Stat 111 Vector Random Variables

Mathematical Statistics 2, Spring 2025 | Caroline Yao

Let's begin the discussion of vector random variables by defining a n dimensional random vector.

Definition 1: We define the **random vector** $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$ where X_1 to X_n are jointly distributed random variables with $\mathbb{E}[X_i] = \mu_i$ and $Cov(X_i, X_j) = \sigma_{ij}$.

Then it is easy to see that the **mean vector** $\mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$. We further define the **covariance matrix** of $\mathbf{X}$, denoted $\mathbf{V}$ to be an $n \times n$ matrix with the ij element σ_{ij}, the covariance of X_i and X_j. Note that $\mathbf{V}$ is symmetric.

We now verify that the usual definition of $Cov(X_i, X_j)$ arises as the expectation of the i, jth element of the random $n \times n$ matrix $(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T$. We first expand the expectation of the matrix term:

$$\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = \mathbb{E}[\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \dots & X_n - \mu_n \end{bmatrix}]$$

$$= \mathbb{E}[\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \dots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \dots & (X_2 - \mu_2)(X_n - \mu_n) \\ \dots & \dots & \dots & \dots \\ (X_n - \mu_n)(X_1 - \mu_1) & (X_n - \mu_n)(X_2 - \mu_2) & \dots & (X_n - \mu_n)^2 \end{bmatrix}]$$

$$= \begin{bmatrix} \mathbb{E}[(X_1 - \mu_1)^2] & \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \dots & \mathbb{E}[(X_1 - \mu_1)(X_n - \mu_n)] \\ \mathbb{E}[(X_2 - \mu_2)(X_1 - \mu_1)] & \mathbb{E}[(X_2 - \mu_2)^2] & \dots & \mathbb{E}[(X_2 - \mu_2)(X_n - \mu_n)] \\ \dots & \dots & \dots & \dots \\ \mathbb{E}[(X_n - \mu_n)(X_1 - \mu_1)] & \mathbb{E}[(X_n - \mu_n)(X_2 - \mu_2)] & \dots & \mathbb{E}[(X_n - \mu_n)^2] \end{bmatrix}$$

$$= \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \dots & Cov(X_2, X_n) \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & Var(X_n) \end{bmatrix}$$

Now let's consider a special case where X is an $n \times 1$ vector of independent random variables with means μ_i and variances σ_i^2 for $i = 1, \dots, n$. Then $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$ and $\mathbf{V} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$.

In this case we have all covariance equal 0 because the n random variables are independent.

With these information in hand, let's try to understand the formulas for the mean and covariance matrix of linear transformations. We define A to be a $m \times n$ matrix and C to be a $m \times 1$ matrix of constants.

Theorem 1: $\mathbb{E}[\mathbf{AX} + \mathbf{C}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{C}$.

Proof. Let $\mathbf{Y} = \mathbf{AX} + \mathbf{C}$. Then for every ith component of Y, we can write $Y_i = \sum_{i=1}^{n} a_{ij} X_j + c_i$. Then by linearity of expectation, we can write $\mathbb{E}[Y_i] = \sum_{i=1}^{n} a_{ij}\mathbb{E}[X_j] + c_i$. We can then turn these equations into matrix form and this completes the proof. $\square$

Note: You may also choose to notice that taking expectation is a linear transformation so by additive and scalar multiplication property of linear transformation, i.e. if T is a linear transformation, then for all x , y in some vector space W, $T(x + y) = T(x) + T(y)$ and for all constants c, $T(cx) = cT(x)$, we can pull out the constants and the expectation will only apply on the input variables, namely X.

Theorem 2: $Cov(\mathbf{AX} + \mathbf{C}) = \mathbf{A}\boldsymbol{V_X}\boldsymbol{A}^T$.

