Cours 9

8 Analyse Bayésienne non paramétrique

Dans le cadre paramétrique, le modèle statistique classique $(\mathscr{X}, \mathscr{A}, P_\theta, \theta \in \Theta)$, où $\mathscr{X}$ est l'espace des observations, $\mathscr{A}$ est la tribu associée, P_θ la loi de l'observation, Θ l'espace des paramètres ($\Theta \subset \mathbb{R}^s$), est transformé en un modèle statistique bayésien par la donnée d'une loi a priori $\pi(.)$ sur θ.
Dans le cadre non paramétrique, l'objet sur lequel on infère est non plus θ mais P_θ la distribution de l'observation ou d'une manière plus générale, une fonctionnelle $g(P_\theta)$. Ce peut être la fonction de répartition, la densité, la fonction de survie (ou de fiabilité), le taux de survie (ou de défaillance), l'intensité d'un processus, etc.
L'approche bayésienne de l'estimation d'une fonctionnelle nécessite donc la construction d'une loi a priori sur P_θ, la mesure probabilité associée à l'observation. Elle est maintenant considérée comme une **mesure aléatoire**.
Soit $\mathbb{M}$ l'ensemble des mesures de Radon sur $\mathbb{R}$ i.e. l'ensemble des mesures μ telle que, pour tout $A \in \mathscr{A}$, $\mu(A) < \infty$.
On peut donner la définition suivante d'une mesure aléatoire :

Définition 17 – *Une mesure aléatoire est une application mesurable de $(\mathscr{X}, \mathscr{A})$ dans $(\mathbb{M}, \mathscr{M})$.*

Soit η une mesure aléatoire, $\forall A \in \mathscr{A}$, $\eta(A)$ est une variable aléatoire.
Le processus associé à cette mesure aléatoire η sera caractérisée par la donnée de la densité jointe de $(\eta(A_1), \eta(A_2), \ldots, \eta(A_{m+1}))$ où $A_1, A_2, \ldots, A_{m+1}$ est une partition de $\mathscr{X}$.
η est un processus stochastique indexé par des éléments de $\mathscr{A}$.

Exemple 1 – Le processus de Poisson est un exemple simple de processus répondant aux conditions énoncées ci-dessus. Considérant, une mesure de comptage η sur $\mathbb{R}^+$ et supposant que pour tout A, $\eta(A)$ est une v.a. discrète qui suit une loi de Poisson de paramètre $\mu(A)$ où μ est une mesure de Radon. On dira que η est une mesure aléatoire de Poisson si :

$$\Pr(\eta(A) = k) = \frac{\mu(A)^k}{k!} \exp\{-\mu(A)\}\ , \quad k \in \mathbb{N}.$$

Et on definira le processus de Poisson d'intensité μ associée à cette mesure η par :

Définition 18 – *Soit μ une mesure finie non nulle sur $\mathbb{R}^+$.*
On dit que la mesure η est un processus de Poisson de paramètre μ, et on note $\eta \in \mathscr{P}(\mu)$, si, pour toute partition mesurable $(A_1, A_2, \ldots, A_{m+1})$ de $\mathbb{R}^+$, la loi de $(\eta(A_1), \eta(A_2), \ldots, \eta(A_m))$ est un produit de lois de Poisson de paramètre respectif $\mu(A_i)$, $i = 1, \cdots, m+1$.

Le processus fondateur de l'approche bayésienne des problèmes non paramétrique est le processus de Dirichlet. Rappelons tout d'abord, définition et propriétés de la loi de Dirichlet.

Définition 19 – *Un vecteur aléatoire $(U_1,U_2,...,U_m)$ à valeurs dans le sous-espace de $\mathbb{R}^m$:*

$$\Delta_m = \left\{ (u_1,u_2,\ldots,u_m) \; ; \; u_i \geq 0 \; , \; i = 1,2,\ldots,m \; ; \; \sum_{i=1}^{m} u_i \leq 1 \right\}$$

suit une loi de Dirichlet de paramètres $(\alpha_1,\alpha_2,\ldots,\alpha_m;\alpha_{m+1})$ si la densité de sa loi s'écrit :

$$\frac{\Gamma(\alpha_1+\alpha_2+\ldots+\alpha_{m+1})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{m+1})}.u_1^{\alpha_1-1}\cdots u_m^{\alpha_m-1}(1-u_1-\ldots-u_m)^{\alpha_{m+1}-1} \tag{11}$$

Remarquons que cette distribution est une manière de généralisation de la loi Bêta. Pour $m = 1$, on retrouve la loi Bêta de paramètres $(\alpha_1;\alpha_2)$.
On définit alors le **processus de Dirichlet** :

Définition 20 – *Soit α une mesure finie non nulle sur $\mathbb{R}^+$.*
On dit que la mesure P est un processus de Dirichlet de paramètre α, et on note $P \in \mathscr{D}(\alpha)$, si, pour toute partition mesurable $(A_1,A_2,\ldots,A_{m+1})$ de $\mathbb{R}^+$, la loi de $(P(A_1),P(A_2),\ldots,P(A_m))$ est une loi de Dirichlet de paramètres $(\alpha(A_1),\alpha(A_2),\ldots,\alpha(A_m);\alpha(A_{m+1}))$.

8.1 Estimateur de Bayes de la fonction de répartition

La loi de Dirichlet possèdent un certain nombre de propriétés qui permettent de construire des résultats sur le processus de Dirichlet qui seront utiles dans la suite.
La preuve du résultat suivant est laissée au soin du lecteur :

Proposition 8 – *Soient $(Y_1,Y_2,\cdots,Y_{m+1})$, $m+1$ variables aléatoires indépendantes de loi Gamma de paramètres respectifs $(\alpha_i,1)$, $i = 1,\cdots,m+1$.*
On considère les variables :

$$X_i = \frac{Y_i}{Y_1+\cdots+Y_{m+1}} \; , \quad i = 1,\cdots,m.$$

Alors $(X_1,\cdots,X_m)$ suit une loi de Dirichlet de paramètres $(\alpha_1,\alpha_2,\cdots,\alpha_m;\alpha_{m+1})$.

