Asymptotic Theory for Linear-Chain Conditional Random Fields - Supplementary Material -

PROOF OF THEOREM 1

The existence of the asymptotic ratios r_{ij} is well-known (Lemma 3.4, Seneta, 2006). Let us establish the geometric rate. For any $\ell \times \ell$-matrix $\boldsymbol{A} = (a_{ij})$, define

$$\phi(\boldsymbol{A}) \quad = \quad \min_{i,j,k,l} \frac{a_{ik}a_{jl}}{a_{jk}a_{il}}.$$

Note that $\phi(\boldsymbol{A}) \leq 1$. Using the concept of Birkhoff's contraction coefficient, one can show that

$$\frac{1-\sqrt{\phi(\boldsymbol{M}_n)}}{1+\sqrt{\phi(\boldsymbol{M}_n)}} \quad \leq \quad \prod_{t=1}^{n} \frac{1-\sqrt{\phi(\boldsymbol{M}(x_t))}}{1+\sqrt{\phi(\boldsymbol{M}(x_t))}}$$

(Chapter 3, Seneta, 2006). With ψ^2 defined in Lemma 1 and using the fact that $\sqrt{\phi(\boldsymbol{M}_n)} \leq 1$, we obtain

$$\frac{1-\sqrt{\phi(\boldsymbol{M}_n)}}{2} \quad \leq \quad \left(\frac{1-\psi}{1+\psi}\right)^n.$$

After a few elementary algebraic manipulations and applying Bernoulli's inequality, we obtain

$$\phi(\boldsymbol{M}_n) \quad \geq \quad 1 - 4\left(\frac{1-\psi}{1+\psi}\right)^n.$$

Now, note that the quantities

$$\max_{k \in \mathcal{Y}} \left(\frac{m_n(i,k)}{m_n(j,k)}\right) \quad \text{and} \quad \min_{k \in \mathcal{Y}} \left(\frac{m_n(i,k)}{m_n(j,k)}\right)$$

are non-increasing and non-decreasing with n, respectively (Lemma 3.1, Seneta, 2006). Moreover, by the definition of $\phi(\cdot)$, the ratio of the minimum to the maximum is greater than $\phi(\boldsymbol{M}_n)$. □

PROOF OF THEOREM 2

We show that c and κ satisfy

$$\begin{aligned} &\big|\, P_{\boldsymbol{\lambda}}^{(-n,n)}(Y_t = y_t, \ldots, Y_{t+k} = y_{t+k} \,|\, \boldsymbol{X} = \boldsymbol{x}) \\ &\quad - P_{\boldsymbol{\lambda}}(Y_t = y_t, \ldots, Y_{t+k} = y_{t+k} \,|\, \boldsymbol{X} = \boldsymbol{x}) \,\big| \;\leq\; c\kappa^n \end{aligned}$$

for all $n \in \mathbb{N}$ such that $-n \leq t$ and $n \geq t+k$. Introduce the vectors $\underline{\boldsymbol{r}}_i(n)$ and $\overline{\boldsymbol{r}}_i(n)$ with the kth components given by

$$\begin{aligned} \underline{r}_{ki}(n) &= \min_{l \in \mathcal{Y}} \left(\frac{g_n(k,l)}{g_n(i,l)}\right), \\ \overline{r}_{ki}(n) &= \max_{l \in \mathcal{Y}} \left(\frac{g_n(k,l)}{g_n(i,l)}\right). \end{aligned}$$

In the same way, we define vectors $\underline{\boldsymbol{s}}_j(n)$ and $\overline{\boldsymbol{s}}_j(n)$ with respect to $\boldsymbol{H}_n$. It is easy to see that

$$\begin{aligned} \underline{\boldsymbol{r}}_i(n)^T \boldsymbol{F} \underline{\boldsymbol{s}}_j(n) &\leq \frac{\boldsymbol{\alpha}_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x})^T \boldsymbol{F} \boldsymbol{\beta}_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x})}{\alpha_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x},i)\,\beta_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x},j)} \\ &\leq \overline{\boldsymbol{r}}_i(n)^T \boldsymbol{F} \overline{\boldsymbol{s}}_j(n). \end{aligned}$$

Furthermore, according to Theorem 1,

$$\underline{\boldsymbol{r}}_i(n)^T \boldsymbol{F} \underline{\boldsymbol{s}}_j(n) \;\leq\; \boldsymbol{r}_i^T \boldsymbol{F} \boldsymbol{s}_j \;\leq\; \overline{\boldsymbol{r}}_i(n)^T \boldsymbol{F} \overline{\boldsymbol{s}}_j(n).$$

Hence,

$$\begin{aligned} &\left| \frac{\boldsymbol{\alpha}_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x})^T \boldsymbol{F} \boldsymbol{\beta}_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x})}{\alpha_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x},i)\,\beta_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x},j)} - \boldsymbol{r}_i^T \boldsymbol{F} \boldsymbol{s}_j \right| \\ &\qquad \leq \left| \left(\overline{\boldsymbol{r}}_i(n) - \underline{\boldsymbol{r}}_i(n)\right)^T \boldsymbol{F} \left(\overline{\boldsymbol{s}}_j(n) - \underline{\boldsymbol{s}}_j(n)\right) \right|. \end{aligned}$$

According to Theorem 1, we obtain

$$\begin{aligned} &\left| \frac{\boldsymbol{\alpha}_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x})^T \boldsymbol{F} \boldsymbol{\beta}_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x})}{\alpha_{-n}^t(\boldsymbol{\lambda},\boldsymbol{x},i)\,\beta_{t+k}^n(\boldsymbol{\lambda},\boldsymbol{x},j)} - \boldsymbol{r}_i^T \boldsymbol{F} \boldsymbol{s}_j \right| \\ &\qquad \leq 16 \,\|\boldsymbol{F}\| \left(\frac{m_{\sup}}{m_{\inf}}\right)^2 \left(\frac{(1-\varphi)(1-\psi)}{(1+\varphi)(1+\psi)}\right)^n \end{aligned}$$

where $\|\boldsymbol{F}\|$ stands for the sum of all components of $\boldsymbol{F}$. Putting all together, we have

$$\begin{aligned} &\big|\, P_{\boldsymbol{\lambda}}^{(-n,n)}(Y_t = y_t, \ldots, Y_{t+k} = y_{t+k} \,|\, \boldsymbol{X} = \boldsymbol{x}) \\ &\qquad - P_{\boldsymbol{\lambda}}(Y_t = y_t, \ldots, Y_{t+k} = y_{t+k} \,|\, \boldsymbol{X} = \boldsymbol{x}) \,\big| \\ &\qquad \leq 16 \,\|\boldsymbol{F}\| \left(\frac{m_{\sup}}{m_{\inf}}\right)^2 \left(\frac{(1-\varphi)(1-\psi)}{(1+\varphi)(1+\psi)}\right)^n \\ &\qquad\quad \times \prod_{i=1}^{k} m_{\boldsymbol{\lambda}}(x_{t+i}, y_{t+i-1}, y_{t+i}), \end{aligned}$$

and now the value for the constant is c obtained by noting that

$$\|\boldsymbol{F}\| \prod_{i=1}^{k} m_{\boldsymbol{\lambda}}(x_{t+i}, y_{t+i-1}, y_{t+i}) \quad \leq \quad \ell^{k+1} m_{\sup}^{2k}.$$

The proof is complete. □

PROOF OF LEMMA 2

Let $\vec{A} = \bigtimes_{t\in\mathbb{Z}} A_t$. Note that $\vec{\tau}^{-1}\vec{A} = \bigtimes_{t\in\mathbb{Z}} A_{t-1}$, and hence $\pi(\tau^{-1}A) = \pi(A)$ for all $A \in \boldsymbol{\mathcal{A}}$ implies

$$\begin{aligned}\vec{\pi}(\vec{\tau}^{-1}\vec{A}) &= \pi\Big(\bigcap_{t\in\mathbb{Z}} \tau^{-t}A_{t-1}\Big)\\ &= \pi\Big(\tau^{-1}\bigcap_{t\in\mathbb{Z}} \tau^{-(t-1)}A_{t-1}\Big)\\ &= \pi\Big(\bigcap_{t\in\mathbb{Z}} \tau^{-(t-1)}A_{t-1}\Big)\\ &= \vec{\pi}(\vec{A}).\end{aligned}$$

