

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

Adrian E. Raftery
University of Washington

Le Bao
University of Washington

Technical Report no. 560
Department of Statistics
University of Washington
Seattle, WA 98195-4322, USA.

August 15, 2009

Abstract

The Joint United Nations Programme on HIV/AIDS (UNAIDS) has decided to use Bayesian melding as the basis for its probabilistic projections of HIV prevalence in countries with generalized epidemics. This combines a mechanistic epidemiological model, prevalence data and expert opinion. Initially, the posterior distribution was approximated by sampling-importance-resampling, which is simple to implement, easy to interpret, transparent to users and gave acceptable results for most countries. For some countries, however, this is not computationally efficient because the posterior distribution tends to be concentrated around nonlinear ridges and can also be multimodal. We propose instead Incremental Mixture Importance Sampling (IMIS), which iteratively builds up a better importance sampling function. This retains the simplicity and transparency of sampling importance resampling, but is much more efficient computationally. It also leads to a simple estimator of the integrated likelihood that is the basis for Bayesian model comparison and model averaging. In simulation experiments and on real data it outperformed both sampling importance resampling and three publicly available generic Markov chain Monte Carlo algorithms for this kind of problem.

Contents

1	Introduction	1
2	The UNAIDS Estimation and Projection Package and Bayesian Melding	3
3	Incremental Mixture Importance Sampling	4
3.1	Basic IMIS Algorithm	4
3.2	IMIS with optimization	6
3.3	Estimating the Integrated Likelihood using IMIS	7
4	Simulated Examples	9
4.1	Methods for Comparison	9
4.2	Ridge-like Simulated Example	10
4.3	Bimodal Example	11
5	Application to EPP for Zimbabwe	14
5.1	Estimation and Projection Results	14
5.2	Model Comparison and Model Averaging	17
6	Discussion	17

List of Tables

1	Efficiencies of Different Methods for the Simulated Examples and EPP for Zimbabwe.	11
2	Efficiencies of Different Methods for the EPP applied to Zimbabwe	15

List of Figures

1	Bimodal Simulated Example	13
2	EPP trajectories for 2004 Zimbabwe Urban Data	14
3	Pairs plots of the posterior inputs for 2004 Zimbabwe Urban Data	16
4	Model selection and model averaging for Zimbabwe urban data	18

1 Introduction

The Joint United Nations Programme on HIV/AIDS (UNAIDS) publishes updated estimates and projections of the number of people living with HIV/AIDS in the countries with generalized epidemics every two years. As part of this, statements of uncertainty are also provided. Generalized epidemics are defined by overall prevalence being above 1%, and the epidemic not being confined to particular subgroups; there are about 38 such countries (Ghys et al. 2004).

UNAIDS bases its estimates and projections on data from two main sources: prevalence estimates from ante-natal clinics and estimates from Demographic and Health Surveys (DHS), representative of the population. The ante-natal clinic data are sparser and less representative, but they are available in all the countries and are typically available for many time periods, allowing estimation of trends. The DHS data are more representative, but they are not available for all countries, and for most countries are available at only one time point. Thus the DHS data give better information about overall prevalence, when they are available, but the ante-natal clinic data give better information about trends.

The quality of the data available varies widely from country to country. As a result, UNAIDS bases its estimates on a relatively simple method that can be supported by the data in all the countries, and so can give estimates that are comparable between countries. This is based on a simple standard epidemiological model with four adjustable parameters. The estimates and projections are produced using a Bayesian melding method that combines the epidemiological model with a hierarchical random effects model for the sampling variability of the data (Alkema et al. 2007; Alkema et al. 2008). Out-of-sample predictive assessments by Alkema et al. (2007) indicated that the overall method performed well over the five-year time horizon of the projections.

This method was incorporated into the Estimation and Projection Package (EPP) produced by UNAIDS and available for download from <http://www.epidem.org>. This is used by the UNAIDS Secretariat and also by national officials producing their own estimates and projections. It was used to produce the 2007 update of the UNAIDS estimates and projections (UNAIDS 2007; Ghys et al. 2008). This update attracted a great deal of attention because it featured a large change, notably a revision of the estimate of the number of people with HIV/AIDS worldwide downwards from 39 million to 33 million.

The Bayesian melding posterior distribution is produced using the sampling-importance-resampling (SIR) algorithm, with the first sample drawn from the prior distribution (Rubin 1987; Rubin 1988; Poole and Raftery 2000; Alkema et al. 2007). This consists of simulating a large number of samples from the prior distribution of the model parameters, weighting these

by their likelihoods, and then resampling them with replacement and with the computed weights. This method is simple to implement and explain, and works well for most of the countries involved. It is not computationally efficient, however, but for most countries this does not matter much: the model runs quickly and so it is possible draw an initial sample of size 200,000, which is enough to give a reasonable estimate of the posterior in most cases.

For some countries, however, this was not enough to give a good picture of the uncertainty. Our goal here is to propose a generic method that is more efficient than the SIR algorithm while retaining its essential simplicity and transparency, and so can be used for all the countries involved. The problem is a difficult one because the posterior distribution tends to feature nonlinear ridges and to be multimodal, and so is unlike posterior distributions typical of standard statistical models. However, these features are common in posterior distributions derived from mechanistic models, because these are often underdetermined by the data, which are typically adequate only to estimate some nonlinear functions of the parameters well (Raftery et al. 1995; Poole and Raftery 2000). This kind of posterior distribution is also challenging for generic Markov chain Monte Carlo (MCMC) algorithms.

We propose a new method, Incremental Mixture Importance Sampling (IMIS). The basic idea was first proposed in a different context by Steele et al. (2006). It is generic, relatively simple to implement and explain, and works well for countries where SIR does not. It also leads to a simple estimator of the integrated likelihood that is the basis for Bayesian model comparison and model averaging; this is challenging to obtain using MCMC (Han and Carlin 2001; Raftery et al. 2007). In simulation experiments constructed to mimic the ridge-like characteristics and multimodality of the posterior, it substantially outperformed SIR and three different generic Metropolis algorithms. It has the advantage of yielding essentially independent samples from the posterior, and so avoiding the autocorrelation and burn-in problems that plague MCMC. It is a kind of adaptive importance sampling, a class of methods that also include sequential Monte Carlo and population Monte Carlo; we discuss links to other methods in Section 6.

In Section 2 we describe the EPP model and methodology used by UNAIDS in more detail. In Section 3 we describe our IMIS method for estimation, and also the simple estimator of the integrated likelihood to which it leads. In Section 4 we give results for some simulated examples, and in Section 5 we give results for estimation and projection for Zimbabwe, one of the countries for which SIR did not perform well.

2 The UNAIDS Estimation and Projection Package and Bayesian Melding

The UNAIDS Estimation and Projection Package is based on a simple susceptible-infected-removed epidemiological model that we will refer to as the “EPP model.” This involves four adjustable parameters or inputs, $\theta = (r, t_0, f_0, \phi)$, where r is the force of infection, t_0 is the start year of the HIV/AIDS epidemic, f_0 is the initial fraction of the adult population at risk of infection, and ϕ is the behavior adjustment parameter. The output ρ is a sequence of yearly HIV prevalence rates.

The model divides the population at time t into three groups: a not-at-risk group $X(t)$, an at-risk group $Z(t)$ and an infected group $Y(t)$. The model assumes a constant non-AIDS mortality rate μ and a constant fertility rate, and does not represent migration or age structure. The rate at which the sizes of the groups change are described by three differential equations:

$$\begin{cases} \frac{dX(t)}{dt} &= \left(1 - f\left(\frac{X(t)}{N(t)}, f_0, \phi\right)\right) E(t) - \mu X(t), \\ \frac{dZ(t)}{dt} &= f\left(\frac{X(t)}{N(t)}, f_0, \phi\right) E(t) - \left(\mu + r \frac{Y(t)}{N(t)} + \lambda(t)\right) Z(t), \\ \frac{dY(t)}{dt} &= \left(r \frac{Y(t)}{N(t)} + \lambda(t)\right) Z(t) - \int_0^t \left(r \frac{Y(\tau)}{N(\tau)} + \lambda(\tau)\right) Z(\tau) g(t - \tau) d\tau, \end{cases} \quad (1)$$

where $N(t) = X(t) + Z(t) + Y(t)$ is the total population, and the function $g(\tau)$ specifies the HIV death rate τ years after infection. The population being modeled is aged 15+. The number of new members at time t , $E(t)$, depends on the population size 15 years ago, the birth rate and the survival rate from birth to age 15. When individuals survive to age 15, they are assigned to either the not-at-risk group $X(t)$ or the at-risk group $Z(t)$. The fraction of the new 15-year-old members entering the at-risk group $Z(t)$ is given by $f\left(\frac{X(t)}{N(t)}, f_0, \phi\right)$.

