CIRM

End-to-end Bayesian learning

Abstract book

25-29th of October 2021, CIRM, Marseille

	Monday	Tuesday	Wednesday	Thursday	Friday
8:45 - 9:00	Introduction				
9:00 - 10:30	Master class I.1 Rémi Bardenet	Master class I.2 Rémi Bardenet	Master class II.1 Elisabeth Gassiat	Master class II.2 Elisabeth Gassiat	Contributed Sebastiano Grazzi Contributed Andrea Bertazzi Contributed Andi Wang
10:30 - 11:00	Coffee Break				
11:00 - 12:30	Contributed Marina Riabiz Contributed Hai-DAng Dau Contributed James Thornton	Contributed Takuo Matsubara Contributed Deborah Sulem Contributed Torben Sell	Contributed Bart Eggen Contributed Gonzalo Mena Contributed Solon Karapanagiotis	Contributed Amani Alahmadi Contributed Luke Kelly	Invited Talk Randal Douc Prize session
12:30 - 14:00	Lunch Break				
14:00 - 15:00	Invited Anthony Lee	Invited Florence Forbes		Invited Gareth Roberts	14:00 - 15:00
15:00 - 15:30	Coffee Break			Coffee Break	
15:30 - 16:30	Invited Judith Rousseau	Invited Eric Moulines		Practical Session Charles Ollion	15:30 - 17:30
16:30 - 17:00	Snooze Time			Snooze Time	
17:00 - 19:30	Poster Session	Speed Mentoring		Social Activities	18:00 - 19:30
19:30 - 21:00	Diner				

Monday

Master Class I.1: 9:00 - 10:30

Rémi Bardenet: A tutorial on Bayesian machine learning: what, why and how

Contributed: 11:00 - 12:30

Marina Riabiz: Optimal Thinning of MCMC Output
The use of heuristics to assess the convergence and compress the output of Markov chain Monte Carlo can be sub-optimal in terms of the empirical approximations that are produced. Typically a number of the initial states are attributed to "burn in" and removed, whilst the remainder of the chain is "thinned" if compression is also required. In this paper we consider the problem of retrospectively selecting a subset of states, of fixed cardinality, from the sample path such that the approximation provided by their empirical distribution is close to optimal. A novel method is proposed, based on greedy minimisation of a kernel Stein discrepancy, that is suitable when the gradient of the log-target can be evaluated and an approximation using a small number of states is required. Theoretical results guarantee consistency of the method and its effectiveness is demonstrated in the challenging context of parameter inference for ordinary differential equations. Software is available in the Stein Thinning package in both Python and MATLAB.
ArXiv preprint: https://arxiv.org/pdf/2005.03952.pdf

Hai-Dang Dau: Waste-free Sequential Monte Carlo
A standard way to move particles in a SMC sampler is to apply several steps of a MCMC (Markov chain Monte Carlo) kernel. Unfortunately, it is not clear how many steps need to be performed for optimal performance. In addition, the output of the intermediate steps are discarded and thus wasted somehow. We propose a new, waste-free SMC algorithm which uses the outputs of all these intermediate MCMC steps as particles. We establish that its output is consistent and asymptotically normal. We use the expression of the asymptotic variance to develop various insights on how to implement the algorithm in practice. We develop in particular a method to estimate, from a single run of the algorithm, the asymptotic variance of any particle estimate. We show empirically, through a range of numerical examples, that waste-free SMC tends to outperform standard SMC samplers, and especially so in situations where the mixing of the considered MCMC kernels decrease across iterations (as in tempering or rare event problems).

James Thornton: Differentiable Particle Filtering via Entropy-Regularized Optimal Transport
Particle Filtering (PF) methods are an established class of procedures for performing inference in non-linear state-space models. Resampling is a key ingredient of PF, necessary to obtain low variance likelihood and states estimates. However, traditional resampling methods result in PF-based loss functions being non-differentiable with respect to model and PF parameters. In a variational inference context, resampling also yields high variance gradient estimates of the PF-based evidence lower bound. By leveraging optimal transport ideas, we introduce a principled differentiable particle filter and provide convergence results. We demonstrate this novel method on a variety of applications.

Invited Talk: 14:00 - 15:00

Anthony Lee: Intermediate distributions and complexity bounds for SMC
It is now fairly common to use Sequential Monte Carlo (SMC) algorithms for normalizing constant

estimation of high-dimensional, complex distributions without any particular structure. In order for the algorithm to give reasonable accuracy, it is well known empirically that one must introduce appropriate intermediate distributions that allow the particle system to "gradually evolve" from a simple initial distribution to the complex target distribution, and one must also specify an appropriate number of particles to control the error. Since both the number of intermediate distributions and the number of particles affect the computational cost of the algorithm, it is crucial to study and attempt to minimize the computational cost of the algorithm subject to constraints on the error. We present three strategies that have been used to specify intermediate distributions and provide bounds on the computational complexity of normalizing constant estimation with well-tuned sequences, which involves obtaining bounds on the length of sequences of intermediate distributions. Although the results for SMC algorithms involve some fairly strong assumptions on the Markov kernels involved, they are to the best of our knowledge the only general results available thus far. This primarily theoretical analysis also suggests where further research is required to tune the approach.

