Example: Matching Problem in Rounds

We have $n$ urns, numbered 1, ..., $n$ and $n$ balls, numbered 1, ..., $n$.

Balls placed randomly in urns, one ball per urn.

Round 1: Remove urns with matching balls.

$X_n = \#$ of balls in matching urn.

Round 2. Remove balls from remaining urns.

Repeat experiment.

Subsequent rounds:
Repeat until all balls have been placed in their matching urns.

Let $Y_n = \#$ of rounds required, starting with $n$ balls / urns.

$R_n = E[Y_n]$.

Condition on $X_n$ and use law of total expectation. We get

$R_n = E[Y_n] = \sum_{k=0}^{n} E(Y_n | X_n = k) P(X_n = k)$

$= \mathbb{P}(X_n = 0)(1 + R_n)
+ \sum_{k=1}^{n} E(Y_n | X_n = k) P(X_n = k) \quad (\dagger)$

Intuitively, since on average each round results in 1 match, it takes on average $n$ rounds to match all $n$ balls, i.e., $R_n = n$ is a guess.
We can use induction to show this rigorously.

\[ P_1 : R_1 = 1 \]

Is this true? Yes, easily seen.

Induction hypothesis : \( P_1, \ldots, P_{n-1} \) are true

where \( P_n : R_n = k \)

Need to show \( P_1, \ldots, P_{n-1} \text{ true} \Rightarrow P_n \text{ true} \)

Then (*) becomes

\[
R_n = P(X_n = 0) (1 + R_n) \\
+ \sum_{k=1}^{n} (1 + (n-k)) P(X_n = k)
\]

\[
R_n (1 - P(X_n = 0)) \\
= P(X_n = 0) + \sum_{k=1}^{n} P(X_n = k) \\
+ n \sum_{k=1}^{n} P(X_n = k) - \sum_{k=1}^{n} k P(X_n = k)
\]

\[
= 1 + n (1 - P(X_n = 0)) - 1 \\
= n (1 - P(X_n = 0)) \quad \Rightarrow \quad R_n = n
\]

Therefore, by induction \( R_n = n \) for all \( n \).
Covariance and Correlation

Let $X$ and $Y$ be two random variables. Their covariance is defined as

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

(when the expectation exists)

A more common way to compute Cov$(X, Y)$ is obtained by multiplying the product and bringing the expectation through:


$$= E(XY) - E[X]E[Y].$$

Covariance is the expected value of the product of the centered $X$ and $Y$.

Cov$(X, Y)$ will be positive when the joint distribution of $(X - E[X], Y - E[Y])$ tends to be concentrated in the 1st and 3rd quadrants of the plane.
If Cov(X, Y) were negative it indicate
a picture more like

Positive covariance: "as X increases, Y tends to increase"

Negative covariance: "as X increases, Y tends to decrease"

Remark: Covariance really measures linear dependence only.

Ex. Let X have dist’n symmetric about 0.
   Let Y = X².

   Then E(X) = 0 and E(X³) = 0, and
   Cov(X, Y) = Cov(X, X²) = E(X³) - E(X)E(X²)
   = 0
If \( X \) and \( Y \) are independent
then \( \operatorname{Cov}(X, Y) = 0 \), since
\[
\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\]
\[
= \mathbb{E}[X] \mathbb{E}[Y] - \mathbb{E}[X] \mathbb{E}[Y] = 0.
\]

The converse is not true (see previous example).

Linearity Properties of Covariance

If \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_m \) are random variables . Let
\[
X = a_1 X_1 + \ldots + a_n X_n,
\]
\[
Y = b_1 Y_1 + \ldots + b_m Y_m,
\]
where \( a_1, \ldots, a_n, b_1, \ldots, b_m \) are constants .

Then
\[
\operatorname{Cov}(X, Y) = \operatorname{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j)
\]
\[
= \mathbb{E}\left[ \sum_{i=1}^{n} a_i X_i \sum_{j=1}^{m} b_j Y_j \right] - \mathbb{E}\left[ \sum_{i=1}^{n} a_i X_i \right] \mathbb{E}\left[ \sum_{j=1}^{m} b_j Y_j \right].
\]
\[
\begin{align*}
\mathbb{E} \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} X_i Y_j \right] & = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \mathbb{E}[X_i Y_j] - \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \mathbb{E}[X_i] \mathbb{E}[Y_j] \\
& = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \mathbb{E}[X_i Y_j] - \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \mathbb{E}[X_i] \mathbb{E}[Y_j] \\
& = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \left( \mathbb{E}[X_i Y_j] - \mathbb{E}[X_i] \mathbb{E}[Y_j] \right) \\
& = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i} b_{j} \text{Cov}(X_i, Y_j)
\end{align*}
\]

