CHAPTER 6

METHODOLOGY

Firstly, this chapter overviews and illustrates the fundamental of Bayesian estimation concept briefly. Secondly, the steps of a Bayesian perspective are required to explain later. In order to keep it a head of Bayesian estimation, the original data and their re-scale are also shown in next section.

6.1 Fundamental of Bayesian Estimation Concept

"...the conduct of monetary policy in the United States has come to involve, at its core, crucial elements of risk management. This conceptual framework emphasizes understanding as much as possible the many sources of risk and uncertainty that policymakers face, quantifying those risks when possible, and assessing the costs associated with each of the risks. In essence, the risk management approach to monetary policymaking is an application of Bayesian decision-making."

By Alan Greenspan at the Meetings of the American Economic Association,
San Diego, California, January 3, 2004:

In a point of view estimating DSGE models, the standard approach in Real Business Cycle theory pioneered by Kydland and Prescott (1982) and Long and Plosser (1983) has been to calibrate parameters and compare moment generated from the model with those of actual data. This method, however, lacks formal statistical foundations (Kim and Pagan, 1994) and hinders testing the result. Sargent (1989) suggest as an alternative maximum likelihood estimation of DSGE models, but potential misspecification due to omitted non-linearities, incorrect assumptions about preferences and technology or incorrectly-specified exogenous shocks easily lead to computational difficulties (Lubik and Schorfheide, 2005).

The Bayesian estimation methodology chosen here follow developments by Dejong et al (2000a,b), Otrok (2001) and Smet and Wouters (2003). Bayesian analysis formally incorporates uncertainty and prior information regarding the parameterization of the model. In other words, in Bayesian analysis parameters of a model are regarded not as unknown fixed values but rather as random variables which take a certain value with some probability - the true value is uncertain. It combines the likelihood with a priori information on the parameters of interest that may have come from earlier microeconometric or macroeconometric studies. By introducing prior information about the structural parameters in the form of probability densities, the likelihood function is reweighed by the prior density. The degree of uncertainty about the prior information can thereby be expressed in terms of the standard deviation of the prior density. Hence, the common practice of fixing some parameters in maximum likelihood estimation has the Bayesian interpretation that no uncertainty exists about the chosen values. Therefore Bayesian approach is taking explicitly account of all uncertainty surrounding parameter estimates

In this section we will introduce the basic mechanics of Bayesian estimation. Firstly, at its most basic level, Bayesian estimation is a bridge between calibration and maximum likelihood. The tradition of calibrating models is inherited through the specification of priors. And the maximum likelihood approach enters through the estimation process based on confronting the model with data. Together, priors can be seen as weights on the likelihood function in order to give more importance to certain areas of the parameter subspace. More technically, these two building blocks - priors and likelihood functions - are tied together by Bayes' rule as following below.

First, priors are described by a density function of the form

$$p\left(\theta_A \middle| A\right) \tag{6.1}$$

where A stands for a specific model, represents the parameters of model A, $p(\bullet)$ stands for a probability density function (pdf) such as a normal, gamma, shifted gamma, inverse gamma, beta, generalized beta, or uniform function.

Second, the likelihood function describes the density of the observed data, given the model and its parameters:

$$L\left(\theta_A \middle| Y_T, A\right) \equiv p\left(Y_T \middle| \theta_A, A\right) \tag{6.2}$$

where Y_T are the observations until period T, and where the likelihood is recursive and can be written as:

$$p\left(Y_T \left|\theta_A, A\right.\right) = p\left(y_0 \left|\theta_A, A\right.\right)\prod_{t=1}^{T} p\left(y_t \left|Y_{t-1}\theta_A, A\right.\right) \qquad (6.3)$$

We now take a step back. Generally speaking, we have a prior density $p(\theta)$ on the one hand, and on the other, a likelihood $p\left(Y_T \left|\theta\right.\right)$. In the end, we are interested in $p\left(\theta \left|Y_T\right.\right)$, the posterior density. Using the Baye's theorem twice we obtain this density of parameters knowing the data. Generally, we have

$$p\left(\theta \left|Y_T\right.\right) = \frac{p\left(\theta;Y_T\right)}{p\left(Y_T\right)} \qquad (6.4)$$

We also know that

$$p\left(Y_T \left|\theta\right.\right) = \frac{p\left(\theta;Y_T\right)}{p(\theta)} \Leftrightarrow p\left(\theta;Y_T\right) = p\left(Y_T \left|\theta\right.\right) \times p(\theta) \qquad (6.5)$$