Invited Talk: 15:30 - 16:30

Judith Rousseau: TBA

Tuesday

Master Class I.2: 9:00 - 10:30

Rémi Bardenet: A tutorial on Bayesian machine learning: what, why and how

Contributed: 11:00 - 12:30

Takuo Matsubara: Robust Generalised Bayesian Inference for Intractable Likelihoods
Generalised Bayesian inference is a framework for updating prior beliefs through the use of a general loss function, rather than a likelihood. This approach can confer robustness to outliers, so it is particularly attractive when working with models that are mis-specified. In this work we consider the use of kernel Stein discrepancies as loss functions, which allows for generalised Bayesian inference with likelihoods involving an intractable normalising constant. Our approach produces pseudo-posteriors that are both (i) computable in closed form for exponential family models, and (ii) robust to model mis-specification. On a theoretical level, we show two fundamental properties, consistency and asymptotic normality of the pseudo-posterior. Additionally, we prove a Bayesian analogue of bias-robustness in the mis-specified context. Then, we provide numerical experiments involving a range of intractable likelihoods, with applications to kernel-based exponential family models and non-Gaussian graphical models.

Deborah Sulem: Bayesian estimation of nonlinear Hawkes process
Multivariate point processes are widely applied to model event-type data such as natural disasters, online message exchanges, financial transactions or neuronal spike trains. One very popular point process model in which the probability of occurrences of new events depend on the past of the process is the Hawkes process. In this work we consider the nonlinear Hawkes process, which notably models excitation and inhibition phenomena between dimensions of the process. In a nonparametric Bayesian estimation framework, we obtain concentration rates of the posterior distribution on the parameters, under mild assumptions on the prior distribution and the model. These results also lead to convergence rates of Bayesian estimators. Another object of interest in event-data modelling is to recover the graph of interaction - or Granger connectivity graph - of the phenomenon. We provide consistency guarantees on Bayesian methods for estimating this quantity; in particular, we prove that the posterior distribution is consistent on the graph adjacency matrix of the process, as well as a Bayesian estimator based on an adequate loss function.
ArXiv preprint: https://arxiv.org/abs/2103.17164

Torben Sell: Dimension-robust function space priors for stochastic control
We discuss a new prior on functions spaces which scales more favourably in the dimension of the function's domain compared to the usual Karhunen-Loéve function space prior, a property we refer to as dimension-robustness. The prior is a Bayesian neural network prior, where each weight and bias has an independent Gaussian prior, but with the key difference that the variances decrease in the width of the network, such that the variances form a summable sequence and the infinite width limit neural network is well defined. We show that the resulting posterior of the unknown function is amenable to sampling using Hilbert space Markov chain Monte Carlo methods. These sampling methods are favoured because they are stable under mesh-refinement, in the sense that the acceptance probability does not shrink to 0 as more parameters are introduced to better approximate the well-defined infinite limit. We show that these priors are competitive and have distinct advantages over other function space priors. Upon defining a suitable likelihood for continuous value functions in a Bayesian approach to reinforcement learning, the prior is used in numerical examples to illustrate its performance and dimension-robustness.
ArXiv preprint: https://arxiv.org/abs/2012.10943

Invited Talk: 14:00 - 15:00

Florence Forbes: Simulation-based Bayesian inference for high dimensional inverse problems

We investigate the use of learning approaches to handle Bayesian inverse problems in a computationally efficient way when the observations to be inverted present a moderately high number of dimensions and are in large number. We propose a tractable inverse regression approach which has the advantage to produce full probability distributions as approximations of the target posterior distributions. These distributions have several interesting features. They provide confidence indices on the predictions and can be combined with importance sampling or approximate Bayesian computations schemes for a better exploration of inverse problems when multiple equivalent solutions exist. They generalised easily to variants that can handle non Gaussian data, dependent or missing observations. The relevance of the proposed approach is illustrated on synthetic examples and on two real data applications, in the context of planetary remote sensing and neuroimaging. The approach shows interesting capabilities both in terms of computational efficiency and multimodal inference.

Invited Talk: 15:30 - 16:30

Eric Moulines: TBA

Wednesday

Master Class II.1: 9:00 - 10:30

Elisabeth Gassiat: Bayesian multiple testing for dependent data and hidden Markov models
Reference for the Master Class:
Multiple testing in non parametric hidden Markov models: an empirical Bayes approach (K. Abraham, I. Castillo, E. Gassiat)
Fundamental limits for learning hidden Markov model parameters (K. Abraham, E. Gassiat, Z. Naulet)

Contributed: 11:00 - 12:30

Bart Eggen: Bayesian Sensitivity Analysis for Causal Inference
In the field of causal inference, sensitivity analysis is very important to assess the robustness of study conclusions when certain assumptions are not satisfied. We will look at the missing data model and perform sensitivity analysis under the assumption that missing outcomes are missing completely at random. Scharfstein et. al. (2003) introduced a fully Bayesian methodology and conjectured a Bernstein-von Mises theorem for the posterior of the outcome, by incorporating prior beliefs about non-identifiable, but interpretable parameters. We prove that their conjecture is correct by providing three Bernstein-von Mises theorems. The results are obtained using the following priors: a Dirichlet process prior on the distribution of the outcome; a Dirichlet process prior on the distribution of the outcome conditional on the subject being treated; a Gaussian process prior on the density of the distribution of the outcome conditional on the subject being treated. We also provide a simulation study, showing the performance of the methods in finite sample scenarios.

