Credit Rating Modeling

Chapter 2: A Markov chain approach to rating migrations

Guglielmo D'Amico
g.damico@unich.it

PhD course activated for the XXXI cycle
Accounting, Management and Finance

Outline of lesson

Outline of lesson

Outline of lesson

Definition

A family of random variables $\{X_t,\ t \in \mathcal{T}\}$ parameterized by index $t \in \mathcal{T}$ is called a stochastic process or random process.

Invariably t will represent *time*, although in general the index t could represent a spatial position or differently.
The set E of possible values of the process is called the *state space* of the random process.

Classification

- ◇ if $\mathcal{T}$ is discrete, then we have a discrete time process
- ◇ if $\mathcal{T}$ is continuous, then we have a continuous time process
- ◇ if E is discrete, then we have a discrete value process
- ◇ if E is continuous, then we have a continuous value process

We will work with discrete value stochastic process in discrete time

A particular realisation of X_t for all $t \in \mathcal{T}$ is called a sample path of the process X_t, so a sample path is an outcome of a stochastic process.

One should keep in mind that in the background there is some sample space Ω on which all the random variables $\{X_t,\ t \in \mathcal{T}\}$, are defined. So a sample path of the process X is the function $t \mapsto X_t(\omega)$ for some fixed $\omega \in \Omega$.

Example

$$C(0) = 100,\ C(t) = C(t-1) + W(t), \quad W(t) \perp W(s)\ \forall s \neq t$$

$$W(t) := \begin{cases} 1 & \text{if Head} \\ -1 & \text{if Tail} \end{cases}$$

Markov chain

Definition

The process $\{X_t,\ t = 0, 1, 2, \ldots\}$ is called a **Markov chain** if, for each t and $i_0, i_1, \ldots, i_{t-1} \in E$ and $j \in E$

$$\begin{aligned} &\mathbb{P}(X_t = j \mid X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \ldots, X_0 = i_0) \\ &= \mathbb{P}(X_t = j \mid X_{t-1} = i_{t-1}) =: p_{i_{t-1},j}(t-1, t) \end{aligned} \tag{1}$$

Intuitively the Markov property says that the future behaviour of a Markov process is determined by where the process currently is but not by where it was in the past.

For a Markov Chain $X_t : t \in \mathbb{N}$, we write for mn

$$p_{i,k}(m,n) = \mathbb{P}(X_n = k | X_m = i)$$

and refer to $(p_{i,k}(m,n); i \in E, k \in E)$ as the transition probabilities from time m to time n. Since the state space E is discrete, it can be put in a bijective correspondence with a subset of the positive integers and each state in E can be identified by an integer. Without loss of generality we assume that the state space is of the form $E = \{1, 2, \ldots, d\}$ for some $d \in \mathbb{N}$. Then we can construct a $d \times d$ matrix $\mathbf{P}(m,n)$ whose (i,k)-th entry is the transition probability $p_{i,k}(m,n)$.

few properties of the matrices $\mathbf{P}(m, n)$

1) $\mathbf{P}(n, n) = \mathbf{I}$

2) $\sum_{k \in E} p_{i,k}(m, n) = 1$

3) The transition probabilities of a discrete time Markov Chain satisfy the Chapman-Kolmogorov equations

$$p_{i,k}(m, n) = \sum_{j \in E} p_{i,j}(m, l) p_{j,k}(l, n)$$

$\forall i, k \in E$ and all times $m \leq l \leq n$. In matrix notation this equality reads as

$$\mathbf{P}(m, n) = \mathbf{P}(m, l)\mathbf{P}(l, n).$$

A consequence of the Chapman-Kolmogorov equations is

$$\mathbf{P}(m,n) = \mathbf{P}(m,m+1)\cdot\mathbf{P}(m+1,m+2)\cdot\ldots\cdot\mathbf{P}(n-2,n-1)\cdot\mathbf{P}(n-1,n).$$

So in order to construct a transition matrix function, one only needs to specify the one step transition matrices $\mathbf{P}(n,n+1)$.

Let $v_i(n) = \mathbb{P}(X_n = j)$ and $\mathbf{v}(n) = (v_1(n), v_2(n), \ldots, v_d(n))$. Then

$$v_j(n) = \sum_{i\in E} \mathbb{P}(X_n = j \mid X_0 = i)\mathbb{P}(X_0 = i),$$

or in matrix notation

$$\mathbf{v}(n) = \mathbf{v}(0)\cdot\mathbf{P}(0,n).$$

Example

$E = \{1, 2, 3\}$, $\mathbf{v}(0) = (0.2, 0.5, 0.3)$.

$$\mathbf{P}(0,1) := \begin{pmatrix} 0.7 & 0.3 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.3 & 0.7 \end{pmatrix}$$

$$\mathbf{P}(1,2) := \begin{pmatrix} 0.0 & 0.4 & 0.6 \\ 0.7 & 0.0 & 0.3 \\ 0.1 & 0.9 & 0.0 \end{pmatrix}$$

Compute $\mathbf{v}(2)$.

Time homogeneous Markov chains

A Markov chain is said to be time homogeneous if the transition probabilities are invariant to time translations, i.e.

