Random vectors

It will be extremely helpful to us if we worked directly with random vectors and not a group of individual random variables. Throughout, all vectors are written columnwise; and so are random ones. Thus, for instance, a random vector $\boldsymbol{X} \in \mathbf{R}^n$ is written columnwise as

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = (X_1\,, \ldots, X_n)'.$$

And even more generally, we might sometimes be interested in random matrices. For instance, a random $m \times n$ matrix is written coordinatewise as

$$\boldsymbol{X} = \begin{pmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{pmatrix}.$$

1. Expectation

If $\boldsymbol{X}$ is a random $m \times n$ matrix, then we define its *expectation* in the most natural possible way as

$$\mathrm{E}\boldsymbol{X} := \begin{pmatrix} \mathrm{E}X_{1,1} & \cdots & \mathrm{E}X_{1,n} \\ \vdots & & \vdots \\ \mathrm{E}X_{m,1} & \cdots & \mathrm{E}X_{m,n} \end{pmatrix}.$$

Many of the properties of expectations continue to hold in the setting of random vectors and/or matrices. The following summarizes some of those properties.

Proposition 1. *Suppose $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are nonrandom matrices, and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random matrices. Then,*

$$\mathrm{E}\,(\boldsymbol{AXB} + \boldsymbol{CYD}) = \boldsymbol{A}\,(\mathrm{E}\boldsymbol{X})\,\boldsymbol{B} + \boldsymbol{C}\,(\mathrm{E}\boldsymbol{Y})\,\boldsymbol{D},$$

provided that the matrix dimensions are sensible.

Proof. Because

$$(\boldsymbol{AXB} + \boldsymbol{CYD})_{i,j} = (\boldsymbol{AXB})_{i,j} + (\boldsymbol{CYD})_{i,j}\,,$$

it suffices to prove that $\mathrm{E}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$. But then, we can work coordinatewise as follows:

$$\begin{aligned}\mathrm{E}\left[(\boldsymbol{AXB})_{i,j}\right] &= \mathrm{E}\sum_{k=1}^{n}\sum_{\ell=1}^{n} A_{i,k}X_{k,\ell}B_{\ell,j} = \sum_{k=1}^{n}\sum_{\ell=1}^{n} A_{i,k}\mathrm{E}(X_{k,\ell})B_{\ell,j}\\ &= \sum_{k=1}^{n}\sum_{\ell=1}^{n} A_{i,k}\left[\mathrm{E}\boldsymbol{X}\right]_{k,\ell} B_{\ell,j} = \left[\boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}\right]_{i,j}.\end{aligned}$$

That is, $\mathrm{E}(\boldsymbol{AXB}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}$ coordinatewise. This proves the result. □

2. Covariance

Suppose $\boldsymbol{X} = (X_1\,,\ldots,X_m)'$ and $\boldsymbol{Y} = (Y_1\,,\ldots,Y_n)'$ are two jointly distributed random vectors. We define their *covariance* as

$$\mathrm{Cov}(\boldsymbol{X}\,,\boldsymbol{Y}) := \begin{pmatrix}\mathrm{Cov}(X_1\,,Y_1) & \cdots & \mathrm{Cov}(X_1\,,Y_n)\\ \vdots & & \vdots\\ \mathrm{Cov}(X_m\,,Y_1) & \cdots & \mathrm{Cov}(X_m\,,Y_n)\end{pmatrix}.$$

Proposition 2. *We always have*

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

Proof. The $(i\,,j)$th entry of the matrix $(\boldsymbol{X}-\mathrm{E}\boldsymbol{X})(\boldsymbol{Y}-\mathrm{E}\boldsymbol{Y})'$ is $(X_i-\mathrm{E}X_i)(Y_j-\mathrm{E}Y_j)$, whose expectation is $\mathrm{Cov}(X_i\,,X_j)$. Because this is true for all $(i\,,j)$, the result holds coordinatewise. □

