Lecture 2 Review:

Elementary Matrix Algebra Review

- rank, trace, transpose, determinants, orthogonality, etc.,
- linear independence, range (column) space, null space,
- spectral theorem/principal axis theorem,
- idempotent matrices, projection matrices, positive definite and positive semi-definite matrices.

RANDOM VECTORS

Definitions:

1. A *random vector* is a vector of random variables
$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$
2. The mean or expectation of $\mathbf{X}$ is defined as
$$E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}.$$
3. A *random matrix* is a matrix of random variables $\mathbf{Z} = (Z_{ij})$. Its expectation is given by $E[\mathbf{Z}] = (E[Z_{ij}])$.

Properties:

1. A constant vector $\mathbf{a}$ and a constant matrix $\mathbf{A}$ satisfy $E[\mathbf{a}] = \mathbf{a}$ and $E[\mathbf{A}] = \mathbf{A}$. ("Constant" means non-random in this context.)
2. $E[\mathbf{X} + \mathbf{Y}] = E[\mathbf{X}] + E[\mathbf{Y}]$.
3. $E[\mathbf{AX}] = \mathbf{A}E[\mathbf{X}]$ for a constant matrix $\mathbf{A}$.
4. More generally (Seber & Lee Theorem 1.1):

$$E[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}E[\mathbf{Z}]\mathbf{B} + \mathbf{C}$$

   if $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are constant matrices.

Definition: If $\mathbf{X}$ is a random vector, the *covariance matrix* of $\mathbf{X}$ is defined as

$$\begin{aligned} \text{cov}(\mathbf{X}) &\equiv [\text{cov}(X_i, X_j)] \\ &\equiv \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{var}(X_n) \end{pmatrix}. \end{aligned}$$

Also called the *variance matrix* or the *variance-covariance matrix*.

Alternatively:

$$\begin{aligned} \text{cov}(\mathbf{X}) &= E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'] \\ &= E\left[\begin{pmatrix} X_1 - E[X_1] \\ \vdots \\ X_n - E[X_n] \end{pmatrix} (X_1 - E[X_1], \cdots, X_n - E[X_n]) \right]. \end{aligned}$$

Example: (Independent random variables.) If $X_1, \ldots, X_n$ are independent then $\text{cov}(\mathbf{X}) = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

If, in addition, the X_i have common variance σ^2, then $\text{cov}(\mathbf{X}) = \sigma^2 \mathbf{I}_n$.

Properties of Covariance Matrices:

1. Symmetric: $\text{cov}(\mathbf{X}) = [\text{cov}(\mathbf{X})]'$.

   Proof: $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$.

2. $\text{cov}(\mathbf{X} + \mathbf{a}) = \text{cov}(\mathbf{X})$ if $\mathbf{a}$ is a constant vector.

3. $\text{cov}(\mathbf{AX}) = \mathbf{A}\text{cov}(\mathbf{X})\mathbf{A}'$ if $\mathbf{A}$ is a constant matrix.

   Proof:

$$\begin{aligned}
\text{cov}(\mathbf{AX}) &= E[(\mathbf{AX} - E[\mathbf{AX}])(\mathbf{AX} - E[\mathbf{AX}])'] \\
&= E[\mathbf{A}(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'\mathbf{A}'] \\
&= \mathbf{A}E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])']\mathbf{A}' \\
&= \mathbf{A}\text{cov}(\mathbf{X})\mathbf{A}'
\end{aligned}$$

4. $\text{cov}(\mathbf{X})$ is positive semi-definite.

   Proof: For any constant vector $\mathbf{a}$, $\mathbf{a}'\text{cov}(\mathbf{X})\mathbf{a} = \text{cov}(\mathbf{a}'\mathbf{X})$.
   But this is just the variance of a random variable:
   $\text{cov}(\mathbf{a}'\mathbf{X}) = \text{var}(\mathbf{a}'\mathbf{X}) \geq 0$.
   (Variances are never negative.)

