Lecture 2. The Wishart distribution

In this lecture, we define the Wishart distribution, which is a family of distributions for symmetric positive definite matrices, and show its relation to Hotelling's T^2 statistic.

2.1 The Wishart distribution

The Wishart distribution is a family of distributions for symmetric positive definite matrices. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent $N_p(\mathbf{0}, \Sigma)$ and form a $p \times n$ data matrix $\mathbf{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n]$. The distribution of a $p \times p$ random matrix $\mathbf{M} = \mathbf{X}\mathbf{X}' = \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i'$ is said to have the Wishart distribution.

Definition 1. The random matrix $\mathbf{M}_{(p\times p)} = \sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i'$ has the Wishart distribution with n degrees of freedom and covariance matrix Σ and is denoted by $\mathbf{M} \sim W_p(n, \Sigma)$. For $n \geq p$, the probability density function of $\mathbf{M}$ is

$$f(\mathbf{M}) = \frac{1}{2^{np/2}\Gamma_p(\frac{n}{2})|\Sigma|^{n/2}} |\mathbf{M}|^{(n-p-1)/2} \exp[-\frac{1}{2}\text{trace}(\Sigma^{-1}\mathbf{M})],$$

with respect to Lebesque measure on the cone of symmetric positive definite matrices. Here, $\Gamma_p(\alpha)$ is the multivariate gamma function.

The precise form of the density is rarely used. Two exceptions are that i) in Bayesian computation, the Wishart distribution is often used as a conjugate prior for the inverse of normal covariance matrix and that ii) when symmetric positive definite matrices are the random elements of interest in diffusion tensor study.

The Wishart distribution is a multivariate extension of χ^2 distribution. In particular, if $M \sim W_1(n, \sigma^2)$, then $M/\sigma^2 \sim \chi^2_n$. For a special case $\Sigma = \mathbb{I}$, $W_p(n, \mathbb{I})$ is called the standard Wishart distribution.

Proposition 1.

i. For $\mathbf{M} \sim W_p(n, \Sigma)$ and $\mathbf{B}_{(p\times m)}$, $\mathbf{B}'\mathbf{M}\mathbf{B} \sim W_m(n, \mathbf{B}'\Sigma\mathbf{B})$.

ii. For $\mathbf{M} \sim W_p(n, \Sigma)$ with $\Sigma > 0$, $\Sigma^{-\frac{1}{2}}\mathbf{M}\Sigma^{-\frac{1}{2}} \sim W_p(n, \mathbb{I}_p)$.

iii. If $\mathbf{M}_i$ are independent $W_p(n_i, \Sigma)$ $(i = 1, \ldots, k)$, then $\sum_{i=1}^k \mathbf{M}_i \sim W_p(n, \Sigma)$, where $n = n_1 + \ldots + n_k$.

iv. For $\mathbf{M}_n \sim W_p(n, \Sigma)$, $E\mathbf{M}_n = n\Sigma$.

v. If $\mathbf{M}_1$ and $\mathbf{M}_2$ are independent and satisfy $\mathbf{M}_1 + \mathbf{M}_2 = \mathbf{M} \sim W_p(n, \Sigma)$ and $\mathbf{M}_1 \sim W_p(n_1, \Sigma)$ then $\mathbf{M}_2 \sim W_p(n - n_1, \Sigma)$.

The law of large numbers and the Cramer-Wold device leads to $\mathbf{M}_n/n \to \Sigma$ in probability as $n \to \infty$.

Corollary 2. *If $\mathbf{M} \sim W_p(n, \Sigma)$ and $\boldsymbol{a} \in \mathbb{R}^p$ is such that $\boldsymbol{a}'\Sigma\boldsymbol{a} \neq 0$*, then*

$$\frac{\boldsymbol{a}'\mathbf{M}\boldsymbol{a}}{\boldsymbol{a}'\Sigma\boldsymbol{a}} \sim \chi^2_n.$$

*The condition $\boldsymbol{a}'\Sigma\boldsymbol{a} \neq 0$ is the same as $\boldsymbol{a} \neq \mathbf{0}$ if $\Sigma > 0$.

