ABCruise
16th to 18th May 2016, Helsinki - Stockholm - Helsinki

Organisers: Jukka Corander, Samuel Kaski, Ritabrata Dutta, Michael Gutmann, Jarno Lintusaari.

Contents

1 Administrative Details

1.1 Webpage

Any further information and announcements will be placed on the workshop webpage *http://www.hiit.fi/ABCruise/*.

1.2 Boat

The workshop will take place on a board of a ferry going from Helsinki to Stockholm via Åland and back.The workshop registration will start at 11:30 am on Monday the 16th of May, and the first talk is at 2:10 pm. Next morning the ferry arrives in Stockholm at 9:45 am. The presentations will continue again on Wednesday at about 2:45 pm, and they will last until 10 pm. We return to Helsinki on the 18th of May 10 am.
If you cannot arrive in Helsinki early enough on the 16th of May, we recommend that the you come on the 30th of May and spend the night in Helsinki (on your own costs).

1.3 Travel

- **Train from the airport:** There is a direct train connection from the airport to the central railway station in downtown Helsinki. You have the option of taking either the I train (west loop, 47 min.) or the P train (east loop, 43 min.), both of which arrive at Helsinki Central Station (Rautatieasema). Tickets cost 5 euros and can be bought at the machines on the platforms or on the train themselves: look for the carriages that state they have ticket sales (lipunmyynti) as some have no ticket sales (ei lipunmyynti).

- **Bus from the airport:** You can also take the local bus 615, which costs the same but takes slightly longer to arrive at the central station. Another option is to take the Finnair Express bus to downtown, which is quicker, but slightly more expensive (6.30 euros). From downtown, you need to come to the ship terminal (Olympiaterminaali), which can be reached with tram 2 (included in train/bus ticket if tram is entered within 90 minutes from buying the ticket, but excluded if you take the Finnair Express bus). The tram trip will take about 13 minutes. Walking the same distance will take about 30 minutes.

- **Website to plan your trip using public transport:** The easiest way to plan the trip from and to the airport is to use the journey planner (*http://www.reittiopas.fi/en/*).
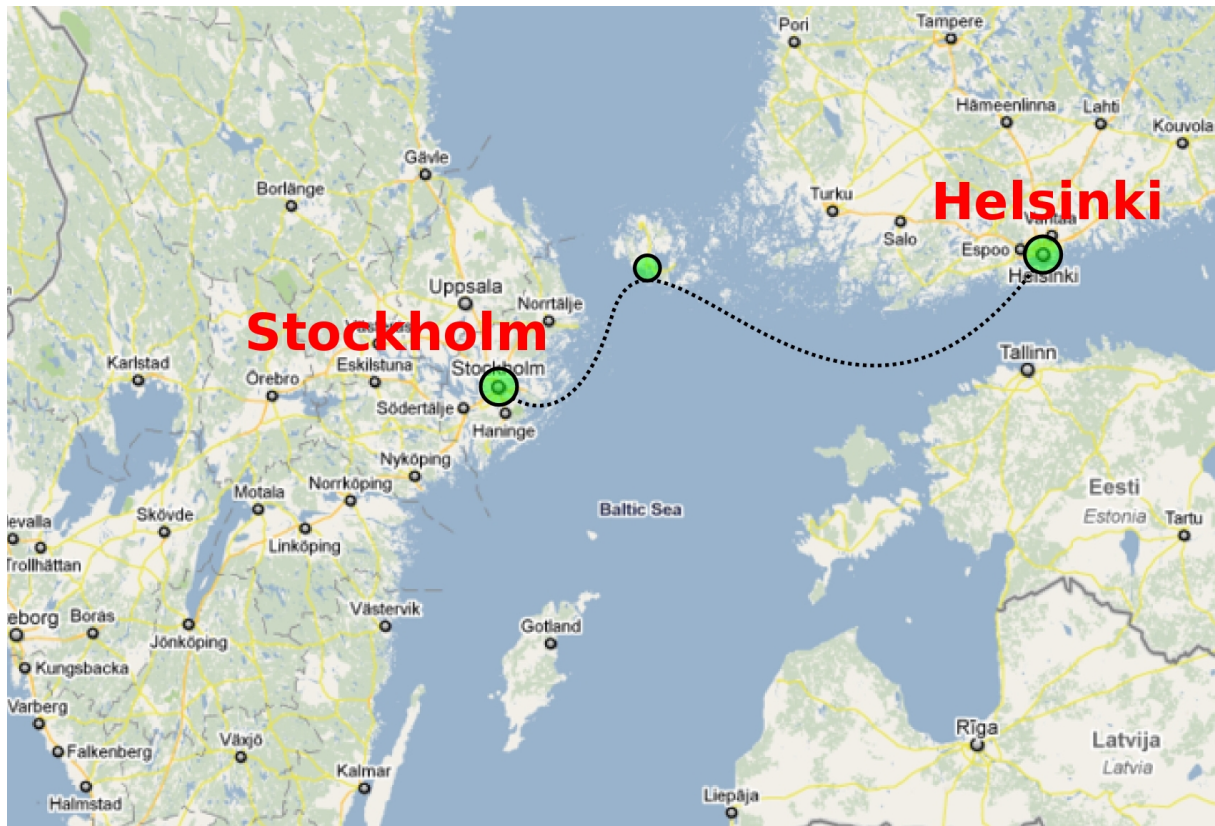
The journey planner will also provide you with maps of the journey.

