What does it mean when we say “a covariance matrix is **bigger** than another”?

jessekelighine.com

December 12, 2023

Consider two covariance matrices $A_{n\times n}$ and $B_{n\times n}$. We say that A is **bigger** than B, often denoted by $A \geq B$ or $A \succeq B$, if $A - B$ is semi-positive definite. Why do we use the “definiteness” of a matrix to compare the size of two covariance matrices?

First, notice that a covariance matrix is not only symmetrical, but also semi-positive definite. Consider a random vector $X = (X_1, ..., X_n)^\intercal$. The covariance matrix is defined by

$$K \coloneqq \mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])^\intercal].$$

Given any constant vector $\mathbf{v}$ of length n, we have

$$\mathbf{v}^\intercal K\mathbf{v} = \mathbf{E}[\mathbf{v}^\intercal (X - \mathbf{E}[X])(\mathbf{v}^\intercal (X - \mathbf{E}[X]))^\intercal] \geq 0$$

by the definition of K. Therefore, the covariance matrix K is semi-positive definite. (In fact, $\mathbf{v}^\intercal K\mathbf{v}$ is zero iff X has no variance at all.)

There is another intuitive way of interpreting the definiteness described above. Consider the same vector $\mathbf{v}$ and the random vector X. The dot product $\mathbf{v}^\intercal X$ is the projection of the random vector from n-dimensional space on a one-dimensional space along the direction of $\mathbf{v}$, i.e., this collapse the n-dimensional random variable to a one-dimensional random variable through some linear combination. If we calculate the variance of the one-dimensional random variable $\mathbf{v}^\intercal X$, we obtain

$$\begin{aligned}\mathrm{Var}[\mathbf{v}^\intercal X] &= \mathbf{E}[\mathbf{v}^\intercal X(\mathbf{v}^\intercal X)^\intercal] - \mathbf{E}[\mathbf{v}^\intercal X]\,\mathbf{E}[\mathbf{v}^\intercal X]^\intercal \\ &= \mathbf{v}^\intercal \big(\mathbf{E}[XX^\intercal] - \mathbf{E}[X]\,\mathbf{E}[X]^\intercal\big)\mathbf{v} \\ &= \mathbf{v}^\intercal K\mathbf{v}.\end{aligned}$$

Notice that the variance assumes the exact form as before. And since variance is non-negative, it is clear that the covariance matrix must be semi-positive definite. That is, for any direction $\mathbf{v}$, the variance of “X projected on that direction” is (clearly) non-negative.

Motivated by the intuitive interpretation, let $X = (X_1, ..., X_n)^\intercal$ and $Y = (Y_1, ..., Y_n)^\intercal$ be random vectors with mean $(0, ..., 0)^\intercal$ for simplicity. Let $A = \mathbf{E}[XX^\intercal]$ and $B = \mathbf{E}[YY^\intercal]$ be the covariance matrices. Our goal is to compare A and B in some meaningful way. Since covariance matrices are multi-dimensional, it is not very straight forward. However, we can project X and Y on a vector $\mathbf{v}$, an then compare the variance (non-negative real number) of the two projections. To make the comparison meaningful, it is reasonable to compare *all* possible projections, i.e., consider all possible choices of $\mathbf{v}$.

Formally, consider any vector $\mathbf{v}$. The projection of X on $\mathbf{v}$ is $\mathbf{v}^\intercal X$. The variance of $\mathbf{v}^\intercal X$ is

$$\begin{aligned}\mathbf{E}[(\mathbf{v}^\intercal X)^2] &= \mathbf{E}[\mathbf{v}^\intercal XX^\intercal \mathbf{v}] \\ &= \mathbf{v}^\intercal\,\mathbf{E}[XX^\intercal]\mathbf{v} = \mathbf{v}^\intercal A\mathbf{v}\end{aligned}$$

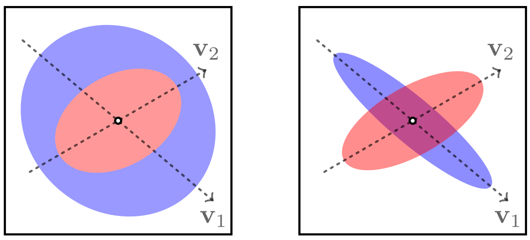
where A is the covariance matrix. Similarly, consider the same for Y. If we find that $\forall\mathbf{v}$,

$$\mathbf{v}^\intercal A\mathbf{v} - \mathbf{v}^\intercal B\mathbf{v} = \mathbf{v}^\intercal (A - B)\mathbf{v} \geq 0,$$

then, by definition, $A - B$ is semi-positive definite. Therefore, we define $A \geq B$ if $A - B$ is semi-positive definite. The interpretation is that “the variance of X is larger than Y in *all directions*”. (This is called the Löwner partial ordering.)

This interpretation of the partial ordering can be understood easily through visualisation. The following are representations of the distributions X and Y where the two random vectors are two-dimensional:

blue distribution: X, covariance matrix: A
red distribution: Y, covariance matrix: B



case 1 case 2

Let X with covariance matrix A be the blue distribution and Y with covariance matrix B be the red distribution. It is clear that in case 1, A is **bigger** than B since the variance of X is bigger that Y’s in *every* direction. (every possible direction of projection) However, the same statement is not true in case 2. In some directions (e.g. $\mathbf{v}_1$), the variance of X is larger; in other directions (e.g. $\mathbf{v}_2$), the variance of Y is larger. Thus, A and B are not comparable by the partial order in case 2. ■