
Adrain E. Raftery and Le Bao

Biometrics 2010
Method - IMIS

**Incremental Mixture Importance Sampling**

- numerical algorithm for sampling from a posterior
- addresses limitations of current algorithm Sampling-Importance-Resampling (SIR) posteriors with multi-modality & 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 $\mathbb{E}[\theta|X]$ for point estimates of $\rho$
    - samples from $\pi(\theta|X)$ for intervals around $\hat{\rho}$
- Bayes Formula: $\pi(\theta|X) = \frac{L(X|M(\theta))\pi(\theta)}{L(X)}$
Bayesian model calibration - early steps

- International Whaling Commission (IWC)
  - 1990s request for quantifying variability
- Bayesian Synthesis - Raftery
  - joint prior: $\pi(\theta, \phi)$
  - $\{(\theta, \phi) : \phi = M(\theta)\} \subset \{(\theta, \phi) : \theta \in \Theta, \phi \in \Phi\}$
  - Borel Paradox
- Bayesian Melding - Raftery
  - $p(\theta, \phi) = p^{[\theta]}(\theta)p^{[\phi]}(\phi) \propto q_1(\theta)q_2(\phi)L_1(D_\theta, \theta)L_2(D_\phi, \phi)$
Sampling-Importance-Resampling

- sample from prior: $S_{\pi_0} = \{\theta_1, \ldots, \theta_N\} \sim \pi_0(\theta)$
- 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 $\theta_1, \ldots, \theta_N = S_{\pi}$
  estimate of posterior: $\pi(\theta|X)$

Note: unique points of posterior will always be a subset of the unique points sampled from prior.
Weighted Sampling - Concept

View multiplication between prior and likelihood

\[ \pi(\theta|X) \propto \frac{L(M(\theta)|X)}{\omega} \pi_0(\theta) \]

as effected by the weighting of a sample from the prior:

\[ \overbrace{\pi(\theta|X)}^{S_\pi} = S(\text{weight} = \omega, \text{set} = \overbrace{\pi_0(\theta)}^{S_{\pi_0}}) \]

Weighted re-sample, \( S_\pi \), can thus be seen as a sample from the posterior.
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
  mixture distribution is new sampling distribution: \( q_k \)
  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 $H_k = N(\theta^k, \Sigma^k)$

- $\theta^k = \arg \max_{\theta_i} \{\omega^{k-1}(\theta_i)\}$
  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 $\propto \omega_i + 1/N_k$

$S_k = S_{k-1} \cup \{\theta_{k,1} \ldots \theta_{k,B}\}$
$N_k = N_0 + kB$
At end of iteration $k$, mixture sampling distribution:

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

weights:

- $\omega_i^{k} \propto L_i(X|M(\theta_i)) \times \frac{\pi_0(\theta_i)}{q_k(\theta_i)} \stackrel{adj}{\propto} \frac{1}{q_k}$ if $q_k = \pi(\theta|X)$, then $\omega_i \propto 1$
Weighted Sampling - Concept

View multiplication between prior and likelihood

\[
\pi(\theta | X) \propto L(\theta | X) \pi_0(\theta)
\]
\[
\propto L(\theta | X) \frac{\pi_0(\theta)}{q_k(\theta)} q_k(\theta)
\]

as effected by the weighting of a sample from the prior:

\[
\hat{\pi}(\theta | X) \left( weight = \omega_k, set = \hat{q}_k(\theta) \right)
\]

Weighted re-sample, \( S_\pi \), can thus be seen as a sample from the posterior (at any iteration \( k \)).
IMIS- Stopping Criteria

- Role of stopping criteria is QA
- Stop when expected fraction of unique points in re-sample $\geq 1 - 1/e$
  - Expected number of unique points in re-sample
    \[ \sum_{i=1}^{N_k} \left\{ 1 - \left( 1 - \omega_i^k \right)^J \right\} \]
- Property arises in idealized scenario
  - when sampling distribution is posterior: $\omega_i = \frac{1}{N_k}$
  - $1/e = \lim_{N_k \to \infty} \left( 1 - 1/N_k \right)^{N_k}$
Simulation - Ridge Set-up

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:

\(\pi_0(\theta) \sim N(\mu_0, \text{Diag}(\sigma_0^2)) \quad \mu_0, \sigma_0^2 \text{ specified}\)

Likelihood:

\(L(X|M(\theta)) \sim N(\mu_L, \text{Diag}(\sigma_L^2)) \quad \mu_L, \sigma_L^2 \text{ specified}\)
Model:

none - evaluating a prior outside of modeling context

Prior:

\[ \pi_0(\theta) \sim Uni([-3, 12]^4) \] specified

Likelihood:

\[ L(X|\theta) \sim \frac{1}{2} N_4(\mu = 0, \text{AR}(-0.95)) + \frac{1}{2} N_4(\mu = 9, \text{AR}(0.95)) \]
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]:

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

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

Reporting efficiency: $\text{ESS}/N_K$ and sample set size $N_k$.  
IMIS: $N_0 = d \times 1,000$, $B = d \times 100$, $J = 3,000$  
SIR: $N = N_k$

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$d$</th>
<th>SIR</th>
<th>IMIS</th>
<th>$N_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge</td>
<td>6</td>
<td>2e-5</td>
<td>0.0675</td>
<td>35,400</td>
</tr>
<tr>
<td>Bimodal</td>
<td>4</td>
<td>0.0002</td>
<td>0.2063</td>
<td>16,800</td>
</tr>
</tbody>
</table>

Table: Results presented in paper.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SIR</th>
<th>IMIS</th>
<th>$N_k$</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge</td>
<td>0.00026</td>
<td>0.092</td>
<td>27,000</td>
<td>Marsh</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.087</td>
<td>27,600</td>
<td>R Pkg</td>
</tr>
<tr>
<td>Bimodal</td>
<td>1.1e-4</td>
<td>0.2555</td>
<td>11,600</td>
<td>Marsh</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2442</td>
<td>11,200</td>
<td>R Pkg</td>
</tr>
</tbody>
</table>

Table: Results by Marsh.
Simulation Results - Variability

Reporting efficiency: $\text{ESS}/N_K$ and sample set size $N_k$.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SIR</th>
<th>IMIS</th>
<th>$N_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge</td>
<td>2e-5</td>
<td>0.0675</td>
<td>35,400</td>
</tr>
<tr>
<td>Bimodal</td>
<td>0.0002</td>
<td>0.2063</td>
<td>16,800</td>
</tr>
</tbody>
</table>

Table: Results presented in paper.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SIR</th>
<th>IMIS</th>
<th>$N_k$ (1,000s)</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ridge</td>
<td>(5e-5)-0.0004</td>
<td>0.082-0.110</td>
<td>22.2-30.6</td>
<td>Marsh</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.081-0.115</td>
<td>22.2-30.6</td>
<td>R Pkg</td>
</tr>
<tr>
<td>Bimodal</td>
<td>(9e-5)-0.0006</td>
<td>0.224-0.349</td>
<td>9.2-12.4</td>
<td>Marsh</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.225-0.341</td>
<td>8.8-12.0</td>
<td>R Pkg</td>
</tr>
</tbody>
</table>

Table: Ranges of results from 100 runs by Marsh.
Promise of IMIS

Good for:

- priors with features like non-linear ridges and bi-modality
  - typical of deterministic scientific models
  - probably applicable to stochastic scientific models
- low ($< 30$) dimensional parameters
- potential efficiency gains