By using these identities, we can combine the prior density and the likelihood function discussed above to get the posterior density:

$$p\left(\theta_A \left|Y_T, A\right.\right) = \frac{p\left(Y_T \left|\theta_A, A\right.\right) p\left(\theta_A, A\right)}{p\left(Y_T \left|A\right.\right)} \qquad (6.6)$$

where $p\left(Y_T \left|A\right.\right)$ is the marginal density of the data conditional on the model

$$p\left(Y_T \left|A\right.\right) = \int_{\theta_A} p\left(\theta_A;Y_T \left|A\right.\right) d\theta_A \qquad (6.7)$$

Finally, the posterior kernel (or un-normalized posterior density, given that the marginal density above is a constant or equal for any parameter), corresponds to the numerator of the posterior density:

$$p\left(\theta_A \left|Y_T, A\right.\right) \propto= p\left(Y_T \left|\theta_A, A\right.\right) p\left(\theta_A, A\right) = \kappa\left(\theta_A \left|Y_T, A\right.\right) \qquad (6.8)$$

This is the fundamental equation that will allow us to rebuild all posterior moments of interest. The trick will be to estimate the likelihood function with the help of the Kalman filter and then simulate the posterior kernel using a sampling-like or Monte Carlo method such as the Metropolis-Hastings which it will be in the next session.

6.2 Estimation Methodology

We adopted a Bayesian estimation approach to bring the model directly to the data. This approach discussed by many authors in the literature in the last few years (e.g. Schortheide 2000, Lubik and Schorfhelde, 2005, Smets and Wouters, 2003). Schematically, we summarize the method which consists of following steps:

1. The non-linear DSGE model is solved via a linear approximation: a linear rational expectation system is obtained that must be obtaining a standard linear model in state apace form.

2. The state-space approximation of the original non-linear model allows the identification of a likelihood function as Kalman recursions and subsequent inference based on Maximum likelihood estimation.

3. Usually theoretical model imply few, well defined shocks; unfortunately this often implies singularities in the determination of the likelihood (in the Kalman filter the number of shocks must at least be as large as the number of observables), implying the introduction of additional structural shocks and/or measurement errors.

4. The Bayesian analysis is performed; prior distributions for model parameters have to be defined, representing the prior beliefs of the analyst on their plausible values, which, in combination with the likelihood function, allows us obtaining the posterior distribution.

5. The Bayesian inference needs the use of stochastic simulations, specifically Markov Chain Monte Carlo (MCMC) techniques, allowing to obtain samples form the posterior joint pdf of the model parameters and subsequently to make an inference in which the parameter uncertainty and the shape of the likelihood are taken into account.

Consequently, we introduce the step of a Bayesian estimation approach as following;

6.2.1 Solving the Model

After we obtain log-linearized the equation, it can be rewritten into a form of a linear rational expectations system (LRE System) as:

$$0 = Ax_t + Bx_{t-1} + Cy_t + Dz_t \tag{6.9}$$

$$0 = FE_t\left(x_{t+1}\right) + Gx_t + Hx_{t-1} + JE_t\left(y_{t+1}\right) + Ky_t + LE_t\left(z_{t+1}\right) + Mz_t \tag{6.10}$$

$$E_t\left(z_{t+1}\right) = Nz_t + E_t\left(\xi_{t+1}\right), E_t\left(\xi_{t+1}\right) = 0 \tag{6.11}$$

for $t = 0,1,2,...$

The vector x_t is the endogenous state vector, y_t is the endogenous vector of unobservable variables (control variable) and z_t is the exogenous stochastic process. We also write it basically in Appendix D.

From using algorithm of Uhlig (1995), the system of above equation can be transformed to the recursive rule for the general equilibrium expressed by the following state model:

$$x_t = Px_{t-1} + Qz_t \tag{6.12}$$

$$y_t = Rx_{t-1} + Sz_t \tag{6.13}$$

$$z_t = Nz_{t-1} + \varepsilon_t \tag{6.14}$$

for $t = 0,1,2,...$ where the equilibrium is stable and described by matrices P, Q, R and S. and x_t, y_t and z_t are denoted in terms of log-deviation from steady state of variables in period t. To calculate the solution, one needs to solve a matrix quadratic equation as following Uhilg (1995).

From the computational point of view, the linear approximation and the solution of the obtained LRE (Linear Rational Expectations) can be done automatically using the DYNARE[1] program (Juillard, 2001).

