Bootstrapping Markov Chains

Adam Kubetta

Presentation at the seminar
Stochastic modelling in Economics and Finance

17th December 2012
Introduction

In the presentation we will use the bootstrap technique to estimate the distribution of the following estimates concerning the discrete time Markov chains:

- the estimate of transition matrix,
- the estimate of stationary distribution,
- the estimate of hitting times distribution,

and for continuous time Markov chains we will be estimating the distribution of

- the estimate of intensity matrix.

Note: We will work only with time homogeneous Markov chains with finite state space $S = \{1, \ldots, s\}$. 
Discrete time Markov chains

Assumptions & the task

Let's suppose we have given a sample \( x = \{x_1, \ldots, x_n\} \) from a Markov chain \( \{X_t\}_{t=0,1,2,\ldots} \), where \( X_t \in S = \{1, \ldots, s\} \) for all \( t = 0, 1, 2, \ldots \), with an unknown transition matrix \( A \) and an initial distribution \( p_0 = p_{t=0} \). Further we assume that the underlying chain is ergodic (positively recurrent, unseparable and aperiodic).

\[
X = \text{RandomMarkovChain}\left[A = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/4 & 3/4 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}, p_0 = 1, n = 10000 \right];
\]

\[
X[1;;10]
\]

\( \{2, 2, 2, 3, 2, 1, 2, 3, 3, 3\} \)

At first, we look at the transition matrix of underlying chain with “canonically ordered states” and analyze the periodicity in individual inseparable components of states to assure that the given sample meets our requirements.
ShowCanonicallyOrderedTransitionMatrix[A, 
  StateNames → {"1", "2", "3", "4"}]
StateClassification[A]
StatePeriods[A] /. {1 → "Aperiodic", ∞ → "Transient"}

{{1, 2, 3, 4}, {}}
{Aperiodic}
Our first goal will be to propose procedures that estimates the following quantities from the given sample $x$.

1) the transition matrix $A$, 
   i.e. the matrix $A_{i,j} = P[X_{t+1} = j \mid X_t = i],$

2) the stationary distribution $\pi$,
   i.e. the distribution on the state space $S$ satisfying $\pi^T = A \pi^T$,

3) the distribution of hitting times $\tau_{i,j}$,
   i.e. the probabilities $\tau_{(i,j),t} = P[\tau_{i,j} = t]$ on time axis $t = 1, 2, \ldots,$
   where $\tau_{i,j}$ is the time of the first enter to $j$ under the condition $P[p_0 = i] = 1.$

After that we will use bootstrapping to obtain the estimate of distribution of all estimated values.
Estimates

The maximum likelihood estimate of $A_{i,j}$ is

$$\hat{A}_{i,j} = \begin{cases} 
\frac{n_{i,j}}{n_i} & \text{for } n_i > 0 \\
0 & \text{for } n_i = 0
\end{cases},$$

where $n_i$ is the number of visits of the state $i$ in the sample $\pi$ and $n_{i,j}$ is number of transitions $i \rightarrow j$ in $\pi$.

It can be shown that assymptotically for $n \rightarrow \infty$

$$\sqrt{n} \left( \hat{A} - A \right) \sim N(0, \Sigma_A),$$

where $\Sigma_A$ is the $(S \times S) \times (S \times S)$ covariance matrix with elements

$$\Sigma_{A,(i,j),(i',j')} = \delta_{i,i'} \left( \delta_{j,j'} A_{i,j} \left( 1 - A_{i,j} \right) - \left( 1 - \delta_{j,j'} \right) A_{i,j} A_{i,j} \right) / \pi_i .$$
Proove

