

Novembre 2001

Numéro 28

Cahiers de l'IMA

Segmentation de séries temporelles à l'aide d'un modèle de mélange

André Berchtold

Institut de Mathématiques Appliquées
Faculté des S.S.P.
Université de Lausanne
BFSH 2
CH-1015 Lausanne

Segmentation de séries temporelles à l'aide d'un modèle de mélange¹

André Berchtold²

Résumé

Les mélanges de lois gaussiennes sont souvent utilisés pour la classification automatique de données. Cependant, dans le contexte des séries temporelles, les modèles utilisés jusqu'à présent étaient généralement basés sur des composantes ayant une espérance et un écart type fixes. L'introduction récente des modèles de la famille MTD a permis de surmonter cette barrière en autorisant une beaucoup plus grande flexibilité dans la modélisation des distributions de probabilité non-gaussiennes. Dans cet article, nous proposons l'utilisation des modèles MTD pour la segmentation de séries temporelles. Après la définition mathématique d'un modèle nommé HMTD généralisant les spécifications proposées précédemment dans la littérature, nous discutons de son estimation, puis nous l'utilisons pour la segmentation de données temporelles.

Mots-clés : Séries temporelles, segmentation, hétéroscédasticité, modèles de mélange, modèles MTD, modèle HMTD.

¹Cet article a été accepté pour une présentation orale dans le cadre des VIIIèmes Rencontres de la Société Francophone de Classification.

²Email : Andre.Berchtold@imaa.unil.ch, Web : www.AndreBerchtold.com

Table des matières

1	Introduction	3
2	Le modèle HMTD	4
3	Estimation	7
4	Segmentation	10
5	Conclusion	12
	Références	14

1 Introduction

Il arrive fréquemment que les données de séries temporelles présentent des comportements fortement hétéroscédastiques pouvant se manifester notamment par une segmentation de l'ensemble de la série en plusieurs zones nettement différenciées, ou encore par de soudaines zones de forte activité au sein d'une période de faible intensité. De telles données sont extrêmement courantes dans le domaine financier (cours des actions, indices boursiers, ...), mais elles se rencontrent également dans les assurances (modélisation des demandes de remboursement) ou lors de l'étude de phénomènes physiques (données sismiques, régime des vents, ...). Les figures I et II présentent des exemples de ce type de données.

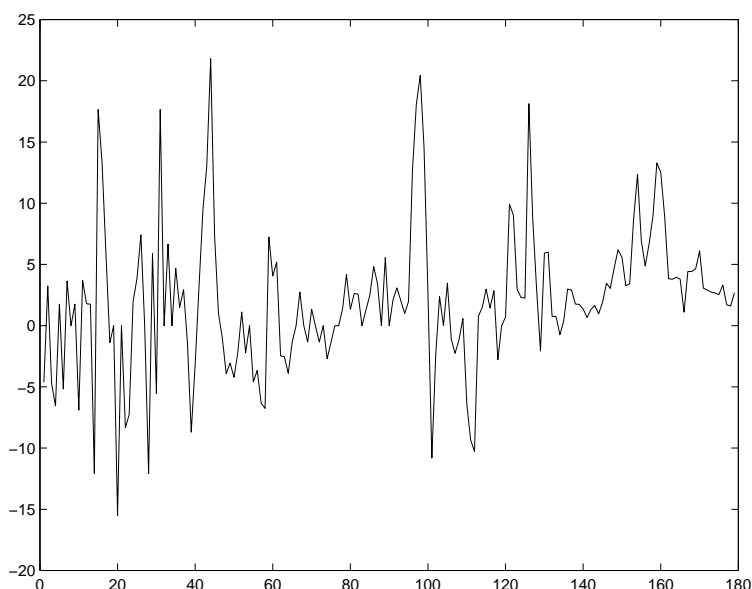


Figure I. *Taux annuel d'inflation aux Etats-Unis entre 1821 et 1999. Nous pouvons observer une très forte variabilité des données avec des zones de forte activité suivies de périodes durant lesquelles l'inflation est relativement faible.*

Lorsque nous sommes confrontés à de telles séries temporelles, il est essentiel de pouvoir identifier les différents types de comportements rencontrés. Par exemple, en finance, il est crucial de déterminer si le cours d'une action est plutôt stable ou au contraire extrêmement variable avant d'effectuer une décision de vente ou d'achat. De même, l'étude de données sismiques pourra permettre de prédire si une catastrophe naturelle est plus ou moins probable dans un avenir proche.