- **Taxi:** Another option is to take taxi from the airport. The price will be about 50 euros, and the trip will take 30 minutes depending on the traffic.

1.4 Practical Information

Some practical details related to the boat workshop are given below.

- We meet on Monday 16.05 at 12:00 a.m. in the Olympia terminal, Olympiaranta 1, located in downtown Helsinki. Put simply this address in google map for directions. It takes about 15 min to walk from the central railway station to the terminal. You can also take tram 2 from the central railway station. The tram stop is just in front of the station. If the railway station is behind you, you should take the tram going to the left.

- The organizers will be in the terminal main lobby (near the group check-in counter). Normally there are relatively few people at that time in the lobby, so it should not

be a problem to spot us. We will give you your tickets/key cards, along with your badge and abstract book.

- You need a passport or EU identity card to enter the ship.

- If you come after 13:00, we leave your key card at the group check-in. You can get it by giving in your name.

- Most likely, the cabins will not be ready when we arrive, so that we will go directly to the conference facilities (the very front of the ship, the entrance is one floor down from the Promenade). There is enough room for luggage there.

- Presentations should be prepared in PDF. We will have a "general" presentation computer, but surely you can use your own laptop. If you use the presentation computer, all presentations and movie files must be downloaded to the presentation computer before the session starts.

- The posters will be in the same room. Blue tack and scotch tape will be provided.

- There is a free WLAN from Silja in the conference facilities, although after the boat leaves the harbour, it may be quite slow. Of course also 4G/3G Internet will work, there is mobile network during almost all of the trip (except for the crossing from Aland to Sweden).

- We arrive in Stockholm at 9:45 a.m. on Tuesday, and the presentations start at 2:45 p.m. You can stay onborad or visit Stockholm during the first part of the day.

- Note that Swedish time is DIFFERENT from Finnish by 1 hour (9 a.m. in Finland is 8 a.m. in Stockholm). We stick to Finnish time on Mon and Wed, and to Swedish time on Tue. This means that you will have an extra hour to sleep the night of Mon/Tue, and one hour less the night of Tue/Wed, so please plan your evening/morning program accordingly.

- We arrive in Helsinki at 10:00 a.m.

- It takes about 1 hour to get to the airport from the ferry terminal.

