In low dimensional statistics

$\frac{p}{n} \to 0$

In high dimensional statistics, could have $p = p_n = n^{\tau+1}$ with τ a large integer, then $\frac{p_n}{n} = \frac{n^{\tau+1}}{n} = n^{\tau}$, and $\hat{\underline{\theta}}$ is not a good estimator of $\underline{\theta}$.

Under regularity conditions that seem to be much too strong, it can sometimes be shown that $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta}) \xrightarrow{P} \underline{0}$.

For example, if nearly all of the $\theta_i = 0$ and the method can find those values, make $\hat{\theta}_i = 0$ if $\theta_i = 0$. then $\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta}) \xrightarrow{P} 0$ is possible, (sparcity principal).

5) Suppose $\underline{x}_1, \ldots, \underline{x}_n$ are iid and $p \times 1$,

Then $\sqrt{n}(\overline{\underline{x}} - \underline{\mu}) \xrightarrow{D} N_p(\underline{0}, \Sigma_{\underline{x}})$ and

if $\hat{\Sigma}_{\underline{x}}^{-1}$ exists and $\hat{\Sigma}_{\underline{x}}^{-1} \xrightarrow{P} \Sigma_{\underline{x}}^{-1}$ then

$$n(\bar{\underline{x}}-\underline{\mu})\hat{\Sigma}_{\underline{x}}^{-1}(\bar{\underline{x}}-\underline{\mu}) \xrightarrow{D} \chi^2_p :$$

one sample Hotelling's T^2 test.

If $n < p$ then $\hat{\Sigma}_x^{-1}$ does not exist since $\hat{\Sigma}_x$ is singular. Replace $\hat{\Sigma}_{\underline{x}}^{-1}$ by

i) $[\text{diag}(\hat{\Sigma}_{\underline{x}})]^{-1} = \begin{pmatrix} 1/\hat{\sigma}_1^2 & & \\ & \ddots & \\ & & \frac{1}{\hat{\sigma}_p^2} \end{pmatrix}$

ii) $\hat{\Sigma}_{\underline{x}}^*$ a generalized inverse that is usually singular

iii) $(\hat{\Sigma}_{\underline{x}} + \lambda I_p)^{-1}$ with $\lambda > 0$

iv) $I_p = I_p^{-1}$

6) If $\sqrt{n}(\bar{\underline{x}}-\underline{\mu}) \xrightarrow{D} \underline{z}$ and $\hat{D}^{-1} \xrightarrow{P} D^{-1}$, then

$$n(\bar{\underline{x}}-\underline{\mu})^T \hat{D}^{-1}(\bar{\underline{x}}-\underline{\mu}) \xrightarrow{D} \underline{z}^T D^{-1} \underline{z}$$ which

usually does not have a χ^2_p distribution. Even $\underline{z}^T \underline{z}$ can have a distribution that is difficult to work with.

7) In simulations with $n = 100$, as $p \uparrow$, using (v) worked better than i) for level (reject Ho with prob δ_n near level 0.05) where something called the nonparametric bootstrap was used to estimate the cutoff of the dist of $\underline{z}^T D^{-1} \underline{z}$.

8) In high dimensions, test statistics that estimate $\underline{\mu}^T \underline{\mu}$ often work well.

Ho $\underline{\mu} = \underline{0}$ vs H_A $\underline{\mu} \neq \underline{0}$ $\Leftrightarrow$ Ho $\underline{\mu}^T \underline{\mu} = 0$, H_A $\underline{\mu}^T \underline{\mu} \neq 0$

$\underline{\mu} = \underline{0}$ iff $\|\underline{\mu}\| = 0$ iff $\|\underline{\mu}\|^2 = 0$ iff $\underline{\mu}^T \underline{\mu} = 0$.

9) Suppose $\underline{x}_1, \ldots, \underline{x}_n$ are iid $p \times 1$ vectors.

To test Ho $\underline{\mu} = \underline{\mu}_0$, use $\underline{w}_i = \underline{x}_i - \underline{\mu}_0$, $i = 1, \ldots, n$.

Then $E(\underline{w}_i) = \underline{0}$ if Ho is true so $E[\underline{x}_i] = \underline{\mu}_0$.

