ASYMPTOTIC PROPERTIES OF MARKOV REGRESSION MODELS FOR LONGITUDINAL CATEGORICAL DATA IN CONTINUOUS TIME

Wei-Hsiung Chao
Department of Statistics

Michael R. Kosorok, PhD
Department of Biostatistics

MADISON, WISCONSIN
ASYMPTOTIC PROPERTIES OF MARKOV REGRESSION MODELS FOR LONGITUDINAL CATEGORICAL DATA IN CONTINUOUS TIME

BY WEI-HSIUNG CHAO AND MICHAEL R. KOSOROK

University of Wisconsin-Madison

We consider regression models for longitudinal categorical data that are generated by piecewise homogeneous continuous-time Markov processes. A general theorem which extends the work of Kaufmann (1987) is developed for the asymptotic properties of the corresponding maximum likelihood estimator when absorbing states are permitted. For covariates and outcomes that are observed at the same time points, the asymptotic properties are shown under mild and easily verifiable conditions which do not involve the unknown parameter. Sufficient conditions are also discussed when covariates are observed more frequently than outcomes.

Key words and phrases. Continuous time Markov chains, maximum likelihood estimation, nonhomogeneous Markov chains, panel data, piecewise homogeneous Markov processes.
1. Introduction. Longitudinal categorical data refer to repeated measurements of categorical outcomes collected from a sample of individuals over a certain period of time. In many research problems, scientific interest centers on the pattern of change in the outcome over time, and it is common in this setting to use Markov regression models to study the predictive relationship between the current values of the outcome and past values of the covariates and outcome. An important class of such models for discrete time are the regression models for nonstationary categorical time series proposed by Fahrmeir and Kaufmann (1987). These authors give a flexible and parsimonious treatment of higher order dependence as well as nonstationarity in their discrete time Markov regression models when absorbing states are not permitted. The asymptotic estimation theory for these models was developed by Kaufmann (1987), and is based on a general theorem for maximum likelihood estimation from a single stochastic process in discrete time.

Our main interest, however, lies in continuous time Markov regression models for longitudinal categorical data. Before proceeding, it is worth noting the difference between life history data and panel data in the continuous time setting. Life history data refer to outcomes that are observed continuously until a censoring time, while panel data are only observed intermittently. In fact, panel data might be observed at different sets of unequally spaced time points for different subjects. Therefore, panel data can also be viewed as a collection of pruned processes where the information of exact transition times is not available. Examples of Markov regression models, in the context of life history data, include the works of Kay (1982) and Anderson, Hansen, and Keiding (1991), where a proportional hazard regression model is assumed for each transition intensity element, and Tuma, Hannan, and Groeneveld (1979), where a log-linear regression model is assumed for each transition rate. In this set-up the (partial) likelihood is written in terms of the available data which includes the imbeded Markov chain and the waiting times between jumps. In contrast, in panel data setting the likelihood for time homogeneous models can be written in terms of transition probability matrices which are matrix exponentials of transition intensities. The matrix exponentials involved in the likelihood can make it challenging to compute deriva-
tives. Kalbfleish and Lawless (1985) derived a formula for the first derivative and proposed a quasi-Newton algorithm for implementing maximum likelihood estimation of time homogeneous Markov regression models for panel data with time independent baseline covariates. No asymptotic theory for maximum likelihood estimation in pruned Markov processes appears in the literature although some discussion on the uniqueness of maximum likelihood estimates has been given by Kalbfleish and Lawless (1985). Our focus in this paper will be with panel data.

Assuming that the underlying continuous time Markov process is homogeneous is unduly restrictive. In many applications, nonhomogeneous Markov models seem more appropriate because changes in time dependent covariates may introduce changes in transition intensities. In settings where covariate information is collected intermittently, such as with panel data, piecewise homogeneous Markov models are an appropriate and practical class of nonhomogeneous models. Piecewise homogeneous Markov process arising from piecewise constant covariates will be the basis for the panel data regression models we describe in this paper. Specifically we will assume that the intensity matrix $Q(t)$ for the underlying continuous time Markov process is a matrix valued function $h$ of the linear combination $Z'(t)\beta$, where the matrix $Z(t)$ is piecewise constant and depends only on the covariate value at time $t$, and where $\beta$ is an unknown parameter vector.

A short introduction to piecewise homogeneous Markov regression models is given in section 2, together with some examples. Statistical inference in these models is based on large sample properties of the maximum likelihood estimator. As mentioned earlier, the likelihood for a pruned continuous time Markov process can be written in terms of transition probabilities as is done with discrete time Markov chains. This approach is presented in section 3 and is accomplished through a multinomial set-up which adapts Kaufmann's (1987) time series setting to our pruned continuous time processes setting. The necessary general theorem on maximum likelihood estimation for multiple pruned continuous time processes is discussed in section 4 which extends Kaufmann's (1987) general theorem for a single discrete time stochastic process. Based on this general theorem we generalize, also in section 4,
Kaufmann's (1987) asymptotic estimation theory for a single nonstationary categorical time series with no absorbing states to multiple pruned continuous time Markov processes with absorbing states. Specific sufficient conditions are given in section 5 for several important settings.

2. **Piecewise Homogeneous Markov Processes.** Let \( \{Y(t), t > 0\} \) be a continuous time stochastic process taking on values in the set \( \{1, \ldots, k\} \). The process \( \{Y(t), t > 0\} \) is said to be a Markov process if for any \( a \in \{1, \ldots, k\} \), and \( s \leq t \),

\[
P\{Y(t) = a|Y(u), u \leq s\} = P\{Y(t) = a|Y(s)\}.
\]

In general, the above conditional probability may depend on both \( t \) and \( s \). If, furthermore,

\[
P\{Y(t) = a|Y(s)\} = P\{Y(t-s) = a|Y(0)\}.
\]

the Markov process is said to be time homogeneous. For \( s \leq t \), denote by \( P(s, t) \) the \( k \times k \) transition probability matrix with entries

\[
p_{ab}(s, t) = P\{Y(t) = b|Y(s) = a\},
\]

for \( a, b = 1, \ldots, k \). By the Markov property, the matrices \( P(s, t) \) satisfy the Chapman-Kolmogorov equations

\[
P(s, t) = P(s, u)P(u, t), \quad s \leq u \leq t,
\]

with side conditions \( p_{ab}(s, t) \geq 0 \), and \( \sum_{c=1}^{k} p_{ac}(s, t) = 1 \) for every \( a, b \), and \( s \leq t \). Conversely, such a solution is the transition probability matrix of a Markov process. Let \( Q(t) = (q_{ab}(t)) \) be the transition intensity matrix defined by

\[
q_{ab}(t) = \lim_{\Delta t \to 0} p_{ab}(t, t + \Delta t)/\Delta t, \quad a \neq b,
\]

and \( q_{aa}(t) = -\sum_{b \neq a} q_{ab}(t), a = 1, \ldots, k \). If \( Q(t) \) is continuous, one can prove that the matrices \( P(s, t), s \leq t \), satisfy the forward and backward Kolmogorov differential equations

\[
\frac{\partial P(s, t)}{\partial t} = P(s, t)Q(t), \quad s \leq t,
\]

\[
\frac{\partial P(s, t)}{\partial s} = -Q(t)P(s, t), \quad s \leq t,
\]
with initial conditions $\mathbf{P}(t, t) = \mathbf{I}$ (see Feller, 1968). Usually an explicit form for the solution $\mathbf{P}(s, t)$ of the above differential equations may not be obtained for general forms of the transition intensity matrix $\mathbf{Q}(t)$, although $\mathbf{P}(s, t)$ can be expressed in terms of a product-integral of $\mathbf{Q}(t)$, as indicated by Gill and Johansen (1990).

In this paper, we are interested in piecewise homogeneous models, where $\mathbf{Q}(t)$ is piecewise constant. In this case, the matrix $\mathbf{P}(s, t)$ can be determined explicitly, by use of the Chapman-Komogorov equations, as follows. Let $\mathbf{Q}(t) = \mathbf{Q}_i$, for $u_{i - 1} < t < u_i$, $i = 1, 2, \ldots$ If $u_{j - 1} \leq s \leq u_j$ and $u_{m - 1} \leq t \leq u_m$, then

$$\mathbf{P}(s, t) = \exp[\mathbf{Q}_j(u_j - s)] \prod_{i=j+1}^{m-1} \exp[\mathbf{Q}_i(u_i - u_{i - 1})] \exp[\mathbf{Q}_m(t - u_{m - 1})],$$

where $\exp[\mathbf{A}t] = \sum_{r=0}^{\infty} \mathbf{A}^r t^r / r!$.

In many longitudinal studies, information on each individual's covariates is often available, and it is desirable to adjust the intensity matrix $\mathbf{Q}(t)$ for the effect of covariates using regression model techniques. We will assume the following structure for the underlying response process $Y(t)$.

(i) $Y(t)$ is a first order Markov process generated by an intensity matrix $\mathbf{Q}(t)$.

(ii) The covariate $\mathbf{Z}(t)$ is a right continuous piecewise constant $p \times n$ nonrandom matrix valued function, which is assumed to be constant between the times we observe it.

(iii) The intensity matrix $\mathbf{Q}(t)$ is determined by the covariate matrix $\mathbf{Z}(t)$ and a $p \times 1$ parameter vector $\mathbf{\beta}$ through an injective matrix valued link function $\mathbf{h}$ of the form $\mathbf{Q}(t) = \mathbf{h}(\mathbf{Z}'(t)\mathbf{\beta})$.

(iv) There might be several absorbing states. Once an absorbing state is observed at a time $s$, no further observation of $Y(t)$, $t > s$, will be obtained.

The right continuity of $\mathbf{Z}(t)$ in (ii) is only required for notational convenience in the following sections. It can be replaced by left continuity, or a mixture of right and left continuities.
The following examples briefly review the past work of piecewise homogeneous Markov regression models, and show how they fit in the framework of the above structure.

(1) The time-homogeneous model. \( Z(t) = Z \) in this situation.

(i) without covariates: \( \beta = \{ q_{ab}; a \neq b, q_{ab} > 0 \} \), and \( Z = I \) in this case.

(ii) with baseline covariates:

(a) The log-linear model has a parametrization of the form \( q_{ab} = \exp(z'\beta_{ab}) \). In this case, \( \beta = \{ \beta_{ab}' \} \) and \( Z = I \otimes z \), where \( \otimes \) denotes the Kronecker product. Many authors have considered this and similar kind of parametrization; for example, see Kay (1986) and Kalbfleish and Lawless (1985). Since each intensity element is modeled as a separate function of the covariates, there are no easily interpretable parameters over all states as in marginal models (Stram, Wei, and Ware 1988). In addition, this may not be a parsimonious model because it may result in numerous parameters even for a moderate number of categories in the outcome variable; see Kosorok and Chao’s (1995) comparison of parsimony that can be achieved by different parametrizations.

(b) Kosorok and Chao (1995) considered the following parametrization for a time reversible ergodic process. \( q_{ab} = \pi_b \mu(a, \land b)(a, v b) \), \( \logit(\sum_{b=1}^m \pi_b) = \alpha_m - z'\beta_{r} \), \( \mu(a, \land b)(a, v b) = \exp(\xi(a, \land b)(a, v b) + z'\gamma) \), where \( a \land b, a \lor b \) denote \( \min\{a, b\} \) and \( \max\{a, b\} \), respectively. In this case, it can be shown that the models can be formulated in terms of \( \beta, h, Z \) in the above general structure. Let there be \( k \) categories for the response variable. It is easy to see that \( \pi = (\pi_1, \ldots, \pi_k) \) is the equilibrium distribution for \( Q \) since it satisfies the equation \( \pi'Q = 0 \) (Ross, 1993). The quantities \( \mu = \{ \mu_{ab}; 1 \leq a < b \leq k \} \) reflect the rates of transfer between pairs of categories. Since \( \pi \) measures the direction of movement (trend) and \( \mu \) measures the rate of movement (diffusion), this model has the capacity of assessing the effect of covariates on the general direction and rate of movement between ordinal categories of the response.
(2) The piecewise time-homogeneous model. By fitting separate models to different time intervals in (1)(i), or by allowing the covariates to be piecewise constant in (1)(ii), we get piecewise time-homogeneous models. Such methods have been considered by Faddy (1976), Kalbfleish and Lawless (1985), Gentleman et. al. (1994), and Kosorok and Chao (1995).

3. Maximum Likelihood Estimation. First we need to introduce some notation. To permit $Z_i(t)$ to change possibly more or less frequently than the rate at which we observe $Y_i(t)$, we consider the following nesting and overlapping structure of observation times. For individual $i, i = 1, \ldots, N$, define the quantities:

- $t_{ij0} \equiv$ observation times of response, $j = 0 \ldots n_i$, and $n_i \geq 1$
- $y_{ija} \equiv$ the indicator of $\{ Y_i(t_{ij0}) = a \}$, where $a = 1, \cdots, k$
- $t_{ijl} \equiv$ observation times of response or covariate, $l = 0, \ldots, m_{ij}$, and $m_{ij} \geq 1$ for $j < n_i$
- $Z_{ijl} \equiv Z_i(t_{ijl})$, and $Z_{ij} = [Z_{ij0} \cdots Z_{ij(m_{ij}-1)}]$
- $Q_{ijl}(\beta) \equiv$ intensity matrix corresponding to $Z_{ijl}$
- $t_{ij(m_{ij})} \equiv t_{i(j+1)0}$ for all $j$, and $m_{i(n_i)} \equiv 0$

There is some redundancy here, in that $Z_{ij(m_{ij})} \equiv Z_{i(j+1)0}$, but this will simplify some of the notation used later.

For each $i = 1, \cdots, N$, and $j = 1, \cdots, n_i$, let $P_{ij}$ be the transition probability matrix from $t_{i(j-1)0}$ to $t_{ij0}$ given by

$$P_{ij} = \prod_{l=1}^{m_{ij}} \exp[Q_{i(j-1)(l-1)}(t_{i(j-1)l} - t_{i(j-1)(l-1)})].$$

Due to the assumption that $Q(t) = h(Z'(t)\beta)$, we can write $P_{ij}$ in the form

$$P_{ij} = g(Z_{i(j-1)}', \Delta t_{ij}), \text{ or } P_{ij} = g(Z_{i(j-1)}', \beta)$$

for simplicity, where the dependence on $\Delta t_{ij} = \{\Delta t_{i(j-1)l}\}_{l=1}^{m_{ij}}, \Delta t_{ijl} = t_{ijl} - t_{ij(l-1)}$, is suppressed. Here $g$ is a link function relating $P_{ij}$ to $Z_{i(j-1)}'\beta$. Also denote

$$y_{ij} = (y_{ij1}, \cdots, y_{ij(k-1)})', \text{ and } p_{ija} = (p_{ija1}, \cdots, p_{ija(k-1)})'. $$
for \(a = 1, \cdots, k\), where \(p_{ijab}\) is the \((a, b)\) entry of the matrix \(P_{ij}\). From expression (3.1), \(P_{ija}\) is a function of \(Z'_{i(j-1)}\beta\), say,
\[
P_{ija} = g_a(Z'_{i(j-1)}\beta).
\]
Let \(\pi_{ij}\) denote the conditional probability vector \(\pi_{ij} = E(y_{ij}|y_{i(j-1)}, \cdots, y_{i0})\), and \(\eta_{ij}\) its link function. If there is no absorbing state,
\[
\pi_{ij} = \sum_{a=1}^{k} y_{i(j-1)a} P_{ija}, \quad \text{and} \quad \eta_{ij} = \sum_{a=1}^{k} y_{i(j-1)a} g_a.
\]
In case there are absorbing states, let \(\Omega\) be the set of nonsorbing states, then
\[
\pi_{ij} = \sum_{a \in \Omega} y_{i(j-1)a} P_{ija}, \quad \text{and} \quad \eta_{ij} = \sum_{a \in \Omega} y_{i(j-1)a} g_a.
\]
Note that \(p_{ijab} > 0\) for all \(i, j, a, b\), where \(a\) is not an absorbing state; see appendix 6.1. Hence \(\pi_{ijb} = E_{\beta}(y_{ijb}|y_{i(j-1)}, \cdots, y_{i0}) > 0\), for all \(b\).

Since, conditioning on \(y_{i0}, \cdots, y_{i(j-1)}\), \(y_{ij}\) is multinomially distributed with probability vector \(\pi_{ij}\) and sample size 1, we can write the log-likelihood increment as
\[
\ell_{ij}(\beta) = \sum_{b=1}^{k} y_{ijb} \log \pi_{ijb}, \quad \pi_{ij} = \eta_{ij}(Z'_{i(j-1)}\beta).
\]

Accordingly, the score increment can be obtained by the chain rule as follows:

(3.2) \[
s_{ij}(\beta) = Z_{i(j-1)} D_{ij}(\beta) \Sigma_{ij}^{-1}(\beta)(y_{ij} - \pi_{ij}(\beta)),
\]

where \(D_{ij}(\beta) = [\partial \eta_{ij}(\gamma)/\partial \gamma']\) is calculated at \(\gamma = Z'_{i(j-1)}\beta\), and where \(\Sigma_{ij}^{-1}(\beta)\) is the inverse of \(\Sigma_{ij}(\beta) = \text{var}_{\beta}(y_{ij}|y_{i(j-1)}, \cdots, y_{i0})\).

The increment of conditional information, defined by
\[
G_{ij}(\beta) = \text{var}_{\beta}(s_{ij}(\beta)|y_{i(j-1)}, \cdots, y_{i0}),
\]
becomes
\[
G_{ij}(\beta) = Z_{i(j-1)} V_{ij}(\beta) Z'_{i(j-1)},
\]
where \(V_{ij}(\beta) = D_{ij}(\beta) \Sigma_{ij}^{-1}(\beta) D'_{ij}(\beta)\).
Let \( u_{ij} = \logit \circ \eta_{ij} \) and \( U_{ij}(\beta) = [\partial u_{ij}(\gamma)/\partial \gamma]' \) calculated at \( \gamma = Z'_{i(i-1)} \beta \). Here \( \logit \) is the \((k-1)\)-dimensional logit function defined by \( \logit(\pi) = (\ln(\pi_1/\pi_k), \ldots, \ln(\pi_{k-1}/\pi_k))' \). Then the score increment can also be expressed as

\[
(3.3) \quad s_{ij}(\beta) = Z_{i(i-1)} U_{ij}(\beta)(y_{ij} - \pi_{ij}(\beta)).
\]

Using this form, the increment \( H_{ij}(\beta) \) of observed information can be conveniently expressed as \( H_{ij}(\beta) = G_{ij}(\beta) - R_{ij}(\beta) \), with the remainder matrix

\[
(3.4) \quad R_{ij}(\beta) = \sum_{a=1}^{k-1} Z_{i(i-1)} W_{ija}(\beta) Z'_{i(i-1)}(y_{ija} - \pi_{ija}(\beta)),
\]

where \( W_{ija}(\beta) = \partial^2 u_{ija}(Z'_{i(i-1)}(\beta))/\partial \gamma \partial \gamma' \) and \( u_{ija} \) is the \( a \)'th component of \( u_{ij} \).

Let \( \nu = \sum_{i=1}^{N} n_{i} \). Denote the log-likelihood, the score, and the conditional information by \( \ell_{\nu}(\beta) \), \( s_{\nu}(\beta) \), \( G_{\nu}(\beta) \), respectively, where

\[
\ell_{\nu}(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_{i}} \ell_{ij}(\beta), \quad s_{\nu}(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_{i}} s_{ij}(\beta), \quad \text{and} \quad G_{\nu}(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{n_{i}} G_{ij}(\beta).
\]

Similarly, \( H_{\nu}(\beta) \) denotes the observed information. The conditional information \( G_{\nu}(\beta) \) plays an important role in the large sample theory of maximum likelihood estimators. By integration, we obtain the information \( E_{\nu}(\beta) = \text{var}_{\beta}(s_{\nu}(\beta)|\{y_{iij}\}_{i=1}^{N}) \) and the unconditional information \( F_{\nu}(\beta) = \text{var}_{\beta}s_{\nu}(\beta) \). Since the distribution of initial responses is not specified, we do not give an explicit expression for \( F_{\nu}(\beta) \).

We will use the following notation for convenience. Let \( A \) be a \( p \times q \) matrix with rows \( A_{1}, \ldots, A_{p} \). Denote \( \overline{A} \) to be the \( 1 \times pq \) row vector \([A_{1} \cdots A_{p}]\), where its \(((i-1)p+j)'\)th (or \(1, (i,j))\) element is the \((i,j)'\)th element of \( A \).

**Regularity Assumptions.** Throughout, we will assume the following regularity assumptions:

(i) The \( Y_{i}(t) \) are independent and follow the structure of a piecewise time-homogeneous Markov regression model as characterized in section 2 for all \( \beta \) out of the open set \( B \).

(ii) \( Z'_{iijl} \beta \) is in the domain \( D \) of \( h \), for all \( i, j, l \), and for all \( \beta \in B \).
(iii) The matrix-valued link function $h$ for the transition intensity matrix $Q$ is two times continuously differentiable, and $\partial h(\gamma)/\partial \gamma$ is of rank $n$.

