Maximum likelihood estimation

$\underset{\sim}{x}_i \sim N_d(\underset{\sim}{\mu}, \Sigma)$, $i=1,\dots,n$, independent, $n-1 \geq d$

$$L(\underset{\sim}{\mu}, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-\frac{d}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})\right)$$

$$= (2\pi)^{-\frac{nd}{2}} (\det \Sigma)^{-\frac{n}{2}} \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})\right)$$

$$\ln L(\underset{\sim}{\mu}, \Sigma) = -\tfrac{nd}{2}\ln(2\pi) - \tfrac{n}{2}\ln\det\Sigma - \tfrac{1}{2}\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})$$

$$\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu}) = \sum_{i=1}^{n} \operatorname{tr}\left((\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})\right)$$

$$= \sum_{i=1}^{n} \operatorname{tr}\left(\Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})^T\right)$$

$$= \operatorname{tr}\left(\Sigma^{-1} \sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})^T\right)$$

$$\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})^T = \sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\bar{x}} + \underset{\sim}{\bar{x}} - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\bar{x}} + \underset{\sim}{\bar{x}} - \underset{\sim}{\mu})^T$$

$$= \sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\bar{x}})(\underset{\sim}{x}_i - \underset{\sim}{\bar{x}})^T + 0 + 0 + n(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})^T$$

$$= Q + n(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})^T$$

$$\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu}) = \operatorname{tr}\left(\Sigma^{-1} Q + n\Sigma^{-1}(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})(\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})^T\right)$$

$$= \operatorname{tr}(\Sigma^{-1}Q) + n\operatorname{tr}\left((\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{\bar{x}} - \underset{\sim}{\mu})\right)$$

$$\geq \operatorname{tr}(\Sigma^{-1}Q)$$

$$\ln L(\underset{\sim}{\mu}, \Sigma) \leq -\tfrac{nd}{2}\ln(2\pi) - \tfrac{n}{2}\ln\det\Sigma - \tfrac{1}{2}\operatorname{tr}(\Sigma^{-1}Q)$$

equality for $\underset{\sim}{\mu} = \underset{\sim}{\bar{x}}$, thus $\hat{\underset{\sim}{\mu}} = \underset{\sim}{\bar{x}}$ for all $\Sigma > 0$

$$\ln L(\hat{\underset{\sim}{\mu}}, \Sigma) = -\tfrac{nd}{2}\ln(2\pi) - \tfrac{n}{2}\left(\ln\det\Sigma + \operatorname{tr}\left(\Sigma^{-1}\tfrac{Q}{n}\right)\right)$$

Take $f(\Sigma) = \ln\det\Sigma + \operatorname{tr}(\Sigma^{-1}A)$, $A > 0$

in the following we use A5.5, A5.4, A1.4, A5.7 and A5.1

$\Sigma^{-1}A$ and $\Sigma^{-\frac{1}{2}}A\Sigma^{-\frac{1}{2}}$ has the same eigenvalues λ_i

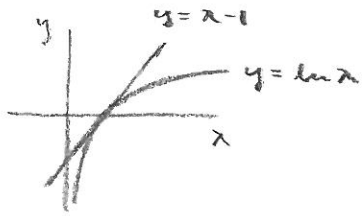
$\Sigma^{-\frac{1}{2}}A\Sigma^{-\frac{1}{2}} > 0 \Rightarrow \lambda_i > 0$, $i = 1, \dots, d$

view the difference

$$f(\Sigma) - f(A) = \ln\det\Sigma + \operatorname{tr}(\Sigma^{-1}A) - \ln\det A - \operatorname{tr} I_d$$

$$= -\ln\det(\Sigma^{-1}A) + \operatorname{tr}(\Sigma^{-1}A) - d$$

$$= -\ln\prod_i \lambda_i + \sum_i \lambda_i - d \qquad \text{cf. A1.2 (a), (b)}$$



$$= \sum_i(-\ln\lambda_i + \lambda_i - 1)$$

$$= \sum_i((\lambda_i - 1) - \ln\lambda_i) \geq 0$$

equality for $\lambda_i = 1$, $i = 1, \dots, d$, thus $\Sigma = A$

(see ** page 7)

$$\ln L(\hat{\mu}, \Sigma) \leq -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\left(\ln\det\frac{Q}{n} + \operatorname{tr} I_d\right)$$

equality for $\Sigma = \frac{Q}{n}$, thus $\hat{\Sigma} = \frac{Q}{n}$

$$\ln L(\hat{\mu}, \hat{\Sigma}) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln\det\hat{\Sigma} - \frac{nd}{2}$$

$$L(\hat{\mu}, \hat{\Sigma}) = (2\pi e)^{-\frac{nd}{2}}(\det\hat{\Sigma})^{-\frac{n}{2}}$$

