MEAN VALUE (EXPECTED VALUE, EXPECTATION)

Definition. Let $X : \Omega \to \mathbb{R}$ be a random variable (r.v.) with distribution function $F(x)$. The **mean value** (or **expected value** or **expectation**) of X is the quantity

$$\mu = E[X] = \int_{-\infty}^{\infty} x \, dF(x)$$

(the integration is in the Riemann-Stieltjes sense) provided the integral is absolutely convergent. If

$$\int_{-\infty}^{\infty} |x| \, dF(x) = \infty,$$

then $E[X]$ does not exist.
The quantity $\mu_n = E[X^n]$ is called the **moment of order** n or the n**-th moment** of X.

Remarks. (i) If X is a discrete r.v. with values $x_1, x_2, x_3, \ldots$ and (mass) probability function $p(x_j)$, then

$$\mu = E[X] = \sum_{x_j} x_j \, p(x_j),$$

provided the sum is absolutely convergent (the sum extends over all values x_j taken by X). If

$$\sum_{x_j} |x_j| \, p(x_j) = \infty,$$

then $E[X]$ does not exist.
(ii) If X is a continuous r.v. with probability density function $f(x)$, then

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx,$$

provided the integral is absolutely convergent. If

$$\int_{-\infty}^{\infty} |x| f(x) dx = \infty,$$

then $E[X]$ does not exist.
(iii) If the graph of $f(x)$ (or $p(x_j)$) has an axis of symmetry $x = x_0$ and $\mu = E[X]$ exists, then $\mu = x_0$.

$\star\star\star$ Properties of the Mean Value:

All properties below hold as long as the mean values exist.

0. If c is a constant, then $E[c] = c$.

1. If $X \leq Y$, then $E[X] \leq E[Y]$.

2. (a) Let $g(x)$ be a real-valued function defined on $\mathbb{R}$.
If X is a discrete r.v. with values $x_1, x_2, x_3, \dots$ and (mass) probability function $p(x_j)$, then

$$E[g(X)] = \sum_{x_j} g(x_j)p(x_j),$$

where the sum extends over all values x_j taken by X.
If X is a continuous r.v. with probability density function $f(x)$, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

(b) Let $g(x, y)$ be a real-valued function defined on $\mathbb{R}^2$.
If X and Y are discrete r.v.'s with values $x_1, x_2, x_3, \dots$ and $y_1, y_2, y_3, \dots$ respectively and joint (mass) probability function $p(x_j, y_k)$, then

$$E[g(X, Y)] = \sum_{y_k} \sum_{x_j} g(x_j, y_k)p(x_j, y_k),$$

where the sums extend over all values x_j of X and all values y_k of Y.
If X and Y are continuous r.v.'s with joint probability density function $f(x, y)$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy,$$

3 (Linearity of the Mean Value)**.** If X, Y are r.v.'s and a, b are constants, then

$$E[aX + bY] = aE[X] + bE[Y]$$

(the property tells us that the mean value respects linear operations).

4. If X, Y are **independent** r.v.'s, then

$$E[XY] = E[X]E[Y].$$

However, if $E[XY] = E[X]E[Y]$, we **cannot** conclude that X, Y are independent.

Let us justify Property 4 for the case where X and Y are continuous r.v.'s with a joint probability density function $f(x, y)$ (the discrete case can be also justified by the appropriate adaptation of the argument).
Since X, Y are independent, we must have $f(x, y) = f_X(x)f_Y(y)$. Thus (by using Property 2(b) with $g(x, y) = xy$)

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = E[X]E[Y]. \end{aligned}$$

VARIANCE

Definition. Let $X : \Omega \to \mathbb{R}$ be a r.v. such that $\mu = E[X]$ exists. The **variance** of X is the quantity

$$\sigma^2 = V[X] = E\left[(X-\mu)^2\right]$$

(it may be ∞). Also, the quantity $\sigma = \sqrt{V[X]}$ is called the **standard deviation** of X.

The variance is a measure of the "spreadness" of the values of X.

$\star\star\star$ **Properties of the Variance:**

0. By its definition we have that $V[X] \geq 0$. If $V[X] = 0$, then $X = \mu$ a.s. (i.e. with probability 1), where $\mu = E[X]$.

1. $V[X] = E[X^2] - E[X]^2$.
This property follows by setting $X = Y$ in Property 2 of the covariance, given below.
An implication of Property 1 is that we always have

$$E[X^2] \geq E[X]^2$$

and equality happens if and only if $V[X] = 0$.

2. If a, b are constants, then

$$V[aX + b] = a^2 V[X].$$

Proof. Let $\mu = E[X]$. Then $E[aX + b] = a\mu + b$, hence the definition of the variance gives

$$V[aX+b] = E\left[\{(aX + b - (a\mu + b)\}^2\right] = E\left[a^2(X-\mu)^2\right] = a^2 E\left[(X-\mu)^2\right]$$

and the proof is finished since $E\left[(X-\mu)^2\right] = V[X]$ by definition. ■

3 (A "Pythagorean Theorem")**.** If X, Y are **independent** r.v.'s, then

$$V[X + Y] = V[X] + V[Y].$$

However, if $V[X + Y] = V[X] + V[Y]$, we **cannot** conclude that X, Y are independent.
This property follows from Property 3 of the covariance, given below.

