

## PARAMÈTRE (C5, G2)

(03 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **modélisation** d'un **phénomène** (aléatoire) implique généralement l'importante notion de **paramètre**. Un paramètre sert à caractériser, en un certain sens, le processus **aléatoire** qui conduit à la réalisation des grandeurs décrivant ce phénomène. Il dépend, en partie, de la façon dont le **statisticien** aborde un **problème statistique** ou formalise un **modèle statistique**. C'est donc une notion relative, mais fondamentale, sur laquelle le statisticien procède à l'**inférence statistique**.

Souvent, un paramètre s'impose de façon naturelle, dans la mesure où son interprétation est évidente ou imposée a priori.

Dans certaines circonstances, un paramètre est parfois appelé « **constante** », par distinction avec la notion de **variable**. Mais cette distinction n'est pas absolue, dans la mesure où un paramètre peut, aussi, varier : eg cadre bayésien (cf **école bayésienne**, **théorie bayésienne**), paramètre dépendant d'**observations** (cf **modèle avec coefficients variables**), etc.

(i) Soit  $(\Omega, \mathcal{F})$  un **espace mesurable** et  $\Theta$  l'**ensemble** des « **indices** » d'une **famille**  $(P_\theta)_{\theta \in \Theta}$  de **probabilités** définies sur  $\mathcal{F}$ .

On définit un **modèle statistique paramétré**  $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$  soit comme un **espace probabilisable**  $(\Omega, \mathcal{F})$  doté d'une famille  $(P_\theta)_{\theta \in \Theta}$ , soit comme une **famille d'espaces probabilisés**  $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$  (cf **modèle statistique**).

On suppose parfois, notamment en **théorie bayésienne**, que  $\Theta$  est muni d'une **tribu de parties**  $\mathcal{B}_\Theta$ . L'espace mesurable  $(\Theta, \mathcal{B}_\Theta)$  est appelé **espace des paramètres**, ou **espace de paramétrage**, du modèle précédent. On parle aussi d'**espace de paramètre**, ou d'**espace de paramétrage**, du modèle.

Tout élément  $\theta \in \Theta$  est appelé **paramètre** du modèle. Cette notion peut être un « objet » simple (eg scalaire : nombre réel, etc) ou multiple (eg paramètre vectoriel, paramètre fonctionnel : **fonction de régression**, etc). C'est l'**homme de l'art**, ou le **statisticien**, qui précise éventuellement la nature du paramètre qui l'intéresse (**paramètre d'intérêt**) ou qui possède une signification concrète.

(ii) Dans certaines **situations statistiques**, la connaissance de la **vraie valeur**  $\theta^*$  prise par le paramètre  $\theta$  dans  $\Theta$  est censée spécifier entièrement l'« expérience » (aléatoire) qui gouverne effectivement le phénomène considéré.

Ainsi, dans une **simulation**, on suppose que  $\theta^*$  est connu, car donné a priori : par suite, la **probabilité**  $P_{\theta^*}$  est aussi connue. Le modèle statistique se réduit alors à un simple espace probabilisé  $(\Omega, \mathcal{F}, P_{\theta^*})$ .

Dans un **schéma de tirage probabiliste** (cf **schéma probabiliste**), on considère aussi que  $P_{\theta^*}$  est unique, mais la démarche suivie est inversée :  $P_{\theta^*}$  n'est pas connue, mais elle est déterminée par calcul, en fonction des particularités du schéma analysé. Autrement dit, on explicite  $P_{\theta^*}$  à partir de ses propriétés ou de particularités qu'on veut lui imposer.

Cependant, une situation aussi favorable est rare, et l'investigation porte essentiellement sur  $\theta^*$ , ie sur la valeur tq  $P_{\theta^*}$  régit effectivement le phénomène (donc les observations auxquelles il donne lieu). L'inférence consiste donc notamment à déterminer (**estimation**), puis à « sélectionner » (**test d'hypothèses**) une valeur  $\tilde{\theta}$  (**estimateur** de  $\theta^*$ ) acceptable du point de vue de la **modélisation** du phénomène.

