Suggested Reference


Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

For more information, see General copyright.
Performance and Limitations of Likelihood based Information Criteria and Leave-one-out Cross-validation Approximation Methods

M.A.C.S. Sampath Fernando
James M. Curran
Renate Meyer

Department of Statistics
University of Auckland
New Zealand

sampathf73@gmail.com

November 21, 2016
Overview

1 Introduction
   - Modelling the behaviour data
   - Bayesian statistical models
   - Electropherogram (EPG)

2 Statistical models for stutter ratio ($SR$)
   - Mean and variance of $SR$
   - Different Models for $SR$

3 Bayesian model comparisons
   - Information criteria
   - Limitations of information criteria
   - LOO-CV and LOO-CV approximations

4 Results and findings
Introduction

- What is a statistical model?
  - A probabilistic system
  - A finite/infinite mixture of probability distributions

- All models are approximations
- “All models are wrong, but some are useful” - George Box (1976)
If we wish to perform statistical inference, or use our model to probabilistically evaluate the behavior of new observations, then we need three steps:

1. Assume that the data are generated from some statistical distribution
2. Write down equations for the parameters of the assumed distribution, e.g. the mean and the standard deviation
3. Use standard techniques to estimate the unknown parameters in 2

Steps 1-3 should be repeated as often as possible to get the "best" model

Most model building consists of steps 2 and 3

Classical and Bayesian approaches
Distribution of data \( (X) \) depends on unknown parameter \( \theta \)

Inference on \( \theta \)

Consists of a parametric statistical model(s) \( f(x|\theta) \)

Prior distribution(s) of parameter \( \theta \)

Different types of Bayesian models

- Non-hierarchical models
- Hierarchical models
- Mixture models
Electropherogram (EPG)
We are interested in modelling the stutter ratio ($SR$)

$$SR = \frac{\text{Observed height of the stutter peak}}{\text{The height of the parent allelic peak}}$$
Behaviour of $SR$

- $SR$ is affected by the longest uninterrupted sequence of the allele, $LUS$
- $SR$ is more variable for smaller values of observed allele height
Mean and Variance of Stutter Ratio

- Mean stutter ratio

\[ \mu_{li} = E(SR_{li}) = \beta_{0l} + \beta_{1l} LUS_{li} \]

- Variance of stutter ratio is inversely proportional to the allele height
  - A common variance for all the loci - models with profile wide variances

\[ \sigma_i^2 = \frac{\sigma^2}{O_{ai}} \]

- Locus specific variance

\[ \sigma_{li}^2 = \frac{\sigma_l^2}{O_{ali}} \]
## Models proposed by Bright et al. (2013)

<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN₀</td>
<td>( \ln(SR_{li}) \sim N(\mu_{li}, \sigma^2_{i}) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li}LUS_{li} )</td>
<td>( \sigma^2_{i} = \frac{\sigma^2_{O_{ai}}}{O_{ali}} )</td>
</tr>
<tr>
<td>LN₁</td>
<td>( \ln(SR_{li}) \sim N(\mu_{li}, \sigma^2_{li}) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li}LUS_{li} )</td>
<td>( \sigma^2_{li} = \frac{\sigma^2_{i}}{O_{ali}} )</td>
</tr>
<tr>
<td>G₀</td>
<td>( SR_{li} \sim \text{Gamma}(\alpha_{li}, \theta_{li}) )</td>
<td>( \mu_{li} = e^{(\beta_{0li}+\beta_{1li}LUS_{li})} )</td>
<td>( \sigma^2_{li} = \frac{\sigma^2_{i}}{O_{ali}} )</td>
</tr>
<tr>
<td>G₁</td>
<td>( SR_{li} \sim \text{Gamma}(\alpha_{li}, \theta_{li}) )</td>
<td>( \mu_{li} = e^{(\beta_{0li}+\beta_{1li}LUS_{li})} )</td>
<td>( \sigma^2_{li} = \frac{\sigma^2_{i}}{O_{ali}} )</td>
</tr>
<tr>
<td>MLN₁</td>
<td>( \ln(SR_{li}) \sim \pi N(\mu_{li}, \sigma^2_{0li}) + (1 - \pi)N(\mu_{li}, \sigma^2_{1li}) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li}LUS_{li} )</td>
<td>( \sigma^2_{0li} = \frac{\sigma^2_{O_{ali}}}{O_{ali}} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \sigma^2_{1li} = \frac{\sigma^2_{0li} + \sigma^2_{1li}}{O_{ali}} )</td>
</tr>
</tbody>
</table>
### Description of the Proposed Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₀</td>
<td>( SR_{li} \sim N(\mu_{li}, \sigma_{i}^2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{i}^2 = \frac{\sigma^2}{O_{aji}} )</td>
</tr>
<tr>
<td>N₁</td>
<td>( SR_{li} \sim N(\mu_{li}, \sigma_{li}^2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{li}^2 = \frac{\sigma^2}{O_{ali}} )</td>
</tr>
<tr>
<td>T₀</td>
<td>( SR_{li} \sim t(\mu_{li}, \sigma_{i}^2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{i}^2 = \frac{\sigma^2}{O_{aji}} )</td>
</tr>
<tr>
<td>T₁</td>
<td>( SR_{li} \sim t(\mu_{li}, \sigma_{li}^2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{li}^2 = \frac{\sigma^2}{O_{ali}} )</td>
</tr>
<tr>
<td>MN₁</td>
<td>( SR_{li} \sim \pi N(\mu_{li}, \sigma_{0li}^2) + (1 - \pi) N(\mu_{li}, \sigma_{1li}^2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{0li}^2 = \frac{\sigma^2}{O_{ali}} )</td>
</tr>
<tr>
<td></td>
<td>( \pi t(\mu_{li}, \sigma_{0li}^2, \nu_1) + (1 - \pi) t(\mu_{li}, \sigma_{1li}^2, \nu_2) )</td>
<td>( \mu_{li} = \beta_{0li} + \beta_{1li} LUS_{li} )</td>
<td>( \sigma_{0li}^2 = \frac{\sigma^2}{O_{ali}} )</td>
</tr>
</tbody>
</table>
Information Criteria

