Module 9: The Multivariate Normal Distribution

Rebecca C. Steorts

Hoff, Section 7.4

Announcements

1. The last day of classes with be April 16, 2019
2. There will be a special lecture on April 18, 2019 by one of my PhD students on mixture models (abstract/title forthcoming).
3. OH will be regularly scheduled until the final exam, April 29, 2019.
4. Your lab sections will serve as extra OH by your TAs until April 29, 2019.
5. The final exam will be April 29, 2019, 9 AM – noon (Old Chem 116).

Agenda

- Moving from univariate to multivariate distributions.
- The multivariate normal (MVN) distribution.
- Conjugate for the MVN distribution.
- The inverse Wishart distribution.
- Conjugate for the MVN distribution (but on the covariance matrix).
- Combining the MVN with inverse Wishart.
- See Chapter 7 (Hoff) for a review of the standard Normal density.

Example: Reading Comprehension

A sample of 22 children are given reading comprehension tests before and after receiving a particular instructional method.[1]

Each student i will then have two scores, $Y_{i,1}$ and $Y_{i,2}$ denoting the pre- and post-instructional scores respectively.

Denote each student's pair of scores by the vector $\boldsymbol{Y}_i$

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} \text{score on first test} \\ \text{score on second test} \end{pmatrix}$$

where $i = 1, \ldots, n$ and $p = 2$.

[1] This example follows Hoff (Section 7.4, p. 112).

Example: Reading Comprehension

What does this data look like that is observed?

$$\boldsymbol{X}_{n\times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{n1} \\ x_{21} & x_{22} & \dots & x_{n2} \\ x_{i1} & x_{i2} & \dots & x_{ni} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

- A row of $\boldsymbol{X}_{n\times p}$ represents a covariate we might be interested in, such as age of a person.
- Denote x_i as the ith row vector of the $X_{n\times p}$ matrix.

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

where its dimension is $p \times 1$.

Example: Reading Comprehension

We may be interested in the population mean $\boldsymbol{\mu}_{p\times 1}$.

$$E[\boldsymbol{Y}] =: E[\boldsymbol{Y}_i] = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

We also may be interested in the population covariance matrix, Σ.

$$\Sigma = Cov(\boldsymbol{Y}) = \begin{pmatrix} E[Y_1^2] - E[Y_1]^2 & E[Y_1 Y_2] - E[Y_1]E[Y_2] \\ E[Y_1 Y_2] - E[Y_1]E[Y_2] & E[Y_2^2] - E[Y_2]^2 \end{pmatrix} \tag{1}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} \tag{2}$$

Remark: $Cov(Y_1) = Var(Y_1) = \sigma_1^2.$ $\quad Cov(Y_1, Y_2) = \sigma_{1,2}.$

General Notation

Assume that $\boldsymbol{y}_{p\times 1} \sim (\mu_{p\times 1}, \Sigma_{p\times p})$.

$$\boldsymbol{y}_{p\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}.$$

$$\boldsymbol{\mu}_{p\times 1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\Sigma_{p\times p} = Cov(\boldsymbol{y}) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix}.$$

Linear Algebra Background

Suppose matrix A is invertible. The

$$\det(A) = \sum_{i=1}^{j=n} a_{ij} A_{ij}.$$

I recommend using the det() commend in R.

Suppose now we have a square matrix $H_{p \times p}$.

$$\text{trace}(H) = \sum_{i} h_{ii},$$

where h_{ii} are the diagonal elements of H.

Linear Algebra Tricks

Suppose that A is $n \times n$ matrix and suppose that B is a $n \times n$ matrix.

Lemma 1:

$$tr(AB) = tr(BA)$$

Proof: Exercise.

Lemma 2:

Suppose $\boldsymbol{x}$ is a vector.

$$\boldsymbol{x}^T A\boldsymbol{x} = tr(\boldsymbol{x}^T A\boldsymbol{x}) = tr(\boldsymbol{x}\boldsymbol{x}^T A) = tr(A\boldsymbol{x}\boldsymbol{x}^T)$$

Proof: Exercise.

Why are these useful? We'll come back to this later in the module.

Notation

- MVN is generalization of univariate normal.
- For the MVN, we write $\boldsymbol{y} \sim \mathcal{MVN}(\boldsymbol{\mu}, \Sigma)$.
- The $(i, j)^{\text{th}}$ component of Σ is the covariance between Y_i and Y_j (so the diagonal of Σ gives the component variances).

Example: $Cov(Y_1, Y_2)$ is just one element of the matrix Σ.

Multivariate Normal

Just as the probability density of a scalar normal is

$$p(x) = \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \tag{3}$$

the probability density of the multivariate normal is

$$p(\vec{x}) = (2\pi)^{-p/2}(\det \Sigma)^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{4}$$

Univariate normal is special case of the multivariate normal with a one-dimensional mean "vector" and a one-by-one variance "matrix."

Standard Multivariate Normal Distribution

Consider

$$Z_1, \ldots, Z_n \overset{iid}{\sim} MVN(0, I)$$

$$f_z(z) = \prod_{i=1}^{n} \frac{1}{2\pi} e^{-z_i^2/2} \tag{5}$$
$$= (\sqrt{2\pi})^{-n} e^{\sum_i -z_i^2/2} \tag{6}$$
$$= (2\pi)^{-n/2} e^{-z^T z/2} \tag{7}$$

Exercise: Why does $\sum_i -z_i^2 = -z^T z$?

- $\mathrm{E[Z]} = 0$
- $\mathrm{Var[Z]} = \mathrm{I}$