Gonzalo Mena: On the choice of priors for the estimation of infection fatality rates in the absence of serological data
Suppose we want to estimate stratified infection fatality rate (IFR) of a new disease. We propose a novel hierarchical bayesian methodology for the estimation of such rates under the assumption that stratified death information is reliable, but case (and infection) information is incomplete. Our estimates are based on a cascade of binomial models linking infection, cases, and deaths. To infer infections from cases we estimate a mapping from a reporting proxy (such as testing rates) to reporting rates. We discuss the identifiability issues that appear as a consequence of incomplete data and over-parameterization. We also discuss how model ensembling can lead to more robust estimates. We apply this model to understand COVID-10 related mortality in Santiago, Chile. From our model we find a strong socioeconomic gradient of IFRs in young age groups.
https://www.medrxiv.org/content/10.1101/2021.01.12.21249682v1

Solon Karapanagiotis: Tailored Bayes: a risk modelling framework under unequal misclassification costs
Risk prediction models are a crucial tool in healthcare and beyond. Risk prediction models with a binary outcome (i.e., binary classification models) are often constructed using methodology which assumes the costs of different classification errors are equal. In many healthcare applications this assumption is not valid, and the differences between misclassification costs can be quite large. For instance, in a diagnostic setting, the cost of misdiagnosing a person with a life-threatening disease as healthy may be larger than the cost of misdiagnosing a healthy person as a patient. In this work, we present Tailored Bayes (TB), a novel Bayesian inference framework which "tailors" model fitting to optimise predictive performance with respect to unbalanced misclassification costs. We use simulation studies to showcase when TB is expected to outperform standard Bayesian methods

in the context of logistic regression. We then apply TB to three real-world applications, a cardiac surgery, a breast cancer prognostication task and a breast cancer tumour classification task, and demonstrate the improvement in predictive performance over standard methods. Finally, we extend the framework to incorporate a variable selection procedure, a ubiquitous challenge in statistical modelling, especially, with the rise of high-dimensional data. We show that TB favours smaller models (with less covariates) compared to standard Bayes paradigm (SB), whilst performing better or no worse than SB. This pattern was seen both in simulated and real data. In addition, we show the relative importance of the covariates changes when we consider unequal misclassification costs.This has implications for risk prediction models since smaller models may result in lower data collection costs and different covariates being selected for further downstream analysis, for instance in genetic fine-mapping and related applications.

Thursday

Master Class II.2: 9:00 - 10:30

Elisabeth Gassiat: Bayesian multiple testing for dependent data and hidden Markov models
Reference for the Master Class:
Multiple testing in non parametric hidden Markov models: an empirical Bayes approach (K. Abraham, I. Castillo, E. Gassiat)
Fundamental limits for learning hidden Markov model parameters (K. Abraham, E. Gassiat, Z. Naulet)

Contributed: 11:00 - 12:30

Amani Alahmadi: SMC ABC estimator for ODE models
Approximate Bayesian Computation (ABC) is a popular tool for estimating the parameters of dynamical systems models, and in particular non-linear differential equation models. It is a Monte Carlo method designed specifically for models in which the likelihood is computationally intractable or expensive, but for which data is relatively easy to simulate. One variant of ABC, known as Sequential Monte Carlo ABC (SMC ABC), shows promise as an efficient methodology for parameter estimation, but some current implementations fail to accurately estimate the posterior distribution of noise variance when applied to Ordinary Differential Equation (ODE) models. Here we present a modified SMC ABC algorithm and propose a new summary statistic that facilitates accurate estimation of noise variance in ODE models. These innovations also result in improved posterior predictive intervals. We apply the proposed method to two ODE epidemiological models, and demonstrate that it outperforms standard SMC ABC in terms of accuracy, and compares favourably with a Markov chain Monte Carlo (MCMC) method in terms of both accuracy and overall computational effort.

Luke Kelly: Coupled MCMC for Bayesian phylogenetic inference
Many modern phylogenetic methods specify a generative model and take a Bayesian approach to inference. However, phylogenetic posterior distributions are intractable functions of the tree and model parameters and are often highly multimodal. Markov chain Monte Carlo is the primary tool for inference, via either extremely long marginal chains or coupled tempered chains, although there have been recent developments using sequential Monte Carlo, piecewise deterministic Markov processes and variational approximations for specific problems. In any case, we lack principled methods to quantify convergence and mixing of Markov schemes on the space of trees, so it is difficult to separate modelling and fitting errors. We extend recent work on couplings of MCMC transition kernels to phylogenetic inference problems in order to construct unbiased estimators and diagnose convergence. The extension is not straightforward due to the complexities of working in tree space and because we couple existing marginal kernels which only operate on a small subset of the state at each iteration. We illustrate our convergence diagnostics and unbiased estimators on a variety of problems and discuss their usefulness compared to other methods.