Le problème posé consiste donc à décomposer une série temporelle en un nombre fini de sous-séries se succédant les unes aux autres et présentant

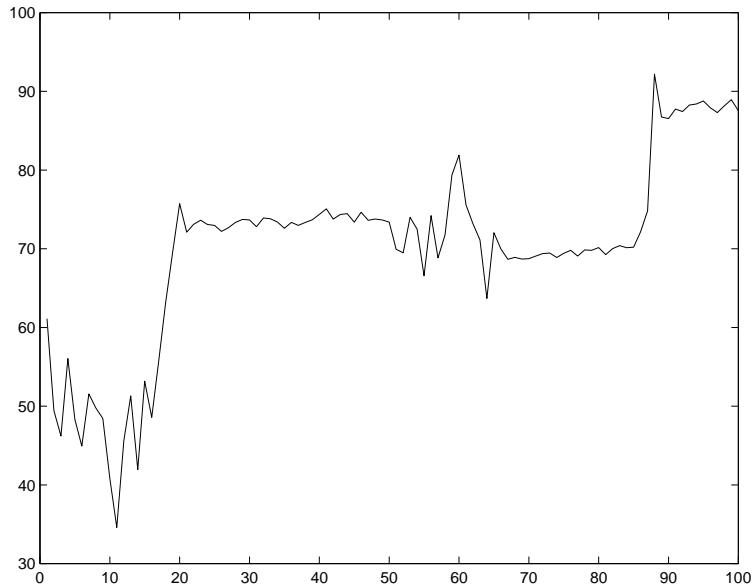


Figure II. *Cette figure présente une série de 100 données simulées où l'on peut remarquer que des séquences de données relativement stables sont entrecoupées par des zones à forte variance.*

des comportements nettement différenciés. Dans cet article, nous commencerons par présenter un modèle de mélange de lois gaussiennes particulièrement adapté à la représentation de séries temporelles. Nous discuterons ensuite de l'estimation de ce modèle, puis de son utilisation en classification.

2 Le modèle HMTD

Les séries temporelles qui nous intéressent présentent des comportements pouvant être très différents d'une période d'observation à une autre. Il est donc possible de supposer qu'une série observée n'a pas été générée par un seul modèle évoluant au cours du temps, mais au contraire par plusieurs modèles différents se succédant. L'utilisation d'un modèle de mélange pour modéliser la série temporelle dans son intégralité est ainsi justifiée.

Nous pouvons noter par ailleurs qu'en raison de l'extrême variabilité des données qui nous intéressent, une estimation ponctuelle de la valeur future prise par une variable, basée sur l'utilisation d'un modèle AR, ARCH ou GARCH (voir par exemple [BOL 1992] pour une revue de ces modèles), n'a que peu de chance de coïncider avec la réalité. Ce problème peut être résolu en ne cherchant pas à estimer la valeur exacte que prendra la variable dans le futur, mais plutôt la distribution de probabilité des différentes valeurs possibles. Le

modèle que nous présentons ici est conçu en ce sens.

Soit une variable continue X_t prenant valeur dans \mathbb{R} et pour laquelle nous disposons de n observations successives. Nous notons X_{t-a}^{t-1} les observations passées de cette variable entre les périodes $t - a$ et $t - 1$. Supposons que les observations de cette variable aient été générées par k sous-modèles différents. Le modèle HMTD (*Heteroskedastic Mixture Transition Distribution model*), estimant la distribution de probabilité de X_t en fonction du passé de cette variable, s'écrit alors

$$F(x_t|x_1^{t-1}) = \sum_{g=1}^k \lambda_g G(x_t|x_{t-r_g}^{t-1})$$

où $F(x_t|x_1^{t-1})$ est la fonction de répartition de la variable X_t conditionnelle à son passé, $G(x_t|x_{t-r_g}^{t-1})$ est une fonction de répartition de la variable X_t conditionnelle à une partie de son passé (de x_{t-r_g} à x_{t-1} , $r_g \geq 1$), et où λ_g est le poids de la g -ième composante du modèle, avec

$$\begin{aligned} \sum_{g=1}^k \lambda_g &= 1 \\ \lambda_g &> 0, \quad \forall g \end{aligned}$$

