C. R. Acad. Sci. Paris, Ser. I 353 (2015) 273–277

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistique

Choix du paramètre de lissage dans l'estimation à noyau d'une matrice de transition d'un processus semi-markovien

Choice of the smoothing parameter in the kernel estimation of the transition matrix of a semi-Markovian process

Mouloud Cherfaoui, Mohamed Boualem, Djamil Aïssani, Smail Adjabi

Laboratoire de modélisation et d'optimisation des systèmes (LAMOS), université de Béjaia, Targa-Ouzamour, 06000 Algérie

INFO ARTICLE

Historique de l'article :
Reçu le 7 août 2013
Accepté après révision le 12 septembre 2014
Disponible sur Internet le 19 janvier 2015

Présenté par le comité de rédaction

RÉSUMÉ

Dans cette Note, nous étudions le problème du choix du paramètre de lissage pour un estimateur à noyau de l'opérateur de transition d'une chaîne de Markov. Pour ce faire, nous avons considéré la file d'attente $GI/M/1/N$. Nous avons constaté que l'estimateur du paramètre de lissage choisi, par la minimisation d'une certaine norme matricielle, donne de meilleurs résultats, en termes de vitesse de convergence de l'erreur quadratique moyenne, que les alternatives classiques.

ABSTRACT

In this Note, we study the problem of choosing the smoothing parameter for a kernel estimator of the transition operator of a Markov chain. To do this, we have considered the $GI/M/1/N$ queue. The proposed smoothing parameter performs better than the existing classical methods in terms of convergence rate of the mean square error.

1. Introduction

Dans la théorie classique de l'estimation paramétrique d'une matrice de transition associée à une chaîne de Markov, nous disposons de plusieurs méthodes, décrites dans [2], qui présentent l'avantage d'être simples à utiliser. Toutefois, il est difficile d'estimer avec précision des matrices de transition modélisant des phénomènes complexes. Pour pallier cette difficulté, nous faisons appel aux méthodes d'estimation non paramétriques. Ces dernières ont fait l'objet de travaux établis par Roussas [9], en utilisant la méthode du noyau. Les résultats obtenus par celui-ci ont été complétés par Masry et Györfi [7] et par Basu et Sahoo [1], parmi d'autres. Par la suite, Laksaci et Yousfate [6] ont étudié un estimateur à noyau de la densité de l'opérateur de transition, vu comme un endomorphisme de L^p, $p \in [1, \infty[$. Cet estimateur permet de construire un

Adresses e-mail : cherfaouimouloud@yahoo.fr (M. Cherfaoui), robertt15dz@yahoo.fr (M. Boualem), lamos_bejaia@hotmail.com (D. Aïssani), adjabi@hotmail.com (S. Adjabi).

http://dx.doi.org/10.1016/j.crma.2014.09.030

estimateur fonctionnel aussi bien pour l'opérateur de transition que pour son adjoint. Leur principal résultat fournit une majoration de la vitesse de convergence (au sens de la norme L^p) de l'estimateur construit.

Dans cette Note, l'objectif principal est d'estimer une matrice de transition inconnue $\mathbb{P}$, associée à une file d'attente, par la méthode de noyau. Plus précisément, notre étude se focalise sur la qualité de l'estimateur de $\mathbb{P}$ par rapport aux procédures de sélection du paramètre de lissage. Afin de réaliser cette étude, nous avons considéré le système particulier $GI/M/1$ (*FIFO*, N). Pour cela, nous avons effectué une étude comparative des résultats obtenus par les méthodes classiques de sélection du paramètre de lissage et par d'autres méthodes qui se basent sur les normes matricielles. Contrairement aux méthodes classiques, ces dernières prennent en considération la loi des temps de service, qui joue le rôle d'une pondération de la distribution générale des inter-arrivées dans les éléments de la matrice de transition du système considéré. Le choix du modèle $GI/M/1$ (*FIFO*, N) est motivé par la disponibilité de ses caractéristiques sous des formes explicites dans la littérature des files d'attente [5]. Par conséquent, afin d'avoir une idée sur la qualité de l'estimateur de $\mathbb{P}$, les matrices de transition peuvent être calculées avec exactitude et les erreurs de calcul numérique (qui peuvent être engendrées par la machine) peuvent ainsi être évitées.

