Lecture 5 — Multivariate distributions

Mohammad Reza Karimi

1 Review of vector operations and notation

A vector $x \in \mathbb{R}^k$ is represented as a tuple $(x_1, \ldots, x_k)$, or a column vector,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}.$$

To make life easier and save space, we often think of x as the *transpose* of a row vector, i.e.,

$$x = (x_1, x_2, \ldots, x_k)^\top.$$

Note that these are just syntactic sugar; a vector is just a tuple.

The *outer* product of a vector x with itself is the matrix

$$xx^\top = \begin{pmatrix} x_1^2 & & \cdots & & x_1 x_k \\ \vdots & \ddots & & & \vdots \\ \vdots & & x_i x_j & & \vdots \\ \vdots & & & \ddots & \vdots \\ x_k x_1 & & \cdots & & x_k^2 \end{pmatrix}.$$

This is the multi-dimensional generalization of the square x^2 of a number x.

The *inner* product between two vectors x and y in $\mathbb{R}^k$ is the scalar (number)

$$\langle x, y \rangle = x \cdot y = x^\top y = \sum_{i=1}^{k} x_i y_i.$$

We mainly use $\langle x, y \rangle$ and $x^\top y$ to denote inner products.
The $k \times k$ *identity* matrix is denoted I_k.

2 Random Vectors

A random vector $X \in \mathbb{R}^k$ is just a vector of random variables $X_1, \ldots, X_k$:

$$X = (X_1, X_2, \ldots, X_k).$$

The **expectation** of $X \in \mathbb{R}^k$ is the vector of expectations of the individual coordinates:

$$\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_k]).$$

It is tempting to define the variance of X as the vector of variances of individual coordinates. But this does not capture pairwise covariances.

Definition 2.1: Covariance between random variables

The covariance between X_i and X_j is

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] = \mathbb{E}[X_i X_j] - \mu_i \mu_j,$$

where $\mu_i = \mathbb{E}[X_i]$ and $\mu_j = \mathbb{E}[X_j]$. Note that the covariance of X_i with itself is just the variance of X_i, i.e., $\mathrm{Cov}(X_i, X_i) = \mathbb{V}[X_i]$.

The **covariance matrix** of a random vector is just the matrix of all pairwise covariances. We get it via an *outer product*.

Definition 2.2: Covariance matrix of a random vector

The covariance Σ of a random vector $X \in \mathbb{R}^k$ is the $k \times k$ matrix of pairwise covariances:

$$\Sigma = \mathbb{V}[X] = \mathbb{E}\left[(X - \mu)(X - \mu)^\top\right] = \mathbb{E}\left[XX^\top\right] - \mu\mu^\top,$$

where $\mu = \mathbb{E}[X]$ is the expectation vector of X.

Note that the ijth entry of Σ is

$$\Sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{Cov}(X_i, X_j).$$

The diagonal entries of Σ are the variances of X_i:

$$\Sigma_{ii} = \mathbb{V}[X_i].$$

Remark.

> The inverse of the covariance matrix is sometimes called the *precision matrix*.

Theorem 2.3: Expectation and covariance of linearly transformed random vectors

Let $X \in \mathbb{R}^k$ be a random vector, with $\mathbb{E}[X] = \mu$, $\mathbb{V}[X] = \Sigma$.

1. Let $a \in \mathbb{R}^k$ be a deterministic vector. Then $a^\top X \in \mathbb{R}$ is a random variable, with $\mathbb{E}[a^\top X] = a^\top \mu$ and $\mathbb{V}[a^\top X] = a^\top \Sigma a$.

2. Let A be a deterministic $k \times \ell$ matrix and $b \in \mathbb{R}^\ell$ be a deterministic vector. Then $A^\top X + b \in \mathbb{R}^\ell$ is a random vector, with $\mathbb{E}[A^\top X + b] = A^\top \mu + b$ and $\mathbb{V}[A^\top X + b] = A^\top \Sigma A$.

Proof. We prove the first statement. For the expectation, we use the linearity of expectation to get $\mathbb{E}[a^\top X] = a^\top \mathbb{E}[X] = a^\top \mu$.

For the variance, we first compute the expectation of $(a^\top X)^2$:

$$\mathbb{E}[(a^\top X)^2] = \mathbb{E}[(a^\top X)(a^\top X)] = \mathbb{E}[a^\top X X^\top a] = a^\top \mathbb{E}[X X^\top] a \tag{1}$$

by linearity. We next compute the square of the expectation:

$$\left(\mathbb{E}[a^\top X]\right)^2 = (a^\top \mu)^2 = a^\top \mu \mu^\top a. \tag{2}$$

Finally, subtract (2) from (1) to get

$$\begin{aligned} \mathbb{V}[a^\top X] &= \mathbb{E}[(a^\top X)^2] - \left(\mathbb{E}[a^\top X]\right)^2 = a^\top \mathbb{E}[X X^\top] a - a^\top \mu \mu^\top a \\ &= a^\top \left(\mathbb{E}[X X^\top] - \mu \mu^\top\right) a = a^\top \Sigma a. \end{aligned}$$

□

Remark.

Note that we didn't have to manipulate indices at all in the above proof. But as a sanity check, let's redo the proof using indices. For the expectation, we get

$$\mathbb{E}[a^\top X] = \mathbb{E}\left[\sum_{i=1}^k a_i X_i\right] = \sum_{i=1}^k a_i \mathbb{E}[X_i] = \sum_{i=1}^k a_i \mu_i = a^\top \mu,$$

using linearity to get the second equality. For the variance, we get

$$\begin{aligned}\mathbb{V}[a^\top X] = \mathbb{V}\left[\sum_{i=1}^k a_i X_i\right] &= \mathrm{Cov}\left(\sum_{i=1}^k a_i X_i, \sum_{j=1}^k a_j X_j\right) \\ &= \sum_{i,j=1}^k a_i a_j \mathrm{Cov}(X_i, X_j) = \sum_{i,j=1}^k a_i a_j \Sigma_{ij} = a^\top \Sigma a.\end{aligned}$$

We used bilinearity of covariance to get the third equality.