Variable Length Markov Chain with Exogenous Covariates

Adriano Zanin Zambom
Cal. State University Northridge

joint work with
Seonjin Kim (Miami University Ohio)
Nancy Lopes Garcia (UNICAMP)

VLMC

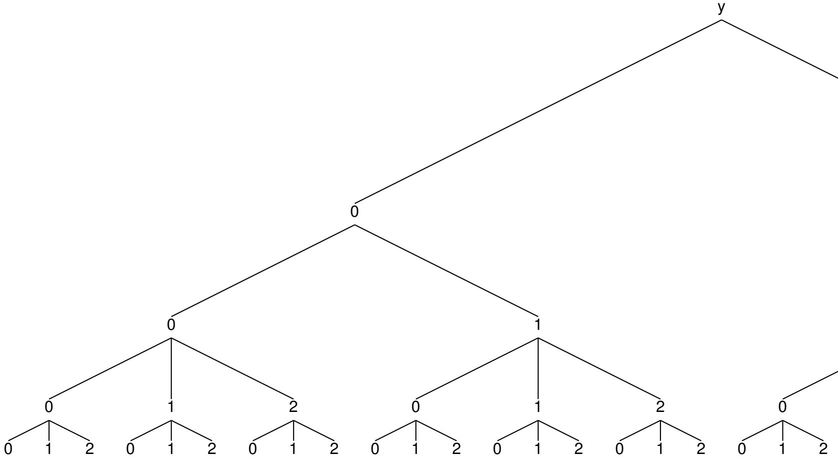
- VLMC - Variable Length Markov Chain
- Buhlmann and Wyner (1999)

VLMC

- VLMC - Variable Length Markov Chain
- Buhlmann and Wyner (1999)
- Full Markov Chain with finite order:
  $P(X_n = x_n) = P(X_{n-1} = x_{n-1}, \ldots, X_{n-k} = x_{n-k})$

VLMC

- VLMC - Variable Length Markov Chain
- Buhlmann and Wyner (1999)
- Full Markov Chain with finite order:
  $P(X_n = x_n) = P(X_{n-1} = x_{n-1}, \ldots, X_{n-k} = x_{n-k})$

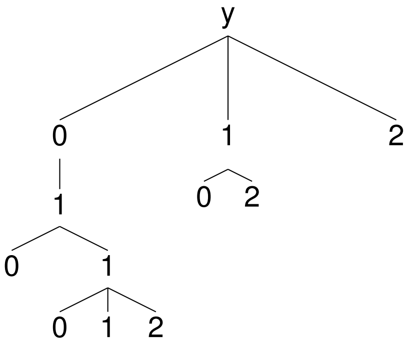


- Curse of Dimensionality!

- Curse of Dimensionality!
- Rissanen (1983) - context algorithm - Computer Science

VLMC

- Curse of Dimensionality!
- Rissanen (1983) - context algorithm - Computer Science
- Lump irrelevant states together



- Buhlmann and Wyner (1999): Probability and Statistics point of view

Definition

(Buhlmann & Wyner, 1999) Let $(X_t)_{t\in\mathbb{Z}}$ *be a stationary process with values* $X_t \in \mathcal{X}$, $|\mathcal{X}| < \infty$. *Denote by* $c : \mathcal{X}^\infty \to \mathcal{X}^\infty$ *a (projection) function which maps*

$$c : x^0_{-\infty} \to x^0_{-\ell+1}, \text{ where } \ell \text{ is defined by}$$
$$\ell = \min\{k : P(X_1 = x_1 | X^0_{-\infty} = x^0_{-\infty}) = P(X_1 = x_1 | X^0_{-k+1} = X^0_{-k+1}) \text{ for all } x_1 \in \mathcal{X}\}$$

Then, $c(\cdot)$ *is called the context function for any* $t \in \mathbb{Z}$, *and* $c(x^{t-1}_{-\infty})$ *is called the context for* x_t.

