39th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering

MaxEnt 2019

Scope

Main topics of the workshop are the application of Bayesian inference and the maximum entropy principle to inverse problems in science, machine learning, information theory and engineering.

Inverse and uncertainty quantification (UQ) problems arise from a large variety of applications, such as earth science, astrophysics, material and plasma science, imaging in geophysics and medicine, nondestructive testing, density estimation, remote sensing, Gaussian process (GP) regression, optimal experimental design, data assimilation and data mining.

The workshop thus invites contributions on all aspects of probabilistic inference, including novel techniques and applications, and work that sheds new light on the foundations of inference.

Advisory Committee

A. Caticha P.M. Goggans K.H. Knuth A. Mohammad-Djafari R. Niven J. Skilling U. von Toussaint G. Verdoolaege University at Albany, USA University of Mississippi, USA University at Albany, USA LSS-CNRS, France UNSW Canberra, Australia MEDC, UK IPP, Germany Ghent Univ. and RMA, Belgium

Local Organizing Committee

U. von Toussaint R. Preuss L. Fahrner / A. Bauer D. Nille / J. Dominguez

Homepage

Further information can be found on our website https://www.ipp.mpg.de/maxent2019

The workshop is sponsored by: E.T. Jaynes Foundation Max-Planck-Institut für Plasmaphysik

Best poster award sponsored by Journal "entropy"

Conference Secretariat

MaxEnt 2019 Max-Planck-Institut für Plasmaphysik Postfach 1322 85741 Garching bei München, Germany

Fax: +49-89-3299-96-1817 email: maxent2019@ipp.mpg.de

Garching bei München

30.6.-5.7.2019

Ξ
ца Ц
δ
5
ā
Φ
õ
E.
ž
fe
Ĕ
2
\mathbf{U}

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
08:15		Registration				
08:50		Opening by H. Zohm (IPP directorate)				
00:60		Tutorial by A. Caticha: Where do Hamiltonians come from	 S. Roberts: Optimization, Quadrature and Inference: pragmatic Bayesianism in practice 	J. Skilling: Galilean Monte Carlo	A. Caldwell: Parallelizing MCMC MCMC	K. Knuth: Estimating Flight Characteristics of Anomalous Aerial Vehicles
10:00		1	N. Carrara: On the Estimation of Mutual Information	W. Henderson: TI-Stan: Adaptively annealed thermodynamic integration with HMC	D. Gencaga: Effects of Neuronal Noise on Neural Communication	F. Uribe: Cross-Entropy Method with Failure- informed Dimension Reduction for Rare Event Estimation
10:30				Break and refreshments		
10:50		K.A. Earle: The Spin Echo, Entropy, and Experimental Design	N. Caticha: Entropic dynamics for learning in NN architectures	W. von der Linden: Nested Sampling in Statistical Physics	R. Niven: Bayesian Identification of Dynamical Systems	B.J. Stokes: Some Bayesian Equidistribution Tests
11:20		N.V. Denisova: Bayesian Approach with Entropy Prior for Open Systems	I. Farcas: Multilevel Adaptive Sparse Leja Approximations for Bayesian Inverse Problems	J. Knollmüller: Metric Gaussian Variational Inference	N. Carrara: Quantum Trajectories in Entropic Dynamics	G. Revillon: A complete Classification and Clustering Model to account for continuous and categorical Data in Presence of Missing Values
11:50		R. Leike: Charting nearby galactic dust: a Bayesian 3d non-parametric tomographic reconstruction	Y. Kalaidzidis: Choice of Prior for Deconvolution of Fluorescent Microscopy Images	J. Madrigal Cianci: Infinite Swapping in Bayesian Inverse Problems	S. Ipek: An Entropic Approach to Geometrodynamics	Conference closing and business meeting
12:20			Lunch	Break		

 Friday									
Thursday	C. Rodriguez: Learning form Data&Prior	A. Caticha The Information Geometry of space time	J. Skilling: Against Information Geometry (ends	16:00)				Departure for conference dinner	Bus back to Garching
Wednesday	A. Mohammad-Djafari: Interaction between Model based Signal and Image processing, Machine Learning and Artificial Intelligence	J. Stern: Randomness in the Courthouse		A. Buchholz: Adaptive Tuning Of Hamiltonian Monte Carlo Within Sequential Monte Carlo	A. Baraov: Electromagnetic induction and relativistic double layer: Mechanism for ball lightning formation	C. Kauffmann: Inference and Regression in Ramachandran Space: MaxEnt Characterization of the Protein Backbone			
Tuesday	P.S. Koutsourelakis: Physics-aware Deep Learning for High- dimensional Uncertainty Quantification without Labeled Data	S. Ranftl: On the Diagnosis of Aortic Dissection with Impedance Cardiography: A Bayesian Feasibility Study with Multi-Fidelity Simulation Data	Break and refreshments	C. Albert: Gaussian Processes for data fulfilling linear Differential equations	Poster session (with	Bavarian beverages)			
Monday	O. Le Maitre: Surrogate models and reduction methods for UQ and inference in large-scale models	M. Van Soom: A new approach to the fromant measurement problem		M. Trassinelli: An unique prob. Func. For quant. and class. phenomena where distributivity is violated	J. Latz: On the well- posedness of Bayesian inverse problems	Y. Lim: Local Bayesian regularization with entropy prior in an ill-posed inverse problem of emission medical tomography			
Sunday	Welcome Reception with beverages	Tutorial by R. Bontekoe:Bayes' Theorem, a toolbox for data analysis (four lectures of 45 min)							
	14:00	15:00	15:30	15:50	16:20	16:50	17:20	17:30	22:15

Š
ē
S
0
Ω_

S. Cameron	C. Chapdelaine	L. Esteves	H. Graziadei	F. Guglielmetti	W. Henderson	R. Jamodien	K. Knuth	M. Lauretto	D. Nille	A. Mohammadpour	R. Preuss	H. Takada	M. Trassinelli	N. Vakili
~	0	ი	4	5	9	7	ø	6	10	1	12	13	4	15

Variational Bayesian Approach in Model-Based Iterative Reconstruction for 3d X-ray Computed Tomography with Gauss-Markov-Potts Prior Nested sampling for atomic physics data: the nested_fit program Markov Chain Monte Carlo Algorithms For Bayesian 3D Reconstruction In Cryo-Electron Microscopy Haphazard Intentional Sampling Techniques in Network Design of Monitoring Stations A Sequential Marginal Likelihood Approximation Using Stochastic Gradients Bayesian determination of parameters for plasma-wall interactions Using entropy to forecast bitcoin's daily conditional value at risk. Bayesian forecasting overdispersed integer-valued time series Empirical algorithm analyses of model comparison methods Distinguishability: The Lesson of the Two Children Problem Deconvolution for events faster than the sampling rate Bayesian Reconstraction with Adaptive Image Notion Pragmatic hypotheses in the evolution of Science Carpets color detection based on their images Radial scales in linear model comparison

Sunday, June 30th

R. Bontekoe

Abstract

This tutorial is aimed at novices in Bayesian data analysis.

Data have little value when their context is lacking. A model is crucial to understand the data. Common data analysis methods, such as least squares fitting, sometimes return undesirable results. The root cause of such failures is often a poorly chosen data model. Another root cause can be the lack of a well defined model for the solution. With the aid of Bayes Theorem both requirements can be reconciled.

Bayes Theorem appears as a simple formula. Originating from the mathematical logic, the formula has to undergo several transformations before it can be used in data analysis. Assumptions about data, model and any background information are to be made. This introduces a seemingly subjective element in the analysis. However, in a Bayesian analysis the effect of every assumption can be quantified, and, if necessary, rectified.

The tutorial covers traditional least squares fitting, overfitting of data, some non-linear fitting, numerical instabilities and remedies. Bayes Theorem is derived as the fundamental formula to manipulate probabilities. Maximum Entropy is shown as a tool to obtain probability distributions from general background information. Bayesian linear regression is covered as well as Bayesian model selection. Objective model selection is the Unique Selling Point of Bayesian data analysis, and lies at the heart of all modern Information Processing.

Monday, July 1st

Where do Hamiltonians come from? A derivation from Entropic Dynamics and Information Geometry

Ariel Caticha*

Department of Physics, University at Albany, NY 12222, USA

I describe new developments in the Entropic Dynamics (ED) approach to Quantum Mechanics. In ED the dynamics of the probability distribution is driven by entropy subject to constraints that are eventually codified into a quantity called the "phase". The central challenge is to identify the relevant physical constraints and, in particular, to specify how those constraints are themselves updated.

In this talk I introduce a new form of Entropic Dynamics in which particles follow smooth differentiable trajectories (as opposed to non-differentiable Brownian paths). I note that the space of probabilities and phases has a natural *symplectic* form (i.e., it is a *phase* space with Hamiltonian flows, Poisson brackets, and all that) and then (using an argument from information geometry) I show that it also has a *metric tensor* (the Fubini-Study metric) and a *complex structure*.

I show that the ED that preserves the symplectic, metric, and complex structures – this is a Hamilton-Killing flow in phase space -- is the linear Schrodinger equation. This explains where the superposition principle comes from, why wave functions are complex and single-valued, and why electric charges are quantized. Finally I observe that Hilbert spaces are not necessary ingredients in this construction. They are a clever but merely optional trick that turns out to be convenient for practical calculations.

* Email: acaticha@albany.edu)

The Spin Echo, Entropy, and Experimental Design

<u>K. A. Earle¹</u>, O. Kazakov¹

(1) University at Albany

kearle@albany.edu, https://earlelab.rit.albany.edu/

Abstract

The spin echo experiment is an important tool in magnetic resonance for exploring the coupling of spin systems to their local environment. Properly interpreted, the spin echo spectrum can yield a wealth of information about chemical structure and dynamic processes in a variety of physical systems. This is particularly true in Electron Spin Resonance (ESR), where the relevant spins are typically weakly localized and strongly coupled to the local environment. This is to be contrasted with Nuclear Magnetic Resonance (NMR) where the spins are generally well-localized and weakly coupled to the local environment. The strong couplings in a typical ESR experiment lead to relaxation effects that are orders of magnitude faster than in analogous cases in NMR. This leads to significant technical constraints on the kinds of time domain experiments that one can perform in ESR vis-à-vis NMR. In particular, the free induction decay (FID) experiment which reports on the correlation function of the interacting spins is often difficult to observe in an ESR experiment, due to spectrometer deadtime and rapid decay rates. The spin echo experiment, on the other hand, relies on a refocusing of the relevant spin magnetization, and can be observed well after the deadtime limitations that can distort and confound the ESR-FID experiment. The refocusing of the spin magnetization where a signal grows out of the baseline to an observable magnitude would seem to be at odds with the second law of thermodynamics. This is not the case, as the observed refocusing is purely a result of conservative, Hamiltonian evolution processes, which leave the entropy unchanged. Relaxation processes play the key role for a proper understanding of entropy production in the spin echo experiment, as we will demonstrate. In contemporary ESR, this has important consequences for the design of pulse sequences where, due to finite excitation bandwidths, contributions from the Hamiltonian dynamics and relaxation processes must be considered together in order to achieve a quantitative treatment of the effects of selective pulses on the spin system under study. We will demonstrate how a nested sampling approach can be used to optimize the selective pulse parameters to achieve a desired spin system response. Although the examples treated are relevant to fluid state systems, the approach is generic and may in addition be expected to be of use for solid state systems, both in NMR and ESR.

Key Words: Magnetic Resonance, entropy, optimization, nested sampling

Bayesian Approach with Entropy Prior for Open Systems

N.V.Denisova

Institute of Theoretical and Applied Mechanics, 630090, Novosibirsk, Russia

This paper discusses the problem of prior probability formulation in the Bayesian Maximum a Posteriori (MAP) approach in the context of imaging problems in positron emission tomography (PET) and single photon emission computer tomography (SPECT). Two probability forms of prior are widely discussed in literature: Entropy and Gibbs priors. These forms are analyzed from the classical statistical physics point of view. The theoretical analysis has shown that both forms are valid in the closed systems approximation. But, living biological objects and all emitting objects are open systems. We have shown that standard approaches with Gibbs and Entropy priors have fundamental limitations due to the closed systems model. In solving inverse ill-posed image reconstruction problem, these limitations lead to a global regularization, in which a single parameter controls the solution in the whole solution area. Global regularization results in over-smoothed images with blurry edges. This problem was discussed in our previous research in solving imaging problems in plasma physics and in medicine [1,2]. Over-smoothing is undesirable, especially in oncology in diagnosis of cancer lesions. The idea about how to overcome the over-smoothing and to improve resolution and sensitivity of reconstructed images, was suggested in the 80-s of the last century, by the authors of the regularization theory [3]. This idea is known as local regularization. The local regularization method was developed and successfully applied within a deterministic approach to clinical X-ray imaging.

