27-08-25

$$\max_{\mu, \Sigma} L(\mu, \Sigma)$$

$\hat{\mu} = \bar{x}$: optimal solⁿ

$\hookrightarrow$ MLE of μ.

$$\max_{\Sigma} L(\hat{\mu}, \Sigma) = \max_{\Sigma} \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{n}{2} \text{Trace}(\Sigma^{-1} S_n)\right)$$

$$= \max_{\eta} \frac{\prod_{i=1}^{p} \eta_i^{n/2}}{(2\pi)^{np/2} |S_n|^{n/2}} \exp\left(-\frac{n}{2} \sum_{i=1}^{p} \eta_i\right)$$

$$= \frac{1}{(2\pi)^{np/2} |S_n|^{n/2}} \max_{\eta} \prod_{i=1}^{p} \eta_i^{n/2} \exp\left(-\frac{n}{2} \eta_i\right)$$

$$= \frac{1}{(2\pi)^{np/2} |S_n|^{n/2}} \prod_{i=1}^{p} \max_{\eta_i} \eta_i^{n/2} \exp\left(-\frac{n}{2} \eta_i\right)$$

Pair of one variable $\eta_i > 0$

$\max_{\eta = (\eta_1, \eta_2)} g(\eta_1, \eta_2) = g_1(\eta_1)\, g_2(\eta_2)$: sep^r of variables

Consider a f^n

$$h(x) = x^{n/2} \exp\left(-\frac{n}{2} x\right), \quad x > 0$$

want to $\max_{x>0} h(x)$ by std. calculus.

$$h'(x) = x^{n/2} \exp\left(-\frac{n}{2} x\right) \times \left(-\frac{n}{2}\right) + \exp\left(-\frac{n}{2} x\right) \frac{n}{2} x^{n/2 - 1} = 0$$

$$\Rightarrow x^{n/2} (-1 + x^{-1}) = 0 \Rightarrow x = 1 \rightarrow \text{maxima}$$

(check)

For likelihood f^n L it means each $\eta_i = 1$, $1 \leq i \leq p$

$$\text{max value of } L(\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2} |S_n|^{n/2}} \exp\left(-\frac{n}{2} p\right)$$

$$\therefore \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{n}{2} \text{Trace}(\Sigma^{-1} S_n)\right) \leq \frac{1}{(2\pi)^{np/2} |S_n|^{n/2}} \exp\left(\frac{-np}{2}\right)$$

$\forall$ symmetric pd Σ matrices.

The upper bound is attained iff $\Sigma = S_n$.

$$S_n^{1/2} \Sigma^{-1} S_n^{1/2} = I_{p \times p} \Rightarrow \Sigma^{-1} = (S_n^{1/2})^{-1} (S_n^{1/2})^{-1} = S_n^{-1}$$

$$\Rightarrow \Sigma = S_n$$

optimal sol^n of the problem

$$\max_{\substack{\Sigma \\ \text{symm. pd} \\ p \times p}} L(\hat{\mu}, \Sigma)$$

is $\hat{\Sigma} = S_n \longrightarrow$ MLE of Σ.

Recall $\quad S_n = \frac{n-1}{n} S$

$S \hookrightarrow p \times p$

variance-covariance matrix of sample data when division is by $n-1$.

unbiased	MLE
$\mu \rightarrow \bar{X}$	$\mu \rightarrow \bar{X}$
$\Sigma \rightarrow S$	$\Sigma \rightarrow S_n = \frac{n-1}{n} S$

Note :- max value $= L(\hat{\mu}, \hat{\Sigma})$

$$= \frac{1}{(2\pi)^{np/2} |S_n|^{n/2}} \exp\left(\frac{-np}{2}\right)$$

$(2\pi)^{np/2} > 0$, $\exp\left(\frac{-np}{2}\right) > 0$

$$\propto \frac{1}{|S_n|^{n/2}} = \frac{1}{\left|\frac{n-1}{n} S\right|^{n/2}}$$

$\Rightarrow L(\hat{\mu}, \hat{\Sigma}) \propto \frac{1}{|S|^{n/2}}$ & $|S|$ = generalized variance of sample data.

The smaller the generalized variance in the sample data, the more accurate is the MLE estimation of $\hat{\Sigma}$ by S_n.

