REVUE DE STATISTIQUE APPLIQUÉE

PH. L'HARDY

Application des méthodes bayesiennes à l'estimation d'élasticités de consommation

Revue de statistique appliquée, tome 14, nº 4 (1966), p. 33-44

<http://www.numdam.org/item?id=RSA_1966__14_4_33_0>

APPLICATION DES MÉTHODES BAYESIENNES A L'ESTIMATION D'ÉLASTICITÉS DE CONSOMMATION

Ph. L'HARDY
Institut National de la Statistique et des Etudes Economiques

Pour résoudre le problème de l'estimation d'un paramètre à partir des résultats d'une observation, l'idée la plus naturelle est d'utiliser le théorème de Bayes (Probabilité des Causes). Mais on voit immédiatement qu'on se heurte à une grave difficulté : Comment choisir la loi de probabilité *a priori* indispensable à cette méthode ? On sait que les méthodes classiques d'estimation (méthode du maximum de vraisemblance) évitent cette difficulté.

Cependant les méthodes Bayesiennes restent séduisantes ; elles suscitent actuellement de nombreuses études en liaison avec la Théorie de la Décision. Des suggestions ont été faites par différents auteurs, en particulier par Raiffa et Schlaifer dans leur "Applied Statistical Decision Theory", pour faciliter le choix de la loi de probabilité *a priori*.

Il nous semble toutefois que les exposés sur les méthodes bayesiennes restent souvent très théoriques. Aussi avons nous voulu tenter une application numérique dans un domaine où la comparaison avec les méthodes classiques soit possible : nous avons choisi pour cela l'estimation d'une élasticité de consommation à partir de séries temporelles.

Après avoir exposé les principes des méthodes d'estimation bayesiennes et les suggestions de Raiffa et Schlaifer, nous montrerons que dans cette méthode, on peut interpréter concrètement l'influence de la loi *a priori* sur le calcul de l'estimation. Traitant alors le problème pratique d'estimer une élasticité, nous caractériserons numériquement la loi de probabilité *a priori* et nous commenterons les résultats numériques obtenus.

PRINCIPE DE LA METHODE DE BAYES

Le résultat d'une expérience ou d'une observation est représenté par la variable x[1]. On suppose qu'il existe un modèle aléatoire dépendant d'un paramètre inconnu θ(1) qui représente correctement les résultats de l'expérience.

Autrement dit, le résultat x de l'expérience apparaît comme une variable aléatoire X dont la loi dépend de θ, et pour chaque valeur de θ les probabilités $P(X = x/\theta)$ sont connues[2].

----

(1) scalaire ou vecteur

(2) $P(X = x/\theta)$ est la vraisemblance de l'observation x

Le problème est alors d'estimer θ quand on connaît le résultat x_0 d'une expérience (c'est-à-dire une réalisation x_0 de la variable aléatoire X). La méthode du maximum de vraisemblance consiste, on le sait, à choisir pour estimateur de θ la valeur $\hat{\theta}$ qui rend $P(X = x_0/\theta)$ maximum.

Dans la méthode de Bayes, on suppose que le paramètre est une variable aléatoire Θ qui peut prendre différentes valeurs θ. On se donne la loi de probabilité *a priori* de Θ. On peut alors calculer les probabilités *a posteriori*(1) de Θ connaissant le résultat de l'observation, en utilisant la formule de Bayes :

$$P(\Theta = \theta / X = x_0) = P(\Theta = \theta). \frac{P(X = x_0/\theta)}{P(X = x_0)}$$

ou encore en utilisant le signe $\sim$ pour représenter la proportionnalité :

$$\boxed{\underbrace{P(\Theta = \theta / X = x_0)}_{\text{Pr. a posteriori}} \sim \underbrace{P(\Theta = \theta)}_{\text{Pr. a priori}} . \underbrace{P(X = x_0/\Theta = \theta)}_{\text{Vraisemblance}}}$$

EXEMPLE. ESTIMATION D'UNE ELASTICITE

L'exemple numérique choisi a été l'estimation de l'élasticité de la consommation alimentaire par tête à prix constant, par rapport à la consommation totale par tête à prix constant, à partir des données annuelles des Comptes des Ménages.

