MINISYMPOSIUM

COMPUTATIONAL STATISTICS AND BAYESIAN INFERENCE FOR STOCHASTIC BIOLOGICAL MODELS

Organizer

Massimiliano Tamborrino

Institute for Stochastics

Johannes Kepler University Linz

Altenbergerstrasse 69, 4040 Linz

Austria

massimiliano.tamborrino@jku.at

Co-organizer

Evelyn Buckwar

Institute for Stochastics

Johannes Kepler University Linz

Altenbergerstrasse 69, 4040 Linz

Austria

evelyn.buckwar@jku.at

Minisymposium Keywords: Bayesian Statistics, Mathematical Neuroscience,

The last decades have been characterised by a notable progress in experimental equipments, technologies and recording techniques, allowing the recording of enormous data sets in several fields of biological applications. In neuroscience, over the last five decades, the number of simultaneously recorded neurons has doubled approximately every 7 years [1], with more than 700 well-isolated single neurons recently recorded from five brain structures in an awake mouse thanks to a new developed silicon probe [2]. Another well known neuronal technique is electroencephalography (EEG), an electrophysiological monitoring method to record electrical activity of the brain, producing massive sets of data. Similarly, advanced experimental techniques have been developed for studying protein folding, a spontaneous process that transforms a disordered polymer into a specific three-dimensional structure, with incorrect foldings associated to a wide range of human diseases. Protein folding data consists of time series with even more than millions of observations.

Such exponential growth in the amount of data was at the origin of a breathtaking expansion of computational statistics, a branch of statistics combining both methodology and algorithms, with terms such as "big data", "data science"and "machine learning"that became more and more familiar in the news and among non-specialist audiences [3]. Newly developed statistical methods should then consider both the computational costs and the potential for more accurate mathematical models for the analysis of biological data. Due to the increasing complexity in the models of interest, the underlying likelihood function, that plays a key role in modern statistics, often cannot be numerically evaluated, requiring the development of innovative likelihood-free statistical techniques, such as Approximate Bayesian Computation (ABC) [4].

The proposed Minisymposium has three main goals. First, we want to present statistical techniques whose use will provide further insights in different biological areas. Second, we want to highlight the key role played by computational statistics and Bayesian statistics for the inference of stochastic biological models. Third, we want to emphasise the existing strong connection and influence between numerical analysis and computational statistics, and the importance of developing suitable (i.e. efficient, convergent, etc.) algorithms when performing statistical inference. We will achieve these goals through the proposed five presentations. All speakers have a mathematical and statistical background, share an expertise in stochastic processes, Bayesian inference and computational statistics, and an interest in tackling statistical problems arising from several fields of application, in particular from biology, epidemiology and neuroscience (both at single and neural network level). In particular, ABC methods, accelerated Markov Chain Monte Carlo algorithms, Bayesian Inference, Bayesian hierarchical modelling approaches and structural preserving numerical methods will be presented, motivated by the fit of single level neuronal data (spike trains),

EEG data, protein folding data and parameters from multi-compartment models. The proposed talks tackle some of the hottest statistical issues, and present novel techniques that can also be applied to any stochastic model yielding the same kind of data.

Minisymposium: Computational statistics and Bayesian inference for stochastic biological models

LIKELIHOOD-FREE METHODS FOR THE INFERENCE OF NON-RENEWAL POINT PROCESSES ARISING FROM NEUROSCIENCE

Massimiliano Tamborrino

massimiliano.tamborrino@jku.at

Institute for Stochastics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

Joint work with Adeline Leclercq-Samson (Department of Probability and Statistics, Universite Grenoble Alpes, 700 Avenue Centrale, 38401 Saint Martin d'Hères, France)

Keywords: Partially observed stochastic processes, Mathematical Neuroscience, Multi-timescale adaptive threshold models, Approximate Bayesian Computation, Non-renewal point processes.

In many signal-processing applications, it is of primary interest to decode or reconstruct the unobserved signal based on some partially observed information. Some examples are all type of recognition (e.g. automatic speech, face, gesture, handwriting), genetics, genomics and neuroscience (ion channels modelling). From a mathematical point of view, this corresponds to estimate model parameters of an unknown coordinate based on discrete observations of one or more other coordinates.