- **Akaike information criterion (AIC)**
  \[
  AIC = -2 \log p(y|\hat{\theta}_{mle}) + 2k
  \]

- **Bayesian information criterion (BIC)**
  \[
  BIC = -2 \log p(y|\hat{\theta}_{mle}) + k \log n
  \]

- **Deviance information criterion (DIC)**
  \[
  DIC = -2 \log p(y|\hat{\theta}_{Bayes}) - p_{DIC}
  \]
  \[
  \hat{p}_{DIC} = 2 \left[ \log p(y|\hat{\theta}_{Bayes}) - \frac{1}{S} \sum_{s=1}^{S} \log p(y|\theta_s) \right]
  \]
  \[
  p_{DIC \text{ alt}} = 2 \text{Var}_{\text{post}} \left[ \log p(y|\theta) \right]
  \]
Widely Available or Watanabe-Akaike Information Criterion (WAIC)

- Log point wise predictive density (lppd)
  \[
  lppd = \sum_{i=1}^{n} \log p_{post}(y_i) = \sum_{i=1}^{n} \log \int_{\theta} p(y_i|\theta) p_{post}(\theta) d\theta
  \]

- Estimated (computed) log point wise predictive density (clppd)
  \[
  clppd = \sum_{i=1}^{n} \log \left[ \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta_s) \right] \\
  WAIC = -2 \left( clppd - p_{WAIC} \right)
  \]

\[
\hat{p}_{WAIC} = 2 \sum_{i=1}^{n} \left\{ \log \left[ \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta_s) \right] - \frac{1}{S} \sum_{s=1}^{S} \log p(y_i|\theta) \right\}
\]

\[
p_{WAIC \text{ alt}} = \sum_{i=1}^{n} \text{Var}_{post} \left[ \log p(y_i|\theta) \right]
\]
Limitations of Information Criteria

- **AIC & BIC**
  - MLE
  - Cannot use with hierarchical models
  - Not recommended for singular models

- **DIC**
  - Cannot use with mixture models (posterior estimates of means are quite delicate)

- **WAIC**
  - Valid if $\text{Var}_{\text{post}}[\log p(y_i|\theta)] \leq 0.4$
  - If $\text{Var}_{\text{post}}[\log p(y_i|\theta)] > 0.4$
    - Then use leave-one-out cross-validation (LOO-CV)
LOO-CV and LOO-CV Approximations

- Leave-one-out cross-validated lppd

\[
\text{lppd}_{\text{loo-cv}} = \sum_{i=1}^{n} \log p_{\text{post}(-i)}(y_i) \quad \text{lppd}_{\text{loo-cv}} = \sum_{i=1}^{n} \log \left[ \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta_{is}) \right]
\]

- Importance sampling LOO-CV (IS-LOO)

\[
r_{is} = \frac{1}{p(y_i|\theta_{is})} \propto \frac{p(\theta_{is}|y_{-i})}{p(\theta_{is}|y)}
\]

- Truncated importance sampling LOO-CV (TIS-LOO)

\[
w_{is} = \min \left( r_{is}, \sqrt{S\bar{r}_i} \right)
\]

- Pareto-smoothed importance sampling LOO-CV (PSIS-LOO)

\[
\tilde{w}_{is} = \min \left( m_{is}, S^{\frac{3}{4}} \bar{m}_i \right)
\]
## Log Predictive Densities (NGM SElect™ Data)