Note that the observed information $H_\nu$ is not guaranteed to be positive definite whatever the link function $h$ is since the link function is applied to the transition intensity matrix $Q$ instead of the transition probability matrix $P$ or its variant. Hence the following asymptotic theory always refers to local maximum likelihood estimators. This is in contrast to the maximum likelihood approach for discrete time Markov processes, where the observed information matrix is positive semi-definite for sure if using a canonical link because it turns out to be the conditional information in this situation: see Farhmeir and Kaufmann (1987), Zeger and Qaqish (1989), and Slud and Kedem (1994).

4. A general theorem on MLE for panel data under a Markov assumption. Kaufmann (1987) presented a general theorem (Kaufmann 1987, Theorem 3) on the asymptotic theory of MLE for a single stochastic process in discrete time which is based on a martingale approach and which is particularly useful for conditional models. Since all that is required is a martingale structure, his general theorem can be extended to multiple pruned stochastic processes in continuous time. This is accomplished by choosing an appropriate martingale so that we can easily check the conditional Lindeberg condition for the martingale central limit theorem (see Hall and Heyde 1980), even in the presence of ties in observation times of responses. In the following we will first construct a filtration so that the resulting martingale is easy to work with in the above sense. Based on this small extension, and in view of the same likelihood form, we will then generalize Kaufmann’s (1987) asymptotic result for a single nonstationary categorical time series without absorbing states to the case of multiple pruned continuous time Markov processes with absorbing states.

4.1 Martingale framework and theorem on MLE. For each $i = 1, \cdots, N$, and $j = 1, \cdots, n_i$, define the $\sigma$-algebra $\mathcal{F}_{i,j-1}$ by

$$\mathcal{F}_{i,j-1} = \sigma\{y_{ij'}|j' = 0, \cdots, j - 1\}.$$
Then $\pi_{ij} = E_\beta(y_{ij} \mid i, j - 1)$. Here and hereafter, dependence on the true parameter vector $\beta$ is mostly suppressed for simplicity. Note that $\{\mathcal{F}_{i,j-1}, j = 1, \ldots, n_i\}$ is a filtration for subject $i$.

Let $m_{ij} = y_{ij} - \pi_{ij}$ (observed minus conditional expected value). Then $\{\sum_{j' = 1}^{j} m_{ij'}\}_{j = 1}^{n_i}$ forms a martingale, with respect to $\{\mathcal{F}_{i,j}\}_{j = 0}^{n_i - 1}$, for each $i$, and so does $\{\sum_{j' = 1}^{j} s_{ij'}\}_{j = 1}^{n_i}$ (see (3.2) or (3.3)). Now, we need to formulate a filtration so that the corresponding score process will be a martingale. A natural choice would be a filtration in continuous time, $\{\mathcal{F}_t\}$, where $\mathcal{F}_t$ is defined by

$$\mathcal{F}_t = \sigma\{y_{ij} \mid t_{ij} \leq t; i = 1, \ldots, N; j = 0, \ldots, n_i\}.$$ 

Let $m_t = \sum_{t_{ij} \leq t} m_{ij}$, $s_t = \sum_{t_{ij} \leq t} s_{ij}$. Note that $\pi_{ij} = E(y_{ij} \mid \mathcal{F}_{i,j-1}) = E(y_{ij} \mid \mathcal{F}_{i,j-1})$, and, for $s < t$,

$$E(\mathcal{F}_t) = m_s + \sum_{s < t_{ij} \leq t} E[E(m_{ij} \mid \mathcal{F}_{t_{ij}}) \mid \mathcal{F}_s] = m_s + 0 = m_s.$$ 

Hence $\{m_t\}$ is a martingale with respect to $\{\mathcal{F}_t\}$, and so is $\{s_t\}$. If there are no ties in observation times of responses, then $\{s_t\}$ and $\{\mathcal{F}_t\}$ can be used to check the conditional Lindeberg conditions described in the martingale central limit theorem. In case of ties, we can construct the following discrete filtration so that the conditional Lindeberg conditions can be easily verified for the corresponding discrete score process. Let $\nu = \sum_{i=1}^{N} n_i$, the total number of observed responses in addition to the initial responses. Consider an enumeration of $\{(i,j); j = 1, \ldots, n_i\}_{i=1}^{N}$ based on observation times of responses in addition to the initial responses. Let

$$a_{ij} = \{(i', j') \mid t_{i'j'} < t_{ij}; j' = 1, \ldots, n_{i'}\}_{i'=1}^{N} \cup \{(i', j') \mid t_{i'j'} = t_{ij}; j' = 1, \ldots, n_{i'}\}_{i'=1}^{i},$$

and $\mu_{ij} = \#a_{ij}$, the number of ordered pairs in the set $a_{ij}$. Define

(4.1) \[ \mathcal{F}_{\mu_{ij}} = \bigvee_{(i', j') \in a_{ij}} \mathcal{F}_{i', j'-1}, \]

$$m_{\mu_{ij}} = \sum_{(i', j') \in a_{ij}} m_{i'j'},$$

and $s_{\mu_{ij}} = \sum_{(i', j') \in a_{ij}} s_{i'j'}$. Since $\pi_{ij} = E(y_{ij} \mid \mathcal{F}_{i,j-1}) = E(y_{ij} \mid \mathcal{F}_{\mu_{ij}}), \nu = \sum_{i=1}^{N} n_i$, and $\{m_{\mu_{ij}}\}_{\mu=1}^{\nu}$ are martingales with respect to $\{\mathcal{F}_{\mu}\}_{\mu=0}^{\nu-1}$. For later use, we define $\{E_{\mu}\}_{\mu=1}^{\nu}$, $\{F_{\mu}\}_{\mu=1}^{\nu}$, $\{H_{\mu}\}_{\mu=1}^{\nu}$, and $\{G_{\mu}\}_{\mu=1}^{\nu}$ similarly, and note that $\{H_{\mu} - G_{\mu}\}_{\mu=1}^{\nu}$ is a.
martingale. Using the filtration \( \mathcal{F}_\nu \) defined in (4.1), \( \Delta s_\nu = s_\nu - s_{\nu-1} \) becomes a single term \( s_{ij} \) for an appropriate \( i, j \). This makes it easy to verify the conditional Lindeberg condition. As mentioned earlier, since all that is required is a martingale structure, Kaufmann’s general theorem on MLE (Kaufmann 1987, Theorem 3) can be applied to multiple pruned continuous time processes such as panel data. There is no need to duplicate the proofs, and we will be satisfied with his general theorem with application to the panel data setting.

Based on this general theorem on MLE, Kaufmann (1987) developed a large sample theory for MLE for a nonstationary categorical time series in his Lemma 3 and Theorem 1. Using the multinomial setup, our likelihood has the same form as Kaufmann’s (1987), and it is possible to generalize his Lemma 3 and Theorem 1 to the pruned continuous time Markov processes setting. Since we want to develop a large sample theory for MLE for the more general situations, like \( \nu \to \infty \), and \( m_{ij} \geq 1 \), we will first give a general theorem on MLE for panel data under a Markov assumption, and then develop the theory for several specific situations in the next section. In addition to the regularity assumptions given in section 3, we need the following conditions to achieve consistency and asymptotic normality. These conditions are the key conditions in the proof of Kaufmann’s Lemma 3 and Theorem 1. These conditions refer to the true probability measure \( P = P_\beta \), and they should be checked for all \( \beta \in B \). In the sequel, \( \lambda_{\text{max}}(A) \) and \( \lambda_{\text{min}}(A) \) denotes the maximum and the minimum eigenvalues, respectively, for a symmetric matrix \( A \). In the following conditions, \( A_\nu \) refers to \( E_\nu \) or \( F_\nu \).

(D) **Divergence:** \( \lambda_{\text{min}}(A_\nu) \to \infty \), as \( \nu \to \infty \). If \( A_\nu = E_\nu \), it holds for any given initial responses.

(B) **Boundedness:** \( EW_\nu \) and \( W_\nu \) are uniformly bounded in \( \nu \), respectively, for \( A_\nu = E_\nu \) and \( F_\nu \), where \( W_\nu = \|A_\nu^{-1/2}(\sum_{i=1}^{N} \sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})A_\nu^{-T/2}\| \).

(C) **Compactness:** The values \( \{Z_{ijl}, \Delta t_{ijl}\} \) lie in a compact set \( C \) such that \( \{\Delta t_{ijl}\} \) is bounded away from zero.
REMARK For a positive definite matrix $A$, we denote $A^{1/2}$ and $A^{T/2}$ to be a left and a right square root, respectively, of $A$ if $A^{1/2}A^{T/2} = A$. Also let $A^{-1/2} = (A^{1/2})^{-1}$, $A^{-T/2} = (A^{T/2})^{-1}$. Note that $A^{1/2}$ and $A^{T/2}$ are unique up to an orthogonal transformation, since $A^{1/2}P$ and $P^T A^{T/2}$ are also left and right square roots, respectively, if $PP^T = I$. An example of a unique continuous version of the square root is the Cholesky square root. The left Cholesky square root is the lower triangular matrix with positive diagonal elements which is also a left square root.

**Lemma 1** Conditions (C) and (D) imply that, as $\nu \to \infty$,

$$
A_{\nu}^{-1/2}G_{\nu}A_{\nu}^{T/2} \to I, \quad A_{\nu}^{-1/2}s_{\nu}^{(2)}A_{\nu}^{T/2} \to I,
$$

in probability unconditionally, where $s_{\nu}^{(2)} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} s_{ij} s_{ij}'$. If $A_{\nu} = E_{\nu}$, the expressions hold also for any given initial responses.

**Theorem 1** Under conditions (D), (B), and (C), the probability that a locally unique MLE exists converges to one. Moreover, there exists a sequence $\{\hat{\beta}_{\nu}\}$ of MLE’s which is consistent and asymptotically normal,

$$
G_{\nu}^{T/2}(\hat{\beta}_{\nu} - \beta) \to_d N(0, I), \quad [s_{\nu}^{(2)}]^{T/2}(\hat{\beta}_{\nu} - \beta) \to_d N(0, I),
$$

as $\nu \to \infty$, with the right Cholesky square roots $G_{\nu}^{T/2}$, $[s_{\nu}^{(2)}]^{T/2}$.

**Remarks.** In brief, Kaufmann’s (1987) assumption $N(i)$ is satisfied from (3.2). Conditions (C) and (D) imply assumptions $N(ii)$ and $N(iii)$. Together with condition (B), assumption $N(iv)$ is satisfied.

