



HAL
open science

Utilisation de familles de mesures de possibilité comme modèle d'information a priori subjective dans l'estimation paramétrique

G. Banon

► **To cite this version:**

G. Banon. Utilisation de familles de mesures de possibilité comme modèle d'information a priori subjective dans l'estimation paramétrique. Bulletin pour les sous-ensembles flous et leurs applications, 1979, 1, pp.36–46. hal-04567376

HAL Id: hal-04567376

<https://hal.science/hal-04567376>

Submitted on 3 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

UTILISATION DE FAMILLES DE MESURES DE POSSIBILITE
COMME MODELES D'INFORMATION A PRIORI SUBJECTIVE
DANS L'ESTIMATION PARAMETRIQUE

G. BANON

L.A.A.S.

Octobre 1979

L.A.A.S. : Laboratoire d'Automatique et d'Analyse des Systèmes
7 avenue du Colonel Roche
31400 Toulouse France

1) INTRODUCTION

Dans ce travail on considère le problème de la modélisation de la connaissance a priori en vue de son utilisation dans les algorithmes d'estimation.

Une des motivations de cette étude est le problème de l'initialisation du Filtre de Kalman utilisé comme technique d'identification. Ce problème sera considéré comme une toile de fond et on se limitera ici à la situation la plus simple correspondant à l'estimation d'une moyenne.

Le problème de l'estimation de la moyenne est très classique et on peut classer les estimateurs connus en deux familles selon qu'ils dépendent ou non d'une information a priori.

Parmi ceux qui sont indépendants d'une connaissance a priori, on peut rappeler la moyenne empirique (parfois appelée estimateur ergodique) et l'estimateur du maximum de vraisemblance non bayésien. Au contraire, l'estimateur du maximum de vraisemblance bayésien ou celui des moindres carrés (parfois appelé estimateur de la variance minimum) dépendent d'une information a priori.

Pour ces deux derniers estimateurs la connaissance a priori est de type probabiliste. C'est-à-dire que l'on suppose que la moyenne à estimer est la valeur prise par une variable aléatoire dont on connaît la loi de probabilité.

A partir d'une analyse critique des modèles de connaissance a priori probabiliste qui conduit à une remise en cause de ces modèles, on montre qu'une famille particulière de modèles non probabilistes convient mieux pour l'estimation de la moyenne lorsque le nombre d'observations est fixé.

A ce niveau, les mesures de possibilité et plus généralement celles de probabilité supérieure (parfois appelées de plausibilité) apparaissent assez naturelles comme modèle d'information a priori, la règle de Dempster généralisant alors celle de Bayes.

Finalement, un des intérêts d'une telle étude doit être d'améliorer les algorithmes d'estimation en augmentant leurs performances pour le cas où l'on dispose seulement d'un petit nombre d'observations (ou d'échantillons) sans pour autant modifier leur comportement asymptotique pour le cas où ce nombre d'observations pourrait devenir infini.

2) ENONCE DU PROBLEME D'ESTIMATION ET RAPPEL DES METHODES CLASSIQUES

Soit (Ω, \mathcal{F}, P) un espace de probabilité et V une variable aléatoire réelle (v.a.r.) définie sur (Ω, \mathcal{F}, P) de loi de probabilité connue, avec en particulier $EV = 0$.

Soit 1_A la v.a.r. définie sur (Ω, \mathcal{F}) par

$$\omega \in \Omega \longmapsto 1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

pour $A \in \mathcal{F}$.

Pour $x \in \mathbb{R}$, on considère la suite des v.a.r. $\{Y_i\}_{i=1}^n$ définie par

$$Y_i = x 1_{\Omega} + V_i \quad i = 1, \dots, n$$

où $\{V_i\}_{i=1}^n$ est une suite de v.a.r. indépendantes et de même loi que V .

Le problème posé est celui de l'estimation de x à partir de $\{Y_i(\omega)\}_{i=1}^n$ en présence ou non d'une information a priori sur x .

