31st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering

MaxEnt 2011

Waterloo, Canada • July 10-15, 2011

PROGRAM AND ABSTRACTS

ORGANISING COMMITTEE

Philip Goyal (Chair), University at Albany (SUNY)
Adom Giffin, Princeton University
Kevin H. Knuth, University at Albany (SUNY)
Edward Vrscay, University of Waterloo

SPONSORS

UNIVERSITY AT ALBANY State University of New York THE WORLD WITHIN REACH

UNIVERSITY OF Waterloo

PERIMETER INSTITUTE FOR THEORETICAL PHYSICS

BOISE STATE UNIVERSITY

www.maxent2011.org

Sunday July 10, 2011

8:30 – 9:30 Registration

Tutorial Session 1 — Chair: John Skilling

9:30 – 10:30	Giuseppe Tenti	"Bayesian data analysis: A gentle introduction"
10:30 – 11:00	*Break*	
11:00 – 12:00	Ariel Caticha	"The Design of Probability Theory"

12:00 – 1:30 Lunch

Tutorial Session 2 — Chair: John Skilling

1:30 – 2:30	Adom Giffin	"MaxEnt: then and now"
2:30 – 3:00	*Break*	
3:00 – 4:00	Philip Goyal	"Information Physics: Towards a New Conception of Physical Reality"
4:00 – 4:30	*Break*	
4:30 – 5:30	Udo von Toussaint	"Numerical Methods in Bayesian Inference"

6:30 – 8:30 Welcome Reception

Monday July 11, 2011

7:30 – 8:15 *Registration*

8:15 – 8:30 *Welcome*

Session 1			Chair: Philip Goyal
8:30	– 9:30	Jos Uffink *(Invited Speaker)*	"Entropy, Entanglement, and Utility"
9:30	– 10:00	Kevin H. Knuth	"Quantification: from inference to physical laws"
10:00	– 10:30	Keith A. Earle	"A Master Equation approach to the `3 + 1' Dirac equation"
10:30	*– 11:00*	*Break*	
11:00	– 11:30	Fabio Mendes	"Bayesian inference in the numerical solution of Laplace's equation"
11:30	– 12:00	Udo von Toussaint	"Beyond Least Squares: Robust Data Analysis"
12:00	*– 1:30*	*Lunch*	

Session 2			Chair: Carlo Cafaro
1:30	– 3:00	Gerald H. Pollack *(Invited Speaker)*	"The Secret Life of Water: E = H_2O"
3:00	*– 3:30*	*Break*	
3:30	– 4:00	Jan Dettmer	"Sequential trans-dimensional Monte Carlo for seabed parameter inference"
4:00	– 4:30	M. Asim Mubeen	"Evidence-Based Filters for Signal Detection: Application to Evoked Brain Responses"
4:30	– 5:00	Michael Betancourt	"Cruising the Simplex: Sampling the Dirichlet Distribution with Hamiltonian Monte Carlo"
5:00	– 5:30	P. G. L. Porta Mana	"A critique of the Maximum Entropy Principle by one of its supporters"

Tuesday July 12, 2011

8:00 – 8:30 Registration

Session 3 — Chair: Robert Niven

8:30 – 9:30	Ralph D. Lorenz *(invited speaker)*	"Full steam ahead, probably"
9:30 – 10:00	Jingfeng Wang	"An Application of the Maximum Entropy Production Principle in Modeling Heat Fluxes over Land Surfaces"
10:00 – 10:30	Benjamin L. Ruddell	"Relationships Between Information Production, Shannon Entropy, Energy Fluxes, and Bounds of Variability in Land Surface Ecosystems"
10:30 – 11:00	*Break*	
11:00 – 12:00	Timothy E. Jupp *(invited speaker)*	"MaxEnt and planetary climates: surely atmospheric dynamics matter?"

12:00 – 2:00 Lunch

Session 4 — Chair: Adom Giffin

2:00 – 3:00	Robert K. Niven *(invited speaker)*	"Application of MaxEnt to Steady-State Flow Systems (and Extremum Entropy Production Principles)"
3:00 – 3:30	Deniz Gencaga	"Difficulties in estimating the information-theoretic quantities from data: a survey paper"
3:30 – 4:00	*Break*	
4:00 – 4:30	Nabin K. Malakar	"Maximum Joint Entropy and Information-Based Collaboration of Automated Learning Machines"
4:30 – 5:00	Julian L. Center, Jr.	"Calibrating and aligning a low-cost vision-inertial navigation system"

6:00 – 10:00 Poster Session & Reception

Wednesday July 13, 2011

8:00 – 8:30 Registration

Session 5			Chair: Keith Earle
8:30	– 9:30	Arieh Ben-Naïm *(invited speaker)*	"Shannon's measure of Information and the thermodynamic Entropy"
9:30	– 10:00	Antoine van de Ven	"Modeling the World by Minimizing Relative Entropy"
10:00	– 10:30	Alexis A. Toda	"Unification of maximum entropy and Bayesian inference via plausible reasoning"
10:30	*– 11:00*	*Break*	
11:00	– 11:30	Francesco Palmieri	"Consistence of sequence classification with entropic priors"
11:30	– 12:00	Kai Krajsek	"Bayesian Inference in Kernel Feature Space"

12:00– 2:00 Lunch

Session 6			Chair: Edward Vrscay
2:00	– 2:30	Yannis Kalaidizis	"Maximum entropy approach for non-supervised parameterization of intracellular vesicle tracking algorithm"
2:30	– 3:00	Mark Ebden	"Soft partitioning in networks via Bayesian Nonnegative Matrix Factorisation"
3:00	– 3:30	Richard P. Mann	"Prawns and Probability: Adventures in Learning Models for Collective Animal Behaviour"
3:30	*– 4:00*	*Break*	

4:00 – 5:00 Panel Discussion (Chair: Adom Giffin)

Thursday July 14, 2011

8:00 – 8:30 Registration

Session 7 — Chair: Philip Goyal

8:30 – 9:30	Robert W. Spekkens *(invited speaker)*	"Almost quantum theory: classical theories with a statistical restriction"
9:30 – 10:00	Marcel Reginatto	"Quantum theory from the geometry of evolving probabilities"
10:00 – 10:30	Carlo Cafaro	"On a differential geometric viewpoint of Jaynes' Maxent method and its quantum extension"
10:30 – 11:00	*Break*	
11:00 – 11:30	Ariel Caticha	"Entropic dynamics and the quantum measurement problem"
11:30 – 12:00	Robin Blume-Kohout	"Likelihood-ratio confidence intervals for quantum states"

12:00 – 1:30 Lunch

Session 8 — Chair: Udo von Touissant

1:30 – 2:30	Radford M. Neal *(invited speaker)*	"New Monte Carlo Methods Based on Hamiltonian Dynamics"
2:30 – 3:30	John Skilling *(invited speaker)*	"Computing Bayes in big spaces"

4:30 – 9:30 Banquet

Friday July 15, 2011

8:00 – 8:30 Registration

Session 9		Chair: Deniz Gencaga
8:30 – 9:00	Edward R. Vrscay	"Fractal-based measure approximation with entropy maximization and sparsity constraints"
9:00 – 9:30	Chris Ferrie	"Minimax estimators for noisy coins"
9:30 – 10:00	Marcelo S. Lauretto	"Estimation and model selection in Dirichlet regression"
10:00 – 10:30	*Break*	
10:30 – 11:00	Paul M. Goggans	"Inference-based design of FIR filters with sum of signed power-of-two coefficients"
11:00 – 11:30	Jonathan Botts	"Bayesian Inference for Acoustic Impedance Boundaries in Room-Acoustic Finite Difference Time-Domain Modeling"
11:30 – 12:00	Brian D. Ziebart	"Process-Conditioned Investing with Incomplete Information using Maximum Causal Entropy"

12:00– 2:00 Lunch

2:00 – 3:00 Business Meeting

Poster Session

Tuesday July 12, 2011   6:00pm – 10:00pm

Arthur H. Baraov	"On the notion of fair games and Bernoulli's concept of moral expectation"
Jingfeng Wang	"On Ignorance Priors and the Principle of Inference"
Sean Alan Ali	"Relating dynamical complexity to quantum entanglement via information geometry and maximum relative entropy methods"
Carlo Cafaro	"An information geometric viewpoint of algorithms in quantum computing"
Adom Giffin	"Insights into the softening of chaotic statistical models by quantum considerations"
Veronica Nieves	"Bayesian Analysis of Scale-Invariant Processes"
Jian Deng	"Maximum partial entropy principle and partial probability-weighted moments"
Noel van Erp	"Parsimonious priors for regression coefficients"
Arthur H. Baraov	"The aircraft carrier problem"
Koki Kyo	"Bayesian estimation of dynamic matching function for u-v analysis in Japan"

Subhadeep Mukhopadhyay	"From Data to Constraints"
Asif Mehmood	"Application of Bayesian Non-Negative Matrix Factorization to Seismic Footstep Signals Separation"
Marcel Reginatto	"Neutron spectrometry at high-energy accelerator facilities: a Bayesian approach using entropic priors"
Nissim Kaufmann	"A note on antidata"
Barrie Stokes	"Maxent alternatives to Pearson family distributions"
Chris Granade	"Adaptive Hamiltonian estimation using Bayesian experimental design"
Do Kester	"A software package for nested sampling"
Marcelo S. Lauretto	"The full Bayesian significance test for symmetry in contingency tables"
Marcelo S. Lauretto	"Reliability analysis in series systems: an empirical comparison between Bayesian and classical estimators"
Sha Zhu	"A hierarchical Bayesian method for synthetic aperture radar image reconstruction"
Michele Pappalardo	"Handling Uncertainty Using Game Theory"
Ryszard P. Kostecki	"Information geometric foundations of quantum theory"

Ning Chu	"Super-resolution image from a sequence of low resolution images based on improved Gauss-Markov-Potts model"
Renaldas Urniezius	"Iteration free vector orientation using maximum relative entropy with observational priors"
Shahid Nawaz	"Momentum and the uncertainty relation in the entropic approach to quantum theory"
Kevin H. Knuth	"From Cox to Emergent Geometry"
Doriano-Boris Pougaza	"New copulas obtained by maximizing Tsallis or Renyi Entropies"
Cameron J. Fackler	"Porous Material Parameter Estimation: A Bayesian Approach"
Adom Giffin	"The error in the two envelope paradox"
Haley A. Maunu	"Maximum entropy production in Daisyworld models"
Keith A. Earle	"Parameter Estimation of Magnetic Resonance Spectra via a Statistical Mechanics Approach"
Mark Ebden	"Soft partitioning in networks via Bayesian Nonnegative Matrix Factorisation"

Abstracts

BAYESIAN DATA ANALYSIS: A GENTLE INTRODUCTION

G. Tenti
Department of Applied Mathematics University of Waterloo

Abstract

The analysis of experimental data is essential to the development of scientific theories, but the type of reasoning used in formulating hypotheses and having them substantiated by the data has been the subject of great controversies for over two hundred years.

Using simple examples—such as coin-tossing—I shall illustrate the source of the disagreements and give reasons why the Bayesian approach is superior to standard ("orthodox") statistical analysis. Time permitting, I shall also show how E.T. Jaynes used Bayesian ideas to derive Fick's law of diffusion.

THE DESIGN OF PROBABILITY THEORY

Ariel Caticha
Department of Physics, University at Albany (SUNY)
(ariel@albany.edu)

Abstract

The problem of coping with uncertainty in the real world is central to all science. As a result of the work by many people a fairly satisfactory solution has been achievedalthough much controversy still remains. The solution involves, first, a scheme that allows one to represent any provisional state of knowledge, and second, a procedure that allows us to revise our beliefs as we acquire new information.

This tutorial is mostly concerned with the first stage—the design of probability theory—as pioneered by Cox. About the second stage—the updating problem—our discussion will be brief, we will only make some brief remarks on Bayes rule and its limitations.

I will follow a very pragmatic approach. Probability theory is neither true nor false; it is designed to be useful, to achieve a certain purpose, to work. Just as in engineering, one is satisfied that a solution works when it performs according to some desired "design specifications" or "design criteria". In engineering there may be many solutions that work fine—they perform the desired function, and in the end, that is all we care about. What is remarkable about the Cox approach is that the design specifications—consistency, universality—are totally restrictive. There is a unique way to handle degrees of belief and this is probability theory.

I will address some of the criticisms that have been raised against the Cox approach. Are these degrees of belief, or plausibility, or credibility, or even degrees of implication? Is there a difference? Should we use a single real number to measure a degree of belief? Can beliefs be compared? Are the Cox design criteria obvious? Are there counter examples to Cox?

Rather than justifying Cox's choices I demonstrate their robustness. I make a different choice of design criteria and derive probability theory as the unique (up to regraduations) consistent representation of the Boolean AND and OR operations.

Key Words: Probability theory, Bayesian Inference, Pragmatism

MAXENT: THEN AND NOW

Adom Giffin[1,2]
(1) Clarkson Univeristy, Potsdam, NY, USA 13699
(2) Princeton University, Princeton, NJ, USA 08544
(physics101@gmail.com)

Abstract

The principle of Maximum Entropy has had a long and controversial history. It was first used by Gibbs to formulate his maximum "ensemble" but more formally introduced by Brillouin in 1952 and then later (and more famously) by Jaynes in 1956. Since then it has had many mathematical tribulations and manifestations. This paper has three objectives: 1) To put the principle in a historical perspective regarding its origin, use, controversies and current status. 2) To pragmatically show how it functions and explain the key concepts it employs. 3) To demonstrate how it is used currently to bring traditional MaxEnt and Bayesian inference under one roof which I call the method of Maximum relative Entropy (MrE).

This paper is not intended to be an authoritative review of the principle of Maximum Entropy. It is meant to ask questions, some of which will be given possible answers. Some will remain open. However, it is my hope that it will at least provide paths for the reader to explore and find the answers themselves.

Key Words: Maximum Entropy, Bayes, Relative Entropy

Information Physics: Towards a New Conception of Physical Reality

P. Goyal
Department of Physics, University at Albany (SUNY)
(pgoyal@albany.edu)

Abstract

The central tenet of information physics is that the concept of information is as fundamental to developing an understanding of the physical universe as are the classical concepts of space and time, matter and energy.

In this talk, I shall sketch the developments—in physics and elsewhere—that have given rise to the field of information physics, and indicate some of the many rather deep insights that 'informational thinking' has provided into the structure of physical theory, in particular into the mathematical structure of quantum theory. I shall also briefly sketch the emerging conception of reality (or ontology) to which these developments seem to naturally lead.

Key Words: information, statistical physics, quantum theory

Numerical Methods in Bayesian Inference

U. von Toussaint[1]
(1) Max-Planck-Institute for Plasmaphysics,
Boltzmannstrasse 2, 85748 Garching, Germany

May 31, 2011

Abstract

Bayesian inference is very simple from a conceptual point of view: Once the likelihood and prior distributions are specified Bayes' theorem allows to derive the posterior probability for every specified parameter vector. However, in most situations the posterior distribution is required primarily for the purpose of evaluating expectation values of a function of interest $f(\boldsymbol{\theta})$ with respect to the posterior,

$$\langle f(\boldsymbol{\theta})\rangle = \int \mathrm{d}\boldsymbol{\theta}\, f(\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathbf{D}, I) = \int \mathrm{d}\boldsymbol{\theta}\, f(\boldsymbol{\theta})\, \frac{p^*(\boldsymbol{\theta})}{Z} \tag{1}$$

The normalization constant of the unnormalized distribution $p^*(\boldsymbol{\theta})$ is given by

$$Z = \int \mathrm{d}\boldsymbol{\theta}\, p^*(\boldsymbol{\theta})\,. \tag{2}$$

These integrals over the parameter space are commonly high-dimensional and analytically intractable, except in very rare circumstances, so that typically neither the expectation value nor the normalization constant are at hand - the latter the key quantity for Bayesian model comparison. Also the marginalization of parameters requires integration in often high-dimensional spaces. There are two different ways to proceed. Either the integrant of Eq. (1) is approximated by a different, more easily accessible function or the integral itself is approximated by numerical integration or by sampling (MCMC) techniques. In the tutorial the key concepts and algorithms to evaluate these integrals are presented and their respective merits are compared using real-world examples.

Key Words: Bayesian Data Analysis, Numerical Methods, Markov Chain Monte Carlo

ENTROPY, ENTANGLEMENT AND UTILITY

J. Uffink
University of Utrecht

Abstract

This talk explores a formal analogy between the study of entanglement in quantum theory, entropy in classical thermodynamics, and utility in decision theory. Roughly speaking, I will argue that in all three cases, the mathematical problem arises of finding and characterizing those functions that respect a given pre-ordering relation, subject to certain auxiliary conditions. Moreover, theorems have been obtained in these three separate areas that might be applied to them in common. It is my main purpose to draw attention to these, and argue how they might be useful in thermodynamics and quantum theory.

QUANTIFICATION: FROM INFERENCE TO PHYSICAL LAWS

Kevin H. Knuth
Departments of Physics and Informatics
University at Albany (SUNY), Albany NY USA
(kknuth@albany.edu)

Abstract

Many are aware that Richard T. Cox's contribution to the foundations of probability theory centers on the realization that the absolute truth values of the Boolean algebra of logical statements can be generalized to degrees of belief [1]. Ed Jaynes, who was inspired by both Cox and Shannon realized that the connections between communication theory and statistical mechanics did not lie in the equations themselves, but rather the ideas that led to the equations [2, p. 4].

