Let $X$ and $Y$ be random variables with finite second moments.

- the correlation: $E[XY]$
- the covariance: $\text{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$
- the correlation coefficient: $\rho_{XY} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)\text{Var}(Y)} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$

### Scaling and Linearity of Covariance

- $\text{Cov}(X+Y, U+V) = \text{Cov}(X,U) + \text{Cov}(X,Y) + \text{Cov}(Y,V) + \text{Cov}(Y,U)$
- $\text{Cov}(cX + cY + d) = c\text{Cov}(X,Y)$

### Important definitions & relationships

1. Unrelated, negatively correlated, positively correlated
2. $\text{Var}(X) = \text{Cov}(X,X)$. In particular, if $X$ and $Y$ are unrelated
   $$\text{Var}(X+Y) = \text{Cov}(X+Y, X+Y) = \text{Cov}(X,X) + \text{Cov}(Y,Y) + \text{Var}(X) + \text{Var}(Y)$$
3. If $X$ & $Y$ are independent, they are unrelated. The reverse does not hold in general.
   (Note: If $X$ & $Y$ are jointly gaussian & unrelated, then they are independent.)

**Today**

- Causality, correlation coefficient & standardized version of random variables
- Sample mean & variance of a data set

- Jointly gaussian random variables
- Causality, correlation coefficient & standardized version of random variables

Notice that $E[XY]$ and $\text{Cov}(X,Y)$ have units, i.e., their units is product of the units of $X$ and $Y$.

E.g., $X$ in km, $Y$ in km, then $\text{Cov}(X,Y)$ and $E[XY]$ are in $(\text{km})^2$.

In particular, changing unit (e.g., from km to m) changes the value of $E[XY]$ & $\text{Cov}(X,Y)$. One solution is to use the standardized version of $X$ and $Y$:

$$\text{Cov} \left( \frac{X-E[X]}{\sigma_X}, \frac{Y-E[Y]}{\sigma_Y} \right) = \text{Cov} \left( \frac{X}{\sigma_X}, \frac{Y}{\sigma_Y} \right) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \rho_{XY}$$

$$E \left[ \left( \frac{X-E[X]}{\sigma_X} \right) \left( \frac{Y-E[Y]}{\sigma_Y} \right) \right] = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} = \rho_{XY}$$
In particular, $\rho_{XY}$ is a number & does not have unit. Notice that $\rho_{aX+b, cY+d} = \rho_{XY}$ for any $a, c > 0$.

The question is, what values can $\rho_{XY}$ take.

**Proposition. Schwartz’s Inequality**

For two random var. $X$ & $Y$, we have:

$$|E[XY]| \leq \sqrt{E[X^2] E[Y^2]}$$

Moreover, if $E[X^2] > 0$, equality holds if & only if $P(Y=cX) = 1$ for some $c$.

**proof.** Let $\lambda = E[XY]/E[X^2]$. We have

$$0 \leq E[(Y-\lambda X)^2] = E[Y^2] - \lambda^2 E[X^2] - 2\lambda E[XY] = E[Y^2] - \frac{E[XY]^2}{E[X^2]}$$

Notice that equality means $E[(Y-\lambda X)^2] = 0$ which is same as $P(Y=\lambda X = 0)$ since $(Y-\lambda X)^2 \geq 0$.

**Cor. 4.8.4.** For two random variables $X$ and $Y$,

$$\text{Cor}(X, Y) \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

Furthermore, if $\text{Var}(X) > 0$ then equality holds if & only if $Y=ax+b$ for some $a$ & $b$.

Consequently, if $\text{Var}(X) \& \text{Var}(Y)$ are not zero, $\rho_{XY}$ is well-defined &

$$|\rho_{XY}| \leq 1$$

$\rho_{XY} = 1$ if & only if $Y=ax+b$ for some $a>0$ & $b$.

$\rho_{XY} = -1$ if & only if $Y=ax+b$ for some $a<0$ & $b$.

**proof.** Using Cauchy-Schwarz for $W = X-E[X] \& Z = Y-E[Y]$, we have

$$|\text{Cor}(X,Y)| = \left|E[WX]\right| \leq \sqrt{E[W^2]E[Z^2]} = \sqrt{\text{Var}(X) \text{Var}(Y)}$$

Hence, $\rho_{XY} = \frac{\text{Cor}(X,Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \in [-1, 1]$ since $|\rho_{XY}| \leq 1$.

If $\rho_{XY} = 1$ then $P(Z=aY) = 1$, i.e., $P(Y=E[Y]=a(X-E[X])) = 1$ which can be written as $P(Y=ax+b) = 1$. Since $\rho_{XY}=1$, we have $\text{Cor}(W,Z) > 0$ which implies $a > 0$.

Similarly, if $\rho_{XY} = -1$, we have $P(Y=ax+b) = 1$ for some $a < 0$.