Here we consider a bivariate stochastic process where available observations are hitting times of one coordinate to the other, and discuss it in the framework of stochastic modelling of single neuron dynamics. The considered multi-timescale adaptive threshold model is not simply an ad-hoc model, but can be derived from the detailed Hodgkin-Huxley model, can accurately predict spike times and incorporate the effects of slow K+ currents, usually mediating adaptation [5]. When performing statistical inference of the underlying model parameters, four difficulties arise: none of the two model components is directly observed; the considered process is not of hidden Markov model type; the underlying likelihood is unknown/intractable; consecutive hitting times are neither independent nor identically distributed. We tackle these statistical issues by considering Approximate Bayesian Computation [4], a likelihood-free method requiring the development of suitable distance criteria to apply, e.g., in an algorithm similar to acceptance-rejection. After presenting the method and proposing several possible distances, I illustrate how to use it on the considered model.

Minisymposium: Computational statistics and Bayesian inference for stochastic biological models

STRUCTURAL PRESERVING NUMERICAL METHOD AND BAYESIAN INFERENCE FOR STOCHASTIC NEURONAL MASS MODELS

IRENE TUBIKANEC

irene.tubikanec@jku.at

Institute for Stochastics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

Joint work with Evelyn Buckwar (Institute for Stochastics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria) and Massimiliano Tamborrino (Institute for Stochastics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria)

The goal of our talk is the statistical inference, using approximate Bayesian computation (ABC) [4], of stochastic neural mass models. These models describe the average properties of the electrical activity of a whole population of neurons, and have been reported to reproduce for example Electroencephalography (EEG), Magnetoencephalography (MEG) and Stereo -Electroencephalography (SEEG) data. Here we focus on a specific reformulation of the Jansen and Rit neural mass model [6] as a stochastic differential equation (SDE) with additive noise [7]. We analyse this new stochastic model through its dynamical and structural properties. In [7], the authors showed that the distribution of the 6-dimensional solution process $X(t) = (X_0(t), \ldots, X_5(t))^T$, $t \in [0, T]$ converges exponentially fast towards a unique invariant measure. Moreover, they developed an efficient numerical splitting method that, differently from standard numerical methods, e.g. Euler's method, preserves this stationary structural model property.

We are interested in the statistical inference of this stochastic model, where we make use of the preservation of the parameter dependent invariant measure. In particular, our goal is the estimation of two important model parameters that are mainly responsible for changes in the structure of the system. The first parameter of interest is the internal connectivity constant C, controlling how the pyramidal cells interact with the excitatory and inhibitory interneurons. The second parameter is the drift μ of one of the six components, describing excitatory inputs from neighbouring or more distant columns. From an experimental point of view, the process $X(t)$ is only partially observed through the EEG-related process $Y(t) = X_1(t) - X_2(t)$, $t \in [0, T]$, making the statistical inference more challenging. Two main difficulties arise. First, due to the fact that the non-linear and multi-dimensional SDE cannot be explicitly solved, the dynamics of the signal process $Y(t)$ can be only computed through the numerical scheme. Second, the corresponding likelihood function is intractable. We tackle this by considering the likelihood-free and simulation based ABC approach. This is a Bayesian technique that necessitates plenty of synthetic data simulations from the original model. We perform these simulations using the structure-preserving and efficient numerical splitting method. The crucial part is then to define reliable distance criteria to compare

the simulated synthetic data with the observed reference data. Here we propose to not directly compare the signal data, but to calculate distances between their related invariant densities, that we approximate with a kernel density estimation approach that uses only one single path of the process $Y(t)$. Using the parameter dependent structural property of the system, the ABC approach, combined with the adopted numerical splitting method, is able to provide satisfactory estimates of the parameters of interest. Our satisfactory results on simulated reference data make the proposed statistical method a good candidate for the fit of the complex stochastic model to EEG measurements.

