

Variable Selection via MKMK

Selezione delle variabili attraverso il metodo MKMK

Roberta Paroli

Università Cattolica del Sacro Cuore, Milano

roberta.paroli@unicatt.it

Luigi Spezia

Università Cà Foscari, Venezia

spezia@dst.unive.it

Riassunto: In questo lavoro viene affrontato, in ambito bayesiano, il problema della scelta delle variabili esogene che più influenzano l'andamento di una serie storica. In letteratura i metodi bayesiani per la selezione delle variabili sono stati studiati soprattutto per il caso dei modelli lineari o lineari generalizzati. Nel presente lavoro si focalizza l'attenzione sul metodo proposto da Kuo e Mallick (1998) e lo si adatta al caso di modelli non lineari per serie storiche dipendenti da una successione di variabili latenti. Poiché, alla luce dei risultati di varie simulazioni, si nota che, all'aumentare sia della complessità del modello che della correlazione tra le variabili esogene, la metodologia adottata non porta a risultati soddisfacenti, se ne propone una sua modifica, denominata Metropolized-Kuo-Mallick (MKMK). La metodologia proposta viene illustrata tramite vari esempi di simulazione.

Keywords: Metropolized-Kuo-Mallick method; Markov Switching Autoregressive models; Non Homogeneous Hidden Markov models.

1. Introduction

In recent years, several methods have been proposed in Bayesian literature for selecting independent variables in regression models: they include “Stochastic Search Variable Selection” (SSVS) by George and McCulloch (1993), unconditional prior Gibbs sampler or “Gibbs sampler Kuo-Mallick” (GKMK) by Kuo and Mallick (1998) and “Gibbs Variable Selection” (GVS) by Dellaportas et al. (2002). The basic idea of these approaches is that the promising predictors are identified by their higher posterior probability, that is the 'best' subset of covariates is that with more frequent appearance in the sequence drawn by the Markov chain Monte Carlo (MCMC) algorithm developed to evaluate the posterior probability. To give a simple idea, we consider the multiple linear regression model with q potential fixed predictors which are candidates for inclusion in the model. Since there are q variables, there are 2^q possible subsets to select. The selection of a subset is also equivalent to set to 0 those regression coefficients corresponding to the covariates that are not included in the final model. Hence, a binary indicator vector $\gamma = (\gamma_1, \dots, \gamma_1, \dots, \gamma_q)'$ is added in the specification of the model and it represents which of the q variables are included. In a Bayesian context the 'best' subset of regressors is identified by analysing the marginal posterior distribution of γ .

The difference among the three above-mentioned methods lies in the way in which the vector γ is involved in the model. In particular, in the SSVS approach, γ is placed in the prior of the regression coefficients, which is a mixture of two normals with different variances, weighted by the indicator variable; in the Kuo-Mallick approach γ is part of the regression equation, while the GVS approach is a "hybrid" of SSVS and GKMK.

In this paper we consider the Kuo and Mallick approach because it is extremely straightforward, since it requires only to specify the prior of the regression coefficients and no pilot runs are required to define the hyperparameters, as in SSVS and GVS. Our aim is twofold: first we adapt the GKMK algorithm to some non linear time series models, such as Markov switching autoregressive models with exogenous variables (MSARX) and non homogeneous hidden Markov models (NHHM); then, we propose the "Metropolization" of the GKMK algorithm that allows us to solve the problems rising when the complexity of the model increases and the exogenous variables have high correlations.

2. GKMK method for non linear time series models

We focus our attention to two special non linear models for time series that depend on unobserved variables by developing GKMK algorithms for MSARX and NHHM.

2.1. MSARX model

MSARX of order $(m; p; q)$ are m different autoregressions of order p , with q exogenous variables, each one depending on one of the m possible regimes, which alternate according to the regime switching, which is driven by an unobserved Markov chain $\{S_t\}$. The equation describing the model is:

$$Y_{t(i)} = \mu_i + \sum_{\tau=1}^p \phi_{\tau(i)} y_{t-\tau} + \sum_{j=1}^q \gamma_{j(i)} \beta_{j(i)} x_{t,j} + E_{t(i)}$$