Intuitively speaking,

$\rho_{XY} > 0$ means $X \& Y$ both tend to be larger than average or both to be smaller than average.
\( \rho_{XY} < 0 \) means \( X \) & \( Y \) tends to be opposite, i.e., \( X \) larger than \( E[X] \) tends to indicate \( Y \) is smaller than \( E[Y] \)

Sample mean & variance of a data set

Example 4.8.7 Suppose \( X_1, \ldots, X_n \) are independent and identically distributed random variables, with mean \( \mu \) and variance \( \sigma^2 \). It might be that the mean and variance are unknown, and that the distribution is not even known to be a particular type, so maximum likelihood estimation is not appropriate. In this case it is reasonable to estimate \( \mu \) and \( \sigma^2 \) by the sample mean and sample variance defined as follows:

\[
\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2.
\]

Note the perhaps unexpected appearance of \( n - 1 \) in the sample variance. Of course, we should have \( n \geq 2 \) to estimate the variance (assuming we don’t know the mean) so it is not surprising that the formula is not defined if \( n = 1 \). Recall that an estimator is called unbiased if the mean of the estimator is equal to the parameter that is being estimated. (a) Is the sample mean an unbiased estimator of \( \mu \)? (b) Find the mean square error, \( E[(\mu - \bar{X})^2] \), for estimation of the mean by the sample mean. (c) Is the sample variance an unbiased estimator of \( \sigma^2 \)?

Solution.

(a) Given data points \( X_1, \ldots, X_n \), our estimation is \( \hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} X_k \). We have

\[
E[\hat{\mu}] = E\left[ \frac{1}{n} \sum_{k=1}^{n} X_k \right] = \frac{1}{n} \sum_{k=1}^{n} E[X_k] = \mu
\]

Here, \( \hat{\mu} \) is an unbiased estimation of \( \mu \).

(b) Notice that

\[
E\left[ (\mu - \bar{X})^2 \right] = E\left[ (\mu - \frac{1}{n} \sum_{k=1}^{n} X_k)^2 \right] = \frac{1}{n^2} E\left[ (\sum_{k=1}^{n} X_k)^2 \right]
\]

\( \frac{1}{n} \sum_{k=1}^{n} E[(\mu - X_k)^2] = \frac{\sigma^2}{n} \)

where \( \sigma^2 \) follows by independence of \( X_k \)'s, i.e. for two independent random variables \( X \) and \( Y \),

\( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \). Hence, the mean square error goes to zero with rate \( \frac{1}{n} \).

(c) Notice that using a single data point, we cannot say anything about variance. Hence \( \frac{1}{n-1} \) makes sense. The main reason, however, is that in the definition of \( \hat{\sigma}^2 \), we plugged in \( \bar{X} \) instead of \( \mu \), i.e., we are using sample mean instead of the actual mean in our estimation.

\[
E\left[ \hat{\sigma}^2 \right] = E\left[ \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2 \right] = E\left[ \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \frac{1}{n} \sum_{i=1}^{n} X_i)^2 \right]
\]

\[
= \frac{1}{n-1} E\left[ \sum_{k=1}^{n} \left( \frac{n-1}{n} X_k - \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \right]
\]
where follows by indepence of $X_i$ s.

\[
\text{Var}(\frac{n-1}{n} X_k + \sum_{i=1, i\neq k}^{n} \frac{X_i}{n}) = \text{Var}(\frac{n-1}{n} X_k) + \sum_{i=1, i\neq k}^{n} \text{Var}(\frac{X_i}{n}) = \frac{(n-1)}{n} \sigma^2 + (n-1) \frac{1}{n} \sigma^2.
\]

Hence, $\hat{\sigma}^2$ is an unbiased estimator of $\sigma^2$.

---

**Def:** We say random variables $X$ and $Y$ are jointly gaussian random variables if any linear combination of $X$ and $Y$ are gaussian random variables, i.e., for any $a,b \in \mathbb{R}$, $aX + bY$ is a gaussian random variable.

**joint distribution of gaussian random variables.** Suppose that $X, Y$ are non-degenerate, i.e., they are not linearly related. Then their joint distribution is given by:

\[
f_{XY}(uv) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp \left( -\frac{(u-\mu_X)^2}{2\sigma_X^2} - \frac{(v-\mu_Y)^2}{2\sigma_Y^2} - \frac{2\rho (u-\mu_X)(v-\mu_Y)}{2\sigma_X \sigma_Y} \right)
\]

\[
= \frac{1}{\sqrt{2\pi} \sigma_X} \exp \left( \frac{(X-\mu_X)^2}{2\sigma_X^2 (1-\rho^2)} \right) \cdot \frac{1}{\sqrt{2\pi} \sigma_Y} \exp \left( \frac{(Y-\mu_Y)^2}{2\sigma_Y^2 (1-\rho^2)} \right) \cdot \frac{1}{\sqrt{1-\rho^2}} \exp \left( -\frac{\rho (X-\mu_X)(Y-\mu_Y)}{(1-\rho^2)} \right)
\]

where

\[
\mu_X = E[X], \quad \sigma_X = \text{Var}(X)
\]

\[
\mu_Y = E[Y], \quad \sigma_Y = \text{Var}(Y)
\]

\[
\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}
\]

**Important remarks:**

1. Setting $a=0$ and $b=1$, implies that $aX + bY = Y$ is $N(\mu_Y, \sigma_Y^2)$. Similarly, $X$ is $N(\mu_X, \sigma_X^2)$.

2. If $\rho = 0$, then $f_{XY}(uv) = f_X(u) f_Y(v)$. Hence uncorrelated jointly gaussian random variables are independent.

3. Let $Z = aX + bY$. Then $X$ and $Z$ are jointly gaussian.

4. If $X$ and $Y$ are gaussian, it does not mean that they are jointly gaussian.
If $X$ and $Y$ are Gaussian, it does not mean that they are jointly Gaussian.