HAL
open science

Capushe : package de sélection de modèle

Vincent Brault, Jean-Patrick Baudry, Cathy Maugis-Rabusseau, Bertrand Michel

To cite this version:

Vincent Brault, Jean-Patrick Baudry, Cathy Maugis-Rabusseau, Bertrand Michel. Capushe : package de sélection de modèle. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717565

HAL Id: hal-00717565

https://hal.science/hal-00717565v1

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Package de sélection de modèle CAPUSHE

J-P. Baudry[a], **V. Brault**[b], **C. Maugis-Rabusseau**[c] and **B. Michel**[d]

[a]Laboratoire de Statistique Théorique et Appliquée
Université Paris 6
4 place Jussieu, 75252 Paris cedex 05
jean-patrick.baudry@upmc.fr

[b]Laboratoire de Mathématiques UMR 8628
Université Paris-Sud 11
F-91405 Orsay cedex
vincent.brault@math.u-psud.fr

[c]Institut de Mathématiques de Toulouse
INSA de Toulouse, Université de Toulouse
118 route de Narbonne F-31062 Toulouse Cedex 9
cathy.maugis@insa-toulouse.fr

[d]Laboratoire de Statistique Théorique et Appliquée
Université Paris 6
4 place Jussieu, 75252 Paris cedex 05
bertrand.michel@upmc.fr

Mots clefs : Estimation de pente guidée par les données – Saut de dimension – Sélection de modèle – Pénalisation – Heuristique de pente

La sélection de modèle est un paradigme général incluant de nombreux problèmes de statistiques. L'une des approches les plus populaires est la minimisation d'un critère pénalisé. Birgé et Massart [2] ont proposé une méthode de calibration où les pénalités sont connues à un facteur multiplicatif près : l'heuristique de pente. Des travaux théoriques valident cette méthode heuristique dans certaines situations et plusieurs articles montrent un comportement prometteur dans des cas pratiques.
Deux méthodes sont actuellement utilisées pour la calibration de cette pénalité : le saut de dimension (figure 2) et l'estimation de pente guidée par les données (figure 1). Baudry, Maugis et Michel [6] ont proposé un package matlab implémentant ces deux méthodes avec des interfaces graphiques. Nous présentons ici le package CAPUSHE pour le logiciel `R` proposant leurs implémentations avec la possibilité de représenter graphiquement les résultats pour les valider. Durant cet exposé, nous ferons un cours rappel de la théorie puis nous expliquerons l'utilisation de nos fonctions à l'aide de l'exemple proposé dans le package. Nous détaillerons également les possibilités que possède l'utilisateur pour évaluer et améliorer la qualité des résultats.
Enfin, nous serons très attentifs aux questions ou suggestions d'améliorations pour le package.

Références

[1] Birgé, L. et Massart, P. (2001). Gaussian model selection. *Journal of the European Mathematical Society*, 3(3):203-268.
[2] Birgé, L. et Massart, P. (2006). Minimal penalties for gaussian model selection. *Probability Theory and Related Fields*, 138(1-2):33-73.

