Estimation par plug-in du taux d'entropie d'un processus markovien de sauts à espace d'état fini

Philippe Regnault

Estimation par plug-in du taux d'entropie d'un processus markovien de sauts à espace d'état fini

Philippe Regnault

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen BP 5186, 14032 CAEN cedex

Résumé

L'entropie d'une loi à valeurs dans un ensemble fini est largement utilisée dans toutes les applications impliquant des variables aléatoires. L'équivalent naturel pour un processus aléatoire est son taux d'entropie, s'exprimant comme une fonction de la probabilité invariante et du générateur pour un processus markovien de sauts homogène, ergodique, à espace d'état fini.

On construit un estimateur par plug-in de ce taux d'entropie dans le cas de l'observation d'une trajectoire du processsus sur une longue période de temps. On démontre que cet estimateur a de bonnes propriétés asymptotiques, il est convergent et asymptotiquement normal ; sa variance asymptotique est explicite dans la plupart des cas.

Le cas des processus à deux états, particulièrement lié à l'étude de durées de vie ou de la fiabilité d'un système, fait l'objet d'une étude numérique détaillée.

Abstract

The entropy of a distribution with finite support is widely used in all applications involving random variables. A natural equivalent for random processes is the entropy rate. For ergodic pure-jump finite state Markov processes, this rate is an explicit function of the asymptotic distribution and the infinitesimal generator .

We estimate the entropy rate by plug-in from the observation of one long trajectory of the process. This estimator is proven to be strongly consistent and asymptotically normal with explicit variance in most of the cases.

The case of two-state Markov processess, widely used in reliability or survival data analysis is detailled and illustrated.

Mots-clés : Statistique mathématique, statistique des processus, processus markoviens de sauts, ergodicité, taux d'entropie, estimation paramétrique.

1 Introduction

L'entropie d'une loi P sur un ensemble fini E,

$$H(P) = -\sum_{x \in E} P(x) \log P(x),$$

a été introduite par Shannon [7] en 1948 dans le cadre de l'étude des codes de compression : il a montré que si $(X_n)_{n \in N}$ est une chaîne de Markov ergodique à espace d'état fini, le quotient $\frac{1}{n} H(P_{(X_1,\dots,X_n)})$ admet une limite $\mathbb{H}(X)$ lorque n tend vers l'infini, appelée taux d'entropie de la chaîne, représentant le taux de compression optimal de codes de compression.

L'utilisation de l'entropie dépasse largement ce résultat original : une boîte à outils statistique complète a été développée et appliquée dans un grand nombre de domaines (voir [3]).

La notion de taux d'entropie introduite par Shannon s'adapte aux processus markoviens de sauts ergodiques à espace d'état fini comme suit.

Soit $X = (X_t)_{t\in\mathbb{R}_+}$ un processus markovien de sauts ergodique à valeurs dans un ensemble fini E. On note $X_{(T)}$ la restriction de X à l'intervalle $[0,T]$.

Définition :
L'entropie partielle de X est $H_T(X) = \int f_{X_{(T)}} \log\left(f_{X_{(T)}}\right) dm$ où $f_{X_{(T)}}$ est la vraisemblance de $P_{X_{(T)}}$ par rapport à une mesure dominante m.
Le taux d'entropie $\mathbb{H}(X)$ de X est la limite de $\frac{1}{T}H_T(X)$ lorsque T tend vers l'infini.

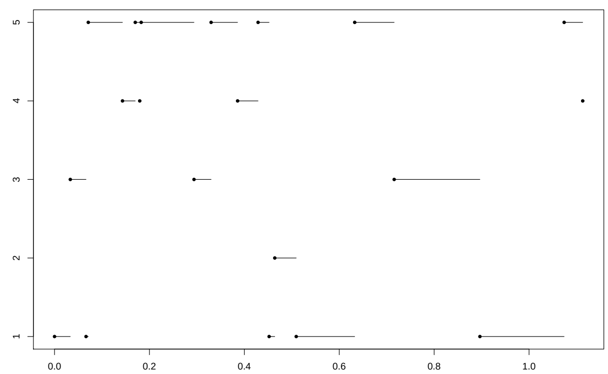

FIG. 1 – Une trajectoire d'un processus de markov à cinq états

Une mesure dominante m de $P_{X_{(T)}}$ a été construite par Albert dans [1] et reprise par Bad Dumitrescu dans [4]. La vraisemblance associée y est explicitée, permettant d'établir une formule explicite du taux d'entropie d'un processus de Markov ergodique, fonction de son générateur $A = (a_{i,j})_{(i,j)\in E^2}$ et de sa probabilité invariante π, soit

$$\mathbb{H}(X) = -\sum_{i\in E}\pi(i)\sum_{j\in E, j\neq i} a_{i,j}\log a_{i,j} + \sum_{i\in E}\pi(i)\sum_{j\in E, j\neq i} a_{i,j}. \tag{1}$$

2 Estimation du taux d'entropie

L'estimation du taux d'entropie d'une chaîne de Markov a été abordée par G. Ciuperca et V. Girardin dans [2] puis par V. Girardin et A. Sesboüé dans [5]. Les auteurs y proposent un estimateur par plug-in du taux d'entropie, basé sur l'estimation de la matrice de transition de la chaîne et de sa loi stationnaire. On adapte ici cette démarche au cas d'un processus à temps continu.