Now suppose $\vec{\tau}^{-1}\vec{A} = \vec{A}$. A necessary condition for this is $A_t = A$ for all $t \in \mathbb{Z}$. Setting $\tilde{A} = \bigcap_{t\in\mathbb{Z}} \tau^{-t}(A)$, we obtain $\vec{\pi}(\vec{A}) = \pi(\tilde{A})$. Now note that $\tau^{-1}\tilde{A} = \tilde{A}$. Thus, if π is τ-ergodic, we have $\pi(\tilde{A}) = 0$ or $\pi(\tilde{A}) = 1$, and hence $\vec{\pi}(\vec{A}) = 0$ or $\vec{\pi}(\vec{A}) = 1$. □

PROOF OF LEMMA 3

The proof that the invariant measure $\mu_{\boldsymbol{\lambda}}$ is unique requires an alternative representation of Markov processes. Write $Q(\boldsymbol{\lambda}, \boldsymbol{x}_1 \ldots \boldsymbol{x}_n, i, j)$ to denote the (i,j)-th component of the product $\boldsymbol{Q}(\boldsymbol{\lambda}, \boldsymbol{x}_1) \ldots \boldsymbol{Q}(\boldsymbol{\lambda}, \boldsymbol{x}_n)$. For $k > 1$ consider the kth iterate of $Q_{\boldsymbol{\lambda}}$:

$$Q^k_{\boldsymbol{\lambda}}(z, C) = \int_{\mathcal{Z}} Q_{\boldsymbol{\lambda}}(z', C)\, Q^{k-1}_{\boldsymbol{\lambda}}(z, dz').$$

Note that

$$\begin{aligned}&Q^k_{\boldsymbol{\lambda}}((\vec{\boldsymbol{x}}, y_0', y_1'),\, \vec{A} \times \{y_0\} \times \{y_1\})\\ &= \begin{cases} Q(\boldsymbol{\lambda}, \boldsymbol{x}_0 \ldots \boldsymbol{x}_{k-2}, y_1', y_0)\, Q(\boldsymbol{\lambda}, \boldsymbol{x}_{k-1}, y_0, y_1) & \text{if } \vec{\tau}^k\vec{\boldsymbol{x}} \in \vec{A},\\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

Now let $L_1 = L_1(\mu_{\boldsymbol{\lambda}})$ denote the space of measurable functions $u : \mathcal{Z} \to \mathbb{R}$ satisfying $\int_{\mathcal{Z}} |u(z)|\, \mu_{\boldsymbol{\lambda}}(dz) < \infty$. For $k \in \mathbb{N}$ let $Q^k_{\boldsymbol{\lambda}}$ be the operator on L_1 defined by

$$Q^k_{\boldsymbol{\lambda}} u(z) = \int_{\mathcal{Z}} u(z') Q^k_{\boldsymbol{\lambda}}(z, dz').$$

Note that, if $k > 1$,

$$\begin{aligned}Q^k_{\boldsymbol{\lambda}} u(\vec{\boldsymbol{x}}, y_0', y_1') &= \sum_{y_0, y_1 \in \mathcal{Y}} u(\vec{\tau}^k\vec{\boldsymbol{x}}, y_0, y_1)\\ &\quad\times Q(\boldsymbol{\lambda}, \boldsymbol{x}_0 \ldots \boldsymbol{x}_{k-2}, y_1', y_0)\, Q(\boldsymbol{\lambda}, \boldsymbol{x}_{k-1}, y_0, y_1).\end{aligned}$$

For the proof that the invariant measure $\mu_{\boldsymbol{\lambda}}$ is unique, let $u_0 \in L_1$ with $u_0 > 0$ and consider the *conservative set* $C^* \subset \mathcal{Z}$ given by

$$C^* = \Big\{ z \in \mathcal{Z} :\ \lim_{n\to\infty} \sum_{k=1}^n Q^k_{\boldsymbol{\lambda}} u_0(z) = \infty \Big\}.$$

Note that the set C^* is independent of the choice of u_0. Furthermore, let $\mathcal{C}_i$ denote the class of *invariant sets*,

$$\mathcal{C}_i = \big\{ C \in \mathcal{C} :\ Q_{\boldsymbol{\lambda}} \mathbf{1}_C = \mathbf{1}_C\ \mu_{\boldsymbol{\lambda}}\text{-almost everywhere} \big\}.$$