6.2.2 Constructing the Kalman Filter

This section describes the construction of the Kalman filter used for evaluating the likelihood function. The rational-expectation solution log-linearized model, as we discuss above, can be express by the following time-varying-coefficients difference equations:

[1] DYNARE is the additional command for the simulation DSGE models, freely available and totally open source. Presently, an estimation module is implemented on DYNARE, to include the most recent developments in Bayesian estimation macro-economic models in an extremely efferent. More details, consulting with Griffoli (2007)

$$X_t = PX_{t-1} + QZ_t \quad (6.15)$$
$$Z_t = NZ_{t-1} + \varepsilon_t \quad (6.16)$$

X_t is the endogenous vector (x_t, y_t), Z_t is the exogenous shocks vector and ε_t is the vector of innovation as $E_t(\xi_{t+1}) = 0$.

The state-space form of the above difference equations is

$$s_t = \Gamma_1 s_{t-1} + \Gamma_2 w_t \quad (6.17)$$

where $s_t' = [X_t', Z_t']$ and

$$\Gamma_1 = \begin{bmatrix} P_t & Q_t P_t \\ 0 & N_t \end{bmatrix} \quad \text{and} \quad \Gamma_2 = \begin{bmatrix} Q_t \\ I \end{bmatrix}$$

The observation equation is given by

$$Y_t = \Lambda s_t + \nu_t \quad (6.18)$$

where

Y_t is a vector of observed variables.

Γ_1 and Γ_2 are matrices of functions of the model's deep parameters (matrices P, Q, R and S) from the state equation representing the dynamic core of the equation.

Λ is a matrix expressing the relationship between observed and state variables.

w_t is a vector of state innovations: $w_t \sim N(0, \Xi)$

ν_t is a vector of measurement errors: $\nu_t \sim N(0, \Upsilon)$

The likelihood function is computed under the assumption of normally distributed disturbances by combining the state-space representation of the model with the measurement equation linking the observed data and the state vector express by equation:

$$s_{t+1} = \Gamma_1 s_t + \Gamma_2 w_{t+1} \quad (6.19)$$
$$Y_t = \Lambda s_t + \nu_t \quad (6.20)$$

As Hamilton (1994, chapter 13) shows, it is possible to use a Kalman filter derived from a time-varying model for likelihood evaluation given the initial some initial stat value $s_0 \sim N\left(s_0, \Sigma_0\right)$

Then we can write the likelihood function of the model[2]:

$$\log L\left(Y^T \left|\Theta\right.\right) = \frac{1}{2}\sum_{t-1}^{T}\left[N\log 2\pi + \log\left|\Omega_{t|t-1}\right| + \sum_{t-1}^{T}\nu_t'\left|\Omega_{t|t-1}^{-1}\right|\nu_t\right] \tag{6.21}$$

where:

$$\Theta = \left\{\Gamma_1, \Gamma_2, \Lambda, \Xi, \Upsilon\right\}$$
$$\Omega_{t|t-1} = \Lambda'\sum\nolimits_{t|t-1}\Lambda + \Upsilon$$
$$\sum\nolimits_{t|t-1} = \Gamma_1\sum\nolimits_{t-1|t-1}\Gamma_1' + \Gamma_2\Xi\Gamma_2'$$

6.3 Data Description

We estimate the model with quarterly data of Thailand and U.S. economies for the period 2000:Q1 to 2007:Q4. Because of the structural breaks in Thai economy, it is, therefore, preventing us from using longer series[3]. Most of data are selected from Bank of Thailand (BOT) and National Economic and Social Development Board (NESDB). For the foreign data, the source is Bureau of Economic Analysis and Board Government of the Federal Reserve System.

Data: The Thai SOE-DSGE NK Model

We choose the following observable eight variables. All variable are de-trended in the given period except the inflation gap which is constructed by the difference of a rate of core inflation and a BOT's target.

[2] The alternative expression for the likelihood function and the Kalman filter step are shown in Appendix D.

[3] The economic crisis was in 1997 and BOT begin to adopt the inflation targeting regime since 2000.

1. y_t is a output gap constructed as a de-trended of log real GDP during the given period.

