High performance calibration of a colorectal cancer natural history model with Incremental Mixture Importance Sampling

Jonathan Ozik, Nicholson Collier, Justin Wozniak
Argonne National Laboratory
Argonne, IL USA

Carolyn Rutter, Jessica Hwang
Rand Corporation
Santa Monica, CA USA

***Abstract*— Microsimulation models are used to predict outcomes under a range of scenarios and are used to inform health policy. Models are driven by several unknown parameters that are selected so that models match observed or expected outcomes, a process called calibration. Our work address statistical and computational challenges faced when calibrating a microsimulation model for colorectal cancer (CRC). We demonstrate the use of the Extreme-scale Model Exploration with Swift/T (EMEWS) framework on high performance computing resources in calibrating a colorectal cancer model (CRC-SPIN) with the use of Incremental Mixture Importance Sampling (IMIS). Adaptive model exploration is required to sample the high-dimensional model parameter space and EMEWS enables the integration of existing statistical and scientific codes while providing scalability to the largest scale supercomputers.**

Keywords—calibration; colorectal cancer; model exploration; screening; workflow

I. (Importance of) CRC Modeling

Microsimulation models synthesize evidence about disease processes to project population-level outcomes under different policy scenarios. Simulation models are increasingly used to inform health policy decision. For example, models have been used to inform U.S. cancer screening guidelines, including guidelines for colorectal cancer (CRC) screening [1].

II. CRC-SPIN

CRC-SPIN simulates individuals that begin in a disease-free state and may progressively transition to an adenoma state, a preclinical CRC state, and clinically detected CRC state, from which they may die from CRC [2] [3]. Individuals may die from other causes at any time. Adenoma risk varies systematically by age and sex, and varies randomly across individuals. Individuals may acquire multiple adenomas and preclinical cancers. The duration of the both the adenoma state (*dwell time,* the time to progression from adenoma to preclinical disease) and the preclinical cancer state (*sojourn time,* the time from preclinical cancer onset to clinical cancer detection) varies randomly across individuals and adenomas within individuals.

CRC-SPIN is written in R and contains 23 unknown parameters that are informed through a process of model calibration. There are 8 calibration targets: statistics from 7 studies and incidence from the SEER registry. A recent validation showed that while CRC-SPIN accurately predicted the impact of screening on CRC-mortality, the simulated sojourn time was likely too short [4], indicating the need for model recalibration.

III. Incremental Mixture Importance Sampling

Calibration of microsimulation models is difficult for two reasons. First, the number of parameters is large relative to the available data. Incremental Mixture Importance Sampling (IMIS) provides a useful tool for exploring the parameter space [5]. IMIS begins with a random search, and incrementally adds points from high likelihood regions. The second challenge is computational. The IMIS algorithm requires evaluation of the likelihood at each sampled parameter vector, but this likelihood is not closed form, and must be simulated. For each sampled point the algorithm uses CRC-SPIN to predict observed outcomes at each calibration target with a relatively large simulation sizes to minimize stochastic error.

IV. EMEWS

Our framework, Extreme-scale Model Exploration with Swift/T (EMEWS) [6] uses the general-purpose parallel scripting language Swift [7] [8] to generate highly concurrent simulation workflows. These workflows enable the integration of external model exploration (ME) algorithms to coordinate the running and evaluation of large numbers of simulations. The general-purpose nature of the programming model allows the user to supplement the workflows with additional analysis and post-processing as well.

EMEWS enables the user to plug in both ME algorithms, such as IMIS, and scientific applications, such as CRC-SPIN. The ME algorithm can be expressed in Python, R, C, C++, Fortran, Julia, Tcl, or any language supported by Swift/T. The scientific application can be implemented as an external application called through the shell, in-memory libraries accessed directly by Swift (for faster invocation), or Python, R, Julia, and JVM language applications. Thus, researchers in various fields who may not be parallel programming experts can simply incorporate existing ME algorithms and run computational experiments on their existing scientific application without explicit parallel programming. A key feature of this approach is that neither the ME algorithm nor the scientific application is modified to fit the framework. This is implemented in a reusable way by connecting the parameter generating ME algorithm and output registration methods to

interprocess communication mechanisms that allow these values to be exchanged with Swift/T. EMEWS currently provides this high-level queue-like interface with three implementations: EQ/Py, EQ/R and EQ/C (EMEWS Queues for Python, R, and C/C++).