Basic stats recall

$\boxed{p=1}$

sampling

population
$\downarrow$

$X_1, \dots, X_n \longrightarrow \underbrace{\text{random}}_{\text{iid}}$ sample

mean $\underset{1\times 1}{\mu}$, variance $\underset{1\times 1}{\sigma^2} > 0$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow rv$$

$$E(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n\mu = \mu$$

$$Var(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}$$

independent samples

Suppose population is $N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \sigma^2/n)$

$$\Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \sim N(0,1) \quad (1)$$

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \rightarrow$ sample variance

$\downarrow$

random variable

We want to see the multivariate case of it.

If we go back to case $p > 1$ & see our previous notes.

$$\begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array} \begin{array}{c} 1 \quad 2 \quad \cdots \quad p \\ \left[\begin{array}{c} \longrightarrow \\ \longrightarrow \\ \\ \longrightarrow \end{array} \right] \end{array}$$

$\underset{p\times 1}{\underset{\sim}{X}_1}, \ldots, \underset{p\times 1}{\underset{\sim}{X}_n}$ be a random sample.
$\underline{\underline{\text{iid}}}$

Let the population be p-dimensional normal $N_p(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma})$

$$\Rightarrow \underset{\sim}{X}_i \sim N_p(\mu, \Sigma), \; 1 \le i \le n$$

$$\Rightarrow \underset{\sim}{V} = \sum_{i=1}^{n} c_i \underset{\sim}{X}_i \text{ random vector}$$

and $\underset{\sim}{V} \sim N_p\left(\sum_{i=1}^{n} c_i \mu_i, \; \sum_{i=1}^{n} c_i^2 \Sigma\right)$

In particular, we take $c_i = \frac{1}{n}$ $\forall$ $i = 1, 2, \dots, n$ {uniform sample}

$$\underset{\sim}{V} = \bar{X} \;\&\; \underset{\sim}{V} \sim N_p\left(\frac{\mu}{n}\sum_{i=1}^{n} 1, \frac{\Sigma}{n^2}\sum_{i=1}^{n} 1\right)$$

$$= N_p\left(\mu, \frac{\Sigma}{n}\right) \quad - \; (2)$$

If $p=1$, then $(2) \Rightarrow (1)$, $\therefore$ (2) is a generalization of (1) in p-dimension.

$\boxed{29-08-25}$

$p=1$

Sample $X_1, \dots, X_n$ random (iid)

$$\bar{X} \sim N(\mu, \sigma^2)$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Consider $p > 1$

$$\bar{X} \sim N_p(\mu_{p\times 1}, \Sigma_{p\times p})$$

if $y \sim \chi^2_\nu$

then

$y = z_1^2 + \dots + z_\nu^2$,

where $z_i \sim N(0,1)\ \forall i$

& $z_1, \dots, z_\nu$ are independent rv.

if $p = 1$

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$(n-1)\frac{s^2}{\sigma^2} = z_1^2 + \dots + z_{n-1}^2, \quad z_i \sim N(0,1)\ \forall i \text{ \& }$$

z_i's are independent rvs.

$$\Rightarrow (n-1)s^2 = (\sigma z_1)^2 + \dots + (\sigma z_{n-1})^2.$$

$$= y_1^2 + \dots + y_{n-1}^2$$

$y_i = \sigma z_i \sim N(0, \sigma^2)\ \forall i$ & $y_1, \dots, y_{n-1}$ are ind. rvs.

Getting motivation from this expression, for general $p > 1$,

$$\underset{p\times p}{(n-1)S} = \underset{p\times 1}{z_1}\,\underset{1\times p}{z_1^T} + \dots + \underset{p\times 1}{z_{n-1}}\,\underset{1\times p}{z_{n-1}^T}$$

$$\underset{p\times 1}{z_i} \sim N_p(0, \underset{p\times p}{\Sigma})\ \forall\ 1 \le i \le n-1$$

and $z_1, z_2, \dots, z_{n-1}$ are independent r.vs.

n : sample size

S : sample variance-covariance matrix.

Wishart Distribution

Let A be a $p \times p$ symmetric random matrix.

If we can write

rank 1 matrix.

$$A = z_1 z_1^T + \dots + z_m z_m^T$$

$$z_i \sim N_p(0, \Sigma), \; 1 \le i \le m \text{ and}$$

$z_1, \dots, z_m$ are independent random vectors.

Then, we say that

$$A \sim W_p(m, \Sigma) \text{ or } W_p(\Sigma, m).$$

i.e., A is p-dim. Wishart distributed & m is degree of freedom.