The Bayesian melding approach (Raftery et al. 1995; Poole and Raftery 2000) was applied to the EPP model by Alkema et al. (2007). It proceeds as follows. A prior distribution is specified for θ . The UNAIDS Reference Group on Estimates, Modelling and Projections agreed on a default prior distribution (see Section 5), but users can specify their own. The observed antenatal clinic and DHS data give the likelihood $L(\rho)$ for the model output, using a hierarchical random effects model. A prior on the model output ρ can also be specified; this is currently restricted to being uniform between specified bounds for specific years.

In the initial UNAIDS implementation, the Bayesian melding procedure computed the posterior distributions using the SIR algorithm of Rubin (1987, 1988), with the prior as importance sampling distribution. It consists of the following stages:

1. Obtain N independent and identically distributed (i.i.d.) samples, $\{\theta_1, \dots, \theta_N\}$, from the input prior distribution $p(\theta)$.

2. For each θ_i , determine the corresponding series of prevalence rates ρ_i by running the EPP model, and calculate the likelihood $L_i = L(\rho_i)$.
3. Set the importance weights for each ρ_i (and thus for each θ_i): $w_i = L_i / \sum_{j=1}^N L_j$.
4. Sample from the multinomial distribution of $\{\theta_1, \dots, \theta_N\}$ with probabilities $\{w_1, \dots, w_N\}$ to approximate the posterior distribution of the inputs and outputs.

The default value of N is 200,000, and the EPP model runs fast enough for this to be feasible.

The SIR algorithm performs acceptably for most of the countries involved. For some, however, it does not work well. In these cases there are a small number of large importance weights, and so the resample is dominated by large numbers of repetitions of the same few points. Then there are few distinct values present in the resample and the target distribution is poorly sampled. This situation occurs because the posterior distribution tends to be concentrated near thin curved manifolds, and in addition is often multimodal. In the most extreme case we have seen, for the urban parts of Ghana, SIR found only 8 unique points out of 3,000 resamples.

3 Incremental Mixture Importance Sampling

3.1 Basic IMIS Algorithm

To overcome this problem, we extend incremental mixture importance sampling (IMIS) to the Bayesian melding problem. IMIS is an iterative generalization of defensive mixture importance sampling (Hesterberg 1995), and was introduced by Steele et al. (2006) in a very different context, that of a high-dimensional discrete distribution.

The basic idea of IMIS is that points with high importance weights are in areas where the target density is underrepresented by the importance sampling distribution. At each iteration, a multivariate normal distribution centered at the point with the highest importance weight is added to the current importance sampling distribution, which thus becomes a mixture of such functions and of the prior. In this way underrepresented parts of the parameter space are successively identified and are given representation, ending up with an iteratively constructed importance sampling distribution that covers the target distribution well.

The algorithm ends when the importance sampling weights are reasonably uniform. Specifically, we end the algorithm when the expected fraction of unique points in the resample is at least $1 - 1/e = 0.632$. This is the expected fraction when the importance

sampling weights are all equal, which is the case when the importance sampling function is the same as the target distribution. The IMIS algorithm is as follows:

1. Initial stage:

- (a) Sample N_0 inputs $\theta_1, \theta_2, \dots, \theta_{N_0}$ from the prior distribution $p(\theta)$.
- (b) For each θ_i , calculate the likelihood L_i , and form the importance weights:

$$w_i^{(0)} = \frac{L_i}{\sum_{j=1}^{N_0} L_j}.$$

2. Importance Sampling Stage: For $k = 1, 2, \dots$, repeat the following steps:

- (a) Choose the current maximum weight input as the center $\theta^{(k)}$. Estimate $\Sigma^{(k)}$ from the weighted covariance of the B inputs with the smallest Mahalanobis distances to $\theta^{(k)}$, where the distances are calculated with respect to the covariance of the prior distribution and the weights are taken to be proportional to the average of the importance weights and $\frac{1}{N_k}$.
- (b) Sample B new inputs from a multivariate Gaussian distribution H_k with covariance matrix $\Sigma^{(k)}$.
- (c) Calculate the likelihood of the new inputs and combine the new inputs with the previous ones. Form the importance weights:

$$w_i^{(k)} = cL_i \times \frac{p(\theta_i)}{q^{(k)}(\theta_i)},$$

where c is chosen so that the weights add to 1, $q^{(k)}$ is the mixture sampling distribution $q^{(k)} = \frac{N_0}{N_k}p + \frac{B}{N_k} \sum_{s=1}^k H_s$, H_s is the s -th multivariate normal distribution, and $N_k = N_0 + Bk$ is the total number of inputs up to iteration k .

3. Resample Stage: Once the stopping criterion is satisfied, resample J inputs with replacement from $\theta_1, \dots, \theta_{N_k}$ with weights w_1, \dots, w_{N_k} , where K is the number of iterations at the importance sampling stage.

The algorithm has several control parameters to be set by the user: the number of initial samples N_0 , the sample size at each importance sampling iteration B , and the number of resamples J . The algorithm is unbiased for any choice of control parameters, because it is an importance sampling algorithm, but the control parameters can affect its efficiency. We have found good results with the choices $N_0 = 1000d$, $B = 100d$ and $J = 3000$, where d is the dimension of the integrand.

A great advantage of importance sampling is that it is effectively self-monitoring, in that poor coverage of the target distribution by the importance sampling is immediately seen by the presence of large importance weights. We use the following specific criteria to assess the performance of the various importance sampling algorithms considered here:

1. The maximum importance weight among the N_k inputs.
2. The variance of the rescaled importance weights, $\hat{V}(w) = \frac{1}{N_k} \sum_{i=1}^{N_k} (N_k w_i - 1)^2$.
3. The entropy of the importance weights relative to uniformity, $\hat{U}(w) = - \sum_{i=1}^{N_k} w_i \frac{\log(w_i)}{\log(N_k)}$.
4. The expected number of unique points after re-sampling, $\hat{Q}(w) = \sum_{i=1}^{N_k} (1 - (1 - w_i)^J)$.
5. The effective sample size, $ESS(w) = \frac{N_k}{1+CV}$, where the coefficient of variation CV is defined as $CV = \frac{Var_q[L(\theta)p(\theta)/q(\theta)]}{E_q^2[L(\theta)p(\theta)/q(\theta)]} \doteq \hat{V}(w)$ (Kong et al. 1994). We can also write $ESS(w) = \frac{N_k}{1+\hat{V}(w)} = \frac{N_k}{1+N_k \sum_{i=1}^{N_k} w_i^2 - 2 \sum_{i=1}^{N_k} w_i + 1} = \frac{1}{\sum_{i=1}^{N_k} w_i^2}$.

3.2 IMIS with optimization

If the prior distribution disagrees strongly with the target distribution, particularly if the posterior is multimodal, the initial sample might completely miss a whole high probability region. In this case, IMIS might not work, as it proceeds essentially by filling in the importance sampling distribution to make it cover the target. If there are no initial points at all in an important unexplored region, IMIS would find it difficult to extend the coverage to that region.

As a remedy for this, we suggest inserting an optimization stage after the initial stage. This yields a mixture of D multivariate normal distributions, each one centered around a local maximum of the target distribution. It works as follows:

(1') Optimization stage:

- (a) Set the maximum weight input from the initial stage as the first starting point and use the optimizer to get a local maximum of the posterior density θ_1^{opt} , and store the inverse of the Hessian matrix as Σ_1^{opt} .
- (b) For $i = 2, \dots, D$, exclude the starting points and the $\frac{N_0}{D}$ inputs that have the smallest Mahalanobis distances from the previous local maximum, $(x - \theta_{i-1}^{opt})' \Sigma_{i-1}^{opt}{}^{-1} (x - \theta_{i-1}^{opt})$. Among the remaining inputs, set the maximum weight input as the new starting point to get θ_i^{opt} and Σ_i^{opt} as in step (a).

