

11

Estimation Non Paramétrique d'un Opérateur de Transition. Cas : La Méthode du Noyau et le Système $GI/M/1/N$

Mouloud CHERFAOUI, Djamil AISSANI and Smail ADJABI

Laboratoire de Modélisation et d'Optimisation des Systèmes LAMOS
Université de Béjaïa 06000, Algérie.

Résumé Dans le présent travail nous avons proposé d'utiliser les normes matricielles afin de sélectionner le paramètre de lissage dans l'estimation d'une matrice de transition d'une chaîne de Markov induite lorsque la loi générale est remplacée par son estimateur à noyau. Une application numérique réalisée sur l'exemple $GI/M/1/N$ confirme l'intérêt d'utilisation des normes matricielles dans l'estimation d'une matrice de transition après une comparaison les procédures classiques de sélection du paramètre de lissage.

Mots clés : Estimation non paramétrique, opérateur de transition, système $GI/M/1/N$, méthode de noyau.

11.1 Introduction et position du problème

Soit T_1, T_2, \dots, T_n un échantillon issu de la variable aléatoire T ayant la densité de probabilité g et la fonction de répartition G . L'estimateur à noyau de Parzen-Rosenblatt classique de la densité $g(t)$ pour tout point $t \in \mathbb{R}$ est donné par :

$$g_h(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - T_i}{h}\right), \quad (11.1)$$

Où K est une fonction de densité symétrique appelée noyau, et h appelé fenêtre ou encore paramètre de lissage. Dans la pratique, l'étape critique dans l'estimation d'une densité par la méthode du noyau est le choix du paramètre h , qui contrôle le lissage de l'estimateur g_h donné par (11.1). Ce problème est largement étudié et pour plus de détail en peut se référer au travail de [7].

Mais supposons que nous désirons estimer la matrice de transition d'une chaîne de Markov induite. Dans ce cas une question se pose : est-il préférable d'estimer le paramètre de lissage h par les méthodes classiques connues dans la littérature ou faut-il prendre en considération la pondération de cette densité ?

La réponse à cette question fait l'objet du présent travail, elle illustrée par des exemples numériques appliqués sur la chaîne de Markov induite du système d'attente $GI/M/1/N$ lorsque la loi des inter-arrivées g remplacé par son estimateur à noyau g_h .

11.2 Choix du noyau et de la méthode de selection

11.2.1 Choix du noyau

Dans plusieurs cas nous sommes face à des situations où la densité g définie sur un support semi-ouvert ou borné. Dans ce cas les noyaux symétriques ne nous fournissent pas de bons estimateurs à cause de la présence du biais aux bornes. Pour remédier ce problème plusieurs méthodes

ont été proposées. On peut citer la méthode du noyau miroir (Schuster), l'utilisation des noyaux asymétrique Gamma, . . . etc.

Noyau de Schuster

L'idée de cette méthode, développée par Deheuvels et Hominal (1979)[?] et Schuster (1985)[?], est d'ajouter une "masse manquante" par réflexion de l'échantillon et qui concerne les données aux frontières. Formellement et sous sa forme la plus simple, il consiste à remplacer $K\left(\frac{t-T_i}{h}\right)$ par $K\left(\frac{t-T_i}{h}\right) + K\left(\frac{t+T_i}{h}\right)$. L'estimateur de la densité est alors de la forme :

$$g_h(t) = \frac{1}{nh} \sum_{i=1}^n \left[K\left(\frac{t-T_i}{h}\right) + K\left(\frac{t+T_i}{h}\right) \right]; \text{ avec } K \text{ est un noyau usuel.} \quad (11.2)$$

Le noyau Gamma

En 2000 Chen [?] propose de remplacer l'estimateur de $g(t)$ par $g_h(t) = \frac{1}{n} \sum_{i=1}^n K(t, h)(T_i)$, ou h est le paramètre de lissage et K est une densité de la loi Gamma donné par :

$$K_{\left(\frac{x}{h}+1, h\right)}(T) = \frac{T^{(t/h)} e^{-(T/h)}}{h^{(t/h)+1} \Gamma((t/h) + 1)}. \quad (11.3)$$

