# Probability_Basics_III

December 15, 2020

Author: DistilledData

## Contents

# 1 Probability Basics III

This notebook continues the discussion of probability from "Advanced_Probabiliity_Basics." In particular, this notebook will focus continue our discussion of expectation with mean, variance and covariance. This notebook will end with a brief discussion of the indicator function.

## 1.1 Mean

### 1.1.1 Sample Mean

Given a set of data $x_1, \cdots, x_n$, the sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

### 1.1.2 Population Mean

Since the mean for a random variable $X$ is given by $\mathbb{E}[X]$, I will refer the reader to the discussion on expected value and its properties in the "Advanced_Probability_Basics" notebook.

### 1.1.3 Relation between Sample and Population Mean

If each $x_i$ in a data set corresponds to the outcome of a trial for the random variable X, and if each $x_i$ is independently drawn, then we expect sample mean $\bar{x}$ to approach the theoretical mean $\mathbb{E}[X]$ as we continue to draw more and more samples. In other words,

$$\lim_{n\to\infty} \bar{x} = \mathbb{E}[X]$$

## 1.2 Variance

### 1.2.1 Sample Variance

Informally, given a particular set of data, variance measures the spread of the data. More formally, given a set of data $x_1, \cdots, x_n$ with mean $\bar{x}$, the sample variance is

$$S^2 = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

### 1.2.2 Population Variance

We can also measure variance for random variables! Given random variable $X$ with mean $\mathbb{E}[X]$, the variance is defined below

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

### 1.2.3 Relation between Sample and Population Variance

If the $x_i$ in the data above are an outcome in the outcome space for $X$ such that each $x_i$ are drawn independently, then we find that

$$\lim_{n\to\infty} S^2 = \mathrm{Var}(X)$$

## 1.3 Properties of Variance

### 1.3.1 Simplified Form

$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

**Proof**

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2 = \mathbb{E}\big[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\big] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

### 1.3.2 Mathematical Properties

Assume real-valued random variables $X$ and $Y$ with means $\mu_X$ and $\mu_Y$. We will also assume some $a, b \in \mathbb{R}$. Properties 4 and 5 use the definition of covariance that we define further below in this notebook.

1. $\mathrm{Var}(a) = 0$
2. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
3. $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
4. $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$
5. $\mathrm{Var}(aX - bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) - 2ab\mathrm{Cov}(X, Y)$

**Proof Property 1**

$$\mathrm{Var}(a) = \mathbb{E}[a^2] - \mathbb{E}[a]^2 = a^2 - a^2 = 0$$

**Property 2**

$$\begin{aligned}\mathrm{Var}(aX) &= \mathbb{E}[(aX)^2] - \mathbb{E}[aX]^2 = \mathbb{E}[(aX)^2] - (a\mathbb{E}[X])^2 \\ &= \mathbb{E}[a^2X^2] - a^2\mathbb{E}[X]^2 = a^2\mathbb{E}[X^2] - a^2\mathbb{E}[X]^2 \\ &= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2\mathrm{Var}(X)\end{aligned}$$

**Property 3**

$$\begin{aligned}
\mathrm{Var}(X+a) &= \mathbb{E}[(X+a)^2] - \mathbb{E}[X+a]^2 = \mathbb{E}[(X+a)^2] - (\mathbb{E}[X] + \mathbb{E}[a])^2 \\
&= \mathbb{E}[(X+a)^2] - (\mathbb{E}[X] + a)^2 \\
&= \mathbb{E}[X^2 + 2aX + a^2] - (\mathbb{E}[X]^2 + 2a\mathbb{E}[X] + a^2) \\
&= \mathbb{E}[X^2] + 2a\mathbb{E}[X] + a^2 - (\mathbb{E}[X]^2 + 2a\mathbb{E}[X] + a^2) \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}(X)
\end{aligned}$$

**Property 4**

$$\begin{aligned}
\mathrm{Var}(aX+bY) &= \mathbb{E}[(aX+bY)^2] - \mathbb{E}[aX+bY]^2 \\
&= \mathbb{E}[(a^2X^2 + 2abXY + b^2Y^2] - (a\mathbb{E}[X] + b\mathbb{E}[Y])^2 \\
&= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[XY] + b^2\mathbb{E}[Y^2] - (a^2\mathbb{E}[X]^2 + 2ab\mathbb{E}[X]\mathbb{E}[Y] + b^2\mathbb{E}[Y]^2) \\
&= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + b^2(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) + 2ab(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)
\end{aligned}$$

