Estimating Cross-validatory Predictive $p$-values with Integrated Importance Sampling for Disease Mapping Models

Longhai Li

Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon, SK, CANADA

Presented on 12 June 2017
SSC Annual Meeting at the University of Manitoba
Acknowledgements

- Joint work with Shi Qiu and Cindy X. Feng.

- The work was supported by grants from Natural Sciences and Engineering Research Council of Canada (NSERC) and Canada Foundation for Innovation (CFI).
Outline

1. Predictive $p$-value
2. A Disease Mapping Model
3. Methods for Computing Predictive $p$-values
   - Posterior Predictive Checking VS LOOCV Checking
   - Non-integrated Importance Sampling
   - Ghosting Method
   - Integrated (Marginalized) Importance Sampling
4. Numerical Comparisons with Two Real Datasets
   - Lip Cancer Data in Scotland
   - Larynx Cancer Data in Germany
5. Conclusions and Future Work
6. References
Section 1

Predictive $p$-value
Predictive $p$-value

- Predictive $p$-value is the tail probability of a predictive distribution:

- Predictive $p$-value can be used to
  - check model
  - identify “outliers” (or divergent regions in disease mapping problems)
Section 2

A Disease Mapping Model
The data represents male lip cancer counts (over the period 1975 - 1980) in the $n = 56$ districts of Scotland. The data includes these columns:

- the number of observed cases of lip cancer, $y_i$;
- the number of expected cases, $E_i$, which are based on age effects, and are proportional to a “population at risk” after such effects have been taken into account;
- the percent of population employed in agriculture, fishing and forestry, $x_i$, used as a covariate; and
- a list of the neighbouring regions.
### A Subset of the Dataset

<table>
<thead>
<tr>
<th>ID</th>
<th>District name</th>
<th>Y</th>
<th>E</th>
<th>SMR</th>
<th>X</th>
<th>Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skye-Lochalsh</td>
<td>9</td>
<td>1.38</td>
<td>6.52</td>
<td>16</td>
<td>5,9,11,19</td>
</tr>
<tr>
<td>2</td>
<td>Banff-Buchan</td>
<td>39</td>
<td>8.66</td>
<td>4.50</td>
<td>16</td>
<td>7,10</td>
</tr>
<tr>
<td>3</td>
<td>Caithness</td>
<td>11</td>
<td>3.04</td>
<td>3.62</td>
<td>10</td>
<td>6,12</td>
</tr>
<tr>
<td>11</td>
<td>Western Isles</td>
<td>13</td>
<td>4.40</td>
<td>2.95</td>
<td>7</td>
<td>1,5,9,12</td>
</tr>
<tr>
<td>15</td>
<td>NE Fife</td>
<td>17</td>
<td>7.84</td>
<td>2.17</td>
<td>7</td>
<td>25,29,50</td>
</tr>
<tr>
<td>17</td>
<td>Badenoch</td>
<td>2</td>
<td>1.07</td>
<td>1.87</td>
<td>10</td>
<td>7,9,13,16,19,29</td>
</tr>
<tr>
<td>26</td>
<td>Dunfermline</td>
<td>15</td>
<td>12.49</td>
<td>1.20</td>
<td>1</td>
<td>25,29,42,43</td>
</tr>
<tr>
<td>38</td>
<td>Monklands</td>
<td>8</td>
<td>9.35</td>
<td>0.86</td>
<td>1</td>
<td>30,42,44,49,51,54</td>
</tr>
<tr>
<td>42</td>
<td>Falkirk</td>
<td>8</td>
<td>15.78</td>
<td>0.51</td>
<td>16</td>
<td>26,30,34,38,43,51</td>
</tr>
<tr>
<td>45</td>
<td>Edinburgh</td>
<td>19</td>
<td>50.72</td>
<td>0.37</td>
<td>1</td>
<td>28,30,33,56</td>
</tr>
<tr>
<td>49</td>
<td>Glasgow</td>
<td>28</td>
<td>88.66</td>
<td>0.32</td>
<td>0</td>
<td>38,40,41,44,47,48,52,53,54</td>
</tr>
<tr>
<td>50</td>
<td>Dundee</td>
<td>6</td>
<td>19.62</td>
<td>0.31</td>
<td>1</td>
<td>15,21,29</td>
</tr>
<tr>
<td>55</td>
<td>Annandale</td>
<td>0</td>
<td>4.16</td>
<td>0</td>
<td>16</td>
<td>18,20,24,27,56</td>
</tr>
<tr>
<td>56</td>
<td>Tweeddale</td>
<td>0</td>
<td>1.76</td>
<td>0</td>
<td>10</td>
<td>18,24,30,33,45,55</td>
</tr>
</tbody>
</table>
A Hierarchical Bayesian Spatial Model for $y_i$’s

