

ESTIMATION NON PARAMÉTRIQUE (H2)

(22 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

On appelle **problème d'estimation non paramétrique** un **problème d'estimation** attaché à un **modèle non paramétrique** (cf aussi **modèle paramétrique**, **paramètre**).

(i) Ce problème n'est pas, par nature, différent d'un **problème d'estimation** « paramétrique » usuel. Cependant, la méthode de résolution des deux problèmes peut différer : notamment, lors de la résolution du **problème d'optimisation** attaché à chacun de ces problèmes. En effet :

(a) dans le cas paramétrique (classique ou bayésien), la **fonction de risque** à minimiser comporte un nombre « fini » de paramètres, principaux ou secondaires (cf **paramètre principal**), d'où la mise en oeuvre des méthodes usuelles de la **programmation mathématique**. La **fonction de risque** est optimisée sur un espace de dimension finie (souvent, un **espace vectoriel** réel) ;

(b) dans le cas non paramétrique, la fonction de risque comporte un nombre a priori « infini » de paramètres : eg estimation d'une **densité de probabilité**, d'une **densité spectrale**, d'une **fonction caractéristique**, etc. Elle doit donc être optimisée sur un espace de dimension infinie (généralement, un **espace fonctionnel**).

Dans les deux cas, cependant, la méthode utilisée définit un estimateur fondé, plus ou moins directement, sur la **loi empirique** associée à l'**échantillon** disponible.

(ii) Soit $(\mathcal{X}, \mathcal{B}, \mathcal{P}^{\mathcal{X}})$ un **modèle statistique** non paramétrique dans lequel X est l'échantillon observé et $(\Gamma, \mathcal{B}_{\Gamma})$ un espace de **caractéristiques** (eg **densité**, **fonction caractéristique**, **quantile**, etc).

On note $c : \mathcal{P}^{\mathcal{X}} \mapsto \Gamma$ une **application « caractéristique »**, ie une application associant, à toute **lp** $P^{\mathcal{X}} \in \mathcal{P}^{\mathcal{X}}$, susceptible de régir le **phénomène** aléatoire étudié, une liste de caractéristiques d'intérêt (densité, fc, quantile, etc).

Un **estimateur (strict)** de la caractéristique $\gamma = c(P^{\mathcal{X}})$ est une **application mesurable** $t : \mathcal{P}^{\mathcal{X}} \mapsto \Gamma$. Une **fonction de perte** $L : \Gamma^2 \mapsto \mathbf{R}_+$ est tq la valeur $L(t(X), c(P^{\mathcal{X}})) = L(T, \gamma)$ représente le « coût » encouru par le **statisticien** lorsqu'il estime la « vraie » valeur γ^* de la caractéristique à l'aide de l'estimateur $T = t(X)$ alors que celle-ci est γ (cf **fonction de coût**).

Comme dans le cas « paramétrique », la **fonction de risque** est définie comme **espérance mathématique** de la fonction de perte précédente, ie :

$$(1) \quad R(t, \gamma) = E L(t(X), \gamma) = \int_{\mathcal{X}} L(t(x), c(P^{\mathcal{X}})) dP^{\mathcal{X}}(x).$$

C'est cette fonction qui doit, en principe, être minimisée pr à γ .

Or, Γ est, en général, un ensemble infini : eg Γ est l'ensemble des densités $f = dP^X / d\mu$ associées aux $lp P^X$ et uniformément dominées par une **mesure positive** μ (généralement, une **mesure σ -finie**), définie sur \mathcal{B} (on suppose que les densités sont μ -équivalentes, ie que l'ensemble des valeurs de \mathcal{X} sur lesquelles elles diffèrent est de μ -mesure nulle : ie on se place directement sur l'**espace quotient** de Γ par la **relation d'équivalence** précédente).

La méthode d'optimisation peut, dans le cas général, être conçue en relation avec un **espace de BANACH**, mais il n'est pas toujours possible de trouver une solution (non triviale) au problème de minimisation de la fonction de risque précédent (cf **différentiabilité, optimisation, calcul des variations**).

(iii) En pratique, on peut :

(a) soit modifier, ou « **pénaliser** », la fonction de risque, en sorte que le problème de minimisation comporte une solution simple, tout en gardant un sens (cf **pénalisation**) ;

(b) soit, le plus souvent, restreindre de façon « raisonnable » l'ensemble Γ sur lequel R doit être minimisée (cf eg **méthode des tamis**).

(iv) L'estimation non paramétrique peut s'étendre, en principe, au contexte bayésien lorsque soit \mathcal{D}^X , soit $c(\mathcal{D}^X)$, soit Γ , peut être muni d'une **mesure de probabilité** a priori (cf **probabilité a priori**). Ainsi, si les **lp** ont leur support dans un **simplexe** (eg **lois discrètes**), on peut munir leur ensemble d'une loi a priori du type **loi de DIRICHLET** (cf **support d'une probabilité**).

(v) La difficulté liée à l'estimation non paramétrique tient au fait que le nombre d'observations est fini : par suite, on ne peut guère (en général) estimer qu'une « partie » de la loi P^X (cf **degré de liberté**).

Ainsi, avec un échantillon $X = (X_1, \dots, X_N)$ et une densité f continue dont le support est $\text{Supp } f = \mathbf{R}$ (cf **support d'une fonction**) :

(a) on peut estimer une caractéristique scalaire (ou vectorielle, si l'espace est de dimension finie), eg estimer le couple $(E \xi, V \xi)$ (**espérance** et **variance**) ou le couple $\{Q_{1/2} \xi, |Q_{3/4} \xi - Q_{1/4} \xi| / |Q_{1/2} \xi|\}$ (**médiane** et **déviations interquartiles** relative) (cf aussi **intervalle interquartile**) ;

(b) mais on ne peut (sans hypothèse supplémentaire) estimer f toute entière avec X . On doit donc ajouter des hypothèses sur f pour réduire le problème à un nombre moindre de caractéristiques (à estimer).

Ainsi, lorsqu'une densité est estimée par la **méthode du noyau**, l'estimateur $f_N \sim$ est de la forme :

$$(2) \quad f_N \sim (x) = (1/h) \sum_{n=1}^N J \{(x - X_n) / h\}, \quad \forall x \in \mathbf{R},$$

où le **noyau** J est donné. On doit donc ici estimer h .

Si le nombre d'observations est infini (eg **trajectoire** d'un **processus** en **temps** continu), il est possible d'estimer « davantage » de caractéristiques de la loi P^X . L'étude des **propriétés asymptotiques** d'un estimateur non paramétrique joue donc un rôle important.