**Property 5**

$$\begin{aligned}
\mathrm{Var}(aX-bY) &= \mathbb{E}[(aX-bY)^2] - \mathbb{E}[aX-bY]^2 \\
&= \mathbb{E}[(a^2X^2 - 2abXY + b^2Y^2] - (a\mathbb{E}[X] + b\mathbb{E}[Y])^2 \\
&= a^2\mathbb{E}[X^2] - 2ab\mathbb{E}[XY] + b^2\mathbb{E}[Y^2] - (a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X]\mathbb{E}[Y] + b^2\mathbb{E}[Y]^2) \\
&= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + b^2(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) - 2ab(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) - 2ab\mathrm{Cov}(X,Y)
\end{aligned}$$

### 1.3.3 Non-Negativity

Given a real-valued random variable $X$, $\mathrm{Var}(X) \geq 0$

**Proof** $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

Suppose $X$ is a real-valued continuous random variable with density function $f_X$ and mean $\mu_X$.

$\mathrm{Var}(X) = \int_{\mathbb{R}} (x-\mu_X)^2 f_X(x) dx$

$\forall x \in \mathbb{R}\ (x-\mu_X)^2 \geq 0$ and $f_X(x) \geq 0 \implies \int_{\mathbb{R}} (X-\mu_X)^2 f_X(x) dx \geq \int_{\mathbb{R}} 0 dx = 0$

Therefore, the integral and hence $\mathrm{Var}(X)$ will also be non-negative

Suppose $X$ is a real-valued discrete random variable with PMF $p_X$ and mean $\mu_X$ $\mathrm{Var}(X) = \sum_x (X - \mu_X)^2 P(X=x)$

For all $x$ in the sum above, $(X-\mu_X)^2 \geq 0$ and $P(X=x) \geq 0$, and hence $(X-\mu_X)^2 P(X=0) \geq 0$. Since all terms in the above sum is greater or equal to 0, $\sum_x (X-\mu_X)^2 P(X=x) = \mathrm{Var}(X) \geq 0$.

#### 1.3.4 Variance of 0

$\text{Var}(X) = 0 \iff \exists a$ such that $P(X = a) = 1$

**Proof** $\implies$:

$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = 0$

Suppose X is a continuous random variable with density function $f_X$ and mean $\mu_X$.

$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (x - \mu_X)^2 f_X(x) dx = 0$

The integral is 0 $\iff$ the integrand $(x - \mu_X)^2 f_X(x)$ is uniformly 0. However, there must be some regions where the density $f_X$ is non-zero, otherwise the laws of probability will be violated. $\forall x \in \mathbb{R}$ such that $f_X(x) \neq 0$, $x = \mu_X$ in order for the integrand to be zero. However, since $\mu_X$ is a constant, there must only be one such value $x \in \mathbb{R}$ where $f_X$ is non-zero. From our discussion of continuous random variables in the "Advanced_Probability_Basics" notebook, we know that one of the properties for a continuous random variable is $\forall x \in \mathbb{R}$ $P(X = x) = 0$, which means that we must be working with a discrete random variable.

Suppose X is a discrete random variable with PMF $p_X$ and mean $\mu_X$.

$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x (X - \mu_X)^2 P(X = x) = 0$

In order for the sum to be 0, each term of the sum must be 0. However, like above, the laws of probability would be violated if $P(X = x)$ were uniformly 0, and hence there is some $x \in \mathbb{R}$ where $P(X = x)$. In this instance, in order for the term in the sum to be 0, $X = \mu_X$. Similar to the argument above, $\mu_X$ is a constant and hence $\mu_X = x$ for only one $x \in \mathbb{R}$. In order for $p_X$ to be a valid PMF, we require that $p_X(x) = P(X = x) = 1$.

