A Location-Mixture Autoregressive Model for Online Forecasting of Lung Tumor Motion

Dan Cervone, Natesh Pillai, Debdeep Pati, Ross Berbecto, John H. Lewis

Harvard Statistics Department and Harvard Medical School

August 6, 2014
External beam radiotherapy:

- Lung tumor patients are given an implant (fiducial) at the location of their tumor.
- X-ray tomography can reveal location of the fiducial, thus the tumor.
- Radiotherapy is applied to the tumor location in a narrow beam, minimizing exposure within healthy tissue.
Introduction

Fiducial

External beam radiotherapy
Introduction

Our problem:

- Patient’s respiration means the tumor is in constant motion.
- Tracking the fidual lags 0.1-2s behind, depending on specific machinery used.
- **We need to forecast the location of the tumor to overcome this latency and ensure concentrated, accurate radiotherapy.**
The data

We consider data from 11 patients:

- multiple days receiving treatment.
- multiple radiotherapy sessions ("beams") per treatment.

![Graph showing tumor motion data](image-url)
The data

We consider data from 11 patients:
- multiple days receiving treatment.
- multiple radiotherapy sessions ("beams") per treatment.
Semiperiodic features

- Signals at different frequencies.
- Fluctuations in location/amplitude/periodicity.
- Repeated motifs.
Location-mixture autoregressive process

Assume $Y_i, i = 0, \ldots, n$ is a time series in $\mathbb{R}$, and for some $p$,

$$Y_n|Y_{n-1}, Y_{n-2}, \ldots \sim \sum_{j=1}^{d_n} \alpha_{n,j} N(\mu_{n,j}, \sigma^2)$$

$$\mu_{n,j} = \tilde{\mu}_{n,j} + \sum_{\ell=1}^{p} \gamma_{\ell} Y_{n-\ell}$$
Location-mixture autoregressive process

Assume $Y_i, i = 0, \ldots, n$ is a time series in $\mathbb{R}$, and for some $p$,

$$Y_n | Y_{n-1}, Y_{n-2}, \ldots \sim \sum_{j=1}^{d_n} \alpha_{n,j} N(\mu_{n,j}, \sigma^2)$$

$$\mu_{n,j} = \tilde{\mu}_{n,j} + \sum_{\ell=1}^{p} \gamma_{\ell} Y_{n-\ell}$$

- $\alpha_{n,j}$ are mixture weights: $\sum_{j=1}^{d_n} \alpha_{n,j} = 1$.
- Mixture means have autoregressive component $\sum_{\ell=1}^{p} \gamma_{\ell} Y_{n-\ell}$.
- Mixture means have location shift component $\tilde{\mu}_{n,j}$.
Location-mixture autoregressive process

Let $\mathbf{Y}_{i-1:p}$ be the subseries $(Y_{i-1} \ldots Y_{i-p})'$, and define mixture components:

$$\alpha_{i,j} = \frac{\exp \left[ -\frac{1}{2} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p})' \Sigma^{-1} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p}) \right]}{\sum_{\ell \leq i-p} \exp \left[ -\frac{1}{2} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})' \Sigma^{-1} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p}) \right]}$$

$$\mu_{i,j} = Y_{i-j} - \gamma' \mathbf{Y}_{i-j-1:p} + \gamma' \mathbf{Y}_{i-1:p}$$

Or, using latent variables,

$$Y_{i|M_i = j, Y_{i-1}, \ldots, Y_0} \sim N(Y_{i-j} + \gamma'(\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-j-1:p}), \sigma^2)$$

$$P(M_i = \ell|Y_{i-1}, \ldots, Y_0) \propto$$

$$\exp \left[ -\frac{1}{2} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p})' \Sigma^{-1} (\mathbf{Y}_{i-1:p} - \mathbf{Y}_{i-\ell-1:p}) \right]$$
Location-mixture autoregressive process

