THÉORÈME DE BAYES (B4, G3)
(27 / 10 / 2020, © Monfort, Dicostat2005, 2005-2020)

Le **théorème de BAYES**, parfois appelé **théorème de la probabilité des causes**, constitue, malgré sa simplicité, l'un des grands résultats du **calcul des probabilités** et de la **Statistique**.

Il se fonde sur le **théorème des probabilités composées** (cf aussi **probabilité conditionnelle**).

La mise en oeuvre du théorème de BAYES constitue le **principe de la probabilité inverse de T. BAYES**, ou **méthode de la probabilité inverse de T. BAYES**. Ce résultat est à l'origine de la **théorie bayésienne**, développée par l'**école bayésienne**.

(i) Le problème initialement posé, appelé **problème de T. BAYES**, est le suivant. On considère un **espace probabilisé** $(\Omega, \mathcal{T}, P)$ et une épreuve aléatoire (cf **expérience aléatoire**) caractérisée par la réalisation de deux **événement** $A \in \mathcal{T}$ et $B \in \mathcal{T}$. On suppose que :

(a) A peut se réaliser à travers l'un des événements (« causes », ou hypothèses) $A_n \in \mathcal{T}$ ($n \in \mathbf{N}$) qui sont supposés incompatibles (ou mutuellement « exclusifs »), ie tq $A_\alpha \cap A_\beta = \varnothing$ si $\beta \neq \alpha$ (cf **incompatibilité**), et tq l'un d'entre eux se produit nécessairement, ie tq $\cup_{n \in \mathbf{N}} A_n = \Sigma_{n \in \mathbf{N}} A_n = \Omega$ (cf **système de constituants**) ;

(b) aucun événement n'est P-négligeable (cf **évènement presque impossible**), ie :

(1) $P(A_n) \neq 0, \forall n \in \mathbf{N}$, et $P(B) \neq 0$.

On établit alors les relations élémentaires :

(2) $B = B \cap \Omega = B \cap (\Sigma_{n \in \mathbf{N}} A_n) = \Sigma_{n \in \mathbf{N}} (B \cap A_n)$,

et le théorème des probabilités composées conduit à :

(3) $P(B) = P\{\Sigma_{n \in \mathbf{N}} (B \cap A_n)\} = \Sigma_{n \in \mathbf{N}} P(B \cap A_n) = \Sigma_{n \in \mathbf{N}} P(B / A_n) P(A_n)$.

Le problème de T. BAYES consiste à calculer la probabilité qu'un événement déterminé A_α ait entraîné l'événement (ou l'**effet**) B sachant que B est réalisé. Il s'agit donc de calculer la probabilité conditionnelle de A_α sachant B. Cette probabilité est appelée **probabilité a posteriori**.

La conclusion de ce théorème s'écrit alors ;

(4) $P(A_\alpha / B) = P(B \cap A_\alpha) / P(B) = R_\alpha / R$,

avec :

$$R_\alpha = P(B / A_\alpha) . P(A_\alpha),$$

et :

$$R = \Sigma_{n \in \mathbf{N}} P(B / A_n) . P(A_n).$$

D'un point de vue terminologique, on dit que :

(a) $P(A_n)$ est la **probabilité a priori** (de réalisation) de la cause A_n (cf **loi a priori**) ;

(b) $P(B / A_n)$ est la **probabilité conditionnelle** que (la réalisation de) la cause A_n soit à l'origine (de la réalisation) du résultat B ;

(c) $P(A_n / B)$ est la **probabilité a posteriori** (de réalisation) de la cause A_n après avoir « observé » le résultat B (cf **loi a posteriori**).

(ii) Le **théorème de T. BAYES** s'exprime aussi dans le cadre des **lois de probabilité conditionnelles**. On considère un espace probabilisé $(\Omega, \mathcal{T}, P)$ et un **couple aléatoire** $(\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times \mathbf{R}^G$ de loi $P^{(\xi, \eta)}$. On note μ (resp ν) une **mesure positive** σ-finie définie sur $\mathcal{B}(\mathbf{R}^K)$ (resp sur $\mathcal{B}(\mathbf{R}^G)$) (cf **mesure σ-finie**), $P(. / \xi)$ la loi de probabilité conditionnelle de η sachant ξ, f_2 la **densité marginale** de la loi $P^{(\xi, \eta)}$ relative à η, ie la **loi marginale** (ou loi propre) P^η de η pr à ν.

Par suite, si $P^{(\xi, \eta)}$ admet une densité par rapport à la **mesure produit** $\mu \otimes \nu$:

$$(5) \qquad dP^{(\xi, \eta)} / d(\mu \otimes \nu)(x, y) = f(x, y), \qquad \forall (x, y) \in \mathbf{R}^K \times \mathbf{R}^G,$$

on établit la **formule de T. BAYES généralisée** suivante :

$$(6) \qquad P(dy / \xi = x) / d\nu(y) = f(x, y) \; / \int \mathbf{1}(\mathbf{R}^G) f(x, y) \, d\nu(y), \qquad \mu\text{-p.p.}, \forall x \in \mathbf{R}^K,$$

qui exprime donc la **densité conditionnelle** $f(y / x) = P(dy / \xi = x) / d\nu(y)$ de η sachant ξ (cf **loi conditionnelle**).

En particulier, la **densité a posteriori** $f(y / x)$ de η sachant ξ s'exprime en fonction de la densité conditionnelle $f(x / y)$ de ξ sachant η et de la densité marginale f_2 de η selon :

$$(4)' \qquad f(y / x) = f(x / y) . f_2(y) \; / \int \mathbf{1}(\mathbf{R}^G) f(x / y) f_2(y) \, d\nu(y), \quad \mu\text{-p.p.}, \qquad \forall x \in \mathbf{R}^K.$$

La formule (4)' résume le **théorème de T. BAYES élémentaire**.

Par symétrie de rôle, il existe une formule analogue liant f (x / y) à f (y / x) et f_1 (densité marginale de $P^{(\xi, \eta)}$ pr à ξ).

La formule (6) permet de calculer les fr conditionnelles de η sachant ξ (resp de ξ sachant η).

On dit que :

(a) la loi conditionnelle P (. / η) (resp la densité conditionnelle f (x / y)) est la **loi a priori**, ou **distribution a priori** (resp la **densité a priori**) ;

(b) la loi conditionnelle P (. / ξ) (resp la densité f (y / x)) est la **loi a posteriori**, ou la **distribution a posteriori** (resp la **densité a posteriori**).

Une **caractéristique** quelconque de ces lois (**moment**, **mode**, **fonction de répartition**, etc) est qualifiée de la même façon.

Les formes élémentaires du théorème de BAYES mettent souvent en jeu les **mesures de LEBESGUE** $\mu = \lambda_K$ et $\mu = \lambda_G$ (cas « continu »), ou encore des **mesures discrètes** (ou des **mesures de comptage**) définies sur $\mathscr{B}(\mathbf{R}^K)$ et $\mathscr{B}(\mathbf{R}^G)$ (cas « discret »).

(iii) Le **théorème de BAYES** est une partie intégrante de la **théorie de la décision** statistique (cf **décision statistique**), car il permet notamment de passer de la (mesure de) probabilité a priori à la (mesure de) probabilité a posteriori portant sur le « **paramètre** » du modèle.

En effet, soit $(\mathscr{X}, \mathscr{B}, (P_\theta^X)_{\theta \in \Theta})$ un **modèle image** sous forme paramétrée, dans lequel P_θ^X est considérée comme **probabilité de transition**. Θ est supposé muni d'une tribu $\mathscr{B}_\Theta$ sur laquelle une **mesure de probabilité** π est donnée a priori : cette mesure est souvent appelée **mesure a priori**.

On pose alors :

(7) $\quad dP_\theta^X / d\mu = f(\,.\,, \theta), \qquad \forall\, \theta \in \Theta\ (\mu\text{-p.p.}),$

ainsi que :

(8) $\quad d\pi / d\nu = p \qquad (\nu\text{-p.p.}),$

où μ et ν sont des mesures positives σ-finies données resp sur $\mathscr{B}$ et sur $\mathscr{B}_\Theta$.

Alors (**théorème de T. BAYES de la théorie de la décision**), la (mesure de) probabilité a posteriori Q_x de la va θ (considérée comme **identité** dans Θ) (cf **application identique**) est définie par la densité suivante :

(9) $\quad q(x, \theta) = f(x, \theta)\, p(\theta) \,/\, \int_\Theta f(x, \theta)\, d\pi(\theta), \qquad (\mu \otimes \nu\text{-p.p.}).$

En pratique, on note souvent f (x / θ), ou f_θ (.), ou encore f_θ , au lieu de f (x, θ) ; par suite, q (θ, x) est souvent notée q (θ / x), ou q_x (.), ou encore q_x .

(iv) Le **théorème de BAYES** (formule (9)) exprime (H. JEFFREYS) que la probabilité a posteriori d'une cause, ou d'une hypothèse, est proportionnelle au produit de sa probabilité a priori par sa **fonction de vraisemblance**.

On appelle **transformation de T. BAYES**, ou **opération de T. BAYES**, l'application f (x / y) $\mapsto$ f (y / x) (resp f (x / θ) $\mapsto$ q (θ / x)) qui transforme la probabilité a priori en probabilité a posteriori.

(v) Deux **contextes statistiques** élémentaires se rencontrent souvent :

(a) **simulation**. Dans cette situation, le modèle statistique consiste à simuler un **échantillon** à partir d'une **variable parente** dont la **loi** est connue. Comme θ est donné, on le considère comme une **va** π-p.s. dégénérée, dont la loi est une **loi de DIRAC** placée en $\theta \in \Theta$ (cf **variable dégénérée**). La loi a posteriori est alors cette même masse δ_θ ;

(b) **principe d'égale ignorance**. La logique de ce principe veut que la loi de θ est inconnue. « Faute de mieux », on suppose que θ suit une **loi uniforme**, dans la mesure où une telle loi peut être définie sur la tribu $\mathscr{B}_\Theta$.

(vi) Dans le cadre bayésien défini par la formule (9), on considère souvent une **famille** $\mathscr{P}^X$ de probabilités P_θ^X , indexées par θ, définies sur $\mathscr{B}$ (eg la famille des **lois de GAUSS-LAPLACE**, celle des **lois multinômiale**, celle des **lois gamma**, etc) et une famille $\mathscr{Q}$ de probabilités π définies sur $\mathscr{B}_\Theta$.

Ces deux familles étant données, si l'application :

(10) $\quad b : \mathscr{P}^X \times \mathscr{Q} \mapsto \mathscr{Q}$.

associe à tout couple (P_θ^X, π) une probabilité Q_x définie par (9) et tq $Q_x \in \mathscr{Q}$ (pour μ-presque tout x $\in \mathscr{X}$), on dit que la famille $\mathscr{Q}$ est « fermée » pr à la famille $\mathscr{P}^X$ pour la transformation b. On dit aussi que $\mathscr{P}^X$ et $\mathscr{Q}$ sont des **familles conjuguées** (**de lois de probabilité**) pr à b (cf **lois conjuguées**).

(vii) La mesure a priori Π est souvent interprétée comme définissant un certain **degré de croyance**, ou **degré de vraisemblance**, attaché aux différentes valeurs $\theta \in \Theta$ (cf aussi **fonction d'utilité**, **utilité**) ; elle est aussi considérée comme traduisant l'**information a priori** dont dispose le **statisticien** avant de résoudre un problème de décision statistique. Dans cette dernière acception, la densité p = dΠ / dν est parfois estimée à partir d'un échantillon θ_1 ,..., θ_M résultant d'un **échantillonnage** ou d'une **expérience aléatoire** (etc) analogues à celle en cours, mais effectué(e)s antérieurement (cf aussi **règle de BAYES empirique**).

(viii) Enfin, et de manière générale, lorsqu'une **procédure statistique** (**test**, **estimation**, **prévision**), notamment une procédure bayésienne, est mise en oeuvre, se pose souvent la question d'apprécier en quoi les caractéristiques ou les « qualités » de cette procédure sont modifiées (ou « altérées ») lorsque la « vraie » loi a priori est différente de celle effectivement retenue.

Cette approche constitue le problème de la **robustesse** d'une procédure pr aux hypothèses relatives à la loi a priori.

(ix) Un exemple classique d'application du théorème de BAYES est celui du **modèle de régression multiple** linéaire $y = X\,b + u$, avec $u \sim \mathcal{N}_N(0\ ,\ \Sigma)$ (où Σ est supposée connue) et $b \sim \mathcal{N}_K(b_0\ ,\ V)$ (loi a priori dont la moyenne b_0 et la variance V sont aussi supposées connues), la loi a posteriori de b conditionnellement à (X, y) est la **loi normale multidimensionnelle** $\mathcal{N}_K$ (m, M), dans laquelle :

$$(11)\quad \begin{aligned} m &= b_0 + V\,X'\,(\Sigma + X\,V\,X')^{-1}\,(y - X\,b_0) \\ M &= v = V\,X'\,(\Sigma + X\,V\,X')^{-1}\,X\,V. \end{aligned}$$