- A model for the observed variables given latent variables

$$y_i | E_i, \lambda_i \sim \text{Poisson}(\lambda_i E_i),$$

where $\lambda_i$ denotes the underlying relative risk for district $i$.

- A model for latent log relative risks $s_i = \log(\lambda_i)$

$$(s_1, \ldots, s_n)^\prime \sim \mathcal{N}_n(\alpha + X\beta, \Phi \tau^2)$$

where $\Phi$ is a matrix modelling spatial dependency with proper conditional auto-regressive (CAR) method.

- A model (prior) for parameters

$$\tau^2 \sim \text{Inv-Gamma}(0.5, 0.0005)$$
$$\beta \sim \mathcal{N}(0, 1000^2)$$
$$\phi \sim \text{Unif}(\phi_0, \phi_1).$$
Section 3

Methods for Computing Predictive $p$-values
Subsection 1

Posterior Predictive Checking VS LOOCV Checking
Posterior Predictive Checking

- The **full data posterior** density of \((s_{1:n}, \theta)\) given observations \(y_{1:n}^{\text{obs}}\) is given by:

\[
P_{\text{post}}(\theta, s_{1:n}|y_{1:n}^{\text{obs}}) = \prod_{j=1}^{n} P_y(y_j^{\text{obs}}|s_j, \theta)P_s(s_{1:n}|\theta)\pi(\theta) / C_1. \tag{1}
\]

- Predictive \(p\)-value for \(y_i\) given parameters and latent variable:

\[
p\text{-value}(y_i^{\text{obs}}|\theta, s_i) = Pr(y_i > y_i^{\text{obs}}|\theta, s_i) + 0.5Pr(y_i = y_i^{\text{obs}}|\theta, s_i). \tag{2}
\]

- The posterior predictive \(p\)-value:

\[
p\text{-value}^{\text{Post}}(y_i^{\text{obs}}) = E_{\text{post}}(p\text{-value}(y_i^{\text{obs}}|\theta, s_i)) \tag{3}
\]

- Given MCMC samples \(\{(\theta^{(t)}, s_{1:n}^{(t)}); t = 1, \ldots, T\}\) from the full data posterior (1), posterior predictive \(p\)-value (3) is computed as follows:

\[
p\text{-value}^{\text{Post}}(y_i^{\text{obs}}) = \frac{\sum_{t=1}^{T} p\text{-value}(y_i^{\text{obs}}|\theta^{(t)}, s_{i}^{(t)})}{T}. \tag{4}
\]
Posterior predictive checking uses the dataset twice: $y_{i}^{\text{obs}}$ is used to obtain the posterior predictive distribution of $y_i$, and is also used to test itself. Generally, $y_{i}^{\text{obs}}$ appears better predictable by the model.

The posterior predictive $p$-values are concentrated around 0.5 rather than uniformly distributed on the interval $(0,1)$.

Failure in identifying “outliers”
LOOCV Predictive $p$-value (Gold Standard)

- **LOOCV distribution** $P_{\text{post}(-i)}(\theta, \mathbf{s}_{1:n}|\mathbf{y}_{-i}^\text{obs})$ is given as follows:

$$
P_{\text{post}(-i)}(\theta, \mathbf{s}_{1:n}|\mathbf{y}_{-i}^\text{obs}) = \prod_{j=1,\ldots,i-1,i+1,\ldots,n} P_y(y_j^\text{obs}|s_j, \theta)P_s(\mathbf{s}_{1:n}|\theta)\pi(\theta) / C_2. \tag{5}
$$

- The **LOOCV predictive $p$-value** for $y_i^{\text{obs}}$

$$
p\text{-value}(y_i^{\text{obs}}|\mathbf{y}_{-i}^\text{obs}) = E_{\text{post}(-i)}(p\text{-value}(y_i^{\text{obs}}|\theta, s_i)). \tag{6}
$$

- Given MCMC samples $\{(\theta^{(t)}, \mathbf{s}_{1:n}^{(t)}); t = 1, \ldots, T\}$ from the LOOCV posterior (5), the **LOOCV predictive $p$-value** is computed as follows:

$$
\hat{p}\text{-value}^{CV}(y_i^{\text{obs}}) = \frac{\sum_{t=1}^{T} p\text{-value}(y_i^{\text{obs}}|\theta^{(t)}, s_i^{(t)})}{T}. \tag{7}
$$

- MCMC sampling needs to be redone for each $y_i^{\text{obs}}$. 