Here I will focus on those ideas [3, 4] by discussing the way in which the concept of order is mathematically formalized using algebras, lattices and partially-ordered sets and generalized through quantification. I will briefly describe how quantifying a partially-ordered set of sequences of measurements leads to a derivation of the Feynman formulation of quantum mechanics [5, 6], which is not only consistent with, but dependent on traditional inference. In addition, I will describe how the ordering of events leads to a derivation of special relativity, and more importantly to a novel concept of space and time as emergent properties of a network of events [7]. In conclusion, I will outline our current research which considers quantum measurements as events and show how we are approaching a deep understanding of relativistic quantum mechanics.

What is surprising is that the same ideas that led to both communication theory and statistical mechanics, taken seriously, lead to quantum mechanics and relativity, and maybe more.

References:

[1] R.T. Cox, *Am. J. Physics* 14, 113 (1946).

[2] E. T. Jaynes, Probability Theory in Science and Engineering, No. 4 in Colloquium Lectures in Pure and Applied Science, Socony-Mobil Oil Co., 1956.

[3] K.H. Knuth, MaxEnt 2003, arXiv:physics/0403031v1 [physics.data-an]

[4] K.H. Knuth, MaxEnt 2009, arXiv:0909.3684v1 [math.GM].

[5] P. Goyal, K.H. Knuth, and J. Skilling. 2010. *Phys. Rev. A* 81, 022109, arXiv:0907.0909 [quant-ph].

[6] P. Goyal, K.H. Knuth. 2011. *Symmetry* 3(2):171-206.

[7] K.H. Knuth, Bahreyni N. 2010. arXiv:1005.4172v2 [math-ph]

Key Words: algebra, lattice, order, physics, poset, quantum mechanics, relativity

A Master Equation Approach to the '3 + 1' Dirac Equation

Keith A. Earle
Physics Department, University at Albany (SUNY)
1400 Washington Ave, Albany NY 12222
kearle@albany.edu, http://earlelab.rit.albany.edu

Abstract

A derivation of the Dirac equation in '3 + 1' dimensions is presented based on a master equation approach originally developed for the '1 + 1' problem by McKeon and Ord. The method of derivation presented here suggests a mechanism by which the work of Knuth and Bahreyni on causal sets may be extended to a derivation of the Dirac equation in the context of an inference problem. The relationship of the approach described here to alternative formulations, such as the unitary cellular automaton of Bialynicki-Birula and the Discrete Time Quantum Walk of Strauch will also be discussed. The formalism developed here will be applied to scattering from a potential step. The time-dependence of the Shannon entropy for the free particle and potential scattering case can be computed via the Born rule. Insights from these computations may be used to deepen understanding of quantum phenomena.

Bayesian inference in the numerical solution of Laplace's equation

Fbio Macdo Mendes[1,2], Edson Alves da Costa Jnior[1,3]
(1) University of Braslia at Gama.
(2) fabiomacedomendes@gmail.com
(3) edsonalves@unb.br

Abstract

Inference is not unrelated to numerical analysis: given partial information about a mathematical problem, one has to estimate the unknown "true solution" and error bars. Many methods of interpolation (least squares, Kriging, Tikhonov regularization, etc) have a probabilistic interpretation. O'Hagan showed that quadratures can also be constructed explicitly as a form of Bayesian inference (O'Hagan, A., BAYESIAN STATISTICS (1992) **4** , pp. 345-363) . In his framework, the integrand is modeled after a Gaussian process. By conditioning the stochastic process to the known values of the integrand in a finite set of points, can can build a reliable estimate for the value of the integral. The present work applies a similar method for the problem of solving Laplace's equation inside a closed boundary. First, one needs a Gaussian process that yields arbitrary harmonic functions. Secondly, the boundaries (Dirichilet or Neumann conditions) are used to update these probabilities and to estimate the solution in the whole domain. This procedure is similar to the widely used *Boundary Element Method*, and it is possible to recover the later as a special case. The language of Bayesian inference gives more flexibility on how the boundary conditions and conservation laws can be handled. This flexibility can be used to attain greater accuracy using a coarser discretization of the boundary. This can open doors to more efficient implementations for solvers of homogeneous parabolic equations.

Key Words: Gaussian Processes, Numerical Analysis, Elliptic Partial Differential Equations

Beyond Least Squares: Robust Data Analysis

U. von Toussaint[1], V. Dose[1]
(1) Max-Planck-Institute for Plasmaphysics,
Boltzmannstrasse 2, 85748 Garching, Germany

April 4, 2011

Abstract

We investigate in a Bayesian framework the performance of two alternative modifications of the 200 years old method of least squares. The first modification considers arbitrary real positive exponents α instead of $\alpha = 2$ in the distance measure. This modification leads to estimates that are less outlier sensitive than traditional least squares. Moreover, even when data are simulated with a Gauss random number generator the optimum exponent α may well deviate from $\alpha = 2$. The second modification consists of abandoning the assumption that data uncertainties entering the distance measure are exact. We replace this assumption by assuming that the experimentally determined uncertainties s_i are point estimates of the unknown true uncertainties σ_i. The remarkable result of this modification is a likelihood which is, unlike traditional least squares, very robust against outliers in case of inconsistent data, but approaches least squares results for consistent data. These properties render data selection by reason of their numerical value unnecessary. Two physics examples of the approach will be discussed.

Key Words: Bayesian Data Analysis, Robust Data Analysis, Least square estimation

The Secret Life of Water: E = H_2O

Gerald H. Pollack, PhD
University of Washington, Seattle
(ghp@u.washington.edu, http://faculty.washington.edu/ghp/)

Abstract

School children learn that water has three phases: solid, liquid and vapor. But we have recently uncovered what appears to be a fourth phase. This phase occurs next to water-loving (hydrophilic) surfaces. It is surprisingly extensive, projecting out from the surface by up to millions of molecular layers.

Of particular significance is the observation that this fourth phase is charged; and, the water just beyond is oppositely charged, creating a battery that can produce current. We found that light recharges this battery. Thus, water can receive and process electromagnetic energy drawn from the environment much like plants. The absorbed light energy can then be exploited for performing work, including electrical and mechanical work. Recent experiments confirm the reality of such energy conversion.

The energy-conversion framework implied above seems rich with implication. Not only does it provide an understanding of how water processes solar and other energies, but also it may provide a foundation for simpler understanding natural phenomena ranging from weather and green energy all the way to biological issues such as the origin of life, transport, and osmosis.

The lecture will present evidence for the presence of this novel phase of water, and will consider the potentially broad implications of this phase for physics, chemistry and biology, as well as some practical applications for engineering (all in one hour!).

SEQUENTIAL TRANS-DIMENSIONAL MONTE CARLO FOR SEABED PARAMETER INFERENCE

J. Dettmer[1], S.E. Dosso[1], C.W. Holland[2]
(1) University of Victoria, Victoria BC, Canada
(2) The Pennsylvania State University, State College PA, USA
e-mail: jand@uvic.ca

Abstract

This paper develops a sequential Monte Carlo algorithm for seabed parameter estimation along tracks in rapidly varying environments. Observations along tracks are in terms of seabed acoustic reflection coefficients as a function of frequency and grazing angle, which are measured using a ship-towed sound source and hydrophone array. Markov chain Monte Carlo methods are applied in combination with sequential techniques (particle filters) to carry out parameter inference for consecutive data sets acquired along a track. The environment is parametrized as a stack of sediment layers, each layer being described by compressional wave velocity, density, and attenuation. Changes in model parametrization along the track (e.g., number of sediment layers) are accounted for with trans-dimensional partition modelling which intrinsically determines the amount of structure supported by the data information content. Challenging issues of rapid environmental change between consecutive data sets and high information content (peaked likelihood) are addressed by bridging distributions implemented using annealed importance sampling. This provides an efficient method to locate high-likelihood regions for new data which are distant and/or disjoint from previous high-likelihood regions. The algorithm is applied to simulated reflection-coefficient data along a track, in which the environment varies rapidly in terms of the number of layers, layer thicknesses, and geoacoustic parameters within layers. In addition, the seabed contains a geologic fault where all layers are offset abruptly, and an erosional channel. Finally, the inversion is applied to data collected on the Malta Plateau, Mediterranean Sea, using a towed chirp-sonar source and a hydrophone array close to the seabed.

References:

[1] J. Dettmer et al. J. Acoust. Soc. Am., 129:1794-1806 (2011).

[2] J. Dettmer et al. J. Acoust. Soc. Am., 128:3393–3405 (2010).

Key Words: Bayesian inference, Sequential Monte Carlo, seabed parameters

Evidence-Based Filters for Signal Detection:Application to Evoked Brain Responses

M. Asim Mubeen[1], Kevin H. Knuth[1,2,3]
(1) Knuth Cyberphysics Lab, Dept. of Physics, University at Albany, Albany NY
(2) Department of Informatics, University at Albany, Albany NY
(3) Autonomous Exploration Inc., Andover MA

Abstract

Template-based signal detection most often relies on computing a correlation, or a dot product, between an incoming data stream and a signal template. Such a correlation results in an ongoing estimate of the magnitude of the signal in the data stream. However, it does not directly indicate the presence or absence of the signal. The problem is really one of model-testing, and the relevant quantity is the Bayesian evidence (marginal likelihood) of the signal model. Given a signal template and an ongoing data stream, we have developed an evidence-based filter that computes the Bayesian evidence that a signal is present in the data.

We demonstrate this algorithm by applying it to brain-machine interface (BMI) data obtained by recording human brain electrical activity, or electroencephalography (EEG). A very popular and effective paradigm in EEG-based BMI is based on the detection of what is called the P300 evoked brain response which is generated in response to particular sensory stimuli. The goal is to detect the presence of a P300 signal in ongoing EEG activity as accurately and as fast as possible. Our algorithm uses a subject-specific P300 template to compute the Bayesian evidence that a sliding window of EEG data contains the signal. The efficacy of this algorithm is demonstrated by comparing receiver operating characteristic (ROC) curves of the evidence-based filter to the usual correlation method. Our results show a significant improvement in single-trial P300 detection. The evidence-based filter promises to improve the accuracy and speed of the detection of evoked brain responses in BMI applications as well the detection of template signals in more general signal processing applications.

Cruising the Simplex: Sampling the Dirichlet Distribution With Hamiltonian Monte Carlo

Michael Betancourt
Massachusetts Institute of Technology
betan@mit.edu

Abstract

The allocation of a conserved quantity is a common feature of many problems in modern statistics, ranging from categorical systems to mixture models and non-parametric estimation. Given its appropriate support, the Dirichlet distribution is a typical component of these analyses, but the very support that makes it useful also makes it difficult to incorporate into compound models. I present a series of transformations that reshapes the canonical Dirichlet distribution to admit efficient Markov Chain Monte Carlo sampling and demonstrate the utility of the sampling with applications common to many physics analyses.

Key Words: Dirichlet Distribution, Markov Chain Monte Carlo

A CRITIQUE OF THE MAXIMUM-ENTROPY PRINCIPLE BY ONE OF ITS SUPPORTERS

P. G. L. Porta Mana
Perimeter Institute for Theoretical Physics, Waterloo

Abstract

What is the relationship between Bayesian theory and the principle of maximum entropy? When does the principle give unreasonable or wrong results? When is it appropriate to use the rule 'expectation = average'? Can Bayesian theory give the same answers as the principle, and better answers when those of the principle are unreasonable?

A supporter of the principle tries to answer these questions by comparing, in a couple of very simple dice-throwing problems, the numerical results given by Bayesian theory and by the principle. And in so doing, he almost becomes an apostate.

FULL STEAM AHEAD, PROBABLY

R. D. Lorenz[1]
(1) JHU Applied Physics Laboratory, Laurel, MD 20723, USA
(ralph.lorenz@jhuapl.edu. http://www.lpl.arizona.edu/ rlorenz)

Abstract

Could the Earth's climate state be selected by thermodynamics [1] and information theory? In particular, there is no a priori reason that the equator to pole temperature gradient, which drives and is determined by the poleward heat transport in the oceans and atmosphere, should be the value we observe. However, it appears to be at a value where the production of thermodynamic entropy by that heat flow is maximized (more or less equivalent to maximizing its potential work output.) This Maximum Entropy Production (MEP) state appears to hold on Titan and, to some extent, Mars, even though their climates are very different in character[2,3].

One rationale is that the greatest number of possible combinations of modes of heat transport (Hadley circulation, eddies, ocean currents, etc.) at steady state will exist where that dissipation is maximized. In other words, that the MEP state is most probable, subject to the constraints that are applied to the system. Chief among these constraints are the pressure (or column mass) of the atmosphere, and the planetary rotation rate. It is recognition of these constraints on MEP that may allow the reconciliation of MEP approaches with a more conventional dynamical meteorological perspective. One obvious application of the principle may be to exoplanets, for which there are generally very little data to constrain more elaborate models.

This talk will review the MEP idea and related topics in planetary science, including the size spectrum of dust devils observed on Mars and Earth. This size distribution may be described by an exponential (suggested on MaxEnt grounds [4]), or perhaps a power law[5], which may arise from self-organized criticality which may in turn be associated with MEP[6].

References:

[1] G. W. Paltridge. Q. J. Roy. Met. Soc. **101**, 475 (1975).
[2] R. D. Lorenz et al. Geophys. Res. Lett. **28**, 415 (2001).
[3] R. D. Lorenz et al. Science. **299**, 837 (2003).
[4] M. V. Kurgansky. Geophys. Res. Lett **33**, L19S06 (2006).
[5] R. D. Lorenz. Icarus **203**, 683 (2009).
[6] R. Dewar, J. Phys. A. **36**, 631 (2003).

Key Words: Climate, Thermodynamics, Maximum Entropy Production

An Application of the Maximum Entropy Production Principle in Modeling Heat Fluxes over Land Surfaces

J. Wang[1], R. L. Bras[2]
(1) University of California, Irvine, CA 92697, USA
(2) Georgia Institute of Technology, Atlanta, GA 30332, USA
([1]jingfenw@uci.edu, [2]rlbras@gatech.edu)

22 March 2011

Abstract

A model of evaporation, sensible and ground heat fluxes over land surfaces is proposed based on the theory of maximum entropy production (MEP) [1,2]. A key element of the proposed MEP model is the formulation of a dissipation (or entropy production) function, whose extremization under the constraint of conservation of energy leads to a unique partition of net radiation into latent, sensible and ground heat fluxes. The solutions are functions of surface soil temperature, surface specific humidity (or its equivalents) and net radiation. The MEP model predicts that surface heat fluxes vary monotonically and smoothly with soil wetness ranging from dry to saturation. The MEP model of transpiration over vegetated surfaces is shown to be a special case of non-vegetated surfaces. A test of the MEP model using field observations indicates that the model performs well bare soils under all soil moisture conditions and no-water-stress canopy. The proposed MEP model is an effective tool in modeling the energy budget over a land surface due to its unique features: (1) the model is built on the state-of-the-art non-equilibrium thermodynamics, (2) the model only needs input of surface variables (i.e. temperature, humidity or its equivalents, and net radiation), (3) the model covers the entire range of water state from dryness to saturation, (4) all model parameters are either physical properties of the system or universal empirical coefficients, and (5) the model formulation allows a unique solution of the fluxes with reduced sensitivity to the uncertainties in the model parameters.

References:

[1] J. Wang et al. Water Resour. Res. **49**, W03521 (2011).

[2] J. Wang et al. Water Resour. Res. **47**, W11422 (2009).

Key Words: maximum entropy production, surface heat fluxes, analytical solution

Relationships Between Information Production, Shannon Entropy, Energy Fluxes, and Bounds of Variability in Land Surface Ecosystems

Benjamin L. Ruddell[1]

(1) Arizona State University, Tempe, AZ

(bruddell@asu.edu, 480-727-5123)

Recent advances in applications of information theory to ecosystem flux data have highlighted the connection between ecosystem productivity, phenology, and information production in a wide range of land surface vegetated ecosystems. We will present a series of preliminary findings expanding on the relationship between these established patterns, adding insights into their origins in the fluxes of energy and the extreme bounds of variability of these systems. The findings have possible applications in the identification of emergent patterns that can be used to predict the adaptation of land surface ecosystems, and other complex systems, under changing climate and forcing dynamics.

MaxEnt and planetary climates: surely atmospheric dynamics matter?

T. E. Jupp[1]
(1) Mathematics Research Institute, University of Exeter, Exeter, UK
(t.e.jupp@exeter.ac.uk)

Abstract

Equator–to–pole heat transport in terrestrial planets results from complex atmospheric motions. Nonetheless, the macroscopic features of this transport can often be predicted simply by applying energy conservation and appealing to the MaxEnt formalism (which in this case corresponds to maximising the rate of thermodynamic entropy production – MEP). The apparent irrelevance of fluid dynamics is worrying - especially to fluid dynamicists. In this talk I shall present some recent results (Jupp & Cox, 2010) which suggest that dynamical constraints do not affect MaxEnt results for Earth, Mars, Venus and Titan and are, in this sense, irrelevant. For planets with different properties, however, it is shown that dynamical constraints would indeed affect the MaxEnt state of the atmosphere.