Theorem 3.1

(i) $\left.\begin{array}{l} \bar{x} \sim N_d(\mu, \frac{\Sigma}{n}) \\ Q \sim W_d(n-1, \Sigma) \end{array}\right\}$ independent $\qquad (Q = (n-1)S)$

(ii) $T^2 = n(\bar{x} - \mu)^T S^{-1}(\bar{x} - \mu) \sim T^2(d, n-1)$

Proof:

(i) Let $y_i = x_i - \mu$, hence $y_i \sim N_d(0, \Sigma)$, $i = 1, \dots, n$ independent

and $P = \frac{1}{n} 1_n 1_n^T$

Now using corollary 3 to theorem 2.4 :

$$\left.\begin{aligned} \bar{y} &= Y^T\left(\tfrac{1}{n}\mathbf{1}_n\right) \sim N_d\left(\mathbf{0}, \tfrac{\Sigma}{n}\right) \\ Q &= Y^T(I_n - P)Y \sim W_d(n-1, \Sigma) \end{aligned}\right\} \text{ independent}$$

as $\operatorname{Var}\bar{y} = \frac{\Sigma}{n}$ cf. corollary to lemma 2.3 (ii)

$$\operatorname{rank}(I_n - P) = n - \operatorname{rank} P = n - 1$$

$$(I_n - P)\tfrac{1}{n}\mathbf{1}_n = \tfrac{1}{n}\mathbf{1}_n - \tfrac{1}{n^2}\mathbf{1}_n\mathbf{1}_n^T\mathbf{1}_n = \mathbf{0}$$

Consequently

$$\left.\begin{aligned} \bar{x} &= \bar{y} + \mu \sim N_d\left(\mu, \tfrac{\Sigma}{n}\right) \\ Q &= X^T(I_n - P)X \sim W_d(n-1, \Sigma) \end{aligned}\right\} \text{ independent}$$

as Q is unchanged, cf. formel (1.11)

(ii) $$T^2 = n(\bar{x} - \mu)^T S^{-1}(\bar{x} - \mu) = n(\bar{x} - \mu)^T\left(\frac{Q}{n-1}\right)^{-1}(\bar{x} - \mu)$$

$$= n(n-1)(\bar{x} - \mu)^T Q^{-1}(\bar{x} - \mu) \sim T^2(d, n-1)$$

cf. corollary to theorem 2.8

Hotellings T^2 test

When $H_0: \mu = \mu_0$ is true

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

$$= \frac{(n-1)d}{n-d} F(d, n-d)$$

i.e. reject when $\frac{n-d}{(n-1)d} T^2_{obs} > F_{1-\alpha}(d, n-d)$

Tecnique: Find the Cholesky decomposition $Q = U^T U$, solve $U^T z = \bar{x} - \mu_0$ wrt z, calculate $n(n-1)\, \underset{\sim}{z}^T \underset{\sim}{z}$

$$= n(n-1)(\underset{\sim}{\bar{x}} - \mu_0)^T U^{-1} (U^T)^{-1} (\underset{\sim}{\bar{x}} - \mu_0)$$

$$= n(n-1)(\underset{\sim}{\bar{x}} - \mu_0)^T Q^{-1} (\underset{\sim}{\bar{x}} - \mu_0)$$

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

Example

Weight (kg)	Height (mm)	Weight (kg)	Height (mm)
71.0	1629	59.5	1513
56.5	1569	61.0	1653
56.0	1561	57.0	1566
61.0	1619	57.5	1580
65.0	1566	74.0	1647
62.0	1639	72.0	1620
53.0	1494	62.5	1637
53.0	1568	68.0	1528
65.0	1540	63.4	1647
57.0	1530	68.0	1605
66.5	1622	69.0	1625
59.1	1486	73.0	1615
64.0	1578	64.0	1640
69.5	1645	65.0	1610
64.0	1648	71.0	1572
56.5	1521	60.2	1534
57.0	1547	55.0	1536
55.0	1505	70.0	1630
57.0	1473	87.0[b]	1542[b]
58.0	1538		