2. Choix du paramètre de lissage

Nous nous intéressons au choix du paramètre de lissage dans l'estimation à noyau, de la matrice de transition, qui minimise un certain critère d'erreur. Pour cela, supposons qu'on dispose d'un n-échantillon $T_1, T_2, \ldots, T_n$, qui représente les durées des inter-arrivées, dans un système $GI/M/1/N$, ayant comme densité de probabilité inconnue g. L'estimation de la matrice de transition, $\mathbb{P}$, de la chaîne de Markov induite associée au système $GI/M/1/N$, donnée par :

$$P_{ij} = \begin{cases} \int_0^\infty \frac{e^{-\mu t}(\mu t)^{i-j+1}}{(i-j+1)!} g(t)\,dt, & \text{si } 1 \le j \le i+1 \le N, \\ \int_0^\infty \frac{e^{-\mu t}(\mu t)^{N-j}}{(N-j)!} g(t)\,dt, & \text{si } 1 \le j \le N \text{ et } i = N, \\ 1 - \sum_{k=1}^N P_{ik}, & \text{si } j = 0, \\ 0, & \text{sinon}, \end{cases} \tag{1}$$

consiste à évaluer la densité inconnue g et de substituer son estimateur, noté g_h, dans les P_{ij}. Dans notre cas, nous optons pour le critère *MISE* et le premier noyau gamma proposé par Chen [4], qui remplacera les noyaux usuels donnés dans [8]. Alors, nous obtenons la formule explicite de g_h, donnée par :

$$g_h(t) = \frac{1}{n}\sum_{i=1}^n K(t,h)(T_i) = \frac{1}{n}\sum_{i=1}^n \frac{T_i^{t/h} e^{-T_i/h}}{h^{(t/h)+1}\Gamma((t/h)+1)}, \tag{2}$$

où K désigne la densité de la loi gamma de paramètres $(\frac{t}{h}+1, h)$ et où $h = h(n)$ est le paramètre de lissage. Le paramètre de lissage optimal h_1^* se calcule comme suit :

$$\begin{aligned} h_1^* &= \arg\min_h MISE(g, g_h) = \arg\min_h \int_0^\infty \mathbb{E}\big(g(t) - g_h(t)\big)^2\,dt \\ &= \arg\min_h \left[h^2 \int_0^\infty \left\{ g'(t) + \frac{1}{2} t g''(t) \right\}^2 dt + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^\infty t^{-1/2} g(t)\,dt + o\big(n^{-1}h^{-1/2} + h^2\big) \right] \\ &\approx \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty t^{-1/2} g(t)\,dt \right]^{2/5} \left[\int_0^\infty \left\{ g'(t) + \frac{1}{2} t g''(t) \right\}^2 dt \right]^{-2/5} 4^{-2/5} n^{-2/5}. \end{aligned} \tag{3}$$

Dans le but de prendre en considération les pondérations, nous proposons l'utilisation des normes matricielles qui ont un impact sur la qualité de l'estimateur de $\mathbb{P}$, noté $\hat{\mathbb{P}}$. En effet, l'utilisation des normes matricielles nous permet d'inclure les pondérations qui sont d'une loi de Poisson, de paramètre $\mu * t$, de la quantité $g(t)$ dans l'expression des P_{ij} lors de l'estimation de $\mathbb{P}$. Dans ce cas, le paramètre de lissage optimal est calculé selon l'une des trois expressions suivantes :

$$h_2^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_1 = \arg\min_h \left[\max_j \left(\sum_{i=0}^N |\hat{P}_{ij} - P_{ij}| \right) \right], \tag{4}$$

$$h_3^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_2 = \arg\min_h \left[\sum_{i=0}^N \sum_{j=0}^N (\hat{P}_{ij} - P_{ij})^2 \right]^{1/2}, \tag{5}$$

$$h_4^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_\infty = \arg\min_h \left[\max_i \left(\sum_{j=0}^N |\hat{P}_{ij} - P_{ij}| \right) \right], \tag{6}$$

où $\hat{P}_{ij}$ est l'estimateur de P_{ij} lorsque nous remplaçons $g(t)$ par son estimateur $g_h(t)$ donné par (2).

