EXPÉRIENCE ALÉATOIRE (B1, B4, E, F, L)
(27 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

Toute phase d'**expérimentation**, destinée à l'étude d'un **phénomène**, passe par la réalisation d'une (ou plusieurs) expérience(s) (cf **plan d'expérience**).

(i) Une **expérience aléatoire** est une expérience dont le **contexte statistique** (notamment le résultat) est soumis à l'effet du « **hasard** ». Elle se décrit par le choix aléatoire d'un élément ω (appelé **élément aléatoire**) appartenant à l'**ensemble « fondamental »** Ω de tous les **événements** (ou résultats) possibles. On confond souvent Ω avec la « description » même de l'expérience.

Un **événement aléatoire** $A \subset \Omega$ est alors représenté par l'ensemble des **unités élémentaires** ω de l'expérience qui le réalisent.

L'élément ω est aussi appelé **événement simple**, tandis que A est appelé **événement complexe**.

Dans le cas d'un plan d'expérience, ω est généralement une **unité statistique** appelée **unité expérimentale**, et A peut représenter le **dispositif expérimental**.

On peut alternativement définir une **expérience aléatoire** comme un **espace probabilisé** $(\Omega, \mathcal{T}, P)$. On supose souvent, implicitement, que la **mesure de probabilité** P définie sur $\mathcal{T}$ appartient à une **famille** $\mathcal{P}$ de probabilités possibles. L'expérience aléatoire se décrit alors à l'aide du **modèle statistique** $(\Omega, \mathcal{T}, \mathcal{P})$.

(ii) Généralement, diverses **variables** caractérisent les éléments ω participant à l'expérience : il s'agit de « **descripteurs** » (**attributs** ou **caractères statistiques**) κ (cf **variable qualitative**) ou de **mesures** numériques ξ (cf **variable numérique**) portant sur le phénomène considéré. Ces variables sont observées pendant l'expérience (cf **observation**), ce qui se traduit mathématiquement par la définition de deux espaces d'observation, $(\mathcal{X}, \mathcal{B})$ et $(\mathcal{K}, \mathcal{E})$, supposés mesurables, d'où l'**espace d'observation** produit $(\mathcal{X} \times \mathcal{K}, \mathcal{B} \otimes \mathcal{E})$, et d'une **variable aléatoire** $\zeta = (\xi, \eta)$ définie sur Ω et à valeurs dans $\mathcal{X} \times \mathcal{K}$. On note $P^{\zeta} = P^{(\xi, \kappa)}$ la loi de ζ (ie l'image de P par ζ) Une « valeur » $(x, k) \in \mathcal{X} \times \mathcal{K}$ tq $(x, k) = (\xi(\omega), \kappa(\omega))$ est aussi appelée « **résultat (d'observation)** », et l'ensemble $B \times E = (\xi(A), \eta(A)) \in \mathcal{B} \times \mathcal{E}$ représente l'ensemble des résultats obtenus à partir du dispositif expérimental $A \in \mathcal{T}$.

Par suite, les **résultats expérimentaux**, ou les **observations expérimentales**, peuvent se représenter à l'aide d'un **modèle image** $(\mathcal{X}_N \times \mathcal{K}_N, \mathcal{B}_N \otimes \mathcal{E}_N, \mathcal{P}^{(X, K)})$ dans lequel $\mathcal{X}_N \times \mathcal{K}_N$ désigne l'espace d'observation d'ensemble, $\mathcal{B}_N \otimes \mathcal{E}_N$ la **tribu engendrée** sur cet espace, (X, K) un ensemble de N observations du **couple aléatoire** (ξ, κ), $P^{(X, K)}$ la loi de (X, K) (image de P par (X, K)) et $\mathcal{P}^{(X, K)}$ la famille de telles lois (on suppose ici fini le nombre N d'observations).

C'est généralement sur ce dernier modèle que l'on applique une **procédure statistique**, en faisant ainsi abstraction du modèle initial (ou modèle de base). Autrement dit, selon le **contexte statistique**, on peut s'intéresser aux **unités expérimentales** (sélection, organisation, etc) aussi bien qu'au observations des variables effectuées sur ces unités : ces observations concernent donc des descripteurs (**attributs** ou **caractères statistiques**) caractérisant ces unités.

En théorie, une expérience peut être conçue comme pouvant être soit déterministe (ou quasiment telle), soit aléatoire :

(a) le caractère déterministe se traduit par le fait que les résultats obtenus sont intangibles (ou considérés comme tels) ;

(b) cependant, il est toujours possible de considérer une expérience déterministe comme une expérience aléatoire particulière : celle où toute **loi de probabilité** supposée intervenir est une **loi dégénérée**, ie une **loi de DIRAC** placée en des points donnés (ie les observations elles-mêmes).

(iii) C'est pourquoi l'on considère très généralement que toute expérience est soumise à un contexte aléatoire, et que toute expérience aléatoire bien spécifiée (cf **spécification**) peut se représenter par l'un des espaces probabilisés, $(\Omega, \mathcal{T}, P)$, $(\mathcal{X} \times \mathcal{K}, \mathcal{B} \otimes \mathcal{E}, P^{(\xi, \kappa)})$ ou $(\mathcal{X}_N \times \mathcal{K}_N, \mathcal{B}_N \otimes \mathcal{E}_N, P^{(X, K)})$ définissant resp les modèles $(\Omega, \mathcal{T}, \mathcal{P})$, $(\mathcal{X} \times \mathcal{K}, \mathcal{B} \otimes \mathcal{E}, \mathcal{P}^{(\xi, \kappa)})$ ou $(\mathcal{X}_N \times \mathcal{K}_N, \mathcal{B}_N \otimes \mathcal{E}_N, \mathcal{P}^{(X, K)})$ précédents.