Dans cet article, nous spécifions les fonctions de répartition G comme étant des distributions gaussiennes :

$$G(x_t|x_{t-r_g}^{t-1}) = \Phi\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right)$$

Afin de tenir compte de la dépendance entre les observations successives de la série temporelle, l'espérance et l'écart type de chaque composante gaussienne sont définis comme des fonctions du passé de X_t . L'espérance de la g -ième composante est spécifiée par le modèle autorégressif suivant :

$$\mu_{g,t} = \varphi_{g,0} + \sum_{i=1}^{p_g} \varphi_{g,i} x_{t-i}, \quad p_g \geq 1 \quad (1)$$

et l'écart type peut être spécifié alternativement par

$$\sigma_{g,t} = \sqrt{\theta_{g,0} + \sum_{j=1}^{q_g} \theta_{g,j} x_{t-j}^2}, \quad q_g \geq 1 \quad (2)$$

ou par

$$\sigma_{g,t} = \sqrt{\theta_{g,0} + \sum_{j=1}^{q_g} \theta_{g,j} (x_{t-j} - \bar{x}_{t-q_g}^{t-1})^2}, \quad q_g \geq 2 \quad (3)$$

où

$$\bar{x}_{t-q_g}^{t-1} = \frac{1}{q_g} \sum_{j=1}^{q_g} x_{t-j}$$

avec

$$\begin{aligned} \theta_{g,0} &> 0, \quad \forall g \\ \theta_{g,j} &\geq 0, \quad \forall g, \quad j = 1, \dots, q_g \end{aligned}$$

La formulation de l'écart type donnée par l'équation (2) présente l'avantage d'être utilisable quel que soit le nombre de retards désiré, de un à l'infini. En revanche, cette formulation a le défaut de ne pas comparer chaque retard à leur moyenne comme dans un vrai écart type. Ainsi, deux séries de données similaires, la seconde prenant des valeurs augmentées systématiquement de dix unités par rapport à la première, aboutiront à des estimations différentes de l'écart type, ce qui n'est pas forcément souhaitable. La seconde formulation de l'écart type (équation (3)) ne souffre pas de ce problème puisque chaque retard est comparé avec la moyenne de tous les retards considérés. En revanche, il est nécessaire d'avoir au moins deux retards. Une troisième possibilité consisterait en le remplacement de $\bar{x}_{t-q_g}^{t-1}$ dans l'équation (3) par l'espérance $\mu_{g,t}$ donnée par l'équation (1). Cependant, ainsi que nous l'avons noté précédemment, l'espérance d'un modèle autorégressif étant rarement une bonne prédiction du comportement futur d'une série temporelle présentant un comportement extrêmement variable, nous déconseillons cette approche. Finalement, pour des applications particulières du modèle, il est toujours possible de définir une ou plusieurs composantes ayant soit une espérance fixe ($p_g=0$), soit un écart type fixe ($q_g=0$), soit même les deux.

Soit $r = \max\{p_1, \dots, p_k, q_1, \dots, q_k\}$, l'ordre maximum de dépendance utilisé pour modéliser l'espérance ou l'écart type d'une composante du modèle et soit f la fonction de densité d'une loi gaussienne. La vraisemblance de x_{r+1}, \dots, x_n conditionnelle à x_1, \dots, x_r s'écrit alors

$$\ell = \prod_{t=r+1}^n \sum_{g=1}^k \lambda_g f\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right)$$

et la log-vraisemblance correspondante s'écrit

$$L = \sum_{t=r+1}^n \log \left[\sum_{g=1}^k \frac{\lambda_g}{\sigma_{g,t}} \exp \left\{ -\frac{(x_t - \mu_{g,t})^2}{2\sigma_{g,t}^2} \right\} \right] - \frac{n-r}{2} \log(2\pi)$$

Le modèle HMTD appartient à la famille des modèles MTD (*Mixture Transition Distribution model*) qui ont tout d'abord été proposés dans le cas de données discrètes par Raftery pour la modélisation de chaînes de Markov d'ordre supérieur à 1 ([RAF 1985]). Dans ce contexte, le principe de base consiste à utiliser une composante différente pour représenter le lien entre chacun des retards du modèle et le présent. Le modèle a ensuite été généralisé au cas de variables continues sous le nom de GMTD (*Gaussian Mixture Transition Distribution model*), en conservant le principe selon lequel chaque retard n'intervient que dans une seule composante du modèle ([LE 1996]). Une généralisation plus avancée a été proposée dans [WON 2000] sous le nom de MAR (*Mixture Autoregression model*), l'idée étant de faire intervenir tous les retards dans la modélisation de l'espérance de chaque composante et donc d'utiliser un modèle similaire à celui de l'équation (1).

