PROBLÈME LINÉAIRE (G2, H1, H3, I3, I9)
(08 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **problème linéaire** est un **problème statistique** basé sur le **modèle de régression multiple** linéaire.

(i) Dans l'approche non paramétrique (ou « semi-paramétrique »), on considère :

(a) un **modèle statistique** de base $(\Omega, \mathcal{T}, \mathcal{P})$;

(b) un **espace d'observation** $(\mathcal{Y}, \mathcal{C})$ tq $\mathcal{Y} = \mathbf{R}^N$ et $\mathcal{C} = \mathcal{B}(\mathbf{R}^N)$ (**tribu borélienne** de $\mathbf{R}^N$) ;

(c) une **va** $y : \Omega \mapsto \mathcal{Y}$ de **loi** $P^y = y\ (P)$, où P parcourt l'ensemble $\mathcal{P}$ des **mesures de probabilité** définies sur $\mathcal{T}$. On note $\mathcal{P}^y = y\ (\mathcal{P})$ la famille image de $\mathcal{P}$ par y ;

(d) un sous-**espace vectoriel** $\mathcal{L}$ de $\mathcal{Y}$ (de dimension Dim $\mathcal{L}$ = K, avec K < N = Dim $\mathcal{Y}$). On note eg $(\xi_1, ..., \xi_K)$ une partie génératrice de $\mathcal{L}$ (eg une **base** de $\mathcal{L}$) ;

(e) une **matrice** $A \in M_{QN}(\mathbf{R})$ (ou l'**opérateur linéaire** associé).

On suppose que, $\forall P \in \mathcal{P}$:

$y \in \mathcal{L}_{\mathbf{R}N}^2 (\Omega, \mathcal{T}, P)$, (où **R**N désigne l'espace $\mathbf{R}^N$),

(1) $E_P\, y = m_P \in \mathcal{L}$,

$\exists\ \sigma_P > 0$ tq $V_P\, y = \sigma_P^2 . I_N$,

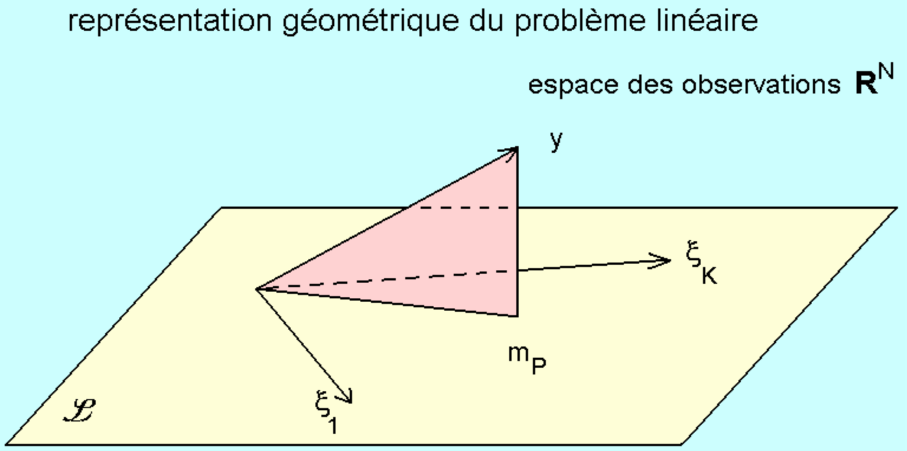
où E_P (resp V_P) désigne l'**espérance** (resp la **variance**) de y calculée avec $P \in \mathcal{P}$.

(ii) Dans son principe, le **problème linéaire** consiste à estimer le nouveau **paramètre** (vectoriel) $g_P = A\ m_P \in \mathbf{R}^Q$ à l'aide d'un **estimateur** t, linéaire pr à y, et vérifiant le **principe de réduction** suivant :

(2) $t\ (\mathcal{L}) = A\ (\mathcal{L})$ (ie $t\ (y) = A\ y$, $\forall\ y \in \mathcal{L}$).

En particulier, si l'on observe $y = m_P$ (espérance de la loi P^y de y elle-même), on estime $A\ m_P$ exactement selon $t\ (m_P) = A\ m_P$.

(iii) Le résultat fondamental relatif au problème linéaire est le **théorème de GAUSS-MARKOV**, où l'on adopte pour critère de choix (entre estimateurs t) une fonction de **perte quadratique** (cf **illustration géométrique**).



(iv) Le problème linéaire s'étend au cas où la **matrice diagonale** $V\ y = \sigma_P{}^2\ I_N$ est remplacée par une **matrice de dispersion** quelconque $V_P\ y = \Sigma_P$. La **projection** à considérer est alors une **projection « oblique »** (ie effectuée selon la **métrique** définie par $\Sigma_P{}^{-1}$). Le résultat de base correspondant à cette situation plus générale est le **théorème de AITKEN-GAUSS-MARKOV**.