$\therefore \text{Var}(X) = 0 \implies \exists a$ such that $P(X = a) = 1$

$\impliedby$:

$\mathbb{E}[X^2] = \sum_x P(X = x) x^2 = 1 \cdot a^2 = a^2$

$\mathbb{E}[X]^2 = \left( \sum_x P(X = x) x \right)^2 = a^2$

$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = a^2 - a^2 = 0$

$\therefore \exists a$ such that $P(X = a) = 1 \implies \text{Var}(X) = 0$

$\therefore \text{Var}(X) = 0 \iff \exists a$ such that $P(X = a) = 1$

## 1.4 Covariance

### 1.4.1 Sample Covariance

Given data with two covariates $X$ and $Y$ with values $(x_1, y_1), \cdots, (x_n, y_n)$, such that the sample mean for $X$ and $Y$ are $\bar{x}$ and $\bar{y}$, respectively, the covariances for the distribution is

$$\text{Cov}(x, y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$$

### 1.4.2 Population Covariance

Intuitively, covariance measures how related one random variable is to another. Given two random variables $X$ and $Y$ that form a joint distribution (X,Y), the covariance, often denoted $\sigma_{XY}$ or $\sigma(X, Y)$ is given by

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

### 1.4.3 Properties of Covariance

### 1.4.4 Simplified Form

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

**Proof:**

$$\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\
&= \mathbb{E}\big[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]\big] \\
&= \mathbb{E}\big[XY\big] - \mathbb{E}\big[Y\mathbb{E}[X]\big] - \mathbb{E}\big[X\mathbb{E}[Y]\big] + \mathbb{E}\big[\mathbb{E}[X]\mathbb{E}[Y]\big] \\
&= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

### 1.4.5 Mathematical Properties

Assume that $X$, $Y$, $W$, $V$ are real-valued random variables, and $a, b, c, d \in \mathbb{R}$

1. $\text{Cov}(X, a) = 0$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
4. $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$
5. $\text{Cov}(aX + bY, cW + dV) = ac\text{Cov}(X, W) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, W) + bd\text{Cov}(Y, V)$

**Proof Property 1**

$$\text{Cov}(X, a) = \mathbb{E}[X \cdot a] - \mathbb{E}[X]\mathbb{E}[a] = a\mathbb{E}[X] - a\mathbb{E}[X] = 0$$

**Property 2**

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[YX] - \mathbb{E}[Y]\mathbb{E}[X] = \mathrm{Cov}(Y,X)$$

**Property 3**

$$\mathrm{Cov}(aX,bY) = \mathbb{E}[aX \cdot bY] - \mathbb{E}[aX]\mathbb{E}[bY] = ab(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) = ab\mathrm{Cov}(X,Y)$$

**Property 4**

$$\begin{aligned}
\mathrm{Cov}(X+a,Y+b) &= \mathbb{E}[(X+a)(Y+b)] - \mathbb{E}[X+a]\mathbb{E}[Y+b] \\
&= \mathbb{E}[XY + aY + bX + ab] - \big(\mathbb{E}[X]\mathbb{E}[Y+b] + \mathbb{E}[a]\mathbb{E}[Y+b]\big) \\
&= \mathbb{E}[XY] + \mathbb{E}[aY] + \mathbb{E}[bX] + \mathbb{E}[ab] - \big(\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[b] + \mathbb{E}[a]\mathbb{E}[Y] + \mathbb{E}[a]\mathbb{E}[b]\big) \\
&= \mathbb{E}[XY] + a\mathbb{E}[Y] + b\mathbb{E}[X] + ab - \big(\mathbb{E}[X]\mathbb{E}[Y] + b\mathbb{E}[X] + a\mathbb{E}[Y] + ab\big) \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathrm{Cov}(X,Y)
\end{aligned}$$

