Transition and Duration Models

Introduction to Markov Chains

Christian Schluter

Markov chains

Consider a discrete time-homogeneous **Markov chain** $\{X_t\}$, where $t = 1, 2, \cdots$ denotes the period, that takes values in a countable state space $S = \{1, 2, \cdots\}$.

The **transition probabilities** are denoted by

$$p_{ij} = \Pr(X_{t+1} = j | X_t = i) = \Pr(X_1 = j | X_0 = i)$$

We collect these transition probabilities in the **transition matrix** P

$$P = [p_{ij}]_{i=1,2,\cdots;j=1,2,\cdots}$$

Denote the probability of going from state i to state j in n time steps (the "n-step" transition probability) by

$$p_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i) = \Pr(X_{n+k} = j \mid X_k = i)$$

Going from i to j could be achieved by a detour via k:

Result (the Chapman-Kolmogorov equation)

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}$$

i.e. $P_{m+n} = P_m P_n$. It follows that $P_n = P^n$.

Proof: Exercise set 2.

Denote the **distribution** of X_t at time t over the various states $s \in S$ by

$$\pi_t = (\Pr(X_t = 1), \Pr(X_t = 2), \cdots).$$

The conditional mean and variance of the process are

$$\begin{aligned} E(s'|s = s_i) &= \sum_j p_{ij} s_j \\ Var(s'|s = s_i) &= \sum_j p_{ij} s_j^2 - [E(s'|s = s_i)]^2 \end{aligned}$$

The Invariant Distribution

The system evolves according to

$$\pi_{t+1} = \pi_t P$$

Iterating, if we start with distribution π_0, after n periods, we have

$$\pi_n = \pi_0 P^n$$

It might be that the system settles down eventually, so that

$$\pi^* = \pi^* P$$

In this case, π^* is called the **invariant distribution**.

We can interpret the invariant distribution in two ways:

(i) π_i^* is the unconditional probability that the chain is currently in state i.

(ii) π_i^* is the probability that the chain will be in state i in t steps as $t \to \infty$.

The long-run mean and variance can be obtained using the invariant distribution

$$\begin{aligned} E(s) &= \sum_i \pi_i^* s_i \\ Var(s) &= \sum_i \pi_i^* s_i^2 - [E(s)]^2 \end{aligned}$$

Under which conditions does a unique invariant distribution exist ?

The chain must not cycle (so it is **aperiodic**, returns to state i occurring at irregular times) and must be such that it is possible to move from any state to any other state in a countable number of periods (so it is said to be **irreducible**).

For a Markov chain that is both irreducible and aperiodic P possesses only one eigenvalue of modulus 1 (that is has one such eigenvalue is a consequence of the Perron-Frobenius theorem).

Result: Under these conditions, the invariant distribution exists (and is unique).

Classification of states and transition matrices

An **irreducible** Markov chain has the property that it is possible to move from any state to any other state in a countable number of periods.

An **absorbing** state is a state i^* such that $p_{i^* i^*} = 1$.

A **persistent state** is a state i^* to which we return with certainty, $\Pr(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$.

A **transient state** is a state i^* with a positive probability that we will never return to i^*, so $\Pr(X_n = i \text{ for some } n \geq 1 | X_0 = i) < 1$.

More formal definitions are as follows.

A state i is said to be **transient** if, given that we start in state i, there is a non-zero probability that we will never return to i.

Let the random variable T_i be the first return time to state i (the "hitting time"):

$$T_i = \inf\{n \geq 1 : X_n = i \mid X_0 = i\}$$

$$f_{ii}^{(n)} = \Pr(T_i = n) = \Pr(X_1 \neq i, X_2 \neq i, \cdots, X_{n-1} \neq i, X_n = i | X_0 = i)$$

denotes the probability that we return to state i for the first time after n steps.

The probability that the chain ever revisits i is

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)}$$

State i is transient if

$$\Pr(T_i < \infty) = \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$$

State i is persistent if $f_{ii} = 1$.

State i is **recurrent** (or **persistent**) if it is not transient, so hitting times are finite.

State i is recurrent if and only

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

This implies that the expected number of visits to this state is infinite:

The mean recurrence time at state i is the expected return time

$$M_i = E[T_i] = \sum_{n=1}^{\infty} n \cdot f_{ii}^{(n)}$$

A state i is said to be **ergodic** if it is aperiodic and positive recurrent, i.e. if it is recurrent, has a period of 1, and has finite mean recurrence time.

If all states in an irreducible Markov chain are ergodic, then the chain is said to be **ergodic**.

Markov's theorem: a Markov chain is ergodic if there is a positive probability to pass from any state to any other state in one step.

("ergodicity" in general refers to invariance. Here, the ergodic MC has an invariant distribution. In time series, ergodicity was defined as a form of asymptotic independence. There, the Ergodic Theorem stated: If $\{y_t\}$ is ergodic and its rth moment μ_r is finite, then $\frac{1}{T}\sum_{t=1}^{T} y_t^r \to^p \mu_r$).

π^* is a stationary distribution if $\pi^* = \pi^* P$.

An irreducible aperiodic Markov chain has a stationary distribution.

In this case

$$\pi_i^* = \frac{1}{M_i}$$

with M_i the mean recurrence time of state i.

Consider the n-step transition probability. For an irreducible aperiodic Markov chain we have

$$p_{ij}^{(n)} \to \frac{1}{M_j}$$

as $n \to \infty$ for all i and j (i.e. the chain forgets its starting point i).

Examples

Our objective is to calculate this invariant distribution, if it exists.

Method 1: Compute

$$\Pi^* = \lim_{t \to \infty} P^t$$

This produces a matrix with identical columns, each of which is the invariant distribution. Under certain assumptions (irreducibility and aperiodicity), this limit exists and is unique.

Method 2: (from Linear Algebra) Compute the (normalised) eigenvector that is associated with the eigenvalue of 1 (which is in fact the largest eigenvalue).

The transition matrix of a Markov chain with state space $\{0, 1, 2\}$ is given by

$$\begin{pmatrix} .2 & .4 & .4 \\ .2 & .5 & .3 \\ .5 & .2 & .3 \end{pmatrix}$$

Calculate the stationary distribution of this Markov chain.

Numerical Examples

```
library(markovchain)
library(expm)
## state space z={0,1,2}
z = 0:2
P = matrix(c(.2,.4,.4,.2,.5,.3
      ,.5,.2,.3), 3,3, byrow=TRUE)

## method 1
pi.star <-  P%^%50  ## %^%t operator from expm library
pi.star

## pi.star
##          [,1]     [,2]     [,3]
## [1,] 0.2989691 0.371134 0.3298969
## [2,] 0.2989691 0.371134 0.3298969
## [3,] 0.2989691 0.371134 0.3298969

pi.star <- pi.star[1,]
## 0.2989691 0.3711340 0.3298969

## check: t(P) %*% pi.star

## method 2
## the same in: library(markovchain)
mcP <- new("markovchain", states=as.character(z),
     transitionMatrix = P)
steadyStates(mcP)

##              1        2         3
## [1,] 0.2989691 0.371134 0.3298969
```

```
## method 3: diagonalisation
matrix.power <- function(A, n) {
   e <- eigen(A)
   M <- e$vectors   # matrix for changing basis
   d <- e$values    # eigen values
   return(M %*% diag(d^n) %*% solve(M))
}
matrix.power(P,50)

## (B) normalised eigenvector of eigenvalue 1

temp <- eigen(t(P))  ## P' p_t = p_t+1  parametrisation
temp$vectors[,1] / sum(temp$vectors[,1]) ## orthonormali
## [1] 0.2989691 0.3711340 0.3298969
```

```
## Analyse the long run mean and variance
##    of the MC process

## theoretical values
## E(y) = sum_i pi.star[i] z[i]
mean.LR <- sum(pi.star * z)
## 1.030928
## Var(y) = sum_i pi.star[i] z[i]^2 -
      (sum_i pi.star[i] z[i])^2
var.LR <- sum(pi.star * z^2) - (mean.LR)^2
## 0.6279094
```

```
## estimate from simulation
outs <- rmarkovchain(n = 10000, object = mcP,
     what = "matrix")
tail(outs)
MCsim <- as.integer(outs)
tail(MCsim)

## need to let the sim. burn in and settle down
##   consider only the last 2000 realisations
##   of the process
mean(tail(MCsim,2000))  ## close to mean.LR
var(tail(MCsim,2000))   ## close to var.LR
```

