Chapter 5 Theory of the Multinormal

Applied Multivariate Statistical Analysis

Course Instructor: Shuen-Lin Jeng

Department of Statistics
National Cheng Kung University

March, 2, 2021

5.1 Elementary Properties of the Multinormal

- The pdf of $X \sim N_p(\mu, \sum)$ is

$$f(x) = |2\pi \sum|^{-1/2} exp\left\{-\frac{1}{2}(x-\mu)^T {\sum}^{-1}(x-\mu)\right\}. \tag{5.1}$$

  The expectation is $\mathbf{E}(X) = \mu$, the covariance can be calculated as $\mathbf{Var}(X) = \mathbf{E}(X-\mu)(X-\mu)^T = \sum$.

- Linear transformations turn normal random variables into normal random variables. If $X \sim N_p(\mu, \sum)$ and $\mathcal{A}(p \times p)$, $c \in \mathbb{R}^p$, then $Y = \mathcal{A}X + c$ is p-variate Normal, i.e.,

$$Y \sim N_p(\mathcal{A}\mu + c, \mathcal{A}\sum \mathcal{A}^T) \tag{5.2}$$

5.1 Elementary Properties of the Multinormal

- If $X \sim N_p(\mu, \sum)$, then the Mahalanobis transformation is

$$Y = \sum^{-1/2}(X - \mu) \sim N_p(0, \mathcal{I}_p) \tag{5.3}$$

and it holds that

$$Y^T Y = (X - \mu)^T \sum^{-1}(X - \mu) \sim \chi^2_p. \tag{5.4}$$

5.1 Elementary Properties of the Multinormal

Theorem 5.1 *Let* $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p(\mu, \sum), X_1 \in \mathbb{R}^r, X_2 \in \mathbb{R}^{p-r}$. *Define* $X_{2.1} = X_2 - (\sum_{21})(\sum_{11}^{-1})X_1$ *from the partitioned covariance matrix*

$$\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$$

Then

$$X_1 \sim N_r\left(\mu_1, \sum\nolimits_{11}\right) \tag{5.5}$$

$$X_{2.1} \sim N_{p-r}\left(\mu_{2.1}, \sum\nolimits_{22.1}\right) \tag{5.6}$$

are independent with

$$\mu_{2.1} = \mu_2 - \sum\nolimits_{21}\sum\nolimits_{11}^{-1}\mu_1, \sum\nolimits_{22.1} = \sum\nolimits_{22} - \sum\nolimits_{21}\sum\nolimits_{11}^{-1}\sum\nolimits_{12}. \tag{5.7}$$

5.1 Elementary Properties of the Multinormal

Corollary 5.1 *Let* $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p(\mu, \sum)$, $\sum = \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}$.
$\sum_{12} = 0$ *if and only if* X_1 *is independent of* X_2.

Corollary 5.2 *If* $X \sim N_p(\mu, \sum)$ *and given some matrices* $\mathcal{A}$ *and* $\mathcal{B}$, *then* $\mathcal{A}X$ *and* $\mathcal{B}X$ *are independent if and only if* $\mathcal{A}\sum\mathcal{B}^T = 0$.

Theorem 5.2 *If* $X \sim N_p(\mu, \sum)$, $\mathcal{A}(q \times p), c \in \mathbb{R}^q$, *and* $q \leq p$, *then* $Y = \mathcal{A}X + c$ *is a q-variate normal, i.e.,*

$$Y \sim N_q(\mathcal{A}\mu + c, \mathcal{A}\sum\mathcal{A}^T).$$

5.1 Elementary Properties of the Multinormal

Theorem 5.3 *The conditional distribution of X_2 given $X_1 = x_1$ is normal with mean $\mu_2 + \sum_{21}\sum_{11}^{-1}(x_1 - \mu_1)$ and covariance $\sum_{22.1}$, i.e.,*

$$(X_2 \mid X_1 = x_1) \sim N_{p-r}\left(\mu_2 + \sum_{21}\sum_{11}^{-1}(x_1 - \mu_1), \sum_{22.1}\right) \tag{5.8}$$

