# Estimators for Markov chains with missing observations

**Wolfgang Wefelmeyer**
University of Cologne
http://www.mi.uni-koeln.de/~wefelm/

Joint with

**Uschi Müller**, Texas A&M University,
**Anton Schick**, Binghamton University

**Avignon, June 15, 2016**

We consider an ergodic, real-valued, first-order Markov chain in discrete time, with transition distribution $Q(x, dy)$ and one-dimensional stationary distribution $\pi(dx)$.

$Q$ and $\pi$ are determined by expectations

$$Ef(X_{i-1}, X_i) = \pi \otimes Qf = \iint f(x, y)\pi(dx)Q(x, dy)$$

for sufficiently many $f$, e.g. by the two-dimensional distribution function.

From consecutive observations $X_0, \ldots, X_n$ such expectations can be estimated by empirical estimators

$$\mathbb{E}_2 f = \frac{1}{n} \sum_{i=1}^{n} f(X_{i-1}, X_i).$$

They are efficient in the nonparametric model.

Suppose we do not have consecutive observations.

We can still estimate one-dimensional expectations $Ef(X_i)$ by the empirical estimator based on the observations.
(Unless the gaps depend on the preceding states).

We do not need to know the time indices (the clock of the chain).

For two-dimensional expectations $Ef(X_{i-1}, X_i)$ we have two problems: We must see some pairs of adjacent realizations, and we must know this (i.e. we must know the clock of the chain).

For example, observe $X_0, X_k, X_{2k}, \ldots$. We can estimate the $k$-step transition distribution $Q^k$, but this does not determine $Q$.

In this talk we assume that we know the clock, and we consider patterns of observations with some adjacent pairs of observations.

We answer the following questions:

How can we use the information in nonadjacent pairs?

If the gaps depend on the preceding states, how can we identify $Q$?

1. Periodic observations

We observe at (known) times $0, j_1, j_1+j_2, \ldots, j_1+\cdots+j_m$ and repeat that pattern $n$ times.
Assume at least one of the steps $j_\mu$ is 1.

Block the observations:

$$\mathbf{Y}_1 = (Y_1, \ldots, Y_m) = (X_{j_1}, \ldots, X_{j_1+\cdots+j_m}), \ldots$$

This is an $m$-dimensional Markov chain with transition distribution

$$S(\mathbf{y}, d\mathbf{z}) = Q^{j_1} \otimes \cdots \otimes Q^{j_m}(\mathbf{y}, d\mathbf{z})$$

which depends only on the last entry $y_m$ of block $\mathbf{y}$.

The stationary distribution is $\pi \otimes Q^{j_2} \cdots \otimes Q^{j_m}(\mathbf{y}, d\mathbf{z})$.

Simplest case: Pattern 0,1,3;4,6,7;... with step sizes 1 and 2.
Observed blocks $\mathbf{Y}_1 = (Y_1, Y_2) = (X_1, X_3)$.
Two-dimensional chain with transition distribution $S = Q \otimes Q^2$.
A simple empirical estimator for $Ef(X_0, X_1)$ is based on observed pairs of adjacent realizations of the chain,

$$\mathbb{E}_2 f = \frac{1}{n} \sum_{i=1}^{n} g(X_{3(i-1)}, X_{3i-2}).$$

Non-adjacent pairs also contain information about $Q$.
Write block as $(X_1, X_2, X_3) = (X, Y, Z)$, with $Y$ unobserved.
Replace $f(X, Y)$ and $f(Y, Z)$ by backward and forward conditional expectations

$$f_{left}(X, Z) = E(f(X, Y)|X, Z), \quad f_{right}(X, Z) = E(f(Y, Z)|X, Z).$$

Express $f_{left}(X,Z) = E(f(X,Y)|X,Z)$ using one- and two-dimensional densities $p$ and $p_2$:

$$f_{left}(x,z) = \frac{\displaystyle\int \frac{p_2(x,y)p_2(y,z)}{p(y)} f(x,y)\,dy}{\displaystyle\int \frac{p_2(x,y)p_2(y,z)}{p(y)} dy}.$$

Plug in kernel density estimators based on the observed pairs.
We obtain a new "empirical estimator" for $Ef(X,Y)$ based on non-adjacent pairs,

$$\mathbb{E}_2^{(2)} \hat{f}_{left} = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_{left}(X_{3i-2}, X_{3i}).$$

Similarly for $f_{right}$.

The "plug-in principle" leads from $o(n^{-1/4})$ rates for the kernel estimators to a $n^{-1/2}$ rate for these "empirical estimators".
Combine estimators $\mathbb{E}_2 f$ with $\mathbb{E}_2^{(2)} \hat{f}_{left}$ and $\mathbb{E}_2^{(2)} \hat{f}_{right}$.

