INTRODUCTION TO LINEAR MODELS AND REGRESSION

Moutushi Chatterjee, notes by Ramdas Singh

Fourth Semester

Contents

1 RANDOM VECTORS 1
  1.1 Definitions and Moments 1
    1.1.1 Moment Generating Function 2
  1.2 Linear Models 3

Index 5

Chapter 1

RANDOM VECTORS

1.1 Definitions and Moments

January 8th.

Recall that for a random variable X, the r-th *raw moment* is defined as

$$\mu_r' = \mathbb{E}[X^r] \tag{1.1}$$

provided the expectation exists. Here, $\mu_1' = \mathbb{E}[X] = \mu$ is the mean. The r-th *central moment* is defined as

$$\mu_r = \mathbb{E}[(X-\mu)^r]. \tag{1.2}$$

Similarly, for a set of sample points $x_1, x_2, \ldots, x_n$, the r-th *sample raw moment* is defined as

$$m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{1.3}$$

and the r-th *sample central moment* is defined as

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r \tag{1.4}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean. Note that $\mu_2 = \operatorname{Var}(X)$ is the variance of X and $\dfrac{\mu_3^2}{\mu_2^3}$ is the *skewness* of X. There is also *coefficient of variation* defined as $\dfrac{\sigma}{\mu}$ where $\sigma = \sqrt{\operatorname{Var}(X)}$ is the standard deviation of X. For a sample, it is defined as $\dfrac{s}{\bar{x}}$ where $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is the sample standard deviation. The *correlation coefficient* between x_i's and y_i's is defined as

$$\frac{\frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2 \cdot \frac{1}{n}\sum(y_i - \bar{y})^2}} \tag{1.5}$$

It is also known as the product-moment correlation coefficient.

These definitions can be extended to random vectors as follows. For a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$, the *mean vector* is defined as

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_p] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}. \tag{1.6}$$

The *dispersion matrix*, or covariance matrix, is defined as

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T] = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_p) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2,X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p,X_1) & \mathrm{Cov}(X_p,X_2) & \cdots & \mathrm{Var}(X_p) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}. \tag{1.7}$$

Similar to the correlation coefficient for random variables, we define the *correlation matrix* for a random vector $\mathbf{X}$ as

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} \tag{1.8}$$

Define D_σ as the diagonal matrix $\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$. Then, we have the relation

$$\boldsymbol{R} = D_\sigma^{-1}\boldsymbol{\Sigma}D_\sigma^{-1}. \tag{1.9}$$

If $\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$ is a random vector where $\mathbf{X}$ is of dimension p and $\mathbf{Y}$ is of dimension q, then we simply have

$$\mathbb{E}[\mathbf{Z}] = \begin{pmatrix} \mathbb{E}[\mathbf{X}] \\ \mathbb{E}[\mathbf{Y}] \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{Z}) = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{X}} & \boldsymbol{\Sigma}_{\mathbf{XY}} \\ \boldsymbol{\Sigma}_{\mathbf{XY}}^\top & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{pmatrix}. \tag{1.10}$$

As for linear transformations, if $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ where $\mathbf{A}$ is a $q \times p$ matrix and $\mathbf{b} \in \mathbb{R}^q$, then

$$\mathbb{E}[\mathbf{Y}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b}, \quad \mathrm{Cov}(\mathbf{Y}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{X})\mathbf{A}^\top. \tag{1.11}$$

In particular, if $\mathbf{a} = \mathbf{A}^\top$ is a p-dimensional vector, then Y is a univariate random variable written as $Y = \mathbf{a}^\top\mathbf{X} + b$.

Mahalanobis Distance

All of the below equations

$$d^2 = (\mathbf{y}_1 - \mathbf{y}_2)^\top \mathbf{S}^{-1}(\mathbf{y}_1 - \mathbf{y}_2), \tag{1.12}$$

$$D^2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}), \tag{1.13}$$

$$D^2 = (\bar{\mathbf{y}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \tag{1.14}$$

are known as the *Mahalanobis distance*. Here, S is the sample covariance matrix and $\boldsymbol{\Sigma}$ is the population covariance matrix. The Mahalanobis distance is a measure of distance that accounts for the correlations between variables and the different scales of the variables. Since $\boldsymbol{\Sigma}$ is positive semi-definite, we set $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}\mathbf{y} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ to get

$$(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{z}^\top\mathbf{z}. \tag{1.15}$$

