Review on Gaussian vectors and the Central Limit Theorem (CLT)

$X = (X_1, X_2, \dots, X_n)$

Def: X is a Gaussian vector in $\mathbb{R}^n$ if there exists $m \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ symmetric positive semidefinite, such that the caracteristic function φ_X of X writes as:

transpose

$$\varphi_X(u) \underset{\text{definition}}{=} \mathbb{E}\left[e^{i\langle u, X\rangle}\right] = \exp\left(i\langle u, m\rangle - \frac{1}{2}u^T \Sigma u\right) \quad \text{for all } u \in \mathbb{R}^n$$

$i^2 = -1$ dot product

In this case, we write $X \sim N_n(m, \Sigma)$

Rk:
- This definition includes the case where X has a Dirac distribution at m $(\Sigma = 0)$.
- If $X \sim N_n(m, \Sigma)$, then $\mathbb{E}[X] = m$ and $Var(X) = \Sigma$
- Any sub-vector of a Gaussian vector is Gaussian.

⚠ The converse isn't true! For instance, take $X = (X_1, X_2)$ where $X_1 \sim N_1(0,1)$ and $X_2 = \varepsilon X_1$, where $\varepsilon \perp\!\!\!\perp X_1$ and $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$.

Then both X_1 and X_2 are Gaussian $(N_1(0,1))$, but (X_1, X_2) is not.

Prop: A random vector is Gaussian if and only if all the linear combinations of its components are Gaussians (in $\mathbb{R}$)

If X is a random vector, then $(X - E(X)) \in \operatorname{Im}(\operatorname{Var}(X))$ almost surely. Hence, X does not have a density if $\det(\operatorname{Var}(X)) = 0$. For a Gaussian vector, the condition "$\det(\operatorname{Var}(X)) \neq 0$" is actually sufficient for it to have a density.

Thm: If $X \sim N_n(m, \Sigma)$ with $\det(\Sigma) \neq 0$, then X has a density, which writes as

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)\right), \text{ for } x \in \mathbb{R}^n.$$

Rk: If $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$ and $X \sim N_n(m, \Sigma)$, then

$$AX + b \sim N_k(Am + b, A \Sigma A^T)$$

This remark is very important since it states that

1) Gaussian vectors are stable through affine transformations,

and

2) it allows to compute the mean and covariance matrix of this transformation.

Let us recall that if Σ is symmetric positive semidefinite, there exists a matrix denoted by $\Sigma^{\frac{1}{2}}$ such that $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma$. $\Sigma^{\frac{1}{2}}$ is called the "square root" of Σ. This matrix is invertible, with inverse $\Sigma^{-\frac{1}{2}}$, if Σ is invertible.

Prop:
- If $\det(\Sigma) \neq 0$ and $X \sim N_n(m, \Sigma)$, then $\Sigma^{-\frac{1}{2}}(X - m) \sim N_n(0, I_n)$
- If $X \sim N_n(0, I_n)$, then $m + \Sigma^{\frac{1}{2}} X \sim N_n(m, \Sigma)$

This proposition allows to "normalise" Gaussian vectors. We'll use it extensively in the course to derive hypothesis testing procedures as well as confidence intervals.

Rk: In general, if two random variables are independent, then their covariance matrix is diagonal.
For Gaussian vectors, the converse is true: If $X = (X_1, \dots, X_n) \sim N(m, \Sigma)$, with Σ being a <u>diagonal</u> matrix, the $X_1, \dots, X_n$ are independent!

Projection of Gaussian vectors

The following theorem is fundamental in all the theory of Gaussian models. It shows up in most of parameter estimation problems involving either Gaussians, or large sample sizes (see CLT later on).

