

Un estimateur récursif de l'entropie

Christophe VIGNAT¹, Jean-François BERCHER^{1,2}

¹Laboratoire Systèmes de Communication,
Université de Marne-la-Vallée,
93 166 Noisy-le-Grand, FRANCE

²Signaux et Télécoms,
ESIEE-Paris

93 162 Noisy-le-Grand, FRANCE

vignat@univ-mlv.fr, bercherj@esiee.fr

Résumé – Nous présentons dans cet article un estimateur récursif de l'entropie instantanée d'un processus aléatoire stationnaire. Cet estimateur repose sur une utilisation particulière d'un histogramme des données. Dans un premier temps, nous donnons une interprétation spectrale de l'histogramme. Dans un second temps, nous introduisons un estimateur non paramétrique de l'entropie, caractérisé par une formule de remise à jour récursive très simple, qui rend sa mise en œuvre économique en calcul. De plus, la possibilité d'introduire un facteur d'oubli lui confère, dans un contexte non-stationnaire, une capacité de poursuite. Des simulations montrent l'intérêt de cet estimateur dans les contextes de détection de ruptures et de séparation de sources.

Abstract – We present in this paper a recursive estimator for the instantaneous entropy of a stationary random process. This estimator is based on a particular use of a data histogram. First, we give a spectral interpretation of data histograms. Second, we introduce a non-parametric estimator of the entropy, characterized by a very simple recursive updating formula, that enables a low cost implementation. Moreover, the possibility of introducing a forgetting factor gives it, in a non-stationary context, an adaptive capacity. Simulations illustrate the interest of such an estimator in applications such as abrupt law changes detection and sources separation.

1 Introduction

L'APPLICATION de la théorie de l'information au traitement du signal connaît, depuis ces dernières années, un succès croissant. L'entropie, notamment, est utilisée avec profit dans des contextes tels que l'égalisation aveugle ou la séparation de sources. Plus généralement, elle autorise le traitement de problèmes pour lesquels les statistiques d'ordre deux ou d'ordre supérieur sont insuffisantes ou d'utilisation malaisée.

Rappelons que l'entropie H_X associée à une loi continue $f_X(x)$ s'écrit :

$$H_X = - \int f_X(x) \log f_X(x) dx. \quad (1)$$

Dans le cas où la loi est discrète, l'entropie, notée cette fois-ci H_X , peut être déduite des probabilités $\{p_i\}$ de la façon suivante :

$$H_X = - \sum_{i=-\infty}^{+\infty} p_i \log p_i. \quad (2)$$

Dans la plupart des cas, cette entropie est calculée à partir de l'expression analytique de la loi des données (lorsque cette loi est connue), ou bien d'un modèle paramétrique de cette loi. Une autre approche consiste à construire un estimateur de l'entropie sans la connaissance a priori de la loi des échantillons, et à partir de leur simple observation. On trouve dans [1] une telle approche, basée sur une estimation préalable de la loi à l'aide des densités de Parzen, d'où est déduite une estimée de l'entropie. La méthode exposée dans [2] consiste à construire un estimateur de l'entropie à partir des statistiques d'ordre.

L'approche que nous proposons est basée sur la construction de l'histogramme de la loi : l'expression de l'entropie associée à cet estimateur est très simple, et fait apparaître une relation de récurrence entre les estimateurs calculés à deux instants consécutifs. Nous montrons de plus comment cet estimateur peut aisément être étendu au cas d'une loi continue. En revanche, cette étude est restreinte, comme c'est très majoritairement le cas, à des lois de probabilité à support borné.

2 Estimateur de l'histogramme

On suppose par la suite disposer d'échantillons $X_n(\omega)$ d'une suite aléatoire stationnaire de loi instantanée $f_X(x)$. Cette loi est de plus supposée à support borné $[a, b]$: cette hypothèse, bien que restrictive, est classique dans le domaine de l'estimation de lois de probabilité. Soit $\{\alpha_k\}_{0 \leq k \leq N-1}$ une subdivision uniforme de l'intervalle $[a, b]$ en N intervalles notés $I_k = [\alpha_k, \alpha_{k+1}]_{0 \leq k \leq N-1}$ avec $\alpha_0 = a$ et $\alpha_N = b$. Chaque intervalle est de longueur $h = \alpha_{k+1} - \alpha_k = \frac{b-a}{N}$. Nous présentons d'abord l'estimation classique par histogramme de la loi $f_X(x)$, basée sur le calcul des probabilités empiriques. Puis nous en donnons, avant de l'appliquer à l'estimation de l'entropie, une interprétation inspirée des techniques d'analyse spectrale.