Les modèles GMTD et MAR ont en commun le fait de ne modéliser que l'espérance de chaque composante, l'écart type étant toujours constant. Le modèle HMTD est donc une étape supplémentaire de généralisation puisqu'il permet d'exprimer non-seulement l'espérance, mais aussi l'écart type de chaque composante gaussienne comme une fonction du passé de la variable observée, ce qui se traduit par une flexibilité d'utilisation bien supérieure. A l'inverse, en fixant à zéro une partie des paramètres du modèle HMTD, il est possible de retrouver les différentes spécifications proposées dans [LE 1996] et [WON 2000].

3 Estimation

Le modèle HMTD étant un mélange de lois gaussiennes, l'utilisation de l'algorithme EM (*Expectation Maximization*) pour estimer les paramètres par maximisation de la log-vraisemblance est une approche logique. Cet algorithme, décrit notamment dans [MCL 1996], considère l'estimation d'un modèle de mélange comme un problème de données incomplètes. Au temps t , nous connaissons la valeur de l'observation X_t , mais nous ne savons pas quelle composante du modèle a servi à la générer. Nous considérons alors un vecteur indicateur Z_t de dimension k dont le g -ième élément prend la valeur un si l'observation a été générée par la g -ième composante du modèle et zéro sinon.

Au temps t , les données complètes se composent donc à la fois de l'observation X_t et du vecteur indicateur Z_t , le contenu exact de ce dernier étant

inconnu. La vraisemblance des données complètes s'écrit alors

$$\ell_c = \prod_{t=r+1}^n \prod_{g=1}^k \lambda_g^{z_{g,t}} f\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right)^{z_{g,t}}$$

et la log-vraisemblance complète correspondante est donnée par

$$\begin{aligned} L_c &= \sum_{g=1}^k \log(\lambda_g) \sum_{t=r+1}^n z_{g,t} - \sum_{t=r+1}^n \sum_{g=1}^k z_{g,t} \log(\sigma_{g,t}) \\ &\quad - \sum_{t=r+1}^n \sum_{g=1}^k z_{g,t} \frac{(x_t - \mu_{g,t})^2}{2\sigma_{g,t}^2} - \frac{n-r}{2} \log(2\pi) \end{aligned} \quad (4)$$

L'algorithme EM procède par itération des deux étapes suivantes :

1. Durant l'étape E, les valeurs des paramètres du modèle sont considérées comme étant fixes et l'espérance de chaque élément $z_{g,t}$ du vecteur Z_t est calculée comme

$$\tilde{z}_{g,t} = \frac{\frac{\lambda_g}{\sigma_{g,t}} f\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right)}{\sum_{g=1}^k \frac{\lambda_g}{\sigma_{g,t}} f\left(\frac{x_t - \mu_{g,t}}{\sigma_{g,t}}\right)}$$

2. Durant l'étape M, les espérances $\tilde{z}_{g,t}$ sont introduites dans l'équation (4) et les paramètres du modèle sont réestimés par maximisation de la log-vraisemblance des données complètes. Les poids des composantes sont réestimés par

$$\hat{\lambda}_g = \frac{\sum_{t=r+1}^n \tilde{z}_{g,t}}{n-r}$$

Pour chaque composante du modèle, les paramètres φ entrant dans le calcul de l'espérance doivent satisfaire les équations suivantes :

$$\sum_{t=r+1}^n \frac{z_{g,t} x_t w_j}{\sigma_{g,t}^2} = \sum_{s=0}^{p_g} \hat{\varphi}_{g,s} \sum_{t=r+1}^n \frac{\tilde{z}_{g,t} w_j w_s}{\sigma_{g,t}^2}, \quad j = 0, \dots, p_g$$

où

$$w_j = \begin{cases} 1, & \text{si } j = 0 \\ x_{t-j}, & \text{sinon} \end{cases}$$