**Property 5**

$$\begin{aligned}
\mathrm{Cov}(aX+bY,cW+dV) &= \mathbb{E}[(aX+bY)(cW+dV)] - \mathbb{E}[aX+bY]\mathbb{E}[cW+dV] \\
&= \mathbb{E}[(aX\cdot cW + bY\cdot cW + aX\cdot dV + bY\cdot dV)] - \big(\mathbb{E}[aX]+\mathbb{E}[bY]\big)\big(\mathbb{E}[cW]+\mathbb{E}[dV]\big) \\
&= \mathbb{E}[aX\cdot cW] + \mathbb{E}[bY\cdot cW] + \mathbb{E}[aX\cdot dV] + \mathbb{E}[bY\cdot dV] \\
&\quad - \big(\mathbb{E}[aX]\mathbb{E}[cW] + \mathbb{E}[aX]\mathbb{E}[dV] + \mathbb{E}[bY]\mathbb{E}[cW] + \mathbb{E}[bY]\mathbb{E}[dV]\big) \\
&= \big(\mathbb{E}[aX\cdot cW] - \mathbb{E}[aX]\mathbb{E}[cW]\big) + \big(\mathbb{E}[bY\cdot cW] - \mathbb{E}[bY]\mathbb{E}[cW]\big) \\
&\quad + \big(\mathbb{E}[aX\cdot dV] - \mathbb{E}[aX]\mathbb{E}[dV]\big) + \big(\mathbb{E}[bY\cdot dV] - \mathbb{E}[bY]\mathbb{E}[dV]\big) \\
&= \big(\mathbb{E}[aX\cdot cW] - \mathbb{E}[aX]\mathbb{E}[cW]\big) + \big(\mathbb{E}[aX\cdot dV] - \mathbb{E}[aX]\mathbb{E}[dV]\big) \\
&\quad + \big(\mathbb{E}[bY\cdot cW] - \mathbb{E}[bY]\mathbb{E}[cW]\big) + \big(\mathbb{E}[bY\cdot dV] - \mathbb{E}[bY]\mathbb{E}[dV]\big) \\
&= \mathrm{Cov}(aX,cW) + \mathrm{Cov}(aX,dV) + \mathrm{Cov}(bY,cW) + \mathrm{Cov}(bY,dV) \\
&= ac\mathrm{Cov}(X,W) + ad\mathrm{Cov}(X,V) + bc\mathrm{Cov}(Y,W) + bd\mathrm{Cov}(Y,V)
\end{aligned}$$

### 1.4.6 Relationship to Variance

$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

### 1.4.7 Independence

If $X$ and $Y$ are independent, then $\mathrm{Cov}(X,Y) = 0$

**Proof** If $X$ and $Y$ are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and hence $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[X] = \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[X] = 0$

### 1.4.8 Uncorrelatedness does NOT imply independence

Two random variables $X$ and $Y$ are said to be *uncorrelated* if $\text{Cov}(X, Y) = 0$. The properties of covariance above tell us that if $X$ and $Y$ are independent, then $\text{Cov}(X, Y) = 0$ and hence $X$ and $Y$ are uncorrelated.

One immediate question that follows is whether uncorrelatedness implies independence. The simple answer is **no**. We will explore a counter-example to demonstrate that the implication does not hold.

Consider the random variables $X$ and $Y = X^2$ where $X$ is a uniform distribution in the range $[-1, 1]$. $X$ and $Y$ are both dependent on the random variable $X$ and hence not independent. Now, we need to show that $X$ and $Y$ are uncorrelated.

Since I am deferring a discussion of common probability distributions to a future notebook, I will provide tthe required information for the computations. Given constants $a, b \in \mathbb{R}$, for a continuous uniform distribution in the range $[a, b]$, $\forall x \in \mathbb{R}$ the PDF is given by

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

In addition, we use the Law of the Unconscious Statistian briefly discussed in the "Probability Basics II" notebook for the computation of $\mathbb{E}[X^3]$.

$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^{1} \frac{x}{2} dx = \frac{x^2}{4}\Bigg|_{x=-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0$

$\mathbb{E}[X] = \int_{-1}^{1} \frac{x^3}{2} dx = \frac{x^4}{8}\Bigg|_{x=-1}^{1} = \frac{1}{8} - \frac{1}{8} = 0$

Now we finally have all the tools we need to compute the covraince between $X$ and $Y$. We will use the simplified form of the covariance discussed above in order to make our calculations a tad easier.

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(X, X^2) \\ &= \mathbb{E}[X \cdot X^2] - \mathbb{E}[X]\mathbb{E}[X^2] \\ &= \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] \\ &= 0 - 0 \cdot \mathbb{E}[X^2] \\ &= 0 \end{aligned}$$

Therefore, we have shown that two random variables can be uncorrelated, but not independent.

## 1.5 Correlation

### 1.5.1 Sample Correlation

The sample correlation is called the "Pearson product-moment correlation coefficient," "Pearson correlation coefficient," or "correlation coefficient."

