Contributed talks

Syed Murtuza Baker (Leibniz Institute of Plant Genetics and Crop Plant Research)

A parameter estimation framework for kinetic models of biological systems

In order to obtain the correct predictive behaviour of a kinetic model, it is crucial to have an accurate and complete set of parameter values. Lack of information regarding these parameters from wet lab experiments has held back the successful use of such models. Therefore these parameters have to be estimated computationally to feature a complete description of the model. In this thesis I propose a novel parameter estimation framework combining different existing approaches with a newly proposed filtering technique for the successful estimation of these unknown parameter values. The framework includes a constrained extension of the square-root unscented Kalman filter to estimate the parameter values within a biologically meaningful parameter space. This framework is capable of addressing both the issues of structural and practical non-identifiablity before performing the final estimation of the parameters. The constrained square-root unscented Kalman filter (CSUKF), guarantees numerical stability by ensuring positive definiteness of the covariance matrix. The CSUKF takes into consideration one of the common features of biological models, noise. Noise is introduced in two ways, in the system due to the uncertainty in the model and in the measurement data due to the inaccuracy in the method or device used to collect the data. By representing the dynamic system as a state space model, the CSUKF jointly estimate the states and the parameters of the non-linear dynamic systems. This makes it possible for the CSUKF to estimate both the parameter values and the hidden variables. CSUKF uses the general probability theory to estimate the parameter values of biological models where reasoning under uncertainty is essential. An identifiability analysis module is included in the framework to identify the non-identifiable parameters. Wherever possible the problem of non-identifiability is resolved through additional, and/or more accurate, measurement data. To assist in resolving the issue, the framework includes ranking of the parameters, determination of the correlation and functional relationship of non-identifiable parameters with other parameters. Finally, when it is not possible to solve the parameter non-identifiability through standard methods, the informed prior is formulated for the unique estimation of parameters even in the presence of non-identifiability. This framework is successfully applied to estimate parameters for three published biological models, the glycolysis model in yeast, the sucrose accumulation model in sugarcane culm tissue and a gene regulatory network.

Christiane Fuchs (Helmholtz Zentrum München)

Bayesian inference on diffusion models for protein dynamics

Stochastic differential equations (SDEs) are a promising instrument to realistically model the time-continuous evolution of natural phenomena in life sciences. The solutions of such SDEs are given by diffusion processes. Statistical inference for diffusions, however, proves to be challenging in practice as the likelihood function is typically intractable. This talk explains and further develops a well-known Bayesian approach which introduces auxiliary observations by means of Markov chain Monte Carlo (MCMC) techniques. This procedure originally suffers from convergence problems which stem from a deterministic link between the model parameters and the quadratic variation of a continuously observed diffusion path. This talk shows a neat modification of the above approach for general multi-dimensional diffusions and provides the mathematical and empirical proof that the so-constructed MCMC scheme converges. The potential of the estimation method is demonstrated using the example of single-cell protein dynamics.

Javier Gonzalez (University of Groningen), Ivan Vujacic and Ernst Wit

Reproducing kernel Hilbert space based estimation of systems of ordinary differential equations

Non-linear systems of differential equations have attracted the interest of researchers in fields like System Biology, Ecology or Biochemistry, due to their flexibility and their ability to describe dynamical systems. Despite the importance of such models in many branches of science they have not been the focus of systematic statistical analysis until the last few years. In this work we propose a general approach to estimate the parameters of systems of differential equations measured with noise. Our methodology is based on the maximization of a penalized likelihood where the differential system of equations is used as a penalty. To do so, we use a Reproducing Kernel Hilbert space approach that allows us to formulate the estimation problem as an unconstrained numeric maximization problem easy to solve. The proposed method is tested in real and simulated examples showing its utility in a wide range of scenarios.

Boumediene Hamzi (Imperial College London)

On parameter estimation of nonlinear stochastic differential equations in reproducing kernel Hilbert spaces

We introduce a data-based approach to estimating key quantities which arise in the study of nonlinear control systems and random nonlinear dynamical systems. Our approach hinges on the observation that much of the existing

linear theory may be readily extended to nonlinear systems - with a reasonable expectation of success - once the nonlinear system has been mapped into a high or infinite dimensional feature space. In particular, we develop computable, consistent estimators approximating controllability and observability energy functions for nonlinear systems, and study the ellipsoids they induce. In all cases the relevant quantities are estimated from simulated or observed data. It is then shown that the controllability energy estimator provides a key means for approximating invariant measures of stochastically forced nonlinear systems. This is joint work with J. Bouvrie (Duke University).

