COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL
ET SES APPLICATIONS

NICE du 16 au 21 JUIN 75

THEORIE DE LA COMMUNICATION
EXEMPLE DES "PROBABILITES D'ESTIMATION"

Hubert DEBART - CIT/ALCATEL

Avenue Aristide Briand 94110 - ARCUEIL

RESUME

On a voulu donner ici un exemple d'application des méthodes générales utilisées dans la théorie de la communication.

La réduction des ensembles de signaux et du bruit à des modèles simples conduit à dégager des notions générales qui à leur tour sont utilisées dans la résolution de problèmes pratiques. On a introduit ici la notion de "probabilités d'estimation" relative à la détection d'un ensemble de signaux en présence de bruit et calculables de façon récurrente. Cette notion dont l'intérêt immédiat n'est pas apparent à première vue conduit à une méthode de calcul des matrices de confusion entre signaux d'un code, ou entre signaux correspondant aux diverses valeurs d'un paramètre en présence d'un bruit élevé ; à une méthode de construction de codes optimaux ; enfin, à une approche de l'estimation de KALMAN-BUCY dans des hypothèses plus générales que les hypothèses courantes.

SUMMARY

This paper consists of an attempt to show an example for the application of general methods of use in communication theory.

Reducing sets of signals and noise to simple patterns leads to enhance general concepts, which are of use in solving actual problems. Here we have introduced the notion of "probability of estimation", related to the detection of a set of signals affected by noise, derivable by a recursive calculation. This concept, whose interest is not obvious at first look, leads to a method for deriving the confusion matrices, for a set of encoded signals, or for a set of signals corresponding to the different values of a parameter, with a high level of noise ; on the other hand, to a method for constructing optimal codes, and to an approach of the KALMAN-BUCY estimation, according to more general hypotheses than the usual ones.

INTRODUCTION -

Le problème majeur de la théorie de la communication est celui de la reconstitution d'un signal envoyé. Il est infiniment divers car il dépend à la fois de la nature de l'ensemble des messages transmis et des contraintes qui pèsent sur le canal de transmission et sur le détecteur.

Depuis l'article fondamental de Shannon en 1948, il a été abordé de deux façons principales. On a recherché d'abord les possibilités extrêmes d'un canal donné, en supposant que tous les moyens de détection possibles étaient disponibles, et ces recherches se poursuivent encore, en faisant varier l'ensemble des messages et les contraintes sur le canal. Par ailleurs, il est certain que les détecteurs réels ont des possibilités très limitées ; on a donc également suivi une autre voie, en supposant qu'on détecte des échantillons successifs des messages, et qu'on améliore le résultat en tenant compte des propriétés générales du message. Cette opération prend des formes très diverses ; elle englobe par exemple l'utilisation des codes correcteurs d'erreurs, quand les messages sont parfaitement définis, et la méthode de Kalman-Bucy, quand on n'en connaît que des caractéristiques aléatoires.

Il serait très long et fastidieux d'essayer de rassembler tous les résultats obtenus. Nous avons simplement voulu montrer ici que l'application de la théorie générale peut revêtir des formes très diverses ; en particulier, il arrive que l'examen des problèmes généraux se traduise par l'introduction d'une notion nouvelle et qu'en retour cette notion aide à la résolution de nombreux problèmes : on peut citer l'exemple de l'entropie de Shannon.

Dans la suite, et à titre d'exemple, nous introduisons la notion de "probabilité d'estimation" et nous recherchons quelle contribution elle peut apporter à la solution de divers problèmes.

I. - PROBABILITES D'ESTIMATION : DEFINITION

Un détecteur a à décider de la présence d'un parmi une famille de n signaux. En fait, quand le canal de transmission est affecté de bruit, cette décision est simplement une action simplificatrice. L'existence du bruit se traduit par une probabilité d'erreur sur la décision. Si le signal 1 a été envoyé, il est interprété comme tel par le détecteur avec une probabilité P, et il existe des probabilités de confusion avec tous les autres signaux de la famille q_{12}, q_{13} ...q_{1n}, avec la relation :

$$P + q_{12} + \ldots + q_{1n} = 1.$$

Ce que le détecteur peut annoncer est donc par exemple "présomption de présence du signal 1 avec une probabilité P, du signal 2 avec une probabilité q_{12}, ... du signal n avec une probabilité q_{1n}, s'il "décide" en faveur du signal 1.

