property on only 2 r.v.'s

cf.

• n pairwisely indep. r.v.'s
• n mutually indep. r.v.'s

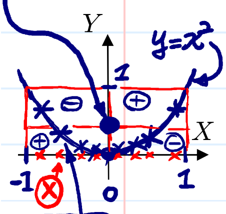
▪ Corollary. If X and Y are independent, then $Cov(X, Y)=0$, i.e., X and Y are uncorrelated. $\rho_{XY}=0$

Proof. When X, Y are independent,

$$E(XY)=E(X)E(Y)=\mu_X\mu_Y.$$

by Corollary in LNp.8-6

$(\mu_X, \mu_Y)=(0, 1/3)$



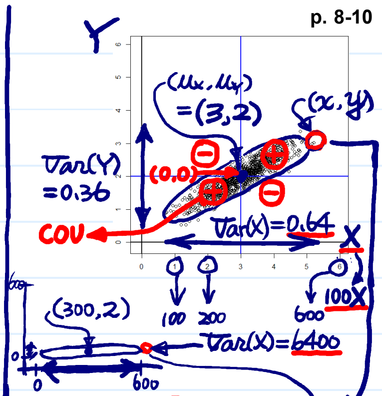
▫ However, the converse statement is *not* necessarily true. "uncorrelated" is a weaker condition than "indep.".

(e.g., let X~Uniform(−1, 1) and $Y=X^2$, then

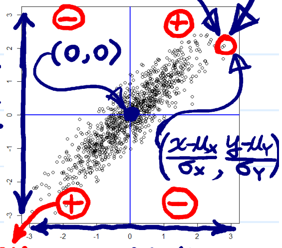
$$Cov(X, Y)=0, \leftarrow E(XY)=E(X^3)=\int_{-1}^{1} x^3 \tfrac{1}{2}dx=0, \quad E(X)E(Y)=0\cdot(1/3)=0$$

but X and Y are *not* independent).

Q: What's the dist. of $Y|X=x$? $P(Y=y|X=x)=\begin{cases}1, & \text{if } y=x^2\\ 0, & \text{o.w.}\end{cases}$



$Y'=\frac{Y-\mu_Y}{\sigma_Y}$, $Var(Y')=1$



cov = cor

$X'=\frac{X-\mu_X}{\sigma_X}$, $Var(X')=1$

Q: Why is cor unit-free?

Note. Y is a function of X. Zero correlation does not mean No relationship. cf. ✩ in LNp.8-9.

X', Y': mean = 0, variance = 1

$E(X'Y')=cov(X',Y')=cor(X',Y') = cor(X,Y)$

standardization (標準化, LNp.6-33)

▪ Corollary.

$$\rho_{XY}=E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right].$$

$E[(X-\mu_X)(Y-\mu_Y)]/\sigma_X\sigma_Y$, $\frac{X-\mu_X}{\sigma_X} = X'$, $\frac{Y-\mu_Y}{\sigma_Y} = Y'$

Proof. By definition.

➢ Example. If $(X_1, \ldots, X_m)$ ~ Multinomial(n, m, p_1, …, p_m), then

$$Cov(X_i, X_j) = -np_ip_j, \quad \text{for } 1 \le i \ne j \le m.$$

▪ Because $(X_1, X_2, X_3+\cdots+X_m)$ ~ Multinomial(n, 3, p_1, p_2, $p_3+\cdots+p_m$), and

$$X_3+\cdots+X_m = n-X_1-X_2, \qquad p_3+\cdots+p_m = 1-p_1-p_2,$$

we have

$0 \le x_1+x_2 \le n$, $0 \le x_1 \le n$, $0 \le x_2 \le n$

$$E(X_1X_2) = \sum x_1x_2\binom{n}{x_1, x_2, n-x_1-x_2}p_1^{x_1}p_2^{x_2}(1-p_1-p_2)^{n-x_1-x_2}$$

$2 \le x_1+x_2 \le n$, $1 \le x_1 \le n-1$, $1 \le x_2 \le n-1$

$$= \sum \cancel{x_1}\cancel{x_2}\frac{n!}{\cancel{x_1}!\cancel{x_2}!(n-x_1-x_2)!}p_1^{x_1}p_2^{x_2}(1-p_1-p_2)^{n-x_1-x_2}$$

(x_1-1), (x_2-1)

$$= n(n-1)p_1p_2\left[\sum \frac{(n-2)!}{(x_1-1)!(x_2-1)!(n-x_1-x_2)!} \times (p_1)^{\underbrace{x_1-1}_{y_1}}(p_2)^{\underbrace{x_2-1}_{y_2}}(1-p_1-p_2)^{\underbrace{n-x_1-x_2}_{(n-2)-y_1-y_2}}\right]$$

$0 \le y_1+y_2 \le n-2$, $0 \le y_1 \le n-2$, $0 \le y_2 \le n-2$

joint pmf of Multinomial($n-2$, 3, p_1, p_2, $1-p_1-p_2$)

$$= n(n-1)p_1p_2.$$

▪ WLOG, we can get $E(X_iX_j) = n(n-1)p_ip_j$, for $i \ne j$.

