$$E(g(X,Y)) = \sum_{\forall x, y} g(x,y) \cdot f(x,y)$$

Example: Let the joint p.f. $f(x,y)$ of (X,Y) be given by the following table:

		x		
	$f(x,y)$	0	1	2
y	1	0.1	0.2	0.3
	2	0.2	0.1	0.1

Calculate $E(XY)$.

$= (1)(0)(0.1) + (1)(1)(0.2) \cdots$

```
sum = 0
for x in range(X):
    for y in range(Y):
        sum += (x)(y)f(x, y)
```

$$E[ag_1(X,Y) + bg_2(X,Y)] = aE[g_1(X,Y)] + bE[g_2(X,Y)]$$

where a and b are constants and g_1 and g_2 are arbitrary functions.

<u>Covariance</u>

Measure of the strength of the relationship between two variables

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

$$= E(xy - x\mu_y - y\mu_x + \mu_x\mu_y)$$

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

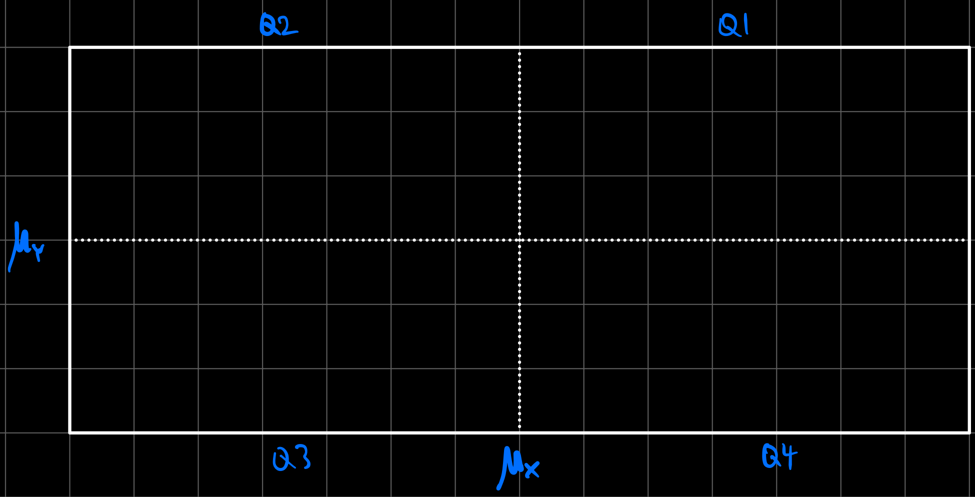
But $E(X) = \mu_X$ and $E(Y) = \mu_Y$

$\Rightarrow E(XY) - E(X)E(Y)$

Note that Cov(X,X) = E(X^2) - E(X)E(X) = E(X^2) - E(X)^2 = Var(X)

Negative covariances are ok

<u>Interpretation of covariance</u>



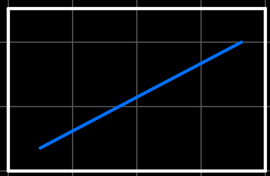
Q1 $\rightarrow (X \overset{+}{-} \mu_X)(Y \overset{+}{-} \mu_Y)$ is positive

Q2 $\rightarrow (X \overset{-}{-} \mu_X)(Y \overset{+}{-} \mu_Y)$ is negative

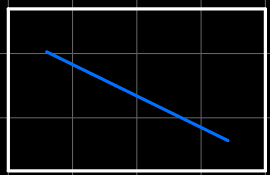
Q3 $\rightarrow (X \overset{-}{-} \mu_X)(Y \overset{-}{-} \mu_Y)$ is positive

Q4 $\rightarrow (X \overset{+}{-} \mu_X)(Y \overset{-}{-} \mu_Y)$ is negative

If most points are in quadrants 1 and 3, covariance is positive



If most points are in quadrants 2 and 4, covariance is negative



If points seem randomly scattered, covariance is roughly 0

<u>Theorem</u>

If random variables X and Y are independent, then E(XY) = E(X)E(Y)

Proof: Recall that if X and Y are independent, then

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

$$E(XY) = \sum_{\forall x, y} xy \cdot f(x, y) \quad \leftarrow \text{a lot of numbers}$$

$$= \sum_{\forall x, y} xy \cdot f_X(x) \cdot f_Y(y)$$

$$= \left[\sum_{\forall x} x \cdot f_X(x)\right]\left[\sum_{\forall x} y \cdot f_Y(y)\right]$$

$$= E(X) \cdot E(Y)$$

If X and Y are independent, then Cov(X,Y) = 0:

- E(XY) = E(X)E(Y)
- 0 = E(XY) - E(X)E(Y)
- Cov(X,Y) = 0

However, if Cov(X,Y) = 0, then X and Y are not necessarily independent

To check if they are independent, check if:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

If this fails, X and Y are not independent. However, if it passes, we still don't really know anything

Correlation

$$\rho = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

No units; measures the *strength of the linear relationship* between X and Y

Essentially a scaled version of the covariance, always lying within the interval [-1,1]

As the correlation moves closer to ±1, the relationship between X and Y becomes more linear

For covariance, interpret the *sign*, and for correlation, interpret both the magnitude and the sign

If the correlation is 0, then the covariance is 0, and X and Y are not necessarily independent

Mean and variance of linear combinations

E(aX + bY) = aE(X) + bE(Y)

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} E(a_i X_i)$$

If $E(X_i) = \mu \quad \forall i$ ← expectation

Then if $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $E(\bar{X}) = \mu$

↑actual values

Expected value of the sample mean should be the population mean, so it is our best estimator

But in practice, they probably won't be equal

Results for covariance:

Cov(X,X) = Var(X)

Suppose W = aX + bY and Z = cU + dV

Then Cov(W,Z) = E(WZ) - E(W)E(Z)

=E[(aX + bY)(cU + dV) - E(aX + bY)E(cU + dV)

When expanding, do aXcU = ac • E(XU)

This eventually gives us

ac • Cov(X, U) + ad • Cov(X, V) + bc • Cov(Y, U) + bd • Cov(Y, V)

More generally, we have

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

Results for Variance:

1. Variance of a linear combination of 2 random variables:

$$Var(aX + bY) \\ = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

Note: X and Y are not independent, so the Cov terms are needed.

If X and Y are independent, then Cov(X,Y) = 0, and:

- Var(X + Y) = Var(X) + Var(Y) + 2(1)(1)(0)
- Var(X - Y) = Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y)
    - Subtraction might produce a negative variance -> red flag

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X)$$

$$= \frac{1}{n^2} \cdot \underbrace{(\sigma^2 + \sigma^2 + \cdots)}_{n \text{ times}}$$

$$= \frac{\sigma^2}{n}$$

So, as $n \to \infty$, $Var(\bar{X}) \to 0$

More information -> less uncertainty