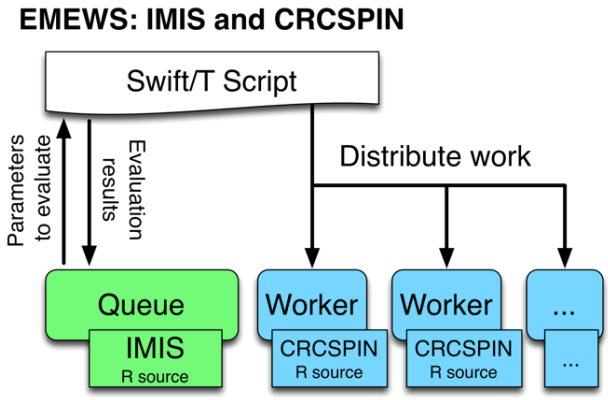

Figure 1: Overall relationship between IMIS and CRC-SPIN within the EMEWS framework.

Figure 1 illustrates the main components of the EMEWS framework as it interacts with IMIS and CRCSPIN. The main user interface is the Swift script, a high-level program. The Swift script receives parameters for evaluation from IMIS via the EMEWS EQ/R queue. Swift then distributes these parameters to parallel worker processes that run the CRC-SPIN model using these parameters. The results of these runs are gathered by the Swift script and passed back to IMIS which in turn produces additional parameters for evaluation based on the results of the CRC-SPIN runs.

V. Calibrating CRC-SPIN with EMEWS

The initial CRCSPIN calibration was run with EMEWS on the Midway cluster at the University of Chicago Research Computing Center. An initial sample of 1000 priors was evaluated, with 2 subsequent samplings of 500 each. Figure 2 shows initial results from this proof-of principle exercise. After a first step, a random draw from the parameter space, the algorithm determined the point with the most likely point (represented in light pink). The second step adds to the sample, using a multivariate normal random draw centered at this most likely point and then determines the next center point to fill in (pictured in red). In this example, points continue to be filled in nearby, though we anticipate greater movement with subsequent iterations. Ultimately, the algorithm will result in a sampling of points from the parameter space with density proportional to the likelihood. These plots also demonstrate the abilty of the algorithm to detect ridges in the likelihood, indicative of colinearity, as shown in the lower righthand plot of mean adenoma risk and change in adenoma risk from the ages of 20 to 50 years old.

VI. Conclusions and Future Work

Between now and November, we will complete the calibration effort, moving from the Midway cluster to the IBM Blue Gene/Q Mira at the Argonne Leadership Computing Facility at Argonne National Laboratory. This will enable us to increase the sampling of the parameter space, with a large initial random draw of 46,000 with subsequent steps of 1,000 additional draws. We expect the program to require approximately 3,000 steps to fully explore the parameter space. The IMIS algorithm will result in a simulated draw from the posterior distribution of model parameters which will provide new information about the uncertainty of calibrated parameters and relationships among parameters.

