

CSE 107

Probability and Statistics for Engineers Variance, Covariance and Correlation

Here we state and prove a few facts regarding the variance, covariance and correlation operators. Let X, Y, Z, U and V be random variables with finite second moments. We will occasionally use the notations $\mu_X = E[X]$, $\sigma_X^2 = \text{Var}(X)$ and $\sigma_X = \sqrt{\text{Var}(X)}$ for the mean, variance and standard deviation of X , respectively.

- (1) If $X \geq 0$, then $E[X] \geq 0$. More generally $X \geq c$ implies $E[X] \geq c$, for any $c \in \mathbb{R}$.
- (2) $\text{Var}(X) \geq 0$.
- (3) If $\text{Var}(X) = 0$, then $P(X = \mu_X) = 1$.
- (4) $\text{Cov}(X, X) = \text{Var}(X)$.
- (5) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (6) $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$.
- (7) $\text{Var}(aX + b) = a^2\text{Var}(X)$.
- (8) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.
- (9) $\text{Cov}(aX + bY, cU + dV) = ac\text{Cov}(X, U) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, U) + bd\text{Cov}(Y, V)$.
- (10) $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$.
- (11) $-1 \leq \rho(X, Y) \leq 1$.

Proof of (1):

If X is discrete, then $E[X] = \sum_x xp_X(x) \geq 0$, since all terms are non-negative. If X is continuous, then $E[X] = \int_0^\infty xf_X(x)dx \geq 0$, since the integrand is non-negative. Also

$$X \geq c \Rightarrow X - c \geq 0 \Rightarrow E[X - c] \geq 0 \Rightarrow E[X] - c \geq 0 \Rightarrow E[X] \geq c$$

as claimed. ■

Proof of (2):

Let $U = (X - \mu_X)^2$. Then $U \geq 0$, whence $\text{Var}(X) = E[U] \geq 0$, by (1). ■

Proof of (3):

Again let $U = (X - \mu_X)^2$, so that $E[U] = \text{Var}(X) = 0$ by hypothesis. Let n be a positive integer and define the event $A_n = \{U \geq 1/n\}$. By the conditional expectation theorem

$$E[U] = E[U|A_n]P(A_n) + E[U|A_n^c]P(A_n^c)$$

The definition of A_n and (1) give us $E[U|A_n] \geq 1/n$. Also $E[U|A_n^c] \geq 0$ since $U \geq 0$. Thus

$$0 = E[U] \geq \frac{1}{n} \cdot P(A_n)$$

which implies $P(A_n) = 0$. Now observe $\{U > 0\} = \bigcup_{n=1}^\infty \{U > 1/n\}$, whence

$$P(U > 0) = P\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty P(A_n) = 0$$

showing that $P(U > 0) = 0$. Therefore $P(U = 0) = 1$, and it follows that

$$\begin{aligned} P(X = \mu_X) &= P(X - \mu_X = 0) \\ &= P((X - \mu_X)^2 = 0) \\ &= P(U = 0) \\ &= 1 \end{aligned}$$

which completes the proof. ■

Note: In the discrete case, $P(X = \mu_X) = 1$ simply means that $X = \mu_X$ is constant, and not actually random. In the continuous case, there may be some $\omega \in \Omega$ for which $X(\omega) \neq \mu_X$, but the event $\{\omega \mid X(\omega) \neq \mu_X\}$ has probability zero. This is usually expressed by saying $X = \mu_X$ *almost surely*, or that X is *essentially constant*.

Proof of (4):

$$\text{Cov}(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2] = \text{Var}(X). \quad \blacksquare$$

Proof of (5):

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[(Y - \mu_Y)(X - \mu_X)] = \text{Cov}(Y, X). \quad \blacksquare$$

Proof of (6):

$$\begin{aligned} \text{Cov}(aX + b, Y) &= E[((aX + b) - (a\mu_X + b))(Y - \mu_Y)] \\ &= E[a(X - \mu_X)(Y - \mu_Y)] \\ &= aE[(X - \mu_X)(Y - \mu_Y)] \\ &= a\text{Cov}(X, Y). \end{aligned} \quad \blacksquare$$

Proof of (7):

$$\begin{aligned} \text{Var}(aX + b) &= \text{Cov}(aX + b, aX + b) && \text{by (4)} \\ &= a\text{Cov}(X, aX + b) && \text{by (6)} \\ &= a\text{Cov}(aX + b, X) && \text{by (5)} \\ &= a^2\text{Cov}(X, X) && \text{by (6)} \\ &= a^2\text{Var}(X) && \text{by (4)}. \end{aligned} \quad \blacksquare$$