En ce sens :

(a) si un espace  $(\Omega, \mathcal{F})$  est muni d'une famille (possible, ou admissible)  $\mathcal{P}$  de probabilités définies sur  $\mathcal{F}$ , on définit un **modèle statistique non paramétré**, ie un modèle de la forme  $(\Omega, \mathcal{F}, \mathcal{P})$ , dont l'étude est l'objet de la **Statistique non paramétrée**. La donnée de la famille  $\mathcal{P} = (P)_{P \in \mathcal{P}}$  ne précise pas d'autre « paramétrage » de  $P$  que par la famille  $\mathcal{P}$  elle-même ;

(b) le modèle  $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$  précédent est un « objet » de la **Statistique paramétrée**. Tout modèle peut donc être écrit sous cette forme ;

(c) dans les situations courantes, on considère que le paramètre est de type numérique (scalaire ou vectoriel), eg  $\Theta \subset \mathbf{R}^Q$ . On parle alors de **Statistique paramétrique** ;

(d) dans d'autres situations, le paramétrage n'est pas précisé ou concerne un espace ni scalaire, ni vectoriel : eg  $\Theta \subset \mathbf{R}^\infty$ , ou encore  $\Theta$  est un espace fonctionnel (**famille de densités de probabilité**, etc) : on parle alors de **Statistique non paramétrique** ;

(e) enfin, une situation intermédiaire est celle d'un modèle dont le paramétrage est en partie paramétrique, en partie non paramétrique. Ainsi, on peut avoir à étudier une **fonction de régression** non linéaire  $(x, \theta) \mapsto E(\eta / \xi = x) = f(x, \theta)$  dont la « forme »  $f$  n'est pas connue, mais dont le paramètre  $\theta$  possède une signification pour le statisticien. On parle alors de **Statistique semi-paramétrique**.

(iii) La différence formelle entre la **Statistique paramétrée** et la **Statistique non paramétrée** peut être vue comme suit :

(a) en posant eg  $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$ , on passe de la seconde à la première ;

(b) en posant  $\Theta = \mathcal{P}$  et  $\theta = P$ , on passe de la première à la seconde, car on peut toujours écrire  $\mathcal{P} = (P)_{P \in \mathcal{P}}$ .

L'usage définit généralement un **modèle paramétrique** comme un modèle paramétré dans lequel  $\Theta \subset V$ , où  $V$  est un **espace vectoriel** de dimension finie ( $\dim V = Q < +\infty$ ), défini le plus souvent sur le corps  $\mathbf{K}$  (avec  $\mathbf{K} = \mathbf{R}$  ou  $\mathbf{K} = \mathbf{C}$ ). Le vecteur  $\theta = (\theta_1, \dots, \theta_Q) \in \Theta$  est aussi appelé **vecteur des paramètres**  $\theta_q$ , avec  $q \in N_Q^*$ .

Par distinction, on appelle **modèle non paramétrique** un modèle statistique pour lequel  $\Theta$  ne peut se mettre sous la forme précédente, ie un modèle non indexable par un espace de dimension finie. Un modèle paramétré (resp paramétrique) est parfois appelé **modèle à indices**. Un modèle non paramétré (resp non paramétrique) est alors, dans cette terminologie, un **modèle à indexation « naturelle »**, ie dont l'indexation se fait à l'aide de la famille de probabilités qui le caractérise elle-même.

(iv) Ce qui précède vaut encore pour (se transpose à) un **modèle image**. Ainsi, si  $(\mathcal{X}, \mathcal{B})$  est un **espace d'observation** et  $X : \Omega \mapsto \mathcal{X}$  une **va** (ou une **statistique**) donnée (eg un **échantillon**), on appelle **modèle image paramétré** un modèle  $(\mathcal{X}, \mathcal{B}, (P_\theta^X)_{\theta \in \Theta})$ , aussi noté  $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$ , dans lequel  $P_\theta^X$  désigne,  $\forall \theta \in \Theta$ , l'image  $X(P_\theta)$  de  $P_\theta$  par  $X$ , ie (par définition) la **loi de probabilité** de  $X$ .

(v) On doit distinguer entre :

(a) l'ensemble  $\Theta$  des valeurs du paramètre  $\theta$  des **mesures de probabilité**  $P_\theta$  ou des **loi de probabilité**  $P_\theta^X$ ;

(b) l'ensemble  $\Gamma$  des valeurs  $\gamma$  des **caractéristiques légales** de ces mêmes  $P_\theta$  ou  $P_\theta^X$ . Ces caractéristiques peuvent être eg une **densité**, un **mode**, un **moment**, un **quantile**, etc.

Ainsi, lorsque  $P_\theta^X = \mathcal{N}_Q(\mu, \Sigma)$  (**loi normale multidimensionnelle**), avec  $Q \geq 1$ , on a seulement  $\Theta = \mathbf{R}^Q \times C_Q(\mathbf{R})$ , où  $C_Q(\mathbf{R}) = M_Q^+(\mathbf{R}) \cap S_Q(\mathbf{R})$  est le **cône** des **matrices symétriques** réelles semi-définies positives. Autrement dit, le couple  $(\mu, \Sigma)$  caractérise (ou détermine) entièrement la loi. Or,  $P_\theta^X$  possède naturellement d'autres caractéristiques que son **espérance** ou sa **dispersion** : **intervalle interquartile** (eg intervalle interquartile), **coefficient de variation**, **densité** pr à  $\lambda_Q$ , etc.

(vi) On peut ne pas s'intéresser à l'ensemble des paramètres caractérisant la famille des lois  $P_\theta^X$ ,  $\forall \theta \in \Theta$ , mais seulement à certains d'entre eux, appelés **paramètres principaux**. Ainsi, dans l'exemple précédent, on peut ne s'intéresser qu'à la **moyenne**  $\mu \in \mathbf{R}^Q$ , le paramètre  $\Sigma$  jouant le rôle de **paramètre secondaire** (ou **paramètre importun**).

Une situation comparable est celle dans laquelle on ne cherche à « connaître » qu'une **fonction du paramètre**, ie la valeur  $\tau = g(\theta)$ , où  $g$  est une application donnée, au lieu du paramètre  $\theta$  lui-même.

(vii) On assimile souvent **méthode non paramétrique**, qui relève de la Statistique non paramétrique, et **méthode affranchie**, parfois dite **méthode libre**, ou **méthode exemptée**. Ces dernières, en effet, sont (en principe) libres de toute hypothèse (autres que mathématiques) relative aux **lp** mises en jeu.

En effet, de nombreuses méthodes non paramétriques nécessitent un certain nombre d'hypothèses minimales : eg **continuité** des densités (ou des **fr**), **indépendance**, **symétrie**, voire **normalité**, etc. Par « **méthodes affranchies** » on entend donc des méthodes qui ne supposent pas l'existence d'un modèle paramétrique particulier (au sens indiqué ci-dessus).

(viii) En pratique, on doit distinguer trois notions :

(a) la « **vraie** » **valeur du paramètre**, notée eg  $\theta^*$ , en général inconnue ;

(b) la « **valeur courante** » du paramètre, simplement notée  $\theta$ , qui sert de variable de calcul (**optimisation**, le plus souvent) ;

(c) l'**estimateur**  $\tilde{\theta} \in \Theta$  **du vrai paramètre**  $\theta^* \in \Theta$ , qui est une **statistique** fonction des observations et vérifiant certaines conditions (eg satisfaisant une **équation estimante**). Si  $N$  désigne la taille d'échantillon, on le note eg  $\theta_N$  pour rappeler qu'il dépend des **observations**.

