Inference for stochastic processes in environmental science

I: Hidden Markov models and precipitation

Peter Guttorp
NRCSE

peter@stat.washington.edu

Collaborators

Enrica Bellone, NCAR
Tamre Cardoso, NRCSE
Stephen Charles, CSIRO
Jim Hughes, NRCSE
Ted Lystig, UW Biostat
Outline

Markov chain
Hidden Markov chain
Nonstationary transition probabilities
Spatial correlation
Measurement error
A state space model for combining data

Why model rainfall?

Input to hydrologic models
Downscaling of general circulation models
Agricultural forecasting
A Markov chain model

\[(R_t \mid R_{t-1}, \ldots, R_1) \sim p_{R_{t-1}, R_t}\]

January data from Snoqualmie Falls, Washington, 1948-1983
325 dry and 791 wet days

<table>
<thead>
<tr>
<th></th>
<th>Today wet</th>
<th>Today dry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yesterday wet</td>
<td>643 ( (643) )</td>
<td>128 ( (223) )</td>
</tr>
<tr>
<td>Yesterday dry</td>
<td>123 ( (223) )</td>
<td>186 ( (91) )</td>
</tr>
<tr>
<td></td>
<td>766</td>
<td>314</td>
</tr>
</tbody>
</table>
Survival function

\[ S(t) = 1 - \Pr(\text{Dry period} \leq t) \]

Dry period \sim \text{Geom}(1/(1-p_{00}))

A spatial Markov model

Three sites, A, B and C, each observing 0 or 1. Notation: AB = (A=1,B=1,C=0)

Markov model:

\[ P(X_t = (X_{A,t}, X_{B,t}, X_{C,t}) = (i,j,k) | X_{t-1} = (l,m,n), ..., X_1) = p_{ijn,ijk} \]

Great Plains data 1949-1984 (Jan-Feb)

<table>
<thead>
<tr>
<th></th>
<th>Dry</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>718</td>
<td>1020</td>
<td>1154</td>
<td>957</td>
<td>866</td>
<td>752</td>
<td>728</td>
<td>657</td>
</tr>
<tr>
<td>MC</td>
<td>722</td>
<td>942</td>
<td>1076</td>
<td>1031</td>
<td>789</td>
<td>750</td>
<td>727</td>
<td>655</td>
</tr>
</tbody>
</table>
A hidden weather state

Two-stage model

\[ C_t \text{ Markov chain, c states} \]

\[ (R_t | C_t, R_{t-1}, C_{t-1}, ..., C_1, R_1) = (R_t | C_t) = \pi_t(C_t) \]

We observe only \( R_1, ..., R_T \).

C clusters similar rainfall patterns. In atmospheric science called a weather state

Likelihood

\[ L(\theta) = \sum_{c \in C} \prod_{i=1}^{T} p(r_i | c_i; \theta)p(c_i; \theta) \]

|C| = 3, \|C\| = 5.2 \times 10^{47}

Forward algorithm: unravel sum recursively

\[ \alpha_t(j) = \Pr(R_1, ..., R_t, C_t = j) \]

\[ = \sum_{i=1}^{\|C\|} \alpha_{t-1}(i)\pi_{t}(j)p_{i,j} \]

\[ L(\theta) = \sum_{j=1}^{\|C\|} \alpha_T(j) \]
Computational algorithm

Lystig (2001): Write

\[
L(\theta) = \Pr(R_1, \ldots, R_T; \theta) = \prod_{t=1}^{T} \Pr(R_t | R_{t-1}, \ldots, R_1; \theta)
\]

\[
\lambda_t(j) = \Pr(R_t, C_t = j | R_{t-1}, \ldots, R_1) = \sum_{i=1}^{[C]} \frac{\lambda_{t-1}(i)p_t(j)p_{i,j}}{\Lambda_{t-1}}
\]

\[
\Lambda_t = \sum_{j=1}^{[C]} \lambda_t(j) = \Pr(R_t | R_{t-1}, \ldots, R_1)
\]

\[
\ell(\theta) = \log L(\theta) = \sum_{t=1}^{T} \log \Lambda_t
\]

Estimating standard errors

The Lystig recursions enable easy calculation of first and second derivatives of the log likelihood, which can be used to estimate standard errors of maximum likelihood estimates of \( \theta \).
Snoqualmie Falls