2.1 Estimateur empirique

À l'instant p , disposant de p échantillons $\{X_i(\omega)\}_{1 \leq i \leq p}$ de la suite aléatoire, on note $U_k^{(p)}$ le nombre d'échantillons qui

appartiennent à l'intervalle I_k . On a évidemment la relation de normalisation

$$\sum_{k=0}^{N-1} U_k^{(p)} = p, \quad (3)$$

qui exprime le fait que les N intervalles se partagent l'ensemble des p échantillons disponibles.

Dans le cas où la loi est discrète, un estimateur de cette loi est simplement

$$\hat{f}^{(p)}(x) = \frac{U_k^{(p)}}{p} \text{ si } x \in I_k. \quad (4)$$

Dans le cas d'une loi continue, cet estimateur s'écrit simplement

$$\hat{f}^{(p)}(x) = \frac{U_k^{(p)}}{hp} \text{ si } x \in I_k, \quad (5)$$

où le terme h^{-1} assure la normalisation de la loi.

A l'instant $(p+1)$, un nouvel échantillon $X_{p+1}(\omega)$, tel que $X_{p+1}(\omega) \in I_{l_{p+1}}$, est disponible, et permet d'enrichir l'estimateur de l'histogramme de la façon suivante :

$$\hat{f}^{(p+1)}(x) = \frac{U_k^{(p+1)}}{p+1} \text{ si } x \in I_k, \quad (6)$$

avec

$$U_k^{(p+1)} = \begin{cases} U_k^{(p)} & \text{si } k \neq l_{p+1}, \\ U_k^{(p)} + 1 & \text{si } k = l_{p+1}. \end{cases} \quad (7)$$

2.2 Une interprétation spectrale de l'estimateur

2.2.1 Un processus sous-jacent

Considérons le processus aléatoire $Z_n(\omega)$, appelé par la suite processus sous-jacent de $X_n(\omega)$, et défini de la façon suivante :

$$Z_n(\omega) = \frac{1}{\sqrt{N}} \sum_{k=1}^N e^{j(nX_k(\omega) + \phi_k(\omega))}, \quad (8)$$

où N est un entier fixé, et $\{\phi_k(\omega)\}$ est une suite i.i.d. de phases aléatoires, de loi uniforme sur l'intervalle $[0, 2\pi]$ et indépendante du processus $X_n(\omega)$. Ce type de processus a été introduit par [3] (sous une forme non-stationnaire et non-ergodique) dans un contexte de filtrage non-linéaire, puis indépendamment par [4] dans le contexte d'estimation paramétrique d'une loi. Il est aisé de vérifier que le processus sous-jacent possède les propriétés suivantes :

- $Z_n(\omega)$ est stationnaire au sens large,
- $\sigma_Z^2 = 1$,
- le spectre $S_Z(f)$ de $Z_n(\omega)$ vérifie

$$S_Z(f) = p_X(f) \quad \forall f \in \left[-\frac{1}{2}, +\frac{1}{2}\right].$$

L'estimation du spectre $S_Z(f)$ de $Z_n(\omega)$ permet ainsi d'estimer la loi de $X_n(\omega)$.

2.2.2 Estimation par un banc de filtres

Le filtrage par un filtre passe-bande, idéal dans la bande $I_q = [\alpha_q, \alpha_{q+1}]$, du processus sous-jacent permet, comme nous allons le montrer, d'estimer la densité de probabilité $p_X(f)$ dans

l'intervalle des valeurs $x \in I_q$. Soit en effet un filtre idéal passe-bande \mathcal{H} dont la réponse fréquentielle vérifie

$$H(f) = \begin{cases} 1 & \text{si } f \in I_q \\ 0 & \text{sinon.} \end{cases} \quad (9)$$