Therefore, $Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j)$

$$= n(n-1)p_ip_j - (np_i)(np_j) = -np_ip_j.$$

- And, for $i \neq j$, Note. $X_1+X_2+X_3+\cdots+X_m=n$ → Why negative

$$\underline{Cor(X_i, X_j)} = \frac{-\cancel{n}\sqrt{p_i p_j}}{\sqrt{\cancel{n}\cancel{p_i}(1-p_i)}\sqrt{\cancel{n}\cancel{p_j}(1-p_j)}} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}.$$

Q: When larger? When smaller? Hint. odds $=\frac{p}{1-p}\uparrow$ as $p\uparrow$. Note. $0\le p_i+p_j\le 1$

⊙ Cov & Cor for Sums of Random Variables

cf.

➢ Notation. In the following, let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be r.v.'s and $-\infty < a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_m < \infty$ are constants.

➢ Recall. $E(a_0+a_1X_1+\cdots+a_nX_n) = a_0+a_1E(X_1)+\cdots+a_nE(X_n)$. ← exchange of E & Σ

LNp.8-4

➢ Theorem (covariance of two sums).

$$Cov(a_0 + a_1X_1 + \cdots + a_nX_n, b_0 + b_1Y_1 + \cdots + b_mY_m)$$

$$= \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j) \to [a_1 \cdots a_n] \begin{bmatrix} cov(X_i, Y_j) \end{bmatrix}_{n\times m} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

a_0, b_0 are gone.

Proof. Let $S = a_0+a_1X_1+\cdots+a_nX_n$, and $T = b_0+b_1Y_1+\cdots+b_mY_m$, then

$$S - \underbrace{E(S)}_{a_0+a_1\mu_{X_1}+\cdots+a_n\mu_{X_n}} = \sum_{i=1}^n a_i(X_i - \mu_{X_i}),$$

$$T - \underbrace{E(T)}_{b_0+b_1\mu_{Y_1}+\cdots+b_m\mu_{Y_m}} = \sum_{j=1}^m b_j(Y_j - \mu_{Y_j}),$$

$$[S - E(S)][T - E(T)] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j (X_i - \mu_{X_i})(Y_j - \mu_{Y_j}).$$

$$\begin{array}{c|cccc} & Y_1 & Y_2 & \cdots & Y_m \\ \hline X_1 & & & & \\ X_2 & & & & \\ \vdots & & & & \\ X_n & & & & \end{array}$$

$cov(X_i, Y_j)$

Therefore, $Cov(S, T) = E\{[S - E(S)][T - E(T)]\}$

$$= \sum_{i=1}^n \sum_{j=1}^m a_i b_j E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})]$$

$$= \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j).$$

cf. mean of sum (LNp.8-4)

➢ Theorem (variance of sum). $Cov(a_0+a_1X_1+\cdots+a_nX_n,\ a_0+a_1X_1+\cdots+a_nX_n)$

presents some useful information about any 2 of the n r.v.'s ← covariance matrix of $X_1, \ldots, X_n$

$$Var(a_0 + \underbrace{a_1X_1}_{Y_1=} + \cdots + \underbrace{a_nX_n}_{Y_n=})$$

gone

$$= \sum_{i=1}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j) \to [a_1, \cdots, a_n] \begin{bmatrix} cov(X_i, X_j) \end{bmatrix}_{n\times n} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$Var(X_i)$; symmetric matrix

$$= \sum_{i=1}^n \underbrace{a_i^2 Var(X_i)}_{= Var(Y_i)}$$

$$+2\sum_{1\le i<j\le n} \underbrace{a_i a_j Cov(X_i, X_j)}_{= Cov(Y_i, Y_j)}.$$

By Thm in LNp.8-9

Proof. $Cov(X_i, X_i) = Var(X_i)$ and $Cov(X_i, X_j) = Cov(X_j, X_i)$.

cf. pairwisely indep. ⇄ ~~⇐~~

- ⊙ Corollary. If $X_1, \ldots, X_n$ are uncorrelated, then

$$Var(a_0 + \underbrace{a_1X_1}_{Y_1} + \cdots + \underbrace{a_nX_n}_{Y_n}) = \sum_{i=1}^n \underbrace{a_i^2 Var(X_i)}_{Var(Y_i)}. \quad \frac{1}{n^2}\times n\sigma^2$$

gone; exchange of Var & Σ

$cor(X_i, X_j) = 0$, $\forall i, j$, $i\neq j$, i.e., $[cov(X_i, X_j)]$ is a diagonal matrix