Benjamin Guedj: Dichotomize and Generalize: PAC-Bayesian Binary Activated Deep Neural Networks
We present a comprehensive study of multilayer neural networks with binary activation, relying on the PAC-Bayesian theory. Our contributions are twofold: (i) we develop an end-to-end framework to train a binary activated deep neural network, (ii) we provide nonvacuous PAC-Bayesian generalization bounds for binary activated deep neural networks. Our results are obtained by minimizing the expected loss of an architecture-dependent aggregation of binary activated deep

neural networks. Our analysis inherently overcomes the fact that binary activation function is non-differentiable. The performance of our approach is assessed on a thorough numerical experiment protocol on real-life datasets.

Invited Talk: 14:00 - 15:00

Gareth Roberts: Regenerative non-reversible MCMC and the Restore algorithm
The talk will describe a flexible new non-reversible MCMC methodology called the Restore algorithm. The algorithm does not require any augmentation of the state space but possesses an embedded regenerative structure which facilitates the proof of good theoretical properties as well as the construction of efficient associated coupling from the past algorithms. This is joint work with Murray Pollock and Andi Wang.

Friday

Contributed: 9:00 - 10:30

Sebastiano Grazzi: Sticky PDMP samplers for sparse and local inference problems
During the talk, I will present the sticky PDMP samplers. These are new efficient Monte Carlo methods based on the simulation of continuous-time piecewise deterministic Markov processes (PDMPs) suitable for inference in high dimensional sparse models, i.e. models for which there is prior knowledge that many coordinates are likely to be exactly 0. This is achieved with the fairly simple idea of endowing existing PDMP samplers with sticky coordinate axes, coordinate planes, etc. Upon hitting those subspaces, an event is triggered, during which the process sticks to the subspace, this way spending some time in a sub-model. That introduces non-reversible jumps between different (sub-)models. This talk is based on the manuscript https://arxiv.org/abs/2103.08478 which is joint work with Joris Bierkens, Frank van der Meulen and Moritz Schauer.

Andrea Bertazzi: Approximations of Piecewise deterministic Markov processes and their convergence properties
Piecewise deterministic Markov processes (PDMP) received substantial interest in recent years as an alternative to classical Markov chain Monte Carlo algorithms. The applicability of PDMP to real world scenarios is currently limited by the fact that these processes can be simulated only when specific properties of the target distribution are known beforehand. In order to overcome this problem, we introduce an Euler-type discretisation scheme for PDMP which does not need such pre-requisite knowledge. For the resulting schemes we study both pathwise convergence to the continuous process as the step size converges to zero and convergence in law to the target measure in the long time limit. Finally, we study the empirical performance of the Euler discretised zig-zag and bouncy particle samplers and compare it to the corresponding continuous time processes. (This is a joint work with Paul Dobson and Joris Bierkens.)

Andi Wang: Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods
In recent years there has been great interest in studying nonreversible MCMC schemes based on piecewise-deterministic Markov processes (PDMPs), since such methods avoid 'diffusive' behaviour which can plague their reversible counterparts. However, the theoretical analysis of such nonreversible methods is considerably more complicated, due to the lack of spectral theory in the nonreversible setting. The study of hypocoercivity is one technique to deliver theoretical high-dimensional guarantees for a range of such samplers, including the Zig-Zag Sampler and the Bouncy Particle Sampler. In my talk I will introduce the study of hypocoercivity, describe their application to PDMP-MCMC schemes, and then discuss a recent extension which covers the sub-geometric case for heavy-tailed target distributions.
ArXiv preprint: https://arxiv.org/abs/2011.09341

Invited Talk: 11:00 - 12:00

Randal Douc: The Kick-Kac teleportation algorithm: boost your favorite Markov Chain Monte Carlo using Kac formula
In this work, we propose to target a given probability measure π by combining two Markov kernels with different invariant probability measures. In its basic form, the mechanism consists in picking up the current position and moving it according to a π-invariant Markov kernel as soon as the proposed move does not fall into a predefined region. If this is the case, then we resort to the last position in this region and move it according to another auxiliary Markov kernel before starting another excursion outside the region with the first kernel. These state dependent interactions allow

to combine smoothly different dynamics that can be taylored to each region while the resulting process still targets the probability measure π thanks to an argument based on the Kac formula. Under weak conditions, we obtain the Law of Large numbers starting from any point of the state space, as a byproduct of the same property for the different implied kernels. Geometric ergodicity and Central Limit theorem are also established. Generalisations where the indicator function on the region target is replaced by an arbitrary acceptance probability are also given and allow to consider any Metropolis Hastings algorithm as a particular case of this general framework. Numerical examples, including mixture of Gaussian distributions are also provided and discussed.