La filtrée par \mathcal{H} du processus sous-jacent s'écrit

$$\begin{aligned} \tilde{Z}_n(\omega) &= \sum_p h_p Z_{n-p}(\omega) \\ &= \frac{1}{\sqrt{N}} \sum_p \sum_{k=1}^N h_p e^{j((n-p)X_k(\omega) + \phi_k(\omega))} \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^N \left(\sum_p h_p e^{-jpX_k(\omega)} \right) e^{j(nX_k(\omega) + \phi_k(\omega))} \\ &= \frac{1}{\sqrt{N}} \sum_{k=1}^N H(X_k(\omega)) e^{j(nX_k(\omega) + \phi_k(\omega))} \end{aligned}$$

où $H(X_k(\omega))$ est la réponse du filtre \mathcal{H} à la fréquence $\frac{X_k(\omega)}{2\pi}$. On en déduit d'après (9) que

$$\tilde{Z}_n(\omega) = \frac{1}{\sqrt{N}} \sum_{\substack{k=1 \\ X_k(\omega) \in I_q}}^N H(X_k(\omega)) e^{j(nX_k(\omega) + \phi_k(\omega))}.$$

La puissance du processus $\tilde{Z}_n(\omega)$ s'écrit donc

$$\sigma_{\tilde{Z}}^2 = \frac{1}{N} \sum_{\substack{k=1 \\ X_k(\omega) \in I_q}}^N H(X_k(\omega)) = \frac{U_k^{(N)}}{N},$$

où $U_k^{(N)}$ est le nombre d'échantillons parmi les N valeurs $X_k(\omega)$ pour $1 \leq k \leq N$ tels que $X_k(\omega) \in I_q$. Ainsi, le calcul de l'histogramme de la loi p_X peut être interprété comme la mesure de la puissance, en sortie d'un banc de filtres idéaux passe-bande.

2.3 Application à la construction d'autres estimateurs

L'analogie précédente peut être étendue en remarquant que toute méthode d'estimation du spectre du processus sous-jacent $Z_k(\omega)$ permet de déduire une méthode d'estimation de la loi p_X des données $X_k(\omega)$. A titre d'exemple, la méthode classique du périodogramme fournit, dans le cas du processus sous-jacent, l'estimateur spectral

$$\hat{S}_Z(\Omega) = \left| \frac{1}{P} \sum_{p=1}^P Z_p(\omega) e^{-jp\Omega} \right|^2,$$

où P est le nombre d'échantillons du processus sous-jacent. Remplaçant chacun des échantillons par sa valeur définie par (8), on en déduit

$$\hat{S}_Z(\Omega) = \frac{1}{P^2} \left| \sum_{p=0}^{P-1} \frac{1}{\sqrt{N}} \sum_{k=1}^N e^{j(pX_k(\omega) + \phi_k(\omega))} e^{-jp\Omega} \right|^2.$$

Pour un nombre de données N assez grand, les échantillons $X_k(\omega)$ étant supposés décorrélés entre eux et indépendants des phases $\phi_k(\omega)$, on peut approcher $\hat{S}_Z(\Omega)$ par

$$\hat{S}_Z(\Omega) \simeq \frac{1}{N} \sum_{k=1}^N \Phi_P(X_k(\omega) - \Omega), \quad (10)$$

où la fonction $\Phi_P(x)$ est le noyau de Fejer d'ordre P , défini par

$$\Phi_P(x) = \frac{1}{P^2} \left(\frac{\sin \frac{P}{2}x}{\sin \frac{1}{2}x} \right)^2,$$

et vérifiant $\Phi_P(x) \geq 0$, $\int_{-\infty}^{+\infty} \Phi_P(x) dx = 1$. L'estimateur (10) du spectre du processus sous-jacent est aussi l'estimateur \hat{p}_X de la loi p_X des données, si bien que

$$\hat{p}_X(\Omega) = \frac{1}{N} \sum_{k=1}^N \Phi_P(X_k(\omega) - \Omega).$$

On reconnaît ici l'estimateur par noyaux (de Fejer) de la loi p_X . Ainsi, à l'estimateur par périodogramme du spectre du processus sous-jacent correspond asymptotiquement l'estimation par noyaux de Fejer de la loi p_X .