As a consequence of Theorem 1, the local asymptotic normality condition (see Ibragimov and Has’minskii 1981 chapter 2) and some efficiency results for $\hat{\beta}_{\nu}$ can be obtained as stated in Kaufmann’s (1987) Lemma 1 and Theorem 2 with his $E_{\nu}$ replaced by our $A_{\nu}$—which could be either $E_{\nu}$ or $F_{\nu}$—and his assumption A replaced by our conditions (D), (B), and (C). This can be done by following the same arguments as Kaufmann.
4.2 Proofs. As shown in Kaufmann (1987) the $\delta$-coefficient technique of Dobrushin (1956) defined on models without absorbing states is very important in proving his Lemma 3, and we will use the same technique in proving our Lemma 1. To permit absorbing states we extend the definition of the $\delta$-coefficient as follows. Let there be $k$ states and $\omega$ nonabsorbing states. Let $\Omega$ denote the set of nonabsorbing states. The $\delta$-coefficient of a $k \times k$ stochastic matrix $P = (p_{ab})$ is defined to be

$$\delta(P) = \frac{1}{2} \max_{a, c \in \Omega} \sum_{b \in \Omega} |p_{ab} - p_{cb}|.$$ 

It can be easily shown that $\delta(P) \leq 1$,

$$\min_{a, b \in \Omega} p_{ab} \geq c \Rightarrow \delta(P) \leq 1 - \omega c,$$  

(4.4)

$$\delta(P_1 P_2) \leq \delta(P_1) \delta(P_2),$$  

(4.5)

for stochastic matrices $P_1, P_2$. The following inequalities can be derived with the $\delta$-coefficient, where the first inequality is a natural extension of Kaufmann's (1987) Lemma 2 to models with absorbing states.

**Lemma 2**  
(i) Let $y_1, y_2$ be random variables with the same finite state space $\{1, \ldots, k\}$ and transition probability matrix $P = (p_{ab})$, where $p_{ab} = P(y_2 = b | y_1 = a)$. If the random variable $x_i$ is a function of $y_i$, $i = 1, 2$, then

$$|\text{cov}(x_1, x_2)| \leq 2\delta(P) E|x_1| \max|x_2|.$$ 

(ii) Let $y_1 \neq y_2, y_3 \neq y_4$ be random variables with the same finite state space $\{1, \ldots, k\}$ and transition probability matrices $P_i = (p^i_{ab})$, where $p^i_{ab} = P(y_i = b | y_{i-1} = a)$, $i = 3, 4$. If the random variable $x_i$ is a function of $(y_{2i-1}, y_{2i})$, $i = 1, 2$, then

$$|\text{cov}(x_1, x_2)| \leq \begin{cases} 2\delta(P_3) E|x_1| \max|x_2|, & y_1, y_2, y_3 \in \Omega, (y_1, y_2) \neq (y_3, y_4), \quad y_2 \neq y_3, \\ 2E|x_1| \max|x_2|, & \text{otherwise}. \end{cases}$$
Proof of Lemma 2. (i) It is obvious that the argument of Kaufmann’s (1987) Lemma 2 applies here with our new definition of the \( \delta \)-coefficient.

(ii) Since it is always true that \( |\text{cov}(x_1, x_2)| \leq 2E|x_1| \max|x_2| \), we will focus on the case where \( y_1, y_2, y_3 \in \Omega, (y_1, y_2) \neq (y_3, y_4) \), and \( y_2 \neq y_3 \). Let \( p_{ab}^{12} \), \( p_{cd}^{34} \) be the marginal distributions of \((y_1, y_2), (y_3, y_4)\), respectively. Using the identity \( \text{cov}(x_1, x_2) = E\{x_1[E(x_2|y_1, y_2) - E(x_2)]\} \), we have

\[
\text{cov}(x_1, x_2) = \sum_{a, b \in \Omega} x_1(a, b) p_{ab}^{12} \sum_{c, d \in \Omega} x_2(c, d) (p_{cd}^{3} - p_{cd}^{34}), \quad \text{and}
\]

\[
|\text{cov}(x_1, x_2)| \leq E|x_1| \max_{c, d \in \Omega} |x_2(c, d)| \max_{b \in \Omega} \sum_{c, d \in \Omega} |p_{cd}^{3} - p_{cd}^{34}|.
\]

Note that \( p_{cd}^{34} = \sum_{a, b \in \Omega} p_{cd}^{3} p_{ab}^{12} \), where \( c \in \Omega \). Therefore,

\[
\sum_{c, d \in \Omega} |p_{cd}^{3} p_{ab}^{12} - p_{cd}^{34}| \leq \sum_{a, b \in \Omega} p_{ab}^{12} \sum_{c, d \in \Omega} |p_{cd}^{3} p_{bc}^{3} - p_{cd}^{34} p_{bc}^{3}| \leq \max_{b \in \Omega} \sum_{c, d \in \Omega} |p_{bc}^{3} - p_{bc}^{34} p_{bc}^{3}|.
\]

The result follows. \( \square \)

Proof of Lemma 1. Assume that \( A_{\nu} = F_{\nu} \). To prove \( F_{\nu}^{-1/2}G_{\nu}F_{\nu}^{-T/2} \to_d I \), it suffices to show that

\[
\lambda' F_{\nu}^{-1/2}G_{\nu}F_{\nu}^{-T/2} \lambda \to_d \lambda' \lambda = 1,
\]

for all \( \lambda \) with \( \|\lambda\| = 1 \). Define

\[
v_{ij} = v_{ij}^{(\nu)} = \lambda' F_{\nu}^{-1/2} Z_{i(j-1)} V_{i} Z_{i(j-1)}' F_{\nu}^{-T/2} \lambda,
\]

for \( i = 1, \ldots, N, j = 1, \ldots, n_i \); then

\[
\sum_{i=1}^{N} \sum_{j=1}^{n_i} v_{ij} = \lambda' F_{\nu}^{-1/2} G_{\nu} F_{\nu}^{-T/2} \lambda \quad \text{and} \quad \sum_{i=1}^{N} \sum_{j=1}^{n_i} E v_{ij} = 1.
\]

Therefore, it suffices to show that

\[
(4.6) \quad \text{var} \sum_{i=1}^{N} \sum_{j=1}^{n_i} v_{ij} \to 0,
\]

as \( \nu \to \infty \). This can be established by using some of the \( \delta \)-coefficient results of Dobrushin (1956). Since the random variable \( v_{ij} (\geq 0) \) depends only on \( y_{i(j-1)} \), it follows from Lemma 3(i) that

\[
|\text{cov}(v_{ij_1}, v_{ij_2})| \leq 2\delta(P_{ij_1} \cdots P_{i(j_2-1)}) E v_{ij_1} M_{vi}, \quad 1 \leq j_1 \leq j_2 \leq n_i,
\]
where $M_{\nu i} = \max_{1 \leq j_2 \leq n_i} v_{ij_2}$, and $P_{ij}$ is the transition probability matrix given in (3.1). The compactness condition (C) guarantees that all entries of $P_{ij}$ are bounded away from zero. Together with (4.4), this ensures the existence of an $r < 1$ with $\delta(P_{ij}) \leq r$, for all $j \geq 1$, and for all $i \geq 1$. Applying (4.5) we get

$$\delta(P_{ij_1} \ldots P_{i(j_2-1)}) \leq r^{j_2-j_1}, \quad 1 \leq j_1 \leq j_2 \leq n_i.$$ 

Therefore,

$$\sum_{j_1=1}^{n_i} \sum_{j_2=j_1}^{n_i} \left| \text{cov}(v_{ij_1}, v_{ij_2}) \right| \leq \frac{2}{1-r} \sum_{j_1=1}^{n_i} E v_{ij_1} M_{\nu i}.$$ 

Now, by independence over subjects, $\text{var}(\sum_{i=1}^{N} \sum_{j=1}^{n_i} v_{ij}) = \sum_{i=1}^{N} \text{var}(\sum_{j=1}^{n_i} v_{ij})$. Using the inequality

$$\text{var} \sum_{j=1}^{n_i} v_{ij} \leq 2 \sum_{j_1=1}^{n_i} \sum_{j_2=j_1}^{n_i} \left| \text{cov}(v_{ij_1}, v_{ij_2}) \right|,$$

we end up with

$$\text{var} \sum_{i=1}^{N} \sum_{j=1}^{n_i} v_{ij} \leq c \sum_{i=1}^{N} \sum_{j_1=1}^{n_i} E v_{ij_1} M_{\nu} = c M_{\nu},$$

where $c = 4/(1-r)$ and $M_{\nu} = \max_{1 \leq i \leq N} M_{\nu i}$. Finally

$$M_{\nu} = \max_{1 \leq i \leq N} \max_{1 \leq j \leq n_i} v_{ij} \leq \sup_{1 \leq i \leq j \leq n_i} \lambda_{\max}(Z_{ij-1}) / \lambda_{\min}(F_{\nu}).$$

The compactness condition (C) implies that the numerator of the right side is finite. Therefore, the first part of (4.2) follows from $\lambda_{\min}(F_{\nu}) \to \infty$.

The second part of (4.2) can be shown by similar arguments, where $v_{ij}$ is now defined as

$$v_{ij} = v_{ij}^{(\nu)} = \lambda' F_{\nu}^{-1/2} S_{ij} S_{ij}' F_{\nu}^{-T/2} \lambda.$$ 

Since the random variable $v_{ij}^{(\nu)} \geq 0$ depends only on $y_{i(j-1)}$ and $y_{ij}$, it follows from Lemma 3(ii) that for $1 \leq j_1 \leq j_2 \leq n_i$,

$$\left| \text{cov}(v_{ij_1}, v_{ij_2}) \right| \leq \begin{cases} 2 \delta(P_{i(j_1+1)} \ldots P_{i(j_2-1)}) E v_{ij_1} M_{\nu i}, & j_1 + 2 \leq j_2, \\ 2 E v_{ij_1} M_{\nu i}, & \text{otherwise}. \end{cases}$$

The remaining arguments can be adjusted accordingly, where the constant $c$ involved in the inequality should be replaced by $4(1 + 1/(1-r))$. 

16
Similar arguments imply the result for $A_{\nu} = E_{\nu}$, for given initial responses, and it yields the unconditional result by integration. □

**Proof of Theorem 1.** With our martingale framework, it can be easily seen from (3.2) that Kaufmann's (1987) assumption N(i) is satisfied. From the previous lemma, Kaufmann’s assumption N(ii) is satisfied with $A_{\nu}(\beta)$ replaced by either $E_{\nu}(\beta)$ or $F_{\nu}(\beta)$, and with $V(\beta) = I$.

