Variable selection using summary statistics

Matti Pirinen[1]

[1] *University of Helsinki, Finland, Matti.pirinen@helsinki.fi*

With increasing capabilities to measure a massive number of variables, efficient variable selection methods are needed to improve our understanding of the underlying data generating processes. This is evident, for example, in human genomics, where genomic regions showing association to a disease may contain thousands of highly correlated variants, while we expect that only a small number of them are truly involved in the disease process.

I discuss ideas that have made variable selection practical in human genomics and demonstrate them through our experiences with the FINEMAP algorithm.

(1) Compressing data to light-weight summaries to avoid logistics and privacy concerns related to complete data sharing and to minimize the computational overhead.

(2) Efficient implementation of sparsity assumptions.

(3) Efficient search algorithms.

(4) Use of public reference databases to complement the available summary statistics.

Deconvolution of drug-response heterogeneity in cancer cell populations

Alvaro Köhn-Luque[1]

[1] *Oslo Centre for Biostatistics and Epidemiology, Faculty of Medicine, University of Oslo, Norway, alvaro.kohn-luque@medisin.uio.no*

Ex vivo drug-sensitivity assays are a basic component of biomedical research. Typically, cells are treated with varying drug concentrations and viable cells are measured at one or more time points. Viability curves, and their characteristics (e.g. IC50), allow to compare drug sensitivity across multiple drugs and cell samples. However, the interpretation of those curves is confounded by the presence of cellular heterogeneity in each sample. The presence of several subclones with different drug sensitivities results in an aggregated drug sensitivity profile that does not represent the cell population complexity, and thus hinders the design of precise treatment strategies.

In this talk I will show how to infer on the presence of cellular subclones with differential drug response, using standard cell viability data at the total population level. We build cell population dynamic models of the evolution of individual subclones over time and dose. We estimate the number of subclones, their proportion and drug-response profiles. We validate the methodology on viability data from admixtures of synthetic and actual cancer cell lines at various known frequencies. Finally, we explore the clinical usefulness of the method by deconvolving drug-response heterogeneity in multiple myeloma patient samples.

This is joint, ongoing work with Jasmine Foo, Kevin Leder and Jasmine Noory (University of Minnesota); Arnoldo Frigessi and Even M. Myklebust (University of Oslo); Shannon M. Mumenthaler (University of Southern California); Dagim. S. Tadele (Cleveland Clinic and Oslo University Hospital), Mariaserena Giliberto, Fredrik Schjesvold, Jorrit Enserink and Kjetil Tasken (Oslo University Hospital).

LilleBror for misspecification-robust likelihood free inference in high dimensions

Owen Thomas[1]

[1] *University of Oslo, Norway*

Likelihood-free inference for simulator-based statistical models has recently grown rapidly from its infancy to a useful tool for practitioners. However, models with more than a very small number of parameters as the target of inference have remained an enigma, in particular for the approximate Bayesian computation (ABC) community. To advance the possibilities for performing likelihood-free inference in high-dimensional parameter spaces, here we introduce an extension of the popular Bayesian optimisation based approach to approximate discrepancy functions in a probabilistic manner which lends itself to an efficient exploration of the parameter space. Our method achieves computational scalability by using separate acquisition procedures for the discrepancies defined for different parameters. These efficient high-dimensional simulation acquisitions are combined with exponentiated loss-likelihoods to provide a misspecification-robust characterisation of the marginal posterior distribution for all model parameters. The method successfully performs computationally efficient inference in a 100-dimensional space on canonical examples and compares favourably to existing Copula-ABC methods. We further illustrate the potential of this approach by fitting a bacterial transmission dynamics model to daycare centre data, which provides biologically coherent results on the strain competition in a 30-dimensional parameter space.