Chapter 7: Properties of Expectations and Related Quantities

Covariance and Correlation

Measures of linear relations b/w random variables

② $\sum x\, p_X(x)$ or $\int_{-\infty}^{\infty} x \cdot p_X(x)\,dx$

The **covariance** of random variables X and Y is

① $$Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y$$

Units = Units of X times Units of Y

$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, p(x,y) = \\ \quad = \sum_x \sum_y xy\, p(x,y) - \mu_X\mu_Y & \text{if } X, Y \text{ are discrete with pmf } p(x,y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\,dx\,dy = \\ \quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy - \mu_X\mu_Y & \text{if } X, Y \text{ are continuous with pdf } f(x,y) \end{cases}$$

Proof:

$E_X = \mu_X$

$$E\left((X-\mu_x)(Y-\mu_y)\right] = E\left[XY - \mu_x Y - \mu_y X + \mu_x \mu_y\right]$$

$$= E(XY) - \mu_x E(Y) - \mu_y E(X) + \mu_x \mu_y = E(XY) - \cancel{\mu_x \mu_y} - \mu_y \mu_x + \cancel{\mu_x \mu_y}$$

③ $$\boxed{E(XY) - \mu_x \mu_y}$$

Example 1: The table shows the joint probability distribution of $X =$ the deductible amounts for homeowner's policy and $Y =$ the deductible amounts for the automobile policy that customers purchased from a certain insurance company. Find the covariance of the two deductibles.

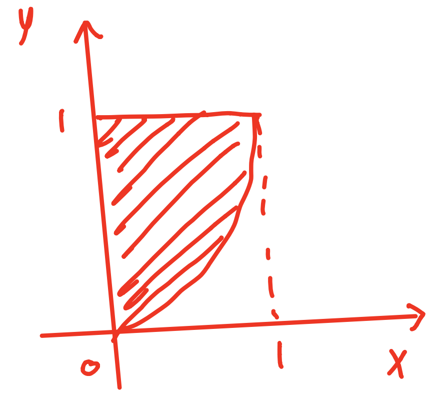
		Y			
		\$0	\$100	\$200	$P_X(x)$
X	\$100	0.10	0.30	0	0.40
	\$250	0.10	0.20	0.30	0.60
$P_Y(y)$		0.20	0.50	0.30	1

$$Cov(X,Y) = E(XY) - \mu_x \mu_y$$

- $E(XY) = \sum_x \sum_y xy\, p(x,y) = 100(0)(0.1) + 100 \cdot 100 \cdot (0.3) + 100(200)(0) + 250(0)(0.10) + 250(100)(0.2) + 250(200)(0.3)$
  $= \boxed{23000} = \$^2\, 23000$
- $E(X) = \mu_x = 100(0.4) + 250(0.6) = \boxed{190} = \$190$
- $E(Y) = \mu_y = 0(0.20) + 100(0.5) + 200(0.3) = 50 + 60 = \boxed{110} = \$100$

$\Rightarrow Cov(X = 23000 - (190 \times 110) = \boxed{2100} = 2100\,\$^2$

Example 2: Let X and Y have the joint pdf $f(x,y) = 4/3$ for $0 < x < 1,\ x^3 < y < 1$. Find the covariance between X and Y.



$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\, dx\, dy$$

$$= \int_0^1 \int_0^{y^{1/3}} xy \cdot \tfrac{4}{3}\, dx\, dy = \tfrac{4}{3}\int_0^1 y \left(\tfrac{x^2}{2}\right)\Big|_0^{y^{1/3}} dy$$

$$= \tfrac{2}{3}\int_0^1 y \cdot y^{2/3}\, dy = \tfrac{2}{3}\int_0^1 y^{5/3}\, dy = \tfrac{2}{3}\,\frac{y^{8/3}}{8/3} = \boxed{0.25}$$

$$f_X(x) = \tfrac{4}{3}(1 - x^3), \quad 0 < x < 1$$

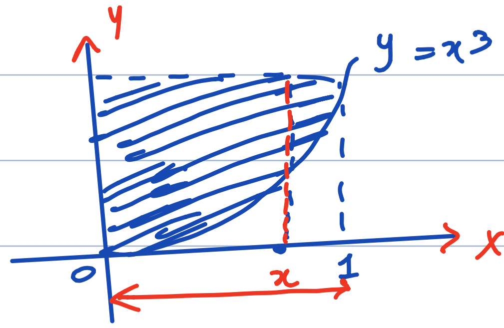
$$E(X) = \int_0^1 x \cdot \tfrac{4}{3}(1 - x^3)\, dx = \boxed{0.4} = \int_0^1\int_0^{y^{1/3}} x \cdot \tfrac{4}{3}\, dx\, dy$$