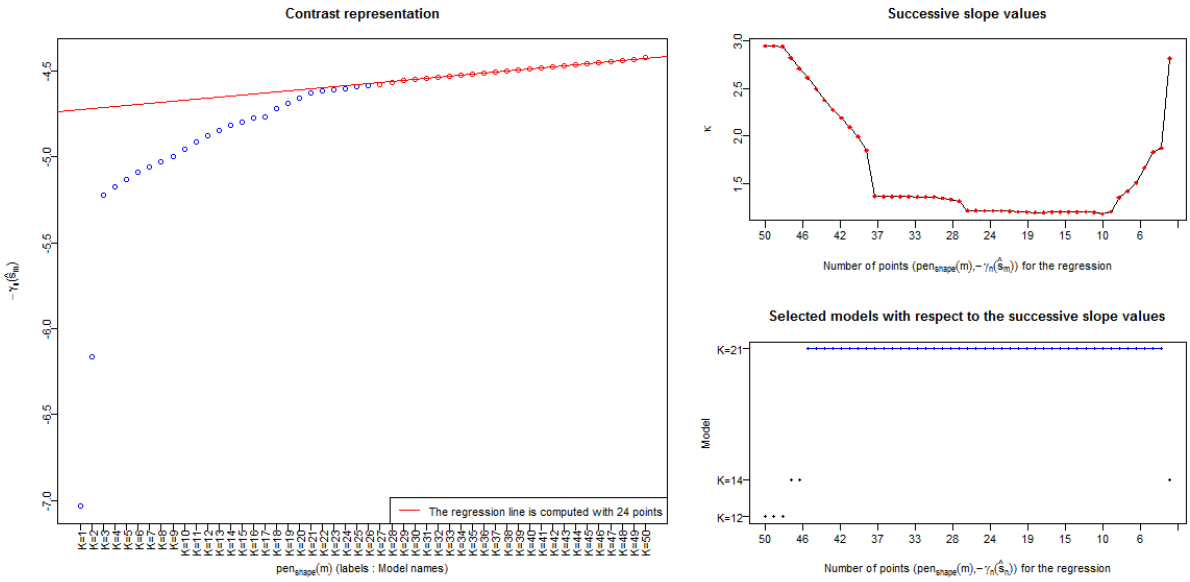

Figure 1: *Représentation graphique proposée par le package* CAPUSHE *pour vérifier la qualité des résultats obtenus par la méthode d'estimation de pente guidée par les données.*

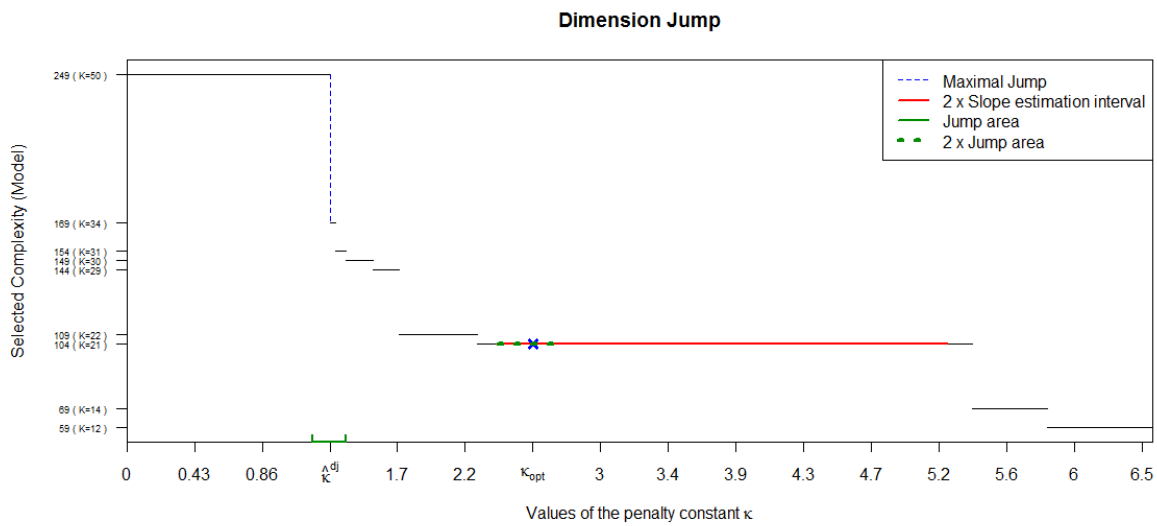

Figure 2: *Représentation graphique proposée par le package* CAPUSHE *pour vérifier la qualité des résultats obtenus par la méthode du saut de dimension.*

[3] Lebarbier, E. (2005). Detexting multiple change-points in the mean of gaussian process by model selection. *Signal Processing*, 85(4):717-736.

[4] Massart, P. (2007). *Concentration Inequalities and Model Selection. École d'été de Probabilités de SAint-Flour 2003. Lecture Notes in Mathematics.* Springer.

[5] Maugis, C. et Michel, B. (2009). A non-asymptotic penalized criterion for gaussian mixture model selection. *ESAIM*: P & S URL http://hal.inria.fr/docs/00/28/50/31/PDF/RR-6549.pdf

[6] Baudry, J.-P., Maugis, C. and Michel, B. (2011) Slope Heuristics: Overview and Implementation. *Statistics and Computing*, Vol 22(2), 455-470.