Mean Pattern Detecting in Hypnogram Data Using Discrete Semi-Markov Process

Haoqun Cao, Xieqing Yu

Abstract

Hypnogram is a plot describing the sleep stages during the sleep process. Hypnogram plays an important role in sleep study. In this work, we propose to model hypnogram data with **nonhomogeneous discrete semi-markov process**. We give a formal definition of such model and prove its likelihood can be decomposed into each transition type through proper approximation. Then we use a novel parametrization for nonhomogeneity which can separate the baseline transition intensity and the effect of nonhomogeneity meanwhile can be efficiently estimated. We propose a method of deriving a mean pattern which is representative and simple compared to the rather noisy data. Our empirical results are consistent with existing medical sleep studies.
Keywords. semi-markov process, competing risk model

1 Introduction

Hypnogram is a plot describing the sleep stages during the sleep process. It can be perceived as a categorical time series. Hypnogram plays an important role in sleep study (Swihart et al., 2008). When given large samples of hypnograms, an efficient way to explore its routine in a statistical way is urgently needed. We also hope to summarise the heterogeneous hypnogram observations into a simple but representative mean pattern.

From the perspective of categorical time series, statistical modeling methods can be summarised into 4 methods. The Multi-State model, the link function approach(Fahrmeir and Kaufmann, 1987), the likelihood-based approach(Fokianos and Kedem, 1998) and spectral envelope approach(Stoffer et al., 1993). In this report, we majorly follow the research line of the Multi-State model. We extend the existing method by proposing weaker assumptions.

The Multi-State Model is defined as a model for a continuous or discrete time stochastic process allowing individuals to move between a finite number of states. It can be generally specified by the transition kernel. However, parametrization and estimation of transition kernel is inefficient. The major way to reduce computational cost is by introducing assumptions. The most widely used assumption is the homogeneous Markov assumption where sojourn times follow a exponential distribution(Andersen and Keiding, 2002). However, the homogeneous Markov assumption can be too restrictive. Weaker assumptions are proposed including the semi-markov assumption where sojourn time can take any distribution.

Existing literature on semi-markov process(SMP) modeling majorly focuses on the homogeneous and continuous case. Kang and Lagakos (2007) proposes to model homogeneous SMP using transition intensity(equivalently cause-specific hazard). The estimation of the transition function is done by jointly optimizing the likelihood function, which is costly. Asanjarani et al. (2021) proposes to decompose the likelihood of continuous homogeneous SMP into each transition type and use the standard survival analysis method to separately estimate transition intensity. Their method is easy to implement, however, homogeneous assumption is still too strong for real-world applications.

Therefore, we introduce the nonhomogeneous discrete semi-markov process to model our hypnogram data. We manage to extend the method of likelihood decomposition to the discrete nonhomogeneous situation through an approximation. Then we explicitly model nonhomogeneity with a time-varying covariate in the proportional hazard model. Through that, we manage to separately specify the baseline transition intensity and the change of intensity caused by nonhomogeneity. Compared to existing method on modeling nonhomogeneity as shown in Fahrmeir and Klinger (1998), our method is easier to implement meanwhile can separate the effect of baseline intensity of effct of nonhomogeneity, therefore more interpretable.

After property modeling our data, we summarise the estimation result by predicting a **mean pattern**. The mean pattern can be perceived as a *mean* of an infinite-dimensional mixed random vector. We propose a way to derive the mean pattern that can capture the major mode of sleep process as shown in the experiments.

Our report covers the following parts in the following order. In section 2.1, we introduce the notations and basic properties of the probabilistic model we use. In section 2.2, we introduce our methods of parametrization for the model and estimation procedure. In section 2.3, we introduce how we derive the mean pattern after requiring the estimation. In section 3, we show our result on the oura ring pregnant women's data.

2 Proposed Modeling Method

2.1 Nonhomogeneous Discrete Semi-Markov Process

Let $\mathcal{S}$ denotes a discrete state set. Denote a sequence of r.v. take value in $\mathcal{S}$ as $\{X_t, t \in \mathbb{N}\}$ representing the state of system at time point t. Denote $T_0 = 0$, $T_h = min\{t, X_t \neq X_{T_{h-1}}\}, h \geq 1$. and $J_h = X_{T_h}, \tau_h = T_h - T_{h-1}$. We call $\{X_t, t \in \mathbb{N}\}$ a discrete-time multi-state process. And J_n and T_h are nth state and corresponding transition time respectively. In the context of hypnogram modeling, we always assume there is a deterministic initial state $J_0 = s_0$. We also assume there is an absorbing state denoted as s_a.

