

Multivariate Statistics

Lecture 07

Fudan University

- 1 Noncentral Chi-Squared Distribution
- 2 Hypothesis Testing for the Mean (Covariance is Known)
- 3 The Generalized T^2 -Statistic

- 1 Noncentral Chi-Squared Distribution
- 2 Hypothesis Testing for the Mean (Covariance is Known)
- 3 The Generalized T^2 -Statistic

Noncentral Chi-Squared Distribution

If y_1, \dots, y_k are independent and each y_i is distributed according to the noncentral chi-squared distribution with n_i degrees of freedom and noncentrality parameter λ_i , then

$$\sum_{i=1}^k y_i \sim \chi_{n_1 + \dots + n_k}^2 \left(\sum_{i=1}^k \lambda_i \right).$$

Theorem 1

If the n -component vector \mathbf{y} is distributed according to $\mathcal{N}(\boldsymbol{\nu}, \mathbf{T})$ with $\mathbf{T} \succ \mathbf{0}$, then

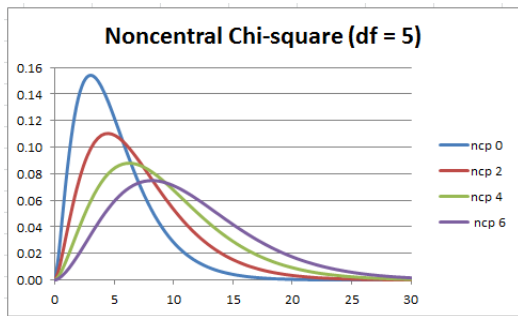
$$\mathbf{y}^\top \mathbf{T}^{-1} \mathbf{y}$$

is distributed according to the noncentral χ^2 -distribution with n degrees of freedom and noncentral parameter $\boldsymbol{\nu}^\top \mathbf{T}^{-1} \boldsymbol{\nu}$. If $\boldsymbol{\nu} = \mathbf{0}$, the distribution is the central χ^2 -distribution.

Noncentral Chi-Squared Distribution

Let $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\lambda}, \mathbf{I})$, then $v = \mathbf{y}^\top \mathbf{y}$ is distributed according to the noncentral χ^2 -distribution with p degrees of freedom and noncentral parameter $\tau^2 = \boldsymbol{\lambda}^\top \boldsymbol{\lambda}$. The probability density function is

$$f(v; p, \tau^2) = \begin{cases} \frac{\exp\left(-\frac{1}{2}(\tau^2 + v)\right) v^{\frac{p}{2}-1}}{2^{\frac{p}{2}} \sqrt{\pi}} \sum_{\beta=0}^{\infty} \frac{\tau^{2\beta} v^{\beta} \Gamma\left(\beta + \frac{1}{2}\right)}{(2\beta)! \Gamma\left(\frac{p}{2} + \beta\right)} & v > 0, \\ 0, & \text{otherwise.} \end{cases}$$



Noncentral Chi-Squared Distribution

Theorem 1

If the n -component vector \mathbf{y} is distributed according to $\mathcal{N}(\boldsymbol{\nu}, \mathbf{T})$ with $\mathbf{T} \succ \mathbf{0}$, then

$$\mathbf{y}^\top \mathbf{T}^{-1} \mathbf{y}$$

is distributed according to the noncentral χ^2 -distribution with n degrees of freedom and noncentral parameter $\boldsymbol{\nu}^\top \mathbf{T}^{-1} \boldsymbol{\nu}$. If $\boldsymbol{\nu} = \mathbf{0}$, the distribution is the central χ^2 -distribution.

For the sample mean $\bar{\mathbf{x}} \sim \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{N}\boldsymbol{\Sigma})$, we have $\sqrt{N}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.

It follows from the theorem that

$$N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$$

has a (central) χ^2 -distribution with p degrees of freedom.

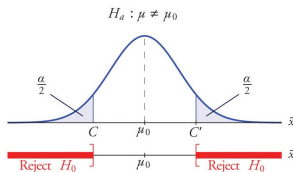
Outline

- 1 Noncentral Chi-Squared Distribution
- 2 Hypothesis Testing for the Mean (Covariance is Known)
- 3 The Generalized T^2 -Statistic

Hypothesis Testing for the Mean (Covariance is Known)

In the univariate case, the difference between the sample mean and the population mean is normally distributed. We consider

$$z = \frac{\sqrt{N}}{\sigma}(\bar{x} - \mu_0).$$



What about multivariate case?