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i), \ \forall n \in \mathbb{N}.$$

Simplified notation for the time homogeneous case:
The r-step transition matrices $\mathbf{P}(0, r), \mathbf{P}(1, r+1), \mathbf{P}(2, r+2), \ldots$ are all equal to each other. Therefore we can use the notation

$$p_{i,k}(r) := p_{i,k}(n, n+r) \quad \mathbf{P}(r) := \mathbf{P}(n, n+r)$$

We refer to $\mathbf{P}(r)$ as the r-step transition matrix. For $r = 1$ we obtain the one-step transition matrix $\mathbf{P}(1)$ denoted by $\mathbf{P}$.

In the case of time homogeneous Markov chains, the Chapman-Kolmogorov equations becomes

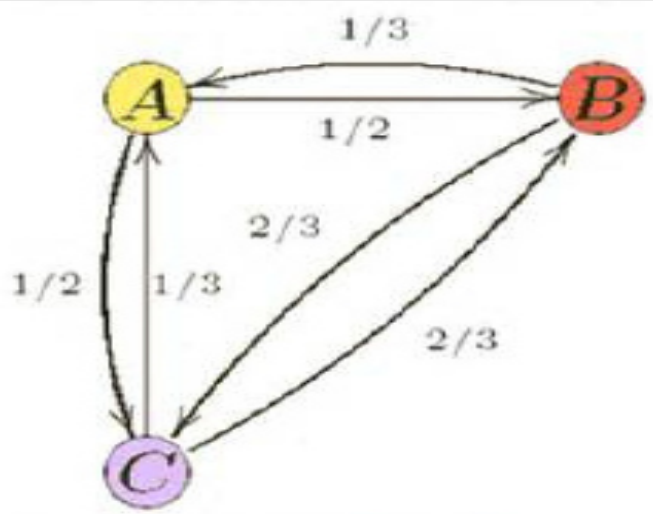
$$\mathbf{P}(n+r) = \mathbf{P}(n) \cdot \mathbf{P}(r), \;\; n, r \geq 0,$$

and consequently

$$\mathbf{P}(n) = \mathbf{P}^n,$$

i.e. the n-step transition matrix is equal to the nth power of the one step transition matrix.

Transition Matrices



$$\mathbf{P} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{pmatrix} \end{array}$$

P1. Given $\mathbf{v}(0) = (0, 1, 0)$ compute $\mathbf{v}(4)$;
P2. Given $\mathbf{v}(0) = (0.2, 0.6, 0.2)$ compute $\mathbf{v}(5)$.

Stationary distribution

Definition

A vector $\pi = (\pi_1, \pi_2, \ldots, \pi_d)$ is called a stationary or invariant (probability) distribution of an HMC with d states and with transition matrix $\mathbf{P}$ if
i) $\pi_i \geq 0$ for all $i \in E$ and $\sum_{i \in E} \pi_i = 1$;
ii) $\pi = \pi\mathbf{P}$ (i.e. π is a left-eigenvector of the matrix $\mathbf{P}$ with eigenvalue equal to 1).

If $\mathbf{v}(0) = \pi$, then $\mathbf{v}(n) = \pi$. Indeed, we have....

Hence the name stationary (or invariant) distribution.

Computing stationary distributions

In order to find the stationary distributions of a MC one needs to solve a linear system of $d+1$ equations in d unknown:

$$\begin{cases} \pi(\mathbf{I}-\mathbf{P}) = 0 \\ \pi \mathbf{1}^T = 1 \end{cases}$$

Alternatively

$$\pi = \mathbf{1} \cdot \left(\mathbf{I} - \mathbf{P} + \mathbf{1}^T \cdot \mathbf{1}\right)^{-1}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$.

Homework

1) Take $E = \{1, 2\}$ and define the transition matrix

$$\mathbf{P} := \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

with $\alpha, \beta \in (0, 1)$. Find the stationary distribution of the MC.

2) Calculate the stationary distribution for

$$\mathbf{P} := \begin{pmatrix} 0.8 & 0.2 & 0.0 \\ 0.4 & 0.5 & 0.1 \\ 0.4 & 0.0 & 0.6 \end{pmatrix}.$$

Asymptotic behavior

Theorem

Let $\mathbf{P}$ *be the transition matrix of an irreducible, aperiodic MC* $\{X_n : n \in \mathbb{N}\}$. *Then it admits a stationary distribution* π *and*

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \pi_k, \quad \forall k \in E;$$

$$\lim_{n\to\infty} p_{i,k}(n) = \pi_k \quad \forall i, k \in E;$$

or in matrix notation

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_d \\ \pi_1 & \pi_2 & \cdots & \pi_d \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_d \end{pmatrix}.$$

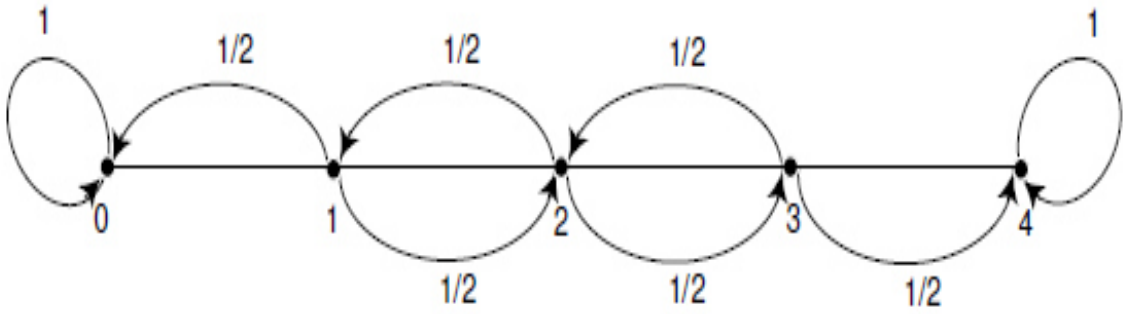
Absorbing chains

Definition

A state $i \in E$ of a Markov chain is called absorbing if it is impossible to leave it (i.e., $p_{i,i} = 1$). A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).

Definition

In an absorbing Markov chain, a state which is not absorbing is called transient.



We form a Markov chain with states 0, 1, 2, 3, and 4. States 0 and 4 are absorbing states. The transition matrix is then

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} .$$

The states 1, 2, and 3 are transient states, and from any of these it is possible to reach the absorbing states 0 and 4. Hence the chain is an absorbing chain. When a process reaches an absorbing state, we shall say that it is *absorbed*. □

Classical Problems

The most obvious question that can be asked about such a chain is:
What is the probability that the process will eventually reach an absorbing state?
Other interesting questions include:
(a) What is the probability that the process will end up in a given absorbing state?
(b) On the average, how long will it take for the process to be absorbed?
(c) On the average, how many times will the process be in each transient state?
The answers to all these questions depend, in general, on the state from which the process starts as well as the transition probabilities.

Canonical Form

Consider an arbitrary absorbing Markov chain. Renumber the states so that the transient states come first. If there are r absorbing states and t transient states, the transition matrix will have the following canonical form

$$\mathbf{P} = \begin{array}{c} \\ TR \\ ABS \end{array} \begin{array}{c} \begin{array}{cc} TR & ABS \end{array} \\ \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{array}$$

Here **I** is an r-by-r indentity matrix, **0** is an r-by-t zero matrix, **R** is a nonzero t-by-r matrix, and **Q** is an t-by-t matrix. The first t states are transient and the last r states are absorbing.

A standard matrix algebra argument shows that $\mathbf{P}^n$ is of the form

$$\mathbf{P}^n = \begin{array}{c} \\ TR \\ ABS \end{array} \begin{array}{c} \begin{array}{cc} TR & ABS \end{array} \\ \begin{pmatrix} \mathbf{Q}^n & * \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{array}$$

where the asterisk $*$ stands for the t-by-r matrix in the upper right-hand corner of $\mathbf{P}^n$. (This submatrix can be written in terms of $\mathbf{Q}$ and $\mathbf{R}$, but the expression is complicated and is not needed at this time.) The form of $\mathbf{P}^n$ shows that the entries of $\mathbf{Q}^n$ give the probabilities for being in each of the transient states after n steps for each possible transient starting state.

Theorem (Probability of Absorption)

In an absorbing Markov chain, the probability that the process will be absorbed is 1 *(i.e.,* $\mathbf{Q}^n \to \mathbf{0}$ *as* $n \to \infty$*).*

Definition (The fundamental matrix)

For an absorbing Markov chain $\mathbf{P}$, the matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ is called the fundamental matrix for $\mathbf{P}$.

Theorem

The fundamental matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ *can be represented by*

$$\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots$$

and the ij-entry $n_{i,j}$ *of the matrix* $\mathbf{N}$ *is the expected number of times the chain is in state j (before absorption), given that it starts in state i. The initial state is counted if* $i = j$*.*

Example

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc|cc} 1 & 2 & 3 & 0 & 4 \end{array} \\ \left(\begin{array}{ccc|cc} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \end{array}.$$

$$\mathbf{Q} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix} \qquad \mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}$$

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array}$$

From the middle row of $\mathbf{N}$, we see that if we start in state 2, then the expected number of times in states 1, 2, and 3 before being absorbed are 1, 2, and 1. □

Time to Absorption

We now consider the question: Given that the chain starts in state i, what is the expected number of steps before the chain is absorbed? The answer is given in the next theorem.

Theorem

Let t_i be the expected number of steps before the chain is absorbed, given that the chain starts in state i, and let $\mathbf{T}$ be the column vector whose i-th entry is t_i. Then

$$\mathbf{T} = \mathbf{N} \cdot \mathbf{1},$$

where $\mathbf{1}$ is a column vector all of whose entries are 1.

Absorption Probabilities

Theorem

Let $b_{i,j}$ be the probability that an absorbing chain will be absorbed in the absorbing state j if it starts in the transient state i. Let $\mathbf{B}$ *be the matrix with entries $b_{i,j}$. Then* $\mathbf{B}$ *is an t-by-r matrix, and*

$$\mathbf{B} = \mathbf{N} \cdot \mathbf{R},$$

where $\mathbf{N}$ *is the fundamental matrix and* $\mathbf{R}$ *is as in the canonical form.*

Example continued....

$$\mathbf{N} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array} .$$

$$\text{T=} \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix} .$$

Thus, starting in states 1, 2, and 3, the expected times to absorption are 3, 4, and 3, respectively.