D'autres équivalences peuvent être mises à jour entre des estimateurs spectraux de $Z_n(\omega)$ et des estimateurs de la loi de p_X [4, 5, 6]. Cependant, le calcul de l'entropie associée à des estimateurs de la loi plus élaborés est souvent impossible, à notre connaissance, à conduire analytiquement. Nous avons toutefois obtenu une expression analytique dans le cas d'une modélisation AR de la densité de probabilité, cf. [4]. Nous nous restreignons ici à l'estimateur du type histogramme défini par (4) ou (5), et dérivons maintenant la formule de remise à jour de l'estimateur de l'entropie qui lui est associé.

3 Estimateur de l'entropie

3.1 Cas d'une loi discrète

L'estimée de l'entropie à l'instant p , notée $\hat{H}^{(p)}$, est calculée comme l'entropie de l'histogramme estimé, si bien que

$$\begin{aligned} \hat{H}^{(p)} &= - \sum_{k=0}^{N-1} \hat{f}_k^{(p)} \log \hat{f}_k^{(p)} \\ &= - \sum_{k=0}^{N-1} \left(\frac{U_k^{(p)}}{p} \right) \log \left(\frac{U_k^{(p)}}{p} \right). \end{aligned}$$

La même définition s'applique à l'instant $(p+1)$. Etant donnée la formule de remise à jour de l'histogramme (7) et la condition de normalisation (3), on en déduit aisément une relation de récurrence sur l'entropie estimée :

$$\hat{H}^{(p+1)} = - \sum_{k=0}^{N-1} \frac{U_k^{(p+1)}}{p+1} \log \frac{U_k^{(p+1)}}{p+1},$$

d'où

$$\begin{aligned} (p+1) \hat{H}^{(p+1)} &= - \sum_{k=0}^{N-1} U_k^{(p+1)} \log \frac{U_k^{(p+1)}}{p+1}, \\ &= - \sum_{\substack{k=0 \\ k \neq l_{p+1}}}^{N-1} U_k^{(p+1)} \log U_k^{(p+1)} - U_{l_{p+1}}^{(p+1)} \log U_{l_{p+1}}^{(p+1)} \\ &\quad + (p+1) \log(p+1), \\ &= - \sum_{k=0}^{N-1} U_k^{(p)} \log U_k^{(p)} + U_{l_{p+1}}^{(p)} \log U_{l_{p+1}}^{(p)} \\ &\quad - \left(U_{l_p}^{(p)} + 1 \right) \log \left(U_{l_p}^{(p)} + 1 \right) + (p+1) \log(p+1). \end{aligned}$$

Or

$$\begin{aligned} p \hat{H}^{(p)} &= - \sum_{k=0}^{N-1} U_k^{(p)} \log \frac{U_k^{(p)}}{p} \\ &= p \log p - \sum_{k=0}^{N-1} U_k^{(p)} \log U_k^{(p)}, \end{aligned}$$

d'où $-\sum_{k=0}^{N-1} U_k^{(p)} \log U_k^{(p)} = -p \log p + p \hat{H}^{(p)}$ si bien que

$$\begin{aligned} (p+1) \hat{H}^{(p+1)} &= p \hat{H}^{(p)} + U_{l_{p+1}}^{(p)} \log U_{l_{p+1}}^{(p)} \\ &\quad - \left(U_{l_p}^{(p)} + 1 \right) \log \left(U_{l_p}^{(p)} + 1 \right) \\ &\quad + (p+1) \log(p+1) - p \log p \\ &= p \hat{H}^{(p)} - g \left(U_{l_{p+1}}^{(p)} \right) + g(p), \end{aligned}$$

d'où finalement

$$\begin{cases} \hat{H}^{(p+1)} = \frac{p}{p+1} \hat{H}^{(p)} + \frac{1}{p+1} \left\{ g(p) - g \left(U_{l_{p+1}}^{(p)} \right) \right\} \\ \hat{H}^{(1)} = 0 \end{cases} \quad (11)$$

avec $g(x) = (x+1) \log(x+1) - x \log x$.

Le choix de l'initialisation est justifié comme suit : à l'instant $p=1$, $X_1(\omega) \in I_{l_1}$ et l'estimateur de la loi s'écrit

$$\hat{f}^{(1)}(x) = \begin{cases} 1 & \text{si } x \in I_{l_1} \\ 0 & \text{sinon} \end{cases}$$

d'où $\hat{H}^{(1)} = 0$.

