Project Paper Update

Tracey Marsh

Group Health Research Institute
University of Washington, Department of Biostatistics

Adrain E. Raftery and Le Bao

Biometrics 2010
Incremental Mixture Importance Sampling

- numerical algorithm for sampling from a posterior
- addresses limitations of current algorithm
  Sampling Importance Resampling (SIR)
  posteriors with multimodality & nonlinear ridges
Bayesian framework in the context of modeling

- **Model**: \( \theta \xrightarrow{M} \rho \)
  - \( M \) deterministic scientific model
    - population dynamics model for bowhead whales
  - \( \theta \) input parameters for model (settings)
    - birth and death rates, initial population size, ....
  - \( \rho = M(\theta) \) model output
    - population size by year

- **Model Calibration (select \( \theta \))**
  - prior on input parameters: \( \pi(\theta) \)
  - \( X \) is observed real data
    - for certain years: observed population counts
  - posterior on input parameters: \( \pi(\theta|X) \)
    - use \( E[\theta|X] \) for point estimates of \( \rho \)
    - samples from \( \pi(\theta|X) \) for intervals around \( \hat{\rho} \)

- **Bayes Formula**: \( \pi(\theta|X) = \frac{L(X|M(\theta))\pi(\theta)}{L(X)} \)
Sampling Importance Re-sampling

- sample from prior: \( \{\theta_1, \ldots, \theta_N\} \sim \pi_0 \)
- run model to get output for each: \( \{M(\theta_1), \ldots, M(\theta_N)\} \)
- calculate likelihood of model output: \( L_i \equiv L(X|M(\theta_i)) \)
- calculate importance weight: \( \omega_i = \frac{L_i}{\sum_j L_j} \)
- weighted re-sample of \( \theta_1, \ldots, \theta_N \)
  estimate of posterior: \( \pi(\theta|X) \)

Note: unique points of posterior will always be a subset of the unique points sampled from prior.
Incremental Mixture Importance Sampling

- Initial sample and weights: \( \{ (\theta_i, \omega_i = \frac{L_i}{\sum_j L_j}) \mid 1 \leq i \leq N_0 \} \)
  resampling now gives SIR estimate of posterior

- 'fill-in' important regions: \( 1 \leq k \leq K \)
  identify underrepresented neighborhood
  add \( B \) points of Normal mass
  use mixture distribution as new prior: \( \pi_k \)
  update weights: \( \omega_i^k \)

- repeat until stopping criteria met: \( K \) times
  expected % unique points in resample \( \geq 1 - 1/e \)

- weighted re-sample from \( \{ (\theta_i, \omega_i^K) \mid 1 \leq i \leq N_0 + KB \} \)
  estimate of posterior: \( \pi(\theta|X) \)
IMIS - Add Normal Mass

Expand sample with B points sampled from \( q_k = \mathcal{N}(\theta^k, \Sigma^k) \)

- \( \theta^k = \arg\max_{\theta_i} \{ \omega^{k-1}(\theta_i) \} \)
  center of important neighborhood
- \( \Sigma^k \)
  weighted covariance of \( B \) points of current sample in the neighborhood of \( \theta^k \)
  Mahalanobis metric w.r.t. \( \pi_0 \)
  weights \( \propto \omega_i + 1/N_k \)

\[
S_k = S_{k-1} \cup \{ \theta_{k,1} \ldots \theta_{k,B} \}
\]
\[
N_k \equiv \#S_k = N_0 + kB
\]
Side Note - Mahalanobis Metric

\[ d(\theta_i, \theta^k) = \sqrt{(\theta_i - \theta^k)^T \Sigma^{-1} (\theta_i - \theta^k)} \]

\( \Sigma \) is covariance matrix for \( \pi_0 \)

Can be thought of as a dissimilarity measure between two points of the same distribution with covariance \( \Sigma \).

Using this to grab all the B points of the sample that form the narrowest percentile range centered at \( \theta^k \).
IMIS - Update weights

At end of iteration $k$,

mixing sampling distribution:

- $\pi_k(\theta) = N_k^{-1} \left( N_0\pi_0(\theta) + B \sum_j q_k(\theta) \right)$

weights:

- $\omega^k_i \propto L_i(X|M(\theta_i)) \times \pi^0(\theta_i)\pi_k(\theta_i)^{-1}$

away from important spots: $adj \approx L_i$

near $\theta^k$: $adj \ll 1$

$\pi_k = \pi(\theta|X)$, then $\omega_i \propto 1$
Replication Goal

Reproduce results for two methods of interest from simulation study.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SIR</th>
<th>IMIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge-Like</td>
<td>2e-5</td>
<td>0.0675</td>
</tr>
<tr>
<td>Bimodal</td>
<td>0.0002</td>
<td>0.2063</td>
</tr>
</tbody>
</table>

Table: ESS/$N_K$

- Prior, model, and likelihood of model output specified
- Evaluated in terms of efficiency: $\frac{ESS}{\#evaluations}$
Simulation - Set-up

Scenario: Ridge-Like
Model:

\[
(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) \xrightarrow{M} \left( \prod_{i=1}^{4} \theta_i, \theta_2\theta_4, \frac{\theta_1}{\theta_5}, \theta_3\theta_6 \right)
\]

Prior:

\[
\pi_0(\theta) \sim \mathcal{N}(\mu_0, \text{Diag}(\sigma_0)) \quad \mu_0, \sigma_0 \text{ specified}
\]

Likelihood:

\[
L(X|M(\theta)) \sim \mathcal{N}(\mu_L, \text{Diag}(\sigma_L)) \quad \mu_L, \sigma_L \text{ specified}
\]
Methods - Evaluation

Evaluated in terms of efficiency: \( \frac{ESS}{N_k} \leq 1 \)

- \( ESS = \) effective sample size of \( N_k \) indirect samples
- compared to \( N_k \) direct samples from \( \pi(\theta | X) \)

Effective sample size [Kong JASA 1994]:

\[
ESS = \frac{N_k}{1 + CV} = \frac{N_k}{1 + \frac{VAR[\omega]}{E^2[\omega]}} = \frac{N_k}{E[\omega^2] E^2[\omega]}
\]

\[
\hat{ESS} = \frac{1}{\sum_i \omega_i^2} \quad \hat{E}[\omega] = 1/N_k
\]
Next Steps

• stopping criteria - understand expected % of unique points $\geq 1 - 1/e$
  value when weights all equal

• simulation
  numerical under/over-flow issues
  covariance estimates - not positive definite
  empirically approximate stopping criteria