- (c) Sample B inputs from each of the D multivariate normal distributions with centers $\theta_1^{opt}, \dots, \theta_D^{opt}$ and covariance matrices $\Sigma_1^{opt}, \dots, \Sigma_D^{opt}$.

In our experiments, we used the L-BFGS-B method (Byrd, Lu, Nocedal, and Zhu 1995) as implemented in the R function `optim`, and we set the maximum number of function evaluations to be 100. When the Hessian matrix was not positive definite, we used the first derivative $g(\theta) = \frac{\partial(L \times p)}{\partial \theta}$ to build a new information matrix, and added it to the inverse of the diagonal of the prior covariance matrix, $diag(\Sigma^{(prior)})^{-1}$. The inverse of the new information matrix is then defined as $\Sigma_i^{opt} = [g'(\theta) \times g(\theta) + diag(\Sigma^{(prior)})^{-1}]^{-1}$. We found good results with $D = 10$. We refer to this version of IMIS as IMIS-opt.

3.3 Estimating the Integrated Likelihood using IMIS

The standard Bayesian approach to model comparison is based on the Bayes factor between two models (Jeffreys 1939; Kass and Raftery 1995), and the standard Bayesian approach to accounting for model uncertainty is Bayesian model averaging (Leamer 1978; Madigan and Raftery 1994; Hoeting et al. 1999; Clyde and George 2004). Both are based on the integrated likelihood of a model M , defined as

$$P(D|M) = \int p(D|\theta)\pi(\theta)d\theta, \quad (2)$$

where D denotes the data, θ denotes the parameter vector (or set of inputs) of the model, $p(D|\theta)$ is the usual likelihood function, and $\pi(\theta)$ is the prior density of θ under the model. We can see from equation (2) that the integrated likelihood is the integral over the parameter space of the product of the likelihood and the prior. The Bayes factor for model M_2 against model M_1 is then

$$B_{21} = \frac{P(D|M_2)}{P(D|M_1)}.$$

It has proven difficult to find a satisfactory general method to estimate the integrated likelihood using standard posterior simulation methods, including direct posterior simulation and MCMC. Many proposals have been made, however, and some work well for specific models; see Raftery et al. (2007) for a review of the literature. IMIS directly yields an estimate of the integrated likelihood that is simple to compute and has good theoretical properties.

We can rewrite (2) as

$$P(D|M) = \int \frac{p(D|\theta)\pi(\theta)}{q(\theta)}q(\theta)d\theta, \quad (3)$$

where $q(\theta)$ is the importance sampling distribution. The IMIS estimator of the integrated likelihood is then

$$\widehat{P(D|M)} = \frac{1}{N_K} \sum_{i=1}^{N_K} \frac{p(D|\theta_i)\pi(\theta_i)}{q(\theta_i)}, \quad (4)$$

where $\{\theta_1, \dots, \theta_{N_K}\}$ is the sample from the final IMIS importance sampling distribution $q(\theta) = q^{(K)}(\theta)$. Then the IMIS estimator has several desirable theoretical properties:

Theorem 1 *Assume that the posterior distribution of θ is proper in general. Then, as an estimator of $p(D|M)$, $\widehat{P(D|M)} = \frac{1}{N_K} \sum_{i=1}^{N_K} \frac{p(D|\theta_i)\pi(\theta_i)}{q^{(K)}(\theta_i)}$ is (i) unbiased, (ii) strongly consistent, and (iii) asymptotically normal, with*

$$\sqrt{N_K} \left(\widehat{P(D|M)} - p(D|M) \right) \sim \mathcal{N}(0, V^{(K)}), \quad (5)$$

where $q^{(K)}$ is the density of a mixture of prior distribution and K multivariate Gaussian distributions, $V^{(K)} = \int \frac{p(D|\theta)^2 \pi(\theta)^2}{q^{(K)}(\theta)} d\theta - p(D|M)^2$, and the asymptotics refer to the amount of simulation increasing, so that N_K tends to infinity at the K th stage.

Proof: Let X_1, \dots, X_{N_K} be realizations of the random variable $X \equiv X(\theta) = p(D|\theta)\pi(\theta)/q^{(K)}(\theta)$, where θ is a draw from $q^{(K)}(\theta)$ and $X = X(\theta_i)$. Then $\widehat{P(D|M)} = \frac{1}{N_K} \sum_{i=1}^{N_K} X_i$. It follows that

$$E \left(\widehat{P(D|M)} \right) = E(X) = E \left[\frac{p(D|\theta)\pi(\theta)}{q^{(K)}(\theta)} \right] = \int p(D|\theta)\pi(\theta)d\theta = p(D|M),$$

where all expectations are with respect to $q^{(K)}$. This proves (i).

Now we show that $E(X^2)$ is finite. This is the case because

$$\begin{aligned} E(X^2) &= \int \frac{p(D|\theta)^2 \pi(\theta)^2}{q^{(K)}(\theta)^2} q^{(K)}(\theta) d\theta = \int \frac{p(D|\theta)^2 \pi(\theta)^2}{q^{(K)}(\theta)} d\theta = \int \frac{p(D|\theta)^2 \pi(\theta)^2}{c\pi(\theta) + (1-c)r(\theta)} d\theta \\ &\quad \text{for some constant } c \text{ between 0 and 1, and for some positive function } r(\theta), \\ &\quad \text{by the definition of } q^{(K)}(\theta), \\ &< \int \frac{p(D|\theta)^2 \pi(\theta)^2}{c\pi(\theta)} d\theta = c^{-1} \int p(D|\theta)^2 \pi(\theta) d\theta < \infty, \end{aligned}$$

as required. The last inequality follows from the fact that $p(D|\theta)^2 \pi(\theta)$ is proportional to the posterior density for a dataset consisting of two replicates of the data D , and so its integral is finite by the general propriety of the posterior.

Then (ii) follows by the strong law of large numbers, and (iii) follows by the central limit theorem, with the asymptotic variance being given by $V^{(K)}$. \square

Furthermore, we define the variance of scaled importance weights $\hat{V}^{(K)}(w)$ as a measure of sampling efficiency, where $w_i = \frac{X_i}{\sum_{i=1}^{N_K} X_i} = \frac{X_i}{N_k \widehat{P(D|M)}}$.

$$\hat{V}^{(K)}(w) = N_k \sum_{i=1}^{N_k} (w_i - \frac{1}{N_k})^2 = N_k \sum_{i=1}^{N_k} w_i^2 - 1 = \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i^2) / \widehat{P(D|M)}^2 - 1 \quad (6)$$

We have shown that $\widehat{P(D|M)} \rightarrow_p P(D|M)$, and also $E(X^2) < \infty$ which leads to $\frac{1}{N_k} \sum_{i=1}^{N_k} (X_i^2) \rightarrow_p E(X^2)$ by the law of large numbers. Therefore, by Slutsky's theorem, $\frac{1}{N_k} \sum_{i=1}^{N_k} (X_i^2) / \widehat{P(D|M)}^2 \rightarrow_p E(X^2) / P(D|M)^2$ and equivalently $\hat{V}^{(K)}(w) \rightarrow_p V^{(K)} / P(D|M)^2$. As a result, the decreasing $\hat{V}^{(K)}(w)$ implies the decreasing asymptotic variance of the estimated integrated likelihood $V^{(K)}$ through the IMIS iterations.

Note that the finite variance of the IMIS estimator is not shared by other importance sampling estimators in general. It arises because the IMIS importance sampling distribution is a mixture of the prior and other distributions. The inclusion of the prior makes IMIS a defensive mixture method in the sense of Hesterberg (1995), and yields this desirable behavior. When the importance sampling distribution is the posterior itself, the importance sampling estimator of the integrated likelihood is the harmonic mean estimator of Newton and Raftery (1994), which is simulation-consistent but whose reciprocal has infinite variance except in special cases (Raftery et al. 2007). This is at the root of the difficulty in estimating the integrated likelihood directly from posterior simulation.