3. Methods for Computing Predictive $p$-values/Posterior Predictive Checking VS LOOCV Checking
Subsection 2

Non-integrated Importance Sampling
Review of Importance Sampling

- Our goal is to find the expectation of a function $a(X)$ when $X$ has a probability density proportional to $f(x)$, denoted by
  \[ E_f (a(X)) \]
- Instead of drawing samples from $f(x)$, we draw samples from an approximate distribution with a probability density $g(x)$.
- Importance weight:
  \[ W(x) = \frac{f(x)}{g(x)} \]
- **Importance reweighting formula**
  \[ E_f (a(X)) = \frac{E_g (a(X)W(X))}{E_g (W(X))}. \] (8)
- The intuitive explanation of the importance reweighting formula (8) is that samples that are more compatible with the target distribution $f$ will be assigned more weight (and vice versa).
Non-integrated Importance Sampling

- We can estimate expectations with respect to $P_{\text{post}(-i)}(\theta, s_{1:n}|y_{-i}^{\text{obs}})$ in (5) by reweighting samples from $P_{\text{post}}(\theta, s_{1:n}|y_{1:n}^{\text{obs}})$ (1):

$$p\text{-value}(y_i^{\text{obs}}|y_{-i}^{\text{obs}}) = \frac{E_{\text{post}}[p\text{-value}(y_i^{\text{obs}}|\theta, s_i) W_i^{\text{nIS}}(\theta, s_i)]}{E_{\text{post}}[W_i^{\text{nIS}}(\theta, s_i)]}, \quad (9)$$

- $W_i^{\text{nIS}}(\theta, s_i)$ is the ratio:

$$W_i^{\text{nIS}}(\theta, s_i) = \frac{P_{\text{post}(-i)}(\theta, s_{1:n}|y_{-i}^{\text{obs}})}{P_{\text{post}}(\theta, s_{1:n}|y_{1:n}^{\text{obs}})} \times \frac{C_2}{C_1} = \frac{1}{P_y(y_i^{\text{obs}}|\theta, s_i)}. \quad (10)$$

- Given MCMC samples $\{(\theta^{(t)}, s_{1:n}^{(t)}); t = 1, \ldots, T\}$ from the full data posterior (1), the nIS predictive $p$-value (9) is computed as follows:

$$\hat{p}\text{-value}^{\text{nIS}} (y_i^{\text{obs}}) = \frac{\sum_t^T \left[ p\text{-value}(y_i^{\text{obs}}|\theta^{(t)}, s_{i}^{(t)}) W_i^{\text{nIS}}(\theta^{(t)}, s_{i}^{(t)}) \right]}{T} / \frac{T}{\sum_t^T W_i^{\text{nIS}}(\theta^{(t)}, s_{i}^{(t)})}. \quad (11)$$
Most of the MCMC samples of $s_i$ are largely bound to regions that fit the observation $y_i^{\text{obs}}$ well.

The estimate (11) is dominated by a single or a few very incompatible MCMC samples.

The “effective” sample size in (11) may be very small.

This leads to the notorious instability problem of importance sampling.
Subsection 3

Ghosting Method
Marshall and Spiegelhalter (2007) propose that the $s_i$ is replaced with a re-generated $s_i$ without reference to $y_{i}^{\text{obs}}$. Technically, draw samples from the “ghosting” distribution of $(\theta, s_{1:n})$:

$$P_{\text{ghost}}(s_{1:n}, \theta) = P_{\text{post}}(\theta, s_{-i}|y_{1:n}^{\text{obs}}) \times P(s_i|s_{-i}, \theta), \quad (12)$$

The “ghosting” predictive $p$-value

$$p\text{-value}_{\text{ghost}}(y_{i}^{\text{obs}}) = E_{\text{ghost}}(p\text{-value}(y_{i}^{\text{obs}}|\theta, s_i)), \quad (13)$$