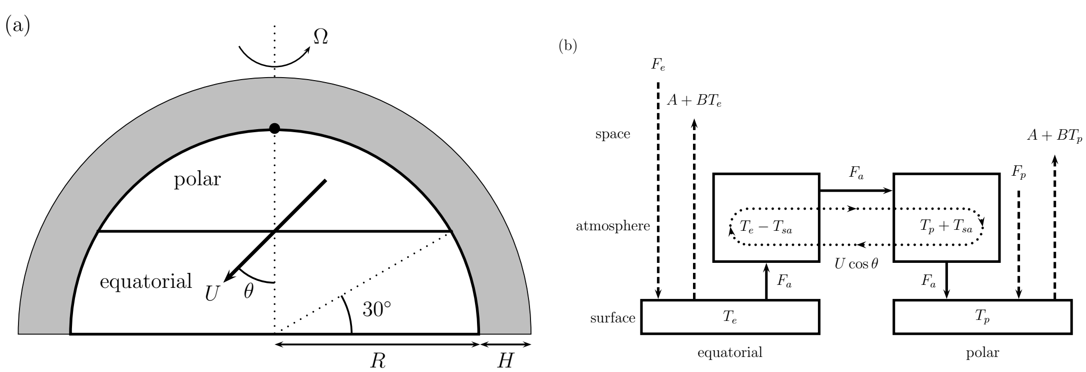


Figure 1: Simple model for equator–to–pole heat transport. (a) A surface wind U blows from pole to equator through an atmosphere of thickness H. (b) Schematic representation of the model. Dashed arrows – radiative energy fluxes, solid arrows – atmospheric energy fluxes, dotted arrows – atmospheric circulation.

References:

[1] T.E. Jupp & P.M. Cox Phil. Trans. R. Soc. B **365**, 1355 – 1365 (2010).

Key Words: MEP, planet, atmosphere, entropy

Application of MaxEnt to Steady-State Flow Systems (and Extremum Entropy Production Principles)

Robert K. Niven[1,2]
(1) School of Engineering and Information Technology,
The University of New South Wales at ADFA, Canberra, ACT 2600,
Australia. (r.niven@adfa.edu.au)
(2) Institut Pprime, CNRS / Universit de Poitiers / ENSMA,
CEAT, Poitiers, France.

15 April 2011

Abstract

Recently, the author gave a MaxEnt-based analysis of steady-state flow systems, using an entropy defined on the set of instantaneous fluxes through an infinitesimal fluid element [1,2]. The formulation is analogous to Gibbs' formulation of equilibrium thermodynamics [3], which expresses the effect of changes in entropy within and outside a system, but is here applied to the steady state of a non-equilibrium flow system. The analysis yields a potential function (negative Massieu function, analogous to a free energy) to be minimised; this in turn can be *approximated* by a maximum or minimum entropy production (MaxEP or MinEP) principle in different circumstances. In this seminar, a generic version of the derivation is first provided, encompassing three seemingly disparate formulations of equilibrium thermodynamics [3], local steady-state flow [1-2] and global steady-state flow [4-5]. The mathematical structure of the analysis, in consequence of Jaynes' framework [6], is first examined, leading into a discussion of the possibility and implications of a scale invariance condition for the application of MaxEnt to flow systems. The consequences of the analysis for several systems are also considered, including (i) the transition between laminar and turbulent flow in a pipe, and (ii) the modelling of planetary climate systems, including solar and extrasolar planets.

References:

[1] R.K. Niven (2009) Phys. Rev. E 80(2): 021113.
[2] R.K. Niven (2010) Phil. Trans. B, 365: 1323-1331.
[3] H.B. Callen, Thermodynamics, 2nd ed., John Wiley, NY, 1985.
[4] R.C. Dewar, J. Phys. A Math. Gen. 36, 631 (2003).
[5] R.C. Dewar, J. Phys. A Math. Gen. 38, L371 (2005).
[6] E.T. Jaynes (G.L. Bretthorst, ed.) Probability Theory: The Logic of Science, Cambridge U.P., Cambridge, 2003.

Key Words: MaxEnt, non-equilibrium system, entropy production, steady state, fluid flow, heat flow

DIFFICULTIES IN ESTIMATING THE INFORMATION-THEORETIC QUANTITIES FROM DATA: A SURVEY PAPER

Deniz Gencaga[1], William B. Rossow[1], Kevin H. Knuth[2]
(1) NOAA-CREST, The City College of New York, NY, USA.
(2) Departments of Physics and Informatics, University at Albany, NY, USA.

Abstract

To quantify the amount of information about a variable, to quantify the amount of information shared between two variables, or even to quantify the amount of information shared along with its direction between coupled variables in a dynamical system, we utilize information-theoretic quantities like entropy, mutual information and transfer entropy, respectively. Although the literature is so rich in methods on estimating these quantities from observational data, to the best of our knowledge these quantities can only be estimated with certain bias and variance, creating a huge difficulty for data users in their fields. Here, we demonstrate main techniques and fill a major gap in the literature by presenting these methods used in very diverse fields from biomedicine [1] to health monitoring in engineering systems [2]. First, we present fixed-width bin histogram based methods [3] and our Bayesian version along with error-bars. Later we demonstrate variable-width bin histogram based methods [4] for mutual information and transfer entropy estimations. Finally, we describe how these quantities are estimated using kernel density estimation techniques [5]. In addition to these, we also demonstrate, how addition and subtraction operations involved in the estimation of the computation of information-theoretical quantities, such as mutual information and transfer entropy, bring external bias to them, and what kind of corrections can be done to avoid these. In conclusion, we believe that this survey will fill a gap in the literature by the demonstration of the most common methods in the estimation of information-theoretic quantities in a single paper.

References:

[1] B. Gourvitch and J. Eggermont, Evaluating Information Transfer Between Auditory Cortical Neurons, J. Neurophysiology, vol. 97, no.3, pp. 2533-2543..

[2] L. Overbay and M. Todd, Dynamic system change detection using a modification of the transfer entropy, Journal of the Sound and Vibration, 322, (2009), 438-453.

[3] R. Moddemeijer, On Estimation of Entropy and Mutual Information of Continuous Distributions, Signal Processing, 1989, vol. 16, nr. 3, pp. 233-246

[4] G. Darbellay and I. Vajda, Estimation of the Information by an adaptive partitioning of the observation space, IEEE Trans. On Information Theory, Vol. 45, No.4, 1999.

[5] S. Sabesan, K. Narayanan, A. Prasad, L. Iasemidis, A. Spanias and K. Tsakalis, Information Flow in Coupled Nonlinear Systems: Application to the Epileptic Human Brain, Data Mining in Biomedicine, Data Mining in Biomedicine, 2007.

Key Words: Information-theoretic quantities, Entropy, Mutual Information, Transfer Entropy, Information Theory

Maximum Joint Entropy and Information-Based Collaboration of Automated Learning Machines

N. K. Malakar[1], K. H. Knuth[2,3] and D. J. Lary[1]

[1]Department of Physics, University of Texas at Dallas
[2]Department of Physics, University at Albany, SUNY
[3]Department of Informatics, University at Albany, SUNY

We are working to develop automated intelligent agents, which can act and react as learning machines with minimal human intervention. To accomplish this, an intelligent agent is viewed as a question-asking machine, which is designed by coupling the processes of inference and inquiry to form a model-based learning unit. In order to select maximally-informative queries, the intelligent agent needs to be able to compute the relevance of a question. This is accomplished by employing the inquiry calculus, which is dual to the probability calculus, and extends information theory by explicitly requiring context. Here we consider the interaction between two question-asking intelligent agents and note that there is a potential information redundancy with respect to the two questions that the agents may choose to pose. We show that the information redundancy is minimized by maximizing the joint entropy of the questions, which simultaneously maximizes the relevance of each question while minimizing the mutual information between them. Maximum joint entropy is therefore an important principle of information-based collaboration, which enables intelligent agents to efficiently learn together.

CALIBRATING AND ALIGNING A LOW-COST VISION-INERTIAL NAVIGATION SYSTEM

Julian L. Center, Jr.[1]
(1) Autonomous Exploration Inc.
AutonomousExploration.com, jcenter@ieee.org

Abstract

The automotive and home gaming industries have driven the cost of Micro Electro Mechanical System (MEMS) accelerometers and gyros to just a few dollars. The personal computer and cell phone industries have driven the cost of relatively high-resolution camera chips to similarly low levels. These developments make it attractive to consider the development of a low-cost robot navigation system using a combination of vision and inertial sensors. Fortunately, the error characteristics of vision and inertial sensors are complementary. MEMS gyros and accelerometers can track high-speed motions, but suffer from long-term drifts that make inertial navigation using only these sensors impractical. Using Bayesian estimation methods, vision information from a stereo camera pair can be used to correct these errors, and produce an accurate navigation system that can operate indoors or in other areas where GPS or other navigation aids are unavailable [1].

To produce an accurate navigation system, we must calibrate the instruments to compensate for stable errors and develop probability models for varying errors. For example, an Inertial Measurement Unit (IMU), consisting of three-axis accelerometers and three-axis gyros, may have misalignments among the instruments that vary from unit to unit, but do not change with time. MEMS inertial components also exhibit biases that may be different every time the instrument is powered up, but constant during a run. MEMS gyros, in particular, are sensitive to temperature variations. Furthermore, all MEMS components exhibit relatively high flicker noise (pink noise). Cameras exhibit low levels of random noise, but also have image distortions due to imperfect optics that vary from camera to camera but are time invariant. Furthermore, to combine camera and IMU information, we must also know the misalignments between the cameras and the IMU coordinate frame.

In this paper, we present a variety of controlled tests and Bayesian estimation methods that can be used to calibrate and characterize both MEMS IMUs and stereo camera pairs. We also present a method for correcting for misalignments between the IMU and the cameras.

References:

[1] J. L. Center and K. H. Knuth "Bayesian Visual Odometry", MaxEnt 2010.

Shannon's measure of Information and the thermodynamic Entropy

Arieh Ben-Naim

Department of Physical Chemistry

The Hebrew University of Jerusalem

Jerusalem, Israel

Email: arieh@fh.huji.ac.il

Abstract

We start with a clear distinction between Shannon's Measure of Information and the Thermodynamic Entropy. The first is defined on any distribution, and therefore it is a very general concept. On the other hand Entropy is defined on a very special set of distributions. Next we show that the Shannon measure of Information (SMI) provides a solid and quantitative basis for the interpretation of the thermodynamic entropy. For an ideal gas the entropy measures the uncertainty in the location and momentum of a particle, as well as two corrections due to the uncertainty principle and the indistinguishability of the particles

Modeling the World by Minimizing Relative Entropy

Antoine van de Ven[1,2], Ben A.M. Schouten[1,2]
(1) Eindhoven University of Technology
(2) Fontys University of Applied Sciences

Abstract

Relative entropy and the principle of minimum relative entropy are proposed as fundamental concepts that can be used for implementing intelligent agents that build a model of the world.

An equation is derived that describes how such an agent builds an internal model of the world by minimizing relative entropy.

Let w represent the world that the agent wants to model, and let m_t represent the beliefs and model of the world by the agent at time t. Assuming that the agent learns about the world by minimum belief updating we show that the following formula will describe the change of the internal model through time

$$KLD\,(w||m_t) = KLD\,(w||m_0) - \sum_{i=1}^{t} KLD\,(m_i||m_{i-1})$$

where KLD stands for the Kullback-Leibler divergence and $KLD\,(w||m)$ can be interpreted as the difference between the world, represented by w and the internal model of the agent, given by m.

With this equation we show that the agent continually improves its model of the world. A main advantage of this equation and the derivation is that we use it to clarify, interpret and show the relations between several seemingly different concepts, interpretations, theorems and approaches from probability theory and information theory. This includes minimum belief updating and inference by minimizing relative entropy, the difference between probability distributions and quantifing (Bayesian) surprise, learning progress, learning rate and curiosity.

References:

[1] Antoine van de Ven. A minimum relative entropy principle for the brain. In Proceedings of the Ninth International Conference on Epigenetic Robotics. Lund University Cognitive Studies, 145, 2009.

[2] Antoine van de Ven and Ben A.M. Schouten, A minimum relative entropy principle for AGI, Proceedings of the Third Conference on Artificial General Intelligence, Advances in Intelligent Systems Research, March 2010, doi:10.2991/agi.2010.26

Key Words: Relative Entropy, Kullback-Leibler divergence, Bayesian Surprise

Unification of Maximum Entropy and Bayesian Inference via Plausible Reasoning

Alexis Akira Toda[1]
(1) Yale University
Email: alexisakira.toda@yale.edu
URL: http://pantheon.yale.edu/~at37/main.html

Abstract

This paper modifies Jaynes's axioms of plausible reasoning and derives the minimum relative entropy principle as well as Bayes's rule from first principles. The new axioms, which I call the *Optimum Information Principle*, can be split into two parts: information gain (continuity and monotonicity, path independence, independence from choice of unit, zero information gain for not updating) and plausible reasoning (taking into account all information, Aristotelian logic, maximum conservatism). The Optimum Information Principle is applicable whenever the decision maker is given the data and the relevant background information. Given that the maximum entropy principle and Bayesian inference are useful methods, the Optimum Information Principle is at least as useful.

References

[1] Richard T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.

[2] Edwin T. Janyes. Information theory and statistical mechanics, I. *Physical Review*, 106(4):620–630, 1957.

[3] Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, U.K., 2003. Edited by G. Larry Bretthorst.

[4] Solomon Kullback. *Information Theory and Statistics*. John Wiley & Sons, New York, 1959.

[5] Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[6] John E. Shore and Rodney W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26(1):26–37, 1980.

[7] Arnold Zellner. Optimal information processing and Bayes's theorem. *American Statistician*, 42(4):278–280, 1988.

Key Words: Axiom, Information gain, Inference

CONSISTENCY OF SEQUENCE CLASSIFICATION WITH ENTROPIC PRIORS

Francesco A. N. Palmieri, Domenico Ciuonzo
SUN - Seconda Universitá di Napoli - ITALY
francesco.palmieri@unina2.it; domenico.ciuonzo@unina2.it

Abstract

Entropic priors, proposed mainly within the context of theoretical physics and for continuos parameter spaces [1-4], represent a very promising approach to "objective" prior determination when such information is not available. Our focus in on the application of the entropic prior idea to Bayesian inference with discrete classes in the context signal processing problems [5-6]. Unfortunately, it is well known that entropic priors, when applied to sequences, may lead to eccessive spreading of the entropy as the number of samples grows. This effect is evident in repeated experiments [2] with priors becoming progressively concentrated on the class that corresponds to the largest conditional entropy. This has been recognized as a main limitation of entropic priors, even if domain-specific containments of prior entropy have been proposed [2]. In this paper we show that the spreading of the entropy may be tolerated if the posterior probabilities remain consistent. A first contribution of this work is the derivation of a condition for posterior consistency. The inequality, derived using the Asymptotic Equiripartition Property (AEP), is based on conditional entropies and KL-divergences. Essentially, unless the likelihoods are sufficiently separated (in the KL sense), entropic priors can lead to convergence to the wrong class as the sample set grows. Furthermore, we show that entropic priors can be modified to force posterior consistency by adding a constraint to joint entropy maximization. The problem is solved applying the standard Karush-Kuhn-Tucker method and an algorithm for coefficient determination is proposed. Simulations on the application of entropic priors to some stochastic sequences are included.

References:

[1] V. Majerník. Marginal probability distribution determined by the maximum entropy method. Reports on Mathematical Physics, 45(2):171 181 (2000).
[2] A. Caticha and R. Preuss. Maximum entropy and bayesian data analysis: Entropic priors. Physical Review, E70, pp. 046127:112, (2004).
[3] A. Giffin and A. Caticha. Updating probabilities with data and moments. In 27th Int. Work. on Bayesian Inf. and Max. Ent. Meth. in Sc. and Eng., volume 954, pages 7484. AIP (2007).
[4] T. Neumann. Bayesian inference featuring entropic priors. In 27th Int. Work. on Bayesian Inf. and Max. Ent. Meth. in Sc. and Eng., volume 954, pages 283292. AIP, 2007.
[5] F. A. N. Palmieri and D. Ciuonzo. Objective Priors from Maximum Entropy in Data Classification, T.R. DII-SUN, tlc.dii.unina2.it/tlc/index.php/research/bayesian-inference, Feb 4 (2011), submitted.
[6] D. Ciuonzo and F. A. N. Palmieri. Entropic Priors for Hidden-Markov Model Classification. To appear on Proc. of IEEE Statistical Signal Processing Workshop (SSP2011), Nice, June 28-30 (2011).

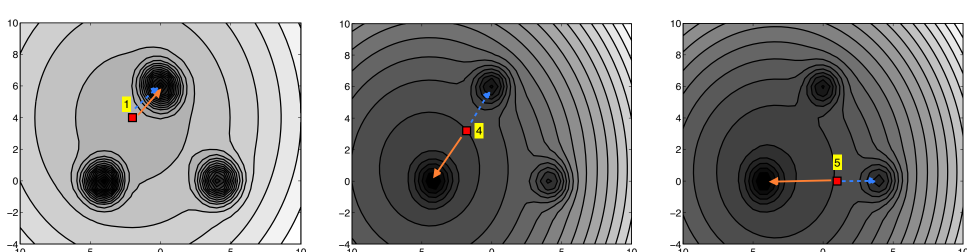
Key Words: Bayesian inference, Entropic priors, Model uncertainty

Bayesian Inference in Kernel Feature Space

Kai Krajsek and Hanno Scharr
Forschungszentrum Jülich GmbH, IBG-2

Abstract

We present a framework for Bayesian estimating in kernel feature space with implicit statistical inference in a high or even infinite dimensional feature space. Kernel based methods have a long tradition in statistical science and since its first steps in the midst of the last century (Schoenberg, 1942) numerous variants have been proposed including kernel density estimation (Parzen, 1963), mean shift (Fukunaga and Hostetler, 1975), spline models (Wahba, 1990), channel smoothing (Felsberg *et al.*, 2006), support vector machines (Cortes and Vapnik, 1995), kernel PCA (Schölkopf *et al.*, 1996) and many others. Bayesian kernel based methods like Gaussian process regression (Neal, 1997) or Bayesian variants of support vector machines (Tipping, 2001) learn a function from an input to an output space. By virtue of the representer theorem (Kimeldorf and Wahba, 1971) this function is given by a finite linear combination of kernels and after learning predictions are then obtained by inserting input data into the kernels. Our Bayesian approach differs from these Bayesian approaches in that we implicitly perform statistical inference in a high (infinite) dimensional feature space. This space is related, like in kernel PCA, by a nonlinear map to the input space which consists of all entities of interests. Inference is performed by means of a Gaussian model in the feature space transforming into a non Gaussian posterior pdf in the input space. Due to the kernel trick only scalar products of elements in the input space need to be computed. We present several experiments demonstrating the merit of our approach. *E.g.* the Figure below illustrates the behavior of our approach (solid arrow) and the kernel PCA (dashed arrow) with respect to a 2D classification problem. Training data consists of three noisy sources and test data of noise data from the lower left source. The first sample is wrongly classified by both methods. Due to updating of the posterior pdf (shown by its contour lines) sample '4', '5' (and later samples) are correctly classified by our approach whereas kernel PCA fails.



