STA437/2005 Methods for Multivariate Data

Gun Ho Jang

Lecture on September 15, 2014

Matrix Algebra

Definition. A $k \times k$ matrix A is *non-negative definite* if and only if $\mathbf{v}^\top A\mathbf{v} \geq 0$ for any $\mathbf{v} \in \mathbb{R}^k$. A $k \times k$ matrix A is positive definite if and only if $\mathbf{v}^\top A\mathbf{v} > 0$ for any $\mathbf{v} \in \mathbb{R}^k \backslash \{0\}$.

Definition. *Eigen values* are solution to $|A - \lambda I| = 0$. For any eigen value λ, there exists an *eigen vector* $\mathbf{v} \neq 0$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

It is possible to be many eigen vectors for a eigen value.

Theorem (Spectral decomposition)**.** Let A be a symmetric $k \times k$ matrix. Then there exist k orthonormal eigen vectors $e_1, \ldots, e_k$ and corresponding eigen values $\lambda_1, \ldots, \lambda_k$ so that

$$A = \lambda_1 e_1 e_1^\top + \cdots + \lambda_k e^k e_k^\top.$$

A sketch proof. Let $\mathbf{e} = (e_1 \ \ldots \ e_k)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$. From the orthonormality, $\mathbf{e}^\top \mathbf{e} = (e_i^\top e_j) = I_k$. The uniqueness of inverse implies $\mathbf{e}\mathbf{e}^\top = I_k$. Since e_i's are eigen vectors, $A\mathbf{e} = (\lambda_1 e_1 \ \ldots \ \lambda_k e_k) = \mathbf{e}\Lambda$. Then $A = (\mathbf{e}\Lambda)\mathbf{e}^{-1} = \mathbf{e}\Lambda\mathbf{e}^\top = \lambda_1 e_1 e_1^\top + \cdots + \lambda_k e_k e_k^\top$. $\square$

Expectation of Random Matrix

For a $n \times p$ random matrix $\mathbf{X} = (X_{ij})$, the expectation is defined by

$$\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_{ij})) = \begin{pmatrix} \mathbb{E}(X_{11}) & \mathbb{E}(X_{12}) & \cdots & \mathbb{E}(X_{1p}) \\ \mathbb{E}(X_{21}) & \mathbb{E}(X_{22}) & \cdots & \mathbb{E}(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}(X_{n1}) & \mathbb{E}(X_{n2}) & \cdots & \mathbb{E}(X_{np}) \end{pmatrix}.$$

Proposition. Let $\mathbf{X}, \mathbf{Y}$ be two $n \times p$ random matrices and A, B be two conformable matrices. Then,
(a) $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$
(b) $\mathbb{E}(A\mathbf{X}B) = A\mathbb{E}(\mathbf{X})B$.

Proof. (a) Let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ so that $Z_{ij} = X_{ij} + Y_{ij}$. Then $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{Z}) = (\mathbb{E}(Z_{ij})) = (\mathbb{E}(X_{ij} + Y_{ij})) = (\mathbb{E}(X_{ij}) + \mathbb{E}(Y_{ij})) = (\mathbb{E}(X_{ij})) + (\mathbb{E}(Y_{ij})) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$.
(b) Let $\mathbf{W} = A\mathbf{X}B$. Then $W_{kl} = \sum_{i=1}^n \sum_{j=1}^p A_{ki} X_{ij} B_{jl}$ and $\mathbb{E}(\mathbf{W}) = (\mathbb{E}(W_{kl})) = (\sum_{i=1}^n \sum_{j=1}^p A_{ki} \mathbb{E}(X_{ij}) B_{jl}) = ([A\mathbb{E}(\mathbf{X})B]_{kl}) = A\mathbb{E}(\mathbf{X})B$. $\square$