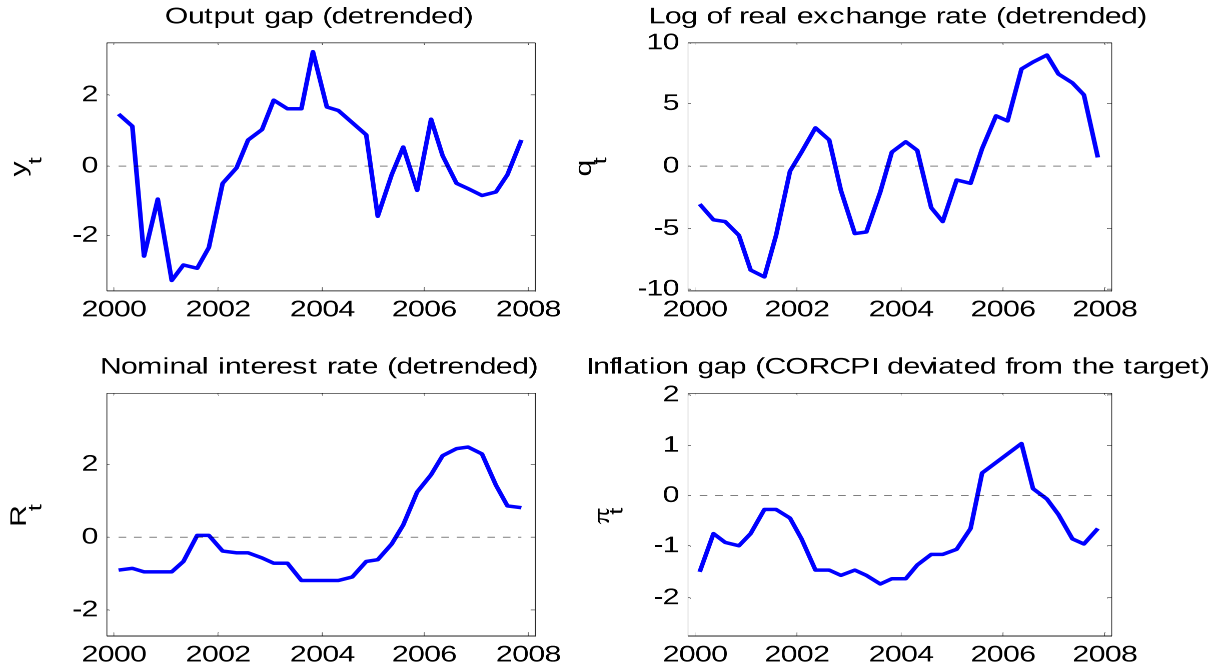
2. q_t is a de-trended real exchange rate using real effective exchange rate[4] (REER) to be a proxy.

3. r_t is a signal policy rate of BOT. RP 14.

4. π_t is the difference of a rate of core inflation and a BOT's target (using an average value of BOT target, 0-3.5 percent).

Figure 6.1

Data to The Thai Open Economy DSGE-based New-Keynes Model, part 1



---

[4] $REER = \sum_{i=1}^{n} w_i \times \frac{FC_i}{HC} \times \frac{P}{P_i}$ where $\sum_{i=1}^{n} w_i = 1$ n = the number of traded partners.

FCi/HC = currency of home country per a unit of foreign currency.

P = the price index for home country.

Pi = the price index of traded partner country i

More details, one can see from the literature of Medhinee Supasawatkul (1999).

5. π_{F_t} is the import price inflation calculated by using the seasonally adjusted import price index.

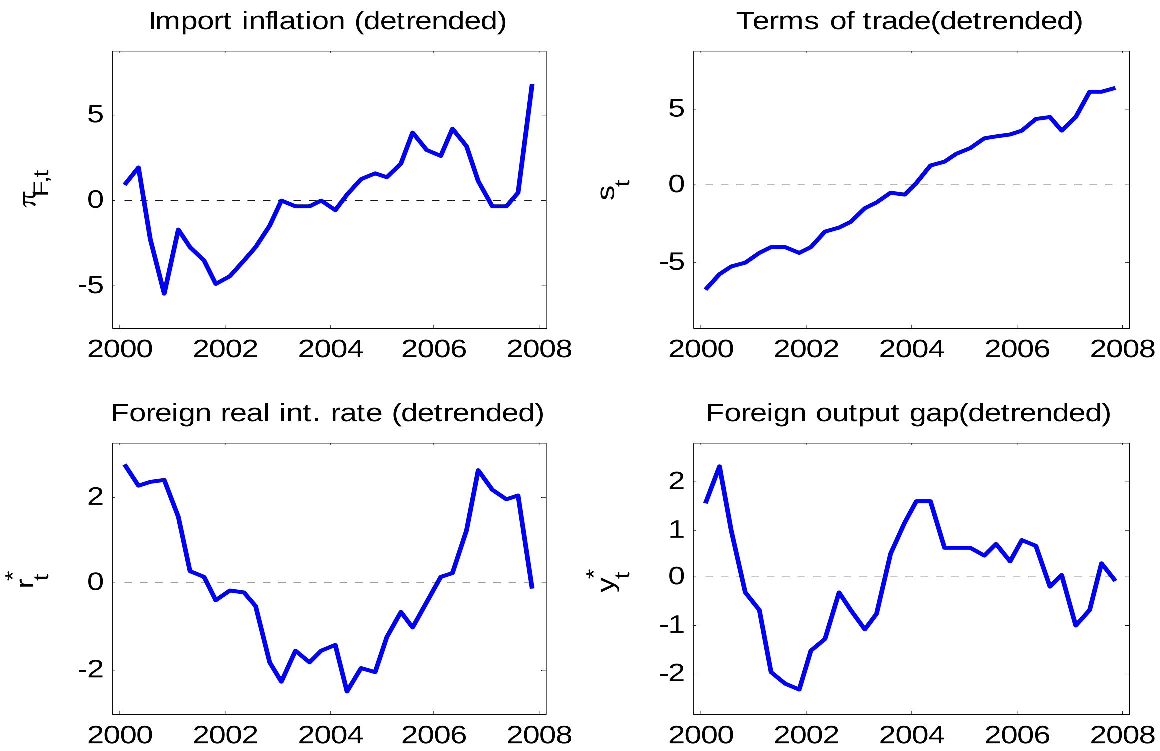
6. s_t is a terms of trade which are defined following the model. It is the difference of log of foreign (U.S.) consumer price index and core consumer price index.

7. y_t^* is a foreign output gap constructed by using the growth rate of US gross domestic product.

8. r_t^* is a real foreign interest rate calculated by the difference of fed fund rate and the overall inflation

Figure 6.2

Data to The Thai Open Economy DSGE-based New-Keynes Model, part 2



6.4 Prior Specifications

As we introduce the Bayesian estimation in the first part, the prior is a part of Bayesian analysis for reflecting our beliefs about the values that parameters can take. Larger prior standard deviations result in diffuse distributions, which mean we have little information in addition to the data. Most earlier attempts to estimate new Keynesian DSGE models with the Bayesian approach use data form developed countries. Therefore, it takes a limit of pre-assumption on developing countries.

For prior value of a habit coefficient, we run simply OLS equation. Elasticity of interest rate to inflation and output, we also assume following the Taylor rule. For Other coefficients, it lines in some literatures (Silveira 2006, Liu 2005, Jakab and Vilagi 2007, Medina and Soto 2006). According to the distributions, some literatures suggest that the gamma and normal distributions can be used interchangeable. The beta distribution is used for parameters constrained on the unit interval. The inverse gamma distribution is used for the shocks

Table 6.1

The Prior Specifications

Parameter	**Definition**	**Domain**	**Density**	**Mean**	**Variance**
h	Habit formation parameter	[0,1]	Beta	0.9	0.1
σ	Inverse elasticity of intertemporal substitution	$\mathbb{R}^{+}$	Normal	1.00	0.50
η	Elasticity of substitution between home and foreign goods	$\mathbb{R}^{+}$	Gamma	1.00	0.50
φ	Inverse elasticity of labor supply	$\mathbb{R}^{+}$	Gamma	1.00	0.50
θ_H	Fraction of non-optimizing firms	[0,1]	Beta	0.50	0.25
θ_F	Fraction of non-optimizing importers	[0,1]	Beta	0.50	0.25
ϕ_1	Elasticity of interest rate to inflation	$\mathbb{R}^{+}$	Gamma	1.50	0.25
ϕ_2	Elasticity of interest rate to output	$\mathbb{R}^{+}$	Gamma	0.50	0.10

Table 6.1

The Prior Specifications (Continued)

Parameter	Definition	Domain	Density	Mean	Variance
ρ_r	Backward-looking parameter for interest rate	[0,1]	Beta	0.50	0.18
ρ_r^*	Foreign real interest rate inertia parameter	[0,1]	Beta	0.50	0.18
ρ_a	Inertia of technology development	[0,1]	Beta	0.50	0.18
λ_1	Foreign output inertia parameter	[0,1]	Beta	0.50	0.18
σ_a	Sd. of productivity shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_s	Sd. of terms of trade shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_q	Sd. Of real exchange rate	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_{π_H}	Sd. of domestic inflation shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_{π_F}	Sd. of import inflation shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_r	Sd. of interest rate shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_{y^*}	Sd. of foreign output shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$
σ_{r^*}	Sd. of foreign real interest rate shock	$\mathbb{R}^{+}$	InvGamma	2.00	$[0,\infty]$