Given “ghosting” samples $\{(\theta^{(t)}, \tilde{s}_{i}^{(t)}); t = 1, \ldots, T\}$, the ghosting predictive $p$-value is computed as follows:

$$\hat{p}\text{-value}_{\text{ghost}}(y_{i}^{\text{obs}}) = \frac{\sum_{t=1}^{T} p\text{-value}(y_{i}^{\text{obs}}|\theta^{(t)}, \tilde{s}_{i}^{(t)})}{T}. \quad (14)$$
Subsection 4

Integrated (Marginalized) Importance Sampling
The “ghosting” predictive $p$-value is not equivalent to the LOOCV $p$-value in theory.

Particularly, the MCMC sample of $(\theta^{(t)}, s^{(t)}_{-i})$ still contain information of $y_{i}^{\text{obs}}$.

$P_{y}(y_{i}^{\text{obs}}|s_{i}, \theta)$ is integrated out with respect to the distribution of $s_{i}$ given $\theta$ without reference to the actual observation $y_{i}^{\text{obs}}$.

The integrated predictive density of $y_{i}^{\text{obs}}$:

$$P(y_{i}^{\text{obs}}|\theta, s_{-i}) = \int P_{y}(y_{i}^{\text{obs}}|s_{i}, \theta)P(s_{i}|s_{-i}, \theta)d s_{i}. \quad (15)$$

We use the integrated predictive density of $y_{i}^{\text{obs}}$ to correct for the bias in $\theta, s_{-i}$ due to inclusion of $y_{i}^{\text{obs}}$ in full data posterior.
iS Predictive $p$-value

- For each MCMC sample, we first generate two sets of new $s_i$ from $P(s_i|s^{(t)}_{-i}, \theta^{(t)})$, denoted by $\{\tilde{s}_i^{(A,k)}; k = 1, \ldots, R\}$ and $\{\tilde{s}_i^{(W,k)}; k = 1, \ldots, R\}$ respectively.

- Computing integrated $p$-value and importance weight:

$$
\hat{A}_i^{(t)} = \frac{\sum_{k=1}^R p\text{-value}(y_i^{\text{obs}}|\theta^{(t)}, \tilde{s}_i^{(A,k)})}{R} \quad (16)
$$

$$
\hat{W}_i^{(t)} = \frac{1}{R} \frac{\sum_{k=1}^R P_y(y_i^{\text{obs}}|\theta^{(t)}, \tilde{s}_i^{(W,k)})}{R} \quad (17)
$$

- The LOOCV $p$-value is then estimated as follows:

$$
\hat{p}\text{-value}^{iS}(y_i^{\text{obs}}) = \frac{\sum_{t=1}^T \hat{A}_i^{(t)} \hat{W}_i^{(t)}}{T} / \frac{\sum_{t=1}^T \hat{W}_i^{(t)}}{T} \quad (18)
$$

- It can be shown that this estimate asymptotically equals to the true LOOCV predictive $p$-value when $T \rightarrow \infty$
Section 4

Numerical Comparisons with Two Real Datasets
Subsection 1

Lip Cancer Data in Scotland
Estimated predictive \( p \)-values\((y_{i}^{obs})\)