POSTERS: Monday 5pm-7:30pm

Nicola Branchini: Optimized Auxiliary Particle Filters
Auxiliary particle filters (APFs) are a class of sequential Monte Carlo (SMC) methods for Bayesian inference in state-space models. In their original derivation, APFs operate in an extended state space using an auxiliary variable to improve inference. In this work, we propose optimized auxiliary particle filters, a framework where the traditional APF auxiliary variables are interpreted as weights in an importance sampling mixture proposal. Under this interpretation, we devise a mechanism for proposing the mixture weights that is inspired by recent advances in multiple and adaptive importance sampling. In particular, we propose to select the mixture weights by formulating a convex optimization problem, with the aim of approximating the filtering posterior at each timestep. Further, we propose a weighting scheme that generalizes previous results on the APF (Pitt et al. 2012), proving unbiasedness and consistency of our estimators. Our framework demonstrates significantly improved estimates on a range of metrics compared to state-of-the-art particle filters at similar computational complexity in challenging and widely used dynamical models.
ArXiv preprint: https://arxiv.org/abs/2011.09317

Daria Bystrova: Bayesian block-diagonal graphical models via the Fiedler prior
We consider the problem of inferring the conditional independence structure between the entries of a Gaussian random vector. Our focus is on finding groups of independent variables. This can be translated into the estimation of a precision matrix (inverse of the covariance matrix) with a block-diagonal structure. We borrow ideas from spectral graph theory and propose a novel prior called Fiedler prior based on the spectrum of a transformation of the precision matrix. This prior shows shrinkage properties and is particularly useful for estimating sparse precision matrices with block-diagonal structure. We compare the shrinkage induced by our prior and the popular Graphical Lasso prior. We evaluate the performance on simulated datasets.

Chris Carmona: Semi-Modular Inference: enhanced learning in multi-modular models by tempering the influence of components
Bayesian statistical inference loses predictive optimality when generative models are misspecified. Working within an existing coherent loss-based generalisation of Bayesian inference, we show existing Modular/Cut-model inference is coherent, and write down a new family of Semi-Modular Inference (SMI) schemes, indexed by an influence parameter, with Bayesian inference and Cut-models as special cases. We give a meta-learning criterion and estimation procedure to choose the inference scheme. The framework applies naturally to Multi-modular models. Cut-model inference allows directed information flow from well-specified modules to misspecified modules, but not vice versa. An existing alternative power posterior method gives tunable but undirected control of information flow, improving prediction in some settings. In contrast, SMI allows tunable and directed information flow between modules. We illustrate our methods on several real datasets from the literature, and show how to target the SMI posterior either via MCMC or a novel Variational-SMI via normalising flows.

Saint-Clair Chabert-Liddell: Learning common structures in a collection of networks with the Stochastic Block Model
Observons une collection de réseaux ayant la même loi d'émission et dont on suppose qu'ils ont une structure commune. Le modèle à blocs stochastiques (SBM) est un modèle modèle probabiliste qui suppose l'existence de variables latentes représentant les classes des nœuds du réseau et qui permet via ses paramètres une description succincte de la structure du réseau à l'échelle mésoscopique. Sur de petits réseaux cette structure mésoscopique est sensible à de faibles variations d'échantillonnage, de plus ces réseaux possèdent une taille est une densité différente rendant la comparaison de leurs structures difficiles. Nous étendons le SBM pour modéliser conjointement une collection de réseaux. Ces réseaux sont supposés indépendants mais possèdent la même struc-

ture mésoscopique à un facteur de densité près. Les paramètres de notre modèle sont estimés par approximation de la vraisemblance via des méthodes variationnelles (algorithme VEM). L'existence d'un bon compromis entre les structures mésoscopiques de ces réseaux n'est pas garantie. Un critère de vraisemblance complète intégrée (ICL) nous permet de sélectionner non seulement le nombre de blocs mais également d'établir la pertinence de l'utilisation de cette structure commune.

Ryan Chan: Hierarchical Monte Carlo Fusion
The Monte Carlo fusion algorithm provides theory and methodology to carry out balanced inferences for the unification of distributed data analyses. This fusion problem arises in many settings, such as expert elicitation and distributed 'big data' problems where traditional Markov chain Monte Carlo methods can scale poorly with increasing amounts of data. While the Monte Carlo fusion algorithm was the first general approach which avoids any form of approximation error in obtaining the unified inference, it can become inefficient in several settings. We present several new approaches with the aim to improve the efficiency and robustness of the Monte Carlo fusion framework.

Stefan Franssen: Bayesian imaging with data-driven priors encoded by neural networks
In the past 5 years there has been a breakthrough in our understanding in the behaviour of (sparse) Deep Neural network regression. For β-Hölder spaces, Schmidt-Hieber gave near minimax convergence rates, and the work has been extend to Besov spaces by Suzuki. These works give guarantees for the square loss of (near) minimizers of the empirical square loss, which imply that Deep Neural Networks following their designs will have good uncertainty quantification. In spite of this progress, there has not been any rigorous way of quantifying uncertainty in the estimates of Deep Neural Networks. We provide both an Empirical Bayesian methodology to provide uncertainty quantification and a theoretical analysis with frequentist coverage guarantees. We also ran a simulation study which illustrates the coverage properties.