In case $\Sigma = I_{p \times p}$, then $W_p(m, I)$ is called as std. form of Wishart distribution.

$$\therefore (n-1)S \sim W_p(n-1, \Sigma)$$

Let $p=1$,

$$A \sim W_1(m, \Sigma)$$

$$\Rightarrow A = z_1^2 + \dots + z_m^2, \quad z_i \sim N(0, \sigma^2); \ z_i\text{'s are ind. r.v.}$$

$$\sim \sigma^2 \chi^2_{(m)}$$

$$\Rightarrow \frac{A}{\sigma^2} \sim \chi^2_{(m)}$$

We can take Wishart distribution as an extension of χ^2-distribution in p-dimension.

pdf of $W_p(m, \Sigma)$ is

$$f_{W_p}(A) = \begin{cases} \dfrac{|A|^{\frac{1}{2}(m-p-1)} \exp\left(-\frac{1}{2}\,\mathrm{trace}(\Sigma^{-1}A)\right)}{2^{\frac{mp}{2}} \pi^{\frac{p(p-1)}{4}} |\Sigma|^{m/2} \prod_{i=1}^{p} \underbrace{\Gamma\left(\frac{m+1-i}{2}\right)}_{\text{gamma } f^n.}} & \text{if } A \text{ is pd.} \\ 0 & \text{otherwise} \end{cases}$$

1. $A_1 \sim W_p(m_1, \Sigma)$

$A_2 \sim W_p(m_2, \Sigma)$

$\Rightarrow A_1 + A_2 \sim W_p(m_1 + m_2, \Sigma)$

2. $A \sim W_p(m, \Sigma)$, $\vec{a} \in \mathbb{R}^p$

$$\vec{a}^T A \vec{a} = \vec{a}^T (z_1 z_1^T + \ldots + z_m z_m^T) \vec{a}$$

$$= (\vec{a}^T z_1)(z_1^T \vec{a}) + \ldots + (\vec{a}^T z_m)(z_m^T \vec{a})$$

$$= y_1^2 + \ldots + y_m^2, \text{ where } y_i = \vec{a}^T z_i.$$

and, $z_i \sim N_p(0, \Sigma)$

$$\Rightarrow \vec{a}^T z_i \sim N_1(0, \vec{a}^T \Sigma \vec{a})$$

$$\Rightarrow y_i \sim N_1(0, \vec{a}^T \Sigma \vec{a}) \quad \forall i$$

$$\Rightarrow \frac{y_i}{\sqrt{a^T \Sigma a}} \sim N_1(0,1) \quad \forall i \text{ provided } \vec{a}^T \Sigma \vec{a} \neq 0.$$

$$\frac{\vec{a}^T A \vec{a}}{a^T \Sigma a} = \underbrace{\left(\frac{y_1}{\sqrt{a^T \Sigma a}}\right)^2}_{N(0,1)} + \cdots + \underbrace{\left(\frac{y_m}{\sqrt{a^T \Sigma a}}\right)^2}_{N(0,1)}$$

$$\Rightarrow \frac{\vec{a}^T A \vec{a}}{a^T \Sigma a} \sim \chi^2_{(m)}$$

$$\boxed{A \sim W_p(m, \Sigma),\ \vec{a} \in \mathbb{R}^p,\ \vec{a} \neq 0 \Rightarrow \frac{\vec{a}^T A \vec{a}}{\vec{a}^T \Sigma \vec{a}} \sim \chi^2_{(m)}}$$

Converse is NOT true., i.e., if $\frac{a^T A a}{a^T \Sigma a} \sim \chi^2_{(m)}$ $\forall \vec{a} \in \mathbb{R}^p$, $\vec{a} \neq 0$.

then, we cannot say that A is Wishart.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

$$\vec{a_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vec{a_2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\left.\begin{array}{l} \vec{a_1}^T A a_1 = a_{11} \\ \vec{a_2}^T A a_2 = a_{22} \end{array}\right\} \text{cannot get } a_{12} \text{ without changing the basis.}$$

Counter-example:

Let $A = W + S$

where $W \sim W_p(m, \Sigma)$

$S = -S^T$ (skew symm.)

$p \times p$

$\because W$ is random $\Rightarrow A$ is random.