2 Timetable

Monday, 16th of May
09:15–11:30 ABC Tutorial (Helsinki University central campus[1])
12:00–13:00 Registration (Olympia Terminal[2], Helsinki)
13:00–14:00 Lunch
14:10–14:50 Richard Everitt
14:50–15:10 Coffee Break
15:10–15:50 Jean-Michel Marin
15:50–16:30 Clara Grazian
16:30–17:20 Departure Break
17:20–18:10 Poster Session, Snacks
18:10–18:50 Thomas Schön
18:50–19:30 Caroline Colijn
19:30–21:30 Dinner Buffet
Tuesday, 17th of May
07:00–09:45 Breakfast
09:45–14:30 Personal tour in Stockholm
14:30–14:45 Coffee
14:45–15:25 Laurent E. Calvet
15:25–16:05 Aki Vehtari
16:05–16:45 Michael Gutmann
16:45–17:00 Boat leaves Stockholm harbor
17:00–19:00 Dinner Buffet
19:00–19:40 Richard Wilkinson
19:40–20:20 Julien Stoehr
20:20–21:00 Ted Meeds
21:00–22:00 Social and poster session, snacks
Wednesday, 18th of May
07:00 Breakfast
10:00 Boat arrives in Helsinki (Olympia Terminal, Helsinki)

---

[1]Lecture hall 4, (3rd floor), Fabianinkatu 33, Helsinki 00170.
[2]Olympiaranta 1, Helsinki 00130

3 Invited talks, in order of appearance

Synthetic Likelihood Sequential Monte Carlo
Richard Everitt (University of Reading).

ABC methods are part of a wider classes of techniques in which an approximated likelihood is used. Standard ABC uses a non-parametric estimate of the distribution of a vector of summary statistics given the parameter, whereas Synthetic Likelihood (Wood, 2010) approximates the same distribution as a multivariate Gaussian. In this talk we discuss the use of approximate likelihoods within importance sampling (IS) and sequential Monte Carlo (SMC) methods, with a particular focus on the use of Synthetic Likelihood.

ABC random forest for parameter estimation
Jean-Michel Marin (University of Montpellier)

When the calculation of the likelihood function is intractable, it is necessary to use approximations. This is a very common problem in some application areas, for instance for population genetic models. Faced with this difficulty, two strategies can be considered. The first is to directly approximate the likelihood function and use conventional inference techniques based on this approximation. That is the case of composite likelihood methods and of variational optimization schemes. Another possibility, in the field of Bayesian statistics, is to use intensive simulation techniques. Approximate Bayesian Computation strategies (ABC) belong to this class. The basic idea is to simulate new data from the model and compare them to the observed one. That is a very active research field which results in a rapprochement between Bayesian inference techniques and statistical learning methods. In Pudlo et al. (2016), the authors show how random forests can be used to discriminate between models. In this talk, we show how this method can extended to the context of parameter estimation and computation of credibility intervals. The idea is to adapt the Quantile Random Forest technique of Meinshausen (2006).
Reference:
Jean-Michel Marin, Pierre Pudlo, Christian P. Robert and Robin Ryder (2012) Approximate Bayesian Computation methods, Statistics and Computing, 22(6), 1167–1180.
Nicolai Meinshausen (2006) Quantile Regression Forests, Journal of Machine Learning Research, 7, 983–999.

Semiparametric estimation via ABC methods
Clara Grazian (Sapienza Università di Roma)

