MULTIVARIATE DISTRIBUTIONS

Covariances and correlations

Covariances of random variables

- The *covariance* between two random variables is a measure of how much the two tend to vary together.
- Note: Covariance is concerned only with "linear" variation.
- Definition 6: If Y_1 and Y_2 are random variables with means μ_1 and μ_2, respectively, then the COVARIANCE of Y_1 and Y_2 is:

$$Cov(Y_1, Y_2) = E[\underbrace{(Y_1 - \mu_1)}\underbrace{(Y_2 - \mu_2)}]$$

$E[Y_1] = \mu_1$

$E[Y_2] = \mu_2$

Positive covariance, negative covariance, no covariance

- If $Cov(Y_1, Y_2) > 0$ then the variables are positively (linearly) related (Large values of Y_1 are associated with large values of Y_2).
- If $Cov(Y_1, Y_2) < 0$ then the variables are negatively (linearly) related (Large values of Y_1 are associated with small values of Y_2 and vice-versa).
- If $Cov(Y_1, Y_2) = 0$ then there is no (linear) association between Y_1 and Y_2.
- Q: What is $Cov(Y_1, Y_1)$?

$E[(Y_1 - \mu_1)(Y_1 - \mu_1)] = E[(Y_1 - \mu_1)^2]$

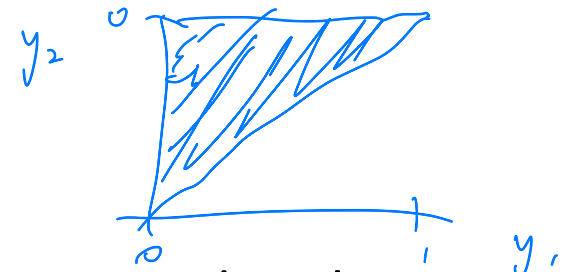
$= Var(Y_1)$

$Var(Y) = E[Y^2] - (E[Y])^2$

Covariances of random variables: An alternative expression

- Theorem 10: For any random variables Y_1 and Y_2,

$$Cov(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1]E[Y_2]$$



Example 7 re-revisited: Y_1 and Y_2 are both proportions and we know that $Y_1 < Y_2$. If their joint pdf is $f(y_1, y_2) = 6y_1$ for $0 < y_1 < y_2 < 1$ (and zero otherwise), find $Cov(Y_1, Y_2)$.

$$E[Y_1] = \int_0^1 \int_0^{y_2} y_1\, 6y_1\, dy_1\, dy_2 = \quad = \frac{1}{2}$$

$$E[Y_2] = \int_0^1 \int_0^{y_2} y_2\, 6y_1\, dy_1\, dy_2 = \quad = \frac{3}{4}$$

$$E[Y_1 Y_2] = \int_0^1 \int_0^{y_2} y_1 y_2\, 6y_1\, dy_1\, dy_2 = \int_0^1 \int_0^{y_2} 6y_1^2 y_2\, dy_1\, dy_2$$

$$= \int_0^1 y_2 \left(2y_1^3 \Big|_0^{y_2}\right) dy_2 = \int_0^1 y_2\, 2y_2^3\, dy_2 = \int_0^1 2y_2^4\, dy_2 = \frac{2}{5} y_2^5 \Big|_0^1 = \frac{2}{5}$$

$$Cov(Y_1, Y_2) = \frac{2}{5} - \frac{1}{2} \cdot \frac{3}{4} = \frac{2}{5} - \frac{3}{8} = \frac{16 - 15}{40} = \boxed{\frac{1}{40}}$$

Independence and covariance

- Theorem 11 (independence and covariance): If Y_1 and Y_2 are independent random variables, then $Cov(Y_1, Y_2) = 0$.

$$Cov(Y_1, Y_2) = \underbrace{E[Y_1 Y_2]}_{} - \overbrace{E[Y_1]E[Y_2]}^{}$$

$$= E[Y_1]\cancel{E[Y_2]} - E[Y_1]\cancel{E[Y_2]} = 0$$

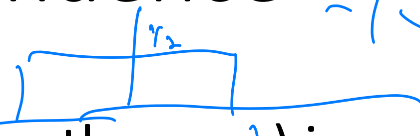
because indep.