Ce problème correspond en pratique au cas où un paramètre x inconnu n'est pas directement accessible à l'expérience et qu'il doit donc être évalué à partir d'un certain nombre d'observations (ou mesures), l'expérimentateur pouvant avoir une connaissance a priori sur le paramètre qu'il cherche à observer.

i) La moyenne empirique

Puisque $x = E Y_i \quad i = 1, \dots, n$, une solution naturelle ne faisant intervenir aucune information a priori est la moyenne empirique

$$\frac{1}{n} \sum_{i=1}^n Y_i, \quad (1)$$

que l'on notera \overline{Y}_n . ■

Supposons que la v.a.r. V possède une densité P_V , alors les v.a.r. Y_1, \dots, Y_n possèdent des densités P_{Y_1}, \dots, P_{Y_n} qui ici sont données par

$$y \in \mathbb{R} \longmapsto P_{Y_i}(y) = P_V(y-x) \quad i=1, \dots, n.$$

A cause de l'indépendance des V_i , la densité conjointe P_{Y_1, \dots, Y_n} des v.a.r. Y_1, \dots, Y_n est donnée par

$$(y_1, \dots, y_n) \in \mathbb{R}^n \longmapsto P_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \prod_{i=1}^n P_V(y_i - x)$$

Soit \mathcal{L} la fonction de vraisemblance $\mathbb{R}^{n+1} \rightarrow \mathbb{R}$ définie par

$$(u, y_1, \dots, y_n) \in \mathbb{R}^{n+1} \longmapsto \mathcal{L}(u, y_1, \dots, y_n) = \prod_{i=1}^n P_V(y_i - u)$$

Le modèle classique pour l'information a priori est de type probabiliste. Dans ce cas on suppose que x est la valeur prise par une v.a.r. X de loi connue. La v.a.r. X est supposée de plus indépendante des V_i . Plus précisément on peut introduire un second espace de probabilité $(\Omega', \mathcal{B}', P')$ et dire que X est une v.a.r. définie sur (Ω', \mathcal{B}') . On considère alors la mesure produit $P \otimes P'$ définie sur $\mathcal{B} \otimes \mathcal{B}'$. Enfin on suppose que X possède une densité π .

iv) Estimateur des moindres carrés

Cet estimateur est obtenu à partir de la solution mesurable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ du problème de la minimisation de la quantité :

$$E (f(Y_1, \dots, Y_n) - X)^2$$

où E se calcule relativement à la probabilité produit $P \otimes P'$.

La solution à ce problème fournit l'estimateur des moindres carrés qui n'est autre que l'espérance mathématique conditionnelle de X connaissant les Y_i et qui s'écrit explicitement

$$\frac{\int_{\mathbb{R}} u \mathcal{L}(u, Y_1, \dots, Y_n) \pi(u) du}{\int_{\mathbb{R}} \mathcal{L}(u, Y_1, \dots, Y_n) \pi(u) du} \quad (2)$$

Cette solution est en général une fonction non linéaire relativement difficile à calculer à partir des lois de X et de V . Aussi, on ne considère souvent que les solutions linéaires du problème ci-dessus. Dans ce cas seuls les moments du premier et du second ordre de X et V apparaissent dans l'estimateur linéaire des moindres carrés qui s'écrit

$$\frac{EX \text{ Var } V}{n \text{ Var } X + \text{Var } V} 1_{\Omega} + \frac{\text{Var } X}{n \text{ Var } X + \text{Var } V} \sum_{i=1}^n Y_i \quad (3) \quad \blacksquare$$

En fait, dans le cas où V est gaussien on peut noter que l'estimateur

$$\frac{\int_{\mathbb{R}} u \mathcal{L}(u, Y_1, \dots, Y_n) du}{\int_{\mathbb{R}} \mathcal{L}(u, Y_1, \dots, Y_n) du} \quad (4)$$

se confond aussi avec la moyenne empirique.

