◦ Useful properties of covarience.

① $Cov(X,Y) = Cov(Y,X)$

② $Cov(X,X) = V(X)$

③ $Cov(aX,Y) = a\,Cov(X,Y)$

④
$$Cov\left(\sum_{i=1}^{n} X_i , \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} cov(X_i , Y_j)$$

3. 증명.
$$\begin{aligned} & Cov(aX,Y) \\ &= E(aXY) - E(aX)\cdot E(Y) \\ &= a\cdot E(XY) - aE(X)E(Y) \\ &= a\cdot(E(XY) - E(X)\cdot E(Y)) = a\cdot Cov(X,Y) \end{aligned}$$

4. 증명.
$$\begin{aligned} Cov\left(\sum_{i=1}^{n} X_i , \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E(X_i)\right)\cdot\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} E(Y_j)\right)\right] \\ &= E\left[\sum_{i=1}^{n}(X_i - E(X_i))\cdot\sum_{j=1}^{m}(Y_j - E(Y_j))\right] \\ &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - E(X_i))(Y_j - E(Y_j))\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} E[(X_i - E(X_i))(Y_j - E(Y_j))] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} cov(X_i , Y_j) \end{aligned}$$

∘Variance of a sum of Random Variables.

Variance of a sum of random variables $X_1, X_2 \cdots, X_n$

can be calculated as follows.

$$Var\left(\sum_{i=1}^{n} X_i\right) = cov\left(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{n} X_j\right)$$ < properties, ; $Cov(X,X) = V(X)$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} cov(X_i, X_j)$$

(i와 j가 같을 때. 그리고 나머지 경우에 대해 고려)

$$= \sum_{i=1}^{n} (V(X_i)) + \sum_{i \neq j} cov(x_i, x_j)$$

$$= \sum_{i=1}^{n} V(x_i) + 2 \cdot \sum_{i<j} cov(x_i, x_j)$$

$Cov(x_2, x_1)$

$$= Var(x_1) + Var(x_2) \cdots Var(x_n), + 2 \cdot \{ Cov(x_1, x_2) + Cov(x_1, x_3) \cdots + Cov(x_{n-1}, x_n) \}$$

n개

몇개?

$x_1 \rightarrow n-1$

$x_2 \rightarrow n-2$

$\vdots$

$x_{n-1} \rightarrow 1$

$$\sum_{k=1}^{n-1} (n-k) = n \cdot (n-1) - \left(\frac{(n-1)\cdot(n-1+1)}{2}\right)$$

$$= n^2 - n - \frac{n^2-n}{2} = \frac{2n^2 - 2n - n^2 + n}{2}$$

$$= \frac{n^2 - n}{2} = \frac{n(n-1)}{2}$$

o Variance of a Sum of Random Variables.

합의 분산 "

Solutions.

$$Var(X_1 + X_2) = Cov(X_1 + X_2,\ X_1 + X_2)$$

상식? 의미적으로 정리.

$$= Cov(X_1 + X_2, X_1) + Cov(X_1 + X_2, X_2)$$

$$= Cov(X_1, X_1) + Cov(X_1, X_2) + Cov(X_1, X_2) + Cov(X_2, X_2)$$

$$= V(X_1) + 2 \cdot Cov(X_1, X_2) + V(X_2)$$

+ (X, Y 가. 독립이라면, $Cov(X, Y) = 0$ 임.

$$= V(X_1) + V(X_2)$$

o 일련의 여러 key들을 가지고,

추가. properties 정리.

o If $X_1, X_2, \cdots, X_n$ are pairwise independent, in that X_i and X_j are independent for $i \neq j$, then.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) \qquad \left(\because \sum_{i>j} cov(X_i, X_j) = 0 \ \forall\ i, j\right)$$

즉, $V(X_1 + X_2 + X_3 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$

o S^2 에 대한 '선' 정리

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random Variables having expected value μ and Variance σ^2.

- Let the sample mean be $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- The quantities $X_i - \bar{X}$, $i = 1, \ldots, n$, are called deviations. as they equal the differences between the individual data and the sample mean.

• The random variable $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$

~~The random variance~~ is called the sample Variance

∘ $Cov(X_i - \bar{X}, \bar{X}) = 0$ 에 대한 조건과. 증명.

Example) Let $X_1, \dots X_n$ be independent[1] and identically distributed[2] random variables having Variance σ^2. show that $Cov(X_i - \bar{X}, \bar{X}) = 0$

Solutions.

$$Cov(X_i - \bar{X}, \bar{X}) = Cov(x_i, \bar{X}) - Cov(\bar{X}, \bar{X})$$

$$= Cov(x_i, \frac{1}{n} \cdot \sum_{i=1}^{n} x_i) - V(\bar{X})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \cdot Cov(x_i, x_i), - V(\bar{x})$$

$$Cov(x_i, \cancel{x_1}) + Cov(\cancel{x_i}, x_2) + \dots + Cov(x_i, x_i) + \dots Cov(\cancel{x_i}, x_n)$$

$$\underset{0}{} $$

$$= \frac{1}{n} \cdot Cov(x_i, x_i) - \frac{\sigma^2}{n}$$

$$= \frac{1}{n} \cdot \sigma^2 - \frac{\sigma^2}{n} = 0$$

✓ 이항분포의 분산에 대한 정의에 앞서, 이항분포의, 유도 및 통계량부터 순차적 접근 _

Bernoulli distribution

절차 : 베르누이 시행 > 성공 or 사건이 발생할 확률 p

> n번 ~~반복~~ 반복 > 성공횟수 x.