Ainsi, un **modèle de régression multiple** (non linéaire) peut s'écrire (dans un **espace d'observation**)  $y = F(b) + u$ ,  $\forall b \in \mathbf{R}^Q$ , mais il devrait être écrit  $y = F(b^*) + u$ , car c'est la vraie valeur  $b^* \in \mathbf{R}^K$  qui est le paramètre d'intérêt, et que c'est elle qui définit la **perturbation aléatoire**  $u$  en fonction des observations  $(X, y)$ .

La **méthode des mco** consiste à minimiser pr à  $b$  la fonction **distance**  $\|y - F(b)\|^2$ , et elle conduit à un estimateur  $\hat{b}$  de  $b$ . En particulier, dans un **modèle de régression** linéaire standard (cf **régression linéaire**)  $y = Xb + u$  (qui devrait être noté  $y = Xb^* + u$ ), avec  $E y / X = 0$  et  $V y / X = \sigma^2 I_N$ , si  $b^* \in \mathbf{R}^K$  est estimé par la méthode des mco, l'estimation obtenue  $\hat{b} = (X'X)^{-1} X'y$  définit un estimateur  $\hat{b} = t_N(X, y)$  calculé en minimisant pr à  $b$  la **forme quadratique**  $q(b) = \|u\|^2 = \|y - Xb\|^2$ . On note parfois, dans ce cas,  $\beta$  la vraie valeur (ou la valeur courante) du paramètre, et  $b$  l'estimation du paramètre.

(ix) L'**estimateur**, ou sa valeur estimée (**estimation**), est diversement noté(e), souvent :

(a) en association avec une **méthode d'estimation** particulière qui le définit : eg  $\hat{\theta} \in \Theta$  pour une **méthode de moindres carrés**, etc ;

(b) en explicitant la taille de l'échantillon : soit « petite », soit « grande » (cf eg **modèle asymptotique**). La notation utilisée comporte alors la taille de l'échantillon qui définit l'estimateur : eg  $\theta_N$  pour un N-échantillon.

(x) Si  $\Theta_0 \subset \Theta$ , on peut souvent « représenter »  $\Theta_0$  de deux façons classiques (cf **contrainte sur les paramètres, paramètre borné, théorie des tests**) :

(a) soit à l'aide de contraintes de la forme (sous-variété implicite) :

$$(1) \quad \Theta_0 = \{\theta \in \Theta : g(\theta) = 0\},$$

qui peut aussi être une inégalité, où  $g : \Theta \mapsto L$  est donnée ;

(b) soit à l'aide d'une « **paramétrisation** » de la forme (sous-variété paramétrée) :

$$(2) \quad \Theta_0 = \{\theta \in \Theta : \theta = h(\lambda), \forall \lambda \in L\},$$

où  $L$  et  $h : L \mapsto \Theta$  sont donnés.

(xi) La façon dont on définit le paramètre ou le « **paramétrage** » dépend donc du problème traité. Un paramètre doit être, de préférence, interprétable dans le contexte du **phénomène**, donc du **domaine de connaissance**, où il intervient.

A titre d'exemples :

(a) physique (thermodynamique) : l'équation élémentaire d'état des gaz parfaits (loi de R. BOYLE - E. MARIOTTE) s'écrit  $P V = R T$ , où  $V$  est le volume molaire,  $T$  la température et  $P$  la pression. La « constante »  $R$  est appelée **constante universelle des gaz** et s'exprime eg à l'aide de l'**unité de mesure** J / mol K (ie JOULE / volume molaire et d° KELVIN). En considérant cette constante comme un paramètre, sa valeur est estimée, après **expérimentation**, à  $R \hat{=} 8,314$  (en J mol<sup>-1</sup> K<sup>-1</sup>) ;