La formule (11) peut être interprétée de la façon suivante : le nouvel estimateur est calculé comme le barycentre de l'ancien estimateur et d'une contribution due au nouvel échantillon. Comme la fonction g est croissante, cette contribution sera d'autant plus faible que l'intervalle $I_{l_{p+1}}$, qui reçoit l'échantillon $X_{p+1}(\omega)$, est déjà peuplé. Dans le cas extrême où tous les échantillons appartiennent au même intervalle, l'incrément est identiquement nul et l'entropie estimée est nulle. Au contraire, si les échantillons visitent "uniformément" tous les intervalles, l'entropie sera grande. Ceci est cohérent avec le résultat selon lequel la loi uniforme $(\alpha_k = \frac{1}{N})_{0 \leq k \leq N-1}$ est, parmi toutes les lois de support borné, celle d'entropie maximale.

3.2 Cas d'une loi continue

L'estimation de l'entropie sous l'hypothèse d'une loi continue est tout aussi immédiate, puisqu'on déduit facilement, à partir des relations (1, 2, 4 et 5), une relation simple entre l'estimateur \hat{H} de l'entropie de la loi continue et l'estimateur \hat{H} de l'entropie de la loi discrète :

$$\hat{H}^{(p)} = \hat{H}^{(p)} + \log h.$$

Nous en déduisons la formule récurrente de calcul de cet estimateur :

$$\begin{cases} \hat{H}^{(p+1)} = \frac{p}{p+1} \hat{H}^{(p)} + \frac{1}{p+1} \left\{ g(p) - g \left(U_{l_{p+1}}^{(p)} \right) + \log h \right\}, \\ \hat{H}^{(1)} = 0. \end{cases} \quad (12)$$

4 Application à la détection de ruptures

Nous montrons ici une des applications immédiates de cet estimateur : la détection de non-stationnarité de la loi.

L'estimateur d'entropie obtenu précédemment, et construit d'après l'hypothèse de données stationnaires, ne peut évidemment être appliqué tel quel dans ce contexte. Cependant, nous tirons parti de sa forme récursive ((11) ou (12)) qui rend possible l'introduction d'un facteur d'oubli, lequel confère à l'estimateur un comportement adaptatif (capacité de poursuite des

non-stationnarités). On montre aisément que l'introduction d'un facteur d'oubli λ conduit à l'estimateur suivant dans le cas discret :

$$\hat{H}^{(p+1)} = \frac{\lambda^{p+1} - \lambda}{\lambda^{p+1} - 1} \hat{H}^{(p)} + \frac{\lambda - 1}{\lambda^{p+1} - 1} \left\{ g(p) - g\left(U_{l_{p+1}}^{(p)}\right) \right\},$$

et dans le cas continu :

$$\hat{H}^{(p+1)} = \frac{\lambda^{p+1} - \lambda}{\lambda^{p+1} - 1} \hat{H}^{(p)} + \frac{\lambda - 1}{\lambda^{p+1} - 1} \left\{ g(p) - g\left(U_{l_{p+1}}^{(p)}\right) + \log h \right\}.$$

La figure (Fig.2) montre l'évolution de l'entropie estimée dans le cas d'un signal $s(n)$ représenté ci-dessous (Fig.1) et composé de 1012 échantillons tirés suivant une loi bi-gaussienne (signal binaire avec bruit additif gaussien), suivis de 1012 échantillons de loi uniforme, pour un facteur d'oubli $\lambda = 0.997$ et $N = 30$ subdivisions.