$$\boxed{Y_1 \,\&\, Y_2 \text{ are indep.} \implies Cov(Y_1, Y_2) = 0}$$

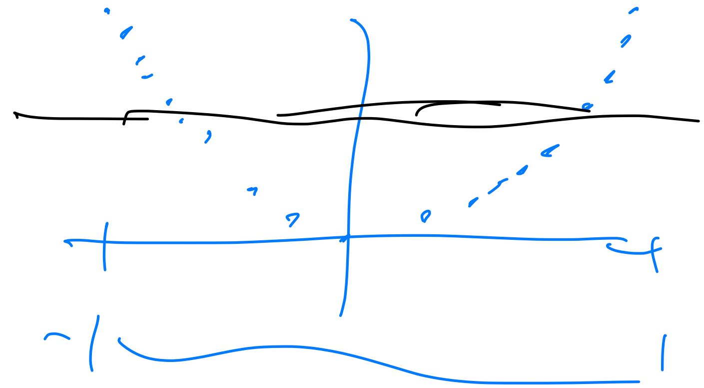
$$Cov(Y_1, Y_2) = 0 \;\not\Longrightarrow\; Y_1 \,\&\, Y_2 \text{ indep?}$$

Covariance and independence

$$f_{Y}(y_1) = \begin{cases} \frac{1}{2} & -1 \le y_1 \le 1 \\ 0 & \text{otherwise} \end{cases}$$



- Note: The converse (of the previous theorem) is not true!
- Example 10: $Y_1 \sim U(-1,1)$ and $Y_2 = Y_1^2$. Find $Cov(Y_1, Y_2)$ and determine whether Y_1 and Y_2 are independent.



$$Cov(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1]E[Y_2]$$

$$E[Y_1 Y_2] = E[Y_1 \cdot Y_1^2] = E[Y_1^3] = \int_{-1}^{1} y_1^3 \frac{1}{2}\, dy_1 = \frac{1}{8} y_1^4 \Big|_{-1}^{1}$$

$$\frac{1}{8}\big((1) - (1)\big) = 0$$

$$E[Y_1] = \int_{-1}^{1} y_1 \frac{1}{2}\, dy_1 = \frac{1}{4} y_1^2 \Big|_{-1}^{1} = \frac{1}{4}(1-1) = 0$$

$$Cov(Y_1, Y_2) = 0 - 0\,E[Y_2] = 0$$

Correlations of random variables

- The covariance will change if either Y_1 or Y_2 is expressed in different units. (Check this on your own.)
- Definition 7: For two random variables Y_1 and Y_2, the CORRELATION COEFFICIENT between Y_1 and Y_2 is
$$Corr(Y_1, Y_2) = \boxed{\rho_{Y_1,Y_2}} = \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}$$
- Theorem 12: For any random variables Y_1 and Y_2,
$$-1 \leq Corr(Y_1, Y_2) \leq 1$$
- A correlation near 1 means a strong positive (linear) relationship between Y_1 and Y_2; near -1 means a strong negative (linear) relationship.

“Perfect” correlations

- <u>Theorem 13</u>: For random variables Y_1 and Y_2, $|Corr(Y_1, Y_2)| = 1$ if and only if there are numbers $a \neq 0$ and b such that $Y_2 = aY_1 + b$. If $a > 0$ then $Corr(Y_1, Y_2) = 1$; if $a < 0$ then $Corr(Y_1, Y_2) = -1$.

Linear functions of random variables

- In statistical analysis, we often use statistics that are linear functions of multiple random variables (data) $Y_1, Y_2, \dots, Y_n$, i.e., functions that can be expressed as $\sum_{i=1}^{n} a_i Y_i = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n$ for constants $a_1 a_2, \dots, a_n$.