(b) physique (électromagnétisme) : la loi de C.A. COULOMB liant deux charges électriques 1 et 2 se traduit par une équation vectorielle (cf modèle d'attraction universelle de I. NEWTON)  $F = Q_1 Q_2 / (4 \pi \varepsilon d^2)$ , dans laquelle  $F$  désigne la force de NEWTON (exprimée en N),  $d$  la distance entre 1 et 2 (exprimée en m),  $Q_i$  la charge électrique de  $i = 1$  ou 2. La constante  $\varepsilon$  est la **permittivité** du milieu ambiant (exprimée en C<sup>2</sup> N<sup>-1</sup> m<sup>-2</sup>). Après expérience dans le vide absolu, la valeur de ce paramètre est estimée à  $\varepsilon \hat{=} 8,854 \times 10^{-12}$  (en C<sup>2</sup> N<sup>-1</sup> m<sup>-2</sup>) ;

(c) physique (optique) : un photon possède une énergie de radiation  $e$  (exprimée en Joule) proportionnelle à sa fréquence  $n$  (exprimée en Hertz), soit  $e = h n$  (cf aussi **énergie d'un système**). Le coefficient de proportionnalité  $h$  est appelé **constante de M. PLANCK** : l'estimation de ce paramètre est égale à  $h \hat{=} 6,6256 \times 10^{-34}$  (en J s<sup>-1</sup>, ie JOULE / seconde) ;

(d) psychologie : l'estimation des capacités intellectuelles (**quotient intellectuel**, ou **QI**) utilise classiquement la **méthode des scores**. La démarche consiste à observer, sur un échantillon d'individus, des réponses à des questions (tests psychométriques sous forme de QCM, etc) censées « révéler » diverses aptitudes (capacités d'analyse, de mémorisation, d'association, de réactivité, de reconnaissance de formes, etc) à partir des réponses aux questions (**tests d'aptitude**). Le schéma causal est synthétisé dans une **fonction score** qui relie les réponses aux questions. Cette fonction est ensuite estimée à partir des observations. L'observation d'une réponse apportée par un nouvel individu permet alors de « situer » celui-ci sur l'**échelle** de la fonction estimée (cf **discrimination**) : cette échelle (gaussienne et scalaire) est en général interprétée à la façon d'une **échelle ordinale** (eg 50 % de la population possède un QI « normal » lorsque  $90 \leq \text{QI} \leq 110$ , etc) ;

(e) sociologie (économie) : une **fonction de consommation** macro-économique peut s'écrire eg sous la forme  $C_t = a + b p_t + c R_t + d C_{t-1} + e P_t$ , avec  $0 \leq C_t \leq R_t$ , dans laquelle  $p_t$  représente le niveau des prix de la période courante  $t$ ,  $C_t$  désigne le flux de consommation finale des ménages pendant cette période,  $R_t$  est le flux de leur revenu disponible brut et  $P_t$  le stock de leur patrimoine en fin de période précédente. Le coefficient  $c = \partial C_t / \partial R_t$ , sans dimension pr aux unités de mesure, est appelé **propension marginale à consommer** (le revenu). Il est estimé, selon les pays, selon les périodes, ou encore selon le type d'observation (**séries temporelles** ou **coupes instantanées**), à une valeur généralement comprise entre 0,6 et 0,9.

(xii) Dans un cadre paramétrique, le paramètre  $\theta$  d'une loi  $P_\theta^X$  spécifie complètement celle-ci : autrement dit, si  $\theta$  est connu et si  $(P_\theta^X)_{\theta \in \Theta}$  est une **famille de lois identifiable**, alors  $P_\theta^X$  est complètement déterminée.