We now establish the conditional Lindeberg condition, Kaufmann’s (1987) assumption N(iii), for our score function. The increment of the score function is $\Delta s_{\nu} = Z_{ij(1)}D_{ij}\Sigma_{ij}^{-1}m_{ij}$, for appropriate $i, j$ such that $\nu = \mu_{ij}$. Since $\{Z_{ij}\}$ is in a compact set, sup$\nu_{1}\|\Delta s_{\nu}\| \leq c < \infty$. Hence $(\Delta s_{\nu})'A_{\nu}^{-1}\Delta s_{\nu} \leq c^2/\lambda_{\min}(A_{\nu})$. Since $\lambda_{\min}(A_{\nu}) \to \infty$, then for fixed $\varepsilon > 0$, we have $I_{\nu\mu}(\varepsilon) = 0, \mu = 1, \ldots, \nu$, if $\nu$ is very large, where $I_{\nu\mu}(\varepsilon)$ is the indicator function of $\{(\Delta s_{\mu})'A_{\nu}^{-1}\Delta s_{\mu} > \varepsilon^2\}$. Therefore, the conditional Lindeberg condition N(iii) is satisfied. The key step when verifying Kaufmann’s assumption N(iv) in his Theorem 1 is the formation of a condition, which is equivalent to our boundedness condition for the case $A_{\nu} = E_{\nu}$.

It is also clear from his argument that the condition should be adjusted to our boundedness condition if $A_{\nu} = F_{\nu}$. The verification of assumption N(iv) is very similar to the corresponding part in the proof of Kaufmann’s (1987) Theorem 1, and is omitted. This completes the proof of the first part of (4.3) with an appropriate square root $G_{\nu}^{T/2}$, based on Kaufmann’s general theorem on MLE (Kaufmann 1987, Theorem 3) with application to panel data. From the Remark (i) of Kaufmann’s (1987) Theorem 3, $G_{\nu}^{T/2}$ is appropriate if $G_{\nu}^{T/2}A_{\nu}^{-1/2}G_{\nu}^{T/2}$ is the right Cholesky square root $A_{\nu}^{-1/2}G_{\nu}^{T/2}A_{\nu}^{-1/2}$. However, (4.2) holds for any version of $A_{\nu}^{1/2}$. If $A_{\nu}^{1/2}$ is chosen as the Cholesky square root, then $G_{\nu}^{T/2}$ will share this property. This proves the first part of (4.3) with the Cholesky square root $G_{\nu}^{T/2}$. If, in addition, $[s_{\nu}(2)]^{1/2}$ is also chosen as the Cholesky square root, (4.2) implies that $[s_{\nu}(2)]^{1/2}G_{\nu}^{T/2} \to I$ in probability, which in turn proves the second part of (4.3) with the Cholesky square root $[s_{\nu}(2)]^{T/2}$. □

5. **Asymptotic Estimation Theory.** Based on Theorem 1 in the preceding section, we will first develop the asymptotic properties for models without absorbing states in the next
two subsections, and then for models with absorbing states in the subsection that follows.

5.1 Models without absorbing states, \( m_{ij} = 1 \). When establishing the large sample properties of MLE for categorical time series, an important step in Kaufmann’s proof (1987) is to formulate the inequality

\[
\inf_{t > i} \lambda_{\min}(V_t) > 0,
\]

where \( V_t \) is a matrix in the categorical time series setting that is similar to our \( V_{ij} \). That is, \( V_t = U_t \Sigma_t U_t' \), with \( \Sigma_t = \text{var}(y_t | \{y_s, s < t\}) \). In Kaufmann’s setting, \( V_t \) is positive definite since \( U_t \) is a nonsingular matrix of the same order as \( \Sigma_t \). In our setting, \( U_{ij} \) is not even a square matrix. In fact, it is an \( n \times (k - 1) \) matrix with \( n \) possibly greater than \( k - 1 \). Hence, it may not be possible to formulate a similar inequality for \( \lambda_{\min}(V_{ij}) \) in our setting. However, in view of the divergence condition (D), similar inequalities for \( E_\beta(V_{ij}(\beta) | \mathcal{F}_{i0}) \) and \( E_\beta V_{ij}(\beta) \) should be enough for our purpose. In the following we will revisit the conditional likelihood to see if another form of conditional information can help us tackle this problem.

A further examination reveals that the log-likelihood increment can also be given by

\[
\ell_{ij}(\beta) = \sum_{a=1}^{k} y_{i(j-1)a} \sum_{b=1}^{k} y_{ijb} \log p_{ijab}.
\]

Accordingly, the score increment can be obtained by the chain rule as follows:

\[
s_{ij}(\beta) = Z_{i(j-1)} \sum_{a=1}^{k} y_{i(j-1)a} D_{ija}(\beta) \Sigma_{ija}^{-1}(\beta)(y_{ij} - p_{ija}(\beta)),
\]

where \( D_{ija}(\beta) = \partial g_a(\gamma)/\partial \gamma' \) evaluated at \( \gamma = Z_{i(j-1)^t} \beta \), and where \( \Sigma_{ija}^{-1}(\beta) \) is the inverse of \( \Sigma_{ija}(\beta) = \text{var}_\beta(y_{ij} | y_{i(j-1)a} = 1) \). The increment of conditional information becomes

\[
G_{ij}(\beta) = Z_{i(j-1)} V_{ij}(\beta) Z_{i(j-1)^t},
\]

where \( V_{ij}(\beta) = \sum_{a=1}^{k} y_{i(j-1)a} D_{ija}(\beta) \Sigma_{ija}^{-1}(\beta) D_{ija}'(\beta) \).

Let \( \tilde{V}_{ij}(\beta) = \sum_{a=1}^{k} D_{ija}(\beta) \Sigma_{ija}^{-1}(\beta) D_{ija}'(\beta) \). Denote \( \tilde{\Sigma}_{ij}(\beta) \) to be the \( k(k - 1) \times k(k - 1) \) block diagonal matrix given by \( \tilde{\Sigma}_{ij}(\beta) = \text{diag}(\{\Sigma_{ija}(\beta)\})_{a=1}^{k} \). Also let \( \tilde{D}_{ij}(\beta) \) be the \( n \times k(k - 1) \)
matrix given by

$$\widetilde{D}_{ij}(\beta) = [D_{ij1}(\beta) \cdots D_{ijk}(\beta)] .$$

Using these notations, $V_{ij}(\beta)$ can be expressed as

$$\widetilde{V}_{ij}(\beta) = \widetilde{D}_{ij}(\beta) \Sigma^{-1}_{ij}(\beta) \widetilde{D}_{ij}(\beta).$$

Since now $\Sigma^{-1}_{ij}(\beta)$ is a square matrix with higher order than $\Sigma^{-1}_{ij}(\beta)$, it is possible to formulate an inequality for $\widetilde{V}_{ij}(\beta)$ parallel to (5.1). Comparing the forms of $V_{ij}(\beta)$ and $\widetilde{V}_{ij}(\beta)$, we will be able to establish inequalities, similar to (5.1), for $E_{\beta}(V_{ij}(\beta)|\mathcal{F}_{i,0})$ and $E_{\beta}V_{ij}(\beta)$. These will be sufficient for the purpose of developing large sample properties of the MLE in our setting.

To deal with the general case, where $N$ might go to infinity, we consider the following condition regarding the behaviour of initial responses.

(P) $\inf_{1 \leq i, a} \ P_{\beta}(y_{i0a} = 1) > 0$, where $a$ is over all states.

Lemma 3 The condition (C) implies that

(5.3) $$\inf_{1 \leq i, 1 \leq j} \lambda_{\min} \widetilde{V}_{ij}(\beta) > 0,$$

(5.4) $$\inf_{1 \leq i, 1 \leq j} \lambda_{\min} E_{\beta}(V_{ij}(\beta)|\mathcal{F}_{i,0}) \geq c_1 \inf_{1 \leq i, 2 \leq j} \lambda_{\min} \widetilde{V}_{ij}(\beta),$$

(5.5) $$\inf_{1 \leq i, 2 \leq j} \lambda_{\min} E_{\beta}V_{ij}(\beta) \geq c_2 \inf_{1 \leq i, 2 \leq j} \lambda_{\min} \widetilde{V}_{ij}(\beta),$$

with some constants $c_1, c_2 > 0$. If, in addition, condition (P) is satisfied, then

(5.6) $$\inf_{1 \leq i, 2 \leq j} \lambda_{\min} E_{\beta}V_{ij}(\beta) \geq c_3 \inf_{1 \leq i, 1 \leq j} \lambda_{\min} \widetilde{V}_{ij}(\beta),$$

with some constant $c_3 > 0$.

Two other conditions on the spread of the covariates are needed which, together with the above lemma, will imply the divergence condition (D) and the boundedness condition (B).

$$S_1 \quad \lambda_{\min} \left( \sum_{i=1}^{N} \sum_{j=1}^{n_i} Z_{ij(i-1)} Z_{ij(i-1)}' \right) \rightarrow \infty.$$
\((S_2)\) \(\lambda_{\min}(\sum_{i=1}^{N} \sum_{j=2}^{n_i} Z_{i(j-1)}' Z_{i(j-1)}) \to \infty\) such that
\[
\frac{\lambda_{\max}(\sum_{i=1}^{N} \sum_{j=1}^{n_i} Z_{i(j-1)} Z_{i(j-1)}')}{\lambda_{\min}(\sum_{i=1}^{N} \sum_{j=2}^{n_i} Z_{i(j-1)} Z_{i(j-1)}')} \leq c, \quad \nu \geq \nu_1,
\]
for some constant \(c > 0\), and some integer \(\nu_1\).

An example of a stronger condition which ensures that \((S_2)\) is satisfied would be
\[
\lambda_{\min}(\sum_{i=1}^{N} \sum_{j=2}^{n_i} Z_{i(j-1)} Z_{i(j-1)}') \geq c \nu, \quad \nu \geq \nu_1,
\]
for some constant \(c > 0\), and some positive integer \(\nu_1\). Based on Theorem 1 and the above lemma, the asymptotic properties are developed below for several scenarios which may be useful in different settings. It is interesting to note that the sufficient conditions given in (i) and (ii) below do not involve the parameter \(\beta\).

**Corollary 1**

(i) For bounded \(N\), under conditions (C) and (S_1), the conditions (B) and (D) are fulfilled with \(A_{\nu}(\beta) = E_{\nu}(\beta)\) or \(F_{\nu}(\beta)\).

(ii) For general \(N\), under conditions (C) and (S_2), the conditions (B) and (D) are fulfilled with \(A_{\nu}(\beta) = E_{\nu}(\beta)\) or \(F_{\nu}(\beta)\).

(iii) For general \(N\), under conditions (C), (S_1), and (P), the conditions (B) and (D) are fulfilled with \(A_{\nu}(\beta) = F_{\nu}(\beta)\).

