CHMM: an R package for coupled Hidden Markov Models

Julie Aubert (1), Xiaoyang Wang (1,2), Emilie Lebarbier (1) & Stéphane Robin (1)
(1) UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France.
(2) School of Mathematics and Statistics, Shandong University, Weihai, China.

Detection of CNV taking into account dependency between individuals

Copy number variations (CNVs) are genomic alterations that result in an abnormal number of copies of one or more genes: duplication (green), normal (blue), deletion (red).

\[
\begin{array}{ccc}
\text{A} & \text{B} & \text{C} \\
\hline
1 & 0.61 & 0.56 \\
0.61 & 1 & 0.75 \\
0.56 & 0.75 & 1
\end{array}
\]

Kinship data

CNV detection of a simulated sample.

Selection criterion [2,3]

\[
\tilde{Q} = \arg \max_Q \mathcal{J}_Q(Y, \tilde{\theta}, \tilde{P}) - \left[1 + Q(Q - 1)\right] \log(IT)/2,
\]

where \(\mathcal{J}_Q(Y, \tilde{\theta}, \tilde{P})\) is the maximized lower bound of the \(Q\)-state model.

Simulation study

Runtime (in second), Weak dependency, \(\sigma = 1\), \(I\): number of lines

<table>
<thead>
<tr>
<th>(I)</th>
<th>HMM-EM</th>
<th>CHMM-EM-CHMM-EM</th>
<th>CHMM-EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.8</td>
<td>0.3</td>
<td>2.0</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
<td>0.5</td>
<td>11.2</td>
</tr>
<tr>
<td>4</td>
<td>1.2</td>
<td>0.5</td>
<td>79.4</td>
</tr>
<tr>
<td>5</td>
<td>1.6</td>
<td>0.8</td>
<td>920.2</td>
</tr>
</tbody>
</table>

Classification accuracy (%) for \(I = 3\)

Coupled HMM applied to the detection of CNV in the maize

Variational inference [1,2]

When \(I\) (the number of individuals) is large, \(P(Z|Y)\) is not computable.

Mean-field approximation

\[
P(Z) = \arg \min_{P_{\mathcal{C}}} \mathcal{K}_C \left[ \tilde{P}(Z); P(Z|Y) \right]
\]

where \(\mathcal{P} = \left\{ \tilde{P}(Z) | \tilde{P}(Z) \propto \prod_i \prod_j P(Z_{i,t} | Z_{j,t-1}) \right\}\) (independent Markov chains)

Forward part of the VE-step

Let denote \(p_{zkr} = P(Z_{i,t} = r | Z_{i-t-1} = q)\), then we obtain a set of fixed point equations for \(p_{zkr}\):

\[
p_{zkr} \propto \pi_{kr} f(Y_{i,t}, \mu_r, \sigma^2) \times \omega^{1 - \pi_{kr}}(1 - \pi_{kr} Z_{j,t})^\omega
\]

References and acknowledgements


This work was supported by the CNR-Mine program funded by the French National Research Agency (ANR-10-GENM-004) and France Agromoteur (11000025).

Conclusions

A model and associated inference for the detection of CNV taken into account dependency.

- Selection criterion
- Heuristic for choosing the value of the parameter \(\omega\).
- CHMM R package available from the CRAN.