sargent

Spencer Lyon

July 10, 2013

Contents

1 RTM4 **2**

1.1 Chapter 2 . . . 2

1.1.1 Markov Chains . . . 2

Definitions . . . 2

Algorithms and Applications . . . 4

Theorems . . . 4

1.1.2 Continuous-state Markov chain . . . 5

Definitions . . . 5

Theorems . . . 6

1.1.3 Stochastic linear difference equations . . . 6

Definitions . . . 6

Algorithms and Applications . . . 7

1.1.4 Population . . . 8

Algorithms and Applications . . . 8

1.1.5 Estimation of Model Parameters . . . 9

Algorithms and Applications . . . 9

1.1.6 The Kalman filter . . . 9

Definitions . . . 9

Algorithms and Applications . . . 9

Theorems . . . 9

2 TODO Marching Orders, number 1 [0/4] **9**

2.1 **TODO** Read Chapter 2 of RMT4 . . . 10

2.2 **TODO** Read the two techincal appendixes . . . 10

2.3 **TODO** Work exercises 2.1-2.5 . . . 10

2.4 **TODO** Think of python examples **:TOM:** 10

1 RTM4

1.1 Chapter 2

1.1.1 Markov Chains

Definitions

- The **Markov Property** is that the conditional probability $P(x_t|x_{t-1}, x_{t-2}, \ldots) = P(x_t|x_{t-1})$
- A matrix is **stochastic** if two things hold: 1. $P_{ij} >= 0 \forall i, j$ and 2. $\sum_j P_{ij} = 1 \forall i$
- A **Markov Chain** is made up of 3 things: 1. the standard basis vectors e_i, 2. transition matrix P, where P_{ij} that defines the probability of moving from state i to state j.
- The probability of moving from state i to state j in k periods is $P_{ij}^{(k)}$. Note that the parenthesis are there to disambiguate what object the power k is operating on. I mean it to say the ij element of the matrix P raised to the k power, not the ij element of P raised to the k power.
- An **unconditional probability** is a $(1xn)$ vector where element i is the probability that $x_t = x_i$. It is indexed over time and evolves by multiplying on the right by P. ($\pi_t = \pi_{t-1}P$)
- A distribution is **stationary** if $\pi_{t+1} = \pi_t$, or in other words $\pi' = \pi' P \to (I - P')\pi = 0$. This makes π an eigenvector of P' corresponding to a unit eigenvalue. (Normally we have $(I\lambda - P')$, but in this case $\lambda = 1$). For all stochastic matrices P, there exists at least one stationary distribution. This distribution is unique iff there is only one unit eigenvalue.
- An **absorbing state** is a state in a stationary distribution that has a probability of 1. This means there can only be one absorbing state in every stationary distribution. An **absorbing subset** is a subset of the entire state that takes up the entire probability. Again, there can only be one absorbing subset, but this subset can contain many items.

- A process is **asymptotically stationary** if, in the limit as $t \to \infty$ the unconditional distribution $\pi_t \to \pi_\infty$. If this π_∞ is the same regardless of the starting value π_0, then the process is said to be asymptotically stationary with a unique invariant distribution.

- The **law of iterated expectations** is illustrated in the fact that when $y_{t+1} = \bar{y}'P'x_t + \bar{y}'v_{t+1}$ (x is state, y is a function of the state, P is transition matrix, v is random shock with $E[v] = 0$) we have that $E[E[Y_{t+1}|x_{t+1}]|x_t] = E[y_{t+1}|x_t]$. More formally, this law says that for any random variable z and two information sets J, I with $j \subset I$, $E[E(z|I)]|J] = E(z|J)

  - Example:

$$Ey_1 = \sum_j \pi_{1,j}\bar{y}_j = \pi_1'\bar{y} - (\pi_0'P)\bar{y} = \pi_0'(P\bar{y})E[E(y_1|x_0 = e_i)] = \sum_i \pi_{0,i} \sum_j P_{ij}\bar{y}_j = \sum_j (\pi_{0,i}P_{ij}$$

- A random variable is said to be **invariant** if it $y_t = y_0, t \geq 0$, for all realizations of $x_t, t \geq 0$ that occur with positive probability under (P, π). In other words, the random variable y_t remains constant at y_0, even while the underlying state x_t moves through the state space X.