$$f_Y(y) = \tfrac{4}{3} y^{1/3}, \quad 0 < y < 1$$

$$E(Y) = \int_0^1 y \cdot \tfrac{4}{3} y^{1/3}\, dy = \boxed{0.571} = \int_0^1\int_0^{y^{1/3}} y \cdot \tfrac{4}{3}\, dx\, dy$$

$$\operatorname{Cov}(X,Y) = 0.25 - (0.4)(0.571) = \boxed{0.0216}$$

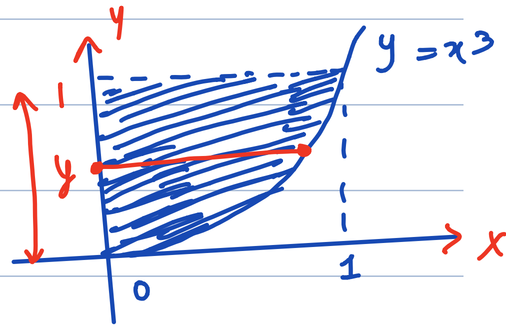
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$



$$= \int_{x^3}^{1} \frac{4}{3}\,dy = \frac{4}{3}\, y\Big|_{x^3}^{1}$$

$$= \frac{4}{3}\left(1-x^3\right), \quad 0<x<1 = S_X$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{0}^{y^{1/3}} \frac{4}{3}\,dx = \frac{4}{3}\, x\Big|_{0}^{y^{1/3}} = \frac{4}{3} y^{1/3} \text{ for } 0<y<1 = S_Y$$



Calculation of marginal is only over shaded support

Properties of Covariance:

- $Cov(X,Y) = Cov(Y,X)$ $\qquad \boxed{\sigma_{X,Y} = \sigma_{Y,X}}$
- $Cov(X,X) = E[(X-\mu_X)^2] = Var(X)$ $\qquad \boxed{E(X^2) - (\mu_X)^2}$ $\qquad \boxed{\sigma_{X,X} = \sigma_X^2}$

4 - $Cov(aX+b, cY+d) = ac\, Cov(X,Y)$

for any constant a, b, c, d

Example: Change of units

$X =$ distance	$Y =$ price/cost	
$m = 1.6\ km$	$\$ = 0.9\ €$	$Cov(X,Y) = 100\ km \cdot €$
$km = \frac{1}{1.6} m$	$€ = \frac{1}{0.9}\ \$$	$Cov = \frac{1}{1.6} \cdot \frac{1}{0.9} \cdot 100 = \frac{100}{1.44}\ m \cdot \$$

strength of linear relationship

The **correlation (coefficient)** of random variables X and Y is

5
$$\boxed{Corr(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y}}$$

Properties of Correlation:

- $Corr(X,Y) = Corr(Y,X)$ $\qquad \boxed{\rho_{X,Y} = \rho_{Y,X}}$

6 - $-1 \le Corr(X,Y) \le 1$ $\qquad \boxed{-1 \le \rho_{X,Y} \le 1}$
- $Corr(aX+b, cY+d) = \begin{cases} Corr(X,Y) & \text{if } a \,\&\, c \text{ have same sign} \\ -Corr(X,Y) & \text{if } a \,\&\, c \text{ have diff signs} \end{cases}$

Interpreting the Correlation Coefficient

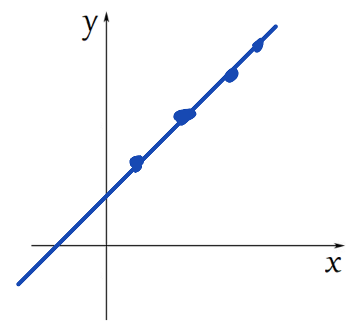
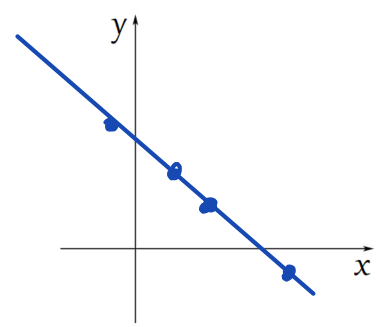
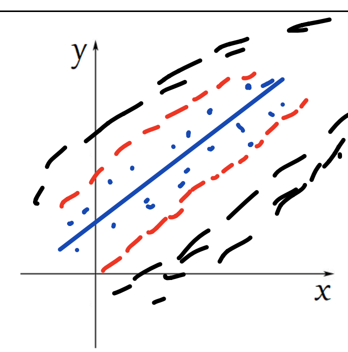
Correlation coefficient measures the **direction and strength of linear relationship** between the two random variables.

can only measure linear & nothing else

Interpretation of the value of $Corr(X, Y)$:

X and Y have a strength direction linear relationship.

