STAT7017 Big Data Statistics	**Semester 2 2018**

Lecture 2: 30 July

Lecturer: Dr Dale Roberts *Scribes: Rui Qiu*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Matrix analysis, Eigenvalues & eigenvectors, and Multivariate Normal distribution.

We denote a set of p random variables $X_i : \Omega \to \mathbb{R},\ i = 1, \cdots, p$ by the (vector-valued) random variable $\mathbb{X} : \Omega \to \mathbb{R}^p$.

$$\mathbb{X}_i = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$$

call it a "random vector".

The mean or expectation of $\mathbb{X}$ is given by

$$\mathbb{E}[\mathbb{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_p] \end{bmatrix} = \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

We always make the assumption that $\mu < \infty$.

A typical measurement involves taking n random samples $\{\mathbb{X}_1, \mathbb{X}_2, \cdots, \mathbb{X}_n\}$.

We can express this in matrix form:

$$\mathbb{X} = (\mathbb{X}_1, \mathbb{X}_2, \cdots, \mathbb{X}_n)' = \begin{pmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \\ \vdots \\ \mathbb{X}_n \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}$$

Notation: transpose of vector $\mathbf{x}$ and matrix A are denoted $\mathbf{x}'$ and A', respectively.

A matrix of random variables is called a random matrix.

Expectation of a random matrix $\mathbb{X} = (X_{ij})$ is given by

$$\mathbb{E}[\mathbb{X}] = (\mathbb{E}(X_{ij}))$$

Lemma 2.1 *Let $\mathbb{X} = (X_{ij})$ and $\mathbb{Y} = (Y_{ij})$ be $n \times p$ random matrices. If A, B and C are constant matrices, then:*

$$\mathbb{E}[\mathbb{X} + \mathbb{Y}] = \mathbb{E}[\mathbb{X}] + \mathbb{E}[\mathbb{Y}]. \tag{2.1}$$

$$\mathbb{E}[A\mathbb{X}B + C] = A\mathbb{E}[\mathbb{X}]B + C. \tag{2.2}$$

Proof: Choosing an arbitrary (i, j)'th element, LHS of (2.1) is

$$\mathbb{E}[X_{ij} + Y_{ij}] = \mathbb{E}[X_{ij}] + \mathbb{E}[Y_{ij}]$$

which is just the RHS of (2.1). Since (i, j) was arbitrary, it holds for all i, j.

In the same way, (i, j)'th element of LHS of (2.2) is

$$\mathbb{E}\left[\sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}x_{kl}b_{lj} + c_{ij}\right] = \sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}\mathbb{E}[\mathbb{X}_{kl}]b_{lj} + c_{ij}$$

which is the RHS of (2.2). ■

<u>**Note:**</u> $A = (a_{ij}) \in \mathbb{R}^{l\times m}$ & $B = (b_{jk}) \in \mathbb{R}^{m\times n}$; then

$$(AB)_{ik} = \sum_{j=1}^{m} a_{ij}b_{jk},\ i = 1, \ldots, l;\ k = 1, \ldots, n.$$

If a $p \times 1$ random vector $\mathbb{X} = (X_1, X_2, \ldots, X_p)'$ has mean $\mu = (\mu_1, \ldots, \mu_p)'$, the <u>covariance matrix of $\mathbb{X}$</u> is defined by

$$\Sigma = Var(\mathbb{X}) = \mathbb{E}\left[(\mathbb{X} - \mu)(\mathbb{X} - \mu)'\right].$$

If a $q\times 1$ random vector $\mathbb{Y} = (Y_1, Y_2, \ldots, Y_q)'$ has mean $\eta = (\eta_1, \eta_2, \ldots, \eta_q)'$, the <u>covariance matrix of $\mathbb{X}$ and $\mathbb{Y}$</u> is defined by

$$Cov(\mathbb{X}, \mathbb{Y}) = \mathbb{E}\left[(\mathbb{X} - \mu)(\mathbb{Y} - \eta)'\right].$$

In particular, $Cov(\mathbb{X}, \mathbb{X}) = Var(\mathbb{X})$.