   Therefore:

5. $\text{cov}(\mathbf{X})$ is positive definite provided no linear combination of the X_i is a constant (Seber & Lee Theorem 1.4)

6. $\text{cov}(\mathbf{X}) = E[\mathbf{XX}'] - E[\mathbf{X}](E[\mathbf{X}])'$

Definition: The *correlation matrix* of $\mathbf{X}$ is defined as

$$\begin{aligned}\operatorname{corr}(\mathbf{X}) &= [\operatorname{corr}(X_i, X_j)] \\ &\equiv \begin{pmatrix} 1 & \operatorname{corr}(X_1, X_2) & \cdots & \operatorname{corr}(X_1, X_n) \\ \operatorname{corr}(X_2, X_1) & 1 & \cdots & \operatorname{corr}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{corr}(X_n, X_1) & \operatorname{corr}(X_n, X_2) & \cdots & 1 \end{pmatrix}.\end{aligned}$$

Denote $\operatorname{cov}(\mathbf{X})$ by $\mathbf{\Sigma} = (\sigma_{ij})$. Then the correlation matrix and covariance matrix are related by

$$\operatorname{cov}(\mathbf{X}) = \operatorname{diag}(\sqrt{\sigma}_{11}, \ldots, \sqrt{\sigma}_{nn}) \operatorname{corr}(\mathbf{X}) \operatorname{diag}(\sqrt{\sigma}_{11}, \ldots, \sqrt{\sigma}_{nn}).$$

This is easily seen using $\operatorname{corr}(X_i, X_j) = \operatorname{cov}(X_i, X_j)/\sqrt{\sigma_{ii}\sigma_{jj}}$.

Example: (Exchangeable random variables.) If $X_1, \ldots, X_n$ are exchangeable, they have a constant variance σ^2 and a constant correlation ρ between any pair of variables. Thus

$$\operatorname{cov}(\mathbf{X}) = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

This is sometimes called an exchangeable covariance matrix.

Definition: If $\mathbf{X}_{m\times 1}$ and $\mathbf{Y}_{n\times 1}$ are random vectors,

$$\begin{aligned}
\text{cov}(\mathbf{X},\mathbf{Y}) &= [\text{cov}(X_i,Y_j)] \\
&\equiv \begin{pmatrix} \text{cov}(X_1,Y_1) & \text{cov}(X_1,Y_2) & \cdots & \text{cov}(X_1,Y_n) \\ \text{cov}(X_2,Y_1) & \text{cov}(X_2,Y_2) & \cdots & \text{cov}(X_2,Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_m,Y_1) & \text{cov}(X_m,Y_2) & \cdots & \text{cov}(X_m,Y_n) \end{pmatrix}.
\end{aligned}$$

Note: We have now defined the covariance matrix for a random vector *and* a covariance matrix for a pair of random vectors.

Alternative form:

$$\begin{aligned}
\text{cov}(\mathbf{X},\mathbf{Y}) &= E[(\mathbf{X}-E[\mathbf{X}])(\mathbf{Y}-E[\mathbf{Y}])'] \\
&= E\left[\begin{pmatrix} X_1 - E[X_1] \\ \vdots \\ X_m - E[X_m] \end{pmatrix} (Y_1 - E[Y_1], \cdots, Y_n - E[Y_n])\right].
\end{aligned}$$

Note: The covariance is defined regardless of the values of m and n.

Theorem: If $\mathbf{A}$ and $\mathbf{B}$ are constant matrices,

$$\text{cov}(\mathbf{AX},\mathbf{BY}) = \mathbf{A}\text{cov}(\mathbf{X},\mathbf{Y})\mathbf{B}'.$$

Proof: Similar to proof of $\text{cov}(\mathbf{AX}) = \mathbf{A}\text{cov}(\mathbf{X})\mathbf{A}'$.