Hence, in either case, the MLE asymptotically exists and is consistent and asymptotically normal.

5.2 Models without absorbing states, \(m_{ij} \geq 1\). In the case \(m_{ij} = 1\), a key step is the formulation of the inequality \(\inf_{1 \leq i, 1 \leq j} \lambda_{\min} \overline{V}_{ij}(\beta) > 0\); this ensures that both conditions (B) and (D) are satisfied. When \(m_{ij} \geq 1\), however, this inequality may no longer hold due to the problem of dimensionality. In this situation \(\overline{D}_{ij}(\beta)\) becomes an \(m_{ij}n \times k(k-1)\) matrix with \(m_{ij}n\) possibly greater than \(k(k-1)\), hence \(\overline{V}_{ij}(\beta) = \overline{D}_{ij}(\beta) \Sigma_{ij}^{-1}(\beta) \overline{D}_{ij}(\beta)'\) is
not ensured to be positive definite. Therefore, we may need to consider a stronger condition under which the conditions (B) and (D) will be met.

Note that, under the compactness condition (C), the inequalities (5.4) and (5.5) of Lemma 3 do hold though (5.3) may not when \( m_{ij} \geq 1 \). In terms of the divergence condition (D), this leads us to consider conditions under which

\[
\lambda_{\min}\left(\sum_{i,j} Z_{i(j-1)} \tilde{D}_{ij} \tilde{\Sigma}_{ij}^{-1} \tilde{D}^t_{ij} Z_{i(j-1)}^t\right) \to \infty,
\]

or equivalently, in view of the compactness condition (C),

\[
(5.7) \quad \lambda_{\min}\left(\sum_{i,j} Z_{i(j-1)} \tilde{D}_{ij} \tilde{D}^t_{ij} Z_{i(j-1)}^t\right) \to \infty,
\]

where \( \tilde{D}_{ij} = \tilde{D}_{ij}(\beta) \) and \( \tilde{\Sigma}_{ij}^{-1} = \tilde{\Sigma}_{ij}^{-1}(\beta) \). This condition is equivalent to (S1) when \( m_{ij} = 1 \) since \( \text{rank}(\tilde{D}_{ij}) = n \); see proof of Lemma 6 below in subsection 5.4. When \( m_{ij} \geq 1 \), however, a single summand \( Z_{i(j-1)} \tilde{D}_{ij} \tilde{D}^t_{ij} Z_{i(j-1)}^t \) becomes

\[
\left\{ \sum_{l=1}^{m_{ij}} Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \right\} \left\{ \sum_{l=1}^{m_{ij}} Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \right\}^t = \sum_{l=1}^{m_{ij}} Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \tilde{D}^t_{i,jl} Z_{i(j-1)(l-1)}^t + \sum_{l \neq l_{i}} Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \tilde{D}^t_{i,jl} Z_{i(j-1)(l-1)}^t,
\]

where \( \tilde{D}_{i,jl} = \partial P_{i,jl}^* / \partial \gamma_l \) evaluated at \( \gamma = Z_{i(j-1)}^t \beta \), \( \gamma_l \) denotes \( Z_{i(j-1)}^t l / \beta \), and where \( P_{i,j}^* \) is obtained from \( P_{i,j} \) by ignoring its last column. Though a similar argument can show that \( \text{rank}(\tilde{D}_{i,jl}) = n \) (see appendix 6.3), this only infers, for example, with conditions (C) and (S1), that

\[
\lambda_{\min}\left(\sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{l=1}^{m_{ij}} Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \tilde{D}^t_{i,jl} Z_{i(j-1)(l-1)}^t\right) \to \infty.
\]

Unlike \( Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \tilde{D}^t_{i,jl} Z_{i(j-1)(l-1)}^t \), the sum of \( Z_{i(j-1)(l-1)} \tilde{D}_{i,jl} \tilde{D}^t_{i,jl} Z_{i(j-1)(l-1)}^t \) and its transpose is only ensured to be symmetric, which is not necessarily positive semidefinite. In brief, in order that (5.7) holds, we need to assume that there is enough useful information in the \( Z_{ij} \)'s for estimation purposes in the following sense:
(S₃) There exists $a > 0$, $c > 0$ and an integer $ν₁$, such that

$$\#\{Z_{i(j−1)} : \|λ'Z_{i(j−1)}D_{ij}(\beta)\| ≥ a, \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i \} ≥ cv, \text{ uniformly in } λ$$

for all $ν ≥ ν₁$, where $\|λ\| = 1$.

(S₄) There exists $a > 0$, $c > 0$ and an integer $ν₁$, such that

$$\#\{Z_{i(j−1)} : \|λ'Z_{i(j−1)}\widetilde{D}_{ij}(\beta)\| ≥ a, \quad i = 1, \ldots, N, \quad j = 2, \ldots, n_i \} ≥ cv, \text{ uniformly in } λ$$

for all $ν ≥ ν₁$, where $\|λ\| = 1$.

These conditions mean that, for estimation of $\beta$, the amount of useful information in the $Z_{i(j−1)}$’s, in terms of $\|λ'Z_{i(j−1)}\widetilde{D}_{ij}(\beta)\| ≥ a$, grows at least as fast as $ν$ in all directions in the parameter space $B$ of $\beta$. Actually, it’s not difficult to show, for example, that condition (S₃) is equivalent to the assertion that

$$λ_{min}(\sum_{i=1}^{N} \sum_{j=1}^{n_i} Z_{i(j−1)}D_{ij}\widetilde{D}_{ij}'Z_{i(j−1)}') ≥ c₁ν, \quad ν ≥ ν₁,$$

for some $c₁ > 0$, and some integer $ν₁$. Although these are not perfect conditions in terms of easy verification, they can provide guidelines for practical use.

**Corollary 2**

(i) For general $N$, under conditions (C) and (S₄), the conditions (B) and (D) are fulfilled with $A_ν(\beta) = E_ν(\beta)$ or $F_ν(\beta)$.

(ii) For general $N$, under conditions (C), (S₃), and (P), the conditions (B) and (D) are fulfilled with $A_ν(\beta) = F_ν(\beta)$.

Hence, in either case, the MLE asymptotically exists and is consistent and asymptotically normal.

5.3 Models with absorbing states. In this situation there can be no asymptotic theory of the MLE if $N$ is bounded, since an absorbing state will be reached almost surely after a finite number of transitions. The asymptotic theory, however, exists if a large number
of independent pruned processes are available. The arguments used in the preceding two subsections may be adapted to provide us with asymptotic results for models with absorbing states. For example, all the subscripts \( a = 1 \ldots k \) involved in subsection 5.1 should be adjusted to \( a \in \Omega \), and \( \bar{\Sigma}_{ij}(\beta) \) and \( D_{ij}(\beta) \) will be \( \omega(k-1) \times \omega(k-1) \) and \( n \times \omega(k-1) \) matrices, respectively. Minor adjustments on the range of \( n_i \) and \( m_{ij} \) in the sufficient conditions are also necessary for Lemma 3, Corollaries 1 and 2 to be adapted to the following Lemma 4, Corollaries 3 and 4.

\((P')\) \( \inf_{1 \leq i, \Omega} P_{\beta}(y_{i0a} = 1) > 0 \), where \( \Omega \) is over all nonabsorbing states.

**Lemma 4** For models with absorbing states and \( m_{ij} = 1 \), the condition (C) implies that

\[ \inf_{1 \leq i, 1 \leq j} \lambda_{\min} \bar{V}_{ij}(\beta) > 0. \]

If, in addition, the set \( \{n_i\} \) is bounded, then

\[ \inf_{1 \leq i, 1 \leq j} \lambda_{\min} E_{\beta} V_{ij}(\beta) \geq c_1 \inf_{1 \leq i, 2 \leq j} \lambda_{\min} \bar{V}_{ij}(\beta), \]

\[ \inf_{1 \leq i, 2 \leq j} \lambda_{\min} E_{\beta} V_{ij}(\beta) \geq c_2 \inf_{1 \leq i, 2 \leq j} \lambda_{\min} \bar{V}_{ij}(\beta), \]

with some constant \( c_1, c_2 > 0 \). If the condition \((P')\) is further satisfied, then

\[ \inf_{1 \leq i, 1 \leq j} \lambda_{\min} E_{\beta} V_{ij}(\beta) \geq c_3 \inf_{1 \leq i, 1 \leq j} \lambda_{\min} \bar{V}_{ij}(\beta), \]

with some constant \( c_3 > 0 \).

**Corollary 3** Assume that \( m_{ij} = 1 \), \( N \rightarrow \infty \) as \( \nu \rightarrow \infty \), and that the set \( \{n_i\} \) is bounded. The conditions (B) and (D) are fulfilled

(i) with \( A_{\nu}(\beta) = E_\nu(\beta) \) or \( F_\nu(\beta) \), under conditions (C) and (S2), or

(ii) with \( A_{\nu}(\beta) = F_\nu(\beta) \), under conditions (C), (S1), and \((P')\).

Hence, in either case, the MLE asymptotically exists and is consistent and asymptotically normal.
COROLLARY 4 Assume that \( m_{ij} \geq 1, N \to \infty \) as \( \nu \to \infty \), and that the set \( \{ n_i, m_{ij} \} \) is bounded. The conditions (B) and (D) are fulfilled

(i) with \( A_\nu(\beta) = E_\nu(\beta) \) or \( F_\nu(\beta) \), under conditions (C) and (S4), or

(ii) with \( A_\nu(\beta) = F_\nu(\beta) \), under conditions (C), (S3), and (P').

Hence, in either case, the MLE asymptotically exists and is consistent and asymptotically normal.

5.4 Proofs. To establish the positive definiteness of the matrix \( \overline{V}_{ij}(\beta) \) we first introduce some matrix notation which will help us to show the rank of \( \overline{D}_{ij}(\beta) \) is \( n \). Let's extend the notion of a matrix to that of an array in an apparent way so that a matrix is an array of dimension two. That is, \( A = (a_{i_1 \cdots i_m}) \) is a \( d_1 \times \cdots \times d_m \) array of dimension \( m \), if \( A \) is a function from the set of \( m \)-tuples \( (i_1, \cdots, i_m) \), where \( i_l = 1, \cdots, d_l, l = 1, \cdots, m \), to the set of real numbers such that \( A(i_1, \cdots, i_m) = a_{i_1 \cdots i_m} \). For our purpose, We will also generalize the notion of matrix multiplication to that of array multiplication in the following special way.

