STOCHASTIC MODELING OF ENVIRONMENTAL TIME SERIES

Richard W. Katz

LECTURE 5

(1) Hidden Markov Models: Applications

(2) Hidden Markov Models: Viterbi Algorithm

(3) Non-Homogeneous Hidden Markov Model

(1) Hidden Markov Models: Applications

- **Duration of Old Faithful Eruptions**
- **Snoqualmie Falls Precipitation Occurrence**
- **Great Plains Precipitation Occurrence**
- **Other Applications**

Duration of Old Faithful Eruptions

- **Data**

-- Recall data set of eruptions of Old Faithful (fit mixture of two normal distributions to waiting time between eruptions)

-- Data set also contains duration of eruptions

299 eruptions (X_t = 0 or 1):

State 0 ("short"): 105 eruptions < 3 min.

State 1 ("long"): 192 eruptions > 68 min.

(2 eruptions of intermediate duration: with 3 ≤ duration ≤ 68; by convention, count as state 1)

- **Results**

-- Fitting binary hidden Markov model

Assume hidden Markov model has only two states

(i) Starting values

-- Used method of moments estimates (taking $X_t \approx J_t$)

$\Pr\{X_t = 1)$: 194/299 = 0.649

$\text{Corr}(X_t, X_{t+1})$: − 0.538

$\delta = 0.65$, $P_{01} = 0.9$, $P_{11} = 0.35$, $p_0 = 0.01$, $p_1 = 0.99$

(ii) Forward probability recursion (for starting values)

$$a_t(i) = \Pr\{X_1 = x_1, \ldots, X_t = x_t, J_t = i\}, \ i = 0, 1; \ t = 1, 2, \ldots, T$$

T	x_t	$a_t(0)$	$a_t(1)$
1	1	0.003500	0.643500
2	0	0.414439	0.002284
3	1	0.000429	0.370056
.	.	.	.
.	.	.	.
.	.	.	.

(iii) Backward probability recursion (for starting values)

$$b_t(i) = \Pr\{X_{t+1} = x_{t+1}, \ldots, X_T = x_T \mid J_t = i\}, \;\; i = 0, 1; \;\; t = T, T-1, \ldots, 1$$

T	x_t	$b_t(0)$	$b_t(1)$
299	0	1.000000	1.000000
298	1	0.108000	0.647000
297	1	0.576585	0.224888
296	0	0.200095	0.081671
.	.	.	.
.	.	.	.
.	.	.	.

(iv) Parameter estimates (EM algorithm)

Iter.	δ	p_0	p_1	P_{01}	P_{11}	log like
1	0.6500	0.0100	0.9900	0.9000	0.3500	−150.136
2	0.9991	0.0240	0.9979	0.9979	0.4411	−133.414
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
30	1.0000	0.2246	1.0000	1.0000	0.1722	−126.7079
31	1.0000	0.2248	1.0000	1.0000	0.1719	−126.7078

- **Model Identification**

-- Use results in MacDonald & Zucchini book

(Different assumption about initial state of hidden Markov model; i.e., $\delta = \pi$)

Model	No. Par.	Log like	AIC	BIC
Independence	1	−193.80	389.60	393.31
First-order Markov chain	2	−134.24	272.48	279.88
Hidden Markov model	4	−127.31	262.62	277.42
Second-order Markov chain	4	−127.12	262.24	277.04