Finalement, les paramètres θ entrant dans le calcul de l'écart type de chaque composante doivent satisfaire les équations suivantes :

$$\begin{aligned} & \left[\sum_{t=r+1}^n \left(\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s \right)^2 \right] \cdot \left[\sum_{t=r+1}^n \tilde{z}_{g,t} u_j \right] \\ &= \left[\sum_{t=r+1}^n \left(\sum_{s=0}^{q_g} \hat{\theta}_{g,s} u_s \right) \right] \cdot \left[\sum_{t=r+1}^n \tilde{z}_{g,t} (x_t - \mu_{g,t})^2 u_j \right], \forall j \quad (5) \end{aligned}$$

où

$$u_j = \begin{cases} 1, & \text{si } j = 0 \\ x_{t-j}^2, & \text{si } j \neq 0 \text{ et si l'écart type est spécifié par (2)} \\ (x_{t-j} - \bar{x}_{t-q_g}^2)^2, & \text{si } j \neq 0 \text{ et si l'écart type est spécifié par (3)} \end{cases}$$

L'équation (5) montre qu'il n'est généralement pas possible de réestimer de façon unique les paramètres θ . Dans le cas particulier où l'écart type de chaque composante est supposé constant et où seuls les paramètres $\theta_{g,0}$ sont donc différents de zéro, nous obtenons

$$\hat{\theta}_{g,0} = \frac{\sum_{t=r+1}^n \tilde{z}_{g,t} (x_t - \mu_{g,t})^2}{\sum_{t=r+1}^n \tilde{z}_{g,t}}, \quad g = 1, \dots, k$$

Dans les autres cas, il n'est pas possible de dériver un estimateur du maximum de vraisemblance. Afin de rester tout-de-même dans le cadre général d'un algorithme EM et d'assurer sa convergence, nous devons alors employer une autre méthode de réestimation des paramètres θ assurant l'obtention d'une log-vraisemblance complète égale ou supérieure à celle atteinte à la fin de l'itération précédente. La méthode suivante, introduite dans [BER 2001c] pour l'estimation de la version discrète du modèle MTD, est une possibilité. Nous considérons séparément chacune des composantes du modèle et l'algorithme suivant est appliqué successivement sur chacune d'entre-elles :

- (a) La dérivée partielle de la log-vraisemblance complète est calculée par rapport à chaque paramètre θ de la composante choisie.
- (b) Le paramètre θ correspondant à la plus grande dérivée partielle prise en valeur absolue est augmenté (dérivée positive) ou diminué (dérivée négative) d'une valeur δ .
- (c) La log-vraisemblance complète est recalculée :
 - Si la nouvelle log-vraisemblance est supérieure ou égale à la valeur précédente, la nouvelle valeur du paramètre θ est acceptée et l'on passe à la réestimation des paramètres de la composante suivante.
 - Sinon, la nouvelle valeur du paramètre θ est rejetée, la valeur de δ est divisée par deux et l'on retourne au début de l'étape (b).

Cette méthode de réestimation des paramètres θ ne peut bien entendu pas être considérée comme optimale puisque seule une partie des paramètres est réestimée à chaque étape (un paramètre θ par composante). Cependant, en pratique, cette méthode s'avère extrêmement fiable et permet d'assurer la convergence de l'algorithme EM.

L'espace des solutions des modèles de type MTD est parfois très complexe, ce qui engendre un autre problème : Si l'algorithme EM est assuré de converger vers un maximum de la log-vraisemblance complète, rien n'assure que ce maximum soit le maximum global plutôt qu'un maximum local. Plusieurs méthodes ont été proposées afin de remédier à cela :

- plusieurs utilisations successives de l'algorithme EM à partir de valeurs de départ différentes ([BIE 2000]) ;
- combinaison de l'algorithme EM avec un algorithme génétique ([BER 2001a], [BER 2001b]) ;
- utilisation simultanée de l'algorithme EM et de la fonction gradient ([BOH 2001]).

Chacune de ces méthodes permet effectivement d'augmenter la probabilité d'atteindre l'optimum global du problème considéré. En revanche, nous n'avons pas connaissance à ce jour d'une étude comparant les efficacités respectives de ces trois approches.

4 Segmentation

Une fois le modèle HMTD estimé, il peut être utilisé afin de déterminer quelle composante a généré chaque donnée observée, ce qui revient à segmenter la série temporelle en un nombre plus ou moins grand de sous-séries. Notre intérêt est bien entendu d'obtenir des sous-séries relativement longues, de façon à dégager une structure générale sous-jacente à la série temporelle.