MAXIMUM ENTROPY APPROACH FOR NON-SUPERVISED PARAMETERIZATION OF INTRACELLULAR VESICLE TRACKING ALGORITHM

Yannis Kalaidizis [1,2], Marino Zerial[1]
(kalaidzi@mpi-cbg.de)
(1) Max Planck Institute Molecular Cell Biology and Genetics, Dresden, Germany,
(2) Belozersky Institute Phys. and Chem. Biology, Moscow State University, Moscow, Russia,

Abstract

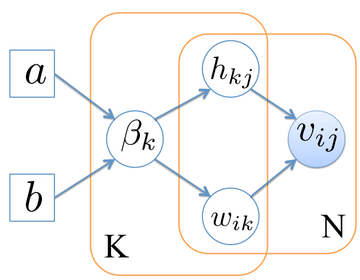
The tracking intracellular vesicle (for example, endosomes) in living cells poses challenging technical difficulties. Following individual particles in a crowded noisy environment requires robust algorithms for particle detection as well as their assignment to tracks. Phototoxicity force put low illumination intensity and results in low signal-to-noise ratio in recorded movies. Given that endosome intensity varies over 2 orders of magnitude and low signal-to-noise ratio, endocytic vesicles detection becomes a challenging task. Vesicle to track assignment, on the other hand, is challenging due to the fast movement of the endosomes, which frequently move more than their own size between successive frames. In addition, the sampling rate often exceeds the characteristic switching time of endosome processive movement and speed. An individual endosomes fission, generation *de novo*, fusion and changing identity (conversion) between two sequential frames of movie makes number of tracked objects variable in time. At the same time, object detection algorithms could produces as false positive as false negative outputs, i.e. count fluctuation of background intensity as a vesicles and loose real vesicles in noisy environment. Altogether it causes complicated and time consuming manual parameterization of tracking algorithms to achieve reasonable quality of tracking. In current work the new Maximum-Entropy based approach is developed to estimate the key parameters of tracking algorithm (Rink et al, 2005) from non-tracked objects joint distributions.

Soft partitioning in networks via Bayesian Nonnegative Matrix Factorization

I. Psorakis[1,2], S. Roberts[1], M. Ebden[1], B. Sheldon[2]
(1) Department of Engineering Science, University of Oxford
(2) Department of Zoology, University of Oxford
(mebden@robots.ox.ac.uk)

Abstract

In this work we present an approach to community detection that utilizes a Bayesian non-negative matrix factorization model to extract overlapping modules from a network.[1] The scheme has the advantage of soft-partitioning solutions, assignment of node participation scores to modules, and an intuitive foundation. We present the performance of the method against benchmark problems and compare it to other algorithms for community detection.



Above is our assumed generative graphical model. The observed variable v_{ij} denotes the nonnegative count of interactions between two individuals i, j in a weighted undirected network with adjacency matrix $\mathbf{V} \in \mathbb{R}_+^{N \times N}$, which we factorize as $\mathbf{V} = \mathbf{W}\mathbf{H}$. In the community detection context, we assume that there are a number K of 'hidden' classes of nodes in the network that affect v_{ij}. Thus we can define allocations of nodes to communities as latent (unobserved) variables that allow us to explain the increased interaction density in certain regions of the network: the more two individuals interact, the more likely they are to belong to the same communities, and vice versa. The components $\mathbf{W}$ and $\mathbf{H}$ have scale hyperparameters β_k, and fixed hyper-hyperparameters a, b.

References:

[1] I. Psorakis, S.J. Roberts, M. Ebden, and B. Sheldon. Overlapping Community Detection using Bayesian Nonnegative Matrix Factorization, *Phys. Rev. E* 2011 (to appear).

Key Words: graphs, nonnegative matrix factorization, community detection

Prawns and Probability: Adventures in Learning Models for Collective Animal Behaviour

R. P. Mann[1], A. Perna[1], D. Strömbom[1],
D. J. T. Sumpter[1], R. Garnett[2], J. E. Herbert-Read[3] and A. J. W. Ward[3]

(1) Department of Mathematics, Uppsala University, Uppsala, Sweden
(2) Robotics Institute, Carnegie Mellon University, Pittsburgh, U.S.A.
(3) School of Biological Sciences, University of Sydney, Sydney, Australia
(email: rmann@math.uu.se)

Abstract

The complex and highly dynamic motion of large animal groups are among the most impressive of natural phenomena. Simulation studies of collective animal behaviour models have repeatedly demonstrated that the much of the large-scale dynamics of such groups can be reproduced from simple local interactions between neighbours within a flock, school, or swarm. The success of such models in producing apparently realistic behaviour, the proliferation of candidate interaction rules and the advent of high-quality tracking data naturally suggests the introduction of rigorous model selection.

We find that large-scale dynamics are insufficient to distinguish between many competing models. Fine-scale recordings of animal movements can be shown to overcome this problem in principle. However, as we show, current models are insufficient to capture the interactions and decisions of real animals, particularly due to the Markovian assumption of simulation studies and the highly structured parametric form of such models.

The complexity of real animals thus poses a strong modelling challenge. We demonstrate that classic models fail to accurately capture the fine-scale interactions between animals and explore how these limitations may be overcome through the inclusion of non-Markovian memory effects, changes to experimental design, and the appropriate use of non-parametric modelling to determine spatial and temporal variables of interest.

Key Words: Animal, Collective Behaviour, Prawns, Swarm

ALMOST QUANTUM THEORY: CLASSICAL THEORIES WITH A STATISTICAL RESTRICTION

Robert W. Spekkens
Perimeter Institute for Theoretical Physics, Waterloo, Canada

Abstract

It is common to assert that the discovery of quantum theory overthrew our classical conception of nature. But what, precisely, was overthrown? In this talk, I demonstrate that a large part of quantum theory can be obtained from a single innovation relative to classical theories, namely, that there is a fundamental restriction on the sorts of statistical distributions over classical states that can be prepared. For both discrete and continuous-variable systems, one can formalize such a restriction using a classical version of complementarity (variables which do not commute according to the Poisson bracket cannot be jointly known) or of Heisenberg's uncertainty principle (products of variances in such variables are non-vanishing) together with a principle of entropy maximization. The toy theories that result from imposing this restriction are found to have a rich structure closely paralleling that of quantum theory and containing analogues of a wide variety of quantum phenomena such as collapse, coherent superposition, entanglement, interference, teleportation, no-cloning, and many others. The diversity and quality of these analogies provides compelling evidence for the view that quantum states are not states of reality – as many interpretations assume – but rather states of incomplete knowledge. I will also discuss the quantum phenomena that are not captured by this principle. Many on this list are found to be instances of a single phenomenon, called contextuality, which I will explain briefly. I will end with a few speculations on what conceptual innovations might underlie the latter set and what might be the origin of the statistical restriction.

References:

[1] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007)

QUANTUM THEORY FROM THE GEOMETRY OF EVOLVING PROBABILITIES

Marcel Reginatto[1], Michael J. W. Hall[2]
(1) Physikalisch-Technische Bundesanstalt,
Bundesallee 100, 38116 Braunschweig, Germany
(2) Theoretical Physics, IAS, Australian National University,
Canberra ACT 0200, Australia

Abstract

Our starting point is the space of probabilities $P(x)$, where the x are coordinates of an n-dimensional configuration space. The action of the translation group, $P(x) \rightarrow P(x-\theta)$, induces a natural Riemannian metric in the space of parameters θ, the Fisher-Rao metric of information geometry, as well as a metric in the space of the probability densities on configuration space.

Our next step is to set the probabilities in motion. To describe the dynamics of P, we enlarge the space and introduce a new field S which is canonically conjugate to P and a symplectic structure; this allows us to define Poisson brackets and Hamiltonian equations of motion. This leads to the theory of ensembles on configuration space (ECS), in which the observables are functionals of P and S which satisfy certain restrictions (see Ref. [1] for a discussion of the ECS formalism in the context of classical, quantum and mixed classical-quantum systems).

We consider the following question: Is it possible to enlarge the metric structure to define a geometry over the full space of the P and S? This is indeed possible, and the geometry that results is one that has both a metric and a symplectic form. To ensure consistency between metric and symplectic structures, we need to introduce a Kähler structure; i.e., a geometry that includes symplectic, metric and complex structures. The Kähler structure imposes a strong restriction on the possible geometries. We show that the simplest geometry that results is a flat Kähler space, and point to assumptions that select this geometry from all other possible geometries.

One remarkable feature of this construction is that the canonical complex coordinates Z of this Kähler geometry are precisely the wave functions $Z = \sqrt{P}e^{iS}$ of quantum mechanics. Furthermore, although we start with a real space, a complex structure is introduced in a natural way. We show that the the kinetic energy observable of quantum mechanics, and hence the Hamiltonian of a free particle, can be derived from the metric in a natural way using geometrical arguments. Finally, it is possible to associate a complex Hilbert space with this Kähler space, and this Hilbert space is the standard one of the quantum theory.

References:

[1] M. J. W. Hall and M. Reginatto, Phys. Rev. **A72**, 062109 (2005).

ON A DIFFERENTIAL GEOMETRIC VIEWPOINT OF JAYNES' MAXENT METHOD AND ITS QUANTUM EXTENSION

Sean Alan Ali[1], Carlo Cafaro[2], Adom Giffin[3], Cosmo Lupo[2] and Stefano Mancini[2]

(1): Department of Arts and Sciences, Albany College of Pharmacy and Health Sciences, 106 New Scotland Avenue, Albany, NY 12208, USA;
(2): School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino, Italy;
(3): Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ 08540, USA.
(e-mail: carlo.cafaro@unicam.it, carlocafaro2000@yahoo.it)

Abstract

It is known that geometric concepts can be introduced in probability theory by defining an inner product between two random variables as the expectation of the product of these variables once a probability distribution has been fixed [L. L. Campbell, "*The Relation Between Information Theory and the Differential Geometry Approach to Statistics*", Inform. Sci. **35**, 199 (1985)]. Then, considering the variation of the probability distribution (either directly or through changing parameters on which the distribution may depend) leads to the appearance of differential geometric ideas within probability calculus.

Following the above-mentioned work of Campbell, we present a classical differential geometric viewpoint of Jaynes' MaxEnt method. Furthermore, inspired by Braunstein [S. L. Braunstein, "*Geometry of quantum inference*", Phys. Lett. **A219**, 169 (1996)], we attempt to construct, in an explicit manner, a suitable differential geometric framework for quantum inference by means of Jaynes' MaxEnt quantum formalism.

Finally, having emphasized some criticisms to Jaynes' quantum MaxEnt approach, we point out additional conceptual problems and computational limitations of this differential geometric extension.

ENTROPIC DYNAMICS AND THE QUANTUM MEASUREMENT PROBLEM

David T. Johnson, Ariel Caticha
Department of Physics, University at Albany – SUNY
Albany, NY 12222, USA
(email: dave@dtjohnson.net, ariel@albany.edu)

Abstract

Since its inception, perhaps the most stubborn objection to quantum mechanics has been the measurement problem. In quantum mechanics, the wave function evolves continuously and deterministically according to the Schrödinger equation. Upon measurement, however, the wave function is said to 'collapse' into an eigenstate of the measurement device. It is only at this interface between the microscopic quantum world and the macroscopic classical world of the measuring device that the probabilistic nature of the wave function is made apparent.

The abrupt, probabilistic change in the wave function upon measurement stands in stark contrast to the evolution described by Schrödinger's equation. This apparent dichotomy has led to many proposed alterations and interpretations of quantum mechanics and even numerous alternative quantum theories. One such theory, built from principles of information, is entropic quantum dynamics.

In this paper, we examine the measurement problem from the perspective of entropic dynamics. We find that there are not two rules to describe the evolution of the wave function (one given by the Schrödinger equation and one for measurement); there is just one – inference. Specifically, we show that the method of maximum entropy leads to the Schrödinger equation when the only information relevant is that time has passed. When information of a different kind is available, the wave function is again updated with the method of maximum entropy. This kind of updating is possible in entropic dynamics because the phase of the wave function is defined purely in terms of probabilities.

The Born rule for position measurements is built into entropic dynamics as position is the only observable. For other types of measurements, Born's rule need not be postulated but can be derived as a convenient method for predicting the position of a quantum system after interacting with a measurement device.

LIKELIHOOD-RATIO CONFIDENCE INTERVALS FOR QUANTUM STATES

R. Blume-Kohout[1]
(1) Theoretical Division, Los Alamos National Laboratory
robin@blumekohout.com

Abstract

Estimation of quantum states – a.k.a. density matrices – is an essential component of experimental quantum information science. Nearly all work to date has focused on *point estimators*, but they cannot support rigorous probabilistic statements about the state – which are critical for the design of fault-tolerant quantum hardware [1]. *Confidence region estimators* (CREs) provide the necessary guarantee. I demonstrate how to construct rigorous and near-optimal CREs for quantum states, at any desired confidence level. First, I show that likelihood-ratio confidence regions (like the Feldman-Cousins construction [2] used in particle physics) are in fact minimum-volume CREs. I modify the canonical likelihood-ratio construction, replacing the parameter-dependent loglikelihood threshold with a uniform (parameter-independent) threshold. The resulting CRE improves on Feldman-Cousins in several ways: it respects the likelihood principle, it is much easier to calculate, and it removes most of the objectionable background-noise-dependence in Feldman-Cousins. Finally, I calculate the uniform threshold analytically by tightly lower-bounding the distribution of the loglikelihood ratio statistic for multinomial and Poisson count data. The end result is a simple and near-optimal prescription for assigning rigorously reliable confidence regions for quantum states, based on any form of tomographic data. Time permitting, I will discuss possible improvements to this result.

References:

[1] J. Preskill, "Fault tolerant quantum computation," in H.-K. Lo, T. Spiller, and S. Popescu (eds.), *Introduction to quantum computation and information*, World Scientific (Singapore), pp. 213-269 (1998).

[2] G. J. Feldman and R. D. Cousins, "A Unified Approach to the Classical Statistical Analysis of Small Signals," Phys. Rev. D **57**, 3873 (1998).

Key Words: Quantum state estimation, confidence intervals, confidence regions, likelihood ratios, multinomial, Poisson

New Monte Carlo Methods Based on Hamiltonian Dynamics

Radford M. Neal
University of Toronto
(radford@utstat.toronto.edu)

Abstract

Hamiltonian dynamics has been used to sample complex distributions for almost as long as the Metropolis algorithm has. Only from the 1980s, however, with the development of the HMC algorithm, has it been applied to distributions other than systems of molecules, such as in my work on Bayesian neural networks. The big advantages of HMC are its suppression of random walk behaviour, which can greatly speed exploration of the state space, and its superior scaling with dimensionality. After reviewing HMC, I will introduce two new Monte Carlo methods based on Hamiltonian dynamics. In “billiard HMC”, the quadratic kinetic energy used in standard HMC is replaced by a piecewise-constant kinetic energy function. Hamiltonian dynamics can then be simulated exactly by solving equations for the times when the trajectory “bounces” off the locations of the discontinuities in the kinetic energy. With exact dynamics, trajectories are never rejected. Furthermore, one can define the kinetic energy in a way that leads to only one state variable changing at a time, which for many distributions allows fast methods for incremental computation to be used. Standard HMC can have difficulty handling multiple modes, and lacks a way of estimating the normalizing constant for the distribution sampled. My “Hamiltonian importance sampling” method aims to address these problems, by exploiting the volume preservation property of Hamiltonian dynamics to create an importance sampling distribution that can closely approximate a complex distribution while still having an easily-computable probability density function, which is necessary for re-weighting the sample points. Several tricks are required to get this idea to work. In particular, for applications to Bayesian inference, it needs to be combined with “slice sampling” for the prior. Athough this methods works very naturally with Hamiltonian dynamics, it can also be adapted for use with other MCMC methods, such as simple Metropolis updates, and so provides a general alternative to methods such as simulated tempering and multicanonical sampling.

COMPUTING BAYES IN BIG SPACES

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

Abstract

The traditional aim of Bayesian computation has been to compute probability distributions (as a set of samples) with accuracy

$$\delta\big(\log \texttt{Evidence}\big) \leq O(1)$$

sufficient to enable models to be ranked whenever the Bayes factors are $O(1)$ or more. In large-scale problems, this ideal is increasingly unattainable except at prohibitive cost. Practicality dictates that computing cost should grow with size only linearly, or not much more. In turn, this dictates compromise with precision.

Statistical mechanics offers an analogy. Here, the number of degrees of freedom is huge, of the order of Avogadro's number ($N \sim 10^{24}$). Entropy $S = \log \Omega$ is a linear ("extensive") variable, so that a thermodynamic system with N degrees of freedom has entropy $S = O(N)$. Fluctuations tend to be $O(N^{1/2})$, so that measuring to precision $\Delta S = O(1)$ would be meaningless. It would also be impossible in practice to attain a precision of 1 part in 10^{24}. Despite this, statistical mechanics is a productive discipline.

