1. Hidden Markov Model
00000
2. Modelling Periodic Variations
0000000
3. Model Estimation
000
4. Conclusions
0000

Flexible Modelling of Diel and Other Periodic Variation in Hidden Markov Models

Marco Zanotti

University Milano-Bicocca

Contents

1. Hidden Markov Model

Motivation

▶ Understanding key patterns of animal movement

▶ Investigating drivers of animal movement and behavior

HMM

Hidden Markov Models (HMMs) are statistical models for time series data involving two stochastic processes:

1. the observation process $x_1, ..., x_T$, the time series
2. the latent state process $S_1, ..., S_T$, not observed

HMM

In the basic framework of HMM

▶ X_t is the state-dependent process

▶ each observation x_t is assumed to be generated by one of N possible distributions $f_j(x)$

▶ the latent state process S_t selects which distributions is active at any time

▶ S_t is assumed to be a Markov process with N states

Markov Property

Being S_t a Markov process, it satisfies the **Markov property**

$$f(S_t|S_1, ..., S_{t-1}) = f(S_t|S_{t-1})$$

that is, the state at time t depends only on the state at time $t-1$. Moreover, the Markov property implies that S_t is fully characterized by:

▶ $\delta_j^{(1)} = P(S_1 = j)$, the initial state probabilities

▶ $\Gamma^{(t)} = [\gamma_{ij}^{(t)}]$, the transition probability matrix, where $\gamma_{ij}^{(t)} = P(S_t = j|S_{t-1} = i)$

2. Modelling Periodic Variations

Problem: Periodicity

In many real applications, time series data are often characterized by **periodicities**, such as diel variations (recurrent patterns over a 24 hours period).

Ignoring the periodicity of the data can **invalidate inference**, for instance the standard errors might be underestimated due to autocorrelation.

Moreover, it may be of interest, instead, to model such periodic variations to comprehensively **understand the process dynamics**, for instance some intra-day behavioral patterns.

Periodic variations can be effectively modeled within the framework of HMM by including **temporal covariates** in the model.

HMM with Covariates

Incorporating covariates in HMM can be done in the
state-transition probabilities.

Expressing state transition probabilities as function of covariates to
infer how state switching depends on external factors.

HMM with Covariates

The covariance-dependence on the categorical distribution of states at time t is typically modeled using a **multinomial logistic regression**, which is achieved by applying the inverse multinomial logit link to each row of the transition probability matrix

$$\gamma_{i1}^{(t)} = logit^{-1}(\tau_{ij}^{(t)}) = \frac{e^{\tau_{ij}^{(t)}}}{\sum_{k=1}^{N} e^{\tau_{ij}^{(t)}}}$$

and the general form of the linear predictor for $\tau_{ij}^{(t)}$ is given by

$$\tau_{ij}^{(t)} = z_t' \beta^{(ij)} = \beta_0^{(ij)} + \beta_1^{(ij)} z_{t1} + ... + \beta_p^{(ij)} z_{tp} \quad (1)$$

HMM with Seasonality

A special type of covariate is **seasonality**, that is a variation that repeats over a specif time period, e.g. within-day or within-year.

Modelling this type of periodicity can be achieved by using:

▶ calendar features (e.g. time of the day)

▶ **trigonometric functions**, with period equal to the cycle length

▶ **cyclic splines**

Trigonometric Modelling

When the aim is to model periodic patterns in the state-transition dynamics, the linear predictor (1) can be extended by including **trigonometric functions** with the desired periodicity.

$$\tau_{ij}^{(t)} = z_t' \beta^{(ij)} + \sum_{k=1}^{K} \omega_k^{(ij)} \, sin\left(\frac{2\pi kt}{l}\right) + \psi_k^{(ij)} \, cos\left(\frac{2\pi kt}{l}\right)$$

where l is the period length (i.e. number of sequential observations to complete a cycle).

The flexibility is of the model increases with K, the number of periodic components.

Cyclic Splines Modelling

To avoid making any assumption on the functional form of the periodic effect, a **nonparametric** modelling of the periodic effect can be obtained replacing trigonometric functions with **splines**.

$$\tau_{ij}^{(t)} = z_t' \beta^{(ij)} + \sum_{q=1}^{Q} a_q^{(ij)} B_q$$

where $a_q^{(ij)}$ are the scaling coefficients and B_q are basis splines wrapped at the desired periodicity (l).

A large value of Q is used to guarantee sufficient flexibility and overfitting is avoided by including a penalty term (P-spline approach).

3. Model Estimation

EM Algorithm

The model is estimated using the **Expectation-Maximization (EM)** algorithm.

The missing data are taken to be the hidden states and the complete-data log-likelihood for the HMM is given by

$$\mathcal{L}_c(\theta) = log \ \delta_{S1}^{(1)} + \sum_{t=2}^{T} log \ \gamma_{S_{t-1},S_t}^{(1)} + \sum_{t=1}^{T} log \ f_{S_t}(x_t)$$

where θ is the set of parameters necessary to define $\delta_j^{(1)}$, $\Gamma^{(t)}$ and the state-dependent distributions $f_j(x)$.

EM Algorithm

The parameter estimates are derived:

▶ Choosing **starting values** for the parameter set θ necessary to define state-dependent distributions and transition probabilities

▶ **E–step** computes conditional expectations of missing data (i.e. the latent states), given the data and current parameter values

▶ **M–step** updates parameters based on the CDLL, replacing the functions of the states by their conditional expectations calculated in the E–step

The E-M steps are repeated until convergence.

1. Hidden Markov Model
00000

2. Modelling Periodic Variations
0000000

3. Model Estimation
000

4. Conclusions
●000

4. Conclusions

Conclusions

A particular class of HMM to model periodic variations is introduced that:

▶ makes use of splines instead of trigonometric functions

▶ increases flexibility, which might be helpful to reveal interesting patterns

▶ applications show that using sine and cosine waves is often enough

Bibliografy

Carlina C. Feldmann, Sina Mews, Angelica Cocculla, Ralf Stanewsky & Roland Langrock (2023), 'Flexible Modelling of Diel and Other Periodic Variation in Hidden Markov Models', Journal of Statistical Theory and Practice, 17:45.

1. Hidden Markov Model
○○○○○

2. Modelling Periodic Variations
○○○○○○○

3. Model Estimation
○○○

4. Conclusions
○○○●

Thank you!