DEGRÉ DE LIBERTÉ (F, G)
(22 / 10 / 2019, © Monfort, Dicostat2005, 2005-2019)

(i) De façon générale, un **système** d'équations quelconque dans lequel il existe davantage de variables que d'équations est dit posséder des **degrés de liberté**.

S'il s'écrit sous la forme :

(1) $f(x) = b$,

avec $f : E \mapsto F$, où $E = \prod_{i=1}^{n} E_i$, $F = \prod_{j=1}^{m} F_j$, et $n > m$, on dit que ce système possède $d = n - m > 0$ **degrés de liberté**.

En effet, la solution de (1) exprime, en général, (cf **théorème de la fonction implicite**) une dépendance entre n variables $x_i \in E_i$ et les n - m autres.

(ii) En **Statistique**, le **nombre de degrés de liberté** est un nombre entier qui représente la différence entre le nombre d'**observations** disponibles (taille d'un **échantillon**, nombre d'**unités expérimentales**) et le nombre de **paramètres** d'une **représentation statistique**. Il relie donc le « nombre d'observations » au nombre de paramètres caractérisant la **loi de probabilité** générant cet échantillon (loi du **phénomène** étudié).

(iii) Soit $(\mathscr{X}, \mathscr{B}, (P_\theta^X)_{\theta \in \Theta})$ un **modèle statistique** paramétré, avec $Q \in \mathscr{O}(\mathbf{R}^Q)$ (famille des ouverts de $\mathbf{R}^Q$), et $X = (X_1 ,..., X_N)$ un **échantillon** de taille N à valeurs dans l'**espace d'observation** produit $\mathscr{X} = \mathbf{R}^N$.

On dit que le **nombre de degrés de liberté du modèle** est :

(2) $d = N - Q$.

(iv) Une justification (heuristique) de la notion provient de la considération suivante. Si le modèle précédent est associé à un **problème d'estimation** dans lequel θ est un **paramètre d'intérêt**, un **estimateur ponctuel** $\theta^\sim$ de θ est défini à l'aide d'une **statistique** $s : \mathscr{X} \mapsto \mathbf{R}^Q$ tq :

(3) $\theta^\sim = s(X) = s(X_1 ,..., X_Q , X_{Q+1} ,..., X_N)$.

Estimer θ par $\theta^\sim$ revient à « utiliser » (ou à « consommer ») un vecteur d'observations $(X_1 ,..., X_Q)$, à valeurs dans un ouvert de $\mathbf{R}^Q$. Le **théorème de la fonction implicite** montre, sous des hypothèses générales, qu'il existe une fonction φ tq, à une **permutation** sur les X_n près :

(3) $\theta^\sim = s(\varphi(X_{Q+1} ,..., X_N) , X_{Q+1} ,..., X_N))$.

Il existe donc $d = N - (Q+1) + 1 = N - Q$ coordonnées de X (ie $X_{Q+1} ,..., X_N$) qui sont arbitraires (ou libres), ie indépendantes (au sens de l'analyse).

(v) Le nombre de degrés de liberté est souvent un paramètre de certaines lp : eg **loi du chi-deux**, **loi de FISHER-SNEDECOR**, etc. Ces lois de probabilité sont notamment suivies par diverses statistiques, tq la précédente.

On a supposé ici que d $\in$ **N***. Le nombre de degrés de liberté n'a été défini que pour des observations en nombre fini et des paramètres en nombre fini Q.

(vi) S'il existe des **contraintes sur les observations** (eg une taille d'échantillon donnée N), le **principe de parcimonie** a pour but d'augmenter d en limitant au maximum le nombre Q de paramètres à prendre en considération dans l'**inférence statistique**.

Des critères de **sélection de modèles** (eg **critère de AKAIKE**, **critère de SCHWARTZ**) utilisent la **réduction de dimensionnalité** Q des modèles pour choisir entre ceux-ci.