HUITIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 1er au 5 JUIN 1981

RELATION ENTRE LA MESURE DE DIVERGENCE ET LA PERFORMANCE RELATIVE D'ESTIMATEURS-CORRELATEURS.

A. Fogel

Faculty of Electrical Engineering, TECHNION - Israel Institute of Technology, Technion City, Haifa 32000, Israel.

RESUME

La convergence d'estimateurs-corrélateurs pour trois types de problèmes de detection, est examinée. Il est demontré que la convergence d'estimateurs au sens de la divergence entraîne la convergence des performances mesurées soit en probabilité d'erreur soit en probabilité de détection, qui sont associées aux détecteurs construits à partir des estimateurs.

On démontre en particulier que dans le cas de la détection d'un signal aléatoire dans du bruit gaussien blanc ou bien dans celui de la détection d'un processus de Poisson doublement stochastique, la distance entre estimateurs qui convient, est l'intégrale de la divergence entre estimateurs.

SUMMARY

Convergence of estimator-correlators is investigated for three classes of detection problems. It is shown that convergence of detectors in probability of error or probability of detection, is implied by convergence of estimates in the sense of the divergence. In particular, the proper distance measure between the estimators when detecting a random signal in white Gaussian noise, or a Doubly Stochastic Poisson Process, is shown to be the integrated divergence between the estimates.

Relation entre la Mesure de Divergence et la Performance Relative d'Estimateurs-Correlateurs.

A. Fogel.

I. Introduction

C'est un fait établi que pour un certain nombre de problèmes de détection, le rapport de vraisemblance généralisé se présente sous une structure d'"estimateur-corrélateur", dans laquelle on effectue une corrélation entre l'observation et l'estimateur des moindres carrés (EMC) du signal à détecter. Cette propriété se retrouve dans la détection de signaux aléatoires dans du bruit blanc gaussien ([1-2]), dans la détection de processus aléatoires ponctuels ([3-4]) et en temps discret lorsque la loi de distribution des observations appartient à la famille exponentielle ([5-6]).

Ces détecteurs, optimum parce qu'ils minimisent la probabilité d'erreur (PE) ou bien parce qu'ils produisent une probabilité de détection (PD) maximale pour une probabilité de fausse alarme (PFA) donnée, ne se prêtent pas en général à une réalisation pratique par manque d'information a priori sur le signal, ce qui rend l'EMC inaccessible. En revanche, il est possible et naturel de réaliser des détecteurs sous-optimaux à partir d'estimateurs réalisables que l'on substitue dans la structure optimale à la place de l'EMC. Une question importante qui se pose alors est d'établir la performance relative d'un tel test mesurée en PE ou PD, vis-à-vis du détecteur optimal pour un certain écart (qui demande à être défini) entre l'EMC et l'estimateur qu'on lui substitue.

Une distance entre estimateurs a déjà été établie ([7]) qui assure que les détecteurs convergent lorsque leurs performances sont évaluées par les bornes de Chernoff et lorsque la distance tend vers zero.

Dans cet article, la convergence des détecteurs est abordée directement sous le critère de la PE et sous le point de vue de la convergence stochastique de mesures. Il sera démontré que pour les trois catégories de problèmes ci-dessus cités, la convergence des détecteurs au sens de la PE est impliquée par la convergence des estimateurs au sens de la divergence ([8]). **Il sera également observé que les distances** obtenues en ([7]) se révèlent être des approximations du premier ordre de celles obtenues à partir de la divergence.

II. Convergence Stochastique de Mesures

A. Définitions

Considérons un espace mesurable $(\Omega,\mathcal{F})$ auquel est adjoint un ensemble $\mathcal{P}$ de mesures de probabilité absolument continues par rapport à une mesure σ-finie $\lambda(x)$. Deux modes de convergence d'une série de mesures de probabilité P_n appartenant à $\mathcal{P}$, **seront** utilisés dans la suite de cet article:

1. Convergence uniforme de P_n à P ([9]). Définition:

   $P_n \xrightarrow[n\to\infty]{} P$ si et seulement si

   $$\lim_{n\to\infty} \sup_{A\in\mathcal{F}} |P_n(A) - P(A)| = 0 .$$

2. Convergence au sens de la divergence. Définition: soient P et Q deux mesures de probabilité. On définit la divergence entre P et Q de la façon suivante ([9-10]):

   $$I(P,Q) = \int p(\omega) \log \frac{p(\omega)}{q(\omega)}\, d\lambda(\omega) \quad , \tag{1}$$

où $p(\omega)$ et $q(\omega)$ représentent les derivées de Radon-Nikodym de P et Q par rapport à $\lambda(\omega)$.