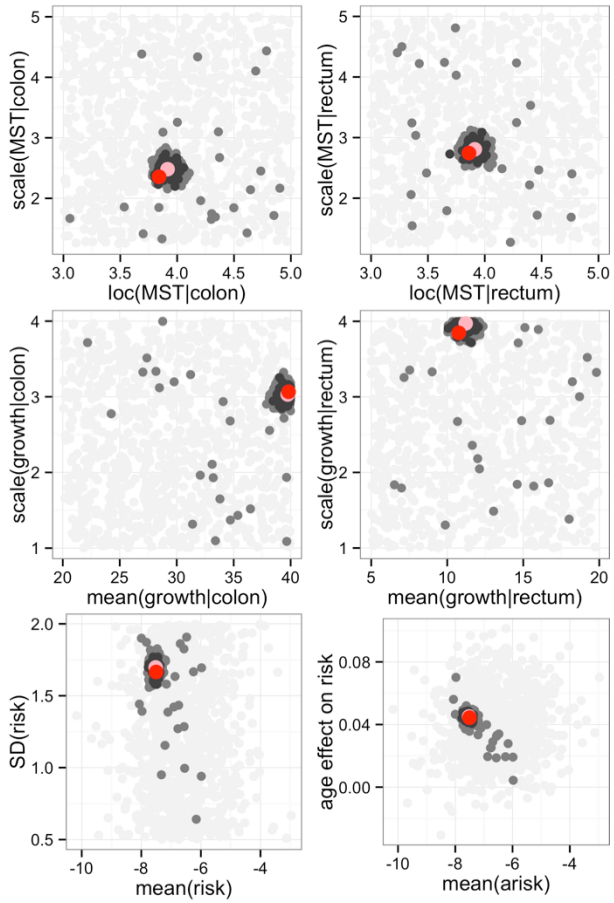

Figure 2: Likelihood plots from preliminary IMIS sampling of CRC-SPIN model runs. The plots show points with non-zero likelihoods in light grey, points with negative log-likelihoods greater than or equal to 65th percentile across points with non-zero likelihoods- in dark grey, points with negative log-likelihoods greater than or equal to 95th percentile in black, and the best point(s) in terms of log likelihood in red.

Acknowledgment

This work was supported by the NIH (award R01GM115839) and the National Cancer Institute (U01-CA-52959, a part of the Cancer Intervention and Surveillance Modeling Network, CISNET). This work was also performed as part of a DOE-NCI collaboration under the Joint Design of Advanced Computing Solutions for Cancer project. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility. This work was completed in part with resources provided by the Research Computing Center at the University of Chicago.

REFERENCES

[1] A. B. Knudsen, A. G. Zauber, C. M. Rutter, S. K. Naber, V. P. Doria-Rose, C. Pabiniak, C. Johanson, S. E. Fischer, I. Lansdorp-Vogelaar and K. M. Kuntz, "Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US Preventive Services Task Force," *JAMA,* vol. 315, no. 23, pp. 2595-2609, June 2016.

[2] C. M. Rutter and J. E. Savarino, "An evidence-based microsimulation model for colorectal cancer: validation and application," *Cancer Epidemiology Biomarkers & Prevention,* vol. 19, no. 8, pp. 1992-2002, 2010.

[3] National Cancer Institute, "CRC-SPIN (RAND)," [Online]. Available: https://resources.cisnet.cancer.gov/registry/packages/crc-spin-rand/. [Accessed 15 Sep 2016].

[4] C. M. Rutter, A. B. Knudsen, T. L. Marsh, V. P. Doria-Rose, E. Johnson, C. Pabiniak, K. M. Kuntz, M. van Ballegooijen, A. G. Zauber and I. Lansdorp-Vogelaar, "Validation of Models Used to Inform Colorectal Cancer Screening Guidelines Accuracy and Implications," *Medical Decision Making,* vol. 36, no. 5, pp. 605-14, Jul 2016.

[5] A. E. Raftery and L. Bao, "Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling," *Biometrics,* vol. 66, no. 4, pp. 1162-73, Dec 2010.

[6] J. Ozik, N. T. Collier, J. M. Wozniak and C. Spagnuolo, "From Desktop To Large-Scale Model Exploration With Swift/T," in *Winter Simulation Conference 2016 Proceedings*, Washington D.C., In Press.

[7] J. M. Wozniak, T. G. Armstrong, M. Wilde, D. S. Katz, E. Lusk and I. T. Foster, "Swift/T: Scalable data flow programming for distributed-memory task-parallel applications," in *Proc. CCGrid 2013*, Delft, 2013.

[8] T. G. Armstrong, J. M. Wozniak, M. Wilde and I. T. Foster, "Compiler Techniques for Massively Scalable Implicit Task Parallelism," in *SC14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2014.