- **Model Properties**

-- Autocorrelation function

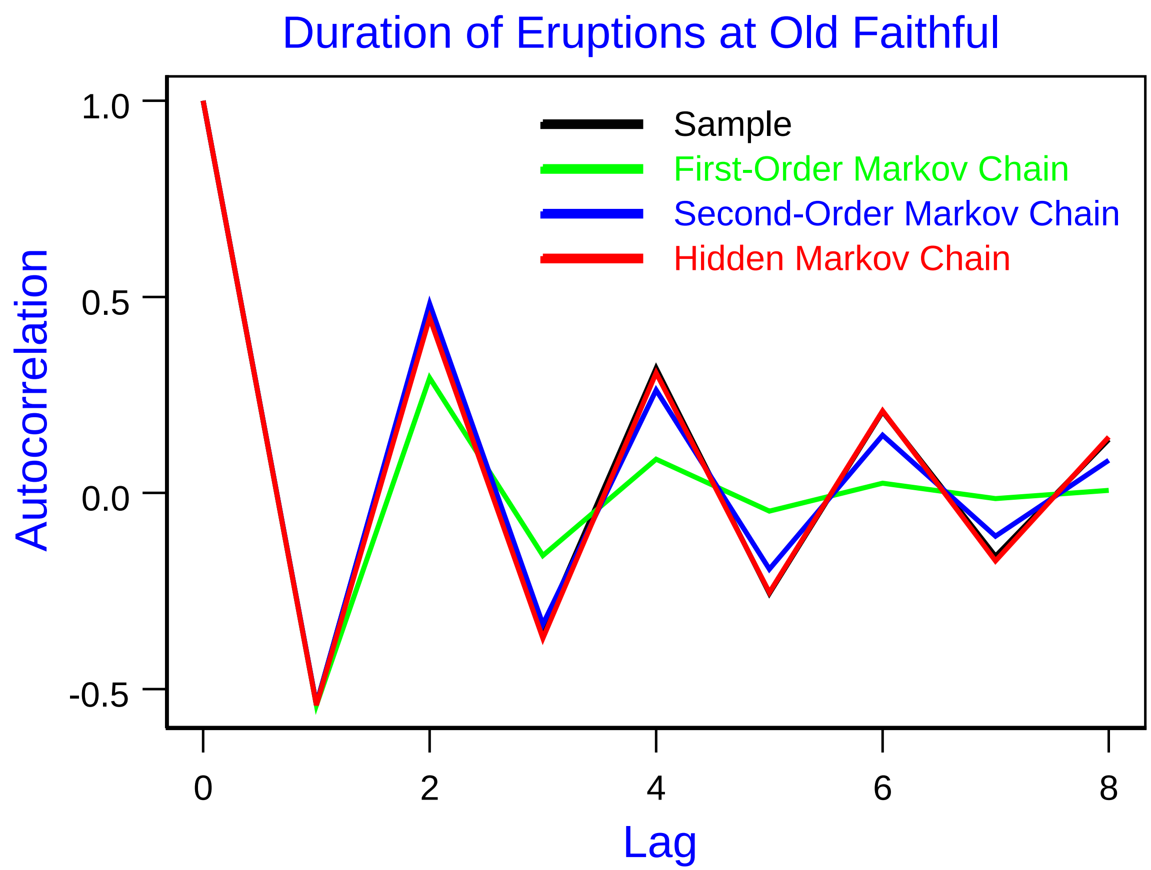
Have expressions for case of:

(1) First-order Markov chain (two-state)

(2) Second-order Markov chain (two-state)

Analogous to first-order Markov chain

(3) Hidden Markov model (binary, two-state)



Duration of Eruptions at Old Faithful
Sample
First-Order Markov Chain
Second-Order Markov Chain
Hidden Markov Chain
Autocorrelation
1.0
0.5
0.0
-0.5
0
2
4
6
8
Lag

Snoqualmie Falls Precipitation Occurrence

- **Data**

-- **Snoqualmie Falls, WA, USA**

-- **Time series of daily precipitation occurrence ($X_t = 0$ or 1)**

-- **January – March, 1948 – 1983 (90 values per yr, 36 yrs)**

-- **Assume each year represents independent time series**

- **Model Fitting**

-- Direct numerical optimization (not EM algorithm) used by Guttorp (1995)

-- Parameter estimates (standard errors from Hughes, 1997)

Parameter	Estimate	Standard error
P_{01}	0.3256	0.03008
P_{11}	0.8578	0.01549
p_0	0.0592	0.03578
p_1	0.9419	0.01614

- **Model Identification**

Model	No. Par.	Log like	AIC	BIC
First-order Markov chain	2	−898.32	1800.64	1812.81
Hidden Markov model	4	−895.10	1798.20	1822.53

- **Model Properties**

-- Distribution of length of dry spells

Geometric for Markov Chain: Underestimates observed frequency (for longer spells)

Hidden Markov model apparently fits better (figure in Guttorp book, p. 108)