The Monte Carlo estimate of the variance of $\widehat{P(D|M)}$ is given by

$$\widehat{\text{Var}}(\widehat{P(D|M)}) = \frac{1}{N_K} \widehat{\text{Var}}_{\theta} \left(\frac{p(D|\theta)\pi(\theta)}{q(\theta)} \right). \quad (7)$$

We can approximate the standard error of $\log(\widehat{P(D|M)})$ by

$$SE(\log(\widehat{P(D|M)})) \approx \frac{SE(\widehat{P(D|M)})}{\widehat{P(D|M)}}. \quad (8)$$

4 Simulated Examples

4.1 Methods for Comparison

We consider two simulated examples, to explore the behavior of IMIS in situations with features similar to those of the EPP problem. The EPP posterior is typically ridge-like and multimodal, so we show one example of simulating from a distribution concentrated around a thin curved manifold, and another of simulating from a bimodal distribution.

We compare IMIS and IMIS-opt with SIR, and also with generic random-walk Metropolis algorithms, which have become perhaps the most common way of simulating from posterior distributions. We include results for three different publicly available algorithms of this kind: a version that is close to theoretically optimal for simulating a multivariate normal distribution, a version that does not use optimization, and a version that was programmed in WinBUGS.

The first one is the random walk Metropolis algorithm implemented by the R function `MCMCmetrop1R` in the `MCMCpack` package (Martin and Quinn 2007). This is close to the theoretically optimal algorithm for simulating from a multivariate normal distribution using a single-block random walk Metropolis algorithm with a multivariate normal proposal distribution (Gelman et al. 1996; Neal and Roberts 2006). It involves numerically optimizing the posterior density to find the posterior mode and evaluating the inverse Hessian at that point. The variance of the proposal distribution is then $2.38/d$ times the estimated inverse Hessian at the posterior mode, where d is the dimension of the target distribution. The theoretical acceptance rate for a multivariate normal target ranges from 0.44 for $d = 1$ to 0.23 as $d \rightarrow \infty$. We ran five chains, all starting from the mode found by the optimizer.

The second generic Metropolis algorithm uses the R function `metrop` in the `mcmc` package (Geyer 2005), with a multivariate normal proposal distribution with variance proportional to the identity matrix. We simulated five starting values from the prior, and ran five chains, tuning the proposal variance so that the acceptance rate was close to 0.23. Finally, we also ran the WinBUGS program (Lunn, Thomas, Best, and Spiegelhalter 2000) using the “zero trick” to specify a new distribution (Cowles 2004), again running five chains from starting values simulated from the prior. This random walk Metropolis algorithm is used in WinBUGS for nonconjugate continuous full conditionals with an unrestricted range.

As our primary criterion for comparing all six methods, we use Efficiency, defined as the Effective Sample Size (ESS) divided by the total number of function evaluations. For the MCMC methods, we used the estimated effective sample size using an autoregressive estimate of the spectrum at zero, as implemented in the R function `effectiveSize` of the `coda` package (Plummer et al. 2006). To estimate the ESS for the MCMC methods, we used the diagnostic of Gelman and Rubin (1992) to assess convergence using the five chains, and then ran one of the chains for as many iterations as used by IMIS, discarding the iterations before the five chains had converged as burn-in.

4.2 Ridge-like Simulated Example

The posterior distributions from mechanistic deterministic models are often ridge-like, that is, concentrated around thin curved manifolds. This is because they are often underdetermined

by the data, with only functions of the parameters being well estimated. We simulate a situation of this kind, with a setup first used by Bates (2001).

The deterministic model which maps the inputs $\theta = (\theta_1, \dots, \theta_6)$ to the outputs $\phi = (\phi_1, \dots, \phi_4)$ is given by

$$\begin{cases} \phi_1 &= \prod_{i=1}^6 \theta_i, \\ \phi_2 &= \theta_2 \theta_4, \\ \phi_3 &= \theta_1 / \theta_5, \\ \phi_4 &= \theta_3 \theta_6. \end{cases} \quad (9)$$

The prior distributions on the inputs are independent and normally distributed with means $\{6.0, 0.5, 5.5, 0.15, 3.0, 0.6\}$ and standard deviations $\{1.3, 0.14, 0.289, 0.029, 0.04, 0.1\}$. In addition, there are independent normal likelihoods on the outputs, with means $\{7.0, 0.0525, 2.0, 4.0\}$ and standard deviations $\{0.5, 0.00144, 0.01, 0.01\}$ respectively.

IMIS converged after 35,400 function evaluations, and IMIS-opt converged after 16,800 evaluations. The efficiencies of the different methods are shown in Table 1. SIR was clearly much less efficient than the other methods. IMIS was far more efficient, and IMIS-opt was about 50% more efficient again, although there were not multiple modes in this case. The MCMC methods were more efficient than SIR, and the `MCMCMetrop1R` method, with its optimality properties, greatly outperformed the `metrop` and `WinBUGS` methods, by a factor of at least 7. However, IMIS and IMIS-opt were much more efficient than the best of the MCMC methods in this case, by a factor of at least 15.

Table 1: Efficiencies of Different Methods for the Simulated Examples and EPP for Zimbabwe. The value 1.0 corresponds to direct independent sampling from the posterior distribution.

Method	Ridge-like	Bimodal $d = 4$	Bimodal $d = 20$	EPP
SIR	2×10^{-5}	.0002	10^{-6}	8×10^{-6}
IMIS	.0675	.2063	.0073	.1361
IMIS-opt	.1040	.4047	.1758	.1553
MCMCMetrop1R	.0045	.0437	.0143	.0031
metrop	.0006	.0047	.0004	NA
WinBUGS	.0005	.0223	.0049	NA

4.3 Bimodal Example

Our second example is bimodal, and the target distribution was a product of a likelihood and a prior. The likelihood was a mixture of two multivariate normal distributions:

$$\frac{1}{2}N_d((0, 0, \dots, 0), AR(-0.95)) + \frac{1}{2}N_d((9, 9, \dots, 9), AR(0.95)),$$

where $\text{AR}(\rho)$ represents a first-order autoregressive correlation matrix with correlation parameter ρ . The prior was $U([-3, 12]^d)$.

We consider two cases: $d = 4$, the same dimension as the EPP model, and $d = 20$, corresponding to larger deterministic models that have been analyzed using Bayesian melding (Green et al 1999, 2000; MacFarlane et al 2000; Radtke et al 2001, 2002; Hong et al 2005). In the four-dimensional case, the high density region for each mode (defined as the sphere centered at the mode with radius three standard deviations) is 4×10^{-6} times the volume of the prior distribution’s support, while in the 20-dimensional case the ratio is about 10^{-35} . This experiment was modeled on an experiment of Warnes (2001).

When $d = 4$, IMIS converged after 13,600 function evaluations and IMIS-opt after 8,600 evaluations. When $d = 20$, the numbers were much larger: 150,000 for IMIS and 41,000 for IMIS-opt. However, when $d = 20$, IMIS had not sampled the second mode, while IMIS-opt was the only one of the methods we consider here that sampled both modes in this challenging example.

Contour and density plots of the simulated values are shown in Figure 1, for the same number of function evaluations for all methods (13,600 for $d = 4$ and 150,000 for $d = 20$). SIR did not perform well: it sampled both modes for $d = 4$, but only sparsely, and for $d = 20$ it missed both modes. IMIS did an excellent job for $d = 4$, but for $d = 20$ only partly explored one mode and missed the other. IMIS-opt did a very good job in both cases.

For $d = 4$, `MCMCMetrop1R` performed best of the three MCMC methods we used. For $d = 20$, however, `MCMCMetrop1R` did not work, but we implemented a similar method using the `metrop` function, and that is shown in Figure 1. For $d = 4$, the MCMC method found both modes, but did not weight them correctly. For $d = 20$, the MCMC method found only one of the modes.

These impressions are confirmed by the efficiencies in Table 1. For both $d = 4$ and $d = 20$, IMIS-opt performed best by far; IMIS was a respectable second for $d = 4$, but far behind for $d = 20$. IMIS-opt was about 10 times more efficient than the best of the three generic MCMC methods in both cases.

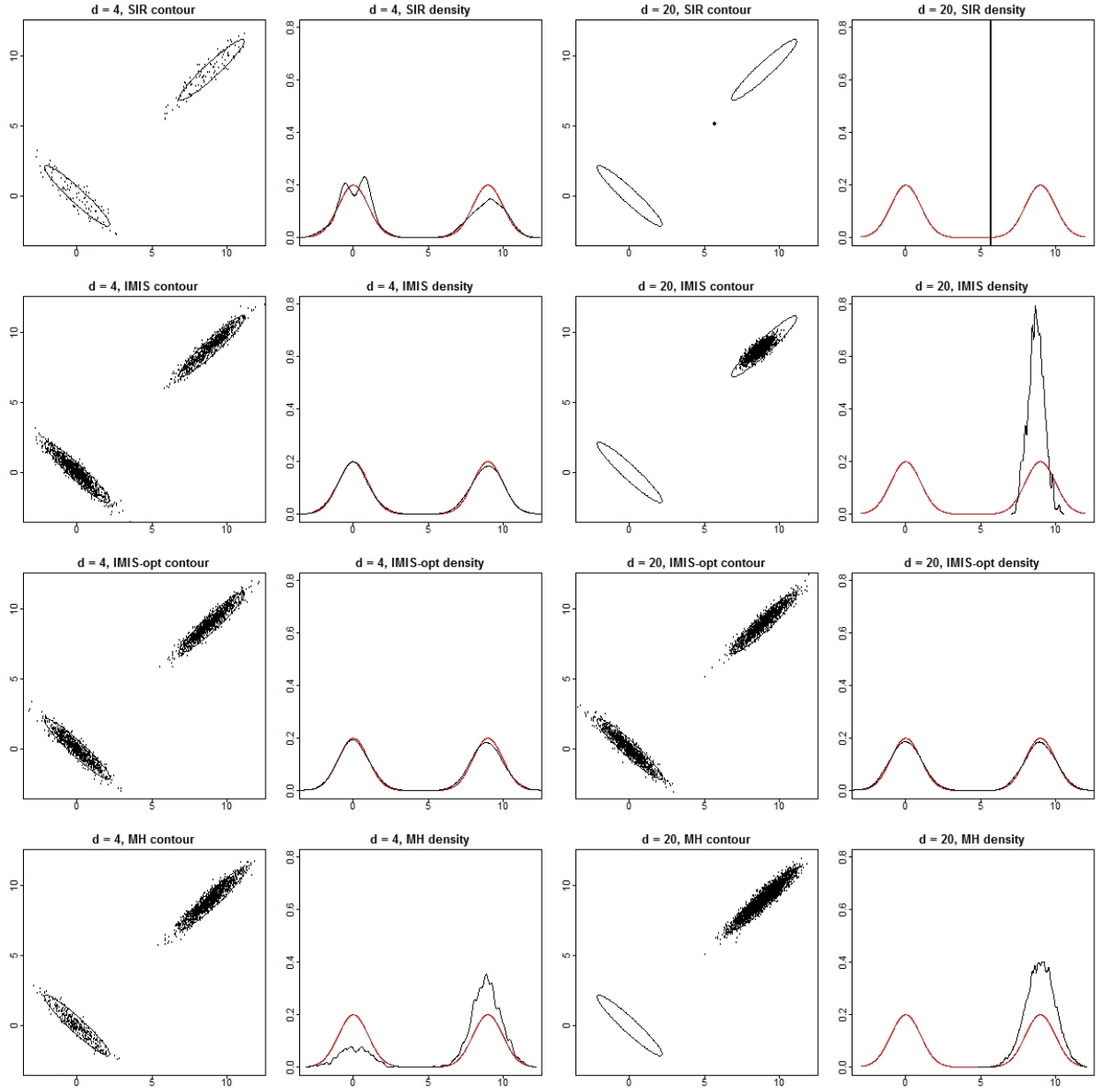


Figure 1: Results for the Bimodal Simulated Example, for SIR, IMIS, IMIS-opt, and the best performing of the three generic Metropolis algorithms considered. The results are for dimensions 4 and 20, and are projected onto two and one dimensions. The true elliptical contours of the densities are shown in black. The true one-dimensional densities are shown in red, and a kernel density estimate from the posterior simulation output is shown in black in each case.

5 Application to EPP for Zimbabwe

5.1 Estimation and Projection Results

We now apply IMIS to Bayesian melding for the EPP model for the urban parts of Zimbabwe. We use prevalence data from 18 antenatal clinics from 1989 to 2004. The observed prevalences are shown by the dots in Figure 2, with different colors corresponding to different clinics.

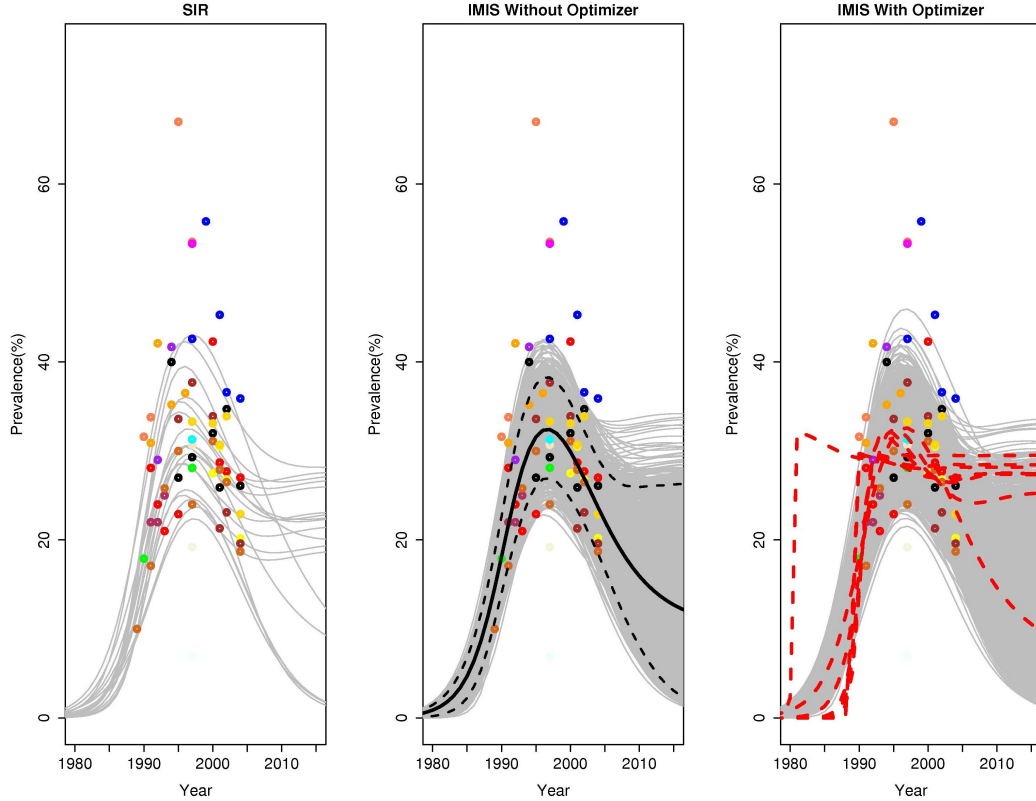


Figure 2: EPP model for 2004 Zimbabwe Urban Data: Uncertainty assessment using SIR (left), IMIS (middle), and IMIS-opt (right). The dots show the observed prevalences, with different colors corresponding to different ante-natal clinics. The gray lines are 3,000 resampled trajectories. In the middle figure, the solid black line is the posterior mean and the dashed black lines are 2.5% and 97.5% posterior quantiles. In the right figure, the red lines are 10 trajectories corresponding to the local optima.

We implemented Bayesian melding using the random effects likelihood function described by Alkema et al. (2007) and the default prior distributions agreed by the UNAIDS Reference Group, namely $\log(r) \sim U[\log(0.5), \log(150)]$, $t_0 \sim \text{Discrete Uniform}[1970, 1971, \dots, 1990]$, $f_0 \sim U[0, 1]$, and $\phi \sim \text{Logistic}(100, 50)$. The starting time of the epidemic, t_0 , is treated

as an integer by the EPP model because the data typically come in yearly increments, but our implementation of IMIS is designed for continuous parameters. We deal with this by treating t_0 as continuous in the importance sampling distribution, with a piecewise constant likelihood.

