BAYESIAN APPROACH ON PARAMETER ESTIMATION IN HIDDEN MARKOV MODEL

Dwi Agustin N.S¹, Septiadi Padmasistra² and Sudartianto²
¹Departement of Industrial Engineering, Widya Mandala Catholic University, Surabaya, Indonesia
²Deparment of Statistics, Padjadjaran University, Bandung, Indonesia
E-Mail: dwi.agustinns@gmail.com

ABSTRACT

This paper presents study about the parameter estimation in hidden markov model. The approach is taken from a Bayesian method, there will be two sources of information, there are information from the likelihood function and the prior function. This approach will be applied to daily rainfall data in Darajat, Garut. The numbers of hidden states are used in this paper based on Schmidt and Fergusson’s climate classification which are suitable to the local conditions. This classification was obtained three types of division in the period of one year where the condition called wet months when monthly rainfall > 100 mm per month, moist months when monthly rainfall between 100 - 60 mm and the dry months when monthly rainfall <60 mm per month. The process estimation of hidden markov parameters is using Gibbs Sampler algorithm.

Keywords: hidden Markov models, Bayesian, Schmidt and Fergusson’s Climate classification, Gibbs sampler.

INTRODUCTION

Modeling of the precipitation has been developed by many researchers; one of them is Coe and Stern (1982) who tried to modeling the daily rainfall data in Zinder region, Nigeria and the Kharja, Jordan. The method used at their research was Generalized Linear Models (GLM) and markov chain. In the 1970's the mathematician Baum and Petrie introduced markov chain development, namely Hidden Markov Model (HMM). HMM was increasingly popular applied in various fields? Rabiner (1989) applied the methods of HMM in speech recognition. Zucchini and Guttorp (1991) applied HMM in the precipitation phenomena. In their research, Zucchini and Guttorp introduced unobserved climate states that influence the occurrence of rain. Thyer and Kuezera (2000) developed a method of HMM to simulate the long-term hydroclimatic data for water resource planning in Australia.

In previous studies, the number of hidden states was observed only two kinds, while in Indonesia itself more widely used climate classification based on Schmidt and Fergusson. In Sudrajat’s research (2009) mentioned that according Schmidt and Fergusson (1951), classification of climate is based on a comparison between the dry and wet months, from this relationship was obtained three types of division in the period of one year where the condition called wet months when monthly rainfall > 100 mm per month, moist months when monthly rainfall between 100 - 60 mm and the dry months when monthly rainfall <60 mm per month.

Garut’s topographic itself is the mountains area where there are protected forests and plantations. The climate classification which are suitable for this area is climate classification based on Schmidt and Fergusson (Sudrajat, 2009).

The process of parameter estimation in hidden markov models in several previous studies were based on a frequentist approach using Baum-Welch algorithm or EM Algorithm. At this approach, the parameters only considered as a fixed value. However, there are other approaches that assume these parameters will form a random variable. It can be happened, because the parameters of hidden markov models, particular in the case of hydrology (rainfall), the influence of time would make the values of these parameters to form a random variable having a probability distribution. So the Bayesian approach can be used in the process of parameter estimation in hidden markov model. So in this paper want to study about parameter estimation in hidden markov models with three different types of hidden climate states in Garut through a Bayesian approach.

HIDDEN MARKOV MODEL (HMM) WITH THREE HIDDEN CLIMATE STATES

Basic Model of HMM

HMM are models in which the distribution that generates an observations depends on the state of an underlying and unobserved markov process (Zucchini and MacDonald, 2009).

The basic model of HMM was illustrated by Ingmar Visser (2011) as shown below:

![Figure-1. Hidden Markov Model Illustrated.](image-url)
Zucchini and MacDonald (2009) are formulating the basic model above becomes:

\[ \Pr(S_t | S^{(t-1)}) = \Pr(S_t | S_{t-1}), \quad t = 2, 3, \ldots \]  

\[ \Pr(Y_t | S^{(t-1)}, S^{(t)}) = \Pr(Y_t | S_t), \quad t = 2, 3, \ldots \]  

with \( Y^{(t)} = (Y_1, Y_2, \ldots, Y_t) \) and \( S^{(t)} = (S_1, S_2, \ldots, S_t) \).

The Equilibrium Probability from HMM with Three Hidden Climate States

Climate states will be following a markov chain with an illustration of this process can be seen in Figure-2.

\[ \Pr(S_t | S_{t-1}) = \Pr(S_t | S_{t-1}), \quad t = 2, 3, \ldots \]  

\[ \Pr(Y_t | S_{t-1}, S_t) = \Pr(Y_t | S_t), \quad t = 2, 3, \ldots \]  

with \( Y^{(t)} = (Y_1, Y_2, \ldots, Y_t) \) and \( S^{(t)} = (S_1, S_2, \ldots, S_t) \).

The relationship between the transition probability matrix with the equilibrium probability from each of climate states are:

\[ \delta = \begin{bmatrix} \delta_D & \delta_M & \delta_W \end{bmatrix} \]  

with \( \delta_D + \delta_M + \delta_W = 1 \)

The joint density of all the data is then:

And the equilibrium probability from each of climate states are:

\[ \delta P = \delta \]

Parameters in HMM with Three Hidden Climate States

According to Tyer dan Kuezera (2000) explained that there is a unknown parameters (\( \theta \)) in the HMM. Parameters (\( \theta \)) will be consist of: the parameters of each climate states distribution (\( \mu \) dan \( \sigma \)), the transition probability of the climate states, and climate states sequence \( S_n = \{s_1, s_2, \ldots, s_n\} \).

So that, its parameters will be consist of:

\[ \theta = \{\mu, \sigma^2, q_1, q_2, q_3\} \]

where: \( \mu = (\mu_D, \mu_M, \mu_W) \), \( \sigma^2 = (\sigma_D^2, \sigma_M^2, \sigma_W^2) \), \( q_i \) is the i-th row of P.

Parameter Estimation in HMM with Three Hidden Climate States

Lambert at al (2003) were estimating parameters through a bayesian equation, by calculating \( \mu \) and \( \sigma \) from posterior distribution of \( \theta \). Based on bayes theorem can be written as:

\[ p(\theta | Y_n) = \frac{P(Y_n | \theta) \times p(\theta)}{P(Y_n)} \]

where:

\[ p(\theta | Y_n) \] = conditional on the entire set of observation \( Y_n \)

\[ P(Y_n | \theta) \] = likelihood function of \( Y_n \)

\[ p(\theta) \] = prior distribution of \( \theta \)

\[ P(Y_n) \] = marginal probability of \( Y_n \)

In Chib (1995) likelihood function of \( Y_n \) can be written as:

\[ P(x, \mu, \sigma^2, \theta) = \phi(x; \mu, \sigma^2) + P(z_1 = \mu | Y_n; \theta) \phi(z_1; \mu, \sigma^2) \]

where \( Y_{t-1} \) represents all the observation up to time \( t-1 \).
\[ P(y|\theta) = \prod_{i=1}^{n} P(y_i|Y_{-i}, \theta) \]
\[ P(y|\theta) = P(y_1|\theta) \prod_{i=2}^{n} P(y_i|Y_{-i}, \theta) \] (11)

**Prior Distribution of \( \theta \) or \( p(\theta) \)**

Chib (1995) define the prior information through the distribution:

\[ \mu_j \sim N(\mu_0, A^{-1}) \]
\[ \sigma_j^2 \sim IG\left(\frac{\nu_0}{2}, \frac{\delta_0}{2}\right) \]
\[ \alpha \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3) \]

**Posterior Distribution of \( \theta \) or \( p(\theta | Y_i) \)**

Chib (1995) define the posterior distribution of \( \theta \) through the distribution:

\[ \pi(\mu|y, z, \sigma^2) = \prod_{i=1}^{d} \phi(\mu_i|\bar{\mu}_i, B_i) \]

with \( \bar{\mu}_i = (A+\sigma_j^2 n_j)^{-1}(A\mu_0 + \sigma_j^2 \bar{y}_j) \) and \( B_i = (A+\sigma_j^2 n_j)^{-1} \).

\[ \pi(\sigma^2|y, z, \mu) = \prod_{i=1}^{d} \left( \frac{1}{\sigma_i^2} \left( \frac{\nu_0+n_j}{2} \right) \right) \]

with \( \delta_j = (y_j - i, \mu_j)^T(y_j - i, \mu_j) \).

\[ \pi(q, \nu_1, \nu_2, \nu_3 | y, z) = \prod_{i=1}^{d} \phi(\nu_i|\nu_i + n_i, \alpha_2 + n_2, \alpha_3 + n_3) \]

So that, the posterior density can be computed from the decomposition:

\[ \pi(\theta | y) = \pi(\mu | y) \times \pi(\sigma^2 | \mu, y) \times \pi(q | \sigma^2, \mu, y) \] (12)

It is difficult when solving bayes equation above, so that required simulation method to obtain the value of parameter \( \theta \). Tyler and Kuzeera (2000) using Markov Chain Monte Carlo (MCMC) in solving this problem.