- ① For $\alpha = 0.05$ and $p = 100$, we have $(1 - \alpha)^p \approx 0.006$.
- ② For $\alpha \approx 0.0005$ and $p = 100$, we have $(1 - \alpha)^p > 0.95$.

Hypothesis Testing for the Mean (Covariance is Known)

What about multivariate case?

$$\frac{\sqrt{N}}{\sigma}(\bar{x} - \mu_0) \implies \frac{N}{\sigma^2}(\bar{x} - \mu_0)^2 \implies N(\bar{x} - \mu_0)^\top \boldsymbol{\Sigma}^{-1}(\bar{x} - \mu_0)$$

Noncentral Chi-Squared Distribution

Theorem 1

If the n -component vector \mathbf{y} is distributed according to $\mathcal{N}(\boldsymbol{\nu}, \mathbf{T})$ with $\mathbf{T} \succ \mathbf{0}$, then

$$\mathbf{y}^\top \mathbf{T}^{-1} \mathbf{y}$$

is distributed according to the noncentral χ^2 -distribution with n degrees of freedom and noncentral parameter $\boldsymbol{\nu}^\top \mathbf{T}^{-1} \boldsymbol{\nu}$. If $\boldsymbol{\nu} = \mathbf{0}$, the distribution is the central χ^2 -distribution.

For the sample mean $\bar{\mathbf{x}} \sim \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{N} \boldsymbol{\Sigma})$, we have $\sqrt{N}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$.

It follows from the theorem that

$$N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

has a (central) χ^2 -distribution with p degrees of freedom.

Hypothesis Testing for the Mean (Covariance is Known)

Let $\chi_p^2(\alpha)$ be the number such that

$$\Pr \left\{ N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) > \chi_p^2(\alpha) \right\} = \alpha.$$

To test the hypothesis that $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0$ is a specified vector, we use as our rejection region (critical region)

$$N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) > \chi_p^2(\alpha).$$

Hypothesis Testing for the Mean (Covariance is Known)

Consider the following statement made on the basis of a sample with mean $\bar{\mathbf{x}}$: “The mean of the distribution satisfies

$$N(\bar{\mathbf{x}} - \boldsymbol{\mu}^*)^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}^*) \leq \chi_p^2(\alpha).$$

as an inequality on $\boldsymbol{\mu}^*$.” This statement is true with probability $1 - \alpha$.

Thus, the set of $\boldsymbol{\mu}^*$ satisfying above inequality is a confidence region for $\boldsymbol{\mu}$ with confidence $1 - \alpha$.

Two-Sample Problems

We suppose

- ① a sample $\{\mathbf{x}_\alpha^{(1)}\}$, $i = 1, \dots, N_1$ from the distribution $\mathcal{N}(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$;
- ② a sample $\{\mathbf{x}_\alpha^{(2)}\}$, $i = 1, \dots, N_2$ from the distribution $\mathcal{N}(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$.

Then the two sample means

$$\bar{\mathbf{x}}^{(1)} = \frac{1}{N_1} \sum_{\alpha=1}^{N_1} \mathbf{x}_\alpha^{(1)} \sim \mathcal{N}\left(\boldsymbol{\mu}^{(1)}, \frac{1}{N_1} \boldsymbol{\Sigma}\right)$$

and

$$\bar{\mathbf{x}}^{(2)} = \frac{1}{N_2} \sum_{\alpha=1}^{N_2} \mathbf{x}_\alpha^{(2)} \sim \mathcal{N}\left(\boldsymbol{\mu}^{(2)}, \frac{1}{N_2} \boldsymbol{\Sigma}\right).$$

are independent.

