

HUITIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS



NICE du 1^{er} au 5 JUIN 1981

RECONSTITUTION D'UNE LOI DE PROBABILITE

Hubert DEBART

SINTRA-ALCATEL, 1, avenue Aristide Briand 94117 ARCUEIL

RESUME

Il est relativement aisé de reconstituer une loi de probabilité dont on connaît la forme générale ; il suffit de fixer la valeur d'un nombre donné de paramètres.

Le problème n'a pas de solution générale si la forme de la loi est totalement inconnue, en particulier si la loi n'est connue que par un nombre restreint de réalisations.

On renonce alors à connaître la forme exacte de la loi. Par contre, on recherche une variable aléatoire dont la variation est minimale, avec une probabilité suffisamment grande. Après avoir traité le problème pour des lois de probabilité connue, on recherche un algorithme conduisant à un résultat approché quand la loi est inconnue ; la transformation ainsi définie est non-linéaire.

La méthode s'étend aux variables à plusieurs dimensions; on peut l'appliquer aux problèmes de reconnaissance de forme.

La méthode de transformation non-linéaire de I. ALKESANDER est apparentée à ce procédé.

SUMMARY

It is rather easy to reconstruct a law of probability whose general (functional) form is given ; it reduces to determination of a given number of parameters.

The problem has no general solution if the law is of totally unknown form, specially if the law is given just by a few realizations.

Then the exact form of the law is unreachable. In this case we search for a random variable whose variation is minimized, with a sufficiently great probability.

We begin with the treatment of this problem for a given law of probability ; then we look for an algorithm leading to an approximate result, if the law is entirely unknown ; the transformation so defined is non-linear.

This method can be extended to many dimensional laws and can be applied to problems of pattern recognition.

The non-linear transformation device of I. ALEKSANDER is acquainted with this method.



I. INTRODUCTION

Le problème le plus courant en statistique est de reconstituer une loi de probabilité dont on connaît la forme générale et dont il faut estimer un certain nombre de paramètres caractéristiques (par exemple la moyenne et la variance d'une loi gaussienne).

Cependant, si on ignore tout de la loi, qui n'est connue que par un certain nombre (souvent très limité) de réalisations, le problème devient très difficile. En particulier, il se pose de cette façon dans la reconnaissance de formes, où on cherche les probabilités qui peuvent être affectées à une figure inconnue.

En fait, la réalisation d'une loi inconnue dont on dispose transporte en général trop d'informations par rapport au résultat cherché. Si par exemple on veut décider qu'une lettre est un A, on ne s'intéresse pas à la graphie particulière de la réalisation présentée à la reconnaissance.

On se pose alors les problèmes suivants :

- quant on connaît une loi de probabilité d'une variable X , il existe un intervalle privilégié dans lequel par exemple la densité $p(X)$ est supérieure à un nombre donné p_0 . Peut-on définir une fonction $f(X)$ dont la dispersion sera minimale dans l'intervalle privilégié.
- Si on ne connaît pas la forme de la loi déterminée seulement par n réalisations, peut-on construire de façon approchée une fonction $f(X)$ qui répondra aux mêmes conditions ?

II. FONCTION A DISPERSION MINIMALE

Une fonction $f(X)$ ne peut être considérée à dispersion minimale dans un intervalle donné de X que si on lui impose une condition supplémentaire. Sinon, il suffit de choisir une constante dont la dispersion est nulle.

Une condition supplémentaire aboutit à la solution d'un problème classique de variation.

Par exemple : on mesure la dispersion d'une fonction f par sa moyenne quadratique :

$$\int_I f^2 P \, dx$$

et on veut avoir une "variation totale" de la fonction donnée dans l'intervalle I : mesurée par :

$$\int f'^2 P \, dx \quad (1)$$

On introduit alors le multiplicateur de LAGRANGE λ et on minimise l'expression

$$\int f^2 P \, dx - \lambda \int f'^2 P \, dx.$$

On doit donc écrire la condition d'Euler :

$$fP = \frac{d}{dx} \left[\lambda f' P \right]$$

où : $fP = \lambda (f''P + f'P')$

où enfin : $\lambda f'' + \frac{\lambda P' f'}{P} - f = 0$.