The semi-markov property holds if and only if,

$$Pr\{J_h = j, \tau_h = t \leq t | J_{h-1} = i, T_{h-1}, ..., T_1, J_0\} = Pr\{J_h = j, \tau_h = t | J_{h-1} = i, T_{h-1}\}, \forall h. \quad (2.1)$$

Further, if the joint conditional probability of J_h, τ_h given J_{h-1}, T_{h-1} is the same for any n, we call $\{X_t, t \in \mathbb{N}\}$ a **nonhomogeneous discrete semi-markov process**.

Due to the semi-markov property, it's straightforward to verify that $\{X_t, t \in \mathbb{N}\}$ can to specified by the conditional joint distribution of J_n, τ_n given J_{h-1}, T_{h-1} which is denoted as,

$$f(j, t|i, T_{h-1}) := Pr(J_h = j, \tau_h = t | J_{h-1} = i, T_{h-1}). \quad (2.2)$$

An alternative to specify the process is by transition intensity function defined as,

$$\alpha_{ij}(t|T_{h-1}) := Pr(\tau_h = t, J_h = j | J_{h-1} = i, \tau_h \geq t, T_{h-1}). \quad (2.3)$$

It's straightforward to verify that the transition intensity function can derive $f(j, t|i, T_{h-1})$, therefore it's equivalent to specifying the process with such functions. It's also notable that the transition intensity function takes exactly the same form as cause-specific function in discrete competing risk models (Tutz et al., 2016).

At the end of this section, we propose a theorem essential to future parametrization and estimation. Denote the conditional survival function of τ_n as $S_i(t|T_{h-1}) := Pr(\tau_n > t|J_{h-1} = i, T_{h-1})$ and denote $S_{ij}(t) := \prod_{t_m \le t} \alpha_{ij}(t_m|T_{h-1})$, then we get the following theorem on likelihood decomposition.

Theorem 2.1. If the following approximation holds,

$$S_i(t|T_{h-1}) \approx \prod_{j \neq i} S_{ij}(t|T_{h-1}) \tag{2.4}$$

, the likelihood of trajectory $\{J_n, \tau_n\}_{h=1}^{H}$ can be expressed as,

$$\begin{aligned}
\mathcal{L} &\approx \prod_{i \in \mathcal{S}} \prod_{j \neq i} \mathcal{L}_{ij} \\
\mathcal{L}_{ij} &= \prod_{h=1}^{H} \mathcal{L}_{ij}^{h} \\
\mathcal{L}_{ij}^{h} &= \Big\{ \prod_{t=1}^{\tau_h - 1} \big[1 - \alpha_{ij}(t|T_{h-1})\big]^{\mathbf{1}_{J_{h-1}=i}} \Big\} \big[\alpha_{ij}(\tau_n|T_{h-1})\big]^{\mathbf{1}_{J_{h-1}=i, J_n=j}},
\end{aligned} \tag{2.5}$$

Remark 2.1. *First, we can observe that the likelihood is decomposed into each transition type, which makes estimation for each transition intensity function separatable. Second, it's noted that for the each decomposed likelihood, it takes the form of binary regression. This form allows for efficient estimation after proper parametrization we will introduce later.*

The idea of decomposition originates in this way. First, we will show the equivalency between discrete semi-markov process and discrete competing risk model. Then, we will show that in discrete competing risk model, a "collapsed likelihood" approach(Lee et al., 2018) is proposed to simply treat observation from other causes as censor when focusing on one cause. We will show that our approximation is the same as such method however originates from different perspectives. The proof of 2.1 is provided in appendix also with the discussions.

2.2 Model Parametrization and Estimation

Suppose we have a total of N individuals with individual covariate(e.g. demographic feature) $\mathbf{z_n}$. Each individual has observations of D_n days. The dth day's observation for individual n is a hypnogram which can be expressed with trajectory $\{J_h^{n,d}, T_h^{n,d}\}_{h=0}^{H^{n,d}}$ where $J_{H^{n,d}}^{n,d} = s_a$. Suppose the observation of day d individual n follows a nonhomogeneous discrete semi-markov process, we model its transition intensity function as,

$$\frac{\alpha_{ij}^{n,d}(t|T_{n-1})}{1 - \alpha_{ij}^{n,d}(t|T_{n-1})} = \frac{\alpha_{ij0}(t)}{1 - \alpha_{ij0}(t)} exp\big[\beta_{ij}(T_{n-1}) + \beta^T \mathbf{Z}_n\big] \tag{2.6}$$

where $\alpha_{ij0}(t)$ is the baseline hazard and $\beta_{ij}(\cdot)$ is a unknown smooth functions.