Latent variables $M_i$ instantiate time series motifs: let

$$\Omega = \begin{pmatrix} \tau^2 & \gamma' \\ \gamma & \Sigma \end{pmatrix},$$

assuming $Y_i|\omega|M_i=j, Y_i-1,\ldots$ yields the LMAR predictive distribution for $Y_i|M_i=j$. 
Location-mixture autoregressive process

Latent variables $M_i$ instantiate time series motifs: let

$$
\Omega = \begin{pmatrix}
\tau^2 & \gamma' \\
\gamma & \Sigma
\end{pmatrix},
$$

then $M_i = j$ implies

$$
(Y_{i-0:p} - Y_{i-j-0:p})' \Omega^{-1} (Y_{i-0:p} - Y_{i-j-0:p}) = \text{small}.
$$
Location-mixture autoregressive process

Latent variables $M_i$ instantiate time series motifs: let

$$\Omega = \begin{pmatrix} \tau^2 & \gamma' \\ \gamma & \Sigma \end{pmatrix},$$

then $M_i = j$ implies

- $$(Y_{i-0:p} - Y_{i-j-0:p})' \Omega^{-1} (Y_{i-0:p} - Y_{i-j-0:p}) = \text{small}.$$  
- Assuming

$$Y_{i-0:p} | M_i = j, Y_{i-j-0:p} \sim N(Y_{i-j-0:p}, \Omega)$$

yields the LMAR predictive distribution for $Y_i | M_i = j, Y_{i-1}, \ldots$
Illustration of time series motifs

- $Y_{i-1:p}$ most recent $p$ values of times series.
- $Y_{i-j-1:p}$ previous instances of this motif.
Illustration of time series motifs

- $Y_{i-1:p}$ most recent $p$ values of times series.
- $Y_{i-j-1:p}$ previous instances of this motif.
Parameter estimation

We focus on estimating $\Omega$:

- 1-1 correspondence between $\Omega$ and LMAR parameters $\gamma, \Sigma, \sigma^2$.

Latent variables $M_i$ yield estimating equation:

$$h(\Omega) = \sum_{i \geq p} \frac{1}{2} \log(|\Omega|) - \frac{1}{2} \sum_{j \leq i - p} 1[M_i = j](Y_{i-0:p} - Y_{i-j-0:p})' \Omega^{-1}(Y_{i-0:p} - Y_{i-j-0:p})$$

- $\hat{\Omega} = \arg\max(h(\Omega))$ computable quickly using EM algorithm.
- Not loglikelihood, but $\hat{\Omega}$ enjoys same large-sample properties as MLE.
Given a point estimate $\hat{\Omega}$, $k$-step ahead predictive distributions can be approximated by marginalizing over future steps 1 to $(k - 1)$:

$$Y_{i+k} \mid Y_i, Y_{i-1}, \ldots \sim \sum_j \alpha_j^k N(\mu_j^k, \sigma_k^2).$$
Given a point estimate \( \hat{\Omega} \), \( k \)-step ahead predictive distributions can be approximated by marginalizing over future steps 1 to \((k - 1)\):

\[
Y_{i+k} | Y_i, Y_{i-1}, \ldots \sim \sum_j \alpha_j^k N(\mu_j^k, \sigma_j^2).
\]

- Mixture components \( \alpha_j^k, \mu_j^k, \sigma_j^2 \) available cheaply, analytically.
- No need for sampling or Monte Carlo.
Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values
Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values

- Predictions must be made in real time (30Hz)!
Predictive performance on clinical data

We tested predictive performance using the following procedure for all clinical observations in our data set.

1. Train model on first 35s of data
2. Fit prediction model in under 5 seconds
3. Simulate predictions for the next 40 seconds of data using model fit
4. Compare predictions with observed values

- Predictions must be made in real time (30Hz)!