Elementwise we have $\Sigma = (\sigma_{ij})$ with

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

The covariance between X_i and X_j.

$$\sigma_{ii} = \mathbb{E}[(X_i - \mu_i)^2] = Var(X_i).$$

We write $\sigma_i^2 = \sigma_{ii}$.

Theorem 2.2 *Let Σ be the covariance matrix of a $p \times 1$ random vector $\mathbb{X}$.*

1. *Σ is positive semidefinite (non-negative definite), that is, for any $p \times 1$ fixed vector $x = (x_1, \ldots, x_p)'$,*

$$x'\Sigma x = \sum_{i=1}^{p}\sum_{j=1}^{p} \sigma_{ij} x_i x_j \geq 0.$$

2. *Let B be a $q \times p$ constant matrix and b be a $q \times 1$ constant vector. Then the covariance matrix $\mathbb{Y} = B\mathbb{X} + b$ is*

$$Var(\mathbb{Y}) = B\Sigma B'.$$

Proof:

1. For any $x \in \mathbb{R}^p$, we have

$$\begin{aligned} Var(x'\mathbb{X}) &= \mathbb{E}[(x'\mathbb{X} - x'\mu)(x'\mathbb{X} - x'\mu)'] \\ &= \mathbb{E}[(x'(\mathbb{X} - \mu))(x'(\mathbb{X} - \mu))'] \\ &= \mathbb{E}[x'(\mathbb{X} - \mu)(\mathbb{X} - \mu)'x] \\ &= x'\mathbb{E}[(\mathbb{X} - \mu)(\mathbb{X} - \mu)']x \\ &= x'\Sigma x. \end{aligned} \tag{2.3}$$

   and we also know that $Var(x'\mathbb{X}) \geq 0$, so the result follows.

2. As $\mathbb{Y} - \mathbb{E}[\mathbb{Y}] = (B\mathbb{X} + b) - (B\mu + b) = B(\mathbb{X} - \mu)$.
   We have

$$\begin{aligned} Var(\mathbb{Y}) &= \mathbb{E}[B(\mathbb{Y} - \mu)(B(\mathbb{Y} - \mu))'] \\ &= \mathbb{E}[B(\mathbb{X} - \mu)(\mathbb{X} - \mu)'B'] \\ &= B\mathbb{E}[(\mathbb{X} - \mu)(\mathbb{X} - \mu)']B' \\ &= B\Sigma B' \end{aligned} \tag{2.4}$$

■

In general, the covariance matrix is positive semidefinite.

We call a matrix A positive definite if for all $x \neq 0$,

$$x'Ax > 0.$$

Recall: a complex number is a number of the form $a + bi$ where i satisfies $i^2 = -1$. We write $Re(a + bi) = a,\ Im(a + bi) = b$.

The complex conjugate of a complex number $z = a + bi$ is given by $\bar{z} = a - bi$. $z \in \mathbb{C}$: "space of complex numbers."

If A is a $m \times n$ matrix with complex entires, then the $n \times m$ matrix A^* is obtained by taking the transpose followed by the complex conjugate of each entry.

$$(A^*)_{ij} = \overline{A_{ij}} \text{ or } A^* = \left(\overline{A}\right)' = \overline{A'}.$$

Example:

$$A = \begin{bmatrix} 1 & -3-i \\ 1+2i & 6i \end{bmatrix},\ A^* = \begin{bmatrix} 1 & 1-2i \\ -3+i & -6i \end{bmatrix}$$

The matrix A^* is called the **conjugate transpose** of A.

Properties:

1. $(A+B)^* = A^* + B^*$, if A, B have the same dimensions.
2. $(rA)^* = \bar{r}A^*$, for $r \in \mathbb{C}$ and matrix A.
3. $(AB)^* = B^*A^*,\ A \in \mathbb{R}^{m\times n},\ B \in \mathbb{R}^{n\times p}$.
4. $(A^*)^* = A,\ A \in \mathbb{R}^{n\times n}$.
5. If $A \in \mathbb{R}^{n\times n}$, then $\det(A^*) = \overline{\det A}$ & $tr(A^*) = \overline{(trA)}$

Classes of matrices

A Hermitian matrix A is a square matrix that satisfies $A = A^*$. That means if if $A = (a_{ij})$ then $a_{ij} = \overline{a_{ji}}$.