Le modèle utilisé est de la forme(2)

$$y = \alpha x + \varepsilon$$

où y est l'accroissement par rapport à l'année de base (1949) de la consommation alimentaire CA_t

$$y_t = \frac{CA_t - CA_{1949}}{CA_{1949}}$$

et x celui de la consommation totale C_t ,

$$x_t = \frac{C_t - C_{1949}}{C_{1949}}$$

L'écart aléatoire ε est supposé être une variable normale de moyenne nulle et d'écart type $\sigma_\varepsilon^2 = 1/h$.

----

(1) "L'estimation" de θ n'est plus alors un nombre, mais la distribution de probabilité *a posteriori* de Θ. On peut si on veut prendre une valeur centrale de cette distribution comme estimation de θ.

(2) En fait α ne correspond pas strictement à la définition de l'élasticité, mais en est une approximation.

Les paramètres à estimer (équivalents à θ dans le cas général) sont ici : $\underline{\alpha}$ et $\underline{h}$.

Les résultats de l'observation sont les quantités $\underline{y}_t$ (les $\underline{x}_t$ ne sont pas aléatoires).

FORME DE LA VRAISEMBLANCE

La vraisemblance des y_t est à un changement de variable près celle des ε_t :

$$V = V(\varepsilon_1 \ldots \varepsilon_N) \sim e^{-\frac{h}{2}\Sigma(\varepsilon_i)^2} \cdot h^{\frac{N}{2}}$$

Posons :

$$S_{xy} = \sum x_i \; y_i$$

$$S_x = \sum (x_i)^2$$

$$S_y = \sum (y_i)^2$$

$$e_i = y_i - a x_i$$

$$a = \frac{S_{xy}}{S_x}$$

$$S_e = \sum e_i^2 = S_y - a^2 S_x$$

$\nu = N - 1$ = degrés de liberté du vecteur $(e_i)_{i=1,n}$

On a alors :

$$\boxed{V = V(S_y, S_{xy}, S_x, N | \alpha, h) \sim e^{-\frac{h}{2}h(S_y - 2\alpha S_{xy} + \alpha^2 S_x)} \cdot h^{\frac{N}{2}}}$$

avec : paramètre à estimer :

$$\theta \begin{cases} \alpha \\ h \end{cases}$$

Variables provenant de l'échantillon :

aléatoires : $\begin{cases} S_y \\ S_{xy} \end{cases}$

non aléatoires : $\begin{cases} S_x \\ N \end{cases}$

CHOIX DE LA PROBABILITE A PRIORI

Désignons par P' les probabilités *a priori* et par P'' les probabilités *a posteriori*.

Avec les notations théoriques, la formule de Bayes s'écrirait :

$$P''(\theta) \sim P'(\theta) \;.\; V(x/\theta)$$

Dans notre exemple, cette formule devient :

$$\boxed{P''(\alpha, h) \sim P'(\alpha, h) \;.\; V(S_y, S_{xy}, S_x, N|\alpha, h)}$$

Le choix des probabilités *a priori* est le point délicat de la méthode de Bayes. Ces probabilités doivent représenter ce que nous avons *a priori*, c'est-à-dire avant l'observation, sur les paramètres à estimer. Le choix étant difficile, on peut commencer par s'imposer la forme de la fonction $P'(\alpha, h)$.

En adoptant la méthode suggérée par Raiffa et Schlaffer, nous choisirons une probabilité *a priori* $P'(\alpha, h)$ de même forme que la vraisemblance, autrement dit nous poserons :

$$P'(\alpha, h) = V(S'_y, S'_{xy}, S'_x, N'|\alpha, h)$$

où S'_y, S'_{xy}, S'_x, N' sont des constantes que nous choisirons d'après nos idées *a priori* sur les distributions de α et h.

Le choix de la forme retenue pour $P'(\alpha, h)$ a des raisons de simplicité des calculs :

Il apparaît en effet immédiatement que :

$$\boxed{\begin{aligned} & P''(\alpha, h) = V(S''_y, S''_{xy}, S''_x, N''|\alpha, h) \\ & \quad \text{avec : } S_y = S'_y + S_y \\ & \qquad\qquad S''_{xy} = S'_{xy} + S_{xy} \\ & \qquad\qquad S''_x = S'_x + S_x \\ & \qquad\qquad N'' = N' + N \end{aligned}}$$

ESTIMATIONS RETENUES POUR a ET h

Nous connaissons la loi *a posteriori* de α et h :

$$P''(\alpha, h) \sim V(S''_y, S''_{xy}, S''_x, N''|\alpha, h).$$

il nous reste à choisir les estimateurs que nous retiendrons : ce seront respectivement les espérances mathématiques des lois marginales de α et h.