Thm (Cochran) Let $X \sim N_n(m, \sigma^2 I_n)$ with $\sigma > 0$, and $L_1 \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} L_p$ a decomposition of $\mathbb{R}^n$ in orthogonal linear subspaces of dimensions $r_1, \dots, r_p$:

Then
1) The orthogonal projections $\tilde{\pi}_1, \dots, \tilde{\pi}_p$ of X onto $L_1, \dots, L_p$ are independent Gaussians
2) for all $i = 1, \dots, p$, $\quad \dfrac{1}{\sigma^2} \|\tilde{\pi}_i\|^2 \sim \chi^2_{r_i}$

Euclidean norm

Chi-squared distribution with $df = r_i$

An example of application of Cochran: Student's test (revisited)

For a sequence of independent Gaussians $X_1, \dots, X_n \sim N_1(m, \sigma^2)$, we recall that

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad , \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Empirical mean

Empirical variance

We also recall that the <u>Student distribution</u> with df degrees of freedom, denoted by T_{df}, is the distribution of $\dfrac{\sqrt{df}\, X}{\sqrt{Y}}$, where

- $X \sim N_1(0,1)$
- $Y \sim \chi^2_{df}$
- $X \perp\!\!\!\perp Y$ "Independent from"

<u>Theorem</u> (Fisher) If $X_1, \dots, X_n \sim N(m, \sigma^2)$ are independent with $\sigma > 0$, then:

- $\overline{X}_n \perp\!\!\!\perp S_n^2$
- $\dfrac{\sqrt{n}\,(\overline{X}_n - m)}{\sigma} \sim N_1(0,1)$, $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$
- $\sqrt{n}\,\dfrac{(\overline{X}_n - m)}{S_n} \sim T_{n-1}$

<u>Proof</u>: without loss of generality, take $m=0$ and $\sigma = 1$ (do affine transformations $Y_i = \dfrac{X_i - m}{\sigma}$ otherwise)

Let us first notice that the vector $X = (X_1, \dots, X_n)$ has distribution $N_n(me, \sigma^2 I_n)$ where $e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \updownarrow n$. Let L be the vector subspace of $\mathbb{R}^n$ spanned by e, and P be the orthogonal projector on L.

We have $PX = \overline{X}_n e$ and $(I_n - P)X = \begin{pmatrix} X_1 - \overline{X}_n \\ \vdots \\ X_n - \overline{X}_n \end{pmatrix}$. Since $(I_n - P)X$ is the orthogonal projection on the orthogonal space $L^\perp$ of L, we deduce from Cochran that $PX \perp\!\!\!\perp (I_n - P)X$, and in particular that $\overline{X}_n \perp\!\!\!\perp \frac{\|(I_n - P)X\|^2}{n-1} = S_n^2$. Finally, again from Cochran (!),

$$(n-1)S_n^2 = \|(I_n - P)X\|^2 \sim \chi^2_{n-1}.$$

Rk: From this result, one derives the Student test and confidence intervals you saw in MATH 181A. :)

Let's now move to non-Gaussian settings, but where a Gaussian approximation can be used because of a large-enough sample size. Namely, let us state the vectorial Central limit theorem.

The Central limit Theorem in $\mathbb{R}^d$

Let $X_1, \dots, X_n$ be random <u>vectors</u> of $\mathbb{R}^d$, $X_i = \begin{pmatrix} X_i^{(1)} \\ \vdots \\ X^{(d)} \end{pmatrix}$, that independent and identically distributed.

From the <u>law of large numbers</u>, you know that

almost sure convergence

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{a.s.} \mathbb{E}(X_1).$$

The CLT tells you how fast this convergence goes.

<u>Thm</u> (Central Limit Theorem)

Let $X_1, \dots, X_n$ be random vectors of $\mathbb{R}^d$, that are independent and identically distributed. Then, as $n \to \infty$, we have

Convergence in distribution

$$\sqrt{n}\left(\overline{X}_n - \mathbb{E}(X_1)\right) \xrightarrow[n\to\infty]{\mathcal{D}} N_d\left(0, Var(X_1)\right).$$

<u>Application</u>: χ^2-tests!

$\hookrightarrow$ let's see what it is.

$$Var(X_1) = \begin{pmatrix} Var(X_1^{(1)}) & Cov(X_1^{(1)}, X_1^{(2)}) & \cdots & Cov(X_1^{(1)}, X_1^{(d)}) \\ Cov(X_1^{(2)}, X_1^{(1)}) & Var(X_2^{(2)}) & \ddots & \vdots \\ \vdots & \ddots & \ddots & Cov(X_1^{(d-1)}, X_1^{(d)}) \\ Cov(X_1^{(d)}, X_1^{(1)}) & \cdots & Cov(X_1^{(d)}, X_1^{(d-1)}) & Var(X_d^{(d)}) \end{pmatrix}$$

$d \times d$ matrix