Microsimulation Model Calibration using Incremental Mixture Approximate Bayesian Computation

Carolyn Rutter
The RAND Corporation
crutter@rand.org

Maria DeYoreo
The RAND Corporation
mdeyoreo@rand.org

Jonathan Ozik
Argonne National Laboratory &
The University of Chicago
jozik@anl.gov

Nicholson Collier
Argonne National Laboratory &
The University of Chicago
ncollier@anl.gov

Abstract

Microsimulation models (MSMs) contain unknown parameters that are selected through calibration, a process of choosing parameters so that the model accurately predicts specific targets. Targets are typically functions of observed data, such as summary statistics. A well-calibrated model can reproduce a wide range of targets. Calibration is often focused on identifying a single best-fitting parameter vector. In contrast, Bayesian calibration methods focus on estimating the joint posterior distribution of model parameters given calibration targets, to capture uncertainty in the model parameters. The estimated posterior distribution can be used to generate model predictions accompanied by estimates of uncertainty. Calibration of MSMs is challenging for multiple reasons. First, it involves searching a high dimensional parameter space to predict many targets accurately, using relatively sparse data. Additionally, it can be difficult to estimate the likelihood accurately, because the distribution of the calibration data is parameterized by an unknown function of the model parameters. In this paper, we develop a new simulation-based method for calibration of MSMs which we refer to as Incremental Mixture Approximate Bayesian Computation (IMABC). We use IMABC to calibrate a MSM for the natural history of colorectal cancer, obtaining approximate posterior distributions for model parameters which can be used to make predictions about hypothetical interventions and future trends in population disease outcomes.

1 Introduction

Microsimulation models (MSMs) are useful tools for estimating population-level effects of medical interventions on health outcomes by combining results from randomized controlled trials, observational studies, and expert opinion. MSMs are characterized by simulation of individual event histories for an idealized population of interest. These individual event histories catalog landmarks in the disease process, such as the development of an incident cancer or a first myocardial infarction. Simulation of event histories is based on mathematical models for key components of the disease process. In general, these processes are not directly observable, though their outcomes are. For example, the process of developing colorectal cancer (CRC) cannot be observed, but the prevalence of both precancerous lesions (adenomas) and preclinical CRC can be estimated from screening

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

trials. Model calibration involves selecting parameter values that result in model predictions that are consistent with observed data and expected findings. Once parameters are selected, MSMs can be used to make predictions about interventions and future trends in population disease outcomes.

The calibration process can be difficult for a variety of reasons. It involves searching a high dimensional parameter space to predict many targets accurately, using relatively sparse data. Many approaches to calibration have been proposed. The simplest calibration method involves perturbing parameters one at a time using an undirected search and evaluating the goodness of fit to calibration data, however this is only feasible when calibrating a few parameters. Directed searches, such as the Nelder-Mead algorithm (Nelder and Mead, 1965), provide a derivative free hill-climb that generally results in a single estimate for each parameter. Kong et al. (2009) used automated parameter search algorithms from engineering (simulated annealing and a genetic algorithm) to find the optimal parameter set that yields the best fit. This involved defining a goodness of fit score determined by a weighted sum of calibration-specific terms, with user defined weights.

Markov Chain Monte Carlo (MCMC) algorithms are the standard tool for simulating from a posterior distribution of model parameters, conditional on observed data. However, due to the large number of targets and the complexity of most MSMs, MCMC is extremely inefficient, and can be difficult to implement. Additionally, it can be difficult and costly to estimate the likelihood, because the distribution of the calibration data is parameterized by an unknown function of the parameters. This requires the use of embedded simulations within each MCMC iteration to obtain an estimate of this function. When combined with the need to discard most draws due to poor mixing and slow convergence, MCMC can be impractical for many MSMs.

Approximate Bayesian Computation (ABC) is a suite of techniques used for simulating draws from the posterior distribution of model parameters given calibration targets, and avoids the need to estimate a likelihood, making it a reasonable alternative to MCMC when likelihoods are not available (Marin et al., 2012, provide a nice review of ABC). However, ABC is inefficient and can fail when the parameter space is high dimensional, when there are many calibration targets, or when the prior distributions are very different from the posterior distributions. McKinley et al. (2017) explore the use of some popular ABC variants in stochastic epidemiological models, and find that in order to get ABC algorithms to fit in a computationally feasible manner, trade-offs in accuracy must be made, and this often leads to approximations that are too poor to make useful inferences.

