STA 360/602L: Module 4.1

Multivariate normal model I

Dr. Olanrewaju Michael Akande

Multivariate data

- So far we have only considered basic models with scalar/univariate outcomes, $Y_1, \ldots, Y_n$.
- In practice however, outcomes of interest are actually often multivariate, e.g.,
  - Repeated measures of weight over time in a weight loss study
  - Measures of multiple disease markers
  - Tumor counts at different locations along the intestine
- Longitudinal data is just a special case of multivariate data.
- Interest then is often on how multiple outcomes are correlated, and on how that correlation may change across outcomes or time points.

Multivariate normal distribution

- The most common model for multivariate outcomes is the multivariate normal distribution.

- Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^T$, where p represents the dimension of the multivariate outcome variable for a single unit of observation.

- For multiple observations, $\boldsymbol{Y_i} = (Y_{i1}, \ldots, Y_{ip})^T$ for $i = 1, \ldots, n$.

- $\boldsymbol{Y}$ follows a multivariate normal distribution, that is, $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, if

$$p(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right\},$$

where $|\Sigma|$ denotes the determinant of Σ.

Multivariate normal distribution

If $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, then

- $\boldsymbol{\mu}$ is the $p \times 1$ mean vector, that is,

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{Y}] = \{\mathbb{E}[Y_1], \ldots, \mathbb{E}[Y_p]\} = (\mu_1, \ldots, \mu_p)^T.$$

- Σ is the $p \times p$ **positive definite and symmetric** covariance matrix, that is,
  - $\Sigma = \{\sigma_{jk}\}$, where σ_{jk} denotes the covariance between Y_j and Y_k.
- $Y_1, \ldots, Y_p$ may be linearly dependent depending on the structure of Σ, which characterizes the association between them.
- For each $j = 1, \ldots, p$, $Y_j \sim \mathcal{N}(\mu_j, \sigma_{jj})$.

Bivariate normal distribution

- In the bivariate case, we have

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_2 \left[\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} = \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} = \sigma_2^2 \end{pmatrix} \right],$$

and $\sigma_{12} = \sigma_{21} = \mathbb{C}\mathrm{ov}[Y_1, Y_2]$.

- The correlation between Y_1 and Y_2 is defined as

$$\rho_{1,2} = \frac{\mathbb{C}\mathrm{ov}[Y_1, Y_2]}{\sqrt{\mathbb{V}\mathrm{ar}[Y_1]}\sqrt{\mathbb{V}\mathrm{ar}[Y_2]}} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}.$$

- $-1 \le \rho_{1,2} \le 1$.
- Correlation coefficient is free of the measurement units.

Back to the multivariate normal

- There are many special properties of the multivariate normal as we will see as we continue to work with the distribution.

- First, dependence between any Y_j and Y_k does not depend on the other $p-2$ variables.

- Second, while generally, **independence implies zero covariance**, for the normal family, the converse is also true. That is, **zero covariance also implies independence**.

- Thus, the covariance Σ carries a lot of information about marginal relationships, especially **marginal independence**.

- If $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I}_p)$, that is, $\epsilon_1, \ldots, \epsilon_p \overset{iid}{\sim} \mathcal{N}(0,1)$, then

$$\boldsymbol{Y} = \boldsymbol{\mu} + A\boldsymbol{\epsilon} \Rightarrow \ \boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$$

  holds for any matrix square root A of Σ, that is, $AA^T = \Sigma$ (see Cholesky decomposition).

Conditional distributions

- Partition $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^T$ as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}_p \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right],$$

  where

  - $\boldsymbol{Y}_1$ and $\boldsymbol{\mu}_1$ are $q \times 1$,
  - $\boldsymbol{Y}_2$ and $\boldsymbol{\mu}_2$ are $(p-q) \times 1$,
  - Σ_{11} is $q \times q$, and
  - Σ_{22} is $(p-q) \times (p-q)$, with $\Sigma_{22} > 0$.

- Then,

$$\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim \mathcal{N}_q \left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

- Marginal distributions are once again normal, that is,

$$\boldsymbol{Y}_1 \sim \mathcal{N}_q \left(\boldsymbol{\mu}_1, \Sigma_{11} \right); \quad \boldsymbol{Y}_2 \sim \mathcal{N}_{p-q} \left(\boldsymbol{\mu}_2, \Sigma_{22} \right).$$

Conditional distributions

- In the bivariate normal case with

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}_2 \left[\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} = \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} = \sigma_2^2 \end{pmatrix} \right],$$

we have

$$Y_1 | Y_2 = y_2 \sim \mathcal{N} \left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(y_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \right).$$

which can also be written as

$$Y_1 | Y_2 = y_2 \sim \mathcal{N} \left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(y_2 - \mu_2), (1 - \rho^2)\sigma_1^2 \right).$$

Multivariate normal likelihood

- Suppose $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^T \sim \mathcal{N}_p(\boldsymbol{\theta}, \Sigma)$, $i = 1, \ldots, n$.
- Write $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)^T$. The resulting likelihood can then be written as

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\boldsymbol{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\boldsymbol{y}_i - \boldsymbol{\theta}) \right\}$$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\theta})^T \Sigma^{-1} (\boldsymbol{y}_i - \boldsymbol{\theta}) \right\}.$$

- It will be super useful to be able to write the likelihood in two different formulations depending on whether we care about the posterior of $\boldsymbol{\theta}$ or Σ.