The nice thing about Hermitian matrices is that they behave a bit like real numbers. Arbitrary square matrices behave like complex numbers (i.e., they can have some quirky behaviour).

Sometimes I will write $A \in \mathbb{M}_p$ to denote that A is a square matrix of size $p \times p$ and if A is Hermitian, I will write $A \in \mathbb{H}_p$.

Notice: $\mathbb{H}_p \subseteq \mathbb{M}_p$.

We can define the (Frobenius) norm of a matrix $A \in \mathbb{M}_p$ as

$$\|A\|_F = \sqrt{\sum_{j=1}^{p}\sum_{k=1}^{p} |a_{jk}|^2},\ A \in \mathbb{M}_p$$

Eigenvalues and Eigenvectors

If $A \in \mathbb{M}_p$ and if $Ae = \lambda e$ and $e \neq 0$ where $e \in \mathbb{R}^p$, then λ and e are called an eigenvalue and an eigenvector of A.

Example:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},\ e = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$Ae = 3e$, so e is eigenvector, $\lambda = 3$. f is not an eigenvector.

If $A \in \mathbb{M}_p$ and $\lambda = (\lambda_1, \ldots, \lambda_p)$, "eigenvalues"; $E = (e_1, e_2, \ldots, e_p) \in \mathbb{M}_p$, "eigenvectors".

Then $AE = \lambda E \iff |A - \lambda I| = 0$ where $|A| = \det(A)$ and E satisfies $(A - \lambda I)E = 0$.

The expression $p(\lambda) = |A - \lambda I|$ is called the characteristic polynomial for A. The equation $p(\lambda) = 0$ is the characteristic equation for A.

If $B \in \mathbb{M}_p$ then the trace is the sum of the diagonal entries, i.e.

$$trB := \sum_{k=1}^{p} b_{kk}, \;\; B \in \mathbb{M}_p.$$

Proposition: If $A \in \mathbb{M}_p$ then $P(\lambda) = |A - \lambda I|$ is a polynomial in λ of degree p and

$$P(\lambda) = \lambda^p - a_1 \lambda^{p-1} + \cdots + (-1)^p a_p$$

where

$$\begin{aligned} a_1 &= tr(A) \\ a_p &= \det(A) \\ a_i &= \text{“sum of i-rowed diagonal mirrors of } A\text{”} \end{aligned} \tag{2.5}$$

E.g. $p = 2, P_2(\lambda) = \lambda^2 - tr(A) + \det(A)$.

Further: $P(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_p - \lambda)$ and $|A| = \prod_{i=1}^{p} \lambda_i$.

Example:

$$B = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix}$$

$$\begin{aligned} |B - \lambda I| &= \begin{vmatrix} 5-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} \\ &= 25 - 4 - 10\lambda + \lambda^2 \\ &= \lambda^2 - 10\lambda + 21 \end{aligned} \tag{2.6}$$

A symmetric matrix is a square matrix that is equal to its transpose, i.e,

$$A = A'.$$

We denote $A \in \mathbb{S}_p$ and $\mathbb{S}_p \subseteq \mathbb{H}_p$.

Proposition: Let $A \in \mathbb{S}_p$, then

1. The characteristic roots $\lambda_1, \ldots, \lambda_p$ are all real. ($\lambda_1, \ldots, \lambda_p$ satisfy $P(\lambda_i) = 0$.

2. If λ_i and λ_j are two distinct characteristic roots of A, the corresponding characteristic vectors e_i and e_j are orthogonal.

Proposition: Let $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$ be the characteristic roots of a matrix $A \in \mathbb{S}_p$. Then

1. $A > 0$ ("positive definite") $\iff \lambda_i > 0,\ i = 1 \ldots, p$.
2. $A \geq 0$ ("positive semidefinite") $\iff \lambda_i \geq 0,\ i = 1, \ldots, p$.