Supposons alors que le signal reçu soit échantillonnable, ce qui pratiquement est toujours le cas, et que les échantillons successifs soient affectés par le bruit de façon indépendante.

On peut voir alors l'opération de détection de la façon suivante : avant toute mesure, les probabilités de présence des différents signaux de la famille sont connues ; en général, elles sont égales ; chacune est égale à $\frac{1}{n}$. Chaque réception et traitement d'échantillon modifie cette situation et donne un nouveau tableau de probabilités de présence, représentable par un vecteur colonne ; l'opération de révision s'arrête à la réception du dernier échantillon.

Le rang du signal dans la famille peut être interprété comme un paramètre j susceptible de prendre les valeurs 1, 2 . . . n. Si nous désignons par x le résultat de la mesure faite sur un échantillon quelconque, l'existence du bruit se traduit par une probabilité conditionnelle p(x / i) que la mesure donne le résultat x, sachant que le signal en cause est le signal de rang i.

Avant la mesure faite sur cet échantillon, le vecteur colonne de probabilités de présence est $q_m(i)$. Si on appelle m le numéro du signal dans la famille et i le rang de l'échantillon dans le temps, si la mesure donne le résultat x sur l'échantillon i + 1, le nouveau vecteur colonne sera déterminé par la règle de BAYES :

$$q_m(i+1) = q_m(i) \times \frac{p(x/i)}{\sum_{i=1}^{i=n} q_m(i)\, p(x/i)}$$

On arrive donc à un vecteur colonne final quand on connaît tous les résultats de mesure $x_1 \ldots x_n$ effectués sur les r échantillons du signal.

Ce vecteur colonne exprime la décision du détecteur.

Supposons alors qu'un signal particulier ait été envoyé, on peut lui affecter le numéro 1 sans nuire à la généralité.

Les vecteurs colonne de probabilités pourront alors porter la désignation de $q_{1,m}(i)$ avec la signification : le signal 1 a été envoyé ; après la réception des i premiers échantillons, le détecteur affecte au signal m la probabilité de présence $q_{1,m}(i)$; le vecteur final $q_{1,m}(s)$ dépend de la succession des résultats de mesure $x_1 \ldots x_s$ qui ont été effectivement trouvés. On peut alors moyenner ce vecteur sur l'ensemble des groupes de résultats $(x_1 \ldots x_s)$ possibles. On définira alors des vecteurs moyens $\Pi_{1,m}(s)$ appelés "vecteurs de probabilités d'estimation" et définis par la relation de récurrence :

$$\Pi_{1,m}(i+1) = \Pi_{1,n}(i) \frac{p(x/1)\, p(x/i)}{\sum_{i=1}^{i=n} \Pi_{1,r}\, p(x/r)}$$

Comme il s'agit d'une définition récurrente, on peut très bien supposer que les probabilités $p(x/i)$ qui résument l'action du bruit varient d'un échantillon à l'autre.

Si on considère l'ensemble des signaux émis, on obtient une matrice carrée

$$\left[\Pi_{q,m}(s)\right] \quad (n \times n)$$

Il correspond à cette matrice une entropie

$$-\sum_{q=1}^{q=n} p_0(q) \left[\sum_{m=1}^{m=n} \Pi_{q,m}(s) \operatorname{Log} \Pi_{q,m}(s)\right]$$

qu'on appellera "entropie de reconstitution".

$p_0(q)$ désigne la probabilité de présence a priori du signal q.

Examinons d'abord le cas où les signaux ne peuvent prendre que des valeurs discrètes.

La quantité "probabilité d'estimation" est alors définie par la loi de récurrence :

$$\Pi_{q,m}(i+1) = \Pi_{q,m}(i) \sum_{x=1}^{x=M} \frac{p(x/q)\, p(x/m)}{\sum_{r=1}^{r=N} p(x/r) \Pi_{q,r}(i))}$$

dans laquelle $p(x/q)$ désigne la probabilité conditionnelle d'observer x quand q a été émis.