Minisymposium: Computational statistics and Bayesian inference for stochastic biological models

BAYESIAN SPECTRAL ESTIMATION METHODS FOR MULTIPLE TIME SERIES AND APPLICATION TO MULTICHANNEL EEG

ANNALISA CADONNA

acadonna@wu.ac.at

Institute for Statistics and Mathematics, Vienna University of Economics and Business, Welthandelsplatz 1,1020 Vienna, Austria

Joint work with Raquel Prado (Department of Applied Mathematics and Statistics, University of California, 1156 High St, 95064 Santa Cruz , California) and Athanasios Kottas (Department of Applied Mathematics and Statistics, University of California, 1156 High St, 95064 Santa Cruz , California)

Keywords: Local Gaussian mixtures, Multi-channel electroencephalography, Log-periodogram, Markov Chain Monte Carlo.

We present a novel Bayesian hierarchical modelling approach to spectral density estimation for multiple time series based on local mixtures of Gaussian distributions. The proposed method presents modelling and computational advantages. In fact, the model is easy to interpret and it allows for quantification of uncertainty. Moreover, it is suitable to data augmentation algorithms for posterior simulation.

The key motivating application is the analysis of multichannel electroencephalographic recordings (EEGs). EEGs record measurements of electrical potential fluctuations at multiple locations on the scalp of a human subject. Identifying which locations lead to electrical brain signals with similar spectral densities, and grouping them based on common spectral features, is particularly meaningful, as it provides insights about the physiological state of the subject and about the spatial structure of cortical brain activity under certain experimental or clinical conditions.

In the proposed model, the log-periodogram of each series is modelled as a mixture of Gaussian distributions with frequency-dependent weights and mean functions. Hence, the implied model for the logspectral density is a mixture of mean functions with frequency-dependent weights. The mixture weights are built through successive differences of a logit-normal distribution function with frequency-dependent parameters. The multiple time series are modelled hierarchically through hyper-priors on the weight parameters, which allow to borrow information across time series. In addition to accommodating flexible spectral density shapes, a practically important feature of the proposed formulation is that it allows for straightforward posterior simulation through a Gibbs sampling algorithm.

Minisymposium: Computational statistics and Bayesian inference for stochastic biological models

BAYESIAN ESTIMATION OF DIFFUSIONS APPLIED TO COMPARTMENT MODELS

MORITZ SCHAUER

m.r.schauer@math.leidenuniv.nl

Mathematical Institute, University of Leiden, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

Joint work with Frank van der Meulen (Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4 2628 CD Delft, The Netherlands)

Keywords: Compartment models, Parameter estimation, Diffusion processes, FitzHugh-Nagumo model, Indirect observations.

A biological system can often be decomposed into a number of homogeneously filled "compartments"and the interactions between them. The compartments need not correspond to a physical space, for example they could represent different species of animals and plants. Interaction between different compartments happen through the exchange of objects or the transformation of material. If those changes happen continuously and are influenced by chance, then a diffusion process can be a good mathematical representation of the model. Examples are SIR-models for infectious disease or biological neuron models. In this talk I consider the problem of recovering information about the interactions from noisy observations of (some of) the compartment populations at discrete times.

References

[1] I. H. Stevenson, K. P. Kording. (2011). *How advances in neural recording affect data analysis*. Nat Neurosci., 14(2): 139–142.

[2] J.J. Jun et al. (2017). *Fully integrated silicon probes for high-density recording of neural activity*. Nature 551, 232–236.

[3] B. Efron, T. Hastie. *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press, Cambridge, 2016.

[4] M.A. Beaumont, W. Zhang, D.J. Balding. (2002). *Approximate Bayesian computation in population genetics*. Genetics, 162(4): 2025–2035.

[5] R. Kobayashi, K. Kitano. (2016) *Impact of slow K^+ currents on spike generation can be described by an adaptive threshold model*. J. Comput. Neurosci., 40(3), 347–362.

[6] B.H. Jansen, V.G. Rit. (1995). *Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns*. Biol. Cybern., 73(4): 357–366.

[7] M. Ableidinger, E. Buckwar, H. Hinterleitner. (2017). *A Stochastic version of the Jansen and Rit neural mass model: analysis and numerics*. J. Math. Neurosc. 7(8).

[8] J. A. Christen, C. Fox. (2005). *Markov chain Monte Carlo using an approximation*. J. Comput. Graph. Stat., 14:795–810.