In the present paper, to overpass the limitation of global regularization in Bayesian method, a new approach to prior probability assignment is suggested. This approach is developed for open systems and is based on Klimontovich's S-theorem for open systems entropy [4]. It is shown that the new approach expanded the limits of standard Gibbs and Entropy prior probabilities for closed system and allowed to justify the method of local statistical regularization. Applications of this method for SPECT imaging are presented in [5].

This work was partly supported by Russian Foundation for Basic Research (grant No. 17-52-14004).

[1] N.V.Denisova IEEE Trans. Nucl Sci. 51 136 (2004)

[2] Denisova N IEEE Trans. Plasma Sci. 37 4 502 (2009).

[3] V.Ya.Arsenin, Yu.A.Kriksin, A.A.Timonov USSR Computational Mathematics and Mathematical Physics 28, 3 113 (1988)

[4] Yu.L.Klimontovich S-Theorem Z.Phys. B 66,125, 1987.)

[5] Y.Lim, N. Denisova About local Bayesian regularization with Entropy prior in an illposed inverse problem of emission medical tomography (submitted for MaxEnt 2019)

*Corresponding author: Tel.: +7 383 3304275 E-mail address: NVDenisova2011@mail.ru

CHARTING NEARBY GALACTIC DUST: A BAYESIAN 3D NON-PARAMETRIC TOMOGRAPHIC RECONSTRUCTION

R. Leike^{1,2}, T. Enßlin^{1,2}

(1) Max Planck Institute for Astrophysics, Karl-Schwarzschildstraße 1, 85748 Garching, Germany (2) Ludwig-Maximilians-Universität, Geschwister-Scholl Platz 1, 80539 Munich, Germany

Abstract

The distribution of galactic dust in the vicinity of our sun is a relevant observable in astrophysics, as it gives constraints on the dynamics of our galaxy, the mechanism of star formation and insights into the origin of cosmic rays. Cosmic dust also obscures objects that are located behind it. For many of these effects, it is insufficient to only know the dust distribution as seen on the celestial sphere, one needs information about the full three dimensional density.

The dust density is only measured indirectly along non-uniformly distributed lines of sight and admits to a complex noise statistics. Thus inferring the dust density poses a challenging inverse problem as the measurement is non-uniform, the noise-statistics is non-Gaussian and the unknown dust density has degrees of freedom for every point in space. An informative prior is required to constrain the infinite degrees of freedom of the dust density in the presence of only finite data.

We present a Bayesian reconstruction [1], using a variant of the variational Bayes algorithm called Metric Gaussian Variational Inference (MGVI) [2] that is implemented in the free software package NIFTy5 [3]. We model the dust density using a log-normal process, with a-priori unknown correlation kernel that is also inferred non-parametrically. Using data from the Gaia collaboration, we infer the dust density and an uncertainty estimate for a 600pc box to a voxel-resolution of 2.3pc. It is the first reconstruction to manifestly show intrinsic dynamic properties of galactic dust, like its filamentary nature, in three dimensions.

References:

[1] R. Leike et al.; Charting nearby Galactic dust clouds using Gaia data only; ArXiv 1901.05971

[2] J. Knollmüller et al.; Metric Gaussian Variational Inference; ArXiv 1901.11033

[3] P. Arras et al.; NIFTy5: Numerical Information Field Theory v5; Astrophysics Source Code Library

Surrogate models and reduction methods for UQ and inference in large-scale models

O. Le Maître Centre de Mathématiques Appliquées, Ecole Polytechnique

Abstract

Uncertainty Quantification (UQ) and Global Sensitivity Analysis (GSA) in numerical models often rely on sampling approaches (either random or deterministic) that call for many resolutions of the model. Even though these computations can usually be carried out in parallel, the application of UQ and GSA methods to large-scale simulations remains challenging, both from the computational, storage and memory points of view. Similarly, Bayesian inference and assimilation problems can be favorably impacted by over-abundant observations, because of overconstrained update problems or numerical issues (overflows, complexity,...), raising the question of observations reduction.

A solution to alleviate the computational burden is to use a surrogate model of the full large scale model, that can be sampled extensively to estimate sensitivity coefficients and characterize the prediction uncertainty. However, building a surrogate for the whole large scale model solution can be extremely demanding and reduction strategies are needed. In this talk, I will introduce several techniques for the reduction of the model output and the construction of its surrogate. Some of these techniques will be illustrated on ocean circulation model simulations. For the reduction of observations, I will discuss and compare few strategies based on information theoretical considerations that have been recently proposed for the Bayesian framework.

A new approach to the formant measurement problem

M. Van Soom* and B. de Boer

Vrije Universiteit Brussel, Artificial Intelligence Lab, Brussels, 1050, Belgium

Formants are characteristic frequency components in human speech that are caused by resonances in the vocal tract during speech production. Fant (1960) systematized the then relatively young science of acoustic phonetics with his acoustic theory of speech production, often called the source-filter model, which has since become the dominant paradigm. The source-filter model provides the theoretical justification for deriving formants from power spectra as prominent local maxima in the power spectral envelope of appropriately windowed and processed speech signals. From this point of view each formant is characterized by three parameters describing the local maximum associated with it: the maximum's center frequency (called the formant frequency), its bandwidth and its peak amplitude.

The concept of a formant is fundamental to phonetics and automated speech processing. For example, formants are considered to be primary features for distinguishing vowel classes, speech perception and for inferring speaker identity, sex and age. Despite this fundamental status – and despite a long history of work on vowel formants - the issue of making accurate measurements of the formant parameters, which we dub "the formant measurement problem" for convenience, is as yet not considered to be fully resolved (e.g. Maurer 2016). Accordingly, a large amount of formant measurement methods exist in the literature. The fundamental reason underlying the formant measurement problem is the fact that most of these methods yield formant frequency estimates (the main quantity of interest) that are sensitive to various user-made choices, such as the form and length of the tapering window or the number of poles in linear predictive analysis. In other words, measuring formants requires careful fine-tuning while speech is notorious for its variability. In addition, there currently seems to be no way to put error bars on the formant frequency, bandwidth and amplitude measurements; in fact, the bandwidth is typically considered to represent the accuracy of the formant frequency, a small bandwidth indicating high accuracy and vice versa.

We believe that perhaps it is time for a fresh start on remedying these issues. Our approach to the formant measurement problem consists of replacing the source-filter model by the arguably more general transient theory of voice production (e.g. Ladefoged 1996) in order to open up the possibility of applying Jaynes' Bayesian spectrum analysis (Bretthorst 1988). We expected a priori that a Bayesian analysis would yield quite sharp formant frequency estimates – now equipped with error bars – since acoustic phonetics gives a wealth of prior information about the problem at hand, including parametric models of the speech waveform, and our preliminary results confirm this expectation. Two important examples of the use of high-accuracy measurements of the formant parameters are forensic speaker identification and medical diagnosis.

G. L. Bretthorst, Bayesian spectrum analysis and parameter estimation. 1988.

G. Fant, Acoustic theory of speech production. Den Haag: Mouton, 1960.

P. Ladefoged, Elements of acoustic phonetics. University of Chicago Press, 1996.

D. Maurer, Acoustics of the Vowel. Bern, Switzerland: Peter Lang, 2016.

*Corresponding author: <u>marnix@ai.vub.ac.be</u> (M. Van Soom)

AN UNIQUE PROBABILITY FUNCTION FOR QUANTUM AND CLASSICAL PHENOMENA WHERE DISTRIBUTIVITY IS VIOLATED

M. Trassinelli¹

(1) Institut des NanoSciences de Paris, CNRS, SU, Paris, France martino.trassinelli@insp.jussieu.fr

Abstract

From a formulation based on logic of statements, extending the definition to the conditional case, we show that a unique definition of the probability function for quantum and classical phenomena can be formulated [1]. This universal definition has however a price: the distributive property of the probability argument is no more valid. Different arguments correspond to different conditions of discernibility and then the presence or absence of interference effects. In the example of Young's slits (see figure), this depends on the possibility to determine from which slit A or B the particle passed before hitting a detector D, where "(A or B true) and (D true)" is different to "(A and D true) or (B and D true)". The presented results are compared and commented in view of recent approaches for the foundation of Quantum Mechanics. In particular, we demonstrate how the third (of three) postulate of Relational Quantum Mechanics [2] can be avoided by the rigorous definition of the probability function [1].



References:

[1] M. Trassinelli, Found. Phys. 48, 1092-1111 (2018).

[2] C. Rovelli, Int. J. Theor. Phys. 35, 1637-1678 (1996)

Key Words: Generalized Probability Function, Quantum Mechanics, Distributive Property

On the well-posedness of Bayesian inverse problems

$\underline{J. Latz}^1$

(1) Zentrum Mathematik, Technische Universität München

Abstract

The subject of this talk (and its corresponding article [1]) is the introduction of a weaker concept of well-posedness of Bayesian inverse problems. The conventional concept of ('Lipschitz') well-posedness in [2] is difficult to verify in practice, especially when considering blackbox models, and probably too strong in many contexts. Our concept replaces the Lipschitz continuity of the posterior measure in the Hellinger distance by just continuity. This weakening is tolerable, since the continuity is in general only used as a stability criterion. The main result of this article is a proof of well-posedness for a large class of Bayesian inverse problems, where very little or no information about the underlying model is available. It includes any Bayesian inverse problem arising when observing finite-dimensional data perturbed by additive, non-degenerate Gaussian noise. Moreover, well-posedness with respect to other probability metrics is investigated, including weak convergence, total variation, Wasserstein, and also the Kullback-Leibler divergence.

References:

- [1] Latz arXiv e-prints **1902.10257** (2019).
- [2] Stuart Acta Numerica **19**, pp. 451-559 (2010).

Keywords: inverse problems, Bayesian inference, well-posedness, supervised learning, Kullback-Leibler divergence

LOCAL BAYESIAN REGULARIZATION WITH ENTROPY PRIOR IN AN ILL-POSED INVERSE PROBLEM OF EMISSION MEDICAL TOMOGRAPHY

<u>Y. Lim</u>¹, N. Denisova^{1,2} (1) Novosibirsk State University (2) Khristianovich Institute of Theoretical and Applied Mechanics Novosibirsk, Russia (yuli-lim@yandex.ru)

Abstract

While solving a reconstruction problem in nuclear medicine one is always restricted by poor statistics of the data. Bayesian regularization with entropy prior is a powerful tool for dealing with stochastic data [1]. The standard Bayesian approach uses global regularization in which a single parameter controls the solution in the whole solution area. Although global regularization allows one to obtain a stable solution, the quality of the reconstructed image is limited. More specifically, small details, which are of great importance for medical diagnostics, cannot be reconstructed with a desirable accuracy.

Here we show that local regularization allows us to overcome the restriction on the reconstruction error mentioned above and to obtain images of higher quality. Properties of the local regularization approach are studied in numerical experiments simulating Single Photon Emission Computer Tomography (SPECT) method [2]. A simple model with "hot spots" and the Hoffman brain phantom were used in simulations. The results obtained allow us to conclude that local regularization is a promising method, especially for "hot spots" diagnostics in nuclear oncology. This work was supported by the Russian Foundation for Basic Research (grant No. 17-52-14004).