Random Vector

Random vector $X = (x_1, \dots, x_n)^\top$ is a $n \times 1$ random matrix. Hence the mean of X is defined by

$$\mathbb{E}(X) = (\mathbb{E}(x_i)) = \begin{pmatrix} \mathbb{E}(x_1) \\ \mathbb{E}(x_2) \\ \vdots \\ \mathbb{E}(x_n) \end{pmatrix}.$$

The *variance* of X is defined by the variance-covariance matrix, that is,

$$\mathbb{V}ar(X) = (\mathrm{Cov}(x_i, x_j)) = (\mathbb{E}((x_i - \mathbb{E}(x_i))(x_j - \mathbb{E}(x_j)))) = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top].$$

In general the *covariance* of two random vectors X and Y is defined by

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

Proposition. Let X, Y be two $n \times 1$ random vectors. Then,
(a) for $a, b \in \mathbb{R}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\mathrm{Cov}(aX + \mathbf{v}, bY + \mathbf{w}) = ab\mathrm{Cov}(X, Y)$,
(b) The mean and variance of $Z = AX$ for a matrix $A \in \mathbb{R}^{k \times n}$ are $\mathbb{E}(Z) = A\mathbb{E}(X)$ and $\mathbb{V}ar(Z) = A\mathbb{V}ar(X)A^\top$.

Proof. (a) $\mathrm{Cov}(aX + \mathbf{v}, bY + \mathbf{w}) = \mathbb{E}((aX + \mathbf{v} - \mathbb{E}(aX + \mathbf{v}))(bY + \mathbf{w} - \mathbb{E}(bY + \mathbf{w}))) = ab\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^\top) = ab\mathrm{Cov}(X, Y)$.
(b) $\mathbb{E}(Z) = \mathbb{E}(AX) = A\mathbb{E}(X)$ and $\mathbb{V}ar(Z) = \mathbb{E}(ZZ^\top) - \mathbb{E}(Z)\mathbb{E}(Z)^\top = \mathbb{E}(AXX^\top A^\top) - A\mathbb{E}(X)\mathbb{E}(X)^\top A^\top = A\mathbb{V}ar(X)A^\top$. □

Partition

Let X be a $n \times 1$ random vector. Consider a partition $X = ({X^{(1)}}^\top, {X^{(2)}}^\top)^\top)$, that is, for some $k, l > 0$ with $k + l = n$, $X^{(1)} = (X_1, \dots, X_k)^\top$ and $X^{(2)} = (X_{k+1}, \dots, X_{k+l})^\top$.

Proposition. The mean and variance of the partition becomes

$$\mathbb{E}(X) = \mathbb{E}\begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X^{(1)}) \\ \mathbb{E}(X^{(2)}) \end{pmatrix}$$
$$\mathbb{V}ar(X) = \begin{pmatrix} \mathrm{Cov}(X^{(1)}, X^{(1)}) & \mathrm{Cov}(X^{(1)}, X^{(2)}) \\ \mathrm{Cov}(X^{(2)}, X^{(1)}) & \mathrm{Cov}(X^{(2)}, X^{(2)}) \end{pmatrix}$$

Proof. Definitions and partition gives

$$\mathbb{E}(X) = \begin{pmatrix} \mathbb{E}(x_1) \\ \vdots \\ \mathbb{E}(x_k) \\ \mathbb{E}(x_{k+1}) \\ \vdots \\ \mathbb{E}(x_{k+l}) \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X^{(1)} \\ \mathbb{E}(X^{(2)}) \end{pmatrix}$$

Similarly

$$\mathbb{V}ar(X) = (\mathrm{Cov}(x_i, x_j)) = \begin{pmatrix} \mathrm{Cov}(x_1, x_1) & \mathrm{Cov}(x_1, x_2) & \cdots & \mathrm{Cov}(x_1, x_n) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Cov}(x_2, x_2) & \cdots & \mathrm{Cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(x_n, x_1) & \mathrm{Cov}(x_n, x_2) & \cdots & \mathrm{Cov}(x_n, x_n) \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(X^{(1)}, X^{(1)}) & \mathrm{Cov}(X^{(1)}, X^{(2)}) \\ \mathrm{Cov}(X^{(2)}, X^{(1)}) & \mathrm{Cov}(X^{(2)}, X^{(2)}) \end{pmatrix}$$

Hence the mean and variance can be computed blockwise. □

Exercise. Let Σ be a symmetric positive definite matrix with partition $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Express Σ^{-1} using Σ_{ij}'s.