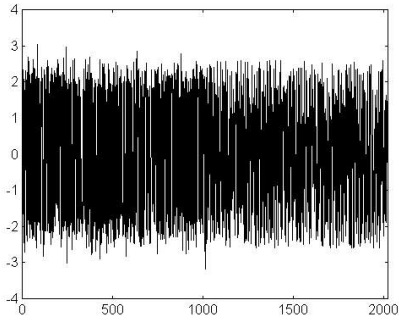


Fig. 1 : signal $s(n)$

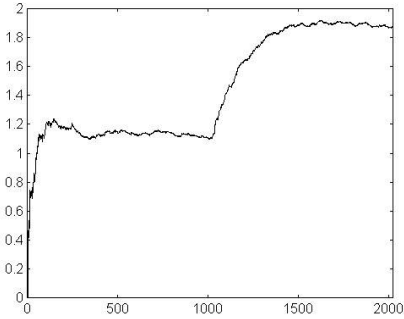


Fig. 2 : entropie estimée \hat{H}

Cette figure indique que l'estimateur empirique converge vers les valeurs théoriques (avec un biais inhérent à l'estimation de l'histogramme), tout en ayant une capacité de poursuite. On remarque de plus que la vitesse de convergence, bien que modeste, reste tout-à-fait acceptable puisque 500 itérations suffisent pour assurer la convergence vers la nouvelle entropie.

5 Application à la séparation de sources

La simulation suivante montre que l'estimateur proposé peut être utilisé pour résoudre le problème de la séparation de sources. Deux sources indépendantes s_1 et s_2 sont mélangées par une matrice A (2×2), ce qui génère les observations x_1 et x_2 . Les sources et la matrice de mélange étant inconnues, il s'agit de construire une estimation y_1 et y_2 des sources en appliquant aux observations une matrice B . Cette matrice B est recherchée de telle sorte que les signaux reconstitués y_1 et y_2 sont indépendants : le produit $A \times B$ sera alors égal à l'identité, à

une matrice de permutation et un scalaire près. On montre que l'indépendance des sources reconstituées peut être assurée par la minimisation de la fonction suivante :

$$C(B) = H_{y_1} + H_{y_2} - \log |\det B|$$

où H_{y_i} représente l'entropie du signal y_i . Nous avons simulé cette méthode dans le cas où ces entropies sont estimées selon la formule (12).

Le tableau suivant indique la valeur de la matrice $M = AB$ (idéalement égale à l'identité) obtenue par cette approche pour différents types de sources, chacune de 1000 échantillons : deux sources uniformes, une source uniforme et une binaire, une source normale et une binaire.

source 1	source 2	M	
$U_{[-0.5, 0.5]}$	$U_{[-0.5, 0.5]}$	1.0032	0.023
		0.0125	0.998
binaire $[-\frac{1}{2}, \frac{1}{2}]$	$U_{[-0.5, 0.5]}$	1	0.0004
		0.0012	0.9984
binaire $[-\frac{1}{2}, \frac{1}{2}]$	$N(0, 1)$	1.03	0.0003
		0.02	0.9937

Ces résultats indiquent que la séparation est réalisée dans tous les cas : le biais inhérent à notre estimateur ne permet cependant pas une convergence exacte. On remarquera toutefois que cette séparation, bien que non exacte, est atteinte à coût de calcul extrêmement faible, ce qui rend cette méthode très attractive, par exemple pour l'initialisation d'algorithmes plus complexes.

6 Conclusion

Nous avons présenté dans cet article un estimateur simple de l'entropie d'un signal, possédant la propriété de récursivité : il peut donc être utilisé dans des contextes non-stationnaires. Cet estimateur, bien que de coût de calcul très réduit, montre de bonnes performances en séparation de sources et en détection de ruptures, comme l'illustrent nos simulations.

Références

- [1] P. Viola, N. Schraudolph and T. Sejnowski, "Empirical Entropy Manipulation for Real-World Problems", Advances in Neural Information Processing Systems, 8, MIT Press, 1996.
- [2] D.T. Pham, "Blind Separation of Instantaneous Mixtures of Sources based on Order Statistics", Technical Report LMC/IMAG, March 1998.
- [3] A. Pagès-Zamora and M.A. Lagunas, "Joint Probability Density Function Estimation by Spectral Estimate Methods", pp. 2936-2939, vol. 5, ICASSP 96, Atlanta
- [4] J.-F. Bercher and C. Vignat, "Estimating the entropy of a Signal with Applications", ICASSP 1999, Phoenix
- [5] S. Kay, "Model-Based Probability Density Function Estimation", IEEE Signal Processing Letters, vol.5, no.12, Dec. 98, pp. 318-320
- [6] J.-F. Bercher et C. Vignat, "Entropy Estimation with Applications to Signal Processing Problems", LSC, Rapport Interne no. 98-057.