Linear functions of random variables

- Theorem 14: Given sequences of random variables $X_1, X_2, \dots, X_m$ and $Y_1, Y_2, \dots, Y_n$ with $E\left[X_j^2\right] < \infty$ for all j and $E\left[Y_i^2\right] < \infty$ for all i, define $U_1 = \sum_{i=1}^{n} a_i Y_i$ and $U_2 = \sum_{j=1}^{m} b_j X_j$ for constants $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_m$. Then

1. $E[U_1] = \sum_{i=1}^{n} a_i E[Y_i] = a_1 E[Y_1] + a_2 E[Y_2] + \cdots + a_n E[Y_n]$
2. $Var(U_1) = \sum_{i=1}^{n} a_i^2 Var(Y_i) + 2 \sum\sum_{i<j} a_i a_j Cov(Y_i, Y_j)$
3. $Cov(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(Y_i, X_j)$

Some "special case" corollaries:

$a_1 = 1 \quad a_2 = 1$

i. $Var(Y_1 + Y_2) = Var(Y_1) + Var(Y_2) + 2Cov(Y_1, Y_2)$

$a_1 = 1$
$a_1 = -1$

ii. $Var(Y_1 - Y_2) = Var(Y_1) + Var(Y_2) - 2Cov(Y_1, Y_2)$

iii. $Var(aY_1 + bY_2) = a^2 Var(Y_1) + b^2 Var(Y_2) + 2abCov(Y_1, Y_2)$

iv. $Cov(aW + bX, cY + dZ) = acCov(W, Y) + adCov(W, Z) + bcCov(X, Y) + bdCov(X, Z)$

def. of Var

Proof (ii)

$$Var(Y_1 - Y_2) = E\left[\left(Y_1 - Y_2 - E[Y_1 - Y_2]\right)^2\right]$$

$$= E\left[\left(Y_1 - Y_2 - E[Y_1] + E[Y_2]\right)^2\right]$$

$$= E\left[\left((Y_1 - E[Y_1]) - (Y_2 - E[Y_2])\right)^2\right]$$

$$= E\left[(Y_1 - E[Y_1])^2 + (Y_2 - E[Y_2])^2 - 2(Y_1 - E[Y_1])(Y_2 - E[Y_2])\right]$$

$$= E\left[(Y_1 - E[Y_1])^2\right] + E\left[(Y_2 - E[Y_2])^2\right] - 2E\left[(Y_1 - E[Y_1])(Y_2 - E[Y_2])\right]$$

$$= Var(Y_1) + Var(Y_2) - 2Cov(Y_1, Y_2) \qquad \square$$

Linear functions of independent random variables

$\Rightarrow Cov(Y_1, Y_2) = 0$

- Corollary: If Y_1 and Y_2 are independent random variables, then

1. $Var(Y_1 + Y_2) = Var(Y_1) + Var(Y_2)$
2. $Var(Y_1 - Y_2) = Var(Y_1) + Var(Y_2)$
3. $Var(aY_1 + bY_2) = a^2Var(Y_1) + b^2Var(Y_2)$

Already done: $E[Y_1] = \frac{1}{2}$ $E(Y_2) = \frac{3}{4}$ $Cov(Y_1, Y_2) = \frac{1}{40}$

$Var(Y_1) = E(Y_1^2) - \underbrace{(E[Y_1])^2}$ .... $= \frac{1}{20}$

Example 7 re-revisited again: Y_1 and Y_2 are both proportions and we know that $Y_1 < Y_2$. If their joint pdf is $f(y_1, y_2) = 6y_1$ for $0 < y_1 < y_2 < 1$ (and zero otherwise), Find $Var(Y_2 - Y_1)$, $Cov(Y_1, Y_2 - Y_1)$, and $Cov(Y_1 + Y_2, Y_2 - Y_1)$.

$Var(Y_2) = \ldots = \frac{3}{80}$

① $Var(Y_2 - Y_1) = Var(Y_2) + Var(Y_1) - 2Cov(Y_1, Y_2)$

$= \frac{3}{80} + \frac{1}{20} - 2\frac{1}{40} =$

② $Cov(Y_1, Y_2 - Y_1) = Cov(Y_1, Y_2) - Cov(Y_1, Y_1) = \frac{1}{40} - \frac{1}{20} = \boxed{-\frac{1}{40}}$

$\overset{\shortparallel}{Var(Y_1)}$

③ $Cov(Y_1 + Y_2, Y_2 - Y_1) = \cancel{Cov(Y_1, Y_2)} - \underbrace{Cov(Y_1, Y_1)}_{Var(Y_1)} + \underbrace{Cov(Y_2, Y_2)}_{Var(Y_2)} - \cancel{Cov(Y_2, Y_1)}$

$-\frac{1}{20} + \frac{3}{80} = \boxed{-\frac{1}{80}}$