Great Plains Precipitation Occurrence

- **Data**

-- **Time series of daily precipitation occurrence**

-- **3 sites in US (Omaha, NE; Des Moines, IO; Grand Island, NE)**

-- **Divided into 6 seasons of about 2 months length (most of length 60 days)**

-- **Period of record: 1949 – 1984 (i.e., 36 yrs)**

- **Model**

-- $X_t(n) = 1$ **if precipitation occurs on** t**th day at** n**th site, n = 1, 2, 3**

-- Assume two-state hidden Markov model $\{J_t\}$

-- Given common hidden state, assume precipitation occurrence at given site conditionally independent of occurrence at other sites

-- Assume each year represents independent time series

-- Let $p_i^{[n]} = \Pr\{X_t(n) = 1 \mid J_t = i\}$, $i = 0, 1;\ n = 1, 2, 3$

- **Model Fitting**

-- Direct numerical optimization (not EM algorithm) used by Guttorp (1995)

-- Parameter estimates (Season 1: Jan-Feb.)

Markov chain: $P_{01} = 0.352$, $P_{11} = 0.665$

Hidden state J_t	$p_i^{[1]}$	$p_i^{[2]}$	$p_i^{[3]}$
$J_t = 0$	0.927	0.874	0.762
$J_t = 1$	0.039	0.211	0.139

- **Model Properties**

-- Common hidden states can induce unconditional spatial dependence (even if assume conditional independence)

-- Table in Guttorp book (p. 109)

Joint probability of precipitation occurrence at pairs of sites (and at all 3 sites) reproduced well by hidden Markov model

-- But not always case that sufficient spatial dependence can be induced

See Zucchini and Guttorp, 1991 paper: modeled additional sites for same region

Other Applications

- **Rainfall Rate** **(Sansom, 1998)**

-- Data essentially instantaneous measurements of precipitation "rate"

-- Fit hidden Markov models to precipitation rate time series (using EM algorithm)

-- Find strong evidence of existence of "breakpoints" between different precipitation rates

-- Able to provide physical interpretation to hidden states

(2) Hidden Markov Models: Viterbi Algorithm

- **Stochastic Dynamic Programming**
- **Viterbi Algorithm**
- **Example**

Stochastic Dynamic Programming

- **Background**

-- **Dynamic optimization problem**

-- **Stochastic (with underlying Markovian structure)**

-- **Use "backwards induction"**

Start with single occasion/ "static" problem

Solve optimization problem by recursion

Viterbi Algorithm

- **"Global Decoding"**

-- Want to determine sequence of states of hidden Markov chain ($i_1, i_2, \ldots, i_T$) maximizing conditional probability:

$$\Pr\{J_1 = i_1, J_2 = i_2, \ldots, J_T = i_T \mid X_1 = x_1, X_2 = x_2 \ldots, X_T = x_T\}$$

-- Initial step

Set

$$\xi_1(i) = \Pr\{J_1 = i, X_1 = x_1\},\ i = 0, 1$$

-- Recursion

Set

$$\xi_t(i) = \max_{i_1, \ldots, i_{t-1}} \Pr\{J_1 = i_1, \ldots, J_{t-1} = i_{t-1}, J_t = i, X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t\}$$

$t = 2, 3, \ldots, T$

Then

$$\xi_{t+1}(j) = \max\{\xi_t(0)P_{0j}, \xi_t(1)P_{1j}\}\, p_j(x_{t+1})$$

$j = 0, 1;\;\; t = 1, 2, \ldots, T-1$

-- **Required states (i_1^*, i_2^*, . . ., i_T^*)**

(1) $t = T$

$$i_T^* = 1 \text{ if } \xi_T(0) < \xi_T(1),$$

$$i_T^* = 0 \text{ otherwise}$$

(2) $t = T - 1, T - 2, \ldots, 1$

$$i_t^* = 1 \text{ if } \xi_t(0)P(0, i_{t+1}^*) < \xi_t(1)P(1, i_{t+1}^*),$$

$$i_t^* = 0 \text{ otherwise}$$

[writing $P_{ij} = P(i, j)$]