Warning. Note where the transpose is: Except in the case that n and m are the same integer, $(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})'(\boldsymbol{Y} - \mathrm{E}\boldsymbol{Y})$ does not even make sense, whereas $(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})(\boldsymbol{Y} - \mathrm{E}\boldsymbol{Y})'$ is always a random $m \times n$ matrix. □

An important special case occurs when we have $\boldsymbol{X} = \boldsymbol{Y}$. In that case we write

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

We call $\mathrm{Var}(\boldsymbol{X})$ the *variance-covariance* matrix of $\boldsymbol{X}$. The terminology is motivated by the fact that

$$\mathrm{Var}(\boldsymbol{X}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1\,,X_2) & \mathrm{Cov}(X_1\,,X_3) & \cdots & \mathrm{Cov}(X_1\,,X_m) \\ \mathrm{Cov}(X_2\,,X_1) & \mathrm{Var}(X_2) & \mathrm{Cov}(X_2\,,X_3) & \cdots & \mathrm{Cov}(X_2\,,X_n) \\ \mathrm{Cov}(X_3\,,X_1) & \mathrm{Cov}(X_3\,,X_2) & \mathrm{Var}(X_3) & \cdots & \mathrm{Cov}(X_3\,,X_m) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_m\,,X_1) & \mathrm{Cov}(X_m\,,X_2) & \mathrm{Cov}(X_m\,,X_3) & \cdots & \mathrm{Var}(X_m) \end{pmatrix}.$$

Note that $\mathrm{Var}(\boldsymbol{X})$ is *always* a square and symmetric matrix; its dimension is $m \times m$ when $\boldsymbol{X}$ is m-dimensional. On-diagonal entries of $\mathrm{Var}(\boldsymbol{X})$ are always nonnegative; off-diagonal entries can be arbitrary real numbers.

3. Mathematical properties of variance and covariance

- Because $(\boldsymbol{X}-\mathrm{E}\boldsymbol{X})(\boldsymbol{X}-\mathrm{E}\boldsymbol{X})' = \boldsymbol{X}\boldsymbol{X}' - \boldsymbol{X}(\mathrm{E}\boldsymbol{X})' - (\mathrm{E}\boldsymbol{X})\boldsymbol{X}' + (\mathrm{E}\boldsymbol{X})(\mathrm{E}\boldsymbol{X})'$, it follows that

  $$\begin{aligned}\mathrm{Var}(\boldsymbol{X}) &= \mathrm{E}\left(\boldsymbol{X}\boldsymbol{X}'\right) - 2(\mathrm{E}\boldsymbol{X})(\mathrm{E}\boldsymbol{X})' + (\mathrm{E}\boldsymbol{X})(\mathrm{E}\boldsymbol{X})' \\ &= \mathrm{E}\left(\boldsymbol{X}\boldsymbol{X}'\right) - (\mathrm{E}\boldsymbol{X})(\mathrm{E}\boldsymbol{X})',\end{aligned}$$

  after expansion. This is a multidimensional extension of the formula $\mathrm{Var}(Z) = \mathrm{E}(Z^2) - (\mathrm{E}Z)^2$, valid for every [univariate] random variable Z.
- If $\boldsymbol{a} \in \mathbf{R}^n$ is nonrandom, then $(\boldsymbol{X} - \boldsymbol{a}) - \mathrm{E}(\boldsymbol{X} - \boldsymbol{a}) = \boldsymbol{X} - \mathrm{E}\boldsymbol{X}$. Therefore,

  $$\mathrm{Var}(\boldsymbol{X} - \boldsymbol{a}) = \mathrm{E}\left[(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})'\right] = \mathrm{Var}(\boldsymbol{X}).$$

  This should be a familiar property in the one-dimensional case.
- If $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are three jointly-distributed random vectors [with the same dimensions], then $\boldsymbol{X}((\boldsymbol{Y}+\boldsymbol{Z})-\mathrm{E}(\boldsymbol{Y}+\boldsymbol{Z}))' = \boldsymbol{X}(\boldsymbol{Y}-\mathrm{E}\boldsymbol{Y})' + \boldsymbol{X}(\boldsymbol{Z} - \mathrm{E}\boldsymbol{Z})'$. Therefore,

  $$\mathrm{Cov}(\boldsymbol{X}\,, \boldsymbol{Y} + \boldsymbol{Z}) = \mathrm{Cov}(\boldsymbol{X}\,, \boldsymbol{Y}) + \mathrm{Cov}(\boldsymbol{X}\,, \boldsymbol{Z}).$$