Sabine Hug (Helmholtz Zentrum München)

Bayesian model selection validates a biokinetic model for zirconium processing in humans

Bayesian model selection methods can help to analyze multi-compartmental models based on ordinary differential equations (ODEs), for example biokinetic models. In the field of radiation protection, such multi-compartmental biokinetic models are the tool of choice for simulating the processing of radioactive substances in the human body. Although easily interpretable, determining the exact compartment structure and interaction mechanisms, i.e. the exact dynamical system, is generally daunting.

We here applied a Markov Chain Monte Carlo (MCMC) approach called thermodynamic integration in combination with a recently developed copula based Metropolis-Hastings sampler to compute Bayes factors for two competing compartmental models for the processing of zirconium in the human body after ingestion. Bayes factors naturally prevent overfitting and provide evidence for any of the two models, even if they are non-nested. Thermodynamic integration uses a smooth transition from the prior to the posterior to stably evaluate the Bayes factor. By applying our method, we could use in vivo measurement data of plasma and urine to evaluate a new model put forward by the Helmholtz Zentrum München (HMGU) against the standard model by the International Commission on Radiological Protection (ICRP).

Jonathan Jaeger (ISBA-UCL) and Philippe Lambert

Bayesian ODE-penalized B-spline model with Gaussian mixture as error distribution

Ordinary differential equations (ODEs) are frequently used to model physical, chemical and biological processes. Currently, the most commonly used estimation procedures rely on nonlinear least squares. These approaches are computationally intensive and often poorly suited for statistical inference.

Alternative estimation methods of the state functions and the ODE parameters

were proposed in Ramsay et al. (2007). It may be viewed as a generalization of the P-spline theory Eilers and Marx (1996) that involves some basis function expansion of each state function and a penalty term expressed using the set of differential equations. Jaeger and Lambert (2011) adapts this approach to a full Bayesian ODE-penalized B-spline approach when the ODEs are affine and the data distribution is assumed Gaussian. The two major drawbacks of the frequentist ODE-penalized smoothing approach are overcome in the Bayesian framework: the selection of the ODE-adhesion parameter is now automatic and uncertainty measures about parameters can simply be obtaining using MCMC. In addition, the possible use of prior information about the ODE parameters is a definite advantage.

The assumption of a Gaussian data distribution is most of the time inappropriate but is very convenient as it enables to marginalize the joint posterior distribution with respect to the spline coefficients and therefore to get rid simply of the inconvenient posterior correlation between the spline coefficients and the ODE parameters. To overcome this limitation, we model homogeneous non-normal data distribution using finite mixture of Gaussian distributions by adapting the approach of Komárek and Lesaffre (2007).

In this talk, we present a fully Bayesian approach to jointly estimate parameters and state function of affine ODE models when the data distribution is homogeneous non-Gaussian. The strategies used to explore the joint posterior distribution with MCMC are presented. Some simulations comparing the performance of the basic Bayesian ODE-penalized approach to the Bayesian Gaussian mixture model are then given. We conclude the presentation with two applications: the selection of competing ODE models using the ODE-adhesion parameters and the analysis of pharmacokinetic data.

References

P Eilers and B Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121, 1996.

J Jaeger and P Lambert. Bayesian generalized profiling estimation in hierarchical linear dynamic systems. IAP Technical Report 11001, 2011.

A Komárek and E Lesaffre. Bayesian accelerated failure time model for correlated interval-censored data with a normal mixture as error distribution. *Statistica Sinica*, 17:549–569, 2007.

J O Ramsay, G Hooker, D Campbell, and J Cao. Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society, Series B*, 69:741–796, 2007.

Anders Jensen (University of Copenhagen)

A Markov Chain Monte Carlo approach to parameter estimation in the FitzHugh-Nagumo model

For all but a few diffusion models an explicit expression for the transition density, and thus the likelihood function, is not available. This leaves the preferred strategy for parameter estimation an open question. There are many methods that deals with this problem, and they tend to become highly complicated to implement in practice, especially when the diffusion is multidimensional. Within the last decade novel Bayesian methods have been developed which can be used for statistical inference and we describe one such Markov Chain Monte Carlo method and adapt it to the two-dimensional stochastic FitzHugh-Nagumo model for parameter inference.

Bartek Knapik (Vrije Universiteit)

Bayesian inverse problems – recovery of the initial condition for the heat equation

Suppose a differential equation describes the evolution of some feature of a system (e.g., heat conduction), depending on its initial value (at time $t = 0$). We observe the feature at time $T > 0$, in the presence of noise or measurement errors, and the aim is to recover the initial condition. Inverse problems of this type are often ill-posed in the sense that the solution operator of the differential equation, which maps the function describing the initial state to the function that describes the state at the later time $T > 0$ at which we observe the system, does typically not have a well-behaved, continuous inverse. This means that in many cases some form of regularization is necessary to solve the inverse problem and to deal with the noise.