Theorem 5.4 *If $X_1 \sim N_r(\mu_1, \sum_{11})$ and $(X_2 \mid X_1 = x_1) \sim N_{p-r}(\mathcal{A}x_1 + b, \Omega)$ where Ω does not depend on x_1 , then $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_p(\mu, \sum)$, where*

$$\mu = \begin{pmatrix} \mu_1 \\ \mathcal{A}\mu_1 + b \end{pmatrix}$$

$$\sum = \begin{pmatrix} \sum_{11} & \sum_{11}\mathcal{A}^T \\ \mathcal{A}\sum_{11}\Omega & \mathcal{A}\sum_{11}\mathcal{A}^T \end{pmatrix}$$

5.1 Elementary Properties of the Multinormal

Conditional Approximations

As we saw in Chap.4(Theorem 4.3), the conditional expectation $\mathbf{E}(X_2 \mid X_1)$ is the mean squared error (MSE) best approximation of X_2 by a function of X_1. We have in this case

$$X_2 = \mathbf{E}(X_2 \mid X_1) + U = \mu_2 + \sum\nolimits_{21} \sum\nolimits_{11}^{-1} (X_1 - \mu_1) + U \tag{5.9}$$

Hence, the best approximation of $X_2 \in \mathbb{R}^{r-p}$ by $X_1 \in \mathbb{R}^r$ is the linear approximation that can be written as

$$X_2 = \beta_0 + \mathcal{B} X_1 + U \tag{5.10}$$

with $\mathcal{B} = \sum_{21} \sum_{11}^{-1}$,$\beta_0 = \mu_2 - B\mu_1$ and $U \sim N(0, \sum_{22.1})$.

5.1 Elementary Properties of the Multinormal

Consider now the particular case where $r = p - 1$. Now $X_2 \in \mathbb{R}$ and $\mathcal{B}$ is a row vector β^T of dimension $(1 \times r)$

$$X_2 = \beta_0 + \beta^T X_1 + U \tag{5.11}$$

The marginal variance of X_2 can be decomposed via (5.11):

$$\sigma_{22} = \beta^T \sum\nolimits_{11} \beta + \sigma_{22.1} = \sigma_{21} \sum\nolimits_{11}^{-1} \sigma_{12} + \sigma_{22.1} \tag{5.12}$$

The ratio

$$\rho^2_{2.1 \ldots r} = \frac{\sigma_{21} \sum_{11}^{-1} \sigma_{12}}{\sigma_{22}} \tag{5.13}$$

5.2 The Wishart Distribution

The Wishart distribution (named after its discoverer) plays a prominent role in the analysis of the estimated covariance matrices. If the mean of $X \sim N_p(\mu, \sum)$ is known to be $\mu = 0$, then for a data matrix $\mathcal{X}(n \times p)$ the estimated covariance matrix is proportional to $\mathcal{X}^{\mathcal{T}}\mathcal{X}$. This is the point where the Wishart distribution comes in, because $\mathcal{M}(p \times p) = \mathcal{X}^{\mathcal{T}}\mathcal{X} = \sum_{i=1}^{n} x_i {x_i}^T$ has a Wishart distribution $W_p(\sum, n)$.

5.2 The Wishart Distribution

Example 5.4 Set $p = 1$, then for $X \sim N_1(0, \sigma^2)$ the data matrix of the observations

$$\mathcal{X} = (x_1, \ldots, x_n)^T \text{ with } \mathcal{M} = \mathcal{X}^T \mathcal{X} = \sum_{i=1}^{n} x_i x_i$$

leads to the Wishart distribution $W_1(\sigma^2, n) = \sigma^2 \chi_n^2$. The one-dimensional Wishart distribution is thus in fact a χ^2 distribution.