- Suppose $\boldsymbol{A}\,, \boldsymbol{B}$ are nonrandom matrices. Then, $(\boldsymbol{A}\boldsymbol{X}-\mathrm{E}(\boldsymbol{A}\boldsymbol{X}))(\boldsymbol{B}\boldsymbol{Y}-\mathrm{E}(\boldsymbol{B}\boldsymbol{Y}))' = \boldsymbol{A}(\boldsymbol{X} - \mathrm{E}\boldsymbol{X})(\boldsymbol{Y} - \mathrm{E}\boldsymbol{Y})'\boldsymbol{B}'$. Therefore,

  $$\mathrm{Cov}(\boldsymbol{A}\boldsymbol{X}\,, \boldsymbol{B}\boldsymbol{Y}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X}\,, \boldsymbol{Y})\boldsymbol{B}'.$$

  The special case that $\boldsymbol{X} = \boldsymbol{Y}$ is worth pointing out: In that case we obtain the identity,

  $$\mathrm{Var}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\mathrm{Var}(\boldsymbol{X})\boldsymbol{A}'.$$

4. A relation to positive-semidefinite matrices

Let $\boldsymbol{a} \in \mathbf{R}^n$ be a nonrandom vector and X be an n-dimensional random vector. Then, the properties of variance-covariance matrices ensure that

$$\mathrm{Var}\left(\boldsymbol{a}'X\right) = \boldsymbol{a}'\mathrm{Var}(X)\boldsymbol{a}.$$

Because $\boldsymbol{a}'X = \sum_{j=1}^{n} a_j X_j$ is univariate, $\mathrm{Var}(\boldsymbol{a}'X) \geq 0$, and hence

$$\boldsymbol{a}'\mathrm{Var}(X)\boldsymbol{a} \geq 0 \qquad \text{for all } \boldsymbol{a} \in \mathbf{R}^n. \tag{1}$$

A real and symmetric $n \times n$ matrix $\boldsymbol{A}$ is said to be *positive semidefinite* if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbf{R}^n$. And $\boldsymbol{A}$ is *positive definite* if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for every nonzero $\boldsymbol{x} \in \mathbf{R}^n$.

Proposition 3. *If X is an n-dimensional random vector, then* $\mathrm{Var}(X)$ *is positive semidefinite. If* $\mathrm{P}\{\boldsymbol{a}'X = b\} = 0$ *for every* $\boldsymbol{a} \in \mathbf{R}^n$ *and* $b \in \mathbf{R}$, *then* $\mathrm{Var}(X)$ *is positive definite.*

Proof. We have seen already in (1) that $\mathrm{Var}(X)$ is positive semidefinite. Now suppose that $\mathrm{P}\{\boldsymbol{a}'X = b\} = 0$, as indicated. Then, $\boldsymbol{a}'X$ is a genuine random variable and hence $\boldsymbol{a}'\mathrm{Var}(X)\boldsymbol{a} = \mathrm{Var}(\boldsymbol{a}'X) > 0$ for all $\boldsymbol{a} \in \mathbf{R}^n$. □

Remark 4. The very same argument can be used to prove the following improvement: Suppose $\mathrm{P}\{\boldsymbol{a}'X = b\} < 1$ for all $b \in \mathbf{R}$ and $\boldsymbol{a} \in \mathbf{R}^n$. Then $\mathrm{Var}(X)$ is positive definite. The proof is the same because $\mathrm{P}\{\boldsymbol{a}'X = \mathrm{E}(\boldsymbol{a}'X)\} < 1$ implies that the variance of the random variable $\boldsymbol{a}'X$ cannot be zero when $\boldsymbol{a} \neq \mathbf{0}$. □