La probabilité invariante π d'un processus de Markov ergodique est caractérisée par l'égalité $\pi.A = 0$. La probabilité invariante est donc une fonction du générateur. A. Albert a établi dans [1] une formule explicite, précisément

$$\pi(i) = \frac{a^{(i,i)}}{\sum_{k \in E} a^{(k,k)}},$$

où $a^{(i,i)}$ est le (i,i)-ième cofacteur de A.

Ce résultat, joint au théorème de Bad Dumitrescu, implique que le taux d'entropie est une fonction du générateur,

$$\mathbb{H}(X) = h(A).$$

On construit alors un estimateur par plug–in de $\mathbb{H}(X)$,

$$\widehat{H}_T = h(\widehat{A}_T),$$

où $\widehat{A}_T$ est un estimateur du générateur.

A. Albert ([1]) a construit un estimateur $\widehat{A}_T$ par maximum de vraisemblance du générateur d'un processus ergodique. Explicitement,

$$\widehat{A}_T(i,j) = \begin{cases} \dfrac{n_T(i,j)}{r_T(i)} & \text{si } i \neq j \text{ et } r_T(i) \neq 0, \\ 0 & \text{si } i \neq j \text{ et } r_T(i) = 0, \\ -\sum\limits_{j \neq i} \widehat{A}_T(i,j) & \text{si } i = j. \end{cases}$$

où $n_T(i,j)$ est le nombre de transitions de l'état i à l'état j et $r_T(i)$ est le temps de séjour en i durant l'intervalle de temps $[0,T]$. Cet estimateur possède de bonnes propriétés asymptotiques :

- $\widehat{A}_T$ converge presque sûrement vers A,
- $\sqrt{T}\left(\widehat{A}_T - A\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_A^2)$ où Σ_A^2 est une matrice diagonale dont les coefficients diagonaux sont $a_{i,j}\rho/a_{i,i}$, où ρ est le produit des valeurs propres non nulles de A.

3 Propriétés asymptotiques de l'estimateur

L'estimateur $\widehat{H}_T = h(\widehat{A}_T)$ hérite alors des propriétés de $\widehat{A}_T$.

Théorème : Aves les notations données plus haut,

1. $\widehat{H}_T$ est fortement consistant, soit $\widehat{H}_T \xrightarrow{\text{p.s.}} \mathbb{H}(X)$,
2. si la dérivée $D_h(A)$ de h en A est non nulle, $\widehat{H}_T$ est asymptotiquement normal et de variance asymptotique explicite, soit
$$\sqrt{T}\left(\widehat{H}_T - \mathbb{H}(X)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_H^2) \text{ où } \Sigma_H^2 = \sum_{(i,j) \in E^2, i \neq j} \frac{a_{i,j}}{a^{(i,i)}} \rho \left(\frac{\partial h}{\partial a_{i,j}}(A)\right)^2,$$
3. Si $D_h(A) = 0$, alors $2T\left(\widehat{H}_T - \mathbb{H}(X)\right) \xrightarrow{\mathcal{L}} \sum_{(i,j) \in E, i \neq j} \lambda_{i,j}\chi^2(1),$
avec $\lambda_{i,j} = a_{i,j}\rho/a^{(i,i)}$.

Démonstration :

1. Puisque $\widehat{A}_T$ converge presque sûrement vers A et que h est continue, $\widehat{H}_T = h(\widehat{A}_T)$ converge presque sûrement vers $h(A) = \mathbb{H}(X)$.

2., 3. Pour les ditributions asymptotiques, on propose la démonstration du cas particulier d'un processus à deux états dans la section suivante. Le lecteur pourra se reporter à [6] pour une démonstration dans le cas général.

Dans le cas où $D_h(A)$ ne s'annule pas, la variance asymptotique est une fonction explicite du générateur A, soit $\Sigma_H^2 = s(A)$. Son estimateur par plug-in $\widehat{\Sigma^2_{HT}} = s(\widehat{A}_T)$ est fortement consistant, d'où le résultat suivant.