We can observe that the nonhomogeneity is modeled through the unknown function $\beta_{ij}(\cdot)$ in a proportional hazard way. In this way, the hazard can be devided into two parts. The first part is the baseline hazard determines the fundamental distribution of time interval. Then the non-homogeneous part expands or squeezes this hazard in a proportional way. This is quiet unlike the literature that adopt counting process theory to model nonhomogeneity (Fahrmeir and Klinger, 1998). Under their method, the baseline and variations caused by nonhomogeneity is not seperable.

From such parametrization, we are making two underlying assumptions here. First, we assume different days' sleep of the same individual are identically distributed. Second, we assume the covariate only affect the proportional part and not affect baseline hazard or $\beta_{ij}(\cdot)$.

The estimation can be done by rewriting the likelihood into form of equation 2.5 and performing standard generalized linear regression with a logit link function.

2.3 Mean Pattern Derivation

After we estimate all the parameters and have the whole probability model in hand, our next job is to derive a mean pattern to summarise the noisy data.

We define the mean pattern to be a predicted hypnogram denoted with $\{\widehat{J}_h, \widehat{T}_h\}_{h=0}^{\widehat{H}}$. However, since J_h is a discrete random variable that does not take value in $\mathbb{R}$, it's not trivial to define its mean value. Therefore, we combine the idea of maximum probability and conventional mean to derive the mean pattern. The mean pattern is computed by first setting $\widehat{J}_0 = s_0$ and $\widehat{T}_0 = 0$, and then compute,

$$\widehat{J}_h = \underset{j}{argmax}\, Pr(J_h = j | J_{h-1} = \widehat{J}_{h-1}, T_{h-1} = \widehat{T}_{h-1}). \tag{2.7}$$

Then compute

$$\widehat{\tau}_h = \mathbb{E}[\tau_h | J_h = \widehat{J}_h, J_{h-1} = \widehat{J}_{h-1}, T_{h-1} = \widehat{T}_{h-1}]. \tag{2.8}$$

Then $\widehat{T}_h = \widehat{T}_{h-1} + \widehat{\tau}_h$. And the mean pattern is ended when the predicted trajectory reaches absorbing state s_a.

3 Real Data Studies

Our data is collected from 17 women wearing oura ring during their pregnancy. Our data consists of consecutive sleep observations from observation of up to 12 months.

We fit the model using the data we gathered. In implementation, $\beta_{ij}(\cdot)$ is estimated using B-spline functions. After acquiring estimation, we adopt the method in section 2.3 and derive the mean pattern. The result is shown in figure 1. When drawing this plot, we did not include covariate into the model.

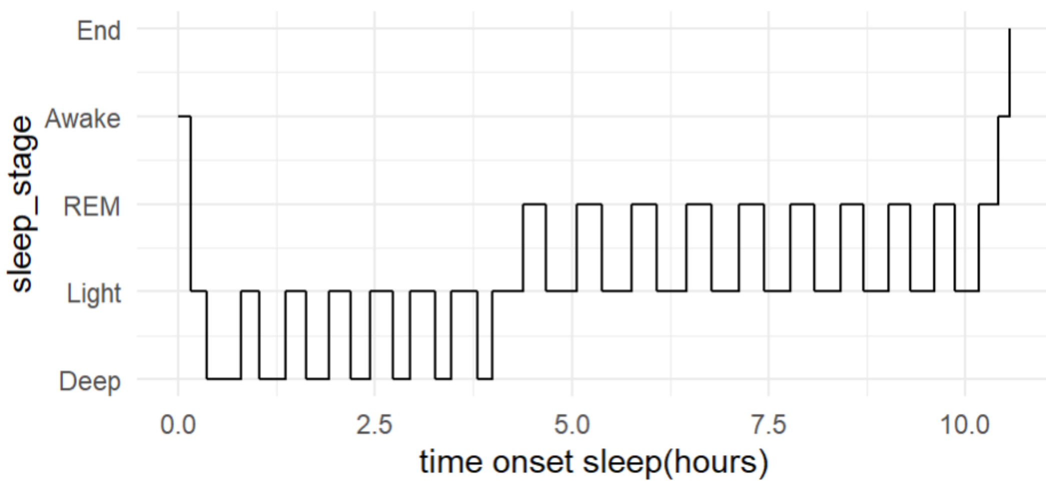

Figure 1: Predicted Mean Pattern

The mean pattern is simple and representative compared to the rather noisy original data. We draw three conclusions from the fitted mean pattern. 1. At early sleep process, one alternates between light and deep sleep for several times. 2. During later half of the sleep process, one alternates between light and REM stage for several times. 3. The duration of each sleep stage undergoes gradual and monotonous change. The duration of deep sleep is increasing while the sleep duration of light sleep first increase then decrease.