Adrien Hairault: Marginal likelihood estimation for Dirichlet Process Mixture models
The Dirichlet Process Mixture (DPM) model, first introduced by Ferguson[1983], has become one of the main tools in the field of Bayesian non-parametrics. Its range of application is broad, from multivariate clustering to Bayesian density estimation. MCMC algorithms for fitting such semi-parametric models are well established (see e.g Neal [2000] and Jainet al. [2007]), but efficient ways of comparing these models to competing semi-parametric and parametric models has been hardly explored so far. Although the Bayesian paradigm offers a straightforward way of comparing any model to a semi-parametric one through the Bayes Factor, the difficulty in computing the marginal likelihood of the latter prevents practitioners from using this approach. Hence, it is a common practice to use the DPM model without considering alternative parametric or semi-parametric models, although it may not always be appropriate. For instance, Miller and Harrison [2014] show that for data assumed to come from a finite mixture with an unknown number of components, the DPM posterior on the number of clusters in the observed data does not converge to the true number of components. While there exist an abundant literature and well-established methods for computing the evidence of parametric models, equivalent methods for non-parametric models such as the DPM model are far from numerous. To our knowledge, only Basu and Chib [2003] address the issue of evidence computation for semi-parametric Bayes by extending the well-established method of Chib [1995]. Nevertheless, this algorithm cannot be easily adapted to the non-conjugate case, and we find that it suffers from poor mixing as the number of observations grows. To overcome these drawbacks, we explore Sequential Monte Carlo (SMC) as an alternative approach. Recently, Griffin [2017] has given a general method for applying SMC to the DPM model using data tempering but focuses on parameter estimation rather than evidence computation. We aim at comparing these two Monte-Carlo methods for marginal likelihood estimation in DPM models, which, to our knowledge, has not yet been done. As benchmark, we also use other estimators such as the Arithmetic Mean Estimator (AME) and the Harmonic Mean Estimator (HME).

Diala Hawat: A new Python toolbox for approximating the structure factor of a point process and testing its hyperuniformity.
Bayesian inference requires computing integrals, and Monte Carlo methods have been developed for that purpose. We consider point processes that could form a new family of Monte Carlo quadratures: hyperuniform (HU) point processes. HU processes are random sets of points such that the variance of the number of points that fall in a large ball is much lower than the volume of that ball, with a ratio going to zero. By definition, HU point processes are thus more efficient than rejection sampling at estimating the volume of a set. There are many candidate HU processes in the physics literature, but rigorously proving that a point process is HU is usually difficult. It is thus desirable to have numerical tests of hyperuniformity. Common practice in statistical physics is to estimate a spectral measure called the structure factor, the behaviour of which around zero is a sign of hyperuniformity. We survey existing estimators of the structure factor and gather them all in a Python toolbox, along with numerical diagnostics of hyperuniformity.

Matthew Holden: Bayesian imaging with data-driven priors encoded by neural networks
We propose a new methodology for performing Bayesian inference in imaging inverse problems where the prior knowledge is available in the form of training data. Following the manifold hypothesis and adopting a generative modelling approach, we construct a data-driven prior that is supported on a sub-manifold of the ambient space, which we can learn from the training data by using a variational autoencoder or a generative adversarial network. We establish the existence and well-posedness of the associated posterior distribution and posterior moments under easily verifiable conditions, providing a rigorous underpinning for Bayesian estimators and uncertainty quantification analyses. Bayesian computation is performed by using a parallel tempered version of the preconditioned Crank-Nicolson algorithm on the manifold, which is shown to be ergodic and robust to the non-convex nature of these data-driven models. In addition to point estimators and uncertainty quantification analyses, we derive a model misspecification test to automatically detect situations where the data-driven prior is unreliable, and explain how to identify the dimension of the latent space directly from the training data. The proposed approach is illustrated with a range of experiments with the MNIST dataset, where it outperforms alternative image reconstruction approaches from the state of the art. A model accuracy analysis suggests that the Bayesian probabilities reported by the data-driven models are also remarkably accurate under a frequentist definition of probability.

Yoann Jayer: Sampling from Coulomb and log-gases with irreversible continuous MCMC methods
Coulomb and log-gases are exchangeable singular Boltzmann–Gibbs measures appearing in mathematical physics in many places and, in particular, in random matrix theory. They can naturally be interpreted as the equilibrium measure of a system of repulsive particles confined by an external potential. As the repulsive interaction between particles are in this case singular, the difficulty is to find a sampler which is both stable and with controlled rejection steps . For this purpose, we propose a MCMC sampler based on a class of continuous irreversible Markov process called Piecewise Deterministic Markov Processes. This sampler has the advantages of being rejection-free and of adapting naturally to the scaling of the problem. We prove, under some assumptions, the invariance and geometrical ergodicity of the sampler and tests its good performances on examples in mathematical physics.