Theorem 3. *If* $\mathbf{M} \sim W_p(n, \Sigma)$ *and* $\boldsymbol{a} \in \mathbb{R}^p$ *and* $n > p - 1$*, then*

$$\frac{\boldsymbol{a}'\Sigma^{-1}\boldsymbol{a}}{\boldsymbol{a}'\mathbf{M}^{-1}\boldsymbol{a}} \sim \chi^2_{n-p+1}.$$

The previous theorem holds for any deterministic $\boldsymbol{a} \in \mathbb{R}^p$, thus holds for any random $\boldsymbol{a}$ provided that the distribution of $\boldsymbol{a}$ is independent of $\mathbf{M}$. This is important in the next subsection.

The following lemma is useful in a proof of Theorem 3

Lemma 4. *For* $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ *invertible, we have* $A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$*, where*

$$\begin{aligned} A^{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \\ A^{12} &= -A^{11}A_{12}A_{22}^{-1}, \\ A^{21} &= -A_{22}^{-1}A_{21}A^{11}, \\ A^{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}. \end{aligned}$$

2.2 Hotelling's T^2 statistic

Definition 2. Suppose $\boldsymbol{X}$ and $\mathbf{S}$ are independent and such that

$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma), \quad m\mathbf{S} \sim W_p(m, \Sigma).$$

Then

$$T_p^2(m) = (\boldsymbol{X} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\boldsymbol{X} - \boldsymbol{\mu})$$

is known as Hotelling's T^2 statistic.

Hotelling's T^2 statistic plays a similar role in multivariate analysis to that of the student's t-statistic in univariate statistical analysis. That is, the application of Hotelling's T^2 statistic is of great practical importance in testing hypotheses about the mean of a multivariate normal distribution when the covariance matrix is unknown.

Theorem 5. *If* $m > p - 1$,

$$\frac{m - p + 1}{mp} T_p^2(m) \sim F_{p, m-p+1}.$$

A special case is when $p = 1$, where Theorem 5 indicates that $T_1^2(m) \sim F_{1,m}$. Where is the connection of the Hotelling's T^2 statistic to the student's t-distribution?

Note that we are indeed abusing the definition of 'statistic' here.

2.3 Samples from a multivariate normal distribution

Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are i.i.d. $N_p(\boldsymbol{\mu}, \Sigma)$. Denote the sample mean and sample variance by

$$\begin{aligned} \bar{\boldsymbol{X}} &= \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_i, \\ \mathbf{S} &= \frac{1}{n-1} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'. \end{aligned}$$

Theorem 6. $\bar{\boldsymbol{X}}$ *and* $\mathbf{S}$ *are independent, with*

$$\begin{aligned} \sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) &\sim N_p(0, \Sigma), \\ (n-1)\mathbf{S} &\sim W_p(n-1, \Sigma). \end{aligned}$$

Corollary 7. *The Hotelling's* T^2 *statistic for MVN sample is defined as*

$$T^2(n-1) = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}),$$

and we have

$$\frac{n-p}{p} \frac{n}{n-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim F_{p, n-p}.$$

(Incomplete) proof of Theorem 6. First note the following decomposition:

$$\sum_i (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})' = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})' + \sum_i (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})';$$

and recall the definition of the Wishart distribution.

It is easy to check the result on $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$. The following argument is for the independence and the distribution of $\mathbf{S}$.

Consider a new set of random vectors $\boldsymbol{Y}_i$ $(i = 1, \ldots, n)$ from a linear combination of $\boldsymbol{X}_i$s by an orthogonal matrix D satisfying

$$D = [d_1, \ldots, d_n], \quad d_1 = \frac{1}{\sqrt{n}}\mathbf{1}_n,$$
$$DD' = D'D = \mathbb{I}_n.$$

Let

$$\boldsymbol{Y}_j = \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu}) d_{ji} = \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i d_{ji} = \tilde{\boldsymbol{X}} d_j,$$

where $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_n]$ is the $p \times n$ matrix of de-meaned random vectors. We claim the following:

1. $\boldsymbol{Y}_j$ are normally distributed.
2. $E(\boldsymbol{Y}_j \boldsymbol{Y}_k') = \begin{cases} \Sigma, & \text{if } j = k; \\ \mathbf{0}, & \text{if } j \neq k. \end{cases}$
3. $\sum_{i=1}^{n} \boldsymbol{Y}_i \boldsymbol{Y}_i' = \sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i'$.
4. $\boldsymbol{Y}_1 \boldsymbol{Y}_1' = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'$.

The facts 1 and 2 show the independence, while facts 3 and 4 give the distribution of $\mathbf{S}$. □

Next lecture is on the inference about the multivariate normal distribution.