Ceci n'est pas toujours le cas, en général. Ainsi, lorsque :

$$(3) \quad \eta = \rho(\xi) + \varepsilon$$

représente un **modèle de régression** non linéaire dans lequel  $\rho$  est la fonction de régression,  $(\xi, \eta)$  est **observable** et  $\varepsilon$  **inobservable**, on peut considérer trois situations :

(a)  $\rho$  est tq  $\rho(\xi) = f(\xi, b)$ , où  $f$  est inconnue et  $b \in \mathbf{R}^Q$  est inconnu : il s'agit d'un **problème semi-paramétrique**, l'un des paramètres étant de dimension infinie (si les indices des observations sont au moins dénombrables), l'autre étant de dimension  $Q$  ;

(b)  $\rho$  est tq  $\rho(\xi) = f(\xi, b)$ , où  $f$  est connue (eg linéaire) et  $b \in \mathbf{R}^Q$  est inconnu : ce contexte est celui d'un **problème paramétrique** usuel (avec  $Q$  paramètres scalaires  $b_q$ ) ;

(c)  $\rho$  est totalement inconnue, ie appartient à l'ensemble des fonctions de régression possibles liant  $\eta$  à  $\xi$  : il s'agit alors d'un **problème non paramétrique**. Parfois, un développement de B. TAYLOR d'ordre 1 (ie une **approximation linéaire**) est considéré(e) comme suffisant(e), ie :

$$(4) \quad \eta = \rho(0) + (1!)^{-1} (D\rho(0))(\xi) + \varepsilon(\xi),$$

où  $D\rho(0) = (\text{Grad } \rho)(0)$  (**gradient** évalué à l'origine  $0 \in \mathbf{R}^K$ ). On se ramène à un **modèle linéaire** usuel (avec terme constant) en posant :

$$(5) \quad \eta = b_0 + b_1 \xi_1 + \dots + b_K \xi_K + \varepsilon(\xi).$$

(xiii) Si  $\mathcal{P}$  (resp  $\mathcal{P}^\xi$ ) est une famille de probabilités (resp de lois) et  $(\Gamma, \mathcal{B}_\Theta)$  un **espace caractéristique** mesurable donné (cf **caractéristique**) (eg  $(\mathbf{R}, \mathcal{B}_\mathbf{R})$ ), on appelle parfois encore, de façon extensive, **paramètre** de  $P \in \mathcal{P}$  (resp de  $P^\xi \in \mathcal{P}^\xi$ ) toute application (mesurable), dite **application caractéristique**,  $c : \mathcal{P} \mapsto \Gamma$  (resp  $\mathcal{P}^\xi \mapsto \Gamma$ ).

(xiv) Enfin, si  $(\mathcal{Z}, \mathcal{D}, P_{\theta^\zeta})_{\theta \in \Theta}$  représente un modèle image sous forme paramétrée, on doit distinguer deux notions de « **loi** » (cf **loi multivariée**, **relation fonctionnelle**, **loi scientifique**) :

(a) la **loi de probabilité**  $P_{\theta^\zeta}$  de la variable  $\zeta$ . L'une de ces lois est supposée gouverner le phénomène étudié (ie le « comportement » de  $\zeta$ ) : cette loi possède donc son paramètre propre  $\theta$  ;

(b) la **loi du phénomène** lui-même, qui implique souvent une notion de **causalité**. Si  $\zeta$  est un **couple aléatoire**  $\zeta = (\xi, \eta)$  et si la notion de **régression** possède un sens (eg en termes d'espérance), on peut s'intéresser à une **fonction de régression** tq :

$$(6) \quad x \mapsto f(x, \psi) \text{ ou } f_\psi(x) = E(\eta / \xi = x) + \varepsilon.$$

Cette fonction comporte ici un paramètre  $\psi$  « incorporé » dans l'**espérance conditionnelle**  $E(\eta / \xi = x) = \int y y \, dP_{\theta^{(\eta/\xi)}}(y)$ , où  $P_{\theta^{(\eta/\xi)}}$  désigne la **loi conditionnelle** de  $\eta$  sachant  $\xi$ . Mais, dans la mesure où elle dérive d'un calcul basé sur la loi  $P_{\theta^\zeta}$  précédente, le paramètre  $\psi$  n'a aucune raison d'être, sauf comme paramètre résultant du calcul d'espérance précédent, donc déduit de  $\theta$ .