10] In low dimensions $\underline{\bar{X}}^T\underline{\bar{X}}$ is a good estimator of $\underline{\mu}^T\underline{\mu}$.

a) If $\underline{x}_i \underset{\text{ind}}{\perp\!\!\!\perp} \underline{x}_j$, then $E[\underline{x}_i^T \underline{x}_j] =$

$$E\left[\sum_k x_{ik} x_{jk}\right] \overset{\text{ind}}{=} \sum_{k=1}^{p} E[x_{ik}]\, E[x_{jk}] \overset{\mu_k = E(x_{ik})}{=}$$

$$\sum_k \mu_k^2 = \underline{\mu}^T\underline{\mu}.$$

b) If $i=j$, $E[\underline{x}_i^T \underline{x}_i] = E\left[\sum_k x_{ik} x_{ik}\right]$

$$= \sum_k E(x_{ik}^2) \underset{\sigma_k^2 = V(x_{ik})}{=} \sum_k [\sigma_k^2 + \mu_k^2] = \underline{\mu}^T\underline{\mu} + \underline{\sigma}^T\underline{\sigma}$$

where $\underline{\sigma} = (\sigma_1, \dots, \sigma_p)$

c) $E[\underline{\bar{X}}^T\underline{\bar{X}}] = E\left[\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^T \frac{1}{n}\sum_{j=1}^{n} \underline{x}_j\right] =$

$$\frac{1}{n^2}\sum_i\sum_j E[\underline{x}_i^T \underline{x}_j] = \frac{1}{n^2}\left(\sum_{i\neq j}^{n}\sum^{n} E[\underline{x}_i^T \underline{x}_j] + \sum_{i=1}^{n} E[\underline{x}_i^T \underline{x}_i]\right)$$

$$= \frac{1}{n^2}\left[(n^2-n)\,\underline{\mu}^T\underline{\mu} + n\,\underline{\mu}^T\underline{\mu} + n\,\underline{\sigma}^T\underline{\sigma}\right] =$$

$$\underline{\mu}^T\underline{\mu} + \frac{\sum_{k=1}^{p} \sigma_k^2}{n}, \text{ and the 2nd term}$$

can be huge if $p >> n$,

11) Several HD tests use

$$T_n = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} \sum^{n} \underline{x}_i^T \underline{x}_j \quad \text{and}$$

showed the estimated Z score

$$\frac{T_n}{\sqrt{\hat{V}(T_n)}} \xrightarrow{D} N(0,1) \quad \text{as } n \to \infty$$

if H_0 $\underline{\mu}^T\underline{\mu} = 0$ is true, where $\hat{V}(T_n)$ estimates $V(T_n)$.

under stronger regularity conditions,

if $n \geq 3$ is fixed, then

$$\frac{T_n}{\sqrt{\hat{V}(T_n)}} \xrightarrow{D} t_k \quad \text{as } p \to \infty, \text{ where}$$

$$k = \frac{n(n-1)}{2} - 1 \quad \text{and } t_k \approx N(0,1) \text{ for}$$

$n \geq 9$ so $k \geq 35$.

12) Estimating $V(T_n)$.

a) Let $m = \left[\frac{n}{2}\right]$. Let

$$W_1, \ldots, W_m = \underbrace{\underline{x}_1^T \underline{x}_2}_{1}, \underbrace{\underline{x}_3^T \underline{x}_4}_{2}, \ldots, \underbrace{\underline{x}_{2m-1}^T \underline{x}_{2m}}_{m}.$$

Then the W_i are iid with $E(W_i) = \underline{\mu}^T \underline{\mu}$ and $V(W_i) = \sigma_W^2$. By the CLT,

$$\sqrt{m}\,(\overline{W}_m - \underline{\mu}^T \underline{\mu}) \xrightarrow{D} N(0, \sigma_W^2).$$

So the usual t-test and t CI applied to $W_1, \ldots, W_m$ works when the $\underline{x}_i$ are high dimensional.

b) T_n is a U statistic, and from the theory of U statistics

$$\text{I)}\quad V(T_n) = \frac{2\sigma_W^2}{n(n-1)} + \overbrace{\frac{4(n-2)}{n(n-1)}}^{\text{usually this term dominates}}\,\theta$$

with $\theta = \text{cov}(W_{ij}, W_{id})$ where $i \neq j$, $i < j$ and $r < d$ and $W_{ij} = \underline{x}_i^T \underline{x}_j$ for $i \neq j$.