beta-VLMC

- Consider the following situation

beta-VLMC

- Consider the following situation
- States: $\mathcal{Y} = \{0, 1\}$

- Consider the following situation
- States: $\mathcal{Y} = \{0, 1\}$
- At each step/transition t we observe a set of covariates $\mathbf{X}_t = (X_{t1}, \ldots, X_{td})$ independently

beta-VLMC

- Consider the following situation
- States: $\mathcal{Y} = \{0, 1\}$
- At each step/transition t we observe a set of covariates $\mathbf{X}_t = (X_{t1}, \dots, X_{td})$ independently
- Assume that for a given context, the observed covariate values are relevant to the transition probability

beta-VLMC

- Consider the following situation
- States: $\mathcal{Y} = \{0, 1\}$
- At each step/transition t we observe a set of covariates $\mathbf{X}_t = (X_{t1}, \ldots, X_{td})$ independently
- Assume that for a given context, the observed covariate values are relevant to the transition probability
- This can be modeled in the following way

beta-VLMC

Definition

Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary process with values $Y_t \in \mathcal{Y}$ and $\mathbf{X}_t \in \mathcal{X} \subseteq \mathbb{R}^d$ a d-dimensional vector of exogenous covariates. Denote by $c : \mathcal{Y}^\infty \to \mathcal{Y}^\infty$ a (projection) function which maps

$$c : y^0_{-\infty} \to y^0_{-\ell+1}, \text{ where } \ell \text{ is defined by}$$
$$\ell = \min\{k : P(Y_1 = 1 | Y^0_{-\infty} = y^0_{-\infty}, \mathbf{X}^0_{-\infty} = \mathbf{x}^0_{-\infty})$$
$$= P_{\boldsymbol\theta}(Y_1 = 1 | Y^0_{-k+1} = y^0_{-k+1}, \mathbf{X}^0_{-h} = \mathbf{x}^0_{-h+1}) \text{ for all } \mathbf{x}^0_{-\infty}, \text{ and } k \geq h\}$$

where, letting $u := y^0_{-\ell+1}$

$$P_{\boldsymbol\theta}(Y_1 = 1 | Y^0_{-\ell+1} = y^0_{-\ell+1}, \mathbf{X}^0_{-h+1} = \mathbf{x}^0_{-h+1}) = \frac{\exp(\alpha^u + \mathbf{x}^0_{-h+1}\boldsymbol\beta^{u\,T})}{1 + \exp(\alpha^u + \mathbf{x}^0_{-h+1}\boldsymbol\beta^{u\,T})}, \tag{1}$$

for $h \leq \ell$ and $\boldsymbol\theta := \boldsymbol\theta^u = (\alpha^u, \boldsymbol\beta^u) = \left(\alpha^u, \boldsymbol\beta^u_0, \ldots, \boldsymbol\beta^u_{(-h+1)}\right)$, defines the vector of coefficients associated with the context (past states) $u = y^0_{-\ell+1}$ for transitioning into state 1, and $\boldsymbol\beta^u_t = (\beta^u_{t1}, \ldots, \beta^u_{td})$ is the vector of coefficients corresponding to the d exogenous covariates at time $t = 0, \ldots, -h+1$. Then, $c(\cdot)$ is called the beta-context function for any $t \in \mathbb{Z}$, and $c(y^{t-1}_{-\infty})$ is called the beta-context for y_t with associated parameter vector $\boldsymbol\theta^u$.

