Hidden Markov model forecasting of earthquakes

Daniel W. Chambers, John E. Ebel, Jenny Baglivo, Alan L. Kafka, Boston College

A HMM consists of a sequence of states and a sequence of observations.

$X_1, X_2, \ldots, X_L$ are state variables.

$Y_1, Y_2, \ldots, Y_L$ are observation variables.

Distribution of Y_t depends on value of X_t.

$X_1, X_2, \ldots, X_L$ are unknown (hidden).

$Y_1, Y_2, \ldots, Y_L$ are observed.

Data= interevent times of L=1227 earthquakes in southern California 1932-2004

Assumptions

SCal region switches between three states (1,2,3)= tendency for (short, moderate, long) time between earthquakes.

$X_1, X_2, \ldots, X_L$ is a Markov chain with initial distribution π and transition probability matrix $A: \; P(X_{t+1} = j | X_t = i) = A_{ij}$.

$Y_1, Y_2, \ldots, Y_L$ are the actual interevent times: $Y_t = \#$ days between earthquakes $t-1$ and t.

Given $X_t = i$, Y_t has an exponential distribution with mean λ_i; $\lambda_1 < \lambda_2 < \lambda_3$.

Goals: estimation and forecasting

(1) Use the data to estimate the parameter set
$\Theta = (\pi, A, \lambda_1, \lambda_2, \lambda_3)$

(2) Given model parameters and interevent times for first t earthquakes, find conditional density of time until next earthquake,

$$P(Y_{t+1} = y \mid y_1, \ldots, y_t, \Theta)$$

Estimation

Estimate Θ using Baum-Welch algorithm:

For time t and state i define forward and backward variables:

$$f_i(t) = P(y_1, \ldots, , y_t, X_t = i|\Theta)$$
$$b_i(t) = P(y_{t+1}, \ldots, y_L|X_t = i, \Theta)$$

These satisfy recursions and can be computed efficiently, given initial parameters Θ. Updated parameter estimates are found from $f_i(t)$ and $b_i(t)$ values and process is repeated; likelihood of observations increases and iterations continue until likelihood converges.

Likelihood of observations $P(y_1, \ldots, y_L|\Theta) = \sum_{i=1}^{3} P(y_1, \ldots, y_L, X_L = i|\Theta) = \sum_{i=1}^{3} f_i(L)$.

Forecasting

At time of t^{th} earthquake, we've seen $y_1, \ldots, y_t$, the times between the first t earthquakes. The time to the next will be Y_{t+1}.

Goal: Given a HMM with parameters Θ and the observations up to the present, find the forecast density

$$P(Y_{t+1} = y \mid y_1, \ldots, y_t, \Theta)$$

Solution: condition on the states X_t, X_{t+1} for earthquakes t and $t+1$. Then

$$P(Y_{t+1} = y \mid y_1, \ldots, y_t, \Theta) = \sum_{j=1}^{3} (1/\lambda_j) e^{-y/\lambda_j} c_{jt},$$

where $c_{jt} = P(X_{t+1} = j \mid y_1, \ldots, y_t, \Theta)$.

That is, forecast density at time t is a weighted sum of the three exponential densities, with weights the conditional probabilities of being in states $1, 2$, or 3 at time $t+1$. These are computable for a given model.

Results

1227 earthquakes in S. California 1932-2004.

2 state and 3 state models were fit.

Model 1 means: $\lambda_1 = 3.29, \; \lambda_2 = 23.32$ days.

Model 2: $\lambda_1 = 1.30, \lambda_2 = 17.42, \lambda_3 = 27.92$

After each earthquake in the catalog, forecast density was found, based on earthquakes seen to that point and integrated to find forecast probability of another earthquake within 7 days.

Forecast results

1226 forecasts made, grouped as follows. For each group of forecasts, catalog was checked and proportion of times an earthquake did occur within 7 days computed:

	[.28,.32)	[.32,.36)	[.36,.50)	[.50,1]
Model 1	.285	.352	.363	.444
Model 2	.292	.300	.421	.625

(For example, under Model 1, of those forecasts for which probability of an earthquake within 7 days fell in range $[.28, .32)$, one did occur in 28.5% of the cases.

Current Work:

- incorporating magnitudes and locations in state variables and observations (magnitude observations given as M4-M4.9 or M5+; location in one of four quadrants).

- After an earthquake, compute forecast probabilities of an earthquake within a given time for each magnitude/location value.

- Scheduled forecasts (daily, weekly) based on seismic history available including elapsed waiting time since previous earthquake.