Example: First-return time

```
z = 0:2
P = matrix(c(.2,.4,.4,.2,.5,.3,.5,.2,.3), 3,3,
     byrow=TRUE)

##      [,1] [,2] [,3]
## [1,]  0.2  0.4  0.4
## [2,]  0.2  0.5  0.3
## [3,]  0.5  0.2  0.3

mcP <- new("markovchain", states =  as.character(z),
         transitionMatrix = P)
pi.star <- steadyStates(mcP)

##              0        1         2
## [1,] 0.2989691 0.371134 0.3298969
```

```
## first return time to state "2"

M <- rep(NA,1000)
for (i in 1:1000) {
    outs <- rmarkovchain(n = 50, object = mcP,
             what = "matrix")
    temp <- which((outs=="2"))
    M[i] <- temp[2] - temp[1]
    ## temp[1] is the first time in the state,
    ## temp[2] is the first return
}

c(mean(M), 1/pi.star[which(z==2)])
## [1] 2.91400 3.03125
```

Applications

In some situations, especially in quantitative macro, it is useful to approximate a stationary continuous state space process by a discrete state space model such as a Markov Chain (with time-invariant transition probabilities).

For instance, it is possible to approximate the AR(1) process

$$y_{t+1} = \rho y_t + \varepsilon_{t+1}$$

where $|\rho| < 1$ and $\varepsilon \sim (0, \sigma^2)$.

The idea is to find a Markov Chain that has the same principal features as the AR(1) process, i.e. the same mean and autocovariances

Example. Consider $y_{t+1} = \rho y_t + \varepsilon_{t+1}$. By recursive substitution

$$y_t = \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} + \rho^t y_0$$

and let $t \to \infty$.

The unconditional mean of an AR(1) is zero.