Cette loi de correspondance est markovienne puisque la connaissance de la matrice $\left[\Pi_{q,m}(i)\right]$ entraîne celle de la matrice $\left[\Pi_{q,m}(i+1)\right]$. Il ne s'agit pas évidemment d'une correspondance markovienne classique, c'est à dire linéaire et homogène, mais le caractère markovien subsiste cependant.

On peut vérifier facilement que la condition

$$\sum_{m=1}^{m=N} \Pi_{q,m}(i) = 1 \qquad \text{entraîne :}$$

$$\sum_{m=1}^{m=N} \Pi_{q,m}(i+1) = 1$$

On peut obtenir une formule analogue si la famille de signaux dépend de façon continue d'un paramètre α. Il faut alors introduire une densité de probabilité d'estimation, notée

$$\varpi_{\alpha_0,\alpha}(i)$$

(α_0 est la vraie valeur du paramètre) et de l'échantillon de rang i à l'échantillon suivant, on a la loi de correspondance

$$\omega_{\alpha_0,\alpha}(i+1) = \omega_{\alpha_0,\alpha}(i) \int \frac{p(x/\alpha_0)\, p(x/\alpha)\, dx}{\int_\beta p(x/\beta)\, \varpi_{\alpha_0,\beta}(i)\, d\beta}$$

formule dans laquelle $p(x / \alpha)$ désigne la densité de probabilité du vecteur de mesure x, connaissant α ; $\int_x$ désigne une intégration sur toutes les valeurs possibles du vecteur de mesure, qui peut être multidimensionnel ; $\int_\beta$ désigne une intégration sur tout l'intervalle de variation du paramètre α.

Dans le cas où les signaux sont constitués par des trains d'impulsions binaires, le calcul est fortement simplifié. On appellera P la probabilité de détection correcte de l'une ou l'autre des deux impulsions élémentaires (on pourrait d'ailleurs supposer que les deux probabilités sont différentes) et p = 1 - P la probabilité d'erreur.

Supposons que l'ensemble des signaux à détecter soit repéré par les nombres 1... n et que le signal envoyé soit le signal 1.

Les formules de récurrence permettant de calculer les probabilités d'estimation $\Pi_{1,m}(i+1)$ quand on connaît $\Pi_{1,m}(i)$ s'écrivent simplement

$$\Pi_{1,1}(i+1) = \Pi_{1,1}(i)\left[\frac{P^2}{P\sum_1 \Pi_{1,m}(i) + p\sum_2 \Pi_{1,m}(i)} + \ldots \right.$$

$$\left. \ldots + \frac{p^2}{p\sum_1 \Pi_{1,m}(i) + P\sum_2 \Pi_{1,m}(i)}\right]$$

formule dans laquelle

$\sum_1$ désigne la somme étendue à l'ensemble des signaux pour lesquels l'élément binaire (i + 1) est le même que celui du signal 1 ;

$\sum_2$ désigne la somme complémentaire.

De même :

$$\Pi_{1,m}(i+1) = \Pi_{1,m}(i)\left[\frac{P^2}{P\sum_1 \Pi_{1,m}(i) + p\sum_2 \Pi_{1,m}(i)} + \ldots \right.$$

$$\left. + \frac{p^2}{p\sum_1 \Pi_{1,m}(i) + P\sum_2 \Pi_{1,m}(i)}\right]$$

ou

$$\Pi_{1,m}(i+1) = \Pi_{1,m}(i)\left[\frac{pP}{P\sum_1 \Pi_{1,m}(i) + p\sum_2 \Pi_{1,m}(i)} + \ldots \right.$$

$$\left. + \frac{pP}{p\sum_1 \Pi_{1,m}(i) + P\sum_2 \Pi_{1,m}(i)}\right]$$

suivant que le signal m appartient à l'ensemble $\sum_1$ ou à l'ensemble $\sum_2$.

On peut dans ces formules remplacer 1 par n'importe quel indice q et on a ainsi la possibilité de calculer $\Pi_{q,m}$.