Énonçons à présent quelques résultats théoriques concernant les propriétés de l'estimateur $\hat{P}_{ij}$.

Proposition 2.1. *Soient $g \in \mathcal{C}^2([0,\infty[)$ une densité de probabilité de la distribution générale des durées des inter-arrivées du système $GI/M/1/N$ et g_h son estimateur à noyau gamma. Sous la condition $\lim_{n\to+\infty} h = 0$, le biais asymptotique de l'estimateur $\hat{P}_{ij}$ s'écrit comme suit :*

$$Biais(\hat{P}_{ij}) = \begin{cases} h\int_0^\infty \frac{(\mu t)^{i-j+1}}{(i-j+1)!}\mathrm{e}^{-\mu t}\{g'(t)+\frac{1}{2}tg''(t)\}\,\mathrm{d}t + o(h), & si\ 1 \le j \le i+1 \le N, \\ h\int_0^\infty \frac{(\mu t)^{N-j}}{(N-j)!}\mathrm{e}^{-\mu t}\{g'(t)+\frac{1}{2}tg''(t)\}\,\mathrm{d}t + o(h), & si\ 1 \le j \le N\ et\ i = N, \\ -\sum_{k=1}^{N} Biais(\hat{P}_{ik}), & si\ j = 0, \\ 0, & sinon. \end{cases} \tag{7}$$

Théorème 2.2. *Soient $g \in \mathcal{C}^2([0,\infty[)$ une densité de probabilité de la distribution générale des durées des inter-arrivées du système $GI/M/1/N$ et g_h son estimateur à noyau gamma.*

$$Si\ \Big(\lim_{n\to+\infty} h = 0\ \ et\ \lim_{n\to+\infty} nh^2 = +\infty\Big), \quad alors\ \big(|\hat{P}_{ij} - P_{ij}| \xrightarrow{P} 0\ lorsque\ h \to 0\big). \tag{8}$$

Remarque 1. La démonstration de la Proposition 2.1 peut se faire en exploitant le résultat de Chen [4] :

$$\mathbb{E}\{g_h(t)\} = g(t) + h\left\{g'(t) + \frac{1}{2}tg''(t)\right\} + o(h).$$

En suivant une démarche similaire à celle de Bouezmarni et Scaillet [3] concernant la consistance des estimateurs à noyau asymétrique d'une densité de probabilité, nous aboutissons au résultat énoncé dans le Théorème 2.2.

3. Application numérique et discussions

Pour répondre à notre objectif, nous avons conçu un algorithme dont les étapes principales sont :

Étape 1 : générer un échantillon de taille n de la loi des inter-arrivées,
Étape 2 : estimer h par la formule (3), calculer $\hat{\mathbb{P}}$ et l'erreur associée,
Étape 3 : estimer h par les formules (4)–(6), calculer $\hat{\mathbb{P}}$ et les erreurs associées.

Pour analyser l'impact de la variation du taux de service μ (poids de g dans les P_{ij}) et la taille de l'échantillon n sur la qualité de l'estimateur $\hat{\mathbb{P}}$, lorsque h est calculé par l'une des formules (3)–(6), dans le système $GI/M/1/N$, nous considérons les situations suivantes : g est une densité d'une loi exponentielle de paramètre $\lambda = 1$, $N = 5$, $\mu = [10/3,\ 5/3,\ 10/9,\ 5/6,\ 2/3]$ et $n = [100,\ 150,\ 200,\ 250,\ 300]$. Les résultats obtenus à l'aide de cet algorithme, en utilisant l'estimateur donné par (2), sont rangés dans le Tableau 1.