La probabilité $P''(\alpha, h)$ se décompose en effet de la façon suivante :

$$P'' \sim \underbrace{(e^{-\frac{1}{2}hS''_e} \cdot h^{\frac{N''-1}{2}})}_{\substack{\text{Loi de } h \text{ :} \\ \chi^2 \text{ à } N'' \text{ ddl}}} \cdot \underbrace{(e^{-\frac{h}{2}(a-\alpha)^2 S''_x} \cdot h^{\frac{1}{2}})}_{\substack{\text{Loi de } \alpha \text{ si } h \text{ :} \\ \text{Loi normale}}}$$

et nos estimations de h et α seront données par les formules suivantes :

Loi de h (1)

$$E(h) = \frac{N'' + 1}{S''_e}$$

$$\mathrm{Var}(h) = \frac{2(N'' + 1)}{(S''_e)^2}$$

Loi de α(1)

La loi marginale de α est une loi de Student de la forme :

$$\frac{\text{constante}}{[S''_e + (a - \alpha)^2 S''_x]^{\frac{N''}{2}+1}}$$

et on a :

$$E(\alpha) = a$$

$$\mathrm{Var}(\alpha) = \frac{S''_e}{S''_x} \frac{1}{N'' - 1}$$

Inversement, ces formules nous servirons pour calculer les caractéristiques S'_x, S'_{xy}, S'_y, N' de la loi *a priori* à partir des idées que nous avons *a priori* de E (h), Var (h), E (α) et Var (α).

CAS PARTICULIER : LOI A PRIORI CONSTANTE

On sait que Jeffreys a préconisé dans certains cas l'utilisation d'une loi *a priori* constante quand on n'a aucune information sur le paramètre :

$$P'(\theta) \sim 1^{(2)}$$

----

(1) Résultats différents de ceux de Raiffa et Schlaifer, qui prennent en fait une forme de loi à priori légèrement différente de la vraisemblance.

(2) Il s'agit en fait d'une loi "dégénérée", dans le cas (qui est le nôtre) où θ prend des valeurs infinies. En effet $\int P'(\theta)d\theta$ diverge alors. Ceci ne doit pas nous arrêter, car il nous suffit que l'opérateur qui nous donne $P''(\theta)$ en fonction de $V(x|\theta)$ soit bien défini, or il s'écrit :

$$P''(\theta) \sim \int P'(\theta) \cdot V(x|\theta)\, dx = E_x V(x|\theta). \text{ intégrale qui converge.}$$

Or il se trouve qu'un tel type de loi *a priori* appartient à la famille que nous avons retenue. En effet :

$$V(0, 0, 0, 0 | \alpha, h) = 1$$

Il suffit donc de prendre $S'_y = S'_{xy} = S'_x = N' = 0$.

On trouve alors :

$$S''_y = S_y$$

$$S''_{xy} = S_{xy}$$

$$S''_x = S_x$$

$$N'' = N$$

$$\boxed{\begin{aligned} E(\alpha) &= a \\ \mathrm{Var}(\alpha) &= \frac{S_e}{S_x} \frac{1}{N-1} \end{aligned}}$$

Ces résultats sont à rapprocher de ceux de la méthode classique (où a est aléatoire et α fixe mais inconnu).

$$\boxed{\begin{aligned} E(a) &= \alpha \\ \mathrm{Var}(a) &= \frac{S_e}{S_x} \frac{1}{N-1} \end{aligned}}$$

MELANGE DE DEUX ECHANTILLONS

Nous allons maintenant chercher à interpréter les formules précédentes en comparant aux résultats des méthodes classiques relatives au mélange de deux échantillons.

Le modèle est toujours : $y = ax + \varepsilon$.