Definition

Let $c(\cdot)$ be a beta-context function of a stationary beta-context model of order k. The beta-context tree τ is defined as

$$\tau \;=\; \tau_c = \{u : u = c(y_{-k+1}^0), y_{-k+1}^0 \in \mathcal{Y}^k\}$$

with an associated parameter tree

$$\tau_\theta \;=\; \{(u, \boldsymbol{\theta}^u) : u \in \tau\}$$

- The Beta-Context Algorithm

1. Start with a maximal beta-context tree

$$\tau^{(0)} = \tau_{\max} = \{u = y^0_{-k+1} : N(y^0_{-k+1}) \geq s(1 + dk)\}$$

   - $1 + dk$ = number param. to be estimated in context y^0_{-k+1}
   - s = tuning parameter

beta-VLMC

- The Beta-Context Algorithm

1. Start with a maximal beta-context tree

$$\tau^{(0)} = \tau_{\max} = \{u = y^0_{-k+1} : N(y^0_{-k+1}) \geq s(1 + dk)\}$$

   - $1 + dk$ = number param. to be estimated in context y^0_{-k+1}
   - s = tuning parameter

2. For each context $u \in \tau^{(0)}$ of length r, test

$$H_0^u : \beta^u_{-r+1} = 0$$

with the LRT

$$-2\left[\log L\left(\tilde{\tau}^u_\theta \middle| y_1^n, \mathbf{x}_1^n\right) - \log L\left(\hat{\tau}^{(0)}_\theta \middle| y_1^n, \mathbf{x}_1^n\right)\right]$$

- The Beta-Context Algorithm

1. Start with a maximal beta-context tree

$$\tau^{(0)} = \tau_{\max} = \{u = y^0_{-k+1} : N(y^0_{-k+1}) \geq s(1 + dk)\}$$

   - $1 + dk$ = number param. to be estimated in context y^0_{-k+1}
   - s = tuning parameter

2. For each context $u \in \tau^{(0)}$ of length r, test

$$H_0^u : \beta^u_{-r+1} = 0$$

with the LRT

$$-2\left[\log L\left(\tilde{\tau}^u_\theta \middle| y_1^n, \mathbf{x}_1^n\right) - \log L\left(\hat{\tau}^{(0)}_\theta \middle| y_1^n, \mathbf{x}_1^n\right)\right]$$

   - If not rejected at level γ_n: prune beta coef. only

3. With respect to the tests performed in **Step 2**:

3. With respect to the tests performed in **Step 2**:
    1. If neither $H_0^{u_1}$ nor $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau_{(0)}$, was rejected then test whether these siblings' final nodes can be dropped

3. With respect to the tests performed in **Step 2**:
   1. If neither $H_0^{u_1}$ nor $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau_{(0)}$, was rejected then test whether these siblings' final nodes can be dropped
   2. If at least one of $H_0^{u_1}$ and $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau^{(0)}$, was rejected, both u_1 and u_2 remain in the tree

3. With respect to the tests performed in **Step 2**:
   1. If neither $H_0^{u_1}$ nor $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau_{(0)}$, was rejected then test whether these siblings' final nodes can be dropped
   2. If at least one of $H_0^{u_1}$ and $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau^{(0)}$, was rejected, both u_1 and u_2 remain in the tree
4. Repeat Steps 2 and 3 with the updated trees $\tau^{(1)}$ and $\tau_{\hat{\theta}}^{(1)}$ for contexts of length $r-1, r-2, \ldots, 1$.

3. With respect to the tests performed in **Step 2**:
   1. If neither $H_0^{u_1}$ nor $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau_{(0)}$, was rejected then test whether these siblings' final nodes can be dropped
   2. If at least one of $H_0^{u_1}$ and $H_0^{u_2}$, for u_1 and u_2 siblings in $\tau^{(0)}$, was rejected, both u_1 and u_2 remain in the tree
4. Repeat Steps 2 and 3 with the updated trees $\tau^{(1)}$ and $\tau_{\hat{\theta}}^{(1)}$ for contexts of length $r-1, r-2, \ldots, 1$.

- Result: $\hat{\tau}$ with associated parameter tree $\hat{\tau}_\theta$ and corresponding beta-context function $\hat{c}(\cdot)$.