Given a covariance matrix Σ (positive semidefinite), it follows that the characteristic roots (eigenvalues) are non-negative and these are denoted by

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0.$$

Let γ_i be the characteristic vector (eigenvector) corresponding to λ_i for $i = 1, \ldots, p$. WLOG, we assume they are orthonormal, i.e.,

$$\gamma_i' \gamma_j = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j. \end{cases}$$

The characteristic roots and vectors satisfy

$$\sum \gamma_i = \lambda_i \gamma_i,\ i = 1, \ldots, p.$$

The relationship $\sum \lambda_i = \lambda_i \gamma_i$ can be expressed as

$$\Sigma \Lambda = \Gamma \Lambda. \tag{2.7}$$

where $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ and $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$.

We assume that the matrix Γ is normalized as $\Gamma'\Gamma = I_p$, then the equation (2.7) implies that

$$\begin{aligned} \Sigma &= \Gamma \Lambda \Gamma' \\ &= \Gamma \Lambda^{1/2} \Lambda^{1/2} \Gamma' \\ &= \Gamma \Lambda^{1/2} \Gamma' \Gamma \Lambda^{1/2} \Gamma' \end{aligned} \tag{2.8}$$

where $\Lambda^{1/2} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})$.

Define $\Sigma^{1/2} = \Gamma \Lambda^{1/2} \Gamma'$ and $C = \Gamma \Lambda^{1/2}$, then

$$\Sigma = (\Sigma^{1/2})^2 = CC'.$$

Note: Here you should be drawing an analogy to variance vs. standard deviation in univariate case as $\sigma^2 = \sigma\sigma$.

The covariance matrix Σ contains the variance of the p variables and the covariances between them. It is desirable to have a measure of "scatter".

Two possibilities are:

Generalized variance given in terms of the determinant.

$$|\Sigma| = |\Gamma\Lambda\Gamma'| = |\Lambda| = \lambda_1 \cdot \lambda_2 \cdots \lambda_p. \tag{2.9}$$

Total variance given in terms of trace.

$$\begin{aligned} tr\Sigma &= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} \\ &= tr(\Gamma\Lambda\Gamma') \\ &= tr(\Lambda\Gamma'\Gamma) \\ &= tr\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_p. \end{aligned} \tag{2.10}$$

Note: Later we will actually consider the term in (2.9) as

$$GV(?) = \frac{1}{p}\log|\Sigma| = \frac{1}{p}\sum_{k=1}^{p}\log(\lambda_k).$$

Characteristic functions

Let $\mathbb{X}$ be a p-dimensional random vector, then

$$P(\mathbb{X} \in E) = \int_E f(\mathbf{x})d\mathbf{x}$$

where $d\mathbf{x} = dx_1 dx_2 \cdots dx_p$. $E \subset \mathbb{R}^p$.

The function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_p)$ is called the density of $\mathbb{X}$.

The characteristic function of $\mathbb{X}$ is

$$C(\theta) = \mathbb{E}\left[e^{i\Theta'\mathbb{X}}\right]$$

where $i := \sqrt{-1}, \Theta := (\theta_1, \theta_2, \ldots, \theta_p)', \theta_i \in \mathbb{R}, i = 1, \ldots, p.$

Theorem 2.3 *There exists a one-to-one correspondence between the distribution of $\mathbb{X}$ and its characteristic function.*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^p}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-i\Theta'\mathbf{x}}C(\Theta)d\theta_1\cdots d\theta_p.$$

You can use the characteristic function to obtain various moments of $\mathbb{X}$.

$$\frac{\partial^m}{\partial_{\theta_1}^{m_1}\partial_{\theta_2}^{m_2}\cdots\partial_{\theta_p}^{m_p}}C(\Theta)=\mathbb{E}\left[(iX_1)^{m_1}(iX_2)^{m_2}\cdots(iX_p)^{m_p}e^{i\Theta'\mathbb{X}}\right]$$

where $m = m_1 + m_2 + \cdots + m_p$.

Taking $\Theta = \mathbf{0}$, we can get the moment $\mathbb{E}[X_1^{m_1}X_2^{m_2}\cdots X_p^{m_p}]$.