In Bayesian calculus, the controlling variable which corresponds to degeneracy Ω is the evidence Z, and the extensive form corresponding to entropy is $\log Z$. For most purposes, when comparing models, the only differences of $\log Z$ that really matter are large,

$$\Delta\big(\log \texttt{Evidence}\big) = O(N)$$

Any difference less than $N^{1/2}$ tends to reflect the particular realisation of noise in the data rather than an important difference between the models, so that computation to the traditional $O(1)$ accuracy would be largely meaningless. It may also be impractically expensive in practice. Despite this, we require Bayesian computation to be a productive methodology, even in large spaces.

On encountering statistical mechanics, most students have a sense of intellectual vertigo at the enormous numbers involved with Ω until they learn to think logarithmically in terms of entropy S. A similar adjustment of mindset is needed for Bayesian computation whenever $N \gg 1$.

I will argue that nested sampling is the natural algorithm for the exploration and quantification of large spaces. Gradient information, where available, can be used. There is also a rather unexpected rôle for curvature, which promises a distinctive new avenue of research.

FRACTAL-BASED MEASURE APPROXIMATION WITH ENTROPY MAXIMIZATION AND SPARSITY CONSTRAINTS

D. La Torre[1], E.R. Vrscay[2]
(1) Department of Economics
University of Milan, Milan, Italy
(2) Department of Applied Mathematics
University of Waterloo, Waterloo, Ontario, Canada

Abstract

Let (X, d) denote a complete metric space. An N-map *iterated function system with probabilities* (IFSP) is a set of N contraction maps $w_i : X \to X$ with associated probabilities p_i. The IFSP, denoted as $(\mathbf{w}, \mathbf{p})$, defines a contractive *Markov operator* M, on the space of of probability measures $\mathcal{M}(X)$ equipped with the Monge-Kantorovich metric d_{MK}. The unique fixed point $\bar{\mu} = M\bar{\mu}$ is referred to as the *invariant measure* of the N-map IFSP.

Here we consider the following inverse problem: Given a target measure μ, find an IFSP $(\mathbf{w}, \mathbf{p})$ with invariant measure $\bar{\mu}$ sufficiently close to μ, i.e., $d_{MK}(\mu, \bar{\mu}) < \epsilon$. From Banach's Theorem, the problem may converted into finding an IFSP with Markov operator M that minimizes the *collage error* $d_{MK}(\mu, M\mu)$.

Nevertheless, the determination of optimal w_i and p_i is still a formidable problem. It was simplified in [1] by employing a fixed, infinite set of maps w_i satisfying a refinement condition on (X, d). This problem was then translated into a moment matching problem that becomes a quadratic programming (QP) problem in the p_i.

In this paper we extend the method developed in [1] along two different directions. First, we search for a set of probabilities p_i that not only minimizes the collage error but also maximizes the entropy of the iterated function system. Second, we include an extra term in the minimization process which takes into account the sparsity of the set of probabilities.

In our new formulations, collage error minimization can be understood as a multi-criteria problem: i.e., collage error, entropy and sparsity. We consider two different methods of solution: (i) *scalarization*, which reduces the multi-criteria program to a single-criteria program by combining all objective functions with different trade-off weights and (ii) *goal programming*, involving the minimization of the distance between each objective function and its goal. Numerical examples show how the two above methods work in practice.

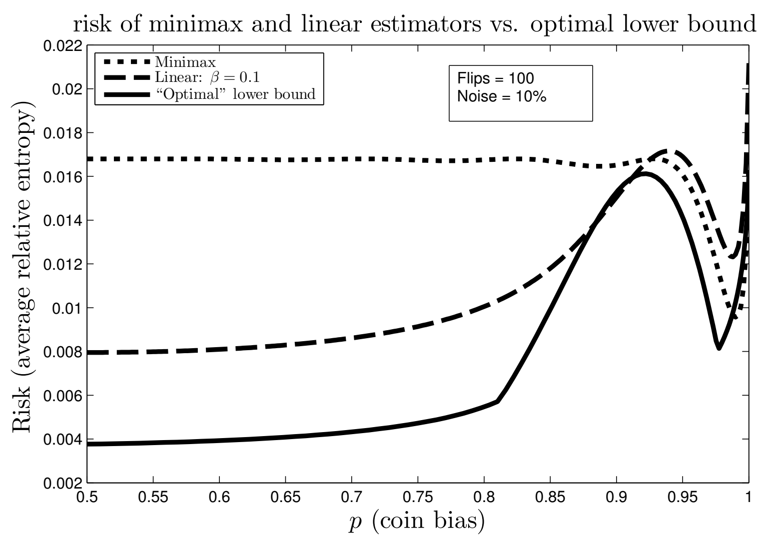
[1] B.Forte and E.R.Vrscay, Solving the inverse problem for measures using iterated function systems: a new approach, *Adv. Appl. Prob.*, 27, 800-820 (1995).

MINIMAX ESTIMATORS FOR NOISY COINS

Chris Ferrie[1], Robin Blume-Kohout[2]
(1) Inst. for Quantum Computing and Dept. of Applied Mathematics, University of Waterloo; csferrie@uwaterloo.ca
(2) Los Alamos National Laboratory; robin@blumekohout.com

Abstract

Optimal estimation of a binomial parameter from count data – e.g., guessing a coin's bias p after observing n heads in N flips – has preoccupied probabilists from Laplace to the present day. Oddly enough, little attention has been paid to the equivalent problem for *noisy* count data. Noisy count data appears in many fields, including particle physics, randomized response, and quantum information science. We analyze the performance of point estimators for "noisy coins" over a wide range of N ($N = 2 \dots 10^4$), with respect to Kullback-Leibler (relative entropy) loss. We focus particularly on minimax estimators. In stark contrast to the well-studied noiseless case, minimax estimators are counterintuitive and nonlinear, but have substantially better worst-case behavior than linear estimators of the form $\hat{p} = (n+\beta)/(N+2\beta)$. However, we also show that the minimax criterion is deeply flawed for noisy count data. Minimax estimators sacrifice tremendous accuracy at most values of p in exchange for tiny improvements around "least favorable" biases. We demonstrate *nearly*-minimax estimators with dramatically improved performance at almost all values of p.



Key Words: Minimax, Bayes Risk, relative entropy loss, Kullback-Leibler, least favorable priors

ESTIMATION AND MODEL SELECTION IN DIRICHLET REGRESSION

A. Camargo[1], J. M. Stern[1], M. S. Lauretto[2]
(1) Institute of Mathematics and Statistics, University of Sao Paulo
(2) School of Arts, Sciences and Humanities, University of Sao Paulo
andreuler@yahoo.com.br, marcelolauretto@usp.br

Abstract

Compositional data consist of vectors whose components are the proportions or percentages of some whole. The peculiarity of this family is that their sum is constrained to be some constant (usually 1). Hence, the corresponding sample space is the simplex, that is quite different from the real Euclidean space associated with unconstrained data. Therefore, attempts to apply statistical methods for unconstrained data often lead to inappropriate inference. Some statistical models for compositional data have been developed since the 70s, particularly for regression analysis.

Here we focus on the Dirichlet Covariate Model, suggested by Campbel and Mosimann (1987). In this model, one considers $\mathbf{y} = (y_1, \ldots, y_D)$ to be a $1 \times k$ positive vector having Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_D)$. In this approach, a Dirichlet regression model is readily obtained by allowing the parameters of a Dirichlet distribution to change with a covariate. For a given covariate vector $\mathbf{x} = (x_1, \ldots, x_C)$, each parameter α_j may be written as a positive-valued function $\lambda_j(\mathbf{x})$ of the covariates $\mathbf{x}$. Thus, a different Dirichlet distribution is modeled for every value of the covariates, resulting in a conditional Dirichlet distribution with $\mathbf{y}|\mathbf{x} \sim \mathcal{D}(\lambda_1(\mathbf{x}) \ldots \lambda_D(\mathbf{x}))$.

In this work, we introduce a new method for estimating the parameters of the Dirichlet Covariate Model, considering $\lambda_j(\mathbf{x})$ as polynomials. We also propose a model selection approach based on the Full Bayesian Significance Test (Pereira and Stern, 1999). Problems of interest are, for example, variable selection and choice of polynomial order.

References:

[1] G.Campbell, J.Mosimann. Multivariate methods for proportional shape. ASA Proceedings of the Section on Statistical Graphics, 10-17 (1987).

[2] R.H.Hijazi, R.W.Jernigan. Modelling Compositional Data Using Dirichlet Regression Models. Journal of Applied Probability & Statistics **4**, 77-91 (2009).

[3] C.A.B.Pereira, J.M.Stern. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. Entropy **1**, 99-110 (1999).

Key Words: Dirichlet regression, model selection

INFERENCE-BASED DESIGN OF FIR FILTERS WITH SUM OF SIGNED POWER-OF-TWO COEFFICIENTS

P. M. Goggans, C.-Y. Chan
University of Mississippi, Electrical Engineering Department
(goggans@olemiss.edu, cychan1@olemiss.edu)

Abstract

The Bayesian inference framework for design has been applied to design linear-phase finite impulse-response (FIR) filters with continuous valued coefficients [1]. In the filter implementation, the continuous valued coefficients are represented by floating or fixed point numbers. Here, we extend the inference-based approach to the design of FIR filters with coefficients expressed as a sum of signed power-of-two (SPoT) terms. The use of SPoT coefficients reduces the implementation cost and power dissipation of the filter, because the multipliers that are used to implement a filter with continuous valued coefficients are replaced by shifters and adders. These advantages are achieved at the expense of a design process with higher complexity, which arises as a result of quantization of filter coefficients into the SPoT space.

In the literature, the design of a FIR filter with SPoT coefficients is commonly formulated as a non-linear optimization problem in a discrete space. The optimization approach has a prominent drawback in that one or more design parameters governing the complexity of a design such as the number of SPoT terms and filter taps used, and the filter length are fixed in the design process. This predetermination of design parameters is very likely to result in filter designs with design complexity higher than required by the design specifications. In contrast, the inference framework has the ability to incorporate all design parameters in a design process and to design filters with design complexity appropriate to the design requirements.

The use of SPoT terms to represent the coefficients results in non-uniformly quantized possible values for the coefficients. In addition, the representation of a particular coefficient is not unique and can be achieved using a vast number of combinations. This redundancy is undesirable and gives rise to a number of difficulties in the inference-based design process. To overcome the problems resulting from redundancy, we have devised and implemented a penalization scheme, which is incorporated into the likelihood.

References:

[1] C.-Y. Chan and P. M. Goggans, Using Bayesian inference for linear phase log FIR filter design, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, P. M. Goggans and C. Y. Chan, Eds., Oxford, Mississippi: American Institute of Physics, 2009, pp. 329 - 335.

Bayesian Inference for Acoustic Impedance Boundaries in Room-Acoustic Finite Difference Time-Domain Modeling

Jonathan Botts, and Ning Xiang
Graduate Program in Architectural Acoustics,
Rensselaer Polytechnic Institute, Troy, New York
(bottsj@rpi.edu)

Abstract

In room acoustics, the finite difference time-domain approach is increasingly being applied to model wave propagation in spaces with complex geometries. For realistic simulation, implementation of frequency-dependent, impedance boundary conditions, is necessary. This paper will demonstrate that the modeling and implementation of acoustic impedance boundaries within the finite difference time-domain approach represents tasks which can be solved by two levels of Bayesian inference. The impedance function is expressed as a parametric model with coefficients of finite order, and the order of the model is directly connected to the accuracy of the calculation, computational expense, and memory requirements. The implicit Occam's razor for boundary impedance model selection, followed by coefficient estimation within the Bayesian framework, can be applied for optimizing computational expense and accuracy achieved in room-acoustic finite difference time-domain modeling.

Key Words: Model selection, parameter estimations, room-acoustic modeling, FDTD, acoustic impedance, boundary conditions

Process-Conditioned Investing with Incomplete Information using Maximum Causal Entropy

B. Ziebart
Carnegie Mellon University
bziebart@cs.cmu.edu

Abstract

Information theory and the principle of maximum entropy serve important roles for making and quantifying growth rate optimal investment [2, 3, 1]. In this work, we extend this line of research to settings where investment outcomes depend on side information variables, $\mathbf{X}_{1:T}$, sequentially revealed from a known process, but where the relationship between $\mathbf{X}$ and investment outcomes $\mathbf{Y}_{1:T}$, are only partially known. Two measures from directed information theory—the causally conditioned probability distribution, $P(\mathbf{Y}||\mathbf{X}) = \prod_{t=1}^{T} P(Y_t|\mathbf{X}_{1:t}, \mathbf{Y}_{1:t-1})$, and the causally conditioned entropy, $H(\mathbf{Y}||\mathbf{X}) = \mathbb{E}_{P(\mathbf{Y},\mathbf{X})}[-\log P(\mathbf{Y}||\mathbf{X})]$—are relevant in this setting. Namely, when the distribution of event outcome Y_t (e.g., a horse race) in terms of sequentially revealed side information variables $\mathbf{x}$ are known and outcomes provide odds $o(y_t|\mathbf{y}_{1:t-1})$ given previous outcomes, the maximum expected investment growth rate is $\mathbb{E}_{P(\mathbf{Y},\mathbf{X})}[\log o(\mathbf{Y})] - H(\mathbf{Y}||\mathbf{X})$ [3].

More generally, the relationship between $\mathbf{X}$ and $\mathbf{Y}$ variables may not be completely understood. Instead, the distribution may be known (or assumed) to satisfy sets of constraints, $\{f_i(P(\mathbf{Y}||\mathbf{X})) = 0\}$ and $\{g_j(P(\mathbf{Y}||\mathbf{X})) \leq 0\}$, for linear functions f_i and convex functions g_j of the causally conditioned probability terms, $P(\mathbf{y}||\mathbf{x})$. By extending the principle of maximum causal entropy [4] to maximize a relative causal entropy that incorporates outcome odds, this work provides a worst-case investment growth rate guarantee for this setting. We present efficient algorithms for obtaining the corresponding worst-case growth rate optimal investment allocation using convex optimization techniques and dynamic programming.

References:

[1] P. D. Grünwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, 32:1367–1433, 2003.

[2] J. Kelly Jr. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956.

[3] H. H. Permuter, Y.-H. Kim, and T. Weissman. On directed information and gambling. In *Proc. IEEE International Symposium on Information Theory*, pages 1403–1407, 2008.

[4] B. D. Ziebart, J. A. Bagnell, and A. K. Dey. Modeling interaction via the principle of maximum causal entropy. In *Proc. International Conf. on Machine Learning*, 2010.

Key Words: Causal entropy, directed information, investing

Poster Abstracts

ON THE NOTION OF FAIR GAMES AND BERNOULLI'S CONCEPT OF MORAL EXPECTATION

Arthur H. Baraov
Hewlett-Packard Company

Abstract

It is commonly believed that Daniel Bernoulli's resolution of the St Petersburg problem provides an estimate of the 'fair' entrance fee for that game. This fundamental misunderstanding of Bernoulli's treatment of the problem didn't escape Edwin Jaynes' attention: *In one of the best known books on probability theory (Feller, 1950, p. 199), Daniel Bernoulli's resolution of the famous St Petersburg paradox is rejected without even being described, except to assure the reader that he 'tried in vain to solve it by the concept of moral expectation'. ... Reading Feller, one finds that he 'resolved' the paradox merely by defining and analyzing a different game.*

For asymmetrical games with two players, it can be quite difficult to come up with a criterion of game fairness everyone finds satisfactory. However, in the case of perfectly symmetrical games, there can be no doubt that such games are fair. Nevertheless, as we demonstrate in this paper, Bernoulli's approach to estimating the 'fair' entrance fee can yield different results for different players even in cases where asymmetry of the game is purely superficial. That shows clearly that the essence of Bernoulli's resolution of the St Petersburg paradox is not about calculating the 'fair' entrance fee at all.

One of the principal objections to Bernoulli's resolution of the St Petersburg paradox is that of justification for the logarithmic assignment of utility: the choice of the logarithm doesn't seem to follow from first principles of probability theory. Indeed, why the utility of amount of money, M, or the 'moral value' as Bernoulli called it, should be taken proportional to $log(M)$? Why the utility is not to be taken proportional to $(M-1)^{1/3}$, or $arctan(M-1)$, which, after all, has an additional advantage of being bounded? The choice of the logarithmic assignment of utility appears as an ad-hoc device with all earmarks of such devices.

In this paper we show that Bernoulli's result can be obtained without resorting to the concept of moral expectation. Using the classical mathematical expectation in combination with arguments dictated by common sense, we obtain a functional equation for the entrance fee in the St Petersburg game. The solution of the equation yields entrance fee as a function of the total fortune of the player, which is in close agreement with that obtained by Bernoulli based on his concept of moral expectation.

On Ignorance Priors and the Principle of Indifference

J. Wang[1], R. L. Bras[2]
(1) University of California, Irvine, CA 92697, USA
(2) Georgia Institute of Technology, Atlanta, GA 30332, USA
([1]jingfenw@uci.edu, [2]rlbras@gatech.edu)

22 March 2011

Abstract

Selection of *ignorance prior* distributions in Bayesian statistics has been a controversial matter. We argue that it is caused by the ambiguity in the probability distribution defined directly on continuous parameters. In this study we show that *Laplace's Principle of Indifference* is a logical, consistent, and sufficient rule in assigning ignorance priors. We further argue that a probability density function consists of two components: a probability function invariant with re-parameterization and a *Limiting Density of Discrete Points* (LDDP). The representation of an ignorance prior is not unique due to the unlimited number of LDDPs through re-parameterization. The usefulness of the concept of LDDP is demonstrated in the resolution of the Borel-Kolmogrove paradox, the Bertrand's problem, and the classical harmonic oscillator.