From the canonical form,

$$\mathbf{R} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 4 \end{array} \\ \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix} \end{array} .$$

Hence,

$$\mathbf{B} = \mathbf{NR} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$= \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 4 \end{array} \\ \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{array} .$$

Here the first row tells us that, starting from state 1, there is probability 3/4 of absorption in state 0 and 1/4 of absorption in state 4. □

Performability Indicators

Let consider a system that moves randomly among a set of states $E = \{1, 2, \ldots, d\}$. We will partition the state space in two subsets U and D:

$$E = U \cup D, \quad U \cap D = \emptyset, \quad U \neq \emptyset, \quad D \neq \emptyset.$$

If the system evolves according to a Markov chain we also need to partition the transition matrix $\mathbf{P}$ and the initial distribution vector α following U and D:

$$\mathbf{P} = \begin{array}{c} \\ U \\ D \end{array} \begin{array}{c} \begin{array}{cc} U & D \end{array} \\ \begin{pmatrix} \mathbf{P}_{UU} & \mathbf{P}_{UD} \\ \mathbf{P}_{DU} & \mathbf{P}_{DD} \end{pmatrix} \end{array}, \quad \alpha = [\alpha_U, \alpha_D]$$

Let us note $\mathbf{1}_{d,r} = (1, \ldots, 1, 0, \ldots, 0)^T$ the d-dimensional column vector whose r first elements are equal to 1 and the others are null. We note $\mathbf{1}_d = \mathbf{1}_{d,d,}$.

Availability function: is the probability that the system is up at moment n

$$A(n) = \mathbb{P}(X_n \in U) = \alpha \mathbf{P}^n \mathbf{1}_{d,r};$$

Reliability function: is the probability that the system is in a working state at time n and that it did not visit the failure states until time n. Let T be the lifetime of system, i.e.

$$T = \inf\{t \in \mathbb{N} : X(t) \in D\},$$

then:

$$R(n) = \mathbb{P}(\forall s \in [0, n], X_s \in U) = \mathbb{P}(T > n) = \alpha_U \mathbf{P}_{UU}^n \mathbf{1}_r;$$

Failure rate: is the conditional probability of failure of the system at moment n, given that it worked until the moment $n-1$:

$$\lambda(n) = \mathbb{P}(T = n \mid T \geq n) = \begin{cases} 1 - \frac{R(n)}{R(n-1)} & \text{if } R(n-1) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

Simulation of a MC

Simple procedure: depending on the initial rating i of the firm, the interval $[0, 1]$ is divided into subintervals according to the migration probabilities $p_{i,j}$ for $j = 1, \ldots, d$. For example, for each rating class i, the intervals are the following:

$$
\begin{aligned}
I_{1,i} &= [0, p_{i,1}) \\
I_{2,i} &= [p_{i,1}, p_{i,1} + p_{i,2}) \\
\ldots & \quad \ldots. \\
I_{j,i} &= [\textstyle\sum_{k=1}^{j-1} p_{i,k}, \sum_{k=1}^{j} p_{i,k}) \\
\ldots & \quad \ldots. \\
I_{K,i} &= [\textstyle\sum_{k=1}^{K-1} p_{i,k}, 1]
\end{aligned}
$$

Then a uniform distributed random variable u_t between 0 and 1 is drawn. Depending on which subinterval the random variable lies in, the company stays in the same rating class i or migrates to rating class j. The migration process for a company or loan in rating class i is determined by the following function $f : [0, 1] \to S$:

$$f_{s_i} = \begin{cases} S_1, & \text{for } u_t \in I_{1,i} \\ S_2 & \text{for } u_t \in I_{2,i} \\ \dots & \dots. \\ \dots & \dots. \\ S_K & \text{for } u_t \in I_{K,i} \end{cases}$$

If more than one time-period is considered and a migration to rating state j occurs, new subintervals have to be calculated based on the migration probabilities $p_{j,k}$ for $k = 1, \dots, d$ and a new random number u_{t+1} is drawn for the following period.

Nonparametric estimation of MC

Let $\mathbf{x}_{n+1} = (x_0, x_1, \ldots, x_L)$ be a realization of length $L+1$ of a MC. The likelihood function is:

$$\mathcal{L}(\theta; \mathbf{x}_{L+1}) := \mathbb{P}_\theta(X_0 = x_0, X_1 = x_1, \ldots, X_L = x_L) \qquad (2)$$

where θ is a value of the parameters of the model and $\theta \in \Theta$ and $\Theta = \{p_{i,j}, i, j \in E : 0 \leq p_{i,j} \leq 1, \sum_{j \in E} p_{i,j} = 1\}$. If we compare the likelihood function at two parameter points and find that

$$\mathbb{P}_{\theta_1}(\mathbf{X}_{L+1} = \mathbf{x}_{L+1}) = \mathcal{L}(\theta_1; \mathbf{x}_{L+1}) > \mathcal{L}(\theta_2; \mathbf{x}_{L+1}) = \mathbb{P}_{\theta_2}(\mathbf{X}_{L+1} = \mathbf{x}_{L+1}),$$

then the sample we actually observed is more likely to have occurred if $\theta = \theta_1$ than if $\theta = \theta_2$, which can be interpreted saying that θ_1 is a more plausible value for the true value of θ than is θ_2.