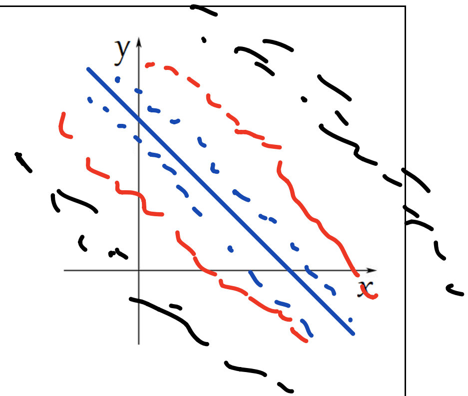
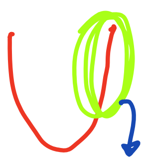
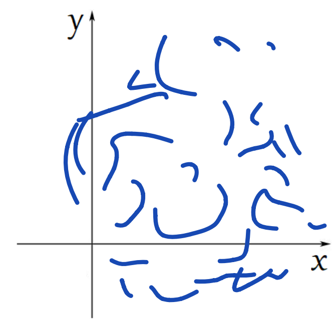
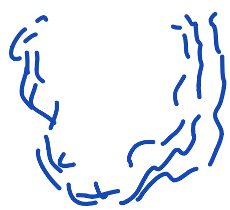
7

Correlation value	Interpretation and illustration
$\rho_{X,Y} = 1$	perfect positive linear relationship 
$\rho_{X,Y} = -1$	Perfect negative linear relationship 
$0 < \rho_{X,Y} < 1$	Strength can be strong|moderate|weak Direction is positive and linear relationship  $1 > \rho_{X,Y} \geq 0.7 \ni$ strong $0.3 \leq \rho_{X,Y} \leq 0.7 \ni$ moderate

$0 < \rho_{X,Y} \leq 0.3 \ni$ weak

Haphazard doesn't appear

linear to us.

<table>
<tr>
<td>$-1 < \rho_{X,Y} < 0$</td>
<td>strong | moderate | weak
negative linear relationship
$-1 < \rho_{X,Y} \le 0.7 \Rightarrow$ strong
$-0.7 \le \rho_{X,Y} \le 0.3 \Rightarrow$ moderate
$-0.3 \le \rho_{X,Y} < 0 \Rightarrow$ weak
</td>
</tr>
<tr>
<td>$\rho_{X,Y} = 0$

picks up linear but not true</td>
<td>• no linear relationship


⇒ quadratic
but correlation won't be able to check ∵ it can only check for linear</td>
</tr>
</table>

Example 1 (cont.): The table shows the joint probability distribution of $X =$ the deductible amounts for homeowner's policy and $Y =$ the deductible amounts for the automobile policy that customers purchased from a certain insurance company. Find and interpret the correlation between the two deductibles.

		Y			
		\$0	\$100	\$200	$P_X(x)$
X	\$100	0.10	0.30	0	0.4
	\$250	0.10	0.20	0.30	0.6
$P_Y(y)$		0.2	0.5	0.3	1

$$Cov(X,Y) = 2100$$

$$Var(X) = \left(100^2(0.4) + (250)^2(0.6)\right) - (190)^2 = \boxed{5400} \ \$^2$$

$$Var(Y) = \left((0)^2(0.2) + (100)^2(0.5) + (200)^2(0.3)\right) - (110)^2$$

$$= \boxed{4900} \ \$^2$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} = \frac{2100}{\sqrt{5400 \cdot 4900}} = \boxed{0.408}$$

no unit

unitless value

Uncorrelated vs Independent:

$\Rightarrow$ X & Y have no linear relationship

Uncorrelation

$\boxed{X \text{ and } Y \text{ are uncorrelated if } Cov(X,Y)=0}$ $\Leftrightarrow$ $\boxed{Cov(X,Y)=0}$

$\Rightarrow$ No linear relationship b/w X & Y if uncorrelated

---

Independence $\Rightarrow$ X & Y have no relation of any kind

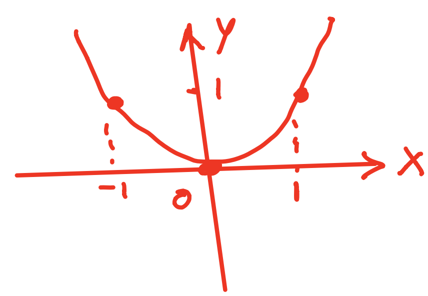
8

$\boxed{\text{If } X \,\&\, Y \text{ are independent} \Rightarrow X,Y \text{ uncorrelated} \Rightarrow Corr(X,Y)=0}$

However, if $Corr(X,Y)=0 \Rightarrow$ Not necessarily X & Y are independent
just means no linear relationship b/w X & Y

Example 3: Suppose X takes values $-1, 0, 1$ with equal probabilities. Let $Y = X^2$. Find correlation between X and Y. Are X and Y independent?

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



x	-1	0	1
$P_X(x)$	$1/3$	$1/3$	$1/3$

$E(X) = -1\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right) = \boxed{0}$

$E(XY) = E(X \cdot X^2) = E(X^3) = (-1)^3 \cdot \frac{1}{3} + 0^3\left(\frac{1}{3}\right) + 1^3\frac{1}{3}$
$= \boxed{0}$

$\boxed{Cov(X,Y) = 0 - 0 \cdot E(Y) = 0}$

$Corr(X,Y) = \dfrac{0}{\sqrt{Var(X)\,Var(Y)}} = 0 \Rightarrow$ X & Y are uncorrelated

$\Downarrow$

X & Y have no linear relationship

X, Y are not independent
$\because$ Y is function of X

Linear Combinations of Random Variables

$a_1X_1 + a_2X_2 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i$ → linear combination of RV's $X_1, X_2 \ldots X_n$ w/ coeff $a_1, \ldots a_n$

Examples:

- sum $X_1 + X_2 + \cdots + X_n$ → linear comb of $X_1, \ldots X_n$ w/ $a_1 = \cdots = a_n = 1$
- sample mean $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ or, sample avg → linear comb of $X_1, \ldots X_n$ w/ coeff $a_1 = \ldots a_n = \frac{1}{n}$
- $-X_1 + 3.4X_2 - 5X_3$ → linear combination of $X_1, X_2, X_3 \ldots$ w/ coeffs $a_1 = -1$, $a_2 = 3.4$, $a_3 = -5$

Mean of a linear combination

9

$$\boxed{E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)}$$

⇓ Always true

Variance of a linear combination

10

- In general, $Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$

$$= a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n) + 2\sum_{i<j}\sum a_ia_jCov(X_iX_j)$$

- If $X_1, X_2, \ldots, X_n$ are all pairwise uncorrelated (i.e. $Cov(X_iX_j) = 0$ for $i \neq j$), then

Same meaning ⇒ $Corr(X_i, X_j) = 0$ $i \neq j$

$$Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n)$$

→ use this if given or you can prove this first

* If one of Cov or Corr = 0 ⇒ other is 0 too
* Var must always be non negative

Example 4: Let X_1 and X_2 be random variables with means $E(X_1) = -4$ and $E(X_2) = 3$, variances $Var(X_1) = 4$ and $Var(X_2) = 9$, and covariance $Cov(X_1, X_2) = -1$. Find the mean and the variance of $Y = 3X_1 - 2X_2$.

linear combination of X_1, X_2 w/ $a_1 = 3$, $a_2 = -2$

$$E(Y) = E(3X_1 - 2X_2) = 3E(X_1) - 2E(X_2) = 3(-4) - (2 \cdot 3) = \boxed{-18}$$

$$Var(Y) = Var(3X_1 - 2X_2)$$

$$= \left(3^2 \cdot Var(X_1) + (-2)^2 Var(X_2)\right) + \left(2 \cdot 3(-2)\, Cov(X_1, X_2)\right)$$

$$= 9 \cdot 4 + 4 \cdot 9 - 12(-1) = \boxed{84}$$

Example 5: X_1 has mean μ_1 and variance σ_1^2, X_2 has mean μ_2 and variance σ_2^2, and their correlation is $Corr(X_1, X_2) = \rho$. Not sure if $\rho = 0$ or not $\therefore$ general formula

a) Find the expressions for the mean and variance of $X_1 + X_2$ and those of $X_1 - X_2$.

