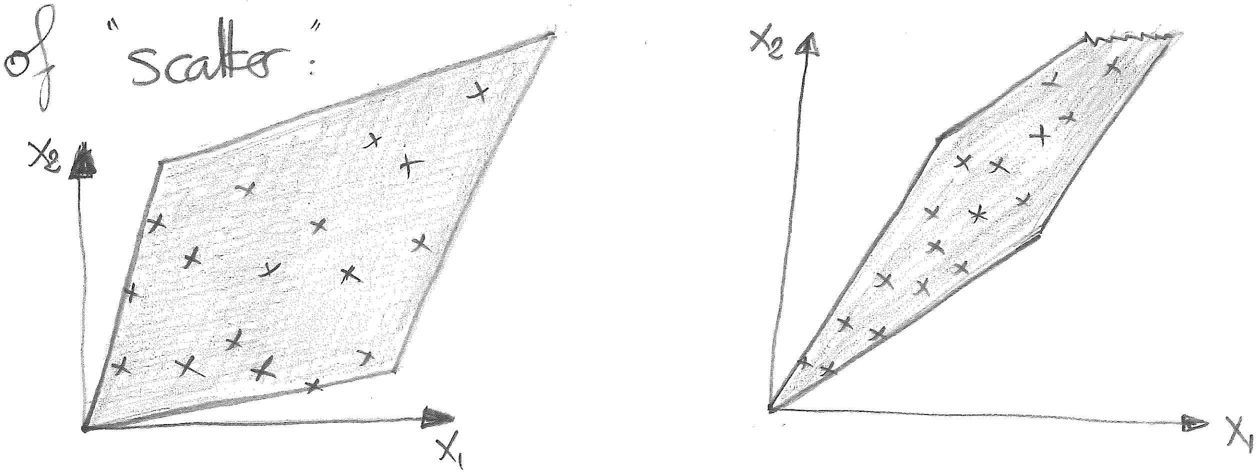
Week 7

This week we will look at two classic measures used to summarise multivariate data. The first is generalised variance which gives a decription of "scatter":



"Scatter" is determined in terms of the volume of a parallelotope which is a generalisation of a parallelepiped to higher dimensions. Eg. a 2-parallelotope is a parallelogon and a 3-parallelotope is a parallelepiped.

It is also another multivariate analogue to the variance σ^2 of a univariate distribution (where the covariance Σ is another)

We will also look at the coefficient of multiple correlation which is a measure of how well a given coordinate in our observation vector can be predicted using a linear function of the other coordinates. It is a score between 0 and 1 where a higher value indicates better predictability of one of the coordinates with respect to the others.

Generalised Variance

The <u>generalised variance</u> of the sample of vectors $\mathbb{X}_1, \mathbb{X}_2, \dots \mathbb{X}_n$ is

$$\det S = |S| := \left| \frac{1}{(n-1)} \sum_{k=1}^{n} (\mathbb{X}_k - \overline{\mathbb{X}})(\mathbb{X}_k - \overline{\mathbb{X}})' \right|.$$

Sometimes we will write $N = (n-1)$ to denote the <u>degrees of freedom</u>.

The GV is used in many likelihood ratio criteria for testing hypothesis. However it suffers from the weakness that by reducing spread to a single number we loose information about correlation. (see workshop).

Here we should distinguish:

* <u>sample GV</u> $|S|$. "random"

* <u>population GV</u>: The "true" GV derived from "true" covariance of population $|\Sigma|$.

We are going to look at estimating the population GV from the sample GV. under 2 different approaches:

- <u>Keep p fixed, take $n \to \infty$</u> to obtain an asymptotic estimate. Use estimate to construct hypothesis test. [CLASSIC].

- Take $p \to \infty$, $n \to \infty$, $y_n := \frac{p}{n} \to y < \infty$. Derive asymptotic estimate, use estimate to construct hypothesis test.

