Wishart distribution
Test for equality of the means

Krzysztof Bartoszek

November 27, 2017

Wishart distribution

Johnson & Wichern Section 4.4 p. 176 (International Edition): Let $\mathbb{R}^p \ni \vec{Z}_i$, $i = 1, \ldots, m$ be all independent and equally distributed as $\mathcal{N}(\vec{0}, \mathbf{\Sigma})$, then we say that

$$\sum_{i=1}^{m} \vec{Z}_i \vec{Z}_i^T \sim W_m(\mathbf{\Sigma})$$

is Wishart–distributed with m–degrees of freedom (and parameter $\mathbf{\Sigma}$).
Johnson & Wichern Section 4.4 p. 176 (International Edition): **FACT**

$$(n-1)\mathbf{S} = \sum_{i=1}^{n} (\vec{X}_i - \vec{\mu})(\vec{X}_i - \vec{\mu})^T \sim W_{n-1}(\mathbf{\Sigma})$$

You are expected to know that definition of a Wishart distribution and what is the distribution of $(n-1)\mathbf{S}$.

Two sample test for equality of the means

We have independent two samples from $\mathbb{R}^p$: $\vec{X}_1^{(1)}, \ldots, \vec{X}_{n_1}^{(1)} \sim \mathcal{N}(\vec{\mu}_1, \mathbf{\Sigma}_1)$ and $\vec{X}_1^{(2)}, \ldots, \vec{X}_{n_2}^{(2)} \sim \mathcal{N}(\vec{\mu}_2, \mathbf{\Sigma}_2)$. One is interested in testing the hypotheses

$$H_0: \quad \vec{\mu}_1 = \vec{\mu}_2 \quad \text{versus} \quad H_1: \quad \vec{\mu}_1 \neq \vec{\mu}_2$$

at the α significance level. Remember that this means that you are asking if the means are equal on all coordinates or if there exists a coordinate on which they differ (**NOT** on which one they differ). Notice that under the null hypothesis, H_0,

- $$\overline{X}_1 - \overline{X}_2 \sim \mathcal{N}\left(\vec{0}, \frac{1}{n_1}\mathbf{\Sigma}_1 + \frac{1}{n_2}\mathbf{\Sigma}_2\right)$$

  The above fact you should be able to know and derive. It comes from the properties of the normal distribution and of the variance–covariance matrix.

The general form of the test statistic is

$$T^2 = (\overline{X}_1 - \overline{X}_2)^T \mathbf{M}^{-1} (\overline{X}_1 - \overline{X}_2).$$

Depending on $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ we will have different values of $\mathbf{M}$ and different distributions for T^2.

- $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ known then $\mathbf{M} = \frac{1}{n_1}\mathbf{\Sigma}_1 + \frac{1}{n_2}\mathbf{\Sigma}_2$ and $T^2 \sim \chi^2(p)$. **THIS YOU ARE EXPECTED TO KNOW FOR THE EXAM** as it is a conclusion from properties of the normal distribution (Johnson & Wichern Result 4.7 p. 163 International Edition).

- Johnson & Wichern Section 6.3 p. 285 (International Edition): $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ assumed equal but unknown then

  $$\mathbf{M} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2\right) =: \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}$$

  and $T^2 \sim \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1}$. Here you do not have to remember the formulae but you are expected to understand the connection with the one sample Hotelling's T^2 test. Also you should remember that here (and in the one sample case) the distribution of the test statistic is an F one.

- $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ not assumed equal and are unknown then $\mathbf{M} = \frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2$ and we do not know the distribution of T^2 in exact form. Approximations are given in the textbook (Subsection "The Two–Sample Situation When $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$" p. 292 (large sample size), 294 (small sample size), International Edition). Here you do not have to remember the formulae. You are expected to know that T^2's distribution is not known in exact form in this case and be able to motivate the expression for $\mathbf{M}$ (as it is a conclusion from how one estimates the variance–covariance matrix).