The unconditional variance is

$$V[y_t] = \frac{\sigma^2}{1-\rho^2}$$

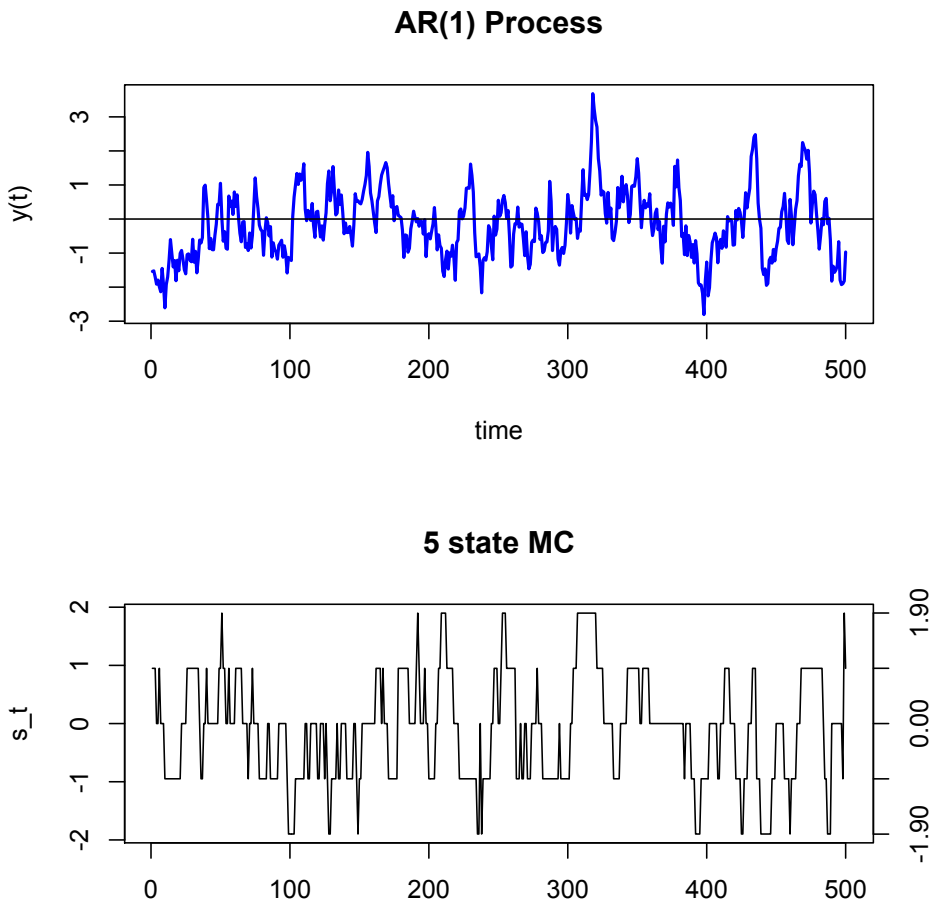
The autocovariances are

$$\gamma_s = E[(y_t - E[(y_t])(y_{t-s} - E[(y_{t_s}])] = \rho^s \frac{\sigma^2}{1-\rho^2}$$

so the autocorrelation function is $\gamma_1 / V[y_t] = \rho$.

The approximation method (Rouwenhorst method) is explained in Kopecky and Suen (2010, REconDyn).

Here we simply look at some of the results.



State-Space Modelling: MC approximation to an AR(1) process

```
rho<-0.85     #parameter of AR(1)
sigma.e<-0.5  #standard deviation of errors in AR(1)
sigma.e^2 / (1-rho^2)
## unconditional variance :   0.9009009

## simulate and visualise the AR(1) process
ar1.sim =  arima.sim(model=list(ar=rho),
              sd=sigma.e,n=500)
ts.plot(ar1.sim,main="AR(1) Process",
   xlab="time",ylab="y(t)", col="blue", lwd=2)
abline(h=0)

acf.sample <- acf(ar1.sim, main="",
               type=c("correlation"))
acf.sample[1]
## compare to theoretical value = rho
##     (=.85); in my sim 0.853
acf(ar1.sim, main="", type="covariance",
    plot=FALSE)[0]  ## in my sim 0.965
```

```
## the discrete state space approx. with 5 states
states <- 5
markov_grid <- seq(-psi,psi, length.out=states)
## psi=1.898316
## the method yields transition matrix P
round(P,3)
      [,1]  [,2]  [,3]  [,4]  [,5]
[1,] 0.732 0.237 0.029 0.002 0.000
[2,] 0.059 0.747 0.179 0.014 0.000
[3,] 0.005 0.119 0.751 0.119 0.005
[4,] 0.000 0.014 0.179 0.747 0.059
[5,] 0.000 0.002 0.029 0.237 0.732

## the invariant distribution
pi.star <- steadyStates(mcB)
##          1    2     3    4      5
## [1,] 0.0625 0.25 0.375 0.25 0.0625

## E(y) = sum_i pi.star[i] z[i]
mean.LR <- sum(pi.star * markov_grid) ## 0
## Var(y) = sum_i pi.star[i] z[i]^2 - (sum_i pi.star[i]
var.LR <- sum(pi.star * markov_grid^2) - (mean.LR)^2
## 0.9009009 = sigma.e^2 / (1-rho^2)

## first order auto-covariance of Markov process (in the
##   E(z'z)= E_{z} {z E{z'|z}} =
##         = sum pi.star_i z_i [sum z' Pr{z'|z}]
(pi.star * markov_grid) %*%  (P %*% markov_grid) ##
## FO auto-correlation
(pi.star * markov_grid) %*%  (P %*% markov_grid) / var.L
## 0.85 = rho
```

Although the discrete state-space Markov Chain approximation to the (de-trended, so stationary) AR(1) process is the most popular approximation, it is not the only one.

The discrete time AR(1) process has an analogue in continuous time, which is the Ornstein Uhlenbeck process:

$$dy = -\gamma y dt + \sigma dB$$

where B denotes Brownian motion $\{B(t) : t \geq 0\}$, the continuous time analogue of the random walk.

Shimer (2005, AER) uses an approximation to the OU process, which exploits the properties of the Poisson process.