5.2 The Wishart Distribution

When we talk about the distribution of a matrix, we mean of course the joint distribution of all its elements. More exactly: since $\mathcal{M} = \mathcal{X}^{\mathcal{T}}\mathcal{X}$ is symmetric we only need to consider the elements of the lower triangular matrix

$$\mathcal{M} = \begin{pmatrix} m_{11} & & & \\ m_{21} & m_{22} & & \\ \vdots & \vdots & \ddots & \\ m_{p1} & m_{p2} & \dots & m_{pp} \end{pmatrix}. \tag{5.14}$$

Hence the Wishart distribution is defined by the distribution of the vector

$$\begin{pmatrix} m_{11} & , \dots, & m_{22} & , \dots, & m_{p2}, & \dots, & m_{pp} \end{pmatrix}^{T} \tag{5.15}$$

5.2 The Wishart Distribution

Theorem 5.5 *If $\mathcal{M} \sim W_p(\sum, n)$ and $\mathcal{B}(p \times q)$, then the distribution of $\mathcal{B}^T \mathcal{M} \mathcal{B}$ is Wishart $W_q(\mathcal{B}^T \sum \mathcal{B}, n)$.*

With this theorem, we can standardize Wishart matrices since with $\mathcal{B} = \sum^{-1/2}$ the distribution of $\sum^{-1/2} \mathcal{M} \sum^{-1/2}$ is $W_p(\mathcal{I}, n)$.

Theorem 5.6 *If $\mathcal{M} \sim W_p(\sum, m)$, and $a \in \mathbb{R}^p$ with $a^T \sum a \neq 0$, then the distribution of $\frac{a^T \mathcal{M} a}{a^T \sum a}$ is χ^2_m*

5.2 The Wishart Distribution

Theorem 5.7 (Cochran) *Let $\mathcal{X}(n \times p)$ be a data matrix from a $N_p(0, \sum)$ distribution and let $\mathcal{C}(n \times n)$ be a symmetric matrix.*

(a) *$\mathcal{X}^T \mathcal{C} \mathcal{X}$ has the distribution of weighted Wishart random variables, i.e.,*

$$\mathcal{X}^{\mathcal{C}} \mathcal{X} = \sum_{i=1}^{n} \lambda_i W_p(\sum, 1),$$

where λ_i, $i = 1, \ldots, n$, are the eigenvalues of $\mathcal{C}$.

5.2 The Wishart Distribution

(b) $\mathcal{X}^T \mathcal{C} \mathcal{X}$ is Wishart if and only if $\mathcal{C}^2 = \mathcal{C}$. In this case,

$$\mathcal{X}^T \mathcal{C} \mathcal{X} \sim W_p(\sum, r)$$

and $r = rank(\mathcal{C}) = tr(\mathcal{C})$.

(c) $n\mathcal{S} = \mathcal{X}^T \mathcal{H} \mathcal{X}$ is distributed as $W_p(\sum, n-1)$ (note that $\mathcal{S}$ is the sample covariance matrix).

(d) $\bar{x}$ and $\mathcal{S}$ are independent.

5.2 The Wishart Distribution

The following properties are useful:

1. If $\mathcal{M} \sim W_p(\sum, n)$, then $\mathbf{E}(\mathcal{M}) = n\sum$.
2. If $\mathcal{M}_i$ are independent Wishart $W_p(\sum, n_i)$ $i = 1, \ldots, k$, then $\mathcal{M} = \sum_{i=1}^{k} \mathcal{M}_i \sim W_p(\sum, n)$ where n $= \sum_{i=1}^{k} n_i$.
3. The density of $W_p(\sum, n-1)$ for a positive definite , $\mathcal{M}$ is given by

$$f_{\sum,n-1}(\mathcal{M}) = \frac{|\mathcal{M}|^{\frac{1}{2}(n-p-2)} e^{-\frac{1}{2}\mathrm{tr}(\mathcal{M}\sum^{-1})}}{2^{\frac{1}{2}p(n-1)} \pi^{\frac{1}{4}p(p-1)} |\sum|^{\frac{1}{2}(n-1)} \prod_{i=1}^{p} \Gamma\left\{\frac{n-i}{2}\right\}}, \qquad (5.16)$$

where Γ is the gamma function: $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

5.3 Hotelling's T^2- Distribution

Suppose that $Y \in \mathbb{R}^p$ is a standard normal random vector, i.e., $Y \sim N_p(0, \mathcal{I})$, independent of the random matrix $\mathcal{M} \sim W_p(\mathcal{I}, n)$. What is the distribution of $Y^T \mathcal{M}^{-1} Y$? The answer is provided by the Hotelling T^2-distribution: n $Y^T \mathcal{M} Y$ is Hotelling $T^2_{p,n}$ distributed.
The Hotelling T^2-distribution is a generalization of the Student t-distribution.