**MCMC Gibbs Sampler**

MCMC Gibbs Sampler aims to find the estimated value of \( \theta_j \) using a conditional posterior distribution which is assumed that \( \theta_{-1} \) is known. The Gibbs sampler algorithm can be summarized by the following steps:

- set up data
- set number of iterations (\( T \))
- set prior parameters: \( \mu_0, A, V_0, \delta_0, \alpha_1, \alpha_2, \alpha_3 \)
- initialize vectors of sampled values \( \mu, \sigma, \alpha \)
- set initial current \( \mu^{(0)}, \sigma^{(0)}, \alpha^{(0)} \)
- for \( t = 1, 2, 3, ..., T \) and \( j = 1, 2, 3 \) (climate states)
  repeat:
  - calculate \( \mu := (A + \sigma_j^2 n_j)^{-1}(A\mu_0 + \sigma_j^2 \bar{y}_j) \)
  - calculate \( B_j := (A + \sigma_j^2 n_j)^{-1} \)
  - generate \( \mu \sim N(\bar{\mu}_j, B_j) \)
  - calculate \( \delta_j = (y_j - i, \mu_j)^T(y_j - i, \mu_j) \)
  - calculate \( b = \frac{\delta_0 + \delta_j}{2} \)
  - generate \( \tau \sim G(a, b) \)
  - calculate \( \sigma = \sqrt{\frac{1}{\tau}} \)
  - calculate \( \alpha = \alpha_j + n_j \)
  - generate \( q \sim \text{Dirichlet}(\alpha) \)
  - end of for loop
  - make autocorrelation plot, trace plots, and ergodic mean plots to see the convergence of algorithm and determine its burn-in period.
  - calculate average parameters of \( \theta \) after burn-in period up to the last iteration:
    \[ \bar{\theta} = \frac{\sum_{i=1}^{m} \theta^{(i)}}{m} \]
  - calculate standard deviation parameters of \( \theta \) after burn-in period up to the last iteration:
    \[ s(\theta) = \frac{\sum_{i=1}^{m} (\theta^{(i)} - \bar{\theta})^2}{m-1} \]
  - finish

**APPLICATION OF HMM WITH THREE HIDDEN CLIMATE STATES**

The data used in this study were daily rainfall data in Darajat Garut from March 1, 2011 until March 27, 2012 and this daily rainfall data expressed in millimeters (mm).

**Simulations Results for Parameter \( \mu \) and \( \sigma^2 \) in Dry Climate State**

Initialization value of each prior parameter distribution for dry climate state is:

\[ \mu_{\text{dry}} \sim N(0.4, 0.25) \]
\[ \sigma_{\text{dry}}^2 \sim IG(10, 10) \]

and iterations were performed as 5000 times. The results of the convergence simulations can be seen in Figure-3 and Figure-4.
Based on Figure-3 and Figure-4 above, MCMC sampling scheme can be said convergence in terms of stationarity. It was seen from the trace plots of each parameter that indicates the sample parameter values have the same periodesitas and also the ergodic means plot has shown a steady value after the first 1000 times iterations or at this case burn in period is 1000 times iterations. So that, the results of estimation parameter $\mu_{\text{dry}}$ and $\sigma^2_{\text{dry}}$ for dry climate state is:

### Table-1. Summary of Results for Parameter $\mu$ and $\sigma^2$ in dry climate state.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\text{dry}}$</td>
<td>0.4</td>
<td>0.399</td>
</tr>
<tr>
<td>$\sigma^2_{\text{dry}}$</td>
<td>1.111</td>
<td>2.511</td>
</tr>
</tbody>
</table>

Same steps performed for each parameter in HMM, so that the result value of the parameter as:
Table 2. Summary of Results for All Parameters in HMM with Three Hidden Climate State.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>prior</th>
<th>posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>stddev</td>
</tr>
<tr>
<td>$\mu_{\text{dry}}$</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>$\sigma^2_{\text{dry}}$</td>
<td>1.111</td>
<td>0.154</td>
</tr>
<tr>
<td>$\mu_{\text{moist}}$</td>
<td>2.5</td>
<td>0.3</td>
</tr>
<tr>
<td>$\sigma^2_{\text{moist}}$</td>
<td>1.111</td>
<td>0.154</td>
</tr>
<tr>
<td>$\mu_{\text{wet}}$</td>
<td>20</td>
<td>17.7</td>
</tr>
<tr>
<td>$\sigma^2_{\text{wet}}$</td>
<td>1.111</td>
<td>0.154</td>
</tr>
<tr>
<td>$\delta_{\text{dry}}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{\text{moist}}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta_{\text{wet}}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES


