1

Statistique bayésienne

Session organisée par Vincent Rivoirard

L'objet est de cette session est de présenter un large panorama de l'utilisation de l'approche bayésienne en statistique. Plus précisément, les exposés de cette session déclineront différentes méthodes de l'inférence bayésienne dans les domaines de l'estimation fonctionnelle, de la statistique semi-paramétrique et de la classification. A travers la réalisation de simulations ou l'utilisation de diverses approches théoriques (théorie minimax ou maxiset, étude de propriétés LAN ou de l'efficacité du second ordre,...), les performances des procédures bayésiennes proposées seront mesurées et comparées à celles d'estimateurs fréquentistes. L'accent sera également mis sur la modélisation à travers de nombreuses discussions portant sur le choix de l'a priori.

Références :

[1] J. K. Ghosh and R. V. Ramamoorthi, *Bayesian nonparametrics*, (2003), Springer-Verlag, New York.

[2] *Bayesian inference in wavelet-based models,* (1999), Eds Peter Müller and Brani Vidakovic, Lecture Notes in Statistics, 141. Springer-Verlag, New York.

[3] A. van der Vaart, *Semiparametric statistics. Lectures on probability theory and statistics (Saint-Flour 1999)*

Adresse de l'organisateur :

Vincent Rivoirard
Laboratoire de Mathématique UMR 8628
Equipe Probabilités, Statistique et Modélisation
Université Paris-Sud 11
Bâtiment 425
91405 Orsay Cedex
France
E-mail : `Vincent.Rivoirard@math.u-psud.fr`
<`http://www.math.u-psud.fr/~rivoirar`>

Quelques résultats sur les tests bayésiens d'hypothèses précises dans un contexte non paramétrique

par **Judith Rousseau**

Les problèmes de tests d'hypothèses nulles précises de la forme : $\theta = \theta_0$ ou $\theta \in \Theta_0$ lorsque Θ_0 est un sous ensemble de Θ de mesure de Lebesgue nulle, sont souvent des approximations de problèmes de tests de la forme : θ est *proche* de θ_0 ou le sous modèle Θ_0 est une approximation *raisonnable* de la réalité. Dans la plupart des cas, il est difficile d'appréhender a priori la notion de proximité ci-dessus, ce qui explique l'utilisation des hypothèses précises.

Dans le cadre bayésien, une des procédures de test les plus utilisées est le facteur de Bayes. Il a été prouvé que le facteur de Bayes associé à une hypothèse ponctuelle pouvait approcher le facteur de Bayes associé à son équivalent moins précise. Nous montrons que lorsque le nombre d'observations augmente ce résultat n'est plus valable, à moins que la tolérance soit extrêmement petite.

Puis nous étudions plus particulièrement le problème de tests d'adéquation. Dans ce cadre nous interprètons deux types de procédures de tests bayésiennes : les facteurs de Bayes et les tests basés sur des distances à l'hypothèse nulle, calibrée par une p-value.

Adresse :

Judith Rousseau
CEREMADE
Université Paris 9 Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
France
E-mail : `Rousseau@ceremade.dauphine.fr`
<`http://www.ceremade.dauphine.fr/~rousseau`>

Représentations probabilistes de la méthode des k plus proches voisins

par **Jean-Michel Marin**

En classification supervisée, on dispose d'un échantillon d'apprentissage constitué d'individus observés sur une variable catégorielle d'intérêt et différentes covariables. Cet échantillon est utilisé afin de construire une fonction de décision qui permet d'affecter tout nouvel individu observé sur les mêmes covariables. La méthode des k plus proches voisins consiste à appliquer une règle d'affectation majoritaire sur les k plus proches voisins dans l'espace des covariables. Traditionnellement, la valeur de k est choisie par validation croisée. Nous présentons des modèles probabilistes associés à cette heuristique. Nous en déduisons des outils statistiques pour le choix de k et du meilleur sous-ensemble de covariables. Une inférence bayésienne sera privilégiée et comparée à la méthodologie du maximum de vraisemblance.