We propose a new calibration approach, Incremental Mixture ABC (IMABC), which begins with a basic rejection-sampling ABC step, like that used in the first genuine rejection-based ABC algorithm of Pritchard et al. (1999). To encourage a more efficient search of parameters that produce simulated targets matching the observed targets, we incrementally update this initial sample using a procedure that is similar to Incremental Mixture Importance Sampling (IMIS; Raftery and Bao, 2010), but does not rely on the likelihood. IMABC can be used to calibrate MSM parameters by simulating draws from the posterior distribution, and is found to be effective in estimating posterior distributions of a high dimensional parameter space, while generating predictions that are consistent with many targets.

2 Microsimulation Model and Calibration Data

2.1 CRC-SPIN Model

The **C**olo**R**ectal **C**ancer **S**imulated **P**opulaton **I**ncidence and **N**atural history model (CRC-SPIN) (Rutter et al., 2009) describes the natural history of CRC based on the adenoma-carcinoma sequence (Muto et al., 1975; Leslie et al., 2002). Four model components describe the natural history of CRC: 1) adenoma risk; 2) adenoma growth; 3) transition from adenoma to preclinical cancer; and 4) transition from preclinical to clinical cancer (sojourn time). CRC-SPIN was previously calibrated using MCMC (Rutter et al., 2009). Recent model validation revealed that while the CRC-SPIN model accurately predicted many aspects of the disease (including clinically detected cancer, cancer mortality, and the action of screening), the model predicted too few preclinical cancers, indicating that sojourn time, the time from transition between preclinical (asymptomatic) colorectal cancer and clinical (symptomatic and detected) cancer, was too short. We have revised the original CRC-SPIN model to address this problem (details of the revised CRC-SPIN model are presented elsewhere).

Here we summarize the updated CRC-SPIN model and its parameters. Adenomas are simulated using a nonhomogeneous Poisson process that depends on age, gender, and individual-level risk,

containing 7 parameters. Adenomas grow according to a Janoshek growth curve with 2 parameters, and may transition to preclinical cancer depending on their size, location, and gender of the person (8 parameters). Finally, time to clinical detection is assumed to follow a Weibull distribution with 2 calibrated parameters. In total, CRC-SPIN contains 19 parameters. These are the parameters we must estimate, and this process is referred to as calibration.

2.2 Calibration Data

The calibration targets represent the data that inform model parameters. We use two types of data to calibrate model parameters: individual-level data and adenoma-level data. Individual-level data is measured from individuals but reported in aggregate, including information from both published studies and registries. Calibration to individual-level data requires simulation of a set of individuals with characteristics that match the calibration data. Such characteristics include age and gender distributions, screening patterns, and time period in which the data was collected. Adenoma-level data is measured for each adenoma and requires simulation of a number of adenomas.

We obtained data points from multiple epidemiological and clinical studies and the Surveillance, Epidemiology, and End Results (SEER) colon and rectal cancer incidence rates in 1975-1979. The SEER data are an example of individual-level targets, which consist of incidence rates within gender and age groups. Adenoma-level targets include preclinical cancer rates in adenomas by adenoma size. In total, we have 44 calibration targets from which to learn the CRC-SPIN model parameters.

2.3 Simulating Calibration Data

To estimate the posterior distribution of model parameters, for a given parameter vector θ, we simulate summary statistics $\pi(\theta)$ that make up our calibration data. To reduce simulation error, we simulate a large number of individuals, m, to estimate $\pi(\theta)$. Simulations used to estimate $\pi(\theta)$ also depend on the age- and sex-distribution of the individuals. Therefore, we must make assumptions about the age and sex distribution of study subjects. In general, studies described their samples by reporting the percentage of men and women, and the average and standard deviation of age. Unless information was provided by sex, we assumed the same age distribution for men and women. We modeled age distributions (used for embedded simulations) using a truncated normal distribution based on the reported mean, standard deviation, and age range. We used a grid search to select a truncated normal with mean and standard deviation closest to the observed values, based on a simple distance measure.