11.2.2 Choix du paramètre de lissage

Dans la littérature, il existe plusieurs méthodes de sélection du paramètre de lissage h qui sont regroupées en deux grandes familles :

- Méthodes plug-in(re-injection) : Choix optimal, Estimateur de Sheather et Jones, . . .
- Méthodes Cross-Validation (Validation Croisée) : UCV (Validation Croisée Non Biaisée), BCV (Validation Croisée Biaisée), . . .

Choix optimal

La décision d'un choix optimal pour le paramètre de lissage suppose la spécification d'un critère d'erreur qui puisse être optimisé. L'optimalité n'est pas un concept absolu : elle est intimement liée au choix du critère, qui peut faire intervenir à la fois la densité inconnue f et l'estimateur g_h (donc h et le noyau K). Dans ce cas, on cherche à minimiser l'Erreur Quadratique Intégrée Moyenne (*MISE*).

$$MISE(g, g_h) = \int \mathbb{E}[g_h(t) - g(t)]^2 dt.$$

Vu la difficulté de manipuler cette dernière quantité, le paramètre de lissage est déterminé d'une manière à minimiser l'Erreur Quadratique Intégrée Moyenne Asymptotique (*AMISE*) qui s'écrit sous forme :

$$AMISE = \frac{h^4}{4} \sigma_K^4 R(g'') + \frac{R(K)}{nh}. \text{ (où } R(f) = \int f^2(t) dt, \text{ Pour toute fonction } f.) \quad (11.4)$$

Ainsi le h optimale est donné par : $h^* = \left[\frac{R(K)}{\sigma_K^4 R(g'')} \right]^{1/5} n^{-1/5}$.

Estimateur de Sheather et Jones

En 1991, Sheather et Jones [?] recommandèrent l'utilisation de l'estimateur naturel $\hat{R}_a(g'')$. En effet, Sheather et Jones choisirent d'estimer $R(g'') = \int_{-\infty}^{+\infty} (g''(x))^2 dt$ par :

$$S(a) = \frac{1}{n(n-1)a^5} \sum_{i=1}^n \sum_{j=1}^n L^{(4)}\left(\frac{t_i - t_j}{a}\right), \quad (11.5)$$

où $L^{(4)}$ désigne la dérivée quatrième d'un noyau suffisamment lisse L et où a est un nouveau paramètre de lissage appelé paramètre pilote. Pour L est un noyau gaussien le paramètre de lissage h_{SJ} est la solution du critère $SJ(h)$ défini par la formule suivante :

$$SJ(\hat{h}) = \left(\frac{1}{2\sqrt{\pi}}\right)^{1/5} S(\hat{\alpha}(\hat{h}))(g'')^{-1/5} n^{-1/5} - \hat{h}, \quad (11.6)$$

avec $\hat{\alpha}(h) = 1.357 \left(\frac{S(a)}{T(b)}\right)^{1/7} h^{5/7}$, $T(b) = \frac{1}{n(n-1)b^7} \sum_{i=1}^n \sum_{j=1}^n L^{(6)}\left(\frac{t_i - t_j}{b}\right)$, $a = 0.920\hat{\lambda}n^{-1/7}$, $b = 0.912\hat{\lambda}n^{-1/9}$, ou $\hat{\lambda}$ représente l'estimateur d'une mesure d'échelle (par exemple l'écart inter

Validation croisée non biaisée

Cette méthode appelée Validation Croisée non Biaisée a été proposée par Rudemo [?] en 1982 et Bowman [?] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$UCV(h) = \int_{\mathbb{R}} [g_h(t) - g(t)]^2 dt - \int_{\mathbb{R}} g^2(t) dt = \int_{\mathbb{R}} g_h^2(t) dt - 2 \int_{\mathbb{R}} g_h(t)g(t) dt.$$

