How to compare the size of covariance matrices?

2025-05-14
jessekelighine.com
Jesse C. Chen 陳捷

Consider two covariance matrices $\mathbf{A}_{n\times n}$ and $\mathbf{B}_{n\times n}$. We say that $\mathbf{A}$ is *larger* than $\mathbf{B}$, denoted by $\mathbf{A} \geq \mathbf{B}$ or $\mathbf{A} \succsim \mathbf{B}$, if $\mathbf{A}-\mathbf{B}$ is semi-positive definite. But why do we use matrix definiteness to compare the "size" of covariance matrices?

To understand this, recall that a covariance matrix is not only symmetric, but also positive semi-definite. Let $\boldsymbol{x} = (x_1, ..., x_n)'$ be a random vector. Its covariance matrix is given by:

$$\mathbf{K} \coloneqq \mathbf{E}\,(\boldsymbol{x} - \mathbf{E}\,\boldsymbol{x})(\boldsymbol{x} - \mathbf{E}\,\boldsymbol{x})'.$$

Now consider any constant vector $\mathbf{v} \in \mathbb{R}^n$. We can examine the quadratic form:

$$\mathbf{v}'\mathbf{K}\mathbf{v} = \mathbf{E}(\mathbf{v}'(\boldsymbol{x} - \mathbf{E}\,\boldsymbol{x})(\mathbf{v}'(\boldsymbol{x} - \mathbf{E}\,\boldsymbol{x}))') \geq 0$$

Therefore, the covariance matrix $\mathbf{K}$ is always positive semi-definite.

There is another intuitive way of understanding the positive semi-definiteness. Consider the same vector $\mathbf{v}$ and the random vector $\boldsymbol{x}$. The dot product $\boldsymbol{y} = \mathbf{v}'\boldsymbol{x}$ is a projection of the random vector from n-dimensional space on a one-dimensional space along the direction of $\mathbf{v}$. The variance of $\boldsymbol{y}$ is given by

$$\begin{aligned}\operatorname{Var}(\boldsymbol{y}) &= \mathbf{E}((\mathbf{v}'\boldsymbol{x})(\mathbf{v}'\boldsymbol{x})') - \mathbf{E}(\mathbf{v}'\boldsymbol{x})\,\mathbf{E}(\mathbf{v}'\boldsymbol{x})' \\ &= \mathbf{v}'\big(\mathbf{E}(\boldsymbol{x}\boldsymbol{x}') - \mathbf{E}(\boldsymbol{x})\,\mathbf{E}(\boldsymbol{x})'\big)\mathbf{v} \\ &= \mathbf{v}'\mathbf{K}\mathbf{v}.\end{aligned}$$

Notice that the variance of $\boldsymbol{y}$ assumes the exact form as before. Since variance is always non-negative, the covariance matrix must be positive semi-definite. Motivated by this intuition, we can now ask: how should we compare two covariance matrices?

With the two interpretations of covariance matrices in mind, let's now compare two covariance matrices. Let $\boldsymbol{x} = (x_1, ..., x_n)'$ and $\boldsymbol{y} = (y_1, ..., y_n)'$ be random vectors, both with mean $(0, ..., 0)'$ for simplicity. Let $\mathbf{A} = \mathbf{E}(\boldsymbol{x}\boldsymbol{x}')$ and $\mathbf{B} = \mathbf{E}(\boldsymbol{y}\boldsymbol{y}')$ denote their respective covariance matrices. Our goal is to compare $\mathbf{A}$ and $\mathbf{B}$ in some meaningful way.

A natural thought is to project both $\boldsymbol{x}$ and $\boldsymbol{y}$ onto an arbitrary direction $\mathbf{v} \in \mathbb{R}^n$, then compare the variances of these one-dimensional projections. Since each projection yields a non-negative scalar variance, we can compare these scalar values across *all* possible directions $\mathbf{v}$.

Formally, consider any vector $\mathbf{v}$, consider the projection $\mathbf{v}'\boldsymbol{x}$. Its variance is

$$\begin{aligned}\mathbf{E}((\mathbf{v}'\boldsymbol{x})^2) &= \mathbf{E}(\mathbf{v}'\boldsymbol{x}\boldsymbol{x}'\mathbf{v}) \\ &= \mathbf{v}'\,\mathbf{E}(\boldsymbol{x}\boldsymbol{x}')\mathbf{v} = \mathbf{v}'\mathbf{A}\mathbf{v}\end{aligned}$$

Similarly, for $\boldsymbol{y}$ we have

$$\mathbf{E}((\mathbf{v}'\boldsymbol{y})^2) = \mathbf{v}'\mathbf{B}\mathbf{v}.$$

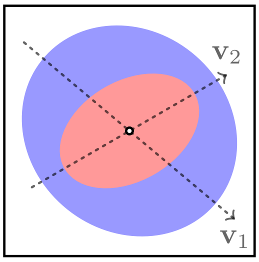
If for all $\mathbf{v}$, we find:

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

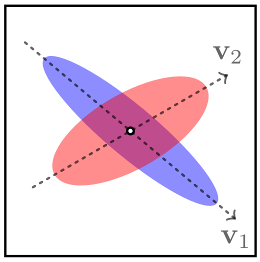
then the matrix $\mathbf{A}-\mathbf{B}$, by definition, is positive semi-definite. In this case, we say that $\mathbf{A}$ is *larger* than $\mathbf{B}$ in the Loewner partial ordering. That is:

> If $\mathbf{A}-\mathbf{B}$ is positive semi-definite, then *for all possible directions* $\mathbf{v}$, the variance of the projection of $\boldsymbol{x}$ exceeds or equals that of $\boldsymbol{y}$. This defines the Loewner ordering.

This interpretation is intuitive when visualized geometrically. Imagine two random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^2$, whose distributions are represented by ellipses:



case 1



case 2

Let the blue ellipse represent the distribution $\boldsymbol{x}$ with covariance matrix $\mathbf{A}$, and let the red ellipse represent the distribution $\boldsymbol{y}$ with covariance matrix $\mathbf{B}$.

In case 1, $\mathbf{A}$ is *larger* than $\mathbf{B}$ since the blue ellipse fully encloses the red one, indicating that in *every direction*, the variance of $\boldsymbol{x}$ exceeds that of $\boldsymbol{y}$. However, the same statement is not true in case 2. In some directions, e.g. along $\mathbf{v}_1$, the variance of $\boldsymbol{x}$ is larger; in other directions, e.g. along $\mathbf{v}_2$, the variance of $\boldsymbol{y}$ is larger. Thus, $\mathbf{A}$ and $\mathbf{B}$ are not comparable by Loewner ordering, meaning that it does not make sense to say that one covariance matrix is larger than the other. ■