Figure 2(a) shows the results of using SIR. The $J = 3,000$ resamples yielded only 19 unique trajectories, each one corresponding to a different set of parameter values. The largest weight was 0.81, so over 2,400 of the 3,000 resampled trajectories were actually the same.

Figures 2(b) and (c) show the IMIS and IMIS-opt results. Clearly they cover the posterior distribution much more fully and smoothly. In IMIS-opt, the red dashed lines correspond to the local optima found; most of these had little support in the posterior.

Table 2 shows performance measures for the various methods used. Among the three MCMC methods we show results only for `MCMCMetrop1R`, as it performed better than the other two in this case. With the IMIS importance sampling distribution, the 3,000 resamples generated 1,929 unique points, which is close to the expected number of unique points when the importance sampling distribution is equal to the target (1,896). IMIS-opt was slightly more efficient than IMIS in this case, but not much more so, reflecting the fact that the secondary modes were much smaller than the largest one. The best of the MCMC methods was far less efficient than IMIS, by a factor of more than 40. In fact, when we ran the five chains from different starting points, they had not converged within $20,400/5 = 4,088$ iterations, according to the Gelman-Rubin diagnostic.

Table 2: Efficiencies of Different Methods for the EPP applied to Zimbabwe

Method	total inputs	variance	max weight	entropy	unique points	ESS
SIR	200,000	13,317	0.81	0.06	24	1.5
IMIS	20,400	6.5	0.0010	0.82	1929	2776
IMIS-opt	17,447	5.5	0.0012	0.84	1915	2709
MCMCMetrop1R()	20,447					62.7

Figure 3 shows pairs plots of the samples from the posterior distribution found by the four methods. The poor coverage of SIR is apparent. It is also clear that MCMC (in this case a single chain starting from the SIR mean and run for over 168,000 iterations) missed significant parts of the posterior distribution.

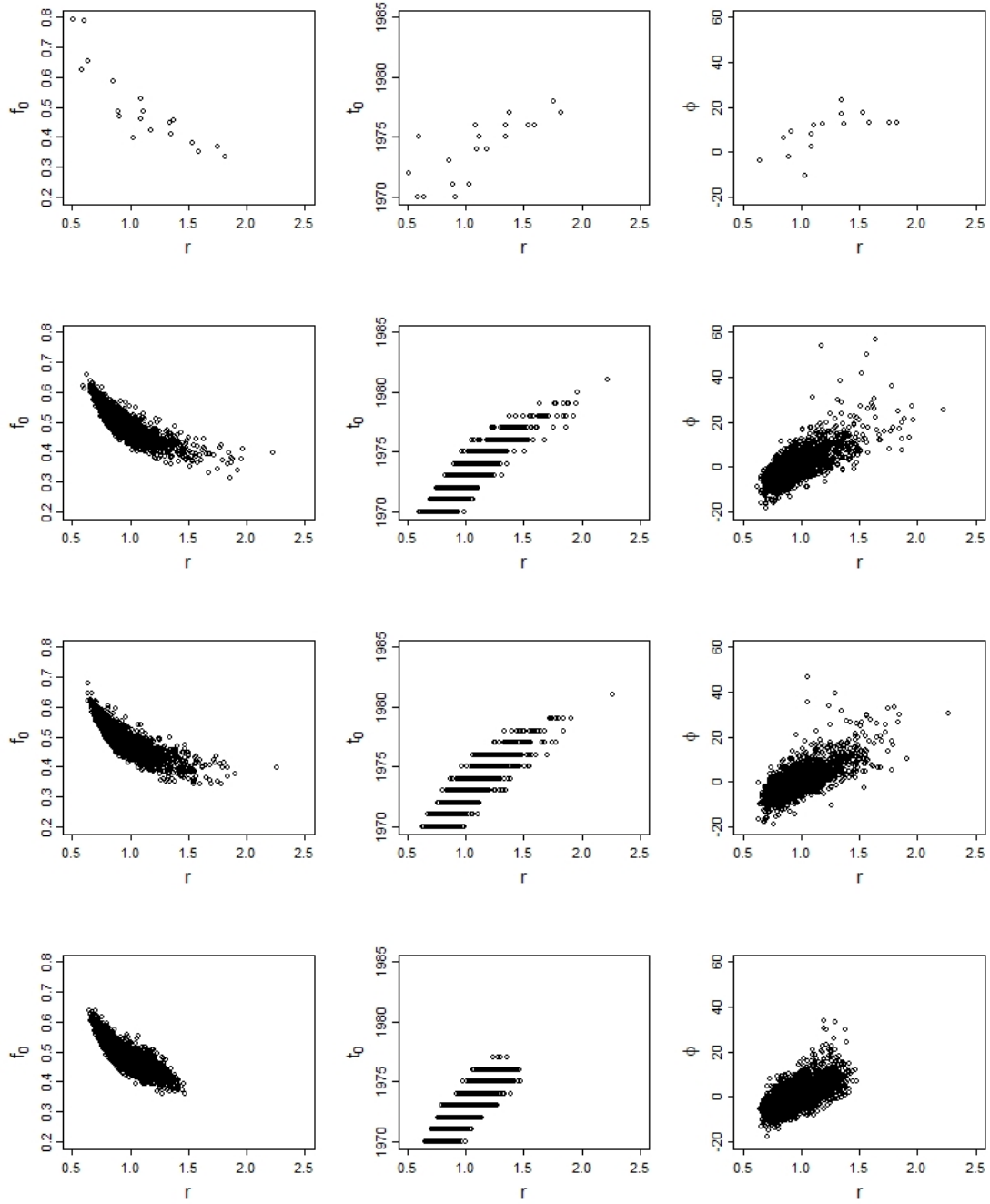


Figure 3: EPP model for Zimbabwe urban data: Comparison of the marginal posterior distribution obtained using SIR with 200,000 inputs (top row), IMIS without optimizers and with 20,400 inputs (second row), IMIS with 10 optimizers and 17,447 inputs (third row) and generic Metropolis random walk starting from the SIR mean and with 168,074 iterations (bottom row).

5.2 Model Comparison and Model Averaging

One question of great interest in the context of HIV/AIDS is whether the pattern of the epidemic has changed, for example due to a change in behavior. One way of representing this is the so-called “R-jump model” proposed by Brown et al. (2008), which allows for a one-time change in the force of infection, r , at time τ . The R-jump model specifies that r depends on time t , denoted by $r(t)$, according to $r(t) = r_1$ if $t_0 \leq t \leq \tau$, and $r(t) = r_2$ if $t > \tau$. We use a joint prior distribution for (t_0, τ) specified by $t_0 \sim U[1970, 1990]$, and $\tau|t_0 \sim U[t_0, 2004]$. The priors for r_1 and r_2 are taken to be independent, with $\log(r_j) \sim U[\log(0.5), \log(150)]$.

This question can be addressed by computing the Bayes factor for the R-jump model, M_2 , against the standard EPP model, M_1 . The estimated values and 95% Bayesian confidence intervals of the log integrated likelihoods after each stage of IMIS are shown in Figure 4. The values increase and then stabilize, which is a result of the “pseudo-bias” that can afflict importance sampling (Ventura 2002). This underlines the importance of running IMIS to convergence.

The estimated log Bayes factor was $\log(B_{21}) = 75.73 - 75.04 = 0.69$. The standard error of the log integrated likelihood was 0.018 for the standard EPP model and 0.020 for the R-jump model, and so the standard error of the log Bayes factor was $\sqrt{0.018^2 + 0.020^2} = 0.027$. Thus the estimated Bayes factor was just 2.0, “evidence worth no more than a bare mention,” in the words of Jeffreys (1939). Thus there is slight evidence for the R-jump model against the standard EPP model for Zimbabwe, but nothing conclusive.

When multiple competing models are considered plausible, Bayesian model averaging can be used to obtain combined inference about the trajectory of the disease prevalence, taking account of the model uncertainty. We illustrate this here by combining the results of the standard EPP and R-jump models. We emphasize that this is purely illustrative, because the scientific plausibility of the R-jump model would have to be more fully accepted before this would be done, and also because there may be other plausible models to be considered. We took $3000 \times \frac{1}{B_{21}+1} = 997$ trajectories from the standard EPP posterior distribution; and $3000 \times \frac{B_{21}}{B_{21}+1} = 2003$ trajectories from the R-jump model. Figure 4(b) shows 50 trajectories randomly selected from this combined sample. In this case the projections for 2015 did not differ greatly between the models, so model uncertainty was not a major component of uncertainty about this quantity of interest.