Let $\vec{a} \in \mathbb{R}^p \Rightarrow \vec{a}^T A a = \vec{a}^T W a + \vec{a}^T S a$

$$= a^T W a + 0$$

$$= a^T W a.$$

$$\Rightarrow \frac{\vec{a}^T A \vec{a}}{\vec{a} \Sigma a} = \frac{a^T W a}{a^T \Sigma a} \sim \chi^2_{(m)}$$

$A = z_1 z_1^T + \dots + z_m z_m^T + S$

$z_i z_i^T$ is symm. matrix & S is skew symm

$\therefore \nexists$ any $y \in \mathbb{R}^p$ s.t. $S = y y^T$ type
($y \neq 0$)

$\therefore$ A is not wishart (not even symm.)

3 - classes missed

9-9-25

$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$.

$$T^2 = n(\bar{x} - \mu_0)^T S^{-1} (\bar{x} - \mu_0)$$

random variable

$$P_{H_0}\{X_{sample} : T^2 > c\} = \alpha$$

c is some +ve constant

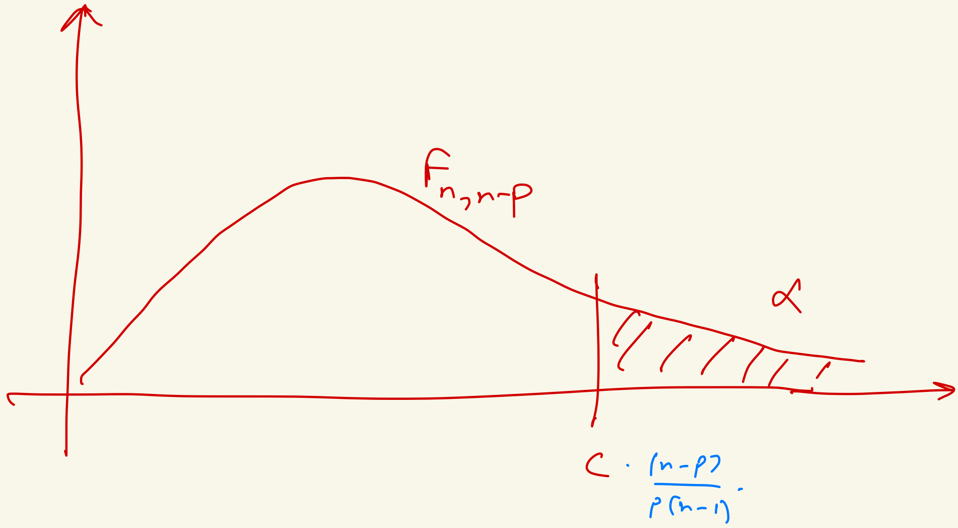
$\in (0,1)$ given level of significance.

$$\frac{n-p}{p(n-1)} T^2 \sim F_{p, n-p, \alpha}$$

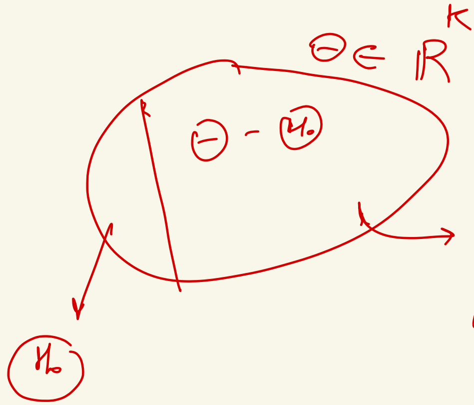
$$P_{H_0}\left\{x : \frac{n-p}{p(n-1)} T^2 > \frac{n-p}{p(n-1)} c\right\} = \alpha$$

critical region

$$c = \frac{p(n-1)}{n-p} F_{p, n-p, \alpha}.$$



Likelihood ratio test (Neymann Pearson Lemma)



θ estimate from sample

$\theta \in \Theta \subseteq \mathbb{R}^k$.

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$.

$$\Lambda = \frac{\sup_{\theta \in H_0} L(\underline{\theta})}{\sup_{\underline{\theta} \in \Theta} L(\underline{\theta})}$$

$$0 \leq \Lambda \leq 1$$

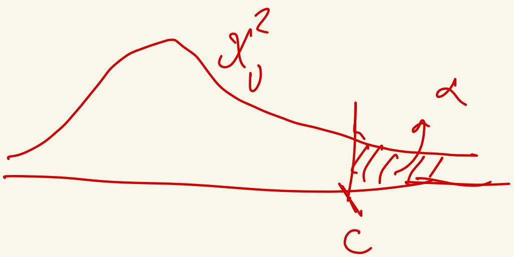
If $\Lambda \to 1$, then it indicates that the sample statistics for $\theta = \theta_0$ is supported by almost all samples.

for critical region
(H_0 is true but reject H_0)
we have to find out const $c \to 0^+$ such that

$P_{H_0}(\Lambda < c) = \alpha$, where $\alpha \in (0,1)$ is the level of significance or area of critical region.