Theorem 5.8 *If $X \sim N_p(\mu, \sum)$ is independent of $\mathcal{M} \sim W_p(\sum, n)$, then*

$$n(X-\mu)^T \mathcal{M}^{-1}(X-\mu) \sim T^2_{p,n}$$

5.3 Hotelling's T^2- Distribution

Corollary 5.3 *If $\bar{x}$ is the mean of a sample drawn from a normal population $N_p(\mu, \sum)$ and $\mathcal{S}$ is the sample covariance matrix, then*

$$(n-1)(\bar{x}-\mu)^T \mathcal{S}^{-1}(\bar{x}-\mu) = n(\bar{x}-\mu)^T \mathcal{S}_u^{-1}(\bar{x}-\mu) \sim T^2_{p,n-1}. \quad (5.17)$$

Recall that $\mathcal{S}_u = \frac{n}{n-1}\mathcal{S}$ is an unbiased estimator of the covariance matrix.

Theorem 5.9

$$T^2_{p,n} = \frac{np}{n-p+1} F_{p,n-p+1}$$

5.3 Hotelling's T^2- Distribution

Example5.5 In the univariate case ($p = 1$), this theorem boils down to the well-known result:

$$(\frac{\bar{x}-\mu}{\sqrt{\mathcal{S}_u}/\sqrt{n}})^2 \sim T^2_{1,n-1} = F_{1,n-1} = t^2_{n-1}$$

Corollary 5.4 Consider a linear transform of $X \sim N_p(\mu, \sum)$, $Y = \mathcal{A}X$ where $\mathcal{A}(q \times p)$ with $(q \leq p)$. If $\bar{x}$ and $\mathcal{S}_x$ are the sample mean and the sample covariance matrix, we have

$$\bar{y} = \mathcal{A}\bar{x} \sim N_q(\mathcal{A}\mu, \frac{1}{n}\mathcal{A}\sum \mathcal{A}^T)$$

$$n\mathcal{S}_Y = n\mathcal{A}\mathcal{S}_X\mathcal{A}^T \sim W_q(\mathcal{A}\sum \mathcal{A}^T, n-1)$$

$$(n-1)(\mathcal{A}\bar{x}-\mathcal{A}\mu)^T(\mathcal{A}\mathcal{S}_X\mathcal{A}^T)^{-1}(\mathcal{A}\bar{x}-\mathcal{A}\mu) \sim T^2_{q,n-1}$$

5.3 Hotelling's T^2- Distribution

We can write (5.17) as

$$T^2 = \sqrt{n}(\bar{x} - \mu)^T \left(\frac{\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})^T}{n-1}\right)^{-1} \sqrt{n}(\bar{x} - \mu)$$

which is of the form

$$\begin{pmatrix}\text{multivariate normal}\\ \text{random vector}\end{pmatrix}^{\top} \left(\frac{\begin{matrix}\text{Wishart random}\\ \text{matrix}\end{matrix}}{\text{degrees of freedom}}\right)^{-1} \begin{pmatrix}\text{multivariate normal}\\ \text{random vector}\end{pmatrix}.$$

5.3 Hotelling's T^2- Distribution

This is analogous to

$$t^2 = \sqrt{n}(\bar{x} - \mu)(s^2)^{-1}\sqrt{n}(\bar{x} - \mu)$$

or

$$\begin{pmatrix}\text{normal}\\ \text{random variable}\end{pmatrix}\left(\frac{\begin{matrix}\chi^2\text{-random}\\ \text{variable}\end{matrix}}{\text{degrees of freedom}}\right)^{-1}\begin{pmatrix}\text{normal}\\ \text{random variable}\end{pmatrix}$$

for the univariate case.

5.4 Spherical and Elliptical Distributions

The multinormal distribution belongs to the large family of elliptical distributions.

Elliptical distributions are often used, particularly in risk management.