Consistency of $\hat{\tau}$ and $\hat{\tau}_\theta$

Consistency of $\hat{\tau}$ and $\hat{\tau}_\theta$

Assumptions:
C1: $\gamma_n \to 0$ such that $n\gamma_n = o(1)$.

beta-VLMC

Consistency of $\hat{\tau}$ and $\hat{\tau}_\theta$

Assumptions:
C1: $\gamma_n \to 0$ such that $n\gamma_n = o(1)$.
C2: $\gamma_n \to 0$ such that $(1/n)\log(1/\gamma_n) = o(1)$.

Consistency of $\hat{\tau}$ and $\hat{\tau}_\theta$

Assumptions:
C1: $\gamma_n \to 0$ such that $n\gamma_n = o(1)$.
C2: $\gamma_n \to 0$ such that $(1/n)\log(1/\gamma_n) = o(1)$.
C3: The order of the initial maximal tree τ_{max} is $r = O(log(n))$.

Consistency of $\hat{\tau}$ and $\hat{\tau}_\theta$

Assumptions:
C1: $\gamma_n \to 0$ such that $n\gamma_n = o(1)$.
C2: $\gamma_n \to 0$ such that $(1/n)\log(1/\gamma_n) = o(1)$.
C3: The order of the initial maximal tree τ_{max} is $r = O(log(n))$.

Theorem

Assume that the beta-context tree τ has finite order k and the contexts $u \in \tau$ are correctly identified, that is, $\hat{\tau} = \tau$, and that its associated parameter tree $\hat{\tau}_\theta$ is pruned sequentially backwards. Then, under conditions C1 and C2, the estimated associated parameter tree is consistent in the sense that

$$\lim_{n \to \infty} P[\hat{\boldsymbol{\theta}}^u = \boldsymbol{\theta}^u, \forall u \in \tau] = 1.$$

Theorem

Assume that the beta-context tree τ has finite order k and the contexts $u \in \tau$ are correctly identified, that is, $\hat{\tau} = \tau$, and that its associated parameter tree $\hat{\tau}_\theta$ is pruned sequentially backwards. Then, under conditions C1 and C2, the estimated associated parameter tree is consistent in the sense that

$$\lim_{n \to \infty} P[\hat{\boldsymbol{\theta}}^u = \boldsymbol{\theta}^u, \forall u \in \tau] = 1.$$

The proof of this theorem is based on showing that the probability of over-fitting and under-fitting tend to 0.

Theorem

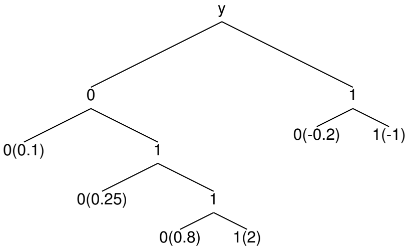
Assume that the beta-context tree τ has finite order k and the contexts $u \in \tau$ are estimated using the beta-context algorithm. Then, under conditions C1 and C2, the estimated context tree is consistent in the sense that

$$\lim_{n \to \infty} P[\hat{\tau} = \tau] = 1.$$

Simulations

Simulations

Model 1



$$
\begin{aligned}
\beta^{00} &= (2,0)' \\
\beta^{010} &= (-1,1,0)' \\
\beta^{0111} &= (1.5,2,0,0)' \\
\beta^{0110} &= (4,3,2,1)' \\
\beta^{10} &= (0,0)' \\
\beta^{11} &= (0,0)'
\end{aligned}
$$

beta-VLMC

Method	BIC	AIC	logLik	No. Params $\hat{\alpha}^u$	No. Params $\hat{\beta}^u$	
beta-VLMC	1128.8	1093.1	-539.3	5.39	7.28	
VLMC	1345.8	1327.5	-660.0	3.72	-	
	order $\hat{\tau}$	order-Covar.	No. Missing $\hat{\tau}$	No. Extra $\hat{\tau}$	Identical τ	Identical τ_θ
beta-VLMC	3.90	3.87	1.28	0.07	0.37	0.04
VLMC	2.30	-	4.63	0.08	0.01	-

Table: Simulation results for Model 1 with $n = 1000$ transitions.

Thank You