<table>
<thead>
<tr>
<th>Model</th>
<th>IS</th>
<th>TIS</th>
<th>PSIS</th>
<th>cllpd</th>
<th>WAIC</th>
<th>WAICalt</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN₀</td>
<td>9002</td>
<td>10447</td>
<td>9722</td>
<td>14188</td>
<td>12575</td>
<td>11492</td>
</tr>
<tr>
<td>LN₁</td>
<td>10207</td>
<td>11576</td>
<td>10855</td>
<td>15295</td>
<td>13680</td>
<td>12600</td>
</tr>
<tr>
<td>MLN₁</td>
<td>9834</td>
<td>11871</td>
<td>11617</td>
<td>16475</td>
<td>14411</td>
<td>13636</td>
</tr>
<tr>
<td>G₀</td>
<td>10509</td>
<td>11649</td>
<td>11048</td>
<td>14961</td>
<td>13470</td>
<td>12529</td>
</tr>
<tr>
<td>G₁</td>
<td>11153</td>
<td>12259</td>
<td>11677</td>
<td>15533</td>
<td>14052</td>
<td>13132</td>
</tr>
<tr>
<td>N₀</td>
<td>11383</td>
<td>12427</td>
<td>11870</td>
<td>15573</td>
<td>14149</td>
<td>13253</td>
</tr>
<tr>
<td>N₁</td>
<td>11873</td>
<td>12784</td>
<td>12208</td>
<td>15926</td>
<td>14492</td>
<td>13596</td>
</tr>
<tr>
<td>MN₁</td>
<td>12053</td>
<td>13133</td>
<td>12609</td>
<td>16501</td>
<td>14903</td>
<td>13961</td>
</tr>
<tr>
<td>T₀</td>
<td>6334</td>
<td>9650</td>
<td>8370</td>
<td>16408</td>
<td>14038</td>
<td>12050</td>
</tr>
<tr>
<td>T₁</td>
<td>8469</td>
<td>11105</td>
<td>10065</td>
<td>16488</td>
<td>14458</td>
<td>12889</td>
</tr>
<tr>
<td>MT₁</td>
<td>9751</td>
<td>11979</td>
<td>11147</td>
<td>16821</td>
<td>14734</td>
<td>13395</td>
</tr>
</tbody>
</table>
## Log Predictive Densities (Identifiler™ Data)

<table>
<thead>
<tr>
<th>Model</th>
<th>IS</th>
<th>TIS</th>
<th>PSIS</th>
<th>clppd</th>
<th>WAIC</th>
<th>WAICalt</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN₀</td>
<td>14902</td>
<td>16962</td>
<td>15929</td>
<td>22452</td>
<td>20079</td>
<td>18498</td>
</tr>
<tr>
<td>LN₁</td>
<td>16601</td>
<td>18627</td>
<td>17600</td>
<td>24000</td>
<td>21652</td>
<td>20101</td>
</tr>
<tr>
<td>MLN₁</td>
<td>17974</td>
<td>20283</td>
<td>19363</td>
<td>26400</td>
<td>23668</td>
<td>21920</td>
</tr>
<tr>
<td>G₀</td>
<td>16640</td>
<td>18335</td>
<td>17448</td>
<td>23226</td>
<td>21038</td>
<td>19639</td>
</tr>
<tr>
<td>G₁</td>
<td>17807</td>
<td>19536</td>
<td>18687</td>
<td>24386</td>
<td>22211</td>
<td>20832</td>
</tr>
<tr>
<td>N₀</td>
<td>16971</td>
<td>18527</td>
<td>17682</td>
<td>23244</td>
<td>21111</td>
<td>19767</td>
</tr>
<tr>
<td>N₁</td>
<td>18283</td>
<td>19846</td>
<td>19022</td>
<td>24563</td>
<td>22425</td>
<td>21085</td>
</tr>
<tr>
<td>MN₁</td>
<td>19761</td>
<td>21378</td>
<td>20753</td>
<td>26725</td>
<td>24127</td>
<td>22623</td>
</tr>
<tr>
<td>T₀</td>
<td>8749</td>
<td>14706</td>
<td>12547</td>
<td>26416</td>
<td>22512</td>
<td>19065</td>
</tr>
<tr>
<td>T₁</td>
<td>10187</td>
<td>15718</td>
<td>13675</td>
<td>26607</td>
<td>22880</td>
<td>19679</td>
</tr>
<tr>
<td>MT₁</td>
<td>14935</td>
<td>18876</td>
<td>17518</td>
<td>26984</td>
<td>23638</td>
<td>21369</td>
</tr>
</tbody>
</table>
Log Predictive Density Profiles (Identifiier™ Data)

![Graph showing various Log Predictive Density Profiles for different models: IS.LOO, TIS.LOO, PSIS.LOO, IS.LOO, clppd, elppd(WAIC), elppd(WAICalt). The x-axis represents Model Name, and the y-axis represents Log Predictive Density. The graph illustrates the performance and limitations of IC & LOO-CV.]

Sampath Fernando (UoA)  Performance & Limitations of IC & LOO-CV  November 21, 2016  20 / 22
Posterior Variances of Lppds (Identifier™ Data)
Thank You!