

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

- 확률변수 X,Y의 선형관계의 유무 및 방향성을 나타내는 척도
- 각 확률변수가 취하는 값의 단위에 의존
- 간편계산법 $Cov(X,Y) = E(XY) - E(X)E(Y)$

6.2 상관계수 (correlation coefficient)

$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)sd(Y)}$

- 확률변수 X,Y의 선형관계의 유무, 방향성 및 강도를 나타내는 척도
- 공분산의 단위에 대한 의존도를 없애준 것

6.3 성질

- $Cov(aX+b, cY+d) = \mathrm{ac}Cov(X,Y)$
- $Corr(aX+b, cY+d) = \mathrm{sign(ac)}\ Corr(X,Y)$
- $\mathrm{Var}(X \pm Y) = \mathrm{Var(X)} + \mathrm{Var(Y)} \pm 2\ Cov(X,Y)$
- $-1 \le \rho \le 1$
- $Y = a + bX$의 관계에 있으면 $\rho=1(b>0)$ 또는 $\rho=-1(b<0)$

6.4 독립인 두 확률변수의 기대값과 성질

(1) E(XY)=E(X)E(Y)

(2) Cov(X,Y)=0, Corr(X,Y)=0

(3) Var(X+Y)=Var(X) + Var(Y)

해보면 나온다.