Inference for Stochastic Processes
1. Markov Chains

By **Robert L. Wolpert**

Duke University ISDS, USA

Basic Setup:

- Initial distribution $S_0 \sim \pi^0(ds)$ on state space $s \in S$
- Discrete: $S = \{s_i\}$, $\pi_i^0 \equiv \pi^0(\{s_i\})$
- Transitions $\mathsf{P}[S_{n+1} = s_j \mid \mathcal{F}_n] = Q_{ij}$, where $s_i \equiv S_n$
- Uncertainty: $Q = Q^\theta$, $\theta \in \Theta$ (ignore uncertainty about π^0)
- Log Likelihood:

$$\begin{aligned} \ell(\theta) &= \log \pi^0(S_0) + \sum_{m<n} \log Q^\theta(S_m, S_{m+1}) \\ &= \log \pi^0(S_0) + \sum_{i,j} N_{ij} \log Q^\theta_{ij}, \end{aligned}$$

where N_{ij} = Number of transitions $s_i \to s_j$ before time n.

Basic Properties:

- Marginal distribution $\pi_i^n \equiv \mathsf{P}[S_n = s_i] = \pi^0 Q_i^n$ (matrix powers)
- Diagonalize: $Q = U \Lambda U^{-1} \Rightarrow Q^n = U \Lambda^n U^{-1}$
- Eigenvalues: $\Lambda = \text{diag}(\vec{\lambda})$, $1 = \lambda_1 \geq |\lambda_2| \geq ... \geq |\lambda_n| \geq 0$
- Properties: Aperiodic & Irreducible $\Rightarrow |\lambda_2| < 1$
- Limits: $\pi^n \to \pi = U^{-1}_{1\cdot}$, first (left, row) eigenvector of Q
- Rate: $\pi^n = \pi + O(|\lambda_2|^n)$ (geometric convergence)

Example:

- $S = \{0, 1\}$; $Q = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$, $0 \leq a, b \leq 1$;
- Eigenvalues: $\det(Q - \lambda I) = (\lambda - 1)(\lambda - 1 + a + b)$, so $\lambda_1 = 1$, $\lambda_2 = 1 - a - b$;
- Properties: $0 < a + b < 2 \Rightarrow \lambda_2 = 1 - a - b$ satisfies $|\lambda_2| < 1$
- Limit: $\pi^n \to \pi = [\frac{b}{a+b}, \frac{a}{a+b}]$, stationary distribution;
- Rate: $\pi^n = \pi + O(|1 - a - b|^n)$, geometric convergence;
- Note: $a + b = 0 \Rightarrow$ reducible, $a + b = 1 \Rightarrow$ periodic; convergence fails in either of these cases.

Basic Inference:

MLE: For $\theta = \mathbf{q} \in \mathbb{R}^{n^2}$, $Q^\theta_{ij} = q_{ij}$, use Lagrange Multipliers:

$$\begin{aligned}
\ell(\mathbf{q}) &\equiv \log \pi^0(S_0) + \sum_{i,j} N_{ij} \log q_{ij} \\
L(\mathbf{q}, \lambda) &\equiv \log \pi^0(S_0) + \sum_{i,j} N_{ij} \log q_{ij} + \sum_i \lambda_i \left[1 - \sum_j q_{ij}\right] \\
\frac{\partial}{q_{ij}} L(\mathbf{q}, \lambda) &= N_{ij}/\log q_{ij} - \lambda_i = 0 \Rightarrow \hat{q}_{ij} = N_{ij}/\hat{\lambda}_i \\
\frac{\partial}{\lambda_i} L(\mathbf{q}, \lambda) &= [1 - \sum_j q_{ij}] = 0 = [1 - \sum_j N_{ij}/\hat{\lambda}_i], \text{ so} \\
\hat{\lambda}_i &= \sum_j N_{ij} = N_{i+} \text{ and} \\
\hat{q}_{ij} &= N_{ij}/\hat{\lambda}_i = N_{ij}/N_{i+}
\end{aligned}$$

Standard Errors:

For large sample-sizes and recurrent Markov chains we have asymptotic normality and so the standard errors SE^2_{ij} for the parameter q_{ij} may be taken from the Fisher information matrix. Let us consider the recurrent case where $\pi^n_i = \mathsf{P}[X_n = i] = [\pi^0 Q^n]_i \to \pi_i$, so for large t,

$$\mathsf{E}[N_{i+} | \pi^0, q] \approx t\pi_i,$$

and the diagonal terms of the Fisher information are

$$\begin{aligned}
I_{ij,ij} &= \mathsf{E}[N_{i+}/q_{ij}(1 - q_{ij})] \\
&\approx t\pi_i/\hat{q}_{ij}(1 - \hat{q}_{ij}),
\end{aligned}$$

so the Standard Error is $\mathrm{SE}_{ij} \approx \sqrt{\hat{q}_{ij}(1 - \hat{q}_{ij})/t\pi_i}$

More Inference:

MLE: Any function $\phi(Q)$ has MLE $\hat{\phi} = \phi(\hat{q})$; for example, $\hat{\lambda}_2$ is just the second eigenvalue of $\hat{q}$, $\hat{\pi}$ is just the first left eigenvector of $\hat{q}$.

Bayesian inference: For independent conjugate (Dirichlet) priors, $q_{i\cdot} \sim \mathsf{Di}(\vec{\alpha}_{i\cdot})$, the posterior distributions are again Dirichlet

$$q_{i\cdot}\big|_N \sim \mathsf{Di}(\alpha_{i+} + N_{i+}),$$

with posterior mean

$$\bar{q}_{ij} = E\left[q_{ij}\big|_{N_{\cdot\cdot}}\right] = \frac{\alpha_{ij} + N_{ij}}{\alpha_{i+} + N_{i+}}$$

(note this tends to the MLE $\hat{q}_{ij} = N_{ij}/N_{i+}$ as $\alpha \to 0$ or as the sample-size increases).

Comparative Inference:

$$\begin{aligned}
\ell(\theta) &\equiv \log \pi^0(S_0) + \sum_{i,j} N_{ij} \log Q^\theta_{ij} \\
\ell(\theta_0) - \ell(\theta_1) &= \sum_{i,j} N_{ij} \log \frac{Q^{\theta_0}_{ij}}{Q^{\theta_1}_{ij}}
\end{aligned}$$