STATISTIQUE DE ANDERSON (I7, K7, K9)

La **statistique de ANDERSON** permet de classer une **observation** supplémentaire dans l'une parmi deux **populations**. Ces populations sont supposées gaussiennes et homogènes (même dispersion) (cf **problème à plusieurs échantillons**, **homogénéité**).

(i) Soit $X^i = (X_{i1}, ..., X_{iN(i)})$ (i = 1 , 2) deux **échantillons aléatoires** (on note de façon équivalente N(i) et N_i). On suppose que ξ_1 et ξ_2 sont des va indépendantes, chacune distribuée selon la **loi normale** $\mathscr{L}(\xi_i) = \mathscr{N}_K(\mu_i, \Sigma)$ (i = 1, 2), et que la **matrice de dispersion** Σ leur est commune (ie ne dépend pas de i). On note X^i un **échantillon iid** issu de ξ_i (i = 1, 2) (donc X^1 et X^2 sont indépendants entre eux). On pose (i = 1, 2) :

$$(1) \quad \begin{aligned} \bar{X}_i &= (N_i)^{-1} \Sigma_{n(i)=1}^{N(i)} X_{in(i)}, \\ S_i^2 &= (N_i - 1)^{-1} . \Sigma_{n(i)=1}^{N(i)} (X_{in(i)} - \bar{X}_i)(X_{in(i)} - \bar{X}_i)', \end{aligned}$$

la **moyenne empirique** vectorielle et la **matrice de dispersion** empirique (corrigée) de l'échantillon n° i, considérés comme estimateurs resp de μ_i et de Σ.

La **variance empirique** d'ensemble « corrigée » est définie comme moyenne (arithmétique) pondérée :

$$(2) \quad S^2 = (N_1 + N_2 - 2)^{-1} . \{(N_1 - 1) S_1^2 + (N_2 - 1) S_2^2\}.$$

S^2 est alors un **estimateur sans biais** de la dispersion commune Σ.

(ii) Si X* représente une observation supplémentaire, on appelle **statistique de T.W. ANDERSON** la **statistique** suivante :

$$(3) \quad W_{N(1)N(2)} = X^{*}{}' S^{-2} (\bar{X}_1 - \bar{X}_2) - (1/2) (\bar{X}_1 + \bar{X}_2)' S^{-2} (\bar{X}_1 - \bar{X}_2),$$

qui s'écrit aussi :

$$(4) \quad W_{N(1)N(2)} = \{X^* - (1/2) (\bar{X}_1 + \bar{X}_2)\}' S^{-2} (\bar{X}_1 - \bar{X}_2).$$

La **règle de décision**, appelée **règle d'affectation** ou **règle de classement** (ou parfois **règle de classification** ou **règle de discrimination**), de X* est alors la suivante :

$$(5) \quad \begin{aligned} W_{N(1)N(2)} > q_{1-\alpha} \quad &\Rightarrow \quad X^* \sim P^{\xi(1)}, \\ W_{N(1)N(2)} < q_{1-\alpha} \quad &\Rightarrow \quad X^* \sim P^{\xi(2)}, \end{aligned}$$

où $q_{1-\alpha}$ est le **quantile** d'ordre $\alpha \in]0, 1[$ de la **loi** $\mathscr{L}(W_{N(1)N(2)})$ de $W_{N(1)N(2)}$.

Cette loi est tabulée et possède des **approximations** asymptotiques (cf **loi asymptotique**).