Background[1]

STA431 Spring 2015

[1] See last slide for copyright information.

Overview

Matrices

- $\mathbf{A} = [a_{ij}]$
- Transpose: $\mathbf{A}^\top = [a_{ji}]$
- Multiplication: $\mathbf{AB} \neq \mathbf{BA}$
- $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$
- Inverse of a *square* matrix: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}$
- $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$
- Positive definite: $\mathbf{v}^\top\mathbf{Av} > 0$ for all $p \times 1$ vectors $\mathbf{v} \neq \mathbf{0}$.

Trace of a square matrix: Sum of the diagonal elements

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{i,i}$$

- Of course $tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$,
- $tr(\mathbf{A}) = tr(\mathbf{A}^\top)$, etc.
- But less obviously, even though $\mathbf{AB} \neq \mathbf{BA}$,
- $tr(\mathbf{AB}) = tr(\mathbf{BA})$

$tr(\mathbf{AB}) = tr(\mathbf{BA})$

Let $\mathbf{A}$ be an $r \times p$ matrix and $\mathbf{B}$ be a $p \times r$ matrix, so that the product matrices $\mathbf{AB}$ and $\mathbf{BA}$ are both defined.

$$
\begin{aligned}
tr(\mathbf{AB}) &= \sum_{i=1}^{r} \left(\sum_{k=1}^{p} a_{i,k} b_{k,i} \right) \\
&= \sum_{k=1}^{p} \left(\sum_{i=1}^{r} b_{k,i} a_{i,k} \right) \\
&= tr(\mathbf{BA})
\end{aligned}
$$

Random vectors

Expected values and variance-covariance matrices

- $E(\mathbf{X}) = [E(X_{i,j})]$
- $E(\mathbf{X}+\mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$
- $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$
- $V(\mathbf{X}) = E\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right\}$
- $V(\mathbf{AX}) = \mathbf{A}V(\mathbf{X})\mathbf{A}^\top$
- $C(\mathbf{X},\mathbf{Y}) = E\left\{(\mathbf{X}-\boldsymbol{\mu}_x)(\mathbf{Y}-\boldsymbol{\mu}_y)^\top\right\}$
- $V(\mathbf{X}+\mathbf{a}) = V(\mathbf{X})$
- $C(\mathbf{X}+\mathbf{a},\mathbf{Y}+\mathbf{b}) = C(\mathbf{X},\mathbf{Y})$

The Centering Rule

Based on $V(\mathbf{X}+\mathbf{a}) = V(\mathbf{X})$

Often, variance and covariance calculations can be simplified by subtracting off constants first.

Denote the *centered* version of $\mathbf{X}$ by $\overset{c}{\mathbf{X}} = \mathbf{X} - E(\mathbf{X})$, so that

- $E(\overset{c}{\mathbf{X}}) = \mathbf{0}$ and
- $V(\overset{c}{\mathbf{X}}) = E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{X}}^\top) = V(\mathbf{X})$

Linear combinations

These are matrices, but they could be scalars

$$
\begin{aligned}
\mathbf{L} &= \mathbf{A}_1\mathbf{X}_1 + \cdots + \mathbf{A}_m\mathbf{X}_m + \mathbf{b} \\
\overset{c}{\mathbf{L}} &= \mathbf{A}_1 \overset{c}{\mathbf{X}}_1 + \cdots + \mathbf{A}_m \overset{c}{\mathbf{X}}_m, \text{ where} \\
\overset{c}{\mathbf{X}}_j &= \mathbf{X}_j - E(\mathbf{X}_j) \text{ for } j = 1, \ldots, m.
\end{aligned}
$$

The centering rule says

$$
\begin{aligned}
V(\mathbf{L}) &= E(\overset{c}{\mathbf{L}}\overset{c}{\mathbf{L}}^\top) \\
C(\mathbf{L}_1, \mathbf{L}_2) &= E(\overset{c}{\mathbf{L}}_1 \overset{c}{\mathbf{L}}_2^\top)
\end{aligned}
$$

In words: To calculate variances and covariances of linear combinations, one may simply discard added constants, center all the random vectors, and take expected values of products.

Example: $V(\mathbf{X}+\mathbf{Y})$

Using the centering rule

$$
\begin{aligned}
V(\mathbf{X}+\mathbf{Y}) &= E(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})^\top \\
&= E(\overset{c}{\mathbf{X}}+\overset{c}{\mathbf{Y}})(\overset{c}{\mathbf{X}}^\top+\overset{c}{\mathbf{Y}}^\top) \\
&= E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{X}}^\top)+E(\overset{c}{\mathbf{Y}}\overset{c}{\mathbf{Y}}^\top)+E(\overset{c}{\mathbf{X}}\overset{c}{\mathbf{Y}}^\top)+E(\overset{c}{\mathbf{Y}}\overset{c}{\mathbf{X}}^\top) \\
&= V(\mathbf{X})+V(\mathbf{Y})+C(\mathbf{X},\mathbf{Y})+C(\mathbf{Y},\mathbf{X})
\end{aligned}
$$

- Does $C(\mathbf{X},\mathbf{Y}) = C(\mathbf{Y},\mathbf{X})$?
- Does $C(\mathbf{X},\mathbf{Y}) = C(\mathbf{Y},\mathbf{X})^\top$?

The Multivariate Normal Distribution

The $p \times 1$ random vector $\mathbf{X}$ is said to have a *multivariate normal distribution*, and we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $\mathbf{X}$ has (joint) density

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu}$ is $p \times 1$ and $\boldsymbol{\Sigma}$ is $p \times p$ symmetric and positive definite.

Σ positive definite

- Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}^\top \mathbf{\Sigma} \mathbf{a} > 0$.
- Since the one-dimensional random variable $Y = \sum_{i=1}^{p} a_i X_i$ may be written as $Y = \mathbf{a}^\top \mathbf{X}$ and $Var(Y) = V(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \mathbf{\Sigma} \mathbf{a}$, it is natural to require that $\mathbf{\Sigma}$ be positive definite.
- All it means is that every non-zero linear combination of $\mathbf{X}$ values has a positive variance.
- And recall $\mathbf{\Sigma}$ positive definite is equivalent to $\mathbf{\Sigma}^{-1}$ positive definite.

Analogies

Multivariate normal reduces to the univariate normal when $p = 1$

- Univariate Normal
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$
  - $E(X) = \mu, Var(X) = \sigma^2$
  - $\frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$
- Multivariate Normal
  - $f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$
  - $E(\mathbf{X}) = \boldsymbol{\mu}$, $V(\mathbf{X}) = \boldsymbol{\Sigma}$
  - $(\mathbf{X}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

More properties of the multivariate normal

- If $\mathbf{c}$ is a vector of constants, $\mathbf{X} + \mathbf{c} \sim N(\mathbf{c} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$
- Linear combinations of multivariate normals are multivariate normal.
- All the marginals (dimension less than p) of $\mathbf{X}$ are (multivariate) normal.
- For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.

Copyright Information