PERTE D'INFORMATION (G)
(24 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

On dit qu'il existe une **perte d'information** lorsque l'**observation** d'un **phénomène** ne peut être réalisée, soit en totalité, soit en partie. On parle d'**observation incomplète** :

(a) **non réponse** à un **sondage** ;

(b) destruction d'**unité expérimentale** ;

(c) **observations** ou **données** inaccessibles (cf **observation manquante**), etc.

Cette notion, qui se rencontre souvent dans diverses **situations statistiques** (eg en **théorie des plans d'expérience**), peut se présenter dans le contexte de la **théorie de la décision**.

(i) Soit ($\mathcal{X}$, $\mathcal{B}$, $(P^X_\theta)_{\theta \in \Theta}$), avec $\Theta \in \mathcal{O}$ ($\mathbf{R}^Q$), un **modèle paramétrique**, X une **statistique** (eg un **échantillon**), S = s (X) et T = t (X) deux statistiques qui en dépendent resp à travers les **applications mesurables** s et t.

L'**information de FISHER** associée à X s'écrit :

(0) $\quad I(\theta) = E_\theta \{D_\theta \operatorname{Log} L(X, \theta) . (D_\theta \operatorname{Log} L(X, \theta)'\}$,

où L désigne la **fonction de vraisemblance** du modèle et D_θ la dérivée partielle pr à θ.

De même, l'information de FISHER associée à S (resp T) est notée $I_S(\theta)$ (resp $I_T(\theta)$) et se calcule eg à l'aide du **modèle image** du modèle donné par S (resp T).

(ii) On montre que :

(a) si S et T sont indépendantes (cf **indépendance stochastique**), on a :

(1) $\quad I_{(S,T)}(\theta) = I_S(\theta) + I_T(\theta)$.

Par suite, si T n'est pas une **statistique exhaustive** (pour θ), le fait d'utiliser T au lieu de X conduit à une **perte d'information** qui peut être mesurée eg selon :

(2) $\quad D_{T/X}(\theta) = I(\theta) - I_T(\theta)$;

(b) si S et T ne sont pas indépendantes, on a, en conditionnant eg pr à T (cf **conditionnement**) (notations évidentes) :

(3) $\quad I_{(S,T)}(\theta) = I_{S/T}(\theta) + I_T(\theta)$.

Par suite, si le couple (S, T) est une statistique exhaustive (pour θ) et si T est de **loi fixe** (pr à θ), on obtient :

(4) $I_{S/T}(\theta) = I_{(S,T)}(\theta) - I_T(\theta) = I(\theta) - 0 = I(\theta)$.

(iii) On peut donc procéder au **recouvrement de l'information**, ou **récupération de l'information**. En effet, si l'on perd de l'information en utilisant une statistique non exhaustive S, on peut la retrouver en conditionnant pr à une statistique exhaustive T.

(iv) Ces notions peuvent s'étendre asymptotiquement (en faisant tendre le nombre N d'observations vers l'infini) (cf **modèle asymptotique**, **propriété asymptotique**).