References:

[1] J. Skilling, S. F. Gull. Bayesian maximum entropy image reconstruction. Spatial statistics and imaging, Institute of Mathematical Statistics, Hayward, CA, 341– 367 (1991)

[2] N. V. Denisova. Regularized versus non-regularized statistical reconstruction techniques. Nuclear Instruments and Methods in Physics Research Section A, **648** 65–67, (2011)

Key Words: global regularization, local regularization, Bayesian approach, maximum entropy, nuclear medicine Tuesday, July 2nd

Optimization, Quadrature and Inference: pragmatic Bayesianism in practice

S. Roberts University of Oxford, UK

Abstract

Over the past decade Bayesian non-parametric models, particularly the Gaussian Process, have been widely researched. The use of flexible function modelling has had impact in Bayesian Optimization and Quadrature as well as in direct data models. This talk will overview the benefits and the drawbacks of Gaussian Processes, showcasing their impact both methodologically and practically.

On the Estimation of Mutual Information

Nicholas Carrara University at Albany. Jesse Ernst University at Albany.

Abstract

Because of the recent interest in using estimates of mutual information (MI) to improve machine-learning (ML) methods[1], it is important to understand the stability of those estimates. ML techniques generally involve coordinate transformations of the data. Further, the input data often include variables that are uncorrelated with either the result classes or with other variables (i.e., noise). We therefore study how coordinate transformations and noise variables impact estimates of MI. We focus our attention on the class of MI estimators, such as Kraskov et al.[2, 3], that rely only on the geometry of the sample, and avoid the need to estimate the underlying probability distribution.

References

- [1] Nicholas Carrara and Jesse A. Ernst. On the Upper Limit of Separability. *arXiv e-prints*, page arXiv:1708.09449, Aug 2017.
- [2] H. Stoegbauer A. Kraskov and P. Grassberger. Estimating mutual information. *Phys. Rev. E*, 69, 2004.
- [3] Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. Efficient estimation of mutual information for strongly dependent variables. CoRR, abs/1411.2003, 2014.

Entropic dynamics for learning in Neural Network architectures

<u>Nestor Caticha</u> Instituto de Física, Universidade de São Paulo,

05315-970, São Paulo, SP, Brazil.

ncaticha@usp.br

April 10, 2019

Abstract

We consider on-line learning in feed-forward Neural Networks from a labeled learning set. Information about the weights of the Neural Network is coded by a member of a family of distributions and its obtained by Maximum Entropy. The location hyperparameters represent the current weights estimates. Learning by Bayesian updates drive the probability distribution away from the family manifold, but determine new constraints on the generators of the family. Hence an update to a new distribution can be done. This is learning by entropic dynamics in neural network architectures (EDNNA). For models with simple architectures this can be calculated and it leads to optimal generalization performance. We show that the hyperparameters change along the gradient of the evidence. For a feed-forward architecture we define the evidence for an internal representation and show that it can be written recursively from the evidence up to the previous layer convoluted with an aggregation kernel, defined by the weights between the layers. We take the continuum limit of layers. This leads to a diffusion-like parabolic PDE equivalent to Wilson's RG but with an aggregation kernel different from those that integrate out ultraviolet degrees of freedom. The RG leads from the microscopic degrees of freedom to the infrared behavior, or thermodynamics. This generalization of the RG leads from the microscopic degrees of freedom of the input data to the macroscopic class or concept.

Key Words: Neural Networks, Entropic dynamics, Renormalization Group

Multilevel Adaptive Sparse Leja Approximations for Bayesian Inverse Problems

 $\begin{array}{ccc} \mbox{Ionut-Gabriel Farcas}^1 & \mbox{Jonas Latz}^2 & \mbox{Elisabeth Ullmann}^2 \\ & \mbox{Tobias Neckel}^1 & \mbox{Hans-Joachim Bungartz}^1 \end{array}$

¹Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany ²Department of Mathematics, Technical University of Munich, Boltzmannstr.

²Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

Deterministic interpolation and quadrature methods are often unsuitable to address Bayesian inverse problems depending on computationally expensive forward mathematical models. While interpolation may give precise posterior approximations, deterministic quadrature is usually unable to efficiently investigate an informative and thus concentrated likelihood. This leads to a large number of required expensive evaluations of the mathematical model. To overcome these challenges, we formulate and test a multilevel adaptive sparse Leja algorithm [1]. At each level, adaptive sparse grid interpolation and quadrature are used to approximate the posterior and perform all quadrature operations, respectively. Specifically, our algorithm uses coarse discretizations of the underlying mathematical model to investigate the parameter space and to identify areas of high posterior probability. Adaptive sparse grid algorithms [2, 3] are then used to place points in these areas, and ignore other areas of small posterior probability. The points are weighted Leja points [4]. As the model discretization is coarse, the construction of the sparse grid is computationally efficient. On this sparse grid, the posterior measure can be approximated accurately with few expensive, fine model discretizations. The efficiency of the algorithm can be enhanced further by exploiting more than two discretization levels. We apply the proposed multilevel adaptive sparse Leja algorithm in numerical experiments involving elliptic inverse problems in 2D and 3D space, in which we compare it with Markov chain Monte Carlo sampling and a standard multilevel approximation.

References

- [1] I.-G. Farcaş, J. Latz, E.Ullmann, T. Neckel and H.-J. Bungartz, *Multilevel Adaptive Sparse Leja Approximations for Bayesian Inverse Problems*. 2019. submitted.
- [2] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), pp. 65–87.
- [3] I.-G. Farcaş, T. Görler, H.-J. Bungartz, F. Jenko, and T. Neckel, Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis, arXiv e-prints, (2018), arXiv:1812.00080, p. arXiv:1812.00080, https://arxiv.org/abs/1812.00080.
- [4] A. Narayan and J. D. Jakeman, Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation, SIAM J. Sci. Comput., 36 (2014), pp. A2952–A2983.
Choice of Prior for Deconvolution of Fluorescent Microscopy Images

Y. Kalaidzidis^{a,b,*}

^aMax Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany ^bFaculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

Deconvolution is widely used in fluorescent microscopy to remove image blurring by light diffraction on objective aperture and resolve micro-organization of intracellular organelles. Tikhonov deconvolution [1] in Fourier space with regularisation term, which is, in essence, a form of prior, is the fastest deconvolution method. Unfortunately, it is prone to produce artefacts like dark areas around bright objects and secondary lobes of intensity like a ripple waves on background. The iterative Bayesian methods (from works of Richardson [2] and Lucy [3]) became popular alternative in the microscopy image deconvolution. Unfortunately, Richardson-Lucy algorithms usually have no clear stop criterion and after few iterations are prone to amplify Poisson noise. To tackle this problem the additional regularisation terms were introduced (e.g. total-variation [4] in image space and PURE-LET [5] in wavelet space). Unfortunately, those algorithms have free parameters to tune (like number of iteration, or weight of regularising term) to balance between image blurring and artefact generation. In this work the Bayesian probabilistic model of fluorescent microscopy image deconvolution (without free tuneable parameters) is proposed. The model includes scattered background light and diffraction-limited image, which combination prevents most usual artefacts of deconvolution. Three forms of priors for intensity distribution were tested on publicly available fluorescent image datasets of conventional and single molecule localization super-resolution fluorescent microscopy [6]. Deconvolution power significantly depended on prior choice. The best prior resulted in resolution comparable with those of single molecule localization microscopy with 10 fold fewer frames as an input and produced no usual artefacts like dark rims or background ripples on conventional microscopy images.

- [1] Tikhonov, A. N. Soviet Mathematics. 4: 1035–1038 (1963)
- [2] Richardson, W, JOSA. 62 (1): 55–59 (1972)
- [3] Lucy, L. B. Astronomical Journal. 79 (6): 745–754 (1974)
- [4] Dey N., Blanc_Feraud L., Zimmer C., Roux P., Kam Z., Olivo-Marin J.-C., Zerubia J., Microsc. Res. Tehc. 69 (4) 260-266 (2006)
- [5] Li J., Luisier F., Blu T., Proc. IEEE Int. Conf. Img. Proc. (ICIP), Phoenix, USA, 2708-2712 (2016)
- [6] Nicolas Olivier D. K., Manley S. [http://bigwww.epfl.ch/smlm/datasets/] (2016)

*Corresponding author: Tel.: +49 351 210-2653 E-mail address: kalaidzi@mpi-cbg.de (Y. Kalaidzidis)

Physics-aware Deep Learning for High-dimensional Uncertainty Quantification without Labeled Data

P.S. Koutsourelakis¹

(1) Department of Mechanical Engineering, Technical University of Munich (p.s.koutsourelakis@tum.de)

Abstract

The development of surrogates/emulators of PDE systems for the purpose of uncertainty quantification tasks is are most often considered as a supervised learning task where input and output data pairs are used for training. The construction of such emulators is by definition a **Small Data** problem which poses challenges to deep learning approaches that have been developed to operate in the big data regime. Such issues are exacerbated when the dimension of the (random) inputs is high as is most often the case in challenging physical applications. This talk presents a methodology that incorporates the governing equations of the PDE-based, physical model in the loss/likelihood function. The resulting physics-aware, probabilistic, deep learning model is trained without any labeled data (e.g. employing only input data) and provide comparable predictive responses with data-driven models while obeying the constraints of the problem at hand. The generalization capability of these models to test inputs which are drastically different from the ones used in training is considered. Quantification and interpretation of the predictive uncertainty is provided for a number of problems.

Key Words: Bayesian, Variational Inference, Emulators, High-dimensional, Uncertainty Quantification

On the Diagnosis of Aortic Dissection with Impedance Cardiography: A Bayesian Feasibility Study with Multi-Fidelity Simulation Data.

S. Ranftl¹, G. M. Melito², V. Badeli³,

A. Reinbacher-Köstinger³, K. Ellermann², W. von der Linden¹
(1) Insitute of Theoretical Physics-Computational Physics *
(2) Institute of Mechanics *
(3) Insitute of Fundamentals and Theory in Electrical Engineering *
*Graz University of Technology
Correspondence: ranftl@tugraz.at
Homepage: biomechaorta.tugraz.at

Abstract

Aortic dissection is a cardiovascular disease with a disconcertingly high mortality. When it comes to diagnosis, medical imaging techniques such as CT, MRT or Ultrasound certainly do the job, but also have their shortcomings. Impedance cardiography is a standard method to monitor a patients heart function and circulatory system by injecting electric currents and measuring voltage drops between electrode pairs attached to the human body. If such measurements distinguished healthy from dissected aortas, one could improve clinical procedures. Experiments are quite difficult, and thus we investigate the feasibility with finite element simulations beforehand. In these simulations, we find uncertain input parameters, e.g. the electrical conductivity of blood. Inference on the state of the aorta from impedance measurements defines an inverse problem in which forward uncertainty propagation through the simulation with vanilla Monte Carlo demands a prohibitively large computational effort. To overcome this limitation, we combine two simulations; one simulation with a high fidelity and another simulation with a low fidelity, [1] and low and high computational costs accordingly. We use the cheap low-fidelity simulation to learn about the expensive high-fidelity simulation. It all boils down to a regression problem - and reduces total computational cost after all.

References:

[1] M.C. Kennedy and A. O'Hagan, Biometrica, 87, 1, pp. 1-13 (2000)

Key Words: Bayesian probability theory, uncertainty quantification, impedance cardiography, aortic dissection

GAUSSIAN PROCESSES FOR DATA FULFILLING LINEAR DIFFERENTIAL EQUATIONS

 $\underline{C.G. Albert}^1$

(1) Max-Planck-Institut für Plasmaphysik, 85748 Garching albert@alumni.tugraz.at

Abstract

Here a new approach to reconstruct fields and source distributions based on Gaussian process (GP) regression [1] is presented for the case where the data are known to fulfil a given linear differential equation $Lu(\mathbf{x}) = q(\mathbf{x})$. The method is applicable to a wide range of data from physical measurements and numerical simulations and allows to estimate parameters of reduced complexity models. Raissi et al. [2,3] have recently demonstrated the application of GPs to data from linear and non-linear differential equations representing physical information. There the covariance function for the source terms q is fixed via the application of the respective differential operator \hat{L} . Here a related but different path is taken that has first been applied to the acoustic Helmholtz equation in [4]. Instead of using a generic covariance function to represent data from an unknown field $u_{\rm h}$, the space of possible covariance functions is restricted to allow only Gaussian random fields that fulfil the homogeneous differential equation $Lu_{\rm h} = 0$. The resulting taylored kernel functions allow to model u_h more reliably from less training data compared to using a generic kernel. For differential equations representing laws of physics such a choice limits realizations of random fields to physically possible solutions. Parameters inside L can be estimated as in Refs. [2,3]. Since the GP is only used to model the homogeneous part $u_{\rm h}$ of the equation, source distributions q have to be imposed separately via the superposition principle. The according inverse problem allows to infer parameters entering q by sampling from the total field $u = u_{\rm h} + u_{\rm p}$. It is demonstrated how cross-validation of the GP for $u_{\rm h} = u - u_{\rm p}$ and varying the particular solution $u_{\rm p}$ by changing parameters in q can be used to perform this task.