We want to know if our estimate for the (population) generalised variance is consistent and unbiased.

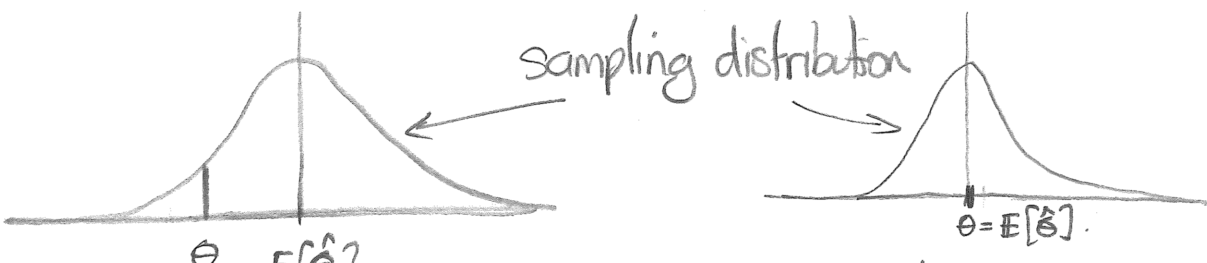
Let $\hat{\theta}_n$ be our estimate, dependent on n.

Def: Let $\hat{\theta}$ be an estimator for a quantity θ. Then $\hat{\theta}$ is an <u>unbiased estimator</u> if

$$\mathbb{E}[\hat{\theta}] = \theta .$$

otherwise $\hat{\theta}$ is said to be <u>biased</u>.

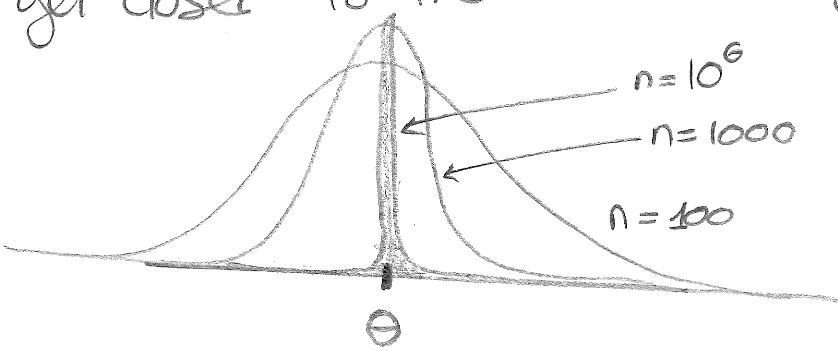
The bias of an estimator $\hat{\theta}$ is given by $B(\hat{\theta}) := \mathbb{E}[\hat{\theta}] - \theta$



positively biased

unbiased.

We would also like to characterise the idea that as we collect more data (ie. $n \to \infty$) our estimator should get closer to the true value of θ.



The estimator $\hat{\theta}_n$ is said to be a consistent estimator of θ if, for any positive number ε,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1.$$

or equivalently,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

The concepts of unbiased and consistency can be linked.

Theorem: An unbiased estimator $\hat{\theta}_n$ for θ is consistent if

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0.$$

Proof: (See any classic stats. book) ■

Wishart Distribution

The Wishart distribution is a multivariate generalisation of the χ^2 (chi-squared) distribution. In general, the distribution of sample covariance matrices and various sums of squares and products of matrices are Wishart distributed provided the underlying dist. is normal.

Let $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n$ be p-dimensional random vectors with p-dimensional multivariate normal distribution $N_p(\mu_k, \Sigma)$

Then the distribution of the pxp random matrix

$$\mathbb{W} = \sum_{k=1}^{n} \mathbb{X}_k \mathbb{X}_k'$$

is called a <u>Wishart distribution with noncentrality Ψ</u>, n degrees of freedom, and covariance matrix Σ.

The non-centrality matrix $\Psi = \mu_1\mu_1' + \mu_2\mu_2' + \cdots + \mu_n\mu_n'$.

