Consistently Estimating Markov Chains with Noisy Aggregate Data

Garrett Bernstein and Daniel Sheldon
University of Massachusetts Amherst

A Extra Proofs

A.1 Additional Details for Proof of Proposition 2

We wish to show that $\Lambda_t = N\big(\operatorname{diag}(\boldsymbol{\mu}_t) + (N-1)\boldsymbol{\mu}_t\boldsymbol{\mu}_t^T\big)$ is invertible. The Sherman-Morrison formula (Sherman and Morrison, 1950) gives the inverse of a matrix that is equal to an invertible matrix ($\operatorname{diag}(\boldsymbol{\mu}_t)$) plus a rank-one matrix (the rank-one outer product of $\boldsymbol{\mu}_t$).

Specifically, let $D = \operatorname{diag}(\boldsymbol{\mu}_t)$ and then we have

$$\begin{aligned}
\Lambda_t^{-1} &= N^{-1}\big(D + (N-1)\boldsymbol{\mu}_t\boldsymbol{\mu}_t^T\big)^{-1} \\
&= N^{-1}\Big(D^{-1} - \frac{(N-1)D^{-1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^T D^{-1}}{1 + (N-1)\boldsymbol{\mu}^T D^{-1}\boldsymbol{\mu}_t}\Big) \\
&= N^{-1}\Big(D^{-1} - \frac{N-1}{N}\mathbf{1}\mathbf{1}^T\Big).
\end{aligned}$$

We have used the fact that $D^{-1}\boldsymbol{\mu}_t = \mathbf{1}$ (the all ones vector) and that $\mathbf{1}^T\boldsymbol{\mu}_t = 1$, since $\boldsymbol{\mu}_t$ is a vector of marginal probabilities.

A.2 Proof of Proposition 3

Proof. Using condition (ii) of the proposition, we write:

$$\mathbb{E}[\mathbf{y}_t] = \mathbb{E}\big[\mathbb{E}[\mathbf{y}_t|\mathbf{n}_t]\big] = \mathbb{E}[A_t\mathbf{n}_t] = A_t\mathbb{E}[\mathbf{n}_t].$$

Since A_t is invertible, we have $\mathbb{E}[\mathbf{n}_t] = A_t^{-1}\mathbb{E}[\mathbf{y}_t]$, which proves conclusion (i).

For the non-central second moments, we have for $s \neq t$:

$$\begin{aligned}
\mathbb{E}\left[\mathbf{y}_s\mathbf{y}_t^T\right] &= \mathbb{E}\big[\mathbb{E}\left[\mathbf{y}_s\mathbf{y}_t^T|\mathbf{n}_s,\mathbf{n}_t\right]\big] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbf{y}_s|\mathbf{n}_s]\cdot\mathbb{E}\left[\mathbf{y}_t^T|\mathbf{n}_t\right]\right] \\
&= \mathbb{E}\left[(A_s\mathbf{n}_s)(A_t\mathbf{n}_t)^T\right] \\
&= A_s\mathbb{E}\left[\mathbf{n}_s\mathbf{n}_t^T\right]A_t^T
\end{aligned} \tag{8}$$

The second line uses condition (i) of the proposition: $\mathbf{y}_s$ and $\mathbf{y}_t$ are conditionally independent given $\mathbf{n}_s$ and $\mathbf{n}_t$ if $s \neq t$. Therefore, we have $\mathbb{E}\left[\mathbf{n}_s\mathbf{n}_t^T\right] = A_s^{-1}\mathbb{E}\left[\mathbf{y}_s\mathbf{y}_t^T\right]A_t^{-T}$, which proves conclusion (iii).

For conclusion (ii), we have for $s \neq t$:

$$\begin{aligned}
\operatorname{Cov}(\mathbf{y}_s,\mathbf{y}_t) &= \mathbb{E}\left[\mathbf{y}_s\mathbf{y}_t^T\right] - \mathbb{E}[\mathbf{y}_s]\,\mathbb{E}[\mathbf{y}_t]^T \\
&= A_s\mathbb{E}\left[\mathbf{n}_s\mathbf{n}_t^T\right]A_t^T - A_s\mathbb{E}[\mathbf{n}_s]\,\mathbb{E}[\mathbf{n}_t]^T A_t^T \\
&= A_s\big(\mathbb{E}\left[\mathbf{n}_s\mathbf{n}_t^T\right] - \mathbb{E}[\mathbf{n}_s]\,\mathbb{E}[\mathbf{n}_t]^T\big)A_t^T \\
&= A_s\operatorname{Cov}(\mathbf{n}_s,\mathbf{n}_t)A_t^T,
\end{aligned}$$

so that $\operatorname{Cov}(\mathbf{n}_s,\mathbf{n}_t) = A_s^{-1}\operatorname{Cov}(\mathbf{y}_s,\mathbf{y}_t)A_t^{-T}$, which completes the proof. □