- Any stochastic process y_t that follows the rule $E[y_{t+1}|x_t] = y_t$ is said to be a **martingale**.

- A stationary Markov chain is said to be **ergodic** if the only invariant functions $\bar{y}$ are constant with probability 1 under the stationary unconditional probability distribution π, i.e. $\bar{y}_i = \bar{y}_j \forall i, j$ with $\pi_i > 0$ and $\pi_j > 0$. Another definition I keep seeing online is that a Markov chain is ergodib or irreducible if it is possible to eventually get from every state to every other state with positive probability.

- The **likelihood** for a stochastic matrix following the Markov property can be written as follows

$$L = \pi_{0,i_0} \prod_i \prod_j P_{i,j}^{n_{ij}}$$

  where n_{ij} is the number of times a one period transition from state i to state j occurs. This function is classified as a *multinomial* distribution.

Algorithms and Applications

- Finding P from one step ahead conditional expectations
  If the transition matrix P is unknown, but you can determine conditional expectations of n independent functions (i.e., n linearly independent vectors $h_1, \ldots, h_n$), you will have uniquely identified P.

- Eigenvalues and left/right eigenvectors of stochastic matrices
  The unit eigenvalues of P have left-eigenvectors that are the stationary distributions of the chain [$(I - P')\pi = 0$] and right eigenvectors that are invariant functions of the chain [$(P - I)\bar{y} = 0$]

  The left eigenvectors of a matrix A are found by solving $(I - A')x = 0$ for x. They are the exact same as the right eigenvectors of A' (right eigenvectors are what I have been finding my whole life)

- Finding stationary distributions
  To find the stationary distributions of a Markov chain P, simply find the left-eigenvectors (right eigenvectors of P') and normalize so that is sums to 1 (find it in terms of simple whole numbers and divide the vector by its sum). To find the invariant functions just find the right eigenvectors of P and then you can multiply out front by any scalar because they don't have to be normalized to 1.

- Markov chain parameter estimation
  Estimation for free parameters θ of a Markov process: Let the transition matrix P and the initial distribution π_0 be functions $P(\theta), \pi_0(\theta)$ of a vector of free parameters θ. Given a sample $\{x_t\}_{t=0}^{T}$, regard the likelihood function as a function of the parameters θ. As the estimator of θ, choose the value that maximizes the likelihood function L (just a very verbose way of saying to MLE).

Theorems

- Let $\bar{y}$ define a random variable as a function of an underlying state x, where x is governed by a stationary Markov chian (P, π). Then

$$\frac{1}{T}\sum_{t=1}^{T} y_t \to E[y_\infty | x_0]$$

  with probability 1.

- Let (P, π) be a stationary Markov chain. If

$$E[y_{t+1}|X + t] = y_t$$

then the random variable $y_t = \bar{y}' x_t$ is invariant.

- Let $\bar{y}$ define a random variable on a stationary and ergodic Markov chain (P, π). Then

$$\frac{1}{T} \sum_{t=1}^{T} y_t \to E[y_0]$$

with probability 1. In other words, the time series average converges t the population mean of the stationary distribution.

1.1.2 Continuous-state Markov chain

Definitions

- **State transitions** are defined the by the cumulative distribution function (cdf)

$$\Pi(s'|s) = Prob(s_{t+1} \geq s'|s_t = s)$$

The initial state s_0 is given by the cdf

$$\Pi_0(s) = Prob(s_0 \leq s)$$

- The **transition density** is

$$\pi(s'|s) = \frac{\partial}{\partial s'} \Pi(s'|s)$$

and the initial density is

$$\pi_0(s) = \frac{\partial}{\partial s} \Pi_0(s)$$

- A **history** is given the notation $s^t = [s_t, s_{t-1}, \ldots, s_0]$ and is just a vector of the value of a variable over time.