In the special case $\Psi = 0$, we say that $\mathbb{W}$ has the <u>Wishart distribution</u> $\mathbb{W} \sim W_p(n, \Sigma)$. The case $W_p(n, I_p)$ is the <u>standard Wishart</u>.

Alternatively, we can view the Wishart distribution $W_p(n, \Sigma)$ as the distribution of $W = Z'Z$ where the rows of $Z : n \times p$ are independent identically distributed as $N_p(0, \Sigma)$.

When $n < p$, $\mathbb{W}$ is <u>singular</u> and $W_p(n, \Sigma)$ does not have a density. When $n \geq p$, a closed-form density exists.

Results on the Wishart distribution can be found in books on multivariate (statistical) analysis.

Distribution of Sample Generalised Variance.

Let $A = \sum_{k=1}^{N} Z_k Z_k'$ $\quad Z_k \sim N_p(0, \Sigma)$, $N = (n-1)$.

Then $|S| \sim \frac{1}{N^p} |A|$.

Let C be the unique square-root of Σ, ie. $CC' = \Sigma$, then define Y_k such that $Z_k =: C Y_k$. In other words,

$$Y_1, Y_2, \ldots, Y_N \sim N_p(0, I_p).$$

or $Y_k = C^{-1} Z_k$.

Now let $B = \sum_{k=1}^{N} Y_k Y_k' = \sum_{k=1}^{N} C^{-1} Z_k Z_k' (C^{-1})' = C^{-1} A (C^{-1})$

then $|A| = |C| \cdot |B| \cdot |C'| = |B| \cdot |\Sigma|$.

Since $\mathbb{B} = \sum_{k=1}^{N} Y_k Y_k'$ and $Y_k \sim N_p(0, I_p)$, it follows that $\mathbb{B} \sim W_p(N, I)$ ie. standard Wishart with N degrees of freedom.

Some properties of Wishart matrices can be found by performing the (Cholesky) decomposition $W = TT'$ where the pxp positive definite matrix W is decomposed into a nonsingular lower triangular matrix T.

<u>Note:</u> If diagonal elements of W are positive then T is uniquely determined.

We have: $\mathbb{W} = TT'$ $\quad T = \begin{pmatrix} T_{11} & & & 0 \\ T_{21} & T_{22} & & \\ \vdots & & \ddots & \\ T_{p1} & T_{p2} & \cdots & T_{pp} \end{pmatrix}$

<u>Theorem:</u> If $\Sigma = I_p$, the elements T_{ij} $(p \geq i \geq j \geq 1)$ are all independent, $T_{ii} \sim \chi^2(N+1-i)$ and $T_{ij} \sim N(0,1)$.
$(W \sim \mathbb{W}_p(N, \Sigma))$

<u>Proof:</u> (see classic stats books) ∎

The previous theorem implies that

$$|B| = \det B = \prod_{i=1}^{p} B_{ii} \sim \prod_{i=1}^{p} t_{ii}^2$$

where $t_{11}^2, t_{22}^2, \ldots, t_{pp}^2$ are χ^2 distributed.

<u>Theorem:</u> The distobution of $|S|$ of the sample $X_1, X_2, \ldots, X_n$ with distribution $N(\mu, \Sigma)$ is the same as the dist. of $|\Sigma|/N^p$ times the product of p independent factors, where the i'th factor is χ^2- dist. with $N+1-i = n-i$ degrees of freedom.

Eg. $p=1$ $\quad |S| \sim |\Sigma| \cdot \chi_N^2 / N$

$p=2$ $\quad |S| \sim \dfrac{|\Sigma|}{N^2} \cdot \chi_N^2 \cdot \chi_{N-1}^2$

In general $\quad |A| = |\Sigma| \cdot \chi_N^2 \cdot \chi_{N-1}^2 \cdots \chi_{N-p+1}^2.$

Asymptotic distribution of sample CV

We shall use a technique called the delta method to get the approximate probability distribution for a function of an asymptotically normal estimator.