Let \( A \) be a \( p \times q \) matrix and \( B \) an \( s \times q \times r \) array. Define \( AB \) to be the \( s \times q \times r \) array with \( s q \times r \) subarrays (matrices) \( AB_i, i = 1, \cdots, s \), where \( B_i \) is the \( i \)th subarray (matrix) of \( B \). Similarly, if \( A \) is an \( s \times p \times q \) array and \( B \) a \( q \times r \) matrix, we define \( AB \) to be the \( s \times q \times r \) array with \( s q \times r \) subarrays (matrices) \( A_iB, i = 1, \cdots, s \), where \( A_i \) is the \( i \)th subarray (matrix) of \( A \).

Let \( A \) be a \( p \times q \times r \) array with \( p q \times r \) subarrays (matrices) \( A_1, \cdots, A_p \). Denote \( \overline{A} \) to be the \( p \times qr \) matrix whose \( i \)th row is \( \overline{A}_i \).

LEMMA 5 Let \( A, B, C \) be matrices so that their multiplication \( ABC \) makes sense. Then \( \overline{ABC} = \overline{B}(A' \otimes C) \), where \( \otimes \) denotes the Kronecker product. The same result holds if exactly one of \( A, B, C \) is an array of dimension 3.

PROOF. For a definition of Kronecker product, see, for example, page 599 of Anderson (1984). We will only prove the result for the matrices, the proof for the 3-dimensional arrays
is similar. Note that

$$\text{the } 1, (i,j) \text{ element of } \overline{ABC} = \text{the } (i,j) \text{ element of } ABC$$

$$= (\text{the } i'\text{th row of } A)B(\text{the } j'\text{th column of } C)$$

$$= B(\text{the } (i,j)'\text{th column of } A' \otimes C)$$

$$= \text{the } 1, (i,j) \text{ element of } \overline{B}(A' \otimes C).$$

Hence the results. $\square$

**Lemma 6** Assume that the $k \times k$ intensity matrix $Q$ is a function of the $n \times 1$ vector $\gamma$, and $P(t) = \exp(Qt)$. Then $\text{rank}\left(\partial P(t)/\partial \gamma\right) = \text{rank}\left(\partial Q/\partial \gamma\right)$ for all $t > 0$.

**Proof.** Let $\partial P(t)/\partial \gamma$ be the $n \times k \times k$ array whose $u$th subarray (matrix) is $\partial P(t)/\partial \gamma_u$, $u = 1, \cdots, n$. To compute $\partial P(t)/\partial \gamma$, we extend Kalbfleish and Lawless's (1985) formula to a general intensity matrix where multiple eigenvalues may arise. Let $Q = AJA^{-1}$ be the Jordan canonical decomposition of $Q$, where $J = \text{diag}(J_1, \cdots, J_h)$ and the $J_i$'s are Jordan blocks. Similar to Kalbfleish and Lawless's (1985) derivation it can be shown that

$$\frac{\partial P(t)}{\partial \gamma_u} = AV_1^{(u)}A^{-1},$$

where

$$V_1^{(u)} = \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} J^l G_1^{(u)} J^{s-1-l} \frac{t^s}{s!}, \quad G_1^{(u)} = A^{-1}(\partial Q/\partial \gamma_u)A.$$

Denote $V_1$ and $G_1$ as the $n \times k \times k$ arrays whose $u$th subarrays (matrices) are $V_1^{(u)}$ and $G_1^{(u)}$, respectively, $u = 1, \cdots, n$. Then, from the above lemma,

$$\frac{\partial P(t)}{\partial \gamma} = \frac{\partial P(t)}{\partial \gamma} = AV_1A^{-1} = V_1(A' \otimes A^{-1}).$$

Since $A$ is invertible, so is $A' \otimes A^{-1}$. Therefore, $\text{rank}\left(\partial P(t)/\partial \gamma\right) = \text{rank}\left(V_1\right)$. As shown in appendix 6.2, $\sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J^l)' \otimes J^{s-1-l} \frac{t^s}{s!}$ is convergent and invertible for all $t > 0$, hence

$$V_1 = G_1 \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J^l)' \otimes J^{s-1-l} \frac{t^s}{s!}.$$
and so \( \text{rank}(\mathbf{V}_1) = \text{rank}(\mathbf{G}_1) \). Finally, we define \( \partial Q / \partial \gamma \) in a manner similar to that of \( \partial P(t) / \partial \gamma \). Arguments similar to those used earlier imply that

\[
\mathbf{G}_1 = \mathbf{A}^{-1} \frac{\partial Q}{\partial \gamma} \mathbf{A} = \frac{\partial Q}{\partial \gamma} ((\mathbf{A}^{-1})' \otimes \mathbf{A}) = \frac{\partial Q}{\partial \gamma} ((\mathbf{A}^{-1})' \otimes \mathbf{A}).
\]

Therefore, \( \text{rank}(\mathbf{G}_1) = \text{rank}(\partial Q / \partial \gamma) \), and we have the desired result. \( \square \)

\textbf{Proof of Lemma 3.} As shown in appendix 6.1, \( p_{ijab} > 0 \) for all \( i, j, a, b \), it follows that \( \Sigma_{ija} = \text{var}(y_{ij} | y_{i(j-1)a} = 1) \) is positive definite, \( a = 1, \cdots, k \). Hence, \( \Sigma_{ij} = \text{diag}\{\Sigma_{ija}\}_{a=1}^{k} \) is positive definite, and so is \( \Sigma_{ij}^{-1} \). Note that \( \sum_{b=1}^{k} p_{ijab} = 1 \) for all \( a \) implies that \( \text{rank}(\mathbf{D}_{ij}) = \text{rank}(\partial P(t) / \partial \gamma) \) for any \( t > 0 \). It follows from the preceding lemma and regularity assumption (iii) that \( \text{rank}(\mathbf{D}_{ij}) = n \) for any \( t > 0 \). Together with the compactness condition (C), we arrive at the desired result (5.3).

The compactness condition (C) implies that \( \inf_{i,j,a} P(y_{ija} = 1 | F_{i,j-1}) > 0 \), subject to \( 1 \leq i, 1 \leq j, 1 \leq a \leq k \). Hence, by induction and the Markov property,

\begin{equation}
\inf_{i,j,i_{j-1}} P(y_{ij} = \mathbf{y} | F_{i,j}) > 0,
\end{equation}

subject to \( 1 \leq i, 1 \leq j, j_{1} \leq j - 1 \), and \( \mathbf{y} \) being taken over the state space. In particular,

\begin{equation}
\inf_{i,j,i_{0}} P(y_{ij} = \mathbf{y} | F_{i,0}) > 0 \quad \text{and} \quad \inf_{i,j,\mathbf{y}} P(y_{ij} = \mathbf{y}) > 0
\end{equation}

hold, subject to \( 1 \leq i, 1 \leq j \), and \( \mathbf{y} \) being taken over the state space. Then the first and second inequalities of (5.9) imply (5.4) and (5.5), respectively. Together with condition (P), the second inequality of (5.9) holds, subject to \( 1 \leq i, 0 \leq j \), and \( \mathbf{y} \) being taken over the state space, and therefore (5.6) follows. \( \square \)

\textbf{Proof of Corollary 1.} From (3.2) and the discussion in subsection 4.1, it follows that \( \{s_{\nu}\} \) is a square integrable zero mean martingale with respect to the filtration \( \{F_{\nu}\} \), relative to \( P \) and \( P(\cdot | F_{\nu}^{(0)}) \), where \( F_{0}^{(\nu)} = V_{i=1}^{N} F_{i,0} \). Therefore, \( \{s_{\nu}\} \) has orthogonal increments \( \{\Delta s_{\nu}\} \), and \( G_{\nu} \), defined by \( \sum_{i,j} Z_{i,j-1} V_{ij} Z_{i,j-1}' \), equals \( \sum_{\nu=1}^{\nu} E[\Delta s_{\nu} (\Delta s_{\nu})' | F_{\nu-1}] \). Further integrations imply that

\begin{equation}
F_{\nu} = E G_{\nu} \quad \text{and} \quad E_{\nu} = E (G_{\nu} | F_{0}^{(\nu)}).
\end{equation}
(i) First note that it is always true that

\[ \lambda_{\min}(F_\nu) \geq \inf_{1 \leq i, j \leq 2} \lambda_{\min}(EV_{i,j}) \lambda_{\min} \left( \sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} \right). \]

Second, conditions (C) and (S1) imply \( \lambda_{\min}(\sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)}) \to \infty \) since \( N \) is bounded. Together with (5.3), (5.5), it follows that \( \lambda_{\min}(F_\nu) \to \infty \) as \( \nu \to \infty \), i.e. the divergence condition (D) is satisfied. To show that the boundedness condition (B) is satisfied, consider the inequality

\[ \|F_\nu^{-1/2} \left( \sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} \right) F_\nu^{-T/2} \| \leq \]

\[ \|F_\nu^{-1/2} \left( \sum_{i=1}^{N} X_{i(0)} X'_{i(0)} \right) F_\nu^{-T/2} \| + \|F_\nu^{-1/2} \left( \sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} \right) F_\nu^{-T/2} \|.
\]

Note that the first term on the right hand side is bounded, so it suffices to show that the second term is uniformly bounded in \( \nu \). However,

\[ \|F_\nu^{-1/2} \left( \sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} \right) F_\nu^{-T/2} \| \leq \]

\[ \frac{\|F_\nu^{-1/2} (\sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} (EV_{i,j}) X'_{i(j-1)} ) F_\nu^{-T/2} \|}{\inf_{1 \leq i, j \leq 2} \lambda_{\min}(EV_{i,j})} \leq \frac{1}{\inf_{1 \leq i, j \leq 2} \lambda_{\min}(EV_{i,j})}.
\]

By (5.3) and (5.5), the right hand side is bounded, hence condition (B) is satisfied. This proves the assertion for \( A_\nu = F_\nu \).

With similar arguments, \( \lambda_{\min}(E_\nu) \to \infty \) can be obtained from (5.3),(5.4), and the condition (S1). Similar arguments as above, together with (5.3) and (5.4), also imply that \( W_\nu = \|E_\nu^{-1/2} (\sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} ) E_\nu^{-T/2} \| \) is uniformly bounded in \( \nu \) for any given set of initial responses. Since \( N \) is bounded and \( k \) is finite, this implies that \( EW_\nu \) is uniformly bounded in \( \nu \). This proves the assertion for \( A_\nu = E_\nu \).