b) Suppose now X_1 and X_2 are uncorrelated. Repeat the tasks in part a).

a) $X_1 + X_2 \rightarrow$ linear combination of X_1, X_2 w/ $a_1 = 1$, $a_2 = 1$

$$E(X_1 + X_2) = E(X_1) + E(X_2) = \mu_1 + \mu_2$$

$$Var(X_1 + X_2) = 1^2 \cdot \sigma_1^2 + 1^2 \cdot \sigma_2^2 + 2 \cdot 1 \cdot 1 \cdot Cov(X_1, X_2)$$

$$\rho = \frac{Cov(X_1, X_2)}{\sigma_1 \sigma_2} \quad ; \quad Cov(X_1, X_2) = \rho \cdot \sigma_1 \cdot \sigma_2$$

$$\boxed{= \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$$

- $X_1 - X_2 \Rightarrow$ linear combination of X_1, X_2 w/ $a_1 = 1$ & $a_2 = -1$

$$E(X_1 - X_2) = E(X_1) - E(X_2) = \mu_1 - \mu_2$$

$$Var(X_1 - X_2) = 1^2 \sigma_1^2 + (-1)^2 \sigma_2^2 + 2 \cdot 1 (-1)\, Cov(X_1, X_2)$$

$$\boxed{= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

b) when X_1 & X_2 are un-correlated $\Rightarrow$ $\boxed{\rho = 0}$

- $X_1 + X_2 \Rightarrow E(X_1 + X_2) = \mu_1 + \mu_2$

  $Var(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$

- $X_1 - X_2 \Rightarrow E(X_1 + X_2) = \mu_1 - \mu_2$

  $Var(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$

Example 6: Let $X_1, X_2, \ldots, X_n$ be a *random sample* from a distribution with mean μ and variance σ^2. Find the expressions for the mean and variance of the

a) sum of these random variables

b) sample mean/average of these random variables

11

∵ random sample

⇒ Uncorrelated

$X_1, \ldots X_n$ are independent & identically distributed

IID

$X_1 \ldots X_n$ iid w/ mean μ & variance σ^2

- Sum $\quad T = \text{Total} = X_1 + \cdots X_n = \sum_{i=1}^{n} X_i$
- Mean $\quad E(T) = \mu + \cdots \mu = \boxed{n\mu}$
- Variance $(T) = 1^2\sigma^2 + \cdots 1^2\sigma^2 = \boxed{n\sigma^2}$
- Sample mean $\quad \bar{X} = \dfrac{X_1 + \cdots X_n}{n} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$

$$E(\bar{X}) = \frac{1}{n}\mu + \cdots \frac{1}{n}\mu = \boxed{\mu}$$

$$\mathrm{Var}(\bar{X}) = \left(\frac{1}{n}\right)^2\sigma^2 + \cdots + \left(\frac{1}{n}\right)^2\sigma^2 = \boxed{\frac{\sigma^2}{n}}$$

12

Covariance of Two Linear Combinations

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

Example 7: $E(X_1) = -4$, $Var(X_1) = 4$, $E(X_2) = 3$, $Var(X_2) = 9$, $E(X_3) = 1$, $Var(X_3) = 5$, $Cov(X_1, X_2) = -1$, $Cov(X_1, X_3) = 2.2$, and X_2 and X_3 are independent. Let $Y = 3X_1 - 2X_2$ and $W = X_2 + X_3 - X_1$. Find $Cov(Y, W)$.

linear combination of X_1, X_2, X_3 with $a_1 = -1$, $a_2 = 1$, $a_3 = 1$

linear comb of X_1, X_2 with coeff $a_1 = 3$, $a_2 = -2$

$$Cov(Y,W) = 3\cdot 1\ Cov(X_1, X_2) - 2\cdot 1\ Cov(X_2, X_2)$$
$$+ 3\cdot 1\ Cov(X_1, X_3) - 2\cdot 1\ Cov(X_2, X_3)$$
$$+ 3(-1)\ Cov(X_1, X_1) + (-2)(-1)\ Cov(X_2, X_1)$$
$$= 3(-1) - 2\cdot(9) + 3\cdot(2.2) - 2\cdot(0)$$
$$- 3\cdot(4) + 2(-1) = \boxed{-28.4}$$

* Rearrange first in correct order