References:

- C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. MIT Press (2006)
- [2] M. Raissi, P. Perdikaris, and G. E. Karniadakis, J. Comput. Phys., 348, 683–693 (2017).
- [3] M. Raissi and G. E. Karniadakis, J. Comput. Phys. 357, 125–141 (2018).
- [4] C. G. Albert, International Congress on Acoustics, Aachen (submitted, 2019).

Key Words: Gaussian process regression, kernel methods, parameter estimation, differential equations

Poster Session

A Sequential Marginal Likelihood Approximation Using Stochastic Gradients

<u>Scott A. Cameron^{1,2}</u>, Hans Eggers^{1,2}, Steve Kroon³

(1) Department of Physics, Stellenbosch University, Matieland 7602, South Africa

(2) National Institute for Theoretical Physics, Matieland 7602, South Africa

(3) Computer Science Division, Stellenbosch University, Matieland 7602, South Africa email: scott.a.cameron@live.co.uk

Abstract

Existing algorithms like nested sampling and annealed importance sampling are able to produce accurate estimates of the marginal likelihood of a model, but tend to scale poorly to large data sets. This is because these algorithms need to recalculate the log-likelihood for each iteration by summing over the whole data set. Efficient scaling to large data sets requires that algorithms only visit small subsets (minibatches) of data on each iteration. To this end, we estimate the marginal likelihood via a sequential decomposition into a product of predictive distributions $p(\mathbf{y}_n|\mathbf{y}_{< n})$. Predictive distributions can be approximated efficiently through Bayesian updating using stochastic gradient Hamiltonian Monte Carlo, which approximates likelihood gradients using mini-batches. Since each data point typically contains little information compared to the whole data set, the convergence to each successive posterior only requires a short burn-in phase. This approach can be viewed as a special case of Sequential Monte Carlo (SMC) with a single particle, but differs from typical SMC methods in that it uses stochastic gradients. We illustrate how this approach scales favourably to large data sets with some simple models.

Variational Bayesian Approach in Model-Based Iterative Reconstruction for 3d X-ray Computed Tomography with Gauss-Markov-Potts Prior

Camille Chapdelaine^{1,2}, Ali Mohammad-Djafari¹, Nicolas Gac¹, Estelle Parra² ¹ Laboratoire des signaux et systèmes, CNRS, CentraleSupélec-Univ Paris Saclay, Gif-sur-Yvette, France ² SAFRAN SA, Safran Tech, Pôle Technologie du Signal et de l'Information, Magny-Les-Hameaux, France

Abstract

3D X-ray Computed Tomography (CT) is used in medicine and non-destructive testing (NDT) for industry to visualize the interior of a volume and control its healthiness. Com- pared to analytical reconstruction methods, model-based iterative reconstruction (MBIR) methods obtain high-quality re- constructions while reducing the dose. Nevertheless, usual Maximum-A-Posteriori (MAP) estimation does not enable to quantify the uncertainties on the reconstruction, which can be useful for the control performed afterwards. Herein, we propose to estimate these uncertainties jointly with the recon- struction by computing Posterior Mean (PM) thanks to Varia- tional Bayesian Approach (VBA). We present our reconstruction algorithm using a Gauss-Markov-Potts prior model on the volume to reconstruct. For PM calculation in VBA, the uncertainties on the reconstruction are given by the variances of the posterior distribution of the volume. To estimate these variances in our algorithm, we need to compute diagonal co- efficients of the posterior covariance matrix. Since this matrix is not available in 3D X-ray CT, we propose an efficient solu- tion to tackle this difficulty, based on the use of a matched pair of projector and backprojector. In our simulations using the Separable Footprint (SF) pair, we compare our PM estimation with MAP estimation. Perspectives for this work are applications to real data as improvement of our GPU implementation of SF pair.

Pragmatic hypotheses in the evolution of Science

Julio M. Stern¹, Rafael Izbicki², Luís G. Esteves¹, Rafael B. Stern²

(1) Institute of Mathematics and Statistics, University of São Paulo, Brazil(2) Department of Statistics, Federal University of São Carlos, Brazil

(e-mails: jstern@ime.usp.br , rafaelizbicki@gmail.com , lgesteves2011@gmail.com , rbstern@gmail.com)

Abstract

This work introduces pragmatic hypotheses associated to precise hypotheses and relates this concept to the spiral of scientific evolution. Previous works determined a characterization of logically consistent Bayesian statistical hypothesis tests and showed that the modal operators obtained from these tests can be represented in the hexagon of oppositions. However, despite the importance of precise hypothesis in science, they cannot be accepted by logically consistent tests. Here, we show that this dilemma can be overcome by the use of pragmatic versions of precise hypotheses. These pragmatic versions allow a level of imprecision in the hypothesis that is small relative to other experimental conditions. The introduction of pragmatic hypotheses allows the evolution of scientific theories based on statistical hypothesis testing to be interpreted using the narratological structure of hexagonal spirals, as defined by Pierre Gallais.

Key Words: Pragmatic hypotheses, Generalized Full Bayesian SignificanceTest,Logically consistent hypothesis testing, Gallais' hexagonal spirals.

BAYESIAN FORECASTING OVERDISPERSED INTEGER-VALUED TIME SERIES

<u>Helton Graziadei</u>¹, Hedibert F. Lopes², Paulo C. Marques F.² (1) University of Sao Paulo, Brazil (2) Insper, Brazil

Abstract

We generalize the Poisson integer autoregressive model of arbitrary order, specifying a Poisson-Geometric mixture distribution, with unknown mixture weight parameter, for the process innovations. This mixture distribution allows the model to account adaptively for overdispersion in a time series of counts. Full conditional distributions of the model parameters are derived in simple closed forms through a data augmentation scheme, allowing us to forecast future counts in a thoroughly probabilistic way. In forecasting burglary occurrences at Washington D.C., the model exhibits a performance gain in the majority of the districts.

Key Words: INAR(p); data-augmentation; overdispersion.

Bayesian Reconstruction with Adaptive Image Notion

F. Guglielmetti

European Southern Observatory (Garching bei München) fgugliel@eso.org

Abstract

Bayesian probability theory is employed to provide a stable and unique solution to the ill-posed inverse problem of image analysis applied to calibrated interferometric data. Specifically, data observed with the Atacama Large Millimeter/Submillimeter Array (ALMA) are exploited. ALMA is an aperture synthesis telescope consisting of 66 antennas that can be positioned in a number of different configurations. It operates over a broad range of observing frequencies in the millimeter and submillimeter regime of the electromagnetic spectrum. Single field interferometry, mosaics, singledish observations, polarization, multiple region modes, solar, VLBI, astrometry and ephemeris observations can be performed with ALMA. The Common Astronomy Software Applications package (CASA) [1] is widely used for ALMA data reduction. CASA stands on a flexible design to process the data via task interface.

Extending the work of [2] to radio interferometric data, a new software is under development with the objective to become a new CASA task. The new algorithm, Bayesian Reconstruction with Adaptive Image Notion (BRAIN), aims at detecting the weak signal in the presence of varying noise also in extreme cases as sparse data and mosaic of images. The technique makes use of Gaussian statistics for a joint source detection and background estimation through a probabilistic mixture-model technique. Statistics is rigorously employed throughout the algorithm, so pixels with low intensity can be handled optimally and accurately, without binning and loss of resolution. Following the work of [3] on experimental spectra, BRAIN makes use of a 2D adaptive kernel deconvolution method to prevent spurious signal that may arise from the deconvolution process of the dirty image from the well-known dirty beam. Continuum, emission and absorption lines detection occurs without an explicit subtraction. An automated decision in the separation between different kind of signals is developed while minimizing spurious detections.

Preliminary results are shown on the application of BRAIN to ALMA simulated data.

References:

[1] CASA Documentation. https://casa.nrao.edu/casadocs-devel.

- [2] F. Guglielmetti et al. MNRAS **396**, 165 (2009).
- [3] R. Fischer et al. Phys. Rev. **E61**, 1152 (2000).

Key Words: methods: data analysis, statistical; techniques: image processing

EMPIRICAL ALGORITHM ANALYSES OF MODEL COMPARISON METHODS

 $\frac{\text{R. Wesley Henderson}^{1}, \text{Paul M. Goggans}^{1}}{(1) \text{ University of Mississippi, USA}}$ (goggans@olemiss.edu)

Abstract

We theoretically and empirically analyze the time complexity and precision of several model comparison algorithms. These algorithms include two parallel implementations of nested sampling and two parallel implementations of adaptively annealed thermodynamic integration. We examine the time complexity and precision of these algorithms as their method parameters are varied, and we examine the parallel speed-up that can be achieved on a multi-CPU shared-memory computer. The empirical tests are conducted using a signal detection problem using simulated data. Our results demonstrate that these methods see significant benefit when being deployed on many-CPU systems. Our results also provide guidance on the time penalties and precision gains encountered when setting method parameters.

Key Words: Model Comparison, MCMC, Empirical Algorithmics, Algorithm Analysis

RADIAL SCALES IN LINEAR MODEL COMPARISON

<u>R. Jamodien¹</u>, H.C. Eggers^{1,2}, M.B. de Kock³
(1) Stellenbosch University, Stellenbosch, South Africa
(2) NITheP, Stellenbosch, South Africa
(3) MPSD, Hamburg, Germany
email: 16113004@sun.ac.za

Abstract

Model comparison relies on evidence; evidence depends directly on priors; priors for models of different parameter space dimensionality K cannot cancel and thereby render evidence calculations and Bayes Factors vulnerable to yielding arbitrary values. We explore the simplest case of this dimensionality conundrum, that of linear regression under conditions of minimal prior information. Guided by the desideratum of maximal K-sensitivity of the likelihood and simultaneous minimal K-sensitivity of the prior, we establish approximate scale relations for the small-K and large-K asymptotic regimes for some typical priors, recasting the conundrum as that of K-hyperspheres centered on the origin and radial scales related to probability mass quantiles. In the case of the likelihood, the question of scales is reduced to a family of possible relations between the prior radial variable and the scale set by the coefficient of determination. The naive origin-centered sphere approach is subsequently supplemented by non-origin-centered spherical symmetry reflecting the generic knowledge of positive definiteness of the coefficient of determination. We compare our results to previously published work and some popular information criteria.

Distinguishability: The Lesson of the Two Children Problem

Kevin H. Knuth

Dept. of Physics, University at Albany, Albany NY, 12222, USA (kknuth@albany.edu)

Abstract

Puzzles designed with an intellectual twist act as an entertaining way of presenting a lesson that can highlight a deficit in our thinking or enlighten us in an often surprising way. However, it seems that the more subtle the lesson embedded within the puzzle, the more disagreement and debate ensues. This is often accompanied by a proliferation of varieties of the puzzle as people consider different yet related problems that they can solve. All this serves to distract from the underlying lesson, which is sometimes lost. Here I consider the problem known as the Two Child Problem where one makes inferences about the sexes of two children based on a spectrum of prior information. I demonstrate that the underlying lesson is that of distinguishability, which is not so unlike independence. When considered from this perspective, the paradoxes of the puzzle are resolved.

