18.650: Fundamentals of Statistics Fall 2023

Lecture 5 — September 15, 2023

Prof. Philippe Rigollet *Scribe: Anya Katsevich*

1 Review of vector operations and notation

A vector $x \in \mathbb{R}^k$ is represented as a column vector,

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

To save space, we often write x as the *transpose* of a row vector, i.e.

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

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

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

The outer product $xx^\top$ of a vector x 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)

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

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 $X_1, \ldots, X_k$:

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

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

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

It is tempting to define the variance of X as the vector of variances of the 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

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

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. $\text{Cov}(X_i, X_i) = \mathbb{V}[X_i]$.

The **covariance matrix** of a random vector is just the matrix of all the 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:

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

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

Note that the ijth entry of Σ is

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

The diagonal entries of Σ are the variances of the 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 $\mathbb{E}[A^\top X] = A^\top \mu + b$ and $\mathbb{V}[A^\top X + b] = A^\top \Sigma A$.

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

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:

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

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

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

$\square$

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.

3 Multivariate Gaussian & multivariate limit theorems

A k-dimensional Gaussian random vector X is denoted $X \sim \mathcal{N}_k(\mu, \Sigma)$, where $\mu = \mathbb{E}[X]$ is the expectation, $\Sigma = \mathbb{V}[X]$ is the covariance matrix, and the subscript k tells you that X is k-dimensional.

The pdf of X is given by

$$f(x) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

A useful exercise is to make sure that for $k = 1$, you get back the pdf of the 1-dimensional Gaussian we covered in lecture 4.

Useful Properties. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$.

1. Linear transformation: if A is a $k \times \ell$ deterministic matrix, and $b \in \mathbb{R}^\ell$ is a vector, then
$$A^\top X + b \sim \mathcal{N}_\ell(A^\top \mu + b, A^\top \Sigma A).$$
2. Standardization: $Z = \Sigma^{-1/2}(X - \mu) \sim \mathcal{N}_k(0, I_k)$ and $X = \Sigma^{1/2} Z + \mu$.

These properties follow from Theorem 2.3.

Theorem 3.1: Multivariate CLT

Let $X_1, \dots, X_n$ be iid random vectors in $\mathbb{R}^k$, with $\mathbb{E}[X_1] = \mu$ and $\mathbb{V}[X_1] = \Sigma$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow \mathcal{N}(0, \Sigma).$$

Theorem 3.2: Multivariate Delta Method

Let $X_1, \dots, X_n$ be iid random vectors in $\mathbb{R}^k$, with $\mathbb{E}[X_1] = \mu$ and $\mathbb{V}[X_1] = \Sigma$, and let $g : \mathbb{R}^k \to \mathbb{R}$. Then

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \rightsquigarrow \mathcal{N}\left(0, \nabla g(\mu)^\top \Sigma \nabla g(\mu)\right),$$

where $\nabla g(\mu)$ is the column vector with ith coordinate $\partial_i g(\mu)$, $i = 1, \dots, k$.

Example.

Let $g : \mathbb{R}^k \to \mathbb{R}$ be given by $g(x) = x_1 x_2 \dots x_k$. Suppose $X_1, \dots, X_n$ are iid random vectors in $\mathbb{R}^k$, with mean $\mu = (2, \dots, 2)^\top$ and covariance $\Sigma = I_k$. We apply the delta method to get the limiting distribution of $g(\bar{X}_n)$.

To do this we need to compute $g(\mu)$, $\nabla g(\mu)$, and $\nabla g(\mu)^\top \Sigma \nabla g(\mu)$, where $\mu = (2, \dots, 2)^\top$ and $\Sigma = I_k$. For $g(\mu)$, we get $g(\mu) = 2^k$. For the gradient, we first compute at a generic x that

$$\nabla g(x) = \begin{pmatrix} x_2 x_3 \dots x_k \\ x_1 x_3 \dots x_k \\ \vdots \\ x_1 x_2 \dots x_{k-1} \end{pmatrix}.$$

Plugging in $x = \mu = (2, \dots, 2)^\top$, we get

$$\nabla g(\mu) = (2^{k-1}, \dots, 2^{k-1})^\top = 2^{k-1} \mathbb{1}_k$$

where $\mathbb{1}_k$ is the k-vector of all ones. Finally,

$$\nabla g(\mu)^\top \Sigma \nabla g(\mu) = \left(2^{k-1} \mathbb{1}_k\right)^\top I_k \left(2^{k-1} \mathbb{1}_k\right) = 2^{2k-2} \mathbb{1}_k^\top \mathbb{1}_k = 2^{2k-2} k.$$

Putting it all together,

$$\sqrt{n}(g(\bar{X}_n) - 2^k) \rightsquigarrow \mathcal{N}\left(0, 2^{2k-2} k\right) \quad \text{as} \quad n \to \infty.$$