Also $\theta \geq 0$.

II) If H_0 $\underline{\mu} = \underline{0}$ is true, then

$$\underline{\theta} = \underline{0} \quad \text{and} \quad V(T_n) = \frac{2\sigma_w^2}{n(n-1)} = V_0.$$

and $S_w^2 = \frac{1}{m-1}\sum_{i=1}^{m}(w_i - \bar{w})^2 \xrightarrow{P} \sigma_w^2$.

$$\hat{\sigma}_w^2 = \frac{1}{n(n-1)} \sum\sum_{i<j} (w_{ij} - T_n)^2 \xrightarrow{P} \sigma_w^2,$$

III) Since $\theta > 0$ if H_0 is false

$$t_1 = \frac{T_n}{\sqrt{\hat{V}_0}} > \frac{T_n}{\sqrt{\hat{V}(T_n)}} = t_2 \text{ when } H_0 \text{ is not true.}$$

so using $\frac{T_n}{\sqrt{\hat{V}_0}}$ gives more power

than using $\frac{T_n}{\sqrt{\hat{V}(T_n)}}$. Let $V = V(T_n)$

and $\hat{V} = \hat{V}(T_n)$, If H_0 is not true,

$\hat{V} \simeq \hat{V}_0 + \frac{4(n-2)\hat{\theta}}{n(n-1)}$. If H_0 is not true and

the second term dominates, $|t_1| \propto \sqrt{n}\, |t_2|$.

13) Suppose we use the interval $\left[T_n - 2\sqrt{\hat{V}_0},\ T_n + 2\sqrt{\hat{V}_0}\right]$

$$= T_n \pm 2\sqrt{\frac{2 S_w^2}{n(n-1)}}$$ and reject

$H_0\ \underline{\mu} = 0$ if 0 is not in the interval.

The HD literature often gives very strong regularity conditions where $\hat{V}_0 \to 0$ if $p = p_n = n^\gamma$ where γ is much larger than 0.5 and $\underline{\mu} = \underline{0}$.

Suppose $\underline{\mu} = \delta \underline{1}$ where $\underline{1}$ is the $p \times 1$ vector of 1's. Then $\underline{\mu}^T \underline{\mu} = \delta^2 p$ and the test using $\hat{V}_0$ may have good power

if $\frac{T_n}{\sqrt{\hat{V}_0}} > 2$ or $\underbrace{T_n > 2\sqrt{\hat{V}_0}}_{T_n - 2\sqrt{\hat{V}_0} > 0}$ or $\underline{\mu}^T\underline{\mu} > 2\sqrt{\hat{V}_0}$

or $\delta^2 p > \frac{2\sqrt{2}\,\sigma_w}{\sqrt{n(n-1)}} \approx \frac{2\sqrt{2}\,\sigma_w}{n}$ or

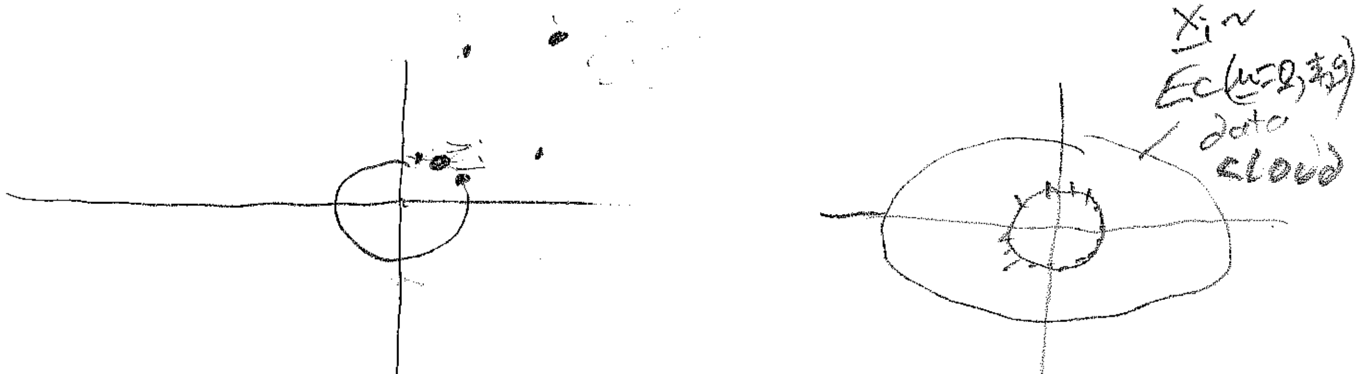
$\delta^2 > \frac{2\sqrt{2}\,\sigma_w}{np}$ ($T_n + 2\sqrt{\hat{V}_0} < 0$ is unlikely since $T_n \approx \delta^2 p$)