HAPHAZARD INTENTIONAL SAMPLING TECHNIQUES IN NETWORK DESIGN OF MONITORING STATIONS

M. S. Lauretto¹, R. B. Stern², F. Nakano¹, C. O. Ribeiro¹, <u>J. M. Stern¹</u> (1) University of Sao Paulo, Brazil (2) Federal University of Sao Carlos, Brazil (jmstern@hotmail.com, www.ime.usp.br/~jstern)

Abstract

In contemporary empirical science, sampling randomization is the golden standard to ensure unbiased, impartial, or fair results, see [1,2]. Randomization works as a technological barrier designed to prevent spurious communication of vested interests or illegitimate interference between parties in the application of interest, that may be a scientific experiment, a legal case, an auditing process, or many other practical applications. In randomized experiments, a common issue is avoiding random allocations yielding groups that differ meaningfully with respect to relevant covariates. This is a critical issue, as the chance of at least one covariate showing a "significant difference" between two treatment groups increases exponentially with the number of covariates. Haphazard Intentional Sampling is a statistical technique developed with the specific purpose of yielding sampling techniques that, on one hand, have all the benefits of standard randomization and, on the other hand, avoid exponentially large (and costly) sample sizes. This approach, proposed at Lauretto et al. [3,4] and Fossaluza et al. [5], combines intentional sampling using goal optimization techniques with random perturbations that induce good decoupling properties. In this work, we illustrate the benefits of Haphazard Intentional Sampling in the planning of statistical experiments and sampling designs, through a case study regarding the re-engineering of the network of measurement stations for atmospheric pollutants in the State of Sao Paulo, Brazil.

References:

- [1] J. Pearl, Causality: Models, Reasoning, and Inference (2000).
- [2] J.M.Stern, Cybernetics & Human Knowing 15, 49 (2008).
- [3] M.S.Lauretto, F.Nakano, C.A.B.Pereira, J.M.Stern, AIP Conf Proc 1490, 189 (2012).
- [4] M.S.Lauretto, R.B.Stern, K.L.Morgan, M.H.Clark, J.M.Stern, AIP Conf Proc 1853, 05003 (2017).
- [5] V.Fossaluza, M.S.Lauretto, C.A.B.Pereira, J.M.Stern, In: A.Polpo et al (Org.). Interdisciplinary Bayesian Statistics 1, 173 (2015).

Deconvolution for events faster than the sampling rate

D. Nille

Max-Planck-Institut für Plasmaphysik,85748 Garching, Germany

Abstract

Inverse problems are wide-spread in math, physics and wherever data are obtained. Noise and other uncertainties in the data from simulation or measurements as well as in the model typically prohibit a mathematical inversion. Instead statistical inference is required to obtain meaningful data and uncertainty quantifiers. A rather tricky inverse problem arises, when intermittent events are part of the signal, which are faster than the available sample rate. The presented example is treating heat loads from a plasma to a solid surface, where the latter acts as a memory for any previous heat influx. While diagnostics are available which can resolve the temporal timing and extent of the fast events, the main diagnostic – an infra-red camera – for quantitative heat load determination is simply too slow for direct observation. In many cases, the events are shorter than the integration time of the used camera, rendering direct inversion approaches useless and their results misguiding. It is shown, how the combination of sequential measurements can be used in a statistical framework to not only infer the amplitudes of the background heat load and the intermittent event, but also the start time and duration of each short event. This allows to quantify the energy content range of such events, which is essential for better understanding of the particle transport and confinement in magnetically confined fusion devices.

CARPETS COLOR DETECTION BASED ON THEIR IMAGES

 <u>S. Marjaneh Hossaini</u>¹, Ali Mohhamad-Djafari², Adel Mohammadpour^{1,3}, Sobhan Mohammadpour⁴, Mohammad Nadi⁵
 (1) Faculty of Math & CS, Amirkabir University of Technology marjanehosseini@aut.ac.ir
 (2) Laboratoire des Signaux & Systèmes (Supélec, CNRS, Univ. Paris Sud) ali.mohhamad-djafari@l2s.centralesupelec.fr
 (3) Dept of Mathematics & Statistics, McGill University adel.mohammadpour@mcgill.ca

(4) Dept of Computer Engineering, Sharif University of Technology mohammadpour@ce.sharif.edu

(5) eCarpetGallery, nadi@ecarpetgallery.ca

Abstract

Automatic color detection is a necessity for carpet resellers. Many well-known color detection algorithms have many shortcomings. Apart from the color itself, neighboring colors, style and pattern also affects how humans perceive color. Most if not all color detection algorithms don't take this into account. Furthermore, the algorithm needed should be invariant to change in brightness, size, and contrast of the image. In a previous experiment, the accuracy of the algorithm was half of the human counterpart. Therefore, we propose a supervised approach to reduce detection errors. We use more than 200000 images from a reseller's database as the learning set to train a Convolutional Neural Network (CNN, or ConvNet) architecture. Finally, we compare the result with other color detection algorithm

Key Words: CONVOLUTIONAL NEURAL NETWORK, COLOR DETECTION, CARPET IMAGES.

Bayesian determination of parameters for plasma-wall interactions

R. Preuss, R. A. Parra, U. von Toussaint Max-Planck-Institut f
ür Plasmaphysik, 85748 Garching, Germany

Abstract

Plasma-wall interactions are of crucial importance in the design of future fusion reactors, since they determine the replacement cycle for the plasma exposed components of the wall. In order to estimate the life-time of those wall components atomistic simulations are essential. Within a Bayesian framework we propose a non-intrusive reduced-order spectral approach (polynomial chaos expansion) to assess the uncertainty of ion-solid interaction simulations. The method not only reduces the number of function evaluations but provides simultaneously a quantitative measure for which combinations of inputs have the most important impact on the result. It is applied to SDTRIM-simulations [1] with several uncertain and Gaussian distributed input parameters, i.e. angle α , projectile energy E_0 and surface binding energy E_{sb} . In combination with recently acquired experimental data the otherwise hardly accessible model parameter E_{sb} can now be estimated.

Key Words: Uncertainty quantification, collocation, non-intrusive, plasma-wall interactions, Bayesian analysis

References

 W. Möller, W. Eckstein, J. Biersack, Tridyn - binary collision simulation of atomic collisions and dynamic composition changes in solids, Comput. Phys. Commun. 51 (3) (1988) 355–368.

USING ENTROPY TO FORECAST BITCOIN'S DAILY CONDITIONAL VALUE AT RISK

Hellinton H. Takada¹, Sylvio X. Azevedo², <u>Julio M. Stern</u>², Celma O. Ribeiro¹
(1) Polytechnic School, University of São Paulo, Brazil
(2) Institute of Mathematics and Statistics, University of São Paulo, Brazil (hellinton@gmail.com)

Abstract

Conditional value at risk (CVaR) is a risk measure for investments introduced by Rockafellar and Uryasev [1]. Yamai and Yoshiba [2] define CVaR as the conditional expectation of loss given that the loss is beyond the value at risk (VaR) level. The VaR is a risk measure that represents how much an investment might lose during usual market conditions with a given probability in a time interval. Rockafellar and Uryasev [1, 3] show that CVaR is superior to VaR in applications related to investment portfolio optimization. The Shannon entropy [4] has been used as an uncertainty measure in investments [5,6] and to forecast the Bitcoin's daily VaR [7,8]. In this paper, we estimate the entropy of intraday distribution of Bitcoin's logreturns and we forecast Bitcoin's daily CVaR using the estimated entropy.

References:

[1] R. T. Rockafellar, S. P. Uryasev, "Optimization of conditional value-at-risk," J. Risk **2**, 21–42 (2000).

[2] Y. Yamai, T. Yoshiba, "Comparative analyses of expected shortfall and valueat-risk: their validity under market stress," *Monetary and Economics Studies*, Bank of Japan, 2002.

[3] R. T. Rockafellar, S. P. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Bank. Finance* **26**, 1443–1471 (2002).

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* **27**, 379–423, (1948).

[5] M. Billio, R. Casarin, M. Costola, A. Pasqualini, "An entropy-based early warning indicator for systemic risk," *J. Int. Financ. Mark. Inst. Money* **45**, 42–59 (2016).

[6] D. T. Pele, E. Lazar, A. Dufour, "Information entropy and measures of market risk," *Entropy* **226**, 19p. (2017).

[7] S. Colucci, "On Estimating Bitcoin Value at Risk: A Comparative Analysis," *Working Paper*, 2018.

[8] D. T. Pele, M. Mazurencu-Marinescu-Pele, "Using High-Frequency Entropy to Forecast Bitcoins Daily Value at Risk," *Entropy*, **102**, 16p. (2019).

Key Words: Entropy, Conditional Value at Risk, Cryptocurrency

NESTED SAMPLING FOR ATOMIC PHYSICS DATA: THE NESTED_FIT PROGRAM

<u>M. Trassinelli¹</u>

(1) Institut des NanoSciences de Paris, CNRS, SU, Paris, France martino.trassinelli@insp.jussieu.fr

Abstract

Nested_fit is a general purpose data analysis code [1] written in Fortran and Python. It is based on the nested sampling algorithm with the implementation of the *lawn mower robot* method for finding new livepoints. The program has especially been designed for the analysis of atomic spectra where different numbers of peaks and line shapes have to be determined. For a given data set and a chosen model, the program provides the *Bayesian evidence*, for the comparison of different hypotheses/models, and the different parameter probability distributions. To give concrete illustration of applications, we consider a spectrum of examples: i) the determination of the potential presence of non-resolved satellite peaks, in a highresolution X-ray spectra of pionic atoms [2] and in a photoemission spectra of gold nanodots [3], ii) the analysis of very low-statistics spectra, in a high-resolution X-ray spectrum of He-like uranium (see fig.) [1] and in a photoemission spectra of carbon nanodots [4]. In cases where the number of components cannot be clearly identified, like for the He-like U case, we show how the main component position can anyway be determined from the probability distributions relative to the single models.



References:

- [1] M. Trassinelli, Nucl. Instrum. Methods B **408**, 301 (2017).
- [2] M. Trassinelli et al., Phys. Lett. B 759, 583-588 (2016).
- [3] A. Lévy et al., submitted to Langmuir.
- [4] I. Papagiannouli et al., J. Phys. Chem. C 122, 14889 (2018).

Key Words: Nested Sampling, Atomic Physics, Model Selection

Markov Chain Monte Carlo Algorithms For Bayesian 3D Reconstruction In Cryo-Electron Microscopy

<u>N. Vakili¹</u>, M. Habeck^{1,2}

 (1) Felix Bernstein Institute for Mathematical Statistics in the Biosciences, Georg August University Goettingen, Germany
 (2) Statistical Inverse Problems in Biophysics, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany
 (N.Vakili: nvakili@gwdg.de, M.Habeck: mhabeck@gwdg.de)

Abstract

In recent years, cryo-electron microscopy (cryo-EM) has emerged as a powerful technique to characterize the structure of large biomolecular assemblies at near atomic resolution. A crucial step in processing cryo-EM data is to reconstruct a volumetric representation of the assembly from thousands to millions of 2D noisy images which are captured in unknown orientations. The 3D reconstruction is typically represented by a voxel grid. We propose to model biomolecular assemblies as clouds of pseudo-atoms represented by mixtures of spherical Gaussian shapes. We develop a Bayesian approach for inferring point clouds from 2D projections. Our approach derives a posterior probability over the unknown pseudo-atom positions and rotational parameters, which we explore with Markov chain Monte Carlo algorithms. Drawing conformational samples from the posterior is challenging due to its high dimensionality and the existence of multiple posterior modes. These challenges can be tackled with Hamiltonian Monte Carlo (HMC) and parallel tempering. We illustrate our approach for biomolecular systems of varying size and complexity.

Key Words: Cryo-Electron Microscopy, 3D Reconstruction, Biomolecular Structure, Hamiltonian Monte Carlo, Parallel Tempering. Wednesday, July 3rd

Galilean Monte Carlo

John Skilling Maximum Entropy Data Consultants, Kenmare, Ireland skilling@eircom.net

Abstract

First introduced in imperfect form in 2011, Galilean Monte Carlo (GMC) is an algorithm for exploring a prior-probability distribution subject to a constraint, as required by likelihood modulation. Samples undergo uniform "Galilean" motion until they attempt to break the constraint, whereupon they reflect off a proxy for the boundary and proceed in a new direction.



Constrained exploration

Currently the most popular algorithm is Hamiltonian Monte Carlo (HMC), which similarly incorporates inertia though it includes forces defined by likelihood gradient. However, the properties of properly formulated GMC are superior to HMC.

Galilean Monte Carlo	Hamiltonian Monte Carlo
No rejection	Trajectories can be rejected
Every step yields a sample \mathbf{x}	Only get accepted destinations
Likelihood $L(\mathbf{x})$ is arbitrary	Likelihood must be differentiable
Step functions OK (nested sampling)	Can not use step functions
Trajectory explores uniformly	Trajectory explores nonuniformly
Needs 2 work vectors	Needs 3 work vectors

Key Words: MCMC, Galileo, Hamiltonian Monte Carlo, nested sampling.