The realisations of the discrete state space process for $\{y : t \geq 0\}$ lie on a symmetric grid

$$Y = \{-n\Delta, -(n-1)\Delta, 0, (n-1)\Delta, n\Delta\}$$

of $2n+1$ gridpoints where Δ denotes the step-size.

If a shock occurs, the process $y(t)$ moves either up, $y(t+1) = y(t) + \Delta$ with probability

$$p(y) = \frac{1}{2}\left(1 - \frac{y}{n\Delta}\right),$$

or it moves down, $y(t+1) = y(t) - \Delta$ with complementary probability $1 - p(y)$.

The advantage of this modelling apporach is that the transition matrix is sparse.

Now assume that shocks arrive according to a Poisson process with arrival rate λ. Consider the behaviour of $y(t+h)$ for a short time period $h \to 0$: The probability that no shock arrives is $1 - h\lambda$, and that one shock arrives is $h\lambda$.

If no shock arrives, $y(t+h) = y(t)$. If one shock arrives, the process moves either down or up. Consider now the increments of the process:

$$\begin{aligned} E\left(y(t+h) - y(t) | y(t)\right) &= 0(1-h\lambda) \\ &+ \Delta h\lambda \frac{1}{2}\left(1 - \frac{y(t)}{n\Delta}\right) - \Delta h\lambda \frac{1}{2}\left(1 + \frac{y(t)}{n\Delta}\right) \\ &= h\frac{\lambda}{n} y(t) \\ &\equiv -h\gamma y(t) \end{aligned}$$

with $\gamma = \lambda / n$.

Similarly, the conditional variance is

$$\begin{aligned} Var & (y(t+h) - y(t)|y(t)) \\ = & E\left((y(t+h) - y(t))^2|y(t)\right) - \left(E\left(y(t+h) - y(t)|y(t)\right)\right)^2 \\ = & h\lambda\Delta^2 \equiv h\sigma^2 \end{aligned}$$

with $\sigma^2 = \lambda\Delta^2$.

Putting everything together

$$\begin{aligned} E\left(y(t+h)-y(t)|y(t)\right) &= -h\gamma y(t) \\ Var\left(y(t+h)-y(t)|y(t)\right) &= h\sigma^2 \end{aligned}$$

which can be represented (divide by h and letting $h \to 0$) by the process

$$dy = -\gamma y(t) + \sigma dx$$

where $x(t)$ represents a white noise error process.

The continuous state space OU process obtains by replacing (Δ, λ, n) by $(\Delta\sqrt{\varepsilon}, \lambda/\varepsilon, n/\varepsilon)$ and letting $\varepsilon \to 0$.

G. Topa (2001), "Social Interactions, Local Spillovers and Unemployment", *The Review of Economic Studies*, Vol. 68, No. 2, 261-295.

A model that explicitly incorporates local interactions and allows agents to exchange information about job openings within their social networks. Agents are more likely to be employed if their social contacts are also employed.

Unemployment in the model evolves according to a Markov process over the set of locations.

At any point in time, employed individuals may become unemployed with some exogenous probability, whereas unemployed individuals may find a job with probability increasing in the number of neighbours currently employed.

This process generates positive spatial covariances of unemployment between nearby locations.

Let S be a finite set of locations. A location in this framework is taken to represent a Census tract. Time flows discretely.

The state of each tract every period, $y_{i,t}$,is the employment rate within each area.

The state variable can only take a finite number of values in the interval $[0, 1]$: $y_{i,t} \in E = \{e_1, \cdots, e_K\} = \{0, 0.1, \cdots, 1\}$.

Distance: the distance d between any two tracts i and j is defined as the number of tract boundaries that one needs to cross to travel from a point inside tract i to a point inside tract j.

The transition probabilities are assumed to be as follows.

Consider first the boundaries. If tract i is at full employment ($y_{it} = 1$), then it may drop to the next lower employment rate with a probability

$$p_d = \Pr(y_{i,t+1} = 0.9 | y_{i,t} = 1)$$

and will stay at the present state with probability $(1 - p_d)$.