Définition: la sèrie P_n converge vers P au sens de la "divergence" si et seulement si

$$\lim_{n\to\infty} I(P_n,P) = 0 \quad .$$

Le concept de divergence a été développé dans le context de la théorie de l'information et celui de l'entropie en thermodynamique. S. Kullback ([10]) en a établi une étude très complète, mais récemment la divergence a trouvé un intérêt nouveau dans la théorie de l'estimation ([11,12]). A cet égard, nous verrons que la divergence lie simultanément estimation et détection.

B. Relation entre les Modes de Convergence

Théorème ([9, p. 57]):

Convergence au sens de la divergence (CD) entraîne la convergence uniforme (CU).

Ce théorème se révélera très utile lorsque l'on démontrera que la convergence de détecteurs est assurée par la convergence uniforme de mesures.

III. Estimateurs-Corrélateurs Optimaux et Sous-Optimaux

A. Estimateur-Corrélateur

Pour un certain nombre de problèmes de détection, il a été démontré que le rapport de vraisemblance (RV)' optimum parce qu'il minimize la PE ou maximize la PD pour une PFA donnée, **se** présente sous une structure d'estimateur-corrélateur qui effectue simultanément le calcul de l'EMC du signal ainsi que le processus de décision quant à la présence de ce signal. En particulier, nous considérons les trois catégories de problèmes suivants:

1. Détection d'un signal aléatoire dans du bruit blanc gaussien.

   Sous l'hypothèse H_0, P_0 est choisie comme étant la mesure de Wiener sur $(\Omega,\mathcal{F})$, ce qui revient à dire que l'observation $\{x(t),\ 0 \leq t \leq T\}$, est un mouvement brownien ("intégrale" de bruit blanc gaussien).

   Sous l'hypothèse H_1, un signal aléatoire $\{\theta(t),\ 0 \leq t \leq T\}$ est ajouté au bruit de sorte que l'on définit une mesure P_1 telle que

   $x(t) - \int_0^t (s)ds$ est un mouvement brownien. On démontre alors que le RV généralisé est une fonctionnelle de l'EMC de $\theta(t)$, à savoir ([1],[13]):

   $$\hat\theta(t) \triangleq E_{H_1}[\theta(t)|x(s),\ o \leq s \leq t] \quad ; \tag{3}$$

   nous avons alors

   $$L(\hat\theta) = E_{H_0}^{\mathcal{F}_T} \frac{dP_1}{dP_0} = \exp\left[\int_0^T \hat\theta(t)dx(t) - \frac{1}{2}\int_0^T \hat\theta^2(t)dt\right], \tag{4}$$

   où $\mathcal{F}_T$ représente la sous algèbre générée par $\{x(s),\ o \leq s \leq T\}$. Il est souvent pratique d'introduire le logarithme de $L(\hat\theta)$, à savoir

   $$\ell(\hat\theta) = \log L(\hat\theta) \quad . \tag{5}$$

Relation entre la Mesure de Divergence et la Performance Relative d'Estimateurs-Correlateurs.

A. Fogel

Le test optimum revient à comparer $\ell(\hat{\theta})$ à un seuil γ, c'est à dire

$$\ell(\hat{\theta}) \underset{H_o}{\overset{H_1}{\gtrless}} \gamma \quad , \tag{6}$$

et γ prend une valeur qui dépend du critère de décision employé (Bayes ou Neyman-Pearson).

Du fait que $\hat{\theta}(t)$ n'est pas calculable en général, il est naturel de lui substituer une approximation dans la structure optimale (6) et d'obtenir ainsi le détecteur sous-optimal:

$$\ell(\hat{\theta}) = \int_o^T \tilde{\theta}(t)dx(t) - \frac{1}{2}\int_o^T \tilde{\theta}^2(t)dt \underset{H_o}{\overset{H_1}{\gtrless}} \gamma. \tag{7}$$

2. Détection de Processus de Poisson doublement stochastique (PPDS)

Dans cette situation, on veut différencier les deux hypothèses suivantes:

H_o: $\{N(t),\ o \leq t \leq T\}$, l'observation, est un processus de Poisson homogène à fonction d'intensité $\lambda_o(t)$.