Random Sample

The p measurements from each sample are supposed to be independent. In other words, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is independent and identically distributed.

The sample mean $\bar{\mathbf{x}} = (\mathbf{x}_1 + \cdots + \mathbf{x}_n)/n$ is unbiased estimator of $\mu = \mathbb{E}(\mathbf{x}_i)$ and its covariance matrix is Σ/n where $\Sigma = \mathbb{V}ar(\mathbf{x}_i)$. The sample covariance matrix $S_n = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top/n$ is a consistent estimator of Σ with bias $-\Sigma/n$. Let $S = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})/(n-1)$ so that S is an unbiased estimator for Σ.

$$
\begin{aligned}
\mathbb{E}(\bar{\mathbf{x}}) &= \mathbb{E}(\sum_{i=1}^n \mathbf{x}_i/n) = \sum_{i=1}^n \mathbb{E}(\mathbf{x}_i)/n = n\mathbb{E}(\mathbf{x}_1)/n = \mathbb{E}(\mathbf{x}_1) = \mu. \\
\mathbb{V}ar(\bar{x}) &= \mathbb{V}ar(\sum_{i=1}^n x_i/n) = n^{-2} \sum_{i=1}^n \mathbb{V}ar(\mathbf{x}_i) = \mathbb{V}ar(\mathbf{x}_1)/n = \Sigma/n. \\
\mathbb{E}(S_n) &= \frac{1}{n}\mathbb{E}(\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top) = \mathbb{E}((\mathbf{x}_1 - \bar{\mathbf{x}})(\mathbf{x}_1 - \bar{\mathbf{x}})^\top) = \mathrm{Cov}(\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_1 - \bar{\mathbf{x}}) \\
&= \mathbb{V}ar(\mathbf{x}_1) - \mathrm{Cov}(\mathbf{x}_1, \bar{\mathbf{x}}) - \mathrm{Cov}(\bar{\mathbf{x}}, \mathbf{x}_1) + \mathbb{V}ar(\bar{\mathbf{x}}) = \Sigma - \Sigma/n - \Sigma/n + \Sigma/n = \Sigma(1 - 1/n). \\
\mathbb{E}(S) &= \mathbb{E}[\frac{n}{n-1} S_n] = \frac{n}{n-1}\Sigma\frac{n-1}{n} = \Sigma.
\end{aligned}
$$

Generalized Variance: Variance-covariance matrix contains $p \times p$ elements which is big to consider simultaneously. Simplified variance might be useful in interpretation.
Suggestion: $|S|$, the determinant of unbiased variance-covariance matrix.
Note: $|S| = (n-1)^{-p}(\text{volume})^2$
Note: If $|S| = 0$, then there exists a linear relationship between variables.

Matrix form:

$$
\bar{\mathbf{x}} = \mathbf{X}^\top(\frac{1}{n}\mathbf{1}), \quad S = \frac{1}{n-1}\mathbf{X}^\top(I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top)\mathbf{X}.
$$

Multivariate Normal Distribution

Normal density: $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$

Let $z_1, \ldots, z_k \sim i.i.d.\ N(0,1)$. then the joint density of $Z = (z_1, \ldots, z_k)$ is

$$
\mathrm{pdf}_Z(\mathbf{z}) = \prod_{i=1}^k (2\pi)^{-1/2}\exp(-z_i^2/2) = |2\pi I_k|^{-1/2}\exp(-\frac{1}{2}\mathbf{z}^\top\mathbf{z}).
$$