<table>
<thead>
<tr>
<th>ID</th>
<th>LOOCV</th>
<th>PCH</th>
<th>GHO</th>
<th>nIS</th>
<th>iIS</th>
<th>ID</th>
<th>LOOCV</th>
<th>PCH</th>
<th>GHO</th>
<th>nIS</th>
<th>iIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.308</td>
<td>0.417</td>
<td>0.310</td>
<td>0.319</td>
<td>0.307</td>
<td>29</td>
<td>0.667</td>
<td>0.547</td>
<td>0.651</td>
<td>0.631</td>
<td>0.664</td>
</tr>
<tr>
<td>2</td>
<td>0.033</td>
<td>0.320</td>
<td>0.050</td>
<td>0.074</td>
<td>0.030</td>
<td>30</td>
<td>0.260</td>
<td>0.367</td>
<td>0.278</td>
<td>0.263</td>
<td>0.262</td>
</tr>
<tr>
<td>3</td>
<td>0.090</td>
<td>0.325</td>
<td>0.096</td>
<td>0.089</td>
<td>0.090</td>
<td>31</td>
<td>0.275</td>
<td>0.359</td>
<td>0.283</td>
<td>0.262</td>
<td>0.274</td>
</tr>
<tr>
<td>4</td>
<td>0.418</td>
<td>0.437</td>
<td>0.423</td>
<td>0.430</td>
<td>0.417</td>
<td>32</td>
<td>0.816</td>
<td>0.601</td>
<td>0.799</td>
<td>0.768</td>
<td>0.818</td>
</tr>
<tr>
<td>5</td>
<td>0.139</td>
<td>0.357</td>
<td>0.155</td>
<td>0.159</td>
<td>0.140</td>
<td>33</td>
<td>0.469</td>
<td>0.455</td>
<td>0.467</td>
<td>0.466</td>
<td>0.463</td>
</tr>
<tr>
<td>6</td>
<td>0.512</td>
<td>0.463</td>
<td>0.512</td>
<td>0.458</td>
<td>0.514</td>
<td>34</td>
<td>0.188</td>
<td>0.317</td>
<td>0.211</td>
<td>0.189</td>
<td>0.190</td>
</tr>
<tr>
<td>7</td>
<td>0.060</td>
<td>0.312</td>
<td>0.072</td>
<td>0.041</td>
<td>0.058</td>
<td>35</td>
<td>0.370</td>
<td>0.414</td>
<td>0.372</td>
<td>0.364</td>
<td>0.370</td>
</tr>
<tr>
<td>8</td>
<td>0.113</td>
<td>0.313</td>
<td>0.114</td>
<td>0.112</td>
<td>0.112</td>
<td>36</td>
<td>0.151</td>
<td>0.284</td>
<td>0.162</td>
<td>0.154</td>
<td>0.149</td>
</tr>
<tr>
<td>9</td>
<td>0.267</td>
<td>0.386</td>
<td>0.281</td>
<td>0.261</td>
<td>0.271</td>
<td>37</td>
<td>0.596</td>
<td>0.524</td>
<td>0.590</td>
<td>0.598</td>
<td>0.601</td>
</tr>
<tr>
<td>10</td>
<td>0.269</td>
<td>0.405</td>
<td>0.279</td>
<td>0.300</td>
<td>0.267</td>
<td>38</td>
<td>0.071</td>
<td>0.221</td>
<td>0.092</td>
<td>0.076</td>
<td>0.073</td>
</tr>
<tr>
<td>11</td>
<td>0.127</td>
<td>0.334</td>
<td>0.137</td>
<td>0.138</td>
<td>0.122</td>
<td>39</td>
<td>0.820</td>
<td>0.627</td>
<td>0.794</td>
<td>0.804</td>
<td>0.821</td>
</tr>
<tr>
<td>12</td>
<td>0.514</td>
<td>0.458</td>
<td>0.518</td>
<td>0.445</td>
<td>0.515</td>
<td>40</td>
<td>0.182</td>
<td>0.285</td>
<td>0.192</td>
<td>0.181</td>
<td>0.178</td>
</tr>
<tr>
<td>13</td>
<td>0.484</td>
<td>0.433</td>
<td>0.485</td>
<td>0.412</td>
<td>0.479</td>
<td>41</td>
<td>0.376</td>
<td>0.413</td>
<td>0.384</td>
<td>0.375</td>
<td>0.376</td>
</tr>
<tr>
<td>14</td>
<td>0.474</td>
<td>0.455</td>
<td>0.472</td>
<td>0.451</td>
<td>0.477</td>
<td>42</td>
<td>0.991</td>
