Project Paper Introduction

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

Context - Modeling

- Joint United Nations Programme on HIV/AIDS
- Estimation and Projection Package
  - IN - force of infection, start year of epidemic, initial proportion at risk, behaviour adjustment
  - OUT - sequences of yearly prevalence rates
  - Informed by two key types of data (antenatal and DHS)
- Current methodology (SIR) works well for *most* countries
- New methodology (IMIS) developed to address un-*most*

This is a naturally occurring feedback loop between a public health agency, a working model, methods development, and application.

Another Modeling Context - Personal Motivation

Cancer Intervention and Surveillance Modeling Network

- Carolyn Rutter (GHRI)
- **C**olo**r**ectal **C**ancer **S**imulated **P**opulation for **I**ncidence and **N**atural history (CRC-SPIN)
  - IN: 23 parameters
  - OUT: individual life histories (birthdate, date of non-CRC death, CRC onset etc.)
  - Informed by multiple external data sources: clinical trials, expert opinion, autopsy studies, SEER data etc.
- Previous methodology - MCMC
- Future methodology - considering IMIS

Model Calibration

Calibration is a process of setting parameter space so that model will produce results reasonably consistent with observed data.

- ad-hoc 'tuning'
- systematic searches over parameter space (un/directed)
- Bayesian framework
  - posterior reflects synthesis of information from multiple sources (prior and data)
  - posterior reflects epidemiological model
  - **calibrated model**: posterior determines inputs
    : posterior enables intervals

Bayesian framework in the context of modeling

- Bayes Formula: $\pi(\theta|D) = \frac{L(D|\rho)\pi(\theta)}{L(D)}$
- Model: $\theta \rightarrow$ M $\rightarrow \rho$
- Bayesian Melding: ???, $L(D|\theta)$, $D = f(\rho, \theta)$, ???

This is currently a major point of confusion on my part.

Addressing with background reading, talking with Carolyn, finding a simple model to work with.

Bayesian Framework - Numerical Analysis

We saw in the 570s that challenges arise quickly when applying numerical methods to Bayesian analysis

- informative priors
- burn-in
- correlated samples
- minimal jumping
- multi-modality *
- non-linear ridges *
- computation time ◦

Mechanistic models can lead to multi-modality and non-linear ridges [sic Raftery & Bao]. Very often only complicated functions of the parameters can be estimated.

Sampling Importance Re-sampling

- sample from prior: $\theta_1, \ldots, \theta_N$
- run model to get output for each: $\rho_1, \ldots, \rho_N$
- calculate likelihood of model output: $L(\rho_1), \ldots, L(\rho_N)$
- calculate importance weight: $\omega_i = \frac{L_i}{\sum_j L_j}$
- weighted re-sample of $\theta_1, \ldots, \theta_N$

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

Incremental Mixture Sampling

- SIR through importance weights: $\omega_i = \frac{L_i}{\sum_j L_j}$
- determine parameter value of greatest importance:
  $\theta_i \ni \omega_i = max\{\omega_1, \ldots, \omega_N\}$
- add Normal mass around θ_i
- use mixture distribution as new prior
- repeat until expected fraction of unique points $\geq 1 - 1/e$

The Competition

Section 4 of article - simulation study used to evaluate methods.

- 6 Methods: IMIS, IMIS-opt, SIR, MCMCMetrop1R, Metrop, WinBUGS
- 2 scenarios: thin manifold, bi-modal
  (prior, model, and likelihood of model output specified)
- Evaluated in terms of efficiency: $\frac{ESS}{\#evaluations}$

Section 5 of article considers application of the methods using EPP and data for Zimbabwe.

Promise of IMIS

IMIS is a modification of the SIR algorithm and purports to keep the strengths of SIR while addressing observed issues with some countries: ridges and multi-modality.

Bonus features: integrated likelihood useful for Bayesian model comparison and averaging.

IMIS focuses on important regions of the parameter space and could be much more computationally efficient than previous Bayesian methods used to calibrate CRC-Spin model.