The probability of going from zero employment to the next higher employment rate is assumed to be

$$p_u = \Pr(y_{i,t+1} = 0.1 | y_{i,t} = 0)$$

If the state of tract i at time t is in the interior

$$\begin{aligned} \Pr(y_{i,t+1} = e_{k-1} | y_{i,t} = e_k) &= \frac{1}{2} p_d \\ \Pr(y_{i,t+1} = e_{k+1} | y_{i,t} = e_k) &= \frac{1}{2} p_u \end{aligned}$$

the state of tract i may remain unchanged during the next period with probability $1 - .5(p_d + p_u)$.

The model generates a first-order Markov process on the map of locations.

The state space Y contains all possible configurations of employment rates over the set of locations. Index each state by $w = 1, ..., W$ where W is the total number of possible states for the whole system.

The system evolves according to

$$\mu_{t+1} = Q\mu_t$$

where Q is the $(W \times W)$ transition matrix, whose entries q_{rs} denote the transition probabilities from state r to state s.

The stationary distribution satisfies

$$\mu^* = Q\mu^*$$

Since we have a MC that is aperiodic and irreducible, it follows that this stationary distribution exists and is unique.

How could one model spatial interactions within this framework ?

Let X_i measure time-invariant characteristics of location i.

Assume (in practice, all functions $(\gamma, \alpha, \lambda)$ are assumed to be linear)

$$\begin{aligned} p_d &= \gamma(X_i) \\ p_u &= \alpha(X_i) + \lambda(X_i) * I_{it} \end{aligned}$$

where I_{it} measures the information that residents of tract i may receive from their employed social contacts in the neighbouring areas. E.g. this could be measured by the mean employment rate in the neighbouring tracts

$$I_{it} = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{j,t}$$

where N_i denotes locations that are neighbours of i.

Ryo Nakajima (2007, ReStud), "Measuring Peer Effects on Youth Smoking Behaviour" proposes a statistical model of social interactions.

The research question is whether there are peer effects in smoking, i.e. whether one's smoking status is influenced by the smoking status one's friends.

If this is so, then peer effects lead to amplifications, so-called *social multipliers*.

Let $y_i^*(t)$ denote the *latent* utility derived from from smoking by person i at time t. The observed smoking status is denoted by $y_i(t)$ equal to 1 if $y_i^*(t) \geq 0$ and -1 otherwise.

Denote the set of friends by I. The influence of friend j on i will be measured by the parameter ρ_{ij}. The social interaction model (for latent utility) is

$$y_i^*(t) = \alpha + \beta x_i + \sum_{j \neq i \in I} \rho_{ij} y_i(t-1) + \varepsilon_i(t)$$

This is a model of adaptive behaviour, since it is the lagged observed smoking status of friends that might have an influence on person i's utility.

Note that it is assumed that decisions are taking sequentially and not simultaneously. Also, we have a model on "on-off-on" events, since over time a person can change between smoking status.

Let $d(t) \in I$ denote the person taking the smoking decision at time t.

Result: If we assume that any person is able to make a decision in each period ($\Pr\{d(t) = i\} > 0 \quad \forall i \in I$), and that the $\varepsilon_i(t)$ are iid across persons and time, then the social interaction process an aperiodic and irreducible Markov Chain.

This then implies that as $t \to \infty$ the process converges to a *unique* steady-state distribution

$$\lim_{t \to \infty} \Pr(y(t) = w | y(0) = w(0), x) = P^*(y = w|x)$$

Assume further that the $\varepsilon_i(t)$ are iid logistic, $\varepsilon_i(t) \sim F(\varepsilon) = \exp(\varepsilon)/[1+\varepsilon]$.

Result: The steady-state distribution is

$$P^*(y = w|x) = \frac{\exp Q(w|x)}{\sum_\eta \exp Q(\eta|x)}$$

where

$$\begin{aligned} Q(w|x) &= \frac{1}{2}\sum w_i(\alpha + \beta x_i) + \frac{1}{2}\sum_i \sum_{j \neq i} \rho_{ij} w_i w_j \\ &= \frac{1}{2}\sum w_i(\alpha + \beta x_i) + \frac{1}{4}\sum_i \sum_j \rho_{ij} w_i w_j \end{aligned}$$