In this talk we study a Bayesian approach to this problem for the particular example of recovering the initial condition for the heat equation. Specifically, we assume we have noisy observations of the solution u to the Dirichlet problem for the heat equation

$$\frac{\partial}{\partial t}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \quad u(x,0) = \mu(x), \quad u(0,t) = u(1,t) = 0,$$

where u is defined on $[0,1] \times [0,T]$ and the function $\mu \in L^2[0,1]$ satisfies $\mu(0) = \mu(1) = 0$. The solution to this problem is given by

$$u(x,t) = \sqrt{2}\sum \mu_i \exp(-i^2\pi^2 t)\sin(i\pi x),$$

where (μ_i) are the coordinates of μ in the basis $e_i = \sqrt{2}\sin(i\pi x)$, for $i \geq 1$.

We assume we observe the solution in white noise of intensity $1/n$. By expanding in the basis (e_i) this is equivalent to observing the sequence of noisy, transformed Fourier coefficients $Y = (Y_1, Y_2, \ldots)$ satisfying

$$Y_i = \kappa_i\mu_i + \frac{1}{\sqrt{n}}Z_i, \qquad i = 1, 2, \ldots,$$

for $\kappa_i = \exp(-i^2\pi^2 T)$ and (μ_i) as above, for $i \geq 1$, and $Z_1, Z_2, \ldots$ independent, standard normal random variables. The aim is to recover the coefficients μ_i, or equivalently, the initial condition $\mu = \sum \mu_i e_i$, under the assumption that the signal-to-noise ratio tends to infinity (so $n \to \infty$).

In order to make inference about μ we put a Gaussian process prior on μ, and compute the corresponding posterior. We consider a class of prior distributions indexed by a parameter quantifying 'smoothness' and show that the corresponding posterior distributions contract around the true parameter at a rate that depends on the smoothness of the true initial condition and the smoothness and scale of the prior. Correct combinations of these characteristics lead to the optimal minimax rate. One type of priors leads to a rate-adaptive Bayesian procedure. The frequentist coverage of credible sets is shown to depend on the combination of the prior and true parameter as well, with smoother priors leading to zero coverage and rougher priors to (extremely) conservative results. In the latter case credible sets are much larger than frequentist confidence sets, in that the ratio of diameters diverges to infinity. The results are numerically illustrated by a simulated data example.

The talk is based on a joint work with Aad van der Vaart and Harry van Zanten.

Umberto Picchini (Lund University)

Inference for SDE models via Approximate Bayesian Computation

Models defined by stochastic differential equations (SDEs) allow for the representation of random variability in dynamical systems. The relevance of this class of models is growing in many applied research areas and is already a standard tool to model e.g. financial, neuronal and population growth dynamics. However inference for multidimensional SDE models is still very challenging from a computational and theoretical point of view. Recent advances in Approximate Bayesian Computation (ABC) allow to perform Bayesian inference for models which are sufficiently complex that the likelihood function is either analytically unavailable or computationally prohibitive to evaluate. We want to consider how Bayesian inference can be effectively applied to complex SDE models via an MCMC ABC algorithm. Focus is on the case where the SDE describes the dynamics of observations affected by measurement error, the latter being a non-negligible source of variability (and inferential complications!) for most biomedical/biostatistical applications. Simulation studies for a simple pharmacokinetic model and a more complex multidimensional SDE for the modelization of stochastic kinetic networks are considered.

Mélanie Prague (Université Bordeaux 2) and Daniel Commenges

Bayesian MAP estimation in models with random effects based on ordinary differential equations applied to treatment monitoring in HIV

Studies of dynamical models based on Ordinary Differential Equations (ODE) considerably enhanced the knowledge about the interaction between HIV and the immune system [1]. Parameters in these models bring valuable information about cells birth and death rates in the population. Moreover, random effects to take into account inter-individual variability open the perspective of treatment individualization based on ODE equilibrium properties.

We use the "Activated T cell model" [2] with random effects on several parameters. A pharmacodynamic function links the treatment dose to the effect of several antiretroviral drugs. We adopt a Bayesian approach because of problems in practical identifiability in such a model. Priors are elicitated so as to cover previously published evaluations. The numerical complexity leads us to choose a Maximum a Posteriori (MAP) estimation method [2] rather than a full Bayesian estimation by MCMC. We will describe the EMRODE algorithm (Estimation in Models with Random effects based on Ordinary Differential Equations) based on a Newton-like algorithm proposed by Guedj et al. [3]. Model prediction abilities will be illustrated on two clinical trials.

ODE equilibrium can be characterized by the basic reproductive number R_0. When $R_0 \leq 1$, infection is controlled. To optimize a treatment, no cost function is necessary; we only have to control the probability that R_0 is below one. A Bayesian algorithm such as Metropolis-Hastings allows us to calculate the a posteriori distribution of R_0 for a specific observed patient. We will present an adaptive drug dose tuning algorithm, which converges toward the critical dose characterized by $R_0 = 1$ when information about the patient increases.