Multivariate normal likelihood

- For inference on $\boldsymbol{\theta}$, it is convenient to write $p(\boldsymbol{Y}|\boldsymbol{\theta},\Sigma)$ as

$$
\begin{aligned}
p(\boldsymbol{Y}|\boldsymbol{\theta},\Sigma) &\propto \underbrace{|\Sigma|^{-\frac{n}{2}}}_{\text{does not involve } \boldsymbol{\theta}} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{\theta})^T\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{\theta})\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i^T-\boldsymbol{\theta}^T)\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{\theta})\right\}\\
&= \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\underbrace{\boldsymbol{y}_i^T\Sigma^{-1}\boldsymbol{y}_i}_{\text{does not involve } \boldsymbol{\theta}} - \underbrace{\boldsymbol{y}_i^T\Sigma^{-1}\boldsymbol{\theta}-\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{y}_i}_{\text{same term}} + \boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{\theta}\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{\theta}-2\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{y}_i\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{\theta}-\frac{1}{2}\sum_{i=1}^{n}(-2)\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{y}_i\right\}\\
&= \exp\left\{-\frac{1}{2}n\boldsymbol{\theta}^T\Sigma^{-1}\boldsymbol{\theta}+\boldsymbol{\theta}^T\Sigma^{-1}\sum_{i=1}^{n}\boldsymbol{y}_i\right\}\\
&= \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T(n\Sigma^{-1})\boldsymbol{\theta}+\boldsymbol{\theta}^T(n\Sigma^{-1}\bar{\boldsymbol{y}})\right\},
\end{aligned}
$$

where $\bar{\boldsymbol{y}} = (\bar{y}_1,\ldots,\bar{y}_p)^T$.

Multivariate normal likelihood

- For inference on Σ, we need to rewrite the likelihood a bit.
- First a few results from matrix algebra:

1. $\mathrm{tr}(\boldsymbol{A}) = \sum_{j=1}^{p} a_{jj}$, where a_{jj} is the jth diagonal element of a square $p \times p$ matrix $\boldsymbol{A}$, where $\mathrm{tr}(\cdot)$ is the **trace function** (sum of diagonal elements).
2. Cyclic property:
$$\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA}) = \mathrm{tr}(\boldsymbol{CAB}),$$
given that the product $\boldsymbol{ABC}$ is a square matrix.
3. If $\boldsymbol{A}$ is a $p \times p$ matrix, then for a $p \times 1$ vector $\boldsymbol{x}$,
$$\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = \mathrm{tr}(\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x})$$
holds by (1), since $\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ is a scalar.
4. $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$.

Multivariate normal likelihood

- It is convenient to rewrite $p(\boldsymbol{Y}|\boldsymbol{\theta}, \Sigma)$ as

$$
\begin{aligned}
p(\boldsymbol{Y}|\boldsymbol{\theta}, \Sigma) &\propto \underbrace{|\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\theta})^T\Sigma^{-1}(\boldsymbol{y}_i - \boldsymbol{\theta})\right\}}_{\text{no algebra/change yet}} \\
&= |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\underbrace{\text{tr}\left[(\boldsymbol{y}_i - \boldsymbol{\theta})^T\Sigma^{-1}(\boldsymbol{y}_i - \boldsymbol{\theta})\right]}_{\text{by result 3}}\right\} \\
&= |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\underbrace{\text{tr}\left[(\boldsymbol{y}_i - \boldsymbol{\theta})(\boldsymbol{y}_i - \boldsymbol{\theta})^T\Sigma^{-1}\right]}_{\text{by cyclic property}}\right\} \\
&= |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\underbrace{\text{tr}\left[\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\theta})(\boldsymbol{y}_i - \boldsymbol{\theta})^T\Sigma^{-1}\right]}_{\text{by result 4}}\right\} \\
&= |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\boldsymbol{S}_\theta\Sigma^{-1}\right]\right\},
\end{aligned}
$$

where $\boldsymbol{S}_\theta = \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\theta})(\boldsymbol{y}_i - \boldsymbol{\theta})^T$ is the residual sum of squares matrix.

What's next?

Move on to the readings for the next module!