Proof. Let $\mathbf{Y} = \mathbf{AX} + \mathbf{C}$.

$$\boldsymbol{V_Y} = \mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^T]$$

We substitute $\mathbf{Y} = \mathbf{AX} + \mathbf{C}$ and $\mathbb{E}[\mathbf{Y}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{C}$ from previous theorem, then

$$\mathbf{Y} - \mathbb{E}[\mathbf{Y}] = \mathbf{AX} + \mathbf{C} - (\mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{C}) = \mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])$$

Then

$$\begin{aligned}
\boldsymbol{V_Y} &= \mathbb{E}[\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T)] \\
&= \mathbb{E}[\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T\mathbf{A}^T] \\
&= \mathbf{A}\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]\mathbf{A}^T \\
&= \mathbf{A}\boldsymbol{V_X}\mathbf{A}^T
\end{aligned}$$

$\square$

Theorem 3: $Var(\sum_{i=1}^{n} X_i) = (\sum_{i=1}^{n} Var(X_i)) + 2(\sum_{i<j} Cov(X_i, X_j))$.

Proof.

$$\begin{aligned}
Var(\sum_{i=1}^{n} X_i) &= \mathbb{E}[(X_1 + \dots X_n)^2] - (\mathbb{E}[X_1 + \dots + X_n])^2 \\
&= \mathbb{E}[X_1^2 + X_2^2 + \dots + X_n^2 + 2X_1X_2 + 2X_1X_3 + \dots + 2X_{n-1}X_n] - (\mathbb{E}[X_1 + \dots X_n])^2 \\
&= \mathbb{E}[\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_iX_j] - (\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n])^2 \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{i<j} \mathbb{E}[X_iX_j] - (\mathbb{E}[X_1]^2 + \dots + \mathbb{E}[X_n]^2 + 2\mathbb{E}[X_1]\mathbb{E}[X_2] + \dots 2\mathbb{E}[X_{n-1}]\mathbb{E}[X_n]) \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{i<j} \mathbb{E}[X_iX_j] - (\sum_{i=1}^{n} \mathbb{E}[X_i]^2 + 2\sum_{i<j} \mathbb{E}[X_i]\mathbb{E}[X_j]) \\
&= (\sum_{i=1}^{n} \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2) + 2(\sum_{i<j} \mathbb{E}[X_iX_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]) \\
&= (\sum_{i=1}^{n} Var(X_i)) + 2(\sum_{i<j} Cov(X_i, X_j))
\end{aligned}$$

□

Theorem 4: 1) $Cov(aX+b, c+dY) = adCov(X,Y)$; 2) $Cov(X+Y,Z) = Cov(X,Z)+Cov(Y,Z)$.

Proof. 1)

$$\begin{aligned}
Cov(aX+b, c+dY) &= \mathbb{E}[(aX+b-\mathbb{E}[aX+b])(c+dY-\mathbb{E}[c+dY])] \\
&= \mathbb{E}[(aX+b-a\mathbb{E}[X]-b)(c+dY-c-d\mathbb{E}[Y])] \\
&= \mathbb{E}[(aX-a\mathbb{E}[X])(dY-d\mathbb{E}[Y])] \\
&= \mathbb{E}[adXY - adX\mathbb{E}[Y] - adY\mathbb{E}[X] + ad\mathbb{E}[X]\mathbb{E}[Y]] \\
&= ad\mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= ad(\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]) \\
&= adCov(X,Y)
\end{aligned}$$

2)

$$\begin{aligned}
Cov(X+Y,Z) &= \mathbb{E}[(X+Y-\mathbb{E}[X+Y])(Z-\mathbb{E}[Z])] \\
&= \mathbb{E}[XZ - X\mathbb{E}[Z] + YZ - Y\mathbb{E}[Z] - Z\mathbb{E}[X+Y] + \mathbb{E}[X+Y]\mathbb{E}[Z]] \\
&= \mathbb{E}[XZ - X\mathbb{E}[Z] + YZ - Y\mathbb{E}[Z] - \mathbb{E}[X] - Z\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Z] + \mathbb{E}[Y]\mathbb{E}[Z]] \\
&= \mathbb{E}[XZ - X\mathbb{E}[Z] - Z\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Z]] + \mathbb{E}[YZ - Y\mathbb{E}[Z] - Z\mathbb{E}[Y] + \mathbb{E}[Y]\mathbb{E}[Z]] \\
&= \mathbb{E}[(X-\mathbb{E}[X])(Z-\mathbb{E}[Z])] + \mathbb{E}[(Y-\mathbb{E}[Y])(Z-\mathbb{E}[Z])] \\
&= Cov(X,Z) + Cov(Y,Z)
\end{aligned}$$