We say that $\mathcal{C}_i$ is *trivial* if $\mu_{\boldsymbol{\lambda}}(C) = 0$ or $\mu_{\boldsymbol{\lambda}}(C) = 1$ for every $C \in \mathcal{C}_i$. A sufficient condition for the existence of at most one invariant probability measure on $(\mathcal{Z}, \mathcal{C})$ is that $C^* = \mathcal{Z}$ (up to a $\mu_{\boldsymbol{\lambda}}$-null set) and $\mathcal{C}_i$ is trivial (Theorem VI.A, Foguel, 1969). We first show that $C^* = \mathcal{Z}$. According to Corollary 1 (ii), we have

$$\begin{aligned}\inf\big\{ Q(\boldsymbol{\lambda}, \boldsymbol{x}_1 \ldots \boldsymbol{x}_n, i, j) :\ n \in \mathbb{N},\, i, j \in \mathcal{Y}\big\}\\ \geq \frac{1}{\ell}\left(\frac{m_{\inf}}{m_{\sup}}\right)^2\end{aligned}$$

for every $\vec{\boldsymbol{x}} = (\boldsymbol{x}_t)_{t\in\mathbb{Z}}$. Hence, for $k > 1$,

$$Q^k_{\boldsymbol{\lambda}} u_0(\vec{\boldsymbol{x}}, y_0', y_1') \geq \frac{1}{\ell^2}\left(\frac{m_{\inf}}{m_{\sup}}\right)^4 \sum_{y_0, y_1 \in \mathcal{Y}} u_0(\vec{\tau}^k\vec{\boldsymbol{x}}, y_0, y_1).$$

Furthermore, since $\vec{P}_{\boldsymbol{X}}$ is $\vec{\tau}$-ergodic on $(\vec{\boldsymbol{\mathcal{X}}}, \vec{\mathcal{A}})$,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n u_0(\vec{\tau}^k\vec{\boldsymbol{x}}, y_0, y_1)\\ = \int_{\vec{\boldsymbol{\mathcal{X}}}} u_0(\vec{\boldsymbol{x}}', y_0, y_1)\, \vec{P}_{\boldsymbol{X}}(d\vec{\boldsymbol{x}}')\end{aligned}$$

for $\vec{P}_{\boldsymbol{X}}$-almost every $\vec{\boldsymbol{x}} \in \vec{\boldsymbol{\mathcal{X}}}$. Now, under the assumption $u_0 > 0$, the integral on the right hand side is strictly greater than 0, hence the unnormalized series on the left hand side would tend to ∞. This argument shows that the series in the definition of C^* diverges for $\mu_{\boldsymbol{\lambda}}$-almost every $z \in \mathcal{Z}$, and hence $C^* = \mathcal{Z}$ up to a $\mu_{\boldsymbol{\lambda}}$-null set.

To show that $\mathcal{C}_i$ is trivial, let $C \in \mathcal{C}_i$ be such that $\mu_{\boldsymbol{\lambda}}(C) > 0$ and $Q_{\boldsymbol{\lambda}}\mathbf{1}_C(z) = \mathbf{1}_C(z)$ for $\mu_{\boldsymbol{\lambda}}$-almost every $z \in \mathcal{Z}$. Note that $Q_{\boldsymbol{\lambda}}\mathbf{1}_C(z) = Q_{\boldsymbol{\lambda}}(z, C)$. If (A1) holds, then all entries of the transition matrix $\boldsymbol{Q}$ are strictly greater than 0, and hence a necessary condition for $Q_{\boldsymbol{\lambda}}(z, C) = 1$ is that $C = \vec{A} \times \mathcal{Y} \times \mathcal{Y}$ for some set $\vec{A} \in \vec{\mathcal{A}}$, which implies that $Q_{\boldsymbol{\lambda}}(z, C) = \mathbf{1}_{\vec{A}}(\vec{\tau}\vec{\boldsymbol{x}})$ and $\mathbf{1}_C(z) = \mathbf{1}_{\vec{A}}(\vec{\boldsymbol{x}})$ for $\mu_{\boldsymbol{\lambda}}$-almost every $z = (\vec{\boldsymbol{x}}, y_0, y_1) \in \mathcal{Z}$. Now note that $\mathbf{1}_{\vec{A}}(\vec{\tau}\vec{\boldsymbol{x}}) = \mathbf{1}_{\vec{A}}(\vec{\boldsymbol{x}})$ is equivalent to $\vec{A} = \vec{\tau}^{-1}\vec{A}$, and if (A2) holds, then $\vec{P}_{\boldsymbol{X}}(\vec{A}) = 0$ or $\vec{P}_{\boldsymbol{X}}(\vec{A}) = 1$ for each set $\vec{A}$ satisfying this condition. □

