Sequential designs for efficiently training emulators of computationally expensive mathematical models

TA Trikalinos, with AG Ellis, R Iskandar, CH Schmid, JB Wong

Many decisions in health must be made under uncertainty or with incomplete understanding of the underlying phenomena. In such cases, mathematical modeling can help decision-makers synthesize evidence from different sources, estimate and aggregate anticipated outcomes while accounting for stakeholder preferences, understand trade-offs, and quantify the impact of uncertainty on decision making. To be informative, a model should be detailed enough to capture salient aspects of the decisional problem at hand, but a highly detailed model can render routine analyses computationally expensive, hindering its usability.

An emulator is a fast-to-evaluate statistical approximation of a detailed mathematical model. When used in lieu of the detailed model, emulators can expedite tasks that require many repeated evaluations, such as model calibration and value-of-information analyses. Emulators are developed using the output of simulators at specific input values (design points). Developing an emulator that closely approximates the simulator can require many design points, which becomes computationally expensive. We describe a self-terminating active learning algorithm to efficiently develop emulators tailored to a specific emulation task. Its postulated advantages over the prevalent approaches include (1) self-termination, (2) development of emulators with smaller mean squared errors, and (3) ability to achieve a target level of accuracy. To explicate, we develop and compare Gaussian Process emulators of a prostate cancer screening model using the adaptive algorithm versus standard approaches.