We also fit the mean pattern for different phases of pregnancy. We use data from 10-19, 20-29, 30-39 months of pregnancy seperately to fit model and to derive mean pattern. The results are shown in figure 2.

We observe that as time passes, pregnant women's sleep quality is deteriorating as they have less deep sleep. They also have trouble getting into deep sleep as observed from the mean pattern of 30-39 weeks of pregnancy.

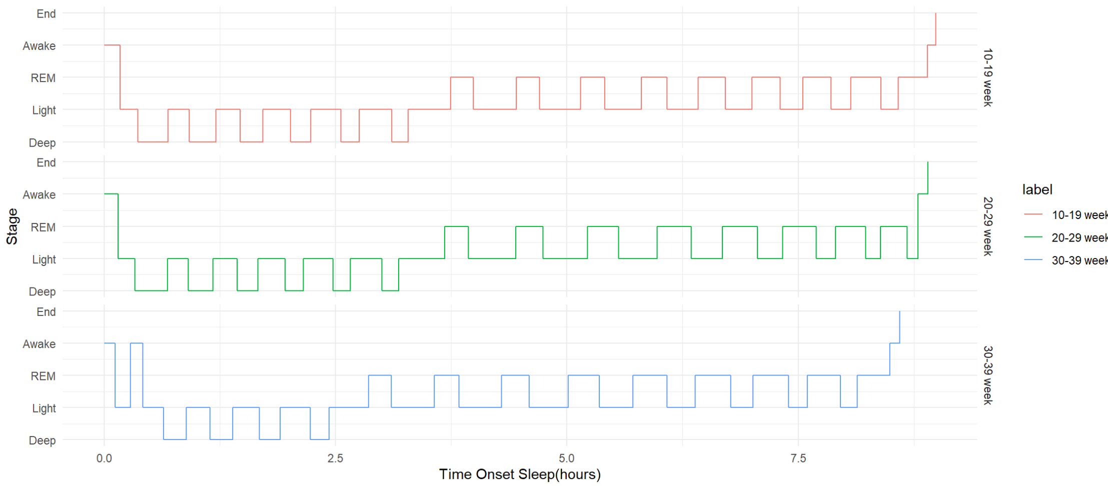

Figure 2: Change of Mean Pattern During Pregnancy

Finally, we fit mean pattern of different individuals. It turns out that there is strong heterogeneity from different individual's sleep, however they all follow the overall mean patten which is first deep and light sleep, and then light and REM sleep. Serious deviations from this pattern may indicate abnormal health conditions. For example, as figure 3 has shown, individual 19 has abnormal mean pattern, which might be explained by the fact that she is over 30 years old and is severely overweight.