H_1: $\{N(t),\ o < t < T\}$ est un PPDS à processus d'intensité $\lambda_1(x(t))$, $x(t)$ étant lui-même un processus stochastique. On démontre alors que le LRV prend la forme ([3],[4]):

$$\ell(\hat{\lambda}) \triangleq -\int_o^T (\hat{\lambda}(t) - \lambda_o(t))dt + \int_o^T \log\frac{\hat{\lambda}(t)}{\lambda_o(t)}dN(t) \underset{H_o}{\overset{H_1}{\gtrless}} \gamma, \tag{8}$$

où

$$\hat{\lambda}(t) = E_{H_1}[\lambda_1(x(t))|N_\sigma,\ o \leq \sigma \leq t] \quad . \tag{9}$$

Comme précédemment, on veut comparer la performance relative à $\ell(\hat{\lambda})$ de tests sous-optimaux de la forme

$$\ell(\tilde{\lambda}) \underset{H_o}{\overset{H_1}{\gtrless}} \gamma \quad , \tag{10}$$

obtenus en substituant $\tilde{\lambda}$ à $\hat{\lambda}$ dans (8).

3. Estimateur-corrélateur en temps discret.

Nous supposons que les observations faites à intervalles discrets sont distribuées suivant un membre de la famille exponentielle[1].

n observations indépendantes et distribuées identiquement sont gouvernées par la densité de probabilité:

$$p(x_i|\theta) = \exp(\theta x_i - b(\theta) + h(x_i)) \quad . \tag{11}$$

Sous H_o, θ est une constante connue θ_o, tandis que sous H_1, θ est une variable aléatoire dont la distribution a priori est désignée par $\pi(\theta)$. Le problème admet une statistique $t = \sum_{i=1}^{n} x_i$ ([14], p. 126) dont la distribution appartient également à la famille exponentielle et s'écrit:

$$p(t|\theta) = \exp(\theta t - nb(\theta) + B(t)) \quad . \tag{12}$$

On demontre alors ([5],[7]) que le LRV s'écrit

$$\ell(\hat{\theta}) = \int_{t_o}^{t} (\hat{\theta}(u) - \theta_o)du + C(\hat{\theta}) \underset{H_o}{\overset{H_1}{\gtrless}} \gamma \; , \tag{13}$$

où t_o est arbitraire, $\hat{\theta}(t)$ est l'EMC de θ (t désignant ici la statistique suffisante) et $C(\hat{\theta})$ est une fonctionnelle de $\hat{\theta}$. Comme précédemment, nous sommes interessés à réaliser des tests sous-optimaux en remplaçant $\hat{\theta}$ par une approximation $\tilde{\theta}$.

B. Convergence d'Estimateurs-Corrélateurs et Convergence Stochastique de Mesures.

Dans les trois catégories de problèmes décrits auparavant, on se doit d'évaluer la performance de détecteurs sous-optimaux du type (7) relativement au détecteur optimal (6).[2] Cela revient à dire que l'on pose la question de la convergence de $\ell(\tilde{\theta})$ vers $\ell(\hat{\theta})$ mesurée en PE ou PD lorsque $\tilde{\theta}$ tend vers $\hat{\theta}$ sous un critère à définir. Pour ce faire, on se restraint aux estimateurs $\tilde{\theta}(t)$ pour lesquels il existe une distribution à priori vis-à-vis de laquelle $\tilde{\theta}(t)$ est l'EMC de $\theta(t)$. En d'autres termes, $\tilde{\theta}(t)$ induit une mesure $\tilde{P}$ sur (Ω,F) vis-à-vis de laquelle $\tilde{\theta}(t)$ est l'EMC de $\theta(t)$. De tels estimateurs existent ([8],[13]).

Par conséquent, poser la question de la convergence de $\ell(\tilde{\theta})$ vers $\ell(\hat{\theta})$ revient à poser la convergence de $\tilde{P}$ vers P.

IV. Convergence Stochastique (CS) et la Divergence

A. Convergence Uniforme des LRV

On considère une série d'estimateurs θ_n tels que les mesures induites P_n tendent vers P (pour laquelle on se souvient que $\hat{\theta}$ est l'EMC) uniformément selon le critère décrit en II.A.2. On démontre alors tres facilement ([8]), que dans la situation de Bayes, la PE associée avec $\ell(\theta_n)$ tend uniformément vers celle associée avec $\ell(\hat{\theta})$, et que l'on obtient un résultat similaire pour les PD dans la situation de Neyman-Pearson.

B. Convergence en "Divergence"

On suppose maintenant que

$$\lim_{n\to\infty} I(P_n,P) = 0 \quad . \tag{14}$$

Le théorème énoncé en II.B et la discussion précédence assurent alors que $\ell(\theta_n)$ tend vers $\ell(\hat{\theta})$ uniformément en PE ou PD pour une PFA donnée.