The measure of correlation for sample data $(x_1, y_1), \cdots, (x_n, y_n)$ for two covariates $x$ and $y$ with sample means $\bar{x}$ and $\bar{y}$ is given by

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

### 1.5.2 Population Correlation

Given two random variables $X$ and $Y$ with expected values $\mu_X$ and $\mu_Y$ and standard deviations $\sigma_X$ and $\sigma_Y$, the population correlation is given by

$$\rho_{X,Y} = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\big[(X - \mu_X)(Y - \mu_Y)\big]}{\sigma_X \sigma_Y}$$

$$\rho_{X,Y} = \mathrm{Corr}(X, Y) = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2}\sqrt{\mathbb{E}[Y^2] - \mathbb{E}[Y]^2}}$$

## 1.6 Properties of Correlation

### 1.6.1 Symmetry

1. $\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$
2. $r_{xy} = r_{yx}$

**Proof Property 1**

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

**Property 2**

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}} = r_{yx}$$

### 1.6.2 Bounds for Sample Correlation

Given real-valued data set $(x_1, y_1), \cdots, (x_n, y_n)$, $-1 \leq r_{xy} \leq 1$

**Proof** The above bound is a direct application of the Cauchy-Schwarz Inequality.

Consider the data set $(x_1, y_1), \cdots, (x_n, y_n)$ with sample mean $(\bar{x}, \bar{y})$ and standard deviation $(\sigma_x, \sigma_y)$.

Consider the vectors $\mathbf{x} = [x_1 - \bar{x}, \cdots, x_n - \bar{x}]$ and $\mathbf{y} = [y_1 - \bar{y}, \cdots, y_n - \bar{y}]$ for the covariates $X$ and $Y$, respectively.

Given $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, denote $\langle \mathbf{a}, \mathbf{b} \rangle$ denote the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$. We will use the following computations in our discussion below.

$$||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$||\mathbf{y}|| = \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle} = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

The Caucy-Schwarz Inequality states that $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^n \ |\langle \mathbf{a}, \mathbf{b} \rangle| \leq ||\mathbf{a}|| \cdot ||\mathbf{b}|| \implies -1 \leq \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{||\mathbf{a}|| \cdot ||\mathbf{b}||} \leq 1$

Plugging in our computations above, we obtain

$-1 \leq r_{xy} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}|| \cdot ||\mathbf{y}||} \leq 1$

### 1.6.3 Bounds for Population Correlation

Given real-valued random variables $X$ and $Y$ with joint distribution $(X, Y)$,

$-1 \leq \rho_{X,Y} \leq 1$

**Proof** **Lemma: Given real Given real-valued random variables $X$ and $Y$ with joint distribution $(X, Y)$, with means $(\mu_X, \mu_Y)$ and variances $(\sigma_X^2, \sigma_Y^2)$, the covariance Cov$(X, Y)$ exists.**

The above statement is a direct result of the Cauchy-Schwarz Inequality that states $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}$. Since we have clear bounds for the value of $\text{Cov}(X, Y)$, it definitely exists.

**Claim: Given real-valued random variables $X$ and $Y$ with joint distribution $(X, Y)$, with means $(\mu_X, \mu_Y)$ and variances $(\sigma_X^2, \sigma_Y^2)$ $-1 \leq \rho_{X,Y} \leq 1$**

As a side note, I really enjoy the proof below because it highlights the creativity often required to show mathematical properties in an elegant fashion.

Consider two real-valued random variables $X$ and $Y$ with joint distribution $(X, Y)$. Furthermore, $X$ has mean $\mu_X$ and variance $\sigma_X^2$, while $Y$ has mean $\mu_Y$ and variance $\sigma_Y^2$.

Now, consider the function below for some $t \in \mathbb{R}$.

$$
\begin{aligned}
g(t) = \mathrm{Var}(Xt+Y) &= \mathbb{E}\Big[\big((X-\mu_X)t+(Y-\mu_Y)\big)^2\Big] \\
&= \mathbb{E}\big[(X-\mu_X)^2t^2+2t(X-\mu_X)(Y-\mu_Y)+(Y-\mu_Y)^2\big] \\
&= \sigma_X^2t^2+2t\mathrm{Cov}(X,Y)+\sigma_Y^2
\end{aligned}
$$

In order for us to employ the non-negativity property of the variance, we first need to show that $\mathrm{Var}(Xt+Y)$ exists. Since the variances for $X$ and $Y$ exist, we know $\mathrm{Cov}(X,Y)$ exists (by our lemma above) and have an exact relation for the variance shown below.