3 Sampling from Posterior Distributions via Incremental Mixture Approximate Bayesian Computation

Approximate Bayesian computation methods are likelihood-free techniques for simulating draws from the posterior distribution of model parameters given data, or in the case of MSMs, given observed calibration targets, $O_1 \dots O_J$. In general terms, the method replaces the calculation of a likelihood with an approximation based on matching model simulations to the observed data using a set of goodness-of-fit metrics (Conlan et al., 2012). To determine whether the simulated data and the observed data are similar, one must define a distance metric, a tolerance level, and in the case of multiple targets, a statistic, or multiple statistics. In the context of MSMs, the published targets are themselves summary statistics, such as means and proportions, and we directly compare the simulated targets to the published targets. In practice, simulating θ from the prior distribution can be very inefficient, because the prior and posterior distributions are often poorly aligned. To address these inefficiencies, many versions of ABC have been developed. Two popular variants are ABC-MCMC (Marjoram et al., 2003) and ABC-SMC (Toni et al., 2009). ABC-MCMC involves an initial rejection-based ABC step, but then generates new samples of θ from a Markov kernel $q(\theta \mid \theta^{(t-1)})$ instead of the prior $\pi(\theta)$. ABC-SMC is a parallel sequential ABC algorithm, based on a sequence of simulated samples, Markov transition kernels, and importance weights. ABC-SMC is adaptive, in that it allows the tolerance level ϵ, which determines whether simulated data is close to the observed data, to vary with each sample $\theta^{(t)}$. As the tolerance level is reduced, the approximate posterior distribution should converge towards the true posterior distribution.

In general, ABC and its variants can be impractical or can fail when the parameter space is high dimensional or there are many summary statistics that the simulated data must approximate (McKinley

et al., 2017). This precludes their use in the context of MSMs, where there are usually many calibration targets. To address the problems with ABC and MCMC in the context of MSMs, while avoiding the calculation of the likelihood, we propose a new simulation based approach that we call incremental mixture approximate Bayesian computation (IMABC). IMABC is simulation-based with an initial ABC step, with later steps of the proposal distribution given by a mixture of normal distributions, with means and covariance matrices that are updated at each iteration. The algorithm shares similarities with IMIS (Steele et al., 2006; Raftery and Bao, 2010), which is an adaptive importance sampling algorithm that addresses the main shortcoming of Sampling Importance Resampling (SIR; Rubin, 1987), which is sensitivity to the proposal distribution (the prior distribution in SIR). In contrast to IMIS, IMABC does not rely on the likelihood, which can be difficult to estimate with sufficient precision for MSMs.

The IMABC algorithm begins in the same way as basic rejection-sampling ABC: A sample of N_0 points, $\theta_1, \ldots, \theta_{N_0}$, is drawn from the prior distribution of model parameters, $\pi(\theta)$. At each sampled point we simulate calibration targets, as described in the previous section. Let S_{ij} be the jth simulated target for the ith sampled point, where $i = 1, \ldots, N_0$, and $j = 1, \ldots, J$. We base the ABC tolerance levels on $(1-\alpha)\%$ confidence intervals for calibration targets, with the α-level of the test akin to an ABC tolerance-level. The ith draw is accepted if all S_{ij}, $j = 1, \ldots, J$, fall within the $(1-\alpha)\%$ confidence interval of the observed targets O_j, $j = 1, \ldots, J$. Simulating targets can be computationally intensive, so we begin by evaluating targets that are not time intensive to compute, and once a target is out of bounds, this entire point is rejected and there is no need to simulate the remaining, computationally intensive, targets. For high dimensional parameter spaces with a large number of targets, the initial α-level, α_0, may need to be very small to obtain acceptable points (e.g., $\alpha < 0.0001$). The use of precisely estimated targets, based on large sample sizes or registry data, may also require very small α-levels.

IMABC then proceeds by adding more points in the neighborhood of the "best" simulated targets (those that are closest to the observed targets) in the current sample. Points are added by sampling from a mixture of normal distributions that is centered at the m closest points with covariance matrices estimated based on points that are nearest to the centers. Like ABC-SMC, IMABC is adaptive, in that we update α_k at each iteration k. Specifically, we allow α_k to increase at each iteration, thereby requiring model predictions to be closer to the targets in order to be deemed acceptable.

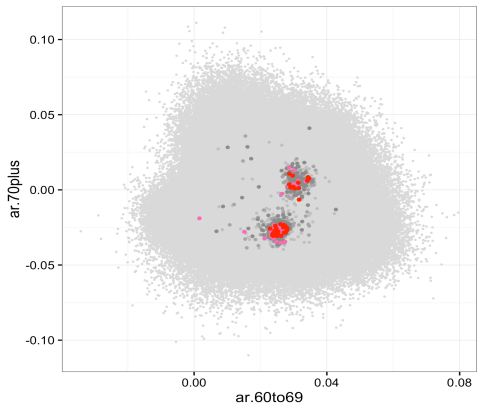

Figure 1: Preliminary results show that the model is learning the CRC-SPIN parameters, with posteriors being shifted from the priors, and in some cases uncovering multi-modal parameter distributions e.g., for changes in adenoma risk, suggesting a mode at a steady state region and a mode at a decreasing risk region.