C. Convergence des Estimateurs

On en vient maintenant à la question de déterminer une distance entre estimateurs qui assure la convergence (14). Nous devons alors différencier entre les trois cas étudiés.

1. Le cas gaussien (voir (4)).

On définit l'hypothèse H^n comme étant l'alternative à P_o lorsque cette alternative est gouvernée par P_n.

[1]. Cette famille comprend en particulier les distributions Binomiale, Beta, Poisson, Gamma, et Gaussienne.

[2]. Dans le cas des processus de Poisson, on doit remplacer $\tilde{\theta}$ et $\hat{\theta}$ par les intensités $\tilde{\lambda}$ et $\hat{\lambda}$, respectivement.

On peut alors écrire

$$I(P_n,P) = E_{H^n}\left[\log\left(\frac{P_n}{P_o}\right) - \log\left(\frac{P}{P_o}\right)\right] \tag{15}$$

de telle sorte que utilisant (4), on a

$$I(P_n,P) = $$
$$= E_{H^n}\left[\int_o^T (\theta_n(t)-\hat\theta(t))dx(t) - \frac{1}{2}\int_o^T (\theta_n^2(t)-\hat\theta^2(t))dt\right] . \tag{16}$$

Le théorème de Girsanov ([13]) permet alors d'écrire

$$dx(t) = \theta_n(t)dt + dW(t) \quad , \tag{17}$$

où $W(t)$ est un mouvement brownien. La substitution de (17) dans (16) entraîne

$$I(P_n,P) = E_{H^n}\left[\int_o^T (\theta_n(t)-\hat\theta(t))dW(t)\right] + $$
$$+ \frac{1}{2}\int_o^T E_{H^n}(\theta_n(t)-\hat\theta(t))^2 dt \quad . \tag{18}$$

Mais le premier terme de (18) est la moyenne d'une intégrale stochastique de Ito et par conséquent est égal à zéro.

Il résulte que nou pouvons définir une distance entre les estimateurs θ_n and $\hat\theta$, à savoir

$$||\theta_n-\hat\theta||^2 = \frac{1}{2}\int_o^T E_{H^n}(\theta_n(t)-\hat\theta(t))^2 dt \quad , \tag{19}$$

telle que

$$I(P_n,P) = ||\theta_n-\hat\theta||^2 \quad . \tag{20}$$

De la discussion en IV.2.B., il ressort donc que

$$||\theta_n-\hat\theta||^2 \underset{n\to\infty}{\longrightarrow} 0 \Longrightarrow \ell(\theta_n) \underset{n\to\infty}{\longrightarrow} \ell(\hat\theta) \quad , \tag{21}$$

et la convergence est uniforme en PE ou PD.

2. Le cas Poisson.

Un développement similaire au précédent ([8]), conduit à déterminer la divergence entre deux variables aléatoires de Poisson d'intensités λ_1 et λ_2 comme

$$I(\lambda_2,\lambda_1) = E_{\lambda_2}\log\frac{P_2}{P_1} = \lambda_2 \log\frac{\lambda_2}{\lambda_1} - (\lambda_2-\lambda_1) \quad , \tag{22}$$

et à définir la distance entre estimateur comme

$$||\lambda_n-\hat\lambda||^2 = \int_o^T E_{H^n}(I(\lambda_n,\hat\lambda)dt \quad , \tag{23}$$

de sorte que

$$||\lambda_n-\hat\lambda|| \underset{n\to\infty}{\longrightarrow} 0 \Longrightarrow \ell(\lambda_n) \underset{n\to\infty}{\longrightarrow} \ell(\hat\lambda) \quad ,$$

uniformément en PE our PD.

Il est facile de démontrer que l'erreur carrée moyenne qui apparaît en (19) a la forme de la divergence entre deux variables aléatoires gaussiennes de sorte que d'après (19) et (23), la distance entre estimateurs qui est compatible avec la détection, est l'intégrale de la divergence entre les estimateurs.