$$\mathrm{Var}(Xt+Y) = t^2\mathrm{Var}(X)+\mathrm{Var}(Y)+2t\mathrm{Cov}(X,Y)$$

Since variance is non-negative and the above is a quadratic function with respect to $t$, we know the discrinimant must be greater than or equal to 0. In other words,

$(2\mathrm{Cov}(X,Y))^2-4\sigma_X^2\sigma_Y^2 = 4\mathrm{Cov}^2(X,Y)-4\sigma_X^2\sigma_Y^2 \geq 0$

$\implies 1 \leq \left(\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}\right)^2 = \rho_{X,Y}^2$

$\therefore -1 \leq \rho_{X,Y} \leq 1$

### 1.6.4 Discussion

The Pearson correlation coefficient plays an important part in the interpretation of linear regression, so I will defere a more detailed discussion of $r_{xy}^2$ to a future notebook.

Since the correlation between two covariates or random variables is bounded between $[-1,1]$, this can offer a measure of relatedness between two variables that may be easier to interpret than its cousin, the corvariance.

Similar to covariance, if two random variables $X$ and $Y$, part of a joint distribution $(X,Y)$ with variances $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$, are independent, then $\mathrm{Corr}(X,Y)=0$. However, if two variables are uncorrelated, we **cannot** say whether $X$ and $Y$ are independent. The mathematical discussion to show this fact is nearly identical to the case of covariance.

## 1.7 Indicator Functions

Given a subset $A$ of set $X$ (i.e. $A \subseteq X$), the indicator function $\nVdash_A$ is a mapping defined below.

$$\nVdash : X \to \{0,1\}$$

$$\nVdash_A(x) = \begin{cases} 1 & if\ x \in A \\ 0 & if\ x \notin A \end{cases}$$

### 1.7.1 Indicator Function for Intersection

$\forall A, B \in X$, $\forall \alpha \in X$, $\mathbb{1}_{A \cap B}(\alpha) = \min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha)$

**Proof** $\forall \alpha \in X, \mathbb{1}_{A \cap B}(\alpha) = \begin{cases} 1 & if\ \alpha \in A \cap B \\ 0 & if\ \alpha \notin A \cap B \end{cases}$

There are four cases to examine depending on whether a given element $\alpha \in X$ is in $A$ or $B$.

1. If $\alpha \in A$ and $\alpha \in B$, then $\alpha \in A \cap B$.

$$\mathbb{1}_{A \cap B}(\alpha) = 1$$
$$\min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \min\{1, 1\} = 1$$
$$\mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha) = 1 \cdot 1 = 1$$

2. If $\alpha \in A$ and $\alpha \notin B$, then $\alpha \notin A \cap B$

$$\mathbb{1}_{A \cap B}(\alpha) = 0$$
$$\min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \min\{1, 0\} = 0$$
$$\mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha) = 1 \cdot 0 = 0$$

3. If $\alpha \notin A$ and $\alpha \in B$, then $\alpha \notin A \cap B$

$$\mathbb{1}_{A \cap B}(\alpha) = 0$$
$$\min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \min\{0, 1\} = 0$$
$$\mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha) = 0 \cdot 1 = 0$$

4. If $\alpha \notin A$ and $\alpha \notin B$, then $\alpha \notin A \cap B$

$$\mathbb{1}_{A \cap B}(\alpha) = 0$$
$$\min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \min\{0, 0\} = 0$$
$$\mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha) = 0 \cdot 0 = 0$$

$\therefore \forall A, B \subseteq X$, $\forall \alpha \in X$ $\mathbb{1}_{A \cap B}(\alpha) = \min\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha)$

### 1.7.2 Indicator Function for Union

$\forall A, B \in X$, $\forall \alpha \in X$, $\mathbb{1}_{A \cup B}(\alpha) = \max\{\mathbb{1}_A(\alpha), \mathbb{1}_B(\alpha)\} = \mathbb{1}_A(\alpha) + \mathbb{1}_B(\alpha) - \mathbb{1}_A(\alpha) \cdot \mathbb{1}_B(\alpha)$

**Proof** $\forall \alpha \in X$, $\nVdash_{A\cup B}(\alpha) = \begin{cases} 1 & if\ \alpha \in A \cup B \\ 0 & if\ \alpha \notin A \cup B \end{cases}$ There are four cases to examine depending on whether a given element $\alpha \in X$ is in $A$ or $B$. 1. If $\alpha \in A$ and $\alpha \in B$, then $\alpha \in A \cup B$.