In the univariate case, we have:

If we have a sequence $(X_n)_{n\geq 1}$ of random variables such that

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

where θ and σ^2 finite. Then if g is a function such that $g'(\theta)$ exists and is non-zero, we have

$$\sqrt{n}(g(X_n) - g(\theta)) \longrightarrow N(0, \sigma^2 [g'(\theta)]^2).$$

We require the multivariate analogue.

Let $(X_n)_{n\geq 1}$ be a sequence of p-dimensional random variables such that

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N_p(0, \Sigma)$$

where θ is p-dimensional constant vector and Σ is a (symmetric positive definite) covariance matrix.

Let $g: \mathbb{R}^p \to \mathbb{R}^r$ have derivative $\nabla g(\theta)$ then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N_p(0, [\nabla g(\theta)]' \Sigma [\nabla g(\theta)])$$

Here, $g(\theta) = (g_1(\theta), \cdots, g_r(\theta))$ ($r \times 1$ vector)

$$\nabla g(\theta) = \left[\frac{\partial g_i(\theta)}{\partial \theta_j}\right]_{ij} \quad (r \times p \text{ matrix})$$

Let's apply this result.

Recall that $|B| \sim \prod_{i=1}^{p} t_{ii}^2$ where $t_{ii}^2 \sim \chi^2_{n-i}$. So

$$\frac{|B|}{N^p} = C_1(n) \times C_2(n) \times \cdots \times C_p(n)$$

$$N C_i(n) \sim \chi^2_{n-i}$$

You can show (or lookup) that since χ^2_{n-i} is distributed as $\sum_{k=1}^{n-i} Z_k^2$, $Z_k \sim N(0,1)$ iid, the CLT states that the "standardised" $NC_i(n)$ is $N(0,1)$ distributed as $n \to \infty$.

$$\mathbb{E}[NC_i(n)] = n-i \quad (\text{since } \chi^2_{n-i}\text{-dist.})$$

$$Var[NC_i(n)] = 2(n-i)$$

$N = n-1$

$n-i = N+1-i$

$$\frac{NC_i(n) - (n-i)}{\sqrt{2(n-i)}} = \frac{N\left[C_i(n) - \frac{(N+1-i)}{N}\right]}{\sqrt{2(N+1-i)}}$$

$$= \sqrt{N}\,\frac{\left[C_i(n) - 1 + \frac{(i-1)}{N}\right]}{\sqrt{2}\sqrt{1-\frac{i-1}{N}}}$$

$$\xrightarrow{d} N(0,1).$$

As $n \to \infty$, $N \to \infty$, we have

$$\frac{i-1}{N} \to 0.$$

$$\sqrt{1 - \frac{i-1}{N}} \to 1.$$

So $\sqrt{N} \, \frac{C_i(n) - 1}{\sqrt{2}} \xrightarrow{d} N(0,1).$

and $\sqrt{N}\,(C_i(n) - 1) \xrightarrow{d} N(0,2).$

We want the asymptotic distribution of $|B|/N^p$ and we obtain this using the delta method:

Let $\mathbb{X}_n := (C_1(n), C_2(n), \dots, C_p(n))'$ $p \times 1$ random vector

$\theta := (1, 1, \dots, 1)'$ $p \times 1$ vector.