Si le choix du noyau s'est fixé au noyau gaussien la formule du critère $UCV(h)$ sera donné par :

$$UCV(h) = \frac{1}{2n^2 h \sqrt{\pi}} \left(n + 2 \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\frac{(T_i - T_j)^2}{2h}\right) \right) - \frac{2}{\sqrt{2\pi} n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\frac{(T_i - T_j)^2}{2h^2}\right).$$

Validation croisée biaisée

Le critère de validation croisée biaisée, a été introduit par Scott et Terrell [?] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le $UCV(h)$ afin de réduire sa variance. Scott et Trell ont donné également la forme du critère $BCV(h)$ (voir [?]) de plus si le noyau utilisé est gaussien dans ce cas le critère BCV sera donnée par la proposition suivante :

$$BCV(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2 h \sqrt{\pi}} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[\left(\frac{T_i - T_j}{h}\right)^4 - 12 \left(\frac{T_i - T_j}{h}\right)^2 + 12 \right] \exp\left[-\frac{(T_i - T_j)^2}{4h^2}\right]. \quad (11.7)$$

11.3 Description du système $GI/M/1 (FIFO, N)$

Considérons le modèle d'attente avec refus de type $GI/M/1 (FIFO, N)$. Nous supposons que les temps d'inter-arrivées sont indépendants et identiquement distribués issus d'une variable aléatoire de moyenne $1/\lambda$ de fonction de répartition $G(t)$, et les durées de service sont indépendantes et sont identiquement et exponentiellement distribuées avec un taux μ . L'ensemble de états de la chaîne $X_n^{(1)}$ qui représente le nombre de clients dans le système $GI/M/1 (FIFO, N)$ juste avant l'arrivée du $n^{ième}$ client est $\{0, 1, \dots, N\}$, et on a : $X_n^{(1)} = \min(X_{n-1} + 1, N) - D_n$, où D_n est le nombre de départs entre les instants d'arrivée consécutifs T_{n-1} et T_n .

Il est très facile de démontrer que le processus $X_n^{(1)}$ est une chaîne de Markov, et ses probabilités de transition s'écrivent comme suite (voir [?]) :

$$\text{Pour } 0 \leq i \leq N - 1 \quad P_{ij}^{(1)} = \begin{cases} a_i & \text{si } j = 0, \\ b_{i-j+1} & \text{si } 1 \leq j \leq i + 1, \\ 0 & \text{si } i + 1 \leq j \leq N. \end{cases} \quad \text{et pour } i = N \quad P_{ij}^{(1)} = \begin{cases} a_{N-1} & \text{si } j = 0, \\ b_{N-j} & \text{si } 1 \leq j \leq N. \end{cases}$$

Où $a_i = 1 - \sum_{k=0}^i b_k$ et $b_k = \int_0^\infty \frac{e^{-\mu t} (\mu t)^k}{k!} dG(t)$.

11.4 Estimation du paramètre de lissage dans le système $GI/M/1/N$

Le paramètre de lissage dans ce cas sera sélectionné d'une manière qu'il minimise l'une des erreurs suivantes :

$$err_1 = \sup_{i,j} \left[\mathbb{E} \left(\hat{P}_{ij} - P_{ij} \right)^2 \right], \quad err_2 = \sum_{i=1}^N \sum_{j=1}^N \left[\mathbb{E} \left(\hat{P}_{ij} - P_{ij} \right)^2 \right], \quad err_3 = \sup_i \sum_{j=1}^N \left[\mathbb{E} \left(\hat{P}_{ij} - P_{ij} \right)^2 \right], \quad \text{ou } err_4 = \sup_j \sum_{i=1}^N \left[\mathbb{E} \left(\hat{P}_{ij} - P_{ij} \right)^2 \right]$$

sachant que \hat{P}_{ij} est l'estimateur de P_{ij} lorsque on remplace $g(t)$ par son estimateur $g_h(t)$.