Two-Sample Problems

Then we have

$$\mathbf{y} = \bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \bar{\mathbf{x}}^{(2)} \end{bmatrix}, \quad \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \bar{\mathbf{x}}^{(2)} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix}, \begin{bmatrix} \frac{1}{N_1} \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \frac{1}{N_2} \boldsymbol{\Sigma} \end{bmatrix} \right)$$

and

$$\mathbf{y} \sim \mathcal{N} \left(\boldsymbol{\nu}, \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \boldsymbol{\Sigma} \right) \quad \text{where} \quad \boldsymbol{\nu} = \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}.$$

Thus

$$\frac{N_1 N_2}{N_1 + N_2} (\mathbf{y} - \boldsymbol{\nu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\nu}) \leq \chi_p^2(\alpha).$$

is a confidence region for the difference $\boldsymbol{\nu}$ of the two mean vectors, vectors, and a critical region for testing the hypothesis $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ is given by

$$\frac{N_1 N_2}{N_1 + N_2} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) > \chi_p^2(\alpha).$$

Outline

- 1 Noncentral Chi-Squared Distribution
- 2 Hypothesis Testing for the Mean (Covariance is Known)
- 3 The Generalized T^2 -Statistic

Student t -Distribution

Let x_1, \dots, x_N be independently and identically drawn from the distribution $\mathcal{N}(\mu, \sigma^2)$, then the random variable

$$t = \frac{\bar{x} - \mu}{s/\sqrt{N}}$$

has student t -distribution with $N - 1$ degrees of freedom, where

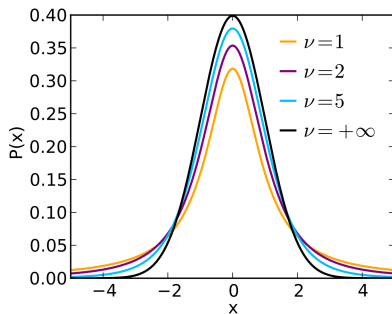
$$\bar{x} = \frac{1}{N} \sum_{\alpha=1}^N x_{\alpha} \quad \text{and} \quad s^2 = \frac{1}{N-1} \sum_{\alpha=1}^N (x_{\alpha} - \bar{x})^2.$$

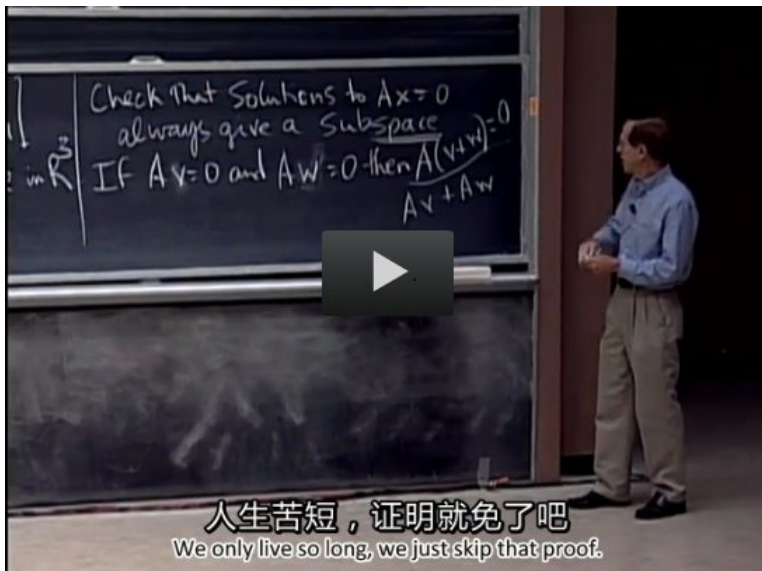
Student t -Distribution

Student's t -distribution has the probability density function given by

$$f(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

where ν is the number of degrees of freedom and Γ is the gamma function.





The Generalized T^2 -Statistic

The multivariate analog of t^2 is

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}),$$

where

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{x}_\alpha \quad \text{and} \quad \mathbf{S} = \frac{1}{N-1} \sum_{\alpha=1}^N (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})^\top.$$

T^2 -Statistic and Likelihood Ratio Criterion

We consider MLE for normal distribution. The likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{pN}{2}} (\det(\boldsymbol{\Sigma}))^{-\frac{N}{2}} \exp \left(-\frac{1}{2} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{\alpha} - \boldsymbol{\mu}) \right).$$

The likelihood ratio criterion is

$$\lambda = \frac{\max_{\boldsymbol{\Sigma} \in \mathbb{S}_p^{++}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{S}_p^{++}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

- 1 The denominator is the maximum over the entire parameter space.
- 2 The numerator is the maximum in the space restricted by the null hypothesis.
- 3 The likelihood ratio test is the procedure of rejecting the null hypothesis when λ is less than a predetermined constant.