Corollaire : Si $D_h(A) \neq 0$ alors $\sqrt{T}(\widehat{H}_T - \mathbb{H}(X))/\widehat{\Sigma_{HT}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

4 Cas d'un processus à deux états

Dans le cas particulier d'un processus à deux états, on peut préciser le résultat du théorème précédent comme suit.

Théorème :

1. Si le générateur n'est pas uniforme, alors $\sqrt{T}(\widehat{H}_T - \mathbb{H}(X)) \to \mathcal{N}(0, \Sigma_H^2)$ quand T tend vers l'infini,
   où $\Sigma_H^2 = \dfrac{a_{1,2}a_{2,1}}{(a_{1,2}+a_{2,1})^3}\big((-a_{1,2} - a_{2,1}\log(a_{1,2}a_{2,1}) + a_{2,1})^2$
   $$+(-a_{2,1} + a_{1,2} - a_{1,2}\log(a_{1,2}a_{2,1})^2)\big).$$

2. Si le générateur est uniforme, alors $2T(\mathbb{H}(X) - \widehat{H}_T) \to \chi^2(2)$ quand T tend vers l'infini.

Démonstration :

1. La dérivée de h en A est nulle si et seulement si le générateur est uniforme ($a_{1,2} = a_{2,1} = 1$). En effet, la formule (1) devient, pour $n = 2$,
   $$\mathbb{H}(X) = \frac{a_{1,2}a_{2,1}}{a_{1,2}+a_{2,1}}(2 - \log(a_{1,2}a_{2,1}))$$
   où $A = \begin{pmatrix} -a_{1,2} & a_{1,2} \\ a_{2,1} & -a_{2,1} \end{pmatrix}$ et $a_{1,2}, a_{2,1} > 0$
   donc
   $$\frac{\partial h}{\partial a_{1,2}}(a_{1,2}, a_{2,1}) = \frac{-a_{1,2}a_{2,1} + a_{2,1}^2 - a_{2,1}^2\log a_{1,2}a_{2,1}}{(a_{1,2}+a_{2,1})^2},$$
   $$\frac{\partial h}{\partial a_{2,1}}(a_{1,2}, a_{2,1}) = \frac{-a_{1,2}a_{2,1} + a_{1,2}^2 - a_{1,2}^2\log a_{1,2}a_{2,1}}{(a_{1,2}+a_{2,1})^2}.$$
   Le système d'équations
   $$\left\{\begin{array}{rcl} -a_{1,2}a_{2,1} + a_{2,1}^2 - a_{2,1}^2\log a_{1,2}a_{2,1} & = & 0 \\ -a_{1,2}a_{2,1} + a_{1,2}^2 - a_{1,2}^2\log a_{1,2}a_{2,1} & = & 0 \end{array}\right.$$

admet pour unique solution $(a_{1,2}, a_{2,1}) = (1, 1)$.

2. Si A n'est pas uniforme, le résultat est une conséquence de la méthode delta.
3. Si A est uniforme, un développement de Taylor de h à l'ordre 2 donne
$$\widehat{H}_T - \mathbb{H}(X) = -\frac{1}{4}\left[\left(\widehat{A}_T(1,2) - a_{1,2}\right)^2 + \left(\widehat{A}_T(2,1) - a_{2,1}\right)^2\right]$$
$$+ o\left(\|\widehat{A}_T - A\|^2\right),$$
la dérivée croisée étant nulle pour le générateur uniforme.
Or $\sqrt{T}\left(\frac{\widehat{A}_T(1,2)-a_{1,2}}{\Sigma_A(1,2)}, \frac{\widehat{A}_T(2,1)-a_{2,1}}{\Sigma_A(2,1)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id)$, donc
$$T\left[\frac{\left(\widehat{A}_T(1,2) - a_{1,2}\right)^2}{\Sigma_A^2(1,2)} + \frac{\left(\widehat{A}_T(2,1) - a_{2,1}\right)^2}{\Sigma_A^2(2,1)}\right] \xrightarrow{\mathcal{L}} \chi^2(2).$$
Le résultat en découle immédiatement puisque $\Sigma_A^2(1,2) = \Sigma_A^2(2,1) = 2$.

Les figures 2 et 3 illustrent respectivement la convergence ponctuelle de l'estimateur $\widehat{H}_T$ dans le cas d'un générateur non uniforme et dans le cas du générateur uniforme. L'estimateur a été calculé à partir de la simulation d'un processus de Markov pour un intervalle de temps $[0, 5000]$. Dans les deux cas, la convergence est très rapide, d'autant plus que le générateur est proche du générateur uniforme (la convergence y étant plus rapide puisque la dérivée y est nulle).