We describe a simple method for making inference on a functional of a multivariate distri-

bution. The method is based on a copula representation of the multivariate distribution, where copula is a flexible probabilistic tool that allows the researcher to model the joint distribution of a random vector in two separate steps: the marginal distributions and a copula function which captures the dependence structure among the vector components. The method is also based on the properties of an approximate Bayesian Monte Carlo algorithm, where the proposed values of the functional of interest are weighted in terms of their empirical likelihood. This method is particularly useful when the likelihood function associated with the working model is too costly to evaluate or when the working model is only partially specified. We consider the general problem of estimating some specific quantities of interest of a generic copula (such as, for example, tail dependence index or Spearman's rho) by adopting an approximate Bayesian approach. In particular, we discuss the use of the an approximate Bayesian computation algorithm based on the empirical likelihood approximation of the marginal likelihood of the quantity of interest. This approach produces an approximation of the posterior distribution of the quantities of interest, based on an approximation of the likelihood function and on a Monte Carlo approximation of the posterior distribution via simulations. Our approach is general, in the sense that it could be adapted both to parametric and nonparametric modelling of the marginal distributions. Also, the use of empirical likelihood avoids the need of choosing a specific parametric copula model.

Particle smoothing for near-degenerate or intractable state transition models
Thomas Schön (Uppsala University)

Smoothing algorithms based on sequential Monte Carlo (i.e., particle smoothers) are key computational tools for addressing the smoothing problem in general nonlinear, non-Gaussian state space models. However, for certain challenging but common model classes the state-of-the-art algorithms still struggle. This work is motivated in particular by two such model classes: (i) models where the state transition kernel is (nearly) degenerate, i.e. (nearly) concentrated on a low-dimensional manifold, and (ii) models where point-wise evaluation of the state transition density is intractable. Both types of models arise in many applications of interest, including tracking, epidemiology, and econometrics. Unfortunately, two of the most successful smoothing strategies–the forward-backward approach and the two-filter approach–cannot be directly applied to these types of models. To alleviate this, we propose a "particle rejuvenation" technique to enable the forward-backward strategy for (nearly) degenerate models and, by extension, for intractable models. The focus of this talk will be on a this extension, where a nearly degenerate approximation of an intractable transition model opens up for using the new particle smoother also for this class of models. The proposed method is essentially a variant of the approximate

Bayesian computation (ABC) technique. However, while ABC is typically used for inference in models with intractable likelihoods for SMC implementations), we use it here to address the issue of intractable transitions.

Joint work with Fredrik Lindsten (Uppsala University), Pete Bunch (University of Cambridge) and Sumeetpal S. Singh (University of Cambridge).

Towards tree topology ABC
Caroline Colijn (Imperial College London)

It has long been accepted that the topology of evolutionary trees is shaped by the underlying evolutionary process. In particular, reconstructed evolutionary trees are more asymmetric than trees from reasonable generative models, such as the Yule or birth-death processes. However, inference of evolution from trees is hindered by a lack of ways to do deep structural comparisons between simulated and observed trees, or between trees from different datasets. I will introduce a metric on tree topologies, based on characterising any possible tree topology as an integer. This metric separates trees derived from several random models known to produce different tree topologies. I apply the metric to data from human influenza A and find that it separates tropical flu from seasonal flu over the same time period. Finally I show in a simple example that it is possible to infer the parameter of a biased speciation model using an ABC approach.

Structural Dynamic Analysis of Systematic Risk
Laurent E. Calvet (HEC Paris)

This paper introduces a structural dynamic factor model (SDFM) for stock returns. In contrast to standard linear factor models, the new approach accounts for nonlinear effects of common factors when the distance-to-default is small. We develop a toolkit of econometric methods for the SDFM, based on indirect inference and Approximate Bayesian Computation (ABC) filtering, which permit to estimate, filter and predict systematic risk. We apply the SDFM to measure the systematic risk of financial institutions and obtain their rating of default and speculative features. Joint work with Veronika Czellar and Christian Gourièroux.

On Gaussian processes, Bayesian optimization and ABC
Aki Vehtari (Aalto University)

I'll first give short introduction of using Gaussian processes inside of inference algorithms, such as Bayesian optimization and ABC. Then I present some examples how the performance of these inference methods can be improved by using additional structural infor-

mation in Gaussian process models.