Geerten Koers: On the non-i.i.d. misspecified Bernstein-Von Mises theorem
The asymptotic behaviour of misspecified posterior distributions is considered in a non-i.i.d. parametric setting. It is shown that a misspecified Bernstein-Von Mises theorem holds, and conditions on the distribution of the data and the likelihood functions of the model are relaxed compared to earlier results. The asymptotic behaviour of the well-specified posterior distribution is compared to that of the misspecified posterior distribution in a non-Gaussian model approximated by Gaussian

likelihoods. Under regularity conditions, the misspecified posterior distribution will concentrate on the true parameter in these models. Natural examples in PDE-theory of models that were not covered by existing literature are analysed. The numerical analysis shows that the misspecified posterior distribution has an incorrect uncertainty quantification, and it is observed that the resulting credible sets over-cover compared to the credible sets coming from the well-specified posterior distribution.

Alice Martin: Backward importance sampling for online estimation of state space models
This paper proposes a new Sequential Monte Carlo algorithm to perform maximum likelihood estimation in partially observed dynamical systems whose dynamics is unknown. Training such generative models and obtaining low variance estimators of the posterior distributions of the unobserved states given the observations is challenging as the transition densities of the states cannot be evaluated point-wise. In this paper, a backward importance sampling step is introduced to estimate such posterior distributions instead of the acceptance-rejection procedure used in the usual approach. In the context of multivariate stochastic differential equations, the proposed algorithm makes use of unbiased estimates of the unknown transition densities under much weaker assumptions than standard alternatives. The performance of this estimator is assessed in the case of a partially observed stochastic Lotka-Volterra model and stochastic recurrent neural networks.

Alessandro Mascaro: A Bayesian approach to causal discovery from soft interventions
Directed Acyclic Graphs (DAGs) are used for encoding conditional independencies in multivariate data as well as for representing causal relations among variables. Causal discovery is the process of identifying the DAG representing the causal structure underlying a sample of data. If the data is only observational and not experimentally designed, the DAG is identifiable only up to its Markov-equivalence class. However, if in addition one uses interventional data, i.e. data in which the causal mechanism generating the data has been altered by an external intervention, one can reduce the size of the original Markov equivalence class and identify the right subclass of Interventional (I-) Markov equivalent DAGs. One can envisage hard or soft interventions, the difference being that the former break the connections between the intervened variables and their parents, while the latter merely perturb the causal mechanism, keeping causal connections intact. In this talk I will propose a Bayesian approach to causal discovery from observational and experimental Gaussian data in which soft interventions are performed. The approach proposed has the substantial advantage of naturally accounting for the uncertainty on the space of DAGs. An extensive comparison with alternative state-of-the-art methodologies is carried out through a simulation study.

Hector McKimm: Hamiltonian Jump Processes Adjusted with Regenerations
We consider enriching a jump process, with discrete-time Markov transition kernel corresponding to approximate Hamiltonian dynamics, with regenerations. The resulting process is non-reversible; when used for Monte Carlo integration, the regenerative structure of the process guarantees it has the correct invariant distribution, so no Metropolis-Hastings accept-reject step is required. The sampler benefits from the efficiency of Hamiltonian dynamics, as well as the advantages of regenerative simulation — including the potential for mode-jumping moves, parallel implementation and an absence of burn-in period. Furthermore, it is possible to recover the normalizing constant of the target density, at no additional computational cost.

Théo Moins: Improving MCMC convergence diagnostic with a local version of $\hat{R}$
Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis. Among the most popular methods, the potential scale reduction factor (commonly named $\hat{R}$) is an indicator that monitors the convergence of all chains to the stationary distribution, based on a comparison of the between- and within-variances of the chains. Several improvements have been suggested since its introduction by Gelman and Rubin (1992). Here, we analyse some properties of the theoretical value R associated to $\hat{R}$ in the case of a localized version that focuses on quantiles

of the distribution. This leads to proposing a new indicator, which is shown to allow both for localizing the MCMC convergence in different quantiles of the distribution, and at the same time for handling some convergence issues not detected by other $\hat{R}$ versions.

Daniel Moss: Efficient estimation and use of cut posterior in semiparametric hidden Markov models
We consider the problem of estimation in Hidden Markov models with finite state space and nonparametric emission distributions. Efficient estimators for the transition matrix are exhibited, and a semiparametric Bernstein-von Mises result is deduced, extending the work of Gassiat et al. (2018) to the HMM setting. Following from this, we adapt the cut posterior approach of Jacob et al. (2017) to jointly estimate the transition matrix and the emission distributions. We also derive a general theorem on contraction rates for such cut posterior approaches, in the spirit of the seminal work of Ghosal and van der Vaart (2007). We then show how the result may be applied to obtain a contraction rate result for the emission distributions in our setting. Finally, simulation studies are provided to illustrate our theoretical results.

Dennis Nieman: Contraction rates of variational Bayes for regression with Gaussian process priors
We consider the Bayesian nonparametric regression model with Gaussian process priors. In practice, sampling from the exact posterior distribution is computationally expensive. We study an approximative procedure called the variational method, which reduces computation time. Of particular interest is a variational framework that has gained popularity in the machine learning literature in the last decade. We investigate frequentist properties of the Bayesian approach: the data are assumed to be generated from a distribution with a true functional parameter, and conditions are given under which the contraction rate does not deteriorate under the approximation. The developed theory is applied to several examples of Gaussian process priors.