References

[1] Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M. and Ho, D.D. (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. *Science*, 5255(271):1582–1586.
[2] M. Prague, D. Commenges, J. Drylewicz and R. Thiébaut. (2012) Treatment monitoring in HIV infected Patients. *Biometrics*, to appear.
[3] J. Guedj, R. Thiébaut and D. Commenges. (2007) Maximum likelihood estimation in dynamical models of HIV. *Biometrics*, 63(4):1198–1206.

Andreas Raue (University of Freiburg)

A study for inference in the presence of non-identifiability: Bayesian MCMC sampling vs. profile likelihood approach

Increasingly complex applications involve large datasets in combination with non-linear and high dimensional dynamical models. In this context, statistical inference is a challenging issue that calls for pragmatic approaches that

take advantage of both Bayesian and frequentist methods. The elegance of Bayesian methodology is founded in the propagation of information content provided by experimental data and prior assumptions to the posterior probability distribution of model predictions. However, for complex applications experimental data and prior assumptions potentially constrain the posterior probability distribution insufficiently. In these situations Bayesian Markov chain Monte Carlo sampling can be infeasible. From a frequentist point of view insufficient experimental data and prior assumptions can be interpreted as non-identifiability. The profile likelihood approach offers to detect and to resolve non-identifiability by experimental design iteratively [Raue et al., 2009]. Therefore, it allows one to better constrain the posterior probability distribution until Markov chain Monte Carlo sampling can be used securely. Using an application from cell biology [Becker et al., 2010] we compare both methods and show that a successive application of both methods facilitates a realistic assessment of uncertainty in model predictions [Raue et al., 2012].

References

A Raue, C Kreutz, T Maiwald, J Bachmann, M Schilling, U Klingmüller, and J Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

V Becker, M Schilling, J Bachmann, U Baumann, A Raue, T Maiwald, J Timmer, and U Klingmueller. Covering a broad dynamic range: information processing at the erythropoietin receptor. *Science*, 328(5984):1404–1408, 2010.

A Raue, C Kreutz, F Theis and J Timmer. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. To appear in *Phil Trans Roy Soc A*, 2012.

Joep Vanlier (TU Eindhoven), Christian Tiemann, Peter Hilbers and Natal van Riel

Targeted experimental design using the posterior predictive distribution

Introduction. Systems biology employs mathematical modeling to further our understanding of biochemical pathways. The complexity of models necessary to describe biological pathways in combination with the limited amount of quantitative data results in large parameter uncertainty which propagates into model predictions. When predictions required to test the hypothesis are insufficiently constrained more data will be required. However, it is often not immediately evident which measurement(s) at which specific time point(s) would be most informative. We focus on designing experiments specifically targeting the variance of quantities of interest that depend on model predictions.

Methods. In this work we used a Bayesian approach to infer a posterior distribution based on the model and the data. Self-normalized importance sampling of the Posterior Predictive Distribution (PPD) was used to perform Optimal Experiment Design (OED).

Results. We proposed a flexible Bayesian method for hypothesis driven experimental design that exploits relations within the posterior predictive distribution whilst considering finite measurement accuracy and model uncertainty. This approach is endowed with the ability to consider multiple measurements under multiple experimental conditions simultaneously. Moreover, the method allows great freedom in terms of quantity of interest. Experiment(s) can be optimized for any quantity that can be expressed in terms of the model and model parameters. We present our method by illustrating its use on a model of the JAK-STAT signaling pathway.

Ivan Vujacic (University of Groningen), Javier Gonzalez and Ernst Wit

A new statistical framework to infer gene regulatory networks with hidden transcription factors

Regulatory networks consist of genes encoding transcription factors (TFs) and the genes they activate or repress. Various types of systems of ordinary differential equations (ODE) have been proposed to model these networks, ranging from linear to Michaelis-Mentin approaches. In practice, a serious drawback to estimate these models is that the TFs are generally unobserved. The reason is the actual lack of high-throughput techniques to measure abundance of proteins in the cell. The challenge is to infer their activity profile together with the kinetic parameters of the ODE using level expression measurements of the genes they regulate. In this work we propose general statistical framework to infer the kinetic parameters of regulatory networks with one or more TFs using time course gene expression data. Our approach is also able to predict the activity levels of the TF. We use a penalized likelihood approach where the ODE is used as a penalty. The main advantage is that the solution of the ODE is not required explicitly as it is common in most proposed methods. This makes our approach computationally efficient and suitable for large systems with many components. We use the proposed method to study a SOS repair system in Escherichia Coli. The reconstructed TF exhibit a similar behavior to experimentally measured profiles and the genetic expression data are fitted properly.

Posters

Miguel Atencia (University of Malaga) and Gonzalo Joya

Parameter estimation for dynamical systems based upon Hopfield and Tank