On peut constater que :

. si les signaux sont équiprobables au départ, et

. si les probabilités de détection correcte des deux impulsions élémentaires sont les mêmes, on a

$$\Pi_{ij} = \Pi_{ji}$$

et la matrice $\left[\Pi_{ij}\right]$ demeure symétrique.

Nous supposerons ces conditions remplies dans la suite, bien qu'elles ne soient pas obligatoires.

II. - PREMIERE APPLICATION : CALCUL DES MATRICES DE CONFUSION

Considérons un ensemble de signaux, formés par des trains d'impulsions binaires. Par exemple, il peut s'agir de l'ensemble des signaux d'un code détecteur d'erreurs ou correcteur d'erreurs. S'il y a n signaux en tout, on appelle P_i la probabilité de réception correcte du signal de rang i et $q_{ji} = q_{ij}$ la probabilité de confusion entre les signaux i et j. La matrice

$$\begin{bmatrix} P_1 & q_{12} \cdots\cdots & q_{1n} \\ q_{21} & P_2 \cdots\cdots & q_{2n} \\ & & \\ q_{n1} & \cdots\cdots\cdots\cdots & P_n \end{bmatrix}$$

dite matrice de confusion, et dont la somme de chaque ligne ou de chaque colonne est égale à 1, exprime l'efficacité de la détection de l'ensemble des signaux.

On peut l'évaluer quand on connaît la matrice des quantités $\| \Pi_{ij} \|$ précédemment définies.

Supposons que le signal 1 ait été envoyé, et qu'on ait calculé, dans des conditions de bruit données, la colonne des éléments Π_{11} ..... Π_{1n}.

Prenons d'abord l'élément Π_{11}, relatif au signal 1. Si le détecteur décide "1 a été envoyé" (décision correcte) il prend cette décision avec une probabilité P_1 et il attribue au signal 1 la probabilité d'existence P_1 ; d'où une contribution P_1^2. Si le détecteur décide "2 a été envoyé" (décision fausse) il prend cette décision avec une probabilité q_{12} et il attribue au signal 1 la probabilité d'existence q_{12} . . . etc. On a donc l'égalité :

$$P_1^2 + q_{12}^2 + \ldots + q_{1n}^2 = \Pi_{11}$$

Prenons ensuite l'élément Π_{12}, relatif au signal 2. Si le détecteur décide "1 a été envoyé" (décision correcte) il prend cette décision avec une probabilité P_1 et il attribue au signal 2 une probabilité d'existence q_{12}, d'où une contribution $P_1 q_{12}$. Si le détecteur décide "2 a été envoyé" (décision fausse) il prend la décision avec une probabilité q_{12} et attribue au signal 2 une probabilité d'existence P_2, d'où une contribution $q_{12} P_2$. S'il décide "3 a été envoyé" (décision fausse) il prend la décision avec une probabilité q_{13} et attribue au signal 2 une probabilité d'existence q_{23}... etc On a alors l'égalité :

$$P_1 q_{12} + q_{12} P_2 + q_{13} q_{23} + q_{14} q_{24} + \ldots + q_{1n} q_{2n} = \Pi_{12}$$

On peut donc écrire un système d'équations à partir de la colonne (Π_{1i}) des probabilités d'estimation, le système est :

$$\left[\begin{array}{l} P_1^2 + q_{12}^2 + \ldots + q_{1n}^2 = \Pi_{11} \\ P_1 q_{12} + q_{12} P_2 + q_{13} q_{23} + \ldots + q_{1n} q_{2n} = \Pi_{12} \\ P_1 q_{13} + q_{12} q_{23} + q_{13} P_3 + \ldots + q_{1n} q_{3n} = \Pi_{13} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ P_1 q_{1n} + q_{12} q_{2n} + \ldots + q_{1n} P_n = \Pi_{1n} \end{array}\right.$$

On peut écrire autant de systèmes analogues qu'il y a de colonnes dans la matrice Π c'est à dire n (avec des équations surabondantes, bien entendu).