En tenant compte des définitions suivantes :

- Δ_n : la différence entre les tailles du $(j+1)$ème et du $j$ème échantillon pour la même valeur de la charge du système ρ,
- Δerr_i : la différence entre les erreurs associées aux $(j+1)$ème et $j$ème échantillon de la $i$ème norme pour une valeur fixe de ρ,
- le rapport $\frac{\Delta err_i}{\Delta n}$ représente la vitesse de convergence définie par la tangente de la courbe de l'erreur, associée à la $i$ème norme, en fonction de la taille de l'échantillon, on constate à partir des résultats du Tableau 1 que :
  - tous les estimateurs considérés convergent en fonction de la taille de l'échantillon. De plus, nous remarquons que la vitesse de convergence $(\frac{\Delta err_i}{\Delta n},\ i = \overline{1,4})$ des estimateurs de $\mathbb{P}$, au sens des normes considérées, est plus importante pour la norme $\|.\|_\infty$, tandis que pour la norme $\|.\|_2$ elle est plus faible ;
  - l'estimateur de $\mathbb{P}$ dépend du taux de service lorsque le paramètre de lissage est calculé par la norme $\|.\|_\infty$ donnée par (6), contrairement au cas où il est calculé par la norme $\|.\|_2$ donnée par (5).

Afin de confirmer le résultat précédent, nous avons évalué, en résolvant l'équation $\pi * \hat{\mathbb{P}} = \pi$, les probabilités stationnaires du système qui permettent l'obtention de $\bar{L}$: le nombre moyen de clients dans le système. Ceci est évidemment effectué en exploitant la matrice de transition estimée dans la deuxième et la troisième étape de l'algorithme. Les résultats obtenus sont résumés dans le Tableau 2.

À partir de ces résultats, en tenant compte des discussions précédentes, nous constatons que les estimateurs de $\bar{L}$, calculés en utilisant les $\hat{\mathbb{P}}$ fournis par les normes matricielles, sont quantitativement proches des valeurs exactes. De plus, nous constatons que l'utilisation de la norme $\|.\|_\infty$ fournit des résultats plus proches de ceux exacts que les normes $\|.\|_1$ et

Tableau 1
Valeurs optimales des h_i^* et les erreurs associées.

Table 1
Optimal h_i^* values and the associated errors.

ρ	n	Classique		Normes matricielles					
		h_1^*	err_1^*	h_2^*	err_2^*	h_3^*	err_3^*	h_4^*	err_4^*
0.3	100	0.2029	0.0062	0.0613	0.0807	0.0632	0.0144	0.0516	0.1955
	150	0.1955	0.0043	0.0647	0.0705	0.0665	0.0091	0.0502	0.1693
	200	0.1766	0.0038	0.0675	0.0606	0.0768	0.0070	0.0612	0.1431
	250	0.1469	0.0038	0.0475	0.0566	0.0487	0.0062	0.0452	0.1382
	300	0.1349	0.0034	0.0547	0.0490	0.0507	0.0050	0.0476	0.1158
0.6	100	0.2387	0.0065	0.1621	0.0558	0.1329	0.0078	0.1269	0.1174
	150	0.2061	0.0044	0.0994	0.0635	0.0901	0.0105	0.0852	0.1413
	200	0.1721	0.0036	0.0987	0.0467	0.0864	0.0061	0.0897	0.0999
	250	0.1346	0.0036	0.0405	0.0488	0.0368	0.0055	0.0366	0.1134
	300	0.1276	0.0026	0.0423	0.0506	0.0358	0.0060	0.0316	0.1133
0.9	100	0.2296	0.0071	0.0918	0.0687	0.0694	0.0113	0.0725	0.1213
	150	0.1998	0.0053	0.0899	0.0513	0.0923	0.0068	0.1026	0.0979
	200	0.1323	0.0042	0.0306	0.0555	0.0273	0.0077	0.0290	0.0965
	250	0.1370	0.0039	0.0462	0.0414	0.0442	0.0049	0.0428	0.0859
	300	0.1540	0.0031	0.0550	0.0464	0.0477	0.0052	0.0416	0.0904
1.2	100	0.2111	0.0063	0.0939	0.0498	0.0974	0.0070	0.1027	0.0866
	150	0.2020	0.0047	0.0975	0.0357	0.0956	0.0037	0.0916	0.0562
	200	0.1826	0.0037	0.0785	0.0390	0.0819	0.0042	0.0815	0.0642
	250	0.1438	0.0033	0.0591	0.0357	0.0558	0.0036	0.0549	0.0528
	300	0.1409	0.0029	0.0601	0.0338	0.0592	0.0030	0.0545	0.0536
1.5	100	0.2755	0.0054	0.1378	0.0429	0.1401	0.0054	0.1517	0.0546
	150	0.1632	0.0046	0.0356	0.0493	0.0370	0.0057	0.0322	0.0651
	200	0.1821	0.0038	0.0779	0.0420	0.0778	0.0049	0.0636	0.0631
	250	0.1551	0.0034	0.0620	0.0400	0.0642	0.0048	0.0630	0.0613
	300	0.1658	0.0025	0.0730	0.0320	0.0725	0.0027	0.0652	0.0530