équation linéaire dont la solution est en général accessible. L'avantage de cette formulation est que la forme analytique de la solution ne dépend pas de l'intervalle I .

Exemple :

distribution exponentielle $P(X) = e^{-X}$. $\frac{P'(X)}{P(X)} = -1$

On a l'équation $\lambda f'' - \lambda f' - f = 0$.
avec une solution exponentielle oscillante ou non suivant la valeur de λ .

La condition (1) peut d'ailleurs être remplacée par bien d'autres. L'inconvénient majeur de cette formulation est qu'elle suppose connue la loi de probabilité.

Pour essayer d'y obvier, on se limitera au cas où $f(X)$ est un polynôme en X .

III. DISPERSION MINIMALE D'UN POLYNOME

On considère une loi de probabilité donnée $P(X)$ d'une variable aléatoire X . On cherche à minimiser, dans un certain intervalle, la moyenne quadratique d'un polynôme de degré n en X : $P_n(X)$. Le problème n'a de sens que si on introduit une normalisation du polynôme :

$$\int_I P_n^2(X) dx = 1.$$

L'intervalle I peut ne pas être simplement convexe, mais formé de la réunion d'intervalles disjoints.

On a : $P(X) = aX + bX^2 + \dots + lX^n$
(il n'y a pas de terme constant car on tomberait sur la solution triviale : $P(X) = \text{cste}$).

On introduit alors le multiplicateur de LAGRANGE λ . Pour une commodité d'écriture, on ramène l'intervalle I à $(0,1)$ sans réduire la généralité ; on a alors à minimiser l'intégrale (1) :

$$\int_0^1 (aX + bX^2 + \dots + lX^n)^2 P(X) dx - \lambda \int_0^1 (aX + \dots + lX^n)^2 dx$$

Ce qui fournit des conditions :

$$aM_2 + \dots + lM_{n+1} = \lambda \left(\frac{a}{3} + \dots + \frac{l}{n+2} \right)$$

$$aM_3 + \dots + lM_{n+2} = \lambda \left(\frac{a}{4} + \dots + \frac{l}{n+3} \right)$$

$$aM_{n+1} + \dots + lM_{2n} = \lambda \left(\frac{a}{n+2} + \frac{l}{2n+1} \right)$$

Où :

$$a(M_2 - \frac{\lambda}{3}) + \dots + l(M_{n+1} - \frac{\lambda}{n+2}) = 0$$

$$a(M_3 - \frac{\lambda}{4}) + \dots + l(M_{n+2} - \frac{\lambda}{n+3}) = 0$$

.

.

$$a(M_{n+1} + \frac{\lambda}{n+2}) + \dots + l(M_{2n} - \frac{\lambda}{2n+1}) = 0$$

Ce sont des équations linéaires homogènes, avec une matrice

$$M = \begin{vmatrix} M_2 - \frac{\lambda}{3} & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & M_{2n} - \frac{\lambda}{2n+1} \end{vmatrix}$$

Le système n'a de solution que si λ est égal à une valeur propre de la matrice.



RECONSTITUTION D'UNE LOI DE PROBABILITE

L'équation aux valeurs propres $D(M) = 0$ n'est pas sous la forme canonique mais on peut s'y ramener par transformation linéaire des variables.

On constate aisément que l'intégrale (1) est nulle si les conditions de minimum sont réalisées. Donc : la moyenne quadratique obtenue est égale à la valeur propre correspondante (puisque $\int_0^1 (aX + \dots + \lambda X^n)^2 dx = 1$)

Il faut donc utiliser la plus petite des valeurs propres de M , (qui sont réelles et positives).