Ce sont des systèmes du second degré, mais ils sont très faciles à résoudre numériquement. En effet, on utilise des approximations successives dont la première s'obtient en annulant les termes quadratiques en q :

$$(P_1)_0^2 = \Pi_{11}$$

$$(q_{12})_0 \left[(P_1)_0 + (P_2)_0 \right] = \Pi_{12}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

etc.

L'approximation suivante s'obtient en remplaçant les termes quadratiques en q par les valeurs calculées précédemment $(q_{ij})_0$ $(q_{ik})_0$ et on converge d'autant plus vite que les quantités q_{ij} sont plus faibles.

On peut citer comme application possible, à part l'étude des codes, le calcul des erreurs de réception d'un signal dépendant d'un paramètre (par exemple le temps) quand ce paramètre est quantifié. On prend comme signaux 1, 2 ... n les différents signaux correspondant à n valeurs du paramètre et on agira comme précédemment. Il n'existe pas de solution différente à ce problème quand le rapport signal/bruit est élevé.

Traitement d'un exemple numérique.

Prenons l'ensemble des trois signaux à 8 bits :

(signal 1)	1	1	1	1	1	1	1	1
(signal 2)	0	1	1	0	1	0	1	1
(signal 3)	1	0	0	1	0	0	0	0

Supposons que la probabilité de détection correcte du bit soit P = 0,6, ce qui correspond à de très mauvaises conditions de réception et que les probabilités a priori d'existence des trois signaux soient les mêmes : $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$.

On calcule la matrice Π_{ij} (8).

Par exemple :

$$\Pi_{11}\ (1) = \frac{1}{3} \left[\frac{0{,}6^2}{0{,}6 \times \frac{2}{3} + 0{,}4 \times \frac{1}{3}} + \frac{0{,}4^2}{0{,}6 \times \frac{1}{3} + 0{,}4 \times \frac{2}{3}} \right] = 0{,}33929$$

$$\Pi_{12}(1)=\frac{1}{3}\left[\frac{0{,}6\times 0{,}4}{0{,}6\times\frac{2}{3}+0{,}4\times\frac{1}{3}}+\frac{0{,}6\times 0{,}4}{0{,}6\times\frac{1}{3}+0{,}4\times\frac{2}{3}}\right]=0{,}32142$$

$$\Pi_{12}(1)=\frac{1}{3}\left[\frac{0{,}6^2}{0{,}6\times\frac{2}{3}+0{,}4\times\frac{1}{3}}+\frac{0{,}4^2}{0{,}6\times\frac{1}{3}+0{,}4\times\frac{2}{3}}\right]=0{,}33929$$

et ainsi de suite.

On trouve ainsi la matrice $\Pi_{ij}(8)$

0,39813	0,32853	0,27334
0,32853	0,41115	0,26032
0,27334	0,26032	0,46634

et on écrit les équations :

$P_1^2 \div q_{12}^2 + q_{13}^2 = 0{,}39813$

$P_1\, q_{12} + q_{12}\, P_2 + q_{13}\, q_{23} = 0{,}32853$

$P_1\, q_{13} + q_{12}\, q_{23} + q_{13}\, P_3 = 0{,}27334$

$q_{21}^2 + P_2^2 + q_{23}^2 = 0{,}41115$

$q_{21}\, q_{31} + P_2\, q_{23} + q_{23}\, P_3 = 0{,}26032$

$q_{31}^2 + q_{32}^2 + P_3^2 = 0{,}46634$

(en ne gardant que les équations distinctes).

On a alors $(P_1)_0 = 0{,}631$ $\quad(q_{12})_0 = 0{,}250$

$(P_2)_0 = 0{,}641$ $\quad(q_{13})_0 = 0{,}210$

$(P_3)_0 = 0{,}683$ $\quad(q_{23})_0 = 0{,}200$

et on converge aisément vers la solution :

$P_1 = 0{,}540$	$q_{12} = 0{,}267$	$q_{13} = 0{,}193$
$q_{12} = 0{,}267$	$P_2 = 0{,}557$	$q_{23} = 0{,}176$
$q_{13} = 0{,}193$	$q_{23} = 0{,}176$	$P_{23} = 0{,}631$

on vérifie aisément que la somme des éléments de chaque ligne est égale à 1.