Definition

For each sample point **x**, let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $\mathcal{L}(\theta \mid \mathbf{x})$ attains its maximum as a function of θ, with **x** held fixed. A maximum likelihood estimator of the parameter θ based on a sample **X** is $\hat{\theta}(\mathbf{X})$.

Then we need to maximize the likelihood function:
- first order condition (set first order derivative (gradient) equal to zero);
- second order condition (second order derivative ≤ 0 (Hessian negative definite));
- often analytical results are not available $\Longrightarrow$ numerical methods;
- often it is advantageous to take the logarithm of the likelihood function.

The use of the Markov property gives

$$\begin{aligned}\mathcal{L}(\theta; \mathbf{x}_{L+1}) &= \mathbb{P}_\theta(X_0 = x_0, X_1 = x_1, \ldots, X_L = x_L) \\ &= v_{x_0}(0) \prod_{k=1}^{L} p_{x_{k-1},x_k} = v_{x_0}(0) \prod_{i,j \in E} p_{i,j}^{n_{i,j}}\end{aligned} \tag{3}$$

where $n_{i,j}$ is the frequency of the one-step transitions $i \to j$ in the sample $\mathbf{x}_{L+1}$.
Clearly the set of d^2 transition frequency $\{n_{i,j}\}_{i,j \in E}$ form a sufficient statistic for the transition matrix $\mathbf{P}$.
It is convenient to maximize the log-likelihood function:

$$\log(\mathcal{L}) = \log(v_{x_0}(0)) + \sum_{i,j \in E} n_{i,j} \log(p_{i,j}). \tag{4}$$

with respect to $p_{i,j}$'s subject to the restriction $\sum_{j \in E} p_{i,j} = 1$.

This yields the maximum likelihood estimators

$$\hat{p}_{i,j} = \frac{n_{i,j}}{n_i}, \tag{5}$$

where $n_i = \sum_{j \in E} n_{i,j}$.

In deriving the above estimator we have assumed that the initil probability distribution $\mathbf{v}(0)$ is non-informative about the transition probabilities $p_{i,j}$. In any case, for large n the effect of the term $\log(v_{x_0}(0))$ can be ignored.

Observation of $K \geq 2$ trajectories

Let assume that we observe $K \geq 2$ independent and identical realizations of the process, each one collected on the fixed time grid $\{0, 1, \ldots, L\}$

Let consider K samples $X_L^{(k)} = \{x^{(k)}(0), x^{(k)}(1), ..., x^{(k)}(L)\}$, observed up to time L, from an homogeneous and ergodic Markov chain $\{X(n)\}_{n \in \mathbb{N}}$ with unknown transition probability matrix $\mathbf{P}$ and state space E.

Let denote by
$n_{i,j}^{(k)}(L) = \sum_{u=1}^{L} 1_{\{X^{(k)}(u-1)=i, X^{(k)}(u)=j\}}$
number of transitions from i to j up to time L in the k-th sample.

$n_i^{(k)}(L) = \sum_{u=0}^{L-1} 1_{\{X^{(k)}(u)=i\}}$
number of visits to i up to time L in the k-th sample.

Due to the Markov hypothesis, the likelihood function assumes the following form:

$$\mathcal{L}(K,L) = \prod_{k=1}^{K} v_{X_0^{(k)}}(0) p_{X_0^{(k)},X_1^{(k)}} \cdots p_{X_{L-1}^{(k)},X_L^{(k)}}$$

$$\prod_{k=1}^{K} v_{X_0^{(k)}}(0) \left(\prod_{(i,j)\in E} \left(p_{(i,j)}\right)^{N_{i,j}^{(k)}(L)} \right)$$

By maximizing the loglikelihood function it is possible to derive the estimator

$$\hat{p}_{i,j}(K;L) = \frac{\sum_{k=1}^{K} n_{i,j}^{(k)}(L)}{\sum_{k=1}^{K} n_i^{(k)}(L)} \tag{6}$$

Theorem (see for example Billingsley ('61))

Let $\xi(L) = (\xi_{i,j}(L))_{i,j \in E}$ *the row random vector with elements*

$$\xi_{i,j}(L) = \frac{1}{\sqrt{N_i(L)}}(N_{i,j}(L) - N_i(L)p_{i,j}).$$

Then $\xi(L)$ *has asymptotically normal distribution, i.e.*

$$\xi(L) \xrightarrow{d} N(0, \Lambda) \quad as \quad L \longrightarrow \infty, \tag{7}$$

where the covariance matrix Λ *has as generic element*

$$\lambda((i,j),(k,l)) = \delta_{ik}(\delta_{jl}p_{i,j} - p_{i,j}p_{i,l}) \tag{8}$$

and δ_{ij} *is the Kronecker delta.*

Proposizione (Basawa and Prakasa Rao ('80))