Tableau 2
Valeurs exactes et estimées de $\bar{L}$.

Table 2
Exact and estimated values of $\bar{L}$.

ρ	n	Exacte	$\bar{L}$ Estimé			
		$\bar{L}$	$\bar{L}_1$	$\bar{L}_2$	$\bar{L}_3$	$\bar{L}_4$
0.3	100	0.4242	0.3599	0.3895	0.3892	0.3906
	150		0.3748	0.4060	0.4049	0.4103
	200		0.3834	0.4128	0.4104	0.4153
	250		0.3822	0.4087	0.4093	0.4104
	300		0.3878	0.4090	0.4094	0.4101
0.6	100	1.2064	1.0715	1.1362	1.1560	1.1662
	150		1.0326	1.1080	1.1155	1.1206
	200		1.0800	1.1402	1.1503	1.1470
	250		1.0673	1.1432	1.1466	1.1481
	300		1.0661	1.1418	1.1476	1.1521
0.9	100	2.1948	1.8948	2.0614	2.0877	2.0866
	150		1.9908	2.1307	2.1273	2.1161
	200		1.9422	2.0730	2.0761	2.0762
	250		1.9982	2.1189	2.1212	2.1251
	300		1.9677	2.0920	2.1011	2.1077
1.2	100	3.0212	2.7759	2.9399	2.9357	2.9304
	150		2.8489	2.9847	2.9870	2.9925
	200		2.8269	2.9554	2.9512	2.9514
	250		2.8465	2.9554	2.9599	2.9607
	300		2.8580	2.9648	2.9657	2.9718
1.5	100	3.5774	3.4275	3.5730	3.5710	3.5591
	150		3.3806	3.5217	3.5206	3.5257
	200		3.3734	3.4860	3.4862	3.5032
	250		3.3840	3.4839	3.4824	3.4842
	300		3.4090	3.5099	3.5106	3.5178

$\|.\|_2$. Par conséquent, il est important de noter que, dans la pratique et pour un calcul plus précis, il est préférable d'utiliser le paramètre de lissage h_4^*.

En guise de conclusion, nous avons mis en évidence, par des simulations, qu'il est préférable, dans l'estimation d'une matrice de transition, d'utiliser les normes matricielles pour le choix du paramètre de lissage plutôt que les autres méthodes classiques de sélection. Notons que cette démarche reste valable pour l'étude d'autres systèmes de files d'attente.

Références

[1] A.K. Basu, D.K. Sahoo, On Berry–Esséen theorem for nonparametric density estimation in Markov sequences, Bull. Inform. Cybernet. 30 (1998) 25–39.
[2] P. Billingsley, Statistical Inference for Markov Processes, University of Chicago Press, Chicago, IL, 1961.
[3] T. Bouezmarni, O. Scaillet, Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, Econom. Theory 21 (2005) 390–412.
[4] S.X. Chen, Probability density functions estimation using Gamma kernels, Ann. Inst. Stat. Math. 52 (2000) 471–480.
[5] L. Kleinrock, Queueing Systems: Theory, Vol. I, John Wiley and Sons, New York, 1975.
[6] A. Laksaci, A. Yousfate, Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 1035–1038.
[7] E. Masry, L. Györfi, Strong consistency and rates for recursive probability density estimators of stationary processes, J. Multivar. Anal. 22 (1987) 79–93.
[8] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat. 33 (1962) 1065–1076.
[9] G.G. Roussas, Nonparametric estimation in Markov processes, Ann. Inst. Stat. Math. 21 (1969) 73–87.