T^2 -Statistic and Likelihood Ratio Criterion

We have

$$\lambda^{\frac{2}{N}} = \frac{1}{1 + T^2/(N-1)},$$

where $T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$.

The likelihood ratio test is defined by the critical region (region of rejection)

$$\lambda \leq \lambda_0, \quad (1)$$

where λ_0 is chosen so that the probability of (1) when the null hypothesis is true is equal to the significance level.

The inequality (1) also equivalent to

$$T^2 \geq T_0^2,$$

where $T_0^2 = (N-1)(\lambda_0^{-2/N} - 1)$.

Invariant Property of t^2 -Test

The Student t -test is invariant w.r.t scale transformations if $\mu = 0$

- ① If $x \sim \mathcal{N}(\mu, \sigma^2)$, then $x^* = cx \sim \mathcal{N}(c\mu, c^2\sigma^2)$ for $c > 0$.
- ② The hypothesis $\mathbb{E}[x] = 0$ is equivalent to $\mathbb{E}[cx] = 0$.
- ③ If observations x_α are transformed to $x_\alpha^* = cx_\alpha$, then

$$t^* = \frac{\bar{x}^* - 0}{s^*/\sqrt{N}} = \frac{\bar{x} - 0}{s/\sqrt{N}} = t.$$

Invariant Property of T^2 -Test

The generalized T^2 -test has a similar property.

- ① If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$, then $\mathbf{x}^* = \mathbf{C}\mathbf{x} \sim \mathcal{N}(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top)$, where $\mathbf{C} \succeq \mathbf{0}$.
- ② The hypothesis $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ is equivalent to the hypothesis $\mathbb{E}[\mathbf{x}^*] = \mathbb{E}[\mathbf{C}\mathbf{x}] = \mathbf{0}$.
- ③ If observations \mathbf{x}_α are transformed to $\mathbf{x}_\alpha^* = \mathbf{C}\mathbf{x}_\alpha$, then T^{*2} computed on \mathbf{x}_α^* is the same as T^2 computed on \mathbf{x}_α .

This follows from $\bar{\mathbf{x}}^* = \mathbf{C}\bar{\mathbf{x}}$, $\mathbf{S}^* = \mathbf{C}\mathbf{S}\mathbf{C}^\top$ and the following lemma.

Lemma 1

For any $p \times p$ non-singular matrices \mathbf{C} and \mathbf{H} and any vector \mathbf{k} , we have

$$\mathbf{k}^\top \mathbf{H}^{-1} \mathbf{k} = (\mathbf{C}\mathbf{k})^\top (\mathbf{C}\mathbf{H}\mathbf{C}^\top)^{-1} (\mathbf{C}\mathbf{k}).$$

The F -distribution with d_1 and d_2 degrees of freedom is the distribution of

$$x = \frac{y_1/d_1}{y_2/d_2} = \frac{d_2 y_1}{d_1 y_2}$$

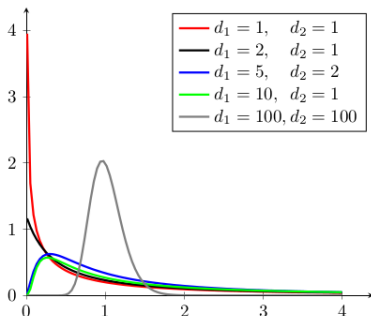
where y_1 and y_2 are independent random variables with Chi-square distributions with respective degrees of freedom d_1 and d_2 .

F-Distribution

The probability density function (pdf) for F-Distribution is

$$f(x; d_1, d_2) = \frac{1}{B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2} x\right)^{-\frac{d_1+d_2}{2}}$$

where $B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$.



Noncentral F -Distribution

If y_1 is a noncentral Chi-squared random variable with noncentrality parameter λ and d_1 degrees of freedom, and y_2 is a (central) Chi-squared random variable with d_2 degrees of freedom that is independent of y_1 , then

$$x = \frac{y_1/d_1}{y_2/d_2}$$

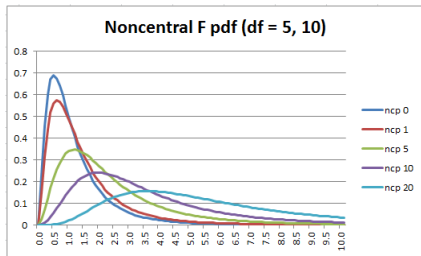
is a noncentral F -distributed random variable.