TI-STAN: ADAPTIVELY ANNEALED THERMODYNAMIC INTEGRATION WITH HMC

 $\frac{\text{R. Wesley Henderson}^{1}, \text{Paul M. Goggans}^{1}}{(1) \text{ University of Mississippi, USA}}$ (goggans@olemiss.edu)

Abstract

We present a novel implementation of the adaptively annealed thermodynamic integration technique using Hamiltonian Monte Carlo (HMC). Thermodynamic integration with importance sampling and adaptive annealing is an especially useful method for estimating model evidence for problems that use physics-based mathematical models. Because it is based on importance sampling, this method requires an efficient way to refresh the ensemble of samples. Existing successful implementations use binary slice sampling on the Hilbert curve to accomplish this task. This implementation works well if the model has few parameters or if it can be broken into separate parts with identical parameter priors that can be refreshed separately. However, for models that are not separable and have many parameters, a different method for refreshing the samples is needed.

We find that HMC, in the form of the MC-Stan package, is effective for jointly refreshing the ensemble under a high-dimensional model. MC-Stan uses automatic differentiation to compute the gradients of the likelihood that HMC requires in about the same amount of time as it computes the likelihood function itself, easing the programming burden compared to implementations of HMC that require explicitly specified gradient functions. We present a description of the overall TI-Stan procedure and results for representative example problems.

Key Words: Model Comparison, MCMC, Thermodynamic Integration, HMC

Nested Sampling in Statistical Physics

<u>W. von der Linden¹</u>, M. Rumetshofer¹, M. Pfeifenberger²
(1) Graz University of Technology
(2) Erich Schmid Institute, Austrian Academy of Sciences (vonderlinden@tugraz.at)

Abstract

A systematic benchmark study is presented for the application of Nested Sampling (NESA) in statistical physics in the frame of the 2-D Potts model. This model exhibits a first or second order phase transition, depending on the model parameter q. The evaluation of the partition function as well as thermodynamic observables poses a major challenge to most standard numerical techniques, such as MCMC. It will be shown that NESA has a couple of important advantages: For the partition function of a system with N sites a) one NESA-run stops after O(N) moves, with a cost of $O(N^2)$ operations, b) a single run suffices to compute the partition function along with the corresponding confidence intervals, c) the confidence intervals of the logarithmic partition function decrease with 1/N, d) a single run allows to compute quantities for all temperatures and, most amazingly, there is no critical slowing down whatsoever, which is in strong contrast to other approaches. Some thermodynamic quantities, like the Helmholtz free energy, the internal energy as well as entropy and heat capacity, can be calculated in the same single NESA-run, performed for the partition function.



Magnetic properties like magnetization and magnetic susceptibility, on the other hand, require additional sampling. Results, scaling behaviour and performance are studied in detail and compared with data obtained by multi-canonical sampling, which is considered the state-of-the-art technique for the Potts model. Eventually, a parallel implementation of nested sampling is discussed.

References:

 M.J. Pfeifenberger, M. Rumetshofer, W. von der Linden J. Comput. Phys. 375, 368-392 (2018)

Key Words: Nested sampling, Potts model, Statistical physics

METRIC GAUSSIAN VARIATIONAL INFERENCE

<u>J. Knollmüller^{1,2}, T. Enßlin^{1,2}</u>

(1) Max-Planck Institute for Astrophysics, Garching(2) Ludwig-Maximilian University, Munich

Abstract

We present Metric Gaussian Variational Inference (MGVI) [1], a method to approximate complex, extremely high-dimensional posterior distributions variationally with a Gaussian. In this approach cross-correlations between all parameters are taken into account, without the necessity to explicitly parametrize the covariance. MGVI achieves this by approximating the covariance with the inverse Fisher metric locally, which is a lower bound to the true covariance and can be motivated from a information geometrical perspective. This covariance can be expressed in terms of implicit operators, which overcomes the necessity of explicit storage. By using Conjugate Gradients the covariance can be approximately applied to vectors and independent samples from the approximate posterior can be drawn efficiently. The accuracy is comparable to an approximation with fully parametrized covariance, while scaling as a mean-field approaches. The parameters required for MGVI scales linearly with the problem size, allowing its application in extremely high dimensional and complex inference problems. We validate the method against Hamiltonian Monte Carlo and present real-world applications.

[1] J. Knollmüller et al. ArXiv 1901.11033 (2019).

INFINITE SWAPPING IN BAYESIAN INVERSE PROBLEMS

Jonas Latz¹, Juan Pablo Madrigal Cianci^{*2}, Fabio Nobile², Raúl Tempone^{3,4}

(1) Zentrum Mathematik, Technische Universität München, Garching b.M., Germany

(2) CSQI, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland.

(3) Computer, Electrical and Mathematical Sciences and Engineering, KAUST, Thuwal, Saudi Arabia.

(4) Fachgruppe Mathematik, Fakultät 1, RWTH Aachen, Germany.

Abstract

Sampling from complicated posterior measures (e.g., high dimensional, multi-modal, or concentrated around a manifold) arising on Bayesian inverse problems (BIP) can be a challenging task for some Markov chain Monte Carlo (MCMC) algorithms such as Metropolis-Hastings. One way of improving the sampling efficiency is to consider a tempering approach such as parallel tempering (PT) [3], where one essentially runs multiple chains at different "temperatures" in parallel, proposing to swap states between chains every N_s steps of the algorithm. Intuitively, tempered chains tend to mix faster, and as such, the swapping step helps to better explore the target measure.

The so-called infinite swapping algorithm (IS) [1,2], originally proposed in the molecular dynamics community, can be understood as an improvement over the PT algorithm. As opposed to PT, IS is a continuous-time Markov process and considers the limit where states between two parallel Markov chains are swapped infinitely often. It is shown in [1], that in a continuous time setting, such approach can be in turn understood as a swap of dynamics, i.e., tempering parameter and Markov kernel (as opposed to states, as in PT) between chains.

In the current work, we present a Markov chain Monte Carlo (MCMC) version of the IS mechanism and use it in the context of a Bayesian inverse problem arising on a seismic source inversion scenario. We implement such algorithm and compare its efficiency with that of more traditional MCMC techniques, such as parallel tempering and random walk Metropolis.

References:

[1] Paul Dupuis, Yufei Liu, Nuria Plattner, and J.D Doll. On the infinite swapping limit for parallel tempering. Multiscale Modeling & Simulation, 10(3):9861022, 2012.

[2] Nuria Plattner, J.D Doll, Paul Dupuis, Hui Wang, Yufei Liu, and J.E Gubernatis. An infinite swapping approach to the rare-event sampling problem. The Journal of chemical physics, 135(13):134111, 2011.

[3] David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. Physical Chemistry Chemical Physics, 7(23):39103916, 2005.

Key Words: Bayesian inversion; parallel tempering; measurement noise; Markov Chain Monte Carlo; uncertainty quantification.

^{*}Corresponding author. juan.madrigalcianci@epfl.ch
Interaction between Model based Signal and Image processing, Machine Learning and Artificial Intelligence

Ali Mohammad-Djafari

International Science Consulting & Training (ISCT) 10 rue de Montjay, 91440 Bures sur Yvette, France https://sites.google.com/view/isct

Abstract

Signal and image processing has always been the main tools in many area and in particular in Medical and Biomedical applications. Nowadays, there are great number of toolboxes, general purpose and very specialized, in which, classical and advanced technics are implemented and can be used: all the transformation based methods (Fourier, Wavelets, Radon, Abel, ... and much more) as well as all the Model Based and iterative regularization methods. Statistical methods have also shown their success in some area when parametric models are available.

Bayesian inference based methods had great success, in particular, when the data are noisy, uncertain, some missing and some outliers and where there is a need to account and to quantify uncertainties.

In some applications, nowadays, we have more and more data. To use these "Big Data" to extract more knowledge, the Machine Learning and Artificial Intelligence tools have shown success and became mandatory. However, even if in many domains of Machine Learning such as classification and clustering these methods have shown success, their use in real scientific problems are limited. The main reasons are twofold: First, the users of these tools can not explain the reasons when the are successful and when they are not. The second is that, in general, these tools can not quantify the remaining uncertainties.

Model based and Bayesian inference approach have been very successful in linear inverse problems. However, adjusting the hyperparameters is complex and the cost of the computation is high. The Convolutional Neural Networks (CNN) and Deep Learning (DL) tools can be useful for pushing farther these limits. At the other side, the Model based methods can be helpful for the selection of the structure of CNN and DL which are crucial in ML success.

In this work, first I give an overview and a survey of the aforementioned methods and explore the possible interactions between them.

Key Words: Signal processing, Image processing, Bayesian inference, Machine Learning, CNN, DL, AI

RANDOMNESS IN THE COURTHOUSE

Olivia Terence Saa¹, Julio Michael Stern¹ (1) IME-USP – University of São Paulo e-mails: olivia@ime.usp.br, jstern@ime.usp.br

Abstract

Randomization is an integral part of well-designed statistical trials, and is also a required procedure in legal systems [3]. Implementation of honest, unbiased, understandable, secure, traceable, auditable and collusion resistant randomization procedures is a mater of great legal, social and political importance [5].

Given the juridical and social importance of randomization, it is important to develop procedures in full compliance with the following desiderata: (a) Statistical soundness and computational efficiency [2]; (b) Procedural, cryptographical and computational security [1,4]; (c) Complete auditability and traceability; (d) Any attempt by participating parties or coalitions to spuriously influence the procedure should be either unsuccessful or be detected; (e) Open-source programming; (f) Multiple hardware platform and operating system implementation; (g) User friendliness and transparency; (h) Flexibility and adaptability for the needs and requirements of multiple application areas (like, for example, clinical trials, selection of jury or judges in legal proceedings, and draft lotteries) [3].

This paper presents a simple and easy to implement randomization protocol that assures, in a formal mathematical setting, full compliance to the aforementioned desiderata for randomization procedures.

References:

[1] D. M. Goldschlag, S. G. Stubblebine (1998). Publicly Verifiable Lotteries: Applications of Delaying Functions. p.214-226 in International Conference on Financial Cryptography. Berlin: Springer.

[2] Haramoto, H., M., Matsumoto, T., Nishimura, F., Panneton, and P. L'Ecuyer. 2008. Efficient Jump Ahead for F2-Linear Random Number Generators. INFORMS Journal on Computing, 20 (3): 290-298.

[3] Marcondes, Diego; Peixoto, Cláudia; Stern, Julio Michael (2019). Assessing Randomness in Case Assignment: The Case Study of the Brazilian Supreme Court. *Law, Probability and Risk*, https://doi.org/10.1093/lpr/mgz006

[4] C. Pierrot, B. Wesolowski (2016) Malleability of the blockchain's entropy. https://eprint.iacr.org/2016/370.pdf

[5] Stern, J. M. (2018) Verstehen (causal/interpretative understanding), Erklären (law-governed description/ prediction), and Empirical Legal Studies. Journal of Institutional and Theoretical Economics (JITE), 174, 105–114.

Key Words: Blockchain entropy, Statistical randomization, Judicial sortition.

Adaptive Tuning Of Hamiltonian Monte Carlo Within Sequential Monte Carlo

<u>A. Buchholz</u>¹, N. Chopin², P. E. Jacob³
(1) MRC-BSU University of Cambridge
(2) ENSAE-CREST
(3) Dept. of Statistics, Harvard University
(alexander.buchholz@mrc-bsu.cam.ac.uk)

Abstract

Since the work of Del Moral et al. (2006) Sequential Monte Carlo (SMC) samplers have become a widely adapted tool for Bayesian inference. However, the quality of the generated samples depends strongly on the choice of the proposal kernel. The default choice of a random walk Metropolis Hastings kernel (RWMH) collapses as the dimension increases. Hamiltonian Monte Carlo (HMC) on the other side scales better with the dimension, but its performance depends on the tuning of some hyper parameters. In this paper we explore the idea of taking advantage of the cloud of particles to adaptively tune the parameters of a HMC kernel within SMC. We build upon the methodology developed by Fearnhead et al. (2013) and suggest another heuristic for choosing the number of move steps to assure a good mixing of the kernel. We illustrate that the obtained tuning procedure achieves a performance comparable to NUTS, see Hoffman and Gelman (2014). Moreover, we show that the sampler is more robust to multi modality and allows the calculation of normalising constants in high dimension. This is underlined by a simulation study on a normal log cox model and a Bayesian binary regression.

