Covariance Algebra and Inference Foundations

Linear Gaussian Models

Sai Sampath Kedari

sampath@umich.edu

Contents

1 Properties of Variance and Covariance

Let X, Y be random variables.

1.1 Variance

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\big[(X-\mathbb{E}[X])^2\big] \\
&= \mathbb{E}\big[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2\big] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

1.2 Covariance

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] \\
&= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

1.3 Covariance with a Constant

Let c be a constant.

$$\begin{aligned}
\mathrm{Cov}(X,c) &= \mathbb{E}\big[(X-\mathbb{E}[X])(c-\mathbb{E}[c])\big] \\
&= \mathbb{E}\big[(X-\mathbb{E}[X])\cdot 0\big] \\
&= 0
\end{aligned}$$

1.4 Shift Property

$$\begin{aligned}
\mathrm{Cov}(X+c,Y) &= \mathbb{E}\big[(X+c-\mathbb{E}[X+c])(Y-\mathbb{E}[Y])\big] \\
&= \mathbb{E}\big[(X+c-\mathbb{E}[X]-c)(Y-\mathbb{E}[Y])\big] \\
&= \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] \\
&= \mathrm{Cov}(X,Y)
\end{aligned}$$

1.5 Scaling Property

$$\begin{aligned}
\mathrm{Cov}(aX,bY) &= \mathbb{E}\big[(aX-\mathbb{E}[aX])(bY-\mathbb{E}[bY])\big] \\
&= \mathbb{E}\big[a(X-\mathbb{E}[X])\,b(Y-\mathbb{E}[Y])\big] \\
&= ab\,\mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big] \\
&= ab\,\mathrm{Cov}(X,Y)
\end{aligned}$$

1.6 Symmetry of Covariance

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[YX] - \mathbb{E}[Y]\mathbb{E}[X] \\
&= \mathrm{Cov}(Y,X)
\end{aligned}$$

1.7 Variance of a Sum

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}\big[(X+Y-\mathbb{E}[X+Y])^2\big]\\ &= \mathbb{E}\big[(X-\mathbb{E}[X]+Y-\mathbb{E}[Y])^2\big]\\ &= \mathbb{E}\big[(X-\mathbb{E}[X])^2\big]\\ &\quad+\mathbb{E}\big[(Y-\mathbb{E}[Y])^2\big]\\ &\quad+2\mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big]\\ &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y)\end{aligned}$$

1.8 Variance of Linear Combination

$$\begin{aligned}\mathrm{Var}(aX+bY) &= \mathrm{Var}(aX)+\mathrm{Var}(bY)+2\mathrm{Cov}(aX,bY)\\ &= a^2\mathrm{Var}(X)+b^2\mathrm{Var}(Y)+2ab\,\mathrm{Cov}(X,Y)\end{aligned}$$

1.9 Covariance of Sums

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

1.10 Multiple Random Variables

Let $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ be random variables.

$$\begin{aligned}\mathrm{Cov}\left(\sum_{i=1}^m a_iX_i,\sum_{j=1}^n b_jY_j\right) &= \mathbb{E}\left[\sum_{i=1}^m a_i(X_i-\mathbb{E}[X_i])\sum_{j=1}^n b_j(Y_j-\mathbb{E}[Y_j])\right]\\ &= \sum_{i=1}^m\sum_{j=1}^n a_ib_j\mathbb{E}\big[(X_i-\mathbb{E}[X_i])(Y_j-\mathbb{E}[Y_j])\big]\\ &= \sum_{i=1}^m\sum_{j=1}^n a_ib_j\mathrm{Cov}(X_i,Y_j)\end{aligned}$$

Define the random vectors

$$\mathbf{X}=\begin{bmatrix}X_1\\ \vdots\\ X_m\end{bmatrix},\qquad \mathbf{Y}=\begin{bmatrix}Y_1\\ \vdots\\ Y_n\end{bmatrix},$$

and deterministic vectors

$$\mathbf{a}=\begin{bmatrix}a_1\\ \vdots\\ a_m\end{bmatrix},\qquad \mathbf{b}=\begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix}.$$

Then

$$\sum_{i=1}^{m} a_i X_i = \mathbf{a}^\top \mathbf{X}, \qquad \sum_{j=1}^{n} b_j Y_j = \mathbf{b}^\top \mathbf{Y}.$$

Define the cross covariance matrix

$$\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}\big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top\big],$$

whose (i, j) entry is $\mathrm{Cov}(X_i, Y_j)$.