**Example.** Let \( X = a_1 X_1 + \ldots + a_n X_n \)

Then
\[
\text{Var}(X) = \text{Cov}(X, X) = \text{Cov} \left( \sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j \right)
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i a_j \text{Cov}(X_i, X_j)
\]
\[
= \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)
\]
\[
= \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)
\]
Correlation

Note that covariance is not a standardized measure. If \( \text{Cov}(X, Y) = 2 \) then \( \text{Cov}(10X, 10Y) = 100 \text{Cov}(X, Y) = 200 \). But the linear dependence between \( X \) and \( Y \) and between \( 10X \) and \( 10Y \) should be the same. Therefore, we should standardize covariance.

Def. For 2 random variables \( X \) and \( Y \) their correlation coefficient is \( \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \).

The correlation coefficient satisfies \( -1 \leq \rho(X, Y) \leq 1 \). This follows from the Cauchy-Schwarz inequality. For 2 random variables \( X \) and \( Y \) the CS-inequality says
\[
\text{E}(XY)^2 \leq \text{E}(X^2)\text{E}(Y^2)
\]
Proof

Consider \((X - XY)^2\). We have

\[
0 \leq E[(X - XY)^2] = E[X^2 - 2\lambda XY + \lambda^2 Y^2] = E[X^2] - 2\lambda E[XY] + \lambda^2 E[Y^2].
\]

Minimize with respect to \(\lambda\):

\[
-2E[XY] + 2\lambda E[Y^2] = 0 \Rightarrow \lambda = \frac{E[XY]}{E[Y^2]}
\]

Plug in this \(\lambda\):

\[
E[X^2] - 2\frac{E[XY]}{E[Y^2]} E[XY] + \frac{E[XY]^2}{E[Y^2]^2} E[Y^2]
\]

\[
= E[X^2] - \frac{E[XY]^2}{E[Y^2]} \geq 0
\]

\[
\Rightarrow E[XY]^2 \leq E[X^2]E[Y^2]
\]

\[
\Rightarrow Cov(X, Y)^2 \leq Var(X)Var(Y)
\]
\[ -\sqrt{\text{Var}(X)\text{Var}(Y)} \leq \text{Cov}(X, Y) \leq \sqrt{\text{Var}(X)\text{Var}(Y)} \]

\[ -1 \leq p(X, Y) \leq 1 \]

Note that \( \text{Cov}(X, Y) = 0 \iff p(X, Y) = 0 \). However, \( p(X, Y) \) does not share the linearity properties of covariance.

**Theorem.** \( p(X, Y) = \pm 1 \) if and only if \( Y = aX + b \) for some constants \( a \neq 0 \) and \( b \).

**Proof.** Suppose \( Y = aX + b \). Then
\[
\text{Cov}(X, Y) = \text{Cov}(X, aX + b) \\
= a \text{Cov}(X, X) + \text{Cov}(X, b) \\
= a \text{Var}(X)
\]
\[
\text{Var}(Y) = \text{Var}(aX + b) \\
= a^2 \text{Var}(X)
\]
Then
\[
p(X, Y) = \frac{a \text{Var}(X)}{\sqrt{\text{Var}(X) a^2 \text{Var}(X)}} \\
= \frac{a}{|a|} = \pm 1
\]
Now suppose \( p(X, Y) = 1 \).

Let \( \sigma_x = \sqrt{\text{Var}(X)} \)
\( \sigma_y = \sqrt{\text{Var}(Y)} \)

Consider
\[
\text{Var}\left( \frac{X}{\sigma_x} - \frac{Y}{\sigma_y} \right) 
= \text{Var}\left( \frac{X}{\sigma_x} \right) + \text{Var}\left( \frac{Y}{\sigma_y} \right) - 2 \text{Cov}\left( \frac{X}{\sigma_x}, \frac{Y}{\sigma_y} \right) 
= 1 + 1 - 2p(X, Y) = 1 + 1 - 2 = 0
\]
\[
\Rightarrow \frac{X}{\sigma_x} - \frac{Y}{\sigma_y} = c, \text{ where } c \text{ is a constant.}
\]

\text{Mar. 12}

Note that \( p(X, X) = \frac{\text{Cov}(X, X)}{\sqrt{\text{Var}(X)} \cdot \text{Var}(X)} = \frac{\text{Var}(X)}{\text{Var}(X)} = 1 \)

which we could also see from that fact that \( X \) is a linear function of \( X \).