Dans les équations, on a posé :

$$M_p = \int_0^1 X^p P(X) dx$$

Exemple : loi exponentielle entre 0 et 1.

$$\text{On a alors : } M_p = \frac{e}{e-1} \int_0^1 X^p e^{-X} dx$$

(pour normaliser les probabilités par $\int_0^1 P(X) dx = 1$)

$$M_1 = 0,418023$$

$$M_2 = 0,25407$$

$$M_3 = 0,180233$$

$$M_4 = 0,138955$$

$$M_5 = 0,112978$$

$$M_6 = 0,094813$$

a) Polynôme du deuxième degré :

On a à résoudre :

$$(M_2 - \frac{\lambda}{3})a + (M_3 - \frac{\lambda}{4})b = 0$$

$$(M_3 - \frac{\lambda}{4})a + (M_4 - \frac{\lambda}{5})b = 0$$

La plus petite valeur propre est $\lambda = 0,66827$.

C'est aussi la moyenne quadratique minimale du polynôme.

b) Polynôme du troisième degré :

Même procédure :

$$(M_2 - \frac{\lambda}{3})a + (M_3 - \frac{\lambda}{4})b + (M_4 - \frac{\lambda}{5})c = 0 \dots \text{etc.}$$

La plus petite valeur propre est : 0,68803. Le gain est très minime.

IV. LOI DE PROBABILITE INCONNUE : RECHERCHE D'UNE SOLUTION APPROCHEE

Tous les calculs précédents supposent évidemment la connaissance des moments c'est-à-dire celle de la loi de probabilité.

Supposons alors que tel n'est pas le cas. On ne connaît la loi que par n réalisations qui ont fourni les valeurs $X_1 \dots X_n$ (rangées dans un ordre croissant). On peut toujours supposer que $X_1 = 0$, $X_n = 1$; on s'y ramène par un simple changement linéaire d'échelle.

On considère alors le polynôme :

$$X(X - X_2) \dots (X - X_n)$$

X est une réalisation ultérieure. Si X est dans l'intervalle $(0,1)$ le polynôme est borné par 1. On peut alors l'utiliser comme détermination approchée du polynôme défini précédemment.

Prenons un exemple pour préciser les choses : c'est celui de la loi exponentielle entre 0 et 1, déjà utilisé.

a) On prend 3 valeurs ramenées à 0, a, 1 et on forme le polynôme :

$$X(X-a)(X-1) = X^3 - (1+a)X^2 + aX.$$

Considérons la moyenne quadratique, ou plutôt la moyenne de cette moyenne quand a prend toutes les valeurs possibles. C'est :

$$\frac{X^6 + (1+a)^2 X^4 + a^2 X^2 - 2(1+a)X^5 + 2aX^4 - 2a(1+a)X^3}{6}$$

C'est-à-dire avec les relations précédentes :

$$M_6 + (1+2M_1+M_2)M_4 + M_2^2 - 2(1+M_1)M_5 + 2M_1M_4 - 2(M_1+M_2)M_3 = 0,00329.$$

L'intégrale du carré du polynôme est en moyenne :

$$\int_0^1 X^2 (X-a)^2 (X-1)^2 dx$$

Soit :

$$\frac{1}{7} + (1+2M_1+M_2)\frac{1}{5} + \frac{M_2}{3} - \frac{2}{6}(1+M_1) + \frac{2M_1}{5} - \frac{2(M_1+M_2)}{4} = 0,00406.$$

Soit un rapport de 0,810 un peu plus élevé que l'optimum trouvé précédemment.

b) Résultat expérimental :

On tire les valeurs au hasard prises dans une loi exponentielle e^{-X}

$$X_1 = 0$$

$$X_2 = 0,076$$

$$X_3 = 0,172$$

$$X_4 = 0,268$$

$$X_5 = 0,268$$

$$X_6 = 0,510$$

$$X_7 = 0,559$$

$$X_8 = 0,667$$

$$X_9 = 0,843$$

$$X_{10} = 1$$

On forme le polynôme :

$$(X-X_1) \dots (X-X_{10})$$

et on calcule sa moyenne quadratique pour 1000 échantillons tirés au hasard dans l'intervalle $(0,1)$ et son intégrale :

$$MQ = 3,396.10^{-10}$$

$$I = 5,0138.10^{-10}$$

Soit un rapport 0,67133, cette fois très proche de l'optimum.