III. - DEUXIEME APPLICATION : CONSTRUCTION DE CODES

Elle dérive facilement de l'application précédente. Considérons un code formé de N signaux et le problème de voir ce code transporter l'information maximale en présence d'un niveau de bruit donné et de signaux équiprobables par exemple.

Supposons le code construit jusqu'au rang i, c'est à dire les i premiers bits de chaque signal définis. Les quantités $\Pi_{qm}(i)$ sont alors calculables. Il y a 2^N façons de déterminer le bit de rang (i + 1). A chacune d'elles correspond une matrice $\left[\Pi_{q,m}(i+1)\right]$ et des probabilités de détection correcte $P_1(i+1)\ldots P_N(i+1)$.

La quantité d'information transmise est de la forme :

$$-\sum_1^N \left(\frac{1}{N}\,\mathrm{Log}\,\frac{1}{N} - P_m\,\mathrm{Log}\,P_m\right)$$

On choisit parmi les 2^N combinaisons possibles celle qui la maximise et on franchit un cran de plus.

Il est donc possible de déterminer ainsi un code de longueur arbitraire, par des opérations numériques longues mais faciles à effectuer.

IV. - TROISIEME APPLICATION : EXTENSION DE LA METHODE DE KALMAN-BUCY

Cette méthode est employée à l'estimation récurrente d'un processus aléatoire markovien, quand il est mesuré avec des erreurs d'observations indépendantes à chaque mesure. On y fait des hypothèses simples sur le comportement du processus markovien : il est supposé être la somme d'un processus déterministe et d'un bruit centré. On passe d'une estimation à la suivante par une opération.

a) de transposition ; l'estimation est transposée dans le temps en supposant que le processus est déterministe (gouverné par des équations linéaires) ;
b) de combinaison de la valeur transposée et de la valeur mesurée. On peut étendre le procédé à un processus markovien plus

général. Supposons que le vecteur d'état définissant le système soit à valeurs discrètes (1 à N). Le système se comporte alors comme une chaîne de Markov, définie par la matrice de transition $\left[P_{ij}\right]$. Par ailleurs, on connaît les probabilités conditionnelles $p(m_s/m)$ de mesurer un point m_s quand la valeur réelle est m.

En partant d'un premier vecteur colonne de probabilités p_0 (1) ... p_0 (N) on élabore successivement les vecteurs colonnes p_1 (1) ...p_1(N) ... puis p_i(1) ... p_i (N) par les opérations que nous allons expliciter.

Considérons alors un système subissant une évolution aléatoire, à des instants discrets. Son état est caractérisé à chaque instant par un nombre n (1 à N). A un instant d'échantillonnage i, on connaît le vecteur colonne des probabilités attachées aux diverses valeurs de n :

$$p_i\ (1) \ldots p_i\ (n) \ldots p_i\ (N) \quad \text{ou} \quad \begin{bmatrix} p_i\ (1) \\ p_i\ (n) \\ p_i\ (N) \end{bmatrix}$$

Cet "état", dans la pratique, désigne évidemment un vecteur dont le nombre de dimensions est suffisant pour qu'on puisse considérer le système défini par lui comme markovien.

On peut alors considérer que le système passe de l'état (n) au temps i à l'état (m) au temps i + 1 avec une probabilité définie par une matrice de transition markovienne $\left[\left[P_{ij}\right]\right]$:

$$\begin{bmatrix} q_{i+1}(1) \\ q_{i+1}(N) \end{bmatrix} = \begin{bmatrix} P_{11} & \ldots & P_{1N} \\ P_{N1} & \ldots & P_{NN} \end{bmatrix} \begin{bmatrix} p_i\ (1) \\ p_i\ (N) \end{bmatrix}$$

Ce qui signifie : la connaissance du système à l'instant i est résumée dans le vecteur colonne $\left[p_i\ (n)\right]$. Cette connaissance se trouve transposée dans le vecteur colonne $\left[q_{i+1}\ (n)\right]$ à l'instant i + 1.