3) INTRODUCTION D'UN MODELE D'INFORMATION A PRIORI SUBJECTIVE NON PROBABILISTE

On peut motiver l'introduction d'un modèle non probabiliste pour l'information a priori par au moins deux remarques. On s'intéresse ici au cas de l'information a priori subjective, c'est-à-dire le cas épistémologique et non physique ou objectif.

i) Remarque 1 : Si l'on compare les formules (2) et (4) on constate que l'estimateur donné par (4) qui correspond au cas d'absence d'information a priori, est un cas particulier de l'estimateur (2) où la fonction π véhiculant la connaissance a priori serait telle que $\pi(u) = 1$ (ou = constante) pour tout $u \in \mathbb{R}$.

Une fonction π identiquement constante sur \mathbb{R} correspondrait donc à une absence d'information a priori, ou encore au cas de l'ignorance totale sur la valeur possible x . Une telle fonction π ne peut pas être alors une densité et par conséquent X ne peut pas être à son tour une v.a.r..

Cette observation n'est pas récente puisque dans (Jeffreys 1961) on prend aussi une fonction constante sur toute la droite réelle pour représenter l'information a priori dans le cas de l'ignorance totale. Il est prouvé (voir par exemple Zellner 1971)) que cette représentation est conforme au principe du maximum d'entropie. Dans ce modèle d'information a priori la notion de probabilité est impropre puisque la mesure correspondante peut prendre la valeur infini.

Si l'on considère que V est une v.a. gaussienne, et que l'information a priori est non probabiliste mais de type gaussien

$$\pi(u) = e^{-\frac{(u - m_0)^2}{2\sigma^2}}, \quad u \in \mathbb{R}$$

(π n'est plus une densité de probabilité),

on constate que l'estimateur donné par (2) se confond avec l'estimateur linéaire des moindres carrés (3). Si enfin on fait $\sigma_0^2 = \infty$, alors le modèle gaussien non probabiliste ci-dessus se confond avec le modèle d'ignorance totale ($\pi(u) = 1 \forall u \in R$) et l'estimateur (2) se confond cette fois avec la moyenne empirique (1).

En somme, en admettant que π puisse appartenir à une classe différente de fonctions on a pu établir un pont entre les estimateurs du type non bayésien et ceux du type bayésien. En conséquence la distinction entre ces deux types d'estimateurs ne serait plus nécessaire. A ce niveau, un autre point de vue (Zacks (1971)) est de considérer les estimateurs non bayésiens comme la limite d'estimateurs bayésiens ; on évite alors d'attaquer de front le problème d'un modèle d'information a priori.

ii) Remarque 2 : Puisque l'on se place dans un contexte épistémologique, on admet que l'on peut donc avoir plusieurs "points de vue" ou modèles d'information a priori, avant de procéder à une évaluation de x .

Afin de comparer ces modèles on considère l'erreur quadratique moyenne résultant de l'application des estimateurs correspondants. Plus précisément si X_n est un des estimateurs étudiés, on considère $E(X_n - x | \Omega)^2$ où E est calculé relativement à la probabilité P .

Enfin on fait l'hypothèse généralement adoptée que V est une v.a.r. de loi $N(0, \sigma^2)$. Tout calcul fait, on trouve pour le moyenne empirique une erreur quadratique moyenne égale à

$$\frac{\sigma^2}{n} ;$$

quant à l'estimateur des moindres carrés, avec l'hypothèse que X est une v.a. de loi $N(m_0, \sigma_0^2)$, on trouve

$$\frac{x^2 \sigma^4 - 2 x m_0 + n \sigma_0^4 \sigma^2 + m_0 \sigma^4}{(n \sigma_0^2 + \sigma^2)^2}$$

Les deux expressions ci-dessus permettent de comparer ces deux estimateurs. Si l'information a priori (m_0, σ_0^2) est bonne, c'est-à-dire si x appartient à l'intervalle $\left[m_0 - \sqrt{2 \sigma_0^2 + \sigma^2/n}, m_0 + \sqrt{2 \sigma_0^2 + \sigma^2/n} \right]$ l'estimateur des moindres carrés donne une erreur quadratique moyenne plus faible que la moyenne empirique. Si par contre l'information a priori n'est pas en accord avec x , on remarque que l'estimateur des moindres carrés peut donner une erreur très grande (croissance quadratique) comparée à celle (constante) de la moyenne empirique.