Hence,

$$\mathrm{Cov}(\mathbf{a}^\top \mathbf{X}, \mathbf{b}^\top \mathbf{Y}) = \mathbf{a}^\top \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) \mathbf{b}.$$

1.11 Covariance of Random Vectors

Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors.

$$\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}\big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top\big]$$

The covariance matrix of a random vector $\mathbf{X}$ is defined as

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\mathbf{X}, \mathbf{X}).$$

1.12 Symmetry of Covariance for Random Vectors

$$\begin{aligned}
\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) &= \mathbb{E}\big[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\top\big] \\
&= \mathbb{E}\big[\big((\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\big)^\top\big] \\
&= \mathrm{Cov}(\mathbf{Y}, \mathbf{X})^\top
\end{aligned}$$

1.13 Variance of Linear Transformation

Let $\mathbf{x}$ be a random vector and A a deterministic matrix.

$$\begin{aligned}
\mathrm{Var}(A\mathbf{x}) &= \mathrm{Cov}(A\mathbf{x}, A\mathbf{x}) \\
&= \mathbb{E}\big[A(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top A^\top\big] \\
&= A \, \mathrm{Cov}(\mathbf{x}) \, A^\top
\end{aligned}$$

1.14 Affine Transformation

Let $\mathbf{y} = A\mathbf{x} + \mathbf{b}$.

$$\begin{aligned}
\mathrm{Cov}(\mathbf{y}) &= \mathrm{Cov}(A\mathbf{x} + \mathbf{b}, A\mathbf{x} + \mathbf{b}) \\
&= \mathrm{Cov}(A\mathbf{x}, A\mathbf{x}) \\
&= A \, \mathrm{Cov}(\mathbf{x}) \, A^\top
\end{aligned}$$

2 Joint and Marginal Gaussian Distributions

Let

$$\mathbf{X} \in \mathbb{R}^n, \qquad \mathbf{Y} \in \mathbb{R}^m$$

be random vectors with finite second moments.

Define the stacked random vector

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \in \mathbb{R}^{n+m}.$$

We say that $(\mathbf{X}, \mathbf{Y})$ is jointly Gaussian if

$$\mathbf{Z} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \Sigma \right),$$

where

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}.$$

By definition of covariance for random vectors,

$$\Sigma = \mathbb{E}\left[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])^\top \right].$$

Expanding the outer product yields

$$\Sigma = \begin{bmatrix} \mathbb{E}\big[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^\top & \mathbb{E}\big[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^\top \\ \mathbb{E}\big[(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{X} - \boldsymbol{\mu}_X)^\top & \mathbb{E}\big[(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{Y} - \boldsymbol{\mu}_Y)^\top \end{bmatrix}.$$

Hence,

$$\Sigma_{XX} = \mathrm{Cov}(\mathbf{X}), \qquad \Sigma_{YY} = \mathrm{Cov}(\mathbf{Y}), \qquad \Sigma_{YX} = \Sigma_{XY}^\top.$$

2.1 Marginal Gaussian Distributions

Define the linear maps

$$A_X = \begin{bmatrix} I_n & 0 \end{bmatrix}, \qquad A_Y = \begin{bmatrix} 0 & I_m \end{bmatrix}.$$