Let $F(L) = (f_{i,j}(L))_{i,j \in E}$ *the row random vector with elements*

$$f_{i,j}(L) = \sqrt{L}(\hat{p}_{i,j}(L) - p_{i,j}).$$

Then $F(L)$ *has asymptotically normal distribution, i.e.*

$$F(L) \xrightarrow{d} \mathcal{N}(0, \Gamma) \quad as \quad L \longrightarrow \infty, \tag{9}$$

where the covariance matrix $\Gamma = \left(\sqrt{\Pi}\right)^{-1} \Lambda \left(\sqrt{\Pi}^{-1}\right)^{T}$ *and*

$$\sqrt{\Pi} = diag\{\sqrt{\pi_1}, \sqrt{\pi_1}, ..., \sqrt{\pi_1}, \sqrt{\pi_2}, \sqrt{\pi_2}, ..., \sqrt{\pi_2}, ..., \sqrt{\pi_d}, \sqrt{\pi_d}, ..., \sqrt{\pi_d}\} \tag{10}$$

Theorem (Anderson and Goodman ('57))

Let $F(K,L) = (f_{i,j}(K,L))_{i,j\in E}$ *the row random vector with elements*

$$f_{i,j}(K,L) = \sqrt{K}(\hat{p}_{i,j}(K,L) - p_{i,j}).$$

Then $F(K,L)$ *has asymptotically normal distribution, i.e.*

$$F(K,L) \xrightarrow{d} \mathcal{N}(0,G) \quad as \quad K \longrightarrow \infty, \tag{11}$$

where the covariance matrix G has as generic element

$$G((i,j),(k,l)) = Cov(f_{i,j}, f_{k,l}) = \frac{1}{\phi(i)}\left(\delta_{ik}(\delta_{jl}p_{i,j} - p_{i,j}p_{i,l})\right) \tag{12}$$

and $\phi(i) = \sum_{k\in E}\sum_{t=1}^{L}\eta_k p_{k,i}^{(t-1)}$ *assuming* $\frac{N_i(K,L)}{\sum_i N_i(K,L)} \to \eta_i$.

Practical consequence

Assume we are interested in estimating a differentiable function $\Phi(\{p_{i,j}\}_{i,j\in A})$ with respect to $p_{i,j}$ where $i,j \in A \subset E$. Then from the Multivariate Delta Mathod it results that:

$$\sqrt{L}\left(\Phi(\{\hat{p}_{i,j}(L)\}_{i,j\in A}) - \Phi(\{p_{i,j}\}_{i,j\in A})\right) \xrightarrow{d} \mathcal{N}(0, \Phi' \Gamma_A \Phi'^T)$$

where Φ' is the vector of first order derivatives $(\frac{\partial \Phi}{\partial p_{i,j}})_{i,j\in A}$.

Test for a specified transition matrix

Consider the hypothesis

$$H : p_{i,j} = t_{i,j}, \;\; i,j \in E, \;\; T = (t_{i,j}) \text{ stocastich matrix}$$

Consider the Kullback-Leibler divergence (relative entropy) for discriminating between $t_{i,j}$ and its MLE $p_{i,j} = \frac{n_{i,j}}{n_i}$ from a sample $\mathbf{x}_{n+1} = (x_0, x_1, \ldots, x_L)$. This is given by

$$D_{KL}(i) = \sum_{j \in E} t_{i,j} \log\left(\frac{t_{i,j}}{\hat{p}_{i,j}}\right) \approx \frac{1}{2} \sum_{j \in E} \frac{(n_{i,j} - n_i t_{i,j})^2}{n_i^2 t_{i,j}}.$$

A weighted average of D_{KL} measures for different $i \in E$ is

$$\sum_{i \in E} n_i D_{KL}(i) \approx \frac{1}{2} \sum_{i,j \in E} \frac{(n_{i,j} - n_i t_{i,j})^2}{n_i t_{i,j}} =: \frac{1}{2} \Phi.$$

The statistic $\sum_{j\in E} \frac{(n_{i,j}-n_i t_{i,j})^2}{n_i t_{i,j}}$ is the χ-squared goodness-of-fit statistic for the multinomial $n_{i,j}$ (n_i fixed) with cell probabilities $n_i t_{i,j}$ and $\sum_{j\in E} n_{i,j} = n_i$. Consequently, the statistic

$$\Phi = \sum_{i,j\in E} \frac{(n_{i,j} - n_i t_{i,j})^2}{n_i t_{i,j}}$$

is a sum of d independent χ-squared statistics corresponding to d rows of the transition matrix. Therefore we find that Φ is asymptotically distributed as χ-squared random variable with $d(d-1)$ dof. Here we have assumed that all the $p_{i,j}$ are positive.

More generally, suppose m_i is the number of positive entries in the ith row of **P** and let $M_i = \{j : p_{i,j} > 0\}$. Then

$$\Phi = \sum_{i,\in E} \sum_{j \in M_i} \frac{(n_{i,j} - n_i t_{i,j})^2}{n_i t_{i,j}}$$

has a χ-squared distribution with $M - d$ dof where M is the total number of positive entries in **P**.

Exercise

$$\mathbf{N} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 70 & 20 & 10 \\ 120 & 60 & 20 \\ 20 & 40 & 40 \end{pmatrix} \end{array} \qquad \mathbf{T} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0.50 & 0.25 & 0.25 \\ 0.60 & 0.10 & 0.30 \\ 0.30 & 0.30 & 0.40 \end{pmatrix} \end{array}$$

Test the hypothesis $H_0 : \mathbf{P} = \mathbf{T}$ at the significance level $\alpha = 0.05$.

What if

$$\mathbf{T} = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0.68 & 0.19 & 0.13 \\ 0.60 & 0.28 & 0.12 \\ 0.21 & 0.41 & 0.38 \end{pmatrix} \end{array}$$

Test for independence

The hypothesis of independence of the random variables $\{X_k\}$, $k = 0, 1, 2, \ldots$ within the stationary Markov dependence assumption takes the following form:

$$H_0 : p_{i,j} = \pi_j, \;\; i, j \in E,$$

where the stationary probabilities $\{\pi_j\}$ are unspecified. Under H_0, we have only d parameters $(\pi_1, \pi_2, \ldots, \pi_d)$ which are subject to a restriction, namely $\sum_{j \in E} \pi_j = 1$.

The MLE of π_j under the null hypothesis is $\hat{\pi}_j = \frac{\sum_{i \in E} n_{i,j}}{n}$.
The unrestricted MLE of $p_{i,j}$ are $\frac{n_{i,j}}{n_i}$.

The following statistic can be used to test the independence hypothesis mentioned before:

$$S = \sum_{i,j \in E} \frac{(n_{i,j} - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i}$$

We have that S has a limiting χ-squared distribution with $(d^2 - d) - (d - 1) = (d - 1)^2$ dof.

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \\ \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \end{array}$$

Exercise

From a sample generated by a 3-state Markov chain we calculated the following frequencies of transitions:

$$N = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 15 & 25 & 50 \\ 32 & 46 & 98 \\ 20 & 32 & 68 \end{pmatrix} \end{array}$$

Execute a test of independence of the random variables $\{X_k\}$, $k = 0, 1, 2, \ldots$ within the stationary Markov dependence assumption, i.e.

$$H_0 : p_{i,j} = \pi_j, \;\; i, j \in E,$$

at the significance level $\alpha = 0.05$.

Test for homogeneity in several samples

Suppose we have r independent samples from r Markov chains with transition probability matrices

$$\mathbf{P}(h) = (p_{i,j}(h))_{i,j \in E}, \quad h = 1, 2, \ldots, r.$$

Let the samples have transition frequencies

$$(n_{i,j}(h)), \quad h = 1, 2, \ldots, r.$$

Then the MLE of $p_{i,j}(h)$ is $\hat{p}_{i,j}(h) = \frac{n_{i,j}(h)}{n_i(h)}$.
Now let us formulate the null hypothesis

$$H_0 : p_{i,j}(1) = p_{i,j}(2) = \ldots = p_{i,j}(r), \quad \forall i, j \in E$$

The MLE of the common transition probability say $p_{i,j}$ is

$$\hat{p}_{i,j} = \frac{\sum_{k=1}^{r} n_{i,j}(h)}{\sum_{k=1}^{r} n_i(h)}.$$

Suppose the samples are of sizes N_1, N_2,...,N_r, where $N = \sum_{k=1}^{r} N_k$. Let $N \to \infty$ in such a way that $\frac{N_k}{N} \to \lambda_k > 0$, $(\sum_{k=1}^{r} \lambda_k = 1)$. Then the χ-squred statistic for testing H_0 is

$$R = \sum_{k=1}^{r} \sum_{i,j \in E} \frac{\left(n_{i,j}(h) - n_i(h)\hat{p}_{i,j}\right)^2}{n_i(h)\hat{p}_{i,j}}.$$

This has a limiting χ^2 distribution with $rd(d-1) - d(d-1) = (r-1)d(d-1)$ dof.

Exercise

From two samples we estimate two 2-states Markov chains obtaining the following frequencies of transitions:

$$N^{(1)} = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} 70 & 20 \\ 120 & 60 \end{pmatrix} \end{array}, \quad N^{(2)} = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} 110 & 30 \\ 175 & 85 \end{pmatrix} \end{array}$$

Execute a test of homogeneity of the two samples, i.e.

$$H_0 : p_{i,j}^{(1)} = p_{i,j}^{(2)}, \ \ i, j \in E,$$

at the significance level $\alpha = 0.05$.

The ratings data

Material from Bangia et al. (2002 JBF)

S&P's rating scale includes the rating modifiers $+/-$ as well as the default (D) and the "not rated" (NR) state.
The modifiers $+/-$ pose two problems:
1) the sample size of issuers per rating class including rating modifiers is not sufficient for low rating categories, causing small sample size concerns that affect statistical inference
2) transition matrices are generally published and applied without modifiers $+/-$, as this format has emerged as an industry standard. Therefore, we exclude the modifiers $+/-$.
Consequently, for example, we consider BBB+ and BBB- ratings as BBB ratings. This methodology reduces the rating categories and ensures sufficient sample sizes for all ratings.

Selection of samples

We estimate both conditional and unconditional ratings transition matrices. By conditional, we refer to conditioning on the stage of the business cycle, expansion or contraction. By unconditional, we refer to averaging across stages of the cycle. We enforce three criteria in the estimation of transition matrices:
1) because our interest is partly in unconditional transition probabilities, the sample period should encompass both expansions and contractions;
2) a minimal sample size has to be met at all times to ensure statistical reliability of the estimates;
3) the time-t transition matrix should reflect the time-t rating universe, as opposed to some outdated sample that is no longer representative.

It might be desirable to estimate transition matrices based on a constant sample over a given period of time. We could, for example, track the 1981 universe from 1981 through 1998, leaving aside new firms that received their first rating after the beginning of 1981. This approach suffers from several problems. First, any given cohort quickly becomes outdated and hence less interesting as new issuers emerge, mergers and acquisitions transpire, and some industries decline while others flourish. Second, the fundamental characteristics of the underlying firms would evolve over time, producing results of dubious interpretation. Third, the statistical size of a cohort would fall below the threshold level as issuers perish, default or retire their rating over time, for example by calling their outstanding debt.

We could attempt to mitigate the problems associated with tracking a fixed cohort simply by using a recently formed cohort. The resulting estimates, however, would reflect only the current economic situation, which might be purely expansion or purely contraction.
Alternatively, we could allow the sample composition to vary over time, incorporating new issuers and discarding those who default. We could, for example, use all issuers outstanding as of January 1, 1981 to estimate the 1981 transition matrix, all issuers outstanding as of January 1, 1982 to estimate the 1982 transition matrix, and so on. This procedure helps ensure that the sample size is always large enough to facilitate sharp statistical inference, that new firms are included in the sample, and that average transition matrices incorporate all states of the economy. We shall use this method.

Treatment of transitions to not rated status

Before proceeding to estimate transition matrices, we must deal with transitions to NR status. Transitions to NR may be due to any of several reasons, including expiration of the debt, calling of the debt, failure to pay the requisite fee to S&P, etc. Unfortunately, however, the details of individual transitions to NR are not known. In particular, it is not known whether any given transition to NR is "benign" or "bad". Bad transitions to NR occur, for example, when a deterioration of credit quality known only to the bond issuer (debtor) leads the issuer to decide to bypass an agency rating.

There are at least three methods for removing NR's from the dataset. The first method is conservative and proceeds by treating transitions to NR as negative information regarding the change in credit quality of the borrower. Here the probability of transitioning to NR is distributed amongst downgraded and defaulted states in proportion to their values by allocating NR values to all cells to the right of diagonal. The second method is liberal and treats transitions to NR status as benign. The probability transitions to NR are distributed among all states, except default, in proportion to their values. This is achieved by allocating the probability of transiting to NR to all but the default column.

The third method, which has emerged as an industry standard, treats transitions to NR status as non-information. The probability of transitions to NR is distributed among all states in proportion to their values. This is achieved by gradually eliminating companies whose ratings are withdrawn. We use this method, which appears sensible and allows for easy comparisons to other studies.
Since reasons for an NR assessment may be both positive and negative, a splitting of the NR state into two new states denoted by NR1 and NR2 is recommended. An instrument is rated NR1 when it enters state NR coming from an investment grade (AAA to BBB). In contrast, an instrument is rated NR2 when it enters state NR coming from a speculative grade (BB to CCC).

Transition horizon

Theoretically, transition matrices can be estimated for any desired transition horizon. Transition matrices estimated over short time periods best reflect the rating process. The shorter the measurement interval, the fewer rating changes are omitted. However, shorter duration also results in less extreme movements, as large movements are often achieved via some intermediary steps. The other factor determining the transition horizon is the application purpose. For the calculation of credit risk exposures by portfolio models, a oneyear transition horizon is standard. Other applications such as the pricing of credit derivatives require shorter horizons; however in practice only annual transition matrices are typically used.
Matrices estimated over longer time periods offer the advantage of less noise inherent in the data, as short-term noise cancels itself out for longer horizons.

Unconditional estimates of **P** Bangia et al. (2002 JBF)

Table 1
Unconditional quarterly migration matrix

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	97.92%	1.95%	0.10%	0.02%	0.01%	–	–	–
AA	0.16%	97.95%	1.75%	0.10%	0.01%	0.02%	0.00%	–
A	0.02%	0.57%	97.91%	1.34%	0.10%	0.06%	0.00%	0.00%
BBB	0.01%	0.07%	1.37%	96.90%	1.38%	0.23%	0.02%	0.03%
BB	0.01%	0.03%	0.17%	1.87%	95.35%	2.26%	0.18%	0.13%
B	–	0.02%	0.07%	0.11%	1.66%	95.72%	1.46%	0.96%
CCC	0.04%	–	0.16%	0.20%	0.41%	3.28%	87.18%	8.72%
							Average default rate	0.284%

Table 2
Unconditional annual migration matrix

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	91.93%	7.46%	0.48%	0.08%	0.04%	–	–	–
AA	0.64%	91.81%	6.75%	0.10%	0.06%	0.12%	0.03%	–
A	0.07%	2.27%	91.69%	5.11%	0.56%	0.25%	0.01%	0.04%
BBB	0.04%	0.27%	5.56%	87.88%	4.83%	1.02%	0.17%	0.24%
BB	0.04%	0.10%	0.61%	7.75%	81.48%	7.89%	1.11%	1.01%
B	–	0.10%	0.28%	0.46%	6.95%	82.80%	3.96%	5.45%
CCC	0.19%	–	0.37%	0.75%	2.43%	12.13%	60.45%	23.69%
							Average default rate	1.215%