Proof of (8):

$$\begin{aligned} \text{Cov}(X + Y, Z) &= E[((X + Y) - (\mu_X + \mu_Y))(Z - \mu_Z)] \\ &= E[(X - \mu_X + Y - \mu_Y)(Z - \mu_Z)] \\ &= E[(X - \mu_X)(Z - \mu_Z) + (Y - \mu_Y)(Z - \mu_Z)] \\ &= E[(X - \mu_X)(Z - \mu_Z)] + E[(Y - \mu_Y)(Z - \mu_Z)] \\ &= \text{Cov}(X, Z) + \text{Cov}(Y, Z). \end{aligned} \quad \blacksquare$$

Proof of (9):

$$\begin{aligned}
\text{Cov}(aX + bY, cU + dV) &= \text{Cov}(aX, cU + dV) + \text{Cov}(bY, cU + dV) && \text{by (8)} \\
&= a\text{Cov}(X, cU + dV) + b\text{Cov}(Y, cU + dV) && \text{by (6)} \\
&= a\text{Cov}(cU + dV, X) + b\text{Cov}(cU + dV, Y) && \text{by (5)} \\
&= a(\text{Cov}(cU, X) + \text{Cov}(dV, X)) + b(\text{Cov}(cU, Y) + \text{Cov}(dV, Y)) && \text{by (8)} \\
&= a(c\text{Cov}(U, X) + d\text{Cov}(V, X)) + b(c\text{Cov}(U, Y) + d\text{Cov}(V, Y)) && \text{by (6)} \\
&= a(c\text{Cov}(X, U) + d\text{Cov}(X, V)) + b(c\text{Cov}(Y, U) + d\text{Cov}(Y, V)) && \text{by (5)} \\
&= ac\text{Cov}(X, U) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, U) + bd\text{Cov}(Y, V).
\end{aligned}$$

■

Proof of (10):

$$\begin{aligned}
\text{Var}(aX + bY) &= \text{Cov}(aX + bY, aX + bY) && \text{by (4)} \\
&= a^2\text{Cov}(X, X) + ab\text{Cov}(X, Y) + ab\text{Cov}(Y, X) + b^2\text{Cov}(Y, Y) && \text{by (9)} \\
&= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y). && \text{by (4) and (5)}
\end{aligned}$$

■

Proof of (11):

The definition of correlation is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Let $a = 1$ and $b = \pm 1$ in (10) to get $\text{Var}(U \pm V) = \text{Var}(U) + \text{Var}(V) \pm 2\text{Cov}(U, V)$. Thus

$$\begin{aligned}
0 &\leq \text{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) && \text{by (2)} \\
&= \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) \pm 2\text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) && \text{letting } U = \frac{X}{\sigma_X} \text{ and } V = \frac{Y}{\sigma_Y} \text{ above} \\
&= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} \pm 2\frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} && \text{by (6) and (7)} \\
&= 1 + 1 \pm 2\rho(X, Y) \\
&= 2(1 \pm \rho(X, Y))
\end{aligned}$$

This gives $0 \leq 1 + \rho(X, Y) \Rightarrow -1 \leq \rho(X, Y)$ and $0 \leq 1 - \rho(X, Y) \Rightarrow \rho(X, Y) \leq 1$. Therefore

$$-1 \leq \rho(X, Y) \leq 1$$

completing the proof. ■

Alternate Proof of (11):

Let $t \in \mathbb{R}$ be chosen arbitrarily. Then (2) and (10) give

$$0 \leq \text{Var}(tU + V) = t^2\text{Var}(U) + 2t\text{Cov}(U, V) + \text{Var}(V)$$

Thus the quadratic polynomial $at^2 + bt + c$, where $a = \text{Var}(U)$, $b = 2\text{Cov}(U, V)$ and $c = \text{Var}(V)$, is never negative, hence has either one real root or two complex roots. Its discriminant is thus non-positive:

$$\begin{aligned} b^2 - 4ac &\leq 0 \Rightarrow b^2 \leq 4ac \\ &\Rightarrow 4\text{Cov}(U, V)^2 \leq 4\text{Var}(U)\text{Var}(V) \\ &\Rightarrow \text{Cov}(U, V)^2 \leq \text{Var}(U)\text{Var}(V) \end{aligned}$$

Now let $U = X/\sigma_X$ and $V = Y/\sigma_Y$. Then by (7)

$$\text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)^2 \leq \text{Var}\left(\frac{X}{\sigma_X}\right)\text{Var}\left(\frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} \cdot \frac{\text{Var}(Y)}{\sigma_Y^2} = 1 \cdot 1 = 1,$$

and by (6)

$$\left(\frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}\right)^2 \leq 1 \Rightarrow (\rho(X, Y))^2 \leq 1 \Rightarrow |\rho(X, Y)| \leq 1,$$

so that $-1 \leq \rho(X, Y) \leq 1$, as claimed. ■