En conséquence, le modèle a priori de type gaussien ne semble donc pas satisfaisant. Peut-on trouver un autre modèle susceptible d'améliorer l'erreur quadratique moyenne lorsque l'information a priori encadre bien x sans risque que l'erreur devienne très grande si la connaissance a priori est mauvaise ?

En fait, on peut trouver une solution simple et relativement naturelle à ce problème en abandonnant une fois de plus le modèle probabiliste. Les remarques ci-dessus suggèrent que l'on choisisse le modèle d'information a priori donné par

$$\pi(u) = \begin{cases} 1 & \text{pour } u \in I \\ t & \text{sinon} \end{cases}$$

où $t \in]0, 1[$ et I un intervalle borné de \mathbb{R} .

Ce modèle peut être interprété en disant que a priori x est considéré comme appartenir à l'intervalle I mais que l'on n'exclut pas avec un degré t que x soit à l'extérieur de cet intervalle.

Soit f_n la fonction définie par

$$z \in \mathbb{R} \mapsto f_n(z) = \frac{\int_I u \mathcal{L}_n(u, z) du + t \int_{\bar{I}} u \mathcal{L}_n(u, z) du}{\int_I \mathcal{L}_n(u, z) du + t \int_{\bar{I}} \mathcal{L}_n(u, z) du}$$

où \mathcal{L}_n est la fonction définie par

$$(u, z) \in \mathbb{R}^2 \mapsto \mathcal{L}_n(u, z) = e^{-\frac{(u-z)^2}{2\sigma^2/n}}$$

Avec l'hypothèse que V est une v.a. gaussienne,

$$X_n = f_n(\bar{Y}_n)$$

se confond avec l'estimateur donné par (2) où π est la fonction définie ci-dessus.

On vérifie maintenant que l'estimateur ci-dessus ne fait courir aucun risque dans le sens où l'erreur quadratique moyenne reste bornée quelle que soit la qualité de l'information a priori.

Théorème 1 (X_n ne fait courir aucun risque)

pour $x = \pm \infty$ $E(X_n - x | \Omega)^2$ est majoré sur R et la borne supérieure est σ^2/n .

On sait donc que X_n ne fait courir aucun risque, il faut maintenant s'assurer qu'il est meilleur que la moyenne empirique lorsque l'information a priori "coïncide" avec la valeur x .

Théorème 2 (X_n peut être meilleur que \bar{Y}_n)

Si x est au milieu de I alors

$$E(X_n - x | \Omega)^2 < \frac{\sigma^2}{n}.$$

La réponse à la question posée ci-dessus est donc positive, il existe bien un modèle a priori conciliant les exigences.

Les remarques précédentes 1 et 2 conduisent donc à introduire un nouveau modèle de connaissance a priori distinct du modèle probabiliste dans le sens où on ne vérifie plus l'hypothèse de densité pour π ($\int_R \pi(u) du = 1$). La seule contrainte sur π est qu'elle soit bornée par 1 (une autre borne pourrait convenir).

En rapprochant ce résultat de (Zadeh 1978), on est tenté de considérer π comme une distribution de possibilité.