6.5 <u>The Markov Chain Monte Carlo Method (MCMC)</u>

After we have estimated the maximum likelihood model and have got the coefficients from constructing the Kalman filter and also have determined the prior distribution from some personal introspection to reflect strongly held beliefs about the validity of economic theories, our task is, then, to adopt the Markov Chain Monte Carlo Method for sampling from probability distributions, according Bayesian approach which is supposing that the parameter θ is a random vector evaluating in time.

Markov Chain Mote Carlo (MCMC)[5] is a method of sampling a target probability distribution by constructing a Markov Chain such that the target distribution is the stationary distribution of the chain, and such that the chain converges in distribution to that stationary distribution. When convergence occurs, realizations of the chain are realizations of the stationary distribution. The task is to construct a chain having a given target as its stationary distribution.

6.5.1 Markov Chains and Transition Kernels

Generally, a Markov Chain is a sequence of random variables X_1, X_2, X_3, ... in which the conditional distribution of a present observations given a set of past observations only depends on the past through the most recent observation. Specifically, if χ is the sample space for the $\{X_t\}$ and A is a subset of a collection of sets on χ then

$$P\left(X_{t+1} \in A \middle| x_0, x_1, ..., x_t\right) = P\left(X_{t+1} \in A \middle| x_t\right) \qquad (6.22)$$

for all $t = 1, 2, 3, ...$ and any such A. The value taken by X_t is called the state of the chain at t. An above expression is called a transition probability. The rule describing how the chain moves from its state at t to its state at $t+1$ is described by the transition kernel. This is a function $K(x, y)$ that for each x provides a probability distribution for y. Thus it is a collection of conditional probability distributions, one for each x. When the sample space, χ, is discrete:

$$K(x, y) = p\left(X_{t+1} = y \middle| X_t = x\right) \qquad x, y \in \chi \qquad (6.23)$$

The probability distribution of X_{t+1} say p_{t+1}, can be described in terms of the transition kernel and the analogous distribution of X_t. This is because the probability that $X_{t+1} = y$ is equal to the sum or integral of the probabilities that $X_t = x$ times the probability that the chain moves to given that it had been in x. Algebraically, when there are M states, this is

---

[5] The phrase monte carlo refers to the use of random number generators to solve mathematical problems. The phrase itself is a reference to the Principality of Monte Carlo in southern France, which is famous for its casino.

$$P\left(X_{t+1}=j\right)=\sum_{t=1}^{M}P\left(X_t=i\right)P\left(X_{t+1}=j\middle|X_t=i\right)\quad ,\ j=1,2,\ldots,M.$$

that is,

$$p_{t+1}\left(j\right)=\sum_{t=1}^{M}p_t\left(i\right)K\left(i,j\right)\quad \text{or}\quad p'_{t+1}=p'_tK\ , \tag{6.24}$$

in terms of continuous sample:

$$p_{t+1}\left(y\right)=\int_{\chi}K\left(x,y\right)p_t\left(x\right)dx.$$

6.5.2 <u>The Metropolis-Hastings (M-H) Algorithms</u>

Metropolis Hasting algorithm is a method to find a kernel (chain) corresponding to a given stationary distribution. In other words, Metropolis-Hastings algorithm is a rejection sampling algorithm used to generate a sequence of samples from a probability distribution.

The algorithm generates a Markov chain in which each state y^{t+1} depends only on the previous state y^t. The step of this algorithm is made as follow. First, we choose an initial value, y_0, and set $t=0$. Next step is draw y^* from $q\left(.\middle|y_t\right)$(a sequence of proposal distribution). The third step is to calculate the ratio $r\left(y_t,y^*\right)=\dfrac{p\left(y^*\right)q\left(y_t\middle|y^*\right)}{p\left(y_t\right)q\left(y^*\middle|y_t\right)}$. If $r\geq 1$, set $y_{t+1}=y^*$; otherwise set $y_{t+1}=y_t$. Next is to increase t by one and then proceed to step 2.

Whether the probability that y^* is accepted depends on as follow

$$\rho\left(y_t,y^*\right)=\min\left(\frac{p\left(y^*\right)q\left(y_t\middle|y^*\right)}{p\left(y_t\right)q\left(y^*\middle|y_t\right)},1\right)$$