11.5 Application numérique

Cette section est consacrée aux résultats numériques obtenus par simulation et les calculs exacts. Enfin les résultats rangés dans les tableaux Tab.11.1 et Tab.11.2 sont obtenus sur le système $M/M/1/N$, ayant comme paramètres $\mu = [10/3 \ 5/3 \ 10/9 \ 5/6 \ 2/3]$, $\lambda = 1$ et $N = 5$, sur des échantillons de différentes tailles et en fonction de la charge du système $\rho = \lambda/\mu$ à l'aide de l'algorithme suivant :

- Étape 1 Fixer les paramètres du système (λ, μ , loi de service, . . .),
- Étape 2 Générer un échantillon de taille n de la loi des inter-arrivées (respectivement loi des durées de service),
- Étape 3 Estimer $h_{classique}^*$ par la méthode classique, et calculer $\hat{P}^{(1)}$ (respectivement $(\hat{P}^{(2)})$),
- Étape 4 Estimer $h_{proposer}^*$ par la norme matricielle, et calculer $\hat{P}^{(1)}$ (respectivement $(\hat{P}^{(2)})$),
- Étape 5 Calculer les erreurs associées à $h_{classique}^*$ et $h_{proposer}^*$.

ρ	n	h_{MISE}^*	Norme 1			Norme 2			Norme 3			Norme 4		
			h_1^*	err_{MISE}^*	err_1^*	h_2^*	err_{MISE}^*	err_2^*	h_3^*	err_{MISE}^*	err_3^*	h_4^*	err_{MISE}^*	err_4^*
0.3	100	0.2029	0.0613	0.0558	0.0807	0.0632	0.0080	0.0144	0.0516	0.0985	0.1955	0.0566	0.0246	0.0403
	150	0.1955	0.0647	0.0487	0.0705	0.0665	0.0043	0.0091	0.0502	0.0824	0.1693	0.0588	0.0188	0.0353
	200	0.1766	0.0675	0.0417	0.0606	0.0768	0.0035	0.0070	0.0612	0.0693	0.1431	0.0656	0.0161	0.0303
	250	0.1469	0.0475	0.0357	0.0566	0.0487	0.0025	0.0062	0.0452	0.0643	0.1382	0.0431	0.0148	0.0283
0.9	100	0.2387	0.1621	0.0384	0.0558	0.1329	0.0035	0.0078	0.1269	0.0533	0.1174	0.1267	0.0145	0.0269
	150	0.1998	0.0899	0.0310	0.0513	0.0923	0.0025	0.0068	0.1026	0.0468	0.0979	0.0962	0.0126	0.0249
	200	0.1323	0.0306	0.0357	0.0555	0.0273	0.0034	0.0077	0.0290	0.0477	0.0965	0.0297	0.0167	0.0274
	250	0.1370	0.0462	0.0225	0.0414	0.0442	0.0014	0.0049	0.0428	0.0399	0.0859	0.0452	0.0100	0.0204
1.5	100	0.1540	0.0550	0.0272	0.0464	0.0477	0.0020	0.0052	0.0416	0.0425	0.0904	0.0423	0.0120	0.0231
	150	0.2755	0.1378	0.0211	0.0429	0.1401	0.0018	0.0054	0.1517	0.0249	0.0546	0.1496	0.0099	0.0207
	200	0.1632	0.0356	0.0259	0.0493	0.0370	0.0019	0.0057	0.0322	0.0325	0.0651	0.0418	0.0119	0.0241
	250	0.1821	0.0779	0.0229	0.0420	0.0778	0.0019	0.0049	0.0636	0.0294	0.0631	0.0737	0.0111	0.0209
300	150	0.1551	0.0620	0.0232	0.0400	0.0642	0.0022	0.0048	0.0630	0.0347	0.0613	0.0658	0.0114	0.0198
	300	0.1658	0.0730	0.0162	0.0320	0.0725	0.0009	0.0027	0.0652	0.0245	0.0530	0.0668	0.0077	0.0158

