FONCTION DISCRIMINANTE (I7, K7, K9)

(27 / 10 / 2020, © Monfort, Dicostat2005, 2005-2019)

Le concept de **fonction discriminante** est une notion de base de la **discrimination**, utilisée dans l'une ou l'autre des **situations statistiques** suivantes :

(a) imputation d'un **échantillon** d'**unités statistiques** à une **population** (inconnue) parmi plusieurs populations possibles (dont l'ensemble est donné). Ces populations sont en concurrence entre elles car, a priori, toutes peuvent avoir généré l'échantillon considéré. Autrement dit, l'échantillon peut avoir été extrait de l'une de ces populations, et il convient de déterminer laquelle ;

(b) imputation d'un échantillon d'**observations** à une **loi de probabilité** (inconnue) parmi d'autres lois possibles (dont l'ensemble est donné). Ces lois sont concurrentes car toutes sont susceptibles avoir engendré l'échantillon. Autrement dit, la loi de l'échantillon est l'une des lois considérées, et l'on cherche à la déterminer.

(i) Soit $\xi = (\xi_1, \dots, \xi_K)$ est une liste de **variables numériques** observées sur des **unités statistiques** $n = 1, \dots, N$. Les observations sont disposées selon une (N, K) -**matrice** X (cf **disposition**).

On dit que $\xi \mapsto \lambda = b' \xi$ est une **fonction discriminante** pour les unités n ssi λ suit un **mélange de lois** différentes.

Une unité n appartient donc à une population et une seule. Autrement dit, $\lambda = b' \xi = \sum_{k=1}^K b_k \xi_k \sim P^\lambda$, avec :

$$(1) \quad P^\xi = \sum_{i \in I} a_i P^{\xi(i)},$$

où les λ_i sont des va implicitement associées au **populations** $i \in I$, et où l'on note, par commodité, $\xi(i)$ pour désigner la va ξ_i .

On doit donc chercher ξ (ie b) tq, au vu de la **matrice d'observation** X de ξ , le mélange P^λ comporte $\text{Card } I < K$ composantes.

(ii) Dans un contexte flou (cf **théorie des parties floues**), l'appartenance d'une unité ou d'une observation associée n'est pas de type dichotomique (oui, non) mais elle est appréciée par une fonction d'appartenance. Cette dernière décrit le degré d'appartenance de l'unité ou de l'observation considérée à chacune des populations ou des lois considérées (degré de vraisemblance).

Ce type de situation statistique se rencontre eg :

(a) biologie : un nouvel agent pathogène peut appartenir à une classe d'agents pathogènes connue (bactéries, champignon, virus, etc), ou peut être susceptible de fonder une nouvelle classe d'agents pathogènes (inconnue jusqu'ici) ;

(b) psychologie : un « dysfonctionnement » comportemental peut parfois être clairement identifiable, mais peut aussi parfois être difficile à classer ;

(c) sociologie (archéologie paléontologique) : des éléments de squelette nouvellement découverts peuvent se rattacher à des catégories charpentières connues, ou non.