References

- Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.
- Fearnhead, P., Taylor, B. M., et al. (2013). An adaptive sequential monte carlo sampler. *Bayesian Analysis*, 8(2):411–438.
- Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Key Words: Hamiltonian Monte Carlo, Sequential Monte Carlo, Normalising Constants, Tuning

ELECTROMAGNETIC INDUCTION AND RELATIVISTIC DOUBLE LAYER: MECHANISM FOR BALL LIGHTNING FORMATION

Arthur H. Baraov Hewlett-Packard Inc.

Abstract

What is the probability that ball lightning is a real phenomenon of Nature? Proponents of the frequency theory of probability would say that this question is not a subject of probability theory, so it cannot answer it. Those who share E. T. Jaynes' take on the theory of probability as an extension of deductive logic will say without hesitation: The answer depends on your prior information. If you are one of those lucky men who had a close encounter with a ball lightning and escaped unscathed, your probability that ball lightning is a real phenomenon equals, of course, unity. On the other hand, if you are a theoretical physicist deeply involved in the problem of controlled thermonuclear fusion, your probability is likely to be zero. The probability can be anywhere in between for all others.



In this presentation, which is based partially on experimental observations, an attempt to raise the likelihood of the reality of ball lightning phenomenon for everyone, professional scientists in the field of plasma physics included, is being made. Ball lightning is conceived here as a stable formation comprised of electrically neutral core plus one or more nested sheaths, each of which is a double electric layer with a voltage drop of million volts or so. The central core and the areas between the sheaths form a lion's share of ball lightning's volume which is filled with weakly ionized air at roughly atmospheric pressure. A number of heuristic arguments backed up by calculations is advanceded in support of a hypothesis that powerful electromagnetic induction, which always accompanies lightning bolt, is a mechanism responsible for charging ball lightning with immense amount of energy that gets accumulated in double electric layers – partly in the form of electrostatic energy but mainly in the form of kinetic energy of ultra-relativistic electrons.

Inference and Regression in Ramachandran Space: MaxEnt Characterization of the Protein Backbone

C. Kauffmann¹, A. Zawadzka-Kazimierczuk², G. Kontaxis¹, R. Konrat¹

(1) Department of Structural and Computational Biology,

Max F. Perutz Laboratories, University of Vienna

(2) Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw

Abstract

Structural biology has long been dominated by the "structure-function" paradigm which ascribes a protein's function to a well-defined 3D structure. While still highly influential, this concept started to be challenged in the late 90s¹. Nowadays, the functional importance of so-called *intrinsically disordered proteins* (IDPs) is generally recognized, their characterization, however, remains challenging. To a large extent, this can be attributed to the inherent ensemble averaging of experimental observables, which are arguably more complex to interpret if a singular average structure cannot be assumed.

Quite intuitively, the information content of experimental evidence strongly depends on the complexity of the underlying ensemble. On a level of tertiary structures, where the sampling of conformational space already poses an intricate challenge, prior assumptions will heavily bias any resulting predictions². For secondary structure, however, this might not necessarily be the case.

Overlooked in IDP characterization until now, we use cross-correlated relaxation experiments to assess the ensemble in terms of its backbone dihedral angles ϕ and ψ : Distributions in (ϕ, ψ) -space appear accessible by experiments alone, suggesting that reasonable prior assumptions can be shifted sufficiently by experimental evidence. The set of experimental observables, however, needs to be chosen with care.

Practical applications of MaxEnt are of course hindered by experimental errors and inaccuracies of the forward model. In an attempt to preserve the simplicity of classical MaxEnt, we propose a (to our knowledge) novel heuristic that exploits similarities between the Lagrange Dual and Least Squares Regression, yielding a (locally) convex and low-dimensional optimization problem. The approach is successfully applied to a simple and well-known protein system, suggesting worthwhile exploration of its validity in other fields and applications.

References:

[1] P. Wright and H. Dyson. J. Mol. Biol. 293.2 (1999), pp. 321-331

[2] R. Rangan et al. J. Chem. Theory Comput. 14.12 (2018), pp. 6632-6641.

Key Words: Structural Biology, Intrinsically Disordered Proteins, NMR Spin Relaxation, MaxEnt, Inference, Regression Thursday, July 4th

Parallelizing MCMC

A. Caldwell Max-Planck-Institut für Physik, 80805 Munich, Germany

Abstract

We present ideas on the parallelization of the Markov Chain Monte Carlo approach via multi-proposal generation and via parameter space partitioning. For the former approach, we summarize developments in weighted sample generation and compare results following techniques presented in the literature. For massive parallelization via parameter space partitioning, the calculation of the marginal likelihood (evidence) is necessary and we solve this task with the Adaptive Harmonic Mean Integration (AHMI) algorithm. We describe the algorithm and it's mathematical properties, and report the results using it on multiple test cases of up to 20 dimensions.

Effects of Neuronal Noise on Neural Communication

Deniz Gençağa^{a,*}, Sevgi Şengül Ayan^b, Joel Poncha Lemayian^a

^aDepartment of Electrical and Electronic Engineering, Antalya Bilim University, 07190 Antalya, Turkey ^bDepartment of Industrial Engineering, Antalya Bilim University, 07190 Antalya, Turkey

In this work, we investigate the effects of neuronal noise sources on the communication between neurons using information-theoretical quantities, such as entropy, Mutual Information (MI) and Transfer Entropy (TE). The electrochemical communication between the neurons is maintained by neural spike trains that are fired at certain time instants as a response to the stimuli on the cells. A spike is fired when the action potential exceeds a certain threshold value. In the literature, neural networks have been mathematically modelled to analyse the relation between ionic currents during action potentials. Integrate and Fire (IF) and Hodgkin Huxley (HH) are two examples among the most common models utilized in these analyses [1].

However, there are also noise sources in the nervous system acting on the above relationships and making the modelling much more complicated. These sources are not treated as nuisance and they can also improve the communication between different neural populations, which is also known as "Stochastic Resonance" [2].

Here, we generalize models, such as HH and IF, in such a way that different neural noise sources with certain probabilistic distributions can act on the action potentials. After the presentation of these noisy models, we explore the statistical relationships between different neurons with and without noise components, by using information-theoretical approaches along with the significance levels of the estimations. Motivated by the successful results of the Bayesian approaches presented in [3], we apply MI to quantify the amount of information shared between certain neurons and then use TE [4] to decide the direction of the coupling between the neurons, i.e. a directional coupling analysis between the pairs.

Finally, we present future research directions for the integration of neuronal noise into a more general model.

References

- [1] Destexhe, Alain, Rudolph-Lilith, Michelle, Neuronal Noise, Springer, 2012.
- [2] F. Moss et al., Clinical Neurophysiology 115 (2004) 267–281.
- [3] Knuth K.H., Gençağa D., Rossow W.B. 2008. Information-Theoretic Methods for Identifying Relationships among Climate Variables, Proceedings of the NASA Earth-Sun System Technology Conference (ESTC 2008), Adelphi, MD.
- [4] Gençağa D. (Ed.), Transfer Entropy, special issue, MDPI Switzerland, 2018, ISBN 978-3-03842-919-7.

*Corresponding author: Tel.: +90 242 245 1394. E-mail address: deniz.gencaga@antalya.edu.tr (Deniz Gençağa)

Bayesian Identification of Dynamical Systems

<u>R.K. Niven¹</u>, M. Abel², M. Quade², L. Cordier³, A. Mohammad-Djafari⁴

(1) The University of New South Wales, Northcott Drive, Canberra, Australia.

(2) Ambrosys GmbH, Potsdam, Germany.

(3) Institut Pprime, Poitiers, France.

(4) Laboratoire des signaux et systèmes (L2S), CentraleSupélec, Gif-sur-Yvette, France.

Abstract

Many inference problems involve a dynamical system, as represented by

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}) \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the state vector and $\mathbf{f} \in \mathbb{R}^n$ is the system function or model. Since the time of Newton, researchers have pondered the problem of *system identification*: how should the user accurately and efficiently identify the model \mathbf{f} – including its functional family and parameter values – from discrete time-series data? For linear systems, an assortment of methods are available including linear regression, Kalman filter and autoregressive moving averages. However, for nonlinear models, such methods can lead to serious errors. In recent years, many machine learning tools have been developed for nonlinear model discovery to overcome this deficiency, usually based on neural networks or evolutionary computational methods, or various classification or order reduction schemes. The first group, while very useful, provide "black box" solutions which are not readily adaptable to new situations, while the second group necessarily involve the sacrificing of data resolution to achieve order reduction.

To address this problem, we propose the use of inverse Bayesian methods for model identification from time-series data. For a typical dynamical system (1), the methods become mathematically identical to Tikhonov regularization, albeit with clear theoretical connections between its residual and regularisation terms, and respectively the Bayesian likelihood and prior. This insight justifies the choice of regularisation term, and can be used to optimise performance and calculate the model uncertainty. Several Bayesian methods including joint maximum a-posteriori (JMAP) and variational Bayesian approximation (VBA) methods are demonstrated by comparison to standard regularization tools, by application to the Lorenz and other dynamical systems.

Key Words: Bayesian inference, dynamical systems, system identification, regularization, sparsification

Quantum Trajectories in Entropic Dynamics

Nicholas Carrara State University of New York at Albany.

Abstract

Entropic Dynamics[1] is a framework for deriving quantum theory with an emphasis on entropic inference as its foundation. When conducting inference, we often appeal to symmetries, and in the case of Entropic Dynamics these symmetries are the various symplectic and metric structures associated with a statistical manifold. The symplectic symmetries give rise to a Hamiltonian dynamics, which coincides with the entropic updating. From here one obtains a clearer picture of what quantum theory is all about, namely, QT is a Hamiltonian dynamics of probabilities. While the interpretation leads to the same macroscopic dynamics as quantum theory, as does other interpretations of quantum mechanics, it also predicts a dynamics of the microstates as being a type of Brownian motion of point particles. This is much like the predictions of stochastic mechanics[2] which also assumes a fluctuating particle as the basis of the microstates. The uniqueness of ED as an entropic inference of particles allows one to smoothly transition between fluctuating particles and the smooth trajectories assumed in Bohmian mechanics[3]. Thus, ED is a generalization of theories of particles which have definite yet unknown positions. Beyond the simple motions of point particles one obtains in a scalar theory, ED has also had recent success in its application of spin [4]. Spin in ED enters not as an extra *internal* degree of freedom of the particle, but rather as a constraint on the motion of the particle (the microstates are still just the positions). This extra contribution to the motion gives interesting insights which we wish to explore in a practical setting.

In this work we explore the consequences of the ED framework by studying the trajectories of particles in the continuum between stochastic and Bohmian limits in the context of many physical examples. These examples include; the Stern-Gerlach experiment, the Aharanov-Bohm effect, the double slit experiment, the Einstein-Podolosky-Rosen experiment, etc. We implement a computational framework using various state of the art methods, such as adaptive unstructured grids, to compute evolutions of the microscopic and macroscopic variables for arbitrary systems of N particles. These methods have been developed mostly by those studying Bohmian trajectories[5] and can be used to solve quantum mechanical problems without perturbative or Quantum Monte Carlo methods[6, 7, 8]. In extending these trajectories to those derived from Entropic Dynamics, we hope to gain further insight into the physics of many particle systems, but also to improve the current trajectory methods[8].