- Corollary. If $X_1, \ldots, X_n$ are uncorrelated and

$$Var(X_1) = \cdots = Var(X_n) \equiv \sigma^2 < \infty,$$

Variance exists; Law of Large Number

$\frac{X_1+\cdots+X_n}{n} \to a_1=\cdots=a_n=\frac{1}{n}$

then $Var(\overline{X}_n) = \boxed{\sigma^2/n}$ ≈ 0 when $n\uparrow\infty$ → i.e. $\overline{X}_n \approx C_n$ when n large enough. C_n a constant → $C_n = ?\ E(\overline{X}_n) = ?$

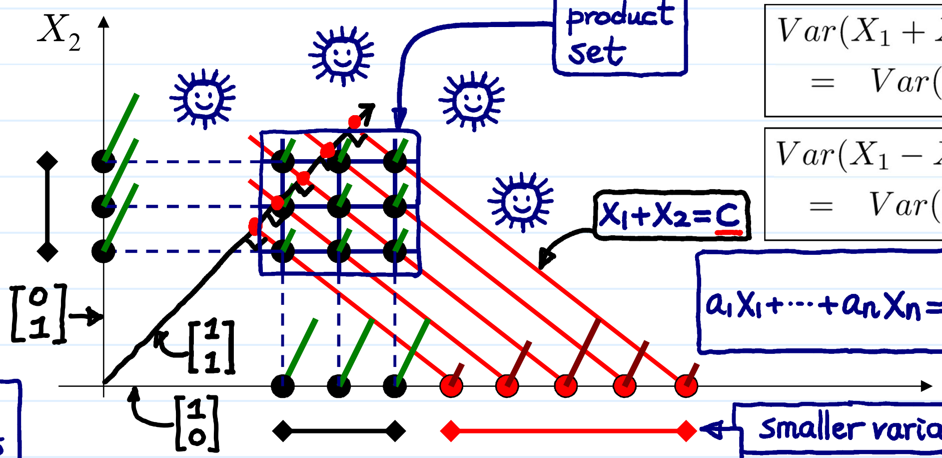
cf. ★ in LNp.8-5

Independent

$Cov(X_1, X_2) = 0$

$\sigma_{X_1}^2 = \sigma_{X_2}^2$

same marginals
different joints



$Var(X_1 + X_2)$
$= Var(X_1) + Var(X_2)$

$Var(X_1 - X_2)$
$= Var(X_1) + Var(X_2)$

$a_1X_1 + \cdots + a_nX_n = [a_1, \cdots, a_n] \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, $[a_1, \cdots, a_n] = \underset{\sim}{a}^T$, $\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \underset{\sim}{x}$

smaller variance

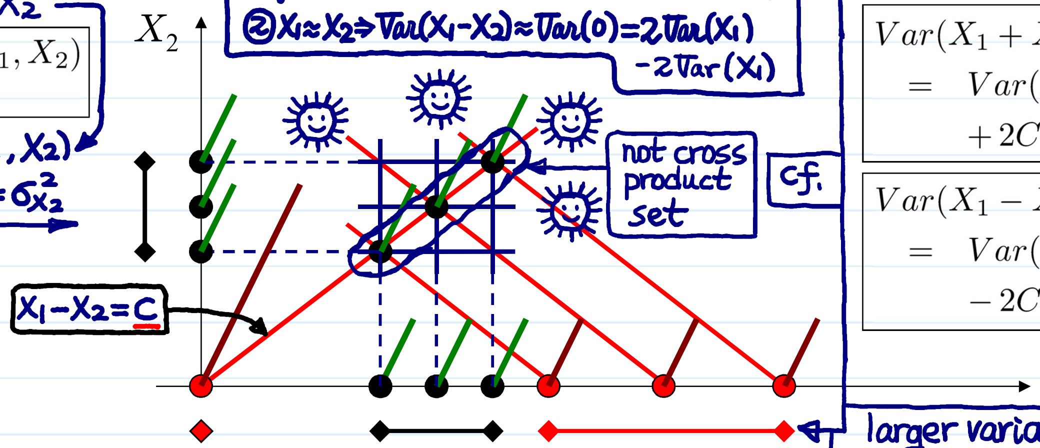
$\underset{\sim}{a}^T \underset{\sim}{x} = |\underset{\sim}{a}| \cdot \frac{\underset{\sim}{a}^T}{|\underset{\sim}{a}|} \cdot \underset{\sim}{x}$

Why? ① $X_1 \approx X_2 \Rightarrow Var(X_1+X_2) \approx Var(2X_1) = 4Var(X_1)$
② $X_1 \approx X_2 \Rightarrow Var(X_1-X_2) \approx Var(0) = 2Var(X_1) - 2Var(X_1)$

$X_1 = X_2$

$Cov(X_1, X_2) > 0$

$Cov(X_1, X_2) = \sigma_{X_1}^2 = \sigma_{X_2}^2$



cf.

$Var(X_1 + X_2)$
$= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

$Var(X_1 - X_2)$
$= Var(X_1) + Var(X_2) - 2Cov(X_1, X_2)$

larger variance