Two-state hidden model
Pr(rain) = 0.059 (0.036) in state 1
0.941(0.016) in state 2
\( \hat{p}_{11} = 0.674 (0.030) \)
\( \hat{p}_{22} = 0.858 (0.016) \)

est corr =
\[
\begin{pmatrix}
1 & -0.40 & 0.24 & -0.71 \\
-0.40 & 1 & -0.79 & 0.01 \\
0.24 & -0.79 & 1 & 0.16 \\
-0.71 & 0.01 & 0.16 & 1
\end{pmatrix}
\]

Survival function
The spatial case

MC: 8 states, 56 parameters
HMM: 2 hidden states (one fairly wet, one fairly dry), 8 parameters, rain conditionally independent at different sites given weather state

<table>
<thead>
<tr>
<th></th>
<th>Dry</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>718</td>
<td>1020</td>
<td>1154</td>
<td>957</td>
<td>866</td>
<td>752</td>
<td>728</td>
<td>657</td>
</tr>
<tr>
<td>HMM</td>
<td>725</td>
<td>1019</td>
<td>1153</td>
<td>956</td>
<td>862</td>
<td>749</td>
<td>728</td>
<td>657</td>
</tr>
<tr>
<td>MC</td>
<td>722</td>
<td>942</td>
<td>1076</td>
<td>1031</td>
<td>789</td>
<td>750</td>
<td>727</td>
<td>655</td>
</tr>
</tbody>
</table>

Nonstationary transition probabilities

Meteorological conditions may affect transition probabilities

\[
\log \left( \frac{p_{ij}(t)}{1 - p_{ij}(t)} \right) = \alpha_{ij} + \beta_{ij}^T A_t
\]
A model for Western Australia rainfall

daily rainfall at 30 stations
Atmospheric variables in model: E-W gradient in 850 hPa geopotential height,
mean sea level pressure, N-S gradient in sea-level pressure
Final model has six weather states (BIC and other diagnostics)
Spatial dependence

Model assessment
Rainfall measurement

Rain gauge (1 hr)
- High wind, low rain rate (evaporation)
- Spatially localized, temporally moderate

Radar reflectivity (6 min)
- Attenuation, not ground measure
- Spatially integrated, temporally fine

Cloud top temp. (satellite, ca 12 hrs)
- Not directly related to precipitation
- Spatially integrated, temporally sparse

Distrometer (drop sizes, 1 min)
- Expensive measurement
- Spatially localized, temporally fine

Radar image
Drop size distribution

n = 1496

n = 7180

n = 5426

n = 1866
Basic relations

Rainfall rate:
\[ R(t) = c_R \pi \int_0^{\infty} D^3 v(D) N(t) f(D) dD \]
v(D) terminal velocity for drop size D
N(t) number of drops at time t
f(D) pdf for drop size distribution

Gauge data:
\[ G(t) \sim N \left( g(w(t)) \int_{t-\Delta}^{t} R(s) ds, \sigma_G^2 \right) \]
g(w) gauge type correction factor
w(t) meteorological variables such as wind speed

Basic relations, cont.

Radar reflectivity:
\[ Z_D(t) = c_Z \int_0^{\infty} D^6 v(D) N(t) f(D) dD \]

Observed radar reflectivity:
\[ Z(t) \sim N(Z_D(t), \sigma_Z^2) \]
Structure of model

Data: \([G|N(D), \theta_G] \quad [Z|N(D), \theta_Z]\)

Processes: \([N|\mu_N, \theta_N] \quad [D|\xi_D, \theta_D]\)

\[\text{log GARCH} \quad \text{LN}\]

Temporal dynamics: \([\mu_N|t, \theta_\mu]\)

\[\text{AR}(1)\]

Model parameters: \([\theta_G, \theta_Z, \theta_N, \theta_\mu, \theta_D | \theta_H]\)

Hyperparameters: \(\theta_H\)
An early result

By estimating the (hidden) state, we can estimate rain rate with or without data.

Future work

Spatial models for continental-scale rainfall
Prediction models for areal rainfall
Spatial models for rainfall measurement
References

Hidden Markov models

Precipitation models

Bayesian hierarchic models