References

- [1] A. Caticha. Entropic dynamics: Quantum mechanics from entropy and information geometry. Annalen der Physik special issue on "Physics of Information", 2017.
- [2] E. Nelson. Quantum Fluctuations. Princeton Series in Physics, 1985.
- [3] P.R. Holland. The Quantum Theory of Motion. Cambridge University Press, 1993.
- [4] A. Caticha and N. Carrara. The entropic dynamics of spin. Forthcoming, 2019.
- [5] R.E.Wyatt. *Quantum Dynamics with Trajectories*. Springer: Interdisciplinary Applied Mathematics, 2005.
- [6] W.C. Schieve T.M. Coffey, R.E. Wyatt. Monte carlo generation of bohmian trajectories. J. Phys. A: Math. Theor., 41, 2008.
- [7] R.E. Wyatt and E.R. Bittner. Quantum mechanics with trajectories: Quantum trajectories and adaptive grids. Computing in Science and Engineering IEEE/AIP special issue on computational chemistry, 2003.
- [8] A. Sanz J. Mompart A. Benseny, G. Albareda and X. Oriols. Applied bohmian mechanics. Eur. Phys. J. D, 68, 2014.

An Entropic approach to Geometrodynamics

Selman Ipek* and Ariel Caticha

Department of Physics, University at Albany, NY 12222, USA

In the Entropic Dynamics (ED) framework quantum theory is derived as an application of entropic methods of inference. The physics is introduced through appropriate choices of variables and of constraints that codify the relevant physical information.

In previous work [1], a manifestly covariant ED of quantum scalar fields in a fixed background spacetime was developed. Manifest relativistic covariance was achieved by imposing constraints in the form of Poisson brackets and of intial conditions to be satisfied by a set of local Hamiltonian generators. Our approach succeeded in extending to the quantum domain the classical framework that originated with Dirac and was later developed by Teitelboim and Kuchar.

In the present work the ED of quantum fields is extended further by allowing the geometry of spacetime to fully partake in the dynamics. The result is a first-principles ED model that in one limit reproduces quantum mechanics and in another limit reproduces classical general relativity.

Our model shares many formal features with the so-called "semi-classical" approach to gravity. Such models have in the past been criticized on the basis of arguments that relied heavily on old interpretations of quantum mechanics and of the measurement problem that have themselves been the subject of controversy. The new perspective offered by our approach to geometrodynamics strongly suggests that an appropriately entropic version of semi-classical gravity remains a viable candidate for unifying quantum and gravitational effects.

[1] S. Ipek. M. Abedi, and A. Caticha, "Entropic Dynamics: Reconstructing Quantum Field Theory in Curved Space-Time"; arXiv:1803.07493 [gr-qc].

*Corresponding author: Email: sipek@albany.edu

LEARNING FROM DATA&PRIOR

Carlos C. Rodriguez

Department of Mathematics and Statistics The University at Albany. SUNY Albany, New York 12222 http://omega.albany.edu:8008

Abstract

Data is prior information and prior information is more data. Meaningful data is always a two part object. The observable (arbitrary) label and the (unobservable) label for the probability distribution that generated the observed label. A meaningful data space is always a DataTheory space containing objects (x, p). We identify the main category of these spaces by defining the morphisms to be sufficient transformations $(x, p) \to (x', p')$, such that from (x', p') we can recover (\tilde{x}, p) where \tilde{x} and x come from the same unobservable p. The only known measures of separation between (unormalized) distributions that are invariant under sufficient transformations were identified by Chentsov and Amari [Va] as the one parameter class of δ -information deviations. The δ -deviations allow us to assign a scalar $A(t,\eta)$ to a pair of distributions where t = t(y) is interpreted as the true distribution of labels and $\eta = \eta(p)$ as a prior over the model. The invariant scalar $A(t, \eta)$ measures the information in η about the true distribution t given the model. The critical points (t,η) of the invariant action produce the most ignorant prior distributions given the model and a guess at the true distribution. When a sample of data is observed, the true distribution is replaced by the empirical producing new ways for processing the observations that include and extend maximum likelihood and bayesian inference. Finally the general theory is applied to the simple logistic regression model, producing new targets and penalties that are shown to out perform the standard methods in simulations using TensorFlow.

Key Words: Information Geometry, Geometric Theory of Ignorance, Maximum Honesty, Deep Learning, Abstraction Sequence, Statistical Manifold, Statistical Category.

The Information Geometry of Space-Time

Ariel Caticha*

Department of Physics, University at Albany, NY 12222, USA

The method of maximum entropy is used to construct a model of curved physical space in which points are blurred; they are defined with a finite resolution. Such a blurred space is a statistical manifold and therefore it is automatically endowed with a metric given by information geometry. Thus *information geometry explains physical distance*.

The dynamics of blurred space, or geometrodynamics, is constructed by requiring that as space undergoes the deformations associated with time evolution, it sweeps a four-dimensional space-time. This reproduces Einstein's equations for gravity with no sources.

Blurred space has a curious hybrid character: it exhibits features that are typically associated to discrete spaces and also to continuous manifolds. For example, there is a minimum length, the volume of a region of space is a measure of the number points within it, and the entropy of space is measured by its volume.

I show that under Lorentz transformations the minimum length suffers a dilation. This "blur dilation" is a new relativistic phenomenon. It is more analogous to a time dilation than to the familiar length contraction.

* Email: acaticha@albany.edu

Against Information Geometry

John Skilling Maximum Entropy Data Consultants, Kenmare, Ireland skilling@eircom.net

Abstract

Information Geometry is inconsistent with basic symmetries so can have no observational consequence.

Friday, July 5th

Estimating Flight Characteristics of Anomalous Aerial Vehicles

Kevin H. Knuth^{1,2}, Robert Powell², Peter Reali² (1) Dept. of Physics, University at Albany, Albany NY, 12222, USA (2) Scientific Coalition for UAP Studies (SCU), Fort Myers FL, 33913, USA (kknuth@albany.edu)

Abstract

A number of Anomalous Aerial Phenomena (AAP) encountered by military, commercial, and civilian aircraft have been reported to be structured craft that exhibit 'impossible' flight characteristics. In this paper we consider a handful of well-documented encounters, including the 2004 encounters with the Nimitz Carrier Group off the coast of California, and estimate minimal accelerations and energies exhibited by the craft during the observed maneuvers. Estimated accelerations range from 100s of gs to 1000s of gs with no observed air disturbance, no sonic booms, and no evidence of excessive heat commensurate with even the minimal estimated energies. In accordance with observations, the estimated parameters describing the behavior of these craft are both anomalous and surprising. The extreme estimated flight characteristics reveal that these observations are either fabricated or seriously in error, or that these craft exhibit technology far more advanced than any known craft on Earth. In many cases the number and quality of witnesses, the variety of roles they played in the encounters, and the equipment used to track and record the craft favor the latter hypothesis that these are indeed technologically advanced craft. The observed flight characteristics of these craft are consistent with the flight characteristics required for interstellar travel. That is, if these observed accelerations were sustainable in space, then these craft could easily reach relativistic speeds within a matter of minutes to hours and cover interstellar distances in a matter of days to weeks, proper time.

CROSS-ENTROPY METHOD WITH FAILURE-INFORMED DIMENSION REDUCTION FOR RARE EVENT ESTIMATION

<u>F. Uribe¹</u>, Y.M. Marzouk², I. Papaioannou¹, D. Straub¹ (1) Technische Universität München (2) Massachusetts Institute of Technology

(felipe.uribe@tum.de,ymarz@mit.edu,iason.papaioannou@tum.de,straub@tum.de)

Abstract

In engineering applications ensuring satisfactory performance of the system under consideration is oftentimes the main concern. Different modes of failure can be grouped in a so-called *limit-state function* (LSF) $g: \Theta \to \mathbb{R}$, depending on a set of uncertain system parameters $\theta \in \Theta \subseteq \mathbb{R}^d$, modeled with a prior cumulative distribution $\mu(\theta)$. We are especially interested in the estimation of the *probability* of failure $p_{\mathcal{F}}$

$$p_{\mathcal{F}} = \mathbb{P}[\mathcal{F}] = \int_{\Theta} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) \, \mathrm{d}\mu(\boldsymbol{\theta}), \tag{1}$$

where $\mathcal{F} = \{ \boldsymbol{\theta} \in \mathbb{R}^d : g(\boldsymbol{\theta}) < 0 \}$ is the failure set, and $\mathbb{1}_{\mathcal{F}} : \mathbb{R}^d \to \{0, 1\}$ stands for the indicator or function, taking values $\mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}) = 1$ when $\boldsymbol{\theta} \in \mathcal{F}$, and zero otherwise.

In this work, we exploit a connection between rare event estimation and Bayesian inversion. Since $p_{\mathcal{F}}$ can be interpreted as a model evidence in the context of Bayesian inference, we can define the Radon-Nikodym derivative

$$\frac{\mathrm{d}\nu_{\mathcal{F}}}{\mathrm{d}\mu}(\boldsymbol{\theta}) = \frac{1}{p_{\mathcal{F}}} \mathbb{1}_{\mathcal{F}}(\boldsymbol{\theta}),\tag{2}$$

where $\nu_{\mathcal{F}}$ can be seen as a 'posterior-failure' distribution of the parameters given the event $\{g(\boldsymbol{\theta}) < 0\}$. This observation allows us to adapt dimensionality reduction techniques that are available in Bayesian inversion, for rare event estimation. We particularly employ the approach in [1], since we expect that the LSF is influential relative to the prior, only on a low-dimensional latent space of $\boldsymbol{\Theta}$. The resulting *failure-informed* space allow us to significantly reduce the dimensionality of the rare event estimation problem. This information is used to build effective low-dimensional biasing distribution models within the *cross-entropy method* [2]. We illustrate the efficiency of the proposed approach with a high-dimensional example problem. References:

[1] O. Zahm et al. Certified dimension reduction in nonlinear Bayesian inverse problems. ArXiv:1807.03712 eprint, (2018), pp. 1–41.

[2] I. Papaioannou et al. Improved cross entropy-based importance sampling with a flexible mixture model. Submitted to Reliability Engineering and System Safety, (2019).

Some Bayesian Equidistribution Tests

B.J Stokes^{a,*}

^aUniversity of Newcastle, NSW, Australia

By the term "equidistribution" we mean either (i) the identity of two given distributions, (ii) that a given dataset can be considered a random sample from a given (parent) distribution, or (iii) that two datasets can be considered as random samples from a common parent distribution. We develop some Bayesian tests of equidistribution, and compare the performance of the sample-vs-sample version with some corresponding frequentist tests.

The sample-vs-distribution version is applied to the resolution of two issues arising from Larry Bretthorst's paper "On the Difference in Means" [1].

All our equidistribution tests are ultimately based on the Sivia-Rawlings Bayesian Uniformity (SRBU) test developed in [2]).

[1] G.L. Bretthorst, Physics and Probability - Essays in Honour of Edwin T. Jaynes. Grandy, W.T & Milonni, P.W. (Eds). 177-194. Cambridge University Press, England, (1993)

[2] D.S. Sivia, S.G. Rawlings, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Knuth, K. et al. (Eds). AIP Conference Proceedings 803, 72-79, (2005)

*Corresponding author: E-mail address: <u>barrie.stokes@newcastle.edu.au</u>
A COMPLETE CLASSIFICATION AND CLUSTERING MODEL TO ACCOUNT FOR CONTINUOUS AND CATEGORICAL DATA IN PRESENCE OF MISSING VALUES AND OUTLIERS

<u>Guillaume Revillon</u>, Ali Mohammad-Djafari (guillaume.revillon@gmail.com; +33-6-19-97-11-17)

Abstract

Classification and clustering problems are closely connected with pattern recognition where many general algorithms have been developed and used in various fields. Depending on the complexity of patterns in data, classification and clustering procedures should take into consideration both continuous and categorical data which can be partially missing and erroneous due to mismeasurements and human errors. However, most algorithms cannot handle missing data and imputation methods are required to generate data to use them. Hence, the main objective of this work is to define a classification and clustering framework that handles both outliers and missing values. Here, an approach based on mixture models is preferred since mixture models provide a mathematically based, flexible and meaningful framework for the wide variety of classification and clustering requirements. More precisely, a scale mixture of Normal distributions is updated to handle outliers and missing data issues for any types of data. Exact inference in that Bayesian approach is unfortunately intractable, therefore a Variational Bayesian (VB) inference is used to find approximate posterior distributions of parameters and to provide a lower bound on the model log evidence used as a criterion for selecting the number of clusters. Eventually, experiments are carried out to exhibit the effectiveness of the proposed model on various datasets.

Key Words: Classification, Clustering, Mixture Models, Bayesian Framework, Outliers, Missing Data

Business Meeting