<td>0.853</td>
<td>0.977</td>
<td>0.987</td>
<td>0.992</td>
</tr>
<tr>
<td>15</td>
<td>0.061</td>
<td>0.280</td>
<td>0.070</td>
<td>0.056</td>
<td>0.062</td>
<td>43</td>
<td>0.880</td>
<td>0.699</td>
<td>0.872</td>
<td>0.866</td>
<td>0.883</td>
</tr>
<tr>
<td>16</td>
<td>0.578</td>
<td>0.496</td>
<td>0.571</td>
<td>0.540</td>
<td>0.578</td>
<td>44</td>
<td>0.599</td>
<td>0.532</td>
<td>0.585</td>
<td>0.588</td>
<td>0.593</td>
</tr>
<tr>
<td>17</td>
<td>0.609</td>
<td>0.473</td>
<td>0.602</td>
<td>0.536</td>
<td>0.606</td>
<td>45</td>
<td>0.962</td>
<td>0.798</td>
<td>0.904</td>
<td>0.973</td>
<td>0.971</td>
</tr>
<tr>
<td>18</td>
<td>0.138</td>
<td>0.303</td>
<td>0.146</td>
<td>0.144</td>
<td>0.136</td>
<td>46</td>
<td>0.802</td>
<td>0.664</td>
<td>0.788</td>
<td>0.807</td>
<td>0.802</td>
</tr>
<tr>
<td>19</td>
<td>0.369</td>
<td>0.422</td>
<td>0.378</td>
<td>0.373</td>
<td>0.366</td>
<td>47</td>
<td>0.510</td>
<td>0.470</td>
<td>0.506</td>
<td>0.506</td>
<td>0.511</td>
</tr>
<tr>
<td>20</td>
<td>0.271</td>
<td>0.366</td>
<td>0.277</td>
<td>0.245</td>
<td>0.271</td>
<td>48</td>
<td>0.687</td>
<td>0.598</td>
<td>0.684</td>
<td>0.692</td>
<td>0.688</td>
</tr>
<tr>
<td>21</td>
<td>0.133</td>
<td>0.309</td>
<td>0.139</td>
<td>0.127</td>
<td>0.129</td>
<td>49</td>
<td>0.987</td>
<td>0.865</td>
<td>0.949</td>
<td>0.983</td>
<td>0.987</td>
</tr>
<tr>
<td>22</td>
<td>0.734</td>
<td>0.572</td>
<td>0.695</td>
<td>0.700</td>
<td>0.744</td>
<td>50</td>
<td>0.954</td>
<td>0.819</td>
<td>0.930</td>
<td>0.951</td>
<td>0.955</td>
</tr>
<tr>
<td>23</td>
<td>0.382</td>
<td>0.427</td>
<td>0.390</td>
<td>0.381</td>
<td>0.384</td>
<td>51</td>
<td>0.590</td>
<td>0.519</td>
<td>0.586</td>
<td>0.581</td>
<td>0.591</td>
</tr>
<tr>
<td>24</td>
<td>0.106</td>
<td>0.278</td>
<td>0.140</td>
<td>0.118</td>
<td>0.109</td>
<td>52</td>
<td>0.574</td>
<td>0.512</td>
<td>0.571</td>
<td>0.576</td>
<td>0.575</td>
</tr>
<tr>
<td>25</td>
<td>0.075</td>
<td>0.259</td>
<td>0.093</td>
<td>0.079</td>
<td>0.073</td>
<td>53</td>
<td>0.757</td>
<td>0.657</td>
<td>0.748</td>
<td>0.750</td>
<td>0.757</td>
</tr>
<tr>
<td>26</td>
<td>0.049</td>
<td>0.224</td>
<td>0.061</td>
<td>0.052</td>
<td>0.048</td>
<td>54</td>
<td>0.847</td>
<td>0.739</td>
<td>0.837</td>
<td>0.841</td>
<td>0.847</td>
</tr>
<tr>
<td>27</td>
<td>0.244</td>
<td>0.348</td>
<td>0.250</td>
<td>0.248</td>
<td>0.244</td>
<td>55</td>
<td>0.990</td>
<td>0.923</td>
<td>0.987</td>
<td>0.990</td>
<td>0.991</td>
</tr>
<tr>
<td>28</td>
<td>0.305</td>
<td>0.383</td>
<td>0.315</td>
<td>0.302</td>
<td>0.308</td>
<td>56</td>
<td>0.841</td>
<td>0.728</td>
<td>0.833</td>
<td>0.826</td>
<td>0.842</td>
</tr>
</tbody>
</table>
Illustration of Optimistic Bias in Posterior Checking

