

## ESPACE D'OBSERVATION (A5, B1)

(03 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **espace d'observation** est un **espace mesurable** spécifique, dont le rôle est fondamental en **Statistique** (cf **observabilité**, **observation**, **espace d'état**).

(i) Soit  $(\Omega, \mathcal{F})$  un **espace probabilisable fondamental**, dans lequel :

(a)  $\Omega$  est un **ensemble fondamental** (univers, **population**) constitué d'**unités statistiques** (éléments, individus). Cet ensemble peut notamment s'associer à une **expérience aléatoire** ou à un **sondage** ;

(b)  $\mathcal{F}$  est une **tribu de parties**, ou **tribu d'événements**. Ces parties ont l'intérêt de représenter pour le **statisticien** des réalisations ou des résultats complexes, etc, en général « **aléatoires** ».

L'analyse statistique porte généralement sur des « **observations** » (**mesures**, etc)  $x = \xi(\omega)$  effectuées sur les éléments  $\omega \in \Omega$ , où  $\xi : \Omega \mapsto \mathcal{X}$  est une **variable d'intérêt** (eg un **attribut**) qui joue le rôle de **variable statistique**. Autrement dit, on définit un **ensemble d'observation**, ou **ensemble des observations**,  $\mathcal{X}$  comme l'ensemble des valeurs prises par  $\xi$  (ie observées sur les individus  $\omega$ ).

On peut munir  $\mathcal{X}$  d'une tribu  $\mathcal{B}$  (eg la **tribu image** de  $\mathcal{F}$  par  $\xi$ ), ce qui définit un **espace probabilisable** appelé, selon la **situation statistique**, **espace d'observation**, ou **espace des observations**, ou **espace d'échantillonnage**, ou encore parfois **espace final**, et noté  $(\mathcal{X}, \mathcal{B})$ .

L'application  $\xi : \Omega \mapsto \mathcal{X}$  est généralement (implicitement) supposée être une **application mesurable**, ie  $(\mathcal{F}, \mathcal{B})$ -mesurable). Elle joue alors le rôle de **variable aléatoire**. Cette variable peut être (cf **loi multivariée**) :

(a) « simple » ou « scalaire » : dans ce cas, elle peut être une **variable qualitative** ou une **variable quantitative** ;

(b) « multiple » : soit sous forme d'une liste de variables, soit sous forme d'une variable vectorielle (**vecteur aléatoire**) ;

(c) « mixte » : au sens où elle « combine » des variables qualitatives et des variables quantitatives, etc.

On peut ainsi considérer :

(a) soit que  $\xi$  est  $(\mathcal{F}, \mathcal{B})$ -mesurable ;

(b) soit que  $\mathcal{B}$  est la tribu image de  $\mathcal{T}$  par  $\xi$  (lorsque  $\mathcal{X}$  n'est pas muni a priori d'une structure d'**espace probabilisable**).

(ii) Si  $P$  est une **mesure de probabilité** définie sur  $\mathcal{T}$ , on peut associer à l'**espace probabilisé** (fondamental) défini par le triplet  $(\Omega, \mathcal{T}, P)$  un « espace probabilisé image ». En effet, la probabilité image de  $P$  par  $\xi$  est la **loi de probabilité**  $P^\xi$  de la **va**  $\xi$  (cf **mesure image**). L'espace probabilisé  $(\mathcal{X}, \mathcal{B}, P^\xi)$  obtenu est, par définition, l'**espace d'observation image** de  $(\Omega, \mathcal{T}, P)$  par  $\xi$  : c'est cet espace qui est généralement pris en compte dans l'analyse statistique.

(iii) Bien que généralement, ou génériquement, appelé « espace d'observation » (cf **variable observable**), il peut arriver :

(a) que certaines **parties** de cet espace ne soient pas, en réalité, « observables » : ie les variables ne sont pas partout observables dans  $\mathcal{X}$  (cf **censure, troncature**) ;

(b) qu'une variable elle-même soit complètement **inobservable** : elle peut simplement servir à l'analyse probabiliste ou à l'analyse statistique. Ainsi, dans certaines théories, aussi bien que dans les **représentations statistiques** associées, on peut définir des **va** inobservables (**variable latente, variable cachée**, etc) qui jouent le rôle de **variables sous-jacentes** au **phénomène** étudié. Le schéma théorique, ainsi que le modèle statistique, résultant de divers calculs, conduisent généralement à une forme « finale », ou forme « terminale », dans laquelle toutes les variables sont des **observables** (cf **observabilité**), ce qui permet de réaliser l'**inférence statistique** ;

(c) qu'une variable soit observée avec **erreur**, ou encore qu'elle soit altérée par d'autres variables non prises en compte (cf **aberration**).

Dans ces divers cas, par abus de langage et par commodité, on peut encore parler d'espace d'observation.

(iv) C'est grâce à l'observation d'un **phénomène** que le scientifique (**homme de l'art** ou **statisticien**) peut en analyser les propriétés (**structure**, fonctionnement, évolution passée) afin d'en tirer parti : **prévision** d'évolution future.

Aussi, la **production statistique** est-elle basée sur un **système d'observation** (cf aussi **système statistique**) qui autorise, a priori, cette observation à divers niveaux : microscopique, mésoscopique ou macroscopique. La définition en termes de « finesse » de ces niveaux relève alors de la nature du phénomène ou du **domaine de connaissance** au sein duquel ce phénomène est repéré et circonscrit.