$$\nVdash_{A\cup B}(\alpha) = 1$$

$$\max\{\nVdash_A(\alpha), \nVdash_B(\alpha)\} = \max\{1, 1\} = 1$$

$$\nVdash_A(\alpha) + \nVdash_B(\alpha) - \nVdash_A(\alpha) \cdot \nVdash_B(\alpha) = 1 + 1 - 1 \cdot 1 = 1$$

2. If $\alpha \in A$ and $\alpha \notin B$, then $\alpha \in A \cup B$

$$\nVdash_{A\cup B}(\alpha) = 1$$

$$\max\{\nVdash_A(\alpha), \nVdash_B(\alpha)\} = \max\{1, 0\} = 1$$

$$\nVdash_A(\alpha) + \nVdash_B(\alpha) - \nVdash_A(\alpha) \cdot \nVdash_B(\alpha) = 1 + 0 - 1 \cdot 0 = 1$$

3. If $\alpha \notin A$ and $\alpha \in B$, then $\alpha \in A \cup B$

$$\nVdash_{A\cup B}(\alpha) = 1$$

$$\max\{\nVdash_A(\alpha), \nVdash_B(\alpha)\} = \max\{0, 1\} = 1$$

$$\nVdash_A(\alpha) + \nVdash_B(\alpha) - \nVdash_A(\alpha) \cdot \nVdash_B(\alpha) = 0 + 1 - 0 \cdot 1 = 1$$

4. If $\alpha \notin A$ and $\alpha \notin B$, then $\alpha \notin A \cup B$

$$\nVdash_{A\cup B}(\alpha) = 0$$

$$\max\{\nVdash_A(\alpha), \nVdash_B(\alpha)\} = \max\{0, 0\} = 0$$

$$\nVdash_A(\alpha) + \nVdash_B(\alpha) - \nVdash_A(\alpha) \cdot \nVdash_B(\alpha) = 0 + 0 - 0 \cdot 0 = 0$$

$\therefore \forall A, B \in X$, $\forall \alpha \in X$ $\nVdash_{A\cup B}(\alpha) = \max\{\nVdash_A(\alpha), \nVdash_B(\alpha)\} = \nVdash_A(\alpha) + \nVdash_B(\alpha) - \nVdash_A(\alpha) \cdot \nVdash_B(\alpha)$

### 1.7.3 Indicator Function of Complement

Given an event $A \subseteq \Omega$ and its complement $A^c = \Omega \backslash A$, $\forall \omega \in A^c$, $\nVdash_{A^c} = 1 - \nVdash_A(\omega)$

**Proof** $\forall \omega \in \Omega$, if $\omega \in A$, then by definition $\omega \notin A$.

$$\nVdash_{A^c}(\omega) = 0$$

$$1 - \nVdash_A(\omega) = 1 - 1 = 0$$

Now suppose $\omega \notin A$, which means that $\omega \in A$.

$$\nVdash_{A^c}(\omega) = 1$$

$$1 - \nVdash_A(\omega) = 1 - 0 = 1$$

### 1.7.4 Powers of Indicator Function

$\forall n \in \mathbb{R} \backslash 0$, given an event $A \subseteq \Omega$, $\forall \omega \in A$, $\left(\nVdash_A(\omega)\right)^n = \nVdash_A(\omega)$

**Proof** For some $\omega \in \Omega$, suppose $\omega \in A$.

$$\left(\nVdash_A(\omega)\right)^n = 1^n = 1 = \nVdash_A(\omega)$$

Now suppose $\omega \notin A$.

$$\left(\nVdash_A(\omega)\right)^n = 0^n = 0 = \nVdash_A(\omega)$$

### 1.7.5 Mean, Variance and Covariance of Indicator Function

For the following discussion, assume we are given a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and an event $A \in \mathcal{F}$. The indicator function $\nVdash_A$ is defined below for the probability space.

$$\nVdash_A : \Omega \to \mathbb{R}$$

$$\nVdash_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

**Mean** $\mathbb{E}[\nVdash_A] = P(A)$

**Proof** In case you decide to peruse the proof below, I want to make an important distinction between functions $p$ and $P$ in the context of probability spaces. While $P$ is a function that maps a given event $E \in \mathcal{F}$ to $[0,1]$, $p$ is a function that maps a given outcome $\omega \in \Omega$ to $[0,1]$.