$g : \mathbb{R}^p \to \mathbb{R}$; $g(\bar{x}) := g(x_1, x_2, \dots, x_p) = x_1 x_2 \cdots x_p$

$\Sigma = 2I_p$ [as $\sqrt{N}\,(C_i(n) - 1) \xrightarrow{d} N(0,2)$ for each i and $C_i(n)$ are all independent.]

$$\nabla g(\theta) = \nabla g(x)\Big|_{x=\theta} \qquad p\times 1 \text{ Matrix.}$$

$$= \left[\frac{\partial g}{\partial x_1}(\theta), \frac{\partial g}{\partial x_2}(\theta), \dots, \frac{\partial g}{\partial x_p}(\theta)\right].$$

$$\frac{\partial g}{\partial x_i} = x_1 x_2 \cdots x_{i-1}\cdot x_{i+1}\cdots x_p. \qquad \frac{\partial g}{\partial x_i}(\theta) = 1.$$

$$\underbrace{[\nabla g(\theta)]'}_{p\times 1}\underset{p\times p}{\Sigma}\underbrace{[\nabla g(\theta)]}_{1\times p} = [1 \ 1 \cdots 1]'\, 2I_p\, [1 \cdots 1]$$

$$= 2p.$$

Hence, $\sqrt{N}\left(\frac{|B|}{N^p} - 1\right) \xrightarrow{d} N(0, 2p).$

<u>Theorem:</u> Let S be a $p\times p$ sample covariance matrix with n degrees of freedom. Then with <u>p fixed</u> and $n \to \infty$,

$$\sqrt{n}\left(\underbrace{|S|/|\Sigma|} - 1\right) \xrightarrow{d} N(0, 2p).$$

Recall we had $|A| = |B|\cdot|\Sigma|$.

$$A = \sum^{N} Z_k Z_k' \quad Z_k \sim N(0, \Sigma) \text{ iid}.$$

High-dimensional asymptotics for GV

We now look at the modern case where

$$n \to \infty, \quad p \to \infty, \quad y_n = \frac{p}{n} \to y \in (0,1).$$

Let's look at the simple case where the sample covariance matrix $\mathbb{S} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{X}_i \mathbb{X}_i'$ and $\mathbb{X}_i \sim N(0, I_p)$

So that $F^{\mathbb{S}} \xrightarrow{d} F_y$ where F_y is Standard Marchenko-Pastur dist.

Remember that the sample GV can be obtained by considering $\frac{1}{p}\log|\mathbb{S}| = \int_0^\infty \log(x)\, dF^{\mathbb{S}}(x)$

But $F^{\mathbb{S}} \xrightarrow{d} F_y$ so

$$\frac{1}{p}\log|\mathbb{S}| \to \int_a^b \log(x)\, dF_y(x) = -1 + \frac{y-1}{y}\log(1-y)$$

$$=: -d(y)$$

(calculated prev lecture)

where $d(u) := 1 + \frac{1-u}{u} \log(1-u)$ which is a positive function. Therefore, we have:

<u>Theorem</u>: Under regime where $p/n \to y \in (0,1)$ as $p, n \to \infty$ and $\mathbb{X}_k \sim N(0, I_p)$,

$$\frac{1}{p} \log |S_n| \xrightarrow{a.s.} -d(y).$$

When the population covariance matrix changes from I_p to Σ the sample GV is multiplied by $|\Sigma|$.

<u>Theorem</u>: Under regime $p/n \to y \in (0,1)$ as $p, n \to \infty$ and $\mathbb{X} \sim N(0, \Sigma)$ we have

$$\frac{1}{p} \log(|S|/|\Sigma|) \xrightarrow{a.s.} -d(y).$$

Notice that in this regime the sample GV is not a consistent estimator of the population GV.

We have the following CLT:

<u>Theorem:</u> Under regime $p/n \to y \in (0,1)$ as $p, n \to \infty$ we have

$$\log\left(\frac{|S|}{|\Sigma|}\right) + \underbrace{p\,d(y_n)}_{(*)} \xrightarrow{d} N(\mu, \sigma^2)$$

where $y_n = \frac{y}{N}$ and

$$\mu = \frac{1}{y}\log(1-y) \quad \sigma^2 = -2\log(1-y).$$

#

Notice that the centering term (*) depends on the sample size n. This occurs because the convergence $y_n \to y$ can occur very slowly.

Unfortunately this means we only know y_n in applications and we need to use y_n to approximate y.