PROOF OF LEMMA 5

We wish to establish that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n E^{(0:n)}_{\boldsymbol{\lambda}}\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big]\\ \sim \frac{1}{n}\sum_{t=1}^n E_{\boldsymbol{\lambda}}\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big].\end{aligned}$$

Let $i, j \in \mathcal{Y}$. Similar to the proof of Theorem 2, we obtain that $P_{\boldsymbol{\lambda}}^{(0:n)}(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x})$ converges to some limit $P_{\boldsymbol{\lambda}}^{(0:\infty)}(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x})$ as n tends to infinity, and there exist constants $c > 0$ and $0 < \kappa < 1$ not depending on $\boldsymbol{x}$ such that

$$\begin{aligned}\big|P_{\boldsymbol{\lambda}}^{(0:n)}&(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x}) - \\ &P_{\boldsymbol{\lambda}}^{(0:\infty)}(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x})\big| \;\le\; c\kappa^{n-t}.\end{aligned}$$

Consequently,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \big|P_{\boldsymbol{\lambda}}^{(0:n)}&(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x}) - \\ &P_{\boldsymbol{\lambda}}^{(0:\infty)}(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x})\big| \;=\; 0\end{aligned}$$

which shows that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} E_{\boldsymbol{\lambda}}^{(0:n)}&\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big] \\ &\sim\; \frac{1}{n}\sum_{t=1}^{n} E_{\boldsymbol{\lambda}}^{(0:\infty)}\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big],\end{aligned}$$

where $E_{\boldsymbol{\lambda}}^{(0:\infty)}$ stands for the conditional expectation with respect to $P_{\boldsymbol{\lambda}}^{(0:\infty)}$. Now, noting that

$$E_{\boldsymbol{\lambda}}^{(0:\infty)}\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big] \sim E_{\boldsymbol{\lambda}}\big[\boldsymbol{f}(X_t, Y_{t-1}, Y_t) \,|\, \boldsymbol{X}\big],$$

we obtain the statement. □

PROOF OF LEMMA 7

Let $\boldsymbol{x} = (x_t)_{t\in\mathbb{Z}}$ be fixed. Using Corollary 1 (ii) and arguments similar to the proof of Theorem 2, it is not difficult to show that the difference between the probabilities $P_{\boldsymbol{\lambda}}(Y_{t-1} = i, Y_t = j, Y_{t+k-1} = l, Y_{t+k} = m \,|\, \boldsymbol{X} = \boldsymbol{x})$ and $P_{\boldsymbol{\lambda}}(Y_{t-1} = i, Y_t = j \,|\, \boldsymbol{X} = \boldsymbol{x}) \times P_{\boldsymbol{\lambda}}(Y_{t+k-1} = l, Y_{t+k} = m \,|\, \boldsymbol{X} = \boldsymbol{x})$ decays at a geometric rate. Since $\boldsymbol{f}$ is bounded, it follows that the covariance of $\boldsymbol{f}(X_t, Y_{t-1}, Y_t)$ and $\boldsymbol{f}(X_{t+k}, Y_{t+k-1}, Y_{t+k})$ conditional on $\boldsymbol{X} = \boldsymbol{x}$ decays component-wise at a geometric rate, and integrating with respect to $P_{\boldsymbol{X}}$ shows that $\gamma_{\boldsymbol{\lambda}}(k)$ decays to 0 at a geometric rate. Consequently, $\sum_{k=1}^{n} \gamma_{\boldsymbol{\lambda}}(k) < \infty$. Similar to the proof of Lemma 5, we obtain that

$$\lim_{n\to\infty} \hat{\gamma}_{\boldsymbol{\lambda}}^{(n)}(k) \quad=\quad \gamma_{\boldsymbol{\lambda}}(k)$$

and

$$\nabla^2 \mathcal{L}_n(\boldsymbol{\lambda}) \quad\sim\quad -\left(\gamma_{\boldsymbol{\lambda}}(0) \,+\, 2\sum_{k=1}^{n} \gamma_{\boldsymbol{\lambda}}(k)\right)$$

$P_{\boldsymbol{\lambda}_0}$-almost surely. The proof is complete. □