Noncentral F -Distribution

The probability density function (pdf) for the noncentral F -distribution is

$$f(x; d_1, d_2, \lambda) = \begin{cases} \sum_{k=0}^{\infty} \frac{\exp(-\frac{\lambda}{2})(\frac{\lambda}{2})^k}{B(\frac{d_2}{2}, \frac{d_1}{2} + k) k!} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2} + k} \left(\frac{d_2}{d_2 + d_1 x}\right)^{\frac{d_1 + d_2}{2} + k} x^{\frac{d_1}{2} - 1 + k}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$.



Distribution of T^2 -Statistics

Theorem 2

Let $T^2 = \mathbf{y}^\top \mathbf{S}^{-1} \mathbf{y}$, where \mathbf{y} is distributed according to $\mathcal{N}_p(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ and $n\mathbf{S}$ is independently distributed as $\sum_{\alpha=1}^n \mathbf{z}_\alpha \mathbf{z}_\alpha^\top$ with $\mathbf{z}_1, \dots, \mathbf{z}_n$ independent, each with distribution $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. Then the random variable

$$\frac{T^2}{n} \cdot \frac{n - p + 1}{p}$$

is distributed as a noncentral F -distribution with p and $n - p + 1$ degrees of freedom and noncentrality parameter $\boldsymbol{\nu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\nu}$. If $\boldsymbol{\nu} = \mathbf{0}$, the distribution is central F .

In the example of likelihood ratio criterion, we consider the special case of $\mathbf{y} = \sqrt{N}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top$, $\boldsymbol{\nu} = \sqrt{N}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^\top$ and $n = N - 1$.

Distribution of T^2 -Statistics

Corollary 2

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ be a sample from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^\top \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0).$$

The distribution of

$$\frac{T^2}{N-1} \cdot \frac{N-p}{p}$$

is noncentral F with p and $N-p$ degrees of freedom and noncentrality parameter $N(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$. If $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ then the F -distribution is central.

For large samples the distribution of T^2 given this corollary is approximately valid even if the parent distribution is not normal.

Distribution of T^2 -Statistics

Theorem 3

Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of independently identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let

$$\hat{\mathbf{x}}_N = \frac{1}{N} \sum_{\alpha=1}^N \mathbf{x}_{\alpha}, \quad \hat{\mathbf{S}}_N = \frac{1}{N-1} \sum_{\alpha=1}^N (\mathbf{x}_{\alpha} - \bar{\mathbf{x}})(\mathbf{x}_{\alpha} - \bar{\mathbf{x}})^{\top}$$

and

$$T_N^2 = N(\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0)^{\top} \mathbf{S}_N^{-1} (\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0).$$

Then the limiting distribution of T_N^2 as $N \rightarrow \infty$ is the χ^2 -distribution with p degrees of freedom if $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Distribution of T^2 -Statistics

Theorem 4

Suppose $\mathbf{y}_1, \dots, \mathbf{y}_m$ are independent with \mathbf{y}_α distributed according to $\mathcal{N}(\mathbf{\Gamma}\mathbf{w}_\alpha, \mathbf{\Phi})$, where \mathbf{w}_α is an r -component vector. Let $\mathbf{H} = \sum_{\alpha=1}^m \mathbf{w}_\alpha \mathbf{w}_\alpha^\top$ assumed non-singular, $\mathbf{G} = \sum_{\alpha=1}^m \mathbf{y}_\alpha \mathbf{w}_\alpha^\top \mathbf{H}^{-1}$ and

$$\mathbf{C} = \sum_{\alpha=1}^m (\mathbf{y}_\alpha - \mathbf{G}\mathbf{w}_\alpha)(\mathbf{y}_\alpha - \mathbf{G}\mathbf{w}_\alpha)^\top = \sum_{\alpha=1}^m \mathbf{y}_\alpha \mathbf{y}_\alpha^\top - \mathbf{G}\mathbf{H}\mathbf{G}^\top.$$

Then \mathbf{C} is distributed as

$$\sum_{\alpha=1}^{m-r} \mathbf{u}_\alpha \mathbf{u}_\alpha^\top$$

where $\mathbf{u}_1, \dots, \mathbf{u}_{m-r}$ are independently distributed according to $\mathcal{N}(\mathbf{0}, \mathbf{\Phi})$ independently of \mathbf{G} .