The density of $X = \mu + \Sigma^{1/2}Z$ can be obtained using the change of variable formula, that is,

$$
\begin{aligned}
\mathrm{pdf}_X(\mathbf{x}) &= \mathrm{pdf}_Z(\Sigma^{-1/2}(\mathbf{x} - \mu)) \cdot \left|\frac{\partial \mathbf{x}}{\partial \mathbf{z}}\right| = |2\pi I_k|^{-1/2}\exp(-\frac{1}{2}(\Sigma^{-1/2}(\mathbf{x} - \mu))^\top(\Sigma^{-1/2}(\mathbf{x} - \mu))) \times |\Sigma^{-1/2}| \\
&= |2\pi I_k|^{-1/2}|\Sigma|^{-1/2}\exp(-\frac{1}{2}(\mathbf{x} - \mu)^\top\Sigma^{-1}(\mathbf{x} - \mu)).
\end{aligned}
$$

Proposition. If $X \sim N(\mu, \Sigma)$, then for a conformable matrix A, $AX \sim N(A\mu, A\Sigma A^\top)$.

Proof. Note that $X = \mu + \Sigma^{1/2}Z$ implies $AX = A(\mu + \Sigma^{1/2}Z) = A\mu + A\Sigma^{1/2}Z$. Hence AX is a normal distribution with mean $A\mu$ and variance $A\Sigma^{1/2}(A\Sigma^{1/2})^\top = A\Sigma A^\top$. □

Proposition. If $X \sim N(\mu, \Sigma)$ with $|\Sigma| > 0$ and $k = \mathrm{rank}(\Sigma)$, then $(X-\mu)^\top \Sigma^{-1}(X-\mu) \sim \chi^2(k)$.

Proof. Note $X = \mu + \Sigma^{1/2}Z$ for $Z \sim N(O, I_k)$. Then $(X-\mu)^\top\Sigma^{-1}(X-\mu) = (\mu+\Sigma^{1/2}Z-\mu)^\top\Sigma^{-1/2}\Sigma^{-1/2}(\mu+\Sigma^{1/2}Z-\mu) = Z^\top Z = Z_1^2 + \cdots + Z_p^2 \sim \chi^2(p)$. □

Using this result, the ellipsoid $C_\gamma = \{\mathbf{x} : (\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu) \le \chi^2_\gamma(p)\}$ has probability γ from $N(\mu, \Sigma)$ for $0 < \gamma < 1$

Exercise. If $x_j \sim N(\mu_j, \Sigma_j)$ are independent, then $x_1 + \cdots + x_k \sim N(\mu_1 + \cdots + \mu_k, \Sigma_1 + \cdots + \Sigma_k)$.

Proposition. The moment generating function of $X \sim N(\mu, \Sigma)$ is $\mathrm{mgf}_X(\mathbf{t}) = \exp(\mathbf{t}^\top\mu + \mathbf{t}^\top\Sigma\mathbf{t}/2)$.

Proof.

$$\mathrm{mgf}_X(\mathbf{t}) = \mathbb{E}[\exp(\mathbf{t}^\top X)] = \mathbb{E}[\exp(\mathbf{t}^\top(\mu + \Sigma^{1/2}Z))] = \exp(\mathbf{t}^\top\mu)\mathrm{mgf}_Z((\mathbf{t}^\top\Sigma^{1/2})^\top).$$

Since $z_1, \ldots, z_k$ are independent, for $\mathbf{u} = (\mathbf{t}^\top\Sigma^{1/2})^\top = \Sigma^{1/2}\mathbf{t}$,

$$\mathrm{mgf}_Z(\mathbf{u}) = \prod_{i=1}^k \exp(u_i^2/2) = \exp(\mathbf{u}^\top\mathbf{u}/2) = \exp((\Sigma^{1/2}\mathbf{t})^\top(\sigma^{1/2}\mathbf{t})/2) = \exp(\mathbf{t}^\top\Sigma\mathbf{t}/2).$$