On constate finalement que la notion de mesure de possibilité (- de F-additivité) remplace avantageusement celle de mesure de probabilité (- d'additivité simple) dans la modélisation de l'information a priori subjective puisqu'elle inclut la sous famille de mesures de t-1 possibilité (cf. (Banon 1979))

$$A \subset R \mapsto \begin{cases} 1 & \text{si } A \cap I \neq \emptyset \\ t & \text{si } A \cap I = \emptyset \\ 0 & \text{si } A = \emptyset \end{cases} \quad t \in]0, 1[,$$

qui peuvent, le cas échéant, améliorer l'évaluation de la moyenne sans faire par ailleurs courir de risque comme on l'a vu par les théorèmes 1 et 2.

4) INTRODUCTION D'UN NOUVEL ESTIMATEUR DE LA MOYENNE

Pour estimer la moyenne, on propose donc d'utiliser l'estimateur donné par (2)

$$\frac{\int_{\mathbb{R}} x \mathcal{L}(x, Y_1, \dots, Y_n) \pi(x) dx}{\int_{\mathbb{R}} \mathcal{L}(x, Y_1, \dots, Y_n) \pi(x) dx}$$

où, cette fois-ci π n'est plus une densité mais une distribution de t-1 possibilité centrée sur un intervalle I de \mathbb{R} .

La moyenne à estimer se présente comme la valeur x prise par une variable floue X qui serait "indépendante" de la suite de v.a.r.

$$\left\{ Y_i \right\}_{i=1}^n$$

Bien que X ne soit plus une variable aléatoire on constate que l'estimateur ci-dessus reste une variable aléatoire. En somme l'estimateur proposé se présente comme l'espérance mathématique d'une variable floue (de type possibiliste) conditionnée par une suite de variables aléatoires.

Cette propriété importante se retrouve dans la règle de Dempster (voir (Dempster 1967) et (Shafer 1976)). Lorsque l'on combine entre elles une mesure floue et une mesure de probabilité par la règle de Dempster on obtient une mesure de probabilité. En fait cette coïncidence n'est pas fortuite. En effet l'estimateur ci-dessus découle de la règle de Bayes lorsque π est une densité de probabilité et on constate avec un peu d'attention que c'est le mécanisme de la règle de Dempster qui est impliqué lorsque π est plus généralement une densité floue (ou une distribution de possibilité). La règle de Dempster se présente d'ailleurs comme une extension de la règle de Bayes.

Afin de se rendre compte du comportement de l'erreur quadratique moyenne consécutive de l'emploi de tel ou tel type d'information a priori on a rédigé un programme Fortran pour le calcul approché de $E(X_n - x | \Omega)^2$ dans le cas particulier où V est une v.a. gaussienne $N(0, n)$.

Sur la figure 1 apparaît 4 courbes correspondant respectivement à $\pi(u) = 1$,

$$\pi(u) = \begin{cases} 1 & \text{si } u \in [-a, a] \\ \frac{1}{2} & \text{sinon} \end{cases}, \quad \pi(u) = \begin{cases} 1 & \text{si } u \in [-a, a] \\ \frac{1}{4} & \text{sinon} \end{cases} \quad \text{et}$$

$\pi \leftrightarrow N(0, \frac{a^2}{3})$, avec $a=1$.