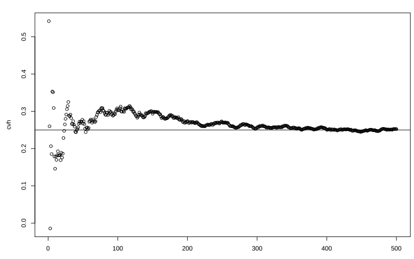

FIG. 2 – Convergence de $\widehat{H}_T$ pour $(a_{1,2}, a_{2,1}) = (2, 3)$

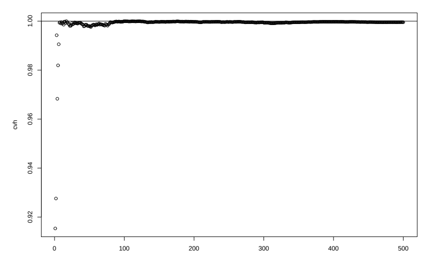

FIG. 3 – Convergence de $\widehat{H}_T$ pour $(a_{1,2}, a_{2,1}) = (1, 1)$

Les figures 4 et 5 illustrent respectivement la convergence de la fonction de répartition empirique de $\sqrt{T}(\widehat{H}_T - H(X))/\Sigma_H$ vers celle d'une loi normale pour le cas d'un processus de générateur non uniforme et la convergence de la fonction de répartition empirique de $2T(\widehat{H}_T - H(X))$ vers celle d'une loi du χ^2 à deux degrés de liberté pour un processus de générateur uniforme. Dans les deux cas, les fonctions de répartition empiriques ont été obtenues après la simulation de 200 trajectoires sur les intervalles de temps $[0, 1000]$, $[0, 2000]$ et $[0, 3000]$.

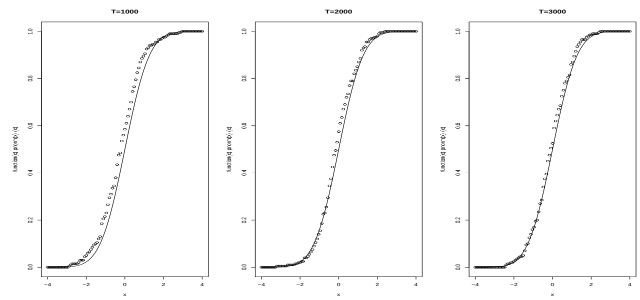

FIG. 4 – Convergence de la fonction de répartition empirique de $\sqrt{T}(\widehat{H}_T - H(X))/\Sigma_H$ vers celle de la loi normale centrée réduite pour $(a_{1,2}, a_{2,1}) = (2, 3)$

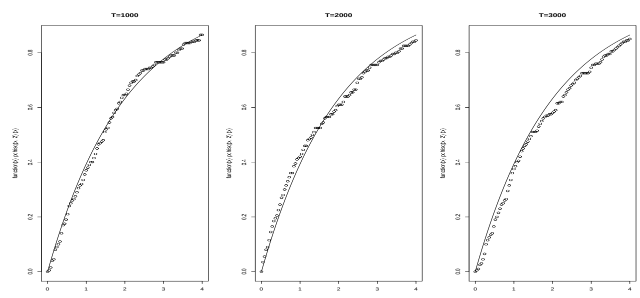

FIG. 5 – Convergence de la fonction de répartition empirique de $2T(\widehat{H}_T - H(X))$ vers celle de la loi du $\chi^2(2)$ pour $(a_{1,2}, a_{2,1}) = (1, 1)$

Références

[1] A. ALBERT. Estimating The Infinitesimal Generator of a Continious Time, Finite State Markov Process. *Annals of mathematical statistics*, Vol. 33, p.727-753. 1962.

[2] G. CIUPERCA and V. GIRARDIN. Estimation of the Entropy Rate of a Countable Markov Chain. *Communications in Statistics - Theory and Methods*, Vol. 36, p. 1-15. 2007.

[3] T. M. COVER and J. A. THOMAS. *Elements of Information Theory.* Edition Wiley. 1991.

[4] M. BAD DUMITRESCU. Some Informational Properties of Markov Pure-Jump Processes. *Casopis Pro Pestovani Matematiky* Vol. 4, p. 429-434. 1986.

[5] V. GIRARDIN and A. SESBOÜÉ. Comparative Construction of Plug-in Estimators of the Entropy Rate of Two-state Markov Chains. A paraître dans *Methodology and Computing in Applied Probability.* 2009.

[6] P. REGNAULT. Etude et estimation du taux d'entropie d'un processus de Markov. *Mémoire de fin de master. Université Paris-Sud.* 2008.

[7] C.E. SHANNON. A Mathematical Theory of Communication. *The Bell System Technical Journal* Vol. 27, p. 379-423, 623-656. 1948.