$\propto \frac{2\sqrt{2}}{n\sqrt{p}}$ if $\sigma_w^2 \propto p$.

14) Let $\underline{z}_i = SS(\underline{x}_i) = \underline{s}(\underline{x}_i) = \begin{cases} \underline{0} & \underline{x}_i = \underline{0} \\ \frac{\underline{x}_i}{\|\underline{x}_i\|} & \underline{x}_i \neq \underline{0} \end{cases}$

be the spatial sign function.

For a continuous dist, $P(\underline{x}_i = \underline{0}) = 0$ and the $\underline{z}_i$ are the projection of the $\underline{x}_i$ on the unit hypersphere centered at $\underline{0}$. This projection preserves the direction of $\underline{x}_i$ but not the magnitude of the $\underline{x}_i$.

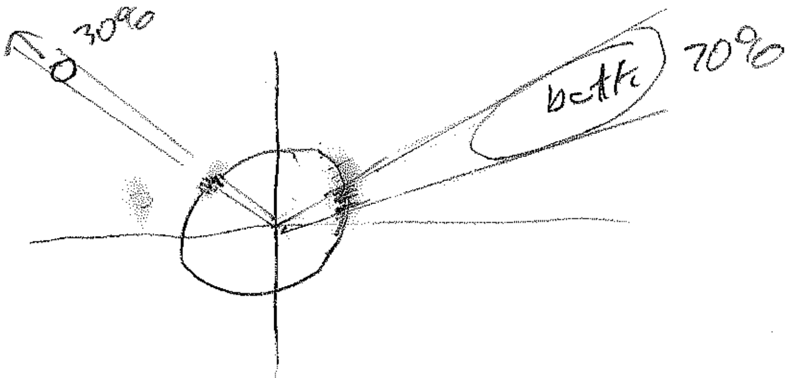


If the $\underline{x}_i$ are iid the $\underline{z}_i = \underline{s}(\underline{x}_i)$ are iid, and $E(\underline{z}_i) = \underline{\mu}_z$ exists even if $E(\underline{x}_i) = \underline{\mu} = \underline{\mu}_x$ does not exist, eg $\underline{x}_i$ iid multivariate cauchy.

If $\underline{\mu}_x = \underline{0}$, $\underline{\mu}_z = \underline{0}$ only if the dist has a lot of symmetry about the origin, such as an EC dist.

Can still test $H_0\ \underline{\mu}_z = \underline{0}$ vs $H_A\ \underline{\mu}_z \neq \underline{0}$ using the statistic $T_n(\underline{z})$ where the $T_n = T_n(\underline{x})$ test was given by 11) – 13).

15) The Z_i have some resistance to outliers.



16) HD outlier detection if $\underline{x} = (x_1, \ldots, x_p)^T$ and each predictor takes on many values.

$$\begin{array}{cccc} & x_1 & & x_p \\ X = & \left(\begin{array}{ccc} x_{11} & \cdots & x_{1p} \\ \vdots & & \\ x_{n1} & & x_{np} \end{array}\right) \end{array} = W$$

e.g. $x_1 = \text{gender} = \begin{cases} 0, & F \\ 1, & M \end{cases}$ is not allowed more than $n/2$ cases in the bulk of the data.

The squared Euclidean distance of x from T is $(\underline{x} - T)^T (\underline{x} - T) = D^2_{\underline{x}}(T, I_p)$.

a) Use Euclidean distances from the coordinatewise median

$$\underbrace{D_i \; (MED(W), I_p)}_{\text{start for the MB estimator}},$$