Key Words: ignorance prior, principle of indifference, limiting density of discrete points

RELATING DYNAMICAL COMPLEXITY TO QUANTUM ENTANGLEMENT VIA INFORMATION GEOMETRY AND MAXIMUM RELATIVE ENTROPY METHODS

Sean Alan Ali[1], Carlo Cafaro[2], Dong-Hoon Kim[3] and Stefano Mancini[2]

(1): Department of Arts and Sciences, Albany College of Pharmacy and Health Sciences, 106 New Scotland Avenue, Albany, NY 12208, USA;
(2): School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino, Italy;
(3): Institute for the Early Universe, Ewha Womans University, Seoul 120-750, South Korea.
(e-mail: seanalanali@gmail.com, carlo.cafaro@unicam.it, carlocafaro2000@yahoo.it)

Abstract

Describing and understanding the essence of quantum entanglement and its connection to dynamical complexity is of great theoretical interest [G. Benenti and G. Casati, "*How complex is quantum motion?*", Phys. Rev **E79**, 025201 (2009)].

In the present work, using information geometry (Riemannian geometry applied to probability theory) and the Maximum relative entropy method we present an information geometric characterization of the quantum entanglement generated by an *s*-wave scattering event between two Gaussian wave-packets.

We conjecture that the scattering induced quantum entanglement between two minimum uncertainty Gaussian wave packets is a macroscopic manifestation emerging from the interaction among specific underlying microscopic statistical structures. We describe the pre and post-collision quantum dynamical scenarios involved in the scattering process by means of uncorrelated and correlated Gaussian statistical models, respectively.

This approach allows us to express the entanglement strength, quantified by the subsystem purity, in terms of scattering potential and incident particle energies. Furthermore, it enables definition of a quantity that serves to quantify the temporal duration over which the entanglement is active. Finally, our approach allows to uncover a quantitative relation between entanglement and information geometric complexity.

AN INFORMATION GEOMETRIC VIEWPOINT OF ALGORITHMS IN QUANTUM COMPUTING

Carlo Cafaro[1] and Stefano Mancini[1]

(1): School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino, Italy

(e-mail: carlo.cafaro@unicam.it, carlocafaro2000@yahoo.it)

Abstract

It is known that Grover's quantum search algorithm can be viewed as a geodesic path on the manifold of Hilbert-space rays where the notion of quantum distinguishability is quantified via the Fubini-Study metric, a gauge invariant metric on the projective Hilbert space [A. Miyake and M. Wadati, "*Geometric strategy for the optimal quantum search*", Phys. Rev. **A64**, 042317 (2001)].

By observing that a parametric quantum wavefunction induces in a natural manner a parametric density operator and by considering its square root, we introduce the concept of Wigner-Yanase metric. Such metric is one among many versions of a so-called quantum Fisher information metric, a metric on manifolds of density operators for both finite and infinite dimensional quantum systems.

We show in an explicit manner that the Wigner-Yanase metric and the Fubini-Study metric differ by a factor of four when considering pure state models. Finally, interpreting the Fubini-Study metric as a quantum version of Fisher metric, we provide an information geometric characterization of Grover's algorithm as a geodesic (shortest length curve) in the parameter space characterizing the pure state model, the manifold of the parametric density operators of pure quantum states.

Our analysis opens up new lines of investigation that may deserve some attention. For instance, it would be worth understanding the relation between computational complexity classes of quantum algorithms and the complexity of the quantum geodesic paths associated with them.

INSIGHTS INTO THE SOFTENING OF CHAOTIC STATISTICAL MODELS BY QUANTUM CONSIDERATIONS

Carlo Cafaro[1], Adom Giffin[2], Cosmo Lupo[1] and Stefano Mancini[1]

(1): School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino, Italy;
(2): Princeton Institute for the Science and Technology of Materials, Princeton University, Princeton, NJ 08540, USA.
(e-mail: agiffin@princeton.edu, carlo.cafaro@unicam.it)

Abstract

A problem of great theoretical interest is understanding how to compare quantum and classical chaos and explaining the reason why the former is weaker than the latter [L. A. Caron et *al*., "*Quantum chaos at finite temperature*", Phys. Lett. **A288**, 145 (2001)]. Indeed, it is commonly conjectured that the weakness of quantum chaos may be a consequence of the Heisenberg uncertainty relation.

It is known that a quantum description of chaos is qualitatively different from a classical description and that the later cannot simply be considered an approximation of the former. Indeed, the only trace of quantum theory which a classical description may retain is the canonical Heisenberg's uncertainty relation, namely a minimum spread of order $\hbar^n$ in the 2n-dimensional phase space.

Inspired by these considerations, we study the information geometry of a Gaussian statistical model when an additional information constraint resembling the canonical minimum uncertainty relation is introduced. We show that the chaoticity of such modified Gaussian statistical model (quantum-like model), quantified by means of the Information Geometric Entropy [C. Cafaro and S. A. Ali, "*Jacobi Fields on Statistical Manifolds of Negative Curvature*", Physica **D234**, 70 (2007)] and the Jacobi vector field intensity, is indeed softened with respect to the chaoticity of the standard Gaussian statistical model (classical-like model).

Our analysis provides evidence that the degree of chaoticity of statistical models is related to the existence of uncertainty relation-like information constraints. Our finding leads us to support the conjecture that quantum chaos is ultimately weaker than classical chaos because of Heisenberg's uncertainty relation, the most important difference between classical and quantum physics.

Bayesian Analysis of Scale-Invariant Processes

V. Nieves[1], J. Wang[1], and R. L. Bras[2]
(1) University of California, Irvine, Irvine, California, USA.
(2) Georgia Institute of Technology, Atlanta, Georgia, USA.
(vnieves@uci.edu)

Abstract

We have demonstrated that the Maximum Entropy (ME) principle in the context of Bayesian probability theory can be used to derive the probability distribution of those processes characterized by its scaling properties including multiscaling moments and geometric mean. We started from a proof-of-concept case of a power-law probability distribution, followed by the general case of multifractality aided by the wavelet representation of the cascade model. The ME formalism leads to the probability distribution of the multiscaling parameter and those of incremental multifractal processes at different scales [1,2]. Compared to other algorithms, the ME method significantly reduces the computational cost. The ME distributions have been evaluated against the empirical histograms derived from the drainage area of a river network, soil moisture and topography. This analysis supports the assertion that the ME principle is a universal and unified framework for modeling processes governed by scale-invariant laws. More importantly, the ME theory opens new possibilities of inferring information of multifractal processes beyond the scales of observation.

References:

[1] V. Nieves et al. Phys. Rev. Lett. **105**, 118701 (2010).
[2] V. Nieves et al. Proc. Natl. Acad. Sci. USA, submitted.

Key Words: Bayesian Statistics, Maximum Entropy, Scale-Invariant Laws, Multifractality, Environmental Sciences.

MAXIMUM PARTIAL ENTROPY PRINCIPLE AND PARTIAL PROBABILITY-WEIGHTED MOMENTS

J. Deng[1], M.D. Pandey[1], W.C. Xie[1]
(1) University of Waterloo, Canada
(E-mail: j7deng@uwaterloo.ca)

Abstract

Maximum entropy (MaxEnt) principle is usually used for estimating the probability density function under specified moment constraints. The density function is then integrated to obtain the cumulative distribution function, which needs to be inverted to obtain a quantile corresponding to some specified probability. In such analysis, consideration of higher order moments is important for accurate modelling of the distribution tail. There are three drawbacks for this conventional methodology: (1) Estimates of higher order (>2) moments from a small sample of data tend to be highly biased; (2) It can merely cope with problems with complete or non-censored samples; (3) Only probability weighted moments of integer orders have been utilized. These difficulties inevitably induce bias and inaccuracy of the resultant quantile estimates and therefore have been the main impediments to the application of the MaxEnt Principle in extreme quantile estimation.

This paper attempts to overcome these problems and presents a distribution free method for estimating the quantile function of a non-negative random variable using the principle of maximum partial entropy subjected to constraints of the partial probability weighted moments estimated from censored sample. The main contributions include: (1) New concepts, i.e., partial entropy, fractional partial probability weighted moments, and partial Kullback-Leibler measure are elegantly defined; (2) The principle of maximum entropy is extended to maximum partial entropy principle, which is defined on a finite interval; (3) New distribution free quantile functions are derived. Numerical analyses are performed to assess the accuracy of extreme value estimates computed from censored samples.

References: [1] M.D. Pandey. Struct. Saf. **22(1)**, 61-79 (2000).

Key Words: Partial entropy, Fractional partial probability-weighted moments, Quantile function, Extreme value analysis

PARSIMONEOUS PRIORS FOR REGRESSION COEFFICIENTS

N. van Erp[1], P. van Gelder[1]
(1) Structural Hydraulic and Probabilistic Design, TU Delft
Delft, The Netherlands

02-04-2011

Abstract

If one wishes to perform model selection for a set of competing regression models, then the obligatory priors can truely be a nuissance. If the number of parameters for the likelihood models differ, one is forced to ponder upon the most suitable bounds of the non-informative priors of the unknown parameters. Since taking a uniform prior with overly large bounds may severely punish the larger likelihood models. How different the situation when we are faced with a problem of parameter estimation. Taking a uniform non-informative prior one can let the bounds of this prior even go to infinity and the parameter estimates will be none the less because of it. Now why is it that a prior that has no discernible influence on our parameter estimations should have such a profound impact on our model selection? And really, should it? We will propose a prior that is in the spirit of Jaynes' scrupulously fair judge who, as by Jaynes' own initial description,[1], is someone "who insists that fairness in comparing models requires that each is delivering the best performance of which it is capable, by giving each the best possible prior probability for its parameters".

References:
[1] E.T. Jaynes, Probability Theory; the logic of science, Jaynes, Cambridge University Press (2003).

Key Words: Bayesian Model Selection, Proper Uniform Priors, Regression Coefficients

THE AIRCRAFT CARRIER PROBLEM

Arthur H. Baraov
Hewlett-Packard Company

Abstract

Congratulations, your name is Ernest J. King, and you are Chief of Naval Operations during World War II! Today you have a decision to make: to build one large aircraft carrier or two small ones. The total cost of building a large aircraft carrier is exactly the same as for two small carriers. Aircraft carriers are needed to accomplish a secret mission. The mission is either accomplished, or the carrier is destroyed by the enemy. The probability to accomplish the mission with a large aircraft carrier is two times higher than that with a small one. Small carriers operate independently of each other: if one is destroyed, the other still can accomplish the mission, and the risk of one small aircraft carrier to be destroyed is independent from the risk of the other. What is your decision - to build one large or two small aircraft carriers?

Uncritical and formal application of probability theory to this decision making problem leads to the following orthodox solution. Let p denote the probability to accomplish the mission with the large carrier, then the corresponding probability for the small one is $p/2$. Therefore the probability to accomplish the mission with two small carriers ought to be $1-(1-p/2)^2 = p - p^2/4$, i. e. building one large aircraft carrier is preferable because it gives a better chance to accomplish the mission.

The above solution seems mathematically impeccable, nevertheless common sense does not agree easily with reasonableness of such a decision. In this paper we argue for the opposite decision: two small aircraft carriers are more likely to succeed in accomplishing the mission. Our solution, which is based on a certain invariance principle, clearly demonstrates the extra power theory of probability acquires when interpreted as calculus of common sense and extension of logic.

BAYESIAN ESTIMATION OF DYNAMIC MATCHING FUNCTION FOR U-V ANALYSIS IN JAPAN

Koki Kyo[1], Hideo Noda[2], Genshiro Kitagawa[3]
(1) Obihiro University of Agriculture and Veterinary Medicine
Inada-cho, Obihiro, Hokkaido 080-8555, Japan
(E-mail: x.q.jiang@m2.dion.ne.jp)
(2) Yamagata University, 1-4-12 Kojirakawa-machi
Yamagata 990-8560, Japan
(3) Research Organization of Information and Systems
4-3-13 Toranomon, Minatoku, Tokyo 105-0001, Japan

Abstract

In this paper we propose a Bayesian method for analyzing the unemployment dynamics. To derive a Beveridge curve for U-V analysis, we introduce Bayesian models based on a matching function for the number of unemployed workers and the number of vacant jobs. In our framework, the efficiency of matching and the elasticities are regarded as time varying parameters. In order to construct a flexible model and obtain reasonable estimates in an under-determined estimation problem, we treat the time varying parameters as random variables and introduce a set of smoothness priors to them from the viewpoint of Bayesian approach. The model is described in a state space model form, so the parameter estimation is carried out using Kalman filter and the fixed interval smoothing. Thereby dynamic features of the labor market can be captured accurately. Focusing on the Japanese labor market for the period 1963-2010, as a practical application of the proposed Bayesian approach, we attempt to decompose the unemployment rate into two components, i.e., the cyclic and the structural components. Consequently, both of these components in the unemployment rate seem to have upward trend, when we examine the trends over the period 1963-2010.

Key Words: BAYESIAN ESTIMATION, BEVERIDGE CURVE, U-V ANALYSIS, MATCHING FUCTION, STATE SPACE MODEL, SMOOTHNESS PRIORS, TIME VARYING PARAMETERS

From Data To Constraints

S.Mukhopadhyay, E.Parzen, S.N.Lahiri
Texas A&M University, Department of Statistics
(Email: deep,eparzen,snlahiri@stat.tamu.edu)

Abstract

Jaynes' Maximum Entropy (MaxEnt) inference starts with the assumption that we have a set of known constraints over on the distribution. In statistical physics, we have a good intuition about the conserved macroscopic variables. It should not be surprising that in real world we have no idea which coordinate to use to specify the state of the system. In other words we only observe *empirical data* and we have to take decision on the constraints from the data. In an effort to circumvent this limitation, we propose a nonparametric quantile based method to extract relevant and significant facts (*sufficient statistics*) for the maximum entropy *exponential model*.

Invariably one of the primary requirements to apply MaxEnt principle : (i) How data can deliver *proper* constraints (statistics) ? (ii) How many constraints to use ?. While the second problem could be linked with the "model selection/parameter estimation problem, the first problem does not have any immediate solution. If we answer this two questions then it is straight forward to apply MaxEnt principle and derive the whole uncertainty distribution of the parameters of interest. The methods set forth in this paper, tailored particularly to address this two question for information processing. This will help the maximum entropy to go beyond the conventional "*exploratory phase*" and become a *objective complete inferential paradigm* for practitioners. Lastly, we illustrate our method with numerical data to demonstrate a unified framework for efficient representation, processing and data analysis.

Main tools:

- Quantile function, quantile density score function;
- Mid-rank transformation of the raw data and Legendre transformation;
- Our novel score function unify the continuous and discrete data type;

References:

[1] Parzen, E. (1979). Nonparametric statistical data modeling. Journal of the American Statistical Association, **74**, 105131.

[2] Parzen, E. (2004). Quantile probability and statistical data modeling. Statistical Scienc, **19**, 652662.

[3] Uffink, J. (1996) The constraint rule of the maximum entropy principle.Studies In History and Philosophy of Modern Physics, **27**, 47-79.

Key Words: Maximum entropy, mid-rank transformations, exponential model, quantile function

Application of Bayesian Non-Negative Matrix Factorization to Seismic Footstep Signals Separation

Asif Mehmood, Raju Damarla and James Sabatier
US Army Research Laboratory
(asif.mehmood1@us.army.mil)

Abstract

Systems employing seismic footstep detection are an important additional layer to perimeter protection and other security systems for numerous applications such as homeland security. This paper reports Bayesian treatment of non-negative matrix factorization (NMF) based on a normal likelihood, and derives an efficient Gibbs sampler to approximate the posterior density of the NMF factors. We discuss how the Gibbs sampler can be used for model selection by estimating the marginal likelihood, and compare with the Bayesian information criterion. NMF techniques have become widely used in audio analysis and source separation. They are typically used on a magnitude spectrogram of a signal, which means that the signal is presented as a sum of components, each having a fixed spectrum and time-varying gain. The main benefit of the non-negative spectrogram factorization techniques is their ability to decompose a complex signal automatically into objects which have a meaningful interpretation. In this paper, a Bayesian based NMF algorithm is developed and implemented on humans and horses footstep seismic data, and results are presented and compared with traditional NMF techniques.

References:

[1] M.N. Schmidt, O. Winther, and L. Hansen, Bayesian non-negative matrix factorization, International Conference on Independent Component Analysis and Signal Separation, (2009).

[2] Knuth K.H. 1998. Bayesian source separation and localization. In: A. Mohammad-Djafari (ed.), SPIE'98 Proceedings: Bayesian Inference for Inverse Problems, San Diego, July 1998, pp. 147-158.

[3] J. Sabatier, A. Ekimov, A review of human signatures in urban environment using seismic and acoustic methods, IEEE Conference on Technologies for Homeland Security, (2008).

Key Words: Bayesian Inference, Non-negative Matrix Factorization, Footstep detection, seismic signals

NEUTRON SPECTROMETRY AT HIGH-ENERGY ACCELERATOR FACILITIES: A BAYESIAN APPROACH USING ENTROPIC PRIORS

Marcel Reginatto
Physikalisch-Technische Bundesanstalt, Braunschweig, Germany
marcel.reginatto@ptb.de

Abstract

Radiation dosimetry at high-energy accelerator facilities is a difficult task due to the complexity of the radiation fields. The dominant radiation outside accelerator shielding is usually due to neutrons, and to estimate dose quantities of interest it is necessary to measure the neutron energy spectrum. Extended-range Bonner sphere spectrometers are widely used for this purpose; however, they typically provide a very limited amount of information and have poor resolving power [1]. The data analysis requires solving an inverse problem which may be handled using Bayesian parameter estimation or maximum entropy deconvolution [2].

