**Stochastic Models and Forecasting Assignment 2**

a)

Overall proportion of days on which there is any rain = 0.392334

Proportion for each month:

January 0.2741935

February 0.2477876

March 0.2919355

April 0.3450000

May 0.4322581

June 0.5008333

July 0.4790323

August 0.5104839

September 0.4783333

October 0.4403226

November 0.3925000

December 0.3072581

Lowest rainfall appears to be December to April.

Above average rainfall is between May and October.

November is roughly average rainfall.

b)

For one hidden state, m=1, we have:

[**Γ**](http://en.wikipedia.org/wiki/Gamma) **=**

For 2 hidden states, m=2, we have:

[**Γ**](http://en.wikipedia.org/wiki/Gamma) **=**

For three hidden states, m=3, we have:

[**Γ**](http://en.wikipedia.org/wiki/Gamma) **= (  )**

c)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **model** | **No of free parameters** | **minus the log-likelihood** | **AIC** | **BIC** |
| 1-state HMM | 1 | 9785.496 | 19572.99 | 19580.58 |
| 2-state HMM | 4 | 9254.946 | 18517.89 | 18548.25 |
| 3-state HMM | 9 | 9208.609 | 18435.22 | 18503.52 |

The 3 state HMM gives the lowest value for the AIC and BIC so m=3 could be considered the best model to use.

For m=3,

[**Γ**](http://en.wikipedia.org/wiki/Gamma) **= (  )**

As and two of the are close to zero, four of the natural parameters are close to their boundary values, which causes problems of convergence.

A modified HMM in which it is assumed as given that and will work better.

iv)

Carry out the processes of local decoding and global decoding for the fitted model in the case m = 3. Construct a dataframe that, to the original dataframe, adds columns for the hidden state probabilities, a column for the hidden states according to local decoding, and a column for the hidden states according to global decoding.

Sample of the new dataframe

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Month | Day | Rain | p1 | p2 | p3 | ld | gd |
| 1971 | 1 | 1 | 1 | 0 | 0.079 | 0.921 | 3 | 3 |
| 1971 | 1 | 2 | 1 | 0 | 0.065 | 0.935 | 3 | 3 |
| 1971 | 1 | 3 | 1 | 0 | 0.047 | 0.953 | 3 | 3 |
| 1971 | 1 | 4 | 1 | 0 | 0.026 | 0.974 | 3 | 3 |
| 1971 | 1 | 5 | 1 | 0 | 0.001 | 0.999 | 3 | 3 |
| 1971 | 1 | 6 | 0 | 0.999 | 0.001 | 0 | 1 | 1 |
| 1971 | 1 | 7 | 0 | 0.999 | 0.001 | 0 | 1 | 1 |
| 1971 | 1 | 8 | 0 | 0.999 | 0.001 | 0 | 1 | 1 |
| 1971 | 1 | 9 | 0 | 0.999 | 0.001 | 0 | 1 | 1 |
| 1971 | 1 | 10 | 0 | 0.999 | 0.001 | 0 | 1 | 1 |

(v) By constructing appropriate tables, check whether the local and global decoding give the same results

The following extract from R (edited in excel for presentation) shows that the local decoding and global decoding do differ in some ways. The biggest difference is the occurrence of State 2. In Global decoding State 2 occurs 832 times, whereas in Local decoding State 2 occurs 1916 times (over twice as often).

> table(ld)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Local decoding** | | |  | |  | |
|  | |  | |  | |
|  | 1 | 2 | | 3 | |
| Freq | 8151 | 1916 | | 4543 | |
| *Rel Freq* | *0.5579* | *0.1311* | | *0.311* | |

> table(gd)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Global decoding** | | |  | |  | |
|  | |  | |  | |
|  | 1 | 2 | | 3 | |
| Freq | 8567 | 832 | | 5211 | |
| *Rel Freq* | *0.5864* | *0.0569* | | *0.3567* | |

> table(ld,gd)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | gd |  |
|  |  | 1 | 2 | 3 |
|  | 1 | 8119 | 32 | 0 |
| ld | 2 | 448 | 743 | 725 |
|  | 3 | 0 | 57 | 4486 |

The state that gives the median value of is state 2.

According to the Global Decoding, this state occurs at the following times.

From 15/6/71 to 28/11/71,

29/6/80 to 15/8/80,

24/5/81 to 22/8/81,

16/8/86 to 1/11/86,

23/5/89 to 15/10/89,

3/6/91 to 27/9/91,

10/8/92 to 10/10/92,

5/6/96 to 5/10/96

*Characterize each of the three hidden states. [5]*

The following tables show whether it is raining or not in each of the three states.

It can be seen that there is no rain in state 1. Hence we can characterise this state as “Dry Days”.

There is only rain in state 3. Hence this state can be characterised as “Wet Days”.

Closely examining the data gives the impression that State 2 could be considered the state where weather is unusually changeable.

> table(Rain,gd)

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  |
| Global Decoding | |  |  |
| Rain | 1 | 2 | 3 |
| 0 | 8567 | 311 | 0 |
| 1 | 0 | 521 | 5211 |

> table(Rain,ld)

|  |  |  |  |
| --- | --- | --- | --- |
| Local Decoding | |  |  |
| Rain | 1 | 2 | 3 |
| 0 | 8151 | 727 | 0 |
| 1 | 0 | 1189 | 4543 |

State 1 – Dry days

State 2 – Unusually changeable periods

State 3 – Wet Days