This shows that the Mahalanobis distance is equivalent to the Euclidean distance in the transformed space where the data has been standardized by the covariance matrix. The covariance matrix of $\mathbf{z}$ is

$$\mathrm{Cov}(\mathbf{z}) = \boldsymbol{\Sigma}^{-1/2}\,\mathrm{Cov}(\mathbf{y})\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1/2}\frac{1}{n}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2} = \frac{1}{n}\mathbf{I}_p. \tag{1.16}$$

Thus, the variables in $\mathbf{z}$ are uncorrelated with equal variances; in other words, they have been standardized.

1.1.1 Moment Generating Function

Recall that in the univariate case, the moment generating function (MGF) of a random variable X is defined as

$$M_X(t) := \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{r=0}^{\infty} \frac{t^r X^r}{r!}\right]. \tag{1.17}$$

We extend this definition to the multivariate case as follows; for a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$, the *moment generating function* is defined as

$$M_{\mathbf{X}}(\mathbf{t}) := \mathbb{E}[e^{\mathbf{t}^\top \mathbf{X}}] = \mathbb{E}\left[e^{\sum_{i=1}^p t_i X_i}\right] \tag{1.18}$$

for all $\mathbf{t} = (t_1, t_2, \ldots, t_p)^T \in \mathbb{R}^p$ such that the expectation exists. When the X_i's are independent, we have

$$M_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^p M_{X_i}(t_i). \tag{1.19}$$

Similar to the univariate case, the MGF can be used to compute moments of the random vector. For example, the first moment (mean vector) can be obtained by taking the gradient of the MGF at $\mathbf{t} = \mathbf{0}$:

$$\mathbb{E}[\mathbf{X}] = \nabla_{\mathbf{t}} M_{\mathbf{X}}(\mathbf{t})\bigg|_{\mathbf{t}=\mathbf{0}}. \tag{1.20}$$

The second moment (covariance matrix) can be obtained by taking the Hessian of the MGF at $\mathbf{t} = \mathbf{0}$:

$$\mathrm{Cov}(\mathbf{X}) = \nabla^2_{\mathbf{t}} M_{\mathbf{X}}(\mathbf{t})\bigg|_{\mathbf{t}=\mathbf{0}} - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{X}]^\top. \tag{1.21}$$

Suppose $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ where $\mathbf{A}$ is a $q \times p$ matrix and $\mathbf{b} \in \mathbb{R}^q$. Then, the MGF of $\mathbf{Y}$ is given by

$$M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \mathbf{b}} M_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t}). \tag{1.22}$$

1.2 Linear Models

Generally, we had

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1.23}$$

where $\epsilon \sim N(0, \sigma^2)$. Here, we note that y is a random variable while x is a fixed variable. Moreover, $\sum_i (y_i - \bar{y}) = 0$ implies that $\sum_i \epsilon_i = 0$. In such a case, $\mathbb{E}[y] = \beta_0 + \beta_1 \mathbb{E}[x]$ and $\mathbb{E}[y \mid x] = \beta_0 + \beta_1 x$. A theorem states that if any one of the regression lines is constant, then $\mathrm{Cov}(X, Y) = 0$.

A simple proof is given as follows: note that $\mathbb{E}[Y \mid X] = k \implies \mathbb{E}[Y] = k$. Also,

$$\mathbb{E}[XY] = \int_{\mathbb{R}} \int_{\mathbb{R}} xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} x \int_{\mathbb{R}} y f_{Y|X}(y \mid x)\,\mathrm{d}y f_X(x)\,\mathrm{d}x = \int_{\mathbb{R}} xkf_X(x)\,\mathrm{d}x = k\mathbb{E}[X]. \tag{1.24}$$

Thus,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = k\mathbb{E}[X] - \mathbb{E}[X]k = 0. \tag{1.25}$$

Index

central moment, 1
coefficient of variation, 1
correlation coefficient, 1
correlation matrix, 2

dispersion matrix, 2

Mahalanobis distance, 2
mean vector, 1
moment generating function, 3

raw moment, 1

sample central moment, 1
sample raw moment, 1
skewness, 1