(a) CV predictive PMF of $y_2$

(b) Posterior Checking PMF of $y_2$
Comparing Estimated $p$-values with LOOCV $p$-values

(c) Posterior checking

(d) Ghosting method
Comparing Estimated $p$-values with LOOCV $p$-values II

4. Numerical Comparisons with Two Real Datasets/Lip Cancer Data in Scotland

(e) Non-integrated IS (nIS)  
(f) Integrated IS (iIS)
Box-plots of Relative Errors in the Estimated $p$-value

$$RE = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \frac{|\hat{p}_i - p_i|}{\min(p_i, 1-p_i)} \times 100,$$

4. Numerical Comparisons with Two Real Datasets/Lip Cancer Data in Scotland
Computation Time

**Table 1:** Comparison of computation time (in seconds). (Abbreviations: LOOCV: actual cross validation, PCH: posterior predictive checking, GHO: Ghosting, nIS: naive importance sampling and iIS: integrated importance sampling).

<table>
<thead>
<tr>
<th></th>
<th>LOOCV</th>
<th>PCH</th>
<th>nIS</th>
<th>GHO</th>
<th>iIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC fitting</td>
<td>1138</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Computing $p$-values</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>84</td>
<td>144</td>
</tr>
<tr>
<td>Total</td>
<td>1139</td>
<td>21</td>
<td>21</td>
<td>104</td>
<td>164</td>
</tr>
</tbody>
</table>
Subsection 2

Larynx Cancer Data in Germany
Dataset Information

- $N = 544$ districts.
- $y_i$: number of larynx cancer mortality counts
- $x_i$: level of smoking consumption
Comparing Estimated $p$-values with LOOCV $p$-values

Integrated Importance Sampling

Posterior Predictive Checking

Non-integrated Importance Sampling

Ghosting Method

4. Numerical Comparisons with Two Real Datasets/Larynx Cancer Data in Germany
### Table 2: Contingency table of districts categorized by cutting predictive $p$-values with 0.1 and 0.9 in German larynx cancer example. The bolded numbers show the mis-categorized districts compared to CV.

<table>
<thead>
<tr>
<th>CV</th>
<th>Posterior predictive checking</th>
<th>Ghosting method</th>
<th>nIS</th>
<th>iIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0,0.1) [0.1,0.9) [0.9,1]</td>
<td>[0,0.1) [0.1,0.9) [0.9,1]</td>
<td>[0,0.1) [0.1,0.9) [0.9,1]</td>
<td>[0,0.1) [0.1,0.9) [0.9,1]</td>
</tr>
<tr>
<td>[0, 0.1)</td>
<td>16 31 0</td>
<td>42 5 0</td>
<td>47 0 0</td>
<td>47 0 0</td>
</tr>
<tr>
<td>[0.1, 0.9)</td>
<td>0 455 0</td>
<td>0 455 0</td>
<td>0 454 1</td>
<td>0 455 0</td>
</tr>
<tr>
<td>[0.9, 1]</td>
<td>0 21 21</td>
<td>0 3 39</td>
<td>0 0 42</td>
<td>0 0 42</td>
</tr>
</tbody>
</table>
## Computation Time

### Table 3: Comparison of computation time in larynx cancer example.

<table>
<thead>
<tr>
<th></th>
<th>LOOCV</th>
<th>PCH</th>
<th>nIS</th>
<th>GHO</th>
<th>iIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC fitting (seconds)</td>
<td>$4.8 \times 10^6$</td>
<td>7816</td>
<td>7816</td>
<td>7816</td>
<td>7816</td>
</tr>
<tr>
<td>Computing $p$-values</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>284</td>
<td>522</td>
</tr>
<tr>
<td>(seconds)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (hours)</td>
<td>1333</td>
<td>2.17</td>
<td>2.17</td>
<td>2.25</td>
<td>2.32</td>
</tr>
<tr>
<td>Total (relative to CV)</td>
<td>1</td>
<td>0.162%</td>
<td>0.162%</td>
<td>0.168%</td>
<td>0.173%</td>
</tr>
</tbody>
</table>
Conclusion and Discussion

- Non-integrated IS by treating latent variables as parameters may give wrong results in predictive model assessment.
- The new proposed iIS can improve the accuracy of IS in assessing Bayesian models with unit-specific latent variables. In our studies, they gave results very close to what given by the actual cross-validation.
- The iIS method can be applied to many other models with correlated or independent random effects provided that the random effect is specific to each test observation or unit, for example the zero-inflated models, which is a special case of mixture models.
- iIS and Ghosting method are not yet applicable in models with complex structure in latent variables. For such models, non-integrated importance sampling is still valid. There is an improved importance sampling that is more widely applicable: “Pareto smoothed importance sampling” (Vehtari and Gelman, 2015).