Un échantillon est un ensemble de couples : $(y_i , x_i)_{i=1,N}$. Pour un tel échantillon, les méthodes classiques donnent :

$$\hat{a} = \frac{S_{xy}}{S_x} \qquad V(\hat{a}) = \frac{S_y - \frac{(S_{xy})^2}{S_x}}{S_x (N-1)}$$

Soient 2 échantillons $(y_i^I , x_i^I)_{i=1,N^I}$ et $(y_i^{II} , x_i^{II})_{i=1,N^{II}}$ et considérons l'échantillon réunion I + II formé de l'ensemble des couples de I de II .

Sur cette réunion on trouve les estimations suivantes :

$$\hat{a} = \frac{S_{xy}^I + S_{xy}^{II}}{S_x^I + S_x^{II}}$$

$$V(\hat{a}) = \frac{(S_y^I + S_y^{II}) - \frac{(S_{xy}^I + S_{xy}^{II})}{S_x^I + S_x^{II}}}{(S_x^I + S_x^{II})(N^I + N^{II} - 1)}$$

Autrement dit on trouve les caractéristiques S_y^{I+II}, S_{xy}^{I+II}, S_x^{I+II}, N^{I+II}, de l'échantillon réunion I + II en sommant celles de chaque échantillon composant.

$$\boxed{\begin{aligned} S_y^{I+II} &= S_y^I + S_y^{II} \\ S_{xy}^{I+II} &= S_{xy}^I + S_{xy}^{II} \\ S_x^{I+II} &= S_x^I + S_x^{II} \\ N^{I+II} &= N^I + N^{II} \end{aligned}}$$

Nous pouvons donc interpréter les formules bayesiennes de la façon suivante :

Tout se passe comme si :

1/ à l'échantillon réellement observé E (de caractéristiques S_y, S_{xy}, S_x, N) on ajoute un échantillon fictif E' (de caractéristiques S'_y, S'_{xy}, S'_x, N') qui représente l'influence des connaissances *a priori* de l'observateur sur les paramètres à estimer.

2/ On applique les méthodes classiques(1) à l'échantillon E + E'.

On voit donc le résultat de la méthode de Raiffa et Schlaifer consistant à prendre une probabilité a priori de même forme que la vraisemblance : Le passage d'un état de connaissance *a priori* à un état de connaissance *a posteriori* par suite de la prise en compte de l'échantillon observé E est assimilé à l'adjonction de l'échantillon E à l'échantillon initial fictif E' correspondant à l'état de connaissance *a priori*.

En particulier cette interprétation justifie le choix d'une loi *a priori* constante ($P'(\alpha, h) \sim 1$) quand on ne sait rien sur les paramètres : l'échantillon E' est alors vide et ne perturbe pas E.

----

(1) Ceci cependant ne doit pas masquer les différences de conception

Point de vue Bayesien : le paramètre α est une variable aléatoire de moyenne

$$a = \frac{S_{xy}}{S_x}$$

Point de vue classique : l'estimateur â, variable aléatoire dont la réalisation pour l'échantillon E vaut

$$a = \frac{S_{xy}}{S_x}$$

a une moyenne α, valeur fixe et inconnue du paramètre.

APPLICATION NUMERIQUE

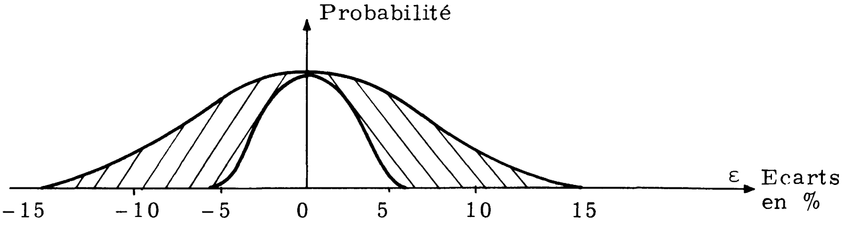
Il nous reste à montrer sur un exemple comment on peut choisir les caractéristiques S'_y, S'_{xy}, S'_x et N' de la loi *a priori* de façon à représenter l'état de connaissance *a priori* que l'on a des paramètres.

Dans le cas d'une élasticité de la consommation alimentaire par rapport à la consommation totale, nous supposerons avoir une idée de cette élasticité : par exemple qu'elle est comprise en gros entre 0,3 et 0,7.

Nous prendrons donc a' = 0,5 et $\sigma'_a = 0,1$.

Cherchons maintenant à fixer nos idées sur la variabilité des écarts :

On peut penser que la courbe de probabilité des écarts ε se trouve dans la région hachurée de la figure ci-dessous :



Les valeurs extrêmes des écarts-types sont donc :

$$\sigma_\varepsilon \# 7 \%$$

$$\sigma_\varepsilon \# 2 \%$$

qui correspondent aux valeurs suivantes de h :

$$h \# \frac{2}{100}$$

$$h \# \frac{25}{100}$$

Nous prendrons donc : $\overline{h}' = \dfrac{1}{10}$

$$\sigma'_h = \frac{0,5}{10}$$

A partir de ces estimations de a', σ'_a, $\overline{h}'$ et σ'_h, nous allons calculer S'_y, S'_{xy}, S'_x, N'.

On a d'abord :

$$a' = \frac{S'_{xy}}{S'_x} \quad \text{d'où} \quad S'_{xy} = 0,5\ S'_x$$

Ensuite : $(\sigma'_a)^2 \# \dfrac{S'_e}{N'} \dfrac{1}{S'_x} = \dfrac{1}{100}$

et $\bar{h}' \# \frac{N'}{S'_e} = \frac{1}{10}$ d'où $S'_x = 1000$ et $S'_{xy} = 500$.

Enfin : $\left(\frac{\sigma'_h}{\bar{h}'}\right)^2 \# \frac{2}{N'}$ d'où $N' = 8$ et $S'_e = 80$

et comme $S'_y = S'_e + a^2 S'_x$, il vient $S'_y = 80 + 250 = 330$.

Les tableaux suivants donnent les résultats des estimations dans cette hypothèse de loi *a priori* (mentionnée dans les tableaux sous le nom d'"hypothèse centrale").

On a également étudié d'autres types d'hypothèses *a priori*.

Types d'hypothèses *a priori* étudiées

Hypothèses	Estimation *a priori* des paramètres
Hypothèse d'ignorance *a priori*.	Echantillon E' vide
Hypothèse centrale	$a' = 0,5$ $\sigma'_a = 0,1$ $\bar{h}' = \frac{1}{10}$ $\frac{\sigma_h}{h} = \frac{1}{2}$
Variante sur a'	$a' = 0,7$ $\sigma'_a = 0,1$ $\bar{h}' = \frac{1}{10}$ $\frac{\sigma_h}{h} = \frac{1}{2}$
Variante sur a' et σ'_a	$a' = 0,7$ $\sigma'_a = 0,3$ $\bar{h}' = \frac{1}{10}$ $\frac{\sigma_h}{h} = \frac{1}{2}$
Variante sur $\bar{h}$	$a' = 0,5$ $\sigma'_a = 0,1$ $\bar{h}' = \frac{1}{100}$ $\frac{\sigma_h}{h} = \frac{1}{2}$
Variante n° 1 sur σ'_h	$a' = 0,5$ $\sigma'_a = 0,1$ $\bar{h}' = \frac{1}{10}$ $\frac{\sigma_h}{h} = 1$
Variante n° 2 sur σ'_h	$a' = 0,5$ $\sigma'_a = 0,1$ $\bar{h}' = \frac{1}{10}$ $\frac{\sigma_h}{h} = \frac{1}{10}$

CONCLUSIONS

Les méthodes de probabilités subjectives paraissent souvent choquantes parce qu'il est nécessaire d'estimer *a priori* les paramètres. Certes ces estimations doivent représenter nos idées *a priori* sur les paramètres. Mais ces idées sont souvent très floues, et les estimations qui en résultent paraissent à juste titre assez arbitraires. Dès lors pour beaucoup une méthode s'appuyant sur des données aussi suspectes paraît peu sérieuse.

Il faut bien voir cependant que dans la méthode de Bayes on introduit les estimations *a priori* des paramètres avec les limites de précision qu'on leur attribue, Si on a peu de confiance, par exemple, dans l'estimation qu'on fait de a', il suffit de prendre un σ'_a assez grand de façon à être certain que l'intervalle $a' \pm \sigma'_a$ englobe la valeur vraie du paramètre.

Caractéristiques des échantillons E' et E'' = E' + E selon les différentes hypothèses *a priori*.

Hypothèses	S'_x S'_y S'_{xy} N'	S''_x S''_y S''_{xy} N''
Ignorance *a priori* sur les paramètres	0 0 0 0	6 400 (1) 1 598 (1) 3 136 (1) 15
Hypothèse centrale	1 000 330 500 8	7 400 1 928 3 636 23
Variante sur a'	1 000 570 700 8	7 400 2 168 3 836 23
Variante sur a' et σ'_a	111 134 78 8	6 511 1 732 3 214 23
Variante sur $\overline{h}'$	10 000 3 300 5 000 8	16 400 4 898 8 136 23
Variante n° 1 sur σ'_h	1 000 270 500 2	7 400 1 868 3 636 17
Variante n° 2 sur σ'	1 000 2 250 500 200	7 400 3 848 3 636 215

(1) ce sont les caractéristiques de l'échantillon E.

On peut même être tout à fait prudent, et prendre l'hypothèse de l'ignorance totale(1). Mais une hypothèse moins négative est presque toujours possible, et elle permet d'améliorer la précision des estimations

----

(1) On obtient alors les mêmes résultats qu'avec les méthodes classiques.

Résultats

Hypothèses	a''	σ''_a	h''	σ_ε	σ''_h
Ignorance *a priori* sur les paramètres	0,490 (1)	0,025 (1)	0,2581	1,968 (1)	0,091
Hypothèse centrale	0,491	0,01227	0,0980	3,194	0,028
Variante sur a'	0,518	0,01066	0,1297	2,777	0,037
Variante sur σ'_a et a'	0,494	0,01000	0,1678	2,441	0,048
Variante sur $\bar{h}'$	0,496	0,01548	0,0278	5,998	0,008
Variante n° 1 sur σ'_h	0,491	0,00844	0,2118	2,173	0,071
Variante n° 2 sur σ_h	0,491	0,03611	0,1046	3,092	0,010

(1) ce sont également les résultats que donnent les méthodes classiques.

finales : par exemple quand on passe de l'hypothèse d'"ignorance totale" à celle baptisée "Hypothèse centrale", σ''_a passe de 0,025 à 0,012[1].

Par contre, ce qui est interdit, c'est d'introduire des données *a priori* invraisemblables, c'est-à-dire de choisir α' et σ'_a de sorte que l'intervalle $\alpha \pm \sigma'_a$ ne contienne pas la valeur vraie du paramètre.

Comparons par exemple les hypothèses "variante sur a'" et "variante sur α' et σ'_a". On sait que le résultat est de l'ordre de 0,49 avec un écart-type σ''_a = 0,03 (Résultat de l'hypothèse "ignorance totale"). Or dans l'hypothèse "variante sur a'" (a' = 0,7 σ'_a = 0,1) les données *a priori* sont peu vraisemblables puisque l'intervalle 0,7 ± 0,1 ne contient pas la valeur vraie, alors que dans l'hypothèse "variante sur a' et σ'_a" (a' = 0,7 σ'_a = 0,3) les données *a priori* sont vraisemblables (l'intervalle 0,7 ± 0,3 contenant la valeur vraie). Aussi dans le premier cas ("variante sur a'") obtient-on des résultats peu compatibles avec ceux de l'hypothèse "ignorance totale", dans le second cas, on trouve des résultats qui améliorent ceux de l'hypothèse "ignorance totale".

On voit donc que les méthodes subjectivistes sont conformes à l'éthique scientifique puisqu'il est permis d'affirmer à condition d'être sûr de ce qu'on avance et qu'il n'est pas interdit de reconnaître son ignorance.

----

(1) Ceci peut paraître contradictoire avec les résultats des hypothèses "variante sur a' et σ'" (hypothèse *a priori* peu précise : a' = 0,7 σ'_a = 0,3, résultat σ''_a = 0,010) et "hypothèse centrale" (hypothèse *a priori* plus précise : a' = 0,5 σ'_a = 0,1, résultat σ''_a = 0,012). Cela tient à ce que a' influe sur σ''_a.

BIBLIOGRAPHIE

JEFFREYS. Theory of Probabilities. Oxford, Clarendon Press.

RAIFFA et SCHLAIFER. Applied Statistical Decision theory. Harvard Univ. Press.