Finally the moment generating function of X is

$$\mathrm{mgf}_X(\mathbf{t}) = \exp(\mathbf{t}^\top\mu + \mathbf{t}^\top\Sigma\mathbf{t}/2).$$

□

Proposition. Suppose that $X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N(0, \Sigma)$ with $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then,
(a) $X^{(i)} \sim N(\mu_i, \Sigma_{ii})$.
(b) If $\Sigma_{12} = O$, then $X^{(1)}$ and $X^{(2)}$ are independent.
(c) $X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ and $X^{(2)}$ are independent.
(d) $X^{(1)} \mid X^{(2)} = x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11\cdot 2})$ where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof. (a) Let $\mathbf{t} = (\mathbf{t}_1^\top, O_{1\times l})^\top$. Then $\mathrm{mgf}_{X^{(1)}}(\mathbf{t}_1) = \mathbb{E}[\exp(\mathbf{t}_1^\top X^{(1)})] = \mathbb{E}[\exp(\mathbf{t}^\top X)] = \mathrm{mgf}_X(\mathbf{t}) = \exp(\mathbf{t}^\top\mu + \mathbf{t}^\top\Sigma\mathbf{t}/2) = \exp(\mathbf{t}_1^\top\mu_1 + \mathbf{t}_1^\top\Sigma_{11}\mathbf{t}_1/2)$. Thus $X^{(1)} \sim N(\mu_1, \Sigma_{11})$. Similarly, $X^{(2)} \sim N(\mu_2, \Sigma_{22})$.
(b) Let $\mathbf{t} = (\mathbf{t}_1^\top, \mathbf{t}_2^\top)^\top$. Then

$$\begin{aligned}
\mathrm{mgf}_{X^{(1)},X^{(2)}}(\mathbf{t}_1, \mathbf{t}_2) &= \mathrm{mgf}_X(\mathbf{t}) = \exp(\mathbf{t}^\top\mu + \mathbf{t}^\top\Sigma\mathbf{t}/2) \\
&= \exp(\mathbf{t}_1^\top\mu_1 + \mathbf{t}_2^\top\mu_2 + \mathbf{t}_1^\top\Sigma_{11}\mathbf{t}_1/2 + \mathbf{t}_1^\top\Sigma_{12}\mathbf{t}_2/2 + \mathbf{t}_2^\top\Sigma_{21}\mathbf{t}_1/2 + \mathbf{t}_2^\top\Sigma_{22}\mathbf{t}_2/2) \\
&= \exp(\mathbf{t}_1^\top\mu_1 + \mathbf{t}_1^\top\Sigma_{11}\mathbf{t}_1/2) \times \exp(\mathbf{t}_2^\top\mu_2 + \mathbf{t}_2^\top\Sigma_{22}\mathbf{t}_2/2) = \mathrm{mgf}_{X^{(1)}}(\mathbf{t}_1) \times \mathrm{mgf}_{X^{(2)}}(\mathbf{t}_2).
\end{aligned}$$

Hence $X^{(1)}$ and $X^{(2)}$ are independent if and only if $\Sigma_{12} = O$.
(c) The covariance between $X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ and $X^{(2)}$ is

$$\mathrm{Cov}(X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}, X^{(2)}) = \mathrm{Cov}(X^{(1)}, X^{(2)}) - \Sigma_{12}\Sigma_{22}^{-1}\mathrm{Cov}(X^{(2)}, X^{(2)}) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = \Sigma_{12} - \Sigma_{12} = O.$$

Hence $X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ and $X^{(2)}$ are independent. (d) From (c), $X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ is independent from $X^{(2)}$ and normally distributed with mean $\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2$ and variance $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. If $X^{(2)} = \mathbf{x}_2$ is given, $X^{(1)} \mid X^{(2)} = \mathbf{x}_2 \equiv^d X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11\cdot 2})$. □