ÉCOLE BAYÉSIENNE (B4, G3)

Cette conception de la **Statistique** comme « science » doit son origine à Thomas BAYES et attache de l'importance aux**idées a priori** que l'**homme de l'art** ou le **statisticien** peuvent avoir sur un **phénomène** et, notamment, sur la **représentation statistique** supposée le régir.

Cette conception (développée eg par G. BOX, H. JEFFREYS, O. KEMPTHORNE, D. LINDLEY, L.J. SAVAGE, C. STEIN, G.C. TIAO ou A. ZELLNER) conduit à admettre l'exister **probabilités a priori**, ou de **probabilités subjectives**, qui traduisent des **degrés de croyance**.

(i) Si le **modèle statistique** est un **modèle image** paramétré $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$, l'**école bayésienne** munit l'ensemble Θ d'une **tribu de parties** $\mathcal{B}_\Theta$, laquelle est alors dotée d'une **mesure de probabilité** Π, appelée **(loi de) probabilité a priori** (cf **loi a priori**, **probabilité a priori**).

Le **problème de décision** (cf **décision statistique**) associé à cette approche permet alors, moyennant la donnée usuelle d'un **espace de décision**, d'un ensemble de **règles de décision** et d'une **fonction de perte**, de définir des **principes bayésiens** **d'induction statistique** : **règle de BAYES**, **règle de BAYES empirique**, etc (cf **inférence statistique**).

(ii) Les limites de cette optique tiennent surtout aux inconvénients suivants :

(a) difficulté d'expliciter Π, car cette mesure de probabilité n'est pas toujours aisée à exhiber ou complètement connue : information « floue » sur les valeurs possibles Θ du **paramètre** θ (cf **théorie des parties floues**) ;

(b) parfois, Π n'est pas une **probabilité**, et l'on peut avoir $\Pi(\Theta) \neq 1$. Ainsi, l'**état de complète ignorance** relativement à la « vraie » valeur θ^* du paramètre se traduit souvent par l'adoption d'une **mesure uniforme** Π sur Θ, mesure qui peut donc être de masse totale $\Pi(\Theta) = +\infty$, notamment dans le cas fréquent où Θ n'est pas un ensemble fini ;

(c) la lourdeur de certains calculs analytiques (passage des lois a priori aux lois a posteriori). Cet inconvénient est aujourd'hui résolu soit par les possibilités du calcul symbolique (qui permet de réaliser des calculs analytiques complexes), soit par les capacités des ordinateurs actuels en termes de calculs numériques.

(ii) Dans un **modèle bayésien** $\{(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}, (D, \mathcal{B}_D), (\Theta, \mathcal{B}_\Theta, \Pi), L\}$, si P_θ^X admet, $\forall\, \theta \in \Theta$, une **densité de probabilité** pr à une **mesure positive** μ donnée :

(1) $dP_\theta^X(x) / d\mu(x) = L(x, \theta)$

(**fonction de vraisemblance** considérée comme **densité conditionnelle** $f(x / \theta)$), et si Π admet pr à une mesure positive ν donnée, définie sur $\mathcal{B}_\Theta$, une **densité a priori** (**dérivée de NIKODYM-RADON**) :

(2) $d\Pi(\theta) / d\nu(\theta) = \pi(\theta)$,

le **théorème de BAYES**, ou **principe de la probabilité inverse**, donne la **densité de probabilité a posteriori** q du paramètre, conditionnellement aux **observations** (lorsque l'expression a un sens) :

(3) $q(\theta / x)$ ou $q_x(\theta) = L(x, \theta)\, \pi(\theta) / \int_\Theta L(x, \theta)\, \pi(\theta)\, d\nu$ ().

(iii) Pour estimer , le **principe du maximum de probabilité a posteriori** (cf aussi **principe bayésien**) consiste à minimiser la fonction $q(., x)$ pr à , ie à maximiser la fonction $\theta \mapsto L(x, \theta)\, \pi(\theta)$ (produit de la **vraisemblance** et de la densité a priori) pr à θ.

En cas d'ignorance absolue relative à Π, le choix d'une probabilité uniforme, ie tq (en supposant que $\nu(\Theta) < +\infty$) :

(4) $\Pi(B) = \nu(B) / \nu(\Theta)$, $\forall\, B \in \mathcal{B}_\Theta$,

est naturel (cf **mesure uniforme**). Ce choix conduit à une densité a priori π tq :

(5) $\pi(\theta) = 1 / \nu(\Theta)$.

Par suite, maximiser $q(. / x)$ pr à θ revient à maximiser $L(x, .)$ pr à θ, et la méthode d'estimation bayésienne précédente se réduit à la **méthode du maximum de vraisemblance**.

(iv) Au lieu de maximiser $q(. / x)$ pr à θ, on peut suivre un autre principe bayésien. Ainsi, on peut (lorsque les opérations ont un sens) considérer une **caractéristique** de **centralité** (**espérance**, **médiane**) de $q(. / x)$ (au lieu du « mode » relativement à θ précédent).

(v) Lorsque Θ n'est pas un **espace vectoriel** de dimension finie (ie lorsque le nombre de paramètres n'est pas fini, comme en Statistique non paramétrique), il est souvent délicat de définir une probabilité sur $\mathcal{B}_\Theta$ (cf **paramétrique**, **modèle paramétrique**).

(iii) L'**inférence bayésienne**, ou la **Statistique bayésienne**, constitue ainsi une approche des problèmes statistiques à partir de la théorie des **probabilités subjectives** : eg un « statisticien bayésien » admet qu'il existe une probabilité égale à $1 - \alpha \in]0,1[$ que la moyenne μ d'une **loi normale** $\mathcal{N}(\mu, \sigma^2)$ appartienne à l'intervalle [a,b] sans référence à la **théorie fréquentiste** des probabilités. Cette inférence se fonde donc essentiellement du **théorème de BAYES** (cf aussi **théorie fiducielle**).

Ce dernier nécessite donc la donnée de la fonction de vraisemblance (ie d'une **spécification** du modèle étudié) et de la loi a priori.

La loi a priori est parfois précisée à l'aide d'**informations** (**observation**, expériences, etc) parallèles, ou antérieures, au problème traité. Ces informations aident à « estimer » cette loi : les **méthodes bayésiennes empiriques** sont destinées à estimer séquentiellement la probabilité a priori (du moins, lorsqu'un cadre séquentiel est possible) (cf **théorie séquentielle**).

Enfin, lorsque la loi a priori est connue seulement de façon floue, il est encore possible d'étudier la **robustesse** de la solution d'un problème bayésien par rapport à des variations de cette loi a priori autour d'une loi supposée.

(iv) Par rapport à l'approche classique, l'approche bayésienne ajoute donc une hypothèse « statistique » supplémentaire, qui porte sur le paramètre du modèle. L'approche classique n'est cependant pas exempte d'hypothèses, notamment celle relative à la **spécification** du modèle précédent, qui correspond aussi à des idées a priori, mais non réellement « statistiques ».