The algorithm is complete once a sufficient number of draws have been obtained, such that the effective sample size (ESS) has reached a certain threshold and α_k is sufficiently large (so that model predictions are close to targets). To estimate the ESS, we calculate likelihood-free importance weights that account for the difference between the prior and mixture proposal distribution. Once the ESS is is large enough, we sample the desired number of draws from the posterior distribution by taking a weighted resample (with replacement) from all points identified as acceptable by IMABC, using the likelihood-free importance weights.

Acknowledgments

The authors gratefully acknowledge the National Cancer Institute and the National Institute of Health for supporting this work. C. Rutter and M. DeYoreo were supported by NCI U01 CA199335. J. Ozik and N. Collier were supported by NIH R01 GM115839. We thank Ifeanyi Edochie, Christopher Maerzluft, and Angel Martinez from RAND, who contributed to other aspects of this work.

References

Conlan, A. J. K., McKinley, T. J., Karolemeas, K., Pollock, E. B., Goodchild, A. V., Mitchell, A. P., Birch, C. P. D., Clifton-Hadley, R. S., and Wood, J. L. N. (2012), "Estimating the Hidden Burden of Bovine Tuberculosis in Great Britain," *PLOS Computational Biology*, 8(10), 1–14.
URL: *https://doi.org/10.1371/journal.pcbi.1002730*

Fearnhead, P., and Prangle, D. (2012), "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 419–474.

Kong, C. Y., McMahon, P. M., and Gazelle, G. S. (2009), "Calibration of Disease Simulation Model Using an Engineering Approach," *Value in Health*, 12(4), 521–529.
URL: *http://dx.doi.org/10.1111/j.1524-4733.2008.00484.x*

Leslie, A., Carey, F. A., Pratt, N. R., and Steele, R. J. C. (2002), "The colorectal adenoma–carcinoma sequence," *British Journal of Surgery*, 89, 845–860.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012), "Approximate Bayesian computational methods," *Statistics and Computing*, 22(6), 1167–1180.

Marjoram, P., Molitor, J., Plagnol, V., and Tavare, S. (2003), "Markov chain Monte Carlo without likelihoods," *Proceedings of the National Academy of Sciences*, 100(26), 15324–15328.

McKinley, T., Vernon, I., Andrianakis, I., McCreesh, N., Oakley, J., Nsubuga, R., Goldstein, M., and White, R. (2017), "Approximate Bayesian Computation and simulation-based inference for complex stochastic epidemic models," *Statistical science*, to appear.

Muto, T., Bussey, H. J. R., and Morson, B. C. (1975), "The Evolution of Cancer in the Colon and Rectum," *Cancer*, 36, 2251–2270.

Nelder, J. A., and Mead, R. (1965), "A simplex method for function minimization," *Computer Journal*, 7, 308–313.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999), "Population growth of human Y chromosomes: a study of Y chromosome microsatellites.," *Molecular Biology and Evolution*, 16(12), 1791–1798.
URL: + *http://dx.doi.org/10.1093/oxfordjournals.molbev.a026091*

Raftery, A., and Bao, L. (2010), "Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling," *Biometrics*, 66, 1162–1173.

Rubin, D. (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, 543–546.

Rutter, C. M., Miglioretti, D. L., and Savarino, J. E. (2009), "Bayesian Calibration of Microsimulation Models," *JASA*, 104, 1338–1350.

Steele, R., Raftery, A., and Edmond, M. (2006), "Computing Normalizing Constants for Finite Mixture Models via Incremental Mixture Importance Sampling (IMIS)," *Journal of Computational and Graphical Statistics*, 15, 712–734.

Tavare, S., Balding, D., Griffiths, R., and Donnelly, P. (1997), "Inferring Coalescence Times from DNA Sequence Data," *Genetics*, 145, 505—518.

Toni, T., Welch, D., Strelkowa, N., and Stumpf, M. (2009), "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems," *Journal of the Royal Society Interface*, 6(31), 187–202.