(ii) Similar arguments, as used in (i), together with (5.3) and (5.4), imply that

\[ \lambda_{\min}(E_\nu) \geq c \lambda_{\min} \left( \sum_{i=1}^{N} \sum_{j=2}^{n_i} X_{i(j-1)} X'_{i(j-1)} \right), \]

\[ \text{Page 27} \]
for some $c > 0$. It then follows from (S2) that $\lambda_{\text{min}}(E_\nu) \to \infty$, as $\nu \to \infty$, and

$$W_\nu = \|E_\nu^{-1/2}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})E_\nu^{-T/2}\| \leq \frac{\lambda_{\text{max}}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})}{\lambda_{\text{min}}(E_\nu)} \leq \frac{\lambda_{\text{max}}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})}{c\lambda_{\text{min}}(\sum_{i=1}^{N}\sum_{j=2}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})},$$

which is uniformly bounded in $\nu$ for any given initial responses. Hence, $E_\nu W_\nu$ is uniformly bounded in $\nu$. This proves the assertion for $A_\nu = E_\nu$. The assertion for $A_\nu = F_\nu$ can be demonstrated with similar arguments.

(iii) The inequality

$$\lambda_{\text{min}}(F_\nu) \geq \inf_{1 \leq i, j \leq 1} \lambda_{\text{min}}(EV_{i,j}) \lambda_{\text{min}}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)}),$$

together with (5.3), (5.6), and (S1), imply that $\lambda_{\text{min}}(F_\nu) \to \infty$, as $\nu \to \infty$. The boundedness condition (B) can be obtained from (5.3), (5.6), and the following inequality

$$\|F_\nu^{-1/2}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})F_\nu^{-T/2}\| \leq \frac{\|F_\nu^{-1/2}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}(EV_{i,j})Z'_{i(j-1)})F_\nu^{-T/2}\|}{\inf_{1 \leq i, j \leq 1} \lambda_{\text{min}}(EV_{i,j})} \leq \frac{1}{\inf_{1 \leq i, j \leq 1} \lambda_{\text{min}}(EV_{i,j})}. \quad \square$$

**Proof of Corollary 2.** (i) First note that, under the compactness condition (C), the inequalities (5.8) and (5.9) still hold in the case of $m_{ij} \geq 1$. Hence

$$\lambda_{\text{min}}(F_\nu) \geq c_1 \lambda_{\text{min}}(\sum_{i=1}^{N}\sum_{j=2}^{n_i} Z_{i(j-1)}D_{ij}D'_{ij}Z'_{i(j-1)}),$$

for some $c_1 > 0$. Condition (S4), therefore, implies that

$$\lambda_{\text{min}}(F_\nu) \geq c_1 c_2 \nu a^2, \quad \nu \geq \nu_1,$$

for some $c_2 > 0$, $a > 0$, and some integer $\nu_1$. Hence $\lambda_{\text{min}}(F_\nu) \to \infty$, as $\nu \to \infty$. Also

$$\|F_\nu^{-1/2}(\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)})F_\nu^{-T/2}\| \leq \frac{\|\sum_{i=1}^{N}\sum_{j=1}^{n_i} Z_{i(j-1)}Z'_{i(j-1)}\|}{\lambda_{\text{min}}(F_\nu)} \leq \frac{\nu \max_{i,j} \|Z_{i(j-1)}Z'_{i(j-1)}\|}{c_1 c_2 \nu a^2} \leq \frac{\max_{i,j} \|Z_{i(j-1)}Z'_{i(j-1)}\|}{c_1 c_2 a^2},$$

28
which is uniformly bounded in \( \nu \). This completes the proof for the case \( A_\nu = F_\nu \). Similar arguments will yield the result for the case \( A_\nu = E_\nu \).

(ii) If both conditions (C) and (P) are satisfied, the second inequality of (5.9) will hold, subject to \( j \geq 0 \), in the case of \( m_{ij} \geq 1 \). Then

\[
\lambda_{\min}(F_\nu) \geq c_1 \lambda_{\min} \left( \sum_{i=1}^{N} \sum_{j=1}^{n_i} Z_{i(i-1)} \tilde{D}_{ij} \tilde{D}_{ij}' Z_{i(i-1)}' \right),
\]

for some \( c_1 > 0 \). The remaining arguments can proceed as in (i) and are omitted. \( \square \)

PROOF OF LEMMA 4. The proof is parallel to that of Lemma 3 with \( 1 \leq a \leq k \) replaced by \( a \in \Omega \). The condition that the set \( \{n_i\} \) is bounded is needed when establishing (5.8), and hence (5.9), for models with absorbing states by noticing that

\[
P(y_{ij} = y|\mathcal{F}_{i,j_1}) \geq c(\omega c)^{j-j_1}, \quad j_1 \leq j - 1,
\]

where \( c = \inf_{i,j,a} P(y_{ija} = 1|\mathcal{F}_{i,j-1}) > 0 \), subject to \( a \in \Omega \). Therefore,

\[
\inf_{i,j,j_1} P(y_{ij} = y|\mathcal{F}_{i,j_1}) \geq c(\omega c)^{n_0},
\]

where \( n_0 = \max_{1 \leq i \leq \infty} n_i \) is finite by the boundedness assumed on the set \( \{n_i\} \). \( \square \)

PROOF OF COROLLARY 3. The proof is parallel to that of Lemma 3 with \( 1 \leq a \leq k \) replaced by \( a \in \Omega \) and is omitted. \( \square \)

PROOF OF COROLLARY 4. The boundedness condition assumed on \( \{m_{ij}\} \) is required due to the same reasoning for \( \{n_i\} \) as given in the proof of Lemma 4. The rest of the proof is parallel to that of Corollary 2 with \( 1 \leq a \leq k \) replaced by \( a \in \Omega \) and is omitted. \( \square \)

6. Appendix.

6.1 Argument of "\( p_{ijab} > 0 \) for all \( i,j,a,b \), where a is not an absorbing state"

Since we are considering a finite number of categories, no instantaneous state will be allowed. Hence, the formula for transition probabilities, for homogeneous Markov processes in continuous time, given in Lemma 8.4.1 of Cinlar (1975) implies the assertion for the case
\[ m_{ij} = 1. \] The assertion for the general case \( m_{ij} \geq 1 \) (piecewise homogeneous processes) then follows by induction and the Markov property. \( \Box \)

6.2 Convergence and Invertibility of \( \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J^l)' \otimes J^{s-1-l} t^s s! \), \( t > 0 \)

Assume that \( J = \text{diag}(J_1, \cdots, J_h) \), and \( J_i \) is an \( m_i \times m_i \) Jordan block of the form \( J_i = d_i I_i + M_i \), \( i = 1, \ldots, h \), where \( I_i = (\delta_{ab}) \) and \( M_i = (\delta_{a,b-1}) \), \( a, b = 1, \ldots, m_i \), and \( m_1 + \cdots + m_h = k \). Clearly \( (J_i)' \otimes J^{s-1-l} = \text{diag}((J_i^l)' \otimes J^{s-1-l}, \ldots, (J_h^l)' \otimes J^{s-1-l}) \) with each \( (J_i^l)' \otimes J^{s-1-l} \) being a lower block triangular matrix since \( J_i \) is an upper triangular matrix. Note that \( M_i^{m_i} = 0 \), and

\[
J_i^l = d_i^l I_i + \binom{l}{i} d_i^{l-i} M_i + \cdots + \binom{l}{m_i-1} d_i^{l-m_i+1} M_i^{m_i-1}.
\]

Hence \( \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J_i^l)' \otimes J^{s-1-l} t^s s! \) is a lower block triangular matrix with nonzero blocks of the form

\[
\sum_{s=n+1}^{\infty} \sum_{l=n}^{s-1} \binom{l}{n} d_i^{l-n} J_j^{s-1-l} t^s s! \quad i, j = 1, \ldots, h, \quad n = 0, \ldots, m_i - 1,
\]

which equals

\[
\sum_{u=1}^{m_j} \sum_{s=n+u}^{\infty} \sum_{l=u}^{s-u} \binom{l}{u} d_i^{l-n} d_j^{s-u-l} t^s s! \quad M_j^{u-1}.
\]

Without loss of generality, we assume that \( d_i, d_j > 1 \). Then for any \( u = 1, \ldots, m_j \),

\[
\sum_{s=n+u}^{\infty} \sum_{l=u}^{s-u} \binom{l}{u} (s-1-l) u! d_i^{l-n} d_j^{s-u-l} t^s s! \leq \sum_{s=n+u}^{\infty} \binom{s-1}{n+u} (d_i d_j)^{s-u-n} t^s s!.
\]

Applying the ratio test to the right side of the above inequality we have shown that every nonzero block of \( \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J_i^l)' \otimes J^{s-1-l} t^s s! \) is convergent for all \( t \), and so is itself.

To show \( \sum_{s=1}^{\infty} \sum_{l=0}^{s-1} (J_i^l)' \otimes J^{s-1-l} t^s s! \) is invertible for all \( t > 0 \), it suffices to show that the blocks on the main block diagonal are all invertible for all \( t > 0 \). These blocks take the following form (replacing \( n \) by \( 0 \) in (6.1))

\[
\sum_{s=1}^{\infty} \sum_{l=0}^{s-1} d_i^{l} J_j^{s-1-l} t^s s!, \quad i, j = 1, \ldots, h,
\]

and are all upper triangular matrices. Thus to show invertibility it suffices to show that the diagonal elements of these blocks are not zero. These diagonal elements can be obtained by
replacing \( n \) by 0 in the summand of (6.2) corresponding to \( u = 1 \) and they have the form

\[
\sum_{s=1}^{\infty} \sum_{t=0}^{s-1} \frac{d_i^l d_j^{s-1-l} t^s}{s!},
\]

which is \( \frac{e^{d_i t} - e^{d_j t}}{d_i - d_j} \) for \( d_i \neq d_j \) and \( te^{d_i t} \) for \( d_i = d_j \). This completes the proof. \( \square \)

6.3 Derivation of \( \text{rank}(\tilde{D}_{ij}) = n \)

Let \( P_{ij} = \exp[Q_{(j-1)(t-1)}(t_{(j-1)(t) - t_{(j-1)(t-1)})}] \), then \( P_{ij} = \prod_{l=1}^{n_{ij}} P_{ijl} \). From Lemma 5,

\[
\frac{\partial P_{ij}}{\partial \gamma_l} = \frac{\partial P_{ijl}}{\partial \gamma_l} \left[ (\prod_{l_1 < l} P_{ijl})^l \otimes \prod_{l_1 > l} P_{ijl} \right].
\]

Since \( (\prod_{l_1 < l} P_{ijl})^l \otimes \prod_{l_1 > l} P_{ijl} \) is invertible, this implies by Lemma 6 that

\[
\text{rank}(\tilde{D}_{ij}) = \text{rank}(\frac{\partial P_{ij}}{\partial \gamma_l}) = \text{rank}(\frac{\partial P_{ijl}}{\partial \gamma_l}) = n. \quad \square
\]
REFERENCES