Par le passé, les mélanges de lois gaussiennes ont souvent été associés à la classification de données ([MCL 1988]), mais il s’agissait toujours de modèles dont chaque composante avait une espérance et un écart type fixes. Par ailleurs, les modèles de la famille MTD ont été surtout utilisés dans le passé pour la représentation la plus précise possible de comportements non-gaussiens. L’originalité de cet article consiste en leur utilisation pour la segmentation de données. Dans cette optique, il est crucial d’éviter une trop grande spécialisation du modèle par rapport à un échantillon de données. Il apparaît notamment que, dans l’optique de la classification, un modèle plus simple, utilisant par exemple moins de retards pour modéliser l’écart type de chaque composante, peut se révéler plus efficace en raison d’une moins grande influence de petites perturbations locales au profit d’une meilleure prise en compte des grandes tendances déterminant l’évolution des données.

En pratique, la détermination de la composante ayant le plus probablement généré chaque observation se fait sur la base des vecteurs indicateurs Z_t . Chaque observation est attribuée à la composante ayant la plus forte probabilité dans ce vecteur. Le résultat peut alors être présenté sous la forme d’un graphique faisant correspondre chaque observation à l’une des composantes du modèle.

A titre d’illustration, nous avons utilisé les données de la figure II. Cette série temporelle présente une situation dans laquelle deux comportements bien distincts se succèdent : des périodes de forte variabilité entrecourent des périodes beaucoup plus stables. Un modèle à deux composantes, avec un ordre de dépendance égal à trois pour les espérances et des variances constantes, permet une bonne séparation des deux types d’observations. Le modèle a les paramètres suivants :

Composante	λ	φ_0	φ_1	φ_2	φ_3	θ_0
1	0.5812	-1.2359	0.6765	0.1867	0.1542	0.3007
2	0.4188	4.9027	0.7126	0.1640	0.0648	40.2413

Ce modèle est caractérisé par deux composantes aux variances très différentes (0.3007 et 40.2413) correspondant aux deux types d’observations apparaissant sur la figure II. La vraisemblance obtenue par ce modèle est égale à -230.77. Elle a été calculée sur les cent observations de la série telles qu’elles apparaissent sur la figure II, conditionnellement à trois observations initiales qui n’ont pas été représentées. La figure III présente la segmentation obtenue à l’aide de ce modèle. Il est à noter que plusieurs modèles dans lesquels l’écart type est également une fonction du passé de la variable observée obtiennent une vraisemblance nettement plus grande que celle du modèle que nous avons retenu ici. Cependant, ces modèles sont moins adaptés à la segmentation, car

ils tendent à être trop influencés par une légère modification de la variabilité des données au sein d'une même zone générale de variance. En revanche, lorsque le but de la modélisation est de pouvoir ensuite estimer le plus précisément possible la distribution de probabilité des événements futurs, ce type de modèle se révèle extrêmement performant.

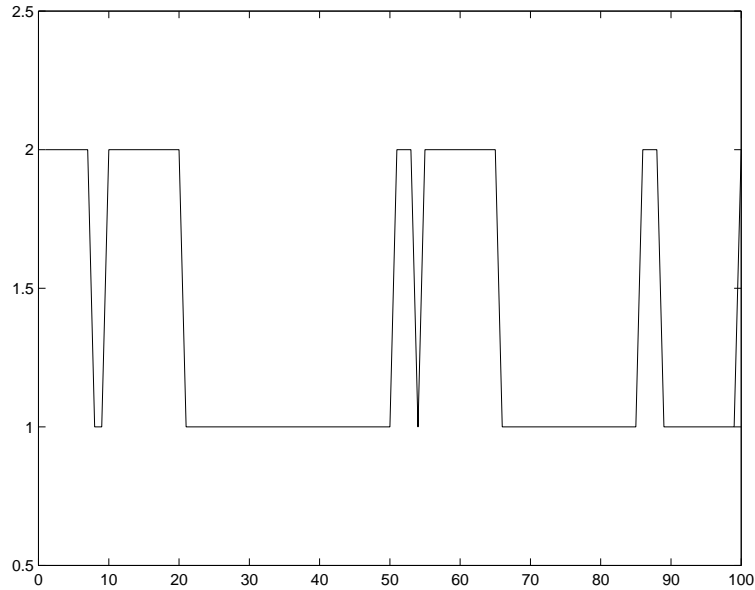


Figure III. Cette figure indique quelle composante du modèle de mélange est la plus probable à chaque instant pour les observations de la figure II. Par exemple, la composante 2 est utilisée pour le début de la série et la composante 1 apparaît pour la première fois à la huitième observation. Il apparaît clairement qu'à de rares exceptions près, le modèle segmente correctement les données entre zones à faible variance (composante 1) et zones à forte variance (composante 2).