Then

$$\mathbf{X} = A_X \mathbf{Z}, \qquad \mathbf{Y} = A_Y \mathbf{Z}.$$

By linearity of expectation,

$$\mathbb{E}[\mathbf{X}] = A_X \mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}_X, \qquad \mathbb{E}[\mathbf{Y}] = A_Y \mathbb{E}[\mathbf{Z}] = \boldsymbol{\mu}_Y.$$

Using the covariance of a linear transformation,

$$\mathrm{Cov}(\mathbf{X}) = A_X \Sigma A_X^\top = \Sigma_{XX}, \qquad \mathrm{Cov}(\mathbf{Y}) = A_Y \Sigma A_Y^\top = \Sigma_{YY}.$$

Hence,

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \Sigma_{XX}), \qquad \mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \Sigma_{YY}).$$

3 Conditional Distribution of a Jointly Gaussian Random Vector

Throughout this section, we work under the joint Gaussian model defined in Section 2. We assume Σ_{YY} is symmetric positive definite.

Our goal is to derive the distribution of

$$\mathbf{X} \mid \mathbf{Y} = \mathbf{y}.$$

3.1 Lemma 1: Joint Gaussian Density

The joint density of $\mathbf{Z}$ is

$$p(\mathbf{z}) = \frac{1}{(2\pi)^{(n+m)/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\right).$$

3.2 Lemma 2: Conditional Density

For random vectors with joint density $p(\mathbf{x}, \mathbf{y})$,

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})}.$$

3.3 Lemma 3: Block Matrix Inversion via Schur Complement

Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \qquad D \text{ invertible.}$$

Define the Schur complement

$$S = A - BD^{-1}C.$$

Assume S is invertible.

Then

$$M^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{bmatrix}.$$

3.4 Applying Lemma 3 to the Covariance Matrix

Identify

$$A = \Sigma_{XX}, \quad B = \Sigma_{XY}, \quad C = \Sigma_{YX}, \quad D = \Sigma_{YY}.$$

Define the conditional covariance

$$\Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}.$$

Then

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{X|Y}^{-1} & -\Sigma_{X|Y}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1} \\ -\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{X|Y}^{-1} & \Sigma_{YY}^{-1} + \Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{X|Y}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1} \end{bmatrix}.$$

3.5 Expansion of the Joint Quadratic Form

Define

$$Q = \begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_X \\ \mathbf{y} - \boldsymbol{\mu}_Y \end{bmatrix}^\top \Sigma^{-1} \begin{bmatrix} \mathbf{x} - \boldsymbol{\mu}_X \\ \mathbf{y} - \boldsymbol{\mu}_Y \end{bmatrix}.$$

Substituting the block inverse yields

$$\begin{aligned} Q = {} & (\mathbf{x} - \boldsymbol{\mu}_X)^\top \Sigma_{X|Y}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) \\ & - (\mathbf{x} - \boldsymbol{\mu}_X)^\top \Sigma_{X|Y}^{-1} \Sigma_{XY} \Sigma_{YY}^{-1} (\mathbf{y} - \boldsymbol{\mu}_Y) \\ & - (\mathbf{y} - \boldsymbol{\mu}_Y)^\top \Sigma_{YY}^{-1} \Sigma_{YX} \Sigma_{X|Y}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X) \\ & + \text{terms independent of } \mathbf{x}. \end{aligned}$$

3.6 Completing the Square

Using

$$x^\top A x - 2x^\top b = (x - A^{-1}b)^\top A (x - A^{-1}b) - b^\top A^{-1} b,$$

define

$$\boldsymbol{\mu}_{X|Y} = \boldsymbol{\mu}_X + \Sigma_{XY} \Sigma_{YY}^{-1} (\mathbf{y} - \boldsymbol{\mu}_Y).$$

Then

$$Q = (\mathbf{x} - \boldsymbol{\mu}_{X|Y})^\top \Sigma_{X|Y}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{X|Y}) + \text{terms independent of } \mathbf{x}.$$

3.7 Conditional Distribution

Therefore,

$$\mathbf{X} \mid \mathbf{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_{X|Y}, \Sigma_{X|Y}\right).$$