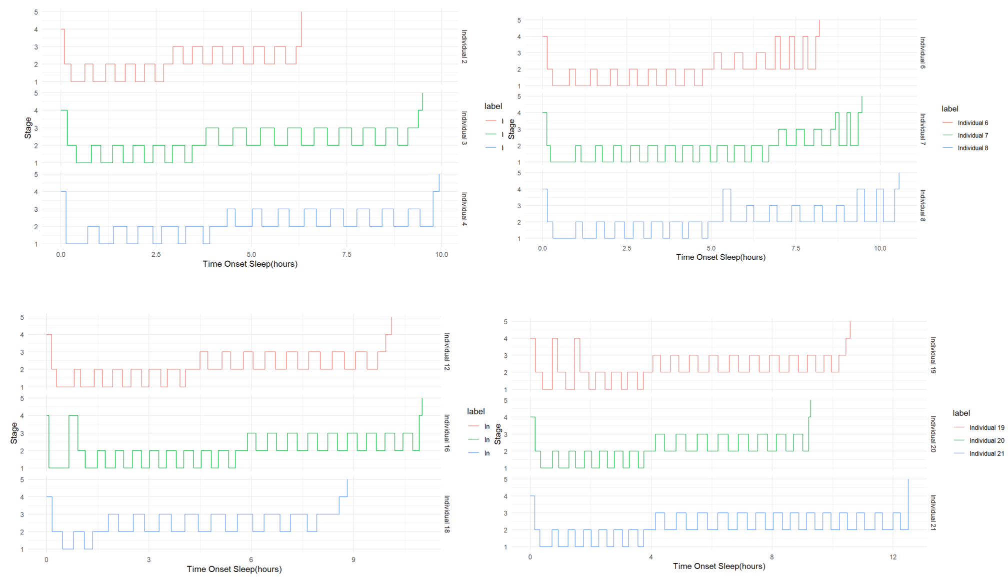

Figure 3: Individual Mean Patterns

4 Conclusion and Discussions

The conclusion and contribution of our work are as follows. Firstly, we give a simple definition to nonhomogeneous discrete semi-markov process. We prove its likelihood can be decomposed into binary regression through proper approximation. Secondly, we use a novel parametrization for nonhomogeneity which can separate the baseline transition intensity and the effect of nonhomogeneity meanwhile can be efficiently estimated. Third, we propose a method of deriving mean pattern and our empirical results are consistent with existing medical sleep studies.

Our drawbacks and future work are as follows: the estimation's consistency and inference have not been discussed in this report. Though Lee et al. (2018) gave a consistency theory in their paper, which demonstrates the estimation is asymptotic normal and unbiased, and they use a sandwiched estimator for the variance. However, their proof directly neglects the approximation error of the likelihood therefore is not the true error under the proposed model. Our future work can include deriving the comprehensive consistency theory for such estimation.

References

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical methods in medical research* **11** 91–115.

Asanjarani, A., Liquet, B. and Nazarathy, Y. (2021). Estimation of semi-markov multistate models: a comparison of the sojourn times and transition intensities approaches. *The international journal of biostatistics* **18** 243–262.

Fahrmeir, L. and Kaufmann, H. (1987). Regression models for non-stationary categorical time series. *Journal of time series Analysis* **8** 147–160.

Fahrmeir, L. and Klinger, A. (1998). A nonparametric multiplicative hazard model for event history analysis. *Biometrika* 581–592.

Fokianos, K. and Kedem, B. (1998). Prediction and classification of non-stationary categorical time series. *Journal of multivariate analysis* **67** 277–296.

Kang, M. and Lagakos, S. W. (2007). Statistical methods for panel data from a semi-markov process, with application to hpv. *Biostatistics* **8** 252–264.

Lee, M., Feuer, E. J. and Fine, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics* **74** 1468–1481.

Stoffer, D. S., Tyler, D. E. and McDougall, A. J. (1993). Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika* **80** 611–622.

Swihart, B. J., Caffo, B., Bandeen-Roche, K. and Punjabi, N. M. (2008). Characterizing sleep structure using the hypnogram. *Journal of Clinical Sleep Medicine* **4** 349–355.

Tutz, G., Schmid, M. et al. (2016). *Modeling discrete time-to-event data*. Springer.

A Proof of Theorem 2.1

Proof. The likelihood can be constructed using the joint conditional distribution of $\{J_h, \tau_h\}$ conditioning on J_{h-1}, T_{h-1}, which is,

$$\mathcal{L} = \prod_{h=1}^{H} f(J_h, \tau_h | J_{h-1}, T_{h-1}). \tag{A.1}$$

Note that there's no need to specify $p(J_0)$ since we always assume J_0 is fixed and known. Now we replace f with transition intensity function and derive,

$$\mathcal{L} = \prod_{h=1}^{H} \alpha_{J_{h-1}J_n}(\tau_h|T_{h-1}) S_{j_{n-1}}(\tau_h|T_{h-1}) \tag{A.2}$$

Then introduce indicator function and derive,

$$\mathcal{L} = \prod_{h=1}^{H} \prod_{i \in \mathcal{S}} \Big[\alpha_{iJ_h}(\tau_h|T_{h-1}) S_i(\tau_h|T_{h-1}) \Big]^{\mathbf{1}_{j_{n-1}=i}} \tag{A.3}$$

Finally, plug in approximation 2.4 and derive,

$$\mathcal{L} \approx \prod_{h=1}^{H} \prod_{i \in \mathcal{S}} \prod_{j \neq i} \Big[S_{ij}(\tau_h|T_{h-1})^{\mathbf{1}_{j_{n-1}=i}} \alpha_{ij}(\tau_h|T_{h-1})^{\mathbf{1}_{J_{n-1}=i, J_n=j}} \Big] \tag{A.4}$$

Finally, plug in the defination of $S_{ij}(\tau_h|T_{h-1})$ and the theorem is proved.

□