INOBSERVABLE, OBSERVABLE (C1, G2)

(20 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Est **observable** (resp **inobservable**) un **paramètre**, une **variable aléatoire**, ou une **statistique** que le **statisticien** peut (resp ne peut) directement observer ou évaluer dans le cadre d'un **phénomène** aléatoire donné : **expérimentation**, **sondage**, **simulation**, etc.

(i) **Paramètre inobservable**. Lorsque $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ est un **modèle statistique** (image), on suppose généralement :

(a) d'une part, que X est une **variable aléatoire observable**, ou une **statistique observable** : eg un **échantillon** constitué de **mesures** effectuées sur des **unités statistiques** ;

(b) la **vraie valeur** θ^* du paramètre (d'intérêt) $\theta \in \Theta$ est **inobservable**. C'est cette dernière qui fait l'objet de l'**inférence statistique** (estimation, tests, etc).

Les paramètres (d'intérêt) qui interviennent dans une **spécification** peuvent donc être totalement, ou seulement partiellement, observables. Ainsi :

(a) dans une **simulation**, les valeurs des paramètres sont non seulement observables mais aussi connues du statisticien, qui en fixe lui-même les valeurs ;

(b) dans la théorie de l'**école bayésienne**, les paramètres sont connus seulement à travers leur **loi a priori** : cette dernière est, par hypothèse, supposée donnée, ou préalablement estimée. Ainsi, une contrainte de la forme $\theta \in B$ (où $B \subset \Theta$) sur le paramètre θ peut s'interpréter comme une condition portant sur la loi a priori Π (qui ne « charge » alors que B) (cf **support d'une probabilité**). Ceci est notamment le cas eg lorsque $B = \{\theta \in \Theta : g(\theta) = 0\}$ ou lorsque $B = \{\theta \in \Theta : g(\theta) \geq 0\}$, où g désigne une fonction précisant la contrainte (cf **identité**).

(ii) **Variable inobservable**. Il peut arriver que certaines **variables aléatoires**, ou certaines statistiques, considérées dans un modèle, soient **totalement inobservables**, ou **partiellement inobservables** (cf a contrario **variable observable**). Cette situation se rencontre notamment :

(a) en cas d'« **observation incomplète** ». Ainsi :

(a)₁ un **échantillonnage** conduit à extraire seulement un **échantillon** A constitué d'**unités** tirées dans une **population** Ω (au lieu de Ω elle-même) ;

(a)₂ il en est de même pour les variables observées sur les unités concernées ;

(a)₃ cette situation d'incomplétude est encore réalisée en cas de **censure** ou de **lacune** ;

(b) lorsque la « théorie » fait intervenir des variables inobservables (cf **variable cachée**) dans la spécification même du modèle, eg :

(b)₁ **facteur** inobservable en **analyse des données** ou même dans un **plan d'expérience** ;

(b)₂ **modèle de structures latentes** comportant des **variables latentes** (cf eg **structure latente, modèle des classes latentes**) ;

(c) en cas d'**erreurs** sur les équations d'un modèle. Ces erreurs peuvent être conceptuellement nécessaires (eg **perturbation aléatoire** intégrée dans une **relation fonctionnelle**, un **modèle de régression** ou un **modèle d'interdépendance**) (cf aussi **modèle à erreurs composées**). Elles peuvent être aussi de simples erreurs d'observation affectant les **variables endogènes** (cf erreurs sur les variables infra) ;

(d) en cas d'erreurs sur les variables d'un modèle, affectant notamment ses **variables exogènes** (cf **modèle à erreurs sur les variables**, etc).