Fast Likelihood-Free Inference via Bayesian Optimization
Michael Gutmann (University of Helsinki)

Statistical models may be specified in terms of stochastic computer program – a simulator – which can generate samples from the model for any configuration of the parameters. While such models support complex data generating mechanisms, the likelihood function is generally incomputable which renders statistical inference difficult. Several likelihood-free inference methods have been proposed which share the basic idea of identifying the model parameters by finding values for which the discrepancy between simulated and observed data is small. Examples are indirect inference and approximate Bayesian computation. A major obstacle to using these methods is their computational cost. The cost is largely due to the need to repeatedly simulate data sets and the lack of knowledge about how the parameters affect the discrepancy. We propose a strategy which combines probabilistic modeling of the discrepancy with optimization to facilitate likelihood-free inference. The strategy is implemented using Bayesian optimization and is shown to accelerate the inference through a reduction in the number of required simulations by several orders of magnitude.
Refereence: M.U. Gutmann and J. Corander, Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models, Journal of Machine Learning Research, in press.

Surrogate modelling and ABC
Richard Wilkinson (University of Sheffield)

When using ABC we are often limited by the available computational resource. An idea that has become popular in the uncertainty quantification literature, particularly when analysing deterministic simulators, is to build a surrogate model, sometimes referred to as an emulator. This is a cheap approximation that can be used in place of the simulator. The same idea has recently been used for stochastic models, but the stochastic situation is far more challenging than the deterministic model setting and it is not yet clear how well this approach will work. This talk will review some of the work in this direction, and highlight some of the challenges and drawbacks of this approach.

Adaptive ABC model choice and geometric summary statistics for hidden Gibbs random fields
Julien Stoehr (University College Dublin)

Selecting between different dependency structures of hidden Markov random fields can be very challenging, due to the intractable normalising constant in the likelihood. We answer this question with approximate Bayesian computation (ABC) which provides a model choice method in the Bayesian paradigm. The present work has to answer the absence of available sufficient statistics for hidden Markov random fields. To make up for the absence of sufficient statistics in that context, we propose geometric summary statistics. The general approach to construct these intuitive statistics relies on a clustering analysis of the sites based on the observed colors and plausible latent graphs. So the question of reliability remains. The efficiency of ABC model choice based on these statistics is evaluated via a local error rate which may be of independent interest. As a byproduct we derived an ABC algorithm that adapts the dimension of the summary statistics to the dataset without distorting the model selection.

Likelihood-free Inference by Controlling Simulator Noise
Ted Meeds (UvA/VU)

Likelihood-free inference, or approximate Bayesian computation (ABC), is a general framework for performing Bayesian inference in simulation-based science. In this talk I will discuss two new approaches to likelihood-free inference that involve explicit control over a simulation's randomness. By re-writing simulation code with two sets of arguments, the simulation parameters and its random numbers, many algorithmic options open up. The first approach, called Optimisation Monte Carlo, in an algorithm that efficiently and independently samples parameters from the posterior by first sampling a set of random numbers from a prior distribution, then running an optimisation algorithm–with fixed random numbers–to match simulation statistics with observed statistics. The second approach is recent and ongoing research on a variational ABC algorithm that has been written in an auto-differentiation language allowing for the gradients of the variational parameters to be computed through the simulation code itself.

4   Posters

Adaptive Summary Statistic Selection within a Sequential Monte Carlo algorithm for Approximate Bayesian Computation
Sophie Watson (University of Bristol)

Sequential Monte Carlo (SMC) algorithms for Approximate Bayesian Computation (ABC) have been shown to improve the accuracy of inference, compared to standard rejection ABC. The reason for this improvement is that standard rejection ABC proposes particles from the prior, whereas the proposal distribution in SMC approaches the posterior. However, as we demonstrate here, standard SMC-ABC methods are still very sensitive to the dimensionality of summary statistics. In addition, such methods can be expensive to implement as they require many simulations from the model. Often a large proportion of simulations are quickly 'rejected' and subsequently ignored. SMC-ABC methods are also affected by the choice of summary statistics: Summarising the data poorly results in worse inference. We present a novel algorithm which uses knowledge gained from all historical simulations from the model to select summary statistics to use for the next iteration of SMC. We show that our algorithm performs favourably on a toy model in which summary statistics have been deliberately poorly chosen and obtains estimates which are remarkably close to those which could be obtained using sufficient statistics.