Or à l'instant (i + 1) on dispose d'une mesure du vecteur d'état, et le résultat de cette mesure est m_s. Elle a été effectuée avec une erreur d'observation dont on connaît la loi de probabilité. On désignera par $\omega\ (m_s/m)$ la probabilité conditionnelle de voir le résultat de mesure donner m_s quand le système se trouve effectivement dans l'état m.

Le problème est d'estimer l'état du système, ou plus exactement d'évaluer le vecteur colonne des probabilités que le système se trouve dans les états 1 ...N à l'instant i + 1 quand on dispose du vecteur colonne transposé $\left[q_{i+1}\ (n)\right]$ et de la mesure m_s.

On aura évidemment le vecteur colonne fourni par l'application de la règle de Bayes :

$$p_{i+1}(m) = q_i\ (m)\ \frac{\omega\ (m_s/m)}{\sum_1^N \omega\ (m_s/m) p_{i+1}(m)} \qquad (2)$$

et l'application des transitions markoviennes (1) et (2) au vecteur colonne $\left[p_i\ (n)\right]$ permet de trouver le vecteur colonne homologue $\left[p_{i+1}\ (n)\right]$. L'opération est alors itérée et on a, par récurrence, les vecteurs colonnes successifs définissant les probabilités des états du système au cours du temps.

On peut alors définir une probabilité d'estimation relative à l'état vrai m_0 (i + 1)

$$\Pi_{i,i+1}\ (m) = q_i(m) \sum_{m_s=1}^{m_s=N} \frac{\omega\ (m_s/\ m_0)\ \omega\ (m_s/\ m)}{\sum_{m=1}^{N} \omega\ (m_s/m)\ p_{i+1}\ (m)}$$

et, par voie de conséquence, une probabilité d'estimation relative à une "trajectoire" donnée : T : m_0 (1) ... m_0 (i), m_0 (i + 1))... et une trajectoire estimée quelconque : $\hat{T}$: m (1) ... m (i), m (i + 1).

La grandeur Π_i (T, $\hat{T}$) signifiera : probabilité d'estimation relative à une trajectoire vraie donnée T et une trajectoire $\hat{T}$, au temps i (entier) en partant d'une situation connue (certaine ou aléatoire) au temps 0.

Cette grandeur se calcule par récurrence. Quand on connaît les grandeurs Π_i (T, $\hat{T}$) cela signifie qu'à l'instant i, toutes les grandeurs Π_i (m_0, m) sont connues. (m_0 est l'état de la trajectoire T, m l'état de la trajectoire $\hat{T}$ à l'instant i).

L'application des formules 1 et 2, où on remplace les p_i (m) par les Π_i (m_0, $\hat{m}$) fournit l'ensemble des grandeurs Π_{i+1} (T, $\hat{T}$).

Le nombre immense des trajectoires possibles interdit de procéder comme au paragraphe précédent.

Cependant, on peut opérer toujours par récurrence pour étudier une trajectoire donnée. On se donne le vecteur colonne des probabilités initiales $\left[p_0 (i) \right]$ et, connaissant l'état m_0 à l'instant 1 de la trajectoire vraie, on en déduit les probabilités d'estimation $\Pi_1 (m_0, m)$. On en déduit aussi la matrice de confusion par le procédé décrit au paragraphe précédent.

Comme on dispose alors des probabilités d'estimation $\Pi_1 (m_0, m)$, on repart de ce vecteur colonne pour franchir un nouveau cran et on étudie ainsi la trajectoire de proche en proche.

Il est bien évident que le caractère récurrent du procédé permet de changer à chaque échantillonnage la matrice de transition $\left[P_{ij} \right]$ de la trajectoire ainsi que de la loi des erreurs d'observation $\omega \ (m_s / m)$.

On peut dire en conclusion que l'application de la théorie de la communication a beaucoup à gagner dans le calcul de quantités récurrentes toutes les fois que cela est possible, en raison des commodités très grandes qui s'attachent à ce genre d'évaluation. Ainsi l'utilisation d'une voie détournée est avantageuse pour calculer des quantités très difficiles à élaborer autrement. Il reste certainement beaucoup à faire dans ce domaine ; la théorie de la communication est toujours une science très récente.