where $Y_{t(i)}$ denotes the conditional variable observed Y_t when $S_t=i$, $i=1, \dots, m$, for any $1 \leq t \leq T$ and for any i ; the autoregressive coefficients $\phi_{\tau(i)}$ depend on the current state of the Markov chain; $x_{t,j}$ is the observation of the j -th exogenous deterministic variable, weighted by the coefficient $\beta_{j(i)}$, that depends on the current state of the Markov chain, and the dummy indicator $\gamma_{j(i)} \forall t, \forall i$. $E_{t(i)}$ denotes the Gaussian noise ($E_{t(i)} \sim N(0; \sigma_i^2)$). The parameters of the model are: the vector μ of the m mean levels μ_i ; the m vectors of the autoregression coefficients $\phi_{(i)} = (\phi_{1(i)}, \phi_{2(i)}, \dots, \phi_{p(i)})'$; the m vectors of the coefficients $\beta_{(i)} = (\beta_{1(i)}, \beta_{2(i)}, \dots, \beta_{q(i)})'$; the m vectors of indicators $\gamma_{(i)} = (\gamma_{1(i)}, \gamma_{2(i)}, \dots, \gamma_{q(i)})'$ and the vector σ^2 of the m variances σ_i^2 and they will be generated by a Gibbs sampling procedure.

We place conjugate and independent priors on any μ_i , $\phi_{(i)}$, $\beta_{(i)}$ and σ_i^2 , for any $i = 1, \dots, m$, while independent Bernoulli priors, $Be(p_{j(i)})$, are placed on each entry of $\gamma_{(i)}$, i.e. $\gamma_{j(i)}$. At any iteration k , the generation of $\mu^{(k)}$, $\phi^{(k)}$, $\beta^{(k)}$, $\sigma^{2(k)}$ follows straight for the conjugacy of the priors, while each $\gamma_{j(i)}^{(k)}$ is generated independently from a Bernoulli distribution with parameter $\pi_{j(i)} = a_{j(i)} / (a_{j(i)} + b_{j(i)})$, where

$$a_{j(i)} = \exp \left\{ - (1/2) \sigma_i^{-2(k-1)} \sum_{t=1}^n (y_t - \mu_i^{(k)} - \sum_{\tau=1}^p \phi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{h=1}^q \theta_{h(i)}^{*(k)} x_{t,h})^2 \right\} p_{j(i)}$$

$$b_{j(i)} = \exp\left\{-(1/2) \sigma_i^{-2(k-1)} \sum_{t=I}^n (y_t - \mu_i^{(k)} - \sum_{\tau=I}^p \phi_{\tau(i)}^{(k)} y_{t-\tau} - \sum_{h=1}^q \theta_{h(i)}^{*(k)} x_{t,h})^2\right\} (1-p_{j(i)})$$

and $\theta_{h(i)}^{*(k)} = \gamma_{h(i)}^{(k)} \beta_{h(i)}^{(k)}$ if $h \neq j$ or $\theta_{h(i)}^{*(k)} = 1$ if $h=j$ and $\theta_{h(i)}^{** (k)} = \gamma_{h(i)}^{(k)} \beta_{h(i)}^{(k)}$ if $h \neq j$, or $\theta_{h(i)}^{** (k)} = 0$ if $h=j$.

2.2. NHHM model

The Gaussian non homogeneous hidden Markov model is a "signal plus noise" model, where it is assumed that the probability density function of every observation, given the contemporary state of the Markov chain, is Gaussian and the transition probabilities vary in time and depend on some exogenous variables. Let $\xi_{ij}^t = P(S_t=j|S_{t-1}=i)$ be the time-varying transition probabilities, for any $i,j = 1, \dots, m$, and at any time $t = 2, \dots, T$, they can be obtained as logistic functions of exogenous deterministic variables. Hence, given the conditional independence and the contemporary dependence conditions, the equations describing the NHHM model are

$$Y_{t(i)} = \mu_i + E_{t(i)}, \quad \text{and} \quad \text{logit}(\xi_{ij}^t) = \ln(\xi_{ij}^t / \xi_{ii}^t) = \beta_{0(i,j)} + \sum_{h=1}^q \gamma_{h(i,j)} \beta_{h(i,j)} x_{t,h}$$

where any vector of parameters $\beta_{i,j} = (\beta_{0(i,j)}, \beta_{1(i,j)}, \dots, \beta_{q(i,j)})'$, if $i \neq j$, is associated to a vector of binary indicators $\gamma_{i,j} = (\gamma_{1(i,j)}, \dots, \gamma_{q(i,j)})'$, of dimension q , of the exogenous variables to be included. The step of the MCMC algorithm for the generation of the indicators must be modified as follows: for any i,j every coefficient $\gamma_{h(i,j)}^{(k)}$ is independently generated for $h=1, \dots, q$, from a Bernoulli distribution with parameter $\pi_{h(i,j)} = a_{h(i,j)} / (a_{h(i,j)} + b_{h(i,j)})$, where

$$a_{h(i,j)} = p_{h(i,j)} \prod_{\{t \geq 2: S_t = i\}} (\exp\{\mathbf{X}_t' \boldsymbol{\theta}_{iS_t}^{*(k)}\}) (1 + \sum_{b \neq i} \exp\{\mathbf{X}_t' \boldsymbol{\theta}_{bS_t}^{*(k)}\})^{-1}$$