Ardjen Pengel: Strong Invariance Principles for Markov Processes
The formulation of a strong invariance principle for a stochastic process asserts that we can quantify the error term of the Brownian approximation in the functional central limit theorem. While these results have many applications, few results for continuous-time settings are available. In this paper, we obtain strong approximation results for Markov processes. Through an application of Nummelin splitting for continuous-time processes by Löcherbach and Loukianova [Stochastic Processes and their Applications, (2008) 1301-1321], we derive a strong invariance principle for a large class of Markov processes. The resulting approximation error is close to the optimal Komlós-Major-Tusnády bound. Furthermore, we show that the required assumptions for this approach can be simplified to more natural ergodicity and moment conditions. Finally, we demonstrate the use of these strong approximation results, by showing that they can be used to analyse spectral variance and batch means estimators of the asymptotic variance of continuous-time Monte Carlo samplers.

Thibault Randrianarisoa: Smoothing and adaptation of shifted Pólya Tree ensembles
Recently, S. Arlot and R. Genuer have shown that a random forest model outperforms its single-tree counterpart in estimating α-Holder functions, $1 \leq \alpha \leq 2$. This backs up the idea that ensembles of tree estimators are smoother estimators than single trees. On the other hand, most positive optimality results on Bayesian tree-based methods assume that $\alpha \leq 1$. Naturally, one wonders whether Bayesian counterparts of forest estimators are optimal on smoother classes, just as observed with frequentist estimators for $\alpha \leq 2$. We focus on density estimation and introduce an ensemble estimator from the classical (truncated) Pólya tree construction in Bayesian nonparametrics. Inspired by the work mentioned above, the resulting Bayesian forest estimator is shown to lead to optimal posterior contraction rates, up to logarithmic terms, for the Hellinger and L1 distances on probability density functions on $[0; 1)$ for arbitrary Holder regularity $\alpha > 0$. This improves upon previous results for constructions related to the Pólya tree prior, whose optimality

was only proven when $\alpha \leq 1$. Also, by adding a hyperprior on the trees' depth, we obtain an adaptive version of the prior that does not require α to be specified to attain optimality.
ArXiv preprint: https://arxiv.org/abs/2010.12299

Louis Rouillard: ADAVI: Automatic Dual Amortized Variational Inference Applied To Pyramidal Bayesian Models
Frequently, population studies feature pyramidally-organized data represented using Hierarchical Bayesian Models (HBM) enriched with plates. These models can become prohibitively large in settings such as neuroimaging, where a sample is composed of a functional MRI signal measured on 64 thousand brain locations, across 4 measurement sessions, and at least tens of subjects. Even a reduced example on a specific cortical region of 300 brain locations features around 1 million parameters, hampering the usage of modern density estimation techniques such as Simulation-Based Inference (SBI). To infer parameter posterior distributions in this challenging class of problems, we designed a novel methodology that automatically produces a variational family dual to a target HBM. This variatonal family, represented as a neural network, consists in the combination of an attention-based hierarchical encoder feeding summary statistics to a set of normalizing flows. Our automatically-derived neural network exploits exchangeability in the plate-enriched HBM and factorizes its parameter space. The resulting architecture reduces by orders of magnitude its parameterization with respect to that of a typical SBI representation, while maintaining expressivity. Our method performs inference on the specified HBM in an amortized setup: once trained, it can readily be applied to a new data sample to compute the parameters' full posterior. We demonstrate the capability of our method on simulated data, as well as a challenging high-dimensional brain parcellation experiment. We also open up several questions that lie at the intersection between SBI techniques and structured Variational Inference.

Achille Thin: NEO: Non Equilibrium Sampling on the Orbit of a Deterministic Transform
Variational auto-encoders (VAE) are popular deep latent variable models which are trained by maximizing an Evidence Lower Bound (ELBO). To obtain tighter ELBO and hence better variational approximations, it has been proposed to use importance sampling to get a lower variance estimate of the evidence. However, importance sampling is known to perform poorly in high dimensions. While it has been suggested many times in the literature to use more sophisticated algorithms such as Annealed Importance Sampling (AIS) and its Sequential Importance Sampling (SIS) extensions, the potential benefits brought by these advanced techniques have never been realized for VAE: the AIS estimate cannot be easily differentiated, while SIS requires the specification of carefully chosen backward Markov kernels. In this paper, we address both issues and demonstrate the performance of the resulting Monte Carlo VAEs on a variety of applications.

Lasse Vuursteen: Exploring optimal distributed testing in high-dimensional Gaussian models using Bayesian techniques
We study the problem of signal detection in Gaussian noise in a distributed setting. We derive a lower bound on the size that the signal needs to have in order to be detectable. Moreover, we exhibit optimal distributed testing strategies that attain the lower bound. In this talk, we will go into Bayesian techniques that allow us to derive testing lower bounds, but also extend to methods achieving these optimal rates. In particular, we will exhibit that priors that are difficult to distinguish from a "no signal" null hypothesis, can at the same time provide likelihood ratio tests are optimal distributed tests.
ArXiv preprint: https://arxiv.org/abs/2012.04957