TABLE 11.1: Comparaison des erreurs et de la valeur du paramètre de lissage de chaque norme

ρ	n	Exacte	N_{MISE}	Norme 1	Norme 2	Norme 3	Norme 4
--------	---	--------	------------	---------	---------	---------	---------

0.3	0.4242	100	0.3599	0.3895	0.3892	0.3906	0.3902
		150	0.3748	0.4060	0.4049	0.4103	0.4086
		200	0.3834	0.4128	0.4104	0.4153	0.4150
		250	0.3822	0.4087	0.4093	0.4104	0.4105
		300	0.3878	0.4090	0.4094	0.4101	0.4102
0.9	2.1948	100	1.8948	2.0614	2.0877	2.0866	2.0952
		150	1.9908	2.1307	2.1273	2.1161	2.1222
		200	1.9422	2.0730	2.0761	2.0762	2.0728
		250	1.9982	2.1189	2.1212	2.1251	2.1206
		300	1.9677	2.0920	2.1011	2.1077	2.1059
1.5	3.5774	100	2.7759	2.9399	2.9357	2.9304	2.9339
		150	3.3806	3.5217	3.5206	3.5257	3.5162
		200	3.3734	3.4860	3.4862	3.5032	3.4907
		250	3.3840	3.4839	3.4824	3.4842	3.4806
		300	3.4090	3.5099	3.5106	3.5178	3.5156

TABLE 11.2: Comparaison du nombre moyen de clients dans le système

Discussion des résultats

A partir des résultats rangés dans le tableau 11.1, on constate que la valeur du AMISE est minime lorsque nous utilisons les normes matricielles quelques soit la taille de l'échantillon des observations des inter-arrivées ainsi que la charge du système ρ , c'est-à-dire les normes matricielles nous offrent de meilleurs estimateurs au sens du AMISE.

Cette constatation est confirmée par la comparaison du nombre moyen de clients dans le système (voir le tableau 11.2) au nombre moyen théorique.

11.6 Conclusion et perspectives

Dans ce travail nous avons soulevé le problème du choix du paramètre de lissage dans l'estimation d'un opérateur de transition d'un système d'attente. A l'aide d'un exemple numérique appliqué sur le système $GI/M/1/N$ (plus exactement le système $M/M/1/N$) nous avons démontré qu'il est préférable d'utiliser les normes matricielles pour le choix du paramètre de lissage que les méthodes de sélection classique (minimisation du MISE).

Le présent travail peut être complété par ce qui suit :

- ✓ Confirmer les résultats précédents par une étude théorique (mathématique)
- ✓ Elaborer des procédures de sélection explicites pour le choix du paramètre de lissage $h^* = \arg \min \|\hat{P} - P\|_v$.
- ✓ Une étude extensive sur d'autre système d'attente ($GI/M/1$, $M/GI/1$, ... etc.).

Références

1. W-R. HEINZELMAN, A. CHANDRAKASAN, and H. BALAKRISHNAN. Energy-efficient communication protocol for wireless sensor networks. In *Proceedings of the IEEE Hawaii International Conference on System Sciences*, pages 3005–3014, Janvier 2000.
2. S. LEE and H.SIN. An energie-efficient distributed unequal clustering protocol for wirless sensor networks. *Proceedings of word Academy of Science Engineering and Technology*, 36, Decembre 2008.
3. N.KHOULALENE. *Regroupement avec Equilibrage de charge dans les Réseaux de Capteurs sans Fil*. Mémoire de magistère en informatique, Université de Béjaia, Algérie, Juin 2007.
4. M. QIN and R. ZIMMERMANN. Vca : An energy-efficient voting-based clustering algorithm for sensor networks. *Journal of Universal Computer Science*, 13(1) :87–109, Janvier 2007.