COVARIANCE

Definition. Let $X, Y : \Omega \to \mathbb{R}$ be r.v.'s, such that the mean values $\mu_X = E[X]$ and $\mu_Y = E[Y]$ exist. The **covariance** of X and Y is the quantity

$$\mathrm{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right].$$

The dimensionless quantity

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{V[X]V[Y]}}$$

is called the **correlation coefficient** of X and Y (it is defined as long as $V[X], V[Y]$ are finite and $\neq 0$). If $\mathrm{Cov}(X, Y) = 0$, we say that X and Y are **uncorrelated**.

If $\mathrm{Cov}(X, Y) > 0$, then Y has a tendency to increase when X increases (e.g., $X =$ height and $Y =$ weight of a person), while if $\mathrm{Cov}(X, Y) < 0$, then Y has a tendency to decrease when X increases. The equation $\mathrm{Cov}(X, Y) = 0$ gives us a first indication of independence of X and Y (see Property 2 below).

Remark. Notice that

$$\mathrm{Cov}(X, X) = V[X].$$

$\star\star\star$ **Properties of the Covariance:**

0 (Symmetry)**.**

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

(this property follows immediately from the definition of the covariance).

1 (Linearity in each argument)**.** If X, X_1, X_2, Y, Y_1, Y_2 are r.v.'s and a_1, a_2, b_1, b_2 are constants, then

$$\mathrm{Cov}(a_1 X_1 + a_2 X_2, Y) = a_1 \mathrm{Cov}(X_1, Y) + a_2 \mathrm{Cov}(X_2, Y),$$

$$\mathrm{Cov}(X, b_1 Y_1 + b_2 Y_2) = b_1 \mathrm{Cov}(X, Y_1) + b_2 \mathrm{Cov}(X, Y_2)$$

(this property follows easily from the definition of the covariance and the linearity of the mean value).

2. $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$
(hence, by Property 4 of the mean value we have that if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$).

Proof. Using the linearity of the mean value we get

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E\left[(X - \mu_X)(Y - \mu_Y)\right] = E\left[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y\right] \\ &= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y\end{aligned}$$

and the property follows since $\mu_X = E[X]$ and $\mu_Y = E[Y]$. ■

3.

$$V[X+Y] = V[X] + V[Y] + 2\,\mathrm{Cov}(X,Y).$$

Proof. Using Property 1 of the variance and the linearity of the mean value we get

$$\begin{aligned} V[X+Y] &= E\left[(X+Y)^2\right] - E[X+Y]^2 = E\left[(X+Y)^2\right] - (E[X]+E[Y])^2 \\ &= E[X^2+Y^2+2XY] - \big(E[X]^2+E[Y]^2+2E[X]E[Y]\big) \\ &= E[X^2]+E[Y^2]+2E[XY]-E[X]^2-E[Y]^2-2E[X]E[Y] \\ &= E[X^2]-E[X]^2+E[Y^2]-E[Y]^2+2\,(E[XY]-E[X]E[Y]) \\ &= V[X]+V[X]+2\,(E[XY]-E[X]E[Y]) \end{aligned}$$

and the property follows from the previous property of the covariance, namely that $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$. ■

4 (Schwarz Inequality)**.**

$$\mathrm{Cov}(X,Y)^2 \le V[X]\,V[Y].$$

Equality happens only when (i) there are constants a and b such that $Y = aX+b$ with probability 1, or (ii) $V[X]=0$.

Proof. If $V[X]=0$, then $X=\mu_X$ with probability 1, hence $\mathrm{Cov}(X,Y)=0$ and our inequality becomes equality. To continue, let us assume $V[X]\neq 0$. For $\lambda\in\mathbb{R}$ we set

$$p(\lambda) = V[\lambda X+Y] = V[\lambda X]+V[Y]+2\,\mathrm{Cov}(\lambda X,Y) = \lambda^2 V[X]+2\lambda\mathrm{Cov}(X,Y)+V[Y].$$

Thus $p(\lambda)$ is a quadratic polynomial satisfying $p(\lambda)\ge 0$, since it is the variance of $\lambda X+Y$. It follows that its discriminant $\Delta = 4\,\mathrm{Cov}(X,Y)^2 - 4\,V[X]\,V[Y]$ must be nonpositive, equivalently

$$\mathrm{Cov}(X,Y)^2 - V[X]\,V[Y] \le 0,$$

which is the inequality we wanted to prove. Now, if we have equality, i.e. $\Delta=0$, then $p(\lambda)$ must have a double real zero, namely

$$p(\lambda) = V[X](\lambda-\lambda_0)^2.$$

Thus, $p(\lambda_0)=0$, i.e. $V[\lambda_0 X+Y]=0$, which implies that there is a constant b such that $Y+\lambda_0 X = b$ with probability 1. ■

Remark. Proposition 4 above implies that the correlation coefficient satisfies

$$-1 \le \rho(X,Y) \le 1.$$

In the extreme case $\rho(X,Y)=1$ we must have $Y=aX+b$ with probability 1, for some constants $a>0$ and b, while if $\rho(X,Y)=-1$, then $Y=aX+b$ with probability 1, for some constants $a<0$ and b.