$$b_{h(i,j)} = (1-p_{h(i,j)}) \prod_{\{t \geq 2: S_t = i\}} (\exp\{\mathbf{X}_t' \boldsymbol{\theta}_{iS_t}^{** (k)}\}) (1 + \sum_{b \neq i} \exp\{\mathbf{X}_t' \boldsymbol{\theta}_{bS_t}^{** (k)}\})^{-1}.$$

with $\mathbf{X}_t$ is the matrix of the observations of the exogenous variables, while $\boldsymbol{\theta}_{iS_t}^{*(k)}$ and $\boldsymbol{\theta}_{iS_t}^{** (k)}$ are defined as in section 2.1.

3. MKMK method for non linear time series models

To perform variable selection with Metropolized-Kuo-Mallick (MKMK) only one step in the basic MCMC algorithm must be modified: in this version of the algorithm, the coefficients of the exogenous variables and the associated dummy variable are generated and accepted in block by a unique Metropolis step.

For any type of model, if $(\beta; \gamma)$ is the pair of the coefficient and the dummy indicator of the exogenous variable, let $(\beta^{(k-1)}; \gamma^{(k-1)})$ be the current state of the MCMC chain and $(\beta^{(k)}; \gamma^{(k)})$ be the candidate drawn from the proposal density $q(\beta^{(k)}; \gamma^{(k)} | \beta^{(k-1)}; \gamma^{(k-1)})$. The new values are accepted with probability

$$\alpha[(\boldsymbol{\beta}^{(k-1)}; \boldsymbol{\gamma}^{(k-1)}); (\boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k)})] = \min \left\{ 1; \frac{\pi(\boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k)}) q(\boldsymbol{\beta}^{(k-1)}; \boldsymbol{\gamma}^{(k-1)} | \boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k)})}{\pi(\boldsymbol{\beta}^{(k-1)}; \boldsymbol{\gamma}^{(k-1)}) q(\boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k)} | \boldsymbol{\beta}^{(k-1)}; \boldsymbol{\gamma}^{(k-1)})} \right\},$$

where $q(\boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k)} | \boldsymbol{\beta}^{(k-1)}; \boldsymbol{\gamma}^{(k-1)})$ can be factorised as $q(\boldsymbol{\gamma}^{(k)} | \boldsymbol{\beta}^{(k-1)}; \boldsymbol{\beta}^{(k)}; \boldsymbol{\gamma}^{(k-1)}) q(\boldsymbol{\beta}^{(k)} | \boldsymbol{\beta}^{(k-1)})$.

4. Simulation examples

Our simulated data are based on the examples discussed by George and McCulloch (1993) and Kuo and Mallick (1998). By these examples, we illustrate the performances of the GKMK and MKMK methods for the time series models presented in the previous section, for problems with small or large number of potential predictors, in the case of independence or correlation. In the original examples for the linear regression model, a dataset of length $n=60$ with $q=5$ predictors is simulated. The exogenous variables $X_1, \dots, X_5$ are obtained as standard normal vectors (independent or correlated), while the regression coefficients are $\boldsymbol{\beta}=(0,0,0,1,1.5)'$, so the indicator vector is $\boldsymbol{\gamma}=(0,0,0,1,1)'$; the variance of the error is $\sigma^2=2.5^2$. These parameters are also used to create the time series generated by our models. For the MSARX model we generate time series of length $n=1000$, for each pairs of $p=1,2,3$ and $m=2,3$ and different coefficient vectors $\boldsymbol{\beta}_{(i)}$ for any states. For the NHHM model the time series generated are of length $n=1000$, for each values of $m=2,3$ and different coefficient vectors $\boldsymbol{\beta}_{(ij)}$ in the logistic function for any states. For each dataset we ran 50000 iterations of the two sampling schemes, after a burn-in period of 10000 iterations, and we tabulated the frequencies of the highest frequency model.

The results show that both the GKMK and the MKMK methods predict the right model with excellent relative frequency, especially in the case of independent exogenous variables. We can also notice that the sampling schemes explore local modes of the posterior distribution because different results are obtained in different runs, although all runs are able to identify the true model.

Moreover, when the correlations among predictors are very high, MKMK is more efficient than traditional methods, especially for the NHHM model.

References

- Dellaportas, P., Forster, J.J., Ntzoufras, I. (2002) On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12, 27-36.
- George, E.I. and McCulloch, R.E. (1993) Variables selection via Gibbs-Sampling. *Journal of the American Statistical Association*, 88, 881-889
- Kuo, L. and Mallick, B. (1998) Variable Selection for Regression Models. *Sankhya*, B, 60, 65-81.