**Lemma: Given an event** $A \subseteq \Omega$**,** $P(A) = \sum\limits_{\omega \in A} p(\omega)$

Given an element $\omega \in \Omega$ and an event $E_\omega = \{\omega\}$, $P(E_\omega) = p(\omega)$. $\forall \omega \in A$, we define the disjoint events $E_\omega$ (the events do not share any outcomes and are thus disjoint) and then employ the $\sigma$-additivity axiom of probability.

$$P(A) = P\left(\bigcup_{\omega \in A} E_\omega\right) = \sum_{\omega \in A} P(E_\omega) = \sum_{\omega \in A} p(\omega)$$

**Claim:** $\forall \omega \in \Omega$**,** $\mathbb{E}[\nVdash_A] = P(A)$

We will use the fact that $\forall \omega \in \Omega$, $\omega \in A$ or $\omega \notin A$ in the proof below.

$$
\begin{aligned}
\mathbb{E}[\mathbb{1}_A] &= \sum_{\omega\in\Omega} p(\omega)\mathbb{1}_A(\omega) \\
&= \sum_{\omega_i\in A} p(\omega_i)\mathbb{1}_A(\omega_i) + \sum_{\omega_j\notin A} p(\omega_j)\mathbb{1}_A(\omega_j) \\
&= \sum_{\omega_i\in A} p(\omega_i)\cdot 1 + \sum_{\omega_i\notin A} p(\omega_i)\cdot 0 \\
&= \sum_{\omega_i\in A} p(\omega_i) \\
&= P(A)
\end{aligned}
$$

**Variance** $\mathrm{Var}(\mathbb{1}_A) = P(A)(1-P(A))$

**Proof**

$$
\begin{aligned}
\mathrm{Var}(\mathbb{1}_A) &= \mathbb{E}\left[\big(\mathbb{1}_A - \mathbb{E}[\mathbb{1}_A]\big)^2\right] \\
&= \mathbb{E}\left[\big(\mathbb{1}_A - P(A)\big)^2\right] \\
&= \mathbb{E}\big[\mathbb{1}_A^2 - 2P(A)\mathbb{1}_A + P(A)^2\big] \\
&= \mathbb{E}\big[\mathbb{1}_A^2\big] - \mathbb{E}\big[2P(A)\mathbb{1}_A\big] + \mathbb{E}\big[P(A)^2\big] \\
&= \mathbb{E}[\mathbb{1}_A] - 2P(A)\mathbb{E}[\mathbb{1}_A] + P(A)^2 \\
&= P(A) - 2P(A)^2 + P(A)^2 = P(A) - P(A)^2 = P(A)(1-P(A))
\end{aligned}
$$

**Covariance** Given two events $A, B \subseteq \Omega$, $\mathrm{Cov}\big(1_A, \mathbb{1}_B\big) = P(A\cap B) - P(A)P(B)$

**Proof**

$$
\begin{aligned}
\mathrm{Cov}\big(\mathbb{1}_A, \mathbb{1}_B\big) &= \mathbb{E}\Big[\big(\mathbb{1}_A - \mathbb{E}[\mathbb{1}_A]\big)\big(\mathbb{1}_B - \mathbb{E}[\mathbb{1}_B]\big)\Big] \\
&= \mathbb{E}\Big[\big(\mathbb{1}_A - P(A)\big)\big(\mathbb{1}_B - P(B)\big)\Big] \\
&= \mathbb{E}\Big[\mathbb{1}_A\cdot\mathbb{1}_B - \mathbb{1}_A P(B) - P(A)\mathbb{1}_B + P(A)P(B)\Big] \\
&= \mathbb{E}\big[\mathbb{1}_A\cdot\mathbb{1}_B\big] - \mathbb{E}\big[\mathbb{1}_A P(B)\big] - \mathbb{E}\big[P(A)\mathbb{1}_B\big] + \mathbb{E}\big[P(A)P(B)\big] \\
&= \mathbb{E}\big[\mathbb{1}_{A\cap B}\big] - P(B)\mathbb{E}\big[\mathbb{1}_A\big] - P(A)\mathbb{E}\big[\mathbb{1}_B\big] + P(A)P(B) \\
&= P(A\cap B) - P(A)P(B) - P(A)P(B) + P(A)P(B) \\
&= P(A\cap B) - P(A)P(B)
\end{aligned}
$$