Ce résultat permet d'établir :

Proposition 9 – *Si $(X_1,X_2,\cdots,X_m)$ suit une loi de Dirichlet de paramètres $(\alpha_1,\cdots,\alpha_m;\alpha_{m+1})$ alors la loi marginale de $(X_{i_1},X_{i_2},\cdots,X_{i_q})$, $q < m$, suit une loi de Dirichlet de paramètres $(\alpha_{i_1},\cdots,\alpha_{i_q};\sum_{\{j\neq i_\ell l;\ell=1,\cdots,q\}}\alpha_j$*

Cette proposition s'obtient en considérant la propriété précédente et en appliquant la propriété d'additivité de la loi Gamma.

Propriété 1 – *Si $Y_l \sim \mathscr{G}(\alpha_l,1)$, alors pour tous nombres entiers $i < j$, $\sum_{\ell=i}^{j} Y_\ell \sim \mathscr{G}(\sum_{\ell=i}^{j}\alpha_l,1))$.*

A partir de ces résultats, on peut montrer que la loi marginale de chaque X_j est une loi Bêta de paramètres $(\alpha_j, \alpha - \alpha_j)$ avec $\alpha = \sum_{j=1}^{m} \alpha_j$. On en déduit l'espérance mathématique de chacune des composantes du vecteur aléatoire $(X_1, \cdots, X_m)$: $E(X_j) = \alpha_j/\alpha \ , \quad j = 1, \cdots, m$. Ces résultats se transposent au cas du processus de Dirichlet et permettent de construire un estimateur de Bayes de la fonction de répartition. Considérons $(X_1, \cdots, X_n)$ un n-échantillon. On note $(X_{(1)}, \cdots, X_{(m)})$, les m observations ordonnées et distinctes. Considérons la partition de $\mathbb{R}$ formée des intervalles dont les bornes sont les $X_{(j)}$, $j = 1, \cdots, m$ avec $X_{(0)} = -\infty$ et $X_{(m+1)} = +\infty$. On pose $I_j = [X_{(j-1)}, X_{(j)}[$, $j = 1, \cdots, m+1$. Considérons maintenant que l'observation consiste en un vecteur $(\varepsilon_X(I_1), \cdots, \varepsilon_X(I_m))$ où $\varepsilon_X(I_j) = \sum_{i=1}^{n} \mathbb{1}_{(X_i \in I_j)}$.
Ce vecteur suit une loi multinomiale de paramètre $(P(I_1), \cdots, P(I_m); P(I_{m+1}))$:

$$f(X \mid P) \propto \prod_{j=1}^{m+1} P(I_j)^{\varepsilon_X(I_j)}.$$

ε_X est donc un processus multinomiale de mesure-paramètre P. Une loi a priori naturelle (loi conjuguée) pour une loi multinomiale est la loi de Dirichlet. On considèrera donc comme loi a priori sur le processus P, un processus de Dirichlet de mesure-paramètre α :

$$\pi(P) = \prod_{j=1}^{m+1} P(I_j)^{\alpha(I_j)-1}.$$

Et la loi a posteriori sera :

$$\pi(P_{/X}) \propto f(X \mid P)\pi(P) = \prod_{j=1}^{m+1} P(I_j)^{(\varepsilon+\alpha)(I_j)-1}.$$

La loi a posteriori de P est donc une loi de Dirichlet ou $P_{/X}$ est un processus de Dirichlet de paramètre-mesure $\varepsilon + \alpha$.
On peut donc énoncer le théorème suivant :

Théorème 1 -

$$P \in \mathscr{D}(\alpha) \Rightarrow P_{/X} \in \mathscr{D}(\alpha + \varepsilon_X)$$

Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes $\tilde{P}_n$ de P est l'espérance a posteriori.

$$\tilde{P}_n(I_j) = \mathbb{E}[P_{/X}(I_j)] = \int P(I_j).\pi(P_{/X}(I_j))dP(I_j) \ , \quad \text{pour tout } I_j.$$

D'après les propriétés de la loi de Dirichlet :
Si $P_{/X} \in \mathscr{D}(\alpha + \varepsilon_X)$ alors

$$E[P_{/X}(I_i)] = \frac{(\alpha + \varepsilon_X)(I_i)}{(\alpha + \varepsilon_X)(\mathbb{R}^+)}$$

.

On peut alors définir l'estimateur de Bayes de P par :

$$\tilde{P}_n(I_i) = \gamma_n \frac{\alpha(I_i)}{\alpha(\mathbb{R}^+)} + (1 - \gamma_n)\hat{P}_n(I_i) \qquad \text{avec} \qquad \gamma_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R}) + n}$$

Ainsi,

$$\tilde{F}_n(x) = \tilde{P}_n([x, +\infty[) = \gamma_n \frac{\alpha(]_\infty, x])}{\alpha(\mathbb{R})} + (1 - \gamma_n)\, F_n(x)$$

puisque : $\dfrac{1}{n}\varepsilon_X(] - \infty, x]) = F_n(x)$.

Remarquons que γ_n se comporte comme le poids de la mesure a priori dans l'estimation. En effet, si $\alpha(\mathbb{R}^+)$ tend vers 0, γ_n tend vers 0 et l'estimateur de Bayes est la fonction de répartition empirique. L'observation est donc privilégiée et la mesure a priori α peut être considérée non informative.

$\diamond$