\textbf{Covariance Matrices and Cross Covariance Matrices}
Let $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_m)^T$ be $2$ random vectors. Define their cross-covariance matrix, $\text{Cov}(X, Y)$ as the $n \times m$ matrix with $(i, j)^{th}$ entry $\text{Cov}(X_i, Y_j)$, \(i = 1, \ldots, n\) \(j = 1, \ldots, m\).

i.e. $\text{Cov}(X, Y) = \text{E}(XY^T) - \text{E}(X)\text{E}(Y^T)$, where expectation is taken component-wise.

Special case: $X = Y$

In this case $\text{Cov}(X, X)$ is called the covariance matrix of $X$. This is more commonly denoted by $\text{Cov}(X)$.

Similarly, we will let $p(X, Y)$ and $p(X)$ denote the $n \times m$ and $n \times n$ correlation matrices given by $p(X, Y)_{ij} = p(X_i, Y_j)$ and $p(X)_{ij} = p(X_i, X_j)$, respectively. Note that the main diagonal of $p(X)$ is all ones.
Recall that if $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are random variables and $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ are real constants then

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

If we let $a = (a_1, \ldots, a_n)^T$ and $b = (b_1, \ldots, b_m)^T$ then in matrix vector format the above is

$$\text{Cov}(a^T X, b^T Y) = a^T \text{Cov}(X, Y) b \quad (\text{write it out}). \quad (*)$$

This generalizes.

Let $A = \begin{bmatrix} A_1 \\ \vdots \\ A_r \end{bmatrix}$ \hspace{1cm} (r x n), \hspace{1cm} $A_i$ is the $i^{th}$ row (1 x n)

Let $B = \begin{bmatrix} B_1 \\ \vdots \\ B_s \end{bmatrix}$ \hspace{1cm} (s x m), \hspace{1cm} $B_j$ is the $j^{th}$ row (1 x m)

Then

$A \text{Cov}(X, Y) B^T$ is the r x s matrix whose $(i,j)^{th}$ entry is

$(i^{th} \text{ row of } A) \text{Cov}(X, Y)(j^{th} \text{ column of } B^T)$

$$A_i \text{Cov}(X, Y) B_j^T = \text{Cov}(A_i X, B_j Y)$$

from $(*)$

$$= \text{Cov}((AX)_i, (BY)_j)$$
\[ = (i,j)\text{th entry of } \text{Cov}(AX, BY) \]

Therefore,
\[ \text{Cov}(AX, BY) = A \text{Cov}(X, Y) B^T. \]

Special case when \( X = Y \) get
\[ \text{Cov}(AX, BX) = A \text{Cov}(X) B^T \]

and if \( A = B \) then
\[ \text{Cov}(AX) = A \text{Cov}(X) A^T. \]

If \( A \) is \( 1 \times n \), i.e., \( A = a^T \) then
\[ \text{Cov}(a^T X, a^T X) = a^T \text{Cov}(X) a = \text{Var}(a^T X). \]

Properties of Covariance matrices

1. \( \text{Cov}(X) \) is nonnegative definite for any random vector \( X \), i.e., \( a^T \text{Cov}(X) a \geq 0 \) for every \( a \in \mathbb{R}^n \).

   Proof Let \( a \in \mathbb{R}^n \), then
\[ a^T \text{Cov}(X) a = \text{Var}(a^T X) \geq 0. \]

2. \( \text{Cov}(X) \) is symmetric since
\[ \text{Cov}(X)_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = \text{Cov}(X)_{ji}. \]
③ If $X$ is a continuous random vector then if $\text{Cov}(X)$ is singular then $X$ has no joint pdf.

Proof. If $\text{Cov}(X)$ is singular then there is a vector $a \in \mathbb{R}^n$ such that $\text{Cov}(X)a = 0$

Then $\text{Var}(a^T X) = a^T \text{Cov}(X)a = 0$

$\Rightarrow P(a^T X = c) = 1$ for some constant $c$.

i.e., $X$ lies on the (0-volume in $\mathbb{R}^n$) hyperplane defined by $\{x \in \mathbb{R}^n : a^T x = c\} = S$

So there can be no density satisfying

$$\int_S f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$$