Distribution of Λ is not precisely known, but what we know is that if sample size $n \to \infty$, then asymptotically

$$-2 \ln \Lambda \xrightarrow{D} \chi^2_{\nu}, \text{ where } \nu = \dim \Theta - \dim \Theta_0$$



Aim : To simplify the expression of Λ (Wilk's lambda)

Recall from notes, assume population to be $N_p(\mu, \Sigma)$, Σ is not known & $\mu_{p \times 1}$ is to be estimated using sampling.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0.$$

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{n}{2}\left[\operatorname{trace}(\Sigma^{-1} S_n) + (\bar{x}-\mu)^T \Sigma^{-1} (\bar{x}-\mu)\right]\right)$$

MLE of $\mu = \hat{\mu} = \bar{x}$

MLE of $\Sigma = \hat{\Sigma} = S_n = \frac{n-1}{n} S$

$$\therefore \sup_{(\mu, \Sigma)} L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(\frac{-np}{2}\right)$$

Σ: symm. pd matrices of $p \times p$

$\mu \in \mathbb{R}^p$.

$$= L(\hat{\mu}, \hat{\Sigma})$$

For numerator in Λ, we assume H_0 is true so $\mu = \mu_0$.

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(\frac{-1}{2} \sum_{i=1}^{n} (x_i - \mu_0)^T \Sigma^{-1} (x_i - \mu_0)\right)$$

original defn.

$$\sup L(\mu, \Sigma) = ?$$

Σ : $p \times p$ symm pd matrices

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \operatorname{trace}\left(\sum_{i=1}^{n} (x_i - \mu_0)^T \Sigma^{-1} (x_i - \mu_0)\right)\right)$$

$$= \quad '' \quad \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \operatorname{trace}\left[(x_i - \mu_0)^T \Sigma^{-1} (x_i - \mu_0)\right]\right)$$

$$= \quad '' \quad \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \operatorname{trace}\left[\Sigma^{-1} (x_i - \mu_0)^T (x_i - \mu_0)\right]\right)$$

$$= \quad '' \quad \exp\left(-\frac{n}{2} \sum_{i=1}^{n} \operatorname{trace}\left(\frac{\Sigma^{-1} (x_i - \mu_0)^T (x_i - \mu_0)}{n}\right)\right)$$

$$= \quad '' \quad \exp\left(-\frac{n}{2} \operatorname{trace} \Sigma^{-1} \sum_{i=1}^{n} \frac{(x_i - \mu_0)(x_i - \mu_0)^T}{n}\right)$$

$$= \quad '' \quad \exp\left(-\frac{n}{2} \operatorname{trace}(\Sigma^{-1} S_0)\right)$$

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)(x_i - \mu_0)^T$$

$$\Rightarrow \quad L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{n}{2} \operatorname{trace}(\Sigma^{-1} S_0)\right)$$

$$\underset{\Sigma}{\text{maximize}} \; L(\mu, \Sigma)$$

Apply the same steps as done while calculating MLE of Σ (using eigenvalues of $\Sigma^{-1} S_0$), we can finally arrive at that, the $\underset{\Sigma}{\text{maximum}} \; L(\mu_0, \Sigma)$ is attained when $\Sigma = S_0$.

$$\therefore \text{Maximum value} = L(\mu_0, S_0) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left(-\frac{np}{2}\right) \quad \text{—— (2)}$$

$$\therefore \Lambda = \frac{(2)}{(1)}$$

$$= \frac{|S_n|^{n/2}}{|S_0|^{n/2}}.$$

$$\Rightarrow \quad \frac{2}{n} \ln \Lambda = \ln|S_n| - \ln|S_0|$$

$$\Rightarrow \boxed{-2\ln\Lambda = n\left(\ln|S_0| - \ln|S_n|\right)}$$

$$S_0 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)(x_i - \mu_0)^T$$

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T.$$