3. Cas du temps discret.

Soit

$$J = \int_{t_o}^t (\theta_n(u) - \hat\theta(u))du \quad . \tag{24}$$

On peut démontrer ([8]) que la distance basée sur la divergence est donnée par

$$||\theta_n-\hat\theta||^2 = E_{Hn}\exp(E_{Hn}(J) - J) \quad . \tag{25}$$

V. Discussion et Conclusion

A. Comparaison avec les Distances Derivées en ([7]).

Des distances entre estimateurs ont été dérivées ([7]) qui assurent que les bornes de Chernoff $B(\theta_n)$ et $B(\hat\theta)$ associées respectivement aux détecteurs sous-optimaux et optimaux, convergent. Ces distances sont de la forme:

1. Cas gaussien

$$||\theta_n-\hat\theta||_c^2 = \int_o^T E_{H_1}(\theta_n(t)-\hat\theta(t)^2 dt \quad . \tag{26}$$

Comparant (19) à (26), on voit que lorsque θ_n tend vers $\hat\theta$, "H^n tend vers H_1", de telle sorte que (26) est une approximation du premier ordre de (19) (sans tenir compte du facteur multiplicateur).

2. Cas de Poisson

La norme derivée en ([7]) était de la forme:

$$||\lambda_n-\hat\lambda||_c^2 = \int_o^T E_{H_1}\frac{(\lambda_n-\hat\lambda)^2}{\hat\lambda(t)}dt \quad . \tag{27}$$

Si l'on suppose λ_n est proche de $\hat\lambda$ dans (23) et que l'on utilise dans (22) l'approximation de log x par $x-1$, on obtient

$$I(\lambda_n,\hat\lambda) \simeq \frac{(\lambda_n-\hat\lambda)^2}{\hat\lambda} \quad , \tag{28}$$

de sorte qu'à nouveau on voit que (27) est une approximation du premier ordere de (23).

3. Cas du temps discret

La distance dérivée dans [7] était de la forme

$$||\theta_n-\hat\theta||_c^2 = \mathrm{var}_{H_1}(J) \quad ,$$

ce que l'on démontre à nouveau être une approximation du premier ordre de (25).

B. Conclusion

Nous avons étudié trois catégories de problèmes de détection dont les détecteurs optimaux se caracterisent par une structure d'estimateur-corrélateur. L'analyse a résulté dans la détermination d'une distance entre estimateurs qui assure que à mesure qu'elle tend vers zéro, les tests induits correspondants convergent uniformément en PE ou PD.

L'étude a produit trois types de mesures basées sur la divergence, deux d'entre elles s'adressant au temps continu et se trouvant égales à l'intégrale de la divergence entre les estimateurs.

References

[1] Kailath, T. "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise", IEEE Trans. Inform. Theory, Vol. IT-15, pp. 350-361, May 1969.

[2] Schweppe, F.C. "Evaluation of Likelihood Functions for Gaussian Signals", IEEE Trans. Inform. Theory, Vol. IT-11, pp. 61-70, Jan. 1965.

[3] Snyder, D.L. "Filtering and Detection for Doubly Stochastic Poisson Processes", IEEE Trans. Inform. Theory, Vol. IT-18, pp. 91-102, Jan. 1972.

[4] Rubin, I. "Regular Point Processes and Their Detection", IEEE Trans. Inform. Theory, Vol. IT-18, pp. 547-557, Sept. 1972.

[5] Schwartz, S.C. "Conditional Mean Estimates and Bayesian Hypothesis Testing", IEEE Trans. Inform. Theory, Vol. IT-21, pp. 663-665, Nov. 1975.

[6] Schwartz, S.C. "The Estimator-Correlator for Discrete-Time Problems", IEEE Trans. Inform. Theory Vol. IT-23, pp. 93-100, Jan. 1977.

[7] Fogel, A., and Schwartz, S.C. "Bounds on Probability of Error and Suboptimal Detectors", submitted to IEEE Trans. Inform. Theory. Also Princeton University Tech. Rep. No. 40, May 1978.

[8] Fogel, A. "On the Divergence Information Measure and the Performance Convergence of Estimator-Correlators", Technion EE Pub. No. 379, Haifa, Israel, June 1980. Submitted to IEEE Trans. on Information Theory.

[9] Luckacs, Stochastic Convergence, A.P., 1975.

[10] Kullback, S. Information and Statistics, Dover, 1968.

[11] Kashyap, R.L. "Minimax Estimation with Divergence Loss Function", Information Sciences 7, pp. 341-364, 1974.

[12] Johnson, R. "Axiomatic Characterization of the Directed Divergences and their Linear Combinations", IEEE Trans, on Inform. Theory, Vol. IT-25, No. 6, pp. 709-716, Nov. 1979.

[13] Wong, E. Stochastic Processes in Information and Dynamical Systems, McGraw-Hill, 1971.

[14] Ferguson, T., Mathematical Statistics, Academic Press, 1967.