CORRECTION D'HÉTÉROGÉNÉITE (C2, C5)

Lorsque la variance d'une **va** dépend de son niveau moyen, il est possible de corriger cette source d'**hétérogénéité** (cf aussi **hétéroscédasticité**).

Soit $\xi \in \mathscr{L}_{\mathbf{R}}^2 (\Omega, \mathscr{T}, P)$ une **vars** de carré intégrable, de **moyenne** $\mu = E\,\xi$ et de **variance** σ^2. S'il existe une relation connue entre variance et moyenne de la forme (hétérogénéité liée au niveau) :

(1) $\sigma = \varphi(\mu)$,

on peut trouver une fonction de correction $f : x \mapsto y = f(x)$ tq la variance $(\sigma^2)^*$ de la va transformée $\eta = f(\xi)$ ne dépende pas (approximativement) du niveau $E\,\eta = \mu^*$.

La **fonction de correction** s'écrit, au premier ordre (cf **formule de TAYLOR stochastique** d'ordre 1) :

(2) $y = f(\mu) + f'(\mu).(x - \mu) + o_P(1) . |x - \mu|$.

En négligeant le reste, on obtient :

(3) $$\mu^* = E\,\eta = f(\mu),$$
$$(\sigma^2)^* = V\,\eta = \sigma^2 \{f'(\mu)\}^2 = \{\varphi(\mu)\}^2 . \{f'(\mu)\}^2,$$

et $(\sigma^2)^*$ ne dépend pas de μ^* s'il existe un nombre $\alpha \in \mathbf{R}$ tq : $\varphi(\mu)\, f'(\mu) = \alpha$. On en déduit que :

(4) $f(x) = \alpha . \int (\varphi(x))^{-1}\, dx$.

En pratique, φ n'est pas connue (cas le plus courant) : on peut l'estimer (eg par **régression** ou **interpolation** polynômiale) à l'aide des **moments empiriques** :

(5) $$m_h = N_h^{-1} \Sigma_{n=1}^{N(h)} X_{nh} ,$$
$$s_h^2 = (N_h - 1)^{-1} \Sigma_{n=1}^{N(h)} (X_{nh} - m_h)^2,$$

calculés sur un **échantillon** $(X_{nh})_{n=1,\ldots,N\,;\,h=1,\ldots,H}$ de taille N égale à $\Sigma_{h=1}^{H} N_h$, issu de la va ξ et réparti en H classes ou **catégories**.