6 Discussion

We have proposed a new computational method for implementing Bayesian melding for the EPP model of the HIV/AIDS epidemic used by UNAIDS. This is incremental mixture

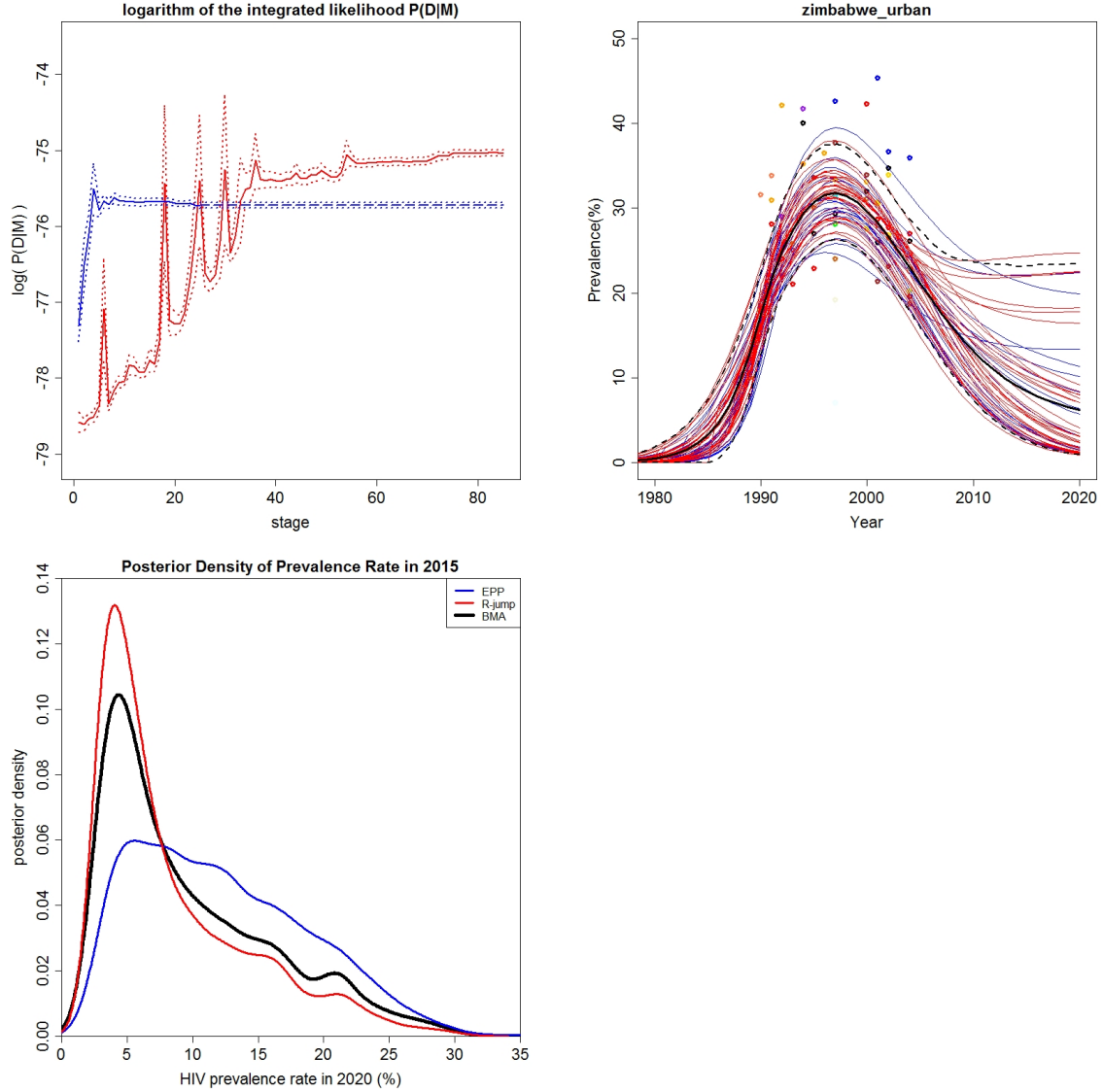


Figure 4: Model selection and model averaging for Zimbabwe urban data: (a) Estimated log integrated likelihood (solid lines) plus or minus twice its approximate standard error (dashed lines). The standard EPP model is in blue and the R-jump model in red. IMIS required 26 stages for the standard EPP model and 86 for the R-jump model. The lines for the standard EPP model are extended horizontally to allow easier visual comparison with the R-jump model. (b) A sample of 50 trajectories from the combined BMA posterior distribution for 2006 Zimbabwe urban data. The trajectories from the standard EPP model are in blue and those from the R-jump model are in red. (c) Kernel density estimate of the posterior predictive distribution of HIV prevalence in 2015. The standard EPP model has median 11.3% and 95% confidence interval (3.0%, 26.0%), the R-jump model has median 6.2% and 95% interval (2.6%, 23.2%), and the BMA combined posterior has median 7.6% and 95% interval (2.7%, 24.9%).

importance sampling (IMIS), a form of adaptive importance sampling. IMIS outperformed both simple SIR and three publicly available variants of generic MCMC in simulations and real data relevant to the problem at hand. IMIS also yields a simple estimate of the integrated likelihood required for Bayesian model comparison and model averaging; this is hard to obtain in general using MCMC. Moreover, the importance sampling framework allows easy assessment of the Monte Carlo error and facilitates parallel implementation.

These results are relevant to problems like the present one, with posterior distributions of moderate dimension (up to 30 or so) that tend to be concentrated close to ridges, or thin curved manifolds, and also can be multimodal. Many problems involving deterministic or mechanistic models have these characteristics, so this method may be more widely useful. It is unlikely to be as useful, however, for some classes of models that are routinely and successfully handled by MCMC, such as Bayesian hierarchical models with hundreds or thousands of parameters.

IMIS is an adaptive important sampling method, and other adaptive importance methods have been proposed, including global adaptive importance sampling (Oh and Berger 1993; West 1992; West 1993; Zhang 1996), local adaptive importance sampling (Givens and Raftery 1996), defensive mixture importance sampling (Hesterberg 1995; Raghavan and Cox 1998; Owen and Zhou 2000), and population Monte Carlo (PMC) (Iba 2000; Cappé et al. 2004; Douc et al. 2007; Cappé et al. 2008). IMIS more directly attacks the fundamental problem of importance sampling: important regions of parameter space that are unrepresented or sparsely represented in the original sample from the important sampling function. The presence of these is identified by large importance weights, and IMIS directly adds mass to the importance sampling function where there are large weights. In this way it learns an effective importance sampling function. Also, the optimization step of IMIS-opt makes it easier to reach modes not initially sampled.

There are also similarities between IMIS and sequential Monte Carlo (SMC), since they are iterative importance sampling schemes. SMC methods were originally introduced to deal with sequential inference problems in statistics; see Liu (2002) and Moral, Doucet, and Jasra (2007) for a review. It was soon pointed out that SMC can also be used for static inference, and several methods for doing this have been proposed, including annealed importance sampling (Neal 2001), resample-move (Gilks and Berzuini 2001), sequential particle filtering (Chopin 2002) and the dynamically weighted importance sampling method (Liang 2002); see Jasra et al. (2007) for a review. A problem for SMC samplers is that the particles may only represent a single mode of the target (Jasra et al. 2008).

The main differences between IMIS and the other adaptive importance sampling methods mentioned (including PMC and SMC methods) are that IMIS updates the importance

sampling function in a different and more direct way, IMIS uses all the values simulated and not just those simulated at the last stage, and IMIS (in the IMIS-opt version) uses optimization to reach local modes not reached by the sample from the initial importance sampling function. The IMIS idea was introduced in a different context by Steele et al. (2006), and the present work extends this to the problem of simulating from a continuous posterior distribution. Cappé et al. (2008) subsequently proposed a PMC method that is an adaptive importance sampling method with an adaptively estimated mixture of multivariate normals and the prior as importance sampling function. IMIS differs from their approach in the way the importance sampling function is updated, its use of all the samples simulated and the optimization step.