□

Finally, we show that for $X_1, \ldots X_n$ iid random variables, we show that $\bar{X}$ is uncorrelated with $X_i - \bar{X}$, for all $i = 1, \ldots, n$.

$$
\begin{aligned}
Cov(\bar{X}, X_i - \bar{X}) &= Cov(X_i - \bar{X}, \bar{X}) && (1)\\
&= Cov(X_i, \bar{X}) - Cov(\bar{X}, \bar{X})\\
&= Cov(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j) - Var(\bar{X})\\
&= \frac{1}{n}\sum_{i=1}^{n} Cov(X_j, X_i) - Var(\bar{X}) && (2)\\
&= \frac{1}{n}Var(X_i) - Var(\bar{X})\\
&= \frac{1}{n}Var(X_i) - Var(\frac{1}{n}\sum_{i=1}^{n} X_i)\\
&= \frac{1}{n}Var(X_i) - \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i)\\
&= \frac{1}{n}Var(X_i) - \frac{1}{n^2}nVar(X_i) && (3)\\
&= \frac{1}{n}Var(X_i) - \frac{1}{n}Var(X_i)\\
&= 0
\end{aligned}
$$

Line (1) is by part 2) of the above theorem. Line (2) is by part 1) of the above theorem. Line (3) is by the fact that X_1 to X_n are iid random variables.

Let's finally use the vector matrix formulation to solve the same problem.
Let X be a random n vector with $\mathbb{E}[\mathbf{X}] = \mu\mathbf{1}$ and $V_{XX} = \sigma^2\mathbf{I}$. Let $Y = \bar{X}$ and let $\mathbf{Z}$ be the vector with ith element $X_i - \bar{X}$. Let's find V_{ZY}. In matrix form,

$$
\begin{aligned}
\mathbf{Z} &= (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\mathbf{X}\\
\mathbf{Y} &= \frac{1}{n}\mathbf{1}^T\mathbf{X}
\end{aligned}
$$

Then by a Corollary of Theorem 2 (see below), we have

$$
V_{ZY} = (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(\sigma^2\mathbf{I})(\frac{1}{n}\mathbf{1}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}
$$

□

Theorem 5: Let $\mathbf{X}$ be a random vector with covariance matrix V_{XX}. If $\mathbf{Y} = \mathbf{AX}$ where $\mathbf{A}$ is $p \times n$ and $\mathbf{Z} = \mathbf{BX}$ where $\mathbf{B}$ is $m \times n$. Then the **cross-covariance** matrix of $\mathbf{Y}$ and $\mathbf{Z}$ is $\boldsymbol{V_{YZ}} = \mathbf{A}\boldsymbol{V_{XX}}\mathbf{B}^T$.

A cross covariance matrix is a matrix whose element in the ijth position is the covariance between the i-th element of a random vector and j-th element of another random vector. When the two random vectors are the same, the cross-covariance matrix is also the covariance matrix.

Proof. By definition

$$\begin{aligned} V_{YZ} &= \mathbb{E}[(Y-\mu_Y)(Z-\mu_Z)^T] \\ &= \mathbb{E}[(AX-\mathbb{E}[AX])(BX-\mathbb{E}[BX])^T] \\ &= \mathbb{E}[(AX-A\mathbb{E}[X])(BX-B\mathbb{E}[X])^T] \\ &= \mathbb{E}[A(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T B^T] \\ &= A\mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T]B^T \\ &= AV_{XX}B^T \end{aligned}$$

□