Adresse :

Jean-Michel MARIN
Laboratoire de Mathématique UMR 8628
Equipe Probabilités, Statistique et Modélisation
Université Paris-Sud 11
Bâtiment 442
91405 Orsay Cedex
France
E-mail : jean-michel.marin@math.u-psud.fr
<http://www.ceremade.dauphine.fr/~marin>

Estimation bayésienne de modèles multifractaux

par Fabrice Gamboa et **Jean-Michel Loubes**

Les fonctions multifractales sont utilisées pour modéliser des signaux dont la régularité varie fortement et très rapidement. Des exemples de tels signaux apparaissent en théorie des fluides ou lors de l'étude du trafic routier par exemple. Un modèle de simulation couramment utilise est de décrire ces fonctions a partir des propriétés de leurs coefficients d'ondelettes. Plus précisément il existe un paramètre de lacunarité η et un paramètre d'intensité α tels que les coefficients sont des réalisations de

$$w_{jk} \sim 2^{(\eta-1)j}\mathcal{N}(2^{-\alpha j}, \sigma_j) + (1 - 2^{(\eta-1)j})\mathcal{N}(0, \sigma_j).$$

Nous construirons un estimateur bayésien d'une réalisation de ces fonctions observée dans un cadre de bruit blanc et montrerons les propriétés LAN d'une telle modélisation.

Adresses :

Fabrice GAMBOA
Laboratoire de Statistique et Probabilités
Université Paul Sabatier
31062 TOULOUSE Cedex 9
France
E-mail : gamboa@math.ups-tlse.fr
<http://www.lsp.ups-tlse.fr/Fp/Gamboa/>

Jean-Michel LOUBES
Université de Montpellier 2
Département de mathématiques
Equipe de Probabilités et Statistique/ UMR 7641
place E. Bataillon, bat 9, bureau 412,
34000 Montpellier
France
E-mail : `loubes@math.univ-montp2.fr`
<`http://www.math.univ-montp2.fr/~loubes`>

Approche Maxiset en estimation bayésienne

par **Florent Autin**, Dominique Picard et Vincent Rivoirard

En utilisant l'approche maxiset comme outil de comparaison de procédures, nous montrons que si les estimateurs bayésiens usuels construits à partir de densités à queues lourdes sont équivalents aux estimateurs de seuillage classiques, il en est de même pour les estimateurs bayésiens construits à partir de densités gaussiennes à grande variance et dont les performances numériques sont très bonnes.

Adresses :
Florent AUTIN
Université d'Aix-Marseille 1
L.A.T.P.
39, rue F. Joliot Curie,
13453 Marseille Cedex 13
France
E-mail : `autin@cmi.univ-mrs.fr`
<`http://www.cmi.univ-mrs.fr/~autin`>

Dominique PICARD
LPMA / UMR 7599
Université Denis Diderot (P7) - Boîte courrier 7012
75251 PARIS Cedex 05
France
E-mail : `picard@math.jussieu.fr`
<`http://www.proba.jussieu.fr/pageperso/picard/picard.html`>

Vincent RIVOIRARD
Laboratoire de Mathématique UMR 8628
Equipe Probabilités, Statistique et Modélisation
Université Paris-Sud 11
Bâtiment 425
91405 Orsay Cedex
E-mail : `Vincent.Rivoirard@math.u-psud.fr`
<`http://www.math.u-psud.fr/~rivoirar`>

Méthode de vraisemblance profilée pénalisée pour l'estimation semi-paramétrique de la période d'une fonction périodique

par **Ismael Castillo**

Dans le modèle d'estimation de période d'une fonction périodique inconnue en bruit blanc gaussien, j'expliquerai comment, par l'intermédiaire d'une méthode de vraisemblance profilée pénalisée, on obtient une famille d'estimateurs de la période paramétrée par une famille de poids. Choisir une suite de poids particulière revient à mettre un a priori sur la fonction inconnue du modèle. De plus, un choix adéquat de poids permet d'obtenir l'efficacité semi-paramétrique au second ordre.

Adresse :

Ismael CASTILLO
Laboratoire de Mathématique UMR 8628
Equipe Probabilités, Statistique et Modélisation
Université Paris-Sud 11
Bâtiment 430
91405 Orsay Cedex
E-mail : `Ismael.Castillo@math.u-psud.fr`
<`http://www.math.u-psud.fr/~castillo`>