5 Conclusion

Dans cet article, nous avons défini un modèle de mélange spécialisé dans l'analyse des séries temporelles et nous avons montré qu'il pouvait être utilisé pour la segmentation de telles données en zones successives ayant des caractéristiques différentes. Bien évidemment, notre traitement n'a pas la prétention d'être exhaustif et de nombreux points méritent d'être étudié plus avant. En particulier, les questions suivantes nécessitent des développements :

- Le modèle actuel est basé sur des composantes gaussiennes. L'utilisation d'autres formes de distributions pourrait augmenter encore la diversité

des situations pouvant être représentées, et donc le champ d'application du modèle.

- L'estimation du modèle n'est pas encore optimale, notamment en ce qui concerne les paramètres relatifs à la variance de chacune des composantes. Des progrès restent à faire dans ce domaine. De plus, une comparaison rigoureuse des différentes stratégies d'optimisation possibles, pouvant par exemple faire appel à des combinaisons d'algorithmes (algorithme génétique, algorithme EM, ...), est nécessaire.
- Ainsi que nous l'avons évoqué dans cet article, le modèle de mélange représentant le plus précisément une série temporelle donnée n'est pas forcément le modèle le mieux adapté à la segmentation de cette même série. La notion de "modèle apte à la segmentation" reste à définir.

Malgré l'importance des développements restant à faire, nous pensons que le modèle HMTD actuel représente déjà une solution performante pour la segmentation de séries temporelles. De par sa flexibilité, il est capable de s'adapter rapidement à des situations très différentes tout en restant dans un cadre conceptuel connu (modèles de mélange, algorithme EM, ...). Il est donc aisé de l'implémenter et de l'utiliser, seul ou en conjonction avec d'autres outils statistiques.

Références

- [BER 2001a] BERCHTOLD A. (2001) Markov chains computation for homogeneous and non-homogeneous data : MARCH 1.1 user's guide, *Journal of Statistical Software*, vol. 6, p. 1–82.
- [BER 2001b] BERCHTOLD A. (2001) Mixture Modeling of Non-Gaussian Probability Distributions, *Proceedings of the 10th International Symposium on Applied Stochastic Models and Data Analysis*, p. 188–193.
- [BER 2001c] BERCHTOLD A. (2001) Estimation in the Mixture Transition Distribution Model, *Journal of Time Series Analysis*, vol. 22, p. 379–397.
- [BIE 2000] BIERNACKI C., CELEUX G., GOVAERT, G. (2000) Stratégies algorithmiques pour maximiser la vraisemblance dans les modèles de mélange, *Actes des XXXIIèmes Journées de Statistique*, p. 140–143.
- [BOH 2001] BÖHNING D. (2001) The Potential of Recent Developments in Nonparametric Mixture Distributions, *Proceedings of the 10th International Symposium on Applied Stochastic Models and Data Analysis*, p. 6–13.
- [BOL 1992] BOLLERSLEV T., CHOU R.Y., KRONER K.F. (1992) ARCH Modeling in Finance, A Review of the Theory and Empirical Evidence, *Journal of Econometrics*, vol. 52, p. 5–59.
- [LE 1996] LE N.D., MARTIN R.D., RAFTERY A.E. (1996) Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models, *Journal of the American Statistical Association*, vol. 91, p. 1504–1515.
- [MCL 1988] MCLACHLAN G.J., BASFORD K.E. (1988) *Mixture Models : Inference and Applications to Clustering*, Marcel Dekker, Inc.
- [MCL 1996] MCLACHLAN G.J., KRISHNAN T. (1996) *The EM Algorithm and Extensions*, John Wiley & Sons.
- [RAF 1985] RAFTERY A.E. (1985) A model for high-order Markov chains, *Journal of the Royal Statistical Society B*, vol. 47, p. 528–539.
- [WON 2000] WONG C.S., LI W.K. (2000) On a mixture autoregression model, *Journal of the Royal Statistical Society B*, vol. 62, p. 92–115.