EXAMPLE 3.1 In Table 3.1 we have bivariate observations on the weights and heights of 39 Peruvian Indians. Suppose we wish to test $H_0: \boldsymbol{\mu} = (63.64, 1615.38)' = \boldsymbol{\mu}_0$ (corresponding to 140 lb and 63 in., respectively). A plot of the data (Fig. 3.1) suggests that the last item is an outlier. Other plots (see Examples 4.3 and 4.6 in Section 4.3) also indicate that the normality assumption is tenable, though there is some suggestion of kurtosis. Ignoring the last item (i.e., $n = 38$), we write

$$\bar{\mathbf{x}} = \begin{pmatrix} 62.5316 \\ 1579.8900 \end{pmatrix}, \qquad \bar{\mathbf{x}} - \boldsymbol{\mu}_0 = \begin{pmatrix} -1.1084 \\ -35.4900 \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} 1331.62 & 7304.83 \\ 7304.83 & 104118 \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} 36.4914 & 200.179 \\ 0 & 253.073 \end{pmatrix}.$$

Solving

$$\begin{pmatrix} 36.4914 & 0 \\ 200.179 & 253.073 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} -1.1084 \\ -35.4900 \end{pmatrix}$$

gives $\mathbf{z}' = (-0.0303749, -0.116191)$. Hence

$$T_0^2 = n(n-1)\mathbf{z}'\mathbf{z} = 20.2798$$

and

$$F_0 = \frac{T_0^2}{n-1} \frac{n-d}{d} = \frac{20.2798}{37} \frac{36}{2} = 9.87.$$

Since $F_{2,30}^{0.001} = 8.77$, $F_0 > F_{2,30}^{0.001} > F_{2,36}^{0.001}$ and we reject H_0 at the 0.1% level of significance. If the outlier is included, $F_0 = 10.50$. □

Linear transformation of observations and hypotheses

$$\underset{\sim}{z} = F\underset{\sim}{x} + \underset{\sim}{c} \qquad E\underset{\sim}{z} = F E\underset{\sim}{x} + \underset{\sim}{c} = F\underset{\sim}{\mu} + \underset{\sim}{c} = \underset{\sim}{\varphi}$$

$$H_0: \underset{\sim}{\varphi} = \underset{\sim}{\varphi}_0 = F\underset{\sim}{\mu}_0 + \underset{\sim}{c}$$

$$T_z^2 = n(\bar{\underset{\sim}{z}} - \underset{\sim}{\varphi}_0)^T S_z^{-1} (\bar{\underset{\sim}{z}} - \underset{\sim}{\varphi}_0) \qquad \text{(see * page 7)}$$

$$= n(F\bar{\underset{\sim}{x}} + \underset{\sim}{c} - F\underset{\sim}{\mu}_0 - \underset{\sim}{c})^T (FSF^T)^{-1} (F\bar{\underset{\sim}{x}} + \underset{\sim}{c} - F\underset{\sim}{\mu}_0 - \underset{\sim}{c})$$

$$= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)^T F^T (F^T)^{-1} S^{-1} F^{-1} F (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)$$

$$= n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)^T S^{-1} (\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0) = T^2$$

Theorem 3.2 :

T^2 is equivalent to the likelihood ratio test statistic for $H_0: \underset{\sim}{\mu} = \underset{\sim}{\mu}_0$ versus $H_1: \underset{\sim}{\mu} \neq \underset{\sim}{\mu}_0$

Proof :

$$\ln L(\Sigma) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln\det\Sigma - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_{i=1}^{n}(\underset{\sim}{x}_i - \underset{\sim}{\mu}_0)(\underset{\sim}{x}_i - \underset{\sim}{\mu}_0)^T\right)$$

$$= -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\left(\ln\det\Sigma + \operatorname{tr}\left(\Sigma^{-1}\frac{Q_0}{n}\right)\right)$$

$$\leq -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\left(\ln\det\frac{Q_0}{n} + \operatorname{tr} I_d\right)$$

equality for $\Sigma = \frac{Q_0}{n}$, thus $\hat{\Sigma}_0 = \frac{Q_0}{n}$

$$\ln L(\hat{\Sigma}_0) = -\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln\det\hat{\Sigma}_0 - \frac{nd}{2}$$

$$\Rightarrow \quad L(\hat{\Sigma}_0) = (2\pi e)^{-\frac{nd}{2}} (\det\hat{\Sigma}_0)^{-\frac{n}{2}}$$