Partitioned variance matrix: Let

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}.$$

Then

$$\text{cov}(\mathbf{Z}) = \begin{pmatrix} \text{cov}(\mathbf{X}) & \text{cov}(\mathbf{X},\mathbf{Y}) \\ \text{cov}(\mathbf{Y},\mathbf{X}) & \text{cov}(\mathbf{Y}) \end{pmatrix}.$$

Expectation of a Quadratic Form:

Theorem: Let $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ and $\mathbf{A}$ be a constant matrix. Then

$$E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}).$$

First Proof (brute force):

$$\begin{aligned} E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] &= E[\sum_i \sum_j a_{ij}(X_i - \mu_i)(X_j - \mu_j)] \\ &= \sum_i \sum_j a_{ij} E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \sum_i \sum_j a_{ij} \text{cov}(X_i, X_j) \\ &= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}). \end{aligned}$$

Second Proof (more clever):

$$\begin{aligned} E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] &= E[\text{tr}\{(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})\}] \\ &= E[\text{tr}\{\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\}] \\ &= \text{tr}\{E[\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']\} \\ &= \text{tr}\{\mathbf{A}E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']\} \\ &= \text{tr}\{\mathbf{A}\boldsymbol{\Sigma}\} \end{aligned}$$

Corollary: $E[\mathbf{X}'\mathbf{A}\mathbf{X}] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

Proof:

$$\mathbf{X}'\mathbf{A}\mathbf{X} = (\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) + \boldsymbol{\mu}'\mathbf{A}\mathbf{X} + \mathbf{X}'\mathbf{A}\boldsymbol{\mu} - \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu},$$

Therefore,

$$E[\mathbf{X}'\mathbf{A}\mathbf{X}] = E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}.$$

Example: Let $X_1, \ldots, X_n$ be independent random variables with common mean μ and variance σ^2. Then the sample variance $s^2 = \sum_i (X_i - \bar{X})^2/(n-1)$ is an unbiased estimate of σ^2.

Proof: Let $\mathbf{X} = (X_1, \ldots, X_n)'$. Then $E[\mathbf{X}] = \mu\mathbf{1}$, $\text{cov}(\mathbf{X}) = \sigma^2\mathbf{I}_{n\times n}$. Let $\mathbf{A} = \mathbf{I}_{n\times n} - \mathbf{1}_n\mathbf{1}_n'/n = \mathbf{I}_n - \bar{\mathbf{J}}_n$. ($\bar{\mathbf{J}}_n = \mathbf{1}_n\mathbf{1}_n'/n$.) Note that

$$(n-1)s^2 = \sum_i (X_i - \bar{X})^2 = \mathbf{X}'\mathbf{A}\mathbf{X}$$

By the corollary

$$\begin{aligned} E[(n-1)s^2] &= E[\mathbf{X}'\mathbf{A}\mathbf{X}] \\ &= \text{tr}(\mathbf{A}\sigma^2\mathbf{I}) + \mu\mathbf{1}'\mathbf{A}\mu\mathbf{1} \\ &= (n-1)\sigma^2 \end{aligned}$$

because $\mathbf{A}\mathbf{1} = \mathbf{0}$.

Independence of Normal Random Variables:

Theorem: For $\mathbf{x} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{x}$ are independently distributed iff $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$.

Proof: Sufficiency (Searle, 1971, §2.5), necessity (Driscol and Gundberg, 1986, American Statistician)

Example: Let $X_1, \ldots, X_n$ be independent random variables with common mean μ and variance σ^2. Show that the sample mean $\bar{X} = \sum_{i=1}^n X_i/n$ and the sample variance S^2 are independently distributed.

Let $\mathbf{x} = (\mathbf{X_1}, \ldots, \mathbf{X_n})'$ so that $\mathbf{x} \sim N(\mu \mathbf{1}_n, \sigma^2 \mathbf{I}_n)$. $S^2 = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A} = \frac{\mathbf{I}_n - \bar{\mathbf{J}}_n}{n-1}$, and $\bar{X} = \mathbf{B}\mathbf{x}$ where $\mathbf{B} = \mathbf{1}_n'/n$.
We now apply the theorem above:

$$\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = (\mathbf{1}_n'/n)(\sigma^2 \mathbf{I}_n)(\frac{\mathbf{I}_n - \bar{\mathbf{J}}_n}{n-1}) = (\frac{\sigma^2}{n(n-1)})(\mathbf{1}_n' - \mathbf{1}_n') = \mathbf{0}.$$

Therefore, S^2 and $\bar{X}$ are independently distributed.