Multivariate Normal Distribution

The simplest case is the bivariate normal distribution.

$$\mathbb{X} = (X_1, X_2)$$

$$P(\mathbb{X} \in E) = P(a < X_1 \le b, c < X_2 \le d) = \int_a^b \int_c^d f(x_1, x_2) dx_1 dx_2.$$

density given by

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-p^2}}\exp\left(\frac{-1}{2(1-p^2)}\Theta\right)$$

and

$$\Theta(x_1,x_2) = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2p\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2.$$

$\mathbb{X} \sim$ Bivariate Normal.

$$\mathbb{E}[X_i] = \mu_i,\ Var[X_i] = \sigma_i^2,\ i = 1, 2.$$

covariance and correlation between X_1 and X_2 are

$$Cov(X_1, X_2) = \rho\sigma_1\sigma_2,\ Cor(X_1, X_2) = \rho.$$

Covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

Let's construct the higher-dimensional case. ($p > 2$).

Recall: $Z \sim N(0,1), p(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, -\infty < z < \infty$.

If $X \sim N(\mu, \sigma^2)$ then $X = \mu + \sigma Z$ and

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}.$$

Take $\mathbb{Z} = (Z_1, Z_2, \ldots, Z_p)'$, $Z_i \sim (N, 0, 1)$ iid then $\mathbb{Z}$ has density

$$\prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^p e^{-\mathbf{z}'\mathbf{z}/2}, \ \mathbf{z} = (z_1, z_2, \ldots, z_p)'.$$

Now consider $\mathbb{E}[\mathbb{X}] = \mu, Cov[\mathbb{X}] = \Sigma$.

Lemma 2.4 *Let $\mathbf{f} = \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_p(\mathbf{x}))'$ be a transformation such that partial derivatives $\partial f_i / \partial x_j$ exist. Then the determinant of the matrix*

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_p} \end{pmatrix}$$

is called the <u>Jacobian determinant</u> and denoted J.

and some useful cases are:

1. $\mathbf{y} = A\mathbf{x} \implies J = |A|, y \in \mathbb{R}^p, x \in \mathbb{R}^p$.
2. $Y = AX \implies J = |A|^q, \mathbf{X} \in \mathbb{R}^{p \times q}, \mathbf{Y} \in \mathbb{R}^{p \times q}$.
3. $\mathbf{Y} = \mathbf{X}B \implies J = |B|^p, A \in \mathbb{M}_p, B \in \mathbb{M}_q$.
4. $\mathbf{Y} = A\mathbf{X}B \implies J = |A|^q |B|^p$.

Now, returning to multivariate normals, the Jacobian for the transformation from $\mathbb{X}$ to $\mathbb{Z}$ is

$$\Sigma^{-1/2}\mathbb{X} - \Sigma^{-1/2}\mu.$$

and $J = |\Sigma^{-1/2}| = |\Sigma|^{-1/2}$.

Theorem 2.5 *If $\mathbb{X}$ is random vector with density $\phi(\mathbf{x})$ then the density of $\mathbb{Y} = g(\mathbb{X})$ is given by*

$$f(y) = \phi(g^{-1}(y))|J|.$$

Therefore, the density of $\mathbb{X}$ is

$$f(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right) \tag{2.11}$$

We say that $\mathbb{X}$ has p-variate Normal distribution if it has density (2.11).

Theorem 2.6 *The following are equivalent:*

1. $\mathbb{X} \sim N_p(\mu, \Sigma)$, $\Sigma \in \mathbb{M}_p$ *positive definite.*
2. $\mathbb{Z} \sim \Sigma^{-1/2}(\mathbb{X} - \mu) \sim N_p(\mathbf{0}, \mathbf{I}_p)$.

Theorem 2.7 $\mathbb{X} \sim N_p(\mu, \Sigma)$

1. $\mathbb{E}[\mathbb{X}] = \mu,\ Var(\mathbb{X}) = \Sigma$.
2. $C_{\mathbb{X}}(\Theta) = \exp\left(i\mu'\Theta - \frac{1}{2}\Theta'\Sigma\Theta\right)$.
3. $B \in \mathbb{R}^{q \times p}$ $, Rank(B) = q,\ b \in \mathbb{R}^q$,

$$\mathbb{Y} = B\mathbb{X} + b \sim N_q(B\mu + b, B\Sigma B').$$