Alternative prediction models considered:

- Our model: LMAR
- Neural networks
- Penalized AR model
- LICORS (Georg and Shalizi, 2012)

For all methods, any tuning parameters were set to patient-independent optimal values, using separate data.
Example 1 \((k = 6)\)
Example 2 ($k = 6$)

![Graph of training data and forecasting models](image)

- **TRAINING DATA**
- **LMAR**
- **NN**
- **PENAL. AR**
- **LICORS**

Dan Cervone (Harvard)
Quantitative summaries (point prediction)

For all methods, predictive performance varies by patient, but some examples:

<table>
<thead>
<tr>
<th>Pat.</th>
<th>Method</th>
<th>0.2s forecast ($k = 6$)</th>
<th>0.4s forecast ($k = 12$)</th>
<th>0.6s forecast ($k = 18$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>MAE</td>
<td>RMSE</td>
</tr>
<tr>
<td>9</td>
<td>LMAR</td>
<td>0.58</td>
<td>0.22</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>0.73</td>
<td>0.32</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>0.81</td>
<td>0.34</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>1.35</td>
<td>0.53</td>
<td>2.20</td>
</tr>
<tr>
<td>10</td>
<td>LMAR</td>
<td>0.88</td>
<td>0.36</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>1.09</td>
<td>0.44</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>0.95</td>
<td>0.45</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>1.62</td>
<td>0.61</td>
<td>2.20</td>
</tr>
<tr>
<td>11</td>
<td>LMAR</td>
<td>1.13</td>
<td>0.44</td>
<td>2.59</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>1.24</td>
<td>0.50</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>1.19</td>
<td>0.63</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>1.64</td>
<td>0.57</td>
<td>3.04</td>
</tr>
</tbody>
</table>
## Quantitative summaries (interval prediction)

### Summaries of interval/distributional predictions:

<table>
<thead>
<tr>
<th>Patient</th>
<th>Method</th>
<th>Coverage</th>
<th>Log PS</th>
<th>Coverage</th>
<th>Log PS</th>
<th>Coverage</th>
<th>Log PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>LMAR</td>
<td>0.89</td>
<td>0.87</td>
<td>0.90</td>
<td>1.65</td>
<td>0.92</td>
<td>2.07</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>0.86</td>
<td>1.02</td>
<td>0.78</td>
<td>2.20</td>
<td>0.80</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>0.81</td>
<td>1.54</td>
<td>0.81</td>
<td>2.21</td>
<td>0.81</td>
<td>2.64</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>0.86</td>
<td>1.62</td>
<td>0.81</td>
<td>1.98</td>
<td>0.79</td>
<td>2.31</td>
</tr>
<tr>
<td>10</td>
<td>LMAR</td>
<td>0.86</td>
<td>1.18</td>
<td>0.88</td>
<td>1.94</td>
<td>0.91</td>
<td>2.33</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>0.84</td>
<td>1.23</td>
<td>0.76</td>
<td>2.25</td>
<td>0.79</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>0.83</td>
<td>1.35</td>
<td>0.84</td>
<td>2.03</td>
<td>0.84</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>0.86</td>
<td>1.61</td>
<td>0.82</td>
<td>2.02</td>
<td>0.81</td>
<td>2.31</td>
</tr>
<tr>
<td>11</td>
<td>LMAR</td>
<td>0.85</td>
<td>1.38</td>
<td>0.87</td>
<td>2.13</td>
<td>0.91</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>NNs</td>
<td>0.87</td>
<td>1.50</td>
<td>0.80</td>
<td>2.70</td>
<td>0.83</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td>Ridge</td>
<td>0.86</td>
<td>1.63</td>
<td>0.85</td>
<td>2.44</td>
<td>0.85</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>LICORS</td>
<td>0.88</td>
<td>1.56</td>
<td>0.83</td>
<td>1.99</td>
<td>0.82</td>
<td>2.25</td>
</tr>
</tbody>
</table>
Conclusions

Future improvements:

- Hierarchical models; information sharing between patients.
- Code optimizations (current pipeline in R/C++)

Further reading: