STATS 100C: Linear Models **Spring 2017**

Lecture 5: April 18

Lecturer: Arash Amini *Scribe: Eric Chuu*

1.1 Multivariate Normal Distribution

A random vector

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

has multivariate normal distribution with parameters

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \Sigma \in \mathbb{R}^{n \times n}$$

if it has the density

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right] \tag{1.1}$$

where Σ is invertible.

Example. Suppose $y_1, y_2, \ldots, y_n \overset{iid}{\sim} N(0,1)$. Then in vector form, the joint density function can be written as

$$f_Y(\mathbf{y}) = \prod_{i=1}^n f_{Y_i}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_i^2}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^n y_i^2\right] = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right]$$

We can see that $\boldsymbol{\mu} = \mathbf{0}, \Sigma = \mathbf{I_n}$, so $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I_n})$. □

1.1.1 Properties of the Multivariate Normal Distribution

Let $\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$.

(1) **Any affine transformation of $\mathbf{y}$, $\mathbf{u} = A\mathbf{y} + \mathbf{b}$, where $A, \mathbf{b}$ are nonrandom, follows a multivariate normal distribution, with**

$$\mathbf{u} \sim \left(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T\right)$$

(2) **The marginal distributions are normal**. suppose we can partition the vector $\mathbf{y}$ into two sub-vectors, $\mathbf{y}_1, \mathbf{y}_2$,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \hline \mathbf{y_2} \end{bmatrix}, \quad \mathbf{y}_1 \in \mathbb{R}^p, \mathbf{y}_2 \in \mathbb{R}^{n-p}$$

Then $\mathbf{y}_1, \mathbf{y}_2$ follow a multivariate normal distribution. We can obtain their respective distributions by noting that

$$\begin{aligned}
\mathbf{y}_1 &= \begin{bmatrix} \mathbf{I_p} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \\
A\mu &= \begin{bmatrix} \mathbf{I_p} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \boldsymbol{\mu}_1 \\
A\Sigma A^T &= \begin{bmatrix} \mathbf{I_p} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I_p} \\ \mathbf{0} \end{bmatrix}
\end{aligned}$$

(3) **Conditional Distributions are Multivariate Normal**
Using the same partition as in property (2), then the conditional distribution

$$\begin{aligned}
&\mathbf{y}_1 | \mathbf{y}_2 \sim N\left(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2}\right) \\
&\boldsymbol{\mu}_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y}_2 - \boldsymbol{\mu}_2) \\
&\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\end{aligned}$$

(4) **Uncorrelatedness is equivalent to Independence.** In general, independence implies uncorrelatedness, but the converse does not necessarily hold. For multivariate normal, the two notions are equivalent. Thus, if we have

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} \sim \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma 21 & \Sigma_{22} \end{bmatrix} \right)$$

and $\Sigma_{12} = 0$, then $\mathbf{y}_1 \perp \mathbf{y}_2$.

Let $\mathbf{x} \sim (\boldsymbol{\mu}, \Sigma)$. Note that since Σ is positive semi-definite, we can define $\Sigma^{\frac{1}{2}}$, where $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$. The decorrelated version of $\mathbf{x}$ is

$$\begin{aligned}
\mathbf{z} &= \Sigma^{-\frac{1}{2}}\mathbf{x} \\
\mathrm{Cov}(\mathbf{z}) &= \Sigma^{-\frac{1}{2}}\mathrm{Cov}(\mathbf{x})\Sigma^{-\frac{1}{2}} = \mathbf{I} \\
&\Rightarrow \mathbf{z} \sim N\left(\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}, \mathbf{I}\right)
\end{aligned}$$

Thus, we can standardize $\mathbf{x}$ by shifting:

$$\begin{aligned}
&\mathbf{z}^* = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu}) \\
&\Rightarrow \mathbf{z}^* \sim N(\mathbf{0}, \mathbf{I})
\end{aligned}$$

1.2 Multiple Linear Regression

Multiple Linear Regression is used to model a functional relationship between a response variable and one or more explanatory/predictor variables. The model is linear in the parameters:

$$Y = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p + \varepsilon \tag{1.2}$$

where

$$\mu := \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$$

is the deterministic component, and ε is random. Some assumptions that we make are:

1. The covariances are assumed to be fixed
2. ε is random variable, associated with noise and unexplained variation, with $\mathbb{E}(\varepsilon) = 0$.

$$\Rightarrow \mathbb{E}(Y) = \mathbb{E}(\mu + \varepsilon) = \mu + \mathbb{E}(\varepsilon) = \mu$$

We can consider the x_i, y_i as a collection of data/observations of independent samples:

$$\{y_i, x_{i1}, x_{i2}, \dots, x_{ip} : i = 1, \dots, n\}$$

The β's remain the same, so we have

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \cdots \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n && (1.3) \\ &= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \dots, n && (1.4) \\ &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} && (1.5) \end{aligned}$$

where

$$X = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_p \\ | & | & & | \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$