Lecture 3

1 Covariance

We have established how to quantify the amount of variability/fluctuations of a scalar random variable X around its mean. We would now like to generalise these notions to the case of vector random variables. After all, our data points will be higher dimensional vectors and, as mentioned before, they will be considered realisations of a vector random variable $\mathbf{X}$. Again, let us start simple by considering 2-dimensional case $d = 2$ and two random variables X_1 and X_2.

1.1 2-D Example

Let $\mathcal{D} = \{\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \ldots, \mathbf{x}^N\} \subset \mathbb{R}^2$ be our 2-D data set, $\mathbf{x}^i = [x_1^i, x_2^i]^T$, $i = 1, 2, ..., N$. We could compute the variance for each data dimension in isolation, but we might be missing an important point! What we should be asking ourselves is if the data coordinates are in some sense 'statistically linked/coupled' while their values fluctuate around their means.

Definition 1.1. Let X_1 and X_2 be random variables. Introduce a new random variable $Z = (X - \mathbb{E}[X_1]) \cdot (Y - \mathbb{E}[X_2])$. The *Co-variance* of X_1 and X_2 is defined by

$$Cov[X_1, X_2] = \mathbb{E}[Z] = \mathbb{E}[(X - \mathbb{E}[X_1]) \cdot (Y - \mathbb{E}[X_2])].$$

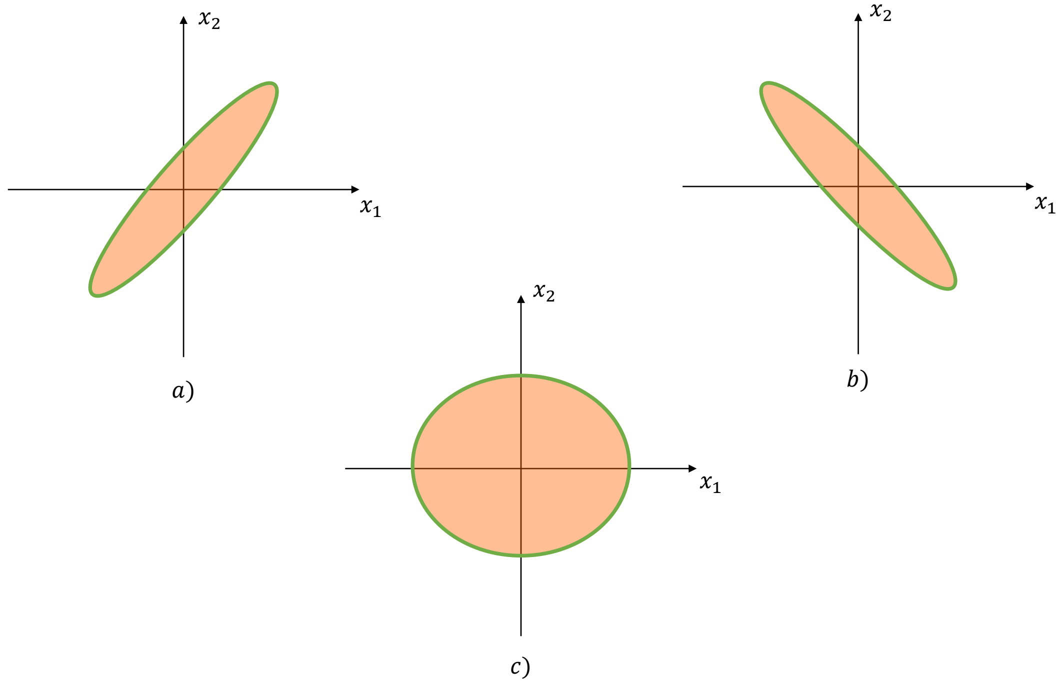
To estimate the co-variance of X_1 and X_2 in practice, if $\mathcal{D}$ is taken to be N realisations of (X_1, X_2),

$$Cov[X_1, X_2] \approx \widehat{Cov[X_1, X_2]} = \frac{1}{N} \sum_{i=1}^{N} \left(x_1^i - \widehat{\mathbb{E}[X_1]} \right) \cdot \left(x_2^i - \widehat{\mathbb{E}[X_2]} \right).$$