A.3 Additional Details for Proof of Theorem 1

We wish to show that $\lim_{k\to\infty}\gamma(k) = 0$, where $\gamma(k) = \operatorname{Cov}\big(n_t(i)n_{t+1}(j), n_{t+k}(i)n_{t+k+1}(j)\big)$. We have

$$\begin{aligned}\gamma(k) &= \mathrm{Cov}(n_t(i)n_{t+1}(j), n_{t+k}(i)n_{t+k+1}(j)) \\ &= \mathrm{Cov}\Big(\sum_{a=1}^{N}\sum_{b=1}^{N}[x_t^{(a)} = i][x_{t+1}^{(b)} = j],\ \sum_{c=1}^{N}\sum_{d=1}^{N}[x_{t+k}^{(c)} = i][x_{t+k+1}^{(d)} = j]\Big) \\ &= \sum_{a,b,c,d=1}^{N} \mathrm{Cov}\Big([x_t^{(a)} = i][x_{t+1}^{(b)} = j],\ [x_{t+k}^{(c)} = i][x_{t+k+1}^{(d)} = j]\Big)\end{aligned}$$

It is enough to show that the covariance in the summand goes to zero for any choice of four individuals $a, b, c, d \in \{1, \ldots, N\}$. Clearly, it is equal to zero when the individuals $\{a, b\}$ do not overlap with $\{c, d\}$, because individuals are independent. We will verify that the covariance goes to zero for the choice $a = b = c = d := m$, which, since it involves only a single individual, is the case with the *greatest* dependence between times t and $t + k$. Verifying the statement for other combinations of a, b, c, d is similar. Because we are considering a single individual m, we now drop the superscript and write $x_t := x_t^{(m)}$. We can rewrite the covariance as:

$$\begin{aligned}&\mathrm{Cov}\Big([x_t = i][x_{t+1} = j],\ [x_{t+k} = i][x_{t+k+1} = j]\Big) \\ &= \mathbb{E}\Big[[x_t = i][x_{t+1} = j][x_{t+k} = i][x_{t+k+1} = j]\Big] - \mathbb{E}\Big[[x_t = i][x_{t+1} = j]\Big]\mathbb{E}\Big[[x_{t+k} = i][x_{t+k+1} = j]\Big] \\ &= \Pr(x_t = i, x_{t+1} = j, x_{t+k} = i, x_{t+k+1} = j) - \Pr(x_t = i, x_{t+1} = j)\Pr(x_{t+k} = i, x_{t+k+1} = j) \\ &= \mu(i,j) \cdot (P^{k-1})_{ji} \cdot P(i,j) \quad - \quad \mu(i,j)^2 . \qquad (9)\end{aligned}$$

In the last line, we apply several facts about the Markov chain. Here, $\mu(i, j) = \Pr(x_t = i, x_{t+1} = j)$ is the (time-independent) pairwise marginal, $(P^{k-1})_{ji} = \Pr(x_{t+k} = i \mid x_{t+1} = j)$ and $P(i, j) = \Pr(x_{t+k+1} = j \mid x_{t+k} = i)$. Since the Markov chain is ergodic, $\lim_{k\to\infty}(P^{k-1})_{ji} = \pi(i)$, so the first term of Equation (9) becomes:

$$\lim_{k\to\infty} \mu(i,j)(P^{k-1})_{ji}P(i,j) = \mu(i,j)\pi(i)P(i,j) = \mu(i,j)^2.$$

Putting it all together, we see that the limit as k goes to infinity of the covariance in Equation (9) is $\mu(i, j)^2 - \mu(i, j)^2 = 0$, as desired. This completes the proof.

A.4 Additional Details for Proof of Theorem 2

We wish to show that $|\gamma(k)|$ decays exponentially to zero as $k \to \infty$, where $|\gamma(k)| = |\mathrm{Cov}\big(n_t(i)n_{t+1}(j), n_{t+k}(i)n_{t+k+1}(j)\big)|$. We follow the exact same steps in Section A.3 up through Equation (9) where we instead desire $|(P^{k-1})_{ji} - \pi(i)| \le C\alpha^k$ for some constants $\alpha \in (0, 1)$ and $C > 0$. This is proved for irreducible and aperiodic P as Theorem 4.9 in (Levin et al., 2009). Using this fact together with Equation (9), we have:

$$\begin{aligned}\Big|\mathrm{Cov}\Big([x_t = i][x_{t+1} = j],\ [x_{t+k} = i][x_{t+k+1} = j]\Big)\Big| &= \big|\mu(i,j) \cdot (P^{k-1})_{ji} \cdot P(i,j) \quad - \quad \mu(i,j)^2\big| \\ &= \big|\mu(i,j) \cdot (P^{k-1})_{ji} \cdot P(i,j) \quad - \quad \mu(i,j) \cdot \pi(i) \cdot P(i,j)\big| \\ &= \mu(i,j) \cdot \big|(P^{k-1})_{ji} - \pi(i)\big| \cdot P(i,j) \\ &\le \mu(i,j) \cdot C\alpha^k \cdot P(i,j) \\ &= C'\alpha^k.\end{aligned}$$

for $C' = \mu(i, j)P(i, j)C$. Thus the result is proved.