2. Observations "missing completely at random"

Jump $j_1, \dots, j_m$ steps with probabilities $w_1, \dots, w_m$.
Assume at least one of the steps $j_\mu$ is 1.

(Compare MCAR in regression.)

This is a one-dimensional Markov chain with transition distribution a mixture of $j_\mu$-step transition distributions

$$S(y, dz) = \sum_{\mu=1}^{m} w_\mu Q^{j_\mu}(y, dz).$$

The stationary distribution is again $\pi$. We observe $X_0, \dots, X_n$ and the step sizes $J_{i-1}$ from $X_{i-1}$ to $X_i$.

For $\mu = 1 \dots, m$ and pairs $X_{i-1}, X_i$ with $J_{i-1} = j_\mu$, construct estimators as before (for periodic patterns).

2. Observations “missing at random”

At any time point and in state $x$, jump $j_1, \ldots, j_m$ steps with probabilities $w_1(x), \ldots, w_m(x)$. possibly depending on the state $x$.
Assume at least one of the steps $j_\mu$ is 1.

(Compare MAR in regression.)

This is a one-dimensional Markov chain wit transition distribution a conditional mixture of $j_\mu$-step transition distributions

$$S(y, dz) = \sum_{\mu=1}^{m} w_\mu(y) Q^{j_\mu}(y, dz).$$

It is still ergodic, but the stationary distribution is not $\pi$, but $\varrho$, say.
Now we can not estimate $\pi \otimes Qf$ directly.
We can however estimate $Q$ as follows.

The stationary distribution of $(X_{i-1}, X_i)$ given $J_{i-1} = j_\mu$ is

$$\varrho_2^{(j_\mu)}(dx, dy) = \varrho(dx) \frac{w_\mu(x)}{\varrho w_\mu} Q^{j_\mu}(x, dy).$$

If we have densities, this can be written

$$r_2^{(j_\mu)}(x, y) = r(x) \frac{w_\mu(x)}{\varrho w_\mu} q^{\otimes j_\mu}(x, y).$$

For step $j_\mu = 1$,

$$r_2(x, y) = r(x) \frac{w_\mu(x)}{\varrho w_\mu} q(x, y),$$

hence

$$q(x, y) = \frac{\varrho w_1}{w_1(x)} \frac{r_2(x, y)}{r(x)}$$

if $w_1(x) > 0$. For such $x$, we can estimate $q(x, y)$ from the observed pairs. Otherwise not, because we never see the next realization of the chain.

For pairs of adjacent observations $(X, Y)$, a kernel estimator for the joint density $r_2(s,t)$ with $w_1(s) > 0$ uses

$$K_b(s-X)K_b(t-Y).$$

Here $K_b(x) = K(x/b)/b$ with kernel $K$ and bandwidth $b$.

The information in pairs of non-adjacent observations can be used as follows. If $Y$ is not observed, but, say, the next observation $Z$, and $w_2(s) > 0$, use instead of $K_b(s-X)K_b(t-Y)$ the conditional expectation

$$m_{left}(s,t)(X,Z) = E\Big(K_b(s-X)K_b(t-Y)|X,Z\Big).$$

It can be expressed in terms of $r$, $r_2$ and $w_1$, which can be estimated from the observations. Similarly for

$$m_{right}(s,t)(X,Z) = E\Big(K_b(s-Y)K_b(t-Z)|X,Z\Big).$$

Estimate $q(s,t)$ by

$$\widehat{q}(s,t) = \frac{\mathbb{E}_2^{(2)} \widehat{m}_{left}(s,t)}{\widehat{r}^{(2)}(s)} = \frac{\sum\limits_{J_{i-1}=2} \widehat{m}_{left}(X_{i-1}, X_i)}{\sum\limits_{J_{i-1}=2} K_b(s - X_{i-1})}.$$

To estimate $m_{left}(X,Z) = E(K_b(s-X)K_b(t-Y)|X,Z)$, write in terms of $r$, $r_2$ and $w_1$, and replace these terms by (kernel) estimators:

$$\begin{aligned} m_{left}(s,t)(X=x,Z=z) &= E\Big(K_b(s-X)K_b(t-Y)|X=x,Z=z\Big) \\ &= \frac{\int q(x,y)(q(y,z)K_b(s-x)K_b(t-y)\,dy}{q^{(2)}(x,z)} \\ &= \frac{\int \frac{r_2(x,y)r_2(y,z)}{w_1(y)r(y)} K_b(s-x)K_b(t-y)\,dy}{\int \frac{r_2(x,y)r_2(y,z)}{w_1(y)r(y)}\,dy}. \end{aligned}$$