The vectors \( \{n_{i,1}, \ldots, n_{i,s}\} \) for each row \( i = 1, \ldots, s \) has multinomial distribution with parameters \( n_i \) and \( i^{th} \) row of transition matrix, i.e. \( \{ \mathbf{A}_{i,1}, \ldots, \mathbf{A}_{i,s} \} \). Therefore, applying central limit theorem we have for each \( i \)

\[
\sqrt{n_i} \left( \{\hat{\mathbf{A}}_{i,1}, \ldots, \hat{\mathbf{A}}_{i,s}\} - \{\mathbf{A}_{i,1}, \ldots, \mathbf{A}_{i,s}\} \right) \xrightarrow{D}_{n_i \to \infty} \mathcal{N}[\mathbf{0}, \Sigma_i],
\]

where

\[
\Sigma_i = \begin{pmatrix}
\mathbf{A}_{i,1}(1 - \mathbf{A}_{i,1}) & -\mathbf{A}_{i,1} \mathbf{A}_{i,2} & \cdots & -\mathbf{A}_{i,1} \mathbf{A}_{i,s} \\
-\mathbf{A}_{i,2} \mathbf{A}_{i,1} & \mathbf{A}_{i,2}(1 - \mathbf{A}_{i,2}) & \cdots & -\mathbf{A}_{i,2} \mathbf{A}_{i,s} \\
\vdots & \vdots & \ddots & \vdots \\
-\mathbf{A}_{i,s} \mathbf{A}_{i,1} & -\mathbf{A}_{i,s} \mathbf{A}_{i,2} & \cdots & \mathbf{A}_{i,s}(1 - \mathbf{A}_{i,s})
\end{pmatrix}.
\]

For number of visits the following convergence holds

\[
\{n_1, \ldots, n_s\} \xrightarrow{\mathcal{P}} \pi,
\]

where \( \pi \) is the only stationary distribution of the chain, because we have assumed its ergodicity and unseparability of its states.
Thus combining these statements we have for all \( i = 1, \ldots, s \)

\[
\sqrt{n} \left( \{\hat{A}_{i,1}, \ldots, \hat{A}_{i,s}\} - \{A_{i,1}, \ldots, A_{i,s}\} \right) \xrightarrow{\mathcal{D}_{n \to \infty}} N(0, \Sigma_i / \pi_i).
\]

Since all rows are independent we can express the covariance between two elements \((i, j)\) and \((i', j')\) of the estimate as

\[
\Sigma_{A_i(i,j), (i', j')} = \delta_{i,i'} \left( \delta_{j,j'} A_{i,j} \left( 1 - A_{i,j} \right) - \left( 1 - \delta_{j,j'} \right) A_{i,j} A_{i,j'} \right) / \pi_i.
\]
\[ \hat{A} = \text{TransitionMatrixEstimate}[\mathbb{x}] \];

\[
\hat{A} = \begin{pmatrix}
0.5 & 0.5 & 0. & 0. \\
0.25 & 0.25 & 0.25 & 0.25 \\
0. & 0.25 & 0.75 & 0. \\
0.333333 & 0.333333 & 0.333333 & 0. \\
\end{pmatrix}
\]

\[
\hat{A} = \begin{pmatrix}
0.508372 & 0.491628 & 0. & 0. \\
0.254167 & 0.261538 & 0.242308 & 0.241987 \\
0. & 0.246603 & 0.753397 & 0. \\
0.350993 & 0.352318 & 0.296689 & 0. \\
\end{pmatrix}
\]
The elements of the stationary distribution we estimate as

\[ \hat{\pi}_i = \frac{n_i}{n}. \]

Unprotect[\(\pi\)]; \(\pi = \text{"\(\pi\)";}\)
StationaryDistribution[A] // First // N
\(\hat{\pi} = \text{StationaryDistributionEstimate}[\(\pi\)] // N\)

\{0.205128, 0.307692, 0.410256, 0.0769231\}

\{0.2151, 0.312, 0.3974, 0.0755\}
The hitting times we will estimate in two ways.

One method requires to use the estimated matrix $\hat{A}$. In order to estimate the probability $\tau_{i,j}(1,\ldots,t) = P[\tau_{i,j} \leq t]$ we plug the $j^{th}$ cannonical (unit) vector $e_j$ to the matrix $\hat{A}$ instead of its $j^{th}$ row (to make the state $j$ absorbing) and then look at the $(i, j)^{th}$ position of the $t^{th}$ power of the refined matrix $j\hat{A}$. Having defined the adjusted matrix we can write

$$\hat{\tau}_{i,j}(1,\ldots,t) = (j\hat{A})_{i,j}^t$$

\textbf{HittingTimeEstimate}[$\mathbb{R}$, 1 $\rightarrow$ 4, $t$ = 3] // N

0.210562
The second approach is more straightforward, does not require estimating the transition matrix first and estimates the distribution of $\tau_{i,j}$ on the whole timeline $\{1, \ldots, t\}$. The procedure consists of the following steps.