If the means of the random variables are centred around zero, then the estimate is simply

$$\widehat{Cov[X_1, X_2]} = \frac{1}{N} \sum_{i=1}^{N} x_1^i x_2^i.$$

To demonstrate how the co-variance of X_1 and X_2 behaves, we consider 3 cases below where the data set produces a) a positive correlation b) a negative correlation and c) no correlation.



The following table provides the parity of $(X_1 - \mathbb{E}[X_1])$ and $(X_2 - \mathbb{E}[X_2])$ and the result of $Cov[X_1, X_2]$ in each case. In case a) if $(X_1 - \mathbb{E}[X_1])$ is a positive value then $(X_2 - \mathbb{E}[X_2])$ is *on average* also positive, resulting in $Cov[X_1, X_2]$ being greater than zero. In case b) if $(X_1 - \mathbb{E}[X_1])$ is a positive value then $(X_2 - \mathbb{E}[X_2])$ is *on average* negative, resulting in negative $Cov[X_1, X_2]$. Finally, in case c) if $(X_1 - \mathbb{E}[X_1])$ is a positive value then $(X_2 - \mathbb{E}[X_2])$ can be positive or negative "with equal measure", resulting in $Cov[X_1, X_2] \approx 0$.

cases	$(X_1 - \mathbb{E}[X_1])$	$(X_2 - \mathbb{E}[X_2])$	$Cov[X_1, X_2]$
a)	$+/-$	$+/-$	> 0
b)	$+/-$	$-/+$	< 0
c)	$+/-$	$+/-$ in both cases	≈ 0

1.2 General Case

It is very unreasonable to think that our data sets will always be two-dimensional, so how we generalise our idea of the co-variance when data dimensionality gets larger than 2? We need to measure covariances of all pairs of coordinates.

Definition 1.2. Consider the vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_d)^T$, where $d \geq 2$. The *co-variance matrix* of $\mathbf{X}$ is a $d \times d$ square and symmetric matrix, defined by

$$Cov[\mathbf{X}] = \begin{bmatrix} Var[X_1] & Cov[X_1, X_2] & Cov[X_1, X_3] & ... & Cov[X_1, X_d] \\ Cov[X_2, X_1] & Var[X_2] & Cov[X_2, X_3] & ... & Cov[X_2, X_d] \\ Cov[X_3, X_1] & Cov[X_3, X_2] & Var[X_3] & ... & Cov[X_3, X_d] \\ . & . & . & ... & . \\ . & . & . & ... & . \\ Cov[X_d, X_1] & Cov[X_d, X_2] & Cov[X_d, X_3] & ... & Var[X_d] \end{bmatrix}$$

Note that, for each $i = 1, 2, \ldots, d$, $Cov[X_i, X_i] = Var[X_i]$. So along the diagonal we have variances.

Suppose now that we have N realisations of the vector random variable $\mathbf{X}$:

$$\begin{aligned} &\mathbf{x}^1 = (x_1^1, x_2^1, ..., x_d^1)^T \\ &\mathbf{x}^2 = (x_1^2, x_2^2, ..., x_d^2)^T \\ &\vdots \\ &\mathbf{x}^N = (x_1^N, x_2^N, ..., x_d^N)^T. \end{aligned}$$

Let $\chi = (\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N)$ be the design matrix for $\mathbf{X}$. Suppose also that each random variable is centred: $\mathbb{E}[X_i] = 0$, $i = 1, 2, \ldots, d$. Then we can estimate the co-variance of $\mathbf{X}$ by

$$Cov[\mathbf{X}] \approx \widehat{Cov[\mathbf{X}]} = \frac{1}{N}\chi\chi^T.$$