Since the estimation of uncertainties is very important, it is desirable to analyze the data using a parameterized model of the spectrum and Bayesian parameter estimation [2]. This approach, however, has the shortcoming that the space of solutions is limited by the model, which may be too simplistic to account for all the relevant structure that may be present in the neutron energy spectrum. Therefore, one would like to allow for departures from the functional form of the model.

To achieve this, the approach presented in Refs. [1,2] has been extended to allow for an entropic prior. The Labesgue measure that is necessary to define the entropy is written in the form of a discretized energy spectrum $\{m_j\} = \{m_j(\{\lambda_a\})\}$ which depends on a set of parameters $\{\lambda_a\}$, and the entropic prior for the discretized energy spectrum $\{a_j\}$ is defined in the standard way; i.e., $\text{prob}(\{a_j\}|\{\lambda_a\}, \alpha, I) \propto \exp(\alpha S(\{a_j\}, \{m_j\}))$ where S is the Skilling expression for the entropy and α a constant. Since $\Delta a_k \equiv a_k - m_k$ is typically small, $|\Delta a_k| \ll m_k$, it is possible to introduce the approximation $\log(a_k/m_k) \approx \Delta a_k/m_k$ in the expression for S; this has advantages for the computation. The calculation is carried out using Bayesian parameter estimation, therefore the parameter α is handled using marginalization.

The method has been tested using simulated data that model measurements made in neutron fields behind shielding at high-energy accelerators.

References:

[1] M. Reginatto et al, Nuclear Technology **168**, 328332, (2009).

[2] M. Reginatto, Radiation Measurements, **44**, 692699, (2009).

Key Words: Bonner sphere spectrometry, Bayesian methods, entropic priors

A NOTE ON ANTIDATA

N. Kaufmann
University at Albany, Albany, NY
tk725472@albany.edu

Abstract

We explore some cases of the antidata phenomenon (see [1]). This is a something seen in Bayesian inference, namely, when an entropic 1-prior encodes an estimate of the unknown distribution parameters (for example the parameters of a Gaussian), this estimate sometimes *reduces* the degrees of freedom in the posterior pdf of the parameters. This is in contrast to the (natural conjugate) 0-prior, where typically the estimate with weight α acts like another α data points added to the actual n data samples, thus increasing the degrees of freedom in the posterior by α and making the 0-posterior more informative. We ask

1. When does the antidata occur?
2. When is it maximized?
3. When is antidata desirable? When should we use which prior?

We correct some previously published formulas for the entropic 1-prior. We produce graphics that allow visualisation of the relevant priors and posteriors.

References: [1] Rodríguez, Carlos C. (2006), *Antidata*, `http://omega.albany.edu:8008`. Also published in AIP Conference Proceedings, Vol. 872, pp. 161 to 178 ("MaxEnt 2006"), editted by Ali Mohammad Djafari; available at `http://proceedings.aip.org`.

Key Words: Antidata, Entropic Prior, Empirical Bayes, Bayesian Inference.

MAXENT ALTERNATIVES TO PEARSON FAMILY DISTRIBUTIONS

Barrie Stokes[1]
(1) Clinical Pharmacology University of Newcastle Australia
(barrie.stokes@newcastle.edu.au)

Abstract

In the spirit of ET Jaynes' Maximum Entropy Principle, a prior distribution conforming, say, to known moments (e.g., mean, standard deviation, skewness, kurtosis) should have the maximum entropy of all such distributions. At a previous MaxEnt conference a method of obtaining MaxEnt univariate distributions under a variety of constraints was presented[1]. The *Mathematica* function `Interpolation`, normally used with numerical data, can also operate with "semi-symbolic" data, and Lagrange Multiplier equations were solved for a set of symbolic ordinates and numerical abscissae describing the required MaxEnt probability density function.

Here we apply a developed version of this approach to finding MaxEnt distributions having prescribed β_1 (skewness squared) and β_2 (kurtosis) values, and compare the entropies of the MaxEnt distributions to the entropies of the corresponding Pearson family distributions having the same β_1 and β_2.

References:

[1] B. Stokes. Continuous MaxEnt Distribution in Mathematica: a "Parameter-Free" Approach. *MaxEnt 2009*, Oxford, Mississippi (2009).

Key Words: MaxEnt Distributions, Pearson Family Distributions, *Mathematica*

ADAPTIVE HAMILTONIAN ESTIMATION USING BAYESIAN EXPERIMENTAL DESIGN

Chris Granade[1,2], Chris Ferrie[1,3], David Cory[1,2,4,5]
(1) Institute for Quantum Computing, University of Waterloo
(2) Department of Physics, University of Waterloo
(3) Department of Applied Mathematics, University of Waterloo
(4) Department of Chemistry, University of Waterloo
(5) Perimeter Institute for Theoretical Physics, Waterloo
(cgranade@gmail.com)

Abstract

Quantum mechanics gives the most accurate description of many physical systems of interest. In turn, the most accurate characterization of a quantum device is given by its quantum mechanical model. Thus, efficient methods for the estimation of the parameters in a quantum mechanical model to the highest level of precision are of utmost importance, not only for building robust quantum technologies, but to reach new regimes of physics.

Bayesian experimental design is a methodology to ascertain the utility of a proposed experiment. Traditionally, Bayesian experimental design does not inform the choice of each individual measurement, but rather an entire experiment (e.g. it answers the question how many samples should I take?). In classical theory the measurement (the one to rule them all!) simply reveals the state of the system at that instant. By contrast, quantum theory presents with the following physical (and conceptual) barrier: no single measurement can reveal the state! Thus, the methodology of experimental design seems tailor-made for quantum theory!

Using Bayesian experimental design techniques, we have shown that for a single two-level system (the fundamental building block of a quantum device) the dynamical parameters of a model Hamiltonian can be estimated with exponentially improved accuracy over standard Fourier estimation. To achieve this, we derive an adaptive protocol which finds the optimal experiments based on previous observations. For more complicated systems (such as those with multi-body coupling and noise) the optimal experiment is computationally difficult to find. However, a course grained search protocol still vastly improves the accuracy of the final estimates.

Key Words: Experimental design, adaptive, parameter estimation, quantum, tomography

A SOFTWARE PACKAGE FOR NESTED SAMPLING.

Do Kester
SRON Netherlands Institute for Space Research,
Landleven 12, 9747 AD Groningen.
email do@sron.nl

Abstract

We present a software implementation of the Nested Sampling algorithm to calculate the evidence the data carries with respect to a model. Simultaneously samples are obtained from the posterior distribution which can be used to calculate inferences from the data given the model. The implementation is largely abstracted from specific prior distributions, likelihood functions, randomization engines and models. Models can be parametrized or not, static or dynamic. A large selection of predefined models is present in the package. Prior distributions on the parameters, likelihoods and engines are pluggable. As standard likelihoods distributions are provided for Gaussian errors, for Laplace errors, for Cauchy errors and for Poisson errors. We have uniform prior distributions, Jeffreys priors and exponential priors. The package contains 6 randomization engines, including a Galilean engine and a birth/death engine for dynamic models. The package is open to other models, priors, likelihoods and engines if need arises.

The nested sampling package is part of ESA's Herschel Interactive Processing Environment (HIPE), a data analysis system which is publicly available under the GNU Lesser General Public License.

Demo's of the package will be part of this poster.

Key Words: Nested Sampling, software

THE FULL BAYESIAN SIGNIFICANCE TEST FOR SYMMETRY IN CONTINGENCY TABLES

G.G.Bernardo[1], M.S.Lauretto[2], J.M.Stern[1]
(1) Institute of Mathematics and Statistics, University of Sao Paulo
(2) School of Arts, Sciences and Humanities, University of Sao Paulo
gustavo.bernardo@gmail.com, marcelolauretto@usp.br, jstern@ime.usp.br

Abstract

The problem of symmetry hypothesis is fundamental in statistics analysis, where the researcher must assess the existence of a certain symmetry condition. In several applications, the state of compliance, normality or health is characterized by the existence of symmetries. In these situations, the lack of symmetry is an indicator of non-compliance, abnormality or illness. The early detection of the lack of symmetry can frequently allow the repair, maintenance or simplified treatment, thus avoiding much more expensive and complex late procedures. This kind of early detection may be helpful in avoiding severe consequences, e.g the breaking of an important part in a machine during its operation. The test for symmetry in contingency tables constitutes a broad and important subarea in Statistics, and several methods have been devised for this problem. In this work we propose the Full Bayesian Significance Test (FBST) for the problems of symmetry and point-symmetry in contingency tables. FBST is an intuitive Bayesian approach which does not assign positive probabilities to zero measure sets when testing sharp hypotheses. Numerical experiments comparing FBST performance to power-divergence statistics suggest that FBST is a good alternative for problems concerning tests for symmetry in contingency tables.

References:

[1] N.A.C.Cressie, T.R.C.Read (1984). Multinomial Goodness-of-Fit Tests. J R Stat Soc B 46(3), 440-464.

[2] C.A.B.Pereira, J.M.Stern (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy,* 1, 99–110.

[3] C.A.P.Pereira, J.M.Stern, S.Wechsler (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis,* 3, 1, 79–100.

[4] T.R.C.Read, N.A.C.Cressie (1988). Goodness-of-Fit Statistics for Discrete Multivariate Data. NY: Springer-Verlag.

Key Words: Contingency tables, FBST, Power divergence, Significance Tests, Symmetry.

RELIABILITY ANALYSIS IN SERIES SYSTEMS: AN EMPIRICAL COMPARISON BETWEEN BAYESIAN AND CLASSICAL ESTIMATORS

A.S.Rodrigues[1], T.C.M.Dias[1] M.S.Lauretto[2], A.P.Campos[1]
(1) Statistics Deparment, Federal University of Sao Carlos
(2) School of Arts, Sciences and Humanities, University of Sao Paulo
polpo@ufscar.br, marcelolauretto@usp.br

April 2011

Abstract

In Reliability Analysis, coherent systems represent a most important struture. In many situations it is found that these systems are arranged in a series configuration, meaning that the system's failure is determined by the first component's failure. A problem of fundamental importance is to estimate the survival function parameters for each component, since this allow to establish policies and specifications that ensure adequate balances among system's reliability and costs. However, reliability data for series systems are usually censured, in the sense that one knowns the system failure time and the responsible component, but it is not known the survival time of the remaining (good) components. In this work, we present a brief revision on reliability analysis in series systems, and compare the performances of three estimation methods: Kaplan-Meier, maximum likelihood and Bayesian estimators. The results of simulation study suggest the maximum likelihood and Bayesian estimators as equivalent, and the Kaplan-Meier performing worse in most cases. References:

[1] E.L.Kaplan, P.Meier. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association **53**, 457-481, 1958.

[2] M.S.Lauretto, S.R.Faria Jr, B.B.Pereira, C.A.B.Pereira, J.M.Stern. The problem of separate hypotheses via mixture models. AIP Conference Proceedings **954**, 268-275, 2007.

[3] A.V.Peterson Jr. Expressing the Kaplan-Meier Estimator as a Function of Empirical Subsurvival Functions. Journal of the American Statistical Association **72(360)**, 854-858, 1977.

[4] A.Polpo, M.Coque-Jr, C.A.B.Pereira. Statistical Analysis for Weibull Distributions in Presence of Right and Left Censoring. The Proceedings of 2009 8th International Conference on Reliability, Maintainability and Safety, **1**, 219-223, 2009.

Key Words: Reliability analysis, series systems, competitive risks, censured reliability data

A HIERARCHICAL BAYESIAN METHOD FOR SYNTHETIC APERTURE RADAR IMAGE RECONSTRUCTION

S. Zhu[1,2], A. Mohammad-Djafari[1], X. Li[2], H.Q. Wang[2]
(1) Laboratoire des Signaux et Systèmes (L2S),
UMR 8506 CNRS-SUPELEC-UPS, 3 Rue Joliot Curie , 91192 Gif-Sur-Yvette, France
(2) Research Institute of Space Electronic Information,
School of Electronic Science and Engineering, NUDT, Changsha 410073, Hunan, China
(Email: {sha.zhu}{djafari}@lss.supelec.fr, lixiangnudt@gmail.com, oliverwhq@tom.com)

Abstract

Synthetic Aperture Radar (SAR) imaging involves an ill-posed inverse problem of reconstructing an image of the unknown scene (target) from partial and truncated information of its Fourier Transform (FT). Conventional deterministic SAR imaging methods based on Inverse Fourier Transform (IFT) are limited by a fundamental assumption that the unmeasured data in the Fourier domain is treated as zero. Indeed, these methods do not account for the prior knowledge of the scene.

Bayesian methods with appropriately chosen priors emerge as promising alternatives. In this paper we develop a Hierarchial Bayesian method for SAR image reconstruction with a generalized Total Variation (TV) prior. We adopt a coordinate-descent optimization method for the MAP estimation. Compared to existing quadratic constraints (equivalent to a Gaussian prior) the proposed method with the TV prior has capability of enhancing the region smoothness and preserving the edges between regions in the reconstructed SAR image. In addition, the Bayesian approach allows the adaptive estimation of the hyperparameters (regularization parameters) compared with the classical deterministic regularization method.

We will also show the relative performances of the proposed method, first on simulated data limiting ourselves on the FT model, then on a more complex forward model going from the scene to the measured data, and finally, on the real laboratory experimental data.

References:

1. M. Cetin and W.C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," IEEE Trans. Image Processing, vol.10, no.4, pp.623-631, 2001.
2. S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," IEEE Trans. Image Processing, vol.17, no.3, pp.326-339, 2008.
3. T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent developments in total variation image restoration," in Mathematical Models of Computer Vision, New York: Springer-Verlag, 2005.
4. N. Bali and A. Mohammad-Djafari, "Hierarchical Markovian Models for Joint Classification, Segmentation and Data Reduction of Hyperspectral Images," In Proc. ESANN, Belgium, Sep. 2006

Key Words: Bayesian inference, Image Reconstruction, Synthetic Aperture Radar, Total Variation, Computed Imaging

Handling Uncertainty Using Game Theory

Michele Pappalardo
DIIN Industrial Department-University of Salerno Italy
mpappalardo@unisa.it

Abstract: The belief functions are a bridge between various models handling different forms of uncertainty. Let be $\mathbb{P}(X)$ the power set of all possible subset of the universal *X. If* $X = \{a, b\}$ then $\mathbb{P}(X) = \{\varnothing, \{a\}, \{b\}, X\}$. The element of power set represent only the states in which this proposition is true. When two pieces of evidence are accepted with unknown values Dempster's rule of combination suggests a model for fusion of different degree of belief. If we analyze the entropy of combination we obtain that the entropy change without a rule. In the condition of highly conflicting the Dempster's rule manages the uncertain without lowering the vagueness. The goal of this paper is to submit the use of *Game Theory* to obtain the interpretation of the belief. Dealing with non-Bayesian sets, the Game Theory gives a tool to reduce the number of subsets representing the frame of discernment. Numerical examples are employed, and through analysis with the existing alternatives we obtain that the approach can effectively be useful.

Keywords: Uncertainty, Entropy, Probability, Belief, Game Theory, MaxEnt Principle

Information geometric foundations of quantum theory

R.P. Kostecki
Institute for Theoretical Physics, University of Warsaw,
Hoza 69, 00-681 Warszawa, Poland

Abstract

We present a new approach to the mathematical and conceptual foundations of quantum theory, grounded in information theoretic principles. Quantum theory is formulated in a new way, without any appeal neither to linear Hilbert space formalism, nor to subjective or frequentist interpretations of probability. On the mathematical level, it is based on the non-linear extension of C^*-algebraic framework using quantum information geometry in the role of kinematics and constrained maximum relative entropy principle in the role of dynamics. On the conceptual level, it is based on information theoretic reformulation of the bayesian approach equipped with a requirement of intersubjective coherence. Apart from resolving the measurement problem and other standard conceptual problems, it provides a novel method of construction of dynamical (interacting) quantum theoretic models without recourse to quantisation procedures, specifying the unique kinematics and unique dynamics directly from a given experimental information. We show that in particular cases our approach recovers lorentzian space-time, Hilbert space quantum theory, and maximum entropy non-equilibrium statistical mechanics.

SUPPER-RESOLUTION IMAGE FROM A SEQUENCE OF LOW RESOLUTION IMAGES BASED ON IMPROVED GUASSS-MARKOV-POTTS MODEL

N. CHU[1,2], A. MOHAMMAD-DJAFARI[1], N.GAC[1]
(1) Laboratoire des signaux et systmes (L2S)
UMR 8506 CNRS-SUPELEC-UPS, 3 rue Joliot-Curie, 91192, France
(2) Supported by China Scholarship Council (CSC)
(Email: {Ning.CHU}{djafari}{nicolas.gac}@lss.supelec.fr)

Abstract

The restoration of super resolution (SR) image from a sequence of low resolution (LR) images can be mainly considered in three steps: motion estimation, point spread function (PSF) estimation, as well as high resolution (HR) image restoration. Typical conventional methods[1] treated this problem in a serial process. However, the estimating errors of motions and PSF can accumulate and seriously affect the SR image restoration. Gauss-Markov-Potts model used in [2] can work much better. It applies the a prior of segmentations or contours in the HR image, which can greatly increase the redundant information for restoration. However, this method needs fix the number of segmentation, which can not adaptively achieve the optimal estimations. In this paper, we firstly apply the Bayesian approach to solve blind deconvolution by adding the proper a priors for both symmetric and some irregular PSF. Then we improve the phase correlation function[2] by accelerating the estimations of the sub-pixel displacements, the arbitrary rotations and zoom. These estimated parameters and PSF can better initialise the SR image restoration, which can both accelerate the iteration and obtain more proper estimations. Finally we improve the model of Gauss-Markov-Potts[2][3] by adding adaptive segmentation variable in order to estimate simultaneously the movements and PSF, as well as reconstruct the HR images and well preserve the segmentation and contour of SR image. Our method is tested and compared to other approaches in both simulations and practical experiments.