- A Markov chain is **stationary** if π_0 satisfies $(\forall s \in S)$

$$\pi_0(s') = \int_s \pi(s'|s) \pi_0(s) ds$$

- A function ϕ of a Markov chain is invariant if

$$\int \phi(s') \pi(s'|s) ds' = \phi(s)$$

Theorems

- Let $y(s)$ be a random variable, a measurable function of s, and let $(\pi(s'|s), \pi_0(s))$ be a stationary and ergodic continuous-state Markov process. Assume that $E|y| < +\infty$. Then

$$\frac{1}{T}\sum_{t=1}^{T} y_t = Ey = \int y(s)\pi_0(s)ds$$

  with probability 1 with respect to the distribution π_0.

1.1.3 Stochastic linear difference equations

Definitions

- A **martingale difference sequence adapted to** $\mathbf{J}_t$ is a sequence z_{t+1} that satisfies the equation $E[z_{t+1}|J_t] = z_t$

- The **first order stochastic linear difference equation** is of the following form $x_{t+1} = A_0 x_t + C w_{t+1}$. w_{t+1} must satisfy one of 3 assumptions (in order of decreasing strictness. Note that if a higher one is satisfied, all lower ones are too. i.e. if A1 then A2 and A3, if A2 then A3):

  - Distributed i.i.d $N \sim (0, I)$
  - $Ew_{t+1}|J_t = 0$ and $Ew_{t+1}w'_{t+1}|J_t = I$, where J_t is all the information at time t and the $E[\cdot|J_t]$ denotes a conditional expectation.
  - $Ew_{t+1} = 0$ and $Ew_{t+1}w'_{t-j} = I$ when $j = 0$ and $Ew_{t+1}w'_{t-j} = 0$ when $j \neq 0$ (this is white noise)

- A **stochastic process** is a sequence of random vectors

- A stochastic process $\{x_t\}$ is said to be **covariance stationary** if it satisfies the following two properties: (a) the mean is independent of time ($Ex_t = Ex_0 \forall t$) and (b) the sequence of autocovariance matrices $E(x_{t+j} - Ex_{t+j})(x_t - Ex_t)'$ depends on the separation between dates j, but not on t.

- A square real valued matrix A_0 is said to be stable if all of its eigenvalues in modulus are strictly less than unity.

Algorithms and Applications

- Linear Stochastic Difference equation form
  To put a stochastic process in the form of a first order linear stochastic difference equation, come up with matrices A_0, C and (optionally) G that satisfy

$$x_{t+1} = A_0 x_t + C w_{t+1}$$

$$y_t = G x_t$$

- Covariance stationary stochastic processes
  Whether or not a stochastic process is covariance stationary often depends on the form in which the process is presented and some initial conditions. We will be working with the form: $\$[x_{\mathrm{1,t+1}}$
  $\mathrm{x}_{\mathrm{2,t+1}} = [1\&00\&\tilde{A}\}]\,[x_{\mathrm{1,t}}$
  $\mathrm{x}_{\mathrm{2t}} + \left[0\tilde{\{C\}}\right] w_{\mathrm{t+1}}\$$. In this form we assume that $\tilde{A}$ is a stable matrix (so that 1 cannot be an eigenvalue of $\tilde{A}$)

  To find the covariance stationary initial conditions for the mean and covariance of a stochastic process, you must follow these steps:

  1. Set your system up in the form described above.
  2. Solve for the stationary mean by taking expected values of both sides of the stochastic linear difference equation and ending up with this equation: $(I - A_i)\mu = 0$. You then need to solve for the eigenvector that corresponds to the single unit eigenvalue of A_0. This vector is the stationary mean vector, or μ.
  3. Solve for the stationary variance by solving the matrix quadratic: $C_x(0) = A_0 C_x(0) A_0' + CC'$. The autocovariance process through time can be found via the relation: $C_x(j) = A_0^j C_x(0)$

  To solve for impulse response functions,
  Re-write the stochastic linear difference equation using the lag operator $Lx_{t+1} = x_t$ to get $(I - A_o L)x_{t+1} = Cw_{t+1}$. Iterating forward from time $t = 0$ leads to the following expressions for x_t and y_t:

$$x_t = A_0^t x_0 + \sum_{j=0}^{t-1} A_0^j C W_{j-t}$$

$$y_t = G A_0^t x_0 + G \sum_{j=0}^{t-1} A_o^j C w_{t-j}$$

The impulse response function for x is $h_j = A_0^j C$ and the impulse response function for y is $\tilde{h}_j = G A_0^j C$