Quand les valeurs tirées sortent de l'intervalle (0,1) la valeur du polynôme est très élevée par rapport aux valeurs de l'intervalle (0,1) d'où un excellent pouvoir de discrimination.

V. PROCESSUS ALEATOIRE

Les considérations précédentes s'appliquent à un processus aléatoire dans un intervalle (0,T) qui est seulement connu par n réalisations $f_1(t) \dots f_n(t)$ ($0 \leq t \leq T$).

On peut supposer que $f(t) = 0$ dans tout l'intervalle (0,T). S'il n'en était pas ainsi, on ajouterait au processus une fonction certaine $M(t)$ et le problème n'en serait pas changé.

Si on forme l'expression :

$$\left[f(t) - f_1(t) \right] \dots \left[f(t) - f_n(t) \right] \quad (1)$$

$f(t)$ étant une réalisation ultérieure les considérations précédentes s'appliquent à tout instant t_0 ($0 < t_0 < T$). On obtient par l'expression (1) une moyenne quadratique $M(t)$ et la quantité $\int_0^T M(t) dt$ est caractéristique de la réalisation $f(t)$.

VI. METHODE DE RECONNAISSANCE DE I. ALEKSANDER

Cette méthode utilise une transformation non linéaire d'un processus. Elle a été appliquée à la reconnaissance des lettres. Une lettre est échantillonnée par n éclairissements. L'éclairissement $f(i)$ définit le processus échantillonné $f(j) \dots f(n)$.

On applique sur le processus une transformation non linéaire qui s'appuie sur n réalisations connues.

On définit une moyenne M du processus : $\overline{f(1)} \dots \overline{f(n)}$ et on introduit une transformation non linéaire $T(f)$ telle que :

$$\begin{aligned} T(f_1) &= M \\ &\vdots \\ T(f_n) &= M \end{aligned}$$

c'est-à-dire de mêmes propriétés que la transformation précédemment définie.

Mais dans la pratique cette transformation est booléenne et utilise les échantillons binaires : $f(i) = 0$ ou 1.

La lettre à reconnaître est analysée en n points sur une matrice d'entrée M_e . On tire au hasard un certain nombre (n exactement) de triplets logiques T_i dont la sortie (0 ou 1) est amenée sur une matrice de sortie M_s .

S'il s'agit de définir les propriétés aléatoires de la lettre A, on définit une figure "A moyenne" et on ajuste les réponses des triplets T_i de façon à obtenir la réponse "A moyenne" pour n entiers représentant n échantillons de A.

On définit ainsi une transformation non linéaire qui a bien les propriétés de la transformation définie au paragraphe précédent.

Une lettre inconnue est présentée à la matrice d'entrée. Il lui correspond une certaine transformée qu'on ramène à l'entrée, et on rejette le cycle jusqu'à obtenir :

- soit A moyenne (reconnaissance)
- soit une autre figure (refus)
- soit un processus cyclique (incertitude).

Les inconvénients de cette réalisation assez brutale de transformation non linéaire sont :

- la saturation des circuits T_i , en cas de nombreuses lettres différentes à connaître
- l'importance égale attribuée à toutes les n réalisations du processus servant à définir le système, une lettre A aberrante provoque un phénomène d'attraction de figures parasites dans la catégorie "A".

VII. CONCLUSION

On sait séparer des variables aléatoires et des processus aléatoires gaussiens. Si cette condition n'est pas réalisée, le problème est d'un ordre de difficulté très supérieure car la connaissance de la corrélation simple ne suffit plus. Par ailleurs, le fait qu'on connaît souvent les processus aléatoires à séparer très imparfaitement et que l'information à en extraire est très rudimentaire par rapport aux contenus des processus, amène naturellement à envisager des transformations non linéaires ; c'est-à-dire non réversibles, qui font perdre une partie de l'information au bénéfice d'un pouvoir discriminant plus grand. Cette technique devrait beaucoup progresser dans les années à venir.