Example

- **Data**

-- Duration of Old Faithful Eruptions

Use parameter estimates from termination of EM algorithm (i.e., iteration no. 31)

Recall posterior probabilities for "local decoding" produced automatically

- **Results**

Time	Observed State	Local decoding	Global decoding
1	1	1 (1.0000)	1
2	0	0 (1.0000)	0
3	1	1 (1.0000)	1
4	0	0 (0.8625)	0
.	.	.	.
.	.	.	.
162	1	0 (0.5895)	0
.	.	.	.
.	.	.	.
299	0	0 (1.0000)	0

(3) Non-Homogeneous Hidden Markov Models

- **Basic Non-Homogeneous Model**

-- **Hughes et al. (1999)**

- **Extensions**

-- **Conditional dependence (Hughes et al., 1999)**

-- **Inclusion of precipitation intensity (Bellone et al., 2000)**

Non-Homogeneous Model

- **Observed covariates**

-- In addition to hidden Markov model

-- Transition probabilities of hidden Markov model depend on covariates

-- Example

Network of sites at which precipitation occurrence modeled (as in Great Plains example)

But now observed meteorological covariates available as well (daily time scale)

- **Model**

-- Notation

Observed variable: $X_t(n) = 1$ **if precipitation occurs on** t**th day at** n**th site**

(Set $\mathbf{X}_t = [X_t(1), X_t(2), \ldots]$**)**

Hidden Markov chain: $\{J_t\}$

Covariates: $\mathbf{Z}_t$ **vector of atmospheric variables at time** t, $t = 1, 2, \ldots, T$

-- Concept

Hidden variable ("weather state") acts as link between two disparate scales (small-scale precipitation and larger-scale atmospheric circulation patterns)

-- Assumption 1

Precipitation occurrence process is conditionally independent given current hidden state:

$$\Pr\{X_t \mid J_1, \ldots, J_T, X_1, \ldots, X_{t-1}, Z_1, \ldots, Z_T\} = \Pr\{X_t \mid J_t\}$$

-- Assumption 2

Given history of hidden state up to time $t - 1$ and entire past and future sequence of atmospheric data, hidden state at time t depends only on previous hidden state and current atmospheric data

$$\Pr\{J_t \mid J_1, \ldots, J_{t-1}, Z_1, \ldots, Z_T\} = \Pr\{J_t \mid J_{t-1}, Z_t\}$$

-- Conditional independence

Retain (for now) assumption that precipitation occurrence is conditionally spatially independent given hidden state

(i.e., as model for $\Pr\{\mathbf{X}_t \mid J_t\}$)

-- Atmospheric variables

Assume multivariate normal distribution

-- Ordinary "homogenous" hidden Markov model is special case

- **Parameter Estimation**

-- EM algorithm

(not feasible to resort to direct maximization of likelihood function)

-- Model selection

Used BIC & physical interpretability of hidden states

- **Application**

-- Precipitation data

15-yr record (1978 – 1992) of daily winter precipitation occurrences at 30 stations in SW Australia

-- Atmospheric variables (gridded data)

Sea-level pressure, geopotential height, air temperature, dew point, wind speed components

-- Interpretation of hidden states

Each day classified into its most likely state using Viterbi algorithm

Extensions

- **Conditional dependence**

-- Relaxing assumption of conditional spatial independence

-- Replace with autologistic model for multivariate binary model

-- Complicates EM algorithm

Both E-step and M-step of EM algorithm become computationally intractable [used EM-MCML (Monte Carlo Maximum Likelihood) algorithm instead]

-- Hughes et al. (1999) applied model to same Australia data set (Improvement in induced spatial dependence in precipitation occurrence)

- **Inclusion of precipitation intensity (Bellone et al., 2000)**

-- Non-homogeneous hidden Markov model

-- Retain conditional spatial independence assumption for precipitation occurrence

-- Given precipitation occurrence, assume precipitation intensity can be modeled as spatially conditionally independent with gamma distribution

-- Used EM algorithm for parameter estimation

-- Used BIC for model identification

-- Used Viterbi algorithm in conjunction with interpretation of hidden states

-- Applied to precipitation occurrence at 24 sites across Washington state, USA (Atmospheric data: pressure heights, temperature, relative humidity)