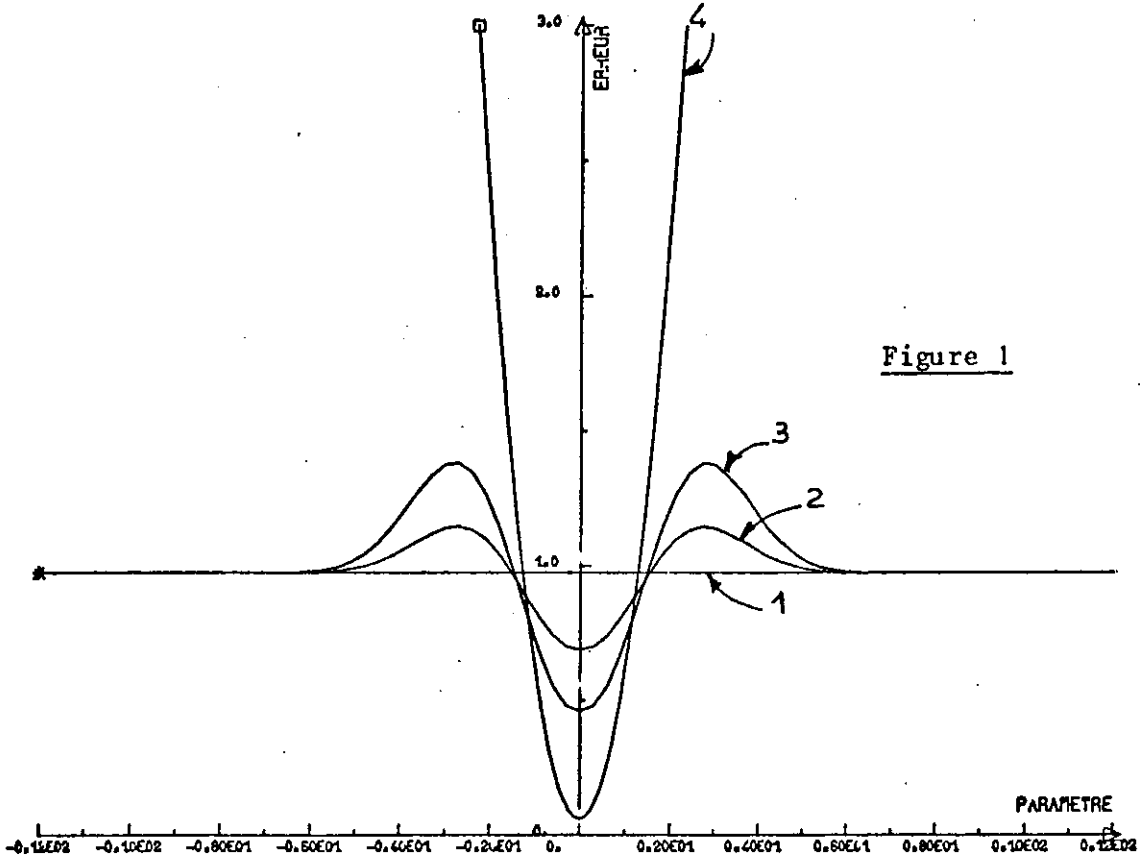


Figure 1

On constate que le comportement de l'erreur est semblable à celui du cas gaussien dans l'intervalle $[-a, a]$ et semblable à celui du cas d'ignorance totale à l'extérieur de cet intervalle, ce qui est certainement très satisfaisant pour les applications.

5 - CONCLUSION

Au cours de cette étude on a montré les inconvénients du modèle probabiliste et en particulier du modèle gaussien comme moyen de représenter l'information a priori subjective en vue de l'estimation d'un paramètre réel (la moyenne). Un modèle plus satisfaisant a été proposés qui correspond à une famille particulière de mesures de possibilité. En fait l'approche choisie ne prend son sens que dans la mesure où l'on suppose que le paramètre à estimer prend ses valeurs sur toute la droite réelle.

REFERENCES

- G. BANON, Distinction between several subsets of fuzzy measures. Soumis à International Journal of Fuzzy Sets and Systems (Avril 1979)
- A.P. DEMPSTER, Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. 38, 1967 , 325-339.
- A.H. JAZWINSKI, Stochastic processes and filtering theory. Academic Press, New York, 1970.
- H. JEFFREYS, Theory of probability. Oxford University Press, Oxford, Third Edition, 1961.
- G. SHAFER, A mathematical theory of evidence. Princeton University Press, Princeton, New Jersey, 1976.
- S. ZACKS, The theory of statistical inference. John Wiley and Sons, New York, 1971.
- L.A. ZADEH, Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1 (1), 1978, 3-28.
- A. ZELLNER, An introduction to bayesian inference in econometrics, John Wiley and Sons, New York, 1971.