1) Select all visits of the state $i$ in the sample $x$ (together with their corresponding time),
2) For each selected state $i$ determine the length of the period up to the first visit of the state $j$,
3) The relative count of the periods of length $t$ gives the estimate $\hat{\tau}_{i,j,t}$ of the probability

$$
\tau_{i,j,t} = P[\tau_{i,j} = t].
$$

$$
\hat{\tau}_{1\rightarrow 4} = \text{HittingTimeEstimate}[x, 1 \rightarrow 4, \text{Out} \rightarrow "PDF"];
\hat{\tau}_{1\rightarrow 4, \text{CDF}} = \text{HittingTimeEstimate}[x, 1 \rightarrow 4, \text{Out} \rightarrow "CDF"];
$$
The following picture shows the estimated density (with respect to the counting measure on $t = 0, 1, 2, ...$).

```
Show[ListPlot[\[Hat]1_{4}, PlotRange \rightarrow \text{All}, AxesOrigin \rightarrow \{0, 0\}],
     Graphics[{PointSize[0.02], Darker[Blue],
               Point[\{ExpectedHittingTime[x, 1 \rightarrow 4], 0\}]})]
```

Printed by Mathematica for Students
The next picture compares the difference between distribution functions calculated using both methods, the first that uses \( \hat{A} \) and the second that does not use \( \hat{A} \).

```math
ListPlot[
{#, HittingTimeEstimate[x, 1 \rightarrow 4, #] // N} & /@ Range[\( \hat{\tau}_{1 \rightarrow 4, \text{CDF}[\{1, 1\}] \)], 
\( \hat{\tau}_{1 \rightarrow 4, \text{CDF}} \) // N}, PlotStyle -> {Blue, Red}]
```
Bootstraping

The idea of bootstraping is to generate further (bootstrap) samples using the (originally) estimated quantity to obtain the distribution of the (original) estimate. Some useful characteristics of the (original) estimate (like variance) can be derived from this distribution.

The method is justified by the fact that the distribution of the bootstrap estimate has (often) assymptotically the same attributes as the distribution of the common (original) estimate.

Note that when estimating from bootstrap samples we use the same procedures defined earlier.

The result of the following code shows the bootstrap samples together with their componentwise (sample) mean.
bootstrapSample = RandomMarkovChain[\(A\), \(p_0\), \(N_n = 1000\)] & /@ Range[\(R = 400\)];

\(\text{bootstrapSample} \// \text{Dimensions;}
\)

\(\bar{A} = \text{TransitionMatrixEstimate} /@ \text{bootstrapSample;}
\)

Mean[\(\bar{A}\)] // N // MatrixForm

\[
\begin{bmatrix}
0.504453 & 0.495547 & 0. & 0. \\
0.253778 & 0.259523 & 0.243617 & 0.243083 \\
0. & 0.248049 & 0.751951 & 0. \\
0.350793 & 0.356503 & 0.292704 & 0.
\end{bmatrix}
\]
ShowMatrixSamples[\[\bar{A}\], ImageSize -> 300, AxesLabel -> None
(*, ColorFunction -> "DarkRainbow", FillingStyle -> Opacity[0.1],
PlotStyle -> Opacity[0.6]*)]
Now we verify, whether the estimated sample (componentwise) variances correspond to the theoretical variances mentioned in the previous subsection.

\[
\text{Map}[\text{Variance}[#] \&, \text{Transpose}[\tilde{A}, \{3, 1, 2\}], \{2\}] \text{ // N} \text{ // MatrixForm}
\]

\[
\frac{1}{\text{A (1 - A) / First@StationaryDistribution[A]}} \text{ // N} \text{ // MatrixForm}
\]

\[
\begin{pmatrix}
0.00126758 & 0.00126758 & 0. & 0. \\
0.00061811 & 0.000614036 & 0.000632537 & 0.000594325 \\
0. & 0.000481738 & 0.000481738 & 0. \\
0.00304564 & 0.00317443 & 0.00282972 & 0.
\end{pmatrix}
\]

\[
\begin{pmatrix}
0.00121875 & 0.00121875 & 0. & 0. \\
0.000609375 & 0.000609375 & 0.000609375 & 0.000609375 \\
0. & 0.000457031 & 0.000457031 & 0. \\
0.00288889 & 0.00288889 & 0.00288889 & 0.
\end{pmatrix}
\]
At last, we bootstrap the stationary distribution and hitting times for fixed time $t$.

\[
\tilde{\pi} = \text{StationaryDistributionEstimate} \, @ \, \text{bootstrapSample} \, // \, \text{Transpose};
\]

\[
\tilde{\pi} \, // \, \text{ListPlot}
\]
StationaryDistribution[A] // N
Mean /@ ð // N
StandardDeviation /@ ð // N

{{0.205128, 0.307692, 0.410256, 0.0769231}}
{0.202818, 0.3111, 0.408925, 0.0771575}
{0.0224254, 0.0156821, 0.033726, 0.00772476}

ð = HittingTimeEstimate[#, 1 → 2, 2] & /@ bootstrapSample;
ð // N // Mean
ð // N // StandardDeviation

0.743734

0.0357162

The theoretical assymptotic properties of the last results are not studied in the underlying article.
Continuous time Markov chains

Quit[]
(* we need the constant π *)

Let's suppose we have given a sample \( z(t) \) for times \( t \in [0, T] \) from time continuous Markov chain (Markov jump process) with an unknown intensity matrix \( A \) and an initial distribution \( p_0 \).

In the first case we assume that the chain’s states are unseparable and reccurent.

\[
A = \begin{pmatrix}
-0.1 & 0.1 & 0 \\
0 & -0.1 & 0.1 \\
0.2 & 0 & -0.2
\end{pmatrix}, \quad p_0 = 1, \quad T = 10000
\]
Before proceeding the calculation we analyze the periodicity of embedded (discrete time) chain.

\[
\text{StatePeriods@ToEmbeddedChainTransitionMatrix@}
\begin{pmatrix}
-0.1 & 0.1 & 0 \\
0 & -0.1 & 0.1 \\
0.2 & 0 & -0.2 \\
\end{pmatrix}
\]

\text{Manipulate[Plot[Evaluate[\[ScriptZ][t]], \{t, 0, T\}, Exclusions \rightarrow \[ScriptZ][1,1,All,2,1]], \{T, 10, 1000\}, Paneled \rightarrow False]}

\{3\}
Estimates

The maximum likelihood estimate of $A_{i,j}$ is

$$
\hat{A}_{i,j} = \begin{cases} 
\frac{n_{i,j}}{\psi(i)} & \text{for } \psi(i) > 0 \\
0 & \text{for } \psi(i) = 0
\end{cases},
$$

where $\psi(i)$ is the total time spent in the state $i$ (more precisely $\psi(i) = \int_0^T \mathbb{1}_{z(t) = i} \, dt$) and $n_{i,j}$ is number of transitions $i \rightarrow j$ in the given sample $z(t)$.

\[
\hat{A} = \text{IntensityMatrixEstimate}[z]; \hat{A} // \text{MatrixForm}
\]

\[
\begin{pmatrix}
-0.0970115 & 0.0970115 & 0. \\
0. & -0.103393 & 0.103393 \\
0.191492 & 0. & -0.191492
\end{pmatrix}
\]
Bootstraping

The following code gives the bootstrap samples together with their componentwise (sample) mean.

```mathematica
bootstrapSample = RandomMarkovChain[\( \hat{A}, p_0, 1000 \) ] & /@ Range[R = 400];
bootstrapSample // Dimensions;
\[ \hat{A} = \text{IntensityMatrixEstimate} /@ \text{bootstrapSample}; \]
Mean[\[ \hat{A} \] ] // N // MatrixForm

\[
\begin{pmatrix}
-0.0983608 & 0.0983608 & 0. \\
0. & -0.104655 & 0.104655 \\
0.196897 & 0. & -0.196897
\end{pmatrix}
\]
```
ShowMatrixSamples[Ê, ImageSize → 300, AxesLabel → None, PlotRange → {-0.5, 0.5}]
We can further use standard Mathematica functions to get e.g. kernel estimate of the distribution estimate of the \((i, j)^{th}\) element of the (original) estimate.

\[
\hat{\theta}_{[A1,1,1]} \quad \text{// ListPlot}
\]

\[
dists = \text{Map}\left[\text{If}\left[\text{Equal}@\hat{\theta}_{[A1,1,1],\#1,\#2}\right],
\text{EmpiricalDistribution}\left[\text{List}@\text{First}@\hat{\theta}_{[A1,1,1],\#1,\#2}\right],
\text{SmoothKernelDistribution}\left[\hat{\theta}_{[A1,1,1],\#1,\#2}\right]\right] \&,
\text{Outer}\left[\text{List}, \text{Range}@\text{Last}@\text{Dimensions}@\hat{\theta}, \text{Range}@\text{Last}@\text{Dimensions}@\hat{\theta}\right],
\{2\}\right];
\]
Map[Plot[CDF[#, x], {x, -0.5, 0.5}, Filling -> Axis, PlotRange -> All] &, dists, {2}] // MatrixForm
# Probabilities of default

When estimating the intensity matrix determining the transition probabilities between rating categories, we usually assume that the state associated with default is absorbing. This does not comply with our assumption of irreducibility and therefore our previous estimate can not be used due to insufficient number of transitions in one sample.

Instead of using one sample $z(t)$ we use a set of samples $\{z_1(t), ..., z_m(t)\}$, which corresponds to the group of companies with the same default intensity matrix.

We use the same function as before, but this time its argument is extended to the set of functions.

```mathematica
sample = RandomMarkovChain[A = {{-0.1, 0.1, 0}, {0, -0.1, 0.1}, {0, 0, 0}}, x0 = {1, 1000, 100}][t];
IntensityMatrixEstimate[sample] // MatrixForm
```

\[
\begin{pmatrix}
-0.0908639 & 0.0908639 & 0 \\
0 & -0.103354 & 0.103354 \\
0 & 0 & 0
\end{pmatrix}
\]
When considering only two states, defaulted / running company, we can estimate the probability of default from the set of 0/1 samples simply as the ratio of zeros, and confidence interval for this estimate can be expressed implicitly using the binomial distribution as follows.

If $X_i \sim \text{Alt}(p)$, then $X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(p, n)$ and we want to find $p_{\min}$ and $p_{\max}$ such that

$$P[X \leq x - 1 \mid X \sim \text{Binomial}(p_{\min}, n)] = 1 - \alpha / 2,$$

$$P[X \leq x \mid X \sim \text{Binomial}(p_{\max}, n)] = \alpha / 2,$$

for chosen confidence level $\alpha$. However, if we observe $x = 0$, then the first equality is not well defined. In this case we can, of course, set $p_{\min} = 0$ and $p_{\max}$ express explicitly (in a closed form) as the solution of

$$P[X = 0 \mid X \sim \text{Binomial}(p_{\max}, n)] = 1 - P[X > 0 \mid X \sim \text{Binomial}(p_{\max}, n)] = \alpha,$$

which is

$$p_{\max} = 1 - \alpha^{1/n}.$$
DefaultProbabilityCI[0, 200, 0.01]

{0., 0., 0.0227628}
Literature

K.B. Athreya, C.D. Fuh:
Bootstrapping Markov chains: countable case.
Journal of Statistical Planning and Inference, Volume 33, Issue 3, December 1992, Pages 311–331

Jens H.E. Christensen, Ernst Hansen, David Lando:
Confidence sets for continuous–time rating transition probabilities.

S. Datta, W.P. McCormick:
Bootstrap for a finite state Markov chain based on i.i.d. resampling
in Exploring the Limits of Bootstrap. Wiley, New York, 1992