Improving approximate Bayesian computation through randomized quasi Monte Carlo
Alexander Buchholz and Nicolas Chopin (CREST)

When the likelihood function of a statistical model is intractable but a simulator of the model is available, approximate Bayesian computation provides us with a tool for statistical inference. Since the computational cost mainly depends on the simulator, one is eager to limit the number of simulations necessary to obtain samples from the posterior. Therefore, we suggest to use randomized quasi Monte Carlo (RQMC) sequences when simulating from the prior distribution instead of ordinary Monte Carlo (MC). We show that the variance of estimators based on posterior samples is reduced. Consequently, a smaller number of observations from the simulator is needed for the same precision. We develop the supporting theory for the asymptotic variance reduction and illustrate our findings via the study of inference in a birth-death-mutation process of [4], commonly used in epidemiology. We compare simulations based on importance sampling (IS), adaptive importance sampling and sequential Monte Carlo. Our work uses results of [1] and [3] for the non-asymptotic variance of RQMC in IS and we extend a central limit theorem of [2] to multivariate sequences that mix MC and QMC. References: [1] S Agapiou, O

Papaspiliopoulos, D Sanz-Alonso, and AM Stuart. Importance sampling: Computational complexity and intrinsic dimension. arXiv preprint arXiv:1511.06196, 2015. [2] Giray Okten, Bruno Tuffin, and Vadim Burago. A central limit theorem and improved error bounds for a hybrid-monte carlo sequence with applications in computational finance. Journal of Complexity, 22(4):435–458, 2006. [3] Art B. Owen. Scrambled net variance for integrals of smooth functions. 25(4):1541–1562. [4] Mark M Tanaka, Andrew R Francis, Fabio Luciani, and SA Sisson. Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data. Genetics, 173(3):1511–1520, 2006.

MCMC for Inverse Problems in Brain Imaging
Philip Maybank (University of Reading)

In Neuroscience, mean-field models are nonlinear dynamical systems that are used to describe the evolution of mean neural population activity, within a given brain region such as the cortex. Mean-field models typically contain 10-100 unknown parameters, and receive high-dimensional noisy input from other brain regions. Here we present preliminary results on inferring mechanistic parameters in the differential equations.

Inferring viral epidemiological features from HIV genomic data using Approximate Bayesian Computation with Sequential Monte Carlo
Neke Ibeh (University of Ottawa)

Understanding the transmission patterns of infectious diseases is critical in order to monitor both their spread and the efficacy of new public health policies. To achieve such a goal, phylogenetic methods have extensively been developed and applied to reconstructing viral ancestral relationships – particularly in the context of HIV epidemics. In this case, initiatives such as the Swiss HIV Cohort Study (SHCS) now contain huge amounts of genomic data. However, current approaches for quantifying the epidemiological dynamics of diseases are computationally intensive and fail to scale well with this magnitude of data. Here we develop an approach based on Approximate Bayesian Computation (ABC) with sequential Monte Carlo (SMC), that circumvents the issue of a costly likelihood function calculation. By means of simulations, we show that our proposed method is capable of inferring key epidemiological parameters of the Swiss HIV epidemic with great accuracy. Our reanalysis of data from the SHCS shows that we are able to distinguish transmission patterns in two culturally different parts of Switzerland. In light of these results, we propose that our method will allow us to evaluate the impact of new public health policies (e.g., implementation of a needle exchange program in the case of HIV) based on genetic data sampled before and after the implementation of a new policy.

Figure 1: Map of 'Silja Symphony': Conference venue are in the head of deck 6