The application of IMIS to the EPP model is implemented as part of the UNAIDS EPP software, publicly and freely available from <http://www.epidem.org>.

Acknowledgements

This research was supported by NICHD grant HD054511. The authors are grateful to Leon-tine Alkema, Tim Brown, Peter Ghys and Eleanor Gouws for helpful discussions and for sharing data.

References

- Alkema, L., A. E. Raftery, and T. Brown (2008). Bayesian melding for estimating uncertainty in national HIV prevalence estimates. *Sexually Transmitted Infections* 84, i11–i16.
- Alkema, L., A. E. Raftery, and S. J. Clark (2007). Probabilistic projections of HIV prevalence using Bayesian melding. *Annals of Applied Statistics* 1, 229–248.
- Bates, S. (2001). *Bayesian inference for deterministic simulation models for environmental assessment*. Ph. D. thesis, University of Washington.
- Brown, T., J. A. Salomon, L. Alkema, A. E. Raftery, and E. Gouws (2008). Progress and challenges in modelling country-level HIV/AIDS epidemics: the UNAIDS Estimation and Projection Package 2007. *Sexually Transmitted Infections* 84, i5–i10.
- Byrd, R. H., P. Lu, J. Nocedal, and C. Y. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16(6), 1190–1208.

- Cappé, O., R. Douc, A. Guillin, J.-M. Marin, and C. Robert (2008, December). Adaptive importance sampling in general mixture classes. *Statistics and Computing* 18(4), 447–459.
- Cappé, O., A. Guillin, J.-M. Marin, C. P. Robert, and C. P. Robertyz (2004). Population Monte Carlo. *Journal of Computational and Graphical Statistics* 13, 907–929.
- Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika* 89(3), 539–551.
- Clyde, M. and E. I. George (2004). Model uncertainty. *Statistical Science* 19, 81–94.
- Cowles, M. K. (2004, November). Review of winbugs 1.4. *The American Statistician* 58, 330–336.
- Douc, R., A. Guillin, J. M. Marin, and C. P. Robert (2007). Convergence of adaptive mixtures of importance sampling schemes. *Annals of Statistics* 35, 420.
- Gelman, A., G. Roberts, and W. Gilks (1996). Efficient Metropolis jumping rules. *Bayesian Statistics* 5, 599–607.
- Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–511.
- Geyer, C. (2005). *MCMC R Package*. At <http://www.stat.umn.edu/geyer/mcmc/>.
- Ghys, P. D., T. Brown, N. C. Grassly, G. Garnett, K. A. Stanecki, J. Stover, and N. Walker (2004). The UNAIDS estimation and projection package: A software package to estimate and project national HIV epidemics. *Sexual Transmitted Infections* 80, i5–i9.
- Ghys, P. D., N. Walker, W. McFarland, R. Miller, and G. P. Garnett (2008). Improved data, methods and tools for the 2007 HIV and AIDS estimates and projections. *Sexually Transmitted Infections* 84, i1–i4.
- Gilks, W. R. and C. Berzuini (2001). Following a moving target Monte Carlo inference for dynamic Bayesian models. *Journal of The Royal Statistical Society Series B* 63(1), 127–146.
- Givens, G. H. and A. E. Raftery (1996). Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association* 91(433), 132–141.
- Green, E. J., D. W. MacFarlane, and H. T. Valentine (2000). Bayesian synthesis for quantifying uncertainty in predictions from process models. *Tree Physiology* 20, 415–419.

- Green, E. J., D. W. MacFarlane, H. T. Valentine, and W. E. Strawderman (1999). Assessing uncertainty in a stand growth model by Bayesian synthesis. *Forest Science* 45, 528–538.
- Han, C. and B. P. Carlin (2001). Markov chain Monte Carlo methods for computing Bayes factors. *Journal of the American Statistical Association* 96, 1122–1132.
- Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37(2), 185–194.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial (with discussion). *Statistical Science* 14, 382–417.
- Hong, B., R. Strawderman, D. P. Swaney, and D. Weinstein (2005). Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. *Water Resources Research* 41, W03007.
- Iba, Y. (2000). Population Monte Carlo algorithms. *Transactions of the Japanese Society for Artificial Intelligence* 16, 279.
- Jasra, A., A. Doucet, D. A. Stephens, and C. C. Holmes (2008). Interacting sequential Monte Carlo samplers for trans-dimensional simulation. *Computational Statistics and Data Analysis* 52(4), 1765–1791.
- Jasra, A., D. A. Stephens, and C. C. Holmes (2007). On population-based simulation for static inference. *Statistics and Computing* 17(3), 263–279.
- Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press.
- Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kong, A., J. Liu, and W. Wong (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association* 89, 278–288.
- Leamer, E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: John Wiley.
- Liang, F. (2002). Dynamically weighted importance sampling in Monte Carlo computation. *Journal of the American Statistical Association* 97, 807–821.
- Liu, J. S. (2002, October). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.
- Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). Winbugs – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 10(4), 325–337.

- MacFarlane, D. W., E. J. Green, and H. T. Valentine (2000). Incorporating uncertainty into the parameters of a forest process model. *Ecological Modelling* 134, 27–40.
- Madigan, D. and A. E. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association* 89, 1335–1346.
- Martin, A. and K. Quinn (2007). *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. At <http://mcmcpack.wustl.edu>.
- Moral, P. D., A. Doucet, and A. Jasra (2007). Sequential Monte Carlo for Bayesian computation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics* 8, pp. 115–148. Oxford University Press.
- Neal, P. and G. Roberts (2006). Optimal scaling for partially updating mcmc algorithms. *Annals of Applied Probability* 16(2), 475–515.
- Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing* 11, 125–139.
- Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B* 56, 3–48.
- Oh, M. and J. Berger (1993). Integration of multimodal functions by Monte Carlo importance sampling. *Journal of the American Statistical Association* 88, 450–456.
- Owen, A. and Y. Zhou (2000). Safe and effective importance sampling. *Journal of the American Statistical Association* 95, 135–143.
- Plummer, M., N. Best, K. Cowles, and K. Vines (2006, March). CODA: Convergence diagnosis and output analysis for MCMC. *R News* 6(1), 7–11.
- Poole, D. and A. E. Raftery (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association* 95(452), 1244–1255.
- Radtke, P. J., T. E. Burk, and P. V. Bolstad (2001). Estimates of the distributions of forest ecosystem model inputs for deciduous forests of eastern North America. *Tree Physiology* 21, 505–512.
- Radtke, P. J., T. E. Burk, and P. V. Bolstad (2002). Bayesian melding of a forest ecosystem model with correlated inputs. *Forest Science* 48, 701–711.

- Raftery, A. E., G. H. Givens, and J. E. Zeh (1995). Inference from a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association* 90(430), 402–416.
- Raftery, A. E., M. A. Newton, J. M. Satagopan, and P. Krivitsky (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity (with discussion). In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 8*, pp. 1–45. Oxford University Press.
- Raghavan, N. and D. Cox (1998). Adaptive mixture importance sampling. *Journal of Statistical Computation and Simulation* 60, 237–259.
- Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association* 82(398), 543–546.
- Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 395–402. Oxford University Press.
- Steele, R. J., A. E. Raftery, and M. J. Emond (2006). Computing normalizing constants for finite mixture models via incremental mixture importance sampling (IMIS). *Journal of Computational and Graphical Statistics* 15, 712–734(23).
- UNAIDS (2007). *2007 AIDS Epidemic Update*. Geneva: UNAIDS. Available from <http://www.unaids.org/en/Publications>.
- Ventura, V. (2002). Non-parametric bootstrap recycling. *Statistics and Computing* 12, 261–271.
- Warnes, G. (2001). The normal kernel coupler: An adaptive Markov chain Monte Carlo method for efficiently sampling from multi-modal distributions. Technical report, Department of Statistics, University of Washington.
- West, M. (1992). Modelling with mixtures. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 503–524. Oxford University Press.
- West, M. (1993). Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society. Series B* 55, 409–422.
- Zhang, P. (1996). Nonparametric importance sampling. *Journal of the American Statistical Association* 91, 1245–1253.