Improving structured post hoc inference via a hidden Markov model

Marie Perrot-Dockès, Gilles Blanchard, Pierre Neuvial, Etienne Roquain

*Email: marie.perrot-dockes@farbonne-universite.fr

Keywords: HMM, Post-hoc bounds, Selective inference, credible intervals, multiple testing

Motivation / Modeling

Copy-Number-Variations (CNV)

- Classical aim: find positions where the CNV differ between the two cancers.
- Selective inference: positions freely selected by the user from the same data.
- Hidden Markov Model

<table>
<thead>
<tr>
<th>Hidden Markov Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0, X) = (\theta, \theta \in \Omega) \sim \text{HMM}, \text{with}</td>
</tr>
<tr>
<td>Hidden states (\theta, {0, 1} ), Markov chain with transition matrix (\Lambda)</td>
</tr>
<tr>
<td>(E(X</td>
</tr>
<tr>
<td>with (f_\theta) known and (\theta) unknown.</td>
</tr>
</tbody>
</table>

Existing inferences

Inferences conditional on \(\theta\)

Multiple-testing:
Prescribe a set of positions with controlled false discoveries.

New bounds

Oracle bound \(V(0, S, \Gamma)\)
Plug-in bound \(V(0, S, \hat{\Gamma})\)

Numerical experiments

Dataset description

Weekly ILI rate (1984-2008)
120 weeks
Two categories: usual (\(X\)) alteration (\(X\))

Application: credible sets for the FDP of various selected sets in the ILI data set

Figure 1: Value of \(V(0, S, \hat{\Gamma}) - V(0, S, \Gamma)\) for different \(\hat{\Gamma}\) and \(\Gamma\) and 300 replications, with \(\alpha = 0.05\), \(S_{\text{ILM}}\) and \(S_{\text{ILM}}\) (\(\leq 0.05\)) being the \(V(0, S, \hat{\Gamma})\) and \(V(0, S, \Gamma)\) best-fitted intervals. The displayed percentages are the proportion of replications where \(V(0, S, \hat{\Gamma}) < V(0, S, \Gamma)\), which is a Monte-Carlo estimate of the violation probability of the bound.

Biblio