Forecasting the conditional covariance matrix
Using the impulse response functions from above we can forecast the expected t period ahead conditional covariance matrix $E(y_t - EY_t|x_0)(y_t - Ey_t|x_0)' = G\left[\sum_{h=0}^{t-1} A_0^h C C' A_0^{h'}\right] G'$

How to apply the Howard improvement algorithm to the evaluation of dynamic criterion.
We will be working with the following equations:

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1}$$

$$u_t = -F_0 x_t$$

$$v(x_0) = -E_0 \sum_{t=0}^{\infty} \beta^t \left[x_t' R x_t + u_t' Q u_t\right]$$

1. Start with some given policy rule F_0 and use it to find $P_0 = R + F_0' Q F_0 + \beta(A - BF_0)' P_0 (A - BF_0)$
2. Use this P_0 to find a $F_1 = \beta(Q + \beta B' P_0 B)^{-1} B' P_0 A$
3. Repeat this sequence using the expressions $P_j = R + F_j' Q F_j + \beta(A - BF_j)' P_j (A - BF_j)$ and \$F$_{j+1}$ = β(Q + β B' P$_j$ B)$^{-1}$B'P$_j$

1.1.4 Population

Algorithms and Applications

- Parameter estimation
  This is simple least squares. If Y is governed by a state-space system and somehow X comes from Y. you can do least squares on them to get a vector β that minimizes the sum of squared errors for the regression. We get that

$$\beta = (EYX')[E(XX')]^{-1}$$

- Multiple Regressors
  If instead of being a scalar Y is a vector of random variables, then you will do multiple regressions. In this case β becomes a matrix and the error term is a vector. The equation for beta is found in the same way.

1.1.5 Estimation of Model Parameters

Algorithms and Applications

- Likelihood function
  The Likelihood function is defined as the joint probability distribution of all the state variables $f(x_t, x_{t-1}, \ldots x_0)$. This distribution can be factored by multiplying successive conditional joint probability distributions

  $$f(x_t, x_{t-1}, \ldots, x_0) = f(x_t|x_{t-1}, \ldots, x_0)f(x_{t-1}|x_{t-2}, \ldots, x_0) \ldots f(x_1|x_0)f(x_0)$$

  Note that for a Markov system the equation becomes $f(x_t|x_{t-1}, \ldots, x_0) = f(x_t|x_{t-1})$ because of the Markov property. This means that the likelihood function becomes

  $$f(x_t, x_{t-1}, \ldots, x_0) = f(x_t|x_{t-1})f(x_{t-1}|x_{t-2}) \ldots f(x_1|x_0)f(x_0)$$

- Special log-likelihood function
  If the w's underlying the stochastic process for Y are Gaussian, then we know what the conditional distribution $f(x_{t+1}|x_t)$ is Gaussian with mean $A_0 x_t$ and covariance matrix CC'. Taking the log of the conditional density of the n dimensional vector x_t becomes

  $$\log f(x_{t+1}|x_t) = -0.5n \log(2\pi) - 0.5 \log det(CC') - 0.5(x_{t+1} - A_0 x_t)'(CC')^{-1}(x_{t+1} - A_0 x_t)$$

1.1.6 The Kalman filter

Definitions

Algorithms and Applications

Theorems

2 TODO Marching Orders, number 1 [0/4]

Clock summary at *2013-07-10 Wed 01:25*

Headline	Time	
Total time	**20:50**	
TODO [#A] Marching Orders, number 1...	20:50	
TODO [#A] Read Chapter 2 of RMT4		15:17
TODO [#A] Work exercises 2.1-2.5		5:33

`DEADLINE:` *2013-07-14 Sun*

2.1 TODO Read Chapter 2 of RMT4

I should probably look over section 2.4.5.2 again. It was a bit complicated and I couldn't replicate its results on my own.

2.2 TODO Read the two techincal appendixes

2.3 TODO Work exercises 2.1-2.5

2.4 TODO Think of python examples :TOM:

Re-create `markov.m` and other Matlab programs