이때 x의 확률분포가. 이항분포.

$$P(x) = {}_nC_x \cdot p^x \cdot (1-p)^{n-x}$$: 이항분포의 함수식

$$= \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}.$$

Binomial distribution.

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

> ~

> 이항분포의 통계량인 평균과. 분산.

$$1.\ E(X) = \sum x \cdot p(x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$$

${}_nC_x$

$$= \sum_{x=0}^{n} x \cdot \frac{n!}{x! \cdot (n-x)!} \cdot p^{x} \cdot (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \cdot \frac{n!}{x! \cdot (n-x)!} \cdot p^{x} \cdot (1-p)^{n-x}$$

변형

$$= \sum_{x=1}^{n} \frac{n \cdot (n-1)!}{x \cdot (x-1)!\,(n-x)!} \cdot p \cdot p^{x-1} \cdot (1-p)^{n-x}$$

$$= np \cdot \sum_{x=1}^{n} \frac{(n-1)!}{x \cdot (x-1)!\,(n-x)!}\, p^{x-1} \cdot (1-p)^{n-x}$$

(치)환, $n-1=m$, $x-1=r$. $\rightarrow n-x=m-r$

$$= np \cdot \sum_{x=1}^{n} \frac{m!}{r! \cdot (n-x)!}\, p^{r} \cdot (1-p)^{n-x}$$

$$= np \cdot \sum_{x=1}^{n} \frac{m!}{r! \cdot (m-r)!}\, p^{r} \cdot (1-p)^{m-r}$$

$$\begin{pmatrix} x \leftarrow 1,\ r=0 \\ x \leftarrow n,\ r=m \end{pmatrix}$$

$$= np \cdot \sum_{r=0}^{m} \frac{m!}{r! \cdot (m-r)!}\, p^{r} \cdot (1-p)^{m-r}$$

$$= n \cdot p \underbrace{\sum_{r=0}^{m} \binom{m}{r} \cdot p^{r} \cdot (1-p)^{m-r}}_{=1}$$

$$X \sim B(n, p) \rightarrow E(X) = np$$

$$2.\ V(X) = E(X^2) - \{E(X)\}^2 = \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x} - n^2p^2$$

$$= \sum_{x=1}^{n} x \cdot x \frac{n \cdot (n-1)!}{x \cdot (x-1)! \cdot (n-x)!}\, p \cdot p^{(x-1)} \cdot (1-p)^{n-x} - n^2p^2$$

>

$m-1-(r+1)$

$$= np \sum_{x=1}^{n} x \cdot \frac{(n-1)!}{(x-1)!(n-x)!} \cdot p^{x-1} \cdot (1-p)^{n-x} - n^2p^2$$

(치환, $n-1=m$, $x-1=r$)

$$= np \sum_{r=0}^{m} \cdot (r+1) \frac{m!}{r!(m-r)!} p^r \cdot (1-p)^{m-r} - n^2p^2$$

$$= np \cdot \underbrace{\sum_{r=0}^{m} r \cdot \frac{m!}{r!(m-r)!} p^r \cdot (1-p)^{m-r}}_{= E(r) = mp} + np \cdot \underbrace{\sum_{r=0}^{m} \frac{m!}{r!(m-r)!} p^r (1-p)^{m-r}}_{=1} - n^2p^2$$

$$= np \cdot mp + np - n^2p^2$$

$np - p$

$$= np \cdot (n-1)p + np - n^2p^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(\underbrace{1-p}_{=q})$$

$$= npq \qquad = V(X)$$

최종 results.

Issue 정리.

- ✓ 확률이 작거나, 클수록 (p) X의 좌 or 우로 이동.
- ✓ n이 30 정도부터 정규분포에 근사해짐.
- ✓ 단. p가. 너무 크거나. 너무 작지 않은 곳에서 가능.

→ 정규분포에 근사하기 어렵다, 스퀴니스 (skewness) 가 크다.

(비대칭도)

+ n이 충분히 클 때 p가. 0에 가까우면 포아송분포.

- Correlation (상관계수)

  - The correlation of two random variable X, Y, denoted by $\rho(X, Y)$ (rho) is defined as.

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \quad -1 \le \rho(X,Y) \le Y.$$

∵ -1 아 1 가. 가까워질수록

r