En pratique, la « vraie » mesure de probabilité P (ou la « vraie » loi $P^{(\xi, \kappa)}$) n'est pas connue. On considère plutôt un **modèle** $(\Omega, \mathcal{T}, P)_{P \in \mathcal{P}}$, dans lequel $\mathcal{P}$ est une famille de probabilités susceptibles de générer les résultats, ou encore un modèle $(\mathcal{X} \times \mathcal{K}, \mathcal{B} \otimes \mathcal{E}, P^{(\xi, \kappa)})$ (avec $P^{(\xi, \kappa)} \in \mathcal{P}^{(\xi, \kappa)}$), où $\mathcal{P}^{(\xi, \kappa)}$ est la famille des lois, image de $\mathcal{P}$ par ζ, dont l'une est censée engendrer les observations issues de l'expérience. Dans l'une ou l'autre de ces familles, on cherche à évaluer et à « repérer » la « vraie probabilité » P ou la « vraie » loi $P\mathcal{P}^{(\xi, \kappa)}$ qui sera considérée comme étant à l'origine des résultats ou des observations de l'expérience (cf **problème d'estimation**, **problème de test**).

En pratique, le repérage d'une « partie » seulement de ces probabilités suffit (cf **caractéristique** d'une loi).

(iv) Une propriété fondamentale d'une expérience ou d'une expérimentation, aléatoire ou non, est la possibilité de sa **répétitivité**, ou **reproductibilité**. On parle aussi parfois de **renouvellement d'une expérience** (cf **renouvellement**).

(a) une **expérience reproductible** est une expérience que l'on peut recommencer et tq sa **répétition** conduit à des résultats ou observations « voisins », aux **erreurs** d'**échantillonnage** près : ainsi, plusieurs échantillons A conduiront à des **moyennes empiriques** peu dispersées (au sens de la **variance** des moyennes).

La **propriété de reproductibilité**, ou **propriété de répétitivité**, **d'une expérience** fonde la **loi des grands nombres** empirique : si une expérience Ω est répétée $N \in$ **N*** fois, alors le **rapport** entre le nombre N_A de réalisations d'un **événement** donné $A \in \mathcal{T}$ et le nombre N de répétitions de l'expérience (ou **fréquence empirique**) tend à se stabiliser. La limite concevable $p(A) = \lim_{N \to +\infty} N_A / N$ est parfois (improprement) appelée **probabilité statistique** (il s'agit d'un estimateur de (A)), et l'application $p : A \mapsto p(A)$ définit une **mesure de probabilité** sur $\mathcal{T}$.

Dans certains **domaines de connaissance** (eg physique), les observations peuvent souvent être engendrées de façon répétitive et l'**homme de l'art** peut faire varier ou contrôler les conditions de l'expérimentation (eg dosages différents) ;

(b) dans d'autres domaines (eg météorologie, biologie, écologie, sociologie), ceci n'est guère possible et les observations constituent souvent des **données préalables** ou **données inamovibles** pour le statisticien. On dit parfois que, dans le domaine considéré, **l'histoire ne se répète pas** (unicité de l'histoire, non retour, irréversibilité, etc). Dans ce cas, il n'est possible d'observer qu'un seul résultat : eg l'ensemble des statistiques décrivant l'évolution, non répétable, d'un **système** (eg météorologique, économique, etc).

L'analyse statistique du **phénomène** (eg **test d'hypothèses**, **estimation**, **prévision**, **classification**, etc) est toujours possible. Les méthodes peuvent cependant différer : on peut dire, dans ce sens, que l'analyse portant sur des **observations non renouvelables** (cf aussi **monographie**) généralise, d'une certaine manière, celle portant sur des observations renouvelables ou contrôlables.

A titre d'exemple, il est possible d'estimer, dans certains cas, les caractéristiques d'intérêt d'un **processus stochastique** en observant une seule **trajectoire** de ce processus, voire même une portion de cette trajectoire. Ceci est souvent le cas pour un processus ergodique (cf **ergodicité**) ou pour un **processus stationnaire**. Dans les deux types de situations rencontrés ci-avant, les conditions générales de déroulement des phénomènes à expliquer sont considérées, la plupart du temps, comme aléatoires.

C'est un des apports majeurs des notions d'expérience aléatoire et de modèle statistique que de permettre l'analyse de phénomènes dont la construction répond à des schémas différents (cf notamment **plan d'expérience**).

(v) D'un point de vue terminologique, le couple $(\Omega, \mathcal{T})$, constitué d'un **ensemble fondamental** Ω et d'une structure mesurable (souvent une **tribu d'événements**) sur Ω, est aussi appelé **épreuve (aléatoire)**. Cette épreuve est dite :

(a) déterministe si $\mathcal{T}$ est dotée d'une probabilité dégénérée $P = \delta_{\omega 0}$ (**loi de DIRAC** au point $\omega_0 \in \Omega$) ;

(b) aléatoire si $\mathcal{T}$ est dotée d'une probabilité non dégénérée.

L'ensemble fondamental Ω composé des résultats élémentaires est encore appelé **catégorie d'épreuves**, et un élément $\omega \in \Omega$ une **épreuve simple** (ou **épreuve élémentaire**). Enfin, $A \in \mathcal{T}$ est appelé **événement complexe** (ou **épreuve complexe**).