References:

[1] G.Roch. and F.Cham. **IEEE Trans.** Image Processing, Vol.15, no.11,(2006).
[2] F.Humb. and A.Djaf. **EURASIP.** Applied Signal Processing, ID 36971,(2006).
[3] A.Djaf. **Computer Journal.** vol.doi:10,1093/comjnl/bxn005, (2008).

Key Words: Super Resolution, Bayesian approach, sub-pixel motion estimation, blind deconvolution, Gauss-Markov-Potts model

ITERATION FREE VECTOR ORIENTATION USING MAXIMUM RELATIVE ENTROPY WITH OBSERVATIONAL PRIORS

Renaldas Urniezius[1], Adom Giffin[2]
(1) Department of Electrical Engineering,
Kaunas University of Technology, Lt-51367 Kaunas, Lithuania
(2) Princeton University, Princeton, NJ, USA 08544
(physics101@gmail.com)

Abstract

The amount of data to be measured and processed becomes extremely large in modern digital systems. As a result, the performance of the method to processing the data is of prime importance. This paper presents an iteration free method using the Maximum relative Entropy method when observed data is introduced as a Gaussian prior. It is shown that the posterior is a Gaussian function, but with updated means, thus making it different from other approaches. Such an approach is capable of performing optimization for a single sample, which makes it applicable to real time applications. A practical example is provided in the paper where updating is performed on vector coordinate variables, where the constraint is a vector length. A special case of this updating is also shown in limit as the standard deviation goes to zero in the priors (essentially making them Dirac delta functions) which leads to a vector normalization formula. In addition, comparisons with current particle filter and MCMC methods will be discussed.

References:

[1] A. Giffin and A. Caticha, Updating probabilities with data and moments, Bayesian Inference and Maximum Entropy Methods in Science and Engineering,eds. K. Knuth, A. Caticha, J. L. Center, A. Giffin, C. C. Rodriguez, AIP Conf.Proc. 954, 74 (2007) (http://arxiv.org/abs/0708.1593).

Key Words: Maximum Entropy, Vector Normalization, Localization, Orientation

MOMENTUM AND THE UNCERTAINTY RELATION IN THE ENTROPIC APPROACH TO QUANTUM THEORY

Shahid Nawaz and Ariel Caticha
Department of Physics, University at Albany – SUNY
Albany, NY 12222, USA
(e-mail: snafridi@gmail.com, ariel@albany.edu)

Abstract

In the Entropic Dynamics (ED) framework quantum theory is derived as an application of the method of maximum entropy. (We are trying to do for quantum mechanics what Jaynes did for statistical mechanics.) The basic assumption is that in addition to the particles of interest the world contains other "extra" variables whose entropy S depends on the positions x of the particles, $S = S(x)$. The Schrödinger equation follows from their coupled non-dissipative dynamics: the entropy $S(x)$ drives the dynamics of the particles x while they in their turn determine the evolution of $S(x)$. An important feature is that the phase of the wave function (and not just its magnitude) receives a statistical interpretation: the phase keeps track of the entropy $S(x)$ of those extra variables.

Entropic Dynamics differs from other information-based approaches to quantum theory in that the position observable assumes a privileged role: the particles have well-defined, albeit unknown, positions. This opens the possibility of *explaining all other observables in purely informational terms*. In this paper our specific goal is to discuss momentum and its corresponding uncertainty principle.

Since particles follow Brownian trajectories that are continuous but non differentiable it is not possible to assign an instantaneous momentum to the particles. Nevertheless, four different notions of momentum can be usefully introduced. They are not associated to the particles but rather to their probability distributions: (1) the *current momentum* is associated to the velocity with which probabilities flow; (2) the *drift momentum* reflects flow along the entropy gradient; (3) the *osmotic momentum* is associated to the velocity with which probabilities diffuse; and (4) the familiar *quantum momentum* is the generator of infinitesimal translations. We find relations among these four momenta: The expectation of the quantum momentum coincides with the expectations of the drift momentum and of the current momentum. The uncertainty in quantum momentum is the uncertainty in osmotic momentum and this leads to the Heisenberg uncertainty relation.

The main conclusion is that *momentum is a statistical concept*. Particles have a position but they do not have a momentum. In ED the momenta are not properties of the particles; they are properties of the probability distributions.

From Cox to Emergent Geometry

Newshaw Bahreyni[1], Kevin H. Knuth[1,2]
(1) Department of Physics, University at Albany, Albany NY USA
(2) Department of Informatics, University at Albany, Albany NY USA

April 18, 2011

Abstract

Knuth has demonstrated that Cox's method of generalizing Boolean algebra to the probability calculus can be extended to distributive algebras and applied to lattices in general [1][2]. More recently Goyal, Knuth and Skilling showed that Cox's ideas can be used to derive Feynman's complex path integral formulation of quantum mechanics [3]. The fundamental concept behind these derivations is based on the quantification of algebraic structures [5], and it is the order-theoretic structure of the underlying algebra that constrains any quantification leading to sets of constraint equations, which we view as laws.

We have taken these ideas further and apply them to partially ordered sets. In our previous paper [4] we showed how applying this quantification method to causally-ordered sets of events results in the scalar being expressed in a Minkowsian form. We also showed how changing the basis from one set of chains into another, under certain conditions, leads to generalized Lorentz transformations for pairs.

We now use the same technique to demonstrate that geometric concepts can be *derived* from order-theoretic concepts. We show how chains in a poset can be used to define points and line segments, and that quantification of the poset leads directly to the Pythagorean theorem as well as the law of cosines and the law of sines, angles and rotations. Thus we show that geometry itself is not fundamental, but rather emergent, through the quantification of a partially ordered set.

References:

[1] Knuth K.H. 2003. Deriving laws from ordering relations. MaxEnt 2003, 204-235.

[2] Knuth K.H. 2009. Measuring on lattices. MaxEnt2009, 132-144.

[3] Goyal P., Knuth K.H., Skilling J. 2010. Why quantum theory is complex, Phys. Rev. A 81, 022109. arXiv:0907.0909v3 [quant-ph]

[4] Knuth K.H. 2010. Information physics: The new frontier. MaxEnt 2010, 3-19. arXiv:1009.5161v1 [math-ph]

[5] Knuth K.H., Bahreyni N. 2010. The order-theoretic origin of special relativity. 115-121.

Key Words: information, partially ordered set, relativity, valuation, geometry

New copulas obtained by maximizing Tsallis or Rényi Entropies

Doriano-Boris Pougaza and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes
UMR 8506 (CNRS-SUPELEC-UNIV PARIS SUD 11)
Plateau de Moulon, 3 rue Joliot Curie,
91192 Gif-sur-Yvette Cedex, France

Abstract

Sklar [1] introduced the notion of copula and solved in this way the problem studied by Fréchet [2] and many other authors on the determination of a joint distribution function when the one dimensional marginal cumulative distributions are prescribed. The same problem arises also in the context of image (the internal density distribution of some physical or biological quantity) reconstruction in X-ray computed tomography when only two orthogonal projections are given. We consider those two previous problems which are the same problem with the restriction that the distributions has bounded support and maximizes the Tsallis-Havrda-Charvát's entropy [3][4] or the Rényi's entropy [5] by rescaling. We focus on the case that the entropy index is equal to 2. We give a theorem and its corollary using the well-known uniform transformation yielding to a genuine method for constructing new families of copulas. We give the expression of some dependence concepts and then provide many examples.

Keywords: Copula, Entropy, Joint density estimation, Shannon, Rényi and Tsallis.
PACS: 02.30.Gp,02.50.Cw, 02.50.Sk

References

[1] A. Sklar, *Fonction de répartition dont les marges sont données, Inst. Stat. Univ. Paris Publ.* **8** (1959), 229 -231.

[2] M.M. Fréchet, *Sur les tableaux de corrélations dont les marges sont données, Ann. Univ. Lyon Sec.A* **14** (1950), 53-77.

[3] J. Havrda and F. Charvát, *Quantification method of classification processes, Kybernetika* **3** (1967), 0-3.

[4] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics, Journal of statistical physics* **52** (1988), 479-487.

[5] A. Rényi, *On measures of dependence, Act. Math. Acad. Sci. Hungar* **10** (1959), 441-451.

Porous Material Parameter Estimation: A Bayesian Approach

Cameron Fackler, Eric Dieckman, and Ning Xiang
Graduate Program in Architectural Acoustics,
Rensselaer Polytechnic Institute, Troy, New York
(facklc@rpi.edu)

Abstract

A Bayesian approach to estimating the physical parameters of rigid-frame porous materials through measurements of the acoustic impedance of such materials is presented. Porous materials are widely used as sound absorbers in many industries. Modeling the effects of porous absorbers and making optimal use of porous absorbent materials requires knowledge of the physical parameters characterizing such materials: porosity, tortuosity, and flow resistivity. For many materials, direct measurement of these parameters requires time-consuming or highly sensitive procedures. Based on some existing models for the characteristic impedance of a porous material in terms of the physical material parameters, Bayesian parameter estimation is used to estimate the physical parameters of a material from a measurement of its complex acoustic impedance. In addition to estimation of the values of the physical parameters, Bayesian analysis provides information on the uncertainties and interdependence of the parameters.

Key Words: Bayesian parameter estimation, porous materials, acoustic impedance

THE ERROR IN THE TWO ENVELOPE PARADOX

Adom Giffin
Princeton University, Princeton, NJ, USA 08544
(physics101@gmail.com)

Abstract

The "paradox" arises in the Two Envelopes Paradox or Exchange Paradox from the incorrect formulation of the argument. The information given is misused and therefore the results are incorrect for the question asked. The key is to be clear on what question we are asking. We must make sure that the question that is asked is the question that is written down. Once that is done, there is no more paradox.

References:

[1] Carlos Rodriguez, Understanding Ignorance, Maximum Entropy and Bayesian Methods," in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, G. J. Erickson and C.R. Smith (eds.), Kluwer Academic Publishers. p.189 (1987)

Key Words: Expectation, exchange paradox, two envelopes, Bayes

MAXIMUM ENTROPY PRODUCTION IN DAISYWORLD MODELS

H.A. Maunu[1], K.H. Knuth[1,2]
(1) University At Albany (SUNY), Department of Physics
(2) University At Albany (SUNY), Department of Informatics
(Email: Haley.Maunu@gmail.com, hm535264@albany.edu)

Abstract

Daisyworld was first introduced in 1983 by Watson and Lovelock as a model that illustrates how life can influence a planet's climate [1]. These models typically involve modeling a planetary surface on which black and white daisies can grow thus influencing the local surface albedo and therefore also the temperature distribution. Since then, variations of daisyworld have been applied to study problems ranging from ecological systems to global climate. Much of the interest in daisyworld models is due to the fact that they enable one to study self-regulating systems. These models are nonlinear, and as such they exhibit sensitive dependence on initial conditions, and depending on the specifics of the model they can also exhibit feedback loops, oscillations, and chaotic behavior. Many daisyworld models are thermodynamic in nature in that they rely on heat flux and temperature gradients. However, what is not well-known is whether, or even why, a daisyworld model might settle into a maximum entropy production (MEP) state. With the aim to better understand these systems, we will discuss what is known about the role of MEP in daisyworld models.

References:
[1] A.J. Watson and J.E. Lovelock, Tellus **35B**, 284-289 (1983).

Key Words: Daisyworld, Maximum Entropy Production

Parameter Estimation of Magnetic Resonance Spectra via a Statistical Mechanics Approach

Keith A. Earle
Physics Department, University at Albany (SUNY)
1400 Washington Ave, Albany NY 12222
kearle@albany.edu, http://earlelab.rit.albany.edu

Abstract

In this work, I develop the connections between parameter fitting and statistical mechanics using the maxent principle of Jaynes. The approach described here leads to an entropy that is extensive in the number of measurements in the average. In addition, I show how to combine measurements from different experiments in an unbiased way in order to maximize the entropy of simultaneous parameter fitting. A pleasing physical picture emerges that interprets fit parameters as generalized coordinates, from which the conjugate forces may be derived via the system partition function. From this perspective, the parameter fitting problem may be interpreted as a process where the system (spectrum) does work against internal stresses (non-optimum model parameters) to achieve a state of minimum free energy/maximum entropy. As an example of the utility of the methods discussed here, this approach will be applied to a detailed analysis of a mixture model that arises in the analysis of multifrequency electron spin resonance spectra.

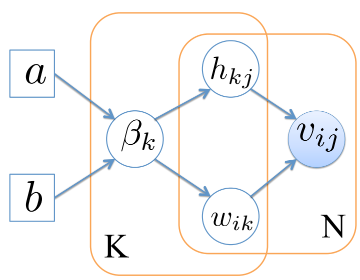
Key Words: Maximum Entropy, Parameter Optimization, Statistical Physics

Soft partitioning in networks via Bayesian Nonnegative Matrix Factorization

I. Psorakis[1,2], S. Roberts[1], M. Ebden[1], B. Sheldon[2]
(1) Department of Engineering Science, University of Oxford
(2) Department of Zoology, University of Oxford
(mebden@robots.ox.ac.uk)

Abstract

In this work we present an approach to community detection that utilizes a Bayesian nonnegative matrix factorization model to extract overlapping modules from a network.[1] The scheme has the advantage of soft-partitioning solutions, assignment of node participation scores to modules, and an intuitive foundation. We present the performance of the method against benchmark problems and compare it to other algorithms for community detection.



Above is our assumed generative graphical model. The observed variable v_{ij} denotes the nonnegative count of interactions between two individuals i, j in a weighted undirected network with adjacency matrix $\mathbf{V} \in \mathbb{R}_+^{N\times N}$, which we factorize as $\mathbf{V} = \mathbf{WH}$. In the community detection context, we assume that there are a number K of 'hidden' classes of nodes in the network that affect v_{ij}. Thus we can define allocations of nodes to communities as latent (unobserved) variables that allow us to explain the increased interaction density in certain regions of the network: the more two individuals interact, the more likely they are to belong to the same communities, and vice versa. The components $\mathbf{W}$ and $\mathbf{H}$ have scale hyperparameters β_k, and fixed hyper-hyperparameters a, b.

References:

[1] I. Psorakis, S.J. Roberts, M. Ebden, and B. Sheldon. Overlapping Community Detection using Bayesian Nonnegative Matrix Factorization, *Phys. Rev. E* 2011 (to appear).

Key Words: graphs, nonnegative matrix factorization, community detection

	10th SUNDAY	11th MONDAY	12th TUESDAY	13th WEDNESDAY	14th THURSDAY	15th FRIDAY
7:30 – 8:00		***Registration***				
8:00 – 8:30		**Welcome**	***Registration***	***Registration***	***Registration***	***Registration***
8:30 – 9:00	***Registration***	**Invited Speaker** J. Uffink 23	**Invited Speaker** R. D. Lorenz 43	**Invited Speaker** A. Ben-Naïm 59	**Invited Speaker** R. W. Spekkens 75	E. R. Vrscay 89
9:00 – 9:30						C. Ferrie 91
9:30 – 10:00	**Tutorial** G. Tenti 13	K. H. Knuth 25	J. Wang 45	A. van de Ven 61	M. Reginatto 77	M. S. Lauretto 93
10:00 – 10:30		K. A. Earle 27	B. L. Ruddell 47	A. A. Toda 63	C. Cafaro 79	Break
10:30 – 11:00	Break	Break	Break	Break	Break	P. M. Goggans 95
11:00 – 11:30	**Tutorial** A. Caticha 15	F. Mendes 29	**Invited Speaker** T. E. Jupp 49	F. Palmieri 65	A. Caticha 81	J. Botts 97
11:30 – 12:00		U. von Toussaint 31		K. Krajsek 67	R. J. Blume-Kohout 83	B. D. Ziebart 99
12:00 – 12:30	***Lunch***	***Lunch***	***Lunch***	***Lunch***	***Lunch***	***Lunch***
12:30 – 1:00						
1:00 – 1:30						
1:30 – 2:00	**Tutorial** A. Giffin 17	**Invited Speaker** G. H. Pollack 33			**Invited Speaker** R. M. Neal 85	
2:00 – 2:30			**Invited Speaker** R. K. Niven 51	Y. Kalaidizis 69		***Business Meeting***
2:30 – 3:00	Break			M. Ebden 71	**Invited Speaker** J. Skilling 87	
3:00 – 3:30	**Tutorial** P. Goyal 19	Break	D. Gencaga 53	R. P. Mann 73		
3:30 – 4:00		J. Dettmer 35	Break	Break		
4:00 – 4:30	Break	M. A. Mubeen 37	N. K. Malakar 55	**Panel Discussion**		
4:30 – 5:00	**Tutorial** U. von Toussaint 21	M. Betancourt 39	J. L. Center 57		***Banquet***	
5:00 – 5:30		P.G.L. Porta Mana 41				
5:30 – 6:00						
6:00 – 6:30			***Poster Session (6-10pm)*** 102–133			
6:30 – 7:00	***Welcome Reception***					
7:00 – 7:30						
7:30 – 8:00						
8:00 – 8:30						