$$\ell = \frac{\sup_{\Sigma} L(\Sigma)}{\sup_{\underset{\sim}{\mu},\Sigma} L(\underset{\sim}{\mu}, \Sigma)} = \frac{(2\pi e)^{-\frac{nd}{2}} (\det\hat{\Sigma}_0)^{-\frac{n}{2}}}{(2\pi e)^{-\frac{nd}{2}} (\det\hat{\Sigma})^{-\frac{n}{2}}} = \left(\frac{\det\hat{\Sigma}}{\det\hat{\Sigma}_0}\right)^{\frac{n}{2}}$$

$$\ell^{\frac{2}{n}} = \frac{\det Q}{\det Q_0} = \frac{\det Q}{\det\left(Q + n(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)(\bar{\underset{\sim}{x}} - \underset{\sim}{\mu}_0)^T\right)}$$

$$\ell^{\frac{2}{n}} = \frac{1}{\det\left(I_d + n\,Q^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})^T\right)}$$

$$= \frac{1}{\det\left(I_1 + n\,(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})^T Q^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})\right)}$$ cf. hint to ex. 2.12 a

$$= \frac{1}{1 + n\,(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})^T Q^{-1}(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})}$$

$$= \frac{1}{1+\frac{1}{n-1}T^2} \quad \Rightarrow \quad T^2 = (n-1)\left(\ell^{-\frac{2}{n}} - 1\right)$$

Note that small values of ℓ correspond to large values of T^2.

Relation to the usual t-test:

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

$$= n \sup_{\underset{\sim}{\ell}} \frac{\left(\underset{\sim}{\ell}^T(\bar{\underset{\sim}{x}}-\underset{\sim}{\mu_0})\right)^2}{\underset{\sim}{\ell}^T S\,\underset{\sim}{\ell}}$$ cf. A 7.6

$$= \sup_{\underset{\sim}{\ell}} \left(\frac{\sqrt{n}\,(\underset{\sim}{\ell}^T\bar{\underset{\sim}{x}} - \underset{\sim}{\ell}^T\underset{\sim}{\mu_0})}{\sqrt{\underset{\sim}{\ell}^T S\,\underset{\sim}{\ell}}}\right)^2$$

$$= \sup_{\underset{\sim}{\ell}} t_\ell^2,$$ where $t_\ell \sim t(n-1)$ is the usual statistic for testing the hypothesis $H_{0\ell}: \underset{\sim}{\ell}^T\underset{\sim}{\mu} = \underset{\sim}{\ell}^T\underset{\sim}{\mu_0}$; $N(\underset{\sim}{\ell}^T\underset{\sim}{\mu}, \underset{\sim}{\ell}^T\Sigma\underset{\sim}{\ell})$

U-∩-test variable

Note that $H_0 = \bigcap_\ell H_{0\ell}$ and $H_1 : \bigcup_\ell H_{1\ell}$

Acceptance area for $H_{0\ell}$ is $\{X \mid t_\ell^2 \leq k\}$

$k = \left(t_{1-\frac{\alpha}{2}}(n-1)\right)^2$

Acceptance area for H_0:

$$\bigcap_\ell \{X \mid t_\ell^2 \leq k\} = \{X \mid \sup_{\underset{\sim}{\ell}} t_\ell^2 \leq k\}$$

$$= \{X \mid T^2 \leq k\}$$

i.e. again Hotellings T^2 as test statistic

$$* \quad Q_z = \sum_{i=1}^{n} (\underset{\sim}{z}_i - \bar{\underset{\sim}{z}})(\underset{\sim}{z}_i - \bar{\underset{\sim}{z}})^T$$

$$= \sum_{i=1}^{n} (F\underset{\sim}{x}_i + \underset{\sim}{c} - F\bar{\underset{\sim}{x}} - \underset{\sim}{c})(F\underset{\sim}{x}_i + \underset{\sim}{c} - F\bar{\underset{\sim}{x}} - \underset{\sim}{c})^T$$

$$= \sum_{i=1}^{n} F(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})^T F^T$$

$$= F\left(\sum_{i=1}^{n} (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})(\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})^T\right) F^T$$

$$= FQF^T$$

$$\Rightarrow S_z = FSF^T$$

$$** \quad \exists\, T \text{ (orthogonal)}: \quad T^T(\Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}})T = I_d$$

$$\Leftrightarrow \quad \Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}} = T I_d T^T = I_d$$

$$\Leftrightarrow \quad \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} = \Sigma^{-\frac{1}{2}} I_d \Sigma^{\frac{1}{2}}$$

$$\Leftrightarrow \quad \Sigma^{-1} A = I_d$$

$$\Leftrightarrow \quad \Sigma = A$$