Project Paper - Final Review

Tracey Marsh

Group Health Research Institute
University of Washington, Department of Biostatistics

Paper

Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling

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
  - θ input parameters for model (settings)
    birth and death rates, initial population size, ...
  - $\rho = M(\theta)$ model output
    population size by year
- Model Calibration (select θ)
  - 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 ρ
    samples from $\pi(\theta|X)$ for intervals around $\widehat{\rho}$
- 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: $\mathcal{S}_{\pi_0} = \{\theta_1, \ldots, \theta_N\} \sim \pi_0(\theta)$
- 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 $\theta_1, \ldots, \theta_N = \mathcal{S}_\pi$
  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

$$\pi(\theta|X) \propto \underbrace{L(M(\theta)|X)}_{\omega}\pi_0(\theta)$$

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

$$\underbrace{\widehat{\pi(\theta|X)}}_{\mathcal{S}_\pi} = \mathcal{S}(weight = \omega, set = \underbrace{\widehat{\pi_0(\theta)}}_{\mathcal{S}_{\pi_0}})$$

Weighted re-sample, $\mathcal{S}_\pi$, can thus be seen as a sample from the posterior.

Incremental Mixture Importance Sampling

- Initial sample and weights: $\left\{\left(\theta_i, \omega_i = \frac{L_i}{\sum_j L_j}\right) 1 \le i \le N_0\right\}$
  resampling now gives SIR estimate of posterior
- 'fill-in' important regions: $1 \le k \le K$
  identify underrepresented neighborhood
  add B points of Normal mass
  mixture distribution is new sampling distribution: q_k
  update weights: ω_i^k
- repeat until stopping criteria met: K times
  expected % unique points in resample $\ge 1 - 1/e$
- weighted re-sample from $\left\{\left(\theta_i, \omega_i^K\right) 1 \le i \le N_0 + KB\right\}$
  estimate of posterior: $\pi(\theta|X)$

IMIS - Add Normal Mass

Expand sample with B points sampled from $H_k = N\left(\theta^k, \Sigma^k\right)$

- $\theta^k = \arg\max_{\theta_i}\{\omega^{k-1}(\theta_i)\}$
  center of important neighborhood
- Σ^k
  weighted covariance of B points of current sample
  in the neighborhood of θ^k
  Mahalanobis metric w.r.t. π_0
  weights $\propto \omega_i + 1/N_k$

$S_k = S_{k-1} \cup \{\theta_{k,1} \ldots \theta_{k,B}\}$
$N_k = N_0 + kB$

IMIS -Update weights

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:

- $\omega_i^k \propto L_i(X|M(\theta_i)) \times \underbrace{\frac{\pi_0(\theta_i)}{q_k(\theta_i)}}_{adj}$

  if $q_k = \pi(\theta|X)$, then $\omega_i \propto 1$

Weighted Sampling - Concept

View multiplication between prior and likelihood

$$\begin{aligned}\pi(\theta|X) &\propto L(\theta|X)\pi_0(\theta)\\ &\propto \underbrace{L(\theta|X)\frac{\pi_0(\theta)}{q_k(\theta)}}_{\omega^k} q_k(\theta)\end{aligned}$$

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

$$\underbrace{\widehat{\pi(\theta|X)}}_{\mathcal{S}_\pi} = \mathcal{S}(weight = \omega^k, set = \underbrace{\widehat{q_k(\theta)}}_{\mathcal{S}_k})$$

Weighted re-sample, $\mathcal{S}_\pi$, 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 $\sum_{i=1}^{N_k}\{1-(1-\omega_i^k)^J\}$
- Property arises in idealized scenario
  - when sampling distribution is posterior: $\omega_i = \frac{1}{N_k}$
  - $1/e = \lim_{N_k \to \infty}(1 - 1/N_k)^{N_k}$

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:

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

Likelihood:

$$L(X|M(\theta)) \sim N(\mu_L, \mathrm{Diag}(\sigma_L^2)) \qquad \mu_L, \sigma_L^2 \text{ specified}$$

Simulation - Bi-modal Set-up

Model:

none - evaluating a prior outside of modeling context

Prior:

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

Likelihood:

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

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]:

$$
\begin{aligned}
\text{ESS} &= \frac{N_K}{1 + CV} \\
&= \frac{N_K}{1 + \frac{VAR[\omega]}{\mathbb{E}^2[\omega]}} \\
&= \frac{N_K}{\frac{\mathbb{E}[\omega^2]}{\mathbb{E}^2[\omega]}} \\
\widehat{\text{ESS}} &= \frac{1}{\sum_i \omega_i^2} \qquad \widehat{\mathbb{E}[\omega]} = 1/N_k
\end{aligned}
$$

Simulation Results

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

Scenario	d	SIR	IMIS	N_k
Ridge	6	2e-5	0.0675	35,400
Bimodal	4	0.0002	0.2063	16,800

Table : Results presented in paper.

Scenario	SIR	IMIS	N_k	Code
Ridge	0.00026	0.092	27,000	Marsh
	-	0.087	27,600	R Pkg
Bimodal	1.1e-4	0.2555	11,600	Marsh
	-	0.2442	11,200	R Pkg

Table : Results by Marsh.

Simulation Results - Variability

Reporting efficiency: ESS/N_K and sample set size N_k.

Scenario	SIR	IMIS	N_k
Ridge	2e-5	0.0675	35,400
Bimodal	0.0002	0.2063	16,800

Table : Results presented in paper.

Scenario	SIR	IMIS	N_k (1,000s)	Code
Ridge	(5e-5)-0.0004	0.082-0.110	22.2-30.6	Marsh
	-	0.081-0.115	22.2-30.6	R Pkg
Bimodal	(9e-5)-0.0006	0.224-0.349	9.2-12.4	Marsh
	-	0.225-0.341	8.8-12.0	R Pkg

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